Hortonworks Data Platform: Release Notes
Copyright © 2012, 2013 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-1.3.0 ........................................................................................................... 1  
   1.1. Product Version: HDP-1.3.0 ......................................................................................... 1  
   1.2. Patch Information ...................................................................................................... 2  
      1.2.1. Patch information for Hadoop ........................................................................... 2  
      1.2.2. Patch information for Ambari ......................................................................... 3  
      1.2.3. Patch information for HBase .......................................................................... 3  
      1.2.4. Patch information for Hive .............................................................................. 5  
      1.2.5. Patch information for HCatalog ....................................................................... 6  
      1.2.6. Patch information for Pig ................................................................................. 6  
      1.2.7. Patch information for ZooKeeper ..................................................................... 6  
      1.2.8. Patch information for Oozie .......................................................................... 6  
      1.2.9. Patch information for Sqoop .......................................................................... 7  
      1.2.10. Patch information for Mahout ....................................................................... 7  
      1.2.11. Patch information for Flume .......................................................................... 7  
   1.3. Minimum System Requirements .................................................................................. 8  
      1.3.1. Hardware Recommendations .............................................................................. 9  
      1.3.2. Operating Systems Requirements ..................................................................... 9  
      1.3.3. Software Requirements ....................................................................................... 9  
      1.3.4. Database Requirements ....................................................................................... 9  
      1.3.5. Virtualization and Cloud Platforms .................................................................... 10  
      1.3.6. Optional: Configure the Local Repositories ......................................................... 10  
   1.4. Upgrading HDP Manually ............................................................................................... 10  
   1.5. Improvements ............................................................................................................. 10  
   1.6. Known Issues .............................................................................................................. 12  
      1.6.1. Known Issues for Hadoop .................................................................................. 12  
      1.6.2. Known Issues for Hive ...................................................................................... 12  
      1.6.3. Known Issues for WebHCatalog ....................................................................... 15  
      1.6.4. Known Issues for HBase ................................................................................. 15  
      1.6.5. Known Issues for Oozie ..................................................................................... 15  
      1.6.6. Known Issues for Ambari ............................................................................... 17  
2. Release Notes HDP-1.2.4 ..................................................................................................... 18  
   2.1. Product Version: HDP-1.2.4 ....................................................................................... 18  
   2.2. Patch Information ...................................................................................................... 19  
      2.2.1. Patch information for Hadoop ........................................................................... 19  
      2.2.2. Patch information for HBase .......................................................................... 20  
      2.2.3. Patch information for Hive .............................................................................. 21  
      2.2.4. Patch information for HCatalog ....................................................................... 22  
      2.2.5. Patch information for Pig ................................................................................. 22  
      2.2.6. Patch information for ZooKeeper ..................................................................... 22  
      2.2.7. Patch information for Oozie .......................................................................... 22  
      2.2.8. Patch information for Sqoop .......................................................................... 23  
      2.2.9. Patch information for Mahout ....................................................................... 23  
      2.2.10. Patch information for Ambari ....................................................................... 23  
   2.3. Minimum System Requirements .................................................................................. 24  
      2.3.1. Hardware Recommendations .............................................................................. 24  
      2.3.2. Operating Systems Requirements ..................................................................... 24  
      2.3.3. Software Requirements ....................................................................................... 24  
      2.3.4. Database Requirements ....................................................................................... 25
iv
7. Release Notes HDP-1.2.0 ................................................................. 66
  7.1. Product Version: HDP-1.2.0 ......................................................... 66
  7.2. Patch Information ................................................................. 67
    7.2.1. Patch information for Hadoop ............................................. 67
    7.2.2. Patch information for HBase .............................................. 67
    7.2.3. Patch information for Hive ................................................ 68
    7.2.4. Patch information for HCatalog ......................................... 69
    7.2.5. Patch information for Pig .................................................. 70
    7.2.6. Patch information for ZooKeeper ........................................ 70
    7.2.7. Patch information for Oozie .............................................. 70
    7.2.8. Patch information for Sqoop ............................................. 70
    7.2.9. Patch information for Ambari ............................................ 71
    7.2.10. Patch information for Mahout .......................................... 71
  7.3. Minimum system requirements .............................................. 71
  7.4. Improvements ........................................................................ 72
  7.5. Known Issues ......................................................................... 73
8. Release Notes HDP-1.1.1.16 .......................................................... 75
  8.1. Product Version: HDP-1.1.1.16 .................................................. 75
  8.2. Patch Information .................................................................. 75
    8.2.1. Patch information for Hadoop ............................................. 75
    8.2.2. Patch information for HBase .............................................. 76
    8.2.3. Patch information for Hive ................................................ 77
    8.2.4. Patch information for HCatalog ......................................... 78
    8.2.5. Patch information for Pig .................................................. 79
    8.2.6. Patch information for Oozie .............................................. 79
    8.2.7. Patch information for Sqoop ............................................. 79
    8.2.8. Patch information for Ambari ............................................ 79
  8.3. Minimum system requirements .............................................. 80
  8.4. Improvements ........................................................................ 81
  8.5. Known Issues ......................................................................... 81
9. Release Notes HDP-1.1.0.15 .......................................................... 83
  9.1. Product Version: HDP-1.1.0.15 .................................................. 83
  9.2. Patch Information .................................................................. 83
    9.2.1. Patch information for Hadoop ............................................. 83
    9.2.2. Patch information for HBase .............................................. 84
    9.2.3. Patch information for Hive ................................................ 85
    9.2.4. Patch information for HCatalog ......................................... 86
    9.2.5. Patch information for Pig .................................................. 86
    9.2.6. Patch information for Oozie .............................................. 86
    9.2.7. Patch information for Sqoop ............................................. 86
    9.2.8. Patch information for Ambari ............................................ 87
  9.3. Minimum system requirements .............................................. 87
  9.4. Improvements ........................................................................ 88
  9.5. Known Issues ......................................................................... 88
10. Release Notes HDP-1.0.1.14 .......................................................... 90
  10.1. Product Version: HDP-1.0.1.14 .................................................. 90
  10.2. Patch Information .................................................................. 90
    10.2.1. Patch information for Hadoop ............................................. 90
    10.2.2. Patch information for HBase .............................................. 91
    10.2.3. Patch information for HCatalog ......................................... 91
    10.2.4. Patch information for Hive .............................................. 91
10.2.5. Patch information for Oozie ............................................................... 91
10.2.6. Patch information for Sqoop .............................................................. 92
10.3. Minimum system requirements ............................................................... 92
10.4. Improvements .............................................................................................. 93
10.5. Known Issues ................................................................................................ 93

11. Release Notes HDP-1.0.0.12 ........................................................................ 95
11.1. Product Version: HDP-1.0.0.12 ................................................................. 95
11.2. Patch Information .......................................................................................... 95
   11.2.1. Patch information for Hadoop ........................................................... 95
   11.2.2. Patch information for HBase .............................................................. 96
   11.2.3. Patch information for HCatalog ......................................................... 96
   11.2.4. Patch information for Hive ................................................................. 96
   11.2.5. Patch information for Oozie ............................................................... 96
   11.2.6. Patch information for Sqoop .............................................................. 97
11.3. Minimum system requirements ............................................................... 97
11.4. Improvements .............................................................................................. 98
11.5. Known Issues ................................................................................................ 98
1. 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

1.1. 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 0.94.6.1
- 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 1.2.3
- 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

1.2. 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

1.2.1. 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.

1.2.2. 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.

1.2.3. Patch information for HBase

HBase is based on Apache HBase 0.94.6.1 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.

1.2.4. 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 `class 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).

1.2.5. Patch information for HCatalog

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

1.2.6. 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.

1.2.7. 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.

1.2.8. 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`.

### 1.2.9. 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.

### 1.2.10. 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.

### 1.2.11. 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**: `JMSource` 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.

1.3. 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).
1.3.1. Hardware Recommendations

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

1.3.2. 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

1.3.3. 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

1.3.4. 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.

1.3.5. 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.

1.3.6. 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.

1.4. 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.

1.5. Improvements

• Apache Hadoop updated to version 1.2.0
• Apache HBase updated to version 0.94.6.1.
• 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.

1.6. 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

1.6.1. 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.

1.6.2. Known Issues for Hive

• Hive Server 2 out of memory due to FileSystem.CACHE leak

Problem: Impersonation. By default, HiveServer2 performs query processing as the user who submitted the query. But if the following parameter is set to false, the query runs as the hiveserver2 process user.

hive.server2.enable.doAs – Impersonate the connected user, default true.

Workaround: To prevent memory leaks in unsecure mode, disable file system caches by setting the following parameters to true:

fs.hdfs.impl.disable.cache – Disable HDFS filesystem cache, default false.
fs.file.impl.disable.cache – Disable local filesystem cache, default false.

• 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:

```
```

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(158|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)
```
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.

• Problem: In Hive 0.9, setting hive.metastore.local = true in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of hive.metastore.uris. But in Hive 0.11, hive.metastore.local is ignored when hive.metastore.uris is set (https://issues.apache.org/jira/browse/HIVE-2585). When
upgrading from HDP 1.0 or HDP 1.1 to HDP 1.3.*, Hive is upgraded from 0.9 to 0.11. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

**Workaround:** To continue to use the embedded metastore after upgrading, clear the hive.metastore.uris setting in hive-site.xml.

### 1.6.3. 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: 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
  ```

### 1.6.4. Known Issues for HBase

- **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/27063b9e4e4bc64da6b3657b5f76ed5 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.namnode.SafeModeException: Cannot complete /apps/hbase/data/test_hbase/3bce795c2ad0713505f20ad3841bc3a2/.tmp/27063b9e4e4bc64da6b3657b5f76ed5. Name node is in safe mode.
  ```

### 1.6.5. 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/hor1n22.gq1.ygridcore.net@HORTON.YGRIDCORE.NET cause:org.apache.hadoop.security.AccessControlException: Client mapred tries to renew a token with renewer specified as jt
```

```
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
4f 52 54 8a 01 3e 41 b9 67 b8 8a 01 3e 65 c5 eb b8 8f 88
Kind: HDFS_DELEGATION_TOKEN, Service: 68.142.244.41:8020. Not rescheduled
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```

