Hortonworks Data Platform

Release Notes

(Oct 22, 2013)

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

List of Tables

1. Release Notes HDP-2.0.6.0

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

In this document:

1.1. Product Version: HDP-2.0.6.0

All HDP 2.0.6.0 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.0.6.0 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.0.6.0.

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

- Apache Hadoop 2.2.0
- Apache HBase 0.96.0
- Apache ZooKeeper 3.4.5
- Apache Pig 0.12.0
- Apache Hive 0.12.0
- Apache HCatalog 0.12.0

Note

Apache HCatalog is now merged with Apache Hive.

- Hue 2.3.0
- Apache Oozie 4.0.0
- Apache Sqoop 1.4.4
- Apache Flume 1.4.0

- Apache Ambari 1.4.1
- Apache Mahout 0.8.0
- Third party components:
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

1.2. Patch Information

In this section:

1.2.1. Patch Information for Hadoop

Hadoop is based on Apache Hadoop 2.2.0 and includes the following additional patches:

- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- **BUG-8178**. Datanodes fail to register with namenode due to minimum version check.

1.2.2. Patch information for HBase

HBase is based on Apache HBase 0.96.0.

1.2.3. Patch information for ZooKeeper

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

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

1.2.4. Patch information for Pig

Pig is based on Apache Pig 0.12.0 and includes the following patches:

- PIG-3518: Need to ship jruby.jar in the release.
- PIG-3517: Fix PermGen error in Pig Unit test on Hadoop 2.
- PIG-3516: Pig does not bring in joda-time as dependency in its pig-template.xml.
- PIG-3512: Reducer estimater is broken by PIG-3497.
- PIG-3257: Add a UUID function to Pig.

1.2.5. Patch information for Hive

Hive is based on Apache Hive 0.12.0

Note

Apache HCatalog is now merged with Apache Hive.

- HIVE-5542: Webhcat is failing to run ddl command on a secure cluster
- HIVE-5511: `percentComplete` returned by job status from WebHCat is null
- HIVE-5496: hcat -e `drop database if exists` fails on authorizing non-existent null db
- HIVE-5485: SBAP errors on null partition being passed into partition level authorization
- HIVE-5484: TestSchemaTool failures when Hive version has more than 3 revision numbers
- HIVE-5480: WebHCat e2e tests for doAs feature are failing
- HIVE-5479:SBAP restricts `hcat -e 'show databases'`
- HIVE-5478: WebHCat e2e testsuite for hcat authorization tests needs some fixes
- HIVE-5474: drop table hangs when `concurrency=true`
- HIVE-5453: `jobsubmission2.conf` should use 'timeout' property
- HIVE-5448:webhcat duplicate test TestMapReduce_2 should be removed
- HIVE-5425: Provide a configuration option to control the default stripe size for ORC
- HIVE-5422: Upgrade Kyro to 2.22 now that it is released
- HIVE-5411: Migrate expression serialization to Kryo
- HIVE-5379: `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- HIVE-5353: job submission that requires access to metastore should not require additional jars to be shipped to target node
- HIVE-5290: Some HCatalog tests have been behaving flaky
- HIVE-5279: Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`

- HIVE-5263: Query Plan cloning time could be improved by using Kryo
- HIVE-5133: webhcat jobs that need to access metastore fails in secure mode
- HIVE-5112: Upgrade protobuf to 2.5 from 2.4
- HIVE-5070: Need to implement `listLocatedStatus()` in ProxyFileSystem for 0.23 shim
- HIVE-4910: Hadoop 2 archives broken
- HIVE-4545: HS2 should return describe table results without space padding
- HIVE-4485: beeline prints null as empty strings
- HIVE-4388: HBase tests fail against Hadoop 2
- HIVE-3815: hive table rename fails if filesystem cache is disabled
- HIVE-1511: Hive plan serialization is slow.

1.2.6. Patch information for HCatalog

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

1.2.7. Patch information for Oozie

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

- OOZIE-1593: Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.

1.2.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.

1.3. Minimum system requirements

In this section:

- Hardware Recommendations
- Operating Systems Requirements
- Software Requirements
- Database Requirements
- Virtualization and Cloud Platforms

- Configure the Local Repositories

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 are supported:

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

Important

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

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

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

1.3.3. Software Requirements

On each of your hosts:

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

1.3.4. Database Requirements

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

1.3.5. 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. 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.

1.4. Improvements

- Added Ubuntu host support.
- Added support to upgrade Ambari.
- Added support for Oracle JDK 1.7 and OpenJDK 7.
- Added the DATE dataytpe.
- Added NameNode HA support (Hadoop 2).
- Enhanced Sqpoop HCatalog support to allow direct mode connection manager implementations.
- Added Security documentation on wire encryption, encrypted shuffle, and JDBC encryption.
- Apache Ambari updated to version 1.4.1. This release of Apache Ambari includes the new features and improvements:
  - Added Hadoop 2 Stack option to Ambari.
  - Added support for enabling NameNode HA (for Hadoop 2 Stack).
  - Added support for Oracle JDK 1.7 and OpenJDK 7.

- Added Ambari Agent host-level Nagios alert.
- Added Host Checks and cleanup script during Cluster Install and Add Hosts wizards.

