

Hortonworks Data Platform

HDP-2.3.0 Release Notes

(Jul 21, 2015)

Hortonworks Data Platform: HDP-2.3.0 Release Notes

Copyright © 2012-2015 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 2.3.0 Release Notes | 1 |
| 1.1. New Features | 2 |
| 1.2. Unsupported Features | 4 |
| 1.2.1. Technical Preview Features | 5 |
| 1.2.2. Community Features | 5 |
| 1.3. HDP 2.3 Repositories | 6 |
| 1.4. Behavior Changes | 8 |
| 1.5. Fixed Issues | 10 |
| 1.6. Apache Patch Information | 17 |
| 1.6.1. Apache Hadoop | 17 |
| 1.6.2. Accumulo | 21 |
| 1.6.3. Atlas | 23 |
| 1.6.4. Calcite | 23 |
| 1.6.5. Falcon | 23 |
| 1.6.6. Flume | 23 |
| 1.6.7. HBase | 25 |
| 1.6.8. Hive | 25 |
| 1.6.9. Kafka | 26 |
| 1.6.10. Knox | 29 |
| 1.6.11. Mahout | 29 |
| 1.6.12. Oozie | 30 |
| 1.6.13. Phoenix | 30 |
| 1.6.14. Pig | 30 |
| 1.6.15. Ranger | 30 |
| 1.6.16. Slider | 32 |
| 1.6.17. Spark | 32 |
| 1.6.18. Sqoop | 33 |
| 1.6.19. Storm | 33 |
| 1.6.20. Tez | 33 |
| 1.7. Known Issues | 35 |
| 1.8. Documentation Errata | 44 |
| 1.8.1. Flume: Kafka Sink | 44 |
| 1.8.2. Hive Sink | 45 |

List of Tables

| | |
|---|----|
| 1.1. Data Governance and Integration | 2 |
| 1.2. Data Access | 2 |
| 1.3. Data Management | 3 |
| 1.4. Security | 4 |
| 1.5. Operations | 4 |
| 1.6. Technical Previews | 5 |
| 1.7. Community Features | 6 |
| 1.8. CentOS 6 | 7 |
| 1.9. CentOS 7 | 7 |
| 1.10. SUSE Linux Enterprise 11 Service Pack 3 | 7 |
| 1.11. Helper Files | 8 |
| 1.12. HBase | 8 |
| 1.13. Spark | 8 |
| 1.14. HDP Search | 8 |
| 1.15. HDFS: High Availability | 8 |
| 1.16. JDK Support | 9 |
| 1.17. Apache HBase | 35 |
| 1.18. Apache Hive | 36 |
| 1.19. Apache Oozie | 39 |
| 1.20. Apache Ranger | 39 |
| 1.21. Apache Slider | 40 |
| 1.22. Apache Spark | 40 |
| 1.23. Apache Tez | 40 |
| 1.24. Apache YARN | 41 |
| 1.25. HDFS and Cloud Deployment | 43 |
| 1.26. Hue | 43 |
| 1.27. Upgrade | 43 |

1. HDP 2.3.0 Release Notes

This document provides you with the latest information about the HDP 2.3.0 release and its product documentation.

Component Versions

The official Apache versions of all HDP 2.3 components are listed below. All 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 2.3.

Official Apache versions for HDP 2.3:

- Apache Hadoop 2.7.1
- 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 HBase 1.1.1
- Apache Hive 1.2.1
- Apache Kafka 0.8.2
- 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 Spark 1.3.1
- Apache Sqoop 1.4.6
- Apache Solr 5.2.1

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

Additional component versions:

- Hue 2.6.1
- Cloudbreak 1.0
- Cascading 3.0.1

1.1. New Features

This section highlights several new features in HDP 2.3.

Table 1.1. Data Governance and Integration

| Component | Feature |
|-----------|--|
| Atlas | <ul style="list-style-type: none"> • General Availability release • Hive integration (ATLAS-75) • Business classification (ATLAS-76) • REST API to create new connections to Atlas (ATLAS-77) |
| Falcon | <ul style="list-style-type: none"> • Replication of Hive and HCat – GA (FALCON-1028) • High Availability (FALCON-1029) • UI to enable entity/process/feed management (FALCON-790) • Search UI and backend support by Free text, Entity and Tag (FALCON-914, FALCON-1095, FALCON-1121, FALCON-1122) |
| Flume | <ul style="list-style-type: none"> • Kafka Source and Kafka Sink support (FLUME-2242) • Hive Sink based on the new Hive Streaming support (FLUME-1734) • An alternative to providing clear text passwords in Flume config (FLUME-2442) |
| Kafka | <ul style="list-style-type: none"> • Generate Hadoop delegation token (KAFKA-1696) • Ranger Integration for Authorization (KAFKA-1688) • Implement SASL/Kerberos (KAFKA-1686) |
| Sqoop | <ul style="list-style-type: none"> • Import sequential datasets from mainframe (SQOOP-1272) • Netezza enhancements: skip control codes, write logs to HDFS (SQOOP-2164) |

Table 1.2. Data Access

| Component | Feature |
|-------------------|---|
| HBase and Phoenix | <ul style="list-style-type: none"> • RPC throttling and quotas (HBASE-11598) |

| Component | Feature |
|-----------|--|
| | <ul style="list-style-type: none"> • Multi-WAL Support (HBASE-5699) • Reliability and performance optimizations (HBASE-12439) • Phoenix SQL improvements: Union All, Date/Decimal types (PHOENIX-1580, PHOENIX-1662) • Phoenix support for Java UDFs (PHOENIX-538) • Phoenix support for HBase Timeline-Consistent Read High Availability (PHOENIX-1683) • Phoenix Tracing Support (PHOENIX-1115) • Phoenix Spark Driver (PHOENIX-1071) |
| Hive | <ul style="list-style-type: none"> • SQL improvements: Current_Date, Current_Timestamp (HIVE-5472) • Other Date/Datetime improvements, such as SerDe support for ISO 8601 format (HIVE-9298, HIVE-9564) • Support UNION (HIVE-9039) and add Interval datatype in expressions (HIVE-9792, HIVE-5021) |
| Pig | <ul style="list-style-type: none"> • Call Hive UDFs from Pig (PIG-3294) • Dynamic Parallelism via Tez (PIG-4434) |
| Spark | <ul style="list-style-type: none"> • General availability of Spark 1.3.1 |
| Solr | <ul style="list-style-type: none"> • Authorization via Ranger (SOLR-7275) • Pluggable authentication framework that proves a Kerberos plugin implementation (SOLR-7468) |
| Storm | <ul style="list-style-type: none"> • Declarative Topology (STORM-561) • Rolling Upgrade (STORM-634) |
| Tez | <ul style="list-style-type: none"> • Secure ATS integration (TEZ-1529) • Enhanced performance and scale (TEZ-776) • Support sort buffers larger than 2GB (TEZ-1803) |

Table 1.3. Data Management

| Component | Feature |
|-----------|---|
| HDFS | <ul style="list-style-type: none"> • Improve distcp efficiency: reduced time and processing power needed to mirror datasets across cluster (HDFS-7535, MAPREDUCE-6248) • Support variable-length blocks (HDFS-3689) • Provide storage quotas per heterogeneous storage types (HDFS-7584) • Pluggable Authorization API (HDFS-6826) • Track and display failed DataNode storage locations in NameNode JMX and UI (HDFS-7604) • Additional DataNode and NameNode operational load metrics available through JMX (HDFS-7773) • HDFS Data at Rest Encryption (HDFS-6134) |
| Slider | <ul style="list-style-type: none"> • Upgrade and reconfiguration without downtime for Slider-based applications (SLIDER-787) |

| Component | Feature |
|-----------|--|
| | <ul style="list-style-type: none"> No packaging required for certain Slider applications (SLIDER-668) |
| YARN | <ul style="list-style-type: none"> Non-exclusive Node Labels - where applications are given preference for the Label they specify, but not exclusive access (YARN-3214) Fair sharing across apps for same user same queue, per queue scheduling policies (YARN-3306) Pluggable authorization for YARN ACLs for integration with Apache Ranger (YARN-3100) |

Table 1.4. Security

| Component | Feature |
|-----------|---|
| Knox | <ul style="list-style-type: none"> Modular architecture - introduce the concept of Knox "stacks", making it easier for customers and partners to add in APIs they wish to protect via Knox (KNOX-481, KNOX-483) Rolling upgrade Support for two-way SSL (KNOX-504) Support for LDAP authentication caching (KNOX-524) Enhance principal mapping across domains supplied with a header |
| Ranger | <ul style="list-style-type: none"> Modular architecture - introduce the concept of Ranger "stacks", making it easier for customers and partners to add in authorization and secure audit support their own component(s) via Ranger (RANGER-203) Extend support for Kafka, YARN, and Solr (RANGER-246, RANGER-248, RANGER-249) Ranger-based KMS support for HDFS encryption (RANGER-247) Use Solr for storing audit logs and querying (RANGER-266) |

Table 1.5. Operations

| Component | Feature |
|------------|--|
| Ambari | <ul style="list-style-type: none"> General Availability of Ambari 2.1 (see the Ambari Release Notes) |
| Oozie | <ul style="list-style-type: none"> HiveServer2 action (OOZIE-1457) Stop jobs by coordinator name (OOZIE-2108) |
| SmartSense | <ul style="list-style-type: none"> A next generation set of tools and services with diagnostics, cluster analytics, and actionable data-driven recommendations. For more information, see Support Tools. (Note: link requires Support Portal access.) |

1.2. Unsupported Features

Some features exist within HDP 2.3, but Hortonworks does not currently support these specific capabilities.