```
security.AccessControlException: Client mapred tries to renew a token with
renewer specified as jt
```
Workaround: Any new job 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.

1.6.6. 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.
2. Release Notes HDP-1.2.4

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
- Improvements
- Known Issues

2.1. Product Version: HDP-1.2.4

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

- Apache Hadoop 1.1.2
- Apache HBase 0.94.5
- Apache Pig 0.10.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog 0.5.0
- Apache Hive 0.10.0
- Apache Oozie 3.2.0
- Apache Sqoop 1.4.2
- Apache Ambari 1.2.3-rc0
- 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
  - Talend Open Studio 5.1.1
2.2. Patch Information

In this section:

- Patch information for Hadoop
- 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 Ambari

2.2.1. Patch information for Hadoop

In this section:

- Patch information for Hadoop (HDP-1.2.4.1)
- Patch information for Hadoop (HDP-1.2.4)

2.2.1.1. Patch information for Hadoop (HDP-1.2.4.1)

- MAPREDUCE-5256: Fixed thread-safe related issues for CombineInputFormat that impacted HiveServer.

2.2.1.2. Patch information for Hadoop (HDP-1.2.4)

Hadoop is based on Apache Hadoop 1.1.2 and includes the following additional patches:

- HDFS-4122: Reduced the size of log messages.
- HADOOP-8832: Added generic service plugin mechanism from HADOOP-5257 to branch-1.
- MAPREDUCE-461: Enabled service-plugins for JobTracker.
- MAPREDUCE-4838: Added locality, avataar, and workflow information to JobHistory.
- MAPREDUCE-4837: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.
- BUG FIXES:
2.2.2. Patch information for HBase

HBase is based on Apache HBase 0.94.5 and includes the following:

- **HBASE-6338**: Cache Method in RPC handler.
- **HBASE-6134**: Improved split-worker to enhance distributed log splitting.
- **HBASE-6508**: Added support to filter out edits at log split time (without breaking backward compatibility).
- **HBASE-7814**: Fixed hbck. hbck can now run on a secure cluster.
- **HBASE-7832**: Added support to use User.getShortName() in FSUtils.
• **HBASE-7851**: Fixed CNFE issues for a guava class.

• **HBASE-6466**: Enabled multi-thread for memstore flush.

• **HBASE-7820**: Added support for multi-realm authentication.

• **HBASE-7913**: Secure REST server should login before getting an instance of REST servlet.

• **HBASE-7915**: Secure ThriftServer needs to login before calling HBaseHandler.

• **HBASE-7920**: Removed isFamilyEssential(byte[] name) from Filter interface in HBase v.0.94.

• **HBASE-8007**: Added TestLoadAndVerify from BigTop.

• **HBASE-8179**: Fixed JSON formatting for cluster status.

### 2.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:

• **HIVE-3802**: Fixed test failures issues for testCliDriver_input39.

• **HIVE-3801**: Fixed test failures issues for testCliDriver_loadpart_err.

• **HIVE-3800**: Fixed test failures issues for testCliDriver_combine2.

• **HIVE-3792**: Fixed compile configurations for Hive pom.xml file.

• **HIVE-3788**: Fixed test failures issues for testCliDriver_repair.

• **HIVE-3782**: Fixed test failures issues for testCliDriver_sample_islocalmode_hook.

• **HIVE-3084**: Fixed build issues caused due to script_broken_pipel.q.

• **HIVE-3760**: Fixed test failures issues for TestNegativeMinimrCliDriver_mapreduce_stack_trace.q.

• **HIVE-3817**: Added namespace for Maven task to fix the deploy issues for the maven-publish target.

• **HIVE-2693**: Added DECIMAL datatype.

• **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.

• **HIVE-3255**: Added high availability support for Hive metastore. Added DBTokenStore to store Delegation Tokens in database.

• **HIVE-3291**: Fixed shims module compilation failures caused due to fs resolvers.

• **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.
• **HIVE-3862**: Added include exclude support to HBase handler.

• **HIVE-3861**: Upgraded HBase dependency to 0.94.2.

• **HIVE-3794**: Fixed Oracle upgrade script for Hive.

• **HIVE-3708**: Added MapReduce workflow information to job configuration.

### 2.2.4. Patch information for HCatalog

HCatalog is based on Apache HCatalog 0.5.0 and includes the following patches:

• **HCATALOG-563**: Improved HCatalog script. HCatalog script can now look in the correct directory for the storage handler JAR files.

### 2.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

• **PIG-3071**: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.

• **PIG-3099**: Pig unit test fixes for TestGrunt (1), TestStore (2), TestEmptyInputDir (3).

• **PIG-3116**: Fixed end to end tests sort command issues for RHEL-6.

• **PIG-3105**: Fixed `TestJobSubmission` unit test failure.

### 2.2.6. 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.

### 2.2.7. Patch information for Oozie

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

• **OOZIE-698**: Enhanced sharelib components.

• **OOZIE-810**: Fixed compilation issues for Oozie documentation.

• **OOZIE-863**: Fixed issues caused due to `JAVA_HOME` settings when `oozie-env.sh` script is invoked.

• **OOZIE-968**: Updated default location of Oozie environment file (`bin/oozie-env.sh`) to `conf/oozie-env.sh` in the `ooziedb.sh` file.

• **OOZIE-1006**: Fixed Hadoop 2.0.2 dependency issues for Oozie.
2.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

- **SQOOP-578**: Fixed issues with sqoop script calls.
- **SQOOP-579**: Improved reuse for custom manager factories.
- **SQOOP-580**: Added support for an open ended job teardown method which is invoked after the job execution.
- **SQOOP-582**: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.
- **SQOOP-462**: Fixed failures for Sqoop HBase test compilation.
- **SQOOP-741**: Enhance OracleConnect getTables() implementation in order to restrict tables to the current user.
- **SQOOP-798**: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.
- **SQOOP-846**: Added Netezza connector for Sqoop.
- **SQOOP-599**: Fixed import operation to HBase for secure cluster.

2.2.9. 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.

2.2.10. Patch information for Ambari

Ambari is based on Apache Ambari 1.2.3-rc0 and includes the following patches:

- **AMBARI-2024**: Fixed issue causing the Ambari Server to become non-responsive after crashing on the HTTP reads on Jersey.
- **AMBARI-1917**: Added support for default LZO properties in the core-site.xml file.
- **AMBARI-1815**: Fixed file corruption issues for core-site.xml file caused after modifying custom configs.
• **AMBARI-1794**: Fixed issues causing Add Host install retry operation to shut down all services in the cluster.

• **AMBARI-1795**: Fixed issues caused when Add Hosts - retrying install operation is performed.

For a complete list of changes visit the Apache Ambari JIRA [here](#).

### 2.3. 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 is **deprecated** as of HDP 1.2.0 and will not be made available in future minor and major releases of HDP. We encourage you to consider Manual Install (RPMs) or Automated Install (Ambari).

### 2.3.1. Hardware Recommendations

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

### 2.3.2. 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

• 64-bit Oracle Linux v5.*, v6.*

### 2.3.3. Software Requirements

On each of your hosts:

• yum
• rpm
• scp
• curl
• wget
• pdsh

2.3.4. Database Requirements

• Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

• Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

• Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

2.3.5. 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.

2.3.6. 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.

2.4. Improvements

In this section:
• Improvements for HDP-1.2.4.1
• Improvements for HDP-1.2.4

2.4.1. Improvements for HDP-1.2.4.1

• Fixed issues for HiveServer. See Patch information for Hadoop
• Updated Ambari upgrade guide. For more information, see Ambari upgrade guide.

2.4.2. Improvements for HDP-1.2.4

• Apache Ambari upgraded to version 1.2.3-rc0. For details, see Patch information for Ambari.

To upgrade Ambari server, follow the instructions provided here.

This version of Apache Ambari supports the following features:
• Added support for Oracle Linux.
• Added support for heterogenous Operating System installs.
• Added support for customizing the Ganglia service account.
• Added support for customizing the Hive Metastore log directory.
• Heatmaps now include HBase. For more information, see here.
• Improved Job Charts. For more information, see here.
• Apache HBase updated to version 0.94.5.
• Apache Flume updated to version 1.3.1.
• Hive upgraded to version 0.10.0.24. Use the following instructions to upgrade Hive:

  • Option I - Upgrade using Ambari:
    1. Stop Hive and WebHCat services using the Ambari UI.
    2. Execute the following commands on the Hive Server machine:

      • For RHEL:
        
        ```
        yum clean all
        yum update hive hcatalog webhcat-tar-hive
        ```

      • For SLES:
        
        ```
        zypper clean all
        zypper up -r Updates-HDP-1.2.1
        ```

      This command will upgrade the hive, hcatalog, and webhcat-tar-hive packages in your environment.
Important

For a multinode cluster, ensure that you also update the host machine where Hive client is installed. On the client machine, execute the following command:

- For RHEL/CentOS:
  ```
  yum update hive catalog
  ```
- For SLES:
  ```
  zypper up -r Updates-HDP-1.2.1
  ```

3. Start Hive and WebHCat services using the Ambari UI.

- Option II - Upgrade manually:
  1. Stop Hive and WebHCat services using the instructions provided here.
  2. Execute the following commands on the Hive Server machine:
     - For RHEL:
       ```
       yum clean all
       yum update hive hcatalog webhcat-tar-hive
       ```
     - For SLES:
       ```
       zypper clean all
       zypper up -r Updates-HDP-1.2.1
       ```

This command will upgrade the hive, hcatalog, and webhcat-tar-hive packages in your environment.

Important

For a multinode cluster, ensure that you also update the host machine where Hive client is installed. On the client machine, execute the following command:

- For RHEL/CentOS:
  ```
  yum update hive catalog
  ```
- For SLES:
  ```
  zypper up -r Updates-HDP-1.2.1
  ```

3. Start Hive and WebHCat services using the Ambari UI using the instructions provided here.

2.5. Known Issues

In this section:
• Known Issues for Hadoop
• Known Issues for Hive
• Known Issues for ZooKeeper
• Known Issues for Oozie
• Known Issues for Sqoop
• Known Issues for Ambari

2.5.1. Known Issues for Hadoop

• If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the CLASSPATH and the native Java libraries in the java.library.path.

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

2.5.2. Known Issues for Hive

• Hive create table operation fails when datanucleus.autoCreateSchema is set to true.

• Problem: In Hive 0.9, setting hive.metastore.local = true in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of hive.metastore.uris. But in Hive 0.10, hive.metastore.local is ignored when hive.metastore.uris is set (https://issues.apache.org/jira/browse/HIVE-2585). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

Workaround: To continue to use the embedded metastore after upgrading, clear the hive.metastore.uris setting in hive-site.xml.

2.5.3. Known Issues for ZooKeeper

• When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the Hosts and the Host Detail pages.

2.5.4. Known Issues for Oozie

• To be able to use Oozie command line client, you must first export JAVA_HOME.
2.5.5. Known Issues for Sqoop

- Sqoop command `list-all-tables` with Teradata connector returns views.
  
  This is caused because the `TeradataConnection listTables` query does not filter out tables alone when it queries the data dictionary.

  The workaround is to remove all views from the schema to use import all tables.

  Note that using import all tables has additional restrictions on the schema (tables cannot have multi-column primary keys etc.).

- Sqoop Teradata connector option `teradata.db.input.target.database` does not work.
  
  The Teradata Hadoop Connector used by Sqoop connector uses incorrect Hive database name while loading rows into Hive tables.

  The workaround is to use default Hive database for Hive imports.

- Sqoop import option `--split-by` is ignored when used with Teradata Sqoop connector.
  
  This issue is caused because the incorrect split table option is passed to the Hadoop connector.

  The workaround is to use the `teradata.db.input.split.by.column` property to specify split columns.

2.5.6. 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. Note that this is true if you chooses to deploy only TaskTracker on a given host.
3. Release Notes HDP-1.2.3.1

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
• Improvements
• Known Issues

3.1. Product Version: HDP-1.2.3.1

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

• Apache Hadoop 1.1.2
• Apache HBase 0.94.5
• Apache Pig 0.10.1
• Apache ZooKeeper 3.4.5
• Apache HCatalog 0.5.0
• Apache Hive 0.10.0
• Apache Oozie 3.2.0
• Apache Sqoop 1.4.2
• Apache Ambari 1.2.2
• 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
  • Talend Open Studio 5.1.1
3.2. Patch Information

In this section:

- Patch information for Hadoop
- 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

3.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.1.2 and includes the following additional patches:

- **HDFS-4122**: Reduced the size of log messages.

- **HADOOP-8832**: Added generic service plugin mechanism from HADOOP-5257 to branch-1.

- **MAPREDUCE-461**: Enabled service-plugins for JobTracker.

- **MAPREDUCE-4838**: Added locality, avataar, and workflow information to JobHistory.

- **MAPREDUCE-4837**: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.

- **BUG FIXES**:

  - **HDFS-4119**: Added slive to branch-1.
  
  - **HDFS-4180**: Updated TestFileCreation for HDFS-4122
  
  - **MAPREDUCE-4478**: Fixed issue with TaskTracker's heartbeat.
  
  - **HDFS-4108**: Fixed dfsnode
dlist to work in secure mode.

  - **HADOOP-8923**: Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.

  - **HADOOP-8164**: Added support to handle paths using backslash character as a path separator (for Windows platform only).
• **HADOOP-9051**: Fixed build failure issues for `ant test`.

• **HADOOP-9036**: Fixed racy test case `TestSinkQueue`.

• **MAPREDUCE-4916**: Fixed `TestTrackerDistributedCacheManager`.

• **HADOOP-7836**: Fixed failures for `TestSaslRPC#testDigestAuthMethodHostBasedToken` when hostname is set to `localhost.localdomain`.

• **HDFS-3402**: Fixed HDFS scripts for secure DataNodes.

• **HADOOP-9296**: Added support to allow users from different realm to authenticate without a trust relationship.

• **MAPREDUCE-4434**: Fixed `JobSplitWriter.java` to handle large `job.split` file.

• **HDFS-4222**: Fixed NameNode issue that caused the NameNode to become unresponsive and resulting in lost heartbeats from DNs when configured to use LDAP.

• **HDFS-3515**: Port HDFS-1457 to branch-1.

• **MAPREDUCE-4843**: Fixed `JobLocalizer when using DefaultTaskController`.

• **MAPREDUCE-2217**: Expire launching task now covers the UNASSIGNED task.

### 3.2.2. Patch information for HBase

HBase is based on Apache HBase 0.94.5 and includes the following:

• **HBASE-6338**: Cache Method in RPC handler.

• **HBASE-6134**: Improved split-worker to enhance distributed log splitting.

• **HBASE-6508**: Added support to filter out edits at log split time (without breaking backward compatibility).

• **HBASE-7814**: Fixed `hbck. hbck can now run on a secure cluster`.

• **HBASE-7832**: Added support to use `User.getShortName()` in `FSUtils`.

• **HBASE-7851**: Fixed CNFE issues for a guava class.

• **HBASE-6466**: Enabled multi-thread for memstore flush.

• **HBASE-7820**: Added support for multi-realm authentication.

• **HBASE-7913**: Secure REST server should login before getting an instance of REST servlet.

• **HBASE-7915**: Secure `ThriftServer needs to login before calling HBaseHandler`.

• **HBASE-7920**: Removed `isFamilyEssential(byte[] name)` from `Filter interface` in HBase v.0.94.

• **HBASE-8007**: Added `TestLoadAndVerify from BigTop`. 
• **HBASE-8179**: Fixed JSON formatting for cluster status.

### 3.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:

- **HIVE-3802**: Fixed test failures issues for `testCliDriver_input39`.
- **HIVE-3801**: Fixed test failures issues for `testCliDriver_loadpart_err`.
- **HIVE-3800**: Fixed test failures issues for `testCliDriver_combine2`.
- **HIVE-3792**: Fixed compile configurations for Hive `pom.xml` file.
- **HIVE-3788**: Fixed test failures issues for `testCliDriver_repair`.
- **HIVE-3782**: Fixed test failures issues for `testCliDriver_sample_islocalmode_hook`.
- **HIVE-3084**: Fixed build issues caused due to `script_broken_pipel.q`.
- **HIVE-3760**: Fixed test failures issues for `TestNegativeMinimrCliDriver_mapreduce_stack_trace.q`.
- **HIVE-3817**: Added namespace for Maven task to fix the deploy issues for the `maven-publish` target.
- **HIVE-2693**: Added `DECIMAL` datatype.
- **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.
- **HIVE-3255**: Added high availability support for Hive metastore. Added `DBTokenStore` to store Delegation Tokens in database.
- **HIVE-3291**: Fixed shims module compilation failures caused due to `fs resolvers`.
- **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.
- **HIVE-3862**: Added include exclude support to HBase handler.
- **HIVE-3861**: Upgraded HBase dependency to 0.94.2.
- **HIVE-3794**: Fixed Oracle upgrade script for Hive.
- **HIVE-3708**: Added MapReduce workflow information to job configuration.

### 3.2.4. Patch information for HCatalog

HCatalog is based on Apache HCatalog 0.5.0 and includes the following patches:

- **HCATALOG-563**: Improved HCatalog script. HCatalog script can now look in the correct directory for the storage handler JAR files.
3.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

- **PIG-3071**: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.

- **PIG-3099**: Pig unit test fixes for TestGrunt(1), TestStore(2), TestEmptyInputDir(3).

- **PIG-3116**: Fixed end to end tests sort command issues for RHEL-6.

- **PIG-3105**: Fixed TestJobSubmission unit test failure.

3.2.6. 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.

3.2.7. Patch information for Oozie

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

- **OOZIE-698**: Enhanced sharelib components.

- **OOZIE-810**: Fixed compilation issues for Oozie documentation.

- **OOZIE-863**: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

- **OOZIE-968**: Updated default location of Oozie environment file (bin/oozie-env.sh) to conf/oozie-env.sh in the ooziedb.sh file.

- **OOZIE-1006**: Fixed Hadoop 2.0.2 dependency issues for Oozie.

- **OOZIE-1048**: Added support to enable propagation of native libraries as a VM argument using java.library.path.

3.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

- **SQOOP-578**: Fixed issues with sqoop script calls.

- **SQOOP-579**: Improved reuse for custom manager factories.

- **SQOOP-580**: Added support for an open ended job teardown method which is invoked after the job execution.
• **SQOOP-582**: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.

• **SQOOP-462**: Fixed failures for Sqoop HBase test compilation.

• **SQOOP-741**: Enhanced `OracleConnect getTables()` implementation in order to restrict tables to the current user.

• **SQOOP-798**: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.

• **SQOOP-846**: Added Netezza connector for Sqoop.

• **SQOOP-599**: Fixed import operation to HBase for secure cluster.

### 3.2.9. 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. 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 is **deprecated** as of HDP 1.2.0 and will not be made available in future minor and major releases of HDP. We encourage you to consider Manual Install (RPMs) or Automated Install (Ambari).

### 3.3.1. Hardware Recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups [here](#).
3.3.2. 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

3.3.3. Software Requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh

3.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

- Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

- Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

3.3.5. 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.

3.3.6. 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.

### 3.4. Improvements

- Apache HBase updated to version 0.94.5.
- Apache Flume updated to version 1.3.1.
- Apache Ambari updated to version 1.2.2.4. This release (1.2.2.5) of Apache Ambari includes the new features and improvements:
  - Host level alerts
  - Paging controls on zoomed graphs
  - Ability to change Ambari Web HTTP port
  - Support for Active Directory-based authentication
  - **AMBARI-1757**: Add support for Stack 1.2.2 to Ambari
  - **AMBARI-1641**: Add support for additional TaskTracker metrics in API
  - **AMBARI-1748**: Custom JDK path added through UI now passed to global parameters

**Fixed issues in Ambari 1.2.2.5 from 1.2.1:**

To see a list of the issues that have been fixed that were noted in the release notes of the last release (1.2.2) of Apache Ambari, use the following query:

```
http://s.apache.org/release_notes_1.2.0_fixed
```

or click [here](http://s.apache.org/release_notes_1.2.0_fixed).

**All Issues fixed in Ambari 1.2.2:**

To see a list of all the issues that have been fixed for Ambari version 1.2.2.5, use the following query:

```
http://s.apache.org/all_issues_fixed_1.2.1
```

or click [here](http://s.apache.org/all_issues_fixed_1.2.1).
3.5. Known Issues

In this section:

• Known Issues for Hadoop

• Known Issues for Hive

• Known Issues for ZooKeeper

• Known Issues for Oozie

• Known Issues for Sqoop

• Known Issues for Ambari

3.5.1. Known Issues for Hadoop

• If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the CLASSPATH and the native Java libraries in the java.library.path.

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

3.5.2. Known Issues for Hive

• Hive create table operation fails when datanucleus.autoCreateSchema is set to true.

• Problem: In Hive 0.9, setting hive.metastore.local = true in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of hive.metastore.uris. But in Hive 0.10, hive.metastore.local is ignored when hive.metastore.uris is set (https://issues.apache.org/jira/browse/HIVE-2585). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

Workaround: To continue to use the embedded metastore after upgrading, clear the hive.metastore.uris setting in hive-site.xml.

3.5.3. Known Issues for ZooKeeper

• When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the Hosts and the Host Detail pages.
3.5.4. Known Issues for Oozie

- To be able to use Oozie command line client, you must first export JAVA_HOME.

3.5.5. Known Issues for Sqoop

- Sqoop command list-all-tables with Teradata connector returns views.
  
  This is caused because the TeradataConnection listTables query does not filter out tables alone when it queries the data dictionary.

  The workaround is to remove all views from the schema to use import all tables.

  Note that using import all tables has additional restrictions on the schema (tables cannot have multi-column primary keys etc.).

- Sqoop Teradata connector option teradata.db.input.target.database does not work.

  The Teradata Hadoop Connector used by Sqoop connector uses incorrect Hive database name while loading rows into Hive tables.

  The workaround is to use default Hive database for Hive imports.

- Sqoop import option --split-by is ignored when used with Teradata Sqoop connector.

  This issue is caused because the incorrect split table option is passed to the Hadoop connector.

  The workaround is to use the teradata.db.input.split.by.column property to specify split columns.

3.5.6. 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.

- To see a list of known open issues in Ambari 1.2.2.5, use the following query:

  http://s.apache.org/all_issues_fixed_1.2.2

  or click here.
4. Release Notes HDP-1.2.3

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
- Patch Information
- Minimum System Requirements
- Improvements
- Known Issues

4.1. Product Version: HDP-1.2.3

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

- Apache Hadoop 1.1.2
- Apache HBase 0.94.5
- Apache Pig 0.10.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog 0.5.0
- Apache Hive 0.10.0
- Apache Oozie 3.2.0
- Apache Sqoop 1.4.2
- Apache Ambari 1.2.2
- 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
  - Talend Open Studio 5.1.1
4.2. Patch Information

In this section:

- Patch information for Hadoop
- 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

4.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.1.2 and includes the following additional patches:

- **HDFS-4122**: Reduced the size of log messages.

- **HADOOP-8832**: Added generic service plugin mechanism from HADOOP-5257 to branch-1.

- **MAPREDUCE-461**: Enabled service-plugins for JobTracker.

- **MAPREDUCE-4838**: Added locality, avataar, and workflow information to JobHistory.

- **MAPREDUCE-4837**: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.

- **BUG FIXES**:

  - **HDFS-4219**: Added slive to branch-1.

  - **HDFS-4180**: Updated TestFileCreation for HDFS-4122

  - **MAPREDUCE-4478**: Fixed issue with TaskTracker's heartbeat.

  - **HDFS-4108**: Fixed dfsnodelist to work in secure mode.

  - **HADOOP-8923**: Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.

  - **HADOOP-8164**: Added support to handle paths using backslash character as a path separator (for Windows platform only).
• **HADOOP-9051**: Fixed build failure issues for ant test.

• **HADOOP-9036**: Fixed racy test case TestSinkQueue.

• **MAPREDUCE-4916**: Fixed TestTrackerDistributedCacheManager.

• **HADOOP-7836**: Fixed failures for TestSaslRPC#testDigestAuthMethodHostBasedToken when hostname is set to localhost.localdomain.

• **HDFS-3402**: Fixed HDFS scripts for secure DataNodes.

• **HADOOP-9296**: Added support to allow users from different realm to authenticate without a trust relationship.

• **MAPREDUCE-4434**: Fixed JobSplitWriter.java to handle large job.split file.

• **HDFS-4222**: Fixed NameNode issue that caused the NameNode to become unresponsive and resulting in lost heartbeats from DNs when configured to use LDAP.

• **HDFS-3515**: Port HDFS-1457 to branch-1.

• **MAPREDUCE-4843**: Fixed JobLocalizer when using DefaultTaskController.

• **MAPREDUCE-2217**: Expire launching task now covers the UNASSIGNED task.

### 4.2.2. Patch information for HBase

HBase is based on Apache HBase 0.94.5 and includes the following:

• **HBASE-6338**: Cache Method in RPC handler.

• **HBASE-6134**: Improved split-worker to enhance distributed log splitting.

• **HBASE-6508**: Added support to filter out edits at log split time (without breaking backward compatibility).

• **HBASE-7814**: Fixed hbck. hbck can now run on a secure cluster.

• **HBASE-7832**: Added support to use User.getShortName() in FSUtils.

• **HBASE-7851**: Fixed CNFE issues for a guava class.

• **HBASE-6466**: Enabled multi-thread for memstore flush.

• **HBASE-7820**: Added support for multi-realm authentication.

• **HBASE-7913**: Secure REST server should login before getting an instance of REST servlet.

• **HBASE-7915**: Secure ThriftServer needs to login before calling HBaseHandler.

• **HBASE-7920**: Removed isFamilyEssential(byte[] name) from Filter interface in HBase v.0.94.

• **HBASE-8007**: Added TestLoadAndVerify from BigTop.
• **HBASE-8179**: Fixed JSON formatting for cluster status.

### 4.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:

- **HIVE-3802**: Fixed test failures issues for `testCliDriver_input39`.
- **HIVE-3801**: Fixed test failures issues for `testCliDriver_loadpart_err`.
- **HIVE-3800**: Fixed test failures issues for `testCliDriver_combine2`.
- **HIVE-3792**: Fixed compile configurations for Hive `pom.xml` file.
- **HIVE-3788**: Fixed test failures issues for `testCliDriver_repair`.
- **HIVE-3782**: Fixed test failures issues for `testCliDriver_sample_islocalmode_hook`.
- **HIVE-3084**: Fixed build issues caused due to `script_broken_pipeline.q`.
- **HIVE-3760**: Fixed test failures issues for `TestNegativeMinimrCliDriver_mapreduce_stack_trace.q`.
- **HIVE-3817**: Added namespace for Maven task to fix the deploy issues for the `maven-publish` target.
- **HIVE-2693**: Added `DECIMAL` datatype.
- **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.
- **HIVE-3255**: Added high availability support for Hive metastore. Added `DBTokenStore` to store Delegation Tokens in database.
- **HIVE-3291**: Fixed shims module compilation failures caused due to `fs resolvers`.
- **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.
- **HIVE-3862**: Added include exclude support to HBase handler.
- **HIVE-3861**: Upgraded HBase dependency to 0.94.2.
- **HIVE-3794**: Fixed Oracle upgrade script for Hive.
- **HIVE-3708**: Added MapReduce workflow information to job configuration.

### 4.2.4. Patch information for HCatalog

HCatalog is based on Apache HCatalog 0.5.0 and includes the following patches:

- **HCATALOG-563**: Improved HCatalog script. HCatalog script can now look in the correct directory for the storage handler JAR files.
4.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

- PIG-3071: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.
- PIG-3099: Pig unit test fixes for TestGrunt(1), TestStore(2), TestEmptyInputDir(3).
- PIG-3116: Fixed end to end tests sort command issues for RHEL-6.
- PIG-3105: Fixed TestJobSubmission unit test failure.

4.2.6. 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.

4.2.7. Patch information for Oozie

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

- OOZIE-698: Enhanced sharelib components.
- OOZIE-810: Fixed compilation issues for Oozie documentation.
- OOZIE-863: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.
- OOZIE-968: Updated default location of Oozie environment file (bin/oozie-env.sh) to conf/oozie-env.sh in the ooziedb.sh file.
- OOZIE-1006: Fixed Hadoop 2.0.2 dependency issues for Oozie.
- OOZIE-1048: Added support to enable propagation of native libraries as a VM argument using java.library.path.

4.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

- SQOOP-578: Fixed issues with sqoop script calls.
- SQOOP-579: Improved reuse for custom manager factories.
- SQOOP-580: Added support for an open ended job teardown method which is invoked after the job execution.
• SQOOP-582: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.

• SQOOP-462: Fixed failures for Sqoop HBase test compilation.

• SQOOP-741: Enhanced OracleConnect getTables() implementation in order to restrict tables to the current user.

• SQOOP-798: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.

• SQOOP-846: Added Netezza connector for Sqoop.

• SQOOP-599: Fixed import operation to HBase for secure cluster.

4.2.9. 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. 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 is deprecated as of HDP 1.2.0 and will not be made available in future minor and major releases of HDP. We encourage you to consider Manual Install (RPMs) or Automated Install (Ambari).

4.3.1. Hardware Recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups here.
4.3.2. 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

4.3.3. Software Requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh

4.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

- Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

- Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

4.3.5. 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.

4.3.6. 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.

### 4.4. Improvements

- Apache HBase updated to version 0.94.5.
- Apache Flume updated to version 1.3.1.
- Apache Ambari updated to version 1.2.2.4. This release (1.2.2.5) of Apache Ambari includes the new features and improvements:
  - Host level alerts
  - Paging controls on zoomed graphs
  - Ability to change Ambari Web HTTP port
  - Support for Active Directory-based authentication
  - **AMBARI-1757**: Add support for Stack 1.2.2 to Ambari
  - **AMBARI-1641**: Add support for additional TaskTracker metrics in API
  - **AMBARI-1748**: Custom JDK path added through UI now passed to global parameters
- **Fixed issues in Ambari 1.2.2.4 from 1.2.1:**

  To see a list of the issues that have been fixed that were noted in the release notes of the last release (1.2.0) of Apache Ambari, use the following query:

  `http://s.apache.org/release_notes_1.2.0_fixed`

  or click [here](http://s.apache.org/release_notes_1.2.0_fixed).

- **All Issues fixed in Ambari 1.2.2:**

  To see a list of the all issues that have been fixed for Ambari version 1.2.2.4, use the following query:

  `http://s.apache.org/all_issues_fixed_1.2.1`

  or click [here](http://s.apache.org/all_issues_fixed_1.2.1).
4.5. Known Issues

In this section:

• Known Issues for Hadoop
• Known Issues for Hive
• Known Issues for ZooKeeper
• Known Issues for Oozie
• Known Issues for Sqoop
• Known Issues for Ambari

4.5.1. Known Issues for Hadoop

• If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the CLASSPATH and the native Java libraries in the java.library.path.

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

4.5.2. Known Issues for Hive

• Hive create table operation fails when datanucleus.autoCreateSchema is set to true.

• Problem: In Hive 0.9, setting hive.metastore.local = true in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of hive.metastore.uris. But in Hive 0.10, hive.metastore.local is ignored when hive.metastore.uris is set (https://issues.apache.org/jira/browse/HIVE-2585). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

Workaround: To continue to use the embedded metastore after upgrading, clear the hive.metastore.uris setting in hive-site.xml.

4.5.3. Known Issues for ZooKeeper

• When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the Hosts and the Host Detail pages.
4.5.4. Known Issues for Oozie

