

Hortonworks Data Platform

HDP 2.3.0 for Windows Release Notes

(Jul 21, 2015)

Hortonworks Data Platform: HDP 2.3.0 for Windows 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 Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, ZooKeeper and Ambari. 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.

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



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

Table of Contents

1. HDP 2.3.0 for Windows Release Notes	1
1.1. Apache Component Support	1
1.1.1. Supported Apache Component Version Information	1
1.1.2. Unsupported Components	2
1.2. New Features	2
1.3. Unsupported Features	5
1.3.1. Technical Preview Features	5
1.3.2. Community Features	5
1.4. Behavior Changes	6
1.5. Apache Patch Information	8
1.5.1. Apache Hadoop	8
1.5.2. Calcite	12
1.5.3. Falcon	12
1.5.4. Flume	12
1.5.5. HBase	13
1.5.6. Hive	13
1.5.7. Knox	14
1.5.8. Mahout	15
1.5.9. Oozie	15
1.5.10. Phoenix	15
1.5.11. Pig	16
1.5.12. Ranger	16
1.5.13. Slider	17
1.5.14. Spark	18
1.5.15. Sqoop	18
1.5.16. Storm	18
1.5.17. Tez	19
1.6. Fixed Issues	20
1.7. Known Issues	26

List of Tables

1.1. Data Governance and Integration	2
1.2. Data Access	3
1.3. Data Management	3
1.4. Security	4
1.5. Operations	4
1.6. Technical Previews	5
1.7. Community Features	5
1.8. HBase	6
1.9. Spark	6
1.10. HDFS: High Availability	6
1.11. JDK Support	7
1.12. Apache HBase	26
1.13. Apache Hive	27
1.14. Apache Oozie	30
1.15. Apache Ranger	31
1.16. Apache Slider	31
1.17. Apache Spark	31
1.18. Apache Tez	32
1.19. Apache YARN	32
1.20. HDFS and Cloud Deployment	34
1.21. Upgrade	34

1. HDP 2.3.0 for Windows Release Notes

The HDP 2.3.0 for Windows Release Notes provide information about Apache component support, new features and changed functionality, Apache patches, and fixed and known issues.

- [Apache Component Support](#)
- [New Features](#)
- [Unsupported Features](#)
- [Behavioral Changes](#)
- [Apache Patch Information](#)
- [Fixed Issues](#)
- [Known Issues](#)

1.1. Apache Component Support

This section provides supported Apache component version information, a list of Apache components not supported on HDP 2.3 for Windows, and third-party tool information.

1.1.1. Supported Apache Component Version Information

All HDP 2.3 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 2.3 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.3.

- Apache Hadoop 2.7.1
- Apache Calcite 1.2.0
- Apache DataFu 1.3.0
- Apache Falcon 0.6.1
- Apache Flume 1.5.2
- Apache HBase 1.1.1
- Apache Hive 1.2.1
- Apache Knox 0.6.0

- Apache Mahout 0.9.0+
- Apache Oozie 4.2.0
- Apache Phoenix 4.4.0
- Apache Pig 0.15.0
- Apache Ranger 0.5.0
- Apache Slider 0.80.0
- Apache Spark 1.3.1
- Apache Sqoop 1.4.6
- Apache Storm 0.10.0
- Apache Tez 0.7.0
- Apache ZooKeeper 3.4.6

1.1.2. Unsupported Components

The following components are supported on HDP 2.3 for Linux, and are not currently supported on HDP 2.3 for Windows:

- Apache Accumulo 1.7.0
- Apache Ambari 2.1
- Apache Atlas 0.5.0
- Apache Kafka 0.8.2
- Apache Solr 5.2.1
- Cascading 3.0.1
- Cloudbreak 1.0
- Hue 2.6.1

1.2. New Features

This section highlights several new features in HDP 2.3.

Table 1.1. Data Governance and Integration

Component	Feature
Falcon	• Replication of Hive and HCat – GA (FALCON-1028)

Component	Feature
	<ul style="list-style-type: none"> High Availability (FALCON-1029) UI to enable entity/process/feed management (FALCON-790) Search UI and backend support by Free text, Entity and Tag (FALCON-914, FALCON-1095, FALCON-1121, FALCON-1122)
Flume	<ul style="list-style-type: none"> Hive Sink based on the new Hive Streaming support (FLUME-1734)
Sqoop	<ul style="list-style-type: none"> Import sequential datasets from mainframe (SQOOP-1272) Netezza enhancements: skip control codes, write logs to HDFS (SQOOP-2164)

Table 1.2. Data Access

Component	Feature
HBase and Phoenix	<ul style="list-style-type: none"> RPC throttling and quotas (HBASE-11598) Multi-WAL Support (HBASE-5699) Reliability and performance optimizations (HBASE-12439) Phoenix SQL improvements: Union All, Date/Decimal types (PHOENIX-1580, PHOENIX-1662) Phoenix support for Java UDFs (PHOENIX-538) Phoenix support for HBase Timeline-Consistent Read High Availability (PHOENIX-1683) Phoenix Tracing Support (PHOENIX-1115) Phoenix Spark Driver (PHOENIX-1071)
Hive	<ul style="list-style-type: none"> SQL improvements: Current_Date, Current_Timestamp (HIVE-5472) Other Date/Datetime improvements, such as SerDe support for ISO 8601 format (HIVE-9298, HIVE-9564) Support UNION (HIVE-9039) and add Interval datatype in expressions (HIVE-9792, HIVE-5021)
Pig	<ul style="list-style-type: none"> Call Hive UDFs from Pig (PIG-3294) Dynamic Parallelism via Tez (PIG-4434)
Spark	<ul style="list-style-type: none"> General availability of Spark 1.3.1
Storm	<ul style="list-style-type: none"> Declarative Topology (STORM-561) Rolling Upgrade (STORM-634)
Tez	<ul style="list-style-type: none"> Secure ATS integration (TEZ-1529) Enhanced performance and scale (TEZ-776) Support sort buffers larger than 2GB (TEZ-1803)

Table 1.3. Data Management

Component	Feature
Slider	<ul style="list-style-type: none"> Upgrade and reconfiguration without downtime for Slider-based applications (SLIDER-787)

