

# Hortonworks Data Platform

## Release Notes

(Jul 2, 2014)

## Hortonworks Data Platform : Release Notes

Copyright © 2012-2014 Hortonworks, Inc. All 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 on Hortonworks technology, see [the Hortonworks Data Platform](#). For more information on Hortonworks services, see either the [Support](#) or [Training](#) page. Feel free to [Contact Us](#) directly to discuss your specific needs.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

## Table of Contents

1. Release Notes HDP-2.1.3 .....	1
1.1. Product Version: HDP-2.1.3 .....	1
1.2. Unsupported Apache components .....	2
1.3. Tech Previews in This Release .....	3
1.4. Fixed in This Release .....	3
1.5. Behavioral Changes .....	3
1.5.1. YARN behavioral changes .....	4
1.5.2. Hive behavioral changes .....	4
1.6. Patch Information .....	4
1.6.1. Patch information for Hadoop Common/HDFS .....	4
1.6.2. Patch information for ZooKeeper .....	6
1.6.3. Patch information for HBase .....	6
1.6.4. Patch information for Pig .....	9
1.6.5. Patch information for Tez .....	10
1.6.6. Patch information for Hive/HCatalog .....	11
1.6.7. Patch information for Oozie .....	12
1.6.8. Patch information for Sqoop .....	13
1.7. Minimum System Requirements .....	15
1.7.1. Hardware recommendations .....	15
1.7.2. Operating systems requirements .....	15
1.7.3. Software requirements .....	16
1.7.4. Database requirements .....	16
1.7.5. Virtualization and cloud platforms .....	16
1.7.6. Configuring the local repositories .....	17
1.8. Upgrading From HDP 2.1.2 to HDP 2.1.3 .....	17
1.9. Improvements .....	20
1.10. Common Vulnerabilities and Exposures .....	20
1.11. Known Issues .....	21
1.11.1. Known Issues for HDP .....	21
1.11.2. Known Issues for YARN .....	22
1.11.3. Known Issues for HBase .....	23
1.11.4. Known Issues for Apache Pig .....	23
1.11.5. Known Issues for Hive .....	23
1.11.6. Known Issues for Hue .....	26
1.11.7. Known Issues for Storm .....	27
1.11.8. Known Issues for Knox .....	28
1.11.9. Known Issues for the Hortonworks Connector for Teradata .....	28
1.11.10. Known Issues for Ambari .....	28
1.12. Deprecated Features .....	28
1.13. Third-party Licenses .....	28
2. Release Notes HDP-2.1.2 .....	29
2.1. Product Version: HDP-2.1.2 .....	29
2.2. Unsupported Apache components .....	30
2.3. Fixed in This Release .....	31
2.4. Behavioral Changes .....	31
2.4.1. Mahout behavioral changes .....	31
2.4.2. HDP 2.1 clusters deployed via Ambari .....	32
2.4.3. Hue behavioral changes .....	32

2.4.4. HBase behavioral changes .....	32
2.4.5. Hive behavioral changes .....	32
2.4.6. Oozie behavioral changes .....	32
2.5. Patch Information .....	32
2.5.1. Patch information for Hadoop Common/HDFS .....	33
2.5.2. Patch information for ZooKeeper .....	33
2.5.3. Patch information for HBase .....	33
2.5.4. Patch information for Pig .....	37
2.5.5. Patch information for Tez .....	37
2.5.6. Patch information for Hive/HCatalog .....	38
2.5.7. Patch information for Oozie .....	39
2.5.8. Patch information for Sqoop .....	40
2.6. Minimum System Requirements .....	41
2.6.1. Hardware recommendations .....	42
2.6.2. Operating systems requirements .....	42
2.6.3. Software requirements .....	42
2.6.4. Database requirements .....	43
2.6.5. Virtualization and cloud platforms .....	43
2.6.6. Configuring the local repositories .....	43
2.7. Upgrading From HDP 2.1.1 to HDP 2.1.2 .....	43
2.8. Improvements .....	46
2.9. Common Vulnerabilities and Exposures .....	46
2.10. Known Issues .....	46
2.10.1. Known Limitation for HDFS Upgrade .....	47
2.10.2. Known Issues for SLES 11 .....	47
2.10.3. Known Issues for HDP .....	47
2.10.4. Known Issues for YARN .....	49
2.10.5. Known Issues for HBase .....	50
2.10.6. Known Issues for Phoenix .....	50
2.10.7. Known Issues for Hive .....	50
2.10.8. Known Issues for Tez .....	55
2.10.9. Known Issues for WebHCat .....	55
2.10.10. Known Issues for Oozie .....	56
2.10.11. Known Issues for Hue .....	57
2.10.12. Known Issues for Flume .....	57
2.10.13. Known Issues for Storm .....	57
2.10.14. Known Issues for Falcon .....	58
2.10.15. Known Issues for Knox .....	60
2.10.16. Known Issues for the Hortonworks Connector for Teradata .....	60
2.10.17. Known Issues for Ambari .....	60
2.11. Deprecated Features .....	61
2.12. Third-party Licenses .....	61
3. Release Notes HDP-2.1.1 .....	62
3.1. Product Version: HDP-2.1.1 .....	62
3.2. Unsupported Apache components .....	63
3.3. Behavioral Changes .....	64
3.3.1. Mahout behavioral changes .....	64
3.3.2. HDP 2.1 clusters deployed via Ambari .....	64
3.3.3. Hue behavioral changes .....	64
3.3.4. HBase behavioral changes .....	64
3.3.5. Hive behavioral changes .....	64

3.3.6. Oozie behavioral changes .....	65
3.4. Patch Information .....	65
3.4.1. Patch information for Hadoop Common/HDFS .....	65
3.4.2. Patch information for ZooKeeper .....	66
3.4.3. Patch information for HBase .....	66
3.4.4. Patch information for Pig .....	69
3.4.5. Patch information for Tez .....	69
3.4.6. Patch information for Hive/HCatalog .....	70
3.4.7. Patch information for Oozie .....	72
3.4.8. Patch information for Sqoop .....	73
3.5. Minimum System Requirements .....	74
3.5.1. Hardware recommendations .....	75
3.5.2. Operating systems requirements .....	75
3.5.3. Software requirements .....	75
3.5.4. Database requirements .....	76
3.5.5. Virtualization and cloud platforms .....	76
3.5.6. Configuring the local repositories .....	76
3.6. Improvements .....	76
3.7. Common Vulnerabilities and Exposures .....	77
3.8. Known Issues .....	77
3.8.1. Known Limitation for HDFS Upgrade .....	78
3.8.2. Known Limitation for Oracle DB Metastore .....	78
3.8.3. Known Issues for SLES 11 .....	78
3.8.4. Known Issues for HDP .....	78
3.8.5. Known Issues for YARN .....	79
3.8.6. Known Issues for HBase .....	80
3.8.7. Known Issues for Phoenix .....	81
3.8.8. Known Issues for Hive .....	82
3.8.9. Known Issues for Tez .....	86
3.8.10. Known Issues for Oozie .....	86
3.8.11. Known Issues for Hue .....	88
3.8.12. Known Issues for Flume .....	88
3.8.13. Known Issues for Storm .....	89
3.8.14. Known Issues for Falcon .....	89
3.8.15. Known Issues for Knox .....	91
3.8.16. Known Issues for the Hortonworks Connector for Teradata .....	91
3.8.17. Known Issues for Ambari .....	91
3.9. Deprecated Features .....	92
3.10. Third-party Licenses .....	92

# List of Tables

- 1.1. Operating Systems mapped to each OS Family ..... 19
- 1.2. Third-party Licenses ..... 28
- 2.1. Operating Systems mapped to each OS Family ..... 45
- 2.2. Third-party Licenses ..... 61
- 3.1. Third-party Licenses ..... 92

# 1. Release Notes HDP-2.1.3

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.3](#)
- [Unsupported Apache components](#)
- [Tech Previews in This Release](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum system requirements](#)
- [Upgrading From HDP 2.1.2 to HDP 2.1.3](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

## 1.1. Product Version: HDP-2.1.3

All HDP 2.1 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 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.



### Note

The minimum level of Apache Ambari to use with HDP 2.1, is version 1.5.1.

This release of Hortonworks Data Platform (HDP) deploys the following Hadoop-related components:

- Apache Hadoop 2.4.0
- Apache HBase 0.98.0
- Apache Pig 0.12.1

- Apache Hive 0.13.1
- Apache Tez 0.4.0
- Apache ZooKeeper 3.4.5
- Hue 2.3.4\*
- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5.0
- Apache Sqoop 1.4.4
- Apache Knox 0.4.0
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0

**Third party components:**

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

**Note**

If you have CentOS 6 and update to 2.1.4, see the [README](#) for any version updates.

## 1.2. Unsupported Apache components

Arvo v 1.7.4 is shipped with HDP 2.1, but is not supported.

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

The following Apache components are shipped as part of HDP 2.1 YARN, but are not supported:



- MapReduce v1 Fair Scheduler ([HADOOP-3746](#))
- YARN Fair Scheduler ([MAPREDUCE-3451](#))
- MapReduce Uber AM ([MAPREDUCE-2405](#))
- MapReduce Eclipse Plugin (for Non-Kerberos and Kerberos cluster)
- Cgroup resource isolation ([YARN-3](#))
- CPU Scheduling ([YARN-2](#))

## 1.3. Tech Previews in This Release

The following features are provided with HDP 2.1.3 as technical previews; that is, they are considered to be still under development, and as such not supported. Do not use the following features in your production systems.

- Preemption – this feature has been found to contain defects that make it unsuitable for a production cluster at this time. For information about Preemption, see [http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk\\_system-admin-guide/content/preemption.html](http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk_system-admin-guide/content/preemption.html).
- Falcon – lag time in displaying lineage information if graph contains many vertices; old graph persists until updated graph is loaded
- Falcon – dependencies, viewed in Graph View

To contact Hortonworks Technical Support, please log a case at <https://support.hortonworks.com/>. If you are currently not an official Hortonworks Customer or Partner, please seek assistance on our Hortonworks Forums at <http://hortonworks.com/community/forums/>.

## 1.4. Fixed in This Release

HDP 2.1.3 is a maintenance release that includes scores of bug fixes and optimizations. It is notable for resuming support for Ubuntu 12.0.4.



### Note

For CentOS 6 operating systems there is a 2.1.4 patch update. For information, see the 2.1.4 [README](#).

## 1.5. Behavioral Changes

The following Apache Components Changed in HDP 2.1.3:

- [What's Changed in YARN](#)
- [What's Changed in Hive](#)

## 1.5.1. YARN behavioral changes

In HDP 2.1.3, when installing with Ambari 1.6.0, go to the Ambari **Services > YARN > Configs** menus and change the value of the YARN `yarn.timeline-service.store-class` property to `org.apache.hadoop.yarn.server.timeline.LevelDbTimelineStore`.

This change is required to ensure that YARN starts during cluster installation.

## 1.5.2. Hive behavioral changes

In HDP 2.1.3 (Hive 0.13.0) the Decimal data type is now treated as the type `Decimal(10,0)`: 10 digits of precision and 0 scale. This is a change from the variable precision and scale that was available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

In order to avoid unintended "rounding" of decimal data, sites that were previously running Hive 0.11.0 and Hive 0.12.0 may need to migrate tables with Decimal columns after upgrading to Hive 0.13.0. For details, see [the Apache Foundation wiki](#). For assistance with upgrades that involve Decimal data, please contact Hortonworks Support.

## 1.6. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)



### Note

Apache YARN, Apache MapReduce and Apache Knox require no additional patches.

### 1.6.1. Patch information for Hadoop Common/HDFS

Hadoop is based on Apache Hadoop 2.4.0 and includes the following additional patches:

- [HADOOP-10475](#): `ConcurrentModificationException` in `AbstractDelegationTokenSelector.selectToken()`.

- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE\_NAME\_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN
- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE
- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#): Edit log corruption due to deferred INode removal

- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible

## 1.6.2. Patch information for ZooKeeper

ZooKeeper is based on Apache ZooKeeper 3.4.5 and includes the following patches:

- [ZOOKEEPER-1702](#): ZooKeeper client may write operation packets before receiving successful response to connection request, can cause TCP RST.

## 1.6.3. Patch information for HBase

HBase is based on Apache HBase 0.98.0. It includes the following patches:

- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94\* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluatiotool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check

- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10688](#): Add a draining\_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList

**Windows Fixes:**

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

**Changes related to HBASE-10070:**

- [HBASE-10875](#) Metas own location should be cached

- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool
- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#)  
TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION\_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.

- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HCK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

## 1.6.4. Patch information for Pig

Pig is based on Apache Pig 0.12.1. It includes the following patches:

- [PIG-3916](#): isEmpty should not be early terminating.
- [PIG-3650](#): Fix for PIG-3100 breaks column pruning
- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

## 1.6.5. Patch information for Tez

Tez is based on Apache Tez 0.4.0 incubating release. It includes the following patches:

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.
- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.
- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V\_ROUTE\_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V\_INIT at INITED.



## 1.6.6. Patch information for Hive/HCatalog

Hive is based on Apache Hive 0.13.1. Apache HCatalog is now merged with Apache Hive. Hive/HCatalog includes the following patches:

- [HIVE-5072](#): Enable directly invoke Sqoop job through WebHCat
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6521](#): WebHCat cannot fetch correct percentComplete for Hive jobs
- [HIVE-6564](#): WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- [HIVE-6602](#): Multi-user HiveServer2 throws error
- [HIVE-6698](#): hcat.py script does not correctly load the hbase storage handler jars
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [HIVE-6931](#): Windows unit test fixes
- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-7006](#): Fix ql\_rewrite\_gbtoidx.q output file
- [HIVE-7009](#): HIVE\_USER\_INSTALL\_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7035](#): Templetion returns 500 for user errors - when job cannot be found
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator
- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges

- [HIVE-7065](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing
- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7166](#): Vectorization with UDFs returns incorrect results
- [HIVE-7167](#): Hive Metastore fails to start with SQLServerException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization
- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers
- [HIVE-7210](#): NPE with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

### 1.6.7. Patch information for Oozie

Oozie is based on Apache Oozie 4.0.0. It includes the following patches:

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [OOZIE-1447](#): Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException.
- [OOZIE-615](#): Support Oozie HA.
- [OOZIE-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [OOZIE-1306](#): Bring cron syntax to coordinator frequency.
- [OOZIE-1460](#): Implement and document oozie HA security.
- [OOZIE-1486](#): cut down on number of small files to track a running action.
- [OOZIE-1491](#): Make sure oozie works with secure ZooKeeper.

- [OOZIE-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [OOZIE-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [OOZIE-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- [OOZIE-1541](#): typo in oozie HA admin-server command line documentation.
- [OOZIE-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [OOZIE-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [OOZIE-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [OOZIE-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [OOZIE-1576](#): Add documentation for oozie sqoop CLI.
- [OOZIE-1587](#): Add "recovery" column to CoordJob table.
- [OOZIE-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.
- [OOZIE-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [OOZIE-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [OOZIE-1618](#): Dryrun should check variable substitution in workflow.xml.
- [OOZIE-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [OOZIE-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [OOZIE-1726](#): Oozie does not support \_HOST when configuring kerberos security.
- [OOZIE-1733](#): Fix test failures by oozie-1722.

### 1.6.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.4 and includes the following patches:

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.

- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): `-password-file` option triggers `FileSystemClosed` exception at end of Oozie action.
- [SQOOP-1260](#): `HADOOP_MAPRED_HOME` should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with `-queries` because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling `autoCommit` on created connections.
- [SQOOP-1246](#): `HBaseImportJob` should add job `authtoken` only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): `Method Configuration#unset` is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against `commons-io` higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option `"-skip-dist-cache"` to allow Sqoop not copying jars in `%SQOOP_HOME%\lib` folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the `-bulk-load-dir` option to support the HBase `doBulkLoad` function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding `*_HOME` when not explicitly defined.

- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying –hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

## 1.7. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

### 1.7.1. Hardware recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

### 1.7.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*



### Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- 64-bit Ubuntu Precise 12.04

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

## 1.7.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

## 1.7.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

## 1.7.5. Virtualization and cloud platforms

HDP is certified and supported when running on virtual or cloud platforms (for example, VMware vSphere or Amazon Web Services EC2) as long as the respective guest OS is supported by HDP and any issues that are detected on these platforms are reproducible on the same supported OS installed on bare metal.

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

## 1.7.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



### Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 1.8. Upgrading From HDP 2.1.2 to HDP 2.1.3

This section describes how to upgrade an existing HDP 2.1.2 installation to HDP 2.1.3.

If you are upgrading from a previous HDP version, such as HDP 2.0, please follow the complete Stack upgrade instruction to HDP 2.1. See:

- **Ambari:** [http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk\\_upgrading\\_Ambari/content/ambari-upgrade-stack-21.html](http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html)
- **Manual:** [http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk\\_installing\\_manually\\_book/content/upgrade\\_2X.html](http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk_installing_manually_book/content/upgrade_2X.html)

### Before You Begin

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

To upgrade from HDP 2.1.2 to HDP 2.1.3, do the following:

#### 1. Download the appropriate hdp.repo file for your OS:

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.3.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.3.0/hdp.repo</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.3.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.3.0/hdp.repo</a>
SLES 11	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.3.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.3.0/hdp.repo</a>

OR **Download the HDP RPMs single repository tarball.** (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.3.0-centos5-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.3.0-centos5-tars-tarball.tar.gz</a>
--------------------------------	---

RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.3.0-centos6-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.3.0-centos6-tars-tarball.tar.gz</a>
SLES 11	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/HDP-2.1.3.0-suse11-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/suse11/HDP-2.1.3.0-suse11-tars-tarball.tar.gz</a>

## 2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.3 Reference Guide](#).



### Note

If you are upgrading an HA NameNode configuration, keep your JournalNodes running while performing this upgrade procedure. Upgrade, rollback and finalization operations on HA NameNodes must be performed with all JournalNodes running.

## 3. Back up Oozie files.

If you have Oozie installed and running, back up the files in the following directories on the Oozie server host, and make sure all files, including `*site.xml` files, are copied:

```
mkdir oozie-conf-bak
cp -R /etc/oozie/conf/* oozie-conf-bak
```

## 4. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux	Remove WebHCat components:	<code>yum erase "webhcat*"</code>
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<code>yum upgrade hive</code> <code>yum erase hcatalog</code> <code>yum install</code> <code>hive-hcatalog</code>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<code>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType &lt;\$databaseType&gt;</code>  The value for <code>\$databaseType</code> can be <code>derby</code> , <code>mysql</code> , <code>oracle</code> , or <code>postgres</code> .
	Upgrade the following components:	<code>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</code>
		<code>yum install webhcat-tar-hive webhcat-tar-pig</code> <code>yum install "hive-*</code> <code>yum install oozie oozie-client</code> <code>rpm -e --nodeps bigtop-jsvcyum install bigtop-jsvc</code>
	Verify that the components were upgraded. Enter:	<code>yum list installed   grep</code> <code>HDP-\$old-stack-version-number</code>
SLES	Remove WebHCat, components:	<code>zypper remove webhcat\*</code>



Operating System	Instructions	Commands
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<code>zypper install hive-hcatalog</code>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<code>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType &lt;\$databaseType&gt;</code>
	Upgrade the following components:	<code>zypper up "collectd*" "epel-release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</code>
		<code>zypper install webhcat-tar-hive webhcat-tar-pig</code> <code>zypper up -r HDP-2.1.3.0</code> <code>zypper install hive\*zypper install oozie oozie-client</code>
	Verify that the components were upgraded:	<code>rpm -qa   grep hadoop, rpm -qa   grep hive and rpm -qa   grep hcatalog</code>
	If components were not upgraded, upgrade them:	<code>yast --update hadoop hcatalog hive</code>

5. If your installation includes tables with Decimal columns that were created on Hive 0.11.0 or 0.12.0, you must migrate those tables. For instructions, see [What's changed in Hive](#). For assistance, contact Hortonworks Customer Support.

6. If you are upgrading from an HA NameNode configuration, restart all JournalNodes. On each JournalNode host, enter the following command:

```
su -l {HDFS_USER} -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh start journalnode"
```

7. Complete the Stack upgrade.

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.3 repositories for HDP and HDP-UTILS. For Ambari 1.6.0 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/{$os}/2.x/updates/2.1.3.0 {$os}
```

where {\$os} is the Operating System Family (OS Family). See the following table:

**Table 1.1. Operating Systems mapped to each OS Family**

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11

8. Finalize the upgrade.

If you are not yet ready to discard your backup, you can start the upgraded HDFS without finalizing the upgrade. (At this stage, you can still roll back if need be.)

Verify your filesystem health. When you are ready to commit to this upgrade (are certain that you will not want to roll back), discard your backup and finalize the upgrade.

As \$HDFS\_USER, execute the following command:

```
hdfs dfsadmin -finalizeUpgrade
```

## 1.9. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- HDP 2.1 supports the use of Oracle DB as a metastore.
- The Storm LogViewer service now allows only worker log files from the Storm log directory to be viewed from the UI, regardless of the privileges of the person running the service.

## 1.10. Common Vulnerabilities and Exposures

- **CVE-2014-0229:** Three HDFS admin commands lack proper privilege checks

**Vendor:** The Apache Software Foundation

**Versions Affected:** Apache Hadoop 0.23 prior to 0.23.11, Apache Hadoop 2.x prior to 2.4.1 (fix included in HDP 2.1.3)

**Impact:** Three HDFS admin commands, refreshNamenodes, deleteBlockPool and shutdownDatanode, lack proper privilege checks in Apache Hadoop 0.23.x prior to 0.23.11, and in 2.x prior to 2.4.1. This allows arbitrary users to create data nodes unnecessarily, refresh its federated NameNode configuration in an untimely manner, delete inactive block pools, or shut itself down.

(The shutdownDatanode command was first introduced in 2.4.0; refreshNamenodes and deleteBlockPool were added in 0.23.0.)

**Recommended Action:** Hadoop 0.23.x users should upgrade to 0.23.11, Hadoop 2.x users should upgrade to HDP 2.1.3, which includes the fix for this vulnerability.

- **CVE-2014-0228:** Apache Hive authorization vulnerability

**Severity:** Moderate

**Vendor:** The Apache Software Foundation

**Versions Affected:** Apache Hive 0.13.0

**Users Affected:** Users who have enabled SQL standards-based authorization mode

**Description:** In SQL standards-based authorization mode, the URIs in Hive queries are expected to be authorized on the file system permissions. However, the directory used in import/export statements is not being authorized. This allows a user who knows the directory to which data has been exported to import that data into his table. This is possible if the user HiveServer2 runs with permissions for that directory and its contents.

**Recommended Action:** Users who use SQL standards-based authorization should upgrade to Hive 0.13.1.

**Credit:** This issue was discovered by Thejas Nair of Hortonworks

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability

**Severity:** Major

**Vendor:** The Apache Software Foundation

**Versions Affected:** Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0

**Users Affected:** Users who have enabled Hadoop's MapReduce security features

**Impact:** Vulnerability allows an unauthorized user to retrieve job details from the job history server

**Recommended Action:** Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to at least 2.3.0 (HDP 2.1.1 includes Hadoop 2.4.0).

**Credit:** This issue was discovered by Koji Noguchi of Yahoo

## 1.11. Known Issues

In this section:

- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Apache Pig](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Hue](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

### 1.11.1. Known Issues for HDP

- **BUG-19593:** Hive user does not have shell access on Deb

**Problem:** When attempting to access Deb as the user `hive`, the following displays:

```
hive:x:113:121:Hive User,,,:/var/lib/hive:/bin/false
```

- **BUG-19531:** ZooKeeper user does not have shell access on Deb

**Problem:** The `/etc/passwd` file has no shell access for ZooKeeper.

- **BUG-16003:** Knox PID directory does not exist on Ubuntu after reboot

**Problem:** Unable to start Knox after Ubuntu reboot

- **BUG-15796:** Sigsegv in mapred history server due to segfault in JniBasedUnixGroupsMapping.

**Problem:** RHEL 6.2 and RHEL 6.3 contain known bugs in nslcd. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by nslcd times out.

**Workaround:** If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.
- Increase the configured nslcd search timeout in `/etc/nslcd.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGroupsMapping` in `core-site.xml`.
- **BUG-15360:HADOOP-10519:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error

## 1.11.2. Known Issues for YARN

- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.

**Problem:** YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.

- **BUG-12327:** ([YARN-90](#)) NM cannot detect when bad disks become healthy again.

**Problem:** If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.

- **BUG-9919:** Shuffle test fails with `mapreduce.map.java.opts` and `mapreduce.reduce.java.opts -Xmx692m`

**Problem:** Shuffle tests fails in ambari with the following configuration:

```
mapred-site.xml
mapreduce.map.memory.mb 1024
mapreduce.map.java.opts -Xmx692m
mapreduce.task.io.sort.mb 200
```

Shuffle test is failing with "java.lang.OutOfMemoryError: Java heap space".

### 1.11.3. Known Issues for HBase

- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

**Problem:** On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.

### 1.11.4. Known Issues for Apache Pig

- **BUG-19608:** ([PIG-4044](#), [PIG-4045](#)) Pig job processing Avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar

**Problem:** If Pig was installed from pig-withouthadoop.jar (which is missing avro-mapred-h2.jar), trying to run Pig with Avro data returns:

```
Error running child : java.lang.IncompatibleClassChangeError: Found
interface org.apache.hadoop.mapreduce.TaskAttemptContext, but class was
expected
```

### 1.11.5. Known Issues for Hive

- **BUG-19152:** Upon start, HiveServer2 doesn't know about the admin users in hive.users.in.admin.role for a while

**Problem:** When HiveServer2 is started it takes a while for it to initialize the users set in the hive.users.in.admin.role property. This causes the first few tests in Hive SQL Standard Auth test suite to fail.

- **BUG-18002:** NullPointerException in OrcInputFormat when Vectorization is turned on

**Problem:** We are unable to verify if the NPE occurs with the normal expected settings.

- **BUG-17850:** HBCK test fails intermittently due to Empty Region Qualifier error

- **BUG-17846:** Hive query using SUM() windowing function fails to complete and stays stuck on reduce task

**Problem:** The query never completes.

- **BUG-17247:** In Hive Cli switching the hive.execution.engine from Tez to MapReduce does not also switch the YARN framework back to MapReduce

**Problem:** If we can't switch the YARN framework back to MR, Hive MR will still run on Tez.

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

**Problem:** The Hive session should shut down cleanly and not kill the app.

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

**Problem:** We force create\_index to run in MR mode when we have a TEZ run, but it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16476:** Oozie-hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder

**Problem:** The issue occurs because Oozie launches the Hive client as the mapreduce user (hadoop in this case). However, the ugi information is that of the user using Oozie (hadoopqa in this case). Therefore, Hive always creates the /tmp/hive-hadoop directory for use as a scratch directory with hadoopqa as the owner. The right fix for this would be to create user specific directories in the first place and should be addressed in [HIVE-6782](#).

**Workaround:** Either wipe out the directory or to set permissions of 777 on the directory.

- **BUG-16393:** Bucketized Table feature fails in some cases.

**Problem:** Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination.

- **BUG-16257:** HBase master fails to start due to BindException

**Problem:** HBase on Suse 11 64 bit, Smoke test fails Intermittent with ERROR main client.ConnectionManager\$HConnectionImplementation: The node /hbase is not in ZooKeeper. Basically HBase default ports clash with the range Linux assigns itself for creating come-and-go ephemeral ports. Therefore, once in awhile we'll see HBase master cannot start due to port binding issue.

- **BUG-15733:** (HIVE-7071) Schema evolution is broken on Tez.

**Problem:** The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
```

```

at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive.ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive.ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive.ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more

```

- **BUG-16476:** Oozie-hive tests run as hadoopqa creates/accesses /tmp/hive-hadoop folder

**Problem:** The issue occurs because Oozie launches the Hive client as the mapreduce user (hadoop in this case). However, the ugi information is that of the user using Oozie (hadoopqa in this case), so Hive always creates the /tmp/hive-hadoop directory for use as a scratch directory with hadoopqa as the owner. The right fix for this would be to create user specific directories in the first place and should be addressed in [HIVE-6782](#).

**Workaround:** Either wipe out the directory or to set permissions of 777 on the directory.

- **BUG-15003:** Hive sink throws exception on shutdown

**Problem:** When using the Hive sink in Flume, you are likely to see the below warning in the logs followed by a stack trace when shutting down the Flume agent:

```

14/03/16 17:39:07 WARN hive.HiveSink: Exception
while closing HiveEndPoint ...

```

There is no current evidence that this exception indicates data loss.

- **BUG-14986:** Region assignments for large number of regions may cause timeouts on windows

**Problem:** On the Windows env, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the master logs.
- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-13551:** Oozie does not understand \_HOST in the kerberos principal name

**Problem:** Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack where we can just send \_HOST and the service at run

time will replace `_HOST` with machine hostname. This is important so that in a HA setup we can push the same configs to all oozie servers.

- **BUG-10512:** Streaming / SELECT TRANSFORM doesn't work with Tez

**Problem:** SELECT TRANSFORM doesn't work with Tez enabled, works in same build with Tez disabled.

- **BUG-8227:** ([HIVE-6638](#)) Hive needs to implement recovery or extend FileOutputComitter.

**Problem:** When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

```
public boolean isRecoverySupported() {
    return false;
}
```

## 1.11.6. Known Issues for Hue

- **BUG-19718:** Traceback error shown when installing Hue on CentOS 5 or SLES 11

**Problem:** During installation of Hue on CentOS 5 and SLES 11, warning messages similar to the following display:

```
warning: user jenkins does not exist - using root (multiple messages)
warning: group jenkins does not exist - using root (multiple messages)
make: Entering directory `/usr/lib/hue/desktop'
--- Regenerating database at /usr/lib/hue/desktop/desktop.db
--- Syncing/updating database at /usr/lib/hue/desktop/desktop.db
Traceback (most recent call last):
  File "/usr/lib/hue/build/env/bin/hue", line 9, in module
    load_entry_point('desktop==2.3.4', 'console_scripts', 'hue')()
  File "/usr/lib/hue/desktop/core/src/desktop/manage_entry.py", line 60, in
  entry
    execute_manager(settings)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/__init__.py", line 438, in execute_manager
    utility.execute()
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/__init__.py", line 379, in execute
    self.fetch_command(subcommand).run_from_argv(self.argv)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 191, in run_from_argv
    self.execute(*args, **options.__dict__)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 219, in execute
    self.validate()
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 249, in validate
    num_errors = get_validation_errors(s, app)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.
  2.3-py2.6.egg/django/core/management/validation.py", line 28, in
  get_validation_errors
    for (app_name, error) in get_app_errors().items():
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/db/models/loading.py", line 146, in get_app_errors
```



```

self._populate()
File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.6.egg/django/db/models/loading.py", line 64, in _populate
self.load_app(app_name)
File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.6.egg/django/db/models/loading.py", line 78, in load_app
models = import_module('.models', app_name)
File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.6.egg/django/utils/importlib.py", line 35, in import_module
__import__(name)
File "/usr/lib/hue/apps/jobbrowser/src/jobbrowser/models.py", line 20, in module
import lxml.html
File "/usr/lib/hue/build/env/lib/python2.6/site-packages/lxml-3.3.5-py2.6-linux-x86_64.egg/lxml/html/__init__.py", line 42, in module
from lxml import etree
ImportError: /usr/lib/hue/build/env/lib/python2.6/site-packages/lxml-3.3.5-py2.6-linux-x86_64.egg/lxml/etree.so: undefined symbol:
_PyByteArray_empty_string
make: *** [/usr/lib/hue/desktop/desktop.db] Error 1

```

**Workaround:** No action is required. The error message above can be ignored. Hue installation proceeds.

- **BUG-19511:** Error message during Hue start after upgrading Hue

**Problem:** After upgrading Hue from Hue 2.3.0 to Hue 2.3.2 and restarting the service, Hue displays

```
Failed to register some apps: Details in /var/log/hue/hue_re_register.log
```

The message is benign, and can be ignored.

**Workaround:** To eliminate this message, enter:

```
chown -R hue:hue /usr/lib/hue
```

- **BUG-19502:** Hue does not ship hive.aux.jars.path

**Problem:** Hue does not ship hive.aux.jars.path. This can cause problems with hbasestorage tables.

**Workaround:** Place the HBase jars on the Hue server host. Enter:

```
ln -s /usr/lib/hbase/lib*.jar /usr/lib/hive/lib/
ln -s /etc/hbase/conf/*.xml /etc/hive/conf
```

## 1.11.7. Known Issues for Storm

- **BUG-19709:** Storm /etc directory not created on Ubuntu after manual installation

**Problem:** After doing the manual install of Storm on Ubuntu (apt-get install storm), as per the [HDP 2.1.3 manual installation instructions](#), no configuration folder is created for Storm under /etc; neither /etc/storm/conf nor /etc/storm/conf.dist are present.

**Workaround:** Create /etc/storm/conf on all the nodes where Storm is deployed, and create the storm.yaml file with contents as noted in [Installing HDP Manually](#).

## 1.11.8. Known Issues for Knox

- **BUG-16003:** Knox PID directory does not exist on Ubuntu after reboot

**Problem:** Unable to start Knox after Ubuntu reboot.

## 1.11.9. Known Issues for the Hortonworks Connector for Teradata

- The Hortonworks Connector for Teradata is not supported at this time. Please check <http://www.hortonworks.com> often for updates, which will be coming soon.

## 1.11.10. Known Issues for Ambari

- See [Ambari Release Notes](#).

## 1.12. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

## 1.13. Third-party Licenses

Table 1.2. Third-party Licenses

HDP Component	License
Phoenix	<a href="#">EPL</a>
Storm	<a href="#">EPL</a>
Accumulo	<a href="#">JCommander</a>
Falcon	<a href="#">CERN</a>
Falcon	<a href="#">Tinkerpop</a>
Knox	<a href="#">ANTLR</a>
Knox	<a href="#">MIT</a>
Knox	<a href="#">EPL</a>
Knox	<a href="#">Bouncy Castle</a>
Knox	<a href="#">OWS</a>

## 2. Release Notes HDP-2.1.2

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.2](#)
- [Unsupported Apache components](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum system requirements](#)
- [Upgrading From HDP 2.1.1 to HDP 2.1.2](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

### 2.1. Product Version: HDP-2.1.2

All HDP 2.1 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 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.



#### Note

The minimum level of Apache Ambari to use with HDP 2.1, is version 1.5.1.