- To be able to use Oozie command line client, you must first export JAVA_HOME.

4.5.5. Known Issues for Sqoop

- Sqoop command `list-all-tables` with Teradata connector returns views.

  This is caused because the TeradataConnection `listTables` query does not filter out tables alone when it queries the data dictionary.

  The workaround is to remove all views from the schema to use `import all tables`.

  Note that using `import all tables` has additional restrictions on the schema (tables cannot have multi-column primary keys etc.).

- Sqoop Teradata connector option `teradata.db.input.target.database` does not work.

  The Teradata Hadoop Connector used by Sqoop connector uses incorrect Hive database name while loading rows into Hive tables.

  The workaround is to use default Hive database for Hive imports.

- Sqoop import option `--split-by` is ignored when used with Teradata Sqoop connector.

  This issue is caused because the incorrect split table option is passed to the Hadoop connector.

  The workaround is to use the `teradata.db.input.split.by.column` property to specify split columns.

4.5.6. Known Issues for Ambari

- To see a list of known open issues in Ambari 1.2.2.4, use the following query:

  [http://s.apache.org/all_issues_fixed_1.2.2](http://s.apache.org/all_issues_fixed_1.2.2)

  or click [here](http://s.apache.org/all_issues_fixed_1.2.2).
5. Release Notes HDP-1.2.2

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

5.1. Product Version: HDP-1.2.2

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

- Apache Hadoop 1.1.2-rc3
- Apache HBase 0.94.2+

**Note**

HBase is based on Apache SVN branch 0.94, revision 1406700, and additional patches as listed [here](#).

- Apache Pig 0.10.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog 0.5.0+

**Note**

HCatalog is based on Apache SVN branch 0.5.0, revision 1425288, and additional patches as listed [here](#).

- Apache Hive 0.10.0
- Apache Oozie 3.2.0
- Apache Sqoop 1.4.2
- Apache Ambari 1.2.2
- Apache Flume 1.3.0
- Apache Mahout 0.7.0
- Third party components:
  - Ganglia 3.2.0
  - GWeb 2.2.0
  - Nagios 3.2.3
  - Talend Open Studio 5.1.1
5.2. Patch Information

5.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.1.2-rc3 and includes the following additional patches:

- **HDFS-4122**: Reduced the size of log messages.
- **HADOOP-8832**: Added generic service plugin mechanism from HADOOP-5257 to branch-1.
- **MAPREDUCE-461**: Enabled service-plugins for JobTracker.
- **MAPREDUCE-4838**: Added locality, avaatar, and workflow information to JobHistory.
- **MAPREDUCE-4837**: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.
- **BUG FIXES**:
  - **HDFS-4219**: Added sliive to branch-1.
  - **HDFS-4180**: Updated \texttt{TestFileCreation} for HDFS-4122
  - **MAPREDUCE-4478**: Fixed issue with TaskTracker's heartbeat.
  - **HDFS-4108**: Fixed \texttt{dfsnodeList} to work in secure mode.
  - **HADOOP-8923**: Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
  - **HADOOP-8164**: Added support to handle paths using backslash character as a path separator (for Windows platform only).
  - **HADOOP-9051**: Fixed build failure issues for \texttt{ant test}.

5.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.94, revision 1406700, and includes the following:

- **HBASE-6338**: Cache Method in RPC handler.
- **HBASE-6134**: Improved split-worker to enhance distributed log splitting.
- **HBASE-7165**: Fixed test failure issues for \texttt{TestSplitLogManager.testUnassignedTimeout}.
- **HBASE-7166**: Fixed test failure issues for \texttt{TestSplitTransactionOnCluster}.
- **HBASE-7177**: Fixed test failure issues for \texttt{TestZooKeeperScanPolicyObserver.testScanPolicyObserver}.
• **HBASE-7235**: Fixed test failure issues for TestMasterObserver.

• **HBASE-7343**: Fixed issues for TestDrainingServer.

• **HBASE-6175**: Fixed issues for TestFSUtils on getFileStatus method in HDFS.

• **HBASE-7398**: Fixed test failures for TestAssignmentManager on CentOS.

### 5.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:

• **HIVE-3814**: Fixed issue for drop partitions operation when using Oracle metastore.

• **HIVE-3775**: Fixed unit test failures caused due to unspecified order of results in `show grant` command

• **HIVE-3815**: Fixed failures for `table rename` operation when filesystem cache is disabled.

• **HIVE-2084**: Upgraded datanucleus from v 2.0.3 to v 3.0.1.

• **HIVE-3802**: Fixed test failures issues for `testCliDriver_input39`.

• **HIVE-3801**: Fixed test failures issues for `testCliDriver_loadpart_err`.

• **HIVE-3800**: Fixed test failures issues for `testCliDriver_combine2`.

• **HIVE-3794**: Fixed Oracle upgrade script for Hive.

• **HIVE-3792**: Fixed compile configurations for Hive `pom.xml` file.

• **HIVE-3788**: Fixed test failures issues for `testCliDriver_repair`.

• **HIVE-3861**: Upgraded HBase dependency to 0.94.2.

• **HIVE-3782**: Fixed test failures issues for `testCliDriver_sample_islocalmode_hook`.

• **HIVE-3084**: Fixed build issues caused due to `script_broken_pipeline`.

• **HIVE-3760**: Fixed test failures issues for `TestNegativeMinimrCliDriver_mapreduce_stack_trace`.

• **HIVE-3717**: Fixed compilation issues caused when using `-Dhadoop.mr.rev=20S` property.

• **HIVE-3862**: Added include exclude support to HBase handler.

• **HIVE-3817**: Added namespace for Maven task to fix the deploy issues for the `maven-publish` target.

• **HIVE-2693**: Added `DECIMAL` datatype.

• **HIVE-3839**: Added support to set up `.git` attributes. This will normalize line endings during cross platform development.
• **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.

• **HIVE-3588**: Added support to use Hive with HBase v0.94.

• **HIVE-3255**: Added high availability support for Hive metastore. Added `DBTokenStore` to store Delegation Tokens in database.

• **HIVE-3708**: Added MapReduce workflow information to job configuration.

• **HIVE-3291**: Fixed shims module compilation failures caused due to `fs` resolvers.

• **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.

• **HIVE-3846**: Fixed Null pointer Exceptions (NPEs) issues caused while executing the `alter view rename` command when authorization is enabled.

### 5.2.4. Patch information for HCatalog

HCatalog is based on Apache SVN branch 0.5.0, revision 1425288 and includes the following patches:

• **HCATALOG-549**: Added changes for WebHCat.

• **HCATALOG-509**: Added support for WebHCat to work with security.

• **HCATALOG-587**: Fixed memory consumption issues for WebHCat Controller Map Task.

• **HCATALOG-588**: Fixed issues with `templeton.log` file for WebHCat server.

• **HCATALOG-589**: Improved dependent library packaging.

• **HCATALOG-577**: Fixed issues for HCatContext that caused persistance of undesired `jobConf` parameters.

• **HCATALOG-584**: Changes in **HCATALOG-538** breaks Pig stores into non-partitioned tables.

• **HCATALOG-580**: Fixed end to end test failures caused by optimizations in **HCATALOG-538**.

• **HCATALOG-583**: Fixed end to end test failures.

• **HCATALOG-590**: Moved `DEFAULT_DATABASE_NAME` constant to `HiveConf` as part of **HIVE-2935**: **HIVE-2935.3.patch.gz**.

• **HCATALOG-573**: Removed version number from `WEBHCAT_JAR` in the `webhcac_config.sh` file.

• **HCATALOG-592**: Improved error message for Hive table/partition not found in order to enable WebHCat to return correct HTTP status code.

• **HCATALOG-583**: Fixed end to end test failures.
5.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

- **PIG-3071**: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.

- **PIG-3099**: Pig unit test fixes for TestGrunt (1), TestStore (2), TestEmptyInputDir (3).

- **PIG-2885**: Fixed test failures for TestJobSubmission and TestHBaseStorage.

- **PIG-3105**: Fixed TestJobSubmission unit test failure.

5.2.6. Patch information for ZooKeeper

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

- **ZOOKEEPER-1598**: Enhanced ZooKeeper version string.

5.2.7. Patch information for Oozie

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

- **OOZIE-698**: Enhanced sharelib components.

- **OOZIE-810**: Fixed compilation issues for Oozie documentation.

- **OOZIE-863**: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

- **OOZIE-968**: Updated default location of Oozie environment file (bin/oozie-env.sh) to conf/oozie-env.sh in the ooziedb.sh file.

- **OOZIE-1006**: Fixed Hadoop 2.0.2 dependency issues for Oozie.

- **OOZIE-1048**: Added support to enable propagation of native libraries as a VM argument using java.library.path.

5.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

- **SQOOP-438**: Added support to allow sourcing of sqoop-env.sh file. This enhancement now allows setting variables directly in the configuration files.

- **SQOOP-462**: Fixed failures for Sqoop HBase test compilation.

- **SQOOP-578**: Fixed issues with sqoop script calls.

- **SQOOP-579**: Improved reuse for custom manager factories.

- **SQOOP-580**: Added support for an open ended job teardown method which is invoked after the job execution.
• **SQOOP-582**: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.

• **SQOOP-741**: Enhanced `OracleConnect.getTables()` implementation in order to restrict tables to the current user.

• **SQOOP-798**: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.

### 5.2.9. Patch information for Mahout

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

• **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.

### 5.3. 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 is **deprecated** as of HDP 1.2.0 and will not be made available in future minor and major releases of HDP. 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

5.3.3. Software Requirements

On each of your hosts:
• yum
• rpm
• scp
• curl
• wget
• pdsh

5.3.4. Database Requirements

• Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

• Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

• Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

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. Improvements

- Apache Ambari updated to version 1.2.2. See *Apache Ambari release notes* for more details.

### 5.5. Known Issues

- Illustrate does not work when using HCatalog loading in Pig scripts.

- **Hive** *create table* operation fails when *datanucleus.autoCreateSchema* is set to *true*.

- When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the **Hosts** and the **Host Detail** pages.

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

- To be able to use Oozie command line client, you must first export *JAVA_HOME*.

- Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.

- If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the classpath and the native Java libraries in the java.library.path.

- **Problem:** In Hive 0.9, setting *hive.metastore.local = true* in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of *hive.metastore.uris*. But in Hive 0.10, *hive.metastore.local* is ignored when *hive.metastore.uris* is set ([https://issues.apache.org/jira/browse/HIVE-2585](https://issues.apache.org/jira/browse/HIVE-2585)). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

  **Workaround:** To continue to use the embedded metastore after upgrading, clear the *hive.metastore.uris* setting in hive-site.xml.
6. Release Notes HDP-1.2.1

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

6.1. Product Version: HDP-1.2.1

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

- Apache Hadoop 1.1.2-rc3
- Apache HBase 0.94.2+

  **Note**
  
  HBase is based on Apache SVN branch 0.94, revision 1406700, and additional patches as listed [here](#).

- Apache Pig 0.10.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog 0.5.0+

  **Note**
  
  HCatalog is based on Apache SVN branch 0.5.0, revision 1425288, and additional patches as listed [here](#).

- Apache Hive 0.10.0
- Apache Oozie 3.2.0
- Apache Sqoop 1.4.2
- Apache Ambari 1.2.1
- Apache Flume 1.3.0
- Apache Mahout 0.7.0
- Third party components:
  
  - Ganglia 3.2.0
  - GWeb 2.2.0
  - Nagios 3.2.3
  - Talend Open Studio 5.1.1
6.2. Patch Information

6.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.1.2-rc3 and includes the following additional patches:

- **HDFS-4122**: Reduced the size of log messages.
- **HADOOP-8832**: Added generic service plugin mechanism from HADOOP-5257 to branch-1.
- **MAPREDUCE-461**: Enabled service-plugins for JobTracker.
- **MAPREDUCE-4838**: Added locality, avataar, and workflow information to JobHistory.
- **MAPREDUCE-4837**: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.
- **BUG FIXES**:
  - **HDFS-4219**: Added slive to branch-1.
  - **HDFS-4180**: Updated `TestFileCreation` for HDFS-4122
  - **MAPREDUCE-4478**: Fixed issue with TaskTracker's heartbeat.
  - **HDFS-4108**: Fixed `dfsnodeList` to work in secure mode.
  - **HADOOP-8923**: Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
