

Hortonworks Data Platform

HDP-2.3.2 Release Notes

(September 4, 2015)

Hortonworks Data Platform: HDP-2.3.2 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 4.0 License.
<http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Table of Contents

1. HDP 2.3.2 Release Notes	1
1.1. New Features	2
1.2. Unsupported Features	3
1.2.1. Technical Preview Features	3
1.2.2. Community Features	3
1.3. Upgrading from HDP 2.3.x to HDP 2.3.2	4
1.3.1. Before you begin	5
1.3.2. Optional: Spark Manual Upgrade Procedure	5
1.3.3. Upgrade Procedure	6
1.3.4. Optional: Spark Manual Downgrade Procedure	9
1.4. Behavior Changes	9
1.5. Apache Patch Information	9
1.5.1. Hadoop	10
1.5.2. Accumulo	17
1.5.3. Atlas	18
1.5.4. Calcite	19
1.5.5. Falcon	19
1.5.6. Flume	23
1.5.7. HBase	25
1.5.8. Hive	25
1.5.9. Kafka	30
1.5.10. Knox	33
1.5.11. Mahout	34
1.5.12. Oozie	34
1.5.13. Phoenix	34
1.5.14. Pig	35
1.5.15. Ranger	36
1.5.16. Slider	38
1.5.17. Spark	39
1.5.18. Sqoop	40
1.5.19. Storm	40
1.5.20. Tez	40
1.6. Common Vulnerabilities and Exposures	43
1.7. Third-party Licenses	44
1.8. Fixed Issues	44
1.9. Known Issues	48
1.10. Documentation Errata	58
1.10.1. Flume: Kafka Sink	58
1.10.2. Hive Sink	59
1.10.3. Configuring Pig Scripts to Use HCatalog in Oozie Workflows	63
1.10.4. Configuring a Sqoop Action to Use Tez to Load Data into a Hive Table	64

List of Tables

- 1.1. Technical Previews 3
- 1.2. Community Features 4
- 1.3. Tez, YARN 9

1. HDP 2.3.2 Release Notes

This document provides you with the latest information about the HDP 2.3.2 release and its product documentation.

Component Versions

The official Apache versions of most HDP 2.3.2 components are unchanged from HDP 2.3.0.0, with the exception of Spark. Spark is upgraded from 1.3.1 to 1.4.1. See more details of Spark 1.4.1 in the [New Features](#) section. All HDP 2.2 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 2.3.2.

Official Apache versions for HDP 2.3.2.

- Apache Accumulo 1.7.0
- Apache Atlas 0.5.0
- Apache Calcite 1.2.0
- Apache DataFu 1.3.0
- Apache Falcon 0.6.1
- Apache Flume 1.5.2
- Apache Hadoop 2.7.1
- Apache HBase 1.1.2
- Apache Hive 1.2.1
- Apache Kafka 0.8.2
- 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 Solr 5.5.0
- Apache Spark 1.4.1
- Apache Sqoop 1.4.6
- Apache Storm 0.10.0-beta
- Apache Tez 0.7.0
- Apache ZooKeeper 3.4.6

Additional component versions:

- Cascading 3.0.1
- Cloudbreak 1.0
- Hue 2.6.1

1.1. New Features

This section highlights several new features in HDP 2.3.2.

Spark 1.4.1 is the default Spark version with HDP 2.3.2. With a new HDP 2.3.2 cluster install, Spark 1.4.1 is installed. With the upgrade of an existing HDP cluster to 2.3.2 using Ambari, the Spark component is automatically upgraded to 1.4.1. For clusters not managed with Ambari, Spark can be manually upgraded from 1.2.1 or 1.3.1 to Spark 1.4.1; see the [Optional Spark Manual Upgrade Procedure](#) for instructions.

If you have upgraded to Spark 1.4.1 and want to downgrade to Spark follow the [Optional Spark Manual Downgrade Procedure](#). The Spark downgrade is only available as a manual step.

- HBase 1.1.2 is now available as a certified component.
- Spark 1.4.1 is now available as a certified component.
 - ORC File Support
 - Spark Core
 - Spark History Server
 - Spark MLLib
 - Spark on YARN
 - Spark on YARN on Kerberos-enabled clusters
 - Support for Hive 0.13.1, including the collect_list UDF
- Support for Debian 6 and Debian 7.

- Support for Ubuntu 12.0.4 and Ubuntu 14.0.4.

1.2. Unsupported Features

Some features exist within HDP 2.3.2, but Hortonworks does not currently support these specific capabilities.

1.2.1. Technical Preview Features

The following features are available within HDP 2.3.2, 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 1.1. 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 Query Server (PHOENIX-971) • Phoenix-Spark Integration • RPC Throttling • Support for <code>init.d</code> scripts
Hive	<ul style="list-style-type: none"> • Hive ACID • Hive Streaming
Kafka	<ul style="list-style-type: none"> • SSL
Slider	<ul style="list-style-type: none"> • Support for Docker-based application packaging (SLIDER-780)
Spark	<ul style="list-style-type: none"> • DataFrame API (SPARK-5097) • Dynamic Executor Allocation • SparkSQL, programmatically with SQLContext (not supported with Thrift Server - JDBC/ODBC) • SparkSQL thrift JDBC/ODBC Server • Spark Streaming
YARN	<ul style="list-style-type: none"> • Add support for network I/O isolation/scheduling for containers (YARN-2140) • NodeManager: add cgroup support for disk I/O isolation (YARN-2619)

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 1.2. Community Features

Component	Feature
Cloudbreak	<ul style="list-style-type: none"> Hosted Cloudbreak (cloudbreak.sequenceiq.com)
Falcon	<ul style="list-style-type: none"> Prism Server User Recipes
HBase	<ul style="list-style-type: none"> HBase Column Family Encryption: use HDFS data at rest encryption instead Use of memcached as block cache is unsupported (HBASE-13170) ZooKeeper-less region assignment
HDFS	<ul style="list-style-type: none"> block-volume device choosing (HDFS-1804) NameNode Federation (HDFS-1052) viewFS (HADOOP-7257)
Kafka	<ul style="list-style-type: none"> Mirror Maker (not supported when Kafka security is active) New Consumer API
Knox	<ul style="list-style-type: none"> Storm REST APIs
Oozie	<ul style="list-style-type: none"> Spark action (OOZIE-1983)
Slider	<ul style="list-style-type: none"> Simplified Application Packaging
Spark	<ul style="list-style-type: none"> GraphX Spark Standalone
YARN	<ul style="list-style-type: none"> Fair Scheduler MapReduce Eclipse Plug-in MapReduce Uber AM

1.3. Upgrading from HDP 2.3.x to HDP 2.3.2

HDP 2.3.2 is a maintenance release of HDP 2.3.x. If you already have HDP 2.3.x installed, upgrading your cluster to HDP 2.3.2 means:

- Keeping the same configuration files you used for HDP 2.3.x
- Keeping the same data and metadata in the same location you used for HDP 2.3.x
- Installing any new components (added for the first time in HDP 2.3.2) side-by-side with existing components

The following table summarizes HDP 2.2.x-to-2.3.2 upgrade options:

Cluster Management	Supporting Doc	Notes
Cluster managed manually	Before you begin [5]	If you have an earlier version of HDP (such as HDP 2.0 or HDP 2.1), see the HDP 2.3.2 Manual Upgrade Guide
Cluster managed via Ambari 1.7.0	Before you begin [5]	
Cluster managed via Ambari 2.0	Use the Upgrading Ambari Guide	Ambari 2.0 supports rolling upgrade between HDP 2.2.x and HDP 2.3.2

Cluster Management	Supporting Doc	Notes
		When upgrading to HDP 2.3.2 using Ambari, Spark 1.3.1 will be automatically upgraded to 1.4.1. If you wish to return to using 1.3.1, use the Spark Manual Downgrade Procedure .
Cluster managed via Ambari 2.1	Use the Upgrading Ambari Guide	Ambari 2.1 supports rolling upgrade between HDP 2.3.x and HDP 2.3.2 When upgrading to HDP 2.3.2 using Ambari, Spark 1.3.1 will be automatically upgraded to 1.4.1. If you wish to return to using 1.3.1, use the Spark Manual Downgrade Procedure .

1.3.1. Before you begin

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation
- Think about whether you are going to upgrade using a [local repository](#) or a [remote repository](#)

1.3.2. Optional: Spark Manual Upgrade Procedure

(Optional) Upgrade Spark from 1.3.1 to 1.4.1. As root:

1. Stop Spark 1.3.1: `su - spark -c "/usr/hdp/current/spark-client/sbin/stop-history-server.sh"`.
2. Remove Spark 1.3.1: `yum erase "spark*"`.
3. Add the node where you want Spark 1.4.1 History Server to run:
 - a. `su - root`
 - b. `wget -nv http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos6/2.x/BUILDS/2.3.2.0-2950/hdpbn.repo -O /etc/yum.repos.d/Spark141TP.repo`
 - c. `yum install spark_2_3_2_0_2950-master -y`
 - d. To use Python: `yum install spark_2_3_2_0_2950-python`
 - e. `conf-select create-conf-dir --package spark --stack-version 2.3.2.0-2950 --conf-version 0`
 - f. `cp /etc/spark/2.3.0.0-2950/0/* /etc/spark/2.3.2.0-2950/0/`
 - g. `conf-select set-conf-dir --package spark --stack-version 2.3.2.0-2950 --conf-version 0`
 - h. `hdp-select set spark-client 2.3.2.0-2950`
 - i. `hdp-select set spark-historyserver 2.3.2.0-2950`

4. Validate the Spark installation. As user spark, run SparkPI example:

- a. `su - spark -c "cd /usr/hdp/current/spark-client"`
- b. `./bin/spark-submit --class org.apache.spark.examples.SparkPi --master yarn-client --num-executors 3 --driver-memory 512m --executor-memory 512m --executor-cores 1 lib/spark-examples*.jar 10`

5. Restart Spark on YARN in either yarn-cluster mode or yarn-client mode:

- **yarn-cluster mode:** `./usr/hdp/current/spark-client/bin/spark-submit --class path.to.your.Class --master yarn-cluster [options] <app jar> [app options]`
- **yarn-client mode:** `./usr/hdp/current/spark-client/bin/spark-shell --master yarn-client`

1.3.3. Upgrade Procedure

Note: When installing Ranger, follow the Manual Upgrade instructions for [Installing Ranger](#).

To upgrade your cluster from HDP 2.3.x to HDP 2.3.2:

1. **Download the appropriate HDP 2.3.2 hdp.repo file for your OS:**

Support for SLES 11 SP1 and Debian 5 has been dropped in HDP 2.3.2.

Operating System	Repository Location
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.3.2.0/hdp.list
Debian 7	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.3.2.0/hdp.list
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.2.0/hdp.repo
RHEL/CentOS/Oracle LINUX 7	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.2.0/hdp.repo
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.2.0/hdp.repo
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.3.2.0/hdp.list
Ubuntu 14	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.3.2.0/hdp.list

or

Download the HDP RPM single repository tarball. (For information on how to install the repositories, see the [local repository](#) instructions.)

