

Hortonworks Data Platform

HDP-2.4.0 Release Notes

(Mar 1, 2016)

Hortonworks Data Platform: HDP-2.4.0 Release Notes

Copyright © 2012-2016 Hortonworks, Inc. Some rights reserved.

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner.

The Hortonworks Data Platform consists of the essential set of Apache Software Foundation projects that focus on the storage and processing of Big Data, along with operations, security, and governance for the resulting system. This includes Apache Hadoop – which includes MapReduce, Hadoop Distributed File System (HDFS), and Yet Another Resource Negotiator (YARN) – along with Ambari, Falcon, Flume, HBase, Hive, Kafka, Knox, Oozie, Phoenix, Pig, Ranger, Slider, Spark, Sqoop, Storm, Tez, and ZooKeeper. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, [training](#) and partner-enablement services. All of our technology is, and will remain, free and open source.

Please visit the [Hortonworks Data Platform](#) page for more information on Hortonworks technology. For more information on Hortonworks services, please visit either the [Support](#) or [Training](#) page. Feel free to [contact us](#) directly to discuss your specific needs.



Except where otherwise noted, this document is licensed under
Creative Commons Attribution ShareAlike 3.0 License.
<http://creativecommons.org/licenses/by-sa/3.0/legalcode>

Table of Contents

1. HDP 2.4.0 Release Notes	1
1.1. New Features	2
1.2. Unsupported Features	2
1.2.1. Technical Preview Features	2
1.2.2. Community Features	3
1.3. Upgrading from HDP 2.3.x to HDP 2.4.0	4
1.4. HDP 2.4.0 Repositories	5
1.5. Behavioral Changes	8
1.6. Apache Patch Information	11
1.6.1. Hadoop	12
1.6.2. Accumulo	25
1.6.3. Atlas	27
1.6.4. Calcite	29
1.6.5. Falcon	31
1.6.6. Flume	36
1.6.7. HBase	38
1.6.8. Hive	40
1.6.9. Kafka	53
1.6.10. Knox	56
1.6.11. Mahout	57
1.6.12. Oozie	57
1.6.13. Phoenix	58
1.6.14. Pig	60
1.6.15. Ranger	61
1.6.16. Slider	67
1.6.17. Spark	68
1.6.18. Sqoop	72
1.6.19. Storm	72
1.6.20. Tez	74
1.6.21. ZooKeeper	79
1.7. Common Vulnerabilities and Exposures	80
1.8. Third-party Licenses	81
1.9. Fixed Issues	81
1.10. Known Issues	83
1.11. Documentation Errata	88
1.11.1. Flume: Kafka Sink	88
1.11.2. Hive Sink	89

List of Tables

1.1. New Features	2
1.2. Technical Previews	2
1.3. Community Features	3
1.4. Debian 6 (Deprecated)	5
1.5. Debian 7	6
1.6. RHEL/CentOS/Oracle LINUX 6	6
1.7. RHEL/CentOS/Oracle LINUX 7	6
1.8. SLES 11 SP3/SP4	7
1.9. Ubuntu 12	7
1.10. Ubuntu 14	7
1.11. Helper Files	8
1.12. Behavioral Changes	8

1. HDP 2.4.0 Release Notes

This document provides you with the latest information about the HDP 2.4.0 release and its product documentation.

Component Versions

The official Apache versions of most HDP 2.4.0 components are unchanged from HDP 2.3.4.0, with the exception of Spark. Spark is upgraded from 1.5.2 to 1.6.0. See more details of Spark 1.6.0 in the [New Features](#) section. All HDP components listed here are official Apache releases of the most recent stable versions available.

Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP components should remain at the following package version levels to ensure a certified and supported copy of HDP 2.4.0.

Official Apache versions for HDP 2.4.0.

- Apache Accumulo 1.7.0
- Apache Atlas 0.5.0
- Apache Calcite 1.2.0
- Apache DataFu 1.3.0
- Apache Falcon 0.6.1
- Apache Flume 1.5.2
- Apache Hadoop 2.7.1
- Apache HBase 1.1.2
- Apache Hive 1.2.1
- Apache Kafka 0.9.0
- Apache Knox 0.6.0
- Apache Mahout 0.9.0+
- Apache Oozie 4.2.0
- Apache Phoenix 4.4.0
- Apache Pig 0.15.0
- Apache Ranger 0.5.0
- Apache Slider 0.80.0
- Apache Solr 5.2.1
- Apache Spark 1.6.0
- Apache Sqoop 1.4.6

- Apache Storm 0.10.0
- Apache Tez 0.7.0
- Apache ZooKeeper 3.4.6

Additional component versions:

- Cascading 3.0.1
- Hue 2.6.1

1.1. New Features

This section highlights the new feature in HDP 2.4.0.

Table 1.1. New Features

Component	Feature
Spark 1.6.0	General availability of Spark 1.6.0 Dynamic executor allocation

1.2. Unsupported Features

Some features exist within HDP 2.4.0, but Hortonworks does not currently support these specific capabilities.

- [Technical Preview Features](#)
- [Community Features](#)

1.2.1. Technical Preview Features

The following features are available within HDP 2.4.0, but are not ready for production deployment. We encourage you to explore these technical preview features in non-production environments and provide feedback on your experiences through the [Hortonworks Community Forums](#).

Table 1.2. Technical Previews

Component	Feature
HBase and Phoenix	<ul style="list-style-type: none">• Phoenix Query Server• Phoenix Query Server (PHOENIX-971)• Phoenix-Spark Integration• RPC Throttling• Support for <code>init.d</code> scripts
Hive	<ul style="list-style-type: none">• Hive Streaming• ACID support
Kafka	<ul style="list-style-type: none">• SSL
Slider	<ul style="list-style-type: none">• Support for Docker-based application packaging (SLIDER-780)

Component	Feature
Spark	<ul style="list-style-type: none"> • GraphX • SparkR • Zeppelin
Storm	<ul style="list-style-type: none"> • Elastic topology via YARN/Slider • JDBC Bolt and Trident • Monitoring of Storm topologies and clusters • Storm-Slider • User Impersonation
YARN	<ul style="list-style-type: none"> • Add support for network I/O isolation/scheduling for containers (YARN-2140) • NodeManager: add cgroup support for disk I/O isolation (YARN-2619) • Spark-ATS integration (BUG-49244)
Zeppelin	<ul style="list-style-type: none"> • Zeppelin 0.6.0 • Install & Manage Zeppelin with Ambari 2.2.1 • PySpark support in Zeppelin - improvements • Zeppelin support for SSL (one way) • Zeppelin support for form-based LDAP authentication

1.2.2. Community Features

The following features are developed and tested by the community, but are not officially supported by Hortonworks. There are variety of reasons that these features are excluded, including: insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, feature deviates from Hortonworks best practices, and more. Do not use them in your production environments.

Table 1.3. Community Features



Component	Feature
Falcon	<ul style="list-style-type: none"> • Prism Server • User Recipes
HBase	<ul style="list-style-type: none"> • HBase Column Family Encryption: use HDFS data at rest encryption instead • Use of memcached as block cache is unsupported (HBASE-13170) • ZooKeeper-less region assignment • Region size balancing (HBASE-13103)
HDFS	<ul style="list-style-type: none"> • block-volume device choosing (HDFS-1804) • NameNode Federation (HDFS-1052) • viewFS (HADOOP-7257)
Kafka	<ul style="list-style-type: none"> • Mirror Maker (not supported when Kafka security is active)


Component	Feature
	<ul style="list-style-type: none"> • New Consumer API
Knox	<ul style="list-style-type: none"> • Storm REST APIs
Oozie	<ul style="list-style-type: none"> • Spark action (OOZIE-1983)
Slider	<ul style="list-style-type: none"> • Simplified Application Packaging
Spark	<ul style="list-style-type: none"> • Spark Standalone • Spark on Mesos • Jupyter/iPython Notebook • Oozie Spark action is not supported, but there is a tech note available for HDP customers
YARN	<ul style="list-style-type: none"> • Fair Scheduler • MapReduce Eclipse Plug-in • MapReduce Uber AM

1.3. Upgrading from HDP 2.3.x to HDP 2.4.0

HDP 2.4.0 is a feature-bearing minor release of HDP. For full instructions on how to upgrade your cluster from HDP 2.3.x to HDP 2.4.0, use the Non-Ambari Cluster Upgrade Guide. [Non-Ambari Cluster Upgrade Guide](#)

The following table summarizes HDP 2.2.x-to-2.4.0 upgrade options:

Cluster Management	Supporting Doc	Notes
Cluster managed manually (HDP 2.1 and earlier)	Non-Ambari Cluster Upgrade Guide	
Cluster managed manually (HDP 2.2 and later)	Non-Ambari Cluster Upgrade Guide	
Cluster managed via Ambari 1.7.0	Upgrading Ambari Guide	
Cluster managed via Ambari 2.0	Upgrading Ambari Guide	<p>Ambari 2.0 supports rolling upgrade between HDP 2.2.x and HDP 2.4.0.</p> <div>  <p>Note</p> <p>Ambari does not support rolling upgrade between HDP 2.1 and 2.3 for Falcon. Use Configure and Validate Falcon to upgrade this component.</p> </div>
Cluster managed via Ambari 2.1	Upgrading Ambari Guide	<p>Ambari 2.1 supports rolling upgrade between HDP 2.3.x and HDP 2.4.0.</p> <div>  <p>Note</p> <p>Ambari does not support rolling upgrade between HDP 2.1 and 2.3 for Falcon. Use Configure and Validate Falcon to upgrade this component.</p> </div>
Cluster managed via Ambari 2.2	Upgrading Ambari Guide	<p>Ambari 2.2 supports rolling upgrade between HDP 2.3.x and HDP 2.4.0.</p>

Cluster Management	Supporting Doc	Notes
		 Note Ambari does not support rolling upgrade between HDP 2.1 and 2.3 for Falcon. Use Configure and Validate Falcon to upgrade this component.

1.4. HDP 2.4.0 Repositories

HDP 2.4.0 for Linux supports the following:

- [Debian 6 \(Deprecated\)](#)
- [Debian 7](#)
- [RHEL/CentOS/Oracle LINUX 6](#)
- [RHEL/CentOS/Oracle LINUX 7](#)
- [SLES 11 SP3/SP4](#)
- [Ubuntu 12](#)
- [Ubuntu 14](#)

You can also view the [Helper Files](#).



Note

The package identifier for HDP 2.4.0 components is **169**. For example:

```
dfs -mkdir -p /hdp/apps/2.4.0.0-$BUILD/hive/
```

would become:

```
dfs -mkdir -p /hdp/apps/2.4.0.0-169/hive/
```

Table 1.4. Debian 6 (Deprecated)

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian6-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian6-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm

Description	Link
Slider app packages	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.5. Debian 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian7-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian7-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.6. RHEL/CentOS/Oracle LINUX 6

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos6-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos6-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.7. RHEL/CentOS/Oracle LINUX 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos7-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos7-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos7/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.8. SLES 11 SP3/SP4

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/HDP-2.4.0.0-suse11sp3-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/HDP-2.4.0.0-suse11sp3-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.9. Ubuntu 12

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu12-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu12-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.10. Ubuntu 14

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu14-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu14-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm

Description	Link
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.11. Helper Files

Description	Link
Companion files	http://public-repo-1.hortonworks.com/HDP/tools/2.4.0.0/hdp_manual_install_rpm_helper_files-2.4.0.0.169.tar.gz

1.5. Behavioral Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.4.0, behavioral changes affect the following Hadoop components.

Table 1.12. Behavioral Changes

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
			Debian 6 support deprecated.	HDP support for Debian 6 is deprecated with HDP 2.4.0. Support for Debian 6 will be removed in a future HDP release.
BUG-43020	Hue		CherryPy is not the the default webserver	<p>Scenario: In previous releases of Hue, CherryPy was not the default webserver, but spawning webserver, which is generally not advisable as CherryPy is multithreaded and scales better than spawning webserver.</p> <p>Previous Behavior: By not commenting out <code>use_cherry_server=false</code> in <code>hue.ini</code>, Hue used spawning webserver instead of Hue.</p> <p>New Behavior: CherryPy is the default webserver for Hue.</p> <p>Workaround/Expected Customer Action: No action required, unless spawning web server is the desired webserver for Hue. In this case, modify <code>use_cherry_server</code> accordingly.</p>

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
BUG-50292	Hue		HUE DataTables Warning in partition columns tab.	<p>Scenario: In previous releases of Hue, an error message appeared when Partition Columns tab was clicked in the Beeswax describe table view.</p> <p>Previous Behavior: When clicking the Partition Columns tab in the Beeswax describe table view, an error message appeared.</p> <p>New Behavior: No error message appears.</p> <p>Workaround: None.</p>
BUG-50775	YARN		Enable collection of debug information by default	<p>Scenario: Container failures during job execution.</p> <p>Previous Behavior: New feature - no previous behaviour.</p> <p>New Behavior: This feature(enabled by default) enables generation of additional information which is useful for debugging container failures. The extra information generated is:</p> <ul style="list-style-type: none"> • A copy of the <code>launch_container.sh</code> script. • A listing of the contents of the container work directory. <p>This information is automatically collected if log-aggregation is enabled, or ignored and discarded if log-aggregation is disabled. The feature can be disabled by setting <code>yarn.nodemanager.log-container-debug-info.enabled</code> to <code>false</code> in <code>yarn-site.xml</code>.</p>

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
BUG-50786	Tez		Increase the default value for MAX_COUNTERS	<p>The default value of the max counter limit has been increased from 2000 to:</p> <ul style="list-style-type: none"> tez.counters.max: 10000 tez.counters.max.groups: 3000
BUG-51480	Spark		We should recommend to run spark history server on top of HDFS	<p>In HDP 2.4, the Spark History Service reads data from HDFS. Prior to HDP 2.4, the data was read from the YARN ATS (Application Timeline Service). Refer to the Spark Guide for more info.</p>
BUG-51481	Spark		Behavior Change for hdp.version for spark configuration from HDP 2.3.4.	<p>Previous Behavior: Before HDP 2.3.4, user need to configure hdp.version variable in spark-defaults.conf:</p> <pre>spark.driver.extraJavaOptions -Dhdp.version= \$VERSION\$BUILD spark.yarn.am.extraJavaOptions -Dhdp.version= \$VERSION\$BUILD</pre> <p>New Behavior: As of HDP 2.4.0: when running HDP, Spark users no longer need to specify hdp.version through the Spark configuration files; it is inferred from the version of HDP installed on the client machine.</p> <p>There are a few exceptions to this rule:</p> <ul style="list-style-type: none"> If you are programmatically submitting jobs in a way that spark-env.sh is not executed during the submit step. If you want to specify a different cluster version than what is installed on the client.
	Ambari, Spark		The Spark shuffle service is enabled by default with this	

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
			release. It is necessary for Spark Dynamic Executor Allocation to work correctly.	

1.6. Apache Patch Information

The following sections list patches in each HDP 2.4.0 component beyond what was fixed in the base version of the Apache component.

- [Hadoop](#)
- [Accumulo](#)
- [Atlas](#)
- [Calcite](#)
- [Falcon](#)
- [Flume](#)
- [HBase](#)
- [Hive](#)
- [Kafka](#)
- [Knox](#)
- [Mahout](#)
- [Oozie](#)
- [Phoenix](#)
- [Pig](#)
- [Ranger](#)
- [Slider](#)
- [Spark](#)
- [Sqoop](#)
- [Storm](#)
- [Tez](#)
- [Zookeeper](#)

1.6.1. Hadoop

HDP 2.4.0 provides the following Apache patches:

- [HADOOP-10406](#): TestIPC.testIpcWithReaderQueuing may fail.
- [HADOOP-12551](#): Introduce FileNotFoundException for WASB FileSystem API.
- [HADOOP-12608](#): Fix exception message in WASB when connecting with anonymous credential.
- [HADOOP-12678](#): Handle empty rename pending metadata file during atomic rename in redo path.
- [HDFS-8729](#): Fix TestFileTruncate#testTruncateWithDataNodesRestartImmediately which occasionally failed.
- [HDFS-9358](#): TestNodeCount#testNodeCount timed out.
- [HDFS-9406](#): FSImage may get corrupted after deleting snapshot.
- [HDFS-9672](#): o.a.h.hdfs.TestLeaseRecovery2 fails intermittently.
- [MAPREDUCE-6566](#): Add retry support to mapreduce CLI tool.
- [MAPREDUCE-6618](#): YarnClientProtocolProvider leaking the YarnClient thread.
- [MAPREDUCE-6621](#): Memory Link in JobClient#submitJobInternal().
- [YARN-3480](#): Remove attempts that are beyond max-attempt limit from state store.
- [YARN-4309](#): Add container launch related debug information to container logs when a container fails.
- [YARN-4497](#): RM might fail to restart when recovering apps whose attempts are missing.
- [YARN-4565](#): Sometimes when sizeBasedWeight FairOrderingPolicy is enabled, under stress appears that cluster is virtually in deadlock.
- [YARN-4584](#): RM startup failure when AM attempts greater than max-attempts.
- [YARN-4625](#): ApplicationSubmissionContext and ApplicationSubmissionContextInfo more consistent.

HDP 2.3.4 provided the following Apache patches:

- [HADOOP-11098](#): [JDK8] Max Non Heap Memory default changed between JDK7 and 8.
- [HADOOP-11628](#): SPNEGO auth does not work with CNAMEs in JDK8.
- [HADOOP-11685](#): StorageException complaining "no lease ID" during HBase distributed log splitting.

- [HADOOP-11918](#): Listing an empty s3a root directory throws FileNotFoundException.
- [HADOOP-11932](#): MetricsSinkAdapter may hang when being stopped.
- [HADOOP-12049](#) Control http authentication cookie persistence via configuration.
- [HADOOP-12089](#): StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- [HADOOP-12186](#) ActiveStandbyElector shouldn't call monitorLockNodeAsync multiple times.
- [HADOOP-12239](#): StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- [HADOOP-12324](#): Better exception reporting in SaslPlainServer.
- [HADOOP-12334](#): Change Mode Of Copy Operation of HBase WAL Archiving to bypass Azure Storage Throttling after retries.
- [HADOOP-12350](#): WASB Logging; Improve WASB Logging around deletes, reads, and writes.
- [HADOOP-12350](#) WASB Logging: Improve WASB Logging around deletes, reads and writes.
- [HADOOP-12407](#): Test failing; hadoop.ipc.TestSaslRPC.
- [HADOOP-12413](#): AccessControlList should avoid calling getGroupNames in isUserInList with empty groups.
- [HADOOP-12437](#) Allow SecurityUtil to lookup alternate hostnames.
- [HADOOP-12438](#): TestLocalFileSystem tests can fail on Windows after HDFS-8767 fix for handling pipe.
- [HADOOP-12440](#): TestRPC#testRPCServerShutdown did not produce the desired thread states before shutting down.
- [HADOOP-12441](#): Fixed kill-command behavior to work correctly across Oses by using bash shell built-in.
- [HADOOP-12463](#) Fix TestShell.testGetSignalKillCommand failure on windows.
- [HADOOP-12484](#): Single File Rename Throws Incorrectly In Potential Race Condition Scenarios.
- [HADOOP-12508](#): delete fails with exception when lease is held on blob.
- [HADOOP-12533](#): Introduce FileNotFoundException in WASB for read and seek API.
- [HADOOP-12540](#): TestAzureFileSystemInstrumentation#testClientErrorMetrics fails intermittently due to assumption that a lease error will be thrown.

- [HADOOP-12542](#): TestDNS fails on Windows after HADOOP-12437.
- [HADOOP-12577](#) Bump up commons-collections version to 3.2.2 to address a security flaw.
- [HADOOP-12617](#) SPNEGO authentication request to non-default realm gets default realm name inserted in target server principal.
- [HBASE-268](#) Rack locality improvement.
- [HDFS-4015](#): Safemode should count and report orphaned blocks.
- [HDFS-4015](#) Safemode should count and report orphaned blocks.
- [HDFS-4366](#) Block Replication Policy Implementation May Skip Higher-Priority Blocks for Lower-Priority Blocks.
- [HDFS-4937](#) ReplicationMonitor can infinite-loop in BlockPlacementPolicyDefault#chooseRandom.
- [HDFS-6481](#) DatanodeManager#getDatanodeStorageInfos() should check the length of storageIDs.
- [HDFS-6581](#) Support for writing to single replica in RAM.
- [HDFS-7390](#) Provide JMX metrics per storage type.
- [HDFS-7483](#) Display information per tier on the Namenode UI.
- [HDFS-7725](#) Incorrect "nodes in service" metrics caused all writes to fail.
- [HDFS-7858](#) Improve HA Namenode Failover detection on the client.
- [HDFS-7928](#): Scanning blocks from disk during rolling upgrade startup takes a lot of time if disks are busy.
- [HDFS-7928](#) Scanning blocks from disk during rolling upgrade startup takes a lot of time if disks are busy.
- [HDFS-8099](#) Change "DFSInputStream has been closed already" message to debug log level.
- [HDFS-8209](#) Support different number of datanode directories in MiniDFSCluster.
- [HDFS-8554](#): TestDatanodeLayoutUpgrade fails on Windows.
- [HDFS-8656](#): Preserve compatibility of ClientProtocol#rollingUpgrade after finalization.
- [HDFS-8696](#): Make the lower and higher watermark in the DN Netty server configurable.
- [HDFS-8778](#): TestBlockReportRateLimiting#testLeaseExpiration can deadlock.
- [HDFS-8785](#) TestDistributedFileSystem is failing in trunk.