1.2.1. Technical Preview Features

The following features are available within HDP 2.3, 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 1.6. Technical Previews

| Component | Feature |
|-------------------|--|
| Cloudbreak | <ul style="list-style-type: none"> Autoscaling (previously called Periscope) of a cluster |
| HBase and Phoenix | <ul style="list-style-type: none"> Phoenix Query Server (PHOENIX-971) Support for <code>init.d</code> scripts Phoenix Query Server RPC Throttling Phoenix-Spark Integration |
| Kafka | <ul style="list-style-type: none"> SSL |
| Ranger | <ul style="list-style-type: none"> Security features for data governance: global policies by metadata tags and searchable security access audit |
| Slider | <ul style="list-style-type: none"> Support for Docker-based application packaging (SLIDER-780) |
| Spark | <ul style="list-style-type: none"> SparkSQL, programmatically with SQLContext (not supported with Thrift Server - JDBC/ODBC) DataFrame API (SPARK-5097) Spark Streaming ML Pipeline API in PySpark (SPARK-3530) ORC file support Dynamic Executor Allocation |
| Storm | <ul style="list-style-type: none"> Elastic topology via YARN/Slider Monitoring of Storm topologies and clusters User Impersonation JDBC Bolt and Trident Storm-Slider |
| YARN | <ul style="list-style-type: none"> NodeManager: add cgroup support for disk I/O isolation (YARN-2619) Add support for network I/O isolation/scheduling for containers (YARN-2140) |

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 1.7. Community Features

| Component | Feature |
|------------|--|
| Cloudbreak | <ul style="list-style-type: none"> Hosted Cloudbreak (cloudbreak.sequenceiq.com) |
| Falcon | <ul style="list-style-type: none"> User Recipes Prism Server |
| 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 |
| HDFS | <ul style="list-style-type: none"> NameNode Federation (HDFS-1052) viewFS (HADOOP-7257) block-volume device choosing (HDFS-1804) |
| Kafka | <ul style="list-style-type: none"> New Consumer API Mirror Maker (not supported when Kafka security is active) |
| 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 GraphX |
| YARN | <ul style="list-style-type: none"> Fair Scheduler MapReduce Uber AM MapReduce Eclipse Plug-in |

1.3. HDP 2.3 Repositories

HDP 2.3.0 for Linux supports CentOS 6, CentOS 7, and SUSE Linux Enterprise 11 Service Pack 3. **Note:** As of HDP 2.3.0, SUSE Linux Enterprise 11 Service Pack 1 is no longer supported.

HDP 2.3.0 for Linux does not support Ubuntu 12.0.4, Ubuntu 14.0.4, Debian 6, or Debian 7; Hortonworks will add support for Ubuntu and Debian in an upcoming maintenance release of HDP 2.3.0. If Debian and Ubuntu are critical to your environment today, use HDP 2.2.6.

Use the following links to download HDP 2.3.



Note

The package identifier for HDP 2.3.0 components is **2557**. For example:

```
dfs -mkdir -p /hdp/apps/2.3.0.0-<$version>/hive/
```

would become:

```
dfs -mkdir -p /hdp/apps/2.3.0.0-2557/hive/
```

Table 1.8. CentOS 6

| Description | Link |
|---------------------|---|
| Component metadata | http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/build_metadata.txt |
| HDP M2 Artifacts | http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/HDP-2.3.0.0-centos6-m2-artifacts.tar |
| HDP repo | http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/hdp.repo |
| HDP RPM tarball | http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/HDP-2.3.0.0-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.noarch.rpm |
| HDP-UTILS | http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/HDP-UTILS-1.1.0.20-centos6.tar.gz |
| Slider app packages | http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/slider-app-packages/pkg-list.txt |

Table 1.9. CentOS 7

| Description | Link |
|---------------------|---|
| Component metadata | http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.0.0/build_metadata.txt |
| HDP M2 Artifacts | http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.0.0/HDP-2.3.0.0-centos7-m2-artifacts.tar |
| HDP repo | http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.0.0/hdp.repo |
| HDP RPM tarball | http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.0.0/HDP-2.3.0.0-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.noarch.rpm |
| HDP-UTILS | http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos7/HDP-UTILS-1.1.0.20-centos7.tar.gz |
| Slider app packages | http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.0.0/slider-app-packages/pkg-list.txt |

Table 1.10. SUSE Linux Enterprise 11 Service Pack 3

| Description | Link |
|--------------------|---|
| Component metadata | http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.0.0/build_metadata.txt |
| HDP M2 Artifacts | http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.0.0/HDP-2.3.0.0-suse11sp3-m2-artifacts.tar |
| HDP repo | http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.0.0/hdp.repo |
| HDP RPM tarball | http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.0.0/HDP-2.3.0.0-suse11sp3-rpm.tar.gz |
| HDP-UTILS | http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/suse11sp3/HDP-UTILS-1.1.0.20-suse11sp3.tar.gz |

| Description | Link |
|---------------------|---|
| Slider app packages | http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.0.0/slider-app-packages/pkg-list.txt |

Table 1.11. Helper Files

| Description | Link |
|-----------------|---|
| Companion files | http://public-repo-1.hortonworks.com/HDP/tools/2.3.0.0/hdp_manual_install_rpm_helper_files-2.3.0.0.2557.tar.gz |

1.4. Behavior Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.3.0, behavioral changes affect the following Hadoop components.

Table 1.12. HBase

| Hortonworks Bug ID | Apache JIRA | Description |
|--------------------|-----------------------------|---|
| BUG-34234 | HBASE-10123 | Hbase default ports have changed in HDP 2.3. All ports numbered "61xxx" should be changed to "16xxx" |

Table 1.13. Spark

| Description |
|---|
| Spark reads data from HDFS/Hive (ORC). |
| <ul style="list-style-type: none"> Upgrade your HDP cluster first, resubmit Spark jobs, and validate job results. |
| API changes: |
| <ul style="list-style-type: none"> SchemaRDD changed to DataFrame SparkSQL implicits package (import sqlContext._ > import sqlContext.implicits._) UDF registration moved to sqlContext.udf |

Table 1.14. HDP Search

| Hortonworks Bug ID | Description |
|--------------------|--|
| BUG-34314 | Solr is now installed via RPM packages (vs. tarballs). |

Table 1.15. HDFS: High Availability

| Hortonworks Bug ID | Problem |
|--------------------|---|
| BUG-22998 | <p>HDFS-6376 allows <code>distcp</code> to copy data between HA clusters. Users can use a new configuration property <code>dfs.internal.nameservices</code> to explicitly specify the name services belonging to the local cluster, while continue using the configuration property <code>dfs.nameservices</code> to specify all of the name services in the local and remote clusters.</p> <p>Steps:</p> <p>Modify the following in the <code>hdfs-site.xml</code> for both cluster A and B:</p> |

| Hortonworks Bug ID | Problem |
|--------------------|---|
| | <ol style="list-style-type: none"> 1. Add both name services to <code>dfs.nameservices = HAA, HAB</code> 2. Add property <code>dfs.internal.nameservices</code> <ul style="list-style-type: none"> • In cluster A: <pre>dfs.internal.nameservices = HAA</pre> • In cluster B: <pre>dfs.internal.nameservices = HAB</pre> 3. Add <code>dfs.ha.namenodes.<nameservice></code> to both clusters <ul style="list-style-type: none"> • in cluster A <pre>dfs.ha.namenodes.HAB = nn1,nn2</pre> • In cluster B <pre>dfs.ha.namenodes.HAA = nn1,nn2</pre> 4. Add property <code>dfs.namenode.rpc-address.<cluster>.<nn></code> <ul style="list-style-type: none"> • In Cluster A <pre>dfs.namenode.rpc-address.HAB.nn1 = <NN1_fqdn>:8020</pre> <pre>dfs.namenode.rpc-address.HAB.nn2 = <NN2_fqdn>:8020</pre> • In Cluster B <pre>dfs.namenode.rpc-address.HAA.nn1 = <NN1_fqdn>:8020</pre> <pre>dfs.namenode.rpc-address.HAA.nn2 = <NN2_fqdn>:8020</pre> 5. Add property <code>dfs.client.failover.proxy.provider.<cluster></code> <ul style="list-style-type: none"> • In cluster A <pre>dfs.client.failover.proxy.provider.HAB = org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider</pre> • In cluster B <pre>dfs.client.failover.proxy.provider.HAA = org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider</pre> 6. Restart the HDFS service. Then run the <code>distcp</code> command using the NameService. For example: <pre>hadoop distcp hdfs://falconG/tmp/testDistcp hdfs://falconE/tmp/</pre> |

Table 1.16. JDK Support

| Description |
|-----------------------------------|
| HDP 2.3 supports JDK 1.7 and 1.8. |

1.5. Fixed Issues

The following features and fixes were contributed back to Apache with the release of HDP 2.3.0.

