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

List of Tables

1. Release Notes HDP-2.0.9.1

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

In this document:

- Product Version: HDP-2.0.9.1

- What's Changed in HDP-2.0.9.1

- Patch Information

- Improvements

- Known Issues

- Third-party Licenses

1.1. Product Version: HDP-2.0.9.1

All HDP 2.0.9.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.0.9.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.0.9.1.

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

- Apache Hadoop 2.2.0

- Apache HBase 0.96.1

- 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 1.4.4

- Apache Mahout 0.8.0

- Third party components:

  - Ganglia 3.5.0

  - Ganglia Web 3.5.7

  - Nagios 3.5.0

1.2. What's Changed in HDP-2.0.9.1

The following HDP components were changed in HDP-2.0.9.1:

- What's Changed in Ambari 1.4.4

1.2.1. What's Changed in Ambari 1.4.4

The following changes were made in Ambari 1.4.4:

ASF Jira Number	Summary
AMBARI-3754	Nagios won't start if HDFS is not selected
AMBARI-4244	NameNode start fails after moving it to another host
AMBARI-4266	Multiuser defaults for GlusterFS
AMBARI-4298	Ambari rpm target should take all DML scripts
AMBARI-4303	Foreign key constraints name too long - during Oracle schema upgrade
AMBARI-4359	Update GlusterFS defaults
AMBARI-4363	Support new version of ambari server in Ambari SCOM server
AMBARI-4391	During Oracle db upgrade, syntax error in Oracle DML scripts
AMBARI-4404	Mapreduce.task.io.sort.mb max value should not exceed 1024mb
AMBARI-4405	Remove property fs.checkpoint.size during upgrade
AMBARI-4418	Skip url replacement for windows build.
AMBARI-4420	ORA-01795: maximum number of expressions in a list is 1000 in Ambari Server log
AMBARI-4436	Ambari-server upgrade takes a long time if there are many command entries in the DB
AMBARI-4440	Package hadoop.sh script for RHEL HA in Ambari Release.
AMBARI-4446	Upgrade schema script for Oracle fails if upgrading from 1.4.1 or 1.4.2 to 1.4.4
AMBARI-4447	Hive and WebHcat won't start with 2.0.6.GlusterFS stack.
AMBARI-4455	Failure saving configuration in a new Host Config Group
AMBARI-4457	Strange behaviour of moving NameNode after enabling HA
AMBARI-4460	Upgrade schema script fails for MySQL if upgrading from M10 or M15 to M20
AMBARI-4463	Need better instruction when ambari-server upgrade fails as sql client is not available
AMBARI-4464	Panel "Secondary NameNode" exists on HDFS Config Tab after HA enabled
AMBARI-4479	Sync configs for core hadoop services for GlusterFS stack

1.3. Patch Information

In this section:

- Patch Information for Hadoop

- Patch Information for HBase

- Patch Information for ZooKeeper

- Patch Information for Pig

- Patch Information for Hive

- Patch Information for HCatalog

- Patch Information for Oozie

1.3.1. Patch Information for Hadoop

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

- HADOOP-10112: har file listing doesn't work with wild card.

- HADOOP-10110: hadoop-auth has a build break due to missing dependency.

- HDFS-5371: Let client retry the same NN when "dfs.client.test.drop.namenode.response.number" is enabled.

- HDFS-5257:addBlock() retry should return LocatedBlock with locations else client will get AIOBE.

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

1.3.2. Patch information for HBase

HBase is based on Apache HBase 0.96.1.

- HBASE-10210: During master startup, RS can be you-are-dead-ed by master in error

- HBASE-10188: Deprecate ServerName constructors, but make it public

1.3.3. 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.

1.3.4. Patch information for Pig

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

- PIG-3522: Remove JSCH jar from PIG

- 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.

1.3.5. Patch information for Hive

Hive is based on Apache Hive 0.12.0

Note

Apache HCatalog is now merged with Apache Hive.

- HIVE-6117: HBase_1 and HBase_2 tests are failing

- HIVE-5601: NPE in ORC's PPD when using select * from table with where predicate

- HIVE-5542: Webhcat is failing to run ddl command on a secure cluster

- HIVE-5515: Writing to an HBase table throws IllegalArgumentException, failing job submission

- HIVE-5511: `percentComplete` returned by job status from WebHCat is null

- HIVE-5496: hcat -e `drop database if exists` fails on authorizing non-existent null db

- HIVE-5485: SBAP errors on null partition being passed into partition level authorization

- HIVE-5484: TestSchemaTool failures when Hive version has more than 3 revision numbers

- HIVE-5480: WebHCat e2e tests for doAs feature are failing

- HIVE-5479:SBAP restricts `hcat -e 'show databases'`

- HIVE-5478: WebHCat e2e testsuite for hcat authorization tests needs some fixes

- HIVE-5474: drop table hangs when `concurrency=true`

- HIVE-5453: `jobsubmission2.conf` should use 'timeout' property

- HIVE-5448:webhcat duplicate test TestMapReduce_2 should be removed

- HIVE-5425: Provide a configuration option to control the default stripe size for ORC

- HIVE-5422: Upgrade 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.

1.3.6. Patch information for HCatalog

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

1.3.7. Patch information for Oozie

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

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

- OOZIE-1563: Fixed colt jar includes GPL licence.

1.3.8. 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.

1.4. Improvements

Apache Ambari update to version 1.4.4. This release of Apache Ambari includes the new features and improvements:

- Support for RHEL HA

- Improvements to Ambari Server upgrade scripts

1.5. Known Issues

In this section:

- Known Issues for Ubuntu 12.04

- Known Issues for SLES 11

- Known Issues for HDP

- Known Issues for HBase

- Known Issues for Oozie

- Known Issues for Hue

- Known Issues for Ambari

1.5.1. Known Issues for Ubuntu 12.04

- Hue does not support Ubuntu hosts at this time.

- Ambari does not support Ubuntu hosts at this time.

1.5.2. Known Issues for SLES 11

- 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
  ```

1.5.3. Known Issues for HDP

- EC2 m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

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

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

1.5.4. Known Issues for HBase

- HBASE-10304 Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString

  Problem: Some MapReduce jobs fail to launch. An exception similar to the following displays:

  ```
  Exception in thread "main" java.lang.IllegalAccessError: class com.google.
  protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.
  protobuf.LiteralByteString
   at java.lang.ClassLoader.defineClass1(Native Method)
   at java.lang.ClassLoader.defineClass(ClassLoader.java:792)
   at java.security.SecureClassLoader.defineClass(SecureClassLoader.java:142)
   at java.net.URLClassLoader.defineClass(URLClassLoader.java:449)
   at java.net.URLClassLoader.access$100(URLClassLoader.java:71)
   at java.net.URLClassLoader$1.run(URLClassLoader.java:361)
   at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
  ```

