

# Hortonworks Data Platform

## Release Notes

(Jan 22, 2015)

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

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-1.3.10 .....	1
1.1. Product Version: HDP-1.3.10 .....	1
1.2. What's Changed in this Release .....	2
1.3. Patch Information .....	2
1.3.1. Patch information for Hadoop .....	3
1.3.2. Patch information for Ambari .....	5
1.3.3. Patch information for HBase .....	5
1.3.4. Patch information for Hive .....	6
1.3.5. Patch information for HCatalog .....	8
1.3.6. Patch information for Pig .....	8
1.3.7. Patch information for ZooKeeper .....	8
1.3.8. Patch information for Oozie .....	8
1.3.9. Patch information for Sqoop .....	9
1.3.10. Patch information for Mahout .....	9
1.3.11. Patch information for Flume .....	9
1.4. Upgrading HDP 1.3.9 to 1.3.10 .....	10
1.5. Known Issues .....	13
1.5.1. Known Issues for Hadoop .....	13
1.5.2. Known Issues for Hive .....	13
1.5.3. Known Issues for WebHCatalog .....	16
1.5.4. Known Issues for HBase .....	16
1.5.5. Known Issues for Oozie .....	16
1.5.6. Known Issues for Ambari .....	17
1.6. Security Fixes .....	17
2. Release Notes HDP-1.3.9 .....	18
2.1. Product Version: HDP-1.3.9 .....	18
2.2. What's Changed in this Release .....	19
2.3. Patch Information .....	19
2.3.1. Patch information for Hadoop .....	19
2.3.2. Patch information for Ambari .....	22
2.3.3. Patch information for HBase .....	22
2.3.4. Patch information for Hive .....	23
2.3.5. Patch information for HCatalog .....	24
2.3.6. Patch information for Pig .....	24
2.3.7. Patch information for ZooKeeper .....	25
2.3.8. Patch information for Oozie .....	25
2.3.9. Patch information for Sqoop .....	25
2.3.10. Patch information for Mahout .....	25
2.3.11. Patch information for Flume .....	26
2.4. Upgrading HDP 1.3.7 to 1.3.9 .....	27
2.5. Known Issues .....	30
2.5.1. Known Issues for Hadoop .....	30
2.5.2. Known Issues for Hive .....	30
2.5.3. Known Issues for WebHCatalog .....	33
2.5.4. Known Issues for HBase .....	33
2.5.5. Known Issues for Oozie .....	33
2.5.6. Known Issues for Ambari .....	34
3. Release Notes HDP-1.3.7 .....	35

3.1. Product Version: HDP-1.3.7 .....	35
3.2. What's Changed in this Release .....	36
3.2.1. What's Changed in Hadoop .....	36
3.2.2. What's Changed in Ambari .....	36
3.2.3. What's Changed in HBase .....	37
3.2.4. What's Changed in Hive .....	37
3.2.5. What's Changed in HCatalog .....	37
3.2.6. What's Changed in Pig .....	37
3.2.7. What's Changed in ZooKeeper .....	37
3.2.8. What's Changed in Oozie .....	37
3.2.9. What's Changed in Sqoop .....	37
3.2.10. What's Changed in Mahout .....	38
3.2.11. What's Changed in Flume .....	38
3.2.12. What's Changed in Hue .....	38
3.3. Patch Information .....	38
3.3.1. Patch information for Hadoop .....	38
3.3.2. Patch information for Ambari .....	40
3.3.3. Patch information for HBase .....	40
3.3.4. Patch information for Hive .....	42
3.3.5. Patch information for HCatalog .....	43
3.3.6. Patch information for Pig .....	43
3.3.7. Patch information for ZooKeeper .....	43
3.3.8. Patch information for Oozie .....	43
3.3.9. Patch information for Sqoop .....	44
3.3.10. Patch information for Mahout .....	44
3.3.11. Patch information for Flume .....	44
3.4. Minimum System Requirements .....	45
3.4.1. Hardware Recommendations .....	46
3.4.2. Operating Systems Requirements .....	46
3.4.3. Software Requirements .....	46
3.4.4. Database Requirements .....	46
3.4.5. Virtualization and Cloud Platforms .....	47
3.4.6. Optional: Configure the Local Repositories .....	47
3.5. Upgrading HDP Manually .....	47
3.6. Improvements .....	48
3.7. Known Issues .....	48
3.7.1. Known Issues for Hadoop .....	48
3.7.2. Known Issues for Hive .....	48
3.7.3. Known Issues for WebHCatalog .....	51
3.7.4. Known Issues for HBase .....	51
3.7.5. Known Issues for Oozie .....	51
3.7.6. Known Issues for Ambari .....	52
4. Release Notes HDP-1.3.3 .....	53
4.1. Product Version: HDP-1.3.3 .....	53
4.2. What's Changed in this Release .....	54
4.2.1. What's Changed in Hadoop .....	54
4.2.2. What's Changed in Ambari .....	55
4.2.3. What's Changed in HBase .....	55
4.2.4. What's Changed in Hive .....	56
4.2.5. What's Changed in Hcatalog .....	56
4.2.6. What's Changed in Pig .....	56

4.2.7. What's Changed in ZooKeeper .....	56
4.2.8. What's Changed in Oozie .....	56
4.2.9. What's Changed in Sqoop .....	57
4.2.10. What's Changed in Mahout .....	57
4.2.11. What's Changed in Flume .....	57
4.2.12. What's Changed in Hue .....	57
4.3. Patch Information .....	57
4.3.1. Patch information for Hadoop .....	58
4.3.2. Patch information for Ambari .....	60
4.3.3. Patch information for HBase .....	60
4.3.4. Patch information for Hive .....	62
4.3.5. Patch information for HCatalog .....	63
4.3.6. Patch information for Pig .....	63
4.3.7. Patch information for ZooKeeper .....	63
4.3.8. Patch information for Oozie .....	63
4.3.9. Patch information for Sqoop .....	63
4.3.10. Patch information for Mahout .....	64
4.3.11. Patch information for Flume .....	64
4.4. Minimum System Requirements .....	65
4.4.1. Hardware Recommendations .....	65
4.4.2. Operating Systems Requirements .....	66
4.4.3. Software Requirements .....	66
4.4.4. Database Requirements .....	66
4.4.5. Virtualization and Cloud Platforms .....	67
4.4.6. Optional: Configure the Local Repositories .....	67
4.5. Upgrading HDP Manually .....	67
4.6. Improvements .....	67
4.7. Known Issues .....	67
4.7.1. Known Issues for Hadoop .....	68
4.7.2. Known Issues for Hive .....	68
4.7.3. Known Issues for WebHCatalog .....	70
4.7.4. Known Issues for HBase .....	71
4.7.5. Known Issues for Oozie .....	71
4.7.6. Known Issues for Ambari .....	72
5. Release Notes HDP-1.3.2 .....	75
5.1. Product Version: HDP-1.3.2 .....	75
5.2. Patch Information .....	76
5.2.1. Patch information for Hadoop .....	76
5.2.2. Patch information for Ambari .....	78
5.2.3. Patch information for HBase .....	78
5.2.4. Patch information for Hive .....	79
5.2.5. Patch information for HCatalog .....	80
5.2.6. Patch information for Pig .....	80
5.2.7. Patch information for ZooKeeper .....	81
5.2.8. Patch information for Oozie .....	81
5.2.9. Patch information for Sqoop .....	81
5.2.10. Patch information for Mahout .....	82
5.2.11. Patch information for Flume .....	82
5.3. Minimum System Requirements .....	83
5.3.1. Hardware Recommendations .....	83
5.3.2. Operating Systems Requirements .....	83

5.3.3. Software Requirements .....	84
5.3.4. Database Requirements .....	84
5.3.5. Virtualization and Cloud Platforms .....	84
5.3.6. Optional: Configure the Local Repositories .....	85
5.4. Upgrading HDP Manually .....	85
5.5. Improvements .....	85
5.6. Known Issues .....	86
5.6.1. Known Issues for Hadoop .....	86
5.6.2. Known Issues for Hive .....	86
5.6.3. Known Issues for WebHCatalog .....	88
5.6.4. Known Issues for HBase .....	89
5.6.5. Known Issues for Oozie .....	90
5.6.6. Known Issues for Ambari .....	92
6. Release Notes HDP-1.3.1 .....	95
6.1. Product Version: HDP-1.3.1 .....	95
6.2. Patch Information .....	96
6.2.1. Patch information for Hadoop .....	96
6.2.2. Patch information for Ambari .....	97
6.2.3. Patch information for HBase .....	97
6.2.4. Patch information for Hive .....	98
6.2.5. Patch information for HCatalog .....	99
6.2.6. Patch information for Pig .....	99
6.2.7. Patch information for ZooKeeper .....	99
6.2.8. Patch information for Oozie .....	100
6.2.9. Patch information for Sqoop .....	100
6.2.10. Patch information for Mahout .....	100
6.2.11. Patch information for Flume .....	100
6.3. Minimum System Requirements .....	101
6.3.1. Hardware Recommendations .....	102
6.3.2. Operating Systems Requirements .....	102
6.3.3. Software Requirements .....	102
6.3.4. Database Requirements .....	102
6.3.5. Virtualization and Cloud Platforms .....	103
6.3.6. Optional: Configure the Local Repositories .....	103
6.4. Upgrading HDP Manually .....	103
6.5. Improvements .....	104
6.6. Known Issues .....	104
6.6.1. Known Issues for Hadoop .....	104
6.6.2. Known Issues for Hive .....	104
6.6.3. Known Issues for WebHCatalog .....	106
6.6.4. Known Issues for HBase .....	107
6.6.5. Known Issues for Oozie .....	107
6.6.6. Known Issues for Ambari .....	109
7. Release Notes HDP-1.3.0 .....	110
7.1. Product Version: HDP-1.3.0 .....	110
7.2. Patch Information .....	111
7.2.1. Patch information for Hadoop .....	111
7.2.2. Patch information for Ambari .....	112
7.2.3. Patch information for HBase .....	112
7.2.4. Patch information for Hive .....	114
7.2.5. Patch information for HCatalog .....	115

---

7.2.6. Patch information for Pig .....	115
7.2.7. Patch information for ZooKeeper .....	115
7.2.8. Patch information for Oozie .....	115
7.2.9. Patch information for Sqoop .....	116
7.2.10. Patch information for Mahout .....	116
7.2.11. Patch information for Flume .....	116
7.3. Minimum System Requirements .....	117
7.3.1. Hardware Recommendations .....	117
7.3.2. Operating Systems Requirements .....	117
7.3.3. Software Requirements .....	118
7.3.4. Database Requirements .....	118
7.3.5. Virtualization and Cloud Platforms .....	118
7.3.6. Optional: Configure the Local Repositories .....	119
7.4. Upgrading HDP Manually .....	119
7.5. Improvements .....	119
7.6. Known Issues .....	120
7.6.1. Known Issues for Hadoop .....	121
7.6.2. Known Issues for Hive .....	121
7.6.3. Known Issues for WebHCatalog .....	123
7.6.4. Known Issues for HBase .....	123
7.6.5. Known Issues for Oozie .....	124
7.6.6. Known Issues for Ambari .....	125

# List of Tables

- 1.1. Operating Systems mapped to each OS Family ..... 13
- 2.1. Operating Systems mapped to each OS Family ..... 29



# 1. Release Notes HDP-1.3.10

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [What's Changed in HDP 1.3.10](#)
- [Patch Information](#)
- [Upgrading HDP Manually](#)
- [Known Issues](#)
- [Known Issues](#)

## 1.1. Product Version: HDP-1.3.10

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6
- Apache Pig 0.11.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog



### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.4.4 (or later)
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Hue 2.2.0
- Third party components:
  - Talend Open Studio for Big Data 5.3
  - Ganglia 3.5.0

- Ganglia Web 3.5.7
- Nagios 3.5.0

## 1.2. What's Changed in this Release

This release contains the following bug fixes:

Component	Apache JIRA	Bug No.	Description
Hadoop Common	N/A, Security Fix	BUG-25265	<a href="#">CVE-2014-3566</a> Custom version of Jetty 6 with bug
HBase	<a href="#">HBASE-9952</a>	BUG-29905	[HBase] After restore snapshot, table data is getting corrupted (test_regionSplitSnapshots_splitNewRegion[table=...])
HDFS	<a href="#">HDFS-7503</a>	BUG-29091	Namenode restart after a large deletion could cause corruption
HDFS	<a href="#">HDFS-5306</a>	BUG-20054	Backport <a href="#">HDFS-5306</a> to 1.3-linux-maint.
HDFS	N/A, Security Fix	BUG-20051	Security Fix. Backport the file browser in branch-2.0
HDFS	<a href="#">HADOOP-10627</a>	BUG-14148	DataNode must support HTTPS on HDP 1.3.x
HDFS	N/A, Security Fix	BUG-20909	Demo of Security fixes.
HDFS	N/A, Security Fix	BUG-20092	Introduce a new option to disable HTTP server on DataNode
HDFS	N/A, Security Fix	BUG-20050	Make 1.3-linux-maint branch builds with Eclipse IDE
HDFS	<a href="#">HDFS-6822</a>	BUG-10706	Namenode and datanode fails to replace "_HOSTNAME" in hadoop.http.authentication.kerberos.principal
HDFS	N/A, Security Fix	BUG-20055	Support HTTPS redirection in webhdfs
HDFS	<a href="#">HADOOP-560</a>	BUG-7169	WebHDFS: openFileWithOffset returns premature EOF
HDFS	<a href="#">HDFS-6277</a>	BUG-17100	WebHdfsFileSystem#toUrl does not perform character encoding
Hive	<a href="#">HIVE-8485</a>	BUG-29405	Hive metastore NullPointerException with Oracle JDBC driver tblproperties/serdeproperties/etc, table not usable
Hive	<a href="#">HIVE-4975</a>	BUG-28335	Backport <a href="#">HIVE-4975</a> to HDP 1.3-*.-maint
Hive	<a href="#">HIVE-6679</a>	BUG-30089	HiveServer2 should support TCP Keepalive and Sendfile
Hive	<a href="#">HIVE-6468</a>	BUG-27968	HiveServer2 and Metastore using SASL out of memory
Hive	<a href="#">HIVE-4516</a>	BUG-19133	Timestamp field is not multi-thread safe
Hive	<a href="#">HIVE-5292</a>	BUG-30026	Hive 0.11 query gives wrong results when group by clause is used
Pig	<a href="#">PIG-4057</a> and <a href="#">PIG-4175</a>	BUG-23916	Group All followed by CROSS with default parallelism

## 1.3. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)

- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 1.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HDFS-7503](#): Namenode restart after large deletions can cause slow processReport (due to logging)
- [HDFS-6822](#): In branch-1, Namenode and datanode fails to replace "\_HOST" to hostname for `hadoop.http.authentication.kerberos.principal`
- [HDFS-6277](#): `WebHdfsFileSystem#toUrl` does not perform character escaping for rename
- [HDFS-5306](#): Datanode https port is not available at the namenode
- [HDFS-3794](#): `WebHDFS: openFileWithOffset` returns premature end of content-length delimited message body
- [HDFS-5306](#): Datanode https port is not available at the namenode
- [HDFS-6822](#): Namenode and datanode fails to replace "\_HOST" to hostname for `hadoop.http.authentication.kerberos.principal`
- [HDFS-6141](#): `WebHdfsFileSystem#toUrl` does not perform character escaping.
- [HDFS-5982](#): Need to update snapshot manager when applying editlog for deleting a snapshottable directory
- [HDFS-2264](#): `NamenodeProtocol` has the wrong value for `clientPrincipal` in `KerberosInfo` annotation
- [HDFS-3404](#): Make `putImage` in `GetImageServlet` infer remote address to fetch from request
- [HADOOP-10627](#): Add documentation for `http/https` policy configuration/setup
- [HADOOP-9509](#): Implemented `ONCRPC` and `XDR`.
- [HADOOP-9515](#): Added general interface for `NFS` and `Mount`.
- [HDFS-4762](#): Added HDFS based `NFSv3` and `Mountd` implementation.
- [HDFS-5038](#): Added the following HDFS branch-2 APIs to HDFS branch-1:
  - `FileSystem#newInstance(Configuration)`
  - `DFSClient#getNamenode()`
  - `FileStatus#isDirectory()`
- [HDFS-4880](#): Added support to print image and edits file loaded by the NameNode, in the logs.