Component	Feature
	<ul style="list-style-type: none"> No packaging required for certain Slider applications (SLIDER-668)
YARN	<ul style="list-style-type: none"> Non-exclusive Node Labels - where applications are given preference for the Label they specify, but not exclusive access (YARN-3214) Fair sharing across apps for same user same queue, per queue scheduling policies (YARN-3306) Pluggable authorization for YARN ACLs for integration with Apache Ranger (YARN-3100)
HDFS	<ul style="list-style-type: none"> Improve distcp efficiency: reduced time and processing power needed to mirror datasets across cluster (HDFS-7535, MAPREDUCE-6248) Support variable-length blocks (HDFS-3689) Provide storage quotas per heterogeneous storage types (HDFS-7584) Pluggable Authorization API (HDFS-6826) Track and display failed DataNode storage locations in NameNode JMX and UI (HDFS-7604) Additional DataNode and NameNode operational load metrics available through JMX (HDFS-7773) HDFS Data at Rest Encryption (HDFS-6134)

Table 1.4. Security

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

Table 1.5. Operations

Component	Feature
Oozie	<ul style="list-style-type: none"> HiveServer2 action (OOZIE-1457) Spark action (OOZIE-1983) Stop jobs by coordinator name (OOZIE-2108)

1.3. Unsupported Features

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

1.3.1. Technical Preview Features

The following features are available within HDP 2.3, but are not ready for production deployment. We encourage you to explore these technical preview features in non-production environments and provide feedback on your experiences through the [Hortonworks Community Forums](#).

Table 1.6. Technical Previews

Component	Feature
HBase and Phoenix	<ul style="list-style-type: none"> Phoenix Query Server (PHOENIX-971)
Ranger	<ul style="list-style-type: none"> Security features for data governance: global policies by metadata tags and searchable security access audit
Spark	<ul style="list-style-type: none"> SparkSQL, programmatically with SQLContext (not supported with Thrift Server - JDBC/ODBC) DataFrame API (SPARK-5097) Spark Streaming ML Pipeline API in PySpark (SPARK-3530) ORC file support Dynamic Executor Allocation
Storm	<ul style="list-style-type: none"> Elastic topology via YARN/Slider Monitoring of Storm topologies and clusters
YARN	<ul style="list-style-type: none"> NodeManager: add cgroup support for disk I/O isolation (YARN-2619) Add support for network I/O isolation/scheduling for containers (YARN-2140)

1.3.2. Community Features

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

Table 1.7. Community Features

Component	Feature
YARN	<ul style="list-style-type: none"> Fair Scheduler MapReduce Uber AM MapReduce Eclipse Plug-in

Component	Feature
HDFS	<ul style="list-style-type: none"> NameNode Federation (HDFS-1052) viewFS (HADOOP-7257) block-volume device choosing (HDFS-1804)
Knox	<ul style="list-style-type: none"> Storm REST APIs
Oozie	<ul style="list-style-type: none"> Spark action (OOZIE-1983)
Spark	<ul style="list-style-type: none"> Spark Standalone GraphX

1.4. Behavior Changes

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

Table 1.8. HBase

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

Table 1.9. Spark

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

Table 1.10. HDFS: High Availability

Hortonworks Bug ID	Problem
BUG-22998	<p>HDFS-6376 allows <code>distcp</code> to copy data between HA clusters. Users can use a new configuration property <code>dfs.internal.nameservices</code> to explicitly specify the name services belonging to the local cluster, while continue using the configuration property <code>dfs.nameservices</code> to specify all of the name services in the local and remote clusters.</p> <p>Steps:</p> <p>Modify the following in the <code>hdfs-site.xml</code> for both cluster A and B:</p> <ol style="list-style-type: none"> Add both name services to <code>dfs.nameservices = HAA, HAB</code>

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

Table 1.11. JDK Support

Description
HDP 2.3 supports JDK 1.7.

1.5. Apache Patch Information

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

1.5.1. Apache Hadoop

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

NEW FEATURES

- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-2571](#) RM to support YARN registry
- [YARN-3345](#) Add non-exclusive node label API.
- [YARN-3365](#) Enhanced NodeManager to support using the 'tc' tool via container-executor for outbound network traffic control.
- [YARN-1376](#) NM need to notify the log aggregation status to RM through heartbeat.
- [YARN-3348](#) Add a 'yarn top' tool to help understand cluster usage.
- [YARN-3347](#) Improve YARN log command to get AMContainer logs as well as running containers logs.
- [YARN-3443](#) Create a 'ResourceHandler' subsystem to ease addition of support for new resource types on the NM.
- [YARN-3361](#) CapacityScheduler side changes to support non-exclusive node labels.
- [YARN-3318](#) Create Initial OrderingPolicy Framework and FifoOrderingPolicy.
- [YARN-3326](#) Support RESTful API for getLabelsToNodes.
- [YARN-3354](#) Add node label expression in ContainerTokenIdentifier to support RM recovery.
- [YARN-1402](#) Update related Web UI and CLI with exposing client API to check log aggregation status.
- [YARN-3463](#) Integrate OrderingPolicy Framework with CapacityScheduler.
- [YARN-3410](#) YARN admin should be able to remove individual application records from RMStateStore.
- [YARN-3225](#) New parameter of CLI for decommissioning node gracefully in RMAdmin CLI.

- [YARN-3366](#) Enhanced NodeManager to support classifying/shaping outgoing network bandwidth traffic originating from YARN containers
- [YARN-3319](#) Implement a FairOrderingPolicy.
- [YARN-2498](#) Respect labels in preemption policy of capacity scheduler for inter-queue preemption.
- [YARN-2619](#) Added NodeManager support for disk io isolation through cgroups.
- [YARN-3448](#) Added a rolling time-to-live LevelDB time line store implementation.
- [YARN-3541](#) Add version information on time line service / generic history web UI and REST API.
- [YARN-3505](#) Node's Log Aggregation Report with SUCCEED should not be cached in RMAppls.