- [HDFS-8809](#): HDFS fsck reports under construction blocks as CORRUPT.
- [HDFS-8829](#): Make SO_RCVBUF and SO_SNDBUF size configurable for DataTransferProtocol sockets and allow configuring auto-tuning.
- [HDFS-8846](#): Add a unit test for INotify functionality across a layout version upgrade.
- [HDFS-8855](#): Webhdfs client leaks active NameNode connections.
- [HDFS-8930](#): Block report lease may leak if the 2nd full block report comes when NN is still in safemode.
- [HDFS-8950](#): NameNode refresh doesn't remove DataNodes that are no longer in the allowed list.
- [HDFS-8965](#): Harden edit log reading code against out of memory errors.
- [HDFS-8965](#): Harden edit log reading code against out of memory errors.
- [HDFS-8969](#): Clean up findbugs warnings for HDFS-8823 and HDFS-8932.
- [HDFS-9008](#): Balancer#Parameters class could use a builder pattern.
- [HDFS-9019](#): Adding informative message to sticky bit permission denied exception.
- [HDFS-9063](#): Correctly handle snapshot path for getContentSummary.
- [HDFS-9082](#): Change the log level in WebHdfsFileSystem.initialize() from INFO to DEBUG.
- [HDFS-9083](#): Replication violates block placement policy.
- [HDFS-9107](#): Prevent NNs unrecoverable death spiral after full GC.
- [HDFS-9112](#): Improve error message for Haadmin when multiple name service IDs are configured.
- [HDFS-9112](#): Improve error message for Haadmin when multiple name service IDs are configured.
- [HDFS-9128](#): TestWebHdfsFileContextMainOperations and TestSWebHdfsFileContextMainOperations fail due to invalid HDFS path on Windows.
- [HDFS-9142](#): Separating Configuration object for namenode(s) in MiniDFSCluster.
- [HDFS-9175](#): Change scope of 'AccessTokenProvider.getAccessToken()' and 'CredentialBasedAccessTokenProvider.getCredential()' abstract methods to public.
- [HDFS-9178](#): Slow datanode I/O can cause a wrong node to be marked bad.
- [HDFS-9184](#): Logging HDFS operation's caller context into audit logs.
- [HDFS-9205](#): Do not schedule corrupt blocks for replication.

- [HDFS-9220](#): Reading small file greater than 512 bytes that is open for append fails due to incorrect checksum.
- [HDFS-9273](#): ACLs on root directory may be lost after NN restart.
- [HDFS-9294](#): DFSClient deadlock when close file and failed to renew lease.
- [HDFS-9305](#): Delayed heartbeat processing causes storm of subsequent heartbeats.
- [HDFS-9311](#): Support optional offload of NameNode HA service health checks to a separate RPC server.
- [HDFS-9343](#): Empty caller context considered invalid.
- [HDFS-9354](#): Fix TestBalancer#testBalancerWithZeroThreadsForMove on Windows.
- [HDFS-9362](#): TestAuditLogger#testAuditLoggerWithCallContext assumes Unix line endings, fails on Windows.
- [HDFS-9364](#): Unnecessary DNS resolution attempts when creating NameNodeProxies.
- [HDFS-9384](#): TestWebHdfsContentLength intermittently hangs and fails due to TCP conversation mismatch between client and server.
- [HDFS-9397](#): Fix typo for readChecksum() LOG.warn in BlockSender.java.
- [HDFS-9413](#): getContentSummary() on standby should throw StandbyException.
- [HDFS-9426](#): Rollingupgrade finalization is not backward compatible.
- [HDFS-9434](#): Recommission a datanode with 500k blocks may pause NN for 30 seconds for printing info log messages.
- [MAPREDUCE-5485](#): Allow repeating job commit by extending OutputCommitter API.
- [MAPREDUCE-6273](#): HistoryFileManager should check whether summaryFile exists to avoid FileNotFoundException causing HistoryFileInfo into MOVE_FAILED state.
- [MAPREDUCE-6302](#): Backport preempt reducers after a configurable timeout irrespective of headroom.
- [MAPREDUCE-6549](#): Multibyte delimiters with LineRecordReader cause duplicate records.
- [YARN-2194](#): Fix bug causing CGroups functionality to fail on RHEL7.
- [YARN-2571](#): RM to support YARN registry.
- [YARN-3467](#): Expose allocatedMB, allocatedVCores, and runningContainers metrics on running Applications in RM Web UI.
- [YARN-3600](#): AM container link is broken (on a killed application, at least).
- [YARN-3727](#): For better error recovery, check if the directory exists before using it for localization.

- [YARN-3751](#) Fixed ApplInfo to check if used resources are null.
- [YARN-3766](#) Fixed the apps table column error of generic history web UI.
- [YARN-3849](#) Too much of preemption activity causing continuous killing of containers across queues.
- [YARN-4140](#) RM container allocation delayed incase of app submitted to Nodelabel partition.
- [YARN-4233](#) YARN Timeline Service plugin: ATS v1.5.
- [YARN-4285](#) Display resource usage as percentage of queue and cluster in the RM UI.
- [YARN-4287](#) Rack locality improvement.
- [YARN-4288](#) Fixed RMProxy to retry on IOException from local host.
- [YARN-4313](#) Race condition in MiniMRYarnCluster when getting history server address.
- [YARN-4345](#) yarn radmin -updateNodeResource doesn't work.
- [YARN-4347](#) Resource manager fails with Null pointer exception.
- [YARN-4349](#) YARN_APPLICATION call to ATS does not have YARN_APPLICATION_CALLER_CONTEXT.
- [YARN-4384](#) updateNodeResource CLI should not accept negative values for resource.
- [YARN-4405](#) Support node label store in non-appendable file system.

HDP 2.3.2 provided the following Apache patches:

NEW FEATURES

- [HDFS-8155](#) Support OAuth2 in WebHDFS.

IMPROVEMENTS

- [HADOOP-10597](#) RPC Server signals backoff to clients when all request queues are full.
- [HADOOP-11960](#) Enable Azure-Storage Client Side logging.
- [HADOOP-12325](#) RPC Metrics: Add the ability track and log slow RPCs.
- [HADOOP-12358](#) Add -safely flag to rm to prompt when deleting many files.
- [HDFS-4185](#) Add a metric for number of active leases.
- [HDFS-4396](#) Add START_MSG/SHUTDOWN_MSG for ZKFC.
- [HDFS-6860](#) BlockStateChange logs are too noisy.

- [HDFS-7923](#) The DataNodes should rate-limit their full block reports by asking the NN on heartbeat messages.
- [HDFS-8046](#) Allow better control of getContentSummary.
- [HDFS-8180](#) AbstractFileSystem Implementation for WebHdfs.
- [HDFS-8278](#) When computing max-size-to-move in Balancer, count only the storage with remaining \geq default block size.
- [HDFS-8432](#) Introduce a minimum compatible layout version to allow downgrade in more rolling upgrade use cases.
- [HDFS-8435](#) Support CreateFlag in WebHDFS.
- [HDFS-8549](#) Abort the balancer if an upgrade is in progress.
- [HDFS-8797](#) WebHdfsFileSystem creates too many connections for pread.
- [HDFS-8818](#) Changes the global moveExecutor to per datanode executors and changes MAX_SIZE_TO_MOVE to be configurable.
- [HDFS-8824](#) Do not use small blocks for balancing the cluster.
- [HDFS-8826](#) In Balancer, add an option to specify the source node list so that balancer only selects blocks to move from those nodes.
- [HDFS-8883](#) NameNode Metrics: Add FSNameSystem lock Queue Length.
- [HDFS-8911](#) NameNode Metric Add Editlog counters as a JMX metric.
- [HDFS-8983](#) NameNode support for protected directories.
- [HDFS-8983](#) NameNode support for protected directories.
- [YARN-2513](#) Host framework UIs in YARN for use with the ATS.
- [YARN-3197](#) Confusing log generated by CapacityScheduler.
- [YARN-3357](#) Move TestFifoScheduler to FIFO package.
- [YARN-3360](#) Add JMX metrics to TimelineDataManager.
- [YARN-3579](#) CommonNodeLabelsManager should support NodeLabel instead of string label name when getting node-to-label/label-to-label mappings.
- [YARN-3978](#) Configurably turn off the saving of container info in Generic AHS.
- [YARN-4082](#) Container shouldn't be killed when node's label updated.
- [YARN-4101](#) RM should print alert messages if Zookeeper and ResourceManager gets connection issue.

- [YARN-4149](#) yarn logs -am should provide an option to fetch all the log files.

BUG FIXES

- [HADOOP-11802](#) DomainSocketWatcher thread terminates sometimes after there is an I/O error during requestShortCircuitShm.
- [HADOOP-12052](#) IPC client downgrades all exception types to IOE, breaks callers trying to use them.
- [HADOOP-12073](#) Azure FileSystem PageBlobInputStream does not return -1 on EOF.
- [HADOOP-12095](#) org.apache.hadoop.fs.shell.TestCount fails.
- [HADOOP-12304](#) Applications using FileContext fail with the default filesystem configured to be wasb/s3/etc.
- [HADOOP-8151](#) Error handling in snappy decompressor throws invalid exceptions.
- [HDFS-6945](#) BlockManager should remove a block from excessReplicateMap and decrement ExcessBlocks metric when the block is removed.
- [HDFS-7608](#) hdfs dfsclient newConnectedPeer has nowrite timeout.
- [HDFS-7609](#) Avoid retry cache collision when Standby NameNode loading edits.
- [HDFS-8309](#) Skip unit test using DataNodeTestUtils#injectDataDirFailure() on Windows.
- [HDFS-8310](#) Fix TestCLI.testAll "help for find" on Windows.
- [HDFS-8311](#) DataStreamer.transfer() should timeout the socket InputStream.
- [HDFS-8384](#) Allow NN to startup if there are files having a lease but are not under construction.
- [HDFS-8431](#) hdfs crypto class not found in Windows.
- [HDFS-8539](#) Hdfs doesn't have class 'debug' in windows.
- [HDFS-8542](#) WebHDFS getHomeDirectory behavior does not match specification.
- [HDFS-8593](#) Calculation of effective layout version mishandles comparison to current layout version in storage.
- [HDFS-8767](#) RawLocalFileSystem.listStatus() returns null for UNIX pipefile.
- [HDFS-8850](#) VolumeScanner thread exits with exception if there is no block pool to be scanned but there are suspicious blocks.
- [HDFS-8863](#) The remaining space check in BlockPlacementPolicyDefault is flawed.
- [HDFS-8879](#) Quota by storage type usage incorrectly initialized upon namenode restart.
- [HDFS-8885](#) ByteRangeInputStream used in webhdfs does not override available().

- [HDFS-8932](#) NPE thrown in NameNode when try to get TotalSyncCount metric before editLogStream initialization.
- [HDFS-8939](#) Test(S)WebHdfsFileContextMainOperations failing on branch-2.
- [HDFS-8969](#) Clean up findbugs warnings for HDFS-8823 and HDFS-8932.
- [HDFS-8995](#) Flaw in registration bookkeeping can make DN die on reconnect.
- [HDFS-9009](#) Send metrics logs to NullAppender by default.
- [YARN-3413](#) Changed Nodelabel attributes (like exclusivity) to be settable only via addToClusterNodeLabels but not changeable at runtime.
- [YARN-3885](#) ProportionalCapacityPreemptionPolicy doesn't preempt if queue is more than 2 level.
- [YARN-3894](#) RM startup should fail for wrong CS xml NodeLabel capacity configuration.
- [YARN-3896](#) RMNode transitioned from RUNNING to REBOOTED because its response id has not been reset synchronously.
- [YARN-3932](#) SchedulerApplicationAttempt#getResourceUsageReport and UserInfo should be based on total-used-resources.
- [YARN-3971](#) Skip RMNodeLabelsManager#checkRemoveFromClusterNodeLabelsOfQueue on nodelabel recovery.
- [YARN-4087](#) Followup fixes after YARN-2019 regarding RM behavior when state-store error occurs.
- [YARN-4092](#) Fixed UI redirection to print useful messages when both RMs are in standby mode.

OPTIMIZATION

- [HADOOP-11772](#) RPC Invoker relies on static ClientCache which has synchronized(this) blocks.
- [HADOOP-12317](#) Applications fail on NM restart on some linux distro because NM container recovery declares AM container as LOST.
- [HADOOP-7713](#) dfs -count -q should label output column.
- [HDFS-8856](#) Make LeaseManager#countPath O(1).
- [HDFS-8867](#) Enable optimized block reports.

HDP 2.3.0 provided the following Apache patches:

NEW FEATURES

- [HDFS-8008](#) Support client-side back off when the datanodes are congested.

- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-1376](#) NM need to notify the log aggregation status to RM through heartbeat.
- [YARN-1402](#) Update related Web UI and CLI with exposing client API to check log aggregation status.
- [YARN-2498](#) Respect labels in preemption policy of capacity scheduler for inter-queue preemption.
- [YARN-2571](#) RM to support YARN registry.
- [YARN-2619](#) Added NodeManager support for disk io isolation through cgroups.
- [YARN-3225](#) New parameter of CLI for decommissioning node gracefully in RMAdmin CLI.
- [YARN-3318](#) Create Initial OrderingPolicy Framework and FifoOrderingPolicy.
- [YARN-3319](#) Implement a FairOrderingPolicy.
- [YARN-3326](#) Support RESTful API for getLabelsToNodes.
- [YARN-3345](#) Add non-exclusive node label API.
- [YARN-3347](#) Improve YARN log command to get AMContainer logs as well as running containers logs.
- [YARN-3348](#) Add a 'yarn top' tool to help understand cluster usage.
- [YARN-3354](#) Add node label expression in ContainerTokenIdentifier to support RM recovery.
- [YARN-3361](#) CapacityScheduler side changes to support non-exclusive node labels.
- [YARN-3365](#) Enhanced NodeManager to support using the 'tc' tool via container-executor for outbound network traffic control.
- [YARN-3366](#) Enhanced NodeManager to support classifying/shaping outgoing network bandwidth traffic originating from YARN containers.
- [YARN-3410](#) YARN admin should be able to remove individual application records from RMStateStore.
- [YARN-3443](#) Create a 'ResourceHandler' subsystem to ease addition of support for new resource types on the NM.
- [YARN-3448](#) Added a rolling time-to-live LevelDB timeline store implementation.
- [YARN-3463](#) Integrate OrderingPolicy Framework with CapacityScheduler.
- [YARN-3505](#) Node's Log Aggregation Report with SUCCEED should not be cached in RMApps.

- [YARN-3541](#) Add version info on timeline service / generic history web UI and REST API.

IMPROVEMENTS

- [HADOOP-10597](#) RPC Server signals backoff to clients when all request queues are full.
- [YARN-1880](#) Cleanup TestApplicationClientProtocolOnHA.
- [YARN-2495](#) Allow admin specify labels from each NM (Distributed configuration for node label).
- [YARN-2696](#) Queue sorting in CapacityScheduler should consider node label.
- [YARN-2868](#) FairScheduler: Metric for latency to allocate first container for an application.
- [YARN-2901](#) Add errors and warning metrics page to RM, NM web UI.
- [YARN-3243](#) CapacityScheduler should pass headroom from parent to children to make sure ParentQueue obey its capacity limits.
- [YARN-3248](#) Display count of nodes blacklisted by apps in the web UI.
- [YARN-3293](#) Track and display capacity scheduler health metrics in web UI.
- [YARN-3294](#) Allow dumping of Capacity Scheduler debug logs via web UI for a fixed time period.
- [YARN-3356](#) Capacity Scheduler FiCaSchedulerApp should use ResourceUsage to track used-resources-by-label.
- [YARN-3362](#) Add node label usage in RM CapacityScheduler web UI.
- [YARN-3394](#) Enrich WebApplication proxy documentation.
- [YARN-3397](#) yarn radmin should skip -failover.
- [YARN-3404](#) Display queue name on application page.
- [YARN-3406](#) Display count of running containers in the RM's Web UI.
- [YARN-3451](#) Display attempt start time and elapsed time on the web UI.
- [YARN-3494](#) Expose AM resource limit and usage in CS QueueMetrics.
- [YARN-3503](#) Expose disk utilization percentage and bad local and log dir counts in NM metrics.
- [YARN-3511](#) Add errors and warnings page to ATS.
- [YARN-3565](#) NodeHeartbeatRequest/RegisterNodeManagerRequest should use NodeLabel object instead of String.
- [YARN-3581](#) Deprecate -directlyAccessNodeLabelStore in RMAdminCLI.

- [YARN-3583](#) Support of NodeLabel object instead of plain String in YarnClient side.
- [YARN-3593](#) Add label-type and Improve "DEFAULT_PARTITION" in Node Labels Page.
- [YARN-3700](#) Made generic history service load a number of latest applications according to the parameter or the configuration.

BUG FIXES

- [HADOOP-11859](#) PseudoAuthenticationHandler fails with httpcomponents v4.4.
- [HADOOP-7713](#) dfs -count -q should label output column.
- [HDFS-27](#) HDFS CLI with -config set to default config complains log file not found error.
- [HDFS-6666](#) Abort NameNode and DataNode startup if security is enabled but block access token is not enabled.
- [HDFS-7645](#) Fix CHANGES.txt.
- [HDFS-7645](#) Rolling upgrade is restoring blocks from trash multiple times.
- [HDFS-7701](#) Support reporting per storage type quota and usage with hadoop/hdfs shell.
- [HDFS-7890](#) Improve information on Top users for metrics in RollingWindowsManager and lower log level.
- [HDFS-7933](#) fsck should also report decommissioning replicas.
- [HDFS-7990](#) IBR delete ack should not be delayed.
- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [HDFS-8055](#) NullPointerException when topology script is missing.
- [HDFS-8144](#) Split TestLazyPersistFiles into multiple tests.
- [HDFS-8152](#) Refactoring of lazy persist storage cases.
- [HDFS-8205](#) CommandFormat#parse() should not parse option as value of option.
- [HDFS-8211](#) DataNode UUID is always null in the JMX counter.
- [HDFS-8219](#) setStoragePolicy with folder behavior is different after cluster restart.
- [HDFS-8229](#) LAZY_PERSIST file gets deleted after NameNode restart.
- [HDFS-8232](#) Missing datanode counters when using Metrics2 sink interface.
- [HDFS-8276](#) LazyPersistFileScrubber should be disabled if scrubber interval configured zero.

- [YARN-2666](#) TestFairScheduler.testContinuousScheduling fails Intermittently.
- [YARN-2740](#) Fix NodeLabelsManager to properly handle node label modifications when distributed node label configuration enabled.
- [YARN-2821](#) Fixed a problem that DistributedShell AM may hang if restarted.
- [YARN-3110](#) Few issues in ApplicationHistory web ui.
- [YARN-3136](#) Fixed a synchronization problem of AbstractYarnScheduler#getTransferredContainers.
- [YARN-3266](#) RMContext#inactiveNodes should have NodeId as map key.
- [YARN-3269](#) Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.
- [YARN-3305](#) Normalize AM resource request on app submission.
- [YARN-3343](#) Increased TestCapacitySchedulerNodeLabelUpdate#testNodeUpdate timeout.
- [YARN-3383](#) AdminService should use "warn" instead of "info" to log exception when operation fails.
- [YARN-3387](#) Previous AM's container completed status couldn't pass to current AM if AM and RM restarted during the same time.
- [YARN-3425](#) NPE from RMNodeLabelsManager.serviceStop when NodeLabelsManager.serviceInit failed.
- [YARN-3435](#) AM container to be allocated Appattempt AM container shown as null.
- [YARN-3459](#) Fix failiure of TestLog4jWarningErrorMetricsAppender.
- [YARN-3517](#) RM web ui for dumping scheduler logs should be for admins only.
- [YARN-3530](#) ATS throws exception on trying to filter results without otherinfo.
- [YARN-3552](#) RM Web UI shows -1 running containers for completed apps.
- [YARN-3580](#) [JDK8] TestClientRMService.testGetLabelsToNodes fails.
- [YARN-3632](#) Ordering policy should be allowed to reorder an application when demand changes.
- [YARN-3654](#) ContainerLogsPage web UI should not have meta-refresh.
- [YARN-3707](#) RM Web UI queue filter doesn't work.
- [YARN-3740](#) Fixed the typo in the configuration name: APPLICATION_HISTORY_PREFIX_MAX_APPS.

1.6.2. Accumulo

HDP 2.4.0 provides Accumulo 1.7.0 and the following Apache patches:

- [ACCUMULO-4080](#) TabletServers should be less aggressively "monitoring RO filesystems".

HDP 2.3.4 provided Accumulo 1.7.0 and the following Apache patches:

- [ACCUMULO-3963](#) Incremental backoff on inability to write to HDFS.
- [ACCUMULO-4053](#) ReplicationOperations.drain() is returning too quickly.
- [ACCUMULO-4060](#) Transient ZooKeeper connection issues kills FATE Runner threads.
- [ACCUMULO-4065](#) Oneway Thrift calls leave message on client's InputStream.
- [ACCUMULO-4069](#) Services failing to renew Kerberos ticket.
- [ACCUMULO-4071](#) BulkImportSequentialRowsIT fails on standalone cluster with HDFS permission errors.

HDP 2.3.2 provided Accumulo 1.7.0 and the following Apache patches:

- [ACCUMULO-3890](#) Use of CredentialProvider results in a lot of NN ops.
- [ACCUMULO-3957](#) Consider moving off getContentSummary in the monitor.
- [ACCUMULO-3967](#) bulk import loses records when loading pre-split table.
- [ACCUMULO-3973](#) ShellServerIT.addauths fails to correctly deal with cached authorizations.
- [ACCUMULO-4001](#) BulkImportSequentialRowsIT fails when using HDFS.