- [HDFS-4944](#): Fixed file path issue with WebHDFS. WebHDFS can now create a file path containing characters that must be URI-encoded (such as space).
- **HDFS Snapshot related changes:**
  - [HDFS-4842](#): Added ability to identify correct prior snapshot before deleting a snapshot under a renamed subtree.
  - [HDFS-4857](#): Enhanced `Snapshot.Root` and `AbstractINodeDiff#snapshotINode`. (`Snapshot.Root` and `AbstractINodeDiff#snapshotINode` should not be put into `INodeMap` when loading `FSImage`.)
  - [HDFS-4863](#): The `root` directory can now be added to the `snapshottable` directory list while loading `fsimage`.
  - [HDFS-4846](#): Enhanced snapshot command line (CLI) commands output stacktrace for invalid arguments.
  - [HDFS-4848](#): Fixed `copyFromLocal` and file rename operations. (While performing `copyFromLocal` operation and/or renaming a file to `.snapshot`, now displays an output message that that `.snapshot` is a reserved name.)
  - [HDFS-4850](#): Fixed `OfflineImageViewer` to work on `fsimages` with empty files or snapshots.
  - [HDFS-4876](#): Fixed JavaDoc for `FileWithSnapshot`.
  - [HDFS-4877](#): Fixed the issues caused while renaming a directory under its prior descendant.
  - [HDFS-4902](#): Fixed path issue for `DFSClient.getSnapshotDiffReport`. `DFSClient.getSnapshotDiffReport` now uses `string` path instead of using the `o.a.h.fs.Path`.
  - [HDFS-4875](#): Added support for testing snapshot file length.
  - [HDFS-5005](#): Moved `SnapshotException` and `SnapshotAccessControlException` to `o.a.h.hdfs.protocol`.
- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-4366](#) Back port MAPREDUCE-4366 to Hadoop 1.2
- [MAPREDUCE-4661](#): Backport HTTPS to WebUIs to branch-1.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.

- [MAPREDUCE-5408](#): Backport [MAPREDUCE-336](#) to branch-1.
- [HDFS-4334](#): Added support to enable adding a unique id to each INode.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.
- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

### 1.3.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.6.1 and contains no additional patches.

### 1.3.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-9952](#): Snapshot restore may fail due to `NullPointerException`
- [HBASE-8148](#): Allow IPC to bind on a specific address.
- [HBASE-8816](#): Added support for loading multiple tables into `LoadTestTool`.
- [HBASE-6338](#): Cache method in RPC handler.
- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)

- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and addendum2+3).
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

### 1.3.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



#### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-5292](#): Join on decimal columns fails to return rows
- [HIVE-4516](#): Fix concurrency bug in `serde/src/java/org/apache/hadoop/hive/serde2/io/TimestampWritable.java`
- [HIVE-6468](#): HS2 & Metastore using SASL out of memory error when curl sends a get request

- [HIVE-6679](#): HiveServer2 should support configurable the server side socket timeout and keepalive for various transports types where applicable
- [HIVE-4975](#): Reading orc file throws exception after adding new column
- [HIVE-8485](#): HMS on Oracle incompatibility
- [HIVE-8042](#): Hive queries execute MoveTasks sequentially which slows down performance significantly in large queries.
- [HIVE-4516](#): Timestamp field is not multi thread safe.
- [HIVE-6716](#): ORC struct throws NPE for tables with inner structs having null values
- [HIVE-5002](#): Loosen readRowIndex visibility in ORC's RecordReaderImpl to package private
- [HIVE-5562](#): Stripe level stats for ORC Files. You can look at stripe stats without having to read the whole stripe.
- [HIVE-2084](#): Upgraded DataNucleus from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.
- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `multit` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.

- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 1.3.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 1.3.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-4175](#): PIG CROSS operation follow by STORE produces non-deterministic results each run
- [PIG-4057](#): Group All followed by CROSS with default parallelism produces wrong results
- [PIG-3236](#): Added support to parametrize snapshot and staging repository ID.
- [PIG-3048](#): Added MapReduce workflow information to job configuration.
- [PIG-3276](#): Changed default value (`/usr/local/hcat/bin/hcat`) for `hcat.bin` to `hcat`.
- [PIG-3277](#): Fixed path to the benchmarks file in the `print` statement.
- [PIG-3071](#): Updated HCatalog JAR file and path to HBase storage handler JAR in the Pig script file.
- [PIG-3262](#): Fixed compilation issues with Pig `contrib 0.11` on certain RPM systems.
- [PIG-2786](#): Enhanced Pig launcher script for HBase/Hcatalog integration.

### 1.3.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

### 1.3.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.3.2 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.



- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.
- [OOZIE-863](#): `JAVA_HOME` must be explicitly set at client because `bin/oozie` does not invoke `oozie-env.sh`.

### 1.3.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-1278](#): Enable Sqoop uncommitted reads in Sqoop for supported databases.
- [SQOOP-979](#): Fixed issues for MySQL direct connector, caused after moving password to credential cache.
- [SQOOP-914](#): Added an abort validation handler.
- [SQOOP-916](#): Enhanced security for passwords in Sqoop 1.x.
- [SQOOP-798](#): Fixed Ant docs failure for RHEL v5.8.

### 1.3.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

### 1.3.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- **JMS Source changes:**
  - [FLUME-924](#): Implemented JMS source for Flume NG.
  - [FLUME-1784](#): Fixed issues with documentation and parameter name.
  - [FLUME-1804](#): JMS source not included in binary distribution.
  - [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes.
  - [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.
  - [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).

- [FLUME-2043](#): JMS Source removed on failure to create configuration
- **Spillable Channel - (Experimental)**
  - [FLUME-1227](#): Introduce some sort of SpillableChannel.
- **Spillable Channel dependencies:**
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.
  - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
  - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
  - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- **Improvements:**
  - [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
  - [FLUME-1891](#): Fast replay runs even when checkpoint exists.
  - [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
  - [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
  - [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
  - [FLUME-1262](#): Move doc generation to a different profile.

## 1.4. Upgrading HDP 1.3.9 to 1.3.10

### Before You Begin

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

### Upgrading to 1.3.10

Use the following instructions to upgrade HDP 1.3.9 to HDP 1.3.10 manually:

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

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/1.x/updates/1.3.10.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos5/1.x/updates/1.3.10.0/hdp.repo</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/1.x/updates/1.3.10.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos6/1.x/updates/1.3.10.0/hdp.repo</a>
SLES 11 SP1	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/1.x/updates/1.3.10.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/suse11/1.x/updates/1.3.10.0/hdp.repo</a>

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

2. RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.10.0-centos5-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.10.0-centos5-tars-tarball.tar.gz</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.10.0-centos6-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.10.0-centos6-tars-tarball.tar.gz</a>
SLES 11 SP1	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.10.0-suse11-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.10.0-suse11-tars-tarball.tar.gz</a>

### 3. 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 have a manually installed cluster, use the instructions provided [here](#).



#### 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.

4. Run the `fsck` command as instructed below and fix any errors. (The resulting file will contain a complete block map of the file system.)

```
su $HDFS_USER
hadoop fsck / -files -blocks -locations > dfs-old-fsck-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

5. Use the following instructions to compare the status before and after the upgrade:



#### Note

The following commands must be executed by the user running the HDFS service (by default, the user is `hdfs`).

- a. Capture the complete namespace of the file system. Run the recursive listing of the root file system:

```
su $HDFS_USER
hadoop dfs -lsr / > dfs-old-lsr-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

- b. Run `report` command to create a list of DataNodes in the cluster.

```
su $HDFS_USER
hadoop dfsadmin -report > dfs-old-report-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

- c. Copy all or unrecoverable data stored in HDFS to a local file system or to a backup instance of HDFS.

d. Optionally, repeat the steps 3 (a) through 3 (c) and compare the results with the previous run to verify that the state of the file system remains unchanged.

6. As an HDFS user, execute the following command to save namespace:

```
su $HDFS_USER
hadoop dfsadmin -safemode enter
hadoop dfsadmin -saveNamespace
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

7. Copy the following checkpoint files into a backup directory:

- `dfs.name.dir/current/edits`
- `dfs.name.dir/image/fsimage`

8. **Stop the HDFS service.**

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

If you are installing manually, use the instructions provided [here](#).



### Note

Verify that all the HDP services in the cluster are stopped.

9. If you have Oozie installed:

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
```

Remove the old Oozie directories on all Oozie server and client hosts

```
rm -R /etc/oozie/conf/* oozie-conf-bak
```

10. Back up the Hive Database.

11. Upgrade the stack on all Agent hosts.

Operating System	Instructions	Commands
RHEL/CentOs/Oracle Linux		
	Upgrade the following components.	<code>yum upgrade "collectd*" "gccxml*" "hadoop*" "hive*" "hbase*" "mahout*" "oozie*" "pig*" "zookeeper*" "webhcat" "hdfs"</code>
	Verify that the components were upgraded.	<code>yum list installed   grep HDP-\$old-stack-version-number</code>
SLES		
	Upgrade the following components.	<code>zypper up "collectd*" "epel-release"</code> <code>zypper up -r HDP-1.3.10.0</code>
	Verify that the components were upgraded.	<code>rpm -qa   grep hadoop, rpm -qa   g</code>

12. Complete the Stack upgrade.

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 1.3.10 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-1.3 http://public-repo-1.hortonworks.com/HDP/
{${os}}/1.x/updates/1.3.10.0 {${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 sp1

### 13 Restart services.

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

If you have a manually installed cluster, use the [Starting HDP Services](#) instructions.



### Note

Remember to restart Hue as the root user: `/etc/init.d/hue restart`

## 1.5. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 1.5.1. Known Issues for Hadoop

There are no known issues for Hadoop in HDP 1.3.10.

### 1.5.2. Known Issues for Hive

- **BUG-10248:** java.lang.ClassCastException while running a join query

**Problem:** when a self join is done with 2 or more columns of different data types. For example: `join tab1.a = tab1.a join tab1.b=tab1.b` and a and b are different

data types. a is double and b is a string for e.g.. Now b cannot be cast into a double. It shouldn't have attempted to use the same serialization for both columns.

**Workaround:** Set the `hive.auto.convert.join.noconditionaltask.size` to a value such that the joins are split across multiple tasks.

- **BUG-5221:** Hive Windowing test Ordering\_1 fails

**Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **BUG-5220:** Hive Windowing test OverWithExpression\_3 fails

**Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN )?])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
```

```

at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target

```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **BUG-5512:** Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```

TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...

```

**Workaround:** The workaround is disable all the memory settings by setting value of the following properties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```

mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1

```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **BUG-4714:** Hive Server 2 Concurrency Failure (`create_index.q`).

**Problem:** While using indexes in Hive, the following error is reported:

```

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask

```

- **BUG-2131, HIVE-5297:** Partition in hive table that is of datatype 'int' is able to accept 'string' entries

**Problem:** Partition in hive table that is of datatype `int` is able to accept `string` entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype `int` can also accept `string` entries while data insertions.

**Workaround:** The workaround is to avoid adding `string` to `int` fields.

### 1.5.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report the correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\failureInfo\":"JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\","runState\":3
```

### 1.5.4. Known Issues for HBase

- **BUG-5723:** HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

### 1.5.5. Known Issues for Oozie

- **BUG-7933:** Oozie failed at smoke test in secured cluster

**Problem:** Oozie fails smoke tests in secured cluster.



**Workaround:**

1. Download the following files attached to, [AMBARI-2879](#):
  - `check_oozie_status.sh`
  - `oozieSmoke.sh`
2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.
3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **BUG-10265:** `TestBundleJobsFilter` test fails on RHEL 6.3, Oracle 6.3 and SUSE clusters with Postgres

**Problem:** `TestBundleJobsFilter` test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of `string integer` to an `integer`. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

## 1.5.6. Known Issues for Ambari

See [Ambari 1.6.1 Known Issues](#).

## 1.6. Security Fixes

This release includes the following security fix.

- [CVE-2014-3566](#)

## 2. Release Notes HDP-1.3.9

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [What's Changed in HDP 1.3.9](#)
- [Patch Information](#)
- [Upgrading HDP Manually](#)
- [Known Issues](#)

### 2.1. Product Version: HDP-1.3.9

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6
- Apache Pig 0.11.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.4.4 (or later)
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Hue 2.2.0
- Third party components:
  - Talend Open Studio for Big Data 5.3

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

## 2.2. What's Changed in this Release

This release contains the following bug fixes:

Project	Apache JIRA	Bug No.	Description
HDFS	<a href="#">HDFS-3794</a>	BUG-7169	WebHDFS: openFileWithOffset returns premature end of content-length delimited message body
HDFS	<a href="#">HDFS-5306</a>	BUG-20054	Datanode https port is not available at the namenode
HDFS	<a href="#">HDFS-6822</a>	BUG-10706	Namenode and datanode fails to replace "_HOST" to hostname for <code>hadoop.http.authentication.kerberos.principal</code>
Security	<a href="#">HDFS-5382</a>	BUG-20051	Backport the file browser in branch-2 to 1.3-linux-maint
Security	<a href="#">HDFS-5536</a>	BUG-20055	Support HTTPS redirection in webhdfs.
Security	<a href="#">HDFS-5536</a>	BUG-20092	Introduce a new option to disable HTTP server on datanode

## 2.3. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 2.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HDFS-3794](#) WebHDFS: `openFileWithOffset` returns premature end of content-length delimited message body
- [HDFS-5306](#) Datanode https port is not available at the namenode
- [HDFS-6822](#) Namenode and datanode fails to replace "\_HOST" to hostname for `hadoop.http.authentication.kerberos.principal`
- [HDFS-6141](#): `WebHdfsFileSystem#toUrl` does not perform character escaping.
- [HDFS-5982](#): Need to update snapshot manager when applying editlog for deleting a snapshotable directory
- [HDFS-2264](#): `NamenodeProtocol` has the wrong value for `clientPrincipal` in `KerberosInfo` annotation
- [HDFS-3404](#): Make `putImage` in `GetImageServlet` infer remote address to fetch from request
- [HADOOP-9509](#): Implemented `ONCRPC` and `XDR`.
- [HADOOP-9515](#): Added general interface for `NFS` and `Mount`.
- [HDFS-4762](#): Added HDFS based `NFSv3` and `Mountd` implementation.
- [HDFS-5038](#): Added the following HDFS branch-2 APIs to HDFS branch-1:
  - `FileSystem#newInstance(Configuration)`
  - `DFSClient#getNamenode()`
  - `FileStatus#isDirectory()`
- [HDFS-4880](#): Added support to print image and edits file loaded by the `NameNode`, in the logs.
- [HDFS-4944](#): Fixed file path issue with WebHDFS. WebHDFS can now create a file path containing characters that must be URI-encoded (such as space).
- 
- **HDFS Snapshot related changes:**
  - [HDFS-4842](#): Added ability to identify correct prior snapshot before deleting a snapshot under a renamed subtree.
  - [HDFS-4857](#): Enhanced `Snapshot.Root` and `AbstractINodeDiff#snapshotINode`. (`Snapshot.Root` and `AbstractINodeDiff#snapshotINode` should not be put into `INodeMap` when loading `FSImage`.)
  - [HDFS-4863](#): The `root` directory can now be added to the `snapshottable` directory list while loading `fsimage`.
  - [HDFS-4846](#): Enhanced snapshot command line (CLI) commands output stacktrace for invalid arguments.

- [HDFS-4848](#): Fixed `copyFromLocal` and file rename operations. (While performing `copyFromLocal` operation and/or renaming a file to `.snapshot`, now displays an output message that that `.snapshot` is a reserved name.)
- [HDFS-4850](#): Fixed `OfflineImageViewer` to work on `fsimages` with empty files or snapshots.
- [HDFS-4876](#): Fixed JavaDoc for `FileWithSnapshot`.
- [HDFS-4877](#): Fixed the issues caused while renaming a directory under its prior descendant.
- [HDFS-4902](#): Fixed path issue for `DFSClient.getSnapshotDiffReport`. `DFSClient.getSnapshotDiffReport` now uses string path instead of using the `o.a.h.fs.Path`.
- [HDFS-4875](#): Added support for testing snapshot file length.
- [HDFS-5005](#): Moved `SnapshotException` and `SnapshotAccessControlException` to `o.a.h.hdfs.protocol`.
- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-4366](#) Back port MAPREDUCE-4366 to Hadoop 1.2
- [MAPREDUCE-4661](#): Backport HTTPS to WebUIs to branch-1.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.
- [MAPREDUCE-5408](#): Backport [MAPREDUCE-336](#) to branch-1.
- [HDFS-4334](#): Added support to enable adding a unique id to each `Inode`.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.
- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.

- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

## 2.3.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.6.1 and contains no additional patches.

## 2.3.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-8148](#): Allow IPC to bind on a specific address.
- [HBASE-8816](#): Added support for loading multiple tables into `LoadTestTool`.
- [HBASE-6338](#): Cache method in RPC handler.
- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)

- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and `addendum2+3`).
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

## 2.3.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-8042](#) Hive queries execute `MoveTasks` sequentially which slows down performance significantly in large queries.
- [HIVE-4516](#) Timestamp field is not multi thread safe.
- [HIVE-6716](#): ORC struct throws NPE for tables with inner structs having null values
- [HIVE-5002](#): Loosen `readRowIndex` visibility in ORC's `RecordReaderImpl` to package private
- [HIVE-5562](#): Stripe level stats for ORC Files. You can look at stripe stats without having to read the whole stripe.
- [HIVE-2084](#): Upgraded `DataNuclues` from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store `Delegation Tokens` in database.
- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.

- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `multit` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 2.3.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 2.3.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3236](#): Added support to parametrize snapshot and staging repository ID.
- [PIG-3048](#): Added MapReduce workflow information to job configuration.
- [PIG-3276](#): Changed default value (`/usr/local/hcat/bin/hcat`) for `hcat.bin` to `hcat`.
- [PIG-3277](#): Fixed path to the benchmarks file in the `print` statement.
- [PIG-3071](#): Updated HCatalog JAR file and path to HBase storage handler JAR in the Pig script file.
- [PIG-3262](#): Fixed compilation issues with Pig `contrib 0.11` on certain RPM systems.



- [PIG-2786](#): Enhanced Pig launcher script for HBase/Hcatalog integration.

### 2.3.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to .m2 repository.

### 2.3.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.3.2 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.
- [OOZIE-863](#): `JAVA_HOME` must be explicitly set at client because `bin/oozie` does not invoke `oozie-env.sh`.

### 2.3.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-1278](#): Enable Sqoop uncommitted reads in Sqoop for supported databases.
- [SQOOP-979](#): Fixed issues for MySQL direct connector, caused after moving password to credential cache.
- [SQOOP-914](#): Added an abort validation handler.
- [SQOOP-916](#): Enhanced security for passwords in Sqoop 1.x.
- [SQOOP-798](#): Fixed Ant docs failure for RHEL v5.8.

### 2.3.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

## 2.3.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- **JMS Source changes:**
  - [FLUME-924](#): Implemented JMS source for Flume NG.
  - [FLUME-1784](#): Fixed issues with documentation and parameter name.
  - [FLUME-1804](#): JMS source not included in binary distribution.
  - [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes.
  - [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.
  - [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
  - [FLUME-2043](#): JMS Source removed on failure to create configuration
- **Spillable Channel - (Experimental)**
  - [FLUME-1227](#): Introduce some sort of `SpillableChannel`.
- **Spillable Channel dependencies:**
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.
  - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
  - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
  - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- **Improvements:**
  - [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
  - [FLUME-1891](#): Fast replay runs even when checkpoint exists.
  - [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
  - [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
  - [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
  - [FLUME-1262](#): Move doc generation to a different profile.

## 2.4. Upgrading HDP 1.3.7 to 1.3.9

### Before You Begin

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

### Upgrading to 1.3.9

Use the following instructions to upgrade HDP 1.3.7 to HDP 1.3.9 manually:

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

RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/1.x/updates/1.3.9.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos5/1.x/updates/1.3.9.0/hdp.repo</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/1.x/updates/1.3.9.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/centos6/1.x/updates/1.3.9.0/hdp.repo</a>
SLES 11 SP1	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/1.x/updates/1.3.9.0/hdp.repo">http://public-repo-1.hortonworks.com/HDP/suse11/1.x/updates/1.3.9.0/hdp.repo</a>

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

2. RHEL/CENTOS/ ORACLE LINUX 5	<a href="http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.9.0-centos5-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.9.0-centos5-tars-tarball.tar.gz</a>
RHEL/CENTOS/ ORACLE LINUX 6	<a href="http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.9.0-centos6-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.9.0-centos6-tars-tarball.tar.gz</a>
SLES 11 SP1	<a href="http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.9.0-suse11-tars-tarball.tar.gz">http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.9.0-suse11-tars-tarball.tar.gz</a>

#### 3. 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 have a manually installed cluster, use the [Stopping HDP Services](#) instructions.



### 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.

#### 4. Stop all services (including MapReduce) and client applications deployed on HDFS.

Use the instructions provided [here](#).

#### 5. Run the `fsck` command as instructed below and fix any errors. (The resulting file will contain a complete block map of the file system.)