  - **HADOOP-8164**: Added support to handle paths using backslash character as a path separator (for Windows platform only).
  - **HADOOP-9051**: Fixed build failure issues for ant test.

6.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.94, revision 1406700, and includes the following:

- **HBASE-6338**: Cache Method in RPC handler.
- **HBASE-6134**: Improved split-worker to enhance distributed log splitting.
- **HBASE-7165**: Fixed test failure issues for `TestSplitLogManager.testUnassignedTimeout`.
- **HBASE-7166**: Fixed test failure issues for `TestSplitTransactionOnCluster`.
- **HBASE-7177**: Fixed test failure issues for `TestZooKeeperScanPolicyObserver.testScanPolicyObserver`. 
• HBASE-7235: Fixed test failure issues for TestMasterObserver.
• HBASE-7343: Fixed issues for TestDrainingServer.
• HBASE-6175: Fixed issues for TestFSUtils on getFileStatus method in HDFS.
• HBASE-7398: Fixed test failures for TestAssignmentManager on CentOS.

6.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:
• HIVE-3814: Fixed issue for drop partitions operation when using Oracle metastore.
• HIVE-3775: Fixed unit test failures caused due to unspecified order of results in show grant command
• HIVE-3815: Fixed failures for table rename operation when filesystem cache is disabled.
• HIVE-2084: Upgraded datanucleus from v 2.0.3 to v 3.0.1.
• HIVE-3802: Fixed test failures issues for testCliDriver_input39.
• HIVE-3801: Fixed test failures issues for testCliDriver_loadpart_err.
• HIVE-3800: Fixed test failures issues for testCliDriver_combine2.
• HIVE-3794: Fixed Oracle upgrade script for Hive.
• HIVE-3792: Fixed compile configurations for Hive pom.xml file.
• HIVE-3788: Fixed test failures issues for testCliDriver_repair.
• HIVE-3861: Upgraded HBase dependency to 0.94.2.
• HIVE-3782: Fixed test failures issues for testCliDriver_sample_islocalmode_hook.
• HIVE-3084: Fixed build issues caused due to script_broken_pipel.q.
• HIVE-3760: Fixed test failures issues for TestNegativeMinimrCliDriver_mapreduce_stack_trace.q.
• HIVE-3717: Fixed compilation issues caused when using -Dhadoop.mr.rev=20S property.
• HIVE-3862: Added include exclude support to HBase handler.
• HIVE-3817: Added namespace for Maven task to fix the deploy issues for the maven-publish target.
• HIVE-2693: Added DECIMAL datatype.
• HIVE-3839: Added support to set up .git attributes. This will normalize line endings during cross platform development.
• **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.

• **HIVE-3588**: Added support to use Hive with HBase v0.94.

• **HIVE-3255**: Added high availability support for Hive metastore. Added `DBTokenStore` to store Delegation Tokens in database.

• **HIVE-3708**: Added MapReduce workflow information to job configuration.

• **HIVE-3291**: Fixed shims module compilation failures caused due to `fs` resolvers.

• **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.

• **HIVE-3846**: Fixed Null pointer Exceptions (NPEs) issues caused while executing the `alter view rename` command when authorization is enabled.

### 6.2.4. Patch information for HCatalog

HCatalog is based on Apache SVN branch 0.5.0, revision 1425288 and includes the following patches:

• **HCATALOG-549**: Added changes for WebHCat.

• **HCATALOG-509**: Added support for WebHCat to work with security.

• **HCATALOG-587**: Fixed memory consumption issues for WebHCat Controller Map Task.

• **HCATALOG-588**: Fixed issues with `templeton.log` file for WebHCat server.

• **HCATALOG-589**: Improved dependent library packaging.

• **HCATALOG-577**: Fixed issues for `HCatContext` that caused persistence of undesired `jobConf` parameters.

• **HCATALOG-584**: Changes in **HCATALOG-538** breaks Pig stores into non-partitioned tables.

• **HCATALOG-580**: Fixed end to end test failures caused by optimizations in **HCATALOG-538**.

• **HCATALOG-583**: Fixed end to end test failures.

• **HCATALOG-590**: Moved `DEFAULT_DATABASE_NAME` constant to `HiveConf` as part of **HIVE-2935**: `HIVE-2935.3.patch.gz`.

• **HCATALOG-573**: Removed version number from `WEBHCAT_JAR` in the `webhcat_config.sh` file.

• **HCATALOG-592**: Improved error message for Hive table/partition not found in order to enable WebHCat to return correct HTTP status code.

• **HCATALOG-583**: Fixed end to end test failures.
6.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

- **PIG-3071**: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.
- **PIG-3099**: Pig unit test fixes for TestGrunt (1), TestStore (2), TestEmptyInputDir (3).
- **PIG-2885**: Fixed test failures for TestJobSumission and TestHBaseStorage.
- **PIG-3105**: Fixed TestJobSubmission unit test failure.

6.2.6. Patch information for ZooKeeper

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

- **ZOOKEEPER-1598**: Enhanced ZooKeeper version string.

6.2.7. Patch information for Oozie

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

- **OOZIE-698**: Enhanced sharelib components.
- **OOZIE-810**: Fixed compilation issues for Oozie documentation.
- **OOZIE-863**: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.
- **OOZIE-968**: Updated default location of Oozie environment file (bin/oozie-env.sh) to conf/oozie-env.sh in the ooziedb.sh file.
- **OOZIE-1006**: Fixed Hadoop 2.0.2 dependency issues for Oozie.
- **OOZIE-1048**: Added support to enable propagation of native libraries as a VM argument using java.library.path.

6.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

- **SQOOP-438**: Added support to allow sourcing of sqoop-env.sh file. This enhancement now allows setting variables directly in the configuration files.
- **SQOOP-462**: Fixed failures for Sqoop HBase test compilation.
- **SQOOP-578**: Fixed issues with sqoop script calls.
- **SQOOP-579**: Improved reuse for custom manager factories.
- **SQOOP-580**: Added support for an open ended job teardown method which is invoked after the job execution.
• **SQOOP-582**: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.

• **SQOOP-741**: Enhanced OracleConnect `getTables()` implementation in order to restrict tables to the current user.

• **SQOOP-798**: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.

### 6.2.9. Patch information for Mahout

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

• **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.

### 6.3. Minimum system requirements

**Hardware Recommendations:**

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

**Operating Systems Requirements:**

The following operating systems 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

**Software Requirements:**

On each of your hosts:

• `yum`

• `rpm`

• `scp`

• `curl`

• `wget`

• `pdsh`

**Database Requirements:**
• Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

• Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

• Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

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.

Important

gsInstaller is deprecated as of HDP 1.2.1 and will not be made available in future minor and major releases of HDP. We encourage you to consider Manual Install (RPMs) or Automated Install (Ambari).

6.4. Improvements

• Apache Ambari updated to version 1.2.1. See Apache Ambari release notes for more details.

6.5. Known Issues

• Illustrate does not work when using HCatalog loading in Pig scripts.

• Hive create table operation fails when datanucleus.autoCreateSchema is set to true.

• When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the Hosts and the
**Host Detail**

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

- To be able to use Oozie command line client, you must first export `JAVA_HOME`.

- Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.

- If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the classpath and the native Java libraries in the java.library.path.

- **Problem:** In Hive 0.9, setting `hive.metastore.local = true` in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of `hive.metastore.uris`. But in Hive 0.10, `hive.metastore.local` is ignored when `hive.metastore.uris` is set (`https://issues.apache.org/jira/browse/HIVE-2585`). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

  **Workaround:** To continue to use the embedded metastore after upgrading, clear the `hive.metastore.uris` setting in hive-site.xml.
7. Release Notes HDP-1.2.0

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

7.1. Product Version: HDP-1.2.0

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

- Apache Hadoop 1.1.2-rc3
- Apache HBase 0.94.2+

**Note**

HBase is based on Apache SVN branch 0.94, revision 1406700, and additional patches as listed here.

- Apache Pig 0.10.1
- Apache ZooKeeper 3.4.5
- Apache HCatalog 0.5.0+

**Note**

HCatalog is based on Apache SVN branch 0.5.0, revision 1425288, and additional patches as listed here.

- Apache Hive 0.10.0
- Apache Oozie 3.2.0
- Apache Sqoop 1.4.2
- Apache Ambari 1.2.0
- Apache Flume 1.3.0
- Apache Mahout 0.7.0
- Third party components:
  - Ganglia 3.2.0
  - GWeb 2.2.0
  - Nagios 3.2.3
  - Talend Open Studio 5.1.1
7.2. Patch Information

7.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.1.2-rc3 and includes the following additional patches:

- **HDFS-4122**: Reduced the size of log messages.
- **HADOOP-8832**: Added generic service plugin mechanism from HADOOP-5257 to branch-1.
- **MAPREDUCE-461**: Enabled service-plugins for JobTracker.
- **MAPREDUCE-4838**: Added locality, avatar, and workflow information to JobHistory.
- **MAPREDUCE-4837**: Added web-service APIs for JobTracker. These APIs can be used to get information on jobs and component tasks.
- **BUG FIXES**:
  - **HDFS-4219**: Added slive to branch-1.
  - **HDFS-4180**: Updated `TestFileCreation` for HDFS-4122
  - **MAPREDUCE-4478**: Fixed issue with TaskTracker's heartbeat.
  - **HDFS-4108**: Fixed `dfsnodelist` to work in secure mode.
  - **HADOOP-8923**: Fixed incorrect rendering of the intermediate web user interface page caused when the authentication cookie (SPENGO/custom) expires.
  - **HADOOP-8164**: Added support to handle paths using backslash character as a path separator (for Windows platform only).
  - **HADOOP-9051**: Fixed build failure issues for `ant test`.

7.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.94, revision 1406700, and includes the following:

- **HBASE-6338**: Cache Method in RPC handler.
- **HBASE-6134**: Improved split-worker to enhance distributed log splitting.
- **HBASE-7165**: Fixed test failure issues for `TestSplitLogManager.testUnassignedTimeout`.
- **HBASE-7166**: Fixed test failure issues for `TestSplitTransactionOnCluster`.
- **HBASE-7177**: Fixed test failure issues for `TestZooKeeperScanPolicyObserver.testScanPolicyObserver`.
• **HBASE-7235**: Fixed test failure issues for TestMasterObserver.
• **HBASE-7343**: Fixed issues for TestDrainingServer.
• **HBASE-6175**: Fixed issues for TestFSUtils on getFileStatus method in HDFS.
• **HBASE-7398**: Fixed test failures for TestAssignmentManager on CentOS.

### 7.2.3. Patch information for Hive

Hive is based on Apache Hive 0.10.0 and includes the following patches:

• **HIVE-3814**: Fixed issue for drop partitions operation when using Oracle metastore.
• **HIVE-3775**: Fixed unit test failures caused due to unspecified order of results in `show grant` command.
• **HIVE-3815**: Fixed failures for `table rename` operation when filesystem cache is disabled.
• **HIVE-2084**: Upgraded datanucleus from v 2.0.3 to v 3.0.1.
• **HIVE-3802**: Fixed test failures issues for `testCliDriver_input39`.
• **HIVE-3801**: Fixed test failures issues for `testCliDriver_loadpart_err`.
• **HIVE-3800**: Fixed test failures issues for `testCliDriver_combine2`.
• **HIVE-3794**: Fixed Oracle upgrade script for Hive.
• **HIVE-3792**: Fixed compile configurations for Hive `pom.xml` file.
• **HIVE-3788**: Fixed test failures issues for `testCliDriver_repair`.
• **HIVE-3861**: Upgraded HBase dependency to 0.94.2.
• **HIVE-3782**: Fixed test failures issues for `testCliDriver_sample_islocalmode_hook`.
• **HIVE-3084**: Fixed build issues caused due to `script_broken_pip1.q`.
• **HIVE-3760**: Fixed test failures issues for `TestNegativeMinimr CliDriver_mapreduce_stack_trace.q`.
• **HIVE-3717**: Fixed compilation issues caused when using `-Dhadoop.mr.rev=20S` property.
• **HIVE-3862**: Added include exclude support to HBase handler.
• **HIVE-3817**: Added namespace for Maven task to fix the deploy issues for the `maven-publish` target.
• **HIVE-2693**: Added DECIMAL datatype.
• **HIVE-3839**: Added support to set up `.git` attributes. This will normalize line endings during cross platform development.
• **HIVE-3678**: Added metastore upgrade scripts for column statistics schema changes for Postgres/MySQL/Oracle/Derby.

• **HIVE-3588**: Added support to use Hive with HBase v0.94.

• **HIVE-3255**: Added high availability support for Hive metastore. Added DBTokenStore to store Delegation Tokens in database.