```
 at java.security.AccessController.doPrivileged(Native Method)
 at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
 at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
 at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
 at org.apache.hadoop.hbase.protobuf.ProtobufUtil.toScan(ProtobufUtil.
java:818)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
convertScanToString(TableMapReduceUtil.java:433)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:186)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:147)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:270)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:100)
...
```

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

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

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

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

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

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

1.5.5. Known Issues for Oozie

- 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.

- 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.

1.5.6. Known Issues for Hue

- Ubuntu hosts not supported at this time.

- 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.

1.5.7. Known Issues for Ambari

- Ambari does not support running or installing stacks on Unbuntu.

- The component version information displayed by Ambari is based on the Ambari Stack definition. If you have applied patches to the Stack and to your software repository, that component version might differ from the actual version installed. There is no functional impact on Ambari if the patch versions mismatch. If you have any questions on component versions, refer to the rpm version installed on the actual host.

- Missing LzoCodec settings in `core-site.xml` file

  Problem: After cluster install, the `io.compression.codecs` property in `$HADOOP_CONF_DIR/core-site.xml` was incorrect. It displays as:

  ```
  <property>
      <name>io.compression.codecs</name>
      <value>org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.
  compress.DefaultCodec</value>
   </property>
  ```

  Workaround: Use Ambari Web to modify the `io.compression.codecs` property. Select **Services > HDFS > Configs > Advanced** and modify to:

  ```
  <property>
      <name>io.compression.codecs</name>
      <value>org.apache.hadoop.io.compress.GzipCodec,com.hadoop.compression.
  lzo.LzoCodec,org.apache.hadoop.io.compress.DefaultCodec</value>
   </property>
  ```

  And add the `io.compression.codec.lzo.class` property to the **Custom** `core-site.xml` section:

  ```
  <property>
      <name>io.compression.codec.lzo.class</name>
      <value>com.hadoop.compression.lzo.LzoCodec</value>
  </property>
  ```

- Incorrect `hive.security.authorization.manager` property after upgrade to Ambari 1.4.4.

  Problem: After upgrading to Ambari 1.4.4, the `hive.security.authorization.manager` property in $HIVE_CONFIG_DIR/hive-site.xml is incorrect. It is set to:

  ```
  <property>
      <name>hive.security.authorization.manager</name>
      <value>org.apache.hcatalog.security.HdfsAuthorizationProvider</value>
  </property>
  ```

  Workaround: Use Ambari Web to modify `hive.security.authorization.manager` property to the correct value. Select **Services > Hive > Configs > Advanced** and make the following changes:

  ```
  <property>
      <name>hive.security.authorization.manager</name>
      <value>org.apache.hadoop.hive.ql.security.authorization.
  StorageBasedAuthorizationProvider</value>
  </property>
  ```

- Upgraded single-node cluster to HDP-2.0.6, may be missing yarn job summary entries.

  Problem: After upgrade of a cluster to HDP 2 using Ambari, you may notice that the yarn job summary entries are missing. This typically happens if the YARN ResourceManager host is shared with MapReduce2 components.

  Workaround: To fix this issue modify the `log4j.properties` file at `/etc/hadoop/conf` on the ResourceManager host by adding the following lines:

Note

Modify the value for `log4j.appender.RMSUMMARY.File` property to contain the actual value of `yarn_log_dir_prefix` and `yarn_user`. You can get the values from the latest global config type. Use `configs.sh` tool to read the global config type.

```
#
# Job Summary Appender
#
# Use following logger to send summary to separate file defined
 by
# hadoop.mapreduce.jobsummary.log.file rolled daily:
# hadoop.mapreduce.jobsummary.logger=INFO,JSA
#
hadoop.mapreduce.jobsummary.logger=${hadoop.root.logger}
hadoop.mapreduce.jobsummary.log.file=hadoop-mapreduce.jobsummary.
log
log4j.appender.JSA=org.apache.log4j.DailyRollingFileAppender
# Set the ResourceManager summary log filename
yarn.server.resourcemanager.appsummary.log.file=hadoop-mapreduce.
jobsummary.log
# Set the ResourceManager summary log level and appender
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.
logger}
#yarn.server.resourcemanager.appsummary.logger=INFO,RMSUMMARY

# To enable AppSummaryLogging for the RM,
# set yarn.server.resourcemanager.appsummary.logger to
# <LEVEL>,RMSUMMARY in hadoop-env.sh

# Appender for ResourceManager Application Summary Log
# Requires the following properties to be set
#     - hadoop.log.dir (Hadoop Log directory)
#     - yarn.server.resourcemanager.appsummary.log.file (resource
 manager app summary log filename)
#     - yarn.server.resourcemanager.appsummary.logger (resource
 manager app summary log level and appender)
log4j.appender.RMSUMMARY=org.apache.log4j.RollingFileAppender
log4j.appender.RMSUMMARY.File=[yarn_log_dir_prefix]/[yarn_user]/
${yarn.server.resourcemanager.appsummary.log.file}
log4j.appender.RMSUMMARY.MaxFileSize=256MB
log4j.appender.RMSUMMARY.MaxBackupIndex=20
log4j.appender.RMSUMMARY.layout=org.apache.log4j.PatternLayout
log4j.appender.RMSUMMARY.layout.ConversionPattern=%d{ISO8601} %p
 %c{2}: %m%n
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.appender.JSA.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss}
 %p %c{2}: %m%n
log4j.appender.JSA.DatePattern=.yyyy-MM-dd
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.logger.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=${yarn.server.resourcemanager.
appsummary.logger}
log4j.additivity.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=false
```

• Unable to start gmond process after upgrade to HDP 2.0.6 Stack from HDP 1.3.2 Stack.

Problem: gmond process fails to start on a host during an upgrade

Workaround: Use the following steps to work around the issue:

1. Log onto the host where gmond fails to start.

2. For the gmond process that fails go to the corresponding directory. For example for HDPSlaves go to:

   ```
   /var/run/ganglia/hdp/HDPSlaves/
   ```

3. Remove the PID in the directory.

4. Stop gmond.

   ```
   service hdp-gmond stop
   ```

5. Start gmond.

   ```
   service hdp-gmond start
   ```

- After upgrading to Ambari 1.4.2, `fs.checkpoint.size` needs to be in appropriate units of bytes, not GBs.

  Problem: Ambari 1.4.1 and earlier assumed this setting to be in GB. The setting is in bytes.

  Workaround: Modify the `fs.checkpoint.size` property using Ambari Web. Select **Services** > **HDFS** > **Configs** > **General** and enter an appropriate integer value in bytes to set the HDFS maximum edit log size for checkpointing. For example, 500000000.

- Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start.

  Problem: The Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start. When the client state became became `installed_and_configured` after Service Start:

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

1.6. Third-party Licenses

Table 1.1. Third-party Licenses

HDP Component	Project Library	Version	License
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-2.0.9.0

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

In this document:

- Product Version: HDP-2.0.9.0

- What's Changed in HDP-2.0.9.0

- Patch Information

- Minimum system requirements

- Improvements

- Known Issues

- Third-party Licenses

2.1. Product Version: HDP-2.0.9.0

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

- Apache Hadoop 2.2.0

- Apache HBase 0.96.1

- 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 1.4.3

- Apache Mahout 0.8.0

- Third party components:

  - Ganglia 3.5.0

  - Ganglia Web 3.5.7

  - Nagios 3.5.0

2.2. What's Changed in HDP-2.0.9.0

The following HDP components were changed in HDP-2.0.9.0:

- HDP changes to 2.0.6.1

- What's Changed in HBase 0.96.1

- What's Changed in this update to Hadoop 2.2.0

- What's Changed in this update to Hive 0.12.0

- What's Changed in this update to Pig 0.12.0

- What's Changed in this update to Oozie 4.0.0

- What's Changed in Ambari 1.4.3

2.2.1. HDP changes to 2.0.6.1

HBase was updated to verson 0.96.1, causing updates to HDP 2.0.6.1 stack for following HDP components:

- Apache Flume 1.4.0

- Apache Hive 0.12.0

- Apache Mahout 0.8.0

- Apache Pig 0.12.0

- Apache Sqoop 1.4.4

To upgrade from HDP 2.0.6.0 to HDP 2.0.6.1, you can follow the HDP 1 to HDP 2 upgrade instructions with the following exceptions.

Do NOT uninstall Oozie as it says here here. Instead run the command:

```
yum erase hadoop-pipes hadoop-sbin hadoop-native
```

And upgrade Oozie instead of re-installing it as it says here. Run

- For RHEL/CentOS:

```
yum upgrade oozie
```

• For SLES:

```
zypper upgrade oozie
```

• For Ubuntu:

```
apt-get updrade oozie
```

2.2.2. What's Changed in HBase 0.96.1

HBase moved to version 0.96.1 and includes the following updates:

Apache JIRA	Summary
HBASE-9750	hbase master failed to start while running chaosmonkey
HBASE-9822	LazyCF test fails with Read 39858 keys; the writer is done and 39859 keys were generated
HBASE-10085	HBase Chaos Monkey Tests and Thrift test all Fail Due to Namespace Manager Error
HBASE-10210	Region server sometimes dies after restart with YouAreDeadException

2.2.3. What's Changed in this update to Hadoop 2.2.0

There was an update to HDFS.

Apache JIRA	Summary
HDFS-5257	Could not complete file error while running load generator on a HA cluster

2.2.4. What's Changed in this update to Hive 0.12.0

Hive did not change version, but was updated due to dependencies on HBase 0.96.1.

Apache JIRA	Summary
HIVE-5515	Writing to an HBase table throws IllegalArgumentException, failing job submission
HIVE-5601	NPE in ORC's PPD when using select * from table with where predicate
HIVE-6117	HBase_1 and HBase_2 tests are failing

2.2.5. What's Changed in this update to Pig 0.12.0

Pig did not change version, but was updated due to dependencies on HBase 0.96.1.

Apache JIRA	Summary
PIG-3522	Remove JSCH jar from PIG

2.2.6. What's Changed in in this update to Oozie

Oozie did not change version, but was updated due to dependencies on HBase 0.96.1.

Apache JIRA	Summary
OOZIE-1563	Remove Colt.jar from the Oozie RPMs

2.2.7. What's Changed in Ambari 1.4.3

The following changes were made in Ambari 1.4.3:

Apache JIRA	Summary
AMBARI-3865	Security Wizard: Add security Step4 controller and Disable security controller can be extended from same parent co
AMBARI-3729	Ganglia monitor started with second or third attempt on secure cluster
AMBARI-3531	Provide host-group and host-group config overrides support
AMBARI-4058	ambari-agent should start automatically upon reboot
AMBARI-3927	Add support for MySQL Ambari DB
AMBARI-3100	dfs.journalnode.edits.dir value should be determined on web-client
AMBARI-3933	ctrl^c during ambari-server setup "Initializing..." leaves postgres in bad state
AMBARI-3628	Restart indicators for services and hosts needed
AMBARI-3727	Need ability to restart host_components in bulk (on services and hosts)
AMBARI-3732, AMBARI-3745	Provide dialog to create new config-group by overriding property
AMBARI-3767	Provide basic config-group management dialog
AMBARI-3751	Provide config-group management bar for services
AMBARI-3765	Create property filtering capability for config-groups
AMBARI-3824	Provide change config-group action on host configs
AMBARI-3883	Provide config-groups capability in installer
AMBARI-3760	Provide config-group support in add-host wizard
AMBARI-3849	Need ability to filter out href field in requests
AMBARI-3725	Check for overlaps in metrics between JMX and Ganglia try to get all point in time metrics from JMX.
AMBARI-3755	Config Refactor: remove unused global properties
AMBARI-3867	JS error after switch to Dashboard page during HDFS starting
AMBARI-3737	HBase -> Configs -> RegionServer -> HBase RegionServers maximum Java heap size: adding Exception fails
AMBARI-3619	getAssociatedResource throws Exception on POST
AMBARI-3650	Poll for host_components which have stale_configs
AMBARI-4110	Avoid duplicating host names in "clusterHostInfo" when preparing tasks
AMBARI-3751	Load config_groups from API when in config pages
AMBARI-3770	Need better error log message when agent unable to reach server
AMBARI-3809	Config group dropdown is empty, and is shown incorrectly.
AMBARI-3736	HBase data directory in hdfs should be parsed from hbase.rootdir property.
AMBARI-3774	Provide ability to add/remove hosts in manage-config-groups dialog
AMBARI-3791	Provide add/remove/rename/duplicate actions in manage-config-groups dialog
AMBARI-3726	Restart indicators for services and hosts disappear after some time.
AMBARI-3890	Background operations: two scrollbars, if width is lower then 1450px
AMBARI-3759	Add host wizard: After successfully bootstrapping host, "next" button is disabled
AMBARI-3772	New config group dialog should should have config-groups populated
AMBARI-3776	Non-default config-group configs should have +Exception action
AMBARI-3796	Config-groups configs should be saved efficiently
AMBARI-3783	Hitting Save in the config-group does not update host membership
AMBARI-3794	Clicking +Exception in Default config-group doesnt add new row
AMBARI-3780	password.dat file should tolerate newlines
AMBARI-3977	Dont request all tasks fields during deploy
AMBARI-3803	For config-groups, rename +Exception action to +Override
AMBARI-3827	Minor message fixes in select/create config group dialog

AMBARI-3831	Add/Remove/Rename/Duplicate action changes not reflected in configs page
AMBARI-3804	Focussing on overridden config field throws JS exception
AMBARI-3815	Remove, Rename and Delete actions enabled for 'Default' config group
AMBARI-3807	Cannot save config-group due to missing service id
AMBARI-3819	Adding hosts to a config-group erases tags and vice-versa
AMBARI-3813	Can override config twice using config-groups
AMBARI-3862	Configs text filter not matching overridden field values
AMBARI-3806	Cannot save Default configs when overridden config present
AMBARI-3816	UI tweaks with stop/start actions in stale_config restart indicator
AMBARI-3832	UI Tweaks in Create Config Group dialog
AMBARI-3835	Manage Configuration Group needs various UI tweaks
AMBARI-3822	Selecting newly created config-group throws JS error in manage config-groups dialog
AMBARI-3833	Duplicate and Rename config-group action UI tweaks
AMBARI-3863	Provide validation for Config Group name
AMBARI-3840	Delete should be disabled if there is no group selected
AMBARI-3844	Error in saving host for newly created config group
AMBARI-3842	Host configs page should properly order services
AMBARI-3838	Services sidebar for host configs needs vertical gap
AMBARI-3836	Unable to close manage-config-groups dialog when only Default group present
AMBARI-3850	Config-groups should be sorted in manage config-groups dialog
AMBARI-3860	Should allow selecting hosts by components in config-groups dialog
AMBARI-3856	Host selection dialog title and description should be passed in
AMBARI-3851	Remove host action is always enabled
AMBARI-3839	DataNode heapsize config override causes DataNode to not start
AMBARI-3861	Modify hover text information for restart indicator
AMBARI-3837	NameNode HA wizard: Configuration changes should be surfaced on ui
AMBARI-3841	After deleting config group, hosts of it become unavailable
AMBARI-3859	No config groups warning should be aligned
AMBARI-3855	YARN Client restart indicator not shown on host configs page
AMBARI-3869	Better messages needed for restart components popup
AMBARI-3876	Add hosts wizard: Hosts should be put into config-groups before installing
AMBARI-3847	Service config properties marked as non-editable does not show description on hover
AMBARI-3915	Stop/Start components action in service configs not calculating correctly
AMBARI-3870	Config groups - Read only configs are being made editable
AMBARI-3902	Admin page->users: password confirmation while adding new user should be computed property
AMBARI-3999	Long host names are inconvenient for viewing in background operations popup
AMBARI-3877	Duplicate config-group action not duplicate configs
AMBARI-3872	Properties with checkboxes are always checked
AMBARI-3886	Config group name validation doesn't include newly added and removed goups
AMBARI-3904	Remove/Rename/Duplicate action changes don't update the link to the Config group
AMBARI-3879	Enable config-groups functionality by default
AMBARI-3889	Post config groups to server on Deploy
AMBARI-3885	Connect config filter to config group tab
AMBARI-3888	Incorrect restart required tooltip view
AMBARI-4020	On Saving Config group, user should be able to modify configs right away
AMBARI-3905	Config groups: Misc UI tweaks
AMBARI-3912	No override for NM heap memory

AMBARI-3946	Install Wizard: not possible to remove config overrides
AMBARI-3913	Description should be shown properly in Manage Config Groups dialog
AMBARI-3996	Installer wizard Config Group: Refreshing the page after saving a config group doesn't persist groups on step-7
AMBARI-3943	Installer: config-group host membership in one service effects others
AMBARI-3967	Config Groups: some fields missing in Select Group Hosts popup
AMBARI-3951	No link for Manage Config groups at Reconfigure sreen
AMBARI-3925	Adding host to multiple groups at the same time fails
AMBARI-3901	Config Groups: restart components should also include clients
AMBARI-3895	Stale_configs flag not updated when host moves back to default config-group
AMBARI-3917	Config Groups: 'Restart required' bar still show up on service config even though the target host had been deleted
AMBARI-3894	Cannot install HDP stack without having config-groups
AMBARI-3897	Restart indicator flags not showing up for HDP 1.3.x stack services
AMBARI-3903	Unit Test broken on trunk
AMBARI-4009	Config Groups: host got lost if fail to add it to new group
AMBARI-3908	Disable 'Add Property...' action in non-default config-groups
AMBARI-3928	Config Groups: Navi tab wrong when going back from host config page
AMBARI-3907, AMBARI-3918	Restart message should have correct message for single hosts/components
AMBARI-4041	Host list changing order in Move wizard
AMBARI-3921	Hovers stay after manage-config-groups dialog is closed
AMBARI-3911	Security Wizard: Service Configuration page is broken
AMBARI-3974	Custom config disappears upon config override
AMBARI-3919	When managing configs for a non-default group, the default configs are non-editable but "Remove" button is still th Custom sections
AMBARI-3914	Add Host wizard stuck on configuration step
AMBARI-3932	Config Group names can be named without any restriction. Need regex for config group name
AMBARI-3916	When trying to duplicate Default Group js error occures.
AMBARI-3923	Stale configs on Clients don't affect on Start/Stop Components buttons
AMBARI-3938	JS error when switching config groups in Hive / Oozie service config pages
AMBARI-3929	AddHost confg-groups sorting
AMBARI-3937	Revise Config Group Management copy to be more descriptive
AMBARI-3944	Service Check tasks are shown incorrectly
AMBARI-3945	Custom hdfs-site.xml section disappears in HDFS Config page
AMBARI-4067	UI improvements for Config groups.
AMBARI-4127	Hosts Filter does not say that a filter is enabled (not possible to figure out if a filter is enabled or not easily)
AMBARI-3942	Click Overview, create new accepts empty name
AMBARI-3940	Manage config groups Gear is clickable but doesn't show dropdown
AMBARI-4018	config group description doesn't preserve multi-line
AMBARI-3955	"Default" config group should be "<Service-name> Default"
AMBARI-3950	Avoid loading extra host-components collections
AMBARI-3957	Add mapper only for services
AMBARI-3992	After making config changes w/o saving, prompt user if they try to navigate away
AMBARI-3968	Installer - Removal of parent config-group custom-config possible
AMBARI-3971	Should not allow remove custom-config if overridden
AMBARI-4049	Registration fail of host should be handled more gracefully
AMBARI-3961	Add flag to enable minimal responses
AMBARI-3991	Manage config group links needed in save config-group confirmation
AMBARI-3973	Reduce influence of host-components quantity on service metrics mapper

AMBARI-4019	On Move Wizard, component layout should show non-moving items as gray
AMBARI-3982	Background operations window, called from wizard doesn't react on 'Do not show this dialog...' flag
AMBARI-4002	Incorrect navigation between pages