```
su $HDFS_USER
hadoop fsck / -files -blocks -locations > dfs-old-fsck-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

#### 6. Use the following instructions to compare the status before and after the upgrade:



## Note

The following commands must be executed by the user running the HDFS service (by default, the user is `hdfs`).

- a. Capture the complete namespace of the file system. Run the recursive listing of the root file system:

```
su $HDFS_USER
hadoop dfs -lsr / > dfs-old-lsr-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

- b. Run report command to create a list of DataNodes in the cluster.

```
su $HDFS_USER
hadoop dfsadmin -report > dfs-old-report-1.log
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

- c. Copy all or unrecoverable data stored in HDFS to a local file system or to a backup instance of HDFS.
- d. Optionally, repeat the steps 3 (a) through 3 (c) and compare the results with the previous run to verify that the state of the file system remains unchanged.

7. As an HDFS user, execute the following command to save namespace:

```
su $HDFS_USER
hadoop dfsadmin -safemode enter
hadoop dfsadmin -saveNamespace
```

where `$HDFS_USER` is the HDFS Service user. For example, `hdfs`.

8. Copy the following checkpoint files into a backup directory:

- `dfs.name.dir/current/edits`
- `dfs.name.dir/image/fsimage`

9. **Stop the HDFS service.**

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

If you are installing manually, use the instructions provided [here](#).



## Note

Verify that all the HDP services in the cluster are stopped.

10. If you have Oozie installed:

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
```

Remove the old Oozie directories on all Oozie server and client hosts

```
rm -R /etc/oozie/conf/* oozie-conf-bak
```

11.Back up the Hive Database.

12.Upgrade the stack on all Agent hosts.

Operating System	Instructions	Commands
RHEL/CentOs/Oracle Linux	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgrade &lt;\$connectionUserName&gt; -passWor</pre> The value for \$databaseType can be de
	Upgrade the following components.	<pre>yum upgrade "collectd*" "gccxml*" "zookeeper*" "hbase*" "webhcat" "h</pre>
	Verify that the components were upgraded.	<pre>yum list installed   grep HDP-\$old-stack-version-number</pre>
SLES	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgrade &lt;\$connectionUserName&gt; -passWor</pre> The value for \$databaseType can be de
	Upgrade the following components.	<pre>zypper up "collectd*" "epel-releas "storm*" "sqoop*" "zookeeper*" "hb</pre>
		<pre>zypper up -r HDP-1.3.9.0</pre>
	Verify that the components were upgraded.	<pre>rpm -qa   grep hadoop, rpm -qa   g</pre>

13.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/bin/hadoop-daemon.sh start
journalnode"
```

14.Complete the Stack upgrade.

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 1.3.9 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-1.3 http://public-repo-1.hortonworks.com/HDP/
{$os}/1.x/updates/1.3.9.0 {$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
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11 sp1

15.Restart services.

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

If you have a manually installed cluster, use the [Starting HDP Services](#) instructions.

#### 16. 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 do NOT want to roll back), discard your backup and finalize the upgrade.

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

```
hdfs dfsadmin -finalizeUpgrade
```

For upgrading from HDP 1.2 to 1.3 manually, see [here](#).

For upgrading Ambari server, follow the instructions provided [here](#) and [here](#).

## 2.5. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 2.5.1. Known Issues for Hadoop

There are no known issues for Hadoop in HDP 1.3.9.

### 2.5.2. Known Issues for Hive

- **BUG-10248:** java.lang.ClassCastException while running a join query

**Problem:** when a self join is done with 2 or more columns of different data types. For example: `join tabl.a = tabl.a join tabl.b=tabl.b` and a and b are different data types. a is double and b is a string for e.g.. Now b cannot be cast into a double. It shouldn't have attempted to use the same serialization for both columns.

**Workaround:** Set the `hive.auto.convert.join.noconditionaltask.size` to a value such that the joins are split across multiple tasks.

- **BUG-5221:** Hive Windowing test Ordering\_1 fails

**Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **BUG-5220:**Hive Windowing test OverWithExpression\_3 fails

**Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN ) ?)])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
```

```
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target
```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **BUG-5512:** Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```
TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...
```

**Workaround:** The workaround is disable all the memory settings by setting value of the following properties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```
mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1
```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **BUG-4714:** Hive Server 2 Concurrency Failure (`create_index.q`).

**Problem:** While using indexes in Hive, the following error is reported:

```
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask
```

- **BUG-2131, HIVE-5297:** Partition in hive table that is of datatype 'int' is able to accept 'string' entries

**Problem:** Partition in hive table that is of datatype int is able to accept string entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype int can also accept string entries while data insertions.

**Workaround:** The workaround is to avoid adding string to int fields.



### 2.5.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report the correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\\"failureInfo\\":\\"JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\\",\\"runState\\":3
```

### 2.5.4. Known Issues for HBase

- **BUG-5723:** HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

### 2.5.5. Known Issues for Oozie

- **BUG-7933:** Oozie failed at smoke test in secured cluster

**Problem:** Oozie fails smoke tests in secured cluster.

**Workaround:**

1. Download the following files attached to, <https://issues.apache.org/jira/browse/AMBARI-2879>:
  - check\_oozie\_status.sh
  - oozieSmoke.sh
2. Replace /var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check\_oozie\_status.sh with the downloaded file.

3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **BUG-10265:** TestBundleJobsFilter test fails on RHEL 6.3, Oracle 6.3 and SUSE clusters with Postgres

**Problem:** TestBundleJobsFilter test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of `string integer` to an `integer`. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

## 2.5.6. Known Issues for Ambari

See [Ambari 1.6.1 Known Issues](#).

## 3. Release Notes HDP-1.3.7

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [What's Changed in HDP 1.3.7](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading HDP Manually](#)
- [Improvements](#)
- [Known Issues](#)

### 3.1. Product Version: HDP-1.3.7

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6
- Apache Pig 0.11.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.4.4 (or later)
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Hue 2.2.0

- Third party components:
  - Talend Open Studio for Big Data 5.3
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

## 3.2. What's Changed in this Release

In this section:

- [What's Changed in Hadoop](#)
- [What's Changed in Ambari](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in HCatalog](#)
- [What's Changed in Pig](#)
- [What's Changed in ZooKeeper](#)
- [What's Changed in Oozie](#)
- [What's Changed in Sqoop](#)
- [What's Changed in Mahout](#)
- [What's Changed in Flume](#)
- [What's Changed in Hue](#)

### 3.2.1. What's Changed in Hadoop

The following updates were made to HDFS for 1.3.7:

- **BUG-16960, HDFS-6141:** WebHdfsFileSystem#toUrl does not perform character escaping.
- **BUG-13793:**Port [HDFS-5982](#) to Hadoop 1.
- **BUG-13753, HDFS-2264, HDFS-3404:** ATT dual homing patched Hadoop Core jar for SNN issue

### 3.2.2. What's Changed in Ambari

HDP 1.3.7 works with [Ambari 1.4.4](#) or later.

### 3.2.3. What's Changed in HBase

The following updates were made to HBase for 1.3.7:

- **BUG-16796, HBASE-8148**: This is a small patch, necessary to allow HBase to work properly in multi-homed servers.

### 3.2.4. What's Changed in Hive

The following updates were made to Hive for 1.3.7:

- **BUG-16580, HIVE-6716**: RecordReader's nextKeyValue() is causing NPE when the value in the struct column is null in ORC table.
- **RMP-1387, HIVE-5002**: Changes to Record Reader Implementation for ORC Made changes to the method for exposing Index information outside of the class.
- **RMP-1367, HIVE-5562**: Stripe level stats for ORC Files. You can look at stripe stats without having to read the whole stripe.
- **RMP-1340, HIVE-5002**: Changes to Fetch Operator implementation:
  - Added a Java UUID to distinguish between distinct calls
  - The table properties are also now serialized into the Job Conf for use by the get Splits

### 3.2.5. What's Changed in HCatalog

The following updates were made to Hcatalog for 1.3.7:

- **BUG-12501, HIVE-6268**: Network resource leak: Hcatalog M/R interface HCatInputFormat

### 3.2.6. What's Changed in Pig

No updates were made to Pig for 1.3.7.

### 3.2.7. What's Changed in ZooKeeper

No updates were made to ZooKeeper for 1.3.7.

### 3.2.8. What's Changed in Oozie

No updates were made to Oozie for 1.3.7.

### 3.2.9. What's Changed in Sqoop

The following updates were made to Sqoop for 1.3.7:

- **BUG-13132, SQOOP-1278**: Enable Sqoop uncommitted reads in Sqoop for supported databases.

## 3.2.10. What's Changed in Mahout

No changes were made to Mahout for 1.3.7.

## 3.2.11. What's Changed in Flume

No updates were made to Flume for 1.3.7.

## 3.2.12. What's Changed in Hue

No updates were made to Hue for 1.3.7.

## 3.3. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 3.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- **HDFS-6141:** WebHdfsFileSystem#toUrl does not perform character escaping.
- **HDFS-5982:** Need to update snapshot manager when applying editlog for deleting a snapshottable directory
- **HDFS-2264:** NamenodeProtocol has the wrong value for clientPrincipal in KerberosInfo annotation
- **HDFS-3404:** Make putImage in GetImageServlet infer remote address to fetch from request

- [HADOOP-9509](#): Implemented ONCRPC and XDR.
- [HADOOP-9515](#): Added general interface for NFS and Mount.
- [HDFS-4762](#): Added HDFS based NFSv3 and Mountd implementation.
- [HDFS-5038](#): Added the following HDFS branch-2 APIs to HDFS branch-1:
  - `FileSystem#newInstance(Configuration)`
  - `DFSCClient#getNamenode()`
  - `FileStatus#isDirectory()`
- [HDFS-4880](#): Added support to print image and edits file loaded by the NameNode, in the logs.
- [HDFS-4944](#): Fixed file path issue with WebHDFS. WebHDFS can now create a file path containing characters that must be URI-encoded (such as space).
- 
- **HDFS Snapshot related changes:**
  - [HDFS-4842](#): Added ability to identify correct prior snapshot before deleting a snapshot under a renamed subtree.
  - [HDFS-4857](#): Enhanced `Snapshot.Root` and `AbstractINodeDiff#snapshotINode`. (`Snapshot.Root` and `AbstractINodeDiff#snapshotINode` should not be put into `INodeMap` when loading `FSImage`.)
  - [HDFS-4863](#): The `root` directory can now be added to the `snapshottable` directory list while loading `fsimage`.
  - [HDFS-4846](#): Enhanced snapshot command line (CLI) commands output stacktrace for invalid arguments.
  - [HDFS-4848](#): Fixed `copyFromLocal` and file rename operations. (While performing `copyFromLocal` operation and/or renaming a file to `.snapshot`, now displays an output message that that `.snapshot` is a reserved name.)
  - [HDFS-4850](#): Fixed `OfflineImageViewer` to work on `fsimages` with empty files or snapshots.
  - [HDFS-4876](#): Fixed JavaDoc for `FileWithSnapshot`.
  - [HDFS-4877](#): Fixed the issues caused while renaming a directory under its prior descendant.
  - [HDFS-4902](#): Fixed path issue for `DFSCClient.getSnapshotDiffReport`. `DFSCClient.getSnapshotDiffReport` now uses `string` path instead of using the `o.a.h.fs.Path`.
  - [HDFS-4875](#): Added support for testing snapshot file length.