HDP 2.3.0 provided Accumulo 1.7.0 and the following Apache patches:

- [ACCUMULO-3809](#) Table problem report has bogus table name for user table.
- [ACCUMULO-3810](#) RandomWalk test, MultiTable fails throwing java.lang.NullPointerException w/ Kerberos on.
- [ACCUMULO-3812](#) T*ProxyIT classes need cleanup.
- [ACCUMULO-3814](#) StandaloneAccumuloClusterControl doesn't set provided ACCUMULO_CONF_DIR on SetGoalState.
- [ACCUMULO-3815](#) StandaloneClusterControl shouldn't use canonical paths.
- [ACCUMULO-3816](#) rpc.sasl.qop not mentioned in Kerberos server-configuration user manual section.
- [ACCUMULO-3821](#) CleanTmpIT fails on dfs.permission enabled HDFS instance.

- [ACCUMULO-3822](#) ImportExportIT fails to write to export directory in HDFS due to permissions.
- [ACCUMULO-3823](#) Support separate client and server ACCUMULO_CONF_DIRS for StandaloneCluster ITs.
- [ACCUMULO-3826](#) User manual accidentally references commerical product.
- [ACCUMULO-3827](#) Default store types for monitor SSL are broken.
- [ACCUMULO-3828](#) SimpleProxyBase ITs failing due to constraint propagation.
- [ACCUMULO-3834](#) ConstraintIT occasionally failing.
- [ACCUMULO-3838](#) ReplicationIT.replicationEntriesPrecludeWalDeletion failed because it missed an expected WAL.
- [ACCUMULO-3839](#) Nonsense error when configuring instance.volumes.replacements.
- [ACCUMULO-3845](#) DurabilityIT failed.
- [ACCUMULO-3846](#) Allow override of C++ compiler through Maven build.
- [ACCUMULO-3847](#) StandaloneClusterControl needs to launch MR jobs locally.
- [ACCUMULO-3849](#) Proxy sets incorrect primary for SASL server transport.
- [ACCUMULO-3850](#) Improve logging in replication code path.
- [ACCUMULO-3852](#) NPE in WorkMaker for non-existent table.
- [ACCUMULO-3853](#) Contention around ConcurrentLinkedListQueue.size() in AsyncSpanReceiver.
- [ACCUMULO-3856](#) ProxyServer.updateAndFlush leaks BatchWriter.
- [ACCUMULO-3858](#) WatchTheWatchCountIT failed with too few watchers.
- [ACCUMULO-3859](#) TabletServer never acknowledged constraint.
- [ACCUMULO-3861](#) DurabilityIT might actually see all results with durability=none.
- [ACCUMULO-3862](#) Improve how AsyncSpanReceiver drops short spans.
- [ACCUMULO-3870](#) Loads of warnings from ClientConfiguration delimiter parsing w/ Kerberos.
- [ACCUMULO-3874](#) Wrong username in exception when user doesn't exist.
- [ACCUMULO-3877](#) TableOperationsIT failed in testCompactEmptyTableWithGeneratorIterator_Splits_Cancel.
- [ACCUMULO-3878](#) Hunt down ClientConfiguration warnings.

- [ACCUMULO-3879](#) MultiInstanceReplicationIT.dataWasReplicatedToThePeer failed.
- [ACCUMULO-3880](#) Malformed Configuration Causes tservers To Shutdown.
- [ACCUMULO-3881](#) T*ProxyITs fail with useKrbForIT=true.
- [ACCUMULO-3882](#) AccumuloOutputFormatIT loads installed client.conf instead of minicluster's.
- [ACCUMULO-3883](#) ITs should not load default ClientConfiguration.
- [ACCUMULO-3886](#) Boolean values in SiteConfiguration must use lower-case starting characters.
- [ACCUMULO-3887](#) Lack of insight into `accumulo admin stop \$tserver`.
- [ACCUMULO-3893](#) ReadWriteIT#sunnyDay fails against Monitor w/ SSL enabled.
- [ACCUMULO-3894](#) KerberosProxyIT too aggressive in waiting for proxy to start.

1.6.3. Atlas

HDP 2.4.0 provides Atlas 0.5.0 and the following Apache patches:

- [ATLAS-448](#): Hive IllegalArgumentException with Atlas hook enabled on SHOW TRANSACTIONS AND SHOW COMPACTIONS.

HDP 2.3.4 provided Atlas 0.5.0 and the following Apache patches:

- [ATLAS-102](#): Issue with SolrIndex.
- [ATLAS-114](#): Upgrade hbase client to 1.1.2.
- [ATLAS-117](#): Build fails on the latest commit.
- [ATLAS-118](#): rename atlas log4j configuration files to avoid config collisions with other projects.
- [ATLAS-128](#): DSL - Add support for comparisons on list type.
- [ATLAS-16](#): jersey jaxb exception.
- [ATLAS-168](#): Atlas UI - Max column in hive 4.
- [ATLAS-17](#): Parameterize schema API query per typeName.
- [ATLAS-179](#): Atlas hook causes mem leak and hive server 2 crashes.
- [ATLAS-180](#): Cleanup atlas doc packaging.
- [ATLAS-195](#): Document Hbase configs.

- [ATLAS-196](#): Fix solr documentation.
- [ATLAS-198](#): Atlas UI Requires Internet Access.
- [ATLAS-238](#): atlas_start.py- the Atlas server won't restart after improper shutdown.
- [ATLAS-279](#): UI not displaying results for certain successful "select" search queries.
- [ATLAS-31](#): Fixed ATLAS build fails with clean repo.
- [ATLAS-33](#): Atlas restart fails.
- [ATLAS-334](#): Update documentation to reflect copying required atlas file on solr installation.
- [ATLAS-335](#): Kerberized cluster: Atlas fails to come up with hbase as backend.
- [ATLAS-344](#): Document HBase permissions for secure cluster.
- [ATLAS-344](#): Document HBase permissions for secure cluster.
- [ATLAS-350](#): Document jaas config details for atlas.
- [ATLAS-350](#): Document jaas config details for atlas.
- [ATLAS-351](#): Solr as indexing backend is not picked up by titan when storage backend is set to hbase.
- [ATLAS-352](#): Fix performance issues with type and entity creation with Hbase as storage backend.
- [ATLAS-355](#): Kerberized cluster: client.properties does not have correct values for the properties.
- [ATLAS-36](#): Need separate persisted properties for HTTP and HTTPS ports.
- [ATLAS-361](#): Add validation when index backends are switched in ATLAS configuration.
- [ATLAS-37](#): atlas repository, webapp, hive-bridge tests fails with Hbase and Solr as Titan storage backend.
- [ATLAS-373](#): Renew TGT from KeyTab for ATLAS service principal.
- [ATLAS-45](#): Entity submit fails.
- [ATLAS-54](#): Rename configs in hive hook.
- [ATLAS-67](#): Validate with secure zookeeper server for titan interaction with hbase and solr.
- [ATLAS-81](#): atlas debian packaing fails in maven build.
- [ATLAS-86](#): Jenkins build failing as of build #41.

- [ATLAS-91](#): Add solr configuration and documentation.
- [ATLAS-92](#): import-hive.sh failed to find HiveMetaStoreBridge.

HDP 2.3.2 provided Atlas 0.5.0 and the following Apache patches:

- [ATLAS-15](#) remove specific version string as default property value.
- [ATLAS-19](#) remove unnecessary docs dir.
- [ATLAS-29](#) create configuration that inherits existing hadoop config.
- [ATLAS-31](#) Fixed ATLAS build fails with clean repo.
- [ATLAS-31](#) Fixed Mixed Index creation fails with Date types.
- [ATLAS-32](#) create HTTP connection in context of invoking user in secure cluster.
- [ATLAS-54](#) Rename configs in hive hook.

HDP 2.3.0 provided Atlas 0.5.0, with no additional Apache patches.

1.6.4. Calcite

HDP 2.4.0 provides Calcite 1.2.0 and the following Apache patches:

- [CALCITE-984](#): NPE in ErrorResponse construction.

HDP 2.3.4 provided Calcite 1.2.0 and the following Apache patches:

- [CALCITE-259](#): Using sub-queries in CASE statement against JDBC tables generates invalid Oracle SQL.
- [CALCITE-429](#): Add statistics SPI for lattic optimization algorithm.
- [CALCITE-522](#): In remote JDBC driver, transmit stataic database properties as a map.
- [CALCITE-585](#): Avatica JDBC methods should throw SQLFeatureNotSupportedException.
- [CALCITE-645](#): Pass server-side exceptions back to the client.
- [CALCITE-661](#): Remote fetch in Calcite JDBC driver.
- [CALCITE-671](#): ByteString does not deserialize properly as a FetchRequest parameterValue.
- [CALCITE-677](#): RemoteDriverTest.testTypeHandling fails east of Greenwich.
- [CALCITE-687](#): Make RemoteDriverTest thread-safe.
- [CALCITE-699](#): In Avatica, synchronize access to Calendar.

- [CALCITE-705](#): DML in Avatica, and split Execute out from Fetch request.
- [CALCITE-708](#): Avatica and Calcite to support DatabaseMetaData getTypeInfo.
- [CALCITE-712](#): Avatica statment execute return all resultset instead of MaxRows from setMaxRows.
- [CALCITE-717](#): Compare BINARY and VARBINARY on unsigned byte values.
- [CALCITE-718](#): Enable fetch to work for Statement.execute<code>*</code>.
- [CALCITE-728](#): Test suite hangs on Windows.
- [CALCITE-730](#): ClassCastException in table from CloneSchema.
- [CALCITE-741](#): Dependencies should not be empty.
- [CALCITE-765](#): RPC server returns JSON data with Content-Type set to text/html.
- [CALCITE-767](#): Commit functionality not exposed by the RPC server.
- [CALCITE-780](#): HTTP error 413 when sending a long string to the Avatica server.
- [CALCITE-789](#): MetaImpl.MetaCatalog should expose TABLE_CAT instead of TABLE_CATALOG.
- [CALCITE-795](#): Loss of precision when sending a decimal number via the remote JSON service.
- [CALCITE-813](#): Upgrade updateCount, maxRows from int to long.
- [CALCITE-825](#): Allow user to specify sort order of an ArrayTable.
- [CALCITE-840](#): Protobuf transport for Avatica.
- [CALCITE-843](#): AvaticaConnection.getAutoCommit throws NullPointerException.
- [CALCITE-865](#): NullPointerException in getTables with PostgreSQL.
- [CALCITE-866](#): Create documentation for RPC message format<code>*</code>.
- [CALCITE-871](#): JdbcResultSet returns incomplete Frame with "default" statement ID.
- [CALCITE-903](#): Enable client to recover from missing server-side state.
- [CALCITE-905](#): getTables returns empty result in JdbcMeta.
- [CALCITE-906](#): Avatica JDbcMeta statement IDs are not unique.
- [CALCITE-908](#): Bump protobuf dependency to protobuf-3.0.0-beta-1.
- [CALCITE-910](#): Improve handling of ARRAY, MULTISET, STRUCT types.

- [CALCITE-912](#): Pass URL connection properties to avatica server.
- [CALCITE-913](#): Avatica's transport of array fields fails.
- [CALCITE-914](#): Missing JsonSubType for ExecuteResponse.
- [CALCITE-919](#): ArithmeticException when querying against decimal field.
- [CALCITE-921](#): ClassCastException byte[] to String with protobuf.
- [CALCITE-927](#): ColumnsRequest Service call doesn't "fix" ResultSetResponse.
- [CALCITE-951](#): Print the server-side stack in the local exception.
- [CALCITE-962](#): Server-side exception (stack trace) not propagated in JdbcMeta.propagate.
- [CALCITE-989](#): Provide generic server metadata in responses.

HDP 2.3.2 provided Calcite 1.2.0, with no additional Apache patches.

HDP 2.3.0 provided Calcite 1.2.0, with no additional Apache patches.

1.6.5. Falcon

HDP 2.4.0 provides Falcon 0.6.1 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Falcon 0.6.1 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.2 provided Falcon 0.6.1 and the following Apache patches:

NEW FEATURES

- [FALCON-1039](#) Add instance dependency API in falcon.
- [FALCON-1188](#) Falcon support for Hive Replication.
- [FALCON-1325](#) Falcon UI.
- [FALCON-796](#) Enable users to triage data processing issues through falcon.

IMPROVEMENTS

- [FALCON-1060](#) Handle transaction failures in Lineage.
- [FALCON-1147](#) Allow _ in the names for name value pair.
- [FALCON-1174](#) Ability to disable oozie dryrun while scheduling or updating the falcon entity.

- [FALCON-1186](#) Add filtering capability to result of instance summary.
- [FALCON-1204](#) Expose default configs for feed late data handling in runtime.properties.
- [FALCON-1317](#) Inconsistent JSON serialization.
- [FALCON-1322](#) Add prefix in runtime.properties.
- [FALCON-1324](#) Pagination API breaks backward compatibility.
- [FALCON-1359](#) Improve output format for Feed Instance Listing.
- [FALCON-1361](#) Default end date should be now.
- [FALCON-1368](#) Improve Falcon server restart time.
- [FALCON-1374](#) Remove the cap on numResults.
- [FALCON-1378](#) Falcon documentation lacks information on how to run Falcon on standalone Oozie/Hadoop setup.
- [FALCON-668](#) FeedReplicator improvement to include more DistCP options.
- [FALCON-676](#) Enable metrics for Titan.
- [FALCON-75](#) Falcon CLI for deleting entities should inform user if entity does not exist.

BUG FIXES

- [FALCON-1038](#) Log mover fails for map-reduce action.
- [FALCON-1101](#) Cluster submission in falcon does not create an owned-by edge.
- [FALCON-1104](#) Exception while adding process instance to graphdb when feed has partition expression.
- [FALCON-1121](#) Backend support for free-text entity search.
- [FALCON-1129](#) In a secure cluster, feed replication fails because of Authentication issues.
- [FALCON-1141](#) Reverse Lookup for feed in prism fails with BadRequest.
- [FALCON-1143](#) Correcting order of entities on reload.
- [FALCON-1144](#) Dynamic partitions not getting registered in Hcat.
- [FALCON-1146](#) feed retention policy deleted everything all the way up to the root.
- [FALCON-1153](#) Instance kill fails intermittently.
- [FALCON-1162](#) Cluster submit succeeds when staging HDFS dir does not have 777.
- [FALCON-1165](#) Falcon restart failed, if defined service in cluster entity is unreachable.

- [FALCON-1244](#) numResults query param in listInstances is ignored when start and end params are not specified.
- [FALCON-1252](#) The parameter "tagkey" should be "tagkeys" in EntityList and FalconCLI twiki.
- [FALCON-1260](#) Instance dependency API produces incorrect results.
- [FALCON-1268](#) Instance Dependency API failure message is not intuitive in distributed mode.
- [FALCON-1282](#) Incorrect hdfs servers property for feed replication in secured environment.
- [FALCON-1310](#) Falcon build fails with Oozie-4.2.0.
- [FALCON-1311](#) Instance dependency API produces inconsistent results in some scenarios.
- [FALCON-1312](#) Falcon post processing action should use Oozie prepared configuration.
- [FALCON-1323](#) Reverse lookup of feeds causes NPE.
- [FALCON-1325](#) Triage API on prism, for an instance at which a process does not exist sends incorrect message.
- [FALCON-1327](#) When using triage on a server for a process which does not exist on that server, a NullPointerException is encountered.
- [FALCON-1328](#) Error in Triage documentation.
- [FALCON-1329](#) Falcon's idempotent behaviour breaks in some cases.
- [FALCON-1344](#) EntityGraph returns null in list of dependent entities.
- [FALCON-1363](#) Fix retry policy example in documentation.
- [FALCON-1398](#) CrossEntityValidations contains incorrect validations.
- [FALCON-1399](#) Property for default number of results is not loaded dynamically.
- [FALCON-1409](#) Update API throws NullPointerException.
- [FALCON-1412](#) Process waits indefinitely and finally timedout even though missing dependencies are met.
- [FALCON-1487](#) In secure cluster setup Hcat process/feed scheduling or replication fails.
- [FALCON-954](#) Secure Kerberos setup: Falcon should periodically revalidate auth token.
- [FALCON-99](#) Adding late data to process doesn't create new coord.

HDP 2.3.0 provided Falcon 0.6.1 and the following Apache patches:

NEW FEATURES

- [FALCON-1039](#) Add instance dependency API in falcon.
- [FALCON-1188](#) Falcon support for Hive Replication.
- [FALCON-790](#) Falcon UI to enable entity/process/feed edits and management.
- [FALCON-796](#) Enable users to triage data processing issues through falcon.

IMPROVEMENTS

- [FALCON-1060](#) Handle transaction failures in Lineage.
- [FALCON-1147](#) Allow _ in the names for name value pair.
- [FALCON-1174](#) Ability to disable oozie dryrun while scheduling or updating the falcon entity.
- [FALCON-1186](#) Add filtering capability to result of instance summary.
- [FALCON-1204](#) Expose default configs for feed late data handling in runtime.properties.
- [FALCON-1317](#) Inconsistent JSON serialization.
- [FALCON-1322](#) Add prefix in runtime.properties.
- [FALCON-1324](#) Pagination API breaks backward compatibility.
- [FALCON-1359](#) Improve output format for Feed Instance Listing.
- [FALCON-1361](#) Default end date should be now.
- [FALCON-1368](#) Improve Falcon server restart time.
- [FALCON-1374](#) Remove the cap on numResults.
- [FALCON-1378](#) Falcon documentation lacks information on how to run Falcon on standalone Oozie/Hadoop setup.
- [FALCON-668](#) FeedReplicator improvement to include more DistCP options.
- [FALCON-676](#) Enable metrics for Titan.
- [FALCON-75](#) Falcon CLI for deleting entities should inform user if entity does not exist.

BUG FIXES

- [FALCON-1101](#) Cluster submission in falcon does not create an owned-by edge.
- [FALCON-1104](#) Exception while adding process instance to graphdb when feed has partition expression.

- [FALCON-1121](#) Backend support for free-text entity search.
- [FALCON-1129](#) In a secure cluster, feed replication fails because of Authentication issues.
- [FALCON-1141](#) Reverse Lookup for feed in prism fails with BadRequest.
- [FALCON-1143](#) Correcting order of entities on reload.
- [FALCON-1144](#) Dynamic partitions not getting registered in Hcat.
- [FALCON-1146](#) feed retention policy deleted everything all the way up to the root.
- [FALCON-1153](#) Instance kill fails intermittently.
- [FALCON-1162](#) Cluster submit succeeds when staging HDFS dir does not have 777.
- [FALCON-1165](#) Falcon restart failed, if defined service in cluster entity is unreachable.
- [FALCON-1244](#) numResults query param in listInstances is ignored when start and end params are not specified.
- [FALCON-1252](#) The parameter "tagkey" should be "tagkeys" in EntityList and FalconCLI twiki.
- [FALCON-1260](#) Instance dependency API produces incorrect results.
- [FALCON-1268](#) Instance Dependency API failure message is not intuitive in distributed mode.
- [FALCON-1282](#) Incorrect hdfs servers property for feed replication in secured environment.
- [FALCON-1310](#) Falcon build fails with Oozie-4.2.0.
- [FALCON-1311](#) Instance dependency API produces inconsistent results in some scenarios.
- [FALCON-1312](#) Falcon post processing action should use Oozie prepared configuration.
- [FALCON-1323](#) Reverse lookup of feeds causes NPE.
- [FALCON-1325](#) Triage API on prism, for an instance at which a process does not exist sends incorrect message.
- [FALCON-1327](#) When using triage on a server for a process which does not exist on that server, a NullPointerException is encountered.
- [FALCON-1328](#) Error in Triage documentation.
- [FALCON-1329](#) Falcon's idempotent behaviour breaks in some cases.
- [FALCON-1344](#) EntityGraph returns null in list of dependent entities.
- [FALCON-1363](#) Fix retry policy example in documentation.

- [FALCON-1398](#) CrossEntityValidations contains incorrect validations.
- [FALCON-1399](#) Property for default number of results is not loaded dynamically.
- [FALCON-1409](#) Update API throws NullPointerException.
- [FALCON-1412](#) Process waits indefinitely and finally timedout even though missing dependencies are met.
- [FALCON-1487](#) In secure cluster setup Hcat process/feed scheduling or replication fails.
- [FALCON-954](#) Secure Kerberos setup : Falcon should periodically revalidate auth token.
- [FALCON-99](#) Adding late data to process doesn't create new coord
- [FALCON-1038](#) Log mover fails for map-reduce action.

HDP 2.3.0 provided Falcon 0.6.1 with no additional Apache patches.

1.6.6. Flume

HDP 2.4.0 provides Flume 1.5.2 and the following Apache patches:

- [FLUME-2865](#) Upgrade thrift version (0.9.2).

HDP 2.3.4 provided Flume 1.5.2 and the following Apache patches:

- [FLUME-2841](#) Upgrade commons-collections to 3.2.2.
- [FLUME-2854](#) Parameterizing jetty version.

HDP 2.3.2 provided Flume 1.5.2, with no additional Apache patches.

HDP 2.3.0 provided Flume 1.5.2 and the following Apache patches:

NEW FEATURES

- [FLUME-1734](#) Hive Sink based on the new Hive Streaming support.
- [FLUME-2442](#) Need an alternative to providing clear text passwords in flume config.

Kafka Sink (preview)

- [FLUME-2251](#) Add support for Kafka Sink.
- [FLUME-2454](#) Support batchSize to allow multiple events per transaction to the Kafka Sink.
- [FLUME-2455](#) Documentation update for Kafka Sink.
- [FLUME-2470](#) Kafka Sink and Source must use camel case for all configs.
- [FLUME-2499](#) Include Kafka Message Key in Event Header, Updated Comments.