AMBARI-4008	Config group names should be unique in a service, not global
AMBARI-4069	Add hosts, if using Local Repository, UI incorrectly says "no"
AMBARI-3984	Config Groups: Background popup show up needs to be integrated when restarting components
AMBARI-3986	YARN and MapReduce2 configs is not displayed
AMBARI-3997	Config-Group POST call should tolerate name reuse
AMBARI-4092	Need tooltip showing error why local repo is bad
AMBARI-4007	In installer overriding config launches the service save confirmation popup
AMBARI-4011	Config description popups linger after filtering
AMBARI-4005	Need to add green color to config-filtering combo entry
AMBARI-4000	mapred-queue-acls properties are shown in mapreduce service advanced section.
AMBARI-4073	Do not block the user from moving forward if the local repo urls dont work - we can WARN at best.
AMBARI-4001	Ambari alerts not even being got from API
AMBARI-4004	Duplicate hosts after closing addServiceWizard
AMBARI-4085	Remove console.logs, which overflow browser console
AMBARI-4046	JS errors after stopping YARN
AMBARI-4029	Installer Step2: Can't proceed next after return from Step7
AMBARI-4063	In installer launching manage config-group dialog throws error in console
AMBARI-4043	In installer, default config-group ends up ordered last in the list of config-groups
AMBARI-3040	In installer, duplicating config-group does not duplicate configs
AMBARI-4040	In installer, behavior of actions in manage config-groups dialog different from reconfigure
AMBARI-4026	hbase.rootdir property should be defined in server side stack definition
AMBARI-4143	Hive Metastore warehouse dir and WebHCat apps dir path is being retrieved from wrong property name
AMBARI-4062	Components in Select Group Hosts popup
AMBARI-4050	Installer: 1-3 steps enabled when deploy is running
AMBARI-4038	Local stack upgrade should update the database as well.
AMBARI-4053	Duplicating config-group in installer doesnt update name in default config-group rendering
AMBARI-4044	Refactor templates and popups
AMBARI-4051	Delete Group button is disabled in Manage Dialog in Installer
AMBARI-4077	Optimize Alert mapper processing
AMBARI-4047	Improve responsiveness of hosts filter in config groups popup
AMBARI-4081	ambari agent unit test failures - TestLiveStatus and TestStackVersionsFileHandler
AMBARI-4109	Manage config-groups dialog should persist only hitting Save (installer and reconfigure)
AMBARI-4048	Minor manage-config-group dialog UI changes
AMBARI-4086	Need clear filter action in configs filter textfield
AMBARI-4088	UI doesn't select all hosts on Assign Slaves step sometimes
AMBARI-4054	Need to show stale-config indicator on hosts page
AMBARI-4052	Rename config-group dialog should allow only description change also
AMBARI-4060	Invalid JournalNodesLive count
AMBARI-4060	HDFS/JournalNodes relations in EmberData
AMBARI-4061	Upgraded single-node M10 cluster has yarn job summary entries missing in /etc/hadoop/conf/log4j.properties
AMBARI-4113	Bootstrap fails without a meaningful message
AMBARI-4129	HA wizard not accessible after upgrade
AMBARI-4089	HDFS/ZKFC relations in EmberData
AMBARI-4074	Ambari sets mapred.healthChecker.script.path and mapred.job.tracker.persist.jobstatus.dir incorrectly

AMBARI-4097	Undo port customization code on branch-1.4.3	
AMBARI-4106	Should not allow blank config group name	
AMBARI-4114	JS error on HA Step 3	
AMBARI-4087	Lazy loading doesn't stop even after data is loaded	
AMBARI-4121	warning on delete host should be at top of dialog	
AMBARI-4093	Installer:Confirm Hosts - improve responsiveness on filtering of hosts	
AMBARI-4102	Host components not showing up or manage'able	
AMBARI-4124	Move masters dialog UI has many problems	
AMBARI-4122	restart icon on hosts page should be same color as used in hosts details + services	
AMBARI-4119	Host Detail page: Add Component acts randomly	
AMBARI-4136	Unsaved Config Changes prompt should say Save, Discard, Cancel	
AMBARI-4184	Update HDP 2.0.6 stack components to match component versions	
AMBARI-4101	OS type check logic is not going to work with centos 6.5+	
AMBARI-4139	Big padding of MapReduce service on Services page if alerts are present	
AMBARI-4125	Ambari MySQL should not configure schema automatically	
AMBARI-4163	Ambari host registering freezes if wrong ssh key provided.	
AMBARI-4132	Stale config indicator not shown when reconfiguring WebHCat, Ganglia, Nagios, ZooKeeper	
AMBARI-4142	Incorrect names for service check operations on "Install, Start and Test" page	
AMBARI-4138	Possibility duplicate one group many times with the same name	
AMBARI-4130	Deleted host still appears in configuration groups	
AMBARI-4161	Move calculation of host status from HostResourceProvider to HeartBeatHandler	
AMBARI-4144	Java heap size, and all other General properties cannot be overridden for all services	
AMBARI-4133	Perf issues on Hosts page - freezes for several seconds and then unfreezes repeatedly on a large cluster	
AMBARI-4140	Install wizard step 5. Components column covers add and remove icons.	
AMBARI-4177	YARN check execute fails on install (CentOS 5)	
AMBARI-4152	Service Config page shows a blank page after enabling security (JS error)	
AMBARI-4151	Incorrect alignment for "Stop Components" and "Start Components" buttons on "Config" tab of any service after property changing	
AMBARI-4175	ambari-server reset doesn't work for MySQL as Ambari DB	
AMBARI-4166	Improve responsiveness on step 9 when deploy in progress	
AMBARI-4169	Default Config Group Duplicate UI not clear	
AMBARI-4191, AMBARI-4202	Action buttons are disabled on Host Details page even though heartbeat is not lost	
AMBARI-4164	Unexpected routing while clicking on Repositories Tab	
AMBARI-4170	Graph time-header missing	
AMBARI-4172	Cluster deploy: inconsistent color of status message	
AMBARI-4179	Ambari UI should not allow to override pid and log dirs	
AMBARI-4189	'DataNode maximum Java heap size' and 'Hadoop maximum Java heap size' overrides for config groups didn't get p during installer	
AMBARI-4182	Reduce number of transactions in HostResourceProvider, HostComponentResourceProvider	
AMBARI-4190	Hosts page shows ip address as "OS NOT SUPPORTED"	
AMBARI-4206	Significant lag between host status update and slave/master component start/stop	
AMBARI-4192	No HBase Masters shown as Active on UI	
AMBARI-4188	Non-admin user can manage config groups	
AMBARI-4195	Misalignment of "Filter" combobox on installer when browser window is narrow	
AMBARI-4193	Assign Masters: plus/minus color and hover cursor are off	
AMBARI-4218	Host config tab layout incorrect	
AMBARI-4201	Enable caching in HostConfigMappingDAO	

AMBARI-4215	Enable caching in ConfigGroupHostMappingDAO
AMBARI-4225	Starting/Stopping components based on restart indicator floods request history and execution queue
AMBARI-4211	Empty content of review step in Add Host wizard
AMBARI-4214	Need to change prompt during mysql setup
AMBARI-4224	When issuing Start/Stop of host components then predicate stale_config=true does not work
AMBARI-4219	Cannot save two config groups for different services but same config type
AMBARI-4244	NameNode start fails after moving it to another host

2.3. Patch Information

In this section:

- Patch Information for Hadoop

- Patch Information for HBase

- Patch Information for ZooKeeper

- Patch Information for Pig

- Patch Information for Hive

- Patch Information for HCatalog

- Patch Information for Oozie

2.3.1. Patch Information for Hadoop

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

- HDFS-5257:addBlock() retry should return LocatedBlock with locations else client will get AIOBE

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

- Datanodes fail to register with namenode due to minimum version check.

2.3.2. Patch information for HBase

HBase is based on Apache HBase 0.96.1.

- HBASE-10188: Deprecate ServerName constructors, but make it public

- HBASE-10210: During master startup, RS can be you-are-dead-ed by master in error

2.3.3. 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.

2.3.4. Patch information for Pig

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

- PIG-3522: Remove JSCH jar from PIG

- 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.

2.3.5. Patch information for Hive

Hive is based on Apache Hive 0.12.0

> **Note**
>
> Apache HCatalog is now merged with Apache Hive.

- HIVE-6117: HBase_1 and HBase_2 tests are failing

- HIVE-5601: NPE in ORC's PPD when using select * from table with where predicate

- HIVE-5542: Webhcat is failing to run ddl command on a secure cluster

- HIVE-5515: Writing to an HBase table throws IllegalArgumentException, failing job submission

- HIVE-5511: `percentComplete` returned by job status from WebHCat is null

- HIVE-5496: hcat -e `drop database if exists` fails on authorizing non-existent null db

- HIVE-5485: SBAP errors on null partition being passed into partition level authorization

- HIVE-5484: TestSchemaTool failures when Hive version has more than 3 revision numbers

- HIVE-5480: WebHCat e2e tests for doAs feature are failing

- HIVE-5479:SBAP restricts hcat `-e 'show databases'`

- HIVE-5478: WebHCat e2e testsuite for hcat authorization tests needs some fixes

- HIVE-5474: drop table hangs when `concurrency=true`

- HIVE-5453: `jobsubmission2.conf` should use 'timeout' property

- HIVE-5448:webhcat duplicate test TestMapReduce_2 should be removed

- HIVE-5425: Provide a configuration option to control the default stripe size for ORC

- HIVE-5422: Upgrade 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.

2.3.6. Patch information for HCatalog

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

2.3.7. Patch information for Oozie

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

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

- OOZIE-1563: Fixed colt jar includes GPL licence.

2.3.8. 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.

2.4. Minimum system requirements

In this section:

- Hardware Recommendations

- Operating Systems Requirements

- Software Requirements

- Database Requirements

- Virtualization and Cloud Platforms

- Configure the Local Repositories

2.4.1. Hardware Recommendations

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

2.4.2. Operating Systems Requirements

The following operating systems are supported:

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

- 64-bit CentOS v5.*, v6.*

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

- 64-bit Oracle Linux v5, v6

- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

- 64-bit Ubuntu Precise 12.04

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

2.4.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

- php-curl (Required for SLES installs.)

2.4.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.

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

- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

2.4.5. Virtualization and Cloud Platforms

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

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

2.4.6. 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.

2.5. Improvements

Apache Ambari update to version 1.4.3. This release of Apache Ambari includes the new features and improvements:

- Ability to create Host Configuration Groups

- Ability to save configuration changes without stopping a service

- Added MySQL Support for Ambari DB

2.6. Known Issues

In this section:

- Known Issues for Ubuntu 12.04

- Known Issues for SLES 11

- Known Issues for HDP

- Known Issues for HBase

- Known Issues for Oozie

- Known Issues for Hue

- Known Issues for Ambari

2.6.1. Known Issues for Ubuntu 12.04

- Hue does not support Ubuntu hosts at this time.

- Ambari does not support Ubuntu hosts at this time.

2.6.2. Known Issues for SLES 11

- php_curl Required for SLES 11 Sp1

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

  Workaround: Install php_curl on your SLES host.