IMPROVEMENTS

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

- [YARN-3503](#) Expose disk utilization percentage and bad local and log dir counts in NM metrics.
- [YARN-3511](#) Add errors and warnings page to ATS.
- [YARN-3406](#) Display count of running containers in the RM's Web UI.
- [YARN-3593](#) Add label-type and Improve "DEFAULT_PARTITION" in Node Labels Page.
- [YARN-3362](#) Add node label usage in RM CapacityScheduler web UI.
- [YARN-3565](#) NodeHeartbeatRequest/RegisterNodeManagerRequest should use NodeLabel object instead of String.
- [YARN-3583](#) Support of NodeLabel object instead of plain String in YarnClient side.
- [YARN-3581](#) Deprecate -directlyAccessNodeLabelStore in RMAAdminCLI.
- [YARN-3700](#) Made generic history service load a number of latest applications according to the parameter or the configuration.

BUG FIXES

- [HDFS-27](#) HDFS CLI with `--config` set to default configuration complains log file not found error.
- [HDFS-7890](#) Improve information on Top users for metrics in RollingWindowsManager and lower log level.
- [HDFS-8229](#) LAZY_PERSIST file gets deleted after NameNode restart.
- [HDFS-8276](#) LazyPersistFileScrubber should be disabled if scrubber interval configured zero.
- [HDFS-8152](#) Refactoring of lazy persist storage cases.
- [HDFS-8144](#) Split TestLazyPersistFiles into multiple tests.
- [HDFS-8219](#) `setStoragePolicy` with folder behavior is different after cluster restart.
- [HDFS-8232](#) Missing datanode counters when using Metrics2 sink interface.
- [HDFS-8205](#) `CommandFormat#parse()` should not parse option as value of option.
- [HDFS-8211](#) DataNode UUID is always null in the JMX counter.
- [HDFS-7990](#) IBR delete ack should not be delayed.
- [HDFS-7645](#) Fix CHANGES.txt
- [HDFS-7645](#) Rolling upgrade is restoring blocks from trash multiple times
- [HDFS-8055](#) NullPointerException when topology script is missing.
- [HDFS-7933](#) fsck should also report decommissioning replicas.

- [HDFS-6666](#) Abort NameNode and DataNode start-up if security is enabled but block access token is not enabled.
- [HADOOP-11859](#) PseudoAuthenticationHandler fails with httpcomponents v4.4.
- [HDFS-7701](#) Support reporting per storage type quota and usage with hadoop/hdfs shell.
- [HADOOP-7713](#) dfs -count -q should label output column
- [HDFS-8008](#) Support client-side back off when the datanodes are congested.
- [HDFS-8009](#) Signal congestion on the DataNode.
- [YARN-3305](#) Normalize AM resource request on app submission.
- [YARN-3269](#) Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.
- [YARN-3383](#) AdminService should use "warn" instead of "info" to log exception when operation fails.
- [YARN-3425](#) NPE from RMNodeLabelsManager.serviceStop when NodeLabelsManager.serviceInit failed.
- [YARN-3435](#) AM container to be allocated Appattempt AM container shown as null.
- [YARN-2666](#) TestFairScheduler.testContinuousScheduling fails Intermittently.
- [YARN-3110](#) Few issues in ApplicationHistory web UI.
- [YARN-3459](#) Fix failure of TestLog4jWarningErrorMetricsAppender.
- [YARN-3266](#) RMContext#inactiveNodes should have NodeId as map key.
- [YARN-3136](#) Fixed a synchronization problem of AbstractYarnScheduler#getTransferredContainers.
- [YARN-3387](#) Previous AM's container completed status couldn't pass to current AM if AM and RM restarted during the same time.
- [YARN-3530](#) ATS throws exception on trying to filter results without otherinfo.
- [YARN-2740](#) Fix NodeLabelsManager to properly handle node label modifications when distributed node label configuration enabled.
- [YARN-3517](#) RM web UI for dumping scheduler logs should be for admins only
- [YARN-3343](#) Increased TestCapacitySchedulerNodeLabelUpdate#testNodeUpdate timeout.
- [YARN-2821](#) Fixed a problem that DistributedShell AM may hang if restarted.
- [YARN-3654](#) ContainerLogsPage web UI should not have meta-refresh.
- [YARN-3552](#) RM Web UI shows -1 running containers for completed apps

- [YARN-3580](#) [JDK8] TestClientRMService.testGetLabelsToNodes fails.
- [YARN-3707](#) RM Web UI queue filter doesn't work.
- [YARN-3632](#) Ordering policy should be allowed to reorder an application when demand changes.
- [YARN-3740](#) Fixed the typo in the configuration name: APPLICATION_HISTORY_PREFIX_MAX_APPS.

1.5.2. Calcite

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

1.5.3. Falcon

HDP 2.3 provides Falcon 0.6.1 with no additional Apache patches.

1.5.4. Flume

HDP 2.3 provides Flume 1.5.2 and the following Apache patches:

NEW FEATURES

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

IMPROVEMENTS

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

- [FLUME-2095](#) JMS source with TIBCO (patch-1)

BUG FIXES

- [FLUME-2451](#) HDFS Sink Cannot Reconnect After NameNode Restart
- [FLUME-2407](#) Spillable Channel sometimes fails on reconfigure
- [FLUME-2358](#) File Channel needs to close BackingStore and EventQueue before deleting files in checkpoint directory
- [FLUME-2402](#) Warning seen when overflow is disabled for Spillable Channel
- [FLUME-2412](#) Improve Logging in Spillable Channel
- [FLUME-2122](#) Minor cleanups of User guide
- [FLUME-2175](#) Update Developer Guide with notes on how to upgrade Protocol Buffer version
- [FLUME-2162](#) TestHDFSEventSinkOnMiniCluster.maxUnderReplicationTest fails on hadoop2
- [FLUME-2530](#) Resource leaks found by Coverity tool
- [FLUME-2541](#) Bug in TestBucketWriter.testSequenceFileCloseRetries
- [FLUME-2441](#) Unit test TestHTTPSSource.java failed with IBM JDK 1.7
- [FLUME-2520](#) HTTP Source should be able to block a prefixed set of protocols.
- [FLUME-2533](#) HTTPS tests fail on Java 6

1.5.5. HBase

HDP 2.3 provides HBase 1.1.1 and the following Apache patches:

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

1.5.6. Hive

HDP 2.3 provides Hive 1.2.1 and the following Apache patches:

INCOMPATIBLE CHANGES

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

NEW FEATURES

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

IMPROVEMENTS

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

BUG FIXES

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

1.5.7. Knox

HDP 2.3 provides Knox 0.6.0 and the following Apache patches:

BUG FIXES

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

IMPROVEMENTS

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

1.5.8. Mahout

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

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

In addition, we have provided the following patch:

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

1.5.9. Oozie

HDP 2.3 provides Oozie 4.2.0 and the following Apache patches:

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

1.5.10. Phoenix

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

- [PHOENIX-2032](#) psql.py is broken after [PHOENIX-2013](#)
- [PHOENIX-2033](#) PQS log environment details on launch
- [PHOENIX-2007](#) java.sql.SQLException: Encountered exception in sub plan [0] execution'
- [PHOENIX-2027](#) Queries with Hints are raising IllegalStateException
- [PHOENIX-2012](#) RowKeyComparisonFilter logs un-encoded data at DEBUG level
- [PHOENIX-2010](#) Properly validate number of arguments passed to the functions in FunctionParseNode#validate
- [PHOENIX-2013](#) Apply [PHOENIX-1995](#) to runnable uberjar as well

- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode (addendum)
- [PHOENIX-2005](#) Connection utilities omit zk client port, parent znode
- [PHOENIX-1996](#) Use BytesStringer instead of ZeroCopyByteString
- [PHOENIX-1995](#) client uberjar does not support dfs
- [PHOENIX-1980](#) CsvBulkLoad cannot load hbase-site.xml from classpath
- [PHOENIX-1976](#) Exit gracefully if addShutdownHook fails.
- [PHOENIX-914](#) Native HBase timestamp support to optimize date range queries in Phoenix

1.5.11. Pig

HDP 2.3 provides Pig 0.15.0 and the following Apache patch:

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

1.5.12. Ranger

HDP 2.3 provides Ranger 0.5.0 and the following Apache patches:

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

BUG

- [RANGER-219](#) Autocomplete behavior of hive tables/columns
- [RANGER-524](#) Hbase plugin: list command should prune the tables returned on user permissions
- [RANGER-529](#) Policy Validation: resources of a policy must match one of the resource hierarchies of the service def.
- [RANGER-533](#) Hbase plugin: if user does not have family-level access to any family in a table then user may be incorrectly denied access done at table/family level during get or scan
- [RANGER-539](#) Rolling downgrade changes

- [RANGER-545](#) Fix js error for lower versions of FF (less than 30)
- [RANGER-548](#) Key rollover command fails
- [RANGER-550](#) Hive plugin: Add audit logging support for metadata queries that have filtering support from hive
- [RANGER-553](#) Default policy creation during service creation should handle service defs with multiple hierarchies, e.g. hive, properly
- [RANGER-554](#) Ranger KMS keys listing page does not support pagination
- [RANGER-555](#) Policy view page (from access audit page) gives 404 with Oracle DB
- [RANGER-558](#) Hbase plugin: unless user has READ access at some level under the table/family being accessed (via scan/get) authorizer should throw an exception and audit
- [RANGER-565](#) Ranger Admin install fails (sometimes) with IO Error when DB used in Oracle
- [RANGER-566](#) Installation of Ranger on Oracle 12c with shared database needs to use private synonym instead of public synonym
- [RANGER-569](#) Enabling Ranger plugin for Hbase should not modify hbase.rpc.protection value
- [RANGER-570](#) Knox plugin: after upgrading ranger from 0.4 to 0.5 the Knox plugin does not work because classes with old names are missing
- [RANGER-571](#) Storm plugin: after upgrading ranger from 0.4 to 0.5 the plugin does not 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

1.5.13. Slider

HDP 2.3 provides Slider 0.80.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [SLIDER-810](#) YARN configuration 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

1.5.14. Spark

HDP 2.3 provides Spark 1.3.1 and the following Apache patches:

IMPROVEMENTS

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

1.5.15. Sqoop

HDP 2.3 provides Sqoop 1.4.6 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

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

1.5.16. Storm

HDP 2.3 provides Storm 0.10.0 and the following Apache patches:

- [STORM-583](#) Add Microsoft Azure Event Hub spout implementations
- [STORM-708](#) CORS support for STORM UI.

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

1.5.17. Tez

HDP 2.3 provides Tez 0.7.0 and the following Apache patches:

IMPROVEMENTS

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

BUG FIXES

- [TEZ-2568](#) auto_sortmerge_join_5 fails in Tez mode
- [TEZ-2548](#) TezClient submitDAG can hang if the AM is in the process of shutting down.
- [TEZ-2475](#) Fix a potential hang in Tez local mode caused by incorrectly handled interrupts.
- [TEZ-2554](#) Tez UI: View log link does not correctly propagate login credential to read log from YARN web.
- [TEZ-2547](#) Tez UI: Download Data fails on secure, cross-origin clusters
- [TEZ-2546](#) Tez UI: Fetch hive query text from timeline if dagInfo is not set.
- [TEZ-2513](#) Tez UI: Allow filtering by DAG ID on All dags table.
- [TEZ-2541](#) DAGClientImpl enable TimelineClient check is wrong.
- [TEZ-2539](#) Tez UI: Pages are not updating in IE.
- [TEZ-2535](#) Tez UI: Failed task attempts link in vertex details page is broken.
- [TEZ-2489](#) Disable warn log for Timeline ACL error when tez.allow.disabled.timeline-domains set to true.
- [TEZ-2528](#) Tez UI: Column selector buttons gets clipped, and table scroll bar not visible in mac.