Kafka Source

- [FLUME-2250](#) Add support for Kafka Source.

IMPROVEMENTS

- [FLUME-2095](#) JMS source with TIBCO (patch-1).
- [FLUME-2226](#) Refactor BlobHandler out of morphline sink and into HTTP source.
- [FLUME-2227](#) Move BlobDeserializer from Morphline Sink to flume-ng-core.
- [FLUME-2337](#) export JAVA_HOME in flume-env.sh.template and increase heap size.
- [FLUME-2450](#) Improve replay index insertion speed.
- [FLUME-2511](#) Allow configuration of enabled protocols in Avro source and Rpc client.
- [FLUME-2586](#) HDFS Sink should have an option to try rename even if close fails.
- [FLUME-2595](#) Add option to checkpoint on file channel shutdown.
- [FLUME-2624](#) Streaming ingest performance improvement.
- [FLUME-2662](#) Upgrade to Commons-IO 2.4.
- [FLUME-2663](#) Address Build warnings of duplicate dependencies listed.
- [FLUME-2665](#) Update documentation for hdfs.closeTries based on [FLUME-2586](#).

BUG FIXES

- [FLUME-2122](#) Minor cleanups of User guide.
- [FLUME-2123](#) Morphline Solr sink missing short type name.
- [FLUME-2162](#) TestHDFSEventSinkOnMiniCluster.maxUnderReplicationTest fails on hadoop2.
- [FLUME-2175](#) Update Developer Guide with notes on how to upgrade Protocol Buffer version.
- [FLUME-2358](#) File Channel needs to close BackingStore and EventQueue before deleting files in checkpoint directory.
- [FLUME-2402](#) Warning seen when overflow is disabled for Spillable Channel.
- [FLUME-2407](#) Spillable Channel sometimes fails on reconfigure.
- [FLUME-2412](#) Improve Logging in Spillable Channel.
- [FLUME-2441](#) Unit test TestHTTPSource.java failed with IBM JDK 1.7.

- [FLUME-2451](#) HDFS Sink Cannot Reconnect After NameNode Restart.
- [FLUME-2501](#) Updating HttpClient lib version to ensure compat with Solr.
- [FLUME-2520](#) HTTP Source should be able to block a prefixed set of protocols.
- [FLUME-2530](#) Resource leaks found by Coverity tool.
- [FLUME-2533](#) HTTPS tests fail on Java 6.
- [FLUME-2541](#) Bug in TestBucketWriter.testSequenceFileCloseRetries.

1.6.7. HBase

HDP 2.4.0 provides HBase 1.1.2 and the following Apache patches:

- [HBASE-14107](#): Administrative Task: Provide an API to List all procedures.
- [HBASE-14108](#): Administrative Task: provide an API to abort a procedure.
- [HBASE-14432](#): Enforce ACL on procedure admin tasks.
- [HBASE-14468](#): Compaction improvements: FIFO compaction policy.
- [HBASE-14471](#): Thrift - HTTP Error 413 full HEAD if using kerberos authentication.
- [HBASE-14487](#): Shell command to list all procedures.
- [HBASE-14488](#): Shell command to abort a procedure.
- [HBASE-14575](#): Relax region read lock for compactions.
- [HBASE-14761](#): Deletes with and without visibility expression do not delete the matching mutation.
- [HBASE-14866](#): VerifyReplication should use peer configuration in peer connection.
- [HBASE-14928](#): Start row should be set for query through HBase REST gateway involving globbing option.
- [HBASE-14963](#): Remove Guava dependency from HBase client code.
- [HBASE-14987](#): Compaction marker whose region name doesn't match current region's needs to be handled.
- [HBASE-15014](#): Fix filterCellByStore in WALsplitter is awful for performance.
- [HBASE-15022](#): replication_admin.rb throws undefined method 'getZooKeeperClusterKey' for ZKUtil.
- [HBASE-15035](#): bulkloading hfiles with tags that require splits do not preserve tags.

- [HBASE-13103](#): [ergonomics] add region size balancing as a feature of master.
- [HBASE-13250](#): chown of ExportSnapshot does not cover all path and files.
- [HBASE-14207](#): Region was hijacked and remained in transition when RS failed to open a region and later regionplan changed to new RS on retry.
- [HBASE-14268](#): Improve KeyLocker.
- [HBASE-14280](#): Bulk Upload from HA cluster to remote HA hbase cluster fails.
- [HBASE-14309](#): Allow load balancer to operate when there is region in transition by adding force flag.
- [HBASE-14314](#): Metrics for block cache should take region replicas into account.
- [HBASE-14342](#): Recursive call in RegionMergeTransactionImpl.getJournal().
- [HBASE-14359](#): HTable#close will hang forever if unchecked error/exception thrown in AsyncProcess#sendMultiAction.
- [HBASE-14361](#): ReplicationSink should create Connection instances lazily.
- [HBASE-14367](#): Add normalization support to shell.
- [HBASE-14445](#): ExportSnapshot does not honor -chmod option.
- [HBASE-14463](#): Severe performance downgrade when parallel reading a single key from BucketCache.
- [HBASE-14475](#): Region split requests are always audited with hbase user rather than request user.
- [HBASE-14497](#): Reverse Scan threw StackOverflow caused by readPt checking.
- [HBASE-14536](#): Balancer and SSH interfering with each other leading to unavailability.
- [HBASE-14565](#): Make ZK connection timeout configurable in MiniZooKeeperCluster.
- [HBASE-14578](#): URISyntaxException during snapshot restore for table with user defined namespace.
- [HBASE-14581](#): Znode cleanup throws auth exception in secure mode.
- [HBASE-14591](#): Region with reference hfile may split after a forced split in IncreasingToUpperBoundRegionSplitPolicy.
- [HBASE-14605](#): Split fails due to 'No valid credentials' error when SecureBulkLoadEndpoint#start tries to access hdfs.
- [HBASE-14621](#): ReplicationLogCleaner stuck on RS crash.
- [HBASE-14624](#): BucketCache.freeBlock is too expensive.

- [HBASE-14631](#): Region merge request should be audited with request user through proper scope of doAs() calls to region observer notifications.
- [HBASE-14632](#): Region server aborts due to unguarded dereference of Reader.
- [HBASE-14680](#): Two configs for snapshot timeout and better defaults.
- [HBASE-14706](#): RegionLocationFinder should return multiple servernames by top host.
- [HBASE-14759](#): Avoid using Math.abs when selecting SyncRunner in FSHLog.
- [HBASE-14788](#): Splitting a region does not support the hbase.rs.evictblocksonclose config when closing source region.
- [HBASE-14799](#): Commons-collections object deserialization remote command execution vulnerability.
- [HBASE-14809](#): Grant / revoke Namespace admin permission to group.
- [HBASE-14885](#): NullPointerException in HMaster#normalizeRegions() due to missing TableDescriptor.
- [HBASE-14893](#): Race between mutation on region and region closing operation leads to NotServingRegionException.

HDP 2.3.2 provided HBase 1.1.2 and the following Apache patches:

- [HBASE-14258](#) Make region_mover.rb script case insensitive with regard to hostname.
- [HBASE-14258](#) Make region_mover.rb script case insensitive with regard to hostname.
- [HBASE-14269](#) FuzzyRowFilter omits certain rows when multiple fuzzy key exist.
- [HBASE-14302](#) TableSnapshotInputFormat should not create back references when restoring snapshot.
- [HBASE-14313](#) After a Connection sees ConnectionClosingException it never recovers.
- [HBASE-14449](#) Rewrite deadlock prevention for concurrent connection close.
- [HBASE-14474](#) DeadLock in RpcClientImpl.Connection.close().

HDP 2.3.0 provided HBase 1.1.1 and the following Apache patches:

- [HBASE-11658](#) Piped commands to hbase shell should return non-zero if shell command failed.
- [HBASE-11940](#) Add utility scripts for snapshotting / restoring all tables in cluster.

1.6.8. Hive

HDP 2.4.0 provides Hive 1.2.1 and the following Apache patches:

- [HIVE-11141](#): Improve RuleRegExp when the Expression node stack gets huge.
- [HIVE-11291](#): Avoid allocation storm while doing rule matching on operator/expression trees.
- [HIVE-11310](#): Avoid expensive AST tree conversion to String for expressions in WHERE clause.
- [HIVE-11311](#): Avoid dumping AST tree String in Explain unless necessary.
- [HIVE-11328](#): Avoid String representation of expression nodes in ConstantPropagateProcFactory unless necessary.
- [HIVE-11330](#): Add early termination for recursion in StatsRulesProcFactory.evaluateExpression.
- [HIVE-11397](#): Parse Hive OR clauses as they are written into the AST.
- [HIVE-11405](#): Add early termination for recursion in StatsRulesProcFactory.evaluateExpression for OR expression.
- [HIVE-11406](#): Vectorization: StringExpr::compare() == 0 is bad for performance.
- [HIVE-11981](#): ORC Schema Evolution Issues (Vectorized, ACID, and Non-Vectorized).
- [HIVE-12625](#): Backport to branch-1 [HIVE-11981](#) ORC Schema Evolution Issues (Vectorized, ACID, and Non-Vectorized).
- [HIVE-12660](#): HS2 memory leak with .hiverc file use.
- [HIVE-12706](#): Incorrect output from from_utc_timestamp()/to_utc_timestamp when local timezone has DST.
- [HIVE-12728](#): Apply DDL restrictions for ORC schema evolution.
- [HIVE-12766](#): TezTask does not close DagClient after execution.
- [HIVE-12799](#): Always use Schema Evolution for ACID.

HDP 2.3.4 provided Hive 1.2.1 and the following Apache patches:

- [HIVE-10397](#): Fix Split Computation for Acid tables with Delta files.
- [HIVE-10592](#): query fails on ORC pdd on timestamp datatype on stripes with all null on the column.
- [HIVE-10752](#): Revert HIVE-5193.
- [HIVE-10755](#): Rework on HIVE-5193 to enhance the column oriented table access.
- [HIVE-10778](#): LLAP: Utilities::gWorkMap needs to be cleaned in HiveServer2.
- [HIVE-10807](#): Invalidate basic stats for insert queries if autogather=false.

- [HIVE-10980](#): Merge of dynamic partitions loads all data to default partition.
- [HIVE-11008](#): webhcat GET /jobs retries on getting job details from history server is too aggressive.
- [HIVE-11016](#): MiniTez mergejoin test fails with Tez input error (issue in merge join under certain conditions).
- [HIVE-11123](#): Fix how to confirm the RDBMS product name at Metastore.
- [HIVE-11149](#): sometimes HashMap in PerfLogger.java hangs.
- [HIVE-11312](#): ORC format: where clause with CHAR data type not returning any rows.
- [HIVE-11371](#): Null pointer exception for nested table query when using ORC versus text.
- [HIVE-11372](#): join with between predicate comparing integer types returns no rows when ORC format used.
- [HIVE-11398](#): Parse wide OR and wide AND trees to flat OR/AND trees.
- [HIVE-11422](#): Join a ACID table with non-ACID table fail with MR.
- [HIVE-11428](#): Performance: Struct IN() clauses are extremely slow (~ 10x slower).
- [HIVE-11432](#): Hive macro give same result for different arguments.
- [HIVE-11448](#): Support vectorization of Multi-OR and Multi-AND.
- [HIVE-11461](#): Transform flat AND/OR into IN struct clause.
- [HIVE-11462](#): GenericUDFStruct should constant fold at compile time.
- [HIVE-11468](#): Vectorize: Struct IN() clauses.
- [HIVE-11497](#): Make sure `-orcfiledump` utility includes `OrcRecordUpdate.AcidStats`.
- [HIVE-11499](#): Hiveserver2 failing with OOM PermGen when using temporary functions due to DataNucleus caching classloaders.
- [HIVE-11510](#): Metatool `updateLocation` warning on views.
- [HIVE-11517](#): Vectorized `auto_smb_mapjoin_14.q` produces different results.
- [HIVE-11523](#): `org.apache.hadoop.hive ql.io.orc.FileDump` should handle errors.
- [HIVE-11540](#): Too many delta files during Compaction OOM.
- [HIVE-11546](#): Projected columns read size should be scaled to split size for ORC Splits.
- [HIVE-11573](#): `PointLookupOptimizer` can be pessimistic at a low nDV.
- [HIVE-11583](#): When PTF is used over a large partitions result could be corrupted.

- [HIVE-11613](#): schematool should return non zero exit status for info command, if state is inconsistent.
- [HIVE-11634](#): Gen Plan Changes to support multi col in clause.
- [HIVE-11668](#): make sure directsql calls pre-query init when needed.
- [HIVE-11669](#): OrcFileDump service should support directories.
- [HIVE-11698](#): Add additional test for PointLookupOptimizer.
- [HIVE-11712](#): Duplicate groupby keys cause ClassCastException.
- [HIVE-11720](#): Allow customers set a custom request/response header size for hiveserver2 using http.
- [HIVE-11723](#): Incorrect string literal escaping.
- [HIVE-11724](#): WebHcat get jobs to order jobs on time order with latest at top.
- [HIVE-11737](#): IndexOutOfBounds compiling query with duplicated groupby keys.
- [HIVE-11745](#): Alter table Exchange partition with multiple partition_spec is not working.
- [HIVE-11748](#): HivePreparedStatement's setTimestamp() does not quote value as required.
- [HIVE-11807](#): Adjust compression buffer size to avoid creation of too many small stripes.
- [HIVE-11825](#): get_json_object(col,'\$.a') is null in where clause didnt work.
- [HIVE-11831](#): TXN tables in Oracle should be created with ROWDEPENDENCIES.
- [HIVE-11835](#): Type decimal(1,1) reads 0.0, 0.00, etc from text file as NULL.
- [HIVE-11892](#): UDTF run in local fetch task does not return rows forwarded during GenericUDTF.close().
- [HIVE-11901](#): StorageBasedAuthorizationProvider requires write permission on table for SELECT statements.
- [HIVE-11902](#): Abort txn cleanup thread throws SyntaxErrorException.
- [HIVE-11914](#): When transactions gets a heartbeat, it doesn't update the lock heartbeat.
- [HIVE-11915](#): BoneCP returns closed connections from the pool.
- [HIVE-11916](#): TxnHandler.getOpenTxnsInfo() and getOpenTxns() may produce inconsistent result.
- [HIVE-11919](#): Hive Union Type Mismatch.
- [HIVE-11920](#): ADD JAR failing with URL schemes other than file/ivy/hdfs.
- [HIVE-11934](#): Transaction lock retry logic results in infinite loop.

- [HIVE-11939](#): TxnDbUtil should turn off jdbc auto commit.
- [HIVE-11940](#): 'INSERT OVERWRITE' query is very slow because it creates one distcp per file to copy data from staging directory to target directory.
- [HIVE-11948](#): Investigate TxnHandler and CompactionTxnHandler to see where we improve concurrency.
- [HIVE-11950](#): WebHCat status file doesn't show UTF8 character.
- [HIVE-11960](#): braces in join conditions are not supported.
- [HIVE-11964](#): RelOptHiveTable.hiveColStatsMap might contain mismatched column stats.
- [HIVE-11975](#): mssql scripts contains invalid 'GO' statement.
- [HIVE-11977](#): Hive should handle an external avro table with zero length files present.
- [HIVE-11983](#): Hive streaming API uses incorrect logic to assign buckets to incoming records.
- [HIVE-11983](#): Hive streaming API : uses incorrect logic to assign buckets to incoming records.
- [HIVE-11988](#): SUMMARY-[security issue with hive & ranger for import table command].
- [HIVE-11990](#): Humboldt: loading data inpath from a temporary table dir fails.
- [HIVE-11990](#): loading data inpath from a temporary table dir fails.
- [HIVE-11995](#): Remove repetitively setting permissions in insert/load overwrite partition.
- [HIVE-11997](#): Add ability to send Compaction Jobs to specific queue.
- [HIVE-11998](#): Improve Compaction process logging.
- [HIVE-12003](#): HIVE-12276: Fix messages in InvalidTable.
- [HIVE-12003](#): Hive Streaming API Add check to ensure table is transactional.
- [HIVE-12011](#): unable to create temporary table using CTAS if regular table with that name already exists.
- [HIVE-12012](#): select query on json table with map containing numeric values fails.
- [HIVE-12021](#): wrong results: HivePreFilteringRule may introduce wrong common operands.
- [HIVE-12057](#): ORC sarg is logged too much.
- [HIVE-12076](#): WebHCat listing jobs after the given JobId even when templeton.jobs.listorder is set to lexicographicaldesc.

- [HIVE-12083](#): HIVE-10965 introduces thrift error if partNames or colNames are empty.
- [HIVE-12084](#): Hive queries with ORDER BY and large LIMIT fails with OutOfMemoryError Java heap space.
- [HIVE-12156](#): expanding view doesn't quote reserved keyword.
- [HIVE-12179](#): Add option to not add spark-assembly.jar to Hive classpath.
- [HIVE-12196](#): NPE when converting bad timestamp value.
- [HIVE-12201](#): Tez settings need to be shown in set -v output when execution engine is tez.
- [HIVE-12204](#): Tez queries stopped running with ApplicationNotRunningException.
- [HIVE-12206](#): using UDF in same session after other queries fails with ClassNotFound error.
- [HIVE-12223](#): Filter on Grouping__ID does not work properly.
- [HIVE-12230](#): custom UDF configure() not called in Vectorization mode.
- [HIVE-12232](#): BucketingSortingReduceSinkOptimizer throws IOB exception for duplicate columns.
- [HIVE-12235](#): a way for admin and jdbc/odbc client to see which hiveserver2 connected in ZooKeeper discovery service.
- [HIVE-12236](#): Enable SimpleFetchOptimizer for more query types.
- [HIVE-12249](#): Improve logging with tez.
- [HIVE-12250](#): query on hive table on hbasestoragehandler leak zookeeper connection.
- [HIVE-12252](#): Streaming API HiveEndPoint can be created w/o partitionVals for partitioned table.
- [HIVE-12254](#): Improve logging with yarn/hdfs.
- [HIVE-12257](#): Enhance ORC FileDump utility to handle flush_length files.
- [HIVE-12261](#): schematool version info exit status should depend on compatibility, not equality.
- [HIVE-12262](#): Session log dir cannot be created in some cases.
- [HIVE-12266](#): When client exists abnormally, it doesn't release ACID locks.
- [HIVE-12273](#): Improve user level explain.
- [HIVE-12276](#): Fix messages in InvalidTable.
- [HIVE-12277](#): Hive macro results on macro_duplicate.q different after adding ORDER BY.

- [HIVE-12280](#): HiveConnection does not try other HS2 after failure for service discovery.
- [HIVE-12282](#): beeline update command printing in verbose mode.
- [HIVE-12295](#): change some logs from info to debug.
- [HIVE-12307](#): Streaming API TransactionBatch.close() must abort any remaining transactions in the batch.
- [HIVE-12312](#): Excessive logging in PPD code.
- [HIVE-12318](#): qtest failing due to NPE in logStats.
- [HIVE-12327](#): WebHCat e2e tests TestJob_1 and TestJob_2 fail.
- [HIVE-12344](#): Wrong types inferred for SemiJoin generation in CBO.
- [HIVE-12345](#): Follow up for HIVE-9013 Hidden conf vars still visible through beeline.
- [HIVE-12357](#): Allow user to set tez job name.
- [HIVE-12364](#): insert into directory query fail, using distcp when data size>hive.exec.copyfile.maxsize (default 32MB).
- [HIVE-12384](#): Union Operator may produce incorrect result on TEZ.
- [HIVE-12387](#): Bug with logging improvements in ATS.
- [HIVE-12387](#): Issues in Hive's use of CallerContext.
- [HIVE-12387](#): When using Hive on Tez or MR, no caller context shows up in YARN audit logs.
- [HIVE-12389](#): CompactionTxnHandler.cleanEmptyAbortedTxns() should safeguard against huge IN clauses.
- [HIVE-12396](#): BucketingSortingReduceSinkOptimizer may still throw IOB exception for duplicate columns.
- [HIVE-12399](#): Filter out NULLs in the Native Vector MapJoin operators.
- [HIVE-12418](#): HiveHBaseTableInputFormat.getRecordReader() causes Zookeeper connection leak.
- [HIVE-12437](#): SMB join in tez fails when one of the tables is empty.
- [HIVE-12444](#): Global Limit optimization on ACID table without base directory may throw exception.
- [HIVE-12450](#): OrcFileMergeOperator does not use correct compression buffer size.
- [HIVE-12465](#): Hive might produce wrong results when (outer) joins are merged.

- [HIVE-12469](#): Bump Commons-Collections dependency from 3.2.1 to 3.2.2. to address vulnerability.
- [HIVE-12476](#): Oracle directSQL NPE error when fetching empty to null for serdeproperties.
- [HIVE-12498](#): ACID: Setting OrcRecordUpdater.OrcOptions.tableProperties() has no effect.
- [HIVE-12500](#): JDBC driver not overlaying params supplied via properties object when reading params from ZK.
- [HIVE-12522](#): Wrong FS error during Tez merge files when warehouse and scratchdir are on different FS.
- [HIVE-12523](#): ATS HIVE_QUERY_ID api returns different dagName than TEZ_DAG_ID api.
- [HIVE-12529](#): HiveTxnManager.acquireLocks() should not block forever.
- [HIVE-12529](#): HiveTxnManager.acquireLocks() should not block forever.
- [HIVE-12556](#): Have an option to kill DAG when user cancels query in HiveServer2.
- [HIVE-12563](#): NullPointerException with 3-way Tez merge join.
- [HIVE-12565](#): VectorUDAFCount.aggregateInputSelection does not handle isRepeated case.
- [HIVE-12567](#): Enhance TxnHandler retry logic to handle ORA-08176.
- [HIVE-12578](#): Hive query failing with error ClassCastException org.apache.hadoop.hive.ql.plan.ExprNodeConstantDesc cannot be cast to org.apache.hadoop.hive.ql.plan.ExprNodeColumnDesc.
- [HIVE-12583](#): HS2 ShutdownHookManager holds extra of Driver instance.
- [HIVE-12584](#): join on char cols with different length returns empty result with vectorization and tez on.
- [HIVE-12585](#): fix TxnHandler connection leak.
- [HIVE-5623](#): ORC accessing array column that's empty will fail with java out of bound exception.
- [HIVE-7723](#): Explain plan for complex query with lots of partitions is slow due to inefficient collection used to find a matching ReadEntity.
- [HIVE-9013](#): beeline (hiveserver2 client) exposes sensitive metastore DB connection info (connection, password).
- [HIVE-9695](#): Redundant filter operator in reducer Vertex when CBO is disabled.

