



Cloudera Enterprise 6 Release Guide

Important Notice

© 2010-2019 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, and any other product or service names or slogans contained in this document are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder. If this documentation includes code, including but not limited to, code examples, Cloudera makes this available to you under the terms of the Apache License, Version 2.0, including any required notices. A copy of the Apache License Version 2.0, including any notices, is included herein. A copy of the Apache License Version 2.0 can also be found here: <https://opensource.org/licenses/Apache-2.0>

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners. Reference to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property. For information about patents covering Cloudera products, see <http://tiny.cloudera.com/patents>.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

Cloudera, Inc.
395 Page Mill Road
Palo Alto, CA 94306
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-362-0488
www.cloudera.com

Release Information

Version: Cloudera Enterprise 6.x
Date: April 9, 2019

Table of Contents

Cloudera Enterprise 6 Release Guide.....	5
Cloudera Enterprise 6 Requirements and Supported Versions.....	5
<i>Hardware Requirements.....</i>	5
<i>Operating System Requirements.....</i>	20
<i>Database Requirements.....</i>	23
<i>Java Requirements.....</i>	26
<i>Networking and Security Requirements.....</i>	27
<i>Data at Rest Encryption Requirements.....</i>	31
<i>Browser Requirements.....</i>	33
<i>Supported Configurations with Virtualization and Cloud Platforms.....</i>	33
<i>Product Compatibility Matrices.....</i>	34
Version, Packaging, and Download Information.....	44
<i>Cloudera Manager 6 Version and Download Information.....</i>	45
<i>CDH 6 Version, Packaging, and Download Information.....</i>	47
<i>Cloudera Navigator 6 Encryption Version and Download Information.....</i>	123
What's New in Cloudera Documentation.....	128
<i>What's New in Cloudera Documentation in March, 2019.....</i>	128
<i>What's New in Cloudera Documentation in February, 2019.....</i>	129
<i>What's New in Cloudera Documentation in January, 2019.....</i>	129
<i>What's New in Cloudera Documentation in December, 2018.....</i>	130
<i>What's New in Cloudera Documentation in November, 2018.....</i>	130
Install and Upgrade Notes.....	131
<i>Upgrades to Cloudera Enterprise 6.x.....</i>	131
<i>TLS Protocol Error with OpenJDK.....</i>	135
<i>Upgrades to Cloudera Enterprise 5.x.....</i>	135
Cloudera Manager 6 Release Notes.....	136
<i>Cloudera Manager 6.2.x Release Notes.....</i>	137
<i>Cloudera Manager 6.1.x Release Notes.....</i>	146
<i>Cloudera Manager 6.0.x Release Notes.....</i>	161
<i>Stale Configurations.....</i>	175
CDH 6 Release Notes.....	177
<i>CDH 6.2.x Release Notes.....</i>	177
<i>CDH 6.1.x Release Notes.....</i>	265
<i>CDH 6.0.x Release Notes.....</i>	431
Cloudera Navigator 6 Data Management Release Notes.....	593
<i>Cloudera Navigator 6.2.0 Data Management Release Notes.....</i>	593
<i>Cloudera Navigator 6.1.x Data Management Release Notes.....</i>	602
<i>Cloudera Navigator 6.0.x Data Management Release Notes.....</i>	622

Cloudera Navigator 6 Encryption Release Notes.....	641
<i>Cloudera Navigator 6.2.x Encryption Release Notes.....</i>	642
<i>Cloudera Navigator 6.1.x Encryption Release Notes.....</i>	648
<i>Cloudera Navigator 6.0.x Encryption Release Notes.....</i>	658
Deprecated Items.....	667
<i>Deprecated Items.....</i>	667
<i>Removed Items.....</i>	668

Appendix: Apache License, Version 2.0.....672

Cloudera Enterprise 6 Release Guide

This guide contains release and download information for installers and administrators. It includes release notes as well as information about installation requirements, supported platforms, and version and download information. The guide also provides a release matrix that shows which major and minor release version of a product is supported with which release version of Cloudera Manager and CDH.

Cloudera Enterprise 6 Requirements and Supported Versions

In an enterprise data hub, Cloudera Manager and CDH interact with several products such as Apache Accumulo, Apache Impala, Hue, Cloudera Search, and Cloudera Navigator. This guide provides information about which major and minor release version of a product is supported with which release version of CDH and Cloudera Manager.

Compatibility across different release versions of Cloudera Manager and CDH must be taken into account, especially when carrying out install/upgrade procedures.

Hardware Requirements

To assess the hardware and resource allocations for your cluster, you need to analyze the types of workloads you want to run on your cluster, and the CDH components you will be using to run these workloads. You should also consider the size of data to be stored and processed, the frequency of workloads, the number of concurrent jobs that need to run, and the speed required for your applications.

As you create the architecture of your cluster, you will need to allocate Cloudera Manager and CDH roles among the hosts in the cluster to maximize your use of resources. Cloudera provides some guidelines about how to assign roles to cluster hosts. See [Recommended Cluster Hosts and Role Distribution](#). When multiple roles are assigned to hosts, add together the total resource requirements (memory, CPUs, disk) for each role on a host to determine the required hardware.

For information about how workloads affect sizing decisions, see the following blog post: [How-to: Select the Right Hardware for Your New Hadoop Cluster](#).

For more information about sizing for a particular component, see the following minimum requirements:

Cloudera Manager

Cloudera Manager Server Storage Requirements

Component	Storage	Notes
Partition hosting /usr	1 GB	
Partition hosting /var	5 GB to 1 TB	Scales according to number of nodes managed. See table below.
Partition hosting /opt	15 GB minimum	Usage grows as the number of parcels downloaded increases.
Cloudera Manager Database Server	5 GB	If the Cloudera Manager Database is shared with the Service Monitor and Host Monitor, more storage space is required to meet the requirements for those components.

Host Based Cloudera Manager Server Requirements

Number of Cluster Hosts	Database Host Configuration	Heap Size	Logical Processors	Cloudera Manager Server /var Directory
Very small (≤ 10)	Shared	2 GB	4	5 GB
Small (≤ 20)	Shared	4 GB	6	20 GB minimum
Medium (≤ 200)	Dedicated	8 GB	6	200 GB minimum
Large (≤ 500)	Dedicated	10 GB	8	500 GB minimum
Extra Large (> 500)	Dedicated	16 GB	16	1 TB minimum



Note: On smaller clusters, the Cloudera Manager Server and Database can share a host. On larger clusters, they must run on separate dedicated hosts.

Service Monitor Requirements

The requirements for the Service Monitor are based on the number of monitored entities. To see the number of monitored entities, perform the following steps:

1. Open the Cloudera Manager Admin Console and click **Clusters > Cloudera Management Service**.
2. Find the **Cloudera Management Service Monitored Entities** chart. If the chart does not exist, add it from the **Chart Library**.

For more information about Cloudera Manager entities, see [Cloudera Manager Entity Types](#).

Clusters with HDFS, YARN, or Impala

Use the recommendations in this table for clusters where the only services with worker roles are HDFS, YARN, or Impala.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-2,000	0-100	1 GB	6 GB
2,000-4,000	100-200	1.5 GB	6 GB
4,000-8,000	200-400	1.5 GB	12 GB
8,000-16,000	400-800	2.5 GB	12 GB
16,000-20,000	800-1,000	3.5 GB	12 GB

Clusters with HBase, Solr, Kafka, or Kudu

Use these recommendations when services such as HBase, Solr, Kafka, or Kudu are deployed in the cluster. These services typically have larger quantities of monitored entities.

Number of Monitored Entities	Number of Hosts	Required Java Heap Size	Recommended Non-Java Heap Size
0-30,000	0-100	2 GB	12 GB
30,000-60,000	100-200	3 GB	12 GB
60,000-120,000	200-400	3.5 GB	12 GB
120,000-240,000	400-800	8 GB	20 GB

Host Monitor

The requirements for the Host Monitor are based on the number of monitored entities.

To see the number of monitored entities, perform the following steps:

1. Open the Cloudera Manager Admin Console and click **Clusters > Cloudera Management Service**.
2. Find the **Cloudera Management Service Monitored Entities** chart. If the chart does not exist, add it from the **Chart Library**.

For more information about Cloudera Manager entities, see [Cloudera Manager Entity Types](#).

For more information about Cloudera Manager entities, see [#unique_11](#).

Number of Hosts	Number of Monitored Entities	Heap Size	Non-Java Heap Size
0-200	<6k	1 GB	2 GB
200-800	6k-24k	2 GB	6 GB
800-1000	24k-30k	3 GB	6 GB

Ensure that you have at least 25 GB of disk space available for the Host Monitor, Service Monitor, Reports Manager, and Events Server databases.

For more information refer to [Host Monitor and Service Monitor Memory Configuration](#).

Reports Manager

The Reports Manager fetches the fsimage from the NameNode at regular intervals. It reads the fsimage and creates a Lucene index for it. To improve the indexing performance, Cloudera recommends provisioning a host as powerful as possible and dedicating an SSD disk to the Reports Manager.

Table 1: Reports Manager

Component	Java Heap	CPU	Disk
Reports Manager	3-4 times the size of the fsimage.	<ul style="list-style-type: none"> • Minimum: 8 cores • Recommended: 16 cores (32 cores, with hyperthreading enabled.) 	1 dedicated disk that is at least 20 times the size of the fsimage. Cloudera strongly recommends using SSD disks.

Agent Hosts

An unpacked parcel requires approximately three times the space of the packed parcel that is stored on the Cloudera Manager Server.

Component	Storage	Notes
Partition hosting /opt	15 GB minimum	Usage grows as new parcels are downloaded to cluster hosts.
/var/log	2 GB per role	Each role running on the host will need at least 2 GB of disk space.

Event Server

The following table lists the minimum requirements for the Event Server:

CPU	RAM	Storage
1 core	256 MB	<ul style="list-style-type: none"> • 5 GB for the Event Database

CPU	RAM	Storage
		<ul style="list-style-type: none"> 20 GB for the Event Server Index Directory. The location of this directory is set by the Event Server Index Directory Event Server configuration property.

Alert Publisher

The following table lists the minimum requirements for the Alert Publisher:

CPU	RAM	Storage
1 core	1 GB	Minimum of 1 disk for log files

Cloudera Navigator

The sizing of Navigator components varies heavily depending on the size of the cluster and the number of audit events generated. Refer to [Minimum Recommended Memory and Disk Space](#) for more information.

Table 2: Cloudera Navigator

Component	Java Heap / Memory	CPU	Disk
Navigator Audit Server	<p>Minimum: 2-3 GB of Java Heap</p> <p>Configure this value using the Java Heap Size of Auditing Server in Bytes configuration property.</p>	Minimum: 1 core	<p>The database used by the Navigator Audit Server must be able to accommodate hundreds of gigabytes (or tens of millions of rows per day). The database size may reach a terabyte.</p> <p>Ideally, the database should not be shared with other services because the audit insertion rate can overwhelm the database server making other services using same database less responsive.</p> <p>See Storage Space Planning for Cloudera Navigator.</p>
Navigator Metadata Server	<ul style="list-style-type: none"> Minimum: 10 GB of Java Heap Recommended: 20 GB Java Heap 	Minimum: 1 core	<ul style="list-style-type: none"> Minimum: 50 GB Recommended: 100-200 GB

Component	Java Heap / Memory	CPU	Disk
	<p>Add 20 GB for operating system buffer cache, however memory requirements can be much higher on a busy cluster and could require provisioning a dedicated host. Navigator logs include estimates based on the number of objects it is tracking.</p> <p>Configure this value using the Java Heap Size of Navigator Metadata Server in Bytes configuration property.</p>		 Note: Data stored by the Metadata server grows indefinitely unless you run the purge function. If you have not set up the purge function to run at a scheduled interval, Cloudera recommends that you run the purge function to reclaim disk space and keep data growth (and the corresponding memory requirement) in check. See Administration (Navigator Console) .

Cloudera Data Science Workbench

Hardware Component	Requirement	Notes
CPU	16+ CPU (vCPU) cores	Allocate at least 1 CPU core per session. 1 CPU core is often adequate for light workloads.
Memory	32 GB RAM	<ul style="list-style-type: none"> As a general guideline, Cloudera recommends nodes with RAM between 60GB and 256GB Allocating less than 2 GB of RAM can lead to out-of-memory errors for many applications.
Disk	<ul style="list-style-type: none"> Root Volume: 100 GB Application Block Device or Mount Point (Master Host Only): 1 TB Docker Image Block Device: 1 TB 	SSDs are strongly recommended for application data storage.

For more information on scaling guidelines and storage requirements for cloud providers such as AWS and Azure, see [Requirements and Supported Platforms](#) in the Cloudera Data Science Workbench documentation.

CDH Accumulo

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Flume

Component	Java Heap	CPU	Disk
Flume	<ul style="list-style-type: none"> Minimum: 1 GB Maximum 4 GB Java Heap size should be greater than the maximum channel capacity <p>Set this value using the Java Heap Size of Agent in Bytes Flume configuration property. See Flume Memory Consumption</p>	Calculate the number of cores using the following formula: $(\text{Number of sources} + \text{Number of sinks}) / 2$	Multiple disks are recommended for file channels, either a JBOD setup or RAID10 (preferred due to increased reliability).

HDFS

Table 3: HDFS

Component	Memory	CPU	Disk
JournalNode	1 GB (default) Set this value using the Java Heap Size of JournalNode in Bytes HDFS configuration property.	1 core minimum	1 dedicated disk
NameNode	<ul style="list-style-type: none"> Minimum: 1 GB (for proof-of-concept deployments) Add an additional 1 GB for each additional 1,000,000 blocks <p>Snapshots and encryption can increase the required heap memory.</p> <p>See Sizing NameNode Heap Memory</p> <p>Set this value using the Java Heap Size of NameNode in</p>	Minimum of 4 dedicated cores; more may be required for larger clusters	<ul style="list-style-type: none"> Minimum of 2 dedicated disks for metadata 1 dedicated disk for log files (This disk may be shared with the operating system.) Maximum disks: 4

Component	Memory	CPU	Disk
	Bytes HDFS configuration property.		
DataNode	<p>Minimum: 4 GB</p> <p>Increase the memory for higher replica counts or a higher number of blocks per DataNode. When increasing the memory, Cloudera recommends an additional 1 GB of memory for every 1 million replicas above 4 million on the DataNodes. For example, 5 million replicas require 5 GB of memory.</p> <p>Set this value using the Java Heap Size of DataNode in Bytes HDFS configuration property.</p>	<p>Minimum: 4 cores. Add more cores for highly active clusters.</p>	<p>Minimum: 4 Maximum: 24</p> <p>The maximum acceptable size will vary depending upon how large average block size is. The DN's scalability limits are mostly a function of the number of replicas per DN, not the overall number of bytes stored. That said, having ultra-dense DNs will affect recovery times in the event of machine or rack failure. Cloudera does not support exceeding 100 TB per data node. You could use 12 x 8 TB spindles or 24 x 4TB spindles. Cloudera does not support drives larger than 8 TB.</p>



Warning: Running CDH on storage platforms other than direct-attached physical disks can provide suboptimal performance. Cloudera Enterprise and the majority of the Hadoop platform are optimized to provide high performance by distributing work across a cluster that can utilize data locality and fast local I/O. Refer to the [Cloudera Enterprise Storage Device Acceptance Criteria Guide](#) for more information about using non-local storage.

HBase

Component	Java Heap	CPU	Disk
Master	<ul style="list-style-type: none"> • 100-10,000 regions: 4 GB • 10,000 or more regions with 200 or more Region Servers: 8 GB • 10,000 or more regions with 300 or more Region Servers: 12 GB <p>Set this value using the Java Heap Size of HBase Master in Bytes HBase configuration property.</p>	<p>Minimum 4 dedicated cores. You can add more cores for larger clusters, when using replication, or for bulk loads.</p>	<p>1 disk for local logs, which can be shared with the operating system and/or other Hadoop logs</p>
Region Server	<ul style="list-style-type: none"> • Minimum: 8 GB • Medium-scale production: 16 GB 	<p>Minimum: 4 dedicated cores</p>	<ul style="list-style-type: none"> • 4 or more spindles for each HDFS DataNode • 1 disk for local logs (this disk can be shared with

Component	Java Heap	CPU	Disk
	<ul style="list-style-type: none"> • Heap larger than 16 GB requires special Garbage Collection tuning. See Configuring the HBase BlockCache <p>Set this value using the Java Heap Size of HBase RegionServer in Bytes HBase configuration property.</p>		the operating system and/or other Hadoop logs
Thrift Server	<p>1 GB - 4 GB</p> <p>Set this value using the Java Heap Size of HBase Thrift Server in Bytes HBase configuration property.</p>	Minimum 2 dedicated cores.	1 disk for local logs, which can be shared with the operating system and other Hadoop logs.



Note: Consider adding more HBase Thrift Servers for production environments and deployments with a large number of Thrift client to scale horizontally.

Hive

Component	Java Heap		CPU	Disk	
HiveServer 2	Single Connection	4 GB	Minimum 4 dedicated cores	Minimum 1 disk	
	2-10 connections	4-6 GB		This disk is required for the following:	
	11-20 connections	6-12 GB		<ul style="list-style-type: none"> • HiveServer2 log files • stdout and stderr output files • Configuration files • Operation logs stored in the <code>operation_logs_dir</code> directory, which is configurable • Any temporary files that might be created by local map tasks under the <code>/tmp</code> directory 	
	21-40 connections	12-16 GB			
	41 to 80 connections	16-24 GB			
	Cloudera recommends splitting HiveServer2 into multiple instances and load balancing them once you start allocating more than 16 GB to HiveServer2. The objective is to adjust the size to reduce the impact of Java garbage collection on active processing by the service.				
	<p>Set this value using the Java Heap Size of HiveServer2 in Bytes Hive configuration property.</p> <p>For more information, see Tuning Hive in CDH.</p>				
Hive Metastore	Single Connection	4 GB	Minimum 4 dedicated cores	Minimum 1 disk	

Component	Java Heap		CPU	Disk
Hive Metastore Server	2-10 connections	4-10 GB		<p>This disk is required so that the Hive metastore can store the following artifacts:</p> <ul style="list-style-type: none"> Logs Configuration files Backend database that is used to store metadata if the database server is also hosted on the same node
	11-20 connections	10-12 GB		
	21-40 connections	12-16 GB		
	41 to 80 connections	16-24 GB		
	Set this value using the Java Heap Size of Hive Metastore Server in Bytes Hive configuration property. For more information, see Tuning Hive in CDH .			
Beeline CLI	Minimum: 2 GB		N/A	N/A

Hive on Spark Executor Nodes

Component	Memory	CPU	Disk
Hive-on-Spark	<ul style="list-style-type: none"> Minimum: 16 GB Recommended: 32 GB for larger data sizes Individual executor heaps should be no larger than 16 GB so machines with more RAM can use multiple executors.	<ul style="list-style-type: none"> Minimum: 4 cores Recommended: 8 cores for larger data sizes 	Disk space requirements are driven by scratch space requirements for Spark spill.

For more information on how to reserve YARN cores and memory that will be used by Spark executors, refer to [Tuning Apache Hive on Spark in CDH](#).

HSM KMS

Component	Memory	CPU	Disk
Navigator HSM KMS	16 GB RAM	Minimum: 2 GHz 64-bit quad core	40 GB, using moderate to high-performance drives.

Hue

Component	Memory	CPU	Disk
Hue Server	<ul style="list-style-type: none"> Minimum: 4 GB Maximum 10 GB If the cluster uses the Hue load balancer, add additional memory 	Minimum: 1 Core to run Django When Hue is configured for high availability, add additional cores	Minimum: 10 GB for the database, which grows proportionally according to the cluster size and workloads. When Hue is configured for high availability, add temp space is required



Note: Hue is limited by cgroup settings. In Cloudera Manager, all memory soft/hard limits are set to -1.

For more information about Hue high availability, see [How to Add a Hue Load Balancer](#).

Impala

Sizing requirements for Impala can vary significantly depending on the size and types of workloads using Impala.

Component	Native Memory	JVM Heap	CPU	Disk
Impala Daemon	<p>Set this value using the Impala Daemon Memory Limit configuration property.</p> <ul style="list-style-type: none"> • Minimum: 32 GB • Recommended: 128 GB 	<p>Set this value using the Java Heap Size of Impala Daemon in Bytes configuration property for the Coordinator Impala Daemons.</p> <ul style="list-style-type: none"> • Minimum: 4 GB • Recommended: 8 GB 	<ul style="list-style-type: none"> • Minimum: 4 • Recommended: 16 or more <p>CPU instruction set: AVX2</p>	<ul style="list-style-type: none"> • Minimum: 1 disk • Recommended: 8 or more
Catalog Server	<p>Set this value using the Java Heap Size of Catalog Server in Bytes configuration property.</p> <ul style="list-style-type: none"> • Minimum: 4 GB • Recommended: 8 GB 		<ul style="list-style-type: none"> • Minimum: 4 • Recommended: 16 or more <p>CPU instruction set: AVX2</p>	<ul style="list-style-type: none"> • Minimum and Recommended: 1 disk

For the networking topology for multi-rack cluster, [Leaf-Spine](#) is recommended for the optimal performance.

Kafka

Kafka requires a fairly small amount of resources, especially with some configuration tuning. By default, Kafka, can run on as little as 1 core and 1GB memory with storage scaled based on requirements for data retention.

CPU is rarely a bottleneck because Kafka is I/O heavy, but a moderately-sized CPU with enough threads is still important to handle concurrent connections and background tasks.

Kafka brokers tend to have a similar hardware profile to HDFS data nodes. How you build them depends on what is important for your Kafka use cases. Use the following guidelines:

To affect performance of these features:	Adjust these parameters:
Message Retention	Disk size
Client Throughput (Producer & Consumer)	Network capacity
Producer throughput	Disk I/O
Consumer throughput	Memory

A common choice for a Kafka node is as follows:

Component	Memory/Java Heap	CPU	Disk
Broker	<ul style="list-style-type: none"> • RAM: 64 GB • Recommended Java heap: 4 GB 	12- 24 cores	<ul style="list-style-type: none"> • 1 HDD For operating system • 1 HDD for Zookeeper dataLogDir

Component	Memory/Java Heap	CPU	Disk
	<p>Set this value using the Java Heap Size of Broker Kafka configuration property.</p> <p>See Other Kafka Broker Properties table.</p>		<ul style="list-style-type: none"> • 10- HDDs, using Raid 10, for Kafka data
MirrorMaker	<p>1 GB heap</p> <p>Set this value using the Java Heap Size of MirrorMaker Kafka configuration property.</p>	1 core per 3-4 streams	<p>No disk space needed on MirrorMaker instance.</p> <p>Destination brokers should have sufficient disk space to store the topics being copied over.</p>

Networking requirements: Gigabit Ethernet or 10 Gigabit Ethernet. Avoid clusters that span multiple data centers.

Key Trustee Server

Component	Memory	CPU	Disk
Key Trustee Server	8 GB	1 GHz 64-bit quad core	20 GB, using moderate to high-performance drives

Key Trustee KMS

Component	Memory	CPU	Disk
Key Trustee KMS	16 GB	2 GHz 64-bit quad core	40 GB, using moderate to high-performance drives

Kudu

Component	Memory	CPU	Disk
Tablet Server	<ul style="list-style-type: none"> • Minimum: 4 GB • Recommended: 10 GB <p>Additional hardware may be required, depending on the workloads running in the cluster. If you are using Impala, see the Impala sizing guidelines.</p>	<p>Kudu currently requires a CPU that supports the SSSE3 and SSE4.2 instruction sets.</p> <p>If you are to run Kudu inside a VM, enable SSE4.2 pass-through to pass through SSE4.2 support into the VM.</p>	1 disk for write-ahead log (WAL). Using an SSD drive may improve performance.
Master	<ul style="list-style-type: none"> • Minimum: 256 MB • Recommended: 1 GB 		1 disk

For more information, see [Kudu Server Management](#).

Oozie

Component	Java Heap	CPU	Disk
Oozie	<ul style="list-style-type: none"> • Minimum: 1 GB (this is the default set by Cloudera Manager). This is sufficient for less than 10 simultaneous workflows, without forking. • If you notice excessive garbage collection, or out-of-memory errors, increase the heap size to 4 GB for medium-size production clusters or to 8 GB for large-size production clusters. • Set this value using the Java Heap Size of Oozie Server in Bytes Oozie configuration property. 	No resources required	No resources required

Additional tuning:

For workloads with many coordinators that run with complex workflows (a **max concurrency reached!** warning appears in the log and the Oozie admin -queuedump command shows a large queue):

- Increase the value of the oozie.service.CallableQueueService.callable.concurrency property to 50.
- Increase the value of the oozie.service.CallableQueueService.threads property to 200.

Do not use a Derby database as a backend database for Oozie.

Search

Component	Java Heap	CPU	Disk
Solr	<ul style="list-style-type: none"> Small workloads, or evaluations: 16 GB Smaller production environments: 32 GB Larger production environments: 96 GB is sufficient for most clusters. <p>Set this value using the Java Heap Size of Solr Server in Bytes Solr configuration property. See</p>	<ul style="list-style-type: none"> Minimum: 4 Recommended: 16 for production workloads 	No requirement. Solr uses HDFS for storage.

Note the following considerations for determining the optimal amount of heap memory:

- Size of searchable material:** The more searchable material you have, the more memory you need. All things being equal, 10 TB of searchable data requires more memory than 1 TB of searchable data.
- Content indexed in the searchable material:** Indexing all fields in a collection of logs, email messages, or Wikipedia entries requires more memory than indexing only the Date Created field.
- The level of performance required:** If the system must be stable and respond quickly, more memory may help. If slow responses are acceptable, you may be able to use less memory.

For more information refer to [Deployment Planning for Cloudera Search](#).

Sentry

Component	Java Heap	CPU	Disk
Sentry Server	<ul style="list-style-type: none"> Minimum: 576 MB Recommended: 2.5 GB per million objects in the Hive database (servers, databases, tables, partitions, columns, URIs, and views) <p>Set this value with the Java Heap Size of Sentry Server in Bytes Sentry configuration property.</p>	Minimum: 4	

For more information about Sentry requirements, see [Before You Install Sentry](#).

Spark

Component	Java Heap	CPU	Disk
Spark History Server	<p>Minimum: 512 MB</p> <p>Set this value using the Java Heap Size of History Server</p>	1	Minimum 1 disk for log files.

Component	Java Heap	CPU	Disk
	<p>in Bytes Spark configuration property.</p>	<p>Important: Cloudera recommends that you adjust the number of CPUs and memory for the Spark History Server based on your specific cluster usage patterns.</p>	

YARN

Component	Java Heap	CPU	Other Recommendations
Job History Server	<ul style="list-style-type: none"> Minimum: 1 GB Increase memory by 1.6 GB for each 100,000 tasks kept in memory. For example: 5 jobs @ 100,000 mappers + 20,000 reducers = 600,000 total tasks requiring 9.6 GB of heap. See the Other Recommendations column for additional tuning suggestions. <p>Set this value using the Java Heap Size of JobHistory Server in Bytes YARN configuration property.</p>	Minimum: 1 core	<ul style="list-style-type: none"> Set the <code>mapred.jobhistory.hist.format</code> property to binary (history files will load about 2x-3x faster with this setting). Available in CDH 5.5.0 or higher only. Set the <code>mapred.jobhistory.tasksize</code> property to a total loaded task count. Using the example in the Java Heap column to the left, of 650,000 total tasks, you can set it to 700,000 to allow for some safety margin. This should also prevent the JobHistoryServer from hanging during garbage collection, since the job count limit does not have a task limit.
NodeManager	<p>Minimum: 1 GB.</p> <p>Configure additional heap memory for the following conditions:</p> <ul style="list-style-type: none"> Large number of containers 	<ul style="list-style-type: none"> Minimum: 8-16 cores Recommended: 32-64 cores 	<p>Disks:</p> <ul style="list-style-type: none"> Minimum: 8 disks Recommended: 12 or more disks <p>Networking:</p> <ul style="list-style-type: none"> Minimum: Dual 1Gbps or faster

Component	Java Heap	CPU	Other Recommendations
	<ul style="list-style-type: none"> Large shxmluffle sizes in Spark or MapReduce <p>Set this value using the Java Heap Size of NodeManager in Bytes YARN configuration property.</p>		<ul style="list-style-type: none"> Recommended: Single/Dual 10 Gbps or faster
ResourceManager	<p>Minimum: 6 GB</p> <p>Configure additional heap memory for the following conditions:</p> <ul style="list-style-type: none"> More jobs Larger cluster size Number of retained finished applications (configured with the <code>yarn.scheduler.maximum-applications</code> property). Scheduler configuration <p>Set this value using the Java Heap Size of ResourceManager in Bytes YARN configuration property.</p>	Minimum: 1 core	
Other Settings	<ul style="list-style-type: none"> Set the ApplicationMaster Memory YARN configuration property to 512 MB Set the Container Memory Minimum YARN configuration property to 1 GB. 	N/A	N/A

For more information, see [Tuning YARN](#).

ZooKeeper

Component	Java Heap	CPU	Disk
ZooKeeper Server	<ul style="list-style-type: none"> Minimum: 1 GB Increase heap size when watching 10,000 - 100,000 ephemeral znodes and are using 1,000 or more clients. <p>Set this value using the Java Heap Size of ZooKeeper Server in Bytes ZooKeeper configuration property.</p>	Minimum: 4 cores	<p>ZooKeeper was not designed to be a low-latency service and does not benefit from the use of SSD drives. The ZooKeeper access patterns – append-only writes and sequential reads – were designed with spinning disks in mind. Therefore Cloudera recommends using HDD drives.</p>

Additional information:

- [Managing ZooKeeper](#)
- [ZooKeeper Administration Guide: Things to Avoid](#)

Operating System Requirements

The following topics describe the operating system requirements for Cloudera software:

Software Dependencies

- **Python** - Cloudera Enterprise 6, with the exception of Hue, is supported on the Python version that is included in the operating system by default, as well as higher versions, but is not compatible with Python 3.0 or higher.

For example, Cloudera Enterprise 6 requires Python 2.6 or higher on RHEL 6 compatible operating systems, but requires Python 2.7 or higher on RHEL 7 compatible operating systems.

Hue in CDH 6 requires Python 2.7 or higher on all operating systems. For RHEL 6 compatible operating systems running Hue, you must [manually install Python 2.7](#).

Python 3 is not supported.

- **Perl** - Cloudera Manager requires [perl](#).
- **python-psycopg2** - Cloudera Manager 6 has a dependency on the package `python-psycopg2`. Hue in CDH 6 requires a higher version of `psycopg2` than is required by the Cloudera Manager dependency. For more information, see [Installing the psycopg2 Python Package](#).
- **iproute package** - Cloudera Enterprise 6 has a dependency on the `iproute` package. Any host that runs the Cloudera Manager Agent requires the package. The required version varies depending on the operating system:

Table 4: iproute package

Operating System	iproute version
RHEL 7 Compatible	iproute-3.10
RHEL 6 Compatible	iproute-2.6
SLES 12 SP2, SP3	iproute2-3.12

CDH and Cloudera Manager Supported Operating Systems

CDH provides 64-bit packages for select versions of RHEL-compatible, SLES, and Ubuntu operating systems.



Important:

In order to be covered by Cloudera Support:

- All CDH hosts in a logical cluster must run on the same major OS release.
- Cloudera supports a *temporarily* mixed OS configuration during an OS upgrade project.
- Cloudera Manager must run on the same OS release as one of the CDH clusters it manages.

Cloudera recommends running the same minor release on all cluster nodes. However, the risk caused by running different minor OS releases is considered lower than the risk of running different major OS releases.

Points to note:

- Red Hat only supports specific upgrades from RHEL 6 to 7. Contact your OS vendor and review the Red Hat article [What are the supported use cases for upgrading to RHEL 7?](#).
- Cloudera does not support CDH cluster deployments in Docker containers.

- Cloudera Enterprise, with the exception of Cloudera Navigator Encrypt, is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in enforcing mode. Cloudera is not responsible for policy support or policy enforcement. If you experience issues with SELinux, contact your OS provider.

Cloudera Enterprise 6.2.x Supported Operating Systems

Operating System	Version (bold=new)
RHEL-compatible	
RHEL/CentOS/OL with RHCK kernel	7.6, 7.5, 7.4, 7.3, 7.2 6.10, 6.9 , 6.8
Oracle Linux (OL)	7.6 , 7.4, 7.3, 7.2 (UEK default) 6.10 (UEK default)
SUSE Linux Enterprise Server	
SLES	12 SP3, 12 SP2
Ubuntu	
Ubuntu	18.04 LTS (Bionic) 16.04 LTS (Xenial)

Cloudera Enterprise 6.1.x Supported Operating Systems

Operating System	Version (bold=new)
RHEL-compatible	
RHEL/CentOS/OL with RHCK kernel	7.6 , 7.5, 7.4, 7.3, 7.2 6.10, 6.9 , 6.8
Oracle Linux (OL)	7.4, 7.3, 7.2 (UEK default) 6.10 (UEK default)
SUSE Linux Enterprise Server	
SLES	12 SP3, 12 SP2
Ubuntu	
Ubuntu	16.04 LTS (Xenial)

Cloudera Enterprise 6.0.x Supported Operating Systems

Operating System	Version (bold=new)
RHEL-compatible	
RHEL/CentOS/OL with RHCK kernel	7.6 , 7.5, 7.4, 7.3, 7.2 6.10 , 6.9 , 6.8
Oracle Linux (OL)	7.4, 7.3, 7.2 (UEK default)
SUSE Linux Enterprise Server	
SLES	12 SP3, 12 SP2
Ubuntu	

Operating System	Version (bold=new)
Ubuntu	16.04 LTS (Xenial)

Filesystem Requirements

Supported Filesystems

The Hadoop Distributed File System (HDFS) is designed to run on top of an underlying filesystem in an operating system. Cloudera recommends that you use either of the following filesystems tested on the [supported operating systems](#):

- **ext3:** This is the most tested underlying filesystem for HDFS.
- **ext4:** This scalable extension of ext3 is supported in more recent Linux releases.



Important: Cloudera does not support in-place upgrades from ext3 to ext4. Cloudera recommends that you format disks as ext4 before using them as data directories.

- **XFS:** This is the default filesystem in RHEL 7.
- **S3:** Amazon Simple Storage Service

Kudu Filesystem Requirements - Kudu is supported on ext4 and XFS. Kudu requires a kernel version and filesystem that supports hole punching. Hole punching is the use of the `fallocate(2)` system call with the `FALLOC_FL_PUNCH_HOLE` option set. For more details, see [Error during hole punch test](#).

File Access Time

Linux filesystems keep metadata that record when each file was accessed. This means that even reads result in a write to the disk. To speed up file reads, Cloudera recommends that you disable this option, called `atime`, using the `noatime` mount option in `/etc/fstab`:

```
/dev/sdb1 /data1 ext4 defaults,noatime 0
```

Apply the change without rebooting:

```
mount -o remount /data1
```

Filesystem Mount Options

The filesystem `mount` options have a `sync` option that allows you to write synchronously.

Using the `sync` filesystem mount option reduces performance for services that write data to disks, such as HDFS, YARN, Kafka and Kudu. In CDH, most writes are already replicated. Therefore, synchronous writes to disk are unnecessary, expensive, and do not measurably improve stability.

NFS and NAS options are not supported for use as DataNode Data Directory mounts, even when using Hierarchical Storage features.

nproc Configuration

Cloudera Manager automatically sets `nproc` configuration in `/etc/security/limits.conf`, but this configuration can be overridden by individual files in `/etc/security/limits.d/`. This can cause [problems](#) with Apache Impala and other components.

Make sure that the `nproc` limits are set sufficiently high, such as 65536 or 262144.

Database Requirements



Important: When you restart processes, the configuration for each of the services is redeployed using information saved in the Cloudera Manager database. If this information is not available, your cluster cannot start or function correctly. You must schedule and maintain regular backups of the Cloudera Manager database to recover the cluster in the event of the loss of this database. For more information, see [Backing Up Databases](#).

Cloudera Manager and CDH come packaged with an embedded PostgreSQL database for use in non-production environments. The embedded PostgreSQL database is not supported in production environments. For production environments, you must configure your cluster to use dedicated external databases.

After installing a database, upgrade to the latest patch and apply appropriate updates. Available updates may be specific to the operating system on which it is installed.

Notes:

- Use UTF8 encoding for all custom databases. MySQL and MariaDB must use the MySQL `utf8` encoding, not `utf8mb4`.
- For MySQL 5.7, you must install the *MySQL-shared-compat* or *MySQL-shared* package. This is required for the Cloudera Manager Agent installation.
- MySQL GTID-based replication is not supported.
- Hue *requires* the default MySQL/MariaDB version (if used) of the operating system on which it is installed. For more information, see [Hue Databases](#).
- Both the Community and Enterprise versions of MySQL are supported, as well as MySQL configured by the AWS RDS service.
- Before upgrading from CDH 5 to CDH 6, check the value of the `COMPATIBLE` initialization parameter in the Oracle Database using the following SQL query:

```
SELECT name, value FROM v$parameter WHERE name = 'compatible'
```

The default value is 12.2.0. If the parameter has a different value, you can set it to the default as shown in the [Oracle Database Upgrade Guide](#).



Note: Before resetting the `COMPATIBLE` initialization parameter to its default value, make sure you consider the effects of this change can have on your system.

Table 5:  MySQL Support across Cloudera Enterprise 6 Releases

MySQL Version	Cloudera Enterprise 6.x
5.7	✓

Table 6:  MariaDB Support across Cloudera Enterprise 6 Releases

MariaDB Version	Cloudera Enterprise 6.2	Cloudera Enterprise 6.0 - 6.1
5.5	✓	✓
10.0	✓	✓
10.1	✓	



Note: PostgreSQL 10 is currently not supported in Cloudera Enterprise 6.0. If your operating system version has PostgreSQL 10 pre-installed, you must downgrade to a supported version.

Table 7: 📈 PostgreSQL Support across Cloudera Enterprise 6 Releases

PostgreSQL Version	Cloudera Enterprise 6.1 - 6.2	Cloudera Enterprise 6.0
8.4	✓	✓
9.2	✓	✓
9.4	✓	✓
9.5	✓	
9.6	✓	
10.x	✓	

Table 8: 📈 Oracle Support across Cloudera Enterprise 6 Releases

Oracle Version	Cloudera Enterprise 6.x
12.2	✓

RDBMS High Availability Support

Various Cloudera components rely on backing RDBMS services as critical infrastructure. You may require Cloudera components to support deployment in environments where RDBMS services are made highly-available. High availability (HA) solutions for RDBMS are implementation-specific, and can create constraints or behavioral changes in Cloudera components.

This section clarifies the support state and identifies known issues and limitations for HA deployments.

High Availability vs. Load Balancing

Understanding the difference between HA and load balancing is important for Cloudera components, which are designed to assume services are provided by a single RDBMS instance. Load balancing distributes operations across multiple RDBMS services in parallel, while HA focuses on service continuity. Load balanced deployments are often used as part of HA strategies to overcome demands of monitoring and failover management in an HA environment. While less easier to implement, load-balanced deployments require applications tailored to the behavior and limitations of the particular technology.

Support Statement: Cloudera components are not designed for and do not support load balanced deployments of any kind. Any HA strategy involving multiple active RDBMS services must ensure all connections are routed to a single RDBMS service at any given time, regardless of vendor or HA implementation/technology.

General High Availability Support

Cloudera supports various RDBMS options, each of which have multiple possible strategies to implement HA. Cloudera cannot reasonably test and certify on each strategy for each RDBMS. Cloudera expects HA solutions for RDBMS to be transparent to Cloudera software, and therefore are not supported and debugged by Cloudera. It is the responsibility of the customer to provision, configure, and manage the RDBMS HA deployment, so that Cloudera software behaves as it would when interfacing with a single, non-HA service. Cloudera will support and help customers troubleshoot issues when a cluster has HA enabled. While diagnosing database-related problems in Cloudera components, customers may be required to temporarily disable or bypass HA mechanisms for troubleshooting purposes. If an HA-related issue is found, it is the responsibility of the customer to engage with the database vendor so that a solution to that issue can be found.

Support Statement: Cloudera Support may require customers to temporarily bypass HA layers and connect directly to supported RDBMS back-ends to troubleshoot issues. Issues observed only when connected through HA layers are the responsibility of the customer DBA staff to resolve.

Vendor-Specific Notes

Oracle RAC:

- Cloudera supports Oracle Exadata and RAC instances when they serve as back-end databases for CDH components without HA. Cloudera software is designed with the assumption of a single database instance, and supports normal operations between Cloudera Enterprise and Oracle Exadata (or RAC) in such an environment.
- Cloudera is an [Oracle Partner Network Gold](#) member, allowing us to download and use Oracle commercial software (such as RAC) for development and testing purposes.

MySQL Asynchronous Replication:

- Supported, tested, and certified
- Master/master or master/slave topographies are acceptable
- You must disable Global Transaction Identifiers (GTID)
- You must use the InnoDB storage engine

MySQL HA with Oracle Clusterware:

- Not tested or certified
- No known or expected problems
- Shared disk, active/passive MySQL hosts
 - Will guard against component failure on MySQL Server host
 - Won't guard against logical or physical corruption or loss
 - Separate DR plan is required
- Resources:
 - [Blog post](#)
 - [Oracle Clusterware product page](#)
 - [Grid Infrastructure Agent documentation](#)

MySQL InnoDB Cluster:

- Prohibited
- Requires enabling GTIDs

MySQL DRBD:

- Older HA tech stack for MySQL, does distributed block writes at OS kernel layer
- Does not add additional semantics or requirements
- Does have performance tradeoffs for write operations
- Poorly suited to write-intensive use cases (e.g. Navigator)

MySQL Cluster (NDB):

- Prohibited
- Very different performance, management and operational characteristics from InnoDB storage engine

Galera Cluster (Percona Cluster, MariaDB Cluster):

- Prohibited
- Adds cluster-wide optimistic locking. This can cause unexpected deadlock errors at commit, or worse, undetected logical database corruption caused by naive retry logic in Cloudera applications

Java Requirements



Note: A Java optimization called [compressed oops](#) (ordinary object pointers) enables a 64-bit JVM to address heap sizes up to about 32 GB using 4-byte pointers. For larger heap sizes, 8-byte pointers are required. This means that a heap size slightly less than 32 GB can hold more objects than a heap size slightly more than 32 GB.

If you do not need more than 32 GB heap, set your heap size to 31GB or less to avoid this issue. If you need 32 GB or more, set your heap size to 48 GB or higher to account for the larger pointers. In general, for heap sizes above 32 GB, multiply the amount of heap you need by 1.5.

Only 64 bit JDKs are supported. Cloudera Manager 6 and CDH 6 do not support JDK 7. Although JDK 7 is supported on all versions of CDH 5, a CDH 5.x cluster that is managed by Cloudera Manager 6.x must use JDK 8 on all cluster hosts. [Oracle JDK 8](#) is supported in Cloudera Manager 6 and CDH 6. JDK 8 is also supported in CDH 5.3 and higher.

OpenJDK 8 is supported in Cloudera Enterprise 6.1.0 and higher, as well as Cloudera Enterprise 5.16.1 and higher. For installation and migration instructions, see [Upgrading the JDK](#).

Applications compiled with JDK 7 are not supported on CDH 6. You must recompile your applications using JDK 8 before upgrading to CDH 6.

Oracle JDK 9 is not supported in any Cloudera Manager or CDH version.

Unless specifically excluded, Cloudera supports later updates to a major JDK release from the release that support was introduced. Cloudera excludes or removes support for select Java updates when security is jeopardized.

Running CDH nodes within the same cluster on different JDK releases is not supported. All cluster hosts must use the same JDK update level.

Supported JDKs

Cloudera Enterprise Version	Supported JDK
5.3 -5.15	Oracle JDK 1.7, Oracle JDK 1.8
5.16 and higher 5.x releases	Oracle JDK 1.7, Oracle JDK 1.8, OpenJDK 1.8
6.0	Oracle JDK 1.8
6.1	Oracle JDK 1.8, OpenJDK 1.8
6.2	Oracle JDK 1.8, OpenJDK 1.8

JDK 8

All JDK 8 updates, from the minimum required version, are supported in Cloudera Enterprise 6 unless specifically excluded. Updates above the minimum that are not listed are supported but not tested. JDK 8 is required for Cloudera Manager 6 and CDH 6.



Warning:

- JDK 8u40, 8u45, and 8u60 are not supported due to JDK issues impacting CDH functionality:
 - JDK 8u40 and 8u45 are affected by [JDK-8077155](#), which affects HTTP authentication for certain web UIs.
 - JDK 8u60 is incompatible with the AWS SDK, and causes problem with DistCP. For more information, see the [KB article](#).
- [Oozie Workflow Graph Display](#) in Hue does not work properly with JDK versions lower than 8u40.

Table 9: Oracle JDK 8 versions that are tested and recommended

Oracle JDK Version	Notes
1.8u181	Recommended / Latest version tested
1.8u162	Recommended
1.8u141	Recommended
1.8u131	Recommended
1.8u121	Recommended
1.8u111	Recommended
1.8u102	Recommended
1.8u91	Recommended
1.8u74	Recommended
1.8u31	Minimum required

Table 10: OpenJDK 8 versions that are tested and recommended

OpenJDK Version	Notes
1.8u181	Minimum required / Latest version tested

JDK 7



Important: JDK 7 and lower are not supported in Cloudera Manager 6.x and CDH 6.x.

Networking and Security Requirements

CDH and Cloudera Manager Supported Transport Layer Security Versions

The following components are supported by the indicated versions of Transport Layer Security (TLS):

Table 11: Components Supported by TLS

Component	Role	Name	Port	Version
Cloudera Manager	Cloudera Manager Server		7182	TLS 1.2
Cloudera Manager	Cloudera Manager Server		7183	TLS 1.2
Flume			9099	TLS 1.2
Flume		Avro Source/Sink		TLS 1.2
Flume		Flume HTTP Source/Sink		TLS 1.2
HBase	Master	HBase Master Web UI Port	60010	TLS 1.2
HDFS	NameNode	Secure NameNode Web UI Port	50470	TLS 1.2
HDFS	Secondary NameNode	Secure Secondary NameNode Web UI Port	50495	TLS 1.2

Component	Role	Name	Port	Version
HDFS	HttpFS	REST Port	14000	TLS 1.1, TLS 1.2
Hive	HiveServer2	HiveServer2 Port	10000	TLS 1.2
Hue	Hue Server	Hue HTTP Port	8888	TLS 1.2
Impala	Impala Daemon	Impala Daemon Beeswax Port	21000	TLS 1.2
Impala	Impala Daemon	Impala Daemon HiveServer2 Port	21050	TLS 1.2
Impala	Impala Daemon	Impala Daemon Backend Port	22000	TLS 1.2
Impala	Impala StateStore	StateStore Service Port	24000	TLS 1.2
Impala	Impala Daemon	Impala Daemon HTTP Server Port	25000	TLS 1.2
Impala	Impala StateStore	StateStore HTTP Server Port	25010	TLS 1.2
Impala	Impala Catalog Server	Catalog Server HTTP Server Port	25020	TLS 1.2
Impala	Impala Catalog Server	Catalog Server Service Port	26000	TLS 1.2
Oozie	Oozie Server	Oozie HTTPS Port	11443	TLS 1.1, TLS 1.2
Solr	Solr Server	Solr HTTP Port	8983	TLS 1.1, TLS 1.2
Solr	Solr Server	Solr HTTPS Port	8985	TLS 1.1, TLS 1.2
Spark	History Server		18080	TLS 1.2
YARN	ResourceManager	ResourceManager Web Application HTTP Port	8090	TLS 1.2
YARN	JobHistory Server	MRv1 JobHistory Web Application HTTP Port	19890	TLS 1.2

CDH and Cloudera Manager Networking and Security Requirements

The hosts in a Cloudera Manager deployment must satisfy the following networking and security requirements:

- **Networking Protocols Support**

CDH requires IPv4. IPv6 is not supported and must be disabled.



Note: Contact your OS vendor for help disabling IPv6.

See also [Configure Network Names](#).

- **Multihoming Support**

Multihoming CDH or Cloudera Manager is not supported outside specifically certified Cloudera partner appliances. Cloudera finds that current Hadoop architectures combined with modern network infrastructures and security practices remove the need for multihoming. Multihoming, however, is beneficial internally in appliance form factors to take advantage of high-bandwidth InfiniBand interconnects.

Although some subareas of the product may work with unsupported custom multihoming configurations, there are known issues with multihoming. In addition, unknown issues may arise because multihoming is not covered by our test matrix outside the Cloudera-certified partner appliances.

- **Entropy**

Data at rest encryption requires sufficient [entropy](#) to ensure randomness.

See entropy requirements in [Data at Rest Encryption Requirements](#) on page 31.

- Cluster hosts must have a working network name resolution system and correctly formatted `/etc/hosts` file. All cluster hosts must have properly configured forward and reverse host resolution through DNS. The `/etc/hosts` files must:
 - Contain consistent information about hostnames and IP addresses across all hosts
 - Not contain uppercase hostnames
 - Not contain duplicate IP addresses

Cluster hosts must not use aliases, either in `/etc/hosts` or in configuring DNS. A properly formatted `/etc/hosts` file should be similar to the following example:

```
127.0.0.1 localhost.localdomain localhost
192.168.1.1 cluster-01.example.com cluster-01
192.168.1.2 cluster-02.example.com cluster-02
192.168.1.3 cluster-03.example.com cluster-03
```

- In most cases, the Cloudera Manager Server must have SSH access to the cluster hosts when you run the installation or upgrade wizard. You must log in using a root account or an account that has password-less `sudo` permission. For authentication during the installation and upgrade procedures, you must either enter the password or upload a public and private key pair for the `root` or `sudo` user account. If you want to use a public and private key pair, the public key must be installed on the cluster hosts before you use Cloudera Manager.

Cloudera Manager uses SSH only during the initial install or upgrade. Once the cluster is set up, you can disable root SSH access or change the root password. Cloudera Manager does not save SSH credentials, and all credential information is discarded when the installation is complete.

- The Cloudera Manager Agent runs as `root` so that it can make sure that the required directories are created and that processes and files are owned by the appropriate user (for example, the `hdfs` and `mapred` users).
- Security-Enhanced Linux (SELinux) must not block Cloudera Manager or CDH operations.



Note: Cloudera Enterprise, with the exception of Cloudera Navigator Encrypt, is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in `enforcing` mode. Cloudera is not responsible for SELinux policy development, support, or enforcement. If you experience issues running Cloudera software with SELinux enabled, contact your OS provider for assistance.

If you are using SELinux in `enforcing` mode, Cloudera Support can request that you disable SELinux or change the mode to `permissive` to rule out SELinux as a factor when investigating reported issues.

- Firewalls (such as `iptables` and `firewalld`) must be disabled or configured to allow access to [ports](#) used by Cloudera Manager, CDH, and related services.
- For RHEL and CentOS, the `/etc/sysconfig/network` file on each host must contain the correct hostname.
- Cloudera Manager and CDH use several user accounts and groups to complete their tasks. The set of user accounts and groups varies according to the components you choose to install. Do not delete these accounts or groups and do not modify their permissions and rights. Ensure that no existing systems prevent these accounts and groups from functioning. For example, if you have scripts that delete user accounts not in a whitelist, add these accounts to the list of permitted accounts. Cloudera Manager, CDH, and managed services create and use the following accounts and groups:

Table 12: Users and Groups

Component (Version)	Unix User ID	Groups	Functionality
Cloudera Manager (all versions)	cloudera-scm	cloudera-scm	Clusters managed by Cloudera Manager run Cloudera Manager Server, monitoring roles, and other Cloudera Server processes as cloudera-scm. Requires keytab file named <code>cmf.keytab</code> because name is hard-coded in Cloudera Manager.
Apache Accumulo	accumulo	accumulo	Accumulo processes run as this user.
Apache Flume	flume	flume	The sink that writes to HDFS as user must have write privileges.
Apache HBase	hbase	hbase	The Master and the RegionServer processes run as this user.
HDFS	hdfs	hdfs, hadoop	The NameNode and DataNodes run as this user, and the HDFS root directory as well as the directories used for edit logs should be owned by it.
Apache Hive	hive	hive	The HiveServer2 process and the Hive Metastore processes run as this user. A user must be defined for Hive access to its Metastore DB (for example, MySQL or Postgres) but it can be any identifier and does not correspond to a Unix uid. This is <code>javax.jdo.option.ConnectionUserName</code> in <code>hive-site.xml</code> .
Apache HCatalog	hive	hive	The WebHCat service (for REST access to Hive functionality) runs as the <code>hive</code> user.
HttpFS	httpfs	httpfs	The HttpFS service runs as this user. See HttpFS Security Configuration for instructions on how to generate the merged <code>httpfs-http.keytab</code> file.
Hue	hue	hue	Hue services run as this user.
Hue Load Balancer	apache	apache	The Hue Load balancer has a dependency on the apache2 package that uses the <code>apache</code> user name. Cloudera Manager does not run processes using this user ID.
Impala	impala	impala, hive	Impala services run as this user.
Apache Kafka	kafka	kafka	Kafka brokers and mirror makers run as this user.
Java KeyStore KMS	kms	kms	The Java KeyStore KMS service runs as this user.
Key Trustee KMS	kms	kms	The Key Trustee KMS service runs as this user.
Key Trustee Server	keytrustee	keytrustee	The Key Trustee Server service runs as this user.
Kudu	kudu	kudu	Kudu services run as this user.
MapReduce	mapred	mapred, hadoop	Without Kerberos, the JobTracker and tasks run as this user. The LinuxTaskController binary is owned by this user for Kerberos.

Component (Version)	Unix User ID	Groups	Functionality
Apache Oozie	oozie	oozie	The Oozie service runs as this user.
Parquet	~	~	No special users.
Apache Pig	~	~	No special users.
Cloudera Search	solr	solr	The Solr processes run as this user.
Apache Spark	spark	spark	The Spark History Server process runs as this user.
Apache Sentry	sentry	sentry	The Sentry service runs as this user.
Apache Sqoop	sqoop	sqoop	This user is only for the Sqoop1 Metastore, a configuration option that is not recommended.
YARN	yarn	yarn, hadoop	Without Kerberos, all YARN services and applications run as this user. The LinuxContainerExecutor binary is owned by this user for Kerberos.
Apache ZooKeeper	zookeeper	zookeeper	The ZooKeeper processes run as this user. It is not configurable.

Data at Rest Encryption Requirements

Encryption comprises several components, each with its own requirements.

Data at rest encryption protection can be applied at a number of levels within Hadoop:

- OS filesystem-level
- Network-level
- HDFS-level (protects both data at rest and in transit)

This section contains the various hardware and software requirements for all encryption products used for Data at Rest Encryption.

For more information on supported operating systems, see [Product Compatibility Matrix for Cloudera Navigator Encryption](#) on page 39.

For more information on the components, concepts, and architecture for encrypting data at rest, see [Encrypting Data at Rest](#).

Entropy Requirements

Cryptographic operations require [entropy](#) to ensure randomness.

You can check the available entropy on a Linux system by running the following command:

```
cat /proc/sys/kernel/random/entropy_avail
```

The output displays the entropy currently available. Check the entropy several times to determine the state of the entropy pool on the system. If the entropy is consistently low (500 or less), you must increase it by installing `rng-tools` and starting the `rngd` service. Run the following commands on RHEL 6-compatible systems:

```
sudo yum install rng-tools
sudo echo 'EXTRAOPTIONS="-r /dev/urandom"' >> /etc/sysconfig/rngd
sudo service rngd start
sudo chkconfig rngd on
```

For RHEL 7, run the following commands:

```
sudo yum install rng-tools
cp /usr/lib/systemd/system/rngd.service /etc/systemd/system/
sed -i -e 's/ExecStart=\/sbin\/rngd -f/ExecStart=\/sbin\/rngd -f -r \/dev\urandom/'
/etc/systemd/system/rngd.service
systemctl daemon-reload
systemctl start rngd
systemctl enable rngd
```

Make sure that the hosts running Key Trustee Server, Key Trustee KMS, and Navigator Encrypt have sufficient entropy to perform cryptographic operations.

Cloudera Manager Requirements

Installing and managing Key Trustee Server using Cloudera Manager requires Cloudera Manager 5.4.0 and higher. Key Trustee Server does not require Cloudera Navigator Audit Server or Metadata Server.

umask Requirements

Key Trustee Server installation requires the default umask of 0022.

Network Requirements

For new Key Trustee Server installations (5.4.0 and higher) and migrated upgrades (see [Migrate Apache Web Server to CherryPy](#) for more information), Key Trustee Server requires the following TCP ports to be opened for inbound traffic:

- 11371

Clients connect to this port over HTTPS.

- 11381 (PostgreSQL)

The passive Key Trustee Server connects to this port for database replication.

For upgrades that are not migrated to the CherryPy web server, the pre-upgrade port settings are preserved:

- 80

Clients connect to this port over HTTP to obtain the Key Trustee Server public key.

- 443 (HTTPS)

Clients connect to this port over HTTPS.

- 5432 (PostgreSQL)

The passive Key Trustee Server connects to this port for database replication.

TLS Certificate Requirements

To ensure secure network traffic, Cloudera recommends obtaining Transport Layer Security (TLS) certificates specific to the hostname of your Key Trustee Server. To obtain the certificate, generate a Certificate Signing Request (CSR) for the fully qualified domain name (FQDN) of the Key Trustee Server host. The CSR must be signed by a trusted Certificate Authority (CA). After the certificate has been verified and signed by the CA, the Key Trustee Server TLS configuration requires:

- The CA-signed certificate
- The private key used to generate the original CSR
- The intermediate certificate/chain file (provided by the CA)

Cloudera recommends not using self-signed certificates. If you use self-signed certificates, you must use the --skip-ssl-check parameter when registering Navigator Encrypt with the Key Trustee Server. This skips TLS hostname

validation, which safeguards against certain network-level attacks. For more information regarding insecure mode, see [Table 1](#).

Browser Requirements

Cloudera Manager, Cloudera Navigator, and Hue are supported on the two most recent [LTS](#) (long term support) or [ESR](#) (extended support release) browsers. Cookies and JavaScript must be enabled.



Important: To see all icons in the Hue Web UI, users with IE and HTTPS must add a Load Balancer.

The following lists the minimum tested versions of the most common browsers: Hue can display in older, and other browsers, but you might not have access to all of its features.

Cloudera Enterprise 6.2

- **Chrome:** 63 ([Version history](#))
- **Firefox:** 59 ([Version history](#))
- **Safari** (Mac only): [Version history](#)
- **Internet Explorer:** 11 ([Version history](#))
- **Microsoft Edge:** 41 ([Version history](#))

Cloudera Enterprise 6.1

- **Chrome:** 63 ([Version history](#))
- **Firefox:** 59 ([Version history](#))
- **Safari** (Mac only): [Version history](#)
- **Internet Explorer:** 11 ([Version history](#))
- **Microsoft Edge:** 41 ([Version history](#))

Cloudera Enterprise 6.0

- **Chrome:** 63 ([Version history](#))
- **Firefox:** 59 ([Version history](#))
- **Safari** (Mac only): [Version history](#)
- **Internet Explorer:** 11 ([Version history](#))
- **Microsoft Edge:** 41 ([Version history](#))

Supported Configurations with Virtualization and Cloud Platforms

This section lists supported cloud and virtualization platforms for deploying Cloudera software. The linked reference architectures are not replacements for statements of support, but are guides to assist with deployment and sizing options.

Amazon Web Services

For information on deploying Cloudera software on a Amazon Web Services (AWS) cloud infrastructure, see the [Cloudera Enterprise Reference Architecture for AWS Deployments](#).

Google Cloud Platform

For information on deploying Cloudera software on a Google Cloud Platform infrastructure, see the [Cloudera Enterprise Reference Architecture for Google Cloud Platform Deployments](#).

Microsoft Azure

For information on deploying Cloudera software on a Microsoft Azure cloud infrastructure, see the [Cloudera Enterprise Reference Architecture for Azure Deployments](#).

Support Limitations for CDH and Cloudera Manager in Microsoft Azure

- Virtual machines must use Cloudera-published CentOS 6.x/7.x OS Images or Red Hat published 6.x/7.x Images. Red Hat images must include bootstrapped configuration as specified in the [Azure Bootstrap Scripts folder](#) in the Cloudera Director GitHub repository.
- Master Node Services *must use* Azure Premium Storage Disks for persistent storage.
- The VM local [temporary disk](#) *must not* be used for any persistent data for HDFS, Kudu, or other services.
- [Microsoft Azure Storage - Block Blob](#) (previously known as Windows Azure Storage Blob, accessed via wasb:// URI's) is only supported for backups using Hadoop's DistCP. Other services are not supported running directly against Azure Storage block blobs.

VMware

For information on deploying Cloudera software on a VMware-based infrastructure, see the [Reference architecture for deploying on VMware](#).

Recommendation when deploying on VMware in the current release:

- Use the part of Hadoop Virtual Extensions that has been implemented in [HADOOP-8468](#). This will prevent data loss when a physical node that hosts two or more DataNodes goes down.

Red Hat OpenStack Platform (OSP) 11

For information on deploying Cloudera software on RedHat OpenStack Platform (OSP) 11, see [Reference Architecture for Deploying CDH 5.x on Red Hat OSP 11](#).

For information on deploying Cloudera software on RedHat OpenStack Platform (OSP) 11 with Ceph storage, see [Reference Architecture for Deploying Cloudera Enterprise 5.x on Red Hat OpenStack Platform 11 with Red Hat Ceph Storage 2.x](#).

Product Compatibility Matrices

For more information on component compatibility across versions, see the following compatibility matrices:

Cloudera Manager and CDH Compatibility

Cloudera uses the following versioning convention: <major>.<minor>.<maintenance>. For example, if a cluster runs Cloudera Manager 6.0.0, the major version is 6, the minor version is 0, and the maintenance version is 0.

The Cloudera Manager <major> + <minor> version must always be equal to or greater than the CDH <major> + <minor> version. Older versions of Cloudera Manager might not support features in newer versions of CDH.

For example:

- Cloudera Manager 5.12.0 can manage CDH 5.12.2 because the <minor> versions are equal. Cloudera Manager 5.12.0 cannot manage CDH 5.14.0 because the Cloudera Manager <minor> version, 12, is less than the CDH <minor> version, 14.
- Cloudera Manager 6.0 can manage clusters running CDH 5.7 up to CDH 5.14 and as long as the <major> + <minor> version of Cloudera Manager is equal or higher than the <major> + <minor> version of CDH.



Important: Using Cloudera Manager 6.0.x to manage a CDH 5.15.x or CDH 5.16 cluster is not a supported configuration.

For more information, see [Supported Upgrade Paths](#).

Product Compatibility Matrix for Apache Accumulo



Warning: Cloudera Manager cannot upgrade Apache Accumulo. Follow the instruction in [Apache Accumulo Installation Guide](#) to upgrade to a CDH 6 compatible Apache Accumulo version.

This matrix contains compatibility information across versions of Apache Accumulo, and CDH and Cloudera Manager. For detailed information on each release, see [Apache Accumulo documentation](#).

Product	Lowest supported Cloudera Manager Version	Lowest supported CDH Version	Lowest supported Impala Version	Lowest supported Search Version	Integrated into CDH
Accumulo 1.9.2	Cloudera Manager 6.0.0	CDH 6.0.0	Not Supported	Not Supported	No
Accumulo 1.7.2	Cloudera Manager 5.0.0	CDH 5.5.0	Not Supported	Not Supported	No
Accumulo 1.6.0	Cloudera Manager 5.0.0	CDH 4.6.0	Not Supported	Not Supported	No

Product Compatibility Matrix for Backup and Disaster Recovery

This matrix contains compatibility information across features of Cloudera Manager Backup and Disaster Recovery and CDH and Cloudera Manager.

Feature	Lowest supported Cloudera Manager Version	Lowest supported CDH Version	Supported Services
Replication	Cloudera Manager 5.0.0	CDH 5.0.0	HDFS, Hive, Impala
Replication to and from Amazon S3*	Cloudera Manager 5.9.0	CDH 5.9.0	HDFS, Hive, Impala
Snapshots	Cloudera Manager 5.0.0	CDH 5.0.0	HDFS, Hive, Impala
Snapshots from Isilon storage	Not Supported	Not Supported	HDFS, Hive, Impala
Replication to and from Microsoft ADLS Gen1	Cloudera Manager 6.1.0 or Cloudera Manager 5.15.0	CDH 5.15	HDFS, Hive, Impala

*BDR does not support S3 as a source or destination when S3 is configured to use SSE-KMS.

Starting in Cloudera Manager 6.1.0, BDR ignores Hive tables backed by Kudu during replication. The change does not affect functionality since BDR does not support tables backed by Kudu. This change was made to guard against data loss due to how the Hive Metastore, Impala, and Kudu interact.

Supported Replication Scenarios

Versions

To replicate data to or from clusters managed by Cloudera Manager 6, the source or destination cluster must be managed by Cloudera Manager 5.14.0 or higher. Note that some functionality may not be available in Cloudera Manager 5.14.0 and higher or 6.0.0 and higher.

Kerberos

BDR supports the following replication scenarios when Kerberos authentication is used on a cluster:

- Secure source to a secure destination.
- Insecure source to an insecure destination.
- Insecure source to a secure destination. Keep the following requirements in mind:
 - In replication scenarios where a destination cluster has multiple source clusters, all the source clusters must either be secure or insecure. BDR does not support replication from a mixture of secure and insecure source clusters.
 - The destination cluster must run Cloudera Manager 6.1.0 or higher.
 - The source cluster must run a compatible Cloudera Manager version.
 - This replication scenario requires additional configuration. For more information, see [Replicating from Insecure to Secure Clusters](#) for Hive and [Replicating from Insecure to Secure Clusters](#) for HDFS.

Cloud Storage

BDR supports replicating to or from Amazon S3 and Microsoft Azure ADLS Gen1.

TLS

You can use TLS with BDR. Additionally, BDR supports replication scenarios where TLS is enabled for non-Hadoop services (Hive/Impala) and TLS is disabled Hadoop services (such as HDFS,YARN, and MapReduce).

Unsupported Replication Scenarios



Note: If you are using Isilon storage for CDH, see [Supported Replication Scenarios for Clusters using Isilon Storage](#) on page 36.

Versions

Replicating to or from Cloudera Manager 6 managed clusters with Cloudera Manager versions earlier than 5.14.0 are not supported.

Kerberos

BDR does not support the following replication scenarios when Kerberos authentication is used on a cluster:

- Secure source to an insecure destination is not supported.

Cloud Storage

BDR does not support replicating to or from ADLS Gen2 Preview.

Supported Replication Scenarios for Clusters using Isilon Storage



Warning: CDH 6.x is currently not supported on Dell EMC Isilon.

Note the following when scheduling replication jobs for clusters that use Isilon storage:

- As of CDH 5.8 and higher, Replication is supported for clusters using Kerberos and Isilon storage on the source or destination cluster, or both. See [Configuring Replication with Kerberos and Isilon](#). Replication between clusters using Isilon storage and Kerberos is not supported in CDH 5.7.
- Make sure that the `hdfs` user is a superuser in the Isilon system. If you specify alternate users with the **Run As** option when creating replication schedules, those users must also be superusers.
- Cloudera recommends that you use the Isilon `root` user for replication jobs. (Specify `root` in the **Run As** field when creating replication schedules.)
- Select the **Skip checksum checks** property when creating replication schedules.
- Clusters that use Isilon storage do not support [snapshots](#). Snapshots are used to ensure data consistency during replications in scenarios where the source files are being modified. Therefore, when replicating from an Isilon cluster, Cloudera recommends that you do not replicate Hive tables or HDFS files that could be modified before the replication completes.

See [Using CDH with Isilon Storage](#).

Product Compatibility Matrix for Cloudera Data Science Workbench

Cloudera Data Science Workbench is a product that enables fast, easy, and secure self-service data science for the enterprise. It allows data scientists to bring their existing skills and tools, such as R, Python, and Scala, to securely run computations on data in Hadoop clusters.

For details about platform requirements for Cloudera Data Science Workbench, refer the [CDSW Requirements and Supported Platforms](#) topic.

Product Compatibility for Dell EMC Isilon



Warning: CDH 6.x is currently not supported on Dell EMC Isilon.

Product Compatibility Matrix for Cloudera Navigator

This matrix contains compatibility information across versions of Cloudera Navigator, Cloudera Manager, and CDH. For detailed information on each release, see [Cloudera Navigator documentation](#). For details on services supported by Navigator in the latest release, see:

- Audit: [Operations by Component](#)
- Metadata: [Service Metadata Entity Types](#)



Note:

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

Product	Feature	Lowest supported Cloudera Manager Version	Lowest supported CDH Version
Cloudera Navigator 6.0.x, 6.1.x, 6.2.x	Auditing, Metadata, Analytics, and Security	Same as Cloudera Navigator	<ul style="list-style-type: none"> • Audit Component <ul style="list-style-type: none"> – CDH 5.4, including HDFS, HBase, Hue, HiveServer2, Impala, Sentry, and Solr • Metadata Component <ul style="list-style-type: none"> – CDH 5.4, including HDFS, HiveServer2, Impala, MapReduce, Oozie, Pig, Sqoop 1, Spark 1 (1.6), YARN – CDH 6.0, including HDFS, HiveServer2, Impala, MapReduce, Oozie, Pig, S3, Spark 1 (1.6) and Spark 2 (2.3), Sqoop 1, YARN

Cloudera Navigator Supported Databases

Cloudera Navigator supports the same databases as CDH and Cloudera Manager. For more information, see [Database Requirements](#) on page 23.

Cloudera Navigator Supported Browsers

See [Browser Requirements](#) on page 33.

Cloudera Navigator Supported CDH and Managed Service Versions

This section describes the CDH and managed service versions supported by the Cloudera Navigator auditing and metadata features.

Cloudera Navigator Auditing

This section describes the audited operations and service versions supported by Cloudera Navigator auditing.

Component	Operations (For details, see Service Audit Events).	Minimum Supported Service Version
HDFS	<ul style="list-style-type: none"> Operations that access or modify a file's or directory's data or metadata Operations denied due to lack of privileges 	CDH 5.0.0
HBase	<ul style="list-style-type: none"> In simple authentication mode, if the HBase Secure RPC Engine property is <code>false</code> (the default), the username in log events is <code>UNKNOWN</code>. To see a meaningful username: <ol style="list-style-type: none"> Click the HBase service. Click the Configuration tab. Select Service-wide > Security. Set the HBase Secure RPC Engine property to <code>true</code>. Save the change and restart the service. 	CDH 5.0.0
Hive	<ul style="list-style-type: none"> Operations (except grant, revoke, and metadata access only) sent to HiveServer2 Operations denied due to lack of privileges <p>Limitations:</p> <ul style="list-style-type: none"> Actions taken against Hive using the Hive CLI are <i>not</i> audited. Therefore if you have enabled auditing you should disable the Hive CLI to prevent actions against Hive that are not audited. In simple authentication mode, the username in log events is the username passed in the HiveServer2 connect command. If you do not pass a username in the connect command, the username in log events is anonymous. 	CDH 5.0.0
Hue	<ul style="list-style-type: none"> Operations (except grant, revoke, and metadata access only) sent through the Beeswax Server 	CDH 5.0.0
	<ul style="list-style-type: none"> User operations such as log in, log out, add and remove user, add and remove LDAP group, add and remove user from LDAP group 	CDH 5.5.0
Impala	<ul style="list-style-type: none"> Queries denied due to lack of privileges Queries that pass analysis 	CDH 5.0.0
Sentry	<ul style="list-style-type: none"> Operations sent to the HiveServer2 and Hive Metastore Server roles and Impala service Adding and deleting roles, assigning roles to groups and removing roles from groups, creating and deleting privileges, granting and revoking privileges Operations denied due to lack of privileges <p>You do not directly configure the Sentry service for auditing. Instead, when you configure the Hive and Impala services for auditing, grant, revoke, and metadata operations appear in the Hive or Impala service audit logs.</p>	CDH 5.1.0
Solr	<ul style="list-style-type: none"> Index creation and deletion Schema and configuration file modification Index, service, document tag access 	CDH 5.4.0

Cloudera Navigator Metadata

This section describes the CDH and managed service versions supported by the Cloudera Navigator metadata feature.

Component	Minimum Supported Version
HDFS. However, federated HDFS is <i>not supported</i> .	CDH 5.0.0
HiveServer2	CDH 5.0.0
Impala	CDH 5.4.0
MapReduce (v1 or v2)	CDH 5.0.0
Oozie. Supported actions: <ul style="list-style-type: none"> • 2.4 - map-reduce, pig, hive, hive2, sqoop • 2.3 and lower - map-reduce, pig, hive, sqoop Unsupported actions include email, shell, and ssh.	CDH 5.0.0
Pig	CDH 5.0.0
Spark	CDH 5.4.0
Spark 2	CDH 5.8.0, Cloudera Manager 6.0
Sqoop 1. All Cloudera connectors are supported.	CDH 5.0.0
YARN	CDH 5.0.0

Product Compatibility Matrix for Cloudera Navigator Encryption

Cloudera Navigator encryption comprises several components.

See below for the individual compatibility matrices for each component:

Cloudera Navigator Key Trustee Server

Because of a change in the ports used by Key Trustee Server, Navigator Encrypt versions lower than 3.7 and Key Trustee KMS versions lower than 5.4 are not supported in Key Trustee Server 5.4 and higher.

Key Trustee Server: Recommended Hardware and Supported Distributions

Recommended Hardware and Supported Distributions

Key Trustee Server must be installed on a dedicated server or virtual machine (VM) that is not used for any other purpose. The backing PostgreSQL database must be installed on the same host as the Key Trustee Server, and must not be shared with any other services. For high availability, the active and passive Key Trustee Servers must not share physical resources. See [Resource Planning for Data at Rest Encryption](#) for more information.

The recommended minimum hardware specifications are as follows:

- Processor: 1 GHz 64-bit quad core
- Memory: 8 GB RAM
- Storage: 20 GB on moderate- to high-performance disk drives

Table 13: Cloudera Navigator Key Trustee Server Compatibility Matrix

Cloudera Navigator Key Trustee Server Version	Supported Operating Systems	Lowest Supported Cloudera Manager Version	Lowest Supported Cloudera Navigator Key HSM Versions	Supported Cloudera Navigator Key Trustee KMS Versions	Supported Cloudera Navigator Encrypt Versions
6.1.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 Oracle Linux: 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 	6.1.x	6.1.x	6.0.x, 6.1.x	6.0.x, 6.1.x
6.0.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.9, 6.8 Oracle Linux: 7.4, 7.3, 7.2, 6.9, 6.8 	6.0.x	6.0.x	6.0.x	6.0.x

Cloudera Navigator Key Trustee KMS

Key Trustee KMS: Recommended Hardware and Supported Distributions

The recommended minimum hardware specifications are as follows:

- Processor: 2 GHz 64-bit quad core
- Memory: 16 GB RAM
- Storage: 40 GB on moderate- to high-performance disk drives

The Key Trustee KMS workload is CPU-intensive. Cloudera recommends using machines with capabilities equivalent to your NameNode hosts, with Intel CPUs that support [AES-NI](#) for optimum performance. Also, Cloudera strongly recommends that you enable TLS for both the HDFS and the Key Trustee KMS services to prevent the passage of plain text key material between the KMS and HDFS data nodes.

Table 14: Cloudera Navigator Key Trustee KMS Compatibility Matrix

Cloudera Navigator Key Trustee KMS Version	Supported Operating Systems	Supported Key Trustee Server Versions	Lowest Supported Cloudera Manager Version	Supported CDH Versions
6.2.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 Oracle Linux: 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 SLES: 12 SP3, 12 SP2 	6.2.x, 6.1.x, 6.0.x	6.2.x	6.2.x, 6.1.x, 6.0.x

Cloudera Navigator Key Trustee KMS Version	Supported Operating Systems	Supported Key Trustee Server Versions	Lowest Supported Cloudera Manager Version	Supported CDH Versions
	<ul style="list-style-type: none"> Ubuntu: 16.04 (Xenial), 18 (Bionic) 			
6.1.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 Oracle Linux: 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 SLES: 12 SP3, 12 SP2 Ubuntu: 16.04 (Xenial) 	6.1.x, 6.0.x	6.1.x	6.1.x, 6.0.x
6.0.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.9, 6.8 Oracle Linux: 7.4, 7.3, 7.2, 6.8 SLES: 12 SP3, 12 SP2 Ubuntu: 16.04 LTS (Xenial) 	6.0.x	6.0.x	6.0.x

Cloudera Navigator HSM KMS

Navigator HSM KMS: Recommended Hardware and Supported Distributions

The recommended minimum hardware specifications are as follows:

- Processor: 2 GHz 64-bit quad core
- Memory: 16 GB RAM
- Storage: 40 GB on moderate- to high-performance disk drives

Supported HSM devices:

- SafeNet Luna
 - HSM software version: 6.2.2-5
 - HSM firmware version: 6.10.9
 - Client: 6.2.2
- Thales nSolo, nConnect
 - Server version: 3.67.11cam4
 - Firmware: 2.65.2
 - Security World Version: 12.30

Table 15: HSM KMS Compatibility Matrix

HSM KMS Version	Supported Operating Systems	Lowest Supported Cloudera Manager Version	Supported CDH Versions
6.2.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 	6.2.x	6.2.x, 6.1.x, 6.0.x
6.1.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 	6.1.x	6.1.x, 6.0.x
6.0.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.5, 7.4, 7.3, 7.2, 6.9, 6.8 	6.0.x	6.0.x

Cloudera Navigator Key HSM

Cloudera Navigator Key HSM must be installed on the same host as Key Trustee Server. Although Key HSM is compatible across all versions of Key Trustee Server, Cloudera strongly recommends also upgrading Key HSM after you upgrade Key Trustee Server.

Key HSM: Recommended Hardware and Supported Distributions

The following are prerequisites for installing Navigator Key HSM:

- Oracle Java Runtime Environment (JRE) 7 or higher with Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files:
 - [JCE for Java SE 7](#)
 - [JCE for Java SE 8](#)



Note: JDK 1.8u161 and higher enable unlimited strength encryption by default, and do not require policy files.

- A supported HSM device:
 - SafeNet Luna
 - HSM firmware version: 6.2.1
 - HSM software version: 5.2.3-1
 - SafeNet KeySecure
 - HSM firmware version: 6.2.1
 - HSM software version: 8.0.1, 8.1.0, 8.7.0
 - Thales nSolo, nConnect
 - HSM firmware version: 11.4.0
 - Client software version: 2.28.9cam136
 - AWS CloudHSM
 - Client software version: 1.1.1
- Key Trustee Server 3.8 or higher



Important: You must install Key HSM on the same host as Key Trustee Server.

Root access is required to install Navigator Key HSM.

Table 16: Cloudera Navigator Key HSM Compatibility Matrix

Cloudera Navigator Key HSM Version	Supported Operating Systems	Lowest Supported Key Trustee Server Version
6.1.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8 	6.1.x, 6.0.x
6.0.x	<ul style="list-style-type: none"> RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.9, 6.8 	6.0.x

Cloudera Navigator Encrypt



Note: Cloudera Enterprise, with the exception of Cloudera Navigator Encrypt, is supported on platforms with Security-Enhanced Linux (SELinux) enabled and in enforcing mode. Cloudera is not responsible for SELinux policy development, support, or enforcement. If you experience issues running Cloudera software with SELinux enabled, contact your OS provider for assistance.

If you are using SELinux in enforcing mode, Cloudera Support can request that you disable SELinux or change the mode to permissive to rule out SELinux as a factor when investigating reported issues.

Supported command-line interpreters:

- sh (Bourne)
- bash (Bash)
- dash (Ubuntu)



Note: Navigator Encrypt does not support installation or use in chroot environments.

Network Requirements

For new Navigator Key Trustee Server installations, Navigator Encrypt initiates TCP traffic over port 11371 (HTTPS) to the Key Trustee Server.

For upgrades, Navigator Encrypt initiates TCP traffic over ports 80 (HTTP) and 443 (HTTPS) to the Navigator Key Trustee Server.

Internet Access

You must have an active connection to the Internet to download many package dependencies, unless you have internal repositories or mirrors containing the dependent packages.

Maintenance Window

Data is not accessible during the encryption process. Plan for system downtime during installation and configuration.

Administrative Access

To enforce a high level of security, all Navigator Encrypt commands require administrative (root) access (including installation and configuration). If you do not have administrative privileges on your server, contact your system administrator before proceeding.

Package Dependencies

Navigator Encrypt requires these packages, which are resolved by your distribution package manager during installation:

- dkms
- keyutils
- ecryptfs-utils
- libkeytrustee
- navencrypt-kernel-module
- openssl
- lsof
- gcc
- cryptsetup

These packages may have other dependencies that are also resolved by your package manager. Installation works with gcc, gcc3, and gcc4.

Table 17: Cloudera Navigator Encrypt Compatibility Matrix

Cloudera Navigator Encrypt Version	Supported Operating Systems	Supported Key Trustee Server Versions
6.2.x	<ul style="list-style-type: none">• RHEL and CentOS: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8• Oracle Linux: 7.6, 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8• SLES: 12 SP2, SP3• Ubuntu: 16.04 LTS (Xenial), 18 (Bionic)	6.2.x, 6.1.x, 6.0.x
6.1.x	<ul style="list-style-type: none">• RHEL and CentOS: 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8• Oracle Linux: 7.5, 7.4, 7.3, 7.2, 6.10, 6.9, 6.8• SLES: 12 SP2, SP3• Ubuntu: 16.04 LTS (Xenial)	6.1.x, 6.0.x
6.0.x	<ul style="list-style-type: none">• RHEL and CentOS: 7.5, 7.4, 7.3, 7.2, 6.9, 6.8• Oracle Linux: 7.4, 7.3, 7.2, 6.9, 6.8• SLES: 12 SP2, SP3• Ubuntu: 16.04 LTS (Xenial)	6.0.x

Version, Packaging, and Download Information

Version and download information for Cloudera Manager, CDH, Impala, and Search can be found in the HTML documentation on the website at [Cloudera Documentation](#). Select the release version number and go to the HTML version of the Release Guide.

Cloudera Manager 6 Version and Download Information

Cloudera Manager is available in the following releases:

- Cloudera Manager 6.2.0 is the current release of Cloudera Manager 6.
- Cloudera Manager 6.1.1 is the previous release of Cloudera Manager.

The 64-bit packages listed here support both Cloudera Express with its extensive set of monitoring and management features, and Cloudera Enterprise with additional functionality. A 60-day trial can be enabled to provide access to the full set of Cloudera Enterprise Cloudera Enterprise features. Cloudera Enterprise can be enabled permanently with the appropriate license. To obtain a Cloudera Enterprise license, fill in this [form](#) or call 866-843-7207.

Cloudera Manager 6.2.0

Release Date: March 29, 2019

Documentation:

- [Cloudera Manager 6.2.0 Release Notes](#) on page 137
- [Cloudera Installation Guide](#)
- [Cloudera Enterprise Upgrade Guide](#)

Table 18: Repositories

Type	Location (baseurl)	Repo File
RHEL 7 Compatible	https://archive.cloudera.com/cm6/6.2.0/redhat7/yum/	cloudera-manager.repo
RHEL6 Compatible	https://archive.cloudera.com/cm6/6.2.0/redhat6/yum/	cloudera-manager.repo
SLES 12	https://archive.cloudera.com/cm6/6.2.0/sles12/yum/	cloudera-manager.repo
Ubuntu Xenial (18.04)	https://archive.cloudera.com/cm6/6.2.0/ubuntu1804/apt/	cloudera-manager.list
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cm6/6.2.0/ubuntu1604/apt/	cloudera-manager.list

Cloudera Manager 6.1.1

Release Date: February 20, 2019

Documentation:

- [Cloudera Manager 6.1.1 Release Notes](#) on page 146
- [Cloudera Installation Guide](#)
- [Cloudera Enterprise Upgrade Guide](#)

Table 19: Repositories

Type	Location (baseurl)	Repo File
RHEL 7 Compatible	https://archive.cloudera.com/cm6/6.1.1/redhat7/yum/	cloudera-manager.repo
RHEL6 Compatible	https://archive.cloudera.com/cm6/6.1.1/redhat6/yum/	cloudera-manager.repo
SLES 12	https://archive.cloudera.com/cm6/6.1.1/sles12/yum/	cloudera-manager.repo

Type	Location (baseurl)	Repo File
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cm6/6.1.1/ubuntu1604/apt/	cloudera-manager.list

Cloudera Manager 6.1.0

Release Date: December 18, 2018

Documentation:

- [Cloudera Manager 6.1.0 Release Notes](#) on page 148
- [Cloudera Installation Guide](#)
- [Cloudera Enterprise Upgrade Guide](#)

Table 20: Repositories

Type	Location (baseurl)	Repo File
RHEL 7 Compatible	https://archive.cloudera.com/cm6/6.1.0/redhat7/yum/	cloudera-manager.repo
RHEL6 Compatible	https://archive.cloudera.com/cm6/6.1.0/redhat6/yum/	cloudera-manager.repo
SLES 12	https://archive.cloudera.com/cm6/6.1.0/sles12/yum/	cloudera-manager.repo
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cm6/6.1.0/ubuntu1604/apt/	cloudera-manager.list

Cloudera Manager 6.0.1

Release Date: October 2018

Documentation:

- [Cloudera Manager 6.0.1 Release Notes](#) on page 161
- [Cloudera Installation Guide](#)
- [Cloudera Enterprise Upgrade Guide](#)

Table 21: Repositories

Type	Location (baseurl)	Repo File
RHEL 7 Compatible	https://archive.cloudera.com/cm6/6.0.1/redhat7/yum/	cloudera-manager.repo
RHEL6 Compatible	https://archive.cloudera.com/cm6/6.0.1/redhat6/yum/	cloudera-manager.repo
SLES 12	https://archive.cloudera.com/cm6/6.0.1/sles12/yum/	cloudera-manager.repo
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cm6/6.0.1/ubuntu1604/apt/	cloudera-manager.list

Cloudera Manager 6.0.0

Release Date: July 2018

Documentation:

- [Cloudera Manager 6.0.0 Release Notes](#) on page 165

- [Cloudera Installation Guide](#)
- [Cloudera Enterprise Upgrade Guide](#)

Table 22: Repositories

Type	Location (<code>baseurl1</code>)	Repo File
RHEL 7 Compatible	https://archive.cloudera.com/cm6/6.0.0/redhat7/yum/	cloudera-manager.repo
RHEL6 Compatible	https://archive.cloudera.com/cm6/6.0.0/redhat6/yum/	cloudera-manager.repo
SLES 12	https://archive.cloudera.com/cm6/6.0.0/sles12/yum/	cloudera-manager.repo
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cm6/6.0.0/ubuntu1604/apt/	cloudera-manager.list

CDH 6 Version, Packaging, and Download Information

For installation instructions for CDH, see [Cloudera Installation](#).

CDH is available in the following releases:

- CDH 6 is based on Apache Hadoop 3. For more information, see [CDH 6 Packaging Information](#) on page 49.
- CDH 5 is based on Apache Hadoop 2.3.0 or later. For information on the exact Apache Hadoop version included in each CDH 5 version, see [CDH 5 Packaging and Tarball Information](#).

CDH 6 Download Information

These topics describe download information for CDH 6.

CDH 6.2.x Download Information

The section describes download information for CDH 6.2.x. For a different release, see [CDH 6 Download Information](#) on page 47.

CDH 6.2.0

Release Date: March 28, 2019

Repository Type	Location (<code>baseurl1</code>)
Parcels	https://archive.cloudera.com/cdh6/6.2.0/parcels/
RHEL 7 Compatible	https://archive.cloudera.com/cdh6/6.2.0/redhat7/yum/
RHEL 6 Compatible	https://archive.cloudera.com/cdh6/6.2.0/redhat6/yum/
SLES 12	https://archive.cloudera.com/cdh6/6.2.0/sles12/yum/
Ubuntu Bionic (18.04)	https://archive.cloudera.com/cdh6/6.2.0/ubuntu1804/apt/
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cdh6/6.2.0/ubuntu1604/apt/

CDH 6.1.x Download Information

The section describes download information for CDH 6.1.x. For a different release, see [CDH 6 Download Information](#) on page 47.

CDH 6.1.1

Release Date: February 20, 2019

Repository Type	Location (<code>baseurl</code>)
Parcels	https://archive.cloudera.com/cdh6/6.1.1/parcels/
RHEL 7 Compatible	https://archive.cloudera.com/cdh6/6.1.1/redhat7/yum/
RHEL 6 Compatible	https://archive.cloudera.com/cdh6/6.1.1/redhat6/yum/
SLES 12	https://archive.cloudera.com/cdh6/6.1.1/sles12/yum/
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cdh6/6.1.1/ubuntu1604/apt/

CDH 6.1.0

Release Date: December 18, 2018

Repository Type	Location (<code>baseurl</code>)
Parcels	https://archive.cloudera.com/cdh6/6.1.0/parcels/
RHEL 7 Compatible	https://archive.cloudera.com/cdh6/6.1.0/redhat7/yum/
RHEL 6 Compatible	https://archive.cloudera.com/cdh6/6.1.0/redhat6/yum/
SLES 12	https://archive.cloudera.com/cdh6/6.1.0/sles12/yum/
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cdh6/6.1.0/ubuntu1604/apt/

CDH 6.0.x Download Information

The section describes download information for CDH 6.0.x. For a different release, see [CDH 6 Download Information](#) on page 47.

CDH 6.0.1

Release Date: October 11, 2018

Repository Type	Location (<code>baseurl</code>)
Parcels	https://archive.cloudera.com/cdh6/6.0.1/parcels/
RHEL 7 Compatible	https://archive.cloudera.com/cdh6/6.0.1/redhat7/yum/
RHEL 6 Compatible	https://archive.cloudera.com/cdh6/6.0.1/redhat6/yum/
SLES 12	https://archive.cloudera.com/cdh6/6.0.1/sles12/yum/
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cdh6/6.0.1/ubuntu1604/apt/

CDH 6.0.0

Release Date: August 30, 2018

Repository Type	Location (<code>baseurl</code>)
Parcels	https://archive.cloudera.com/cdh6/6.0.0/parcels/
RHEL 7 Compatible	https://archive.cloudera.com/cdh6/6.0.0/redhat7/yum/
RHEL 6 Compatible	https://archive.cloudera.com/cdh6/6.0.0/redhat6/yum/
SLES 12	https://archive.cloudera.com/cdh6/6.0.0/sles12/yum/
Ubuntu Xenial (16.04)	https://archive.cloudera.com/cdh6/6.0.0/ubuntu1604/apt/

CDH 6 Packaging Information

Each CDH release series is made up of a collection of CDH project packages that are known to work together. The package version numbers of the CDH projects in each CDH release are listed in the following table.



Important: To see the details of all the changes and bug-fixes for a given component in a given release, make sure you read the Changes information as well as the Release Notes, following the links in the tables below.

Component Versions

For a complete list of the individual component versions, see the `manifest.json` file in the `parcels` directory for your chosen release. The component version numbers are the same in parcels as the package distributions for each release.

1. Go to <https://archive.cloudera.com/cdh6/>.
2. Click through to the `parcels` directory for your release (for example, <https://archive.cloudera.com/cdh6/6.2.0/parcels/>).
3. Open the `manifest.json` file.

For each component, the `pkg_version` is a concatenation of `<component_base_version>+<cdh_version>+0` where:

- `<component-base_version>` is the base version of the open source component included in the CDH package.
- `<cdh_version>` is the version of the CDH package.

For example, in <https://archive.cloudera.com/cdh6/6.0.0/parcels/>, this entry for Hadoop shows that the upstream version is 3.0.0, the CDH version is 6.0.0.

```
{
  "name": "hadoop",
  "pkg_release": "1.cdh6.0.0.p.284270",
  "pkg_version": "3.0.0+cdh6.0.0+0",
  "version": "3.0.0-cdh6.0.0-SNAPSHOT"
},
```

Build and Release Numbering

If you are installing CDH 6 with a package manager, you will also see build and release information as part of the file name. The build and package release fields follow the patch level: for example, `hbase-2.0.0+cdh6.0.0+0-1.cdh6.0.0.p.284270.e17.x86_64.rpm`. The suffix `-1.cdh6.0.0.p.284270.e17.x86_64` represents:

- the base of the release field (1)
- the CDH release (cdh6.0.0)
- a legacy patch identifier (p)
- a unique CDH build number (284270)
- the distribution (e17 = RHEL 7 compatible, e16 = RHEL 6 compatible). SLES 12 packages omit this.
- the processor architecture (x86_64, noarch, i386, amd_64). noarch means that the packages are not architecture-specific.

External Documentation



Note: This page contains references to CDH 5 components or features that have been removed from CDH 6. These references are only applicable if you are managing a CDH 5 cluster with Cloudera Manager 6. For more information, see [Deprecated Items](#) on page 667.

Cloudera provides documentation for CDH as a whole, whether your CDH cluster is managed by Cloudera Manager or not. In addition, you may find it useful to refer to documentation for the individual components included in CDH. Where possible, these links point to the main documentation for a project, in the Cloudera release archive. This ensures that

you are looking at the correct documentation for the version of a project included in CDH. Otherwise, the links may point to the project's main site.

- [Apache Avro](#)
- [Apache Crunch](#)
- [Apache Flume](#)
- [Apache Hadoop](#)
- [Apache HBase](#)
- [Apache Hive](#)
- [Hue](#)
- [Kite](#)
- [Apache Oozie](#)
- [Apache Parquet](#)
- [Apache Pig](#)
- [Apache Sentry](#)
- [Apache Solr](#)
- [Apache Spark](#)
- [Apache Sqoop](#)
- [Apache ZooKeeper](#)

CDH 6.2.x Packaging

The package version numbers of the projects comprising each CDH 6.2.x release are listed in the following tables. For the same type of information for other CDH releases, see [CDH 6 Packaging Information](#) on page 49. To view the overall release notes for CDH 6, see [CDH 6 Release Notes](#) on page 177.



Important: To see the details of all the changes and bug-fixes for a given component in a given release, make sure you read the Changes information as well as the Release Notes, following the links in the tables below.

CDH 6.2.0 Packaging

Component	Component Version	Changes Information
Apache Avro	1.8.2	Changes
Apache Flume	1.9.0	Changes
Apache Hadoop	3.0.0	Changes
Apache HBase	2.1.2	Changes
HBase Indexer	1.5	Changes
Apache Hive	2.1.1	Changes
Hue	4.3.0	Changes
Apache Impala	3.2.0	Changes
Apache Kafka	2.1.0	Changes
Kite SDK	1.0.0	
Apache Kudu	1.9.0	Changes
Apache Solr	7.4.0	Changes
Apache Oozie	5.1.0	Changes
Apache Parquet	1.9.0	Changes

Component	Component Version	Changes Information
Parquet-format	2.3.1	Changes
Apache Pig	0.17.0	Changes
Apache Sentry	2.1.0	Changes
Apache Spark	2.4.0	Changes
Apache Sqoop	1.4.7	Changes
Apache ZooKeeper	3.4.5	Changes

CDH 6.1.x Packaging

The package version numbers of the projects comprising each CDH 6.1.x release are listed in the following tables. For the same type of information for other CDH releases, see [CDH 6 Packaging Information](#) on page 49. To view the overall release notes for CDH 6, see [CDH 6 Release Notes](#) on page 177.



Important: To see the details of all the changes and bug-fixes for a given component in a given release, make sure you read the Changes information as well as the Release Notes, following the links in the tables below.

CDH 6.1.1 Packaging

Component	Component Version	Changes Information
Apache Avro	1.8.2	Changes
Apache Flume	1.8.0	Changes
Apache Hadoop	3.0.0	Changes
Apache HBase	2.1.1	Changes
HBase Indexer	1.5	Changes
Apache Hive	2.1.1	Changes
Hue	4.3.0	Changes
Apache Impala	3.1.0	Changes
Apache Kafka	2.0	Changes
Kite SDK	1.0.0	
Apache Kudu	1.8.0	Changes
Apache Solr	7.4	Changes
Apache Oozie	5.0.0	Changes
Apache Parquet	1.9.0	Changes
Parquet-format	2.3.1	Changes
Apache Pig	0.17.0	Changes
Apache Sentry	2.1.0	Changes
Apache Spark	2.4	Changes
Apache Sqoop	1.4.7	Changes
Apache ZooKeeper	3.4.5	Changes

CDH 6.1.0 Packaging

Component	Component Version	Changes Information
Apache Avro	1.8.2	Changes
Apache Flume	1.8.0	Changes
Apache Hadoop	3.0.0	Changes
Apache HBase	2.1.1	Changes
HBase Indexer	1.5	Changes
Apache Hive	2.1.1	Changes
Hue	4.3.0	Changes
Apache Impala	3.1.0	Changes
Apache Kafka	2.0	Changes
Kite SDK	1.0.0	
Apache Kudu	1.8.0	Changes
Apache Solr	7.4	Changes
Apache Oozie	5.0.0	Changes
Apache Parquet	1.9.0	Changes
Parquet-format	2.3.1	Changes
Apache Pig	0.17.0	Changes
Apache Sentry	2.1.0	Changes
Apache Spark	2.4	Changes
Apache Sqoop	1.4.7	Changes
Apache ZooKeeper	3.4.5	Changes

CDH 6.0.x Packaging

The package version numbers of the projects comprising each CDH 6.0.x release are listed in the following tables. For the same type of information for other CDH releases, see [CDH 6 Packaging Information](#) on page 49. To view the overall release notes for CDH 6, see [CDH 6 Release Notes](#) on page 177.



Important: To see the details of all the changes and bug-fixes for a given component in a given release, make sure you read the Changes information as well as the Release Notes, following the links in the tables below.

CDH 6.0.1 Packaging

Component	Component Version	Changes Information
Apache Avro	1.8.2	Changes
Apache Flume	1.8.0	Changes
Apache Hadoop	3.0.0	Changes
Apache HBase	2.0.2	Changes
HBase Indexer	1.5	Changes
Apache Hive	2.1.1	Changes

Component	Component Version	Changes Information
Hue	4.2.0	Changes
Apache Impala	3.0.0	Changes
Apache Kafka	1.0.1	Changes
Kite SDK	1.0.0	
Apache Kudu	1.6.0	Changes
Apache Solr	7.0.0	Changes
Apache Oozie	5.0.0	Changes
Apache Parquet	1.9.0	Changes
Parquet-format	2.3.1	Changes
Apache Pig	0.17.0	Changes
Apache Sentry	2.0.0	Changes
Apache Spark	2.2.0	Changes
Apache Sqoop	1.4.7	Changes
Apache ZooKeeper	3.4.5	Changes

CDH 6.0.0 Packaging

Component	Component Version	Changes Information
Apache Avro	1.8.2	Changes
Apache Flume	1.8.0	Changes
Apache Hadoop	3.0.0	Changes
Apache HBase	2.0.0	Changes
HBase Indexer	1.5	Changes
Apache Hive	2.1.1	Changes
Hue	4.2.0	Changes
Apache Impala	3.0.0	Changes
Apache Kafka	1.0.1	Changes
Kite SDK	1.0.0	
Apache Kudu	1.6.0	Changes
Apache Solr	7.0.0	Changes
Apache Oozie	5.0.0	Changes
Apache Parquet	1.9.0	Changes
Parquet-format	2.3.1	Changes
Apache Pig	0.17.0	Changes
Apache Sentry	2.0.0	Changes
Apache Spark	2.2.0	Changes
Apache Sqoop	1.4.7	Changes

Component	Component Version	Changes Information
Apache ZooKeeper	3.4.5	Changes

Using the CDH 6 Maven Repository

If you want to build applications or tools for use with CDH 6 components and you are using Maven or Ivy for dependency management, you can pull the CDH 6 artifacts from the Cloudera Maven repository. The repository is available at <https://repository.cloudera.com/artifactory/cloudera-repos/>.



Important: When you build an application JAR, *do not* include CDH JARs, because they are already provided. If you do, upgrading CDH can break your application. To avoid this situation, set the Maven dependency scope to provided. If you have already built applications which include the CDH JARs, update the dependency to set scope to provided and recompile.

The following is a sample POM (`pom.xml`) file:

```
<project xmlns="http://maven.apache.org/POM/4.0.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0
  http://maven.apache.org/maven-v4_0_0.xsd">
  <repositories>
    <repository>
      <id>cloudera</id>
      <url>https://repository.cloudera.com/artifactory/cloudera-repos/</url>
    </repository>
  </repositories>
</project>
```

Maven Artifacts for CDH 6.2.x Releases

CDH 6.2.0

Coming soon.

Maven Artifacts for CDH 6.1.x Releases

CDH 6.1.1

The following table lists the project name, groupId, artifactId, and version required to access each CDH artifact.

Project	groupId	artifactId	version
Apache Avro	org.apache.avro	avro	1.8.2-cdh6.1.1
	org.apache.avro	avro-compiler	1.8.2-cdh6.1.1
	org.apache.avro	avro-guava-dependencies	1.8.2-cdh6.1.1
	org.apache.avro	avro-ipc	1.8.2-cdh6.1.1
	org.apache.avro	avro-mapred	1.8.2-cdh6.1.1
	org.apache.avro	avro-maven-plugin	1.8.2-cdh6.1.1
	org.apache.avro	avro-protobuf	1.8.2-cdh6.1.1
	org.apache.avro	avro-service-archetype	1.8.2-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.avro	avro-thrift	1.8.2-cdh6.1.1
	org.apache.avro	avro-tools	1.8.2-cdh6.1.1
	org.apache.avro	trevni-avro	1.8.2-cdh6.1.1
	org.apache.avro	trevni-core	1.8.2-cdh6.1.1
Apache Crunch	org.apache.crunch	crunch-archetype	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-contrib	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-core	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-examples	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-hbase	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-hive	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-scrunch	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-spark	0.11.0-cdh6.1.1
	org.apache.crunch	crunch-test	0.11.0-cdh6.1.1
Apache Flume 1.x	org.apache.flume	flume-checkstyle	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-auth	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-config-filter-api	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-configuration	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-core	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-embedded-agent	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-environment-variable-config-filter	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-external-process-config-filter	1.8.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.flume	flume-ng-hadoop-credential-store-config-filter	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-node	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-sdk	1.8.0-cdh6.1.1
	org.apache.flume	flume-ng-tests	1.8.0-cdh6.1.1
	org.apache.flume	flume-tools	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-channels	flume-file-channel	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-channels	flume-jdbc-channel	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-channels	flume-kafka-channel	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-channels	flume-spillable-memory-channel	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-clients	flume-ng-log4jappender	1.8.0-cdh6.1.1
	org.apache.flume.flume- ng-sources	flume-avro-source	1.8.0-cdh6.1.1
	org.apache.flume.flume- ng-sources	flume-thrift-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-dataset-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-hdfs-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-hive-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-http-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-irc-sink	1.8.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.flume.flume-ng-sinks	flume-ng-hbase2-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-ng-kafka-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sinks	flume-ng-morphline-solr-sink	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sources	flume-jms-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sources	flume-kafka-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sources	flume-scribe-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sources	flume-taildir-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-ng-sources	flume-twitter-source	1.8.0-cdh6.1.1
	org.apache.flume.flume-shared	flume-shared-kafka	1.8.0-cdh6.1.1
	org.apache.flume.flume-shared	flume-shared-kafka-test	1.8.0-cdh6.1.1
Apache Hadoop	org.apache.hadoop	hadoop-aliyun	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-annotations	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-archive-logs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-archives	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-assemblies	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-auth	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-aws	3.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-azure	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-azure-datalake	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-build-tools	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-client	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-client-api	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-client-integration-tests	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-client-minicluster	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-client-runtime	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-cloud-storage	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-common	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-datajoin	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-distcp	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-extras	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-gridmix	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-hdfs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-hdfs-client	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-hdfs-httpfs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-hdfs-native-client	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-hdfs-nfs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-kafka	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-kms	3.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-mapreduce-client-app	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-common	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-core	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-hs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-hs-plugins	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-jobclient	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-nativetask	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-shuffle	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-client-uploader	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-mapreduce-examples	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-maven-plugins	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-minicluster	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-minikdc	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-nfs	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-openstack	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-resourceestimator	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-rumen	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-sls	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-streaming	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-api	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-applications-distributedshell	3.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-yarn-applications-unmanaged-am-launcher	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-client	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-common	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-registry	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-applicationhistoryservice	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-common	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-nodemanager	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-resourcemanager	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-router	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-sharedcachemanager	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-tests	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-timeline-pluginstorage	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase-tests	3.0.0-cdh6.1.1
	org.apache.hadoop	hadoop-yarn-server-web-proxy	3.0.0-cdh6.1.1
Apache HBase	org.apache.hbase	hbase-annotations	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-checkstyle	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-client	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-client-project	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-common	2.1.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hbase	hbase-endpoint	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-error-prone	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-examples	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-external-blockcache	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-hadoop-compat	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-hadoop2-compat	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-http	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-it	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-mapreduce	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-metrics	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-metrics-api	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-procedure	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-protocol	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-protocol-shaded	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-replication	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-resource-bundle	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-rest	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-rsgroup	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-server	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-shaded-client	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-shaded-client-byo-hadoop	2.1.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hbase	hbase-shaded-client-project	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-shaded-mapreduce	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-shell	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-spark	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-spark-it	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-testing-util	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-thrift	2.1.0-cdh6.1.1
	org.apache.hbase	hbase-zookeeper	2.1.0-cdh6.1.1
HBase Indexer	com.ngdata	hbase-indexer-all	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-cli	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-common	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-demo	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-dist	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-engine	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-model	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-morphlines	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-mr	1.5-cdh6.1.1
	com.ngdata	hbase-indexer-server	1.5-cdh6.1.1
	com.ngdata	hbase-sep-api	1.5-cdh6.1.1
	com.ngdata	hbase-sep-demo	1.5-cdh6.1.1
	com.ngdata	hbase-sep-impl	1.5-cdh6.1.1
	com.ngdata	hbase-sep-tools	1.5-cdh6.1.1
Apache Hive	org.apache.hive	hive-accumulo-handler	2.1.1-cdh6.1.1
	org.apache.hive	hive-ant	2.1.1-cdh6.1.1
	org.apache.hive	hive-beeline	2.1.1-cdh6.1.1
	org.apache.hive	hive-classification	2.1.1-cdh6.1.1
	org.apache.hive	hive-cli	2.1.1-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hive	hive-common	2.1.1-cdh6.1.1
	org.apache.hive	hive-contrib	2.1.1-cdh6.1.1
	org.apache.hive	hive-exec	2.1.1-cdh6.1.1
	org.apache.hive	hive-hbase-handler	2.1.1-cdh6.1.1
	org.apache.hive	hive-hplsql	2.1.1-cdh6.1.1
	org.apache.hive	hive-jdbc	2.1.1-cdh6.1.1
	org.apache.hive	hive-kryo-registrator	2.1.1-cdh6.1.1
	org.apache.hive	hive-llap-client	2.1.1-cdh6.1.1
	org.apache.hive	hive-llap-common	2.1.1-cdh6.1.1
	org.apache.hive	hive-llap-ext-client	2.1.1-cdh6.1.1
	org.apache.hive	hive-llap-server	2.1.1-cdh6.1.1
	org.apache.hive	hive-llap-tez	2.1.1-cdh6.1.1
	org.apache.hive	hive-metastore	2.1.1-cdh6.1.1
	org.apache.hive	hive-orc	2.1.1-cdh6.1.1
	org.apache.hive	hive-serde	2.1.1-cdh6.1.1
	org.apache.hive	hive-service	2.1.1-cdh6.1.1
	org.apache.hive	hive-service-rpc	2.1.1-cdh6.1.1
	org.apache.hive	hive-shims	2.1.1-cdh6.1.1
	org.apache.hive	hive-spark-client	2.1.1-cdh6.1.1
	org.apache.hive	hive-storage-api	2.1.1-cdh6.1.1
	org.apache.hive	hive-testutils	2.1.1-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.hive.hcatalog	hive-hcatalog-core	2.1.1-cdh6.1.1
	org.apache.hive.hcatalog	hive-hcatalog-pig-adapter	2.1.1-cdh6.1.1
	org.apache.hive.hcatalog	hive-hcatalog-server-extensions	2.1.1-cdh6.1.1
	org.apache.hive.hcatalog	hive-hcatalog-streaming	2.1.1-cdh6.1.1
	org.apache.hive.hcatalog	hive-webhcat	2.1.1-cdh6.1.1
	org.apache.hive.hcatalog	hive-webhcat-java-client	2.1.1-cdh6.1.1
	org.apache.hive.shims	hive-shims-0.23	2.1.1-cdh6.1.1
	org.apache.hive.shims	hive-shims-common	2.1.1-cdh6.1.1
	org.apache.hive.shims	hive-shims-scheduler	2.1.1-cdh6.1.1
Apache Kafka	org.apache.kafka	connect-api	2.0.0-cdh6.1.1
	org.apache.kafka	connect-basic-auth-extension	2.0.0-cdh6.1.1
	org.apache.kafka	connect-file	2.0.0-cdh6.1.1
	org.apache.kafka	connect-json	2.0.0-cdh6.1.1
	org.apache.kafka	connect-runtime	2.0.0-cdh6.1.1
	org.apache.kafka	connect-transforms	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-clients	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-examples	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-log4j-appender	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-streams	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-streams-examples	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-streams-scala_2.11	2.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.kafka	kafka-streams-scala_2.12	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-streams-test-utils	2.0.0-cdh6.1.1
	org.apache.kafka	kafka-tools	2.0.0-cdh6.1.1
	org.apache.kafka	kafka_2.11	2.0.0-cdh6.1.1
	org.apache.kafka	kafka_2.12	2.0.0-cdh6.1.1
Kite SDK	org.kitesdk	kite-data-core	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-crunch	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-hbase	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-hive	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-mapreduce	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-oozie	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-s3	1.0.0-cdh6.1.1
	org.kitesdk	kite-data-spark	1.0.0-cdh6.1.1
	org.kitesdk	kite-hadoop-compatibility	1.0.0-cdh6.1.1
	org.kitesdk	kite-maven-plugin	1.0.0-cdh6.1.1
	org.kitesdk	kite-minicluster	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-avro	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-core	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-hadoop-core	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-hadoop-parquet-avro	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-hadoop-rcfile	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-hadoop-sequencefile	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-json	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-maxmind	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-metrics-scalable	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-metrics-servlets	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-protobuf	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-saxon	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-solr-cell	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-solr-core	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-tika-core	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-tika-decompress	1.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.kitesdk	kite-morphlines-twitter	1.0.0-cdh6.1.1
	org.kitesdk	kite-morphlines-useragent	1.0.0-cdh6.1.1
	org.kitesdk	kite-tools	1.0.0-cdh6.1.1
Apache Kudu	org.apache.kudu	kudu-client	1.8.0-cdh6.1.1
	org.apache.kudu	kudu-client-tools	1.8.0-cdh6.1.1
	org.apache.kudu	kudu-flume-sink	1.8.0-cdh6.1.1
	org.apache.kudu	kudu-mapreduce	1.8.0-cdh6.1.1
	org.apache.kudu	kudu-spark2-tools_2.11	1.8.0-cdh6.1.1
	org.apache.kudu	kudu-spark2_2.11	1.8.0-cdh6.1.1
Apache Oozie	org.apache.oozie	oozie-client	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-core	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-examples	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-server	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-distcp	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-hcatalog	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-hive	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-hive2	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-oozie	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-pig	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-spark	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-sqoop	5.0.0-cdh6.1.1
	org.apache.oozie	oozie-sharelib-streaming	5.0.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.oozie	oozie-tools	5.0.0-cdh6.1.1
	org.apache.oozie.test	oozie-mini	5.0.0-cdh6.1.1
Apache Pig	org.apache.pig	pig	0.17.0-cdh6.1.1
	org.apache.pig	piggybank	0.17.0-cdh6.1.1
	org.apache.pig	pigsmoke	0.17.0-cdh6.1.1
	org.apache.pig	pigunit	0.17.0-cdh6.1.1
Cloudera Search	com.cloudera.search	search-crunch	1.0.0-cdh6.1.1
	com.cloudera.search	search-mr	1.0.0-cdh6.1.1
Apache Sentry	com.cloudera.cdh	solr-upgrade	1.0.0-cdh6.1.1
	org.apache.sentry	sentry-binding-hbase-indexer	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-hive	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-hive-common	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-hive-conf	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-hive-follower	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-kafka	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-binding-solr	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-core-common	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-core-model-db	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-core-model-indexer	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-core-model-kafka	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-core-model-solr	2.1.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.sentry	sentry-dist	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-hdfs-common	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-hdfs-dist	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-hdfs-namenode-plugin	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-hdfs-service	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-policy-common	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-policy-engine	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-provider-cache	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-provider-common	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-provider-db	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-provider-file	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-service-api	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-service-client	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-service-server	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-shaded	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-shaded-miscellaneous	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-spi	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-tests-hive	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-tests-kafka	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-tests-solr	2.1.0-cdh6.1.1
	org.apache.sentry	sentry-tools	2.1.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.sentry	solr-sentry-handlers	2.1.0-cdh6.1.1
Apache Solr	org.apache.lucene	lucene-analyzers-common	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-icu	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-kuromoji	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-morfologik	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-nori	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-opennlp	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-phonetic	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-smartcn	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-stempel	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-analyzers-uima	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-backward-codecs	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-benchmark	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-classification	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-codecs	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-core	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-demo	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-expressions	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-facet	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-grouping	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-highlighter	7.4.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.lucene	lucene-join	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-memory	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-misc	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-queries	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-queryparser	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-replicator	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-sandbox	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-spatial	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-spatial-extras	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-spatial3d	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-suggest	7.4.0-cdh6.1.1
	org.apache.lucene	lucene-test-framework	7.4.0-cdh6.1.1
	org.apache.solr	solr-analysis-extras	7.4.0-cdh6.1.1
	org.apache.solr	solr-analytics	7.4.0-cdh6.1.1
	org.apache.solr	solr-cell	7.4.0-cdh6.1.1
	org.apache.solr	solr-clustering	7.4.0-cdh6.1.1
	org.apache.solr	solr-core	7.4.0-cdh6.1.1
	org.apache.solr	solr-dataimporthandler	7.4.0-cdh6.1.1
	org.apache.solr	solr-dataimporthandler-extras	7.4.0-cdh6.1.1
	org.apache.solr	solr-jetty-customizations	7.4.0-cdh6.1.1
	org.apache.solr	solr-langid	7.4.0-cdh6.1.1
	org.apache.solr	solr-ltr	7.4.0-cdh6.1.1
	org.apache.solr	solr-prometheus-exporter	7.4.0-cdh6.1.1
	org.apache.solr	solr-security-util	7.4.0-cdh6.1.1
	org.apache.solr	solr-sentry-audit-logging	7.4.0-cdh6.1.1
	org.apache.solr	solr-solrj	7.4.0-cdh6.1.1
	org.apache.solr	solr-test-framework	7.4.0-cdh6.1.1
	org.apache.solr	solr-uima	7.4.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.solr	solr-velocity	7.4.0-cdh6.1.1
Apache Spark	org.apache.spark	spark-avro_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-catalyst_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-core_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-graphx_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-hadoop-cloud_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-hive_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-kvstore_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-launcher_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-mllib-local_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-mllib_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-network-common_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-network-shuffle_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-network-yarn_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-repl_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-sketch_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-sql-kafka-0-10_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-sql_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-streaming-flume-assembly_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-streaming-flume-sink_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-streaming-flume_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-streaming-kafka-0-10-assembly_2.11	2.4.0-cdh6.1.1

Project	groupId	artifactId	version
	org.apache.spark	spark-streaming-kafka-0-10_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-streaming_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-tags_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-unsafe_2.11	2.4.0-cdh6.1.1
	org.apache.spark	spark-yarn_2.11	2.4.0-cdh6.1.1
Apache Sqoop	org.apache.sqoop	sqoop	1.4.7-cdh6.1.1
Apache ZooKeeper	org.apache.zookeeper	zookeeper	3.4.5-cdh6.1.1

CDH 6.1.0

The following table lists the project name, groupId, artifactId, and version required to access each CDH artifact.

Project	groupId	artifactId	version
Apache Avro	org.apache.avro	avro	1.8.2-cdh6.1.0
	org.apache.avro	avro-compiler	1.8.2-cdh6.1.0
	org.apache.avro	avro-guava-dependencies	1.8.2-cdh6.1.0
	org.apache.avro	avro-ipc	1.8.2-cdh6.1.0
	org.apache.avro	avro-mapred	1.8.2-cdh6.1.0
	org.apache.avro	avro-maven-plugin	1.8.2-cdh6.1.0
	org.apache.avro	avro-protobuf	1.8.2-cdh6.1.0
	org.apache.avro	avro-service-archetype	1.8.2-cdh6.1.0
	org.apache.avro	avro-thrift	1.8.2-cdh6.1.0
	org.apache.avro	avro-tools	1.8.2-cdh6.1.0
	org.apache.avro	trevni-avro	1.8.2-cdh6.1.0
	org.apache.avro	trevni-core	1.8.2-cdh6.1.0

Project	groupId	artifactId	version
Apache Crunch	org.apache.crunch	crunch-archetype	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-contrib	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-core	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-examples	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-hbase	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-hive	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-scrunch	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-spark	0.11.0-cdh6.1.0
	org.apache.crunch	crunch-test	0.11.0-cdh6.1.0
Apache Flume 1.x	org.apache.flume	flume-checkstyle	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-auth	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-config-filter-api	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-configuration	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-core	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-embedded-agent	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-environment-variable-config-filter	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-external-process-config-filter	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-hadoop-credential-store-config-filter	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-node	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-sdk	1.8.0-cdh6.1.0
	org.apache.flume	flume-ng-tests	1.8.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.flume	flume-tools	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-channels	flume-file-channel	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-channels	flume-jdbc-channel	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-channels	flume-kafka-channel	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-channels	flume-spillable-memory-channel	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-clients	flume-ng-log4jappender	1.8.0-cdh6.1.0
	org.apache.flume.flume-legacy-sources	flume-avro-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-legacy-sources	flume-thrift-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-dataset-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-hdfs-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-hive-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-http-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-irc-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-ng-hbase2-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sinks	flume-ng-kafka-sink	1.8.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.flume.flume-ng-sinks	flume-ng-morphline-solr-sink	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sources	flume-jms-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sources	flume-kafka-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sources	flume-scribe-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sources	flume-taildir-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-ng-sources	flume-twitter-source	1.8.0-cdh6.1.0
	org.apache.flume.flume-shared	flume-shared-kafka	1.8.0-cdh6.1.0
	org.apache.flume.flume-shared	flume-shared-kafka-test	1.8.0-cdh6.1.0
Apache Hadoop	org.apache.hadoop	hadoop-aliyun	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-annotations	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-archive-logs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-archives	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-assemblies	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-auth	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-aws	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-azure	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-azure-datalake	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-build-tools	3.0.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-client	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-client-api	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-client-integration-tests	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-client-minicluster	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-client-runtime	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-cloud-storage	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-common	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-datajoin	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-distcp	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-extras	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-gridmix	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-hdfs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-hdfs-client	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-hdfs-httfs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-hdfs-native-client	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-hdfs-nfs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-kafka	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-kms	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-app	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-common	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-core	3.0.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-mapreduce-client-hs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-hs-plugins	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-jobclient	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-nativetask	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-shuffle	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-client-uploader	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-mapreduce-examples	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-maven-plugins	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-minicluster	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-minikdc	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-nfs	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-openstack	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-resourceestimator	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-rumen	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-sls	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-streaming	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-api	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-applications-distributedshell	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-applications-unmanaged-am-launcher	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-client	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-common	3.0.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-yarn-registry	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-applicationhistoryservice	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-common	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-nodemanager	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-resourcemanager	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-router	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-sharedcachemanager	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-tests	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-timeline-pluginstorage	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase-tests	3.0.0-cdh6.1.0
	org.apache.hadoop	hadoop-yarn-server-web-proxy	3.0.0-cdh6.1.0
Apache HBase	org.apache.hbase	hbase-annotations	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-checkstyle	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-client	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-client-project	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-common	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-endpoint	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-error-prone	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-examples	2.1.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hbase	hbase-external-blockcache	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-hadoop-compat	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-hadoop2-compat	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-http	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-it	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-mapreduce	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-metrics	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-metrics-api	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-procedure	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-protocol	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-protocol-shaded	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-replication	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-resource-bundle	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-rest	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-rsgroup	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-server	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-shaded-client	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-shaded-client-byo-hadoop	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-shaded-client-project	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-shaded-mapreduce	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-shell	2.1.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hbase	hbase-spark	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-spark-it	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-testing-util	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-thrift	2.1.0-cdh6.1.0
	org.apache.hbase	hbase-zookeeper	2.1.0-cdh6.1.0
HBase Indexer	com.ngdata	hbase-indexer-all	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-cli	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-common	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-demo	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-dist	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-engine	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-model	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-morphlines	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-mr	1.5-cdh6.1.0
	com.ngdata	hbase-indexer-server	1.5-cdh6.1.0
	com.ngdata	hbase-sep-api	1.5-cdh6.1.0
	com.ngdata	hbase-sep-demo	1.5-cdh6.1.0
	com.ngdata	hbase-sep-impl	1.5-cdh6.1.0
	com.ngdata	hbase-sep-tools	1.5-cdh6.1.0
Apache Hive	org.apache.hive	hive-accumulo-handler	2.1.1-cdh6.1.0
	org.apache.hive	hive-ant	2.1.1-cdh6.1.0
	org.apache.hive	hive-beeline	2.1.1-cdh6.1.0
	org.apache.hive	hive-classification	2.1.1-cdh6.1.0
	org.apache.hive	hive-cli	2.1.1-cdh6.1.0
	org.apache.hive	hive-common	2.1.1-cdh6.1.0
	org.apache.hive	hive-contrib	2.1.1-cdh6.1.0
	org.apache.hive	hive-exec	2.1.1-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hive	hive-hbase-handler	2.1.1-cdh6.1.0
	org.apache.hive	hive-hplsql	2.1.1-cdh6.1.0
	org.apache.hive	hive-jdbc	2.1.1-cdh6.1.0
	org.apache.hive	hive-kryo-registrator	2.1.1-cdh6.1.0
	org.apache.hive	hive-llap-client	2.1.1-cdh6.1.0
	org.apache.hive	hive-llap-common	2.1.1-cdh6.1.0
	org.apache.hive	hive-llap-ext-client	2.1.1-cdh6.1.0
	org.apache.hive	hive-llap-server	2.1.1-cdh6.1.0
	org.apache.hive	hive-llap-tez	2.1.1-cdh6.1.0
	org.apache.hive	hive-metastore	2.1.1-cdh6.1.0
	org.apache.hive	hive-orc	2.1.1-cdh6.1.0
	org.apache.hive	hive-serde	2.1.1-cdh6.1.0
	org.apache.hive	hive-service	2.1.1-cdh6.1.0
	org.apache.hive	hive-service-rpc	2.1.1-cdh6.1.0
	org.apache.hive	hive-shims	2.1.1-cdh6.1.0
	org.apache.hive	hive-spark-client	2.1.1-cdh6.1.0
	org.apache.hive	hive-storage-api	2.1.1-cdh6.1.0
	org.apache.hive	hive-testutils	2.1.1-cdh6.1.0
	org.apache.hive.hcatalog	hive-hcatalog-core	2.1.1-cdh6.1.0
	org.apache.hive.hcatalog	hive-hcatalog-pig-adapter	2.1.1-cdh6.1.0
	org.apache.hive.hcatalog	hive-hcatalog-server-extensions	2.1.1-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.hive.hcatalog	hive-hcatalog-streaming	2.1.1-cdh6.1.0
	org.apache.hive.hcatalog	hive-webhcat	2.1.1-cdh6.1.0
	org.apache.hive.hcatalog	hive-webhcat-java-client	2.1.1-cdh6.1.0
	org.apache.hive.shims	hive-shims-0.23	2.1.1-cdh6.1.0
	org.apache.hive.shims	hive-shims-common	2.1.1-cdh6.1.0
	org.apache.hive.shims	hive-shims-scheduler	2.1.1-cdh6.1.0
Apache Kafka	org.apache.kafka	connect-api	2.0.0-cdh6.1.0
	org.apache.kafka	connect-basic-auth-extension	2.0.0-cdh6.1.0
	org.apache.kafka	connect-file	2.0.0-cdh6.1.0
	org.apache.kafka	connect-json	2.0.0-cdh6.1.0
	org.apache.kafka	connect-runtime	2.0.0-cdh6.1.0
	org.apache.kafka	connect-transforms	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-clients	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-examples	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-log4j-append	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-streams	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-streams-examples	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-streams-scala_2.11	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-streams-scala_2.12	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-streams-test-utils	2.0.0-cdh6.1.0
	org.apache.kafka	kafka-tools	2.0.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.kafka	kafka_2.11	2.0.0-cdh6.1.0
	org.apache.kafka	kafka_2.12	2.0.0-cdh6.1.0
Kite SDK	org.kitesdk	kite-data-core	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-crunch	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-hbase	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-hive	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-mapreduce	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-oozie	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-s3	1.0.0-cdh6.1.0
	org.kitesdk	kite-data-spark	1.0.0-cdh6.1.0
	org.kitesdk	kite-hadoop-compatibility	1.0.0-cdh6.1.0
	org.kitesdk	kite-maven-plugin	1.0.0-cdh6.1.0
	org.kitesdk	kite-minicluster	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-avro	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-core	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-hadoop-core	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-hadoop-parquet-avro	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-hadoop-rcfile	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-hadoop-sequencefile	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-json	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-maxmind	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-metrics-scalable	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-metrics-servlets	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-protobuf	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-saxon	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-solr-cell	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-solr-core	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-tika-core	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-tika-decompress	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-twitter	1.0.0-cdh6.1.0
	org.kitesdk	kite-morphlines-useragent	1.0.0-cdh6.1.0
	org.kitesdk	kite-tools	1.0.0-cdh6.1.0
Apache Kudu	org.apache.kudu	kudu-client	1.8.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.kudu	kudu-client-tools	1.8.0-cdh6.1.0
	org.apache.kudu	kudu-flume-sink	1.8.0-cdh6.1.0
	org.apache.kudu	kudu-mapreduce	1.8.0-cdh6.1.0
	org.apache.kudu	kudu-spark2-tools_2.11	1.8.0-cdh6.1.0
	org.apache.kudu	kudu-spark2_2.11	1.8.0-cdh6.1.0
Apache Oozie	org.apache.oozie	oozie-client	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-core	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-examples	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-server	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-distcp	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-hcatalog	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-hive	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-hive2	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-oozie	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-pig	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-spark	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-sqoop	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-sharelib-streaming	5.0.0-cdh6.1.0
	org.apache.oozie	oozie-tools	5.0.0-cdh6.1.0
	org.apache.oozie.test	oozie-mini	5.0.0-cdh6.1.0
Apache Pig	org.apache.pig	pig	0.17.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.pig	piggybank	0.17.0-cdh6.1.0
	org.apache.pig	pigsmoke	0.17.0-cdh6.1.0
	org.apache.pig	pigunit	0.17.0-cdh6.1.0
Cloudera Search	com.cloudera.search	search-crunch	1.0.0-cdh6.1.0
	com.cloudera.search	search-mr	1.0.0-cdh6.1.0
	com.cloudera.cdh	solr-upgrade	1.0.0-cdh6.1.0
Apache Sentry	org.apache.sentry	sentry-binding-hbase-indexer	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-hive	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-hive-common	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-hive-conf	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-hive-follower	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-kafka	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-binding-solr	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-core-common	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-core-model-db	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-core-model-indexer	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-core-model-kafka	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-core-model-solr	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-dist	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-hdfs-common	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-hdfs-dist	2.1.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.sentry	sentry-hdfs-namenode-plugin	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-hdfs-service	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-policy-common	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-policy-engine	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-provider-cache	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-provider-common	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-provider-db	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-provider-file	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-service-api	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-service-client	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-service-server	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-shaded	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-shaded-miscellaneous	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-spi	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-tests-hive	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-tests-kafka	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-tests-solr	2.1.0-cdh6.1.0
	org.apache.sentry	sentry-tools	2.1.0-cdh6.1.0
	org.apache.sentry	solr-sentry-handlers	2.1.0-cdh6.1.0
Apache Solr	org.apache.lucene	lucene-analyzers-common	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-icu	7.4.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.lucene	lucene-analyzers-kuromoji	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-morfologik	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-nori	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-opennlp	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-phonetic	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-smartcn	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-stempel	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-analyzers-uima	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-backward-codecs	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-benchmark	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-classification	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-codecs	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-core	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-demo	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-expressions	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-facet	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-grouping	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-highlighter	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-join	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-memory	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-misc	7.4.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.lucene	lucene-queries	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-queryparser	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-replicator	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-sandbox	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-spatial	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-spatial-extras	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-spatial3d	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-suggest	7.4.0-cdh6.1.0
	org.apache.lucene	lucene-test-framework	7.4.0-cdh6.1.0
org.apache.solr	solr-analysis-extras		7.4.0-cdh6.1.0
org.apache.solr	solr-analytics		7.4.0-cdh6.1.0
org.apache.solr	solr-cell		7.4.0-cdh6.1.0
org.apache.solr	solr-clustering		7.4.0-cdh6.1.0
org.apache.solr	solr-core		7.4.0-cdh6.1.0
org.apache.solr	solr-dataimporthandler		7.4.0-cdh6.1.0
org.apache.solr	solr-dataimporthandler-extras		7.4.0-cdh6.1.0
org.apache.solr	solr-jetty-customizations		7.4.0-cdh6.1.0
org.apache.solr	solr-langid		7.4.0-cdh6.1.0
org.apache.solr	solr-ltr		7.4.0-cdh6.1.0
org.apache.solr	solr-prometheus-exporter		7.4.0-cdh6.1.0
org.apache.solr	solr-security-util		7.4.0-cdh6.1.0
org.apache.solr	solr-sentry-audit-logging		7.4.0-cdh6.1.0
org.apache.solr	solr-solrj		7.4.0-cdh6.1.0
org.apache.solr	solr-test-framework		7.4.0-cdh6.1.0
org.apache.solr	solr-uima		7.4.0-cdh6.1.0
org.apache.solr	solr-velocity		7.4.0-cdh6.1.0
Apache Spark	org.apache.spark	spark-avro_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-catalyst_2.11	2.4.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.spark	spark-core_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-graphx_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-hadoop-cloud_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-hive_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-kvstore_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-launcher_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-mllib-local_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-mllib_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-network-common_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-network-shuffle_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-network-yarn_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-repl_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-sketch_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-sql-kafka-0-10_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-sql_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming-flume-assembly_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming-flume-sink_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming-flume_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming-kafka-0-10-assembly_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming-kafka-0-10_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-streaming_2.11	2.4.0-cdh6.1.0

Project	groupId	artifactId	version
	org.apache.spark	spark-tags_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-unsafe_2.11	2.4.0-cdh6.1.0
	org.apache.spark	spark-yarn_2.11	2.4.0-cdh6.1.0
Apache Sqoop	org.apache.sqoop	sqoop	1.4.7-cdh6.1.0
Apache ZooKeeper	org.apache.zookeeper	zookeeper	3.4.5-cdh6.1.0

Maven Artifacts for CDH 6.0.x Releases**CDH 6.0.1**

The following table lists the project name, groupId, artifactId, and version required to access each CDH artifact.

Project	groupId	artifactId	version
Apache Hadoop	org.apache.hadoop	hadoop-aliyun	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-annotations	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-archive-logs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-archives	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-assemblies	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-auth	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-aws	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-azure	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-azure-datalake	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-build-tools	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-client	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-client-api	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-client-integration-tests	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-client-minicluster	3.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-client-runtime	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-cloud-storage	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-common	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-datajoin	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-distcp	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-extras	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-gridmix	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-hdfs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-hdfs-client	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-hdfs-httpfs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-hdfs-native-client	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-hdfs-nfs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-kafka	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-kms	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-app	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-common	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-core	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-hs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-hs-plugins	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-jobclient	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-nativetask	3.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-mapreduce-client-shuffle	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-client-uploader	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-mapreduce-examples	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-maven-plugins	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-minicluster	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-minikdc	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-nfs	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-openstack	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-resourceestimator	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-rumen	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-sls	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-streaming	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-api	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-applications-distributedshell	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-applications-unmanaged-am-launcher	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-client	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-common	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-registry	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-applicationhistoryservice	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-common	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-nodemanager	3.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-yarn-server-resourcemanager	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-router	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-sharedcachemanager	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-tests	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-timeline-pluginstorage	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase-tests	3.0.0-cdh6.0.1
	org.apache.hadoop	hadoop-yarn-server-web-proxy	3.0.0-cdh6.0.1
Apache Hive	org.apache.hive	hive-accumulo-handler	2.1.1-cdh6.0.1
	org.apache.hive	hive-ant	2.1.1-cdh6.0.1
	org.apache.hive	hive-beeline	2.1.1-cdh6.0.1
	org.apache.hive	hive-classification	2.1.1-cdh6.0.1
	org.apache.hive	hive-cli	2.1.1-cdh6.0.1
	org.apache.hive	hive-common	2.1.1-cdh6.0.1
	org.apache.hive	hive-contrib	2.1.1-cdh6.0.1
	org.apache.hive	hive-exec	2.1.1-cdh6.0.1
	org.apache.hive	hive-hbase-handler	2.1.1-cdh6.0.1
	org.apache.hive	hive-hplsql	2.1.1-cdh6.0.1
	org.apache.hive	hive-jdbc	2.1.1-cdh6.0.1
	org.apache.hive	hive-llap-client	2.1.1-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hive	hive-llap-common	2.1.1-cdh6.0.1
	org.apache.hive	hive-llap-ext-client	2.1.1-cdh6.0.1
	org.apache.hive	hive-llap-server	2.1.1-cdh6.0.1
	org.apache.hive	hive-llap-tez	2.1.1-cdh6.0.1
	org.apache.hive	hive-metastore	2.1.1-cdh6.0.1
	org.apache.hive	hive-orc	2.1.1-cdh6.0.1
	org.apache.hive	hive-serde	2.1.1-cdh6.0.1
	org.apache.hive	hive-service	2.1.1-cdh6.0.1
	org.apache.hive	hive-service-rpc	2.1.1-cdh6.0.1
	org.apache.hive	hive-shims	2.1.1-cdh6.0.1
	org.apache.hive	hive-spark-client	2.1.1-cdh6.0.1
	org.apache.hive	hive-storage-api	2.1.1-cdh6.0.1
	org.apache.hive	hive-testutils	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-hcatalog-core	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-hcatalog-pig-adapter	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-hcatalog-server-extensions	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-hcatalog-streaming	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-webhcat	2.1.1-cdh6.0.1
	org.apache.hive.hcatalog	hive-webhcat-java-client	2.1.1-cdh6.0.1
	org.apache.hive.shims	hive-shims-0.23	2.1.1-cdh6.0.1
	org.apache.hive.shims	hive-shims-common	2.1.1-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hive.shims	hive-shims-scheduler	2.1.1-cdh6.0.1
Apache HBase	org.apache.hbase	hbase-annotations	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-checkstyle	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-client	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-client-project	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-common	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-endpoint	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-error-prone	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-examples	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-external-blockcache	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-hadoop-compat	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-hadoop2-compat	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-http	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-it	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-mapreduce	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-metrics	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-metrics-api	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-procedure	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-protocol	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-protocol-shaded	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-replication	2.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.hbase	hbase-resource-bundle	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-rest	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-rsgroup	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-server	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-shaded-client	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-shaded-client-byo-hadoop	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-shaded-client-project	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-shaded-mapreduce	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-shell	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-spark	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-spark-it	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-testing-util	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-thrift	2.0.0-cdh6.0.1
	org.apache.hbase	hbase-zookeeper	2.0.0-cdh6.0.1
Apache ZooKeeper	org.apache.zookeeper	zookeeper	3.4.5-cdh6.0.1
Apache Sqoop	org.apache.sqoop	sqoop	1.4.7-cdh6.0.1
Apache Pig	org.apache.pig	pig	0.17.0-cdh6.0.1
	org.apache.pig	piggybank	0.17.0-cdh6.0.1
	org.apache.pig	pigsmoke	0.17.0-cdh6.0.1
	org.apache.pig	pigunit	0.17.0-cdh6.0.1
Apache Flume 1.x	org.apache.flume	flume-checkstyle	1.8.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.flume	flume-ng-auth	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-configuration	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-core	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-embedded-agent	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-node	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-sdk	1.8.0-cdh6.0.1
	org.apache.flume	flume-ng-tests	1.8.0-cdh6.0.1
	org.apache.flume	flume-tools	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-channels	flume-file-channel	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-channels	flume-jdbc-channel	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-channels	flume-kafka-channel	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-channels	flume-spillable-memory-channel	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-clients	flume-ng-log4jappender	1.8.0-cdh6.0.1
	org.apache.flume.flume-legacy-sources	flume-avro-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-legacy-sources	flume-thrift-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-dataset-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-hdfs-sink	1.8.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.flume.flume-ng-sinks	flume-hive-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-http-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-irc-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-ng-hbase2-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-ng-kafka-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sinks	flume-ng-morphline-solr-sink	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sources	flume-jms-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sources	flume-kafka-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sources	flume-scribe-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sources	flume-taildir-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-ng-sources	flume-twitter-source	1.8.0-cdh6.0.1
	org.apache.flume.flume-shared	flume-shared-kafka-test	1.8.0-cdh6.0.1
Kafka	org.apache.kafka	connect-api	1.0.1-cdh6.0.1
	org.apache.kafka	connect-file	1.0.1-cdh6.0.1
	org.apache.kafka	connect-json	1.0.1-cdh6.0.1
	org.apache.kafka	connect-runtime	1.0.1-cdh6.0.1
	org.apache.kafka	connect-transforms	1.0.1-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.kafka	kafka-clients	1.0.1-cdh6.0.1
	org.apache.kafka	kafka-examples	1.0.1-cdh6.0.1
	org.apache.kafka	kafka-log4j-appender	1.0.1-cdh6.0.1
	org.apache.kafka	kafka-streams	1.0.1-cdh6.0.1
	org.apache.kafka	kafka-streams-examples	1.0.1-cdh6.0.1
	org.apache.kafka	kafka-tools	1.0.1-cdh6.0.1
	org.apache.kafka	kafka_2.11	1.0.1-cdh6.0.1
Apache Oozie	org.apache.oozie	oozie-client	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-core	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-examples	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-server	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-distcp	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-hcatalog	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-hive	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-hive2	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-oozie	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-pig	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-spark	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-sqoop	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-sharelib-streaming	5.0.0-cdh6.0.1
	org.apache.oozie	oozie-tools	5.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.oozie.test	oozie-mini	5.0.0-cdh6.0.1
Apache Sentry	com.cloudera.cdh	solr-upgrade	1.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-hbase-indexer	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-hive	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-hive-common	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-hive-conf	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-hive-follower	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-kafka	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-binding-solr	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-core-common	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-core-model-db	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-core-model-indexer	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-core-model-kafka	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-core-model-solr	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-dist	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-hdfs-common	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-hdfs-dist	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-hdfs-namenode-plugin	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-hdfs-service	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-policy-common	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-policy-engine	2.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.sentry	sentry-provider-cache	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-provider-common	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-provider-db	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-provider-file	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-shaded-miscellaneous	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-tests-hive	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-tests-kafka	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-tests-solr	2.0.0-cdh6.0.1
	org.apache.sentry	sentry-tools	2.0.0-cdh6.0.1
	org.apache.sentry	solr-sentry-handlers	2.0.0-cdh6.0.1
Apache Spark	org.apache.spark	spark-catalyst_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-core_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-graphx_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-hadoop-cloud_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-hive_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-launcher_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-mllib-local_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-mllib_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-network-common_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-network-shuffle_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-network-yarn_2.11	2.2.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.spark	spark-repl_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-sketch_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-sql-kafka-0-10_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-sql_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming-flume-assembly_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming-flume-sink_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming-flume_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming-kafka-0-10-assembly_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming-kafka-0-10_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-streaming_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-tags_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-unsafe_2.11	2.2.0-cdh6.0.1
	org.apache.spark	spark-yarn_2.11	2.2.0-cdh6.0.1
Apache Crunch	org.apache.crunch	crunch-archetype	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-contrib	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-core	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-examples	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-hbase	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-hive	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-scrunch	0.11.0-cdh6.0.1
	org.apache.crunch	crunch-spark	0.11.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.crunch	crunch-test	0.11.0-cdh6.0.1
Apache Avro	org.apache.avro	avro	1.8.2-cdh6.0.1
	org.apache.avro	avro-compiler	1.8.2-cdh6.0.1
	org.apache.avro	avro-guava-dependencies	1.8.2-cdh6.0.1
	org.apache.avro	avro-ipc	1.8.2-cdh6.0.1
	org.apache.avro	avro-mapred	1.8.2-cdh6.0.1
	org.apache.avro	avro-maven-plugin	1.8.2-cdh6.0.1
	org.apache.avro	avro-protobuf	1.8.2-cdh6.0.1
	org.apache.avro	avro-service-archetype	1.8.2-cdh6.0.1
	org.apache.avro	avro-thrift	1.8.2-cdh6.0.1
	org.apache.avro	avro-tools	1.8.2-cdh6.0.1
	org.apache.avro	trevni-avro	1.8.2-cdh6.0.1
	org.apache.avro	trevni-core	1.8.2-cdh6.0.1
Kite SDK	org.kitesdk	kite-data-core	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-crunch	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-hbase	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-hive	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-mapreduce	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-oozie	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-s3	1.0.0-cdh6.0.1
	org.kitesdk	kite-data-spark	1.0.0-cdh6.0.1
	org.kitesdk	kite-hadoop-compatibility	1.0.0-cdh6.0.1
	org.kitesdk	kite-maven-plugin	1.0.0-cdh6.0.1
	org.kitesdk	kite-minicluster	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-avro	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-core	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-hadoop-core	1.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.kitesdk	kite-morphlines-hadoop-parquet-avro	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-hadoop-rcfile	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-hadoop-sequencefile	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-json	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-maxmind	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-metrics-scalable	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-metrics-servlets	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-protobuf	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-saxon	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-solr-cell	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-solr-core	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-tika-core	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-tika-decompress	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-twitter	1.0.0-cdh6.0.1
	org.kitesdk	kite-morphlines-useragent	1.0.0-cdh6.0.1
	org.kitesdk	kite-tools	1.0.0-cdh6.0.1
Apache Solr	org.apache.lucene	lucene-analyzers-common	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-icu	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-kuromoji	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-morfologik	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-phonetic	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-smartcn	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-stempel	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-analyzers-uima	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-backward-codecs	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-benchmark	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-classification	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-codecs	7.0.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.lucene	lucene-core	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-demo	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-expressions	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-facet	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-grouping	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-highlighter	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-join	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-memory	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-misc	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-queries	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-queryparser	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-replicator	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-sandbox	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-spatial	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-spatial-extras	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-spatial3d	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-suggest	7.0.0-cdh6.0.1
	org.apache.lucene	lucene-test-framework	7.0.0-cdh6.0.1
	org.apache.solr	solr-analysis-extras	7.0.0-cdh6.0.1
	org.apache.solr	solr-analytics	7.0.0-cdh6.0.1
	org.apache.solr	solr-cell	7.0.0-cdh6.0.1
	org.apache.solr	solr-clustering	7.0.0-cdh6.0.1
	org.apache.solr	solr-core	7.0.0-cdh6.0.1
	org.apache.solr	solr-dataimporthandler	7.0.0-cdh6.0.1

Project	groupId	artifactId	version
Apache Solr	org.apache.solr	solr-dataimporthandler-extras	7.0.0-cdh6.0.1
	org.apache.solr	solr-jetty-customizations	7.0.0-cdh6.0.1
	org.apache.solr	solr-langid	7.0.0-cdh6.0.1
	org.apache.solr	solr-ltr	7.0.0-cdh6.0.1
	org.apache.solr	solr-security-util	7.0.0-cdh6.0.1
	org.apache.solr	solr-sentry-audit-logging	7.0.0-cdh6.0.1
	org.apache.solr	solr-solrj	7.0.0-cdh6.0.1
	org.apache.solr	solr-test-framework	7.0.0-cdh6.0.1
	org.apache.solr	solr-uima	7.0.0-cdh6.0.1
	org.apache.solr	solr-velocity	7.0.0-cdh6.0.1
Cloudera Search	com.cloudera.search	search-crunch	1.0.0-cdh6.0.1
	com.cloudera.search	search-mr	1.0.0-cdh6.0.1
HBase Indexer	com.ngdata	hbase-indexer-all	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-cli	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-common	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-demo	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-dist	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-engine	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-model	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-morphlines	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-mr	1.5-cdh6.0.1
	com.ngdata	hbase-indexer-server	1.5-cdh6.0.1
	com.ngdata	hbase-sep-api	1.5-cdh6.0.1
	com.ngdata	hbase-sep-demo	1.5-cdh6.0.1
	com.ngdata	hbase-sep-impl	1.5-cdh6.0.1
	com.ngdata	hbase-sep-tools	1.5-cdh6.0.1
Apache Kudu	org.apache.kudu	kudu-client	1.6.0-cdh6.0.1
	org.apache.kudu	kudu-client-tools	1.6.0-cdh6.0.1
	org.apache.kudu	kudu-flume-sink	1.6.0-cdh6.0.1
	org.apache.kudu	kudu-mapreduce	1.6.0-cdh6.0.1
	org.apache.kudu	kudu-spark2-tools_2.11	1.6.0-cdh6.0.1

Project	groupId	artifactId	version
	org.apache.kudu	kudu-spark2_2.11	1.6.0-cdh6.0.1

CDH 6.0.0

The following table lists the project name, groupId, artifactId, and version required to access each CDH artifact.

Project	groupId	artifactId	version
Apache Hadoop	org.apache.hadoop	hadoop-aliyun	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-annotations	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-archive-logs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-archives	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-assemblies	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-auth	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-aws	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-azure	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-azure-datalake	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-build-tools	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-client	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-client-api	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-client-integration-tests	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-client-minicluster	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-client-runtime	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-cloud-storage	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-common	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-datajoin	3.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-distcp	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-extras	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-gridmix	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-hdfs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-hdfs-client	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-hdfs-httpfs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-hdfs-native-client	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-hdfs-nfs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-kafka	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-kms	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-app	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-common	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-core	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-hs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-hs-plugins	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-jobclient	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-nativetask	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-shuffle	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-client-uploader	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-mapreduce-examples	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-maven-plugins	3.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-minicluster	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-minikdc	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-nfs	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-openstack	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-resourceestimator	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-rumen	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-sls	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-streaming	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-api	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-applications-distributedshell	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-applications-unmanaged-am-launcher	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-client	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-common	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-registry	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-applicationhistoryservice	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-common	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-nodemanager	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-resourcemanager	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-router	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-sharedcachemanager	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-tests	3.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hadoop	hadoop-yarn-server-timeline-pluginstorage	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-timelineservice-hbase-tests	3.0.0-cdh6.0.0
	org.apache.hadoop	hadoop-yarn-server-web-proxy	3.0.0-cdh6.0.0
Apache Hive	org.apache.hive	hive-accumulo-handler	2.1.1-cdh6.0.0
	org.apache.hive	hive-ant	2.1.1-cdh6.0.0
	org.apache.hive	hive-beeline	2.1.1-cdh6.0.0
	org.apache.hive	hive-classification	2.1.1-cdh6.0.0
	org.apache.hive	hive-cli	2.1.1-cdh6.0.0
	org.apache.hive	hive-common	2.1.1-cdh6.0.0
	org.apache.hive	hive-contrib	2.1.1-cdh6.0.0
	org.apache.hive	hive-exec	2.1.1-cdh6.0.0
	org.apache.hive	hive-hbase-handler	2.1.1-cdh6.0.0
	org.apache.hive	hive-hplsql	2.1.1-cdh6.0.0
	org.apache.hive	hive-jdbc	2.1.1-cdh6.0.0
	org.apache.hive	hive-llap-client	2.1.1-cdh6.0.0
	org.apache.hive	hive-llap-common	2.1.1-cdh6.0.0
	org.apache.hive	hive-llap-ext-client	2.1.1-cdh6.0.0
	org.apache.hive	hive-llap-server	2.1.1-cdh6.0.0
	org.apache.hive	hive-llap-tez	2.1.1-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hive	hive-metastore	2.1.1-cdh6.0.0
	org.apache.hive	hive-orc	2.1.1-cdh6.0.0
	org.apache.hive	hive-serde	2.1.1-cdh6.0.0
	org.apache.hive	hive-service	2.1.1-cdh6.0.0
	org.apache.hive	hive-service-rpc	2.1.1-cdh6.0.0
	org.apache.hive	hive-shims	2.1.1-cdh6.0.0
	org.apache.hive	hive-spark-client	2.1.1-cdh6.0.0
	org.apache.hive	hive-storage-api	2.1.1-cdh6.0.0
	org.apache.hive	hive-testutils	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-hcatalog-core	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-hcatalog-pig-adapter	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-hcatalog-server-extensions	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-hcatalog-streaming	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-webhcat	2.1.1-cdh6.0.0
	org.apache.hive.hcatalog	hive-webhcat-java-client	2.1.1-cdh6.0.0
	org.apache.hive.shims	hive-shims-0.23	2.1.1-cdh6.0.0
	org.apache.hive.shims	hive-shims-common	2.1.1-cdh6.0.0
	org.apache.hive.shims	hive-shims-scheduler	2.1.1-cdh6.0.0
Apache HBase	org.apache.hbase	hbase-annotations	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-checkstyle	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-client	2.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hbase	hbase-client-project	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-common	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-endpoint	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-error-prone	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-examples	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-external-blockcache	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-hadoop-compat	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-hadoop2-compat	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-http	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-it	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-mapreduce	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-metrics	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-metrics-api	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-procedure	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-protocol	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-protocol-shaded	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-replication	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-resource-bundle	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-rest	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-rsgroup	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-server	2.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.hbase	hbase-shaded-client	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-shaded-client-byo-hadoop	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-shaded-client-project	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-shaded-mapreduce	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-shell	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-spark	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-spark-it	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-testing-util	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-thrift	2.0.0-cdh6.0.0
	org.apache.hbase	hbase-zookeeper	2.0.0-cdh6.0.0
Apache ZooKeeper	org.apache.zookeeper	zookeeper	3.4.5-cdh6.0.0
Apache Sqoop	org.apache.sqoop	sqoop	1.4.7-cdh6.0.0
Apache Pig	org.apache.pig	pig	0.17.0-cdh6.0.0
	org.apache.pig	piggybank	0.17.0-cdh6.0.0
	org.apache.pig	pigsmoke	0.17.0-cdh6.0.0
	org.apache.pig	pigunit	0.17.0-cdh6.0.0
Apache Flume 1.x	org.apache.flume	flume-checkstyle	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-auth	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-configuration	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-core	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-embedded-agent	1.8.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.flume	flume-ng-node	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-sdk	1.8.0-cdh6.0.0
	org.apache.flume	flume-ng-tests	1.8.0-cdh6.0.0
	org.apache.flume	flume-tools	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-channels	flume-file-channel	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-channels	flume-jdbc-channel	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-channels	flume-kafka-channel	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-channels	flume-spillable-memory-channel	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-clients	flume-ng-log4jappender	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sources	flume-avro-source	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sources	flume-thrift-source	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sinks	flume-dataset-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sinks	flume-hdfs-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sinks	flume-hive-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sinks	flume-http-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume- ng -sinks	flume-irc-sink	1.8.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.flume.flume-ng-sinks	flume-ng-hbase2-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sinks	flume-ng-kafka-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sinks	flume-ng-morphline-solr-sink	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sources	flume-jms-source	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sources	flume-kafka-source	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sources	flume-scribe-source	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sources	flume-taildir-source	1.8.0-cdh6.0.0
	org.apache.flume.flume-ng-sources	flume-twitter-source	1.8.0-cdh6.0.0
	org.apache.flume.flume-shared	flume-shared-kafka-test	1.8.0-cdh6.0.0
Kafka	org.apache.kafka	connect-api	1.0.1-cdh6.0.0
	org.apache.kafka	connect-file	1.0.1-cdh6.0.0
	org.apache.kafka	connect-json	1.0.1-cdh6.0.0
	org.apache.kafka	connect-runtime	1.0.1-cdh6.0.0
	org.apache.kafka	connect-transforms	1.0.1-cdh6.0.0
	org.apache.kafka	kafka-clients	1.0.1-cdh6.0.0
	org.apache.kafka	kafka-examples	1.0.1-cdh6.0.0
	org.apache.kafka	kafka-log4j-appender	1.0.1-cdh6.0.0
	org.apache.kafka	kafka-streams	1.0.1-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.kafka	kafka-streams-examples	1.0.1-cdh6.0.0
	org.apache.kafka	kafka-tools	1.0.1-cdh6.0.0
	org.apache.kafka	kafka_2.11	1.0.1-cdh6.0.0
Apache Oozie	org.apache.oozie	oozie-client	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-core	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-examples	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-server	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-distcp	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-hcatalog	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-hive	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-hive2	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-oozie	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-pig	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-spark	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-sqoop	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-sharelib-streaming	5.0.0-cdh6.0.0
	org.apache.oozie	oozie-tools	5.0.0-cdh6.0.0
	org.apache.oozie.test	oozie-mini	5.0.0-cdh6.0.0
Apache Sentry	com.cloudera.cdh	solr-upgrade	1.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-hbase-indexer	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-hive	2.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.sentry	sentry-binding-hive-common	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-hive-conf	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-hive-follower	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-kafka	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-binding-solr	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-core-common	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-core-model-db	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-core-model-indexer	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-core-model-kafka	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-core-model-solr	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-dist	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-hdfs-common	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-hdfs-dist	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-hdfs-namenode-plugin	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-hdfs-service	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-policy-common	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-policy-engine	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-provider-cache	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-provider-common	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-provider-db	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-provider-file	2.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.sentry	sentry-shaded-miscellaneous	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-tests-hive	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-tests-kafka	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-tests-solr	2.0.0-cdh6.0.0
	org.apache.sentry	sentry-tools	2.0.0-cdh6.0.0
	org.apache.sentry	solr-sentry-handlers	2.0.0-cdh6.0.0
Apache Spark	org.apache.spark	spark-catalyst_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-core_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-graphx_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-hadoop-cloud_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-hive_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-launcher_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-mllib-local_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-mllib_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-network-common_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-network-shuffle_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-network-yarn_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-repl_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-sketch_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-sql-kafka-0-10_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-sql_2.11	2.2.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.spark	spark-streaming-flume-assembly_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-streaming-flume-sink_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-streaming-flume_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-streaming-kafka-0-10-assembly_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-streaming-kafka-0-10_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-streaming_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-tags_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-unsafe_2.11	2.2.0-cdh6.0.0
	org.apache.spark	spark-yarn_2.11	2.2.0-cdh6.0.0
Apache Crunch	org.apache.crunch	crunch-archetype	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-contrib	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-core	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-examples	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-hbase	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-hive	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-scrunch	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-spark	0.11.0-cdh6.0.0
	org.apache.crunch	crunch-test	0.11.0-cdh6.0.0
Apache Avro	org.apache.avro	avro	1.8.2-cdh6.0.0
	org.apache.avro	avro-compiler	1.8.2-cdh6.0.0
	org.apache.avro	avro-guava-dependencies	1.8.2-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.avro	avro-ipc	1.8.2-cdh6.0.0
	org.apache.avro	avro-mapred	1.8.2-cdh6.0.0
	org.apache.avro	avro-maven-plugin	1.8.2-cdh6.0.0
	org.apache.avro	avro-protobuf	1.8.2-cdh6.0.0
	org.apache.avro	avro-service-archetype	1.8.2-cdh6.0.0
	org.apache.avro	avro-thrift	1.8.2-cdh6.0.0
	org.apache.avro	avro-tools	1.8.2-cdh6.0.0
	org.apache.avro	trevni-avro	1.8.2-cdh6.0.0
	org.apache.avro	trevni-core	1.8.2-cdh6.0.0
Kite SDK	org.kitesdk	kite-data-core	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-crunch	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-hbase	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-hive	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-mapreduce	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-oozie	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-s3	1.0.0-cdh6.0.0
	org.kitesdk	kite-data-spark	1.0.0-cdh6.0.0
	org.kitesdk	kite-hadoop-compatibility	1.0.0-cdh6.0.0
	org.kitesdk	kite-maven-plugin	1.0.0-cdh6.0.0
	org.kitesdk	kite-minicluster	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-avro	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-core	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-hadoop-core	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-hadoop-parquet-avro	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-hadoop-rccfile	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-hadoop-sequencefile	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-json	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-maxmind	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-metrics-scalable	1.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.kitesdk	kite-morphlines-metrics-servlets	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-protobuf	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-saxon	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-solr-cell	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-solr-core	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-tika-core	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-tika-decompress	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-twitter	1.0.0-cdh6.0.0
	org.kitesdk	kite-morphlines-useragent	1.0.0-cdh6.0.0
	org.kitesdk	kite-tools	1.0.0-cdh6.0.0
Apache Solr	org.apache.lucene	lucene-analyzers-common	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-icu	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-kuromoji	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-morfologik	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-phonetic	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-smartcn	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-stempel	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-analyzers-uima	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-backward-codecs	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-benchmark	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-classification	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-codecs	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-core	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-demo	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-expressions	7.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.lucene	lucene-facet	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-grouping	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-highlighter	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-join	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-memory	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-misc	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-queries	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-queryparser	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-replicator	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-sandbox	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-spatial	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-spatial-extras	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-spatial3d	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-suggest	7.0.0-cdh6.0.0
	org.apache.lucene	lucene-test-framework	7.0.0-cdh6.0.0
	org.apache.solr	solr-analysis-extras	7.0.0-cdh6.0.0
	org.apache.solr	solr-analytics	7.0.0-cdh6.0.0
	org.apache.solr	solr-cell	7.0.0-cdh6.0.0
	org.apache.solr	solr-clustering	7.0.0-cdh6.0.0
	org.apache.solr	solr-core	7.0.0-cdh6.0.0
	org.apache.solr	solr-dataimporthandler	7.0.0-cdh6.0.0
	org.apache.solr	solr-dataimporthandler-extras	7.0.0-cdh6.0.0
	org.apache.solr	solr-jetty-customizations	7.0.0-cdh6.0.0
	org.apache.solr	solr-langid	7.0.0-cdh6.0.0
	org.apache.solr	solr-ltr	7.0.0-cdh6.0.0
	org.apache.solr	solr-security-util	7.0.0-cdh6.0.0

Project	groupId	artifactId	version
	org.apache.solr	solr-sentry-audit-logging	7.0.0-cdh6.0.0
	org.apache.solr	solr-solrj	7.0.0-cdh6.0.0
	org.apache.solr	solr-test-framework	7.0.0-cdh6.0.0
	org.apache.solr	solr-uima	7.0.0-cdh6.0.0
	org.apache.solr	solr-velocity	7.0.0-cdh6.0.0
Cloudera Search	com.cloudera.search	search-crunch	1.0.0-cdh6.0.0
	com.cloudera.search	search-mr	1.0.0-cdh6.0.0
HBase Indexer	com.ngdata	hbase-indexer-all	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-cli	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-common	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-demo	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-dist	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-engine	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-model	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-morphlines	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-mr	1.5-cdh6.0.0
	com.ngdata	hbase-indexer-server	1.5-cdh6.0.0
	com.ngdata	hbase-sep-api	1.5-cdh6.0.0
	com.ngdata	hbase-sep-demo	1.5-cdh6.0.0
	com.ngdata	hbase-sep-impl	1.5-cdh6.0.0
	com.ngdata	hbase-sep-tools	1.5-cdh6.0.0
Apache Kudu	org.apache.kudu	kudu-client	1.6.0-cdh6.0.0
	org.apache.kudu	kudu-client-tools	1.6.0-cdh6.0.0
	org.apache.kudu	kudu-flume-sink	1.6.0-cdh6.0.0
	org.apache.kudu	kudu-mapreduce	1.6.0-cdh6.0.0
	org.apache.kudu	kudu-spark2-tools_2.11	1.6.0-cdh6.0.0
	org.apache.kudu	kudu-spark2_2.11	1.6.0-cdh6.0.0

Cloudera Navigator 6 Encryption Version and Download Information

The 64-bit packages listed here support both Cloudera Express with its extensive set of monitoring and management features, and Cloudera Enterprise with additional functionality. A 60-day trial can be enabled to provide access to the

full set of Cloudera Enterprise Cloudera Enterprise features. Cloudera Enterprise can be enabled permanently with the appropriate license. To obtain a Cloudera Enterprise license, fill in this [form](#) or call 866-843-7207.

Cloudera Navigator Key Trustee Server Version and Download Information

You can install Key Trustee Server using parcels in Cloudera Manager or using packages.

For more information on parcels, see [Parcels](#).

Navigator Key Trustee Server 6.1.0

- Parcel repository tarball:
 - Go to the Cloudera Navigator Key Trustee Server [download page](#) and select **Parcels** from the **Package or Parcel** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Installing Cloudera Navigator Key Trustee Server](#).
- Package repository tarball:
 - Go to the Cloudera Navigator Key Trustee Server [download page](#) and select **Packages** from the **Package or Parcel** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Package installation command:

```
yum install keytrustee-server-6.1.0
```

- Set up an internal repository, following the instructions in [Installing Cloudera Navigator Key Trustee Server](#).

- Release Notes:

[Cloudera Navigator 6.1.0 Encryption Release Notes](#) on page 649

Navigator Key Trustee Server 6.0.0

- Parcel repository tarball:
 - Go to the Cloudera Navigator Key Trustee Server [download page](#) and select **Parcels** from the **Package or Parcel** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Installing Cloudera Navigator Key Trustee Server](#).
- Package repository tarball:
 - Go to the Cloudera Navigator Key Trustee Server [download page](#) and select **Packages** from the **Package or Parcel** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Package installation command:

```
yum install keytrustee-server-6.0.0
```

- Set up an internal repository, following the instructions in [Installing Cloudera Navigator Key Trustee Server](#).

- Release Notes:

[Cloudera Navigator 6.0.0 Encryption Release Notes](#) on page 659

Cloudera Navigator Key HSM Version and Download Information

Cloudera Navigator Key HSM 6.1.0

RHEL or CentOS

- Package repository tarball:
 - Visit the Cloudera Navigator Key HSM [download page](#). Select your OS version, and click **DOWNLOAD NOW**.
- Package installation command:

```
yum install keytrustee-keyhsm-6.1.0
```

- Release notes:

[Cloudera Navigator 6.1.0 Encryption Release Notes](#) on page 649

Cloudera Navigator Key HSM 6.0.0

RHEL or CentOS

- Package repository tarball:
 - Visit the Cloudera Navigator Key HSM [download page](#). Select your OS version, and click **DOWNLOAD NOW**.
- Package installation command:

```
yum install keytrustee-keyhsm-6.0.0
```

- Release notes:

[Cloudera Navigator 6.0.0 Encryption Release Notes](#) on page 659

Cloudera Navigator Key Trustee KMS Version and Download Information

You can install Key Trustee KMS using parcels in Cloudera Manager or using packages.

For more information on parcels, see [Parcels](#).

Key Trustee KMS 6.2.0

- Parcel repository tarball:
 - Visit the Key Trustee KMS [download page](#).
 - Select **Parcels** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:
 - Visit the Key Trustee KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Select your operating system from the **SELECT AN OS** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Key Trustee KMS](#)
- Release notes:

[Cloudera Navigator 6.2.0 Encryption Release Notes](#) on page 642

Key Trustee KMS 6.1.0

- Parcel repository tarball:

Cloudera Enterprise 6 Release Guide

- Visit the Key Trustee KMS [download page](#).
- Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
- Click **DOWNLOAD NOW**.
- Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:
 - Visit the Key Trustee KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Select your operating system from the **SELECT AN OS** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Key Trustee KMS](#)
- Release notes:

[Cloudera Navigator 6.1.0 Encryption Release Notes](#) on page 649

Key Trustee KMS 6.0.0

- Parcel repository tarball:
 - Visit the Key Trustee KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:
 - Visit the Key Trustee KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Select your operating system from the **SELECT AN OS** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Key Trustee KMS](#)
- Release notes:

[Cloudera Navigator 6.0.0 Encryption Release Notes](#) on page 659

Cloudera Navigator HSM KMS Version and Download Information

You can install HSM KMS using parcels in Cloudera Manager or using packages.

For more information on parcels, see [Parcels](#).

Cloudera Navigator HSM KMS 6.2.0

- Parcel repository tarball:
 - Visit the HSM KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.

- Click **DOWNLOAD NOW**.
- Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:
 - Visit the HSM KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Select your operating system from the **SELECT AN OS** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Navigator HSM KMS Backed by Thales HSM](#)
 - [Installing Navigator HSM KMS Backed by Luna HSM](#)
- Release notes:

[Cloudera Navigator 6.2.0 Encryption Release Notes](#) on page 642

Cloudera Navigator HSM KMS 6.1.0

- Parcel repository tarball:
 - Visit the HSM KMS [download page](#).
 - Select **Parcels** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:
 - Visit the HSM KMS [download page](#).
 - Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Select your operating system from the **SELECT AN OS** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Navigator HSM KMS Backed by Thales HSM](#)
 - [Installing Navigator HSM KMS Backed by Luna HSM](#)
- Release notes:

[Cloudera Navigator 6.1.0 Encryption Release Notes](#) on page 649

Cloudera Navigator HSM KMS 6.0.0

- Parcel repository tarball:
 - Visit the HSM KMS [download page](#).
 - Select **Parcels** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
 - Click **DOWNLOAD NOW**.
 - Set up an internal repository, following the instructions in [Creating and Using a Remote Parcel Repository for Cloudera Manager](#).
- Package repository tarball:

Cloudera Enterprise 6 Release Guide

- Visit the HSM KMS [download page](#).
- Select **Packages** from the **CHOOSE DOWNLOAD TYPE** drop-down menu.
- Select your operating system from the **SELECT AN OS** drop-down menu.
- Click **DOWNLOAD NOW**.
- Set up an internal repository, following the instructions in [Creating and Using a Package Repository for Cloudera Manager](#).
- Installation instructions:
 - [Installing Navigator HSM KMS Backed by Thales HSM](#)
 - [Installing Navigator HSM KMS Backed by Luna HSM](#)
- Release notes:
[Cloudera Navigator 6.0.0 Encryption Release Notes](#) on page 659

Cloudera Navigator Encrypt Version and Download Information

Navigator Encrypt 6.2.0

Release notes: [Cloudera Navigator 6.2.0 Encryption Release Notes](#) on page 642.

Package repository tarball: Visit the Cloudera Navigator Encrypt [download page](#). Select your OS version, and then click **DOWNLOAD NOW**.

For instructions on setting up the repository and installing Navigator Encrypt, see [Installing Cloudera Navigator Encrypt](#).

Navigator Encrypt 6.1.0

Release notes: [Cloudera Navigator 6.1.0 Encryption Release Notes](#) on page 649.

Package repository tarball: Visit the Cloudera Navigator Encrypt [download page](#). Select your OS version, and then click **DOWNLOAD NOW**.

For instructions on setting up the repository and installing Navigator Encrypt, see [Installing Cloudera Navigator Encrypt](#).

Navigator Encrypt 6.0.0

Release notes: [Cloudera Navigator 6.0.0 Encryption Release Notes](#) on page 659

Package repository tarball: Visit the Cloudera Navigator Encrypt [download page](#). Select your OS version, and then click **DOWNLOAD NOW**.

For instructions on setting up the repository and installing Navigator Encrypt, see [Installing Cloudera Navigator Encrypt](#).

What's New in Cloudera Documentation

This page describes new topics added and changes made to Cloudera documentation.

What's New in Cloudera Documentation in March, 2019

This section describes new topics added and major changes made to Cloudera documentation in March, 2019:

Table 23:

Product	What's New	Link
Cloudera Data Science Workbench	Published a new Security Overview for Cloudera Data Science Workbench. This topic goes over the basics of the CDSW security model, the wildcard DNS requirement, and how authentication, authorization, and wire	CDSW Security Overview

Product	What's New	Link
	encryption work in Cloudera Data Science Workbench.	
Apache Flume	Added information on Flume Kudu sink, which is a Flume sink that reads events from a channel and writes them to a Kudu table.	Flume Kudu Sink
Apache Hbase	Added information about how to validate HBase compatibility when upgrading a CDH 5 cluster to CDH 6.	Migrating Apache HBase Before Upgrading to CDH 6
	Added information about how to configure the BucketCache IO Engine, which defines where to store the content of the BucketCache.	Configuring the BucketCache IO Engine
YARN	Added information about how to enable GPU as a resource.	Enabling GPU Using Cloudera Manager
	Added information about how to create and enable a custom resource.	Create Custom Resource Using Cloudera Manager

What's New in Cloudera Documentation in February, 2019

This section describes new topics added and major changes made to Cloudera documentation in February, 2019:

Table 24:

Product	What's New	Link
Cloud Storage	CDH 6.1.0 supports Google Cloud Storage as the persistent storage layer for CDH clusters.	Configuring Google Cloud Storage Connectivity
Apache Impala	Added a new section on load-balancing proxy in TLS-enabled cluster.	Special Proxy Considerations for TLS/SSL-Enabled Clusters
	Added a section on enabling LDAP in Cloudera Manager.	Enabling LDAP in Cloudera Manager
	Updated docs to reflect decoupling of compute and storage in Hadoop clusters.	Components of the Impala Server
Apache Kudu	Added a recommendation to use nscd (name service caching daemon) for all name resolutions.	Slow Name Resolution and nscd
Hue	The Hue Guide has been completely re-organized and additional content added. Look for further improvements in the areas of performance tuning and reference architectures to be released soon.	<ul style="list-style-type: none"> • CDH 6.1 Hue Guide • CDH 6.0 Hue Guide

What's New in Cloudera Documentation in January, 2019

This section describes new topics added and major changes made to Cloudera documentation in January, 2019:

Table 25:

Product	What's New	Link
Cloudera Data Science Workbench	<p>Released Cloudera Data Science Workbench 1.5.0.</p> <p>This release includes direct integration with Hortonworks Data Platform (HDP), adds support for Cloudera Enterprise 6.1 (and higher), and ships some noteworthy bug fixes and enhancements.</p>	<ul style="list-style-type: none"> • 1.5.0 Release Notes • Deploying CDSW on HDP
Apache Spark ML	Learn how to configure Spark ML to use native math libraries that accelerate model training speed for algorithms like Alternating Least Squares (ALS).	Using Native Math Libraries to Accelerate Spark Machine Learning Applications
Workload XM	The new Workload View feature enables you to break down workloads by specific criteria to perform deep-dive analysis on the queries. For example, you can use the Workload View feature to determine which users are executing workloads that do not adhere to SLAs. You can also examine how queries being sent to specific databases or that use specific pools are performing against SLAs.	<ul style="list-style-type: none"> • Classifying Workloads for Analysis with Workload Views • Video: Classifying Workloads to Gain Insights

What's New in Cloudera Documentation in December, 2018

This section describes new topics added and major changes made to Cloudera documentation in December, 2018:

Table 26:

Product	What's New	Link
Cloudera Data Science Workbench	<p>Released Cloudera Data Science Workbench 1.4.3.</p> <p>This release adds new customizations and fixes some critical bugs, including a permanent fix for TSB-346: Risk of Data Loss on Cloudera Data Science Workbench Shutdown and Restart. Please read the Release Notes carefully before you start upgrading or perform a shutdown/restart operation on any previous version of CDSW.</p>	Cloudera Data Science Workbench 1.4.3 Release Notes
OpenJDK	OpenJDK is now supported with Cloudera Manager and CDH 6.1 and higher.	<ul style="list-style-type: none"> • Manually Installing OpenJDK • Manually Migrating from Oracle JDK to OpenJDK

What's New in Cloudera Documentation in November, 2018

This section describes new topics added and major changes made to Cloudera documentation in November, 2018:

Table 27:

Product	What's New	Link
Apache HBase	Information about how to move the HBase Master role from one host to another.	Moving HBase Master Role to Another Host
Navigator Encrypt	When configuring a Navigator Encrypt mount point, you can use either the device name or the UUID to configure the mount point. Because the UUID mount point configuration choice is preferable for Navigator Encrypt, there is a new procedure for migrating from a device name to UUID.	Converting from Device Names to UUIDs for Encrypted Devices
Workload Experience Manager (Workload XM)	With the release of Cloudera Manager 5.16, new functionality has been added to Workload XM to redact logs and queries and for proxy server support. These new features can be enabled by configuring the Telemetry Publisher service in Cloudera Manager. In addition to these new features, now you can download the SQL commands to address "Corrupt Table Statistics" and "Missing Table Statistics" query health checks and numerous usability enhancements have also been added.	What's New from Workload XM
	Information about how to configure Workload XM to tunnel through a firewall in your environment.	Configuring a Firewall for Workload XM
	Detailed description of the diagnostic data collection performed by Workload XM.	Workload XM Diagnostic Data Collection

Install and Upgrade Notes

The notes in this topic contain important information about installing and upgrading Cloudera Enterprise. You should review these notes before installing or upgrading your software. For general release notes about Cloudera Enterprise, see [Cloudera Enterprise 6 Release Guide](#) on page 5.

Upgrades to Cloudera Enterprise 6.x

Restart of Impala and Hive required for Cloudera Manager 6.2 upgrade with ADLS

After upgrading to Cloudera Manager 6.2 or higher, Impala and Hive will be marked as stale for users running CDH 6.1 and using the ADLS Service. You will need to restart Hive and Impala before being able to connect to ADLS Gen2, but all previous functionality will continue to work without a restart. The configurations that will be marked stale are:

- fs.azure.account.auth.type
- fs.azure.account.oauth.provider.type
- fs.azure.account.oauth2.client.endpoint
- fs.azure.account.oauth2.client.id
- fs.azure.account.oauth2.client.secret.

Cloudera Bug: OPSAPS-47436

TSB-359 Backup and Disaster Recovery (BDR) HDFS and Hive Replications will fail on clusters running Cloudera Manager 6.1.0

Backup and Disaster Recovery (BDR) HDFS and Hive Replications will fail when replicating from secured (Kerberized) source clusters to destination clusters that have been upgraded to Cloudera Manager 6.1.0.

This also affects new installations of Cloudera Manager 6.1.0 on the destination cluster if an admin restarts the Cloudera Manager service.

Products affected: Cloudera Manager Backup and Disaster Recovery in a secure (Kerberized) environment

Releases affected: Cloudera Manager 6.1.0 (when used as the destination cluster of HDFS and/or Hive replication)

Users affected: Customers using HDFS or Hive Replication

Severity (Low/Medium/High): High

Root Cause and Impact:

In HDFS and Hive Replication, Cloudera Manager first runs a process on the destination cluster to verify if the replication is possible. Due to a bug, the source cluster is treated as an insecure (non-kerberized) cluster. As a result, replication fails.

You will see the exception `javax.security.sasl.SaslException: GSS initiate failed [Caused by GSSEException: No valid credentials provided (Mechanism level: Fail to create credential. (63) - No service creds)]` in the process stderr logs.

Immediate action required: If you use BDR, do not upgrade a destination cluster to Cloudera Manager 6.1.0. Upgrade to Cloudera Manager 6.1.1 or higher when it becomes available.

If you have already upgraded your destination cluster to Cloudera Manager to 6.1.0, use the following workaround:

1. For an existing HDFS or Hive replication schedule, select **Actions > Edit Configuration**.
2. Save the schedule.

Please note that you will need to edit only one schedule even if you have multiple schedules.

Note: This workaround is not persistent. That is, if you restart the Cloudera Manager service, you must repeat the above workaround.

Cloudera Issue: OPSAPS-48865

Fixed in Cloudera Manager 6.1.1

[Upgrades from Cloudera Enterprise 5.15 or 5.16 to 6.0x are not supported](#)

You cannot upgrade to Cloudera Manager or CDH 6.0.0 from Cloudera Manager or CDH 5.15 or 5.16.

[Upgrading to CDH 6.1.0 Enables Direct SQL mode in Hive service by default](#)

For details about the Cloudera Manager `Enable Direct SQL` option, refer to [Hive Metastore Database](#).

[Upgrades from Cloudera Enterprise 6.0 Beta Release to 6.x General Release Not supported](#)

You cannot upgrade to any Cloudera Manager or CDH 6.x general release from the Cloudera Manager or CDH 6.0 Beta release.

[Cloudera Express License Enforcement](#)

Use of Cloudera Express is limited to a total of 100 hosts running CDH6.0 or later across all environments used by an organization..

Note the following:

- Cloudera Manager will not allow you to add hosts to a CDH 6.x cluster if the total number of hosts across all CDH 6.x clusters will exceed 100.

- Cloudera Manager will not allow you to upgrade any cluster to CDH 6.x if the total number of managed CDH6.x cluster hosts will exceed 100. If an upgrade from Cloudera Manager 6.0 to 6.1 fails due to this limitation, you must downgrade Cloudera Manager to version 6.0, remove some hosts so that the number of hosts is less than 100, then retry the upgrade.