Operating System	Tarball Location
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.3.2.0/HDP-2.3.2.0-debian6-deb.tar.gz

Operating System	Tarball Location
Debian 7	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.3.2.0/HDP-2.3.2.0-debian7-deb.tar.gz
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.2.0/HDP-2.3.2.0-centos6-rpm.tar.gz
RHEL/CentOS/Oracle LINUX 7	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3.2.0/HDP-2.3.2.0-centos7-rpm.tar.gz
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3.2.0/HDP-2.3.2.0-suse11sp3-rpm.tar.gz
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.3.2.0/HDP-2.3.2.0-ubuntu12-deb.tar.gz
Ubuntu 14	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.3.2.0/HDP-2.3.2.0-ubuntu14-deb.tar.gz

2. Run an update:

```
apt-get update
```

3. Install the HDP 2.3.2 bits:

Operating System	Commands
RHEL/CentOS/Oracle LINUX	<p>Install HDP 2.3.2 components on relevant nodes, according to the services that run on those hosts:</p> <pre>yum install "hadoop_2_3_2_0_2950*" "oozie_2_3_2_0_2950*" "pig_2_3_2_0_2950*" "sqoop_2_3_2_0_2950*" "zookeeper_2_3_2_0_2950*" "hbase_2_3_2_0_2950*" "hive_2_3_2_0_2950*" "tez_2_3_2_0_2950*" "storm_2_3_2_0_2950*" "falcon_2_3_2_0_2950*" "flume_2_3_2_0_2950*" "phoenix_2_3_2_0_2950*" "accumulo_2_3_2_0_2950*" "mahout_2_3_2_0_2950*"</pre>
SLES	<p>Install HDP 2.3.2 components on relevant nodes, according to the services that run on those hosts:</p> <pre>zypper install "hadoop_2_3_2_2950*" "oozie_2_3_2_2950*" "pig_2_3_2_2950*" "sqoop_2_3_2_2950*" "zookeeper_2_3_2_2950*" "hbase_2_3_2_2950*" "hive_2_3_2_2950*" "tez_2_3_2_2950*" "storm_2_3_2_2950*" "falcon_2_3_2_2950*" "flume_2_3_2_2950*" "phoenix_2_3_2_2950*" "accumulo_2_3_2_2950*" "mahout_2_3_2_2950*"</pre>
Ubuntu/Debian	<p>Install HDP 2.3.2 components on relevant nodes, according to the services that run on those hosts:</p> <pre>apt-get install "hadoop-2-3-2-2950*" "oozie-2-3-2-2950*" "pig-2-3-2-2950*" "sqoop-2-3-2-2950*" "zookeeper-2-3-2-2950*" "hbase-2-3-2-2950*" "hive-2-3-2-2950*" "tez-2-3-2-2950*" "storm-2-3-2-2950*" "falcon-2-3-2-2950*" "flume-2-3-2-2950*" "phoenix-2-3-2-2950*" "accumulo-2-3-2-2950*" "mahout-2-3-2-2950*"</pre>

4. Stop all HDP 2.3.x Services.

If you are managing your cluster manually, stop all HDP 2.3.x Services. See "Controlling HDP Services Manually" in the *HDP Reference Guide*.

If you are managing your cluster with Ambari 1.7.0, do the following:

- a. Open Ambari Web
- b. Browse to **Services**
- c. Use **Service Actions** to stop each service

For all services, switch the active version to HDP 2.3.2.

On each host in the cluster, use `hdp-select` to switch all services to the HDP 2.3.2 version:

```
hdp-select set all <hdp2.3.2 version>
```

5. (Ambari 1.7.0-managed clusters only) Complete the Stack Upgrade.

If you are managing your cluster with Ambari 1.7.0, update the repository Base URLs to use the HDP 2.3.2 repositories for HDP and HDP-UTILS:

- a. Open Ambari Web
- b. Browse to **Admin > Repositories**
- c. Edit the Base URLs

6. Start all HDP 2.3.2 services, in the following order:

a. ZooKeeper

```
su - zookeeper export ZOOCFGDIR=/usr/hdp/current/zookeeper-server/conf ; export ZOOCFG=zoo.cfg; source /usr/hdp/current/zookeeper-server/conf/zookeeper-env.sh ; /usr/hdp/current/zookeeper-server/bin/zkServer.sh start
```

b. (HA NameNode upgrade only) ZooKeeper Failover Controller Daemons

```
/usr/hdp/current/hadoop-hdfs-namenode/.. /hadoop/sbin/hadoop-daemon.sh start zkfc
```

c. (HA NameNode upgrade only) JournalNodes

```
su - hdfs /usr/hdp/current/hadoop-hdfs-journalnode/.. /hadoop/sbin/hadoop-daemon.sh start journalnode
```

d. HDFS NameNode(s)

Start the HDFS NameNode(s). Because there is no metadata schema update for this upgrade, start the NameNode(s) in normal mode:

```
su - hdfs /usr/hdp/current/hadoop-hdfs-namenode/.. /hadoop/sbin/hadoop-daemon.sh start namenode
```

e. Remaining Services

Start the rest of the HDP services. On each host in the cluster, start the services that are relevant to that cluster. To identify the start commands for all services, see "Controlling HDP Services Manually" in the *HDP Reference Guide*.

You now have an upgraded cluster. Ensure that your workloads run correctly on this upgraded cluster.

1.3.4. Optional: Spark Manual Downgrade Procedure

When upgrading to HDP 2.3.2 using Ambari, Spark 1.3.1 is automatically upgraded to 1.4.1. However, if you wish to return to using 1.3.1:

1. Remove Spark 1.4.1 from your HDP cluster using Ambari:

```
curl -u admin:admin -H "X-Requested-By: ambari" -X DELETE
```

```
http://<AMBARI_HOST>:8080/api/v1/clusters/<CLUSTER_NAME>/services/SPARK
```

2. Manually install Spark 1.3.1 with [HDP 2.3.0 Installing HDP Manually: Installing and Configuring Apache Spark](#).

1.4. Behavior Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.3.2, behavioral changes affect the following Hadoop components.

Table 1.3. Tez, YARN

Hortonworks Bug ID	Apache JIRA	Description
BUG-41435	YARN-4149	If the user sets the option "ALL" for the parameter <i>logFiles</i> , they will get all the logs for the container. In previous versions, the user would get the data from the log file named "ALL" if such a log file existed. No user action is required.
BUG-44387		Changed the value of <code>yarn.timeline-service.generic-application-history.save-non-am-container-meta-info</code> to <code>false</code> . This config defines if YARN will save container meta-info in the application history store.

1.5. Apache Patch Information

The following sections list patches in each HDP 2.3.2 component beyond what was fixed in the base version of the Apache component.

1.5.1. Hadoop

HDP 2.3.2 provides the following Apache patches:

NEW FEATURES

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

IMPROVEMENTS

- [HADOOP-10597](#) RPC Server signals backoff to clients when all request queues are full.
- [HADOOP-11960](#) Enable Azure-Storage Client Side logging.
- [HADOOP-12325](#) RPC Metrics: Add the ability track and log slow RPCs.
- [HADOOP-12358](#) Add -safely flag to rm to prompt when deleting many files.
- [HDFS-4185](#) Add a metric for number of active leases.
- [HDFS-4396](#) Add START_MSG/SHUTDOWN_MSG for ZKFC.
- [HDFS-6860](#) BlockStateChange logs are too noisy.
- [HDFS-7923](#) The DataNodes should rate-limit their full block reports byasking the NN on heartbeat messages.
- [HDFS-8046](#) Allow better control of getContentSummary.
- [HDFS-8180](#) AbstractFileSystem Implementation for WebHdfs.
- [HDFS-8278](#) When computing max-size-to-move in Balancer, count only the storage with remaining \geq default block size.
- [HDFS-8432](#) Introduce a minimum compatible layout version to allow downgrade in more rolling upgrade use cases.
- [HDFS-8435](#) Support CreateFlag in WebHDFS.
- [HDFS-8549](#) Abort the balancer if an upgrade is in progress.
- [HDFS-8797](#) WebHdfsFileSystem creates too many connections for pread.
- [HDFS-8818](#) Changes the global moveExecutor to per datanode executors and changes MAX_SIZE_TO_MOVE to be configurable.
- [HDFS-8824](#) Do not use small blocks for balancing the cluster.
- [HDFS-8826](#) In Balancer, add an option to specify the source node list so that balancer only selects blocks to move from those nodes.
- [HDFS-8883](#) NameNode Metrics: Add FSNameSystem lock Queue Length.
- [HDFS-8911](#) NameNode Metric Add Editlog counters as a JMX metric.
- [HDFS-8983](#) NameNode support for protected directories.

- [HDFS-8983](#) NameNode support for protected directories.
- [YARN-2513](#) Host framework UIs in YARN for use with the ATS.
- [YARN-3197](#) Confusing log generated by CapacityScheduler.
- [YARN-3357](#) Move TestFifoScheduler to FIFO package.
- [YARN-3360](#) Add JMX metrics to TimelineDataManager.
- [YARN-3579](#) CommonNodeLabelsManager should support NodeLabel instead of string label name when getting node-to-label/label-to-label mappings.
- [YARN-3978](#) Configurably turn off the saving of container info in Generic AHS.
- [YARN-4082](#) Container shouldn't be killed when node's label updated.
- [YARN-4101](#) RM should print alert messages if ZooKeeper and ResourceManager gets connection issue.
- [YARN-4149](#) yarn logs -am should provide an option to fetch all the log files.

BUG FIXES

- [HADOOP-11802](#) DomainSocketWatcher thread terminates sometimes after there is an I/O error during requestShortCircuitShm.
- [HADOOP-12052](#) IPC client downgrades all exception types to IOE, breaks callers trying to use them.
- [HADOOP-12073](#) Azure FileSystem PageBlobInputStream does not return -1 onEOF.
- [HADOOP-12095](#) org.apache.hadoop.fs.shell.TestCount fails.
- [HADOOP-12304](#) Applications using FileContext fail with the default filesystem configured to be wasb/s3/etc.
- [HADOOP-8151](#) Error handling in snappy decompressor throws invalidexceptions.
- [HDFS-6945](#) BlockManager should remove a block from excessReplicateMap and decrement ExcessBlocks metric when the block is removed.
- [HDFS-7608](#) hdfs dfsclient newConnectedPeer has nowrite timeout.
- [HDFS-7609](#) Avoid retry cache collision when Standby NameNode loading edits.
- [HDFS-8309](#) Skip unit test using DataNodeTestUtils#injectDataDirFailure() on Windows.
- [HDFS-8310](#) Fix TestCLI.testAll "help for find" on Windows.
- [HDFS-8311](#) DataStreamer.transfer() should timeout the socket InputStream.
- [HDFS-8384](#) Allow NN to startup if there are files having a lease but are not under construction.
- [HDFS-8431](#) hdfs crypto class not found in Windows.

- [HDFS-8539](#) Hdfs doesn't have class 'debug' in windows.
- [HDFS-8542](#) WebHDFS getHomeDirectory behavior does not match specification.
- [HDFS-8593](#) Calculation of effective layout version mishandles comparison to current layout version in storage.
- [HDFS-8767](#) RawLocalFileSystem.listStatus() returns null for UNIX pipefile.
- [HDFS-8850](#) VolumeScanner thread exits with exception if there is no blockpool to be scanned but there are suspicious blocks.
- [HDFS-8863](#) The remaining space check in BlockPlacementPolicyDefault is flawed.
- [HDFS-8879](#) Quota by storage type usage incorrectly initialized upon namenoderestart.
- [HDFS-8885](#) ByteRangInputSteam used in webhdfs does not overrideavailable().
- [HDFS-8932](#) NPE thrown in NameNode when try to get TotalSyncCount metricbefore editLogStream initialization.
- [HDFS-8939](#) Test(S)WebHdfsFileContextMainOperations failing on branch-2.
- [HDFS-8969](#) Clean up findbugs warnings for HDFS-8823 and HDFS-8932.
- [HDFS-8995](#) Flaw in registration bookkeeping can make DN die on reconnect.
- [HDFS-9009](#) Send metrics logs to NullAppender by default.
- [YARN-3413](#) Changed Nodelabel attributes (like exclusivity) to be settable only via addToClusterNodeLabelsbut not changeable at runtime.
- [YARN-3885](#) ProportionalCapacityPreemptionPolicy doesn't preempt if queue is more than 2 level.
- [YARN-3894](#) RM startup should fail for wrong CS xml NodeLabel capacity configuration.
- [YARN-3896](#) RMNode transitioned from RUNNING to REBOOTED because its response idhas not been reset synchronously.
- [YARN-3932](#) SchedulerApplicationAttempt#getResourceUsageReport and UserInfo should based on total-used-resources.
- [YARN-3971](#) Skip RMNodeLabelsManager#checkRemoveFromClusterNodeLabelsOfQueue on nodelabel recovery.
- [YARN-4087](#) Followup fixes after YARN-2019 regarding RM behavior when state-store error occurs.
- [YARN-4092](#) Fixed UI redirection to print useful messages when both RMs are in standby mode.

OPTIMIZATION

- [HADOOP-11772](#) RPC Invoker relies on static ClientCache which has synchronized(this) blocks.

- [HADOOP-12317](#) Applications fail on NM restart on some Linux distro because NM container recovery declares AM container as LOST.
- [HADOOP-7713](#) dfs -count -q should label output column.
- [HDFS-8856](#) Make LeaseManager#countPath O(1).
- [HDFS-8867](#) Enable optimized block reports.