Potential Data Loss

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|-----------------------------|--|
| HBase | BUG-33249 | HBASE-13576 | HBCK failed to recover certain regions |

Security

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|----------------------|--------------------|-----------------------------|---|
| Falcon | BUG-32768 | FALCON-954 | Secure Kerberos setup : Falcon should periodically revalidate security credentials. |
| Falcon Documentation | BUG-31545 | | Add details of Falcon directory permissions to documentation |
| HBase | BUG-30856 | HBASE-11869 | Support snapshot owner other than Global Admins in Hbase |
| HBase | BUG-33278 | HBASE-13239 | HBASE grants at specific column level does not work for Groups |
| HBase | BUG-36565 | HBASE-13734 | deleteall behavior changes after applying org.apache.hadoop.hbase.security.visibility.VisibilityController to HBase |
| Hive | BUG-33338 | HIVE-10528 | Hiveserver2 in HTTP mode is not applying auth_to_local rules |
| Hive | BUG-35992 | HIVE-10875 | SELECT a.* FROM (SELECT * FROM source_view) a results in permission denied error. |
| Hue | BUG-28794 | | Security scan – The Hue web application is vulnerable to stored Cross-Site Scripting (XSS) |
| Hue | BUG-30961 | | Security scan – The Hue web application is vulnerable to stored Cross-Site Scripting (XSS) |
| Hue | BUG-32176 | | Django Vulnerabilities in Hue |
| Hue | BUG-32729 | | Directory listing is enabled for HUE |
| Hue | BUG-33163 | | Clear Text Password Shows In Hue UI for Oozie jobs and job log files. |
| Knox | BUG-33708 | KNOX-525 | persisted service registry is not updated to support HA after upgrade |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|--------------|--------------------|----------------------------|---|
| Knox, Ranger | BUG-36431 | | Found Jersey client API incompatibility that prevent HTTPS Knox from working |
| Oozie | BUG-33621 | | Clear Text Password Shows In Oozie workflow configuration |
| Ranger | BUG-30108 | RANGER-375 | Ranger - when backend DB not reachable error thrown is "incorrect password/username" when trying to log into Ranger |
| Ranger | BUG-31425 | | PolicyManager throws 404 error when creating HBase repo in Ranger |
| Ranger | BUG-33819 | RANGER-483 | Provide an option to create password using alternate message digest algorithm |
| YARN, HADOOP | BUG-41100 | | Yarn services Tomcat instance upgraded to 6.0.44 |

Incorrect Results

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|-----------------------------|---|
| HCatalog | BUG-22370 | HIVE-5545 | HCatRecord getInteger method returns String when used on Partition columns of type INT |
| Hive | BUG-31914 | HIVE-9278 | arithmetic operators return incorrect results from certain operand types |
| Hive | BUG-33275 | HIVE-8746 | ORC timestamp columns are sensitive to daylight savings time |
| Hive | BUG-34210 | HIVE-9278 | multiple built-in date functions return incorrect results in same where clause |
| Hive | BUG-34971 | HIVE-10481 | ACID table update finishes but values not really updated if column names are not all lower case |
| Hive | BUG-36223 | | wrong results for 2 left outer joins with overlapping join on keys, filter IS NULL pushed to left side table wrongly. |
| Hive | BUG-40100 | HIVE-11104 | Insert overwrite query does not return expected value |
| Hue | BUG-33631 | | Hue 2.6.1-2041 not displaying tables in alphabetical order |
| Phoenix | BUG-34506 | PHOENIX-896 | Unable to load tab delimited data via Phoenix |
| Pig | BUG-34138 | | Receiving exception 'No such file or directory' when using 'pig -useHCatalog' |
| Pig | BUG-36853 | PIG-4541 | Skewed full outer join does not return records if any |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|--------------------------|---|
| | | | relation is empty. Outer join does not return any record if left relation is empty |
| Pig | BUG-37485 | PIG-4556 | Pig tests abort with Check of prerequisites failed: <Failed running /usr/bin/pig -e fs -ls /> |

Stability

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|-----------------------------|--|
| Falcon | BUG-34020 | FALCON-1165 | Falcon will fail to start, if a cluster entity that was defined is not reachable. |
| HBase | BUG-21924 | HBASE-13555 | HBase web gives a 500 error when you attempt to view table details and it's not the master |
| HBase | BUG-29350 | HBASE-12791 | HBase does not attempt to clean up an aborted SPLIT when the Regions Server is shutting down. |
| HBase | BUG-33349 | HBASE-13608 | 500 Error with Stargate through Knox, using AD, SPNEGO, and Pre-Auth |
| HDFS | BUG-34175 | HDFS-8072 | Non-HDFS disk space is not reclaimed until after Datanode restart |
| HDFS | BUG-5700 | | NameNode should shut down if out of memory error occurs |
| Hue | BUG-30224 | | django.db.utils.DatabaseError: current transaction is aborted when setting up Hue on PostgreSQL database |
| Hue | BUG-33284 | | Hue - "/usr/lib/hue/build/env/bin/hue syncdb --noinput" in MySQL5.6 will result in an error |
| Knox | BUG-33488 | KNOX-530 | Running Oozie jobs through Knox on a cluster with HDFS HA does not use proper namenode host name. |
| Oozie | BUG-26984 | OOZIE-1728 | Queue configuration is not working for distcp action |
| Storm | BUG-28119 | STORM-563 | Kafka Spout doesn't pick up from the beginning of the queue unless forceFromStart specified |

Query Failure

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|----------------------------|---|
| HCatalog | BUG-30038 | | Sqoop Import on External Hcatalog Table |
| Hive | BUG-27636 | HIVE-10500 | intermittent acid_concurrency |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|----------------------------|---|
| | | | test failures due to NoSuchLockException |
| Hive | BUG-29427 | HIVE-9235 | Turn off Parquet Vectorization until all data types work: DECIMAL, DATE, TIMESTAMP, CHAR, and VARCHAR |
| Hive | BUG-30901 | HIVE-10559 | with constant propagation off, left join and join, hive.tez.dynamic.partition.pruning throw compile error IndexOutOfBoundsException |
| Hive | BUG-33145 | | HS2 HTTP Mode - Beeline Hanging after upgrading from HDP 2.1 to HDP 2.2 |
| Hive | BUG-33857 | HIVE-10273 | Select * on Table View (with UNION ALL) throws NullPointerException if Tez execution engine is used |
| Hive | BUG-34809 | HIVE-10242 | ACID: insert overwrite prevents create table command |
| Hive | BUG-34872 | HIVE-10559 | Hive tez dynamic pruning throws IndexOutOfBoundsException exception on certain queries |
| Hive | BUG-34956 | Hive-10677 | Analyze table compute stats for columns, ColumnStatsTask fail when hive.exec.parallel=true (default false) |
| Hive | BUG-35048 | HIVE-10483 | insert overwrite table with self join gets into a deadlock state in ACID DBTxnManager |
| Hive | BUG-35795 | HIVE-10286 | orc ppd does not typecast of string to timestamp when evaluate predicate for timestamp column |
| Hive | BUG-36111 | HIVE-8470 | Orc writer cant handle column of type void |
| Hive | BUG-37304 | HIVE-9950 | Hive query using Cuckoo hashing causing NullPointerException |
| Hive | BUG-37429 | HIVE-9937 | vectorization with ACID table, count(1) fails with ArrayIndexOutOfBoundsException |
| Hive | BUG-38292 | HIVE-10929 | In Tez mode,dynamic partitioning query with union all fails at moveTask,Invalid partition key & values |
| Hive | BUG-38817 | HIVE-11031 | Hive Alter table, concatenate partition files, throws error |
| Hive | BUG-39159 | HIVE-11027 | Hive Incorrect query results with Bucket Map joins vs Shuffle Joins. |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|----------------------------|---|
| Hive | BUG-39868 | HIVE-11051 | Tez Optimized MapJoin: Tables with Array<String> causes task failures |
| Hive | BUG-40036 | PIG-4624 | empty ORC file without schema produced by Hive |

Upgrade

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|---------------------------|--|
| HDFS | BUG-34508 | HDFS-8127 | NameNode Failover during HA upgrade can cause DataNode to finalize upgrade that will fail datanode -rollback |
| Hue | BUG-32999 | | After upgrading to HDP 2.2 from HDP 2.1. HUE now throws an error when trying to delete users |