Note: If you downgrade from Cloudera Enterprise to Cloudera Express and the number of managed hosts exceeds 100, Cloudera Manager will disable all cluster management commands except for commands used to stop a cluster. You will not be able to restart or otherwise use clusters while the total number of hosts exceeds 100. Use the Cloudera Manager Admin Console to remove some hosts so that the number of hosts is less than 100.

Affected Versions: CM 6.1 and higher

Cloudera Issue: OPSAPS-46868

Cloudera Data Science Workbench is Not Supported with Cloudera Enterprise 6.0

Cloudera Data Science Workbench is not supported with Cloudera Enterprise 6.0.x. Cloudera Data Science Workbench 1.5.0 (and higher) is supported with Cloudera Manager 6.1.x (and higher) and CDH 6.1.x (and higher).

Impala roles with SELECT or INSERT privileges receive REFRESH privileges during the upgrade

Due to the Sentry and Impala fine grained privileges feature in 5.16.0, if a role has the `SELECT` or `INSERT` privilege on an object in Impala before upgrading to CDH 5.16.0, that role will automatically get the `REFRESH` privilege during the upgrade.

Hue requires manual installation of psycopg2

If you are installing or upgrading to CDH 6.0.0 and using the PostgreSQL database for the Hue database, you must install psycopg2 2.5.4 or higher on all Hue hosts. See [Installing the psycopg2 Python package](#).

Cloudera Issue: OPSAPS-47080

CDH Upgrade fails to delete Solr data from HDFS

The CDH upgrade process fails to delete Solr data from HDFS and the recreated collections fail to be initialized due to the existing indexes.

Workaround: Perform the following steps *after* you run the CDH Upgrade wizard and *before* you finalize the HDFS upgrade:

1. Log in to the Cloudera Manager Admin Console.
2. Go to the Solr service page.
3. Stop the Solr service and dependent services. Click **Actions > Stop**.
4. Click **Actions > Reinitialize Solr State for Upgrade**.
5. Click **Actions > Bootstrap Solr Configuration**.
6. Start the Solr and dependent services. Click **Actions > Start**.
7. Click **Actions > Bootstrap Solr Collections**.

Affected Versions: CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47502

Package Installation of CDH Fails

When you install CDH with packages from a custom repository, ensure that the version of CDH you select for **Select the version of CDH** matches the version of CDH for the custom repository. Selecting the CDH version and specifying a custom repository are done during the **Select Repository** stage of installation.

If the versions do not match, installation fails.

Affected Versions: Cloudera Manager 6.x

Fixed Versions: N/A

Apache Issue: N/A

Cloudera Issue: OPSAPS-45703

Uninstall CDH 5 Sqoop connectors for Teradata and Netezza before upgrading to CDH 6

Sqoop includes two connectors, one for Teradata and one for Netezza. The connectors are released in separate parcels and tarballs and can be installed in Cloudera Manager or manually. The versioning of the connectors takes the form <connector_version>c<major_cdh_version>. For example, 1.6c5 refers to the connector 1.6 for CDH 5. The manifest files do not prohibit installing the CDH 5 connectors on CDH 6, but they are not compatible with CDH 6.

If you have CDH 5 connectors installed, they will not be automatically upgraded during the CDH upgrade, and they are not compatible with CDH 6, so they should be uninstalled before the upgrade. Keeping the CDH 5 connectors will not cause the upgrade to fail, but instead will cause a failure to occur during Sqoop runtime. Cloudera will release the connectors for CDH 6 at a later time.

For more information about the Teradata and Netezza connectors, go to [Cloudera Enterprise Connector Documentation](#) and choose the connector and version to see the documentation for your connector.

Unsupported Sqoop options cause upgrade failures

New [fail-fast](#) checks for unsupported options were introduced in CDH 6. Users should check the jobs stored in their Sqoop metastore and remove all unsupported options. Some unsupported options were silently ignored in earlier CDH versions during upgrades, but in CDH 6, the same options fail instantly. See the following JIRAs in [Apache Sqoop Incompatible Changes](#):

Generated Avro code from CDH 5 should be regenerated when upgrading

Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.

Upgrading Apache Parquet to CDH 6

Parquet packages and the project's group ID were renamed, and some of the class methods were removed.

If you directly consumes the Parquet API instead of using Parquet through a CDH component, your need to update and recompile your code. See [Parquet API Change](#) for details of the changes.

No HBase Replication Peer Configuration Change During Rolling Update

When doing a rolling upgrade from a CDH 6 version to a higher version, do not do any replication peer configuration changes. This includes removing a peer, adding a peer, and changing the configuration on a peer.

Oracle Database Initialization

Before upgrading from CDH 5 to CDH 6, check the value of the COMPATIBLE initialization parameter in the Oracle Database using the following SQL query:

```
SELECT name, value FROM v$parameter WHERE name = 'compatible'
```

The default value is 12.2.0. If the parameter has a different value, you can set it to the default as shown in the [Oracle Database Upgrade Guide](#).



Note: Before resetting the COMPATIBLE initialization parameter to its default value, make sure you consider the effects of this change can have on your system.

TLS Protocol Error with OpenJDK

If you are using an older version of OpenJDK 1.8 and have enabled SSL/TLS for the Cloudera Manager Admin Console, you may encounter a TLS protocol error when connecting to the Admin Console, stating that there are no ciphers in common. This is because older versions of OpenJDK may not implement certain TLS ciphers, causing an inability to log into the Cloudera Manager Admin Console when TLS is enabled.

Workaround:

You can workaround this issue by doing one of the following:

- Upgrade OpenJDK to a [supported version of OpenJDK](#) that is higher than version 1.8.0_181.
- If it is not possible to upgrade OpenJDK, enable less secure TLS ciphers in Cloudera Manager. You can do this by opening the `/etc/default/cloudera-scm-server` in a text editor and adding the following line:

```
export CMF_OVERRIDE_TLS_CIPHERS=<cipher_list>
```

Where `<cipher_list>` is a list of TLS cipher suites separated by colons. For example:

```
export
```

Cloudera Bug: OPSAPS-49578

Upgrades to Cloudera Enterprise 5.x

Flume Kafka client incompatible changes in CDH 5.8

Due to the change of offset storage from ZooKeeper to Kafka in the CDH 5.8 Flume Kafka client, data might not be consumed by the Flume agents, or might be duplicated (if `kafka.auto.offset.reset=smallest`) during an upgrade to CDH 5.8.

Cloudera Issue: TSB-173

Upgrade to CDH 5.13 or higher Requires Pre-installation of Spark 2.1 or Spark 2.2

If your cluster has Spark 2.0 or Spark 2.1 installed and you want to upgrade to CDH 5.13 or higher, you must first upgrade to Spark 2.1 release 2 or later before upgrading CDH. To install these versions of Spark, do the following before running the CDH Upgrade Wizard:

1. Install the Custom Service Descriptor (CSD) file. See

- [Installing Spark 2.1](#)
- [Installing Spark 2.2](#)



Note:

Spark 2.2 requires that JDK 1.8 be deployed throughout the cluster. JDK 1.7 is not supported for Spark 2.2.

See [Step 2: Install Java Development Kit](#).

2. Download, distribute, and activate the Parcel for the version of Spark that you are installing:

- **Spark 2.1 release 2:** The parcel name includes "cloudera2" in its name.
- **Spark 2.2 release 1:** The parcel name includes "cloudera1" in its name.

See [Managing Parcels](#).

Affected versions: CDH 5.13.0 and higher

Cloudera Issue: CDH-56775

Sentry may require increased Java heap settings before upgrading CDH to 5.13

Before upgrading to CDH 5.13 or higher, you may need to increase the size of the Java heap for Sentry. A warning will be displayed during upgrade, but it is the user's responsibility to ensure this setting is adjusted properly before proceeding. See [Performance Guidelines](#).

Affected versions: CDH 5.13 or higher

Cloudera Issue: OPSAPS-42541

[Apache MapReduce Jobs May Fail During Rolling Upgrade to CDH 5.11.0 or CDH 5.11.1](#)

In CDH 5.11, Cloudera introduced four new counters that are reported by MapReduce jobs. During a rolling upgrade from a cluster running CDH 5.10.x or lower to CDH 5.11.0 or CDH 5.11.1, a MapReduce job with an application master running on a host running CDH 5.10.x or lower may launch a map or reduce task on one of the newly-upgraded CDH 5.11.0 or CDH 5.11.1 hosts. The new task will attempt to report the new counter values, which the old application master will not understand, causing an error in the logs similar to the following:

```
2017-06-08 17:43:37,173 WARN [Socket Reader #1 for port 41187]
org.apache.hadoop.ipc.Server: Unable to read call parameters for client 10.17.242.22 on
connection protocol org.apache.hadoop.mapred.TaskUmbilicalProtocol for rpcKind
RPC_WRITABLE
java.lang.ArrayIndexOutOfBoundsException: 23
    at
...
...
```

This error could cause the task and the job to fail.

Workaround:

Avoid performing a rolling upgrade to CDH 5.11.0 or CDH 5.11.1 from CDH 5.10.x or lower. Instead, skip CDH 5.11.0 and CDH 5.11.1 if you are performing a rolling upgrade, and upgrade to CDH 5.12 or higher, or CDH 5.11.2 or higher when the release becomes available.

Cloudera Issue: DOCS-2384, TSB-241

[Cloudera Manager set catalogd default jvm memory to 4G can cause out of memory error on upgrade to Cloudera Manager 5.7 or higher](#)

After upgrading to 5.7 or higher, you might see a reduced Java heap maximum on Impala Catalog Server due to a change in its default value. Upgrading from Cloudera Manager lower than 5.7 to Cloudera Manager 5.8.2 no longer causes any effective change in the Impala Catalog Server Java Heap size.

When upgrading from Cloudera Manager 5.7 or later to Cloudera Manager 5.8.2, if the Impala Catalog Server Java Heap Size is set at the default (4GB), it is automatically changed to either 1/4 of the physical RAM on that host, or 32GB, whichever is lower. This can result in a higher or a lower heap, which could cause additional resource contention or out of memory errors, respectively.

Cloudera Issue: OPSAPS-34039

Cloudera Manager 6 Release Notes

These Release Notes provide information on the new features and known issues and limitations for Cloudera Manager 6. These Release Notes also include fixed issues for releases starting from Cloudera Manager 6.0.0.

For information about supported operating systems, and other requirements for using Cloudera Manager, see [Cloudera Enterprise 6 Requirements and Supported Versions](#) on page 5.

To view the Release Notes for a specific Cloudera Manager release, see below:

Cloudera Manager 6.2.x Release Notes

To view release notes for specific Cloudera Manager 6.2.x releases, see the following:

[Cloudera Manager 6.2.0 Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for Cloudera Manager 6.2.0:

[New Features and Changes in Cloudera Manager 6.2.0](#)

The following sections describe new and changed features for Cloudera Manager 6.2.0:

[*Virtual Private Clusters - Separation of Compute and Storage services*](#)

A Virtual Private Cluster uses the Cloudera Shared Data Experience (SDX) to simplify deployment of both on-premise and cloud-based applications and enable workloads running in different clusters to securely and flexibly share data.

A new type of cluster is available in CDH 6.2, called a *Compute cluster*. A Compute cluster runs computational services such as Impala, Spark, or YARN but you configure these services to access data hosted in another Regular CDH cluster, called the *Base cluster*. Using this architecture you can separate compute and storage resources in a variety of ways to flexibly maximize resources.

See [Virtual Private Clusters and Cloudera SDX](#).

[*Ubuntu 18 Support*](#)

Support for Ubuntu 18.04 has been added for Cloudera Manager and CDH 6.2 and higher.

Cloudera Issue: OPSAPS-48410

[*Backup and Disaster Recovery \(BDR\)*](#)

Hive Direct Replication to S3/ADLS Backed Cluster

BDR now supports Hive direct replication from on-premise to S3/ADLS clusters and metadata replication to the Hive Metastore.

Using a single replication process, BDR enables Hive data to be pulled from HDFS to S3/ADLS clusters and use the "Hive-on-cloud" mode, where the target Hive Metastore updates the table locations to point to S3/ADLS clusters. This process facilitates easy data migration and synchronisation between the cloud and on-premise clusters.

For more information, see [Hive/Impala Replication](#).

Replication to and from ADLS Gen2

You can now replicate HDFS files and Hive data to and from Microsoft ADLS Gen2. To use ADLS Gen2 as the source or destination, you must add Azure credentials to Cloudera Manager. Note that the URI format for ADLS Gen2 is not the same as ADLS Gen1. For ADLS Gen2 use the following URI format:

`abfs[s]://<file_system>@<account_name>.dfs.core.windows.net/<path>/.`

[*Hosts*](#)

Duplicate Host Detection and Hostname Migration

Cloudera Manager now detects and rejects duplicate hosts from joining a cluster and gracefully tolerates changes in hostnames for managed hosts, better supporting automated deployments

[*Installation*](#)

Accumulo Initialization

An Initialize Accumulo checkbox now displays in the Installation wizard.

Cloudera Issue: OPSAPS-48619

JDBC URL for the Hive Metastore Database Connection

You can now specify a JDBC URL when establishing a connection from the Hive service to a supported backend database (MySQL, PostgreSQL, or OracleDB). Enter the JDBC URL on the Setup Database page in the Create Cluster and Create Service wizards in Cloudera Manager.

Cloudera Enterprise 6 Release Guide

Cloudera Issue: OPSAPS-48668

Licensing

Start and Deactivation Dates for Cloudera Enterprise Licenses

Cloudera Enterprise licenses now include a start date and a deactivation date. Enterprise-only features are enabled on the start date and will be disabled after the deactivation date. If you install the license before the start date, a banner displays in the Cloudera Manager Admin console showing the number of days until the license becomes effective.

Cloudera Issue: OPSAPS-47500

Enhanced License Enforcement - Node Limit

When an Enterprise license expires, Cloudera Manager reverts to the Express version. This includes enforcing a maximum of 100 nodes across all CDH 6 clusters.

Cloudera Issue: OPSAPS-48611

Enhanced License Enforcement - Feature Availability

Features only available with a Cloudera Enterprise license are turned off after the deactivation date has passed. For legacy licenses that do not have a deactivation date, the features are turned off on the expiration date.

Cloudera Issue: OPSAPS-46864

Enhanced License Enforcement - KMS Configuration

Cloudera Manager will not allow KMS configuration changes after the deactivation date specified in the new license file although the KMS will remain functional. For legacy licenses, the deactivation date defaults to the expiration date specified in the license.

Cloudera Issue: OPSAPS-48501

Cloudera Manager API

Cross-Cluster Network Bandwidth Test

Cloudera Manager now has an API to test network bandwidth between clusters, helping determine if the infrastructure is suitable for separating storage and compute services.

For more information, see the following entries in the [Cloudera Manager REST API documentation](#):

- [POST /cm/commands/clustersPerfInspector](#)
- [ApiClustersPerfInspectorArgs](#)
- [ApiPerfInspectorBandwidthArgs](#)

API for Managing Expiring Cloudera Manager Sessions

There is a new Cloudera Manager API endpoint, `/users/expireSessions/{UserName}` that can be invoked by a user with the Full administrator or User administrator role that expires all of a particular user's active Cloudera Manager Admin console sessions - local or external. Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-43756

Service Type Information in the ApiServiceRef

The Cloudera Manager API endpoint `ApiServiceRef` now returns the service type. Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-48369

API to Emit All Features Available

A new attribute/property `features` has been added to the API endpoint `/cm/license`. It lists all the features that are available in the product for the given license. For example:

```
{ "owner" : """John Smith""", "uuid" : "12c8052f-d78f-4a8e-bba4-a55a2d141fcc" ,  
"features" : [ { "name" : "PEERS", "description" : "Peers" }, { "name" :
```

```
  "BDR" , "description" : "BDR" } , { "name" : "KERBEROS" , "description" :
  "Kerberos" } , . . .
```

Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-49060

New Name Attribute for ApiAuthRole

ApiAuthRole entities can now be specified and looked up with a name string for the role, as specified in the API documentation. Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-46780

Kafka Configuration and Monitoring

New Kafka Metrics

The following metrics have been added:

- kafka_topic_unclean_leader_election_enable_rate_and_time_ms
- kafka_incremental_fetch_session_evictions_rate -
- kafka_num_incremental_fetch_partitions_cached -
- kafka_num_incremental_fetch_sessions
- kafka_groups_completing_rebalance
- kafka_groups_dead
- kafka_groups_empty
- kafka_groups_preparing_rebalance
- kafka_groups_stable
- kafka_zookeeper_request_latency
- kafka_zookeeper_auth_failures
- kafka_zookeeper_disconnects
- kafka_zookeeper_expires
- kafka_zookeeper_read_only_connects
- kafka_zookeeper_sasl_authentications
- kafka_zookeeper_sync_connects

The following metric is deprecated: kafka_responses_being_sent

Cloudera Issue: OPSAPS-48911, OPSAPS-48798, OPSAPS-48311, OPSAPS-48656

Kafka Broker ID Display

Kafka Broker IDs are now displayed on the Cloudera Manager's Kafka Instances page.

Cloudera Issue: OPSAPS-44331

Kafka Topics in the diagnostic bundle

Diagnostic bundles for Kafka will now include the output of the following commands:

- kafka-topics --describe
- kafka-topics --list

Cloudera Issue: OPSAPS-36755

Kafka Configuration Properties for Delegation Tokens

The following new configuration parameters required to configure Kafka delegation tokens have been added:

- delegation.token.max.lifetime.ms

The token has a maximum lifetime beyond which it cannot be renewed anymore. Default value 7 days.

- Delegation.token.expiry.time.ms

The token validity time in seconds before the token needs to be renewed. Default value 1 day.

Cloudera Issue: OPSAPS-47051

Enhanced Security for Kafka in Zookeeper with ACLs

A new script, `zookeeper-security-migration.sh` script is now available to lock down Kafka data in Zookeeper. See [Kafka Security Hardening with Zookeeper ACLs](#).

Cloudera Issue: OPSAPS-47988

Hive Server 2

New Graph for the Compilation Metrics

A new graph, **Operations Awaiting Compilation** for HiveServer2 compilation metrics has been added.

Cloudera Issue: OPSAPS-47506

Secured ADLS Credentials for HS2

ADLS credentials are now stored securely via Cloudera Manager for use with HS2. This enables multi-user Hive-with-ADLS clusters.

Learn more at [Configuring ADLS Access Using Cloudera Manager](#).

Cloudera Issue: OPSAPS-49076

Secured S3 Credentials HS2 on S3

S3 credentials are now stored securely by Cloudera Manager for use with Hive. This enables multi-user Hive-on-S3 clusters.

Learn more at [Configuring the Amazon S3 Connector](#).

The following sub-tasks are related to this feature:

- Distribute the path of the HDFS credential store file and decryption password to HS2

Adds job credstore path and decryption password propagation for HS2.

Cloudera Issue: OPSAPS-48662

- Manage an encrypted credential store in HDFS for HS2

Adds a job specific credstore for HS2.

Cloudera Issue: OPSAPS-48661

- Rotate the password and the encrypted credential file in HDFS on every HS2 restart

Adds password and credstore file rotation on every HS2 role restart.

Cloudera Issue: OPSAPS-48663

delegation.token.master.key Generation

`delegation.token.master.key` is now automatically generated by Cloudera Manager/.

Cloudera Issue: OPSAPS-48525

New Warning for Hue Advanced Configuration Snippet

Warnings will be emitted if the values for **Hue Service Advanced Configuration Snippet** or **Hue Server Advanced Configuration Snippet** are not formatted properly. For example, if it does not contain a configuration section like `[desktop]`.

Cloudera Issue: OPSAPS-27606

[Increased Default Value for dfs.client.block.write.locateFollowingBlock.retries configuration](#)

The default value for the HDFS configuration `dfs.client.block.write.locateFollowingBlock.retries` configuration's has been changed from 5 to 7.

Cloudera Issue: OPSAPS-48170

[Support GPU Scheduling and Isolation for YARN](#)

Added support to enable usage of GPUs in YARN applications and for custom YARN resource types.

Cloudera Issue: OPSAPS-48685

[Health Test for Erasure Coding Policies](#)

A new **Verify Erasure Coding Policies For Cluster Topology** health test has been introduced. The health test fails with a yellow status if there are not enough data nodes or racks to support all enabled erasure coding policies.

Cloudera Issue: OPSAPS-48526

[Disk Caching Configurations in Spark Service](#)

Disk caching for the Spark History Server can now be enabled from Cloudera Manager.

Cloudera Issue: OPSAPS-48385

[Decimal Support for Sqoop Clients](#)

Sqoop decimal support for Parquet and Avro imports will now be turned on by default for new CDH 6.2 (or higher) clusters. In the case of an newly upgraded cluster, decimal support must be enabled manually.

- Setting the following property to enable decimal support in Avro:
`sqoop.avro.logical_types.decimal.enable=true`
- Setting the following properties to enable decimal support in Parquet:
`sqoop.parquet.logical_types.decimal.enable=true`
`parquet.job.configurator.implementation=hadoop`

Please note that changing any of these properties might break existing Sqoop jobs, or alter their output in a way that disrupts consumers further down the chain.

Cloudera Issue: OPSAPS-48938

[TLS](#)

Apply Auto-TLS Configuration to Existing Services

You can now use Auto-TLS to add TLS to an existing cluster. This functionality is available in both the Cloudera Manager Admin Console and by using the API. See [Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS](#),

There is a new cluster Cloudera Manager API command `ConfigureAutoTlsServices` which will enable Auto-TLS for services in a single cluster. Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-47349

Support for TLS proto/ciphers in Custom Service Descriptors (CSD)

Added the ability to specify the TLS protocol and the TLS cipher suites in CSDs.

Cloudera Issue: OPSAPS-48214

Expose the configurations to use TLS encryption to the Hive Metastore Database on the Hive Metastore (Hive) Configurations Page

Exposes properties that can be used to configure TLS from the Hive Metastore Server to the Hive Metastore Database. As a minimum configuration requirement, the **Enable TLS/SSL to the Hive Metastore Database** checkbox must be enabled. (The default value is disabled.) If the **Hive Metastore TLS/SSL Client Truststore** properties are provided, then those will be used. Otherwise, the default list of well-known certificate authorities will be used. Additionally, ability to

Cloudera Enterprise 6 Release Guide

provide a JDBC URL override to use when connecting to the database is also exposed. This will override all other values used to create the JDBC URL. This is an advanced configuration option and should only be used as a safety-valve.

Cloudera Issue: OPSAPS-48666

Enable Auto-TLS Globally

There is now a Cloudera Manager API command, `GenerateCmcaCommand`, which will enable Auto-TLS for an existing Cloudera Manager deployment. This command creates an internal Cloudera Manager Certificate Authority (CMCA) and certificates for all existing hosts. Please refer to the [Cloudera Manager REST API documentation](#) for more information.

Cloudera Issue: OPSAPS-43102

Kafka/Flume Auto-TLS enhancements

Flume now supports Auto-TLS when used with Kafka.

Cloudera Issue: OPSAPS-46339

License Enforcement - Auto TLS

Auto-TLS is not available when using a Trial license. To enable Auto-TLS, you must have an Enterprise license.

Cloudera Issue: OPSAPS-48981

Custom certificates for Cloudera Manager Certificate Authority (CMCA)

When using Auto-TLS with custom certificates, you can use the new `AddCustomCerts` command to add certificates associated with a hostname to the Auto-TLS certificate database. Please refer to the [Cloudera Manager REST API documentation](#) for more information. details.

Cloudera Issue: OPSAPS-48678

Fixed Issues in Cloudera Manager 6.2.0

The following sections describes issue fixed in Cloudera Manager 6.2.0:

Default thresholds for swap warnings are too low

The default process swap size warning threshold has been increased from 0 to 200Mb. A new host level configuration property, **Default Process Swap Memory Thresholds**, has been added for process swap size alert thresholds. This allows for bulk updates of process level alerts that use the default settings.

Cloudera Bug: OPSAPS-44904

Service name auto-generation code is broken, causing a constraint violation

Fixed an issue where Cloudera Manager incorrectly auto-generated the Service Display Name when adding services.

Cloudera Bug: OPSAPS-48672

Hive replication does not copy an empty database

Empty databases without metadata or Hive/Impala UDFs are now replicated during Hive replication.

Cloudera Bug: OPSAPS-47224

Cluster template import causes state corruption

Fixed an issue with importing Cluster Templates. When hosts that are used in another cluster are, by mistake, specified in the template, then these hosts are dissociated with their cluster and the cluster is left in an invalid state and beyond repair.

Cloudera Bug: OPSAPS-48680

Summary and full views in the Replication API display the same result

Fixed an issue where using the Cloudera Manager API to return a summary view for a replication job returns a summary with limited information instead of the full view.

Cloudera Bug: OPSAPS-47182

Setting displayName in cluster template does not impact display name

Fixed an issue where setting the display name for a service using a cluster template did not change the service name.

Cloudera Bug: OPSAPS-46056

Server fails to contact agent when TLS is on

Fixed an issue with TLS encryption. When TLS encryption is enabled for agent-server communication, but no key and certificate is configured for an agent host, then the agent might erroneously report that it is serving the agent status server (port 9000) over HTTPS when it is not. This caused failures during diagnostic bundle collection.

To configure the agent status server to use HTTPS, configure the agent with a client_cert_file and client_key_file in the /etc/cloudera-scm-agent/config.ini file.

Cloudera Bug: OPSAPS-48958

Host inspector incorrectly reports bad version of PSYCOPG2

Fixed an issue where the Host Inspector may incorrectly report that an incompatible version of PSYCOPG2 was in use.

Cloudera Bug: OPSAPS-48649

Diagnostic bundle has zero-length agent log .zip files

Fixed an issue where zero-length agent log .zip files were included in diagnostic bundles.

Cloudera Bug: OPSAPS-49208

Import/Export cluster deployment through API for inbuilt/custom roles

Fixed an issue where using the Cloudera Manager API to import and export deployments failed when using custom roles.

Cloudera Bug: OPSAPS-48104

Cannot negotiate TLS to Cloudera Manager Admin Console on OpenJDK 1.8 on Centos 7.2

Fixed an issue where older versions of OpenJDK may not implement certain TLS ciphers, causing an inability to log into the Cloudera Manager Admin Console when TLS is enabled. Either upgrade the version of OpenJDK, or allow Cloudera Manager to use less secure ciphers by editing /etc/default/cloudera-scm-server and uncommenting the line that contains "export CMF_OVERRIDE_TLS_CIPHERS=".

Cloudera Bug: OPSAPS-49578

YARN jobs failure when using encrypted shuffle

On some operating systems, the default file mask may cause the Auto-TLS truststore file to not be world-readable, causing YARN jobs to fail when encrypted shuffle is enabled. This issue has been fixed.

Cloudera Bug: OPSAPS-48731

Uploading a valid license via the Cloudera Manager API Swagger client returns 500 Server Error

Fixed the update_license() method in ClouderaManagerResourceApi in the Cloudera Manager Python Swagger-based API client that allows updating the license file for Cloudera Manager. Without the fix, the update_license() method fails with the following error: "No multipart with content id license found" HTTP error 500.

Cloudera Bug: OPSAPS-49116

Syntax highlighting for Java API client docs is broken

Fixed an issue with syntax highlighting for the Cloudera Manager Java API client SDK documentation.

Cloudera Bug: OPSAPS-49406

Disable Hive > Actions > Update Hive Metastore NameNodes if high availability is not enabled for HDFS

Fixed an issue where re-running the Update Hive Metastore NameNodes Command when HDFS High Availability was disabled corrupted the URI paths in Hive tables by appending an additional port number.

Cloudera Bug: OPSAPS-46970

Cloudera Navigator Limitation when using Virtual Private Clusters

Navigator does not capture auditing and lineage information from Compute clusters in version 6.2.0.

Cloudera Bug: OPSAPS-48694

Upgrade CDH screen does not list all the databases that need to be backed up

When you run the Upgrade Cluster wizard to upgrade from CDH 6.1 to 6.2, Hive now appears in the list of databases to be backed up.

Cloudera Bug: OPSAPS-49314

dfs.client.read.shortcircuit needs to be set correctly for each Impala instance

When Impala daemons are not co-located with a DataNode, Advanced Configuration Snippets are no longer necessary to disable short circuit reads when using CDH 6.2 or higher.

Cloudera Bug: OPSAPS-46971

Maven dependency issue for jaxrs

Fixed an issue where some of the Cloudera Manager API endpoints stopped working in version 6.1.0.

Cloudera Bug: OPSAPS-49159

Java runtime error when setting static pool through Cloudera Manager

Fixed a RuntimeException that occurs when configuring Static Pools in Cloudera Manager.

Cloudera Bug: OPSAPS-48476

Snapshot Policies screen hangs

Fixed an issue where the Snapshots Policies screen in the Cloudera Manager Admin Console hangs when there are scheduled replications with snapshots.

Cloudera Bug: OPSAPS-49511

Make sure Spark app log collection works with authentication on

Fixed several issues with the YARN diagnostic bundle, which could not collect Spark event logs when SSL or Kerberos authentication was enabled for the Spark History Server.

Cloudera Bug: OPSAPS-39280

New compile lock related configuration values to Cloudera Manager

The following HiveServer2 configuration parameters for compile locking have been added:

- hive.driver.parallel.compilation
- hive.driver.parallel.compilation.global.limit

Cloudera Bug: OPSAPS-47503

Add pre-upgrade confirmation box to HBase upgrade

Added two checkboxes to the HBase upgrade screen for validating co-processors and tables.

Cloudera Bug: OPSAPS-48815

Diagnostic bundle creation slows down Cloudera Managers

Fixed an issue where running a diagnostic bundle command slows down other commands running in Cloudera Manager.

Cloudera Bug: OPSAPS-49233

Broken Link on Replication History page

Fixed a broken link on the Replication History page to the respective YARN jobs.

Cloudera Bug: OPSAPS-29486

Incompatible Changes in Cloudera Manager 6.2

See below for incompatible changes in Cloudera Manager 6.2.0:

Default thresholds for swap warnings are too low

The default process swap size warning threshold has been increased from 0 to 200Mb. A new host level configuration property, **Default Process Swap Memory Thresholds**, has been added for process swap size alert thresholds. This allows for bulk updates of process level alerts that use the default settings.

Cloudera Bug: OPSAPS-44904

Known Issues and Limitations in Cloudera Manager 6.2.0

The following sections describe known issues and limitations for Cloudera Manager 6.2.0:

Cloudera Manager 6.2.0 does not have the correct license notification

Cloudera Manager 6.2.0 contains the third-party license notification for a previous release. You can view the correct license notification file [here](#) or on the [Third-party License Page for Cloudera Manager](#).

Limitations for Virtual Private Clusters

There are a number of limitations and considerations for running Virtual Private Clusters, including the types of services you can run on a Compute cluster and supported versions of CDH. See [Compatibility Considerations for Virtual Private Clusters](#).

Restart of Impala and Hive required for Cloudera Manager 6.2 upgrade with ADLS

After upgrading to Cloudera Manager 6.2 or higher, Impala and Hive will be marked as stale for users running CDH 6.1 and using the ADLS Service. You will need to restart Hive and Impala before being able to connect to ADLS Gen2, but all previous functionality will continue to work without a restart. The configurations that will be marked stale are:

- fs.azure.account.auth.type
- fs.azure.account.oauth.provider.type
- fs.azure.account.oauth2.client.endpoint
- fs.azure.account.oauth2.client.id
- fs.azure.account.oauth2.client.secret.

Cloudera Bug: OPSAPS-47436

Add Hive Execution Service on Compute Cluster for Hue

To enable Hue to run Hive queries on a Compute cluster, you must install the **Hive Execution Service** on the Compute cluster.

Alternately, you can disable the Hive editor in Hue to prevent users from using it (it will not work correctly) by doing the following:

1. In the Cloudera Manager Admin console, go to the Hue service on the Compute cluster.
2. Open the **Hue Web UI**.
3. Select **Admin > Manage Users**.
4. Select the **Group** tab.
5. Click on the row containing the **default** group.
6. De-select the **beeswax.access:Launch this application** permission.
7. Click **Update Group**.

Cloudera Bug: DOCS-4438 OPSAPS-49062

BDR invalidate metadata command

When running a Hive replication job, the **invalidate metadata** command is run automatically by the replication job and runs as the **impala** user. If an administrator has configured a different user with permissions to run the **invalidate metadata** command, the command fails.

Workaround: Grant the Impala user permission to run the **invalidate metadata** command.

Cloudera Bug: OPSAPS-49215

TLS Protocol Error with OpenJDK

If you are using an older version of OpenJDK 1.8 and have enabled SSL/TLS for the Cloudera Manager Admin Console, you may encounter a TLS protocol error when connecting to the Admin Console, stating that there are no ciphers in common. This is because older versions of OpenJDK may not implement certain TLS ciphers, causing an inability to log into the Cloudera Manager Admin Console when TLS is enabled.

Workaround:

You can workaround this issue by doing one of the following:

- Upgrade OpenJDK to a [supported version of OpenJDK](#) that is higher than version 1.8.0_181.
- If it is not possible to upgrade OpenJDK, enable less secure TLS ciphers in Cloudera Manager. You can do this by opening the `/etc/default/cloudera-scm-server` in a text editor and adding the following line:

```
export CMF_OVERRIDE_TLS_CIPHERS=<cipher_list>
```

Where `<cipher_list>` is a list of TLS cipher suites separated by colons. For example:

```
export
```

Cloudera Bug: OPSAPS-49578

Cloudera Manager 6.1.x Release Notes

To view release notes for specific Cloudera Manager 6.1.x releases, see the following:

[Cloudera Manager 6.1.1 Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for Cloudera Manager 6.1.1:

[New Features in Cloudera Manager 6.1.1](#)

The following sections describe new and changed features for Cloudera Manager 6.1.1:

There are no new features in Cloudera Manager 6.1.1. See also [Known Issues and Limitations in Cloudera Manager 6.1.1](#) on page 147.

[Fixed Issues in Cloudera Manager 6.1.1](#)

The following sections describes issues fixed in Cloudera Manager 6.1.1:

[Backup and disaster recovery \(BDR\) HDFS and Hive replications will fail on clusters running Cloudera Manager 6.1.0](#)

This issue caused BDR HDFS and Hive replications to fail when you replicated from secured (Kerberized) source clusters to destination clusters that have been upgraded to Cloudera Manager 6.1.0.

This issue has been fixed in this release. After upgrade, the schedules will correctly identify the source as either a secure or unsecure cluster. In the case of multiple peers, one secure and another unsecure peer is supported.

Cloudera Issue: OPSAPS-48865

[The Add Hosts link on the Cloudera Director page is broken in the public cloud scenario](#)

When you add hosts to Cloudera Manager deployed in AWS, a page that describes Cloudera Director is displayed. There is a link to the classic Add Hosts wizard. This wizard was broken in version 6.0.0 and 6.0.1. This issue is fixed in this release.

Cloudera Issue: OPSAPS-48627

BDR fails if dfs.nameservices is overridden with multiple nameservice names

This fix:

1. Adds a feature flag that you can use to disable the replace nameservice feature by calling the following API call:

```
http://cm_host:cm_port/api/v19/cm/config
```

with the body:

```
{ "items": [ { "name": "feature_flag_bdr_replace_nameservice", "value": "false" } ] }
```

2. Handles the scenario for multiple nameservices in dfs.nameservices configuration. Now the feature cross-checks that with the fs.defaultFS configured in core-site.xml.

Cloudera Issue: OPSAPS-48579

The diff format is not displayed on the configuration revisions diff page

The delta between the old and the new configuration value was previously shown using a red background color to indicate removal, and a green color to indicate addition. This was not displaying correctly in Cloudera Manager 6.1.0. This issue is fixed in this release.

Cloudera Issue: OPSAPS-48544

Decommissioning a DataNode during a current decommission (in parallel) never completes

With this change, you can decommission DataNodes in parallel. They won't appear to be stuck as incomplete on the user interface.

Cloudera Issue: OPSAPS-39746

Known Issues and Limitations in Cloudera Manager 6.1.1

The following sections describe known issues and limitations for Cloudera Manager 6.1.1:

Backup and Disaster Recovery (BDR) performance regression after upgrading to CDH 6.0.0

Hive replication with BDR experiences a performance regression when comparing CDH 6.0.0 and CDH 5.14.4. The slowdown occurs during the import step. For example, the performance regression may only be 10% for 4 million partitions. As the number of partitions goes down though, the performance impact becomes more visible. For example, 100,000 partitions may experience a 20% performance regression.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0; CDH 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47520

Cloudera Manager allows more than a single space in YARN Admin ACLs

When adding a YARN Admin ACL in Cloudera Manager, you are allowed to enter multiple spaces in the entry. The space is the separator between the user and group lists, and only a single space should be allowed in the entry. All entries that appear after a second single space in a YARN Admin ACL will be ignored.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47688

Integer data types map to Float in Swagger API client

Integer data types show up as floating point numbers when using the Cloudera Manager API Python client.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-45689

User Sessions page doesn't update with a newly logged in SAML user

If you log into Cloudera Manager as the administrator user, and then log into Cloudera Manager with a SAML user through a different browser, the SAML user does not appear on the **User Sessions** page.

Affected Versions: 6.0.0, 6.0.1

Cloudera Issue: OPSAPS-47025

Package Installation of CDH Fails

When you install CDH with packages from a custom repository, ensure that the version of CDH you select for **Select the version of CDH** matches the version of CDH for the custom repository. Selecting the CDH version and specifying a custom repository are done during the **Select Repository** stage of installation.

If the versions do not match, installation fails.

Affected Versions: Cloudera Manager 6.x

Fixed Versions: N/A

Apache Issue: N/A

Cloudera Issue: OPSAPS-45703

Cloudera Manager 6.1.0 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for Cloudera Manager 6.1.0:

New Features and Changes in Cloudera Manager 6.1.0

The following sections describe new and changed features for Cloudera Manager 6.1.0:

Accumulo

Accumulo installations now use the Hadoop Credential Provider to handle sensitive properties. For example, the instance secret and trace user password.

Agents

The Cloudera Manager Admin console now displays a message if the agent is hearbeating with an invalid CM_GUID

API endpoints for Roles

For API documentation and the new Swagger-based API client, roles can be accessed from Roles Resource instead of Services Resource. This does not change the roles endpoint and does not impact those accessing the Cloudera Manager API endpoints directly using tools like curl.

Audit Events

Cloudera Manager logs events in the Audits database table when the API is accessed either from the Cloudera Manager Admin Console or from any other client. When the API is accessed at a rapid rate, the Audits database table grows rapidly, negatively impacting Cloudera Manager performance.

Cloudera Manager now collects similar audit events that occur during a configurable period into a unique audit entry in the Audits database table. This can prevent the Audits table from being filled at a rapid rate. This feature can be configured by setting arguments to **CMF_JAVA_OPTS** in **cloudera-scm-server.properties**:

- **com.cloudera.cmf.persist.cmAuditTrackerConfig.timeToLiveMs** : Period during which similar audit entries will be coalesced into one. Default is 10000 milliseconds. Setting this value to 0 disables this feature entirely
- **com.cloudera.cmf.persist.cmEventCoalescer.maxTrackedEvents**: Number of maximum events that can be candidates for coalescing in a certain period. Default is 1024. If this limit is reached, then the oldest event is removed.

Auto-TLS

Certificate Handling

The **certmanager** can now use the following option to automatically skip invalid certificates and import the rest of the bundle: `--skip-invalid-ca-certs`. Previously, if one or more of the certificates in a bundle were invalid, then the entire setup operation failed.

Randomization of Sequential Certificate Authority Serial Numbers

Previously, certificates generated by Auto-TLS always started at serial number 0. Now, certificates will start from a random serial number. This affects only new deployments using Auto-TLS. Existing deployments using Auto-TLS are unaffected.

Supported Services

Auto-TLS now supports the following services: Flume, Java Keystore KMS, KeyTrustee server, KeyTrustee KMS, Thales HSM KMS, and Luna HSM KMS. When adding these services while Auto-TLS is enabled, TLS configuration will be added automatically.

Backup and Disaster Recovery (BDR)

Insecure Cluster to Secure Cluster Replication

You can now use BDR to replicate data from an insecure cluster that does not use Kerberos authentication, to a secure cluster that uses Kerberos. Note that the reverse is not true. BDR does not support replicating from a secure cluster to an insecure cluster.

To perform the replication, the destination cluster must be managed by Cloudera Manager 6.1.0 or higher. The source cluster must run Cloudera Manager 5.14.0 or higher in order to replicate to Cloudera Manager 6.

For more information, see [Replicating from Insecure to Secure Clusters](#) for Hive or [Replicating from Insecure to Secure Clusters](#) for HDFS.

Invalidate Metadata

BDR enhanced the **Invalidate Metadata** option so that the command is issued per Impala service after replication. This ensures that if a cluster has multiple Impala services, only the target Impala's metadata cash will be invalidated and require refresh, which can impact performance.

Kudu

BDR now ignores Hive tables backed by Kudu during replication. The change does not affect functionality since BDR does not support replicating Kudu tables. This change was made to guard against data loss due to how the Hive Mestastore, Imapla, and Kudu interact.

Log Retention

Previously, Cloudera Manager retained logs for BDR Replication jobs indefinitely. Now, Cloudera Manager retains BDR logs for 90 days by default. You can now change the number of days Cloudera Manager retains logs for or disable log retention completely with the **Backup and Disaster Log Retention** property of the HDFS Service.

Faster Incremental Replication using HDFS Snapshot-diff Report

This feature compares two HDFS snapshots to reduce the number of files scanned during the copy-listing phase of replication to only those files that have known changes between runs. This can speed up replication performance dramatically when large number of files are unchanged between replications.

This feature relies on the immutable snapshot feature of HDFS. This feature existed in prior releases of CDH, it is now on by default in 6.1. You can also configure replication jobs to abort on snapshot diff failure when you create or edit a replication schedule. This can happen if files that are in the scope of replication have been added, changed or deleted on the destination cluster, which is generally unsupported by BDR. However, BDR will fall back to an exhaustive comparison of files, and you can use various options for conflict resolution in this case, such as "delete policy".

See the following pages for guidelines on using snapshot diff-based replication: [Hive Guidelines](#) and [HDFS Guidelines](#).

PostgreSQL 10 Support

Added support for PostgreSQL version 10 for databases Cloudera Manager uses to store configuration, monitoring, and reporting data and for managed services that require a database.

Cloudera Express License Enforcement

A Cloudera Express license is only valid when less than 100 hosts are used across an organization.

Note the following:

- Cloudera Manager will not allow you to add hosts to a CDH 6.x cluster if the total number of hosts across all CDH 6.x clusters will exceed 100.
- Cloudera Manager will not allow you to upgrade any cluster to CDH 6.x if the total number of managed CDH 6.x cluster hosts will exceed 100. If an upgrade from Cloudera Manager 6.0 to 6.1 fails due to this limitation, you must downgrade Cloudera Manager to version 6.0, remove some hosts so that the number of hosts is less than 100, then retry the upgrade.



Note: If you downgrade from Cloudera Enterprise to Cloudera Express and the number of managed hosts exceeds 100, Cloudera Manager will disable all cluster management commands except for commands used to stop a cluster. You will not be able to restart or otherwise use clusters while the total number of hosts exceeds 100. Use the Cloudera Manager Admin Console to remove some hosts so that the number of hosts is less than 100.

Affected Versions:Cloudera Manager 6.1 and higher

Diagnostic Bundles

The diagnostic bundle has been improved in the following ways:

- The `dmesg` host command output collected as part of diagnostic bundles now includes formatted timestamps, if the host operating system supports it.
- Diagnostic bundles now capture information about all network interfaces on each host, regardless of name.

HBase CDH 5 to CDH 6 upgrade checks for hbase_prefix_tree_encoding

Added an upgrade check for upgrades from CDH 5 to CDH 6 that checks whether HBase tables are using `PREFIX_TREE_ENCODING` and warns the user.

Cloudera Bug: OPSAPS-44701

HDFS

You can now configure `nfs.export.point` as part of the HDFS configurations for NFSGateway.

Hive

Size of Hive query locks in ZooKeeper

When taking locks on a table, Hive creates a Zookeeper object for each such lock which contains the full query string. This query string is only used to display locks with `SHOW LOCKS EXTENDED` command. It has no impact on the actual locking process.

However, this often created huge memory pressure on the ZooKeeper instance. For example, for a query string of 1MB in size, if the locks are acquired on 10000 partitions of a table, then this requires 10GB of memory on ZooKeeper. To alleviate this pressure, the maximum query length stored in ZooKeper lock object has been limited to 10000 characters by default via `hive.locks.query.string.max.length` property. To reiterate, this does not affect any behavior except for how queries are displayed in the output of the `SHOW LOCKS EXTENDED` command. This configuration value can be increased to a maximum of 1 million, which is the data limit of a znode (1 MB).

Hive Metastore Connection Retries

A new configuration parameter, `hive.metastore.connect.retries`, has been added for HiveServer2 with an increased default value.

Hue

For RedHat7 and compatible platforms, if Hue uses Postgres (including the Cloudera Manager embedded database for proof-of-concept installations), the appropriate version of `psycopg2` will be automatically installed by Cloudera Manager.

Hue Logs

Cloudera Manager can now parse httpd log files, including those used by Hue, meaning they will be included in diagnostic bundles, log search, and visible for browsing in the Cloudera Manager UI.

Impala

New Impala configuration parameters for idle query timeout and idle session timeout

Cloudera Manager now supports configuring the Impala `idle_query_timeout` and `idle_session_timeout` parameters.

New Impala daemon configuration property for JVM heap size

A new Impala Daemon configuration parameter, **Java Heap Size of Impala Daemon in Bytes** has been added to configure the JVM heap size. It defaults to 4 GB, and, like all memory parameters, may require tuning.

Impalad JVM usage plots are now on the Impala Daemon's role status page

Impala Daemon's JVM Heap Usage plots are now available on the Impala Daemon's Status page on the Cloudera Manager Admin Console.

Cloudera Bug: OPSAPS-47832

Impala Metrics

Impala exposes additional metrics about the JVM and GC now. GC metric charts for the Impala Daemon's embedded JVM will now be seen on the Impala Daemon's role status page in the **Cloudera Manager Admin Console**.

Impala Health Checks

Added two new health checks:

- JVM pause time
- Maximum capacity for concurrent client connections for the Impala Daemon. You can configure this health check with the **Impala Daemon Max Client Connections** parameter.

Impala Chart Library

The Impala predefined charts have been updated to include more meaningful metrics and remove rarely used plots.

Impala Resource Pools

Impala resource pools now contain minimum/maximum allowed memory limit (`MEM_LIMIT`) values for queries submitted to a particular pool. This change also adds validations for those attributes. For more information about these attributes, see [IMPALA-7349](#).

Intel's MKL Repository

The **Intel Math Kernel Library (MKL)** parcel is now included in the default parcel repositories starting in Cloudera Manager 6.1. This parcel can accelerate certain machine learning workloads. The parcel is available, but not downloaded or activated on clusters by default. Read more about it here:

<https://software.intel.com/en-us/articles/installing-intel-mkl-cloudera-cdh-parcel>

Kafka

Kafka Data Retention Parameter

The Kafka Broker parameter **Data Retention Hours** (`data.retention.hours`) was removed from the **Cloudera Manager Admin Console**. Use the **Data Retention Time** (`data.retention.ms`) parameter instead.

Kafka Broker Network Threads Parameter

A new configuration property, `num.network.threads` has been added to the Kafka broker configuration parameters. The default value is based on the upstream version.

Kafka Broker Performance Defaults

For CDH 6.1 and higher installations, default values for the following configuration parameters of the Kafka service have been changed based on production recommendations: - `num.replica.fetchers=4` and `num.network.threads=8`

Kafka Metrics

The following metrics have been added for BrokerTopic:

- kafka_fetch_message_conversions_per_sec
- kafka_produce_message_conversions_per_sec
- kafka_replication_bytes_in_per_sec
- kafka_replication_bytes_out_per_sec
- kafka_total_fetch_requests_per_sec
- kafka_total_produce_requests_per_sec

The following metrics have been added for the Controller:

- kafka_auto_leader_balance_rate_and_time_ms
- kafka_controlled_shutdown_rate_and_time_ms
- kafka_controller_change_rate_and_time_ms
- kafka_isr_change_rate_and_time_ms
- kafka_leader_and_isr_response_received_rate_and_time_ms
- kafka_log_dir_change_rate_and_time_ms
- kafka_manual_leader_balance_rate_and_time_ms
- kafka_partition_reassignment_rate_and_time_ms
- kafka_topic_change_rate_and_time_ms
- kafka_topic_deletion_rate_and_time_ms
- kafka_controller_state
- kafka_global_partition_count
- kafka_global_topic_count

The following metrics have been added for the ReplicaManager:

- kafka_failed_isr_updates
- kafka_offline_replica_count
- kafka_under_min_isr_partition_count

The following metrics have been added for the LogCleaner:

- kafka_logcleaner_cleaner_recopy_percent
- kafka_logcleaner_max_buffer_utilization_percent
- kafka_logcleaner_max_clean_time_secs
- kafka_logcleaner_max_dirty_percent
- kafka_logcleaner_time_since_last_run_ms
- kafka_logcleaner_offline_log_directory_count

Kafka Shutdown and Recovery

The graceful stop timeout of the Kafka service has been increased to 120 seconds, and a new configuration property, num.recovery.threads.per.data.dir has been added.

JBOD-related metrics

New metrics have been added that show the number of offline log directories and offline partitions in Kafka

Improved Redaction of Kerberos Credentials

Enhanced the behavior of the Import KDC Account Manager Credentials command. If the command fails, the currently configured redaction policy is now applied to the command's error output. User names and passwords are always redacted from the output.

New cluster Metrics for Cloud storage

The amount of data read and written through S3 and Azure Data Lake storage by MapReduce jobs can now be viewed as cluster metrics. For example: s3a_bytes_read and adl_bytes_written.

Cloudera Bug: OPSAPS-44748

Network Performance Inspector

The **Network Performance Inspector** allows you to examine the latency among the hosts managed by Cloudera Manager. You can use this tool to diagnose latency issues that can significantly affect the performance of workloads such as MapReduce jobs, Spark jobs, and Hive and Impala queries, particularly when using remote storage.

The inspector runs ping commands from each host to all other hosts, and reports the average ping time and packet loss percentage. You can use this information to identify problematic hosts or networking infrastructure issues so that you take corrective action. You can run the inspector on-demand, and it is also available when adding a new cluster. You can also run the inspector using the Cloudera Manager API.

See [Inspecting Network Performance](#)

OpenJDK

OpenJDK is now supported for Cloudera Manager and CDH 6.1 and higher.

For more information, see [Java Requirements](#) on page 26 and [Manually Migrating from Oracle JDK to OpenJDK](#).

Sentry

- A new Sentry configuration has been added to the SENTRY configuration. This enables Sentry OWNER privileges and is disabled by default.
- A new Sentry configuration for OWNER Privileges is added with ALL_WITH_GRANT as the default

YARN Fair Scheduler Properties

Two existing YARN configuration parameters have now been exposed in Cloudera Manager. Fair Scheduler Dynamic Max Assign has been added, which allows the ResourceManager to allocate up to half the available resources on a node during node heartbeat, as long as the Fair Scheduler Assign Multiple Tasks setting is true. The default value is true. Also, the Fair Scheduler Max Assign property has also been added, which sets the number of containers allocated by the ResourceManager with each node heartbeat, as long as Fair Scheduler Assign Multiple Tasks is true and Fair Scheduler Dynamic Max Assign is false. The default value is -1 which is equivalent to unlimited. These changes should not have any effect on YARN behavior as they are just being shown in Cloudera Manager and the default values are unchanged.

System User Group Membership

The Host inspector will now display a warning if various Linux system users (e.g. 'yarn','hdfs','hue','sentry') are not members of a group of the same name, which is required, particularly when Kerberos authentication is enabled. For more information, see [Hadoop Users \(user:group\) and Kerberos Principals](#).

TLS

You can now set the TLS cipher suites for Hadoop with the `ssl.server.exclude.cipher.list` property.

ZooKeeper

Enable Kerberos Authentication and **Enable Server to Server SASL Authentication** settings in ZooKeeper have been linked together since both should be either turned on or off. If either is switched on or off, the other automatically follows.

This change automates steps that were manually required to address CVE-2018-8012 .

Fixed Issues in Cloudera Manager 6.1.0

The following sections describes issue fixed in Cloudera Manager 6.1.0:

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

Technical Service Bulletin 2019-310 (TSB)

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service,

or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper

Releases affected: Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower

Users affected: All

Date/time of detection: June 7, 2018

Severity (Low/Medium/High): 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))

Impact: Remote code execution

CVE: CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings “Enable TLS/SSL for ZooKeeper JMX” and “Enable TLS client authentication for JMX port” on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting “Enable JMX Agent” on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0

Upgrade fails during checkJavaComponent (DbHostHeartbeat.java:177)

Fixed an issue with Java version parsing during Cloudera Manager upgrade.

Cloudera Issue: OPSAPS-47620

CDH 6 upgrade validator fails when no Sentry service is available

Fixed an issue where a CDH 5 cluster with a Keystore Indexer without a Sentry service. When attempting to upgrade to CDH 6 an empty error message displays. Note that when Sentry Policy File is enabled, users must either disable it or add a Sentry service, so that the policy file can be migrated automatically.

Cloudera Issue: OPSAPS-47617

creds.localjceks marked stale with an empty diff

Fixed an issue where `creds.localjceks`, the encrypted keystore used for the Hadoop Credentials provider, might be shown under the list of stale configuration files, but the contents did not actually change. When a role instance is shown as stale and its files include `creds.localjceks`, then this file will also be marked stale. This fix eliminates false reports of staleness.

Cloudera Issue: OPSAPS-47511

Search upgrade reinitialize does not use config for hdfs command

Fixed an issue where on CDH upgrade, Solr index files were not getting deleted from HDFS. This caused Solr to fail to start since files had an old index scheme.

Cloudera Issue: OPSAPS-47502

Export cluster template API returns failure

Fixed the cluster template export failure when the Hue configuration HDFS Web Interface Role (`webhdfs_url`) is using or pointing to the httpfs load balancer rather than to an HDFS role.

Cloudera Issue: OPSAPS-47060

Kafka should use the Garbage first garbage collector by default

Fixed an issue where Kafka broker and MirrorMaker processes did not use the Garbage-First (G1) garbage collector.

Cloudera Issue: OPSAPS-45956*Externally authenticated users cannot view their roles or previous session*

Fixed a display issue where a user could not see their assigned roles and most recent successful login by navigating to <Username> > **My Profile** in the **Cloudera Manager Admin Console**. This issue did not affect functionality.

Cloudera Issue: OPSAPS-46996, OPSAPS-47025*Cloudera Manager not detecting available physical memory correctly*

Fixed an issue with incorrect reporting of used physical memory on host nodes with a significant amount of Shared Memory in use. Cloudera Manager now takes usage of Shared Memory into account when reporting the physical memory used on a host node.

Cloudera Issue: OPSAPS-47396*HDFS_CLIENT_CONFIG_JAVA_OPTS has hbase in the template name*

Changed the API name to fix the wrong name in the parameter.

Old Name	New Name
hbase_client_java_opts	hdfs_client_java_opts

This parameter configures the Client Java Configuration Options found under the HDFS Gateway role configuration. Any API scripts or cluster templates referencing these old names need to be updated to use the new names.

Cloudera Issue: OPSAPS-24569*Fix typos in a "detecton_window" API names*

Changed the API names to fix typos in the following parameters:

Old Name	New Name
hbase_active_master_detecton_window	hbase_active_master_detection_window
hdfs_active_namenode_detecton_window	hdfs_active_namenode_detection_window
mapreduce_active_jobtracker_detecton_window	mapreduce_active_jobtracker_detection_window
yarn_active_resourcemanager_detecton_window	yarn_active_resourcemanager_detection_window

These parameters tune the behavior of health test checking. The affected entities are: HBase Master, HDFS NameNode, MapReduce JobTracker, YARN ResourceManager. Any API scripts or cluster templates referencing these old names need to be updated to use the new names.

Cloudera Issue: OPSAPS-39223*CDH 6 Spark CSD does not support Auto-TLS*

Fixed an issue where Auto-TLS settings were not applied to the Spark service when Auto-TLS was enabled.

Cloudera Issue: OPSAPS-47925*Impala shell does not display the port number*

Fixed an issue where the Impala shell command in the Cloudera Manager Admin Console was missing the port number required to connect to the Impala shell.

Cloudera Issue: OPSAPS-47589*Enable ZooKeeper fix for CVE-2018-8012*

Enable Kerberos Authentication and **Enable Server to Server SASL Authentication** settings in ZooKeeper have been linked together since both should be either turned on or off. If either is switched on or off, the other automatically follows.

This change automates steps that address CVE-2018-8012. Previously, the solution required manual steps.

Cloudera Issue: OPSAPS-46628

Combine audit entries

Fixed an issue that occurs when the API is accessed at a rapid rate. This can cause the Audits database table to grow rapidly, negatively impacting Cloudera Manager performance.

Cloudera Manager logs events in the Audits database table when the API is accessed either from the Admin Console or from any other client. You can now configure a time period during which similar events are combined into one log entry. For more information, see [Audit Events](#) on page 148.

Cloudera Issue: OPSAPS-46898

CMF_SERVER_ARGS if given a configuration file results in staleness for Cloudera Manager

Fixed an issue where applying a configuration change with CMF_SERVER_ARGS arguments (using the /etc/default/cloudera-scm-server configuration file) led to a staleness warning after a Cloudera Manager server restart.

Cloudera Issue: OPSAPS-47240

Kudu package missing from libs/common/src/main/java/com/cloudera/cmf/CDHResources

Fixes an issue where Cloudera Manager did not install Kudu packages when CDH was installed with packages instead of parcels.

Cloudera Issue: OPSAPS-45692

Restart warnings are incorrect after starting role with outdated configuration

Fixed an issue where some roles that required restarts were not correctly identified after starting a role marked as **Started with Outdated Configuration**.

Cloudera Issue: OPSAPS-45237

Type in HiveServer2 load balancer API name

Fixed typos in the following parameter. This change affects Hive services when Hive Server 2 is configured for High Availability.

Table 28: API Names

Old Name	New Name
hiveserver2_load_balancer	hiveserver2_load_balancer

Any API scripts or cluster templates referencing these old names will need to be updated to use the new names.

Cloudera Issue: OPSAPS-33266

Traceback seen in ImpalaRoleDiagnosticsCollection and HBaseRoleDiagnosticsCollectionprocess

Fixed an issue that caused an exception to occur in the Cloudera Manager Agent during diagnostic bundle collection if the process had exited previously.

Cloudera Issue: OPSAPS-47354

Fix kafka_network_processor_avg_idle metric

Fixed an issue where the **kafka_network_processor_avg_idle** metric shows **NO DATA**.

Cloudera Issue: OPSAPS-45816

Sentry fails on first run, due to a pending command

When starting Sentry for the first time after the service was added, the "Creating Sentry Database Tables" step in the Start Service command may fail with the error: "There is already a pending command on this entity". This issue has been fixed and starting Sentry for the first time after the service was added no longer fails due to a pending command.

Cloudera Issue: OPSAPS-48426*HDFS Canary with HA nameservice in a non-federated cluster fails*

The HDFS canary no longer erroneously reports UNKNOWN health status.

Cloudera Issue: OPSAPS-48337*Server and Daemon RPM installation scripts do not work well with Puppet installs*

If you have installed the JDK at a non-standard location, set the `JAVA_HOME` environment variable before installing Cloudera Manager. If you cannot set `JAVA_HOME` in your environment, create an empty file with the path `/etc/cloudera-pre-install/CLOUDERA_SKIP_JAVA_INSTALL_CHECK` to skip any Java checks during package installation of Cloudera Manager Server and Daemon packages.

Cloudera Issue: OPSAPS-47908*Cannot stop Kafka broker*

Fixed an issue where the Kafka Broker could not be stopped if **Automatically Restart Process** is enabled. Because of a misconfiguration in process monitoring, the Cloudera Manager Agent would also restart the process when a legitimate stop was requested. Additionally, without automatic restarts, once the process was stopped, the health check for **Unexpected Exits** would eventually show the process in bad health. Note that this bug affected all CSD-based services where a graceful stop behavior was enabled at the role-level.

Cloudera Issue: OPSAPS-45029*Database connection error.*

Fixed a database connection leak issue that caused the following error: `java.lang.IllegalStateException: currentCmfEntityManager already in transaction.`

Cloudera Issue: OPSAPS-45829*CSD role creation logic fixed for second instance of service*

Fixes the automatic role creation logic when adding a second instance of a service. Adding a second instance of a service could result in extra roles being generated for the first instance of a service.

Cloudera Issue: OPSAPS-47766*Agent should download key bundles when behind proxy (plain HTTP)*

Even if a proxy server was configured for Cloudera Manager, it was not used to download the package signing key during host installs, leading to installation failures. This has been fixed so that downloading the package signing key will use the configured proxy, but only if it is a plain HTTP proxy. Proxies requiring authentication or HTTPS are not currently supported. As a workaround, you can mirror the package repository locally to avoid needing a proxy.

Cloudera Issue: OPSAPS-47830*Incompatible Changes in Cloudera Manager 6.1*

See below for incompatible changes in Cloudera Manager 6.1.0:

API Name Changes

`HDFS_CLIENT_CONFIG_JAVA_OPTS` has `hbase` in the template name

Changed the API name to fix the wrong name in the parameter.

Old Name	New Name
<code>hbase_client_java_opts</code>	<code>hdfs_client_java_opts</code>

This parameter configures the Client Java Configuration Options found under the HDFS Gateway role configuration. Any API scripts or cluster templates referencing these old names need to be updated to use the new names.

Cloudera Issue: OPSAPS-24569

Fix typos in a "detecton_window" API names

Changed the API names to fix typos in the following parameters:

Old Name	New Name
hbase_active_master_detecton_window	hbase_active_master_detection_window
hdfs_active_namenode_detecton_window	hdfs_active_namenode_detection_window
mapreduce_active_jobtracker_detecton_window	mapreduce_active_jobtracker_detection_window
yarn_active_resourcemanager_detecton_window	yarn_active_resourcemanager_detection_window

These parameters tune the behavior of health test checking. The affected entities are: HBase Master, HDFS NameNode, MapReduce JobTracker, YARN ResourceManager. Any API scripts or cluster templates referencing these old names need to be updated to use the new names.

Cloudera Issue: OPSAPS-39223

Roles-related Cloudera Manager API endpoints are now accessed using RolesResource

For API documentation and the new, Swagger-based API client, roles can be accessed from Roles Resource instead of Services Resource. This does not change the roles endpoint and does not impact those accessing the Cloudera Manager API endpoints directly using tools like curl.

Cloudera Bug: OPSAPS-47787

Typo in HiveServer2 load balancer API name

Fixed typos in the following parameter. This change affects Hive services when Hive Server 2 is configured for High Availability.

Table 29: API Names

Old Name	New Name
hiveserver2_load_balancer	hiveserver2_load_balancer

Any API scripts or cluster templates referencing these old names will need to be updated to use the new names.

Cloudera Issue: OPSAPS-33266

Known Issues and Limitations in Cloudera Manager 6.1.0

The following sections describe known issues and limitations for Cloudera Manager 6.1.0:

HBase Failure on Upgrade to Cloudera Manager 6.1

After upgrading from Cloudera Manager 5.14 (and??) to Cloudera Manager 6.1 and restarting the Hive service, running an HBase query cause the following error in beeline:

```
Error: java.io.IOException: org.apache.hadoop.hbase.client.RetryExhaustedException:  
Failed after attempts=36, exceptions:
```

Workaround:

1. In the Cloudera Manager Admin console, go to the Hive service.
2. Select the **Configuration** tab.
3. Search for the following property: HiveServer2 Environment Advanced Configuration Snippet (Safety Valve).

4. Add the following property and value:

```
HADOOP_CLASSPATH=/etc/hbase/conf
```

5. Click **Save Changes**.

6. Restart HiveServer2:

- a. Go to the Hive Service.
- b. Click the **Instances** tab.
- c. Click the **HiveServer2** link in the table.
- d. Click **Actions > Restart this HiveServer2**.

Cloudera Issue: OPSAPS 49330

Cloudera Manager Server and Agents encounter TLS issues when the Agent config.ini file is misconfigured

The Cloudera Manager Server and Agents incorrectly thinks that TLS is enabled for the Agent if you set `use_tls=1` in the Agent `config.ini` but do not provide a key, certificate, or truststore. This causes several issues, including log messages that indicate TLS is enabled when it is not, missing files from diagnostic bundles, and unavailable logs.

Workaround: Ensure that you specify a key, certificate, and truststore when you configure `use_tls=1` in the Agent `config.ini` file.

Affected Versions: Cloudera Manager 6.1.x

Cloudera Issue: OPSAPS-48898, OPSAPS-48897

The restart Cloudera Manager Agent command does not restart the Agent Listener

When you restart the Cloudera Manager Agent, the Agent Listener (the `status_server` process) does not restart. This can cause issues if you make changes to TLS for Agents after you have installed Cloudera Manager since the Agent Listener needs to be restarted for TLS changes to take effect.

Workaround: Run the following command on every Agent host with `supervisorctl`:

```
/opt/cloudera/cm-agent/bin/supervisorctl -c
/var/run/cloudera-scm-agent/supervisor/supervisord.conf restart status_server
```

This command requires root access to the host. `supervisorctl` is owned by the `cloudera-scm` user, but `supervisord.conf` is owned by `root`.

Affected Versions: Cloudera Manager 6.0.x, 6.1.x

Cloudera Issue: OPSAPS-48886

TSB-359 Backup and Disaster Recovery (BDR) HDFS and Hive Replications will fail on clusters running Cloudera Manager 6.1.0

Backup and Disaster Recovery (BDR) HDFS and Hive Replications will fail when replicating from secured (Kerberized) source clusters to destination clusters that have been upgraded to Cloudera Manager 6.1.0.

This also affects new installations of Cloudera Manager 6.1.0 on the destination cluster if an admin restarts the Cloudera Manager service.

Products affected: Cloudera Manager Backup and Disaster Recovery in a secure (Kerberized) environment

Releases affected: Cloudera Manager 6.1.0 (when used as the destination cluster of HDFS and/or Hive replication)

Users affected: Customers using HDFS or Hive Replication

Severity (Low/Medium/High): High

Root Cause and Impact:

In HDFS and Hive Replication, Cloudera Manager first runs a process on the destination cluster to verify if the replication is possible. Due to a bug, the source cluster is treated as an insecure (non-kerberized) cluster. As a result, replication fails.

You will see the exception `javax.security.sasl.SaslException: GSS initiate failed [Caused by GSSEException: No valid credentials provided (Mechanism level: Fail to create credential. (63) - No service creds)]` in the process stderr logs.

Immediate action required: If you use BDR, do not upgrade a destination cluster to Cloudera Manager 6.1.0. Upgrade to Cloudera Manager 6.1.1 or higher when it becomes available.

If you have already upgraded your destination cluster to Cloudera Manager to 6.1.0, use the following workaround:

1. For an existing HDFS or Hive replication schedule, select **Actions > Edit Configuration**.
2. Save the schedule.

Please note that you will need to edit only one schedule even if you have multiple schedules.

Note: This workaround is not persistent. That is, if you restart the Cloudera Manager service, you must repeat the above workaround.

Cloudera Issue: OPSAPS-48865

Fixed in Cloudera Manager 6.1.1

BDR Job Mapper shows the following warning: AuthenticationException: GSSEException: No valid credentials provided

A BDR Job Mapper might succeed with the following stack trace in the log message:

```
2018-12-05 13:57:03,475 WARN [main]
org.apache.hadoop.crypto.key.kms.LoadBalancingKMSClientProvider: KMS provider at
[https://src-3.example.com:16000/kms/v1/] threw an IOException:
java.io.IOException:
org.apache.hadoop.security.authentication.client.AuthenticationException: GSSEException:
  No valid credentials provided (Mechanism level: Fail to create credential. (63) - No
service creds)
at
org.apache.hadoop.crypto.key.kms.KMSClientProvider.createConnection(KMSClientProvider.java:492)
at
org.apache.hadoop.crypto.key.kms.KMSClientProvider.decryptEncryptedKey(KMSClientProvider.java:793)
```

The warning appears if you try to replicate from an encrypted source cluster that has multiple KMS instances.

Workaround: You can safely ignore this message because the client succeeds upon fail over.

Affected Versions: CDH 6.1.0

Cloudera Issue: CDH-76053

Diff Format changed for Configuration History page

Color formatting for the diff display omits the red and green colors that indicate what was removed and added.

Affected Versions: Cloudera Manager 6.1.0

Cloudera Issue: OPSAPS-48544

Backup and Disaster Recovery (BDR) performance regression after upgrading to CDH 6.0.0

Hive replication with BDR experiences a performance regression when comparing CDH 6.0.0 and CDH 5.14.4. The slowdown occurs during the import step. For example, the performance regression may only be 10% for 4 million partitions. As the number of partitions goes down though, the performance impact becomes more visible. For example, 100,000 partitions may experience a 20% performance regression.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0; CDH 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47520

Cloudera Manager allows more than a single space in YARN Admin ACLs

When adding a YARN Admin ACL in Cloudera Manager, you are allowed to enter multiple spaces in the entry. The space is the separator between the user and group lists, and only a single space should be allowed in the entry. All entries that appear after a second single space in a YARN Admin ACL will be ignored.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47688

Integer data types map to Float in Swagger API client

Integer data types show up as floating point numbers when using the Cloudera Manager API Python client.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-45689

Package Installation of CDH Fails

When you install CDH with packages from a custom repository, ensure that the version of CDH you select for **Select the version of CDH** matches the version of CDH for the custom repository. Selecting the CDH version and specifying a custom repository are done during the **Select Repository** stage of installation.

If the versions do not match, installation fails.

Affected Versions: Cloudera Manager 6.x

Fixed Versions: N/A

Apache Issue: N/A

Cloudera Issue: OPSAPS-45703

Backup and Disaster Recovery replication to/from Cloudera Manager 6 clusters require Cloudera Manager 5.14.0 or higher

You can only use BDR to replicate to/from clusters managed by Cloudera Manager 6 with Cloudera Manager 5.14.0 or higher.

Affected versions: Cloudera Manager 6.x

Cloudera Issue: OPSAPS-42207

Cloudera Manager 6.0.x Release Notes

To view release notes for specific Cloudera Manager 6.0.x releases, see the following:

[Cloudera Manager 6.0.1 Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for Cloudera Manager 6.0.1:

[New Features in Cloudera Manager 6.0.1](#)

The following sections describe new and changed features for Cloudera Manager 6.0.1:

[New Features in Cloudera Manager 6.0.1](#)

Diagnostic Bundles

- Diagnostic bundles now contain the DB_READ_LATENCY metric.

Service Monitor

- Improved the performance of the Service Monitor by reducing memory consumption.

PostgreSQL 10 Support

- Added support for PostgreSQL version 10 for databases Cloudera Manager uses to store configuration, monitoring, and reporting data and for managed services that require a database.

Upgrade

- Added an upgrade check from CDH 5 to CDH 6 to check if HBase tables are using `PREFIX_TREE_ENCODING`.

Fixed Issues in Cloudera Manager 6.0.1

The following sections describes issue fixed in Cloudera Manager 6.0.1 releases:

X-Frame-Options not set in latest C6

Fixed an issue where a page in the Cloudera Manager Admin Console was not sending an X-Frame-Options header while other pages in the Admin Console did. The header is now sent.

Cloudera Issue: OPSAPS-47252

Auto-TLS support for commands like Oozie's Upload Sharelib

Fixed an issue that occurred when the Oozie Upload Sharelib command would fail due to missing TLS configuration if Auto-TLS is enabled.

Cloudera Issue: OPSAPS-47084

CDH upgrade fails with checkJavaComponent(DbHostHeartbeat.java:177)

Fixed a Java version parsing issue during Cloudera Manager upgrade.

Cloudera Issue: OPSAPS-47620

CDH 6.0.0 upgrade validator fails when no Sentry service is available

Fixed an issue that occurs when upgrading CDH 5 clusters with a Keystore Indexer but no Sentry service. A validator shows an empty error message in a popup when trying to upgrade to CDH 6.0.0. Note that when Sentry Policy File is enabled, users must either disable it or add a Sentry service in order for the policy file to be migrated automatically.

Cloudera Issue: OPSAPS-47617

Agent install fails with "Installing hue-plugins package"

Fixed an issue where a CDH 6.0.0 installation with packages fails.

Cloudera Issue: OPSAPS-47105

CDH Upgrade fails to delete Solr data from HDFS

Fixed an issue where where Solr index files were not getting deleted from HDFS during the upgrade process. This causes Solr to fail to start since files have an old index scheme.

Cloudera Issue: OPSAPS-47502

Cloudera Manager wizard shows CDH 5.15.0 as a package choice

Fixed an issue where CDH 5.15.0 was shown as a valid package choice for Cloudera Manager 6.0.0. For more information about valid CDH versions for Cloudera Manager 6.0.0, see [Upgrade paths](#).

Cloudera Issue: OPSAPS-47200

Host inspector shows psycopg2 version error even after psycopg2-2.6.2 is installed

Fixed an issue where hosts that run Redhat, Centos, or OEL showed incorrect host inspector results for the Python version check and for psycopg2 version check.

Cloudera Issue: OPSAPS-47217

Fix display vcore and memory values when they are not integers

Fixed the display of vcore and memory values to support values other than integers.

Cloudera Issue: OPSAPS-47271

Upgrade Service Inspector shows it passed, but the solr check failed

Fixed issue where the "Service Inspector" on the upgrade page was showing Solr in good health even though Solr had failures.

Cloudera Issue: OPSAPS-46958**Known Issues and Limitations in Cloudera Manager 6.0.1**

The following sections describe known issues and limitations for Cloudera Manager 6.0.1:

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service, or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper**Releases affected:** Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower**Users affected:** All**Date/time of detection:** June 7, 2018**Severity (Low/Medium/High):** 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))**Impact:** Remote code execution**CVE:** CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings “Enable TLS/SSL for ZooKeeper JMX” and “Enable TLS client authentication for JMX port” on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting “Enable JMX Agent” on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0***The restart Cloudera Manager Agent command does not restart the Agent Listener***

When you restart the Cloudera Manager Agent, the Agent Listener (the `status_server` process) does not restart. This can cause issues if you make changes to TLS for Agents after you have installed Cloudera Manager since the Agent Listener needs to be restarted for TLS changes to take effect.

Workaround: Run the following command on every Agent host with `supervisorctl`:

```
/opt/cloudera/cm-agent/bin/supervisorctl -c /var/run/cloudera-scm-agent/supervisor/supervisord.conf restart status_server
```

This command requires root access to the host. `supervisorctl` is owned by the `cloudera-scm` user, but `supervisord.conf` is owned by `root`.

Affected Versions: Cloudera Manager 6.0.x, 6.1.x**Cloudera Issue:** OPSAPS-48886

Backup and Disaster Recovery (BDR) performance regression after upgrading to CDH 6.0.0

Hive replication with BDR experiences a performance regression when comparing CDH 6.0.0 and CDH 5.14.4. The slowdown occurs during the import step. For example, the performance regression may only be 10% for 4 million partitions. As the number of partitions goes down though, the performance impact becomes more visible. For example, 100,000 partitions may experience a 20% performance regression.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0; CDH 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47520

Cloudera Manager allows more than a single space in YARN Admin ACLs

When adding a YARN Admin ACL in Cloudera Manager, you are allowed to enter multiple spaces in the entry. The space is the separator between the user and group lists, and only a single space should be allowed in the entry. All entries that appear after a second single space in a YARN Admin ACL will be ignored.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47688

Integer data types map to Float in Swagger API client

Integer data types show up as floating point numbers when using the Cloudera Manager API Python client.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-45689

Apache Accumulo is not supported with Cloudera Manager

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1

Fixed Versions: Cloudera Manager 6.1.0

Cloudera Issue: OPSAPS-42807, OPSAPS-42814

Cloudera Data Science Workbench is not supported with Cloudera Manager 6.0.x

Cloudera Data Science Workbench is not supported with Cloudera Manager 6.0.x. Cloudera Data Science Workbench 1.5.0 (and higher) is supported with Cloudera Manager 6.1.x (and higher).

Affected Versions: Cloudera Manager 6.0.x

Cloudera Issue: DSE-2769

Externally authenticated users cannot view their roles or previous session

Usually, a user can see their assigned roles and most recent successful login by navigating to <Username> > **My Profile** in the Cloudera Manager Admin Console. The fields appear blank for users who use an external authentication method, such as SAML.

This issue is only a display issue and does not affect any functionality. The user can perform any tasks available to their assigned roles.

Workaround: User accounts with roles that can view the **Roles** page, such as a Full Administrator, can view the roles assigned to all Cloudera Manager user accounts and their session information.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1

Fixed Versions: Cloudera Manager 6.1.0

Cloudera Issue: OPSAPS-46996, OPSAPS-47025

Package Installation of CDH Fails

When you install CDH with packages from a custom repository, ensure that the version of CDH you select for **Select the version of CDH** matches the version of CDH for the custom repository. Selecting the CDH version and specifying a custom repository are done during the **Select Repository** stage of installation.

If the versions do not match, installation fails.

Affected Versions: Cloudera Manager 6.x

Fixed Versions: N/A

Apache Issue: N/A

Cloudera Issue: OPSAPS-45703

Backup and Disaster Recovery replication to/from Cloudera Manager 6 clusters require Cloudera Manager 5.14.0 or higher

You can only use BDR to replicate to/from clusters managed by Cloudera Manager 6 with Cloudera Manager 5.14.0 or higher.

Affected versions: Cloudera Manager 6.x

Cloudera Issue: OPSAPS-42207

[Cloudera Manager 6.0.0 Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for Cloudera Manager 6.0.0:

[New Features in Cloudera Manager 6.0.0](#)

The following sections describe new and changed features for Cloudera Manager 6.0.0:

[New Features in Cloudera Manager 6.0.0](#)

API

You can now access the Cloudera Manager Swagger API user interface through the Cloudera Manager Admin Console. Navigate to **Support > API Explorer** to open Swagger.

Cloudera Manager 6.0 introduces new Python and Java API clients based on the [Swagger](#) API. These new API clients support all Cloudera Manager API versions.

Compatibility with Older Versions

The [Older Python client](#) [Older Java client](#) are still supported when the API version is lower than 30. Therefore, older Python and Java API clients can still be used with Cloudera Manager version 6.0 and higher as long as API version 19 or lower is used.

For example, a customer can use old Cloudera Manager API client version 5.14 with Cloudera Manager version 6.0, which by default invokes API version 19. If a customer wants to use new features that were introduced in Cloudera Manager 6.0, (API version 30) then customer must use the new API clients.

Older Python and Java clients and new Swagger-based Python and Java clients can co-exist in an application to allow for incremental transition to new the Swagger-based Python and Java clients.

Auto-TLS

Auto-TLS simplifies configuring TLS for clusters managed by Cloudera Manager. Cloudera Manager can now add hosts with TLS certificates automatically generated. Additionally, when you add new CDH services, Cloudera Manager populates the TLS configuration for the service.

You can use self-signed certificates created by Cloudera Manager's internal certificate authority, or you can use certificates you already have from a trusted public CA or your own internal CA.

An Enterprise or Trial license is required to enable Auto-TLS.

For more information, see [Configuring TLS Encryption for Cloudera Manager and CDH Using Auto-TLS](#).

Cluster-specific User Roles

You can now assign privileges for specific clusters to the following user roles: Cluster Administrator, Operator, Limited Operator, Configurator, and Read-Only.

For example, the user account `lucy` has the Cluster Administrator role with privileges for a cluster named `Cluster1`. `lucy` can only perform the Cluster Administrator actions on `Cluster1`. She cannot perform actions on any other cluster managed by Cloudera Manager.

For more information, see [Cloudera Manager User Roles](#).

Common Service Descriptors (CSD)

- CSDs can now specify more than one repository for parcels. A list of URLs can be specified via the new property `additionalRepoUrls` of the parcel descriptor.
- CSD authors can declare a set of invalid values for numeric parameters for the following types: long, double, port, and memory.

HBase

New command to create the HBase HDFS WAL directory in a separate location. By default the HBase HDFS WAL directory is created in `/hbase/WALs`. New command allows the HBase HDFS WAL directory to be created a different location. After setting the separate WAL directory, the Hbase service needs to be restarted.

Cell-Level ACL Checks

Newly deployed Cloudera Manager managed clusters will now optimize for HBase Cell-Level ACL checks by default with the `hbase.security.access.early_out` property. The property controls whether certain checks can be skipped for performance reasons

Impala

Cloudera Manager now collects more metrics that are helpful to Impala administrators for monitoring Catalog size and Impala Daemon health. Metrics from Impala Daemons help track the amount of memory used by the Java Virtual Machine (JVM) embedded in the Impala Daemon process. Use the metrics to understand memory consumption, particularly the memory consumption of the Catalog cache stored in coordinator Impala Daemons. The new metrics are:

- `impala_jvm_heap_committed_usage_byte`
- `impala_jvm_heap_current_usage_bytes`
- `impala_jvm_heap_init_usage_bytes`
- `impala_jvm_heap_max_usage_bytes`

Kafka

- Added `num.network.threads` as a configuration setting for Kafka brokers.
- Kafka's broker heap size is now configurable in the Cloudera Manager Admin Console wizard for Kafka.

Security

Cloudera Manager now passes SSL keystore credentials to SOLR through the Hadoop Credential Store.

Scalability

- The performance for the cluster status table loading time on the **Cloudera Manager Admin Console** home page for large clusters has been improved.
- The cluster restart command/operation can be retried now. This is especially helpful in a large cluster where the user has an opportunity to fix the cluster restart failure and retry the failed command.
- When Cloudera Manager sees multiple agents with the same hostname or IP appearing, it will ask users to adjust the UUID on the host and remove the agent that has no roles running.

Upgrade

Cloudera Manager has the following upgrade improvements:

- New upgrade wizard and documentation.

The [Cloudera Enterprise Upgrade Guide](#) allows you to create a customized version of the guide that only includes the steps required for your upgrade. You can use a form at the top of pages in the guide to select your Cloudera Enterprise versions, operating system versions, databases, and other information about your upgrade. The information you enter is retained on each page in the guide.

- You can now run the Host and Service inspectors up to two days prior to an upgrade. This allows a long running inspection to complete prior to starting the upgrade.
- Rolling upgrades for CDH have been improved. YARN jobs running MapReduce2 are now configured to read MapReduce JARs from HDFS instead of from local disk. This makes jobs more robust during rolling upgrade when the local binaries are modified while a job is executing. Clusters created in or upgrading to CDH 6.0 will use this new behavior.

Changes in Cloudera Manager 6.0.0

Agents

Because of changes to Cloudera Manager, the commands used for a hard stop and hard restart have changed.

For more information, see [Starting, Stopping, and Restarting Cloudera Manager Agents](#).

API names

The following API names have changed to fix typos:

- `hiveserver2_load_balancer` has been changed to `hiveserver2_load_balancer`
- `hbase_client_java_opts` has been changed to `hdfs_client_java_opts`
- `hbase_active_master_detecton_window` has been changed to `hbase_active_master_detection_window`
- `hdfs_active_namenode_detecton_window` has been changed to `hdfs_active_namenode_detection_window`
- `mapreduce_active_jobtracker_detecton_window` has been changed to `mapreduce_active_jobtracker_detection_window`
- `yarn_active_resourcemanager_detecton_window` has been changed to `yarn_active_resourcemanager_detection_window`

The `hiveserver2_load_balancer` change affects Hive services when HiveServer 2 is configured for High Availability.

The `hdfs_client_java_opts` parameter configures the Client Java Configuration Options, found under the HDFS Gateway role configuration.

The other parameters tune the behavior of health test checking for the HBase Master, HDFS NameNode, MapReduce JobTracker, and YARN ResourceManager respectively.

Any API scripts or cluster templates referencing the old names will need to be updated to use the new names.

Cloudera Issue: OPSAPS-33266, OPSAPS-39223, and OPSAPS-24569

Client Configurations

Downloading the client configuration for a service now requires a user account that meets the following requirements: the user account must be assigned a user role that has permission to perform the action and has privileges for the specific cluster because of the new cluster-specific user role feature.

External Authentication

Previously, Cloudera Manager, by default, mapped specific values from an external authentication method to Cloudera Manager user roles.

For example, if the authentication method for Cloudera Manager is a SAML Script, Cloudera Manager automatically mapped exit codes 0 to 11 to the user roles that Cloudera Manager ships with. If you upgrade to Cloudera Manager 6, these mappings are preserved. You can continue using these default mappings, create additional ones, or map different values.



Important: If you create external authentication entities, such as a new LDAP group, and do not map it to a Cloudera Manager user role, the users in that group will default to no access. Users in the group cannot perform any actions on the cluster.

If you perform a fresh installation of Cloudera Manager, values must be mapped to user roles manually in Cloudera Manager.

Additionally, LDAP Group, SAML Attribute, and External Program to user role mappings are no longer done through the **Administration > Settings** page. Instead, like the exit codes for SAML Scripts, they are configured on the new **Administration > Users & Roles (previously Users) > <Authentication Method>** page.

For more information, see [Mapping External Authentication to a Role](#).

HBase

- Updated property values:

Updated the default values for the following properties to match the upstream defaults:

- hbase.snapshot.region.timeout
- hbase.snapshot.master.timeout.millis
- hbase.client.retries.number (all roles)
- hbase.hstore.blockingStoreFiles (regionserver)

Removed the following values:

- hbase.snapshot.master.timeoutMillis
- hbase.fs.tmp.dir (all roles)
- hbase.bucketcache.combinedcache.enabled (regionserver)
- hbase.bulkload.staging.dir (regionserver)
- hbase.regionserver.hlog.blocksize (regionserver)

- The HBase Thrift Server now turns on Framed Transport and Compact Protocols by default out of the box for safety reasons. This may require code changes to any custom client-side thrift programs that connect to the HBase Thrift Server to continue working.

Kafka

- **Data Retention Hours property** - The Kafka Broker parameter **Data Retention Hours** (`data.retention.hours`) was removed from the Cloudera Manager Admin console. Use **Data Retention Time** (`data.retention.ms`) parameter instead.
- **Default Kafka minimum heap** - The default minimum allowed heap for Kafka has been increased to 256 MB. The recommended minimum heap is 512 MB. If your Kafka broker heap size is set to a value less than 256 MB, increase it to 256 MB or higher.

Menu Names

The following list describes changes to menu names:

- **Users**

The **Users** page that is accessed from the **Administration** menu has been renamed to **Users & Roles**.

New Users

If you create a user and do not assign a role to it, the user defaults to no access. The user cannot perform any actions on the cluster.

Reporting

The default HDFS block count reporting threshold for Cloudera Manager has been changed to 1000000 from 500000. When you upgrade, the configuration will be updated to the new default if you are using the default.

User Roles

The Dashboard and Auditor user roles can now view the Solr Collection Statistics and the HBase Table Statistics pages.

Fixed Issues in Cloudera Manager 6.0.0

The following sections describes issue fixed in Cloudera Manager 6.0.0 releases:

Open Redirect and XSS in Cloudera Manager

Technical Service Bulletin 2018-321 (TSB)

One type of page in Cloudera Manager uses a `returnUrl` parameter to redirect the user to another page in Cloudera Manager once a wizard is completed. The validity of this parameter was not checked. As a result, the user could be automatically redirected to an attacker's external site or perform a malicious JavaScript function that results in cross-site scripting (XSS).

With this fix, Cloudera Manager no longer allows any value in the `returnUrl` parameter with patterns such as `http://`, `https://`, `//`, or `javascript`. The only exceptions to this rule are the SAML login/logout URLs, since they are explicitly configured and are not passed via the `returnUrl` parameter.

Products affected: Cloudera Manager

Releases affected:

- 5.15.0 and all earlier releases

Users affected: The following Cloudera Manager roles: "cluster administrator", "full administrators", and "configurators".

Date/time of detection: June 20, 2018

Detected by: Mohit Rawat & Ekta Mittal

Severity (Low/Medium/High): 8.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:R/S:U/C:H/I:H/A:H](#))

Impact: [Open redirects](#) can silently redirect a victim to an attacker's site. XSS vulnerabilities can be used to steal credentials or to perform arbitrary actions as the targeted user.

CVE: CVE-2018-15913

Immediate action required: Upgrade to Cloudera Manager 5.15.1 or higher

Addressed in release/refresh/patch:

- Cloudera Manager 5.15.1 and higher
- Cloudera Manager 6.0.0

Hard Restart of Cloudera Manager Agents May Cause Subsequent Service Errors

If a "hard restart" or "hard stop" operation is performed on a Cloudera Manager Agent, the restarted agent will erroneously restart roles that existed prior to the restart and, subsequently, 60 days later, these roles may experience errors or be killed.

Affected Versions: All versions of Cloudera Manager 5.x

Cloudera Issue: OPSAPS-43550, TSB-308

Knowledge base: For the latest update on this issue, see the corresponding Knowledge article: TSB 2018-308: Hard Restart of Cloudera Manager Agents May Cause Subsequent Service Errors

Logging issue slows down Backup and Disaster Recovery Hive and HDFS Replication jobs

Fixed the issue described in TSB-289. For more information, see the TSB.

Cloudera Issue: OPSAPS-44160

Cloudera Manager upgrade workflow incorrectly requires deploying some optional management roles

Fixed the issue described in TSB-290 where you could not proceed through the upgrade process without adding certain optional management roles. For more information, see TSB-290.

Cloudera Issue: OPSAPS-44629

Microsoft Azure Credentials in Log Files

Fixed an issue where Microsoft Azure credentials might appear in Hive audit logs.

Cloudera Issue: CDH-56241

Non-production installation of Cloudera Manager on SLES 12 does not work

Fixed an issue where the non-production installation of Cloudera Manager did not work on SLES 12.

Impala and Kudu logs missing from diagnostic bundle

Fixed an issue where Impala and Kudu logs were missing from the diagnostic bundle if their log directories have broken symlinks.

Cloudera Issue: OPSAPS-41194

Services die due to HDFS taking too long to start

Fixed an issue where HDFS takes a long time to come up after a restart, causing some dependent services to fail to start.

Cloudera Issue: CDH-54889

Instances and Hosts page refresh when a command dialog is closed

Fixed an issue where the **Instances** and **All Hosts** pages reloads reload when a command finishes.

Cloudera Issue: OPSAPS-45761

Spark cross-realm authentication fails

Spark now correctly respects auth_to_local name rules for HDFS services with cross-realm trust configured.

Cloudera Issue: OPSAPS-46103

Error "Mismatched input PATTERN expecting EOF" the detailUsage page for the Resource Manager

Fixed the issue where a user sees an error message about Mismatched input PATTERN.

Cloudera Issue: OPSAPS-42437

Upgrading a license finishes on the wrong page

The Enable Trial workflow previously ended up on the upgrade page. Now it goes to the Home page upon completion.

Cloudera Issue: OPSAPS-45444

Open Redirect in Cloudera Manager Add Service

Fixed an issue where Cloudera Manager redirected to arbitrary URLs upon the completion of a workflow. Cloudera Manager now limits it to paths on the same host/port

Cloudera Issue: OPSAPS-46681

Kafka broker and MirrorMaker should only listen on the loopback interface for JMX connections

Kafka broker and MirrorMaker processes now listen on only the loopback interface for JMX connections. The fix causes Kafka brokers and MirrorMaker to be marked as stale after upgrading to Cloudera Manager 6.0.0 or later.

Perform a rolling restart of Kafka brokers and MirrorMaker.

Cloudera Issue: OPSAPS-46633

Remove the IMPALA_ASSIGNMENT_LOCALITY Impala check

This check was removed.

Cloudera Issue: OPSAPS-46807

Inconsistent handling of case sensitivity for cluster names in URLs

Fixed an issue where cases sensitivity for cluster names was not handled consistently with the API, mainly related to the cluster name. For example, the end point "/api/v6/clusters/cluster 1/services" and "/api/v6/clusters/Cluster 1/services" are equivalent.

Cloudera Issue: OPSAPS-43691

HBase Indexer can possibly emit sentry client configs even if sentry isn't directly configured

On a KeyValue Store Indexer service, Sentry was enabled if the Solr dependency was using Sentry, even if the KeyValue Store Indexer was set to none in its Sentry dependency configuration. This is now corrected for CDH 5.14 or higher clusters.

After upgrading Cloudera Manager, clusters on CDH 5.14 or higher will be marked as stale if you have Sentry enabled for Solr but not enabled for KeyValue Store Indexer. If you are affected by this issue, restart the stale services to apply the fix.

Cloudera Issue: OPSAPS-43695

GenerateHostCerts command doesn't use passphrase for SSH key auth

When using the generateHostCerts command API, the password field was being used instead of the passphrase field for SSH keypair-based authentication. This is now fixed so that the userName and password fields are used for username/password authentication, and the privateKey and passphrase fields are used for keypair-based authentication.

Cloudera Issue: OPSAPS-45514

dfs.client.block.write.replace-datanode-on-failure.enable property

HBase will respect HDFS settings for `dfs.client.block.write.replace-datanode-on-failure`.

Cloudera Issue: OPSAPS-36611

API names

The following API names have changed to fix typos:

- `hiveserver2_load_balancer` has been changed to `hiveserver2_load_balancer`
- `hbase_client_java_opts` has been changed to `hdfs_client_java_opts`
- `hbase_active_master_detecton_window` has been changed to `hbase_active_master_detection_window`
- `hdfs_active_namenode_detecton_window` has been changed to `hdfs_active_namenode_detection_window`
- `mapreduce_active_jobtracker_detecton_window` has been changed to `mapreduce_active_jobtracker_detection_window`
- `yarn_active_resourcemanager_detecton_window` has been changed to `yarn_active_resourcemanager_detection_window`

The `hiveserver2_load_balancer` change affects Hive services when HiveServer 2 is configured for High Availability.

The `hdfs_client_java_opts` parameter configures the Client Java Configuration Options, found under the HDFS Gateway role configuration.

The other parameters tune the behavior of health test checking for the HBase Master, HDFS NameNode, MapReduce JobTracker, and YARN ResourceManager respectively.

Any API scripts or cluster templates referencing the old names will need to be updated to use the new names.

Cloudera Issue: OPSAPS-33266, OPSAPS-39223, and OPSAPS-24569

Cloudera Manager fails to enable Kerberos if TLS is configured

Fixed an issue where the wizard for Kerberos fails if TLS is enabled. When enabling Kerberos to a cluster running TLS, the system cannot use the privileged ports (<1024). Instead, the wizard will prompt the user to use the appropriate port values.

Cloudera Issue: OPSAPS-33345

Cloudera Manager Agent install or upgrade hangs

During Cloudera Manager agent installs or upgrades, Cloudera Manager accesses both Cloudera and non-Cloudera repositories. Fixed an issue where the installation or upgrade could hang due to a misconfigured or problematic third party repository.

Cloudera Issue: OPSAPS-45576

CDH did not install Kudu when using packages

Fixes an issue where Cloudera Manager did not install Kudu packages when CDH was installed using packages instead of parcels.

Cloudera Issue: OPSAPS-45692

"create" option in nestedUserQueue allocation rule is added to the wrong part of the allocation rules in the fair scheduler configuration

The Dynamic Resource Pools user interface now supports the following placement rules and pool creation policy can be separately configured for the parent group as well as the individual user group:

- root.primaryGroup.username
- root.secondaryExistingGroup.username
- root.[pool name].username

Previously, only the create="true|false" flag could be added to the inner element of the nestedUserQueue element. This meant that a root.primaryGroup or root.secondaryExistingGroup pool could be created, which was not correct. Now, you can add the create="true|false" flag to the actual nestedUserQueue element as well as the inner element of the nestedUserQueue element. An additional restriction is that if root.<parent>.username should use an existing pool (create = false), then root.<parent> must also use an existing pool.

Cloudera Issue: OPSAPS-42803

Display steady fairshare that correspond to weight in YARN Dynamic Resource Pool Configuration

Two columns are added to the Dynamic Resources Pool Configuration 'Resource Pool' table - Fair Share Cpu and Memory. These display the resources allocated to each pool, based on the % of resources allocated via their fair share weights. If min resources are specified for pools, the fair share values will not accurately reflect resource allocation. These values are displayed only for pools that do not have any sub-pools.

Cloudera Issue: OPSAPS-45188

[oozie] Emit correct port in load balancer urls

The 'oozie_load_balancer' CM configuration parameter has been changed. Previously it was specified as '<hostname>:<port>' format. In CM 5.15 and later the format is simply '<hostname>'. As this format change is incompatible, please note that any client reading this value via API should also read as necessary the load balancer port configuration parameters ('oozie_load_balancer_http_port' and 'oozie_load_balancer_https_port'); the correct port parameter to use depends on whether SSL is enabled (value of 'oozie_use_ssl')

Cloudera Issue: OPSAPS-43846

Yarn NodeManager stale due to missing CCgroups

Fixed an issue when using YARN with CGroups. The YARN NodeManager may show as being stale due to System Resources even when it is not. The diff of it will show named-cpu as having changed even when it was not modified.

Cloudera Issue: OPSAPS-43973

Upgraded Jetty version

Jetty updated to version 9.4.6.v20170531 to fix CVE-2017-9735.

Cloudera Issue: OPSAPS-42317

Impala Dynamic Resource Pools wrongly gives everyone access to root pool (and all child pools)

Fixed an issue where all users had access to all Impala resource pools if no users or groups were specified in the root pool. Now, no users get access to a pool if no users or groups is specified.

Cloudera Issue: OPSAPS-45046***YARN Dynamic Resource Pools wrongly gives everyone access to root pool (and all child pools)***

Fixed an issue where all users had access to all YARN resource pools if no users or groups were specified in the root pool. Now, no users get access to a pool if no users or groups is specified.

Cloudera Issue: OPSAPS-44949**Known Issues and Limitations in Cloudera Manager 6.0.0**

The following sections describe known issues and limitations for Cloudera Manager 6.0.0:

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service, or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper

Releases affected: Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower

Users affected: All

Date/time of detection: June 7, 2018

Severity (Low/Medium/High): 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))

Impact: Remote code execution

CVE: CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings “Enable TLS/SSL for ZooKeeper JMX” and “Enable TLS client authentication for JMX port” on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting “Enable JMX Agent” on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0

The restart Cloudera Manager Agent command does not restart the Agent Listener

When you restart the Cloudera Manager Agent, the Agent Listener (the `status_server` process) does not restart. This can cause issues if you make changes to TLS for Agents after you have installed Cloudera Manager since the Agent Listener needs to be restarted for TLS changes to take effect.

Workaround: Run the following command on every Agent host with `supervisorctl`:

```
/opt/cloudera/cm-agent/bin/supervisorctl -c
/var/run/cloudera-scm-agent/supervisor/supervisord.conf restart status_server
```

This command requires root access to the host. `supervisorctl` is owned by the `cloudera-scm` user, but `supervisord.conf` is owned by `root`.

Affected Versions: Cloudera Manager 6.0.x, 6.1.x

Cloudera Issue: OPSAPS-48886

Backup and Disaster Recovery (BDR) performance regression after upgrading to CDH 6.0.0

Hive replication with BDR experiences a performance regression when comparing CDH 6.0.0 and CDH 5.14.4. The slowdown occurs during the import step. For example, the performance regression may only be 10% for 4 million partitions. As the number of partitions goes down though, the performance impact becomes more visible. For example, 100,000 partitions may experience a 20% performance regression.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0; CDH 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47520

Cloudera Manager allows more than a single space in YARN Admin ACLs

When adding a YARN Admin ACL in Cloudera Manager, you are allowed to enter multiple spaces in the entry. The space is the separator between the user and group lists, and only a single space should be allowed in the entry. All entries that appear after a second single space in a YARN Admin ACL will be ignored.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-47688

Integer data types map to Float in Swagger API client

Integer data types show up as floating point numbers when using the Cloudera Manager API Python client.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1, 6.1.0

Cloudera Issue: OPSAPS-45689

Solr Service reports stale configurations even after restart

Solr reports stale configurations, and the Solr Server role fails to start with the following error: Role failed to start due to error: The archive already contains creds.localjceks. The issue occurs if your deployment has Solr and HDFS uses LDAP Group Mapping.

Workaround: If you have a CDH 5 cluster and use LDAP Group Mapping, do not upgrade to CDH 6.0.0. If you have a CDH 6.0.0 cluster, disable LDAP Group Mappings.

Affected Versions: Cloudera Manager 6.0.0 and CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47321

Apache Accumulo is not supported with Cloudera Manager

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1

Fixed Versions: Cloudera Manager 6.1.0

Cloudera Issue: OPSAPS-42807, OPSAPS-42814

Cloudera Data Science Workbench is not supported with Cloudera Manager 6.0.x

Cloudera Data Science Workbench is not supported with Cloudera Manager 6.0.x. Cloudera Data Science Workbench 1.5.0 (and higher) is supported with Cloudera Manager 6.1.x (and higher).

Affected Versions: Cloudera Manager 6.0.x

Cloudera Issue: DSE-2769

Externally authenticated users cannot view their roles or previous session

Usually, a user can see their assigned roles and most recent successful login by navigating to <Username> > **My Profile** in the Cloudera Manager Admin Console. The fields appear blank for users who use an external authentication method, such as SAML.

This issue is only a display issue and does not affect any functionality. The user can perform any tasks available to their assigned roles.

Workaround: User accounts with roles that can view the **Roles** page, such as a Full Administrator, can view the roles assigned to all Cloudera Manager user accounts and their session information.

Affected Versions: Cloudera Manager 6.0.0, 6.0.1

Fixed Versions: Cloudera Manager 6.1.0

Cloudera Issue: OPSAPS-46996, OPSAPS-47025

Package Installation of CDH Fails

When you install CDH with packages from a custom repository, ensure that the version of CDH you select for **Select the version of CDH** matches the version of CDH for the custom repository. Selecting the CDH version and specifying a custom repository are done during the **Select Repository** stage of installation.

If the versions do not match, installation fails.

Affected Versions: Cloudera Manager 6.x

Fixed Versions: N/A

Apache Issue: N/A

Cloudera Issue: OPSAPS-45703

Backup and Disaster Recovery replication to/from Cloudera Manager 6 clusters require Cloudera Manager 5.14.0 or higher

You can only use BDR to replicate to/from clusters managed by Cloudera Manager 6 with Cloudera Manager 5.14.0 or higher.

Affected versions: Cloudera Manager 6.x

Cloudera Issue: OPSAPS-42207

Stale Configurations

When you upgrade Cloudera Manager, configurations for some services may become stale. This page describes those configurations and whether a restart can be delayed.

Sentry and Solr stale configuration

On a KeyValue Store Indexer service, Sentry was enabled if the Solr dependency was using Sentry, even if the KeyValue Store Indexer was set to none in its Sentry dependency configuration. This is now corrected for CDH 5.14 or higher clusters.

After upgrading Cloudera Manager, clusters on CDH 5.14 or higher will be marked as stale if you have Sentry enabled for Solr but not enabled for KeyValue Store Indexer. If you are affected by this bug, restart the stale services to pick up this fix.

Cloudera Issue: OPSAPS-43695

Kafka Broker and MirrorMaker stale configuration

Kafka broker and MirrorMaker processes now listen on only the loopback interface for JMX connections. The fix causes Kafka brokers and MirrorMaker to be marked as stale after upgrading to Cloudera Manager 6.0.0 or later.

Perform a rolling restart of Kafka.

Cloudera Issue: OPSAPS-46633

HDFS DataNode stale configuration

After upgrading to Cloudera Manager 5.10 or higher, the following configuration for HDFS DataNodes will be marked as stale: `dfs.datanode.balance.max.concurrent.moves`.

The staleness is caused by the following new feature: HDFS balancer can now be configured to specify which hosts are included and excluded or which hosts are used as sources for transferring replicas. Additional properties for tuning the performance of the balancer can now also be configured starting with CDH 5.10.0.

You can safely ignore this warning and defer restarting.

Cloudera Issue: OPSAPS-36642

YARN MapReduce Job History stale configuration

When upgrading to Cloudera Manager 5.10, YARN will be marked as having stale configuration due to `mapreduce.jobhistory.loadedjob.tasks.max`. Unless you change this parameter and want the non-default value to take effect (only takes effect in CDH 5.9+), you can simply ignore this staleness and defer restarting YARN.

Cloudera Issue: OPSAPS-32132

Change in Hue Load Balancer version causes Hue Load Balancer and Server to be marked with stale configuration

After upgrading Cloudera Manager, the Hue Load Balancer and Server roles will be marked as having a stale configuration if all of these are true:

- The **Enable TLS/SSL for Hue** property is set to true.
- The Hue load balancer is enabled

Workaround: If your cluster uses Apache httpd 2.4 as the Hue load balancer, restart the Hue service promptly. If your cluster uses an earlier version of httpd, there is no urgency to restart the Hue service. (Apache httpd 2.4 is installed automatically by some recent versions of Linux, or may have been explicitly installed.)

Cloudera Issue: OPSAPS-40700, OPSAPS-41850

Hue Load Balancer SSL Handshake error

The Hue load balancer previously set the `ProxyPreserveHost` directive to `On`, when it should have been set to `Off`. This causes problems making SSL connections when using Apache httpd 2.4 or higher. The error caused problems when verifying the CN, which older versions of Apache httpd did not encounter because they did not properly verify the CN.

When upgrading Cloudera Manager, the Hue load balancer may be marked as having a stale configuration. If you are experiencing issues connecting to Hue with SSL, restart the Hue service to update the configuration.

Cloudera Issue: OPSAPS-40700

Maintenance State Minimal Block Replication staleness after upgrade

Upgrading to Cloudera Manager 5.12 or later may show Maintenance State Minimal Block Replication as a stale configuration under HDFS, suggesting a restart. It is safe to ignore this warning and delay restart.

Cloudera Issue: OPSAPS-39102

YARN ACL configuration property staleness after upgrade

After upgrading to Cloudera Manager 5.12 or higher, the following YARN configuration properties may show staleness warnings:

- ACL for viewing a job - `mapreduce.job.acl-view-job`
- ACL for modifying a job - `mapreduce.job.acl-modify-job`
- Enable MapReduce ACLs - `mapreduce.cluster.acls.enabled`

It is safe to defer restart if you are not using YARN job view/modify ACLs.

Cloudera Issue: OPSAPS-33586

CDH 6 Release Notes

These Release Notes provide information on new features, fixed issues, known issues, limitations, and incompatible changes for CDH 6.

If you are using CDH 5, see the [CDH 5 Release Notes](#).

To view the Release Notes for specific CDH 6 releases, see the following:

For links to the detailed change lists that describe the bug fixes and improvements to all of the CDH 6 projects, see [CDH Packaging Information](#).

For more information about installing and configuring CDH 6, see [Cloudera Installation Guide](#).

CDH 6.2.x Release Notes

To view release notes for specific CDH 6.2.x releases, see the following:

CDH 6.2.0 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for CDH 6.2.0:

New Features in CDH 6.2.0

See below for new features in CDH 6.2.0, grouped by component:

Apache Accumulo

There are no notable new features in this release.

Apache Avro

There are no notable new features in this release.

Apache Crunch

There are no notable new features in this release.

Apache Flume

In CDH 6.2, Flume is rebased on Apache Flume 1.9, which provides a number of improvements, including:

- Flume HDFS Sink retries close a configurable number of times with a configurable interval in between before attempting the recoverLease.
- Global SSL keystore parameters can be specified through the standard -D JSSE system properties or in environment variables. Component-level configuration is also possible.
- Update to Kafka 2.0 client.
- SSL/TLS support for syslog and multi port syslog sources has been added.
- The new default of hdfs.callTimeout is 30 sec.
- Two new interfaces for getting the transaction capacity and the batch size fields have been added to prevent source batch size larger than the channel transaction capacity.

For more information on new features and improvements added in Flume 1.9, see the [Apache Flume 1.9 Release Notes](#).

Apache Hadoop

Hadoop Common

There are no notable new features in this release.

HDFS

The following are some of the new features in this release of HDFS:

JournalNode Synchronization

CDH now supports synchronizing the contents of JournalNodes in the cluster. This capability helps in maintaining consistency in the contents of all the JournalNodes across the cluster.

For more information, see [Synchronizing the contents of JournalNodes](#).

Option for fixing misreplicated blocks

The `hdfs fsck` command now includes the `-replicate` option which triggers the replication of misreplicated data blocks.

For more information, see [Fixing Block Inconsistencies](#).

MapReduce

There are no notable new features in this release.

YARN

The following are some of the notable new features in this release of YARN:

GPU Usage

CDH supports NVIDIA GPU as a resource for YARN. GPU use can be enabled with Cloudera Manager.

For more information, see [Enable GPU Using Cloudera Manager](#).

Custom Resource Types

CDH supports the definition and management of custom resources. This means that the resource system in YARN is configurable. Resources can be created with Cloudera Manager.

For more information, see [Create Custom Resource Using Cloudera Manager](#).

Apache HBase

The following are some of the notable new features in this release of HBase:

HBase Pre-Upgrade Tools Checkbox

There are three pre-upgrade tools that help you to validate HBase compatibility when upgrading a CDH 5 cluster to CDH 6:

- `hbase pre-upgrade validate-dbe` and `hbase pre-upgrade validate-hfile`: These tools validate that none of your tables or snapshots uses the PREFIX_TREE Data Block Encoding.
- `hbase pre-upgrade validate-cp`: This tool validates that your co-processors are compatible with the upgrade.

When you are attempting to upgrade from a CDH 5 cluster to a CDH 6 cluster checkboxes appear to ensure that you have performed all the HBase related pre-upgrade migration steps. For more information, see [Migrating Apache HBase Before Upgrading to CDH 6](#).

HBase Serial Replication

Serial replication allows the HBase Replication to send updates to a remote cluster in an ordered way. For example, it can send the updates in the same order as the change was received by the source. There are two ways to enable this feature:

- Specify the `SERIAL => true` flag when a new peer is created:

```
hbase> add_peer 'serialpeer1', CLUSTER_KEY => "cluster.example.com:2181:/hbase", SERIAL => true
```

- Modify an existing peer:

```
hbase> set_peer_serial 'serialpeer1', true
```

If Lily HBase NRT Indexer Service is used, Cloudera recommends not to use HBase Serial Replication as it causes additional delays to propagate updates to Solr.

Additional IO Engine Support

Two new bucket cache io engine types are supported:

- `mmap`: Stores and accesses cache through memory mapping to a file under a specified path.
- `pmem`: Uses the direct access capabilities from a persistent memory devices. It can be configured only for paths mounted on DC PMEM devices.

These two engines can be configured only in Cloudera Manager using safety valve. For more information, see [Configuring the Off-heap BucketCache](#).

Apache Hive / Hive on Spark / HCatalog

Apache Hive

The following are some of the notable new features in this release of Hive:

Compile Lock Removal

Hive now supports the removal of the query compilation lock. Deactivating the compilation lock enables a controlled number of queries to compile in parallel. The default degree of parallelism (number of workers) is three, and users can configure this in Cloudera Manager depending on their needs.

Learn more about this feature in [Removing the Hive Compilation Lock](#).

Secured S3 Credentials for Hive

S3 credentials are now stored securely by Cloudera Manager for use with Hive. This enables multi-user Hive-on-S3 clusters.

Learn more at [Configuring the Amazon S3 Connector](#).

Secured ADLS Credentials for Hive

ADLS credentials are now stored securely via Cloudera Manager for use with Hive. This enables multi-user Hive-with-ADLS clusters.

Learn more at [Configuring ADLS Access Using Cloudera Manager](#).

Hive on Spark

There are no notable new features in this release.

Hue

The following are some of the notable new features in this release of Hue.

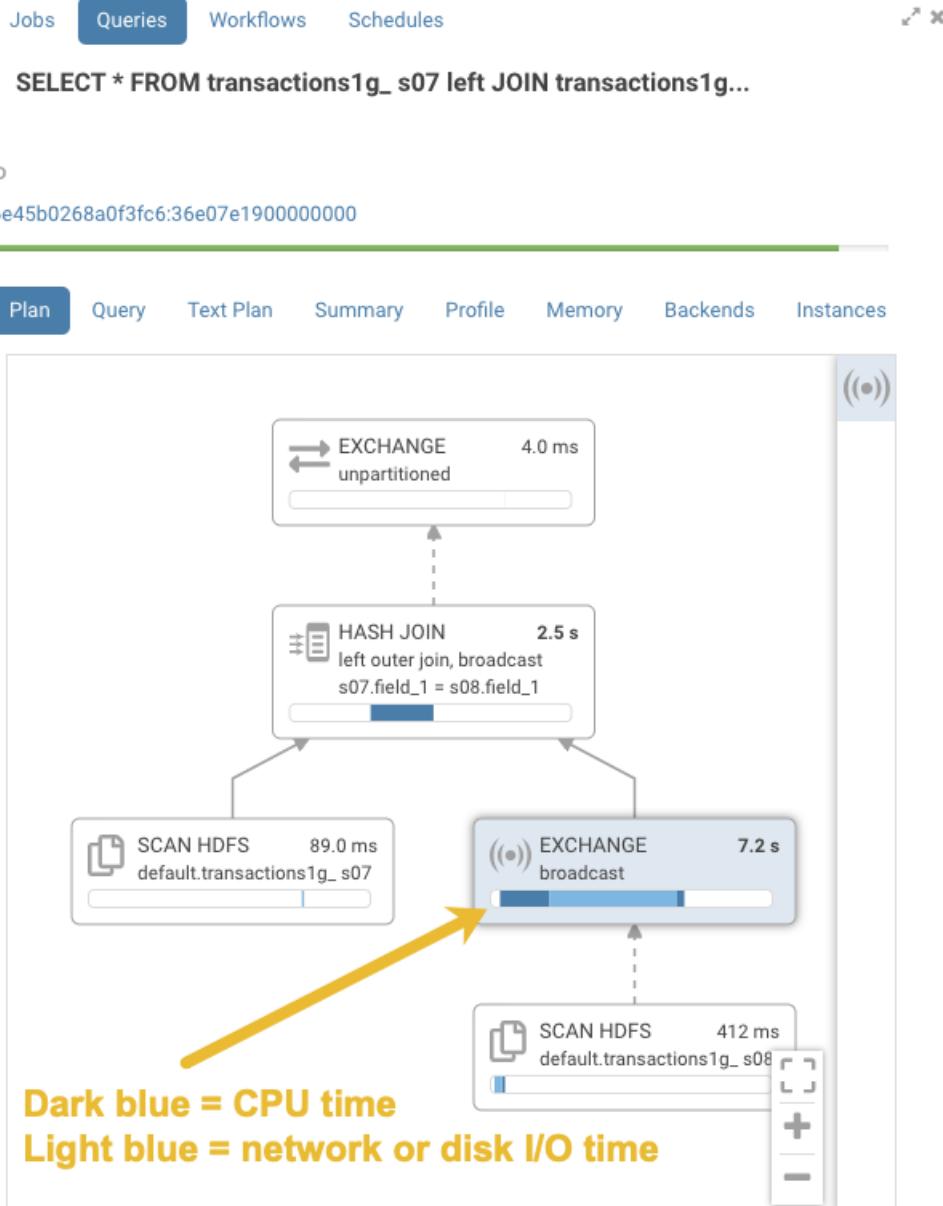
Apache Tez Integration Improvements

Now when you are using Tez as the query execution engine for Hive, jobs are displayed in the Hue Job Browser. The query ID is printed and query progress is displayed.

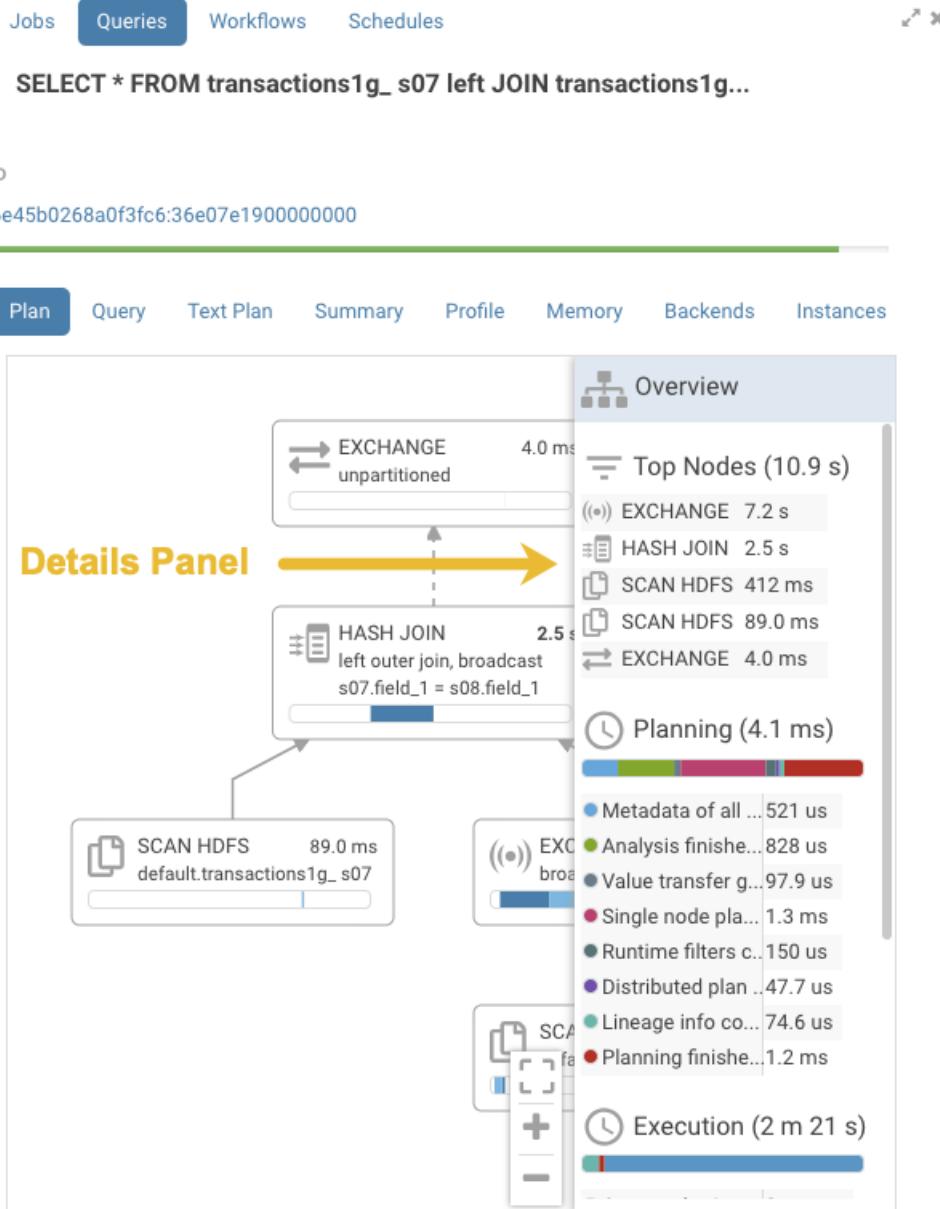
For more information about this improvement, see the [Hue blog](#).

Enhanced Impala SQL Query Troubleshooting

Graphical displays of Impala SQL query profiles have been enhanced with greater detail. This added information helps you understand where and why query bottlenecks occur and how to optimize your queries to eliminate them. For example, detailed information is now provided about CPU processing time and network or disk I/O time for each node of query execution:



In addition, click the header of the pane to open a details panel:



To access this feature:

1. Run a query in the Hue Impala editor.
2. From the menu on the left, launch the Job Browser.
3. In the Job Browser, select the **Queries** tab.
4. In the list of queries, click on the query that you just ran to launch the graphical display of the query.

For more information about this new feature, see the [Hue blog](#).

Apache Impala

The following are some of the notable new features in this release of Impala.

Multi-cluster Support

- Remote File Handle Cache

Impala can now cache remote HDFS file handles when the `--cache_remote_file_handles` impalad flag is set to `true`. This feature does not apply to non-HDFS tables, such as Kudu or HBase tables, and does not apply to the tables that store their data on cloud services such as S3 or ADLS.

See [Scalability Considerations for NameNode Traffic with File Handle Caching](#) for file handle caching in Impala.

Enhancements in Resource Management and Admission Control

- Admission Debug page is available in [Impala Daemon web UI](#) at `/admission` and provides the following information about Impala resource pools.
 - Pool configuration
 - Relevant pool stats
 - Queued queries in order of being queued (local to the coordinator)
 - Running queries (local to this coordinator)
 - Histogram of the distribution of peak memory usage by admitted queries
- A new query option, `NUM_ROWS_PRODUCED_LIMIT`, was added to limit the number of rows returned from queries.

Impala will cancel a query if the query produces more rows than the limit specified by this query option. The limit applies only when the results are returned to a client, e.g. for a `SELECT` query, but not an `INSERT` query. This query option is a guardrail against users accidentally submitting queries that return a large number of rows.

Metadata Performance Improvements

- [Automatic Metadata Sync using Hive Metastore Notification Events](#)

When enabled, the `catalogd` polls Hive Metastore (HMS) notifications events at a configurable interval and syncs with HMS. You can use the new web UI pages of the `catalogd` to check the state of the automatic invalidate event processor.

Note: This is a preview feature in CDH 6.2, and you should not use the feature without the guidance of Cloudera Support. If you are interested in using the feature, file a support ticket and work through Cloudera Support.

Compatibility and Usability Enhancements

- Impala can now read the `TIMESTAMP_MILLIS` and `TIMESTAMP_MICROS` Parquet types.
See [Using the Parquet File Format with Impala Tables](#) for the Parquet support in Impala.
- The `LEVENSHTEIN` string function is supported.

The function returns the Levenshtein distance between two input strings, the minimum number of single-character edits required to transform one string to another.

- The `IF NOT EXISTS` clause is supported in the [ALTER TABLE](#) statement.
- The new `DEFAULT_FILE_FORMAT` query option allows you to set the default table file format.

This removes the need for the `STORED AS <format>` clause. Set this option if you prefer a value that is not `TEXT`. The supported formats are:

- `TEXT`
 - `RC_FILE`
 - `SEQUENCE_FILE`
 - `AVRO`
 - `PARQUET`
 - `KUDU`
 - `ORC`
- Extended or verbose `EXPLAIN` output includes the following new information for queries:
 - The text of the analyzed query that may have been rewritten to include various optimizations and implicit casts.

- The implicit casts and literals shown with the actual types.
- CPU resource utilization (user, system, iowait) metrics were added to the [Impala profile](#) outputs.

Security Enhancement

The [REFRESH AUTHORIZATION](#) statement was implemented for refreshing authorization data.

Apache Kafka

The following are some of the notable new features in this release of Kafka CDH 6.2.0.

Rebase on Apache Kafka 2.1.0

The Kafka version in CDH 6.2.0 is based on Apache Kafka 2.1.0.

Apache Kafka 2.1.0 introduces the following notable changes:

- The internal schema used to store consumer offsets has been changed.



Warning: As a result of this change, downgrading to previous versions is not be possible after upgrade.

- Support for Zstandard compression has been added.
- Unclean leader election is automatically enabled by the controller when `unclean.leader.election.enable` config is dynamically updated by using per-topic config override.
- A new method `AdminClient#metrics()`, has been added to `AdminClient`. This allows any application using the `AdminClient` to gain more information and insight by viewing the metrics captured from the `AdminClient`.

For upstream release notes, see Apache Kafka version [2.1.0](#) release notes.

New Metrics

A number of new metrics are introduced for Kafka. The following list is only a summary, for full list of metrics, see [Metrics Reference](#).

Broker Metrics related to the following:

- Zookeeper Request Latency
- Consumer Groups Completing Rebalance
- Consumer Groups Dead
- Consumer Groups Empty
- Consumer Groups Preparing Rebalance
- Consumer Groups Stable
- Zookeeper Auth Failures
- Zookeeper Disconnects
- Zookeeper Expires
- Zookeeper Read Only Connects
- Zookeeper Sasl Authentications
- Zookeeper Sync Connects
- Incremental Fetch Session Evictions Rate
- Number of Incremental Fetch Partitions Cached

- Number of Incremental Fetch Sessions
- Unclean Leader Election Enable Rate and Time

Support for Authentication with Delegation Tokens

As of CDH 6.2.0, Cloudera supports delegation token based authentication on Kafka clusters. Delegation token based authentication is a lightweight authentication method designed to complement existing SASL authentication. Although Kafka delegation tokens make use of the SCRAM authentication model, SCRAM is not supported. For more information, see [Kafka Delegation Tokens](#).

Broker IDs are visible on the instance page in Cloudera Manager

Broker IDs can now be easily viewed in Cloudera Manager. To view broker IDs select the Kafka service and go to **Instances**. The broker IDs can be found next to each Kafka broker instance enclosed in brackets.

Default Behaviour Changes

Kafka CDH 6.2.0. Introduces the following default behaviour changes:

- Unclean leader election is automatically enabled by the controller when `unclean.leader.election.enable` config is dynamically updated by using per-topic config override.
- Diagnostic data bundles collected by Cloudera Manager from now on include information on Kafka topics. The bundle includes the information exposed by the following two commands:
 - `kafka-topics --describe`
 - `kafka-topics --list`

Apache Kudu

The following are some of the notable new features and enhancements in this release of Kudu:

- Kudu now supports location awareness. The rack assignments made in Cloudera Manager will be used in Kudu automatically.

The `kudu cluster rebalance` tool has been updated to act in accordance with the placement policy of a location-aware Kudu.

Upon upgrading, if rack locations are assigned, you should run the `kudu cluster rebalance` tool to ensure your existing tables are in compliance with the rack awareness placement policy.

See [Kudu Administration](#) for more information about this feature.

- When creating a table, the master now enforces a restriction on the total number of replicas rather than the total number of partitions. If manually overriding `--max_create_tablets_per_ts`, the maximum size of a new table has effectively been cut by a factor of its replication factor. Note that partitions can still be added after table creation.
- The compaction policy has been updated to favor reducing the number of rowsets. This can lead to faster scans and lower bootup times, particularly in the face of a “trickling inserts” workload, where rows are inserted slowly in primary key order.
- A new tablet-level metric, `average_diskrowset_height`, shows how much a replica needs to be compacted, as indicated by the average number of rowsets per unit of keyspace.
- Scans which read multiple columns of tables undergoing a heavy `UPDATE` workload are now more CPU efficient. In some cases, scan performance of such tables may be several times faster upon upgrading to this release.
- Kudu-Spark users can now provide the short “kudu” format alias to Spark. This enables using `.format("kudu")` in places where you would have needed to provide the fully qualified name like `.format("org.apache.kudu.spark.kudu")` or imported `org.apache.kudu.spark.kudu._` and used the implicit `.kudu` functions. See [Kudu Integration with Spark](#) for detail.

- The KuduSink class has been added to the Spark integration as a StreamSinkProvider, allowing structured streaming writes into Kudu.
- The amount of server-side logging has been greatly reduced for Kudu's consensus implementation and background processes. This logging was determined to be not useful and unnecessarily verbose.
- The Kudu web UI now clearly indicates which columns are a part of the primary keys.
- The new kudu table describe tool describes table attributes, including schema, partitioning, replication factor, column encodings, compressions, and default values.
- The new kudu table scan tool scans rows from a table, supporting comparison, in-list, and is-null predicates.
- The new kudu locate_row tool allows users to determine what tablet a given primary key belongs to, and whether a row exists for that primary key.
- The new kudu diagnose dump_mem_trackers tool allows users to output the contents of the /mem-trackers web UI page in the CSV format.

Apache Oozie

There are no notable new features in this release.

Apache Parquet

The following are some of the notable new features in this release of Parquet:

Support for New Logical Type Representation

A new, more flexible logical type API has been introduced in parquet-mr—based on the Thrift field in parquet-format, which has been available for a while. The new API allows storage and retrieval of different type attributes, for example, timestamp semantics and precision.

The new logical types are represented by the `LogicalTypeAnnotation` class and are entirely forward- and backward-compatible with the previous logical types. Files written using the old API can be read using the new API, and as long as no new types are used, files written using the new API can also be read using the old API.

Apache Pig

There are no notable new features in this release.

Cloudera Search

There are no notable new features in this release.

Apache Sentry

There are no notable new features in this release.

Apache Spark

Spark Driver Logs

The Spark service collects Spark driver logs when Spark applications are run in YARN-client mode or with the Spark Shell. This feature is enabled by default, and the logs are persisted to an HDFS directory and included in YARN Diagnostic Bundles.

To disable this feature, uncheck **Persist Driver Logs to Dfs** on the **Configuration** page of your Spark service.

For more information, see [Collecting Spark Driver Logs](#).

Spark Structured Streaming reference application for CDH

The Spark structured streaming reference application is a project that includes sample applications that demonstrate an Apache Kafka -> Apache Spark Structured Streaming -> Apache Kudu pipeline for ingestion. The main goal of the project is to aid customers in building a structured streaming application on CDH. For more information, visit [Spark Structured Streaming Reference Application for CDH](#) on GitHub.

Apache Sqoop

The following are some of the notable new features and enhancements in this release of Sqoop:

Support Decimal Type from Parquet Import and Export

Sqoop now supports the import and export of DECIMAL type correctly for both HDFS and Hive import. This feature is turned on by default in new CDH 6.2 clusters. It is turned off by default in older (upgraded) clusters.

Learn more at [Configuring ADLS Access Using Cloudera Manager](#).

Importing Data into Microsoft Azure Data Lake Store (Gen1 and Gen2) Using Sqoop

CDH 6.2 supports using Apache Sqoop with both generations of ADLS. You can use Sqoop to efficiently transfer bulk data between Apache Hadoop and structured datastores such as relational databases. You can use Sqoop to import data from any relational database that has a JDBC adaptor such as SQL Server, MySQL, and others, to the ADLS file system.

For more information, see [Importing Data into Microsoft Azure Data Lake Store Using Sqoop](#).

Apache Zookeeper

There are no notable new features in this release.

Fixed Issues in CDH 6.2.0

CDH 6.2.0 fixes the following issues:

Kafka JMX Tool Cannot Connect to JMX

The Kafka JMX tool cannot connect to the JMX agent of the Kafka Broker or MirrorMaker if the specified address of the JMX remote connector is bound to 127.0.0.1.

Workaround:

1. In Cloudera Manager go to **Kafka > Instances** and select the affected broker.
2. Find the **Additional Broker Java Options** and **Additional MirrorMaker Java Options** properties and add the following Java option to the configuration:

```
-Djava.rmi.server.hostname=127.0.0.1
```



Note: Configuring the **Additional MirrorMaker Java Options** property is only required if you are using JMX with MirrorMaker.

3. Restart the affected brokers.

Affected Versions: CDH 6.0.0 and higher

Fixed Versions: CDH 6.2.0

Cloudera Issue: OPSAPS-48695

Upstream Issues Fixed

The following upstream issues are fixed in CDH 6.2.0:

Apache Accumulo

There are no notable fixed issues in this release.

Apache Avro

There are no notable fixed issues in this release.

Apache Crunch

There are no notable fixed issues in this release.

Apache Flume

The following issues are fixed in CDH 6.2.0:

- [FLUME-2050](#) - Upgrade to Log4j 2.10.0
- [FLUME-2071](#) - Flume Context doesn't support float or double configuration values.
- [FLUME-2464](#) - Remove hadoop and hbase profiles.
- [FLUME-2653](#) - Allow hdfs sink inUseSuffix to be empty
- [FLUME-2698](#) - Upgrade Jetty Version
- [FLUME-2786](#), [FLUME-3056](#), [FLUME-3117](#) - Application enters a deadlock when stopped while handleConfigurationEvent
- [FLUME-2799](#) - Kafka Source - Add message offset to headers
- [FLUME-2894](#) - Flume components should stop in the correct order
- [FLUME-2976](#) - Exception when JMS source tries to connect to a Weblogic server without authentication
- [FLUME-2988](#) - Kafka Sink metrics missing eventDrainAttemptCount
- [FLUME-2989](#) - Added 2 KafkaChannel metrics
- [FLUME-3046](#) - Kafka Sink and Source Configuration Improvements
- [FLUME-3087](#) - Change log level from WARN to INFO
- [FLUME-3101](#) - Add maxBatchCount config property to Taildir Source.
- [FLUME-3115](#) - Update netty library
- [FLUME-3133](#) - Add client IP / hostname headers to Syslog sources.
- [FLUME-3142](#) - Adding HBase2 sink
- [FLUME-3158](#) - Upgrade surefire version and config
- [FLUME-3183](#) - Maven: generate SHA-512 checksum during deploy
- [FLUME-3186](#) - Make asyncHbaseClient config parameters available from Flume config
- [FLUME-3194](#) - Upgrade derby to the latest version
- [FLUME-3201](#) - Fix SyslogUtil to handle RFC3164 format in December correctly
- [FLUME-3223](#) - Flume HDFS Sink should retry close prior recover lease
- [FLUME-3228](#) - Incorrect parameter name in timestamp interceptor docs
- [FLUME-3243](#) - hdfs.callTimeout deafault increased and deprecated
- [FLUME-3246](#) - Validate flume configuration to prevent larger source batchsize than
- [FLUME-3253](#) - Update jackson-databind dependency to the latest version
- [FLUME-3270](#) - Close JMS resources in JMSMessageConsumer constructor in
- [FLUME-3281](#) - Update to Kafka 2.0
- [FLUME-3282](#) - Use slf4j in every component
- [FLUME-3294](#) - Fix polling logic in TaildirSource
- [FLUME-3296](#) - Revert log4j 2 upgrade on 1.9 branch
- [FLUME-3298](#) - Make hadoop-common optional in hadoop-credential-store-config-filter
- [FLUME-3299](#) - Fix log4j scopes in pom files
- [FLUME-3302](#) - Fix issues discovered during the release
- [FLUME-3314](#) - Fixed NPE in Kafka source/channel during offset migration

Apache Hadoop

The following issues are fixed in CDH 6.2.0:

- [HADOOP-9567](#) - Provide auto-renewal for keytab based logins.
- [HADOOP-11100](#) - Support to configure ftpClient.setControlKeepAliveTimeout.
- [HADOOP-14314](#) - The OpenSolaris taxonomy link is dead in InterfaceClassification.md.
- [HADOOP-14970](#) - MiniHadoopClusterManager does not respect the lack of format option.
- [HADOOP-15214](#) - Make Hadoop compatible with Guava 21.0.
- [HADOOP-15813](#) - Enable a more reliable SSL connection reuse.
- [HADOOP-15823](#) - ABFS: Stop requiring client ID and tenant ID for MSI.

- [HADOOP-15832](#) - Upgrade BouncyCastle to 1.60.
- [HADOOP-15860](#) - ABFS: Throw IllegalArgumentException when a directory or a file name ends with a period.

HDFS

The following issues are fixed in CDH 6.2.0:

- [HDFS-12498](#) - JournalNodeSyncer is not started in a federated HA cluster.
- [HDFS-12579](#) - JournalNodeSyncer should use fromUrl field of EditLogManifestResponse to construct the servlet Url.
- [HDFS-12716](#) - The 'dfs.datanode.failed.volumes.tolerated' property to support minimum number of volumes that should be available.
- [HDFS-12886](#) - Ignore minReplication for block recovery.
- [HDFS-12946](#) - Add a tool to check the rack configuration against EC policies.
- [HDFS-13023](#) - JournalNodeSyncer does not work on a secure cluster.
- [HDFS-13626](#) - Fix incorrect username when the setOwner operation is denied.
- [HDFS-13744](#) - OIV tool should better handle control characters present in file or directory names.
- [HDFS-13761](#) - Add toString method to the AclFeature class.
- [HDFS-13818](#) - Extend OIV to detect FSImage corruption.
- [HDFS-13996](#) - Make HttpFS ACLs RegEx-configurable.
- [HDFS-14008](#) - NameNode should log the snapshotdiff report.
- [HDFS-14015](#) - Improve error handling in hdfsThreadDestructructor in the native thread local storage.
- [HDFS-14027](#) - DFSStripedOutputStream should implement both the hsync methods.
- [HDFS-14028](#) - The HDFS OIV temporary directory deletes a folder.
- [HDFS-14053](#) - Provide ability for NameNode to re-replicate based on topology changes.
- [HDFS-14061](#) - Check if the cluster topology supports the EC policy before setting, enabling, or adding it.
- [HDFS-14125](#) - Use a parameterized log format in ECTopologyVerifier.
- [HDFS-14140](#) - JournalNodeSyncer authentication is failing in a secure cluster.
- [HDFS-14188](#) - Make the hdfs ec -verifyClusterSetup command accept an EC policy as a parameter.
- [HDFS-14231](#) - DataXceiver#run() should not log exceptions caused by InvalidTokenException as an error.

MapReduce 2

The following issues are fixed in CDH 6.2.0:

- [MAPREDUCE-4669](#) - MRAM web UI does not work with HTTPS.
- [MAPREDUCE-7125](#) - JobResourceUploader creates LocalFileSystem when it's not necessary.

YARN

The following issues are fixed in CDH 6.2.0:

- [YARN-7396](#) - NPE when accessing container logs due to null dirsHandler
- [YARN-8582](#) - Document YARN support for HTTPS in AM Web server.
- [YARN-8865](#) - RMStateStore contains large number of expired RMDelegationToken
- [YARN-8899](#) - Fixed minicluster dependency on yarn-server-web-proxy.
- [YARN-8908](#) - Fix errors in yarn-default.xml related to GPU/FPGA.
- [YARN-9087](#) - Improve logging for initialization of Resource plugins.
- [YARN-9095](#) - Removed Unused field from Resource: NUM_MANDATORY_RESOURCES
- [YARN-9213](#) - RM Web UI v1 does not show custom resource allocations for containers page
- [YARN-9318](#) - Resources#multiplyAndRoundUp does not consider Resource Types
- [YARN-9322](#) - Store metrics for custom resource types into FSQueueMetrics and query them in FairSchedulerQueueInfo
- [YARN-9323](#) - FSLeafQueue#computeMaxAMResource does not override zero values for custom resources

Apache HBase

The following issues are fixed in CDH 6.2.0:

- [HBASE-17356](#) - Add replica get support
- [HBASE-18735](#) - Provide an option to kill a MiniHBaseCluster without waiting on shutdown
- [HBASE-19695](#) - Handle disabled table for async client
- [HBASE-19722](#) - Meta query statistics metrics source
- [HBASE-20220](#) - [RSGroup] Check if table exists in the cluster before moving it to the specified regionserver group
- [HBASE-20604](#) - ProtobufLogReader#readNext can incorrectly loop to the same position in the stream until the the WAL is rolled
- [HBASE-20917](#) - MetaTableMetrics#stop references uninitialized requestsMap for non-meta region
- [HBASE-21178](#) - [BC break] : Get and Scan operation with a custom converter_class not working
- [HBASE-21215](#) - Figure how to invoke hbck2; make it easy to find
- [HBASE-21247](#) - Custom Meta WAL Provider doesn't default to custom WAL Provider whose configuration value is outside the enums in Providers
- [HBASE-21281](#) - Upgrade bouncycastle to latest
- [HBASE-21282](#) - Upgrade to latest jetty 9.2 and 9.3 versions
- [HBASE-21297](#) - ModifyTableProcedure can throw TNDE instead of IOE in case of REGION_REPLICATION change
- [HBASE-21300](#) - Fix the wrong reference file path when restoring snapshots for tables with MOB columns
- [HBASE-21314](#) - The implementation of BitSetNode is not efficient
- [HBASE-21321](#) - HBASE-21278 to branch-2.1 and branch-2.0
- [HBASE-21322](#) - Add a scheduleServerCrashProcedure() API to HbckService
- [HBASE-21336](#) - Simplify the implementation of WALProcedureMap
- [HBASE-21338](#) - Warn if balancer is an ill-fit for cluster size
- [HBASE-21342](#) - FileSystem in use may get closed by other bulk load call in secure bulkLoad
- [HBASE-21345](#) - [hbck2] Allow version check to proceed even though master is 'initializing'.
- [HBASE-21349](#) - Do not run CatalogJanitor or Nomalizer when cluster is shutting down
- [HBASE-21354](#) - Procedure may be deleted improperly during master restarts resulting in 'Corrupt'
- [HBASE-21355](#) - HStore's storeSize is calculated repeatedly which causing the confusing region split
- [HBASE-21356](#) - bulkLoadHFile API should ensure that rs has the source hfile's write permissions
- [HBASE-21363](#) - Rewrite the buildingHoldCleanupTracker method in WALProcedureStore
- [HBASE-21364](#) - Procedure holds the lock should put to front of the queue after restart
- [HBASE-21371](#) - Hbase unable to compile against Hadoop trunk (3.3.0-SNAPSHOT) due to license error
- [HBASE-21372](#) -) Set hbase.assignment.maximum.attempts to Long.MAX
- [HBASE-21375](#) - Revisit the lock and queue implementation in MasterProcedureScheduler
- [HBASE-21377](#) - Add debug log for procedure stack id related operations
- [HBASE-21384](#) - Procedure with holdlock=false should not be restored lock when restarts
- [HBASE-21385](#) - HTable.delete request use rpc call directly instead of AsyncProcess
- [HBASE-21387](#) - Race condition surrounding in progress snapshot handling in snapshot cache leads to loss of snapshot files
- [HBASE-21388](#) - No need to instantiate MemStoreLAB for master which not carry table
- [HBASE-21391](#) - RefreshPeerProcedure should also wait master initialized before executing
- [HBASE-21395](#) - Abort split/merge procedure if there is a table procedure of the same table going on
- [HBASE-21401](#) - Sanity check when constructing the KeyValue
- [HBASE-21407](#) - Resolve NPE in backup Master UI
- [HBASE-21410](#) - A helper page that help find all problematic regions and procedures
- [HBASE-21413](#) - Empty meta log doesn't get split when restart whole cluster
- [HBASE-21421](#) - Do not kill RS if reportOnlineRegions fails
- [HBASE-21423](#) - Procedures for meta table/region should be able to execute in separate workers
- [HBASE-21437](#) - Bypassed procedure throw IllegalArgumentException when its state is WAITING_TIMEOUT
- [HBASE-21439](#) - RegionLoads aren't being used in RegionLoad cost functions
- [HBASE-21440](#) - Assign procedure on the crashed server is not properly interrupted
- [HBASE-21445](#) - CopyTable by bulkload will write hfile into yarn's HDFS

- [HBASE-21468](#) - separate workers for meta table is not working
- [HBASE-21473](#) - RowIndexSeekerV1 may return cell with extra two \x00\x00 bytes which has no tags
- [HBASE-21480](#) - Taking snapshot when RS crashes prevent we bring the regions online
- [HBASE-21485](#) - Add more debug logs for remote procedure execution
- [HBASE-21490](#) - WALProcedure may remove proc wal files still with active procedures
- [HBASE-21492](#) - CellCodec Written To WAL Before It's Verified
- [HBASE-21498](#) - Master OOM when SplitTableRegionProcedure new CacheConfig and instantiate a new BlockCache
- [HBASE-21511](#) - Remove in progress snapshot check in SnapshotFileCache#getUnreferencedFiles
- [HBASE-21524](#) - Fix logging in ConnectionImplementation.isTableAvailable()
- [HBASE-21545](#) - NEW_VERSION_BEHAVIOR breaks Get/Scan with specified columns
- [HBASE-21551](#) - Memory leak when use scan with STREAM at server side -
- [HBASE-21554](#) - Show replication endpoint classname for replication peer on master web UI
- [HBASE-21567](#) - Allow overriding configs starting up the shell
- [HBASE-21568](#) - Use CacheConfig.DISABLED where we don't expect to have blockcache running
- [HBASE-21570](#) - Add write buffer periodic flush support for AsyncBufferedMutator
- [HBASE-21580](#) - Support getting Hbck instance from AsyncConnection
- [HBASE-21582](#) - If call HBaseAdmin#snapshotAsync but forget call isSnapshotFinished, then SnapshotHFileCleaner will skip to run every time
- [HBASE-21590](#) - Optimize trySkipToNextColumn in StoreScanner a bit.
- [HBASE-21592](#) - quota.addGetResult(r) throw NPE
- [HBASE-21610](#) - , numOpenConnections metric is set to -1 when zero server channel exist
- [HBASE-21620](#) - Problem in scan query when using more than one column prefix filter in some cases
- [HBASE-21629](#) - draining_servers.rb is broken
- [HBASE-21630](#) - [shell] Define ENDKEY == STOPROW
- [HBASE-21631](#) - list_quotas should print human readable values for LIMIT
- [HBASE-21639](#) - maxHeapUsage value not read properly from config during EntryBuffers initialization
- [HBASE-21645](#) - Perform sanity check and disallow table creation/modification with region replication < 1
- [HBASE-21662](#) - Add append_peer_exclude_namespaces and remove_peer_exclude_namespaces shell commands
- [HBASE-21663](#) - Add replica scan support
- [HBASE-21682](#) - Support getting from specific replica
- [HBASE-21694](#) - Add append_peer_exclude_tableCFs and remove_peer_exclude_tableCFs shell commands
- [HBASE-21704](#) - The implementation of DistributedHBaseCluster.getServerHoldingRegion is incorrect
- [HBASE-21705](#) - Should treat meta table specially for some methods in AsyncAdmin
- [HBASE-21712](#) - Make submit-patch.py python3 compatible
- [HBASE-21732](#) - Should call toUpperCase before using Enum.valueOf in some methods for ColumnFamilyDescriptor
- [HBASE-21738](#) - Remove all the CLSM#size operation in our memstore because it's an quite time consuming.
- [HBASE-21746](#) - Fix two concern cases in RegionMover
- [HBASE-21843](#) - RegionGroupingProvider breaks the meta wal file name pattern which may cause data loss for meta region
- [HBASE-21862](#) - IPCUtil.wrapException should keep the original exception types for all the connection exceptions
- [HBASE-21915](#) - Make FileLinkInputStream implement CanUnbuffer
- [HBASE-21960](#) - Ensure RESTServletContainer used by RESTServer

Apache Hive

The following issues are fixed in CDH 6.2.0:

Code Changes Might Be Required

The following fixes might require code changes for the CDH 6.2.0 release of Apache Hive:

Code Changes Should Not Be Required

The following fixes should not require code changes, but they contain improvements that might enhance your deployment:

- [HIVE-15884](#) - Optimize not between for vectorization
- [HIVE-16839](#) - Unbalanced calls to openTransaction/commitTransaction when alter the same partition concurrently
- [HIVE-18238](#) - Driver execution may not have configuration changing side-effects
- [HIVE-18652](#) - Expose remoteBytesReadToDisk via HoS
- [HIVE-19564](#) - Vectorization: Fix NULL / Wrong Results issues in Arithmetic
- [HIVE-20306](#) - Implement projection spec for fetching only requested fields from partitions
- [HIVE-20307](#) - Add support for filterspec to the getPartitions with projection API
- [HIVE-20330](#) - HCatLoader cannot handle multiple InputJobInfo objects for a job with multiple inputs
- [HIVE-20331](#) - Query with union all, lateral view and Join fails with "cannot find parent in the child operator"
- [HIVE-20484](#) - Disable Block Cache By Default With HBase SerDe
- [HIVE-20535](#) - Add new configuration to set the size of the global compile lock
- [HIVE-20661](#) - Dynamic partitions loading calls add partition for every partition 1-by-1
- [HIVE-20722](#) - Switch HS2 CompileLock to use fair locks
- [HIVE-20737](#) - Local SparkContext is shared between user sessions and should be closed only when there is no active
- [HIVE-20776](#) - HMS filterHooks on server-side in addition to client-side
- [HIVE-20796](#) - jdbc URL can contain sensitive information that should not be logged
- [HIVE-20818](#) - Views created with a WHERE subquery will regard views referenced in the subquery as direct input
- [HIVE-20843](#) - Properly detect RELY constraint in primary keys and foreign keys
- [HIVE-20914](#) - MRScratchDir permission denied when "hive.server2.enable.doAs", "hive.exec.submitviachild" are set to "true" and impersonated/proxy user is used
- [HIVE-20924](#) - Property 'hive.driver.parallel.compilation.global.limit' should be immutable at runtime
- [HIVE-20992](#) - Split the config "hive.metastore.dbaccess.ssl.properties" into more meaningful configs
- [HIVE-21015](#) - HCatLoader can't provide statistics for tables not in default DB
- [HIVE-21028](#) - get_table_meta should use a fetch plan to avoid race conditions ending up in NucleusObjectNotFoundException
- [HIVE-21030](#) - Add credential store env properties redaction in JobConf
- [HIVE-21035](#) - Race condition in SparkUtilities#getSparkSession
- [HIVE-21044](#) - Add SLF4J reporter to the metastore metrics systems
- [HIVE-21035](#): Add HMS total api count stats and connection pool stats to metrics
- [HIVE-21077](#) - Database and Catalogs should have creation time
- [HIVE-21083](#) - Remove the requirement to specify the truststore location when TLS to the database is turned on
- [HIVE-21116](#) - HADOOP_CREDSTORE_PASSWORD is not populated under yarn.app.mapreduce.am.admin.user.env
- [HIVE-21320](#) - Portget_fields() and get_tables_by_type() are not protected by HMS server access control

Hue

The following issues are fixed in CDH 6.2.0:

- [HUE-7128](#) - [core] Apply config ENABLE_DOWNLOAD to search dashboard download
- [HUE-7258](#) - [jb] Add config check for Spark history server URL
- [HUE-7919](#) - oozie error 'NoneType' object has no attribute 'is_superuser'
- [HUE-8140](#) - [editor] Stabilize multi-statement execution
- [HUE-8330](#) - [core] Multi cluster support of namespaces and compute
- [HUE-8564](#) - Avro viewer for File Browser
- [HUE-8577](#) - [autocomplete] Update Hive and Impala autocomplete to the latest version
- [HUE-8584](#) - [useradmin] Exposing errors for Add Sync Ldap Group
- [HUE-8585](#) - [useradmin] Exposing errors for Add Sync Ldap Users
- [HUE-8587](#) - Enable queries in Job Browser to work with Smart Connection Pool
- [HUE-8598](#) - [autocomplete] Improve autocomplete for CREATE statements

- [HUE-8605](#) - [metadata] Only show the Table Privilege tab when Sentry is enabled
- [HUE-8610](#) - [tb] The sample call from the Table Browser fails for computes other than default
- [HUE-8616](#) - [cluster] getNamespaces for impala returns namespace with hive compute
- [HUE-8617](#) - [frontend] Support multi cluster in invalidate metadata
- [HUE-8638](#) - [importer] Add autocompletion to query editor in second step of importer
- [HUE-8641](#) - [frontend] Trigger a namespace refresh when the context catalog is cleared
- [HUE-8645](#) - [assist] Improve namespace listing after cluster creation
- [HUE-8648](#) - [importer] Sqoop-configured RDBMS fails
- [HUE-8649](#) - [frontend] Add a performance graph component
- [HUE-8651](#) - [editor] Add a dedicated execution analysis tab in the editor
- [HUE-8657](#) - [frontend] Improve create and configure cluster forms
- [HUE-8659](#) - [importer] Fix js exception with the field editor
- [HUE-8661](#) - [assist] Enable scrollbars in context popover view sql
- [HUE-8664](#) - [importer] Fixed Flume source import properties initialization
- [HUE-8665](#) - [editor] Add basic execution analysis for Impala
- [HUE-8666](#) - [autocomplete] Fix timing issue with "... ? from table" completion
- [HUE-8667](#) - [autocomplete] Fix issue where order by and group by suggestions aren't displayed properly
- [HUE-8668](#) - [editor] Add table names to syntax checker suggestions
- [HUE-8670](#) - [cluster] Adding auto resize option to the update cluster API
- [HUE-8679](#) - [jb] Support query interface in multi cluster node
- [HUE-8680](#) - [core] Fill in Impalad WEBUI username passwords automatically if needed
- [HUE-8681](#) - [assist] Include unopened topics in the language ref filter
- [HUE-8682](#) - [backend] Change PAM lib to python-pam-1.8.4
- [HUE-8685](#) - [importer] DB importer always shows DB already exists
- [HUE-8688](#) - Update Chinese language code to enable localization
- [HUE-8690](#) - Fix Hue allows unsigned SAML assertions
- [HUE-8691](#) - [useradmin] Add/sync group does not add users if the objectClass posixGroup already exists in the group LDAP entry
- [HUE-8692](#) - [useradmin] Group sync fails if all group members are not found
- [HUE-8693](#) - [useradmin] Security app only displays 100 users in the impersonate list
- [HUE-8694](#) - [frontend] Fix scroll in the database drop-down menu
- [HUE-8695](#) - [importer] Do not show the command but submit when clicking on submit button

Apache Impala

The following issues are fixed in CDH 6.2.0:

- [IMPALA-341](#) - Remote profiles are no longer ignored by the coordinator for the queries with the `LIMIT` clause.
- [IMPALA-941](#) - Impala supports fully qualified table names that start with a number.
- [IMPALA-1048](#) - The query execution summary now includes the total time taken and memory consumed by the data sink at the root of each query fragment.
- [IMPALA-3323](#) - Fixed the issue where valid impala-shell options, such as `--ldap_password_cmd`, were unrecognized when the `--config_file` option was specified.
- [IMPALA-5397](#) - If a query has a dedicated coordinator, its end time is now set when the query releases its admission control resources. With no dedicated coordinator, the end time is set on un-registration.
- [IMPALA-5474](#) - Fixed an issue where adding a trivial subquery to a query with an error turns the error into a warning.
- [IMPALA-6521](#) - When set, experimental flags are now shown in `/varz` in web UI and log files.
- [IMPALA-6900](#) - `INVALIDATE_METADATA` operation is no longer ignored when HMS is empty.
- [IMPALA-7446](#) - Impala enables buffer pool garbage collection when near process memory limit to prevent queries from spilling to disk earlier than necessary.
- [IMPALA-7659](#) - In `COMPUTE STATS`, Impala counts the number of `NULL` values in a table
- [IMPALA-7857](#) - Logs more information about StateStore failure detection.

- [IMPALA-7928](#) - To increase the efficiency of the HDFS file handle cache, remote reads for a particular file are scheduled to a consistent set of executor nodes.
- [IMPALA-7929](#) - Impala query on tables created via Hive and mapped to HBase failed with an internal exception because the qualifier of the HBase key column is null in the mapped table. Impala relaxed the requirement and allows a NULL qualifier.
- [IMPALA-7960](#) - Impala now returns a correct result when comparing TIMESTAMP to a string literal in a binary predicate where the TIMESTAMP is casted to VARCHAR of smaller length.
- [IMPALA-7961](#) - Fixed an issue where queries running with the SYNC_DDL query option can fail when the Catalog Server is under a heavy load with concurrent catalog operations of long-running DDLs.
- [IMPALA-8026](#) - Impala query profile now reports correct row counts for all nested loop join modes.
- [IMPALA-8061](#) - Impala correctly initializes S3_ACCESS_VALIDATED variable to zero when TARGET_FILESYSTEM=3.
- [IMPALA-8154](#) - Disabled the Kerberos auth_to_local setting to prevent connection issues between impalads.
- [IMPALA-8188](#) - Impala now correctly detects an NVME device name and handles it.
- [IMPALA-8245](#) - Added hostname to the timeout error message to enable the user to easily identify the host which has reached a bad connection state with the HDFS NameNode.
- [IMPALA-8254](#) - COMPUTE STATS failed if COMPRESSION_CODEC is set.

Apache Kafka

The following issues are fixed in CDH 6.2.0:

- [KAFKA-3514](#) - Stream timestamp computation needs some further thoughts.
- [KAFKA-4932](#) - Add support for UUID serialization and deserialization
- [KAFKA-5690](#) - Add support to list ACLs for a given principal
- [KAFKA-5975](#) - No response when deleting topics and delete.topic.enable=false
- [KAFKA-6082](#) - Fence zookeeper updates with controller epoch zkVersion
- [KAFKA-6123](#) - Give client MetricsReporter auto-generated client.id
- [KAFKA-6195](#) - Resolve DNS aliases in bootstrap.server (KIP-235)
- [KAFKA-6684](#) - Support casting Connect values with bytes schema to string
- [KAFKA-6753](#) - Updating the OfflinePartitions count only when necessary
- [KAFKA-6835](#) - Enable topic unclean leader election to be enabled without controller change
- [KAFKA-6863](#) - Kafka clients should try to use multiple DNS resolved IP
- [KAFKA-6914](#) - Set parent classloader of DelegatingClassLoader same as the worker's
- [KAFKA-6923](#) - Refactor Serializer/Deserializer for KIP-336
- [KAFKA-6926](#) - Simplified some logic to eliminate some suppressions of NPath complexity checks
- [KAFKA-6950](#) - Delay response to failed client authentication to prevent potential DoS issues (KIP-306)
- [KAFKA-6998](#) - Disable Caching when max.cache.bytes are zero.
- [KAFKA-7080](#) - and KAFKA-7222: Cleanup overlapping KIP changes
- [KAFKA-7096](#) - Clear buffered data for partitions that are explicitly unassigned by user
- [KAFKA-7117](#) - Support AdminClient API in AclCommand (KIP-332)
- [KAFKA-7134](#) - KafkaLog4jAppender exception handling with ignoreExceptions
- [KAFKA-7139](#) - Support option to exclude the internal topics in kafka-topics.sh
- [KAFKA-7196](#) - Remove heartbeat delayed operation for those removed consumers at the end of each rebalance
- [KAFKA-7211](#) - MM should handle TimeoutException in commitSync
- [KAFKA-7215](#) - Improve LogCleaner Error Handling
- [KAFKA-7223](#) - In-Memory Suppression Buffering
- [KAFKA-7240](#) - total metrics in Streams are incorrect
- [KAFKA-7277](#) - Migrate Streams API to Duration instead of longMs times
- [KAFKA-7299](#) - Batch LeaderAndIsr requests for AutoLeaderRebalance
- [KAFKA-7311](#) - Reset next batch expiry time on each poll loop
- [KAFKA-7313](#) - StopReplicaRequest should attempt to remove future replica for the partition only if future replica exists
- [KAFKA-7324](#) - NPE due to lack of SASLExtensions in SASL/OAUTHBEARER

- [KAFKA-7326](#) - KStream.print() should flush on each line for PrintStream
- [KAFKA-7332](#) - Update CORRUPT_MESSAGE exception message description
- [KAFKA-7333](#) - Protocol changes for KIP-320
- [KAFKA-7338](#) - Specify AES128 default encryption type for Kerberos tests
- [KAFKA-7366](#) - Make topic configs segment.bytes and segment.ms to take effect immediately
- [KAFKA-7379](#) - [streams] send.buffer.bytes should be allowed to set -1 in KafkaStreams
- [KAFKA-7394](#) - OffsetsForLeaderEpoch supports topic describe access
- [KAFKA-7395](#) - Add fencing to replication protocol (KIP-320)
- [KAFKA-7396](#) - Materialized, Serialized, Joined, Consumed and Produced with implicit Serdes
- [KAFKA-7399](#) - KIP-366, Make FunctionConversions deprecated
- [KAFKA-7400](#) - Compacted topic segments that precede the log start offset
- [KAFKA-7403](#) - Use default timestamp if no expire timestamp set in offset commit value
- [KAFKA-7406](#) - Name join group repartition topics
- [KAFKA-7409](#) - Validate message format version before creating topics or altering configs
- [KAFKA-7415](#) - Persist leader epoch and start offset on becoming a leader
- [KAFKA-7428](#) - ConnectionStressSpec: add "action", allow multiple clients
- [KAFKA-7429](#) - Enable key/truststore update with same filename/password
- [KAFKA-7437](#) - Persist leader epoch in offset commit metadata
- [KAFKA-7439](#) - Replace EasyMock and PowerMock with Mockito in clients module
- [KAFKA-7441](#) - Allow LogCleanerManager.resumeCleaning() to be used concurrently
- [KAFKA-7456](#) - Serde Inheritance in DSL
- [KAFKA-7462](#) - Make token optional for OAuthBearerLoginModule
- [KAFKA-7464](#) - catch exceptions in "leaderEndpoint.close()" when shutting down ReplicaFetcherThread
- [KAFKA-7467](#) - NoSuchElementException is raised because controlBatch is empty
- [KAFKA-7475](#) - capture remote address on connection authentication errors, and log it
- [KAFKA-7476](#) - Fix Date-based types in SchemaProjector
- [KAFKA-7477](#) - Improve Streams close timeout semantics
- [KAFKA-7481](#) - Add upgrade/downgrade notes for 2.1.x
- [KAFKA-7482](#) - LeaderAndIsrRequest should be sent to the shutting down broker
- [KAFKA-7483](#) - Allow streams to pass headers through Serializer.
- [KAFKA-7496](#) - Handle invalid filters gracefully in KafkaAdminClient#describeAcls
- [KAFKA-7498](#) - Remove references from `common.requests` to `clients`
- [KAFKA-7501](#) - Fix producer batch double deallocation when receiving message too large error on expired batch
- [KAFKA-7505](#) - Process incoming bytes on write error to report SSL failures
- [KAFKA-7519](#) - Clear pending transaction state when expiration fails
- [KAFKA-7532](#) - Clean-up controller log when shutting down brokers
- [KAFKA-7534](#) - Error in flush calling close may prevent underlying store from closing
- [KAFKA-7535](#) - KafkaConsumer doesn't report records-lag if isolation.level is read_committed
- [KAFKA-7560](#) - PushHttpMetricsReporter should not convert metric value to double
- [KAFKA-7742](#) - Fixed removing hmac entry for a token being removed from DelegationTokenCache

Apache Kudu

The following issues are fixed in CDH 6.2.0:

- The Kudu Python client now detects and reports on conflicting/incorrect initialization of the OpenSSL library to avoid glitches and undefined behavior.
- [KUDU-1678](#) - Fixed a crash caused by a race condition between altering tablet schemas and deleting tablet replicas.
- [KUDU-2680](#) - Now the `kudu fs update_dirs` tool can correctly remove directories in the presence of tablet tombstones.

- [KUDU-2195](#) - Now you can use the `--cmeta_force_fsync` flag to fsync Kudu's consensus metadata more aggressively. Setting this to `true` may decrease Kudu's performance, but will improve its durability in the face of power failures and forced shutdowns.
- [KUDU-2684](#) - Fixed an issue that would cause an excessive amount of RPC traffic from Kudu masters if the tablet servers were configured with duplicated master addresses.
- [KUDU-2688](#) - Fixed an issue that would cause the `kudu cluster rebalance` tool to run indefinitely in the case of tables with a replication factor of 2.
- [KUDU-2690](#) - Fixed an issue that could lead to a failure to bootstrap tablet replicas that were a part of workloads with many alter table operations.
- [KUDU-2710](#) - Fixed an issue with the Java scanner's `keepAlive` that could lead to a permanent hang in the scanner.
- [KUDU-2706](#) - Fixed an issue that would cause undefined behavior upon connecting to a secure cluster concurrently from multiple C++ clients.

Apache Oozie

The following issues are fixed in CDH 6.2.0:

- [OOZIE-1393](#) - Allow sending emails via TLS
- [OOZIE-2211](#) - Remove OozieCLI#validateCommandV41
- [OOZIE-2339](#) - [fluent-job] Minimum Viable Fluent Job API
- [OOZIE-2352](#) - Unportable shebang in shell scripts
- [OOZIE-2494](#) - Cron syntax not handling DST properly
- [OOZIE-2684](#) - Bad database schema error for WF_ACTIONS table
- [OOZIE-2718](#) - Improve -dryrun for bundles
- [OOZIE-2791](#) - ShareLib installation may fail on busy Hadoop clusters
- [OOZIE-2826](#) - Upgrade joda-time to 2.9.9
- [OOZIE-2829](#) - Improve sharelib upload to accept multiple source folders
- [OOZIE-2937](#) - Remove redundant groupId from the child POMs
- [OOZIE-2942](#) - [examples] Fix Findbugs warnings
- [OOZIE-2949](#) - Fix and backportEscape quotes whitespaces in Sqoop <command> field
- [OOZIE-3109](#) - [log-streaming] Escape HTML-specific characters
- [OOZIE-3134](#) - Potential inconsistency between the in-memory SLA map and the Oozie database
- [OOZIE-3155](#) - [ui] Job DAG is not refreshed when a job is finished
- [OOZIE-3156](#) - Retry SSH action check when cannot connect to remote host
- [OOZIE-3160](#) - PriorityDelayQueue put()/take() can cause significant CPU load due to busy waiting
- [OOZIE-3178](#) - /bin/mkdirsh -Papache-release fails due to javadoc errors
- [OOZIE-3185](#) - Upgrade org.apache.derby to 10.11.1.1
- [OOZIE-3193](#) - Applications are not killed when submitted via subworkflow
- [OOZIE-3208](#) - "It should never happen" error messages should be more specific to root cause
- [OOZIE-3209](#) - XML schema error when submitting pyspark example
- [OOZIE-3210](#) - [build] Revision information is empty
- [OOZIE-3219](#) - Cannot compile with hadoop 3.1.0
- [OOZIE-3224](#) - Upgrade Jetty to 9.3
- [OOZIE-3227](#) - Eliminate duplicate dependencies when using Hadoop 3 DistributedCache
- [OOZIE-3229](#) - [client] [ui] Improved SLA filtering options
- [OOZIE-3233](#) - Remove DST shift from the coordinator job's end time
- [OOZIE-3235](#) - Upgrade ActiveMQ to 5.15.3
- [OOZIE-3260](#) - [sla] Remove stale item above max retries on JPA related errors from in-memory SLA map
- [OOZIE-3278](#) - Oozie fails to start with Hadoop 2.6.0
- [OOZIE-3297](#) - Retry logic does not handle the exception from BulkJPAExecutor properly
- [OOZIE-3298](#) - [MapReduce action] External ID is not filled properly and failing MR job is treated as SUCCEEDED
- [OOZIE-3303](#) - Oozie UI does not work after Jetty 9.3 upgrade
- [OOZIE-3304](#) - Parsing sharelib timestamps is not threadsafe

Cloudera Enterprise 6 Release Guide

- [OOZIE-3307](#) - [core] Limit heap usage of LauncherAM
- [OOZIE-3309](#) - Runtime error during /v2/sla filtering for bundle
- [OOZIE-3310](#) - SQL error during /v2/sla filtering
- [OOZIE-3330](#) - [spark-action] Remove double quotes inside plain option values
- [OOZIE-3331](#) - [spark-action] Inconsistency while parsing quoted Spark options
- [OOZIE-3334](#) - Don't use org.apache.hadoop.hbase.security.User in HDFSCredentials
- [OOZIE-3340](#) - [fluent-job] Create error handler ACTION only if needed
- [OOZIE-3348](#) - [Hive action] Remove dependency hive-contrib
- [OOZIE-3354](#) - [core] [SSH action] SSH action gets hung
- [OOZIE-3369](#) - [core] Upgrade guru.nidi:graphviz-java to 0.7.0
- [OOZIE-3370](#) - Property filtering is not consistent across job submission
- [OOZIE-3389](#) - Getting input dependency list on the UI throws NPE
- [OOZIE-3390](#) - [Shell action] STDERR contains a bogus error message
- [OOZIE-3400](#) - [core] Fix PurgeService sub-sub-workflow checking

Apache Parquet

The following issues are fixed in CDH 6.2.0:

- [PARQUET-196](#) - parquet-tools command for row count & size
- [PARQUET-852](#) - Slowly ramp up sizes of byte in ByteBasedBitPackingEncoder
- [PARQUET-969](#) - Decimal datatype support for parquet-tools output
- [PARQUET-1336](#) - PrimitiveComparator should implements Serializable
- [PARQUET-1407](#) - Avro: Fix binary values returned from dictionary encoding
- [PARQUET-1421](#) - InternalParquetRecordWriter logs debug messages at the INFO level
- [PARQUET-1440](#) - Parquet-tools: Parse int32 or int64 decimal values to big decimals with the proper scale
- [PARQUET-1472](#) - Parquet-tools: Parse int32 or int64 decimal values to big decimals with the proper scale
- [PARQUET-1475](#) - Fix lack of cause propagation in DirectCodecFactory.ParquetCompressionCodecException
- [PARQUET-1510](#) - Fix notEq for optional columns with null values
- [PARQUET-1527](#) - [parquet-tools] cat command throw java.lang.ClassCastException

Apache Pig

There are no notable fixed issues in this release.

Cloudera Search

The following issues are fixed in CDH 6.2.0:

- [SOLR-2834](#) - Handle CharacterFilters in Solr
- [SOLR-8207](#) - Collections with underscores in name no longer cause a crash the Cloud->Nodes UI
- [SOLR-8207](#) - Add "Nodes" view to the Admin UI "Cloud" tab, listing nodes and key metrics
- [SOLR-8207](#) - Nodes view support for shard_1_1_1 format and replica1, replica_1 format. Show core state in label if not 'active'
- [SOLR-12570](#) - OpenNLPExtractNamedEntitiesUpdateProcessor cannot support multi fields because pattern replacement doesn't work correctly
- [SOLR-12597](#) - Migrate API should fail requests that do not specify split.key parameter
- [SOLR-12649](#) - CloudSolrClient retries requests unnecessarily exception from server
- [SOLR-12670](#) - RecoveryStrategy logs wrong wait time when retrying recovery
- [SOLR-12679](#) - MiniSolrCloudCluster.stopJettySolrRunner should remove jetty from the internal list
- [SOLR-12679](#) - MiniSolrCloudCluster.startJettySolrRunner method should not add a duplicate jetty instance to the list
- [SOLR-12770](#) - Make it possible to configure a host whitelist for distributed search
- [SOLR-12776](#) - Setting of TMP in solr.cmd causes invisibility of Solr to JDK tools

Apache Sentry

The following issues are fixed in CDH 6.2.0:

- [SENTRY-1797](#) - SentryKerberosContext should use periodic executor instead of managing periodic execution via run() method.
- [SENTRY-2146](#) - Add better error handling to ResourceAuthorizationProvider and improve logging in related classes
- [SENTRY-2205](#) - Improve Sentry NN Logging.
- [SENTRY-2301](#) - Log where sentry stands in the snapshot fetching process, periodically
- [SENTRY-2329](#) - Integrate sentry with Hadoop 3.1.1
- [SENTRY-2372](#) - SentryStore should not implement grantOptionCheck
- [SENTRY-2428](#) - Skip null partitions or partitions with null sds entries
- [SENTRY-2431](#) - Update Solr permission mapping to include Metric history reading permission
- [SENTRY-2432](#) - The case of a username is ignored when determining object ownership
- [SENTRY-2433](#) - Dropping object privileges does not include update of dropping user privileges
- [SENTRY-2436](#) - Add annotations for classes that are used in binding as public
- [SENTRY-2437](#) - When granting privileges a single transaction per grant causes long delays
- [SENTRY-2441](#) - When MAuthzPathsMapping is deleted all associated MPaths should be deleted automatically..
- [SENTRY-2444](#) - SigUtils signal handler needs a way to unregister functions.
- [SENTRY-2458](#) - Separate Web UI and service from service-server to prevent circular dependencies
- [SENTRY-2460](#) - Export sentry permission information to HDFS location.
- [SENTRY-2466](#) - PortCreate generic sentry store metrics
- [SENTRY-2477](#) - When requesting for deltas check if nn seq num is 1 more than latest sequence num
- [SENTRY-2481](#) - Filter HMS server-side objects based on HMS user authorization (Sergio Pena, reviewed by Na Li, Arjun Mishra).
- [SENTRY-2483](#) - Implement HMS PreReadEvent support in MetastoreAuthzBinding
- [SENTRY-2486](#) - Wrong user name when sentry HMSFollower gets full snapshot from HMS at insecure mode
- [SENTRY-2488](#) - PortAdd privilege cache to sentry hive bindings in DefaultAccessValidator
- [SENTRY-2490](#) - PortWhen building a full perm update for each object we only build 1 privilege per role
- [SENTRY-2492](#) - Consecutive ALL grants get deleted when multiple roles have ALL grants on that object
- [SENTRY-2493](#) - Sentry store api's for path mapping should handle empty/null paths.
- [SENTRY-2495](#) - Support Conjunctive Matching in Solr QueryDocAuthorizationComponent
- [SENTRY-2496](#) - Support multi-field attribute based document level controls for Solr.
- [SENTRY-2497](#) - show grant role results should handle case where URI doesn't have a defined scheme. (Haley Reeve reviewed by Na Li).
- [SENTRY-2498](#) - Exception while deleting paths that doesn't exist
- [SENTRY-2500](#) - CREATE on server does not provide HMS server side read authorization for get_all_tables(database_name)
- [SENTRY-2501](#) - Add cache for HMS server filtering hook
- [SENTRY-2502](#) - Sentry NN plug-in stops fetching updates from sentry server.
- [SENTRY-2503](#) - Failed to revoke the privilege from impala-shell if the privilege added from beeline cli on multi-clusters
- [SENTRY-2503](#) - Failed to revoke the privilege from impala-shell if the privilege added from beeline cli on multi-clusters. Filter grant option correctly

Apache Spark

The following issues are fixed in CDH 6.2.0:

- [SPARK-22148](#) - [SPARK-15815][SCHEDULER] Acquire new executors to avoid hang because of blacklisting
- [SPARK-23257](#) - [K8S] Kerberos Support for Spark on K8S
- [SPARK-23781](#) - [CORE] Merge token renewer functionality into HadoopDelegationTokenManager.
- [SPARK-23831](#) - Revert "[SQL] Add org.apache.derby to IsolatedClientLoader"
- [SPARK-24434](#) - [K8S] pod template files
- [SPARK-24553](#) - [UI][FOLLOWUP][2.4 BACKPORT] Fix unnecessary UI redirect

- [SPARK-24920](#) - [CORE] Allow sharing Netty's memory pool allocators
- [SPARK-24958](#) - [CORE] Add memory from procfs to executor metrics.
- [SPARK-25003](#) - [PYSPARK] Use SessionExtensions in Pyspark
- [SPARK-25023](#) - Clarify Spark security documentation
- [SPARK-25118](#) - [CORE] Persist Driver Logs in Client mode to Hdfs
- [SPARK-25222](#) - [K8S] Improve container status logging
- [SPARK-25451](#) - [SPARK-26100][CORE] Aggregated metrics table doesn't show the right number of the total tasks
- [SPARK-25501](#) - [SS] Add kafka delegation token support.
- [SPARK-25515](#) - [K8S] Adds a config option to keep executor pods for debugging
- [SPARK-25560](#) - [SQL] Allow FunctionInjection in SparkExtensions
- [SPARK-25682](#) - [K8S] Package example jars in same target for dev and distro images.
- [SPARK-25689](#) - [CORE] Follow up: don't get delegation tokens when kerberos not available.
- [SPARK-25689](#) - [YARN] Make driver, not AM, manage delegation tokens.
- [SPARK-25730](#) - [K8S] Delete executor pods from kubernetes after figuring out why they died
- [SPARK-25745](#) - [K8S] Improve docker-image-tool.sh script
- [SPARK-25778](#) - WriteAheadLogBackedBlockRDD in YARN Cluster Mode Fails ...
- [SPARK-25786](#) - [CORE] If the ByteBuffer.hasArray is false , it will throw UnsupportedOperationException for Kryo
- [SPARK-25815](#) - [K8S] Support kerberos in client mode, keytab-based token renewal.
- [SPARK-25828](#) - [K8S] Bumping Kubernetes-Client version to 4.1.0
- [SPARK-25837](#) - [CORE] Fix potential slowdown in AppStatusListener when cleaning up stages
- [SPARK-25875](#) - [K8S] Merge code to set up driver command into a single step.
- [SPARK-25876](#) - [K8S] Simplify kubernetes configuration types.
- [SPARK-25877](#) - [K8S] Move all feature logic to feature classes.
- [SPARK-25905](#) - [CORE] When getting a remote block, avoid forcing a conversion to a ChunkedByteBuffer
- [SPARK-25922](#) - [K8] Spark Driver/Executor "spark-app-selector" label mismatch
- [SPARK-25957](#) - [K8S] Make building alternate language binding docker images optional
- [SPARK-25960](#) - [K8S] Support subpath mounting with Kubernetes
- [SPARK-26002](#) - [SQL] Fix day of year calculation for Julian calendar days
- [SPARK-26011](#) - [SPARK-SUBMIT] Yarn mode pyspark app without python main resource does not honor "spark.jars.packages"
- [SPARK-26029](#) - [BUILD][2.4] Bump previousSparkVersion in MimaBuild.scala to be 2.3.0
- [SPARK-26094](#) - [CORE][STREAMING] createNonEcFile creates parent dirs.
- [SPARK-26109](#) - [WEBUI] Duration in the task summary metrics table and the task table are different
- [SPARK-26119](#) - [CORE][WEBUI] Task summary table should contain only successful tasks' metrics
- [SPARK-26186](#) - [SPARK-26184][CORE] Last updated time is not getting updated for the Inprogress application
- [SPARK-26194](#) - [K8S] Auto generate auth secret for k8s apps.
- [SPARK-26219](#) - [CORE][BRANCH-2.4] Executor summary should get updated for failure jobs in the history server UI
- [SPARK-26236](#) - [SS] Add kafka delegation token support documentation.
- [SPARK-26239](#) - File-based secret key loading for SASL.
- [SPARK-26256](#) - [K8S] Fix labels for pod deletion
- [SPARK-26267](#) - [SS] Retry when detecting incorrect offsets from Kafka
- [SPARK-26304](#) - [SS] Add default value to spark.kafka.sasl.kerberos.service.name parameter
- [SPARK-26307](#) - [SQL] Fix CTAS when INSERT a partitioned table using Hive serde
- [SPARK-26322](#) - [SS] Add spark.kafka.sasl.token.mechanism to ease delegation token configuration.
- [SPARK-26493](#) - [SQL] Allow multiple spark.sql.extensions
- [SPARK-26592](#) - [SS] Throw exception when kafka delegation token tried to obtain with proxy user
- [SPARK-26595](#) - [CORE] Allow credential renewal based on kerberos ticket cache.
- [SPARK-26694](#) - [CORE] Progress bar should be enabled by default for spark-shell
- [SPARK-26726](#) - Synchronize the amount of memory used by the broadcast variable to the UI display

- [SPARK-26745](#) - [SPARK-24959][SQL][BRANCH-2.4] Revert count optimization in JSON datasource by
- [SPARK-26753](#) - [CORE] Fixed custom log levels for spark-shell by using Filter instead of Threshold
- [SPARK-26873](#) - [SQL] Use a consistent timestamp to build Hadoop Job IDs.