  ```
  zypper install php-curl
  ```

2.6.3. Known Issues for HDP

- EC2 m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

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

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

2.6.4. Known Issues for HBase

- HBASE-10304 Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString

  Problem: Some MapReduce jobs fail to launch. An exception similar to the following displays:

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

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

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

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

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

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

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

2.6.5. Known Issues for Oozie

- 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.

- 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.

2.6.6. Known Issues for Hue

- Ubuntu hosts not supported at this time.

- 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.

2.6.7. Known Issues for Ambari

- Ambari does not support running or installing stacks on Unbuntu.

- The component version information displayed by Ambari is based on the Ambari Stack definition. If you have applied patches to the Stack and to your software repository, that component version might differ from the actual version installed. There is no functional impact on Ambari if the patch versions mismatch. If you have any questions on component versions, refer to the rpm version installed on the actual host.

- Upgrading Ambari server from v1.3.x/1.4.1 to 1.4.3 using local repository may fail

  Problem: Upgrading the ambari server using a local repository may fail. Error displays as:

  "Checking database integrity... Database is consistent. Adjusting ambari-server permissions and ownership... ERROR: Exiting with exit code 1. Reason: /var/lib/ambari-server/resources/upgrade/dml/Ambari-DML-Postgres-FIX_LOCAL_REPO.sql: No such file or directory."

  Workaround: Load fix scripts on the Ambari server, in the following directory:

  ```
  cd /var/lib/ambari-server/resources/upgrade/dml
  ```

  Download fix scripts, using the following commands:

  ```
  wget https://raw2.github.com/apache/ambari/branch-1.4.3/ambari-server/src/
  main/resources/upgrade/dml/Ambari-DML-Postgres-FIX_LOCAL_REPO.sql
  wget https://raw2.github.com/apache/ambari/branch-1.4.3/ambari-server/src/
  main/resources/upgrade/dml/Ambari-DML-Oracle-FIX_LOCAL_REPO.sql
  ```

- Missing LzoCodec settings in `core-site.xml` file

  Problem: After cluster install, the `io.compression.codecs` property in `$HADOOP_CONF_DIR/core-site.xml` was incorrect. It displays as:

  ```
  <property>
      <name>io.compression.codecs</name>
      <value>org.apache.hadoop.io.compress.GzipCodec,org.apache.hadoop.io.
  compress.DefaultCodec</value>
   </property>
  ```

  Workaround: Use Ambari Web to modify the `io.compression.codecs` property. Select **Services** > **HDFS** > **Configs** > **Advanced** and modify to:

  ```
  <property>
  ```

```
    <name>io.compression.codecs</name>
    <value>org.apache.hadoop.io.compress.GzipCodec,com.hadoop.compression.
lzo.LzoCodec,org.apache.hadoop.io.compress.DefaultCodec</value>
 </property>
```

And add the `io.compression.codec.lzo.class` property to the **Custom** `core-site.xml` section:

```
<property>
    <name>io.compression.codec.lzo.class</name>
    <value>com.hadoop.compression.lzo.LzoCodec</value>
</property>
```

- Exception shown when deleting Host Config Group.

  Problem: If a host is associated with a Host Config Group and you delete that host from the cluster, you see the following exception dialog:

  ```
  Transaction rolled back because transaction was set to RollbackOnly. 500
  status code.
  ```

  The host is deleted from the cluster as expected, but the Config Group still shows the deleted host association.

  Workaround: Delete the host from the Config Group.

- Incorrect `hive.security.authorization.manager` property after upgrade to Ambari 1.4.3.

  Problem: After upgrading to Ambari 1.4.3, the `hive.security.authorization.manager` property in $HIVE_CONFIG_DIR/hive-site.xml is incorrect. It is set to:

  ```
  <property>
      <name>hive.security.authorization.manager</name>
      <value>org.apache.hcatalog.security.HdfsAuthorizationProvider</value>
  </property>
  ```

  Workaround: Use Ambari Web to modify `hive.security.authorization.manager` property to the correct value. Select **Services** > **Hive** > **Configs** > **Advanced** and make the following changes:

  ```
  <property>
      <name>hive.security.authorization.manager</name>
      <value>org.apache.hadoop.hive.ql.security.authorization.
  StorageBasedAuthorizationProvider</value>
  </property>
  ```

- Upgraded single-node cluster to HDP-2.0.6, may be missing yarn job summary entries.

  Problem: After upgrade of a cluster to HDP 2 using Ambari, you may notice that the yarn job summary entries are missing. This typically happens if the YARN ResourceManager host is shared with MapReduce2 components.

  Workaround: To fix this issue modify the `log4j.properties` file at `/etc/hadoop/conf` on the ResourceManager host by adding the following lines:

Note

Modify the value for `log4j.appender.RMSUMMARY.File` property to contain the actual value of `yarn_log_dir_prefix` and `yarn_user`. You can get the values from the latest global config type. Use `configs.sh` tool to read the global config type.

