Hortonworks Data Platform

Release Notes

docs.hortonworks.com

Hortonworks Data Platform : Release Notes

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.

Table of Contents

1. 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

1.1. 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

1.2. 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

1.2.1. 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 incorect 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.

1.2.2. 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.

1.2.3. 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_pipe1.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.

1.2.4. 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.

1.2.5. 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.

1.2.6. 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.

1.2.7. 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`.

1.2.8. 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: Enhanaced `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.

1.2.9. 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.

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 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).

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

1.3.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

1.3.4. 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.

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. 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 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.

- **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_1.2.2
  ```

  or click here.

- 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 HDP-1.2.1.1
         ```

         This command will install 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:

         ```
         zypper up -r HDP-1.2.1.1
         ```

    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 HDP-1.2.1.1
         ```

This command will install 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:
>
> ```
> zypper up -r HDP-1.2.1.1
> ```

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

1.5. 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

1.5.1. 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.

1.5.2. 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.

1.5.3. 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.

1.5.4. Known Issues for Oozie

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

1.5.5. Known Issues for Sqoop

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

  This issue occurs because the Teradata Connector 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 `import-all-tables` has additional restrictions on the schema (tables cannot have multi-column primary keys, etc.) as described in the Sqoop User Guide, Chapter 8.

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

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

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

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

  This issue occurs 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.

1.5.6. 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
```

or click here.

2. 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.

2.1. 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

2.2. Patch Information

2.2.1. Patch information for Hadoop

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

- 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 incorect 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`.

2.2.2. 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.

2.2.3. 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_pipe1.q`.

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

- **HIVE-3717**: Fixed cimpilation 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.

2.2.4. 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.

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-2885: Fixed test failures for `TestJobSumission` and `TestHBaseStorage`.

- 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.

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.

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

2.2.8. 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: Enhanaced `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.

2.2.9. 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.

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

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

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

2.5. 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). 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. 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.

3.1. 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

3.2. Patch Information

3.2.1. Patch information for Hadoop

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

- 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 incorect 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.`

3.2.2. 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.

3.2.3. 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_pipe1.q`.

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

- **HIVE-3717**: Fixed cimpilation 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.

3.2.4. 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.

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-2885: Fixed test failures for `TestJobSumission` and `TestHBaseStorage`.

- 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.

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-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: Enhanaced `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.

3.2.9. 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.

3.3. 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).

3.4. Improvements

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

3.5. 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). 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. 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.

4.1. 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

4.2. Patch Information

4.2.1. Patch information for Hadoop

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

- 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 incorect 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`.

4.2.2. 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.

4.2.3. 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_pipe1.q`.

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

- **HIVE-3717**: Fixed cimpilation 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.

4.2.4. 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.

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-2885: Fixed test failures for `TestJobSumission` and `TestHBaseStorage`.

- 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.

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-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: Enhanaced `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.

4.2.9. 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.

4.2.10. 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.

4.3. 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).

4.4. 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. .

4.5. 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). 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.

5. Release Notes HDP-1.1.1.16

Hortonworks Data Platform with Hortonworks Management Console powered by Apache Hadoop

5.1. 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

5.2. Patch Information

5.2.1. Patch information for Hadoop

Hadoop is patched to include the following:

- High Availability (HA) enhancements: HDFS-3522, HDFS-3521, HDFS-1108, HDFS-3551, HDFS-528, HDFS-3667, HDFS-3516, HDFS-3696, HDFS-3658, MAPREDUCE-4328, MAPREDUCE-3837, MAPREDUCE-4328, MAPREDUCE-4603, and HADOOP-8656.

- Performance improvements: HDFS-2465, HDFS-2751, HDFS-496, MAPREDUCE-782, MAPREDUCE-1906, MAPREDUCE-4399, MAPREDUCE-4400, MAPREDUCE-3289, MAPREDUCE-3278, HADOOP-7753, and HADOOP-8617.

- 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 ontaining the delegation token.

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

- MAPREDUCE-3837: Enhanced JobTracker job recivery 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.

5.2.2. 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 permissons 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_SPLITTING`node) 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.

- 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.

5.2.3. 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 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 imcreased to 2000.

5.2.4. 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 instructiosn on using the CTAS workaround for Hive with JSON SerDe.

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

- HCATALOG-482: Added documentation for instructions on shipping **libjars** from HDFS. This otion 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.

5.2.5. 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.

5.2.6. 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 .

5.2.7. 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.

5.2.8. 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.

5.3. 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.

5.4. 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.

5.5. 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.

6. Release Notes HDP-1.1.0.15

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

6.1. 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

6.2. Patch Information

6.2.1. Patch information for Hadoop

Hadoop is patched to include the following:

- High Availability (HA) enhancements: HDFS-3522, HDFS-3521, HDFS-1108, HDFS-3551, HDFS-528, HDFS-3667, HDFS-3516, HDFS-3696, HDFS-3658, MAPREDUCE-4328, MAPREDUCE-3837, MAPREDUCE-4328, MAPREDUCE-4603, and HADOOP-8656.

- Performance improvements: HDFS-2465, HDFS-2751, HDFS-496, MAPREDUCE-782, MAPREDUCE-1906, MAPREDUCE-4399, MAPREDUCE-4400, MAPREDUCE-3289, MAPREDUCE-3278, HADOOP-7753, and HADOOP-8617.

- 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 ontaining the delegation token.

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

- MAPREDUCE-3837: Enhanced JobTracker job recivery 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.

6.2.2. 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 permissons 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_SPLITTING`node) 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.

- 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.

6.2.3. 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 Relese 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 imcreased to 2000.

6.2.4. 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.

6.2.5. 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.

6.2.6. 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.

6.2.7. 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.

6.2.8. 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.

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.*

> ### 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`.

6.4. 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) .

6.5. 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.

7. Release Notes HDP-1.0.1.14

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

7.1. 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

7.2. Patch Information

7.2.1. 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 ontaining the delegation token.

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

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

7.2.2. 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_SPLITTING`node) that prevented Region splitting.

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

7.2.3. 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.

7.2.4. 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 imcreased to 2000.

7.2.5. 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 .

7.2.6. 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.

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.*

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.

7.4. 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.

7.5. 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.

8. Release Notes HDP-1.0.0.12

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

8.1. 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

8.2. Patch Information

8.2.1. 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 ontaining the delegation token.

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

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

8.2.2. 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_SPLITTING`node) that prevented Region splitting.

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

8.2.3. 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.

8.2.4. 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 imcreased to 2000.

8.2.5. 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 .

8.2.6. 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.

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

- 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.

8.4. 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.

8.5. 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.