1.5. Known Issues

In this section:

1.5.1. Known Issues for HDP Installation

- **BUG-11344:** Invalid entry in core-site.xml file

  Problem: After extracting the companion files, the included core-site.xml file contains the following invalid entry:

```
{code}
<property>
    <name>fs.defaultFS</name>
    <value>TODO-NAMENODE-HOSTNAME</value>
  </property>
{code}
```

  Workaround: Edit the <value> element as follows:

```
{code}
<property>
    <name>fs.defaultFS</name>
    <value>hdfs://TODO-NAMENODE-HOSTNAME:PORT</value>
  </property>
{code}
```

  This includes the protocol and port together with the Namenode hostname.

1.5.2. Known Issues for Ubuntu 12.04

- Hue does not support Ubuntu hosts at this time.
- Ambari does not support Ubuntu hosts at this time.

1.5.3. Known Issues for SLES 11

- **BUG-9904:** php_curl Required for SLES 11 Sp1

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

  Workaround: Install php_curl on your SLES host.

  ```
  zypper install php-curl
  ```

1.5.4. Known Issues for HDFS

- **BUG-10264:** Could not complete file error while running load generator on a HA cluster.

  Problem: When a client sends a request to allocate a block to write to and the NameNode fails over at that time, client can fail to write to the file. The client is stuck permanently trying to recover from this error condition. This condition should rarely occur given failover must occur right when the `addBlock` operation is in progress.

  Workaround: This bug has been fixed in Apache as part of HDFS-5257.

1.5.5. Known Issues for Hive

- **BUG-10175:** NullPointerExceptions in ORC's PPD when using `select *` from table with a `where` predicate

  Problem: ORC predicate pushdown will fail with NullPointerException for `select *` queries with `where` predicate.

  Workaround: Disable the predicate pushdown feature using the following Hive config:

  ```
  SET hive.optimize.index.filter=false;
  ```

1.5.6. Known Issues for Oozie

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

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

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

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

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

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

  Problem: From the oozie log:

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

But this job and the application complete successfully.

1.5.7. Known Issues for Hue

- Ubuntu hosts not supported at this time.
- **BUG-9734:** Data loss during Migration of Hue DB from default (sqlite) to Oracle DB:

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

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

     Result: Hue starts fine and continues working, but there is loss of data.
  5. Some tables are lost in HCatalog.
  6. Some pig scripts do not show up on UI.

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

1.5.8. Known Issues for Ambari

- Ubuntu hosts not supported at this time.
- **BUG-10115:** The `HostCleanup` script does not appear to run when running distributed.

  Problem: During the host check step of the `Cluster Install` wizard or the `Add hosts` wizard, if warnings or errors are detected in your environments (such as installed packages or running processes), you will be provided information on how to execute the `HostCleanup` script. If you attempt to execute the `HostCleanup` script (using SSH, for example) distributed across all the hosts in your cluster without user interaction, the execution appears to hang since the script prompts for responses during execution.

  Workaround: Do not execute the script without user interaction. Execute the script on each host while attending the execute to be able to follow and respond to any

prompts. `python /usr/lib/python2.6/site-packages/ambari_agent/HostCleanup.py -k "users"` To delete all resources, ignore option -k. Use "-s" for silent cleanup.

- **BUG-9969:** Default Authorization Provider needs to be set in Hive configs.

  Problem: If you set up a secure cluster in Ambari and do not manually set the Default Authorization Provider in your Hive configurations, you will see errors.

  Workaround: Select Ambari Web > Services > Hive > Configs and set following properties for hive-site.xml:

  ```
  hive.security.authorization.enabled=true
  hive.security.authorization.manager=org.apache.hadoop.hive.ql.security.authorization.StorageBasedAuthorizationProvider
  hive.security.metastore.authorization.manager=org.apache.hadoop.hive.ql.security.authorization.StorageBasedAuthorizationProvider
  hive.security.authenticator.manager=org.apache.hadoop.hive.ql.security.ProxyUserAuthenticator
  ```

- **BUG-9797:** Enable Security fails when the Ambari setup is re-run to set JAVA_HOME to Oracle JDK7.

  Problem:

  1. Run ambari-setup -s
  2. Install the Oracle JDK : `/usr/jdk1.7.0_40`
  3. Run setup: `ambari-setup -j /usr/jdk1.7.0_40`
  4. Install `jdk1.7.0_40` on all hosts at JAVA_HOME path specified in above step.
  5. Install jce-7 policy on all hosts and unzip it at `/usr/jdk1.7.0_40/jre/lib/security`.
  6. Go through installation wizard and then Enable security wizard.
  7. Enable security fails as Ambari overrides all manually downloaded and unzipped jce-7 policy files with jce-6 policy files.

  Workaround: If you change your JDK, please remove the jce policy files.

- **BUG-9606:** Firewall issues display during Host Checks at Install or at Add New Host on CentOS 5 and SLES 11.

  Problem:

  - Start "Add new host" wizard trying to add host with iptables running on Centos05 or SLES 11 host.
  - After host confirmed host checks display firewall issue.
  - Stop iptables on host manually.
  - Rerun checks. Host checks still report about firewall issue.

- Refresh the page. Confirm process will repeat and finish without warnings this time, but with the message:

```
All host checks passed on 1 registered hosts. Click here to see the check
 results.
```

- Select **Click here** to see the check results. Host checks still report the firewall issue.

Workaround: Confirm iptables is disabled or all necessary ports are open on all cluster hosts. Disable iptables manually or configure your network for the necessary ports (see Configuring Ports for Hadoop 1.x and Configuring Ports for Hadoop 2.x).