• **HIVE-3708**: Added MapReduce workflow information to job configuration.

• **HIVE-3291**: Fixed shims module compilation failures caused due to fs resolvers.

• **HIVE-2935**: Implemented HiveServer2 (Hive Server 2). Added JDBC/ODBC support over HiveServer2.

• **HIVE-3846**: Fixed Null pointer Exceptions (NPEs) issues caused while executing the `alter view rename` command when authorization is enabled.

### 7.2.4. Patch information for HCatalog

HCatalog is based on Apache SVN branch 0.5.0, revision 1425288 and includes the following patches:

• **HCATALOG-549**: Added changes for WebHCat.

• **HCATALOG-509**: Added support for WebHCat to work with security.

• **HCATALOG-587**: Fixed memory consumption issues for WebHCat Controller Map Task.

• **HCATALOG-588**: Fixed issues with `templeton.log` file for WebHCat server.

• **HCATALOG-589**: Improved dependent library packaging.

• **HCATALOG-577**: Fixed issues for `HCatContext` that caused persistance of undesired `jobConf` parameters.

• **HCATALOG-584**: Changes in `HCATALOG-538` breaks Pig stores into non-partitioned tables.

• **HCATALOG-580**: Fixed end to end test failures caused by optimizations in `HCATALOG-538`.

• **HCATALOG-583**: Fixed end to end test failures.

• **HCATALOG-590**: Moved `DEFAULT_DATABASE_NAME` constant to `HiveConf` as part of HIVE-2935: HIVE-2935.3.patch.gz.

• **HCATALOG-573**: Removed version number from `WEBHCAT_JAR` in the `webhcat_config.sh` file.

• **HCATALOG-592**: Improved error message for Hive table/partition not found in order to enable WebHCat to return correct HTTP status code.

• **HCATALOG-583**: Fixed end to end test failures.
7.2.5. Patch information for Pig

Pig is based on Apache Pig 0.10.1 and includes the following patches:

• **PIG-3071**: Updated Pig script file. The script file now has modified HCatalog JAR file and PATH that points to HBase storage handler JAR file.

• **PIG-3099**: Pig unit test fixes for TestGrunt(1), TestStore(2), TestEmptyInputDir(3).

• **PIG-2885**: Fixed test failures for TestJobSumission and TestHBaseStorage.

• **PIG-3105**: Fixed TestJobSubmission unit test failure.

7.2.6. Patch information for ZooKeeper

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

• **ZOOKEEPER-1598**: Enhanced ZooKeeper version string.

7.2.7. Patch information for Oozie

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

• **OOZIE-698**: Enhanced sharelib components.

• **OOZIE-810**: Fixed compilation issues for Oozie documentation.

• **OOZIE-863**: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

• **OOZIE-968**: Updated default location of Oozie environment file (bin/oozie-env.sh) to conf/oozie-env.sh in the ooziedb.sh file.

• **OOZIE-1006**: Fixed Hadoop 2.0.2 dependency issues for Oozie.

• **OOZIE-1048**: Added support to enable propagation of native libraries as a VM argument using java.library.path.

7.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.2 and includes the following patches:

• **SQOOP-438**: Added support to allow sourcing of sqoop-env.sh file. This enhancement now allows setting variables directly in the configuration files.

• **SQOOP-462**: Fixed failures for Sqoop HBase test compilation.

• **SQOOP-578**: Fixed issues with sqoop script calls.

• **SQOOP-579**: Improved reuse for custom manager factories.

• **SQOOP-580**: Added support for an open ended job teardown method which is invoked after the job execution.
• **SQOOP-582**: Added a template method for job submission in Export/Import JobBase. A connector can now submit a job and also complete other tasks simultaneously while the on-going job is in progress.

• **SQOOP-741**: Enhanced `OracleConnect getTables()` implementation in order to restrict tables to the current user.

• **SQOOP-798**: Fixed issue for ANT docs for RedHat Enterprise Linux (RHEL) v5.8.

### 7.2.9. Patch information for Ambari

Ambari is based on Apache Ambari 1.2.0 and includes the following patches:

• For a complete list of changes, visit the Apache Ambari JIRA [here](#).

### 7.2.10. Patch information for Mahout

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

• **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.

### 7.3. Minimum system requirements

**Hardware Recommendations:**

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

**Operating Systems Requirements:**

The following operating systems 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

**Software Requirements:**

On each of your hosts:

• `yum`

• `rpm`

• `scp`

• `curl`
• wget
• pdsh

Database Requirements:

• Hive and HCatalog require a database to use as a metadata store. MySQL 5.x or Oracle 11gr2 are supported. You may provide access to an existing database, or the Ambari and gsInstaller installers will install MySQL for you if you want.

• Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x or Oracle 11gr2 are also supported.

• Ambari requires a database to use as a metadata store, but comes with Postgres 8.x. This is the only database supported in this version.

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.

Important

gsInstaller is deprecated as of HDP 1.2.0 and will not be made available in future minor and major releases of HDP. We encourage you to consider Manual Install (RPMs) or Automated Install (Ambari).

7.4. Improvements

• Fixed incorrect host mappings for Hive causing failure of Hive smoke tests.

• Hadoop updated to version 1.1.2.

• HBase updated to version 0.94.2.

• Pig updated to version 0.10.1.

• ZooKeeper updated to version 3.4.5.

• HCatalog updated to version 0.5.0.
• Hive updated to version 0.10.0.
• Oozie updated to upstream version 3.2.0.
• Added support for Apache Mahout.
• Talend Open Studio updated to upstream version 5.1.2.
• Ambari updated to version 1.2.0.
• Added WebHCat functionality to HCatalog.
• Added authentication support for Ambari Web, with default local user auth provider.
• Added more options on cluster provisioning, including ability to set more than one ZooKeeper server, specify installation of DataNode, TaskTracker and Client components, and customize service user accounts.
• Enhanced Ambari web user interface. For a complete list of changes visit the Apache Ambari JIRA here.
• Added ability to visualize cluster heatmaps in Ambari. Cluster heatmaps provide a unified view of key metrics for all the nodes in your cluster.
• Added ability to view information on individual cluster hosts using Ambari web user interface.
• Added ability to run service smoke tests from the Ambari management console.
• Added ability to use RESTful APIs for cluster metrics in Ambari.
• Added support for SLES 11 SP1 (64-bit) and added browser support for IE9, Chrome, and Safari (refer to the product documentation for full list of supported platforms).
• Added support for SQL Server and Oracle.

7.5. Known Issues

• Illustrate does not work when using HCatalog loading in Pig scripts.
• Hive create table operation fails when datanucleus.autoCreateSchema is set to true.
• When at least one ZooKeeper Server becomes non-responsive, the host status for the other hosts with the ZooKeeper Servers may be displayed incorrectly on the

<table>
<thead>
<tr>
<th>Hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>and the</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>pages.</td>
</tr>
</tbody>
</table>
• Use of init.d scripts for starting or stopping Hadoop services, is not recommended.
• To be able to use Oozie command line client, you must first export `JAVA_HOME`.

• Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.

• If you are using Talend 5.1.1, you need to include the new hadoop-core.jar, hadoop-lzo.jar and /etc/hadoop/conf in the classpath and the native Java libraries in the java.library.path.

• **Problem:** In Hive 0.9, setting `hive.metastore.local = true` in hive-site.xml meant that the embedded metastore would ALWAYS be used regardless of the setting of `hive.metastore.uris`. But in Hive 0.10, `hive.metastore.local` is ignored when `hive.metastore.uris` is set (https://issues.apache.org/jira/browse/HIVE-2585). When upgrading from HDP 1.0 or HDP 1.1 to HDP 1.2.*, Hive is upgraded from 0.9 to 0.10. Therefore, the embedded metastore may no longer be used after upgrading without adjusting the hive-site.xml settings.

  **Workaround:** To continue to use the embedded metastore after upgrading, clear the `hive.metastore.uris` setting in hive-site.xml.
8. Release Notes HDP-1.1.1.16

Hortonworks Data Platform with Hortonworks Management Console powered by Apache Hadoop

8.1. Product Version: HDP-1.1.1.16

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

- Apache Hadoop 1.0.3
- Apache HBase 0.92.1+
- Apache Pig 0.9.2
- Apache ZooKeeper 3.3.4
- Apache HCatalog 0.4.0
- Apache Hive 0.9.0
- Templeton 0.1.4
- Apache Oozie 3.1.3
- Apache Sqoop 1.4.2
- Hortonworks Management Center (HMC) 1.0.2
- Apache Flume 1.2.0
- HA-Monitor 0.1.1

Third party components:

- Ganglia 3.2.0
- Nagios 3.2.3
- Talend Open Studio 5.1.1

8.2. Patch Information

8.2.1. Patch information for Hadoop

Hadoop is patched to include the following:


• Bug Fixes: HDFS-3846 and MAPREDUCE-4558.

• HADOOP-5464: Added support for disabling write timeout. To do this, you can set zero values for dfs.socket.timeout and dfs.datanode.socket.write.timeout parameters.

• HDFS-2617: Replaced Kerberized SSL for image transfer and fsck with SPNEGO-based solution.

• HDFS-3466: Fixed SPNEGO filter to use the DFS_WEB_AUTHENTICATION_KERBEROS_KEYTAB_KEY to find the keytab.

• HDFS-3461: Fixed HFTP to use the same port and protocol while obtaining the delegation token.

• HADOOP-6947: Fixed Kerberos relogin to configure the refreshKrb5Config correctly.

• MAPREDUCE-3837: Enhanced JobTracker job recovery mechanism in the event of a crash.

• HDFS-3652: Fixed edit stream issues caused in the event of FSEditlog failure.

• MAPREDUCE-4399: Fixed performance regression in shuffle.

• HADOOP-7154: Added support to set MALLOC_ARENA_MAX in hadoop-config.sh file.

• HDFS-3652: Fixed edit stream issues caused in the event of FSEditlog failure.

• MAPREDUCE-4399: Fixed performance regression in shuffle.

• HADOOP-7154: Added support to set MALLOC_ARENA_MAX in hadoop-config.sh file.

8.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.92, revision 1344056 and includes the following:

• HBASE-6447: Fixed issues with ZooKeeper test failures.

• HBASE-6334: Improvements for RegionServer tests.

• HBASE-4470: Fixed issues related to ServerNotRunning exception with HBase master.

• HBASE-6460: Fixed issues with hbck -repairHoles command.

• HBASE-6552: Fixed issues related to TestAcidGuarantees system tests.

• HBASE-6512: Fixed issues related to incorrect log name for OfflineMetaRepair.

• HBASE-6308: Improved coprocessors to prevent dependency conflicts with HBase.

• HBASE-6576: Fixed issues with HBaseAdmin.createTable method.
• HBASE-6565: Fixed issues with coprocessor in multithreading environments.

• HBASE-6538: Removed copy_table.rb script file.

• HBASE-6608: Fixes for HBASE-6160.

• HBASE-6503: Updated HBase Shell documentation.

• HBASE-5714: Enhanced permission checks to ensure that write permissions are checked before hbck is used to modify HDFS.

• HBASE-6631: Fixed failure issues for TestHMasterRPCException.

• HBASE-6632: Fixed failure issues for testCreateTableRPCTimeOut method.

• HBASE-6054: Fixed build issues caused because of missing commons-io.

• HBASE-5986: Resolved issues caused while executing large scale ingestion tests. The .META table updates are now atomic when regions are split.

• HBASE-6088: Fixed ZooKeeper exceptions (caused when creating RS_2K_SPLITTINGnode) that prevented Region splitting.

• HBASE-6107: Fixed issues with distributed log splitting.

• HBASE-6450: Added support to set MALLOC_ARENA_MAX in hbase-config.sh file. This fix resolves the issue of RegionServer crashes on RHEL 6.x due to memory.

8.2.3. Patch information for Hive

Hive includes the following patches:

• HIVE-2928: Added support for Oracle-backed Hive-Metastore (longvarchar to clob in package.jdo).

• HIVE-3082: Enhanced Oracle Metastore schema script to include DDL for DN internal tables.

• HIVE-3008: Fixed issues with memory leak in TUGIContainingTransport.

• HIVE-3063: Fixed failures with drop partition for non-string columns.

• HIVE-3076: Fixed failures with drop partition for non-partition columns.

• HIVE-3168: Fixed issues caused due to additional characters returned with ByteArrayRef.

• HIVE-3246: Changed the internal representation of binary type within Hive. UDFs which were earlier using either the binary type and or the Java representation of binary data in Hive (ByteArrayRef) must now be updated to reflect the new representation - byte[]. Also note that this does not change the format for on-disk data.
• HIVE-3153: Improved Relese codecs and output streams.