Usability

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|---|--|
| Falcon | BUG-34630 | AMBARI-11140 | Falcon UI does not allow login by any user other than ambari-qa |
| Falcon | BUG-35141 | | Regression in the Falcon UI; Only 10 results shown for all entities. |
| Flume | BUG-35598 | FLUME-2095 | JMS source setup fails with "org.apache.flume. FlumeException: Could not lookup ConnectionFactory" |
| HBase | BUG-35338 | HBASE-13555 | HBase web gives a 500 error when you attempt to view Thread Stacks |
| HCatalog | BUG-30186 | HIVE-9381 | Cannot load data into specific Hive table via pig HCatStorer |
| Hive | BUG-23032 | | Jars added in Hive shell whilst in Tez mode are not recognized |
| Hive | BUG-31113 | | Hive "SLF4J: Class path contains multiple SLF4J bindings." error |
| Hive | BUG-31899 | HIVE-10271 | remove hive.server2.thrift.http.min/max.worker.threads properties |
| Hive | BUG-32859 | HIVE-9977 | Hive metastore auto major compaction not happening if data is loaded through dynamic partitioning insert |
| Hive | BUG-33876 | HIVE-10226 | ANALYZE TABLE (compute_stats) UDAF doesn't have Date support |
| Hive | BUG-35285 | HIVE-3682 , HIVE-5672 | cannot use separator of choice for Hive export in HDFS |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|------------------------------|--|
| Hive | BUG-35511 | | hive.support.concurrency=false doesn't turn of Transactions |
| Hive, Hue | BUG-29899 | HIVE-9223 | HS2/hue/tez: multiple concurrent queries not allowed in single tez session |
| Hue | BUG-32301 | | Not able to change Hue users passwords via the Hue UI |
| Hue | BUG-33632 | | Hue 2.6.1-2041 does not allow you to copy data displayed when doing a query |
| Hue | BUG-33693 | | Hue - Enable Parameterization doesn't work |
| Hue | BUG-35310 | | Not able to access/edit the workflow from Oozie Job Designer when the workflow is not in Page 1 |
| Hue | BUG-38607 | | Hue does not respect Support dfs.umaskmode, fs.permissions.umask-mode when creating files or folders |
| Pig | BUG-29616 | PIG-4381 | On PIG grunt shell DEFINE commands fails when it spans multiple lines. |
| Ranger | BUG-21763 | RANGER-451 | Policy Admin User/Group page doesn't show groups for users when the user belongs to large number of groups |
| Ranger | BUG-22120 | RANGER-244 | Option to disable/hide imported users/groups |
| Ranger | BUG-28060 | RANGER-281 | Support for PostgreSQL as DB server for XA Secure (Ranger, Argus) |
| Ranger | BUG-29900 | RANGER-318 | Not able to add user with only numbers |
| Ranger | BUG-31749 | RANGER-320 | Usersync NPE when object does not have userNameAttribute |
| Ranger | BUG-31782 | RANGER-337 | Ranger gives error when using hyphen or space in name field |
| Ranger | BUG-37467 | AMBARI-10924 | Ambari replaces ranger-admin install.properties with default after every start making it harder to debug |
| Spark | BUG-39567 | SPARK-8383 | Spark History Server shows Last Updated as 1969/12/31 when SparkPI application completed |
| WebHCat | BUG-7691 | | Provide/extend API to support table views |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|---------------------------|--|
| YARN | BUG-19508 | YARN-2246 | Job History Link in RM UI is redirecting to the URL which contains Job Id twice |
| YARN | BUG-27558 | YARN-3526 | ResourceManager probe on standby server in HA mode doesn't complete the redirect |
| YARN | BUG-9512 | YARN-2238 | Applications sometimes are not visible in RM web UI |

Performance

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|---|---|
| HDFS | BUG-38249 | HDFS-7435 | PB encoding of block reports is very inefficient |
| Hive | BUG-25165 | HIVE-9709 | Performance Issue Beeline > Knox > HS2 |
| Hive | BUG-33592 | HIVE-9644 , HIVE-9645 | case/when query not optimized, partition pruning not happening |
| Hive | BUG-37846 | | unix_timestamp non-deterministic causing less performance, current_timestamp not available in HDP 2.2.x |
| Hive, Tez | BUG-39394 | HIVE-10746 | Mapred.TextInputFormat defaults to 1 byte FileSplits |
| Hue | BUG-33579 | | Character Limit in Group Names unable to sync more than 30 characters. |
| Pig | BUG-33853 | PIG-4488 | Pig on Tez only uses default queue |
| Ranger | BUG-21901 | RANGER-192 | XASecure loading groups for user hang if there are too many groups for the user |
| Ranger | BUG-34139 | | Ranger is running out of database connection when looking up user/groups |
| Tez | BUG-23856 | TEZ-14 | Support for speculative execution of slow tasks |

Other

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-------------|--------------------|---------------------------|---|
| HDFS | BUG-31287 | HDFS-8055 | NullPointerException when topology script is missing. |
| HDP / Stack | BUG-33764 | | hdp-select needs to handle user created directories in /usr/hdp |
| Hue | BUG-26357 | | Display Hue server details |
| Hue | BUG-29981 | | Support HiveServer2 HTTP/S transport in hue |
| Hue | BUG-30145 | | Support LDAP authentication via Hue to Hiveserver2 |

| Component | Hortonworks Bug ID | Apache JIRA | Summary |
|-----------|--------------------|-------------|---|
| Hue | BUG-31350 | | Hue UTF8 Filter Decoding |
| Hue | BUG-32315 | | Users cannot use the job browser to investigate FAILED jobs. It works for other job types. |
| Hue | BUG-33469 | | Hue - log empty for create and insert statements |
| Hue | BUG-34408 | | hue-plugins.x86_64 has hadoop dependency |
| Hue | BUG-35047 | | Can't see Tez job details from Hue job browser while Tez job is running |
| Hue | BUG-35847 | | Select has a 1024 fd limit. If the FD number goes above 1024 for some reason (long running process), then all apps relying on thrift will fail. |
| Hue | BUG-36121 | | In HUE, clicking on useradmin -> groups gives "NoReverseMatch at / useradmin/groups" if group name with URL encoding crossing 80 characters |
| Kafka | BUG-32049 | | No Kafka Documentation Available at docs.hortonworks.com |

1.6. Apache Patch Information

The following sections list patches in each HDP 2.3 component beyond what was fixed in the base version of the Apache component.

1.6.1. Apache Hadoop

HDP 2.3 provides Apache Hadoop 2.7.1 and the following Apache patches for Hadoop-core, HDFS, and YARN:

NEW FEATURES

- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-2571](#) RM to support YARN registry
- [YARN-3345](#) Add non-exclusive node label API.
- [YARN-3365](#) Enhanced NodeManager to support using the 'tc' tool via container-executor for outbound network traffic control.
- [YARN-1376](#) NM need to notify the log aggregation status to RM through heartbeat.
- [YARN-3348](#) Add a 'yarn top' tool to help understand cluster usage.

- [YARN-3347](#) Improve YARN log command to get AMContainer logs as well as running containers logs.
- [YARN-3443](#) Create a 'ResourceHandler' subsystem to ease addition of support for new resource types on the NM.
- [YARN-3361](#) CapacityScheduler side changes to support non-exclusive node labels.
- [YARN-3318](#) Create Initial OrderingPolicy Framework and FifoOrderingPolicy.
- [YARN-3326](#) Support RESTful API for getLabelsToNodes.
- [YARN-3354](#) Add node label expression in ContainerTokenIdentifier to support RM recovery.
- [YARN-1402](#) Update related Web UI and CLI with exposing client API to check log aggregation status.
- [YARN-3463](#) Integrate OrderingPolicy Framework with CapacityScheduler.
- [YARN-3410](#) YARN admin should be able to remove individual application records from RMStateStore.
- [YARN-3225](#) New parameter of CLI for decommissioning node gracefully in RMAdmin CLI.
- [YARN-3366](#) Enhanced NodeManager to support classifying/shaping outgoing network bandwidth traffic originating from YARN containers
- [YARN-3319](#) Implement a FairOrderingPolicy.
- [YARN-2498](#) Respect labels in preemption policy of capacity scheduler for inter-queue preemption.
- [YARN-2619](#) Added NodeManager support for disk io isolation through cgroups.
- [YARN-3448](#) Added a rolling time-to-live LevelDB timeline store implementation.
- [YARN-3541](#) Add version info on timeline service / generic history web UI and REST API.
- [YARN-3505](#) Node's Log Aggregation Report with SUCCEED should not be cached in RMApps.

IMPROVEMENTS

- [HADOOP-10597](#) RPC Server signals backoff to clients when all request queues are full.
- [YARN-1880](#) Cleanup TestApplicationClientProtocolOnHA
- [YARN-3243](#) CapacityScheduler should pass headroom from parent to children to make sure ParentQueue obeys its capacity limits.
- [YARN-3356](#) Capacity Scheduler FiCaSchedulerApp should use ResourceUsage to track used-resources-by-label.
- [YARN-2868](#) FairScheduler: Metric for latency to allocate first container for an application.

- [YARN-3397](#) yarn rmdadmin should skip -failover.
- [YARN-2495](#) Allow admin specify labels from each NM (Distributed configuration for node label).
- [YARN-3248](#) Display count of nodes blacklisted by apps in the web UI.
- [YARN-2901](#) Add errors and warning metrics page to RM, NM web UI.
- [YARN-3294](#) Allow dumping of Capacity Scheduler debug logs via web UI for a fixed time period.
- [YARN-3293](#) Track and display capacity scheduler health metrics in web UI.
- [YARN-3394](#) Enrich WebApplication proxy documentation.
- [YARN-3404](#) Display queue name on application page.
- [YARN-2696](#) Queue sorting in CapacityScheduler should consider node label.
- [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-3406](#) Display count of running containers in the RM's Web UI.
- [YARN-3593](#) Add label-type and Improve "DEFAULT_PARTITION" in Node Labels Page.
- [YARN-3362](#) Add node label usage in RM CapacityScheduler web UI.
- [YARN-3565](#) NodeHeartbeatRequest/RegisterNodeManagerRequest should use NodeLabel object instead of String.
- [YARN-3583](#) Support of NodeLabel object instead of plain String in YarnClient side.
- [YARN-3581](#) Deprecate -directlyAccessNodeLabelStore in RMAdminCLI.
- [YARN-3700](#) Made generic history service load a number of latest applications according to the parameter or the configuration.

BUG FIXES

- [HDFS-27](#) HDFS CLI with --config set to default config complains log file not found error.
- [HDFS-7890](#) Improve information on Top users for metrics in RollingWindowsManager and lower log level.
- [HDFS-8229](#) LAZY_PERSIST file gets deleted after NameNode restart.
- [HDFS-8276](#) LazyPersistFileScrubber should be disabled if scrubber interval configured zero.