- **BUG-9597:** Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start.

  Problem: The Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start. It is caused by service client's behavior when its state became `installed_and_configured` after Service Start:

```
{'hdp-hadoop::client': stage => 2, service_state =>
 installed_and_configured}
```

- **BUG-8898:** Ambari no longer stops iptables on Ambari Server or Ambari Agent start.

  Problem: Prior to HDP 2.0, the Ambari server and agents automatically stopped iptables if they were already running. With the release of HDP 2.0, Ambari does not stop iptables.

  Workaround: Disable iptables manually or configure your network for the necessary ports (see Configuring Ports for Hadoop 1.x and Configuring Ports for Hadoop 2.x

1.6. Third-party Licenses

Table 1.1. Third-party Licenses

HDP Component	Project Library	Version	License
Oozie	CERN Colt Project	1.2.0	Colt License Agreement
Pig	JRuby	1.6.7	CPL 1.0
ZooKeeper	jsoup	1.7.1	jsoup License (MIT License)
Mahout	XML Pull Parser (XPP)	1.1.4c	Indiana University Extreme! Lab Software License
Mahout	JAXB API	2.2.2	CDDL 1.1
Mahout	JAX-WS		CDDL 1.1

2. Release Notes HDP-2.0.5.0 Beta

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

In this document:

2.1. Product Version: HDP-2.0.5.0 Beta

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

- Apache Hadoop 2.1.0
- Apache HBase 0.95.2
- Apache ZooKeeper 3.4.5
- Apache Pig 0.11.1
- Apache Hive 0.11.0
- Apache HCatalog 0.11.0

Note

Apache HCatalog is now merged with Apache Hive.

- Apache Oozie 4.0.0
- Apache Sqoop 1.4.4
- Apache Sqoop2 1.99.2

Note

Apache Sqoop2 is still under development and not yet meant for deployment.

- Apache Flume 1.4.0
- Third party components:
  - Hue 2.2.0

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

2.2. Patch Information

In this section:

2.2.1. Patch Information for Hadoop

Hadoop is based on Apache Hadoop 2.1.0 and includes the following additional patches:

- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.

2.2.2. Patch information for HBase

HBase is based on Apache HBase 0.95 and includes the following patches:

- HBASE-9348: `TerminatedWrapper` error decoding, skipping skippable types.
- HBASE-9332: `OrderedBytes` does not decode Strings correctly.
- HBASE-9324: `TestProcedureMember#testMemberCommitException`, `testSimpleRun`, and `testMemberCommitCommsFailure` are flaky on Suse OS
- HBASE-9318: `Procedure#waitForLatch` may not throw error even there is one.
- HBASE-9315: `TestLruBlockCache.testBackgroundEvictionThread` fails on SuSE.
- HBASE-9311: Create a migration script that will move data from 0.94.x to 0.96.
- HBASE-9303: Snapshot restore of table which splits after snapshot was taken encounters 'Region is not online'
- HBASE-9302: NPE when granting permission on table.

- HBASE-9285: User who created table cannot scan the same table due to insufficient permissions.
- HBASE-9268: Client does not recover from a stalled region server.
- HBASE-9276: List tables API should filter with isSystemTable.
- HBASE-9257: `TestAsyncProcess#testFailAndSuccess` fails sometime due to a race condition.
- HBASE-9251: `list_namespace_tables` seems to fail.
- HBASE-9250: Lease sleep time can throw an exception.
- HBASE-9238: Bug in `Mutation::getFamilyMap`
- HBASE-9234: Rebuilding user regions should ignore system tables.
- HBASE-9227: `RESTServer` should handle the `loginUser` correctly.
- HBASE-9226: Thrift host and port are hardcoded in thrift2 `DemoClient.java`
- HBASE-9225: `TestAsyncProcess#testErrorsServers` is flaky on SuSE Linux.
- HBASE-9222: Thrift DemoClient failed with error IllegalArgument(message:Row length is 0).
- HBASE-9210: `hbase shell -d` does not print out exception stack trace.
- HBASE-9023: TestIOFencing.testFencingAroundCompactionAfterWALSync occasionally fails.
- HBASE-8760: Possible loss of data in snapshot taken after region split.
- HBASE-8409: Addendum. Part of the patch that got missed in the 0.95 commit of the original patch.
- HBASE-8165: Update the ProtoBuf libraries to 2.5.
- HBASE-7658: Grant with an empty string as permission should throw an exception.

2.2.3. Patch information for ZooKeeper

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

- ZOOKEEPER-1598: Ability to support more digits in the version string.
- ZOOKEEPER-1584: Added `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

2.2.4. Patch information for Pig

Pig is based on Apache Pig 0.11.1 + Apache Pig trunk SVN 1504206 and includes the following patches:

- PIG-3425: Hive jdo API jar referenced in Pig script throws error.
- PIG-3257: Add a UUID function to Pig.
- PIG-3247: Changes to Over functionality to allow user to declare datatype it will return.
- PIG-3247: Fixed error in percent_rank calculation.
- PIG-3247: Slightly modified to work with Pig 0.10.
- PIG-3048: Added MapReduce workflow information to job configuration.

2.2.5. Patch information for Hive

Hive is based on Apache Hive 0.11 + Apache Hive trunk SVN r1514119 and includes the following patches:

Note

Apache HCatalog is now merged with Apache Hive.