HDP 2.3.2 provided Hive 1.2.1 and the following Apache patches:

IMPROVEMENTS

- [HIVE-11037](#) HiveOnTez: make explain user level = true as default.

BUG FIXES

- [HIVE-10140](#): Window boundary is not compared correctly.
- [HIVE-10453](#): Reverted.
- [HIVE-10569](#): Hive CLI gets stuck when hive.exec.parallel=true; and some exception happens during SessionState.start.
- [HIVE-10571](#): HiveMetaStoreClient should close existing thrift connection before its reconnect.
- [HIVE-10620](#): ZooKeeperHiveLock overrides equal() method but not hashCode().
- [HIVE-10646](#): ColumnValue does not handle NULL_TYPE.
- [HIVE-10651](#): ORC file footer cache should be bounded.
- [HIVE-10698](#): query on view results fails with table not found error if view is created with subquery alias (CTE).
- [HIVE-10714](#): Bloom filter column names specification should be case insensitive.
- [HIVE-10722](#): external table creation with msck in Hive can create unusable partition.
- [HIVE-10726](#): Hive JDBC setQueryTimeout should not throw exception to make it work with JMeter.
- [HIVE-10731](#): NullPointerException in HiveParser.g.
- [HIVE-10732](#): Hive JDBC driver does not close operation for metadata queries.
- [HIVE-10771](#): "separatorChar" has no effect in "CREATE TABLE AS SELECT" statement.
- [HIVE-10781](#): HadoopJobExecHelper Leaks RunningJobs.
- [HIVE-10790](#): orc write on viewFS throws exception.
- [HIVE-10793](#): HIVE-11587 Hybrid Hybrid Grace Hash Join: Don't allocate all hash table memory upfront.
- [HIVE-10802](#): Table join query with some constant field in select fails.
- [HIVE-10808](#): Inner join on Null throwing Cast Exception.
- [HIVE-10835](#): Concurrency issues in JDBC driver.
- [HIVE-10880](#): The bucket number is not respected in insert overwrite.
- [HIVE-10925](#): Non-static threadlocals in metastore code can potentially cause memory leak.

- [HIVE-10963](#): Hive throws NPE rather than meaningful error message when window is missing.
- [HIVE-10972](#): DummyTxnManager always locks the current database in shared mode, which is incorrect.
- [HIVE-11013](#): MiniTez tez_join_hash test on the branch fails with NPE (initializeOp not called?).
- [HIVE-11024](#): Error inserting a date value via parameter marker (PreparedStatement.setDate).
- [HIVE-11029](#): hadoop.proxyuser.mapr.groups does not work to restrict the groups that can be impersonated.
- [HIVE-11054](#): Read error: Partition Varchar column cannot be cast to string.
- [HIVE-11079](#): Fix qfile tests that fail on Windows due to CR/character escape differences.
- [HIVE-11087](#): DbTxnManager exceptions should include txnid.
- [HIVE-11090](#): ordering issues with windows unit test runs.
- [HIVE-11095](#): SerDeUtils another bug ,when Text is reused.
- [HIVE-11102](#): ReaderImpl: getColumnIndicesFromNames does not work for some cases.
- [HIVE-11112](#): ISO-8859-1 text output has fragments of previous longer rows appended.
- [HIVE-11135](#): Fix the Beeline set and save command in order to avoid the NullPointerException.
- [HIVE-11151](#): Calcite transitive predicate inference rule should not transitively add not null filter on non-nullable input.
- [HIVE-11152](#): Swapping join inputs in ASTConverter.
- [HIVE-11157](#): Hive.get(HiveConf) returns same Hive object to different user sessions.
- [HIVE-11171](#): Join reordering algorithm might introduce projects between joins.
- [HIVE-11172](#): Vectorization wrong results for aggregate query with where clause without group by.
- [HIVE-11174](#): Hive does not treat floating point signed zeros as equal (-0.0 should equal 0.0 according to IEEE floating point spec).
- [HIVE-11176](#): Caused by: java.lang.ClassCastException: org.apache.hadoop.hive.serde2.lazybinary.LazyBinaryStruct cannot be cast to [Ljava.lang.Object;.
- [HIVE-11193](#): ConstantPropagateProcCtx should use a Set instead of a List to hold operators to be deleted.

- [HIVE-11198](#): Fix load data query file format check for partitioned tables.
- [HIVE-11203](#): Beeline force option doesn't force execution when errors occurred in a script.
- [HIVE-11211](#): Reset the fields in JoinStatsRule in StatsRulesProcFactory.
- [HIVE-11216](#): UDF GenericUDFMapKeys throws NPE when a null map value is passed in.
- [HIVE-11221](#): In Tez mode, alter table concatenate orc files can intermittently fail with NPE.
- [HIVE-11255](#): get_table_objects_by_name() in HiveMetaStore.java needs to retrieve table objects in multiple batches.
- [HIVE-11258](#): The function drop_database_core() of HiveMetaStore.java may not drop all the tables.
- [HIVE-11271](#): java.lang.IndexOutOfBoundsException when union all with if function.
- [HIVE-11301](#): thrift metastore issue when getting stats results in disconnect.
- [HIVE-11303](#): Getting Tez LimitExceededException after dag execution on large query.
- [HIVE-11317](#): ACID: Improve transaction Abort logic due to timeout.
- [HIVE-11320](#): ACID enable predicate pushdown for insert-only delta file.
- [HIVE-11344](#): HIVE-9845 makes HCatSplit.write modify the split so that PartInfo objects are unusable after it.
- [HIVE-11356](#): SMB join on tez fails when one of the tables is empty.
- [HIVE-11357](#): ACID enable predicate pushdown for insert-only delta file 2.
- [HIVE-11375](#): Broken processing of queries containing NOT (x IS NOT NULL and x 0).
- [HIVE-11407](#): JDBC DatabaseMetaData.getTables with large no of tables call leads to HS2 OOM.
- [HIVE-11429](#): Increase default JDBC result set fetch size (# rows it fetches in one RPC call) to 1000 from 50.
- [HIVE-11433](#): NPE for a multiple inner join query.
- [HIVE-11442](#): Remove commons-configuration.jar from Hive distribution.
- [HIVE-11449](#): HIVE-11587 "Capacity must be a power of two" error when HybridHashTableContainer memory threshold is too low.
- [HIVE-11456](#): HCatStorer should honor mapreduce.output.basename.
- [HIVE-11467](#): HIVE-11587 WriteBuffers rounding wbSize to next power of 2 may cause OOM.

- [HIVE-11493](#): Predicate with integer column equals double evaluates to false.
- [HIVE-11502](#): Map side aggregation is extremely slow.
- [HIVE-11581](#): HiveServer2 should store connection params in ZK when using dynamic service discovery for simpler client connection string.
- [HIVE-11587](#): Fix memory estimates for mapjoin hashtable.
- [HIVE-11592](#): ORC metadata section can sometimes exceed protobuf message size limit.
- [HIVE-11600](#): Hive Parser to Support multi col in clause (x,y..) in ((..),..., ()).
- [HIVE-11605](#): Incorrect results with bucket map join in tez.
- [HIVE-11606](#): Bucket map joins fail at hash table construction time.
- [HIVE-11607](#): Export tables broken for data > 32 MB.
- [HIVE-11658](#): Load data file format validation does not work with directories.
- [HIVE-11727](#): (BUG-44285). Hive on Tez through Oozie: Some queries fail with fnf exception.
- [HIVE-11755](#): Incorrect method called with Kerberos enabled in AccumuloStorageHandler.
- [HIVE-11820](#): export tables with size of >32MB throws "java.lang.IllegalArgumentException: Skip CRC is valid only with update options".
- [HIVE-11836](#): ORC SARG creation throws NPE for null constants with void type.
- [HIVE-11839](#): Vectorization wrong results with filter of (CAST AS CHAR).
- [HIVE-11849](#): NPE in HiveHBaseTableShapshotInputFormat in query with just count(*).
- [HIVE-11852](#): numRows and rawDataSize table properties are not replicated.
- [HIVE-11875](#): JDBC Driver does not honor delegation token mechanism when readings params from ZooKeeper.
- [HIVE-11897](#): JDO rollback can throw pointless exceptions.
- [HIVE-11928](#): ORC footer section can also exceed protobuf message limit.
- [HIVE-11936](#): Support SQLAnywhere as a backing DB for the hive metastore.
- [HIVE-5277](#): HBase handler skips rows with null valued first cells when only row key is selected.
- [HIVE-6727](#): Table level stats for external tables are set incorrectly.
- [HIVE-7476](#): CTAS does not work properly for s3.

- [HIVE-8529](#): HiveSessionImpl#fetchResults should not try to fetch operation log when hive.server2.logging.operation.enabled is false.
- [HIVE-9566](#): HiveServer2 fails to start with NullPointerException.
- [HIVE-9625](#): Delegation tokens for HMS are not renewed.
- [HIVE-9811](#): Hive on Tez leaks WorkMap objects.
- [HIVE-9974](#): Sensitive data redaction: data appears in name of mapreduce job

HDP 2.3.0 provided Hive 1.2.1 and the following Apache patches:

INCOMPATIBLE CHANGES

- [HIVE-11118](#) Load data query should validate file formats with destination tables.

NEW FEATURES

- [HIVE-10233](#) Hive on Tez: memory manager for grace hash join.

IMPROVEMENTS

- [HIVE-11164](#) WebHCat should log contents of HiveConf on startup [HIVE-11037](#)
HiveOnTez: make explain user level = true as default.

BUG FIXES

- [HIVE-10251](#) [HIVE-9664](#) makes hive depend on ivysettings.xml (using [HIVE-10251](#).simple.patch).
- [HIVE-10996](#) Aggregation / Projection over Multi-Join Inner Query producing incorrect results.
- [HIVE-11028](#) Tez: table self join and join with another table fails with IndexOutOfBoundsException.
- [HIVE-11048](#) Make test cbo_windowing robust.
- [HIVE-11050](#) testCliDriver_vector_outer_join.* failures in Unit tests due to unstable data creation queries.
- [HIVE-11051](#) Hive 1.2.0 MapJoin w/Tez - LazyBinaryArray cannot be cast to [Ljava.lang.Object;.
- [HIVE-11059](#) hcatalog-server-extensions tests scope should depend on hive-exec.
- [HIVE-11060](#) Make test windowing.q robust.
- [HIVE-11066](#) Ensure tests don't share directories on FS.
- [HIVE-11074](#) Update tests for [HIVE-9302](#) after removing binaries.

- [HIVE-11076](#) Explicitly set `hive.cbo.enable=true` for some tests.
- [HIVE-11083](#) Make test `cbo_windowing` robust.
- [HIVE-11104](#) Select operator doesn't propagate constants appearing in expressions.
- [HIVE-11147](#) MetaTool doesn't update FS root location for partitions with space in name.

1.6.9. Kafka

HDP 2.4.0 provides Kafka 0.9.0 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Kafka 0.9.0 and the following Apache patches:

- [KAFKA-2803](#): Add hard bounce system test for Kafka Connect.
- [KAFKA-2812](#): improve consumer integration tests.
- [KAFKA-2862](#): Fix MirrorMaker's `message.handler.args` description.
- [KAFKA-2872](#): unite sink nodes with parent nodes in `addSink`.
- [KAFKA-2877](#): handle request timeout in sync group.
- [KAFKA-2878](#): Guard against `OutOfMemory` in Kafka broker.
- [KAFKA-2879](#): Make `MiniKDC` test service slightly more generic.
- [KAFKA-2880](#): consumer should handle `disconnect/timeout` for metadata requests.
- [KAFKA-2881](#): Improve Consumer Configs and API Documentation.
- [KAFKA-2882](#): Add constructor cache for Snappy and LZ4 Output/Input streams in `Compressor.java`.
- [KAFKA-2892](#): Consumer Docs Use Wrong Method.
- [KAFKA-2899](#): improve logging when unexpected exceptions thrown in reading local log.
- [KAFKA-2906](#): Fix Connect javadocs, restrict only to api subproject, and clean up javadoc warnings.
- [KAFKA-2913](#): missing partition check when removing groups from cache.
- [KAFKA-2942](#): inadvertent auto-commit when pre-fetching can cause message loss.
- [KAFKA-2950](#): Fix performance regression in the producer.

HDP 2.3.2 provided Kafka 0.8.2, with no additional Apache patches.

HDP 2.3.0 provided Kafka 0.8.2 and the following Apache patches:

- [KAFKA-1005](#) Shutdown consumer at the end of consumer performance test.
- [KAFKA-1416](#) Unify sendMessages in TestUtils.
- [KAFKA-1461](#) Replica fetcher thread does not implement any back-off behavior.
- [KAFKA-1499](#) trivial follow-up (remove unnecessary parentheses).
- [KAFKA-1501](#) Let the OS choose the port in unit tests to avoid collisions.
- [KAFKA-1517](#) Messages is a required argument to Producer Performance Test.
- [KAFKA-1546](#) Automate replica lag tuning;.
- [KAFKA-1634](#) Bump up Offset Commit Request to v2 to add global retention and remove per-partition commit timestamp.
- [KAFKA-1664](#) Kafka does not properly parse multiple ZK nodes with non-root chroot.
- [KAFKA-1683](#) add Session concept in SocketServer.PlainTextTransportLayer fixes.
- [KAFKA-1684](#) Kerberos/SASL implementation.
- [KAFKA-1688](#) Adding all public entities for adding a pluggable authorizer to kafka.
- [KAFKA-1755](#) Reject compressed and unkeyed messages sent to compacted topics.
- [KAFKA-1809](#) Refactor brokers to allow listening on multiple ports and IPs.
- [KAFKA-1824](#) ConsoleProducer - properties key.separator and parse.key no longer work.
- [KAFKA-1845](#) KafkaConfig should use ConfigDef.
- [KAFKA-1852](#) Reject offset commits to unknown topics.
- [KAFKA-1863](#) Add docs for possible thrown exception in Callback;.
- [KAFKA-1865](#) Add a flush() method to the producer.
- [KAFKA-1866](#) LogStartOffset gauge throws exceptions after log.delete().
- [KAFKA-1910](#) Refactor new consumer and fixed a bunch of corner cases / unit tests.
- [KAFKA-1926](#) Replace kafka.utils.Utils with o.a.k.common.utils.Utils.
- [KAFKA-1961](#) Prevent deletion of _consumer_offsets topic.
- [KAFKA-1973](#) Remove the accidentally created LogCleanerManager.scala.orig.
- [KAFKA-1982](#) (add missing files) change kafka.examples.Producer to use the new java producer.
- [KAFKA-1982](#) change kafka.examples.Producer to use the new java producer.

- [KAFKA-1986](#) Request failure rate should not include invalid message size and offset out of range.
- [KAFKA-1988](#) Fix `org.apache.kafka.common.utils.Utils.abs` and add `Partitioner.toPositive`.
- [KAFKA-1989](#) New purgatory design; patched by Yasuhiro Matsuda.
- [KAFKA-1990](#) Add unlimited time-based log retention.
- [KAFKA-1992](#) `checkEnoughReplicasReachOffset` doesn't need to get `requiredAcks`.
- [KAFKA-1994](#) Evaluate performance effect of chroot check on Topic creation.
- [KAFKA-1996](#) Fix scaladoc error.
- [KAFKA-1997](#) Refactor `MirrorMaker` based on [KIP-3](#);
- [KAFKA-2002](#) `Mx4JLoader` doesn't disable when `kafka_mx4jenable=false`.
- [KAFKA-2009](#) Fix two minor bugs in mirror maker.
- [KAFKA-2013](#) benchmark test for the purgatory.
- [KAFKA-2016](#) `RollingBounceTest` takes long.
- [KAFKA-2024](#) Log compaction can generate unindexable segments.
- [KAFKA-2033](#) Small typo in documentation.
- [KAFKA-2034](#) `sourceCompatibility` not set in Kafka build.gradle.
- [KAFKA-2039](#) Update Scala to 2.10.5 and 2.11.6.
- [KAFKA-2042](#) Update topic list of the metadata regardless of cluster information;
- [KAFKA-2043](#) `CompressionType` is passed in each `RecordAccumulator` append.
- [KAFKA-2044](#) Support requests and responses from `o.a.k.common` in `KafkaApis`.
- [KAFKA-2047](#) Move the stream creation into concurrent mirror maker threads.
- [KAFKA-2048](#) Change lock synchronized to `inLock()` for `partitionMapCond`.
- [KAFKA-2050](#) Avoid calling `.size()` on linked list.
- [KAFKA-2056](#) Fix transient `testRangePartitionAssignor` failure.
- [KAFKA-2088](#) `kafka-console-consumer.sh` should not create zookeeper path when no brokers found and `chroot` was set in `zookeeper.connect`.
- [KAFKA-2090](#) Remove duplicate check to `metadataFetchInProgress`.
- [KAFKA-2096](#) Enable `keepalive` socket option for broker to prevent socket leak.

- [KAFKA-2099](#) BrokerEndPoint file, methods and object names should match.
- [KAFKA-2104](#) testDuplicateListeners() has a typo.
- [KAFKA-2109](#) Support retries in KafkaLog4jAppender.
- [KAFKA-2112](#) make overflowWheel volatile.
- [KAFKA-2113](#) TestPurgatoryPerformance does not compile using IBM JDK.
- [KAFKA-2114](#) Unable to change min.insync.replicas default.
- [KAFKA-2115](#) Error updating metrics in RequestChannel.
- [KAFKA-2117](#) Use the correct metadata field for reading offset struct.
- [KAFKA-2118](#) Cleaner cannot clean after shutdown during replaceSegments.
- [KAFKA-2119](#) ConsumerRecord key() and value() methods should not have throws Exception.
- [KAFKA-2121](#) Close internal modules upon client shutdown.
- [KAFKA-2122](#) Remove controller.message.queue.size Config.
- [KAFKA-2128](#) kafka.Kafka should return non-zero exit code when caught exception.
- [KAFKA-2131](#) Update new producer javadocs with correct documentation links.
- [KAFKA-2138](#) Fix producer to honor retry backoff.
- [KAFKA-2140](#) Improve code readability.
- [KAFKA-527](#) Compression support does numerous byte copies;

1.6.10. Knox

HDP 2.4.0 provides Knox 0.6.0 and the following Apache patches:

- [KNOX-647](#): Rename LDAP artifacts from test to demo.

HDP 2.3.4 provided Knox 0.6.0 and the following Apache patches:

- [KNOX-566](#): Make the Default Ephemeral DH Key Size 2048 for TLS.
- [KNOX-579](#): Regex based identity assertion provider with static dictionary lookup.
- [KNOX-581](#): Hive dispatch not propagating effective principal name.
- [KNOX-633](#): Upgrade apache commons-collections.

HDP 2.3.2 provided Knox 0.6.0 and the following Apache patches:

- [KNOX-598](#): Concurrent JDBC clients via KNOX to Kerberized HiveServer2 causes HTTP 401 error (due to Kerberos Replay attack error).
- [KNOX-599](#): Template with {**} in queries are expanded with =null for query params without a value.

HDP 2.3.0 provided Knox 0.6.0 and the following Apache patches:

BUG FIXES

- [KNOX-476](#) implementation for X-Forwarded-* headers support and population.
- [KNOX-546](#) Consuming intermediate response during kerberos request dispatching.
- [KNOX-550](#) reverting back to original hive kerberos dispatch behavior.
- [KNOX-559](#) renaming service definition files.

IMPROVEMENTS

- [KNOX-545](#) Simplify Keystore Management for Cluster Scaleout.
- [KNOX-561](#) Allow Knox pid directory to be configured via the Knox-env.sh file.

1.6.11. Mahout

In HDP-2.3.x and 2.4.x, instead of shipping a specific Apache release of Mahout, we synchronized to a particular revision point on Apache Mahout trunk. This revision point is after the 0.9.0 release, but before the 0.10.0 release. This provides a large number of bug fixes and functional enhancements over the 0.9.0 release, but provides a stable release of the Mahout functionality before the complete conversion to new Spark-based Mahout in 0.10.0. In the future, after the Spark-based Mahout functionality has stabilized, HDP plans to ship with it.

The revision point chosen for Mahout in HDP 2.3.x and 2.4.x is from the "mahout-0.10.x" branch of Apache Mahout, as of 19 December 2014, revision 0f037cb03e77c096 in GitHub.

In addition, we have provided the following patches:

- [MAHOUT-1493](#) Port Naive Bayes to Scala DSL.
- [MAHOUT-1589](#) mahout.cmd has duplicated content.

1.6.12. Oozie

HDP 2.4.0 provides Oozie 4.2.0 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Oozie 4.2.0 with the patches specified below. No new additional Apache patches have been included in this release or in HDP 2.3.2.