- [HDFS-8152](#) Refactoring of lazy persist storage cases.
- [HDFS-8144](#) Split TestLazyPersistFiles into multiple tests.
- [HDFS-8219](#) setStoragePolicy with folder behavior is different after cluster restart.
- [HDFS-8232](#) Missing datanode counters when using Metrics2 sink interface.
- [HDFS-8205](#) CommandFormat#parse() should not parse option as value of option.
- [HDFS-8211](#) DataNode UUID is always null in the JMX counter.
- [HDFS-7990](#) IBR delete ack should not be delayed.
- [HDFS-7645](#) Fix CHANGES.txt
- [HDFS-7645](#) Rolling upgrade is restoring blocks from trash multiple times
- [HDFS-8055](#) NullPointerException when topology script is missing.
- [HDFS-7933](#) fsck should also report decommissioning replicas.
- [HDFS-6666](#) Abort NameNode and DataNode startup if security is enabled but block access token is not enabled.
- [HADOOP-11859](#) PseudoAuthenticationHandler fails with httpcomponents v4.4.
- [HDFS-7701](#) Support reporting per storage type quota and usage with hadoop/hdfs shell.
- [HADOOP-7713](#) dfs -count -q should label output column
- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-3305](#) Normalize AM resource request on app submission.
- [YARN-3269](#) Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.
- [YARN-3383](#) AdminService should use "warn" instead of "info" to log exception when operation fails.
- [YARN-3425](#) NPE from RMNodeLabelsManager.serviceStop when NodeLabelsManager.serviceInit failed.
- [YARN-3435](#) AM container to be allocated Appattempt AM container shown as null.
- [YARN-2666](#) TestFairScheduler.testContinuousScheduling fails Intermittently.
- [YARN-3110](#) Few issues in ApplicationHistory web ui.
- [YARN-3459](#) Fix failiure of TestLog4jWarningErrorMetricsAppender.
- [YARN-3266](#) RMContext#inactiveNodes should have NodeId as map key.

- [YARN-3136](#) Fixed a synchronization problem of AbstractYarnScheduler#getTransferredContainers.
- [YARN-3387](#) Previous AM's container completed status couldn't pass to current AM if AM and RM restarted during the same time.
- [YARN-3530](#) ATS throws exception on trying to filter results without otherinfo.
- [YARN-2740](#) Fix NodeLabelsManager to properly handle node label modifications when distributed node label configuration enabled.
- [YARN-3517](#) RM web ui for dumping scheduler logs should be for admins only
- [YARN-3343](#) Increased TestCapacitySchedulerNodeLabelUpdate#testNodeUpdate timeout.
- [YARN-2821](#) Fixed a problem that DistributedShell AM may hang if restarted.
- [YARN-3654](#) ContainerLogsPage web UI should not have meta-refresh.
- [YARN-3552](#) RM Web UI shows -1 running containers for completed apps
- [YARN-3580](#) [JDK8] TestClientRMService.testGetLabelsToNodes fails.
- [YARN-3707](#) RM Web UI queue filter doesn't work.
- [YARN-3632](#) Ordering policy should be allowed to reorder an application when demand changes.
- [YARN-3740](#) Fixed the typo in the configuration name: APPLICATION_HISTORY_PREFIX_MAX_APPS.

1.6.2. Accumulo

HDP 2.3 provides 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 commercial 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 ConcurrentLinkedQueue.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 2.3 provides Atlas 0.5.0 and the following Apache patches:

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

1.6.4. Calcite

HDP 2.3 provides Calcite 1.2.0, with no additional Apache patches.

1.6.5. Falcon

HDP 2.3 provides Falcon 0.6.1 with no additional Apache patches.

1.6.6. Flume

HDP 2.3 provides 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-2455](#) Documentation update for Kafka Sink
- [FLUME-2454](#) Support batchSize to allow multiple events per transaction to the 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-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-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-2586](#) HDFS Sink should have an option to try rename even if close fails
- [FLUME-2665](#) Update documentation for hdfs.closeTries based on [FLUME-2586](#)
- [FLUME-2095](#) JMS source with TIBCO (patch-1)

BUG FIXES

- [FLUME-2451](#) HDFS Sink Cannot Reconnect After NameNode Restart
- [FLUME-2407](#) Spillable Channel sometimes fails on reconfigure
- [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-2412](#) Improve Logging in Spillable Channel
- [FLUME-2122](#) Minor cleanups of User guide
- [FLUME-2175](#) Update Developer Guide with notes on how to upgrade Protocol Buffer version

- [FLUME-2123](#) Morphline Solr sink missing short type name
- [FLUME-2162](#) TestHDFSEventSinkOnMiniCluster.maxUnderReplicationTest fails on hadoop2
- [FLUME-2501](#) Updating HttpClient lib version to ensure compat with Solr
- [FLUME-2530](#) Resource leaks found by Coverity tool
- [FLUME-2541](#) Bug in TestBucketWriter.testSequenceFileCloseRetries
- [FLUME-2441](#) Unit test TestHTTPSource.java failed with IBM JDK 1.7
- [FLUME-2520](#) HTTP Source should be able to block a prefixed set of protocols.
- [FLUME-2533](#) HTTPS tests fail on Java 6

1.6.7. HBase

HDP 2.3 provides 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 2.3 provides 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-11147](#) MetaTool doesn't update FS root location for partitions with space in name
- [HIVE-11104](#) Select operator doesn't propagate constants appearing in expressions
- [HIVE-11074](#) Update tests for [HIVE-9302](#) after removing binaries
- [HIVE-11051](#) Hive 1.2.0 MapJoin w/Tez - LazyBinaryArray cannot be cast to [Ljava.lang.Object;
- [HIVE-11083](#) Make test cbo_windowing robust

- [HIVE-10996](#) Aggregation / Projection over Multi-Join Inner Query producing incorrect results
- [HIVE-11076](#) Explicitly set `hive.cbo.enable=true` for some tests
- [HIVE-11060](#) Make test windowing.q robust
- [HIVE-11059](#) hcatalog-server-extensions tests scope should depend on hive-exec
- [HIVE-11066](#) Ensure tests don't share directories on FS
- [HIVE-11050](#) `testCliDriver_vector_outer_join.*` failures in Unit tests due to unstable data creation queries
- [HIVE-11048](#) Make test `cbo_windowing` robust
- [HIVE-11028](#) Tez: table self join and join with another table fails with `IndexOutOfBoundsException`
- [HIVE-10251](#) [HIVE-9664](#) makes hive depend on `ivysettings.xml` (using [HIVE-10251](#).simple.patch)

1.6.9. Kafka

HDP 2.3 provides Kafka 0.8.2 and the following Apache patches:

- [KAFKA-1688](#) Adding all public entities for adding a pluggable authorizer to kafka.
- [KAFKA-1683](#) add Session concept in `SocketServer.PlainTextTransportLayer` fixes.
- [KAFKA-1684](#) Kerberos/SASL implementation.
- [KAFKA-2118](#) Cleaner cannot clean after shutdown during `replaceSegments`.
- [KAFKA-2114](#) Unable to change `min.insync.replicas` default.
- [KAFKA-2128](#) `kafka.Kafka` should return non-zero exit code when caught exception.
- [KAFKA-2140](#) follow up, checking in newly renamed file `ConsumerRebalanceFailedException`.
- [KAFKA-2122](#) Remove `controller.message.queue.size` Config
- [KAFKA-2140](#) Improve code readability
- [KAFKA-2034](#) `sourceCompatibility` not set in Kafka build.gradle
- [KAFKA-2138](#) Fix producer to honor retry backoff
- [KAFKA-2121](#) Close internal modules upon client shutdown
- [KAFKA-1990](#) Add unlimited time-based log retention
- [KAFKA-2131](#) Update new producer javadocs with correct documentation links

- [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-1994](#) Evaluate performance effect of chroot check on Topic creation
- [KAFKA-2088](#) kafka-console-consumer.sh should not create zookeeper path when no brokers found and chroot was set in zookeeper.connect.
- [KAFKA-2119](#) ConsumerRecord key() and value() methods should not have throws Exception
- [KAFKA-2113](#) TestPurgatoryPerformance does not compile using IBM JDK
- [KAFKA-2056](#) Fix transient testRangePartitionAssignor failure
- [KAFKA-1416](#) Unify sendMessages in TestUtils
- [KAFKA-2117](#) Use the correct metadata field for reading offset struct
- [KAFKA-2115](#) Error updating metrics in RequestChannel
- [KAFKA-2112](#) make overflowWheel volatile
- [KAFKA-2090](#) Remove duplicate check to metadataFetchInProgress
- [KAFKA-2096](#) Enable keepalive socket option for broker to prevent socket leak
- [KAFKA-1989](#) New purgatory design; patched by Yasuhiro Matsuda
- [KAFKA-2109](#) Support retries in KafkaLog4jAppender
- [KAFKA-2104](#) testDuplicateListeners() has a typo
- [KAFKA-1517](#) Messages is a required argument to Producer Performance Test
- [KAFKA-1973](#) Remove the accidentally created LogCleanerManager.scala.orig
- [KAFKA-1910](#) Follow-up; Revert the no-offset-committed error code
- [KAFKA-1461](#) Implement per-partition back-off for replica fetcher
- [KAFKA-1992](#) checkEnoughReplicasReachOffset doesn't need to get requiredAcks
- [KAFKA-2033](#) Small typo in documentation
- [KAFKA-2099](#) BrokerEndPoint file, methods and object names should match
- [KAFKA-2043](#) CompressionType is passed in each RecordAccumulator append
- [KAFKA-1926](#) Replace kafka.utils.Utils with o.a.k.common.utils.Utils
- [KAFKA-1809](#) Refactor brokers to allow listening on multiple ports and IPs
- [KAFKA-1005](#) Shutdown consumer at the end of consumer performance test.