- HCATALOG-631: HBase e2e tests on single nodes on Hadoop 2.0.3 with "dfs.client.read.shortcircuit" turning on for HBase
- HIVE-7910: Fixed `webhcat_config.sh` checks for env variables being set before sourcing `webhcat-env.sh`.
- HIVE-5137: Fixed a Hive SQL query should not return a ResultSet when the underlying plan does not include a FetchTask.
- HIVE-5136: Fixed HCatalog HBase Storage handler fails test with `protbuf2.5`.
- HIVE-5133: Fixed WebHCat jobs that need to access metastore fails in secure mode.
- HIVE-5129: Fixed Multiple table insert fails on count(distinct).
- HIVE-5117: Fixed orc_dictionary_threshold is not deterministic.
- HIVE-5113: Fixed WebHCat should allow configuring memory used by templetoncontroller map job in hadoop2.
- HIVE-5112: Upgraded protobuf to 2.5 from 2.4.
- HIVE-5011: Dynamic partitioning in HCatalog broken on external tables.
- HIVE-5085: Fixed Hive Metatool errors out if HIVE_OPTS is set.
- HIVE-4679: Fixed WebHCat can deadlock Hadoop if the number of concurrently running tasks if higher or equal than the number of mappers.
- Remove path windowing function.
- HIVE-4611: Fixed SMB join failures because of conflicts in bigtable selection policy.
- HIVE-4601: WebHCat, Templeton support for proxy users.

- HIVE-4545: Fixed HS2 should return describe table results without space padding.
- HIVE-4542: Fixed `estJdbcDriver2.testMetaDataGetSchemas` fails because of unexpected database.
- HIVE-4524: Added support for `Hive HBaseStorageHandler` to work with HCatalog.
- HIVE-4485: beeline prints null as empty strings.
- HIVE-4388: HBase tests fail against Hadoop 2.
- HIVE-4246: Implemented predicate pushdown for ORC.
- HIVE-4214: OVER accepts general expression instead of just function.
- HIVE-4171: Current database in metastore. Hive is not consistent with `SessionState`.
- HIVE-3846: Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- HIVE-3815: Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- HIVE-2084: Upgraded DataNuclues from v2.0.3 to v3.0.1.

2.2.6. Patch information for HCatalog

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

2.2.7. Patch information for Oozie

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

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

2.2.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.

2.3. Minimum system requirements

In this section:

- Hardware Recommendations
- Operating Systems Requirements
- Software Requirements
- Database Requirements
- Virtualization and Cloud Platforms
- Configure the Local Repositories

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 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.

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

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 and come with an embedded Derby database by default.

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

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. 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.

2.4. Improvements

- Added the HDP 2.0 installation using the Ambari Automated Installer.
- Added the DATE dataytpe.
- Added NameNode HA support (HDP Stack only)
- Added encrypted shuffle.
- Enhanced Sqpoop HCatalog support to allow direct mode connection manager implementations.
- Added Sqoop2.
- Added Security documentation on wire encryption, encrypted shuffle, and JDBC encryption.

2.5. Known Issues

In this section:

- Known Issues for Hosts
- Known Issues for Hadoop
- Known Issues for Pig
- Known Issues for Ambari

2.5.1. Known Issues for Hosts

- Only RHEL6/CentOS6 hosts are supported for Beta. RHEL5/CentOS5 and SuSE11 are not supported at this time.

2.5.2. Known Issues for Hadoop

- YARN-1107: Restart secure RM with recovery enabled while oozie jobs are running causes the RM to fail during startup.

2.5.3. Known Issues for Pig

- Pig AvroStorage with snappy compression does not work on CentOS 5.

  Problem: Snappy does not work on CenotOS 5 (x86_64) configurations:

  ```
  libstdc++-4.1.2-52.el5_8.1
  ```

  ```
  glibc-2.5-81.el5_8.7
  ```

  Workaround: Snappy Java provides a workaround here.

- For Beta, the pig-0.11.2.2.0.5.0-67.tar.gz file installs Pig 0.11.1. This discrepancy is scheduled to be resolved in GA.

2.5.4. Known Issues for Ambari

- Do not select **Enable Security** because Secure Deployment Support is not available for Beta.
- HA is not yet intended for deployment and use with Ambari. For Beta, enabling HA requires that you shut down Ambari.

3. Release Notes HDP-2.0.4.0 (Community Preview)

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

In this document:

3.1. Product Version: HDP-2.0.4.0 (Community Preview)

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

- Apache Hadoop 2.1.0
- Apache HBase 0.94.6
- Apache ZooKeeper 3.4.5
- Apache Pig 0.11.1
- Apache Hive 0.11
- Tez 0.1
- Apache HCatalog 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3

3.2. Patch Information

In this section:

- Patch Information for HBase
- Patch Information for ZooKeeper
- Patch Information for Pig
- Patch Information for Hive
- Patch Information for HCatalog
- Patch Information for Oozie

3.2.1. Patch Information for Hadoop

Hadoop is based on Apache Hadoop 2.1.0 and includes the following additional patches:

- YARN-321: Generic Application History.