- [TEZ-2391](#) TestVertexImpl timing out at times on Jenkins builds.
- [TEZ-2509](#) YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.
- [TEZ-2527](#) Tez UI: Application hangs on entering erroneous RegEx in counter table search box
- [TEZ-2523](#) Tez UI: derive applicationId from dag/vertex id instead of relying on Json data
- [TEZ-2505](#) PipelinedSorter uses Comparator objects concurrently from multiple threads.
- [TEZ-2504](#) Tez UI: tables - show status column without scrolling, numeric 0 shown as Not available
- [TEZ-2478](#) Move OneToOne routing to store events in Tasks.
- [TEZ-2482](#) Tez UI: Mouse events not working on IE11
- [TEZ-1529](#) ATS and TezClient integration in secure Kerberos enabled cluster.
- [TEZ-2481](#) Tez UI: graphical view does not render properly on IE11
- [TEZ-2474](#) The old taskNum is logged incorrectly when parallelism is changed
- [TEZ-2460](#) Temporary solution for issue due to [YARN-2560](#)
- [TEZ-2455](#) Tez UI: Dag view caching, error handling and minor layout changes
- [TEZ-2453](#) Tez UI: show the dagInfo is the application has set the same.
- [TEZ-2447](#) Tez UI: Generic changes based on feedbacks.
- [TEZ-2409](#) Allow different edges to have different routing plugins

1.6. Fixed Issues

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

Potential Data Loss

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

Security

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-30108	RANGER-375	Ranger	Ranger - when backend DB not reachable error thrown is "incorrect password/"

Hortonworks Bug ID	Apache JIRA	Component	Summary
			username" when trying to log into Ranger
BUG-30856	HBASE-11869	HBase	Support snapshot owner other than Global Admins in Hbase
BUG-31425		Ranger	PolicyManager throws 404 error when creating HBase repo in Ranger
BUG-31545		Falcon Documentation	Add details of Falcon directory permissions to documentation
BUG-32768	FALCON-954	Falcon	Secure Kerberos setup : Falcon should periodically revalidate security credentials.
BUG-33278	HBASE-13239	HBase	HBASE grants at specific column level does not work for Groups
BUG-33338	HIVE-10528	Hive	Hiveserver2 in HTTP mode is not applying auth_to_local rules
BUG-33621		Oozie	Clear Text Password Shows In Oozie workflow configuration
BUG-33708	KNOX-525	Knox	persisted service registry is not updated to support HA after upgrade
BUG-33819	RANGER-483	Ranger	Provide an option to create password using alternate message digest algorithm
BUG-35992	HIVE-10875	Hive	SELECT a.* FROM (SELECT * FROM source_view) a results in permission denied error.
BUG-36431		Knox, Ranger	Found Jersey client API incompatibility that prevent HTTPS Knox from working
BUG-36565	HBASE-13734	HBase	deleteall behavior changes after applying org.apache.hadoop.hbase.security.visibility.VisibilityController to HBase
BUG-41100		YARN, HADOOP	Yarn services Tomcat instance upgraded to 6.0.44

Incorrect Results

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-22370	HIVE-5545	HCatalog	HCatRecord getInteger method returns String when used on Partition columns of type INT
BUG-31914	HIVE-9278	Hive	arithmetic operators return incorrect results from certain operand types
BUG-33275	HIVE-8746	Hive	ORC timestamp columns are sensitive to daylight savings time

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-34138		Pig	Receiving exception 'No such file or directory' when using 'pig -useHCatalog'
BUG-34210	HIVE-9278	Hive	multiple built-in date functions return incorrect results in same where clause
BUG-34506	PHOENIX-896	Phoenix	Unable to load tab delimited data via Phoenix
BUG-34971	HIVE-10481	Hive	ACID table update finishes but values not really updated if column names are not all lower case
BUG-36223		Hive	wrong results for 2 left outer joins with overlapping join on keys, filter IS NULL pushed to left side table wrongly.
BUG-36853	PIG-4541	Pig	Skewed full outer join does not return records if any relation is empty. Outer join does not return any record if left relation is empty
BUG-37485	PIG-4556	Pig	Pig tests abort with Check of prerequisites failed: <Failed running /usr/bin/pig -e fs -ls />
BUG-40100	HIVE-11104	Hive	Insert overwrite query does not return expected value

Stability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-21924	HBASE-13555	HBase	HBase web gives a 500 error when you attempt to view table details and it's not the master
BUG-26984	OOZIE-1728	Oozie	Queue configuration is not working for distcp action
BUG-29350	HBASE-12791	HBase	HBase does not attempt to clean up an aborted SPLIT when the Regions Server is shutting down.
BUG-33349	HBASE-13608	HBase	500 Error with Stargate through Knox, using AD, SPNEGO, and Pre-Auth
BUG-33488	KNOX-530	Knox	Running Oozie jobs through Knox on a cluster with HDFS HA does not use proper namenode host name.
BUG-34020	FALCON-1165	Falcon	Falcon will fail to start, if a cluster entity that was defined is not reachable.
BUG-34175	HDFS-8072	HDFS	Non-HDFS disk space is not reclaimed until after Datanode restart
BUG-5700		HDFS	NameNode should shut down if out of memory error occurs

Query Failure

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-27636	HIVE-10500	Hive	intermittent acid_concurrency test failures due to NoSuchLockException
BUG-29427	HIVE-9235	Hive	Turn off Parquet Vectorization until all data types work: DECIMAL, DATE, TIMESTAMP, CHAR, and VARCHAR
BUG-30038		HCatalog	Sqoop Import on External Hcatalog Table
BUG-30901	HIVE-10559	Hive	with constant propagation off, left join and join, hive.tez.dynamic.partition.pruning throw compile error IndexOutOfBoundsException
BUG-33145		Hive	HS2 HTTP Mode - Beeline Hanging after upgrading from HDP 2.1 to HDP 2.2
BUG-33857	HIVE-10273	Hive	Select * on Table View (with UNION ALL) throws NullPointerException if Tez execution engine is used
BUG-34809	HIVE-10242	Hive	ACID: insert overwrite prevents create table command
BUG-34872	HIVE-10559	Hive	Hive tez dynamic pruning throws IndexOutOfBoundsException exception on certain queries
BUG-34956	Hive-10677	Hive	Analyze table compute stats for columns, ColumnStatsTask fail when hive.exec.parallel=true (default false)
BUG-35048	HIVE-10483	Hive	insert overwrite table with self join gets into a deadlock state in ACID DBTxnManager
BUG-35795	HIVE-10286	Hive	orc ppd does not typecast of string to timestamp when evaluate predicate for timestamp column
BUG-36111	HIVE-8470	Hive	Orc writer cant handle column of type void
BUG-37304	HIVE-9950	Hive	Hive query using Cuckoo hashing causing NullPointerException
BUG-37429	HIVE-9937	Hive	vectorization with ACID table, count(1) fails with ArrayIndexOutOfBoundsException
BUG-38292	HIVE-10929	Hive	In Tez mode,dynamic partitioning query with union all fails at moveTask,Invalid partition key & values