- [KAFKA-1996](#) Fix scaladoc error.
- [KAFKA-2024](#) Log compaction can generate unindexable segments.
- [KAFKA-2002](#) Mx4JLoader doesn't disable when kafka_mx4jenable=false.
- [KAFKA-2050](#) Avoid calling .size() on linked list.
- [KAFKA-1501](#) Let the OS choose the port in unit tests to avoid collisions
- [KAFKA-1546](#) Automate replica lag tuning;
- [KAFKA-1961](#) Prevent deletion of _consumer_offsets topic
- [KAFKA-2016](#) RollingBounceTest takes long
- [KAFKA-2013](#) benchmark test for the purgatory
- [KAFKA-2039](#) Update Scala to 2.10.5 and 2.11.6
- [KAFKA-2044](#) Support requests and responses from o.a.k.common in KafkaApis
- [KAFKA-1634](#) Bump up Offset Commit Request to v2 to add global retention and remove per-partition commit timestamp
- [KAFKA-527](#) Use in-place decompression enabled inner iterator to replace old decompress function
- [KAFKA-2047](#) Move the stream creation into concurrent mirror maker threads
- [KAFKA-527](#) Compression support does numerous byte copies;
- [KAFKA-2048](#) Change lock synchronized to inLock() for partitionMapCond
- [KAFKA-2042](#) Update topic list of the metadata regardless of cluster information;
- [KAFKA-1910](#) Fix two bugs on MemoryRecords and KafkaConsumer;
- [KAFKA-1997](#) Hopefully last follow-up fix to get messageHandlerArgs right
- [KAFKA-1997](#) Follow-up to add the shutdown hook before starting the consumers;
- [KAFKA-1863](#) Add docs for possible thrown exception in Callback;
- [KAFKA-1997](#) Refactor MirrorMaker based on [KIP-3](#);
- [KAFKA-1461](#) Replica fetcher thread does not implement any back-off behavior
- [KAFKA-1910](#) Refactor new consumer and fixed a bunch of corner cases / unit tests
- [KAFKA-2009](#) Fix two minor bugs in mirror maker.
- [KAFKA-1845](#) KafkaConfig should use ConfigDef
- [KAFKA-1988](#) Fix org.apache.kafka.common.utils.Utils.abs and add Partitioner.toPositive

- [KAFKA-1986](#) Request failure rate should not include invalid message size and offset out of range
- [KAFKA-1499](#) trivial follow-up (remove unnecessary parentheses)
- [KAFKA-1852](#) Reject offset commits to unknown topics
- [KAFKA-1755](#) Reject compressed and unkeyed messages sent to compacted topics
- [KAFKA-1865](#) Add a flush() method to the producer.
- [KAFKA-1824](#) ConsoleProducer - properties key.separator and parse.key no longer work
- [KAFKA-1866](#) LogStartOffset gauge throws exceptions after log.delete()
- [KAFKA-1664](#) Kafka does not properly parse multiple ZK nodes with non-root chroot

1.6.10. Knox

HDP 2.3 provides 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-561](#) Allow Knox pid directory to be configured via the Knox-env.sh file
- [KNOX-545](#) Simplify Keystore Management for Cluster Scaleout

1.6.11. Mahout

In HDP-2.3, 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 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 patch:

- [MAHOUT-1589](#) mahout.cmd has duplicated content

1.6.12. Oozie

HDP 2.3 provides Oozie 4.2.0 and the following Apache patches:

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

1.6.13. Phoenix

HDP 2.3 provides Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- [PHOENIX-2032](#) psql.py is broken after [PHOENIX-2013](#)
- [PHOENIX-2033](#) PQS log environment details on launch
- [PHOENIX-2007](#) java.sql.SQLException: Encountered exception in sub plan [0] execution'
- [PHOENIX-2027](#) Queries with Hints are raising IllegalStateException
- [PHOENIX-2012](#) RowKeyComparisonFilter logs unencoded data at DEBUG level
- [PHOENIX-2010](#) Properly validate number of arguments passed to the functions in FunctionParseNode#validate
- [PHOENIX-2013](#) Apply [PHOENIX-1995](#) to runnable uberjar as well
- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode (addendum)
- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode
- [PHOENIX-1996](#) Use ByteString instead of ZeroCopyByteString
- [PHOENIX-1995](#) client uberjar doesn't support dfs
- [PHOENIX-1980](#) CsvBulkLoad cannot load hbase-site.xml from classpath
- [PHOENIX-1976](#) Exit gracefully if addShutdownHook fails.
- [PHOENIX-914](#) Native HBase timestamp support to optimize date range queries in Phoenix

1.6.14. Pig

HDP 2.3 provides Pig 0.15.0 and the following Apache patch:

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

1.6.15. Ranger

HDP 2.3 provides 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

- [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 2.3 provides 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 2.3 provides 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 2.3 provides 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 2.3 provides Storm 0.10.0 and the following Apache patches:

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

1.6.20. Tez

HDP 2.3 provides Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [TEZ-2568](#) auto_sortmerge_join_5 fails in Tez mode
- [TEZ-2548](#) TezClient submitDAG can hang if the AM is in the process of shutting down.

- [TEZ-2475](#) Fix a potential hang in Tez local mode caused by incorrectly handled interrupts.
- [TEZ-2554](#) Tez UI: View log link does not correctly propagate login credential to read log from YARN web.
- [TEZ-2547](#) Tez UI: Download Data fails on secure, cross-origin clusters
- [TEZ-2546](#) Tez UI: Fetch hive query text from timeline if dagInfo is not set.
- [TEZ-2513](#) Tez UI: Allow filtering by DAG ID on All dags table.
- [TEZ-2541](#) DAGClientImpl enable TimelineClient check is wrong.
- [TEZ-2539](#) Tez UI: Pages are not updating in IE.
- [TEZ-2535](#) Tez UI: Failed task attempts link in vertex details page is broken.
- [TEZ-2489](#) Disable warn log for Timeline ACL error when tez.allow.disabled.timeline-domains set to true.
- [TEZ-2528](#) Tez UI: Column selector buttons gets clipped, and table scroll bar not visible in mac.
- [TEZ-2391](#) TestVertexImpl timing out at times on jenkins builds.
- [TEZ-2509](#) YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.
- [TEZ-2527](#) Tez UI: Application hangs on entering erroneous RegEx in counter table search box
- [TEZ-2523](#) Tez UI: derive applicationId from dag/vertex id instead of relying on json data
- [TEZ-2505](#) PipelinedSorter uses Comparator objects concurrently from multiple threads.
- [TEZ-2504](#) Tez UI: tables - show status column without scrolling, numeric 0 shown as Not available
- [TEZ-2478](#) Move OneToOne routing to store events in Tasks.
- [TEZ-2482](#) Tez UI: Mouse events not working on IE11
- [TEZ-1529](#) ATS and TezClient integration in secure kerberos enabled cluster.
- [TEZ-2481](#) Tez UI: graphical view does not render properly on IE11
- [TEZ-2474](#) The old taskNum is logged incorrectly when parallelism is changed
- [TEZ-2460](#) Temporary solution for issue due to [YARN-2560](#)
- [TEZ-2455](#) Tez UI: Dag view caching, error handling and minor layout changes
- [TEZ-2453](#) Tez UI: show the dagInfo is the application has set the same.
- [TEZ-2447](#) Tez UI: Generic changes based on feedbacks.

- [TEZ-2409](#) Allow different edges to have different routing plugins

1.7. Known Issues

Table 1.17. Apache HBase

| | |
|--------------------|---|
| Apache JIRA | |
| Hortonworks Bug ID | BUG-45664 |
| Description | Memory leak in Kafka Broker caused by leak in instance of ConcurrentHashMap/socketContainer |
| Apache JIRA | KAFKA-2012 |
| Hortonworks Bug ID | BUG-45688 |
| Description | kafka index file corruption |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-42355 |
| Description | <p>Moved application from HDP 2.2 to HDP 2.3 and now ACLs don't appear to be functioning the same</p> <p>Workaround: Set <code>hbase.security.access.early_out=false</code>, as in the following example:</p> <pre><property> <name>hbase.security.access.early_out</name> <value>false</value> </property></pre> |
| Apache JIRA | HBASE-13330 , HBASE-13647 |
| Hortonworks Bug ID | BUG-36817 |
| Description | test_IntegrationTestRegionReplica Replication[IntegrationTestRegion ReplicaReplication] fails with READ FAILURES |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-39322 |
| Description | <p>The HBase bulk load process is a MapReduce job that typically runs under the user ID who owns the source data. HBase data files created as a result of the job are then bulk-loaded into HBase RegionServers. During this process, HBase RegionServers move the bulk-loaded files from the user's directory, and moves (renames) the files under the HBase <code>root.dir (/apps/hbase/data)</code>. When HDFS data encryption is used, HDFS cannot rename across encryption zones with different keys.</p> <p>Workaround: Run the MapReduce job as the <code>hbase</code> user, and specify an output directory in the same encryption zone as the HBase root directory.</p> |
| Apache JIRA | HBASE-13832 , HDFS-8510 |
| Hortonworks Bug ID | BUG-40536 |
| Description | When rolling upgrade is performed for HDFS, sometimes the HBase Master might run out of datanodes on which to keep its write-pipeline active. When this occurs, the HBase |