HDP 2.3.0 provided the following Apache patches:

NEW FEATURES

- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-1376](#) NM need to notify the log aggregation status to RM through heartbeat.
- [YARN-1402](#) Update related Web UI and CLI with exposing client API to check log aggregation status.
- [YARN-2498](#) Respect labels in preemption policy of capacity scheduler for inter-queue preemption.
- [YARN-2571](#) RM to support YARN registry
- [YARN-2619](#) Added NodeManager support for disk IO isolation through cgroups.
- [YARN-3225](#) New parameter of CLI for decommissioning node gracefully in RMAdmin CLI.
- [YARN-3318](#) Create Initial OrderingPolicy Framework and FifoOrderingPolicy.
- [YARN-3319](#) Implement a FairOrderingPolicy.
- [YARN-3326](#) Support RESTful API for getLabelsToNodes.
- [YARN-3345](#) Add non-exclusive node label API.
- [YARN-3347](#) Improve YARN log command to get AMContainer logs as well as running containers logs.
- [YARN-3348](#) Add a 'yarn top' tool to help understand cluster usage.
- [YARN-3354](#) Add node label expression in ContainerTokenIdentifier to support RM recovery.
- [YARN-3361](#) CapacityScheduler side changes to support non-exclusive node labels.
- [YARN-3365](#) Enhanced NodeManager to support using the 'tc' tool via container-executor for outbound network traffic control.
- [YARN-3366](#) Enhanced NodeManager to support classifying/shaping outgoing network bandwidth traffic originating from YARN containers
- [YARN-3410](#) YARN admin should be able to remove individual application records from RMStateStore.

- [YARN-3443](#) Create a 'ResourceHandler' subsystem to ease addition of support for new resource types on the NM.
- [YARN-3448](#) Added a rolling time-to-live LevelDB timeline store implementation.
- [YARN-3463](#) Integrate OrderingPolicy Framework with CapacityScheduler.
- [YARN-3505](#) Node's Log Aggregation Report with SUCCEED should not be cached in RMApps.
- [YARN-3541](#) Add version info on timeline service / generic history web UI and REST API.

IMPROVEMENTS

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

- [YARN-3511](#) Add errors and warnings page to ATS.
- [YARN-3565](#) NodeHeartbeatRequest/RegisterNodeManagerRequest should use NodeLabel object instead of String.
- [YARN-3581](#) Deprecate -directlyAccessNodeLabelStore in RMAdminCLI.
- [YARN-3583](#) Support of NodeLabel object instead of plain String in YarnClient side.
- [YARN-3593](#) Add label-type and Improve "DEFAULT_PARTITION" in Node Labels Page.
- [YARN-3700](#) Made generic history service load a number of latest applications according to the parameter or the configuration.

BUG FIXES

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

- [HDFS-8276](#) LazyPersistFileScrubber should be disabled if scrubber interval configured zero.
- [YARN-2666](#) TestFairScheduler.testContinuousScheduling fails Intermittently.
- [YARN-2740](#) Fix NodeLabelsManager to properly handle node label modifications when distributed node label configuration enabled.
- [YARN-2821](#) Fixed a problem that DistributedShell AM may hang if restarted.
- [YARN-3110](#) Few issues in ApplicationHistory web UI.
- [YARN-3136](#) Fixed a synchronization problem of AbstractYarnScheduler#getTransferredContainers.
- [YARN-3266](#) RMContext#inactiveNodes should have NodeId as map key.
- [YARN-3269](#) Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.
- [YARN-3305](#) Normalize AM resource request on app submission.
- [YARN-3343](#) Increased TestCapacitySchedulerNodeLabelUpdate#testNodeUpdate timeout.
- [YARN-3383](#) AdminService should use "warn" instead of "info" to log exception when operation fails.
- [YARN-3387](#) Previous AM's container completed status couldn't pass to current AM if AM and RM restarted during the same time.
- [YARN-3425](#) NPE from RMNodeLabelsManager.serviceStop when NodeLabelsManager.serviceInit failed.
- [YARN-3435](#) AM container to be allocated Appattempt AM container shown as null.
- [YARN-3459](#) Fix failure of TestLog4jWarningErrorMetricsAppender.
- [YARN-3517](#) RM web UI for dumping scheduler logs should be for admins only
- [YARN-3530](#) ATS throws exception on trying to filter results without otherinfo.
- [YARN-3552](#) RM Web UI shows -1 running containers for completed apps
- [YARN-3580](#) [JDK8] TestClientRMService.testGetLabelsToNodes fails.
- [YARN-3632](#) Ordering policy should be allowed to reorder an application when demand changes.
- [YARN-3654](#) ContainerLogsPage web UI should not have meta-refresh.
- [YARN-3707](#) RM Web UI queue filter doesn't work.
- [YARN-3740](#) Fixed the typo in the configuration name: APPLICATION_HISTORY_PREFIX_MAX_APPS.

1.5.2. Accumulo

HDP 2.3.2 provides Accumulo 1.7.0 and the following Apache patches:

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

HDP 2.3.0 provided Accumulo 1.7.0 and the following Apache patches:

- [ACCUMULO-3809](#) Table problem report has bogus table name for user table
- [ACCUMULO-3810](#) RandomWalk test, MultiTable fails throwing java.lang.NullPointerException w/ Kerberos on
- [ACCUMULO-3812](#) T*ProxyIT classes need cleanup
- [ACCUMULO-3814](#) StandaloneAccumuloClusterControl doesn't set provided ACCUMULO_CONF_DIR on SetGoalState
- [ACCUMULO-3815](#) StandaloneClusterControl shouldn't use canonical paths
- [ACCUMULO-3816](#) rpc.sasl.qop not mentioned in Kerberos server-configuration user manual section
- [ACCUMULO-3821](#) CleanTmpIT fails on dfs.permission enabled HDFS instance
- [ACCUMULO-3822](#) ImportExportIT fails to write to export directory in HDFS due to permissions
- [ACCUMULO-3823](#) Support separate client and server ACCUMULO_CONF_DIRS for StandaloneCluster ITs
- [ACCUMULO-3826](#) User manual accidentally references commerical product
- [ACCUMULO-3827](#) Default store types for monitor SSL are broken
- [ACCUMULO-3828](#) SimpleProxyBase ITs failing due to constraint propagation
- [ACCUMULO-3834](#) ConstraintIT occasionally failing
- [ACCUMULO-3838](#) ReplicationIT.replicationEntriesPrecludeWalDeletion failed because it missed an expected WAL
- [ACCUMULO-3839](#) Nonsense error when configuring instance.volumes.replacements
- [ACCUMULO-3845](#) DurabilityIT failed

- [ACCUMULO-3846](#) Allow override of C++ compiler through Maven build
- [ACCUMULO-3847](#) StandaloneClusterControl needs to launch MR jobs locally
- [ACCUMULO-3849](#) Proxy sets incorrect primary for SASL server transport
- [ACCUMULO-3850](#) Improve logging in replication code path
- [ACCUMULO-3852](#) NPE in WorkMaker for non-existent table
- [ACCUMULO-3853](#) Contention around 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

1.5.3. Atlas

HDP 2.3.0 provided Atlas 0.5.0 and the following Apache patches:

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

1.5.4. Calcite

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

1.5.5. Falcon

HDP 2.3.2 provides Falcon 0.6.1 and the following Apache patches:

NEW FEATURES

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

IMPROVEMENTS

- [FALCON-1060](#) Handle transaction failures in Lineage.
- [FALCON-1147](#) Allow _ in the names for name value pair.
- [FALCON-1174](#) Ability to disable oozie dryrun while scheduling or updating the falcon entity.
- [FALCON-1186](#) Add filtering capability to result of instance summary.
- [FALCON-1204](#) Expose default configs for feed late data handling in runtime.properties.
- [FALCON-1317](#) Inconsistent JSON serialization.
- [FALCON-1322](#) Add prefix in runtime.properties.
- [FALCON-1324](#) Pagination API breaks backward compatibility.
- [FALCON-1359](#) Improve output format for Feed Instance Listing.
- [FALCON-1361](#) Default end date should be now.

- [FALCON-1368](#) Improve Falcon server restart time.
- [FALCON-1374](#) Remove the cap on numResults.
- [FALCON-1378](#) Falcon documentation lacks information on how to run Falcon on standalone Oozie/Hadoop setup.
- [FALCON-668](#) FeedReplicator improvement to include more DistCP options.
- [FALCON-676](#) Enable metrics for Titan.
- [FALCON-75](#) Falcon CLI for deleting entities should inform user if entity does not exist.

BUG FIXES

- [FALCON-1038](#) Log mover fails for map-reduce action.
- [FALCON-1101](#) Cluster submission in falcon does not create an owned-by edge.
- [FALCON-1104](#) Exception while adding process instance to graphdb when feed has partition expression.
- [FALCON-1121](#) Backend support for free-text entity search.
- [FALCON-1129](#) In a secure cluster, feed replication fails because of Authentication issues.
- [FALCON-1141](#) Reverse Lookup for feed in prism fails with BadRequest.
- [FALCON-1143](#) Correcting order of entities on reload.
- [FALCON-1144](#) Dynamic partitions not getting registered in Hcat.
- [FALCON-1146](#) feed retention policy deleted everything all the way up to the root.
- [FALCON-1153](#) Instance kill fails intermittently.
- [FALCON-1162](#) Cluster submit succeeds when staging HDFS dir does not have 777.
- [FALCON-1165](#) Falcon restart failed, if defined service in cluster entity is unreachable.
- [FALCON-1244](#) numResults query param in listInstances is ignored when start and end params are not specified.
- [FALCON-1252](#) The parameter "tagkey" should be "tagkeys" in EntityList and FalconCLI twiki.
- [FALCON-1260](#) Instance dependency API produces incorrect results.
- [FALCON-1268](#) Instance Dependency API failure message is not intuitive in distributed mode.
- [FALCON-1282](#) Incorrect hdfs servers property for feed replication in secured environment.
- [FALCON-1310](#) Falcon build fails with Oozie-4.2.0.

- [FALCON-1311](#) Instance dependency API produces inconsistent results in some scenarios.
- [FALCON-1312](#) Falcon post processing action should use Oozie prepared configuration.
- [FALCON-1323](#) Reverse lookup of feeds causes NPE.
- [FALCON-1325](#) Triage API on prism, for an instance at which a process does not exist sends incorrect message.
- [FALCON-1327](#) When using triage on a server for a process which does not exist on that server, a NullPointerException is encountered.
- [FALCON-1328](#) Error in Triage documentation.
- [FALCON-1329](#) Falcon's idempotent behaviour breaks in some cases.
- [FALCON-1344](#) EntityGraph returns null in list of dependent entities.
- [FALCON-1363](#) Fix retry policy example in documentation.
- [FALCON-1398](#) CrossEntityValidations contains incorrect validations.
- [FALCON-1399](#) Property for default number of results is not loaded dynamically.
- [FALCON-1409](#) Update API throws NullPointerException.
- [FALCON-1412](#) Process waits indefinitely and finally timed out even though missing dependencies are met.
- [FALCON-1487](#) In secure cluster setup Hcat process/feed scheduling or replication fails.
- [FALCON-954](#) Secure Kerberos setup: Falcon should periodically revalidate auth token.
- [FALCON-99](#) Adding late data to process doesn't create new coord.