3.2.2. Patch information for HBase

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

- HBASE-6338: Cache Method in RPC handler.
- HBASE-6134: Improvement for split-worker to speed up distributed log splitting.
- HBASE-6508: 0.89-fb Filter out edits at log split time (without breaking backward compatibility).
- HBASE-6466: Enabled multi-thread for memstore flush.
- HBASE-7820: Support for multi-realm authentication.
- HBASE-8179: JSON formatting for cluster status is sort of broken.
- HBASE-8081: Backport HBASE-7213 - Separate hlog for meta tables.
- HBASE-8158: Backport HBASE-8140 - Use JarFinder more aggressively when resolving MR dependencies.
- HBASE-8260: Created generic integration test for HBase trunk and 0.94 branch that is more deterministic, can be run for longer durations, and is less aggressive
- HBASE-8274: Backport to 94: HBASE-7488 Implement HConnectionManager.locateRegions which is currently returning null
- HBASE-8146: IntegrationTestBigLinkedList does not work on distributed setup
- HBASE-8207: Replication could have data loss when machine name contains hyphen character (-).
- HBASE-8106: Test to check replication log znodes move is done correctly

- HBASE-8246: Backport HBASE-6318 to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- HBASE-8276: Backport HBASE-6738 to HBase 0.94 branch - Too aggressive task resubmission from the distributed log manager.
- HBASE-8270: Backport HBASE-8097 to HBase 0.94 branch - `MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.
- HBASE-8326: `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- HBASE-8352: Renamed `.snapshot` directory (to `.hbase-snapshot`).
- HBASE-8377: `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- HBASE-8505: References to split daughters should not be deleted separately from parent META entry (Patch file: `hbase-8505_v2-0.94-reduce.patch`).
- HBASE-8550: 0.94 ChaosMonkey grep for master is too broad.
- HBASE-8547: Fixed `java.lang.RuntimeException: Cached an already cached block` (Patch file: hbase-8547_v2-0.94-reduced.patch, and addendum2+3)
- HBASE-7410: Added snapshot/clone/restore/export documentation to reference guide.
- HBASE-8530: Refined error message from `ExportSnapshot` when there is leftover snapshot in target cluster.

3.2.3. 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: Added `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

3.2.4. Patch information for Pig

Pig is based on Apache Pig 0.11.1 and includes the following patches:

- PIG-3048: Added MapReduce workflow information to job configuration.
- PIG-3248: Upgraded Hadoop-2.0.0-alpha to Hadoop-2.0.3-alpha.
- PIG-3335: Fixed failures for `TestErrorHandling.tesNegative7` on MR2.
- PIG-3360: Fixed failures for intermittent end to end tests on Hadoop 2.

- PIG-3361: Improved Hadoop version detection ability for Pig unit test.

3.2.5. Patch information for Hive

Hive is based on Apache Hive 0.11 and includes the following:

Note

Apache HCatalog is now merged with Apache Hive.

- HIVE-4160: Added support for Vectorized Query Execution in Hive.
- HIVE-4677: Fixed HCatalog WebHCat end to end test failures on Hadoop 2.
- HIVE-2084: Upgraded DataNuclues from v2.0.3 to v3.0.1.
- HIVE-3815: Fixed failures for `hive table rename` operation when filesystem cache is disabled.
- HIVE-3846: Fixed null pointer exceptions (NPEs) for `alter view rename` operations when authorization is enabled.
- HIVE-3255: Added `DBTokenStore` to store Delegation Tokens in database.
- HIVE-4171: Current database in metastore. Hive is not consistent with `SessionState`.
- HIVE-4392: Fixed `Illogical InvalidObjectException`when using `mulit aggregate` functions with star columns.
- HIVE-4343: Fixed HiveServer2 with Kerberos - local task for map join fails.
- HIVE-4485: Fixed beeline prints null as empty strings.
- HIVE-4502: NPE - subquery smb joins fails.
- HIVE-4510: Fixed HiveServer2 nested exceptions.
- HIVE-4513: Added support to disable Hive history logs by default.
- HIVE-4521: Fixed auto join conversion failures
- HIVE-4540: Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- HIVE-4611: Fixed SMB join failures because of conflicts in bigtable selection policy.
- HIVE-5542: Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- HIVE-3255: Fixed Metastore upgrade scripts failures for PostgreSQL version less than 9.1.
- HIVE-4486: Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.

- Removed `npath` windowing function.
- HIVE-4465: Fixed issues for WebHCatalog end to end tests for the `exitvalue`.
- HIVE-4524: Added support for `Hive HBaseStorageHandler` to work with HCatalog.
- HIVE-4551: Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).
- HIVE-4246: Implemented predicate pushdown for ORC.
- HIVE-4579: Created a SARG interface for RecordReaders.
- HIVE-4660: Let there be Tez.
- HIVE-4478: In ORC, added boolean noNulls flag to column stripe metadata.

3.2.6. Patch information for HCatalog

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

3.2.7. Patch information for Oozie

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

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

3.2.8. Patch information for Sqoop

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

- SQOOP-931: HCatalog integration with Sqoop.
- SQOOP-916: Added an abort validation handler.
- SQOOP-798: Fixed failures for Ant docs on RedHat Enterprise Linux v5.8.

3.3. Minimum system requirements

In this section:

- Hardware Recommendations
- Operating Systems Requirements

- Software Requirements
- Database Requirements
- Virtualization and Cloud Platforms
- Configure the Local Repositories

3.3.1. Hardware Recommendations

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

3.3.2. Operating Systems Requirements

The following operating systems 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.

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

3.3.3. Software Requirements

On each of your hosts:

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

3.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and comes with embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with embedded Derby database by default.

3.3.5. Virtualization and Cloud Platforms

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

See Operating Systems Requirements for the list of supported operating systems for HDP.