This release of Hortonworks Data Platform (HDP) deploys the following Hadoop-related components:

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.0

- Apache Tez 0.4
- Apache ZooKeeper 3.4.5
- Hue 2.3.1
- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0
- Third party components:
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

## 2.2. Unsupported Apache components

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation (Apache JIRA HDFS-1052)
- viewFS (Apache JIRA HADOOP-7257)
- viewFS (Apache JIRA HADOOP-7257)

The following Apache Components are shipped as part of HDP 2.1 YARN, but are not supported:

- Application Timeline Server (Hive-on-Tez metrics)
- AM failure/restart resiliency
- MapReduce Uber AM
- YARN CGroup resource isolation

- Admin Node labels
- CPU Scheduling
- Fair Scheduler
- MapReduce Eclipse Plug-in

## 2.3. Fixed in This Release

HDP 2.1.2 includes the following fixes:

BUG-17210	<a href="#">STORM-269</a>	Any readable file is exposed via the log viewer UI
BUG-17186	<a href="#">HADOOP-813</a>	Allow wildcard bind for additional YARN/MR endpoints
BUG-17060	<a href="#">HIVE-6952</a>	Hive Record Writer is not backward-compatible on HDP-2.1 release
BUG-16976	<a href="#">HIVE-6945</a>	Dropping ranges of partitions returns an error after succeeding on Oracle
BUG-16890	<a href="#">HIVE-6957</a>	Hive SQL standard auth calls accessing local or HDFS urls fails in Kerberos secure cluster with binary HS2 transport
BUG-16809, BUG-16008	<a href="#">HDFS-6273</a>	Provide service binding address lists for HA services currently lacking them (HDFS/Namenode in 16808, MR/Yarn in 16809)
BUG-16797	<a href="#">HIVE-4576</a>	templeton.hive.properties does not allow values with commas

## 2.4. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- [What's Changed in Mahout](#)
- [HDP 2.1 Clusters Deployed via Ambari](#)
- [What's Changed in Hue](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in Oozie](#)

### 2.4.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

Deprecated algorithms were removed ([MAHOUT-1296](#)) without Frequent Pattern Mining. Multilayer Perceptron was added ([MAHOUT-1265](#)).

## 2.4.2. HDP 2.1 clusters deployed via Ambari

Freshly-installed HDP 2.1 clusters deployed via Ambari will have the new Hive authorization system turned on by default. Manual installs and upgrades are not affected by this, only fresh HDP 2.1 installs via Ambari. This default behavior will be modified in Ambari 1.6.0.

## 2.4.3. Hue behavioral changes

Hue now supports the optional ability to input Unix usernames in lowercase letters and have Active Directory return usernames in upper case, in cases where LDAP/Active directory is being used as the back end.

## 2.4.4. HBase behavioral changes

In HDP 2.1, if the user does not have read privileges to a table and scans that table he will get an empty result set back.

## 2.4.5. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable `hive.server2.enable.doAs` is set to true, before the user starts the HiveServer2 process, they should create a scratch directory, `/tmp/hive-<username>`, on HDFS, where `<username>` is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

## 2.4.6. Oozie behavioral changes

When Oozie has been installed manually, before a site can execute any Oozie actions, the Oozie shared libraries must be made explicitly available. Add the following information to the file `oozie-site.xml`:

```
<property>
<name>oozie.service.WorkflowAppService.system.libpath</name>
<value>/user/$username/share/lib</value>
</property>
```

## 2.5. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)

- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)



### Note

Apache YARN, Apache MapReduce and Apache Knox require no additional patches.

## 2.5.1. Patch information for Hadoop Common/HDFS

Hadoop is based on Apache Hadoop 2.4 and includes the following additional patches:

- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken().
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.

## 2.5.2. Patch information for ZooKeeper

ZooKeeper is based on Apache ZooKeeper 3.4.5 and includes the following patches:

- [ZOOKEEPER-1702](#): ZooKeeper client may write operation packets before receiving successful response to connection request, can cause TCP RST.

## 2.5.3. Patch information for HBase

HBase is based on Apache HBase 0.98.0. It includes the following patches:

- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10592](#): Refactor PerformanceEvaluationTool
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10548](#): Correct commons-math dependency version

- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10688](#): Add a draining\_node script to manage nodes in draining mode
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94\* to hbase0.96+
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table



- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

#### Windows Fixes:

- [HBASE-10799](#) [WINDOWS]  
TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows

#### Changes related to HBASE-10070:

- [HBASE-10875](#) Metas own location should be cached
- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool
- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#)  
TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times

- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION\_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HCK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage

- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

## 2.5.4. Patch information for Pig

Pig is based on Apache Pig 0.12.1. It includes the following patches:

- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

## 2.5.5. Patch information for Tez

Tez is based on Apache Tez 0.4.0 incubating release. It includes the following patches:

- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V\_ROUTE\_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.

- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitonException: Invalid event: V\_INIT at INITED.

## 2.5.6. Patch information for Hive/HCatalog

Hive is based on Apache Hive 0.13.0. Apache HCatalog is now merged with Apache Hive. Hive/HCatalog includes the following patches:

- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6919](#): hive sql std auth select query fails on partitioned tables
- [HIVE-6915](#): Hive Hbase queries fail on secure Tez cluster
- [HIVE-6898](#): Functions in hive are failing with java.lang.ClassNotFoundException on Tez
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6883](#): Dynamic partitioning optimization does not honor sort order or order by
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6828](#): Hive tez bucket map join conversion interferes with map join conversion
- [HIVE-6826](#): Hive-tez has issues when different partitions work off of different input types
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6726](#): Hcat cli does not close SessionState

- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive

## 2.5.7. Patch information for Oozie

Oozie is based on Apache Oozie 4.0.0 and includes the following patches:

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [Oozie-615](#): Support Oozie HA.
- [Oozie-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [Oozie-1306](#): Bring cron syntax to coordinator frequency.
- [Oozie-1460](#): Implement and document oozie HA security.
- [Oozie-1486](#): cut down on number of small files to track a running action.
- [Oozie-1491](#): Make sure oozie works with secure ZooKeeper.
- [Oozie-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [Oozie-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [Oozie-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- [Oozie-1541](#): typo in oozie HA admin-server command line documentation.
- [Oozie-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [Oozie-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [Oozie-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [Oozie-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [Oozie-1576](#): Add documentation for oozie sqoop CLI.
- [Oozie-1587](#): Add "recovery" column to CoordJob table.
- [Oozie-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.

- [Oozie-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [Oozie-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [Oozie-1618](#): Dryrun should check variable substitution in workflow.xml.
- [Oozie-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [Oozie-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [Oozie-1726](#): Oozie does not support \_HOST when configuring kerberos security.
- [Oozie-1733](#): Fix test failures by oozie-1722.

## 2.5.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.4 and includes the following patches:

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): -password-file option triggers FileSystemClosed exception at end of Oozie action.
- [SQOOP-1260](#): HADOOP\_MAPRED\_HOME should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with -queries because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling autoCommit on created connections.

- [SQOOP-1246](#): HBaseImportJob should add job authtoken only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP\_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding \*\_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

## 2.6. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

## 2.6.1. Hardware recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

## 2.6.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*



### Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

## 2.6.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)



## 2.6.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

## 2.6.5. Virtualization and cloud platforms

HDP is certified and supported when running on virtual or cloud platforms (for example, VMware vSphere or Amazon Web Services EC2) as long as the respective guest OS is supported by HDP and any issues that are detected on these platforms are reproducible on the same supported OS installed on bare metal.

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

## 2.6.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



### Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 2.7. Upgrading From HDP 2.1.1 to HDP 2.1.2

This section describes how to upgrade an existing HDP 2.1.1 installation to HDP 2.1.2.

If you are upgrading from a previous HDP version, such as HDP 2.0, please upgrade to HDP 2.1.1 before beginning this process. See:

- **Ambari:** [http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk\\_upgrading\\_Ambari/content/ambari-upgrade-stack-21.html](http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html)
- **Manual:** [http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.1/bk\\_installing\\_manually\\_book/content/upgrade\\_2X.html](http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.1/bk_installing_manually_book/content/upgrade_2X.html)

### Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.

- Think about whether you're going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.1 to HDP 2.1.2, do the following:

### 1. Download the appropriate hdp.repo file for your OS:

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.2.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.2.0/hdp.repo</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.2.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.2.0/hdp.repo</a>
SLES 11	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.2.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.2.0/hdp.repo</a>

OR Download the HDP RPMs, and hdp.repo as a single tar file. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/2.x/GA/2.1-latest/HDP-2.1-latest-centos5-rpm.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos5/2.x/GA/2.1-latest/HDP-2.1-latest-centos5-rpm.tar.gz</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/2.x/GA/2.1-latest/HDP-2.1-latest-centos6-rpm.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos6/2.x/GA/2.1-latest/HDP-2.1-latest-centos6-rpm.tar.gz</a>
SLES 11	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/2.x/GA/2.1-latest/HDP-2.1-latest-suse11-rpm.tar.gz">http://public-repo-1.hortonworks.com/HDP/suse11/2.x/GA/2.1-latest/HDP-2.1-latest-suse11-rpm.tar.gz</a>

### 2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.2 Reference Guide](#).



### Note

If you are upgrading an HA NameNode configuration, make sure you only stop the services and do not stop any of the JournalNodes. If any of your JournalNodes are not running while performing any upgrade, rollback or finalization operation, or else the operation will fail.

### 3. If you have Oozie installed/running:

Back up the files in the following directories on the Oozie server host, and make sure all files, including *site.xml files, are copied	<pre>mkdir oozie-conf-bak cp -R /etc/oozie/conf/* oozie-conf-bak</pre>
Remove the old Oozie directories on all Oozie server and client hosts	<pre>rm -R /etc/oozie/conf/*</pre>

### 4. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux	Remove WebHCat components:	<pre>yum erase "webhcat*"</pre>
	If you haven't done so already, stop the Hive Metastore.	

Operating System	Instructions	Commands
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<pre>yum upgrade hive yum erase hcatalog yum install hive-hcatalog</pre>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType &lt; \$databaseType&gt;</pre> <p>The value for \$databaseType can be derby, mysql, oracle, or postgres.</p>
	Upgrade the following components:	<pre>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "squoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</pre>
		<pre>yum install webhcat-tar-hive webhcat-tar-pig yum install "hive-*" yum install oozie oozie-client rpm -e --nodeps bigtop-jsvcyum install bigtop-jsvc</pre>
	Verify that the components were upgraded. Enter:	<pre>yum list installed   grep HDP-\$old-stack-version-number</pre>
SLES	Remove WebHCat, components:	<pre>zypper remove webhcat\*</pre>
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<pre>zypper install hive-hcatalog</pre>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType &lt; \$databaseType&gt;</pre>
	Upgrade the following components:	<pre>zypper up "collectd*" "epel-release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</pre>
		<pre>zypper install webhcat-tar-hive webhcat-tar-pig zypper up -r HDP-2.1.2.0 zypper install hive\*zypper install oozie oozie-client</pre>
	Verify that the components were upgraded:	<pre>rpm -qa   grep hadoop, rpm -qa   grep hive and rpm -qa   grep hcatalog</pre>
	If components were not upgraded, upgrade them:	<pre>yast --update hadoop hcatalog hive</pre>

5. If you are upgrading from an HA NameNode configuration, restart all JournalNodes. On each JournalNode host, enter the following command:

```
su -l {HDFS_USER} -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh start
journalnode"
```

6. Complete the Stack upgrade.

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.3 repositories for HDP and HDP-UTILS. For Ambari 1.6.0 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/
{${os}}/2.x/updates/2.1.3.0 {${os}}
```

where {\${os}} is the Operating System Family (OS Family). See the following table:

**Table 2.1. Operating Systems mapped to each OS Family**

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5

OS Family	Operating System
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11

## 7. Finalize the upgrade.

If you are not yet ready to discard your backup, you can start the upgraded HDFS without finalizing the upgrade. (At this stage, you can still roll back if need be.)

Verify your filesystem health. When you are ready to commit to this upgrade (are certain that you will not want to roll back), discard your backup and finalize the upgrade.

As \$HDFS\_USER, execute the following command:

```
hdfs dfsadmin -finalizeUpgrade
```

## 2.8. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- HDP 2.1 supports the use of Oracle DB as a metastore.
- The Storm LogViewer service now allows only worker log files from the Storm log directory to be viewed from the UI, regardless of the privileges of the person running the service.

## 2.9. Common Vulnerabilities and Exposures

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability
- **Severity:** Major
- **Vendor:** The Apache Software Foundation
- **Versions Affected:** Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0
- **Users Affected:** Users who have enabled Hadoop's MapReduce security features
- **Impact:** Vulnerability allows an unauthorized user to retrieve job details from the job history server
- **Mitigation:** Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to 2.3.0
- **Credit:** This issue was discovered by Koji Noguchi of Yahoo

## 2.10. Known Issues

In this section:

- [Known Limitation for HDFS Upgrade](#)
- [Known Issues for SLES 11](#)

- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for WebHCat](#)
- [Known Issues for Oozie](#)
- [Known Issues for Hue](#)
- [Known Issues for Flume](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

### 2.10.1. Known Limitation for HDFS Upgrade

- **BUG-17431:** ([HDFS-5526](#)) If you upgrade from HDP 2.0.9.1 to HDP 2.1, the DataNode cannot be rolled back to HDP 2.0.9.1.

Contact Support for an update to fix this HDP 2.0.9.1 issue.

### 2.10.2. Known Issues for SLES 11

- **BUG-9904:** php\_curl Required for SLES 11 Sp1

**Problem:** Several alerts return with `Return code of 255 is out of bounds` while trying to install a cluster on SLES because `php_curl` is not installed.

**Workaround:** Install `php_curl` on your SLES host.