HDP 2.3.0 provided Falcon 0.6.1 and the following Apache patches:

NEW FEATURES

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

IMPROVEMENTS

- [FALCON-1060](#) Handle transaction failures in Lineage
- [FALCON-1147](#) Allow _ in the names for name value pair
- [FALCON-1174](#) Ability to disable oozie dryrun while scheduling or updating the falcon entity
- [FALCON-1186](#) Add filtering capability to result of instance summary

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

BUG FIXES

- [FALCON-1101](#) Cluster submission in falcon does not create an owned-by edge
- [FALCON-1104](#) Exception while adding process instance to graphdb when feed has partition expression
- [FALCON-1121](#) Backend support for free-text entity search
- [FALCON-1129](#) In a secure cluster, feed replication fails because of Authentication issues
- [FALCON-1141](#) Reverse Lookup for feed in prism fails with BadRequest
- [FALCON-1143](#) Correcting order of entities on reload
- [FALCON-1144](#) Dynamic partitions not getting registered in Hcat
- [FALCON-1146](#) feed retention policy deleted everything all the way up to the root
- [FALCON-1153](#) Instance kill fails intermittently
- [FALCON-1162](#) Cluster submit succeeds when staging HDFS dir does not have 777
- [FALCON-1165](#) Falcon restart failed, if defined service in cluster entity is unreachable
- [FALCON-1244](#) numResults query param in listInstances is ignored when start and end params are not specified
- [FALCON-1252](#) The parameter "tagkey" should be "tagkeys" in EntityList and FalconCLI twiki
- [FALCON-1260](#) Instance dependency API produces incorrect results

- [FALCON-1268](#) Instance Dependency API failure message is not intuitive in distributed mode
- [FALCON-1282](#) Incorrect hdfs servers property for feed replication in secured environment
- [FALCON-1310](#) Falcon build fails with Oozie-4.2.0
- [FALCON-1311](#) Instance dependency API produces inconsistent results in some scenarios
- [FALCON-1312](#) Falcon post processing action should use Oozie prepared configuration
- [FALCON-1323](#) Reverse lookup of feeds causes NPE
- [FALCON-1325](#) Triage API on prism, for an instance at which a process does not exist sends incorrect message
- [FALCON-1327](#) When using triage on a server for a process which does not exist on that server, a NullPointerException is encountered
- [FALCON-1328](#) Error in Triage documentation
- [FALCON-1329](#) Falcon's idempotent behaviour breaks in some cases
- [FALCON-1344](#) EntityGraph returns null in list of dependent entities
- [FALCON-1363](#) Fix retry policy example in documentation
- [FALCON-1398](#) CrossEntityValidations contains incorrect validations
- [FALCON-1399](#) Property for default number of results is not loaded dynamically
- [FALCON-1409](#) Update API throws NullPointerException
- [FALCON-1412](#) Process waits indefinitely and finally timed out even though missing dependencies are met
- [FALCON-1487](#) In secure cluster setup Hcat process/feed scheduling or replication fails
- [FALCON-954](#) Secure Kerberos setup : Falcon should periodically revalidate auth token
- [FALCON-99](#) Adding late data to process doesn't create new coord
[FALCON-1038](#) Log mover fails for map-reduce action

HDP 2.3.0 provided Falcon 0.6.1 with no additional Apache patches.

1.5.6. Flume

HDP 2.3.0 provided Flume 1.5.2 and the following Apache patches:

NEW FEATURES

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

Kafka Sink (preview)

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

Kafka Source

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

IMPROVEMENTS

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

BUG FIXES

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

- [FLUME-2407](#) Spillable Channel sometimes fails on reconfigure
- [FLUME-2412](#) Improve Logging in Spillable Channel
- [FLUME-2441](#) Unit test TestHTTPSource.java failed with IBM JDK 1.7
- [FLUME-2451](#) HDFS Sink Cannot Reconnect After NameNode Restart
- [FLUME-2501](#) Updating HttpClient lib version to ensure compatibility with Solr
- [FLUME-2520](#) HTTP Source should be able to block a prefixed set of protocols.
- [FLUME-2530](#) Resource leaks found by Coverity tool
- [FLUME-2533](#) HTTPS tests fail on Java 6
- [FLUME-2541](#) Bug in TestBucketWriter.testSequenceFileCloseRetries

1.5.7. HBase

HDP 2.3.2 provides HBase 1.1.2 and the following Apache patches:

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

HDP 2.3.0 provided HBase 1.1.1 and the following Apache patches:

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

1.5.8. Hive

HDP 2.3.2 provides Hive 1.2.1 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [HIVE-10140](#): Window boundary is not compared correctly

- [HIVE-10453](#): Reverted
- [HIVE-10569](#): Hive CLI gets stuck when `hive.exec.parallel=true`; and some exception happens during `SessionState.start`
- [HIVE-10571](#): `HiveMetaStoreClient` should close existing thrift connection before its reconnect
- [HIVE-10620](#): `ZooKeeperHiveLock` overrides `equal()` method but not `hashCode()`
- [HIVE-10646](#): `ColumnValue` does not handle `NULL_TYPE`
- [HIVE-10651](#): ORC file footer cache should be bounded
- [HIVE-10698](#): query on view results fails with table not found error if view is created with subquery alias (CTE).
- [HIVE-10714](#): Bloom filter column names specification should be case insensitive
- [HIVE-10722](#): external table creation with `msck` in Hive can create unusable partition
- [HIVE-10726](#): Hive JDBC `setQueryTimeout` should not throw exception to make it work with `JMeter`
- [HIVE-10731](#): `NullPointerException` in `HiveParser.g`
- [HIVE-10732](#): Hive JDBC driver does not close operation for metadata queries
- [HIVE-10771](#): `"separatorChar"` has no effect in `"CREATE TABLE AS SELECT"` statement
- [HIVE-10781](#): `HadoopJobExecHelper` Leaks `RunningJobs`
- [HIVE-10790](#): orc write on viewFS throws exception
- [HIVE-10793](#): HIVE-11587 Hybrid Grace Hash Join: Don't allocate all hash table memory upfront
- [HIVE-10802](#): Table join query with some constant field in select fails
- [HIVE-10808](#): Inner join on Null throwing Cast Exception
- [HIVE-10835](#): Concurrency issues in JDBC driver
- [HIVE-10880](#): The bucket number is not respected in insert overwrite.
- [HIVE-10925](#): Non-static threadlocals in metastore code can potentially cause memory leak
- [HIVE-10963](#): Hive throws NPE rather than meaningful error message when window is missing
- [HIVE-10972](#): `DummyTxnManager` always locks the current database in shared mode, which is incorrect.
- [HIVE-11013](#): MiniTez `tez_join_hash` test on the branch fails with NPE (`initializeOp` not called?)

- [HIVE-11024](#): Error inserting a date value via parameter marker (PreparedStatement.setDate)
- [HIVE-11029](#): hadoop.proxyuser.mapr.groups does not work to restrict the groups that can be impersonated
- [HIVE-11054](#): Read error: Partition Varchar column cannot be cast to string
- [HIVE-11079](#): Fix qfile tests that fail on Windows due to CR/character escape differences
- [HIVE-11087](#): DbTxnManager exceptions should include txnid
- [HIVE-11090](#): ordering issues with windows unit test runs
- [HIVE-11095](#): SerDeUtils another bug ,when Text is reused
- [HIVE-11102](#): ReaderImpl: getColumnIndicesFromNames does not work for some cases
- [HIVE-11112](#): ISO-8859-1 text output has fragments of previous longer rows appended
- [HIVE-11135](#): Fix the Beeline set and save command in order to avoid the NullPointerException
- [HIVE-11151](#): Calcite transitive predicate inference rule should not transitively add not null filter on non-nullable input
- [HIVE-11152](#): Swapping join inputs in ASTConverter
- [HIVE-11157](#): Hive.get(HiveConf) returns same Hive object to different user sessions
- [HIVE-11171](#): Join reordering algorithm might introduce projects between joins
- [HIVE-11172](#): Vectorization wrong results for aggregate query with where clause without group by
- [HIVE-11174](#): Hive does not treat floating point signed zeros as equal (-0.0 should equal 0.0 according to IEEE floating point spec)
- [HIVE-11176](#): Caused by: java.lang.ClassCastException: org.apache.hadoop.hive.serde2.lazybinary.LazyBinaryStruct cannot be cast to [Ljava.lang.Object;
- [HIVE-11193](#): ConstantPropagateProcCtx should use a Set instead of a List to hold operators to be deleted
- [HIVE-11198](#): Fix load data query file format check for partitioned tables
- [HIVE-11203](#): Beeline force option doesn't force execution when errors occurred in a script.
- [HIVE-11211](#): Reset the fields in JoinStatsRule in StatsRulesProcFactory
- [HIVE-11216](#): UDF GenericUDFMapKeys throws NPE when a null map value is passed in
- [HIVE-11221](#): In Tez mode, alter table concatenate orc files can intermittently fail with NPE

- [HIVE-11255](#): `get_table_objects_by_name()` in `HiveMetaStore.java` needs to retrieve table objects in multiple batches
- [HIVE-11258](#): The function `drop_database_core()` of `HiveMetaStore.java` may not drop all the tables
- [HIVE-11271](#): `java.lang.IndexOutOfBoundsException` when union all with if function
- [HIVE-11301](#): thrift metastore issue when getting stats results in disconnect
- [HIVE-11303](#): Getting Tez `LimitExceededException` after dag execution on large query
- [HIVE-11317](#): ACID: Improve transaction Abort logic due to timeout
- [HIVE-11320](#): ACID enable predicate pushdown for insert-only delta file
- [HIVE-11344](#): HIVE-9845 makes `HCatSplit.write` modify the split so that `PartInfo` objects are unusable after it
- [HIVE-11356](#): SMB join on tez fails when one of the tables is empty
- [HIVE-11357](#): ACID enable predicate pushdown for insert-only delta file 2
- [HIVE-11375](#): Broken processing of queries containing NOT (`x IS NOT NULL` and `x 0`)
- [HIVE-11407](#): JDBC `DatabaseMetaData.getTables` with large no of tables call leads to HS2 OOM
- [HIVE-11429](#): Increase default JDBC result set fetch size (# rows it fetches in one RPC call) to 1000 from 50
- [HIVE-11433](#): NPE for a multiple inner join query
- [HIVE-11442](#): Remove `commons-configuration.jar` from Hive distribution
- [HIVE-11449](#): HIVE-11587 "Capacity must be a power of two" error when `HybridHashTableContainer` memory threshold is too low
- [HIVE-11456](#): `HCatStorer` should honor `mapreduce.output.basename`
- [HIVE-11467](#): HIVE-11587 `WriteBuffers` rounding `wbSize` to next power of 2 may cause OOM
- [HIVE-11493](#): Predicate with integer column equals double evaluates to false
- [HIVE-11502](#): Map side aggregation is extremely slow
- [HIVE-11581](#): `HiveServer2` should store connection params in ZK when using dynamic service discovery for simpler client connection string.
- [HIVE-11587](#): Fix memory estimates for mapjoin hashtable
- [HIVE-11592](#): ORC metadata section can sometimes exceed protobuf message size limit
- [HIVE-11600](#): Hive Parser to Support multi col in clause (`x,y..`) in (`(..),..., ()`)

- [HIVE-11605](#): Incorrect results with bucket map join in tez.
- [HIVE-11606](#): Bucket map joins fail at hash table construction time
- [HIVE-11607](#): Export tables broken for data > 32 MB
- [HIVE-11658](#): Load data file format validation does not work with directories
- [HIVE-11727](#): (BUG-44285). Hive on Tez through Oozie: Some queries fail with fnf exception
- [HIVE-11755](#): Incorrect method called with Kerberos enabled in AccumuloStorageHandler
- [HIVE-11820](#): export tables with size of >32MB throws "java.lang.IllegalArgumentException: Skip CRC is valid only with update options"
- [HIVE-11836](#): ORC SARG creation throws NPE for null constants with void type
- [HIVE-11839](#): Vectorization wrong results with filter of (CAST AS CHAR)
- [HIVE-11849](#): NPE in HiveHBaseTableShapshotInputFormat in query with just count(*)
- [HIVE-11852](#): numRows and rawDataSize table properties are not replicated
- [HIVE-11875](#): JDBC Driver does not honor delegation token mechanism when readings params from ZooKeeper
- [HIVE-11897](#): JDO rollback can throw pointless exceptions
- [HIVE-11928](#): ORC footer section can also exceed protobuf message limit
- [HIVE-11936](#): Support SQLAnywhere as a backing DB for the hive metastore
- [HIVE-5277](#): HBase handler skips rows with null valued first cells when only row key is selected
- [HIVE-6727](#): Table level stats for external tables are set incorrectly
- [HIVE-7476](#): CTAS does not work properly for s3
- [HIVE-8529](#): HiveSessionImpl#fetchResults should not try to fetch operation log when hive.server2.logging.operation.enabled is false.
- [HIVE-9566](#): HiveServer2 fails to start with NullPointerException
- [HIVE-9625](#): Delegation tokens for HMS are not renewed
- [HIVE-9811](#): Hive on Tez leaks WorkMap objects
- [HIVE-9974](#): Sensitive data redaction: data appears in name of mapreduce job

HDP 2.3.0 provided Hive 1.2.1 and the following Apache patches:

INCOMPATIBLE CHANGES

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

NEW FEATURES

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

IMPROVEMENTS

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

BUG FIXES

- [HIVE-10251](#) [HIVE-9664](#) makes hive depend on ivysettings.xml (using [HIVE-10251](#).simple.patch)
- [HIVE-10996](#) Aggregation / Projection over Multi-Join Inner Query producing incorrect results
- [HIVE-11028](#) Tez: table self join and join with another table fails with IndexOutOfBoundsException
- [HIVE-11048](#) Make test cbo_windowing robust
- [HIVE-11050](#) testCliDriver_vector_outer_join.* failures in Unit tests due to unstable data creation queries
- [HIVE-11051](#) Hive 1.2.0 MapJoin w/Tez - LazyBinaryArray cannot be cast to [Ljava.lang.Object;
- [HIVE-11059](#) hcatalog-server-extensions tests scope should depend on hive-exec
- [HIVE-11060](#) Make test windowing.q robust
- [HIVE-11066](#) Ensure tests don't share directories on FS
- [HIVE-11074](#) Update tests for [HIVE-9302](#) after removing binaries
- [HIVE-11076](#) Explicitly set hive.cbo.enable=true for some tests
- [HIVE-11083](#) Make test cbo_windowing robust
- [HIVE-11104](#) Select operator doesn't propagate constants appearing in expressions
- [HIVE-11147](#) MetaTool doesn't update FS root location for partitions with space in name

1.5.9. Kafka

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

HDP 2.3.0 provided Kafka 0.8.2 and the following Apache patches:

- [KAFKA-1005](#) Shutdown consumer at the end of consumer performance test.
- [KAFKA-1416](#) Unify sendMessages in TestUtils
- [KAFKA-1461](#) Implement per-partition back-off for replica fetcher

- [KAFKA-1461](#) Replica fetcher thread does not implement any back-off behavior
- [KAFKA-1499](#) trivial follow-up (remove unnecessary parentheses)
- [KAFKA-1501](#) Let the OS choose the port in unit tests to avoid collisions
- [KAFKA-1517](#) Messages is a required argument to Producer Performance Test
- [KAFKA-1546](#) Automate replica lag tuning;
- [KAFKA-1634](#) Bump up Offset Commit Request to v2 to add global retention and remove per-partition commit timestamp
- [KAFKA-1664](#) Kafka does not properly parse multiple ZK nodes with non-root chroot
- [KAFKA-1683](#) add Session concept in SocketServer.PlainTextTransportLayer fixes.
- [KAFKA-1684](#) Kerberos/SASL implementation.
- [KAFKA-1688](#) Adding all public entities for adding a pluggable authorizer to Kafka.
- [KAFKA-1755](#) Reject compressed and unkeyed messages sent to compacted topics
- [KAFKA-1809](#) Refactor brokers to allow listening on multiple ports and IPs
- [KAFKA-1824](#) ConsoleProducer - properties key.separator and parse.key no longer work
- [KAFKA-1845](#) KafkaConfig should use ConfigDef
- [KAFKA-1852](#) Reject offset commits to unknown topics
- [KAFKA-1863](#) Add docs for possible thrown exception in Callback;
- [KAFKA-1865](#) Add a flush() method to the producer.
- [KAFKA-1866](#) LogStartOffset gauge throws exceptions after log.delete()
- [KAFKA-1910](#) Fix two bugs on MemoryRecords and KafkaConsumer;
- [KAFKA-1910](#) Follow-up; Revert the no-offset-committed error code
- [KAFKA-1910](#) Refactor new consumer and fixed a bunch of corner cases / unit tests
- [KAFKA-1926](#) Replace kafka.utils.Utils with o.a.k.common.utils.Utils
- [KAFKA-1961](#) Prevent deletion of _consumer_offsets topic
- [KAFKA-1973](#) Remove the accidentally created LogCleanerManager.scala.orig
- [KAFKA-1982](#) (add missing files) change kafka.examples.Producer to use the new java producer
- [KAFKA-1982](#) change kafka.examples.Producer to use the new java producer
- [KAFKA-1986](#) Request failure rate should not include invalid message size and offset out of range

- [KAFKA-1988](#) Fix org.apache.kafka.common.utils.Utils.abs and add Partitioner.toPositive
- [KAFKA-1989](#) New purgatory design; patched by Yasuhiro Matsuda
- [KAFKA-1990](#) Add unlimited time-based log retention
- [KAFKA-1992](#) checkEnoughReplicasReachOffset doesn't need to get requiredAcks
- [KAFKA-1994](#) Evaluate performance effect of chroot check on Topic creation
- [KAFKA-1996](#) Fix scaladoc error.
- [KAFKA-1997](#) Follow-up to add the shutdown hook before starting the consumers;
- [KAFKA-1997](#) Hopefully last follow-up fix to get messageHandlerArgs right
- [KAFKA-1997](#) Refactor MirrorMaker based on [KIP-3](#);
- [KAFKA-2002](#) Mx4JLoader doesn't disable when kafka_mx4jenable=false.
- [KAFKA-2009](#) Fix two minor bugs in mirror maker.
- [KAFKA-2013](#) benchmark test for the purgatory
- [KAFKA-2016](#) RollingBounceTest takes long
- [KAFKA-2024](#) Log compaction can generate unindexable segments.
- [KAFKA-2033](#) Small typo in documentation
- [KAFKA-2034](#) sourceCompatibility not set in Kafka build.gradle
- [KAFKA-2039](#) Update Scala to 2.10.5 and 2.11.6
- [KAFKA-2042](#) Update topic list of the metadata regardless of cluster information;
- [KAFKA-2043](#) CompressionType is passed in each RecordAccumulator append
- [KAFKA-2044](#) Support requests and responses from o.a.k.common in KafkaApis
- [KAFKA-2047](#) Move the stream creation into concurrent mirror maker threads
- [KAFKA-2048](#) Change lock synchronized to inLock() for partitionMapCond
- [KAFKA-2050](#) Avoid calling .size() on linked list.
- [KAFKA-2056](#) Fix transient testRangePartitionAssignor failure
- [KAFKA-2088](#) kafka-console-consumer.sh should not create zookeeper path when no brokers found and chroot was set in zookeeper.connect.
- [KAFKA-2090](#) Remove duplicate check to metadataFetchInProgress
- [KAFKA-2096](#) Enable keepalive socket option for broker to prevent socket leak
- [KAFKA-2099](#) BrokerEndPoint file, methods and object names should match

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

1.5.10. Knox

HDP 2.3.0 provided Knox 0.6.0 and the following Apache patches:

BUG FIXES

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

IMPROVEMENTS

- [KNOX-545](#) Simplify Keystore Management for Cluster Scaleout

- [KNOX-561](#) Allow Knox pid directory to be configured via the `knox-env.sh` file

1.5.11. Mahout

In HDP-2.3.2, 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 2.3.2 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

1.5.12. Oozie

HDP 2.3.0 provided Oozie 4.2.0 and the following Apache patches:

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

1.5.13. Phoenix

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

- [PHOENIX-1659](#) PHOENIXDatabaseMetaData.getColumns does not return REMARKS column.
- [PHOENIX-2074](#) StackOverflowError with RoundRobinResultIterator.
- [PHOENIX-2096](#) Tweak criteria for when round robin iterator is used.
- [PHOENIX-1978](#) UDF ArgumentTypeMismatchException.
- [PHOENIX-2011](#) Default, min, and max values should not require quotes around it in create function.
- [PHOENIX-2151](#) Two different UDFs called on same column return values from first UDF only.
- [PHOENIX-2022](#) BaseRegionScanner.next should be abstract.
- [PHOENIX-2073](#) Two bytes character in LIKE expression is not allowed.
- [PHOENIX-2131](#) CastParseNode.toSQL omits closing parenthesis.

- [PHOENIX-2066](#) Existing client fails initialization due to upgrade attempting to create column with no name.
- [PHOENIX-2075](#) MR integration uses single mapper unless table is salted.
- [PHOENIX-2254](#) zookeeper.znode.parent value is not taking affect in MR integration job.

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

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

1.5.14. Pig

HDP 2.3.2 provided Pig 0.15.0 and the following Apache patches:

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

- [PIG-4679](#) Performance degradation due to InputSizeReducerEstimator since [PIG-3754](#)

HDP 2.3.0 provided Pig 0.15.0 and the following Apache patch:

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

1.5.15. Ranger

HDP 2.3.2 provides Ranger 0.5.0 and the following Apache patches:

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

BUG FIXES

- [RANGER-560](#) Policy validation: Provide user friendly error messages about validation failures
- [RANGER-580](#) HBase plugin: Plugin may not work after upgrade
- [RANGER-584](#) Service validation: Provide user friendly error messages about validation failures
- [RANGER-587](#) ranger-admin-site.xml not getting updated when ranger.authentication.method is changed
- [RANGER-588](#) Take care of Ranger KMS installation even if 'java' is not in PATH
- [RANGER-593](#) Service def validation: Provide user friendly error messages about validation failures
- [RANGER-594](#) Policy Validation: Change the logic to generate friendly error messages to be like used for Service and Service def
- [RANGER-598](#) Update Ranger config migration script to work with Ranger 0.5
- [RANGER-615](#) Audit to db: Truncate all string values of audit record so that writing of audit does not fail
- [RANGER-618](#) KMS gets slower in key creation once Database grows
- [RANGER-621](#) Solr service-def JSON has incorrect impliedGrants for solr_admin permission
- [RANGER-622](#) Hive plugin: Add jar via beeline throws NPE
- [RANGER-623](#) Enable plugin scripts should handle file permissions for certain umask value
- [RANGER-624](#) Windows installation broken after SQLAnywhere support
- [RANGER-625](#) Change db flavor input parameter value from SQLAnywhere to SQLA
- [RANGER-627](#) Processing done by Audit Shutdown hooks can confuse someone looking at logs to think that shutdown of a service is held up due to Ranger plugin
- [RANGER-628](#) Make filters for ranger-admin search binds configurable

- [RANGER-630](#) Data consistency across API and UI
- [RANGER-632](#) Policy validation error messages produced by the server are not seen by the user
- [RANGER-637](#) Make REFERRAL property in Ranger User sync configurable
- [RANGER-638](#) Ranger admin should redirect back to login page when session cookies expires
- [RANGER-639](#) Storm plugin - commons-lang is a required dependency and hence should be packaged as part of storm plugin
- [RANGER-641](#) Ranger kms start fails if java is not set and started using service keyword
- [RANGER-642](#) Update USERSEARCHFILTER for Ranger Authentication on Windows
- [RANGER-653](#) Move delegated admin check to mgr layer from service layer for XPermMap and XAuditMap

HDP 2.3.0 provided Ranger 0.5.0 and the following Apache patches:

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

BUG FIXES

- [RANGER-219](#) Autocomplete behavior of hive tables/columns
- [RANGER-524](#) HBase plugin: list command should prune the tables returned on user permissions
- [RANGER-529](#) Policy Validation: resources of a policy must match one of the resource hierarchies of the service def.
- [RANGER-533](#) HBase plugin: if user does not have family-level access to any family in a table then user may be incorrectly denied access done at table/family level during get or scan
- [RANGER-539](#) Rolling downgrade changes
- [RANGER-545](#) Fix js error for lower versions of FF (less than 30)
- [RANGER-548](#) Key rollover command fails
- [RANGER-550](#) Hive plugin: Add audit logging support for metadata queries that have filtering support from hive

- [RANGER-553](#) Default policy creation during service creation should handle service defs with multiple hierarchies, e.g. hive, properly
- [RANGER-554](#) Ranger KMS keys listing page does not support pagination
- [RANGER-555](#) Policy view page (from access audit page) gives 404 with Oracle DB
- [RANGER-558](#) HBase plugin: unless user has READ access at some level under the table/family being accessed (via scan/get) authorizer should throw an exception and audit
- [RANGER-565](#) Ranger Admin install fails (sometimes) with IO Error when DB used in Oracle
- [RANGER-566](#) Installation of Ranger on Oracle 12c with shared database needs to use private synonym instead of public synonym
- [RANGER-569](#) Enabling Ranger plugin for HBase should not modify hbase.rpc.protection value
- [RANGER-570](#) Knox plugin: after upgrading ranger from 0.4 to 0.5 the Knox plugin won't work because classes with old names are missing
- [RANGER-571](#) Storm plugin: after upgrading ranger from 0.4 to 0.5 the plugin won't work because classes with old names are missing
- [RANGER-575](#) Allow KMS policies to be assigned to all users
- [RANGER-576](#) Storm audit not showing access type in the Ranger Admin Audit UI

HDP CHANGES

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

1.5.16. Slider

HDP 2.3.2 provides Slider 0.80.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

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

- [SLIDER-941](#) Add JAAS config templates for HBase.