• HIVE-3291: Fixed failures for the shims module.

• HIVE-3098: Fixed memory leak from large number of FileSystem instance in FileSystem.Cache.

• HIVE-2084: Datanucleus is upgraded to upstream version 3.0.1.

• HIVE-2918: Fixed exceptions caused for Hive Dynamic Partition Insert when the number of partitions are created even after the default value of hive.exec.max.dynamic.partitions is increased to 2000.

8.2.4. Patch information for HCatalog

HCatalog includes the following patches:

• HCATALOG-485: Added document for storage-based security. The storage based security now ignores GRANT/REVOKE statements

• HCATALOG-431: Added document for instructions on mapping HCatalog type to either a Java class or a Pig type.

• HCATALOG-492: Added document for instruction on using the CTAS workaround for Hive with JSON SerDe.

• HCATALOG-442: Updated documentation for instructions on using HCatalog with Pig.

• HCATALOG-482: Added documentation for instructions on shipping libjars from HDFS. This option allows reusing distributed cache entries.

• HCATALOG-481: Fixed command line interface (CLI) usage syntax and also updated HCatalog documentation.

• HCATALOG-444: Added documentation for using Reader and Writer Interfaces.

• HCATALOG-427: Added documentation for storage based authorization.

• HCATALOG-448: Performance improvements for HCatStorer.

• HCATALOG-350: Added support to write binary data to the HCatRecord.

• HCATALOG-436: Fixed incorrect naming for JSON SerDe column on CTAS.

• HCATALOG-471: Fixed issues with HCat_ShowDes_1 test failures.

• HCATALOG-464: Upgraded datanucleus for HCatalog.

• HCATALOG-412: Added support for HCatalog to publish artifacts to the Maven repository.

• HCATALOG-410: Added support for proxy user in HCatalog client.

• HCATALOG-420: Added HCATALOG-363 patch to the HCatalog 0.4 branch.
8.2.5. Patch information for Pig

Pig includes the following patches:

• PIG-2766: Introduced a new command line parameter for Pig - -useHCatalog. This parameter imports the appropriate JAR files for Pig's use with HCatalog. If the user has setup the home directories for Hive or HCatalog, those settings would override the default values.

8.2.6. Patch information for Oozie

Oozie is patched to include the following:

• OOZIE-698: Enhanced sharelib components.
• OOZIE-697: Added OOZIE-77 patch to Oozie 3.1 branch.
• OOZIE-810: Fixed compilation issues for Oozie documentation.
• OOZIE-863: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

8.2.7. Patch information for Sqoop

Sqoop is patched to include the following:

• SQOOP-438: Added support to allow sourcing of sqop-env.sh file. This enhancement now allows setting variables directly in the configuration files.
• SQOOP-462: Fixed failures for Sqoop HBase test compilation.
• SQOOP-578: Fixed issues with sqoop script calls.
• SQOOP-579: Improved reuse for custom manager factories.
• SQOOP-580: Added support for an open ended job teardown method which is invoked after the job execution.
• SQOOP-582: Added a template method for job submission in Export/Import JobBase. This will enable a connector to submit a job and also complete other tasks simultaneously while the job is in progress.

8.2.8. Patch information for Ambari

Ambari includes the following patches:

• AMBARI-664: Fixed mapred io sort mb and heap size for Map/Reduce.
• AMBARI-641: Fixed issues for the status.dat file for Nagios.
• AMBARI-628: Fixed issues for hdp-nagios and hdp-monitoring files
• AMBARI-633: Fixed invalid HTML markup for monitoring dashboard.
• AMBARI-597: Removed RPM dependency on the /usr/bin/php scripts.

8.3. Minimum system requirements

Hardware Recommendations:
Although there is no single hardware requirement for installing HDP, there are some basic guidelines. You can see sample setups here.

Operating Systems Requirements:
The following operating systems are supported:
• 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
• 64-bit CentOS v5.*, v6.*

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

Graphics Requirements:
The HMC deployment wizard runs as a browser-based Web app. You must have a machine capable of running a graphical browser to use this tool.

Software Requirements:
On each of your hosts:
• yum
• rpm
• scp
• curl
• wget
• pdsh

On the machine from which you will run HMC:
• Firefox v.12+

Database Requirements:
Hive or HCatalog requires a MySQL database for its use. You can choose to use a current instance or let the HMC deployment wizard create one for you.

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 repos. 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)” repo. If this repo is disabled, the installation is unable to access the **rubygems** package, which is necessary for HMC to operate. If you encounter problems with base OS repos being unavailable, please contact your system administrator to arrange for these additional repos to be proxied or mirrored.

### 8.4. Improvements

- Fixed incorrect host mappings for Hive causing failure of Hive smoke tests.
- Templeton updated to upstream version 0.1.4.
- HA-monitor updated to upstream version 1.1.0.
- Fixed HDFS log corruption when disk gets filled.
- Added support for pluggable components. This feature will enable export of DFS functionality using arbitrary protocols.
- Added support to enable service plugins for JobTracker.

### 8.5. Known Issues

- The **ALTER INDEX** command will fail for Hive if used in an automated script that also contains the **CREATE INDEX** command. The workaround is to either use the **ALTER INDEX** command in an interactive shell or add it to a separate script file.

- Hive and HCatalog authorizations are based on permissions in the underlying storage system and so are not affected by account-management DDL statements such as **GRANT** and **REVOKE**. See HCatalog documentation of “Authorizations for HCatalog”.

- Preview of the mount point directories during HDP installation will display the Oozie and ZooKeeper directories even if the corresponding services are not enabled. For details, see **AMBARI-572**.

- In some cases, while finalizing the bootstrap nodes for HMC the update shows incorrect message.

- HMC installation currently does not support Hadoop security.

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

- To be able to use Oozie command line client, you must first export **JAVA_HOME**.
• Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.
9. Release Notes HDP-1.1.0.15

RELEASE NOTES:Hortonworks Data Platform with Hortonworks Management Console powered by Apache Hadoop

9.1. Product Version: HDP-1.1.0.15

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

- Apache Hadoop 1.0.3
- Apache HBase 0.92.1+
- Apache Pig 0.9.2
- Apache ZooKeeper 3.3.4
- Apache HCatalog 0.4.0
- Apache Hive 0.9.0
- Templeton 0.1.4
- Apache Oozie 3.1.3
- Apache Sqoop 1.4.2
- Hortonworks Management Center (HMC) 1.0.2
- Apache Flume 1.2.0
- HA-Monitor 0.1.0
- Third party components:
  - Ganglia 3.2.0
  - Nagios 3.2.3
  - Talend Open Studio 5.1.1

9.2. Patch Information

9.2.1. Patch information for Hadoop

Hadoop is patched to include the following:

**9.2.2. Patch information for HBase**

HBase is based on Apache SVN branch 0.92, revision 1344056 and includes the following:

- **HBASE-6447**: Fixed issues with ZooKeeper test failures.
- **HBASE-6334**: Improvements for RegionServer tests.
- **HBASE-4470**: Fixed issues related to `ServerNotRunning` exception with HBase master.
- **HBASE-6460**: Fixed issues with `hbck -repairHoles` command.
- **HBASE-6552**: Fixed issues related to `TestAcid Guarantees` system tests.
- **HBASE-6512**: Fixed issues related to incorrect log name for `OfflineMetaRepair`.
- **HBASE-6308**: Improved coprocessors to prevent dependency conflicts with HBase.
- **HBASE-6576**: Fixed issues with `HBaseAdmin.createTable` method.
• HBASE-6565: Fixed issues with coprocessor in multithreading environments.

• HBASE-6538: Removed copy_table.rb script file.

• HBASE-6608: Fixes for HBASE-6160.

• HBASE-6503: Updated HBase Shell documentation.

• HBASE-5714: Enhanced permission checks to ensure that write permissions are checked before hbck is used to modify HDFS.

• HBASE-6631: Fixed failure issues for TestHMasterRPCException.

• HBASE-6632: Fixed failure issues for testCreateTableRPCTimeOut method.

• HBASE-6054: Fixed build issues caused because of missing commons-io.

• HBASE-5986: Resolved issues caused while executing large scale ingestion tests. The .META table updates are now atomic when regions are split.

• HBASE-6088: Fixed ZooKeeper exceptions (caused when creating RS_ZK_SPLITTINGnode) that prevented Region splitting.

• HBASE-6107: Fixed issues with distributed log splitting.

• HBASE-6450: Added support to set MALLOC_ARENA_MAX in hbase-config.sh file. This fix resolves the issue of RegionServer crashes on RHEL 6.x due to memory.

9.2.3. Patch information for Hive

Hive includes the following patches:

• HIVE-3008: Fixed issues with memory leak in TUGIContainingTransport.

• HIVE-3063: Fixed failures with drop partition for non-string columns.

• HIVE-3076: Fixed failures with drop partition for non-partition columns.

• HIVE-3168: Fixed issues caused due to additional characters returned with ByteArrayRef.

• HIVE-3246: Changed the internal representation of binary type within Hive. UDFs which were earlier using either the binary type and or the Java representation of binary data in Hive (ByteArrayRef) must now be updated to reflect the new representation - byte[]. Also note that this does not change the format for on-disk data.

• HIVE-3153: Improved Release codecs and output streams.

• HIVE-3291: Fixed failures for the shims module.

• HIVE-3098: Fixed memory leak from large number of FileSystem instanse in FileSystem.Cache.

• HIVE-2084: Datanucleus is upgraded to upstream version 3.0.1.
• HIVE-2918: Fixed exceptions caused for Hive Dynamic Partition Insert when the number of partitions are created even after the default value of hive.exec.max.dynamic.partitions is increased to 2000.

### 9.2.4. Patch information for HCatalog

HCatalog includes the following patches:

- HCATALOG-448: Performance improvements for HCatStorer.
- HCATALOG-350: Added support to write binary data to the HCatRecord.
- HCATALOG-436: Fixed incorrect naming for JSON SerDe column on CTAS.
- HCATALOG-471: Fixed issues with HCat_ShowDes_1 test failures.
- HCATALOG-464: Upgraded datanucleus for HCatalog.
- HCATALOG-412: Added support for HCatalog to publish artifacts to the Maven repository.
- HCATALOG-410: Added support for proxy user in HCatalog client.
- HCATALOG-420: Added HCATALOG-363 patch to the HCatalog 0.4 branch.

### 9.2.5. Patch information for Pig

Pig includes the following patches:

- PIG-2766: Introduced a new command line parameter for Pig `-useHCatalog`. This parameter imports the appropriate JAR files for Pig's use with HCatalog. If the user has setup the home directories for Hive or HCatalog, those settings would override the default values.

### 9.2.6. Patch information for Oozie

Oozie is patched to include the following:

- OOZIE-698: Enhanced sharelib components.
- OOZIE-697: Added OOZIE-77 patch to Oozie 3.1 branch.
- OOZIE-810: Fixed compilation issues for Oozie documentation.
- OOZIE-863: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

### 9.2.7. Patch information for Sqoop

Sqoop is patched to include the following:

- SQOOP-438: Added support to allow sourcing of sqoop-env.sh file. This enhancement now allows setting variables directly in the configuration files.
• SQOOP-462: Fixed failures for Sqoop HBase test compilation.

• SQOOP-578: Fixed issues with sqoop script calls.

• SQOOP-579: Improved reuse for custom manager factories.

• SQOOP-580: Added support for an open ended job teardown method which is invoked after the job execution.

• SQOOP-582: Added a template method for job submission in Export/Import JobBase. This will enable a connector to submit a job and also complete other tasks simultaneously while the job is in progress.

9.2.8. Patch information for Ambari

Ambari includes the following patches:

• AMBARI-664: Fixed mapred io sort mb and heap size for Map/Reduce.

• AMBARI-641: Fixed issues for the status.dat file for Nagios.

• AMBARI-628: Fixed issues for hdp-nagios and hdp-monitoring files

• AMBARI-633: Fixed invalid HTML markup for monitoring dashboard.

• AMBARI-597: Removed RPM dependency on the /usr/bin/php scripts.

9.3. Minimum system requirements

Hardware Recommendations:

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

Operating Systems Requirements:

The following operating systems are supported:

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

• 64-bit CentOS v5.*, v6.*

Important

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

Graphics Requirements:

The HMC deployment wizard runs as a browser-based Web app. You must have a machine capable of running a graphical browser to use this tool.