Apache Sqoop

The following issues are fixed in CDH 6.2.0:

- [SQOOP-3237](#) - Mainframe FTP transfer option to insert custom FTP commands prior to transfer
- [SQOOP-3382](#) - Add parquet numeric support for Parquet in hdfs import
- [SQOOP-3396](#) - Add parquet numeric support for Parquet in Hive import

Apache Zookeeper

There are no notable fixed issues in this release.

Unsupported Features in CDH 6.2.0

This page lists the unsupported features in CDH 6.2.x. For the complete list of Known Issues and Limitations, see [Known Issues and Limitations in CDH 6.2.0](#) on page 236.

Apache Hadoop Unsupported Features

The following sections list unsupported features in Hadoop common components:

- [HDFS Unsupported Features](#) on page 199
- [YARN Unsupported Features](#) on page 199

HDFS Unsupported Features

The following HDFS features are not supported in CDH 6.x:

- ACLs for the NFS gateway
- Aliyun Cloud Connector
- HDFS Router Based Federation
- HDFS truncate
- More than two NameNodes
- Openstack Swift
- Quota support for Storage Types
- SFTP FileSystem
- Upgrade Domain
- Variable length block
- ZStandard Compression Codec

YARN Unsupported Features

The following YARN features are not supported in CDH 6.2.x:

- Application Timeline Server v2 (ATSV2)
- Cgroup Memory Enforcement
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Docker on YARN (DockerContainerExecutor)
- Native Services
- New Aggregated Log File Format
- Node Labels
- Pluggable Scheduler Configuration
- Reservation REST APIs
- Resource Estimator Service
- Resource Profiles

- Rolling Log Aggregation
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- YARN WebUI v2

Apache HBase Unsupported Features

The following HBase features are not supported in CDH 6.2.x:

- Master hosting meta
- Cloudera does not provide support for user-provided custom coprocessors of any kind.
- Server-side encryption of HFiles. You should configure HDFS client-side encryption.
- In-memory compaction
- Visibility labels
- Stripe compaction
- Clients setting priority on operations
- Specifying a custom asynchronous connection implementation
- Client tarball
- Rolling upgrade from CDH 5 HBase versions

Apache Hive Unsupported Features

The following Hive features are not supported in CDH 6.2.x:

- AccumuloStorageHandler ([HIVE-7068](#))
- ACID ([HIVE-5317](#))
- Built-in `version()` function is not supported (CDH-40979)
- Cost-based Optimizer (CBO)
- Explicit Table Locking
- HCatalog - HBase plugin
- Hive Authorization (Instead, use Apache Sentry.)
- Hive on Apache Tez
- Hive Local Mode Execution
- Hive Metastore - Derby
- Hive Web Interface (HWI)
- HiveServer1 / JDBC 1
- HiveServer2 Dynamic Service Discovery (HS2 HA) ([HIVE-8376](#))
- HiveServer2 - HTTP Mode (Use THRIFT mode.)
- HPL/SQL ([HIVE-11055](#))
- LLAP (Live Long and Process framework)
- Scalable Dynamic Partitioning and Bucketing Optimization ([HIVE-6455](#))
- Session-level Temporary Tables ([HIVE-7090](#))
- Table Replication Across HCatalog Instances ([HIVE-7341](#))

Apache Kafka Unsupported Features

The following Kafka features are not supported in CDH 6.2.x:

- Idempotent and transactional capabilities in the producer are currently an unsupported beta feature given their maturity and complexity. This feature will be supported in a future release.
- Kafka Connect is included in CDH 6.2.0, but is not supported. Flume and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. See [Flafka: Apache Flume Meets Apache Kafka for Event Processing](#) for more information.

- Kafka Streams is included in CDH 6.2.0, but is not supported. Instead, use Spark and Spark Streaming have a fully functional ETL/stream processing pipeline. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for more information.
- The Kafka default authorizer is included in CDH 6.2.0, but is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.

Apache Oozie Unsupported Features

The following Oozie feature is not supported in CDH 6.2.x:

- Conditional coordinator input logic.

Apache Pig Unsupported Features

The following Pig features are not supported in CDH 6.2.x:

- Pig on Tez is not supported in CDH 6.1.x ([PIG-3446](#) / [PIG-3419](#)).
- Pig on Spark.

Cloudera Search Unsupported Features

The following Search features are not supported in CDH 6.2.x:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation
- Saving search results

Apache Sentry Unsupported Features

The following Sentry features are not supported in CDH 6.2.x:

- Import and export of Sentry metadata to and from Sentry servers
- Sentry shell command line for Hive
- Relative URI paths ([Known Issue](#))
- Object types Server and URL in

```
show grant role <role name> on object <object name>
```

- ([Known Issue](#))
- ALTER and DROP privileges for Hive and Impala

In addition, as of CDH 6.0.x, Sentry policy files have been removed. See the Sentry [Incompatible Changes](#) for more information.

Apache Spark Unsupported Features

The following Spark features are not supported in CDH 6.2.x:

- Apache Spark experimental features/APIs are not supported unless stated otherwise.
- Using the JDBC Datasource API to access Hive or Impala is not supported
- ADLS not Supported for All Spark Components. Microsoft Azure Data Lake Store (ADLS) is a cloud-based filesystem that you can access through Spark applications. Spark with Kudu is not currently supported for ADLS data. (Hive on Spark is available for ADLS in CDH 5.12 and higher.)
- IPython / Jupyter notebooks is not supported. The IPython notebook system (renamed to Jupyter as of IPython 4.0) is not supported.
- Certain Spark Streaming features not supported. The `mapWithState` method is unsupported because it is a nascent unstable API.
- Thrift JDBC/ODBC server is not supported
- Spark SQL CLI is not supported

- GraphX is not supported
- SparkR is not supported
- Structured Streaming is supported, but the following features of it *are not*:
 - Continuous processing, which is still experimental, is not supported.
 - Stream static joins with HBase have not been tested and therefore are not supported.
- Spark cost-based optimizer (CBO) not supported.

Apache Swoop Unsupported Features

The following Swoop feature is not supported in CDH 6.2.x:

- [import-mainframe](#)

Incompatible Changes in CDH 6.2.0



Important:

In addition to incompatible changes, CDH 6 also deprecated or removed support for several components, including Spark 1 and MapReduce v1. For information about components, sub-components, or functionality that are deprecated or no longer supported, see [Deprecated Items](#) on page 667.

See below for incompatible changes in CDH 6.2.0, grouped by component:

Apache Accumulo

CDH 6.2.0 introduces no new incompatible changes for Apache Accumulo.

Apache Avro

API Changes

One method was removed in CDH 6.0.0:

```
GenericData.toString (Object datum, StringBuilder buffer)
```

Incompatible Changes from Avro 1.8.0

- Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.
- [AVRO-997](#): Generic API requires GenericEnumSymbol - likely to break current Generic API users that often have String or Java Enum for these fields
- [AVRO-1502](#): Avro Objects now Serializable - IPC needs to be regenerated/recompiled
- [AVRO-1602](#): removed Avro internal RPC tracing, presumed unused. Current rec would be HTrace
- [AVRO-1586](#): Compile against Hadoop 2 - probably not an issue since we've been compiling against Hadoop 2 for C5.
- [AVRO-1589](#): [Java] ReflectData.AllowNulls will create incompatible Schemas for primitive types - may need a KI since it used to fail at runtime but now will fail earlier.

Apache Crunch



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

The following changes are introduced in CDH 6.0.0, and are not backward compatible:

- Crunch is available only as Maven artifacts from the Cloudera Maven repository. It is not included as part of CDH. For more information, see [Apache Crunch Guide](#).

- Crunch supports only Spark 2 and higher releases.
- Crunch supports only HBase 2 and higher releases.
 - The API methods in Crunch-HBase use HBase 2 API types and methods.

Apache Flume

AsyncHBaseSink and HBaseSink

CDH 6 uses HBase 2.0. AsyncHBaseSink is incompatible with HBase 2.0 and is not supported in CDH 6. HBaseSink has been replaced with HBase2Sink. HBase2Sink works the same way as HBaseSink. The only difference is that it is compatible with HBase 2.0. The only additional configuration required to use HBase2Sink is to replace the component name in your configuration.

For example, replace this text:

```
agent.sinks.my_hbase_sink.type = hbase
```

With this:

```
agent.sinks.my_hbase_sink.type = hbase2
```

Or, if you use the FQN of the sink class, replace this text:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase.HBaseSink
```

With this:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase2.HBase2Sink
```

For more information about how to configure HBase2Sink, see [Importing Data Into HBase](#).

com.google.common.collect.ImmutableMap

Flume has removed `com.google.common.collect.ImmutableMap` from the `org.apache.flume.Context` API and replaced it with `java.util.Map` due to Guava compatibility issues ([FLUME-2957](#)). Plugins using the `Context.getParameters()` and `Context.getSubProperties()` APIs will need to assign the return value of those methods to a `Map<String, String>` variable instead of an `ImmutableMap<String, String>` variable, if they do not already do so. Most usages in the Flume codebase already used `Map<String, String>` at the time of this change.

Apache Hadoop

- [HDFS Incompatible Changes](#) on page 203
- [MapReduce](#) on page 204
- [YARN](#) on page 205

HDFS Incompatible Changes

CDH 6.2.0 introduces no new incompatible changes for HDFS.

CDH 6.1.0 introduces no new incompatible changes for HDFS.

CDH 6.0.1 introduces no new incompatible changes for HDFS.

CDH 6.0.0, introduces the following incompatible changes for HDFS:

- HFTP has been removed.
- The S3 and S3n connectors have been removed. Users should now use the S3a connector.
- The BookkeeperJournalManager has been removed.
- Changes were made to the structure of the HDFS JAR files to better isolate clients from Hadoop library dependencies. As a result, client applications that depend on Hadoop's library dependencies may no longer work. In these cases, the client applications will need to include the libraries as dependencies directly.

- Several library dependencies were upgraded. Clients that depend on those libraries may break because the library version changes. In these cases, the client applications will need to either be ported to the new library versions or include the libraries as dependencies directly.
- [HDFS-6962](#) changes the behavior of ACL inheritance to better align with POSIX ACL specifications, which states that the umask has no influence when a default ACL propagates from parent to child. Previously, HDFS ACLs applied the client's umask to the permissions when inheriting a default ACL defined on a parent directory. Now, HDFS can ignore the umask in these cases for improved compliance with POSIX. This behavior is on by default due to the inclusion of [HDFS-11957](#). It can be configured by `settingdfs.namenode posix.acl.inheritance.enabled` in `hdfs-site.xml`. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-11957](#) changes the default behavior of ACL inheritance introduced by [HDFS-6962](#). Previously, the behavior was disabled by default. Now, the feature is enabled by default. Any code expecting the old ACL inheritance behavior will have to be updated. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-6252](#) removed `dfshealth.jsp` since it is part of the old NameNode web UI. By default, Cloudera Manager links to the new NameNode web UI, which has an equivalent health page at `dfshealth.html`.
- [HDFS-11100](#) changes the behavior of deleting files protected by a sticky bit. Now, the deletion fails.
- [HDFS-10689](#) changes the behavior of the `hdfs dfs chmod` command. Now, the command resets sticky bit permission on a file/directory when the leading sticky bit is omitted in the octal mode (like 644). When a file or directory permission is applied using octal mode and sticky bit permission needs to be preserved, then it has to be explicitly mentioned in the permission bits (like 1644).
- [HDFS-10650](#) changes the behavior of `DFSClient#mkdirs` and `DFSClient#primitiveMkdir`. Previously, they create a new directory with the default permissions 00666. Now, they will create a new directory with permission 00777.
- [HADOOP-8143](#) changes the default behavior of `distcp`. Previously, the `-pb` option was not used by default, which may have caused some checksums to fail when block sizes did not match. Now, the `-pb` option is included by default to preserve block size when using `distcp`.
- [HADOOP-10950](#) changes several heap management variables:
 - `HADOOP_HEAPSIZE` variable has been deprecated. Use `HADOOP_HEAPSIZE_MAX` and `HADOOP_HEAPSIZE_MIN` instead to set `Xmx` and `Xms`
 - The internal variable `JAVA_HEAP_MAX` has been removed.
 - Default heap sizes have been removed. This will allow for the JVM to use auto-tuning based upon the memory size of the host. To re-enable the old default, configure `HADOOP_HEAPSIZE_MAX="1g"` in `hadoop-env.sh`.
 - All global and daemon-specific heap size variables now support units. If the variable is only a number, the size is assumed to be in megabytes.
- [HADOOP-14426](#) upgrades the version of Kerby from 1.0.0-RC2 to 1.0.0
- [HDFS-10970](#) updates the version of Jackson from 1.9.13 to 2.x in `hadoop-hdfs`.
- [HADOOP-9613](#) updates the Jersey version to the latest 1.x release.
- [HADOOP-10101](#) updates Guava dependency to 21.0
- [HADOOP-14225](#) removes the `xmlenc` dependency. If you rely on the transitive dependency, you need to set the dependency explicitly in your code after this change.
- [HADOOP-13382](#) remove unneeded `commons-httpclient` dependencies from POM files in Hadoop and sub-projects. This incompatible change may affect projects that have undeclared transitive dependencies on `commons-httpclient`, which used to be provided by `hadoop-common` or `hadoop-client`.
- [HADOOP-13660](#) upgrades the `commons-configuration` version from 1.6 to 2.1.
- [HADOOP-12064](#) upgrades the following dependencies:
 - Guice from 3.0 to 4.0
 - cglib from 2.2 to 3.2.0
 - asm from 3.2 to 5.0.4

MapReduce

CDH 6.2.0 introduces no new incompatible changes for MapReduce.

CDH 6.0.0, introduces the following incompatible changes:

- Support for MapReduce v1 has been dropped from CDH 6.0.0.
- CDH 6 supports applications compiled against CDH 5.7.0 and higher MapReduce frameworks. Make sure to not to include the CDH jars with your application by marking them as "provided" in the `pom.xml` file.

YARN

CDH 6.2.0 introduces no new incompatible changes for YARN.

CDH 6.1.0 introduces no new incompatible changes for YARN.

CDH 6.0.1 introduces no new incompatible changes for YARN.

CDH 6.0.0 introduces no new incompatible changes for YARN.

Apache HBase

CDH 6.2.x contains the following downstream HBase incompatible change:

hbase.security.authorization

The default value for `hbase.security.authorization` has been changed from true to false. Secured clusters should make sure to explicitly set it to true in XML configuration file before upgrading to one of these versions ([HBASE-19483](#)). True as the default value of `hbase.security.authorization` was changed because not all clusters need authorization. (History: [HBASE-13275](#)) Rather, only the clusters which need authorization should set this configuration as true.

Incompatible Changes

For more information about upstream incompatible changes, see the Apache Reference Guide [Incompatible Changes](#) and [Upgrade Paths](#).

CDH 6.1.0 introduces the following incompatible changes for HBase:

- [HBASE-20270](#): Error triggered command help is no longer available.
- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.

CDH 6.0.x introduces the following upstream HBase incompatible changes:

- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.
- Public interface API changes:
 - [HBASE-15607](#): Admin
 - [HBASE-19112](#), [HBASE-18945](#): Cell
 - Region, Store, HBaseTestingUtility
- [HBASE-18792](#): hbase-2 needs to defend against hbck operations
- [HBASE-15982](#): Interface ReplicationEndpoint extends Guava's Service.
- [HBASE-18995](#): Split CellUtil into public CellUtil and PrivateCellUtil for Internal use only.
- [HBASE-19179](#): Purged the hbase-prefix-tree module and all references from the code base.
- [HBASE-17595](#): Add partial result support for small/limited scan; Now small scan and limited scan could also return partial results.
- [HBASE-16765](#): New default split policy, SteppingSplitPolicy.
- [HBASE-17442](#): Move most of the replication related classes from hbase-client to hbase-replication package.
- [HBASE-16196](#): The bundled JRuby 1.6.8 has been updated to version 9.1.9.0.
- [HBASE-18811](#): Filters have been moved from Public to LimitedPrivate.
- [HBASE-18697](#): Replaced hbase-shaded-server jar with hbase-shaded-mapreduce jar.
- [HBASE-18640](#): Moved mapreduce related classes out of hbase-server into separate hbase-mapreduce jar .
- [HBASE-19128](#): Distributed Log Replay feature has been removed.
- [HBASE-19176](#): Hbase-native-client has been removed.
- [HBASE-17472](#): Changed semantics of granting new permissions. Earlier, new grants would override previous permissions, but now, the new and existing permissions get merged.
- [HBASE-18374](#): Previous "mutate" latency metrics has been renamed to "put" metrics.
- [HBASE-15740](#): Removed Replication metric source.shippedKBs in favor of source.shippedBytes.

- [HBASE-13849](#): Removed restore and clone snapshot from the WebUI.
- [HBASE-13252](#): The concept of managed connections in HBase (deprecated before) has now been extinguished completely, and now all callers are responsible for managing the lifecycle of connections they acquire.
- [HBASE-14045](#): Bumped thrift version to 0.9.2.
- [HBASE-5401](#): Changes to number of tasks PE runs when clients are mapreduce. Now tasks == client count. Previous we hardcoded ten tasks per client instance.

Changed Behavior

CDH 6.1.x contains the following HBase behavior changes:

- [HBASE-14350](#): Assignment Manager v2 - Split/Merge have moved to the Master; it runs them now. Hooks around Split/Merge are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-18271](#): Moved to internal shaded netty.
- [HBASE-17343](#): Default MemStore to be CompactingMemStore instead of DefaultMemStore. In-memory compaction of CompactingMemStore demonstrated sizable improvement in HBase's write amplification and read/write performance.
- [HBASE-19092](#): Make Tag IA.LimitedPrivate and expose for CPs.
- [HBASE-18137](#): Replication gets stuck for empty WALs.
- [HBASE-17513](#): Thrift Server 1 uses different QOP settings than RPC and Thrift Server 2 and can easily be misconfigured so there is no encryption when the operator expects it.
- [HBASE-16868](#): Add a replicate_all flag to replication peer config. The default value is true, which means all user tables (REPLICATION_SCOPE != 0) will be replicated to peer cluster.
- [HBASE-19341](#): Ensure Coprocessors can abort a server.
- [HBASE-18469](#): Correct RegionServer metric of totalRequestCount.
- [HBASE-17125](#): Marked Scan and Get's setMaxVersions() and setMaxVersions(int) as deprecated. They are easy to misunderstand with column family's max versions, so use readAllVersions() and readVersions(int) instead.
- [HBASE-16567](#): Core is now up on protobuf 3.1.0 (Coprocessor Endpoints and REST are still on protobuf 2.5.0).
- [HBASE-14004](#): Fix inconsistency between Memstore and WAL which may result in data in remote cluster that is not in the origin (Replication).
- [HBASE-18786](#): FileNotFoundException opening a StoreFile in a primary replica now causes a RegionServer to crash out where before it would be ignored (or optionally handled via close/reopen).
- [HBASE-17956](#): Raw scans will also read TTL expired cells.
- [HBASE-17017](#): Removed per-region latency histogram metrics.
- [HBASE-19483](#): Added ACL checks to RSGroup commands - On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands.
- [HBASE-19358](#): Added ACL checks to RSGroup commands (HBASE-19483): On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands. Improved stability of splitting log when do failover.
- [HBASE-18883](#): Updated our Curator version to 4.0 - Users who experience classpath issues due to version conflicts are recommended to use either the hbase-shaded-client or hbase-shaded-mapreduce artifacts.
- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.
- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.
- [HBASE-15968](#): This issue resolved two long-term issues in HBase: 1) Puts may be masked by a delete before them, and 2) Major compactions change query results. Offers a new behavior to fix this issue with a little performance reduction. Disabled by default. See the issue for details and caveats.
- [HBASE-13701](#): SecureBulkLoadEndpoint has been integrated into HBase core as default bulk load mechanism. It is no longer needed to install it as a coprocessor endpoint.
- [HBASE-9774](#): HBase native metrics and metric collection for coprocessors.
- [HBASE-18294](#): Reduce global heap pressure: flush based on heap occupancy.

Apache Hive/Hive on Spark/HCatalog

Apache Hive

The following changes are introduced to Hive in CDH 6.0.0, and are not backwards compatible:

- [UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting](#) on page 533
- [Support for UNION DISTINCT](#) on page 534
- [OFFLINE and NO_DROP Options Removed from Table and Partition DDL](#) on page 534
- [DESCRIBE Query Syntax Change](#) on page 535
- [CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names](#) on page 535
- [Reserved and Non-Reserved Keyword Changes in HiveQL](#) on page 535
- [Apache Hive API Changes in CDH 6.0.0](#) on page 537
- [Apache Hive Configuration Changes in CDH 6.0.0](#) on page 538
- [HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes](#) on page 540
- [Values Returned for Decimal Numbers Are Now Padded with Trailing Zeros to the Scale of the Specified Column](#) on page 541
- [Hive Logging Framework Switched to SLF4J/Log4j 2](#) on page 541
- [Deprecated Parquet Java Classes Removed from Hive](#) on page 541
- [Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms](#) on page 541
- [S3N Connector Is Removed from CDH 6.0](#) on page 542
- [Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request](#) on page 542
- [Support Added for Escaping Carriage Returns and New Line Characters for Text Files \(LazySimpleSerDe\)](#) on page 542
- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 542

Changing Table File Format from ORC with the ALTER TABLE Command Not Supported in CDH 6

Changing the table file format from ORC to another file format with the ALTER TABLE command is not supported in CDH 6 (it returns an error).

UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting

Prior to this change, Hive performed implicit casts when data types from different type groups were specified in queries that use UNION_ALL. For example, before CDH 6.0, if you had the two following tables:

Table "one"

one.col_1	one.col_2	one.col_3
21	hello_all	b

Where col_1 datatype is int, col_2 datatype is string, and col_3 datatype is char(1).

Table "two"

two.col_4	two.col_5	two.col_6
75.0	abcde	45

Where col_4 datatype is double, col_5 datatype is varchar(5), and col_6 datatype is int.

And you ran the following UNION_ALL query against these two tables:

```
SELECT * FROM one UNION ALL SELECT col_4 AS col_1, col_5 AS col_2, col_6 AS
```

```
col_3 FROM two;
```

You received the following result set:

_u1.col_1	_u1.col_2	_u1.col_3
75.0	abcde	4
21.0	hello	b

Note that this statement implicitly casts the values from table `one` with the following errors resulting in data loss:

- `one.col_1` is cast to a `double` datatype
- `one.col_2` is cast to a `varchar(5)` datatype, which truncates the original value from `hello_all` to `hello`
- `one.col_3` is cast to a `char(1)` datatype, which truncates the original value from 45 to 4

In CDH 6.0, no implicit cast is performed across different type groups. For example, `STRING`, `CHAR`, and `VARCHAR` are in one type group, and `INT`, `BIGINT`, and `DECIMAL` are in another type group, and so on. So, in CDH 6.0 and later, the above query that uses `UNION ALL`, returns an exception for the columns that contain datatypes that are not part of a type group. In CDH 6.0 and later, Hive performs the implicit cast only *within* type groups and not *across* different type groups. For more information, see [HIVE-14251](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
 - The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
 - The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

OFFLINE and NO_DROP Options Removed from Table and Partition DDL

Support for Hive table and partition protection options have been removed in CDH 6.0, which includes removal of the following functionality:

- Support has been removed for:

- `ENABLE | DISABLE NO_DROP [CASCADE]`
- `ENABLE | DISABLE OFFLINE`
- `ALTER TABLE ... IGNORE PROTECTION`

- The following support has also been removed from the `HiveMetastoreClient` class:

The `ignoreProtection` parameter has been removed from the `dropPartitions` methods in the `IMetaStoreClient` interface.

For more information, see [HIVE-1145](#).

Cloudera recommends that you use Apache Sentry to replace most of this functionality. Although Sentry governs permissions on `ALTER TABLE`, it does not include permissions that are specific to a partition. See [Authorization Privilege Model for Hive and Impala](#) and [Configuring the Sentry Service](#).

DESCRIBE Query Syntax Change

In CDH 6.0 syntax has changed for `DESCRIBE` queries as follows:

- `DESCRIBE` queries where the column name is separated by the table name using a period is no longer supported:

```
DESCRIBE testTable.testColumn;
```

Instead, the table name and column name must be separated with a space:

```
DESCRIBE testTable testColumn;
```

- The `partition_spec` must appear *after* the table name, but *before* the optional column name:

```
DESCRIBE default.testTable PARTITION (part_col = 100) testColumn;
```

For more details, see the [Apache wiki](#) and [HIVE-12184](#).

CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names

In CDH 6.0, `CREATE TABLE` statements fail if any of the specified column names contain a period or a colon. For more information, see [HIVE-10120](#) and the [Apache wiki](#).

Reserved and Non-Reserved Keyword Changes in HiveQL

Hive reserved and non-reserved keywords have changed in CDH 6.0. *Reserved keywords* cannot be used as table or column names unless they are enclosed with back ticks (for example, `data`). *Non-reserved keywords* can be used as table or column names without enclosing them with back ticks. Non-reserved keywords have proscribed meanings in HiveQL, but can still be used as table or column names. For more information about the changes to reserved and non-reserved words listed below, see [HIVE-6617](#) and [HIVE-14872](#).

In CDH 6.0, the following changes have been introduced to Hive reserved and non-reserved keywords and are not backwards compatible:

- [Hive New Reserved Keywords Added in CDH 6.0](#) on page 535
- [Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0](#) on page 536
- [Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0](#) on page 536
- [Hive New Non-Reserved Keywords Added in CDH 6.0](#) on page 536
- [Hive Non-Reserved Keyword Removed in CDH 6.0](#) on page 537

Hive New Reserved Keywords Added in CDH 6.0

The following table contains new reserved keywords that have been added:

COMMIT	CONSTRAINT	DEC	EXCEPT
FOREIGN	INTERVAL	MERGE	NUMERIC
ONLY	PRIMARY	REFERENCES	ROLLBACK

START

Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0

The following table contains non-reserved keywords that have been converted to be reserved keywords:

ALL	ALTER	ARRAY	AS
AUTHORIZATION	BETWEEN	BIGINT	BINARY
BOOLEAN	BOTH	BY	CREATE
CUBE	CURSOR	DATE	DECIMAL
DOUBLE	DELETE	DESCRIBE	DROP
EXISTS	EXTERNAL	FALSE	FETCH
FLOAT	FOR	FULL	GRANT
GROUP	GROUPING	IMPORT	IN
INT	INNER	INSERT	INTERSECT
INTO	IS	LATERAL	LEFT
LIKE	LOCAL	NONE	NULL
OF	ORDER	OUT	OUTER
PARTITION	PERCENT	PROCEDURE	RANGE
READS	REGEXP	REVOKE	RIGHT
RLIKE	ROLLUP	ROW	ROWS
SET	SMALLINT	TABLE	TIMESTAMP
TO	TRIGGER	TRUNCATE	UNION
UPDATE	USER	USING	VALUES
WITH	TRUE		

Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0

The following table contains reserved keywords that have been converted to be non-reserved keywords:

CURRENT_DATE	CURRENT_TIMESTAMP	HOLD_DDLTIME	IGNORE
NO_DROP	OFFLINE	PROTECTION	READONLY

Hive New Non-Reserved Keywords Added in CDH 6.0

The following table contains new non-reserved keywords that have been added:

ABORT	AUTOCOMMIT	CACHE	DAY
DAYOFWEEK	DAYS	DETAIL	DUMP
EXPRESSION	HOUR	HOURS	ISOLATION
KEY	LAST	LEVEL	MATCHED
MINUTE	MINUTES	MONTH	MONTHS
NORELY	NOVALIDATE	NULLS	OFFSET
OPERATOR	RELY	SECOND	SECONDS

SNAPSHOT	STATUS	SUMMARY	TRANSACTION
VALIDATE	VECTORIZATION	VIEWS	WAIT
WORK	WRITE	YEAR	YEARS

Hive Non-Reserved Keyword Removed in CDH 6.0

The following non-reserved keyword has been removed:

DEFAULT

Apache Hive API Changes in CDH 6.0.0

The following changes have been introduced to the Hive API in CDH 6.0, and are not backwards compatible:

- [AddPartitionMessage.getPartitions\(\) Can Return NULL](#) on page 537
- [DropPartitionEvent and PreDropPartitionEvent Class Changes](#) on page 537
- [GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date](#) on page 537
- [GenericUDF.getConstantLongValue Has Been Removed](#) on page 537
- [Increased Width of Hive Metastore Configuration Columns](#) on page 537

AddPartitionMessage.getPartitions() Can Return NULL

The `getPartitions()` method has been removed from the `AddPartitionEvent` class in the `org.apache.hadoop.hive.metastore.events` interface. It was removed to prevent out-of-memory errors when the list of partitions is too large.

Instead use the `getPartitionIterator()` method. For more information, see [HIVE-9609](#) and the [AddPartitionEvent documentation](#).

DropPartitionEvent and PreDropPartitionEvent Class Changes

The `getPartitions()` method has been removed and replaced by the `getPartitionIterator()` method in the `DropPartitionEvent` class and the `PreDropPartitionEvent` class.

In addition, the `(Partition partition, boolean deleteData, HiveMetastore.HMSHandler handler)` constructors have been deleted from the `PreDropPartitionEvent` class. For more information, see [HIVE-9674](#) and the [PreDropPartitionEvent documentation](#).

GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date

The `getTimestampValue` method in the `GenericUDF` class now returns a `TIMESTAMP` value instead of a `DATE` value. For more information, see [HIVE-10275](#) and the [GenericUDF documentation](#).

GenericUDF.getConstantLongValue Has Been Removed

The `getConstantLongValue` method has been removed from the `GenericUDF` class. It has been noted by the community that this method is not used in Hive. For more information, see [HIVE-10710](#) and the [GenericUDF documentation](#).

Increased Width of Hive Metastore Configuration Columns

The columns used for configuration values in the Hive metastore have been increased in width, resulting in the following incompatible changes in the `org.apache.hadoop.hive.metastore.api` interface.

This change introduced an incompatible change to the `get_table_names_by_filter` method of the `ThriftHiveMetastore` class. Before this change, this method accepts a string filter, which allows clients to filter a table by its `TABLEPROPERTIES` value. For example:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 <> \"yellow\";

org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 <> \"yellow\";
```

```
PARAMS + "test_param_1 = \"yellow\" ;
```

After this change, the `TABLE_PARAMS.PARAM_VALUE` column is now a CLOB data type. Depending on the type of database that you use (for example, MySQL, Oracle, or PostgreSQL), the semantics may have changed and operators like "`=`", "`<>`", and "`!=`" might not be supported. Refer to the documentation for your database for more information. You must use operators that are compatible with CLOB data types. There is no equivalent "`<>`" operator that is compatible with CLOB. So there is no equivalent operator for the above example that uses the "`<>`" inequality operator. The equivalent for "`=`" is the `LIKE` operator so you would rewrite the second example above as:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 LIKE \"yellow\" ;
```

For more information, see [HIVE-12274](#).

Apache Hive Configuration Changes in CDH 6.0.0

The following configuration property changes have been introduced to Hive in CDH 6.0, and are not backwards compatible:

- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 538
- [Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters](#) on page 538
- [Hive Strict Checks Have Been Re-factored To Be More Granular](#) on page 539
- [Java XML Serialization Has Been Removed](#) on page 539
- [Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties](#) on page 540
- [HiveServer2 Impersonation Property \(`hive.server2.enable.impersonation`\) Removed](#) on page 540
- [Changed Default File Format for Storing Intermediate Query Results](#) on page 540

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on `hive.enforce.bucketing`](#) and the [topic on `hive.enforce.sorting`](#).

Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters

Directories in HDFS can contain unprintable or unsupported characters that are not visible even when you run the `hadoop fs -ls` command on the directories. When external tables are created with the `MSCK REPAIR TABLE` command, the partitions using these HDFS directories that contain unsupported characters are unusable for Hive. To avoid this, the configuration parameter `hive.msck.path.validation` has been added. This configuration property controls the behavior of the `MSCK REPAIR TABLE` command, enabling you to set whether validation checks are run on the HDFS directories when `MSCK REPAIR TABLE` is run.

The property `hive.msck.path.validation` can be set to one of the following values:

Value Name	Description
throw	Causes Hive to throw an exception when it tries to process an HDFS directory that contains unsupported characters with the <code>MSCK REPAIR TABLE</code> command. This is the default setting for <code>hive.msck.path.validation</code> .
skip	Causes Hive to skip the skip the directories that contain unsupported characters, but still repairs the others.
ignore	Causes Hive to completely skip any validation of HDFS directories when the <code>MSCK REPAIR TABLE</code> command is

Value Name	Description
	run. This setting can cause bugs because unusable partitions are created.

By default, the `hive.msck.path.validation` property is set to `throw`, which causes Hive to throw an exception when MSCK REPAIR TABLE is run and HDFS directories containing unsupported characters are encountered. To work around this, set this property to `skip` until you can repair the HDFS directories that contain unsupported characters.

To set this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the **Configuration** tab.
3. Search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting.
4. In the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting, add the **Name** of the property, the **Value** (`throw`, `skip`, or `ignore`), and a **Description** of the setting.
5. Click **Save Changes** and restart the service.

For more information, see [HIVE-10722](#).

Hive Strict Checks Have Been Re-factored To Be More Granular

Originally, the configuration property `hive.mapred.mode` was added to restrict certain types of queries from running. Now it has been broken down into more fine-grained configurations, one for each type of restricted query pattern. The configuration property `hive.mapred.mode` has been removed and replaced with the following configuration properties, which provide more granular control of Hive strict checks:

Configuration Property	Description	Default Value
<code>hive.strict.checks.bucketing</code>	When set to <code>true</code> , running <code>LOAD DATA</code> queries against bucketed tables is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.type.safety</code>	When set to <code>true</code> , comparing <code>bigint</code> to <code>string</code> data types or <code>bigint</code> to <code>double</code> data types is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.orderby.no.limit</code>	When set to <code>true</code> , prevents queries from being run that contain an <code>ORDER BY</code> clause with no <code>LIMIT</code> clause.	<code>false</code>
<code>hive.strict.checks.no.partition.filter</code>	When set to <code>true</code> , prevents queries from being run that scan a partitioned table but do not filter on the partition column.	<code>false</code>
<code>hive.strict.checks.cartesian.product</code>	When set to <code>true</code> , prevents queries from being run that contain a Cartesian product (also known as a cross join).	<code>false</code>

All of these properties can be set with Cloudera Manager in the following configuration settings for the Hive service:

- **Restrict LOAD Queries Against Bucketed Tables** (`hive.strict.checks.bucketing`)
- **Restrict Unsafe Data Type Comparisons** (`hive.strict.checks.type.safety`)
- **Restrict Queries with ORDER BY but no LIMIT clause** (`hive.strict.checks.orderby.no.limit`)
- **Restrict Partitioned Table Scans with no Partitioned Column Filter** (`hive.strict.checks.no.partition.filter`)
- **Restrict Cross Joins (Cartesian Products)** (`hive.strict.checks.cartesian.product`)

For more information about these configuration properties, see [HIVE-12727](#), [HIVE-15148](#), [HIVE-18251](#), and [HIVE-18552](#).

Java XML Serialization Has Been Removed

The configuration property `hive.plan.serialization.format` has been removed. Previously, this configuration property could be set to either `javaXML` or `kryo`. Now the default is `kryo` serialization, which cannot be changed. For more information, see [HIVE-12609](#) and the [Apache wiki](#).

Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties

The configuration property `hive.groupby.orderby.position.alias`, which enabled using column position with the `GROUP BY` and the `ORDER BY` clauses has been removed and replaced with the following two configuration properties. These configuration properties enable using column position with `GROUP BY` and `ORDER BY` separately:

Configuration Property Name	Description/Default Setting	Possible Values
<code>hive.groupby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>GROUP BY</code> clauses in queries. Default Setting: <code>false</code> . This behavior is turned off by default.	<code>true</code> <code>false</code>
<code>hive.orderby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>ORDER BY</code> clauses in queries. Default Setting: <code>true</code> . This behavior is turned on by default.	<code>true</code> <code>false</code>

For more information, see [HIVE-15797](#) and the Apache wiki entries for [configuration properties](#), [GROUP BY syntax](#), and [ORDER BY syntax](#).

HiveServer2 Impersonation Property (`hive.server2.enable.impersonation`) Removed

In earlier versions of CDH, the following two configuration properties could be used to set impersonation for HiveServer2:

- `hive.server2.enable.impersonation`
- `hive.server2.enable.doAs`

In CDH 6.0, `hive.server2.enable.impersonation` is removed. To configure impersonation for HiveServer2, use the configuration property `hive.server2.enable.doAs`. To set this property in Cloudera Manager, select the Hive service and click on the **Configuration** tab. Then search for the **HiveServer2 Enable Impersonation** setting and select the checkbox to enable HiveServer2 impersonation. This property is enabled by default in CDH 6.

For more information about this property, see the [Apache wiki documentation for HiveServer2 configuration properties](#).

Changed Default File Format for Storing Intermediate Query Results

The configuration property `hive.query.result.fileformat` controls the file format in which a query's intermediate results are stored. In CDH 6, the default setting for this property has been changed from `TextFile` to `SequenceFile`.

To change this configuration property in Cloudera Manager:

1. In the Admin Console, select the Hive service and click on the **Configuration** tab.
2. Then search for the **Hive Service Advanced Configuration Snippet (Safety Valve) for `hive-site.xml`** setting and add the following information:
 - **Name:** `hive.query.result.fileformat`
 - **Value:** Valid values are `TextFile`, `SequenceFile` (default), or `RCfile`
 - **Description:** Sets the file format in which a query's intermediate results are stored.
3. After you add this information, click **Save Changes** and restart the Hive service.

For more information about this parameter, see the [Apache wiki](#).

HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes

HiveServer2 Thrift API code has been repackaged in CDH 6.0, resulting in the following changes:

- All files generated by the Thrift API for HiveServer2 have moved from the following *old* namespace:

`org.apache.hive.service.cli.thrift`

To the following *new* namespace:

`org.apache.hive.service.rpc.thrift`

- All files generated by the Thrift API for HiveServer2 have moved into a separate jar file called `service-rpc`.

As a result of these changes, all Java classes such as `TCLIService.java`, `TOpenSessionReq.java`, `TSessionHandle.java`, and `TGetSchemasReq.java` have changed locations. For more information, see [HIVE-12442](#).

Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column

Decimal values that are returned in query results are now padded with trailing zeroes to match the specified scale of the corresponding column. For example, *before* this change, when Hive read a decimal column with a specified scale of 5, the value returned for zero was returned as 0. *Now*, the value returned for zero is 0.00000. For more information, see [HIVE-12063](#).

Hive Logging Framework Switched to SLF4J/Log4j 2

The logging framework for Hive has switched to [SLF4J \(Simple Logging Facade for Java\)](#) and now uses [Log4j 2](#) by default. Use of Log4j 1.x, Apache Commons Logging, and `java.util.logging` have been removed. To accommodate this change, write all Log4j configuration files to be compatible with Log4j 2.

For more information, see [HIVE-12237](#), [HIVE-11304](#), and the [Apache wiki](#).

Deprecated Parquet Java Classes Removed from Hive

The deprecated parquet classes, `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` have been removed from Hive because they resided outside of the `org.apache` namespace. Any existing tables that use these classes are automatically migrated to the new SerDe classes when the metastore is upgraded.

Use one of the following options for specifying the Parquet SerDe for new Hive tables:

- Specify in the `CREATE TABLE` statement that you want it stored as Parquet. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING) STORED AS PARQUET;
```

- Set the `INPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat` and set the `OUTPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat`. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING)
STORED AS
  INPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat"
  OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat";
```

For more information, see [HIVE-6757](#) and the [Apache wiki](#).

Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms

Support for JDBC, counter-based, and HBase-based statistics collection mechanisms has been removed from Hive. The following configuration properties are no longer supported:

- `hive.stats.dbclass`
- `hive.stats.retries.wait`
- `hive.stats.retries.max`
- `hive.stats.jdbc.timeout`

- `hive.stats.dbconnectionstring`
- `hive.stats.jdbcdrive`
- `hive.stats.key.prefix.reserve.length`

This change also removed the `cleanUp(String keyPrefix)` method from the [StatsAggregator interface](#).

Now all Hive statistics are collected on the default file system. For more information, see [HIVE-12164](#), [HIVE-12411](#), [HIVE-12005](#), and the [Apache wiki](#).

S3N Connector Is Removed from CDH 6.0

The [S3N connector](#), which is used to connect to the Amazon S3 file system from Hive has been removed from CDH 6.0. To connect to the S3 file system from Hive in CDH 6.0, you must now use the S3A connector. There are a number of differences between the S3N and the S3A connectors, including configuration differences. See the [Apache wiki page on integrating with Amazon Web Services](#) for details.

Migration involves making the following changes:

- Changing all metastore data containing URIs that start with `s3n://` to `s3a://`. This change is performed automatically when you upgrade the Hive metastore.
- Changing all scripts containing URIs that start with `s3n://` to `s3a://`. You must perform this change manually.

Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request

Six additional columns have been added to the `TRowSet` that is returned by the `TCLIService#GetTables` request. These columns were added to comply with the official JDBC API. For more information, see the documentation for [java.sql.DatabaseMetaData](#).

The columns added are:

Column Name	Description
REMARKS	Explanatory comment on the table.
TYPE_CAT	Types catalog.
TYPE_SCHEMA	Types schema.
TYPE_NAME	Types name.
SELF_REFERENCING_COL_NAME	Name of the designed identifier column of a typed table.
REF_GENERATION	Specifies how values in the <code>SELF_REFERENCING_COL_NAME</code> column are created.

For more information, see [HIVE-7575](#).

Support Added for Escaping Carriage Returns and New Line Characters for Text Files (LazySimpleSerDe)

Support has been added for escaping carriage returns and new line characters in text files by modifying the `LazySimpleSerDe` class. Without this change, carriage returns and new line characters are interpreted as delimiters, which causes incorrect query results.

This feature is controlled by the SerDe property `serialization.escape.crlf`. It is enabled (set to `true`) by default. If `serialization.escape.crlf` is enabled, '`r`' or '`n`' cannot be used as separators or field delimiters.

This change only affects text files and removes the `getNullString` method from the [LazySerDeParameters class](#). For more information, see [HIVE-11785](#).

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to `false`, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to `true`. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on hive.enforce.bucketing](#) and the [topic on hive.enforce.sorting](#).

Hue

There are no incompatible changes in this release.

Apache Impala

Port Change for SHUTDOWN Command

The SHUTDOWN command for shutting down a remote server used the backend port in CDH 6.1. Starting in CDH 6.2, the command uses the KRPC port, e.g. :shutdown('host100:27000').

Apache Kafka

Incompatible Changes Introduced in CDH 6.2.0



Warning: Kafka in CDH 6.2.0 is based on Apache Kafka 2.1.0, which contains a change to the internal schema used to store consumer offsets. As a result of this change, downgrading Kafka to a version lower than CDH 6.2.0 is **NOT** possible once Kafka has been upgraded to CDH 6.2.0 or higher.

Default Behaviour Changes

Kafka CDH 6.2.0. Introduces the following default behaviour changes:

- Unclean leader election is automatically enabled by the controller when `unclean.leader.election.enable` config is dynamically updated by using per-topic config override.
- Diagnostic data bundles collected by Cloudera Manager from now on include information on Kafka topics. The bundle includes the information exposed by the following two commands:
 - `kafka-topics --describe`
 - `kafka-topics --list`

Incompatible Changes Introduced in CDH 6.1.1

CDH 6.1.1 introduces no new incompatible changes for Kafka.

Incompatible Changes Introduced in CDH 6.1.0

Scala-based Client API Removed

Scala-based clients were deprecated in a previous release and are removed as of CDH 6.1.0.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are effected:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Default Behaviour Change

KAFKA-7050: The default value for `request.timeout.ms` is decreased to 30 seconds. In addition, a new logic is added that makes the `JoinGroup` requests ignore this timeout.

Incompatible Changes Introduced in CDH 6.0.1

CDH 6.0.1 introduces no new incompatible changes for Kafka.

Incompatible Changes Introduced in CDH 6.0.0

Kafka is now bundled as part of CDH. The following sections describe incompatible changes between the previous, separately installed Kafka (CDK powered by Apache Kafka version 3.1) and the CDH 6.0.0 Kafka version. These changes affect clients built with CDH 6.0.0 libraries. Cloudera recommends upgrading clients to the new release; however clients built with previous versions of Kafka will continue to function.

Packaging

CDH and previous distributions of Kafka (CDK Powered by Apache Kafka) cannot coexist in the same cluster.

Deprecated Scala-based Client API and New Java Client API

Scala-based clients are deprecated in this release and will be removed in an upcoming release.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are deprecated and unsupported as of CDH 6.0.0:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Command Line Options Removed

Some command line tools are affected by the deprecation of old clients (see the previous entry). The following options have been removed and are not recognized as valid options:

- `--new-consumer`
- `--old-consumer`
- `--old-producer`

The tools affected use the new clients.

Command Line Tools Removed

The following command line tools and runnable classes are removed:

- `kafka-replay-log-producer`
- `kafka-simple-consumer-shell`
- `kafka.tools.ReplayLogProducer`
- `kafka.tools.SimpleConsumerShell`
- `kafka.tools.ExportZkOffset`
- `kafka.tools.ImportZkOffset`
- `kafka.tools.SimpleConsumerPerformance`
- `kafka.tools.UpdateOffsetsInZK`
- `kafka.tools.VerifyConsumerRebalance`
- `kafka.tools.ProducerPerformance`

Consumer API Changes

Consumer methods invoked with unassigned partitions now raise an `IllegalStateException` instead of an `IllegalArgumentException`.

Previous versions of the Consumer method `poll(long)` would wait for metadata updates regardless of timeout parameter. This behavior is expected to change in future releases; make sure your client applications include an appropriate timeout parameter and do not rely on the previous behavior.

Exception Classes Removed

The following exceptions were deprecated in a previous release and are not thrown anymore are removed:

- `GroupCoordinatorNotAvailableException`
- `GroupLoadInProgressException`
- `NotCoordinatorForGroupException`
- `kafka.common.KafkaStorageException`

Metrics Updated

Kafka consumers' per-partition metrics were changed to use tags for topic and partition rather than the metric name. For more information see [KIP-225](#).

Apache Kudu

There are no incompatible changes in this release.

Apache Oozie

There are no incompatible changes in this release.

Apache Parquet

Packages and Group ID Renamed

As a part of the Apache incubation process, all Parquet packages and the project's group ID were renamed as follows:

Parquet Version	1.6.0 and lower (CDH 5.x)	1.7.0 and higher (CDH 6.x)
Java Package Names	parquet.*	org.apache.parquet.*
Group ID	com.twitter	org.apache.parquet

If you directly consume the Parquet API, instead of using Parquet through Hive, Impala or other CDH component, you need to update your code to reflect these changes:

Update *.java files:

Before	After
import parquet.*;	import org.apache.parquet.*;

Update pom.xml:

Before	After
<pre><dependency> <groupId> com.twitter </groupId> <version> \${parquet.version} </version> </dependency></pre>	<pre><dependency> <groupId> org.apache.parquet </groupId> <version> \${parquet.version} </version> </dependency></pre>

API Methods Removed

In Parquet 1.6, a number of API methods were removed from the `parquet.hadoop.ParquetInputSplit` class that depended on reading metadata on the client side. Metadata should be read on the task side instead.

Removed Method	New Method to Use
<code>parquet.hadoop.ParquetInputSplit.getFileSchema</code>	<code>org.apache.parquet.hadoop.api.InitContext.getFileSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getRequestedSchema</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getRequestedSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getReadSupportMetadata</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getReadSupportMetadata</code>
<code>parquet.hadoop.ParquetInputSplit.getBlocks</code>	<code>org.apache.parquet.hadoop.metadata.ParquetMetadata.getBlocks</code>
<code>parquet.hadoop.ParquetInputSplit.getExtraMetadata</code>	-

Apache Pig

The following change is introduced to Pig in CDH 6.0 and is not a backwards compatible change. You must modify your Pig scripts as described below.

Removal of the Apache DataFu Pig JAR from CDH 6

Apache DataFu Pig is a collection of user-defined functions that can be used with Pig for data mining and statistical analysis on large-scale data. The DataFu JAR was included in CDH 5, but due to very low adoption rates, the JAR was deprecated in CDH 5.9 and is being removed from CDH 6, starting with CDH 6.0. It is no longer supported.

Recommended Migration Strategy

A simple way to assess what DataFu functions you are using in your Pig scripts is to use the `grep` utility to search for occurrences of "datafu" in your code. When DataFu functions are used in Pig scripts, you must use a function definition entry that contains "datafu" like the following example:

```
define <function_name> datafu.pig... .<class_name>();
```

Use `grep` to search for the string "datafu" in your scripts and that will identify where the DataFu JAR is used.

Cloudera recommends migrating to Hive UDFs or operators wherever it is possible. However, if there are cases where it is impossible to replace DataFu functions with Hive functions, download the upstream version of the DataFu Pig libraries and place them on the node where the Pig front end is used. To preserve compatibility, use the version 1.1.0 JAR, which was the version included in CDH 5. You can download the JAR file [here](#). However, Cloudera does not support using this upstream DataFu JAR file.

Mapping DataFu UDFs to Hive UDFs

The following Hive UDFs map to DataFu UDFs and can be used instead in Pig scripts with the caveats that are listed:

Table 30: Hive Functions That Map to DataFu Functions

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
MD5 (hash)	Computes the MD5 value of a string and outputs a hex value by default.	md5	None
SHA (hash)	Computes the SHA value of a string and outputs a hex value by default.	sha/sha1	None
RandInt (random)	Generates a uniformly distributed integer between two bounds.	rand	None
VAR (stats)	Generates the variance of a set of values.	variance	None
Median (stats)	Computes the median for a sorted input bag. A special case of the Quantile function.	percentile(0.5)	See Limitations When Substituting Quantile and Median DataFu Functions on page 221.
Quantile (stats)	Computes quantiles for a sorted input bag.	percentile	See Limitations When Substituting Quantile and Median DataFu Functions on page 221.

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
Coalesce (util)	Returns the first non-null value from a tuple, like COALESCE in SQL.	coalesce	None
InUDF (util)	Similar to the SQL IN function, this function provides a convenient way to filter using a logical disjunction over many values.	IN	None

For more information about using Hive UDFs, see https://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_hive_udf.html.

Limitations When Substituting Quantile and Median DataFu Functions

With the exception of Median and Quantile all Hive functions specified in the above table should work as expected in Pig scripts. Median extends Quantile in DataFu functions and the equivalent Hive functions have a similar relationship. However, there is an important difference in how you use percentile and how you use Quantile. The differences are summarized in the following table:

Table 31: Differences Between Usage of DataFu 'Quantile' and Hive 'percentile'

Function:	Quantile	percentile
Input must be sorted:	Yes	No
Nulls are allowed in input:	No	Yes
Examples:	<pre> register datafu-1.2.0.jar; define median datafu.pig.stats.Quantile('0.5'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B = group A by name; C = foreach B { sorted = order A by n; generate group, flatten (median(sorted.n)); } </pre>	<pre> define percentile HiveUDAF ('percentile'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B2 = foreach A generate name, n, 0.5 as perc; C2 = GROUP B2 by name; D2 = FOREACH C2 generate group, percentile (B2. (n, perc)); </pre>

Although DataFu StreamingQuantile and StreamingMedian might appear to match Hive's percentile_approx function, Pig cannot consume percentile_approx.

DataFu Functions with No Hive Function or Operator Equivalent

The following general limitations apply when mapping DataFu UDFs to Hive UDFs:

- Many DataFu functions operate on a custom Pig data structure called a *bag*. No Hive UDFs can operate on Pig bags, so there are no equivalents for these DataFu functions.
- Some DataFu functions are custom functions that do not have Hive UDF equivalents. For example, the DataFu functions that calculate geographic distances, run the PageRank algorithm, or that do sampling. There are no equivalent Hive UDFs for these DataFu functions either.

Table 32: DataFu Functions with No Hive UDF Equivalent

AppendToBag (bags)	AssertUDF (util)	BagConcat (bags)
BagGroup (bags)	BagLeftOuterJoin (bags)	BagSplit (bags)
BoolToInt (util)	CountEach (bags)	DistinctBy (bags)
EmptyBagToNull (bags)	EmptyBagToNullFields (bags)	Enumerate (bags)
FirstTupleFromBag (bags)	HaversineDistInMiles (geo)	IntToBool (util)
MarkovPairs	NullToEmptyBag (bags)	PageRank (linkanalysis)
PrependToBag (bags)	ReservoirSample (sampling)*	ReverseEnumerate (bags)
SampleByKey (sampling)*	SessionCount (sessions)	Sessionize (sessions)
SetIntersect (sets)	SetUnion (sets)	SimpleRandomSample (sampling)*
TransposeTupleToBag (util)	UnorderedPairs (bags)	UserAgentClassify (urls)
WeightedSample (sampling)*	WilsonBinConf (stats)	—

* These DataFu functions might be replaced with TABLESAMPLE in HiveQL. See the [Apache Hive wiki](#).

Cloudera Search

The following changes are introduced in CDH 6.1

In CDH 6.1 Cloudera Search is rebased on Apache Solr 7.4.

Deprecations

- Enabling/disabling autoAddReplicas cluster-wide with the API is deprecated. Use suspend/resume trigger APIs with name=".auto_add_replicas" instead.
- In the ReplicationHandler, the master.commitReserveDuration sub-element is deprecated. Configure a direct commitReserveDuration element instead for use in all modes (leader, follower, cloud).

Removals

- The old Leader-In-Recovery implementation (implemented in Solr 4.9) has been removed and replaced. Solr supports rolling upgrades from old 7.x versions of Solr to future 7.x releases until the last release of the 7.x major version. This means that to upgrade to Solr 8, you will have to be on Solr 7.3 or higher.
- The throttling mechanism used to limit the rate of processed autoscaling events has been removed. This deprecates the actionThrottlePeriodSeconds setting in the set-properties command of Autoscaling API. Use the triggerCooldownPeriodSeconds parameter to pause event processing.
- The RunExecutableListener event listener was removed for security reasons. If you want to listen to events caused by updates and commits, or you want to optimize, write your own listener as native Java class as part of a Solr plugin.

For more information see the [Apache Solr 7.4 Release Notes](#).

The following changes are introduced in CDH 6.0

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has many incompatibilities with the 4.10 version of Apache Solr used in recent CDH 5 releases, such as the following:

- Solr 7 uses a managed schema by default. Generating an instance directory no longer generates schema.xml. For instructions on switching to a managed schema, see [Switching from schema.xml to Managed Schema](#) in *Apache Solr Reference Guide*.
- Creating a collection using solrctl collection --create without specifying the -c <configName> parameter now uses a default configuration set (named _default) instead of a configuration set with the same name as the collection. To avoid this, always specify the -c <configName> parameter when creating new collections.

For the full list of changes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

Apache Sentry contains the following incompatible change in CDH 6.0.0:

- Sentry no longer supports policy file authorization. You must migrate policy files to the database-backed Sentry service before you upgrade to CDH 6.0.0 unless you are using Sentry policy files for Solr. If you are using Sentry policy files for Solr, you must migrate to the database-backed Sentry service after you upgrade.

For information about migrating policy files before you upgrade, see [Migrating from Sentry Policy Files to the Sentry Service](#). For information about migrating policy files for Solr after you upgrade, see [Migrating Sentry Privileges for Solr After Upgrading to CDH 6](#).

Apache Spark

The following sections describe changes in Spark support in CDH 6 that might require special handling during upgrades, or code changes within existing applications.

- All Spark applications built against Spark 1.6 in CDH 5 must be rebuilt against Spark 2.x in CDH 6.
- Spark 2 in CDH 6 works with Java 8, not Java 7. If this change produces any Java code incompatibilities, update your Java code and rebuild the application.
- Spark 2 in CDH 6 works with Scala 2.11, not Scala 2.10. If this change produces any Scala code incompatibilities, update your Scala code and rebuild the application.
- `HiveContext` and `SQLContext` have been removed, although those variables still work for backward compatibility. Use the `SparkSession` object to replace both of these handles.
- `DataFrames` have been removed from the Scala API. `DataFrame` is now a special case of `Dataset`.

Since compile-time type-safety in Python and R is not a language feature, the concept of `Dataset` does not apply to these languages' APIs. Instead, `DataFrame` remains the primary programming abstraction.

- Spark 2.0 and higher do not use an assembly JAR for standalone applications.
- If you have event logs created in CDH 5.3 or lower, you cannot read those logs using Spark in CDH 6.0 or higher.

Apache Sqoop

CDH 6.2.0 introduces no new incompatible changes for Apache Sqoop.

The following changes are introduced in CDH 6.0, and are not backwards compatible:

- All classes in `com.cloudera.sqoop` packages have been removed in CDH 6.0. Use the corresponding classes from `org.apache.sqoop` packages. For example, use `org.apache.sqoop.SqoopOptions` instead of `com.cloudera.sqoop.SqoopOptions`.



Note: This change only affects customers who build their own application on top of Sqoop classes. Sqoop CLI users are not affected.

- Because of changes introduced in the Sqoop metastore logic, the metastore database created by Sqoop CDH 6 cannot be used by earlier versions. The metastore database created by Sqoop CDH 5 can be used by both Sqoop CDH 5 and Sqoop CDH 6.

Require an explicit option to be specified with `--split-by` for a String column

Using the --split-by option with a CHAR or VARCHAR column does not always work properly, so Sqoop now requires the user to set the `org.apache.sqoop.splitter.allow_text_splitter` property to true to confirm that they are aware of this risk.

Example:

```
sqoop import -Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect $MYCONN  
--username $MYUSER --password $MYPWD --table "test_table" --split-by "string_column"
```

For more information, see [SQOOP-2910](#).

Make Sqoop fail if user specifies --direct connector when it is not available

The --direct option is only supported with the following databases: MySQL, PostgreSQL, Oracle, Netezza.

In earlier releases, Sqoop silently ignored this option if it was specified for other databases, but it now throws an error.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table  
"direct_import" --direct
```

The command fails with the following error message:

```
Was called with the --direct option, but no direct connector available.
```

For more information, see [SQOOP-2913](#).

Sqoop does not allow --as-parquetfile with hcatalog jobs or when hive import with create-hive-table is used

The --create-hive-table option is not supported when the user imports into Hive in Parquet format. Earlier this option was silently ignored, and the data was imported even if the Hive table existed. Sqoop will now fail if the --create-hive-table option is used with the --as-parquetfile option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table  
"test_table" --hive-import --as-parquetfile --create-hive-table
```

The command fails with the following error message:

```
Hive import and create hive table is not compatible with importing into ParquetFile  
format.
```

For more information, see [SQOOP-3010](#).

Create fail fast for export with --hcatalog-table <HIVE_VIEW>

Importing into and exporting from a Hive view using HCatalog is not supported by Sqoop. A fail fast check was introduced so that now Sqoop throws a descriptive error message if the user specified a Hive view in the value of the --hcatalog-table option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table"  
--hcatalog-table "test_view"
```

The command fails with the following error message:

```
Reads/Writes from and to Views are not supported by HCatalog
```

For more information, see [SQOOP-3027](#).

Simplify Unicode character support in source files

Simplify Unicode character support in source files (introduced by [SQOOP-3074](#)) by defining explicit locales instead of using EscapeUtils. The Java source files generated by Sqoop will be encoded in UTF-8 format.

For more information, see [SQOOP-3075](#).

Columns added to MySql after initial Sqoop import, export back to table with same schema fails

If we export from HDFS to an RDBMS table and the file on HDFS has no value for some of the columns defined in the table, Sqoop will use the values of --input-null-string and --input-null-non-string options. Earlier this scenario was not supported and Sqoop failed.

For more information, see [SQOOP-3158](#).

Sqoop fails if the user tries to encode a null value when using --direct connector and a MySQL database

The MySQL direct connector does not support the --null-string, --null-non-string, --input-null-string, and --input-null-non-string options. These options were silently ignored earlier, but Sqoop now throws an error if these options are used with MySQL direct imports and exports.

For more information, see [SQOOP-3206](#).

Apache Zookeeper

There are no incompatible changes in this release.

Timezone Names Unsupported in Impala in CDH 6.2.0

The following table lists the time zone names / aliases no longer supported in Impala along with the canonical names you can use to replace the unsupported aliases with.

Deprecated	Replace with
ACDT	Australia/South
ACST	Australia/Darwin
ACWST	Australia/Eucla
ADT	America/Thule
AEDT	Australia/Sydney
AEST	Australia/Sydney
AFT	Asia/Kabul
AKDT	America/Anchorage
AKST	America/Anchorage
ALMT	Asia/Almaty
AMST	America/Cuiaba
AMT	Asia/Yerevan
ANAT	Asia/Anadyr
AQTT	Asia/Aqtau
AWST	Australia/West
AZOST	Atlantic/Azores
AZOT	Atlantic/Azores
AZST	Asia/Baku
AZT	Asia/Baku
Acre Time	Brazil/Acre

Afghanistan Time	Asia/Kabul
Alaska Daylight Time	America/Anchorage
Alaska Standard Time	America/Anchorage
Alma-Ata Time	Asia/Almaty
Amazon Summer Time	America/Cuiaba
Amazon Time	Brazil/West
Anadyr Time	Asia/Anadyr
Aqtau Time	Asia/Aqtau
Aqtobe Time	Asia/Aqtobe
Arabia Standard Time	Asia/Aden
Argentine Time	America/Argentina/Buenos_Aires
Armenia Time	Asia/Yerevan
Asia/Riyadh87	-
Asia/Riyadh88	-
Asia/Riyadh89	-
Atlantic Daylight Time	America/Thule
Atlantic Standard Time	America/Puerto_Rico
Australian Central Daylight Time (South Australia)	Australia/South
Australian Central Daylight Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Standard Time (Northern Territory)	Australia/Darwin
Australian Central Standard Time (South Australia)	Australia/South
Australian Central Standard Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Western Standard Time	Australia/Eucla
Australian Eastern Daylight Time (New South Wales)	Australia/Sydney
Australian Eastern Daylight Time (Tasmania)	Australia/Hobart
Australian Eastern Daylight Time (Victoria)	Australia/Victoria
Australian Eastern Standard Time (New South Wales)	Australia/Sydney
Australian Eastern Standard Time (Queensland)	Australia/Brisbane
Australian Eastern Standard Time (Tasmania)	Australia/Hobart
Australian Eastern Standard Time (Victoria)	Australia/Victoria
Australian Western Standard Time	Australia/West
Azerbaijan Summer Time	Asia/Baku
Azerbaijan Time	Asia/Baku
Azores Summer Time	Atlantic/Azores
Azores Time	Atlantic/Azores

BDT	Asia/Dhaka
BNT	Asia/Brunei
BOT	America/La_Paz
BRST	America/Sao_Paulo
BRT	America/Sao_Paulo
BTT	Asia/Thimbu
Bangladesh Time	Asia/Dhaka
Bhutan Time	Asia/Thimbu
Bolivia Time	America/La_Paz
Bougainville Standard Time	Pacific/Bougainville
Brasilia Summer Time	America/Sao_Paulo
Brasilia Time	America/Sao_Paulo
British Summer Time	GB
Brunei Time	Asia/Brunei
CCT	Indian/Cocos
CDT	America/Chicago
CEST	CET
CHADT	NZ-CHAT
CHAST	NZ-CHAT
CHOT	Asia/Choibalsan
CHUT	Pacific/Yap
CKT	Pacific/Rarotonga
CLST	America/Santiago
CLT	America/Santiago
COT	America/Bogota
CVT	Atlantic/Cape_Verde
CXT	Indian/Christmas
Canada/East-Saskatchewan	America/Regina
Cape Verde Time	Atlantic/Cape_Verde
Central African Time	Africa/Harare
Central Daylight Time	America/Chicago
Central European Summer Time	CET
Central European Time	CET
Central Indonesia Time	Asia/Makassar
Central Standard Time	America/Chicago
ChST	Pacific/Guam

Chamorro Standard Time	Pacific/Guam
Chatham Daylight Time	NZ-CHAT
Chatham Standard Time	NZ-CHAT
Chile Summer Time	America/Santiago
Chile Time	America/Santiago
China Standard Time	Asia/Shanghai
Choibalsan Time	Asia/Choibalsan
Christmas Island Time	Indian/Christmas
Chuuk Time	Pacific/Yap
Cocos Islands Time	Indian/Cocos
Colombia Time	America/Bogota
Cook Is. Time	Pacific/Rarotonga
Coordinated Universal Time	UCT
Cuba Daylight Time	Cuba
Cuba Standard Time	Cuba
DAVT	Antarctica/Davis
DDUT	Antarctica/DumontDUrville
Davis Time	Antarctica/Davis
Dumont-d'Urville Time	Antarctica/DumontDUrville
EASST	Pacific/Easter
EAST	Pacific/Easter
EDT	America/Indiana/Indianapolis
EEST	Africa/Cairo
EGST	America/Scoresbysund
EGT	America/Scoresbysund
East Indonesia Time	Asia/Jayapura
Easter Is. Summer Time	Pacific/Easter
Easter Is. Time	Pacific/Easter
Eastern African Time	Africa/Addis_Ababa
Eastern Daylight Time	America/Indiana/Indianapolis
Eastern European Summer Time	Africa/Cairo
Eastern European Time	Africa/Cairo
Eastern Greenland Summer Time	America/Scoresbysund
Eastern Greenland Time	America/Scoresbysund
Eastern Standard Time	EST
Ecuador Time	America/Guayaquil

FJST	Pacific/Fiji
FJT	Pacific/Fiji
FKT	Atlantic/Stanley
FNT	America/Noronha
Falkland Is. Time	Atlantic/Stanley
Fernando de Noronha Time	America/Noronha
Fiji Summer Time	Pacific/Fiji
Fiji Time	Pacific/Fiji
French Guiana Time	America/Cayenne
French Southern & Antarctic Lands Time	Indian/Kerguelen
GALT	Pacific/Galapagos
GAMT	Pacific/Gambier
GET	Asia/Tbilisi
GFT	America/Cayenne
GILT	Pacific/Tarawa
GMT+00:00	GMT0
GMT+01:00	Etc/GMT-1
GMT+02:00	Etc/GMT-2
GMT+03:00	Etc/GMT-3
GMT+04:00	Etc/GMT-4
GMT+05:00	Etc/GMT-5
GMT+06:00	Etc/GMT-6
GMT+07:00	Etc/GMT-7
GMT+08:00	Etc/GMT-8
GMT+09:00	Etc/GMT-9
GMT+10:00	Etc/GMT-10
GMT+11:00	Etc/GMT-11
GMT+12:00	Etc/GMT-12
GMT+13:00	Etc/GMT-13
GMT+14:00	Etc/GMT-14
GMT-01:00	Etc/GMT+1
GMT-02:00	Etc/GMT+2
GMT-03:00	Etc/GMT+3
GMT-04:00	Etc/GMT+4
GMT-05:00	Etc/GMT+5
GMT-06:00	Etc/GMT+6

GMT-07:00	Etc/GMT+7
GMT-08:00	Etc/GMT+8
GMT-09:00	Etc/GMT+9
GMT-10:00	Etc/GMT+10
GMT-11:00	Etc/GMT+11
GMT-12:00	Etc/GMT+12
GST	Asia/Dubai
GYT	America/Guyana
Galapagos Time	Pacific/Galapagos
Gambier Time	Pacific/Gambier
Georgia Time	Asia/Tbilisi
Ghana Mean Time	Africa/Accra
Gilbert Is. Time	Pacific/Tarawa
Greenwich Mean Tim	GB
Gulf Standard Time	Asia/Dubai
Guyana Time	America/Guyana
HADT	US/Aleutian
HAST	US/Aleutian
HKT	Hongkong
HOVT	Asia/Hovd
Hawaii Standard Time	HST
Hawaii-Aleutian Daylight Time	US/Aleutian
Hawaii-Aleutian Standard Time	US/Aleutian
Hong Kong Time	Hongkong
Hovd Time	Asia/Hovd
ICT	Asia/Ho_Chi_Minh
IDT	Israel
IOT	Indian/Chagos
IRDT	Iran
IRKT	Asia/Chita
IRST	Iran
India Standard Time	Asia/Kolkata
Indian Ocean Territory Time	Indian/Chagos
Indochina Time	Asia/Ho_Chi_Minh
Iran Daylight Time	Iran
Iran Standard Time	Iran

Irish Summer Time	Eire
Irkutsk Time	Asia/Chita
Israel Daylight Time	Israel
Israel Standard Time	Israel
Japan Standard Time	Asia/Tokyo
KGT	Asia/Bishkek
KOST	Pacific/Kosrae
KRAT	Asia/Krasnoyarsk
KST	ROK
Khandyga Time	Asia/Khandyga
Kirgizstan Time	Asia/Bishkek
Korea Standard Time	ROK
Kosrae Time	Pacific/Kosrae
Krasnoyarsk Time	Asia/Krasnoyarsk
LHDT	Australia/LHI
LHST	Australia/LHI
LINT	Pacific/Kiritimati
Line Is. Time	Pacific/Kiritimati
Lord Howe Daylight Time	Australia/LHI
Lord Howe Standard Time	Australia/LHI
MAGT	Asia/Magadan
MART	Pacific/Marquesas
MAWT	Antarctica/Mawson
MDT	Navajo
MEST	MET
MHT	Kwajalein
MIST	Antarctica/Macquarie
MMT	Asia/Rangoon
MSK	W-SU
MUT	Indian/Mauritius
MVT	Indian/Maldives
MYT	Asia/Kuching
Macquarie Island Standard Time	Antarctica/Macquarie
Magadan Time	Asia/Magadan
Malaysia Time	Asia/Kuching
Maldives Time	Indian/Maldives

Marquesas Time	Pacific/Marquesas
Marshall Islands Time	Kwajalein
Mauritius Time	Indian/Mauritius
Mawson Time	Antarctica/Mawson
Middle Europe Summer Time	MET
Middle Europe Time	MET
Mideast/Riyadh87	-
Mideast/Riyadh88	-
Mideast/Riyadh89	-
Moscow Standard Time	W-SU
Mountain Daylight Time	Navajo
Mountain Standard Time	MST
Myanmar Time	Asia/Rangoon
NCT	Pacific/Noumea
NDT	America/St_Johns
NFT	Pacific/Norfolk
NOVT	Asia/Novosibirsk
NPT	Asia/Katmandu
NRT	Pacific/Nauru
NUT	Pacific/Niue
NZDT	NZ
NZST	NZ
Nauru Time	Pacific/Nauru
Nepal Time	Asia/Katmandu
New Caledonia Time	Pacific/Noumea
New Zealand Daylight Time	NZ
New Zealand Standard Time	NZ
Newfoundland Daylight Time	America/St_Johns
Newfoundland Standard Time	America/St_Johns
Niue Time	Pacific/Niue
Norfolk Time	Pacific/Norfolk
Novosibirsk Time	Asia/Novosibirsk
OMST	Asia/Omsk
ORAT	Asia/Oral
Omsk Time	Asia/Omsk
Oral Time	Asia/Oral

PDT	America/Los_Angeles
PET	America/Lima
PETT	Asia/Kamchatka
PGT	Pacific/Port_Moresby
PHOT	Pacific/Enderbury
PHT	Asia/Manila
PKT	Asia/Karachi
PMDT	America/Miquelon
PMST	America/Miquelon
PONT	Pacific/Ponape
PWT	Pacific/Palau
PYST	America/Asuncion
PYT	America/Asuncion
Pacific Daylight Time	America/Los_Angeles
Pacific Standard Time	America/Los_Angeles
Pakistan Time	Asia/Karachi
Palau Time	Pacific/Palau
Papua New Guinea Time	Pacific/Port_Moresby
Paraguay Summer Time	America/Asuncion
Paraguay Time	America/Asuncion
Peru Time	America/Lima
Petropavlovsk-Kamchatski Time	Asia/Kamchatka
Philippines Time	Asia/Manila
Phoenix Is. Time	Pacific/Enderbury
Pierre & Miquelon Daylight Time	America/Miquelon
Pierre & Miquelon Standard Time	America/Miquelon
Pitcairn Standard Time	Pacific/Pitcairn
Pohnpei Time	Pacific/Ponape
QYZT	Asia/Qyzylorda
Qyzylorda Time	Asia/Qyzylorda
RET	Indian/Reunion
ROTT	Antarctica/Rothera
Reunion Time	Indian/Reunion
Rothera Time	Antarctica/Rothera
SAKT	Asia/Sakhalin
SAMT	Europe/Samara

SAST	Africa/Maseru
SBT	Pacific/Guadalcanal
SCT	Indian/Mahe
SGT	Singapore
SRET	Asia/Srednekolymsk
SRT	America/Paramaribo
SYOT	Antarctica/Syowa
Sakhalin Time	Asia/Sakhalin
Samara Time	Europe/Samara
Samoa Standard Time	US/Samoa
Seychelles Time	Indian/Mahe
Singapore Time	Singapore
Solomon Is. Time	Pacific/Guadalcanal
South Africa Standard Time	Africa/Maseru
South Georgia Standard Time	Atlantic/South_Georgia
Srednekolymsk Time	Asia/Srednekolymsk
Suriname Time	America/Paramaribo
Syowa Time	Antarctica/Syowa
SystemV/AST4	America/Puerto_Rico
SystemV/AST4ADT	Canada/Atlantic
SystemV/CST6	Canada/Saskatchewan
SystemV/CST6CDT	US/Central
SystemV/EST5	America/Cayman
SystemV/EST5EDT	America/New_York
SystemV/HST10	US/Hawaii
SystemV/MST7	US/Arizona
SystemV/MST7MDT	America/Denver
SystemV/PST8	Pacific/Pitcairn
SystemV/PST8PDT	US/Pacific
SystemV/YST9	Pacific/Gambier
SystemV/YST9YDT	US/Alaska
TAHT	Pacific/Tahiti
TFT	Indian/Kerguelen
TJT	Asia/Dushanbe
TKT	Pacific/Fakaofo
TLT	Asia/Dili

TMT	Asia/Ashgabat
TOT	Pacific/Tongatapu
TVT	Pacific/Funafuti
Tahiti Time	Pacific/Tahiti
Tajikistan Time	Asia/Dushanbe
Timor-Leste Time	Asia/Dili
Tokelau Time	Pacific/Fakaofo
Tonga Time	Pacific/Tongatapu
Turkmenistan Time	Asia/Ashgabat
Tuvalu Time	Pacific/Funafuti
ULAT	Asia/Ulan_Bator
UYST	America/Montevideo
UYT	America/Montevideo
UZT	Asia/Tashkent
Ulaanbaatar Time	Asia/Ulan_Bator
Uruguay Summer Time	America/Montevideo
Uruguay Time	America/Montevideo
Ust-Nera Time	Asia/Ust-Nera
Uzbekistan Time	Asia/Tashkent
VET	America/Caracas
VLAT	Asia/Ust-Nera
VOST	Antarctica/Vostok
VUT	Pacific/Efate
Vanuatu Time	Pacific/Efate
Venezuela Time	America/Caracas
Vladivostok Time	Asia/Vladivostok
Vostok Time	Antarctica/Vostok
WAKT	Pacific/Wake
WAST	Africa/Windhoek
WAT	Africa/Lagos
WEST	WET
WFT	Pacific/Wallis
WGST	America/Godthab
WGT	America/Godthab
WIB	Asia/Jakarta
WIT	Asia/Jayapura

WITA	Asia/Makassar
WSDT	Pacific/Apia
WSST	Pacific/Apia
Wake Time	Pacific/Wake
Wallis & Futuna Time	Pacific/Wallis
West Indonesia Time	Asia/Jakarta
West Samoa Daylight Time	Pacific/Apia
West Samoa Standard Time	Pacific/Apia
Western African Summer Time	Africa/Windhoek
Western African Time	Africa/Lagos
Western European Summer Time	WET
Western European Time	WET
Western Greenland Summer Time	America/Godthab
Western Greenland Time	America/Godthab
XJT	Asia/Urumqi
Xinjiang Standard Time	Asia/Urumqi
YAKT	Asia/Yakutsk
YEKT	Asia/Yekaterinburg
Yakutsk Time	Asia/Yakutsk
Yekaterinburg Time	Asia/Yekaterinburg

Known Issues and Limitations in CDH 6.2.0

The following sections describe the known issues in CDH 6.2.0, grouped by component:

Operating System Known Issues

Known issues and workarounds related to operating systems are listed below.

Linux kernel security patch and CDH services crashes CVE-2017-10000364

After applying a recent Linux kernel security patch for [CVE-2017-1000364](#), CDH services that use the JSVC set of libraries crash with a Java Virtual Machine (JVM) error such as:

```
A fatal error has been detected by the Java Runtime Environment:
SIGBUS (0x7) at pc=0x00007fe91ef6cebc, pid=30321, tid=0x00007fe930c67700
```

Cloudera services for HDFS and Impala cannot start after applying the patch.



Important: If you have not upgraded your Linux kernel using the distribution's patch for CVE-2017-1000364, do **not** apply the patch.

Commonly used Linux distributions are shown in the table below. However, the issue affects any CDH release that runs on RHEL, CentOS, Oracle Linux, SUSE Linux, or Ubuntu and that has had the Linux kernel security patch for CVE-2017-1000364 applied.

If you have already applied the patch for your OS according to the advisories for CVE-2017-1000364, apply the kernel update that contains the fix for your operating system (some of which are listed in the table). If you cannot apply the kernel update, you can workaround the issue by increasing the Java thread stack size as detailed in the steps below.

Distribution	Advisories for CVE-2017-1000364	Advisory updates
Oracle Linux 6	ELSA-2017-1486	Oracle has fixed this problem in ELSA-2017-1723 .
Oracle Linux 7	ELSA-2017-1484	Oracle has also added the fix for Oracle Linux 7 in ELBA-2017-1674 .
RHEL 6	RHSA-2017-1486	RedHat has fixed this problem for RHEL 6, marked this as outdated and superseded by RHSA-2017-1723 .
RHEL 7	RHSA-2017-1484	RedHat has fixed this problem for RHEL 7 and has marked this patch as outdated and superseded by RHBA-2017-1674 .
SLES	CVE-2017-1000364	SUSE has also fixed this problem and the patch names are included in this same advisory.

Workaround

If you cannot apply the kernel update, you can set the Java thread stack size to `-Xss1280k` for the affected services using the appropriate Java configuration option or the environment advanced configuration snippet, as detailed below.

For role instances that have specific Java configuration options properties:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Type `java` in the search field to display Java related configuration parameters. The **Java Configuration Options for Catalog Server** property field displays. Type `-Xss1280k` in the entry field, adding to any existing settings.
4. Click **Save Changes**.
5. Navigate to the HDFS service by selecting **Clusters > HDFS**.
6. Click the **Configuration** tab.
7. Click the Scope filter **DataNode**. The **Java Configuration Options for DataNode** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
8. Click **Save Changes**.
9. Select the Scope filter **NFS Gateway**. The **Java Configuration Options for NFS Gateway** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
- 10 Click **Save Changes**.
- 11 Restart the affected roles (or configure the safety valves in next section and restart when finished with all configurations).

For role instances that do not have specific Java configuration options:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Click the Scope filter **Impala Daemon** and Category filter **Advanced**.
4. Type `impala daemon` environment in the search field to find the safety valve entry field.
5. In the **Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

6. Click **Save Changes**.
7. Click the Scope filter **Impala StateStore** and Category filter **Advanced**.
8. In the **Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

9. Click Save Changes.**10** Restart the affected roles.

The table below summarizes the parameters that can be set for the affected services:

Service	Settable Java Configuration Option
HDFS DataNode	Java Configuration Options for DataNode
HDFS NFS Gateway	Java Configuration Options for NFS Gateway
Impala Catalog Server	Java Configuration Options for Catalog Server
Impala Daemon	Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K
Impala StateStore	Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K

Cloudera Issue: CDH-55771

Leap-Second Events



Note: The [next leap-second event](#) is unknown at this time. The last leap-second event occurred on December 16, 2016 at 23:59:60 UTC.

Impact: After a leap-second event, Java applications (including CDH services) using older Java and Linux kernel versions, may consume almost 100% CPU. See <https://access.redhat.com/articles/15145>.

Leap-second events are tied to the time synchronization methods of the Linux kernel, the Linux distribution and version, and the Java version used by applications running on affected kernels.

Although Java is increasingly agnostic to system clock progression (and less susceptible to a kernel's mishandling of a leap-second event), using JDK 7 or 8 should prevent issues at the CDH level (for CDH components that use the Java Virtual Machine).

Immediate action required:

(1) Ensure that the kernel is up to date.

- **RHEL6/7, CentOS 6/7** - 2.6.32-298 or higher
- **Oracle Enterprise Linux (OEL)** - Kernels built in 2013 or later
- **SLES12** - No action required.

(2) Ensure that your Java JDKs are current (especially if the kernel is not up to date and cannot be upgraded).

- **Java 8** - No action required.

(3) Ensure that your systems use either NTP or PTP synchronization.

For systems not using time synchronization, update both the OS tzdata and Java tzdata packages to the tzdata-2016g version, at a minimum. For OS tzdata package updates, contact OS support or check updated OS repositories. For Java tzdata package updates, see Oracle's [Timezone Updater Tool](#).

Cloudera Issue: CDH-44788, TSB-189

Apache Accumulo Known Issues

There are no notable known issues in this release of Apache Accumulo.

Apache Crunch Known Issues

Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

Apache Flume Known Issues

Fast Replay does not work with encrypted File Channel

If an encrypted file channel is set to use fast replay, the replay will fail and the channel will fail to start.

Workaround: Disable fast replay for the encrypted channel by setting `use-fast-replay` to false.

Apache Issue: [FLUME-1885](#)

Apache Hadoop Known Issues

This page includes known issues and related topics, including:

Deprecated Properties

Several Hadoop and HDFS properties have been deprecated as of Hadoop 3.0 and later. For details, see [Deprecated Properties](#).

Hadoop Common

Error when executing Java classes from a CDH cluster running on Ubuntu 18

Using the `hadoop` command-line interface for executing Java classes that are not in the default package results in error messages similar to the following:

```
#hadoop org.apache.hadoop.conf.Configuration
/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.914039/bin/../lib/hadoop/libexec//hadoop-functions.sh:
line 2366: HADOOP_ORG.APACHE.HADOOP.CONF.CONFIGURATION_USER: bad substitution
/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.x.p0.914039/bin/../lib/hadoop/libexec//hadoop-functions.sh:
line 2331: HADOOP_ORG.APACHE.HADOOP.CONF.CONFIGURATION_USER: bad substitution
/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.x.p0.914039/bin/../lib/hadoop/libexec//hadoop-functions.sh:
line 2426: HADOOP_ORG.APACHE.HADOOP.CONF.CONFIGURATION_OPTS: bad substitution
```

This issue occurs only in CDH 6.2 clusters running on Ubuntu 18 and the error messages can be safely ignored.

Workaround: Run the `java` command directly using `hadoop classpath` to get the classpath. For example, instead of `hadoop org.apache.hadoop.conf.Configuration`, you can run `java -cp `hadoop classpath` org.apache.hadoop.conf.Configuration`.

Affected Versions: CDH 6.2.0

Fixed Versions: N/A

Apache Issue: [HADOOP-16167](#)

Hadoop LdapGroupsMapping does not support LDAPS for self-signed LDAP server

Hadoop LdapGroupsMapping does not work with LDAP over SSL (LDAPS) if the LDAP server certificate is self-signed. This use case is currently not supported even if Hadoop User Group Mapping LDAP TLS/SSL Enabled, Hadoop User Group Mapping LDAP TLS/SSL Truststore, and Hadoop User Group Mapping LDAP TLS/SSL Truststore Password are filled properly.

Affected Versions: All CDH versions

Apache Issue: [HADOOP-12862](#)

Cloudera Issue: CDH-37926

HDFS

OIV ReverseXML processor fails

The HDFS OIV ReverseXML processor fails if the XML file contains escaped characters.

Affected Versions: CDH 6.x

Apache Issue: [HDFS-12828](#)

Cannot move encrypted files to trash

With HDFS encryption enabled, you cannot move encrypted files or directories to the trash directory.

Workaround: To remove encrypted files/directories, use the following command with the `-skipTrash` flag specified to bypass trash.

```
rm -r -skipTrash /testdir
```

Affected Versions: All CDH versions

Apache Issue: [HADOOP-10902](#)

HDFS NFS gateway and CDH installation (using packages) limitation

HDFS NFS gateway works as shipped ("out of the box") only on RHEL-compatible systems, but not on SLES or Ubuntu. Because of a bug in native versions of `portmap/rpcbind`, the HDFS NFS gateway does not work out of the box on SLES or Ubuntu systems when CDH has been installed from the command-line, using packages. It does work on supported versions of RHEL-compatible systems on which `rpcbind-0.2.0-10.el6` or later is installed, and it does work if you use Cloudera Manager to install CDH, or if you start the gateway as root.

Workarounds and caveats:

- On Red Hat and similar systems, make sure `rpcbind-0.2.0-10.el6` or later is installed.
- On SLES and Ubuntu systems, do one of the following:
 - Install CDH using Cloudera Manager; or
 - Start the NFS gateway as root; or
 - [Start the NFS gateway without using packages](#); or
 - You can use the gateway by running `rpcbind` in insecure mode, using the `-i` option, but keep in mind that this allows anyone from a remote host to bind to the portmap.

Upstream Issue: [731542](#) (Red Hat), [823364](#) (SLES)

No error when changing permission to 777 on .snapshot directory

Snapshots are read-only; running `chmod 777` on the `.snapshot`s directory does not change this, but does not produce an error (though other illegal operations do).

Affected Versions: All CDH versions

Apache Issue: [HDFS-4981](#)

Snapshot operations are not supported by ViewFileSystem

Affected Versions: All CDH versions

Snapshots do not retain directories' quotas settings

Affected Versions: All CDH versions

Apache Issue: [HDFS-4897](#)

Permissions for `dfs.namenode.name.dir` incorrectly set

Hadoop daemons should set permissions for the `dfs.namenode.name.dir` (or `dfs.name.dir`) directories to `drwx-----` (700), but in fact these permissions are set to the file-system default, usually `drwxr-xr-x` (755).

Workaround: Use `chmod` to set permissions to 700.

Affected Versions: All CDH versions

Apache Issue: [HDFS-2470](#)

`hadoop fsck -move` does not work in a cluster with host-based Kerberos

Workaround: Use `hadoop fsck -delete`

Affected Versions: All CDH versions

Apache Issue: None

Block report can exceed maximum RPC buffer size on some DataNodes

On a DataNode with a large number of blocks, the block report may exceed the maximum RPC buffer size.

Workaround: Increase the value `ipc.maximum.data.length` in `hdfs-site.xml`:

```
<property>
  <name>ipc.maximum.data.length</name>
  <value>268435456</value>
</property>
```

Affected Versions: All CDH versions

Apache Issue: None

MapReduce2 and YARN

YARN's Continuous Scheduling can cause slowness in Oozie

When Continuous Scheduling is enabled in Yarn, this can cause slowness in Oozie due to long delays in communicating with Yarn. In Cloudera Manager 5.9.0 and higher, Enable Fair Scheduler Continuous Scheduler is turned off by default.

Workaround: Turn off Enable Fair Scheduler Continuous Scheduling in Cloudera Manager YARN Configuration. To keep equivalent benefits of this feature, turn on Fair Scheduler Assign Multiple Tasks.

Affected Versions: All CDH versions

Cloudera Issue: CDH-60788

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

Affected Versions: All CDH versions

Apache Issue: None

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Routable IP address required by ResourceManager

ResourceManager requires routable `host:port` addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard `0.0.0.0` address.

Workaround: Set the address, in the form `host:port`, either in the client-side configuration, or on the command line when you submit the job.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-6808

Amazon S3 copy may time out

The Amazon S3 filesystem does not support renaming files, and performs a copy operation instead. If the file to be moved is very large, the operation can time out because S3 does not report progress during the operation.

Workaround: Use `-Dmapred.task.timeout=15000000` to increase the MR task timeout.

Affected Versions: All CDH versions

Apache Issue: [MAPREDUCE-972](#)

Cloudera Issue: CDH-17955

GPU and Custom Resource Types Are Not Added to the YARN Client's Configuration File When Enabled

When GPU or other custom resource type is configured in Cloudera Manager, the appropriate resource (for example `yarn.io/gpu`) is not added to the YARN client's configuration (`yarn-site.xml`) file. As a result, jobs that use GPU or the configured custom resource type will fail.

Workaround: Add the appropriate resource manually to the **YARN Client Advanced Configuration Snippet (Safety Valve) for `yarn-site.xml`:**

1. In Cloudera Manager select YARN service and go to **Configuration**.
2. Search for *YARN Client Advanced Configuration Snippet (Safety Valve) for `yarn-site.xml`*
3. Add the following snippet:

```
<property>
<name>yarn.resource-types</name>
<value>yarn.io/gpu</value>
</property>
```

Affected Versions: CDH 6.2.0

Cloudera Issue: OPSAPS-49507

GPU or Custom Resource Type User Jobs can Fail After Recovery

When a GPU or other custom resource goes offline when it has containers that use that particular resource and they have not reached completion, after the restart the application will start to recover. However, if the resource is not available anymore the job that uses that resource will fail.

Workaround: N/A

Affected Versions: CDH 6.2.0

Cloudera Issue: CDH-77649

NodeManager Fails if GPU Use Is Enabled without any Configured GPU

When **Enable GPU Usage** is enabled for a NodeManager and there is no properly configured GPU device in that node, the NodeManager will not start.

Workaround: Disable **Enable GPU Usage** for that NodeManager in Cloudera Manager.

Affected Versions: CDH 6.2.0

Apache Issue: [YARN-9217](#)

Apache HBase Known Issues

IOException from Timeouts

CDH 5.12.0 includes the fix [HBASE-16604](#), where the internal scanner that retries in case of IOException from timeouts could potentially miss data. Java clients were properly updated to account for the new behavior, but thrift clients will now see exceptions where the previous missing data would be.

Workaround: Create a new scanner and retry the operation when encountering this issue.

`IntegrationTestReplication` fails if replication does not finish before the `verify` phase begins

During `IntegrationTestReplication`, if the `verify` phase starts before the `replication` phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the `-t` flag to set the timeout value before starting verification.

Cloudera Issue: None.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

`ExportSnapshot` or `DistCP` operations may fail on the Amazon `s3a://` protocol

`ExportSnapshot` or `DistCP` operations may fail on AWS when using certain JDK 8 versions, due to an incompatibility between the AWS Java SDK 1.9.x and the `joda-time` date-parsing module.

Workaround: Use `joda-time 2.8.1` or higher, which is included in AWS Java SDK 1.10.1 or higher.

Cloudera Issue: None.

An operating-system level tuning issue in RHEL7 causes 30% latency regressions

Workaround: Avoid using RHEL 7 if you have a latency-critical workload. For a cached workload, consider tuning the C-state (power-saving) behavior of your CPUs.

Cloudera Issue: CDH-32642

Severity: Medium

A RegionServer under extreme duress due to back-to-back garbage collection combined with heavy load on HDFS can lock up while attempting to append to the WAL

The RegionServer appears operational to ZooKeeper, and continues to host regions, but cannot complete any writes. The most obvious symptom is that all writes to regions on this RegionServer time out, and the RegionServer log shows no progress other than queuing of flushes that never complete. Log messages such as the following may occur:

```
124028 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.Sleeper: We slept
42911ms instead of 3000ms,
this is likely due to a long garbage collecting pause and it's usually bad, see
http://hbase.#
124029 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.JvmPauseMonitor: Detected
pause in JVM or
host machine (eg GC): pause of approximately 41110ms
```

```
1806 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:
Slow sync cost: 2734 ms, current pipeline:
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
1807 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:
Slow sync cost: 2963 ms, current pipeline:
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
```

Apache Issue: [HBASE-14374](#)

Workaround: Restart the RegionServer. To avoid the problem, adjust garbage-collection settings, give the RegionServer more RAM, and reduce the load on HDFS.

Export to Azure Blob Storage (the `wasb://` or `wasbs://` protocol) is not supported

CDH 5.3 and higher supports Azure Blob Storage for some applications. However, a null pointer exception occurs when you specify a `wasb://` or `wasbs://` location in the `--copy-to` option of the `ExportSnapshot` command or as the output directory (the second positional argument) of the `Export` command.

Workaround: None.

Apache Issue: [HADOOP-12717](#)

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a `Delete Table` fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.
- If `hbaseAdmin.modifyTable()` is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The `postOperation` is implemented only for `postDeleteColumn()`.
- If `Create Table` fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: None

Apache Issue: [HBASE-6992](#)

Apache Hive/HCatalog/Hive on Spark/Hive Metastore Known Issues

This topic contains:

- [Hive Known Issues](#)
- [HCatalog Known Issues](#)
- [Hive on Spark Known Issues](#)
- [Hive Metastore Known Issues](#) on page 246

Hive Known Issues

Hive Jobs Are Submitted to a Single Queue When Sentry is Deployed

Hive jobs are not submitted into the correct YARN queue when Hive is using Sentry because Hive does not use the YARN API to resolve the user or group of the job's original submitter. This causes the job to be placed in a queue using the placement rules based on the Hive user. The HiveServer2 fair scheduler queue mapping used for "non-impersonation" mode does not handle the primary-secondary queue mappings correctly.

Affected Version: CDH 6.0.0 and higher

Workaround: If you are a Hive and Sentry user, do not upgrade to CDH 6.0.0 or 6.0.1. This issue will be fixed as soon as possible. If you must use Hive and Sentry in CDH 6.0.0 or 6.0.1, see [YARN Dynamic Resource Pools Do Not Work with Hive When Sentry Is Enabled](#) for additional workarounds.

When vectorization is enabled on any file type (ORC, Parquet) queries that divide by zero using the modulo operator (%) return an error

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query divides by zero using the modulo operator (%), it returns the following error: `Arithmetic exception [divide by] 0.` For example, if you run the following query this issue is triggered: `SELECT 100 % column_c1 FROM table_t1;` and the value in `column_c1` is zero. The divide operator (/) is not affected by this issue.

Workaround: Disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-19564](#)

Cloudera Issue: CDH-71211

When vectorization is enabled for Hive on any file type (ORC, Parquet) queries that perform comparisons in the `SELECT` clause on large values in columns with the data type of `BIGINT` might return wrong results

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query performs a comparison operation between very large values in columns that are `BIGINT` data types in the `SELECT` clause of the query, incorrect results might be returned. Comparison operators include `==`, `!=`, `<`, `<=`, `>`, and `>=`. This issue does not occur when the

comparison operation is performed in the filtering clause of the query. This issue can also occur when the difference of values in such columns is out of range for a `LONG` (64-bit) data type. For example, if `column_c1` stores `8976171455044006767` and `column_c2` stores `-7272907770454997143`, a query such as `SELECT column_c1 < column_c2 FROM table_test` returns `true` instead of `false` because the difference (`8976171455044006767 - (-7272907770454997143)`) is `1.6249079225499E19` which is greater than `9.22337203685478E18`, which is the maximum possible value that a `LONG` (64-bit) data type can hold.

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-20207](#)

Cloudera Issue: CDH-70996

Workaround: Use a `DECIMAL` type instead of `BIGINT` for columns that might contain very large values. Another option is to disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Specified column position in the ORDER BY clause is not supported for SELECT * queries

When column positions are specified in `ORDER BY` clauses, they are not honored for `SELECT *` queries and an error is returned as shown in the following example:

```
CREATE TABLE decimal_1 (id decimal(5,0));
SELECT * FROM decimal_1 ORDER BY 1 limit 100;
Error while compiling statement: FAILED: SemanticException [Error 10219]: Position in
ORDER BY is not supported when using SELECT *
```

Instead the query must list out the columns it is selecting.

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-68550

DirectSQL with PostgreSQL

Hive doesn't support Hive direct SQL queries with PostgreSQL database. It only supports this feature with MySQL, MariaDB, and Oracle. With PostgreSQL, direct SQL is disabled as a precaution, since there have been issues reported upstream where it is not possible to fallback on DataNucleus in the event of some failures, plus other non-standard behaviors. For more information, see [Hive Configuration Properties](#).

Affected Versions: All CDH versions

Cloudera Issue: CDH-49017

`ALTER PARTITION ... SET LOCATION` does not work on Amazon S3 or between S3 and HDFS

Cloudera recommends that you do not use `ALTER PARTITION ... SET LOCATION` on S3 or between S3 and HDFS. The rest of the `ALTER PARTITION` commands work as expected.

Affected Versions: All CDH versions

Cloudera Issue: CDH-42420

Commands run against an Oracle-backed metastore might fail

Commands run against an Oracle-backed Metastore fail with error:

```
javax.jdo.JDODataStoreException Incompatible data type for column TBLS.VIEW_EXPANDED_TEXT
: was CLOB (datastore),
but type expected was LONGVARCHAR (metadata). Please check that the type in the datastore
and the type specified in the MetaData are consistent.
```

This error might occur if the metastore is run on top of an Oracle database with the configuration property `datanucleus.validateColumns` set to true.

Workaround: Set `datanucleus.validateColumns=false` in the `hive-site.xml` configuration file.

Affected Versions: All CDH versions

Cannot create archive partitions with external HAR (Hadoop Archive) tables

`ALTER TABLE ... ARCHIVE PARTITION` is not supported on external tables.

Affected Versions: All CDH versions

Cloudera Issue: CDH-9638

Object types Server and URI are not supported in "SHOW GRANT ROLE `roleName` on OBJECT `objectName`" statements

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Cloudera Issue: CDH-19430

HCatalog Known Issues



Note: As of CDH 5, HCatalog is part of Apache Hive.

There are no notable known issues in this release of HCatalog.

Hive on Spark (HoS) Known Issues

Hive on Spark queries fail with "Timed out waiting for client to connect" for an unknown reason

If this exception is preceded by logs of the form "client.RpcRetryingCaller: Call exception...", then this failure is due to an unavailable HBase service. On a secure cluster, `spark-submit` will try to obtain delegation tokens from HBase, even though Hive on Spark might not need them. So if HBase is unavailable, `spark-submit` throws an exception.

Workaround: Fix the HBase service, or set `spark.yarn.security.tokens.hbase.enabled` to false.

Affected Versions: CDH 5.7.0 and higher

Cloudera Issues: CDH-59591, CDH-59599

Hive Metastore Known Issues

HMS Read Authorization: `get_num_partitions_by_filter` Ignores Authorization

A user can get the number of partitions of a table regardless of the user's permissions

HMS Read Authorization: `Get_Partitions_With_Auth` Returns All Partitions in a Table When User Has Select Access to One Column

When a user does not have any privilege on that table, including access to any of its columns, then the user has no access to the table's partition metadata. If a user has select permission on any one column of a table, then the can get all metadata of the partitions for the table, including columns that the user does not have any permission on.

HMS Read Authorization: `Partition_Name_To_Vals` Is Not Protected by Read Authorization

A user can get the partition values of a valid partition name regardless of their permissions.

HMS Notifications API is Not Protected by Server Side Read Authorization

A user without any privileges cannot get metadata of a database or a table from the HMS server. However, it is possible for such a user to access the metadata changes.

Hue Known Issues

The following sections describe known issue and workaround in Hue for CDH 6.2.0:

Table Browser Must Be Refreshed to View Tables Created with the Data Import Wizard

When you create a new table from a file by using the Data Import Wizard, the newly created table columns do not display in the Table Browser until you refresh it.

For example:

1. Create a table from the sample file `web_logs_2.csv` by clicking the plus sign in the left panel, which launches the Data Import Wizard:

The screenshot shows the Hue Table Browser interface. On the left, there's a sidebar with icons for databases, tables, and queries. Below that is a list of tables under the 'default' database: 'customers', 'sample_07', 'sample_08', and 'web_logs'. At the top right of this list is a button labeled '(4) +' with a circular refresh icon next to it. A large yellow arrow points upwards towards this button.

2. After you define the table, click **Submit** to generate the new table:

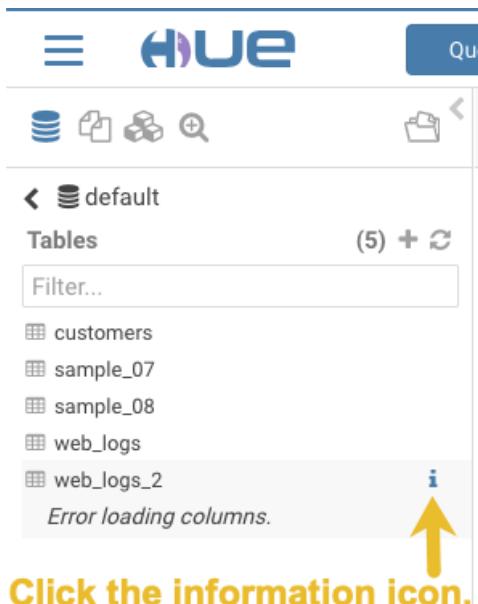
The screenshot shows the Hue Data Import Wizard. The top navigation bar includes 'Query' and a search bar. The main area is titled 'Import to table'. On the left, there's a sidebar with 'Tables' and a list of existing tables: 'customers', 'sample_07', 'sample_08', and 'web_logs'. The central part of the screen shows a preview of the data from the file '/user/admin/2015_11_19/web_logs_2.csv'. Below this is a 'DESTINATION' section with a 'Name' field set to 'default.web_logs_2'. Under 'PROPERTIES', the 'Format' is set to 'Text'. In the 'FIELDS' section, four fields are defined: 'field_1' (bigint), 'field_2' (string), 'field_3' (bigint), and 'field_4' (string). A yellow arrow points to the 'Submit' button at the bottom of the form. Another yellow arrow points to the 'Define table characteristics.' text above the 'Fields' section.

3. After you generate the new table, you can see it listed on the left assist panel, but when you click the table name to display the columns, an error displays:

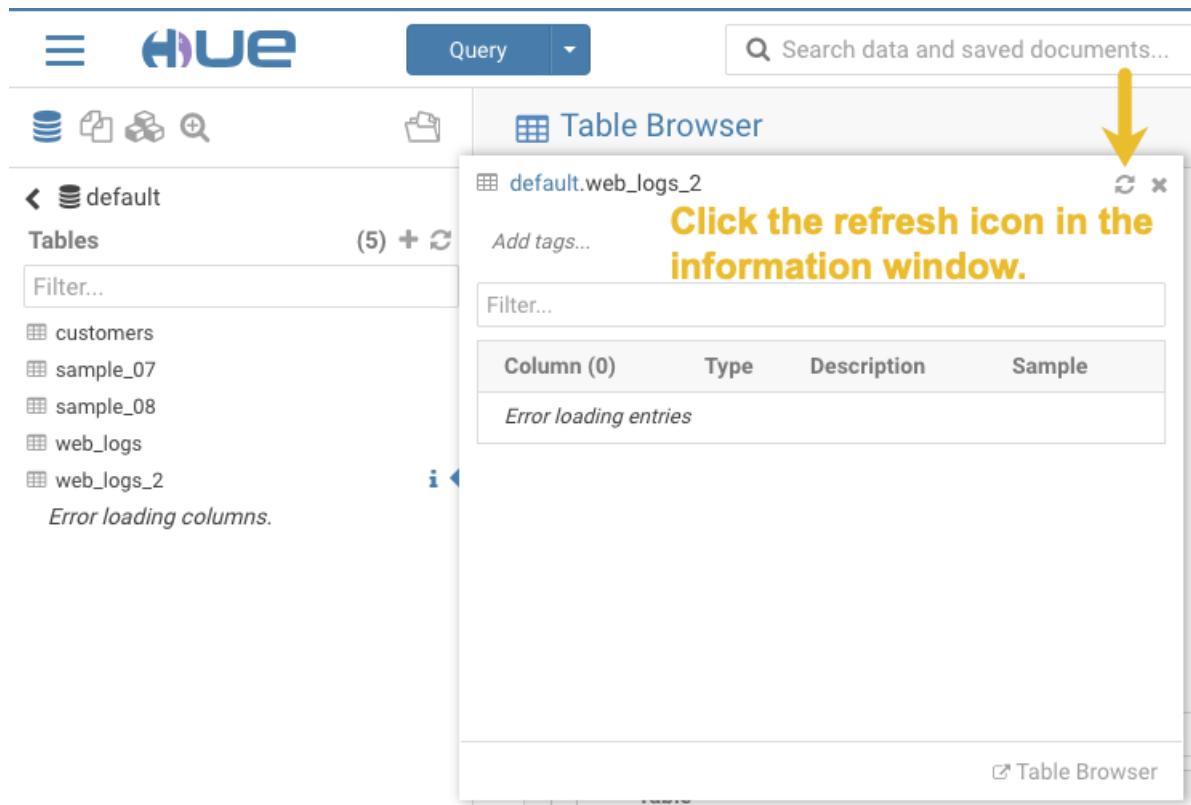


Workaround:

1. Click the information icon that is adjacent to the new table:



2. In the information window that opens, click the refresh icon in the upper right corner to view the table columns:



Using the refresh icon in the information window is the least expensive way to refresh the page so performance is not affected.

Affected Version(s): CDH 6.2.0

Cloudera Issue: CDH-77238

Hue does not support the Spark App

Hue does not currently support the Spark application.

Connecting to PostgreSQL Database Fails with Error "No module named psycopg2"

When configuring Hue to use a PostgreSQL database, the connection fails with the following error:

```
Error loading psycopg2 module: No module named psycopg2
```

Workaround: Install the `psycopg2` Python package as documented in [Installing the psycopg2 Python Package](#).

Affected Versions: All CDH 6 versions

Fixed Versions: None

Apache Issue: N/A

Cloudera Issue: CDH-65804

[Apache Impala Known Issues](#)

The following sections describe known issues and workarounds in Impala, as of the current production release. This page summarizes the most serious or frequently encountered issues in the current release, to help you make planning decisions about installing and upgrading. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.



Note: The online issue tracking system for Impala contains comprehensive information and is updated in real time. To verify whether an issue you are experiencing has already been reported, or which release an issue is fixed in, search on the [Impala JIRA tracker](#).

Impala Known Issues: Startup

These issues can prevent one or more Impala-related daemons from starting properly.

Impala requires FQDN from hostname command on kerberized clusters

The method Impala uses to retrieve the host name while constructing the Kerberos principal is the `gethostname()` system call. This function might not always return the fully qualified domain name, depending on the network configuration. If the daemons cannot determine the FQDN, Impala does not start on a kerberized cluster.

Workaround: Test if a host is affected by checking whether the output of the `hostname` command includes the FQDN. On hosts where `hostname` only returns the short name, pass the command-line flag `--hostname=fully_qualified_domain_name` in the startup options of all Impala-related daemons.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4978](#)

Impala Known Issues: Crashes and Hangs

These issues can cause Impala to quit or become unresponsive.

Unable to view large catalog objects in catalogd Web UI

In `catalogd` Web UI, you can list metadata objects and view their details. These details are accessed via a link and printed to a string formatted using thrift's `DebugProtocol`. Printing large objects (> 1 GB) in Web UI can crash `catalogd`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6841](#)

Impala Known Issues: Performance

These issues involve the performance of operations such as queries or DDL statements.

Metadata operations block read-only operations on unrelated tables

Metadata operations that change the state of a table, like `COMPUTE STATS` or `ALTER RECOVER PARTITIONS`, may delay metadata propagation of unrelated unloaded tables triggered by statements like `DESCRIBE` or `SELECT` queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6671](#)

Impala Known Issues: Security

These issues relate to security features, such as Kerberos authentication, Sentry authorization, encryption, auditing, and redaction.

Impala does not support Heimdal Kerberos

Heimdal Kerberos is not supported in Impala.

Apache Issue: [IMPALA-7072](#)

Affected Versions: All CDH 6 versions

System-wide auth-to-local mapping not applied correctly to Kudu service account

Due to system `auth_to_local` mapping, the principal may be mapped to some local name.

When running with Kerberos enabled, you may hit the following error message where `<random-string>` is some random string which doesn't match the primary in the Kerberos principal.

```
WARNINGS: TransmitData() to X.X.X.X:27000 failed: Remote error: Not authorized:
{username='<random-string>', principal='impala/redacted'} is not allowed to access
DataStreamService
```

Workaround: Start Impala with the `--use_system_auth_to_local=false` flag to ignore the system-wide auth_to_local mappings configured in `/etc/krb5.conf`.

Apache Issue: KUDU-2198

Affected Versions: CDH 5.15, CDH 6.1 and higher

Impala Known Issues: Resources

These issues involve memory or disk usage, including out-of-memory conditions, the spill-to-disk feature, and resource management features.

Handling large rows during upgrade to CDH 5.13 / Impala 2.10 or higher

After an upgrade to CDH 5.13 / Impala 2.10 or higher, users who process very large column values (long strings), or have increased the `--read_size` configuration setting from its default of 8 MB, might encounter capacity errors for some queries that previously worked.

Resolution: After the upgrade, follow the instructions in [Handling Large Rows During Upgrade to CDH 5.13 / Impala 2.10 or Higher](#) to check if your queries are affected by these changes and to modify your configuration settings if so.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6028](#)

Configuration to prevent crashes caused by thread resource limits

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal error: Unable to allocate
section memory!
terminate called after throwing an instance of
'boost::exception_detail::clone_impl<boost::exception_detail::error_info_injector<boost::thread_resource_error>
>'
```

Workaround:

In CDH 6.0 and lower versions of CDH, configure each host running an `impalad` daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

In CDH 6.1 and higher versions, it is unlikely that you will hit the thread resource limit. Configure each host running an `impalad` daemon with the following setting:

```
echo 8000000 > /proc/sys/vm/max_map_count
```

To make the above settings durable, refer to your OS documentation. For example, on RHEL 6.x:

1. Add the following line to `/etc/sysctl.conf`:

```
vm.max_map_count=8000000
```

2. Run the following command:

```
sysctl -p
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-5605](#)

Breakpad minidumps can be very large when the thread count is high

The size of the breakpad minidump files grows linearly with the number of threads. By default, each thread adds 8 KB to the minidump size. Minidump files could consume significant disk space when the daemons have a high number of threads.

Workaround: Add `--minidump_size_limit_hint_kb=size` to set a soft upper limit on the size of each minidump file. If the minidump file would exceed that limit, Impala reduces the amount of information for each thread from 8 KB to 2 KB. (Full thread information is captured for the first 20 threads, then 2 KB per thread after that.) The minidump file can still grow larger than the "hinted" size. For example, if you have 10,000 threads, the minidump file can be more than 20 MB.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3509](#)

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the impalad daemon.

Workaround: To monitor overall memory usage, use the top command, or add the memory figures in the Impala web UI /memz tab to JVM memory usage shown on the /metrics tab.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-691](#)

Impala Known Issues: Correctness

These issues can cause incorrect or unexpected results from queries. They typically only arise in very specific circumstances.

Incorrect result due to constant evaluation in query with outer join

An OUTER JOIN query could omit some expected result rows due to a constant such as FALSE in another join clause. For example:

```
explain SELECT 1 FROM alltypestiny a1
    INNER JOIN alltypesagg a2 ON a1.smallint_col = a2.year AND false
    RIGHT JOIN alltypes a3 ON a1.year = a3.bigint_col;
+-----+
| Explain String
+-----+
| Estimated Per-Host Requirements: Memory=1.00KB Vcores=1
| 00:EMPTYSET
+-----+
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3094](#)

% escaping does not work correctly in a LIKE clause

If the final character in the RHS argument of a LIKE operator is an escaped \% character, it does not match a % final character of the LHS argument.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2422](#)

Crash: impala::Coordinator::ValidateCollectionSlots

A query could encounter a serious error if includes multiple nested levels of INNER JOIN clauses involving subqueries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2603](#)

Impala Known Issues: Interoperability

These issues affect the ability to interchange data between Impala and other systems. They cover areas such as data types and file formats.

Queries Stuck on Failed HDFS Calls and not Timing out

If the following error appears multiple times in a short duration while running a query, it would mean that the connection between the `impalad` and the HDFS NameNode is in a bad state and hence the `impalad` would have to be restarted:

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed to finish before the
<hdfs_operation_timeout_sec> second timeout "
```

Workaround: Restart the `impalad` in the bad state.

Affected Versions: All versions of Impala

Apache Issue: [HADOOP-15720](#)

Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink `hdfs.writeFormat` must be set to `Text`, rather than its default value of `Writable`. The `hdfs.writeFormat` setting must be changed to `Text` before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

Resolution: This information has been requested to be added to the upstream Flume documentation.

Affected Versions: All CDH 6 versions

Cloudera Issue: CDH-13199

Avro Scanner fails to parse some schemas

The default value in Avro schema must match the first union type. For example, if the default value is `null`, then the first type in the `UNION` must be "`null`".

Workaround: Swap the order of the fields in the schema specification. For example, use `["null", "string"]` instead of `["string", "null"]`. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-635](#)

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1024](#)

Incorrect results with basic predicate on CHAR typed column

When comparing a `CHAR` column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1652](#)

Impala Known Issues: Limitations

These issues are current limitations of Impala that require evaluation as you plan how to integrate Impala into your data management workflow.

Set limits on size of expression trees

Very deeply nested expressions within queries can exceed internal Impala limits, leading to excessive memory usage.

Workaround: Avoid queries with extremely large expression trees. Setting the query option `disable_codegen=true` may reduce the impact, at a cost of longer query runtime.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4551](#)

Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

Workaround: Use standard HDFS on all Impala nodes.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-77](#)

Hue and BDR require separate parameters for Impala Load Balancer

Cloudera Manager supports a single parameter for specifying the Impala Daemon Load Balancer. However, because BDR and Hue need to use different ports when connecting to the load balancer, it is not possible to configure the load balancer value so that BDR and Hue will work correctly in the same cluster.

Workaround: To configure BDR with Impala, use the load balancer configuration either without a port specification or with the Beeswax port.

To configure Hue, use the **Hue Server Advanced Configuration Snippet (Safety Valve)** for `impalad_flags` to specify the load balancer address with the HiveServer2 port.

Affected Versions: CDH versions from 5.11 to 6.0.1

Cloudera Issue: [OPSAPS-46641](#)

Impala Known Issues: Miscellaneous / Older Issues

These issues do not fall into one of the above categories or have not been categorized yet.

A failed CTAS does not drop the table if the insert fails

If a `CREATE TABLE AS SELECT` operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed `CREATE TABLE AS SELECT`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2005](#)

Casting scenarios with invalid/inconsistent results

Using a `CAST` function to convert large literal values to smaller types, or to convert special values such as `Nan` or `Inf`, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1821](#)

Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

Workaround: Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) for details about modifying these environment settings.

Resolution: Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-532](#)

EMC Isilon Known Issues

CDH 6.0 is not currently supported on EMC Isilon.

Affected Versions: All CDH 6 versions

Apache Kafka Known Issues

Topics Created with the "kafka-topics" Tool Might Not Be Secured

Topics that are created and deleted via Kafka are secured (for example, auto created topics). However, most topic creation and deletion is done via the `kafka-topics` tool, which talks directly to ZooKeeper or some other third-party tool that talks directly to ZooKeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. Anyone with access to ZooKeeper can create and delete topics. They will not be able to describe, read, or write to the topics even if they can create them.

The following commands talk directly to ZooKeeper and therefore are not secured via Kafka:

- `kafka-topics.sh`
- `kafka-configs.sh`
- `kafka-preferred-replica-election.sh`
- `kafka-reassign-partitions.sh`

"`offsets.topic.replication.factor`" Must Be Less Than or Equal to the Number of Live Brokers

The `offsets.topic.replication.factor` broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a `GROUP_COORDINATOR_NOT_AVAILABLE` error until the cluster size meets this replication factor requirement.

Requests Fail When Sending to a Nonexistent Topic with "`auto.create.topics.enable`" Set to True

The first few `produce` requests fail when sending to a nonexistent topic with `auto.create.topics.enable` set to true.

Workaround: Increase the number of retries in the Producer configuration setting `retries`.

Custom Kerberos Principal Names Cannot Be Used for Kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start.

Workaround: None. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Performance Degradation When SSL Is Enabled

Significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with `ssl.secure.random.implementation = SHA1PRNG`. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Affected Versions: CDK 2.x and later

Fixed Versions: None

Apache Issue: KAFKA-2561

Cloudera Issue: None

Kafka Broker Fails to Start Due to Slow Sentry and HMS startup

This issue is encountered on cluster startup and is caused by misalignment between Kafka, Sentry, and HMS. The slow startup of HMS slows down Sentry startup which consequently makes the Kafka connection to Sentry time out. Ultimately, the Kafka broker will be unable to start.

Workaround: Manually increase the number of remote procedure call retries between Sentry and Kafka through the **Sentry Client Advanced Configuration Snippet (Safety Valve)** for `sentry-site.xml` property.

1. Go to **Sentry > Configuration** and find the **Sentry Client Advanced Configuration Snippet (Safety Valve) for `sentry-site.xml`** property.
2. Click on the add button.
3. Enter the following data:
 - Name: `sentry.service.client.rpc.retry-total`
 - Value: 20
4. Enter a **Reason for change**, and then click **Save Changes** to commit the changes.
5. Return to the Home page by clicking the Cloudera Manager logo.
6. Click the restart stale services icon next to the Sentry service to invoke the cluster restart wizard.
7. Click **Restart Stale Services**.
8. Click **Restart Now**.
9. Click **Finish**.

Affected Versions: CDH 6.1.0 and higher

Fixed Versions: N/A

Cloudera Issue: CDH-74713

Connections with Expired Delegation Tokens Remain Active

Connections with expired delegation tokens stay alive even if the token expires. The connection will only terminate if the client disconnects. Once the client is disconnected it will not be able to reconnect with the expired token.

Workaround: N/A

Affected Versions: CDH 6.2.0 and higher

Fixed Versions: N/A

Apache Issue: [KAFKA-7352](#)

Cloudera Issue: N/A

Apache Kudu Known Issues

The following are known bugs and issues in Kudu. Note that this list is not exhaustive, and is meant to communicate only the most important known issues.

Timeout Possible with Log Force Synchronization Option

If the Kudu master is configured with the `-log_force_fsync_all` option, tablet servers and clients will experience frequent timeouts, and the cluster may become unusable.

Affected Versions: All CDH 6 versions

Longer Startup Times with a Large Number of Tablets

If a tablet server has a very large number of tablets, it may take several minutes to start up. It is recommended to limit the number of tablets per server to 1000 or fewer. The maximum allowed number of tablets is 2000 per server. Consider this limitation when pre-splitting your tables. If you notice slow start-up times, you can monitor the number of tablets per server in the web UI.

Affected Versions: All CDH 6 versions

Descriptions for Kudu TLS/SSL Settings in Cloudera Manager

Use the descriptions in the following table to better understand the TLS/SSL settings in the Cloudera Manager Admin Console.

Field	Usage Notes
Kerberos Principal	Set to the default principal, kudu.

Field	Usage Notes
Enable Secure Authentication And Encryption	Select this checkbox to enable authentication and RPC encryption between all Kudu clients and servers, as well as between individual servers. Only enable this property after you have configured Kerberos.
Master TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu master host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to the Kudu master web UI.
Tablet Server TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu tablet server host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to Kudu tablet server web UIs.
Master TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu master host's private key (set in Master TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Tablet Server TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu tablet server host's private key (set in Tablet Server TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Master TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Tablet Server TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Enable TLS/SSL for Master Server	Enables HTTPS encryption on the Kudu master web UI.
Enable TLS/SSL for Tablet Server	Enables HTTPS encryption on the Kudu tablet server web UIs.

Affected Versions: All CDH 6 versions

Apache Oozie Known Issues

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used due to a sys table change in PostgreSQL from version 9.5 to 9.6. The failure only happens if Oozie uses a JDBC driver earlier than 9.4.1209.

Workaround:

- After the parcels of the new version are distributed, replace the PostgreSQL JDBC driver with a newer one (version 9.4.1209 or higher) in the new parcel, at the following locations:
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/lib/
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/libtools/
- Perform the upgrade.



Note: If you already started the upgrade and the process stops with an error message about missing columns, you can change the drivers at that point of the process as well, and resume the upgrade.

If your cluster is installed from packages, you must change the drivers at the following locations:

- /usr/lib/oozie/libtools/
- /usr/lib/oozie/lib/



Note: You can change the driver after the packages installation, but before running the CDH upgrade wizard. You can also do it during the update process, when the error occurs.

You can download the driver from the [PostgreSQL JDBC driver homepage](#).

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-75951

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the `resume` command to tell Oozie to continue the workflow from the point at which it left off.

Affected Versions: CDH 5 and higher

Cloudera Issue: CDH-14623

Oozie works with MapReduce or YARN, but not both

The Oozie server works with a MapReduce (MRv1) cluster or a YARN (MRv2) cluster, but not both at the same time.

Workaround: Use two different Oozie servers.

Affected Versions: All

Apache Parquet Known Issues

There are no known issues in Parquet.

Apache Pig Known Issues

There are no known issues in this release.

Cloudera Search Known Issues

The current release includes the following known limitations:

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: `solr.sentry.enableSqlQuery=true`
- Graph: `solr.sentry.enableGraphQuery=true`
- Stream: `solr.sentry.enableStreams=true`

Workaround: None

Affected Versions: All CDH 6 releases

Cloudera Issue: CDH-66345

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search for CDH 5.5.0 includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-34050

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with CrunchIndexerTool, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-22190

The `quickstart.sh` file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if ZooKeeper and the NameNode are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the NameNode and ZooKeeper are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the MapReduce jobs and Solr commands that are run by the `quickstart`.

Affected Versions: All

Cloudera Issue: CDH-19923

Field value class guessing and Automatic schema field addition are not supported with the MapReduceIndexerTool nor the HBaseMapReduceIndexerTool

The MapReduceIndexerTool and the HBaseMapReduceIndexerTool can be used with a Managed Schema created via NRT indexing of documents or via the Solr Schema API. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the MapReduceIndexerTool or HBaseMapReduceIndexerTool. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the Solr Schema API or index sample documents using NRT indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect using the List Fields API command.

Affected Versions: All

Cloudera Issue: CDH-26856

The `Browse` and `Spell` Request Handlers are not enabled in schemaless mode

The `Browse` and `Spell` Request Handlers require certain fields be present in the schema. Since those fields cannot be guaranteed to exist in a Schemaless setup, the `Browse` and `Spell` Request Handlers are not enabled by default.

Workaround: If you require the “Browse” and “Spell” Request Handlers, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Affected Versions: All

Cloudera Issue: CDH-19407

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. Blockcache writing is disabled by default.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-17978

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given page explaining that access is denied, and instead receive a web page that never finishes loading.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-58276

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection.

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Affected Versions: All

Cloudera Issue: CDH-15441

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Affected Versions: All

Cloudera Issue: CDH-58694

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-21162

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-11357

Apache Sentry Known Issues

SHOW ROLE GRANT GROUP raises exception for a group that was never granted a role

If you run the command SHOW ROLE GRANT GROUP for a group that has never been granted a role, beeline raises an exception. However, if you run the same command for a group that does not have any roles, but has at one time been granted a role, you do not get an exception, but instead get an empty list of roles granted to the group.

Workaround: Adding a role will prevent the exception.

Affected Versions:

- CDH 5.16.0
- CDH 6.0.0

Cloudera Issue: CDH-71694

GRANT/REVOKE operations could fail if there are too many concurrent requests

Under a significant workload, Grant/Revoke operations can have issues.

Workaround: If you need to make many privilege changes, plan them at a time when you do not need to do too many at once.

Affected Versions: CDH 5.13.0 and above

Apache Issue: [SENTRY-1855](#)

Cloudera Issue: CDH-56553

Creating large set of Sentry roles results in performance problems

Using more than a thousand roles/permissions might cause significant performance problems.

Workaround: Plan your roles so that groups have as few roles as possible and roles have as few permissions as possible.

Affected Versions: CDH 5.13.0 and above

Cloudera Issue: CDH-59010

Users can't track jobs with Hive and Sentry

As a prerequisite of enabling Sentry, Hive impersonation is turned off, which means all YARN jobs are submitted to the Hive job queue, and are run as the `hive` user. This is an issue because the YARN History Server now has to block users from accessing logs for their own jobs, since their own usernames are not associated with the jobs. As a result, end users cannot access any job logs unless they can get `sudo` access to the cluster as the `hdfs`, `hive` or other admin users.

In CDH 5.8 (and higher), Hive overrides the default configuration, `mapred.job.queuename`, and places incoming jobs into the connected user's job queue, even though the submitting user remains `hive`. Hive obtains the relevant queue/username information for each job by using YARN's `fair-scheduler.xml` file.

Affected Versions: CDH 5.2.0 and above

Cloudera Issue: CDH-22890

Column-level privileges are not supported on Hive Metastore views

GRANT and REVOKE for column level privileges is not supported on Hive Metastore views.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-754](#)

SELECT privilege on all columns does not equate to SELECT privilege on table

Users who have been explicitly granted the SELECT privilege on all columns of a table, will *not* have the permission to perform table-level operations. For example, operations such as `SELECT COUNT (1)` or `SELECT COUNT (*)` will not work even if you have the SELECT privilege on all columns.

There is one exception to this. The `SELECT * FROM TABLE` command will work even if you do not have explicit table-level access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-838](#)

The EXPLAIN SELECT operation works without table or column-level privileges

Users are able to run the EXPLAIN SELECT operation, exposing metadata for all columns, even for tables/columns to which they weren't explicitly granted access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-849](#)

Cloudera Enterprise 6 Release Guide

Object types Server and URI are not supported in SHOW GRANT ROLE *roleName* on OBJECT *objectName*

Workaround: Use SHOW GRANT ROLE *roleName* to list all privileges granted to the role.

Affected Versions: All CDH versions

Apache Issue: N/A

Cloudera Issue: CDH-19430

Relative URI paths not supported by Sentry

Sentry supports only absolute (not relative) URI paths in permission grants. Although some early releases (for example, CDH 5.7.0) might not have raised explicit errors when relative paths were set, upgrading a system that uses relative paths causes the system to lose Sentry permissions.

Affected Versions: All versions. Relative paths are not supported in Sentry for permission grants.

Resolution: Revoke privileges that have been set using relative paths, and grant permissions using absolute paths before upgrading.

Absolute (Use this form)	Relative (Do not use this form)
hdfs://absolute/path/	hdfs://relative/path
s3a://bucketname/	s3a://bucketname

Apache Spark Known Issues

The following sections describe the current known issues and limitations in Apache Spark 2.x as distributed with CDH 6.1.x. In some cases, a feature from the upstream Apache Spark project is currently not considered reliable enough to be supported by Cloudera.

RDD.repartition() has different failure handling in Spark 2.4 and may cause job failures

The RDD.repartition() transformation, which reshuffles data in the RDD randomly to create either more or fewer partitions and then balances it across the partitions, was using a round-robin method to distribute data that caused incorrect answers to be returned for RDD jobs. This issue has been corrected, but it introduced a behavior change in RDD job failure handling. Now, Spark actively fails a job if there is a fetch failure that was caused by a node failure after repartitioning.

Workaround: Use the RDD.checkpoint() method to save the intermediate RDD data to HDFS. First, call `SparkContext.setCheckpointDir(directory: String)` to set the checkpoint directory where the intermediate data will be saved. Note that the directory must be an HDFS path. Then mark the RDD for checkpointing by calling `RDD.checkpoint()` when you use the RDD.repartition() transformation.

Apache Issue: [SPARK-23243](#)

Cloudera Issue: CDH-76413

PySpark broadcast variables fail when disk encryption is enabled

When disk encryption is enabled, PySpark broadcast variables fail with the following stack trace:

```
Traceback (most recent call last): File "broadcast.py", line 37, in <module>
words_new.value File "/pyspark.zip/pyspark/broadcast.py", line 137, in value
File "pyspark.zip/pyspark/broadcast.py", line 122, in load_from_path File
"pyspark.zip/pyspark/broadcast.py", line 128, in load EOFError: Ran out of input
```

Workaround: None

Affected Versions: CDH 6.0.1, CDH 6.1.0, CDH 6.2.0

Apache Issue: [SPARK-26201](#)

Cloudera Issue: CDH-76116

Structured Streaming exactly-once fault tolerance constraints

In Spark Structured Streaming, the exactly-once fault tolerance for file sink is valid only for files that are in the manifest. These files are located in the _spark_metadata subdirectory of the file sink output directory. Only process files that have file names starting with digits. Other temporary files can also appear in this directory, but they should not be processed. Typically, these temporary file file names start with a period (".").

You can list the valid manifest files, excluding the temporary files, by using a command like the following, which assumes your output directory is located at /tmp/output. As the appropriate user, run the following command to list the valid manifest files:

```
hadoop fs -ls /tmp/output/_spark_metadata/[0-9]*
```

Workaround: None

Affected Versions: CDH 6.1.0 and higher

Cloudera Issue: CDH-75191

Spark SQL does not respect size limit for the varchar type

Spark SQL treats varchar as a string (that is, there no size limit). The observed behavior is that Spark reads and writes these columns as regular strings; if inserted values exceed the size limit, no error will occur. The data will be truncated when read from Hive, but not when read from Spark.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Apache Issue: [SPARK-5918](#)

Cloudera Issue: CDH-33642

Spark SQL does not prevent you from writing key types not supported by Avro tables

Spark allows you to declare DataFrames with any key type. Avro supports only string keys and trying to write any other key type to an Avro table will fail.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33648

Spark SQL does not support timestamp in Avro tables

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33649

Spark SQL does not respect Sentry ACLs when communicating with Hive metastore

Even if user is configured via Sentry to not have read permission to a Hive table, a Spark SQL job running as that user can still read the table's metadata directly from the Hive metastore. **Cloudera Issue:** CDH-33658

Dynamic allocation and Spark Streaming

If you are using Spark Streaming, Cloudera recommends that you disable dynamic allocation by setting spark.dynamicAllocation.enabled to false when running streaming applications.

Limitation with Region Pruning for HBase Tables

When SparkSQL accesses an HBase table through the HiveContext, region pruning is not performed. This limitation can result in slower performance for some SparkSQL queries against tables that use the HBase SerDes than when the same table is accessed through Impala or Hive.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-56330

Running spark-submit with --principal and --keytab arguments does not work in client mode

The spark-submit script's --principal and --keytab arguments do not work with Spark-on-YARN's client mode.

Workaround: Use cluster mode instead.

Affected Versions: All

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Apache Sqoop Known Issues

Column names cannot start with a number when importing data with the --as-parquetfile option.

Currently, Sqoop is using an Avro schema when writing data as a parquet file. The Avro schema requires that column names do not start with numbers, therefore Sqoop is renaming the columns in this case, prepending them with an underscore character. This can lead to issues when one wants to reuse the data in other tools, such as Impala.

Workaround: Rename the columns to comply with Avro limitations (start with letters or underscore, as specified in the [Avro documentation](#)).

Cloudera Issue: None

MySQL JDBC driver shipped with CentOS 6 systems does not work with Sqoop

CentOS 6 systems currently ship with version 5.1.17 of the MySQL JDBC driver. This version does not work correctly with Sqoop.

Workaround: Install version 5.1.31 of the JDBC driver as detailed in [Installing the JDBC Drivers for Sqoop 1](#).

Affected Versions: MySQL JDBC 5.1.17, 5.1.4, 5.3.0

Cloudera Issue: CDH-23180

MS SQL Server "integratedSecurity" option unavailable in Sqoop

The integratedSecurity option is not available in the Sqoop CLI.

Workaround: None

Cloudera Issue: None

Sqoop1 (doc import + --as-parquetfile) limitation with KMS/KTS Encryption at Rest

Due to a limitation with Kite SDK, it is not possible to use (sqoop import --as-parquetfile) with KMS/KTS Encryption zones. See the following example.

```
sqoop import --connect jdbc:db2://djaxludb1001:61035/DDBAT003 --username=dh810202 --P
--target-dir /data/hive_scratch/ASDISBURSEMENT --delete-target-dir -ml --query "select
disbursementnumber,disbursementdate,xmldata FROM DB2dba.ASDISBURSEMENT where
DISBURSEMENTNUMBER = 2011113210000115311 AND \$CONDITIONS" -hive-import --hive-database
adminserver -hive-table asdisbursement_dave --map-column-java XMLDATA=String
--as-parquetfile

16/12/05 12:23:46 INFO mapreduce.Job: map 100% reduce 0%
```

```
16/12/05 12:23:46 INFO mapreduce.Job: Job job_1480530522947_0096 failed with state FAILED
due to: Job commit failed: org.kitesdk.data.DatasetIOException: Could not move contents
of
hdfs://AJAX01-ns/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096
to hdfs://AJAX01-ns/data/RetiredApps/INS/AdminServer/asdisbursement_dave
<SNIP>
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.IOException):
/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096/5ddcac42-5d69-4e46-88c2-17bedac4858.parquet
can't be moved into an encryption zone.
```

Workaround: If you use the Parquet Hadoop API based implementation for importing into Parquet, specify a `--target-dir` which is the same encryption zone as the Hive warehouse directory.

If you use the Kite Dataset API based implementation, use an alternate data file type, for example text or avro.

Apache Issue: SQUOP-2943

Cloudera Issue: CDH-40826

Doc import as Parquet files may result in out-of-memory errors

Out-of-memory (OOM) errors can be caused in the following two cases:

- With many very large rows (multiple megabytes per row) before initial-page-run check (ColumnWriter)
- When rows vary significantly by size so that the next-page-size check is based on small rows and is set very high followed by many large rows

Apache Issue: PARQUET-99

Workaround: None, other than restructuring the data.

Apache ZooKeeper Known Issues

There are no known issues in this release.

CDH 6.1.x Release Notes

To view release notes for specific CDH 6.1.x releases, see the following:

CDH 6.1.1 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for CDH 6.1.1:

What's New in CDH 6.1.1

This is a maintenance release that fixes some important issues. For details, see [Fixed Issues in CDH 6.1.1](#) on page 265.

Fixed Issues in CDH 6.1.1

The Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager.

When ResourceManager high availability is enabled the Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager. This causes the following issues in Cloudera Manager:

- If **Enable Kerberos Authentication for HTTP Web-Console** is disabled: Cloudera Manager shows statistics for the wrong server.
- If **Enable Kerberos Authentication for HTTP Web-Console** is enabled: connection from the agent to the standby fails with the `HTTPError: HTTP Error 401: Authentication required` error message. As a result, the health of the Standby Resource Manager will become bad.

Workaround: N/A

Affected Versions: CDH 6.0.x, CDH 6.1.0

Fixed Version: CDH 6.1.1

Cloudera Issue: CDH-76040

Upstream Issues Fixed

See below for issues fixed in CDH 6.1.1, grouped by component:

Apache Accumulo

There are no notable fixed issues in this release.

Apache Avro

There are no notable fixed issues in this release.

Apache Crunch

There are no notable fixed issues in this release.

Apache Flume

There are no notable fixed issues in this release.

Apache Hadoop HDFS

The following issues are fixed in CDH 6.1.1:

- [HADOOP-15717](#) - Fixed an issue where an IOException related to `tgt.getEndTime()` was not correctly logged.
- [HADOOP-15823](#) - Fixed an issue with ALDS Gen2 (ABFS) that required the user to configure the client ID and tenant ID for MSI.
- [HADOOP-15973](#) - Fixed an issue where configuration resources are not cached if they are a stream.

MapReduce 2

The following issues are fixed in CDH 6.1.1:

- [MAPREDUCE-7131](#) - Fixed a race condition where the Job History Server moves files from intermediate to finished but thinks the files are in intermediate.
- [MAPREDUCE-7156](#) - Fixed a NullPointerException when you reach the max shuffle connections.
- [MAPREDUCE-7159](#) - Enhanced the FrameworkUploader to ensure proper permissions of generated framework tar.gz if restrictive umask is used.

YARN

There are no notable fixed issues in this release.

Apache HBase

The following issues are fixed in CDH 6.1.1:

- [HBASE-21237](#) - Use CompatRemoteProcedureResolver to dispatch open/close region requests to RS
- [HBASE-21351](#) - The force update thread may have race with PE worker when the procedure is rolling back
- [HBASE-21503](#) - Replication normal source can get stuck due potential race conditions between source wal reader and wal provider initialization threads.
- [HBASE-21504](#) - If enable FIFOCompactionPolicy, a compaction may write a "empty" hfile whose maxTimeStamp is long max. This kind of hfile will never be archived.
- [HBASE-21618](#) - Scan with the same startRow(inclusive=true) and stopRow(inclusive=false) returns one result
- [HBASE-21621](#) - Reversed scan does not return expected number of rows
- [HBASE-21683](#) - Reset readsEnabled flag after successfully flushing the primary region

Apache Hive

The following issues are fixed in CDH 6.1.1:

- [HIVE-14557](#) - Nullpointer When both SkewJoin and Mapjoin Enabled
- [HIVE-20168](#) - ReduceSinkOperator Logging Hidden
- [HIVE-20169](#) - Print Final Rows Processed in MapOperator

Hue

The following issues are fixed in CDH 6.1.1:

- [HUE-8631](#) - [hbase] pull thrift transport from hbase-site.xml
- [HUE-8675](#) - [core] Fix external users created as superuser

Apache Impala

The following issues are fixed in CDH 6.1.1:

- [IMPALA-6661](#) - Treats NaN values to be equal when grouping, putting all NaN values in one group.
- [IMPALA-7777](#) - Fixed a crash due to arithmetic overflows in the Exchange Node.
- [IMPALA-5474](#) - Fixed an issue where adding a trivial subquery to a query with an error turns the error into a warning.
- [IMPALA-7939](#) - Fixed an issue in Impala Shell that would not run a valid CREATE TABLE statement when there is a word, "update", in the expression.
- [IMPALA-7960](#) - Fixed incorrect comparisons of TIMESTAMP when they were cast to shorter VARCHAR and STRING.
- [IMPALA-8026](#) - Now correctly calculates the number of rows for nested loop joins in query profiles.
- [IMPALA-7857](#) - Logs more information about the StateStore failure detection.

Apache Kafka

There are no notable fixed issues in this release.

Apache Kudu

The following issue is fixed in CDH 6.1.1:

- [KUDU-1678](#) - Fixed a rare crash caused by a race condition when a replica is shutting down while processing an alter table.

Apache Oozie

The following issues are fixed in CDH 6.1.1:

- [OOZIE-3382](#) - [SSH action] Optimize process streams draining

Apache Parquet

The following issues are fixed in CDH 6.1.1:

- [PARQUET-1305](#) - Backward incompatible change introduced in 1.8
- [PARQUET-1407](#) - Avro: Fix binary values returned from dictionary encoding
- [PARQUET-1472](#) - Dictionary filter fails on FIXED_LEN_BYTE_ARRAY

Apache Pig

The following issues are fixed in CDH 6.1.1:

- [PIG-5373](#) - InterRecordReader might skip records if certain sync markers are used
- [PIG-5374](#) - Use CircularFifoBuffer in InterRecordReader

Cloudera Search

The following issues are fixed in CDH 6.1.1:

- [SOLR-12615](#) - HashQParserPlugin won't throw an NPE for string hash key and documents with empty value
- [SOLR-12674](#) - RollupStream should not use the HashQueryParser for 1 worker

Apache Sentry

The following issues are fixed in CDH 6.1.1:

- [SENTRY-2428](#) - Skip null partitions or partitions with null sds entries
- [SENTRY-2464](#) - Catch exception thrown on first reload for UpdatableCache

Apache Spark

The following issues are fixed in CDH 6.1.1:

Cloudera Enterprise 6 Release Guide

- [SPARK-25767](#) - [SQL] Fix lazily evaluated stream of expressions in code generation
- [SPARK-26079](#) - [SQL] Ensure listener event delivery in StreamingQueryListenersConfSuite.
- [SPARK-26118](#) - [WEB UI] Introducing spark.ui.requestHeaderSize for setting HTTP requestHeaderSize
- [SPARK-26201](#) - Fix python broadcast with encryption
- [SPARK-26605](#) - [YARN] Update AM's credentials when creating tokens.
- [SPARK-26680](#) - [SQL] Eagerly create inputVars while conditions are appropriate

Apache Sqoop

There are no notable fixed issues in this release.

Apache Zookeeper

There are no notable fixed issues in this release.

Unsupported Features in CDH 6.1.1

This page lists the unsupported features in CDH 6.1.x. For the complete list of Known Issues and Limitations, see [Known Issues and Limitations in CDH 6.1.1](#) on page 305.

Apache Hadoop Unsupported Features

The following sections list unsupported features in Hadoop common components:

- [HDFS Unsupported Features](#) on page 268
- [YARN Unsupported Features](#) on page 268

HDFS Unsupported Features

The following HDFS features are not supported in CDH 6.x:

- ACLs for the NFS gateway
- Aliyun Cloud Connector
- HDFS Router Based Federation
- HDFS truncate
- More than two NameNodes
- Openstack Swift
- Quota support for Storage Types
- SFTP FileSystem
- Upgrade Domain
- Variable length block
- ZStandard Compression Codec

YARN Unsupported Features

The following YARN features are not supported in CDH 6.1.x:

- Application Timeline Server v2 (ATSV2)
- Cgroup Memory Enforcement
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Docker on YARN (DockerContainerExecutor)
- Native Services
- New Aggregated Log File Format
- Node Labels
- Pluggable Scheduler Configuration
- Reservation REST APIs
- Resource Estimator Service
- Resource Profiles

- Rolling Log Aggregation
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- YARN WebUI v2

Apache HBase Unsupported Features

The following HBase features are not supported in CDH 6.1.x:

- Master hosting meta
- Cloudera does not provide support for user-provided custom coprocessors of any kind.
- Server-side encryption of HFiles. You should configure HDFS client-side encryption.
- In-memory compaction
- Visibility labels
- Stripe compaction
- Clients setting priority on operations
- Specifying a custom asynchronous connection implementation
- Client tarball
- Serial replication
- Rolling upgrade from CDH 5 HBase versions

Apache Hive Unsupported Features

The following Hive features are not supported in CDH 6.1.x:

- AccumuloStorageHandler ([HIVE-7068](#))
- ACID ([HIVE-5317](#))
- Built-in `version()` function is not supported (CDH-40979)
- Cost-based Optimizer (CBO)
- Explicit Table Locking
- HCatalog - HBase plugin
- Hive Authorization (Instead, use Apache Sentry.)
- Hive on Apache Tez
- Hive Local Mode Execution
- Hive Metastore - Derby
- Hive Web Interface (HWI)
- HiveServer1 / JDBC 1
- HiveServer2 Dynamic Service Discovery (HS2 HA) ([HIVE-8376](#))
- HiveServer2 - HTTP Mode (Use THRIFT mode.)
- HPL/SQL ([HIVE-11055](#))
- LLAP (Live Long and Process framework)
- Scalable Dynamic Partitioning and Bucketing Optimization ([HIVE-6455](#))
- Session-level Temporary Tables ([HIVE-7090](#))
- Table Replication Across HCatalog Instances ([HIVE-7341](#))

Apache Kafka Unsupported Features

The following Kafka features are not supported in CDH 6.1.x:

- Idempotent and transactional capabilities in the producer are currently an unsupported beta feature given their maturity and complexity. This feature will be supported in a future release.
- Kafka Connect is included in CDH 6.1.0, but is not supported. Flume and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. See [Flafka: Apache Flume Meets Apache Kafka for Event Processing](#) for more information.

- Kafka Streams is included in CDH 6.0.0, but is not supported. Instead, use Spark and Spark Streaming have a fully functional ETL/stream processing pipeline. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for more information.
- The Kafka default authorizer is included in CDH 6.1.0, but is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.
- The legacy Scala clients (producer and consumer) that are under the `kafka.producer.*` and `kafka.consumer.*` package are removed in CDH 6.1.0. See [Deprecated Scala-based Client API and New Java Client API](#) on page 544.

Apache Oozie Unsupported Features

The following Oozie feature is not supported in CDH 6.1.x:

- Conditional coordinator input logic.

Apache Pig Unsupported Features

The following Pig features are not supported in CDH 6.1.x:

- Pig on Tez is not supported in CDH 6.1.x ([PIG-3446](#) / [PIG-3419](#)).
- Pig on Spark.

Cloudera Search Unsupported Features

The following Search features are not supported in CDH 6.1.x:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation
- Saving search results

Apache Sentry Unsupported Features

The following Sentry features are not supported in CDH 6.1.x:

- Import and export of Sentry metadata to and from Sentry servers
- Sentry shell command line for Hive
- Relative URI paths ([Known Issue](#))
- Object types Server and URL in

```
show grant role <role name> on object <object name>
```

- ([Known Issue](#))
- ALTER and DROP privileges for Hive and Impala

In addition, as of CDH 6.0.x, Sentry policy files have been removed. See the Sentry [Incompatible Changes](#) for more information.

Apache Spark Unsupported Features

The following Spark features are not supported in CDH 6.1.x:

- Apache Spark experimental features/APIs are not supported unless stated otherwise.
- Using the JDBC Datasource API to access Hive or Impala is not supported
- ADLS not Supported for All Spark Components. Microsoft Azure Data Lake Store (ADLS) is a cloud-based filesystem that you can access through Spark applications. Spark with Kudu is not currently supported for ADLS data. (Hive on Spark is available for ADLS in CDH 5.12 and higher.)
- IPython / Jupyter notebooks is not supported. The IPython notebook system (renamed to Jupyter as of IPython 4.0) is not supported.
- Certain Spark Streaming features not supported. The `mapWithState` method is unsupported because it is a nascent unstable API.

- Thrift JDBC/ODBC server is not supported
- Spark SQL CLI is not supported
- GraphX is not supported
- SparkR is not supported
- Structured Streaming is supported, but the following features of it *are not*:
 - Continuous processing, which is still experimental, is not supported.
 - Stream static joins with HBase have not been tested and therefore are not supported.
- Spark cost-based optimizer (CBO) not supported.

Apache Sqoop Unsupported Features

The following Sqoop feature is not supported in CDH 6.1.x:

- [import-mainframe](#)

Incompatible Changes in CDH 6.1.1



Important:

In addition to incompatible changes, CDH 6 also deprecated or removed support for several components, including Spark 1 and MapReduce v1. For information about components, sub-components, or functionality that are deprecated or no longer supported, see [Deprecated Items](#) on page 667.

See below for incompatible changes in CDH 6.1.1, grouped by component:

Apache Accumulo

CDH 6.1.1 introduces no new incompatible changes for Apache Accumulo.

Apache Avro

API Changes

One method was removed in CDH 6.0.0:

```
GenericData.toString (Object datum, StringBuilder buffer)
```

Incompatible Changes from Avro 1.8.0

- Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.
- [AVRO-997](#): Generic API requires GenericEnumSymbol - likely to break current Generic API users that often have String or Java Enum for these fields
- [AVRO-1502](#): Avro Objects now Serializable - IPC needs to be regenerated/recompiled
- [AVRO-1602](#): removed Avro internal RPC tracing, presumed unused. Current rec would be HTrace
- [AVRO-1586](#): Compile against Hadoop 2 - probably not an issue since we've been compiling against Hadoop 2 for C5.
- [AVRO-1589](#): [Java] ReflectData.AllowNulls will create incompatible Schemas for primitive types - may need a KI since it used to fail at runtime but now will fail earlier.

Apache Crunch



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

The following changes are introduced in CDH 6.0.0, and are not backward compatible:

- Crunch is available only as Maven artifacts from the Cloudera Maven repository. It is not included as part of CDH. For more information, see [Apache Crunch Guide](#).
- Crunch supports only Spark 2 and higher releases.
- Crunch supports only HBase 2 and higher releases.
 - The API methods in Crunch-HBase use HBase 2 API types and methods.

Apache Flume

AsyncHBaseSink and HBaseSink

CDH 6 uses HBase 2.0. AsyncHBaseSink is incompatible with HBase 2.0 and is not supported in CDH 6. HBaseSink has been replaced with HBase2Sink. HBase2Sink works the same way as HBaseSink. The only difference is that it is compatible with HBase 2.0. The only additional configuration required to use HBase2Sink is to replace the component name in your configuration.

For example, replace this text:

```
agent.sinks.my_hbase_sink.type = hbase
```

With this:

```
agent.sinks.my_hbase_sink.type = hbase2
```

Or, if you use the FQN of the sink class, replace this text:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase.HBaseSink
```

With this:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase2.HBase2Sink
```

For more information about how to configure HBase2Sink, see [Importing Data Into HBase](#).

com.google.common.collect.ImmutableMap

Flume has removed `com.google.common.collect.ImmutableMap` from the `org.apache.flume.Context` API and replaced it with `java.util.Map` due to Guava compatibility issues ([FLUME-2957](#)). Plugins using the `Context.getParameters()` and `Context.getSubProperties()` APIs will need to assign the return value of those methods to a `Map<String, String>` variable instead of an `ImmutableMap<String, String>` variable, if they do not already do so. Most usages in the Flume codebase already used `Map<String, String>` at the time of this change.

Apache Hadoop

- [HDFS](#) on page 272
- [MapReduce](#) on page 273
- [YARN](#) on page 274

HDFS

CDH 6.1.1 introduces no new incompatible changes for HDFS.

CDH 6.1.0 introduces no new incompatible changes for HDFS.

CDH 6.0.1 introduces no new incompatible changes for HDFS.

CDH 6.0.0 introduces the following incompatible changes for HDFS:

- HFTP has been removed.
- The S3 and S3n connectors have been removed. Users should now use the S3a connector.
- The BookkeeperJournalManager has been removed.

- Changes were made to the structure of the HDFS JAR files to better isolate clients from Hadoop library dependencies. As a result, client applications that depend on Hadoop's library dependencies may no longer work. In these cases, the client applications will need to include the libraries as dependencies directly.
- Several library dependencies were upgraded. Clients that depend on those libraries may break because the library version changes. In these cases, the client applications will need to either be ported to the new library versions or include the libraries as dependencies directly.
- [HDFS-6962](#) changes the behavior of ACL inheritance to better align with POSIX ACL specifications, which states that the umask has no influence when a default ACL propagates from parent to child. Previously, HDFS ACLs applied the client's umask to the permissions when inheriting a default ACL defined on a parent directory. Now, HDFS can ignore the umask in these cases for improved compliance with POSIX. This behavior is on by default due to the inclusion of [HDFS-11957](#). It can be configured by setting `dfs.namenode posix.acl.inheritance.enabled` in `hdfs-site.xml`. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-11957](#) changes the default behavior of ACL inheritance introduced by [HDFS-6962](#). Previously, the behavior was disabled by default. Now, the feature is enabled by default. Any code expecting the old ACL inheritance behavior will have to be updated. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-6252](#) removed `dfshealth.jsp` since it is part of the old NameNode web UI. By default, Cloudera Manager links to the new NameNode web UI, which has an equivalent health page at `dfshealth.html`.
- [HDFS-11100](#) changes the behavior of deleting files protected by a sticky bit. Now, the deletion fails.
- [HDFS-10689](#) changes the behavior of the `hdfs dfs chmod` command. Now, the command resets sticky bit permission on a file/directory when the leading sticky bit is omitted in the octal mode (like 644). When a file or directory permission is applied using octal mode and sticky bit permission needs to be preserved, then it has to be explicitly mentioned in the permission bits (like 1644).
- [HDFS-10650](#) changes the behavior of `DFSClient#mkdirs` and `DFSClient#primitiveMkdir`. Previously, they create a new directory with the default permissions 00666. Now, they will create a new directory with permission 00777.
- [HADOOP-8143](#) changes the default behavior of `distcp`. Previously, the `-pb` option was not used by default, which may have caused some checksums to fail when block sizes did not match. Now, the `-pb` option is included by default to preserve block size when using `distcp`.
- [HADOOP-10950](#) changes several heap management variables:
 - `HADOOP_HEAPSIZE` variable has been deprecated. Use `HADOOP_HEAPSIZE_MAX` and `HADOOP_HEAPSIZE_MIN` instead to set `Xmx` and `Xms`
 - The internal variable `JAVA_HEAP_MAX` has been removed.
 - Default heap sizes have been removed. This will allow for the JVM to use auto-tuning based upon the memory size of the host. To re-enable the old default, configure `HADOOP_HEAPSIZE_MAX="1g"` in `hadoop-env.sh`.
 - All global and daemon-specific heap size variables now support units. If the variable is only a number, the size is assumed to be in megabytes.
- [HADOOP-14426](#) upgrades the version of Kerby from 1.0.0-RC2 to 1.0.0
- [HDFS-10970](#) updates the version of Jackson from 1.9.13 to 2.x in `hadoop-hdfs`.
- [HADOOP-9613](#) updates the Jersey version to the latest 1.x release.
- [HADOOP-10101](#) updates Guava dependency to 21.0
- [HADOOP-14225](#) removes the `xmlenc` dependency. If you rely on the transitive dependency, you need to set the dependency explicitly in your code after this change.
- [HADOOP-13382](#) remove unneeded `commons-httpclient` dependencies from POM files in Hadoop and sub-projects. This incompatible change may affect projects that have undeclared transitive dependencies on `commons-httpclient`, which used to be provided by `hadoop-common` or `hadoop-client`.
- [HADOOP-13660](#) upgrades the `commons-configuration` version from 1.6 to 2.1.
- [HADOOP-12064](#) upgrades the following dependencies:
 - Guice from 3.0 to 4.0
 - cglib from 2.2 to 3.2.0
 - asm from 3.2 to 5.0.4

CDH 6.1.1 introduces no new incompatible changes for MapReduce.

CDH 6.1.0 introduces no new incompatible changes for MapReduce.

CDH 6.0.1 introduces no new incompatible changes for MapReduce.

CDH 6.0.0 introduces the following incompatible changes:

- Support for MapReduce v1 has been dropped from CDH 6.0.0.
- CDH 6 supports applications compiled against CDH 5.7.0 and higher MapReduce frameworks. Make sure to not to include the CDH jars with your application by marking them as "provided" in the `pom.xml` file.

YARN

CDH 6.1.1 introduces no new incompatible changes for YARN.

CDH 6.1.0 introduces no new incompatible changes for MapReduce.

CDH 6.0.1 introduces no new incompatible changes for YARN.

CDH 6.0.0 introduces no new incompatible changes for YARN.

Apache HBase

CDH 6.1.x contains the following downstream HBase incompatible change:

`hbase.security.authorization`

The default value for `hbase.security.authorization` has been changed from true to false. Secured clusters should make sure to explicitly set it to true in XML configuration file before upgrading to one of these versions ([HBASE-19483](#)). True as the default value of `hbase.security.authorization` was changed because not all clusters need authorization. (History: [HBASE-13275](#)) Rather, only the clusters which need authorization should set this configuration as true.

Incompatible Changes

For more information about upstream incompatible changes, see the Apache Reference Guide [Incompatible Changes](#) and [Upgrade Paths](#).

CDH 6.1.0 introduces the following incompatible changes for HBase:

- [HBASE-20270](#): Error triggered command help is no longer available.
- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.

CDH 6.0.x introduces the following upstream HBase incompatible changes:

- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.
- Public interface API changes:
 - [HBASE-15607](#): Admin
 - [HBASE-19112](#), [HBASE-18945](#): Cell
 - Region, Store, HBaseTestingUtility
- [HBASE-18792](#): hbase-2 needs to defend against hbck operations
- [HBASE-15982](#): Interface ReplicationEndpoint extends Guava's Service.
- [HBASE-18995](#): Split CellUtil into public CellUtil and PrivateCellUtil for Internal use only.
- [HBASE-19179](#): Purged the hbase-prefix-tree module and all references from the code base.
- [HBASE-17595](#): Add partial result support for small/limited scan; Now small scan and limited scan could also return partial results.
- [HBASE-16765](#): New default split policy, SteppingSplitPolicy.
- [HBASE-17442](#): Move most of the replication related classes from hbase-client to hbase-replication package.
- [HBASE-16196](#): The bundled JRuby 1.6.8 has been updated to version 9.1.9.0.
- [HBASE-18811](#): Filters have been moved from Public to LimitedPrivate.
- [HBASE-18697](#): Replaced hbase-shaded-server jar with hbase-shaded-mapreduce jar.
- [HBASE-18640](#): Moved mapreduce related classes out of hbase-server into separate hbase-mapreduce jar .

- [HBASE-19128](#): Distributed Log Replay feature has been removed.
- [HBASE-19176](#): Hbase-native-client has been removed.
- [HBASE-17472](#): Changed semantics of granting new permissions. Earlier, new grants would override previous permissions, but now, the new and existing permissions get merged.
- [HBASE-18374](#): Previous "mutate" latency metrics has been renamed to "put" metrics.
- [HBASE-15740](#): Removed Replication metric source.shippedKBs in favor of source.shippedBytes.
- [HBASE-13849](#): Removed restore and clone snapshot from the WebUI.
- [HBASE-13252](#): The concept of managed connections in HBase (deprecated before) has now been extinguished completely, and now all callers are responsible for managing the lifecycle of connections they acquire.
- [HBASE-14045](#): Bumped thrift version to 0.9.2.
- [HBASE-5401](#): Changes to number of tasks PE runs when clients are mapreduce. Now tasks == client count. Previous we hardcoded ten tasks per client instance.

Changed Behavior

CDH 6.1.x contains the following HBase behavior changes:

- [HBASE-14350](#): Assignment Manager v2 - Split/Merge have moved to the Master; it runs them now. Hooks around Split/Merge are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-18271](#): Moved to internal shaded netty.
- [HBASE-17343](#): Default MemStore to be CompactingMemStore instead of DefaultMemStore. In-memory compaction of CompactingMemStore demonstrated sizable improvement in HBase's write amplification and read/write performance.
- [HBASE-19092](#): Make Tag IA.LimitedPrivate and expose for CPs.
- [HBASE-18137](#): Replication gets stuck for empty WALs.
- [HBASE-17513](#): Thrift Server 1 uses different QOP settings than RPC and Thrift Server 2 and can easily be misconfigured so there is no encryption when the operator expects it.
- [HBASE-16868](#): Add a replicate_all flag to replication peer config. The default value is true, which means all user tables (REPLICATION_SCOPE != 0) will be replicated to peer cluster.
- [HBASE-19341](#): Ensure Coprocessors can abort a server.
- [HBASE-18469](#): Correct RegionServer metric of totalRequestCount.
- [HBASE-17125](#): Marked Scan and Get's setMaxVersions() and setMaxVersions(int) as deprecated. They are easy to misunderstand with column family's max versions, so use readAllVersions() and readVersions(int) instead.
- [HBASE-16567](#): Core is now up on protobuf 3.1.0 (Coprocessor Endpoints and REST are still on protobuf 2.5.0).
- [HBASE-14004](#): Fix inconsistency between Memstore and WAL which may result in data in remote cluster that is not in the origin (Replication).
- [HBASE-18786](#): FileNotFoundException opening a StoreFile in a primary replica now causes a RegionServer to crash out where before it would be ignored (or optionally handled via close/reopen).
- [HBASE-17956](#): Raw scans will also read TTL expired cells.
- [HBASE-17017](#): Removed per-region latency histogram metrics.
- [HBASE-19483](#): Added ACL checks to RSGroup commands - On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands.
- [HBASE-19358](#): Added ACL checks to RSGroup commands (HBASE-19483): On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands. Improved stability of splitting log when do failover.
- [HBASE-18883](#): Updated our Curator version to 4.0 - Users who experience classpath issues due to version conflicts are recommended to use either the hbase-shaded-client or hbase-shaded-mapreduce artifacts.
- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.
- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.
- [HBASE-15968](#): This issue resolved two long-term issues in HBase: 1) Puts may be masked by a delete before them, and 2) Major compactions change query results. Offers a new behavior to fix this issue with a little performance reduction. Disabled by default. See the issue for details and caveats.

- [HBASE-13701](#): SecureBulkLoadEndpoint has been integrated into HBase core as default bulk load mechanism. It is no longer needed to install it as a coprocessor endpoint.
- [HBASE-9774](#): HBase native metrics and metric collection for coprocessors.
- [HBASE-18294](#): Reduce global heap pressure: flush based on heap occupancy.

Apache Hive/Hive on Spark/HCatalog

Apache Hive

The following changes are introduced to Hive in CDH 6.0.0, and are not backwards compatible:

- [UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting](#) on page 533
- [Support for UNION DISTINCT](#) on page 534
- [OFFLINE and NO_DROP Options Removed from Table and Partition DDL](#) on page 534
- [DESCRIBE Query Syntax Change](#) on page 535
- [CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names](#) on page 535
- [Reserved and Non-Reserved Keyword Changes in HiveQL](#) on page 535
- [Apache Hive API Changes in CDH 6.0.0](#) on page 537
- [Apache Hive Configuration Changes in CDH 6.0.0](#) on page 538
- [HiveServer2 Thrift API Repackaged Resulting in Class File Location Changes](#) on page 540
- [Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column](#) on page 541
- [Hive Logging Framework Switched to SLF4J/Log4j 2](#) on page 541
- [Deprecated Parquet Java Classes Removed from Hive](#) on page 541
- [Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms](#) on page 541
- [S3N Connector Is Removed from CDH 6.0](#) on page 542
- [Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request](#) on page 542
- [Support Added for Escaping Carriage Returns and New Line Characters for Text Files \(LazySimpleSerDe\)](#) on page 542
- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 542

Changing Table File Format from ORC with the ALTER TABLE Command Not Supported in CDH 6

Changing the table file format from ORC to another file format with the ALTER TABLE command is not supported in CDH 6 (it returns an error).

UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting

Prior to this change, Hive performed implicit casts when data types from different type groups were specified in queries that use UNION_ALL. For example, before CDH 6.0, if you had the two following tables:

Table "one"

one.col_1	one.col_2	one.col_3
21	hello_all	b

Where col_1 datatype is int, col_2 datatype is string, and col_3 datatype is char(1).

Table "two"

two.col_4	two.col_5	two.col_6
75.0	abcde	45

```
+-----+-----+-----+-----+
```

Where `col_4` datatype is `double`, `col_5` datatype is `varchar(5)`, and `col_6` datatype is `int`.

And you ran the following `UNION ALL` query against these two tables:

```
SELECT * FROM one UNION ALL SELECT col_4 AS col_1, col_5 AS col_2, col_6 AS col_3 FROM two;
```

You received the following result set:

_u1.col_1	_u1.col_2	_u1.col_3
75.0	abcde	4
21.0	hello	b

Note that this statement implicitly casts the values from table `one` with the following errors resulting in data loss:

- `one.col_1` is cast to a `double` datatype
- `one.col_2` is cast to a `varchar(5)` datatype, which truncates the original value from `hello_all` to `hello`
- `one.col_3` is cast to a `char(1)` datatype, which truncates the original value from `45` to `4`

In CDH 6.0, no implicit cast is performed across different type groups. For example, `STRING`, `CHAR`, and `VARCHAR` are in one type group, and `INT`, `BIGINT`, and `DECIMAL` are in another type group, and so on. So, in CDH 6.0 and later, the above query that uses `UNION ALL`, returns an exception for the columns that contain datatypes that are not part of a type group. In CDH 6.0 and later, Hive performs the implicit cast only *within* type groups and not *across* different type groups. For more information, see [HIVE-14251](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
- The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
- The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

OFFLINE and NO_DROP Options Removed from Table and Partition DDL

Support for Hive table and partition protection options have been removed in CDH 6.0, which includes removal of the following functionality:

- Support has been removed for:
 - ENABLE | DISABLE NO_DROP [CASCADE]
 - ENABLE | DISABLE OFFLINE
 - ALTER TABLE ... IGNORE PROTECTION

- The following support has also been removed from the `HiveMetastoreClient` class:

The `ignoreProtection` parameter has been removed from the `dropPartitions` methods in the `IMetaStoreClient` interface.

For more information, see [HIVE-11145](#).

Cloudera recommends that you use Apache Sentry to replace most of this functionality. Although Sentry governs permissions on `ALTER TABLE`, it does not include permissions that are specific to a partition. See [Authorization Privilege Model for Hive and Impala](#) and [Configuring the Sentry Service](#).

DESCRIBE Query Syntax Change

In CDH 6.0 syntax has changed for `DESCRIBE` queries as follows:

- `DESCRIBE` queries where the column name is separated by the table name using a period is no longer supported:

```
DESCRIBE testTable.testColumn;
```

Instead, the table name and column name must be separated with a space:

```
DESCRIBE testTable testColumn;
```

- The `partition_spec` must appear *after* the table name, but *before* the optional column name:

```
DESCRIBE default.testTable PARTITION (part_col = 100) testColumn;
```

For more details, see the [Apache wiki](#) and [HIVE-12184](#).

CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names

In CDH 6.0, `CREATE TABLE` statements fail if any of the specified column names contain a period or a colon. For more information, see [HIVE-10120](#) and the [Apache wiki](#).

Reserved and Non-Reserved Keyword Changes in HiveQL

Hive reserved and non-reserved keywords have changed in CDH 6.0. *Reserved keywords* cannot be used as table or column names unless they are enclosed with back ticks (for example, `data`). *Non-reserved keywords* can be used as table or column names without enclosing them with back ticks. Non-reserved keywords have proscribed meanings in HiveQL, but can still be used as table or column names. For more information about the changes to reserved and non-reserved words listed below, see [HIVE-6617](#) and [HIVE-14872](#).

In CDH 6.0, the following changes have been introduced to Hive reserved and non-reserved keywords and are not backwards compatible:

- [Hive New Reserved Keywords Added in CDH 6.0](#) on page 535
- [Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0](#) on page 536
- [Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0](#) on page 536
- [Hive New Non-Reserved Keywords Added in CDH 6.0](#) on page 536
- [Hive Non-Reserved Keyword Removed in CDH 6.0](#) on page 537

Hive New Reserved Keywords Added in CDH 6.0

The following table contains new reserved keywords that have been added:

COMMIT	CONSTRAINT	DEC	EXCEPT
FOREIGN	INTERVAL	MERGE	NUMERIC
ONLY	PRIMARY	REFERENCES	ROLLBACK
START			

Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0

The following table contains non-reserved keywords that have been converted to be reserved keywords:

ALL	ALTER	ARRAY	AS
AUTHORIZATION	BETWEEN	BIGINT	BINARY
BOOLEAN	BOTH	BY	CREATE
CUBE	CURSOR	DATE	DECIMAL
DOUBLE	DELETE	DESCRIBE	DROP
EXISTS	EXTERNAL	FALSE	FETCH
FLOAT	FOR	FULL	GRANT
GROUP	GROUPING	IMPORT	IN
INT	INNER	INSERT	INTERSECT
INTO	IS	LATERAL	LEFT
LIKE	LOCAL	NONE	NULL
OF	ORDER	OUT	OUTER
PARTITION	PERCENT	PROCEDURE	RANGE
READS	REGEXP	REVOKE	RIGHT
RLIKE	ROLLUP	ROW	ROWS
SET	SMALLINT	TABLE	TIMESTAMP
TO	TRIGGER	TRUNCATE	UNION
UPDATE	USER	USING	VALUES
WITH	TRUE		

Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0

The following table contains reserved keywords that have been converted to be non-reserved keywords:

CURRENT_DATE	CURRENT_TIMESTAMP	HOLD_DDLTIME	IGNORE
NO_DROP	OFFLINE	PROTECTION	READONLY

Hive New Non-Reserved Keywords Added in CDH 6.0

The following table contains new non-reserved keywords that have been added:

ABORT	AUTOCOMMIT	CACHE	DAY
DAYOFWEEK	DAYS	DETAIL	DUMP
EXPRESSION	HOUR	HOURS	ISOLATION

KEY	LAST	LEVEL	MATCHED
MINUTE	MINUTES	MONTH	MONTHS
NORELY	NOVALIDATE	NULLS	OFFSET
OPERATOR	RELY	SECOND	SECONDS
SNAPSHOT	STATUS	SUMMARY	TRANSACTION
VALIDATE	VECTORIZATION	VIEWS	WAIT
WORK	WRITE	YEAR	YEARS

Hive Non-Reserved Keyword Removed in CDH 6.0

The following non-reserved keyword has been removed:

DEFAULT

Apache Hive API Changes in CDH 6.0.0

The following changes have been introduced to the Hive API in CDH 6.0, and are not backwards compatible:

- [AddPartitionMessage.getPartitions\(\) Can Return NULL](#) on page 537
- [DropPartitionEvent and PreDropPartitionEvent Class Changes](#) on page 537
- [GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date](#) on page 537
- [GenericUDF.getConstantLongValue Has Been Removed](#) on page 537
- [Increased Width of Hive Metastore Configuration Columns](#) on page 537

AddPartitionMessage.getPartitions() Can Return NULL

The `getPartitions()` method has been removed from the `AddPartitionEvent` class in the `org.apache.hadoop.hive.metastore.events` interface. It was removed to prevent out-of-memory errors when the list of partitions is too large.

Instead use the `getPartitionIterator()` method. For more information, see [HIVE-9609](#) and the [AddPartitionEvent documentation](#).

DropPartitionEvent and PreDropPartitionEvent Class Changes

The `getPartitions()` method has been removed and replaced by the `getPartitionIterator()` method in the `DropPartitionEvent` class and the `PreDropPartitionEvent` class.

In addition, the `(Partition partition, boolean deleteData, HiveMetastore.HMSHandler handler)` constructors have been deleted from the `PreDropPartitionEvent` class. For more information, see [HIVE-9674](#) and the [PreDropPartitionEvent documentation](#).

GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date

The `getTimestampValue` method in the `GenericUDF` class now returns a `TIMESTAMP` value instead of a `DATE` value. For more information, see [HIVE-10275](#) and the [GenericUDF documentation](#).

GenericUDF.getConstantLongValue Has Been Removed

The `getConstantLongValue` method has been removed from the `GenericUDF` class. It has been noted by the community that this method is not used in Hive. For more information, see [HIVE-10710](#) and the [GenericUDF documentation](#).

Increased Width of Hive Metastore Configuration Columns

The columns used for configuration values in the Hive metastore have been increased in width, resulting in the following incompatible changes in the `org.apache.hadoop.hive.metastore.api` interface.

This change introduced an incompatible change to the `get_table_names_by_filter` method of the `ThriftHiveMetastore` class. Before this change, this method accepts a string filter, which allows clients to filter a table by its TABLEPROPERTIES value. For example:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 <> \"yellow\"";  
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 = \"yellow\"";
```

After this change, the `TABLE_PARAMS.PARAM_VALUE` column is now a `CLOB` data type. Depending on the type of database that you use (for example, MySQL, Oracle, or PostgreSQL), the semantics may have changed and operators like `=`, `<>`, and `!=` might not be supported. Refer to the documentation for your database for more information. You must use operators that are compatible with `CLOB` data types. There is no equivalent `<>` operator that is compatible with `CLOB`. So there is no equivalent operator for the above example that uses the `<>` inequality operator. The equivalent for `=` is the `LIKE` operator so you would rewrite the second example above as:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 LIKE \"yellow\"";
```

For more information, see [HIVE-12274](#).

Apache Hive Configuration Changes in CDH 6.0.0

The following configuration property changes have been introduced to Hive in CDH 6.0, and are not backwards compatible:

- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 538
- [Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters](#) on page 538
- [Hive Strict Checks Have Been Re-factored To Be More Granular](#) on page 539
- [Java XML Serialization Has Been Removed](#) on page 539
- [Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties](#) on page 540
- [HiveServer2 Impersonation Property \(`hive.server2.enable.impersonation`\) Removed](#) on page 540
- [Changed Default File Format for Storing Intermediate Query Results](#) on page 540

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on `hive.enforce.bucketing`](#) and the [topic on `hive.enforce.sorting`](#).

Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters

Directories in HDFS can contain unprintable or unsupported characters that are not visible even when you run the `hadoop fs -ls` command on the directories. When external tables are created with the `MSCK REPAIR TABLE` command, the partitions using these HDFS directories that contain unsupported characters are unusable for Hive. To avoid this, the configuration parameter `hive.msck.path.validation` has been added. This configuration property controls the behavior of the `MSCK REPAIR TABLE` command, enabling you to set whether validation checks are run on the HDFS directories when `MSCK REPAIR TABLE` is run.

The property `hive.msck.path.validation` can be set to one of the following values:

Value Name	Description
throw	Causes Hive to throw an exception when it tries to process an HDFS directory that contains unsupported characters

Value Name	Description
	with the MSCK REPAIR TABLE command. This is the default setting for <code>hive.msck.path.validation</code> .
skip	Causes Hive to skip the skip the directories that contain unsupported characters, but still repairs the others.
ignore	Causes Hive to completely skip any validation of HDFS directories when the MSCK REPAIR TABLE command is run. This setting can cause bugs because unusable partitions are created.

By default, the `hive.msck.path.validation` property is set to `throw`, which causes Hive to throw an exception when MSCK REPAIR TABLE is run and HDFS directories containing unsupported characters are encountered. To work around this, set this property to `skip` until you can repair the HDFS directories that contain unsupported characters.

To set this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the **Configuration** tab.
3. Search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting.
4. In the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting, add the **Name** of the property, the **Value** (`throw`, `skip`, or `ignore`), and a **Description** of the setting.
5. Click **Save Changes** and restart the service.

For more information, see [HIVE-10722](#).

Hive Strict Checks Have Been Re-factored To Be More Granular

Originally, the configuration property `hive.mapred.mode` was added to restrict certain types of queries from running. Now it has been broken down into more fine-grained configurations, one for each type of restricted query pattern. The configuration property `hive.mapred.mode` has been removed and replaced with the following configuration properties, which provide more granular control of Hive strict checks:

Configuration Property	Description	Default Value
<code>hive.strict.checks.bucketing</code>	When set to <code>true</code> , running LOAD DATA queries against bucketed tables is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.type.safety</code>	When set to <code>true</code> , comparing bigint to string data types or bigint to double data types is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.orderby.no.limit</code>	When set to <code>true</code> , prevents queries from being run that contain an ORDER BY clause with no LIMIT clause.	<code>false</code>
<code>hive.strict.checks.no.partition.filter</code>	When set to <code>true</code> , prevents queries from being run that scan a partitioned table but do not filter on the partition column.	<code>false</code>
<code>hive.strict.checks.cartesian.product</code>	When set to <code>true</code> , prevents queries from being run that contain a Cartesian product (also known as a cross join).	<code>false</code>

All of these properties can be set with Cloudera Manager in the following configuration settings for the Hive service:

- **Restrict LOAD Queries Against Bucketed Tables** (`hive.strict.checks.bucketing`)

- **Restrict Unsafe Data Type Comparisons** (`hive.strict.checks.type.safety`)
- **Restrict Queries with ORDER BY but no LIMIT clause** (`hive.strict.checks.orderby.no.limit`)
- **Restrict Partitioned Table Scans with no Partitioned Column Filter**
(`hive.strict.checks.no.partition.filter`)
- **Restrict Cross Joins (Cartesian Products)** (`hive.strict.checks.cartesian.product`)

For more information about these configuration properties, see [HIVE-12727](#), [HIVE-15148](#), [HIVE-18251](#), and [HIVE-18552](#).

Java XML Serialization Has Been Removed

The configuration property `hive.plan.serialization.format` has been removed. Previously, this configuration property could be set to either `javaXML` or `kryo`. Now the default is `kryo` serialization, which cannot be changed. For more information, see [HIVE-12609](#) and the [Apache wiki](#).

Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties

The configuration property `hive.groupby.orderby.position.alias`, which enabled using column position with the `GROUP BY` and the `ORDER BY` clauses has been removed and replaced with the following two configuration properties. These configuration properties enable using column position with `GROUP BY` and `ORDER BY` separately:

Configuration Property Name	Description/Default Setting	Possible Values
<code>hive.groupby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>GROUP BY</code> clauses in queries. Default Setting: <code>false</code> . This behavior is turned off by default.	<code>true</code> <code>false</code>
<code>hive.orderby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>ORDER BY</code> clauses in queries. Default Setting: <code>true</code> . This behavior is turned on by default.	<code>true</code> <code>false</code>

For more information, see [HIVE-15797](#) and the Apache wiki entries for [configuration properties](#), [GROUP BY syntax](#), and [ORDER BY syntax](#).

HiveServer2 Impersonation Property (`hive.server2.enable.impersonation`) Removed

In earlier versions of CDH, the following two configuration properties could be used to set impersonation for HiveServer2:

- `hive.server2.enable.impersonation`
- `hive.server2.enable.doAs`

In CDH 6.0, `hive.server2.enable.impersonation` is removed. To configure impersonation for HiveServer2, use the configuration property `hive.server2.enable.doAs`. To set this property in Cloudera Manager, select the Hive service and click on the **Configuration** tab. Then search for the **HiveServer2 Enable Impersonation** setting and select the checkbox to enable HiveServer2 impersonation. This property is enabled by default in CDH 6.

For more information about this property, see the [Apache wiki documentation for HiveServer2 configuration properties](#).

Changed Default File Format for Storing Intermediate Query Results

The configuration property `hive.query.result.fileformat` controls the file format in which a query's intermediate results are stored. In CDH 6, the default setting for this property has been changed from `TextFile` to `SequenceFile`.

To change this configuration property in Cloudera Manager:

1. In the Admin Console, select the Hive service and click on the **Configuration** tab.
2. Then search for the **Hive Service Advanced Configuration Snippet (Safety Valve) for `hive-site.xml`** setting and add the following information:
 - **Name:** `hive.query.result.fileformat`
 - **Value:** Valid values are `TextFile`, `SequenceFile` (default), or `RCfile`

- **Description:** Sets the file format in which a query's intermediate results are stored.

3. After you add this information, click **Save Changes** and restart the Hive service.

For more information about this parameter, see the [Apache wiki](#).

HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes

HiveServer2 Thrift API code has been repackaged in CDH 6.0, resulting in the following changes:

- All files generated by the Thrift API for HiveServer2 have moved from the following *old* namespace:

`org.apache.hive.service.cli.thrift`

To the following *new* namespace:

`org.apache.hive.service.rpc.thrift`

- All files generated by the Thrift API for HiveServer2 have moved into a separate jar file called `service-rpc`.

As a result of these changes, all Java classes such as `TCLIService.java`, `TOpenSessionReq.java`, `TSessionHandle.java`, and `TGetSchemasReq.java` have changed locations. For more information, see [HIVE-12442](#).

Values Returned for Decimal Numbers Are Now Padded with Trailing Zeros to the Scale of the Specified Column

Decimal values that are returned in query results are now padded with trailing zeroes to match the specified scale of the corresponding column. For example, *before* this change, when Hive read a decimal column with a specified scale of 5, the value returned for zero was returned as 0. *Now*, the value returned for zero is 0.00000. For more information, see [HIVE-12063](#).

Hive Logging Framework Switched to SLF4J/Log4j 2

The logging framework for Hive has switched to [SLF4J \(Simple Logging Facade for Java\)](#) and now uses [Log4j 2](#) by default. Use of Log4j 1.x, Apache Commons Logging, and `java.util.logging` have been removed. To accommodate this change, write all Log4j configuration files to be compatible with Log4j 2.

For more information, see [HIVE-12237](#), [HIVE-11304](#), and the [Apache wiki](#).

Deprecated Parquet Java Classes Removed from Hive

The deprecated `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` have been removed from Hive because they resided outside of the `org.apache` namespace. Any existing tables that use these classes are automatically migrated to the new SerDe classes when the metastore is upgraded.

Use one of the following options for specifying the Parquet SerDe for new Hive tables:

- Specify in the `CREATE TABLE` statement that you want it stored as Parquet. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING) STORED AS PARQUET;
```

- Set the `INPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat` and set the `OUTPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat`. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING)
STORED AS
  INPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat"
  OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat";
```

For more information, see [HIVE-6757](#) and the [Apache wiki](#).

Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms

Support for JDBC, counter-based, and HBase-based statistics collection mechanisms has been removed from Hive. The following configuration properties are no longer supported:

- `hive.stats.dbclass`
- `hive.stats.retries.wait`
- `hive.stats.retries.max`
- `hive.stats.jdbc.timeout`
- `hive.stats.dbconnectionstring`
- `hive.stats.jdbcdrive`
- `hive.stats.key.prefix.reserve.length`

This change also removed the `cleanUp(String keyPrefix)` method from the [StatsAggregator interface](#).

Now all Hive statistics are collected on the default file system. For more information, see [HIVE-12164](#), [HIVE-12411](#), [HIVE-12005](#), and the [Apache wiki](#).

S3N Connector Is Removed from CDH 6.0

The [S3N connector](#), which is used to connect to the Amazon S3 file system from Hive has been removed from CDH 6.0. To connect to the S3 file system from Hive in CDH 6.0, you must now use the S3A connector. There are a number of differences between the S3N and the S3A connectors, including configuration differences. See the [Apache wiki page on integrating with Amazon Web Services](#) for details.

Migration involves making the following changes:

- Changing all metastore data containing URIs that start with `s3n://` to `s3a://`. This change is performed automatically when you upgrade the Hive metastore.
- Changing all scripts containing URIs that start with `s3n://` to `s3a://`. You must perform this change manually.

Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request

Six additional columns have been added to the `TRowSet` that is returned by the `TCLIService#GetTables` request. These columns were added to comply with the official JDBC API. For more information, see the documentation for [java.sql.DatabaseMetaData](#).

The columns added are:

Column Name	Description
REMARKS	Explanatory comment on the table.
TYPE_CAT	Types catalog.
TYPE_SCHEMA	Types schema.
TYPE_NAME	Types name.
SELF_REFERENCING_COL_NAME	Name of the designed identifier column of a typed table.
REF_GENERATION	Specifies how values in the <code>SELF_REFERENCING_COL_NAME</code> column are created.

For more information, see [HIVE-7575](#).

Support Added for Escaping Carriage Returns and New Line Characters for Text Files (LazySimpleSerDe)

Support has been added for escaping carriage returns and new line characters in text files by modifying the `LazySimpleSerDe` class. Without this change, carriage returns and new line characters are interpreted as delimiters, which causes incorrect query results.

This feature is controlled by the SerDe property `serialization.escape.crlf`. It is enabled (set to `true`) by default. If `serialization.escape.crlf` is enabled, '`r`' or '`n`' cannot be used as separators or field delimiters.

This change only affects text files and removes the `getNullString` method from the [LazySerDeParameters class](#). For more information, see [HIVE-11785](#).

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on `hive.enforce.bucketing`](#) and the [topic on `hive.enforce.sorting`](#).

Hue

There are no incompatible changes in this release.

Apache Impala

There are no incompatible changes in this release.

Apache Kafka

Incompatible Changes Introduced in CDH 6.1.1

CDH 6.1.1 introduces no new incompatible changes for Kafka.

Incompatible Changes Introduced in CDH 6.1.0

Scala-based Client API Removed

Scala-based clients were deprecated in a previous release and are removed as of CDH 6.1.0.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are effected:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Default Behaviour Change

[KAFKA-7050](#): The default value for `request.timeout.ms` is decreased to 30 seconds. In addition, a new logic is added that makes the `JoinGroup` requests ignore this timeout.

Incompatible Changes Introduced in CDH 6.0.1

CDH 6.0.1 introduces no new incompatible changes for Kafka.

Incompatible Changes Introduced in CDH 6.0.0

Kafka is now bundled as part of CDH. The following sections describe incompatible changes between the previous, separately installed Kafka (CDK powered by Apache Kafka version 3.1) and the CDH 6.0.0 Kafka version. These changes affect clients built with CDH 6.0.0 libraries. Cloudera recommends upgrading clients to the new release; however clients built with previous versions of Kafka will continue to function.

Packaging

CDH and previous distributions of Kafka (CDK Powered by Apache Kafka) cannot coexist in the same cluster.

Deprecated Scala-based Client API and New Java Client API

Scala-based clients are deprecated in this release and will be removed in an upcoming release.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are deprecated and unsupported as of CDH 6.0.0:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Command Line Options Removed

Some command line tools are affected by the deprecation of old clients (see the previous entry). The following options have been removed and are not recognized as valid options:

- `--new-consumer`
- `--old-consumer`
- `--old-producer`

The tools affected use the new clients.

Command Line Tools Removed

The following command line tools and runnable classes are removed:

- `kafka-replay-log-producer`
- `kafka-simple-consumer-shell`
- `kafka.tools.ReplayLogProducer`
- `kafka.tools.SimpleConsumerShell`
- `kafka.tools.ExportZkOffset`
- `kafka.tools.ImportZkOffset`
- `kafka.tools.SimpleConsumerPerformance`
- `kafka.tools.UpdateOffsetsInZK`
- `kafka.tools.VerifyConsumerRebalance`
- `kafka.tools.ProducerPerformance`

Consumer API Changes

Consumer methods invoked with unassigned partitions now raise an `IllegalStateException` instead of an `IllegalArgumentException`.

Previous versions of the Consumer method `poll(long)` would wait for metadata updates regardless of timeout parameter. This behavior is expected to change in future releases; make sure your client applications include an appropriate timeout parameter and do not rely on the previous behavior.

Exception Classes Removed

The following exceptions were deprecated in a previous release and are not thrown anymore are removed:

- `GroupCoordinatorNotAvailableException`
- `GroupLoadInProgressException`
- `NotCoordinatorForGroupException`
- `kafka.common.KafkaStorageException`

Metrics Updated

Kafka consumers' per-partition metrics were changed to use tags for topic and partition rather than the metric name. For more information see [KIP-225](#).

Apache Kudu

Client Library Compatibility

- The Kudu 1.8 Java client library is API- and ABI-compatible with Kudu 1.7. Applications written against Kudu 1.7 will compile and run against the Kudu 1.8 client library and vice-versa.
- The Kudu 1.8 C++ client is API- and ABI-forward-compatible with Kudu 1.7. Applications written and compiled against the Kudu 1.7 client library will run without modification against the Kudu 1.8 client library. Applications written and compiled against the Kudu 1.8 client library will run without modification against the Kudu 1.7 client library.

- The Kudu 1.8 Python client is API-compatible with Kudu 1.7. Applications written against Kudu 1.7 will continue to run against the Kudu 1.8 client and vice-versa.

Apache Oozie

There are no incompatible changes in this release.

Apache Parquet

Packages and Group ID Renamed

As a part of the Apache incubation process, all Parquet packages and the project's group ID were renamed as follows:

Parquet Version	1.6.0 and lower (CDH 5.x)	1.7.0 and higher (CDH 6.x)
Java Package Names	parquet.*	org.apache.parquet.*
Group ID	com.twitter	org.apache.parquet

If you directly consume the Parquet API, instead of using Parquet through Hive, Impala or other CDH component, you need to update your code to reflect these changes:

Update *.java files:

Before	After
import parquet.*;	import org.apache.parquet.*;

Update pom.xml:

Before	After
<pre><dependency> <groupId> com.twitter </groupId> <version> \${parquet.version} </version> </dependency></pre>	<pre><dependency> <groupId> org.apache.parquet </groupId> <version> \${parquet.version} </version> </dependency></pre>

API Methods Removed

In Parquet 1.6, a number of API methods were removed from the `parquet.hadoop.ParquetInputSplit` class that depended on reading metadata on the client side. Metadata should be read on the task side instead.

Removed Method	New Method to Use
<code>parquet.hadoop.ParquetInputSplit.getFileSchema</code>	<code>org.apache.parquet.hadoop.api.InitContext.getFileSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getRequestedSchema</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getRequestedSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getReadSupportMetadata</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getReadSupportMetadata</code>
<code>parquet.hadoop.ParquetInputSplit.getBlocks</code>	<code>org.apache.parquet.hadoop.metadata.ParquetMetadata.getBlocks</code>
<code>parquet.hadoop.ParquetInputSplit.getExtraMetadata</code>	-

Apache Pig

The following change is introduced to Pig in CDH 6.0 and it is not backwards compatible. You must modify your Pig scripts as described below.

Removal of the Apache DataFu Pig JAR from CDH 6

Apache DataFu Pig is a collection of user-defined functions that can be used with Pig for data mining and statistical analysis on large-scale data. The DataFu JAR was included in CDH 5, but due to very low adoption rates, the JAR was deprecated in CDH 5.9 and is being removed from CDH 6, starting with CDH 6.0. It is no longer supported.

Recommended Migration Strategy

A simple way to assess what DataFu functions you are using in your Pig scripts is to use the `grep` utility to search for occurrences of "datafu" in your code. When DataFu functions are used in Pig scripts, you must use a function definition entry that contains "datafu" like the following example:

```
define <function_name> datafu.pig... .<class_name>();
```

Use `grep` to search for the string "datafu" in your scripts and that will identify where the DataFu JAR is used.

Cloudera recommends migrating to Hive UDFs or operators wherever it is possible. However, if there are cases where it is impossible to replace DataFu functions with Hive functions, download the upstream version of the DataFu Pig libraries and place them on the node where the Pig front end is used. To preserve compatibility, use the version 1.1.0 JAR, which was the version included in CDH 5. You can download the JAR file [here](#). However, Cloudera does not support using this upstream DataFu JAR file.

Mapping DataFu UDFs to Hive UDFs

The following Hive UDFs map to DataFu UDFs and can be used instead in Pig scripts with the caveats that are listed:

Table 33: Hive Functions That Map to DataFu Functions

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
MD5 (hash)	Computes the MD5 value of a string and outputs a hex value by default.	md5	None
SHA (hash)	Computes the SHA value of a string and outputs a hex value by default.	sha/sha1	None
RandInt (random)	Generates a uniformly distributed integer between two bounds.	rand	None
VAR (stats)	Generates the variance of a set of values.	variance	None
Median (stats)	Computes the median for a sorted input bag. A special case of the Quantile function.	percentile(0.5)	See Limitations When Substituting Quantile and Median DataFu Functions on page 290.
Quantile (stats)	Computes quantiles for a sorted input bag.	percentile	See Limitations When Substituting Quantile and Median DataFu Functions on page 290.
Coalesce (util)	Returns the first non-null value from a tuple, like COALESCE in SQL.	coalesce	None
InUDF (util)	Similar to the SQL IN function, this function provides a convenient way	IN	None

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
	to filter using a logical disjunction over many values.		

For more information about using Hive UDFs, see
https://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_hive_udf.html.

Limitations When Substituting Quantile and Median DataFu Functions

With the exception of `Median` and `Quantile` all Hive functions specified in the above table should work as expected in Pig scripts. `Median` extends `Quantile` in DataFu functions and the equivalent Hive functions have a similar relationship. However, there is an important difference in how you use `percentile` and how you use `Quantile`. The differences are summarized in the following table:

Table 34: Differences Between Usage of DataFu 'Quantile' and Hive 'percentile'

Function:	Quantile	percentile
Input must be sorted:	Yes	No
Nulls are allowed in input:	No	Yes
Examples:	<pre>register datafu-1.2.0.jar; define median datafu.pig.stats.Quantile('0.5'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B = group A by name; C = foreach B { sorted = order A by n; generate group, flatten (median(sorted.n)); }</pre>	<pre>define percentile HiveUDAF ('percentile'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B2 = foreach A generate name, n, 0.5 as perc; C2 = GROUP B2 by name; D2 = FOREACH C2 generate group, percentile (B2. (n, perc));</pre>

Although DataFu `StreamingQuantile` and `StreamingMedian` might appear to match Hive's `percentile_approx` function, Pig cannot consume `percentile_approx`.

DataFu Functions with No Hive Function or Operator Equivalent

The following general limitations apply when mapping DataFu UDFs to Hive UDFs:

- Many DataFu functions operate on a custom Pig data structure called a *bag*. No Hive UDFs can operate on Pig bags, so there are no equivalents for these DataFu functions.
- Some DataFu functions are custom functions that do not have Hive UDF equivalents. For example, the DataFu functions that calculate geographic distances, run the PageRank algorithm, or that do sampling. There are no equivalent Hive UDFs for these DataFu functions either.

Table 35: DataFu Functions with No Hive UDF Equivalent

AppendToBag (bags)	AssertUDF (util)	BagConcat (bags)
BagGroup (bags)	BagLeftOuterJoin (bags)	BagSplit (bags)
BoolToInt (util)	CountEach (bags)	DistinctBy (bags)

<code>EmptyBagToNull (bags)</code>	<code>EmptyBagToNullFields (bags)</code>	<code>Enumerate (bags)</code>
<code>FirstTupleFromBag (bags)</code>	<code>HaversineDistInMiles (geo)</code>	<code>IntToBool (util)</code>
<code>MarkovPairs</code>	<code>NullToEmptyBag (bags)</code>	<code>PageRank (linkanalysis)</code>
<code>PrependToBag (bags)</code>	<code>ReservoirSample (sampling)*</code>	<code>ReverseEnumerate (bags)</code>
<code>SampleByKey (sampling)*</code>	<code>SessionCount (sessions)</code>	<code>Sessionize (sessions)</code>
<code>SetIntersect (sets)</code>	<code>SetUnion (sets)</code>	<code>SimpleRandomSample (sampling)*</code>
<code>TransposeTupleToBag (util)</code>	<code>UnorderedPairs (bags)</code>	<code>UserAgentClassify (urls)</code>
<code>WeightedSample (sampling)*</code>	<code>WilsonBinConf (stats)</code>	—

* These DataFu functions might be replaced with TABLESAMPLE in HiveQL. See the [Apache Hive wiki](#).

Cloudera Search

The following changes are introduced in CDH 6.1

In CDH 6.1 Cloudera Search is rebased on Apache Solr 7.4.

Deprecations

- Enabling/disabling `autoAddReplicas` cluster-wide with the API is deprecated. Use `suspend/resume` trigger APIs with `name= ".auto_add_replicas"` instead.
- In the `ReplicationHandler`, the `master.commitReserveDuration` sub-element is deprecated. Configure a direct `commitReserveDuration` element instead for use in all modes (leader, follower, cloud).

Removals

- The old Leader-In-Recovery implementation (implemented in Solr 4.9) has been removed and replaced. Solr supports rolling upgrades from old 7.x versions of Solr to future 7.x releases until the last release of the 7.x major version. This means that to upgrade to Solr 8, you will have to be on Solr 7.3 or higher.
- The throttling mechanism used to limit the rate of processed autoscaling events has been removed. This deprecates the `actionThrottlePeriodSeconds` setting in the set-properties command of Autoscaling API. Use the `triggerCooldownPeriodSeconds` parameter to pause event processing.
- The `RunExecutableListener` event listener was removed for security reasons. If you want to listen to events caused by updates and commits, or you want to optimize, write your own listener as native Java class as part of a Solr plugin.

For more information see the [Apache Solr 7.4 Release Notes](#).

The following changes are introduced in CDH 6.0

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has many incompatibilities with the 4.10 version of Apache Solr used in recent CDH 5 releases, such as the following:

- Solr 7 uses a managed schema by default. Generating an instance directory no longer generates `schema.xml`. For instructions on switching to a managed schema, see [Switching from schema.xml to Managed Schema](#) in *Apache Solr Reference Guide*.
- Creating a collection using `solrctl collection --create` without specifying the `-c <configName>` parameter now uses a default configuration set (`named _default`) instead of a configuration set with the same name as the collection. To avoid this, always specify the `-c <configName>` parameter when creating new collections.

For the full list of changes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

Apache Sentry contains the following incompatible change in CDH 6.0.0:

- Sentry no longer supports policy file authorization. You must migrate policy files to the database-backed Sentry service before you upgrade to CDH 6.0.0 unless you are using Sentry policy files for Solr. If you are using Sentry policy files for Solr, you must migrate to the database-backed Sentry service after you upgrade.

For information about migrating policy files before you upgrade, see [Migrating from Sentry Policy Files to the Sentry Service](#). For information about migrating policy files for Solr after you upgrade, see [Migrating Sentry Privileges for Solr After Upgrading to CDH 6](#).

Apache Spark

The following sections describe changes in Spark support in CDH 6 that might require special handling during upgrades, or code changes within existing applications.

- All Spark applications built against Spark 1.6 in CDH 5 must be rebuilt against Spark 2.x in CDH 6.
- Spark 2 in CDH 6 works with Java 8, not Java 7. If this change produces any Java code incompatibilities, update your Java code and rebuild the application.
- Spark 2 in CDH 6 works with Scala 2.11, not Scala 2.10. If this change produces any Scala code incompatibilities, update your Scala code and rebuild the application.
- `HiveContext` and `SQLContext` have been removed, although those variables still work for backward compatibility. Use the `SparkSession` object to replace both of these handles.
- `DataFrames` have been removed from the Scala API. `DataFrame` is now a special case of `Dataset`.

Since compile-time type-safety in Python and R is not a language feature, the concept of `Dataset` does not apply to these languages' APIs. Instead, `DataFrame` remains the primary programming abstraction.

- Spark 2.0 and higher do not use an assembly JAR for standalone applications.
- If you have event logs created in CDH 5.3 or lower, you cannot read those logs using Spark in CDH 6.0 or higher.

Apache Sqoop

CDH 6.1.1 introduces no new incompatible changes for Apache Sqoop.

The following changes are introduced in CDH 6.0, and are not backwards compatible:

- All classes in `com.cloudera.sqoop` packages have been removed in CDH 6.0. Use the corresponding classes from `org.apache.sqoop` packages. For example, use `org.apache.sqoop.SqoopOptions` instead of `com.cloudera.sqoop.SqoopOptions`.



Note: This change only affects customers who build their own application on top of Sqoop classes. Sqoop CLI users are not affected.

- Because of changes introduced in the Sqoop metastore logic, the metastore database created by Sqoop CDH 6 cannot be used by earlier versions. The metastore database created by Sqoop CDH 5 can be used by both Sqoop CDH 5 and Sqoop CDH 6.

Require an explicit option to be specified with `--split-by` for a String column

Using the `--split-by` option with a CHAR or VARCHAR column does not always work properly, so Sqoop now requires the user to set the `org.apache.sqoop.splitter.allow_text_splitter` property to `true` to confirm that they are aware of this risk.

Example:

```
sqoop import -Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect $MYCONN
--username $MYUSER --password $MYPWD --table "test_table" --split-by "string_column"
```

For more information, see [SQOOP-2910](#).

Make Sqoop fail if user specifies --direct connector when it is not available

The --direct option is only supported with the following databases: MySQL, PostgreSQL, Oracle, Netezza.

In earlier releases, Sqoop silently ignored this option if it was specified for other databases, but it now throws an error.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table
"direct_import" --direct
```

The command fails with the following error message:

```
Was called with the --direct option, but no direct connector available.
```

For more information, see [SQOOP-2913](#).