HDP 2.3.0 provided Slider 0.80.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [SLIDER-810](#) YARN config changes to enable partial logs upload for long running services (default include/exclude patterns does not upload any files)
- [SLIDER-877](#) move SLIDER_HOME assignment to slider.py
- [SLIDER-878](#) Slider cannot support JDK 1.8 for command slider registry --getconf hbase-site --name hb1
- [SLIDER-888](#) intermittent errors when accessing key store password during localization of cert stores
- [SLIDER-901](#) AgentClientProvider should use File.separator in paths for platform independence
- [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

1.5.17. Spark

HDP 2.3.2 provides Spark 1.4.1 and the following Apache patches:

NEW FEATURES

- [SPARK-1537](#) Add integration with Yarn's Application Timeline Server.
- [SPARK-6112](#) Provide external block store support through HDFS RAM_DISK.

BUG FIXES

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

HDP 2.3.0 provided Spark 1.3.1 and the following Apache patches:

IMPROVEMENTS

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

1.5.18. Sqoop

HDP 2.3.2 provides Sqoop 1.4.6 and the following Apache patches:

IMPROVEMENTS

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

HDP 2.3.0 provided Sqoop 1.4.6 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

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

1.5.19. Storm

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

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

1.5.20. Tez

HDP 2.3.2 provides Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [TEZ-2602](#) Throwing EOFException when launching MR job.
- [TEZ-2745](#) ClassNotFoundException in InputInitializer causes AM to crash.
- [TEZ-2768](#) Log a useful error message when the summary stream cannot be closed when shutting down an AM.
- [TEZ-2549](#) Reduce Counter Load on the Timeline Server.
- [TEZ-2291](#) TEZ UI. Improper vertex name in tables.
- [TEZ-2211](#) Tez UI. Allow users to configure timezone.
- [TEZ-2754](#) Tez UI. StartTime and EndTime is not displayed with right format in Graphical View.
- [TEZ-2761](#) Tez UI. update the progress on the dag and vertices pages with info from AM.
- [TEZ-2766](#) Tez UI. Add vertex in-progress info in DAG details.
- [TEZ-2761](#) Addendum fix build failure for java 6.
- [TEZ-2752](#) logUnsuccessful completion in Attempt should write original finish time to ATS.
- [TEZ-2807](#) Log data in the finish event instead of the start event.
- [TEZ-2792](#) Add AM web service API for tasks.
- [TEZ-2792](#) Addendum fix build failure for java 6.
- [TEZ-2780](#) Tez UI, Update All Tasks page while in progress.

- [TEZ-2660](#) Tez UI, need to show application page even if system metrics publish is disabled.
- [TEZ-2825](#) Report progress in terms of completed tasks to reduce load on AM for Tez UI.
- [TEZ-2663](#) SessionNotRunning exceptions are wrapped in a ServiceException from a dying AM.
- [TEZ-2842](#) Tez UI, Update Tez App details page while in-progress.
- [TEZ-2483](#) Tez UI, Show error if in progress fails due to AM not reachable.
- [TEZ-2847](#) Tez UI, Task details doesn't gets updated on manual refresh after job complete.
- [TEZ-814](#) Improve heuristic for determining a task has failed outputs.
- [TEZ-2812](#) Preemption sometimes does not respect heartbeats between pre-emptions.
- [TEZ-2834](#) Make Tez preemption resilient to incorrect free resource reported by YARN.
- [TEZ-814](#) Improve heuristic for determining a task has failed outputs.
- [TEZ-2816](#) Preemption sometimes does not respect heartbeats between pre-emptions.
- [TEZ-2853](#) Tez UI, task attempt page is coming empty.

INCOMPATIBLE CHANGES

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

HDP 2.3.0 provided Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [TEZ-1529](#) ATS and TezClient integration in secure kerberos enabled cluster.
- [TEZ-2391](#) TestVertexImpl timing out at times on Jenkins builds.
- [TEZ-2409](#) Allow different edges to have different routing plugins
- [TEZ-2447](#) Tez UI: Generic changes based on feedbacks.
- [TEZ-2453](#) Tez UI: show the dagInfo is the application has set the same.
- [TEZ-2455](#) Tez UI: Dag view caching, error handling and minor layout changes
- [TEZ-2460](#) Temporary solution for issue due to [YARN-2560](#)
- [TEZ-2474](#) The old taskNum is logged incorrectly when parallelism is changed

- [TEZ-2475](#) Fix a potential hang in Tez local mode caused by incorrectly handled interrupts.
- [TEZ-2478](#) Move OneToOne routing to store events in Tasks.
- [TEZ-2481](#) Tez UI: graphical view does not render properly on IE11
- [TEZ-2482](#) Tez UI: Mouse events not working on IE11
- [TEZ-2489](#) Disable warn log for Timeline ACL error when tez.allow.disabled.timeline-domains set to true.
- [TEZ-2504](#) Tez UI: tables - show status column without scrolling, numeric 0 shown as Not available
- [TEZ-2505](#) PipelinedSorter uses Comparator objects concurrently from multiple threads.
- [TEZ-2509](#) YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.
- [TEZ-2513](#) Tez UI: Allow filtering by DAG ID on All dags table.
- [TEZ-2523](#) Tez UI: derive applicationId from dag/vertex id instead of relying on json data
- [TEZ-2527](#) Tez UI: Application hangs on entering erroneous RegEx in counter table search box
- [TEZ-2528](#) Tez UI: Column selector buttons gets clipped, and table scroll bar not visible in mac.
- [TEZ-2535](#) Tez UI: Failed task attempts link in vertex details page is broken.
- [TEZ-2539](#) Tez UI: Pages are not updating in IE.
- [TEZ-2541](#) DAGClientImpl enable TimelineClient check is wrong.
- [TEZ-2546](#) Tez UI: Fetch hive query text from timeline if dagInfo is not set.
- [TEZ-2547](#) Tez UI: Download Data fails on secure, cross-origin clusters
- [TEZ-2548](#) TezClient submitDAG can hang if the AM is in the process of shutting down.
- [TEZ-2554](#) Tez UI: View log link does not correctly propagate login credential to read log from YARN web.
- [TEZ-2568](#) auto_sortmerge_join_5 fails in Tez mode

1.6. Common Vulnerabilities and Exposures

- **CVE-2015-5167:** Restrict REST API data access for non-admin users
Severity: Important
Vendor: Hortonworks
Versions Affected: All HDP 2.3.x releases prior to 2.3.2

Users Affected: All users of ranger policy admin tool.

Impact: See [BUG-41604](#) and [RANGER-630](#). Data access restrictions via REST API are not consistent with restrictions in policy admin UI. Non-admin users can access some ranger data restricted for admin users by calling REST API.

Recommended Action: Upgrade to HDP 2.3.2+.

1.7. Third-party Licenses

Global: [Apache 2.0](#)

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

1.8. Fixed Issues

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

Potential Data Loss

Component	Hortonworks Bug ID	Apache JIRA	Summary
Hive	BUG-42935	HIVE-11607	Fix and backport HIVE-11607 for 2.3-maint (Export tables broken for data > 32 MB)
Storm	BUG-42792	STORM-960	While using a storm/hive/bolt to populate an ACID ORC table, it causes data loss and scan jobs fails with 'java.io.EOFException' error

Security

Component	Hortonworks Bug ID	Apache JIRA	Summary
HDFS	BUG-43223	HDFS-8155	Support OAuth2 in WebHDFS
Kafka	BUG-42312		Kafka GetOffsetShell doesn't work in Kerberos Environment

Incorrect Results

Component	Hortonworks Bug ID	Apache JIRA	Summary
Hive	BUG-44508	HIVE-11839	Vectorization wrong results with filter of (CAST AS CHAR)
Hive	BUG-43168	HIVE-11605	Hive + Tez: map side join giving different query results
Knox	BUG-44504	KNOX-598	Concurrent JDBC clients via KNOX to Kerberized HiveServer2 - causes HTTP 401 error (due to Kerberos Replay attack error)
Ranger	BUG-43030	RANGER-615	Exceptions related to XA secure audit record - cloned for Grainger

Stability

Component	Hortonworks Bug ID	Apache JIRA	Summary
Accumulo	BUG-43482	ACCUMULO-3957	Accumulo monitor's getContentSummary calls may cause NN pauses
Falcon	BUG-40941	FALCON-1373	Hive DR not work on destination cluster
HBase	BUG-42155	HBASE-14196	Thrift server connection idle timeout issue
HCatalog, Hive	BUG-41418	HIVE-11317	ACID: Improve transaction Abort logic due to timeout
HDFS	BUG-43896	HDFS-8797	WebHdfsFileSystem creates too many connections for pread
HDFS	BUG-43306	HDFS-8809	HDFS fsck reports under construction blocks as "CORRUPT"
HDFS	BUG-43657	HDFS-8995	Flaw in registration bookkeeping can make DN die on reconnect
Hive	BUG-34774	HIVE-10231	Compute partition column stats fails if partition col type is date
Hive	BUG-43270	HIVE-10651	ORC : OrcProto \$StripeStatistics cache grows causing severe GC in Application master
Kafka	BUG-42865		Customer upgrade resulted in java.io.FileNotFoundException: /usr/hdp/2.3.0.0-2557/Kafka/bin/./config/tools-log4j.properties (No such file or directory)
Sqoop	BUG-41537	SQOOP-2387	NPE during sqoop import when column contains a special character
Storm	BUG-41546	STORM-951	Storm Hive Bolt leaking a transaction for every txnBatch

Component	Hortonworks Bug ID	Apache JIRA	Summary
Tez	BUG-43220	TEZ-2745	ClassNotFoundException in InputInitializer causes AM to crash
YARN	BUG-41417	YARN-2019	RM crashes with FATAL error STATE_STORE_OP_FAILED

Query Failure

Component	Hortonworks Bug ID	Apache JIRA	Summary
Hive	BUG-44673	HIVE-11193	query plan fails, tez dynamic partition pruning "fail to find child from parent " in this case (left outer join union all with constant on different data types)
Hive	BUG-43315	HIVE-11344	While reading table using hive HCatalog api throws null pointer exception.
Hive	BUG-42919	HIVE-11606	bucket mapjoin error "capacity must be power of two" when #rows < #buckets
Hive, Tez	BUG-41584	HIVE-11356	Error thrown when Empty tables and SMB (OUTER) JOINS are used
Hue	BUG-44101		When accessing a Table which has more than 4K columns, Beeswax and HCatalog fails to get the content and results in error

Upgrade

None.

Usability

Component	Hortonworks Bug ID	Apache JIRA	Summary
HDFS	BUG-43767	HDFS-8180	AbstractFileSystem Implementation for WebHdfs
HDFS	BUG-43225	HDFS-8435	Support CreateFlag in WebHdfs
Hive	BUG-43159	HIVE-11581	HS2 URL parameters should be stored in zookeeper
Hive	BUG-42834	HIVE-11658	Load data inpath fails with SemanticException when used with ORC files
Phoenix	BUG-43284	Phoenix-2031	Unable to process timestamp data loaded via Phoenix org.apache.phoenix.pig.PhoenixHBaseLoader
Phoenix	BUG-41321	Phoenix-1659	Backport Phoenix-1659 to 2.3.x maintenance
Ranger	BUG-43276	RANGER-628	RangerAuthenticationProvider is using anonymous bind for search when credentials are specified

Component	Hortonworks Bug ID	Apache JIRA	Summary
Ranger	BUG-43313	RANGER-638	Ranger admin should redirect back to login page when login cookies are no longer valid
Spark	BUG-42252	SPARK-6918	Secure HBase with Kerberos does not work over YARN
Storm	BUG-41859	STORM-512	KafkaBolt doesn't handle ticks properly
Storm	BUG-43462	STORM-833 , STORM-848	Build same version of slf4j-api and log4j-over-slf4j
Storm	BUG-43141	STORM-848	Clean up dependencies and shade as much as possible
Storm	BUG-42794	STORM-938	Add an option to the storm hive-bolt to flush at a certain interval

Performance

Component	Hortonworks Bug ID	Apache JIRA	Summary
HDFS	BUG-42941	HDFS-8826	Balancer may not move blocks efficiently in some cases.
Hive	BUG-43118	HIVE-6727	table fast stats numFiles, totalSize set to 0 when table has non-default location
Pig	BUG-44276	PIG-4679	Drastic performance degradation due to InputSizeReducerEstimator since PIG-3754
Ranger	BUG-41359	RANGER-591	CLONE - UserSync process performance issue due to user group cache inconsistency.