3.3.6. 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.

3.4. Improvements

- Added support for Tez.
- Added support for ORCFile. For more details, see ORC File.
- Hadoop updated to version 2.1.0.
- Pig updated to version 0.11.1.
- Hive updated to version 0.11.0.
- Added support for Oozie.
- Added support for Sqoop.

3.5. Known Issues

In this section:

- Known Issues for Hadoop
- Known Issues for Hive

- Known Issues for HBase

3.5.1. Known Issues for Hadoop

- CapacityScheduler does not activate applications when configuration is refreshed. For more details, see YARN-474.

3.5.2. Known Issues for Hive

- Vectorization should be disabled for tables with unsupported columns types.

  Workaround: Workaround is to use supported column types (`tinyint`, `smallint`, `int`, `bigint`, `float`, `double`, `boolean`, and `timestamp`). To disable vectorization edit `hive-site.xml` file and update the following property:

  ```
  hive.vectorized.execution.enabled=false
  ```

- Mapreduce task from Hive dynamic partitioning query is killed.

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

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

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

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

  To change these values using the UI, use the instructions provided here to update these properties.

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

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

  the following error is reported in the Hive log file:

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

  Workaround: The workaround is to use the following query:

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

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

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

the following error is reported in the Hive log file:

```
NoViableAltException(15@[129:7: ( ( ( KW_AS )? identifier ) | ( KW_AS LPAREN
 identifier ( COMMA identifier )* RPAREN ) )?])
at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
at org.antlr.runtime.DFA.predict(DFA.java:116)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectItem(HiveParser_SelectClauseParser.java:2298)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectList(HiveParser_SelectClauseParser.java:1042)
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.
selectClause(HiveParser_SelectClauseParser.java:779)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.
java:30649)
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.
java:28851)
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.
java:28766)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.
java:28306)
at org.apache.hadoop.hive.ql.parse.HiveParser.
queryStatementExpression(HiveParser.java:28100)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.
java:1213)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:928)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.
java:39)
at sun.reflect.DelegatingMethodAccessorImpl.
invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'
 'from' in selection target
```

Workaround: The workaround is to use the following query:

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

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

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

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

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

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

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

3.5.3. Known Issues for HBase

- For HBase, `ReplicationZookeeper.copyQueuesFromRSUsingMulti` returns queues when it fails to execute. For more details, see HBASE-8099.

  The workaround is to disable the `hbase.zookeeper.useMulti` property.

- When using HBase, memory issues might happen if short circuit read feature in HDFS is enabled. The workaround is to disable short circuit reads in HDFS.

4. Release Notes HDP-2.0.0.2 (Alpha)

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

4.1. Product Version: HDP-2.0.0.2 (Alpha)

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

- Apache Hadoop 2.0.3
- Apache HBase 0.94.5
- Apache ZooKeeper 3.4.5
- Apache Pig 0.10.1
- Apache Hive 0.10.0+

Note

Hive is based on Apache Hive 0.10.0 plus the work in progress on Hive trunk, SVN revision 1251437, and additional patches as listed here.

- Tez 0.1
- Apache HCatalog 0.5.0

4.2. Patch Information

In this section:

4.2.1. Patch Information for Hadoop

Hadoop is based on Apache Hadoop 2.0.3 and includes the following additional patches:

- HDFS-4541: Set `hadoop.log.dir` and `hadoop.id.str` when starting secure DataNode to write the logs to correct directory by default.

- YARN-429: Added missing Capacity Scheduler configurations to yarn-test artifact.
- HDFS-4540: Fixed Namenode HTTP server to use the web authentication keytab for `spnego` principal.

4.2.2. Patch information for HBase

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

- HBASE-6338: Cache Method in RPC handler.
- HBASE-6134: Improved split-worker to enhance distributed log splitting.
- HBASE-6508: Added support to filter out edits at log split time.
- HBASE-7814: Port HBASE-6963 (Unable to run `hbck` on a secure cluster), to 0.94 branch.
- HBASE-7832: Added support to use `User.getShortName()` in `FSUtils`.
- HBASE-7851: Included the `guava` classes as a dependency for jobs using `mapreduce.TableMapReduceUtil`.
- HBASE-6466: Added support to enable multi-threading 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: Moved `isFamilyEssential(byte[] name)` out of `Filter` interface in HBase 0.94.
- HBASE-8007: Added `TestLoadAndVerify` from BigTop.

4.2.3. 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: Added `mvn-install` target for deploying the ZooKeeper artifacts to `.m2` repository.

4.2.4. Patch information for Pig

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

- PIG-3116: Fixed end to end tests sort command issues for RHEL-6.
- PIG-2885: Fixed test failures for `TestJobSumission` and `TestHBaseStorage`.
- PIG-3105: Fixed `TestJobSubmission` unit test failure.

- PIG-3099: Pig unit test fixes for `TestGrunt(1)`, `TestStore(2)`, `TestEmptyInputDir(3)`.
- 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-3248: Upgraded Hadoop-2.0.0-alpha to Hadoop-2.0.3-alpha.