```
#
# Job Summary Appender
#
# Use following logger to send summary to separate file defined
 by
# hadoop.mapreduce.jobsummary.log.file rolled daily:
# hadoop.mapreduce.jobsummary.logger=INFO,JSA
#
hadoop.mapreduce.jobsummary.logger=${hadoop.root.logger}
hadoop.mapreduce.jobsummary.log.file=hadoop-mapreduce.jobsummary.
log
log4j.appender.JSA=org.apache.log4j.DailyRollingFileAppender
# Set the ResourceManager summary log filename
yarn.server.resourcemanager.appsummary.log.file=hadoop-mapreduce.
jobsummary.log
# Set the ResourceManager summary log level and appender
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.
logger}
#yarn.server.resourcemanager.appsummary.logger=INFO,RMSUMMARY

# To enable AppSummaryLogging for the RM,
# set yarn.server.resourcemanager.appsummary.logger to
# <LEVEL>,RMSUMMARY in hadoop-env.sh

# Appender for ResourceManager Application Summary Log
# Requires the following properties to be set
#     - hadoop.log.dir (Hadoop Log directory)
#     - yarn.server.resourcemanager.appsummary.log.file (resource
 manager app summary log filename)
#     - yarn.server.resourcemanager.appsummary.logger (resource
 manager app summary log level and appender)
log4j.appender.RMSUMMARY=org.apache.log4j.RollingFileAppender
log4j.appender.RMSUMMARY.File=[yarn_log_dir_prefix]/[yarn_user]/
${yarn.server.resourcemanager.appsummary.log.file}
log4j.appender.RMSUMMARY.MaxFileSize=256MB
log4j.appender.RMSUMMARY.MaxBackupIndex=20
log4j.appender.RMSUMMARY.layout=org.apache.log4j.PatternLayout
log4j.appender.RMSUMMARY.layout.ConversionPattern=%d{ISO8601} %p
 %c{2}: %m%n
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.appender.JSA.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss}
 %p %c{2}: %m%n
log4j.appender.JSA.DatePattern=.yyyy-MM-dd
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.logger.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=${yarn.server.resourcemanager.
appsummary.logger}
log4j.additivity.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=false
```

• Hive check execute fails after upgrading from Ambari 1.4.1 to Ambari 1.4.2.

Problem: When Ambari is upgraded to 1.4.2 and security is enabled, the Hive service check can fail due to a conflicting combination of authorization properties.

Workaround: Disable authorization. Using Ambari UI, set `hive.security.authorization.enabled` to false. Or, verify that the correct combination of authorization properties are used. For example:

```
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
```

• Enable NameNode HA wizard freezes after you add a host to the cluster.

Problem: After upgrading from Ambari 1.4.1 to Ambari 1.4.3 and from HDP 1.3.3 to HDP 2.0.6 Stack, the Enable NameNode HA wizard freezes when you add a new host to the cluster. The Enable HA Wizard outputs the following JS errors:

```
Uncaught TypeError: Cannot set property 'addNNHost' of undefined db.js:428
App.db.setRollBackHighAvailabilityWizardAddNNHost db.js:428
module.exports.Em.Route.extend.step2.Em.Route.extend.next
 high_availability_routes.js:145
Ember.StateManager.Ember.State.extend.sendRecursively ember-latest.js:15579
Ember.StateManager.Ember.State.extend.send ember-latest.js:15564
App.WizardStep5Controller.Em.Controller.extend.submit step5_controller.
js:646
ActionHelper.registeredActions.(anonymous function).handler ember-latest.
js:19458
(anonymous function) ember-latest.js:11250
f.event.dispatch jquery-1.7.2.min.js:3
h.handle.i jquery-1.7.2.min.js:3
```

Workaround: Close the other open windows and login again from the current window.

• Unable to start gmond process after upgrade to HDP 2.0.6 Stack from HDP 1.3.2 Stack.

Problem: gmond process fails to start on a host during an upgrade

Workaround: Use the following steps to work around the issue:

1. Log onto the host where gmond fails to start.

2. For the gmond process that fails go to the corresponding directory. For example for HDPSlaves go to:

   ```
   /var/run/ganglia/hdp/HDPSlaves/
   ```

3. Remove the PID in the directory.

4. Stop gmond.

   ```
   service hdp-gmond stop
   ```

5. Start gmond.