Sqoop does not allow --as-parquetfile with hcatalog jobs or when hive import with create-hive-table is used

The --create-hive-table option is not supported when the user imports into Hive in Parquet format. Earlier this option was silently ignored, and the data was imported even if the Hive table existed. Sqoop will now fail if the --create-hive-table option is used with the --as-parquetfile option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table
"test_table" --hive-import --as-parquetfile --create-hive-table
```

The command fails with the following error message:

```
Hive import and create hive table is not compatible with importing into ParquetFile
format.
```

For more information, see [SQOOP-3010](#).

Create fail fast for export with --hcatalog-table <HIVE_VIEW>

Importing into and exporting from a Hive view using HCatalog is not supported by Sqoop. A fail fast check was introduced so that now Sqoop throws a descriptive error message if the user specified a Hive view in the value of the --hcatalog-table option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table"
--hcatalog-table "test_view"
```

The command fails with the following error message:

```
Reads/Writes from and to Views are not supported by HCatalog
```

For more information, see [SQOOP-3027](#).

Simplify Unicode character support in source files

Simplify Unicode character support in source files (introduced by [SQOOP-3074](#)) by defining explicit locales instead of using EscapeUtils. The Java source files generated by Sqoop will be encoded in UTF-8 format.

For more information, see [SQOOP-3075](#).

Columns added to MySQL after initial Sqoop import, export back to table with same schema fails

If we export from HDFS to an RDBMS table and the file on HDFS has no value for some of the columns defined in the table, Sqoop will use the values of `--input-null-string` and `--input-null-non-string` options. Earlier this scenario was not supported and Sqoop failed.

For more information, see [SQOOP-3158](#).

Sqoop fails if the user tries to encode a null value when using `--direct` connector and a MySQL database

The MySQL direct connector does not support the `--null-string`, `--null-non-string`, `--input-null-string`, and `--input-null-non-string` options. These options were silently ignored earlier, but Sqoop now throws an error if these options are used with MySQL direct imports and exports.

For more information, see [SQOOP-3206](#).

Apache Zookeeper

There are no incompatible changes in this release.

Timezone Names Unsupported in Impala in CDH 6.1.1

The following table lists the time zone names / aliases no longer supported in Impala along with the canonical names you can use to replace the unsupported aliases with.

Deprecated	Replace with
ACDT	Australia/South
ACST	Australia/Darwin
ACWST	Australia/Eucla
ADT	America/Thule
AEDT	Australia/Sydney
AEST	Australia/Sydney
AFT	Asia/Kabul
AKDT	America/Anchorage
AKST	America/Anchorage
ALMT	Asia/Almaty
AMST	America/Cuiaba
AMT	Asia/Yerevan
ANAT	Asia/Anadyr
AQTT	Asia/Aqttau
AWST	Australia/West
AZOST	Atlantic/Azores
AZOT	Atlantic/Azores
AZST	Asia/Baku
AZT	Asia/Baku
Acre Time	Brazil/Acre
Afghanistan Time	Asia/Kabul
Alaska Daylight Time	America/Anchorage

Alaska Standard Time	America/Anchorage
Alma-Ata Time	Asia/Almaty
Amazon Summer Time	America/Cuiaba
Amazon Time	Brazil/West
Anadyr Time	Asia/Anadyr
Aqtau Time	Asia/Aqtau
Aqtobe Time	Asia/Aqtobe
Arabia Standard Time	Asia/Aden
Argentine Time	America/Argentina/Buenos_Aires
Armenia Time	Asia/Yerevan
Asia/Riyadh87	-
Asia/Riyadh88	-
Asia/Riyadh89	-
Atlantic Daylight Time	America/Thule
Atlantic Standard Time	America/Puerto_Rico
Australian Central Daylight Time (South Australia)	Australia/South
Australian Central Daylight Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Standard Time (Northern Territory)	Australia/Darwin
Australian Central Standard Time (South Australia)	Australia/South
Australian Central Standard Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Western Standard Time	Australia/Eucla
Australian Eastern Daylight Time (New South Wales)	Australia/Sydney
Australian Eastern Daylight Time (Tasmania)	Australia/Hobart
Australian Eastern Daylight Time (Victoria)	Australia/Victoria
Australian Eastern Standard Time (New South Wales)	Australia/Sydney
Australian Eastern Standard Time (Queensland)	Australia/Brisbane
Australian Eastern Standard Time (Tasmania)	Australia/Hobart
Australian Eastern Standard Time (Victoria)	Australia/Victoria
Australian Western Standard Time	Australia/West
Azerbaijan Summer Time	Asia/Baku
Azerbaijan Time	Asia/Baku
Azores Summer Time	Atlantic/Azores
Azores Time	Atlantic/Azores
BDT	Asia/Dhaka
BNT	Asia/Brunei

BOT	America/La_Paz
BRST	America/Sao_Paulo
BRT	America/Sao_Paulo
BTT	Asia/Thimbu
Bangladesh Time	Asia/Dhaka
Bhutan Time	Asia/Thimbu
Bolivia Time	America/La_Paz
Bougainville Standard Time	Pacific/Bougainville
Brasilia Summer Time	America/Sao_Paulo
Brasilia Time	America/Sao_Paulo
British Summer Time	GB
Brunei Time	Asia/Brunei
CCT	Indian/Cocos
CDT	America/Chicago
CEST	CET
CHADT	NZ-CHAT
CHAST	NZ-CHAT
CHOT	Asia/Choibalsan
CHUT	Pacific/Yap
CKT	Pacific/Rarotonga
CLST	America/Santiago
CLT	America/Santiago
COT	America/Bogota
CVT	Atlantic/Cape_Verde
CXT	Indian/Christmas
Canada/East-Saskatchewan	America/Regina
Cape Verde Time	Atlantic/Cape_Verde
Central African Time	Africa/Harare
Central Daylight Time	America/Chicago
Central European Summer Time	CET
Central European Time	CET
Central Indonesia Time	Asia/Makassar
Central Standard Time	America/Chicago
ChST	Pacific/Guam
Chamorro Standard Time	Pacific/Guam
Chatham Daylight Time	NZ-CHAT

Chatham Standard Time	NZ-CHAT
Chile Summer Time	America/Santiago
Chile Time	America/Santiago
China Standard Time	Asia/Shanghai
Choibalsan Time	Asia/Choibalsan
Christmas Island Time	Indian/Christmas
Chuuk Time	Pacific/Yap
Cocos Islands Time	Indian/Cocos
Colombia Time	America/Bogota
Cook Is. Time	Pacific/Rarotonga
Coordinated Universal Time	UCT
Cuba Daylight Time	Cuba
Cuba Standard Time	Cuba
DAVT	Antarctica/Davis
DDUT	Antarctica/DumontDUrville
Davis Time	Antarctica/Davis
Dumont-d'Urville Time	Antarctica/DumontDUrville
EASST	Pacific/Easter
EAST	Pacific/Easter
EDT	America/Indiana/Indianapolis
EEST	Africa/Cairo
EGST	America/Scoresbysund
EGT	America/Scoresbysund
East Indonesia Time	Asia/Jayapura
Easter Is. Summer Time	Pacific/Easter
Easter Is. Time	Pacific/Easter
Eastern African Time	Africa/Addis_Ababa
Eastern Daylight Time	America/Indiana/Indianapolis
Eastern European Summer Time	Africa/Cairo
Eastern European Time	Africa/Cairo
Eastern Greenland Summer Time	America/Scoresbysund
Eastern Greenland Time	America/Scoresbysund
Eastern Standard Time	EST
Ecuador Time	America/Guayaquil
FJST	Pacific/Fiji
FJT	Pacific/Fiji

FKT	Atlantic/Stanley
FNT	America/Noronha
Falkland Is. Time	Atlantic/Stanley
Fernando de Noronha Time	America/Noronha
Fiji Summer Time	Pacific/Fiji
Fiji Time	Pacific/Fiji
French Guiana Time	America/Cayenne
French Southern & Antarctic Lands Time	Indian/Kerguelen
GALT	Pacific/Galapagos
GAMT	Pacific/Gambier
GET	Asia/Tbilisi
GFT	America/Cayenne
GILT	Pacific/Tarawa
GMT+00:00	GMT0
GMT+01:00	Etc/GMT-1
GMT+02:00	Etc/GMT-2
GMT+03:00	Etc/GMT-3
GMT+04:00	Etc/GMT-4
GMT+05:00	Etc/GMT-5
GMT+06:00	Etc/GMT-6
GMT+07:00	Etc/GMT-7
GMT+08:00	Etc/GMT-8
GMT+09:00	Etc/GMT-9
GMT+10:00	Etc/GMT-10
GMT+11:00	Etc/GMT-11
GMT+12:00	Etc/GMT-12
GMT+13:00	Etc/GMT-13
GMT+14:00	Etc/GMT-14
GMT-01:00	Etc/GMT+1
GMT-02:00	Etc/GMT+2
GMT-03:00	Etc/GMT+3
GMT-04:00	Etc/GMT+4
GMT-05:00	Etc/GMT+5
GMT-06:00	Etc/GMT+6
GMT-07:00	Etc/GMT+7
GMT-08:00	Etc/GMT+8

GMT-09:00	Etc/GMT+9
GMT-10:00	Etc/GMT+10
GMT-11:00	Etc/GMT+11
GMT-12:00	Etc/GMT+12
GST	Asia/Dubai
GYT	America/Guyana
Galapagos Time	Pacific/Galapagos
Gambier Time	Pacific/Gambier
Georgia Time	Asia/Tbilisi
Ghana Mean Time	Africa/Accra
Gilbert Is. Time	Pacific/Tarawa
Greenwich Mean Tim	GB
Gulf Standard Time	Asia/Dubai
Guyana Time	America/Guyana
HADT	US/Aleutian
HAST	US/Aleutian
HKT	Hongkong
HOVT	Asia/Hovd
Hawaii Standard Time	HST
Hawaii-Aleutian Daylight Time	US/Aleutian
Hawaii-Aleutian Standard Time	US/Aleutian
Hong Kong Time	Hongkong
Hovd Time	Asia/Hovd
ICT	Asia/Ho_Chi_Min
IDT	Israel
IOT	Indian/Chagos
IRDT	Iran
IRKT	Asia/Chita
IRST	Iran
India Standard Time	Asia/Kolkata
Indian Ocean Territory Time	Indian/Chagos
Indochina Time	Asia/Ho_Chi_Min
Iran Daylight Time	Iran
Iran Standard Time	Iran
Irish Summer Time	Eire
Irkutsk Time	Asia/Chita

Israel Daylight Time	Israel
Israel Standard Time	Israel
Japan Standard Time	Asia/Tokyo
KGT	Asia/Bishkek
KOST	Pacific/Kosrae
KRAT	Asia/Krasnoyarsk
KST	ROK
Khandyga Time	Asia/Khandyga
Kirgizstan Time	Asia/Bishkek
Korea Standard Time	ROK
Kosrae Time	Pacific/Kosrae
Krasnoyarsk Time	Asia/Krasnoyarsk
LHDT	Australia/LHI
LHST	Australia/LHI
LINT	Pacific/Kiritimati
Line Is. Time	Pacific/Kiritimati
Lord Howe Daylight Time	Australia/LHI
Lord Howe Standard Time	Australia/LHI
MAGT	Asia/Magadan
MART	Pacific/Marquesas
MAWT	Antarctica/Mawson
MDT	Navajo
MEST	MET
MHT	Kwajalein
MIST	Antarctica/Macquarie
MMT	Asia/Rangoon
MSK	W-SU
MUT	Indian/Mauritius
MVT	Indian/Maldives
MYT	Asia/Kuching
Macquarie Island Standard Time	Antarctica/Macquarie
Magadan Time	Asia/Magadan
Malaysia Time	Asia/Kuching
Maldives Time	Indian/Maldives
Marquesas Time	Pacific/Marquesas
Marshall Islands Time	Kwajalein

Mauritius Time	Indian/Mauritius
Mawson Time	Antarctica/Mawson
Middle Europe Summer Time	MET
Middle Europe Time	MET
Mideast/Riyadh87	-
Mideast/Riyadh88	-
Mideast/Riyadh89	-
Moscow Standard Time	W-SU
Mountain Daylight Time	Navajo
Mountain Standard Time	MST
Myanmar Time	Asia/Rangoon
NCT	Pacific/Noumea
NDT	America/St_Johns
NFT	Pacific/Norfolk
NOVT	Asia/Novosibirsk
NPT	Asia/Katmandu
NRT	Pacific/Nauru
NUT	Pacific/Niue
NZDT	NZ
NZST	NZ
Nauru Time	Pacific/Nauru
Nepal Time	Asia/Katmandu
New Caledonia Time	Pacific/Noumea
New Zealand Daylight Time	NZ
New Zealand Standard Time	NZ
Newfoundland Daylight Time	America/St_Johns
Newfoundland Standard Time	America/St_Johns
Niue Time	Pacific/Niue
Norfolk Time	Pacific/Norfolk
Novosibirsk Time	Asia/Novosibirsk
OMST	Asia/Omsk
ORAT	Asia/Oral
Omsk Time	Asia/Omsk
Oral Time	Asia/Oral
PDT	America/Los_Angeles
PET	America/Lima

PETT	Asia/Kamchatka
PGT	Pacific/Port_Moresby
PHOT	Pacific/Enderbury
PHT	Asia/Manila
PKT	Asia/Karachi
PMDT	America/Miquelon
PMST	America/Miquelon
PONT	Pacific/Ponape
PWT	Pacific/Palau
PYST	America/Asuncion
PYT	America/Asuncion
Pacific Daylight Time	America/Los_Angeles
Pacific Standard Time	America/Los_Angeles
Pakistan Time	Asia/Karachi
Palau Time	Pacific/Palau
Papua New Guinea Time	Pacific/Port_Moresby
Paraguay Summer Time	America/Asuncion
Paraguay Time	America/Asuncion
Peru Time	America/Lima
Petropavlovsk-Kamchatski Time	Asia/Kamchatka
Philippines Time	Asia/Manila
Phoenix Is. Time	Pacific/Enderbury
Pierre & Miquelon Daylight Time	America/Miquelon
Pierre & Miquelon Standard Time	America/Miquelon
Pitcairn Standard Time	Pacific/Pitcairn
Pohnpei Time	Pacific/Ponape
QYZT	Asia/Qyzylorda
Qyzylorda Time	Asia/Qyzylorda
RET	Indian/Reunion
ROTT	Antarctica/Rothera
Reunion Time	Indian/Reunion
Rothera Time	Antarctica/Rothera
SAKT	Asia/Sakhalin
SAMT	Europe/Samara
SAST	Africa/Maseru
SBT	Pacific/Guadalcanal

SCT	Indian/Mahe
SGT	Singapore
SRET	Asia/Srednekolymsk
SRT	America/Paramaribo
SYOT	Antarctica/Syowa
Sakhalin Time	Asia/Sakhalin
Samara Time	Europe/Samara
Samoa Standard Time	US/Samoa
Seychelles Time	Indian/Mahe
Singapore Time	Singapore
Solomon Is. Time	Pacific/Guadalupe
South Africa Standard Time	Africa/Maseru
South Georgia Standard Time	Atlantic/South_Georgia
Srednekolymsk Time	Asia/Srednekolymsk
Suriname Time	America/Paramaribo
Syowa Time	Antarctica/Syowa
SystemV/AST4	America/Puerto_Rico
SystemV/AST4ADT	Canada/Atlantic
SystemV/CST6	Canada/Saskatchewan
SystemV/CST6CDT	US/Central
SystemV/EST5	America/Cayman
SystemV/EST5EDT	America/New_York
SystemV/HST10	US/Hawaii
SystemV/MST7	US/Arizona
SystemV/MST7MDT	America/Denver
SystemV/PST8	Pacific/Pitcairn
SystemV/PST8PDT	US/Pacific
SystemV/YST9	Pacific/Gambier
SystemV/YST9YDT	US/Alaska
TAHT	Pacific/Tahiti
TFT	Indian/Kerguelen
TJT	Asia/Dushanbe
TKT	Pacific/Fakaofo
TLT	Asia/Dili
TMT	Asia/Ashgabat
TOT	Pacific/Tongatapu

TVT	Pacific/Funafuti
Tahiti Time	Pacific/Tahiti
Tajikistan Time	Asia/Dushanbe
Timor-Leste Time	Asia/Dili
Tokelau Time	Pacific/Fakaofo
Tonga Time	Pacific/Tongatapu
Turkmenistan Time	Asia/Ashgabat
Tuvalu Time	Pacific/Funafuti
ULAT	Asia/Ulan_Bator
UYST	America/Montevideo
UYT	America/Montevideo
UZT	Asia/Tashkent
Ulaanbaatar Time	Asia/Ulan_Bator
Uruguay Summer Time	America/Montevideo
Uruguay Time	America/Montevideo
Ust-Nera Time	Asia/Ust-Nera
Uzbekistan Time	Asia/Tashkent
VET	America/Caracas
VLAT	Asia/Ust-Nera
VOST	Antarctica/Vostok
VUT	Pacific/Efate
Vanuatu Time	Pacific/Efate
Venezuela Time	America/Caracas
Vladivostok Time	Asia/Vladivostok
Vostok Time	Antarctica/Vostok
WAKT	Pacific/Wake
WAST	Africa/Windhoek
WAT	Africa/Lagos
WEST	WET
WFT	Pacific/Wallis
WGST	America/Godthab
WGT	America/Godthab
WIB	Asia/Jakarta
WIT	Asia/Jayapura
WITA	Asia/Makassar
WSDT	Pacific/Apia

WSST	Pacific/Apia
Wake Time	Pacific/Wake
Wallis & Futuna Time	Pacific/Wallis
West Indonesia Time	Asia/Jakarta
West Samoa Daylight Time	Pacific/Apia
West Samoa Standard Time	Pacific/Apia
Western African Summer Time	Africa/Windhoek
Western African Time	Africa/Lagos
Western European Summer Time	WET
Western European Time	WET
Western Greenland Summer Time	America/Godthab
Western Greenland Time	America/Godthab
XJT	Asia/Urumqi
Xinjiang Standard Time	Asia/Urumqi
YAKT	Asia/Yakutsk
YEKT	Asia/Yekaterinburg
Yakutsk Time	Asia/Yakutsk
Yekaterinburg Time	Asia/Yekaterinburg

Known Issues and Limitations in CDH 6.1.1

The following sections describe the known issues in CDH 6.1.1, grouped by component:

Operating System Known Issues

Known issues and workarounds related to operating systems are listed below.

Linux kernel security patch and CDH services crashes CVE-2017-10000364

After applying a recent Linux kernel security patch for [CVE-2017-1000364](#), CDH services that use the JSVC set of libraries crash with a Java Virtual Machine (JVM) error such as:

```
A fatal error has been detected by the Java Runtime Environment:
SIGBUS (0x7) at pc=0x00007fe91ef6cebc, pid=30321, tid=0x00007fe930c67700
```

Cloudera services for HDFS and Impala cannot start after applying the patch.



Important: If you have not upgraded your Linux kernel using the distribution's patch for CVE-2017-1000364, do **not** apply the patch.

Commonly used Linux distributions are shown in the table below. However, the issue affects any CDH release that runs on RHEL, CentOS, Oracle Linux, SUSE Linux, or Ubuntu and that has had the Linux kernel security patch for CVE-2017-1000364 applied.

If you have already applied the patch for your OS according to the advisories for CVE-2017-1000364, apply the kernel update that contains the fix for your operating system (some of which are listed in the table). If you cannot apply the kernel update, you can workaround the issue by increasing the Java thread stack size as detailed in the steps below.

Distribution	Advisories for CVE-2017-1000364	Advisory updates
Oracle Linux 6	ELSA-2017-1486	Oracle has fixed this problem in ELSA-2017-1723 .
Oracle Linux 7	ELSA-2017-1484	Oracle has also added the fix for Oracle Linux 7 in ELBA-2017-1674 .
RHEL 6	RHSA-2017-1486	RedHat has fixed this problem for RHEL 6, marked this as outdated and superseded by RHSA-2017-1723 .
RHEL 7	RHSA-2017-1484	RedHat has fixed this problem for RHEL 7 and has marked this patch as outdated and superseded by RHBA-2017-1674 .
SLES	CVE-2017-1000364	SUSE has also fixed this problem and the patch names are included in this same advisory.

Workaround

If you cannot apply the kernel update, you can set the Java thread stack size to `-Xss1280k` for the affected services using the appropriate Java configuration option or the environment advanced configuration snippet, as detailed below.

For role instances that have specific Java configuration options properties:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Type `java` in the search field to display Java related configuration parameters. The **Java Configuration Options for Catalog Server** property field displays. Type `-Xss1280k` in the entry field, adding to any existing settings.
4. Click **Save Changes**.
5. Navigate to the HDFS service by selecting **Clusters > HDFS**.
6. Click the **Configuration** tab.
7. Click the Scope filter **DataNode**. The **Java Configuration Options for DataNode** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
8. Click **Save Changes**.
9. Select the Scope filter **NFS Gateway**. The **Java Configuration Options for NFS Gateway** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
10. Click **Save Changes**.
11. Restart the affected roles (or configure the safety valves in next section and restart when finished with all configurations).

For role instances that do not have specific Java configuration options:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Click the Scope filter **Impala Daemon** and Category filter **Advanced**.
4. Type `impala daemon` environment in the search field to find the safety valve entry field.
5. In the **Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

6. Click **Save Changes**.
7. Click the Scope filter **Impala StateStore** and Category filter **Advanced**.
8. In the **Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

9. Click **Save Changes**.
10. Restart the affected roles.

The table below summarizes the parameters that can be set for the affected services:

Service	Settable Java Configuration Option
HDFS DataNode	Java Configuration Options for DataNode
HDFS NFS Gateway	Java Configuration Options for NFS Gateway
Impala Catalog Server	Java Configuration Options for Catalog Server
Impala Daemon	Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K
Impala StateStore	Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K

Cloudera Issue: CDH-55771

Leap-Second Events



Note: The [next leap-second event](#) is unknown at this time. The last leap-second event occurred on December 16, 2016 at 23:59:60 UTC.

Impact: After a leap-second event, Java applications (including CDH services) using older Java and Linux kernel versions, may consume almost 100% CPU. See <https://access.redhat.com/articles/15145>.

Leap-second events are tied to the time synchronization methods of the Linux kernel, the Linux distribution and version, and the Java version used by applications running on affected kernels.

Although Java is increasingly agnostic to system clock progression (and less susceptible to a kernel's mishandling of a leap-second event), using JDK 7 or 8 should prevent issues at the CDH level (for CDH components that use the Java Virtual Machine).

Immediate action required:

(1) Ensure that the kernel is up to date.

- **RHEL6/7, CentOS 6/7** - 2.6.32-298 or higher
- **Oracle Enterprise Linux (OEL)** - Kernels built in 2013 or later
- **SLES12** - No action required.

(2) Ensure that your Java JDKs are current (especially if the kernel is not up to date and cannot be upgraded).

- **Java 8** - No action required.

(3) Ensure that your systems use either NTP or PTP synchronization.

For systems not using time synchronization, update both the OS tzdata and Java tzdata packages to the tzdata-2016g version, at a minimum. For OS tzdata package updates, contact OS support or check updated OS repositories. For Java tzdata package updates, see Oracle's [Timezone Updater Tool](#).

Cloudera Issue: CDH-44788, TSB-189

Apache Accumulo Known Issues

There are no notable known issues in this release of Apache Accumulo.

Apache Crunch Known Issues



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

Apache Flume Known Issues

Fast Replay does not work with encrypted File Channel

If an encrypted file channel is set to use fast replay, the replay will fail and the channel will fail to start.

Workaround: Disable fast replay for the encrypted channel by setting `use-fast-replay` to false.

Apache Issue: [FLUME-1885](#)

Apache Hadoop Known Issues

This page includes known issues and related topics, including:

Deprecated Properties

Several Hadoop and HDFS properties have been deprecated as of Hadoop 3.0 and later. For details, see [Deprecated Properties](#).

Hadoop Common

Hadoop LdapGroupsMapping does not support LDAPS for self-signed LDAP server

Hadoop LdapGroupsMapping does not work with LDAP over SSL (LDAPS) if the LDAP server certificate is self-signed. This use case is currently not supported even if Hadoop User Group Mapping LDAP TLS/SSL Enabled, Hadoop User Group Mapping LDAP TLS/SSL Truststore, and Hadoop User Group Mapping LDAP TLS/SSL Truststore Password are filled properly.

Affected Versions: All CDH versions

Apache Issue: [HADOOP-12862](#)

Cloudera Issue: CDH-37926

HDFS

Clusters running CDH 5.16.1, 6.1.0, or 6.1.1 can lose some HDFS file permissions any time the Name Node is restarted

When a cluster is upgraded to 5.16.1, 6.1.0, or 6.1.1 roles with SELECT and/or INSERT privileges on an Impala database or table will have the REFRESH privilege added as part of the upgrade process. HDFS ACLs for roles with the REFRESH privilege get set with empty permissions whenever the Name Node is restarted. This can cause any jobs or queries run by users within affected roles to fail because they will no longer be able to access affected Impala database or tables.

Products Affected: HDFS and components that access files in HDFS

Affected Versions: CDH 5.16.1, 6.1.0, 6.1.1

Users Affected: Clusters with Impala and HDFS ACLs managed by Sentry upgrading from any release to CDH 5.16.1, 6.1.0, and 6.1.1.

Severity (Low/Medium/High): High

Root Cause and Impact: The new privilege REFRESH was introduced in CDH 5.16 and 6.1 and applies to Impala databases and tables. When a cluster is upgraded to 5.16.1, 6.1.0, or 6.1.1, roles with SELECT or INSERT privileges on an Impala database or table will have the REFRESH privilege added during the upgrade.

HDFS ACLs for roles with the REFRESH privilege get set with empty permissions whenever the Name Node is restarted. The Name Node is restarted during the upgrade.

For example if a group `appdev` is in role `appdev_role` and has SELECT access to the Impala table "project" the HDFS ACLs prior to the upgrade would look similar to:

```
group: appdev
      group: r--
```

After the upgrade the HDFS ACLs will be set with no permissions and will look like this:

```
group: appdev
group: :---
```

Any jobs or queries run by users within affected roles will fail because they will no longer be able to access affected Impala database or tables. This impacts any SQL client accessing the affected databases and tables. For example, if a Hive client is used to access a table created in Impala it will also fail. Jobs accessing the files directly through HDFS, e.g. via Spark, will also be impacted.

The HDFS ACLs will get reset whenever the Name Node is restarted.

Immediate action required: If possible, do not upgrade to releases CDH 5.16.1, 6.1.0, or 6.1.1 if Impala is used and Sentry manages HDFS ACLs within your environment. Subsequent CDH releases will resolve the problem with a product fix under [SENTRY-2490](#).

If an upgrade is being considered, reach out to your account team to discuss other possibilities, and to receive additional insight into future product release schedules.

If an upgrade must be executed, contact Cloudera Support indicating the upgrade plan and why an upgrade is being executed. Options are available to assist with the upgrade if necessary.

Addressed in release/refresh/patch: Patches for 5.16.1, 6.1.0 and 6.1.1 are available for major supported operating systems. Customers are encouraged to contact Cloudera Support for a patch. The patch should be applied immediately after upgrade to any of the affected versions.

The fix for this TSB will be included in 6.1.2, 6.2.0, 5.16.2, and 5.17.0.

OIV ReverseXML processor fails

The HDFS OIV ReverseXML processor fails if the XML file contains escaped characters.

Affected Versions: CDH 6.x

Apache Issue: [HDFS-12828](#)

Cannot move encrypted files to trash

With HDFS encryption enabled, you cannot move encrypted files or directories to the trash directory.

Workaround: To remove encrypted files/directories, use the following command with the `-skipTrash` flag specified to bypass trash.

```
rm -r -skipTrash /testdir
```

Affected Versions: All CDH versions

Apache Issue: [HADOOP-10902](#)

HDFS NFS gateway and CDH installation (using packages) limitation

HDFS NFS gateway works as shipped ("out of the box") only on RHEL-compatible systems, but not on SLES or Ubuntu. Because of a bug in native versions of `portmap/rpcbind`, the HDFS NFS gateway does not work out of the box on SLES or Ubuntu systems when CDH has been installed from the command-line, using packages. It does work on supported versions of RHEL-compatible systems on which `rpcbind-0.2.0-10.el6` or later is installed, and it does work if you use Cloudera Manager to install CDH, or if you start the gateway as root.

Workarounds and caveats:

- On Red Hat and similar systems, make sure `rpcbind-0.2.0-10.el6` or later is installed.
- On SLES and Ubuntu systems, do one of the following:
 - Install CDH using Cloudera Manager; or
 - Start the NFS gateway as root; or
 - [Start the NFS gateway without using packages](#); or

- You can use the gateway by running `rpcbind` in insecure mode, using the `-i` option, but keep in mind that this allows anyone from a remote host to bind to the portmap.

Upstream Issue: [731542](#) (Red Hat), [823364](#) (SLES)

No error when changing permission to 777 on `.snapshot` directory

Snapshots are read-only; running `chmod 777` on the `.snapshot`s directory does not change this, but does not produce an error (though other illegal operations do).

Affected Versions: All CDH versions

Apache Issue: [HDFS-4981](#)

Snapshot operations are not supported by ViewFileSystem

Affected Versions: All CDH versions

Snapshots do not retain directories' quotas settings

Affected Versions: All CDH versions

Apache Issue: [HDFS-4897](#)

Permissions for `dfs.namenode.name.dir` incorrectly set

Hadoop daemons should set permissions for the `dfs.namenode.name.dir` (or `dfs.name.dir`) directories to `drwx-----` (700), but in fact these permissions are set to the file-system default, usually `drwxr-xr-x` (755).

Workaround: Use `chmod` to set permissions to 700.

Affected Versions: All CDH versions

Apache Issue: [HDFS-2470](#)

`hadoop fsck -move` does not work in a cluster with host-based Kerberos

Workaround: Use `hadoop fsck -delete`

Affected Versions: All CDH versions

Apache Issue: None

Block report can exceed maximum RPC buffer size on some DataNodes

On a DataNode with a large number of blocks, the block report may exceed the maximum RPC buffer size.

Workaround: Increase the value `ipc.maximum.data.length` in `hdfs-site.xml`:

```
<property>
  <name>ipc.maximum.data.length</name>
  <value>268435456</value>
</property>
```

Affected Versions: All CDH versions

Apache Issue: None

MapReduce2 and YARN

YARN's Continuous Scheduling can cause slowness in Oozie

When Continuous Scheduling is enabled in Yarn, this can cause slowness in Oozie due to long delays in communicating with Yarn. In Cloudera Manager 5.9.0 and higher, Enable Fair Scheduler Continuous Scheduler is turned off by default.

Workaround: Turn off Enable Fair Scheduler Continuous Scheduling in Cloudera Manager YARN Configuration. To keep equivalent benefits of this feature, turn on Fair Scheduler Assign Multiple Tasks.

Affected Versions: All CDH versions

Cloudera Issue: CDH-60788

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

Affected Versions: All CDH versions

Apache Issue: None

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Routable IP address required by ResourceManager

ResourceManager requires routable `host:port` addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard `0.0.0.0` address.

Workaround: Set the address, in the form `host:port`, either in the client-side configuration, or on the command line when you submit the job.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-6808

Amazon S3 copy may time out

The Amazon S3 filesystem does not support renaming files, and performs a copy operation instead. If the file to be moved is very large, the operation can time out because S3 does not report progress during the operation.

Workaround: Use `-Dmapred.task.timeout=15000000` to increase the MR task timeout.

Affected Versions: All CDH versions

Apache Issue: [MAPREDUCE-972](#)

Cloudera Issue: CDH-17955

Apache HBase Known Issues

IOException from Timeouts

CDH 5.12.0 includes the fix [HBASE-16604](#), where the internal scanner that retries in case of IOException from timeouts could potentially miss data. Java clients were properly updated to account for the new behavior, but thrift clients will now see exceptions where the previous missing data would be.

Workaround: Create a new scanner and retry the operation when encountering this issue.

IntegrationTestReplication fails if replication does not finish before the verify phase begins

During `IntegrationTestReplication`, if the `verify` phase starts before the `replication` phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the `-t` flag to set the timeout value before starting verification.

Cloudera Issue: None.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

`ExportSnapshot` or `DistCp` operations may fail on the Amazon `s3a://` protocol

`ExportSnapshot` or `DistCP` operations may fail on AWS when using certain JDK 8 versions, due to an incompatibility between the AWS Java SDK 1.9.x and the `joda-time` date-parsing module.

Workaround: Use `joda-time 2.8.1` or higher, which is included in AWS Java SDK 1.10.1 or higher.

Cloudera Issue: None.

An operating-system level tuning issue in RHEL7 causes 30% latency regressions

Workaround: Avoid using RHEL 7 if you have a latency-critical workload. For a cached workload, consider tuning the C-state (power-saving) behavior of your CPUs.

Cloudera Issue: CDH-32642

Severity: Medium

A RegionServer under extreme duress due to back-to-back garbage collection combined with heavy load on HDFS can lock up while attempting to append to the WAL

The RegionServer appears operational to ZooKeeper, and continues to host regions, but cannot complete any writes. The most obvious symptom is that all writes to regions on this RegionServer time out, and the RegionServer log shows no progress other than queuing of flushes that never complete. Log messages such as the following may occur:

```
124028 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.Sleeper: We slept  
42911ms instead of 3000ms,  
this is likely due to a long garbage collecting pause and it's usually bad, see  
http://hbase.#  
124029 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.JvmPauseMonitor: Detected  
pause in JVM or  
host machine (eg GC): pause of approximately 41110ms
```

```
1806 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2734 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#  
1807 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2963 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
```

Apache Issue: [HBASE-14374](#)

Workaround: Restart the RegionServer. To avoid the problem, adjust garbage-collection settings, give the RegionServer more RAM, and reduce the load on HDFS.

Export to Azure Blob Storage (the `wasb://` or `wasbs://` protocol) is not supported

CDH 5.3 and higher supports Azure Blob Storage for some applications. However, a null pointer exception occurs when you specify a `wasb://` or `wasbs://` location in the `--copy-to` option of the `ExportSnapshot` command or as the output directory (the second positional argument) of the `Export` command.

Workaround: None.

Apache Issue: [HADOOP-12717](#)

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a `Delete Table` fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.

- If `hbaseAdmin.modifyTable()` is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The `postOperation` is implemented only for `postDeleteColumn()`.
- If `Create Table` fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: None

Apache Issue: [HBASE-6992](#)

Apache Hive/HCatalog/Hive on Spark Known Issues

This topic contains:

- [Hive Known Issues](#)
- [HCatalog Known Issues](#)
- [Hive on Spark Known Issues](#)

Hive Known Issues

Hive Jobs Are Submitted to a Single Queue When Sentry is Deployed

Hive jobs are not submitted into the correct YARN queue when Hive is using Sentry because Hive does not use the YARN API to resolve the user or group of the job's original submitter. This causes the job to be placed in a queue using the placement rules based on the Hive user. The HiveServer2 fair scheduler queue mapping used for "non-impersonation" mode does not handle the primary-secondary queue mappings correctly.

Affected Version: CDH 6.0.0 and higher

Workaround: If you are a Hive and Sentry user, do not upgrade to CDH 6.0.0 or 6.0.1. This issue will be fixed as soon as possible. If you must use Hive and Sentry in CDH 6.0.0 or 6.0.1, see [YARN Dynamic Resource Pools Do Not Work with Hive When Sentry Is Enabled](#) for additional workarounds.

When vectorization is enabled on any file type (ORC, Parquet) queries that divide by zero using the modulo operator (%) return an error

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query divides by zero using the modulo operator (%), it returns the following error: `Arithmetic exception [divide by] 0.` For example, if you run the following query this issue is triggered: `SELECT 100 % column_c1 FROM table_t1;` and the value in `column_c1` is zero. The divide operator (/) is not affected by this issue.

Workaround: Disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-19564](#)

Cloudera Issue: CDH-71211

When vectorization is enabled for Hive on any file type (ORC, Parquet) queries that perform comparisons in the `SELECT` clause on large values in columns with the data type of `BIGINT` might return wrong results

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query performs a comparison operation between very large values in columns that are `BIGINT` data types in the `SELECT` clause of the query, incorrect results might be returned. Comparison operators include `==`, `!=`, `<`, `<=`, `>`, and `>=`. This issue does not occur when the comparison operation is performed in the filtering clause of the query. This issue can also occur when the difference of values in such columns is out of range for a `LONG` (64-bit) data type. For example, if `column_c1` stores `8976171455044006767` and `column_c2` stores `-7272907770454997143`, a query such as `SELECT column_c1 < column_c2 FROM table_test` returns `true` instead of `false` because the difference (`8976171455044006767 - (-7272907770454997143)`) is `1.6249079225499E19` which is greater than `9.22337203685478E18`, which is the maximum possible value that a `LONG` (64-bit) data type can hold.

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE_20207](#)

Cloudera Issue: CDH-70996

Workaround: Use a DECIMAL type instead of BIGINT for columns that might contain very large values. Another option is to disable vectorization for the query that is triggering this at either the session level by using the SET statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Specified column position in the ORDER BY clause is not supported for SELECT * queries

When column positions are specified in ORDER BY clauses, they are not honored for SELECT * queries and an error is returned as shown in the following example:

```
CREATE TABLE decimal_1 (id decimal(5,0));
SELECT * FROM decimal_1 ORDER BY 1 limit 100;
Error while compiling statement: FAILED: SemanticException [Error 10219]: Position in
ORDER BY is not supported when using SELECT *
```

Instead the query must list out the columns it is selecting.

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-68550

DirectSQL with PostgreSQL

Hive doesn't support Hive direct SQL queries with PostgreSQL database. It only supports this feature with MySQL, MariaDB, and Oracle. With PostgresSQL, direct SQL is disabled as a precaution, since there have been issues reported upstream where it is not possible to fallback on DataNucleus in the event of some failures, plus other non-standard behaviors. For more information, see [Hive Configuration Properties](#).

Affected Versions: All CDH versions

Cloudera Issue: CDH-49017

ALTER PARTITION ... SET LOCATION does not work on Amazon S3 or between S3 and HDFS

Cloudera recommends that you do not use ALTER PARTITION ... SET LOCATION on S3 or between S3 and HDFS. The rest of the ALTER PARTITION commands work as expected.

Affected Versions: All CDH versions

Cloudera Issue: CDH-42420

Commands run against an Oracle-backed metastore might fail

Commands run against an Oracle-backed Metastore fail with error:

```
javax.jdo.JDODataStoreException Incompatible data type for column TBLS.VIEW_EXPANDED_TEXT
: was CLOB (datastore),
but type expected was LONGVARCHAR (metadata). Please check that the type in the datastore
and the type specified in the MetaData are consistent.
```

This error might occur if the metastore is run on top of an Oracle database with the configuration property `datanucleus.validateColumns` set to true.

Workaround: Set `datanucleus.validateColumns=false` in the `hive-site.xml` configuration file.

Affected Versions: All CDH versions

Cannot create archive partitions with external HAR (Hadoop Archive) tables

ALTER TABLE ... ARCHIVE PARTITION is not supported on external tables.

Affected Versions: All CDH versions

Cloudera Issue: CDH-9638

Object types Server and URI are not supported in "SHOW GRANT ROLE *roleName* on OBJECT *objectName*" statements

Workaround: Use SHOW GRANT ROLE *roleName* to list all privileges granted to the role.

Affected Versions: All CDH versions

Cloudera Issue: CDH-19430

HCatalog Known Issues



Note: As of CDH 5, HCatalog is part of Apache Hive.

There are no notable known issues in this release of HCatalog.

Hive on Spark (HoS) Known Issues

Hive on Spark queries fail with "Timed out waiting for client to connect" for an unknown reason

If this exception is preceded by logs of the form "client.RpcRetryingCaller: Call exception...", then this failure is due to an unavailable HBase service. On a secure cluster, spark-submit will try to obtain delegation tokens from HBase, even though Hive on Spark might not need them. So if HBase is unavailable, spark-submit throws an exception.

Workaround: Fix the HBase service, or set spark.yarn.security.tokens.hbase.enabled to false.

Affected Versions: CDH 5.7.0 and higher

Cloudera Issues: CDH-59591, CDH-59599

Hue Known Issues

Hue does not support the Spark App

Hue does not currently support the Spark application.

Connecting to PostgreSQL Database Fails with Error "No module named psycopg2"

When configuring Hue to use a PostgreSQL database, the connection fails with the following error:

```
Error loading psycopg2 module: No module named psycopg2
```

Workaround: Install the psycopg2 Python package as documented in [Installing the psycopg2 Python Package](#).

Affected Versions: All CDH 6 versions

Fixed Versions: None

Apache Issue: N/A

Cloudera Issue: CDH-65804

Apache Impala Known Issues

The following sections describe known issues and workarounds in Impala, as of the current production release. This page summarizes the most serious or frequently encountered issues in the current release, to help you make planning decisions about installing and upgrading. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.



Note: The online issue tracking system for Impala contains comprehensive information and is updated in real time. To verify whether an issue you are experiencing has already been reported, or which release an issue is fixed in, search on the [Impala JIRA tracker](#).

Impala Known Issues: Startup

These issues can prevent one or more Impala-related daemons from starting properly.

Impala requires FQDN from hostname command on kerberized clusters

The method Impala uses to retrieve the host name while constructing the Kerberos principal is the `gethostname()` system call. This function might not always return the fully qualified domain name, depending on the network configuration. If the daemons cannot determine the FQDN, Impala does not start on a kerberized cluster.

Workaround: Test if a host is affected by checking whether the output of the `hostname` command includes the FQDN. On hosts where `hostname` only returns the short name, pass the command-line flag `--hostname=fully_qualified_domain_name` in the startup options of all Impala-related daemons.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4978](#)

Impala Known Issues: Crashes and Hangs

These issues can cause Impala to quit or become unresponsive.

Unable to view large catalog objects in catalogd Web UI

In `catalogd` Web UI, you can list metadata objects and view their details. These details are accessed via a link and printed to a string formatted using thrift's `DebugProtocol`. Printing large objects (> 1 GB) in Web UI can crash `catalogd`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6841](#)

Impala Known Issues: Performance

These issues involve the performance of operations such as queries or DDL statements.

Metadata operations block read-only operations on unrelated tables

Metadata operations that change the state of a table, like `COMPUTE STATS` or `ALTER RECOVER PARTITIONS`, may delay metadata propagation of unrelated unloaded tables triggered by statements like `DESCRIBE` or `SELECT` queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6671](#)

Impala Known Issues: Security

These issues relate to security features, such as Kerberos authentication, Sentry authorization, encryption, auditing, and redaction.

Impala does not support Heimdal Kerberos

Heimdal Kerberos is not supported in Impala.

Apache Issue: [IMPALA-7072](#)

Affected Versions: All CDH 6 versions

System-wide auth-to-local mapping not applied correctly to Kudu service account

Due to system `auth_to_local` mapping, the principal may be mapped to some local name.

When running with Kerberos enabled, you may hit the following error message where `<random-string>` is some random string which doesn't match the primary in the Kerberos principal.

```
WARNINGS: TransmitData() to X.X.X.X:27000 failed: Remote error: Not authorized: {username='<random-string>', principal='impala/redacted'} is not allowed to access DataStreamService
```

Workaround: Start Impala with the `--use_system_auth_to_local=false` flag to ignore the system-wide `auth_to_local` mappings configured in `/etc/krb5.conf`.

Apache Issue: [KUDU-2198](#)

Affected Versions: CDH 5.15, CDH 6.1 and higher**Impala Known Issues: Resources**

These issues involve memory or disk usage, including out-of-memory conditions, the spill-to-disk feature, and resource management features.

Handling large rows during upgrade to CDH 5.13 / Impala 2.10 or higher

After an upgrade to CDH 5.13 / Impala 2.10 or higher, users who process very large column values (long strings), or have increased the --read_size configuration setting from its default of 8 MB, might encounter capacity errors for some queries that previously worked.

Resolution: After the upgrade, follow the instructions in [Handling Large Rows During Upgrade to CDH 5.13 / Impala 2.10 or Higher](#) to check if your queries are affected by these changes and to modify your configuration settings if so.

Affected Versions: All CDH 6 versions**Apache Issue:** [IMPALA-6028](#)**Configuration to prevent crashes caused by thread resource limits**

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal error: Unable to allocate section memory!
terminate called after throwing an instance of
'boost::exception_detail::clone_impl<boost::exception_detail::error_info_injector<boost::thread_resource_error>'
```

Workaround:

In CDH 6.0 and lower versions of CDH, configure each host running an impalad daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

In CDH 6.1 and higher versions, it is unlikely that you will hit the thread resource limit. Configure each host running an impalad daemon with the following setting:

```
echo 8000000 > /proc/sys/vm/max_map_count
```

To make the above settings durable, refer to your OS documentation. For example, on RHEL 6.x:

1. Add the following line to /etc/sysctl.conf:

```
vm.max_map_count=8000000
```

2. Run the following command:

```
sysctl -p
```

Affected Versions: All CDH 6 versions**Apache Issue:** [IMPALA-5605](#)**Breakpad minidumps can be very large when the thread count is high**

The size of the breakpad minidump files grows linearly with the number of threads. By default, each thread adds 8 KB to the minidump size. Minidump files could consume significant disk space when the daemons have a high number of threads.

Workaround: Add `--minidump_size_limit_hint_kb=size` to set a soft upper limit on the size of each minidump file. If the minidump file would exceed that limit, Impala reduces the amount of information for each thread from 8 KB to 2 KB. (Full thread information is captured for the first 20 threads, then 2 KB per thread after that.) The minidump file can still grow larger than the "hinted" size. For example, if you have 10,000 threads, the minidump file can be more than 20 MB.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3509](#)

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the impalad daemon.

Workaround: To monitor overall memory usage, use the top command, or add the memory figures in the Impala web UI /memz tab to JVM memory usage shown on the /metrics tab.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-691](#)

Impala Known Issues: Correctness

These issues can cause incorrect or unexpected results from queries. They typically only arise in very specific circumstances.

Incorrect result due to constant evaluation in query with outer join

An OUTER JOIN query could omit some expected result rows due to a constant such as FALSE in another join clause. For example:

```
explain SELECT 1 FROM alltypestiny a1
    INNER JOIN alltypesagg a2 ON a1.smallint_col = a2.year AND false
    RIGHT JOIN alltypes a3 ON a1.year = a1.bigint_col;
+-----+
| Explain String
+-----+
| Estimated Per-Host Requirements: Memory=1.00KB Vcores=1
| 00:EMPTYSET
+-----+
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3094](#)

% escaping does not work correctly in a LIKE clause

If the final character in the RHS argument of a LIKE operator is an escaped \% character, it does not match a % final character of the LHS argument.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2422](#)

Crash: impala::Coordinator::ValidateCollectionSlots

A query could encounter a serious error if includes multiple nested levels of INNER JOIN clauses involving subqueries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2603](#)

Impala Known Issues: Metadata

These issues affect how Impala interacts with metadata. They cover areas such as the metastore database and the Impala Catalog Server daemon.

Concurrent catalog operations with heavy DDL workloads can cause queries with SYNC_DDL to fail fast

When Catalog Server is under a heavy load with concurrent catalog operations of long running DDLs, queries running with the `SYNC_DDL` query option can fail with the following message:

```
ERROR: CatalogException: Couldn't retrieve the catalog topic
version for the SYNC_DDL operation after 3 attempts. The operation has
been successfully executed but its effects may have not been
broadcast to all the coordinators.
```

The catalog operation is actually successful as the change has been committed to HMS and Catalog Server cache, but when Catalog Server notices a longer than expected time for it to broadcast the changes, it fails fast.

The coordinator daemons eventually sync up in the background.

Affected Versions: CDH versions 6.0 and 6.1

Apache Issue: [IMPALA-7961](#) / CDH-76345

Impala Known Issues: Interoperability

These issues affect the ability to interchange data between Impala and other systems. They cover areas such as data types and file formats.

Queries Stuck on Failed HDFS Calls and not Timing out

If the following error appears multiple times in a short duration while running a query, it would mean that the connection between the `impalad` and the HDFS NameNode is in a bad state and hence the `impalad` would have to be restarted:

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed to finish before the
<hdfs_operation_timeout_sec> second timeout "
```

Workaround: Restart the `impalad` in the bad state.

Affected Versions: All versions of Impala

Apache Issue: [HADOOP-15720](#)

Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink `hdfs.writeFormat` must be set to `Text`, rather than its default value of `Writable`. The `hdfs.writeFormat` setting must be changed to `Text` before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

Resolution: This information has been requested to be added to the upstream Flume documentation.

Affected Versions: All CDH 6 versions

Cloudera Issue: CDH-13199

Avro Scanner fails to parse some schemas

The default value in Avro schema must match the first union type. For example, if the default value is `null`, then the first type in the `UNION` must be "`null`".

Workaround: Swap the order of the fields in the schema specification. For example, use `["null", "string"]` instead of `["string", "null"]`. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-635](#)

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1024](#)

Incorrect results with basic predicate on CHAR typed column

When comparing a CHAR column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1652](#)

Impala Known Issues: Limitations

These issues are current limitations of Impala that require evaluation as you plan how to integrate Impala into your data management workflow.

Set limits on size of expression trees

Very deeply nested expressions within queries can exceed internal Impala limits, leading to excessive memory usage.

Workaround: Avoid queries with extremely large expression trees. Setting the query option `disable_codegen=true` may reduce the impact, at a cost of longer query runtime.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4551](#)

Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

Workaround: Use standard HDFS on all Impala nodes.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-77](#)

Hue and BDR require separate parameters for Impala Load Balancer

Cloudera Manager supports a single parameter for specifying the Impala Daemon Load Balancer. However, because BDR and Hue need to use different ports when connecting to the load balancer, it is not possible to configure the load balancer value so that BDR and Hue will work correctly in the same cluster.

Workaround: To configure BDR with Impala, use the load balancer configuration either without a port specification or with the Beeswax port.

To configure Hue, use the **Hue Server Advanced Configuration Snippet (Safety Valve)** for `impalad_flags` to specify the load balancer address with the HiveServer2 port.

Affected Versions: CDH versions from 5.11 to 6.0.1

Cloudera Issue: [OPSAPS-46641](#)

Impala Known Issues: Miscellaneous / Older Issues

These issues do not fall into one of the above categories or have not been categorized yet.

A failed CTAS does not drop the table if the insert fails

If a `CREATE TABLE AS SELECT` operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed `CREATE TABLE AS SELECT`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2005](#)

Casting scenarios with invalid/inconsistent results

Using a `CAST` function to convert large literal values to smaller types, or to convert special values such as `NaN` or `Inf`, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1821](#)

Impala Parser issue when using fully qualified table names that start with a number

A fully qualified table name starting with a number could cause a parsing error. In a name such as `db.571_market`, the decimal point followed by digits is interpreted as a floating-point number.

Workaround: Surround each part of the fully qualified name with backticks (`` ``).

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-941](#)

Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

Workaround: Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) for details about modifying these environment settings.

Resolution: Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-532](#)

EMC Isilon Known Issues

CDH 6.0 is not currently supported on EMC Isilon.

Affected Versions: All CDH 6 versions

Apache Kafka Known Issues

Topics Created with the "kafka-topics" Tool Might Not Be Secured

Topics that are created and deleted via Kafka are secured (for example, auto created topics). However, most topic creation and deletion is done via the `kafka-topics` tool, which talks directly to ZooKeeper or some other third-party tool that talks directly to ZooKeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. Anyone with access to ZooKeeper can create and delete topics. They will not be able to describe, read, or write to the topics even if they can create them.

The following commands talk directly to ZooKeeper and therefore are not secured via Kafka:

- `kafka-topics.sh`
- `kafka-configs.sh`
- `kafka-preferred-replica-election.sh`
- `kafka-reassign-partitions.sh`

"`offsets.topic.replication.factor`" Must Be Less Than or Equal to the Number of Live Brokers

The `offsets.topic.replication.factor` broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a `GROUP_COORDINATOR_NOT_AVAILABLE` error until the cluster size meets this replication factor requirement.

Requests Fail When Sending to a Nonexistent Topic with "`auto.create.topics.enable`" Set to True

The first few `produce` requests fail when sending to a nonexistent topic with `auto.create.topics.enable` set to true.

Workaround: Increase the number of retries in the Producer configuration setting `retries`.

Custom Kerberos Principal Names Cannot Be Used for Kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start.

Workaround: None. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Performance Degradation When SSL Is Enabled

Significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with `ssl.secure.random.implementation = SHA1PRNG`. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Affected Versions: CDK 2.x and later

Fixed Versions: None

Apache Issue: KAFKA-2561

Cloudera Issue: None

Kafka Broker Fails to Start Due to Slow Sentry and HMS startup

This issue is encountered on cluster startup and is caused by misalignment between Kafka, Sentry, and HMS. The slow startup of HMS slows down Sentry startup which consequently makes the Kafka connection to Sentry time out. Ultimately, the Kafka broker will be unable to start.

Workaround: Manually increase the number of remote procedure call retries between Sentry and Kafka through the **Sentry Client Advanced Configuration Snippet (Safety Valve)** for `sentry-site.xml` property.

1. Go to **Sentry > Configuration** and find the **Sentry Client Advanced Configuration Snippet (Safety Valve) for `sentry-site.xml`** property.
2. Click on the add button.
3. Enter the following data:
 - Name: `sentry.service.client.rpc.retry-total`
 - Value: 20
4. Enter a **Reason for change**, and then click **Save Changes** to commit the changes.
5. Return to the Home page by clicking the Cloudera Manager logo.
6. Click the restart stale services icon next to the Sentry service to invoke the cluster restart wizard.
7. Click **Restart Stale Services**.
8. Click **Restart Now**.
9. Click **Finish**.

Affected Versions: CDH 6.1.0 and higher

Fixed Versions: N/A

Cloudera Issue: CDH-74713

Kafka JMX Tool Cannot Connect to JMX

The Kafka JMX tool cannot connect to the JMX agent of the Kafka Broker or MirrorMaker if the specified address of the JMX remote connector is bound to 127.0.0.1.

Workaround:

1. In Cloudera Manager go to **Kafka > Instances** and select the affected broker.
2. Find the **Additional Broker Java Options** and **Additional MirrorMaker Java Options** properties and add the following Java option to the configuration:

```
-Djava.rmi.server.hostname=127.0.0.1
```



Note: Configuring the **Additional MirrorMaker Java Options** property is only required if you are using JMX with MirrorMaker.

3. Restart the affected brokers.

Affected Versions: CDH 6.0.0 and higher

Fixed Versions: CDH 6.2.0

Cloudera Issue: OPSAPS-48695

Apache Kudu Known Issues

The following are known bugs and issues in Kudu. Note that this list is not exhaustive, and is meant to communicate only the most important known issues.

C++ Client Fails to Re-acquire Authentication Token in Multi-master Clusters

A security-related issue can cause Impala queries to start failing on busy clusters in the following scenario:

- The cluster runs with the `--rpc_authentication` set as optional or required. The default is optional. Secure clusters use required.
- The cluster is using multiple masters.
- Impala queries happen frequently enough that the leader master connection to some `impalad` isn't idle-closed (more than 1 query per 65 seconds).
- The connection stays alive for longer than the authentication token timeout (1 week by default).
- A master leadership change occurs after the authentication token expiration.

Impala queries will start failing with errors in the `impalad` logs like:

```
I0904 13:53:08.748968 95857 client-internal.cc:283] Unable to determine the new leader
Master: Not authorized: Client connection negotiation failed: client connection to
10.164.44.13:7051: FATAL_INVALID_AUTHENTICATION_TOKEN: Not authorized: authentication
token expired
I0904 13:53:10.389009 95861 status.cc:125] Unable to open Kudu table: Timed out:
GetTableSchema timed out after deadline expired
@ 0x95b1e9 impala::Status::Status()
@ 0xff22d4 impala::KuduScanNodeBase::Open()
@ 0xff101e impala::KuduScanNode::Open()
@ 0xb73ced impala::FragmentInstanceState::Open()
@ 0xb7532b impala::FragmentInstanceState::Exec()
@ 0xb64ae8 impala::QueryState::ExecFInstance()
@ 0xd15193 impala::Thread::SuperviseThread()
@ 0xd158d4 boost::detail::thread_data<>::run()
@ 0x129188a (unknown)
@ 0x7f717ceade25 start_thread
@ 0x7f717cbdb34d __clone
```

Impala shell queries will fail with a message like:

```
Unable to open Kudu table: Timed out: GetTableSchema timed out after deadline expired
```

Workaround:

- Restart the affected Impala Daemons. Restarting a daemon ensures the problem will not reoccur for at least the authentication token lifetime, which defaults to one week.
- Increase the authentication token lifetime (`--authn_token_validity_seconds`). Beware that raising this lifetime increases the window of vulnerability of the cluster if a client is compromised. It is recommended that you keep the token lifetime at one month maximum for a secure cluster. For unsecured clusters, a longer token lifetime is acceptable, and a 3 month lifetime is recommended.

Affected Versions: From CDH 5.11 through CDH 6.0.1

Apache Issue: [KUDU-2580](#)

Timeout Possible with Log Force Synchronization Option

If the Kudu master is configured with the `-log_force_fsync_all` option, tablet servers and clients will experience frequent timeouts, and the cluster may become unusable.

Affected Versions:

All CDH 6 versions

Longer Startup Times with a Large Number of Tablets

If a tablet server has a very large number of tablets, it may take several minutes to start up. It is recommended to limit the number of tablets per server to 1000 or fewer. The maximum allowed number of tablets is 2000 per server. Consider this limitation when pre-splitting your tables. If you notice slow start-up times, you can monitor the number of tablets per server in the web UI.

Affected Versions:

All CDH 6 versions

Descriptions for Kudu TLS/SSL Settings in Cloudera Manager

Use the descriptions in the following table to better understand the TLS/SSL settings in the Cloudera Manager Admin Console.

Field	Usage Notes
Kerberos Principal	Set to the default principal, <code>kudu</code> .
Enable Secure Authentication And Encryption	Select this checkbox to enable authentication <i>and</i> RPC encryption between all Kudu clients and servers, as well as between individual servers. Only enable this property after you have configured Kerberos.
Master TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu master host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to the Kudu master web UI.
Tablet Server TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu tablet server host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to Kudu tablet server web UIs.
Master TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu master host's private key (set in Master TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Tablet Server TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu tablet server host's private key (set in Tablet Server TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Master TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Tablet Server TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Enable TLS/SSL for Master Server	Enables HTTPS encryption on the Kudu master web UI.
Enable TLS/SSL for Tablet Server	Enables HTTPS encryption on the Kudu tablet server web UIs.

Affected Versions:

All CDH 6 versions

Apache Oozie Known Issues

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used due to a sys table change in PostgreSQL from version 9.5 to 9.6. The failure only happens if Oozie uses a JDBC driver earlier than 9.4.1209.

Workaround:

1. After the parcels of the new version are distributed, replace the PostgreSQL JDBC driver with a newer one (version 9.4.1209 or higher) in the new parcel, at the following locations:
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/lib/
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/libtools/
2. Perform the upgrade.



Note: If you already started the upgrade and the process stops with an error message about missing columns, you can change the drivers at that point of the process as well, and resume the upgrade.

If your cluster is installed from packages, you must change the drivers at the following locations:

- /usr/lib/oozie/libtools/
- /usr/lib/oozie/lib/



Note: You can change the driver after the packages installation, but before running the CDH upgrade wizard. You can also do it during the update process, when the error occurs.

You can download the driver from the [PostgreSQL JDBC driver homepage](#).

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-75951

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the `resume` command to tell Oozie to continue the workflow from the point at which it left off.

Affected Versions: CDH 5 and higher

Cloudera Issue: CDH-14623

Oozie works with MapReduce or YARN, but not both

The Oozie server works with a MapReduce (MRv1) cluster or a YARN (MRv2) cluster, but not both at the same time.

Workaround: Use two different Oozie servers.

Affected Versions: All

[Apache Parquet Known Issues](#)

There are no known issues in Parquet.

[Apache Pig Known Issues](#)

There are no known issues in this release.

[Cloudera Search Known Issues](#)

The current release includes the following known limitations:

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: `solr.sentry.enableSqlQuery=true`
- Graph: `solr.sentry.enableGraphQuery=true`
- Stream: `solr.sentry.enableStreams=true`

Workaround: None

Affected Versions: All CDH 6 releases

Cloudera Issue: CDH-66345

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search for CDH 5.5.0 includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-34050

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with `CrunchIndexerTool`, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-22190

The `quickstart.sh` file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if ZooKeeper and the NameNode are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the NameNode and ZooKeeper are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the MapReduce jobs and Solr commands that are run by the `quickstart`.

Affected Versions: All

Cloudera Issue: CDH-19923

Field value class guessing and Automatic schema field addition are not supported with the `MapReduceIndexerTool` nor the `HBaseMapReduceIndexerTool`

The `MapReduceIndexerTool` and the `HBaseMapReduceIndexerTool` can be used with a Managed Schema created via NRT indexing of documents or via the Solr Schema API. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the `MapReduceIndexerTool` or `HBaseMapReduceIndexerTool`. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the Solr Schema API or index sample documents using NRT indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect using the List Fields API command.

Affected Versions: All

Cloudera Issue: CDH-26856

The `Browse` and `Spell` Request Handlers are not enabled in schemaless mode

The `Browse` and `Spell` Request Handlers require certain fields be present in the schema. Since those fields cannot be guaranteed to exist in a Schemaless setup, the `Browse` and `Spell` Request Handlers are not enabled by default.

Workaround: If you require the “Browse” and “Spell” Request Handlers, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Affected Versions: All

Cloudera Issue: CDH-19407

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. Blockcache writing is disabled by default.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-17978

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given page explaining that access is denied, and instead receive a web page that never finishes loading.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-58276

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection.

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Affected Versions: All

Cloudera Issue: CDH-15441

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Affected Versions: All

Cloudera Issue: CDH-58694

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-21162

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-11357

Apache Sentry Known Issues

When granting privileges, a single transaction per grant causes long delays

Sentry takes a long time to grant or revoke a large number of column-level privileges that are requested in a single statement. For example if you execute the following command:

```
GRANT SELECT(col1, col2, ...) ON TABLE table1;
```

Sentry applies the grants to each column separately and the refresh process causes long delays.

Workaround: Split the grant statement up into smaller chunks. This prevents the refresh process from causing delays.

Affected Versions:

- CDH: 5.14.4
- CDH: 5.15.1
- CDH: 5.16.0
- CDH: 6.1.0

Fixed Versions:

- CDH 5.16.1 and above
- CDH 6.2.0 and above

Cloudera Issue: CDH-74982

SHOW ROLE GRANT GROUP raises exception for a group that was never granted a role

If you run the command SHOW ROLE GRANT GROUP for a group that has never been granted a role, beeline raises an exception. However, if you run the same command for a group that does not have any roles, but has at one time been granted a role, you do not get an exception, but instead get an empty list of roles granted to the group.

Workaround: Adding a role will prevent the exception.

Affected Versions:

- CDH 5.16.0
- CDH 6.0.0

Cloudera Issue: CDH-71694

GRANT/REVOKE operations could fail if there are too many concurrent requests

Under a significant workload, Grant/Revoke operations can have issues.

Workaround: If you need to make many privilege changes, plan them at a time when you do not need to do too many at once.

Affected Versions: CDH 5.13.0 and above

Apache Issue: [SENTRY-1855](#)

Cloudera Issue: CDH-56553

Creating large set of Sentry roles results in performance problems

Using more than a thousand roles/permissions might cause significant performance problems.

Workaround: Plan your roles so that groups have as few roles as possible and roles have as few permissions as possible.

Affected Versions: CDH 5.13.0 and above

Cloudera Issue: CDH-59010

Users can't track jobs with Hive and Sentry

As a prerequisite of enabling Sentry, Hive impersonation is turned off, which means all YARN jobs are submitted to the Hive job queue, and are run as the `hive` user. This is an issue because the YARN History Server now has to block users from accessing logs for their own jobs, since their own usernames are not associated with the jobs. As a result, end users cannot access any job logs unless they can get `sudo` access to the cluster as the `hdfs`, `hive` or other admin users.

In CDH 5.8 (and higher), Hive overrides the default configuration, `mapred.job.queuename`, and places incoming jobs into the connected user's job queue, even though the submitting user remains `hive`. Hive obtains the relevant queue/username information for each job by using YARN's `fair-scheduler.xml` file.

Affected Versions: CDH 5.2.0 and above

Cloudera Issue: CDH-22890

Column-level privileges are not supported on Hive Metastore views

`GRANT` and `REVOKE` for column level privileges is not supported on Hive Metastore views.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-754](#)

`SELECT` privilege on all columns does not equate to `SELECT` privilege on table

Users who have been explicitly granted the `SELECT` privilege on all columns of a table, will *not* have the permission to perform table-level operations. For example, operations such as `SELECT COUNT (1)` or `SELECT COUNT (*)` will not work even if you have the `SELECT` privilege on all columns.

There is one exception to this. The `SELECT * FROM TABLE` command will work even if you do not have explicit table-level access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-838](#)

The `EXPLAIN SELECT` operation works without table or column-level privileges

Users are able to run the `EXPLAIN SELECT` operation, exposing metadata for all columns, even for tables/columns to which they weren't explicitly granted access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-849](#)

Object types Server and URI are not supported in `SHOW GRANT ROLE roleName on OBJECT objectName`

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Apache Issue: N/A

Cloudera Issue: CDH-19430

Relative URI paths not supported by Sentry

Sentry supports only absolute (not relative) URI paths in permission grants. Although some early releases (for example, CDH 5.7.0) might not have raised explicit errors when relative paths were set, upgrading a system that uses relative paths causes the system to lose Sentry permissions.

Affected Versions: All versions. Relative paths are not supported in Sentry for permission grants.

Resolution: Revoke privileges that have been set using relative paths, and grant permissions using absolute paths before upgrading.

Absolute (Use this form)	Relative (Do not use this form)
<code>hdfs://absolute/path/</code>	<code>hdfs://relative/path</code>
<code>s3a://bucketname/</code>	<code>s3a://bucketname</code>

Apache Spark Known Issues

The following sections describe the current known issues and limitations in Apache Spark 2.x as distributed with CDH 6.1.x. In some cases, a feature from the upstream Apache Spark project is currently not considered reliable enough to be supported by Cloudera.

RDD.repartition() has different failure handling in Spark 2.4 and may cause job failures

The `RDD.repartition()` transformation, which reshuffles data in the RDD randomly to create either more or fewer partitions and then balances it across the partitions, was using a round-robin method to distribute data that caused incorrect answers to be returned for RDD jobs. This issue has been corrected, but it introduced a behavior change in RDD job failure handling. Now, Spark actively fails a job if there is a fetch failure that was caused by a node failure after repartitioning.

Workaround: Use the `RDD.checkpoint()` method to save the intermediate RDD data to HDFS. First, call `SparkContext.setCheckpointDir(directory: String)` to set the checkpoint directory where the intermediate data will be saved. Note that the directory must be an HDFS path. Then mark the RDD for checkpointing by calling `RDD.checkpoint()` when you use the `RDD.repartition()` transformation.

Apache Issue: [SPARK-23243](#)

Cloudera Issue: CDH-76413

Spark Streaming write-ahead logs do not run on HDFS directories with Erasure Coding enabled

Spark Streaming write-ahead logs (WALs) cannot run on HDFS directories when Erasure Coding is enabled. Erasure Coding does not support `hflush()`, `hsync()`, and `append()`, which prevents the WALs from running.

Workaround: Configure Spark Streaming with a checkpoint directory that does not have Erasure Coding enabled on it. You can set the checkpoint directory with `ssc.checkpoint("directory_name")`. For example:

```
ssc.checkpoint("_checkpoint")
```

Affected Versions: CDH 6.1.0

Fixed Versions: CDH 6.2.0

Apache Issue: [SPARK-26094](#)

Cloudera Issue: CDH-61127

PySpark broadcast variables fail when disk encryption is enabled

When disk encryption is enabled, PySpark broadcast variables fail with the following stack trace:

```
Traceback (most recent call last): File "broadcast.py", line 37, in <module>
words_new.value File "/pyspark.zip/pyspark/broadcast.py", line 137, in value
File "pyspark.zip/pyspark/broadcast.py", line 122, in load_from_path File
"pyspark.zip/pyspark/broadcast.py", line 128, in load EOFError: Ran out of input
```

Workaround: None

Affected Versions: CDH 6.0.1, CDH 6.1.0, CDH 6.2.0

Apache Issue: [SPARK-26201](#)

Cloudera Issue: CDH-76116

Structured Streaming exactly-once fault tolerance constraints

In Spark Structured Streaming, the exactly-once fault tolerance for `file sink` is valid only for files that are in the manifest. These files are located in the `_spark_metadata` subdirectory of the `file sink` output directory. Only process files that have file names starting with digits. Other temporary files can also appear in this directory, but they should not be processed. Typically, these temporary file file names start with a period (".").

You can list the valid manifest files, excluding the temporary files, by using a command like the following, which assumes your output directory is located at `/tmp/output`. As the appropriate user, run the following command to list the valid manifest files:

```
hadoop fs -ls /tmp/output/_spark_metadata/[0-9]*
```

Workaround: None

Affected Versions: CDH 6.1.0 and higher

Cloudera Issue: CDH-75191

Spark SQL does not respect size limit for the varchar type

Spark SQL treats `varchar` as a string (that is, there no size limit). The observed behavior is that Spark reads and writes these columns as regular strings; if inserted values exceed the size limit, no error will occur. The data will be truncated when read from Hive, but not when read from Spark.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Apache Issue: [SPARK-5918](#)

Cloudera Issue: CDH-33642

Spark SQL does not prevent you from writing key types not supported by Avro tables

Spark allows you to declare DataFrames with any key type. Avro supports only string keys and trying to write any other key type to an Avro table will fail.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33648

Spark SQL does not support timestamp in Avro tables

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33649

Spark SQL does not respect Sentry ACLs when communicating with Hive metastore

Even if user is configured via Sentry to not have read permission to a Hive table, a Spark SQL job running as that user can still read the table's metadata directly from the Hive metastore. **Cloudera Issue:** CDH-33658

Dynamic allocation and Spark Streaming

If you are using Spark Streaming, Cloudera recommends that you disable dynamic allocation by setting `spark.dynamicAllocation.enabled` to `false` when running streaming applications.

Limitation with Region Pruning for HBase Tables

When SparkSQL accesses an HBase table through the HiveContext, region pruning is not performed. This limitation can result in slower performance for some SparkSQL queries against tables that use the HBase SerDes than when the same table is accessed through Impala or Hive.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-56330

Running `spark-submit` with `--principal` and `--keytab` arguments does not work in client mode

The `spark-submit` script's `--principal` and `--keytab` arguments do not work with Spark-on-YARN's client mode.

Workaround: Use `cluster` mode instead.

Affected Versions: All

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Apache Sqoop Known Issues

Column names cannot start with a number when importing data with the `--as-parquetfile` option.

Currently, Sqoop is using an Avro schema when writing data as a parquet file. The Avro schema requires that column names do not start with numbers, therefore Sqoop is renaming the columns in this case, prepending them with an underscore character. This can lead to issues when one wants to reuse the data in other tools, such as Impala.

Workaround: Rename the columns to comply with Avro limitations (start with letters or underscore, as specified in the [Avro documentation](#)).

Cloudera Issue: None

MySQL JDBC driver shipped with CentOS 6 systems does not work with Sqoop

CentOS 6 systems currently ship with version 5.1.17 of the MySQL JDBC driver. This version does not work correctly with Sqoop.

Workaround: Install version 5.1.31 of the JDBC driver as detailed in [Installing the JDBC Drivers for Sqoop 1](#).

Affected Versions: MySQL JDBC 5.1.17, 5.1.4, 5.3.0

Cloudera Issue: CDH-23180

MS SQL Server "integratedSecurity" option unavailable in Sqoop

The `integratedSecurity` option is not available in the Sqoop CLI.

Workaround: None

Cloudera Issue: None

Sqoop1 (doc import + `--as-parquetfile`) limitation with KMS/KTS Encryption at Rest

Due to a limitation with Kite SDK, it is not possible to use (`sqoop import --as-parquetfile`) with KMS/KTS Encryption zones. See the following example.

```
sqoop import --connect jdbc:db2://djaxludb1001:61035/DBBAT003 --username=dh810202 --password=XXXXXXXXXX --target-dir /data/hive_scratch/ASDISBURSEMENT --delete-target-dir -m1 --query "select disbursementnumber,disbursementdate,xmldata FROM DB2dba.ASDISBURSEMENT where DISBURSEMENTNUMBER = 2011113210000115311 AND \$CONDITIONS" -hive-import --hive-database adminserver -hive-table asdisbursement_dave --map-column-java XMLDATA=String --as-parquetfile

16/12/05 12:23:46 INFO mapreduce.Job: map 100% reduce 0%
16/12/05 12:23:46 INFO mapreduce.Job: Job job_1480530522947_0096 failed with state FAILED
due to: Job commit failed: org.kitesdk.data.DatasetIOException: Could not move contents
of
hdfs://AJAX01-ns/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096
to hdfs://AJAX01-ns/data/RetiredApps/INS/AdminServer/asdisbursement_dave
<SNIP>
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.IOException):
```

```
/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096/5ddcac42-5d69-4e46-88c2-17bedac4858.parquet  
can't be moved into an encryption zone.
```

Workaround: If you use the Parquet Hadoop API based implementation for importing into Parquet, specify a `--target-dir` which is the same encryption zone as the Hive warehouse directory.

If you use the Kite Dataset API based implementation, use an alternate data file type, for example text or avro.

Apache Issue: SQUOP-2943

Cloudera Issue: CDH-40826

Doc import as Parquet files may result in out-of-memory errors

Out-of-memory (OOM) errors can be caused in the following two cases:

- With many very large rows (multiple megabytes per row) before initial-page-run check (ColumnWriter)
- When rows vary significantly by size so that the next-page-size check is based on small rows and is set very high followed by many large rows

Apache Issue: PARQUET-99

Workaround: None, other than restructuring the data.

Apache ZooKeeper Known Issues

There are no known issues in this release.

[CDH 6.1.0 Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for CDH 6.1.0:

[New Features in CDH 6.1.0](#)

See below for new features in CDH 6.1.0, grouped by component:

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.1.x cluster is supported. Apache Accumulo is shipped separately from CDH.

Apache Avro

There are no notable new features in this release.

Apache Crunch

There are no notable new features in this release.

Apache Flume

The following list shows what's new and changed in Apache Flume for CDH 6.1.0:

- Flume JMS support for TLS connections including client certificate authentication. For details, see [Configuring TLS/SSL for Flume JMS Source](#).
- Flume Auto-TLS support. For details, see [Configuring TLS/SSL for Flume](#).

Apache Hadoop

Hadoop Common

There are no notable new features in this release.

HDFS

ADLS Gen2

CDH supports using ADLS Gen2 as a storage layer for MapReduce, Hive on MapReduce, Hive on Spark, Spark, Oozie, and Impala.

Cloudera Enterprise 6 Release Guide

For more information, see the ADLS Gen2 documentation. For information about configuring CDH and ADLS Gen2, see [Configuring ADLS Gen2 Connectivity](#).

Google Cloud Storage

CDH supports using Google Cloud Storage (GCS) as a storage layer for Hive, MapReduce, and Spark. To use GCS, you must download the connector and distribute it to your cluster. For more information about how to do this and limitations, see [Configuring Google Cloud Storage Connectivity](#).

CacheReplicationMonitor

You can now disable the CacheReplicationMonitor with the following **Advanced Configuration Snippet (Safety Valve)** for `hdfs-site.xml`: `dfs.namenode.caching.enabled`. To maintain backwards compatibility, the default value is `true` to enable the default caching. To disable the CacheReplicationMonitor, set the value to `false` when you add the **Advanced Configuration Snippet (Safety Valve)** for `hdfs-site.xml`:

Erasure Coding

CDH 6.1.0 supports Erasure Coding (EC). EC is an alternative to the standard 3x replication that HDFS uses by default. When an HDFS cluster uses EC, no additional direct copies of the data are generated. Instead, data is striped into blocks and encoded to generate parity blocks. If there are any missing or corrupt blocks, HDFS uses the remaining data and parity blocks to reconstruct the missing pieces in the background. This process provides a similar level of data durability to 3x replication but at a lower storage cost.

For more information, see [Data Durability](#).

Snapshots

You can now enable immutable snapshots for HDFS with Cloudera Manager. Enabling this feature also enables snapshot diff-based copy listing for BDR.

In the **Cloudera Manager Admin Console**, navigate to **Clusters > <HDFS cluster> > Configuration** and search for **Enable Immutable Snapshots**.

This feature is off by default.

MapReduce

There are no notable new features in this release.

YARN

There are no notable new features in this release.

Apache HBase

[Replication Status of WAL Groups in Web UI](#)

New sections are added to Web UI to show the status of replication:

- **Peers**: Shows all replication peers and some of their configuration, including peer id, cluster key, state, bandwidth, the size of current log, the log queue size, the replication offset and which namespace or table it replicates.
- **Replication status of all Region Server**: Shows the delay of replication, including the `AgeOfLastShippedOp`, `SizeOfLogQueue` and `ReplicationLag` for each Region Server.

If the replication offset shows `-1` and replication delay is `UNKNOWN`, that means replication is not started. There are two common reasons for this: the peer is disabled or the `replicationEndpoint` is sleeping.

[Default Behavior Change](#)

[HBASE-20856](#): By default the meta WAL provider (`hbase.wal.meta_provider`) is set to the same as the normal WAL (`hbase.wal.provider`).

[Apache Hive / Hive on Spark / HCatalog](#)

Apache Hive

The following are some of the notable new features in this release of Hive:

Erasure Coding Support

You can now use Erasure Coding (EC) with your infrequently accessed Hive tables and partitions. Learn how to plan, evaluate, and activate Erasure Coding in our [Best Practices for Using Hive with Erasure Coding](#) guide.

Query Plan Graph View for Hive Web UI

You can now view your query plans on an informative and visual graph. Learn how to activate the graph view, start understanding your query plans, follow the MapReduce progress bar, and pinpoint errors easily with the [Query Plan Graph View for Hive Web UI](#).

Fine Grained Privileges

Sentry and Hive introduced fine grained privileges to provide object-level privileges to roles.

Fine grained privileges adds the CREATE privilege, which allow users to create databases and tables. See the [Sentry Privileges](#) documentation for more information about the new privileges.

Object Ownership

Object ownership designates an owner for a database, table, or view in Sentry. The owner of an object has the equivalent of the ALL privilege on the object. See the [Object Ownership](#) documentation for information about enabling object ownership.

Because of the new object ownership feature, HMS stores the user that creates a table or database in Hive as the object owner. If object ownership is enabled, Sentry grants the user the OWNER privilege. Whether or not object ownership is enabled, HMS stores the user that creates the object as the object owner. Previously, HMS stored the hive user as the object owner.

The following statements were added to Hive to support object ownership via Sentry:

- ALTER DATABASE SET OWNER
- ALTER TABLE SET OWNER
- SHOW GRANT USER

Hive on Spark

There are no notable new features in this release.

Hue

The following are some of the notable new features in this release of Hue:

- Language Reference built-in, Column Sampling, black theme for Editor
- Simplifying the end user Data Catalog search
- Improved SQL Exploration

For more information, see <http://gethue.com/additional-sql-improvements-in-hue-4-3/>.

Apache Impala

Enhancements in Authorization

Fine-grained Privileges

Sentry and Impala introduced fine-grained privileges to provide object-level privileges to roles.

Fine-grained privileges include the REFRESH and CREATE privileges, which allow users to create databases and tables, and to execute commands that update metadata information on Impala databases and tables. See [Impala Sentry documentation](#) for the new privileges and the scopes of the objects that you can grant the new privileges on.

The following new privileges were added:

- The REFRESH privilege
- The CREATE privilege
- The SELECT and INSERT privileges on SERVER

If a role has `SELECT` or `INSERT` privilege on an object in Impala before upgrading to CDH 6.1, that role will automatically get the `REFRESH` privilege during the upgrade.

Object Ownership

Object ownership designates an owner for a database, table, or view in Sentry. The owner of an object has the `OWNER` privilege which is the equivalent of the `ALL` privilege on the object. See [Object Ownership](#) for information about enabling object ownership.

If the object ownership feature is enabled, Sentry grants the user the `OWNER` privilege. Whether or not object ownership is enabled, HMS stores the object creator as the default object owner. Previously, HMS stored the Kerberos user as the object owner.

The following statements were added in Impala to support object ownership via Sentry:

- `ALTER DATABASE SET OWNER`
- `ALTER TABLE SET OWNER`
- `ALTER VIEW SET OWNER`
- `SHOW GRANT USER`

Enhancements in Admission Control and Resource Management

The following is a list of noteworthy improvements made in Impala in resource management and admission control.

- Starting in CDH 6.1 / Impala 3.1, Impala automatically chooses how much memory to give a query based on the memory estimate from the planner and bounded by the min/max guardrails that you configure for resource pools. In previous versions, you were required to set a single memory limit (via the `mem_limit` setting) per resource pool.

The following new resource pool settings can be configured in Cloudera Manager or in the admission control configuration files:

- Minimum Query Memory Limit (`min-query-mem-limit`)
- Maximum Query Memory Limit (`max-query-mem-limit`)
- Clamp MEM_LIMIT Query Option (`clamp-mem-limit-query-option`)

See [Admission Control and Query Queuing](#) for detail information.

- Improvements to prevent scan operators running out of memory in many more scenarios.
- Many improvements for more accurate memory estimates in admission control.
- New query options to reject complex queries.

The options are enforced by admission control based on planner resource requirements and the schedule.

- `THREAD RESERVATION LIMIT` limits the total number of reserved threads in fragments scheduled on a single backend.
- `THREAD RESERVATION AGGREGATE LIMIT` limits the sum of reserved threads across all fragments.

IANA Time Zones Support

Now you have an option to customize time zone databases in Impala with well-known sources, such as IANA.

- The `--hdfs_zone_info_zip` startup flag specifies the path to a zip archive that contains the IANA time zone database. The default location of the time zone database is the `/usr/share/zoneinfo` folder. See [Customizing Time Zones](#) for more information and the steps to set the flag.
- The `--hdfs_zone_alias_conf` startup flag specifies the path to a configuration file that contains definitions for non-standard timezone aliases. See [Customizing Time Zones](#) for more information and the steps to set the flag.
- The new [`TIMEZONE` Query Option \(CDH 6.1 / Impala 3.1 or higher only\)](#) query option defines the local time zone to be used for conversions between UTC and the local time. By default, the coordinator node's time zone is used as the local time zone. See [TIMESTAMP Data Type](#) for detail.

General Performance Improvements

A new query option, [SHUFFLE_DISTINCT_EXPRS](#), controls the shuffling behavior when a query has both grouping and distinct expressions.

Metadata Performance Improvements

- Incremental Stats

The following enhancements improve Impala stability. The features reduce chances of having catalogd and impalad crash due to be out of memory when using incremental stats.

- Incremental stats are now compressed in memory in catalogd, reducing memory footprint in catalogd.
- Incremental stats are fetched on demand from catalogd by impalad coordinators. This enhancement reduces memory footprint for impalad coordinators and statestored and also reduces network requirements to broadcast metadata.

See [#unique_1253](#) for details.

- Automatic Invalidiation of Metadata



Note: This feature is experimental and not recommended for use in production clusters.

To keep the size of metadata bounded and to reduce the chances of catalogd cache running out of memory, this release introduces an automatic metadata invalidation feature with time-based and memory-based invalidation.

Automatic invalidation of metadata provides more stability with lower chances of running out of memory, but could potentially cause performance risks. The feature is turned off by default.

See [#unique_1254](#) for details.

Compatibility and Usability Enhancements

- Additional separators are supported between date and time in default `TIMESTAMP` format, specifically, the multi-space separator and the 'T' separator. See [TIMESTAMP Data Type](#) for more information on `TIMESTAMP` format.
- New hint placement is supported for `INSERT` statements. See [Optimizer Hints in Impala](#) for detail.
- The `REGEX_ESCAPE()` function was implemented for escaping special characters to treat them literally in string literals.
- `SHOW CREATE VIEW` was implemented with the same functionality as `SHOW CREATE TABLE`.
- The [SHUTDOWN Statement](#) SQL command was implemented for a graceful shutdown of Impala.
- A query can contain multiple `DISTINCT` operators.
- Impala Shell can connect directly to `impalad` when configured with proxy load balancer and Kerberos. See [impala-shell Configuration Options](#) for the new flag that enables the direct connection.
- Impala can read and write data in Azure Data Lake Storage Gen2.

By default, TLS is enabled when ADLS Gen2 is accessed via HTTP and HTTPS.

Apache Kafka

The following are some of the notable new features in this release of Kafka CDH 6.1.0.

Rebase on Apache Kafka 2.0.0

The Kafka version in CDH 6.1.0 is based on Apache Kafka 2.0.0.

Apache Kafka 2.0.0 provides a number of improvements including:

- An improved replication protocol that lessens log divergence between leader and follower during fast leader failover.
- An improved and reworked controller.

Cloudera Enterprise 6 Release Guide

- Support for more partitions per cluster.
- Incremental fetch requests, which improves replication for large partitions.
- A new configuration option for the Kafka consumer to avoid indefinite blocking.

For upstream release notes, see Apache Kafka version [1.0.2](#), [1.1.0](#), [1.1.1](#), and [2.0.0](#) release notes.

New Metrics

A high number of new metrics are introduced for Kafka. The following list is only a summary, for full list of metrics, see [Metrics Reference](#).

Broker Metrics related to the following:

- Controller State
- Global Partition Count
- Global Topic Count
- Kafka Log Cleaner
- Auto Leader Balance Rate and Time
- Controlled Shutdown Rate and Time
- Controller Change Rate and Time
- ISR Change Rate and Time
- Leader and ISR Response Received Rate and Time
- Log Dir Change Rate and Time
- Manual Leader Balance Rate and Time
- Partition Reassignment Rate and Time
- Topic Change Rate and Time
- Topic Deletion Rate and Time

Broker Topic Metrics related to the following:

- Fetch Message Conversion
- Produce Message Conversion
- Incoming Replication rate
- Outgoing Replication Rate
- Total Fetch Requests per Second
- Total Produce Requests per Second

Replica Metrics related to the following:

- Failed ISR Updates
- Offline Replica Count
- Under Min ISR Partition Count

JBOD Support

As of CDH 6.1.0, Cloudera officially supports Kafka clusters with nodes using JBOD configurations.

JBOD support introduces a new command line tool and improves an existing tool:

- A new tool, `kafka-log-dirs`, is added. The tool allows users to query partition assignment information.
- The `kafka-reassign-partitions` tool is expanded with a new functionality that allows users to reassign partitions between log directories. Users can move partitions to a different log directory on the same broker as well as to log directories on other brokers.

Security Improvements

Dependencies for third-party libraries containing security vulnerabilities are updated. Kafka in CDH 6.1.0 is shipped with third-party libraries that do not contain any known security vulnerabilities.

The properties required for enabling remote JMX authentication on Kafka brokers are available in Cloudera Manager. Users are no longer required to carry out setup through a command line interface.

Default Behavior Changes

[KAFKA-7050](#): The default value for `request.timeout.ms` is decreased to 30 seconds. In addition, a new logic is added that makes the `JoinGroup` requests ignore this timeout.

Apache Kudu

The following list describes new features in Apache Kudu for CDH 6.1.0:

- Examples showcasing functionality in C++, Java, and Python, previously hosted in a separate repository have been added. They can be found in the [examples/](#) top-level subdirectory.
- Added `kudu diagnose parse_stacks`, a tool to parse sampled stack traces out of a diagnostics log. See [KUDU-2353](#).
- Added support for `IS NULL` and `IS NOT NULL` predicates to the Kudu Python client. See [KUDU-2399](#).
- Introduced [manual data rebalancer](#) into the Kudu CLI tool. The rebalancer can be used to redistribute table replicas among tablet servers. The rebalancer can be run via `kudu cluster rebalance` sub-command. Using the new tool, it's possible to rebalance Kudu clusters of version 1.4.0 and newer.
- Added `kudu tserver get_flags` and `kudu master get_flags`, two tools that allow superusers to retrieve all the values of command line flags from remote Kudu processes. The `get_flags` tools support filtering the returned flags by tag, and by default will return only flags that were explicitly set.
- Added `kudu tablet unsafe_replace_tablet`, a tool to replace a tablet with a new one. This tool is meant to be used to recover a table when one of its tablets has permanently lost all replicas. The data in the tablet that is replaced is lost, so this tool should only be used as a last resort. See [KUDU-2290](#).

The following list describes optimizations and improvements in Apache Kudu for CDH 6.1.0:

- There is a new metric for each tablet replica tracking the number of election failures since the last successful election attempt and the time since the last heartbeat from the leader. See [KUDU-2287](#).
- Kudu now supports building and running on Ubuntu 18.04 (“Bionic Beaver”). See [KUDU-2427](#).
- Kudu now supports building and running against OpenSSL 1.1. See [KUDU-1889](#).
- Added Kerberos support to the Kudu Flume sink. See [KUDU-2012](#).
- The Kudu Spark connector now supports Spark Streaming DataFrames. See [KUDU-2539](#).
- Added `-tables` filtering argument to `kudu table list`. See [KUDU-2529](#).
- Clients now support setting a limit on the number of returned rows in scans. See [KUDU-16](#).
- Added Pandas support to the Python client. See [KUDU-1276](#).
- Enabled configuration of mutation buffer in the Python client. See [KUDU-2441](#).
- Added a `keepAlive` API call to the `KuduScanner` and `AsyncKuduScanner` in the Java client. This API can be used to keep the scanners alive on the server when processing of messages will take longer than the scanner TTL. See [KUDU-2095](#).
- The Kudu Spark integration now uses the `keepAlive` API when reading data. By default it will call `keepAlive` on a scanner with a period of 15 seconds. This will ensure that Spark jobs with large batch sizes or slow processing times do not fail with scanner not found errors. See [KUDU-2563](#).
- Number of reactor threads in the C++ client is now configurable. See [KUDU-2368](#).
- Added an optimization to avoid bottlenecks on `getpwuid_r()` in libnss during a Raft leader election storm. See [KUDU-2395](#).
- Improved rowset tree pruning making scans with open-ended intervals on primary key. See [KUDU-2566](#).
- The `kudu perf loadgen` toll now supports generating range-partitioned tables. The `-table_num_buckets` configuration is now removed in favor of `-table_num_hash_partitions` and `-table_num_range_partitions`. See [KUDU-1861](#).
- CFile checksum failures will now cause the affected tablet replicas to be failed and re-replicated elsewhere. See [KUDU-2469](#).
- Servers are now able to start up with data directories missing on disk. See [KUDU-2359](#).
- The `kudu perf loadgen` tool now creates tables with a period-separated database name, for example: `default.loadgen_auto_abc123`. This new behavior does not take effect if the `--table` flag is provided. The database of the table can be changed using a new `--auto_database` flag. This change is made in anticipation of an eventual Kudu/HMS integration. See [KUDU-2191](#).

- Introduced FAILED_UNRECOVERABLE replica health status. This is to mark replicas which are not able to catch up with the leader due to GC-collected segments of WAL and other unrecoverable cases like disk failure. With that, the replica management scheme becomes hybrid: the system evicts replicas with FAILED_UNRECOVERABLE health status before adding a replacement if it anticipates that it can commit the transaction, while in other cases it first adds a non-voter replica and removes the failed one only after promoting a newly-added replica to voter role.
- Two additional configuration parameters, `socketReadTimeoutMs` and `ScanRequestTimeout`, have been added to the Spark connector to allow better tuning to avoid scan timeouts under high load.
- The `kudu table` tool now supports two new options to rename tables and columns: `rename_table` and `rename_column`.
- Kudu will now wait for the clock to become synchronized at startup, controlled by the new flag `-ntp_initial_sync_wait_secs`. See [KUDU-2242](#).
- Tablet deletions are now throttled, which will help Kudu clusters remain stable even when many tablets are deleted at once. The number of tablets that a tablet server will delete at once is controlled by the new flag `-num_tablets_to_delete_simultaneously`. See [KUDU-2289](#).
- The `kudu cluster ksck` tool has been significantly enhanced. It now checks master health and consensus status, displays any unsafe or hidden flags set in the cluster, and produces a summary of the Kudu versions running on the master and tablet servers. In addition, it now supports JSON output, both in pretty-printed and compact form. The output format is controlled by the `-ksck_format` flag.

Apache Oozie

There are no notable new features in this release.

Apache Parquet

There are no notable new features in this release.

Apache Pig

There are no notable new features in this release.

Cloudera Search

In CDH 6.1 Cloudera Search is rebased on Apache Solr 7.4, which has added new features since the 7.0 version of Apache Solr used in the CDH 6.0 release.



Important: Log4j2 originally available in Apache Solr 7.4 has been excluded from CDH 6.1. Solr delivered with CDH uses Log4j 1.2.17.

Some features included in Apache Solr 7.4 are not supported in Cloudera Search in CDH 6.1. For more information, see [Cloudera Search Unsupported Features](#) on page 368.

For detailed information on the new features added in Solr 7.4, see the [Apache Solr 7.4 Release Notes](#).

Changes in Configuration Structure

- The top-level `<highlighting>` element in `solrconfig.xml` is now officially deprecated in favor of the equivalent `<searchComponent>` syntax. This element has been out of use in default Solr installations for several releases already.
- Shard and cluster metric reporter configuration now require a `class` attribute.
 - If a reporter configures the `group="shard"` attribute, also configure the `class="org.apache.solr.metrics.reporters.solr.SolrShardReporter"` attribute.
 - If a reporter configures the `group="cluster"` attribute, also configure the `class="org.apache.solr.metrics.reporters.solr.SolrClusterReporter"` attribute.

See [Shard and Cluster Reporters](#) in the Apache Solr Reference Guide for more information.

Changes in Default Configuration Values

- The default value of `autoReplicaFailoverWaitAfterExpiration`, used with the `AutoAddReplicas` feature, has increased to 120 seconds from the previous default of 30 seconds. This affects how soon Solr adds new replicas to replace the replicas on nodes that have either crashed or shutdown.
- The default Solr log file size have been raised to 32MB, and the number of backups is now 10. See [Configuring Logging](#) for more information on how to change this default logging configuration.
- The `eDisMax` parser, by default, doesn't allow subqueries that specify a Solr parser using either local parameters, or the older `_query_` magic field trick.

For example, `{ !prefix f=myfield v=enterp} or _query_ : "{ !prefix f=myfield v=enterp}"` are not supported by default. If you want to allow power-users to do this, set `uf=*` query or some other value that includes `_query_`.

If you need full backwards compatibility for the time being, use `luceneMatchVersion=7.1.0` or an earlier version.

- In the XML query parser (`defType=xmlparser` or `{ !xmlparser ... }`), the resolving of external entities is now disallowed by default.

Changes in Default Behavior

- Configuring `slowQueryThresholdMillis` now logs slow requests to a separate file named `solr_slow_requests.log`. Earlier, slow requests were logged in the `solr.log` file.
- In the leader-follower model of scaling Solr, a follower no longer commits an empty index when a completely new index is detected on the leader during replication. To resume earlier behavior, pass `false` to `skipCommitOnMasterVersionZero` in the follower section of replication handler configuration, or pass it to the `fetchindex` command.
- Collections created without specifying a configset name use a copy of the `_default` configset since Solr 7.0. Before 7.3, the copied configset was named the same as the collection name. From 7.3 onwards, it is named with a new ".AUTOCREATED" suffix to prevent overwriting custom configset names.
- The `rq` parameter used with Learning to Rank rerank query parsing no longer considers the `defType` parameter. See [Running a Rerank Query](#) for more information about this parameter.
- Replicas that are not up-to-date are no longer allowed to become leader. Use the `FORCELEADER` command of the Collections API to allow these replicas become leader.
- The behaviour of the autoscaling system now pauses all triggers from execution between the start of actions and the end of a cool down period. The triggers will resume after the cool down period expires. Earlier, the cool down period was a fixed period that started after actions for a trigger event got completed. During this time all triggers continued to run, but any events were rejected and tried later.
- Starting a query string with local parameters `{ !myparser ... }` is used to switch from one query parser to another. It is intended to be used by Solr system developers, not end users doing searches. To reduce negative side-effects of unintended hackability, Solr now limits the cases when local parameters is parsed to contexts in which the default parser is "lucene" or "func".
 - If `defType=edismax`, `q={ !myparser ... }` doesn't work. In this example, put the desired query parser into the `defType` parameter.
 - If `defType=edismax`, `hl.q= { !myparser ... }` doesn't work. In this example, either put the desired query parser into the `hl.qparser` parameter or set `hl.qparser=lucene`.
- The feature to add replicas automatically if a replica goes down was earlier available only when storing indexes in HDFS. It has been ported to the autoscaling framework, and `AutoAddReplicas` is now available to all users even if their indexes are on local disks.
- Changing the `autoAddReplicas` property from disabled (`false`) to enabled (`true`) using `MODIFYCOLLECTION` API no longer replaces down replicas for the collection immediately. Instead, replicas are only added if a node containing them went down while `AutoAddReplicas` was enabled. The parameters `autoReplicaFailoverBadNodeExpiration` and `autoReplicaFailoverWorkLoopDelay` are no longer used.
- All Stream Evaluators in `solrj.io.eval` have been refactored to have a simpler and more robust structure. This simplifies and condenses the code required to implement a new Evaluator and makes it much easier for evaluators to handle different data types (primitives, objects, arrays, lists, and so forth).

Apache Sentry

The following new features have been added to Apache Sentry in CDH 6.1.0:

Fine Grained Privileges

The CREATE and REFRESH (Impala only) privileges have been introduced to allow users to create databases, tables and functions, and to execute commands that update metadata information on Impala databases and tables.

For more information about the new privileges, see [Sentry Privileges](#).

Object Ownership

Object ownership designates an owner for a database, table, or view in Sentry. The owner of an object has the equivalent of the ALL privilege on the object.

In CDH 6.1.0, object ownership is enabled by default with a new CDH installation. For information about enabling and using object ownership, see [Object Ownership](#).

No Group Name Case Restrictions

Sentry no longer normalizes group name characters to be lowercase. Therefore, operating system group names do not need to be treated as case insensitive. In previous versions, Sentry modified capital letters in operating system group names to be lowercase.

Apache Spark

The following list describes what's new and changed in Apache Spark for CDH 6.1.0, which is based on Apache Spark 2.4 upstream version:

- Support for [Structured Streaming](#) with some limitations. For details, check [Apache Spark Unsupported Features](#) and [Apache Spark Known Issues](#).
- Support for the Microsoft ADLS Gen2 storage service. For more information, see [Accessing Data Stored in Azure Data Lake Store \(ADLS\) through Spark](#).
- Support for Erasure Coding metrics for Hive on Spark. For more information, see [Tools Related to Erasure Coding in Hive](#) in the Hive Component Guide.

Apache Sqoop

The following new features have been added to Apache Sqoop in CDH 6.1.0:

Incremental Import NULL Column Updates into HBase

This feature implements the --hbase-null-incremental-mode option for the `sqoop-import` tool, which allows users to specify how NULL column updates are handled during incremental imports. For more information, see [Importing Data Into HBase Using Sqoop](#).

Automatic Compile Directory Clearing

This feature implements the --delete-compile-dir option for the `sqoop-import` tool, which enables users to automatically delete the generated class and jar files from the disk after the job finishes.

By default all temporary files generated by the `ClassWriter` are left behind on disk in the `/tmp/sqoop-username/compile` directory. Because the table schema can be extracted from these files, Cloudera recommends that you use the --delete-compile-dir option to delete these files.

Parquet Hadoop API Based Implementation for Importing Data Into Parquet Format

Support for the Hadoop API based implementation for importing data into Parquet has been added. This feature implements a new option, --parquet-configurator-implementation, which allows users to specify which implementation used for importing data into Parquet files. For more information, see [Importing Data into Parquet Format Using Sqoop](#).

HiveServer2 Support

This feature implements support for importing data into Hive through HiveServer2.

This feature adds three new options to the `sqoop import` tool:

- `--hs2-url`
- `--hs2-user`
- `--hs2-keytab`

The feature does not introduce any changes to the default behavior of Hive imports. When the user specifies the `--hs2-url` option, commands are sent to HiveServer2 through a JDBC connection. The data itself is not transferred via the JDBC connection. It is written directly to HDFS and moved to the Hive warehouse using the `LOAD DATA INPATH` command just like in the case of the default Hive import.

HiveServer2 provides proper Sentry authorization. As a result, Cloudera recommends importing data into Hive through HiveServer2 instead of the default method. Currently, Sqoop can authenticate to HiveServer2 using Kerberos only.

For more information, see [Importing Data into Hive with Sqoop Through HiveServer2](#).

Support for Import into Amazon S3

Sqoop now supports import from RDBMS into Amazon S3 exploiting the capabilities of the Hadoop-Amazon Web Services integration. For more information about the Hadoop-AWS module, see [Hadoop-AWS module: Integration with Amazon Web Services](#).

Default Precision and Scale in Avro Import

Support to specify a default precision and scale to be used in the avro schema when a table contains numeric data in Oracle, or numeric or decimal data in Postgres, has been added. This feature implements two new properties, `sqoop.avro.logical_types.decimal.default.precision` and `sqoop.avro.logical_types.decimal.scale` to specify the default precision and scale. For more information about Importing Avro in Sqoop, see [Importing Avro Data Files in Sqoop](#).

Behavior Changes

MS SQL Connector Concerning Connection Resets

The recovery logic of the MS-SQL connector proved to be unreliable; therefore, the default behavior was changed from resilient to non-resilient. In other words, the recovery logic is now turned off by default.

The recovery logic can be turned on with the `--resilient` option.

The `--non-resilient` option, which was previously used to turn the recovery logic off, is now ignored.

The resilient operation of the MS-SQL connector requires the split-by column to contain unique values in ascending order only. Otherwise, using the `--resilient` option can lead to duplicate or missing records in the output.

Examples

Importing from a table:

```
sqoop import ... --table custom_table --split-by id -- --resilient
```

Importing via a query:

```
sqoop import ... --query "SELECT ... WHERE $CONDITIONS" --split-by ordered_column -- --resilient
```

Apache Zookeeper

The following list shows what's new and changed in Apache Zookeeper for CDH 6.1.0:

- A new metric is available to let you monitor the size of generated responses to see how to set the client's `jute.maxbuffer` property correctly. For more information, see [ZOOKEEPER-2940](#).
- A new metric is available to track the number of slow fsyncs. For more information, see [ZOOKEEPER-3019](#).
- A tool has been added to recover log and snapshot entries with CRC errors . For more information, see [ZOOKEEPER-2994](#).

Fixed Issues in CDH 6.1.0

CDH 6.1.0 fixes the following issues:

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

Technical Service Bulletin 2019-310 (TSB)

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service, or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper

Releases affected: Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower

Users affected: All

Date/time of detection: June 7, 2018

Severity (Low/Medium/High): 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))

Impact: Remote code execution

CVE: CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings "Enable TLS/SSL for ZooKeeper JMX" and "Enable TLS client authentication for JMX port" on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting "Enable JMX Agent" on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0

Spark Streaming jobs loop if missing Kafka topic

Spark jobs can loop endlessly if the Kafka topic is deleted while a Kafka streaming job (which uses KafkaSource) is in progress.

Cloudera Issue: CDH-57903, CDH-64513

Long-running Spark applications on a secure cluster might fail if driver is restarted

If you submit a long-running app on a secure cluster using the --principal and --keytab options in cluster mode, and a failure causes the driver to restart after 7 days (the default maximum HDFS delegation token lifetime), the new driver fails with an error similar to the following:

```
Exception in thread "main"
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken):
  token <token_info> can't be found in cache
```

Apache Issue: [SPARK-23361](#)

Cloudera Issue: CDH-64865

Kafka May Be Stuck with Under-replicated Partitions after ZooKeeper Session Expires

This problem can occur when your Kafka cluster includes a large number of under-replicated Kafka partitions. One or more broker logs include messages such as the following:

```
[2016-01-17 03:36:00,888] INFO Partition [__samza_checkpoint_event-creation_1,3] on broker 3: Shrinking ISR for partition [__samza_checkpoint_event-creation_1,3] from 6,5 to 5 (kafka.cluster.Partition)
[2016-01-17 03:36:00,891] INFO Partition [__samza_checkpoint_event-creation_1,3] on broker 3: Cached zkVersion [66] not equal to that in zookeeper, skip updating ISR (kafka.cluster.Partition)
```

There will also be an indication of the ZooKeeper session expiring in one or more Kafka broker logs around the same time as the previous errors:

```
INFO zookeeper state changed (Expired) (org.I0Itec.zkclient.ZkClient)
```

The log is typically in `/var/log/kafka` on each host where a Kafka broker is running. The location is set by the property `kafka.log4j.dir` in Cloudera Manager. The log name is `kafka-broker-hostname.log`. In diagnostic bundles, the log is under `logs/hostname-ip-address/`.

Workaround: To move forward after seeing this problem, restart the affected Kafka brokers. You can restart individual brokers from the **Instances** tab in the Kafka service page in Cloudera Manager.



Note: If restarting the brokers does not resolve the problem, you might not have this issue; see [KAFKA-3083 A soft failure in controller may leave a topic partition in an inconsistent state](#). This problem also involves the ZooKeeper session expiring, but will not involve the error message with `Cached zkVersion [XX] not equal to that in zookeeper`.

To reduce the chances of this issue happening again, do what you can to make sure ZooKeeper sessions do not expire:

- Reduce the potential for long garbage collection pauses by brokers:
 - Use a better garbage collection mechanism in the JVM, such as G1GC. You can do this by adding `-XX:+UseG1GC` in the `broker_java_opts`.
 - Increase broker heap size if it is too small (`broker_max_heap_size`). Be careful that you don't choose a heap size that can cause out-of-memory problems given all the services running on the node.
- Increase the ZooKeeper session timeout configuration on brokers (`zookeeper.session.timeout.ms`), to reduce the likelihood that sessions expire.
- Ensure ZooKeeper itself is well resourced and not overwhelmed so it can respond. For example, it is highly recommended to locate the ZooKeeper log directory on its own disk.

Affected Versions: CDK 1.4.x, 2.0.x, 2.1.x, 2.2.x

Fixed Versions:

- **Full Fix:** CDH 6.1.0
- **Partial Fix:** CDH 6.0.0, Kafka implementations with CDH 6.0.0 are less likely to encounter this issue.

Apache Issue: [KAFKA-2729](#)

Cloudera Issue: CDH-42514

Upstream Issues Fixed

The following upstream issues are fixed in CDH 6.1.0:

Apache Accumulo

There are no notable fixed issues in this release.

Apache Avro

Cloudera Enterprise 6 Release Guide

There are no notable fixed issues in this release.

Apache Crunch

There are no notable fixed issues in this release.

Apache Flume

The following issues are fixed in CDH 6.1.0:

- [FLUME-2442](#) - Need an alternative to providing clear text passwords in flume config
- [FLUME-2973](#) - Deadlock in hdfs sink
- [FLUME-2977](#) - Upgrade RAT to 0.12
- [FLUME-3050](#) - add counters for error conditions and expose to monitor URL
- [FLUME-3182](#) - add support for SSL/TLS for syslog (tcp) sources
- [FLUME-3222](#) - Fix for NoSuchFileException thrown when files are being deleted
- [FLUME-3223](#) - Flume HDFS Sink should retry close prior recover lease
- [FLUME-3227](#) - Add Rate Limiter to stresssource
- [FLUME-3239](#) - Do not rename files in SpoolDirectorySource
- [FLUME-3246](#) - Validate flume configuration to prevent larger source batchsize than
- [FLUME-3269](#) - Support JSSE keystore/trustore -D system properties
- [FLUME-3278](#) - Handling -D keystore parameters in Kafka components

Apache Hadoop HDFS

The following issues are fixed in CDH 6.1.0:

- [HADOOP-9214](#) - Enhance the hadoop fs touchz command so that it can now modify atime and mtime.
- [HADOOP-12502](#) - Fixed an issue where setting the replication of a HDFS folder recursively can run out of memory.
- [HADOOP-13649](#) - s3guard: Implement time-based (TTL) expiry for LocalMetadataStore.
- [HADOOP-13761](#) - S3Guard: Implement retries for DDB failures and throttling; translate exceptions.
- [HADOOP-14212](#) - Expose SecurityEnabled boolean field in JMX for other services besides NameNode.
- [HADOOP-14507](#) - Extend per-bucket secret key config with explicit getPassword() on fs.s3a.\$bucket.secret.key.
- [HADOOP-14758](#) - Improve S3GuardTool.prune to handle UnsupportedOperationException.
- [HADOOP-14759](#) - Improve S3GuardTool.prune to prune specific bucket entries.
- [HADOOP-14913](#) - Implement sticky bit for rename() operation in Azure WASB.
- [HADOOP-14935](#) - Fix an issue where Azure POSIX permissions are taking effect in access() method even when authorization is enabled.
- [HADOOP-14965](#) - Change the S3a input stream "normal" fadvise mode to be adaptive.
- [HADOOP-15054](#) - Upgrade hadoop dependency on commons-codec to 1.11.
- [HADOOP-15086](#) - Fix an issue where the NativeAzureFileSystem file rename is not atomic.
- [HADOOP-15121](#) - Fix a NullPointerException when using DecayRpcScheduler.
- [HADOOP-15141](#) - Support IAM Assumed roles in S3A.
- [HADOOP-15143](#) - Fix an NPE due to Invalid KerberosTicket in UGI.
- [HADOOP-15151](#) - Fix an issue where the MapFile.fix creates a wrong index file in case of block-compressed data file.
- [HADOOP-15176](#) - Enhance IAM Assumed Role support in S3A client.
- [HADOOP-15206](#) - Fix an issue where BZip2 drops and duplicates records when input split size is small.
- [HADOOP-15209](#) - Enhance DistCp to eliminate needless deletion of files under already deleted directories.
- [HADOOP-15212](#) - Add independent secret manager method for logging expired tokens.
- [HADOOP-15215](#) - Enhance s3guard set-capacity command to fail on read/write of 0.
- [HADOOP-15217](#) - Enhance FsURLConnection to handle paths with spaces.
- [HADOOP-15250](#) - Fix an issue where a multiHomed server network cluster Network IPC Client binds the wrong address.
- [HADOOP-15267](#) - S3A multipart upload fails when SSE-C encryption is enabled.

- [HADOOP-15391](#) - Add missing CSS file in hadoop-aws, hadoop-aliyun, hadoop-azure and hadoop-azure-datalake modules.
 - [HADOOP-15423](#) - Merge fileCache and dirCache into one single cache in LocalMetadataStore
 - [HADOOP-15441](#) - Log kms url and token service at debug level.
 - [HADOOP-15446](#) - WASB: PageBlobInputStream.skip breaks HBASE replication.
 - [HADOOP-15449](#) - Increase default timeout of ZK session to avoid frequent NameNode failover.
 - [HADOOP-15469](#) - Fix an issue where the S3A directory committer commit job fails if _temporary directory created under destination.
 - [HADOOP-15478](#) - Fix an issue with WASB that caused an hflush() and hsync() regression.
 - [HADOOP-15541](#) - Fix an issue where the AWS SDK can mistake stream timeouts for EOF and throw SdkClientExceptions.
 - [HADOOP-15598](#) - Fix an issue where the DataChecksum calculate checksum experiences contention on hashtable synchronization.
 - [HADOOP-15612](#) - Improve exception when tfile fails to load LzoCodec.
 - [HADOOP-15633](#) - Fix an issue where fs.TrashPolicyDefault cannot create trash directory.
 - [HADOOP-15679](#) - Enhance ShutdownHookManager shutdown time to be configurable & extended.
 - [HADOOP-15684](#) - Fix an issue where triggerActiveLogRoll stuck on dead NameNode when ConnectTimeoutException happens.
 - [HADOOP-15719](#) - Fail-fast when using OAuth over http.
 - [HADOOP-15850](#) - Enhance CopyCommitter#concatFileChunks to check that the blocks per chunk is not 0.
 - [HADOOP-15861](#) - Move DelegationTokenIssuer to the correct path.
 - [HDFS-9049](#) - Make Datanode Netty reverse proxy port configurable.
 - [HDFS-10183](#) - Prevent race condition during class initialization.
 - [HDFS-11701](#) - Fix an issue where NPE from Unresolved Host causes permanent DFSInputStream failures.
 - [HDFS-11719](#) - Enhance Arrays.fill() wrong index in BlockSender.readChecksum() exception handling.
 - [HDFS-11900](#) - Fix an issue where hedged reads thread pool creation not synchronized.
 - [HDFS-12070](#) - Fix an issue where failed block recovery leaves files open indefinitely and at risk for data loss.
 - [HDFS-12574](#) - Add CryptoInputStream to WebHdfsFileSystem read call.
 - [HDFS-12907](#) - Allow read-only access to reserved raw for non-superusers.
 - [HDFS-12978](#) - Add fine-grained locking while consuming journal stream.
 - [HDFS-13027](#) - Handle possible NPEs due to deleted blocks in race condition.
- [HDFS-13048](#) - Fix an issue where the LowRedundancyReplicatedBlocks metric can be negative.
- [HDFS-13052](#) - Add support for snapshot diff with WebHDFS.
 - [HDFS-13060](#) - Add a BlacklistBasedTrustedChannelResolver for TrustedChannelResolver.
- [HDFS-13081](#) - Allow SASL and privileged HTTP with Datanode#checkSecureConfig.
- [HDFS-13087](#) - Make snapshotted encryption zone information immutable.
 - [HDFS-13145](#) - Fix an issue where an SBN crash occurs when transitioning to ANN with in-progress edit tailing enabled.
- [HDFS-13225](#) - Fix an issue where StripeReader#checkMissingBlocks()'s IOException info is incomplete.
- [HDFS-13280](#) - Fix NPE in get snaphottable directory list call.
 - [HDFS-13330](#) - Fix an issue where ShortCircuitCache#fetchOrCreate never retries.
- [HDFS-13448](#) - Ignore locality for First Block Replica.
- [HDFS-13493](#) - Reduce the HttpServer2 thread count on DataNodes.
 - [HDFS-13641](#) - Add metrics for edit log tailing.
- [HDFS-13658](#) - Expose HighestPriorityLowRedundancy blocks statistics.
- [HDFS-13668](#) - FSPermissionChecker may throw rrayIndexOutOfBoundsException when checking inode permission.
 - [HDFS-13686](#) - Add overall metrics for FSNamesystemLock.
- [HDFS-13728](#) - Fix an issue where the Disk Balancer fails if volume usage is greater than capacity.

- [HDFS-13731](#) - Fix an issue where ReencryptionUpdater fails with ConcurrentModificationException during processCheckpoints.
- [HDFS-13738](#) - Fix an issue where fsck -list-corruptfileblocks encounters an infinite loop if the user is not privileged.
- [HDFS-13758](#) - Enhance DatanodeManager to throw exception if it has BlockRecoveryCommand but the block is not under construction.
- [HDFS-13820](#) - Add an ability to disable CacheReplicationMonitor.
- [HDFS-13830](#) - Add support for getting snashpottable directory list.
- [HDFS-13831](#) - Make block increment deletion number configurable.
- [HDFS-13833](#) - Improve BlockPlacementPolicyDefault's consider load logic.
- [HDFS-13838](#) - Fix an issue where WebHdfsFileSystem.getFileStatus() does not return correct "snapshot enabled" status.
- [HDFS-13846](#) - Fix an issue where safe blocks counter is not decremented correctly if the block is striped.
- [HDFS-13868](#) - Fix an NPE with the GETSNAPSHOTDIFF API when the parameter "snapshotname" is given but "oldsnapshotname" is not.
- [HDFS-13876](#) - Implement ALLOWSNAPSHOT/DISALLOWSNAPSHOT for HttpFS.
- [HDFS-13877](#) - Implement GETSNAPSHOTDIFF for HttpFS.
- [HDFS-13878](#) - Implement GETSNAPSHOTTABLEDIRECTORYLIST for HttpFS.
- [HDFS-13882](#) - Set a maximum delay for retrying locateFollowingBlock.
- [HDFS-13885](#) - Add debug logs in dfsclient around decrypting EDEK.
- [HDFS-13886](#) - Fix an issue where HttpFSFileSystem.getFileStatus() doesn't return "snapshot enabled" bit.
- [HDFS-14009](#) - Fix an issue where FileStatus#setSnapshotEnabledFlag throws InvocationTargetException when attribute set is emptySet.

MapReduce 2

The following issues are fixed in CDH 6.1.0:

- [MAPREDUCE-6861](#) - Add metrics tags for ShuffleClientMetrics.
- [MAPREDUCE-7150](#) - Optimize collections used by MR JHS to reduce its memory.

YARN

The following issues are fixed in CDH 6.1.0:

- [YARN-7159](#) - Normalize unit of resource objects in ResourceManager to avoid unit conversion in critical path.
- [YARN-7237](#) - Cleanup usages of ResourceProfiles.
- [YARN-7728](#) - Expose container preemptions related information in Capacity Scheduler queue metrics.
- [YARN-7738](#) - CapacityScheduler: Support refresh maximum allocation for multiple resource types .
- [YARN-7948](#) - Enable fair scheduler to refresh maximum allocation for multiple resource types.
- [YARN-8338](#) - Fixed an issue where TimelineService V1.5 does not come up after HADOOP-15406.
- [YARN-8566](#) - Add diagnostic message for unschedulable containers .
- [YARN-8842](#) - Expose metrics for custom resource types in QueueMetrics.
- [YARN-8990](#) - Fix fair scheduler race condition in app submit and queue cleanup.

Apache HBase

The following issues are fixed in CDH 6.1.0:

- [HBASE-18451](#) - PeriodicMemstoreFlusher should inspect the queue before adding a delayed flush request, fix logging
- [HBASE-18549](#) - Add metrics for failed replication queue recovery
- [HBASE-19418](#) - configurable range of delay in PeriodicMemstoreFlusher
- [HBASE-20193](#) - Basic Replication Web UI - Regionserver
- [HBASE-20375](#) - Remove use of getCurrentUserCredentials in hbase-spark module
- [HBASE-20469](#) - Directory used for sidelining old recovered edits files should be made configurable
- [HBASE-20732](#) - Shutdown scan pool when master is stopped
- [HBASE-20734](#) - Colocate recovered edits directory with hbase.wal.dir

- [HBASE-20741](#) - Split of a region with replicas creates all daughter regions
- [HBASE-20792](#) - info:servername and info:sn inconsistent for OPEN region
- [HBASE-20808](#) - (Addendum) Remove duplicate calls for cancelling of chores
- [HBASE-20846](#) - Restore procedure locks when master restarts
- [HBASE-20857](#) - balancer status tag in jmx metrics
- [HBASE-20865](#) - CreateTableProcedure is stuck in retry loop in CREATE_TABLE_WRITE_FS_LAYOUT state
- [HBASE-20892](#) - [UI] Start / End keys are empty on table.jsp
- [HBASE-20942](#) - Revert "Fix Array Index Out Of Bounds Exception for RpcServer TRACE logging"
- [HBASE-20965](#) - Separate region server report requests to new handlers
- [HBASE-20985](#) - add two attributes when we do normalization
- [HBASE-20986](#) - Separate the config of block size when we do log splitting and write Hlog
- [HBASE-21001](#) - ReplicationObserver fails to load in HBase 2.0.0
- [HBASE-21023](#) - Added bypassProcedure() API to HbckService
- [HBASE-21032](#) - ScanResponses contain only one cell each
- [HBASE-21055](#) - NullPointerException when balanceOverall() but server balance info is null
- [HBASE-21072](#) - Addendum do not write lock file when running TestHBaseFsckReplication
- [HBASE-21073](#) - Redo concept of maintenance mode
- [HBASE-21095](#) - The timeout retry logic for several procedures are broken after master restarts
- [HBASE-21125](#) - 'HBASE-20942 Improve RpcServer TRACE logging' to branch-2.1
- [HBASE-21126](#) - "Add ability for HBase Canary to ignore a configurable number of ZooKeeper down nodes" to branch-2.1
- [HBASE-21127](#) - TableRecordReader need to handle cursor result too
- [HBASE-21132](#) - return wrong result in rest multiget
- [HBASE-21144](#) - AssignmentManager.waitForAssignment is not stable
- [HBASE-21155](#) - Save on a few log strings and some churn in wal splitter by skipping out early if no logs in dir
- [HBASE-21156](#) - [hbck2] Queue an assign of hbase:meta and bulk assign/unassign
- [HBASE-21158](#) - Empty qualifier cell is always returned when using QualifierFilter
- [HBASE-21164](#) - reportForDuty should do backoff rather than retry
- [HBASE-21171](#) - [amv2] Tool to parse a directory of MasterProcWALs standalone
- [HBASE-21172](#) - Reimplement the retry backoff logic for ReopenTableRegionsProcedure
- [HBASE-21174](#) - [REST] Failed to parse empty qualifier in TableResource#getScanResource
- [HBASE-21179](#) - Fix the number of actions in responseTooSlow log
- [HBASE-21181](#) - Use the same filesystem for wal archive directory and wal directory
- [HBASE-21182](#) - Failed to execute start-hbase.sh
- [HBASE-21185](#) - WALPrettyPrinter: Additional useful info to be printed by wal printer tool, for debugability purposes
- [HBASE-21190](#) - Log files and count of entries in each as we load from the MasterProcWAL store
- [HBASE-21191](#) - Add a holding-pattern if no assign for meta or namespace (Can happen if masterprocwals have been cleared).
- [HBASE-21196](#) - HTableMultiplexer clears the meta cache after every put operation
- [HBASE-21200](#) - Memstore flush doesn't finish because of seekToPreviousRow() in memstore scanner.
- [HBASE-21204](#) - NPE when scan raw DELETE_FAMILY_VERSION and codec is not set
- [HBASE-21206](#) - Scan with batch size may return incomplete cells
- [HBASE-21207](#) - Add client side sorting functionality in master web UI for table and region server details
- [HBASE-21208](#) - Bytes#toShort doesn't work without unsafe
- [HBASE-21212](#) - Wrong flush time when update flush metric
- [HBASE-21214](#) - [hbck2] setTableState just sets hbase:meta state, not in-memory state
- [HBASE-21223](#) - [amv2] Remove abort_procedure from shell
- [HBASE-21228](#) - Memory leak since AbstractFSWAL caches Thread object and never clean later
- [HBASE-21232](#) - Show table state in Tables view on Master home page
- [HBASE-21233](#) - Allow the procedure implementation to skip persistence of the state after a execution

- [HBASE-21242](#) - Revert "[amv2] Miscellaneous minor log and assign procedure create improvements; ADDENDUM Fix TestHRegionInfo"
- [HBASE-21248](#) - Implement exponential backoff when retrying for ModifyPeerProcedure
- [HBASE-21249](#) - Add jitter for ProcedureUtil.getBackoffTimeMs
- [HBASE-21250](#) - Addendum remove unused modification in hbase-server module
- [HBASE-21250](#) - Refactor WALProcedureStore and add more comments for better understanding the implementation
- [HBASE-21254](#) - Need to find a way to limit the number of proc wal files
- [HBASE-21259](#) - [amv2] Revived deadservers; recreated serverstatenode
- [HBASE-21260](#) - The whole balancer plans might be aborted if there are more than one plans to move a same region
- [HBASE-21263](#) - Mention compression algorithm along with other storefile details
- [HBASE-21266](#) - Not running balancer because processing dead regionservers, but empty dead rs list
- [HBASE-21280](#) - Add anchors for each heading in UI
- [HBASE-21287](#) - Allow configuring test master initialization wait time.
- [HBASE-21288](#) - HostingServer in UnassignProcedure is not accurate
- [HBASE-21292](#) - IdLock.getLockEntry() may hang if interrupted
- [HBASE-21299](#) - List counts of actual region states in master UI tables section
- [HBASE-21303](#) - [shell] clear_deadservers with no args fails
- [HBASE-21323](#) - Revert "Should not skip force updating for a sub procedure even if"
- [HBASE-21425](#) - 2.1.1 fails to start over 1.x data; namespace not assigned

Apache Hive

The following issues are fixed in CDH 6.1.0:

Code Changes Might Be Required

The following fixes might require code changes for the CDH 6.1.0 release of Apache Hive:

- [HIVE-14388](#) - Add number of rows inserted message after insert command in Beeline
- [HIVE-17799](#) - Add Ellipsis For Truncated Query In Hive Lock
- [HIVE-19344](#) - Change default value of msck.repair.batch.size

Code Changes Should Not Be Required

The following fixes should not require code changes, but they contain improvements that might enhance your deployment:

- [HIVE-6980](#) - Drop table by using direct SQL
- [HIVE-10296](#) - Cast exception observed when hive runs a multi-join query on metastore (postgres), since postgres pushes the filter into the join, and ignores the condition before applying cast
- [HIVE-13900](#) - HiveStatement.executeAsync() may not work properly when hive.server2.async.exec.async.compile is turned on
- [HIVE-14162](#) - Allow disabling of a long-running job on Hive On Spark On YARN
- [HIVE-14560](#) - Support exchange partition between s3 and HDFS tables
- [HIVE-14690](#) - Query fail when hive.exec.parallel=true, with conflicting session dir
- [HIVE-14984](#) - Hive-WebUI access results in Request is a replay (34) attack
- [HIVE-15104](#) - Hive on Spark generate more shuffle data than hive on mr
- [HIVE-15180](#) - Extend JSONMessageFactory to store additional information about metadata objects on different table events
- [HIVE-15250](#) - Reuse partitions info generated in MoveTask to its subscribers (StatsTask)
- [HIVE-15712](#) - New HiveConf in SQLOperation.getSerDe() impacts CPU on Hiveserver2
- [HIVE-15995](#) - Syncing metastore table with serde schema
- [HIVE-16071](#) - Hos RPCServer misuses the timeout in its RPC handshake
- [HIVE-16143](#) - Improve msck repair batching
- [HIVE-16172](#) - Switch to a fairness lock to synchronize HS2 thrift client

- [HIVE-16219](#) - Metastore notification_log contains serialized message with non-functional fields
- [HIVE-16285](#) - Servlet for dynamically configuring log levels
- [HIVE-16346](#) - inheritPerms should be conditional based on the target filesystem
- [HIVE-16348](#) - HoS query is canceled but error message shows RPC is closed
- [HIVE-16431](#) - Support Parquet StatsNoJobTask for Spark & Tez engine
- [HIVE-16607](#) - ColumnStatsAutoGatherContext regenerates HiveConf.HIVEQUERYID
- [HIVE-16664](#) - Add join related Hive blobstore tests
- [HIVE-16736](#) - General Improvements to BufferedRows
- [HIVE-17300](#) - WebUI query plan graphs
- [HIVE-17401](#) - Hive session idle timeout doesn't function properly
- [HIVE-17747](#) - HMS DropTableMessage should include the full table object
- [HIVE-18031](#) - Support replication for Alter Database operation
- [HIVE-18118](#) - Explain Extended should indicate if a file being read is an EC file
- [HIVE-18652](#) - Print Spark metrics on console
- [HIVE-18690](#) - Integrate with Spark OutputMetrics
- [HIVE-18696](#) - The partition folders might not get cleaned up properly in the HiveMetaStore.add_partitions_core method if an exception occurs
- [HIVE-18705](#) - Improve HiveMetaStoreClient.dropDatabase
- [HIVE-18743](#) - CREATE TABLE on S3 data can be extremely slow.DO_NOT_UPDATE_STATS workaround is buggy
- [HIVE-18766](#) - Race condition during shutdown of RemoteDriver, error messages aren't always sent
- [HIVE-18778](#) - Needs to capture input/output entities in explain
- [HIVE-18906](#) - Lower Logging for "Using direct SQL".
- [HIVE-18916](#) - SparkClientImpl doesn't error out if spark-submit fails.
- [HIVE-19008](#) - Improve Spark session id logging
- [HIVE-19053](#) - RemoteSparkJobStatus#getSparkJobInfo treats all exceptions as timeout errors
- [HIVE-19079](#) - Add extended query string to Spark job description
- [HIVE-19370](#) - Issue: ADD Months function on timestamp datatype fields in Hive
- [HIVE-19371](#) - Add table ownerType to HMS thrift API
- [HIVE-19372](#) - Add table ownerType to JDO/SQL and ObjectStore
- [HIVE-19374](#) - Parse and process ALTER TABLE SET OWNER command syntax
- [HIVE-19477](#) - Hiveserver2 in HTTP mode not emitting metric default.General.open_connections
- [HIVE-19486](#) - Discrepancy in HikariCP config naming
- [HIVE-19508](#) - SparkJobMonitor getReport doesn't print stage progress in order
- [HIVE-19525](#) - Spark task logs print PLAN PATH excessive number of times
- [HIVE-19559](#) - SparkClientImpl shouldn't name redirector thread RemoteDriver
- [HIVE-19718](#) - Adding partitions in bulk also fetches table for each partition
- [HIVE-19733](#) - RemoteSparkJobStatus#getSparkStageProgress inefficient implementation
- [HIVE-19766](#) - Show the number of rows inserted when execution engine is Spark
- [HIVE-19783](#) - Retrieve only locations in HiveMetaStore.dropPartitionsAndGetLocations
- [HIVE-19786](#) - RpcServer cancelTask log message is incorrect
- [HIVE-19787](#) - Log message when spark-submit has completed
- [HIVE-19814](#) - RPC Server port is always random for spark
- [HIVE-19899](#) - Support stored as JsonFile
- [HIVE-19937](#) - Intern fields in MapWork on deserialization
- [HIVE-19942](#) - Hive Notification: All events for indexes should have table name
- [HIVE-19986](#) - Add logging of runtime statistics indicating when Hdfs Erasure Coding is used by MR
- [HIVE-20032](#) - Don't serialize hashCode for repartitionAndSortWithinPartitions
- [HIVE-20056](#) - SparkPartitionPruner shouldn't be triggered by Spark tasks
- [HIVE-20098](#) - Statistics: NPE when getting Date column partition statistics
- [HIVE-20212](#) - Hiveserver2 in http mode emitting metric default.General.open_connections incorrectly

- [HIVE-20374](#) - Write Hive version information to Parquet footer
- [HIVE-20466](#) - Improve org.apache.hadoop.hive.ql.exec.FunctionTask Experience
- [HIVE-20505](#) - upgrade org.openjdk.jmh:jmh-core to 1.21
- [HIVE-20544](#) - TOpenSessionReq logs password and username
- [HIVE-20545](#) - Exclude parameters that can have potentially large size from HMS notification message JSON
- [HIVE-20601](#) - EnvironmentContext null in ALTER_PARTITION event in DbNotificationListener
- [HIVE-20603](#) - "Wrong FS" error when inserting to partition after changing table location filesystem
- [HIVE-20678](#) - HiveHBaseTableOutputFormat should implement HiveOutputFormat to ensure compatibility
- [HIVE-20695](#) - HoS Query fails with hive.exec.parallel=true.
- [HIVE-20711](#) - Race Condition when Multi-Threading in SessionState.createRoothDFSDir
- [HIVE-20742](#) - SparkSessionManagerImpl maintenance thread only cleans up session once

Hue

The following issues are fixed in CDH 6.1.0:

- [HUE-7407](#) - [useradmin] Added superuser group priv to useradmin
- [HUE-7698](#) - [oozie] Added warning when there is a space in the shell action
- [HUE-7698](#) - [oozie] Files of a Shell document action in a workflow are not being generated in the XML
- [HUE-7860](#) - [core] Update greenlet from 0.4.12 to 0.4.15
- [HUE-7860](#) - [core] Add monotonic 1.5
- [HUE-7860](#) - [core] Update Gunicorn from 19.7.1 to 19.9.0
- [HUE-7860](#) - [core] Update eventlet from 0.21.0 to 0.24.1
- [HUE-7860](#) - [core] Add dnspython 1.15.0
- [HUE-8139](#) - [core] Fix django-debug-toolbar 1.9.1 to work with django_debug_panel
- [HUE-8140](#) - [editor] Automatically continue execution after DDL statements in batch mode
- [HUE-8330](#) - [cluster] Keep only external cluster configs in [[clusters]]
- [HUE-8330](#) - [core] API should not check for remote cloud clusters if they are not configured
- [HUE-8339](#) - [impala] Fix typo in smart pooling ini configuration
- [HUE-8391](#) - [importer] Improve Create table from File UX when loading data from parent directory not readable by hive/impala
- [HUE-8488](#) - [fb] Disable drag&drop when show_upload_button=false
- [HUE-8507](#) - [editor] Add types to sqlalchemy results.
- [HUE-8507](#) - [editor] SQL alchemy result set column headers are missing.
- [HUE-8509](#) - [oozie] Schedule repetitive remote jobs
- [HUE-8509](#) - [oozie] Support sending a SQL query to a remote cluster
- [HUE-8509](#) - [jb] Clean-up of the listing of remote jobs
- [HUE-8509](#) - [oozie] Properly set the capture output flag of shell document action
- [HUE-8509](#) - [oozie] Remote job action
- [HUE-8509](#) - [kafka] Do not break left panel
- [HUE-8514](#) - [core] Log metrics when calling is_alive
- [HUE-8516](#) - [cluster] List more namespaces and filter out invalid ones
- [HUE-8518](#) - [editor] Fix sample Kudu
- [HUE-8519](#) - [jb] Impala API can now directly return json
- [HUE-8521](#) - [auth] Protect against empty LDAP login username
- [HUE-8522](#) - [jb] Make paused tasks more obvious. Add queued state to Impala
- [HUE-8523](#) - [jb] Display Impala backends & instances
- [HUE-8524](#) - [impala] Provide the root cause of INVALIDATE METADATA failures
- [HUE-8527](#) - [editor] Fix concatenation type exception in namespace call
- [HUE-8528](#) - [frontend] Temporarily disable namespace caching
- [HUE-8529](#) - [frontend] Create a context selector component
- [HUE-8531](#) - [sqoop] Properly name the table import job

- [HUE-8532](#) - [core] Fix database migration test.
- [HUE-8533](#) - [importer] Properly displayed failed import progress bar as red and not orange
- [HUE-8534](#) - [jb] Django url name does not exist and breaks page
- [HUE-8535](#) - [sqoop] Use the proper engine name and not the connection nice name as jdbc prefix
- [HUE-8536](#) - [sqoop] Include hive-site.xml automatically when importing data to hive
- [HUE-8537](#) - [sqoop] List the proper column type when importing to a hive table
- [HUE-8538](#) - [sqoop] Allow table preview from manual input not JDBC
- [HUE-8538](#) - [importer] Automatically fill-up the db driver list when selecting sqoop
- [HUE-8539](#) - [importer] Clean-up configuration and turn sqoop and solr imports to on by default
- [HUE-8540](#) - [sqoop] Add ability to set default jdbc driver path for any sqoop job
- [HUE-8541](#) - [oozie] Workflow rerun does not restart polling for job status
- [HUE-8542](#) - [frontend] Add a custom left nav for multi cluster mode
- [HUE-8542](#) - [frontend] Polish cloud cluster and require multi cluster mode to be on
- [HUE-8544](#) - [importer] Support sending file data into a kafka topic
- [HUE-8545](#) - [search] Fix filtering in the index selection dropdown
- [HUE-8546](#) - [assist] Limit assist refresh to the active namespace for DDL statement executions
- [HUE-8546](#) - [assist] Make sure the assist gets refreshed after multiple DDL statement executions
- [HUE-8547](#) - [jb] Fix navigation from create schedule to view schedule.
- [HUE-8547](#) - [jb] Fix refresh on coordinator page.
- [HUE-8548](#) - [jb] Fix invalid date in workflow task
- [HUE-8549](#) - [autocomplete] Improve CTE alias suggestions when there's a trailing ";"
- [HUE-8550](#) - [jb] Use the context selector component in the job browser
- [HUE-8550](#) - [frontend] Make last selected compute and namespace sticky
- [HUE-8550](#) - [jb] Default to the last selected type of compute in the job browser
- [HUE-8550](#) - [jb] Refresh job browser tabs on compute selection
- [HUE-8551](#) - [importer] Support setting basic Flume configs
- [HUE-8553](#) - [kafka] Link create topic API to the UI
- [HUE-8553](#) - [kafka] Add a workaround API for creating a topic
- [HUE-8554](#) - [indexer] Protect against empty sample data that can be null
- [HUE-8554](#) - [importer] Support latest Spark version 2 natively
- [HUE-8554](#) - [manager] Adding a check if service is installed API
- [HUE-8554](#) - [cluster] Create data warehouse cluster skeleton
- [HUE-8554](#) - [core] Support dist Spark installed when running envelope via shell
- [HUE-8554](#) - [cluster] Avoid double escapating of data warehouse results
- [HUE-8554](#) - [cluster] Rename analytic cluster API command to dataware
- [HUE-8555](#) - [cluster] Do not submit remote coordinator jobs by default
- [HUE-8555](#) - [jb] Refactor job browser preview to support multi cluster
- [HUE-8555](#) - [jb] Support killing data warehouse cluster
- [HUE-8555](#) - [jb] Sort clusters with the most recent first
- [HUE-8555](#) - [jb] List data warehouse clusters
- [HUE-8555](#) - [jb] Auto select the first cluster if possible at init
- [HUE-8556](#) - [fb] Overuse of trash folder checking
- [HUE-8557](#) - [sqoop] DB name and table names variables were already present
- [HUE-8557](#) - [sqoop] Offer to rename the table or selected a different existing Hive database
- [HUE-8558](#) - [jb] Add tracking URL to Spark Jobs and remove url and killUrl
- [HUE-8559](#) - [jb] Hue shows incorrect color for failed oozie jobs
- [HUE-8560](#) - [tb] Make sure the default DB is opened by default in the Table Browser
- [HUE-8560](#) - [tb] Stick to the same view when switching namespaces in the Table Browser
- [HUE-8561](#) - [editor] Don't show databases for spark editor
- [HUE-8562](#) - [frontend] Make sure the context popover is shown above the jobs panel

- [HUE-8564](#) - [useradmin] Fix last activity update for notebook/api/check_status
- [HUE-8564](#) - [useradmin] Fix last activity update for jobbrowser/api/jobs requests
- [HUE-8565](#) - [fb] Parent directory should not be selectable
- [HUE-8565](#) - [fb] Current directory should not be deletable.
- [HUE-8566](#) - [useradmin] Update message for duplicate user creation.
- [HUE-8567](#) - [jb] Fix id max length in mini jb
- [HUE-8568](#) - [jb] Prevent mini jb actions from taking content width
- [HUE-8568](#) - [jb] Activate smart file links from the logs by also checking for prefixes
- [HUE-8570](#) - [assist] Extract a separate column sample component
- [HUE-8570](#) - [frontend] Right align the Hue dropdown when rendered outside the window
- [HUE-8570](#) - [assist] Add distinct as an option for column samples in the context popover
- [HUE-8570](#) - [editor] Enable click to insert from sample popover to SQL variables
- [HUE-8570](#) - [assist] Add inline autocomplete for column samples
- [HUE-8570](#) - [editor] Enable optional operation on the sample API endpoints
- [HUE-8570](#) - [assist] Limit context popover sample operations to Impala and Hive
- [HUE-8570](#) - [assist] Add min and max to column sample popover
- [HUE-8571](#) - [sentry] navigator_api ERROR for
PRIVILEGE_HIERARCHY[hierarchy[server][SENTRY_PRIVILEGE_KEY]['action']]
- [HUE-8572](#) - [cluster] Bubble up authentication errors on remote clusters
- [HUE-8572](#) - [tb] Fix JS exception when clearing table browser selection via pubsub
- [HUE-8572](#) - [tb] Add compute and namespace to DROP table endpoint
- [HUE-8572](#) - [tb] Fix log overflow in history panel
- [HUE-8573](#) - [sqoop] Out of the box import of a MySQL table
- [HUE-8573](#) - [sqoop] Avoid unrelated casting error when testing the connection
- [HUE-8574](#) - [importer] Adding Flume flows
- [HUE-8574](#) - [flume] Support updating Flume agent config
- [HUE-8574](#) - [importer] Nav Kafka stream import to Solr and Kudu part 1
- [HUE-8574](#) - [importer] Allow audit logs to be sent to Solr
- [HUE-8574](#) - [importer] Setup automatically a Flume grapping Hue HTTPD logs and put into the sample collection
- [HUE-8574](#) - [importer] Feature flag for showing the Field Editor
- [HUE-8574](#) - [cluster] Auto scaling data warehouse cluster API skeleton
- [HUE-8574](#) - [importer] Button caret to call for getting the job config
- [HUE-8575](#) - [importer] Add external multi table support
- [HUE-8575](#) - [importer] Fix file to table import.
- [HUE-8576](#) - [editor] Add backticked suggestion to the syntax checked for reserved keywords
- [HUE-8577](#) - [autocomplete] Add all currently reserved keywords for Impala
- [HUE-8577](#) - [editor] Rebuild Ace with updated dependencies
- [HUE-8577](#) - [autocomplete] Add support for Impala SHOW GRANT ROLE/USER statements
- [HUE-8577](#) - [autocomplete] Add support for Impala ALTER TABLE/VIEW SET OWNER
- [HUE-8577](#) - [autocomplete] Add Impala METHOD to reserved keywords
- [HUE-8577](#) - [autocomplete] Make previously non-reserved keywords reserved for Impala
- [HUE-8577](#) - [autocomplete] Fix issue where the statement type location is added twice
- [HUE-8578](#) - [importer] Auto select id column if present in Kudu tables
- [HUE-8578](#) - [importer] Implement Flume output
- [HUE-8578](#) - [importer] Get basic Flume ingest step integrated
- [HUE-8578](#) - [manager] Restrict API calls to admin
- [HUE-8579](#) - [core] Blacklisting certain apps like filebrowser and oozie can fail
- [HUE-8580](#) - [editor] Fix jdbc assist.
- [HUE-8580](#) - [importer] Improve usability of table import
- [HUE-8580](#) - [importer] Fix RDBMS support for scoop configured import.

- [HUE-8581](#) - [importer] Fix timing related JS exceptions
- [HUE-8581](#) - [importer] Improve query type selection layout for the field editor
- [HUE-8581](#) - [importer] Fix JS error on target namespace selection and improve layout for table import
- [HUE-8581](#) - [importer] Improve the stream import form layout
- [HUE-8581](#) - [importer] Allow typed paths in the hivechooser binding
- [HUE-8581](#) - [importer] Fix JS error for field query editor in importer
- [HUE-8582](#) - [jb] Make back button from editing a file more obvious
- [HUE-8583](#) - [fb] Surface too many buckets error
- [HUE-8588](#) - [core] Fix PAM backend has conflict with timer metrics
- [HUE-8589](#) - [core] Split cluster listing to its own API
- [HUE-8589](#) - [jb] Switch from compute to the cluster API endpoint in the job browser
- [HUE-8591](#) - [cluster] Integration skeleton for Data Warehouse v2 API
- [HUE-8591](#) - [impala] Properly pickup the selected compute cluster
- [HUE-8591](#) - [cluster] Remove extra debug info
- [HUE-8591](#) - [cluster] Step of logic simplification of the multi cluster configuration
- [HUE-8591](#) - [cluster] Display impalad hostname
- [HUE-8591](#) - [impala] Properly point to the selected cluster hostname
- [HUE-8591](#) - [cluster] Protect against override of cluster name
- [HUE-8591](#) - [core] Showing up S3 browser by default in cloud mode
- [HUE-8591](#) - [cluster] Move port to 21050
- [HUE-8591](#) - [cluster] Safeguard against localhost
- [HUE-8591](#) - [cluster] Properly use the correct cluster hostname in the editor
- [HUE-8591](#) - [cluster] Add hostname check in the cluster hostname log trace
- [HUE-8591](#) - [cluster] Split cluster template between static and dynamic clusters
- [HUE-8591](#) - [cluster] Clear the compute cache on namespace refresh from left assist
- [HUE-8591](#) - [cluster] Avoid failing when cluster is None
- [HUE-8591](#) - [cluster] Wire in API for listing and creating k8 clusters
- [HUE-8591](#) - [cluster] Plug in the list of clusters
- [HUE-8591](#) - [cluster] Adding cluster resize capabilities on the cluster page
- [HUE-8591](#) - [cluster] Add Thrift client used for the specific query server
- [HUE-8591](#) - [cluster] Refresh the context selector when namespaces are refreshed
- [HUE-8591](#) - [cluster] Hook in remote Impala coordinator URL of selected cluster
- [HUE-8591](#) - [cluster] Use default port if ont in a selected remote cluster
- [HUE-8591](#) - [cluster] Add impalad link to cluster page
- [HUE-8591](#) - [cluster] Add logic to get the corresponding Impalad name
- [HUE-8591](#) - [cluster] Add some progress bar color and effect on cluster resize
- [HUE-8591](#) - [cluster] Add proper cluster page
- [HUE-8591](#) - [cluster] Use name as clusterName throughout the calls
- [HUE-8591](#) - [cluster] Move API url to a config property
- [HUE-8591](#) - [cluster] Prevent red error popups
- [HUE-8591](#) - [cluster] Fix name of default cluster
- [HUE-8591](#) - [cluster] Use properly Impala Thrift Client on remote Impala cluster direct connection
- [HUE-8592](#) - [frontend] Enable default click to navigate for catalog entries table
- [HUE-8592](#) - [frontend] Add option to automatically refresh samples in the catalog entries table
- [HUE-8592](#) - [frontend] Create a polling catalog entries list component that waits until an entity exists
- [HUE-8594](#) - [editor] Avoid js error when lastSelectedCompute does not exist
- [HUE-8595](#) - [flume] Collect and ingest Hue balancer logs out of the box
- [HUE-8597](#) - [frontend] Use the default SQL interpreter as source type in the global search results
- [HUE-8599](#) - [frontend] Add pubSub to force clear the context catalog from the job browser
- [HUE-8599](#) - [frontend] Improve stability of the context selector

- [HUE-8600](#) - [tb] Limit Table Browser namespace selection to namespaces with active computes
- [HUE-8601](#) - [jb] Fix issue where context selector in mini jb is hidden behind expand text
- [HUE-8602](#) - [sentry] Remove ALTER and DROP table privileges for now
- [HUE-8602](#) - [sentry] Remove ALTER and DROP in the Hive section
- [HUE-8603](#) - [editor] Always show the query compatibility check results
- [HUE-8604](#) - [frontend] Use the latest opened database by default throughout
- [HUE-8606](#) - [s3] Opening S3 browser makes a call to HDFS
- [HUE-8607](#) - [tb] Include namespace when querying a table from the table browser
- [HUE-8607](#) - [tb] Fix broken drop table action in the table browser
- [HUE-8607](#) - [tb] Fix query and view table actions in the table browser
- [HUE-8609](#) - [tb] Fix exception in describe table call from the Table Browser
- [HUE-8610](#) - [tb] Make sure the created notebook for samples requests has the provided compute
- [HUE-8610](#) - [tb] Include compute in stats and describe table calls from the table browser
- [HUE-8610](#) - [core] Always send the full cluster instead of id to the APIs
- [HUE-8610](#) - [tb] Include compute when fetching samples from the table browser
- [HUE-8611](#) - [assist] Send cluster parameter with the invalidate calls
- [HUE-8612](#) - [editor] Improve the editor shortcut search to show results from all categories
- [HUE-8612](#) - [editor] Add missing keyboard shortcuts to the editor help
- [HUE-8613](#) - [tb] Send cluster when dropping databases from the table browser
- [HUE-8614](#) - [tb] Fix the create new database action in the Table Browser
- [HUE-8615](#) - [frontend] Make sure namespaces and computes always have a name in the context selector
- [HUE-8617](#) - [frontend] Add pubSub to the context selector for setting cluster/compute/namespace
- [HUE-8618](#) - [editor] Prevent js exception when typing while the context is loading
- [HUE-8619](#) - [tb] Include cluster in the partitions API call
- [HUE-8619](#) - [tb] Switch to POST for partitions API call
- [HUE-8621](#) - [editor] Add a custom Ace mode for the dark theme
- [HUE-8621](#) - [editor] Add keyboard shortcut to toggle dark mode
- [HUE-8621](#) - [editor] Add ace option to toggle dark mode
- [HUE-8621](#) - [editor] Add dark mode keyboard shortcut to the editor help
- [HUE-8623](#) - [frontend] Send cluster when checking if a table or database exists in the importer
- [HUE-8624](#) - [beeswax] Fix tests on create database to redirect on a v4 page
- [HUE-8625](#) - [editor] Prevent js exception when dragging from top search to the editor after visiting the importer
- [HUE-8626](#) - [security] Fix navigation issues after visiting the security app
- [HUE-8627](#) - [frontend] Add partition result view to the top search
- [HUE-8628](#) - [assist] Indicate context in the left assist filter placeholder
- [HUE-8629](#) - [assist] Don't show a database icon in the breadcrumb of non sql type assist panels
- [HUE-8629](#) - [assist] Customise the assist icons for streams
- [HUE-8629](#) - [assist] Add a dedicated streams assist panel
- [HUE-8629](#) - [assist] Make sure entries are loaded in left assist for non sql types
- [HUE-8629](#) - [assist] Improve assist context menu for kafka
- [HUE-8630](#) - [core] Fix TestMetastoreWithHadoop.test_basic_flow_get_apps
- [HUE-8630](#) - Fix TestRdbmsIndexer missing RdbmsIndexer
- [HUE-8630](#) - [fb] Fix TestFileBrowserWithHadoop.test_index_home_directory
- [HUE-8634](#) - HUE-8111 [core] Perform 4.3 release
- [HUE-8635](#) - [editor] Add the correct styles to the language reference context popover
- [HUE-8639](#) - [metadata] Do not do Sentry filtering when Sentry is not configured
- [HUE-8639](#) - [metadata] Include the docstring into the configuration
- [HUE-8650](#) - [importer] Fix make_notebook default namespace & compute
- [HUE-8652](#) - [frontend] Fix JS exception in jquery.hiveautocomplete when no namespaces are returned
- [HUE-8654](#) - [editor] Prevent setting empty object for namespace and compute

- [HUE-8654](#) - [editor] Guarantee a namespace and compute is set in single cluster mode
- [HUE-8655](#) - [editor] Have the location handler wait for a compute and namespace to be set
- [HUE-8656](#) - [tb] Make sure a compute is always set in the table browser
- [HUE-8660](#) - [assist] Fix file preview in left assist for files with # in the name
- [HUE-8660](#) - [core] Fix page routing issues with file browser paths containing #
- [HUE-8660](#) - [assist] Support multiple # in file names for assist preview
- [HUE-8662](#) - [core] Fix missing static URLs

Apache Impala

The following issues are fixed in CDH 6.1.0:

- [IMPALA-6202](#) - The mod() function now behaves the same as the % operator.
- [IMPALA-6373](#) - Allow primitive type widening on parquet tables. Impala only supports conversion to those types without any loss of precision:
 - TINYINT (INT32) -> SMALLINT (INT32), INT (INT32), BIGINT (INT64), DOUBLE
 - SMALLINT (INT32) -> INT (INT32), BIGINT (INT64), DOUBLE
 - INT (INT32) -> BIGINT (INT64), DOUBLE
 - FLOAT -> DOUBLE
- [IMPALA-6442](#) - Fixed the incorrectly reported Parquet file offset in error messages.
- [IMPALA-6568](#) - The Query Compilation section was added to profile outputs.
- [IMPALA-6844](#) - Impala now correctly handles a possible null pointer in the to_date() function.
- [IMPALA-7272](#) - Fixed potential crash when a min-max runtime filter is generated for a string value.
- [IMPALA-7449](#) - Fixed network throughput calculation by measuring the network throughput of each individual RPC and uses a summary counter to track avg/min/max of network throughputs.
- [IMPALA-7585](#) - Now Impala always explicitly sets user credentials after creating RPC proxy.
- [IMPALA-7668](#) - Now Impala closes URLClassLoader instances and cleans up any open temporary jar files to avoid file descriptor leaks and disk space issues.
- [IMPALA-7824](#) - Running INVALIDATE METADATA with authorization enabled no longer causes a hang when Sentry is unavailable.

Apache Kafka

The following issues are fixed in CDH 6.1.0:

- [KAFKA-2983](#) - Remove Scala consumers and related code
- [KAFKA-3702](#) - Change log level of SSL close_notify failure
- [KAFKA-4950](#) - Fix ConcurrentModificationException on assigned-partitions metric update
- [KAFKA-5098](#) - KafkaProducer should reject sends to invalid topics
- [KAFKA-5588](#) - Remove deprecated --new-consumer tools option
- [KAFKA-5697](#) - Use nonblocking poll in Streams
- [KAFKA-5891](#) - Proper handling of LogicalTypes in Cast
- [KAFKA-5919](#) - Adding checks on "version" field for tools using it
- [KAFKA-6054](#) - Add 'version probing' to Kafka Streams rebalance
- [KAFKA-6264](#) - Split log segments as needed if offsets overflow the indexes
- [KAFKA-6538](#) - Changes to enhance ByteStore exceptions thrown from RocksDBStore with more human readable info
- [KAFKA-6546](#) - Use LISTENER_NOT_FOUND error for missing listener
- [KAFKA-6562](#) - Make jackson-databind an optional clients dependency
- [KAFKA-6648](#) - Fetcher.getTopicMetadata() should return all partitions for each requested topic
- [KAFKA-6697](#) - Broker should not die if getCanonicalPath fails
- [KAFKA-6704](#) - InvalidStateStoreException from IQ when StreamThread closes store
- [KAFKA-6711](#) - GlobalStateManagerImpl should not write offsets of in-memory stores in checkpoint file
- [KAFKA-6726](#) - Fine Grained ACL for CreateTopics (KIP-277)

- [KAFKA-6730](#) - Simplify State Store Recovery
- [KAFKA-6743](#) - ConsumerPerformance fails to consume all messages [KIP-281]
- [KAFKA-6749](#) - Fixed TopologyTestDriver to process stream processing guarantee as exactly once
- [KAFKA-6750](#) - Add listener name to authentication context (KIP-282)
- [KAFKA-6760](#) - Fix response logging in the Controller
- [KAFKA-6782](#) - solved the bug of restoration of aborted messages for GlobalStateStore and KGlobalTable
- [KAFKA-6805](#) - Enable broker configs to be stored in ZK before broker start
- [KAFKA-6809](#) - Count inbound connections in the connection-creation metric
- [KAFKA-6813](#) - return to double-counting for count topology names
- [KAFKA-6841](#) - Support Prefixed ACLs (KIP-290)
- [KAFKA-6859](#) - Do not send LeaderEpochRequest for undefined leader epochs
- [KAFKA-6860](#) - Fix NPE in Kafka Streams with EOS enabled
- [KAFKA-6884](#) - Consumer group command should use new admin client
- [KAFKA-6897](#) - Prevent KafkaProducer.send from blocking when producer is closed
- [KAFKA-6906](#) - MINOR: code cleanup follow up for
- [KAFKA-6906](#) - Fixed to commit transactions if data is produced via wall clock punctuation
- [KAFKA-6927](#) - Chunked down-conversion to prevent out of memory errors on broker [KIP-283]
- [KAFKA-6935](#) - Add config for allowing optional optimization
- [KAFKA-6936](#) - Implicit materialized for aggregate, count and reduce
- [KAFKA-6944](#) - Add system tests testing the new throttling behavior using older clients/brokers
- [KAFKA-6946](#) - Keep the session id for incremental fetch when fetch responses are throttled
- [KAFKA-6949](#) - alterReplicaLogDirs() should grab partition lock when accessing log of the future replica
- [KAFKA-6955](#) - Use Java AdminClient in DeleteRecordsCommand
- [KAFKA-6967](#) - TopologyTestDriver does not allow pre-populating state stores that have change logging
- [KAFKA-6973](#) - Validate topic config message.timestamp.type
- [KAFKA-6975](#) - Fix replica fetching from non-batch-aligned log start offset
- [KAFKA-6979](#) - Add `default.api.timeout.ms` to KafkaConsumer (KIP-266)
- [KAFKA-6981](#) - Move the error handling configuration properties into the ConnectorConfig and SinkConnectorConfig classes
- [KAFKA-6986](#) - Export Admin Client metrics through Stream Threads
- [KAFKA-6991](#) - Fix ServiceLoader issue with PluginClassLoader
- [KAFKA-6997](#) - Exclude test-sources.jar when \$INCLUDE_TEST_JARS is FALSE
- [KAFKA-7001](#) - Rename errors.allowed.max property in Connect to errors.tolerance
- [KAFKA-7002](#) - Add a config property for DLQ topic's replication factor
- [KAFKA-7003](#) - Set error context in message headers
- [KAFKA-7005](#) - Remove duplicate resource class.
- [KAFKA-7006](#) - remove duplicate Scala ResourceNameType in preference to...
- [KAFKA-7007](#) - Use JSON for /kafka-acl-extended-changes path
- [KAFKA-7010](#) - Rename ResourceNameType to PatternType
- [KAFKA-7011](#) - Remove ResourceNameType field from Java Resource class.
- [KAFKA-7012](#) - Don't process SSL channels without data to process
- [KAFKA-7019](#) - Make reading metadata lock-free by maintaining an atomically-updated read snapshot
- [KAFKA-7021](#) - Reuse source based on config
- [KAFKA-7021](#) - Update upgrade guide section for reusing source topic
- [KAFKA-7023](#) - Move prepareForBulkLoad() call after customized RocksDBConfigSetter
- [KAFKA-7028](#) - Properly authorize custom principal objects
- [KAFKA-7029](#) - Update ReplicaVerificationTool not to use SimpleConsumer
- [KAFKA-7030](#) - Add configuration to disable message down-conversion (KIP-283)
- [KAFKA-7031](#) - Connect API shouldn't depend on Jersey
- [KAFKA-7032](#) - The TimeUnit is neglected by KafkaConsumer#close(long, Tim...

- [KAFKA-7039](#) - Create an instance of the plugin only it's a Versioned Plugin
- [KAFKA-7043](#) - Modified plugin isolation whitelist with recently added converters
- [KAFKA-7044](#) - Fix Fetcher.fetchOffsetsByTimes and NPE in describe consumer group
- [KAFKA-7047](#) - Added SimpleHeaderConverter to plugin isolation whitelist
- [KAFKA-7048](#) - NPE when creating connector
- [KAFKA-7050](#) - Decrease default consumer request timeout to 30s
- [KAFKA-7055](#) - Update InternalTopologyBuilder to throw TopologyException if a processor or sink is added with no upstream node attached
- [KAFKA-7056](#) - Moved Connects new numeric converters to runtime
- [KAFKA-7058](#) - Comparing schema default values using Objects#deepEquals()
- [KAFKA-7066](#) - added better logging in case of Serialisation issue
- [KAFKA-7068](#) - Handle null config values during transform
- [KAFKA-7076](#) - Skip rebuilding producer state when using old message format
- [KAFKA-7080](#) - pass segmentInterval to CachingWindowStore
- [KAFKA-7082](#) - Concurrent create topics may throw NodeExistsException
- [KAFKA-7091](#) - AdminClient should handle FindCoordinatorResponse errors
- [KAFKA-7097](#) - HOTFIX:; Set create time default to -1L in VerifiableProducer
- [KAFKA-7097](#) - VerifiableProducer does not work properly with --message-create-time argument
- [KAFKA-7104](#) - More consistent leader's state in fetch response
- [KAFKA-7111](#) - Log error connecting to node at a higher log level
- [KAFKA-7112](#) - MINOR:Only resume restoration if state is still PARTITIONS_ASSIGNED after poll
- [KAFKA-7119](#) - Handle transient Kerberos errors as non-fatal exceptions
- [KAFKA-7119](#) - Handle transient Kerberos errors on server side
- [KAFKA-7126](#) - Reduce number of rebalance for large consumer group after a topic is created
- [KAFKA-7128](#) - Follower has to catch up to offset within current leader epoch to join ISR
- [KAFKA-7136](#) - Avoid deadlocks in synchronized metrics reporters
- [KAFKA-7144](#) - Fix task assignment to be even
- [KAFKA-7147](#) - ReassignPartitionsCommand should be able to connect to broker over SSL
- [KAFKA-7164](#) - Follower should truncate after every missed leader epoch change
- [KAFKA-7168](#) - Treat connection close during SSL handshake as retriable
- [KAFKA-7182](#) - SASL/OAUTHBEARER client response missing %x01 seps
- [KAFKA-7185](#) - Allow empty resource name when matching ACLs
- [KAFKA-7192](#) - Follow-up: update checkpoint to the reset beginning offset
- [KAFKA-7192](#) - Wipe out if EOS is turned on and checkpoint file does not exist
- [KAFKA-7194](#) - Fix buffer underflow if onJoinComplete is retried after failure
- [KAFKA-7216](#) - Ignore unknown ResourceTypes while loading acl cache
- [KAFKA-7228](#) - Set errorHandlingMetrics for dead letter queue
- [KAFKA-7231](#) - Ensure NetworkClient uses overridden request timeout
- [KAFKA-7242](#) - Reverse xform configs before saving
- [KAFKA-7250](#) - switch scala transform to TransformSupplier
- [KAFKA-7250](#) - fix transform function in scala DSL to accept TranformerSupplier
- [KAFKA-7255](#) - Fix timing issue with create/update in SimpleAclAuthorizer
- [KAFKA-7261](#) - Record 1.0 for total metric when Count stat is used for rate
- [KAFKA-7278](#) - replaceSegments() should not call asyncDeleteSegment() for segments which have been removed from segments list
- [KAFKA-7280](#) - Synchronize consumer fetch request/response handling
- [KAFKA-7284](#) - streams should unwrap fenced exception
- [KAFKA-7285](#) - Create new producer on each rebalance if EOS enabled
- [KAFKA-7286](#) - Avoid getting stuck loading large metadata records
- [KAFKA-7287](#) - Set open ACL for old consumer znode path

- [KAFKA-7296](#) - Handle coordinator loading error in TxnOffsetCommit
- [KAFKA-7298](#) - Raise UnknownProducerIdException if next sequence number is unknown
- [KAFKA-7301](#) - Fix streams Scala join ambiguous overload
- [KAFKA-7316](#) - Fix Streams Scala filter recursive call #5538
- [KAFKA-7322](#) - Fix race condition between log cleaner thread and log retention thread when topic cleanup policy is updated
- [KAFKA-7347](#) - Return not leader error for OffsetsForLeaderEpoch requests to non-replicas
- [KAFKA-7353](#) - Connect logs 'this' for anonymous inner classes
- [KAFKA-7354](#) - Fix IdlePercent and NetworkProcessorAvgIdlePercent metric
- [KAFKA-7369](#) - Handle retriable errors in AdminClient list groups API
- [KAFKA-7385](#) - Fix log cleaner behavior when only empty batches are retained
- [KAFKA-7386](#) - streams-scala should not cache serdes
- [KAFKA-7388](#) - equal sign in property value for password
- [KAFKA-7414](#) - Out of range errors should never be fatal for follower
- [KAFKA-7434](#) - Fix NPE in DeadLetterQueueReporter
- [KAFKA-7453](#) - Expire registered channels not selected within idle timeout
- [KAFKA-7454](#) - Use lazy allocation for SslTransportLayer buffers and null them on close
- [KAFKA-7459](#) - Use thread-safe Pool for RequestMetrics.requestRateInternal
- [KAFKA-7460](#) - Fix Connect Values converter date format pattern

Apache Kudu

The following issues are fixed in CDH 6.1.0:

- [KUDU-844](#) - [webui]and other /tablet-rowsetlayout-svg improvements
- [KUDU-972](#) - Fixed an issue where Kudu's block cache memory tracking (as seen on the /mem-trackers web UI page) wasn't accounting for all of the overhead of the cache itself.
- [KUDU-1038](#) - When a tablet is deleted, its write-ahead log recovery directory is also deleted, if it exists.
- [KUDU-2179](#) - Fixed an issue where kudu cluster ksck running a snapshot checksum scan would use a single snapshot timestamp for all tablets. This caused the checksum process to fail if the checksum process took a long time and the number of tablets was sufficiently large. The tool should now be able to checksum tables even if the process takes many hours.
- [KUDU-2260](#) - Fixed a rare issue where system failure could leave unexpected null bytes at the end of metadata files, causing Kudu to be unable to restart.
- [KUDU-2293](#) - Fixed an issue with failed tablet copies that would cause subsequent tablet copies to crash the tablet server.
- [KUDU-2322](#) - Fixed a bug where leader logged excessively when the followers fell behind.
- [KUDU-2324](#) - Add gflags to disable individual maintenance ops.
- [KUDU-2335](#) - Fixed reporting of leader health during lifecycle transitions.
- [KUDU-2364](#) - When a tablet server was wiped and recreated with the same RPC address, ksck listed it twice, both as healthy, even though only one of them was there. This bug is now fixed by verifying the UUID of the server.
- [KUDU-2406](#) - Fixed an issue preventing Kudu from starting when using Vormetric's encrypted filesystem (secrets2) on ext4.
- [KUDU-2414](#) - Fixed an issue where the C++ client would fail to reopen an expired scanner; instead, the client would retry in a tight loop and eventually timeout.
- [KUDU-2437](#) - Split a tablet into primary key ranges by size.
- [KUDU-2443](#) - Fixed moving single-replica tablets.
- [KUDU-2447](#) - Fixed a tablet server crash when a tablet is scanned with two predicates on its primary key and the predicates do not overlap.
- [KUDU-2463](#) - Fixed a bug in which incorrect results would be returned in scans following a server restart.
- [KUDU-2509](#) - Fixed use-after-free in case of WAL replay error.
- [KUDU-2510](#) - Fixed symmetric difference logging.
- [KUDU-2521](#) - Java Implementation for BloomFilter.

- [KUDU-2525](#) - Fixed an issue where the Kudu MapReduce connector's KuduTableInputFormat may exhaust its scan too early.
- [KUDU-2531](#) - (part 1) Ignore invalid tablet metadata files.
- [KUDU-2531](#) - (part 2) Add -nobackup flag to pbc edit tool.
- [KUDU-2540](#) - Fixed a bug causing a tablet server crash when a write batch request from a client failed coarse-grained authorization.
- [KUDU-2580](#) - Fixed authentication token reacquisition in the C++ client.
- [KUDU-2601](#) - Correctly print newly created files.
- Fixed an error that would cause the Kudu CLI tool to unexpectedly exit when the connection to the master or tserver was abruptly closed.

Apache Oozie

The following issues are fixed in CDH 6.1.0:

- [OOZIE-2427](#) - [Kerberos] Authentication failure for the javascript resources under /ext-2.2
- [OOZIE-2791](#) - ShareLib installation may fail on busy Hadoop clusters
- [OOZIE-2867](#) - [Coordinators] Emphasize Region/City timezone format
- [OOZIE-2883](#) - ProxyUserService: invalid configuration error message is misleading
- [OOZIE-2914](#) - Consolidate trim calls
- [OOZIE-2934](#) - [sharelib/spark] Fix Findbugs error
- [OOZIE-2967](#) - TestStatusTransitService.testBundleStatusCoordSubmitFails fails intermittently in Apache Oozie Core 5.0.0-SNAPSHOT
- [OOZIE-2968](#) - TestJavaActionExecutor.testCredentialsSkip fails intermittently
- [OOZIE-3134](#) - Potential inconsistency between the in-memory SLA map and the Oozie database
- [OOZIE-3155](#) - [ui] Job DAG is not refreshed when a job is finished
- [OOZIE-3217](#) - Enable definition of admin users using oozie-site.xml
- [OOZIE-3221](#) - Rename DEFAULT_LAUNCHER_MAX_ATTEMPS
- [OOZIE-3224](#) - Upgrade Jetty to 9.3
- [OOZIE-3229](#) - [client] [ui] Improved SLA filtering options
- [OOZIE-3229](#) - [build] test-patch-30-distro improvement
- [OOZIE-3233](#) - Remove DST shift from the coordinator job's end time
- [OOZIE-3235](#) - Upgrade ActiveMQ to 5.15.3
- [OOZIE-3251](#) - Disable JMX for ActiveMQ in the tests
- [OOZIE-3257](#) - TestHiveActionExecutor#testHiveAction still fails
- [OOZIE-3260](#) - [sla] Remove stale item above max retries on JPA related errors from in-memory SLA map
- [OOZIE-3298](#) - [MapReduce action] External ID is not filled properly and failing MR job is treated as SUCCEEDED
- [OOZIE-3303](#) - Oozie UI does not work after Jetty 9.3 upgrade
- [OOZIE-3309](#) - Runtime error during /v2/sla filtering for bundle
- [OOZIE-3310](#) - SQL error during /v2/sla filtering
- [OOZIE-3348](#) - [Hive action] Remove dependency hive-contrib
- [OOZIE-3354](#) - [core] [SSH action] SSH action gets hung
- [OOZIE-3369](#) - [core] Upgrade guru.nidi:graphviz-java to 0.7.0
- [OOZIE-3370](#) - Property filtering is not consistent across job submission
- [OOZIE-3376](#) - [tests] TestGraphGenerator should assume JDK8 minor version at least 1.8.0_u40
- [OOZIE-3378](#) - Coordinator action's status is SUBMITTED after E1003 error

Apache Parquet

The following issues are fixed in CDH 6.1.0:

- [PARQUET-952](#) - Avro union with single type fails with 'is not a group'
- [PARQUET-1417](#) - BINARY_AS_SIGNED_INTEGER_COMPARATOR fails with IOBE for the same arrays with the different length

Apache Pig

There are no notable fixed issues in this release.

Cloudera Search

The following issues are fixed in CDH 6.1.0:

- [SOLR-12541](#) - Metrics handler throws an error if there are transient cores.
- [SOLR-12594](#) - MetricsHistoryHandler.getOverseerLeader fails when hostname contains hyphen.
- [SOLR-12683](#) - HashQuery will throw an exception if more than 4 partitionKeys is specified.
- [SOLR-12704](#) - Guard AddSchemaFieldsUpdateProcessorFactory against null field names and field values.
- [SOLR-12750](#) - Migrate API should lock the collection instead of shard.
- [SOLR-12765](#) - Incorrect format of JMX cache stats.
- [SOLR-12836](#) - ZkController creates a cloud solr client with no connection or read timeouts.

For more information on the fixes, see the upstream release notes:

- [Apache Solr 7.1 Release Notes](#)
- [Apache Solr 7.2 Release Notes](#)
- [Apache Solr 7.3 Release Notes](#)
- [Apache Solr 7.4 Release Notes](#)

Apache Sentry

The following issues are fixed in CDH 6.1.0:

- [SENTRY-853](#) - Handle show grant on auth failure correctly
- [SENTRY-1572](#) - SentryMain() shouldn't dynamically load tool class
- [SENTRY-1896](#) - Optimize retrieving entities by other entity types
- [SENTRY-1944](#) - Optimize DelegateSentryStore.getGroupsByRoles() and update SentryGenericPolicyProcessor to retrieve roles to group mapping in a single transaction
- [SENTRY-2085](#) - Sentry error handling exposes SentryGroupNotFoundException externally.
- [SENTRY-2092](#) - Drop Role log message shows "Creating role"
- [SENTRY-2115](#) - Incorrect behavior of HMsFollower when HDFSsync feature is disabled.
- [SENTRY-2127](#) - Fix unstable unit test TestColumnEndToEnd.testCrossDbTableOperations
- [SENTRY-2141](#) - Sentry Privilege TimeStamp is not converted to grantTime in HivePrivilegeInfo correctly
- [SENTRY-2143](#) - Table renames should synchronize with Sentry
- [SENTRY-2168](#) - Altering table will not update sentry permissions when HDFS sync is disabled
- [SENTRY-2194](#) - Upgrade Sentry hadoop-version dependency to 2.7.5
- [SENTRY-2198](#) - Update to Kafka 1.0.0.
- [SENTRY-2199](#) - Bump Hive version from 2.3.2 to 2.3.3
- [SENTRY-2200](#) - Migrate 3.x Datanucleus unsupported configurations to 4.1 Datanucleus
- [SENTRY-2209](#) - Incorrect class in SentryHdfsMetricsUtil.java.
- [SENTRY-2210](#) - AUTHZ_PATH should have index on the foreign key AUTHZ_OBJ_ID
- [SENTRY-2213](#) - Increase schema version from 2.0.0 to 2.1.0
- [SENTRY-2214](#) - Sentry should not allow URI grants to EMPTY or NULL locations
- [SENTRY-2224](#) - Support SHOW GRANT on HIVE_OBJECT
- [SENTRY-2231](#) - Fix URI check on List Privileges by Provider in SentryStore
- [SENTRY-2238](#) - Explicitly set Database on SentryHivePrivilegeObjectDesc
- [SENTRY-2244](#) - Alter sentry role or user at granting privilege can avoid extra query to database
- [SENTRY-2245](#) - Remove privileges that do not associate with a role or a user
- [SENTRY-2251](#) - Update user privileges based on changes to authorizables
- [SENTRY-2252](#) - Normalize the Sentry store API's to handle both user/role privileges
- [SENTRY-2255](#) - alter table set owner command can be executed only by user with proper privilege
- [SENTRY-2258](#) - Remove user when it is not associated with other objects

- [SENTRY-2259](#) - SQL CONSTRAINT name is too long for Oracle 11.2
- [SENTRY-2261](#) - Implement JSONAlterDatabaseMessage to write HMS alter database events
- [SENTRY-2262](#) - Sentry client is not compatible when connecting to Sentry 2.0
- [SENTRY-2264](#) - It is possible to elevate privileges from DROP using alter table rename
- [SENTRY-2270](#) - Illegal privileges on columns can be granted on Hive
- [SENTRY-2271](#) - Wrong log messages/method names in SentrySchema related classes.
- [SENTRY-2273](#) - Create the SHOW GRANT USER task for Hive
- [SENTRY-2280](#) - The request received in SentryPolicyStoreProcessor.sentry_notify_hms_event is null
- [SENTRY-2281](#) - list_privileges_by_user() fails with a JDODetachedFieldAccessException
- [SENTRY-2293](#) - Fix logging parameters on SentryHDFSServiceProcessor
- [SENTRY-2294](#) - Add requestorUsername to client.notifyHmsEvent() method
- [SENTRY-2295](#) - Owner privileges should not be granted to sentry admin users
- [SENTRY-2307](#) - Avoid HMS event synchronization while sentry is fetching full snapshot
- [SENTRY-2309](#) - Port ModifiedCatch NPE thrown when fetching Partitions with no corresponding SDS entry
- [SENTRY-2310](#) - Sentry is not be able to fetch full update subsequently, when there is HMS restart in the snapshot process.
- [SENTRY-2312](#) - Update owner privileges for table when owner is changed.
- [SENTRY-2313](#) - alter database set owner command can be executed only by user with proper privilege
- [SENTRY-2315](#) - The grant all operation is not dropping the create/alter/drop/index/lock privileges
- [SENTRY-2324](#) - Allow sentry to fetch configurable notifications from HMS
- [SENTRY-2332](#) - Load hadoop default configuration when starting sentry service
- [SENTRY-2333](#) - Create index AUTHZ_PATH_FK_IDX at table AUTHZ_PATH for Postgres only when it does not exist
- [SENTRY-2352](#) - User roles with ALTER on a table can not show or describe the table on which they have ALTER
- [SENTRY-2359](#) - Object owner is unable to grant privileges: SentryAccessDeniedException
- [SENTRY-2373](#) - Incorrect WARN message when processing add partition messages
- [SENTRY-2376](#) - Bump Jackson libraries versions to 1.9.13 and 2.9.6
- [SENTRY-2392](#) - Add metrics statistics to list_user_privileges and list_role_privileges API
- [SENTRY-2395](#) - ALTER VIEW AS SELECT is asking for CREATE privileges instead of ALTER
- [SENTRY-2403](#) - Incorrect naming in RollingFileWithoutDeleteAppender
- [SENTRY-2406](#) - Make sure inputHierarchy and outputHierarchy have unique values
- [SENTRY-2409](#) - ALTER TABLE SET OWNER does not allow to change the table if using only the table name
- [SENTRY-2417](#) - LocalGroupMappingService class docs do not accurately reflect required INI format
- [SENTRY-2419](#) - Log where sentry stands in the process of persisting the snapshot
- [SENTRY-2423](#) - Increase the allocation size for auto-increment of id's for Snapshot tables.
- [SENTRY-2427](#) - Use Hadoop KerberosName class to derive shortName
- [SENTRY-2429](#) - Transfer database owner drops table owner
- [SENTRY-2432](#) - PortThe case of a username is ignored when determining object ownership
- [SENTRY-2433](#) - Dropping object privileges does not include update of dropping user privileges

Apache Spark

The following issues are fixed in CDH 6.1.0:

- [SPARK-4502](#) - [SQL] Rename to spark.sql.optimizer.nestedSchemaPruning.enabled
- [SPARK-19355](#) - Revert[SPARK-25352]
- [SPARK-19724](#) - [SQL] allowCreatingManagedTableUsingNonemptyLocation should have legacy prefix
- [SPARK-20327](#) - [YARN] Follow up: fix resource request tests on Hadoop 3.
- [SPARK-20327](#) - [CORE][YARN] Add CLI support for YARN custom resources, like GPUs
- [SPARK-20360](#) - [PYTHON] reprs for interpreters
- [SPARK-20594](#) - Adjust fix forfor CDH version of Hive.
- [SPARK-21318](#) - [SQL] Improve exception message thrown by `lookupFunction`
- [SPARK-21402](#) - [SQL] Fix java array of structs deserialization

- [SPARK-22666](#) - [ML][FOLLOW-UP] Improve testcase to tolerate different schema representation
- [SPARK-23401](#) - [PYTHON][TESTS] Add more data types for PandasUDFTests
- [SPARK-23429](#) - [CORE] Add executor memory metrics to heartbeat and expose in executors REST API
- [SPARK-23549](#) - [SQL] Rename config spark.sql.legacy.compareDateTimestampInTimestamp
- [SPARK-23715](#) - Revert "[SQL] the input of to/from_utc_timestamp can not have timezone
- [SPARK-23907](#) - [SQL] Revert regr_* functions entirely
- [SPARK-23972](#) - Revert "[BUILD][SQL] Update Parquet to 1.10.0."
- [SPARK-24157](#) - [SS][FOLLOWUP] Rename to spark.sql.streaming.noDataMicroBatches.enabled
- [SPARK-24324](#) - [PYTHON][FOLLOW-UP] Rename the Conf to
spark.sql.legacy.execution.pandas.groupedMap.assignColumnsByName
- [SPARK-24518](#) - Revert "[CORE] Using Hadoop credential provider API to store password"
- [SPARK-24519](#) - [CORE] Compute SHUFFLE_MIN_NUM_PARTS_TO_HIGHLY_COMPRESS only once
- [SPARK-24709](#) - [SQL][FOLLOW-UP] Make schema_of_json's input json as literal only
- [SPARK-24709](#) - [SQL][2.4] use str instead of basestring in isinstance
- [SPARK-24777](#) - [SQL] Add write benchmark for AVRO
- [SPARK-24787](#) - [CORE] Revert hsync in EventLoggingListener and make FsHistoryProvider to read
lastBlockBeingWritten data for logs
- [SPARK-24918](#) - [CORE] Executor Plugin API
- [SPARK-25021](#) - [K8S][BACKPORT] Add spark.executor.pyspark.memory limit for K8S
- [SPARK-25044](#) - [FOLLOW-UP] Change ScalaUDF constructor signature
- [SPARK-25314](#) - [SQL] Fix Python UDF accessing attributes from both side of join in join conditions
- [SPARK-25318](#) - Add exception handling when wrapping the input stream during the the fetch or stage retry in
response to a corrupted block
- [SPARK-25321](#) - [ML] Fix local LDA model constructor
- [SPARK-25384](#) - [SQL] Clarify fromJsonForceNullableSchema will be removed in Spark 3.0
- [SPARK-25416](#) - [SQL] ArrayPosition function may return incorrect result when right expression is implicitly down
casted
- [SPARK-25417](#) - [SQL] ArrayContains function may return incorrect result when right expression is implicitly down
casted
- [SPARK-25422](#) - [CORE] Don't memory map blocks streamed to disk.
- [SPARK-25425](#) - [SQL][BACKPORT-2.4] Extra options should override session options in DataSource V2
- [SPARK-25450](#) - [SQL] PushProjectThroughUnion rule uses the same exprId for project expressions in each Union
child, causing mistakes in constant propagation
- [SPARK-25454](#) - [SQL] add a new config for picking minimum precision for integral literals
- [SPARK-25460](#) - [BRANCH-2.4][SS] DataSourceV2: SS sources do not respect SessionConfigSupport
- [SPARK-25468](#) - [WEBUI] Highlight current page index in the spark UI
- [SPARK-25469](#) - [SQL] Eval methods of Concat, Reverse and ElementAt should use pattern matching only once
- [SPARK-25495](#) - [SS] FetchedData.reset should reset all fields
- [SPARK-25502](#) - [CORE][WEBUI] Empty Page when page number exceeds the reatinedTask size.
- [SPARK-25503](#) - [CORE][WEBUI] Total task message in stage page is ambiguous
- [SPARK-25505](#) - [SQL] The output order of grouping columns in Pivot is different from the input order
- [SPARK-25505](#) - [SQL][FOLLOWUP] Fix for attributes cosmetically different in Pivot clause
- [SPARK-25509](#) - [CORE] Windows doesn't support POSIX permissions
- [SPARK-25519](#) - [SQL] ArrayRemove function may return incorrect result when right expression is implicitly
downcasted.
- [SPARK-25521](#) - [SQL] Job id showing null in the logs when insert into command Job is finished.
- [SPARK-25522](#) - [SQL] Improve type promotion for input arguments of elementAt function
- [SPARK-25533](#) - [CORE][WEBUI] AppSummary should hold the information about succeeded Jobs and completed
stages only
- [SPARK-25535](#) - [CORE] Work around bad error handling in commons-crypto.
- [SPARK-25536](#) - [CORE] metric value for METRIC_OUTPUT_RECORDS_WRITTEN is incorrect

- [SPARK-25538](#) - [SQL] Zero-out all bytes when writing decimal
- [SPARK-25543](#) - [K8S] Print debug message iff execIdsRemovedInThisRound is not empty.
- [SPARK-25546](#) - [CORE] Don't cache value of EVENT_LOG_CALLSITE_LONG_FORM.
- [SPARK-25568](#) - [CORE] Continue to update the remaining accumulators when failing to update one accumulator
- [SPARK-25579](#) - [SQL] Use quoted attribute names if needed in pushed ORC predicates
- [SPARK-25591](#) - [PYSPARK][SQL] Avoid overwriting deserialized accumulator
- [SPARK-25601](#) - [PYTHON] Register Grouped aggregate UDF Vectorized UDFs for SQL Statement
- [SPARK-25602](#) - [SQL] SparkPlan.getByteArrayRdd should not consume the input when not necessary
- [SPARK-25636](#) - [CORE] spark-submit cuts off the failure reason when there is an error connecting to master
- [SPARK-25644](#) - [SS] Fix java foreachBatch in DataStreamWriter
- [SPARK-25660](#) - [SQL] Fix for the backward slash as CSV fields delimiter
- [SPARK-25669](#) - [SQL] Check CSV header only when it exists
- [SPARK-25673](#) - [BUILD] Remove Travis CI which enables Java lint check
- [SPARK-25674](#) - [SQL] If the records are incremented by more than 1 at a time, the number of bytes might rarely ever get updated
- [SPARK-25674](#) - [FOLLOW-UP] Update the stats for each ColumnarBatch
- [SPARK-25690](#) - [SQL] Analyzer rule HandleNullInputsForUDF does not stabilize and can be applied infinitely
- [SPARK-25697](#) - [CORE] When zstd compression enabled, InProgress application is throwing Error in the history webui
- [SPARK-25704](#) - [CORE] Allocate a bit less than Int.MaxValue
- [SPARK-25708](#) - [SQL] HAVING without GROUP BY means global aggregate
- [SPARK-25714](#) - Fix Null Handling in the Optimizer rule BooleanSimplification
- [SPARK-25718](#) - [SQL] Detect recursive reference in Avro schema and throw exception
- [SPARK-25727](#) - [SQL] Add outputOrdering to otherCopyArgs in InMemoryRelation
- [SPARK-25738](#) - [SQL] Fix LOAD DATA INPATH for hdfs port
- [SPARK-25741](#) - [WEBUI] Long URLs are not rendered properly in web UI
- [SPARK-25768](#) - [SQL] fix constant argument expecting UDAFs
- [SPARK-25776](#) - [CORE] The disk write buffer size must be greater than 12
- [SPARK-25793](#) - [ML] call SaveLoadV2_0.load for classNameV2_0
- [SPARK-25816](#) - [SQL] Fix attribute resolution in nested extractors
- [SPARK-25822](#) - [PYSPARK] Fix a race condition when releasing a Python worker
- [SPARK-25827](#) - [CORE] Avoid converting incoming encrypted blocks to byte buffers
- [SPARK-25840](#) - [BUILD] `make-distribution.sh` should not fail due to missing LICENSE-binary
- [SPARK-25842](#) - [SQL] Deprecate rangeBetween APIs introduced in SPARK-21608
- [SPARK-25854](#) - [BUILD] fix `build/mvn` not to fail during Zinc server shutdown
- [SPARK-25855](#) - [CORE] Don't use erasure coding for event logs by default
- [SPARK-25871](#) - [STREAMING] Don't use EC for streaming WAL
- [SPARK-25904](#) - [CORE] Allocate arrays smaller than Int.MaxValue
- [SPARK-25918](#) - [SQL] LOAD DATA LOCAL INPATH should handle a relative path

Apache Sqoop

The following issues are fixed in CDH 6.1.0:

- [SQOOP-2567](#) - SQOOP import for Oracle fails with invalid precision/scale for decimal
- [SQOOP-2949](#) - SQL Syntax error when split-by column is of character type and min or max value has single quote inside it
- [SQOOP-3042](#) - Sqoop does not clear compile directory under /tmp/sqoop-username/compile automatically
- [SQOOP-3052](#) - Introduce Gradle based build for Sqoop to make it more developer friendly / open
- [SQOOP-3082](#) - Sqoop import fails after TCP connection reset if split by datetime column
- [SQOOP-3224](#) - Mainframe FTP transfer should have an option to use binary mode for transfer
- [SQOOP-3225](#) - Mainframe module FTP listing parser should cater for larger datasets on disk

- [SQOOP-3267](#) - Incremental import to HBase deletes only last version of column
- [SQOOP-3288](#) - Changing OracleManager to use CURRENT_TIMESTAMP instead of
- [SQOOP-3300](#) - Implement JDBC and Kerberos tools for HiveServer2 support
- [SQOOP-3309](#) - Implement HiveServer2 client
- [SQOOP-3326](#) - Mainframe FTP listing for GDG should filter out non-GDG datasets in a heterogeneous listing
- [SQOOP-3327](#) - Mainframe FTP needs to include "Migrated" datasets when parsing the FTP list
- [SQOOP-3328](#) - Implement an alternative solution for Parquet reading and writing
- [SQOOP-3330](#) - Sqoop --append does not work with -Dmapreduce.output.basename
- [SQOOP-3331](#) - Add Mainframe FTP integration test for GDG dataset.
- [SQOOP-3333](#) - Change default behavior of the MS SQL connector to non-resilient.
- [SQOOP-3335](#) - Add Hive support to the new Parquet writing implementation
- [SQOOP-3353](#) - Sqoop should not check incremental constraints for HBase imports
- [SQOOP-3378](#) - Error during direct Netezza import/export can interrupt process in uncontrolled ways

Apache Zookeeper

The following issues are fixed in CDH 6.1.0:

- [ZOOKEEPER-706](#) - Large numbers of watches can cause session re-establishment to fail
- [ZOOKEEPER-1382](#) - Zookeeper server holds onto dead/expired session ids in the watch data structures

Unsupported Features in CDH 6.1.0

This page lists the unsupported features in CDH 6.1.x. For the complete list of Known Issues and Limitations, see [Known Issues and Limitations in CDH 6.1.0](#) on page 403.

Apache Hadoop Unsupported Features

The following sections list unsupported features in Hadoop common components:

- [HDFS Unsupported Features](#) on page 366
- [YARN Unsupported Features](#) on page 366

HDFS Unsupported Features

The following HDFS features are not supported in CDH 6.x:

- ACLs for the NFS gateway
- Aliyun Cloud Connector
- HDFS Router Based Federation
- HDFS truncate
- More than two NameNodes
- Openstack Swift
- Quota support for Storage Types
- SFTP FileSystem
- Upgrade Domain
- Variable length block
- ZStandard Compression Codec

YARN Unsupported Features

The following YARN features are not supported in CDH 6.1.x:

- Application Timeline Server v2 (ATSV2)
- Cgroup Memory Enforcement
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Docker on YARN (DockerContainerExecutor)

- Native Services
- New Aggregated Log File Format
- Node Labels
- Pluggable Scheduler Configuration
- Reservation REST APIs
- Resource Estimator Service
- Resource Profiles
- Rolling Log Aggregation
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- YARN WebUI v2

Apache HBase Unsupported Features

The following HBase features are not supported in CDH 6.1.x:

- Master hosting meta
- Cloudera does not provide support for user-provided custom coprocessors of any kind.
- Server-side encryption of HFiles. You should configure HDFS client-side encryption.
- In-memory compaction
- Visibility labels
- Stripe compaction
- Clients setting priority on operations
- Specifying a custom asynchronous connection implementation
- Client tarball
- Serial replication
- Rolling upgrade from CDH 5 HBase versions

Apache Hive Unsupported Features

The following Hive features are not supported in CDH 6.1.x:

- AccumuloStorageHandler ([HIVE-7068](#))
- ACID ([HIVE-5317](#))
- Built-in `version()` function is not supported (CDH-40979)
- Cost-based Optimizer (CBO)
- Explicit Table Locking
- HCatalog - HBase plugin
- Hive Authorization (Instead, use Apache Sentry.)
- Hive on Apache Tez
- Hive Local Mode Execution
- Hive Metastore - Derby
- Hive Web Interface (HWI)
- HiveServer1 / JDBC 1
- HiveServer2 Dynamic Service Discovery (HS2 HA) ([HIVE-8376](#))
- HiveServer2 - HTTP Mode (Use THRIFT mode.)
- HPL/SQL ([HIVE-11055](#))
- LLAP (Live Long and Process framework)
- Scalable Dynamic Partitioning and Bucketing Optimization ([HIVE-6455](#))
- Session-level Temporary Tables ([HIVE-7090](#))
- Table Replication Across HCatalog Instances ([HIVE-7341](#))

Apache Kafka Unsupported Features

The following Kafka features are not supported in CDH 6.1.x:

- Idempotent and transactional capabilities in the producer are currently an unsupported beta feature given their maturity and complexity. This feature will be supported in a future release.
- Kafka Connect is included in CDH 6.1.0, but is not supported. Flume and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. See [Flafka: Apache Flume Meets Apache Kafka for Event Processing](#) for more information.
- Kafka Streams is included in CDH 6.0.0, but is not supported. Instead, use Spark and Spark Streaming have a fully functional ETL/stream processing pipeline. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for more information.
- The Kafka default authorizer is included in CDH 6.1.0, but is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.
- The legacy Scala clients (producer and consumer) that are under the `kafka.producer.*` and `kafka.consumer.*` package are removed in CDH 6.1.0. See [Deprecated Scala-based Client API and New Java Client API](#) on page 544.

Apache Oozie Unsupported Features

The following Oozie feature is not supported in CDH 6.1.x:

- Conditional coordinator input logic.

Apache Pig Unsupported Features

The following Pig features are not supported in CDH 6.1.x:

- Pig on Tez is not supported in CDH 6.1.x ([PIG-3446](#) / [PIG-3419](#)).
- Pig on Spark.

Cloudera Search Unsupported Features

The following Search features are not supported in CDH 6.1.x:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation
- Saving search results

Apache Sentry Unsupported Features

The following Sentry features are not supported in CDH 6.1.x:

- Import and export of Sentry metadata to and from Sentry servers
- Sentry shell command line for Hive
- Relative URI paths ([Known Issue](#))
- Object types Server and URL in