4.2.5. Patch information for Hive

Hive is based on Apache Hive 0.10.0 plus the work in progress on Hive trunk, SVN revision 1251437, and includes the following:

- HIVE-2340: Optimized `orderby` operation followed by a `groupby` operation.
- HIVE-4143: Fixed incorrect column mappings with over clause.
- HIVE-4191: Fixed `describe table/show columns` for HiveServer2 merge.
- HIVE-896: Added `LEAD`, `LAG`, `FIRST`, and `LAST` analytical windowing functions to Hive.
- HIVE-4140: Added supprt to specify alias for windowing function.
- HIVE-4126: Removed support for lead/lag UDFs outside of UDAF args.
- HIVE-4124: Added tests for windowing.
- HIVE-4139: Fixed `MiniDFS shim`.
- HIVE-4127: Fixed test failures for ORCs `TestFileDump` for Hadoop 2.x.
- HIVE-4138: Fixed issue with ORC's union object inspector for `TypeInfoUtils`.
- HIVE-4081: Added support for expressions with `over` clause. .
- HIVE-4105: Fixed deserialization issues for Hive `MapJoinOperator` for the join-keys.
- HIVE-4015: Added ORC file as a file format to the grammar.
- HIVE-4098: Fixed issue with `OrcInputFormat` caused because `createValue` function is not invoked by default.
- HIVE-4097: Fixed issue for ORC file caused when `hive.io.file.readcolumn.ids` are empty.
- HIVE-3874: Created a new Optimized Row Columnar file format for Hive.
- HIVE-3952: Merged map-job followed by map-reduce job.
- HIVE-4106: Fixed multi-way joins failures for SMB joins.
- HIVE-4071: Fixed issues with Map-join outer join.
- HIVE-3996: Added support to enforce the memory limit on the multi-table map-join.

- HIVE-3891: Changes to physical optimizer for auto sort-merge join.
- HIVE-4103: Removed `System.gc()` method from the map-join local-task loop.
- HIVE-4105: Fixed Hive `localtask` that caused buffer issues for disk-writes or reads.
- HIVE-4094: Fixed failures for `decimal_3.q` and `decimal_serde.q` on Hadoop 2.x.
- HIVE-3846: Fixed failures for `alter view rename` operation that caused null pointer exception (NPE) when authorization is enabled.
- HIVE-2084: Upgraded DataNuclueus to v 3.0.1.
- HIVE-4182: Fixed issues for doAS to work with HiveServer2 in Non-Kerberos mode with local job.
- HIVE-3775: Fixed issues with testcases that have non-specified order of output.
- HIVE-3815: Fixed failures for Hive table `rename` operation when filesystem cache is disabled.
- HIVE-4167: Fixed conversion of Bucket Map join to SMB join caused when tables are not sorted.
- HIVE-3861: Upgraded HBase dependency to 0.94.2.
- HIVE-3862: Fixed `testHBaseNegativeCliDriver_cascade_dbdrop` test failures.
- HIVE-3708: Added MapReduce workflow information to job configuration.
- HIVE-2935: Implement HiveServer2.
- HIVE-3717: Fixed compilation errors when `-Dhadoop.mr.rev` property is set to 20S.
- HIVE-3937: Improved Hive Profiler.

4.2.6. Patch information for HCatalog

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

- HCATALOG-624: Fix HCatalog for for Hadoop 2.0.
- HCATALOG-555: HCatalog script should look for `hcatalog-core` JAR file and add `HCAT_PREFIX/conf` as a config location that is checked.
- HCATALOG-573: Removed version number from `WEBHCAT_JAR` in `webhcat_config.sh`.
- HCATALOG-631: Fixed HBase end to endtests on single nodes on Hadoop 2.0.3 when `dfs.client.read.shortcircuit` is enabled for HBase.

4.3. Minimum system requirements

In this section:

- Hardware Recommendations
- Operating Systems Requirements
- Software Requirements
- Database Requirements
- Virtualization and Cloud Platforms
- Configure the Local Repositories

4.3.1. Hardware Recommendations

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

4.3.2. Operating Systems Requirements

The following operating systems 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.

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

4.3.3. Software Requirements

On each of your hosts:

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

4.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x is supported. You may provide access to an existing database, or install MySQL instance using the instructions provided here.

- Oozie requires a database to use as a metadata store, but comes with embedded Derby database by default. MySQL 5.x is supported.

4.3.5. Virtualization and Cloud Platforms

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

See Operating Systems Requirements for the list of supported operating systems for HDP.

4.3.6. 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.

4.4. Improvements

- Added support for Tez.
- Added support for ORCFile. For more details, see ORC File.
- Hadoop updated to version 2.0.3 (Alpha).
- HBase updated to version 0.94.5.
- ZooKeeper updated to version 3.4.5.
- Pig updated to version 0.10.1.
- Hive updated to version 0.10.0.
- HCatalog updated to version 0.5.0

4.5. Known Issues

In this section:

- Known Issues for Hadoop
- Known Issues for Hive
- Known Issues for HBase

4.5.1. Known Issues for Hadoop

- CapacityScheduler does not activate applications when configuration is refreshed. For more details, see YARN-474.

4.5.2. Known Issues for Hive

- If `hive.auto.convert.join=true`, queries with MapJoin will fail with the following message:

```
FAILED: SemanticException [Error 10227]: Not all clauses are supported with
 mapjoin hint. Please remove mapjoin hint.
```

  If this property is set to false, most queries (except for the `union` and `union_all` operations) will accept MapJoin hints.

  If `hive.auto.convert.join=true`, for queries with `union` operations, it will result in the `SemanticException` error as shown above.

- Hive ORC files currently do not work with HCatalog. For more details, see HCATALOG-632.