```
zypper install php-curl
```

### 2.10.3. Known Issues for HDP

- **BUG-15796:** Sigsegv in mapred history server due to segfault in `JniBasedUnixGroupsMapping`.

**Problem:** RHEL 6.2 and RHEL 6.3 contain known bugs in `nslcd`. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by `nslcd` times out.

**Workaround:** If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.

- Increase the configured nslcd search timeout in `/etc/nslcd.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGro` in `core-site.xml`.
- **BUG-15360:HADOOP-10519:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error

**Problem:**

- **BUG-825:ED2** m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

**Problem:** Directories and disks that you assign for logging in HDP do not have enough space to maintain logs during HDP operations.

**Workaround:** Designate least 10 GB of free space on a disk that will be used by HDP logging.

### 2.10.3.1. Known issues for HDFS

- **BUG-14542:** HDP 2.1 exception during namenode service work.

**Problem:** After the start of the NameNode service, the following exception occurred:

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_00000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msec
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
9 more
```

### 2.10.3.2. Known issues for MapReduce

- **BUG-15360: (HADOOP-10519)** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.
- **BUG-14749:** `CombineFileInputFormat.getSplits()` including directories in its results.

**Problem:** This is causing Hive test root\_dir\_external\_table.q to fail when running against hadoop-2. Opened Apache Jira [MAPREDUCE-5756](#) Created in Monarch as <https://hwxmonarch.atlassian.net/browse/HADOOP-801>, creating equivalent bug for HDP-2.1.

- **BUG-12005:** ([MAPREDUCE-2306](#)) Mapreduce.task.io.sort.mb is capped at 2047.

**Problem:** mapreduce.task.io.sort.mb is hardcoded to not allow values larger than 2047. If you enter a value larger then this the map tasks will always crash at this line:

```
https://github.com/apache/hadoop-mapreduce/blob/HDFS-641/src/java/org/apache/hadoop/mapred/MapTask.java?source=cc#L746
```

## 2.10.4. Known Issues for YARN

- **BUG-15834:** YARN and/or mapred client should add tokens for default filesystem.

**Problem:** As noticed in BUG-15360 if jobs are using webhdfs then they could run into an issue where the job only has webhdfs tokens and yarn jobs would fail as it tries to talk to hdfs over default fs since tokens for default fs as not part of the job.

- **BUG-15376:** ([YARN-1892](#)) CS fast scheduling patch ends up causing excessive logging.

**Problem:** Seeing about 1 GB of logs per hour.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.

**Problem:** YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.

- **BUG-12327:** ([YARN-90](#)) NM cannot detect when bad disks become healthy again.

**Problem:** If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.

- **BUG-9919:** Shuffle test fails with mapreduce.map.java.opts and mapreduce.reduce.java.opts -Xmx692m

**Problem:** Shuffle tests fails in ambari with the following configuration:

```
mapred-site.xml
mapreduce.map.memory.mb 1024
mapreduce.map.java.opts -Xmx692m
mapreduce.task.io.sort.mb 200
```

Shuffle test is failing with "java.lang.OutOfMemoryError: Java heap space".

- **BUG-7531:** ([YARN-991](#)) Hadoop metrics link does not contain correct content.

**Problem:** In the Resource Manager UI Tools section, clicking on Logs and Metrics opens pages that do not contain correct information.

## 2.10.5. Known Issues for HBase

- **BUG-16900: (HBASE-11036)** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.

**Problem:** The Big Linked List Test with Chaos Monkey Test run shows Not Serving Region Exceptions in the Yarn logs, causing failure.

- **BUG-16513:** HBCK Tests Fail Intermittently Due to NotServingRegionException.

**Problem:** The HBCK Tool tests fails intermittently due to a NotServingRegionException, noted in the Master Logs.

- **BUG-16257: (HBASE-10123)** HBase master fails to start due to BindException.

**Problem:** HBase on Suse 11 64 bit, smoke test fails intermittently with:

```
ERROR [main] client.ConnectionManager$HConnectionImplementation: The node /
hbase is not in ZooKeeper.
```

- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

**Problem:** On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.

## 2.10.6. Known Issues for Phoenix

- **BUG-16484:** Phoenix ZooKeeper quorum string cannot contain the port number.

**Problem:** HDP 2.1 defines port numbers in hbase.zookeeper.quorum in hbase-site.xml, which causes conflicts when you use Phoenix on HBase. This results in an error message similar to the following:

```
java.sql.SQLException: ERROR 102 (08001): Malformed connection url.
```

- **Workaround:** Remove the port number from hbase.zookeeper.quorum in hbase-site.xml, and include the port number in the JDBC connector string:

```
jdbc:phoenix [ :<zookeeper quorum> [ :<port number > ] [ :/hbase ]]
```

## 2.10.7. Known Issues for Hive

- **BUG-17773:** Hive-on-Tez does not work with encrypted shuffle.

**Problem:** The Encrypted Shuffle capability enables the MapReduce shuffle to be encrypted using HTTPS, with optional client authentication. (This is also known as bi-directional HTTPS, or HTTPS with client certificates.) Hive-on-Tez does not work in an environment where encrypted shuffle is enabled.



- **BUG-17603:** ([HIVE-6985](#)) Grant privileges on PUBLIC role are not being honored.

**Problem:** When a privilege is granted to public role, the privilege is supposed to be applicable for all users. However, the privilege check fails for users, even if they have public role in the list of current roles. (This issue only exists with the Public role; it does not affect the granting of privileges to other roles.)

- **BUG-16945:** HiveSysTest-testTableCreationAndQuery fails with error "Permission denied inode="/tmp/hive-hadoop"".

**Problem:** HiveSysTest-testTableCreationAndQuery fails with error "Permission denied inode="/tmp/hive-hadoop"". Hadoop user is not cleaned up as guest user and thus the test fails. Verified hiveserver2.doas property is set to true in hive-site.xml.

- **BUG-16890:** ([HIVE-6957](#)) Hive SQL standard auth calls accessing local or HDFS URLs fail in Kerberos secure cluster with binary HS2 transport.

**Problem:** This is blocking all CREATE table calls where we access LOCAL or HDFS uri.

```
>>> create external table studenttab10k(
name string,
age int,
gpa double)
row format delimited
fields terminated by '\t'
stored as textfile
location '/user/hcat/tests/data/studenttab10k';
2014-04-17 00:12:13,627 DEBUG [main] transport.TSaslTransport: writing data
length: 297
2014-04-17 00:12:13,657 DEBUG [main] transport.TSaslTransport: CLIENT:
reading data length: 351
Error: Error while compiling statement: FAILED: HiveAccessControlException
Permission denied.
Principal [name=hrt_qa@HORTON.YGRIDCORE.NET, type=USER] does not have
following privileges on Object
[type=DFS_URI, name=/user/hcat/tests/data/studenttab10k] : [INSERT, DELETE,
OBJECT OWNERSHIP] (state=42000,code=40000)
```

- **BUG-16660:** On Tez setup, Hive jobs in webhcat run in default mr mode even in Hive.

**Problem:** Currently when we run Hive jobs through Webhcat we always run in MR mode even though we are running them in a cluster where Hive queries would have run in Tez mode. This is only on Linux installs. The problem here is that we run hive queries using hive.tar.gz on HDFS and specifying explicit hive configurations, here are the properties that we use in webhcat-site.xml:

templeton.hive.archive	hdfs:///apps/webhcat/hive.tar.gz
templeton.hive.path	value: hive.tar.gz/hive/bin/hive
templeton.hive.home	value: hive.tar.gz/hive
templeton.hive.properties	hive.metastore.local=false, hive.metastore.uris=thrift://hivehost:9083, hive.metastore.sasl.enabled=false, hive.metastore.execute.setugi=true

When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add "hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.
- **BUG-16608: (FALCON-390)** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** The job fails with the following permission error:

```
Copying data from hdfs://arbit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data/dt=2010-01-01-20
Copying file: hdfs://arbit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(message:Got exception: org.apache.hadoop.security.AccessControlException Permission denied: user=hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/HCatReplication_oneSourceOneTarget_hyphen":arbit:hdfs:drwxr-xr-x at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

**Problem:** Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16864:** When Hive standard authorization is enabled, the owner of the table backing index is missing.

**Problem:** The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
(SessionState.java:printError(546)) - FAILED: HiveAccessControlException
Permission denied. Principal [name=hrt_qa, type=USER] does not have
following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index_] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
name=default__missing_ddl_3_missing_ddl_3_index_] : [OBJECT
OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
```

```

at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...

```

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

**Problem:** The Hive session should shut down cleanly and not kill the app.

- **BUG-16771:** ([HIVE-6867](#)) Hive table has multiple copies of streaming data when testing the Hive Server restart scenario.

**Problem:** When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

**Problem:** We force create\_index to run in MR mode when we have a TEZ run, but it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16393:** Bucketized Table feature fails in some cases.

**Problem:** Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination. Example follows:

```
CREATE TABLE P1(key STRING, val STRING)
CLUSTERED BY (key) SORTED BY (key) INTO 2 BUCKETS STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/Users/jpullokkaran/apache-hive1/data/files/P1.txt'
INTO TABLE P1;

-- perform an insert to make sure there are 2 files
INSERT OVERWRITE TABLE P1 select key, val from P1;
```

- **Workaround:** Avoid loading data for bucketed table.
- **BUG-16391:** Streaming transactions fail on MSSQL.

**Problem:** After creating tables using the MSSQL composite script provided by BUG-15827 running Flume, Hive Sink tests failed because no data made it into Hive tables.

- **BUG-15733:** Schema evolution is broken on Tez.

**Problem:** The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
```

```

at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more

```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-8227: (HIVE-6638)** Hive needs to implement recovery or extend FileOutputComitter.

**Problem:** When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

```

public boolean isRecoverySupported() {
    return false;
}

```

## 2.10.8. Known Issues for Tez

- **BUG-17773:** Hive-on-Tez does not work with encrypted shuffle.
- **BUG-15376: (YARN-1892)** CS fast scheduling patch ends up causing excessive logging.

**Problem:** Seeing about 1 GB of logs per hour.

## 2.10.9. Known Issues for WebHCat

- **BUG-16660:** On Tez setup, Hive jobs in webhcat run in default mr mode even in Hive.

**Problem:** Currently when we run Hive jobs through Webhcat we always run in MR mode even though we are running them in a cluster where Hive queries would have run in Tez mode. This is only on Linux installs. The problem here is that we run hive queries using hive.tar.gz on HDFS and specifying explicit hive configurations, here are the properties that we use in webhcat-site.xml:

templeton.hive.archive	hdfs://apps/webhcat/hive.tar.gz
templeton.hive.path	value: hive.tar.gz/hive/bin/hive
templeton.hive.home	value: hive.tar.gz/hive
templeton.hive.properties	hive.metastore.local=false, hive.metastore.uris=thrift://hivehost:9083, hive.metastore.sasl.enabled=false, hive.metastore.execute.setugi=true

When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add

"hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

## 2.10.10. Known Issues for Oozie

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTask. MetaException(message:Got exception: org.apache.hadoop.security.AccessControlException Permission denied: user=hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

**Problem:** Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16397:** Documentation does not describe how to renew an expired Oozie authorization token.

**Problem:** Due to the introduction of [HADOOP-10416](#), users may encounter situations where they are not authorized to perform certain actions because they are in an unsecured environment and their Oozie authorization token has expired.

**Workaround:** Remove the oozie auth token cache file at `System.getProperty("user.home")/.oozie-auth-token`, then re-run the Oozie command to renew the token.

- **BUG-13551:** Oozie does not understand `_HOST` in the Kerberos principal name.

**Problem:** Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack, where we can just send `_HOST` and at run time the service replaces `_HOST` with machine hostname. This is important so that in a HA setup we can push the same configs to all Oozie servers.

- **BUG-10400:** Oozie workflows that contain Hive queries which run mapreduce jobs fail on secure clusters.

**Problem:** There is a bug in Hive ([HIVE-5618](#)) where delegation tokens are requested for a user who does not have the ability to do so (such as when it is launched from Oozie).

**Workaround:** Set the configuration parameter before any query statements in the script file are launched as part of the Hive action.