```
show grant role <role name> on object <object name>
```

- ([Known Issue](#))
- ALTER and DROP privileges for Hive and Impala

In addition, as of CDH 6.0.x, Sentry policy files have been removed. See the Sentry [Incompatible Changes](#) for more information.

Apache Spark Unsupported Features

The following Spark features are not supported in CDH 6.1.x:

- Apache Spark experimental features/APIs are not supported unless stated otherwise.

- Using the JDBC Datasource API to access Hive or Impala is not supported
- ADLS not Supported for All Spark Components. Microsoft Azure Data Lake Store (ADLS) is a cloud-based filesystem that you can access through Spark applications. Spark with Kudu is not currently supported for ADLS data. (Hive on Spark is available for ADLS in CDH 5.12 and higher.)
- IPython / Jupyter notebooks is not supported. The IPython notebook system (renamed to Jupyter as of IPython 4.0) is not supported.
- Certain Spark Streaming features not supported. The `mapWithState` method is unsupported because it is a nascent unstable API.
- Thrift JDBC/ODBC server is not supported
- Spark SQL CLI is not supported
- GraphX is not supported
- SparkR is not supported
- Structured Streaming is supported, but the following features of it *are not*:
 - Continuous processing, which is still experimental, is not supported.
 - Stream static joins with HBase have not been tested and therefore are not supported.
- Spark cost-based optimizer (CBO) not supported.

Apache Sqoop Unsupported Features

The following Sqoop feature is not supported in CDH 6.1.x:

- [import-mainframe](#)

Incompatible Changes in CDH 6.1.0



Important:

In addition to incompatible changes, CDH 6 also deprecated or removed support for several components, including Spark 1 and MapReduce v1. For information about components, sub-components, or functionality that are deprecated or no longer supported, see [Deprecated Items](#) on page 667.

See below for incompatible changes in CDH 6.1.0, grouped by component:

Apache Accumulo

CDH 6.1.0 introduces no new incompatible changes for Apache Accumulo.

Apache Avro

API Changes

One method was removed in CDH 6.0.0:

```
GenericData.toString (Object datum, StringBuilder buffer)
```

Incompatible Changes from Avro 1.8.0

- Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.
- [AVRO-997](#): Generic API requires GenericEnumSymbol - likely to break current Generic API users that often have String or Java Enum for these fields
- [AVRO-1502](#): Avro Objects now Serializable - IPC needs to be regenerated/recompiled
- [AVRO-1602](#): removed Avro internal RPC tracing, presumed unused. Current rec would be HTrace
- [AVRO-1586](#): Compile against Hadoop 2 - probably not an issue since we've been compiling against Hadoop 2 for C5.
- [AVRO-1589](#): [Java] ReflectData.AllowNulls will create incompatible Schemas for primitive types - may need a KI since it used to fail at runtime but now will fail earlier.

Apache Crunch



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

The following changes are introduced in CDH 6.0.0, and are not backward compatible:

- Crunch is available only as Maven artifacts from the Cloudera Maven repository. It is not included as part of CDH. For more information, see [Apache Crunch Guide](#).
- Crunch supports only Spark 2 and higher releases.
- Crunch supports only HBase 2 and higher releases.
 - The API methods in Crunch-HBase use HBase 2 API types and methods.

Apache Flume

AsyncHBaseSink and HBaseSink

CDH 6 uses HBase 2.0. AsyncHBaseSink is incompatible with HBase 2.0 and is not supported in CDH 6. HBaseSink has been replaced with HBase2Sink. HBase2Sink works the same way as HBaseSink. The only difference is that it is compatible with HBase 2.0. The only additional configuration required to use HBase2Sink is to replace the component name in your configuration.

For example, replace this text:

```
agent.sinks.my_hbase_sink.type = hbase
```

With this:

```
agent.sinks.my_hbase_sink.type = hbase2
```

Or, if you use the FQN of the sink class, replace this text:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase.HBaseSink
```

With this:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase2.HBase2Sink
```

For more information about how to configure HBase2Sink, see [Importing Data Into HBase](#).

com.google.common.collect.ImmutableMap

Flume has removed `com.google.common.collect.ImmutableMap` from the `org.apache.flume.Context` API and replaced it with `java.util.Map` due to Guava compatibility issues ([FLUME-2957](#)). Plugins using the `Context.getParameters()` and `Context.getSubProperties()` APIs will need to assign the return value of those methods to a `Map<String, String>` variable instead of an `ImmutableMap<String, String>` variable, if they do not already do so. Most usages in the Flume codebase already used `Map<String, String>` at the time of this change.

Apache Hadoop

- [HDFS Incompatible Changes](#) on page 370
- [MapReduce](#) on page 372
- [YARN](#) on page 372

HDFS Incompatible Changes

CDH 6.1.0 introduces no new incompatible changes for HDFS.

CDH 6.0.1 introduces no new incompatible changes for HDFS.

CDH 6.0.0, introduces the following incompatible changes for HDFS:

- HFTP has been removed.
- The S3 and S3n connectors have been removed. Users should now use the S3a connector.
- The BookkeeperJournalManager has been removed.
- Changes were made to the structure of the HDFS JAR files to better isolate clients from Hadoop library dependencies. As a result, client applications that depend on Hadoop's library dependencies may no longer work. In these cases, the client applications will need to include the libraries as dependencies directly.
- Several library dependencies were upgraded. Clients that depend on those libraries may break because the library version changes. In these cases, the client applications will need to either be ported to the new library versions or include the libraries as dependencies directly.
- [HDFS-6962](#) changes the behavior of ACL inheritance to better align with POSIX ACL specifications, which states that the umask has no influence when a default ACL propagates from parent to child. Previously, HDFS ACLs applied the client's umask to the permissions when inheriting a default ACL defined on a parent directory. Now, HDFS can ignore the umask in these cases for improved compliance with POSIX. This behavior is on by default due to the inclusion of [HDFS-11957](#). It can be configured by setting `dfs.namenode posix.acl.inheritance.enabled` in `hdfs-site.xml`. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-11957](#) changes the default behavior of ACL inheritance introduced by [HDFS-6962](#). Previously, the behavior was disabled by default. Now, the feature is enabled by default. Any code expecting the old ACL inheritance behavior will have to be updated. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-6252](#) removed `dfshealth.jsp` since it is part of the old NameNode web UI. By default, Cloudera Manager links to the new NameNode web UI, which has an equivalent health page at `dfshealth.html`.
- [HDFS-11100](#) changes the behavior of deleting files protected by a sticky bit. Now, the deletion fails.
- [HDFS-10689](#) changes the behavior of the `hdfs dfs chmod` command. Now, the command resets sticky bit permission on a file/directory when the leading sticky bit is omitted in the octal mode (like 644). When a file or directory permission is applied using octal mode and sticky bit permission needs to be preserved, then it has to be explicitly mentioned in the permission bits (like 1644).
- [HDFS-10650](#) changes the behavior of `DFSClient#mkdirs` and `DFSClient#primitiveMkdir`. Previously, they create a new directory with the default permissions 00666. Now, they will create a new directory with permission 00777.
- [HADOOP-8143](#) changes the default behavior of `distcp`. Previously, the `-pb` option was not used by default, which may have caused some checksums to fail when block sizes did not match. Now, the `-pb` option is included by default to preserve block size when using `distcp`.
- [HADOOP-10950](#) changes several heap management variables:
 - `HADOOP_HEAPSIZE` variable has been deprecated. Use `HADOOP_HEAPSIZE_MAX` and `HADOOP_HEAPSIZE_MIN` instead to set `Xmx` and `Xms`
 - The internal variable `JAVA_HEAP_MAX` has been removed.
 - Default heap sizes have been removed. This will allow for the JVM to use auto-tuning based upon the memory size of the host. To re-enable the old default, configure `HADOOP_HEAPSIZE_MAX= "1g"` in `hadoop-env.sh`.
 - All global and daemon-specific heap size variables now support units. If the variable is only a number, the size is assumed to be in megabytes.
- [HADOOP-14426](#) upgrades the version of Kerby from 1.0.0-RC2 to 1.0.0
- [HDFS-10970](#) updates the version of Jackson from 1.9.13 to 2.x in `hadoop-hdfs`.
- [HADOOP-9613](#) updates the Jersey version to the latest 1.x release.
- [HADOOP-10101](#) updates Guava dependency to 21.0
- [HADOOP-14225](#) removes the `xmlenc` dependency. If you rely on the transitive dependency, you need to set the dependency explicitly in your code after this change.
- [HADOOP-13382](#) remove unneeded `commons-httpclient` dependencies from POM files in Hadoop and sub-projects. This incompatible change may affect projects that have undeclared transitive dependencies on `commons-httpclient`, which used to be provided by `hadoop-common` or `hadoop-client`.
- [HADOOP-13660](#) upgrades the `commons-configuration` version from 1.6 to 2.1.
- [HADOOP-12064](#) upgrades the following dependencies:
 - Guice from 3.0 to 4.0
 - cglib from 2.2 to 3.2.0

- asm from 3.2 to 5.0.4

MapReduce

CDH 6.0.1 introduces no new incompatible changes for MapReduce.

CDH 6.0.0, introduces the following incompatible changes:

- Support for MapReduce v1 has been dropped from CDH 6.0.0.
- CDH 6 supports applications compiled against CDH 5.7.0 and higher MapReduce frameworks. Make sure to not to include the CDH jars with your application by marking them as "provided" in the `pom.xml` file.

YARN

CDH 6.1.0 introduces no new incompatible changes for YARN.

CDH 6.0.1 introduces no new incompatible changes for YARN.

CDH 6.0.0 introduces no new incompatible changes for YARN.

Apache HBase

CDH 6.1.x contains the following downstream HBase incompatible change:

hbase.security.authorization

The default value for `hbase.security.authorization` has been changed from true to false. Secured clusters should make sure to explicitly set it to true in XML configuration file before upgrading to one of these versions ([HBASE-19483](#)). True as the default value of `hbase.security.authorization` was changed because not all clusters need authorization. (History: [HBASE-13275](#)) Rather, only the clusters which need authorization should set this configuration as true.

Incompatible Changes

For more information about upstream incompatible changes, see the Apache Reference Guide [Incompatible Changes](#) and [Upgrade Paths](#).

CDH 6.1.0 introduces the following incompatible changes for HBase:

- [HBASE-20270](#): Error triggered command help is no longer available.
- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.

CDH 6.0.x introduces the following upstream HBase incompatible changes:

- [HBASE-20406](#): Prevent Thrift in HTTP mode to accept the TRACE and OPTIONS methods.
- Public interface API changes:
 - [HBASE-15607](#): Admin
 - [HBASE-19112](#), [HBASE-18945](#): Cell
 - Region, Store, HBaseTestingUtility
- [HBASE-18792](#): hbase-2 needs to defend against hbck operations
- [HBASE-15982](#): Interface ReplicationEndpoint extends Guava's Service.
- [HBASE-18995](#): Split CellUtil into public CellUtil and PrivateCellUtil for Internal use only.
- [HBASE-19179](#): Purged the hbase-prefix-tree module and all references from the code base.
- [HBASE-17595](#): Add partial result support for small/limited scan; Now small scan and limited scan could also return partial results.
- [HBASE-16765](#): New default split policy, SteppingSplitPolicy.
- [HBASE-17442](#): Move most of the replication related classes from hbase-client to hbase-replication package.
- [HBASE-16196](#): The bundled JRuby 1.6.8 has been updated to version 9.1.9.0.
- [HBASE-18811](#): Filters have been moved from Public to LimitedPrivate.
- [HBASE-18697](#): Replaced hbase-shaded-server jar with hbase-shaded-mapreduce jar.
- [HBASE-18640](#): Moved mapreduce related classes out of hbase-server into separate hbase-mapreduce jar .
- [HBASE-19128](#): Distributed Log Replay feature has been removed.

- [HBASE-19176](#): Hbase-native-client has been removed.
- [HBASE-17472](#): Changed semantics of granting new permissions. Earlier, new grants would override previous permissions, but now, the new and existing permissions get merged.
- [HBASE-18374](#): Previous "mutate" latency metrics has been renamed to "put" metrics.
- [HBASE-15740](#): Removed Replication metric source.shippedKBs in favor of source.shippedBytes.
- [HBASE-13849](#): Removed restore and clone snapshot from the WebUI.
- [HBASE-13252](#): The concept of managed connections in HBase (deprecated before) has now been extinguished completely, and now all callers are responsible for managing the lifecycle of connections they acquire.
- [HBASE-14045](#): Bumped thrift version to 0.9.2.
- [HBASE-5401](#): Changes to number of tasks PE runs when clients are mapreduce. Now tasks == client count. Previous we hardcoded ten tasks per client instance.

Changed Behavior

CDH 6.1.x contains the following HBase behavior changes:

- [HBASE-14350](#): Assignment Manager v2 - Split/Merge have moved to the Master; it runs them now. Hooks around Split/Merge are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-18271](#): Moved to internal shaded netty.
- [HBASE-17343](#): Default MemStore to be CompactingMemStore instead of DefaultMemStore. In-memory compaction of CompactingMemStore demonstrated sizable improvement in HBase's write amplification and read/write performance.
- [HBASE-19092](#): Make Tag IA.LimitedPrivate and expose for CPs.
- [HBASE-18137](#): Replication gets stuck for empty WALS.
- [HBASE-17513](#): Thrift Server 1 uses different QOP settings than RPC and Thrift Server 2 and can easily be misconfigured so there is no encryption when the operator expects it.
- [HBASE-16868](#): Add a replicate_all flag to replication peer config. The default value is true, which means all user tables (REPLICATION_SCOPE != 0) will be replicated to peer cluster.
- [HBASE-19341](#): Ensure Coprocessors can abort a server.
- [HBASE-18469](#): Correct RegionServer metric of totalRequestCount.
- [HBASE-17125](#): Marked Scan and Get's setMaxVersions() and setMaxVersions(int) as deprecated. They are easy to misunderstand with column family's max versions, so use readAllVersions() and readVersions(int) instead.
- [HBASE-16567](#): Core is now up on protobuf 3.1.0 (Coprocessor Endpoints and REST are still on protobuf 2.5.0).
- [HBASE-14004](#): Fix inconsistency between Memstore and WAL which may result in data in remote cluster that is not in the origin (Replication).
- [HBASE-18786](#): FileNotFoundException opening a StoreFile in a primary replica now causes a RegionServer to crash out where before it would be ignored (or optionally handled via close/reopen).
- [HBASE-17956](#): Raw scans will also read TTL expired cells.
- [HBASE-17017](#): Removed per-region latency histogram metrics.
- [HBASE-19483](#): Added ACL checks to RSGroup commands - On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands.
- [HBASE-19358](#): Added ACL checks to RSGroup commands (HBASE-19483): On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands. Improved stability of splitting log when do failover.
- [HBASE-18883](#): Updated our Curator version to 4.0 - Users who experience classpath issues due to version conflicts are recommended to use either the hbase-shaded-client or hbase-shaded-mapreduce artifacts.
- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.
- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.
- [HBASE-15968](#): This issue resolved two long-term issues in HBase: 1) Puts may be masked by a delete before them, and 2) Major compactions change query results. Offers a new behavior to fix this issue with a little performance reduction. Disabled by default. See the issue for details and caveats.
- [HBASE-13701](#): SecureBulkLoadEndpoint has been integrated into HBase core as default bulk load mechanism. It is no longer needed to install it as a coprocessor endpoint.

- [HBASE-9774](#): HBase native metrics and metric collection for coprocessors.
- [HBASE-18294](#): Reduce global heap pressure: flush based on heap occupancy.

Apache Hive/Hive on Spark/HCatalog

Apache Hive

The following changes are introduced to Hive in CDH 6.0.0, and are not backwards compatible:

- [UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting](#) on page 533
- [Support for UNION DISTINCT](#) on page 534
- [OFFLINE and NO_DROP Options Removed from Table and Partition DDL](#) on page 534
- [DESCRIBE Query Syntax Change](#) on page 535
- [CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names](#) on page 535
- [Reserved and Non-Reserved Keyword Changes in HiveQL](#) on page 535
- [Apache Hive API Changes in CDH 6.0.0](#) on page 537
- [Apache Hive Configuration Changes in CDH 6.0.0](#) on page 538
- [HiveServer2 Thrift API Repackaged Resulting in Class File Location Changes](#) on page 540
- [Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column](#) on page 541
- [Hive Logging Framework Switched to SLF4J/Log4j 2](#) on page 541
- [Deprecated Parquet Java Classes Removed from Hive](#) on page 541
- [Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms](#) on page 541
- [S3N Connector Is Removed from CDH 6.0](#) on page 542
- [Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request](#) on page 542
- [Support Added for Escaping Carriage Returns and New Line Characters for Text Files \(LazySimpleSerDe\)](#) on page 542
- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 542

Changing Table File Format from ORC with the ALTER TABLE Command Not Supported in CDH 6

Changing the table file format from ORC to another file format with the ALTER TABLE command is not supported in CDH 6 (it returns an error).

UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting

Prior to this change, Hive performed implicit casts when data types from different type groups were specified in queries that use UNION ALL. For example, before CDH 6.0, if you had the two following tables:

Table "one"

one.col_1	one.col_2	one.col_3
21	hello_all	b

Where col_1 datatype is int, col_2 datatype is string, and col_3 datatype is char(1).

Table "two"

two.col_4	two.col_5	two.col_6
75.0	abcde	45

Where `col_4` datatype is `double`, `col_5` datatype is `varchar(5)`, and `col_6` datatype is `int`.

And you ran the following `UNION ALL` query against these two tables:

```
SELECT * FROM one UNION ALL SELECT col_4 AS col_1, col_5 AS col_2, col_6 AS col_3 FROM two;
```

You received the following result set:

_u1.col_1	_u1.col_2	_u1.col_3
75.0	abcde	4
21.0	hello	b

Note that this statement implicitly casts the values from table `one` with the following errors resulting in data loss:

- `one.col_1` is cast to a `double` datatype
- `one.col_2` is cast to a `varchar(5)` datatype, which truncates the original value from `hello_all` to `hello`
- `one.col_3` is cast to a `char(1)` datatype, which truncates the original value from 45 to 4

In CDH 6.0, no implicit cast is performed across different type groups. For example, `STRING`, `CHAR`, and `VARCHAR` are in one type group, and `INT`, `BIGINT`, and `DECIMAL` are in another type group, and so on. So, in CDH 6.0 and later, the above query that uses `UNION ALL`, returns an exception for the columns that contain datatypes that are not part of a type group. In CDH 6.0 and later, Hive performs the implicit cast only *within* type groups and not *across* different type groups. For more information, see [HIVE-14251](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
- The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
- The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

OFFLINE and NO_DROP Options Removed from Table and Partition DDL

Support for Hive table and partition protection options have been removed in CDH 6.0, which includes removal of the following functionality:

- Support has been removed for:

- `ENABLE | DISABLE NO_DROP [CASCADE]`

- ENABLE | DISABLE OFFLINE
 - ALTER TABLE ... IGNORE PROTECTION
- The following support has also been removed from the `HiveMetastoreClient` class:

The `ignoreProtection` parameter has been removed from the `dropPartitions` methods in the `IMetaStoreClient` interface.

For more information, see [HIVE-11145](#).

Cloudera recommends that you use Apache Sentry to replace most of this functionality. Although Sentry governs permissions on `ALTER TABLE`, it does not include permissions that are specific to a partition. See [Authorization Privilege Model for Hive and Impala](#) and [Configuring the Sentry Service](#).

DESCRIBE Query Syntax Change

In CDH 6.0 syntax has changed for `DESCRIBE` queries as follows:

- `DESCRIBE` queries where the column name is separated by the table name using a period is no longer supported:

```
DESCRIBE testTable.testColumn;
```

Instead, the table name and column name must be separated with a space:

```
DESCRIBE testTable testColumn;
```

- The `partition_spec` must appear *after* the table name, but *before* the optional column name:

```
DESCRIBE default.testTable PARTITION (part_col = 100) testColumn;
```

For more details, see the [Apache wiki](#) and [HIVE-12184](#).

CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names

In CDH 6.0, `CREATE TABLE` statements fail if any of the specified column names contain a period or a colon. For more information, see [HIVE-10120](#) and the [Apache wiki](#).

Reserved and Non-Reserved Keyword Changes in HiveQL

Hive reserved and non-reserved keywords have changed in CDH 6.0. *Reserved keywords* cannot be used as table or column names unless they are enclosed with back ticks (for example, `data`). *Non-reserved keywords* can be used as table or column names without enclosing them with back ticks. Non-reserved keywords have proscribed meanings in HiveQL, but can still be used as table or column names. For more information about the changes to reserved and non-reserved words listed below, see [HIVE-6617](#) and [HIVE-14872](#).

In CDH 6.0, the following changes have been introduced to Hive reserved and non-reserved keywords and are not backwards compatible:

- [Hive New Reserved Keywords Added in CDH 6.0](#) on page 535
- [Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0](#) on page 536
- [Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0](#) on page 536
- [Hive New Non-Reserved Keywords Added in CDH 6.0](#) on page 536
- [Hive Non-Reserved Keyword Removed in CDH 6.0](#) on page 537

Hive New Reserved Keywords Added in CDH 6.0

The following table contains new reserved keywords that have been added:

COMMIT	CONSTRAINT	DEC	EXCEPT
FOREIGN	INTERVAL	MERGE	NUMERIC
ONLY	PRIMARY	REFERENCES	ROLLBACK
START			

Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0

The following table contains non-reserved keywords that have been converted to be reserved keywords:

ALL	ALTER	ARRAY	AS
AUTHORIZATION	BETWEEN	BIGINT	BINARY
BOOLEAN	BOTH	BY	CREATE
CUBE	CURSOR	DATE	DECIMAL
DOUBLE	DELETE	DESCRIBE	DROP
EXISTS	EXTERNAL	FALSE	FETCH
FLOAT	FOR	FULL	GRANT
GROUP	GROUPING	IMPORT	IN
INT	INNER	INSERT	INTERSECT
INTO	IS	LATERAL	LEFT
LIKE	LOCAL	NONE	NULL
OF	ORDER	OUT	OUTER
PARTITION	PERCENT	PROCEDURE	RANGE
READS	REGEXP	REVOKE	RIGHT
RLIKE	ROLLUP	ROW	ROWS
SET	SMALLINT	TABLE	TIMESTAMP
TO	TRIGGER	TRUNCATE	UNION
UPDATE	USER	USING	VALUES
WITH	TRUE		

Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0

The following table contains reserved keywords that have been converted to be non-reserved keywords:

CURRENT_DATE	CURRENT_TIMESTAMP	HOLD_DDLTIME	IGNORE
NO_DROP	OFFLINE	PROTECTION	READONLY

Hive New Non-Reserved Keywords Added in CDH 6.0

The following table contains new non-reserved keywords that have been added:

ABORT	AUTOCOMMIT	CACHE	DAY
DAYOFWEEK	DAYS	DETAIL	DUMP
EXPRESSION	HOUR	HOURS	ISOLATION
KEY	LAST	LEVEL	MATCHED

MINUTE	MINUTES	MONTH	MONTHS
NORELY	NOVALIDATE	NULLS	OFFSET
OPERATOR	RELY	SECOND	SECONDS
SNAPSHOT	STATUS	SUMMARY	TRANSACTION
VALIDATE	VECTORIZATION	VIEWS	WAIT
WORK	WRITE	YEAR	YEARS

Hive Non-Reserved Keyword Removed in CDH 6.0

The following non-reserved keyword has been removed:

DEFAULT

Apache Hive API Changes in CDH 6.0.0

The following changes have been introduced to the Hive API in CDH 6.0, and are not backwards compatible:

- [AddPartitionMessage.getPartitions\(\) Can Return NULL](#) on page 537
- [DropPartitionEvent and PreDropPartitionEvent Class Changes](#) on page 537
- [GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date](#) on page 537
- [GenericUDF.getConstantLongValue Has Been Removed](#) on page 537
- [Increased Width of Hive Metastore Configuration Columns](#) on page 537

AddPartitionMessage.getPartitions() Can Return NULL

The `getPartitions()` method has been removed from the `AddPartitionEvent` class in the `org.apache.hadoop.hive.metastore.events` interface. It was removed to prevent out-of-memory errors when the list of partitions is too large.

Instead use the `getPartitionIterator()` method. For more information, see [HIVE-9609](#) and the [AddPartitionEvent documentation](#).

DropPartitionEvent and PreDropPartitionEvent Class Changes

The `getPartitions()` method has been removed and replaced by the `getPartitionIterator()` method in the `DropPartitionEvent` class and the `PreDropPartitionEvent` class.

In addition, the `(Partition partition, boolean deleteData, HiveMetastore.HMSHandler handler)` constructors have been deleted from the `PreDropPartitionEvent` class. For more information, see [HIVE-9674](#) and the [PreDropPartitionEvent documentation](#).

GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date

The `getTimestampValue` method in the `GenericUDF` class now returns a `TIMESTAMP` value instead of a `DATE` value. For more information, see [HIVE-10275](#) and the [GenericUDF documentation](#).

GenericUDF.getConstantLongValue Has Been Removed

The `getConstantLongValue` method has been removed from the `GenericUDF` class. It has been noted by the community that this method is not used in Hive. For more information, see [HIVE-10710](#) and the [GenericUDF documentation](#).

Increased Width of Hive Metastore Configuration Columns

The columns used for configuration values in the Hive metastore have been increased in width, resulting in the following incompatible changes in the `org.apache.hadoop.hive.metastore.api` interface.

This change introduced an incompatible change to the `get_table_names_by_filter` method of the `ThriftHiveMetastore` class. Before this change, this method accepts a string filter, which allows clients to filter a table by its TABLEPROPERTIES value. For example:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 <> \"yellow\"";  
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 = \"yellow\"";
```

After this change, the `TABLE_PARAMS.PARAM_VALUE` column is now a `CLOB` data type. Depending on the type of database that you use (for example, MySQL, Oracle, or PostgreSQL), the semantics may have changed and operators like `=`, `<>`, and `!=` might not be supported. Refer to the documentation for your database for more information. You must use operators that are compatible with `CLOB` data types. There is no equivalent `<>` operator that is compatible with `CLOB`. So there is no equivalent operator for the above example that uses the `<>` inequality operator. The equivalent for `=` is the `LIKE` operator so you would rewrite the second example above as:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 LIKE \"yellow\"";
```

For more information, see [HIVE-12274](#).

Apache Hive Configuration Changes in CDH 6.0.0

The following configuration property changes have been introduced to Hive in CDH 6.0, and are not backwards compatible:

- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 538
- [Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters](#) on page 538
- [Hive Strict Checks Have Been Re-factored To Be More Granular](#) on page 539
- [Java XML Serialization Has Been Removed](#) on page 539
- [Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties](#) on page 540
- [HiveServer2 Impersonation Property \(`hive.server2.enable.impersonation`\) Removed](#) on page 540
- [Changed Default File Format for Storing Intermediate Query Results](#) on page 540

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on `hive.enforce.bucketing`](#) and the [topic on `hive.enforce.sorting`](#).

Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters

Directories in HDFS can contain unprintable or unsupported characters that are not visible even when you run the `hadoop fs -ls` command on the directories. When external tables are created with the `MSCK REPAIR TABLE` command, the partitions using these HDFS directories that contain unsupported characters are unusable for Hive. To avoid this, the configuration parameter `hive.msck.path.validation` has been added. This configuration property controls the behavior of the `MSCK REPAIR TABLE` command, enabling you to set whether validation checks are run on the HDFS directories when `MSCK REPAIR TABLE` is run.

The property `hive.msck.path.validation` can be set to one of the following values:

Value Name	Description
throw	Causes Hive to throw an exception when it tries to process an HDFS directory that contains unsupported characters

Value Name	Description
	with the MSCK REPAIR TABLE command. This is the default setting for <code>hive.msck.path.validation</code> .
skip	Causes Hive to skip the skip the directories that contain unsupported characters, but still repairs the others.
ignore	Causes Hive to completely skip any validation of HDFS directories when the MSCK REPAIR TABLE command is run. This setting can cause bugs because unusable partitions are created.

By default, the `hive.msck.path.validation` property is set to `throw`, which causes Hive to throw an exception when MSCK REPAIR TABLE is run and HDFS directories containing unsupported characters are encountered. To work around this, set this property to `skip` until you can repair the HDFS directories that contain unsupported characters.

To set this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the **Configuration** tab.
3. Search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting.
4. In the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting, add the **Name** of the property, the **Value** (`throw`, `skip`, or `ignore`), and a **Description** of the setting.
5. Click **Save Changes** and restart the service.

For more information, see [HIVE-10722](#).

Hive Strict Checks Have Been Re-factored To Be More Granular

Originally, the configuration property `hive.mapred.mode` was added to restrict certain types of queries from running. Now it has been broken down into more fine-grained configurations, one for each type of restricted query pattern. The configuration property `hive.mapred.mode` has been removed and replaced with the following configuration properties, which provide more granular control of Hive strict checks:

Configuration Property	Description	Default Value
<code>hive.strict.checks.bucketing</code>	When set to <code>true</code> , running LOAD DATA queries against bucketed tables is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.type.safety</code>	When set to <code>true</code> , comparing bigint to string data types or bigint to double data types is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.orderby.no.limit</code>	When set to <code>true</code> , prevents queries from being run that contain an ORDER BY clause with no LIMIT clause.	<code>false</code>
<code>hive.strict.checks.no.partition.filter</code>	When set to <code>true</code> , prevents queries from being run that scan a partitioned table but do not filter on the partition column.	<code>false</code>
<code>hive.strict.checks.cartesian.product</code>	When set to <code>true</code> , prevents queries from being run that contain a Cartesian product (also known as a cross join).	<code>false</code>

All of these properties can be set with Cloudera Manager in the following configuration settings for the Hive service:

- **Restrict LOAD Queries Against Bucketed Tables** (`hive.strict.checks.bucketing`)

- **Restrict Unsafe Data Type Comparisons** (`hive.strict.checks.type.safety`)
- **Restrict Queries with ORDER BY but no LIMIT clause** (`hive.strict.checks.orderby.no.limit`)
- **Restrict Partitioned Table Scans with no Partitioned Column Filter**
(`hive.strict.checks.no.partition.filter`)
- **Restrict Cross Joins (Cartesian Products)** (`hive.strict.checks.cartesian.product`)

For more information about these configuration properties, see [HIVE-12727](#), [HIVE-15148](#), [HIVE-18251](#), and [HIVE-18552](#).

Java XML Serialization Has Been Removed

The configuration property `hive.plan.serialization.format` has been removed. Previously, this configuration property could be set to either `javaXML` or `kryo`. Now the default is `kryo` serialization, which cannot be changed. For more information, see [HIVE-12609](#) and the [Apache wiki](#).

Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties

The configuration property `hive.groupby.orderby.position.alias`, which enabled using column position with the `GROUP BY` and the `ORDER BY` clauses has been removed and replaced with the following two configuration properties. These configuration properties enable using column position with `GROUP BY` and `ORDER BY` separately:

Configuration Property Name	Description/Default Setting	Possible Values
<code>hive.groupby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>GROUP BY</code> clauses in queries. Default Setting: <code>false</code> . This behavior is turned off by default.	<code>true</code> <code>false</code>
<code>hive.orderby.position.alias</code>	When set to <code>true</code> , specifies that columns can be referenced with their position when using <code>ORDER BY</code> clauses in queries. Default Setting: <code>true</code> . This behavior is turned on by default.	<code>true</code> <code>false</code>

For more information, see [HIVE-15797](#) and the Apache wiki entries for [configuration properties](#), [GROUP BY syntax](#), and [ORDER BY syntax](#).

HiveServer2 Impersonation Property (`hive.server2.enable.impersonation`) Removed

In earlier versions of CDH, the following two configuration properties could be used to set impersonation for HiveServer2:

- `hive.server2.enable.impersonation`
- `hive.server2.enable.doAs`

In CDH 6.0, `hive.server2.enable.impersonation` is removed. To configure impersonation for HiveServer2, use the configuration property `hive.server2.enable.doAs`. To set this property in Cloudera Manager, select the Hive service and click on the **Configuration** tab. Then search for the **HiveServer2 Enable Impersonation** setting and select the checkbox to enable HiveServer2 impersonation. This property is enabled by default in CDH 6.

For more information about this property, see the [Apache wiki documentation for HiveServer2 configuration properties](#).

Changed Default File Format for Storing Intermediate Query Results

The configuration property `hive.query.result.fileformat` controls the file format in which a query's intermediate results are stored. In CDH 6, the default setting for this property has been changed from `TextFile` to `SequenceFile`.

To change this configuration property in Cloudera Manager:

1. In the Admin Console, select the Hive service and click on the **Configuration** tab.
2. Then search for the **Hive Service Advanced Configuration Snippet (Safety Valve) for `hive-site.xml`** setting and add the following information:
 - **Name:** `hive.query.result.fileformat`
 - **Value:** Valid values are `TextFile`, `SequenceFile` (default), or `RCfile`

- **Description:** Sets the file format in which a query's intermediate results are stored.

3. After you add this information, click **Save Changes** and restart the Hive service.

For more information about this parameter, see the [Apache wiki](#).

HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes

HiveServer2 Thrift API code has been repackaged in CDH 6.0, resulting in the following changes:

- All files generated by the Thrift API for HiveServer2 have moved from the following *old* namespace:

`org.apache.hive.service.cli.thrift`

To the following *new* namespace:

`org.apache.hive.service.rpc.thrift`

- All files generated by the Thrift API for HiveServer2 have moved into a separate jar file called `service-rpc`.

As a result of these changes, all Java classes such as `TCLIService.java`, `TOpenSessionReq.java`, `TSessionHandle.java`, and `TGetSchemasReq.java` have changed locations. For more information, see [HIVE-12442](#).

Values Returned for Decimal Numbers Are Now Padded with Trailing Zeros to the Scale of the Specified Column

Decimal values that are returned in query results are now padded with trailing zeroes to match the specified scale of the corresponding column. For example, *before* this change, when Hive read a decimal column with a specified scale of 5, the value returned for zero was returned as 0. *Now*, the value returned for zero is 0.00000. For more information, see [HIVE-12063](#).

Hive Logging Framework Switched to SLF4J/Log4j 2

The logging framework for Hive has switched to [SLF4J \(Simple Logging Facade for Java\)](#) and now uses [Log4j 2](#) by default. Use of Log4j 1.x, Apache Commons Logging, and `java.util.logging` have been removed. To accommodate this change, write all Log4j configuration files to be compatible with Log4j 2.

For more information, see [HIVE-12237](#), [HIVE-11304](#), and the [Apache wiki](#).

Deprecated Parquet Java Classes Removed from Hive

The deprecated `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` have been removed from Hive because they resided outside of the `org.apache` namespace. Any existing tables that use these classes are automatically migrated to the new SerDe classes when the metastore is upgraded.

Use one of the following options for specifying the Parquet SerDe for new Hive tables:

- Specify in the `CREATE TABLE` statement that you want it stored as Parquet. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING) STORED AS PARQUET;
```

- Set the `INPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat` and set the `OUTPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat`. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING)
STORED AS
  INPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat"
  OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat";
```

For more information, see [HIVE-6757](#) and the [Apache wiki](#).

Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms

Support for JDBC, counter-based, and HBase-based statistics collection mechanisms has been removed from Hive. The following configuration properties are no longer supported:

- `hive.stats.dbclass`
- `hive.stats.retries.wait`
- `hive.stats.retries.max`
- `hive.stats.jdbc.timeout`
- `hive.stats.dbconnectionstring`
- `hive.stats.jdbcdrive`
- `hive.stats.key.prefix.reserve.length`

This change also removed the `cleanUp(String keyPrefix)` method from the [StatsAggregator interface](#).

Now all Hive statistics are collected on the default file system. For more information, see [HIVE-12164](#), [HIVE-12411](#), [HIVE-12005](#), and the [Apache wiki](#).

S3N Connector Is Removed from CDH 6.0

The [S3N connector](#), which is used to connect to the Amazon S3 file system from Hive has been removed from CDH 6.0. To connect to the S3 file system from Hive in CDH 6.0, you must now use the S3A connector. There are a number of differences between the S3N and the S3A connectors, including configuration differences. See the [Apache wiki page on integrating with Amazon Web Services](#) for details.

Migration involves making the following changes:

- Changing all metastore data containing URIs that start with `s3n://` to `s3a://`. This change is performed automatically when you upgrade the Hive metastore.
- Changing all scripts containing URIs that start with `s3n://` to `s3a://`. You must perform this change manually.

Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request

Six additional columns have been added to the `TRowSet` that is returned by the `TCLIService#GetTables` request. These columns were added to comply with the official JDBC API. For more information, see the documentation for [java.sql.DatabaseMetaData](#).

The columns added are:

Column Name	Description
REMARKS	Explanatory comment on the table.
TYPE_CAT	Types catalog.
TYPE_SCHEMA	Types schema.
TYPE_NAME	Types name.
SELF_REFERENCING_COL_NAME	Name of the designed identifier column of a typed table.
REF_GENERATION	Specifies how values in the <code>SELF_REFERENCING_COL_NAME</code> column are created.

For more information, see [HIVE-7575](#).

Support Added for Escaping Carriage Returns and New Line Characters for Text Files (LazySimpleSerDe)

Support has been added for escaping carriage returns and new line characters in text files by modifying the `LazySimpleSerDe` class. Without this change, carriage returns and new line characters are interpreted as delimiters, which causes incorrect query results.

This feature is controlled by the SerDe property `serialization.escape.crlf`. It is enabled (set to `true`) by default. If `serialization.escape.crlf` is enabled, '`r`' or '`n`' cannot be used as separators or field delimiters.

This change only affects text files and removes the `getNullString` method from the [LazySerDeParameters class](#). For more information, see [HIVE-11785](#).

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on `hive.enforce.bucketing`](#) and the [topic on `hive.enforce.sorting`](#).

Hue

There are no incompatible changes in this release.

Apache Impala

There are no incompatible changes in this release.

Apache Kafka

Incompatible Changes Introduced in CDH 6.1.0

Scala-based Client API Removed

Scala-based clients were deprecated in a previous release and are removed as of CDH 6.1.0.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are effected:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Default Behaviour Change

[KAFKA-7050](#): The default value for `request.timeout.ms` is decreased to 30 seconds. In addition, a new logic is added that makes the `JoinGroup` requests ignore this timeout.

Incompatible Changes Introduced in CDH 6.0.1

CDH 6.0.1 introduces no new incompatible changes for Kafka.

Incompatible Changes Introduced in CDH 6.0.0

Kafka is now bundled as part of CDH. The following sections describe incompatible changes between the previous, separately installed Kafka (CDK powered by Apache Kafka version 3.1) and the CDH 6.0.0 Kafka version. These changes affect clients built with CDH 6.0.0 libraries. Cloudera recommends upgrading clients to the new release; however clients built with previous versions of Kafka will continue to function.

Packaging

CDH and previous distributions of Kafka (CDK Powered by Apache Kafka) cannot coexist in the same cluster.

Deprecated Scala-based Client API and New Java Client API

Scala-based clients are deprecated in this release and will be removed in an upcoming release.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are deprecated and unsupported as of CDH 6.0.0:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Command Line Options Removed

Some command line tools are affected by the deprecation of old clients (see the previous entry). The following options have been removed and are not recognized as valid options:

- `--new-consumer`
- `--old-consumer`
- `--old-producer`

The tools affected use the new clients.

Command Line Tools Removed

The following command line tools and runnable classes are removed:

- `kafka-replay-log-producer`
- `kafka-simple-consumer-shell`
- `kafka.tools.ReplayLogProducer`
- `kafka.tools.SimpleConsumerShell`
- `kafka.tools.ExportZkOffset`
- `kafka.tools.ImportZkOffset`
- `kafka.tools.SimpleConsumerPerformance`
- `kafka.tools.UpdateOffsetsInZK`
- `kafka.tools.VerifyConsumerRebalance`
- `kafka.tools.ProducerPerformance`

Consumer API Changes

Consumer methods invoked with unassigned partitions now raise an `IllegalStateException` instead of an `IllegalArgumentException`.

Previous versions of the Consumer method `poll(long)` would wait for metadata updates regardless of timeout parameter. This behavior is expected to change in future releases; make sure your client applications include an appropriate timeout parameter and do not rely on the previous behavior.

Exception Classes Removed

The following exceptions were deprecated in a previous release and are not thrown anymore are removed:

- `GroupCoordinatorNotAvailableException`
- `GroupLoadInProgressException`
- `NotCoordinatorForGroupException`
- `kafka.common.KafkaStorageException`

Metrics Updated

Kafka consumers' per-partition metrics were changed to use tags for topic and partition rather than the metric name. For more information see [KIP-225](#).

Apache Kudu

Client Library Compatibility

- The Kudu 1.8 Java client library is API- and ABI-compatible with Kudu 1.7. Applications written against Kudu 1.7 will compile and run against the Kudu 1.8 client library and vice-versa.
- The Kudu 1.8 C++ client is API- and ABI-forward-compatible with Kudu 1.7. Applications written and compiled against the Kudu 1.7 client library will run without modification against the Kudu 1.8 client library. Applications written and compiled against the Kudu 1.8 client library will run without modification against the Kudu 1.7 client library.
- The Kudu 1.8 Python client is API-compatible with Kudu 1.7. Applications written against Kudu 1.7 will continue to run against the Kudu 1.8 client and vice-versa.

Apache Oozie

There are no incompatible changes in this release.

Apache Parquet

Packages and Group ID Renamed

As a part of the Apache incubation process, all Parquet packages and the project's group ID were renamed as follows:

Parquet Version	1.6.0 and lower (CDH 5.x)	1.7.0 and higher (CDH 6.x)
Java Package Names	parquet.*	org.apache.parquet.*
Group ID	com.twitter	org.apache.parquet

If you directly consume the Parquet API, instead of using Parquet through Hive, Impala or other CDH component, you need to update your code to reflect these changes:

Update *.java files:

Before	After
import parquet.*;	import org.apache.parquet.*;

Update pom.xml:

Before	After
<pre><dependency> <groupId> com.twitter </groupId> <version> \${parquet.version} </version> </dependency></pre>	<pre><dependency> <groupId> org.apache.parquet </groupId> <version> \${parquet.version} </version> </dependency></pre>

API Methods Removed

In Parquet 1.6, a number of API methods were removed from the `parquet.hadoop.ParquetInputSplit` class that depended on reading metadata on the client side. Metadata should be read on the task side instead.

Removed Method	New Method to Use
<code>parquet.hadoop.ParquetInputSplit.getFileSchema</code>	<code>org.apache.parquet.hadoop.api.InitContext.getFileSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getRequestedSchema</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getRequestedSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getReadSupportMetadata</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getReadSupportMetadata</code>
<code>parquet.hadoop.ParquetInputSplit.getBlocks</code>	<code>org.apache.parquet.hadoop.metadata.ParquetMetadata.getBlocks</code>
<code>parquet.hadoop.ParquetInputSplit.getExtraMetadata</code>	-

Apache Pig

The following change is introduced to Pig in CDH 6.0 and is not a backwards compatible change. You must modify your Pig scripts as described below.

Removal of the Apache DataFu Pig JAR from CDH 6

Apache DataFu Pig is a collection of user-defined functions that can be used with Pig for data mining and statistical analysis on large-scale data. The DataFu JAR was included in CDH 5, but due to very low adoption rates, the JAR was deprecated in CDH 5.9 and is being removed from CDH 6, starting with CDH 6.0. It is no longer supported.

Recommended Migration Strategy

A simple way to assess what DataFu functions you are using in your Pig scripts is to use the `grep` utility to search for occurrences of "datafu" in your code. When DataFu functions are used in Pig scripts, you must use a function definition entry that contains "datafu" like the following example:

```
define <function_name> datafu.pig... .<class_name>();
```

Use `grep` to search for the string "datafu" in your scripts and that will identify where the DataFu JAR is used.

Cloudera recommends migrating to Hive UDFs or operators wherever it is possible. However, if there are cases where it is impossible to replace DataFu functions with Hive functions, download the upstream version of the DataFu Pig libraries and place them on the node where the Pig front end is used. To preserve compatibility, use the version 1.1.0 JAR, which was the version included in CDH 5. You can download the JAR file [here](#). However, Cloudera does not support using this upstream DataFu JAR file.

Mapping DataFu UDFs to Hive UDFs

The following Hive UDFs map to DataFu UDFs and can be used instead in Pig scripts with the caveats that are listed:

Table 36: Hive Functions That Map to DataFu Functions

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
MD5 (hash)	Computes the MD5 value of a string and outputs a hex value by default.	md5	None
SHA (hash)	Computes the SHA value of a string and outputs a hex value by default.	sha/sha1	None
RandInt (random)	Generates a uniformly distributed integer between two bounds.	rand	None
VAR (stats)	Generates the variance of a set of values.	variance	None
Median (stats)	Computes the median for a sorted input bag. A special case of the Quantile function.	percentile(0.5)	See Limitations When Substituting Quantile and Median DataFu Functions on page 388.
Quantile (stats)	Computes quantiles for a sorted input bag.	percentile	See Limitations When Substituting Quantile and Median DataFu Functions on page 388.
Coalesce (util)	Returns the first non-null value from a tuple, like COALESCE in SQL.	coalesce	None
InUDF (util)	Similar to the SQL IN function, this function provides a convenient way to filter using a logical disjunction over many values.	IN	None

For more information about using Hive UDFs, see
https://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_hive_udf.html.

Limitations When Substituting Quantile and Median DataFu Functions

With the exception of Median and Quantile all Hive functions specified in the above table should work as expected in Pig scripts. Median extends Quantile in DataFu functions and the equivalent Hive functions have a similar relationship. However, there is an important difference in how you use percentile and how you use Quantile. The differences are summarized in the following table:

Table 37: Differences Between Usage of DataFu 'Quantile' and Hive 'percentile'

Function:	Quantile	percentile
Input must be sorted:	Yes	No
Nulls are allowed in input:	No	Yes
Examples:	<pre> register datafu-1.2.0.jar; define median datafu.pig.stats.Quantile('0.5'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B = group A by name; C = foreach B { sorted = order A by n; generate group, flatten (median(sorted.n)); } </pre>	<pre> define percentile HiveUDAF ('percentile'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B2 = foreach A generate name, n, 0.5 as perc; C2 = GROUP B2 by name; D2 = FOREACH C2 generate group, percentile (B2. (n, perc)); </pre>

Although DataFu StreamingQuantile and StreamingMedian might appear to match Hive's percentile_approx function, Pig cannot consume percentile_approx.

DataFu Functions with No Hive Function or Operator Equivalent

The following general limitations apply when mapping DataFu UDFs to Hive UDFs:

- Many DataFu functions operate on a custom Pig data structure called a *bag*. No Hive UDFs can operate on Pig bags, so there are no equivalents for these DataFu functions.
- Some DataFu functions are custom functions that do not have Hive UDF equivalents. For example, the DataFu functions that calculate geographic distances, run the PageRank algorithm, or that do sampling. There are no equivalent Hive UDFs for these DataFu functions either.

Table 38: DataFu Functions with No Hive UDF Equivalent

AppendToBag (bags)	AssertUDF (util)	BagConcat (bags)
BagGroup (bags)	BagLeftOuterJoin (bags)	BagSplit (bags)
BoolToInt (util)	CountEach (bags)	DistinctBy (bags)
EmptyBagToNull (bags)	EmptyBagToNullFields (bags)	Enumerate (bags)
FirstTupleFromBag (bags)	HaversineDistInMiles (geo)	IntToBool (util)
MarkovPairs	NullToEmptyBag (bags)	PageRank (linkanalysis)
PrependToBag (bags)	ReservoirSample (sampling)*	ReverseEnumerate (bags)
SampleByKey (sampling)*	SessionCount (sessions)	Sessionize (sessions)

SetIntersect (sets)	SetUnion (sets)	SimpleRandomSample (sampling)*
TransposeTupleToBag (util)	UnorderedPairs (bags)	UserAgentClassify (urls)
WeightedSample (sampling)*	WilsonBinConf (stats)	—

* These DataFu functions might be replaced with TABLESAMPLE in HiveQL. See the [Apache Hive wiki](#).

Cloudera Search

The following changes are introduced in CDH 6.1

In CDH 6.1 Cloudera Search is rebased on Apache Solr 7.4.

Deprecations

- Enabling/disabling autoAddReplicas cluster-wide with the API is deprecated. Use suspend/resume trigger APIs with name=".auto_add_replicas" instead.
- In the ReplicationHandler, the master.commitReserveDuration sub-element is deprecated. Configure a direct commitReserveDuration element instead for use in all modes (leader, follower, cloud).

Removals

- The old Leader-In-Recovery implementation (implemented in Solr 4.9) has been removed and replaced. Solr supports rolling upgrades from old 7.x versions of Solr to future 7.x releases until the last release of the 7.x major version. This means that to upgrade to Solr 8, you will have to be on Solr 7.3 or higher.
- The throttling mechanism used to limit the rate of processed autoscaling events has been removed. This deprecates the actionThrottlePeriodSeconds setting in the set-properties command of Autoscaling API. Use the triggerCooldownPeriodSeconds parameter to pause event processing.
- The RunExecutableListener event listener was removed for security reasons. If you want to listen to events caused by updates and commits, or you want to optimize, write your own listener as native Java class as part of a Solr plugin.

For more information see the [Apache Solr 7.4 Release Notes](#).

The following changes are introduced in CDH 6.0

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has many incompatibilities with the 4.10 version of Apache Solr used in recent CDH 5 releases, such as the following:

- Solr 7 uses a managed schema by default. Generating an instance directory no longer generates schema.xml. For instructions on switching to a managed schema, see [Switching from schema.xml to Managed Schema](#) in *Apache Solr Reference Guide*.
- Creating a collection using solrctl collection --create without specifying the -c <configName> parameter now uses a default configuration set (named _default) instead of a configuration set with the same name as the collection. To avoid this, always specify the -c <configName> parameter when creating new collections.

For the full list of changes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

Apache Sentry contains the following incompatible change in CDH 6.0.0:

- Sentry no longer supports policy file authorization. You must migrate policy files to the database-backed Sentry service before you upgrade to CDH 6.0.0 unless you are using Sentry policy files for Solr. If you are using Sentry policy files for Solr, you must migrate to the database-backed Sentry service after you upgrade.

For information about migrating policy files before you upgrade, see [Migrating from Sentry Policy Files to the Sentry Service](#). For information about migrating policy files for Solr after you upgrade, see [Migrating Sentry Privileges for Solr After Upgrading to CDH 6](#).

Apache Spark

The following sections describe changes in Spark support in CDH 6 that might require special handling during upgrades, or code changes within existing applications.

- All Spark applications built against Spark 1.6 in CDH 5 must be rebuilt against Spark 2.x in CDH 6.
- Spark 2 in CDH 6 works with Java 8, not Java 7. If this change produces any Java code incompatibilities, update your Java code and rebuild the application.
- Spark 2 in CDH 6 works with Scala 2.11, not Scala 2.10. If this change produces any Scala code incompatibilities, update your Scala code and rebuild the application.
- `HiveContext` and `SQLContext` have been removed, although those variables still work for backward compatibility. Use the `SparkSession` object to replace both of these handles.
- `DataFrames` have been removed from the Scala API. `DataFrame` is now a special case of `Dataset`.

Since compile-time type-safety in Python and R is not a language feature, the concept of `Dataset` does not apply to these languages' APIs. Instead, `DataFrame` remains the primary programming abstraction.

- Spark 2.0 and higher do not use an assembly JAR for standalone applications.
- If you have event logs created in CDH 5.3 or lower, you cannot read those logs using Spark in CDH 6.0 or higher.

Apache Sqoop

CDH 6.1.0 introduces no new incompatible changes for Apache Sqoop.

The following changes are introduced in CDH 6.0, and are not backwards compatible:

- All classes in `com.cloudera.sqoop` packages have been removed in CDH 6.0. Use the corresponding classes from `org.apache.sqoop` packages. For example, use `org.apache.sqoop.SqoopOptions` instead of `com.cloudera.sqoop.SqoopOptions`.



Note: This change only affects customers who build their own application on top of Sqoop classes. Sqoop CLI users are not affected.

- Because of changes introduced in the Sqoop metastore logic, the metastore database created by Sqoop CDH 6 cannot be used by earlier versions. The metastore database created by Sqoop CDH 5 can be used by both Sqoop CDH 5 and Sqoop CDH 6.

Require an explicit option to be specified with `--split-by` for a String column

Using the `--split-by` option with a CHAR or VARCHAR column does not always work properly, so Sqoop now requires the user to set the `org.apache.sqoop.splitter.allow_text_splitter` property to `true` to confirm that they are aware of this risk.

Example:

```
sqoop import -Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect $MYCONN
--username $MYUSER --password $MYPWD --table "test_table" --split-by "string_column"
```

For more information, see [SQOOP-2910](#).

Make Sqoop fail if user specifies `--direct` connector when it is not available

The `--direct` option is only supported with the following databases: MySQL, PostgreSQL, Oracle, Netezza.

In earlier releases, Sqoop silently ignored this option if it was specified for other databases, but it now throws an error.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "direct_import" --direct
```

The command fails with the following error message:

```
Was called with the --direct option, but no direct connector available.
```

For more information, see [SQOOP-2913](#).

Sqoop does not allow --as-parquetfile with hcatalog jobs or when hive import with create-hive-table is used

The --create-hive-table option is not supported when the user imports into Hive in Parquet format. Earlier this option was silently ignored, and the data was imported even if the Hive table existed. Sqoop will now fail if the --create-hive-table option is used with the --as-parquetfile option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table" --hive-import --as-parquetfile --create-hive-table
```

The command fails with the following error message:

```
Hive import and create hive table is not compatible with importing into ParquetFile format.
```

For more information, see [SQOOP-3010](#).

Create fail fast for export with --hcatalog-table <HIVE_VIEW>

Importing into and exporting from a Hive view using HCatalog is not supported by Sqoop. A fail fast check was introduced so that now Sqoop throws a descriptive error message if the user specified a Hive view in the value of the --hcatalog-table option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table" --hcatalog-table "test_view"
```

The command fails with the following error message:

```
Reads/Writes from and to Views are not supported by HCatalog
```

For more information, see [SQOOP-3027](#).

Simplify Unicode character support in source files

Simplify Unicode character support in source files (introduced by [SQOOP-3074](#)) by defining explicit locales instead of using EscapeUtils. The Java source files generated by Sqoop will be encoded in UTF-8 format.

For more information, see [SQOOP-3075](#).

Columns added to MySql after initial Sqoop import, export back to table with same schema fails

If we export from HDFS to an RDBMS table and the file on HDFS has no value for some of the columns defined in the table, Sqoop will use the values of --input-null-string and --input-null-non-string options. Earlier this scenario was not supported and Sqoop failed.

For more information, see [SQOOP-3158](#).

Sqoop fails if the user tries to encode a null value when using --direct connector and a MySQL database

The MySQL direct connector does not support the `--null-string`, `--null-non-string`, `--input-null-string`, and `--input-null-non-string` options. These options were silently ignored earlier, but Sqoop now throws an error if these options are used with MySQL direct imports and exports.

For more information, see [SQOOP-3206](#).

Apache Zookeeper

There are no incompatible changes in this release.

Timezone Names Unsupported in Impala in CDH 6.1.0

The following table lists the time zone names / aliases no longer supported in Impala along with the canonical names you can use to replace the unsupported aliases with.

Deprecated	Replace with
ACDT	Australia/South
ACST	Australia/Darwin
ACWST	Australia/Eucla
ADT	America/Thule
AEDT	Australia/Sydney
AEST	Australia/Sydney
AFT	Asia/Kabul
AKDT	America/Anchorage
AKST	America/Anchorage
ALMT	Asia/Almaty
AMST	America/Cuiaba
AMT	Asia/Yerevan
ANAT	Asia/Anadyr
AQTT	Asia/Aqtau
AWST	Australia/West
AZOST	Atlantic/Azores
AZOT	Atlantic/Azores
AZST	Asia/Baku
AZT	Asia/Baku
Acre Time	Brazil/Acre
Afghanistan Time	Asia/Kabul
Alaska Daylight Time	America/Anchorage
Alaska Standard Time	America/Anchorage
Alma-Ata Time	Asia/Almaty
Amazon Summer Time	America/Cuiaba
Amazon Time	Brazil/West
Anadyr Time	Asia/Anadyr
Aqtau Time	Asia/Aqtau

Aqtobe Time	Asia/Aqtobe
Arabia Standard Time	Asia/Aden
Argentine Time	America/Argentina/Buenos_Aires
Armenia Time	Asia/Yerevan
Asia/Riyadh87	-
Asia/Riyadh88	-
Asia/Riyadh89	-
Atlantic Daylight Time	America/Thule
Atlantic Standard Time	America/Puerto_Rico
Australian Central Daylight Time (South Australia)	Australia/South
Australian Central Daylight Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Standard Time (Northern Territory)	Australia/Darwin
Australian Central Standard Time (South Australia)	Australia/South
Australian Central Standard Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Western Standard Time	Australia/Eucla
Australian Eastern Daylight Time (New South Wales)	Australia/Sydney
Australian Eastern Daylight Time (Tasmania)	Australia/Hobart
Australian Eastern Daylight Time (Victoria)	Australia/Victoria
Australian Eastern Standard Time (New South Wales)	Australia/Sydney
Australian Eastern Standard Time (Queensland)	Australia/Brisbane
Australian Eastern Standard Time (Tasmania)	Australia/Hobart
Australian Eastern Standard Time (Victoria)	Australia/Victoria
Australian Western Standard Time	Australia/West
Azerbaijan Summer Time	Asia/Baku
Azerbaijan Time	Asia/Baku
Azores Summer Time	Atlantic/Azores
Azores Time	Atlantic/Azores
BDT	Asia/Dhaka
BNT	Asia/Brunei
BOT	America/La_Paz
BRST	America/Sao_Paulo
BRT	America/Sao_Paulo
BTT	Asia/Thimbu
Bangladesh Time	Asia/Dhaka
Bhutan Time	Asia/Thimbu

Bolivia Time	America/La_Paz
Bougainville Standard Time	Pacific/Bougainville
Brasilia Summer Time	America/Sao_Paulo
Brasilia Time	America/Sao_Paulo
British Summer Time	GB
Brunei Time	Asia/Brunei
CCT	Indian/Cocos
CDT	America/Chicago
CEST	CET
CHADT	NZ-CHAT
CHAST	NZ-CHAT
CHOT	Asia/Choibalsan
CHUT	Pacific/Yap
CKT	Pacific/Rarotonga
CLST	America/Santiago
CLT	America/Santiago
COT	America/Bogota
CVT	Atlantic/Cape_Verde
CXT	Indian/Christmas
Canada/East-Saskatchewan	America/Regina
Cape Verde Time	Atlantic/Cape_Verde
Central African Time	Africa/Harare
Central Daylight Time	America/Chicago
Central European Summer Time	CET
Central European Time	CET
Central Indonesia Time	Asia/Makassar
Central Standard Time	America/Chicago
ChST	Pacific/Guam
Chamorro Standard Time	Pacific/Guam
Chatham Daylight Time	NZ-CHAT
Chatham Standard Time	NZ-CHAT
Chile Summer Time	America/Santiago
Chile Time	America/Santiago
China Standard Time	Asia/Shanghai
Choibalsan Time	Asia/Choibalsan
Christmas Island Time	Indian/Christmas

Chuuk Time	Pacific/Yap
Cocos Islands Time	Indian/Cocos
Colombia Time	America/Bogota
Cook Is. Time	Pacific/Rarotonga
Coordinated Universal Time	UCT
Cuba Daylight Time	Cuba
Cuba Standard Time	Cuba
DAVT	Antarctica/Davis
DDUT	Antarctica/DumontDUrville
Davis Time	Antarctica/Davis
Dumont-d'Urville Time	Antarctica/DumontDUrville
EASST	Pacific/Easter
EAST	Pacific/Easter
EDT	America/Indiana/Indianapolis
EEST	Africa/Cairo
EGST	America/Scoresbysund
EGT	America/Scoresbysund
East Indonesia Time	Asia/Jayapura
Easter Is. Summer Time	Pacific/Easter
Easter Is. Time	Pacific/Easter
Eastern African Time	Africa/Addis_Ababa
Eastern Daylight Time	America/Indiana/Indianapolis
Eastern European Summer Time	Africa/Cairo
Eastern European Time	Africa/Cairo
Eastern Greenland Summer Time	America/Scoresbysund
Eastern Greenland Time	America/Scoresbysund
Eastern Standard Time	EST
Ecuador Time	America/Guayaquil
FJST	Pacific/Fiji
FJT	Pacific/Fiji
FKT	Atlantic/Stanley
FNT	America/Noronha
Falkland Is. Time	Atlantic/Stanley
Fernando de Noronha Time	America/Noronha
Fiji Summer Time	Pacific/Fiji
Fiji Time	Pacific/Fiji

French Guiana Time	America/Cayenne
French Southern & Antarctic Lands Time	Indian/Kerguelen
GALT	Pacific/Galapagos
GAMT	Pacific/Gambier
GET	Asia/Tbilisi
GFT	America/Cayenne
GILT	Pacific/Tarawa
GMT+00:00	GMT0
GMT+01:00	Etc/GMT-1
GMT+02:00	Etc/GMT-2
GMT+03:00	Etc/GMT-3
GMT+04:00	Etc/GMT-4
GMT+05:00	Etc/GMT-5
GMT+06:00	Etc/GMT-6
GMT+07:00	Etc/GMT-7
GMT+08:00	Etc/GMT-8
GMT+09:00	Etc/GMT-9
GMT+10:00	Etc/GMT-10
GMT+11:00	Etc/GMT-11
GMT+12:00	Etc/GMT-12
GMT+13:00	Etc/GMT-13
GMT+14:00	Etc/GMT-14
GMT-01:00	Etc/GMT+1
GMT-02:00	Etc/GMT+2
GMT-03:00	Etc/GMT+3
GMT-04:00	Etc/GMT+4
GMT-05:00	Etc/GMT+5
GMT-06:00	Etc/GMT+6
GMT-07:00	Etc/GMT+7
GMT-08:00	Etc/GMT+8
GMT-09:00	Etc/GMT+9
GMT-10:00	Etc/GMT+10
GMT-11:00	Etc/GMT+11
GMT-12:00	Etc/GMT+12
GST	Asia/Dubai
GYT	America/Guyana

Galapagos Time	Pacific/Galapagos
Gambier Time	Pacific/Gambier
Georgia Time	Asia/Tbilisi
Ghana Mean Time	Africa/Accra
Gilbert Is. Time	Pacific/Tarawa
Greenwich Mean Tim	GB
Gulf Standard Time	Asia/Dubai
Guyana Time	America/Guyana
HADT	US/Aleutian
HAST	US/Aleutian
HKT	Hongkong
HOVT	Asia/Hovd
Hawaii Standard Time	HST
Hawaii-Aleutian Daylight Time	US/Aleutian
Hawaii-Aleutian Standard Time	US/Aleutian
Hong Kong Time	Hongkong
Hovd Time	Asia/Hovd
ICT	Asia/Ho_Chi_Minh
IDT	Israel
IOT	Indian/Chagos
IRDT	Iran
IRKT	Asia/Chita
IRST	Iran
India Standard Time	Asia/Kolkata
Indian Ocean Territory Time	Indian/Chagos
Indochina Time	Asia/Ho_Chi_Minh
Iran Daylight Time	Iran
Iran Standard Time	Iran
Irish Summer Time	Eire
Irkutsk Time	Asia/Chita
Israel Daylight Time	Israel
Israel Standard Time	Israel
Japan Standard Time	Asia/Tokyo
KGT	Asia/Bishkek
KOST	Pacific/Kosrae
KRAT	Asia/Krasnoyarsk

KST	ROK
Khandyga Time	Asia/Khandyga
Kirgizstan Time	Asia/Bishkek
Korea Standard Time	ROK
Kosrae Time	Pacific/Kosrae
Krasnoyarsk Time	Asia/Krasnoyarsk
LHDT	Australia/LHI
LHST	Australia/LHI
LINT	Pacific/Kiritimati
Line Is. Time	Pacific/Kiritimati
Lord Howe Daylight Time	Australia/LHI
Lord Howe Standard Time	Australia/LHI
MAGT	Asia/Magadan
MART	Pacific/Marquesas
MAWT	Antarctica/Mawson
MDT	Navajo
MEST	MET
MHT	Kwajalein
MIST	Antarctica/Macquarie
MMT	Asia/Rangoon
MSK	W-SU
MUT	Indian/Mauritius
MVT	Indian/Maldives
MYT	Asia/Kuching
Macquarie Island Standard Time	Antarctica/Macquarie
Magadan Time	Asia/Magadan
Malaysia Time	Asia/Kuching
Maldives Time	Indian/Maldives
Marquesas Time	Pacific/Marquesas
Marshall Islands Time	Kwajalein
Mauritius Time	Indian/Mauritius
Mawson Time	Antarctica/Mawson
Middle Europe Summer Time	MET
Middle Europe Time	MET
Mideast/Riyadh87	-
Mideast/Riyadh88	-

Mideast/Riyadh89	-
Moscow Standard Time	W-SU
Mountain Daylight Time	Navajo
Mountain Standard Time	MST
Myanmar Time	Asia/Rangoon
NCT	Pacific/Noumea
NDT	America/St_Johns
NFT	Pacific/Norfolk
NOVT	Asia/Novosibirsk
NPT	Asia/Katmandu
NRT	Pacific/Nauru
NUT	Pacific/Niue
NZDT	NZ
NZST	NZ
Nauru Time	Pacific/Nauru
Nepal Time	Asia/Katmandu
New Caledonia Time	Pacific/Noumea
New Zealand Daylight Time	NZ
New Zealand Standard Time	NZ
Newfoundland Daylight Time	America/St_Johns
Newfoundland Standard Time	America/St_Johns
Niue Time	Pacific/Niue
Norfolk Time	Pacific/Norfolk
Novosibirsk Time	Asia/Novosibirsk
OMST	Asia/Omsk
ORAT	Asia/Oral
Omsk Time	Asia/Omsk
Oral Time	Asia/Oral
PDT	America/Los_Angeles
PET	America/Lima
PETT	Asia/Kamchatka
PGT	Pacific/Port_Moresby
PHOT	Pacific/Enderbury
PHT	Asia/Manila
PKT	Asia/Karachi
PMDT	America/Miquelon

PMST	America/Miquelon
PONT	Pacific/Ponape
PWT	Pacific/Palau
PYST	America/Asuncion
PYT	America/Asuncion
Pacific Daylight Time	America/Los_Angeles
Pacific Standard Time	America/Los_Angeles
Pakistan Time	Asia/Karachi
Palau Time	Pacific/Palau
Papua New Guinea Time	Pacific/Port_Moresby
Paraguay Summer Time	America/Asuncion
Paraguay Time	America/Asuncion
Peru Time	America/Lima
Petropavlovsk-Kamchatski Time	Asia/Kamchatka
Philippines Time	Asia/Manila
Phoenix Is. Time	Pacific/Enderbury
Pierre & Miquelon Daylight Time	America/Miquelon
Pierre & Miquelon Standard Time	America/Miquelon
Pitcairn Standard Time	Pacific/Pitcairn
Pohnpei Time	Pacific/Ponape
QYZT	Asia/Qyzylorda
Qyzylorda Time	Asia/Qyzylorda
RET	Indian/Reunion
ROTT	Antarctica/Rothera
Reunion Time	Indian/Reunion
Rothera Time	Antarctica/Rothera
SAKT	Asia/Sakhalin
SAMT	Europe/Samara
SAST	Africa/Maseru
SBT	Pacific/Guadalcanal
SCT	Indian/Mahe
SGT	Singapore
SRET	Asia/Srednekolymsk
SRT	America/Paramaribo
SYOT	Antarctica/Syowa
Sakhalin Time	Asia/Sakhalin

Samara Time	Europe/Samara
Samoa Standard Time	US/Samoa
Seychelles Time	Indian/Mahe
Singapore Time	Singapore
Solomon Is. Time	Pacific/Guadalcanal
South Africa Standard Time	Africa/Maseru
South Georgia Standard Time	Atlantic/South_Georgia
Srednekolymsk Time	Asia/Srednekolymsk
Suriname Time	America/Paramaribo
Syowa Time	Antarctica/Syowa
SystemV/AST4	America/Puerto_Rico
SystemV/AST4ADT	Canada/Atlantic
SystemV/CST6	Canada/Saskatchewan
SystemV/CST6CDT	US/Central
SystemV/EST5	America/Cayman
SystemV/EST5EDT	America/New_York
SystemV/HST10	US/Hawaii
SystemV/MST7	US/Arizona
SystemV/MST7MDT	America/Denver
SystemV/PST8	Pacific/Pitcairn
SystemV/PST8PDT	US/Pacific
SystemV/YST9	Pacific/Gambier
SystemV/YST9YDT	US/Alaska
TAHT	Pacific/Tahiti
TFT	Indian/Kerguelen
TJT	Asia/Dushanbe
TKT	Pacific/Fakaofo
TLT	Asia/Dili
TMT	Asia/Ashgabat
TOT	Pacific/Tongatapu
TVT	Pacific/Funafuti
Tahiti Time	Pacific/Tahiti
Tajikistan Time	Asia/Dushanbe
Timor-Leste Time	Asia/Dili
Tokelau Time	Pacific/Fakaofo
Tonga Time	Pacific/Tongatapu

Turkmenistan Time	Asia/Ashgabat
Tuvalu Time	Pacific/Funafuti
ULAT	Asia/Ulan_Bator
UYST	America/Montevideo
UYT	America/Montevideo
UZT	Asia/Tashkent
Ulaanbaatar Time	Asia/Ulan_Bator
Uruguay Summer Time	America/Montevideo
Uruguay Time	America/Montevideo
Ust-Nera Time	Asia/Ust-Nera
Uzbekistan Time	Asia/Tashkent
VET	America/Caracas
VLAT	Asia/Ust-Nera
VOST	Antarctica/Vostok
VUT	Pacific/Efate
Vanuatu Time	Pacific/Efate
Venezuela Time	America/Caracas
Vladivostok Time	Asia/Vladivostok
Vostok Time	Antarctica/Vostok
WAKT	Pacific/Wake
WAST	Africa/Windhoek
WAT	Africa/Lagos
WEST	WET
WFT	Pacific/Wallis
WGST	America/Godthab
WGT	America/Godthab
WIB	Asia/Jakarta
WIT	Asia/Jayapura
WITA	Asia/Makassar
WSDT	Pacific/Apia
WSST	Pacific/Apia
Wake Time	Pacific/Wake
Wallis & Futuna Time	Pacific/Wallis
West Indonesia Time	Asia/Jakarta
West Samoa Daylight Time	Pacific/Apia
West Samoa Standard Time	Pacific/Apia

Western African Summer Time	Africa/Windhoek
Western African Time	Africa/Lagos
Western European Summer Time	WET
Western European Time	WET
Western Greenland Summer Time	America/Godthab
Western Greenland Time	America/Godthab
XJT	Asia/Urumqi
Xinjiang Standard Time	Asia/Urumqi
YAKT	Asia/Yakutsk
YEKT	Asia/Yekaterinburg
Yakutsk Time	Asia/Yakutsk
Yekaterinburg Time	Asia/Yekaterinburg

Known Issues and Limitations in CDH 6.1.0

The following sections describe the known issues in CDH 6.1.0, grouped by component:

Operating System Known Issues

Known issues and workarounds related to operating systems are listed below.

Linux kernel security patch and CDH services crashes CVE-2017-10000364

After applying a recent Linux kernel security patch for [CVE-2017-1000364](#), CDH services that use the JSVC set of libraries crash with a Java Virtual Machine (JVM) error such as:

```
A fatal error has been detected by the Java Runtime Environment:
SIGBUS (0x7) at pc=0x00007fe91ef6cebc, pid=30321, tid=0x00007fe930c67700
```

Cloudera services for HDFS and Impala cannot start after applying the patch.



Important: If you have not upgraded your Linux kernel using the distribution's patch for CVE-2017-1000364, do **not** apply the patch.

Commonly used Linux distributions are shown in the table below. However, the issue affects any CDH release that runs on RHEL, CentOS, Oracle Linux, SUSE Linux, or Ubuntu and that has had the Linux kernel security patch for CVE-2017-1000364 applied.

If you have already applied the patch for your OS according to the advisories for CVE-2017-1000364, apply the kernel update that contains the fix for your operating system (some of which are listed in the table). If you cannot apply the kernel update, you can workaround the issue by increasing the Java thread stack size as detailed in the steps below.

Distribution	Advisories for CVE-2017-1000364	Advisory updates
Oracle Linux 6	ELSA-2017-1486	Oracle has fixed this problem in ELSA-2017-1723 .
Oracle Linux 7	ELSA-2017-1484	Oracle has also added the fix for Oracle Linux 7 in ELBA-2017-1674 .
RHEL 6	RHSA-2017-1486	RedHat has fixed this problem for RHEL 6, marked this as outdated and superseded by RHSA-2017-1723 .

Distribution	Advisories for CVE-2017-1000364	Advisory updates
RHEL 7	RHSA-2017-1484	RedHat has fixed this problem for RHEL 7 and has marked this patch as outdated and superseded by RHBA-2017-1674 .
SLES	CVE-2017-1000364	SUSE has also fixed this problem and the patch names are included in this same advisory.

Workaround

If you cannot apply the kernel update, you can set the Java thread stack size to `-Xss1280k` for the affected services using the appropriate Java configuration option or the environment advanced configuration snippet, as detailed below.

For role instances that have specific Java configuration options properties:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Type `java` in the search field to display Java related configuration parameters. The **Java Configuration Options for Catalog Server** property field displays. Type `-Xss1280k` in the entry field, adding to any existing settings.
4. Click **Save Changes**.
5. Navigate to the HDFS service by selecting **Clusters > HDFS**.
6. Click the **Configuration** tab.
7. Click the Scope filter **DataNode**. The **Java Configuration Options for DataNode** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
8. Click **Save Changes**.
9. Select the Scope filter **NFS Gateway**. The **Java Configuration Options for NFS Gateway** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
10. Click **Save Changes**.
11. Restart the affected roles (or configure the safety valves in next section and restart when finished with all configurations).

For role instances that do not have specific Java configuration options:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Click the Scope filter **Impala Daemon** and Category filter **Advanced**.
4. Type `impala daemon` environment in the search field to find the safety valve entry field.
5. In the **Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

6. Click **Save Changes**.
7. Click the Scope filter **Impala StateStore** and Category filter **Advanced**.
8. In the **Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

9. Click **Save Changes**.
10. Restart the affected roles.

The table below summarizes the parameters that can be set for the affected services:

Service	Settable Java Configuration Option
HDFS DataNode	Java Configuration Options for DataNode
HDFS NFS Gateway	Java Configuration Options for NFS Gateway

Service	Settable Java Configuration Option
Impala Catalog Server	Java Configuration Options for Catalog Server
Impala Daemon	Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K
Impala StateStore	Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)
	JAVA_TOOL_OPTIONS=-Xss1280K

Cloudera Issue: CDH-55771

Leap-Second Events



Note: The [next leap-second event](#) is unknown at this time. The last leap-second event occurred on December 16, 2016 at 23:59:60 UTC.

Impact: After a leap-second event, Java applications (including CDH services) using older Java and Linux kernel versions, may consume almost 100% CPU. See <https://access.redhat.com/articles/15145>.

Leap-second events are tied to the time synchronization methods of the Linux kernel, the Linux distribution and version, and the Java version used by applications running on affected kernels.

Although Java is increasingly agnostic to system clock progression (and less susceptible to a kernel's mishandling of a leap-second event), using JDK 7 or 8 should prevent issues at the CDH level (for CDH components that use the Java Virtual Machine).

Immediate action required:

(1) Ensure that the kernel is up to date.

- **RHEL6/7, CentOS 6/7** - 2.6.32-298 or higher
- **Oracle Enterprise Linux (OEL)** - Kernels built in 2013 or later
- **SLES12** - No action required.

(2) Ensure that your Java JDKs are current (especially if the kernel is not up to date and cannot be upgraded).

- **Java 8** - No action required.

(3) Ensure that your systems use either NTP or PTP synchronization.

For systems not using time synchronization, update both the OS tzdata and Java tzdata packages to the tzdata-2016g version, at a minimum. For OS tzdata package updates, contact OS support or check updated OS repositories. For Java tzdata package updates, see Oracle's [Timezone Updater Tool](#).

Cloudera Issue: CDH-44788, TSB-189*Apache Accumulo Known Issues*

There are no notable known issues in this release of Apache Accumulo.

Apache Crunch Known Issues

Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

Apache Flume Known Issues

Fast Replay does not work with encrypted File Channel

If an encrypted file channel is set to use fast replay, the replay will fail and the channel will fail to start.

Workaround: Disable fast replay for the encrypted channel by setting `use-fast-replay` to false.

Apache Issue: [FLUME-1885](#)

Apache Hadoop Known Issues

This page includes known issues and related topics, including:

Deprecated Properties

Several Hadoop and HDFS properties have been deprecated as of Hadoop 3.0 and later. For details, see [Deprecated Properties](#).

Hadoop Common

Hadoop LdapGroupsMapping does not support LDAPS for self-signed LDAP server

Hadoop LdapGroupsMapping does not work with LDAP over SSL (LDAPS) if the LDAP server certificate is self-signed. This use case is currently not supported even if Hadoop User Group Mapping LDAP TLS/SSL Enabled, Hadoop User Group Mapping LDAP TLS/SSL Truststore, and Hadoop User Group Mapping LDAP TLS/SSL Truststore Password are filled properly.

Affected Versions: All CDH versions

Apache Issue: [HADOOP-12862](#)

Cloudera Issue: CDH-37926

HDFS

Clusters running CDH 5.16.1, 6.1.0, or 6.1.1 can lose some HDFS file permissions any time the Name Node is restarted

When a cluster is upgraded to 5.16.1, 6.1.0, or 6.1.1 roles with SELECT and/or INSERT privileges on an Impala database or table will have the REFRESH privilege added as part of the upgrade process. HDFS ACLs for roles with the REFRESH privilege get set with empty permissions whenever the Name Node is restarted. This can cause any jobs or queries run by users within affected roles to fail because they will no longer be able to access affected Impala database or tables.

Products Affected: HDFS and components that access files in HDFS

Affected Versions: CDH 5.16.1, 6.1.0, 6.1.1

Users Affected: Clusters with Impala and HDFS ACLs managed by Sentry upgrading from any release to CDH 5.16.1, 6.1.0, and 6.1.1.

Severity (Low/Medium/High): High

Root Cause and Impact: The new privilege REFRESH was introduced in CDH 5.16 and 6.1 and applies to Impala databases and tables. When a cluster is upgraded to 5.16.1, 6.1.0, or 6.1.1, roles with SELECT or INSERT privileges on an Impala database or table will have the REFRESH privilege added during the upgrade.

HDFS ACLs for roles with the REFRESH privilege get set with empty permissions whenever the Name Node is restarted. The Name Node is restarted during the upgrade.

For example if a group `appdev` is in role `appdev_role` and has SELECT access to the Impala table "project" the HDFS ACLs prior to the upgrade would look similar to:

```
group: appdev
      group::r--
```

After the upgrade the HDFS ACLs will be set with no permissions and will look like this:

```
group: appdev
      group::---
```

Any jobs or queries run by users within affected roles will fail because they will no longer be able to access affected Impala database or tables. This impacts any SQL client accessing the affected databases and tables. For example, if a Hive client is used to access a table created in Impala it will also fail. Jobs accessing the files directly through HDFS, e.g. via Spark, will also be impacted.

The HDFS ACLs will get reset whenever the Name Node is restarted.

Immediate action required: If possible, do not upgrade to releases CDH 5.16.1, 6.1.0, or 6.1.1 if Impala is used and Sentry manages HDFS ACLs within your environment. Subsequent CDH releases will resolve the problem with a product fix under [SENTRY-2490](#).

If an upgrade is being considered, reach out to your account team to discuss other possibilities, and to receive additional insight into future product release schedules.

If an upgrade must be executed, contact Cloudera Support indicating the upgrade plan and why an upgrade is being executed. Options are available to assist with the upgrade if necessary.

Addressed in release/refresh/patch: Patches for 5.16.1, 6.1.0 and 6.1.1 are available for major supported operating systems. Customers are encouraged to contact Cloudera Support for a patch. The patch should be applied immediately after upgrade to any of the affected versions.

The fix for this TSB will be included in 6.1.2, 6.2.0, 5.16.2, and 5.17.0.

OIV ReverseXML processor fails

The HDFS OIV ReverseXML processor fails if the XML file contains escaped characters.

Affected Versions: CDH 6.x

Apache Issue: [HDFS-12828](#)

Cannot move encrypted files to trash

With HDFS encryption enabled, you cannot move encrypted files or directories to the trash directory.

Workaround: To remove encrypted files/directories, use the following command with the `-skipTrash` flag specified to bypass trash.

```
rm -r -skipTrash /testdir
```

Affected Versions: All CDH versions

Apache Issue: [HADOOP-10902](#)

HDFS NFS gateway and CDH installation (using packages) limitation

HDFS NFS gateway works as shipped ("out of the box") only on RHEL-compatible systems, but not on SLES or Ubuntu. Because of a bug in native versions of `portmap/rpcbind`, the HDFS NFS gateway does not work out of the box on SLES or Ubuntu systems when CDH has been installed from the command-line, using packages. It does work on supported versions of RHEL-compatible systems on which `rpcbind-0.2.0-10.el6` or later is installed, and it does work if you use Cloudera Manager to install CDH, or if you start the gateway as root.

Workarounds and caveats:

- On Red Hat and similar systems, make sure `rpcbind-0.2.0-10.el6` or later is installed.
- On SLES and Ubuntu systems, do one of the following:
 - Install CDH using Cloudera Manager; or
 - Start the NFS gateway as root; or
 - [Start the NFS gateway without using packages](#); or
 - You can use the gateway by running `rpcbind` in insecure mode, using the `-i` option, but keep in mind that this allows anyone from a remote host to bind to the portmap.

Upstream Issue: [731542](#) (Red Hat), [823364](#) (SLES)

No error when changing permission to 777 on .snapshot directory

Snapshots are read-only; running `chmod 777` on the `.snapshot`s directory does not change this, but does not produce an error (though other illegal operations do).

Affected Versions: All CDH versions

Apache Issue: [HDFS-4981](#)

Snapshot operations are not supported by ViewFileSystem

Affected Versions: All CDH versions

Snapshots do not retain directories' quotas settings

Affected Versions: All CDH versions

Apache Issue: [HDFS-4897](#)

Permissions for dfs.namenode.name.dir incorrectly set

Hadoop daemons should set permissions for the `dfs.namenode.name.dir` (or `dfs.name.dir`) directories to `drwx-----` (700), but in fact these permissions are set to the file-system default, usually `drwxr-xr-x` (755).

Workaround: Use `chmod` to set permissions to 700.

Affected Versions: All CDH versions

Apache Issue: [HDFS-2470](#)

`hadoop fsck -move` does not work in a cluster with host-based Kerberos

Workaround: Use `hadoop fsck -delete`

Affected Versions: All CDH versions

Apache Issue: None

Block report can exceed maximum RPC buffer size on some DataNodes

On a DataNode with a large number of blocks, the block report may exceed the maximum RPC buffer size.

Workaround: Increase the value `ipc.maximum.data.length` in `hdfs-site.xml`:

```
<property>
  <name>ipc.maximum.data.length</name>
  <value>268435456</value>
</property>
```

Affected Versions: All CDH versions

Apache Issue: None

MapReduce2 and YARN

The Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager.

When ResourceManager high availability is enabled the Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager. This causes the following issues in Cloudera Manager:

- If **Enable Kerberos Authentication for HTTP Web-Console** is disabled: Cloudera Manager shows statistics for the wrong server.
- If **Enable Kerberos Authentication for HTTP Web-Console** is enabled: connection from the agent to the standby fails with the `HTTPError: HTTP Error 401: Authentication required` error message. As a result, the health of the Standby Resource Manager will become bad.

Workaround: N/A

Affected Versions: CDH 6.0.x, CDH 6.1.0

Fixed Version: CDH 6.1.1

Cloudera Issue: CDH-76040

YARN's Continuous Scheduling can cause slowness in Oozie

When Continuous Scheduling is enabled in Yarn, this can cause slowness in Oozie due to long delays in communicating with Yarn. In Cloudera Manager 5.9.0 and higher, Enable Fair Scheduler Continuous Scheduler is turned off by default.

Workaround: Turn off Enable Fair Scheduler Continuous Scheduling in Cloudera Manager YARN Configuration. To keep equivalent benefits of this feature, turn on Fair Scheduler Assign Multiple Tasks.

Affected Versions: All CDH versions

Cloudera Issue: CDH-60788

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

Affected Versions: All CDH versions

Apache Issue: None

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Routable IP address required by ResourceManager

ResourceManager requires routable `host : port` addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard `0.0.0.0` address.

Workaround: Set the address, in the form `host : port`, either in the client-side configuration, or on the command line when you submit the job.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-6808

Amazon S3 copy may time out

The Amazon S3 filesystem does not support renaming files, and performs a copy operation instead. If the file to be moved is very large, the operation can time out because S3 does not report progress during the operation.

Workaround: Use `-Dmapred.task.timeout=15000000` to increase the MR task timeout.

Affected Versions: All CDH versions

Apache Issue: [MAPREDUCE-972](#)

Cloudera Issue: CDH-17955

Apache HBase Known Issues

IOException from Timeouts

CDH 5.12.0 includes the fix [HBASE-16604](#), where the internal scanner that retries in case of IOException from timeouts could potentially miss data. Java clients were properly updated to account for the new behavior, but thrift clients will now see exceptions where the previous missing data would be.

Workaround: Create a new scanner and retry the operation when encountering this issue.

IntegrationTestReplication fails if replication does not finish before the verify phase begins

Cloudera Enterprise 6 Release Guide

During `IntegrationTestReplication`, if the `verify` phase starts before the `replication` phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the `-t` flag to set the timeout value before starting verification.

Cloudera Issue: None.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

`ExportSnapshot` or `DistCp` operations may fail on the Amazon `s3a://` protocol

`ExportSnapshot` or `DistCp` operations may fail on AWS when using certain JDK 8 versions, due to an incompatibility between the AWS Java SDK 1.9.x and the `joda-time` date-parsing module.

Workaround: Use `joda-time 2.8.1` or higher, which is included in AWS Java SDK 1.10.1 or higher.

Cloudera Issue: None.

An operating-system level tuning issue in RHEL7 causes 30% latency regressions

Workaround: Avoid using RHEL 7 if you have a latency-critical workload. For a cached workload, consider tuning the C-state (power-saving) behavior of your CPUs.

Cloudera Issue: CDH-32642

Severity: Medium

A RegionServer under extreme duress due to back-to-back garbage collection combined with heavy load on HDFS can lock up while attempting to append to the WAL

The RegionServer appears operational to ZooKeeper, and continues to host regions, but cannot complete any writes. The most obvious symptom is that all writes to regions on this RegionServer time out, and the RegionServer log shows no progress other than queuing of flushes that never complete. Log messages such as the following may occur:

```
124028 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.Sleeper: We slept  
42911ms instead of 3000ms,  
this is likely due to a long garbage collecting pause and it's usually bad, see  
http://hbase.%  
124029 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.JvmPauseMonitor: Detected  
pause in JVM or  
host machine (eg GC): pause of approximately 41110ms
```

```
1806 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2734 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#  
1807 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2963 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
```

Apache Issue: [HBASE-14374](#)

Workaround: Restart the RegionServer. To avoid the problem, adjust garbage-collection settings, give the RegionServer more RAM, and reduce the load on HDFS.

Export to Azure Blob Storage (the `wasb://` or `wasbs://` protocol) is not supported

CDH 5.3 and higher supports Azure Blob Storage for some applications. However, a null pointer exception occurs when you specify a `wasb://` or `wasbs://` location in the `--copy-to` option of the `ExportSnapshot` command or as the output directory (the second positional argument) of the `Export` command.

Workaround: None.

Apache Issue: [HADOOP-12717](#)

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a Delete Table fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.
- If hbaseAdmin.modifyTable() is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The postOperation is implemented only for postDeleteColumn().
- If Create Table fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: None

Apache Issue: [HBASE-6992](#)

Apache Hive/HCatalog/Hive on Spark Known Issues

This topic contains:

- [Hive Known Issues](#)
- [HCatalog Known Issues](#)
- [Hive on Spark Known Issues](#)

Hive Known Issues

Hive Jobs Are Submitted to a Single Queue When Sentry is Deployed

Hive jobs are not submitted into the correct YARN queue when Hive is using Sentry because Hive does not use the YARN API to resolve the user or group of the job's original submitter. This causes the job to be placed in a queue using the placement rules based on the Hive user. The HiveServer2 fair scheduler queue mapping used for "non-impersonation" mode does not handle the primary-secondary queue mappings correctly.

Affected Version: CDH 6.0.0 and higher

Workaround: If you are a Hive and Sentry user, do not upgrade to CDH 6.0.0 or 6.0.1. This issue will be fixed as soon as possible. If you must use Hive and Sentry in CDH 6.0.0 or 6.0.1, see [YARN Dynamic Resource Pools Do Not Work with Hive When Sentry Is Enabled](#) for additional workarounds.

When vectorization is enabled on any file type (ORC, Parquet) queries that divide by zero using the modulo operator (%) return an error

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query divides by zero using the modulo operator (%), it returns the following error: Arithmetic exception [divide by] 0. For example, if you run the following query this issue is triggered: `SELECT 100 % column_c1 FROM table_t1;` and the value in column_c1 is zero. The divide operator (/) is not affected by this issue.

Workaround: Disable vectorization for the query that is triggering this at either the session level by using the SET statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-19564](#)

Cloudera Issue: CDH-71211

When vectorization is enabled for Hive on any file type (ORC, Parquet) queries that perform comparisons in the SELECT clause on large values in columns with the data type of BIGINT might return wrong results

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query performs a comparison operation between very large values in columns that are BIGINT data types in the SELECT clause of the query, incorrect results might be returned. Comparison operators include ==, !=, <, <=, >, and >=. This issue does not occur when the comparison operation is performed in the filtering clause of the query. This issue can also occur when the difference of values in such columns is out of range for a LONG (64-bit) data type. For example, if column_c1 stores

8976171455044006767 and column_c2 stores -7272907770454997143, a query such as `SELECT column_c1 < column_c2 FROM table_test` returns `true` instead of `false` because the difference (`8976171455044006767 - (-7272907770454997143)`) is `1.6249079225499E19` which is greater than `9.22337203685478E18`, which is the maximum possible value that a `LONG` (64-bit) data type can hold.

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE_20207](#)

Cloudera Issue: CDH-70996

Workaround: Use a `DECIMAL` type instead of `BIGINT` for columns that might contain very large values. Another option is to disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Specified column position in the ORDER BY clause is not supported for SELECT * queries

When column positions are specified in `ORDER BY` clauses, they are not honored for `SELECT *` queries and an error is returned as shown in the following example:

```
CREATE TABLE decimal_1 (id decimal(5,0));
SELECT * FROM decimal_1 ORDER BY 1 limit 100;
Error while compiling statement: FAILED: SemanticException [Error 10219]: Position in
ORDER BY is not supported when using SELECT *
```

Instead the query must list out the columns it is selecting.

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-68550

DirectSQL with PostgreSQL

Hive doesn't support Hive direct SQL queries with PostgreSQL database. It only supports this feature with MySQL, MariaDB, and Oracle. With PostgreSQL, direct SQL is disabled as a precaution, since there have been issues reported upstream where it is not possible to fallback on DataNucleus in the event of some failures, plus other non-standard behaviors. For more information, see [Hive Configuration Properties](#).

Affected Versions: All CDH versions

Cloudera Issue: CDH-49017

`ALTER PARTITION ... SET LOCATION` does not work on Amazon S3 or between S3 and HDFS

Cloudera recommends that you do not use `ALTER PARTITION ... SET LOCATION` on S3 or between S3 and HDFS. The rest of the `ALTER PARTITION` commands work as expected.

Affected Versions: All CDH versions

Cloudera Issue: CDH-42420

Commands run against an Oracle-backed metastore might fail

Commands run against an Oracle-backed Metastore fail with error:

```
javax.jdo.JDODataStoreException Incompatible data type for column TBLS.VIEW_EXPANDED_TEXT
: was CLOB (datastore),
but type expected was LONGVARCHAR (metadata). Please check that the type in the datastore
and the type specified in the MetaData are consistent.
```

This error might occur if the metastore is run on top of an Oracle database with the configuration property `datanucleus.validateColumns` set to true.

Workaround: Set `datanucleus.validateColumns=false` in the `hive-site.xml` configuration file.

Affected Versions: All CDH versions

Cannot create archive partitions with external HAR (Hadoop Archive) tables

```
ALTER TABLE ... ARCHIVE PARTITION is not supported on external tables.
```

Affected Versions: All CDH versions**Cloudera Issue:** CDH-9638

Object types Server and URI are not supported in "SHOW GRANT ROLE *roleName* on OBJECT *objectName*" statements

Workaround: Use SHOW GRANT ROLE *roleName* to list all privileges granted to the role.

Affected Versions: All CDH versions**Cloudera Issue:** CDH-19430

HCatalog Known Issues



Note: As of CDH 5, HCatalog is part of Apache Hive.

There are no notable known issues in this release of HCatalog.

Hive on Spark (HoS) Known Issues

Hive on Spark queries fail with "Timed out waiting for client to connect" for an unknown reason

If this exception is preceded by logs of the form "client.RpcRetryingCaller: Call exception...", then this failure is due to an unavailable HBase service. On a secure cluster, spark-submit will try to obtain delegation tokens from HBase, even though Hive on Spark might not need them. So if HBase is unavailable, spark-submit throws an exception.

Workaround: Fix the HBase service, or set spark.yarn.security.tokens.hbase.enabled to false.

Affected Versions: CDH 5.7.0 and higher**Cloudera Issues:** CDH-59591, CDH-59599*Hue Known Issues*

Hue does not support the Spark App

Hue does not currently support the Spark application.

Connecting to PostgreSQL Database Fails with Error "No module named psycopg2"

When configuring Hue to use a PostgreSQL database, the connection fails with the following error:

```
Error loading psycopg2 module: No module named psycopg2
```

Workaround: Install the psycopg2 Python package as documented in [Installing the psycopg2 Python Package](#).

Affected Versions: All CDH 6 versions**Fixed Versions:** None**Apache Issue:** N/A**Cloudera Issue:** CDH-65804*Apache Impala Known Issues*

The following sections describe known issues and workarounds in Impala, as of the current production release. This page summarizes the most serious or frequently encountered issues in the current release, to help you make planning decisions about installing and upgrading. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.



Note: The online issue tracking system for Impala contains comprehensive information and is updated in real time. To verify whether an issue you are experiencing has already been reported, or which release an issue is fixed in, search on the [Impala JIRA tracker](#).

Impala Known Issues: Startup

These issues can prevent one or more Impala-related daemons from starting properly.

Impala requires FQDN from hostname command on kerberized clusters

The method Impala uses to retrieve the host name while constructing the Kerberos principal is the `gethostname()` system call. This function might not always return the fully qualified domain name, depending on the network configuration. If the daemons cannot determine the FQDN, Impala does not start on a kerberized cluster.

Workaround: Test if a host is affected by checking whether the output of the `hostname` command includes the FQDN. On hosts where `hostname` only returns the short name, pass the command-line flag `--hostname=fully_qualified_domain_name` in the startup options of all Impala-related daemons.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4978](#)

Impala Known Issues: Crashes and Hangs

These issues can cause Impala to quit or become unresponsive.

Unable to view large catalog objects in catalogd Web UI

In `catalogd` Web UI, you can list metadata objects and view their details. These details are accessed via a link and printed to a string formatted using thrift's `DebugProtocol`. Printing large objects (> 1 GB) in Web UI can crash `catalogd`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6841](#)

Impala Known Issues: Performance

These issues involve the performance of operations such as queries or DDL statements.

Metadata operations block read-only operations on unrelated tables

Metadata operations that change the state of a table, like `COMPUTE STATS` or `ALTER RECOVER PARTITIONS`, may delay metadata propagation of unrelated unloaded tables triggered by statements like `DESCRIBE` or `SELECT` queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6671](#)

Impala Known Issues: Security

These issues relate to security features, such as Kerberos authentication, Sentry authorization, encryption, auditing, and redaction.

Impala does not support Heimdal Kerberos

Heimdal Kerberos is not supported in Impala.

Apache Issue: [IMPALA-7072](#)

Affected Versions: All CDH 6 versions

System-wide auth-to-local mapping not applied correctly to Kudu service account

Due to system `auth_to_local` mapping, the principal may be mapped to some local name.

When running with Kerberos enabled, you may hit the following error message where `<random-string>` is some random string which doesn't match the primary in the Kerberos principal.

```
WARNINGS: TransmitData() to X.X.X.X:27000 failed: Remote error: Not authorized:
{username='<random-string>', principal='impala/redacted'} is not allowed to access
DataStreamService
```

Workaround: Start Impala with the `--use_system_auth_to_local=false` flag to ignore the system-wide auth_to_local mappings configured in `/etc/krb5.conf`.

Apache Issue: KUDU-2198

Affected Versions: CDH 5.15, CDH 6.1 and higher

Impala Known Issues: Resources

These issues involve memory or disk usage, including out-of-memory conditions, the spill-to-disk feature, and resource management features.

Handling large rows during upgrade to CDH 5.13 / Impala 2.10 or higher

After an upgrade to CDH 5.13 / Impala 2.10 or higher, users who process very large column values (long strings), or have increased the `--read_size` configuration setting from its default of 8 MB, might encounter capacity errors for some queries that previously worked.

Resolution: After the upgrade, follow the instructions in [Handling Large Rows During Upgrade to CDH 5.13 / Impala 2.10 or Higher](#) to check if your queries are affected by these changes and to modify your configuration settings if so.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6028](#)

Configuration to prevent crashes caused by thread resource limits

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal error: Unable to allocate
section memory!
terminate called after throwing an instance of
'boost::exception_detail::clone_impl<boost::exception_detail::error_info_injector<boost::thread_resource_error>
>'
```

Workaround:

In CDH 6.0 and lower versions of CDH, configure each host running an `impalad` daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

In CDH 6.1 and higher versions, it is unlikely that you will hit the thread resource limit. Configure each host running an `impalad` daemon with the following setting:

```
echo 8000000 > /proc/sys/vm/max_map_count
```

To make the above settings durable, refer to your OS documentation. For example, on RHEL 6.x:

1. Add the following line to `/etc/sysctl.conf`:

```
vm.max_map_count=8000000
```

2. Run the following command:

```
sysctl -p
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-5605](#)

Breakpad minidumps can be very large when the thread count is high

The size of the breakpad minidump files grows linearly with the number of threads. By default, each thread adds 8 KB to the minidump size. Minidump files could consume significant disk space when the daemons have a high number of threads.

Workaround: Add `--minidump_size_limit_hint_kb=size` to set a soft upper limit on the size of each minidump file. If the minidump file would exceed that limit, Impala reduces the amount of information for each thread from 8 KB to 2 KB. (Full thread information is captured for the first 20 threads, then 2 KB per thread after that.) The minidump file can still grow larger than the "hinted" size. For example, if you have 10,000 threads, the minidump file can be more than 20 MB.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3509](#)

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the impalad daemon.

Workaround: To monitor overall memory usage, use the top command, or add the memory figures in the Impala web UI /memz tab to JVM memory usage shown on the /metrics tab.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-691](#)

Impala Known Issues: Correctness

These issues can cause incorrect or unexpected results from queries. They typically only arise in very specific circumstances.

Incorrect result due to constant evaluation in query with outer join

An OUTER JOIN query could omit some expected result rows due to a constant such as FALSE in another join clause. For example:

```
explain SELECT 1 FROM alltypestiny a1
    INNER JOIN alltypesagg a2 ON a1.smallint_col = a2.year AND false
    RIGHT JOIN alltypes a3 ON a1.year = a3.bigint_col;
+-----+
| Explain String
+-----+
| Estimated Per-Host Requirements: Memory=1.00KB Vcores=1
| 00:EMPTYSET
+-----+
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3094](#)

% escaping does not work correctly in a LIKE clause

If the final character in the RHS argument of a LIKE operator is an escaped \% character, it does not match a % final character of the LHS argument.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2422](#)

Crash: impala::Coordinator::ValidateCollectionSlots

A query could encounter a serious error if includes multiple nested levels of INNER JOIN clauses involving subqueries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2603](#)

Impala Known Issues: Metadata

These issues affect how Impala interacts with metadata. They cover areas such as the metastore database and the Impala Catalog Server daemon.

Concurrent catalog operations with heavy DDL workloads can cause queries with SYNC_DDL to fail fast

When Catalog Server is under a heavy load with concurrent catalog operations of long running DDLs, queries running with the SYNC_DDL query option can fail with the following message:

```
ERROR: CatalogException: Couldn't retrieve the catalog topic
version for the SYNC_DDL operation after 3 attempts. The operation has
been successfully executed but its effects may have not been
broadcast to all the coordinators.
```

The catalog operation is actually successful as the change has been committed to HMS and Catalog Server cache, but when Catalog Server notices a longer than expected time for it to broadcast the changes, it fails fast.

The coordinator daemons eventually sync up in the background.

Affected Versions: CDH versions 6.0 and 6.1

Apache Issue: [IMPALA-7961](#) / CDH-76345

Impala Known Issues: Interoperability

These issues affect the ability to interchange data between Impala and other systems. They cover areas such as data types and file formats.

Queries Stuck on Failed HDFS Calls and not Timing out

If the following error appears multiple times in a short duration while running a query, it would mean that the connection between the impalad and the HDFS NameNode is in a bad state and hence the impalad would have to be restarted:

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed to finish before the
<hdfs_operation_timeout_sec> second timeout "
```

Workaround: Restart the impalad in the bad state.

Affected Versions: All versions of Impala

Apache Issue: [HADOOP-15720](#)

Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink hdfs.writeFormat must be set to Text, rather than its default value of Writable. The hdfs.writeFormat setting must be changed to Text before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

Resolution: This information has been requested to be added to the upstream Flume documentation.

Affected Versions: All CDH 6 versions

Cloudera Issue: CDH-13199

Avro Scanner fails to parse some schemas

The default value in Avro schema must match the first union type. For example, if the default value is null, then the first type in the UNION must be "null".

Workaround: Swap the order of the fields in the schema specification. For example, use ["null", "string"] instead of ["string", "null"]. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-635](#)

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1024](#)

Incorrect results with basic predicate on CHAR typed column

When comparing a `CHAR` column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1652](#)

Impala Known Issues: Limitations

These issues are current limitations of Impala that require evaluation as you plan how to integrate Impala into your data management workflow.

Set limits on size of expression trees

Very deeply nested expressions within queries can exceed internal Impala limits, leading to excessive memory usage.

Workaround: Avoid queries with extremely large expression trees. Setting the query option `disable_codegen=true` may reduce the impact, at a cost of longer query runtime.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4551](#)

Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

Workaround: Use standard HDFS on all Impala nodes.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-77](#)

Hue and BDR require separate parameters for Impala Load Balancer

Cloudera Manager supports a single parameter for specifying the Impala Daemon Load Balancer. However, because BDR and Hue need to use different ports when connecting to the load balancer, it is not possible to configure the load balancer value so that BDR and Hue will work correctly in the same cluster.

Workaround: To configure BDR with Impala, use the load balancer configuration either without a port specification or with the Beeswax port.

To configure Hue, use the **Hue Server Advanced Configuration Snippet (Safety Valve)** for `impalad_flags` to specify the load balancer address with the HiveServer2 port.

Affected Versions: CDH versions from 5.11 to 6.0.1

Cloudera Issue: [OPSAPS-46641](#)

Impala Known Issues: Miscellaneous / Older Issues

These issues do not fall into one of the above categories or have not been categorized yet.

A failed CTAS does not drop the table if the insert fails

If a `CREATE TABLE AS SELECT` operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed `CREATE TABLE AS SELECT`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2005](#)**Casting scenarios with invalid/inconsistent results**

Using a `CAST` function to convert large literal values to smaller types, or to convert special values such as `Nan` or `Inf`, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1821](#)

Impala Parser issue when using fully qualified table names that start with a number

A fully qualified table name starting with a number could cause a parsing error. In a name such as `db.571_market`, the decimal point followed by digits is interpreted as a floating-point number.

Workaround: Surround each part of the fully qualified name with backticks (`` ``).

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-941](#)

Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

Workaround: Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) for details about modifying these environment settings.

Resolution: Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-532](#)*EMC Isilon Known Issues*

CDH 6.0 is not currently supported on EMC Isilon.

Affected Versions: All CDH 6 versions

Apache Kafka Known Issues

Topics Created with the "kafka-topics" Tool Might Not Be Secured

Topics that are created and deleted via Kafka are secured (for example, auto created topics). However, most topic creation and deletion is done via the `kafka-topics` tool, which talks directly to ZooKeeper or some other third-party tool that talks directly to ZooKeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. Anyone with access to ZooKeeper can create and delete topics. They will not be able to describe, read, or write to the topics even if they can create them.

The following commands talk directly to ZooKeeper and therefore are not secured via Kafka:

- `kafka-topics.sh`
- `kafka-configs.sh`
- `kafka-preferred-replica-election.sh`
- `kafka-reassign-partitions.sh`

"`offsets.topic.replication.factor`" Must Be Less Than or Equal to the Number of Live Brokers

The `offsets.topic.replication.factor` broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a `GROUP_COORDINATOR_NOT_AVAILABLE` error until the cluster size meets this replication factor requirement.

Requests Fail When Sending to a Nonexistent Topic with "auto.create.topics.enable" Set to True

The first few produce requests fail when sending to a nonexistent topic with `auto.create.topics.enable` set to true.

Workaround: Increase the number of retries in the Producer configuration setting `retries`.

Cloudera Enterprise 6 Release Guide

Custom Kerberos Principal Names Cannot Be Used for Kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start.

Workaround: None. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Performance Degradation When SSL Is Enabled

Significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with `ssl.secure.random.implementation = SHA1PRNG`. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Affected Versions: CDK 2.x and later

Fixed Versions: None

Apache Issue: KAFKA-2561

Cloudera Issue: None

Kafka Broker Fails to Start Due to Slow Sentry and HMS startup

This issue is encountered on cluster startup and is caused by misalignment between Kafka, Sentry, and HMS. The slow startup of HMS slows down Sentry startup which consequently makes the Kafka connection to Sentry time out. Ultimately, the Kafka broker will be unable to start.

Workaround: Manually increase the number of remote procedure call retries between Sentry and Kafka through the **Sentry Client Advanced Configuration Snippet (Safety Valve)** for `sentry-site.xml` property.

1. Go to **Sentry > Configuration** and find the **Sentry Client Advanced Configuration Snippet (Safety Valve) for `sentry-site.xml`** property.
2. Click on the add button.
3. Enter the following data:
 - Name: `sentry.service.client.rpc.retry-total`
 - Value: 20
4. Enter a **Reason for change**, and then click **Save Changes** to commit the changes.
5. Return to the Home page by clicking the Cloudera Manager logo.
6. Click the restart stale services icon next to the Sentry service to invoke the cluster restart wizard.
7. Click **Restart Stale Services**.
8. Click **Restart Now**.
9. Click **Finish**.

Affected Versions: CDH 6.1.0 and higher

Fixed Versions: N/A

Cloudera Issue: CDH-74713

Kafka JMX Tool Cannot Connect to JMX

The Kafka JMX tool cannot connect to the JMX agent of the Kafka Broker or MirrorMaker if the specified address of the JMX remote connector is bound to 127.0.0.1.

Workaround:

1. In Cloudera Manager go to **Kafka > Instances** and select the affected broker.
2. Find the **Additional Broker Java Options** and **Additional MirrorMaker Java Options** properties and add the following Java option to the configuration:

```
-Djava.rmi.server.hostname=127.0.0.1
```



Note: Configuring the **Additional MirrorMaker Java Options** property is only required if you are using JMX with MirrorMaker.

3. Restart the affected brokers.

Affected Versions: CDH 6.0.0 and higher

Fixed Versions: CDH 6.2.0

Cloudera Issue: OPSAPS-48695

Apache Kudu Known Issues

The following are known bugs and issues in Kudu. Note that this list is not exhaustive, and is meant to communicate only the most important known issues.

C++ Client Fails to Re-acquire Authentication Token in Multi-master Clusters

A security-related issue can cause Impala queries to start failing on busy clusters in the following scenario:

- The cluster runs with the `--rpc_authentication` set as optional or required. The default is optional. Secure clusters use required.
- The cluster is using multiple masters.
- Impala queries happen frequently enough that the leader master connection to some `impalad` isn't idle-closed (more than 1 query per 65 seconds).
- The connection stays alive for longer than the authentication token timeout (1 week by default).
- A master leadership change occurs after the authentication token expiration.

Impala queries will start failing with errors in the `impalad` logs like:

```
I0904 13:53:08.748968 95857 client-internal.cc:283] Unable to determine the new leader
Master: Not authorized: Client connection negotiation failed: client connection to
10.164.44.13:7051: FATAL_INVALID_AUTHENTICATION_TOKEN: Not authorized: authentication
token expired
I0904 13:53:10.389009 95861 status.cc:125] Unable to open Kudu table: Timed out:
GetTableSchema timed out after deadline expired
@ 0x95b1e9 impala::Status::Status()
@ 0xff22d4 impala::KuduScanNodeBase::Open()
@ 0xff101e impala::KuduScanNode::Open()
@ 0xb73ced impala::FragmentInstanceState::Open()
@ 0xb7532b impala::FragmentInstanceState::Exec()
@ 0xb64ae8 impala::QueryState::ExecFInstance()
@ 0xd15193 impala::Thread::SuperviseThread()
@ 0xd158d4 boost::detail::thread_data<>::run()
@ 0x129188a (unknown)
@ 0x7f717ceade25 start_thread
@ 0x7f717cbdb34d __clone
```

Impala shell queries will fail with a message like:

```
Unable to open Kudu table: Timed out: GetTableSchema timed out after deadline expired
```

Workaround:

- Restart the affected Impala Daemons. Restarting a daemon ensures the problem will not reoccur for at least the authentication token lifetime, which defaults to one week.
- Increase the authentication token lifetime (`--authn_token_validity_seconds`). Beware that raising this lifetime increases the window of vulnerability of the cluster if a client is compromised. It is recommended that you keep the token lifetime at one month maximum for a secure cluster. For unsecured clusters, a longer token lifetime is acceptable, and a 3 month lifetime is recommended.

Affected Versions: From CDH 5.11 through CDH 6.0.1

Apache Issue: [KUDU-2580](#)

Timeout Possible with Log Force Synchronization Option

If the Kudu master is configured with the `-log_force_fsync_all` option, tablet servers and clients will experience frequent timeouts, and the cluster may become unusable.

Affected Versions:

All CDH 6 versions

Longer Startup Times with a Large Number of Tablets

If a tablet server has a very large number of tablets, it may take several minutes to start up. It is recommended to limit the number of tablets per server to 1000 or fewer. The maximum allowed number of tablets is 2000 per server. Consider this limitation when pre-splitting your tables. If you notice slow start-up times, you can monitor the number of tablets per server in the web UI.

Affected Versions:

All CDH 6 versions

Descriptions for Kudu TLS/SSL Settings in Cloudera Manager

Use the descriptions in the following table to better understand the TLS/SSL settings in the Cloudera Manager Admin Console.

Field	Usage Notes
Kerberos Principal	Set to the default principal, <code>kudu</code> .
Enable Secure Authentication And Encryption	Select this checkbox to enable authentication <i>and</i> RPC encryption between all Kudu clients and servers, as well as between individual servers. Only enable this property after you have configured Kerberos.
Master TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu master host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to the Kudu master web UI.
Tablet Server TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu tablet server host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to Kudu tablet server web UIs.
Master TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu master host's private key (set in Master TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Tablet Server TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu tablet server host's private key (set in Tablet Server TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Master TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Tablet Server TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Enable TLS/SSL for Master Server	Enables HTTPS encryption on the Kudu master web UI.
Enable TLS/SSL for Tablet Server	Enables HTTPS encryption on the Kudu tablet server web UIs.

Affected Versions:

All CDH 6 versions

Apache Oozie Known Issues

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used due to a sys table change in PostgreSQL from version 9.5 to 9.6. The failure only happens if Oozie uses a JDBC driver earlier than 9.4.1209.

Workaround:

1. After the parcels of the new version are distributed, replace the PostgreSQL JDBC driver with a newer one (version 9.4.1209 or higher) in the new parcel, at the following locations:
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/lib/
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/libtools/
2. Perform the upgrade.



Note: If you already started the upgrade and the process stops with an error message about missing columns, you can change the drivers at that point of the process as well, and resume the upgrade.

If your cluster is installed from packages, you must change the drivers at the following locations:

- /usr/lib/oozie/libtools/
- /usr/lib/oozie/lib/



Note: You can change the driver after the packages installation, but before running the CDH upgrade wizard. You can also do it during the update process, when the error occurs.

You can download the driver from the [PostgreSQL JDBC driver homepage](#).

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-75951

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the `resume` command to tell Oozie to continue the workflow from the point at which it left off.

Affected Versions: CDH 5 and higher

Cloudera Issue: CDH-14623

Oozie works with MapReduce or YARN, but not both

The Oozie server works with a MapReduce (MRv1) cluster or a YARN (MRv2) cluster, but not both at the same time.

Workaround: Use two different Oozie servers.

Affected Versions: All

Apache Parquet Known Issues

There are no known issues in Parquet.

Apache Pig Known Issues

There are no known issues in this release.

Cloudera Search Known Issues

The current release includes the following known limitations:

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: `solr.sentry.enableSqlQuery=true`
- Graph: `solr.sentry.enableGraphQuery=true`
- Stream: `solr.sentry.enableStreams=true`

Workaround: None

Affected Versions: All CDH 6 releases

Cloudera Issue: CDH-66345

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search for CDH 5.5.0 includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-34050

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with `CrunchIndexerTool`, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-22190

The `quickstart.sh` file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if ZooKeeper and the NameNode are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the NameNode and ZooKeeper are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the MapReduce jobs and Solr commands that are run by the `quickstart`.

Affected Versions: All

Cloudera Issue: CDH-19923

Field value class guessing and Automatic schema field addition are not supported with the `MapReduceIndexerTool` nor the `HBaseMapReduceIndexerTool`

The `MapReduceIndexerTool` and the `HBaseMapReduceIndexerTool` can be used with a Managed Schema created via NRT indexing of documents or via the Solr Schema API. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the `MapReduceIndexerTool` or `HBaseMapReduceIndexerTool`. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the Solr Schema API or index sample documents using NRT indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect using the List Fields API command.

Affected Versions: All

Cloudera Issue: CDH-26856

The `Browse` and `Spell` Request Handlers are not enabled in schemaless mode

The `Browse` and `Spell` Request Handlers require certain fields be present in the schema. Since those fields cannot be guaranteed to exist in a Schemaless setup, the `Browse` and `Spell` Request Handlers are not enabled by default.

Workaround: If you require the “Browse” and “Spell” Request Handlers, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Affected Versions: All

Cloudera Issue: CDH-19407

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. Blockcache writing is disabled by default.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-17978

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given page explaining that access is denied, and instead receive a web page that never finishes loading.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-58276

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection.

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Affected Versions: All

Cloudera Issue: CDH-15441

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Affected Versions: All

Cloudera Issue: CDH-58694

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-21162

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-11357

Apache Sentry Known Issues

When granting privileges, a single transaction per grant causes long delays

Sentry takes a long time to grant or revoke a large number of column-level privileges that are requested in a single statement. For example if you execute the following command:

```
GRANT SELECT(col1, col2, ...) ON TABLE table1;
```

Sentry applies the grants to each column separately and the refresh process causes long delays.

Workaround: Split the grant statement up into smaller chunks. This prevents the refresh process from causing delays.

Affected Versions:

- CDH: 5.14.4
- CDH: 5.15.1
- CDH: 5.16.0
- CDH: 6.1.0

Fixed Versions:

- CDH 5.16.1 and above
- CDH 6.2.0 and above

Cloudera Issue: CDH-74982

SHOW ROLE GRANT GROUP raises exception for a group that was never granted a role

If you run the command SHOW ROLE GRANT GROUP for a group that has never been granted a role, beeline raises an exception. However, if you run the same command for a group that does not have any roles, but has at one time been granted a role, you do not get an exception, but instead get an empty list of roles granted to the group.

Workaround: Adding a role will prevent the exception.

Affected Versions:

- CDH 5.16.0
- CDH 6.0.0

Cloudera Issue: CDH-71694

GRANT/REVOKE operations could fail if there are too many concurrent requests

Under a significant workload, Grant/Revoke operations can have issues.

Workaround: If you need to make many privilege changes, plan them at a time when you do not need to do too many at once.

Affected Versions: CDH 5.13.0 and above

Apache Issue: [SENTRY-1855](#)

Cloudera Issue: CDH-56553

Creating large set of Sentry roles results in performance problems

Using more than a thousand roles/permissions might cause significant performance problems.

Workaround: Plan your roles so that groups have as few roles as possible and roles have as few permissions as possible.

Affected Versions: CDH 5.13.0 and above

Cloudera Issue: CDH-59010

Users can't track jobs with Hive and Sentry

As a prerequisite of enabling Sentry, Hive impersonation is turned off, which means all YARN jobs are submitted to the Hive job queue, and are run as the `hive` user. This is an issue because the YARN History Server now has to block users from accessing logs for their own jobs, since their own usernames are not associated with the jobs. As a result, end users cannot access any job logs unless they can get `sudo` access to the cluster as the `hdfs`, `hive` or other admin users.

In CDH 5.8 (and higher), Hive overrides the default configuration, `mapred.job.queuename`, and places incoming jobs into the connected user's job queue, even though the submitting user remains `hive`. Hive obtains the relevant queue/username information for each job by using YARN's `fair-scheduler.xml` file.

Affected Versions: CDH 5.2.0 and above

Cloudera Issue: CDH-22890

Column-level privileges are not supported on Hive Metastore views

`GRANT` and `REVOKE` for column level privileges is not supported on Hive Metastore views.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-754](#)

`SELECT` privilege on all columns does not equate to `SELECT` privilege on table

Users who have been explicitly granted the `SELECT` privilege on all columns of a table, will *not* have the permission to perform table-level operations. For example, operations such as `SELECT COUNT (1)` or `SELECT COUNT (*)` will not work even if you have the `SELECT` privilege on all columns.

There is one exception to this. The `SELECT * FROM TABLE` command will work even if you do not have explicit table-level access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-838](#)

The `EXPLAIN SELECT` operation works without table or column-level privileges

Users are able to run the `EXPLAIN SELECT` operation, exposing metadata for all columns, even for tables/columns to which they weren't explicitly granted access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-849](#)

Object types Server and URI are not supported in `SHOW GRANT ROLE roleName on OBJECT objectName`

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Apache Issue: N/A

Cloudera Issue: CDH-19430

Relative URI paths not supported by Sentry

Sentry supports only absolute (not relative) URI paths in permission grants. Although some early releases (for example, CDH 5.7.0) might not have raised explicit errors when relative paths were set, upgrading a system that uses relative paths causes the system to lose Sentry permissions.

Affected Versions: All versions. Relative paths are not supported in Sentry for permission grants.

Resolution: Revoke privileges that have been set using relative paths, and grant permissions using absolute paths before upgrading.

Absolute (Use this form)	Relative (Do not use this form)
<code>hdfs://absolute/path/</code>	<code>hdfs://relative/path</code>
<code>s3a://bucketname/</code>	<code>s3a://bucketname</code>

Apache Spark Known Issues

The following sections describe the current known issues and limitations in Apache Spark 2.x as distributed with CDH 6.1.x. In some cases, a feature from the upstream Apache Spark project is currently not considered reliable enough to be supported by Cloudera.

RDD.repartition() has different failure handling in Spark 2.4 and may cause job failures

The `RDD.repartition()` transformation, which reshuffles data in the RDD randomly to create either more or fewer partitions and then balances it across the partitions, was using a round-robin method to distribute data that caused incorrect answers to be returned for RDD jobs. This issue has been corrected, but it introduced a behavior change in RDD job failure handling. Now, Spark actively fails a job if there is a fetch failure that was caused by a node failure after repartitioning.

Workaround: Use the `RDD.checkpoint()` method to save the intermediate RDD data to HDFS. First, call `SparkContext.setCheckpointDir(directory: String)` to set the checkpoint directory where the intermediate data will be saved. Note that the directory must be an HDFS path. Then mark the RDD for checkpointing by calling `RDD.checkpoint()` when you use the `RDD.repartition()` transformation.

Apache Issue: [SPARK-23243](#)

Cloudera Issue: CDH-76413

Spark Streaming write-ahead logs do not run on HDFS directories with Erasure Coding enabled

Spark Streaming write-ahead logs (WALs) cannot run on HDFS directories when Erasure Coding is enabled. Erasure Coding does not support `hflush()`, `hsync()`, and `append()`, which prevents the WALs from running.

Workaround: Configure Spark Streaming with a checkpoint directory that does not have Erasure Coding enabled on it. You can set the checkpoint directory with `ssc.checkpoint("directory_name")`. For example:

```
ssc.checkpoint("_checkpoint")
```

Affected Versions: CDH 6.1.0

Fixed Versions: CDH 6.2.0

Apache Issue: [SPARK-26094](#)

Cloudera Issue: CDH-61127

PySpark broadcast variables fail when disk encryption is enabled

When disk encryption is enabled, PySpark broadcast variables fail with the following stack trace:

```
Traceback (most recent call last): File "broadcast.py", line 37, in <module>
words_new.value File "/pyspark.zip/pyspark/broadcast.py", line 137, in value
File "pyspark.zip/pyspark/broadcast.py", line 122, in load_from_path File
"pyspark.zip/pyspark/broadcast.py", line 128, in load EOFError: Ran out of input
```

Workaround: None

Affected Versions: CDH 6.0.1, CDH 6.1.0, CDH 6.2.0

Apache Issue: [SPARK-26201](#)

Cloudera Issue: CDH-76116

Structured Streaming exactly-once fault tolerance constraints

In Spark Structured Streaming, the exactly-once fault tolerance for `file sink` is valid only for files that are in the manifest. These files are located in the `_spark_metadata` subdirectory of the `file sink` output directory. Only process files that have file names starting with digits. Other temporary files can also appear in this directory, but they should not be processed. Typically, these temporary file file names start with a period (".").

You can list the valid manifest files, excluding the temporary files, by using a command like the following, which assumes your output directory is located at `/tmp/output`. As the appropriate user, run the following command to list the valid manifest files:

```
hadoop fs -ls /tmp/output/_spark_metadata/[0-9]*
```

Workaround: None

Affected Versions: CDH 6.1.0 and higher

Cloudera Issue: CDH-75191

Spark SQL does not respect size limit for the varchar type

Spark SQL treats `varchar` as a string (that is, there no size limit). The observed behavior is that Spark reads and writes these columns as regular strings; if inserted values exceed the size limit, no error will occur. The data will be truncated when read from Hive, but not when read from Spark.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Apache Issue: [SPARK-5918](#)

Cloudera Issue: CDH-33642

Spark SQL does not prevent you from writing key types not supported by Avro tables

Spark allows you to declare DataFrames with any key type. Avro supports only string keys and trying to write any other key type to an Avro table will fail.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33648

Spark SQL does not support timestamp in Avro tables

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33649

Spark SQL does not respect Sentry ACLs when communicating with Hive metastore

Even if user is configured via Sentry to not have read permission to a Hive table, a Spark SQL job running as that user can still read the table's metadata directly from the Hive metastore. **Cloudera Issue:** CDH-33658

Dynamic allocation and Spark Streaming

If you are using Spark Streaming, Cloudera recommends that you disable dynamic allocation by setting `spark.dynamicAllocation.enabled` to `false` when running streaming applications.

Limitation with Region Pruning for HBase Tables

When SparkSQL accesses an HBase table through the HiveContext, region pruning is not performed. This limitation can result in slower performance for some SparkSQL queries against tables that use the HBase SerDes than when the same table is accessed through Impala or Hive.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-56330

Running `spark-submit` with `--principal` and `--keytab` arguments does not work in client mode

Cloudera Enterprise 6 Release Guide

The `spark-submit` script's `--principal` and `--keytab` arguments do not work with Spark-on-YARN's client mode.

Workaround: Use `cluster` mode instead.

Affected Versions: All

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Apache Sqoop Known Issues

Column names cannot start with a number when importing data with the `--as-parquetfile` option.

Currently, Sqoop is using an Avro schema when writing data as a parquet file. The Avro schema requires that column names do not start with numbers, therefore Sqoop is renaming the columns in this case, prepending them with an underscore character. This can lead to issues when one wants to reuse the data in other tools, such as Impala.

Workaround: Rename the columns to comply with Avro limitations (start with letters or underscore, as specified in the [Avro documentation](#)).

Cloudera Issue: None

MySQL JDBC driver shipped with CentOS 6 systems does not work with Sqoop

CentOS 6 systems currently ship with version 5.1.17 of the MySQL JDBC driver. This version does not work correctly with Sqoop.

Workaround: Install version 5.1.31 of the JDBC driver as detailed in [Installing the JDBC Drivers for Sqoop 1](#).

Affected Versions: MySQL JDBC 5.1.17, 5.1.4, 5.3.0

Cloudera Issue: CDH-23180

MS SQL Server "integratedSecurity" option unavailable in Sqoop

The `integratedSecurity` option is not available in the Sqoop CLI.

Workaround: None

Cloudera Issue: None

Sqoop1 (doc import + `--as-parquetfile`) limitation with KMS/KTS Encryption at Rest

Due to a limitation with Kite SDK, it is not possible to use (`sqoop import --as-parquetfile`) with KMS/KTS Encryption zones. See the following example.

```
sqoop import --connect jdbc:db2://djaxludb1001:61035/DBBAT003 --username=dh810202 --password=XXXXXXXXXX --target-dir /data/hive_scratch/ASDISBURSEMENT --delete-target-dir -m1 --query "select disbursementnumber,disbursementdate,xmldata FROM DB2dba.ASDISBURSEMENT where DISBURSEMENTNUMBER = 2011113210000115311 AND \$CONDITIONS" -hive-import --hive-database adminserver -hive-table asdisbursement_dave --map-column-java XMLDATA=String --as-parquetfile

16/12/05 12:23:46 INFO mapreduce.Job: map 100% reduce 0%
16/12/05 12:23:46 INFO mapreduce.Job: Job job_1480530522947_0096 failed with state FAILED
due to: Job commit failed: org.kitesdk.data.DatasetIOException: Could not move contents
of
hdfs://AJAX01-ns/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096
to hdfs://AJAX01-ns/data/RetiredApps/INS/AdminServer/asdisbursement_dave
<SNIP>
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.IOException):
```

```
/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096/5ddcac42-5d69-4e46-88c2-17bedac4858.parquet  
can't be moved into an encryption zone.
```

Workaround: If you use the Parquet Hadoop API based implementation for importing into Parquet, specify a `--target-dir` which is the same encryption zone as the Hive warehouse directory.

If you use the Kite Dataset API based implementation, use an alternate data file type, for example text or avro.

Apache Issue: SQUOP-2943

Cloudera Issue: CDH-40826

Doc import as Parquet files may result in out-of-memory errors

Out-of-memory (OOM) errors can be caused in the following two cases:

- With many very large rows (multiple megabytes per row) before initial-page-run check (ColumnWriter)
- When rows vary significantly by size so that the next-page-size check is based on small rows and is set very high followed by many large rows

Apache Issue: PARQUET-99

Workaround: None, other than restructuring the data.

Apache ZooKeeper Known Issues

There are no known issues in this release.

CDH 6.0.x Release Notes

To view release notes for specific CDH 6.0.x releases, see the following:

CDH 6.0.1 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for CDH 6.0.1:

What's New in CDH 6.0.1

This is a maintenance release that fixes some important issues. For details, see [Fixed Issues in CDH 6.0.1](#) on page 431.

Fixed Issues in CDH 6.0.1

See below for issues fixed in CDH 6.0.1, grouped by component:

Apache Flume

The following issues are fixed in CDH 6.0.1:

- [FLUME-3237](#) - Handling RuntimeExceptions coming from the JMS provider in JMSSource

Apache Hadoop

The following issues are fixed in CDH 6.0.1:

- [HADOOP-13597](#) - Fixed an issue where the KMS consumes memory and increases File Descriptors.
- [HADOOP-15311](#) - You can now configure the acceptor/selector count for HttpServer2.
- [HADOOP-15473](#) - Fixed an UnrecoverableKeyException caused by JDK-8189997.
- [HADOOP-15593](#) - Fixed NPE in UGI spawnAutoRenewalThreadForUserCreds.
- [HADOOP-15609](#) - The KMS now retries when an SSLHandshakeException occurs.
- [HADOOP-15638](#) - The KMS Accept Queue Size default changed from 500 to 128 in Hadoop 3.x.
- [HADOOP-15655](#) - Enhance KMS client retry behavior to retry on timeout.
- [HADOOP-15696](#) - Fixed an issue where the KMS experiences performance regression due to too many open file descriptors after Jetty migration.
- [HADOOP-15698](#) - Fixed an issue where the KMS log4j is not initialized properly at startup.
- [HADOOP-15708](#) - Fixed an issue where reading values from Configuration before adding deprecations makes it impossible to read the values with a deprecated key.

HBase

The following issues are fixed in CDH 6.0.1:

- [HBASE-19572](#) - RegionMover should use the configured default port number and not the one from HConstants
- [HBASE-19722](#) - Meta query statistics metrics source
- [HBASE-19764](#) - Fix Checkstyle errors in hbase-endpoint
- [HBASE-20244](#) - NoSuchMethodException when retrieving private method decryptEncryptedDataEncryptionKey from DFSClient
- [HBASE-20401](#) - Make MAX_WAIT and waitIfNotFinished in CleanerContext configurable
- [HBASE-20403](#) - Fix race between prefetch task and non-pread HFile reads
- [HBASE-20474](#) - Show non-RPC tasks on master/regionserver Web UI by default
- [HBASE-20538](#) - Upgrade our hadoop versions to 2.7.7 and 3.0.3
- [HBASE-20565](#) - ColumnRangeFilter combined with ColumnPaginationFilter can produce incorrect result
- [HBASE-20614](#) - REST scan API with incorrect filter text file throws HTTP 503 Service Unavailable error
- [HBASE-20642](#) - Clients should re-use the same nonce across DDL operations
- [HBASE-20648](#) - HBASE-19364 "Truncate_preserve fails with table when replica region > 1" for master branch
- [HBASE-20649](#) - Validate HFiles do not have PREFIX_TREE DataBlockEncoding; ADDEDNDUM ADD MISSING FILE
- [HBASE-20649](#) - Validate HFiles do not have PREFIX_TREE DataBlockEncoding
- [HBASE-20681](#) - Explicitly include hamcrest in binary tarball
- [HBASE-20691](#) - Change the default WAL storage policy back to "NONE""
- [HBASE-20697](#) - Can't cache All region locations of the specify table by calling table.getRegionLocator().getAllRegionLocations()
- [HBASE-20705](#) - Having RPC quota on a table now no longer prevents Space Quota to be recreate/removed
- [HBASE-20706](#) - Prevent MTP from trying to reopen non-OPEN regions
- [HBASE-20723](#) - Custom hbase.wal.dir results in data loss because we write recovered edits into a different place than where the recovering region server looks for them
- [HBASE-20745](#) - Log when master proc wal rolls
- [HBASE-20752](#) - Make sure the regions are truly reopened after ReopenTableRegionsProcedure
- [HBASE-20770](#) - WAL cleaner logs way too much; gets clogged when lots of work to do
- [HBASE-20772](#) - Controlled shutdown fills Master log with the disturbing message 'No matching procedure found for rit=OPEN, location=ZZZZ, table=YYYYY, region=XXXX transition to CLOSED'
- [HBASE-20777](#) - RpcConnection could still remain opened after we shutdown the NettyRpcServer
- [HBASE-20780](#) - ServerRpcConnection logging cleanup Get rid of one of the logging lines in ServerRpcConnection by amalgamating all into one new-style log line.
- [HBASE-20781](#) - Save recalculating families in a WALEdit batch of Cells
- [HBASE-20794](#) - add INFO level log to createTable operation
- [HBASE-20795](#) - Allow option in BBKVComparator.compare to do comparison without sequence id
- [HBASE-20806](#) - Split style journal for flushes and compactions
- [HBASE-20810](#) - Include the procedure id in the exception message in HBaseAdmin for better debugging
- [HBASE-20812](#) - Add defaults to Table Interface so implementors don't have to
- [HBASE-20813](#) - Removed RPC quotas when the associated table/Namespace is dropped off
- [HBASE-20817](#) - Infinite loop when executing ReopenTableRegionsProcedure
- [HBASE-20825](#) - Fix pre and post hooks of CloneSnapshot and RestoreSnapshot for Access checks
- [HBASE-20826](#) - Truncate really long RpcServer warnings unless TRACE is on
- [HBASE-20829](#) - Remove the addFront assertion in MasterProcedureScheduler.doAdd
- [HBASE-20833](#) - Modify pre-upgrade coprocessor validator to support table level coprocessors
- [HBASE-20839](#) - Fallback to FSHLog if we can not instantiated AsyncFSWAL when user does not specify AsyncFSWAL explicitly
- [HBASE-20853](#) - Polish "Add defaults to Table Interface so Implementors don't have to"
- [HBASE-20856](#) - PITA having to set WAL provider in two places
- [HBASE-20860](#) - Merged region's RIT state may not be cleaned after master restart

- [HBASE-20867](#) - RS may get killed while master restarts
- [HBASE-20869](#) - Endpoint-based Export use incorrect user to write to destination
- [HBASE-20875](#) - MemStoreLABImp::copyIntoCell uses 7% CPU when writing
- [HBASE-20878](#) - Data loss if merging regions while ServerCrashProcedure executing
- [HBASE-20882](#) - HBASE-20616 "TruncateTableProcedure is stuck in retry loop in TRUNCATE_TABLE_CREATE_FS_LAYOUT state" to branch-2.0
- [HBASE-20885](#) - Removed entry for RPC quota from hbase:quota when RPC quota is removed
- [HBASE-20887](#) - HBASE-20865 "CreateTableProcedure is stuck in retry loop in CREATE_TABLE_WRITE_FS_LAYOUT state"
- [HBASE-20903](#) - HBASE-20792 "info:servername and info:sn inconsistent for OPEN region" to branch-2.0
- [HBASE-20914](#) - Trim Master memory usage
- [HBASE-20921](#) - Possible NPE in ReopenTableRegionsProcedure
- [HBASE-20924](#) - "HBASE-20846 Restore procedure locks when master restarts"
- [HBASE-20935](#) - HStore.removeCompactedFiles should log in case it is unable to delete a file
- [HBASE-20939](#) - There will be race when we call suspendIfNotReady and then throw ProcedureSuspendedException
- [HBASE-20940](#) - HStore.cansplit should not allow split to happen if it has references
- [HBASE-20941](#) - Created and implemented HbckService in master
- [HBASE-20942](#) - Fix ArrayIndexOutOfBoundsException for RpcServer TRACE logging
- [HBASE-20975](#) - Lock may not be taken or released while rolling back procedure
- [HBASE-20978](#) - [amv2] Worker terminating UNNATURALLY during MoveRegionProcedure
- [HBASE-20981](#) - Rollback stateCount accounting thrown-off when exception out of rollbackState
- [HBASE-20989](#) - Minor, miscellaneous logging fixes
- [HBASE-21004](#) - to branch-2.0 HBASE-20708 "Remove the usage of RecoverMetaProcedure"
- [HBASE-21007](#) - Memory leak in HBase REST server
- [HBASE-21018](#) - RS crashed because AsyncFS was unable to update HDFS data encryption key
- [HBASE-21029](#) - Miscount of memstore's heap/offheap size if same cell was put
- [HBASE-21031](#) - Memory leak if replay edits failed during region opening
- [HBASE-21041](#) - Memstore's heap size will be decreased to minus zero after flush
- [HBASE-21047](#) - Object creation of StoreFileScanner thru constructor and close may leave refCount to -1
- [HBASE-21050](#) - Exclusive lock may be held by a SUCCESS state procedure forever
- [HBASE-21062](#) - Correctly use the defaultProvider value on the Providers enum when constructing a WALProvider
- [HBASE-21072](#) - Block out HBCK1 in hbase2
- [HBASE-21078](#) - [amv2] CODE-BUG NPE in RTP doing Unassign
- [HBASE-21083](#) - Introduce a mechanism to bypass the execution of a stuck procedure
- [HBASE-21088](#) - HStoreFile should be closed in HStore#hasReferences
- [HBASE-21120](#) - MoveRegionProcedure makes no progress; goes to STUCK

Region Server occasionally fails when HDFS data transport encryption is enabled

In rare cases, an HBase RegionServer on a Hadoop Data Transfer Encryption enabled cluster (dfs.encrypt.data.transfer = true) may crash because it is not able to update the encryption key.

Workaround: Restart the RegionServer.

Affected Versions: CDH 6.0.0

Fixed Versions: 6.0.1

Apache Issue: [HBASE-21018](#)

Cloudera Issue: CDH-71613

Prefetch sometimes doesn't work with encrypted file system

If HBase prefetch is enabled (hbase.rs.prefetchblocksonopen = true) on an encrypted HDFS cluster, HBase RegionServer may crash due to memory corruption.

Workaround: Disable HBase prefetch (hbase.rs.prefetchblocksonopen = false).

Affected Versions: CDH 6.0.0

Fixed Versions: 6.0.1

Apache Issue: [HBASE-20403](#)

Cloudera Issue: CDH-68666

Apache HDFS

The following issues are fixed in CDH 6.0.1:

- [HDFS-5040](#) - You can now see an audit log for admin commands and output the log of all DFS admin commands.
- [HDFS-10240](#) - Fixed an issue where the race between close/recoverLease leads to missing blocks.
- [HDFS-10453](#) - Fixed an issue where the ReplicationMonitor thread could get stuck for a long time due to the race between replication and delete of the same file in a large cluster.
- [HDFS-13051](#) - Fixed an issue where a deadlock occurs when rolleeditlog rpc call happens and editPendingQ is full.
- [HDFS-13178](#) - Add a force option to DiskBalancer Execute command
- [HDFS-13181](#) - Add a configuration to DiskBalancer for valid plan hours
- [HDFS-13281](#) - Fixed an issue where the Namenode#createFile was not ./reserved/raw/ aware.
- [HDFS-13314](#) - NameNode optionally exits if it detects FslImage corruption
- [HDFS-13322](#) - FUSE lib now recognizes the change of the Kerberos ticket cache path if it was changed with the KRB5CCNAME environment variable during the same user session.
- [HDFS-13339](#) - Fixed an issue where volume reference cannot be released and may lead to deadlock when DataXceiver does a check volume
- [HDFS-13721](#) - Fixed an NPE in DataNode due to an uninitialized DiskBalancer.
- [HDFS-13727](#) - The DiskBalancer now logs a full stack trace if it exits with an unhandled exception.
- [HDFS-13813](#) - Added a check to see if a child inode exists in the global FSDirectory dir when saving (serializing) INodeDirectorySection.