Other

Component	Hortonworks Bug ID	Apache JIRA	Summary
Accumulo	BUG-42799	ACCUMULO-3890	Use of CredentialProvider results in a lot of NameNode operations
HDFS	BUG-42942	HDFS-8278	HDFS Balancer should consider remaining storage % when checking for under-utilized machines.
HDFS	BUG-43786	HDFS-8542	WebHDFS getHomeDirectory behavior does not match specification
HDFS	BUG-41192	HDFS-8767	HDFS listStatus on Unix pipe file fails with NPE (hive and other components can not use it)
HDFS	BUG-43227	HDFS-8885	ByteRangeInputStream used in webhdfs does not override available()
HDFS	BUG-43787	HDFS-8939	Test(S)WebHdfsFileContextMainOperations failing on branch-2

Component	Hortonworks Bug ID	Apache JIRA	Summary
Hive	BUG-42396	HIVE-11442	commons-configuration-1.6.jar is backwards incompatible and regressive

1.9. Known Issues

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-28296	STORM-642	Storm	Benchmarking for trident filter and join
BUG-28297	STORM-642	Storm	Benchmarking for trident sliding windows and enrichment
BUG-28298	STORM-642	Storm	benchmarking for storm core watchlist type filter
BUG-28448	KNOX-474	Knox	Improved Kerberos config validation and diagnostics at startup
BUG-28904	SLIDER-713	Slider	BUG-14598 Implement the cluster flex operation via REST
BUG-28905	SLIDER-714	Slider	BUG-14598 Expose REST view of containers/ component instances with DELETE operation
BUG-28907	SLIDER-716	Slider	BUG-14598 Offer REST upgrade operation
BUG-28908	SLIDER-717	Slider	BUG-14598 Migrate slider client to slider REST API
BUG-28909	SLIDER-170	Slider	BUG-14598 Migrate slider client to REST YarnClient implementation
BUG-28910	SLIDER-718	Slider	BUG-14598 Slider client to support remote file operations via WebHDFS
BUG-28914	SLIDER-720	Slider	BUG-14598 Slider client to switch to YARN Registry REST API
BUG-29589	STORM-676	IOT, Storm	Sliding Time Windows
BUG-29603	STORM-650	Storm	Kafka Spout/Bolt rewrite! , support for Kafka 0.8.2 as it breaks backward compatibility for consumers
BUG-31629	TEZ-2086	Tez	TEZ-UI Support Task/ TaskAttempt level log
BUG-32936	FALCON-1111	Falcon	Add a new Java action in the process workflow start to update Graph DB
BUG-34711	HIVE-10939	Hive	BUG-31340 Hive 2.3.x UT (Windows): ORC failures
BUG-35363		Hive	ntile (order by) fails of java array limit when column values more 2,147,483,647

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-36632	HBASE-13832 , HBASE-8510	HDFS	After Rolling Upgrade From HDP 2.2 to 2.3 HBase Master Goes Down
BUG-36817	HBASE-13330 , HBASE-13647	HBase	test_IntegrationTestRegionReplicaReplication[Inte fails with READ FAILURES
BUG-37012	RANGER-526	Ranger	Provide REST API to change user role
BUG-38046		Spark	Spark ATS is missing Kill event
BUG-38054	RANGER-577	Ranger	ranger should not change hive config if authorization is disabled
BUG-38299	CALCITE-645	Phoenix	In case of query failures phoenix query server exception/error message details are not propagated to sqlline-thin.py/client
BUG-38471	HDFS-8512	HDFS	Storage type inside LocatedBlock object is not fully exposed for GETFILESTATUS
BUG-38980	HBASE-14223	HBase	Meta WALs are not split or cleared
BUG-39160	SPARK-8275	Spark	Spark History Server will not updated incomplete applications once viewed
BUG-39265	OOZIE-2311	Oozie	NPE in oozie logs while running feed replication tests causes jobs to fail.
BUG-39338		Spark	lpython pyspark does not work on SUSE 11.3
BUG-39344	PHOENIX-1993	Phoenix	Phoenix is raising nullPointerExceptions on queries.
BUG-39424	YARN-2194	YARN	NM fails to come with error "Not able to enforce CPU weights; cannot write to cgroup" [Cgroups]
BUG-39463	AMBARI-12445	Ranger	ranger-admin and ranger-usersync scripts missing in /etc/init.d when installed through ambari.
BUG-39615	HIVE-11036	Hive	Hive query got stuck as HS2 server throwing "java.io.IOException: Incomplete HDFS URI, no host: hdfs://NAMENODE_HOSTNAME:8020/ranger/audit/hiveServer2/..."
BUG-39796	SPARK-7889	Spark	Spark HistoryServer caches incomplete App UIs
BUG-39823		HDFS	Restart NFSGateway is failed After changing fs.defaultFS default value at webhdfs://Host_name:50070 (Secured cluster)

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-39988	HIVE-11110	Hive	CBO : Default partition filter is from MetaStore query causing TPC-DS to regress by 3x
BUG-40313	HIVE-11068	Hive	Hive : OOM in Hive client during plan generation, regression introduced by HIVE-10940
BUG-40481	AMBARI-12436	Ambari, Falcon	Falcon service check failed / daemon crashed during RU
BUG-40682	SLIDER-909	Ambari	Slider HBase app package fails in secure cluster with wire-encryption on
BUG-40722		Kafka	Ability to catch and deal with OffsetOutOfRangeException
BUG-40883	HADOOP-10027	Hadoop Common	Backport libhadoop.so zlib SEGV fix
BUG-41038	SLIDER-777	Slider, YAaS	Provide slider dependencies as a self contained versioned tarball
BUG-41151		HDP / Stack, Hue	HDP version not displaying in Hue post upgrade
BUG-41366		Hue	Hue by default is using spawning instead of cherrypy
BUG-41369		HDP / Stack, Hue	HDP version not displayed on centos6, Hadoop version displayed is incorrect
BUG-41606		YARN	AM Container link in YARN RM UI links to a 404 page after application completion
BUG-41621		HDFS	HDFS: Track DFSClient name in hdfs-audit.log
BUG-42072	PHOENIX-2067 , PHOENIX-2120	Phoenix	Backport PHOENIX-2067 and PHOENIX-2120
BUG-42569	HIVE-10022	Hive	'create database [db_name] location / tmp/[db_name].db' via beeline throws [hrt_qa] does not have [WRITE] privilege although hive.server2.enable.doAs=false when Ranger or StdAuth is on.
BUG-42608		Hive	no logging about progress in moveTask, during long time of large partitioned table loading
BUG-42681		HCatalog, Pig	HCatStorer does not store full result of union unless preceded by reducer operation
BUG-42828		Hive	multi-insert fails with merge tasks when hive.exec.parallel=true

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-42880		Knox	Knox performance improvements under load
BUG-43057	AMBARI-12844	Ambari, Ranger	after enabling SSL for ranger plugin Quick link is not loading on Ambari UI
BUG-43391	AMBARI-12403	YARN	Unable to submit application to YARN from external client on Kerberos-enabled cluster
BUG-43655		Hadoop Common	hadoop cassini UT failure
BUG-43691	HIVE-11713	Hive	Map instances with null keys are not written unless table file format is ORC.
BUG-43699	TEZ-2768	Tez	Log a useful error message when the summary stream cannot be closed when shutting down an AM
BUG-43711		Kafka	KafkaBroker should handle disk failures for log.dir gracefully
BUG-43982		MapReduce, YARN	Yarn application logs are showing the following error after upgrade.
BUG-44016		Storm	Storm Nimbus and DRPC Server die with java.lang.OutOfMemoryError: Java heap space
BUG-44038	HIVE-11762	Hive	TestHCatLoaderEncryption failures in 2.3-maint: DFSClient.setKeyProvider() signature changed between Hadoop 2.6 and 2.7
BUG-44103		Storm	Storm Nimbus and DRPC Server die with java.lang.OutOfMemoryError: Java heap space
BUG-44140	HIVE-11875		Oozie-Hive test fail due to JDBC not using the delegation token provided by the client Workaround: Specify an HS2 URL directly when using Oozie.
BUG-44153	HDFS-9046	HDFS	Any Error during BPOfferService run can leads to Missing DN (HDFS-9046)
BUG-44157	RANGER-629	Hive	[Ranger][hive] security issue with hive & ranger for import table command
BUG-44166	HIVE-5277	Hive	Backport HIVE-5277(HBase handler skips rows with null valued first cells when only row key is selected)
BUG-44191		HBase, Slider	test_runintegrationtestzkandhdfspermissions failed on HBase-Slider

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-44253	RANGER-656	Ranger	Ranger UI - KMS Need to handle 404 error when click on breadcrumb
BUG-44378		Ranger	Ranger-Solr: Issue with Mapping Principals to user names "solr"
BUG-44382		Hive	wrong results: non-mapjoin with auto.convert.join.noconditionaltask=false and hive.optimize.reducededuplication=true wrong plan for join followed with group by
BUG-44385	RANGER-643	Ranger	Audit page: filter by Tags column does not work
BUG-44386	RANGER-644	Ranger	Update Solr audit source/ destination to support 'Tags' field
BUG-44388	RANGER-645	Ranger	Tag DB Store should be available in all supported DB flavors
BUG-44390	HIVE-11831 , HIVE-11833	Hive	Hive Server2 concurrency tests failing on Oracle 12 with error "ORA-08177: can't serialize access for this transaction"
BUG-44392	RANGER-658	Ranger	make ranger_credential_helper.py available in usersync dir
BUG-44426	HIVE-11915	Hive	HS2LongRunning: 'create table' fail intermittently as it failed to acquire locks
BUG-44489	HIVE-11892	Hive	Local fetch task does not work with UDTF
BUG-44505		Tez	[Tez View] Dags do not get updated without hitting refresh
BUG-44506	AMBARI-13138	Ambari, Ranger	Ranger installed with HDP 2.2 doesn't handle alerts when http, HTTPS ports value are changed
BUG-44646	HIVE-11935	Hive	HS2LongRunningHTTP: Describe table intermittently throws NPE
BUG-44739		Phoenix	Still having problems after fixing "Duplicate Column Name" error when creating new table
BUG-44837	HIVE-11902	Hive	Abort txn cleanup thread throws SyntaxErrorException
BUG-44928		Kafka, Storm	need to include storm-kafka-0.10.0.2.3.0.0-2557.jar in the storm/lib sub-directory
BUG-45023	HBASE-14501	HBase	HBase replication in TDE clusters fails with NullPointerException.

Hortonworks Bug ID	Apache JIRA	Component	Summary
			<p>Replication live tails the active WAL of the region server that is writing the WAL files. Due to a bug in handling EOF in decoding cells from the WAL files, a NPE can be observed.</p> <p>Workaround: Replication and async WAL replication for region replicas cannot be used together with HDFS TDE.</p>
BUG-45054	TEZ-2398	Tez	Sometime faulttolerance test application is sometimes hanging
BUG-45083		HCatalog, Pig	More than 66 columns breaks MR but not tez (Pig HCat)
BUG-45108	RANGER-664	Ranger	PolicyRefresher connection and read timeout should be configurable
BUG-45118	HADOOP-11618 , HADOOP-12304	Falcon, Oozie	(Patch) When WASB is set as default FS then oozie service check fails
BUG-45181		Ranger	Ranger does validations on group names while reading group_users that are not enforced during letting the groups into the system
BUG-45190		Slider	<p>HBase Slider App Creation issue</p> <p>Creation/start of a Slider application via the Ambari Slider view might fail with the following error in the AM log:</p> <pre>java.lang. AbstractMethodError: javax.ws.rs.core. UriBuilder.uri(Ljava/ lang/String;)</pre> <p>Workaround: Perform the following:</p> <ol style="list-style-type: none"> 1. Stop and start the application again from the Ambari Slider view. <p>Stop and start might have to be attempted couple of times. If this works, ignore all additional steps below.</p> <ol style="list-style-type: none"> 2. If Step 1 fails: stop the application from the Ambari Slider view. 3. Find the user that was set as the value of the