```
hive.server2.enable.doAs = false
```

This parameter instructs Hive not to request delegation tokens, which should not be done when running under Oozie.

## 2.10.11. Known Issues for Hue

- **BUG-9734:** Data loss during Migration of Hue DB from default (sqlite) to Oracle DB:

**Problem:** Migration of data and tables from SQLite to Oracle does not work and needs to be performed manually.

1. Install Hue and start Hue (Hue creates table in sqlite db).
2. Do NOT perform any tasks (such as uploading files, pig jobs, or hcat jobs) on the HDP stack from Hue UI.
3. Stop Hue, configure Oracle.
4. Start Hue.

Result: Hue starts fine and continues working, but there is loss of data.

5. Some tables are lost in HCatalog.
6. Some pig scripts do not show up on UI.

**Workaround:** Manually migrate the data and tables from SQLite to Oracle.

## 2.10.12. Known Issues for Flume

- **BUG-16771:** ([HIVE-6867](#)) Hive table returns multiple copies of streaming data when testing the Hive Server restart scenario.

**Problem:** When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

## 2.10.13. Known Issues for Storm

- **BUG-16232:** Storm python support can use wrong version of python if supervisor host has more than one version of python installed.

**Problem:** Storm requires the default system python interpreter to be version 2.6 or higher. Earlier versions of python can see this conflict.

**Workaround:** Ensure that the default system python interpreter is version 2.6 or higher.

- **BUG-15960:** Worker node gets 'FileNotFoundException : stormconf.ser'.

**Problem:** While running Storm-HDFS topologies in a secure environment, the following error was observed in the worker node:

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
initialization")
```

## 2.10.14. Known Issues for Falcon

- **BUG-16608: (FALCON-390)** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** Falcon generated hive-action does not pass the hive-site.xml with the right configuration parameters. One manifestation of the problem will be the failure in table import job where user "hive" will be used to write to a directory owned by the table owner. This is because hive.metastore.execute.setugi parameter is not being passed as part of the hive action.

**Workaround:** Add a Hive default configuration to Oozie.



Stop the Oozie service.



### Warning

This change allows you to work with Hive tables and Oozie workflows, but will impact all Hive actions, including non-Falcon Oozie workflows.

Under the oozie configuration directory (typically `/etc/oozie/conf`), there will be a subdirectory called `action-conf`. Under that directory, either create or modify the file `hive-site.xml` and add the following:

```
<property>
  <name>hive.metastore.execute.setugi</name>
  <value>true</value>
</property>
```

After making this change restart the Oozie service. If Oozie is configured for HA, perform this configuration change on all Oozie server nodes.

- **BUG-16290: (FALCON-389)** Oozie config changes needed to support HCat replication in Falcon

**Problem:** Oozie config changes are needed before Falcon can handle HCat replication.

**Workaround:** Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.
2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to `/etc/hadoop/conf-1` and the other to `/etc/hadoop/conf-2`.
3. For each `oozie-site.xml` file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

For example, if Falcon connects to 3 clusters, specify:

```

<property>
  <name>oozie.service.HadoopAccessorService.hadoop.configurations</
name>
  <value>*/etc/hadoop/
conf,$NameNode:$rpcPortNN=$hadoopConfDir1,$ResourceManager1:$rpcPortRM=$hadoopConfDir1,$
:$rpcPortNN=$hadoopConfDir3,$ResourceManager3:$rpcPortRM
=$hadoopConfDir3</value>
  <description>
    Comma separated AUTHORITY=HADOOP_CONF_DIR, where AUTHORITY is
the HOST:PORT of
the Hadoop service (JobTracker, HDFS). The wildcard '*'
configuration is
used when there is no exact match for an authority. The
HADOOP_CONF_DIR contains
the relevant Hadoop *-site.xml files. If the path is relative is
looked within
the Oozie configuration directory; though the path can be
absolute (i.e. to point
to Hadoop client conf/ directories in the local filesystem.
  </description>
</property>

```

4. Restart the Oozie service on all clusters.

## 2.10.15. Known Issues for Knox

- **BUG-16592:** When accessing through Knox Gateway, Oozie not supported for HDP for Windows or Linux single-node clusters.

**Problem:** When accessing a Hadoop cluster through an Apache Knox Gateway, Oozie is not supported for HDP for Windows or HDP for Linux single-node clusters.

**Workaround:** Either use Oozie only when accessing a multinode HDP cluster in Linux, or access Oozie without going through Knox.

- **BUG-14461:** Knox has Kerberos config at the global level rather than cluster topology level.

**Problem:** The gateway.hadoop.kerberos.secured=false Boolean flag indicates whether the Hadoop cluster protected by Gateway is secured with Kerberos in gateway-site.xml.

## 2.10.16. Known Issues for the Hortonworks Connector for Teradata

- The Hortonworks Connector for Teradata is not supported at this time. Please check <http://www.hortonworks.com> often for updates, which will be coming soon.

## 2.10.17. Known Issues for Ambari

- See [Ambari Release Notes](#).

## 2.11. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

## 2.12. Third-party Licenses

Table 2.2. Third-party Licenses

HDP Component	License
Phoenix	<a href="#">EPL</a>
Storm	<a href="#">EPL</a>
Accumulo	<a href="#">JCommander</a>
Falcon	<a href="#">CERN</a>
Falcon	<a href="#">Tinkerpop</a>
Knox	<a href="#">ANTLR</a>
Knox	<a href="#">MIT</a>
Knox	<a href="#">EPL</a>
Knox	<a href="#">Bouncy Castle</a>
Knox	<a href="#">OWS</a>

## 3. Release Notes HDP-2.1.1

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.1](#)
- [Unsupported Apache components](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

### 3.1. Product Version: HDP-2.1.1

All HDP 2.1 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 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.



#### Note

The minimum level of Apache Ambari to use with HDP 2.1, is version 1.5.1.

This release of Hortonworks Data Platform (HDP) deploys the following Hadoop-related components:

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Tez 0.4
- Apache ZooKeeper 3.4.5
- Hue 2.3.1

- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0
- Third party components:
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

## 3.2. Unsupported Apache components

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation (Apache JIRA HDFS-1052)
- viewFS (Apache JIRA HADOOP-7257)
- viewFS (Apache JIRA HADOOP-7257)

The following Apache Components are shipped as part of HDP 2.1 YARN, but are not supported:

- Application Timeline Server (Hive-on-Tez metrics)
- AM failure/restart resiliency
- MapReduce Uper AM
- YARN CGroup resource isolation
- Admin Node labels
- CPU Scheduling

- Fair Scheduler
- MapReduce Eclipse Plug-in

## 3.3. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- [What's Changed in Mahout](#)
- [HDP 2.1 Clusters Deployed via Ambari](#)
- [What's Changed in Hue](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in Oozie](#)

### 3.3.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

Deprecated algorithms were removed (<https://issues.apache.org/jira/browse/MAHOUT-1296>) without Frequent Pattern Mining. Multilayer Perceptron was added (<https://issues.apache.org/jira/browse/MAHOUT-1265>).

### 3.3.2. HDP 2.1 clusters deployed via Ambari

Freshly-installed HDP 2.1 clusters deployed via Ambari will have the new Hive authorization system turned on by default. Manual installs and upgrades are not affected by this, only fresh HDP 2.1 installs via Ambari. This default behavior will be modified in Ambari 1.6.0.

### 3.3.3. Hue behavioral changes

Hue now supports the optional ability to input Unix usernames in lowercase letters and have Active Directory return usernames in upper case, in cases where LDAP/Active directory is being used as the back end.

### 3.3.4. HBase behavioral changes

In HDP 2.1, if the user does not have read privileges to a table and scans that table he will get an empty result set back.

### 3.3.5. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable `hive.server2.enable.doAs` is set to true, before the user starts the HiveServer2 process, they should create a scratch

directory, `/tmp/hive-<username>`, on HDFS, where `<username>` is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

### 3.3.6. Oozie behavioral changes

When Oozie has been installed manually, before a site can execute any Oozie actions, the Oozie shared libraries must be made explicitly available. Add the following information to the file `oozie-site.xml`:

```
<property>
  <name>oozie.service.WorkflowAppService.system.libpath</name>
  <value>/user/$username/share/lib</value>
</property>
```

## 3.4. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive//HCat](#)
- [Patch Information for Oozie](#)



### Note

Apache YARN, Apache MapReduce and Apache Knox require no additional patches.

### 3.4.1. Patch information for Hadoop Common/HDFS

Hadoop is based on Apache Hadoop 2.4 and includes the following additional patches:

- [HDFS-5257](#): `addBlock()` retry should return `LocatedBlock` with locations else client will get AIOBE
- [HDFS-5089](#): When a `LayoutVersion` supports `SNAPSHOT`, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- [HADOOP-10475](#): `ConcurrentModificationException` in `AbstractDelegationTokenSelector.selectToken()`.

- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.

### 3.4.2. Patch information for ZooKeeper

ZooKeeper is based on Apache ZooKeeper 3.4.5 and includes the following patches:

- [ZOOKEEPER-1702](#): ZooKeeper client may write operation packets before receiving successful response to connection request, can cause TCP RST.

### 3.4.3. Patch information for HBase

HBase is based on Apache HBase 0.98.0. It includes the following patches:

- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10592](#): Refactor PerformanceEvaluationtool
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10688](#): Add a draining\_node script to manage nodes in draining mode
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10670](#): HBaseFsck#connect() should use new connection



- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94\* to hbase0.96+
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

#### Windows Fixes:

- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows

#### Changes related to HBASE-10070:

- [HBASE-10875](#) Metas own location should be cached

- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool
- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#)  
TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION\_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.

- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

### 3.4.4. Patch information for Pig

Pig is based on Apache Pig 0.12.1. It includes the following patches:

- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

### 3.4.5. Patch information for Tez

Tez is based on Apache Tez 0.4.0 incubating release. It includes the following patches:

- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V\_ROUTE\_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V\_INIT at INITED.