Apache Hive

Code Changes Should Not Be Required

The following fixes should not require code changes, but they contain improvements that might enhance your deployment:

- [HIVE-13696](#) - Modify fair-scheduler.xml and automatically update/validate jobs submitted to fair-scheduler
- [HIVE-15387](#) - NPE in HiveServer2 webUI Historical SQL Operations section
- [HIVE-16483](#) - Hive on Spark should populate split-related configurations to HiveConf
- [HIVE-17213](#) - Hive on Spark: file merging doesn't work for union all
- [HIVE-18977](#) - Listing partitions returns different results with JDO and direct SQL
- [HIVE-19048](#) - Initscript errors are ignored
- [HIVE-19133](#) - HiveServer2 WebUI phase-wise performance metrics not showing correctly
- [HIVE-19202](#) - CBO failed due to NullPointerException in HiveAggregate.isBucketedInput()
- [HIVE-19251](#) - ObjectStore.getNextNotification with LIMIT should use less memory
- [HIVE-19259](#) - CREATE VIEW that uses UNION ALL fails with 'Table not found'
- [HIVE-19265](#) - Potential NPE returned instead of actual exception in Hive#copyFiles
- [HIVE-19424](#) - NPE In MetaDataFormatters
- [HIVE-19668](#) - HiveServer2 service hanging due to over 30% of the heap wasted by duplicate org.antlr.runtime.CommonToken's and duplicate strings
- [HIVE-19752](#) - PerfLogger integration for critical Hive-on-S3 paths
- [HIVE-19870](#) - HCatalog dynamic partition query can fail if the table path is managed by Sentry
- [HIVE-19891](#) - inserting into external tables with custom partition directories may cause data loss
- [HIVE-20183](#) - Inserting from bucketed table can cause data loss if the source table contains an empty bucket
- [HIVE-20226](#) - HMS getNextNotification throws exception when request maxEvents exceed table's max_rows

- [HIVE-20345](#) - Drop database may hang if the tables get deleted from a different call

Hue

The following issue is fixed in CDH 6.0.1:

Queries from ImpalaDaemonApi failing when Impala is configured with webserver_htpasswd

When `webserver_htpassword_username` and `webserver_htpassword_password` are used to authenticate the Impala web UIs, the Hue JobBrowser Impala Queries' page returns a 404 error, even with Kerberos authentication.

Workaround: None

Fixed Versions: 6.0.1 6.1.0

Cloudera Issue: CDH-71138

Apache Impala

The following issues are fixed in CDH 6.0.1:

- [IMPALA-4908](#) - NULL floats with different value fields now compare equal.
- [IMPALA-7014](#) - Disabled stacktrace symbolisation by default.
- [IMPALA-7078](#) - Improved memory consumption of Avro scans of wide tables.
- [IMPALA-7145](#) - Fixed a memory leak in OpenSSL when spill-to-disk encryption is enabled.
- [IMPALA-7225](#) - REFRESH PARTITION on a single partition no longer resets its row count to -1.
- [IMPALA-7330](#) - After LOAD DATA, Impala now only refreshes the affected partition.
- [IMPALA-7360](#) - Avro scanner sometimes skips blocks when skip marker is on the HDFS block boundary.
- [IMPALA-7559](#) - Disabled Parquet stat filtering for UTC-normalized timestamp columns.

Apache Kudu

The following issues are fixed in CDH 6.0.1:

- [KUDU-2312](#) - Fixed a crash that could occur on some systems when a query had more than 16 predicates.
- [KUDU-2509](#) - Fixed use-after-free in case of a WAL replay error.

Apache Oozie

The following issues are fixed in CDH 6.0.1:

- [OOZIE-3193](#) - Applications are not killed when submitted via subworkflow
- [OOZIE-3330](#) - and OOZIE-3331 - Spark options parsing bugfix

Cloudera Search

The following issues are fixed in CDH 6.0.1:

- [SOLR-11590](#) - Synchronize ZK connect/disconnect handling so that they are processed in linear order
- [SOLR-12343](#) - JSON Field Facet refinement can return incorrect counts/stats for sorted buckets
- [SOLR-12450](#) - Don't allow referral to external resources in various config files
- [SOLR-12516](#) - JSON range facets can incorrectly refine subfacets for buckets

CDH Upgrade fails to delete Solr data from HDFS

The CDH upgrade process fails to delete Solr data from HDFS and the recreated collections fail to be initialized due to the existing indexes.

Workaround: Perform the following steps *after* you run the CDH Upgrade wizard and *before* you finalize the HDFS upgrade:

1. Log in to the Cloudera Manager Admin Console.
2. Go to the Solr service page.
3. Stop the Solr service and dependent services. Click **Actions > Stop**.
4. Click **Actions > Reinitialize Solr State for Upgrade**.
5. Click **Actions > Bootstrap Solr Configuration**.

6. Start the Solr and dependent services. Click **Actions > Start**.
7. Click **Actions > Bootstrap Solr Collections**.

Affected Versions: CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47502

Solr Service reports stale configurations even after restart

Solr reports stale configurations, and the Solr Server role fails to start with the following error: Role failed to start due to error: The archive already contains creds.localjceks. The issue occurs if your deployment has Solr and HDFS uses LDAP Group Mapping.

Workaround: If you have a CDH 5 cluster and use LDAP Group Mapping, do not upgrade to CDH 6.0.0. If you have a CDH 6.0.0 cluster, disable LDAP Group Mappings.

Affected Versions: Cloudera Manager 6.0.0 and CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47321

Cloudera Search configuration migration script fails to detect incompatible SecureAdminHandlers request handler

The SecureAdminHandlers request handler is incompatible with Apache Solr 7, which is used in CDH 6. The Cloudera Search configuration migration script fails to detect this incompatibility.

Workaround: Remove SecureAdminHandlers request handlers from the solrconfig.xml files of any configuration set that uses them during the [pre-upgrade configuration migration](#).

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Cloudera Issue: CDH-72239

Apache Spark

The following issues are fixed in CDH 6.0.1:

- [SPARK-21525](#) - [STREAMING] Check error code from supervisor RPC.
- [SPARK-23679](#) - [YARN] Setting RM_HA_URLS for AmIpFilter to avoid redirect failure in YARN mode
- [SPARK-25253](#) - [PYSPARK] Refactor local connection & auth code

Apache YARN

The following issues are fixed in CDH 6.0.1:

- [YARN-6966](#) - NodeManager metrics may return wrong negative values when NM restart.
- [YARN-7542](#) - Fix issue that causes some Running Opportunistic Containers to be recovered as PAUSED.
- [YARN-8436](#) - FSParentQueue: Comparison method violates its general contract.
- [YARN-8518](#) - test-container-executor test_is_empty() is broken
- [YARN-8605](#) - TestDominantResourceFairnessPolicy.testModWhileSorting is flaky.

Unsupported Features in CDH 6.0.1

This page lists the unsupported features in CDH 6.0.1. For the complete list of Known Issues and Limitations, see [Known Issues and Limitations in CDH 6.0.1](#) on page 473.

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.x cluster is not currently supported. If you try to upgrade to CDH 6.0.x you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. If you try to upgrade to CDH 6.0.x, you will be asked to remove the CDSW service from your cluster.

Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Apache Hadoop Unsupported Features

The following sections list unsupported features in Hadoop common components:

- [HDFS Unsupported Features](#) on page 437
- [YARN Unsupported Features](#) on page 437

HDFS Unsupported Features

The following HDFS features are not supported in CDH 6.0.x:

- ACLs for the NFS gateway
- Aliyun Cloud Connector
- Erasure Coding
- HDFS NameNode Federation
- HDFS Router Based Federation
- HDFS truncate
- More than two NameNodes
- Openstack Swift
- Quota support for Storage Types
- SFTP FileSystem
- Upgrade Domain
- Variable length block
- ZStandard Compression Codec

YARN Unsupported Features

The following YARN features are not supported in CDH 6.0.x:

- Application Timeline Server v2 (ATSV2)
- Cgroup Memory Enforcement
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Docker on YARN (DockerContainerExecutor)
- Native Services
- New Aggregated Log File Format
- Node Labels
- Pluggable Scheduler Configuration
- Reservation REST APIs
- Resource Estimator Service
- Resource Profiles
- Rolling Log Aggregation
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- YARN WebUI v2

Apache HBase Unsupported Features

The following HBase features are not supported in CDH 6.0.x:

- Master hosting meta
- Cloudera does not provide support for user-provided custom coprocessors of any kind.
- Server-side encryption of HFiles. You should configure HDFS client-side encryption.
- In-memory compaction

The following features, introduced upstream in HBase, are not supported in CDH:

- Visibility labels
- Stripe compaction
- Clients setting priority on operations
- Specifying a custom asynchronous connection implementation

Apache Hive Unsupported Features

The following Hive features are not supported in CDH 6.0.x:

- AccumuloStorageHandler ([HIVE-7068](#))
- ACID ([HIVE-5317](#))
- Built-in `version()` function is not supported (CDH-40979)
- Cost-based Optimizer (CBO)
- Explicit Table Locking
- HCatalog - HBase plugin
- Hive Authorization (Instead, use Apache Sentry.)
- Hive on Apache Tez
- Hive Local Mode Execution
- Hive Metastore - Derby
- Hive Web Interface (HWI)
- HiveServer1 / JDBC 1
- HiveServer2 Dynamic Service Discovery (HS2 HA) ([HIVE-8376](#))
- HiveServer2 - HTTP Mode (Use THRIFT mode.)
- HPL/SQL ([HIVE-11055](#))
- LLAP (Live Long and Process framework)
- Scalable Dynamic Partitioning and Bucketing Optimization ([HIVE-6455](#))
- Session-level Temporary Tables ([HIVE-7090](#))
- Table Replication Across HCatalog Instances ([HIVE-7341](#))

Apache Kafka Unsupported Features

The following Kafka feature is not supported in CDH 6.0.x:

- Idempotent and transactional capabilities in the producer are currently an unsupported beta feature given their maturity and complexity. This feature will be supported in a future release.
- Kafka Connect is included in CDH 6.0.0, but is not supported. Flume and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. See [Flafka: Apache Flume Meets Apache Kafka for Event Processing](#) for more information.
- Kafka Streams is included in CDH 6.0.0, but is not supported. Instead, use Spark and Spark Streaming have a fully functional ETL/stream processing pipeline. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for more information.
- The Kafka default authorizer is included in CDH 6.0.0, but is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.
- Using Kafka with a JBOD setup is an unsupported beta feature given its maturity and complexity. Using JBOD in production will be supported only in a later release.
- The legacy Scala clients (producer and consumer) that are under the `kafka.producer.*` and `kafka.consumer.*` package are deprecated in CDH 6.0.0. See [Deprecated Scala-based Client API and New Java Client API](#) on page 544.

Apache Oozie Unsupported Features

The following Oozie feature is not supported in CDH 6.0.x:

- Conditional coordinator input logic.

Apache Pig Unsupported Features

The following Pig features are not supported in CDH 6.0.x:

- Pig on Tez is not supported in CDH 6.0.x ([PIG-3446](#) / [PIG-3419](#)).
- Pig on Spark.

Cloudera Search Unsupported Features

The following Search features are not supported in CDH 6.0.x:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation is not supported
- Saving search results is not supported

Apache Sentry Unsupported Features

The following Sentry features are not supported in CDH 6.0.x:

- Import and export of Sentry metadata to and from Sentry servers
- Sentry shell command line for Hive
- Relative URI paths ([Known Issue](#))
- Object types Server and URL in

```
show grant role <role name> on object <object name>
```

([Known Issue](#))

- ALTER and DROP privileges for Hive and Impala

In addition, as of CDH 6.0.x, Sentry policy files have been removed. See the Sentry [Incompatible Changes](#) for more information.

Apache Spark Unsupported Features

The following Spark features are not supported in CDH 6.0.x:

- Apache Spark experimental features/APIs are not supported unless stated otherwise.
- Using the JDBC Datasource API to access Hive or Impala is not supported
- ADLS not Supported for All Spark Components. Microsoft Azure Data Lake Store (ADLS) is a cloud-based filesystem that you can access through Spark applications. Spark with Kudu is not currently supported for ADLS data. (Hive on Spark is available for ADLS in CDH 5.12 and higher.)
- IPython / Jupyter notebooks is not supported. The IPython notebook system (renamed to Jupyter as of IPython 4.0) is not supported.
- Certain Spark Streaming features not supported. The `mapWithState` method is unsupported because it is a nascent unstable API.
- Thrift JDBC/ODBC server is not supported
- Spark SQL CLI is not supported
- GraphX is not supported
- SparkR is not supported
- Structured Streaming is not supported.
- Spark cost-based optimizer (CBO) not supported.

Apache Sqoop Unsupported Features

The following Sqoop feature is not supported in CDH 6.0.x:

- [import-mainframe](#)

Incompatible Changes in CDH 6.0.1



Important:

In addition to incompatible changes, CDH 6 also deprecated or removed support for several components, including Spark 1 and MapReduce v1. For information about components, sub-components, or functionality that are deprecated or no longer supported, see [Deprecated Items](#) on page 667.

See below for incompatible changes in CDH 6.0.1, grouped by component:

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.x cluster is not currently supported. If you try to upgrade to CDH 6.0.x you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Apache Avro

API Changes

One method was removed in CDH 6.0.0:

```
GenericData.toString (Object datum, StringBuilder buffer)
```

Incompatible Changes from Avro 1.8.0

- Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.
- [AVRO-997](#): Generic API requires GenericEnumSymbol - likely to break current Generic API users that often have String or Java Enum for these fields
- [AVRO-1502](#): Avro Objects now Serializable - IPC needs to be regenerated/recompiled
- [AVRO-1602](#): removed Avro internal RPC tracing, presumed unused. Current rec would be HTrace
- [AVRO-1586](#): Compile against Hadoop 2 - probably not an issue since we've been compiling against Hadoop 2 for C5.
- [AVRO-1589](#): [Java] ReflectData.AllowNulls will create incompatible Schemas for primitive types - may need a KI since it used to fail at runtime but now will fail earlier.

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. If you try to upgrade to CDH 6.0.x, you will be asked to remove the CDSW service from your cluster.

Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Cloudera Issue: DSE-2769

Apache Crunch



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

The following changes are introduced in CDH 6.0.0, and are not backward compatible:

- Crunch is available only as Maven artifacts from the Cloudera Maven repository. It is not included as part of CDH. For more information, see [Apache Crunch Guide](#).

- Crunch supports only Spark 2 and higher releases.
- Crunch supports only HBase 2 and higher releases.
 - The API methods in Crunch-HBase use HBase 2 API types and methods.

Apache Flume

AsyncHBaseSink and HBaseSink

CDH 6 uses HBase 2.0. AsyncHBaseSink is incompatible with HBase 2.0 and is not supported in CDH 6. HBaseSink has been replaced with HBase2Sink. HBase2Sink works the same way as HBaseSink. The only difference is that it is compatible with HBase 2.0. The only additional configuration required to use HBase2Sink is to replace the component name in your configuration.

For example, replace this text:

```
agent.sinks.my_hbase_sink.type = hbase
```

With this:

```
agent.sinks.my_hbase_sink.type = hbase2
```

Or, if you use the FQN of the sink class, replace this text:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase.HBaseSink
```

With this:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase2.HBase2Sink
```

For more information about how to configure HBase2Sink, see [Importing Data Into HBase](#).

com.google.common.collect.ImmutableMap

Flume has removed `com.google.common.collect.ImmutableMap` from the `org.apache.flume.Context` API and replaced it with `java.util.Map` due to Guava compatibility issues ([FLUME-2957](#)). Plugins using the `Context.getParameters()` and `Context.getSubProperties()` APIs will need to assign the return value of those methods to a `Map<String, String>` variable instead of an `ImmutableMap<String, String>` variable, if they do not already do so. Most usages in the Flume codebase already used `Map<String, String>` at the time of this change.

Apache Hadoop

- [HDFS Incompatible Changes](#) on page 441
- [MapReduce](#) on page 442
- [YARN](#) on page 443

HDFS Incompatible Changes

CDH 6.0.1 introduces no new incompatible changes for HDFS.

CDH 6.0.0, introduces the following incompatible changes for HDFS:

- HFTP has been removed.
- The S3 and S3n connectors have been removed. Users should now use the S3a connector.
- The BookkeeperJournalManager has been removed.
- Changes were made to the structure of the HDFS JAR files to better isolate clients from Hadoop library dependencies. As a result, client applications that depend on Hadoop's library dependencies may no longer work. In these cases, the client applications will need to include the libraries as dependencies directly.
- Several library dependencies were upgraded. Clients that depend on those libraries may break because the library version changes. In these cases, the client applications will need to either be ported to the new library versions or include the libraries as dependencies directly.

- [HDFS-6962](#) changes the behavior of ACL inheritance to better align with POSIX ACL specifications, which states that the umask has no influence when a default ACL propagates from parent to child. Previously, HDFS ACLs applied the client's umask to the permissions when inheriting a default ACL defined on a parent directory. Now, HDFS can ignore the umask in these cases for improved compliance with POSIX. This behavior is on by default due to the inclusion of [HDFS-11957](#). It can be configured by setting `dfs.namenode posix.acl.inheritance.enabled` in `hdfs-site.xml`. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-11957](#) changes the default behavior of ACL inheritance introduced by [HDFS-6962](#). Previously, the behavior was disabled by default. Now, the feature is enabled by default. Any code expecting the old ACL inheritance behavior will have to be updated. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-6252](#) removed `dfshealth.jsp` since it is part of the old NameNode web UI. By default, Cloudera Manager links to the new NameNode web UI, which has an equivalent health page at `dfshealth.html`.
- [HDFS-11100](#) changes the behavior of deleting files protected by a sticky bit. Now, the deletion fails.
- [HDFS-10689](#) changes the behavior of the `hdfs dfs chmod` command. Now, the command resets sticky bit permission on a file/directory when the leading sticky bit is omitted in the octal mode (like 644). When a file or directory permission is applied using octal mode and sticky bit permission needs to be preserved, then it has to be explicitly mentioned in the permission bits (like 1644).
- [HDFS-10650](#) changes the behavior of `DFSClient#mkdirs` and `DFSClient#primitiveMkdir`. Previously, they create a new directory with the default permissions 00666. Now, they will create a new directory with permission 00777.
- [HADOOP-8143](#) changes the default behavior of `distcp`. Previously, the `-pb` option was not used by default, which may have caused some checksums to fail when block sizes did not match. Now, the `-pb` option is included by default to preserve block size when using `distcp`.
- [HADOOP-10950](#) changes several heap management variables:
 - `HADOOP_HEAPSIZE` variable has been deprecated. Use `HADOOP_HEAPSIZE_MAX` and `HADOOP_HEAPSIZE_MIN` instead to set `Xmx` and `Xms`
 - The internal variable `JAVA_HEAP_MAX` has been removed.
 - Default heap sizes have been removed. This will allow for the JVM to use auto-tuning based upon the memory size of the host. To re-enable the old default, configure `HADOOP_HEAPSIZE_MAX="1g"` in `hadoop-env.sh`.
 - All global and daemon-specific heap size variables now support units. If the variable is only a number, the size is assumed to be in megabytes.
- [HADOOP-14426](#) upgrades the version of Kerby from 1.0.0-RC2 to 1.0.0
- [HDFS-10970](#) updates the version of Jackson from 1.9.13 to 2.x in `hadoop-hdfs`.
- [HADOOP-9613](#) updates the Jersey version to the latest 1.x release.
- [HADOOP-10101](#) updates Guava dependency to 21.0
- [HADOOP-14225](#) removes the `xmlenc` dependency. If you rely on the transitive dependency, you need to set the dependency explicitly in your code after this change.
- [HADOOP-13382](#) remove unneeded `commons-httpclient` dependencies from POM files in Hadoop and sub-projects. This incompatible change may affect projects that have undeclared transitive dependencies on `commons-httpclient`, which used to be provided by `hadoop-common` or `hadoop-client`.
- [HADOOP-13660](#) upgrades the `commons-configuration` version from 1.6 to 2.1.
- [HADOOP-12064](#) upgrades the following dependencies:
 - Guice from 3.0 to 4.0
 - cglib from 2.2 to 3.2.0
 - asm from 3.2 to 5.0.4

MapReduce

CDH 6.0.1 introduces no new incompatible changes for MapReduce.

CDH 6.0.0, introduces the following incompatible changes:

- Support for MapReduce v1 has been dropped from CDH 6.0.0.

- CDH 6 supports applications compiled against CDH 5.7.0 and higher MapReduce frameworks. Make sure to not to include the CDH jars with your application by marking them as "provided" in the `pom.xml` file.

YARN

CDH 6.0.1 introduces no new incompatible changes for YARN.

CDH 6.0.0 introduces no new incompatible changes for YARN.

Apache HBase

CDH 6.0.x contains the following downstream HBase incompatible change:

`hbase.security.authorization`

The default value for `hbase.security.authorization` has been changed from true to false. Secured clusters should make sure to explicitly set it to true in XML configuration file before upgrading to one of these versions ([HBASE-19483](#)). True as the default value of `hbase.security.authorization` was changed because not all clusters need authorization. (History: [HBASE-13275](#)) Rather, only the clusters which need authorization should set this configuration as true.

Incompatible Changes

For more information about upstream incompatible changes, see the Apache Reference Guide [Incompatible Changes](#) and [Upgrade Paths](#).

CDH 6.0.x contains the following upstream HBase incompatible changes:

- Public interface API changes:
 - [HBASE-15607](#): Admin
 - [HBASE-19112](#), [HBASE-18945](#): Cell
 - Region, Store, HBaseTestingUtility
- [HBASE-18792](#): hbase-2 needs to defend against hbck operations
- [HBASE-15982](#): Interface ReplicationEndpoint extends Guava's Service.
- [HBASE-18995](#): Split CellUtil into public CellUtil and PrivateCellUtil for Internal use only.
- [HBASE-19179](#): Purged the hbase-prefix-tree module and all references from the code base.
- [HBASE-17595](#): Add partial result support for small/limited scan; Now small scan and limited scan could also return partial results.
- [HBASE-16765](#): New default split policy, SteppingSplitPolicy.
- [HBASE-17442](#): Move most of the replication related classes from hbase-client to hbase-replication package.
- [HBASE-16196](#): The bundled JRuby 1.6.8 has been updated to version 9.1.9.0.
- [HBASE-18811](#): Filters have been moved from Public to LimitedPrivate.
- [HBASE-18697](#): Replaced hbase-shaded-server jar with hbase-shaded-mapreduce jar.
- [HBASE-18640](#): Moved mapreduce related classes out of hbase-server into separate hbase-mapreduce jar .
- [HBASE-19128](#): Distributed Log Replay feature has been removed.
- [HBASE-19176](#): Hbase-native-client has been removed.
- [HBASE-17472](#): Changed semantics of granting new permissions. Earlier, new grants would override previous permissions, but now, the new and existing permissions get merged.
- [HBASE-18374](#): Previous "mutate" latency metrics has been renamed to "put" metrics.
- [HBASE-15740](#): Removed Replication metric source.shippedKBs in favor of source.shippedBytes.
- [HBASE-13849](#): Removed restore and clone snapshot from the WebUI.
- [HBASE-13252](#): The concept of managed connections in HBase (deprecated before) has now been extinguished completely, and now all callers are responsible for managing the lifecycle of connections they acquire.
- [HBASE-14045](#): Bumped thrift version to 0.9.2.
- [HBASE-5401](#): Changes to number of tasks PE runs when clients are mapreduce. Now tasks == client count. Previous we hardcoded ten tasks per client instance.

Changed Behavior

CDH 6.0.x contains the following HBase behavior changes:

- [HBASE-14350](#): Assignment Manager v2 - Split/Merge have moved to the Master; it runs them now. Hooks around Split/Merge are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-18271](#): Moved to internal shaded netty.
- [HBASE-17343](#): Default MemStore to be CompactingMemStore instead of DefaultMemStore. In-memory compaction of CompactingMemStore demonstrated sizable improvement in HBase's write amplification and read/write performance.
- [HBASE-19092](#): Make Tag IA.LimitedPrivate and expose for CPs.
- [HBASE-18137](#): Replication gets stuck for empty WALs.
- [HBASE-17513](#): Thrift Server 1 uses different QOP settings than RPC and Thrift Server 2 and can easily be misconfigured so there is no encryption when the operator expects it.
- [HBASE-16868](#): Add a replicate_all flag to replication peer config. The default value is true, which means all user tables (REPLICATION_SCOPE != 0) will be replicated to peer cluster.
- [HBASE-19341](#): Ensure Coprocessors can abort a server.
- [HBASE-18469](#): Correct RegionServer metric of totalRequestCount.
- [HBASE-17125](#): Marked Scan and Get's setMaxVersions() and setMaxVersions(int) as deprecated. They are easy to misunderstand with column family's max versions, so use readAllVersions() and readVersions(int) instead.
- [HBASE-16567](#): Core is now up on protobuf 3.1.0 (Coprocessor Endpoints and REST are still on protobuf 2.5.0).
- [HBASE-14004](#): Fix inconsistency between Memstore and WAL which may result in data in remote cluster that is not in the origin (Replication).
- [HBASE-18786](#): FileNotFoundException opening a StoreFile in a primary replica now causes a RegionServer to crash out where before it would be ignored (or optionally handled via close/reopen).
- [HBASE-17956](#): Raw scans will also read TTL expired cells.
- [HBASE-17017](#): Removed per-region latency histogram metrics.
- [HBASE-19483](#): Added ACL checks to RSGroup commands - On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands.
- [HBASE-19358](#): Added ACL checks to RSGroup commands (HBASE-19483): On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands. Improved stability of splitting log when do failover.
- [HBASE-18883](#): Updated our Curator version to 4.0 - Users who experience classpath issues due to version conflicts are recommended to use either the hbase-shaded-client or hbase-shaded-mapreduce artifacts.
- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.
- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.
- [HBASE-15968](#): This issue resolved two long-term issues in HBase: 1) Puts may be masked by a delete before them, and 2) Major compactions change query results. Offers a new behavior to fix this issue with a little performance reduction. Disabled by default. See the issue for details and caveats.
- [HBASE-13701](#): SecureBulkLoadEndpoint has been integrated into HBase core as default bulk load mechanism. It is no longer needed to install it as a coprocessor endpoint.
- [HBASE-9774](#): HBase native metrics and metric collection for coprocessors.
- [HBASE-18294](#): Reduce global heap pressure: flush based on heap occupancy.

Apache Hive/Hive on Spark/HCatalog

Apache Hive

The following changes are introduced to Hive in CDH 6.0.0, and are not backwards compatible:

- [UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting](#) on page 533
- [Support for UNION DISTINCT](#) on page 534
- [OFFLINE and NO_DROP Options Removed from Table and Partition DDL](#) on page 534
- [DESCRIBE Query Syntax Change](#) on page 535
- [CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names](#) on page 535
- [Reserved and Non-Reserved Keyword Changes in HiveQL](#) on page 535

- [Apache Hive API Changes in CDH 6.0.0](#) on page 537
- [Apache Hive Configuration Changes in CDH 6.0.0](#) on page 538
- [HiveServer2 Thrift API Repackaged Resulting in Class File Location Changes](#) on page 540
- [Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column](#) on page 541
- [Hive Logging Framework Switched to SLF4J/Log4j 2](#) on page 541
- [Deprecated Parquet Java Classes Removed from Hive](#) on page 541
- [Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms](#) on page 541
- [S3N Connector Is Removed from CDH 6.0](#) on page 542
- [Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request](#) on page 542
- [Support Added for Escaping Carriage Returns and New Line Characters for Text Files \(LazySimpleSerDe\)](#) on page 542
- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 542

Changing Table File Format from ORC with the ALTER TABLE Command Not Supported in CDH 6

Changing the table file format from ORC to another file format with the ALTER TABLE command is not supported in CDH 6 (it returns an error).

UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting

Prior to this change, Hive performed implicit casts when data types from different type groups were specified in queries that use UNION ALL. For example, before CDH 6.0, if you had the two following tables:

Table "one"

one.col_1	one.col_2	one.col_3
21	hello_all	b

Where col_1 datatype is int, col_2 datatype is string, and col_3 datatype is char(1).

Table "two"

two.col_4	two.col_5	two.col_6
75.0	abcde	45

Where col_4 datatype is double, col_5 datatype is varchar(5), and col_6 datatype is int.

And you ran the following UNION ALL query against these two tables:

```
SELECT * FROM one UNION ALL SELECT col_4 AS col_1, col_5 AS col_2, col_6 AS col_3 FROM two;
```

You received the following result set:

_u1.col_1	_u1.col_2	_u1.col_3
75.0	abcde	4
21.0	hello	b

```
+-----+-----+-----+-----+
```

Note that this statement implicitly casts the values from table `one` with the following errors resulting in data loss:

- `one.col_1` is cast to a `double` datatype
- `one.col_2` is cast to a `varchar(5)` datatype, which truncates the original value from `hello_all` to `hello`
- `one.col_3` is cast to a `char(1)` datatype, which truncates the original value from `45` to `4`

In CDH 6.0, no implicit cast is performed across different type groups. For example, `STRING`, `CHAR`, and `VARCHAR` are in one type group, and `INT`, `BIGINT`, and `DECIMAL` are in another type group, and so on. So, in CDH 6.0 and later, the above query that uses `UNION ALL`, returns an exception for the columns that contain datatypes that are not part of a type group. In CDH 6.0 and later, Hive performs the implicit cast only *within* type groups and not *across* different type groups. For more information, see [HIVE-14251](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
- The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
- The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

OFFLINE and NO_DROP Options Removed from Table and Partition DDL

Support for Hive table and partition protection options have been removed in CDH 6.0, which includes removal of the following functionality:

- Support has been removed for:

- `ENABLE | DISABLE NO_DROP [CASCADE]`
- `ENABLE | DISABLE OFFLINE`
- `ALTER TABLE ... IGNORE PROTECTION`

- The following support has also been removed from the `HiveMetastoreClient` class:

The `ignoreProtection` parameter has been removed from the `dropPartitions` methods in the `IMetaStoreClient` interface.

For more information, see [HIVE-11145](#).

Cloudera recommends that you use Apache Sentry to replace most of this functionality. Although Sentry governs permissions on `ALTER TABLE`, it does not include permissions that are specific to a partition. See [Authorization Privilege Model for Hive and Impala](#) and [Configuring the Sentry Service](#).

DESCRIBE Query Syntax Change

In CDH 6.0 syntax has changed for `DESCRIBE` queries as follows:

- `DESCRIBE` queries where the column name is separated by the table name using a period is no longer supported:

```
DESCRIBE testTable.testColumn;
```

Instead, the table name and column name must be separated with a space:

```
DESCRIBE testTable testColumn;
```

- The `partition_spec` must appear *after* the table name, but *before* the optional column name:

```
DESCRIBE default.testTable PARTITION (part_col = 100) testColumn;
```

For more details, see the [Apache wiki](#) and [HIVE-12184](#).

CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names

In CDH 6.0, `CREATE TABLE` statements fail if any of the specified column names contain a period or a colon. For more information, see [HIVE-10120](#) and the [Apache wiki](#).

Reserved and Non-Reserved Keyword Changes in HiveQL

Hive reserved and non-reserved keywords have changed in CDH 6.0. *Reserved keywords* cannot be used as table or column names unless they are enclosed with back ticks (for example, `data`). *Non-reserved keywords* can be used as table or column names without enclosing them with back ticks. Non-reserved keywords have proscribed meanings in HiveQL, but can still be used as table or column names. For more information about the changes to reserved and non-reserved words listed below, see [HIVE-6617](#) and [HIVE-14872](#).

In CDH 6.0, the following changes have been introduced to Hive reserved and non-reserved keywords and are not backwards compatible:

- [Hive New Reserved Keywords Added in CDH 6.0](#) on page 535
- [Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0](#) on page 536
- [Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0](#) on page 536
- [Hive New Non-Reserved Keywords Added in CDH 6.0](#) on page 536
- [Hive Non-Reserved Keyword Removed in CDH 6.0](#) on page 537

Hive New Reserved Keywords Added in CDH 6.0

The following table contains new reserved keywords that have been added:

COMMIT	CONSTRAINT	DEC	EXCEPT
FOREIGN	INTERVAL	MERGE	NUMERIC
ONLY	PRIMARY	REFERENCES	ROLLBACK
START			

Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0

The following table contains non-reserved keywords that have been converted to be reserved keywords:

ALL	ALTER	ARRAY	AS
AUTHORIZATION	BETWEEN	BIGINT	BINARY
BOOLEAN	BOTH	BY	CREATE

CUBE	CURSOR	DATE	DECIMAL
DOUBLE	DELETE	DESCRIBE	DROP
EXISTS	EXTERNAL	FALSE	FETCH
FLOAT	FOR	FULL	GRANT
GROUP	GROUPING	IMPORT	IN
INT	INNER	INSERT	INTERSECT
INTO	IS	LATERAL	LEFT
LIKE	LOCAL	NONE	NULL
OF	ORDER	OUT	OUTER
PARTITION	PERCENT	PROCEDURE	RANGE
READS	REGEXP	REVOKE	RIGHT
RLIKE	ROLLUP	ROW	ROWS
SET	SMALLINT	TABLE	TIMESTAMP
TO	TRIGGER	TRUNCATE	UNION
UPDATE	USER	USING	VALUES
WITH	TRUE		

Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0

The following table contains reserved keywords that have been converted to be non-reserved keywords:

CURRENT_DATE	CURRENT_TIMESTAMP	HOLD_DDLTIME	IGNORE
NO_DROP	OFFLINE	PROTECTION	READONLY

Hive New Non-Reserved Keywords Added in CDH 6.0

The following table contains new non-reserved keywords that have been added:

ABORT	AUTOCOMMIT	CACHE	DAY
DAYOFWEEK	DAYS	DETAIL	DUMP
EXPRESSION	HOUR	HOURS	ISOLATION
KEY	LAST	LEVEL	MATCHED
MINUTE	MINUTES	MONTH	MONTHS
NORELY	NOVALIDATE	NULLS	OFFSET
OPERATOR	RELY	SECOND	SECONDS
SNAPSHOT	STATUS	SUMMARY	TRANSACTION
VALIDATE	VECTORIZATION	VIEWS	WAIT
WORK	WRITE	YEAR	YEARS

Hive Non-Reserved Keyword Removed in CDH 6.0

The following non-reserved keyword has been removed:

DEFAULT

Apache Hive API Changes in CDH 6.0.0

The following changes have been introduced to the Hive API in CDH 6.0, and are not backwards compatible:

- [AddPartitionMessage.getPartitions\(\) Can Return NULL](#) on page 537
- [DropPartitionEvent and PreDropPartitionEvent Class Changes](#) on page 537
- [GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date](#) on page 537
- [GenericUDF.getConstantLongValue Has Been Removed](#) on page 537
- [Increased Width of Hive Metastore Configuration Columns](#) on page 537

AddPartitionMessage.getPartitions() Can Return NULL

The `getPartitions()` method has been removed from the `AddPartitionEvent` class in the `org.apache.hadoop.hive.metastore.events` interface. It was removed to prevent out-of-memory errors when the list of partitions is too large.

Instead use the `getPartitionIterator()` method. For more information, see [HIVE-9609](#) and the [AddPartitionEvent documentation](#).

DropPartitionEvent and PreDropPartitionEvent Class Changes

The `getPartitions()` method has been removed and replaced by the `getPartitionIterator()` method in the `DropPartitionEvent` class and the `PreDropPartitionEvent` class.

In addition, the `(Partition partition, boolean deleteData, HiveMetastore.HMSHandler handler)` constructors have been deleted from the `PreDropPartitionEvent` class. For more information, see [HIVE-9674](#) and the [PreDropPartitionEvent documentation](#).

GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date

The `getTimestampValue` method in the `GenericUDF` class now returns a `TIMESTAMP` value instead of a `DATE` value. For more information, see [HIVE-10275](#) and the [GenericUDF documentation](#).

GenericUDF.getConstantLongValue Has Been Removed

The `getConstantLongValue` method has been removed from the `GenericUDF` class. It has been noted by the community that this method is not used in Hive. For more information, see [HIVE-10710](#) and the [GenericUDF documentation](#).

Increased Width of Hive Metastore Configuration Columns

The columns used for configuration values in the Hive metastore have been increased in width, resulting in the following incompatible changes in the `org.apache.hadoop.hive.metastore.api` interface.

This change introduced an incompatible change to the `get_table_names_by_filter` method of the `ThriftHiveMetastore` class. Before this change, this method accepts a string filter, which allows clients to filter a table by its `TABLEPROPERTIES` value. For example:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 <> \"yellow\";
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 = \"yellow\";
```

After this change, the `TABLE_PARAMS.PARAM_VALUE` column is now a CLOB data type. Depending on the type of database that you use (for example, MySQL, Oracle, or PostgreSQL), the semantics may have changed and operators like `"=", "<>",` and `"!="` might not be supported. Refer to the documentation for your database for more information. You must use operators that are compatible with CLOB data types. There is no equivalent `<>` operator that is compatible with CLOB. So there is no equivalent operator for the above example that uses the `<>` inequality operator. The equivalent for `"=` is the `LIKE` operator so you would rewrite the second example above as:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
PARAMS + "test_param_1 LIKE \"yellow\";
```

```
PARAMS + "test_param_1 LIKE \"yellow\"";
```

For more information, see [HIVE-12274](#).

Apache Hive Configuration Changes in CDH 6.0.0

The following configuration property changes have been introduced to Hive in CDH 6.0, and are not backwards compatible:

- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 538
- [Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters](#) on page 538
- [Hive Strict Checks Have Been Re-factored To Be More Granular](#) on page 539
- [Java XML Serialization Has Been Removed](#) on page 539
- [Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties](#) on page 540
- [HiveServer2 Impersonation Property \(hive.server2.enable.impersonation\) Removed](#) on page 540
- [Changed Default File Format for Storing Intermediate Query Results](#) on page 540

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on hive.enforce.bucketing](#) and the [topic on hive.enforce.sorting](#).

Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters

Directories in HDFS can contain unprintable or unsupported characters that are not visible even when you run the `hadoop fs -ls` command on the directories. When external tables are created with the `MSCK REPAIR TABLE` command, the partitions using these HDFS directories that contain unsupported characters are unusable for Hive. To avoid this, the configuration parameter `hive.msck.path.validation` has been added. This configuration property controls the behavior of the `MSCK REPAIR TABLE` command, enabling you to set whether validation checks are run on the HDFS directories when `MSCK REPAIR TABLE` is run.

The property `hive.msck.path.validation` can be set to one of the following values:

Value Name	Description
throw	Causes Hive to throw an exception when it tries to process an HDFS directory that contains unsupported characters with the <code>MSCK REPAIR TABLE</code> command. This is the default setting for <code>hive.msck.path.validation</code> .
skip	Causes Hive to skip the skip the directories that contain unsupported characters, but still repairs the others.
ignore	Causes Hive to completely skip any validation of HDFS directories when the <code>MSCK REPAIR TABLE</code> command is run. This setting can cause bugs because unusable partitions are created.

By default, the `hive.msck.path.validation` property is set to `throw`, which causes Hive to throw an exception when `MSCK REPAIR TABLE` is run and HDFS directories containing unsupported characters are encountered. To work around this, set this property to `skip` until you can repair the HDFS directories that contain unsupported characters.

To set this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the **Configuration** tab.

3. Search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting.
4. In the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting, add the **Name** of the property, the **Value** (throw, skip, or ignore), and a **Description** of the setting.
5. Click **Save Changes** and restart the service.

For more information, see [HIVE-10722](#).

Hive Strict Checks Have Been Re-factored To Be More Granular

Originally, the configuration property `hive.mapred.mode` was added to restrict certain types of queries from running. Now it has been broken down into more fine-grained configurations, one for each type of restricted query pattern. The configuration property `hive.mapred.mode` has been removed and replaced with the following configuration properties, which provide more granular control of Hive strict checks:

Configuration Property	Description	Default Value
<code>hive.strict.checks.bucketing</code>	When set to <code>true</code> , running <code>LOAD DATA</code> queries against bucketed tables is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.type.safety</code>	When set to <code>true</code> , comparing <code>bigint</code> to <code>string</code> data types or <code>bigint</code> to <code>double</code> data types is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.orderby.no.limit</code>	When set to <code>true</code> , prevents queries from being run that contain an <code>ORDER BY</code> clause with no <code>LIMIT</code> clause.	<code>false</code>
<code>hive.strict.checks.no.partition.filter</code>	When set to <code>true</code> , prevents queries from being run that scan a partitioned table but do not filter on the partition column.	<code>false</code>
<code>hive.strict.checks.cartesian.product</code>	When set to <code>true</code> , prevents queries from being run that contain a Cartesian product (also known as a cross join).	<code>false</code>

All of these properties can be set with Cloudera Manager in the following configuration settings for the Hive service:

- **Restrict LOAD Queries Against Bucketed Tables** (`hive.strict.checks.bucketing`)
- **Restrict Unsafe Data Type Comparisons** (`hive.strict.checks.type.safety`)
- **Restrict Queries with ORDER BY but no LIMIT clause** (`hive.strict.checks.orderby.no.limit`)
- **Restrict Partitioned Table Scans with no Partitioned Column Filter** (`hive.strict.checks.no.partition.filter`)
- **Restrict Cross Joins (Cartesian Products)** (`hive.strict.checks.cartesian.product`)

For more information about these configuration properties, see [HIVE-12727](#), [HIVE-15148](#), [HIVE-18251](#), and [HIVE-18552](#).

Java XML Serialization Has Been Removed

The configuration property `hive.plan.serialization.format` has been removed. Previously, this configuration property could be set to either `javaXML` or `kryo`. Now the default is `kryo` serialization, which cannot be changed. For more information, see [HIVE-12609](#) and the [Apache wiki](#).

Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties

The configuration property `hive.groupby.orderby.position.alias`, which enabled using column position with the `GROUP BY` and the `ORDER BY` clauses has been removed and replaced with the following two configuration properties. These configuration properties enable using column position with `GROUP BY` and `ORDER BY` separately:

Configuration Property Name	Description/Default Setting	Possible Values
hive.groupby.position.alias	When set to true, specifies that columns can be referenced with their position when using GROUP BY clauses in queries. Default Setting: false. This behavior is turned off by default.	true false
hive.orderby.position.alias	When set to true, specifies that columns can be referenced with their position when using ORDER BY clauses in queries. Default Setting: true. This behavior is turned on by default.	true false

For more information, see [HIVE-15797](#) and the Apache wiki entries for [configuration properties](#), [GROUP BY syntax](#), and [ORDER BY syntax](#).

HiveServer2 Impersonation Property (hive.server2.enable.impersonation) Removed

In earlier versions of CDH, the following two configuration properties could be used to set impersonation for HiveServer2:

- `hive.server2.enable.impersonation`
- `hive.server2.enable.doAs`

In CDH 6.0, `hive.server2.enable.impersonation` is removed. To configure impersonation for HiveServer2, use the configuration property `hive.server2.enable.doAs`. To set this property in Cloudera Manager, select the Hive service and click on the **Configuration** tab. Then search for the **HiveServer2 Enable Impersonation** setting and select the checkbox to enable HiveServer2 impersonation. This property is enabled by default in CDH 6.

For more information about this property, see the [Apache wiki documentation for HiveServer2 configuration properties](#).

Changed Default File Format for Storing Intermediate Query Results

The configuration property `hive.query.result.fileformat` controls the file format in which a query's intermediate results are stored. In CDH 6, the default setting for this property has been changed from `TextFile` to `SequenceFile`.

To change this configuration property in Cloudera Manager:

1. In the Admin Console, select the Hive service and click on the **Configuration** tab.
2. Then search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting and add the following information:
 - **Name:** `hive.query.result.fileformat`
 - **Value:** Valid values are `TextFile`, `SequenceFile` (default), or `RCfile`
 - **Description:** Sets the file format in which a query's intermediate results are stored.
3. After you add this information, click **Save Changes** and restart the Hive service.

For more information about this parameter, see the [Apache wiki](#).

HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes

HiveServer2 Thrift API code has been repackaged in CDH 6.0, resulting in the following changes:

- All files generated by the Thrift API for HiveServer2 have moved from the following *old* namespace:
`org.apache.hive.service.cli.thrift`
To the following *new* namespace:
`org.apache.hive.service.rpc.thrift`
- All files generated by the Thrift API for HiveServer2 have moved into a separate jar file called `service-rpc`.

As a result of these changes, all Java classes such as `TCLIService.java`, `TOpenSessionReq.java`, `TSessionHandle.java`, and `TGetSchemasReq.java` have changed locations. For more information, see [HIVE-12442](#).

Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column

Decimal values that are returned in query results are now padded with trailing zeroes to match the specified scale of the corresponding column. For example, *before* this change, when Hive read a decimal column with a specified scale of 5, the value returned for zero was returned as 0. *Now*, the value returned for zero is 0 . 00000. For more information, see [HIVE-12063](#).

Hive Logging Framework Switched to SLF4J/Log4j 2

The logging framework for Hive has switched to [SLF4J \(Simple Logging Facade for Java\)](#) and now uses [Log4j 2](#) by default. Use of Log4j 1.x, Apache Commons Logging, and `java.util.logging` have been removed. To accommodate this change, write all Log4j configuration files to be compatible with Log4j 2.

For more information, see [HIVE-12237](#), [HIVE-11304](#), and the [Apache wiki](#).

Deprecated Parquet Java Classes Removed from Hive

The deprecated parquet classes, `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` have been removed from Hive because they resided outside of the `org.apache` namespace. Any existing tables that use these classes are automatically migrated to the new SerDe classes when the metastore is upgraded.

Use one of the following options for specifying the Parquet SerDe for new Hive tables:

- Specify in the `CREATE TABLE` statement that you want it stored as Parquet. For example:

```
CREATE TABLE <parquet_table_name> (coll INT, col2 STRING) STORED AS PARQUET;
```

- Set the `INPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat` and set the `OUTPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat`. For example:

```
CREATE TABLE <parquet_table_name> (coll INT, col2 STRING)
STORED AS
  INPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat"
  OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat";
```

For more information, see [HIVE-6757](#) and the [Apache wiki](#).

Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms

Support for JDBC, counter-based, and HBase-based statistics collection mechanisms has been removed from Hive. The following configuration properties are no longer supported:

- `hive.stats.dbclass`
- `hive.stats.retries.wait`
- `hive.stats.retries.max`
- `hive.stats.jdbc.timeout`
- `hive.stats.dbconnectionstring`
- `hive.stats.jdbcdrive`
- `hive.stats.key.prefix.reserve.length`

This change also removed the `cleanUp(String keyPrefix)` method from the [StatsAggregator interface](#).

Now all Hive statistics are collected on the default file system. For more information, see [HIVE-12164](#), [HIVE-12411](#), [HIVE-12005](#), and the [Apache wiki](#).

S3N Connector Is Removed from CDH 6.0

The [S3N connector](#), which is used to connect to the Amazon S3 file system from Hive has been removed from CDH 6.0. To connect to the S3 file system from Hive in CDH 6.0, you must now use the S3A connector. There are a number of differences between the S3N and the S3A connectors, including configuration differences. See the [Apache wiki page on integrating with Amazon Web Services](#) for details.

Migration involves making the following changes:

- Changing all metastore data containing URIs that start with `s3n://` to `s3a://`. This change is performed automatically when you upgrade the Hive metastore.
- Changing all scripts containing URIs that start with `s3n://` to `s3a://`. You must perform this change manually.

Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request

Six additional columns have been added to the `TRowSet` that is returned by the `TCLIService#GetTables` request. These columns were added to comply with the official JDBC API. For more information, see the documentation for [java.sql.DatabaseMetaData](#).

The columns added are:

Column Name	Description
REMARKS	Explanatory comment on the table.
TYPE_CAT	Types catalog.
TYPE_SCHEMA	Types schema.
TYPE_NAME	Types name.
SELF_REFERENCING_COL_NAME	Name of the designed identifier column of a typed table.
REF_GENERATION	Specifies how values in the <code>SELF_REFERENCING_COL_NAME</code> column are created.

For more information, see [HIVE-7575](#).

Support Added for Escaping Carriage Returns and New Line Characters for Text Files (LazySimpleSerDe)

Support has been added for escaping carriage returns and new line characters in text files by modifying the `LazySimpleSerDe` class. Without this change, carriage returns and new line characters are interpreted as delimiters, which causes incorrect query results.

This feature is controlled by the SerDe property `serialization.escape.crlf`. It is enabled (set to `true`) by default. If `serialization.escape.crlf` is enabled, '`r`' or '`n`' cannot be used as separators or field delimiters.

This change only affects text files and removes the `getNullString` method from the [LazySerDeParameters class](#). For more information, see [HIVE-11785](#).

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on hive.enforce.bucketing](#) and the [topic on hive.enforce.sorting](#).

Hue

There are no incompatible changes in this release.

Apache Impala

List of Reserved Words Updated

The list of [reserved words](#) in Impala was updated in CDH 6.0.

If you need to use the reserved words from previous versions of CDH, set the `impalad` and `catalogd` startup option, `reserved_words_version`, to "2.11.0".

Decimal V2 Used by Default

In Impala, two different behaviors of `DECIMAL` types are supported. In CDH 6.0, `DECIMAL V2` is used by default. See [DECIMAL Type](#) for detail information.

If you need to continue using the first version of the `DECIMAL` type for the backward compatibility of your queries, set the `DECIMAL_V2` query option to `FALSE`.

Behavior of Column Aliases Changed

To conform to the SQL standard, Impala no longer performs alias substitution in the subexpressions of `GROUP BY`, `HAVING`, and `ORDER BY`.

For example, the following statements will now result in syntax errors.

```
SELECT int_col / 2 AS x
FROM functional.alltypes
GROUP BY x / 2;

SELECT int_col / 2 AS x
FROM functional.alltypes
ORDER BY -x;

SELECT int_col / 2 AS x
FROM functional.alltypes
GROUP BY x
HAVING x > 3;
```

Default PARQUET_ARRAY_RESOLUTION Changed

The `PARQUET_ARRAY_RESOLUTION` query option controls the path-resolution behavior for Parquet files with nested arrays. The default value for the `PARQUET_ARRAY_RESOLUTION` was changed to `THREE_LEVEL` in CDH 6.0. Review your queries to see if the default value change result in different result sets.

See [PARQUET_ARRAY_RESOLUTION Query Option](#) for the information about the option.

Non-standard Timezone Names Unsupported

As the initial step for IANA timezone integration in the coming release, Impala will drop the support for non-standard timezone aliases in CDH 6.0.

Impala supports a majority of the [IANA time zones](#) with the following exceptions of time zones not supported: America/Fort_Nelson, America/Punta_Arenas, Asia/Atyrau, Asia/Barnaul, Asia/Famagusta, Asia/Tomsk, Asia/Yangon, Europe/Astrakhan, Europe/Kirov, Europe/Saratov, Europe/Ulyanovsk, GMT+0, GMT-0, ROC

See [Unsupported Time Zone](#) for the list of time zone aliases no longer supported and the canonical names you can use to replace the unsupported aliases with.

Return Type Changed for EXTRACT and DATE_PART Functions in CDH 6.0 / Impala 3.0

The following changes were made to the `EXTRACT` and `DATE_PART` functions:

- The output type of the `EXTRACT` and `DATE_PART` functions was changed to `BIGINT`.
- Extracting the millisecond part from a `TIMESTAMP` returns the seconds component and the milliseconds component. For example, `EXTRACT (CAST('2006-05-12 18:27:28.123456789' AS TIMESTAMP), 'MILLISECOND')` will return 28123.

Apache Kafka

Kafka is now bundled as part of CDH. The following sections describe incompatible changes between the previous, separately installed Kafka (CDK powered by Apache Kafka version 3.1) and the CDH 6.0.0 Kafka version. These changes affect clients built with CDH 6.0.0 libraries. Cloudera recommends upgrading clients to the new release; however clients built with previous versions of Kafka will continue to function.

Packaging

CDH and previous distributions of Kafka (CDK Powered by Apache Kafka) cannot coexist in the same cluster.

Deprecated Scala-based Client API and New Java Client API

Scala-based clients are deprecated in this release and will be removed in an upcoming release.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are deprecated and unsupported as of CDH 6.0.0:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Command Line Options Removed

Some command line tools are affected by the deprecation of old clients (see [previous entry](#)). The following options have been removed and are not recognized as valid options:

- `--new-consumer`
- `--old-consumer`
- `--old-producer`

The tools affected use the new clients.

Command Line Tools Removed

The following command line tools and runnable classes are removed:

- `kafka-replay-log-producer`
- `kafka-simple-consumer-shell`
- `kafka.tools.ReplayLogProducer`
- `kafka.tools.SimpleConsumerShell`
- `kafka.tools.ExportZkOffset`
- `kafka.tools.ImportZkOffset`
- `kafka.tools.SimpleConsumerPerformance`
- `kafka.tools.UpdateOffsetsInZK`
- `kafka.tools.VerifyConsumerRebalance`
- `kafka.tools.ProducerPerformance`

Consumer API Changes

Consumer methods invoked with unassigned partitions now raise an `IllegalStateException` instead of an `IllegalArgumentException`.

Previous versions of the Consumer method `poll(long)` would wait for metadata updates regardless of timeout parameter. This behavior is expected to change in future releases; make sure your client applications include an appropriate timeout parameter and do not rely on the previous behavior.

Exception Classes Removed

The following exceptions were deprecated in a previous release and are not thrown anymore are removed:

- `GroupCoordinatorNotAvailableException`
- `GroupLoadInProgressException`
- `NotCoordinatorForGroupException`
- `kafka.common.KafkaStorageException`

Metrics Updated

Kafka consumers' per-partition metrics were changed to use tags for topic and partition rather than the metric name. For more information see [KIP-225](#).

Apache Kudu

There are no incompatible changes in this release.

Apache Oozie

There are no incompatible changes in this release.

Apache Parquet

Packages and Group ID Renamed

As a part of the Apache incubation process, all Parquet packages and the project's group ID were renamed as follows:

Parquet Version	1.6.0 and lower (CDH 5.x)	1.7.0 and higher (CDH 6.x)
Java Package Names	parquet.*	org.apache.parquet.*
Group ID	com.twitter	org.apache.parquet

If you directly consume the Parquet API, instead of using Parquet through Hive, Impala or other CDH component, you need to update your code to reflect these changes:

Update *.java files:

Before	After
import parquet.*;	import org.apache.parquet.*;

Update pom.xml:

Before	After
<pre><dependency> <groupId> com.twitter </groupId> <version> \${parquet.version} </version> </dependency></pre>	<pre><dependency> <groupId> org.apache.parquet </groupId> <version> \${parquet.version} </version> </dependency></pre>

API Methods Removed

In Parquet 1.6, a number of API methods were removed from the `parquet.hadoop.ParquetInputSplit` class that depended on reading metadata on the client side. Metadata should be read on the task side instead.

Removed Method	New Method to Use
<code>parquet.hadoop.ParquetInputSplit.getFileSchema</code>	<code>org.apache.parquet.hadoop.api.InitContext.getFileSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getRequestedSchema</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getRequestedSchema</code>
<code>parquet.hadoop.ParquetInputSplit.getReadSupportMetadata</code>	<code>org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getReadSupportMetadata</code>
<code>parquet.hadoop.ParquetInputSplit.getBlocks</code>	<code>org.apache.parquet.hadoop.metadata.ParquetMetadata.getBlocks</code>
<code>parquet.hadoop.ParquetInputSplit.getExtraMetadata</code>	-

Apache Pig

The following change is introduced to Pig in CDH 6.0 and is not a backwards compatible change. You must modify your Pig scripts as described below.

Removal of the Apache DataFu Pig JAR from CDH 6

Apache DataFu Pig is a collection of user-defined functions that can be used with Pig for data mining and statistical analysis on large-scale data. The DataFu JAR was included in CDH 5, but due to very low adoption rates, the JAR was deprecated in CDH 5.9 and is being removed from CDH 6, starting with CDH 6.0. It is no longer supported.

Recommended Migration Strategy

A simple way to assess what DataFu functions you are using in your Pig scripts is to use the `grep` utility to search for occurrences of "datafu" in your code. When DataFu functions are used in Pig scripts, you must use a function definition entry that contains "datafu" like the following example:

```
define <function_name> datafu.pig... .<class_name>() ;
```

Use `grep` to search for the string "datafu" in your scripts and that will identify where the DataFu JAR is used.

Cloudera recommends migrating to Hive UDFs or operators wherever it is possible. However, if there are cases where it is impossible to replace DataFu functions with Hive functions, download the upstream version of the DataFu Pig libraries and place them on the node where the Pig front end is used. To preserve compatibility, use the version 1.1.0 JAR, which was the version included in CDH 5. You can download the JAR file [here](#). However, Cloudera does not support using this upstream DataFu JAR file.

Mapping DataFu UDFs to Hive UDFs

The following Hive UDFs map to DataFu UDFs and can be used instead in Pig scripts with the caveats that are listed:

Table 39: Hive Functions That Map to DataFu Functions

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
MD5 (hash)	Computes the MD5 value of a string and outputs a hex value by default.	md5	None
SHA (hash)	Computes the SHA value of a string and outputs a hex value by default.	sha/sha1	None
RandInt (random)	Generates a uniformly distributed integer between two bounds.	rand	None
VAR (stats)	Generates the variance of a set of values.	variance	None
Median (stats)	Computes the median for a sorted input bag. A special case of the Quantile function.	percentile(0.5)	See Limitations When Substituting Quantile and Median DataFu Functions on page 459.
Quantile (stats)	Computes quantiles for a sorted input bag.	percentile	See Limitations When Substituting Quantile and Median DataFu Functions on page 459.
Coalesce (util)	Returns the first non-null value from a tuple, like COALESCE in SQL.	coalesce	None
InUDF (util)	Similar to the SQL IN function, this function provides a convenient way	IN	None

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
	to filter using a logical disjunction over many values.		

For more information about using Hive UDFs, see

https://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_hive_udf.html.

Limitations When Substituting Quantile and Median DataFu Functions

With the exception of `Median` and `Quantile` all Hive functions specified in the above table should work as expected in Pig scripts. `Median` extends `Quantile` in DataFu functions and the equivalent Hive functions have a similar relationship. However, there is an important difference in how you use `percentile` and how you use `Quantile`. The differences are summarized in the following table:

Table 40: Differences Between Usage of DataFu 'Quantile' and Hive 'percentile'

Function:	Quantile	percentile
Input must be sorted:	Yes	No
Nulls are allowed in input:	No	Yes
Examples:	<pre>register datafu-1.2.0.jar; define median datafu.pig.stats.Quantile('0.5'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B = group A by name; C = foreach B { sorted = order A by n; generate group, flatten (median(sorted.n)); }</pre>	<pre>define percentile HiveUDAF ('percentile'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B2 = foreach A generate name, n, 0.5 as perc; C2 = GROUP B2 by name; D2 = FOREACH C2 generate group, percentile (B2. (n, perc));</pre>

Although DataFu `StreamingQuantile` and `StreamingMedian` might appear to match Hive's `percentile_approx` function, Pig cannot consume `percentile_approx`.

DataFu Functions with No Hive Function or Operator Equivalent

The following general limitations apply when mapping DataFu UDFs to Hive UDFs:

- Many DataFu functions operate on a custom Pig data structure called a *bag*. No Hive UDFs can operate on Pig bags, so there are no equivalents for these DataFu functions.
- Some DataFu functions are custom functions that do not have Hive UDF equivalents. For example, the DataFu functions that calculate geographic distances, run the PageRank algorithm, or that do sampling. There are no equivalent Hive UDFs for these DataFu functions either.

Table 41: DataFu Functions with No Hive UDF Equivalent

AppendToBag (bags)	AssertUDF (util)	BagConcat (bags)
BagGroup (bags)	BagLeftOuterJoin (bags)	BagSplit (bags)
BoolToInt (util)	CountEach (bags)	DistinctBy (bags)

<code>EmptyBagToNull (bags)</code>	<code>EmptyBagToNullFields (bags)</code>	<code>Enumerate (bags)</code>
<code>FirstTupleFromBag (bags)</code>	<code>HaversineDistInMiles (geo)</code>	<code>IntToBool (util)</code>
<code>MarkovPairs</code>	<code>NullToEmptyBag (bags)</code>	<code>PageRank (linkanalysis)</code>
<code>PrependToBag (bags)</code>	<code>ReservoirSample (sampling)*</code>	<code>ReverseEnumerate (bags)</code>
<code>SampleByKey (sampling)*</code>	<code>SessionCount (sessions)</code>	<code>Sessionize (sessions)</code>
<code>SetIntersect (sets)</code>	<code>SetUnion (sets)</code>	<code>SimpleRandomSample (sampling)*</code>
<code>TransposeTupleToBag (util)</code>	<code>UnorderedPairs (bags)</code>	<code>UserAgentClassify (urls)</code>
<code>WeightedSample (sampling)*</code>	<code>WilsonBinConf (stats)</code>	—

* These DataFu functions might be replaced with TABLESAMPLE in HiveQL. See the [Apache Hive wiki](#).

Cloudera Search

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has many incompatibilities with the 4.10 version of Apache Solr used in recent CDH 5 releases, such as the following:

- Solr 7 uses a managed schema by default. Generating an instance directory no longer generates `schema.xml`. For instructions on switching to a managed schema, see [Switching from schema.xml to Managed Schema](#) in *Apache Solr Reference Guide*.
- Creating a collection using `solrctl collection --create` without specifying the `-c <configName>` parameter now uses a default configuration set (`named _default`) instead of a configuration set with the same name as the collection. To avoid this, always specify the `-c <configName>` parameter when creating new collections.

For the full list of changes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

Apache Sentry contains the following incompatible change in CDH 6.0.0:

- Sentry no longer supports policy file authorization. You must migrate policy files to the database-backed Sentry service before you upgrade to CDH 6.0.0 unless you are using Sentry policy files for Solr. If you are using Sentry policy files for Solr, you must migrate to the database-backed Sentry service after you upgrade.

For information about migrating policy files before you upgrade, see [Migrating from Sentry Policy Files to the Sentry Service](#). For information about migrating policy files for Solr after you upgrade, see [Migrating Sentry Privileges for Solr After Upgrading to CDH 6](#).

Apache Spark

The following sections describe changes in Spark support in CDH 6 that might require special handling during upgrades, or code changes within existing applications.

- All Spark applications built against Spark 1.6 in CDH 5 must be rebuilt against Spark 2.x in CDH 6.
- Spark 2 in CDH 6 works with Java 8, not Java 7. If this change produces any Java code incompatibilities, update your Java code and rebuild the application.
- Spark 2 in CDH 6 works with Scala 2.11, not Scala 2.10. If this change produces any Scala code incompatibilities, update your Scala code and rebuild the application.
- `HiveContext` and `SQLContext` have been removed, although those variables still work for backward compatibility. Use the `SparkSession` object to replace both of these handles.
- `DataFrames` have been removed from the Scala API. `DataFrame` is now a special case of `Dataset`.

Since compile-time type-safety in Python and R is not a language feature, the concept of Dataset does not apply to these languages' APIs. Instead, DataFrame remains the primary programming abstraction.

- Spark 2.0 and higher do not use an assembly JAR for standalone applications.
- If you have event logs created in CDH 5.3 or lower, you cannot read those logs using Spark in CDH 6.0 or higher.

Apache Sqoop

The following changes are introduced in CDH 6.0, and are not backwards compatible:

- All classes in `com.cloudera.sqoop` packages have been removed in CDH 6.0. Use the corresponding classes from `org.apache.sqoop` packages. For example, use `org.apache.sqoop.SqoopOptions` instead of `com.cloudera.sqoop.SqoopOptions`.



Note: This change only affects customers who build their own application on top of Sqoop classes. Sqoop CLI users are not affected.

- Because of changes introduced in the Sqoop metastore logic, the metastore database created by Sqoop CDH 6 cannot be used by earlier versions. The metastore database created by Sqoop CDH 5 can be used by both Sqoop CDH 5 and Sqoop CDH 6.

Require an explicit option to be specified with `--split-by` for a String column

Using the `--split-by` option with a CHAR or VARCHAR column does not always work properly, so Sqoop now requires the user to set the `org.apache.sqoop.splitter.allow_text_splitter` property to `true` to confirm that they are aware of this risk.

Example:

```
sqoop import --Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect $MYCONN
--username $MYUSER --password $MYPWD --table "test_table" --split-by "string_column"
```

For more information, see [SQOOP-2910](#).

Make Sqoop fail if user specifies `--direct` connector when it is not available

The `--direct` option is only supported with the following databases: MySQL, PostgreSQL, Oracle, Netezza.

In earlier releases, Sqoop silently ignored this option if it was specified for other databases, but it now throws an error.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table
"direct_import" --direct
```

The command fails with the following error message:

```
Was called with the --direct option, but no direct connector available.
```

For more information, see [SQOOP-2913](#).

Sqoop does not allow `--as-parquetfile` with hcatalog jobs or when hive import with `create-hive-table` is used

The `--create-hive-table` option is not supported when the user imports into Hive in Parquet format. Earlier this option was silently ignored, and the data was imported even if the Hive table existed. Sqoop will now fail if the `--create-hive-table` option is used with the `--as-parquetfile` option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table
"test_table" --hive-import --as-parquetfile --create-hive-table
```

The command fails with the following error message:

```
Hive import and create hive table is not compatible with importing into ParquetFile  
format.
```

For more information, see [SQOOP-3010](#).

Create fail fast for export with --hcatalog-table <HIVE_VIEW>

Importing into and exporting from a Hive view using HCatalog is not supported by Sqoop. A fail fast check was introduced so that now Sqoop throws a descriptive error message if the user specified a Hive view in the value of the --hcatalog-table option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table"  
--hcatalog-table "test_view"
```

The command fails with the following error message:

```
Reads/Writes from and to Views are not supported by HCatalog
```

For more information, see [SQOOP-3027](#).

Simplify Unicode character support in source files

Simplify Unicode character support in source files (introduced by [SQOOP-3074](#)) by defining explicit locales instead of using EscapeUtils. The Java source files generated by Sqoop will be encoded in UTF-8 format.

For more information, see [SQOOP-3075](#).

Columns added to MySQL after initial Sqoop import, export back to table with same schema fails

If we export from HDFS to an RDBMS table and the file on HDFS has no value for some of the columns defined in the table, Sqoop will use the values of --input-null-string and --input-null-non-string options. Earlier this scenario was not supported and Sqoop failed.

For more information, see [SQOOP-3158](#).

Sqoop fails if the user tries to encode a null value when using --direct connector and a MySQL database

The MySQL direct connector does not support the --null-string, --null-non-string, --input-null-string, and --input-null-non-string options. These options were silently ignored earlier, but Sqoop now throws an error if these options are used with MySQL direct imports and exports.

For more information, see [SQOOP-3206](#).

Apache Zookeeper

There are no incompatible changes in this release.

Timezone Names Unsupported in Impala in CDH 6.0.1

The following table lists the time zone names / aliases no longer supported in Impala along with the canonical names you can use to replace the unsupported aliases with.

Deprecated	Replace with
ACDT	Australia/South
ACST	Australia/Darwin
ACWST	Australia/Eucla
ADT	America/Thule
AEDT	Australia/Sydney

AEST	Australia/Sydney
AFT	Asia/Kabul
AKDT	America/Anchorage
AKST	America/Anchorage
ALMT	Asia/Almaty
AMST	America/Cuiaba
AMT	Asia/Yerevan
ANAT	Asia/Anadyr
AQTT	Asia/Aqtai
AWST	Australia/West
AZOST	Atlantic/Azores
AZOT	Atlantic/Azores
AZST	Asia/Baku
AZT	Asia/Baku
Acre Time	Brazil/Acre
Afghanistan Time	Asia/Kabul
Alaska Daylight Time	America/Anchorage
Alaska Standard Time	America/Anchorage
Alma-Ata Time	Asia/Almaty
Amazon Summer Time	America/Cuiaba
Amazon Time	Brazil/West
Anadyr Time	Asia/Anadyr
Aqtai Time	Asia/Aqtai
Aqtobe Time	Asia/Aqtobe
Arabia Standard Time	Asia/Aden
Argentine Time	America/Argentina/Buenos_Aires
Armenia Time	Asia/Yerevan
Asia/Riyadh87	-
Asia/Riyadh88	-
Asia/Riyadh89	-
Atlantic Daylight Time	America/Thule
Atlantic Standard Time	America/Puerto_Rico
Australian Central Daylight Time (South Australia)	Australia/South
Australian Central Daylight Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Standard Time (Northern Territory)	Australia/Darwin
Australian Central Standard Time (South Australia)	Australia/South

Australian Central Standard Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Western Standard Time	Australia/Eucla
Australian Eastern Daylight Time (New South Wales)	Australia/Sydney
Australian Eastern Daylight Time (Tasmania)	Australia/Hobart
Australian Eastern Daylight Time (Victoria)	Australia/Victoria
Australian Eastern Standard Time (New South Wales)	Australia/Sydney
Australian Eastern Standard Time (Queensland)	Australia/Brisbane
Australian Eastern Standard Time (Tasmania)	Australia/Hobart
Australian Eastern Standard Time (Victoria)	Australia/Victoria
Australian Western Standard Time	Australia/West
Azerbaijan Summer Time	Asia/Baku
Azerbaijan Time	Asia/Baku
Azores Summer Time	Atlantic/Azores
Azores Time	Atlantic/Azores
BDT	Asia/Dhaka
BNT	Asia/Brunei
BOT	America/La_Paz
BRST	America/Sao_Paulo
BRT	America/Sao_Paulo
BTT	Asia/Thimbu
Bangladesh Time	Asia/Dhaka
Bhutan Time	Asia/Thimbu
Bolivia Time	America/La_Paz
Bougainville Standard Time	Pacific/Bougainville
Brasilia Summer Time	America/Sao_Paulo
Brasilia Time	America/Sao_Paulo
British Summer Time	GB
Brunei Time	Asia/Brunei
CCT	Indian/Cocos
CDT	America/Chicago
CEST	CET
CHADT	NZ-CHAT
CHAST	NZ-CHAT
CHOT	Asia/Choibalsan
CHUT	Pacific/Yap
CKT	Pacific/Rarotonga

CLST	America/Santiago
CLT	America/Santiago
COT	America/Bogota
CVT	Atlantic/Cape_Verde
CXT	Indian/Christmas
Canada/East-Saskatchewan	America/Regina
Cape Verde Time	Atlantic/Cape_Verde
Central African Time	Africa/Harare
Central Daylight Time	America/Chicago
Central European Summer Time	CET
Central European Time	CET
Central Indonesia Time	Asia/Makassar
Central Standard Time	America/Chicago
ChST	Pacific/Guam
Chamorro Standard Time	Pacific/Guam
Chatham Daylight Time	NZ-CHAT
Chatham Standard Time	NZ-CHAT
Chile Summer Time	America/Santiago
Chile Time	America/Santiago
China Standard Time	Asia/Shanghai
Choibalsan Time	Asia/Choibalsan
Christmas Island Time	Indian/Christmas
Chuuk Time	Pacific/Yap
Cocos Islands Time	Indian/Cocos
Colombia Time	America/Bogota
Cook Is. Time	Pacific/Rarotonga
Coordinated Universal Time	UCT
Cuba Daylight Time	Cuba
Cuba Standard Time	Cuba
DAVT	Antarctica/Davis
DDUT	Antarctica/DumontDUrville
Davis Time	Antarctica/Davis
Dumont-d'Urville Time	Antarctica/DumontDUrville
EASST	Pacific/Easter
EAST	Pacific/Easter
EDT	America/Indiana/Indianapolis

EEST	Africa/Cairo
EGST	America/Scoresbysund
EGT	America/Scoresbysund
East Indonesia Time	Asia/Jayapura
Easter Is. Summer Time	Pacific/Easter
Easter Is. Time	Pacific/Easter
Eastern African Time	Africa/Addis_Ababa
Eastern Daylight Time	America/Indiana/Indianapolis
Eastern European Summer Time	Africa/Cairo
Eastern European Time	Africa/Cairo
Eastern Greenland Summer Time	America/Scoresbysund
Eastern Greenland Time	America/Scoresbysund
Eastern Standard Time	EST
Ecuador Time	America/Guayaquil
FJST	Pacific/Fiji
FJT	Pacific/Fiji
FKT	Atlantic/Stanley
FNT	America/Noronha
Falkland Is. Time	Atlantic/Stanley
Fernando de Noronha Time	America/Noronha
Fiji Summer Time	Pacific/Fiji
Fiji Time	Pacific/Fiji
French Guiana Time	America/Cayenne
French Southern & Antarctic Lands Time	Indian/Kerguelen
GALT	Pacific/Galapagos
GAMT	Pacific/Gambier
GET	Asia/Tbilisi
GFT	America/Cayenne
GILT	Pacific/Tarawa
GMT+00:00	GMT0
GMT+01:00	Etc/GMT-1
GMT+02:00	Etc/GMT-2
GMT+03:00	Etc/GMT-3
GMT+04:00	Etc/GMT-4
GMT+05:00	Etc/GMT-5
GMT+06:00	Etc/GMT-6

GMT+07:00	Etc/GMT-7
GMT+08:00	Etc/GMT-8
GMT+09:00	Etc/GMT-9
GMT+10:00	Etc/GMT-10
GMT+11:00	Etc/GMT-11
GMT+12:00	Etc/GMT-12
GMT+13:00	Etc/GMT-13
GMT+14:00	Etc/GMT-14
GMT-01:00	Etc/GMT+1
GMT-02:00	Etc/GMT+2
GMT-03:00	Etc/GMT+3
GMT-04:00	Etc/GMT+4
GMT-05:00	Etc/GMT+5
GMT-06:00	Etc/GMT+6
GMT-07:00	Etc/GMT+7
GMT-08:00	Etc/GMT+8
GMT-09:00	Etc/GMT+9
GMT-10:00	Etc/GMT+10
GMT-11:00	Etc/GMT+11
GMT-12:00	Etc/GMT+12
GST	Asia/Dubai
GYT	America/Guyana
Galapagos Time	Pacific/Galapagos
Gambier Time	Pacific/Gambier
Georgia Time	Asia/Tbilisi
Ghana Mean Time	Africa/Accra
Gilbert Is. Time	Pacific/Tarawa
Greenwich Mean Tim	GB
Gulf Standard Time	Asia/Dubai
Guyana Time	America/Guyana
HADT	US/Aleutian
HAST	US/Aleutian
HKT	Hongkong
HOVT	Asia/Hovd
Hawaii Standard Time	HST
Hawaii-Aleutian Daylight Time	US/Aleutian

Hawaii-Aleutian Standard Time	US/Aleutian
Hong Kong Time	Hongkong
Hovd Time	Asia/Hovd
ICT	Asia/Ho_Chi_Minh
IDT	Israel
IOT	Indian/Chagos
IRDT	Iran
IRKT	Asia/Chita
IRST	Iran
India Standard Time	Asia/Kolkata
Indian Ocean Territory Time	Indian/Chagos
Indochina Time	Asia/Ho_Chi_Minh
Iran Daylight Time	Iran
Iran Standard Time	Iran
Irish Summer Time	Eire
Irkutsk Time	Asia/Chita
Israel Daylight Time	Israel
Israel Standard Time	Israel
Japan Standard Time	Asia/Tokyo
KGT	Asia/Bishkek
KOST	Pacific/Kosrae
KRAT	Asia/Krasnoyarsk
KST	ROK
Khandyga Time	Asia/Khandyga
Kirgizstan Time	Asia/Bishkek
Korea Standard Time	ROK
Kosrae Time	Pacific/Kosrae
Krasnoyarsk Time	Asia/Krasnoyarsk
LHDT	Australia/LHI
LHST	Australia/LHI
LINT	Pacific/Kiritimati
Line Is. Time	Pacific/Kiritimati
Lord Howe Daylight Time	Australia/LHI
Lord Howe Standard Time	Australia/LHI
MAGT	Asia/Magadan
MART	Pacific/Marquesas

MAWT	Antarctica/Mawson
MDT	Navajo
MEST	MET
MHT	Kwajalein
MIST	Antarctica/Macquarie
MMT	Asia/Rangoon
MSK	W-SU
MUT	Indian/Mauritius
MVT	Indian/Maldives
MYT	Asia/Kuching
Macquarie Island Standard Time	Antarctica/Macquarie
Magadan Time	Asia/Magadan
Malaysia Time	Asia/Kuching
Maldives Time	Indian/Maldives
Marquesas Time	Pacific/Marquesas
Marshall Islands Time	Kwajalein
Mauritius Time	Indian/Mauritius
Mawson Time	Antarctica/Mawson
Middle Europe Summer Time	MET
Middle Europe Time	MET
Mideast/Riyadh87	-
Mideast/Riyadh88	-
Mideast/Riyadh89	-
Moscow Standard Time	W-SU
Mountain Daylight Time	Navajo
Mountain Standard Time	MST
Myanmar Time	Asia/Rangoon
NCT	Pacific/Noumea
NDT	America/St_Johns
NFT	Pacific/Norfolk
NOVT	Asia/Novosibirsk
NPT	Asia/Katmandu
NRT	Pacific/Nauru
NUT	Pacific/Niue
NZDT	NZ
NZST	NZ

Nauru Time	Pacific/Nauru
Nepal Time	Asia/Katmandu
New Caledonia Time	Pacific/Noumea
New Zealand Daylight Time	NZ
New Zealand Standard Time	NZ
Newfoundland Daylight Time	America/St_Johns
Newfoundland Standard Time	America/St_Johns
Niue Time	Pacific/Niue
Norfolk Time	Pacific/Norfolk
Novosibirsk Time	Asia/Novosibirsk
OMST	Asia/Omsk
ORAT	Asia/Oral
Omsk Time	Asia/Omsk
Oral Time	Asia/Oral
PDT	America/Los_Angeles
PET	America/Lima
PETT	Asia/Kamchatka
PGT	Pacific/Port_Moresby
PHOT	Pacific/Enderbury
PHT	Asia/Manila
PKT	Asia/Karachi
PMDT	America/Miquelon
PMST	America/Miquelon
PONT	Pacific/Ponape
PWT	Pacific/Palau
PYST	America/Asuncion
PYT	America/Asuncion
Pacific Daylight Time	America/Los_Angeles
Pacific Standard Time	America/Los_Angeles
Pakistan Time	Asia/Karachi
Palau Time	Pacific/Palau
Papua New Guinea Time	Pacific/Port_Moresby
Paraguay Summer Time	America/Asuncion
Paraguay Time	America/Asuncion
Peru Time	America/Lima
Petropavlovsk-Kamchatski Time	Asia/Kamchatka

Philippines Time	Asia/Manila
Phoenix Is. Time	Pacific/Enderbury
Pierre & Miquelon Daylight Time	America/Miquelon
Pierre & Miquelon Standard Time	America/Miquelon
Pitcairn Standard Time	Pacific/Pitcairn
Pohnpei Time	Pacific/Ponape
QYJT	Asia/Qyzylorda
Qyzylorda Time	Asia/Qyzylorda
RET	Indian/Reunion
ROTT	Antarctica/Rothera
Reunion Time	Indian/Reunion
Rothera Time	Antarctica/Rothera
SAKT	Asia/Sakhalin
SAMT	Europe/Samara
SAST	Africa/Maseru
SBT	Pacific/Guadalcanal
SCT	Indian/Mahe
SGT	Singapore
SRET	Asia/Srednekolymsk
SRT	America/Paramaribo
SYOT	Antarctica/Syowa
Sakhalin Time	Asia/Sakhalin
Samara Time	Europe/Samara
Samoa Standard Time	US/Samoa
Seychelles Time	Indian/Mahe
Singapore Time	Singapore
Solomon Is. Time	Pacific/Guadalcanal
South Africa Standard Time	Africa/Maseru
South Georgia Standard Time	Atlantic/South_Georgia
Srednekolymsk Time	Asia/Srednekolymsk
Suriname Time	America/Paramaribo
Syowa Time	Antarctica/Syowa
SystemV/AST4	America/Puerto_Rico
SystemV/AST4ADT	Canada/Atlantic
SystemV/CST6	Canada/Saskatchewan
SystemV/CST6CDT	US/Central

SystemV/EST5	America/Cayman
SystemV/EST5EDT	America/New_York
SystemV/HST10	US/Hawaii
SystemV/MST7	US/Arizona
SystemV/MST7MDT	America/Denver
SystemV/PST8	Pacific/Pitcairn
SystemV/PST8PDT	US/Pacific
SystemV/YST9	Pacific/Gambier
SystemV/YST9YDT	US/Alaska
TAHT	Pacific/Tahiti
TFT	Indian/Kerguelen
TJT	Asia/Dushanbe
TKT	Pacific/Fakaofo
TLT	Asia/Dili
TMT	Asia/Ashgabat
TOT	Pacific/Tongatapu
TVT	Pacific/Funafuti
Tahiti Time	Pacific/Tahiti
Tajikistan Time	Asia/Dushanbe
Timor-Leste Time	Asia/Dili
Tokelau Time	Pacific/Fakaofo
Tonga Time	Pacific/Tongatapu
Turkmenistan Time	Asia/Ashgabat
Tuvalu Time	Pacific/Funafuti
ULAT	Asia/Ulan_Bator
UYST	America/Montevideo
UYT	America/Montevideo
UZT	Asia/Tashkent
Ulaanbaatar Time	Asia/Ulan_Bator
Uruguay Summer Time	America/Montevideo
Uruguay Time	America/Montevideo
Ust-Nera Time	Asia/Ust-Nera
Uzbekistan Time	Asia/Tashkent
VET	America/Caracas
VLAT	Asia/Ust-Nera
VOST	Antarctica/Vostok

VUT	Pacific/Efate
Vanuatu Time	Pacific/Efate
Venezuela Time	America/Caracas
Vladivostok Time	Asia/Vladivostok
Vostok Time	Antarctica/Vostok
WAKT	Pacific/Wake
WAST	Africa/Windhoek
WAT	Africa/Lagos
WEST	WET
WFT	Pacific/Wallis
WGST	America/Godthab
WGT	America/Godthab
WIB	Asia/Jakarta
WIT	Asia/Jayapura
WITA	Asia/Makassar
WSDT	Pacific/Apia
WSST	Pacific/Apia
Wake Time	Pacific/Wake
Wallis & Futuna Time	Pacific/Wallis
West Indonesia Time	Asia/Jakarta
West Samoa Daylight Time	Pacific/Apia
West Samoa Standard Time	Pacific/Apia
Western African Summer Time	Africa/Windhoek
Western African Time	Africa/Lagos
Western European Summer Time	WET
Western European Time	WET
Western Greenland Summer Time	America/Godthab
Western Greenland Time	America/Godthab
XJT	Asia/Urumqi
Xinjiang Standard Time	Asia/Urumqi
YAKT	Asia/Yakutsk
YEKT	Asia/Yekaterinburg
Yakutsk Time	Asia/Yakutsk
Yekaterinburg Time	Asia/Yekaterinburg

Known Issues and Limitations in CDH 6.0.1

The following sections describe the known issues in CDH 6.0.1, grouped by component:

Operating System Known Issues

Known issues and workarounds related to operating systems are listed below.

Linux kernel security patch and CDH services crashes CVE-2017-10000364

After applying a recent Linux kernel security patch for [CVE-2017-1000364](#), CDH services that use the JSVC set of libraries crash with a Java Virtual Machine (JVM) error such as:

```
A fatal error has been detected by the Java Runtime Environment:  
SIGBUS (0x7) at pc=0x00007fe91ef6cebc, pid=30321, tid=0x00007fe930c67700
```

Cloudera services for HDFS and Impala cannot start after applying the patch.



Important: If you have not upgraded your Linux kernel using the distribution's patch for CVE-2017-1000364, do **not** apply the patch.

Commonly used Linux distributions are shown in the table below. However, the issue affects any CDH release that runs on RHEL, CentOS, Oracle Linux, SUSE Linux, or Ubuntu and that has had the Linux kernel security patch for CVE-2017-1000364 applied.

If you have already applied the patch for your OS according to the advisories for CVE-2017-1000364, apply the kernel update that contains the fix for your operating system (some of which are listed in the table). If you cannot apply the kernel update, you can workaround the issue by increasing the Java thread stack size as detailed in the steps below.

Distribution	Advisories for CVE-2017-1000364	Advisory updates
Oracle Linux 6	ELSA-2017-1486	Oracle has fixed this problem in ELSA-2017-1723 .
Oracle Linux 7	ELSA-2017-1484	Oracle has also added the fix for Oracle Linux 7 in ELBA-2017-1674 .
RHEL 6	RHSA-2017-1486	RedHat has fixed this problem for RHEL 6, marked this as outdated and superseded by RHSA-2017-1723 .
RHEL 7	RHSA-2017-1484	RedHat has fixed this problem for RHEL 7 and has marked this patch as outdated and superseded by RHBA-2017-1674 .
SLES	CVE-2017-1000364	SUSE has also fixed this problem and the patch names are included in this same advisory.

Workaround

If you cannot apply the kernel update, you can set the Java thread stack size to `-Xss1280k` for the affected services using the appropriate Java configuration option or the environment advanced configuration snippet, as detailed below.

For role instances that have specific Java configuration options properties:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Type `java` in the search field to display Java related configuration parameters. The **Java Configuration Options for Catalog Server** property field displays. Type `-Xss1280k` in the entry field, adding to any existing settings.
4. Click **Save Changes**.
5. Navigate to the HDFS service by selecting **Clusters > HDFS**.
6. Click the **Configuration** tab.
7. Click the Scope filter **DataNode**. The **Java Configuration Options for DataNode** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
8. Click **Save Changes**.
9. Select the Scope filter **NFS Gateway**. The **Java Configuration Options for NFS Gateway** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.

10 Click **Save Changes**.

11 Restart the affected roles (or configure the safety valves in next section and restart when finished with all configurations).

For role instances that do not have specific Java configuration options:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Click the Scope filter **Impala Daemon** and Category filter **Advanced**.
4. Type `impala daemon` environment in the search field to find the safety valve entry field.
5. In the **Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

6. Click **Save Changes**.

7. Click the Scope filter **Impala StateStore** and Category filter **Advanced**.

8. In the **Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

9. Click **Save Changes**.

10 Restart the affected roles.

The table below summarizes the parameters that can be set for the affected services:

Service	Settable Java Configuration Option
HDFS DataNode	Java Configuration Options for DataNode
HDFS NFS Gateway	Java Configuration Options for NFS Gateway
Impala Catalog Server	Java Configuration Options for Catalog Server
Impala Daemon	Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)
	<code>JAVA_TOOL_OPTIONS=-Xss1280K</code>
Impala StateStore	Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)
	<code>JAVA_TOOL_OPTIONS=-Xss1280K</code>

Cloudera Issue: CDH-55771

Leap-Second Events



Note: The [next leap-second event](#) is unknown at this time. The last leap-second event occurred on December 16, 2016 at 23:59:60 UTC.

Impact: After a leap-second event, Java applications (including CDH services) using older Java and Linux kernel versions, may consume almost 100% CPU. See <https://access.redhat.com/articles/15145>.

Leap-second events are tied to the time synchronization methods of the Linux kernel, the Linux distribution and version, and the Java version used by applications running on affected kernels.

Although Java is increasingly agnostic to system clock progression (and less susceptible to a kernel's mishandling of a leap-second event), using JDK 7 or 8 should prevent issues at the CDH level (for CDH components that use the Java Virtual Machine).

Immediate action required:

- (1) Ensure that the kernel is up to date.

Cloudera Enterprise 6 Release Guide

- **RHEL6/7, CentOS 6/7** - 2.6.32-298 or higher
- **Oracle Enterprise Linux (OEL)** - Kernels built in 2013 or later
- **SLES12** - No action required.

(2) Ensure that your Java JDKs are current (especially if the kernel is not up to date and cannot be upgraded).

- **Java 8** - No action required.

(3) Ensure that your systems use either NTP or PTP synchronization.

For systems not using time synchronization, update both the OS tzdata and Java tzdata packages to the tzdata-2016g version, at a minimum. For OS tzdata package updates, contact OS support or check updated OS repositories. For Java tzdata package updates, see Oracle's [Timezone Updater Tool](#).

Cloudera Issue: CDH-44788, TSB-189

Apache Accumulo Known Issues

Running Apache Accumulo on top of a CDH 6.0.x cluster is not currently supported. If you try to upgrade to CDH 6.0.x you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Affected Versions: CDH 6.0.x

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Cloudera Issue: DSE-2769

Apache Crunch Known Issues



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

Apache Flume Known Issues

Fast Replay does not work with encrypted File Channel

If an encrypted file channel is set to use fast replay, the replay will fail and the channel will fail to start.

Workaround: Disable fast replay for the encrypted channel by setting `use-fast-replay` to false.

Apache Issue: [FLUME-1885](#)

Apache Hadoop Known Issues

This page includes known issues and related topics, including:

Deprecated Properties

Several Hadoop and HDFS properties have been deprecated as of Hadoop 3.0 and later. For details, see [Deprecated Properties](#).

Hadoop Common

Hadoop LdapGroupsMapping does not support LDAPS for self-signed LDAP server

Hadoop LdapGroupsMapping does not work with LDAP over SSL (LDAPS) if the LDAP server certificate is self-signed. This use case is currently not supported even if Hadoop User Group Mapping LDAP TLS/SSL Enabled, Hadoop User Group Mapping LDAP TLS/SSL Truststore, and Hadoop User Group Mapping LDAP TLS/SSL Truststore Password are filled properly.

Affected Versions: All CDH versions

Apache Issue: [HADOOP-12862](#)

Cloudera Issue: CDH-37926

HDFS

OIV ReverseXML processor fails

The HDFS OIV ReverseXML processor fails if the XML file contains escaped characters.

Affected Versions: CDH 6.x**Apache Issue:** [HDFS-12828](#)

Cannot move encrypted files to trash

With HDFS encryption enabled, you cannot move encrypted files or directories to the trash directory.

Workaround: To remove encrypted files/directories, use the following command with the `-skipTrash` flag specified to bypass trash.

```
rm -r -skipTrash /testdir
```

Affected Versions: All CDH versions**Apache Issue:** [HADOOP-10902](#)

HDFS NFS gateway and CDH installation (using packages) limitation

HDFS NFS gateway works as shipped ("out of the box") only on RHEL-compatible systems, but not on SLES or Ubuntu. Because of a bug in native versions of `portmap/rpcbind`, the HDFS NFS gateway does not work out of the box on SLES or Ubuntu systems when CDH has been installed from the command-line, using packages. It does work on supported versions of RHEL-compatible systems on which `rpcbind-0.2.0-10.el6` or later is installed, and it does work if you use Cloudera Manager to install CDH, or if you start the gateway as root.

Workarounds and caveats:

- On Red Hat and similar systems, make sure `rpcbind-0.2.0-10.el6` or later is installed.
- On SLES and Ubuntu systems, do one of the following:
 - Install CDH using Cloudera Manager; or
 - Start the NFS gateway as root; or
 - [Start the NFS gateway without using packages](#); or
 - You can use the gateway by running `rpcbind` in insecure mode, using the `-i` option, but keep in mind that this allows anyone from a remote host to bind to the portmap.

Upstream Issue: [731542](#) (Red Hat), [823364](#) (SLES)

No error when changing permission to 777 on .snapshot directory

Snapshots are read-only; running `chmod 777` on the `.snapshot`s directory does not change this, but does not produce an error (though other illegal operations do).

Affected Versions: All CDH versions**Apache Issue:** [HDFS-4981](#)

Snapshot operations are not supported by ViewFileSystem

Affected Versions: All CDH versions

Snapshots do not retain directories' quotas settings

Affected Versions: All CDH versions**Apache Issue:** [HDFS-4897](#)Permissions for `dfs.namenode.name.dir` incorrectly set

Hadoop daemons should set permissions for the `dfs.namenode.name.dir` (or `dfs.name.dir`) directories to `drwx-----` (700), but in fact these permissions are set to the file-system default, usually `drwxr-xr-x` (755).

Workaround: Use `chmod` to set permissions to 700.

Affected Versions: All CDH versions

Apache Issue: [HDFS-2470](#)

hadoop fsck -move does not work in a cluster with host-based Kerberos

Workaround: Use `hadoop fsck -delete`

Affected Versions: All CDH versions

Apache Issue: None

Block report can exceed maximum RPC buffer size on some DataNodes

On a DataNode with a large number of blocks, the block report may exceed the maximum RPC buffer size.

Workaround: Increase the value `ipc.maximum.data.length` in `hdfs-site.xml`:

```
<property>
  <name>ipc.maximum.data.length</name>
  <value>268435456</value>
</property>
```

Affected Versions: All CDH versions

Apache Issue: None

MapReduce2 and YARN

The Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager.

When ResourceManager high availability is enabled the Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager. This causes the following issues in Cloudera Manager:

- If **Enable Kerberos Authentication for HTTP Web-Console** is disabled: Cloudera Manager shows statistics for the wrong server.
- If **Enable Kerberos Authentication for HTTP Web-Console** is enabled: connection from the agent to the standby fails with the `HTTPError: HTTP Error 401: Authentication required` error message. As a result, the health of the Standby Resource Manager will become bad.

Workaround: N/A

Affected Versions: CDH 6.0.x, CDH 6.1.0

Fixed Version: CDH 6.1.1

Cloudera Issue: CDH-76040

YARN's Continuous Scheduling can cause slowness in Oozie

When Continuous Scheduling is enabled in Yarn, this can cause slowness in Oozie due to long delays in communicating with Yarn. In Cloudera Manager 5.9.0 and higher, Enable Fair Scheduler Continuous Scheduler is turned off by default.

Workaround: Turn off `Enable Fair Scheduler Continuous Scheduling` in Cloudera Manager YARN Configuration. To keep equivalent benefits of this feature, turn on `Fair Scheduler Assign Multiple Tasks`.

Affected Versions: All CDH versions

Cloudera Issue: CDH-60788

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

Affected Versions: All CDH versions

Apache Issue: None

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Routable IP address required by ResourceManager

ResourceManager requires routable host : port addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard 0.0.0.0 address.

Workaround: Set the address, in the form host : port, either in the client-side configuration, or on the command line when you submit the job.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-6808

Amazon S3 copy may time out

The Amazon S3 filesystem does not support renaming files, and performs a copy operation instead. If the file to be moved is very large, the operation can time out because S3 does not report progress during the operation.

Workaround: Use `-Dmapred.task.timeout=15000000` to increase the MR task timeout.

Affected Versions: All CDH versions

Apache Issue: [MAPREDUCE-972](#)

Cloudera Issue: CDH-17955

Apache HBase Known Issues

IOException from Timeouts

CDH 5.12.0 includes the fix [HBASE-16604](#), where the internal scanner that retries in case of IOException from timeouts could potentially miss data. Java clients were properly updated to account for the new behavior, but thrift clients will now see exceptions where the previous missing data would be.

Workaround: Create a new scanner and retry the operation when encountering this issue.

IntegrationTestReplication fails if replication does not finish before the verify phase begins

During IntegrationTestReplication, if the verify phase starts before the replication phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the `-t` flag to set the timeout value before starting verification.

Cloudera Issue: None.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

Cloudera Enterprise 6 Release Guide

ExportSnapshot or DistCp operations may fail on the Amazon s3a:// protocol

ExportSnapshot or DistCP operations may fail on AWS when using certain JDK 8 versions, due to an incompatibility between the AWS Java SDK 1.9.x and the joda-time date-parsing module.

Workaround: Use joda-time 2.8.1 or higher, which is included in AWS Java SDK 1.10.1 or higher.

Cloudera Issue: None.

An operating-system level tuning issue in RHEL7 causes 30% latency regressions

Workaround: Avoid using RHEL 7 if you have a latency-critical workload. For a cached workload, consider tuning the C-state (power-saving) behavior of your CPUs.

Cloudera Issue: CDH-32642

Severity: Medium

A RegionServer under extreme duress due to back-to-back garbage collection combined with heavy load on HDFS can lock up while attempting to append to the WAL

The RegionServer appears operational to ZooKeeper, and continues to host regions, but cannot complete any writes. The most obvious symptom is that all writes to regions on this RegionServer time out, and the RegionServer log shows no progress other than queuing of flushes that never complete. Log messages such as the following may occur:

```
124028 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.Sleeper: We slept  
42911ms instead of 3000ms,  
    this is likely due to a long garbage collecting pause and it's usually bad, see  
    http://hbase.#  
124029 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.JvmPauseMonitor: Detected  
    pause in JVM or  
    host machine (eg GC): pause of approximately 41110ms
```

```
1806 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
    Slow sync cost: 2734 ms, current pipeline:  
    [DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#  
1807 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
    Slow sync cost: 2963 ms, current pipeline:  
    [DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
```

Apache Issue: [HBASE-14374](#)

Workaround: Restart the RegionServer. To avoid the problem, adjust garbage-collection settings, give the RegionServer more RAM, and reduce the load on HDFS.

Export to Azure Blob Storage (the wasb:// or wasbs:// protocol) is not supported

CDH 5.3 and higher supports Azure Blob Storage for some applications. However, a null pointer exception occurs when you specify a wasb:// or wasbs:// location in the --copy-to option of the ExportSnapshot command or as the output directory (the second positional argument) of the Export command.

Workaround: None.

Apache Issue: [HADOOP-12717](#)

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a Delete Table fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.
- If hbaseAdmin.modifyTable() is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The postOperation is implemented only for postDeleteColumn().
- If Create Table fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: None

Apache Issue: [HBASE-6992](#)

Apache Hive / HCatalog / Hive on Spark Known Issues

This topic also contains:

- [HCatalog Known Issues](#)
- [Hive on Spark Known Issues](#)

Hive Jobs Are Submitted to a Single Queue When Sentry is Deployed

Hive jobs are not submitted into the correct YARN queue when Hive is using Sentry because Hive does not use the YARN API to resolve the user or group of the job's original submitter. This causes the job to be placed in a queue using the placement rules based on the Hive user. The HiveServer2 fair scheduler queue mapping used for "non-impersonation" mode does not handle the primary-secondary queue mappings correctly.

Affected Version: CDH 6.0.0 and higher

Workaround: If you are a Hive and Sentry user, do not upgrade to CDH 6.0.0 or 6.0.1. This issue will be fixed as soon as possible. If you must use Hive and Sentry in CDH 6.0.0 or 6.0.1, see [YARN Dynamic Resource Pools Do Not Work with Hive When Sentry Is Enabled](#) for additional workarounds.

When vectorization is enabled on any file type (ORC, Parquet) queries that divide by zero using the modulo operator (%) return an error

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query divides by zero using the modulo operator (%), it returns the following error: Arithmetic exception [divide by] 0. For example, if you run the following query this issue is triggered: `SELECT 100 % column_c1 FROM table_t1;` and the value in `column_c1` is zero. The divide operator (/) is not affected by this issue.

Workaround: Disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-19564](#)**Cloudera Issue:** CDH-71211

When vectorization is enabled for Hive on any file type (ORC, Parquet) queries that perform comparisons in the `SELECT` clause on large values in columns with the data type of `BIGINT` might return wrong results

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query performs a comparison operation between very large values in columns that are `BIGINT` data types in the `SELECT` clause of the query, incorrect results might be returned. Comparison operators include `==`, `!=`, `<`, `<=`, `>`, and `>=`. This issue does not occur when the comparison operation is performed in the filtering clause of the query. This issue can also occur when the difference of values in such columns is out of range for a `LONG` (64-bit) data type. For example, if `column_c1` stores 8976171455044006767 and `column_c2` stores -7272907770454997143, a query such as `SELECT column_c1 < column_c2 FROM table_test` returns `true` instead of `false` because the difference (8976171455044006767 - (-7272907770454997143)) is 1.6249079225499E19 which is greater than 9.22337203685478E18, which is the maximum possible value that a `LONG` (64-bit) data type can hold.

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE_20207](#)**Cloudera Issue:** CDH-70996

Workaround: Use a `DECIMAL` type instead of `BIGINT` for columns that might contain very large values. Another option is to disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Specified column position in the ORDER BY clause is not supported for SELECT * queries

When column positions are specified in ORDER BY clauses, they are not honored for SELECT * queries and an error is returned as shown in the following example:

```
CREATE TABLE decimal_1 (id decimal(5,0));
SELECT * FROM decimal_1 ORDER BY 1 limit 100;
Error while compiling statement: FAILED: SemanticException [Error 10219]: Position in
ORDER BY is not supported when using SELECT *
```

Instead the query must list out the columns it is selecting.

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-68550

DirectSQL with PostgreSQL

Hive doesn't support Hive direct SQL queries with PostgreSQL database. It only supports this feature with MySQL, MariaDB, and Oracle. With PostgresSQL, direct SQL is disabled as a precaution, since there have been issues reported upstream where it is not possible to fallback on DataNucleus in the event of some failures, plus other non-standard behaviors. For more information, see [Hive Configuration Properties](#).

Affected Versions: All CDH versions

Cloudera Issue: CDH-49017

ALTER PARTITION ... SET LOCATION does not work on Amazon S3 or between S3 and HDFS

Cloudera recommends that you do not use ALTER PARTITION ... SET LOCATION on S3 or between S3 and HDFS. The rest of the ALTER PARTITION commands work as expected.

Affected Versions: All CDH versions

Cloudera Issue: CDH-42420

Commands run against an Oracle-backed metastore might fail

Commands run against an Oracle-backed Metastore fail with error:

```
javax.jdo.JDODataStoreException Incompatible data type for column TBLS.VIEW_EXPANDED_TEXT
  : was CLOB (datastore),
but type expected was LONGVARCHAR (metadata). Please check that the type in the datastore
and the type specified in the MetaData are consistent.
```

This error might occur if the metastore is run on top of an Oracle database with the configuration property `datanucleus.validateColumns` set to true.

Workaround: Set `datanucleus.validateColumns=false` in the `hive-site.xml` configuration file.

Affected Versions: All CDH versions

Cannot create archive partitions with external HAR (Hadoop Archive) tables

ALTER TABLE ... ARCHIVE PARTITION is not supported on external tables.

Affected Versions: All CDH versions

Cloudera Issue: CDH-9638

Object types Server and URI are not supported in "SHOW GRANT ROLE roleName on OBJECT objectName" statements

Workaround: Use SHOW GRANT ROLE roleName to list all privileges granted to the role.

Affected Versions: All CDH versions

Cloudera Issue: CDH-19430

HCatalog Known Issues



Note: As of CDH 5, HCatalog is part of Apache Hive.

There are no notable known issues in this release of HCatalog.

Hive on Spark (HoS) Known Issues

Hive on Spark queries fail with "Timed out waiting for client to connect" for an unknown reason

If this exception is preceded by logs of the form "client.RpcRetryingCaller: Call exception...", then this failure is due to an unavailable HBase service. On a secure cluster, `spark-submit` will try to obtain delegation tokens from HBase, even though Hive on Spark might not need them. So if HBase is unavailable, `spark-submit` throws an exception.

Workaround: Fix the HBase service, or set `spark.yarn.security.tokens.hbase.enabled` to `false`.

Affected Versions: CDH 5.7.0 and higher

Cloudera Issues: CDH-59591, CDH-59599

Hue Known Issues

Hue does not support the Spark App

Hue does not currently support the Spark application.

Connecting to PostgreSQL Database Fails with Error "No module named psycopg2"

When configuring Hue to use a PostgreSQL database, the connection fails with the following error:

```
Error loading psycopg2 module: No module named psycopg2
```

Workaround: Install the `psycopg2` Python package as documented in [Installing the psycopg2 Python Package](#).

Affected Versions: All CDH 6 versions

Fixed Versions: None

Apache Issue: N/A

Cloudera Issue: CDH-65804

Apache Impala Known Issues

The following sections describe known issues and workarounds in Impala, as of the current production release. This page summarizes the most serious or frequently encountered issues in the current release, to help you make planning decisions about installing and upgrading. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.



Note: The online issue tracking system for Impala contains comprehensive information and is updated in real time. To verify whether an issue you are experiencing has already been reported, or which release an issue is fixed in, search on the [Impala JIRA tracker](#).

Impala Known Issues: Startup

These issues can prevent one or more Impala-related daemons from starting properly.

Impala requires FQDN from hostname command on kerberized clusters

The method Impala uses to retrieve the host name while constructing the Kerberos principal is the `gethostname()` system call. This function might not always return the fully qualified domain name, depending on the network configuration. If the daemons cannot determine the FQDN, Impala does not start on a kerberized cluster.

Workaround: Test if a host is affected by checking whether the output of the `hostname` command includes the FQDN. On hosts where `hostname` only returns the short name, pass the command-line flag `--hostname=fully_qualified_domain_name` in the startup options of all Impala-related daemons.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4978](#)

Impala Known Issues: Crashes and Hangs

These issues can cause Impala to quit or become unresponsive.

Unable to view large catalog objects in catalogd Web UI

In catalogd Web UI, you can list metadata objects and view their details. These details are accessed via a link and printed to a string formatted using thrift's DebugProtocol. Printing large objects (> 1 GB) in Web UI can crash catalogd.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6841](#)

Impala Known Issues: Performance

These issues involve the performance of operations such as queries or DDL statements.

Metadata operations block read-only operations on unrelated tables

Metadata operations that change the state of a table, like COMPUTE STATS or ALTER RECOVER PARTITIONS, may delay metadata propagation of unrelated unloaded tables triggered by statements like DESCRIBE or SELECT queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6671](#)

Slow queries for Parquet tables with convert_legacy_hive_parquet_utc_timestamps=true

The configuration setting convert_legacy_hive_parquet_utc_timestamps=true uses an underlying function that can be a bottleneck on high volume, highly concurrent queries due to the use of a global lock while loading time zone information. This bottleneck can cause slowness when querying Parquet tables, up to 30x for scan-heavy queries. The amount of slowdown depends on factors such as the number of cores and number of threads involved in the query.



Note:

The slowdown only occurs when accessing TIMESTAMP columns within Parquet files that were generated by Hive, and therefore require the on-the-fly timezone conversion processing.

Workaround: Store the TIMESTAMP values as strings in one of the following formats:

- YYYY-MM-dd
- YYYY-MM-dd HH:mm:ss
- YYYY-MM-dd HH:mm:ss.SSSSSSSS

The date can have the 1-9 digits in the fractional part.

Impala implicitly converts such string values to TIMESTAMP in calls to date/time functions.

Affected Versions: All CDH 6 versions

Fixed Versions: CDH 6.1.0

Apache Issue: [IMPALA-3316](#)

Impala Known Issues: Security

These issues relate to security features, such as Kerberos authentication, Sentry authorization, encryption, auditing, and redaction.

In Impala with Sentry enabled, REVOKE ALL ON SERVER does not remove the privileges granted with the GRANT option

If you grant a role the ALL privilege at the SERVER scope with the WITH GRANT OPTION clause, you cannot revoke the privilege. Although the SHOW GRANT ROLE command will show that the privilege has been revoked immediately after you run the command, the ALL privilege will reappear when you run the SHOW GRANT ROLE command after Sentry refreshes.

Immediate Action Required: Once the privilege has been granted, the only way to remove it is to delete the role.

Affected Versions: CDH 6.0.0, CDH 6.0.1, CDH 5.15.0, CDH 5.15.1, CDH 5.14.x and all prior releases

Fixed Versions: CDH 6.1.0, CDH 6.0.2, CDH 5.16.0, CDH 5.15.2

Cloudera Issue: TSB-341

Impala does not support Heimdal Kerberos

Heimdal Kerberos is not supported in Impala.

Apache Issue: [IMPALA-7072](#)

Affected Versions: All CDH 6 versions

Impala Known Issues: Resources

These issues involve memory or disk usage, including out-of-memory conditions, the spill-to-disk feature, and resource management features.

Handling large rows during upgrade to CDH 5.13 / Impala 2.10 or higher

After an upgrade to CDH 5.13 / Impala 2.10 or higher, users who process very large column values (long strings), or have increased the --read_size configuration setting from its default of 8 MB, might encounter capacity errors for some queries that previously worked.

Resolution: After the upgrade, follow the instructions in [Handling Large Rows During Upgrade to CDH 5.13 / Impala 2.10 or Higher](#) to check if your queries are affected by these changes and to modify your configuration settings if so.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6028](#)

Configuration to prevent crashes caused by thread resource limits

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal error: Unable to allocate section memory!
terminate called after throwing an instance of
'boost::exception_detail::clone_impl<boost::exception_detail::error_info_injector<boost::thread_resource_error>'
```

Workaround:

In CDH 6.0 and lower versions of CDH, configure each host running an impalad daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

In CDH 6.1 and higher versions, it is unlikely that you will hit the thread resource limit. Configure each host running an impalad daemon with the following setting:

```
echo 8000000 > /proc/sys/vm/max_map_count
```

To make the above settings durable, refer to your OS documentation. For example, on RHEL 6.x:

1. Add the following line to /etc/sysctl.conf:

```
vm.max_map_count=8000000
```

2. Run the following command:

```
sysctl -p
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-5605](#)

Breakpad minidumps can be very large when the thread count is high

The size of the breakpad minidump files grows linearly with the number of threads. By default, each thread adds 8 KB to the minidump size. Minidump files could consume significant disk space when the daemons have a high number of threads.

Workaround: Add `--minidump_size_limit_hint_kb=size` to set a soft upper limit on the size of each minidump file. If the minidump file would exceed that limit, Impala reduces the amount of information for each thread from 8 KB to 2 KB. (Full thread information is captured for the first 20 threads, then 2 KB per thread after that.) The minidump file can still grow larger than the "hinted" size. For example, if you have 10,000 threads, the minidump file can be more than 20 MB.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3509](#)

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the impalad daemon.

Workaround: To monitor overall memory usage, use the top command, or add the memory figures in the Impala web UI `/memz` tab to JVM memory usage shown on the `/metrics` tab.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-691](#)

Impala Known Issues: Correctness

These issues can cause incorrect or unexpected results from queries. They typically only arise in very specific circumstances.

Incorrect result due to constant evaluation in query with outer join

An `OUTER JOIN` query could omit some expected result rows due to a constant such as `FALSE` in another join clause.

For example:

```
explain SELECT 1 FROM alltypestiny a1
    INNER JOIN alltypesagg a2 ON a1.smallint_col = a2.year AND false
    RIGHT JOIN alltypes a3 ON a1.year = a3.bigint_col;
+-----+
| Explain String
+-----+
| Estimated Per-Host Requirements: Memory=1.00KB Vcores=1
| 00:EMPTYSET
+-----+
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3094](#)

BST between 1972 and 1995

The calculation of start and end times for the BST (British Summer Time) time zone could be incorrect between 1972 and 1995. Between 1972 and 1995, BST began and ended at 02:00 GMT on the third Sunday in March (or second Sunday when Easter fell on the third) and fourth Sunday in October. For example, both function calls should return 13, but actually return 12, in a query such as:

```
select
  extract(from_utc_timestamp(cast('1970-01-01 12:00:00' as timestamp), 'Europe/London'),
  "hour") summer70start,
  extract(from_utc_timestamp(cast('1970-12-31 12:00:00' as timestamp), 'Europe/London'),
  "hour") summer70end;
```

Affected Versions: All CDH 6 versions

Fixed Versions: CDH 6.1

Apache Issue: [IMPALA-3082](#)

% escaping does not work correctly in a LIKE clause

If the final character in the RHS argument of a `LIKE` operator is an escaped `\%` character, it does not match a `%` final character of the LHS argument.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2422](#)

Crash: impala::Coordinator::ValidateCollectionSlots

A query could encounter a serious error if includes multiple nested levels of `INNER JOIN` clauses involving subqueries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2603](#)

Impala Known Issues: Metadata

These issues affect how Impala interacts with metadata. They cover areas such as the metastore database and the Impala Catalog Server daemon.

Concurrent catalog operations with heavy DDL workloads can cause queries with `SYNC_DDL` to fail fast

When Catalog Server is under a heavy load with concurrent catalog operations of long running DDLs, queries running with the `SYNC_DDL` query option can fail with the following message:

```
ERROR: CatalogException: Couldn't retrieve the catalog topic
version for the SYNC_DDL operation after 3 attempts. The operation has
been successfully executed but its effects may have not been
broadcast to all the coordinators.
```

The catalog operation is actually successful as the change has been committed to HMS and Catalog Server cache, but when Catalog Server notices a longer than expected time for it to broadcast the changes, it fails fast.

The coordinator daemons eventually sync up in the background.

Affected Versions: CDH versions 6.0 and 6.1

Apache Issue: [IMPALA-7961](#) / CDH-76345

Impala Known Issues: Interoperability

These issues affect the ability to interchange data between Impala and other systems. They cover areas such as data types and file formats.

Queries Stuck on Failed HDFS Calls and not Timing out

If the following error appears multiple times in a short duration while running a query, it would mean that the connection between the `impalad` and the HDFS NameNode is in a bad state and hence the `impalad` would have to be restarted:

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed to finish before the
<hdfs_operation_timeout_sec> second timeout "
```

Workaround: Restart the `impalad` in the bad state.

Affected Versions: All versions of Impala

Apache Issue: [HADOOP-15720](#)

Deviation from Hive behavior: Out of range values float/double values are returned as maximum allowed value of type (Hive returns NULL)

Impala behavior differs from Hive with respect to out of range float/double values. Out of range values are returned as maximum allowed value of type (Hive returns NULL).

Workaround: None

Affected Versions: All CDH 6 versions

Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink `hdfs.writeFormat` must be set to `Text`, rather than its default value of `Writable`. The `hdfs.writeFormat` setting must be changed to `Text` before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

Resolution: This information has been requested to be added to the upstream Flume documentation.

Affected Versions: All CDH 6 versions

Cloudera Issue: CDH-13199

Avro Scanner fails to parse some schemas

The default value in Avro schema must match the first union type. For example, if the default value is `null`, then the first type in the `UNION` must be "`null`".

Workaround: Swap the order of the fields in the schema specification. For example, use `["null", "string"]` instead of `["string", "null"]`. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-635](#)

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1024](#)

Incorrect results with basic predicate on CHAR typed column

When comparing a `CHAR` column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1652](#)

Impala Known Issues: Limitations

These issues are current limitations of Impala that require evaluation as you plan how to integrate Impala into your data management workflow.

Set limits on size of expression trees

Very deeply nested expressions within queries can exceed internal Impala limits, leading to excessive memory usage.

Workaround: Avoid queries with extremely large expression trees. Setting the query option `disable_codegen=true` may reduce the impact, at a cost of longer query runtime.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4551](#)

Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

Workaround: Use standard HDFS on all Impala nodes.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-77](#)

Hue and BDR require separate parameters for Impala Load Balancer

Cloudera Manager supports a single parameter for specifying the Impala Daemon Load Balancer. However, because BDR and Hue need to use different ports when connecting to the load balancer, it is not possible to configure the load balancer value so that BDR and Hue will work correctly in the same cluster.

Workaround: To configure BDR with Impala, use the load balancer configuration either without a port specification or with the Beeswax port.

To configure Hue, use the **Hue Server Advanced Configuration Snippet (Safety Valve)** for `impalad_flags` to specify the load balancer address with the HiveServer2 port.

Affected Versions: CDH versions from 5.11 to 6.0.1

Cloudera Issue: [OPSAPS-46641](#)

Impala Known Issues: Miscellaneous / Older Issues

These issues do not fall into one of the above categories or have not been categorized yet.

A failed CTAS does not drop the table if the insert fails

If a `CREATE TABLE AS SELECT` operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed `CREATE TABLE AS SELECT`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2005](#)

Casting scenarios with invalid/inconsistent results

Using a `CAST` function to convert large literal values to smaller types, or to convert special values such as `Nan` or `Inf`, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1821](#)

Impala Parser issue when using fully qualified table names that start with a number

A fully qualified table name starting with a number could cause a parsing error. In a name such as `db.571_market`, the decimal point followed by digits is interpreted as a floating-point number.

Workaround: Surround each part of the fully qualified name with backticks (`\``).

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-941](#)

Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

Workaround: Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) for details about modifying these environment settings.

Resolution: Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-532](#)

EMC Isilon Known Issues

CDH 6.0 is not currently supported on EMC Isilon.

Affected Versions: CDH 6.0.x

Apache Kafka Known Issues

Topics Created with the "kafka-topics" Tool Might Not Be Secured

Topics that are created and deleted via Kafka are secured (for example, auto created topics). However, most topic creation and deletion is done via the `kafka-topics` tool, which talks directly to ZooKeeper or some other third-party tool that talks directly to ZooKeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. Anyone with access to ZooKeeper can create and delete topics. They will not be able to describe, read, or write to the topics even if they can create them.

The following commands talk directly to ZooKeeper and therefore are not secured via Kafka:

- `kafka-topics.sh`
- `kafka-configs.sh`
- `kafka-preferred-replica-election.sh`
- `kafka-reassign-partitions.sh`

"`offsets.topic.replication.factor`" Must Be Less Than or Equal to the Number of Live Brokers

The `offsets.topic.replication.factor` broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a `GROUP_COORDINATOR_NOT_AVAILABLE` error until the cluster size meets this replication factor requirement.

Kafka May Be Stuck with Under-replicated Partitions after ZooKeeper Session Expires

This problem can occur when your Kafka cluster includes a large number of under-replicated Kafka partitions. One or more broker logs include messages such as the following:

```
[2016-01-17 03:36:00,888] INFO Partition [__samza_checkpoint_event-creation_1,3] on
broker 3: Shrinking ISR for partition [__samza_checkpoint_event-creation_1,3] from 6,5
to 5 (kafka.cluster.Partition)
[2016-01-17 03:36:00,891] INFO Partition [__samza_checkpoint_event-creation_1,3] on
broker 3: Cached zkVersion [66] not equal to that in zookeeper, skip updating ISR
(kafka.cluster.Partition)
```

There will also be an indication of the ZooKeeper session expiring in one or more Kafka broker logs around the same time as the previous errors:

```
INFO zookeeper state changed (Expired) (org.I0Itec.zkclient.ZkClient)
```

The log is typically in `/var/log/kafka` on each host where a Kafka broker is running. The location is set by the property `kafka.log4j.dir` in Cloudera Manager. The log name is `kafka-broker-hostname.log`. In diagnostic bundles, the log is under `logs/hostname-ip-address/`.

Workaround: To move forward after seeing this problem, restart the affected Kafka brokers. You can restart individual brokers from the **Instances** tab in the Kafka service page in Cloudera Manager.



Note: If restarting the brokers does not resolve the problem, you might not have this issue; see [KAFKA-3083 A soft failure in controller may leave a topic partition in an inconsistent state](#). This problem also involves the ZooKeeper session expiring, but will not involve the error message with Cached zkVersion [xx] not equal to that in zookeeper.

To reduce the chances of this issue happening again, do what you can to make sure ZooKeeper sessions do not expire:

- Reduce the potential for long garbage collection pauses by brokers:
 - Use a better garbage collection mechanism in the JVM, such as G1GC. You can do this by adding -XX:+UseG1GC in the broker_java_opts.
 - Increase broker heap size if it is too small (broker_max_heap_size). Be careful that you don't choose a heap size that can cause out-of-memory problems given all the services running on the node.
- Increase the ZooKeeper session timeout configuration on brokers (zookeeper.session.timeout.ms), to reduce the likelihood that sessions expire.
- Ensure ZooKeeper itself is well resourced and not overwhelmed so it can respond. For example, it is highly recommended to locate the ZooKeeper log directory on its own disk.

Affected Versions: CDK 1.4.x, 2.0.x, 2.1.x, 2.2.x

Fixed Versions:

- **Full Fix:** CDH 6.1.0
- **Partial Fix:** CDH 6.0.0, Kafka implementations with CDH 6.0.0 are less likely to encounter this issue.

Apache Issue: [KAFKA-2729](#)

Cloudera Issue: CDH-42514

Requests Fail When Sending to a Nonexistent Topic with "auto.create.topics.enable" Set to True

The first few produce requests fail when sending to a nonexistent topic with auto.create.topics.enable set to true.

Workaround: Increase the number of retries in the Producer configuration setting retries.

Custom Kerberos Principal Names Cannot Be Used for Kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start.

Workaround: None. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Performance Degradation When SSL Is Enabled

Significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with ssl.secure.random.implementation = SHA1PRNG. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Affected Versions: CDK 2.x and later

Fixed Versions: None

Apache Issue: KAFKA-2561

Cloudera Issue: None

Apache Kudu Known Issues

The following are known bugs and issues in Kudu. Note that this list is not exhaustive, and is meant to communicate only the most important known issues.

CFile Checksum Failure Causes Queries to Fail

When a CFile checksum fails, for example, due to a underlying disk corruption, queries against the replica will fail with an error message, such as this:

```
Unable to advance iterator: Corruption: checksum error on CFile block
```

Workaround: Remove the corrupted replica from the tablet's Raft configuration. See [Kudu Troubleshooting Guide](#) for the detailed steps.

Affected Versions: CDH 6.0.x and lower

Apache Issue: [KUDU-2469](#)

C++ Client Fails to Re-acquire Authentication Token in Multi-master Clusters

A security-related issue can cause Impala queries to start failing on busy clusters in the following scenario:

- The cluster runs with the `--rpc_authentication` set as optional or required. The default is optional. Secure clusters use required.
- The cluster is using multiple masters.
- Impala queries happen frequently enough that the leader master connection to some `impalad` isn't idle-closed (more than 1 query per 65 seconds).
- The connection stays alive for longer than the authentication token timeout (1 week by default).
- A master leadership change occurs after the authentication token expiration.

Impala queries will start failing with errors in the `impalad` logs like:

```
I0904 13:53:08.748968 95857 client-internal.cc:283] Unable to determine the new leader
Master: Not authorized: Client connection negotiation failed: client connection to
10.164.44.13:7051: FATAL_INVALID_AUTHENTICATION_TOKEN: Not authorized: authentication
token expired
I0904 13:53:10.389009 95861 status.cc:125] Unable to open Kudu table: Timed out:
GetTableSchema timed out after deadline expired
@ 0x95b1e9 impala::Status::Status()
@ 0xff22d4 impala::KuduScanNodeBase::Open()
@ 0xff101e impala::KuduScanNode::Open()
@ 0xb73ced impala::FragmentInstanceState::Open()
@ 0xb7532b impala::FragmentInstanceState::Exec()
@ 0xb64ae8 impala::QueryState::ExecFInstance()
@ 0xd15193 impala::Thread::SuperviseThread()
@ 0xd158d4 boost::detail::thread_data<>::run()
@ 0x129188a (unknown)
@ 0x7f717ceade25 start_thread
@ 0x7f717cbdb34d __clone
```

Impala shell queries will fail with a message like:

```
Unable to open Kudu table: Timed out: GetTableSchema timed out after deadline expired
```

Workaround:

- Restart the affected Impala Daemons. Restarting a daemon ensures the problem will not reoccur for at least the authentication token lifetime, which defaults to one week.
- Increase the authentication token lifetime (`--authn_token_validity_seconds`). Beware that raising this lifetime increases the window of vulnerability of the cluster if a client is compromised. It is recommended that you keep the token lifetime at one month maximum for a secure cluster. For unsecured clusters, a longer token lifetime is acceptable, and a 3 month lifetime is recommended.

Affected Versions: From CDH 5.11 through CDH 6.0.1

Apache Issue: [KUDU-2580](#)

Timeout Possible with Log Force Synchronization Option

If the Kudu master is configured with the `-log_force_fsync_all` option, tablet servers and clients will experience frequent timeouts, and the cluster may become unusable.

Affected Versions: All CDH 6 versions**Longer Startup Times with a Large Number of Tablets**

If a tablet server has a very large number of tablets, it may take several minutes to start up. It is recommended to limit the number of tablets per server to 1000 or fewer. The maximum allowed number of tablets is 2000 per server. Consider this limitation when pre-splitting your tables. If you notice slow start-up times, you can monitor the number of tablets per server in the web UI.

Affected Versions: All CDH 6 versions**Descriptions for Kudu TLS/SSL Settings in Cloudera Manager**

Use the descriptions in the following table to better understand the TLS/SSL settings in the Cloudera Manager Admin Console.

Field	Usage Notes
Kerberos Principal	Set to the default principal, kudu.
Enable Secure Authentication And Encryption	Select this checkbox to enable authentication <i>and</i> RPC encryption between all Kudu clients and servers, as well as between individual servers. Only enable this property after you have configured Kerberos.
Master TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu master host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to the Kudu master web UI.
Tablet Server TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu tablet server host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to Kudu tablet server web UIs.
Master TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu master host's private key (set in Master TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Tablet Server TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu tablet server host's private key (set in Tablet Server TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Master TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Tablet Server TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Enable TLS/SSL for Master Server	Enables HTTPS encryption on the Kudu master web UI.
Enable TLS/SSL for Tablet Server	Enables HTTPS encryption on the Kudu tablet server web UIs.

Affected Versions: All CDH 6 versions**Apache Oozie Known Issues****External ID of MapReduce action not filled properly and failing MR job treated as SUCCEEDED**

When a MapReduce action is launched from Oozie, the external ID field is not filled properly. It gets populated with the YARN ID of the LauncherAM, not with the ID of the actual MR job. If the MR job is submitted successfully and then fails, it will be treated as a successfully executed action.

Affected Versions: CDH 6.0.0 and higher**Fixed Version:** CDH 6.1.0 and higher**Apache Issue:** [OOZIE-3298](#)

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the `resume` command to tell Oozie to continue the workflow from the point at which it left off.

Affected Versions: CDH 5 and higher

Cloudera Issue: CDH-14623

Oozie works with MapReduce or YARN, but not both

The Oozie server works with a MapReduce (MRv1) cluster or a YARN (MRv2) cluster, but not both at the same time.

Workaround: Use two different Oozie servers.

Affected Versions: All

Apache Parquet Known Issues

There are no known issues in Parquet.

Apache Pig Known Issues

There are no known issues in this release.

Cloudera Search Known Issues

The current release includes the following known limitations:

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: `solr.sentry.enableSqlQuery=true`
- Graph: `solr.sentry.enableGraphQuery=true`
- Stream: `solr.sentry.enableStreams=true`

Workaround: None

Affected Versions: All CDH 6 releases

Cloudera Issue: CDH-66345

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search for CDH 5.5.0 includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-34050

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with CrunchIndexerTool, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None

Affected Versions: All**Cloudera Issue:** CDH-22190

The `quickstart.sh` file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if ZooKeeper and the NameNode are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the NameNode and ZooKeeper are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the MapReduce jobs and Solr commands that are run by the `quickstart`.

Affected Versions: All**Cloudera Issue:** CDH-19923

Field value class guessing and Automatic schema field addition are not supported with the `MapReduceIndexerTool` nor the `HBaseMapReduceIndexerTool`

The `MapReduceIndexerTool` and the `HBaseMapReduceIndexerTool` can be used with a Managed Schema created via NRT indexing of documents or via the Solr Schema API. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the `MapReduceIndexerTool` or `HBaseMapReduceIndexerTool`. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the Solr Schema API or index sample documents using NRT indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect using the List Fields API command.

Affected Versions: All**Cloudera Issue:** CDH-26856

The `Browse` and `Spell` Request Handlers are not enabled in schemaless mode

The `Browse` and `Spell` Request Handlers require certain fields be present in the schema. Since those fields cannot be guaranteed to exist in a Schemaless setup, the `Browse` and `Spell` Request Handlers are not enabled by default.

Workaround: If you require the “`Browse`” and “`Spell`” Request Handlers, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Affected Versions: All**Cloudera Issue:** CDH-19407

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. Blockcache writing is disabled by default.

Workaround: None

Affected Versions: All**Cloudera Issue:** CDH-17978

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given page explaining that access is denied, and instead receive a web page that never finishes loading.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-58276

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection.

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Affected Versions: All

Cloudera Issue: CDH-15441

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Affected Versions: All

Cloudera Issue: CDH-58694

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-21162

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-11357

Apache Sentry Known Issues

When granting privileges, a single transaction per grant causes long delays

Sentry takes a long time to grant or revoke a large number of column-level privileges that are requested in a single statement. For example if you execute the following command:

```
GRANT SELECT(col1, col2, ...) ON TABLE table1;
```

Sentry applies the grants to each column separately and the refresh process causes long delays.

Workaround: Split the grant statement up into smaller chunks. This prevents the refresh process from causing delays.

Affected Versions:

- CDH: 5.14.4
- CDH: 5.15.1
- CDH: 5.16.0
- CDH: 6.1.0

Fixed Versions:

- CDH 5.16.1 and above
- CDH 6.2.0 and above

Cloudera Issue: CDH-74982

SHOW ROLE GRANT GROUP raises exception for a group that was never granted a role

If you run the command SHOW ROLE GRANT GROUP for a group that has never been granted a role, beeline raises an exception. However, if you run the same command for a group that does not have any roles, but has at one time been granted a role, you do not get an exception, but instead get an empty list of roles granted to the group.

Workaround: Adding a role will prevent the exception.

Affected Versions:

- CDH 5.16.0
- CDH 6.0.0

Cloudera Issue: CDH-71694**GRANT/REVOKE operations could fail if there are too many concurrent requests**

Under a significant workload, Grant/Revoke operations can have issues.

Workaround: If you need to make many privilege changes, plan them at a time when you do not need to do too many at once.

Affected Versions: CDH 5.13.0 and above

Apache Issue: [SENTRY-1855](#)**Cloudera Issue:** CDH-56553**Creating large set of Sentry roles results in performance problems**

Using more than a thousand roles/permissions might cause significant performance problems.

Workaround: Plan your roles so that groups have as few roles as possible and roles have as few permissions as possible.

Affected Versions: CDH 5.13.0 and above

Cloudera Issue: CDH-59010

Users can't track jobs with Hive and Sentry

As a prerequisite of enabling Sentry, Hive impersonation is turned off, which means all YARN jobs are submitted to the Hive job queue, and are run as the `hive` user. This is an issue because the YARN History Server now has to block users from accessing logs for their own jobs, since their own usernames are not associated with the jobs. As a result, end users cannot access any job logs unless they can get `sudo` access to the cluster as the `hdfs`, `hive` or other admin users.

In CDH 5.8 (and higher), Hive overrides the default configuration, `mapred.job.queuename`, and places incoming jobs into the connected user's job queue, even though the submitting user remains `hive`. Hive obtains the relevant queue/username information for each job by using YARN's `fair-scheduler.xml` file.

Affected Versions: CDH 5.2.0 and above

Cloudera Issue: CDH-22890

Column-level privileges are not supported on Hive Metastore views

GRANT and REVOKE for column level privileges is not supported on Hive Metastore views.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-754](#)

SELECT privilege on all columns does not equate to SELECT privilege on table

Users who have been explicitly granted the `SELECT` privilege on all columns of a table, will *not* have the permission to perform table-level operations. For example, operations such as `SELECT COUNT (1)` or `SELECT COUNT (*)` will not work even if you have the `SELECT` privilege on all columns.

There is one exception to this. The `SELECT * FROM TABLE` command will work even if you do not have explicit table-level access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-838](#)

The EXPLAIN SELECT operation works without table or column-level privileges

Users are able to run the `EXPLAIN SELECT` operation, exposing metadata for all columns, even for tables/columns to which they weren't explicitly granted access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-849](#)

Object types Server and URI are not supported in `SHOW GRANT ROLE roleName on OBJECT objectName`

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Apache Issue: N/A

Cloudera Issue: CDH-19430

Relative URI paths not supported by Sentry

Sentry supports only absolute (not relative) URI paths in permission grants. Although some early releases (for example, CDH 5.7.0) might not have raised explicit errors when relative paths were set, upgrading a system that uses relative paths causes the system to lose Sentry permissions.

Affected Versions: All versions. Relative paths are not supported in Sentry for permission grants.

Resolution: Revoke privileges that have been set using relative paths, and grant permissions using absolute paths before upgrading.

Absolute (Use this form)	Relative (Do not use this form)
<code>hdfs://absolute/path/</code>	<code>hdfs://relative/path</code>
<code>s3a://bucketname/</code>	<code>s3a://bucketname</code>

Apache Spark Known Issues

The following sections describe the current known issues and limitations in Apache Spark 2.x as distributed with CDH 6. In some cases, a feature from the upstream Apache Spark project is currently not considered reliable enough to be supported by Cloudera.

PySpark broadcast variables fail when disk encryption is enabled

When disk encryption is enabled, PySpark broadcast variables fail with the following stack trace:

```
Traceback (most recent call last): File "broadcast.py", line 37, in <module>
words_new.value File "/pyspark.zip/pyspark/broadcast.py", line 137, in value
File "pyspark.zip/pyspark/broadcast.py", line 122, in load_from_path File
"pyspark.zip/pyspark/broadcast.py", line 128, in load EOFError: Ran out of input
```

Workaround: None

Affected Versions: CDH 6.0.1

Apache Issue: [SPARK-26201](#)

Cloudera Issue: CDH-76116

Spark Streaming jobs loop if missing Kafka topic

Spark jobs can loop endlessly if the Kafka topic is deleted while a Kafka streaming job (which uses KafkaSource) is in progress.

Workaround: Stop a job before deleting a Kafka topic.

Affected Versions: All

Cloudera Issue: CDH-57903, CDH-64513

Spark SQL does not respect size limit for the varchar type

Spark SQL treats `varchar` as a string (that is, there no size limit). The observed behavior is that Spark reads and writes these columns as regular strings; if inserted values exceed the size limit, no error will occur. The data will be truncated when read from Hive, but not when read from Spark.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Apache Issue: [SPARK-5918](#)

Cloudera Issue: CDH-33642

Spark SQL does not prevent you from writing key types not supported by Avro tables

Spark allows you to declare DataFrames with any key type. Avro supports only string keys and trying to write any other key type to an Avro table will fail.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33648

Spark SQL does not support timestamp in Avro tables

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33649

Spark SQL does not respect Sentry ACLs when communicating with Hive metastore

Even if user is configured via Sentry to not have read permission to a Hive table, a Spark SQL job running as that user can still read the table's metadata directly from the Hive metastore. **Cloudera Issue:** CDH-33658

Dynamic allocation and Spark Streaming

If you are using Spark Streaming, Cloudera recommends that you disable dynamic allocation by setting `spark.dynamicAllocation.enabled` to `false` when running streaming applications.

Limitation with Region Pruning for HBase Tables

When SparkSQL accesses an HBase table through the HiveContext, region pruning is not performed. This limitation can result in slower performance for some SparkSQL queries against tables that use the HBase SerDes than when the same table is accessed through Impala or Hive.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-56330

Running `spark-submit` with `--principal` and `--keytab` arguments does not work in client mode

The `spark-submit` script's `--principal` and `--keytab` arguments do not work with Spark-on-YARN's client mode.

Workaround: Use `cluster` mode instead.

Affected Versions: All

Long-running apps on a secure cluster might fail if driver is restarted

If you submit a long-running app on a secure cluster using the `--principal` and `--keytab` options in cluster mode, and a failure causes the driver to restart after 7 days (the default maximum HDFS delegation token lifetime), the new driver fails with an error similar to the following:

```
Exception in thread "main"
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken):
  token <token_info> can't be found in cache
```

Workaround: None

Affected Versions: CDH 6.0

Apache Issue: [SPARK-23361](#)

Cloudera Issue: CDH-64865

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Apache Sqoop Known Issues

Column names cannot start with a number when importing data with the `--as-parquetfile` option.

Currently, Sqoop is using an Avro schema when writing data as a parquet file. The Avro schema requires that column names do not start with numbers, therefore Sqoop is renaming the columns in this case, prepending them with an underscore character. This can lead to issues when one wants to reuse the data in other tools, such as Impala.

Workaround: Rename the columns to comply with Avro limitations (start with letters or underscore, as specified in the [Avro documentation](#)).

Cloudera Issue: None

MySQL JDBC driver shipped with CentOS 6 systems does not work with Sqoop

CentOS 6 systems currently ship with version 5.1.17 of the MySQL JDBC driver. This version does not work correctly with Sqoop.

Workaround: Install version 5.1.31 of the JDBC driver as detailed in [Installing the JDBC Drivers for Sqoop 1](#).

Affected Versions: MySQL JDBC 5.1.17, 5.1.4, 5.3.0

Cloudera Issue: CDH-23180

MS SQL Server "integratedSecurity" option unavailable in Sqoop

The `integratedSecurity` option is not available in the Sqoop CLI.

Workaround: None

Cloudera Issue: None

Sqoop1 (doc import + `--as-parquetfile`) limitation with KMS/KTS Encryption at Rest

Due to a limitation with Kite SDK, it is not possible to use (`sqoop import --as-parquetfile`) with KMS/KTS Encryption zones. See the following example.

```
sqoop import --connect jdbc:db2://djaxludb1001:61035/DBBAT003 --username=dh810202 --P
--target-dir /data/hive_scratch/ASDISBURSEMENT --delete-target-dir -ml --query "select
disbursementnumber,disbursementdate,xmldata FROM DB2dba.ASDISBURSEMENT where
DISBURSEMENTNUMBER = 2011113210000115311 AND \$CONDITIONS" -hive-import --hive-database
```

```

adminserver -hive-table asdisbursement_dave --map-column-java XMLDATA=String
--as-parquetfile

16/12/05 12:23:46 INFO mapreduce.Job: map 100% reduce 0%
16/12/05 12:23:46 INFO mapreduce.Job: Job job_1480530522947_0096 failed with state FAILED
due to: Job commit failed: org.kitesdk.data.DatasetIOException: Could not move contents
of
hdfs://AJAX01-ns/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096
to hdfs://AJAX01-ns/data/RetiredApps/INS/AdminServer/asdisbursement_dave
<SNIP>
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.IOException):
/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096/5ddcac42-5d69-4e46-88c2-17bedac4858.parquet
can't be moved into an encryption zone.

```

Workaround: If you use the Parquet Hadoop API based implementation for importing into Parquet, specify a `--target-dir` which is the same encryption zone as the Hive warehouse directory.

If you use the Kite Dataset API based implementation, use an alternate data file type, for example text or avro.

Apache Issue: SQUOP-2943

Cloudera Issue: CDH-40826

Doc import as Parquet files may result in out-of-memory errors

Out-of-memory (OOM) errors can be caused in the following two cases:

- With many very large rows (multiple megabytes per row) before initial-page-run check (ColumnWriter)
- When rows vary significantly by size so that the next-page-size check is based on small rows and is set very high followed by many large rows

Apache Issue: PARQUET-99

Workaround: None, other than restructuring the data.

Apache ZooKeeper Known Issues

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

Technical Service Bulletin 2019-310 (TSB)

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service, or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper

Releases affected: Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower

Users affected: All

Date/time of detection: June 7, 2018

Severity (Low/Medium/High): 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))

Impact: Remote code execution

CVE: CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings “Enable TLS/SSL for ZooKeeper JMX” and “Enable TLS client authentication for JMX port” on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting “Enable JMX Agent” on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0

CDH 6.0.0 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for CDH 6.0.0:

New Features in CDH 6.0.0

See below for new features in CDH 6.0.0, grouped by component:

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Apache Avro

The following are new features from the upstream release Avro 1.8.2 that are available in the CDH 6.0.0 release. For more information on Avro 1.8.2, see [Apache Avro 1.8.2 Documentation](#).

- [AVRO-834](#): Data File corruption recovery tool
- [AVRO-1502](#): Avro objects should implement Serializable
- [AVRO-1684](#): Add date, time, and timestamp to specific object model classes
- [AVRO-1402](#): Support for DECIMAL type
- [AVRO-1439](#): MultipleInputs equivalent for Avro MR

Apache Crunch

There are no notable new features in this release.

Apache Flume

Apache Flume contains the following new features:

- CDH 6.0.0 uses HBase 2.0.
- HBaseSink is replaced with HBase2Sink. For more information about Flume HBase2Sink changes in CDH 6.0.0, see [Incompatible Changes in CDH 6.0.0](#) on page 528.

Serializer Class Names

The HBase serializer classes have been renamed to change `Hbase` to `HBase` and to include the version number.

The following class names have been updated:

New HBase2 Name	Legacy Name
org.apache.flume.sink.hbase2.HBase2EventSerializer	org.apache.flume.sink.hbase.HbaseEventSerializer
org.apache.flume.sink.hbase2.SimpleHBase2EventSerializer	org.apache.flume.sink.hbase.SimpleHbaseEventSerializer
org.apache.flume.sink.hbase2.RegexHBase2EventSerializer	org.apache.flume.sink.hbase.RegexHbaseEventSerializer
org.apache.flume.sink.hbase2.HBase2Sink	org.apache.flume.sink.hbase.HBaseSink

In CDH, the legacy names have been kept for backward compatibility and are aliases to the new names. Existing configurations will work without changes, but new configurations should use the HBase 2.x serializer class names. In the upstream Apache Flume version, the legacy names refer to HBase 1.x classes and the new names refer to HBase 2.x classes.

For example, the following configuration contains the new class name and will use HBase2Sink for CDH and upstream Apache Flume:

```
host1.sinks.sink1.type = org.apache.flume.sink.hbase.HBase2Sink # or hbase2
host1.sinks.sink1.serializer = org.apache.flume.sink.hbase.SimpleHBase2EventSerializer
```

However, the following configuration, which contains the legacy name, will use HBase2Sink for CDH, but will use the old HBaseSink for upstream Apache Flume:

```
host1.sinks.sink1.type = org.apache.flume.sink.hbase.HBaseSink # or hbase
host1.sinks.sink1.serializer = org.apache.flume.sink.hbase.SimpleHbaseEventSerializer
```

Apache Hadoop

Hadoop Common

As part of the Hadoop 3 rebase, CDH 6.0.0 supports the updated version of the Hadoop shell discussed in [HADOOP-9902](#). This introduces several incompatible changes as noted [HADOOP-9902](#). Customers using the Hadoop shell may need to make adjustments to their scripts.

For more information, see [HADOOP-9902](#) and the Apache Hadoop shell documentation.

HDFS

There are no notable new features in this release.

MapReduce

There are no notable new features in this release.

YARN

Resource Types

Support has been added for YARN resource types. In addition to existing CPU and memory resource-based scheduling, you can now utilize user-defined countable resources like GPUs. Use this feature to schedule YARN tasks based on an available resource.

For more information, see [YARN Resource Configuration](#).

YARN JARs

YARN jobs running MapReduce2 are now configured to read MapReduce JARs from HDFS instead of from local disk. This makes jobs more robust during rolling upgrades when the local binaries are modified while a job is executing. Clusters created in or upgrading to CDH 6.0 will use this new behavior.

Apache HBase

CDH 6.0.0 includes most of the features introduced in HBase 2.0 and ensures that they are fully compatible with other CDH components. However, HBase 2.0 may include features that are not supported by Cloudera as part of CDH 6.0.0.

Supportability

CDH 6.0.0 contains a significant number of enhancements designed to make HBase easier to configure and to support.

Assignment Manager

CDH 6.0.0 has some significant architectural changes to the assignment manager. The assignment manager is a component in the HBase master that manages region to region server assignment and ensures that any one region replica is assigned to just one region server. The overall goal is to make HBase more resilient through providing a unified place within HBase to handle some tricky scenarios related to region assignment including:

- Multi step procedures with rollback and rollforward capabilities in case of failure.
- Notifications across multiple region servers.
- Coordination of long-running and/or computationally expensive procedures.
- Procedures that require precise coordination across multiple machines (e.g., snapshots).

Note that HBCK is read-only in CDH 6.0.0. [HBASE-18792](#) is listed in the CDH HBase [Incompatible Changes](#).

Replication

Replication peers can be defined for entire namespaces instead of for each individual table. For clusters with hundreds or thousands of tables defining replication configuration settings for each table is cumbersome and error prone and being able to define configuration at the namespace level is a significant improvement.

- [HBASE-16447](#): Replication by namespaces config in peer; Set a namespace in peer config means that all tables in this namespace will be replicated.
- [HBASE-19293](#): Support adding a new replication peer in disabled state.
- [HBASE-17314](#): (Replication) Limit total buffered size for all replication sources - Add a conf "replication.total.buffer.quota" to limit total size of buffered entries in all replication peers. It will prevent server getting OOM if there are many peers. Default value is 256MB.
- [HBASE-17296](#): Peer level throttling for replication - Add the bandwidth upper limit to ReplicationPeerConfig and a new shell cmd set_peer_bandwidth to update the bandwidth in need.
- [HBASE-16466](#): HBase snapshots supported by VerifyReplication tool - verifyrep can now compare source table snapshot against peer table snapshot which reduces load on RS by reading data from HDFS directly using Snapshot scanners.

Balancer Improvements

Periodically, the HBase master will look at load across region servers and attempt to balance it out by reassigning regions. The handling of this cluster maintenance has improved in CDH 6.0.0:

- [HBASE-18164](#): Much faster stochastic load balancer; Improvements in locality cost function and candidate generator
- [HBASE-17178](#): Region balancing throttling - New config to protect availability of cluster by specifying maximum number of regions (in percentage) that can be in transition at any time.
- [HBASE-14309](#): Added boolean parameter, force, to 'balancer' command so that admin can force region balancing even when there is region (other than hbase:meta) in transition.

Shell

HBase's command line shell has been enhanced with several new commands that allow for easier cluster administration. New commands allow for administrators to better understand the current compaction state of a table and to clear the compaction queue if needed. There is also a new command to display metadata associated with a region.

New shell commands:

- [HBASE-17928](#): Clear compaction queues : clear_compaction_queues.
- [HBASE-16147](#): Get compaction state of a table : compaction_state.
- [HBASE-14925](#): Displaying the table's region info : list_regions.

Major Compaction Tool

HBase deployments with a heavy write workload often disable HBase's automatic handling of major compactions in order to reduce IO during peak cluster load. HBase now ships a tool to make it easier for operators who then have to ensure major compactions can happen during off-hours.

- [HBASE-19528](#): Major Compaction Tool - Tool allows you to compact a cluster with given concurrency of regionservers compacting at a given time.

Metrics

In CDH 6.0.0, there are a significant number of new metrics available. Additionally, client side metrics are now collected by default. Some examples include:

- Block cache metrics for the primary region replica instead of lumping metrics for the primary, secondary, and tertiary replicas.
- Per table metrics that describe the memstore, storefiles, and region size.
- Metrics that track assignment manager responsibilities like merging and splitting regions, assigning and unassigning regions to region servers, and state information that exists at the time in the event of a server crash.
- Latency metrics for checkAndPut, checkAndDelete, putBatch and deleteBatch.
- Counts for hedgedReads and hedgedWrites Latency histograms on a per-region basis.

- Metrics for tracking coprocessor usage.

New metrics:

- [HBASE-14314](#): 3 new Block cache metrics for primary region replica.
- [HBASE-15671](#): Per-table metrics on memstore, storefile and regionsize.
- [HBASE-15518](#): Per-table metrics aggregated from per-region metrics in region server metrics.
- [HBASE-16549](#): Few AM metrics.
- [HBASE-18374](#): Latency metrics for checkAndPut, checkAndDelete, putBatch and deleteBatch.
- [HBASE-12220](#): hedgedReads and hedgedReadWins counts.
- [HBASE-19285](#): Per-RegionServer table latency histograms have been returned to HBase.

Metrics changes:

- [HBASE-15943](#): New "Process Metrics" tab that dumps mbean -- mostly jvm -- metrics.
- [HBASE-14583](#): Client side metrics are enabled by default - Sets the default value of hbase.client.metrics.enable=true.

Spark 2 Integration

Spark integration for accessing data in HBase has been updated to work with the Spark 2 version that ships with CDH 6.x. It should provide equivalent functionality to the CDH 5.x integration with Spark 1.6.

Improved Performance

CDH 6.0.0 has an extensive set of enhancements aimed at improving overall performance. The following sections describe some of the improvements, but this list is not exhaustive.

Off-Heap Write Cache

CDH 6.0.0 includes the ability to define an off-heap write cache. Off-heap caches allow HBase to manage memory directly instead of delegating memory management to java. HBase has much more context on how to safely reclaim memory it's using compared to the java garbage collection process which allows HBase to more efficiently manage memory usage. Off-heap caching allows HBase to use more memory than would be practical if delegating the responsibility to java where "stop the world" garbage collection can cause large spikes in latency. One consequence of allowing significantly more write side cache memory through an off-heap cache is an increase in density of regions per region server.

Off-Heap Read Path

It is now possible to configure HBase so that data remains off-heap along the entire read path, avoiding copies of the data onto the java heap. With this feature enabled cached data will have a similar latency to on-heap caching in earlier versions of HBase but without putting GC pressure on the JVM. For more information on configuring and running a deployment with an off-heap read path see the Apache HBase documentation reference guide section on [Offheap Read Path](#).

Improved Cleaning of Old Files

CDH 6.0.0 has an option to allow HFileCleaner to use multiple threads. HFileCleaner is a background process that removes files containing dereferenced data. This new configuration is useful when a system is under substantial write load to ensure that clean up operations complete in a timely manner to prevent the archive directory from ballooning in size.

- [HBASE-18083](#): Configure HFileCleaner to use multiple threads for large/small (archived) hfile cleaning.
- [HBASE-17215](#): Separate small/large file delete threads in HFileCleaner to accelerate archived hfile cleanup speed.
- [HBASE-18309](#): CleanerChore (for cleaning up HFiles and old log files) now supports multiple threads.

Streaming Scan API

An additional performance enhancement in CDH 6.0.0 is the introduction of a "streaming scan" API. This API can improve the performance of intensive workloads by allowing multiple chunks of data to be prefetched concurrently. Previously, prefetch operations were synchronous, i.e, they required the previous prefetch operation to complete. Synchronous prefetches can result in a "stop-and-wait" access pattern where an application has finished processing

its previous chunk of data and is now waiting for the next chunk of data to arrive. Prefetching a second chunk of data before the first is entirely processed helps eliminate the “stop-and-wait” access pattern.

- [HBASE-13071](#): A pipelined scan API for speeding up applications that combine massive data traversal with compute-intensive processing.

New RPC Server

A new RPC server based on Netty4 is included in CDH 6.0.0. This RPC server option can increase the performance of workloads with large amounts of random reads and writes. This new RPC server option is on by default.

- [HBASE-17263](#), [HBASE-19323](#): NettyRpcServer- A new RPC server based on Netty4 which can improve random read (get) performance. It is also the default RPC server replacing SimpleRpcServer.
- [HBASE-15136](#): New RPC scheduler - CoDel - The purpose is to prevent long standing call queues caused by discrepancy between request rate and available throughput, caused by kernel/disk IO/networking stalls.

Note:

- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.

These changes should eliminate a number of issues in earlier versions of HBase commonly referred to as "regions stuck in transition."

- Split/Merge have moved to the Master; it runs them now. RegionServer hooks are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-16414](#): Improve performance for RPC encryption with Apache Common Crypto.
- [HBASE-16023](#): Fastpath for the FIFO rpcscheduler - Will shine best when high random read workload (YCSB workload for instance).
- [HBASE-15994](#): Allow selection of RpcSchedulers.

Improved MapReduce over Snapshots

HBase provides users the ability to run map/reduce jobs over HBase snapshots. Snapshots are immutable, and by running map/reduce jobs directly against them clients can avoid performance overhead associated with going through the region servers. In prior HBase versions exactly one map task per region was created. In CDH 6.0.0, the ability to assign multiple mappers to a particular region is possible. This change can further improve the performance of map/reduce jobs over HBase snapshots and eliminates the need to design schemas optimized for the prior limitation.

- [HBASE-18090](#): TableSnapshotInputFormat now supports multiple mappers per region.

Heterogeneous Storage Management

HDFS has supported heterogeneous storage management (HSM) for several releases. This feature allows users to specify that certain HDFS files or directories should reside on specific physical media. In CDH 6.0.0, users can do something very similar at the logical, column-family level. Currently supported storage policies include ALL_SSD/ONE_SSD/HOT/WARM/COLD.

Note that HBase has supported setting storage policies on WALS since CDH 5.7.

- [HBASE-14061](#): Column Family-level Storage Policy - Currently supported storage policies include ALL_SSD/ONE_SSD/HOT/WARM/COLD
- [HBASE-15172](#): Support setting storage policy in bulkload

Client Robustness Against Slow Region Servers

Clients can now optionally elect to handle slow region servers in application code. By setting the configuration value for `hbase.client.perserver.requests.threshold` an application can control how many concurrent requests to a single region server are allowed. After reaching this threshold the HBase client code will throw a `ServerTooBusyException` rather than attempting to establish an additional connection. If an HBase cluster has a small number of region servers that are having trouble this will allow an application to react in some way rather than simply waiting for the server to eventually respond. Because using this feature requires application side changes to handle retry logic you must opt-in to it; by default `hbase.client.perserver.requests.threshold` is set to practically unlimited.

- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.

Maven Archetypes

- [HBASE-14877](#): Maven archetype: client application - Introduces a new infrastructure for creation and maintenance of Maven archetypes in the context of the hbase project, and it also introduces the first archetype, which end-users may utilize to generate a simple hbase-client dependent project.

JIRAs

CDH 6.0.0 contains the following new features:

- [HBASE-18519](#): Use builder pattern to create cell
- [HBASE-13259](#): mmap() based BucketCache IOEngine - Can be configured using the property hbase.bucketcache.ioengine.
- [HBASE-14247](#): Separate the old WALs into different regionserver directories - Disabled by default. Can be enabled by using the configuration hbase.separate.oldlogdir.by.regionserver.
- [HBASE-12706](#): Support multiple port numbers in ZK quorum string - hbase.zookeeper.quorum configuration now allows servers together with client ports consistent with the way Zookeeper java client accepts the quorum string. In this case, using hbase.zookeeper.clientPort is not needed. eg.
hbase.zookeeper.quorum=myserver1:2181,myserver2:20000,myserver3:31111.
- [HBASE-15806](#): A coprocessor service based export tool - org.apache.hadoop.hbase.coprocessor.Export.
- [HBASE-15536](#): Changed default WALProvider to AsyncFSWALProvider.
- [HBASE-18533](#): New configurations for BucketCache.
- [HBASE-17583](#): Add inclusive/exclusive support for startRow and endRow of scan for sync client.
- [HBASE-16672](#): Added option for bulk load to always copy hfile(s) instead of renaming.
- [HBASE-19336](#): Two new RSGroup commands to operate on namespace level - move_namespaces_rsgroup and move_servers_namespaces_rsgroup.
- [HBASE-17437](#): Support WAL directory outside of the root directory - Useful for cloud deployments because it allows keeping WALs on a faster network drive than hfiles, giving better latency and throughput.
- [HBASE-16213](#): New data block encoding for fast random gets.
- [HBASE-15265](#): Asynchronous WALs.
- [HBASE-14969](#): Throughput controller for flush - Added means of throttling flush throughput. By default there is no limit (NoLimitThroughputController). An alternative controller, PressureAwareFlushThroughputController, allows specifying throughput bounds. See PressureAwareFlushThroughputController.java class for detail.
- [HBASE-15187](#): Added CSRF prevention filter to REST gateway - Disabled by default, can be enabled by using the configuration hbase.rest.csrf.enabled.
- [HBASE-15711](#): Added client side property to provide more logging details for batch errors.
- [HBASE-15511](#): New ClusterStatus.Option field to limit the scope of response.
- [HBASE-15576](#): Scanning cursor to prevent blocking long time on ResultScanner.next().
- [HBASE-15921](#): New AsyncTable client.
- [HBASE-11344](#): Hide row keys from the web UIs
- [HBASE-3462](#): UI pages for splitting/merging now operate by taking a row key prefix from the user rather than a full region name.
- [HBASE-8410](#): Basic quota support for namespaces - Namespace auditor provides basic quota support for namespaces in terms of number of tables and number of regions.

Apache Hive / Hive on Spark / HCatalog

Apache Hive

See below for the new features added to Hive in CDH 6.0.0:

- [Query Vectorization Support for Parquet Files](#) on page 508
- [Support for UNION DISTINCT](#) on page 508
- [Support for NULLS FIRST/NULLS LAST](#) on page 508
- [Added Support for Windowing and Analytics Functions](#) on page 509

- [Table or Partition Statistics Editing](#) on page 509
- [SHOW CREATE DATABASE Support](#) on page 509
- [Support for Multiple-Column IN Clause](#) on page 509
- [Support for More Hive Functions](#) on page 509

Query Vectorization Support for Parquet Files

By default, the Hive query execution engine processes one row of a table at a time. The single row of data goes through all the operators in the query before the next row is processed, resulting in very inefficient CPU usage. In vectorized query execution, data rows are batched together and represented as a set of column vectors. The query engine then processes these vectors of columns, which greatly reduces CPU usage for typical query operations like scans, filters, aggregates, and joins.

Hive query vectorization is enabled by setting the `hive.vectorized.execution.enabled` property to `true`. In both CDH 5 and CDH 6, this property is set to true by default. But in CDH 5, vectorized query execution in Hive is only possible on ORC-formatted tables, which Cloudera recommends that you do not use for overall compatibility with the CDH platform. Instead, Cloudera recommends that you use tables in the Parquet format because all CDH components support this format and can be consumed by all CDH components. In CDH 6, query vectorization is supported for Parquet tables in Hive.

For more information, see [Enabling Query Vectorization](#) and [Apache Parquet Tables with Hive in CDH](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
- The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
- The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

Support for NULLS FIRST/NULLS LAST

Support has been added for `NULLS FIRST` and `NULLS LAST` options. These options can be used to determine whether null values appear before or after non-null data values when the `ORDER BY` clause is used. Hive follows the [SQL:2003 standard](#) for this feature, but the SQL standard does not specify the behavior by default. By default in Hive, null values are sorted as if lower than non-null values. This means that `NULLS FIRST` is the default behavior for `ASC` order, and `NULLS LAST` is the default behavior for `DESC` order. See [Syntax of Order By](#) on the Apache Hive wiki and [HIVE-12994](#) for further details.

Here are some usage examples:

```
SELECT x.* FROM table1 x ORDER BY a ASC NULLS FIRST;
SELECT x.* FROM table1 x ORDER BY a ASC NULLS LAST;
```

Added Support for Windowing and Analytics Functions

Support for the following has been added to CDH 6.0:

- Using DISTINCT with windowing functions. See [HIVE-9534](#) for details.
- Support for ORDER BY and a windowing clause when DISTINCT is used in a partitioning clause. See [HIVE-13453](#) for details.
- Support to reference aggregate functions within the OVER clause. See [HIVE-13475](#) for details.

For further details, see the [Apache Language Manual on Windowing and Analytics](#).

Table or Partition Statistics Editing

Support has been added for editing the statistics information that is stored for a table or a partition. For example, you can run the following statement to set the number of rows for a table to 1000:

```
ALTER TABLE table1 UPDATE STATISTICS SET (' numRows'='1000');
```

For more information, see [HIVE-12730](#) and the [Apache wiki](#).

SHOW CREATE DATABASE Support

Support has been added for the SHOW CREATE DATABASE command, which prints out the DDL statement that was used to create a database:

```
SHOW CREATE DATABASE database1;
```

For more information, see [HIVE-11706](#)

Support for Multiple-Column IN Clause

Support has been added so that the IN clause in queries operates over multiple column references. The new syntax is boldfaced in the following example:

```
CREATE TABLE test (col1 int, col2 int);
INSERT INTO TABLE test VALUES (1, 1), (1, 2), (2, 1), (2, 2);
SELECT * FROM test;
+-----+-----+
| test.col1 | test.col2 |
+-----+-----+
| 1          | 1          |
| 1          | 2          |
| 2          | 1          |
| 2          | 2          |
+-----+-----+
SELECT * FROM test WHERE (col1, col2) IN ((1, 1));
+-----+-----+
| test.col1 | test.col2 |
+-----+-----+
| 1          | 1          |
+-----+-----+
```

For more information, see [HIVE-11600](#)

Support for More Hive Functions

Support has been added for the following Hive UDFs:

bround	chr	factorial
floor_day	floor_hours	floor_minute

floor_month	floor_quarter	floor_second
floor_week	floor_year	grouping
mask	mask_first_n	mask_hash
mask_last_n	mask_show_first_n	mask_show_last_n
quarter	replace	sha1
sha	shiftleft	shiftright
shiftrightunsigned	substring_index	

All built-in Hive functions can be listed with the command `SHOW FUNCTIONS;` and a short description that explains what a function does is returned with the command `DESCRIBE <function_name>;` For more information about Hive functions, see the [Apache wiki](#) and [Managing UDFs](#) in the Cloudera enterprise documentation.

Hive on Spark

See below for the new features added to Hive on Spark in CDH 6.0:

- [Dynamic RDD Caching for Hive on Spark](#) on page 510
- [Optimized Hash Tables Enabled for Hive on Spark](#) on page 511

Dynamic RDD Caching for Hive on Spark

An optimization has been added to Hive on Spark that enables automatic caching of reused RDDs (Resilient Distributed Datasets). This optimization can improve query performance when the query or sub-query must scan a table multiple times. For example, [TPC-DS](#) query 39 is a query that requires multiple table scans. This optimization is disabled by default in CDH 6.0, but can be enabled by setting the `hive.combine.equivalent.work.optimization` property to `true` in the `hive-site.xml` file.



Important: While dynamic RDD caching can improve performance, using Spark's RDD cache may add additional memory pressure to Spark executors. This might increase the chance that a Spark executor runs out of memory and crashes.

To configure this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the Configuration tab.
3. Search for the **HiveServer2 Advanced Configuration Snippet (Safety Valve) for `hive-site.xml`**.
4. Enter the following property configuration information:

- **Name:** `hive.combine.equivalent.work.optimization`
- **Value:** `true`
- **Description:** Enables dynamic RDD caching for HoS

To disable this configuration, set the **Value** field to `false`.

To set this configuration property in the XML editor, enter the following code:

```
<property>
  <name>hive.combine.equivalent.work.optimization</name>
  <value>true</value>
  <description>Disables dynamic RDD caching for HoS</description>
</property>
```

5. Click **Save Changes**, and restart the service.

For more information see [HIVE-10844](#) and [HIVE-10550](#).

Optimized Hash Tables Enabled for Hive on Spark

Support has been added for optimized hash tables for Hive on Spark to reduce memory overhead. This feature is enabled by default in CDH 6.0, but can be disabled by setting the `hive.mapjoin.optimized.hashtable` property to `false` in the `hive-site.xml` file. To configure this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the Configuration tab.
3. Search for the **HiveServer2 Advanced Configuration Snippet (Safety Valve) for `hive-site.xml`**.
4. Enter the following property configuration information:

- **Name:** `hive.mapjoin.optimized.hashtable`
- **Value:** `false`
- **Description:** Disables optimized hash tables for HoS

To enable this configuration, set the **Value** field to `true`.

To set this configuration property in the XML editor, enter the following code:

```
<property>
  <name>hive.mapjoin.optimized.hashtable</name>
  <value>false</value>
  <description>Disables optimized hash tables for HoS</description>
</property>
```

5. Click **Save Changes**, and restart the service.

For more details, see [HIVE-11182](#) and [HIVE-6430](#).

Hue

There are no notable new features in this release.

Apache Impala

The following are some of the significant new features in this release of Impala.

List of Reserved Words Updated with SQL:2016 Reserved Words

The list of reserved words in Impala was updated in CDH 6.0 to be inline with the SQL:2016 standard.

See [Impala Reserved Words](#) for the updated list of reserved words.

Decimal V2 Used by Default

In Impala, two different behaviors of `DECIMAL` types are supported. In CDH 6.0, the V2 behavior of `DECIMAL` is used by default. See [DECIMAL Type](#) for detail information.

Behavior of Column Aliases Changed

To conform to the SQL standard, Impala no longer performs alias substitution in the subexpressions of `GROUP BY`, `HAVING`, and `ORDER BY`. See [for detail information](#).

Default PARQUET_ARRAY_RESOLUTION Changed

The default value of the `PARQUET_ARRAY_RESOLUTION` was changed to `THREE_LEVEL` in CDH 6.0.

The `PARQUET_ARRAY_RESOLUTION` setting controls the path-resolution behavior for Parquet files with nested arrays. See [PARQUET_ARRAY_RESOLUTION Query Option](#) for the information about the option.

Clustering Hint for Inserts Enabled by Default

In CDH 6.0, the `CLUSTERED` hint is enabled by default. The hint inserts a local sort by the partitioning columns to a query plan.

The `clustered` hint is only effective for HDFS and Kudu tables.

Apache Kafka

The following are some of the notable new features in this release of Kafka CDH 6.0.0.

Kafka Bundled with CDH

As of CDH 6.0.0, Apache Kafka is now included in the CDH bundle. You can choose to install Kafka from the list of custom services or you can add Kafka later to an existing cluster.

The Kafka version in CDH 6.0.0 is based on Apache Kafka 1.0.1.

Cloudera documentation for Apache Kafka in CDH 6.0.0 is part of the CDH documentation; see the [Kafka Guide](#).

New Metrics

Base Metrics

- Global Topic Count
- In-Sync Replica Counts
 - Under Min ISR Partition Count
 - Under Min ISR

Broker Metrics

- Health Checks: Error rates
 - Message conversion rate and time
 - Request size and temporary memory size
 - Authentication success and failure rates
 - ZooKeeper status and latency

Security Improvements

This release includes improved diagnostics for SASL and SSL authentication failures as described in [KIP-152](#).

This release includes an upgrade to PrincipalBuilder interface that improves SSL authentication and extends the functionality to SASL authentication. See [KAFKA-5783](#) for details.

Log Context added to logger messages

This release includes using a `LogContext` object that automatically adds a log prefix to every message written by loggers constructed from it.

- The consumer log messages now contain the consumer group and client IDs, which is very helpful when multiple consumers are run on the same instance.
- The producer log messages are automatically prefixed with client ID and transactional ID.
- The `NetworkClient` log messages now include client IDs.

Default Behavior Changes

- [KAFKA-3356](#) - Removed `ConsumerOffsetChecker`, which was deprecated in Apache Kafka version 0.9.
- [KAFKA-5384](#) - KIP-162: Enabled topic deletion by default. This release changes the default of `delete.topic.enable` to true.

Apache Kudu

The following are some of the notable new features in this release of Kudu 1.6 / CDH 6.0.

- Cloudera Manager now supports installing Kudu as part of the Express Wizard installation.
- Cloudera Manager has improved monitoring of Kudu node health. Cloudera Manager will indicate whether the health of the service is concerning or bad according to the number of master or tablet server roles that are not running.
- The disk failure tolerance feature is enabled by default on tablet servers, and the capability was extended to handle data directory failures at runtime. In the event of a runtime disk failure, any tablets with data on a failed

disk will be shut down and restarted on another tablet server. By default, tablets are striped across all available disks. Note that the first configured data directory and the WAL directory cannot currently tolerate disk failures. This will be further improved in future Kudu releases.

- Kudu servers can now adopt new data directories via the `kudu fs update_dirs` tool. The new directory will be used by new tablet replicas only. Removing directories is not yet supported ([KUDU-2202](#)).
- Kudu servers have two new flags to control the webui TLS/HTTPS settings: `--webserver_tls_ciphers` and `--webserver_tls_min_protocol`. These flags allow the advertised TLS ciphers and TLS protocol versions to be configured. Additionally, the webserver now excludes insecure legacy ciphers by default ([KUDU-2190](#)).
- The strategy Kudu uses for automatically healing tablets which have lost a replica has been improved. When a tablet loses a replica due to server or disk failures, the new re-replication strategy, or replica management scheme, first adds a replacement tablet replica before evicting the failed one. With the previous replica management scheme, the system first evicts the failed replica and then adds a replacement. The new replica management scheme allows for much faster recovery of tablets in scenarios where one tablet server goes down and then returns back shortly after 5 minutes or so. The new scheme also provides substantially better overall stability on clusters with frequent server failures ([KUDU-1097](#)).

The following are the notable optimizations improvements made in Kudu.

- Kudu servers can now tolerate short interruptions in NTP clock synchronization. NTP synchronization is still required when any Kudu daemon starts up. If NTP synchronization is not available, diagnostic information is now logged to help pinpoint the issue ([KUDU-1578](#)).
- Tablet server startup time has been improved significantly on servers containing large numbers of blocks.
- The log block manager now performs disk data deletion in batches. This optimization can significantly reduce the time taken to delete data on a tablet.
- The usage of sensitive data redaction flag has been changed. By setting the `--redact=log` flag, redaction will be disabled in the web UI but retained for server logs. Alternatively, `--redact=none` can be used to disable redaction completely.
- The Spark DataSource integration now can take advantage of scan locality for better scan performance. The scan will take place at the closest replica instead of going to the leader.
- Various optimizations were made to reduce the 99th percentile latency of writes on tablet servers. This can also improve throughput on certain write workloads, particularly on larger clusters.
- You can configure Kudu to ignore the system-wide auth_to_local mappings configured in /etc/krb5.conf by setting the configuration flag `--use_system_auth_to_local=false` ([KUDU-2198](#)).
- The performance of the compaction scheduler has been improved. In previous versions, certain types of time series workloads could cause compaction scheduling to take tens of seconds. These workloads now schedule compactations an order of magnitude more efficiently.
- The compaction scheduler has been improved to avoid running a compaction when the benefit of that compaction is extremely small.
- Tablet servers now consider the health of all replicas of a tablet before deciding to evict one. This can improve stability of the Kudu cluster after experiencing multiple simultaneous daemon failures ([KUDU-2048](#)).
- Several performance improvements have been made to the Kudu master, particularly in concurrency of clients opening tables. This should improve performance in highly concurrent workloads.
- The on-disk size metric for a tablet now includes all data and metadata. Previously, it excluded WAL segments and consensus metadata ([KUDU-1755](#)).
- Added a verbose mode for the Kudu cluster `ksck` command to enable output of detailed information on the cluster's metadata, even when no errors are detected.

Apache Oozie

The following new features and improvements are available in this release of Oozie:

- Major revamp of the Hadoop components: Hadoop 3, HBase 2, Hive 2, Spark 2
- Launch actions are now available via YARN (instead of MapReduce)
- Switch from Tomcat 6 (end-of-life) to Jetty 9
- Local ShareLib improvements:
 - New option to use local paths in the ShareLib instead of HDFS paths
 - Usability improvement: users don't have to upload .jar files to HDFS
 - After first-time cluster deployment, high volume of HDFS roundtrips are reduced on container resource localization
 - Improved version management
 - Automatic updates .jar files on cluster machines
- Supportability improvements:
 - Transient database outage resiliency
 - Spark action wrap-up:
 - Extended parameter parsing
 - Spark 2 support
 - Database migration tool can cope with large databases
 - Integration test framework improvements
 - Diagnostic bundle collector
 - Coordinator improvements

Apache Parquet

The new features and improvements in Apache Parquet versions 1.6.0 through 1.9.0 are included in the CDH 6.0 release. For the complete list of new features and enhancements that are available upstream in Parquet, see [Apache Parquet Release Notes](#).

In addition to the contents of the upstream Parquet 1.9.0 release, the Parquet in CDH 6.0 includes new features and bug fixes that were added to the upstream Parquet after the latest release. The following are the noteworthy new features added on top of the Parquet 1.9.0.

PARQUET-386: parquet-tools Prints Column Metadata Statistics

When you run `parquet-tools` with `schema` as an argument, the tool will print the statistics data of columns.

See [Using the Parquet File Format with Impala Tables](#) for using `parquet-tools` to examine the structure and data of Parquet files.

PARQUET-1025: New min-max Statistics in parquet-mr

The new min-max statistics is supported in `parquet-mr`. The old min and max statistics were only applicable to integer and floating point types. Additionally, the new `min_value` and `max_value` fields allow efficient filtering of decimal and string values as well. You can read more about these new statistics fields in [Cloudera Engineering blog post](#).

PARQUET-1115: Warning Message in parquet-tools merge

To prevent users from using the `parquet-tools merge` command inappropriately, a warning was added to the help text of the `parquet-tools` command and to the output of the `parquet-tools merge` command if the size of the original row groups are below a threshold.

Many users are tempted to use the new `parquet-tools merge` functionality, because they want to achieve good performance and historically that has been associated with large Parquet files. However, in practice Hive performance won't change significantly after using `parquet-tools merge`, but Impala performance will be much worse. The reason for that is that good performance is not a result of large files but large rowgroups instead (up to the HDFS block size).

However, `parquet-tools merge` does not merge rowgroups, it just places them one after the other. It was intended to be used for Parquet files that are already arranged in row groups of the desired size. When used to merge many small files, the resulting file will still contain small row groups and one loses most of the advantages of larger files (the only one that remains is that it takes a single HDFS operation to read them).

Additional Features and Enhancements

The additional new features and enhancements added to CDH 6.0 are:

- [PARQUET-321](#): Default maximum block padding to 8MB.
- [PARQUET-801](#): Allow UserDefinedPredicates in DictionaryFilter.
- [PARQUET-822](#): Upgrade the Java dependencies.
- [PARQUET-1026](#): Allow unsigned binary stats when the min value equals the max value.
- [PARQUET-1170](#): String values are represented based on logical type and represented properly in tools/logs.

Apache Pig

- Hive UDF support ([PIG-3294](#))
- Auto Local Mode ([PIG-3463](#))

Cloudera Search



Important: Some features included in Apache Solr 7 are not supported in Cloudera Search in CDH 6.0. For more information, see [Cloudera Search Unsupported Features](#) on page 527.

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has added many new features since the 4.10 version of Apache Solr used in recent CDH 5 releases.

For information on the new features added in Solr 5, Solr 6, and Solr 7, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Some new feature highlights include:

Solr JSON Request API

You can now create search queries in JSON format, which is much easier to read and write than the previous Solr query syntax.

More information:

- Apache Solr Reference Guide: [JSON Request API](#)
- Yonik Seeley's blog post: [Solr JSON Request API](#)

JSON Facet API

You can now write facet queries in JSON format. Facet structure is inherently nested, and is much easier to navigate in JSON format.

More information:

- [SOLR-7214](#)
- Yonik Seeley's blog post: [JSON Facet API](#)

Nested Faceting

You can now switch the facet domain between parents and children for the purposes of calculating facets.

More information:

- [SOLR-7676](#)
- Yonik Seeley's blog post: [Nested Objects in Solr](#)

Cloudera Enterprise 6 Release Guide

Per-Collection `clusterstate.json`

In CDH 5, Solr stores the states of all cores and collections in a single file in ZooKeeper, named `clusterstate.json`. This creates a single point of failure and a point of contention in large clusters with many actively changing collections. In CDH 6, Solr stores the state of cores in a single `clusterstate.json` file for each collection.

More information:

- [SOLR-5473](#)

Improved Integration with Apache Sentry

Cloudera Search in CDH 6 supports pluggable authorization modules, including Apache Sentry. The Sentry privilege model for Cloudera Search has been improved to support more granular control over privileges.

More information:

- [Authorization Privilege Model for Cloudera Search](#)
- Apache Solr Reference Guide: [Authentication and Authorization Plugins](#)
- [SENTRY-1475](#)

HyperLogLog Based Distributed Cardinality

This is a faster (but less accurate) way to calculate the number of distinct values on high cardinality fields. This implementation uses less memory than the previous one.

More information:

- [SOLR-6968](#)
- [SOLR-7553](#)
- Yonik Seeley's blog post: [Count Distinct in Solr](#)

Improved Support for DocValues

More information:

- Apache Solr Reference Guide: [DocValues](#)

Distributed IDF

In CDH 5, Solr calculates TF-IDF at the shard level. This results in misleading relevancy when TF-IDF varies a lot among shards. In CDH 6, Solr can use global TF-IDF statistics.

More information:

- Apache Solr Reference Guide:
 - [Distributed Requests](#)
 - [Distributed Search with Index Sharding](#)
- [SOLR-1632](#)

Jetty 9

In CDH 6, Solr runs inside a Jetty container instead of Tomcat.

Apache Sentry

There are no notable new features in this release.

Apache Spark

The following list describes what's new and changed in Apache Spark for CDH 6. Because CDH 6 includes features from Spark 2.x, this list includes the features across several releases of the CDS Powered By Apache Spark parcel, which was available separately for CDH 5.

- More flexibility to interpret `TIMESTAMP` values written by Impala. Setting the `spark.sql.parquet.int96TimestampConversion` configuration setting to `true` makes Spark interpret `TIMESTAMP` values, when reading from Parquet files written by Impala, without applying any adjustment from

the UTC to the local time zone of the server. This behavior provides better interoperability for Parquet data written by Impala, which does not apply any time zone adjustment to `TIMESTAMP` values when reading or writing them.

- The Spark “blacklisting” feature is now turned on by default. This feature takes hosts out of the pool of executors when they exhibit unreliable behavior, even if those hosts are not completely down. See the [Blacklisting in Apache Spark](#) blog post for details.
- Hive On Spark now runs on top of Apache Spark 2.x.
- Spark On HBase now runs on top of Apache Spark 2.x.
- The Oozie Spark action now runs on top of Apache Spark 2.x.
- Support for using Spark 2 jobs to read and write data on the Azure Data Lake Store (ADLS) cloud service.
- Version 2.2 or higher of CDS Powered By Apache Spark requires JDK 8.
- New direct connector to Kafka that uses the new Kafka consumer API. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for details.
- New `SparkSession` object replaces `HiveContext` and `SQLContext`.
 - Most of the Hive logic has been reimplemented in Spark.
 - Some Hive dependencies still exist:
 - SerDe support.
 - UDF support.
- Added support for the unified Dataset API.
- Faster Spark SQL achieved with whole stage code generation.
- More complete SQL syntax now supports subqueries.
- Adds the `spark-csv` library.
- Backport of SPARK-5847. The root for metrics is now the app name (`spark.app.name`) instead of the app ID. The app ID requires investigation to match to the app name, and changes when streaming jobs are stopped and restarted.

Apache Sqoop

The following are the new features in this release of Sqoop. All the new features available upstream in Sqoop 1.4.7 have been included in the CDH 6.0 release. For more information on Sqoop 1.4.7, see [Release Notes for Sqoop 1.4.7](#).

Sqoop connector parcel locations

Starting in CDH 6.0.0, Sqoop Connector parcels can be found in two separate parcel repositories:
<https://archive.cloudera.com/sqoop-teradata-connector1/latest/> and
<https://archive.cloudera.com/sqoop-netezza-connector1/latest/>.

SQOOP-816: Add support for external Hive tables

You can specify the importing of data into an external table in Hive (instead of a managed table, which is the default behavior). In order to do this, use the `--external-table-dir` option to specify the path. Include the `--hive-import` flag in the command line arguments if you use the import tool.

Example commands:

```
# Importing from Oracle RDBMS server into external HIVE table:
sqoop import --hive-import --connect $CONN --table $TABLENAME --username $USER --password
$PASS --external-table-dir /tmp/external_table_example

# Creating a hive table with a different name than in the database:
sqoop create-hive-table --connect $CONN --table $TABLENAME --username $USER --password
$PASS --external-table-dir /tmp/foobar_example --hive-table foobar
```

SQOOP-1904: Add support for DB2's XML data type

Added support for DB2's XML data type when importing to HDFS. Avro and Parquet files are not supported. The `--as-textfile` and `--as-sequencefile` options work as expected.

Example commands:

```
# the import command works as expected:  
sqoop import --connect $CONN --table $TABLENAME --username $USER --password $PASS  
  
# export back to table in the database:  
sqoop export --connect $CONN --table TEST_FOOBAR --username $USER --password $PASS  
--export-dir path/to/imported/data
```

SQOOP-1905: Add `--schema` option for import-all-tables and list-tables for DB2 connector

Sqoop now supports the `--schema` tool option for DB2. If the option is not present, then the schema of the current user will be used as default.



Note: This option does not work for other tools, such as the `import` tool or the `export` tool.

Example commands:

```
sqoop list-tables --connect $CONN --username $USER --password $PASS -- --schema DB2INST2  
  
sqoop import-all-tables --connect $CONN --username $USER --password $PASS -- --schema  
DB2INST2
```

SQOOP-2936: Provide Apache Atlas integration for hcatalog-based exports

Sqoop already supported publishing information about its data operations via the `SqoopJobDataPublisher` class before this change. However, this class is now extended to allow Apache Atlas to identify data lineage when importing or exporting from or to Hive, by publishing the Hive database and table names along with the previously published data.

SQOOP-2976: Flag to expand decimal values to fit Avro schema

The `sqoop.avro.decimal_padding.enable` flag was added to Sqoop to allow padding decimal values when importing data into Avro files. This flag must be used together with `sqoop.avro.logical_types.decimal.enable` set to `true`. The padding enables the user to import decimal and number types (with a *declared precision and scale*) into Avro files when using databases that do not store the decimal values padded, such as the Oracle RDBMS.

Example command:

```
sqoop import -Dsqoop.avro.decimal_padding.enable=true  
-Dsqoop.avro.logical_types.decimal.enable=true --connect $CON --username $USER --password  
$PASS --query "select * from table_name where \$CONDITIONS" --target-dir  
hdfs://nameservice1//etl/target_path --as-avrodatafile --verbose -m 1
```

SQOOP-3178: Incremental Merging for Parquet File Format

Sqoop now supports the incremental merging of parquet files. The merge tool combines two datasets where entries in one dataset should overwrite entries of an older dataset.