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-38817	HIVE-11031	Hive	Hive Alter table, concatenate partition files, throws error
BUG-39159	HIVE-11027	Hive	Hive Incorrect query results with Bucket Map joins vs Shuffle Joins.
BUG-39868	HIVE-11051	Hive	Tez Optimized MapJoin: Tables with Array<String> causes task failures
BUG-40036	PIG-4624	Hive	empty ORC file without schema produced by Hive

Upgrade

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-34508	HDFS-8127	HDFS	NameNode Failover during HA upgrade can cause DataNode to finalize upgrade that will fail datanode -rollback

Usability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-19508	YARN-2246	YARN	Job History Link in RM UI is redirecting to the URL which contains Job Id twice
BUG-21763	RANGER-451	Ranger	Policy Admin User/Group page doesn't show groups for users when the user belongs to large number of groups
BUG-22120	RANGER-244	Ranger	Option to disable/hide imported users/groups
BUG-23032		Hive	Jars added in Hive shell whilst in Tez mode are not recognized
BUG-27558	YARN-3526	YARN	ResourceManager probe on standby server in HA mode doesn't complete the redirect
BUG-28060	RANGER-281	Ranger	Support for PostgreSQL as DB server for XA Secure (Ranger, Argus)
BUG-29616	PIG-4381	Pig	On PIG grunt shell DEFINE commands fails when it spans multiple lines.
BUG-29899	HIVE-9223	Hive	HS2/tez: multiple concurrent queries not allowed in single tez session
BUG-29900	RANGER-318	Ranger	Not able to add user with only numbers
BUG-30186	HIVE-9381	HCatalog	Cannot load data into specific Hive table via pig HCatStorer
BUG-31113		Hive	Hive "SLF4J: Class path contains multiple SLF4J bindings." error

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-31749	RANGER-320	Ranger	Usersync NPE when object does not have userNameAttribute
BUG-31782	RANGER-337	Ranger	Ranger gives error when using hyphen or space in name field
BUG-31899	HIVE-10271	Hive	remove hive.server2.thrift.http.min/max.worker.threads properties
BUG-32859	HIVE-9977	Hive	Hive metastore auto major compaction not happening if data is loaded through dynamic partitioning insert
BUG-33876	HIVE-10226	Hive	ANALYZE TABLE (compute_stats) UDAF doesn't have Date support
BUG-35141		Falcon	Regression in the Falcon UI; Only 10 results shown for all entities.
BUG-35285	HIVE-3682 , HIVE-5672	Hive	cannot use separator of choice for Hive export in HDFS
BUG-35338	HBASE-13555	HBase	HBase web gives a 500 error when you attempt to view Thread Stacks
BUG-35511		Hive	hive.support.concurrency=false doesn't turn of Transactions
BUG-35598	FLUME-2095	Flume	JMS source setup fails with "org.apache.flume.FlumeException: Could not lookup ConnectionFactory"
BUG-39567	SPARK-8383	Spark	Spark History Server shows Last Updated as 1969/12/31 when SparkPI application completed
BUG-7691		WebHCat	Provide/extend API to support table views
BUG-9512	YARN-2238	YARN	Applications sometimes are not visible in RM web UI

Performance

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-21901	RANGER-192	Ranger	XASecure loading groups for user hang if there are too many groups for the user
BUG-23856	TEZ-14	Tez	Support for speculative execution of slow tasks
BUG-25165	HIVE-9709	Hive	Performance Issue Beeline > Knox > HS2
BUG-33592	HIVE-9644 , HIVE-9645	Hive	case/when query not optimized, partition pruning not happening
BUG-33853	PIG-4488	Pig	Pig on Tez only uses default queue

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-34139		Ranger	Ranger is running out of database connection when looking up user/groups
BUG-37846		Hive	unix_timestamp non-deterministic causing less performance, current_timestamp not available in HDP 2.2.x
BUG-38249	HDFS-7435	HDFS	PB encoding of block reports is very inefficient
BUG-39394	HIVE-10746	Hive, Tez	Mapred.TextInputFormat defaults to 1 byte FileSplits

Other

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-31287	HDFS-8055	HDFS	NullPointerException when topology script is missing.
BUG-33764		HDP / Stack	hdps-select needs to handle user created directories in /usr/hdp

1.7. Known Issues

Table 1.12. Apache HBase

Apache JIRA	
Hortonworks Bug ID	BUG-42355
Description	Moved application from HDP 2.2 to HDP 2.3 and now ACLs don't appear to be functioning the same Workaround: Set <code>hbase.security.access.early_out=false</code> , as in the following example: <pre><property> <name>hbase.security.access.early_out</name> <value>false</value> </property></pre>
Apache JIRA	HBASE-13330 , HBASE-13647
Hortonworks Bug ID	BUG-36817
Description	test_IntegrationTestRegionReplica Replication[IntegrationTestRegion ReplicaReplication] fails with READ FAILURES
Apache JIRA	
Hortonworks Bug ID	BUG-39322
Description	The HBase bulk load process is a MapReduce job that typically runs under the user ID who owns the source data. HBase data files created as a result of the job are then bulk-loaded into HBase RegionServers. During this process, HBase RegionServers move the bulk-loaded files from the user's directory, and moves (renames) the files under the HBase <code>root.dir (/apps/hbase/data)</code> . When HDFS data encryption is used, HDFS cannot rename across encryption zones with different keys.