- [HDFS-5005](#): Moved `SnapshotException` and `SnapshotAccessControlException` to `o.a.h.hdfs.protocol`.
- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-4661](#): Backport HTTPS to WebUIs to branch-1.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.
- [MAPREDUCE-5408](#): Backport [MAPREDUCE-336](#) to branch-1.
- [HDFS-4334](#): Added support to enable adding a unique id to each `Inode`.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.
- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

### 3.3.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.4.4 and contains no additional patches.

### 3.3.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-8148](#): Allow IPC to bind on a specific address.
- [HBASE-8816](#): Added support for loading multiple tables into `LoadTestTool`.
- [HBASE-6338](#): Cache method in RPC handler.



- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and addendum2+3).

- **HBASE-7410**: [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- **HBASE-8530**: Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

### 3.3.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



#### Note

Apache HCatalog is now merged with Apache Hive.

- **HIVE-6716**: ORC struct throws NPE for tables with inner structs having null values
- **HIVE-5002**: Loosen `readRowIndex` visibility in ORC's `RecordReaderImpl` to package private
- **HIVE-5562**: Stripe level stats for ORC Files. You can look at stripe stats without having to read the whole stripe.
- **HIVE-2084**: Upgraded DataNucleus from v2.0.3 to v3.0.1.
- **HIVE-3815**: Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- **HIVE-3846**: Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- **HIVE-3255**: Added `DBTokenStore` to store Delegation Tokens in database.
- **HIVE-4171**: Current database in metastore. Hive is not consistent with `SessionState`.
- **HIVE-4392**: Fixed `Illogical InvalidObjectException` when using `multit` aggregate functions with star columns.
- **HIVE-4343**: Fixed HiveServer2 with Kerberos - local task for map join fails.
- **HIVE-4485**: Fixed beeline prints null as empty strings.
- **HIVE-4510**: Fixed HiveServer2 nested exceptions.
- **HIVE-4513**: Added support to disable Hive history logs by default.
- **HIVE-4521**: Fixed auto join conversion failures
- **HIVE-4540**: Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- **HIVE-4611**: Fixed SMB join failures because of conflicts in bigtable selection policy.
- **HIVE-5542**: Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- **HIVE-3255**: Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.

- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for `WebHCatalog` end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for `Hive HBaseStorageHandler` to work with `HCatalog`.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 3.3.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 3.3.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3236](#): Added support to parametrize snapshot and staging repository ID.
- [PIG-3048](#): Added MapReduce workflow information to job configuration.
- [PIG-3276](#): Changed default value (`/usr/local/hcat/bin/hcat`) for `hcat.bin` to `hcat`.
- [PIG-3277](#): Fixed path to the benchmarks file in the `print` statement.
- [PIG-3071](#): Updated HCatalog JAR file and path to HBase storage handler JAR in the Pig script file.
- [PIG-3262](#): Fixed compilation issues with `Pig contrib 0.11` on certain RPM systems.
- [PIG-2786](#): Enhanced Pig launcher script for HBase/Hcatalog integration.

### 3.3.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

### 3.3.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.3.2 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.

- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.
- [OOZIE-863](#): `JAVA_HOME` must be explicitly set at client because `bin/oozie` does not invoke `oozie-env.sh`.

### 3.3.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-1278](#): Enable Sqoop uncommitted reads in Sqoop for supported databases.
- [SQOOP-979](#): Fixed issues for MySQL direct connector, caused after moving password to credential cache.
- [SQOOP-914](#): Added an abort validation handler.
- [SQOOP-916](#): Enhanced security for passwords in Sqoop 1.x.
- [SQOOP-798](#): Fixed Ant docs failure for RHEL v5.8.

### 3.3.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

### 3.3.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- **JMS Source changes:**
  - [FLUME-924](#): Implemented JMS source for Flume NG.
  - [FLUME-1784](#): Fixed issues with documentation and parameter name.
  - [FLUME-1804](#): JMS source not included in binary distribution.
  - [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes.
  - [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.

- [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
- [FLUME-2043](#): JMS Source removed on failure to create configuration
- **Spillable Channel - (Experimental)**
  - [FLUME-1227](#): Introduce some sort of SpillableChannel.
- **Spillable Channel dependencies:**
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.
  - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
  - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
  - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- **Improvements:**
  - [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
  - [FLUME-1891](#): Fast replay runs even when checkpoint exists.
  - [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
  - [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
  - [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
  - [FLUME-1262](#): Move doc generation to a different profile.

## 3.4. Minimum System Requirements

In this section:

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



### Note

gsInstaller was **deprecated** as of HDP 1.2.0 and is no longer being made available in 1.3.0 or in future releases.

We encourage you to consider [Manual Install \(RPMs\)](#) or [Automated Install \(Ambari\)](#).

## 3.4.1. Hardware Recommendations

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

## 3.4.2. Operating Systems Requirements

The following operating systems (OS) are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- Oracle Linux 5 and 6

## 3.4.3. Software Requirements

On each of your hosts:

- yum (RHEL/CentOS)
- zypper (SLES)



### Note

Ensure that the Zypper version is 1.3.14.

- rpm
- scp
- curl
- wget
- pdsh

## 3.4.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and by default uses embedded Derby database. MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are supported.

You may provide access to an existing database, or you can use Ambari installer to deploy MySQL instance for your environment. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Oozie requires a database to use as a metadata store and by default uses embedded Derby database.

MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Ambari requires a database to store information about cluster topology and configuration.

The default database is Postgres 8.x and Oracle 11gr2 is also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

### 3.4.5. 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.

### 3.4.6. Optional: Configure 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. For example, if you are using RHEL 6 your hosts must be able to access the “Red Hat Enterprise Linux Server 6 Optional (RPMs)” repository. If this repository is disabled, the installation is unable to access the `rubygems` package. If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 3.5. Upgrading HDP Manually

Use the following instructions to upgrade HDP manually:

For upgrading manually, see [here](#).

For upgrading Ambari server, follow the instructions provided [here](#) and [here](#).

## 3.6. Improvements

The following improvements were made in HDP 1.3.7:

- **BUG-14150, HADOOP-10379:** HTTP Cookies are now marked HttpOnly and Secure.
- **BUG-14153, HDFS-6084:** Delegation tokens not be exposed via HTTP Referer header.

## 3.7. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 3.7.1. Known Issues for Hadoop

There are no known issues for Hadoop in HDP 1.3.7.

### 3.7.2. Known Issues for Hive

- **BUG-17088:** HiveServer2Concurr test fails intermittently throwing `org.apache.thrift.transport.TTransportException: Peer indicated failure: GSS initiate failed`

**Problem:** When we open two different connections at the same time, it results in a replay error from the GSS-API. From kerberos v5 RFC (<http://www.ietf.org/rfc/rfc4120.txt>): Implementation note: If a client generates multiple requests to the KDC with the same timestamp, including the microsecond field, all but the first of the requests received will be rejected as replays. This might happen, for example, if the resolution of the client's clock is too coarse. Client implementations SHOULD ensure that the timestamps are not reused, possibly by incrementing the microseconds field in the time stamp when the clock returns the same time for multiple requests.

**Workaround:** Add some delay on the client side when issuing multiple concurrent request. One way to achieve this is to synchronize the connection creation on the client side, which will introduce the required delay.

- **BUG-10248:** `java.lang.ClassCastException` while running a join query

**Problem:** when a self join is done with 2 or more columns of different data types. For example: `join tabl.a = tabl.a join tabl.b=tabl.b` and a and b are different



data types. a is double and b is a string for e.g.. Now b cannot be cast into a double. It shouldn't have attempted to use the same serialization for both columns.

**Workaround:** Set the `hive.auto.convert.join.noconditionaltask.size` to a value such that the joins are split across multiple tasks.

- **BUG-5221:** Hive Windowing test Ordering\_1 fails

**Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **BUG-5220:** Hive Windowing test OverWithExpression\_3 fails

**Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN )?])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
```

```

at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target

```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **BUG-5512:** Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```

TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...

```

**Workaround:** The workaround is disable all the memory settings by setting value of the following properties to -1 in the mapred-site.xml file on the JobTracker and TaskTracker host machines in your cluster:

```

mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1

```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **BUG-4714:** Hive Server 2 Concurrency Failure (create\_index.q).

**Problem:** While using indexes in Hive, the following error is reported:

```

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask

```

- **BUG-2131, HIVE-5297:** Partition in hive table that is of datatype 'int' is able to accept 'string' entries

**Problem:** Partition in hive table that is of datatype `int` is able to accept `string` entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype `int` can also accept `string` entries while data insertions.