```
# As with avro data file, the merge tool requires two already existing parquet imports:  
sqoop import --connect $CONN --table $TABLENAME --username $USER --password $PASS  
--target-dir incr_merge_test_1 --as-parquetfile
```

```
sqoop import --connect $CONN --table $TABLENAME --username $USER --password $PASS
--target-dir incr_merge_test_2 --as-parquetfile
```

```
# the merge command itself looks like this:
sqoop merge --new-data incr_merge_test_2 --onto incr_merge_test_1 --target-dir merged
--jar-file datatypes.jar --class-name Foo --merge-key id
```

The feature also works with incremental imports (`--incremental lastmodified`).

SQOOP-3216: Expanded Metastore support for MySql, Oracle, Postgresql, Microsoft SQL Server, and DB2

In addition to HSQLDB, Sqoop can now store metastore metadata in MySql, Oracle, Postgresql, Microsoft SQL Server, and DB2.

The JDBC connection string of the database storing the metastore has to be configured with the `--meta-connect` option or using the `sqoop.metastore.client.autoconnect.url` property in `sqoop-site.xml`.

The following table includes the supported connection string formats:

Service	Connect String Format
MySQL	<code>jdbc:mysql://<server>:<port>/<dbname></code>
HSQLDB	<code>jdbc:hsqldb:hsq1://<server>:<port>/<dbname></code>
PostgreSQL	<code>jdbc:postgresql://<server>:<port>/<dbname></code>
Oracle	<code>jdbc:oracle:thin:@//<server>:<port>/<SID></code>
DB2	<code>jdbc:db2://<server>:<port>/<dbname></code>
Microsoft SQL Server	<code>jdbc:sqlserver://<server>:<port>;database=<dbname></code>

If the `--meta-connect` option is present, then Sqoop tries to connect to the metastore database specified in this parameter value. Sqoop uses the username and password specified in `--meta-username` and `--meta-password` parameters. If they are not present, Sqoop uses empty username and password values. If the database in the connection string is not supported, Sqoop throws an exception.

If the `--meta-connect` parameter is not present, Sqoop checks if `sqoop.metastore.client.enable.autoconnect` configuration parameter is set to false. The default value is true. If the parameter is false, Sqoop throws an error, since there are no applicable metastore implementations.

Next, Sqoop checks the `sqoop.metastore.client.autoconnect.url` configuration parameter. In this parameter, users can specify a JDBC connection string to a metastore database. Supported databases are HSQLDB, MySQL, Oracle, PostgreSQL, DB2 and SQL Server. If this parameter is not present, Sqoop creates an embedded HSQLDB metastore in the home directory of the OS user that executes the `sqoop` command. The username and password can be specified in `sqoop.metastore.client.autoconnect.username` (default value: SA) and `sqoop.metastore.client.autoconnect.password` (default value is the empty string). If the connection string in `sqoop.metastore.client.autoconnect.url` is not supported, Sqoop throws an exception.



Note: Sqoop needs the corresponding JDBC drivers to be installed to be able to connect to the metastore (except when using HSQLDB metastore). The drivers for the metastore have to be installed the same way drivers are installed for connectors.

As in previous versions, the metastore tool can only start a shared HSQLDB instance. To use a metastore backed by a database other than HSQLDB, start and configure it manually and provide the correct JDBC connection string to Sqoop. Upon the first connection, Sqoop creates the necessary metadata tables, so the database user specified in the `--meta-username` option has to have CREATE TABLE permission.

Example commands

- Listing available jobs in the metastore:

```
sqoop job --list --meta-connect jdbc:oracle:thin:@//myhost:1521/ORCLCDB --meta-username ms_user --meta-password ms_password
```

- Creating a new job in the metastore:

```
sqoop job --create myjob1 --meta-connect jdbc:oracle:thin:@//myhost:1521/ORCLCDB  
--meta-username ms_user --meta-password ms_password -- import --connect  
jdbc:mysql://mysqlhost:3306/sqoop --username sqoop --password sqoop --table "TestTable"  
-m 1
```

- Executing an existing job:

```
sqoop job --exec myjob1 --meta-connect jdbc:oracle:thin:@//myhost:1521/ORCLCDB  
--meta-username ms_user --meta-password ms_password
```

- Showing the definition of an existing job:

```
sqoop job --show myjob2 --meta-connect jdbc:oracle:thin:@//myhost:1521/ORCLCDB  
--meta-username ms_user --meta-password ms_password
```

- Deleting an existing job:

```
sqoop job --delete myjob1 --meta-connect jdbc:oracle:thin:@//myhost:1521/ORCLCDB  
--meta-username ms_user --meta-password ms_password
```

Apache Zookeeper

There are no notable new features in this release.

Fixed Issues in CDH 6.0.0

See below for issues fixed in CDH 6.0.0, grouped by component:

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Apache Avro

There are no notable fixed issues in this release.

Apache Crunch

There are no notable fixed issues in this release.

Apache Flume

There are no notable fixed issues in this release.

Apache Hadoop

HDFS

- [HADOOP-12267](#) - DistCp to S3a fails due to integer overflow in retry timer.

MapReduce 2 and YARN

- [YARN-4212](#) (CDH-31358) - Jobs in pool with DRF policy will not run if root pool is FAIR.
- [MAPREDUCE-6638](#) (CDH-37412) - Jobs with encrypted spills do not recover if the Application Master goes down.
- [YARN-1558](#) - Moving jobs between queues not persistent after restart.

Apache HBase

In CDH 6.0.0, the default values for properties that are required to enable cell-level ACLs have changed. Previously, you needed to modify the properties to enable cell-level ACLs. In CDH 6.0.0, you do not need to modify them. The properties and their new default values are listed below:

```
hbase.security.exec.permission.checks => true
hbase.security.access.early_out => false
hfile.format.version => 3
```

For information about the upstream fixes, see the [Apache HBase JIRAs](#).

The following JIRA has also been closed:

- [HBASE-7621](#) - RemoteHTable now supports binary row keys with any character or byte by properly encoding request URLs.

Apache Hive / HCatalog / Hive on Spark

In CDH 6.0, Hive fixed issues resulted in new features and incompatible changes. For details on these fixed issues, see the following sections of the CDH 6.0 Release Notes for Hive:

- [New Features in Hive CDH 6.0](#)
- [Incompatible Changes in Hive CDH 6.0](#)

Hue

There are no notable fixed issues in this release.

Apache Impala

There are no notable fixed issues in this release

Apache Kafka

Authenticated Kafka clients may impersonate other users

Authenticated Kafka clients may impersonate any other user via a manually crafted protocol message with SASL/PLAIN or SASL/SCRAM authentication when using the built-in PLAIN or SCRAM server implementations in Apache Kafka.

Note that the SASL authentication mechanisms that apply to this issue are neither recommended nor supported by Cloudera. In Cloudera Manager (CM) there are [four choices](#): PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL. The SASL/PLAIN option described in this issue is not the same as SASL_PLAINTEXT option in CM. That option uses Kerberos and is not affected. As a result it is highly unlikely that Kafka is susceptible to this issue when managed by CM unless the authentication protocol is overridden by an Advanced Configuration Snippet (Safety Valve).

Products affected: CDK Powered by Apache Kafka

Releases affected: CDK 2.1.0 to 2.2.0, CDK 3.0.0

Users affected: All users

Detected by: Rajini Sivaram (rsivaram@apache.org)

Severity (Low/Medium/High):8.3 (High) ([CVSS](#):3.0/AV:N/AC:L/PR:L/UI:N/S:U/C:L/I:H/A:H)

Impact:Privilege escalation.

CVE: CVE-2017-12610

Immediate action required: Upgrade to a newer version of CDK Powered by Apache Kafka where the issue has been fixed.

Addressed in release/refresh/patch: CDK 3.1, CDH 6.0 and higher

Knowledge article: For the latest update on this issue see the corresponding Knowledge article: [TSB 2018-332: Two Kafka Security Vulnerabilities: Authenticated Kafka clients may impersonate other users and and may interfere with data replication](#)

Authenticated clients may interfere with data replication

Cloudera Enterprise 6 Release Guide

Authenticated Kafka users may perform an action reserved for the Broker via a manually created fetch request interfering with data replication, resulting in data loss.

Products affected: CDK Powered by Apache Kafka

Releases affected: CDK 2.0.0 to 2.2.0, CDK 3.0.0

Users affected: All users

Detected by: Rajini Sivaram (rsivaram@apache.org)

Severity (Low/Medium/High): 6.3 (Medium) ([CVSS:3.0/AV:N/AC:L/PR:L/UI:N/S:U/C:L/I:L/A:L](#))

Impact: Potential data loss due to improper replication.

CVE: CVE-2018-1288

Immediate action required: Upgrade to a newer version of CDK Powered by Apache Kafka where the issue has been fixed.

Addressed in release/refresh/patch: CDK 3.1, CDH 6.0 and higher

Knowledge article: For the latest update on this issue see the corresponding Knowledge article: [TSB 2018-332: Two Kafka Security Vulnerabilities: Authenticated Kafka clients may impersonate other users and and may interfere with data replication](#)

Upstream Issues Fixed

Metrics

- [KAFKA-6252](#) - A metric named 'XX' already exists, can't register another one.
- [KAFKA-5987](#) - Kafka metrics templates (`MetricNameTemplate` and the `Metric.toHtmlTable`) used in document generation should maintain order of tags.
- [KAFKA-5968](#) - Remove all broker metrics during shutdown.
- [KAFKA-5746](#) - New metrics to support health checks, including:

Broker-side metrics

- Error rates
- Message conversion rate and time
- Request size and temporary memory size
- Authentication success and failure rates
- ZooKeeper status and latency

Client-side metrics

- Client versions exposed as a metric
- [KAFKA-5738](#) - Add cumulative count attribute for all Kafka rate metrics to make Kafka metrics more compatible with other metrics such as Yammer.
 - [KAFKA-5597](#) - Auto-generate Producer sender metrics.
 - [KAFKA-5461](#) - KIP-168: New metric ("GlobalTopicCount") track the total topic count per cluster.
 - [KAFKA-5341](#) - New metrics ("UnderMinIsrPartitionCount" and per-partition "UnderMinIsr") track the number of partitions whose in-sync replicas count is less than the minimum configured for in-sync replicas (`min.insync.replicas`).

Security Supportability

- [KAFKA-6258](#) - SSLTransportLayer should keep reading from the socket until either the buffer is full or the socket has no more data.
- [KAFKA-5920](#) - Handle SSL authentication failures as non-retrievable exceptions in clients marked CONNECTED and DISCONNECTED at the same time.
- [KAFKA-5854](#) - KIP-152: Handle SASL authentication failures as non-retrievable exceptions in clients.
- [KAFKA-5783](#) - KIP-189: Implement KafkaPrincipalBuilder interface with support for SASL.

- [KAFKA-5720](#) - In Jenkins, kafka.api.SaslSslAdminClientIntegrationTest failed with org.apache.kafka.common.errors.TimeoutException.
- [KAFKA-5417](#) - Clients get inconsistent connection states when SASL/SSL connection is marked CONNECTED and DISCONNECTED at the same time.
- [KAFKA-4764](#) - KIP-152: Improve diagnostics for SASL authentication failures.

Kafka Client

- [KAFKA-6287](#) - Inconsistent protocol type for empty consumer groups.
- [KAFKA-5856](#) - KIP 195: Increase the number of partitions of a topic using AdminClient.createPartitions().
- [KAFKA-5763](#) - Refactor NetworkClient to use LogContext.
- [KAFKA-5762](#) - Refactor AdminClient to use LogContext.
- [KAFKA-5755](#) - Refactor KafkaProducer to use LogContext.
- [KAFKA-5737](#) - KafkaAdminClient thread should be a daemon.
- [KAFKA-5726](#) - The KafkaConsumer method subscribe overload that takes just *pattern* without ConsumerRebalanceListener.
- [KAFKA-5629](#) - ConsoleConsumer overrides auto.offset.reset property when provided on the command line without warning about it.
- [KAFKA-5556](#) - The KafkaConsumer method commitSync throws the exception IllegalStateException: Attempt to retrieve exception from future which hasn't failed
- [KAFKA-5534](#) - The KafkaConsumer method offsetsForTimes should include partitions in result even if no offset could be found.
- [KAFKA-5512](#) - KafkaConsumer: High memory allocation rate when idle.
- [KAFKA-4856](#) - Calling KafkaProducer.close() from multiple threads may cause spurious error.
- [KAFKA-4767](#) - KafkaProducer is not joining its IO thread properly.
- [KAFKA-4669](#) - KafkaProducer.flush hangs when NetworkClient.handleCompletedReceives throws an exception.
- [KAFKA-2105](#) - NullPointerException in client on metadataRequest.

Apache Kudu

There are no notable fixed issues in this release

Apache Oozie

There are no notable fixed issues in this release.

Apache Parquet

- [PARQUET-1217](#) - Incorrect handling of missing values in Statistics.
- [PARQUET-357](#) - Parquet-thrift generates wrong schema for Thrift binary fields.
- [PARQUET-686](#) - Clarifications about min-max stats.
- [PARQUET-753](#) - Fixed GroupType.union() to handle original type.
- [PARQUET-765](#) - Upgrade Avro to 1.8.1.
- [PARQUET-783](#) - Close the underlying stream when an H2SeekableInputStream is closed.
- [PARQUET-791](#) - Add missing column support for UserDefinedPredicate.
- [PARQUET-806](#) - Parquet-tools silently suppresses error messages.
- [PARQUET-825](#) - Static analyzer findings (NPEs, resource leaks).
- [PARQUET-1005](#) - Fix DumpCommand parsing to allow column projection.
- [PARQUET-1064](#) - Deprecate type-defined sort ordering for INTERVAL type.
- [PARQUET-1065](#) - Deprecate type-defined sort ordering for INT96 type.
- [PARQUET-1133](#) - Add int96 support by returning bytearray, Skip originalType comparison for map types when originalType is null.
- [PARQUET-1141](#) - Fix field ID handling.
- [PARQUET-1152](#) - Parquet-thrift doesn't compile with Thrift 0.9.3.
- [PARQUET-1153](#) - Parquet-thrift doesn't compile with Thrift 0.10.0.

Cloudera Enterprise 6 Release Guide

- [PARQUET-1185](#) - TestBinary#testBinary unit test fails after PARQUET-1141.
- [PARQUET-1191](#) - Type.hashCode() takes originalType into account but Type.equals() does not.
- [PARQUET-1208](#) - Occasional endless loop in unit test.
- [PARQUET-1217](#) - Incorrect handling of missing values in Statistics.
- [PARQUET-1246](#) - Parquet ignores float/double statistics in case of NaN.

Apache Pig

There are no notable fixed issues in this release of Apache Pig.

Cloudera Search

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has fixed many issues since the 4.10 version of Apache Solr used in recent CDH 5 releases.

For information on the fixes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

There are no notable fixed issues in this release

Apache Spark

CDH 6.0.0 uses CDS 2.2 Release 2 Powered By Apache Spark. The fixed issues listed in the [release notes for CDS 2.2 Release 2](#) have been incorporated in CDH 6.0.0.

Apache Sqoop

CDH 6.0 uses Sqoop 1.4.7. The fixed issues listed in the [release notes for Sqoop 1.4.7](#) have been incorporated in CDH 6.0. In addition, Cloudera has backported the following JIRAs into Sqoop in CDH 6.0.0, which are not yet released upstream:

- [SQOOP-3273](#): Removing com.cloudera.sqoop packages
- [SQOOP-3275](#): HBase test cases should start mini DFS cluster as well
- [SQOOP-3255](#): Sqoop ignores metastore properties defined in sqoop-site.xml
- [SQOOP-3241](#): ImportAllTablesTool uses the same SqoopOptions object for every table import
- [SQOOP-3153](#): Sqoop export with --as-<spec_file_format> error message could be more verbose
- [SQOOP-3266](#): Update 3rd party and manual test running related info in COMPILING.txt
- [SQOOP-3233](#): SqoopHCatImportHelper.convertNumberTypes check for Varchar instead of Char
- [SQOOP-3257](#): Sqoop must not log database passwords
- [SQOOP-3243](#): Importing BLOB data causes 'Stream closed' error on encrypted HDFS
- [SQOOP-3229](#): Document how to run third party tests manually with databases running in docker
- [SQOOP-3216](#): Expanded Metastore support for MySql, Oracle, Postgresql, MSSql, and DB2
- [SQOOP-3014](#): Sqoop with HCatalog import loose precision for large numbers that does not fit into double
- [SQOOP-3232](#): Remove Sqoop dependency on deprecated HBase APIs
- [SQOOP-3222](#): Test HBase kerberized connectivity
- [SQOOP-3195](#): SQLServerDatatypeImportDelimitedFileTest can fail in some environments
- [SQOOP-3196](#): Modify MySQLAuthTest to use configurable test database parameters
- [SQOOP-3139](#): sqoop tries to re execute select query during import in case of a connection reset error and this is causing lots of duplicate records from source
- [SQOOP-3218](#): Make sure the original ClassLoader is restored when running HCatalog tests
- [SQOOP-3178](#): Incremental Merging for Parquet File Format
- [SQOOP-3206](#): Make sqoop fail if user uses --direct connector and tries to encode a null value when using a MySQL database

Excluded JIRA

[SQOOP-3149](#) is part of Sqoop 1.4.7, but because it contains a serious bug, Cloudera has excluded it from the CDH 6.0 release.

Apache Zookeeper

There are no notable fixed issues in this release.

Unsupported Features in CDH 6.0.0

This page lists the unsupported features in CDH 6.0.0. For the complete list of Known Issues and Limitations, see [Known Issues and Limitations in CDH 6.0.0](#) on page 562.

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. If you try to upgrade to CDH 6.0.x, you will be asked to remove the CDSW service from your cluster.

Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Apache Hadoop Unsupported Features

The following sections list unsupported features in Hadoop common components:

- [HDFS Unsupported Features](#) on page 525
- [YARN Unsupported Features](#) on page 525

HDFS Unsupported Features

The following HDFS features are not supported in CDH 6.0.0:

- ACLs for the NFS gateway
- Aliyun Cloud Connector
- Erasure Coding
- HDFS NameNode Federation
- HDFS Router Based Federation
- HDFS truncate
- More than two NameNodes
- Openstack Swift
- Quota support for Storage Types
- SFTP FileSystem
- Upgrade Domain
- Variable length block
- ZStandard Compression Codec

YARN Unsupported Features

The following YARN features are not supported in CDH 6.0.0:

- Application Timeline Server v2 (ATSV2)
- Cgroup Memory Enforcement
- Container Resizing
- Distributed or Centralized Allocation of Opportunistic Containers
- Distributed Scheduling
- Docker on YARN (DockerContainerExecutor)
- Native Services
- New Aggregated Log File Format

- Node Labels
- Pluggable Scheduler Configuration
- Reservation REST APIs
- Resource Profiles
- Resource Estimator Service
- Rolling Log Aggregation
- (non-Zookeeper) ResourceManager State Store
- Shared Cache
- YARN Federation
- YARN WebUI v2

Apache HBase Unsupported Features

The following HBase features are not supported in CDH 6.0.0:

- Master hosting meta
- Cloudera does not provide support for user-provided custom coprocessors of any kind.
- Server-side encryption of HFiles. You should configure HDFS client-side encryption.
- In-memory compaction
- hbck read-write repair mode. See [Checking Consistency in HBase Tables](#) for more information.

The following features, introduced upstream in HBase, are not supported in CDH:

- Visibility labels
- Stripe compaction
- Clients setting priority on operations
- Specifying a custom asynchronous connection implementation

Apache Hive Unsupported Features

The following Hive features are not supported in CDH 6.0.0:

- AccumuloStorageHandler ([HIVE-7068](#))
- ACID ([HIVE-5317](#))
- Built-in `version()` function is not supported (CDH-40979)
- Cost-based Optimizer (CBO)
- Explicit Table Locking
- HCatalog - HBase plugin
- Hive Authorization (Instead, use Apache Sentry.)
- Hive on Apache Tez
- Hive Local Mode Execution
- Hive Metastore - Derby
- Hive Web Interface (HWI)
- HiveServer1 / JDBC 1
- HiveServer2 Dynamic Service Discovery (HS2 HA) ([HIVE-8376](#))
- HiveServer2 - HTTP Mode (Use THRIFT mode.)
- HPL/SQL ([HIVE-11055](#))
- LLAP (Live Long and Process framework)
- Scalable Dynamic Partitioning and Bucketing Optimization ([HIVE-6455](#))
- Session-level Temporary Tables ([HIVE-7090](#))
- Table Replication Across HCatalog Instances ([HIVE-7341](#))

Apache Kafka Unsupported Features

The following Kafka feature is not supported in CDH 6.0.0:

- Idempotent and transactional capabilities in the producer are currently an unsupported beta feature given their maturity and complexity. This feature will be supported in a future release.
- Kafka Connect is included in CDH 6.0.0, but is not supported. Flume and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. See [Flafka: Apache Flume Meets Apache Kafka for Event Processing](#) for more information.
- Kafka Streams is included in CDH 6.0.0, but is not supported. Instead, use Spark and Spark Streaming have a fully functional ETL/stream processing pipeline. See [Using Kafka with Apache Spark Streaming for Stream Processing](#) for more information.
- The Kafka default authorizer is included in CDH 6.0.0, but is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.
- Using Kafka with a JBOD setup is an unsupported beta feature given its maturity and complexity. Using JBOD in production will be supported only in a later release.
- The legacy Scala clients (producer and consumer) that are under the `kafka.producer.*` and `kafka.consumer.*` package are deprecated in CDH 6.0.0. See [Deprecated Scala-based Client API and New Java Client API](#) on page 544.

Apache Oozie Unsupported Features

The following Oozie feature is not supported in CDH 6.0.0:

- Conditional coordinator input logic.

Apache Pig Unsupported Features

The following Pig features are not supported in CDH 6.0.0:

- Pig on Tez is not supported in CDH 6.0.0 ([PIG-3446](#) / [PIG-3419](#)).
- Pig on Spark.

Cloudera Search Unsupported Features

The following Search features are not supported in CDH 6.0.0:

- [Solr SQL/JDBC](#)
- [Graph Traversal](#)
- [Cross Data Center Replication \(CDCR\)](#)
- [SolrCloud Autoscaling](#)
- HDFS Federation is not supported
- Saving search results is not supported

Apache Sentry Unsupported Features

The following Sentry features are not supported in CDH 6.0.0:

- Import and export of Sentry metadata to and from Sentry servers
- Sentry shell command line for Hive
- Relative URI paths ([Known Issue](#))
- Object types Server and URL in

```
show grant role <role name> on object <object name>
```

([Known Issue](#))

- ALTER and DROP privileges for Hive and Impala

In addition, as of CDH 6.0.0, Sentry policy files have been removed. See the Sentry [Incompatible Changes](#) for more information.

Apache Spark Unsupported Features

The following Spark features are not supported in CDH 6.0.0:

- Apache Spark experimental features/APIs are not supported unless stated otherwise.
- Using the JDBC Datasource API to access Hive or Impala is not supported

- ADLS not Supported for All Spark Components. Microsoft Azure Data Lake Store (ADLS) is a cloud-based filesystem that you can access through Spark applications. Spark with Kudu is not currently supported for ADLS data. (Hive on Spark is available for ADLS in CDH 5.12 and higher.)
- IPython / Jupyter notebooks is not supported. The IPython notebook system (renamed to Jupyter as of IPython 4.0) is not supported.
- Certain Spark Streaming features not supported. The `mapWithState` method is unsupported because it is a nascent unstable API.
- Thrift JDBC/ODBC server is not supported
- Spark SQL CLI is not supported
- GraphX is not supported
- SparkR is not supported
- Structured Streaming is not supported.
- Spark cost-based optimizer (CBO) not supported.

Apache Sqoop Unsupported Features

The following Sqoop feature is not supported in CDH 6.0.0:

- [import-mainframe](#)

Incompatible Changes in CDH 6.0.0



Important:

In addition to incompatible changes, CDH 6 also deprecated or removed support for several components, including Spark 1 and MapReduce v1. For information about components, sub-components, or functionality that are deprecated or no longer supported, see [Deprecated Items](#) on page 667.

See below for incompatible changes in CDH 6.0.0, grouped by component:

Apache Accumulo

Running Apache Accumulo on top of a CDH 6.0.0 cluster is not currently supported. If you try to upgrade to CDH 6.0.0 you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Apache Avro

API Changes

One method was removed in CDH 6.0.0:

```
GenericData.toString (Object datum, StringBuilder buffer)
```

Incompatible Changes from Avro 1.8.0

- Changes in logical types cause code generated in Avro with CDH 6 to differ from code generated in Avro with CDH 5. This means that old generated code will not necessarily work in CDH 6. Cloudera recommends that users regenerate their generated Avro code when upgrading.
- [AVRO-997](#): Generic API requires GenericEnumSymbol - likely to break current Generic API users that often have String or Java Enum for these fields
- [AVRO-1502](#): Avro Objects now Serializable - IPC needs to be regenerated/recompiled
- [AVRO-1602](#): removed Avro internal RPC tracing, presumed unused. Current rec would be HTrace
- [AVRO-1586](#): Compile against Hadoop 2 - probably not an issue since we've been compiling against Hadoop 2 for C5.
- [AVRO-1589](#): [Java] ReflectData.AllowNulls will create incompatible Schemas for primitive types - may need a KI since it used to fail at runtime but now will fail earlier.

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. If you try to upgrade to CDH 6.0.x, you will be asked to remove the CDSW service from your cluster.

Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Cloudera Issue: DSE-2769

Apache Crunch



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

The following changes are introduced in CDH 6.0.0, and are not backward compatible:

- Crunch is available only as Maven artifacts from the Cloudera Maven repository. It is not included as part of CDH. For more information, see [Apache Crunch Guide](#).
- Crunch supports only Spark 2 and higher releases.
- Crunch supports only HBase 2 and higher releases.
 - The API methods in Crunch-HBase use HBase 2 API types and methods.

Apache Flume

AsyncHBaseSink and HBaseSink

CDH 6 uses HBase 2.0. AsyncHBaseSink is incompatible with HBase 2.0 and is not supported in CDH 6. HBaseSink has been replaced with HBase2Sink. HBase2Sink works the same way as HBaseSink. The only difference is that it is compatible with HBase 2.0. No additional configuration is required when HBase2Sink is used, but you can replace the component type in your configuration.

For example, replace this text:

```
agent.sinks.my_hbase_sink.type = hbase
```

With this:

```
agent.sinks.my_hbase_sink.type = hbase2
```

Or, if you use the FQN of the sink class, replace this text:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase.HBaseSink
```

With this:

```
agent.sinks.my_hbase_sink.type = org.apache.flume.sink.hbase2.HBase2Sink
```

For more information about how to configure HBase2Sink, see [Importing Data Into HBase](#).

For more information about the use of legacy names, see [Serializer Class Names](#).

com.google.common.collect.ImmutableMap

Flume has removed `com.google.common.collect.ImmutableMap` from the `org.apache.flume.Context` API and replaced it with `java.util.Map` due to Guava compatibility issues ([FLUME-2957](#)). Plugins using the `Context.getParameters()` and `Context.getSubProperties()` APIs will need to assign the return value of those methods to a `Map<String, String>` variable instead of an `ImmutableMap<String, String>` variable, if they do not already do so. Most usages in the Flume codebase already used `Map<String, String>` at the time of this change.

Apache Hadoop

- [HDFS Incompatible Changes](#) on page 530

- [MapReduce](#) on page 531
- [YARN](#) on page 531

HDFS Incompatible Changes

- HFTP has been removed.
- The S3 and S3n connectors have been removed. Users should now use the S3a connector.
- The BookkeeperJournalManager has been removed.
- Changes were made to the structure of the HDFS JAR files to better isolate clients from Hadoop library dependencies. As a result, client applications that depend on Hadoop's library dependencies may no longer work. In these cases, the client applications will need to include the libraries as dependencies directly.
- Several library dependencies were upgraded. Clients that depend on those libraries may break because the library version changes. In these cases, the client applications will need to either be ported to the new library versions or include the libraries as dependencies directly.
- [HDFS-6962](#) changes the behavior of ACL inheritance to better align with POSIX ACL specifications, which states that the umask has no influence when a default ACL propagates from parent to child. Previously, HDFS ACLs applied the client's umask to the permissions when inheriting a default ACL defined on a parent directory. Now, HDFS can ignore the umask in these cases for improved compliance with POSIX. This behavior is on by default due to the inclusion of [HDFS-11957](#). It can be configured by `settingdfs.namenode posix.acl.inheritance.enabled` in `hdfs-site.xml`. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-11957](#) changes the default behavior of ACL inheritance introduced by [HDFS-6962](#). Previously, the behavior was disabled by default. Now, the feature is enabled by default. Any code expecting the old ACL inheritance behavior will have to be updated. See the Apache Hadoop HDFS Permissions Guide for more information.
- [HDFS-6252](#) removed `dfshealth.jsp` since it is part of the old NameNode web UI. By default, Cloudera Manager links to the new NameNode web UI, which has an equivalent health page at `dfshealth.html`.
- [HDFS-11100](#) changes the behavior of deleting files protected by a sticky bit. Now, the deletion fails.
- [HDFS-10689](#) changes the behavior of the `hdfs dfs chmod` command. Now, the command resets sticky bit permission on a file/directory when the leading sticky bit is omitted in the octal mode (like 644). When a file or directory permission is applied using octal mode and sticky bit permission needs to be preserved, then it has to be explicitly mentioned in the permission bits (like 1644).
- [HDFS-10650](#) changes the behavior of `DFSClient#mkdirs` and `DFSClient#primitiveMkdir`. Previously, they create a new directory with the default permissions 00666. Now, they will create a new directory with permission 00777.
- [HADOOP-8143](#) changes the default behavior of `distcp`. Previously, the `-pb` option was not used by default, which may have caused some checksums to fail when block sizes did not match. Now, the `-pb` option is included by default to preserve block size when using `distcp`.
- [HADOOP-10950](#) changes several heap management variables:
 - `HADOOP_HEAPSIZE` variable has been deprecated. Use `HADOOP_HEAPSIZE_MAX` and `HADOOP_HEAPSIZE_MIN` instead to set `Xmx` and `Xms`
 - The internal variable `JAVA_HEAP_MAX` has been removed.
 - Default heap sizes have been removed. This will allow for the JVM to use auto-tuning based upon the memory size of the host. To re-enable the old default, configure `HADOOP_HEAPSIZE_MAX="1g"` in `hadoop-env.sh`.
 - All global and daemon-specific heap size variables now support units. If the variable is only a number, the size is assumed to be in megabytes.
- [HADOOP-14426](#) upgrades the version of Kerby from 1.0.0-RC2 to 1.0.0
- [HDFS-10970](#) updates the version of Jackson from 1.9.13 to 2.x in `hadoop-hdfs`.
- [HADOOP-9613](#) updates the Jersey version to the latest 1.x release.
- [HADOOP-10101](#) updates Guava dependency to 21.0
- [HADOOP-14225](#) removes the `xmlenc` dependency. If you rely on the transitive dependency, you need to set the dependency explicitly in your code after this change.
- [HADOOP-13382](#) remove unneeded `commons-httpclient` dependencies from POM files in Hadoop and sub-projects. This incompatible change may affect projects that have undeclared transitive dependencies on `commons-httpclient`, which used to be provided by `hadoop-common` or `hadoop-client`.

- [HADOOP-13660](#) upgrades the commons-configuration version from 1.6 to 2.1.
- [HADOOP-12064](#) upgrades the following dependencies:
 - Guice from 3.0 to 4.0
 - cglib from 2.2 to 3.2.0
 - asm from 3.2 to 5.0.4

MapReduce

- Support for MapReduce v1 has been dropped from CDH 6.0.0.
- CDH 6 supports applications compiled against CDH 5.7.0 and higher MapReduce frameworks. Make sure to not to include the CDH jars with your application by marking them as "provided" in the `pom.xml` file.

YARN

There are no incompatible changes in this release.

Apache HBase

CDH 6.0.0 contains the following downstream HBase incompatible change:

hbase.security.authorization

The default value for `hbase.security.authorization` has been changed from true to false. Secured clusters should make sure to explicitly set it to true in XML configuration file before upgrading to one of these versions ([HBASE-19483](#)). True as the default value of `hbase.security.authorization` was changed because not all clusters need authorization. (History: [HBASE-13275](#)) Rather, only the clusters which need authorization should set this configuration as true.

Incompatible Changes

For more information about upstream incompatible changes, see the Apache Reference Guide [Incompatible Changes](#) and [Upgrade Paths](#).

CDH 6.0.0 contains the following upstream HBase incompatible changes:

- Public interface API changes:
 - [HBASE-15607](#): Admin
 - [HBASE-19112](#), [HBASE-18945](#): Cell
 - Region, Store, HBaseTestingUtility
- [HBASE-18792](#): hbase-2 needs to defend against hbck operations
- [HBASE-15982](#): Interface ReplicationEndpoint extends Guava's Service.
- [HBASE-18995](#): Split CellUtil into public CellUtil and PrivateCellUtil for Internal use only.
- [HBASE-19179](#): Purged the hbase-prefix-tree module and all references from the code base.
- [HBASE-17595](#): Add partial result support for small/limited scan; Now small scan and limited scan could also return partial results.
- [HBASE-16765](#): New default split policy, SteppingSplitPolicy.
- [HBASE-17442](#): Move most of the replication related classes from hbase-client to hbase-replication package.
- [HBASE-16196](#): The bundled JRuby 1.6.8 has been updated to version 9.1.9.0.
- [HBASE-18811](#): Filters have been moved from Public to LimitedPrivate.
- [HBASE-18697](#): Replaced hbase-shaded-server jar with hbase-shaded-mapreduce jar.
- [HBASE-18640](#): Moved mapreduce related classes out of hbase-server into separate hbase-mapreduce jar .
- [HBASE-19128](#): Distributed Log Replay feature has been removed.
- [HBASE-19176](#): Hbase-native-client has been removed.
- [HBASE-17472](#): Changed semantics of granting new permissions. Earlier, new grants would override previous permissions, but now, the new and existing permissions get merged.
- [HBASE-18374](#): Previous "mutate" latency metrics has been renamed to "put" metrics.
- [HBASE-15740](#): Removed Replication metric source.shippedKBs in favor of source.shippedBytes.
- [HBASE-13849](#): Removed restore and clone snapshot from the WebUI.

- [HBASE-13252](#): The concept of managed connections in HBase (deprecated before) has now been extinguished completely, and now all callers are responsible for managing the lifecycle of connections they acquire.
- [HBASE-14045](#): Bumped thrift version to 0.9.2.
- [HBASE-5401](#): Changes to number of tasks PE runs when clients are mapreduce. Now tasks == client count. Previous we hardcoded ten tasks per client instance.

Changed Behavior

CDH 6.0.0 contains the following HBase behavior changes:

- [HBASE-14350](#): Assignment Manager v2 - Split/Merge have moved to the Master; it runs them now. Hooks around Split/Merge are now noops. To intercept Split/Merge phases, CPs need to intercept on MasterObserver.
- [HBASE-18271](#): Moved to internal shaded netty.
- [HBASE-17343](#): Default MemStore to be CompactingMemStore instead of DefaultMemStore. In-memory compaction of CompactingMemStore demonstrated sizable improvement in HBase's write amplification and read/write performance.
- [HBASE-19092](#): Make Tag IA.LimitedPrivate and expose for CPs.
- [HBASE-18137](#): Replication gets stuck for empty WALS.
- [HBASE-17513](#): Thrift Server 1 uses different QOP settings than RPC and Thrift Server 2 and can easily be misconfigured so there is no encryption when the operator expects it.
- [HBASE-16868](#): Add a replicate_all flag to replication peer config. The default value is true, which means all user tables (REPLICATION_SCOPE != 0) will be replicated to peer cluster.
- [HBASE-19341](#): Ensure Coprocessors can abort a server.
- [HBASE-18469](#): Correct RegionServer metric of totalRequestCount.
- [HBASE-17125](#): Marked Scan and Get's setMaxVersions() and setMaxVersions(int) as deprecated. They are easy to misunderstand with column family's max versions, so use readAllVersions() and readVersions(int) instead.
- [HBASE-16567](#): Core is now up on protobuf 3.1.0 (Coprocessor Endpoints and REST are still on protobuf 2.5.0).
- [HBASE-14004](#): Fix inconsistency between Memstore and WAL which may result in data in remote cluster that is not in the origin (Replication).
- [HBASE-18786](#): FileNotFoundException opening a StoreFile in a primary replica now causes a RegionServer to crash out where before it would be ignored (or optionally handled via close/reopen).
- [HBASE-17956](#): Raw scans will also read TTL expired cells.
- [HBASE-17017](#): Removed per-region latency histogram metrics.
- [HBASE-19483](#): Added ACL checks to RSGroup commands - On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands.
- [HBASE-19358](#): Added ACL checks to RSGroup commands (HBASE-19483): On a secure cluster, only users with ADMIN rights will be able to execute RSGroup commands. Improved stability of splitting log when do failover.
- [HBASE-18883](#): Updated our Curator version to 4.0 - Users who experience classpath issues due to version conflicts are recommended to use either the hbase-shaded-client or hbase-shaded-mapreduce artifacts.
- [HBASE-16388](#): Prevent client threads being blocked by only one slow region server - Added a new configuration to limit the max number of concurrent request to one region server.
- [HBASE-15212](#): New configuration to limit RPC request size to protect the server against very large incoming RPC requests. All requests larger than this size will be immediately rejected before allocating any resources.
- [HBASE-15968](#): This issue resolved two long-term issues in HBase: 1) Puts may be masked by a delete before them, and 2) Major compactions change query results. Offers a new behavior to fix this issue with a little performance reduction. Disabled by default. See the issue for details and caveats.
- [HBASE-13701](#): SecureBulkLoadEndpoint has been integrated into HBase core as default bulk load mechanism. It is no longer needed to install it as a coprocessor endpoint.
- [HBASE-9774](#): HBase native metrics and metric collection for coprocessors.
- [HBASE-18294](#): Reduce global heap pressure: flush based on heap occupancy.

Apache Hive/Hive on Spark/HCatalog

Apache Hive

The following changes are introduced to Hive in CDH 6.0.0, and are not backwards compatible:

- [UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting](#) on page 533
- [Support for UNION DISTINCT](#) on page 534
- [OFFLINE and NO_DROP Options Removed from Table and Partition DDL](#) on page 534
- [DESCRIBE Query Syntax Change](#) on page 535
- [CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names](#) on page 535
- [Reserved and Non-Reserved Keyword Changes in HiveQL](#) on page 535
- [Apache Hive API Changes in CDH 6.0.0](#) on page 537
- [Apache Hive Configuration Changes in CDH 6.0.0](#) on page 538
- [HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes](#) on page 540
- [Values Returned for Decimal Numbers Are Now Padded with Trailing Zeroes to the Scale of the Specified Column](#) on page 541
- [Hive Logging Framework Switched to SLF4J/Log4j 2](#) on page 541
- [Deprecated Parquet Java Classes Removed from Hive](#) on page 541
- [Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms](#) on page 541
- [S3N Connector Is Removed from CDH 6.0](#) on page 542
- [Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request](#) on page 542
- [Support Added for Escaping Carriage Returns and New Line Characters for Text Files \(LazySimpleSerDe\)](#) on page 542
- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 542

Changing Table File Format from ORC with the ALTER TABLE Command Not Supported in CDH 6

Changing the table file format from ORC to another file format with the ALTER TABLE command is not supported in CDH 6 (it returns an error).

UNION ALL Statements Involving Data Types from Different Type Groups No Longer Use Implicit Type Casting

Prior to this change, Hive performed implicit casts when data types from different type groups were specified in queries that use UNION_ALL. For example, before CDH 6.0, if you had the two following tables:

Table "one"

one.col_1	one.col_2	one.col_3
21	hello_all	b

Where col_1 datatype is int, col_2 datatype is string, and col_3 datatype is char(1).

Table "two"

two.col_4	two.col_5	two.col_6
75.0	abcde	45

Where col_4 datatype is double, col_5 datatype is varchar(5), and col_6 datatype is int.

And you ran the following UNION_ALL query against these two tables:

```
SELECT * FROM one UNION ALL SELECT col_4 AS col_1, col_5 AS col_2, col_6 AS col_3 FROM two;
```

You received the following result set:

_u1.col_1	_u1.col_2	_u1.col_3
75.0	abcde	4
21.0	hello	b

Note that this statement implicitly casts the values from table `one` with the following errors resulting in data loss:

- `one.col_1` is cast to a `double` datatype
- `one.col_2` is cast to a `varchar(5)` datatype, which truncates the original value from `hello_all` to `hello`
- `one.col_3` is cast to a `char(1)` datatype, which truncates the original value from 45 to 4

In CDH 6.0, no implicit cast is performed across different type groups. For example, `STRING`, `CHAR`, and `VARCHAR` are in one type group, and `INT`, `BIGINT`, and `DECIMAL` are in another type group, and so on. So, in CDH 6.0 and later, the above query that uses `UNION ALL`, returns an exception for the columns that contain datatypes that are not part of a type group. In CDH 6.0 and later, Hive performs the implicit cast only *within* type groups and not *across* different type groups. For more information, see [HIVE-14251](#).

Support for UNION DISTINCT

Support has been added for the `UNION DISTINCT` clause in HiveQL. See [HIVE-9039](#) and [the Apache wiki](#) for more details. This feature introduces the following incompatible changes to HiveQL:

- **Behavior in CDH 5:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses either before a `UNION ALL` clause or at the end of the query, resulting in the following behaviors:
 - When specified before, these clauses are applied to the query before `UNION ALL` is applied.
 - When specified at the end of the query, these clauses are applied to the query after `UNION ALL` is applied.
- The `UNION` clause is equivalent to `UNION ALL`, in which no duplicates are removed.

- **Behavior in CDH 6:**

- `SORT BY`, `CLUSTER BY`, `ORDER BY`, `LIMIT`, and `DISTRIBUTE BY` can be specified without delineating parentheses *only* at the end of the query, resulting in the following behaviors:
 - These clauses are applied to the entire query.
 - Specifying these clauses before the `UNION ALL` clause results in a parsing error.
- The `UNION` clause is equivalent to `UNION DISTINCT`, in which all duplicates are removed.

OFFLINE and NO_DROP Options Removed from Table and Partition DDL

Support for Hive table and partition protection options have been removed in CDH 6.0, which includes removal of the following functionality:

- Support has been removed for:

- `ENABLE | DISABLE NO_DROP [CASCADE]`
- `ENABLE | DISABLE OFFLINE`
- `ALTER TABLE ... IGNORE PROTECTION`

- The following support has also been removed from the `HiveMetastoreClient` class:

The `ignoreProtection` parameter has been removed from the `dropPartitions` methods in the `IMetaStoreClient` interface.

For more information, see [HIVE-11145](#).

Cloudera recommends that you use Apache Sentry to replace most of this functionality. Although Sentry governs permissions on `ALTER TABLE`, it does not include permissions that are specific to a partition. See [Authorization Privilege Model for Hive and Impala](#) and [Configuring the Sentry Service](#).

DESCRIBE Query Syntax Change

In CDH 6.0 syntax has changed for `DESCRIBE` queries as follows:

- `DESCRIBE` queries where the column name is separated by the table name using a period is no longer supported:

```
DESCRIBE testTable.testColumn;
```

Instead, the table name and column name must be separated with a space:

```
DESCRIBE testTable testColumn;
```

- The `partition_spec` must appear *after* the table name, but *before* the optional column name:

```
DESCRIBE default.testTable PARTITION (part_col = 100) testColumn;
```

For more details, see the [Apache wiki](#) and [HIVE-12184](#).

CREATE TABLE Change: Periods and Colons No Longer Allowed in Column Names

In CDH 6.0, `CREATE TABLE` statements fail if any of the specified column names contain a period or a colon. For more information, see [HIVE-10120](#) and the [Apache wiki](#).

Reserved and Non-Reserved Keyword Changes in HiveQL

Hive reserved and non-reserved keywords have changed in CDH 6.0. *Reserved keywords* cannot be used as table or column names unless they are enclosed with back ticks (for example, `data`). *Non-reserved keywords* can be used as table or column names without enclosing them with back ticks. Non-reserved keywords have proscribed meanings in HiveQL, but can still be used as table or column names. For more information about the changes to reserved and non-reserved words listed below, see [HIVE-6617](#) and [HIVE-14872](#).

In CDH 6.0, the following changes have been introduced to Hive reserved and non-reserved keywords and are not backwards compatible:

- [Hive New Reserved Keywords Added in CDH 6.0](#) on page 535
- [Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0](#) on page 536
- [Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0](#) on page 536
- [Hive New Non-Reserved Keywords Added in CDH 6.0](#) on page 536
- [Hive Non-Reserved Keyword Removed in CDH 6.0](#) on page 537

Hive New Reserved Keywords Added in CDH 6.0

The following table contains new reserved keywords that have been added:

COMMIT	CONSTRAINT	DEC	EXCEPT
FOREIGN	INTERVAL	MERGE	NUMERIC
ONLY	PRIMARY	REFERENCES	ROLLBACK
START			

Hive Non-Reserved Keywords Converted to Reserved Keywords in CDH 6.0

The following table contains non-reserved keywords that have been converted to be reserved keywords:

ALL	ALTER	ARRAY	AS
AUTHORIZATION	BETWEEN	BIGINT	BINARY
BOOLEAN	BOTH	BY	CREATE
CUBE	CURSOR	DATE	DECIMAL
DOUBLE	DELETE	DESCRIBE	DROP
EXISTS	EXTERNAL	FALSE	FETCH
FLOAT	FOR	FULL	GRANT
GROUP	GROUPING	IMPORT	IN
INT	INNER	INSERT	INTERSECT
INTO	IS	LATERAL	LEFT
LIKE	LOCAL	NONE	NULL
OF	ORDER	OUT	OUTER
PARTITION	PERCENT	PROCEDURE	RANGE
READS	REGEXP	REVOKE	RIGHT
RLIKE	ROLLUP	ROW	ROWS
SET	SMALLINT	TABLE	TIMESTAMP
TO	TRIGGER	TRUNCATE	UNION
UPDATE	USER	USING	VALUES
WITH	TRUE		

Hive Reserved Keywords Converted to Non-Reserved Keywords in CDH 6.0

The following table contains reserved keywords that have been converted to be non-reserved keywords:

CURRENT_DATE	CURRENT_TIMESTAMP	HOLD_DDLTIME	IGNORE
NO_DROP	OFFLINE	PROTECTION	READONLY

Hive New Non-Reserved Keywords Added in CDH 6.0

The following table contains new non-reserved keywords that have been added:

ABORT	AUTOCOMMIT	CACHE	DAY
DAYOFWEEK	DAYS	DETAIL	DUMP
EXPRESSION	HOUR	HOURS	ISOLATION
KEY	LAST	LEVEL	MATCHED
MINUTE	MINUTES	MONTH	MONTHS
NORELY	NOVALIDATE	NULLS	OFFSET
OPERATOR	RELY	SECOND	SECONDS
SNAPSHOT	STATUS	SUMMARY	TRANSACTION

VALIDATE	VECTORIZATION	VIEWS	WAIT
WORK	WRITE	YEAR	YEARS

Hive Non-Reserved Keyword Removed in CDH 6.0

The following non-reserved keyword has been removed:

DEFAULT

Apache Hive API Changes in CDH 6.0.0

The following changes have been introduced to the Hive API in CDH 6.0, and are not backwards compatible:

- [AddPartitionMessage.getPartitions\(\) Can Return NULL](#) on page 537
- [DropPartitionEvent and PreDropPartitionEvent Class Changes](#) on page 537
- [GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date](#) on page 537
- [GenericUDF.getConstantLongValue Has Been Removed](#) on page 537
- [Increased Width of Hive Metastore Configuration Columns](#) on page 537

AddPartitionMessage.getPartitions() Can Return NULL

The `getPartitions()` method has been removed from the `AddPartitionEvent` class in the `org.apache.hadoop.hive.metastore.events` interface. It was removed to prevent out-of-memory errors when the list of partitions is too large.

Instead use the `getPartitionIterator()` method. For more information, see [HIVE-9609](#) and the [AddPartitionEvent documentation](#).

DropPartitionEvent and PreDropPartitionEvent Class Changes

The `getPartitions()` method has been removed and replaced by the `getPartitionIterator()` method in the `DropPartitionEvent` class and the `PreDropPartitionEvent` class.

In addition, the `(Partition partition, boolean deleteData, HiveMetastore.HMSHandler handler)` constructors have been deleted from the `PreDropPartitionEvent` class. For more information, see [HIVE-9674](#) and the [PreDropPartitionEvent documentation](#).

GenericUDF.getTimestampValue Method Now Returns Timestamp Instead of Date

The `getTimestampValue` method in the `GenericUDF` class now returns a `TIMESTAMP` value instead of a `DATE` value. For more information, see [HIVE-10275](#) and the [GenericUDF documentation](#).

GenericUDF.getConstantLongValue Has Been Removed

The `getConstantLongValue` method has been removed from the `GenericUDF` class. It has been noted by the community that this method is not used in Hive. For more information, see [HIVE-10710](#) and the [GenericUDF documentation](#).

Increased Width of Hive Metastore Configuration Columns

The columns used for configuration values in the Hive metastore have been increased in width, resulting in the following incompatible changes in the `org.apache.hadoop.hive.metastore.api` interface.

This change introduced an incompatible change to the `get_table_names_by_filter` method of the `ThriftHiveMetastore` class. Before this change, this method accepts a string filter, which allows clients to filter a table by its `TABLEPROPERTIES` value. For example:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
    PARAMS + "test_param_1 <> \"yellow\";

org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_
    PARAMS + "test_param_1 = \"yellow\";
```

After this change, the `TABLE_PARAMS.PARAM_VALUE` column is now a `CLOB` data type. Depending on the type of database that you use (for example, MySQL, Oracle, or PostgreSQL), the semantics may have changed and operators like "`=`", "`<>`", and "`!=`" might not be supported. Refer to the documentation for your database for more information. You must use operators that are compatible with `CLOB` data types. There is no equivalent "`<>`" operator that is compatible with `CLOB`. So there is no equivalent operator for the above example that uses the "`<>`" inequality operator. The equivalent for "`=`" is the `LIKE` operator so you would rewrite the second example above as:

```
org.apache.hadoop.hive.metastore.api.hive_metastoreConstants.HIVE_FILTER_FIELD_PARAMS + "test_param_1 LIKE \\"yellow\\";
```

For more information, see [HIVE-12274](#).

Apache Hive Configuration Changes in CDH 6.0.0

The following configuration property changes have been introduced to Hive in CDH 6.0, and are not backwards compatible:

- [Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables](#) on page 538
- [Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters](#) on page 538
- [Hive Strict Checks Have Been Re-factored To Be More Granular](#) on page 539
- [Java XML Serialization Has Been Removed](#) on page 539
- [Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties](#) on page 540
- [HiveServer2 Impersonation Property \(hive.server2.enable.impersonation\) Removed](#) on page 540
- [Changed Default File Format for Storing Intermediate Query Results](#) on page 540

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to false, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to true. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on hive.enforce.bucketing](#) and the [topic on hive.enforce.sorting](#).

Hive Throws an Exception When Processing HDFS Directories Containing Unsupported Characters

Directories in HDFS can contain unprintable or unsupported characters that are not visible even when you run the `hadoop fs -ls` command on the directories. When external tables are created with the `MSCK REPAIR TABLE` command, the partitions using these HDFS directories that contain unsupported characters are unusable for Hive. To avoid this, the configuration parameter `hive.msck.path.validation` has been added. This configuration property controls the behavior of the `MSCK REPAIR TABLE` command, enabling you to set whether validation checks are run on the HDFS directories when `MSCK REPAIR TABLE` is run.

The property `hive.msck.path.validation` can be set to one of the following values:

Value Name	Description
throw	Causes Hive to throw an exception when it tries to process an HDFS directory that contains unsupported characters with the <code>MSCK REPAIR TABLE</code> command. This is the default setting for <code>hive.msck.path.validation</code> .
skip	Causes Hive to skip the skip the directories that contain unsupported characters, but still repairs the others.
ignore	Causes Hive to completely skip any validation of HDFS directories when the <code>MSCK REPAIR TABLE</code> command is run. This setting can cause bugs because unusable partitions are created.

By default, the `hive.msck.path.validation` property is set to `throw`, which causes Hive to throw an exception when MSCK REPAIR TABLE is run and HDFS directories containing unsupported characters are encountered. To work around this, set this property to `skip` until you can repair the HDFS directories that contain unsupported characters.

To set this property in Cloudera Manager:

1. In the Admin Console, select the Hive service.
2. Click the **Configuration** tab.
3. Search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting.
4. In the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting, add the **Name** of the property, the **Value** (`throw`, `skip`, or `ignore`), and a **Description** of the setting.
5. Click **Save Changes** and restart the service.

For more information, see [HIVE-10722](#).

Hive Strict Checks Have Been Re-factored To Be More Granular

Originally, the configuration property `hive.mapred.mode` was added to restrict certain types of queries from running. Now it has been broken down into more fine-grained configurations, one for each type of restricted query pattern. The configuration property `hive.mapred.mode` has been removed and replaced with the following configuration properties, which provide more granular control of Hive strict checks:

Configuration Property	Description	Default Value
<code>hive.strict.checks.bucketing</code>	When set to <code>true</code> , running <code>LOAD DATA</code> queries against bucketed tables is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.type.safety</code>	When set to <code>true</code> , comparing <code>bigint</code> to <code>string</code> data types or <code>bigint</code> to <code>double</code> data types is not allowed.	<code>true</code> . This is a backwards incompatible change.
<code>hive.strict.checks.orderby.no.limit</code>	When set to <code>true</code> , prevents queries from being run that contain an <code>ORDER BY</code> clause with no <code>LIMIT</code> clause.	<code>false</code>
<code>hive.strict.checks.no.partition.filter</code>	When set to <code>true</code> , prevents queries from being run that scan a partitioned table but do not filter on the partition column.	<code>false</code>
<code>hive.strict.checks.cartesian.product</code>	When set to <code>true</code> , prevents queries from being run that contain a Cartesian product (also known as a cross join).	<code>false</code>

All of these properties can be set with Cloudera Manager in the following configuration settings for the Hive service:

- **Restrict LOAD Queries Against Bucketed Tables** (`hive.strict.checks.bucketing`)
- **Restrict Unsafe Data Type Comparisons** (`hive.strict.checks.type.safety`)
- **Restrict Queries with ORDER BY but no LIMIT clause** (`hive.strict.checks.orderby.no.limit`)
- **Restrict Partitioned Table Scans with no Partitioned Column Filter** (`hive.strict.checks.no.partition.filter`)
- **Restrict Cross Joins (Cartesian Products)** (`hive.strict.checks.cartesian.product`)

For more information about these configuration properties, see [HIVE-12727](#), [HIVE-15148](#), [HIVE-18251](#), and [HIVE-18552](#).

Java XML Serialization Has Been Removed

The configuration property `hive.plan.serialization.format` has been removed. Previously, this configuration property could be set to either `javaXML` or `kryo`. Now the default is `kryo` serialization, which cannot be changed. For more information, see [HIVE-12609](#) and the [Apache wiki](#).

Configuration Property Enabling Column Position Usage with GROUP BY and ORDER BY Separated into Two Properties

The configuration property `hive.groupby.orderby.position.alias`, which enabled using column position with the GROUP BY and the ORDER BY clauses has been removed and replaced with the following two configuration properties. These configuration properties enable using column position with GROUP BY and ORDER BY separately:

Configuration Property Name	Description/Default Setting	Possible Values
<code>hive.groupby.position.alias</code>	When set to true, specifies that columns can be referenced with their position when using GROUP BY clauses in queries. Default Setting: false. This behavior is turned off by default.	true false
<code>hive.orderby.position.alias</code>	When set to true, specifies that columns can be referenced with their position when using ORDER BY clauses in queries. Default Setting: true. This behavior is turned on by default.	true false

For more information, see [HIVE-15797](#) and the Apache wiki entries for [configuration properties](#), [GROUP BY syntax](#), and [ORDER BY syntax](#).

HiveServer2 Impersonation Property (`hive.server2.enable.impersonation`) Removed

In earlier versions of CDH, the following two configuration properties could be used to set impersonation for HiveServer2:

- `hive.server2.enable.impersonation`
- `hive.server2.enable.doAs`

In CDH 6.0, `hive.server2.enable.impersonation` is removed. To configure impersonation for HiveServer2, use the configuration property `hive.server2.enable.doAs`. To set this property in Cloudera Manager, select the Hive service and click on the **Configuration** tab. Then search for the **HiveServer2 Enable Impersonation** setting and select the checkbox to enable HiveServer2 impersonation. This property is enabled by default in CDH 6.

For more information about this property, see the [Apache wiki documentation for HiveServer2 configuration properties](#).

Changed Default File Format for Storing Intermediate Query Results

The configuration property `hive.query.result.fileformat` controls the file format in which a query's intermediate results are stored. In CDH 6, the default setting for this property has been changed from `TextFile` to `SequenceFile`.

To change this configuration property in Cloudera Manager:

1. In the Admin Console, select the Hive service and click on the **Configuration** tab.
2. Then search for the **Hive Service Advanced Configuration Snippet (Safety Valve)** for `hive-site.xml` setting and add the following information:
 - **Name:** `hive.query.result.fileformat`
 - **Value:** Valid values are `TextFile`, `SequenceFile` (default), or `RCfile`
 - **Description:** Sets the file format in which a query's intermediate results are stored.
3. After you add this information, click **Save Changes** and restart the Hive service.

For more information about this parameter, see the [Apache wiki](#).

HiveServer2 Thrift API Code Repackaged Resulting in Class File Location Changes

HiveServer2 Thrift API code has been repackaged in CDH 6.0, resulting in the following changes:

- All files generated by the Thrift API for HiveServer2 have moved from the following *old* namespace:

`org.apache.hive.service.cli.thrift`

To the following *new* namespace:

```
org.apache.hive.service.rpc.thrift
```

- All files generated by the Thrift API for HiveServer2 have moved into a separate jar file called `service-rpc`.

As a result of these changes, all Java classes such as `TCLIService.java`, `TOpenSessionReq.java`, `TSessionHandle.java`, and `TGetSchemasReq.java` have changed locations. For more information, see [HIVE-12442](#).

Values Returned for Decimal Numbers Are Now Padded with Trailing Zeros to the Scale of the Specified Column

Decimal values that are returned in query results are now padded with trailing zeroes to match the specified scale of the corresponding column. For example, *before* this change, when Hive read a decimal column with a specified scale of 5, the value returned for zero was returned as 0. *Now*, the value returned for zero is 0.00000. For more information, see [HIVE-12063](#).

Hive Logging Framework Switched to SLF4J/Log4j 2

The logging framework for Hive has switched to [SLF4J \(Simple Logging Facade for Java\)](#) and now uses [Log4j 2](#) by default. Use of Log4j 1.x, Apache Commons Logging, and `java.util.logging` have been removed. To accommodate this change, write all Log4j configuration files to be compatible with Log4j 2.

For more information, see [HIVE-12237](#), [HIVE-11304](#), and the [Apache wiki](#).

Deprecated Parquet Java Classes Removed from Hive

The deprecated parquet classes, `parquet.hive.DeprecatedParquetInputFormat` and `parquet.hive.DeprecatedParquetOutputFormat` have been removed from Hive because they resided outside of the `org.apache` namespace. Any existing tables that use these classes are automatically migrated to the new SerDe classes when the metastore is upgraded.

Use one of the following options for specifying the Parquet SerDe for new Hive tables:

- Specify in the `CREATE TABLE` statement that you want it stored as Parquet. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING) STORED AS PARQUET;
```

- Set the `INPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat` and set the `OUTPUTFORMAT` to `org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat`. For example:

```
CREATE TABLE <parquet_table_name> (col1 INT, col2 STRING)
STORED AS
  INPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat"
  OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat";
```

For more information, see [HIVE-6757](#) and the [Apache wiki](#).

Removed JDBC, Counter-based, and HBase-based Statistics Collection Mechanisms

Support for JDBC, counter-based, and HBase-based statistics collection mechanisms has been removed from Hive. The following configuration properties are no longer supported:

- `hive.stats.dbclass`
- `hive.stats.retries.wait`
- `hive.stats.retries.max`
- `hive.stats.jdbc.timeout`
- `hive.stats.dbconnectionstring`
- `hive.stats.jdbcdrive`
- `hive.stats.key.prefix.reserve.length`

This change also removed the `cleanUp(String keyPrefix)` method from the [StatsAggregator interface](#).

Now all Hive statistics are collected on the default file system. For more information, see [HIVE-12164](#), [HIVE-12411](#), [HIVE-12005](#), and the [Apache wiki](#).

S3N Connector Is Removed from CDH 6.0

The [S3N connector](#), which is used to connect to the Amazon S3 file system from Hive has been removed from CDH 6.0. To connect to the S3 file system from Hive in CDH 6.0, you must now use the S3A connector. There are a number of differences between the S3N and the S3A connectors, including configuration differences. See the [Apache wiki page on integrating with Amazon Web Services](#) for details.

Migration involves making the following changes:

- Changing all metastore data containing URIs that start with `s3n://` to `s3a://`. This change is performed automatically when you upgrade the Hive metastore.
- Changing all scripts containing URIs that start with `s3n://` to `s3a://`. You must perform this change manually.

Columns Added to TRowSet Returned by the Thrift TCLIService#GetTables Request

Six additional columns have been added to the `TRowSet` that is returned by the `TCLIService#GetTables` request. These columns were added to comply with the official JDBC API. For more information, see the documentation for [java.sql.DatabaseMetaData](#).

The columns added are:

Column Name	Description
REMARKS	Explanatory comment on the table.
TYPE_CAT	Types catalog.
TYPE_SCHEMA	Types schema.
TYPE_NAME	Types name.
SELF_REFERENCING_COL_NAME	Name of the designed identifier column of a typed table.
REF_GENERATION	Specifies how values in the SELF_REFERENCING_COL_NAME column are created.

For more information, see [HIVE-7575](#).

Support Added for Escaping Carriage Returns and New Line Characters for Text Files (LazySimpleSerDe)

Support has been added for escaping carriage returns and new line characters in text files by modifying the `LazySimpleSerDe` class. Without this change, carriage returns and new line characters are interpreted as delimiters, which causes incorrect query results.

This feature is controlled by the SerDe property `serialization.escape.crlf`. It is enabled (set to `true`) by default. If `serialization.escape.crlf` is enabled, '`r`' or '`n`' cannot be used as separators or field delimiters.

This change only affects text files and removes the `getNullString` method from the [LazySerDeParameters class](#). For more information, see [HIVE-11785](#).

Bucketing and Sorting Enforced by Default When Inserting Data into Hive Tables

The configuration properties `hive.enforce.sorting` and `hive.enforce.bucketing` have been removed. When set to `false`, these configurations disabled enforcement of sorted and bucketed tables when data was inserted into a table. Removing these configuration properties effectively sets these properties to `true`. In CDH 6.0, bucketing and sorting are enforced on Hive tables during insertions and cannot be turned off. For more information, see the Apache wiki [topic on hive.enforce.bucketing](#) and the [topic on hive.enforce.sorting](#).

Hue

There are no incompatible changes in this release.

Apache Impala

List of Reserved Words Updated

The list of [reserved words](#) in Impala was updated in CDH 6.0.

If you need to use the reserved words from previous versions of CDH, set the impalad and catalogd startup option, `reserved_words_version`, to "2.11.0".

Decimal V2 Used by Default

In Impala, two different behaviors of `DECIMAL` types are supported. In CDH 6.0, `DECIMAL` V2 is used by default. See [DECIMAL Type](#) for detail information.

If you need to continue using the first version of the `DECIMAL` type for the backward compatibility of your queries, set the [DECIMAL_V2](#) query option to `FALSE`.

Behavior of Column Aliases Changed

To conform to the SQL standard, Impala no longer performs alias substitution in the subexpressions of `GROUP BY`, `HAVING`, and `ORDER BY`.

For example, the following statements will now result in syntax errors.

```
SELECT int_col / 2 AS x
FROM functional.alltypes
GROUP BY x / 2;

SELECT int_col / 2 AS x
FROM functional.alltypes
ORDER BY -x;

SELECT int_col / 2 AS x
FROM functional.alltypes
GROUP BY x
HAVING x > 3;
```

Default PARQUET_ARRAY_RESOLUTION Changed

The `PARQUET_ARRAY_RESOLUTION` query option controls the path-resolution behavior for Parquet files with nested arrays. The default value for the `PARQUET_ARRAY_RESOLUTION` was changed to `THREE_LEVEL` in CDH 6.0. Review your queries to see if the default value change result in different result sets.

See [PARQUET_ARRAY_RESOLUTION Query Option](#) for the information about the option.

Non-standard Timezone Names Unsupported

As the initial step for IANA timezone integration in the coming release, Impala will drop the support for non-standard timezone aliases in CDH 6.0.

Impala supports a majority of the [IANA time zones](#) with the following exceptions of time zones not supported: America/Fort_Nelson, America/Punta_Arenas, Asia/Atyrau, Asia/Barnaul, Asia/Famagusta, Asia/Tomsk, Asia/Yangon, Europe/Astrakhan, Europe/Kirov, Europe/Saratov, Europe/Ulyanovsk, GMT+0, GMT-0, ROC

See [Unsupported Time Zone](#) for the list of time zone aliases no longer supported and the canonical names you can use to replace the unsupported aliases with.

Return Type Changed for EXTRACT and DATE_PART Functions in CDH 6.0 / Impala 3.0

The following changes were made to the `EXTRACT` and `DATE_PART` functions:

- The output type of the `EXTRACT` and `DATE_PART` functions was changed to `BIGINT`.
- Extracting the millisecond part from a `TIMESTAMP` returns the seconds component and the milliseconds component. For example, `EXTRACT (CAST('2006-05-12 18:27:28.123456789' AS TIMESTAMP), 'MILLISECOND')` will return 28123.

Apache Kafka

Kafka is now bundled as part of CDH. The following sections describe incompatible changes between the previous, separately installed Kafka (CDK powered by Apache Kafka version 3.1) and the CDH 6.0.0 Kafka version. These changes

Cloudera Enterprise 6 Release Guide

affect clients built with CDH 6.0.0 libraries. Cloudera recommends upgrading clients to the new release; however clients built with previous versions of Kafka will continue to function.

Packaging

CDH and previous distributions of Kafka (CDK Powered by Apache Kafka) cannot coexist in the same cluster.

Deprecated Scala-based Client API and New Java Client API

Scala-based clients are deprecated in this release and will be removed in an upcoming release.

The following Scala-based client implementations from package `kafka.*` (known as 'old clients') are deprecated and unsupported as of CDH 6.0.0:

- `kafka.consumer.*`
- `kafka.producer.*`
- `kafka.admin.*`

Client applications making use of these implementations must be migrated to corresponding Java clients available in `org.apache.kafka.*` (the 'new clients') package. Existing command line options and tools now use the new clients package.

Command Line Options Removed

Some command line tools are affected by the deprecation of old clients (see [previous entry](#)). The following options have been removed and are not recognized as valid options:

- `--new-consumer`
- `--old-consumer`
- `--old-producer`

The tools affected use the new clients.

Command Line Tools Removed

The following command line tools and runnable classes are removed:

- `kafka-replay-log-producer`
- `kafka-simple-consumer-shell`
- `kafka.tools.ReplayLogProducer`
- `kafka.tools.SimpleConsumerShell`
- `kafka.tools.ExportZkOffset`
- `kafka.tools.ImportZkOffset`
- `kafka.tools.SimpleConsumerPerformance`
- `kafka.tools.UpdateOffsetsInZK`
- `kafka.tools.VerifyConsumerRebalance`
- `kafka.tools.ProducerPerformance`

Consumer API Changes

Consumer methods invoked with unassigned partitions now raise an `IllegalStateException` instead of an `IllegalArgumentException`.

Previous versions of the Consumer method `poll(long)` would wait for metadata updates regardless of timeout parameter. This behavior is expected to change in future releases; make sure your client applications include an appropriate timeout parameter and do not rely on the previous behavior.

Exception Classes Removed

The following exceptions were deprecated in a previous release and are not thrown anymore are removed:

- `GroupCoordinatorNotAvailableException`
- `GroupLoadInProgressException`
- `NotCoordinatorForGroupException`

- kafka.common.KafkaStorageException

Metrics Updated

Kafka consumers' per-partition metrics were changed to use tags for topic and partition rather than the metric name. For more information see [KIP-225](#).

Apache Kudu

There are no incompatible changes in this release.

Apache Oozie

There are no incompatible changes in this release.

Apache Parquet

Packages and Group ID Renamed

As a part of the Apache incubation process, all Parquet packages and the project's group ID were renamed as follows:

Parquet Version	1.6.0 and lower (CDH 5.x)	1.7.0 and higher (CDH 6.x)
Java Package Names	parquet.*	org.apache.parquet.*
Group ID	com.twitter	org.apache.parquet

If you directly consume the Parquet API, instead of using Parquet through Hive, Impala or other CDH component, you need to update your code to reflect these changes:

Update *.java files:

Before	After
import parquet.*;	import org.apache.parquet.*;

Update pom.xml:

Before	After
<pre><dependency> <groupId> com.twitter </groupId> <version> \${parquet.version} </version> </dependency></pre>	<pre><dependency> <groupId> org.apache.parquet </groupId> <version> \${parquet.version} </version> </dependency></pre>

API Methods Removed

In Parquet 1.6, a number of API methods were removed from the parquet.hadoop.ParquetInputSplit class that depended on reading metadata on the client side. Metadata should be read on the task side instead.

Removed Method	New Method to Use
parquet.hadoop.ParquetInputSplit.getFileSchema	org.apache.parquet.hadoop.api.InitContext.getFileSchema
parquet.hadoop.ParquetInputSplit.getRequestedSchema	org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getRequestedSchema
parquet.hadoop.ParquetInputSplit.getReadSupportMetadata	org.apache.parquet.hadoop.api.ReadSupport.ReadContext.getReadSupportMetadata
parquet.hadoop.ParquetInputSplit.getBlocks	org.apache.parquet.hadoop.metadata.ParquetMetadata.getBlocks
parquet.hadoop.ParquetInputSplit.getExtraMetadata	-

Apache Pig

The following change is introduced to Pig in CDH 6.0 and is not a backwards compatible change. You must modify your Pig scripts as described below.

Removal of the Apache DataFu Pig JAR from CDH 6

Apache DataFu Pig is a collection of user-defined functions that can be used with Pig for data mining and statistical analysis on large-scale data. The DataFu JAR was included in CDH 5, but due to very low adoption rates, the JAR was deprecated in CDH 5.9 and is being removed from CDH 6, starting with CDH 6.0. It is no longer supported.

Recommended Migration Strategy

A simple way to assess what DataFu functions you are using in your Pig scripts is to use the `grep` utility to search for occurrences of "datafu" in your code. When DataFu functions are used in Pig scripts, you must use a function definition entry that contains "datafu" like the following example:

```
define <function_name> datafu.pig... .<class_name>();
```

Use `grep` to search for the string "datafu" in your scripts and that will identify where the DataFu JAR is used.

Cloudera recommends migrating to Hive UDFs or operators wherever it is possible. However, if there are cases where it is impossible to replace DataFu functions with Hive functions, download the upstream version of the DataFu Pig libraries and place them on the node where the Pig front end is used. To preserve compatibility, use the version 1.1.0 JAR, which was the version included in CDH 5. You can download the JAR file [here](#). However, Cloudera does not support using this upstream DataFu JAR file.

Mapping DataFu UDFs to Hive UDFs

The following Hive UDFs map to DataFu UDFs and can be used instead in Pig scripts with the caveats that are listed:

Table 42: Hive Functions That Map to DataFu Functions

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
MD5 (hash)	Computes the MD5 value of a string and outputs a hex value by default.	md5	None
SHA (hash)	Computes the SHA value of a string and outputs a hex value by default.	sha/sha1	None
RandInt (random)	Generates a uniformly distributed integer between two bounds.	rand	None
VAR (stats)	Generates the variance of a set of values.	variance	None
Median (stats)	Computes the median for a sorted input bag. A special case of the Quantile function.	percentile(0.5)	See Limitations When Substituting Quantile and Median DataFu Functions on page 547.
Quantile (stats)	Computes quantiles for a sorted input bag.	percentile	See Limitations When Substituting Quantile and Median DataFu Functions on page 547.

DataFu Function (package)	Description	Hive UDF or Operator Equivalent	Caveats
Coalesce (util)	Returns the first non-null value from a tuple, like COALESCE in SQL.	coalesce	None
InUDF (util)	Similar to the SQL IN function, this function provides a convenient way to filter using a logical disjunction over many values.	IN	None

For more information about using Hive UDFs, see https://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_hive_udf.html.

Limitations When Substituting Quantile and Median DataFu Functions

With the exception of Median and Quantile all Hive functions specified in the above table should work as expected in Pig scripts. Median extends Quantile in DataFu functions and the equivalent Hive functions have a similar relationship. However, there is an important difference in how you use percentile and how you use Quantile. The differences are summarized in the following table:

Table 43: Differences Between Usage of DataFu 'Quantile' and Hive 'percentile'

Function:	Quantile	percentile
Input must be sorted:	Yes	No
Nulls are allowed in input:	No	Yes
Examples:	<pre> register datafu-1.2.0.jar; define median datafu.pig.stats.Quantile('0.5'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B = group A by name; C = foreach B { sorted = order A by n; generate group, flatten (median(sorted.n)); } </pre>	<pre> define percentile HiveUDAF ('percentile'); A = LOAD 'nums.txt' AS (name:chararray, n:long); B2 = foreach A generate name, n, 0.5 as perc; C2 = GROUP B2 by name; D2 = FOREACH C2 generate group, percentile (B2. (n, perc)); </pre>

Although DataFu StreamingQuantile and StreamingMedian might appear to match Hive's percentile_approx function, Pig cannot consume percentile_approx.

DataFu Functions with No Hive Function or Operator Equivalent

The following general limitations apply when mapping DataFu UDFs to Hive UDFs:

- Many DataFu functions operate on a custom Pig data structure called a *bag*. No Hive UDFs can operate on Pig bags, so there are no equivalents for these DataFu functions.
- Some DataFu functions are custom functions that do not have Hive UDF equivalents. For example, the DataFu functions that calculate geographic distances, run the PageRank algorithm, or that do sampling. There are no equivalent Hive UDFs for these DataFu functions either.

Table 44: DataFu Functions with No Hive UDF Equivalent

AppendToBag (bags)	AssertUDF (util)	BagConcat (bags)
BagGroup (bags)	BagLeftOuterJoin (bags)	BagSplit (bags)
BoolToInt (util)	CountEach (bags)	DistinctBy (bags)
EmptyBagToNull (bags)	EmptyBagToNullFields (bags)	Enumerate (bags)
FirstTupleFromBag (bags)	HaversineDistInMiles (geo)	IntToBool (util)
MarkovPairs	NullToEmptyBag (bags)	PageRank (linkanalysis)
PrependToBag (bags)	ReservoirSample (sampling)*	ReverseEnumerate (bags)
SampleByKey (sampling)*	SessionCount (sessions)	Sessionize (sessions)
SetIntersect (sets)	SetUnion (sets)	SimpleRandomSample (sampling)*
TransposeTupleToBag (util)	UnorderedPairs (bags)	UserAgentClassify (urls)
WeightedSample (sampling)*	WilsonBinConf (stats)	—

* These DataFu functions might be replaced with TABLESAMPLE in HiveQL. See the [Apache Hive wiki](#).

Cloudera Search

Cloudera Search in CDH 6.0 is rebased on Apache Solr 7.0, which has many incompatibilities with the 4.10 version of Apache Solr used in recent CDH 5 releases, such as the following:

- Solr 7 uses a managed schema by default. Generating an instance directory no longer generates schema.xml. For instructions on switching to a managed schema, see [Switching from schema.xml to Managed Schema](#) in *Apache Solr Reference Guide*.
- Creating a collection using solrctl collection --create without specifying the -c <configName> parameter now uses a default configuration set (named _default) instead of a configuration set with the same name as the collection. To avoid this, always specify the -c <configName> parameter when creating new collections.

For the full list of changes, see the upstream release notes:

- [Apache Solr 5 Release Notes](#)
- [Apache Solr 6 Release Notes](#)
- [Apache Solr 7 Release Notes](#)

Apache Sentry

Apache Sentry contains the following incompatible change in CDH 6.0.0:

- Sentry no longer supports policy file authorization. You must migrate policy files to the database-backed Sentry service before you upgrade to CDH 6.0.0 unless you are using Sentry policy files for Solr. If you are using Sentry policy files for Solr, you must migrate to the database-backed Sentry service after you upgrade.

For information about migrating policy files before you upgrade, see [Migrating from Sentry Policy Files to the Sentry Service](#). For information about migrating policy files for Solr after you upgrade, see [Migrating Sentry Privileges for Solr After Upgrading to CDH 6](#).

Apache Spark

The following sections describe changes in Spark support in CDH 6 that might require special handling during upgrades, or code changes within existing applications.

- All Spark applications built against Spark 1.6 in CDH 5 must be rebuilt against Spark 2.x in CDH 6.
- Spark 2 in CDH 6 works with Java 8, not Java 7. If this change produces any Java code incompatibilities, update your Java code and rebuild the application.

- Spark 2 in CDH 6 works with Scala 2.11, not Scala 2.10. If this change produces any Scala code incompatibilities, update your Scala code and rebuild the application.
- `HiveContext` and `SQLContext` have been removed, although those variables still work for backward compatibility. Use the `SparkSession` object to replace both of these handles.
- `DataFrames` have been removed from the Scala API. `DataFrame` is now a special case of `Dataset`.

Since compile-time type-safety in Python and R is not a language feature, the concept of `Dataset` does not apply to these languages' APIs. Instead, `DataFrame` remains the primary programming abstraction.

- Spark 2.0 and higher do not use an assembly JAR for standalone applications.
- If you have event logs created in CDH 5.3 or lower, you cannot read those logs using Spark in CDH 6.0 or higher.

Apache Sqoop

The following changes are introduced in CDH 6.0, and are not backwards compatible:

- All classes in `com.cloudera.sqoop` packages have been removed in CDH 6.0. Use the corresponding classes from `org.apache.sqoop` packages. For example, use `org.apache.sqoop.SqoopOptions` instead of `com.cloudera.sqoop.SqoopOptions`.



Note: This change only affects customers who build their own application on top of Sqoop classes. Sqoop CLI users are not affected.

- Because of changes introduced in the Sqoop metastore logic, the metastore database created by Sqoop CDH 6 cannot be used by earlier versions. The metastore database created by Sqoop CDH 5 can be used by both Sqoop CDH 5 and Sqoop CDH 6.

Require an explicit option to be specified with `--split-by` for a String column

Using the `--split-by` option with a CHAR or VARCHAR column does not always work properly, so Sqoop now requires the user to set the `org.apache.sqoop.splitter.allow_text_splitter` property to true to confirm that they are aware of this risk.

Example:

```
sqoop import -Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect $MYCONN
--username $MYUSER --password $MYPWD --table "test_table" --split-by "string_column"
```

For more information, see [SQOOP-2910](#).

Make Sqoop fail if user specifies `--direct` connector when it is not available

The `--direct` option is only supported with the following databases: MySQL, PostgreSQL, Oracle, Netezza.

In earlier releases, Sqoop silently ignored this option if it was specified for other databases, but it now throws an error.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table
"direct_import" --direct
```

The command fails with the following error message:

```
Was called with the --direct option, but no direct connector available.
```

For more information, see [SQOOP-2913](#).

Sqoop does not allow `--as-parquetfile` with hcatalog jobs or when hive import with create-hive-table is used

The --create-hive-table option is not supported when the user imports into Hive in Parquet format. Earlier this option was silently ignored, and the data was imported even if the Hive table existed. Sqoop will now fail if the --create-hive-table option is used with the --as-parquetfile option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table" --hive-import --as-parquetfile --create-hive-table
```

The command fails with the following error message:

```
Hive import and create hive table is not compatible with importing into ParquetFile format.
```

For more information, see [SQOOP-3010](#).

Create fail fast for export with --hcatalog-table <HIVE_VIEW>

Importing into and exporting from a Hive view using HCatalog is not supported by Sqoop. A fail fast check was introduced so that now Sqoop throws a descriptive error message if the user specified a Hive view in the value of the --hcatalog-table option.

Example:

```
sqoop import --connect $MYCONN --username $MYUSER --password $MYPWD --table "test_table" --hcatalog-table "test_view"
```

The command fails with the following error message:

```
Reads/Writes from and to Views are not supported by HCatalog
```

For more information, see [SQOOP-3027](#).

Simplify Unicode character support in source files

Simplify Unicode character support in source files (introduced by [SQOOP-3074](#)) by defining explicit locales instead of using EscapeUtils. The Java source files generated by Sqoop will be encoded in UTF-8 format.

For more information, see [SQOOP-3075](#).

Columns added to MySQL after initial Sqoop import, export back to table with same schema fails

If we export from HDFS to an RDBMS table and the file on HDFS has no value for some of the columns defined in the table, Sqoop will use the values of --input-null-string and --input-null-non-string options. Earlier this scenario was not supported and Sqoop failed.

For more information, see [SQOOP-3158](#).

Sqoop fails if the user tries to encode a null value when using --direct connector and a MySQL database

The MySQL direct connector does not support the --null-string, --null-non-string, --input-null-string, and --input-null-non-string options. These options were silently ignored earlier, but Sqoop now throws an error if these options are used with MySQL direct imports and exports.

For more information, see [SQOOP-3206](#).

Apache Zookeeper

There are no incompatible changes in this release.

Timezone Names Unsupported in Impala in CDH 6.0.0

The following table lists the time zone names / aliases no longer supported in Impala along with the canonical names you can use to replace the unsupported aliases with.

Deprecated	Replace with
------------	--------------

ACDT	Australia/South
ACST	Australia/Darwin
ACWST	Australia/Eucla
ADT	America/Thule
AEDT	Australia/Sydney
AEST	Australia/Sydney
AFT	Asia/Kabul
AKDT	America/Anchorage
AKST	America/Anchorage
ALMT	Asia/Almaty
AMST	America/Cuiaba
AMT	Asia/Yerevan
ANAT	Asia/Anadyr
AQTT	Asia/Aqtau
AWST	Australia/West
AZOST	Atlantic/Azores
AZOT	Atlantic/Azores
AZST	Asia/Baku
AZT	Asia/Baku
Acre Time	Brazil/Acre
Afghanistan Time	Asia/Kabul
Alaska Daylight Time	America/Anchorage
Alaska Standard Time	America/Anchorage
Alma-Ata Time	Asia/Almaty
Amazon Summer Time	America/Cuiaba
Amazon Time	Brazil/West
Anadyr Time	Asia/Anadyr
Aqtau Time	Asia/Aqtau
Aqtobe Time	Asia/Aqtobe
Arabia Standard Time	Asia/Aden
Argentine Time	America/Argentina/Buenos_Aires
Armenia Time	Asia/Yerevan
Asia/Riyadh87	-
Asia/Riyadh88	-
Asia/Riyadh89	-
Atlantic Daylight Time	America/Thule

Atlantic Standard Time	America/Puerto_Rico
Australian Central Daylight Time (South Australia)	Australia/South
Australian Central Daylight Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Standard Time (Northern Territory)	Australia/Darwin
Australian Central Standard Time (South Australia)	Australia/South
Australian Central Standard Time (South Australia/New South Wales)	Australia/Yancowinna
Australian Central Western Standard Time	Australia/Eucla
Australian Eastern Daylight Time (New South Wales)	Australia/Sydney
Australian Eastern Daylight Time (Tasmania)	Australia/Hobart
Australian Eastern Daylight Time (Victoria)	Australia/Victoria
Australian Eastern Standard Time (New South Wales)	Australia/Sydney
Australian Eastern Standard Time (Queensland)	Australia/Brisbane
Australian Eastern Standard Time (Tasmania)	Australia/Hobart
Australian Eastern Standard Time (Victoria)	Australia/Victoria
Australian Western Standard Time	Australia/West
Azerbaijan Summer Time	Asia/Baku
Azerbaijan Time	Asia/Baku
Azores Summer Time	Atlantic/Azores
Azores Time	Atlantic/Azores
BDT	Asia/Dhaka
BNT	Asia/Brunei
BOT	America/La_Paz
BRST	America/Sao_Paulo
BRT	America/Sao_Paulo
BTT	Asia/Thimbu
Bangladesh Time	Asia/Dhaka
Bhutan Time	Asia/Thimbu
Bolivia Time	America/La_Paz
Bougainville Standard Time	Pacific/Bougainville
Brasilia Summer Time	America/Sao_Paulo
Brasilia Time	America/Sao_Paulo
British Summer Time	GB
Brunei Time	Asia/Brunei
CCT	Indian/Cocos
CDT	America/Chicago

CEST	CET
CHADT	NZ-CHAT
CHAST	NZ-CHAT
CHOT	Asia/Choibalsan
CHUT	Pacific/Yap
CKT	Pacific/Rarotonga
CLST	America/Santiago
CLT	America/Santiago
COT	America/Bogota
CVT	Atlantic/Cape_Verde
CXT	Indian/Christmas
Canada/East-Saskatchewan	America/Regina
Cape Verde Time	Atlantic/Cape_Verde
Central African Time	Africa/Harare
Central Daylight Time	America/Chicago
Central European Summer Time	CET
Central European Time	CET
Central Indonesia Time	Asia/Makassar
Central Standard Time	America/Chicago
ChST	Pacific/Guam
Chamorro Standard Time	Pacific/Guam
Chatham Daylight Time	NZ-CHAT
Chatham Standard Time	NZ-CHAT
Chile Summer Time	America/Santiago
Chile Time	America/Santiago
China Standard Time	Asia/Shanghai
Choibalsan Time	Asia/Choibalsan
Christmas Island Time	Indian/Christmas
Chuuk Time	Pacific/Yap
Cocos Islands Time	Indian/Cocos
Colombia Time	America/Bogota
Cook Is. Time	Pacific/Rarotonga
Coordinated Universal Time	UCT
Cuba Daylight Time	Cuba
Cuba Standard Time	Cuba
DAVT	Antarctica/Davis

DDUT	Antarctica/DumontDUrville
Davis Time	Antarctica/Davis
Dumont-d'Urville Time	Antarctica/DumontDUrville
EASST	Pacific/Easter
EAST	Pacific/Easter
EDT	America/Indiana/Indianapolis
EEST	Africa/Cairo
EGST	America/Scoresbysund
EGT	America/Scoresbysund
East Indonesia Time	Asia/Jayapura
Easter Is. Summer Time	Pacific/Easter
Easter Is. Time	Pacific/Easter
Eastern African Time	Africa/Addis_Ababa
Eastern Daylight Time	America/Indiana/Indianapolis
Eastern European Summer Time	Africa/Cairo
Eastern European Time	Africa/Cairo
Eastern Greenland Summer Time	America/Scoresbysund
Eastern Greenland Time	America/Scoresbysund
Eastern Standard Time	EST
Ecuador Time	America/Guayaquil
FJST	Pacific/Fiji
FJT	Pacific/Fiji
FKT	Atlantic/Stanley
FNT	America/Noronha
Falkland Is. Time	Atlantic/Stanley
Fernando de Noronha Time	America/Noronha
Fiji Summer Time	Pacific/Fiji
Fiji Time	Pacific/Fiji
French Guiana Time	America/Cayenne
French Southern & Antarctic Lands Time	Indian/Kerguelen
GALT	Pacific/Galapagos
GAMT	Pacific/Gambier
GET	Asia/Tbilisi
GFT	America/Cayenne
GILT	Pacific/Tarawa
GMT+00:00	GMT0

GMT+01:00	Etc/GMT-1
GMT+02:00	Etc/GMT-2
GMT+03:00	Etc/GMT-3
GMT+04:00	Etc/GMT-4
GMT+05:00	Etc/GMT-5
GMT+06:00	Etc/GMT-6
GMT+07:00	Etc/GMT-7
GMT+08:00	Etc/GMT-8
GMT+09:00	Etc/GMT-9
GMT+10:00	Etc/GMT-10
GMT+11:00	Etc/GMT-11
GMT+12:00	Etc/GMT-12
GMT+13:00	Etc/GMT-13
GMT+14:00	Etc/GMT-14
GMT-01:00	Etc/GMT+1
GMT-02:00	Etc/GMT+2
GMT-03:00	Etc/GMT+3
GMT-04:00	Etc/GMT+4
GMT-05:00	Etc/GMT+5
GMT-06:00	Etc/GMT+6
GMT-07:00	Etc/GMT+7
GMT-08:00	Etc/GMT+8
GMT-09:00	Etc/GMT+9
GMT-10:00	Etc/GMT+10
GMT-11:00	Etc/GMT+11
GMT-12:00	Etc/GMT+12
GST	Asia/Dubai
GYT	America/Guyana
Galapagos Time	Pacific/Galapagos
Gambier Time	Pacific/Gambier
Georgia Time	Asia/Tbilisi
Ghana Mean Time	Africa/Accra
Gilbert Is. Time	Pacific/Tarawa
Greenwich Mean Tim	GB
Gulf Standard Time	Asia/Dubai
Guyana Time	America/Guyana

HADT	US/Aleutian
HAST	US/Aleutian
HKT	Hongkong
HOVT	Asia/Hovd
Hawaii Standard Time	HST
Hawaii-Aleutian Daylight Time	US/Aleutian
Hawaii-Aleutian Standard Time	US/Aleutian
Hong Kong Time	Hongkong
Hovd Time	Asia/Hovd
ICT	Asia/Ho_Chi_Minh
IDT	Israel
IOT	Indian/Chagos
IRDT	Iran
IRKT	Asia/Chita
IRST	Iran
India Standard Time	Asia/Kolkata
Indian Ocean Territory Time	Indian/Chagos
Indochina Time	Asia/Ho_Chi_Minh
Iran Daylight Time	Iran
Iran Standard Time	Iran
Irish Summer Time	Eire
Irkutsk Time	Asia/Chita
Israel Daylight Time	Israel
Israel Standard Time	Israel
Japan Standard Time	Asia/Tokyo
KGT	Asia/Bishkek
KOST	Pacific/Kosrae
KRAT	Asia/Krasnoyarsk
KST	ROK
Khandyga Time	Asia/Khandyga
Kirgizstan Time	Asia/Bishkek
Korea Standard Time	ROK
Kosrae Time	Pacific/Kosrae
Krasnoyarsk Time	Asia/Krasnoyarsk
LHDT	Australia/LHI
LHST	Australia/LHI

LINT	Pacific/Kiritimati
Line Is. Time	Pacific/Kiritimati
Lord Howe Daylight Time	Australia/LHI
Lord Howe Standard Time	Australia/LHI
MAGT	Asia/Magadan
MART	Pacific/Marquesas
MAWT	Antarctica/Mawson
MDT	Navajo
MEST	MET
MHT	Kwajalein
MIST	Antarctica/Macquarie
MMT	Asia/Rangoon
MSK	W-SU
MUT	Indian/Mauritius
MVT	Indian/Maldives
MYT	Asia/Kuching
Macquarie Island Standard Time	Antarctica/Macquarie
Magadan Time	Asia/Magadan
Malaysia Time	Asia/Kuching
Maldives Time	Indian/Maldives
Marquesas Time	Pacific/Marquesas
Marshall Islands Time	Kwajalein
Mauritius Time	Indian/Mauritius
Mawson Time	Antarctica/Mawson
Middle Europe Summer Time	MET
Middle Europe Time	MET
Mideast/Riyadh87	-
Mideast/Riyadh88	-
Mideast/Riyadh89	-
Moscow Standard Time	W-SU
Mountain Daylight Time	Navajo
Mountain Standard Time	MST
Myanmar Time	Asia/Rangoon
NCT	Pacific/Noumea
NDT	America/St_Johns
NFT	Pacific/Norfolk

NOVT	Asia/Novosibirsk
NPT	Asia/Katmandu
NRT	Pacific/Nauru
NUT	Pacific/Niue
NZDT	NZ
NZST	NZ
Nauru Time	Pacific/Nauru
Nepal Time	Asia/Katmandu
New Caledonia Time	Pacific/Noumea
New Zealand Daylight Time	NZ
New Zealand Standard Time	NZ
Newfoundland Daylight Time	America/St_Johns
Newfoundland Standard Time	America/St_Johns
Niue Time	Pacific/Niue
Norfolk Time	Pacific/Norfolk
Novosibirsk Time	Asia/Novosibirsk
OMST	Asia/Omsk
ORAT	Asia/Oral
Omsk Time	Asia/Omsk
Oral Time	Asia/Oral
PDT	America/Los_Angeles
PET	America/Lima
PETT	Asia/Kamchatka
PGT	Pacific/Port_Moresby
PHOT	Pacific/Enderbury
PHT	Asia/Manila
PKT	Asia/Karachi
PMDT	America/Miquelon
PMST	America/Miquelon
PONT	Pacific/Ponape
PWT	Pacific/Palau
PYST	America/Asuncion
PYT	America/Asuncion
Pacific Daylight Time	America/Los_Angeles
Pacific Standard Time	America/Los_Angeles
Pakistan Time	Asia/Karachi

Palau Time	Pacific/Palau
Papua New Guinea Time	Pacific/Port_Moresby
Paraguay Summer Time	America/Asuncion
Paraguay Time	America/Asuncion
Peru Time	America/Lima
Petropavlovsk-Kamchatski Time	Asia/Kamchatka
Philippines Time	Asia/Manila
Phoenix Is. Time	Pacific/Enderbury
Pierre & Miquelon Daylight Time	America/Miquelon
Pierre & Miquelon Standard Time	America/Miquelon
Pitcairn Standard Time	Pacific/Pitcairn
Pohnpei Time	Pacific/Ponape
QYZT	Asia/Qyzylorda
Qyzylorda Time	Asia/Qyzylorda
RET	Indian/Reunion
ROTT	Antarctica/Rothera
Reunion Time	Indian/Reunion
Rothera Time	Antarctica/Rothera
SAKT	Asia/Sakhalin
SAMT	Europe/Samara
SAST	Africa/Maseru
SBT	Pacific/Guadalcanal
SCT	Indian/Mahe
SGT	Singapore
SRET	Asia/Srednekolymsk
SRT	America/Paramaribo
SYOT	Antarctica/Syowa
Sakhalin Time	Asia/Sakhalin
Samara Time	Europe/Samara
Samoa Standard Time	US/Samoa
Seychelles Time	Indian/Mahe
Singapore Time	Singapore
Solomon Is. Time	Pacific/Guadalcanal
South Africa Standard Time	Africa/Maseru
South Georgia Standard Time	Atlantic/South_Georgia
Srednekolymsk Time	Asia/Srednekolymsk

Suriname Time	America/Paramaribo
Syowa Time	Antarctica/Syowa
SystemV/AST4	America/Puerto_Rico
SystemV/AST4ADT	Canada/Atlantic
SystemV/CST6	Canada/Saskatchewan
SystemV/CST6CDT	US/Central
SystemV/EST5	America/Cayman
SystemV/EST5EDT	America/New_York
SystemV/HST10	US/Hawaii
SystemV/MST7	US/Arizona
SystemV/MST7MDT	America/Denver
SystemV/PST8	Pacific/Pitcairn
SystemV/PST8PDT	US/Pacific
SystemV/YST9	Pacific/Gambier
SystemV/YST9YDT	US/Alaska
TAHT	Pacific/Tahiti
TFT	Indian/Kerguelen
TJT	Asia/Dushanbe
TKT	Pacific/Fakaofo
TLT	Asia/Dili
TMT	Asia/Ashgabat
TOT	Pacific/Tongatapu
TVT	Pacific/Funafuti
Tahiti Time	Pacific/Tahiti
Tajikistan Time	Asia/Dushanbe
Timor-Leste Time	Asia/Dili
Tokelau Time	Pacific/Fakaofo
Tonga Time	Pacific/Tongatapu
Turkmenistan Time	Asia/Ashgabat
Tuvalu Time	Pacific/Funafuti
ULAT	Asia/Ulan_Bator
UYST	America/Montevideo
UYT	America/Montevideo
UZT	Asia/Tashkent
Ulaanbaatar Time	Asia/Ulan_Bator
Uruguay Summer Time	America/Montevideo

Uruguay Time	America/Montevideo
Ust-Nera Time	Asia/Ust-Nera
Uzbekistan Time	Asia/Tashkent
VET	America/Caracas
VLAT	Asia/Ust-Nera
VOST	Antarctica/Vostok
VUT	Pacific/Efate
Vanuatu Time	Pacific/Efate
Venezuela Time	America/Caracas
Vladivostok Time	Asia/Vladivostok
Vostok Time	Antarctica/Vostok
WAKT	Pacific/Wake
WAST	Africa/Windhoek
WAT	Africa/Lagos
WEST	WET
WFT	Pacific/Wallis
WGST	America/Godthab
WGT	America/Godthab
WIB	Asia/Jakarta
WIT	Asia/Jayapura
WITA	Asia/Makassar
WSDT	Pacific/Apia
WSST	Pacific/Apia
Wake Time	Pacific/Wake
Wallis & Futuna Time	Pacific/Wallis
West Indonesia Time	Asia/Jakarta
West Samoa Daylight Time	Pacific/Apia
West Samoa Standard Time	Pacific/Apia
Western African Summer Time	Africa/Windhoek
Western African Time	Africa/Lagos
Western European Summer Time	WET
Western European Time	WET
Western Greenland Summer Time	America/Godthab
Western Greenland Time	America/Godthab
XJT	Asia/Urumqi
Xinjiang Standard Time	Asia/Urumqi

YAKT	Asia/Yakutsk
YEKT	Asia/Yekaterinburg
Yakutsk Time	Asia/Yakutsk
Yekaterinburg Time	Asia/Yekaterinburg

Known Issues and Limitations in CDH 6.0.0

The following sections describe the known issues in CDH 6.0.0, grouped by component:

Operating System Known Issues

Known issues and workarounds related to operating systems are listed below.

Linux kernel security patch and CDH services crashes CVE-2017-10000364

After applying a recent Linux kernel security patch for [CVE-2017-1000364](#), CDH services that use the JSVC set of libraries crash with a Java Virtual Machine (JVM) error such as:

```
A fatal error has been detected by the Java Runtime Environment:  
SIGBUS (0x7) at pc=0x00007fe91ef6cebc, pid=30321, tid=0x00007fe930c67700
```

Cloudera services for HDFS and Impala cannot start after applying the patch.



Important: If you have not upgraded your Linux kernel using the distribution's patch for CVE-2017-1000364, do **not** apply the patch.

Commonly used Linux distributions are shown in the table below. However, the issue affects any CDH release that runs on RHEL, CentOS, Oracle Linux, SUSE Linux, or Ubuntu and that has had the Linux kernel security patch for CVE-2017-1000364 applied.

If you have already applied the patch for your OS according to the advisories for CVE-2017-1000364, apply the kernel update that contains the fix for your operating system (some of which are listed in the table). If you cannot apply the kernel update, you can workaround the issue by increasing the Java thread stack size as detailed in the steps below.

Distribution	Advisories for CVE-2017-1000364	Advisory updates
Oracle Linux 6	ELSA-2017-1486	Oracle has fixed this problem in ELSA-2017-1723 .
Oracle Linux 7	ELSA-2017-1484	Oracle has also added the fix for Oracle Linux 7 in ELBA-2017-1674 .
RHEL 6	RHSA-2017-1486	RedHat has fixed this problem for RHEL 6, marked this as outdated and superseded by RHSA-2017-1723 .
RHEL 7	RHSA-2017-1484	RedHat has fixed this problem for RHEL 7 and has marked this patch as outdated and superseded by RHBA-2017-1674 .
SLES	CVE-2017-1000364	SUSE has also fixed this problem and the patch names are included in this same advisory.

Workaround

If you cannot apply the kernel update, you can set the Java thread stack size to `-Xss1280k` for the affected services using the appropriate Java configuration option or the environment advanced configuration snippet, as detailed below.

For role instances that have specific Java configuration options properties:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.

3. Type `java` in the search field to display Java related configuration parameters. The **Java Configuration Options for Catalog Server** property field displays. Type `-Xss1280k` in the entry field, adding to any existing settings.
4. Click **Save Changes**.
5. Navigate to the HDFS service by selecting **Clusters > HDFS**.
6. Click the **Configuration** tab.
7. Click the Scope filter **DataNode**. The **Java Configuration Options for DataNode** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
8. Click **Save Changes**.
9. Select the Scope filter **NFS Gateway**. The **Java Configuration Options for NFS Gateway** field displays among the properties listed. Enter `-Xss1280k` into the field, adding to any existing properties.
10. Click **Save Changes**.
11. Restart the affected roles (or configure the safety valves in next section and restart when finished with all configurations).

For role instances that do not have specific Java configuration options:

1. Log in to Cloudera Manager Admin Console.
2. Select **Clusters > Impala**, and then click the **Configuration** tab.
3. Click the Scope filter **Impala Daemon** and Category filter **Advanced**.
4. Type `impala daemon` environment in the search field to find the safety valve entry field.
5. In the **Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

6. Click **Save Changes**.
7. Click the Scope filter **Impala StateStore** and Category filter **Advanced**.
8. In the **Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)**, enter:

```
JAVA_TOOL_OPTIONS=-Xss1280K
```

9. Click **Save Changes**.
10. Restart the affected roles.

The table below summarizes the parameters that can be set for the affected services:

Service	Settable Java Configuration Option
HDFS DataNode	Java Configuration Options for DataNode
HDFS NFS Gateway	Java Configuration Options for NFS Gateway
Impala Catalog Server	Java Configuration Options for Catalog Server
Impala Daemon	Impala Daemon Environment Advanced Configuration Snippet (Safety Valve)
	<code>JAVA_TOOL_OPTIONS=-Xss1280K</code>
Impala StateStore	Impala StateStore Environment Advanced Configuration Snippet (Safety Valve)
	<code>JAVA_TOOL_OPTIONS=-Xss1280K</code>

Cloudera Issue: CDH-55771

Leap-Second Events



Note: The [next leap-second event](#) is unknown at this time. The last leap-second event occurred on December 16, 2016 at 23:59:60 UTC.

Impact: After a leap-second event, Java applications (including CDH services) using older Java and Linux kernel versions, may consume almost 100% CPU. See <https://access.redhat.com/articles/15145>.

Leap-second events are tied to the time synchronization methods of the Linux kernel, the Linux distribution and version, and the Java version used by applications running on affected kernels.

Although Java is increasingly agnostic to system clock progression (and less susceptible to a kernel's mishandling of a leap-second event), using JDK 7 or 8 should prevent issues at the CDH level (for CDH components that use the Java Virtual Machine).

Immediate action required:

(1) Ensure that the kernel is up to date.

- **RHEL6/7, CentOS 6/7** - 2.6.32-298 or higher
- **Oracle Enterprise Linux (OEL)** - Kernels built in 2013 or later
- **SLES12** - No action required.

(2) Ensure that your Java JDKs are current (especially if the kernel is not up to date and cannot be upgraded).

- **Java 8** - No action required.

(3) Ensure that your systems use either NTP or PTP synchronization.

For systems not using time synchronization, update both the OS tzdata and Java tzdata packages to the tzdata-2016g version, at a minimum. For OS tzdata package updates, contact OS support or check updated OS repositories. For Java tzdata package updates, see Oracle's [Timezone Updater Tool](#).

Cloudera Issue: CDH-44788, TSB-189

Apache Accumulo Known Issues

Running Apache Accumulo on top of a CDH 6.0.x cluster is not currently supported. If you try to upgrade to CDH 6.0.x you will be asked to remove the Accumulo service from your cluster. Running Accumulo on top of CDH 6 will be supported in a future release.

Affected Versions: CDH 6.0.0

Cloudera Data Science Workbench

Cloudera Data Science Workbench is not supported with CDH 6.0.x. Cloudera Data Science Workbench 1.5.0 (and higher) is supported with CDH 6.1.x (and higher).

Cloudera Issue: DSE-2769

Apache Crunch Known Issues



Warning: As of CDH 6.0.0, Apache Crunch is deprecated, and will be removed in a future release. For more information, see [Deprecated Items](#) on page 667.

[Back to top](#)

Apache Flume Known Issues

Fast Replay does not work with encrypted File Channel

If an encrypted file channel is set to use fast replay, the replay will fail and the channel will fail to start.

Workaround: Disable fast replay for the encrypted channel by setting `use-fast-replay` to false.

Apache Issue: [FLUME-1885](#)

Apache Hadoop Known Issues

This page includes known issues and related topics, including:

Deprecated Properties

Several Hadoop and HDFS properties have been deprecated as of Hadoop 3.0 and later. For details, see [Deprecated Properties](#).

Hadoop Common

KMS consumes memory and increases File Descriptors

Due to an internal implementation change ([HADOOP-13597](#)), the KMS may consume a greater heap size and users may observe additional open file descriptors. Consequently, the KMS may return warnings ("Too Many File Descriptors") in its log, or may exit as the result of an `OutOfMemoryError`. Due to the extra JVM GC activities, the CDH 6 KMS may experience up to 20% throughput reduction (compared to CDH 5). In a production cluster, a KMS server may consume twice as much heap usage, and retain more than 20,000 open File Descriptors.

Workaround:

- Increase the KMS user maximum open file descriptors to 32,768 or more, and ensure the KMS runs on a dedicated host.
- Increase the KMS heap size to 4GB or more.
- Install additional KMS roles.

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Apache Issue: [HADOOP-13597](#)

Cloudera Issue: CDH-71528

Hadoop LdapGroupsMapping does not support LDAPS for self-signed LDAP server

Hadoop LdapGroupsMapping does not work with LDAP over SSL (LDAPS) if the LDAP server certificate is self-signed. This use case is currently not supported even if Hadoop User Group Mapping LDAP TLS/SSL Enabled, Hadoop User Group Mapping LDAP TLS/SSL Truststore, and Hadoop User Group Mapping LDAP TLS/SSL Truststore Password are filled properly.

Affected Versions: All CDH versions

Apache Issue: [HADOOP-12862](#)

Cloudera Issue: CDH-37926

HDFS

Deadlock occurs when rolleeditlog rpc call happens and editPendingQ is full

CDH 6 enables async edit logger by default. A known issue in the async edit logger may cause a deadlock in the NameNode.

Workaround: Disable async edit logger by setting `dfs.namenode.edits.asynclogging` to false in the following property in Cloudera Manager: NameNode Advanced Configuration Snippet (Safety Valve) for `hdfs-site.xml`.

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Apache Issue: [HDFS-13051](#)

Cloudera Issue: CDH-71921, CDH-64325

User Group Information (UGI) Ticket-Granting Ticket (TGT) renewal fails

When the UGI attempts to renew its TGT, the renewal fails with a `NullPointerException`. This issue can happen to a number of service roles, including YARN resource manager, Impala Daemon and Spark queries.

Workaround: The YARN ResourceManager and ImpalaD require a manual restart. Spark queries retry upon failure, so the issue would only be noticeable after several failures.

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Apache Issue: [HADOOP-15593](#)

OIV ReverseXML processor fails

The HDFS OIV ReverseXML processor fails if the XML file contains escaped characters.

Affected Versions: CDH 6.x

Apache Issue: [HDFS-12828](#)

UnrecoverableKeyException occurs

An UnrecoverableKeyException error occurs due to the new Enhanced KeyStore Mechanisms feature in Oracle JDK 8u171 or later. For more information, see the Oracle JDK release notes for version 8u171.

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Apache Issue: [HADOOP-15473](#)

Cannot move encrypted files to trash

With HDFS encryption enabled, you cannot move encrypted files or directories to the trash directory.

Workaround: To remove encrypted files/directories, use the following command with the `-skipTrash` flag specified to bypass trash.

```
rm -r -skipTrash /testdir
```

Affected Versions: All CDH versions

Apache Issue: [HADOOP-10902](#)

HDFS NFS gateway and CDH installation (using packages) limitation

HDFS NFS gateway works as shipped ("out of the box") only on RHEL-compatible systems, but not on SLES or Ubuntu. Because of a bug in native versions of `portmap/rpcbind`, the HDFS NFS gateway does not work out of the box on SLES or Ubuntu systems when CDH has been installed from the command-line, using packages. It does work on supported versions of RHEL-compatible systems on which `rpcbind-0.2.0-10.el6` or later is installed, and it does work if you use Cloudera Manager to install CDH, or if you start the gateway as root.

Workarounds and caveats:

- On Red Hat and similar systems, make sure `rpcbind-0.2.0-10.el6` or later is installed.
- On SLES and Ubuntu systems, do one of the following:
 - Install CDH using Cloudera Manager; or
 - Start the NFS gateway as root; or
 - [Start the NFS gateway without using packages](#); or
 - You can use the gateway by running `rpcbind` in insecure mode, using the `-i` option, but keep in mind that this allows anyone from a remote host to bind to the portmap.

Upstream Issue: [731542](#) (Red Hat), [823364](#) (SLES)

No error when changing permission to 777 on .snapshot directory

Snapshots are read-only; running `chmod 777` on the `.snapshot` directory does not change this, but does not produce an error (though other illegal operations do).

Affected Versions: All CDH versions

Apache Issue: [HDFS-4981](#)

Snapshot operations are not supported by ViewFileSystem

Affected Versions: All CDH versions

Snapshots do not retain directories' quotas settings

Affected Versions: All CDH versions

Apache Issue: [HDFS-4897](#)

Permissions for dfs.namenode.name.dir incorrectly set

Hadoop daemons should set permissions for the `dfs.namenode.name.dir` (or `dfs.name.dir`) directories to `drwx-----` (700), but in fact these permissions are set to the file-system default, usually `drwxr-xr-x` (755).

Workaround: Use `chmod` to set permissions to 700.

Affected Versions: All CDH versions

Apache Issue: [HDFS-2470](#)

`hadoop fsck -move` does not work in a cluster with host-based Kerberos

Workaround: Use `hadoop fsck -delete`

Affected Versions: All CDH versions

Apache Issue: None

Block report can exceed maximum RPC buffer size on some DataNodes

On a DataNode with a large number of blocks, the block report may exceed the maximum RPC buffer size.

Workaround: Increase the value `ipc.maximum.data.length` in `hdfs-site.xml`:

```
<property>
  <name>ipc.maximum.data.length</name>
  <value>268435456</value>
</property>
```

Affected Versions: All CDH versions

Apache Issue: None

MapReduce2 and YARN

The Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager.

When ResourceManager high availability is enabled the Standby Resource Manager redirects /jmx and /metrics requests to the Active Resource Manager. This causes the following issues in Cloudera Manager:

- If **Enable Kerberos Authentication for HTTP Web-Console** is disabled: Cloudera Manager shows statistics for the wrong server.
- If **Enable Kerberos Authentication for HTTP Web-Console** is enabled: connection from the agent to the standby fails with the `HTTPError: HTTP Error 401: Authentication required` error message. As a result, the health of the Standby Resource Manager will become bad.

Workaround: N/A

Affected Versions: CDH 6.0.x, CDH 6.1.0

Fixed Version: CDH 6.1.1

Cloudera Issue: CDH-76040

YARN's Continuous Scheduling can cause slowness in Oozie

When Continuous Scheduling is enabled in Yarn, this can cause slowness in Oozie due to long delays in communicating with Yarn. In Cloudera Manager 5.9.0 and higher, Enable Fair Scheduler Continuous Scheduler is turned off by default.

Workaround: Turn off Enable Fair Scheduler Continuous Scheduling in Cloudera Manager YARN Configuration. To keep equivalent benefits of this feature, turn on Fair Scheduler Assign Multiple Tasks.

Affected Versions: All CDH versions

Cloudera Issue: CDH-60788

JobHistory URL mismatch after server relocation

After moving the JobHistory Server to a new host, the URLs listed for the JobHistory Server on the ResourceManager web UI still point to the old JobHistory Server. This affects existing jobs only. New jobs started after the move are not affected.

Workaround: For any existing jobs that have the incorrect JobHistory Server URL, there is no option other than to allow the jobs to roll off the history over time. For new jobs, make sure that all clients have the updated `mapred-site.xml` that references the correct JobHistory Server.

Affected Versions: All CDH versions

Apache Issue: None

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165

Routable IP address required by ResourceManager

ResourceManager requires routable `host : port` addresses for `yarn.resourcemanager.scheduler.address`, and does not support using the wildcard `0.0.0.0` address.

Workaround: Set the address, in the form `host : port`, either in the client-side configuration, or on the command line when you submit the job.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-6808

Amazon S3 copy may time out

The Amazon S3 filesystem does not support renaming files, and performs a copy operation instead. If the file to be moved is very large, the operation can time out because S3 does not report progress during the operation.

Workaround: Use `-Dmapred.task.timeout=15000000` to increase the MR task timeout.

Affected Versions: All CDH versions

Apache Issue: [MAPREDUCE-972](#)

Cloudera Issue: CDH-17955

Apache HBase Known Issues

HBase suffers data loss during system recovery when a custom WAL directory is configured

[HBASE-20723](#) covers a critical data loss bug. It is present when an HBase deployment is configured to use a non-default location for storing its write-ahead-log. If `hbase.wal.dir` is set to a different location than `hbase.rootdir`, then the recovery process will mistakenly believe there are no edits to replay in the event of process failure of a region server.

Products affected: HBase

Releases affected:

- CDH 5.11.x-5.14.x
- CDH 5.15.0, 5.15.1
- CDH 6.0.0

User affected: Anyone setting the configuration value `hbase.wal.dir` to a setting other than the default. For Cloudera Manager users, this would require setting a safety valve for the `hbase-site.xml` file.

User with non-default setting can determine if they are affected by looking for INFO log messages that indicates edits have been skipped. The following is an example of such message:

```
2018-06-12 22:08:40,455 INFO [RS_LOG_REPLY_OPS-wn2-duohba:16020-0-Writer-1]
wal.WALSplitter: This region's directory doesn't exist:
hdfs://mycluster/walontest/data/default/tbl/b7fd7db5694eb71190955292b3ff7648. It is very
likely that it was already split so it's safe to discard those edits.
```

Note that the above message is normally harmless, but in this specific edge case the recovery code is looking at the incorrect location to determine region status.

Severity (Low/Medium/High): High

Impact: Data loss is unrecoverable once write-ahead-logs have been deleted as part of routine system processes. There is one exception, if the cluster uses data center replication to ship edits to another cluster and that cluster had not experienced similar data loss.

Immediate action required: Upgrade to a CDH version with the fix.

Addressed in release/refresh/patch: CDH 5.15.2 and higher, CDH 5.16.1 and higher; CDH 6.0.1 and higher

Knowledge article: For the latest update on this issue see the corresponding Knowledge article - [TSB 2019-320: HBase suffers data loss during system recovery when a custom WAL directory is configured](#)

Region Server occasionally fails when HDFS data transport encryption is enabled

In rare cases, an HBase RegionServer on a Hadoop Data Transfer Encryption enabled cluster (`dfs.encrypt.data.transfer = true`) may crash because it is not able to update the encryption key.

Workaround: Restart the RegionServer.

Affected Versions: CDH 6.0.0

Fixed Versions: 6.0.1

Apache Issue: [HBASE-21018](#)

Cloudera Issue: CDH-71613

Prefetch sometimes doesn't work with encrypted file system

If HBase prefetch is enabled (`hbase.rs.prefetchblocksonopen = true`) on an encrypted HDFS cluster, HBase RegionServer may crash due to memory corruption.

Workaround: Disable HBase prefetch (`hbase.rs.prefetchblocksonopen = false`).

Affected Versions: CDH 6.0.0

Fixed Versions: 6.0.1

Apache Issue: [HBASE-20403](#)

Cloudera Issue: CDH-68666

Hosting regions on Master is not supported

[HBASE-18511](#) added a feature that allows you to host regions on Master. However, this feature is unreliable and the performance gains are not well established, and because of this it is not supported in CDH.

Apache Issue: [HBASE-18511](#)

Cloudera Issue: CDH-53864

IOException from Timeouts

CDH 5.12.0 includes the fix [HBASE-16604](#), where the internal scanner that retries in case of IOException from timeouts could potentially miss data. Java clients were properly updated to account for the new behavior, but thrift clients will now see exceptions where the previous missing data would be.

Workaround: Create a new scanner and retry the operation when encountering this issue.

IntegrationTestReplication fails if replication does not finish before the verify phase begins

During IntegrationTestReplication, if the verify phase starts before the replication phase finishes, the test will fail because the target cluster does not contain all of the data. If the HBase services in the target cluster does not have enough memory, long garbage-collection pauses might occur.

Workaround: Use the -t flag to set the timeout value before starting verification.

Cloudera Issue: None.

HDFS encryption with HBase

Cloudera has tested the performance impact of using HDFS encryption with HBase. The overall overhead of HDFS encryption on HBase performance is in the range of 3 to 4% for both read and update workloads. Scan performance has not been thoroughly tested.

ExportSnapshot or DistCp operations may fail on the Amazon s3a:// protocol

ExportSnapshot or DistCP operations may fail on AWS when using certain JDK 8 versions, due to an incompatibility between the AWS Java SDK 1.9.x and the joda-time date-parsing module.

Workaround: Use joda-time 2.8.1 or higher, which is included in AWS Java SDK 1.10.1 or higher.

Cloudera Issue: None.

An operating-system level tuning issue in RHEL 7 causes 30% latency regressions

Workaround: Avoid using RHEL 7 if you have a latency-critical workload. For a cached workload, consider tuning the C-state (power-saving) behavior of your CPUs.

Cloudera Issue: CDH-32642

Severity: Medium

A RegionServer under extreme duress due to back-to-back garbage collection combined with heavy load on HDFS can lock up while attempting to append to the WAL

The RegionServer appears operational to ZooKeeper, and continues to host regions, but cannot complete any writes. The most obvious symptom is that all writes to regions on this RegionServer time out, and the RegionServer log shows no progress other than queuing of flushes that never complete. Log messages such as the following may occur:

```
124028 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.Sleeper: We slept  
42911ms instead of 3000ms,  
this is likely due to a long garbage collecting pause and it's usually bad, see  
http://hbase.%  
124029 2015-11-14 05:54:48,659 WARN org.apache.hadoop.hbase.util.JvmPauseMonitor: Detected  
pause in JVM or  
host machine (eg GC): pause of approximately 41110ms
```

```
1806 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2734 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#  
1807 2015-11-14 04:58:09,952 INFO org.apache.hadoop.hbase.regionserver.wal.FSHLog:  
Slow sync cost: 2963 ms, current pipeline:  
[DatanodeInfoWithStorage[10.17.198.17:20002,DS-56e2cf88-f267-43a8-b964-b29858#
```

Apache Issue: [HBASE-14374](#)

Workaround: Restart the RegionServer. To avoid the problem, adjust garbage-collection settings, give the RegionServer more RAM, and reduce the load on HDFS.

Export to Azure Blob Storage (the wasb:// or wasbs:// protocol) is not supported

CDH 5.3 and higher supports Azure Blob Storage for some applications. However, a null pointer exception occurs when you specify a wasb:// or wasbs:// location in the --copy-to option of the ExportSnapshot command or as the output directory (the second positional argument) of the Export command.

Workaround: None.

Apache Issue: [HADOOP-12717](#)

AccessController postOperation problems in asynchronous operations

When security and Access Control are enabled, the following problems occur:

- If a Delete Table fails for a reason other than missing permissions, the access rights are removed but the table may still exist and may be used again.
- If hbaseAdmin.modifyTable() is used to delete column families, the rights are not removed from the Access Control List (ACL) table. The postOperation is implemented only for postDeleteColumn().
- If Create Table fails, full rights for that table persist for the user who attempted to create it. If another user later succeeds in creating the table, the user who made the failed attempt still has the full rights.

Workaround: None

Apache Issue: [HBASE-6992](#)

Apache Hive / HCatalog / Hive on Spark Known Issues

This topic also contains:

- [HCatalog Known Issues](#)
- [Hive on Spark Known Issues](#)

Hive Jobs Are Submitted to a Single Queue When Sentry is Deployed

Hive jobs are not submitted into the correct YARN queue when Hive is using Sentry because Hive does not use the YARN API to resolve the user or group of the job's original submitter. This causes the job to be placed in a queue using the placement rules based on the Hive user. The HiveServer2 fair scheduler queue mapping used for "non-impersonation" mode does not handle the primary-secondary queue mappings correctly.

Affected Version: CDH 6.0.0 and higher

Workaround: If you are a Hive and Sentry user, do not upgrade to CDH 6.0.0 or 6.0.1. This issue will be fixed as soon as possible. If you must use Hive and Sentry in CDH 6.0.0 or 6.0.1, see [YARN Dynamic Resource Pools Do Not Work with Hive When Sentry Is Enabled](#) for additional workarounds.

When vectorization is enabled on any file type (ORC, Parquet) queries that divide by zero using the modulo operator (%) return an error

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query divides by zero using the modulo operator (%), it returns the following error: Arithmetic exception [divide by] 0. For example, if you run the following query this issue is triggered: SELECT 100 % column_c1 FROM table_t1; and the value in column_c1 is zero. The divide operator (/) is not affected by this issue.

Workaround: Disable vectorization for the query that is triggering this at either the session level by using the SET statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-19564](#)

Cloudera Issue: CDH-71211

When vectorization is enabled for Hive on any file type (ORC, Parquet) queries that perform comparisons in the SELECT clause on large values in columns with the data type of BIGINT might return wrong results

When vectorization is enabled for Hive on any file type, including ORC and Parquet, if the query performs a comparison operation between very large values in columns that are BIGINT data types in the SELECT clause of the query, incorrect results might be returned. Comparison operators include ==, !=, <, <=, >, and >=. This issue does not occur when the comparison operation is performed in the filtering clause of the query. This issue can also occur when the difference of values in such columns is out of range for a LONG (64-bit) data type. For example, if column_c1 stores 8976171455044006767 and column_c2 stores -7272907770454997143, a query such as `SELECT column_c1 < column_c2 FROM table_test` returns true instead of false because the difference (8976171455044006767 - (-7272907770454997143)) is 1.6249079225499E19 which is greater than 9.22337203685478E18, which is the maximum possible value that a LONG (64-bit) data type can hold.

Affected Versions: When query vectorization is enabled for Hive, this issue affects Hive ORC tables in all versions of CDH and affects Hive Parquet tables in CDH 6.0 and later

Apache Issue: [HIVE-20207](#)

Cloudera Issue: CDH-70996

Workaround: Use a DECIMAL type instead of BIGINT for columns that might contain very large values. Another option is to disable vectorization for the query that is triggering this at either the session level by using the `SET` statement or at the server level by disabling the property with Cloudera Manager. For information about how to enable or disable query vectorization, see [Enabling Hive Query Vectorization](#).

Specified column position in the ORDER BY clause is not supported for SELECT * queries

When column positions are specified in ORDER BY clauses, they are not honored for SELECT * queries and an error is returned as shown in the following example:

```
CREATE TABLE decimal_1 (id decimal(5,0));
SELECT * FROM decimal_1 ORDER BY 1 limit 100;
Error while compiling statement: FAILED: SemanticException [Error 10219]: Position in
ORDER BY is not supported when using SELECT *
```

Instead the query must list out the columns it is selecting.

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-68550

DirectSQL with PostgreSQL

Hive doesn't support Hive direct SQL queries with PostgreSQL database. It only supports this feature with MySQL, MariaDB, and Oracle. With PostgresSQL, direct SQL is disabled as a precaution, since there have been issues reported upstream where it is not possible to fallback on DataNucleus in the event of some failures, plus other non-standard behaviors. For more information, see [Hive Configuration Properties](#).

Affected Versions: All CDH versions

Cloudera Issue: CDH-49017

ALTER PARTITION ... SET LOCATION does not work on Amazon S3 or between S3 and HDFS

Cloudera recommends that you do not use ALTER PARTITION ... SET LOCATION on S3 or between S3 and HDFS. The rest of the ALTER PARTITION commands work as expected.

Affected Versions: All CDH versions

Cloudera Issue: CDH-42420

Commands run against an Oracle-backed metastore might fail

Commands run against an Oracle-backed Metastore fail with error:

```
javax.jdo.JDODataStoreException Incompatible data type for column TBLS.VIEW_EXPANDED_TEXT
: was CLOB (datastore),
```

but type expected was LONGVARCHAR (metadata). Please check that the type in the datastore and the type specified in the MetaData are consistent.

This error might occur if the metastore is run on top of an Oracle database with the configuration property `datanucleus.validateColumns` set to true.

Workaround: Set `datanucleus.validateColumns=false` in the `hive-site.xml` configuration file.

Affected Versions: All CDH versions

Cannot create archive partitions with external HAR (Hadoop Archive) tables

`ALTER TABLE ... ARCHIVE PARTITION` is not supported on external tables.

Affected Versions: All CDH versions

Cloudera Issue: CDH-9638

Object types `Server` and `URI` are not supported in "SHOW GRANT ROLE `roleName` on OBJECT `objectName`" statements

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Cloudera Issue: CDH-19430

HCatalog Known Issues



Note: As of CDH 5, HCatalog is part of Apache Hive.

There are no notable known issues in this release of HCatalog.

[Back to top](#)

Hive on Spark (HoS) Known Issues

Hive on Spark queries fail with "Timed out waiting for client to connect" for an unknown reason

If this exception is preceded by logs of the form "client.RpcRetryingCaller: Call exception...", then this failure is due to an unavailable HBase service. On a secure cluster, `spark-submit` will try to obtain delegation tokens from HBase, even though Hive on Spark might not need them. So if HBase is unavailable, `spark-submit` throws an exception.

Workaround: Fix the HBase service, or set `spark.yarn.security.tokens.hbase.enabled` to false.

Affected Versions: CDH 5.7.0 and higher

Cloudera Issues: CDH-59591, CDH-59599

Hue Known Issues

Hue does not support the Spark App

Hue does not currently support the Spark application.

Connecting to PostgreSQL Database Fails with Error "No module named `psycopg2`"

When configuring Hue to use a PostgreSQL database, the connection fails with the following error:

Error loading `psycopg2` module: No module named `psycopg2`

Workaround: Install the `psycopg2` Python package as documented in [Installing the psycopg2 Python Package](#).

Affected Versions: All CDH 6 versions

Fixed Versions: None

Apache Issue: N/A

Cloudera Issue: CDH-65804

Apache Impala Known Issues

The following sections describe known issues and workarounds in Impala, as of the current production release. This page summarizes the most serious or frequently encountered issues in the current release, to help you make planning decisions about installing and upgrading. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.



Note: The online issue tracking system for Impala contains comprehensive information and is updated in real time. To verify whether an issue you are experiencing has already been reported, or which release an issue is fixed in, search on the [Impala JIRA tracker](#).

Impala Known Issues: Startup

These issues can prevent one or more Impala-related daemons from starting properly.

Impala requires FQDN from hostname command on kerberized clusters

The method Impala uses to retrieve the host name while constructing the Kerberos principal is the `gethostname()` system call. This function might not always return the fully qualified domain name, depending on the network configuration. If the daemons cannot determine the FQDN, Impala does not start on a kerberized cluster.

Workaround: Test if a host is affected by checking whether the output of the `hostname` command includes the FQDN. On hosts where `hostname` only returns the short name, pass the command-line flag `--hostname=fully_qualified_domain_name` in the startup options of all Impala-related daemons.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4978](#)

Impala Known Issues: Crashes and Hangs

These issues can cause Impala to quit or become unresponsive.

Unable to view large catalog objects in catalogd Web UI

In `catalogd` Web UI, you can list metadata objects and view their details. These details are accessed via a link and printed to a string formatted using thrift's `DebugProtocol`. Printing large objects (> 1 GB) in Web UI can crash `catalogd`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6841](#)

Impala Known Issues: Performance

These issues involve the performance of operations such as queries or DDL statements.

Metadata operations block read-only operations on unrelated tables

Metadata operations that change the state of a table, like `COMPUTE STATS` or `ALTER RECOVER PARTITIONS`, may delay metadata propagation of unrelated unloaded tables triggered by statements like `DESCRIBE` or `SELECT` queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6671](#)

Slow queries for Parquet tables with `convert_legacy_hive_parquet_utc_timestamps=true`

The configuration setting `convert_legacy_hive_parquet_utc_timestamps=true` uses an underlying function that can be a bottleneck on high volume, highly concurrent queries due to the use of a global lock while loading time zone information. This bottleneck can cause slowness when querying Parquet tables, up to 30x for scan-heavy queries. The amount of slowdown depends on factors such as the number of cores and number of threads involved in the query.

**Note:**

The slowdown only occurs when accessing `TIMESTAMP` columns within Parquet files that were generated by Hive, and therefore require the on-the-fly timezone conversion processing.

Workaround: Store the `TIMESTAMP` values as strings in one of the following formats:

- `YYYY-MM-dd`
- `YYYY-MM-dd HH:mm:ss`
- `YYYY-MM-dd HH:mm:ss.SSSSSSSSSS`

The date can have the 1-9 digits in the fractional part.

Impala implicitly converts such string values to `TIMESTAMP` in calls to date/time functions.

Affected Versions: All CDH 6 versions

Fixed Versions: CDH 6.1.0

Apache Issue: [IMPALA-3316](#)

Impala Known Issues: JDBC and ODBC Drivers

These issues affect applications that use the JDBC or ODBC APIs, such as business intelligence tools or custom-written applications in languages such as Java or C++.

Impala Known Issues: Security

These issues relate to security features, such as Kerberos authentication, Sentry authorization, encryption, auditing, and redaction.

In Impala with Sentry enabled, `REVOKE ALL ON SERVER` does not remove the privileges granted with the `GRANT` option

If you grant a role the `ALL` privilege at the `SERVER` scope with the `WITH GRANT OPTION` clause, you cannot revoke the privilege. Although the `SHOW GRANT ROLE` command will show that the privilege has been revoked immediately after you run the command, the `ALL` privilege will reappear when you run the `SHOW GRANT ROLE` command after Sentry refreshes.

Immediate Action Required: Once the privilege has been granted, the only way to remove it is to delete the role.

Affected Versions: CDH 6.0.0, CDH 6.0.1, CDH 5.15.0, CDH 5.15.1, CDH 5.14.x and all prior releases

Fixed Versions: CDH 6.1.0, CDH 6.0.2, CDH 5.16.0, CDH 5.15.2

Cloudera Issue: TSB-341

Impala does not support Heimdal Kerberos

Heimdal Kerberos is not supported in Impala.

Apache Issue: [IMPALA-7072](#)

Affected Versions: All CDH 6 versions

Impala Known Issues: Resources

These issues involve memory or disk usage, including out-of-memory conditions, the spill-to-disk feature, and resource management features.

Handling large rows during upgrade to CDH 5.13 / Impala 2.10 or higher

After an upgrade to CDH 5.13 / Impala 2.10 or higher, users who process very large column values (long strings), or have increased the `--read_size` configuration setting from its default of 8 MB, might encounter capacity errors for some queries that previously worked.

Resolution: After the upgrade, follow the instructions in [Handling Large Rows During Upgrade to CDH 5.13 / Impala 2.10 or Higher](#) to check if your queries are affected by these changes and to modify your configuration settings if so.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-6028](#)

Configuration to prevent crashes caused by thread resource limits

Impala could encounter a serious error due to resource usage under very high concurrency. The error message is similar to:

```
F0629 08:20:02.956413 29088 llvm-codegen.cc:111] LLVM hit fatal error: Unable to allocate section memory!
terminate called after throwing an instance of
'boost::exception_detail::clone_impl<boost::exception_detail::error_info_injector<boost::thread_resource_error>'>
```

Workaround:

In CDH 6.0 and lower versions of CDH, configure each host running an `impalad` daemon with the following settings:

```
echo 2000000 > /proc/sys/kernel/threads-max
echo 2000000 > /proc/sys/kernel/pid_max
echo 8000000 > /proc/sys/vm/max_map_count
```

In CDH 6.1 and higher versions, it is unlikely that you will hit the thread resource limit. Configure each host running an `impalad` daemon with the following setting:

```
echo 8000000 > /proc/sys/vm/max_map_count
```

To make the above settings durable, refer to your OS documentation. For example, on RHEL 6.x:

1. Add the following line to `/etc/sysctl.conf`:

```
vm.max_map_count=8000000
```

2. Run the following command:

```
sysctl -p
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-5605](#)

Breakpad minidumps can be very large when the thread count is high

The size of the breakpad minidump files grows linearly with the number of threads. By default, each thread adds 8 KB to the minidump size. Minidump files could consume significant disk space when the daemons have a high number of threads.

Workaround: Add `--minidump_size_limit_hint_kb=size` to set a soft upper limit on the size of each minidump file. If the minidump file would exceed that limit, Impala reduces the amount of information for each thread from 8 KB to 2 KB. (Full thread information is captured for the first 20 threads, then 2 KB per thread after that.) The minidump file can still grow larger than the "hinted" size. For example, if you have 10,000 threads, the minidump file can be more than 20 MB.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3509](#)

Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the `impalad` daemon.

Workaround: To monitor overall memory usage, use the `top` command, or add the memory figures in the Impala web UI `/memz` tab to JVM memory usage shown on the `/metrics` tab.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-691](#)

Impala Known Issues: Correctness

These issues can cause incorrect or unexpected results from queries. They typically only arise in very specific circumstances.

Incorrect result due to constant evaluation in query with outer join

An OUTER JOIN query could omit some expected result rows due to a constant such as FALSE in another join clause.

For example:

```
explain SELECT 1 FROM alltypestiny a1
    INNER JOIN alltypesagg a2 ON a1.smallint_col = a2.year AND false
    RIGHT JOIN alltypes a3 ON a1.year = a3.bigint_col;
+-----+
| Explain String
+-----+
| Estimated Per-Host Requirements: Memory=1.00KB Vcores=1
| 00:EMPTYSET
+-----+
```

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-3094](#)

BST between 1972 and 1995

The calculation of start and end times for the BST (British Summer Time) time zone could be incorrect between 1972 and 1995. Between 1972 and 1995, BST began and ended at 02:00 GMT on the third Sunday in March (or second Sunday when Easter fell on the third) and fourth Sunday in October. For example, both function calls should return 13, but actually return 12, in a query such as:

```
select
    extract(from_utc_timestamp(cast('1970-01-01 12:00:00' as timestamp), 'Europe/London'),
    "hour") summer70start,
    extract(from_utc_timestamp(cast('1970-12-31 12:00:00' as timestamp), 'Europe/London'),
    "hour") summer70end;
```

Affected Versions: All CDH 6 versions

Fixed Versions: CDH 6.1

Apache Issue: [IMPALA-3082](#)

% escaping does not work correctly in a LIKE clause

If the final character in the RHS argument of a LIKE operator is an escaped \% character, it does not match a % final character of the LHS argument.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2422](#)

Crash: impala::Coordinator::ValidateCollectionSlots

A query could encounter a serious error if includes multiple nested levels of INNER JOIN clauses involving subqueries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2603](#)

Impala Known Issues: Metadata

These issues affect how Impala interacts with metadata. They cover areas such as the metastore database and the Impala Catalog Server daemon.

Concurrent catalog operations with heavy DDL workloads can cause queries with SYNC_DDL to fail fast

When Catalog Server is under a heavy load with concurrent catalog operations of long running DDLs, queries running with the SYNC_DDL query option can fail with the following message:

```
ERROR: CatalogException: Couldn't retrieve the catalog topic  
version for the SYNC_DDL operation after 3 attempts. The operation has  
been successfully executed but its effects may have not been  
broadcast to all the coordinators.
```

The catalog operation is actually successful as the change has been committed to HMS and Catalog Server cache, but when Catalog Server notices a longer than expected time for it to broadcast the changes, it fails fast.

The coordinator daemons eventually sync up in the background.

Affected Versions: CDH versions 6.0 and 6.1

Apache Issue: [IMPALA-7961](#) / CDH-76345

Impala Known Issues: Interoperability

These issues affect the ability to interchange data between Impala and other systems. They cover areas such as data types and file formats.

Queries Stuck on Failed HDFS Calls and not Timing out

If the following error appears multiple times in a short duration while running a query, it would mean that the connection between the impalad and the HDFS NameNode is in a bad state and hence the impalad would have to be restarted:

```
"hdfsOpenFile() for <filename> at backend <hostname:port> failed to finish before the  
<hdfs_operation_timeout_sec> second timeout "
```

Workaround: Restart the impalad in the bad state.

Affected Versions: All versions of Impala

Apache Issue: [HADOOP-15720](#)

Deviation from Hive behavior: Out of range values float/double values are returned as maximum allowed value of type (Hive returns NULL)

Impala behavior differs from Hive with respect to out of range float/double values. Out of range values are returned as maximum allowed value of type (Hive returns NULL).

Workaround: None

Affected Versions: All CDH 6 versions

Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink hdfs.writeFormat must be set to Text, rather than its default value of Writable. The hdfs.writeFormat setting must be changed to Text before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

Resolution: This information has been requested to be added to the upstream Flume documentation.

Affected Versions: All CDH 6 versions

Cloudera Issue: CDH-13199

Avro Scanner fails to parse some schemas

The default value in Avro schema must match the first union type. For example, if the default value is null, then the first type in the UNION must be "null".

Workaround: Swap the order of the fields in the schema specification. For example, use ["null", "string"] instead of ["string", "null"]. Note that the files written with the problematic schema must be rewritten with the new schema because Avro files have embedded schemas.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-635](#)

Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

Workaround: Remove trailing semicolon from the Avro schema.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1024](#)

Incorrect results with basic predicate on CHAR typed column

When comparing a `CHAR` column value to a string literal, the literal value is not blank-padded and so the comparison might fail when it should match.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1652](#)

Impala Known Issues: Limitations

These issues are current limitations of Impala that require evaluation as you plan how to integrate Impala into your data management workflow.

Set limits on size of expression trees

Very deeply nested expressions within queries can exceed internal Impala limits, leading to excessive memory usage.

Workaround: Avoid queries with extremely large expression trees. Setting the query option `disable_codegen=true` may reduce the impact, at a cost of longer query runtime.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-4551](#)

Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

Workaround: Use standard HDFS on all Impala nodes.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-77](#)

Hue and BDR require separate parameters for Impala Load Balancer

Cloudera Manager supports a single parameter for specifying the Impala Daemon Load Balancer. However, because BDR and Hue need to use different ports when connecting to the load balancer, it is not possible to configure the load balancer value so that BDR and Hue will work correctly in the same cluster.

Workaround: To configure BDR with Impala, use the load balancer configuration either without a port specification or with the Beeswax port.

To configure Hue, use the **Hue Server Advanced Configuration Snippet (Safety Valve)** for `impalad_flags` to specify the load balancer address with the HiveServer2 port.

Affected Versions: CDH versions from 5.11 to 6.0.1

Cloudera Issue: [OPSAPS-46641](#)

Impala Known Issues: Miscellaneous / Older Issues

These issues do not fall into one of the above categories or have not been categorized yet.

A failed CTAS does not drop the table if the insert fails

If a `CREATE TABLE AS SELECT` operation successfully creates the target table but an error occurs while querying the source table or copying the data, the new table is left behind rather than being dropped.

Workaround: Drop the new table manually after a failed `CREATE TABLE AS SELECT`.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-2005](#)

Casting scenarios with invalid/inconsistent results

Using a `CAST` function to convert large literal values to smaller types, or to convert special values such as `Nan` or `Inf`, produces values not consistent with other database systems. This could lead to unexpected results from queries.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-1821](#)

Impala Parser issue when using fully qualified table names that start with a number

A fully qualified table name starting with a number could cause a parsing error. In a name such as `db.571_market`, the decimal point followed by digits is interpreted as a floating-point number.

Workaround: Surround each part of the fully qualified name with backticks (`` ``).

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-941](#)

Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

Workaround: Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) for details about modifying these environment settings.

Resolution: Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

Affected Versions: All CDH 6 versions

Apache Issue: [IMPALA-532](#)

EMC Isilon Known Issues

CDH 6.0 is not currently supported on EMC Isilon.

Affected Versions: CDH 6.0.0

Apache Kafka Known Issues

Topics Created with the "kafka-topics" Tool Might Not Be Secured

Topics that are created and deleted via Kafka are secured (for example, auto created topics). However, most topic creation and deletion is done via the `kafka-topics` tool, which talks directly to ZooKeeper or some other third-party tool that talks directly to ZooKeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. Anyone with access to ZooKeeper can create and delete topics. They will not be able to describe, read, or write to the topics even if they can create them.

The following commands talk directly to ZooKeeper and therefore are not secured via Kafka:

- `kafka-topics.sh`
- `kafka-configs.sh`
- `kafka-preferred-replica-election.sh`
- `kafka-reassign-partitions.sh`

"offsets.topic.replication.factor" Must Be Less Than or Equal to the Number of Live Brokers

The `offsets.topic.replication.factor` broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a `GROUP_COORDINATOR_NOT_AVAILABLE` error until the cluster size meets this replication factor requirement.

Kafka May Be Stuck with Under-replicated Partitions after ZooKeeper Session Expires

This problem can occur when your Kafka cluster includes a large number of under-replicated Kafka partitions. One or more broker logs include messages such as the following:

```
[2016-01-17 03:36:00,888] INFO Partition [__samza_checkpoint_event-creation_1,3] on broker 3: Shrinking ISR for partition [__samza_checkpoint_event-creation_1,3] from 6,5 to 5 (kafka.cluster.Partition)
[2016-01-17 03:36:00,891] INFO Partition [__samza_checkpoint_event-creation_1,3] on broker 3: Cached zkVersion [66] not equal to that in zookeeper, skip updating ISR (kafka.cluster.Partition)
```

There will also be an indication of the ZooKeeper session expiring in one or more Kafka broker logs around the same time as the previous errors:

```
INFO zookeeper state changed (Expired) (org.I0Itec.zkclient.ZkClient)
```

The log is typically in `/var/log/kafka` on each host where a Kafka broker is running. The location is set by the property `kafka.log4j.dir` in Cloudera Manager. The log name is `kafka-broker-hostname.log`. In diagnostic bundles, the log is under `logs/hostname-ip-address/`.

Workaround: To move forward after seeing this problem, restart the affected Kafka brokers. You can restart individual brokers from the **Instances** tab in the Kafka service page in Cloudera Manager.



Note: If restarting the brokers does not resolve the problem, you might not have this issue; see [KAFKA-3083 A soft failure in controller may leave a topic partition in an inconsistent state](#). This problem also involves the ZooKeeper session expiring, but will not involve the error message with `Cached zkVersion [XX] not equal to that in zookeeper`.

To reduce the chances of this issue happening again, do what you can to make sure ZooKeeper sessions do not expire:

- Reduce the potential for long garbage collection pauses by brokers:
 - Use a better garbage collection mechanism in the JVM, such as G1GC. You can do this by adding `-XX:+UseG1GC` in the `broker_java_opts`.
 - Increase broker heap size if it is too small (`broker_max_heap_size`). Be careful that you don't choose a heap size that can cause out-of-memory problems given all the services running on the node.
- Increase the ZooKeeper session timeout configuration on brokers (`zookeeper.session.timeout.ms`), to reduce the likelihood that sessions expire.
- Ensure ZooKeeper itself is well resourced and not overwhelmed so it can respond. For example, it is highly recommended to locate the ZooKeeper log directory on its own disk.

Affected Versions: CDK 1.4.x, 2.0.x, 2.1.x, 2.2.x

Fixed Versions:

- **Full Fix:** CDH 6.1.0
- **Partial Fix:** CDH 6.0.0, Kafka implementations with CDH 6.0.0 are less likely to encounter this issue.

Apache Issue: [KAFKA-2729](#)

Cloudera Issue: CDH-42514

Requests Fail When Sending to a Nonexistent Topic with "auto.create.topics.enable" Set to True

The first few `produce` requests fail when sending to a nonexistent topic with `auto.create.topics.enable` set to true.

Workaround: Increase the number of retries in the Producer configuration setting `retries`.

Custom Kerberos Principal Names Cannot Be Used for Kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start.

Workaround: None. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

Performance Degradation When SSL Is Enabled

Significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, and message size. Consumers are typically more affected than producers.

Workaround: Configure brokers and clients with `ssl.secure.random.implementation = SHA1PRNG`. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Affected Versions: CDK 2.x and later

Fixed Versions: None

Apache Issue: [KAFKA-2561](#)

Cloudera Issue: None

Apache Kudu Known Issues

The following are known bugs and issues in Kudu. Note that this list is not exhaustive, and is meant to communicate only the most important known issues.

CFile Checksum Failure Causes Queries to Fail

When a CFile checksum fails, for example, due to a underlying disk corruption, queries against the replica will fail with an error message, such as this:

```
Unable to advance iterator: Corruption: checksum error on CFile block
```

Workaround: Remove the corrupted replica from the tablet's Raft configuration. See [Kudu Troubleshooting Guide](#) for the detailed steps.

Affected Versions: CDH 6.0.x and lower

Apache Issue: [KUDU-2469](#)

C++ Client Fails to Re-acquire Authentication Token in Multi-master Clusters

A security-related issue can cause Impala queries to start failing on busy clusters in the following scenario:

- The cluster runs with the `--rpc_authentication` set as optional or required. The default is optional. Secure clusters use required.
- The cluster is using multiple masters.
- Impala queries happen frequently enough that the leader master connection to some `impalad` isn't idle-closed (more than 1 query per 65 seconds).
- The connection stays alive for longer than the authentication token timeout (1 week by default).
- A master leadership change occurs after the authentication token expiration.

Impala queries will start failing with errors in the `impalad` logs like:

```
I0904 13:53:08.748968 95857 client-internal.cc:283] Unable to determine the new leader
Master: Not authorized: Client connection negotiation failed: client connection to
10.164.44.13:7051: FATAL_INVALID_AUTHENTICATION_TOKEN: Not authorized: authentication
token expired
I0904 13:53:10.389009 95861 status.cc:125] Unable to open Kudu table: Timed out:
GetTableSchema timed out after deadline expired
@ 0x95b1e9 impala::Status::Status()
@ 0xff22d4 impala::KuduScanNodeBase::Open()
@ 0xff101e impala::KuduScanNode::Open()
@ 0xb73ced impala::FragmentInstanceState::Open()
@ 0xb7532b impala::FragmentInstanceState::Exec()
@ 0xb64ae8 impala::QueryState::ExecFInstance()
@ 0xd15193 impala::Thread::SuperviseThread()
@ 0xd158d4 boost::detail::thread_data<>::run()
```

```

@ 0x129188a (unknown)
@ 0x7f717ceade25 start_thread
@ 0x7f717cbdb34d __clone

```

Impala shell queries will fail with a message like:

```
Unable to open Kudu table: Timed out: GetTableSchema timed out after deadline expired
```

Workaround:

- Restart the affected Impala Daemons. Restarting a daemon ensures the problem will not reoccur for at least the authentication token lifetime, which defaults to one week.
- Increase the authentication token lifetime (`--authn_token_validity_seconds`). Beware that raising this lifetime increases the window of vulnerability of the cluster if a client is compromised. It is recommended that you keep the token lifetime at one month maximum for a secure cluster. For unsecured clusters, a longer token lifetime is acceptable, and a 3 month lifetime is recommended.

Affected Versions: From CDH 5.11 through CDH 6.0.1

Apache Issue: [KUDU-2580](#)

Timeout Possible with Log Force Synchronization Option

If the Kudu master is configured with the `-log_force_fsync_all` option, tablet servers and clients will experience frequent timeouts, and the cluster may become unusable.

Affected Versions: All CDH 6 versions

Longer Startup Times with a Large Number of Tablets

If a tablet server has a very large number of tablets, it may take several minutes to start up. It is recommended to limit the number of tablets per server to 1000 or fewer. The maximum allowed number of tablets is 2000 per server. Consider this limitation when pre-splitting your tables. If you notice slow start-up times, you can monitor the number of tablets per server in the web UI.

Affected Versions: All CDH 6 versions

Descriptions for Kudu TLS/SSL Settings in Cloudera Manager

Use the descriptions in the following table to better understand the TLS/SSL settings in the Cloudera Manager Admin Console.

Field	Usage Notes
Kerberos Principal	Set to the default principal, <code>kudu</code> .
Enable Secure Authentication And Encryption	Select this checkbox to enable authentication <i>and</i> RPC encryption between all Kudu clients and servers, as well as between individual servers. Only enable this property after you have configured Kerberos.
Master TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu master host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to the Kudu master web UI.
Tablet Server TLS/SSL Server Private Key File (PEM Format)	Set to the path containing the Kudu tablet server host's private key (PEM-format). This is used to enable TLS/SSL encryption (over HTTPS) for browser-based connections to Kudu tablet server web UIs.
Master TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu master host's private key (set in Master TLS/SSL Server Private Key File). The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Tablet Server TLS/SSL Server Certificate File (PEM Format)	Set to the path containing the signed certificate (PEM-format) for the Kudu tablet server host's private key (set in Tablet Server TLS/SSL Server Private Key File).

Field	Usage Notes
	The certificate file can be created by concatenating all the appropriate root and intermediate certificates required to verify trust.
Master TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Tablet Server TLS/SSL Server CA Certificate (PEM Format)	Disregard this field.
Enable TLS/SSL for Master Server	Enables HTTPS encryption on the Kudu master web UI.
Enable TLS/SSL for Tablet Server	Enables HTTPS encryption on the Kudu tablet server web UIs.

Affected Versions: All CDH 6 versions

Apache Oozie Known Issues

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used

Oozie database upgrade fails when PostgreSQL version 9.6 or higher is used due to a sys table change in PostgreSQL from version 9.5 to 9.6. The failure only happens if Oozie uses a JDBC driver earlier than 9.4.1209.

Workaround:

1. After the parcels of the new version are distributed, replace the PostgreSQL JDBC driver with a newer one (version 9.4.1209 or higher) in the new parcel, at the following locations:
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/lib/
 - /opt/cloudera/parcels/\${newparcel.version}/lib/oozie/libtools/
2. Perform the upgrade.



Note: If you already started the upgrade and the process stops with an error message about missing columns, you can change the drivers at that point of the process as well, and resume the upgrade.

If your cluster is installed from packages, you must change the drivers at the following locations:

- /usr/lib/oozie/libtools/
- /usr/lib/oozie/lib/



Note: You can change the driver after the packages installation, but before running the CDH upgrade wizard. You can also do it during the update process, when the error occurs.

You can download the driver from the [PostgreSQL JDBC driver homepage](#).

Affected Versions: CDH 6.0.0 and higher

Cloudera Issue: CDH-75951

External ID of MapReduce action not filled properly and failing MR job treated as SUCCEEDED

When a MapReduce action is launched from Oozie, the external ID field is not filled properly. It gets populated with the YARN ID of the LauncherAM, not with the ID of the actual MR job. If the MR job is submitted successfully and then fails, it will be treated as a successfully executed action.

Affected Versions: CDH 6.0.0 and higher

Fixed Version: CDH 6.1.0 and higher

Apache Issue: [OOZIE-3298](#)

Oozie jobs fail (gracefully) on secure YARN clusters when JobHistory server is down

If the JobHistory server is down on a YARN (MRv2) cluster, Oozie attempts to submit a job, by default, three times. If the job fails, Oozie automatically puts the workflow in a SUSPEND state.

Workaround: When the JobHistory server is running again, use the `resume` command to tell Oozie to continue the workflow from the point at which it left off.

Affected Versions: CDH 5 and higher

Cloudera Issue: CDH-14623

Oozie works with MapReduce or YARN, but not both

The Oozie server works with a MapReduce (MRv1) cluster or a YARN (MRv2) cluster, but not both at the same time.

Workaround: Use two different Oozie servers.

Affected Versions: All

[Apache Parquet Known Issues](#)

There are no known issues in Parquet.

[Apache Pig Known Issues](#)

There are no known issues in this release.

[Cloudera Search Known Issues](#)

The current release includes the following known limitations:

CDH Upgrade fails to delete Solr data from HDFS

The CDH upgrade process fails to delete Solr data from HDFS and the recreated collections fail to be initialized due to the existing indexes.

Workaround: Perform the following steps *after* you run the CDH Upgrade wizard and *before* you finalize the HDFS upgrade:

1. Log in to the Cloudera Manager Admin Console.
2. Go to the Solr service page.
3. Stop the Solr service and dependent services. Click **Actions > Stop**.
4. Click **Actions > Reinitialize Solr State for Upgrade**.
5. Click **Actions > Bootstrap Solr Configuration**.
6. Start the Solr and dependent services. Click **Actions > Start**.
7. Click **Actions > Bootstrap Solr Collections**.

Affected Versions: CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47502

Solr Service reports stale configurations even after restart

Solr reports stale configurations, and the Solr Server role fails to start with the following error: Role failed to start due to error: The archive already contains `creds.localjceks`. The issue occurs if your deployment has Solr and HDFS uses LDAP Group Mapping.

Workaround: If you have a CDH 5 cluster and use LDAP Group Mapping, do not upgrade to CDH 6.0.0. If you have a CDH 6.0.0 cluster, disable LDAP Group Mappings.

Affected Versions: Cloudera Manager 6.0.0 and CDH 6.0.0

Fixed Versions: Cloudera Manager 6.0.1

Cloudera Issue: OPSAPS-47321

Solr SQL, Graph, and Stream Handlers are Disabled if Collection Uses Document-Level Security

The Solr SQL, Graph, and Stream handlers do not support document-level security, and are disabled if document-level security is enabled on the collection. If necessary, these handlers can be re-enabled by setting the following Java system properties, but document-level security is not enforced for these handlers:

- SQL: `solr.sentry.enableSqlQuery=true`
- Graph: `solr.sentry.enableGraphQuery=true`
- Stream: `solr.sentry.enableStreams=true`

Workaround: None

Affected Versions: All CDH 6 releases

Cloudera Issue: CDH-66345

Collection Creation No Longer Supports Automatically Selecting A Configuration If Only One Exists

Before CDH 5.5.0, a collection could be created without specifying a configuration. If no `-c` value was specified, then:

- If there was only one configuration, that configuration was chosen.
- If the collection name matched a configuration name, that configuration was chosen.

Search for CDH 5.5.0 includes multiple built-in configurations. As a result, there is no longer a case in which only one configuration can be chosen by default.

Workaround: Explicitly specify the collection configuration to use by passing `-c <configName>` to `solrctl collection --create`.

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-34050

CrunchIndexerTool which includes Spark indexer requires specific input file format specifications

If the `--input-file-format` option is specified with CrunchIndexerTool, then its argument must be `text`, `avro`, or `avroParquet`, rather than a fully qualified class name.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-22190

The `quickstart.sh` file does not validate ZooKeeper and the NameNode on some operating systems

The `quickstart.sh` file uses the `timeout` function to determine if ZooKeeper and the NameNode are available. To ensure this check can be complete as intended, the `quickstart.sh` determines if the operating system on which the script is running supports `timeout`. If the script detects that the operating system does not support `timeout`, the script continues without checking if the NameNode and ZooKeeper are available. If your environment is configured properly or you are using an operating system that supports `timeout`, this issue does not apply.

Workaround: This issue only occurs in some operating systems. If `timeout` is not available, the `quickstart` continues and final validation is always done by the MapReduce jobs and Solr commands that are run by the `quickstart`.

Affected Versions: All

Cloudera Issue: CDH-19923

Field value class guessing and Automatic schema field addition are not supported with the MapReduceIndexerTool nor the HBaseMapReduceIndexerTool

The MapReduceIndexerTool and the HBaseMapReduceIndexerTool can be used with a Managed Schema created via NRT indexing of documents or via the Solr Schema API. However, neither tool supports adding fields automatically to the schema during ingest.

Workaround: Define the schema before running the MapReduceIndexerTool or HBaseMapReduceIndexerTool. In non-schemaless mode, define in the schema using the `schema.xml` file. In schemaless mode, either define the schema using the Solr Schema API or index sample documents using NRT indexing before invoking the tools. In either case, Cloudera recommends that you verify that the schema is what you expect using the List Fields API command.

Affected Versions: All**Cloudera Issue:** CDH-26856

The *Browse* and *Spell* Request Handlers are not enabled in schemaless mode

The *Browse* and *Spell* Request Handlers require certain fields be present in the schema. Since those fields cannot be guaranteed to exist in a Schemaless setup, the *Browse* and *Spell* Request Handlers are not enabled by default.

Workaround: If you require the “Browse” and “Spell” Request Handlers, add them to the `solrconfig.xml` configuration file. Generate a non-schemaless configuration to see the usual settings and modify the required fields to fit your schema.

Affected Versions: All**Cloudera Issue:** CDH-19407

Enabling blockcache writing may result in unusable indexes

It is possible to create indexes with `solr.hdfs.blockcache.write.enabled` set to `true`. Such indexes may appear corrupt to readers, and reading these indexes may irrecoverably corrupt indexes. Blockcache writing is disabled by default.

Workaround: None

Affected Versions: All**Cloudera Issue:** CDH-17978

Users with insufficient Solr permissions may receive a "Page Loading" message from the Solr Web Admin UI

Users who are not authorized to use the Solr Admin UI are not given page explaining that access is denied, and instead receive a web page that never finishes loading.

Workaround: None

Affected Versions: All**Cloudera Issue:** CDH-58276

Using MapReduceIndexerTool or HBaseMapReduceIndexerTool multiple times may produce duplicate entries in a collection.

Repeatedly running the MapReduceIndexerTool on the same set of input files can result in duplicate entries in the Solr collection. This occurs because the tool can only insert documents and cannot update or delete existing Solr documents. This issue does not apply to the HBaseMapReduceIndexerTool unless it is run with more than zero reducers.

Workaround: To avoid this issue, use HBaseMapReduceIndexerTool with zero reducers. This must be done without Kerberos.

Affected Versions: All**Cloudera Issue:** CDH-15441

Deleting collections might fail if hosts are unavailable

It is possible to delete a collection when hosts that host some of the collection are unavailable. After such a deletion, if the previously unavailable hosts are brought back online, the deleted collection may be restored.

Workaround: Ensure all hosts are online before deleting collections.

Affected Versions: All**Cloudera Issue:** CDH-58694

Saving search results is not supported

Cloudera Search does not support the ability to save search results.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-21162

HDFS Federation is not supported

Cloudera Search does not support HDFS Federation.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-11357

Cloudera Search configuration migration script fails to detect incompatible SecureAdminHandlers request handler

The `SecureAdminHandlers` request handler is incompatible with Apache Solr 7, which is used in CDH 6. The Cloudera Search configuration migration script fails to detect this incompatibility.

Workaround: Remove `SecureAdminHandlers` request handlers from the `solrconfig.xml` files of any configuration set that uses them during the [pre-upgrade configuration migration](#).

Affected Versions: CDH 6.0.0

Fixed Versions: CDH 6.0.1

Cloudera Issue: CDH-72239

Apache Sentry Known Issues

When granting privileges, a single transaction per grant causes long delays

Sentry takes a long time to grant or revoke a large number of column-level privileges that are requested in a single statement. For example if you execute the following command:

```
GRANT SELECT(col1, col2, ...) ON TABLE table1;
```

Sentry applies the grants to each column separately and the refresh process causes long delays.

Workaround: Split the grant statement up into smaller chunks. This prevents the refresh process from causing delays.

Affected Versions:

- CDH: 5.14.4
- CDH: 5.15.1
- CDH: 5.16.0
- CDH: 6.1.0

Fixed Versions:

- CDH 5.16.1 and above
- CDH 6.2.0 and above

Cloudera Issue: CDH-74982

SHOW ROLE GRANT GROUP raises exception for a group that was never granted a role

If you run the command `SHOW ROLE GRANT GROUP` for a group that has never been granted a role, beeline raises an exception. However, if you run the same command for a group that does not have any roles, but has at one time been granted a role, you do not get an exception, but instead get an empty list of roles granted to the group.

Workaround: Adding a role will prevent the exception.

Affected Versions:

- CDH 5.16.0
- CDH 6.0.0

Cloudera Issue: CDH-71694

GRANT/REVOKE operations could fail if there are too many concurrent requests

Under a significant workload, Grant/Revoke operations can have issues.

Workaround: If you need to make many privilege changes, plan them at a time when you do not need to do too many at once.

Affected Versions: CDH 5.13.0 and above

Apache Issue: [SENTRY-1855](#)

Cloudera Issue: CDH-56553

Creating large set of Sentry roles results in performance problems

Using more than a thousand roles/permissions might cause significant performance problems.

Workaround: Plan your roles so that groups have as few roles as possible and roles have as few permissions as possible.

Affected Versions: CDH 5.13.0 and above

Cloudera Issue: CDH-59010

Users can't track jobs with Hive and Sentry

As a prerequisite of enabling Sentry, Hive impersonation is turned off, which means all YARN jobs are submitted to the Hive job queue, and are run as the `hive` user. This is an issue because the YARN History Server now has to block users from accessing logs for their own jobs, since their own usernames are not associated with the jobs. As a result, end users cannot access any job logs unless they can get `sudo` access to the cluster as the `hdfs`, `hive` or other admin users.

In CDH 5.8 (and higher), Hive overrides the default configuration, `mapred.job.queueName`, and places incoming jobs into the connected user's job queue, even though the submitting user remains `hive`. Hive obtains the relevant queue/username information for each job by using YARN's `fair-scheduler.xml` file.

Affected Versions: CDH 5.2.0 and above

Cloudera Issue: CDH-22890

Column-level privileges are not supported on Hive Metastore views

`GRANT` and `REVOKE` for column level privileges is not supported on Hive Metastore views.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-754](#)

SELECT privilege on all columns does not equate to SELECT privilege on table

Users who have been explicitly granted the `SELECT` privilege on all columns of a table, will *not* have the permission to perform table-level operations. For example, operations such as `SELECT COUNT (1)` or `SELECT COUNT (*)` will not work even if you have the `SELECT` privilege on all columns.

There is one exception to this. The `SELECT * FROM TABLE` command will work even if you do not have explicit table-level access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-838](#)

The EXPLAIN SELECT operation works without table or column-level privileges

Users are able to run the `EXPLAIN SELECT` operation, exposing metadata for all columns, even for tables/columns to which they weren't explicitly granted access.

Affected Versions: All CDH versions

Apache Issue: [SENTRY-849](#)

Object types Server and URI are not supported in SHOW GRANT ROLE roleName on OBJECT objectName

Workaround: Use `SHOW GRANT ROLE roleName` to list all privileges granted to the role.

Affected Versions: All CDH versions

Apache Issue: N/A

Cloudera Issue: CDH-19430

Relative URI paths not supported by Sentry

Sentry supports only absolute (not relative) URI paths in permission grants. Although some early releases (for example, CDH 5.7.0) might not have raised explicit errors when relative paths were set, upgrading a system that uses relative paths causes the system to lose Sentry permissions.

Affected Versions: All versions. Relative paths are not supported in Sentry for permission grants.

Resolution: Revoke privileges that have been set using relative paths, and grant permissions using absolute paths before upgrading.

Absolute (Use this form)	Relative (Do not use this form)
hdfs://absolute/path/	hdfs://relative/path
s3a://bucketname/	s3a://bucketname

Apache Spark Known Issues

The following sections describe the current known issues and limitations in Apache Spark 2.x as distributed with CDH 6. In some cases, a feature from the upstream Apache Spark project is currently not considered reliable enough to be supported by Cloudera.

Spark Streaming jobs loop if missing Kafka topic

Spark jobs can loop endlessly if the Kafka topic is deleted while a Kafka streaming job (which uses KafkaSource) is in progress.

Workaround: Stop a job before deleting a Kafka topic.

Affected Versions: All

Cloudera Issue: CDH-57903, CDH-64513

Spark SQL does not respect size limit for the varchar type

Spark SQL treats `varchar` as a string (that is, there no size limit). The observed behavior is that Spark reads and writes these columns as regular strings; if inserted values exceed the size limit, no error will occur. The data will be truncated when read from Hive, but not when read from Spark.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Apache Issue: [SPARK-5918](#)

Cloudera Issue: CDH-33642

Spark SQL does not prevent you from writing key types not supported by Avro tables

Spark allows you to declare DataFrames with any key type. Avro supports only string keys and trying to write any other key type to an Avro table will fail.

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33648

Spark SQL does not support timestamp in Avro tables

Workaround: None

Affected Versions: CDH 5.5.0 and higher

Cloudera Issue: CDH-33649

Spark SQL does not respect Sentry ACLs when communicating with Hive metastore

Even if user is configured via Sentry to not have read permission to a Hive table, a Spark SQL job running as that user can still read the table's metadata directly from the Hive metastore. **Cloudera Issue:** CDH-33658

Dynamic allocation and Spark Streaming

If you are using Spark Streaming, Cloudera recommends that you disable dynamic allocation by setting `spark.dynamicAllocation.enabled` to `false` when running streaming applications.

Limitation with Region Pruning for HBase Tables

When SparkSQL accesses an HBase table through the HiveContext, region pruning is not performed. This limitation can result in slower performance for some SparkSQL queries against tables that use the HBase SerDes than when the same table is accessed through Impala or Hive.

Workaround: None

Affected Versions: All

Cloudera Issue: CDH-56330

Running `spark-submit` with `--principal` and `--keytab` arguments does not work in client mode

The `spark-submit` script's `--principal` and `--keytab` arguments do not work with Spark-on-YARN's client mode.

Workaround: Use `cluster` mode instead.

Affected Versions: All

Long-running apps on a secure cluster might fail if driver is restarted

If you submit a long-running app on a secure cluster using the `--principal` and `--keytab` options in cluster mode, and a failure causes the driver to restart after 7 days (the default maximum HDFS delegation token lifetime), the new driver fails with an error similar to the following:

```
Exception in thread "main"
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken):
  token <token_info> can't be found in cache
```

Workaround: None

Affected Versions: CDH 6.0

Apache Issue: [SPARK-23361](#)**Cloudera Issue:** CDH-64865

History link in ResourceManager web UI broken for killed Spark applications

When a Spark application is killed, the history link in the ResourceManager web UI does not work.

Workaround: To view the history for a killed Spark application, see the Spark HistoryServer web UI instead.

Affected Versions: All CDH versions

Apache Issue: None

Cloudera Issue: CDH-49165*Apache Sqoop Known Issues*

Column names cannot start with a number when importing data with the `--as-parquetfile` option.

Currently, Sqoop is using an Avro schema when writing data as a parquet file. The Avro schema requires that column names do not start with numbers, therefore Sqoop is renaming the columns in this case, prepending them with an underscore character. This can lead to issues when one wants to reuse the data in other tools, such as Impala.

Cloudera Enterprise 6 Release Guide

Workaround: Rename the columns to comply with Avro limitations (start with letters or underscore, as specified in the [Avro documentation](#)).

Cloudera Issue: None

MySQL JDBC driver shipped with CentOS 6 systems does not work with Sqoop

CentOS 6 systems currently ship with version 5.1.17 of the MySQL JDBC driver. This version does not work correctly with Sqoop.

Workaround: Install version 5.1.31 of the JDBC driver as detailed in [Installing the JDBC Drivers for Sqoop 1](#).

Affected Versions: MySQL JDBC 5.1.17, 5.1.4, 5.3.0

Cloudera Issue: CDH-23180

MS SQL Server "integratedSecurity" option unavailable in Sqoop

The `integratedSecurity` option is not available in the Sqoop CLI.

Workaround: None

Cloudera Issue: None

Sqoop1 (doc import + --as-parquetfile) limitation with KMS/KTS Encryption at Rest

Due to a limitation with Kite SDK, it is not possible to use (`sqoop import --as-parquetfile`) with KMS/KTS Encryption zones. See the following example.