| | |
|--------------------|---|
| | <p>Master Aborts after a few attempts to keep the pipeline going. To avoid this situation:</p> <p>Workaround:</p> <ol style="list-style-type: none"> 1. Before performing the rolling upgrade of HDFS, update the HBase configuration by setting "dfs.client.block.write.replace-datanode-on-failure.best.effort" to true. 2. Restart the HBase Master. 3. Perform the rolling upgrade of HDFS. <p>Undo the configuration change done in Step 1.</p> <p>Restart the HBase Master.</p> <p><i>Note:</i> There is a window of time during the rolling upgrade of HDFS when the HBase Master might be working with just one node and if that node fails, the WAL data might be lost. In practice, this is an extremely rare situation.</p> <p>Alternatively, the HBase Master can be turned off during the rolling upgrade of HDFS to avoid the above procedure. If this strategy is taken, client DDL operations and RegionServer failures cannot be handled during this time.</p> <p>A final alternative if the HBase Master fails during rolling upgrade of HDFS, a manual start can be performed.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-42186 |
| Description | <p>HDP 2.3 HBase install needs MapReduce class path modified for HBase functions to work</p> <p>Cluster that have Phoenix enabled placed the following config in hbase-site.xml:</p> <pre>Property: hbase.rpc.controllerfactory.class Value:org.apache.hadoop.hbase.ipc.controller.ServerRpcControllerFactory</pre> <p>This property points to a class found only in phoenix-server jar. To resolve this class at run time for the above listed Mapreduce Jobs, it needs to be part of the MapReduce classpath.</p> <p>Workaround: Update mapreduce.application.classpath property in mapred-site.xml file to point to /usr/hdp/current/phoenix-client/phoenix-server.jar file.</p> |

Table 1.18. Apache Hive

| | |
|--------------------|---|
| Apache JIRA | HIVE-11587 |
| Hortonworks Bug ID | BUG-42500 |
| Description | <p>Hive Hybrid Grace MapJoin can cause OutOfMemory Issues</p> <p>Hive Hybrid Grace Mapjoin is a new feature in HDP 2.3 (Hive 1.2). Mapjoin joins two tables, holding the smaller one in memory. Grace Hybrid Mapjoin spills parts of the small table to disk when the Map Join does not fit in memory at runtime. Right now there is a bug in the code that can cause this implementation to use too much</p> |

| | |
|--------------------|--|
| | <p>memory, causing an OutOfMemory error. This applies to the Tez execution engine only.</p> <p>Workaround: Turn off hybrid grace map join by setting this property in hive-site.xml:</p> <ul style="list-style-type: none"> • Navigate to Hive>Configs>Advanced>Custom hive-site. • Set <code>hive.mapjoin.hybridgrace.hashtable=false</code>. |
| Apache JIRA | HIVE-11110 |
| Hortonworks Bug ID | BUG-39988 |
| Description | CBO: Default partition filter is from MetaStore query causing TPC-DS to regress by 3x. |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-39412 |
| Description | <p>Users should not use <code>datanucleus.identifierFactory = datanucleus2</code> in hive config.</p> <p>Setting <code>datanucleus.identifierFactory</code> to <code>datanucleus2</code> can potentially lead to data corruption if <code>directSql</code> is enabled. Avoid using this setting if you are setting up a new metastore. If you are migrating an old metastore with this configuration parameter already set, contact Support for a few steps to address the issue.</p> |
| Apache JIRA | HIVE-10978 |
| Hortonworks Bug ID | BUG-39282 |
| Description | <p>When HDFS is encrypted (data at rest encryption is enabled) and the Hadoop Trash feature is enabled, DROP TABLE and DROP PARTITION have unexpected behavior.</p> <p>(The Hadoop Trash feature is enabled by setting <code>fs.trash.interval > 0</code> in <code>core-site.xml</code>.)</p> <p>When Trash is enabled, the data file for the table should be "moved" to the Trash bin, but if the table is inside an Encryption Zone, this "move" operation is not allowed.</p> <p>Workaround: Here are two ways to work around this issue:</p> <ol style="list-style-type: none"> 1. Use PURGE, as in DROP TABLE ... PURGE. This skips the Trash bin even if Trash is enabled. 2. set <code>fs.trash.interval = 0</code>. Caution: this configuration change must be done in <code>core-site.xml</code>. Setting it in <code>hive-site.xml</code> may lead to data corruption if a table with the same name is created later. |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-38785 |
| Description | <p>With RHEL7, the <code>cpu</code> and <code>cpuacct</code> controllers are managed together by default. The default directory is <code>/sys/fs/cgroup/cpu,cpuacct</code>. The presence of the comma leads to failures when initializing the NodeManager (when using the LinuxContainerExecutor).</p> |

| | |
|--------------------|--|
| | <p>Workaround: Create your own directory(such as <code>/sys/fs/cgroup/hadoop/cpu</code>) and set <code>yarn.nodemanager.linux-container-executor.cgroups.mount</code> to <code>true</code>. This will allow the NodeManager to mount the cpu controller, and YARN will be able to enforce CPU limits for you.</p> <p>If you wish to mount the <code>cgroups</code> yourself (or provide a mount point), please set <code>yarn.nodemanager.linux-container-executor.cgroups.mount</code> to <code>false</code> and ensure that the hierarchy specified in <code>yarn.nodemanager.linux-container-executor.cgroups.hierarchy</code> exists in the mount location. Make sure there are no commas in your pathnames.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-37042 |
| Description | <p>Limitations while using <code>timestamp.formats serde</code> parameter.</p> <p>Two issues involving the <code>timestamp.formats SerDe</code> parameter:</p> <ul style="list-style-type: none"> Displays only 3 decimal digits when it returns values, but it accepts more decimal digits. <p>For example, if you run the following commands:</p> <pre>drop table if exists src_hbase_ts; create table src_hbase_ts(rowkey string, ts1 string, ts2 string, ts3 string, ts4 string) STORED BY 'org.apache.hadoop.hive. hbase. HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4') TBLPROPERTIES ('hbase.table.name' = 'hbase_ts'); insert into src_hbase_ts values ('1','2011-01-01T01:01: 01.111111111', '2011-01-01T01:01: 01.123456111', '2011-01-01T01:01: 01.111111111', '2011-01-01T01:01: 01.134567890'); drop table if exists hbase_ts_1; create external table hbase_ts_1(rowkey string, ts1 timestamp, ts2 timestamp, ts3 timestamp, ts4 timestamp) STORED BY 'org.apache.hadoop.hive. hbase. HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4', 'timestamp.formats' = "yyyy-MM- dd'T'HH:mm:ss.SSSSSSSS") TBLPROPERTIES ('hbase.table.name' = 'hbase_ts'); select * from hbase_ts_1;</pre> <p>The <code>timestamp.formats</code> parameter displays:</p> <pre>1 2011-01-01 01:01:01.111 2011-01-01 01:01:01. 123 2011-01-01 01:01:01.111 2011-01-01 01:01:01. 134</pre> <p>When the expected output is:</p> |

| | |
|--|--|
| | <pre>1 2011-01-01 01:01:01.111111111 2011-01-01 01:01:01.123456111 2011-01-01 01:01:01. 111111111 2011-0</pre> <ul style="list-style-type: none"> The yyyy-MM-dd'T'HH:mm:ss.SSSSSSSS format accepts any timestamp data up to .SSSSSSSS decimal digits (9 places to the left of the decimal) instead of only reading data with .SSSSSSSS decimal digits (9 places to the left of the decimal). <p>For example, if you run the following commands:</p> <pre>drop table if exists src_hbase_ts; create table src_hbase_ts(rowkey string, ts1 string, ts2 string, ts3 string, ts4 string) STORED BY 'org.apache.hadoop. hive. hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4') TBLPROPERTIES ('hbase.table.name' = 'hbase_ts'); insert into src_hbase_ts values ('1','2011-01-01T01:01: 01.111111111', '2011-01-01T01:01: 01.111', '2011-01-01T01:01: 01.11', '2011-01-01T01:01:01.1'); drop table if exists hbase_ts_1; create external table hbase_ts_1(rowkey string, ts1 timestamp, ts2 timestamp, ts3 timestamp, ts4 timestamp) STORED BY 'org.apache.hadoop. hive. hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4', 'timestamp.formats' = "yyyy-MM- dd'T'HH:mm:ss.SSSSSSSS") TBLPROPERTIES ('hbase.table.name' = 'hbase_ts'); select * from hbase_ts_1;</pre> <p>The actual output is:</p> <pre>1 2011-01-01 01:01:01.111 2011-01-01 01:01:01. 111 2011-01-01 01:01:01.11 2011-01-01 01:01:01.1</pre> <p>When the expected output is:</p> <pre>1 2011-01-01 01:01:01.111 NULL NULL NULL</pre> |
|--|--|

Table 1.19. Apache Oozie

| | |
|--------------------|---|
| Apache JIRA | OOZIE-2311 |
| Hortonworks Bug ID | BUG-39265 |
| Description | NPE in oozie logs while running feed replication tests causes jobs to fail. |

Table 1.20. Apache Ranger

| | |
|--------------------|---|
| Apache JIRA | RANGER_577 |
| Hortonworks Bug ID | BUG-38054 |
| Description | Ranger should not change Hive config if authorization is disabled |

Table 1.21. Apache Slider

| | |
|--------------------|--|
| Apache JIRA | SLIDER-909 |
| Hortonworks Bug ID | BUG-40682 |
| Description | Slider HBase app package fails in secure cluster with wire-encryption on |