**Workaround:** The workaround is to avoid adding `string` to `int` fields.

### 3.7.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report the correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\failureInfo\":"JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\", \"runState\":3
```

### 3.7.4. Known Issues for HBase

- **BUG-5723:** HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

### 3.7.5. Known Issues for Oozie

- **Problem:** Oozie fails smoke tests in secured cluster.

**Workaround:**

1. Download the following files attached to, <https://issues.apache.org/jira/browse/AMBARI-2879>:
  - check\_oozie\_status.sh
  - oozieSmoke.sh
2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.
3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **BUG-10265:** TestBundleJobsFilter test fails on RHEL 6.3, Oracle 6.3 and SUSE clusters with Postgres

**Problem:** TestBundleJobsFilter test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of `string integer` to an `integer`. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

### 3.7.6. Known Issues for Ambari

See [Ambari 1.4.4 Known Issues](#).

## 4. Release Notes HDP-1.3.3

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [What's Changed in HDP 1.3.3](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading HDP Manually](#)
- [Improvements](#)
- [Known Issues](#)

### 4.1. Product Version: HDP-1.3.3

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6.1
- Apache Pig 0.11.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.4.1
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Hue 2.2.0
- Third party components:

- Talend Open Studio for Big Data 5.3
- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

## 4.2. What's Changed in this Release

In this section:

- [What's Changed in Hadoop](#)
- [What's Changed in Ambari](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in HCatalog](#)
- [What's Changed in Pig](#)
- [What's Changed in ZooKeeper](#)
- [What's Changed in Oozie](#)
- [What's Changed in Sqoop](#)
- [What's Changed in Mahout](#)
- [What's Changed in Flume](#)
- [What's Changed in Hue](#)

### 4.2.1. What's Changed in Hadoop

The following updates were made to Hadoop Core for 1.3.3:

- **BUG-9530, MAPREDUCE-5508**:HDP1.3.2/HA mapreduce stops working after kerberos enabled
- **BUG-8838, HAMONITOR-8838**:HA monitor fails to provide kerberos credentials, mistaking services as down and thus intermittently tries to restart namenode/jobtracker services

The following updates were made to MapReduce for 1.3.3:

- **BUG-10178, MAPREDUCE-1238** : negative finishedMapTasks counter hangs JobTracker
- **BUG-8838, HAMONITOR-8838**:HA monitor fails to provide kerberos credentials, mistaking services as down and thus intermittently tries to restart namenode/jobtracker services

- **BUG-8384, MAPREDUCE-5490**:MapReduce doesn't pass the classpath to child processes
- **BUG-7991,MAPREDUCE-5508**:JT ui/jmx/rpc calls hang some times in secure mode

The following updates were made to HDFS for 1.3.3:

- **BUG-9225, HDFS-5245** : Distcp from hdp1(webhdfs/hftp) to hdp2(webhdfs) throws a 'File does not exist' error (HDFS-5245)
- **BUG-8838, HAMONITOR-8838**: HA monitor fails to provide kerberos credentials, mistaking services as down and thus intermittently tries to restart namenode/jobtracker services
- **BUG-6041, HDFS-4794** Browsing filesystem via webui throws kerberos exception when NN service RPC is enabled in a secure cluster

## 4.2.2. What's Changed in Ambari

The following updates were made to Ambari for 1.3.3:

- **BUG-10727, AMBARI-3752**: MR jobs are hanging on a 2-node cluster with default configuration
- **BUG-10660, AMBARI-3743**:YARN dynamic configs generate 0 values nodes
- **BUG-10634, AMBARI-3719** : YARN default number of containers calculation skewed towards heavier resources
- **BUG-10584, AMBARI-3708** : Reconfigure of dynamic configs not showing modified values
- **BUG-10573, AMBARI-3707**: Smoke tests are broken on trunk for mr2, pig, oozie due to invalid JVM config
- **BUG-10565, AMBARI-3722**: Dynamic configs code needs testcases
- **BUG-10557, AMBARI-3697**: JS exception trying to calculate dynamic configs in branch-1.4
- **BUG-10495, AMBARI-3687**: Deploying on EC2 with hosts that have 7.3 GB ram, default MR2 task mem defaults are bad
- **BUG-10481, AMBARI-3675**: Default value of 'Default virtual memory for a job's map-task' is not valid
- **BUG-10324, AMBARI-3647**: Fix the version numbers for all stack components for Stack 1.3.3
- **BUG-10169, AMBARI-3579**: Add new stack definition for 1.3.3 to the stack (along with 1.3.2).

## 4.2.3. What's Changed in HBase

The following updates were made to HBase for 1.3.3:

- No updates were made to HBase for 1.3.3.

## 4.2.4. What's Changed in Hive

The following updates were made to Hive for 1.3.3:

- **BUG-10216, HIVE-5989**: Hive metastore authorization check can check using wrong user, resulting hive job failures.
- **BUG-10170, HIVE-5256**: ArrayIndexOutOfBoundsException Exception while inserting data into hive table
- **BUG-9889, AMBARI-5223**: hive-env.sh overwrites user value of HIVE\_AUX\_JARS\_PATH
- **BUG-9364, HIVE-5122**: Backport HIVE-5122 Add partition for multiple partition ignores locations for non-first partitions
- **BUG-9363, HIVE-4689 and HIVE-4781**: Backport HIVE-4781 / HIVE-4689 LEFT SEMI JOIN generates wrong results when the number of rows belonging to a single key of the right table exceed hive.join.emit.interval
- **BUG-9362, HIVE-4845**: Backport HIVE-4845 Correctness issue with MapJoins using the null safe operator
- **BUG-9361, HIVE-4789**: Backport HIVE-4789 FetchOperator fails on partitioned Avro data
- **BUG-9360, HIVE-3953**: Backport HIVE-3953 Reading of partitioned Avro data fails because of missing properties
- **BUG-9123, HIVE-3807**: Authorization does not work in Kerberos Hive, authorization uses short name that is different than kerberos principal.
- **BUG-8944, HIVE-4547**: Hive scripts used in version 0.9 with the CAST function within a view statement no longer execute successfully in hive 0.11 if columns are not in single quotes.

## 4.2.5. What's Changed in Hcatalog

The following updates were made to Hcatalog for 1.3.3:

- **BUG-9438, HIVE-5636**: InputJobInfo.getTableInfo() is returning NULL - in HDP 1.3.2

## 4.2.6. What's Changed in Pig

No updates were made to Pig for 1.3.3.

## 4.2.7. What's Changed in ZooKeeper

No updates were made to ZooKeeper for 1.3.3.

## 4.2.8. What's Changed in Oozie

No updates were made to Oozie for 1.3.3.



## 4.2.9. What's Changed in Sqoop

The following updates were made to Sqoop for 1.3.3:

- **BUG-9785**, **SQOOP-931**: Sqoop can not load hcatalog table to named database2
- Sqoop Connector for Teradata connector 1.0.9a

## 4.2.10. What's Changed in Mahout

No updates were made to Mahout for 1.3.3.

## 4.2.11. What's Changed in Flume

No updates were made to Flume for 1.3.3.

## 4.2.12. What's Changed in Hue

The following updates were made to Hue for 1.3.3:

- **BUG-10444**: Removed limitation of only showing the last 10 jobs
- **BUG-9326**: Resolved issue when copying a file or folder in file browser
- **BUG-9246**: Improved performance of page loading
- **BUG-9040**: Changed behavior of job browser to accurately show the current state of a job after it has been killed outside of Hue
- **BUG-8954**: Resolved issue when deleting trash
- **BUG-8769**: Resolved ability to kill Oozie workflows that are submitted by a user other than hue
- **BUG-8413**: Added support for providing more than one argument to Pig scripts
- **BUG-8106**: Resolved disappearing "Kill Job" option in Pig editor following re-authentication

## 4.3. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)

- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 4.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HADOOP-9509](#): Implemented `ONCRPC` and `XDR`.
- [HADOOP-9515](#): Added general interface for NFS and Mount.
- [HDFS-4762](#): Added HDFS based `NFSv3` and `Mountd` implementation.
- [HDFS-5038](#): Added the following HDFS branch-2 APIs to HDFS branch-1:
  - `FileSystem#newInstance(Configuration)`
  - `DFSClient#getNamenode()`
  - `FileStatus#isDirectory()`
- [HDFS-4880](#): Added support to print image and edits file loaded by the NameNode, in the logs.
- [HDFS-4944](#): Fixed file path issue with WebHDFS. WebHDFS can now create a file path containing characters that must be URI-encoded (such as space).
- **HDFS Snapshot related changes:**
  - [HDFS-4842](#): Added ability to identify correct prior snapshot before deleting a snapshot under a renamed subtree.
  - [HDFS-4857](#): Enhanced `Snapshot.Root` and `AbstractINodeDiff#snapshotINode`. (`Snapshot.Root` and `AbstractINodeDiff#snapshotINode` should not be put into `INodeMap` when loading `FSImage`.)
  - [HDFS-4863](#): The `root` directory can now be added to the `snapshottable` directory list while loading `fsimage`.
  - [HDFS-4846](#): Enhanced snapshot command line (CLI) commands output stacktrace for invalid arguments.
  - [HDFS-4848](#): Fixed `copyFromLocal` and file rename operations. (While performing `copyFromLocal` operation and/or renaming a file to `.snapshot`, now displays an output message that that `.snapshot` is a reserved name.)

- [HDFS-4850](#): Fixed `OfflineImageViewer` to work on `fsimages` with empty files or snapshots.
- [HDFS-4876](#): Fixed JavaDoc for `FileWithSnapshot`.
- [HDFS-4877](#): Fixed the issues caused while renaming a directory under its prior descendant.
- [HDFS-4902](#): Fixed path issue for `DFSCClient.getSnapshotDiffReport`. `DFSCClient.getSnapshotDiffReport` now uses string path instead of using the `o.a.h.fs.Path`.
- [HDFS-4875](#): Added support for testing snapshot file length.
- [HDFS-5005](#): Moved `SnapshotException` and `SnapshotAccessControlException` to `o.a.h.hdfs.protocol`.
- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-4661](#): Backport HTTPS to WebUIs to branch-1.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.
- [MAPREDUCE-5408](#): Backport [MAPREDUCE-336](#) to branch-1.
- [HDFS-4334](#): Added support to enable adding a unique id to each `INode`.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.
- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.
- NFS Updates:
  - [BUG-5966](#): fix wrong XDR method names.

- [BUG-5750](#): Input/output error after restarting Namenode while uploading data.
- [BUG-5934](#): RuntimeException causes data loading failure.
- Changed some trace level and add lock/unlock trace.
- [BUG-6111](#): use final clause to guarantee OpenFileCtx lock is unlocked eventually.
- Fixed readdir/readdirplus response to copy dirents to response
- [BUG-6609](#): Reduce lock granularity in OpenFileCtx
- [BUG-6352](#): Should resend response for some repeated write request
- Other updates:
  - [BUG-7668](#): Fix backwards-incompatible change for ClientProtocol#fsync.
  - [BUG-10178](#): Fix negative value for JobInProgress.finishedMapTasks.

### 4.3.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.4.1 and includes the following:

- [AMBARI-3752](#): MR jobs are hanging on a 2-node cluster with default configuration
- [AMBARI-3743](#): YARN dynamic configs generate 0 values nodes
- [AMBARI-3722](#): Dynamic configs code needs testcases
- [AMBARI-3719](#): YARN default number of containers calculation skewed towards heavier resources
- [AMBARI-3708](#): Reconfigure of dynamic configs not showing modified values
- [AMBARI-3707](#): YARN dynamic configs generate 0 map/reduce memory on 2GB machines
- [AMBARI-3697](#): JS exception trying to calculate dynamic configs in branch-1.4
- [AMBARI-3678](#): Change Oozie default log level to INFO
- [AMBARI-3675](#): Default value of 'Default virtual memory for a job's map-task' is not valid
- [AMBARI-3647](#): Fix the version numbers for all stack components for Stack.
- [AMBARI-3579](#): Adding new stack based on hadoop 1.0.

### 4.3.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-8816](#): Added support for loading multiple tables into LoadTestTool.
- [HBASE-6338](#): Cache method in RPC handler.

- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and addendum2+3).
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).

- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

### 4.3.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



#### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-2084](#): Upgraded DataNucleus from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.
- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `mult` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 4.3.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 4.3.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3236](#): Added support to parametrize snapshot and staging repository ID.
- [PIG-3048](#): Added MapReduce workflow information to job configuration.
- [PIG-3276](#): Changed default value (`/usr/local/hcat/bin/hcat`) for `hcat.bin` to `hcat`.
- [PIG-3277](#): Fixed path to the benchmarks file in the `print` statement.
- [PIG-3071](#): Updated HCatalog JAR file and path to HBase storage handler JAR in the Pig script file.
- [PIG-3262](#): Fixed compilation issues with Pig `contrib 0.11` on certain RPM systems.
- [PIG-2786](#): Enhanced Pig launcher script for HBase/Hcatalog integration.

### 4.3.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

### 4.3.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.3.2 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.
- [OOZIE-863](#): `JAVA_HOME` must be explicitly set at client because `bin/oozie` does not invoke `oozie-env.sh`.

### 4.3.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-979](#): Fixed issues for MySQL direct connector, caused after moving password to credential cache.
- [SQOOP-914](#): Added an abort validation handler.
- [SQOOP-916](#): Enhanced security for passwords in Sqoop 1.x.
- [SQOOP-798](#): Fixed Ant docs failure for RHEL v5.8.

### 4.3.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed NullPointerException in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

### 4.3.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- **JMS Source changes:**
  - [FLUME-924](#): Implemented JMS source for Flume NG.
  - [FLUME-1784](#): Fixed issues with documentation and parameter name.
  - [FLUME-1804](#): JMS source not included in binary distribution.
  - [FLUME-1777](#): `AbstractSource` does not provide enough implementation for subclasses.
  - [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.
  - [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
  - [FLUME-2043](#): JMS Source removed on failure to create configuration
- **Spillable Channel - (Experimental)**
  - [FLUME-1227](#): Introduce some sort of `SpillableChannel`.
- **Spillable Channel dependencies:**
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.



- [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
- [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
- [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- **Improvements:**
  - [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
  - [FLUME-1891](#): Fast replay runs even when checkpoint exists.
  - [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
  - [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
  - [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
  - [FLUME-1262](#): Move doc generation to a different profile.

## 4.4. Minimum System Requirements

In this section:

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



### Note

`gsInstaller` was **deprecated** as of HDP 1.2.0 and is no longer being made available in 1.3.0 or in future releases.

We encourage you to consider [Manual Install \(RPMs\)](#) or [Automated Install \(Ambari\)](#).

### 4.4.1. Hardware Recommendations

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

## 4.4.2. Operating Systems Requirements

The following operating systems (OS) are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- Oracle Linux 5 and 6

## 4.4.3. Software Requirements

On each of your hosts:

- yum (RHEL/CentOS)
- zypper (SLES)



### Note

Ensure that the Zypper version is 1.3.14.

- rpm
- scp
- curl
- wget
- pdsh

## 4.4.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and by default uses embedded Derby database. MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are supported. You may provide access to an existing database, or you can use Ambari installer to deploy MySQL instance for your environment. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).
- Oozie requires a database to use as a metadata store and by default uses embedded Derby database.

MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Ambari requires a database to store information about cluster topology and configuration.

The default database is Postgres 8.x and Oracle 11gr2 is also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

## 4.4.5. 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.

## 4.4.6. Optional: Configure 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. For example, if you are using RHEL 6 your hosts must be able to access the “Red Hat Enterprise Linux Server 6 Optional (RPMs)” repository. If this repository is disabled, the installation is unable to access the `rubygems` package. If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 4.5. Upgrading HDP Manually

Use the following instructions to upgrade HDP manually:

For upgrading manually, see [here](#).

For upgrading Ambari server, follow the instructions provided [here](#) and [here](#).

## 4.6. Improvements

- Backported several Hadoop 2.0 APIs to Comanche.
- **BUG-5284**: Fixed delegation Token renewal exception in jobtracker logs.
- **BUG-5483**: NFS: file upload fails to upload in NFS-MountDir fixed.
- **BUG-5774**: Fixed taking snapshot on Oracle Linux and Java version 1.6.0\_31, may occasionally result in spurious `Timeout` error.
- **HDFS-4880**: Print the image and edits file loaded by the namenode in the logs.

## 4.7. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 4.7.1. Known Issues for Hadoop

- JobTracker UI, JMX queries, and RPC calls sometimes hang in HA mode.

**Problem:** JobTracker becomes slow and non-responsive in HA mode because `dfs.client.retry.policy.enabled` is not set to 'final' and 'false' for the JobTracker.

**Workaround:** Set `dfs.client.retry.policy.enabled` to 'final' and 'false' *only* for the JobTracker. Clients (such as MapReduce, Pig, Hive, Oozie) should still be set to 'true' in HA mode.

- Use of `init.d` scripts for starting or stopping Hadoop services, is not recommended.

### 4.7.2. Known Issues for Hive

- **BUG-10248:** `java.lang.ClassCastException` while running a join query

**Problem:** when a self join is done with 2 or more columns of different data types. For example: `join tabl.a = tabl.a join tabl.b=tabl.b` and a and b are different data types. a is double and b is a string for e.g.. Now b cannot be cast into a double. It shouldn't have attempted to use the same serialization for both columns.

**Workaround:** Set the `hive.auto.convert.join.noconditionaltask.size` to a value such that the joins are split across multiple tasks.

- **BUG-5512:** Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```
TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...
```

**Workaround:** The workaround is disable all the memory settings by setting value of the following properties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```
mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1
```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **BUG-5221:**Hive Windowing test Ordering\_1 fails

**Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **BUG-5220:**Hive Windowing test OverWithExpression\_3 fails

**Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN ) ?)])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
```

```

at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target

```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **Problem:** While using indexes in Hive, the following error is reported:

```
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask
```

- **Problem:** Partition in hive table that is of datatype int is able to accept string entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype int can also accept string entries while data insertions.

**Workaround:** The workaround is to avoid adding string to int fields.

### 4.7.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```

FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1

```

- **Problem:** Failure to report correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\failureInfo\":"\JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\","runState\":3
```

## 4.7.4. Known Issues for HBase

- HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSCliet: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

## 4.7.5. Known Issues for Oozie

- **Problem:** Oozie fails smoke tests in secured cluster.

**Workaround:**

1. Download the following files attached to, <https://issues.apache.org/jira/browse/AMBARI-2879>:

- check\_oozie\_status.sh
- oozieSmoke.sh

2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.

3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.

5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/
oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`

- `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **BUG-10265:** `TestBundleJobsFilter` test fails on RHEL 6.3, Oracle 6.3 and SUSE clusters with Postgres

**Problem:** `TestBundleJobsFilter` test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of `string integer` to an `integer`. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

## 4.7.6. Known Issues for Ambari

- **Problem:** Oozie fails smoke tests in secured cluster.

### Workaround:

1. Download the following files attached to <https://issues.apache.org/jira/browse/AMBARI-2879>:
  - `check_oozie_status.sh`
  - `oozieSmoke.sh`
2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.
3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **Problem:** The `ambari-server` command displays invalid options for setting up Ganglia and Nagios.

On the Ambari server host machine, when you execute the following command:

```
ambari-server
```



You see the following output:

```
Using python /usr/bin/python2.6
Usage: /usr/sbin/ambari-server {start|stop|restart|setup|upgrade|status|
upgradestack|setup-ldap|setup-https|setup-ganglia_https|setup-nagios_https|
encrypt-passwords} [options]
```

**Workaround:** The `setup-ganglia_https` and `setup-nagios_https` are not valid options.

Use `setup-ganglia-https` and `setup-nagios-https` with the `ambari-server` command to set up Ganglia and Nagios.

- **Problem:** `ntpd` service warning might be displayed as part of host check at the bootstrap stage.

**Workaround:** Verify that `ntpd` is running on all nodes. Execute the following command on all the nodes:

- For RHEL/CentOS:

```
service ntpd status
```

- For SLES:

```
service ntp status
```

- **Problem:** Selecting **Use local software repository** option causes Ambari to deploy default stack version. The default stack version for HDP v1.3.3 is `HDP-1.3.3`
- **Workaround:** To install previous version of HDP with local repository option, complete the following instructions:

1. SSH into the Ambari Server host machine and execute the following commands:

```
cd /usr/lib/ambari-server/web/javascripts
```

```
rm app.js
```

```
gunzip app.js.gz
```

```
vi app.js
```

2. Change the value of `App.defaultLocalStackVersion` parameter in `app.js` file to the expected value of HDP release,

For example, to install HDP 1.3.0, change the `App.defaultLocalStackVersion` parameter as shown below:

```
App.defaultLocalStackVersion = 'HDPLocal-1.3.0';
```

3. Execute the following command:

```
gzip app.js
```

4. Clear the browser cache and log in to Ambari Web.

- **Problem:** Ganglia RRD database requires a large amount of disk space. Ganglia collects Hadoop and System metrics for the hosts in the cluster. These metrics are stored in the RRD database. Based on the number of Services you install, and the number of hosts in your clusters, the RRD database could become quite large.

**Workaround:** During cluster install, on the Customize Services page, select the **Misc** tab and set the base directory where RRD stores the collected metrics. Choose a directory location that has a minimum of 16 GB disk space available.

**Workaround:** You can also minimize the space used by Ganglia.

To reduce the Ganglia metrics collection granularity and reduce the overall disk space used by Ganglia, perform these steps after successfully completing your cluster install:

1. Download the following utility script [configs.sh](#).
2. From Ambari, stop the Ganglia service and wait for it to stop completely.
3. Get the existing directory path for Ganglia RRD files (the "rrds" folder) using the `configs.sh` script:

```
./configs.sh get $my.ambari.server $clustername global | grep  
"rrdcached_base_dir"
```

where,

`$my.ambari.server` is the Ambari Server host and  
`$clustername` is the name of the cluster.

4. Log into the Ganglia Server host.
5. Backup the content of the `rrds` folder and then clean the folder.
6. Edit the `gmetadLib.sh` file

```
vi /var/lib/ambari-agent/puppet/modules/hdp-ganglia/files/gmetadLib.sh
```

7. Comment out the existing RRAs entry and enter the following:

```
RRAs "RRA:AVERAGE:0.5:1:244" "RRA:AVERAGE:0.5:24:244" "RRA:AVERAGE:0.  
5:168:244" "RRA:AVERAGE:0.5:672:244" "RRA:AVERAGE:0.5:5760:374"
```

8. From Ambari, start the Ganglia service.
9. To confirm your change is applied, on the Ganglia Server host, you should see the line from above in the `gmetad.conf` file:

```
more /etc/ganglia/hdp/gmetad.conf
```



### Note

You may need to wait for 5-10 minutes to see the metrics populate.

## 5. Release Notes HDP-1.3.2

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading HDP Manually](#)
- [Improvements](#)
- [Known Issues](#)

### 5.1. Product Version: HDP-1.3.2

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6.1
- Apache Pig 0.11
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.2.5
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Hue 2.2.0

- Third party components:
  - Talend Open Studio for Big Data 5.3
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

## 5.2. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 5.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HADOOP-9509](#): Implemented ONCRPC and XDR.
- [HADOOP-9515](#): Added general interface for NFS and Mount.
- [HDFS-4762](#): Added HDFS based NFSv3 and Mountd implementation.
- [HDFS-5038](#): Added the following HDFS branch-2 APIs to HDFS branch-1:
  - `FileSystem#newInstance(Configuration)`
  - `DFSClient#getNamenode()`
  - `FileStatus#isDirectory()`

- [HDFS-4880](#): Added support to print image and edits file loaded by the NameNode, in the logs.
- [HDFS-4944](#): Fixed file path issue with WebHDFS. WebHDFS can now create a file path containing characters that must be URI-encoded (such as space).
- **HDFS Snapshot related changes:**
  - [HDFS-4842](#): Added ability to identify correct prior snapshot before deleting a snapshot under a renamed subtree.
  - [HDFS-4857](#): Enhanced `Snapshot.Root` and `AbstractINodeDiff#snapshotINode`. (`Snapshot.Root` and `AbstractINodeDiff#snapshotINode` should not be put into `INodeMap` when loading `FSImage`.)
  - [HDFS-4863](#): The `root` directory can now be added to the `snaphottable` directory list while loading `fsimage`.
  - [HDFS-4846](#): Enhanced snapshot command line (CLI) commands output stacktrace for invalid arguments.
  - [HDFS-4848](#): Fixed `copyFromLocal` and file rename operations. (While performing `copyFromLocal` operation and/or renaming a file to `.snapshot`, now displays an output message that that `.snapshot` is a reserved name.)
  - [HDFS-4850](#): Fixed `OfflineImageViewer` to work on `fsimages` with empty files or snapshots.
  - [HDFS-4876](#): Fixed JavaDoc for `FileWithSnapshot`.
  - [HDFS-4877](#): Fixed the issues caused while renaming a directory under its prior descendant.
  - [HDFS-4902](#): Fixed path issue for `DFSClient.getSnapshotDiffReport`. `DFSClient.getSnapshotDiffReport` now uses `string` path instead of using the `o.a.h.fs.Path`.
  - [HDFS-4875](#): Added support for testing snapshot file length.
  - [HDFS-5005](#): Moved `SnapshotException` and `SnapshotAccessControlException` to `o.a.h.hdfs.protocol`.
- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.