HDP 2.3.0 provided Oozie 4.2.0 and the following Apache patches:

- [OOZIE-2289](#) hive-jdbc dependency in core/pom.xml should be compile.
- [OOZIE-2290](#) Oozie db version update should happen after all DDL tweak.
- [OOZIE-2291](#) Hive2 workflow.xml.security should have "cred" in action tag instead of "hive2" tag.

1.6.13. Phoenix

HDP 2.4.0 provides Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- [PHOENIX-2608](#): Incompatibility between Jackson1 version shipped with Phoenix, YARN.

HDP 2.3.4 provided Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- [PHOENIX-1277](#): CSVCommonsLoader not allowing null CHAR values (non PK).
- [PHOENIX-1734](#): Local index improvements.
- [PHOENIX-1975](#): Detect and use HBASE_HOME when set.
- [PHOENIX-2014](#): WHERE search condition ignored when also using row value constructor in view.
- [PHOENIX-2027](#): Subqueries with no data are raising IllegalStateException.
- [PHOENIX-2029](#): Queries are making two rpc calls for getTable.
- [PHOENIX-2030](#): CsvBulkLoadTool should use logical name of the table for output directory suffix.
- [PHOENIX-2040](#): Mark spark/scala dependencies as 'provided'.
- [PHOENIX-2059](#): MR index build does not handle table with a schema name correctly.
- [PHOENIX-2087](#): Ensure predictable column position during alter table.
- [PHOENIX-2118](#): Remove/modfiy usages of Guava StopWatch and deprecated ComparisonChain methods.
- [PHOENIX-2125](#): ORDER BY on full PK on salted table does not work.
- [PHOENIX-2139](#): LIKE '%' is not filtering out null columns.
- [PHOENIX-2141](#): ComparisonExpression should return Boolean null if either operand is null.
- [PHOENIX-2149](#): MAX Value of Sequences not honored when closing Connection between calls to NEXT VALUE FOR.
- [PHOENIX-2205](#): Group by a divided value (e.g., time/10) returns NULL.

- [PHOENIX-2207](#): Load scanner caches in parallel when using stats and round robin iterator.
- [PHOENIX-2238](#): Bulk dataload Hive data with \001 delimiter.
- [PHOENIX-2274](#): Sort-merge join could not optimize out the sort on the right table.
- [PHOENIX-2313](#): TypeMismatchException thrown while querying a table that has an index with a Boolean (BUG-46167).
- [PHOENIX-2359](#): Configuration for PQS to use Protobuf serialization instead of JSON.
- [PHOENIX-2369](#): Some UDF tests failed due to not finding Hadoop classpath.
- [PHOENIX-2372](#): Check for null before passing value to Calendar.
- [PHOENIX-2448](#): Fix quoting for sqlline-thin.py on Windows.
- [PHOENIX-2449](#): QueryServer needs Hadoop configuration on classpath with Kerberos.
- [PHOENIX-2477](#): ClassCastException in IndexedWALEditCodec after HBASE-14501.

HDP 2.3.2 provided Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- [PHOENIX-1659](#) PHOENIXDatabaseMetaData.getColumns does not return REMARKS column.
- [PHOENIX-1978](#) UDF ArgumentTypeMismatchException.
- [PHOENIX-2011](#) Default, min, and max values should not require quotes around it in create function.
- [PHOENIX-2022](#) BaseRegionScanner.next should be abstract.
- [PHOENIX-2066](#) Existing client fails initialization due to upgrade attempting to create column with no name.
- [PHOENIX-2073](#) Two bytes character in LIKE expression is not allowed.
- [PHOENIX-2074](#) StackOverflowError with RoundRobinResultIterator.
- [PHOENIX-2075](#) MR integration uses single mapper unless table is salted.
- [PHOENIX-2096](#) Tweak criteria for when round robin iterator is used.
- [PHOENIX-2131](#) CastParseNode.toSQL omits closing parenthesis.
- [PHOENIX-2151](#) Two different UDFs called on same column return values from first UDF only.
- [PHOENIX-2254](#) zookeeper.znode.parent value is not taking affect in MR integration job.

HDP 2.3.0 provided Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- [PHOENIX-1395](#) ResultSpooler spill files are left behind in /tmp folder.
- [PHOENIX-1976](#) Exit gracefully if addShutdownHook fails.
- [PHOENIX-1980](#) CsvBulkLoad cannot load hbase-site.xml from classpath.
- [PHOENIX-1995](#) client uberjar doesn't support dfs.
- [PHOENIX-1996](#) Use ByteString instead of ZeroCopyByteString.
- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode.
- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode (addendum).
- [PHOENIX-2007](#) java.sql.SQLException: Encountered exception in sub plan [0] execution'.
- [PHOENIX-2010](#) Properly validate number of arguments passed to the functions in FunctionParseNode#validate.
- [PHOENIX-2012](#) RowKeyComparisonFilter logs unencoded data at DEBUG level.
- [PHOENIX-2013](#) Apply [PHOENIX-1995](#) to runnable uberjar as well.
- [PHOENIX-2027](#) Queries with Hints are raising IllegalStateException.
- [PHOENIX-2031](#) Unable to process timestamp/Date data loaded via PHOENIX org.apache.PHOENIX.pig.PHOENIXHBaseLoader.
- [PHOENIX-2032](#) psql.py is broken after [PHOENIX-2013](#).
- [PHOENIX-2033](#) PQS log environment details on launch.
- [PHOENIX-2063](#) Row value constructor doesn't work when used in COUNT.
- [PHOENIX-2181](#) HPOENIXHBaseLoader doesn't work with salted tables.
- [PHOENIX-914](#) Native HBase timestamp support to optimize date range queries in Phoenix.

1.6.14. Pig

HDP 2.4.0 provides Pig 0.15.0 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Pig 0.15.0 and the following Apache patches:

- [PIG-4635](#): NPE while running pig script in tez mode(pig 0.15 with tez 0.7).
- [PIG-4683](#): Nested order is broken after PIG-3591 in some cases.
- [PIG-4688](#): Limit followed by POPartialAgg can give empty or partial results in Tez.

- [PIG-4696](#): Empty map returned by a streaming_python udf wrongly contains a null key.
- [PIG-4703](#): TezOperator.stores shall not ship to backend.
- [PIG-4707](#): [Pig on Tez] Streaming job hangs with pig.exec.mapPartAgg=true.
- [PIG-4712](#): [Pig on Tez] NPE in Bloom UDF after Union.
- [PIG-4714](#): Improve logging across multiple components with callerId.
- [PIG-4744](#): Honor tez.staging-dir setting in tez-site.xml.

HDP 2.3.2 provided Pig 0.15.0 and the following Apache patches:

- [PIG-4627](#) [Pig on Tez] Self join does not handle null values correctly.
- [PIG-4628](#) Pig 0.14 job with order by fails in mapreduce mode with Oozie.
- [PIG-4649](#) [Pig on Tez] Union followed by HCatStorer misses some data.
- [PIG-4679](#) Performance degradation due to InputSizeReducerEstimator since [PIG-3754](#).

HDP 2.3.0 provided Pig 0.15.0 and the following Apache patch:

- [PIG-4624](#) Error on ORC empty file without schema,

1.6.15. Ranger

HDP 2.4.0 provides Ranger 0.5.0 and the following Apache patches:

- [RANGER-173](#): Utility scripts to create HDFS audit folders and policies.
- [RANGER-725](#): Add the right .gitignore file to the newly added projects so that directory listing is clean after a build.
- [RANGER-767](#): Refactor UserGroupSink implementation and consolidate performance improvements.
- [RANGER-772](#): Hive plugin Update Ranger authorizer to mimic changes made by Hive standard authorizer for the case when IMPORT can end up creating a table.
- [RANGER-773](#): Fix newly found Coverity scan issues for Ranger KMS.
- [RANGER-778](#): Fix user update issue.
- [RANGER-809](#): Audit framework need to cache the getHostName() values to reuse for successive calls.

HDP 2.3.4 provided Ranger 0.5.0 and the following Apache patches:

- [RANGER-246](#): Need to update the current implementation for recent changes in Kafka.

- [RANGER-526](#): Provide REST API to change user role.
- [RANGER-586](#): Ranger plugins should not add dependent libraries to component's CLASSPATH.
- [RANGER-590](#): Escape spaces in the user and group names which are part of rest call uri in UserSync process.
- [RANGER-602](#): Solr client in SolrCloud mode should work with zookeeper settings also.
- [RANGER-607](#): Unable to create multiple policyItems for same user or group.
- [RANGER-608](#): Denied access to list a directory does not generate audit.
- [RANGER-652](#): LDAP configuration tool.
- [RANGER-656](#): Ranger UI - KMS Need to handle 404 error when clicked on breadcrumb.
- [RANGER-658](#): Package ranger_credential_helper.py with Ranger Usersync assembly.
- [RANGER-661](#): Plugin receives empty policy list though the service has policies.
- [RANGER-663](#): Race condition during policy update causes policy to get in an bad state.
- [RANGER-664](#): Ranger PolicyRefresh REST Client timeout parameter should be configurable.
- [RANGER-665](#): ranger.ldap.ad.referral property is not getting updated in RANGER-admin-site.xml.
- [RANGER-666](#): Ranger to support Azure SQL Database.
- [RANGER-671](#): Add support to retrieve permissions for the logged in user from UserSession rather going to database every time.
- [RANGER-673](#): Setup changes to allow Ranger service to installed using custom service user.
- [RANGER-674](#): Ranger public rest api gives 200 response for wrong credential instead of 401.
- [RANGER-677](#): Ranger Admin fails to render policies referring to groups that contain "." in name.
- [RANGER-680](#): Remove public group by default in default policy for KMS repo.
- [RANGER-681](#): Update default sync intervals for LDAP and UNIX.
- [RANGER-682](#): Ranger to support Azure Blob Datastore as an audit destination via HDFS audit handler.
- [RANGER-684](#): Ranger Usersync - Add Ability to transform user/group names.

- [RANGER-687](#): after each 30 seconds audit is getting updated in plugin tab.
- [RANGER-688](#): Handle scenario where ids of XUser and XPortalUser are not in sync.
- [RANGER-697](#): KeyAdmin role user should see only KMS related audit access logs in Audit tab.
- [RANGER-700](#): Provide a wrapper shell script to run the FileSourceUserGroupBuilder process.
- [RANGER-701](#): Update setup scripts to allow special characters in passwords.
- [RANGER-702](#): Optimize policy download performance.
- [RANGER-705](#): Ranger Usersync should provide summary logs on the sync progress instead of not logging any details after 2000 users.
- [RANGER-706](#): Optimize audit db upgrade patches to minimize timeout issues.
- [RANGER-712](#): Create a new project which can serve as a template to write ranger extensions.
- [RANGER-713](#): Knox-plugin failed to enable after plugin modification for not to add dependent libraries to component's CLASSPATH.
- [RANGER-714](#): Enhancements to the db admin setup scripts.
- [RANGER-715](#): Fix issues reported by coverity test in Ranger Plugin ClassLoader.
- [RANGER-717](#): Hive and Hbase ranger plugin Audit to DB failed to log after plugin modification for not to add dependent libraries to component's CLASSPATH.
- [RANGER-720](#): Ldap discovery tool doesn't seem to be working as expected.
- [RANGER-724](#): AuditBatchQueue: prevQueueSize not recomputed after initial assignment - static code analyzer flagged issue.
- [RANGER-725](#): Add the right .gitignore file to the newly projects so that directory listing is clean after a build.
- [RANGER-727](#): Knox Plugin failed to AuditToSpool file when Audit Destination is down.
- [RANGER-731](#): Ranger plugin for YARN doesn't seem to be able to write audit to Kerberized HDFS.
- [RANGER-733](#): Implement best coding practices to resolve issues found during code scan.
- [RANGER-739](#): Ranger HBase Plugin returning null for RegionObserver.preCompact calls causing hbase:acl issue.
- [RANGER-740](#): Kafka Authorizer interface has added close() method. Ranger should also implement it.

- [RANGER-741](#): Fix installation script to skip Audit DB password check if audit source is SOLR.
- [RANGER-742](#): Ranger usersync fails after syncing 500 users from AD or ldap server when paged results is enabled.
- [RANGER-743](#): External users with Admin Role should be allowed to create/update users.
- [RANGER-744](#): Kafka Authorizer has updated how IP/Host is passed.
- [RANGER-745](#): Upgrade Apache commons-collections.
- [RANGER-747](#): RangerAdmin is considering "none" as valid ZK Host Name for Solr.
- [RANGER-748](#): Users in policy got changed after upgrade.
- [RANGER-749](#): Ranger KMS to support multiple KMS instances with keys across multiple clusters.
- [RANGER-754](#): Ranger YARN Plugin lookup and test connection should support SPENGO enabled HTTP Authentication.
- [RANGER-755](#): ldap run.sh script fails since auth directory does not exist.
- [RANGER-756](#): ldapTool fails with -r option to retrieve only users/group/all.
- [RANGER-757](#): [LDAP tool] authentication fails if use -d option to search only users.
- [RANGER-758](#): Handle special characters in passwords starting from -r.
- [RANGER-761](#): Transaction logs not getting generated under audit menu admin tab if policy name is changed.
- [RANGER-766](#): Yarn Plugin Config hadoop.security.authentication should be non-mandatory with default value.
- [RANGER-767](#): Refactor UserGroupSink implementation and consolidate performance improvements.

HDP 2.3.2 provided Ranger 0.5.0 and the following Apache patches:

- [RANGER-551](#) Policy Validation: If resource levels are not valid for any hierarchy then checks about missing mandatory levels should be skipped

BUG FIXES

- [RANGER-560](#) Policy validation: Provide user friendly error messages about validation failures
- [RANGER-580](#) Hbase plugin: Plugin may not work after upgrade
- [RANGER-584](#) Service validation: Provide user friendly error messages about validation failures

- [RANGER-587](#) ranger-admin-site.xml not getting updated when ranger.authentication.method is changed
- [RANGER-588](#) Take care of Ranger KMS installation even if 'java' is not in PATH
- [RANGER-593](#) Service def validation: Provide user friendly error messages about validation failures
- [RANGER-594](#) Policy Validation: Change the logic to generate friendly error messages to be like used for Service and Service def
- [RANGER-598](#) Update Ranger config migration script to work with Ranger 0.5
- [RANGER-615](#) Audit to db: Truncate all string values of audit record so that writing of audit does not fail
- [RANGER-618](#) KMS gets slower in key creation once Database grows
- [RANGER-621](#) Solr service-def JSON has incorrect impliedGrants for solr_admin permission
- [RANGER-622](#) Hive plugin: Add jar via beeline throws NPE
- [RANGER-623](#) Enable plugin scripts should handle file permissions for certain umask value
- [RANGER-624](#) Windows installation broken after SQLAnywhere support
- [RANGER-625](#) Change db flavor input parameter value from SQLAnywhere to SQLA
- [RANGER-627](#) Processing done by Audit Shutdown hooks can confuse someone looking at logs to think that shutdown of a service is held up due to Ranger plugin
- [RANGER-628](#) Make filters for ranger-admin search binds configurable
- [RANGER-630](#) Data consistency across API and UI
- [RANGER-632](#) Policy validation error messages produced by the server are not seen by the user
- [RANGER-637](#) Make REFERRAL property in Ranger User sync configurable
- [RANGER-638](#) Ranger admin should redirect back to login page when session cookies expires
- [RANGER-639](#) Storm plugin - commons-lang is a required dependency and hence should be packaged as part of storm plugin
- [RANGER-641](#) Ranger kms start fails if java is not set and started using service keyword
- [RANGER-642](#) Update USERSEARCHFILTER for Ranger Authentication on Windows
- [RANGER-653](#) Move delegated admin check to mgr layer from service layer for XPermMap and XAuditMap

HDP 2.3.0 provided Ranger 0.5.0 and the following Apache patches:

- [RANGER-422](#) Add additional database columns to support aggregation
- [RANGER-423](#) Support audit log aggregation in Ranger Admin UI
- [RANGER-513](#) Policy validation: resource hierarchies check does not work with single-node hierarchies as in HDFS
- [RANGER-551](#) Policy Validation: If resource levels are not valid for any hierarchy then checks about missing mandatory levels should be skipped.
- [RANGER-564](#) Add incubating to the release name

BUG FIXES

- [RANGER-219](#) Autocomplete behavior of hive tables/columns
- [RANGER-524](#) Hbase plugin: list command should prune the tables returned on user permissions
- [RANGER-529](#) Policy Validation: resources of a policy must match one of the resource hierarchies of the service def.
- [RANGER-533](#) Hbase plugin: if user does not have family-level access to any family in a table then user may be incorrectly denied access done at table/family level during get or scan
- [RANGER-539](#) Rolling downgrade changes
- [RANGER-545](#) Fix js error for lower versions of FF (less than 30)
- [RANGER-548](#) Key rollover command fails
- [RANGER-550](#) Hive plugin: Add audit logging support for metadata queries that have filtering support from hive
- [RANGER-553](#) Default policy creation during service creation should handle service defs with multiple hierarchies, e.g. hive, properly
- [RANGER-554](#) Ranger KMS keys listing page does not support pagination
- [RANGER-555](#) Policy view page (from access audit page) gives 404 with Oracle DB
- [RANGER-558](#) Hbase plugin: unless user has READ access at some level under the table/family being accessed (via scan/get) authorizer should throw an exception and audit
- [RANGER-565](#) Ranger Admin install fails (sometimes) with IO Error when DB used in Oracle
- [RANGER-566](#) Installation of Ranger on Oracle 12c with shared database needs to use private synonym instead of public synonym

- [RANGER-569](#) Enabling Ranger plugin for Hbase should not modify hbase.rpc.protection value
- [RANGER-570](#) Knox plugin: after upgrading ranger from 0.4 to 0.5 the knox plugin won't work because classes with old names are missing
- [RANGER-571](#) Storm plugin: after upgrading ranger from 0.4 to 0.5 the plugin won't work because classes with old names are missing
- [RANGER-575](#) Allow KMS policies to be assigned to all users
- [RANGER-576](#) Storm audit not showing access type in the Ranger Admin Audit UI

HDP CHANGES

- [RANGER-450](#) Failed to install Ranger component due to Ranger policyManager script failures

1.6.16. Slider

HDP 2.4.0 provides Slider 0.80.0 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Slider 0.80.0 and the following Apache patches:

- [SLIDER-128](#) Support graceful stop of component instances.
- [SLIDER-777](#) Provide slider dependencies as a self contained versioned tarball.
- [SLIDER-916](#) Fix Docker related bugs.

HDP 2.3.2 provided Slider 0.80.0 and the following Apache patches:

IMPROVEMENTS

- [SLIDER-812](#) Make component configurations in appConfig available on the SliderAgent side.

BUG FIXES

- [SLIDER-481](#) giving registry log messages meaningful text.
- [SLIDER-911](#) remove surplus jax rs jsr311-api JAR.
- [SLIDER-912](#) x-insecure rest API should be off by default.
- [SLIDER-923](#) switch to TryOnceThenFail retry policy on IPC (needed for Hadoop 2.8+).
- [SLIDER-931](#) Security permissions on set up ZK path are too lax.
- [SLIDER-941](#) Add JAAS config templates for HBase.

HDP 2.3.0 provided Slider 0.80.0 and the following Apache patches:

IMPROVEMENTS

- [SLIDER-812](#) Making component configurations in appConfig available on the SliderAgent side
- [SLIDER-891](#) Add ability to set Slider AM launch environment during cluster create/start

BUG FIXES

- [SLIDER-810](#) YARN config changes to enable partial logs upload for long running services (default include/exclude patterns does not upload any files)
- [SLIDER-877](#) move SLIDER_HOME assignment to slider.py
- [SLIDER-878](#) Slider cannot support jdk 1.8 for command slider registry --getconf hbase-site --name hb1
- [SLIDER-888](#) intermittent errors when accessing key store password during localization of cert stores
- [SLIDER-901](#) AgentClientProvider should use File.separator in paths for platform independency
- [SLIDER-902](#) add config to client cert gen command
- [SLIDER-904](#) Resource leak reported by coverity scan results
- [SLIDER-905](#) Container request fails when Slider requests container with node label and host constraints

1.6.17. Spark

HDP 2.4.0 provides Spark 1.6.0 and the following Apache patches:

- [SPARK-11314](#): Add service API and test service for Yarn Cluster schedulers.
- [SPARK-11315](#): Add YARN extension service to publish Spark events to YARN timeline service.
- [SPARK-11323](#): Add History Service Provider to service application histories from YARN timeline server.
- [SPARK-12417](#): Support to have ORC bloom filters during write code path.
- [SPARK-12898](#): Consider having dummyCallSite for HiveTableScan.
- [SPARK-12925](#): Improve HiveInspectors.unwrap for StringObjectInspector.getPrimitiveWritableObject.