### 3.4.6. Patch information for Hive/HCatalog

Hive is based on Apache Hive 0.13.0. Apache HCatalog is now merged with Apache Hive. Hive includes the following patches:

- [HIVE-6117](#): HBase\_1 and HBase\_2 tests are failing
- [HIVE-5601](#): NPE in ORC's PPD when using SELECT \* from table with WHERE predicate
- [HIVE-5542](#): WebHCat is failing to run ddl command on a secure cluster
- [HIVE-5515](#): Writing to an HBase table throws IllegalArgumentException, failing job submission

- [HIVE-5511](#): `percentComplete` returned by job status from WebHCat is null
- [HIVE-5496](#): `hcat -e drop database if exists` fails on authorizing non-existent null db
- [HIVE-5485](#): SBAP errors on null partition being passed into partition level authorization
- [HIVE-5484](#): TestSchemaTool failures when Hive version has more than 3 revision numbers
- [HIVE-5480](#): WebHCat e2e tests for doAs feature are failing
- [HIVE-5479](#): SBAP restricts `hcat -e 'show databases'`
- [HIVE-5478](#): WebHCat e2e testsuite for hcat authorization tests needs some fixes
- [HIVE-5474](#): drop table hangs when `concurrency=true`
- [HIVE-5453](#): `jobsubmission2.conf` should use 'timeout' property
- [HIVE-5448](#): WebHCat duplicate test TestMapReduce\_2 should be removed
- [HIVE-5425](#): Provide a configuration option to control the default stripe size for ORC
- [HIVE-5422](#): Upgrade Kryo to 2.22 now that it is released
- [HIVE-5411](#): Migrate expression serialization to Kryo
- [HIVE-5379](#): `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- [HIVE-5353](#): job submission that requires access to metastore should not require additional jars to be shipped to target node
- [HIVE-5290](#): Some HCatalog tests have been behaving flaky
- [HIVE-5279](#): Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`
- [HIVE-5263](#): Query Plan cloning time could be improved by using Kryo
- [HIVE-5133](#): webhcat jobs that need to access metastore fails in secure mode
- [HIVE-5112](#): Upgrade protobuf to 2.5 from 2.4
- [HIVE-5070](#): Need to implement `listLocatedStatus()` in `ProxyFileSystem` for 0.23 shim
- [HIVE-4910](#): Hadoop 2 archives broken
- [HIVE-4545](#): HS2 should return describe table results without space padding
- [HIVE-4485](#): beeline prints null as empty strings
- [HIVE-4388](#): HBase tests fail against Hadoop 2
- [HIVE-3815](#): hive table rename fails if filesystem cache is disabled

- [HIVE-1511](#): Hive plan serialization is slow.

### 3.4.7. Patch information for Oozie

Oozie is based on Apache Oozie 4.0.0 and includes the following patches:

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [Oozie-615](#): Support Oozie HA.
- [Oozie-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [Oozie-1306](#): Bring cron syntax to coordinator frequency.
- [Oozie-1460](#): Implement and document oozie HA security.
- [Oozie-1486](#): cut down on number of small files to track a running action.
- [Oozie-1491](#): Make sure oozie works with secure ZooKeeper.
- [Oozie-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [Oozie-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [Oozie-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- [Oozie-1541](#): typo in oozie HA admin-server command line documentation.
- [Oozie-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [Oozie-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [Oozie-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [Oozie-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [Oozie-1576](#): Add documentation for oozie sqoop CLI.
- [Oozie-1587](#): Add "recovery" column to CoordJob table.
- [Oozie-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.
- [Oozie-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [Oozie-1608](#): update curator to 2.4.0 when it's available to fix security hole.

- [Oozie-1618](#): Dryrun should check variable substitution in workflow.xml.
- [Oozie-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [Oozie-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [Oozie-1726](#): Oozie does not support \_HOST when configuring kerberos security.
- [Oozie-1733](#): Fix test failures by oozie-1722.

### 3.4.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.4 and includes the following patches:

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): -password-file option triggers FileSystemClosed exception at end of Oozie action.
- [SQOOP-1260](#): HADOOP\_MAPRED\_HOME should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with -queries because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling autoCommit on created connections.
- [SQOOP-1246](#): HBaseImportJob should add job authToken only if HBase is secured.

- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP\_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding \*\_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

## 3.5. Minimum System Requirements

In this section:



- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

### 3.5.1. Hardware recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

### 3.5.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*



#### Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).

### 3.5.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

### 3.5.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

### 3.5.5. Virtualization and cloud platforms

HDP is certified and supported when running on virtual or cloud platforms (for example, VMware vSphere or Amazon Web Services EC2) as long as the respective guest OS is supported by HDP and any issues that are detected on these platforms are reproducible on the same supported OS installed on bare metal.

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

### 3.5.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



#### Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 3.6. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- Storm
- Falcon
- Tez
- YARN
- Phoenix
- MapReduce2

- Improved SQL compliance (GRANT, REVOKE, WHERE clause subqueries, common table expressions, data type support)

## 3.7. Common Vulnerabilities and Exposures

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability
- **Severity:** Major
- **Vendor:** The Apache Software Foundation
- **Versions Affected:** Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0
- **Users Affected:** Users who have enabled Hadoop's MapReduce security features
- **Impact:** Vulnerability allows an unauthorized user to retrieve job details from the job history server
- **Mitigation:** Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to 2.3.0
- **Credit:** This issue was discovered by Koji Noguchi of Yahoo

## 3.8. Known Issues

In this section:

- [Known Limitation for Oracle DB Metastore](#)
- [Known Issues for SLES 11](#)
- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for Oozie](#)
- [Known Issues for Hue](#)
- [Known Issues for Flume](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

### 3.8.1. Known Limitation for HDFS Upgrade

**BUG-17431:** ([HDFS-5526](#)) If you upgrade from HDP 2.0.9.1 to HDP 2.1, the DataNode cannot be rolled back to HDP 2.0.9.1. Contact Support for an update to fix this HDP 2.0.9.1 issue.

### 3.8.2. Known Limitation for Oracle DB Metastore

HDP 2.1.1 does not support the use of Oracle DB as a metastore. Upgrade to HDP 2.1.2 or higher.

### 3.8.3. Known Issues for SLES 11

- **BUG-9904:** php\_curl Required for SLES 11 Sp1

**Problem:** Several alerts return with `Return code of 255 is out of bounds` while trying to install a cluster on SLES because `php_curl` is not installed.

**Workaround:** Install `php_curl` on your SLES host.

```
zypper install php-curl
```

### 3.8.4. Known Issues for HDP

- **BUG-15796:** Sigsegv in mapred history server due to segfault in `JniBasedUnixGroupsMapping`.

**Problem:** RHEL 6.2 and RHEL 6.3 contain known bugs in `nsld`. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by `nsld` times out.

**Workaround:** If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.
- Increase the configured `nsld` search timeout in `/etc/nsld.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGro` in `core-site.xml`.
- **BUG-825:** EC2 m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

**Problem:** Directories and disks that you assign for logging in HDP do NOT have enough space to maintain logs during HDP operations.

**Workaround:** Designate least 10 GB of free space on a disk that will be used by HDP logging.

#### 3.8.4.1. Known issues for HDFS

- **BUG-14542:** HDP 2.1 exception during namenode service work.

**Problem:** After the start of the NameNode service, the following exception occurred:

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_00000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msec
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
9 more
```

### 3.8.4.2. Known issues for MapReduce

- **BUG-12005:** Mapreduce.task.io.sort.mb is capped at 2047.

**Problem:** mapreduce.task.io.sort.mb is hardcoded to not allow values larger than 2047. If you enter a value larger then this the map tasks will always crash at this line:

```
https://github.com/apache/hadoop-mapreduce/blob/HDFS-641/src/java/org/apache/hadoop/mapred/MapTask.java?source=cc#L746
```

- **BUG-14749:** CombineFileInputFormat.getSplits() including directories in its results.

**Problem:** This is causing Hive test root\_dir\_external\_table.q to fail when running against hadoop-2. Opened Apache Jira MAPREDUCE-5756 Created in Monarch as <https://hwxmonarch.atlassian.net/browse/HADOOP-801>, creating equivalent bug for Baikal.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

### 3.8.5. Known Issues for YARN

- **BUG-158341:** YARN and/or mapred client should add tokens for default filesystem.

**Problem:** As noticed in BUG-15360 if jobs are using webhdfs then they could run into an issue where the job only has webhdfs tokens and yarn jobs would fail as it tries to talk to hdfs over default fs since tokens for default fs as not part of the job.

- **BUG-15376:** {YARN-1892} CS fast scheduling patch ends up causing excessive logging.

**Problem:** Seeing about 1 GB of logs per hour.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.
- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.  
**Problem:** YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.
- **BUG-12327:** [Yarn-90] NM cannot detect when bad disks become healthy again.  
**Problem:** If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.
- **BUG-7531:** Hadoop metrics link does not contain correct content.  
**Problem:** In the Resource Manager UI Tools section, clicking on Logs and Metrics opens pages that do not contain correct information.

### 3.8.6. Known Issues for HBase

- **BUG-16900 (HBASE-11036):** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.  
**Problem:** The Big Linked List Test with Chaos Monkey Test run fails with a Not Serving Region exception in the YARN logs.
- **BUG-16513:** HBCK Tests Fail Intermittently Due to NotServingRegionException.  
**Problem:** The HBCK Tool tests fails intermittently due to a NotServingRegionException, noted in the Master Logs.
- **BUG-16257:** HBase master fails to start due to BindException.  
**Problem:** HBase on Suse 11 64 bit, smoke test fails intermittently with:

```
ERROR [main] client.ConnectionManager$HConnectionImplementation: The node /
hbase is not in ZooKeeper.
```
- **BUG-14986:** HBase Bloomberg HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.  
**Problem:** On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.
- **BUG-12167, HBASE-10304** Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString  
**Problem:** Some MapReduce jobs fail to launch. An exception similar to the following displays:

```
Exception in thread "main" java.lang.IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString
    at java.lang.ClassLoader.defineClass1(Native Method)
    at java.lang.ClassLoader.defineClass(ClassLoader.java:792)
    at java.security.SecureClassLoader.defineClass(SecureClassLoader.java:142)
    at java.net.URLClassLoader.defineClass(URLClassLoader.java:449)
    at java.net.URLClassLoader.access$100(URLClassLoader.java:71)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:361)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
    at org.apache.hadoop.hbase.protobuf.ProtobufUtil.toScan(ProtobufUtil.java:818)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.convertScanToString(TableMapReduceUtil.java:433)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:186)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:147)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:270)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:100)
    ...
```

This issue occurs because of an optimization introduced in [HBASE-9867](#) that inadvertently introduced a classloader dependency. This affects both jobs using the `-libjars` option and "fat jar," jobs which package their runtime dependencies in a nested lib folder.

**Workaround:** To satisfy the new classloader requirements, include `hbase-protocol.jar` in Hadoop's classpath. For a system-wide resolution, include a reference to the `hbase-protocol.jar` in Hadoop's lib directory, using a symlink or by copying the jar into the new location.

To resolve on a per-job launch basis, specify a value for `HADOOP_CLASSPATH` at job submission time. If you are launching jobs that package their dependencies, all three of the following job launching commands satisfy this requirement:

```
$ HADOOP_CLASSPATH=/path/to/hbase-protocol.jar:/path/to/hbase/conf hadoop
  jar MyJob.jar MyJobMainClass
$ HADOOP_CLASSPATH=$(hbase mapredcp):/path/to/hbase/conf hadoop jar MyJob.
  jar MyJobMainClass
$ HADOOP_CLASSPATH=$(hbase classpath) hadoop jar MyJob.jar MyJobMainClass
```

If you are using jars that do not package their dependencies, use the following command structure:

```
$ HADOOP_CLASSPATH=$(hbase mapredcp):/etc/hbase/conf hadoop jar MyApp.jar
  MyJobMainClass -libjars $(hbase mapredcp | tr ':' ',') ...
```

### 3.8.7. Known Issues for Phoenix

- **BUG-16484:** Phoenix ZooKeeper quorum string cannot contain the port number.

**Problem:** HDP 2.1 defines port numbers in `hbase.zookeeper.quorum` in `hbase-site.xml`, which causes conflicts when you use Phoenix on HBase. This results in an error message similar to the following:

```
java.sql.SQLException: ERROR 102 (08001): Malformed connection url.
```

- **Workaround:** Remove the port number from `hbase.zookeeper.quorum` in `hbase-site.xml`, and include the port number in the JDBC connector string

```
jdbc:phoenix [ :<zookeeper quorum> [:<port number > ] [ :/hbase ]]
```

### 3.8.8. Known Issues for Hive

- **BUG-17603: (HIVE-6985)**(HIVE-6985) Grant privileges on PUBLIC role are not being honored.

**Problem:** When a privilege is granted to public role, the privilege is supposed to be applicable for all users. However, the privilege check fails for users, even if they have public role in the list of current roles. (This issue only exists with the Public role; it does not affect the granting of privileges to other roles.)

- **BUG-16890:** Hive SQL standard auth calls accessing local or HDFS URLs fail in Kerberos secure cluster with binary HS2 transport.

**Problem:** This is blocking all CREATE table calls where we access LOCAL or HDFS uri.