	Workaround: Run the MapReduce job as the <code>hbase</code> user, and specify an output directory in the same encryption zone as the HBase root directory.
Apache JIRA	HBASE-13832 , HDFS-8510
Hortonworks Bug ID	BUG-40536
Description	<p>When rolling upgrade is performed for HDFS, sometimes the HBase Master might run out of datanodes on which to keep its write-pipeline active. When this occurs, the HBase Master Aborts after a few attempts to keep the pipeline going. To avoid this situation:</p> <p>Workaround:</p> <ol style="list-style-type: none"> 1. Before performing the rolling upgrade of HDFS, update the HBase configuration by setting <code>"dfs.client.block.write.replace-datanode-on-failure.best.effort"</code> to true. 2. Restart the HBase Master. 3. Perform the rolling upgrade of HDFS. <p>Undo the configuration change done in Step 1.</p> <p>Restart the HBase Master.</p> <p><i>Note:</i> There is a window of time during the rolling upgrade of HDFS when the HBase Master might be working with just one node and if that node fails, the WAL data might be lost. In practice, this is an extremely rare situation.</p> <p>Alternatively, the HBase Master can be turned off during the rolling upgrade of HDFS to avoid the above procedure. If this strategy is taken, client DDL operations and RegionServer failures cannot be handled during this time.</p> <p>A final alternative if the HBase Master fails during rolling upgrade of HDFS, a manual start can be performed.</p>
Apache JIRA	
Hortonworks Bug ID	BUG-42186
Description	<p>HDP 2.3 HBase install needs MapReduce class path modified for HBase functions to work</p> <p>Cluster that have Phoenix enabled placed the following config in <code>hbase-site.xml</code>:</p> <pre>Property: hbase.rpc.controllerfactory.class Value:org.apache.hadoop.hbase.ipc.controller.ServerRpcControllerFactory</pre> <p>This property points to a class found only in <code>phoenix-server.jar</code>. To resolve this class at run time for the above listed Mapreduce Jobs, it needs to be part of the MapReduce classpath.</p> <p>Workaround: Update <code>mapreduce.application.classpath</code> property in <code>mapred-site.xml</code> file to point to <code>/usr/hdp/current/phoenix-client/phoenix-server.jar</code> file.</p>

Table 1.13. Apache Hive

Apache JIRA	HIVE-11587
--------------------	----------------------------

Hortonworks Bug ID	BUG-42500
Description	<p>Hive Hybrid Grace MapJoin can cause OutOfMemory Issues</p> <p>Hive Hybrid Grace Mapjoin is a new feature in HDP 2.3 (Hive 1.2). Mapjoin joins two tables, holding the smaller one in memory. Grace Hybrid Mapjoin spills parts of the small table to disk when the Map Join does not fit in memory at runtime. Right now there is a bug in the code that can cause this implementation to use too much memory, causing an OutOfMemory error. This applies to the Tez execution engine only.</p> <p>Workaround: Turn off hybrid grace map join by setting this property in hive-site.xml:</p> <ul style="list-style-type: none"> • Navigate to Hive>Configs>Advanced>Custom hive-site. • Set <code>hive.mapjoin.hybridgrace.hashtable=false</code>.
Apache JIRA	HIVE-11110
Hortonworks Bug ID	BUG-39988
Description	CBO: Default partition filter is from MetaStore query causing TPC-DS to regress by 3x.
Apache JIRA	
Hortonworks Bug ID	BUG-39412
Description	<p>Users should not use <code>datanucleus.identifierFactory = datanucleus2</code> in hive config.</p> <p>Setting <code>datanucleus.identifierFactory</code> to <code>datanucleus2</code> can potentially lead to data corruption if <code>directSql</code> is enabled. Avoid using this setting if you are setting up a new metastore. If you are migrating an old metastore with this configuration parameter already set, contact Support for a few steps to address the issue.</p>
Apache JIRA	HIVE-10978
Hortonworks Bug ID	BUG-39282
Description	<p>When HDFS is encrypted (data at rest encryption is enabled) and the Hadoop Trash feature is enabled, DROP TABLE and DROP PARTITION have unexpected behavior.</p> <p>(The Hadoop Trash feature is enabled by setting <code>fs.trash.interval > 0</code> in <code>core-site.xml</code>.)</p> <p>When Trash is enabled, the data file for the table should be "moved" to the Trash bin, but if the table is inside an Encryption Zone, this "move" operation is not allowed.</p> <p>Workaround: Here are two ways to work around this issue:</p> <ol style="list-style-type: none"> 1. Use PURGE, as in DROP TABLE ... PURGE. This skips the Trash bin even if Trash is enabled. 2. set <code>fs.trash.interval = 0</code>. Caution: this configuration change must be done in <code>core-site.xml</code>. Setting it in <code>hive-site.xml</code> may lead to data corruption if a table with the same name is created later.

Apache JIRA	
Hortonworks Bug ID	BUG-38785
Description	<p>With RHEL7, the <code>cpu</code> and <code>cpuacct</code> controllers are managed together by default. The default directory is <code>/sys/fs/cgroup/cpu,cpuacct</code>. The presence of the comma leads to failures when initializing the NodeManager (when using the LinuxContainerExecutor).</p> <p>Workaround: Create your own directory (such as <code>/sys/fs/cgroup/hadoop/cpu</code>) and set <code>yarn.nodemanager.linux-container-executor.cgroups.mount</code> to <code>true</code>. This will allow the NodeManager to mount the <code>cpu</code> controller, and YARN will be able to enforce CPU limits for you.</p> <p>If you wish to mount the <code>cgroups</code> yourself (or provide a mount point), please set <code>yarn.nodemanager.linux-container-executor.cgroups.mount</code> to <code>false</code> and ensure that the hierarchy specified in <code>yarn.nodemanager.linux-container-executor.cgroups.hierarchy</code> exists in the mount location. Make sure there are no commas in your pathnames.</p>
Apache JIRA	
Hortonworks Bug ID	BUG-37042
Description	<p>Limitations while using <code>timestamp.formats serde</code> parameter.</p> <p>Two issues involving the <code>timestamp.formats SerDe</code> parameter:</p> <ul style="list-style-type: none"> • Displays only 3 decimal digits when it returns values, but it accepts more decimal digits. <p>For example, if you run the following commands:</p> <pre>drop table if exists src_hbase_ts; create table src_hbase_ts(rowkey string, ts1 string, ts2 string, ts3 string, ts4 string) STORED BY 'org.apache.hadoop.hive. hbase. HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4') TBLPROPERTIES ('hbase.table.name' = 'hbase_ts'); insert into src_hbase_ts values ('1','2011-01-01T01:01: 01.111111111', '2011-01-01T01:01: 01.123456111', '2011-01-01T01:01: 01.111111111', '2011-01-01T01:01: 01.134567890'); drop table if exists hbase_ts_1; create external table hbase_ts_1(rowkey string, ts1 timestamp, ts2 timestamp, ts3 timestamp, ts4 timestamp) STORED BY 'org.apache.hadoop.hive. hbase. HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.columns.mapping' = 'm:ts1,m:ts2,m:ts3,m:ts4', 'timestamp.formats' = "yyyy-MM-</pre>

```
dd'T'HH:mm:ss.SSSSSSSS") TBLPROPERTIES
('hbase.table.name' = 'hbase_ts');

select * from hbase_ts_1;
```