HDP 2.3.4 provided Spark 1.5.2 and the following Apache patches:

- [SPARK-10058](#): CORE, TESTS, Fix the flaky tests in HeartbeatReceiverSuite.
- [SPARK-10389](#): SQL, [1.5 support order by non-attribute grouping expression on Aggregate.
- [SPARK-10515](#): When killing executor, the pending replacement executors should not be lost.
- [SPARK-10534](#): SQL, ORDER BY clause allows only columns that are present in the select projection list.
- [SPARK-10577](#): PYSPARK, DataFrame hint for broadcast join.
- [SPARK-10581](#): DOCS, Groups are not resolved in scaladoc in sql classes.
- [SPARK-10619](#): Can't sort columns on Executor Page.
- [SPARK-10741](#): SQL, Hive Query Having/OrderBy against Parquet table is not working
- [SPARK-10790](#): YARN, Fix initial executor number not set issue and consolidate the codes.
- [SPARK-10812](#): YARN, Fix shutdown of token renewer.
- [SPARK-10812](#): YARN, Spark hadoop util support switching to yarn.
- [SPARK-10825](#): CORE, TESTS, Fix race conditions in StandaloneDynamicAllocationSuite.
- [SPARK-10829](#): SQL, Fix 2 bugs for filter on partitioned columns.
- [SPARK-10833](#): BUILD, Inline, organize BSD/MIT licenses in LICENSE.
- [SPARK-10845](#): SQL, Makes spark.sql.hive.version a SQLConfEntry.
- [SPARK-10858](#): YARN: archives/jar/files rename with # doesn't work unl.
- [SPARK-10859](#): SQL, fix stats of StringType in columnar cache.
- [SPARK-10871](#): include number of executor failures in error msg.
- [SPARK-10885](#): STREAMING, Display the failed output op in Streaming UI.
- [SPARK-10889](#): STREAMING, Bump KCL to add MillisBehindLatest metric.
- [SPARK-10901](#): YARN, spark.yarn.user.classpath.first doesn't work.
- [SPARK-10904](#): SPARKR, Fix to support `select(df, c("col1", "col2"))`.
- [SPARK-10914](#): UnsafeRow serialization breaks when two machines have different Oops size.
- [SPARK-10932](#): PROJECT INFRA, Port two minor changes to release-build.sh from scripts' old repo.

- [SPARK-10934](#): SQL, handle hashCode of unsafe array correctly.
- [SPARK-10952](#): Only add hive to classpath if HIVE_HOME is set.
- [SPARK-10955](#): STREAMING, Add a warning if dynamic allocation for Streaming applications.
- [SPARK-10959](#): PYSPARK, StreamingLogisticRegressionWithSGD does not t....
- [SPARK-10959](#): PYSPARK, StreamingLogisticRegressionWithSGD does not train with given regParam and convergenceTol parameters.
- [SPARK-10960](#): SQL, SQL with windowing function should be able to refer column in inner select.
- [SPARK-10971](#): SPARKR, RRunner should allow setting path to Rscript.
- [SPARK-10973](#): ML, PYTHON, Fix IndexError exception on SparseVector when asked for index after the last non-zero entry.
- [SPARK-10980](#): SQL, fix bug in create Decimal.
- [SPARK-10981](#): SPARKR, SparkR Join improvements.
- [SPARK-11009](#): SQL, fix wrong result of Window function in cluster mode.
- [SPARK-11023](#): YARN, Avoid creating URIs from local paths directly.
- [SPARK-11026](#): YARN, spark.yarn.user.classpath.first does work for 'SPARK-submit -jars hdfs://user/foo.jar'.
- [SPARK-11032](#): SQL, correctly handle having.
- [SPARK-11039](#): DOCS, WEBUI, Document additional ui configurations.
- [SPARK-11047](#): Internal accumulators miss the internal flag when replaying events in the history server.
- [SPARK-11051](#): CORE, Do not allow local checkpointing after the RDD is materialized and checkpointed.
- [SPARK-11056](#): Improve documentation of SBT build.
- [SPARK-11063](#): STREAMING, Change preferredLocations of Receiver's RDD to hosts rather than hostports.
- [SPARK-11066](#): Update DAGScheduler's "misbehaved ResultHandler".
- [SPARK-11094](#): Strip extra strings from Java version in test runner.
- [SPARK-11103](#): SQL, Filter applied on Merged Parquet shema with new column fail.
- [SPARK-11104](#): STREAMING, Fix a deadlock in StreamingContext.stop.

- [SPARK-11126](#): SQL, Fix a memory leak in SQLListener._stageIdToStageMetrics.
- [SPARK-11126](#): SQL, Fix the potential flaky test.
- [SPARK-11135](#): SQL, Exchange incorrectly skips sorts when existing ordering is non-empty subset of required ordering.
- [SPARK-11153](#): SQL, Disables Parquet filter push-down for string and binary columns.
- [SPARK-11188](#): SQL, Elide stacktraces in bin/SPARK-sql for AnalysisExceptions.
- [SPARK-11233](#): SQL, register cosh in function registry.
- [SPARK-11244](#): SPARKR, sparkR.stop() should remove SQLContext.
- [SPARK-11246](#): SQL, Table cache for Parquet broken in 1.5.
- [SPARK-11251](#): Fix page size calculation in local mode.
- [SPARK-11264](#): bin/SPARK-class can't find assembly jars with certain GREP_OPTIONS set.
- [SPARK-11270](#): STREAMING, Add improved equality testing for TopicAndPartition from the Kafka Streaming API.
- [SPARK-11287](#): Fixed class name to properly start TestExecutor from deploy.client.TestClient.
- [SPARK-11294](#): SPARKR, Improve R doc for read.df, write.df, saveAsTable.
- [SPARK-11299](#): DOC, Fix link to Scala DataFrame Functions reference.
- [SPARK-11302](#): MLLIB, 2) Multivariate Gaussian Model with Covariance matrix returns incorrect answer in some cases.
- [SPARK-11303](#): SQL, filter should not be pushed down into sample.
- [SPARK-11417](#): SQL, no @Override in codegen.
- [SPARK-11424](#): Guard against double-close() of RecordReaders.
- [SPARK-11434](#): SQL, Fix test "Filter applied on merged Parquet schema with new column fails".
- [SPARK-5966](#): WIP, SPARK-submit deploy-mode cluster is not compatible with master local>.
- [SPARK-8386](#): SQL, add write.mode for insertIntoJDBC when the parm overwrite is false.

HDP 2.3.2 provided Spark 1.4.1 and the following Apache patches:

NEW FEATURES

- [SPARK-1537](#) Add integration with Yarn's Application Timeline Server.

- [SPARK-6112](#) Provide external block store support through HDFS RAM_DISK.

BUG FIXES

- [SPARK-10623](#) NoSuchElementException thrown when ORC predicate push-down is turned on.

HDP 2.3.0 provided Spark 1.3.1 and the following Apache patches:

IMPROVEMENTS

- [SPARK-7326](#) (Backport) Performing window() on a WindowedDStream doesn't work all the time JDK 1.7 repackaging

1.6.18. Sqoop

HDP 2.4.0 provides Sqoop 1.4.6 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided Sqoop 1.4.6 and the following Apache patch:

- [SQOOP-2399](#) Sqoop import into Hbase - BigDecimalSplitter java.lang.ArrayIndexOutOfBoundsException.

HDP 2.3.2 provided Sqoop 1.4.6 and the following Apache patches:

IMPROVEMENTS

- [SQOOP-2387](#) Sqoop should support importing from table with column names containing some special character.
- [SQOOP-2457](#) Add option to automatically compute statistics after loading data into a hive table.

HDP 2.3.0 provided Sqoop 1.4.6 and the following Apache patches:

IMPROVEMENTS

- [SQOOP-2370](#) Netezza - need to support additional options for full control character handling.

BUG FIXES

- [SQOOP-2326](#) Fix Netezza trunc-string option handling and unnecessary log directory during imports.

1.6.19. Storm

HDP 2.4.0 provides Storm 0.10.0-beta and the following Apache patches:

- [STORM-1476](#): Filter -c options from args and add them as part of storm.options.

- [STORM-422](#): Allow more arguments to be passed to storm jar.
- [STORM-745](#): Fix storm.cmd to evaluate 'shift' correctly with 'storm jar'.

HDP 2.3.4 provided Storm 0.10.0-beta and the following Apache patches:

- [STORM-1017](#): If ignoreZkOffsets set true, KafkaSpout will reset zk offset when recover from failure.
- [STORM-1030](#): Hive Connector Fixes.
- [STORM-1102](#): Add a default flush interval for HiveBolt.
- [STORM-1147](#): STORM JDBC Bolt should add validation to ensure either insertQuery or table name is specified and not both.
- [STORM-1203](#): Worker metadata file creation doesn't use STORM.log.dir config.
- [STORM-139](#): hashCode does not work for byte[].
- [STORM-412](#): Allow users to modify logging levels of running topologies.
- [STORM-742](#): Let ShellBolt treat all messages to update heartbeat.
- [STORM-992](#): A bug in the timer.clj might cause unexpected delay to schedule new event.
- [STORM-996](#): netty-unit-tests/test-batch demonstrates out-of-order delivery.

HDP 2.3.2 provided Storm 0.10.0-beta, with no additional Apache patches.

HDP 2.3.0 provided Storm 0.10.0-beta and the following Apache patches:

- [STORM-166](#) Highly Available Nimbus
- [STORM-583](#) Add Microsoft Azure Event Hub spout implementations
- [STORM-691](#) Add basic lookup / persist bolts
- [STORM-703](#) With hash key option for RedisMapState, only get values for keys in batch
- [STORM-708](#) CORS support for STORM UI.
- [STORM-711](#) All connectors should use collector.reportError and tuple anchoring.
- [STORM-713](#) Include topic information with Kafka metrics.
- [STORM-714](#) Make CSS more consistent with self, prev release
- [STORM-724](#) Document RedisStoreBolt and RedisLookupBolt which is missed.
- [STORM-727](#) Storm tests should succeed even if a storm process is running locally.
- [STORM-741](#) Allow users to pass a config value to perform impersonation.

1.6.20. Tez

HDP 2.4.0 provides Tez 0.7.0 and the following Apache patches:

- [TEZ-2307](#): Possible wrong error message when submitting new dag.
- [TEZ-2898](#): Tez tools : swimlanes.py is broken.
- [TEZ-2900](#): Ignore V_INPUT_DATA_INFORMATION when vertex is in Failed/Killed/Error.
- [TEZ-3017](#): HistoryACLManager does not have a close method for cleanup.
- [TEZ-3025](#): InputInitializer creation should use the dag ugi.
- [TEZ-3032](#): Incorrect start time in different events for DAG history events.
- [TEZ-3037](#): History URL should be set regardless of which history logging service is enabled.

HDP 2.3.4 provided Tez 0.7.0 and the following Apache patches:

- [TEZ-1314](#): Port MAPREDUCE-5821 to Tez.
- [TEZ-2398](#): Flaky test: TestFaultTolerance.
- [TEZ-2436](#): Tez UI: Add cancel button in column selector.
- [TEZ-2538](#): ADDITIONAL_SPILL_COUNT wrongly populated for DefaultSorter with multiple partitions.
- [TEZ-2553](#): Tez UI: Tez UI Nits.
- [TEZ-2561](#): Port for TaskAttemptListenerImpTezDag should be configurable.
- [TEZ-2758](#): Remove append API in RecoveryService after TEZ-1909.
- [TEZ-2781](#): Fallback to send only TaskAttemptFailedEvent if taskFailed heartbeat fails.
- [TEZ-2787](#): Tez AM should have java.io.tmpdir=./tmp to be consistent with tasks.
- [TEZ-2808](#): Race condition between preemption and container assignment.
- [TEZ-2829](#): Tez UI: minor fixes to in-progress update of UI from AM.
- [TEZ-2850](#): Tez MergeManager OOM for small Map Outputs.
- [TEZ-2851](#): Support a way for upstream applications to pass in a caller context to Tez.
- [TEZ-2866](#): Tez UI: Newly added columns wont be displayed by default in tables.
- [TEZ-2874](#): Improved logging for caller context.

- [TEZ-2878](#): Tez UI: AM error handling - Make the UI handle cases in which AM returns unexpected/no data.
- [TEZ-2882](#): Consider improving fetch failure handling.
- [TEZ-2885](#): Remove counter logs from AMWebController.
- [TEZ-2887](#): Tez build failure due to missing dependency in pom files.
- [TEZ-2893](#): Tez UI: Retain vertex info displayed in DAG details page even after completion.
- [TEZ-2894](#): Tez UI: Disable sorting for few columns while in progress. Display an alert on trying to sort them.
- [TEZ-2895](#): Tez UI: Add option to enable and disable in-progress.
- [TEZ-2896](#): Fix thread names used during Input/Output initialization.
- [TEZ-2899](#): Backport graphical view fix from TEZ-2899.
- [TEZ-2907](#): NPE in IFile.Reader.getLength during final merge operation.
- [TEZ-2908](#): Tez UI: Errors are logged, but not displayed in the UI when AM fetch fails.
- [TEZ-2909](#): Tez UI: Application link in All DAGs table is disable when applicationhistory is unavailable.
- [TEZ-2910](#): Tez should invoke HDFS Client API to set up caller context.
- [TEZ-2915](#): Tez UI: Getting back to the DAG details page is difficult.
- [TEZ-2923](#): Tez Live UI counters view empty for vertices, tasks, attempts.
- [TEZ-2927](#): Tez UI: Graciously fail when system-metrics-publisher is disabled.
- [TEZ-2929](#): Tez UI: Dag details page displays vertices to be running even when dag have completed.
- [TEZ-2930](#): Tez UI: Parent controller is not polling at times.
- [TEZ-2933](#): Tez UI: Load application details from RM when available.
- [TEZ-2936](#): Support HDFS-based Timeline writer.
- [TEZ-2946](#): Tez UI: At times RM return a huge error message making the yellow error bar to fill the whole screen.
- [TEZ-2947](#): Tez UI: Timeline, RM : AM requests gets into a consecutive loop in counters page without any delay.
- [TEZ-2949](#): Allow duplicate dag names within session for Tez.
- [TEZ-2960](#): Tez UI: Move hardcoded url namespace to the configuration file.

- [TEZ-2968](#): Counter limits exception causes AM to crash.
- [TEZ-2970](#): Re-localization in TezChild does not use correct UGI.
- [TEZ-2975](#): Bump up apache commons dependency.
- [TEZ-2995](#): Timeline primary filter should only be on callerId and not type.

HDP 2.3.2 provided Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

- [TEZ-2719](#) Consider reducing logs in unordered fetcher with shared-fetch option.
- [TEZ-2767](#) Make TezMxBeanResourceCalculator the default resource calculator.
- [TEZ-2789](#) Backport events added in TEZ-2612 to branch-0.7.
- [TEZ-2812](#) Tez UI, Update task and attempt tables while in progress.
- [TEZ-2813](#) Tez UI, add counter data for rest api calls to AM Web Services v2.
- [TEZ-2817](#) Tez UI, update in progress counter data for the dag vertices and tasks table.
- [TEZ-2817](#) Tez UI, update in progress counter data for the dag vertices and tasks table.
- [TEZ-2830](#) Backport TEZ-2774 to branch-0.7. Improvements to logging in the AM and part of the runtime.
- [TEZ-2844](#) Backport TEZ-2775 to branch-0.7. Improve and consolidate logging in Runtime components.
- [TEZ-2876](#) Tez UI, Update vertex, task and attempt details page while in progress.

BUG FIXES

- [TEZ-2211](#) Tez UI. Allow users to configure timezone.
- [TEZ-2291](#) TEZ UI. Improper vertex name in tables.
- [TEZ-2483](#) Tez UI, Show error if in progress fails due to AM not reachable.
- [TEZ-2549](#) Reduce Counter Load on the Timeline Server.
- [TEZ-2602](#) Throwing EOFException when launching MR job.
- [TEZ-2660](#) Tez UI, need to show application page even if system metrics publish is disabled.
- [TEZ-2663](#) SessionNotRunning exceptions are wrapped in a ServiceException from a dying AM.
- [TEZ-2745](#) ClassNotFound in InputInitializer causes AM to crash.

- [TEZ-2752](#) logUnsuccessful completion in Attempt should write original finish time to ATS.
- [TEZ-2754](#) Tez UI. StartTime and EndTime is not displayed with right format in Graphical View.
- [TEZ-2761](#) Addendum fix build failure for java 6.
- [TEZ-2761](#) Tez UI. update the progress on the dag and vertices pages with info from AM.
- [TEZ-2766](#) Tez UI. Add vertex in-progress info in DAG details.
- [TEZ-2768](#) Log a useful error message when the summary stream cannot be closed when shutting down an AM.
- [TEZ-2780](#) Tez UI, Update All Tasks page while in progress.
- [TEZ-2792](#) Add AM web service API for tasks.
- [TEZ-2792](#) Addendum fix build failure for java 6.
- [TEZ-2807](#) Log data in the finish event instead of the start event.
- [TEZ-2812](#) Preemption sometimes does not respect heartbeats between preemptions.
- [TEZ-2816](#) Preemption sometimes does not respect heartbeats between preemptions.
- [TEZ-2825](#) Report progress in terms of completed tasks to reduce load on AM for Tez UI.
- [TEZ-2834](#) Make Tez preemption resilient to incorrect free resource reported by YARN.
- [TEZ-2842](#) Tez UI, Update Tez App details page while in-progress.
- [TEZ-2847](#) Tez UI, Task details doesn't gets updated on manual refresh after job complete.
- [TEZ-2853](#) Tez UI, task attempt page is coming empty.
- [TEZ-814](#) Improve heuristic for determining a task has failed outputs.
- [TEZ-814](#) Improve heuristic for determining a task has failed outputs.

INCOMPATIBLE CHANGES

- [TEZ-2768](#) Log a useful error message when the summary stream cannot be closed when shutting down an AM.

HDP 2.3.0 provided Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

- [TEZ-2076](#) Tez framework to extract/analyze data stored in ATS for specific dag.
- [TEZ-2461](#) tez-history-parser compile fails with hadoop-2.4.

BUG FIXES

- [TEZ-1529](#) ATS and TezClient integration in secure kerberos enabled cluster.
- [TEZ-2391](#) TestVertexImpl timing out at times on jenkins builds.
- [TEZ-2409](#) Allow different edges to have different routing plugins
- [TEZ-2447](#) Tez UI: Generic changes based on feedbacks.
- [TEZ-2453](#) Tez UI: show the dagInfo is the application has set the same.
- [TEZ-2455](#) Tez UI: Dag view caching, error handling and minor layout changes
- [TEZ-2460](#) Temporary solution for issue due to [YARN-2560](#)
- [TEZ-2474](#) The old taskNum is logged incorrectly when parallelism is changed
- [TEZ-2475](#) Fix a potential hang in Tez local mode caused by incorrectly handled interrupts.
- [TEZ-2478](#) Move OneToOne routing to store events in Tasks.
- [TEZ-2481](#) Tez UI: graphical view does not render properly on IE11
- [TEZ-2482](#) Tez UI: Mouse events not working on IE11
- [TEZ-2489](#) Disable warn log for Timeline ACL error when tez.allow.disabled.timeline-domains set to true.
- [TEZ-2504](#) Tez UI: tables - show status column without scrolling, numeric 0 shown as Not available
- [TEZ-2505](#) PipelinedSorter uses Comparator objects concurrently from multiple threads.
- [TEZ-2509](#) YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.
- [TEZ-2513](#) Tez UI: Allow filtering by DAG ID on All dags table.
- [TEZ-2523](#) Tez UI: derive applicationId from dag/vertex id instead of relying on json data
- [TEZ-2527](#) Tez UI: Application hangs on entering erroneous RegEx in counter table search box
- [TEZ-2528](#) Tez UI: Column selector buttons gets clipped, and table scroll bar not visible in mac.
- [TEZ-2535](#) Tez UI: Failed task attempts link in vertex details page is broken.
- [TEZ-2539](#) Tez UI: Pages are not updating in IE.
- [TEZ-2541](#) DAGClientImpl enable TimelineClient check is wrong.

- [TEZ-2546](#) Tez UI: Fetch hive query text from timeline if dagInfo is not set.
- [TEZ-2547](#) Tez UI: Download Data fails on secure, cross-origin clusters
- [TEZ-2548](#) TezClient submitDAG can hang if the AM is in the process of shutting down.
- [TEZ-2554](#) Tez UI: View log link does not correctly propagate login credential to read log from YARN web.
- [TEZ-2568](#) auto_sortmerge_join_5 fails in Tez mode

1.6.21. ZooKeeper

HDP 2.4.0 provides ZooKeeper 3.4.6 with the patches specified below. No new additional Apache patches have been included in this release.

HDP 2.3.4 provided ZooKeeper 3.4.6 and the following Apache patches:

- [ZOOKEEPER-706](#) Large numbers of watches can cause session re-establishment to fail.

HDP 2.3.2 provided ZooKeeper 3.4.6 and the following Apache patches:

- [ZOOKEEPER-1506](#): Re-try DNS hostname -> IP resolution if node connection fails.
- [ZOOKEEPER-1574](#): Set svn:eol-style property to native.
- [ZOOKEEPER-1575](#): adding .gitattributes to prevent CRLF and LF mismatches for source and text files.
- [ZOOKEEPER-1791](#): ZooKeeper package includes unnecessary jars that are part of the package.
- [ZOOKEEPER-1848](#): [WINDOWS] Java NIO socket channels does not work with Windows ipv6 on JDK6.
- [ZOOKEEPER-1876](#): Add support for installing windows services in .cmd scripts.
- [ZOOKEEPER-1888](#): ZkCli.cmd commands fail with "'java' is not recognized as an internal or external command".
- [ZOOKEEPER-1897](#): ZK Shell/Cli not processing commands.
- [ZOOKEEPER-1926](#): Unit tests should only use build/test/data for data.
- [ZOOKEEPER-1952](#): Default log directory and file name can be changed.
- [ZOOKEEPER-2053](#): Make scripts use ZOOKEEPER_HOME if set.
- [ZOOKEEPER-2064](#): Prevent resource leak in various classes.

HDP 2.3.0 provided ZooKeeper 3.4.6, with no additional Apache patches.

1.7. Common Vulnerabilities and Exposures

- **CVE-2015-5167:** Restrict REST API data access for non-admin users

Severity: Important

Vendor: Hortonworks

Versions Affected: All HDP 2.3.x releases prior to 2.3.4

Users Affected: All users of ranger policy admin tool.