```
>>> create external table studenttab10k(
name string,
age int,
gpa double)
row format delimited
fields terminated by '\t'
stored as textfile
location '/user/hcat/tests/data/studenttab10k';
2014-04-17 00:12:13,627 DEBUG [main] transport.TSaslTransport: writing data
length: 297
2014-04-17 00:12:13,657 DEBUG [main] transport.TSaslTransport: CLIENT:
reading data length: 351
Error: Error while compiling statement: FAILED: HiveAccessControlException
Permission denied.
Principal [name=hrt_qa@HORTON.YGRIDCORE.NET, type=USER] does not have
following privileges on Object
[type=DFS_URI, name=/user/hcat/tests/data/studenttab10k] : [INSERT, DELETE,
OBJECT OWNERSHIP] (state=42000,code=40000)
```

- **BUG-16660:** On Tez setup, Hive jobs in webhcat run in default mr mode even in Hive.

**Problem:** Currently when we run Hive jobs through Webhcat we always run in MR mode even though we are running them in a cluster where Hive queries would have run in Tez mode. This is only on Linux installs. The problem here is that we run hive queries using `hive.tar.gz` on HDFS and specifying explicit hive configurations, here are the properties that we use in `webhcat-site.xml`:

templeton.hive.archive	hdfs:///apps/webhcat/hive.tar.gz
------------------------	----------------------------------



templeton.hive.path	value: hive.tar.gz/hive/bin/hive
templeton.hive.home	value: hive.tar.gz/hive
templeton.hive.properties	hive.metastore.local=false, hive.metastore.uris=thrift://hivehost:9083, hive.metastore.sasl.enabled=false, hive.metastore.execute.setugi=true

When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add "hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.
- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** The job fails with the following permission error:

```
Copying data from hdfs://arbit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data/dt=2010-01-01-20
Copying file: hdfs://arbit-falcon-2.cs1cloud.internal:8020/projects/ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(message:Got exception: org.apache.hadoop.security.AccessControlException Permission denied: user=hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/HCatReplication_oneSourceOneTarget_hyphen":arbit:hdfs:drwxr-xr-x at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

**Problem:** Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16864:** When Hive standard authorization is enabled, the owner of the table backing index is missing.

**Problem:** The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
(SessionState.java:printError(546)) - FAILED: HiveAccessControlException
```

```

Permission denied. Principal [name=hrt_qa, type=USER] does not have
following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
name=default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT
OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...

```

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

**Problem:** The Hive session should shut down cleanly and not kill the app.

- **BUG-16771:** (Apache Bug: HIVE-6867) Hive table has multiple copies of streaming data when testing the Hive Server restart scenario.

**Problem:** When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

**Problem:** We force create\_index to run in MR mode when we have a TEZ run. But it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16393:** Bucketized Table feature fails in some cases.

**Problem:** Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination. Example follows:

- ```
CREATE TABLE P1(key STRING, val STRING)
  CLUSTERED BY (key) SORTED BY (key) INTO 2 BUCKETS STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/Users/jpullokkaran/apache-hive1/data/files/P1.txt'
  INTO TABLE P1;

-- perform an insert to make sure there are 2 files
INSERT OVERWRITE TABLE P1 select key, val from P1;
```

- **Workaround:** Avoid loading data for bucketed table.

- **BUG-16391:** Streaming transactions fail on MSSQL.

**Problem:** After creating tables using the MSSQL composite script provided by BUG-15827 running Flume, Hive Sink tests failed because no data made it into Hive tables.

- **BUG-15733:** Schema evolution is broken on Tez.

**Problem:** The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
  java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
```

```

at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more

```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-8227:** Hive needs to implement recovery or extend FileOutputComitter.

**Problem:** When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

```

public boolean isRecoverySupported() {
    return false;
}

```

### 3.8.9. Known Issues for Tez

- **BUG-15376:** {YARN-1892} CS fast scheduling patch ends up causing excessive logging.

**Problem:** Seeing about 1 GB of logs per hour.

### 3.8.10. Known Issues for Oozie

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.
ql.exec.DDLTask. MetaException(message:Got exception: org.apache.
hadoop.security.AccessControlException Permission denied: user=
hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/
HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.
checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

**Problem:** Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16397:** Documentation does not describe how to renew an expired Oozie authorization token.

**Problem:** Due to the introduction of [HADOOP-10416](#), users may encounter situations where they are not authorized to perform certain actions because they are in an unsecured environment and their Oozie authorization token has expired.

**Workaround:** Remove the oozie auth token cache file at `System.getProperty("user.home")/.oozie-auth-token`, then re-run the Oozie command to renew the token.

- **BUG-13551:** Oozie does not understand \_HOST in the Kerberos principal name.

**Problem:** Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack, where we can just send \_HOST and at run time the service replaces \_HOST with machine hostname. This is important so that in a HA setup we can push the same configs to all Oozie servers.

- **BUG-10177:** Oozie workflows that contain Hive queries which run mapreduce jobs fail on secure clusters.

**Problem:** There is a bug in Hive ([HIVE-5618](#)) where delegation tokens are requested for a user who does not have the ability to do so (such as when it is launched from Oozie).

**Workaround:** Set the configuration parameter before any query statements in the script file are launched as part of the Hive action.

```
hive.server2.enable.doAs = false
```

This parameter instructs Hive not to request delegation tokens, which should not be done when running under Oozie.

- **BUG-9671:** Oozie reports the job as failed when the app and job completed successfully when RM is restarted multiple times

**Problem:** From the Oozie log:

```
2013-10-05 23:04:58,952 DEBUG HadoopAccessorService:545 - USER[hrt_qa]
GROUP[-] TOKEN[] APP[wordcount-wf] JOB[0000003-131005052220011-oozie-oozi-
W] ACTION[0000003-131005052220011-oozie-oozi-W@wc] Checking
if filesystem hdfs is supported
2013-10-05 23:04:58,954 WARN MapReduceActionExecutor:542 - USER[hrt_qa]
GROUP[-] TOKEN[] APP[wordcount-wf] JOB[0000003-131005052220011-oozie-oozi-
W] ACTION[0000003-131005052220011-oozie-oozi-W@wc] Launch
erMapper died, check Hadoop log for job [hor12n01.gq1.ygridcore.
net:8032:job_1381013595258_0001]
```

But this job and the application complete successfully.

### 3.8.11. Known Issues for Hue

- **BUG-9734:** Data loss during Migration of Hue DB from default (sqlite) to Oracle DB:

**Problem:** Migration of data and tables from SQLite to Oracle does not work and needs to be performed manually.

1. Install Hue and start Hue, (Hue creates table in sqlite db).
2. Do NOT perform any tasks (such as uploading files, pig jobs, or hcat jobs) on the HDP stack from Hue UI.
3. Stop Hue, configure Oracle.
4. Start Hue.

Result: Hue starts fine and continues working, but there is loss of data.

5. Some tables are lost in HCatalog.
6. Some pig scripts do not show up on UI.

**Workaround:** Manually migrate the data and tables from SQLite to Oracle.

### 3.8.12. Known Issues for Flume

- **BUG-16771:** Hive table returns multiple copies of streaming data when testing the Hive Server restart scenario.

**Problem:** When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

### 3.8.13. Known Issues for Storm

- **BUG-16232:** Storm python support can use wrong version of python if supervisor host has more than one version of python installed.

**Problem:** Storm requires the default system python interpreter to be version 2.6 or higher. Earlier versions of python can see this conflict.

**Workaround:** Ensure that the default system python interpreter is version 2.6 or higher.

- **BUG-15960:** Worker node gets 'FileNotFoundException : stormconf.ser'.

**Problem:** While running Storm-HDFS topologies in a secure environment, the following error was observed in the worker node:

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$_main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
initialization")
```

### 3.8.14. Known Issues for Falcon

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

**Problem:** Falcon generated hive-action does not pass the hive-site.xml with the right configuration parameters. One manifestation of the problem will be the failure in table import job where user "hive" will be used to write to a directory owned by the table

owner. This is because `hive.metastore.execute.setugi` parameter is not being passed as part of the hive action.

**Workaround:** Add a Hive default configuration to Oozie.

Stop the Oozie service.



### Warning

This change allows you to work with Hive tables and Oozie workflows, but will impact all Hive actions, including non-Falcon Oozie workflows.

Under the oozie configuration directory (typically `/etc/oozie/conf`), there will be a subdirectory called `action-conf`. Under that directory, either create or modify the file `hive-site.xml` and add the following:

```
<property>
  <name>hive.metastore.execute.setugi</name>
  <value>true</value>
</property>
```

After making this change restart the Oozie service. If Oozie is configured for HA, perform this configuration change on all Oozie server nodes.

- **BUG-16290:** Strange delegation token issues in secure clusters

**Problem:** Inconsistencies in rules for `hadoop.security.auth_to_local` can lead to issues with delegation token renewals in secure clusters.

**Workaround:** Verify that `hadoop.security.auth_to_local` in `core-site.xml` is consistent across all clusters.

- **BUG-16290, FALCON-389:** Oozie config changes needed to support HCat replication in Falcon

**Problem:** Oozie config changes are needed before Falcon can handle HCat replication.

**Workaround:** Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.
2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to `/etc/hadoop/conf-1` and the other to `/etc/hadoop/conf-2`.
3. For each `oozie-site.xml` file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

For example, if Falcon connects to 3 clusters, specify:



```

<property>
  <name>oozie.service.HadoopAccessorService.hadoop.configurations</
name>
  <value>*/etc/hadoop/
conf,$NameNode:$rpcPortNN=$hadoopConfDir1,$ResourceManager1:$rpcPortRM=$hadoopConfDir1,$
:$rpcPortNN=$hadoopConfDir3,$ResourceManager3:$rpcPortRM
=$hadoopConfDir3</value>
  <description>
    Comma separated AUTHORITY=HADOOP_CONF_DIR, where AUTHORITY is
the HOST:PORT of
the Hadoop service (JobTracker, HDFS). The wildcard '*'
configuration is
used when there is no exact match for an authority. The
HADOOP_CONF_DIR contains
the relevant Hadoop *-site.xml files. If the path is relative is
looked within
the Oozie configuration directory; though the path can be
absolute (i.e. to point
to Hadoop client conf/ directories in the local filesystem.
  </description>
</property>

```

4. Restart the Oozie service on all clusters.

### 3.8.15. Known Issues for Knox

- **BUG-16592:** When accessing through Knox Gateway, Oozie not supported for HDP for Windows or Linux single-node clusters.

**Problem:** When accessing a Hadoop cluster through an Apache Knox Gateway, Oozie is not supported for HDP for Windows or HDP for Linux single-node clusters.

**Workaround:** Either use Oozie only when accessing a multinode HDP cluster in Linux, or access Oozie without going through Knox.

- **BUG-14461:** Knox has Kerberos config at the global level rather than cluster topology level.

**Problem:** The gateway.hadoop.kerberos.secured=false Boolean flag indicates whether the Hadoop cluster protected by Gateway is secured with Kerberos in gateway-site.xml.

### 3.8.16. Known Issues for the Hortonworks Connector for Teradata

- The Hortonworks Connector for Teradata is not supported at this time. Please check <http://www.hortonworks.com> often for updates, which will be coming soon.

### 3.8.17. Known Issues for Ambari

- See [Ambari Release Notes](#).

## 3.9. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

## 3.10. Third-party Licenses

Table 3.1. Third-party Licenses

HDP Component	License
Phoenix	<a href="#">EPL</a>
Storm	<a href="#">EPL</a>
Accumulo	<a href="#">JCommander</a>
Falcon	<a href="#">CERN</a>
Falcon	<a href="#">Tinkerpop</a>
Knox	<a href="#">ANTLR</a>
Knox	<a href="#">MIT</a>
Knox	<a href="#">EPL</a>
Knox	<a href="#">Bouncy Castle</a>
Knox	<a href="#">OWS</a>