Software Requirements:

On each of your hosts:
• yum
• rpm
• scp
• curl
• wget
• pdsh

• On the machine from which you will run HMC:
  • Firefox v.12+

**Database Requirements:**
Hive or HCatalog requires a MySQL database for its use. You can choose to use a current instance or let the HMC deployment wizard create one for you.

**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.

**Note**
If you use the Hortonworks repository tarball image to copy the repository to your local mirror, the name of the gsInstaller file in that local copy will be HDP-gsInstaller-1.1.0.15-2.tar.gz instead of HDP-gsInstaller-1.1.0.15.tar.gz.

**9.4. Improvements**

• Introduced storage-based authorization for Hive with HCatalog.

• Introduced high availability feature using VMware and Red Hat Enterprise Linux. See High Availability for Hadoop.

• Added support for Apache Flume. For details, see, Installing Apache Flume.

• Added support to install HDP manually using RPMs. For details, see, Manually Deploying HDP(Using RPMs).

**9.5. Known Issues**

• The **ALTER INDEX** command will fail for Hive if used in an automated script that also contains the **CREATE INDEX** command. The workaround is to either use the **ALTER INDEX** command in an interactive shell or add it to a separate script file.
- Hive and HCatalog authorizations are based on permissions in the underlying storage system and so are not affected by account-management DDL statements such as GRANT and REVOKE. See HCatalog documentation of "Authorizations for HCatalog".

- Preview of the mount point directories during HDP installation will display the Oozie and ZooKeeper directories even if the corresponding services are not enabled. For details, see AMBARI-572.

- In some cases, while finalizing the bootstrap nodes for HMC the update shows incorrect message.

- HMC installation currently does not support Hadoop security.

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

- To be able to use Oozie command line client, you must first export JAVA_HOME.

- Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.
10. Release Notes HDP-1.0.1.14

RELEASE NOTES: Hortonworks Data Platform with Hortonworks Management Console powered by Apache Hadoop

10.1. Product Version: HDP-1.0.1.14

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

• Apache Hadoop 1.0.3
• Apache HBase 0.92.1+
• Apache Pig 0.9.2
• Apache ZooKeeper 3.3.4
• Apache HCatalog 0.4.0
• Apache Hive 0.9.0
• Templeton 0.1.4
• Apache Oozie 3.1.3
• Apache Sqoop 1.4.1
• Apache Ambari 0.9
• Third party components:
  • Ganglia 3.2.0
  • Nagios 3.2.3
  • Talend Open Studio 5.1.1

10.2. Patch Information

10.2.1. Patch information for Hadoop

Hadoop is patched to include the following:

• HDFS-3652: Fixed edit stream issues caused in the event of FSEditlog failure.
• MAPREDUCE-4399: Fixed performance regression in shuffle.
• HADOOP-7154: Added support to set MALLOC_ARENA_MAX in hadoop-config.sh file.
• HADOOP-5464: Added support for disabling write timeout. To do this, you can set zero values for dfs.socket.timeout and dfs.datanode.socket.write.timeout parameters.
• HDFS-2617: Replaced Kerberized SSL for image transfer and fsck with SPNEGO-based solution.
• HDFS-3466: Fixed SPNEGO filter to use the DFS_WEB_AUTHENTICATION_KERBEROS_KEYTAB_KEY to find the keytab.
• HDFS-3461: Fixed HFTP to use the same port and protocol while obtaining the delegation token.
• HADOOP-6947: Fixed Kerberos relogin to configure the refreshKrb5Config correctly.
• MAPREDUCE-3837: Enhanced JobTracker job recovery mechanism in the event of a crash.

10.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.92, revision 1344056 and includes the following:
• HBASE-6450: Added support to set MALLOC_ARENA_MAX in hbase-config.sh file. This fix resolves the issue of RegionServer crashes on RHEL 6.x due to memory.
• HBASE-6054: Fixed build issues caused because of missing commons-io.
• HBASE-5986: Resolved issues caused while executing large scale ingestion tests. The .META table updates are now atomic when regions are split.
• HBASE-6088: Fixed ZooKeeper exceptions (caused when creating RS_ZK_SPLITTINGnode) that prevented Region splitting.
• HBASE-6107: Fixed issues with distributed log splitting.

10.2.3. Patch information for HCatalog

HCatalog is patched to include the following:
• HCATALOG-412: Added support for HCatalog to publish artifacts to the Maven repository.
• HCATALOG-410: Added support for proxy user in HCatalog client.
• HCATALOG-420: Added HCATALOG-363 patch to the HCatalog 0.4 branch.

10.2.4. Patch information for Hive

Hive is patched to include the following:
• HIVE-2084: Datanucleus is upgraded to upstream version 3.0.1.
• HIVE-2918: Fixed exceptions caused for Hive Dynamic Partition Insert when the number of partitions are created even after the default value of hive.exec.max.dynamic.partitions is increased to 2000.

10.2.5. Patch information for Oozie

Oozie is patched to include the following:
• OOZIE-698: Enhanced sharelib components.

• OOZIE-697: Added OOZIE-77 patch to Oozie 3.1 branch.

• OOZIE-810: Fixed compilation issues for Oozie documentation.

• OOZIE-863: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.

### 10.2.6. Patch information for Sqoop

Sqoop is patched to include the following:

• SQOOP-438: Added support to allow sourcing of sqoop-env.sh file. This enhancement now allows setting variables directly in the configuration files.

• SQOOP-462: Fixed failures for Sqoop HBase test compilation.

### 10.3. Minimum system requirements

#### Hardware Recommendations:

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

#### Operating Systems Requirements:

The following operating systems are supported:

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

• 64-bit CentOS v5.*, v6.*

#### Important

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

#### Graphics Requirements:

The HMC deployment wizard runs as a browser-based Web app. You must have a machine capable of running a graphical browser to use this tool.

#### Software Requirements:

On each of your hosts:

• yum

• rpm

• scp

• curl
• wget
• pdsh

• On the machine from which you will run HMC:
  • Firefox v.12+

Database Requirements:
Hive or HCatalog requires a MySQL database for its use. You can choose to use a current instance or let the HMC deployment wizard create one for you.

10.4. Improvements

• Added support for RHEL v6.x and CentOS v6.x for Hortonworks Management Center (HMC).
  • HMC is the graphical user interface (GUI) based installer for managing and monitoring end-to-end Hadoop deployments.
  • For more details, see: Using HMC.

• Improved configuration options for HMC.

• Added support for additional HBase configuration parameters.

• Improved validation tests for invalid parameter values for configuring the services.

• Fixed LZO library unavailability during runtime.

• Improved support for testing client side validation errors.

• Fixed the log generation issue caused due to undefined variables.

• Fixed Templeton configuration parameters.

• Added support for Talend Open Studio.
  • HDP packages Talend Open Studio to provide a graphical interface for Extract, Transform, and Load (ETL).
  • Talend utilizes HDP's HCatalog metadata management capability to import raw data into Hadoop, create and manage schemas on the raw data, and facilitate transformational queries on that data.

See: Using Data Integration Services Powered By Talend.

10.5. Known Issues

• HMC installation currently does not support Hadoop security.

• Use of init.d scripts for starting or stopping Hadoop services, is not recommended.
• To be able to use Oozie command line client, you must first export JAVA_HOME.

• Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.
11. Release Notes HDP-1.0.0.12

RELEASE NOTES: Hortonworks Data Platform with Hortonworks Management Console powered by Apache Hadoop

11.1. Product Version: HDP-1.0.0.12

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

- Apache Hadoop 1.0.3
- Apache HBase 0.92.1+
- Apache Pig 0.9.2
- Apache ZooKeeper 3.3.4
- Apache HCatalog 0.4.0
- Apache Hive 0.9.0
- Templeton 0.1.4
- Apache Oozie 3.1.3
- Apache Sqoop 1.4.1

- Third party components:
  - Ganglia 3.2.0
  - Nagios 3.2.3

11.2. Patch Information

11.2.1. Patch information for Hadoop

Hadoop is patched to include the following:

- HADOOP-5464: Added support for disabling write timeout. To do this, you can set zero values for dfs.socket.timeout and dfs.datanode.socket.write.timeout parameters.

- HDFS-2617: Replaced Kerberized SSL for image transfer and fsck with SPNEGO-based solution.

- HDFS-3466: Fixed SPNEGO filter to use the DFS_WEB_AUTHENTICATION_KERBEROS_KEYTAB_KEY to find the keytab.

- HDFS-3461: Fixed HFTP to use the same port and protocol while obtaining the delegation token.
• HADOOP-6947: Fixed Kerberos relogin to configure the refreshKrb5Config correctly.
• MAPREduce-3837: Enhanced JobTracker job recovery mechanism in the event of a crash.

### 11.2.2. Patch information for HBase

HBase is based on Apache SVN branch 0.92, revision 1344056 and includes the following:

• HBASE-6054: Fixed build issues caused because of missing commons-io.

• HBASE-5986: Resolved issues caused while executing large scale ingestion tests. The .META table updates are now atomic when regions are split.

• HBASE-6088: Fixed ZooKeeper exceptions (caused when creating RS_ZK_SPLITTINGnode) that prevented Region splitting.

• HBASE-6107: Fixed issues with distributed log splitting.

### 11.2.3. Patch information for HCatalog

HCatalog is patched to include the following:

• HCATALOG-412: Added support for HCatalog to publish artifacts to the Maven repository.

• HCATALOG-410: Added support for proxy user in HCatalog client.

• HCATALOG-420: Added HCATALOG-363 patch to the HCatalog 0.4 branch.

### 11.2.4. Patch information for Hive

Hive is patched to include the following:

• HIVE-2084: Datanucleus is upgraded to upstream version 3.0.1.

• HIVE-2918: Fixed exceptions caused for Hive Dynamic Partition Insert when the number of partitions are created even after the default value of hive.exec.max.dynamic.partitions is increased to 2000.

### 11.2.5. Patch information for Oozie

Oozie is patched to include the following:

• OOZIE-698: Enhanced sharelib components.

• OOZIE-697: Added OOZIE-77 patch to Oozie 3.1 branch.

• OOZIE-810: Fixed compilation issues for Oozie documentation.

• OOZIE-863: Fixed issues caused due to JAVA_HOME settings when oozie-env.sh script is invoked.
11.2.6. Patch information for Sqoop

Sqoop is patched to include the following:

• SQOOP-438: Added support to allow sourcing of `sqoop-env.sh` file. This enhancement now allows to set variables directly in the configuration files.

• SQOOP-462: Fixed failures for Sqoop HBase test compilation.

11.3. Minimum system requirements

Hardware Recommendations:

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

Operating Systems Requirements:

The following operating systems are supported:

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

Important

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

Graphics Requirements:

The HMC deployment wizard runs as a browser-based Web app. You must have a machine capable of running a graphical browser to use this tool.

Software Requirements:

On each of your hosts:

• yum
• rpm
• scp
• curl
• wget
• pdsh
• net-snmp
• net-snmp-utils

• On the machine from which you will run HMC:
• Firefox v.12+

**Database Requirements:**

Hive or HCatalog requires a MySQL database for its use. You can choose to use a current instance or let the HMC deployment wizard create one for you.

### 11.4. Improvements

• Introduced Hortonworks Management Center (HMC).

  • HMC is the graphical user interface (GUI) based installer for managing and monitoring end-to-end Hadoop deployments.
  
  • For more details, see: Using HMC.

• Upgraded multiple components.

• Added support for Talend Open Studio.

  • HDP packages Talend Open Studio to provide a graphical interface for Extract, Transform, and Load (ETL).
  
  • Talend utilizes HDP’s HCatalog metadata management capability to import raw data into Hadoop, create and manage schemas on the raw data, and facilitate transformational queries on that data.

  See: Using Data Integration Services Powered By Talend.

### 11.5. Known Issues

• HMC installation currently does not support Hadoop security.

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

• To be able to use Oozie command line client, you must first export JAVA_HOME.

• Pig jobs submitted via Templeton fail. The workaround for this issue is available here.

• The Sqoop client deployed by HMC does not have the neccessary MySQL connector JAR file. The workaround for this issue is available here.

• Pig or MapReduce jobs get incorrect data when reading binary data type from the HCatalog table. For details, see: HCATALOG-430.