Table 1.22. Apache Spark

| | |
|--------------------|--|
| Apache JIRA | |
| Hortonworks Bug ID | BUG-41644, BUG-41484 |
| Description | <p>Apache and custom Spark builds need an HDP specific configuration. See the Troubleshooting Spark: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.0/bk_spark-quickstart/content/ch_troubleshooting-spark-quickstart.html section for more details.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-38046 |
| Description | <p>Spark ATS is missing Kill event</p> <p>If a running Spark application is killed in the YARN ATS (<code>yarn application -kill <appid></code>), the log will not list the outcome of the kill operation.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-39468 |
| Description | <p>When accessing an HDFS file from pyspark, the HADOOP_CONF_DIR environment must be set. For example:</p> <pre>export HADOOP_CONF_DIR=/etc/hadoop/conf [hrt_qa@ip-172-31-42-188 spark]\$ pyspark [hrt_qa@ip-172-31-42-188 spark]\$ >>>lines = sc.textFile("hdfs://ip-172-31-42-188.ec2.internal:8020/tmp/PySparkTest/file-01")</pre> <p>If HADOOP_CONF_DIR is not set properly, you might receive the following error:</p> <pre>Py4JJavaError: An error occurred while calling z:org.apache.spark.api.python.PythonRDD. collectAndServe. org.apache.hadoop.security.AccessControlException: SIMPLE authentication is not enabled. Available:[TOKEN, KERBEROS] at sun.reflect. NativeConstructorAccessorImpl.newInstance0(Native Method) at sun.reflect.NativeConstructorAccessorImpl. newInstance(NativeConstructorAccessorImpl.java:57)</pre> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-39674 |
| Description | <p>Spark does not yet support wire encryption, dynamic executor allocation, SparkR, GraphX, Spark Streaming, iPython, or Zeppelin.</p> |

Table 1.23. Apache Tez

| | |
|-------------|--|
| Apache JIRA | |
|-------------|--|

| | |
|--------------------|---|
| Hortonworks Bug ID | BUG-40608 |
| Description | <p>Tez UI View/Download link fails if URL does not match cookie.</p> <p>Workaround: Tez UI View/Download link will work if a browser accesses a URL that matches the cookie.</p> <p>Example: MapReduce JHS cookie is set with an external IP address. If a user clicks on the link from their internal cluster, the URL will differ and the request will fail with a <code>dr.who</code> error.</p> |

Table 1.24. Apache YARN

| | |
|--------------------|--|
| Apache JIRA | YARN-2194 |
| Hortonworks Bug ID | BUG-39424 |
| Description | NM fails to come with error "Not able to enforce cpu weights; cannot write to cgroup." |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-39756 |
| Description | NM web UI cuts <code>?user.name</code> when redirecting URL to MR JHS. |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-35942 |
| Description | <p>Users must manually configure ZooKeeper security with ResourceManager High Availability.</p> <p>Right now, the default value of <code>yarn.resourcemanager.zk-acl</code> is <code>world:any:rwcd,a</code>. That means anyone can read/write/create/delete/setPermission for the znode which is not secure and not acceptable.</p> <p>To make it more secure, we can rely on Kerberos to do the authentication for us. We could configure sasl authentication and only Kerberos authenticated user can access to <code>zkrmstatestore</code>.</p> <p>ZooKeeper Configuration</p> <p><i>Note:</i> This step of securing ZooKeeper is to be done <i>once</i> for the HDP cluster. If this has been done to secure HBase, for example, then you do not need to repeat these ZooKeeper steps if Apache YARN ResourceManager High Availability is to use the same ZooKeeper.</p> <ol style="list-style-type: none"> 1. Create a keytab for zookeeper called <code>zookeeper.service.keytab</code> and save it in <code>/etc/security/keytabs</code>. 2. Add following contents in <code>zoo.cfg</code>: <pre> authProvider.1=org.apache.zookeeper.server.auth.SASLAuthenticationProvider jaasLoginRenew=3600000 kerberos.removeHostFromPrincipal=true kerberos.removeRealmFromPrincipal=true </pre> 3. Create <code>zookeeper_client_jaas.conf</code>: |

```
Client {
com.sun.security.auth.module.Krb5LoginModule
  required
useKeyTab=false
useTicketCache=true;
};
```

4. Create zookeeper_jaas.conf:

```
Server {
com.sun.security.auth.module.Krb5LoginModule
  required
useKeyTab=true
storeKey=true
useTicketCache=false
keyTab="$PATH_TO_ZOOKEEPER_KEYTAB"
(such as "/etc/security/keytabs/zookeeper.
service.keytab")
principal="zookeeper/$HOST";
(such as "zookeeper/xuan-sec-yarn-ha-2.
novalocal@SCL42.HORTONWORKS.COM";)
};
```

5. Add the following contents in zookeeper-env.sh:

```
export CLIENT_JVMFLAGS="-Djava.security.
auth.login.config=/etc/zookeeper/conf/
zookeeper_client_jaas.conf"
export SERVER_JVMFLAGS="-Xmx1024m -Djava.
security.auth.login.config=/etc/zookeeper/conf/
zookeeper_jaas.conf"
```

Apache YARN Configuration

The following applies to HDP 2.2 and HDP 2.3.

Note: All nodes which launched the ResourceManager (active / standby) should make these changes.

1. Create a new configuration file: yarn_jaas.conf under the directory that houses the Hadoop Core configurations - if this is /etc/hadoop/conf, then put in that directory.

```
Client {
com.sun.security.auth.module.Krb5LoginModule
  required
useKeyTab=true
storeKey=true
useTicketCache=false
keyTab="$PATH_TO_RM_KEYTAB"
(such as "/etc/security/keytabs/rm.service.
keytab")
principal="rm/$HOST";
(such as "rm/xuan-sec-yarn-ha-1.
novalocal@EXAMPLE.COM";)
};
```

2. Add a new property in yarn-site.xml. Assuming that ResourceManager logs in with a Kerberos principle of the form rm/_HOST@DOMAIN.COM.

```
<property>
  <name>yarn.resourcemanager.zk-acl</name>
  <value>sasl:rm:rwcd</value>
</property>
```

3. Add a new YARN_OPTS into yarn-env.sh, and make sure this YARN_OPTS will be picked up when we start ResourceManagers.

```
YARN_OPTS="$YARN_OPTS -Dzookeeper.sasl.client=
true -Dzookeeper.sasl.client.username=zookeeper
-Djava.security.auth.login.config=/etc/
hadoop/conf/yarn_jaas.conf -Dzookeeper.sasl.
clientconfig=Client"
```

| | |
|--|---|
| | HDFS Configuration Note: This applies to HDP 2.1, 2.2, and 2.3. 1. In <code>hdfs-site.xml</code> , set the following property, for security of ZooKeeper based fail-over controller, when NameNode HA is enabled: <pre> <property> <name>ha.zookeeper.acl</name> <value>sasl:nn:rwcd</value> </property> </pre> |
|--|---|

Table 1.25. HDFS and Cloud Deployment

| | |
|--------------------|--|
| Apache JIRA | HADOOP-11618 , HADOOP-12304 |
| Hortonworks Bug ID | BUG-42065 |
| Description | <p>HDP 2.3: Cannot set non HDFS FS as default. This prevents S3, WASB, and GCC from working.</p> <p>HDP cannot be configured to use an external file system as the default file system - such as Azure WASB, Amazon S3, Google Cloud Storage. The default file system is configured in <code>core-site.xml</code> using the <code>fs.defaultFS</code> property. Only HDFS can be configured as the default file system.</p> <p>These external file systems can be configured for access as an optional file system, just not as the default file system.</p> |

Table 1.26. Hue

| | |
|--------------------|---|
| Apache JIRA | |
| Hortonworks Bug ID | BUG-40761 |
| Description | <p>Hue is not supported in CentOS 7.</p> <p>Workaround: Deploy Hue on CentOS 6.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-41369 |
| Description | <p>Hue About Page may not display the correct version information.</p> |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-41366 |
| Description | <p>Hue by default is using <code>spawning</code> instead of <code>cherrypy</code>.</p> <p>Impact: May cause performance impact</p> <p>Steps to reproduce: Install Hue in a cluster. View the <code>hue.ini</code> file and see <code>use_cherrypy_server=false</code></p> <p>Workaround: Modify the <code>hue.ini</code> file in <code>/etc/hue/conf</code>. Change from <code>use_cherrypy_server=false</code> to <code>use_cherrypy_server=true</code></p> |

Table 1.27. Upgrade

| | |
|--------------------|---|
| Apache JIRA | HDFS-8782 |
| Hortonworks Bug ID | BUG-41215 |
| Description | Upgrade to block ID-based DN storage layout delays DN registration. |

| | |
|--------------------|---|
| | When upgrading from a pre-HDP-2.2 release, a DataNode with a lot of disks, or with blocks that have random block IDs, can take a long time (potentially hours). The DataNode will not register to the NameNode until it finishes upgrading the storage directory. |
| Apache JIRA | |
| Hortonworks Bug ID | BUG-32401 |
| Description | Rolling upgrade/downgrade should not be used if truncate is turned on. Workaround: Before starting a rolling upgrade or downgrade process, turn truncate off. |

1.8. Documentation Errata

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

1.8.1. 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 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
```

1.8.2. 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 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 |

| Property Name | Default | Description |
|----------------------|---------|---|
| | | 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 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 |

| Property Name | Default | Description |
|---------------------------|---------|--|
| | | 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) |
| %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 useLocalTimeStamp is 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.