- [HDFS-4334](#): Added support to enable adding a unique id to each INode.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.
- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

## 5.2.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.2.5 and includes no patches.

## 5.2.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-8816](#): Added support for loading multiple tables into `LoadTestTool`.
- [HBASE-6338](#): Cache method in RPC handler.
- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)

- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and `addendum2+3`).
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

## 5.2.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-2084](#): Upgraded DataNuclues from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.

- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.
- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `multit` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

## 5.2.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

## 5.2.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3236](#): Added support to parametrize snapshot and staging repository ID.



- [PIG-3048](#): Added MapReduce workflow information to job configuration.
- [PIG-3276](#): Changed default value (`/usr/local/hcat/bin/hcat`) for `hcat.bin` to `hcat`.
- [PIG-3277](#): Fixed path to the benchmarks file in the `print` statement.
- [PIG-3071](#): Updated HCatalog JAR file and path to HBase storage handler JAR in the Pig script file.
- [PIG-3262](#): Fixed compilation issues with Pig `contrib 0.11` on certain RPM systems.
- [PIG-2786](#): Enhanced Pig launcher script for HBase/Hcatalog integration.

### 5.2.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to .m2 repository.

### 5.2.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.3.2 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.
- [OOZIE-863](#): `JAVA_HOME` must be explicitly set at client because `bin/oozie` does not invoke `oozie-env.sh`.

### 5.2.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-979](#): Fixed issues for MySQL direct connector, caused after moving password to credential cache.
- [SQOOP-914](#): Added an abort validation handler.
- [SQOOP-916](#): Enhanced security for passwords in Sqoop 1.x.
- [SQOOP-798](#): Fixed Ant docs failure for RHEL v5.8.

## 5.2.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

## 5.2.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- **JMS Source changes:**
  - [FLUME-924](#): Implemented JMS source for Flume NG.
  - [FLUME-1784](#): Fixed issues with documentation and parameter name.
  - [FLUME-1804](#): JMS source not included in binary distribution.
  - [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes.
  - [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.
  - [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
  - [FLUME-2043](#): JMS Source removed on failure to create configuration
- **Spillable Channel - (Experimental)**
  - [FLUME-1227](#): Introduce some sort of `SpillableChannel`.
- **Spillable Channel dependencies:**
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.
  - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
  - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
  - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- **Improvements:**

- [FLUME-1878](#): FileChannel replay should print status every 10000 events.
- [FLUME-1891](#): Fast replay runs even when checkpoint exists.
- [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
- [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
- [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
- [FLUME-1262](#): Move doc generation to a different profile.

## 5.3. Minimum System Requirements

In this section:

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



### Note

gsInstaller was **deprecated** as of HDP 1.2.0 and is no longer being made available in 1.3.0 or in future releases.

We encourage you to consider [Manual Install \(RPMs\)](#) or [Automated Install \(Ambari\)](#).

### 5.3.1. Hardware Recommendations

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

### 5.3.2. Operating Systems Requirements

The following operating systems (OS) are supported:

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

- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- Oracle Linux 5 and 6

### 5.3.3. Software Requirements

On each of your hosts:

- yum (RHEL/CentOS)
- zypper (SLES)



#### Note

Ensure that the Zypper version is 1.3.14.

- rpm
- scp
- curl
- wget
- pdsh

### 5.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and by default uses embedded Derby database. MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are supported. You may provide access to an existing database, or you can use Ambari installer to deploy MySQL instance for your environment. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).
- Oozie requires a database to use as a metadata store and by default uses embedded Derby database.

MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Ambari requires a database to store information about cluster topology and configuration.

The default database is Postgres 8.x and Oracle 11gr2 is also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

### 5.3.5. 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.

### 5.3.6. Optional: Configure 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. For example, if you are using RHEL 6 your hosts must be able to access the “Red Hat Enterprise Linux Server 6 Optional (RPMs)” repository. If this repository is disabled, the installation is unable to access the `rubygems` package. If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 5.4. Upgrading HDP Manually

Use the following instructions to upgrade HDP manually:

For upgrading manually, see [here](#).

For upgrading Ambari server, follow the instructions provided [here](#) and [here](#).

## 5.5. Improvements

- Added support for deploying Hue. For more details, see [Installing and Configuring Hue](#).
- Apache HBase updated to version 0.94.6.1.
- Talend Open Studio for Big Data updated to version 5.3.
- Ganglia updated to version 3.5.0.
- Ganglia Web updated to version 3.5.7
- Nagios updated to version 3.5.0.
- Apache Ambari updated to version 1.2.5. This release of Apache Ambari includes the new features and improvements:
  - Added support to setup Ganglia and Nagios HTTPS
  - Added support to run Ambari Server as non-root account.
  - Added ability to manage Kerberos Secure Cluster.
  - Added support to setup Ambari Server HTTPS.

- Enabled Ambari Server configuration property encryption.
- Added support to configure Ambari Server-Agent Two-Way SSL Communication.
- Added ability to customize Dashboard Widgets.
- Improved Host Checks during Install Wizard

## 5.6. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 5.6.1. Known Issues for Hadoop

- JobTracker UI, JMX queries, and RPC calls sometimes hang in HA mode.

**Problem:** JobTracker becomes slow and non-responsive in HA mode because `dfs.client.retry.policy.enabled` is not set to 'final' and 'false' for the JobTracker.

**Workaround:** Set `dfs.client.retry.policy.enabled` to 'final' and 'false' *only* for the JobTracker. Clients (such as MapReduce, Pig, Hive, Oozie) should still be set to 'true' in HA mode.

**Problem:** While uploading files to NFS-MountDir, the following error is reported in the DataNode log file:

```
INFO org.apache.hadoop.hdfs.nfs.nfs3.OpenFileCtx: requested offset=4980736
and current filesize=0
```

**Workaround:** On some environments, especially for virtualized environments, copying large files of size close to 1GB fails intermittently. This issue is expected to be addressed in the upcoming release.

- Use of init.d scripts for starting or stopping Hadoop services, is not recommended.

### 5.6.2. Known Issues for Hive

- Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```
TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...
```

**Workaround:** The workaround is disable all the memory settings by setting value of the following perperties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```
mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1
```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN ) ?)])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
```

```

at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target

```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **Problem:** While using indexes in Hive, the following error is reported:

```
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask
```

- **Problem:** Partition in hive table that is of datatype int is able to accept string entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype int can also accept string entries while data insertions.

**Workaround:** The workaround is to avoid adding string to int fields.

### 5.6.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.



The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\\"failureInfo\\":\\"JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\\",\\"runState\\":3
```

- **Problem:** WebHCatalog configuration `templeton.libjars` property value is incorrect. For more information see [AMBARI-2862](#).

**Workaround:** Change the value for the property in the `/usr/lib/hcatalog/conf/webhcat-site.xml` file, as shown below:

```
<property>
  <name>templeton.libjars</name>
  <value>/usr/lib/zookeeper/zookeeper.jar,/usr/lib/hcatalog/share/
hcatalog/hcatalog-core.jar,/usr/lib/hive/lib/hive-exec-0.11.0.1.3.2.0-97.
jar,/usr/lib/hive/lib/hive-metastore-0.11.0.1.3.2.0-97.jar,/usr/lib/hive/
lib/libfb303-0.9.0.jar,/usr/lib/hive/lib/libthrift-0.9.0.jar,/usr/lib/
hive/lib/jdo2-api-2.3-ec.jar,/usr/lib/hive/lib/slf4j-api-1.6.1.jar,/usr/
lib/hcatalog/share/webhcat/svr/lib/antlr-runtime-3.4.jar,/usr/lib/hive/
lib/datanucleus-api-jdo-3.0.7.jar,/usr/lib/hive/lib/datanucleus-api-jdo-3.
0.7.jar,/usr/lib/hive/lib/datanucleus-core-3.0.9.jar,/usr/lib/hive/lib/
datanucleus-enhancer-3.0.1.jar,/usr/lib/hive/lib/datanucleus-rdbms-3.0.8.
jar</value>
</property>
```

## 5.6.4. Known Issues for HBase

- HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

- Taking snapshot on Oracle Linux and Java version 1.6.0\_31, may occasionally result in spurious Timeout error.

The workaround is to retry taking the snapshot.

## 5.6.5. Known Issues for Oozie

- **Problem:** Oozie fails smoke tests in secured cluster.

### Workaround:

1. Download the following files attached to, <https://issues.apache.org/jira/browse/AMBARI-2879>:
  - check\_oozie\_status.sh
  - oozieSmoke.sh
2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.
3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **Problem:** `TestBundleJobsFilter` test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of `string integer` to an `integer`. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

- **Problem:** Delegation Token renewal exception in JobTracker logs.

The following exception is reported in the JobTracker log file when executing a long running job on Oozie in secure mode:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:jt/horln22.gq1.ygridcore.net@HORTON.
YGRIDCORE.NET cause:org.apache.hadoop.security.AccessContr
olException: org.apache.hadoop.security.AccessControlException: Client
mapred tries to renew a token with renewer specified as jt
2013-04-25 15:09:41,543 ERROR org.apache.hadoop.mapreduce.security.token.
DelegationTokenRenewal: Exception renewing tokenId: 00 06 68 72 74 5f 71
61 02 6a 74 34 6f 6f 7a 69 65 2f 68 6f 72 31 6e 32 34 2e 67 71 31 2e 79 67
```

```

72 69 64 63 6f 72 65 2e 6e 65 74 40 48 4f 52 54 4f 4e 2e 59 47 52 49 44 43
4f 52 45 2e 4e 45 54 8a 01 3e 41 b9 67 b8 8a 01 3e 65 c5 eb b8 8f 88 8f 9c,
Kind: HDFS_DELEGATION_TOKEN, Service: 68.142.244.41:8020. Not rescheduled
org.apache.hadoop.security.AccessControlException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native
Method)
    at sun.reflect.NativeConstructorAccessorImpl.
newInstance(NativeConstructorAccessorImpl.java:39)
    at sun.reflect.DelegatingConstructorAccessorImpl.
newInstance(DelegatingConstructorAccessorImpl.java:27)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:513)
    at org.apache.hadoop.ipc.RemoteException.
instantiateException(RemoteException.java:95)
    at org.apache.hadoop.ipc.RemoteException.
unwrapRemoteException(RemoteException.java:57)
    at org.apache.hadoop.hdfs.DFSClient$Renewer.renew(DFSClient.
java:678)
    at org.apache.hadoop.security.token.Token.renew(Token.java:309)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:221)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:217)
    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:1195)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask.run(DelegationTokenRenewal.java:216)
    at java.util.TimerThread.mainLoop(Timer.java:512)
    at java.util.TimerThread.run(Timer.java:462)
Caused by: org.apache.hadoop.ipc.RemoteException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at org.apache.hadoop.security.token.
delegation.AbstractDelegationTokenSecretManager.
renewToken(AbstractDelegationTokenSecretManager.java:267)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.
renewDelegationToken(FSNamesystem.java:6280)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.
renewDelegationToken(NameNode.java:652)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.
invoke(NativeMethodAccessorImpl.java:39)
    at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
    at java.lang.reflect.Method.invoke(Method.java:597)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:578)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1405)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1401)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:396)
    at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1195)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:1399)

    at org.apache.hadoop.ipc.Client.call(Client.java:1118)
    at org.apache.hadoop.ipc.RPC$Invoker.invoke(RPC.java:229)
    at $Proxy7.renewDelegationToken(Unknown Source)

```

```
at org.apache.hadoop.hdfs.DFSCClient$Renewer.renew(DFSCClient.
java:676)
... 9 more
```

**Workaround:** Any new job on secure cluster that runs longer than the validity of the Kerberos ticket (typically 24 hours) will fail as the delegation token will not be renewed.

## 5.6.6. Known Issues for Ambari

- **Problem:** Oozie fails smoke tests in secured cluster.

### Workaround:

1. Download the following files attached to <https://issues.apache.org/jira/browse/AMBARI-2879>:
  - check\_oozie\_status.sh
  - oozieSmoke.sh
2. Replace `/var/lib/ambari-agent/puppet/modules/hdp-nagios/files/check_oozie_status.sh` with the downloaded file.
3. On the Nagios Server host machine, restart Nagios using the following command:

```
service nagios start
```

4. Replace `/var/lib/ambari-agent/puppet/modules/hdp-oozie/files/oozieSmoke.sh` with the downloaded file all the hosts in your cluster.
5. Restart Oozie on the Oozie Server host machine using the following command:

```
sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR/log; /usr/lib/oozie/bin/oozie-start.sh"
```

where:

- `$OOZIE_USER` is the Oozie Service user. For example, `oozie`
  - `$OOZIE_LOG_DIR` is the directory where Oozie log files are stored (for example: `/var/log/oozie`).
- **Problem:** The `ambari-server` command displays invalid options for setting up Ganglia and Nagios.

On the Ambari server host machine, when you execute the following command:

```
ambari-server
```

You see the following output:

```
Using python /usr/bin/python2.6
Usage: /usr/sbin/ambari-server {start|stop|restart|setup|upgrade|status|
upgradestack|setup-ldap|setup-https|setup-ganglia_https|setup-nagios_https|
encrypt-passwords} [options]
```

**Workaround:** The `setup-ganglia_https` and `setup-nagios_https` are not valid options.

Use `setup-ganglia-https` and `setup-nagios-https` with the `ambari-server` command to set up Ganglia and Nagios.

- **Problem:** `ntpd` service warning might be displayed as part of host check at the bootstrap stage.

**Workaround:** Verify that `ntpd` is running on all nodes. Execute the following command on all the nodes:

- For RHEL/CentOS:

```
service ntpd status
```

- For SLES:

```
service ntp status
```

- **Problem:** Selecting **Use local software repository** option causes Ambari to deploy default stack version. The default stack version for HDP v1.3.2 is `HDP-1.3.2`
- **Workaround:** To install previous version of HDP with local repository option, complete the following instructions:

1. SSH into the Ambari Server host machine and execute the following commands:

```
cd /usr/lib/ambari-server/web/javascripts
```

```
rm app.js
```

```
gunzip app.js.gz
```

```
vi app.js
```

2. Change the value of `App.defaultLocalStackVersion` parameter in `app.js` file to the expected value of HDP release,

For example, to install HDP 1.3.0, change the `App.defaultLocalStackVersion` parameter as shown below:

```
App.defaultLocalStackVersion = 'HDPLocal-1.3.0';
```

3. Execute the following command:

```
gzip app.js
```

4. Clear the browser cache and log in to Ambari Web.

- **Problem:** Ganglia RRD database requires a large amount of disk space. Ganglia collects Hadoop and System metrics for the hosts in the cluster. These metrics are stored in the RRD database. Based on the number of Services you install, and the number of hosts in your clusters, the RRD database could become quite large.

**Workaround:** During cluster install, on the Customize Services page, select the **Misc** tab and set the base directory where RRD stores the collected metrics. Choose a directory location that has a minimum of 16 GB disk space available.

**Workaround:** You can also minimize the space used by Ganglia.

To reduce the Ganglia metrics collection granularity and reduce the overall disk space used by Ganglia, perform these steps after successfully completing your cluster install:

1. Download the following utility script [configs.sh](#).
2. From Ambari, stop the Ganglia service and wait for it to stop completely.
3. Get the existing directory path for Ganglia RRD files (the "rrds" folder) using the `configs.sh` script:

```
./configs.sh get $my.ambari.server $clustername global | grep  
"rrdcached_base_dir"
```

where,

`$my.ambari.server` is the Ambari Server host and  
`$clustername` is the name of the cluster.

4. Log into the Ganglia Server host.
5. Backup the content of the `rrds` folder and then clean the folder.
6. Edit the `gmetadLib.sh` file

```
vi /var/lib/ambari-agent/puppet/modules/hdp-ganglia/files/gmetadLib.sh
```

7. Comment out the existing RRAs entry and enter the following:

```
RRAs "RRA:AVERAGE:0.5:1:244" "RRA:AVERAGE:0.5:24:244" "RRA:AVERAGE:0.  
5:168:244" "RRA:AVERAGE:0.5:672:244" "RRA:AVERAGE:0.5:5760:374"
```

8. From Ambari, start the Ganglia service.
9. To confirm your change is applied, on the Ganglia Server host, you should see the line from above in the `gmetad.conf` file:

```
more /etc/ganglia/hdp/gmetad.conf
```



### Note

You may need to wait for 5-10 minutes to see the metrics populate.

## 6. Release Notes HDP-1.3.1

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading HDP Manually](#)
- [Improvements](#)
- [Known Issues](#)

### 6.1. Product Version: HDP-1.3.1

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6
- Apache Pig 0.11
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.2.4
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Third party components:
  - Ganglia 3.2.0
  - GWeb 2.2.0

- [Nagios 3.2.3](#)

## 6.2. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 6.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.
- [HDFS-4334](#): Added support to enable adding a unique id to each `Inode`.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for `inode ID` to `inode map`.
- [HDFS-4785](#): Fixed issue for `Concat` operation that affected removal of the concatenated files from `InodeMap`.
- [HDFS-4784](#): Fixed `Null Pointer Exception (NPE)` in `FSDirectory.resolvePath()`.



- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

## 6.2.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.2.4 and includes no patches.

## 6.2.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-6338](#): Cache method in RPC handler.
- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)

- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and `addendum2+3`).
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.
- [HBASE-8350](#): Added support to enable ChaosMonkey to run commands as different users.
- [HBASE-8405](#): Added new custom options to how `ClusterManager` runs commands.
- [HBASE-8465](#): Added support for auto-drop rollback snapshot for snapshot restore.
- [HBASE-8455](#): Updated `ExportSnapshot` to reflect changes in [HBASE-7419](#).
- [HBASE-8413](#): Fixed `Snapshot verify region` will always fail if the HFile has been archived.
- [HBASE-8259](#): Snapshot backport in 0.94.6 breaks rolling restarts.
- [HBASE-8213](#): Fixed global authorization may lose efficacy.

## 6.2.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-2084](#): Upgraded DataNucleus from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.

- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `multit` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 6.2.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 6.2.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3048](#): Added MapReduce workflow information to job configuration.

### 6.2.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

## 6.2.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.2.0 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.

## 6.2.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-931](#): Added support to integrate Apache HCatalog with Apache Sqoop.  
This Sqoop-HCatalog connector supports storage formats abstracted by HCatalog.
- [SQOOP-916](#): Added an abort validation handler.
- [SQOOP-798](#): Ant docs fail to work on RHEL v5.8.

## 6.2.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

## 6.2.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- [FLUME-924](#): Implemented JMS source for Flume NG.
- [FLUME-1784](#): `JMSsource` Fixed minor documentation problem and parameter name.
- [FLUME-1804](#): JMS source not included in binary distribution.
- [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes
- [FLUME-1886](#): Added JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSsource`.

- [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
- [FLUME-2043](#): JMS Source removed on failure to create configuration
- [FLUME-1227](#): Introduce some sort of SpillableChannel ([Spillable Channel - Experimental]).
- Spillable Channel dependencies:
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.
  - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
  - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
  - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
- [FLUME-1891](#): Fast replay runs even when checkpoint exists.
- [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
- [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
- [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
- [FLUME-1262](#): Move doc generation to a different profile.

## 6.3. Minimum System Requirements

In this section:

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



### Note

`gsInstaller` was **deprecated** as of HDP 1.2.0 and is no longer being made available in 1.3.0 or in future releases.

We encourage you to consider [Manual Install \(RPMs\)](#) or [Automated Install \(Ambari\)](#).

### 6.3.1. Hardware Recommendations

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

### 6.3.2. Operating Systems Requirements

The following operating systems (OS) are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- Oracle Linux 5 and 6

### 6.3.3. Software Requirements

On each of your hosts:

- yum (RHEL/CentOS)
- zypper (SLES)



#### Note

Ensure that the Zypper version is 1.3.14.

- rpm
- scp
- curl
- wget
- pdsh

### 6.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and by default uses embedded Derby database. MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are supported. You may provide access to an existing database, or you can use Ambari installer to deploy MySQL instance for your environment. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).
- Oozie requires a database to use as a metadata store and by default uses embedded Derby database.

MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Ambari requires a database to store information about cluster topology and configuration.

The default database is Postgres 8.x and Oracle 11gr2 is also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

### 6.3.5. 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.

### 6.3.6. Optional: Configure 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. For example, if you are using RHEL 6 your hosts must be able to access the “Red Hat Enterprise Linux Server 6 Optional (RPMs)” repository. If this repository is disabled, the installation is unable to access the `rubygems` package. If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 6.4. Upgrading HDP Manually

Use the following instructions to upgrade HDP manually:

1. For SUSE, you must uninstall before updating the repo file. The instructions to uninstall HDP are provided [here](#).
2. For RHEL/CentOS, use one of the following options to upgrade HDP:
  - Option I:
    - a. Uninstall HDP using the instructions provided [here](#).
    - b. Install HDP using the instructions provided [here](#).
  - Option II: Update the repo using the instructions provided [here](#).

## 6.5. Improvements

- Apache Ambari updated to version 1.2.4. This release of Apache Ambari includes the new features and improvements:

- Ambari requires a database to store information about cluster topology and configuration.

The default database is Postgres 8.x and Oracle 11gr2 is also supported. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Added support for configuring Oracle 11gr2 for Oozie and Hive metastores. For more information, see [Supported Database Matrix for Hortonworks Data Platform](#).
- Added support for non-root SSH install option.
- Added support to use either HDP-1.2.1 or HDP-1.3.0 stack.

## 6.6. Known Issues

In this section:

- [Known Issues for Hadoop](#)
- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

### 6.6.1. Known Issues for Hadoop

- File upload fails to upload in NFS-MountDir.

**Problem:** While uploading files to NFS-MountDir, the following error is reported in the DataNode log file:

```
INFO org.apache.hadoop.hdfs.nfs.nfs3.OpenFileCtx: requested offset=4980736  
and current filesize=0
```

**Workaround:** On some environments, especially for virtualized environments, copying large files of size close to 1GB fails intermittently. This issue is expected to be addressed in the upcoming release.

- Use of init.d scripts for starting or stopping Hadoop services, is not recommended.

### 6.6.2. Known Issues for Hive

- Mapreduce task from Hive dynamic partitioning query is killed.



**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```
TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...
```

**Workaround:** The workaround is disable all the memory settings by setting value of the following perperties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```
mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1
```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN ) ?)])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
```

```

at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target

```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **Problem:** While using indexes in Hive, the following error is reported:

```
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask
```

- **Problem:** Partition in hive table that is of datatype int is able to accept string entries. For example,

```
CREATE TABLE tab1 (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype int can also accept string entries while data insertions.

**Workaround:** The workaround is to avoid adding string to int fields.

### 6.6.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\failureInfo\":"\JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\","\runState\":3
```

## 6.6.4. Known Issues for HBase

- HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

## 6.6.5. Known Issues for Oozie

- TestBundleJobsFilter test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of string integer to an integer. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

- Delegation Token renewal exception in JobTracker logs.

**Problem:** The following exception is reported in the JobTracker log file when executing a long running job on Oozie in secure mode:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:jt/horln22.gql.ygridcore.net@HORTON.
YGRIDCORE.NET cause:org.apache.hadoop.security.AccessContr
olException: org.apache.hadoop.security.AccessControlException: Client
mapred tries to renew a token with renewer specified as jt
2013-04-25 15:09:41,543 ERROR org.apache.hadoop.mapreduce.security.token.
DelegationTokenRenewal: Exception renewing tokenIdent: 00 06 68 72 74 5f 71
61 02 6a 74 34 6f 6f 7a 69 65 2f 68 6f 72 31 6e 32 34 2e 67 71 31 2e 79 67
72 69 64 63 6f 72 65 2e 6e 65 74 40 48 4f 52 54 4f 4e 2e 59 47 52 49 44 43
```

```
4f 52 45 2e 4e 45 54 8a 01 3e 41 b9 67 b8 8a 01 3e 65 c5 eb b8 8f 88 8f 9c,
Kind: HDFS_DELEGATION_TOKEN, Service: 68.142.244.41:8020. Not rescheduled
org.apache.hadoop.security.AccessControlException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native
Method)
    at sun.reflect.NativeConstructorAccessorImpl.
newInstance(NativeConstructorAccessorImpl.java:39)
    at sun.reflect.DelegatingConstructorAccessorImpl.
newInstance(DelegatingConstructorAccessorImpl.java:27)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:513)
    at org.apache.hadoop.ipc.RemoteException.
instantiateException(RemoteException.java:95)
    at org.apache.hadoop.ipc.RemoteException.
unwrapRemoteException(RemoteException.java:57)
    at org.apache.hadoop.hdfs.DFSClient$Renewer.renew(DFSClient.
java:678)
    at org.apache.hadoop.security.token.Token.renew(Token.java:309)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:221)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:217)
    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:1195)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask.run(DelegationTokenRenewal.java:216)
    at java.util.TimerThread.mainLoop(Timer.java:512)
    at java.util.TimerThread.run(Timer.java:462)
Caused by: org.apache.hadoop.ipc.RemoteException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at org.apache.hadoop.security.token.
delegation.AbstractDelegationTokenSecretManager.
renewToken(AbstractDelegationTokenSecretManager.java:267)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.
renewDelegationToken(FSNamesystem.java:6280)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.
renewDelegationToken(NameNode.java:652)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.
invoke(NativeMethodAccessorImpl.java:39)
    at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
    at java.lang.reflect.Method.invoke(Method.java:597)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:578)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1405)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1401)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:396)
    at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1195)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:1399)

    at org.apache.hadoop.ipc.Client.call(Client.java:1118)
    at org.apache.hadoop.ipc.RPC$Invoker.invoke(RPC.java:229)
    at $Proxy7.renewDelegationToken(Unknown Source)
```