- The Hive ORC file format in this Alpha release is NOT guaranteed to be compatible with future versions of Hive. A future version of Hive might not be able to read an ORC file created with the Hive version in this Alpha release.

  There is no intent to provide upgraders to convert ORC files from this Alpha release to the format used in the future. Future compatibility of ORC file formats will be supported after Hive is formally released by Apache.

- Queries with OVER and LEAD currently fail.

- Queries with a window specification (ie. using ROWS BETWEEN or RANGE BETWEEN) might return incorrect results.

- Queries with multiple ORDER BY columns in a window specification return incorrect results.

- Some queries with an OVER clause might take a long time to execute.

- Currently `Windowing_Checkin_2` test fails.

- For the OVER clause, some queries with GROUP BY, ORDER BY, and OVER clause produce incorrect results.

- Currently Hive provides incorrect default window frame in queries like: `select avg(a) over (partition by b order by c);`.

The workaround is to rewrite the query as `select avg(a) over (partition by b order by c rows unbounded preceding);`.

This query is semantically equivalent and produces correct results. For more details, see HIVE-4190

4.5.3. Known Issues for HBase

- For HBase, `ReplicationZookeeper.copyQueuesFromRSUsingMulti` returns queues when it fails to execute. For more details, see HBASE-8099.

  The workaround is to disable the `hbase.zookeeper.useMulti` property.

- When using HBase, memory issues might happen if short circuit read feature in HDFS is enabled. The workaround is to disable short circuit reads in HDFS.

5. Release Notes HDP-2.0.0.1 (Alpha)

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

5.1. Product Version: HDP-2.0.0.1 (Alpha)

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

- Apache Hadoop 2.0.2
- Apache HBase 0.94.2
- Apache Pig 0.9.2
- Apache ZooKeeper 3.4.3
- Apache HCatalog 0.4.0
- Apache Hive 0.9.0
- Apache Oozie 3.2.0
- Hortonworks Management Center (HMC) 2.0.0.1
- Third party components:
  - Ganglia 3.2.0
  - Nagios 3.2.3

5.2. Patch Information

Pig includes the following patches:

- PIG-2766: Improved Pig-HCatalog usability.
- PIG-2791: Fixed issue with using ViewFileSystem for Pig.

Hive includes the following patches:

- HIVE-2084: Upgraded DataNucleus to upstream version.
- HIVE-2918: Fixed issue with Hive Dynamic Partition Insert. The move task will now consider the value of **hive.exec.max.dynamic.partitions** parameter.
- HIVE-3008: Fixed memory leak in TUGIContainingTransport.
- HIVE-3063: Fixed failures when using drop partition for non-string columns.
- HIVE-3076: Fixed failures when using drop partition for non-partition columns.

- HIVE-3168: Fixed copy issue for **LazyBinaryObjectInspector.getPrimitiveJavaObject**.
- HIVE-3246: Fixed binary datatype for Java primitive type.
- HIVE-3153: Added improvements to reduce memory consumption for RCFile.Writer.
- HIVE-3291: Fixed issues with fs resolvers.
- HIVE-3098: Fixed memory leak issue caused by large number of FileSystem instances in FileSystem.CACHE.
- HIVE-2928: Added support for Oracle-backed Hive-Metastore (`longvarcharto clob` in `package.jdo`).
- HIVE-3082: Added support for Oracle Metastore schema script to include DDL for DataNucleus' internal tables.

HCatalog includes the following patches:

- HCATALOG-412: HCatalog now publishes artifacts to the local M2 cache.
- HCATALOG-410: Added support for proxy user in HCatalog client.
- HCATALOG-420: Backport HCATALOG-363 for 0.4 branch.
- 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: Updated datanucleus to upstream version.
- HCATALOG-350: Fixed dependecy for HCatRecord. Writing BINARY data to HCatRecord now does not depend on a Hive class.
- HCATALOG-436: Fixed incorrect naming for JSON SerDe column on CTAS.
- HCATALOG-471: Fixed Test HCat_ShowDes_1[1-3] failures.
- HCATALOG-375: Added support for HCatalog to be interoperable with Hadoop 0.23.

Oozie includes the following patches:

- OOZIE-698: Added support to configure version dependency for the sharelib components.
- OOZIE-810: Updated Oozie POM file to use Doxia 9.2y from the available repository.
- OOZIE-863: Fixed failures when invoking the `oozie-env.sh` script file. It is not required to explicitly set `JAVA_HOME` at client.
- OOZIE-968: Added support to source Oozie environment from conf in Oozie db setup script file.
- OOZIE-947: Forward porting OOZIE-733 to 3.2 and trunk.
- OOZIE-1006: Added support for Oozie to be interoperable with Hadoop 2.0.2.

Ambari includes the following patches:

- AMBARI-664: Fixed mapred io sort mb and heap size for Map/Reduce.
- AMBARI-641: Added support to change the location of Nagios' `status.dat` file according to the underlying platform.
- AMBARI-628: Fixed configuration and permission 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.
- AMBARI-701: Added support to handle the pre-setup user-supplied Hive Metastore.

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

- Hadoop updated to upstream version 2.0.2 (Alpha).
- HBase updated to upstream version 0.94.2.
- ZooKeeper updated to upstream version 3.4.3.

- Oozie updated to upstream version 3.2.0.
- HMC updated to upstream version 2.0.0.1.

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