Hortonworks Bug ID	Apache JIRA	Component	Summary
			<p>property <code>slider.user</code> in the Ambari Slider view.</p> <p>4. Log in to a host which has Slider client installed.</p> <p>5. Run <code>slider start <app_name></code> from the command line (which was found in step 3).</p>
BUG-45236		Upgrade	<p>If a downgrade is made to HDP 2.2.8.x, the VERSION table in metastore rdbms needs to be updated to set the version to 0.14.0:</p> <ul style="list-style-type: none"> MySQL: <pre data-bbox="1182 680 1427 1129"> UPDATE VERSION SET SCHEMA_VERSION='1.2.0', VERSION_COMMENT='Hive release version 1.2.0' where VER_ID=1; SELECT 'Finished setting MetaStore schema to 1.2.0' AS ' '; UPDATE VERSION SET SCHEMA_VERSION='0.14.0', VERSION_COMMENT='Hive release version 0.14.0' where VER_ID=1; SELECT 'Finished setting MetaStore schema to 0.14.0' AS ' '; </pre> Postgres: <pre data-bbox="1182 1199 1427 1648"> UPDATE "VERSION" SET "SCHEMA_VERSION"='1.2.0', "VERSION_COMMENT"='Hive release version 1.2.0' where "VER_ID"=1; SELECT 'Finished setting MetaStore schema to 1.2.0'; UPDATE "VERSION" SET "SCHEMA_VERSION"='0.14.0', "VERSION_COMMENT"='Hive release version 0.14.0' where "VER_ID"=1; SELECT 'Finished setting MetaStore schema to 0.14.0'; </pre> Oracle:

Hortonworks Bug ID	Apache JIRA	Component	Summary
			<pre> UPDATE VERSION SET SCHEMA_VERSION='1.2.0', VERSION_COMMENT='Hive release version 1.2.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 1.2.0' AS Status from dual; UPDATE VERSION SET SCHEMA_VERSION='0.14.0', VERSION_COMMENT='Hive release version 0.14.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 0.14.0' AS Status from dual; </pre> <ul style="list-style-type: none"> • azuredb: <pre> UPDATE [dbo].[VERSION] SET SCHEMA_VERSION='1.2.0', VERSION_COMMENT='Hive release version 1.2.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 1.2.0' AS MESSAGE; UPDATE [dbo].[VERSION] SET SCHEMA_VERSION='0.14.0', VERSION_COMMENT='Hive release version 0.14.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 0.14.0' AS MESSAGE; </pre> • mssql: <pre> UPDATE VERSION SET SCHEMA_VERSION='1.2.0', VERSION_COMMENT='Hive release version 1.2.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 1.2.0' AS MESSAGE; UPDATE VERSION SET SCHEMA_VERSION='0.14.0', VERSION_COMMENT='Hive release version 0.14.0' where VER_ID=1; SELECT 'Finished upgrading MetaStore schema to 0.14.0' AS MESSAGE; </pre> <p>If an upgrade is being performed after a downgrade to HDP 2.2.8.x, and the above instructions</p>

Hortonworks Bug ID	Apache JIRA	Component	Summary
			have been run, the VERSION table in metastore rdbms should be changed to 1.2.0 before the upgrade is performed.
BUG-45664		Kafka	Memory leak in Kafka Broker caused by leak in instance of ConcurrentHashMap/socketContainer
BUG-45688	KAFKA-2012	Kafka	Kafka index file corruption
BUG-46186	N/A	Ranger	<p>When configuring the Ranger HBase plugin in a Kerberized Environment, it may be necessary to add an auth-to-local property to the plugin in order for the Test Connection button to work properly.</p> <p>The connection test can fail with the error Illegal Principal name even though the principal for the Ranger repository config user is correct and working.</p> <p>To resolve this issue:</p> <ol style="list-style-type: none"> 1. Log in to the Ranger Admin UI. 2. Locate the HBase plugin and click Edit. 3. In the Add New Configurations section of the Edit Service page, add a <code>hadoop.security.auth_to_local</code> rule and include only one rule that matches the Ranger Repository Config User. <p>For example, if the config user principal is 'hbaseuser@AD.EXAMPLE.COM', then the rule will look like this: <code>RULE:1:\$1@\$0(hbaseuser@EXAMPLE.COM)s/.*\/hbaseuser/.</code></p> <ol style="list-style-type: none"> 4. Save the configuration and attempt to click Test Connection again.
BUG-46214		Spark	SparkR - SparkR is missing from our package in Spark 1.4.1; it will be made available in an upcoming maintenance release of HDP 2.3.x.

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-46823		Knox	KNOX_GATEWAY START was failed when NN HA was enabling
BUG-50531		Kafka	Kafka file system support Issue: Encrypted file systems such as SafenetFS are not supported for Kafka. Index file corruption can occur. For more information, see: Install Kafka .
BUG-55196	HIVE-12937	Hive	DbNotificationListener unable to clean up old notification events

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
TSB-405	N/A	N/A	Impact of LDAP Channel Binding and LDAP signing changes in Microsoft Active Directory Microsoft has introduced changes in LDAP Signing and LDAP Channel Binding to increase the security for communications between LDAP clients and Active Directory domain controllers. These optional changes will have an impact on how 3rd party products integrate with Active Directory using the LDAP protocol. Workaround Disable LDAP Signing and LDAP Channel Binding features in Microsoft Active Directory if they are enabled For more information on this issue, see the corresponding Knowledge article: TSB-2021 405: Impact of LDAP Channel Binding and LDAP signing changes in Microsoft Active Directory
TSB-406	N/A	HDFS	CVE-2020-9492 Hadoop filesystem bindings (ie: webhdfs) allows credential stealing WebHDFS clients might send SPNEGO authorization header to remote URL without proper verification. A maliciously crafted request can trigger services to send server credentials to a webhdfs path (ie: webhdfs://...) for capturing the service principal For more information on this issue, see the corresponding Knowledge article: TSB-2021 406: CVE-2020-9492 Hadoop filesystem bindings (ie: webhdfs) allows credential stealing
TSB-434	HADOOP-17208 , HADOOP-17304	Hadoop	KMS Load Balancing Provider Fails to invalidate Cache on Key Delete For more information on this issue, see the corresponding Knowledge article: TSB 2020-434: KMS Load Balancing Provider Fails to invalidate Cache on Key Delete
TSB-465	N/A	HBase	Corruption of HBase data stored with MOB feature For more information on this issue, see the corresponding Knowledge article: TSB 2021-465: Corruption of HBase data stored with MOB feature on upgrade from CDH 5 and HDP 2
TSB-497	N/A	Solr	CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
			<p>The Apache Solr ReplicationHandler (normally registered at "/replication" under a Solr core) has a "masterUrl" (also "leaderUrl" alias) parameter. The "masterUrl" parameter is used to designate another ReplicationHandler on another Solr core to replicate index data into the local core. To help prevent the CVE-2021-27905 SSRF vulnerability, Solr should check these parameters against a similar configuration used for the "shards" parameter.</p> <p>For more information on this issue, see the corresponding Knowledge article: TSB 2021-497: CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler</p>
TSB-512	N/A	HBase	<p>HBase MOB data loss</p> <p>HBase tables with the MOB feature enabled may encounter problems which result in data loss.</p> <p>For more information on this issue, see the corresponding Knowledge article: TSB 2021-512: HBase MOB data loss</p>

1.10. Documentation Errata

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

- [Flume: Kafka Sink \[58\]](#)
- [Hive Sink \[59\]](#)
- [Configuring Pig Scripts to Use HCatalog in Oozie Workflows \[63\]](#)
- [Configuring a Sqoop Action to Use Tez to Load Data into a Hive Table \[64\]](#)

1.10.1. 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.

Property Name	Default	Description
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.

```

a1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink a1.sinks.k1.topic =
mytopic
a1.sinks.k1.brokerList = localhost:9092
a1.sinks.k1.requiredAcks = 1
a1.sinks.k1.batchSize = 20
a1.sinks.k1.channel = c1

```

1.10.2. 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.

Property Name	Default	Description
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 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 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)
%l	Hour (01..12)

Alias	Description
%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 useLocalTimestamps 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.

1.10.3. Configuring Pig Scripts to Use HCatalog in Oozie Workflows

To access HCatalog with a Pig action in an Oozie workflow, you need to modify configuration information to point to the Hive metastore URIs.

There are two methods for providing this configuration information. Which method you use depends upon how often your Pig scripts access the HCatalog.

1.10.3.1. Configuring Individual Pig Actions to Access HCatalog

If only a few individual Pig actions access HCatalog, do the following:

1. Identify the URI (host and port) for the Thrift metastore server.
 - a. In Ambari, click **Hive > Configs > Advanced**.
 - b. Make note of the URI in the `hive.metastore.uris` field in the General section.

This information is also stored in the `hive.default.xml` file.

2. Add the following two properties to the `<configuration>` elements in each Pig action.



Note

Replace `[host:port(default:9083)]` in the example below with the host and port for the Thrift metastore server.

```
<configuration>
  <property>
    <name>hive.metastore.uris</name>
    <value>thrift://[host:port(default:9083)]</value>
    <description>A comma separated list of metastore uris the client can
    use to contact the
    metastore server.</description>
  </property>
  <property>
    <name>oozie.action.sharelib.for.pig</name>
    <value>pig,hive,hcatalog</value>
    <description>A comma separated list of libraries to be used by the
    Pig action.</description>
  </property>
</configuration>
```

1.10.3.2. Configuring All Pig Actions to Access HCatalog

If all of your Pig actions access HCatalog, do the following:

1. Add the following line to the `job.properties` files, located in your working directory:

```
oozie.action.sharelib.for.pig=pig,hive,hcatalog
```

```
<!-- A comma separated list of libraries to be used by the Pig action.-->
```

2. Identify the URI (host and port) for the Thrift metastore server.
 - a. In Ambari, click **Hive > Configs > Advanced**.
 - b. Make note of the URI in the `hive.metastore.uris` field in the General section.

This information is also stored in the `hive.default.xml` file.
3. Add the following property to the `<configuration>` elements in each Pig action.



Note

Replace `[host:port(default:9083)]` in the example below with the host and port for the Thrift metastore server.

```
<configuration>
  <property>
    <name>hive.metastore.uris</name>
    <value>thrift://[host:port(default:9083)]</value>
    <description>A comma separated list of metastore uris the client can
    use to contact the
    metastore server.</description>
  </property>
</configuration>
```

1.10.4. Configuring a Sqoop Action to Use Tez to Load Data into a Hive Table

You can use the Tez execution engine to load data into a Hive table using the `--hive-import` option,

In the code example in each step, replace the sample text in [square brackets] with the appropriate information for your configuration.

1. Create a workflow directory.

```
hdfs dfs -mkdir -p [/user/dummy/app]
```

2. Create a `lib` directory in the workflow directory.

```
hdfs dfs -mkdir -p [/user/dummy/app/lib]
```

3. Copy the database JDBC driver jar file to the `lib` directory.

```
hadoop fs -copyFromLocal [/usr/share/java/mysql-connector-java.jar]
[/user/dummy/app/lib]
```

4. Copy the `hive-site.xml` and `tez-site.xml` files to a location accessible by the workflow. For example:

```
hadoop fs -copyFromLocal [/etc/oozie/conf/action-conf/hive/hive-site.xml /
user/dummy/app]
hadoop fs -copyFromLocal [/etc/oozie/conf/action-conf/hive/tez-site.xml /
user/dummy/app]
```

5. In the Sqoop action of the workflow, do the following:

- Add `hive-site` and `tez-site` resources in the `<file>` element of the Sqoop action in the workflow.

```
<file>/user/dummy/app/hive-site.xml#hive-site.xml</file>
<file>/user/dummy/app/tez-site.xml#tez-site.xml</file>
```

- Include the `--hive-import` option in the `<command>` element.

```
<command>import --connect [jdbc:mysql://db_host:port/database] --
username [user]
--password [pwd] --driver [com.mysql.jdbc.Driver] --table [table_name]
--hive-import -m 1 </command>
```

6. Add the following into the `job.properties` file.

```
oozie.use.system.libpath=true
oozie.action.sharelib.for.sqoop=sqoop,hive
```

More information regarding the Sqoop parameters can be found in the Apache documentation at https://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html#_importing_data_into_hive

Example Workflow Action

Replace all sample text in [square brackets] in the example below with the appropriate workflow name, URI, paths, file names, etc. for your configuration.

```
<action name="sqoop-node">
  <sqoop xmlns="uri:oozie:sqoop-action:0.2">
    <job-tracker>${jobTracker}</job-tracker>
    <name-node>${nameNode}</name-node>
    <configuration>
      <property>
        <name>mapred.job.queue.name</name>
        <value>${queueName}</value>
      </property>
    </configuration>
    <command>import --connect [jdbc:mysql://db_host:port/database] --
username [user]
--password [pwd] --driver [com.mysql.jdbc.Driver] --table [table_name] --hive-
import -m 1</command>
    <file>/user/dummy/app/hive-site.xml#hive-site.xml</file>
    <file>/user/dummy/app/tez-site.xml#tez-site.xml</file>
  </sqoop>
  <ok to="end"/>
  <error to="killJob"/>
</action>
```