```
at org.apache.hadoop.hdfs.DFSClient$Renewer.renew(DFSClient.  
java:676)  
... 9 more
```

**Workaround:** Any new job on secure cluster that runs longer than the validity of the Kerberos ticket (typically 24 hours) will fail as the delegation token will not be renewed.

## 6.6.6. Known Issues for Ambari

- Nagios assumes that DataNode is deployed on all the host machine in your cluster.

The Nagios server displays DataNode alert on all the host machines even if a particular slave machine does not host a DataNode daemon.

- Ambari user interface (UI) allows adding existing properties to custom `core-site.xml` and `hdfs-site.xml` settings. For more information, see [AMBARI-2313](#)

For more information on the specific issues for Ambari, see [Troubleshooting - Specific Issues](#) section.

## 7. Release Notes HDP-1.3.0

This chapter provides information on the product version, patch information for various components, improvements, and known issues (if any) for the current release.

This document contains:

- [Product Version](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading HDP Manually](#)
- [Improvements](#)
- [Known Issues](#)

### 7.1. Product Version: HDP-1.3.0

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

- Apache Hadoop 1.2.0
- Apache HBase 0.94.6
- Apache Pig 0.11
- Apache ZooKeeper 3.4.5
- Apache HCatalog



#### Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Ambari 1.2.3
- Apache Flume 1.3.1
- Apache Mahout 0.7.0
- Third party components:

- [Ganglia 3.2.0](#)
- [GWeb 2.2.0](#)
- [Nagios 3.2.3](#)

## 7.2. Patch Information

In this section:

- [Patch information for Hadoop](#)
- [Patch information for Ambari](#)
- [Patch information for HBase](#)
- [Patch information for Hive](#)
- [Patch information for HCatalog](#)
- [Patch information for Pig](#)
- [Patch information for ZooKeeper](#)
- [Patch information for Oozie](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)
- [Patch information for Flume](#)

### 7.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches:

- [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.
- [HDFS-4750](#): Added support for NFSv3 interface to HDFS.
- [MAPREDUCE-5109](#): Added support to apply `Job view-acl` to job lists on JobTracker and also to the JobHistory listings.
- [MAPREDUCE-5217](#): Fixed issues for `DistCP` when launched by Oozie on a secure cluster.
- [MAPREDUCE-5256](#): Improved `CombineInputFormat` to make it thread safe. This issue was affecting HiveServer.
- [HDFS-4334](#): Added support to enable adding a unique id to each INode.
- [HDFS-4635](#): Move `BlockManager#computeCapacity` to `LightWeightGSet`.
- [HDFS-4434](#): Added support for inode ID to inode map.

- [HDFS-4785](#): Fixed issue for Concat operation that affected removal of the concatenated files from InodeMap.
- [HDFS-4784](#): Fixed Null Pointer Exception (NPE) in `FSDirectory.resolvePath()`.
- [HADOOP-8923](#): Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
- [HDFS-4108](#): Fixed `dfsnodelist` to work in secure mode.
- [HADOOP-9296](#): Added support to allow users from different realm to authenticate without a trust relationship.

## 7.2.2. Patch information for Ambari

Ambari is based on Apache Ambari 1.2.3 and includes the following:

- [AMBARI-1983](#): Added new parameters to improve HBase Mean Time To Recover (MTTR).
- [AMBARI-2136](#): Fixed incorrect HOME paths in `/etc/sqoop/conf/sqoop-env.sh` file.
- [AMBARI-2198](#): Avoid using `fqdn` in Puppet which uses the FQDN value from `puppet/facter`, instead pass the hostname parameter from Python `socket.getfqdn()`.
- [AMBARI-2110](#): Updated the `fs.file.impl.disable.cache=true` property in the `hive-site.xml` file.
- [AMBARI-2134](#): Fixed Oozie proxy test failure on Ambari deployed cluster.
- [AMBARI-2149](#): Fixed incorrect GC log directory path for HBase process.
- [AMBARI-2141](#): Fixed when HBase user is changed, HBase fails to start after upgrade.
- [AMBARI-2146](#): Fixed when Hive and Oozie users have been changed, after upgrade Hive metastore and Oozie fail to start.
- [AMBARI-2165](#): Ambari server upgrade fails when a user tries to upgrade for the second time.
- [AMBARI-2164](#): `START_FAILED` and `STOP_FAILED` no longer exist, the upgrade script should repair `hostcomponentstate` table to convert these to `INSTALLED`.
- [AMBARI-2117](#): Updated `mapred.jobtracker.retirejob.interval` to 21600000 (6 hours).
- [AMBARI-2116](#): Updated default configurations for Ambari to improve performance.

## 7.2.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following:

- [HBASE-6338](#): Cache method in RPC handler.



- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "-".
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8505](#): References to split daughters should not be deleted separately from parent META entry (patch file: `hbase-8505_v2-0.94-reduce.patch`).
- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and addendum2+3).

- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.
- [HBASE-8350](#): Added support to enable `ChaosMonkey` to run commands as different users.
- [HBASE-8405](#): Added new custom options to how `ClusterManager` runs commands.
- [HBASE-8465](#): Added support for auto-drop rollback snapshot for snapshot restore.
- [HBASE-8455](#): Updated `ExportSnapshot` to reflect changes in [HBASE-7419](#).
- [HBASE-8413](#): Fixed `Snapshot verify region` will always fail if the HFile has been archived.
- [HBASE-8259](#): Snapshot backport in 0.94.6 breaks rolling restarts.
- [HBASE-8213](#): Fixed global authorization may lose efficacy.

## 7.2.4. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches:



### Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-2084](#): Upgraded DataNucleus from v2.0.3 to v3.0.1.
- [HIVE-3815](#): Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- [HIVE-3846](#): Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- [HIVE-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.
- [HIVE-4171](#): Current database in metastore. Hive is not consistent with `SessionState`.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `mult` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- [HIVE-4510](#): Fixed HiveServer2 nested exceptions.
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4521](#): Fixed auto join conversion failures

- [HIVE-4540](#): Fixed failures for GROUPBY/DISTINCT operations when `mapjoin.mapred=true`.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataSetSchemas` failures.
- [HIVE-3255](#): Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.
- Removed `npath` windowing function.
- [HIVE-4465](#): Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).

### 7.2.5. Patch information for HCatalog

Apache HCatalog is now merged with Apache Hive. For details on the list of patches, see [Patch information for Hive](#).

### 7.2.6. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches:

- [PIG-3048](#): Added MapReduce workflow information to job configuration.

### 7.2.7. Patch information for ZooKeeper

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

- [ZOOKEEPER-1598](#): Enhanced ZooKeeper version string.
- [ZOOKEEPER-1584](#): Adding `mvn-install` target for deploying the ZooKeeper artifacts to .m2 repository.

### 7.2.8. Patch information for Oozie

Oozie is based on Apache Oozie 3.2.0 and includes the following patches:

- [OOZIE-1356](#): Fixed issue with the Bundle job in `PAUSEWITHERROR` state that fails change to `SUSPENDEDWITHERROR` state on suspending the job.
- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.

## 7.2.9. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches:

- [SQOOP-931](#): Added support to integrate Apache HCatalog with Apache Sqoop.

This Sqoop-HCatalog connector supports storage formats abstracted by HCatalog.

- [SQOOP-916](#): Added an abort validation handler.
- [SQOOP-798](#): Ant docs fail to work on RHEL v5.8.

## 7.2.10. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches:

- [MAHOUT-958](#): Fixed `NullPointerException` in `RepresentativePointsMapper` when running `cluster-reuters.sh` example with `kmeans`.
- [MAHOUT-1102](#): Fixed Mahout build failures for default profile caused when `hadoop.version` is passed as an argument.
- [MAHOUT-1120](#): Fixed execution failures for Mahout examples script for RPM based installations.

## 7.2.11. Patch information for Flume

Flume is based on Apache Flume 1.3.1 and includes the following patches:

- [FLUME-924](#): Implement a JMS source for Flume NG.
- [FLUME-1784](#): `JMSSource` fix minor documentation problem and parameter name.
- [FLUME-1804](#): JMS source not included in binary distribution.
- [FLUME-1777](#): `AbstractSource` does not provide enough implementation for subclasses
- [FLUME-1886](#): Add a JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSSource`.
- [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
- [FLUME-2043](#): JMS Source removed on failure to create configuration
- [FLUME-1227](#): Introduce some sort of `SpillableChannel` ([`Spillable Channel - Experimental`]).
- Spillable Channel dependencies:
  - [FLUME-1630](#): Improved Flume configuration code.
  - [FLUME-1502](#): Support for running simple configurations embedded in host process.

- [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from `configure` method.
- [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
- [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
- [FLUME-1891](#): Fast replay runs even when checkpoint exists.
- [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
- [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
- [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
- [FLUME-1262](#): Move doc generation to a different profile.

## 7.3. Minimum System Requirements

In this section:

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



### Note

`gsInstaller` was **deprecated** as of HDP 1.2.0 and is no longer being made available in 1.3.0 or in future releases.

We encourage you to consider [Manual Install \(RPMs\)](#) or [Automated Install \(Ambari\)](#).

### 7.3.1. Hardware Recommendations

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

### 7.3.2. Operating Systems Requirements

The following operating systems (OS) are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.\*, v6.\*
- 64-bit CentOS v5.\*, v6.\*
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- Oracle Linux 5 and 6

### 7.3.3. Software Requirements

On each of your hosts:

- yum (RHEL/CentOS)
- zypper (SLES)



#### Note

Ensure that the Zypper version is 1.3.14.

- rpm
- scp
- curl
- wget
- pdsh

### 7.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and by default uses embedded Derby database. MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are supported. You may provide access to an existing database, or you can use Ambari installer to deploy MySQL instance for your environment.
- Oozie requires a database to use as a metadata store and by default uses embedded Derby database.

MySQL 5.x, Oracle 11gr2, or PostgreSQL 8.x are also supported.

- Ambari requires a database to use as a metadata store and uses Postgres 8.x. This is the only database supported in this version.

### 7.3.5. 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.

## 7.3.6. Optional: Configure 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. For example, if you are using RHEL 6 your hosts must be able to access the “Red Hat Enterprise Linux Server 6 Optional (RPMs)” repository. If this repository is disabled, the installation is unable to access the `rubygems` package. If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

## 7.4. Upgrading HDP Manually

Use the following instructions to upgrade HDP manually:

1. For SUSE, you must uninstall before updating the repo file. The instructions to uninstall HDP are provided [here](#).
2. For RHEL/CentOS, use one of the following options to upgrade HDP:
  - Option I:
    - a. Uninstall HDP using the instructions provided [here](#).
    - b. Install HDP using the instructions provided [here](#).
  - Option II: Update the repo using the instructions provided [here](#).

## 7.5. Improvements

- Apache Hadoop updated to version 1.2.0
- Apache HBase updated to version 0.94.6.
- Apache Pig updated to version 0.11.
- Apache Hive updated to version 0.11.
- Apache Oozie updated to version 3.3.2.
- Apache Sqoop updated to version 1.4.3.
- Added support for PostgreSQL v.8.x for Hive Metastore, Oozie, and Sqoop. For more details, see [Supported Database Matrix for Hortonworks Data Platform](#).

- Added the following to Apache Hadoop:
  - [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.  
  
Snapshots are point in time images of parts of the filesystem or the entire filesystem. Snapshots can be a read-only or a read-write point in time copy of the filesystem. There are several use cases for snapshots in HDFS. For details, see [User Guide - HDFS Snapshots](#).
  - [HDFS-4750](#): Added support for NFSv3 interface to HDFS. NFS interface support provides the ability for HDFS to have seamless integration with client's file system. For details, see [User Guide - HDFS NFS Gateway](#).
- Added the following to Apache Flume NG:
  - Implemented a JMS source for Apache Flume NG. See [FLUME-924](#), [FLUME-1784](#), [FLUME-1804](#), [FLUME-1777](#), [FLUME-1886](#), [FLUME-1976](#), and [FLUME-2043](#). Also see [Apache Flume Documentation](#).
  - Added SpillableChannel (experimental) to Apache Flume NG. See [FLUME-1227](#) for more details.  
  
Also see [FLUME-1630](#), [FLUME-1502](#), [FLUME-1772](#), [FLUME-1852](#), and [FLUME-1849](#) for SpillableChannel dependencies.
  - Improvements to Flume NG: [FLUME-1878](#), [FLUME-1891](#), and [FLUME-1762](#).
  - Bug fixes: [FLUME-1870](#), [FLUME-1918](#), and [FLUME-1262](#).
- Added support to integrate Apache HCatalog with Apache Sqoop.  
  
This Sqoop-HCatalog connector supports storage formats abstracted by HCatalog. For more information, see [SQOOP-931](#).
- Apache Ambari updated to version 1.2.3. This release of Apache Ambari includes the new features and improvements:
  - Added support for Oracle Linux 5 and 6 (64-bit)
  - Added support for heterogenous OS clusters.
  - Added support to customize "Ganglia" user account.
  - Added support to customize Hive Metastore log directory.
  - Added support for HBase Heatmaps.
  - Improved Monitoring and Analysis Job Diagnostics Visualization.

## 7.6. Known Issues

In this section:

- [Known Issues for Hadoop](#)



- [Known Issues for Hive](#)
- [Known Issues for WebHCatalog](#)
- [Known Issues for HBase](#)
- [Known Issues for Oozie](#)
- [Known Issues for Ambari](#)

## 7.6.1. Known Issues for Hadoop

- File upload fails to upload in NFS-MountDir.

**Problem:** While uploading files to NFS-MountDir, the following error is reported in the DataNode log file:

```
INFO org.apache.hadoop.hdfs.nfs.nfs3.OpenFileCtx: requested offset=4980736
and current filesize=0
```

**Workaround:** On some environments, especially for virtualized environments, copying large files of size close to 1GB fails intermittently. This issue is expected to be addressed in the upcoming release.

- Use of init.d scripts for starting or stopping Hadoop services, is not recommended.

## 7.6.2. Known Issues for Hive

- Mapreduce task from Hive dynamic partitioning query is killed.

**Problem:** When using the Hive script to create and populate the partitioned table dynamically, the following error is reported in the TaskTracker log file:

```
TaskTree [pid=30275,tipID=attempt_201305041854_0350_m_000000_0]
is running beyond memory-limits. Current usage : 1619562496bytes.
Limit : 1610612736bytes. Killing task. TaskTree [pid=30275,tipID=
attempt_201305041854_0350_m_000000_0] is running beyond memory-limits.
Current usage : 1619562496bytes. Limit : 1610612736bytes. Killing task.
Dump of the process-tree for attempt_201305041854_0350_m_000000_0 : |-
PID PPID PGRPID SESSID CMD_NAME USER_MODE_TIME(MILLIS) SYSTEM_TIME(MILLIS)
VMEM_USAGE(BYTES) RSSMEM_USAGE(PAGES) FULL_CMD_LINE |- 30275 20786 30275
30275 (java) 2179 476 1619562496 190241 /usr/jdk64/jdk1.6.0_31/jre/bin/
java ...
```

**Workaround:** The workaround is disable all the memory settings by setting value of the following properties to -1 in the `mapred-site.xml` file on the JobTracker and TaskTracker host machines in your cluster:

```
mapred.cluster.map.memory.mb = -1
mapred.cluster.reduce.memory.mb = -1
mapred.job.map.memory.mb = -1
mapred.job.reduce.memory.mb = -1
mapred.cluster.max.map.memory.mb = -1
mapred.cluster.max.reduce.memory.mb = -1
```

To change these values using the UI, use the instructions provided [here](#) to update these properties.

- **Problem:** While executing the following query:

```
select s, avg(d) over (partition by i order by f, b) from over100k;
```

the following error is reported in the Hive log file:

```
FAILED: SemanticException Range based Window Frame can have only 1 Sort Key
```

**Workaround:** The workaround is to use the following query:

```
select s, avg(d) over (partition by i order by f, b rows unbounded
preceding) from over100k;
```

- **Problem:** While executing the following query:

```
select s, i, avg(d) over (partition by s order by i) / 10.0 from over100k;
```

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
identifier ( COMMA identifier )* RPAREN ) ?)])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.
java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
```

```
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
'from' in selection target
```

**Workaround:** The workaround is to use the following query:

```
select s, i, avg(d) / 10.0 over (partition by s order by i) from over100k;
```

- **Problem:** While using indexes in Hive, the following error is reported:

```
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
MapRedTask
```

- **Problem:** Partition in hive table that is of datatype int is able to accept string entries. For example,

```
CREATE TABLE tabl (id1 int,id2 string) PARTITIONED BY(month string,day int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
```

In the above example, the partition day of datatype int can also accept string entries while data insertions.

**Workaround:** The workaround is to avoid adding string to int fields.

### 7.6.3. Known Issues for WebHCatalog

- **Problem:** WebHCat is unable to submit Hive jobs when running in secure mode. All Hive operations will fail.

The following error is reported in the Hive log file:

```
FAILED: Error in metadata: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.metastore.HiveMetaStoreClient
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.
DDLTask
templeton: job failed with exit code 1
```

- **Problem:** Failure to report correct state for the killed job in WebHCatalog.

The following error is reported in the WebHCatalog log file:

```
\\"failureInfo\\":\\"JobCleanup Task Failure, Task:
task_201304012042_0406_m_000002\\",\\"runState\\":3
```

### 7.6.4. Known Issues for HBase

- HBase RegionServers fails to shutdown.

**Problem:** RegionServers may fail to shutdown. The following error is reported in the RegionServer log file:

```
INFO org.apache.hadoop.hdfs.DFSCClient: Could not complete /apps/
hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/
27063b9e4ebc4644adb36571b5f76ed5 retrying...
```

and the following error is reported in the NameNode log file:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:hbase cause:org.apache.hadoop.hdfs.server.
namenode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/
3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4ebc4644adb36571b5f76ed5. Name
node is in safe mode.
```

## 7.6.5. Known Issues for Oozie

- TestBundleJobsFilter test fails on RHEL v6.3, Oracle v6.3, and SUSE clusters with PostgreSQL.

This issue is caused due to the strict typing of PostgreSQL which restricts the auto casting of string integer to an integer. The issue is reported when string representation of integer values is substituted into a query for PostgreSQL on the JPA layer.

- Delegation Token renewal exception in JobTracker logs.

**Problem:** The following exception is reported in the JobTracker log file when executing a long running job on Oozie in secure mode:

```
ERROR org.apache.hadoop.security.UserGroupInformation:
PrivilegedActionException as:jt/horln22.gq1.ygridcore.net@HORTON.
YGRIDCORE.NET cause:org.apache.hadoop.security.AccessContr
olException: org.apache.hadoop.security.AccessControlException: Client
mapred tries to renew a token with renewer specified as jt
2013-04-25 15:09:41,543 ERROR org.apache.hadoop.mapreduce.security.token.
DelegationTokenRenewal: Exception renewing tokenId: 00 06 68 72 74 5f 71
61 02 6a 74 34 6f 6f 7a 69 65 2f 68 6f 72 31 6e 32 34 2e 67 71 31 2e 79 67
72 69 64 63 6f 72 65 2e 6e 65 74 40 48 4f 52 54 4f 4e 2e 59 47 52 49 44 43
4f 52 45 2e 4e 45 54 8a 01 3e 41 b9 67 b8 8a 01 3e 65 c5 eb b8 8f 88 8f 9c,
Kind: HDFS_DELEGATION_TOKEN, Service: 68.142.244.41:8020. Not rescheduled
org.apache.hadoop.security.AccessControlException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native
Method)
    at sun.reflect.NativeConstructorAccessorImpl.
newInstance(NativeConstructorAccessorImpl.java:39)
    at sun.reflect.DelegatingConstructorAccessorImpl.
newInstance(DelegatingConstructorAccessorImpl.java:27)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:513)
    at org.apache.hadoop.ipc.RemoteException.
instantiateException(RemoteException.java:95)
    at org.apache.hadoop.ipc.RemoteException.
unwrapRemoteException(RemoteException.java:57)
    at org.apache.hadoop.hdfs.DFSClient$Renewer.renew(DFSClient.
java:678)
    at org.apache.hadoop.security.token.Token.renew(Token.java:309)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:221)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask$1.run(DelegationTokenRenewal.java:217)
    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:1195)
    at org.apache.hadoop.mapreduce.security.token.DelegationTokenRenewal
$RenewalTimerTask.run(DelegationTokenRenewal.java:216)
```

```
    at java.util.TimerThread.mainLoop(Timer.java:512)
    at java.util.TimerThread.run(Timer.java:462)
Caused by: org.apache.hadoop.ipc.RemoteException: org.apache.hadoop.
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
    at org.apache.hadoop.security.token.
delegation.AbstractDelegationTokenSecretManager.
renewToken(AbstractDelegationTokenSecretManager.java:267)
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.
renewDelegationToken(FSNamesystem.java:6280)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.
renewDelegationToken(NameNode.java:652)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.
invoke(NativeMethodAccessorImpl.java:39)
    at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
    at java.lang.reflect.Method.invoke(Method.java:597)
    at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:578)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1405)
    at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1401)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:396)
    at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1195)
    at org.apache.hadoop.ipc.Server$Handler.run(Server.java:1399)

    at org.apache.hadoop.ipc.Client.call(Client.java:1118)
    at org.apache.hadoop.ipc.RPC$Invoker.invoke(RPC.java:229)
    at $Proxy7.renewDelegationToken(Unknown Source)
    at org.apache.hadoop.hdfs.DFSClient$Renewer.renew(DFSClient.
java:676)
    ... 9 more
```

**Workaround:** Any new job on secure cluster that runs longer than the validity of the Kerberos ticket (typically 24 hours) will fail as the delegation token will not be renewed.

## 7.6.6. Known Issues for Ambari

- Nagios assumes that DataNode is deployed on all the host machine in your cluster.

The Nagios server displays DataNode alert on all the host machines even if a particular slave machine does not host a DataNode daemon.

For more information on the specific issues for Ambari, see [Troubleshooting - Specific Issues](#) section.