The timestamp.formats parameter displays:

```
1 2011-01-01 01:01:01.111 2011-01-01 01:01:01.
123 2011-01-01 01:01:01.111 2011-01-01 01:01:01.
134
```

When the expected output is:

```
1 2011-01-01 01:01:01.111111111 2011-01-01
01:01:01.123456111 2011-01-01 01:01:01.
111111111 2011-0
```

- The yyyy-MM-dd'T'HH:mm:ss.SSSSSSSS format accepts any timestamp data up to .SSSSSSSS decimal digits (9 places to the left of the decimal) instead of only reading data with .SSSSSSSS decimal digits (9 places to the left of the decimal).

For example, if you run the following commands:

```
drop table if exists src_hbase_ts; create
table src_hbase_ts( rowkey string, ts1
string, ts2 string, ts3 string, ts4
string ) STORED BY 'org.apache.hadoop.
hive. hbase.HBaseStorageHandler' WITH
SERDEPROPERTIES ('hbase.columns.mapping'
= 'm:ts1,m:ts2,m:ts3,m:ts4')
TBLPROPERTIES ('hbase.table.name' =
'hbase_ts');
```

```
insert into src_hbase_ts values
('1','2011-01-01T01:01: 01.111111111',
'2011-01-01T01:01: 01.111',
'2011-01-01T01:01: 01.11',
'2011-01-01T01:01:01.1');
```

```
drop table if exists hbase_ts_1;
```

```
create external table hbase_ts_1( rowkey
string, ts1 timestamp, ts2 timestamp,
ts3 timestamp, ts4 timestamp )
STORED BY 'org.apache.hadoop. hive.
hbase.HBaseStorageHandler' WITH
SERDEPROPERTIES ( 'hbase.columns.mapping'
= 'm:ts1,m:ts2,m:ts3,m:ts4',
'timestamp.formats' = "yyyy-MM-
dd'T'HH:mm:ss.SSSSSSSS") TBLPROPERTIES
('hbase.table.name' = 'hbase_ts');
```

```
select * from hbase_ts_1;
```

The actual output is:

```
1 2011-01-01 01:01:01.111 2011-01-01 01:01:01.
111 2011-01-01 01:01:01.11 2011-01-01 01:01:01.1
```

When the expected output is:

```
1 2011-01-01 01:01:01.111 NULL NULL NULL
```

Table 1.14. Apache Oozie

Apache JIRA	OOZIE-2311
Hortonworks Bug ID	BUG-39265

Description	NPE in oozie logs while running feed replication tests causes jobs to fail.
-------------	---

Table 1.15. Apache Ranger

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

Table 1.16. Apache Slider

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

Table 1.17. Apache Spark

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

Apache JIRA	
Hortonworks Bug ID	BUG-39674
Description	Spark does not yet support wire encryption, dynamic executor allocation, SparkR, GraphX, Spark Streaming, iPython, or Zeppelin.

Table 1.18. Apache Tez

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

Table 1.19. Apache YARN

Apache JIRA	YARN-2194
Hortonworks Bug ID	BUG-39424
Description	NM fails to come with error "Not able to enforce cpu weights; cannot write to cgroup."
Apache JIRA	
Hortonworks Bug ID	BUG-39756
Description	NM web UI cuts <code>?user.name</code> when redirecting URL to MR JHS.
Apache JIRA	
Hortonworks Bug ID	BUG-35942
Description	<p>Users must manually configure ZooKeeper security with ResourceManager High Availability.</p> <p>Right now, the default value of <code>yarn.resourcemanager.zk-acl</code> is <code>world:any:rwcd</code>. That means anyone can read/write/create/delete/setPermission for the znode which is not secure and not acceptable.</p> <p>To make it more secure, we can rely on Kerberos to do the authentication for us. We could configure sasl authentication and only Kerberos authenticated user can access to <code>zkrmstatestore</code>.</p> <p>ZooKeeper Configuration</p> <p><i>Note:</i> This step of securing ZooKeeper is to be done <i>once</i> for the HDP cluster. If this has been done to secure HBase, for example, then you do not need to repeat these ZooKeeper steps if Apache YARN ResourceManager High Availability is to use the same ZooKeeper.</p>

1. Create a keytab for zookeeper called `zookeeper.service.keytab` and save it in `/etc/security/keytabs`.

2. Add following contents in `zoo.cfg`:

```
authProvider.1=org.apache.zookeeper.server.auth.SASLAuthenticationProvider
jaasLoginRenew=360000
kerberos.removeHostFromPrincipal=true
kerberos.removeRealmFromPrincipal=true
```

3. Create `zookeeper_client_jaas.conf`:

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

4. Create `zookeeper_jaas.conf`:

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

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

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

Apache YARN Configuration

The following applies to HDP 2.2 and HDP 2.3.

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

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

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

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

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

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

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

HDFS Configuration

Note: This applies to HDP 2.1, 2.2, and 2.3.

1. In `hdfs-site.xml`, set the following property, for security of ZooKeeper based fail-over controller, when NameNode HA is enabled:

```
<property>
  <name>ha.zookeeper.acl</name>
  <value>sasl:nn:rwcd</value>
</property>
```

Table 1.20. HDFS and Cloud Deployment

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

Table 1.21. Upgrade

Apache JIRA	HDFS-8782
Hortonworks Bug ID	BUG-41215
Description	<p>Upgrade to block ID-based DN storage layout delays DN registration.</p> <p>When upgrading from a pre-HDP-2.2 release, a DataNode with a lot of disks, or with blocks that have random block IDs, can take a long time (potentially hours). The DataNode will not register to the NameNode until it finishes upgrading the storage directory.</p>
Apache JIRA	
Hortonworks Bug ID	BUG-32401
Description	<p>Rolling upgrade/downgrade should not be used if truncate is turned on. Workaround: Before starting a rolling upgrade or downgrade process, turn truncate off.</p>