Impact: See BUG-41604 and [RANGER-630](#). Data access restrictions via REST API are not consistent with restrictions in policy admin UI.

Recommended Action: Upgrade to 2.3.4.x+ or HDP 2.4.0+.

- **CVE-2015-7521:** Apache Hive authorization bug disclosure

Severity: Important

Vendor: The Apache Software Foundation

Versions Affected: Apache Hive 1.0.0 - 1.0.1, Apache Hive 1.1.0 - 1.1.1, and Apache Hive 1.2.0 - 1.2.1

Description: Some partition-level operations exist that do not explicitly also authorize privileges of the parent table. This can lead to issues when the parent table would have denied the operation, but no denial occurs because the partition-level privilege is not checked by the authorization framework, which defines authorization entities only from the table level upwards. This issue is known to affect Hive clusters protected by both Ranger as well as SqlStdHiveAuthorization.

Mitigation: For Hive 1.0, 1.1 and 1.2, a separate jar is being made available, which users can put in their `${HIVE_HOME}/lib/`, and this provides a hook for administrators to add to their `hive-site.xml`, by setting `hive.semantic.analyzer.hook=org.apache.hadoop.hive.ql.parse.ParentTableAuthorizationHook`. This parameter is a comma-separated-list and this hook can be appended to an existing list if one already exists in the setup. You will then want to make sure that you protect the `hive.semantic.analyzer.hook` parameter from being changed at runtime by adding it to `hive.conf.restricted.list`. This jar and associated source tarball are available for download over at : <https://hive.apache.org/downloads.html> along with their gpg-signed .asc signatures, as well as the md5sums for verification in the `hive-parent-auth-hook/` directory. This issue has already been patched in all Hive branches that are affected, and any future release will not need these mitigation steps.

Hortonworks Bug ID: BUG-50827

- **CVE-2016-0733:** Ranger Admin authentication issue

Severity: Important

Vendor: Hortonworks

Versions Affected: All HDP 2.3.x releases prior to 2.3.4

Users Affected: All users of ranger policy admin tool.

Impact: See BUG-50669 and [RANGER-835](#). Malicious Users can gain access to ranger admin UI without proper authentication.

Recommended Action: Upgrade to 2.3.4.x+ or HDP 2.4.0+.

- **CVE-2016-0735:** In some cases, presence of an exclude policy at a level can give the user access at its parent level.

Severity: Critical

Vendor: Hortonworks

Versions Affected: All HDP 2.3.0+.

Users Affected: All users that use Ranger to authorize HBase, Hive, and Knox.

Impact: See BUG-50558. In some cases, presence of an exclude policy at a level can give the user access at its parent level. For example, if a hive policy excludes access for a user to a particular column, then such a user would be able to alter the name of that table. Only a user who has access at the table level should be able to do so. Due to this bug however, the user is able to do the operation which is caused by presence of an exclude policy at the column-level for that table. Recommended Action: Upgrade to HDP 2.4.0.0+ or contact Hortonworks support team.

Recommended Action: Upgrade to 2.3.7+ or HDP 2.4.0+.

1.8. Third-party Licenses

Global: [Apache 2.0](#)

Component	Subcomponents	License
Accumulo	JCommander	JCommander
Falcon	cern.colt* , cern.jet* , cern.clhep	CERN
Knox	ApacheDS, Groovy	ANTLR
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle
Oozie	JDOM Oro	
Phoenix		EPL
Storm	Logback	EPL

1.9. Fixed Issues

Fixed issues represents selected issues that were previously logged via Hortonworks Support, but are now addressed in the current release. These issues may have been reported in previous versions within the Known Issues section; meaning they were reported by customers or identified by Hortonworks Quality Engineering team.

Potential Data Loss

None.

Security

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-46910		Hue	Request to add HUE-1663 to fix LDAP referral handling for authentication.
BUG-50350	HBASE-14471	HBase	Thrift - HTTP Error 413 full HEAD if using kerberos authentication.
BUG-51013	TEZ-3025	Tez	InputInitializer creation should use the dag ugi.

Incorrect Results

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-36030	HIVE-12706	Hive	Hive from_utc_timestamp() function doesn't return the right result.

Stability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-41224		Hue	Workaround Hiveserver2 memory Leak.
BUG-50215	FALCON-1595	Falcon	In secure cluster, Falcon server loses ability to communicate with HDFS over time.
BUG-50700	HIVE-12660	Hive	HS2 memory leak with .hiverc file use.

Query Failure

None.

Upgrade

None.

Usability

None.

Performance

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-42479	HBASE-14575	HBase	Relax region read lock for compactions.

Other

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-50101		Hadoop Common	Remove ivy jar from hadoop-izo source code.

1.10. Known Issues

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-23260	HIVE-11421 , HIVE-11981	Hive	Insert after alter table on Acid table causes <code>ArrayIndexOutOfBoundsException</code> Issue: Schema evolution is not supported for Acid tables. Alter Table commands on Acid table may make the table unreadable. There is not w/a after that.
BUG-30022	HIVE-7693	Hive	Invalid column ref error in order by when using column alias in select clause and using having
BUG-30556		Hive	Killed job is not display correctly by ATS hook failure
BUG-38054	RANGER-577	Ranger	Ranger performs authorization check, even if Hive authorization is disabled When Ranger is Enabled for authorization, it would do the authorization check even if the hive authorization is disabled (via <code>hive.security.authorization.enabled = false</code>).
BUG-38928	HIVE-11599	Hive	Alter table <code>transactional=false</code> on ACID table misleading user. create transactional table would be better Issue: Once an ACID table is created (with <code>transactional=true</code>) and data is written to it, changing the property to <code>transactional=false</code> may cause wrong data to be returned by queries over this table.
BUG-39663		Hive	Hive compactor fails when 'hive' user does not have x permissions on the table directory
BUG-39988	HIVE-11110	Hive	CBO : Default partition filter is from MetaStore query causing TPC-DS to regress by 3x
BUG-43128	HIVE-12724	Hive	Query wrong results: non-ACID to ACID, major compaction does not compact all pre-ACID data into base dir
BUG-43129		Hive	ACID update/delete query fails with

Hortonworks Bug ID	Apache JIRA	Component	Summary
			ArrayIndexOutOfBoundsException when there is null struct
BUG-45846	HIVE-11605	Hive	Incorrect results with bucket map join in tez
BUG-45857	HIVE-5623	Hive	ORC accessing array column that's empty will fail with java out of bound exception
BUG-46092	HIVE-7193	Hive	We need to filter users who can connect on our HiveServer2 to those in a particular LDAP group (with nested groups)
BUG-46128		Storm	Kafka spout showing errors in the log streaming workflow
BUG-46771	HIVE-11716	Hive	Reading ACID table from non-acid session should raise an error
BUG-47626		HBase, Hive, Phoenix	Creating Hive table on Hbase fails in HDP 2.3.x with Phoenix Dependencies
BUG-48217		YARN	After restarting Yarn services, applications stays in ACCEPTED state
BUG-48416		Falcon	Unable to save a feed via Falcon UI
BUG-48979		YARN	<p>Issue: Jobs fail with javax.security.sasl.SaslException</p> <p>Cause: OpenJDK 7u91 is stricter about validating ticket caches; if there is any ticket in the cache without a default realm, SASL will fail to work.</p> <p>Workaround: To avoid this problem, the /etc/krb5.conf file must contain a domain_realm section.</p>
BUG-48990	HIVE-12837	Hive	MicroStrategy query 117 fails with out of memory error
BUG-49726	HIVE-10632	Hive	Hive compactor processing partitions of table that does not exist
BUG-49784	HDFS-9445	HDFS	Datanode may deadlock while handling a bad volume
BUG-49949		Hive	Query with duplicate columns in GROUP BY clause fails
BUG-49955	PHOENIX-2531	Phoenix	<p>Phoenix thin-client jar has incorrect META-INF/services/java.sql.Driver file</p> <p>Issue: The Phoenix Thin Client Driver, org.apache.phoenix.queryserver.client.Driver,</p>

Hortonworks Bug ID	Apache JIRA	Component	Summary
			<p>is not automatically registered in the JDBC DriverManager.</p> <p>Workaround:</p> <ol style="list-style-type: none"> 1. Add the following to your Java application before trying to obtain the driver: <pre> Properties props = new Properties(); String url = "jdbc:phoenix:thin:url= http:// localhost:8765"; try { Class.forName("org. apache.phoenix. queryserver.client. Driver"); } catch (Exception e) { throw new RuntimeException(e); } DriverManager. getConnection(url, props); </pre> 2. Manually register the driver.
BUG-50073	HDFS-9557	HDFS	Reduce object allocation in PB conversion
BUG-50089	HDFS-9572	HDFS	Prevent DataNode log spam if a client connects on the data transfer port but sends no data.
BUG-50302	HBASE-15039	HBase	<p>HMaster and RegionServers should try to refresh token keys from zk when facing InvalidToken</p> <p>Issue: Many customers use importTsv to load data into HBase, but the cluster network may not always be good. Sometimes they face many map tasks inside importTsv throwing InvalidToken exception.</p> <p>Workaround: Restart the region server.</p>
BUG-50479	HADOOP-11252	Hadoop Common	RPC client does not time out by default
BUG-50531		Kafka	<p>Kafka file system support</p> <p>Issue: Encrypted file systems such as SafenetFS are not supported for Kafka. Index file corruption can occur.</p> <p>Workaround: N/A</p>

Hortonworks Bug ID	Apache JIRA	Component	Summary
			For more information, see: Install Kafka .
BUG-50576	SQOOP-2779	Sqoop	Sqoop metastore doesn't seem to recognize --schema option
BUG-50599	HDFS-9600	HDFS	Do not check replication if the block is under construction
BUG-50652	HDFS-9574	HDFS	Reduce client failures during datanode restart
BUG-51239		Ranger	NameNode shutdown hangs due to Ranger HDFS plugin
BUG-51292		Kafka, Spark	Spark Streaming Kafka direct API unit test failure in HDP release Workaround: None.
BUG-51293	HDFS-9655	HDFS	NN should start JVM pause monitor before loading fsimage
BUG-51423	HDFS-9313	HDFS	Possible NullPointerException in BlockManager if no excess replica can be chosen
BUG-51424	HDFS-9314	HDFS	Improve BlockPlacementPolicyDefault's picking of excess replicas
BUG-51426	HDFS-9574	HDFS	Reduce client failures during datanode restart
BUG-51427	HDFS-8647 , HDFS-9600	HDFS	Do not check replication if the block is under construction
BUG-51428	HADOOP-11252	Hadoop Common	RPC client does not time out by default
BUG-51433	HDFS-9572	HDFS	Prevent DataNode log spam if a client connects on the data transfer port but sends no data.
BUG-51436	HDFS-9557	HDFS	Reduce object allocation in PB conversion
BUG-51439	HDFS-9445	HDFS	Datanode may deadlock while handling a bad volume
BUG-51442	HDFS-9655	HDFS	NN should start JVM pause monitor before loading fsimage
BUG-51463	HDFS-8898	HDFS	Create API and command-line argument to get quota and quota usage without detailed content summary
BUG-51657		Ranger	Cyclic symlink for ranger usersync configuration after upgrade Issue: When manually upgrading from 2.3.4 to 2.4.0, a cyclic symlink gets created for usersync conf.

Hortonworks Bug ID	Apache JIRA	Component	Summary
			<p>Cause: This issue arises due to wrongly created symlinks which make the conf links cyclic and invalid.</p> <p>Workaround:</p> <ul style="list-style-type: none"> For Ranger-Admin <ol style="list-style-type: none"> Delete the conf links from below paths: <pre> /usr/hdp/current/ranger-admin/ /usr/hdp/current/ranger-admin/ews/webapp/WEB-INF/classes/ /etc/ranger/admin/ </pre> Run the <code>set_globals.sh</code> script present under <code>/usr/hdp/current/ranger-admin/</code> to create appropriate symlinks. Re-Install the application from Ambari. For Ranger-Usersync: <ol style="list-style-type: none"> Delete the conf links from below paths: <pre> /usr/hdp/current/ranger-usersync/ /etc/ranger/usersync/ </pre> Run the <code>set_globals.sh</code> script present under <code>/usr/hdp/current/ranger-usersync/</code> to create appropriate symlinks. Re-Install the application from Ambari.
BUG-51712		Ranger	<p>After upgrade getting <code>java.lang.NoClassDefFoundError</code></p> <p>Issue: <code>java.lang.NoClassDefFoundError</code> error found in ranger usersync log after upgrade.</p> <p>Workaround: Copy the below jar files into ranger usersync lib folder and restart the service.</p>

Hortonworks Bug ID	Apache JIRA	Component	Summary
			cp /usr/hdp/current/hadoop-client/lib/commons-httpclient-3.1.jar /usr/hdp/current/ranger-usersync/lib/cp /usr/hdp/current/hadoop-client/lib/commons-codec-1.4.jar /usr/hdp/current/ranger-usersync/lib/

1.11. Documentation Errata

The following section contains late additions or corrections to the product documentation.

- [Flume: Kafka Sink](#)
- [Hive Sink](#)

1.11.1. Flume: Kafka Sink

This is a Flume Sink implementation that can publish data to a Kafka topic. One of the objectives is to integrate Flume with Kafka so that pull-based processing systems can process the data coming through various Flume sources. This currently supports Kafka 0.8.x series of releases.

Property Name	Default	Description
type	-	Must be set to org.apache.flume.sink.kafka.KafkaSink.
brokerList	-	List of brokers Kafka-Sink will connect to, to get the list of topic partitions. This can be a partial list of brokers, but we recommend at least two for HA. The format is a comma separated list of hostname:port.
topic	default-flume-topic	The topic in Kafka to which the messages will be published. If this parameter is configured, messages will be published to this topic. If the event header contains a "topic" field, the event will be published to that topic overriding the topic configured here.
batchSize	100	How many messages to process in one batch. Larger batches improve throughput while adding latency.
requiredAcks	1	How many replicas must acknowledge a message before it is considered successfully written. Accepted values are 0 (Never wait for acknowledgement), 1 (wait for leader only), -1 (wait for all replicas) Set this to -1 to avoid data loss in some cases of leader failure.
Other Kafka Producer Properties	-	These properties are used to configure the Kafka Producer. Any producer property supported by Kafka can be used. The only requirement is to prepend the property name with

Property Name	Default	Description
		the prefix "kafka.". For example: kafka.producer.type.

Note: Kafka Sink uses the topic and key properties from the FlumeEvent headers to send events to Kafka. If the topic exists in the headers, the event will be sent to that specific topic, overriding the topic configured for the Sink. If key exists in the headers, the key will be used by Kafka to partition the data between the topic partitions. Events with same key will be sent to the same partition. If the key is null, events will be sent to random partitions.

An example configuration of a Kafka sink is given below. Properties starting with the prefix kafka (the last 3 properties) are used when instantiating the Kafka producer. The properties that are passed when creating the Kafka producer are not limited to the properties given in this example. It is also possible include your custom properties here and access them inside the preprocessor through the Flume Context object passed in as a method argument.

```
al.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink al.sinks.k1.topic = mytopic
al.sinks.k1.brokerList = localhost:9092
al.sinks.k1.requiredAcks = 1
al.sinks.k1.batchSize = 20
al.sinks.k1.channel = c1
```

1.11.2. Hive Sink

This sink streams events containing delimited text or JSON data directly into a Hive table or partition. Events are written using Hive transactions. As soon as a set of events are committed to Hive, they become immediately visible to Hive queries. Partitions to which flume will stream to can either be pre-created or, optionally, Flume can create them if they are missing. Fields from incoming event data are mapped to corresponding columns in the Hive table.

Property Name	Default	Description
channel	-	
type	-	The component type name, needs to be hive.
hive.metastore	-	Hive metastore URI (eg thrift://a.b.com:9083).
hive.database	-	Hive database name
hive.table	-	Hive table name.
hive.partition	-	Comma separated list of partition values identifying the partition to write to. May contain escape sequences. E.g: If the table is partitioned by (continent: string, country :string, time : string) then 'Asia,India,2014-02-26-01-21' will indicate continent=Asia, country=India, time=2014-02-26-01-21.
hive.txnsPerBatchAsk	100	Hive grants a batch of transactions instead of single transactions to streaming clients like Flume. This setting configures the number of desired transactions per Transaction

Property Name	Default	Description
		Batch. Data from all transactions in a single batch end up in a single file. Flume will write a maximum of batchSize events in each transaction in the batch. This setting in conjunction with batchSize provides control over the size of each file. Note that eventually Hive will transparently compact these files into larger files.
heartBeatInterval	240	(In seconds) Interval between consecutive heartbeats sent to Hive to keep unused transactions from expiring. Set this value to 0 to disable heartbeats .
autoCreatePartitions	true	Flume will automatically create the necessary Hive partitions to stream to.
batchSize	15000	Max number of events written to Hive in a single Hive transaction.
maxOpenConnections	500	Allow only this number of open connections. If this number is exceeded, the least recently used connection is closed.
callTimeout	10000	(In milliseconds) Timeout for Hive & HDFS I/O operations, such as openTxn, write, commit, abort.
serializer	–	Serializer is responsible for parsing out field from the event and mapping them to columns in the hive table. Choice of serializer depends upon the format of the data in the event. Supported serializers: DELIMITED and JSON.
roundUnit	minute	The unit of the round down value - second, minute or hour.
roundValue	1	Rounded down to the highest multiple of this (in the unit configured using hive.roundUnit), less than current time.
timeZone	Local	Name of the timezone that should be used for resolving the escape sequences in partition, e.g. Time America/Los_Angeles.
useLocalTimeStamp	false	Use the local time (instead of the timestamp from the event header) while replacing the escape sequences.

Following serializers are provided for Hive sink:

- **JSON:** Handles UTF8 encoded Json (strict syntax) events and requires no configuration. Object names in the JSON are mapped directly to columns with the same name in the Hive table. Internally uses org.apache.hive.hcatalog.data.JsonSerDe but is independent of the Serde of the Hive table. This serializer requires HCatalog to be installed.
- **DELIMITED:** Handles simple delimited textual events. Internally uses LazySimpleSerde but is independent of the Serde of the Hive table.

Property Name	Default	Description
serializer.delimiter	,	(Type: string) The field delimiter in the incoming data. To use special

Property Name	Default	Description
		characters, surround them with double quotes like “\t”.
serializer.fieldnames	–	The mapping from input fields to columns in hive table. Specified as a comma separated list (no spaces) of hive table columns names, identifying the input fields in order of their occurrence. To skip fields leave the column name unspecified. Eg. ‘time,,ip,message’ indicates the 1st, 3rd and 4th fields in input map to time, ip and message columns in the hive table.
serializer.serdeSeparator	Ctrl-A	(Type: character) Customizes the separator used by underlying serde. There can be a gain in efficiency if the fields in serializer.fieldnames are in same order as table columns, the serializer.delimiter is same as the serializer.serdeSeparator and number of fields in serializer.fieldnames is less than or equal to number of table columns, as the fields in incoming event body do not need to be reordered to match order of table columns. Use single quotes for special characters like ‘\t’. Ensure input fields do not contain this character. Note: If serializer.delimiter is a single character, preferably set this to the same character.

The following are the escape sequences supported:

Alias	Description
%{host}	Substitute value of event header named “host”. Arbitrary header names are supported.
%t	Unix time in milliseconds .
%a	Locale’s short weekday name (Mon, Tue, ...)
%A	Locale’s full weekday name (Monday, Tuesday, ...)
%b	Locale’s short month name (Jan, Feb, ...)
%B	Locale’s long month name (January, February, ...)
%c	Locale’s date and time (Thu Mar 3 23:05:25 2005)
%d	Day of month (01)
%D	Date; same as %m/%d/%y
%H	Hour (00..23)
%I	Hour (01..12)
%j	Day of year (001..366)
%k	Hour (0..23)
%m	Month (01..12)
%M	Minute (00..59)
%p	Locale’s equivalent of am or pm
%s	Seconds since 1970-01-01 00:00:00 UTC
%S	Second (00..59) %y last two digits of year (00..99)

Alias	Description
%Y	Year (2015)
%z	+hhmm numeric timezone (for example, -0400)

Example Hive table:

```
create table weblogs ( id int , msg string )
partitioned by (continent string, country string, time string)
clustered by (id) into 5 buckets
stored as orc;
```

Example for agent named a1:

```
a1.channels = c1
a1.channels.c1.type = memory
a1.sinks = k1
a1.sinks.k1.type = hive
a1.sinks.k1.channel = c1
a1.sinks.k1.hive.metastore = thrift://127.0.0.1:9083
a1.sinks.k1.hive.database = logsdb
a1.sinks.k1.hive.table = weblogs
a1.sinks.k1.hive.partition = asia,{country},%Y-%m-%d-%H-%M
a1.sinks.k1.useLocalTimeStamp = false
a1.sinks.k1.round = true
a1.sinks.k1.roundValue = 10
a1.sinks.k1.roundUnit = minute
a1.sinks.k1.serializer = DELIMITED
a1.sinks.k1.serializer.delimiter = "\t"
a1.sinks.k1.serializer.serdeSeparator = '\t'
a1.sinks.k1.serializer.fieldnames =id,msg
```

Note: For all of the time related escape sequences, a header with the key “timestamp” must exist among the headers of the event (unless useLocalTimestamps set to true). One way to add this automatically is to use the TimestampInterceptor.

The above configuration will round down the timestamp to the last 10th minute. For example, an event with timestamp header set to 11:54:34 AM, June 12, 2012 and ‘country’ header set to ‘india’ will evaluate to the partition (continent=‘asia’,country=‘india’,time=‘2012-06-12-11-50’. The serializer is configured to accept tab separated input containing three fields and to skip the second field.