```
sqoop import --connect jdbc:db2://djaxludb1001:61035/DBBAT003 --username=dh810202 --P  
--target-dir /data/hive_scratch/ASDISBURSEMENT --delete-target-dir -m1 --query "select  
disbursementnumber,disbursementdate,xmldata FROM DB2dba.ASDISBURSEMENT where  
DISBURSEMENTNUMBER = 2011113210000115311 AND \$CONDITIONS" -hive-import --hive-database  
adminserver -hive-table asdisbursement_dave --map-column-java XMLDATA=String  
--as-parquetfile  
  
16/12/05 12:23:46 INFO mapreduce.Job: map 100% reduce 0%  
16/12/05 12:23:46 INFO mapreduce.Job: Job job_1480530522947_0096 failed with state FAILED  
due to: Job commit failed: org.kitesdk.data.DatasetIOException: Could not move contents  
of  
hdfs://AJAX01-ns/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096  
to hdfs://AJAX01-ns/data/RetiredApps/INS/AdminServer/asdisbursement_dave  
<SNIP>  
Caused by: org.apache.hadoop.ipc.RemoteException(java.io.IOException):  
/tmp/adminserver/.temp/job_1480530522947_0096/mr/job_1480530522947_0096/5ddcac42-5d69-4e46-88c2-17bedac4858.parquet  
can't be moved into an encryption zone.
```

Workaround: If you use the Parquet Hadoop API based implementation for importing into Parquet, specify a `--target-dir` which is the same encryption zone as the Hive warehouse directory.

If you use the Kite Dataset API based implementation, use an alternate data file type, for example text or avro.

Apache Issue: SQQOP-2943

Cloudera Issue: CDH-40826

Doc import as Parquet files may result in out-of-memory errors

Out-of-memory (OOM) errors can be caused in the following two cases:

- With many very large rows (multiple megabytes per row) before initial-page-run check (ColumnWriter)
- When rows vary significantly by size so that the next-page-size check is based on small rows and is set very high followed by many large rows

Apache Issue: PARQUET-99

Workaround: None, other than restructuring the data.

Apache ZooKeeper Known Issues

ZooKeeper JMX did not support TLS when managed by Cloudera Manager

Technical Service Bulletin 2019-310 (TSB)

The ZooKeeper service optionally exposes a JMX port used for reporting and metrics. By default, Cloudera Manager enables this port, but prior to Cloudera Manager 6.1.0, it did not support mutual TLS authentication on this connection. While JMX has a password-based authentication mechanism that Cloudera Manager enables by default, weaknesses have been found in the authentication mechanism, and Oracle now advises JMX connections to enable mutual TLS authentication in addition to password-based authentication. A successful attack may leak data, cause denial of service, or even allow arbitrary code execution on the Java process that exposes a JMX port. Beginning in Cloudera Manager 6.1.0, it is possible to configure mutual TLS authentication on ZooKeeper's JMX port.

Products affected: ZooKeeper

Releases affected: Cloudera Manager 6.1.0 and lower, Cloudera Manager 5.16 and lower

Users affected: All

Date/time of detection: June 7, 2018

Severity (Low/Medium/High): 9.8 High ([CVSS:3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H](#))

Impact: Remote code execution

CVE: CVE-2018-11744

Immediate action required: Upgrade to Cloudera Manager 6.1.0 and enable TLS for the ZooKeeper JMX port by turning on the configuration settings “Enable TLS/SSL for ZooKeeper JMX” and “Enable TLS client authentication for JMX port” on the ZooKeeper service and configuring the appropriate TLS settings. Alternatively, disable the ZooKeeper JMX port via the configuration setting “Enable JMX Agent” on the ZooKeeper service.



Note: Disabling the ZooKeeper JMX port prevents Cloudera Manager from performing health checks on the ZooKeeper service.

Addressed in release/refresh/patch: Cloudera Manager 6.1.0

Cloudera Navigator 6 Data Management Release Notes



Note: For Cloudera Navigator 6 encryption component release notes, see [Cloudera Navigator 6 Encryption Release Notes](#) on page 641.

To view release notes for the data management components of a specific Cloudera Navigator 6 release, see the following:

Cloudera Navigator 6.2.0 Data Management Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the data management components of Cloudera Navigator 6.2.0:

New Features in Cloudera Navigator 6.2.0

The following sections describe what's new and what has changed in the data management components of Cloudera Navigator 6.2.0:

Virtual Private Cluster Support

Navigator continues to support audit and metadata extraction from services on the base cluster in an environment running both base and compute clusters. However, no audit events or metadata are extracted from services running on compute clusters. You can expect the following behavior for services included in a data context:

- **HDFS, S3, ADLS.** HDFS audits and metadata changes that affect file system directories and files are collected in Navigator. Operations and operation executions that are performed on the compute cluster against HDFS entities

are not collected. This means Navigator will show the HDFS entities, but if operations are performed against them from the compute cluster, Navigator will not include metadata for the operation execution or lineage from the HDFS entity to any input or output.

- **Hive.** HMS audits and metadata changes that describe Hive databases, tables, views, partitions, and columns are collected in Navigator. Operations and operation executions performed by HiveServer2, Impala, and SparkSQL on the compute cluster against Hive entities are not collected. For example, if a user runs a SELECT against a table from the compute cluster, a SELECT event is not collected; if a user creates a table and loads data into it on the compute cluster, Navigator will create an entity for the table when extracting metadata from HMS on the base cluster. It will not include metadata for the create operation or create lineage to indicate the source of data for the table.
- **Sentry.** Services that check user access against Sentry policies have the same behavior on the compute cluster as on the base cluster. However, Navigator will not collect audits for when services on the compute cluster check Sentry policies for data access. For example, when a user performs a SELECT against a table from the compute cluster and the user does not have access to the table or one or more columns, the denied access event would be created by the service on the compute cluster and is therefore not collected. If the same actions occur on the base cluster, the Navigator audits will include the events.

To summarize, Navigator does collect audit events and extract metadata from shared services in the data context. Navigator does not collect audits from services running on compute clusters. Navigator does not extract metadata for services running on compute clusters. For more information, see [Virtual Private Clusters and Cloudera SDX](#).

Deterministic Metadata Purge Operation

In this release, the metadata purge operation has a high priority relative to other Navigator Metadata Server tasks. When a purge operation runs, it first stops any other Navigator Metadata Server tasks. After the purge completes (or reaches the time limit for when it can run), it reschedules the extraction tasks that were stopped. This behavior applies to metadata clean up jobs scheduled in the Navigator console (**Administration > Purge Settings**) or through the Navigator API and to jobs to clear deleted managed metadata properties (**Administration > Managed Properties > Purge deleted Properties**). Note that policy tasks interrupted by a purge operation are not restarted.

For more information, see [Best Practices for Clearing Metadata using Purge](#).

Bulk Metadata Update API

This release provides a bulk interface for updating Navigator entity metadata. Use the `PUT /entities/bulk/` API to update metadata for many entities in the same call. This API is faster than the single `PUT` API because it uses a single HTTP request to apply the metadata rather than an HTTP request for each entity.

For more information see the interactive API documentation available from the Help menu in the Navigator console or [Updating Metadata for Entities in Bulk: PUT /entities/bulk](#).

Metadata for Hive columns now includes the column number

The technical metadata for Hive table columns now includes the ordinal position for the columns as "Field Index", the index starting at zero. You can include "Field Index" as a search facet for Hive columns or specify "fieldIndex" in a metadata search query. In addition, columns are ordered by their field index when listed in the table metadata.

For new installations, this additional metadata appears as metadata for Hive assets are extracted. For upgraded installations, the additional metadata must be re-extracted from Hive assets that are already in Navigator; you may not see the field index values immediately after the upgrade.

Issues Fixed in Cloudera Navigator 6.2.0

The following sections describe the issues fixed in the data management components of Cloudera Navigator 6.2.0:

Navigator Metadata Server purge jobs might not run if there were policies configured

Navigator Metadata Server purge could produce messages such as "Checking if maintenance is running" and then failed to run during the available window. This problem occurred when a scheduled purge job waits for extraction tasks to finish but while waiting, a policy job starts, preventing the purge job from running. In this release, policy jobs will not be triggered if purge jobs are scheduled and waiting to start.

Cloudera Issue: NAV-7037

Support for Compute-Only Clusters

Navigator support for compute-only clusters in Cloudera Manager includes audit and metadata extraction from services on the base cluster. No audit events or metadata are extracted from services running on the compute cluster.

For more information, see [Virtual Private Cluster Support](#) on page 593.

Cloudera Issue: NAV-7028

International characters in tag names

Navigator tags now support UNICODE characters beyond ASCII. Only ASCII text can be used in the name of a user-defined or managed property. Property values can include international characters.

Cloudera Issue: NAV-7011

Bulk update API

This release provides a bulk interface for updating Navigator entity metadata. Use the `PUT /entities/bulk/` API to update metadata for many entities in the same call. This API is faster than the single `PUT` API because it uses a single HTTP request to apply the metadata rather than an HTTP request for each entity.

For more information see the interactive API documentation available from the Help menu in the Navigator console or [Updating Metadata for Entities in Bulk: PUT /entities/bulk](#).

Security vulnerabilities addressed for Thrift

The Apache Thrift client used by Navigator has been upgraded to 0.12.0 to resolve a security vulnerability indicated by [CVE-2018-1320](#).

Cloudera Issue: NAV-6998

Metadata purge priority over other tasks

This release changes the behavior of the metadata purge operation. When the scheduled purge operation starts, Navigator Metadata Server stops any running extraction and policy tasks so that the purge operation can start immediately. The extraction tasks are automatically rescheduled when the purge completes; the policy tasks are not.

For more information, see [Best Practices for Clearing Metadata using Purge](#).

Cloudera Issue: NAV-6959

Reduction in the amount of memory used by Navigator Metadata Server

This release includes changes that improve how Navigator Metadata Server uses memory, reducing the overall heap required.

Cloudera Issue: NAV-6958

Console display of purge history was not sorted

The Navigator console content showing the history of purge jobs is now sorted by time in descending order.

Cloudera Issue: NAV-6916

Security vulnerabilities addressed for Jetty

The Eclipse Jetty web server used by Navigator has been upgraded to 9.4.11 to resolve security vulnerabilities indicated by: [CVE-2015-2080](#), [CVE-2016-4800](#), [CVE-2017-7657](#), [CVE-2017-7658](#), and [CVE-2017-9735](#).

Cloudera Issue: NAV-6901, NAV-6865

Lineage input and output count corrected

Lineage output counts were not displayed correctly when an entity is both an input and output for the current entity. Both the input and output counts are off by one.

In this release, the input and output counts include all leaf nodes that are not operations and not marked deleted.

Cloudera Issue: NAV-6892

[Index added to Hive column metadata](#)

The technical metadata for Hive columns now includes the ordinal position of the column as "Field Index", where the index starts at zero. Columns are now listed by their field index when shown in the parent table or view details. Previously, there was no specific order. The additional metadata for existing Hive column entities must be extracted from the Hive sources, so you will not see it immediately after upgrading.

Cloudera Issue: NAV-6815

[Navigator didn't recognize local files in Spark jobs](#)

Spark jobs can use files on the local filesystem as job inputs or outputs. Navigator only supported HFDS, Hive, and S3 assets as job inputs or outputs. As of this release, Navigator now handles local source types when extracting metadata from Spark jobs.

Cloudera Issue: NAV-6811

[Swagger interactive documentation fixed for GET /audits API](#)

The Navigator API interactive documentation for GET /audits had the wrong selector name for databases. The reference to `database_name` is now correct.

To explore the Swagger interactive API interface, open **API Documentation** from the help menu in the Navigator console.

Cloudera Issue: NAV-6411

[Kite Datasets support deprecated](#)

Kite Dataset API was deprecated for Cloudera Manager in 6.0.0. Navigator support for metadata extraction from Kite Datasets is now deprecated and will be removed in a future release. To avoid log messages indicating Kite Dataset issues, you can disable Kite extraction using a Cloudera Manager safety valve. Set the following in the **Navigator Metadata Server Advanced Configuration Snippet (Safety Valve) for cloudera-navigator.properties**:

```
nav.extractor.hdfs.datasets.enabled
```

Cloudera Issue: NAV-6368

[Navigator Audit Server "CannotSendRequest" error resolved](#)

When an error occurred in the audit pipeline, it was possible that not all components in the Navigator Audit Server recovered fully. One symptom of such a failure is the "CannotSendRequest" error in the Navigator Audit Server log. This problem is fixed in this release.

Cloudera Issue: NAV-6321

[Corrected exception produced during S3 extraction](#)

Enabling or disabling S3 extraction caused a null pointer exception (NPE) indicating that some of the information needed was null. The error did not affect the S3 extraction functionality. This release improves the interchange to avoid the error.

Cloudera Issue: NAV-6246

[Known Issues in Cloudera Navigator 6.2.0](#)

The following known issues and limitations from a previous release affect the data management components of Cloudera Navigator 6.2.0:

[Authentication and Authorization](#)

[*SAML authentication fails with "Cloudera Manager Only" setting*](#)

With the following combination of Cloudera Manager configuration properties set, authentication to Navigator fails:

- Authentication Backend Order: Cloudera Manager Only
- External Authentication Type: SAML

Workaround: To configure Navigator for SAML authentication, use an **Authentication Backend Order** other than "Cloudera Manager Only".

Authentication Backend Order nav.auth.backend.order	Navigator Metadata Server Default Group  <input checked="" type="radio"/> External Only <input type="radio"/> External then Cloudera Manager <input type="radio"/> Cloudera Manager Only <input type="radio"/> Cloudera Manager then External
---	--

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-6211

Cloudera Manager Configuration

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

At this time, there is no workaround.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-5331

Hive, Hue, Impala

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

Affected Versions: Cloudera Navigator 6.0.0 and later

Workaround: To fix this problem, manually merge the original HiveServer2 safety valve content for `hive.exec.post.hooks` with the new value. For example, in the case of Unravel, the new safety valve would look like the following:

```
<property>
  <name>hive.exec.post.hooks</name>
  <value>com.unaveldata.flowhive.hivePostHook,com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger</value>
  <description>for Unravel, from unraveldata.com</description>
</property>
```

Cloudera Issue: NAV-5331*Viewing Navigator tags in Hue overloads Metadata Server heap*

When viewing Cloudera Navigator tags through Hue, Navigator uses more memory than usual and does not release the memory after logging out of Hue. Eventually, the calls between Hue and Navigator will occupy the majority of the heap space allocated to Navigator Metadata Server.

Workaround: Restart the Navigator Metadata Server periodically to clear the heap usage.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4326*Lineage not generated for Pig operations on Hive tables using HCatalog loader*

When accessing a Hive table using Pig, lineage is generated in Navigator when using physical file loads, such as:

```
A = LOAD '/user/hive/warehouse/navigator_demo.db/salesdata';
B = LIMIT A 16;
STORE B INTO '/user/hive/warehouse/navigator_demo.db/salesdata_sample_file' using
PigStorage (';');
```

However, when accessing the Hive table using the HCatalog load, lineage for the Pig operation is not generated when browsing the source table lineage. Such as:

```
A = LOAD 'navigator_demo.salesdata' using org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 16;
STORE B INTO 'navigator_demo.salesdata_sample_hcatalog' using
org.apache.hive.hcatalog.pig.HCatStorer();
```

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3411*HiveServer1 and Hive CLI support removed*

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

If you use Cloudera Navigator to capture auditing, lineage, and metadata for Hive operations, upgrade to HiveServer2 if you have not done so already.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: TSB-185*Navigator Audit Server**Logging Threshold setting is not honored*

The value for Navigator Audit Server Logging Threshold found in Cloudera Manager is not honored. Instead, messages are logged at trace level and displayed at DEBUG syslog level. This configuration property is set in **Cloudera Management Service > Configuration > Navigator Audit Server**.

Cloudera Issue: NAV-3737

Navigator Metadata Server*Navigator Embedded Solr can reach its limit on number of documents it can store*

Important: This issue is critical when upgrading Cloudera Navigator deployments from Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0.

Navigator Metadata Server extracts HDFS entities by performing a one-time bulk extraction and then switching to incremental extraction. In Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0 (Navigator releases 2.9.0, 2.9.1, and 2.10.0), a problem causes HDFS bulk extraction to be run more than one time, resulting in duplicate relations created for HDFS. Over time, embedded Solr runs out of document IDs that it can assign to new relations and fails with following error:

```
"Caused by: java.lang.IllegalArgumentException: Too many documents, composite IndexReaders cannot exceed 2147483519"
```

When this happens, Navigator stops any more extraction of data as no new documents can be added to Solr.

After upgrading to this release, there is an additional recover step as described in "Repairing metadata in the storage directory after upgrading" in [Troubleshooting Navigator Data Management](#).

Affected Versions: Versions prior to Cloudera Manager 5.10 upgraded to Cloudera Manager 5.10 or higher

Fixed Versions: N/A

Cloudera Issue: NAV-5600

Log includes the error "EndPoint1 must not be null"

The following error may appear in the Navigator Metadata Server log in systems upgraded from Cloudera Manager version 5.x:

```
2017-10-17 13:00:23,007 ERROR com.cloudera.nav.hive.extractor.AbstractHiveExtractor [CDHExecutor-0-CDHUrlClassLoader@14784b7b]: Unable to parse hive view query *: EndPoint1 must not be null or empty
java.lang.IllegalStateException: EndPoint1 must not be null or empty
```

This error occurs because the Hive pull extraction for creating a Hive view produces an incorrect lineage relationship for the Hive view. However, Navigator also receives information for the view creation through the push extractor, which correctly produces the lineage relation. You can safely ignore this error.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4224

Purge

Policy specifications and cluster names affect purge

Policies cannot use cluster names in queries. Cluster name is a derived attribute and cannot be used as-is.

Workaround: When setting move actions for Cloudera Navigator, if there is only one cluster known to the Navigator instance, remove the `clusterName` clause.

If there is more than one cluster known to the Navigator instance, replace `clusterName` with `sourceId`. To get the `sourceId`, issue a query in this format:

```
curl '<nav-url>/api/v9/entities/?query=type%3Asource&limit=100&offset=0'
```

Use the identity of the matching HDFS service for this cluster as the `sourceId`.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3537

Spark

Spark Lineage Limitations and Requirements

Spark lineage diagrams are supported in the Cloudera Navigator 6.0 release. Spark lineage is supported for Spark 1.6 and Spark 2.3. Lineage is not available for Spark when Cloudera Manager is running in single user mode. In addition to these requirements, Spark lineage has the following limitations:

- Lineage is produced only for data that is read/written and processed using the Dataframe and SparkSQL APIs. Lineage is not available for data that is read/written or processed using Spark's RDD APIs.
- Lineage information is not produced for calls to aggregation functions such as `groupBy()`.
- The default lineage directory for Spark on Yarn is `/var/log/spark/lineage`. No process or user should write files to this directory—doing so can cause agent failures. In addition, changing the Spark on Yarn lineage directory has no effect: the default remains `/var/log/spark/lineage`.

Spark extractor enabled using safety valve deprecated

The Spark extractor included prior to CDH 5.11 and enabled by setting the safety valve, `nav.spark.extraction.enable=true` is being deprecated, and could be removed completely in a future release. If you are upgrading from CDH 5.10 or earlier and were using the extractor configured with this safety valve, be sure to remove the setting when you upgrade.

Upgrade Issues and Limitations

Upgrading Cloudera Navigator from Cloudera Manager 5.9 or Earlier Can be Extremely Slow

Upgrading a cluster running Cloudera Navigator to Cloudera Manager 5.10 (or higher) can be extremely slow due to an internal change made to the Solr schema in Cloudera Navigator 2.9. A Solr instance is embedded in Cloudera Navigator and supports its search capabilities. The Solr schema used by Cloudera Navigator has been modified in the 2.9 release to use datatype `long` rather than `string` for an internal `id` field. This change makes Cloudera Navigator far more robust and scalable over the long term.

However, the upgrade process itself can take a significantly long time because the existing Solr documents—the indexed and searchable data structures used by Solr that are contained in the Cloudera Navigator storage directory—are migrated to the new schema. This change to the Solr schema affects only those Cloudera Navigator deployments that use the metadata and lineage features.



Note: Cloudera Navigator deployments that use only the auditing features of the product—not metadata or lineage—are not affected by this issue.

The upgrade process for Cloudera Navigator starts automatically at the end of the Cloudera Manager upgrade, and the migration to the new schema occurs automatically as part of that upgrade process. The Navigator Metadata Server and Navigator console are not available during the upgrade. Navigator Audit Server runs normally. The amount of time that administrators should allow for this process depends on the quantity stored at the Navigator Metadata Server Storage Dir (`nav.data.dir`, or simply "storage directory") location as listed here:

Metadata and lineage usage	Description
None	Deployments that use Cloudera Navigator audit capability only —without metadata or lineage—do not have the issue. Backup the Navigator Metadata Server storage directory and then delete it before upgrading.
storage directory < 60 GB	Deployments with relatively small Navigator Metadata Server data directories may take 1 to 2 days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.
storage directory > 60 GB	Deployments with relatively large Navigator Metadata Server data directories may take several days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Workaround: To reduce the time required for upgrading the Navigator Metadata Server data directories for deployments currently running Cloudera Navigator 2.8 that uses its metadata and lineage features, consider removing unneeded entries from the metadata before the upgrade. The Navigator Purge feature allows you to remove metadata for deleted entities and for entities and operations older than a specified date. For more information on what metadata you can remove with Purge, see [Managing Metadata Storage with Purge](#).

Run purge before starting the Cloudera Manager upgrade (to Cloudera Manager 5.10), following the steps below.



Warning: These steps may mitigate but do not fully resolve the issue. Follow these steps **before** starting the Cloudera Manager upgrade for any Cloudera Manager 5.9 or earlier cluster that currently uses the Cloudera Navigator metadata and lineage features.

- Check the Navigator Metadata Server storage directory size. The path is `/var/lib/cloudera-scm-navigator` (default) unless configured otherwise. If you need to check the setting:
 - Log in to Cloudera Manager Admin Console.
 - Select **Clusters > Cloudera Management Service**.
 - Click **Configuration** and then click the **Navigator Metadata Server** Scope filter:

Navigator Metadata Server	Navigator Metadata Server (node-1)
Storage Dir	
nav.data.dir	/var/lib/cloudera-scm-navigator

- Confirm that the cluster uses the default configuration, or make a note of the location specified and the node name.
- Check the size of the actual directory contents. The following example shows a freshly installed system and so it is virtually empty.

```
[root@node-1 ~]# cd /var/lib/cloudera-scm-navigator
[root@node-1 cloudera-scm-navigator]# ls -l
total 12
drwxr-x--- 2 cloudera-scm cloudera-scm 113 Jul 12 06:56 diagnosticData
drwxr-x--- 2 cloudera-scm cloudera-scm 4096 Jul 12 09:16 extractorState
-rw-r----- 1 cloudera-scm cloudera-scm 36 Jul 12 05:54 instance.uuid
drwxr-x--- 4 cloudera-scm cloudera-scm 60 Jul 12 04:18 solr
drwxr-x--- 7 cloudera-scm cloudera-scm 4096 Jul 12 07:26 temp
[root@node-1 cloudera-scm-navigator]# cd solr
[root@node-1 solr]# ls -l
total 4
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_elements
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_relations
-rw-r----- 1 cloudera-scm cloudera-scm 450 Jul 6 16:13 solr.xml
[root@node-1 solr]#
```

- Back up the contents of the directory. Use Cloudera Manager BDR or your preferred method.
- Schedule a purge process as described in [Scheduling the Purge Process](#).



Note: Users and processes cannot access Cloudera Navigator while purge is running.

Set options to purge metadata for deleted HDFS entities and any operations.

- Check the storage directory size again. If needed, re-run the purge with a shorter time span to retain metadata. If the storage directory consumption cannot be reduced below 60GB, do not start the Cloudera Manager upgrade. Instead, contact [Cloudera support](#) to help you with this upgrade.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: NAV-5046

Cloudera Navigator 6.1.x Data Management Release Notes

To view release notes for the data management components of specific Cloudera Navigator 6.1.x releases, see the following:

Cloudera Navigator 6.1.1 Release Notes

The following topics describe new features (none), fixed issues, and known issues for the data management components of Cloudera Navigator 6.1.1:

New Features in Cloudera Navigator 6.1.1

There are no new features in the data management components of Cloudera Navigator. See also [Known Issues in Cloudera Navigator 6.1.1](#) on page 602.

Issues Fixed in Cloudera Navigator 6.1.1

The following sections describe the issues fixed in the data management components of Cloudera Navigator 6.1.1:

Correction to lineage input and output counts

Lineage output counts were not displayed correctly in the Navigator console when an entity was both an input and output for the current entity. Both the input and output counts were off by one.

In this release, the input and output counts include all leaf nodes that are not operations and are not marked deleted.

Cloudera Issue: NAV-6892

Duplicate data flow relations after Sqoop operation

After using Sqoop to transfer data to HDFS, Navigator Metadata Server may have duplicated relations between YARN operation execution entities and HDFS directory entities. This issue is fixed in this release.

Cloudera Issue: NAV-6446

Corrected exception produced during S3 extraction

Enabling or disabling S3 extraction caused a null pointer exception (NPE) indicating that some of the information needed was null. The error did not affect the S3 extraction functionality. This release improves the interchange to avoid the error.

Cloudera Issue: NAV-6246

Known Issues in Cloudera Navigator 6.1.1

The following known issues and limitations from a previous release affect the data management components of Cloudera Navigator 6.1.1:

Authentication and Authorization

SAML authentication fails with "Cloudera Manager Only" setting

With the following combination of Cloudera Manager configuration properties set, authentication to Navigator fails:

- Authentication Backend Order: Cloudera Manager Only
- External Authentication Type: SAML

Workaround: To configure Navigator for SAML authentication, use an **Authentication Backend Order** other than "Cloudera Manager Only".

Authentication Backend Order nav.auth.backend.order	Navigator Metadata Server Default Group  <input checked="" type="radio"/> External Only <input type="radio"/> External then Cloudera Manager <input type="radio"/> Cloudera Manager Only <input type="radio"/> Cloudera Manager then External
---	--

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-6211

Cloudera Manager Configuration

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

At this time, there is no workaround.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-5331

Hive, Hue, Impala

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

Affected Versions: Cloudera Navigator 6.0.0 and later

Workaround: To fix this problem, manually merge the original HiveServer2 safety valve content for `hive.exec.post.hooks` with the new value. For example, in the case of Unravel, the new safety valve would look like the following:

```
<property>
  <name>hive.exec.post.hooks</name>
  <value>com.navdata.datflowhive.hook.HivePostHook,com.cloudera.navigator.audit.hive.HiveBackendContext,org.apache.hadoop.hive.ql.hooks.LineageLogger</value>
  <description>for Unravel, from unraveldata.com</description>
</property>
```

Cloudera Issue: NAV-5331

Viewing Navigator tags in Hue overloads Metadata Server heap

When viewing Cloudera Navigator tags through Hue, Navigator uses more memory than usual and does not release the memory after logging out of Hue. Eventually, the calls between Hue and Navigator will occupy the majority of the heap space allocated to Navigator Metadata Server.

Workaround: Restart the Navigator Metadata Server periodically to clear the heap usage.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4326

Lineage not generated for Pig operations on Hive tables using HCatalog loader

When accessing a Hive table using Pig, lineage is generated in Navigator when using physical file loads, such as:

```
A = LOAD '/user/hive/warehouse/navigator_demo.db/salesdata';
B = LIMIT A 16;
STORE B INTO '/user/hive/warehouse/navigator_demo.db/salesdata_sample_file' using
PigStorage (';');
```

However, when accessing the Hive table using the HCatalog load, lineage for the Pig operation is not generated when browsing the source table lineage. Such as:

```
A = LOAD 'navigator_demo.salesdata' using org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 16;
STORE B INTO 'navigator_demo.salesdata_sample_hcatalog' using
org.apache.hive.hcatalog.pig.HCatStorer();
```

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3411

HiveServer1 and Hive CLI support removed

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

If you use Cloudera Navigator to capture auditing, lineage, and metadata for Hive operations, upgrade to HiveServer2 if you have not done so already.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: TSB-185

Navigator Audit Server

Logging Threshold setting is not honored

The value for Navigator Audit Server Logging Threshold found in Cloudera Manager is not honored. Instead, messages are logged at trace level and displayed at DEBUG syslog level. This configuration property is set in **Cloudera Management Service > Configuration > Navigator Audit Server**.

Cloudera Issue: NAV-3737

Navigator Metadata Server

International characters not supported in tags or property names

Navigator tags and the key portion of user-defined and managed properties do not support UNICODE characters beyond ASCII. Only ASCII text can be used in the text of a tag or the name of a property. Property values can include international characters.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: Tags are fixed in Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7011, NAV-7044

Navigator Embedded Solr can reach its limit on number of documents it can store



Important: This issue is critical when upgrading Cloudera Navigator deployments from Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0.

Navigator Metadata Server extracts HDFS entities by performing a one-time bulk extraction and then switching to incremental extraction. In Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0 (Navigator releases 2.9.0, 2.9.1, and 2.10.0), a problem causes HDFS bulk extraction to be run more than one time, resulting in duplicate relations created for HDFS. Over time, embedded Solr runs out of document IDs that it can assign to new relations and fails with following error:

```
"Caused by: java.lang.IllegalArgumentException: Too many documents, composite IndexReaders cannot exceed 2147483519"
```

When this happens, Navigator stops any more extraction of data as no new documents can be added to Solr.

After upgrading to this release, there is an additional recover step as described in "Repairing metadata in the storage directory after upgrading" in [Troubleshooting Navigator Data Management](#).

Affected Versions: Versions prior to Cloudera Manager 5.10 upgraded to Cloudera Manager 5.10 or higher

Fixed Versions: N/A

Cloudera Issue: NAV-5600

Log includes the error "EndPoint1 must not be null"

The following error may appear in the Navigator Metadata Server log in systems upgraded from Cloudera Manager version 5.x:

```
2017-10-17 13:00:23,007 ERROR com.cloudera.nav.hive.extractor.AbstractHiveExtractor [CDHExecutor-0-CDHUrlClassLoader@14784b7b]: Unable to parse hive view query *: EndPoint1 must not be null or empty
java.lang.IllegalStateException: EndPoint1 must not be null or empty
```

This error occurs because the Hive pull extraction for creating a Hive view produces an incorrect lineage relationship for the Hive view. However, Navigator also receives information for the view creation through the push extractor, which correctly produces the lineage relation. You can safely ignore this error.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4224

Purge

Navigator Metadata Server purge jobs may not run if there are policies configured

Navigator Metadata Server purge can produce messages such as "Checking if maintenance is running" and then fail to run during the available window. This problem occurs when a scheduled purge job waits for extraction tasks to finish but while waiting, a policy job starts, preventing the purge job from running.

Workaround: Specifically, to stop policies from running when you are trying to ensure that purge jobs will run, you can delete policies or temporarily change them to run infrequently to give the purge job time to run. When purge jobs have caught up to their backlog of work, you can change the policies back to running more frequently. Note that simply disabling the policies is not sufficient.

More generally, if you find that scheduled purge jobs are not running because there are other Navigator tasks in progress, consider stopping Navigator extractors and setting policies to run much less frequently. Then manually run the metadata purge using an API call to match your Navigator version:

```
curl -X POST -u user:password "https://navigator_host:7187/api/vXX/maintenance/purge?deleteTimeThresholdMinutes=duration"
```

API versions correspond to Navigator versions as described [Mapping API Versions to Product Versions](#).

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7037

Policy specifications and cluster names affect purge

Policies cannot use cluster names in queries. Cluster name is a derived attribute and cannot be used as-is.

Workaround: When setting move actions for Cloudera Navigator, if there is only one cluster known to the Navigator instance, remove the `clusterName` clause.

If there is more than one cluster known to the Navigator instance, replace `clusterName` with `sourceId`. To get the `sourceId`, issue a query in this format:

```
curl '<nav-url>/api/v9/entities/?query=type%3Asource&limit=100&offset=0'
```

Use the identity of the matching HDFS service for this cluster as the `sourceId`.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3537

Spark

Spark Lineage Limitations and Requirements

Spark lineage diagrams are supported in the Cloudera Navigator 6.0 release. Spark lineage is supported for Spark 1.6 and Spark 2.3. Lineage is not available for Spark when Cloudera Manager is running in single user mode. In addition to these requirements, Spark lineage has the following limitations:

- Lineage is produced only for data that is read/written and processed using the Dataframe and SparkSQL APIs. Lineage is not available for data that is read/written or processed using Spark's RDD APIs.
- Lineage information is not produced for calls to aggregation functions such as `groupBy()`.
- The default lineage directory for Spark on Yarn is `/var/log/spark/lineage`. No process or user should write files to this directory—doing so can cause agent failures. In addition, changing the Spark on Yarn lineage directory has no effect: the default remains `/var/log/spark/lineage`.

Navigator doesn't recognize local files in Spark jobs

Spark jobs can use files on the local filesystem as job inputs or outputs. Navigator, however, only supports HFDS, Hive, and S3 assets as job inputs or outputs. When Navigator extracts metadata from Spark and encounters a local source type, the metadata is discarded and the following error appears in the Navigator Metadata Server log:

```
2018-10-11 12:14:26,192 WARN com.cloudera.nav.api.ApiExceptionMapper  
[qtp1574898980-23815]: Unexpected exception.  
java.lang.RuntimeException: Source LOCAL isn't supported for Spark Lineage
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-6811

Spark extractor enabled using safety valve deprecated

The Spark extractor included prior to CDH 5.11 and enabled by setting the safety valve, `nav.spark.extraction.enable=true` is being deprecated, and could be removed completely in a future release. If you are upgrading from CDH 5.10 or earlier and were using the extractor configured with this safety valve, be sure to remove the setting when you upgrade.

Upgrade Issues and Limitations

Upgrading Cloudera Navigator from Cloudera Manager 5.9 or Earlier Can be Extremely Slow

Upgrading a cluster running Cloudera Navigator to Cloudera Manager 5.10 (or higher) can be extremely slow due to an internal change made to the Solr schema in Cloudera Navigator 2.9. A Solr instance is embedded in Cloudera Navigator and supports its search capabilities. The Solr schema used by Cloudera Navigator has been modified in the 2.9 release to use datatype `long` rather than `string` for an internal `id` field. This change makes Cloudera Navigator far more robust and scalable over the long term.

However, the upgrade process itself can take a significantly long time because the existing Solr documents—the indexed and searchable data structures used by Solr that are contained in the Cloudera Navigator storage directory—are migrated to the new schema. This change to the Solr schema affects only those Cloudera Navigator deployments that use the metadata and lineage features.



Note: Cloudera Navigator deployments that use only the auditing features of the product—not metadata or lineage—are not affected by this issue.

The upgrade process for Cloudera Navigator starts automatically at the end of the Cloudera Manager upgrade, and the migration to the new schema occurs automatically as part of that upgrade process. The Navigator Metadata Server and Navigator console are not available during the upgrade. Navigator Audit Server runs normally. The amount of time that administrators should allow for this process depends on the quantity stored at the Navigator Metadata Server Storage Dir (`nav.data.dir`, or simply "storage directory") location as listed here:

Metadata and lineage usage	Description
None	Deployments that use Cloudera Navigator audit capability only —without metadata or lineage—do not have the issue. Backup the Navigator Metadata Server storage directory and then delete it before upgrading.
storage directory < 60 GB	Deployments with relatively small Navigator Metadata Server data directories may take 1 to 2 days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.
storage directory > 60 GB	Deployments with relatively large Navigator Metadata Server data directories may take several days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Workaround: To reduce the time required for upgrading the Navigator Metadata Server data directories for deployments currently running Cloudera Navigator 2.8 that uses its metadata and lineage features, consider removing unneeded entries from the metadata before the upgrade. The Navigator Purge feature allows you to remove metadata for deleted entities and for entities and operations older than a specified date. For more information on what metadata you can remove with Purge, see [Managing Metadata Storage with Purge](#).

Run purge before starting the Cloudera Manager upgrade (to Cloudera Manager 5.10), following the steps below.



Warning: These steps may mitigate but do not fully resolve the issue. Follow these steps **before** starting the Cloudera Manager upgrade for any Cloudera Manager 5.9 or earlier cluster that currently uses the Cloudera Navigator metadata and lineage features.

- Check the Navigator Metadata Server storage directory size. The path is `/var/lib/cloudera-scm-navigator` (default) unless configured otherwise. If you need to check the setting:
 - Log in to Cloudera Manager Admin Console.

- Select **Clusters > Cloudera Management Service**.
- Click **Configuration** and then click the **Navigator Metadata Server** Scope filter:

Navigator Metadata Server	Navigator Metadata Server (node-1)
Storage Dir	/var/lib/cloudera-scm-navigator
nav.data.dir	

- Confirm that the cluster uses the default configuration, or make a note of the location specified and the node name.
- Check the size of the actual directory contents. The following example shows a freshly installed system and so it is virtually empty.

```
[root@node-1 ~]# cd /var/lib/cloudera-scm-navigator
[root@node-1 cloudera-scm-navigator]# ls -l
total 12
drwxr-x--- 2 cloudera-scm cloudera-scm 113 Jul 12 06:56 diagnosticData
drwxr-x--- 2 cloudera-scm cloudera-scm 4096 Jul 12 09:16 extractorState
-rw-r----- 1 cloudera-scm cloudera-scm 36 Jul 12 05:54 instance.uuid
drwxr-x--- 4 cloudera-scm cloudera-scm 60 Jul 12 04:18 solr
drwxr-x--- 7 cloudera-scm cloudera-scm 4096 Jul 12 07:26 temp
[root@node-1 cloudera-scm-navigator]# cd solr
[root@node-1 solr]# ls -l
total 4
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_elements
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_relations
-rw-r----- 1 cloudera-scm cloudera-scm 450 Jul 6 16:13 solr.xml
[root@node-1 solr]#
```

- Back up the contents of the directory. Use Cloudera Manager BDR or your preferred method.
- Schedule a purge process as described in [Scheduling the Purge Process](#).



Note: Users and processes cannot access Cloudera Navigator while purge is running.

Set options to purge metadata for deleted HDFS entities and any operations.

- Check the storage directory size again. If needed, re-run the purge with a shorter time span to retain metadata. If the storage directory consumption cannot be reduced below 60GB, do not start the Cloudera Manager upgrade. Instead, contact [Cloudera support](#) to help you with this upgrade.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: NAV-5046

Cloudera Navigator 6.1.0 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the data management components of Cloudera Navigator 6.1:

New Features in Cloudera Navigator 6.1.0

The following sections describe what's new and what has changed in the data management components of Cloudera Navigator 6.1.0:

Significant improvements in metadata extraction and purge performance

HDFS metadata extraction can take up significant resources, particularly if the initial build extraction is not completed in a timely interval. Extractors running for long periods then cause the metadata purge operation to be cut short or never run at all. This release addresses these problems by significantly improving metadata extraction for HDFS and

HMS and reducing the time required to purge metadata. Together, these improvements will make your Navigator Metadata Server more robust and the resource management more predictable.

- **Incremental metadata extraction from HMS**

This release introduces incremental extraction logic for extracting metadata from Hive Metastore (HMS). The CDH 6.1.0 distribution of Hive makes available additional timing information, which Navigator takes advantage of to perform incremental extraction after the first full extraction is complete. Incremental metadata extraction has the potential to substantially improve the HMS extraction performance. With incremental extraction, Navigator can check for HMS metadata changes much more often, substantially reducing the time required for Hive metadata changes to be reflected in Navigator and Hue.

This feature is disabled by default. Cloudera testing for production data volumes is ongoing.

- **HDFS metadata extraction performance improvement**

Previously, Navigator processing for HDFS metadata extraction could take long enough that Navigator could not finish indexing the `fsimage` before the edit log was checkpointed into a new `fsimage`. This caused extraction to get stuck in a loop of always parsing the `fsimage` in bulk extraction mode (slow) rather than moving into incremental extraction from the edit log (fast). This release includes changes that allow Navigator to avoid this "bulk extraction loop," therefore reducing the duration of HDFS extraction by as much as half for the initial extraction. This change improves the performance of subsequent extractions significantly, both by making it much more likely that extraction will be incremental and from additional performance improvements in this fix.

- **Improved metadata purge performance**

This release includes changes that significantly reduce the time required to complete purging for HDFS entities and relations. The purge behavior is changed such that HDFS entities are not included in the purge if they are referenced as an endpoint in a data flow lineage relation. The performance improvement will shorten the time required to run the HDFS portion of the purge operation; combined that smaller window with the improved extraction performance and the purge jobs are much more likely to run and run to completion.

Improvements to audit extraction filters

This release introduces changes to default HDFS audit filters to better focus the collection of audit events. This change increases the scalability of your Navigator Audit Server system by removing unnecessary audit events.

The following rules were added or updated:

- All HDFS denied access events are accepted. Previously, HDFS denied access events from filtered users and filtered HDFS directories were not recorded.
- HDFS events from the Hive, Spark, and Impala staging directories are discarded. HDFS events from the following job history directories are also discarded:
 - `/user/history/done_intermediate`
 - `/user/spark/applicationHistory`
 - `/user/spark/spark2ApplicationHistory`
- All HDFS delete and rename operations from directories that are not already filtered are accepted.

In addition at the end of the filter list, Cloudera Manager provides a rule that filters events from HDFS `getfileinfo` operations. By default, this filter rule has no affect. Cloudera recommends that you change this rule to **Discard** to stop capturing events for this HDFS operation that indicates access to file *metadata* but does not indicate access to file *data*. This change alone may provide up to 30% reduction in HDFS audit data size.

When upgrading to this release, you may not see the new filters if you have customized your HDFS audit filters previously:

- If HDFS filters have been changed from the default, the new filters will NOT take effect.
- If HDFS filters have not been changed from the default, the new filters will take effect.

Whether or not the new filters are applied, you'll notice that the Cloudera Manager interface has improved controls for adding the audit filters: you have options for setting the order and for duplicating an existing filter to use it as a starting point on a new filter.

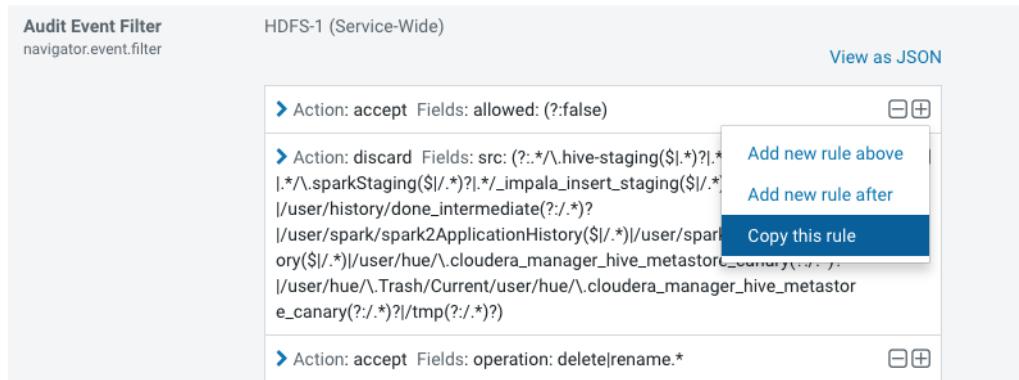


Figure 1: Audit Filter Control in Cloudera Manager Service Configuration

To review the recommended filters and determine further optimizations for your system's requirements, see [Recommended Audit Filters](#) and [Reviewing Default Audit Filters](#).

New Cloudera Manager health alerts for Navigator Metadata Server

This release includes new checks for Navigator Metadata Server health. If the checks fail, they trigger a health alert in Cloudera Manager for the Navigator Metadata Server. The checks include "Solr Element Count Threshold" and "Solr Relation Count Threshold". These checks trigger an alert if documents in Navigator Metadata Server's embedded Solr collection are threatening to exceed the maximum allowed number for either the element core or the relation core. Typically, if either of these alerts trigger, Navigator has encountered a problem that produces more relations than it should. The triggers let you know that this problem is occurring while it can be addressed efficiently.

For more information, see [Navigator Metadata Server Solr Element Count Threshold Test](#) and [Navigator Metadata Server Solr Relation Count Threshold Test](#).

Console shows column metadata in the table entity

This release includes improvements to the Navigator console, including improvements to the layout and information in the columns list. When you are viewing the detail page for a table or view, columns show inline descriptions and managed properties. A popover menu lets you see additional column metadata. The inline search in the **Column** box now matches for managed metadata and description text in addition to name and type.

Auto-suggestions in Navigator console search

The Navigator console Search field will now auto-suggest names of entities—tables, views, fields and databases. When you select one of the suggested entities, the console opens the entity details page.

This auto-complete feature also can be used with `db.table.field` search syntax.

New audit events for Fine Grained Privileges and Ownership

Navigator Audit Server receives audit events from Sentry, Hive, and Impala that are produced when an administrator or user changes ownership for a database, table, or view. For more information, see Sentry's [Object Ownership](#) and Navigator's list of [audit events for Sentry](#).

New audit events for HDFS Erasure Coding

Navigator Audit Server receives audit events from HDFS that are produced when an administrator performs a command to manage an HDFS erasure coding policy. For more information, see HDFS's [Erasure Coding](#) introduction and Navigator's list of [audit events for HDFS](#).

Issues Fixed in Cloudera Navigator 6.1.0

The following sections describe the issues fixed in the data management components of Cloudera Navigator 6.1.0:

Null Pointer Exception when processing Impala audits

Navigator Audit Server produced a Null Pointer Exception error and failed to process Impala audits. The error was similar to the following:

```
2018-09-26 01:53:25,341 ERROR
com.cloudera.nav.analytics.dataservices.etl.tasks.audits.AbstractAuditETLTask
[pool-12-thread-2]: Error encountered in executing the task New Databases ETL Task for
service impala. Will continue with remaining sources. java.lang.NullPointerException
```

This problem is fixed in this release.

Cloudera Issue: NAV-6889

Error in Navigator Audit Server log: "IOException: Unexpected length (too long)"

Audit Analytics in the Navigator console were disabled by default in the previous release; however, this change caused certain common operations to throw exceptions in the Navigator Audit Server log:

```
java.io.IOException: Unexpected length (too long): 2064261152, max:10485760
```

This issue is fixed in this release.

Cloudera Issue: NAV-6876

Purge API call produced HTTP error 500

Previously, if the purge API call is made without supplying `deleteTimeThresholdMinutes` or `runtimeCapMinutes`, Navigator Metadata Server returns an HTTP error 500. As of this release, Navigator returns an HTTP error 400.

Cloudera Issue: NAV-6818

Newlines respected in Description field

The entity Description field in the Navigator console now respects newlines and whitespace included in the field text.

Cloudera Issue: NAV-6803

Connection timeout increased in Navigator Audit Server to avoid broken pipe errors

This release of Navigator increases the connection idle time for the connection between Navigator Audit Server and each Cloudera Manager agent that sends audits to Navigator from audited services.

This change will help avoid audit pipeline "broken pipe" errors that can occur when increased load on the audit server node causes Navigator Audit Server not to respond quickly enough.

Cloudera Issue: NAV-6796

Console filter is no longer case sensitive

The audit event filter in the Navigator console was case sensitive for database or table names when audit events are stored in PostgreSQL databases. This release resolves the problem.

Cloudera Issue: NAV-6141, NAV-6795

Incremental metadata extraction from HMS

This release includes incremental extraction for HMS metadata. See [New Features](#) for more information.

Navigator left additional duplicate spark relations

This change removes all duplicate Spark data flow relations, extending the fix provided by [Navigator Metadata Server has high CPU usage, linking taking too long after Spark extraction](#).

Cloudera Issue: NAV-6751

Tables that have been deleted in Hive still show up in Navigator

Navigator metadata extraction from Hive uses an extractor state to keep track of the Hive entities committed to the Navigator Metadata Server data storage in Solr. In some cases, such as when the extractor state file is deleted or when

Cloudera Enterprise 6 Release Guide

Navigator Metadata Server is shut down during extraction from Hive, the extractor state no longer represents the state of the Hive extraction. When this happens, there may be entities listed in Navigator that are not accounted for in the extractor state. When this happens, Navigator no longer updates the Hive entities that are in Solr but not in the extractor state.

This release includes a check to make sure that the extractor state is synchronized with the content of Hive entities in Solr. To enable the check set the following property in the **Navigator Metadata Server Advanced Configuration Snippet (Safety Valve)**:

```
nav.extractor.hive.validate_state=true
```

Cloudera Issue: NAV-6745

Better error message for purging deleted properties

When purging deleted properties from the Navigator console, occasionally the error "Failed to start purge. Reason: API: Service Unavailable" appeared. The error occurred when a purge job is already scheduled and another the purge job could not be added to the queue. In this release, the message is changed to better indicate the problem: "Extractors are running. Waiting for jobs to finish before starting maintenance."

Cloudera Issue: NAV-6739

Swagger "Try it out" option for the /audits API fails

Using the interactive Navigator API for the /audits endpoint failed with the error:

```
No serializer found for class sun.net.www.protocol.http.HttpURLConnection$HttpInputStream  
and  
no properties discovered to create BeanSerializer (to avoid exception, disable  
SerializationConfig.SerializationFeature.FAIL_ON_EMPTY_BEANS) )"
```

Also, the documentation accompanying the audit API says a selector is `table` but selector is actually `table_name`.

Work-around: Change `limit` to a value less than 10000.

Cloudera Issue: NAV-6664

Support for PostgreSQL 10

Navigator Audit Server and Navigator Metadata Server support using MySQL version 10.x for their databases.

Cloudera Issue: NAV-6619

New controls for setting the order of audit filters in Cloudera Manager

The Cloudera Manager interface for adding Navigator audit filters is improved so that it is easier to control the order of the filters and to duplicate an existing filter.

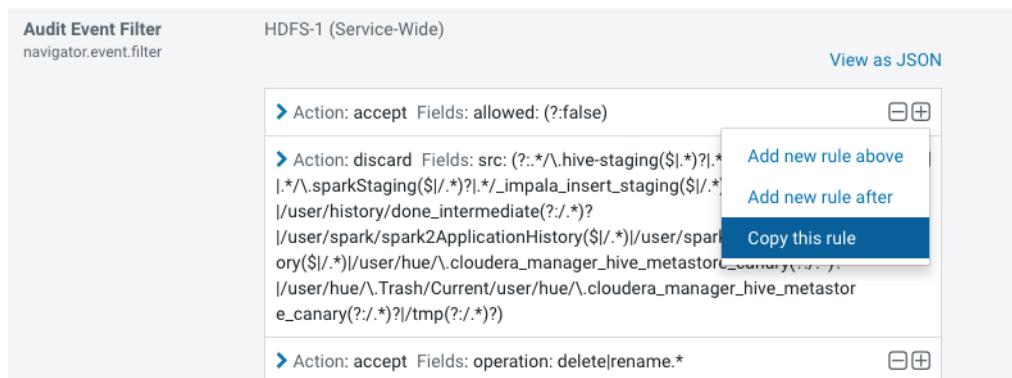


Figure 2: Audit Filter Controls in Cloudera Manager Service Configuration

Cloudera Issue: NAV-6616

Pre-registration was not working for Hive entities

Pre-registration of Hive entities (table, column, or database) fails causing exceptions in the logs such as:

```
java.lang.ClassCastException: com.cloudera.nav.core.model.GenericEntity cannot be cast
to com.cloudera.nav.hive.model.HTable
```

Cloudera Issue: NAV-6608***Navigator Metadata Server had high CPU usage, linking taking too long after Spark extraction***

In some cases, Navigator Metadata Server processing required an extraordinary amount of memory for longer than expected while extracting and linking metadata when Spark (Spark 1 or Spark 2) extractions are enabled.

A symptom of the problem is a flood of Navigator Metadata Server log messages such as:

```
Finished linking relation with id: <id>
```

This change removes duplicate relations from SPARK data flow relations that were causing the system to use the additional resources.

Cloudera Issue: NAV-6591***Reduced noise of logging for SparkLinker***

The Navigator Metadata Server logs included too much logging detail at the `INFO` level coming from the linking process that runs against Spark entities. Many of the log entries were moved to `DEBUG`.

Cloudera Issue: NAV-6560***"ConsoleAppender" error in Hive Server 2 log***

After running a query in Hive, the Hive Server 2 standard error log included the following error:

```
log4j:ERROR A "org.apache.log4j.ConsoleAppender" object is not assignable to a
"com.cloudera.navigator.shaded.log4j.Appender" variable.
log4j:ERROR The class "com.cloudera.navigator.shaded.log4j.Appender" was loaded by
log4j:ERROR [sun.misc.Launcher$AppClassLoader@47d384ee] whereas object of type
log4j:ERROR "org.apache.log4j.ConsoleAppender" was loaded by
[sun.misc.Launcher$AppClassLoader@47d384ee].
log4j:ERROR Could not instantiate appender named "out".
```

This problem occurred because the Navigator plugin for Hive Server 2 expects a different LOG4J integration than what Hive Server 2 is using. The result was that no Navigator audit messages appear in the Hive Server 2 logs. Hive Server 2 auditing was not affected by this problem.

Cloudera Issue: NAV-6523***Auto-suggestions in Navigator console search***

The Navigator console Search field will now auto-suggest names of entities—tables, views, fields and databases. When you select one of the suggested entities, the console opens the entity details page.

This auto-complete feature also can be used with `db.table.field` search syntax.

Cloudera Issue: NAV-6496***Navigator did not mark HDFS entities as deleted when bulk extraction took too long to complete***

In large HDFS deployments, the `fsimage` takes a long time to index. For Navigator, the fallout of this delay was that no HDFS entities deleted in the cluster were marked as deleted in Navigator. This problem is fixed in this release. See also [HDFS metadata extraction performance improvement](#) on page 614.

Cloudera Issue: NAV-6456

Duplicate data flow relations after Sqoop operation

After using Sqoop to transfer data to HDFS, Navigator Metadata Server may have duplicated relations between YARN OPERATION_EXECUTION entities and HDFS directory entities.

Cloudera Issue: NAV-6446

Cap on lineage relations

Navigator capped the number of lineage parents that could be returned, however this cap included lineage relations with second-level parents (columns). When reaching the limit, the Navigator Metadata Server log included an error such as Lineage Diagram is limited to 3000 entities.

Work-around: Set `nav.capacity.max_nodes_limit=1000` in **Navigator Metadata Server Advanced Configuration Snippet (Safety Valve) for cloudera-navigator.properties** configuration. Restart Navigator Metadata Server and validate that the lineage appears correctly. This workaround assumes that lineage touches a maximum of 1000 tables; increase the limit if necessary to accommodate lineage relations in your environment.

Cloudera Issue: NAV-6371

New behavior in default audit filters

This release includes new audit filters, including denied access events. See [Improvements to audit extraction filters](#) on page 609.

Cloudera Issue: NAV-6290

Improved metadata purge performance

This release significantly reduces the time required to complete purging for HDFS files. The purge behavior is changed such that HDFS entities are not included in the purge if they are referenced as an endpoint in a data flow lineage relation.

Cloudera Issue: NAV-6284

Console improvements for viewing entity metadata

This release includes improvements to the Navigator console, including improvements to the layout and information in the columns list. When you are viewing the detail page for a table or view, columns show inline descriptions and managed properties. A popover menu lets you see additional column metadata. The inline search in the **Column** box now matches for managed metadata and description text in addition to name and type.

Cloudera Issue: NAV-6234

Multiple Cluster Extractor Failure

When Navigator Metadata Server is started with a fresh storage directory on Cloudera Manager versions 5.12.0 or later, extractors were not working when a single Cloudera Manager managed multiple clusters.

This problem is fixed in this release.

Cloudera Issue: NAV-6145

New Health Alerts for Navigator Metadata Server Solr cores

This release includes new checks for Navigator Metadata Server health. See [New Cloudera Manager health alerts for Navigator Metadata Server](#) on page 610.

Cloudera Issue: NAV-6116

HDFS metadata extraction performance improvement

Navigator processing for HDFS metadata extraction can take long enough that Navigator cannot finish indexing the `fsimage` before the edit log is checkpointed into a new `fsimage`. This causes extraction to get stuck in a loop of always parsing the `fsimage` in bulk extraction mode (which is slow) rather than moving into incremental extraction from the edit log. This release includes changes that allow Navigator to avoid the bulk extraction loop, therefore reducing the duration of HDFS extraction by as much as half for the initial extraction; performance of subsequent extractions can improve significantly, benefiting both by moving to incremental extraction and by improvements in this fix.

Cloudera Issue: NAV-6033*[jQuery version upgrade for security improvements](#)*

As of this release, Navigator uses jQuery 3.x, which includes security improvements from the 2.x version. This change does not affect the behavior of the Navigator console.

Cloudera Issue: NAV-6003*[ClassCastException in Navigator Metadata Server log](#)*

When a Hive view is created, then dropped, and then subsequently recreated as a table with the same name as that of the original view, the Hive extraction process shows this exception in Navigator Metadata Server logs:

```
java.lang.ClassCastException: com.cloudera.nav.hive.model.HView cannot be cast to
com.cloudera.nav.hive.model.HTable
```

Cloudera Issue: NAV-5939*[Security headers added for Navigator resources](#)*

Security headers are now included on Navigator console resources (such as JavaScript files) so that Navigator console pages pass through security scanners without additional validation.

Cloudera Issue: NAV-5929*[Content Security Policy added for Navigator console pages](#)*

The Navigator console now includes the Content Security Policy (CSP) support, which introduces an addition layer of security for browsers that support it. See [Mozilla's Policy Browser Compatibility](#)

Cloudera Issue: NAV-5928*[Authentication failed with "javax.naming.PartialResultException" when using LDAP](#)*

Previously, if Navigator Metadata Server is configured with **External Authentication Type** "LDAP" and the LDAP server is Active Directory, when the **LDAP Group Search Base** property is configured with the root of the directory tree, authentication fails in the console with the error "Incorrect username or password."

Navigator Metadata Server logs show the following exception:

```
2018-10-24 12:21:04,619 ERROR com.cloudera.nav.auth.DelegatingNavAuthProvider
[qtp1989132530-59]:
External Authentication Service threw exception during authentication process.
org.springframework.ldap.PartialResultException: Unprocessed Continuation Reference(s);
nested exception is javax.naming.PartialResultException: Unprocessed Continuation
Reference(s);
remaining name 'DC=ad,DC=sec,DC=example,DC=com'
```

There were two common workarounds for this problem: use the Global Catalog or configure a more specific location in the directory tree for the **LDAP Group Search Base**. In this release, the root-level **LDAP Group Search Base** is correctly evaluated and no work-around is required. If you have configured one of these workarounds, you can keep the configuration as is with no change of behavior.

Cloudera Issue: NAV-5669*[New audit events for Fine Grained Privileges and Ownership](#)*

Navigator Audit Server receives audit events from Sentry, Hive, and Impala that are produced when an administrator or user changes ownership for a database, table, or view. For more information, see Sentry's [Object Ownership](#) and Navigator's list of [audit events for Sentry](#).

Cloudera Issue: CDH-65161

New audit events for HDFS Erasure Coding

Navigator Audit Server receives audit events from HDFS that are produced when an administrator performs a command to manage an HDFS erasure coding policy. For more information, see HDFS's [Erasure Coding](#) introduction and Navigator's list of [audit events for HDFS](#).

Cloudera Issue: NAV-5266

Impala and Hive audit events failed to be captured when any audit event included 4-byte characters, such as Emoji characters

This problem applies when the Navigator Audit Server database is MySQL configured to use `utf8` character set encoding.

When a query includes an Emoji or other Unicode supplementary-plane character that is encoded as four bytes in UTF-8, Navigator Audit Server failed to process the event and any following events from the same service.

For new installations, you can avoid this issue by configuring MySQL v5.5 or later to use the `utf8mb4` character set encoding. Converting existing MySQL databases for Navigator Metadata Server and Navigator Audit Server to `utf8mb4` requires updating the master tables to meet MySQL restrictions encountered after the change to 4-byte encoding. Contact Cloudera Technical Support for details.

Cloudera Issue: NAV-4845

Missing audit events for one-off processes

Navigator Audit Server did not write audit events for activity from one-off processes spawn by Cloudera Manager. The missing audit events could be for HDFS or HBase, with server log messages as follows:

```
2018-04-07 10:35:09,159 INFO com.cloudera.navigator.analytics.load.AnalyticsStagingService
[AnalyticsStagingService]: Could not load HDFS activity data for analytics
java.lang.IllegalStateException:
Number of files must be > 0
```

or

```
2018-11-29 12:34:25,616 WARN com.cloudera.navigator.NavigatorServer
[653696675@NavigatorServer-1]:
Error persisting events for service hbase java.lang.IllegalArgumentException: No property
'tableName' in
class 'com.cloudera.navigator.model.GenericAuditEvent'.
```

Cloudera Issue: NAV-4006

Audit reports Download option showed incorrect API version

The API version was not correct in the Cloudera Navigator download option for audit reports.

Cloudera Issue: NAV-3699

Easier viewing of SQL text in audit logs

For certain audit events containing SQL, the Navigator console audit entries for SQL now show a preview and an option to expand and format the SQL.

Cloudera Issue: NAV-2331

Known Issues in Cloudera Navigator 6.1.0

The following known issues and limitations from a previous release affect the data management components of Cloudera Navigator 6.1.0:

Authentication and Authorization

SAML authentication fails with "Cloudera Manager Only" setting

With the following combination of Cloudera Manager configuration properties set, authentication to Navigator fails:

- Authentication Backend Order: Cloudera Manager Only
- External Authentication Type: SAML

Workaround: To configure Navigator for SAML authentication, use an **Authentication Backend Order** other than "Cloudera Manager Only".

Authentication Backend Order nav.auth.backend.order	Navigator Metadata Server Default Group  <input checked="" type="radio"/> External Only <input type="radio"/> External then Cloudera Manager <input type="radio"/> Cloudera Manager Only <input type="radio"/> Cloudera Manager then External
---	--

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-6211

Cloudera Manager Configuration

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

At this time, there is no workaround.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-5331

Hive, Hue, Impala

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

Affected Versions: Cloudera Navigator 6.0.0 and later

Workaround: To fix this problem, manually merge the original HiveServer2 safety valve content for `hive.exec.post.hooks` with the new value. For example, in the case of Unravel, the new safety valve would look like the following:

```
<property>
  <name>hive.exec.post.hooks</name>
  <value>com.unaveldata.flowhive.hivePostHook,com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger</value>
  <description>for Unravel, from unraveldata.com</description>
</property>
```

Cloudera Issue: NAV-5331

Viewing Navigator tags in Hue overloads Metadata Server heap

When viewing Cloudera Navigator tags through Hue, Navigator uses more memory than usual and does not release the memory after logging out of Hue. Eventually, the calls between Hue and Navigator will occupy the majority of the heap space allocated to Navigator Metadata Server.

Workaround: Restart the Navigator Metadata Server periodically to clear the heap usage.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4326

Lineage not generated for Pig operations on Hive tables using HCatalog loader

When accessing a Hive table using Pig, lineage is generated in Navigator when using physical file loads, such as:

```
A = LOAD '/user/hive/warehouse/navigator_demo.db/salesdata';
B = LIMIT A 16;
STORE B INTO '/user/hive/warehouse/navigator_demo.db/salesdata_sample_file' using
PigStorage (';');
```

However, when accessing the Hive table using the HCatalog load, lineage for the Pig operation is not generated when browsing the source table lineage. Such as:

```
A = LOAD 'navigator_demo.salesdata' using org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 16;
STORE B INTO 'navigator_demo.salesdata_sample_hcatalog' using
org.apache.hive.hcatalog.pig.HCatStorer();
```

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3411

HiveServer1 and Hive CLI support removed

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

If you use Cloudera Navigator to capture auditing, lineage, and metadata for Hive operations, upgrade to HiveServer2 if you have not done so already.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: TSB-185

Navigator Audit Server

Logging Threshold setting is not honored

The value for Navigator Audit Server Logging Threshold found in Cloudera Manager is not honored. Instead, messages are logged at trace level and displayed at DEBUG syslog level. This configuration property is set in **Cloudera Management Service > Configuration > Navigator Audit Server**.

Cloudera Issue: NAV-3737

Navigator Metadata Server

International characters not supported in tags or property names

Navigator tags and the key portion of user-defined and managed properties do not support UNICODE characters beyond ASCII. Only ASCII text can be used in the text of a tag or the name of a property. Property values can include international characters.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: Tags are fixed in Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7011, NAV-7044

Navigator Embedded Solr can reach its limit on number of documents it can store



Important: This issue is critical when upgrading Cloudera Navigator deployments from Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0.

Navigator Metadata Server extracts HDFS entities by performing a one-time bulk extraction and then switching to incremental extraction. In Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0 (Navigator releases 2.9.0, 2.9.1, and 2.10.0), a problem causes HDFS bulk extraction to be run more than one time, resulting in duplicate relations created for HDFS. Over time, embedded Solr runs out of document IDs that it can assign to new relations and fails with following error:

```
"Caused by: java.lang.IllegalArgumentException: Too many documents, composite IndexReaders cannot exceed 2147483519"
```

When this happens, Navigator stops any more extraction of data as no new documents can be added to Solr.

After upgrading to this release, there is an additional recover step as described in "Repairing metadata in the storage directory after upgrading" in [Troubleshooting Navigator Data Management](#).

Affected Versions: Versions prior to Cloudera Manager 5.10 upgraded to Cloudera Manager 5.10 or higher

Fixed Versions: N/A

Cloudera Issue: NAV-5600

Log includes the error "EndPoint1 must not be null"

The following error may appear in the Navigator Metadata Server log in systems upgraded from Cloudera Manager version 5.x:

```
2017-10-17 13:00:23,007 ERROR com.cloudera.nav.hive.extractor.AbstractHiveExtractor [CDHExecutor-0-CDHUrlClassLoader@14784b7b]: Unable to parse hive view query *: EndPoint1 must not be null or empty
java.lang.IllegalStateException: EndPoint1 must not be null or empty
```

This error occurs because the Hive pull extraction for creating a Hive view produces an incorrect lineage relationship for the Hive view. However, Navigator also receives information for the view creation through the push extractor, which correctly produces the lineage relation. You can safely ignore this error.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4224

Purge

Navigator Metadata Server purge jobs may not run if there are policies configured

Navigator Metadata Server purge can produce messages such as "Checking if maintenance is running" and then fail to run during the available window. This problem occurs when a scheduled purge job waits for extraction tasks to finish but while waiting, a policy job starts, preventing the purge job from running.

Workaround: Specifically, to stop policies from running when you are trying to ensure that purge jobs will run, you can delete policies or temporarily change them to run infrequently to give the purge job time to run. When purge jobs

have caught up to their backlog of work, you can change the policies back to running more frequently. Note that simply disabling the policies is not sufficient.

More generally, if you find that scheduled purge jobs are not running because there are other Navigator tasks in progress, consider stopping Navigator extractors and setting policies to run much less frequently. Then manually run the metadata purge using an API call to match your Navigator version:

```
curl -X POST -u user:password  
"https://navigator_host:7187/api/vXX/maintenance/purge?deleteTimeThresholdMinutes=duration"
```

API versions correspond to Navigator versions as described [Mapping API Versions to Product Versions](#).

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7037

Policy specifications and cluster names affect purge

Policies cannot use cluster names in queries. Cluster name is a derived attribute and cannot be used as-is.

Workaround: When setting move actions for Cloudera Navigator, if there is only one cluster known to the Navigator instance, remove the `clusterName` clause.

If there is more than one cluster known to the Navigator instance, replace `clusterName` with `sourceId`. To get the `sourceId`, issue a query in this format:

```
curl '<nav-url>/api/v9/entities/?query=type%3Asource&limit=100&offset=0'
```

Use the identity of the matching HDFS service for this cluster as the `sourceId`.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3537

Spark

Spark Lineage Limitations and Requirements

Spark lineage diagrams are supported in the Cloudera Navigator 6.0 release. Spark lineage is supported for Spark 1.6 and Spark 2.3. Lineage is not available for Spark when Cloudera Manager is running in single user mode. In addition to these requirements, Spark lineage has the following limitations:

- Lineage is produced only for data that is read/written and processed using the Dataframe and SparkSQL APIs. Lineage is not available for data that is read/written or processed using Spark's RDD APIs.
- Lineage information is not produced for calls to aggregation functions such as `groupBy()`.
- The default lineage directory for Spark on Yarn is `/var/log/spark/lineage`. No process or user should write files to this directory—doing so can cause agent failures. In addition, changing the Spark on Yarn lineage directory has no effect: the default remains `/var/log/spark/lineage`.

Navigator doesn't recognize local files in Spark jobs

Spark jobs can use files on the local filesystem as job inputs or outputs. Navigator, however, only supports HFDS, Hive, and S3 assets as job inputs or outputs. When Navigator extracts metadata from Spark and encounters a local source type, the metadata is discarded and the following error appears in the Navigator Metadata Server log:

```
2018-10-11 12:14:26,192 WARN com.cloudera.nav.api.ApiExceptionMapper  
[qtp1574898980-23815]: Unexpected exception.  
java.lang.RuntimeException: Source LOCAL isn't supported for Spark Lineage
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-6811

Spark extractor enabled using safety valve deprecated

The Spark extractor included prior to CDH 5.11 and enabled by setting the safety valve, `nav.spark.extraction.enable=true` is being deprecated, and could be removed completely in a future release. If you are upgrading from CDH 5.10 or earlier and were using the extractor configured with this safety valve, be sure to remove the setting when you upgrade.

Upgrade Issues and Limitations

Upgrading Cloudera Navigator from Cloudera Manager 5.9 or Earlier Can be Extremely Slow

Upgrading a cluster running Cloudera Navigator to Cloudera Manager 5.10 (or higher) can be extremely slow due to an internal change made to the Solr schema in Cloudera Navigator 2.9. A Solr instance is embedded in Cloudera Navigator and supports its search capabilities. The Solr schema used by Cloudera Navigator has been modified in the 2.9 release to use datatype `long` rather than `string` for an internal `id` field. This change makes Cloudera Navigator far more robust and scalable over the long term.

However, the upgrade process itself can take a significantly long time because the existing Solr documents—the indexed and searchable data structures used by Solr that are contained in the Cloudera Navigator storage directory—are migrated to the new schema. This change to the Solr schema affects only those Cloudera Navigator deployments that use the metadata and lineage features.



Note: Cloudera Navigator deployments that use only the auditing features of the product—not metadata or lineage—are not affected by this issue.

The upgrade process for Cloudera Navigator starts automatically at the end of the Cloudera Manager upgrade, and the migration to the new schema occurs automatically as part of that upgrade process. The Navigator Metadata Server and Navigator console are not available during the upgrade. Navigator Audit Server runs normally. The amount of time that administrators should allow for this process depends on the quantity stored at the Navigator Metadata Server Storage Dir (`nav.data.dir`, or simply "storage directory") location as listed here:

Metadata and lineage usage	Description
None	Deployments that use Cloudera Navigator audit capability only —without metadata or lineage—do not have the issue. Backup the Navigator Metadata Server storage directory and then delete it before upgrading.
storage directory < 60 GB	Deployments with relatively small Navigator Metadata Server data directories may take 1 to 2 days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.
storage directory > 60 GB	Deployments with relatively large Navigator Metadata Server data directories may take several days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Workaround: To reduce the time required for upgrading the Navigator Metadata Server data directories for deployments currently running Cloudera Navigator 2.8 that uses its metadata and lineage features, consider removing unneeded entries from the metadata before the upgrade. The Navigator Purge feature allows you to remove metadata for deleted entities and for entities and operations older than a specified date. For more information on what metadata you can remove with Purge, see [Managing Metadata Storage with Purge](#).

Run purge before starting the Cloudera Manager upgrade (to Cloudera Manager 5.10), following the steps below.



Warning: These steps may mitigate but do not fully resolve the issue. Follow these steps **before** starting the Cloudera Manager upgrade for any Cloudera Manager 5.9 or earlier cluster that currently uses the Cloudera Navigator metadata and lineage features.

- Check the Navigator Metadata Server storage directory size. The path is `/var/lib/cloudera-scm-navigator` (default) unless configured otherwise. If you need to check the setting:
 - Log in to Cloudera Manager Admin Console.
 - Select **Clusters > Cloudera Management Service**.
 - Click **Configuration** and then click the **Navigator Metadata Server** Scope filter:

Navigator Metadata Server	Navigator Metadata Server (node-1)
Storage Dir	/var/lib/cloudera-scm-navigator
nav.data.dir	

- Confirm that the cluster uses the default configuration, or make a note of the location specified and the node name.
- Check the size of the actual directory contents. The following example shows a freshly installed system and so it is virtually empty.

```
[root@node-1 ~]# cd /var/lib/cloudera-scm-navigator
[root@node-1 cloudera-scm-navigator]# ls -l
total 12
drwxr-x--- 2 cloudera-scm cloudera-scm 113 Jul 12 06:56 diagnosticData
drwxr-x--- 2 cloudera-scm cloudera-scm 4096 Jul 12 09:16 extractorState
-rw-r----- 1 cloudera-scm cloudera-scm 36 Jul 12 05:54 instance.uuid
drwxr-x--- 4 cloudera-scm cloudera-scm 60 Jul 12 04:18 solr
drwxr-x--- 7 cloudera-scm cloudera-scm 4096 Jul 12 07:26 temp
[root@node-1 cloudera-scm-navigator]# cd solr
[root@node-1 solr]# ls -l
total 4
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_elements
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_relations
-rw-r----- 1 cloudera-scm cloudera-scm 450 Jul 6 16:13 solr.xml
[root@node-1 solr]#
```

- Back up the contents of the directory. Use Cloudera Manager BDR or your preferred method.
- Schedule a purge process as described in [Scheduling the Purge Process](#).



Note: Users and processes cannot access Cloudera Navigator while purge is running.

Set options to purge metadata for deleted HDFS entities and any operations.

- Check the storage directory size again. If needed, re-run the purge with a shorter time span to retain metadata. If the storage directory consumption cannot be reduced below 60GB, do not start the Cloudera Manager upgrade. Instead, contact [Cloudera support](#) to help you with this upgrade.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: NAV-5046

Cloudera Navigator 6.0.x Data Management Release Notes

To view release notes for the data management components of specific Cloudera Navigator 6.0.x releases, see the following:

Cloudera Navigator 6.0.1 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the data management components of Cloudera Navigator 6.0.1:

New Features in Cloudera Navigator 6.0.1

There are no new features in the data management components of Cloudera Navigator. See also [Known Issues in Cloudera Navigator 6.0.1](#) on page 623.

Issues Fixed in Cloudera Navigator 6.0.1

There are no fixed issues in the data management components of Cloudera Navigator. See also [Known Issues in Cloudera Navigator 6.0.1](#) on page 623.

Known Issues in Cloudera Navigator 6.0.1

The following known issues and limitations from a previous release affect the data management components of Cloudera Navigator 6.0.1:

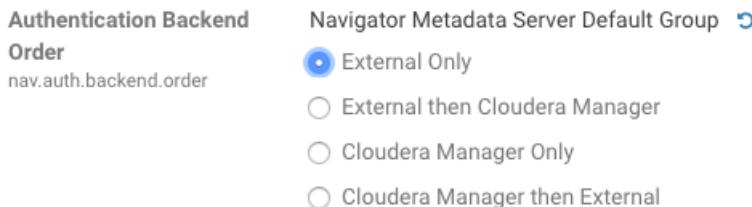
Authentication and Authorization

SAML authentication fails with "Cloudera Manager Only" setting

With the following combination of Cloudera Manager configuration properties set, authentication to Navigator fails:

- Authentication Backend Order: Cloudera Manager Only
- External Authentication Type: SAML

Workaround: To configure Navigator for SAML authentication, use an **Authentication Backend Order** other than "Cloudera Manager Only".



Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-6211

Cloudera Manager Configuration

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

At this time, there is no workaround.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-5331

Cloudera Manager audit events case-sensitive when using PostgreSQL

When Cloudera Manager and Navigator Audit Server are installed using PostgreSQL databases, the behavior of queries run from the Navigator console is different between the two databases. The result is that Cloudera Manager events

are returned only if the query values match the case of the event values as they are stored in the Cloudera Manager database.

For example, the Hive operation "HiveReplicationCommand" is audited by Cloudera Manager; the audit log shows the command as HIVEREPLICATIONCOMMAND but querying with upper case fails to return the corresponding audit events. However, querying as operation = HiveReplicationCommand does return results.

Audit events other than those for Cloudera Manager are not affected.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-6141, NAV-6795

Hive, Hue, Impala

ClassCastException in Navigator Metadata Server log

When a Hive view is created, then dropped, and then subsequently recreated as a table with the same name as that of the original view, the Hive extraction process shows this exception in Navigator Metadata Server logs:

```
java.lang.ClassCastException: com.cloudera.nav.hive.model.HView cannot be cast to  
com.cloudera.nav.hive.model.HTable
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-5939

ConsoleAppender error in Hive Server 2 log

After running a query in Hive, the HiveServer2 stderr log includes the following error:

```
log4j:ERROR A "org.apache.log4j.ConsoleAppender" object is not assignable to a  
"com.cloudera.navigator.shaded.log4j.Appender" variable.  
log4j:ERROR The class "com.cloudera.navigator.shaded.log4j.Appender" was loaded by  
log4j:ERROR [sun.misc.Launcher$AppClassLoader@47d384ee] whereas object of type  
log4j:ERROR "org.apache.log4j.ConsoleAppender" was loaded by  
[sun.misc.Launcher$AppClassLoader@47d384ee].  
log4j:ERROR Could not instantiate appender named "out".
```

This problem occurs because the Navigator Audit Server plugin for Hive Server 2 expects a different log4j integration than what HiveServer 2 is using. The result is that no Navigator Audit messages appear in the Hive Server 2 logs. Hive Server 2 auditing is not affected by this problem.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-6523

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to  
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'  
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

Affected Versions: Cloudera Navigator 6.0.0 and later

Workaround: To fix this problem, manually merge the original HiveServer2 safety valve content for `hive.exec.post.hooks` with the new value. For example, in the case of Unravel, the new safety valve would look like the following:

```
<property>
  <name>hive.exec.post.hooks</name>
  <value>com.navadata.flowhive.hive.hivePostHook,com.cloudera.navigator.audit.hive.HiveAuditEventAggregation.hiveql_hooks.LineageLogger</value>
  <description>for Unravel, from unraveldata.com</description>
</property>
```

Cloudera Issue: NAV-5331

Impala and Hive audit events fail to be captured when one audit event includes 4-byte characters, such as an emoji

This problem applies when the Navigator Audit Server database is a MySQL database configured to use the "UTF8" character set.

When a query includes an emoji or other Unicode supplementary-plane character that is encoded as four bytes in UTF-8, Navigator Audit Server fails to process the event and any following events from the same service.

Workaround: You can resolve this problem by configuring MySQL v5.5 or later to use the "UTF8MB4" character set. The error is described in this Stack Overflow article:

<https://stackoverflow.com/questions/13653712/java-sql-sqlexception-incorrect-string-value-xf0-x9f-x91-xbd-xf0-x9f>

The solution is described in the MySQL documentation topic [The UTF8MB4 Character Set \(4-Byte UTF-8 Unicode Encoding\)](#). Changing the character set requires restarting the MySQL server. It doesn't affect the Navigator Audit Server data.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-4845

Viewing Navigator tags in Hue overloads Metadata Server heap

When viewing Cloudera Navigator tags through Hue, Navigator uses more memory than usual and does not release the memory after logging out of Hue. Eventually, the calls between Hue and Navigator will occupy the majority of the heap space allocated to Navigator Metadata Server.

Workaround: Restart the Navigator Metadata Server periodically to clear the heap usage.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4326

Lineage not generated for Pig operations on Hive tables using HCatalog loader

When accessing a Hive table using Pig, lineage is generated in Navigator when using physical file loads, such as:

```
A = LOAD '/user/hive/warehouse/navigator_demo.db/salesdata';
B = LIMIT A 16;
STORE B INTO '/user/hive/warehouse/navigator_demo.db/salesdata_sample_file' using
PigStorage ('');
```

However, when accessing the Hive table using the HCatalog load, lineage for the Pig operation is not generated when browsing the source table lineage. Such as:

```
A = LOAD 'navigator_demo.salesdata' using org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 16;
STORE B INTO 'navigator_demo.salesdata_sample_hcatalog' using
org.apache.hive.hcatalog.pig.HCatStorer();
```

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3411

HiveServer1 and Hive CLI support removed

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

If you use Cloudera Navigator to capture auditing, lineage, and metadata for Hive operations, upgrade to HiveServer2 if you have not done so already.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: TSB-185

Navigator Audit Server

Logging Threshold setting is not honored

The value for Navigator Audit Server Logging Threshold found in Cloudera Manager is not honored. Instead, messages are logged at trace level and displayed at DEBUG syslog level. This configuration property is set in **Cloudera Management Service > Configuration > Navigator Audit Server**.

Cloudera Issue: NAV-3737

Audit reports download option shows incorrect API version

The API version is not correct in the Cloudera Navigator download option for audit reports.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-3699

Pending audit events may be lost after role migration

Pending and deleted audit events may be lost during role migration. Pending audit events are those events that have not yet transferred from a source node through the Cloudera Navigator plug-in for the respective role to the Navigator Audit Server. If the Navigator Audit Server is unreachable due to network issues or other issues, the pending audits—which should automatically transfer at the conclusion of the role migration—may be lost.

Workaround: Before migrating a role to a different node:

- Check the state of the Navigator Audit Server (using Cloudera Manager Admin Console).
- Proceed with the role migration only if Navigator Audit Server is running.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-1006

Audit CSV has extra columns and is missing some data

Audit details exported as CSV files do not contain Sentry data. In addition, some of the column names display twice (Operation Text, Database Name, Object Type, for example), although the actual details display only of the duplicate columns.

Workaround: Export audits to JSON format to obtain Sentry data.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-708*Navigator Metadata Server*

Navigator does not mark HDFS entities as deleted when in bulk extraction takes too long to complete

In large HDFS deployments, the `fsimage` takes a long time to index. When an HDFS checkpoint occurs it creates a new `fsimage`. However if the previous `fsimage` is still in the process of being indexed, Navigator cannot use the incremental changes found in the `inotify` stream because it refers to the newly created `fsimage`.

When this happens, Navigator attempts to start indexing the newer `fsimage`, creating a loop where Navigator can never take advantage of the more efficient change processing through `inotify`. The immediate fallout of this delay is that no HDFS entities deleted in the cluster will be marked as deleted in Navigator.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-6456

International characters not supported in tags or property names

Navigator tags and the key portion of user-defined and managed properties do not support UNICODE characters beyond ASCII. Only ASCII text can be used in the text of a tag or the name of a property. Property values can include international characters.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: Tags are fixed in Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7011, NAV-7044

Navigator Embedded Solr can reach its limit on number of documents it can store



Important: This issue is critical when upgrading Cloudera Navigator deployments from Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0.

Navigator Metadata Server extracts HDFS entities by performing a one-time bulk extraction and then switching to incremental extraction. In Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0 (Navigator releases 2.9.0, 2.9.1, and 2.10.0), a problem causes HDFS bulk extraction to be run more than one time, resulting in duplicate relations created for HDFS. Over time, embedded Solr runs out of document IDs that it can assign to new relations and fails with following error:

"Caused by: java.lang.IllegalArgumentException: Too many documents, composite IndexReaders cannot exceed 2147483519"

When this happens, Navigator stops any more extraction of data as no new documents can be added to Solr.

After upgrading to this release, there is an additional recover step as described in "Repairing metadata in the storage directory after upgrading" in [Troubleshooting Navigator Data Management](#).

Affected Versions: Versions prior to Cloudera Manager 5.10 upgraded to Cloudera Manager 5.10 or higher

Fixed Versions: N/A

Cloudera Issue: NAV-5600

Log includes the error "EndPoint1 must not be null"

The following error may appear in the Navigator Metadata Server log in systems upgraded from Cloudera Manager version 5.x:

```
2017-10-17 13:00:23,007 ERROR com.cloudera.nav.hive.extractor.AbstractHiveExtractor
[CDHExecutor-0-CDHUrlClassLoader@14784b7b]: Unable to parse hive view query *: EndPoint1
must not be null or empty
java.lang.IllegalStateException: EndPoint1 must not be null or empty
```

This error occurs because the Hive pull extraction for creating a Hive view produces an incorrect lineage relationship for the Hive view. However, Navigator also receives information for the view creation through the push extractor, which correctly produces the lineage relation. You can safely ignore this error.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4224

Purge

Navigator Metadata Server purge jobs may not run if there are policies configured

Navigator Metadata Server purge can produce messages such as "Checking if maintenance is running" and then fail to run during the available window. This problem occurs when a scheduled purge job waits for extraction tasks to finish but while waiting, a policy job starts, preventing the purge job from running.

Workaround: Specifically, to stop policies from running when you are trying to ensure that purge jobs will run, you can delete policies or temporarily change them to run infrequently to give the purge job time to run. When purge jobs have caught up to their backlog of work, you can change the policies back to running more frequently. Note that simply disabling the policies is not sufficient.

More generally, if you find that scheduled purge jobs are not running because there are other Navigator tasks in progress, consider stopping Navigator extractors and setting policies to run much less frequently. Then manually run the metadata purge using an API call to match your Navigator version:

```
curl -X POST -u user:password  
"https://navigator_host:7187/api/vXX/maintenance/purge?deleteTimeThresholdMinutes=duration"
```

API versions correspond to Navigator versions as described [Mapping API Versions to Product Versions](#).

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7037

Policy specifications and cluster names affect purge

Policies cannot use cluster names in queries. Cluster name is a derived attribute and cannot be used as-is.

Workaround: When setting move actions for Cloudera Navigator, if there is only one cluster known to the Navigator instance, remove the `clusterName` clause.

If there is more than one cluster known to the Navigator instance, replace `clusterName` with `sourceId`. To get the `sourceId`, issue a query in this format:

```
curl '<nav-url>/api/v9/entities/?query=type%3Asource&limit=100&offset=0'
```

Use the identity of the matching HDFS service for this cluster as the `sourceId`.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3537

Spark

Spark Lineage Limitations and Requirements

Spark lineage diagrams are supported in the Cloudera Navigator 6.0 release. Spark lineage is supported for Spark 1.6 and Spark 2.3. Lineage is not available for Spark when Cloudera Manager is running in single user mode. In addition to these requirements, Spark lineage has the following limitations:

- Lineage is produced only for data that is read/written and processed using the Dataframe and SparkSQL APIs. Lineage is not available for data that is read/written or processed using Spark's RDD APIs.
- Lineage information is not produced for calls to aggregation functions such as `groupBy()`.

- The default lineage directory for Spark on Yarn is `/var/log/spark/lineage`. No process or user should write files to this directory—doing so can cause agent failures. In addition, changing the Spark on Yarn lineage directory has no effect: the default remains `/var/log/spark/lineage`.

Navigator doesn't recognize local files in Spark jobs

Spark jobs can use files on the local filesystem as job inputs or outputs. Navigator, however, only supports HFDS, Hive, and S3 assets as job inputs or outputs. When Navigator extracts metadata from Spark and encounters a local source type, the metadata is discarded and the following error appears in the Navigator Metadata Server log:

```
2018-10-11 12:14:26,192 WARN com.cloudera.nav.api.ApiExceptionMapper
[qtp1574898980-23815]: Unexpected exception.
java.lang.RuntimeException: Source LOCAL isn't supported for Spark Lineage
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-6811

Spark extractor enabled using safety valve deprecated

The Spark extractor included prior to CDH 5.11 and enabled by setting the safety valve, `nav.spark.extraction.enable=true` is being deprecated, and could be removed completely in a future release. If you are upgrading from CDH 5.10 or earlier and were using the extractor configured with this safety valve, be sure to remove the setting when you upgrade.

Upgrade Issues and Limitations

Upgrading Cloudera Navigator from Cloudera Manager 5.9 or Earlier Can be Extremely Slow

Upgrading a cluster running Cloudera Navigator to Cloudera Manager 5.10 (or higher) can be extremely slow due to an internal change made to the Solr schema in Cloudera Navigator 2.9. A Solr instance is embedded in Cloudera Navigator and supports its search capabilities. The Solr schema used by Cloudera Navigator has been modified in the 2.9 release to use datatype `long` rather than `string` for an internal `id` field. This change makes Cloudera Navigator far more robust and scalable over the long term.

However, the upgrade process itself can take a significantly long time because the existing Solr documents—the indexed and searchable data structures used by Solr that are contained in the Cloudera Navigator storage directory—are migrated to the new schema. This change to the Solr schema affects only those Cloudera Navigator deployments that use the metadata and lineage features.



Note: Cloudera Navigator deployments that use only the auditing features of the product—not metadata or lineage—are not affected by this issue.

The upgrade process for Cloudera Navigator starts automatically at the end of the Cloudera Manager upgrade, and the migration to the new schema occurs automatically as part of that upgrade process. The Navigator Metadata Server and Navigator console are not available during the upgrade. Navigator Audit Server runs normally. The amount of time that administrators should allow for this process depends on the quantity stored at the Navigator Metadata Server Storage Dir (`nav.data.dir`, or simply "storage directory") location as listed here:

Metadata and lineage usage	Description
None	Deployments that use Cloudera Navigator audit capability only —without metadata or lineage—do not have the issue. Backup the Navigator Metadata Server storage directory and then delete it before upgrading.
storage directory < 60 GB	Deployments with relatively small Navigator Metadata Server data directories may take 1 to 2 days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Metadata and lineage usage	Description
storage directory > 60 GB	Deployments with relatively large Navigator Metadata Server data directories may take several days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Workaround: To reduce the time required for upgrading the Navigator Metadata Server data directories for deployments currently running Cloudera Navigator 2.8 that uses its metadata and lineage features, consider removing unneeded entries from the metadata before the upgrade. The Navigator Purge feature allows you to remove metadata for deleted entities and for entities and operations older than a specified date. For more information on what metadata you can remove with Purge, see [Managing Metadata Storage with Purge](#).

Run purge before starting the Cloudera Manager upgrade (to Cloudera Manager 5.10), following the steps below.



Warning: These steps may mitigate but do not fully resolve the issue. Follow these steps **before** starting the Cloudera Manager upgrade for any Cloudera Manager 5.9 or earlier cluster that currently uses the Cloudera Navigator metadata and lineage features.

- Check the Navigator Metadata Server storage directory size. The path is `/var/lib/cloudera-scm-navigator` (default) unless configured otherwise. If you need to check the setting:
 - Log in to Cloudera Manager Admin Console.
 - Select **Clusters > Cloudera Management Service**.
 - Click **Configuration** and then click the **Navigator Metadata Server** Scope filter:

Navigator Metadata Server	Navigator Metadata Server (node-1)
Storage Dir	/var/lib/cloudera-scm-navigator
nav.data.dir	

- Confirm that the cluster uses the default configuration, or make a note of the location specified and the node name.

- Check the size of the actual directory contents. The following example shows a freshly installed system and so it is virtually empty.

```
[root@node-1 ~]# cd /var/lib/cloudera-scm-navigator
[root@node-1 cloudera-scm-navigator]# ls -l
total 12
drwxr-x--- 2 cloudera-scm cloudera-scm 113 Jul 12 06:56 diagnosticData
drwxr-x--- 2 cloudera-scm cloudera-scm 4096 Jul 12 09:16 extractorState
-rw-r----- 1 cloudera-scm cloudera-scm 36 Jul 12 05:54 instance.uuid
drwxr-x--- 4 cloudera-scm cloudera-scm 60 Jul 12 04:18 solr
drwxr-x--- 7 cloudera-scm cloudera-scm 4096 Jul 12 07:26 temp
[root@node-1 cloudera-scm-navigator]# cd solr
[root@node-1 solr]# ls -l
total 4
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_elements
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_relations
-rw-r----- 1 cloudera-scm cloudera-scm 450 Jul 6 16:13 solr.xml
[root@node-1 solr]#
```

- Back up the contents of the directory. Use Cloudera Manager BDR or your preferred method.
- Schedule a purge process as described in [Scheduling the Purge Process](#).



Note: Users and processes cannot access Cloudera Navigator while purge is running.

Set options to purge metadata for deleted HDFS entities and any operations.

- Check the storage directory size again. If needed, re-run the purge with a shorter time span to retain metadata. If the storage directory consumption cannot be reduced below 60GB, do not start the Cloudera Manager upgrade. Instead, contact [Cloudera support](#) to help you with this upgrade.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: NAV-5046

Cloudera Navigator 6.0.0 Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the data management components of Cloudera Navigator 6.0.0:

New Features in Cloudera Navigator 6.0.0

The following sections describe what's new and what has changed in the data management components of Cloudera Navigator 6.0.0:

HDFS Analytics disabled by default

The HDFS Analytics available in the Navigator console is now disabled by default.

To enable the HDFS Analytics, set the following properties in the Cloudera Management Service configuration:

- Set navigator.analytics.enabled to be true in the **Navigator Audit Server Advanced Configuration Snippet (Safety Valve) for cloudera-navigator.properties**.
- Set nav.analytics.audit.enabled to be true in the **Navigator Metadata Server Advanced Configuration Snippet (Safety Valve) for cloudera-navigator.properties**.

Cloudera Issue: NAV-6372

Navigator Metadata is now called "Properties"

Some of the Navigator terminology has changed in this release.

Old Term	New Term
Metadata	Property
Managed Metadata	Managed Property
Custom Metadata	User-defined Property

"User-defined metadata" continues to include properties and tags; we encourage customers to migrate from tags to managed properties. For more information, see [Navigator Business Metadata](#).

Cloudera Issue: NAV-6149

Pre-register metadata for operations

Applications can now pre-register operations in Navigator Metadata Server using a POST /entities REST call with a generated operation source ID. For details on generating the ID before making the POST call, see the Navigator SDK example:

<https://github.com/cloudera/navigator-sdk/blob/master/examples/src/main/java/com/cloudera/nav/sdk/examples/hivelineage/CustomHiveLineageCreator.java#L50>

Cloudera Issue: NAV-5940

Navigator version numbering now corresponds to Cloudera Manager versions

As of Cloudera Manager version 6, Cloudera Navigator version numbering tracks with the Cloudera Manager numbering. In this new system, the release that follows Navigator 2.x will be Navigator 6.x. Specifically, functionality from Navigator 2.13

Navigator releases associated with Cloudera Manager 5.x will continue with the previous numbering sequence.

Cloudera Issue: NAV-5368

Hive Metastore instances identified using UUIDs

Navigator used a combination of the Hive HMS database connection URL and driver name configuration values to uniquely identify the HMS and entities source from that HMS. This way of identifying the Hive source could change unexpectedly in cases where the database is migrated to another server.

Fresh installations of Navigator 6.x use the UUID in the metastore database as a unique identifier. In this way, multiple HMS instances can be uniquely identified even in cases where the data is coming from the same backing database.

For new installations, the Hive database UUID will identify the HMS sources. However, Navigator will continue to use the database URL as the Hive source identifier when deployments are upgraded from the Navigator from Cloudera Manager 5.x releases.

Cloudera Issue: NAV-4602

Preventing concurrent logins

An option is now available to limit the number of simultaneous sessions authenticated against the same user name. The property `nav.max.concurrent.sessions` takes an integer value as a limit; set the value to -1 (default) to turn off the limit. To change the default behavior, in Cloudera Manager, add the property with a new value to the "Navigator Metadata Server Advanced Configuration Snippet (Safety Valve) for `cloudera-navigator.properties`." Restart the Navigator Metadata Server to apply the change.

Cloudera Issue: NAV-6020

Name and description lengths increased

This release increases the number of characters stored for entity names and descriptions in Navigator Metadata Server:

- Name length was 40 characters and is now 500.
- Description length was 500 and is now not limited.

Cloudera Issue: NAV-6228

Issues Fixed in Cloudera Navigator 6.0.0

The following sections describe the issues fixed in the data management components of Cloudera Navigator 6.0.0:

Trust Store password encrypted

In previous releases, when Cloudera Manager starts Navigator Metadata Server role, it passed trust store password un-encrypted. In this release, when configured to use TLS, the trust store password is encrypted when passed between Cloudera Manager and Navigator.

Cloudera Issue: NAV-6382

Automated clean-up when Navigator Metadata Server storage directory is cleared

When Navigator Metadata Server role starts, it now checks if the Solr storage directory is empty or not. If it is empty, this check resets the ordinal numbers and starts a new Navigator maintenance history record in the Navigator Metadata Server database.

Cloudera Issue: NAV-6267

Web Server and Platform version disclosure removed

This release includes a change to omit the server and platform versions from the web server response header.

Cloudera Issue: NAV-6019

Security headers improved with Cache-Control

This release includes Cache-Control cookie headers with the setting "no-cache" to ensure that no browser content is cached. Navigator Metadata Server, the web server for the Navigator console, already uses the no-cache Cache-Control header for API interactions. This change extends the security coverage to additional content such as JavaScript.

Cloudera Issue: NAV-6018

Less Secure Credentials Protection Policy could expose Azure credentials in audit logs

When you use Cloudera Manager to configure the ADLS Connector service using the Less Secure option for the Credentials Protection Policy, it is possible for Hive audit logs to include Microsoft Azure credentials. If you are using Navigator Audit Server, these credentials may appear in audit reports. To mitigate this problem, make sure that access to Hive logs is appropriately controlled and that Navigator users with Auditing Viewer roles are cleared to have access to the Hive credentials.

Cloudera Issue: NAV-5861

Multi-cluster support for Hive lineage

When there were multiple Hive MetaStore (HMS) sources associated with a single Hive service, the method Navigator used to determine the source was not accurate, resulting in Impala and Spark extractors creating lineage relations that referred to the wrong sources and cross-cluster lineage relations were not generated.

Cloudera Issue: NAV-5859

Table-to-HDFS links were not established when Navigator supported multiple, high-availability clusters

When Navigator extracts metadata for multiple clusters and when the clusters were configured for high availability operation, Navigator does not correctly link tables to their HDFS backing data. The result is that lineage between Hive tables and their physical data are not created. In addition, some Hive table metadata that is derived from the backing files is not available.

You may see errors in the log such as the following:

```
2018-02-19 09:01:32,999 ERROR com.cloudera.nav.persist.impl.CompositeLinker
[CDHExecutor-0-CDHUrlClassLoader@01010d7e]: Internal error while
linking.java.lang.IllegalArgumentException: expected one element but was:
<com.cloudera.nav.core.model.Source@b22e68ba, com.cloudera.nav.core.model.Source@5b1bac2c>
```

Cloudera Issue: NAV-5749

Accurate dashboard metric for New Queries Executed

In some situations, the data reported for the Navigator Dashboard item **New Queries Executed** was shown to be larger than the actual number when Impala queries were included in the metrics. This problem has been resolved.

Cloudera Issue: NAV-5742

Known Issues in Cloudera Navigator 6.0.0

The following sections describe known issues in the data management components of Cloudera Navigator 6.0.0:

Authentication and Authorization

SAML authentication fails with "Cloudera Manager Only" setting

With the following combination of Cloudera Manager configuration properties set, authentication to Navigator fails:

- Authentication Backend Order: Cloudera Manager Only
- External Authentication Type: SAML

Workaround: To configure Navigator for SAML authentication, use an **Authentication Backend Order** other than "Cloudera Manager Only".

Authentication Backend Order nav.auth.backend.order	Navigator Metadata Server Default Group  <input checked="" type="radio"/> External Only <input type="radio"/> External then Cloudera Manager <input type="radio"/> Cloudera Manager Only <input type="radio"/> Cloudera Manager then External
---	--

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-6211

Cloudera Manager Configuration

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to  
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'  
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

At this time, there is no workaround.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-5331

Cloudera Manager audit events case-sensitive when using PostgreSQL

When Cloudera Manager and Navigator Audit Server are installed using PostgreSQL databases, the behavior of queries run from the Navigator console is different between the two databases. The result is that Cloudera Manager events are returned only if the query values match the case of the event values as they are stored in the Cloudera Manager database.

For example, the Hive operation "HiveReplicationCommand" is audited by Cloudera Manager; the audit log shows the command as `HIVEREPLICATIONCOMMAND` but querying with upper case fails to return the corresponding audit events. However, querying as `operation = HiveReplicationCommand` does return results.

Audit events other than those for Cloudera Manager are not affected.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-6141, NAV-6795

Hive, Hue, Impala

ClassCastException in Navigator Metadata Server log

When a Hive view is created, then dropped, and then subsequently recreated as a table with the same name as that of the original view, the Hive extraction process shows this exception in Navigator Metadata Server logs:

```
java.lang.ClassCastException: com.cloudera.nav.hive.model.HView cannot be cast to  
com.cloudera.nav.hive.model.HTable
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-5939

ConsoleAppender error in Hive Server 2 log

After running a query in Hive, the HiveServer2 `stderr` log includes the following error:

```
log4j:ERROR A "org.apache.log4j.ConsoleAppender" object is not assignable to a  
"com.cloudera.navigator.shaded.log4j.Appender" variable.  
log4j:ERROR The class "com.cloudera.navigator.shaded.log4j.Appender" was loaded by  
log4j:ERROR [sun.misc.Launcher$AppClassLoader@47d384ee] whereas object of type
```

```
log4j:ERROR "org.apache.log4j.ConsoleAppender" was loaded by
[sun.misc.Launcher$AppClassLoader@47d384ee].
log4j:ERROR Could not instantiate appender named "out".
```

This problem occurs because the Navigator Audit Server plugin for Hive Server 2 expects a different log4j integration than what HiveServer 2 is using. The result is that no Navigator Audit messages appear in the Hive Server 2 logs. Hive Server 2 auditing is not affected by this problem.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-6523

Overriding safety valve settings disables audit and lineage features

Customers or third party applications such as Unravel may require that `hive.exec.post.hooks` is configured in a HiveServer2 safety valve. Cloudera Manager will comment out the `hive.exec.post.hooks` value that is configured if audit or lineage is enabled for Hive. The safety valve content shows the commented code:

```
<!--'hive.exec.post.hooks', originally set to
'com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger'
(non-final), is overridden below by a safety valve-->
```

This automated change disables Navigator's auditing and lineage features without notification.

Affected Versions: Cloudera Navigator 6.0.0 and later

Workaround: To fix this problem, manually merge the original HiveServer2 safety valve content for `hive.exec.post.hooks` with the new value. For example, in the case of Unravel, the new safety valve would look like the following:

```
<property>
  <name>hive.exec.post.hooks</name>
  <value>com.cloudera.dataflow.hive.hook.HivePostHook,com.cloudera.navigator.audit.hive.HiveExecHookContext,org.apache.hadoop.hive.ql.hooks.LineageLogger</value>
  <description>for Unravel, from unraveldata.com</description>
</property>
```

Cloudera Issue: NAV-5331

Impala and Hive audit events fail to be captured when one audit event includes 4-byte characters, such as an emoji

This problem applies when the Navigator Audit Server database is a MySQL database configured to use the "UTF8" character set.

When a query includes an emoji or other Unicode supplementary-plane character that is encoded as four bytes in UTF-8, Navigator Audit Server fails to process the event and any following events from the same service.

Workaround: You can resolve this problem by configuring MySQL v5.5 or later to use the "UTF8MB4" character set. The error is described in this Stack Overflow article:

<https://stackoverflow.com/questions/13653712/java-sql-sqlexception-incorrect-string-value-xf0-x9f-x91-xbd-xf0-x9f>

The solution is described in the MySQL documentation topic [The UTF8MB4 Character Set \(4-Byte UTF-8 Unicode Encoding\)](#). Changing the character set requires restarting the MySQL server. It doesn't affect the Navigator Audit Server data.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-4845

Viewing Navigator tags in Hue overloads Metadata Server heap

When viewing Cloudera Navigator tags through Hue, Navigator uses more memory than usual and does not release the memory after logging out of Hue. Eventually, the calls between Hue and Navigator will occupy the majority of the heap space allocated to Navigator Metadata Server.

Workaround: Restart the Navigator Metadata Server periodically to clear the heap usage.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4326

Lineage not generated for Pig operations on Hive tables using HCatalog loader

When accessing a Hive table using Pig, lineage is generated in Navigator when using physical file loads, such as:

```
A = LOAD '/user/hive/warehouse/navigator_demo.db/salesdata';
B = LIMIT A 16;
STORE B INTO '/user/hive/warehouse/navigator_demo.db/salesdata_sample_file' using
PigStorage (';');
```

However, when accessing the Hive table using the HCatalog load, lineage for the Pig operation is not generated when browsing the source table lineage. Such as:

```
A = LOAD 'navigator_demo.salesdata' using org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 16;
STORE B INTO 'navigator_demo.salesdata_sample_hcatalog' using
org.apache.hive.hcatalog.pig.HCatStorer();
```

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3411

HiveServer1 and Hive CLI support removed

Cloudera Navigator requires HiveServer2 for complete governance Hive queries. Cloudera Navigator does not capture audit events for queries that are run on HiveServer1/Hive CLI, and lineage is not captured for certain types of operations that are run on HiveServer1.

If you use Cloudera Navigator to capture auditing, lineage, and metadata for Hive operations, upgrade to HiveServer2 if you have not done so already.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: TSB-185

Navigator Audit Server

Logging Threshold setting is not honored

The value for Navigator Audit Server Logging Threshold found in Cloudera Manager is not honored. Instead, messages are logged at trace level and displayed at DEBUG syslog level. This configuration property is set in **Cloudera Management Service > Configuration > Navigator Audit Server**.

Cloudera Issue: NAV-3737

Audit reports download option shows incorrect API version

The API version is not correct in the Cloudera Navigator download option for audit reports.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: 6.1.0

Cloudera Issue: NAV-3699

Pending audit events may be lost after role migration

Pending and deleted audit events may be lost during role migration. Pending audit events are those events that have not yet transferred from a source node through the Cloudera Navigator plug-in for the respective role to the Navigator Audit Server. If the Navigator Audit Server is unreachable due to network issues or other issues, the pending audits—which should automatically transfer at the conclusion of the role migration—may be lost.

Workaround: Before migrating a role to a different node:

- Check the state of the Navigator Audit Server (using Cloudera Manager Admin Console).
- Proceed with the role migration only if Navigator Audit Server is running.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-1006

Audit CSV has extra columns and is missing some data

Audit details exported as CSV files do not contain Sentry data. In addition, some of the column names display twice (Operation Text, Database Name, Object Type, for example), although the actual details display only of the duplicate columns.

Workaround: Export audits to JSON format to obtain Sentry data.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-708

Navigator Metadata Server

Navigator does not mark HDFS entities as deleted when in bulk extraction takes too long to complete

In large HDFS deployments, the `fsimage` takes a long time to index. When an HDFS checkpoint occurs it creates a new `fsimage`. However if the previous `fsimage` is still in the process of being indexed, Navigator cannot use the incremental changes found in the `inotify` stream because it refers to the newly created `fsimage`.

When this happens, Navigator attempts to start indexing the newer `fsimage`, creating a loop where Navigator can never take advantage of the more efficient change processing through `inotify`. The immediate fallout of this delay is that no HDFS entities deleted in the cluster will be marked as deleted in Navigator.

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1

Fixed Versions: Cloudera Navigator 6.1.0

Cloudera Issue: NAV-6456

International characters not supported in tags or property names

Navigator tags and the key portion of user-defined and managed properties do not support UNICODE characters beyond ASCII. Only ASCII text can be used in the text of a tag or the name of a property. Property values can include international characters.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: Tags are fixed in Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7011, NAV-7044

Navigator Embedded Solr can reach its limit on number of documents it can store



Important: This issue is critical when upgrading Cloudera Navigator deployments from Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0.

Navigator Metadata Server extracts HDFS entities by performing a one-time bulk extraction and then switching to incremental extraction. In Cloudera Manager releases 5.10.0, 5.10.1 and 5.11.0 (Navigator releases 2.9.0, 2.9.1, and 2.10.0), a problem causes HDFS bulk extraction to be run more than one time, resulting in duplicate relations created for HDFS. Over time, embedded Solr runs out of document IDs that it can assign to new relations and fails with following error:

```
"Caused by: java.lang.IllegalArgumentException: Too many documents, composite IndexReaders cannot exceed 2147483519"
```

When this happens, Navigator stops any more extraction of data as no new documents can be added to Solr.

After upgrading to this release, there is an additional recover step as described in "Repairing metadata in the storage directory after upgrading" in [Troubleshooting Navigator Data Management](#).

Affected Versions: Versions prior to Cloudera Manager 5.10 upgraded to Cloudera Manager 5.10 or higher

Fixed Versions: N/A

Cloudera Issue: NAV-5600

Log includes the error "EndPoint1 must not be null"

The following error may appear in the Navigator Metadata Server log in systems upgraded from Cloudera Manager version 5.x:

```
2017-10-17 13:00:23,007 ERROR com.cloudera.nav.hive.extractor.AbstractHiveExtractor [CDHExecutor-0-CDHUrlClassLoader@14784b7b]: Unable to parse hive view query *: EndPoint1 must not be null or empty  
java.lang.IllegalStateException: EndPoint1 must not be null or empty
```

This error occurs because the Hive pull extraction for creating a Hive view produces an incorrect lineage relationship for the Hive view. However, Navigator also receives information for the view creation through the push extractor, which correctly produces the lineage relation. You can safely ignore this error.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-4224

Purge

Navigator Metadata Server purge jobs may not run if there are policies configured

Navigator Metadata Server purge can produce messages such as "Checking if maintenance is running" and then fail to run during the available window. This problem occurs when a scheduled purge job waits for extraction tasks to finish but while waiting, a policy job starts, preventing the purge job from running.

Workaround: Specifically, to stop policies from running when you are trying to ensure that purge jobs will run, you can delete policies or temporarily change them to run infrequently to give the purge job time to run. When purge jobs have caught up to their backlog of work, you can change the policies back to running more frequently. Note that simply disabling the policies is not sufficient.

More generally, if you find that scheduled purge jobs are not running because there are other Navigator tasks in progress, consider stopping Navigator extractors and setting policies to run much less frequently. Then manually run the metadata purge using an API call to match your Navigator version:

```
curl -X POST -u user:password "https://navigator_host:7187/api/vXX/maintenance/purge?deleteTimeThresholdMinutes=duration"
```

API versions correspond to Navigator versions as described [Mapping API Versions to Product Versions](#).

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-7037

Policy specifications and cluster names affect purge

Policies cannot use cluster names in queries. Cluster name is a derived attribute and cannot be used as-is.

Workaround: When setting move actions for Cloudera Navigator, if there is only one cluster known to the Navigator instance, remove the `clusterName` clause.

If there is more than one cluster known to the Navigator instance, replace `clusterName` with `sourceId`. To get the `sourceId`, issue a query in this format:

```
curl '<nav-url>/api/v9/entities/?query=type%3Asource&limit=100&offset=0'
```

Use the identity of the matching HDFS service for this cluster as the `sourceId`.

Affected Versions: Cloudera Navigator 6.0.0 and later

Fixed Versions: N/A

Cloudera Issue: NAV-3537

Spark

Spark Lineage Limitations and Requirements

Spark lineage diagrams are supported in the Cloudera Navigator 6.0 release. Spark lineage is supported for Spark 1.6 and Spark 2.3. Lineage is not available for Spark when Cloudera Manager is running in single user mode. In addition to these requirements, Spark lineage has the following limitations:

- Lineage is produced only for data that is read/written and processed using the Dataframe and SparkSQL APIs. Lineage is not available for data that is read/written or processed using Spark's RDD APIs.
- Lineage information is not produced for calls to aggregation functions such as `groupBy()`.
- The default lineage directory for Spark on Yarn is `/var/log/spark/lineage`. No process or user should write files to this directory—doing so can cause agent failures. In addition, changing the Spark on Yarn lineage directory has no effect: the default remains `/var/log/spark/lineage`.

Navigator doesn't recognize local files in Spark jobs

Spark jobs can use files on the local filesystem as job inputs or outputs. Navigator, however, only supports HDFS, Hive, and S3 assets as job inputs or outputs. When Navigator extracts metadata from Spark and encounters a local source type, the metadata is discarded and the following error appears in the Navigator Metadata Server log:

```
2018-10-11 12:14:26,192 WARN com.cloudera.nav.api.ApiExceptionMapper
[qtp1574898980-23815]: Unexpected exception.
java.lang.RuntimeException: Source LOCAL isn't supported for Spark Lineage
```

Affected Versions: Cloudera Navigator 6.0.0, 6.0.1, 6.1.0, 6.1.1

Fixed Versions: Cloudera Navigator 6.2.0

Cloudera Issue: NAV-6811

Spark extractor enabled using safety valve deprecated

The Spark extractor included prior to CDH 5.11 and enabled by setting the safety valve, `nav.spark.extraction.enable=true` is being deprecated, and could be removed completely in a future release. If you are upgrading from CDH 5.10 or earlier and were using the extractor configured with this safety valve, be sure to remove the setting when you upgrade.

Upgrade Issues and Limitations

Upgrading Cloudera Navigator from Cloudera Manager 5.9 or Earlier Can be Extremely Slow

Upgrading a cluster running Cloudera Navigator to Cloudera Manager 5.10 (or higher) can be extremely slow due to an internal change made to the Solr schema in Cloudera Navigator 2.9. A Solr instance is embedded in Cloudera Navigator and supports its search capabilities. The Solr schema used by Cloudera Navigator has been modified in the 2.9 release to use datatype `long` rather than `string` for an internal `id` field. This change makes Cloudera Navigator far more robust and scalable over the long term.

However, the upgrade process itself can take a significantly long time because the existing Solr documents—the indexed and searchable data structures used by Solr that are contained in the Cloudera Navigator storage directory—are migrated to the new schema. This change to the Solr schema affects only those Cloudera Navigator deployments that use the metadata and lineage features.



Note: Cloudera Navigator deployments that use only the auditing features of the product—not metadata or lineage—are not affected by this issue.

The upgrade process for Cloudera Navigator starts automatically at the end of the Cloudera Manager upgrade, and the migration to the new schema occurs automatically as part of that upgrade process. The Navigator Metadata Server and Navigator console are not available during the upgrade. Navigator Audit Server runs normally. The amount of time that administrators should allow for this process depends on the quantity stored at the Navigator Metadata Server Storage Dir (nav.data.dir, or simply "storage directory") location as listed here:

Metadata and lineage usage	Description
None	Deployments that use Cloudera Navigator audit capability only —without metadata or lineage—do not have the issue. Backup the Navigator Metadata Server storage directory and then delete it before upgrading.
storage directory < 60 GB	Deployments with relatively small Navigator Metadata Server data directories may take 1 to 2 days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.
storage directory > 60 GB	Deployments with relatively large Navigator Metadata Server data directories may take several days for the upgrade process to complete. See the workaround below for steps to take before upgrading to Cloudera Manager 5.10 to possibly reduce the upgrade time.

Workaround: To reduce the time required for upgrading the Navigator Metadata Server data directories for deployments currently running Cloudera Navigator 2.8 that uses its metadata and lineage features, consider removing unneeded entries from the metadata before the upgrade. The Navigator Purge feature allows you to remove metadata for deleted entities and for entities and operations older than a specified date. For more information on what metadata you can remove with Purge, see [Managing Metadata Storage with Purge](#).

Run purge before starting the Cloudera Manager upgrade (to Cloudera Manager 5.10), following the steps below.



Warning: These steps may mitigate but do not fully resolve the issue. Follow these steps **before** starting the Cloudera Manager upgrade for any Cloudera Manager 5.9 or earlier cluster that currently uses the Cloudera Navigator metadata and lineage features.

- Check the Navigator Metadata Server storage directory size. The path is /var/lib/cloudera-scm-navigator (default) unless configured otherwise. If you need to check the setting:
 - Log in to Cloudera Manager Admin Console.
 - Select **Clusters > Cloudera Management Service**.
 - Click **Configuration** and then click the **Navigator Metadata Server** Scope filter:

Navigator Metadata Server
Storage Dir
nav.data.dir

Navigator Metadata Server (node-1)
/var/lib/cloudera-scm-navigator

- Confirm that the cluster uses the default configuration, or make a note of the location specified and the node name.

- Check the size of the actual directory contents. The following example shows a freshly installed system and so it is virtually empty.

```
[root@node-1 ~]# cd /var/lib/cloudera-scm-navigator
[root@node-1 cloudera-scm-navigator]# ls -l
total 12
drwxr-x--- 2 cloudera-scm cloudera-scm 113 Jul 12 06:56 diagnosticData
drwxr-x--- 2 cloudera-scm cloudera-scm 4096 Jul 12 09:16 extractorState
-rw-r---- 1 cloudera-scm cloudera-scm 36 Jul 12 05:54 instance.uuid
drwxr-x--- 4 cloudera-scm cloudera-scm 60 Jul 12 04:18 solr
drwxr-x--- 7 cloudera-scm cloudera-scm 4096 Jul 12 07:26 temp
[root@node-1 cloudera-scm-navigator]# cd solr
[root@node-1 solr]# ls -l
total 4
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_elements
drwxr-x--- 4 cloudera-scm cloudera-scm 28 Jul 12 05:56 nav_relations
-rw-r---- 1 cloudera-scm cloudera-scm 450 Jul 6 16:13 solr.xml
[root@node-1 solr]#
```

- Back up the contents of the directory. Use Cloudera Manager BDR or your preferred method.
- Schedule a purge process as described in [Scheduling the Purge Process](#).



Note: Users and processes cannot access Cloudera Navigator while purge is running.

Set options to purge metadata for deleted HDFS entities and any operations.

- Check the storage directory size again. If needed, re-run the purge with a shorter time span to retain metadata. If the storage directory consumption cannot be reduced below 60GB, do not start the Cloudera Manager upgrade. Instead, contact [Cloudera support](#) to help you with this upgrade.

Affected Versions: Cloudera Navigator 6.x

Fixed Versions: N/A

Cloudera Issue: NAV-5046

Cloudera Navigator 6 Encryption Release Notes



Note: For Cloudera Navigator 6 data management component release notes, see [Cloudera Navigator 6 Data Management Release Notes](#) on page 593.

These Release Notes provide information on the new features and known issues and limitations for the following Cloudera Navigator encryption components:

- Cloudera Navigator Key Trustee Server
- Cloudera Navigator Key HSM
- Cloudera Navigator Key Trustee KMS
- Cloudera Navigator HSM KMS
- Cloudera Navigator Encrypt

For information about supported operating systems, and other requirements for using Cloudera Navigator encryption components, see [Cloudera Enterprise 6 Requirements and Supported Versions](#) on page 5.

For more information about installing and configuring Cloudera Navigator encryption components, see [Cloudera Installation Guide](#).

To view release notes for the encryption components of a specific Cloudera Navigator 6 release, see the following:

Cloudera Navigator 6.2.x Encryption Release Notes

To view release notes for the encryption components of specific Cloudera Navigator 6.2.x releases, see the following:

[Cloudera Navigator 6.2.0 Encryption Release Notes](#)

The following topics describe new features, fixed issues, incompatible changes, and known issues for the encryption components of Cloudera Navigator 6.2.0:

[New Features in Cloudera Navigator 6.2.0 Encryption](#)

See below for new features in the encryption components of Cloudera Navigator 6.2.0, grouped by component:

[*New Features in Cloudera Navigator Key Trustee Server 6.2.0*](#)

Navigator Key Trustee Server is not released in 6.2.0.

[*New Features in Cloudera Navigator Key HSM 6.2.0*](#)

Navigator Key HSM is not released in 6.2.0.

[*New Features in Cloudera Navigator Key Trustee KMS 6.2.0*](#)

This release of Cloudera Navigator Key Trustee KMS provides support for the following platforms:

- RHEL and CentOS 7.6
- Oracle Linux 7.6
- Ubuntu 18 (Bionic)

[*New Features in Cloudera Navigator HSM KMS 6.2.0*](#)

This release of Cloudera Navigator HSM KMS provides support for the following platforms:

- RHEL and CentOS 7.6

[*New Features in Cloudera Navigator Encrypt 6.2.0*](#)

This release of Cloudera Navigator Encrypt provides support for the following platforms:

- RHEL and CentOS 7.6
- Oracle Linux 7.6
- Ubuntu 18 (Bionic)

[Issues Fixed in Cloudera Navigator 6.2.0 Encryption](#)

See below for issues fixed in the encryption components of Cloudera Navigator 6.2.0:

[*Issues Fixed in Navigator Key Trustee KMS 6.2.0*](#)

There are no new fixed issues for Navigator Key Trustee KMS in 6.2.0.

[*Issues Fixed in Navigator HSM KMS 6.2.0*](#)

There are no new fixed issues for Navigator HSM KMS in 6.2.0.

[*Issues Fixed in Navigator Encrypt 6.2.0*](#)

There are no new fixed issues for Navigator Encrypt in 6.2.0.

[Known Issues in Cloudera Navigator 6.2.0 Encryption](#)

The following sections describes known issues in the encryption components of Cloudera Navigator 6.2.0, grouped by component:

Known Issues in Cloudera Navigator Key Trustee Server 6.2.0**Warning:**

Interrupting deposit migration from Key Trustee Server to Key HSM can result in lost data

Workaround: Do not interrupt deposit migration to Key HSM.

Additional Key Trustee Server process appears after key creation when passive database is stopped

If a key is created while the passive Key Trustee Server database is down, the key creation will fail with the message, "Database write timed out." In this case, even after the passive database comes up, a hanging Key Trustee Server process thread may remain on the system attempting to complete the write. To view which Key Trustee Server processes are running, enter:

```
ps -ef | grep keytrustee | grep -v postgres
```

In normal operations there should be only one Key Trustee Server process running.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6684

Workaround: These extra Key Trustee Server processes do not cause any errors with the Key Trustee Server, but they will not be cleaned up properly when the Key Trustee Server is shut down. To fully shut down the Key Trustee Server when there are extra processes running, stop the Key Trustee Server service, and then kill any remaining processes.

Key Trustee Server active database setup fails when command line configuration specifies non-default database port

If the Key Trustee Server is configured from the command line to use a non-default database port (the default port is 11381), then when the Key Trustee Server service is added to Cloudera Manager, the first database startup fails.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6238

Workaround:

1. Log in to the Key Trustee Server and manually stop/start the databases:

```
service keytrustee-db stop
```

Make sure that the `postmaster.pid` file, which is located in the database directory, no longer exists. If necessary, replace `/var/lib/keytrustee/db` in the following command with the appropriate database directory for your system:

```
# ls /var/lib/keytrustee/db/postmaster.pid
```

If `postmaster.pid` has not been cleaned up, then use the `pg_ctl` utility to stop the database directly:

```
# pg_ctl stop
```

2. Return to the Cloudera Manager home page where the Key Trustee Server service is listed.
3. Go into the Key Trustee Server service configuration and for **Key Trustee Server Database Port**, specify the port that was configured during the command line configuration.
4. Redeploy the Key Trustee Server configuration and restart the service.

Key Trustee KMS cannot connect to Key Trustee Server using TLS versions other than 1.0 on JDK 7

If you have configured Key Trustee Server to use a TLS version other than 1.0, Key Trustee KMS fails to connect to Key Trustee Server, and key operations fail when using JDK 7.

Workaround: Use TLS version 1.0 only, or JDK 8.

[Key Trustee Server cannot use TLS version 1.2 on RHEL 6](#)

Configuring Key Trustee Server to use TLS version 1.2 causes Key Trustee Server to be unable to start.

Workaround: Use your operating system package manager to upgrade the `pyOpenSSL` package to version 1.4 or higher, or do not configure Key Trustee Server to use TLS version 1.2.

[Key Trustee Server PKCS8 private key cannot communicate with Key HSM](#)

If its private key is in PKCS8 format, Key Trustee Server cannot communicate with Key HSM.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-3172

Workaround: Convert the Key Trustee Server private key to raw RSA format.

[Known Issues in Cloudera Navigator Key HSM 6.2.0](#)

[Roll key command throws an exception and cannot retrieve metadata for key](#)

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if there is significant (> 15 ms ping time) network latency between the Key Trustee Server and the HSM device, then EDEK generation errors can occur during the roll key operation. These errors manifest in the KMS log as errors on the `generateEncryptedKey` operation. The KMS will recover from these errors on its own, but they may represent a nuisance to the operator.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-5646

Workaround: When these errors occur, you can use the `hadoop key list -metadata` command to confirm whether or not the key roll was successful, despite the error condition.

[KeySecure HSM not supported](#)

The KeySecure HSM is not supported in this release.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6001

Workaround: None.

[Key HSM Luna setup not showing the correct login status](#)

When running the `keyhsm setup luna` command, you are prompted for the Luna HSM slot number and login password. Key HSM then attempts to log into the Luna HSM to verify these settings are correct. In some circumstances, the setup script reports that the login was successful, even if it failed.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6623

Workaround: Any incorrect settings will cause the Key HSM service to throw an exception upon startup and exit. If the Key HSM service does not start correctly, check the log for the message: "Unable to sign into Luna HSM server. Please rerun application with 'setup' option to configure password." If this message appears, re-run the `keyhsm setup luna` command, and enter the correct slot number and login password.

[The `keyhsm trust` command fails](#)

When performing setup of Key HSM, the `ktadmin keyhsm --server http://server:port --trust` command may fail with the following message: "Unable to connect to Key HSM server at this address. Is the server running, and

is this KeyTrustee instance trusted (by running `keyhsm trust`)?" This failure can occur even if you have already successfully executed the `keyhsm trust` command.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6131

Workaround: Re-execute the `keyhsm trust` command, and then retry the `ktadmin keyhsm` command.

Upgrading Key HSM removes init script and binary

Upgrading Key HSM from 1.4.x to 1.5.x and higher removes the Key HSM init script and `/usr/bin/keyhsm` binary.

Workaround: Reinstall Key HSM:

```
sudo yum reinstall keytrustee-keyhsm
```

Key HSM cannot trust Key Trustee Server certificate if it has extended attributes

Key HSM cannot trust the Key Trustee Server certificate if it has extended attributes, and therefore cannot integrate with Key Trustee Server.

Workaround: Import the Key Trustee Server certificate to the Key HSM trust store using Java `keytool` instead of the `keyhsm trust` command.

Known Issues in Cloudera Navigator Key Trustee KMS 6.2.0

The Key Trustee KMS service fails to start if the Trust Store is configured without also configuring the Keystore

If you configure the Key Trustee KMS service **Key Management Server Proxy TLS/SSL Certificate Trust Store File** and **Key Management Server Proxy TLS/SSL Certificate Trust Store Password** parameters without also configuring the **Key Management Server Proxy TLS/SSL Server JKS Keystore File Location** and **Key Management Server Proxy TLS/SSL Server JKS Keystore File Password** parameters, the Key Trustee KMS service does not start.

Workaround: Configure all Trust Store and Keystore parameters.

Key Trustee KMS backup script fails if PostgreSQL versions lower than 9.3 are installed

If PostgreSQL versions lower than 9.3 are installed on the Key Trustee KMS host, the `ktbackup.sh` script fails with an error similar to the following:

```
pg_dump: server version: 9.3.11; pg_dump version: 9.2.14
pg_dump: aborting because of server version mismatch
```

Workaround: Uninstall the lower PostgreSQL version.

Known Issues in Cloudera Navigator HSM KMS 6.2.0

Encryption zone key is not deleted after migrating from Key Trustee KMS to HSM KMS

After migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS service should be restarted. Until it is restarted, any keys that are deleted after the migration may still be cached. If a deleted key is cached, then the data encrypted with that key can still be accessed even though the key has been deleted.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6434

Workaround: Restart the HSM KMS service after the key migration is complete.

KT KMS migration to HSM KMS fails if a key is not found on HSM

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if a key version is deleted on the HSM device directly without also deleting and purging that key version on the Key Trustee KMS, then attempts to migrate from the KT KMS to the HSM KMS will fail.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-5671

Workaround: None.

HSM KMS proxy fails to start during key migration

When migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS Proxy may fail to start and return a fatal error.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6431

Workaround: Navigate to the **Additional Java Configuration Options for KMS** setting in the HSM KMS configuration, and ensure that the HSM KMS truststore matches the truststore used by the Key Trustee KMS:

`-Djavax.net.ssl.trustStore=/path/to/truststore`

You can remove this option after completing the migration.

HSM KMS Thales Proxy role fails to start with non-default Thales HSM Server Port

Port 9001 is used by Cloudera Manager services, and it is also the default privileged port for the Thales HSM KMS. If you use port 9001 for the Thales HSM KMS, it will prevent the CM 6.0.0 upgrade from completing successfully.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6403

Workaround: If the Thales HSM KMS already exists, then before upgrading to Cloudera Manager 6.0.0, you must change the privileged Thales HSM KMS port; the recommended port is 11501. The non-privileged port default is 9000, and does not need to be changed.

If you are newly installing the Thales HSM KMS on a 6.0.0 system, then you must set the port to a non-default value before adding the HSM KMS backed by Thales service in Cloudera Manager.

To change the privileged port, log into the Thales HSM KMS machine(s), and run the following commands:

```
# sudo /opt/nfast/bin/config-serverstartup --enable-tcp --enable-privileged-tcp  
--privport=11501  
[server_settings] change successful; you must restart the hardserver for this to take  
effect  
# sudo /opt/nfast/sbin/init.d-ncipher restart  
-- Running shutdown script 90ncsnmpd  
  
-- Running shutdown script 60raserv  
  
...  
  
'ncsnmpd' server now running
```

After successfully running the commands, restart the Thales HSM KMS.

Key store key password specified in Cloudera Manager is not properly used by HSM KMS Service

If the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore File Password and the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore Key Password are not both specified or differ from each other, then the HSM KMS inter-node handshake fails and HSM KMS High Availability is not configured.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6020

Workaround: Use the same password for the keystore file and the keystore key file.

Timeout error during encryption zone key creation

There are situations where the key cache is synchronously populated to capacity during the create encryption zone operation. The expected behavior is that the key cache is synchronously populated only to the low watermark level (the rest of the keys should be created asynchronously).

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-5296

Workaround: On the HSM KMS, in the field **HSM KMS Proxy Advanced Configuration Snippet (Safety Valve)** for kms-site.xml:

- hadoop.security.kms.encrypted.key.cache.low.watermark .05
- hadoop.security.kms.encrypted.key.cache.size 30

On the HDFS, in the field **HDFS Cluster-wide Advanced Configuration Snippet (Safety Valve)** for core-site.xml:

- hadoop.security.kms.client.encrypted.key.cache.size 30
- hadoop.security.kms.client.encrypted.key.cache.low-watermark .05

HSM KMS Luna may need to be restarted if inactive for extended period

If Hadoop key operations return `com.safenetinc.luna.exception.LunaCryptokiException` after the KMS has been running without activity for an extended period time, the Luna session may have been dropped.

Affected Version(s): 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-5018

Workaround: Restart the KMS service.

Creating multiple instances of HSM KMS on the same host and port causes an error upon delete

Creating a KMS role instance on a host that previously hosted a KMS role instance in the same role group that had its data directories deleted results in errors when attempting to run Hadoop key delete operations.

Affected Version(s): 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-4992

Workaround: This workaround requires the assistance of Cloudera support; request assistance with issue KT-4992

Incorrect status for "Restart stale services" step in HDFS encryption wizard post-service installation

There are times when completion of the HDFS Encryption Wizard does not show an active "Restart stale services and redeploy client configuration" link.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-4987

Workaround: Refresh the page and the link should become active.

The encryption wizard continues to fail if there is a failure during initial configuration run

The encryption wizard continues to fail if there was a failure during the initial run configuring HSM KMS.

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-4909

Workaround: Open Cloudera Manager in another browser tab, and manually stop the installed KMS by clicking the arrow next to the KMS and selecting **Stop**. Then retry the installation in the new tab after correcting the cause of the install failure.

Before installing the Thales backed HSM KMS, you must add the KMS user to the `nfast` group

After installation of the Thales HSM client, and before installing Navigator HSM KMS backed by Thales HSM, you must add the KMS user to the `nfast` group..

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-4618

Workaround: Run the following command to manually add the KMS user to the `nfast` group:
`usermod -a -G nfast kms`

Known Issues in Cloudera Navigator Encrypt 6.2.0

Navigator Encrypt cannot create an ACL when Key Trustee Server is down

Navigator Encrypt cannot create new ACLs when the Key Trustee Server is down. Navigator Encrypt will attempt to verify the master key with the Key Trustee Server when adding ACLs, even if the master key should already have been cached on the local system. If it can't communicate with the Key Trustee Server, the add ACL request will fail.

Affected Version: 6.1.0, 6.2.0

Cloudera Bug: KT-6390

Workaround: Make sure the Key Trustee Server is running before making modifications to Navigator Encrypt ACLs.

Issue with `nav encrypt-mount` service status message

Affected Version: 6.0.0, 6.1.0, 6.2.0

Cloudera Bug: KT-6309

On RHEL 7, the service `nav encrypt-mount` `stop` command might return a successful exit status message, even if there was a problem stopping Navigator Encrypt. This is not an issue with the `nav encrypt-mount` service command; rather, it is a problem with the message output.

You can verify that `nav encrypt-mount` stopped successfully by checking for the presence of the `nav encryptfs` module in the `lsmod` output:

```
# lsmod | grep nav encrypt  
nav encryptfs 101407 0
```

Alternatively, you can verify that the `nav encrypt` mount points successfully dismounted by checking the output of the `df` or `mount` commands.

Workaround: None.

Cloudera Navigator 6.1.x Encryption Release Notes

To view release notes for the encryption components of specific Cloudera Navigator 6.1.x releases, see the following:

Cloudera Navigator 6.1.1 Encryption Release Notes

No Cloudera Navigator Encryption products were updated or released for 6.1.1.

New Features in Cloudera Navigator 6.1.1 Encryption

No Cloudera Navigator Encryption products were released in 6.1.1.

Issues Fixed in Cloudera Navigator 6.1.1 Encryption

No Cloudera Navigator Encryption products were released in 6.1.1.

Known Issues in Cloudera Navigator 6.1.1 Encryption

No Cloudera Navigator Encryption products were released in 6.1.1.

Cloudera Navigator 6.1.0 Encryption Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the encryption components of Cloudera Navigator 6.1.0:

New Features in Cloudera Navigator 6.1.0 Encryption

See below for new features in the encryption components of Cloudera Navigator 6.1.0, grouped by component:

New Features in Cloudera Navigator Key Trustee Server 6.1.0

This release of Cloudera Navigator Key Trustee Server provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator Key Trustee Server.

New Features in Cloudera Navigator Key HSM 6.1.0

This release of Cloudera Navigator Key HSM provides the following new functionality:

- You can now use the following commands to stop the Key HSM service: `service keyhsm stop` and `keyhsm stop`. Previously, you could use the `shutdown` option, which does not match Unix service conventions.
- Key HSM now supports AWS CloudHSM.

New Features in Cloudera Navigator Key Trustee KMS 6.1.0

This release of Cloudera Navigator Key Trustee KMS provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator Key Trustee KMS.

New Features in Cloudera Navigator HSM KMS 6.1.0

This release of Cloudera Navigator HSM KMS provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator HSM KMS.

New Features in Cloudera Navigator Encrypt 6.1.0

This release of Cloudera Navigator Encrypt provides the following new functionality:

- For configurations where a device name was used to configure a mount point, and you wish to convert to a UUID, you have the option to use the UUID conversion utility (`nav encrypt-prepare --convert-uuid`), which converts existing mount points from using the device name to using the UUID. See [Converting from Device Names to UUIDs for Encrypted Devices](#).

Issues Fixed in Cloudera Navigator 6.1.0 Encryption

See below for issues fixed in the encryption components of Cloudera Navigator 6.1.0:

Issues Fixed in Navigator Key Trustee KMS 6.1.0

New Key Trustee KMS failed after being added to an environment that previously had a single Key Trustee KMS instance

When adding a new Navigator Key Trustee KMS instance to an environment that previously only had a single Key Trustee KMS instance, the new Key Trustee KMS periodically failed to start and returned the following message:

"Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synced between all KMS hosts. Aborting to prevent data inconsistency."

Cloudera Issue: KT-6231

CDH upgrade failure

When upgrading to Key Trustee KMS 6.0.0 from Key Trustee KMS 5.14.0 or lower, *and* performing a rolling restart (instead of a full restart), the first Key Trustee KMS instance to restart may fail to come up and present the error: "Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synched between all KMS hosts. Aborting to prevent data inconsistency."

Cloudera Bug: KT-6547

When a mount point is added Navigator Encrypt updates configuration files before the action completes

The `navcrypt-prepare` command sometimes performed updates to the `/etc/navcrypt/control` and `/etc/navcrypt/ztab` files before the command completed.

Cloudera Issue: KT-6383

Issues Fixed in Navigator Key HSM 6.1.0

Key HSM Luna setup not showing the correct login status

When running the `keyhsm setup luna` command, you are prompted for the Luna HSM slot number and login password. Key HSM then attempts to log into the Luna HSM to verify these settings are correct. In some circumstances, the setup script reports that the login was successful, even if it failed.

Cloudera Issue: KT-6623

Too many keys on Luna HSM causes Key HSM startup to fail

If there are too many keys on the Luna HSM, Key HSM startup will fail with a Java core dump because it times out querying the keys.

Cloudera Issue: KT-6129

Issues Fixed in Navigator Encrypt 6.1.0

Navigator Encrypt-related packages should be downgraded with the Navigator Encrypt package

The `navcrypt downgrade` command only downgraded the `navcrypt` package. It did not downgrade the associated `navcrypt-kernel-module` and `libkeytrustee` packages.

Cloudera Issue: KT-6381

When a mount point is added Navigator Encrypt updates configuration files before the action completes

The `navcrypt-prepare` command sometimes performed updates to the `/etc/navcrypt/control` and `/etc/navcrypt/ztab` files before the command completed. In such cases, if there was an error with the mounting or unmounting of the `navcrypt` mount point, then the updated `control` and `ztab` files did not accurately reflect which mount points existed.

Cloudera Issue: KT-6383

Navigator Encrypt will not build on RHEL kernel 3.10.0-862.14.4

The Navigator Encrypt kernel module will not build on RHEL kernel 3.10.0-862.14.4. This impacts new installations, and existing installations that are upgrading to kernel 3.10.0-862.14.4. This issue prevents the `navcrypt-mount` service from running and Navigator Encrypt mount points from being accessible.

Cloudera Issue: KT-6677

Upgrade from Navigator Encrypt 3.x to 6.0.0 does not trigger a navencryptfs recompile

After upgrading from Navigator Encrypt 3.x to Navigator Encrypt 6.0.0, the kernel module may not be rebuilt, even if the upgrade and installation commands indicate that it was built successfully.

Cloudera Bug: KT-6382

Known Issues in Cloudera Navigator 6.1.0 Encryption

The following sections describes known issues in the encryption components of Cloudera Navigator 6.1.0, grouped by component:

Known Issues in Cloudera Navigator Key Trustee Server 6.1.0



Warning:

Interrupting deposit migration from Key Trustee Server to Key HSM can result in lost data

Workaround: Do not interrupt deposit migration to Key HSM.

Additional Key Trustee Server process appears after key creation when passive database is stopped

If a key is created while the passive Key Trustee Server database is down, the key creation will fail with the message, "Database write timed out." In this case, even after the passive database comes up, a hanging Key Trustee Server process thread may remain on the system attempting to complete the write. To view which Key Trustee Server processes are running, enter:

```
ps -ef | grep keytrustee | grep -v postgres
```

In normal operations there should be only one Key Trustee Server process running.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6684

Workaround: These extra Key Trustee Server processes do not cause any errors with the Key Trustee Server, but they will not be cleaned up properly when the Key Trustee Server is shut down. To fully shut down the Key Trustee Server when there are extra processes running, stop the Key Trustee Server service, and then kill any remaining processes.

Key Trustee Server active database setup fails when command line configuration specifies non-default database port

If the Key Trustee Server is configured from the command line to use a non-default database port (the default port is 11381), then when the Key Trustee Server service is added to Cloudera Manager, the first database startup fails.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6238

Workaround:

1. Log in to the Key Trustee Server and manually stop/start the databases:

```
service keytrustee-db stop
```

Make sure that the `postmaster.pid` file, which is located in the database directory, no longer exists. If necessary, replace `/var/lib/keytrustee/db` in the following command with the appropriate database directory for your system:

```
# ls /var/lib/keytrustee/db/postmaster.pid
```

If `postmaster.pid` has not been cleaned up, then use the `pg_ctl` utility to stop the database directly:

```
# pg_ctl stop
```

2. Return to the Cloudera Manager home page where the Key Trustee Server service is listed.
3. Go into the Key Trustee Server service configuration and for **Key Trustee Server Database Port**, specify the port that was configured during the command line configuration.
4. Redeploy the Key Trustee Server configuration and restart the service.

[Key Trustee KMS cannot connect to Key Trustee Server using TLS versions other than 1.0 on JDK 7](#)

If you have configured Key Trustee Server to use a TLS version other than 1.0, Key Trustee KMS fails to connect to Key Trustee Server, and key operations fail when using JDK 7.

Workaround: Use TLS version 1.0 only, or JDK 8.

[Key Trustee Server cannot use TLS version 1.2 on RHEL 6](#)

Configuring Key Trustee Server to use TLS version 1.2 causes Key Trustee Server to be unable to start.

Workaround: Use your operating system package manager to upgrade the `pyOpenSSL` package to version 1.4 or higher, or do not configure Key Trustee Server to use TLS version 1.2.

[Key Trustee Server PKCS8 private key cannot communicate with Key HSM](#)

If its private key is in PKCS8 format, Key Trustee Server cannot communicate with Key HSM.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-3172

Workaround: Convert the Key Trustee Server private key to raw RSA format.

[*Known Issues in Cloudera Navigator Key HSM 6.1.0*](#)

[Roll key command throws an exception and cannot retrieve metadata for key](#)

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if there is significant (> 15 ms ping time) network latency between the Key Trustee Server and the HSM device, then EDEK generation errors can occur during the roll key operation. These errors manifest in the KMS log as errors on the `generateEncryptedKey` operation. The KMS will recover from these errors on its own, but they may represent a nuisance to the operator.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-5646

Workaround: When these errors occur, you can use the `hadoop key list -metadata` command to confirm whether or not the key roll was successful, despite the error condition.

[KeySecure HSM not supported](#)

The KeySecure HSM is not supported in this release.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6001

Workaround: None.

[Too many keys on Luna HSM causes Key HSM startup to fail](#)

If there are too many keys on the Luna HSM, Key HSM startup will fail with a Java core dump because it times out querying the keys.

Affected Version: 6.0.0

Cloudera Bug: KT-6129

Fixed in Version: 6.1.0

Workaround: Increase the value in the `keyhsm.countdown.time` property in `application.properties` (the default is 45 seconds) until Key HSM starts successfully.

Key HSM Luna setup not showing the correct login status

When running the `keyhsm setup luna` command, you are prompted for the Luna HSM slot number and login password. Key HSM then attempts to log into the Luna HSM to verify these settings are correct. In some circumstances, the setup script reports that the login was successful, even if it failed.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6623

Workaround: Any incorrect settings will cause the Key HSM service to throw an exception upon startup and exit. If the Key HSM service does not start correctly, check the log for the message: "Unable to sign into Luna HSM server. Please rerun application with 'setup' option to configure password." If this message appears, re-run the `keyhsm setup luna` command, and enter the correct slot number and login password.

The `keyhsm trust` command fails

When performing setup of Key HSM, the `ktadmin keyhsm --server http://server:port --trust` command may fail with the following message: "Unable to connect to Key HSM server at this address. Is the server running, and is this KeyTrustee instance trusted (by running `keyhsm trust`)?" This failure can occur even if you have already successfully executed the `keyhsm trust` command.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6131

Workaround: Re-execute the `keyhsm trust` command, and then retry the `ktadmin keyhsm` command.

Upgrading Key HSM removes init script and binary

Upgrading Key HSM from 1.4.x to 1.5.x and higher removes the Key HSM init script and `/usr/bin/keyhsm` binary.

Workaround: Reinstall Key HSM:

```
sudo yum reinstall keytrustee-keyhsm
```

Key HSM cannot trust Key Trustee Server certificate if it has extended attributes

Key HSM cannot trust the Key Trustee Server certificate if it has extended attributes, and therefore cannot integrate with Key Trustee Server.

Workaround: Import the Key Trustee Server certificate to the Key HSM trust store using Java `keytool` instead of the `keyhsm trust` command.

Known Issues in Cloudera Navigator Key Trustee KMS 6.1.0

CDH upgrade failure

When upgrading to Key Trustee KMS 6.0.0 from Key Trustee KMS 5.14.0 or lower, *and* performing a rolling restart (instead of a full restart), the first Key Trustee KMS instance to restart may fail to come up and present the error: "Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synched between all KMS hosts. Aborting to prevent data inconsistency."

Affected Versions: 6.0.0

Cloudera Bug: KT-6547

Fixed Version: 6.1.0

Workaround: If possible, perform a full restart instead of a rolling restart.

If you cannot execute a full restart, then add the following line to the `/var/lib/kms-keytrustee/keytrustee/.keytrustee/keytrustee.conf` file on all Key Trustee KMS instances, and then restart the Key Trustee KMS that failed:

```
"FINGERPRINT_VALIDATED": "True"
```

New Key Trustee KMS may fail after being added to an environment that previously had a single Key Trustee KMS instance

When adding a new Navigator Key Trustee KMS instance to an environment that previously only had a single Key Trustee KMS instance, the new Key Trustee KMS may fail to start and return the following message:

"Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synced between all KMS hosts. Aborting to prevent data inconsistency."

This message correctly appears if GPG private keys have not been synced across all Key Trustee KMS instances. However, it may erroneously appear when the GPG private keys have been synced, but the original Key Trustee KMS instance has not been restarted to pick up the new configuration.

If you encounter this error message and have either recently upgraded the Key Trustee KMS, or have installed a new Key Trustee KMS service running in HA mode, refer to [Key Trustee KMS Encryption Issues](#) for troubleshooting steps.

Affected Version: 6.0.0

Fixed Version: 6.1.0

Cloudera Bug: KT-6231

Workaround: Restart the original Key Trustee KMS instance, and then start the new Key Trustee KMS instance.

The Key Trustee KMS service fails to start if the Trust Store is configured without also configuring the Keystore

If you configure the Key Trustee KMS service **Key Management Server Proxy TLS/SSL Certificate Trust Store File** and **Key Management Server Proxy TLS/SSL Certificate Trust Store Password** parameters without also configuring the **Key Management Server Proxy TLS/SSL Server JKS Keystore File Location** and **Key Management Server Proxy TLS/SSL Server JKS Keystore File Password** parameters, the Key Trustee KMS service does not start.

Workaround: Configure all Trust Store and Keystore parameters.

Key Trustee KMS backup script fails if PostgreSQL versions lower than 9.3 are installed

If PostgreSQL versions lower than 9.3 are installed on the Key Trustee KMS host, the `ktbackup.sh` script fails with an error similar to the following:

```
pg_dump: server version: 9.3.11; pg_dump version: 9.2.14
pg_dump: aborting because of server version mismatch
```

Workaround: Uninstall the lower PostgreSQL version.

Known Issues in Cloudera Navigator HSM KMS 6.1.0

Encryption zone key is not deleted after migrating from Key Trustee KMS to HSM KMS

After migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS service should be restarted. Until it is restarted, any keys that are deleted after the migration may still be cached. If a deleted key is cached, then the data encrypted with that key can still be accessed even though the key has been deleted.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6434

Workaround: Restart the HSM KMS service after the key migration is complete.

KT KMS migration to HSM KMS fails if a key is not found on HSM

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if a key version is deleted on the HSM device directly without also deleting and purging that key version on the Key Trustee KMS, then attempts to migrate from the KT KMS to the HSM KMS will fail.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-5671

Workaround: None.

HSM KMS proxy fails to start during key migration

When migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS Proxy may fail to start and return a fatal error.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6431

Workaround: Navigate to the **Additional Java Configuration Options for KMS** setting in the HSM KMS configuration, and ensure that the HSM KMS truststore matches the truststore used by the Key Trustee KMS:

`-Djavax.net.ssl.trustStore=/path/to/truststore`

You can remove this option after completing the migration.

HSM KMS Thales Proxy role fails to start with non-default Thales HSM Server Port

Port 9001 is used by Cloudera Manager services, and it is also the default privileged port for the Thales HSM KMS. If you use port 9001 for the Thales HSM KMS, it will prevent the CM 6.0.0 upgrade from completing successfully.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6403

Workaround: If the Thales HSM KMS already exists, then before upgrading to Cloudera Manager 6.0.0, you must change the privileged Thales HSM KMS port; the recommended port is 11501. The non-privileged port default is 9000, and does not need to be changed.

If you are newly installing the Thales HSM KMS on a 6.0.0 system, then you must set the port to a non-default value before adding the HSM KMS backed by Thales service in Cloudera Manager.

To change the privileged port, log into the Thales HSM KMS machine(s), and run the following commands:

```
# sudo /opt/nfast/bin/config-serverstartup --enable-tcp --enable-privileged-tcp
--privport=11501
[server_settings] change successful; you must restart the hardserver for this to take
effect
# sudo /opt/nfast/sbin/init.d-ncipher restart
  -- Running shutdown script 90ncsnmpd
  -- Running shutdown script 60raserv
...
'ncsnmpd' server now running
```

After successfully running the commands, restart the Thales HSM KMS.

Key store key password specified in Cloudera Manager is not properly used by HSM KMS Service

If the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore File Password and the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore Key Password are not both specified or differ from each other, then the HSM KMS inter-node handshake fails and HSM KMS High Availability is not configured.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6020

Workaround: Use the same password for the keystore file and the keystore key file.

Timeout error during encryption zone key creation

There are situations where the key cache is synchronously populated to capacity during the create encryption zone operation. The expected behavior is that the key cache is synchronously populated only to the low watermark level (the rest of the keys should be created asynchronously).

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-5296

Workaround: On the HSM KMS, in the field **HSM KMS Proxy Advanced Configuration Snippet (Safety Valve)** for kms-site.xml:

- hadoop.security.kms.encrypted.key.cache.low.watermark .05
- hadoop.security.kms.encrypted.key.cache.size 30

On the HDFS, in the field **HDFS Cluster-wide Advanced Configuration Snippet (Safety Valve)** for core-site.xml:

- hadoop.security.kms.client.encrypted.key.cache.size 30
- hadoop.security.kms.client.encrypted.key.cache.low-watermark .05

HSM KMS Luna may need to be restarted if inactive for extended period

If Hadoop key operations return com.safenetinc.luna.exception.LunaCryptokiException after the KMS has been running without activity for an extended period time, the Luna session may have been dropped.

Affected Version(s): 6.0.0, 6.1.0

Cloudera Bug: KT-5018

Workaround: Restart the KMS service.

Creating multiple instances of HSM KMS on the same host and port causes an error upon delete

Creating a KMS role instance on a host that previously hosted a KMS role instance in the same role group that had its data directories deleted results in errors when attempting to run Hadoop key delete operations.

Affected Version(s): 6.0.0, 6.1.0

Cloudera Bug: KT-4992

Workaround: This workaround requires the assistance of Cloudera support; request assistance with issue KT-4992

Incorrect status for "Restart stale services" step in HDFS encryption wizard post-service installation

There are times when completion of the HDFS Encryption Wizard does not show an active "Restart stale services and redeploy client configuration" link.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-4987

Workaround: Refresh the page and the link should become active.

The encryption wizard continues to fail if there is a failure during initial configuration run

The encryption wizard continues to fail if there was a failure during the initial run configuring HSM KMS.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-4909

Workaround: Open Cloudera Manager in another browser tab, and manually stop the installed KMS by clicking the arrow next to the KMS and selecting **Stop**. Then retry the installation in the new tab after correcting the cause of the install failure.

Before installing the Thales backed HSM KMS, you must add the KMS user to the `nfast` group

After installation of the Thales HSM client, and before installing Navigator HSM KMS backed by Thales HSM, you must add the KMS user to the `nfast` group..

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-4618

Workaround: Run the following command to manually add the KMS user to the `nfast` group:
`usermod -a -G nfast kms`

Known Issues in Cloudera Navigator Encrypt 6.1.0

Navigator Encrypt cannot create an ACL when Key Trustee Server is down

Navigator Encrypt cannot create new ACLs when the Key Trustee Server is down. Navigator Encrypt will attempt to verify the master key with the Key Trustee Server when adding ACLs, even if the master key should already have been cached on the local system. If it can't communicate with the Key Trustee Server, the add ACL request will fail.

Affected Version: 6.1.0

Cloudera Bug: KT-6390

Workaround: Make sure the Key Trustee Server is running before making modifications to Navigator Encrypt ACLs.

Navigator Encrypt-related packages should be downgraded with the Navigator Encrypt package

The `navencrypt downgrade` command will only downgrade the `navencrypt` package. It will not downgrade the associated `navencrypt-kernel-module` and `libkeytrustee` packages.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6381

Workaround: To fully downgrade Navigator Encrypt, manually downgrade all of the associated Navigator Encrypt packages (in the order listed):

1. `navencrypt`
2. `navencrypt-kernel-module`
3. `libkeytrustee`

Upgrade from Navigator Encrypt 3.x to 6.0.0 does not trigger a `navcryptofs` recompile

After upgrading from Navigator Encrypt 3.x to Navigator Encrypt 6.0.0, the kernel module may not be rebuilt, even if the upgrade and installation commands indicate that it was built successfully.

Affected Version: 6.0.0

Cloudera Bug: KT-6382

Fixed in Version: 6.1.0

Workaround: If the version currently loaded is not the desired version, use the following commands to update it. In this case, you must manually load the updated kernel module.

To identify which version of the kernel module is being used, look for the currently loaded version in the `dmesg` output (in this case, the output reveals that kernel module version was not rebuilt):

```
navencrypt: time=1531428589 level=INFO app="kernel" Cloudera navcryptofs v3.14.0-405
loaded
```

For RHEL/UBUNTU:

```
service navencrypt-mount stop  
navencrypt-module-setup  
service navencrypt-mount start
```

For SLES:

```
service navencrypt-mount stop  
rmmod navcryptfs  
modprobe -v navcryptfs  
service navencrypt-mount start
```

When a mount point is added Navigator Encrypt updates configuration files before the action completes

The `navencrypt-prepare` command sometimes performs updates to the `/etc/navencrypt/control` and `/etc/navencrypt/ztab` files before the command has completed. In such cases, if there is an error with the mounting or unmounting of the `navencrypt` mount point, then the updated control and `ztab` files will not accurately reflect which mount points exist. Do *not* manually change these files without having a known working backup.

Affected Version: 6.0.0

Cloudera Bug: KT-6383

Fixed in Version: 6.1.0

Workaround: None.

Issue with `navencrypt-mount` service status message

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6309

On RHEL 7, the `service navencrypt-mount stop` command might return a successful exit status message, even if there was a problem stopping Navigator Encrypt. This is not an issue with the `navencrypt-mount` service command; rather, it is a problem with the message output.

You can verify that `navencrypt-mount` stopped successfully by checking for the presence of the `navcryptfs` module in the `lsmod` output:

```
# lsmod | grep navencrypt  
navcryptfs 101407 0
```

Alternatively, you can verify that the `navencrypt` mount points successfully dismounted by checking the output of the `df` or `mount` commands.

Workaround: None.

Cloudera Navigator 6.0.x Encryption Release Notes

To view release notes for the encryption components of specific Cloudera Navigator 6.0.x releases, see the following:

[Cloudera Navigator 6.0.1 Encryption Release Notes](#)

No Cloudera Navigator Encryption products were updated or released for 6.0.1.

[New Features in Cloudera Navigator 6.0.1 Encryption](#)

No Cloudera Navigator Encryption products were released in 6.0.1.

[Issues Fixed in Cloudera Navigator 6.0.1 Encryption](#)

No Cloudera Navigator Encryption products were released in 6.0.1.

Known Issues in Cloudera Navigator 6.0.1 Encryption

No Cloudera Navigator Encryption products were released in 6.0.1.

Cloudera Navigator 6.0.0 Encryption Release Notes

The following topics describe new features, fixed issues, incompatible changes, and known issues for the encryption components of Cloudera Navigator 6.0.0:

New Features in Cloudera Navigator 6.0.0 Encryption

See below for new features in the encryption components of Cloudera Navigator 6.0.0, grouped by component:

New Features in Cloudera Navigator Key Trustee Server 6.0.0

This release of Cloudera Navigator Key Trustee Server provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator Key Trustee Server.

New Features in Cloudera Navigator Key HSM 6.0.0

This release of Cloudera Navigator Key HSM provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator Key HSM.

New Features in Cloudera Navigator Key Trustee KMS 6.0.0

This release of Cloudera Navigator Key Trustee KMS provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator Key Trustee KMS.

New Features in Cloudera Navigator HSM KMS 6.0.0

This release of Cloudera Navigator HSM KMS provides the following new functionality:

- There are no notable new features in this release of Cloudera Navigator HSM KMS.

New Features in Cloudera Navigator Encrypt 6.0.0

This release of Cloudera Navigator Encrypt provides the following new functionality:

- When changing the master key, there is a new option for the `nav encrypt key --change` command: the `--rsa-oaep` parameter updates the master key to use RSA OAEP padding. While this is optional, it is a best practice to use the new, more secure padding. See the Navigator Encrypt Upgrade documentation for more details.
- Support for RHEL/CentOS 7.5.

Issues Fixed in Cloudera Navigator 6.0.0 Encryption

See below for issues fixed in the encryption components of Cloudera Navigator 6.0.0, grouped by component:

Issues Fixed in Cloudera Navigator Key Trustee Server 6.0.0

There are no notable fixed issues in this release of Cloudera Navigator Key Trustee Server.

Issues Fixed in Cloudera Navigator Key HSM 6.0.0

There are no notable fixed issues in this release of Cloudera Navigator Key HSM.

Issues Fixed in Cloudera Navigator Key Trustee KMS 6.0.0

Validation fails if `hostname` command returns `shortname`

If the `hostname` command on the OS returns `shortname`, and the `core-site.xml` of the KMS process has a `hadoop.security.key.provider.path` with a fully qualified domain name (FQDN), then the znodes will be created with the `shortname`. Consequently, when KMS 1 checks the fingerprint of KMS 2, it will expect the FQDN as the znode, and fail the validation.

Cloudera Issue: KT-6412

Issues Fixed in Cloudera Navigator HSM KMS 6.0.0

There are no notable fixed issues in this release of Cloudera Navigator HSM KMS.

New Features in Cloudera Navigator Encrypt 6.0.0

There are no notable fixed issues in this release of Cloudera Navigator Encrypt.

Known Issues in Cloudera Navigator 6.0.0 Encryption

The following sections describes known issues in the encryption components of Cloudera Navigator 6.0, grouped by component:

Known Issues in Cloudera Navigator Key Trustee Server 6.0.0



Warning:

Interrupting deposit migration from Key Trustee Server to Key HSM can result in lost data

Workaround: Do not interrupt deposit migration to Key HSM.

Key Trustee Server active database setup fails when command line configuration specifies non-default database port

If the Key Trustee Server is configured from the command line to use a non-default database port (the default port is 11381), then when the Key Trustee Server service is added to Cloudera Manager, the first database startup fails.

Affected Version: 6.0.0

Cloudera Bug: KT-6238

Workaround:

1. Log in to the Key Trustee Server and manually stop/start the databases:

```
service keytrustee-db stop
```

Make sure that the `postmaster.pid` file, which is located in the database directory, no longer exists. If necessary, replace `/var/lib/keytrustee/db` in the following command with the appropriate database directory for your system:

```
# ls /var/lib/keytrustee/db/postmaster.pid
```

If `postmaster.pid` has not been cleaned up, then use the `pg_ctl` utility to stop the database directly:

```
# pg_ctl stop
```

2. Return to the Cloudera Manager home page where the Key Trustee Server service is listed.
3. Go into the Key Trustee Server service configuration and for **Key Trustee Server Database Port**, specify the port that was configured during the command line configuration.
4. Redeploy the Key Trustee Server configuration and restart the service.

Key Trustee KMS cannot connect to Key Trustee Server using TLS versions other than 1.0 on JDK 7

If you have configured Key Trustee Server to use a TLS version other than 1.0, Key Trustee KMS fails to connect to Key Trustee Server, and key operations fail when using JDK 7.

Workaround: Use TLS version 1.0 only, or JDK 8.

Key Trustee Server cannot use TLS version 1.2 on RHEL 6

Configuring Key Trustee Server to use TLS version 1.2 causes Key Trustee Server to be unable to start.

Workaround: Use your operating system package manager to upgrade the `pyOpenSSL` package to version 1.4 or higher, or do not configure Key Trustee Server to use TLS version 1.2.

Key Trustee Server PKCS8 private key cannot communicate with Key HSM

If its private key is in PKCS8 format, Key Trustee Server cannot communicate with Key HSM.

Affected Version: 6.0.0

Cloudera Bug: KT-3172

Workaround: Convert the Key Trustee Server private key to raw RSA format.

Known Issues in Cloudera Navigator Key HSM 6.0.0

Roll key command throws an exception and cannot retrieve metadata for key

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if there is significant (> 15 ms ping time) network latency between the Key Trustee Server and the HSM device, then EDEK generation errors can occur during the roll key operation. These errors manifest in the KMS log as errors on the `generateEncryptedKey` operation. The KMS will recover from these errors on its own, but they may represent a nuisance to the operator.

Affected Version: 6.0.0

Cloudera Bug: KT-5646

Workaround: When these errors occur, you can use the `hadoop key list -metadata` command to confirm whether or not the key roll was successful, despite the error condition.

KeySecure HSM not supported

The KeySecure HSM is not supported in this release.

Affected Version: 6.0.0

Cloudera Bug: KT-6001

Workaround: None.

Too many keys on Luna HSM causes Key HSM startup to fail

If there are too many keys on the Luna HSM, Key HSM startup will fail with a Java core dump because it times out querying the keys.

Affected Version: 6.0.0

Cloudera Bug: KT-6129

Fixed in Version: 6.1.0

Workaround: Increase the value in the `keyhsm.countdown.time` property in `application.properties` (the default is 45 seconds) until Key HSM starts successfully.

Key HSM Luna setup not showing the correct login status

When running the `keyhsm setup luna` command, you are prompted for the Luna HSM slot number and login password. Key HSM then attempts to log into the Luna HSM to verify these settings are correct. In some circumstances, the setup script reports that the login was successful, even if it failed.

Affected Version: 6.0.0

Cloudera Bug: KT-6623

Workaround: Any incorrect settings will cause the Key HSM service to throw an exception upon startup and exit. If the Key HSM service does not start correctly, check the log for the message: "Unable to sign into Luna HSM server."

Please rerun application with 'setup' option to configure password." If this message appears, re-run the `keyhsm setup luna` command, and enter the correct slot number and login password.

The `keyhsm trust` command fails

When performing setup of Key HSM, the `ktadmin keyhsm --server http://server:port --trust` command may fail with the following message: "Unable to connect to Key HSM server at this address. Is the server running, and is this KeyTrustee instance trusted (by running `keyhsm trust`)?" This failure can occur even if you have already successfully executed the `keyhsm trust` command.

Affected Version: 6.0.0

Cloudera Bug: KT-6131

Workaround: Re-execute the `keyhsm trust` command, and then retry the `ktadmin keyhsm` command.

Upgrading Key HSM removes init script and binary

Upgrading Key HSM from 1.4.x to 1.5.x and higher removes the Key HSM init script and `/usr/bin/keyhsm` binary.

Workaround: Reinstall Key HSM:

```
sudo yum reinstall keytrustee-keyhsm
```

Key HSM cannot trust Key Trustee Server certificate if it has extended attributes

Key HSM cannot trust the Key Trustee Server certificate if it has extended attributes, and therefore cannot integrate with Key Trustee Server.

Workaround: Import the Key Trustee Server certificate to the Key HSM trust store using Java `keytool` instead of the `keyhsm trust` command.

Known Issues in Cloudera Navigator Key Trustee KMS 6.0.0

CDH upgrade failure

When upgrading to Key Trustee KMS 6.0.0 from Key Trustee KMS 5.14.0 or lower, *and* performing a rolling restart (instead of a full restart), the first Key Trustee KMS instance to restart may fail to come up and present the error: "Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synched between all KMS hosts. Aborting to prevent data inconsistency."

Affected Versions: 6.0.0

Cloudera Bug: KT-6547

Workaround: If possible, perform a full restart instead of a rolling restart.

If you cannot execute a full restart, then add the following line to the `/var/lib/kms-keytrustee/keytrustee/.keytrustee/keytrustee.conf` file on all Key Trustee KMS instances, and then restart the Key Trustee KMS that failed:

```
"FINGERPRINT_VALIDATED" : "True"
```

New Key Trustee KMS may fail after being added to an environment that previously had a single Key Trustee KMS instance

When adding a new Navigator Key Trustee KMS instance to an environment that previously only had a single Key Trustee KMS instance, the new Key Trustee KMS may fail to start and return the following message:

"Unable to verify private key match between KMS hosts. If the system has been recently upgraded, DO NOT TAKE FURTHER ACTION and contact your support representative as soon as possible. If this is a new installation, verify private key files have been synced between all KMS hosts. Aborting to prevent data inconsistency."

This message correctly appears if GPG private keys have not been synced across all Key Trustee KMS instances. However, it may erroneously appear when the GPG private keys have been synced, but the original Key Trustee KMS instance has not been restarted to pick up the new configuration.

If you encounter this error message and have either recently upgraded the Key Trustee KMS, or have installed a new Key Trustee KMS service running in HA mode, refer to [Key Trustee KMS Encryption Issues](#) for troubleshooting steps.

Affected Version: 6.0.0

Cloudera Bug: KT-6231

Workaround: Restart the original Key Trustee KMS instance, and then start the new Key Trustee KMS instance.

The Key Trustee KMS service fails to start if the Trust Store is configured without also configuring the Keystore

If you configure the Key Trustee KMS service **Key Management Server Proxy TLS/SSL Certificate Trust Store File** and **Key Management Server Proxy TLS/SSL Certificate Trust Store Password** parameters without also configuring the **Key Management Server Proxy TLS/SSL Server JKS Keystore File Location** and **Key Management Server Proxy TLS/SSL Server JKS Keystore File Password** parameters, the Key Trustee KMS service does not start.

Workaround: Configure all Trust Store and Keystore parameters.

Key Trustee KMS backup script fails if PostgreSQL versions lower than 9.3 are installed

If PostgreSQL versions lower than 9.3 are installed on the Key Trustee KMS host, the `ktbackup.sh` script fails with an error similar to the following:

```
pg_dump: server version: 9.3.11; pg_dump version: 9.2.14
pg_dump: aborting because of server version mismatch
```

Workaround: Uninstall the lower PostgreSQL version.

Known Issues in Cloudera Navigator HSM KMS 6.0.0

Encryption zone key is not deleted after migrating from Key Trustee KMS to HSM KMS

After migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS service should be restarted. Until it is restarted, any keys that are deleted after the migration may still be cached. If a deleted key is cached, then the data encrypted with that key can still be accessed even though the key has been deleted.

Affected Version: 6.0.0, 6.1.0

Cloudera Bug: KT-6434

Workaround: Restart the HSM KMS service after the key migration is complete.

KT KMS migration to HSM KMS fails if a key is not found on HSM

When using Key Trustee KMS with Key Trustee Server and Key HSM (backed by an HSM device), if a key version is deleted on the HSM device directly without also deleting and purging that key version on the Key Trustee KMS, then attempts to migrate from the KT KMS to the HSM KMS will fail.

Affected Version: 6.0.0

Cloudera Bug: KT-5671

Workaround: None.

HSM KMS proxy fails to start during key migration

When migrating keys from the Key Trustee KMS to the HSM KMS, the HSM KMS Proxy may fail to start and return a fatal error.

Affected Version: 6.0.0

Cloudera Bug: KT-6431

Workaround: Navigate to the **Additional Java Configuration Options for KMS** setting in the HSM KMS configuration, and ensure that the HSM KMS truststore matches the truststore used by the Key Trustee KMS:

```
-Djavax.net.ssl.trustStore=/path/to/truststore
```

You can remove this option after completing the migration.

HSM KMS Thales Proxy role fails to start with non-default Thales HSM Server Port

Port 9001 is used by Cloudera Manager services, and it is also the default privileged port for the Thales HSM KMS. If you use port 9001 for the Thales HSM KMS, it will prevent the CM 6.0.0 upgrade from completing successfully.

Affected Version: 6.0.0

Cloudera Bug: KT-6403

Workaround: If the Thales HSM KMS already exists, then before upgrading to Cloudera Manager 6.0.0, you must change the privileged Thales HSM KMS port; the recommended port is 11501. The non-privileged port default is 9000, and does not need to be changed.

If you are newly installing the Thales HSM KMS on a 6.0.0 system, then you must set the port to a non-default value before adding the HSM KMS backed by Thales service in Cloudera Manager.

To change the privileged port, log into the Thales HSM KMS machine(s), and run the following commands:

```
# sudo /opt/nfast/bin/config-serverstartup --enable-tcp --enable-privileged-tcp
--privport=11501
[server_settings] change successful; you must restart the hardserver for this to take
effect
# sudo /opt/nfast/sbin/init.d-ncipher restart
-- Running shutdown script 90ncsnmpd

-- Running shutdown script 60raserv

...
'ncsnmpd' server now running
```

After successfully running the commands, restart the Thales HSM KMS.

Key store key password specified in Cloudera Manager is not properly used by HSM KMS Service

If the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore File Password and the Navigator HSM KMS Proxy TLS/SSL Server JKS Keystore Key Password are not both specified or differ from each other, then the HSM KMS inter-node handshake fails and HSM KMS High Availability is not configured.

Affected Version: 6.0.0

Cloudera Bug: KT-6020

Workaround: Use the same password for the keystore file and the keystore key file.

Timeout error during encryption zone key creation

There are situations where the key cache is synchronously populated to capacity during the create encryption zone operation. The expected behavior is that the key cache is synchronously populated only to the low watermark level (the rest of the keys should be created asynchronously).

Affected Version: 6.0.0

Cloudera Bug: KT-5296

Workaround: On the HSM KMS, in the field **HSM KMS Proxy Advanced Configuration Snippet (Safety Valve)** for kms-site.xml:

- hadoop.security.kms.encrypted.key.cache.low.watermark .05
- hadoop.security.kms.encrypted.key.cache.size 30

On the HDFS, in the field **HDFS Cluster-wide Advanced Configuration Snippet (Safety Valve)** for core-site.xml:

- hadoop.security.kms.client.encrypted.key.cache.size 30
- hadoop.security.kms.client.encrypted.key.cache.low-watermark .05

HSM KMS Luna may need to be restarted if inactive for extended period

If Hadoop key operations return com.safenetinc.luna.exception.LunaCryptokiException after the KMS has been running without activity for an extended period time, the Luna session may have been dropped.

Affected Version(s): 6.0.0

Cloudera Bug: KT-5018

Workaround: Restart the KMS service.

Creating multiple instances of HSM KMS on the same host and port causes an error upon delete

Creating a KMS role instance on a host that previously hosted a KMS role instance in the same role group that had its data directories deleted results in errors when attempting to run Hadoop key delete operations.

Affected Version(s): 6.0.0

Cloudera Bug: KT-4992

Workaround: This workaround requires the assistance of Cloudera support; request assistance with issue KT-4992

Incorrect status for "Restart stale services" step in HDFS encryption wizard post-service installation

There are times when completion of the HDFS Encryption Wizard does not show an active "Restart stale services and redeploy client configuration" link.

Affected Version: 6.0.0

Cloudera Bug: KT-4987

Workaround: Refresh the page and the link should become active.

The encryption wizard continues to fail if there is a failure during initial configuration run

The encryption wizard continues to fail if there was a failure during the initial run configuring HSM KMS.

Affected Version: 6.0.0

Cloudera Bug: KT-4909

Workaround: Open Cloudera Manager in another browser tab, and manually stop the installed KMS by clicking the arrow next to the KMS and selecting **Stop**. Then retry the installation in the new tab after correcting the cause of the install failure.

Before installing the Thales backed HSM KMS, you must add the KMS user to the nfast group

After installation of the Thales HSM client, and before installing Navigator HSM KMS backed by Thales HSM, you must add the KMS user to the nfast group..

Affected Version: 6.0.0

Cloudera Bug: KT-4618

Workaround: Run the following command to manually add the KMS user to the nfast group:
`usermod -a -G nfast kms`

Known Issues in Cloudera Navigator Encrypt 6.0.0

Navigator Encrypt-related packages should be downgraded with the Navigator Encrypt package

The navencrypt downgrade command will only downgrade the navencrypt package. It will not downgrade the associated navencrypt-kernel-module and libkeytrustee packages.

Affected Version: 6.0.0

Cloudera Bug: KT-6381

Workaround: To fully downgrade Navigator Encrypt, manually downgrade all of the associated Navigator Encrypt packages (in the order listed):

1. navencrypt
2. navencrypt-kernel-module
3. libkeytrustee

Navigator Encrypt will not build on RHEL kernel 3.10.0-862.14.4

The Navigator Encrypt kernel module will not build on RHEL kernel 3.10.0-862.14.4. This impacts new installations, and existing installations that are upgrading to kernel 3.10.0-862.14.4. This issue prevents the navencrypt-mount service from running and Navigator Encrypt mount points from being accessible.

Affected Versions: 6.0.0

Fixed in Version: 6.1.0

Cloudera Bug: KT-6677

Workaround: Upgrade to Navigator Encrypt 6.1.0 before upgrading to RHEL kernel 3.10.0-862.14.4.

Upgrade from Navigator Encrypt 3.x to 6.0.0 does not trigger a navcryptfs recompile

After upgrading from Navigator Encrypt 3.x to Navigator Encrypt 6.0.0, the kernel module may not be rebuilt, even if the upgrade and installation commands indicate that it was built successfully.

Affected Version: 6.0.0

Cloudera Bug: KT-6382

Workaround: If the version currently loaded is not the desired version, use the following commands to update it. In this case, you must manually load the updated kernel module.

To identify which version of the kernel module is being used, look for the currently loaded version in the `dmesg` output (in this case, the output reveals that kernel module version was not rebuilt):

```
navencrypt: time=1531428589 level=INFO app="kernel" Cloudera navcryptfs v3.14.0-405
loaded
```

For RHEL/UBUNTU:

```
service navencrypt-mount stop
navencrypt-module-setup
service navencrypt-mount start
```

For SLES:

```
service navencrypt-mount stop
rmmod navcryptfs
modprobe -v navcryptfs
service navencrypt-mount start
```

When a mount point is added Navigator Encrypt updates configuration files before the action completes

The `navencrypt-prepare` command sometimes performs updates to the `/etc/navencrypt/control` and `/etc/navencrypt/ztab` files before the command has completed. In such cases, if there is an error with the mounting or unmounting of the navencrypt mount point, then the updated control and `ztab` files will not accurately reflect which mount points exist. Do *not* manually change these files without having a known working backup.

Affected Version: 6.0.0

Cloudera Bug: KT-6383

Workaround: None.

Issue with `navencrypt-mount` service status message

Affected Version: 6.0.0

Cloudera Bug: KT-6309

On RHEL 7, the service `navencrypt-mount stop` command might return a successful exit status message, even if there was a problem stopping Navigator Encrypt. This is not an issue with the `navencrypt-mount` service command; rather, it is a problem with the message output.

You can verify that `navencrypt-mount` stopped successfully by checking for the presence of the `navcryptofs` module in the `lsmod` output:

```
# lsmod | grep navencrypt
navcryptofs 101407 0
```

Alternatively, you can verify that the `navencrypt` mount points successfully dismounted by checking the output of the `df` or `mount` commands.

Workaround: None.

Deprecated Items

This page lists operating systems, Java versions, databases, platforms, CDH components and subcomponents, and product functionality that have been deprecated or removed.

Deprecated Items

A deprecated item is a feature, component, platform, or functionality that Cloudera is planning to remove in a future release. Cloudera supports items that are deprecated until they are removed, and the deprecation gives customers time to plan for removal.

The following table lists deprecated items:

- [Table 45: CDH Components, Subcomponents, and Product Functionality](#) on page 667

Table 45: CDH Components, Subcomponents, and Product Functionality

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support Is Removed
Activity Monitor	Activity Monitor is only used and deployed by Cloudera Manager when the MapReduce service (MRv1) is deployed.	5.9.0	7.0.0
Apache Crunch	Apache Crunch is deprecated, and will be	6.0.0	To be determined.

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support Is Removed
	removed in a future release. Cloudera recommends using Spark 2 instead. Additionally, as of CDH 6.0.0, Crunch is available only as Maven artifacts from the Cloudera Maven repository. For more information, see Apache Crunch Guide .		
Hive CLI	About Hive	5.0.0	To be determined.
Kite Dataset API	Kite Dataset API is deprecated, and will be removed in a future release.	6.0.0	To be determined.
Kudu Flume sink configuration parameters	The <code>producer.skipMissingColumn</code> , <code>producer.skipBadColumnValue</code> , and <code>producer.warnUnmatchedRows</code> Kudu Flume sink configuration parameters have been deprecated in favor of <code>producer.missingColumnPolicy</code> , <code>producer.badColumnValuePolicy</code> , and <code>producer.unmatchedRowPolicy</code> respectively.	6.1.0	
Python 2.6	Python 2.6, packaged with RHEL6, is deprecated. Key Trustee Server uses the utilities <code>ktadmin</code> and <code>keytrustee-orgtool</code> , which use the native version of Python that is packaged with the host OS. <ul style="list-style-type: none"> • Initializing Standalone Key Trustee Server • Managing Key Trustee Server Organizations 	6.0.0	To be determined.

Table 46: Cloudera Manager Components, Subcomponents, and Product Functionality

Item	Related Information	Release in Which Item Is Removed	Release in Which Support Is Removed
Single User Mode		6.0.0	TBD

Removed Items

A removed item is a feature, component, platform, or functionality that has been removed from the product and is no longer supported. Documentation for the Feature, component, platform, or functionality has also been removed.

The following tables list the items removed in version 6.0:

- [Table 47: Operating Systems, Java Versions, Databases, and Platforms](#) on page 669
- [Table 48: CDH Components, Subcomponents, and Product Functionality](#) on page 670
- [Table 49: Cloudera Manager Components, Subcomponents, and Product Functionality](#) on page 671

Table 47: Operating Systems, Java Versions, Databases, and Platforms

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support Is Removed
Operating System			
• Ubuntu 14.04	CDH and Cloudera Manager Supported Operating Systems on page 20	5.8.0	5.16.1, 6.0.0
Operating System			
• SLES 11	CDH and Cloudera Manager Supported Operating Systems on page 20	5.9.0	6.0.0
Operating System			
• RHEL 5, CentOS 5, Oracle Enterprise Linux 5	CDH and Cloudera Manager Supported Operating Systems on page 20	July 2015	6.0.0
• Ubuntu 10.04 (already EOL by Canonical) and 12.04			
• Debian 7			
Java 7	Java Requirements on page 26	5.0.0	6.0.0
Database			
• MySQL 5.0		July 2015	6.0.0
• PostgreSQL 8.1			
Database	Database Requirements on page 23	5.9.0	6.0.0
• MySQL 5.1			
• PostgreSQL 8.4			
Database		5.9.0	6.0.0
• Oracle 11g			
Tarball			
CDH Tarball Distribution		5.9.0	6.0.0
Filesystem	S3 and S3n are replaced by S3a. Storing HBase Snapshots on Amazon S3 and Copying Cluster Data Using DistCp		
• Amazon S3 and S3n connectors			
Cloudera Manager Tarball Distribution		5.9.0	6.0.0

Table 48: CDH Components, Subcomponents, and Product Functionality

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support Is Removed
AsyncHBaseSink	AsyncHBaseSink is incompatible with HBase 2.0 and you can no longer use AsyncHBaseSink with Apache Flume. For information about using HBase2Sink with Apache Flume, see Using Flume .	6.0.0	6.0.0
DataFu	See Removal of the Apache DataFu Pig JAR from CDH 6 on page 546	5.9.0	6.0.0
HBaseSink	HBaseSink has been replaced with HBase2Sink, which is compatible with HBase 2.0. For information about using HBase2Sink with Apache Flume, see Using Flume .	6.0.0	6.0.0
hbck read-write repair mode	hbck is only available in a read-only inconsistency identifying mode. See Checking Consistency in HBase Tables for more information.	6.0.0	6.0.0
HFTP	Use WebHDFS	5.10.1	6.0.0
HiveServer1		5.3.0	6.0.0
Key HSM Debug Startup	To get debug information during start up, set the root log level to <code>debug</code> in the <code>src/main/resources/conf/kms-env.sh</code> file.	6.1.0	6.1.0
kudu perf loadgen tool configuration options	The <code>-table_num_buckets</code> configuration option of the kudu perf loadgen tool is now removed in favor of <code>-table_num_hash_partitions</code> and <code>-table_num_range_partitions</code>	6.1.0	
Legacy Scala clients for Kafka (consumer and producer)	The legacy Scala clients (producer and consumer) that are under the <code>kafka.producer.*</code> and <code>kafka.consumer.*</code> package.	CDK 3.0.0 and CDH 6.0.0	6.1.0
Llama		5.5.0	6.0.0

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support Is Removed
MRv1, MapReduce v1 APIs, MapReduce service	Migrating from MapReduce 1 (MRv1) to MapReduce 2 (MRv2)	5.0.0	6.0.0
Mahout		5.5.0	6.0.0
Management of Key Trustee Server without Cloudera Manager	Cloudera Navigator Key Trustee Server	5.9.0	6.0.0
MR Pipes		5.9.0	6.0.0
Navigator Encrypt Filesystem-Level Encryption Using eCryptfs	Block-Level Encryption with dm-crypt There is no support for creating new eCryptfs mount points. Previously existing eCryptfs mount points are not affected.	July 2015	6.0.0
Navigator Encrypt migration command	The navencrypt-migration command is deprecated, and has been removed.	February 1, 2018	6.0.0
Old NameNode UI		5.5.0	6.0.0
Oozie Hive Action	Oozie Hive 2 Action Extension	5.7.0	6.0.0
Parquet library with group ID com.twitter	Using Apache Parquet Data Files with CDH	6.0.0	6.0.0
Parquet methods for reading metadata on the client side	Parquet Incompatible Changes	6.0.0	6.0.0
Sentry policy files	Migrating from Sentry Policy Files to the Sentry Service	5.8.0	6.0.0
Spark Standalone		5.5.0	6.0.0
Spark 1.x		5.13	6.0.0
Sqoop2		5.9.0	6.0.0
Whirr		5.5.0	6.0.0
YARN Capacity Scheduler	Use FairScheduler	5.9.0	6.0.0

Table 49: Cloudera Manager Components, Subcomponents, and Product Functionality

Item	Related Information	Release in Which Item Is Deprecated	Release in Which Support is Removed
Multi Cloudera Manager Dashboard		5.11.1	6.0
Cloudera Manager API Versions v1 - 5	Cloudera Manager API Documentation	5.13	6.0

Appendix: Apache License, Version 2.0

SPDX short identifier: Apache-2.0

Apache License
Version 2.0, January 2004
<http://www.apache.org/licenses/>

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims

licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

1. You must give any other recipients of the Work or Derivative Works a copy of this License; and
2. You must cause any modified files to carry prominent notices stating that You changed the files; and
3. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
4. If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions.

Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions.

Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks.

This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty.

Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability.

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability.

Appendix: Apache License, Version 2.0

While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

APPENDIX: How to apply the Apache License to your work

To apply the Apache License to your work, attach the following boilerplate notice, with the fields enclosed by brackets "[]" replaced with your own identifying information. (Don't include the brackets!) The text should be enclosed in the appropriate comment syntax for the file format. We also recommend that a file or class name and description of purpose be included on the same "printed page" as the copyright notice for easier identification within third-party archives.

```
Copyright [yyyy] [name of copyright owner]

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
```