```
service hdp-gmond start
```

- Add a step to modify fs.defaultFS in default companion file doc

    Problem:

- After upgrading to Ambari 1.4.2, `fs.checkpoint.size` needs to be in appropriate units of bytes, not GBs.

    Problem: Ambari 1.4.1 and earlier assumed this setting to be in GB. The setting is in bytes.

    Workaround: Modify the `fs.checkpoint.size` property using Ambari Web. Select **Services** > **HDFS** > **Configs** > **General** and enter an appropriate integer value in bytes to set the HDFS maximum edit log size for checkpointing. For example, 500000000.

- Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start.

    Problem: The Log4j property file is overwritten during HDFS/ZooKeeper/Oozie services Start. When the client state became became `installed_and_configured` after Service Start:

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

2.7. Third-party Licenses

Table 2.1. Third-party Licenses

HDP Component	Project Library	Version	License
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

3. Release Notes HDP-2.0.8.0

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

In this document:

- Product Version: HDP-2.0.8.0

- What's Changed in HDP-2.0.8.0

- Patch Information

- Minimum system requirements

- Improvements

- Known Issues

- Third-party Licenses

3.1. Product Version: HDP-2.0.8.0

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 1.4.2

- Apache Mahout 0.8.0

- Third party components:

  - Ganglia 3.5.0

  - Ganglia Web 3.5.7

  - Nagios 3.5.0

3.2. What's Changed in HDP-2.0.8.0

The following HDP components were changed in HDP-2.0.8.0:

- What's Changed in Ambari 1.4.2

3.2.1. What's Changed in Ambari 1.4.2

The following changes were made in Ambari 1.4.2:

- BUG-11614 AMBARI-4059 NameNode/YARN max heap is not showing in Web

- BUG-11500 AMBARI-4025 Smoke tests for HBase fail on a 4 node cluster when smoke tests are run on a non RS/Master host.

- BUG-11467 AMBARI-4022 Config upgrade helper is not working.

- BUG-11464 AMBARI-4024 HiveSchema file for Hive should be hive-schema-0.12.0.oracle.sql

- BUG-11407 AMBARI-3966 TaskTracker fails to start on 1.3.2 in Secure mode

- BUG-11394 AMBARI-3987 Resource providers are set for wrong stack version

- BUG-11380 AMBARI-3980 WebHCat PID Dir is /etc/run/webhcat for HDP-1.x stack

- BUG-11348 AMBARI-3970 Ambari Oracle DDL upgrade script is not working with non default ambari username

- BUG-11346 AMBARI-3998, AMBARI-4012 NameNode max heap is not showing in HDP 1.3.2 stack (and possibly others)

- BUG-11338 AMBARI-3964 Set value of mapred.jobtracker.completeuserjobs.maximum=0 in Ambari, currently it is set to 5 in HDP 1.3.x

- BUG-11336 AMBARI-3966 Secure cluster: task-controller binary should be owned by the group specified in taskcontroller.cfg

- BUG-11280 AMBARI-3965 "Add New Hosts" wizard drops out to step "Install, start and test" of installer.

- BUG-11278 AMBARI-3958 Metrics of HDFS, YARN and HBase are not available.

- BUG-11275 AMBARI-3963 Ambari host registering freezes if wrong ssh key provided.

- BUG-11260 AMBARI-3948 HBase Multi-Master setup shows HBase Master alerts under HDP 1.3.x stack

- BUG-11259 AMBARI-3953 HBase Master alerts are confusing in multi-master environment

- BUG-11253 AMBARI-3954 hbase.zookeeper.quorum changing inconsistently on hosts after adding ZookeeperServer

- BUG-11246 AMBARI-3936 Ambari UI won't load after upgrading Ambari with Oracle

- BUG-11211 AMBARI-3981 Services mysteriously disappear after Stack upgrade

- BUG-11202 AMBARI-3979 nagios alerts are not adjusted after delete host

- BUG-11201 AMBARI-3910 NodeManager live dashboard widget does not update after delete host

- BUG-11113 AMBARI-3892 UI shows "Directory: /" on trying to navigate to "Login" page

- BUG-11106 AMBARI-3882 Background operations popup window minimum size should be fixed when narrowing down the browser

- BUG-11100 AMBARI-3881 UI incorrect behavior during upgrade

- BUG-11086 AMBARI-3899 "HDFS Short-circuit read" config property is repeated

- BUG-11080 AMBARI-3880 HDFS and HBase time series and other data are not available through the API

- BUG-11076 AMBARI-3866 UI pointing to Ambari incubator site instead of TLP site

- BUG-11057 AMBARI-3909 Upgrade scenario for Local Repo needs to be addressed when user was using a local repo and moves to the new local repo setup.

- BUG-11055 AMBARI-3864 JS Error in 'Add Host Wizard' if we proceed with failed registered hosts

- BUG-11054 AMBARI-3857 Clicking on Settings link navigates to login page for a non-admin user.

- BUG-11039 AMBARI-3848 Remove Host JS Error in 'add host wizard'

- BUG-11026 AMBARI-3845 Incorrect warning message when trying to delete host with running components

- BUG-11022 AMBARI-3871 When adding nodes to an existing cluster yum should specify a specific ambari-agent version

- BUG-11018 AMBARI-3858 "Add New Hosts" wizard appears automatically after it closed and re-login to UI

- BUG-10997 AMBARI-3846 Nagios alerts when returning bad json breaks the UI really badly.

- BUG-10976 AMBARI-3834 Wrong response for requests with OR on components

- BUG-10969 AMBARI-3820 Security wizard: Reloading wizard fails due to missing user credentials.

- BUG-10945 AMBARI-3830 Incorrect behavior of Ambari with logged on user without admin rights.

- BUG-10904 AMBARI-3808 If hdfs-exclude-file config type already exists then installing DATANODE (e.g. while adding new host) fails

- BUG-10876 AMBARI-3829 Custom user name is not displayed

- BUG-10875 AMBARI-3828 JS error on Confirm Hosts page

- BUG-10856 AMBARI-3773 Ambari server runs into error due to the new stack 2.0.

- BUG-10844 AMBARI-3800 Please add stack 2.0.8 into Ambari

- BUG-10837 AMBARI-3781 JS error on trying to login after launching Security Wizard

- BUG-10835 AMBARI-3801 hostname wrapping on Hosts page

- BUG-10830 AMBARI-3769 HbaseMaster visible in components list after it was removed

- BUG-10826 AMBARI-3792 after reboot, Unable to connect to: https://server:8441/agent/ v1/heartbeat/agent_hostname due to No JSON object could be decoded"

- BUG-10743 AMBARI-3757 Unable to delete host or host-components without warning

- BUG-10741 AMBARI-3814 Enable data locality in YARN

- BUG-10632 AMBARI-3735 Reassign Master: moving the NameNode while NameNode HA is enabled breaks HBase

- BUG-10628 AMBARI-3763 Incorrect href for a POST response using associated resources

- BUG-10583 AMBARI-3711 Installer wizard: Zookeeper service should be a requirement for Hive+WebHcat service.

- BUG-10549 AMBARI-3715 Reassign Master Wizard does not display folder and hosts on "Manual commands" page after browser reopening

- BUG-10532 AMBARI-3686 NameNode HA wizard (Configure Components step): Task "Reconfigure HDFS" always fail, and user cannot proceed to next step

- BUG-10530 AMBARI-3670 Get hosts API is broken

- BUG-10523 AMBARI-3671 "Uncaught exception" in JS while navigating through services on Services page

- BUG-10488 AMBARI-3702 background ops dialog showing "no hosts to show" even with hosts

- BUG-10484 AMBARI-3690 Typo in label text on Hosts page

- BUG-10483 AMBARI-3649 If there is no slave component when we click on link to host js error occurs.

- BUG-10482 AMBARI-3695 "Confirm hosts" shows "ntpd not running" warning, but it's running on host

- BUG-10473 AMBARI-3680 Hive Smoke Test fails after installing clients on an added host

- BUG-10472 AMBARI-3878 ResourceManager Heap metrics is not correct on Ambari console

- BUG-10463 AMBARI-3648 Failed to start Hive Metastore (centos5.8, Stack 2.0)

- BUG-10431 AMBARI-3651 Reassign wizard "Configure Component": Task "Stop All Services" failed, but after retry is marked as successful.

- BUG-10429 AMBARI-3640 Reassign Master: manual commands step disappear after logout/login (HA enabled)

- BUG-10419 AMBARI-3626 Cleanup dialog for unable to delete host (masters installed)

- BUG-10418 AMBARI-3583 Incorrect progress calculation of request

- BUG-10416 AMBARI-3682 Invalid UI behaviour after editing Hive/Oozie database URL

- BUG-10410 AMBARI-3682 Cleanup dialog for unable to delete host when no heartbeat

- BUG-10395 AMBARI-3621 cleanup dialog for unable to delete host (components running)

- BUG-10392 AMBARI-3621AMBARI-3699 App.ServiceConfig needs 'configsValidator' validating across service values

- BUG-10391 AMBARI-3644 App.ServiceConfig needs 'defaultsProvider' mechanism

- BUG-10384 AMBARI-3615 Ambari agent creates empty folder /var/ambari-agent

- BUG-10372 AMBARI-3583 Requests API rollup tasks information improperly

- BUG-10370 AMBARI-3616 After deleting host with last slave component charts throw js errors

- BUG-10363 AMBARI-3605 Hive client installation failed on configured for Ambari server HTTPS

- BUG-10362 AMBARI-3712 Web UI incorectly shows host as 'Heartbeat lost' or 'slaves down' after closing add host wizard

- BUG-10352 AMBARI-3613 Enable HA wizard loads after sign in

- BUG-10350 AMBARI-3696 Smoke test for YARN uses the hostname in yarn.resourcemanager.webapp.address for smoke tests.

- BUG-10348 AMBARI-3678 Change Oozie default log level to INFO

- BUG-10343 AMBARI-3691 host detail page missing padding under tabs

- BUG-10342 AMBARI-3632 need ability to remove hbase master

- BUG-10325 AMBARI-3618 host actions UI changes based on new stop/start all and delete func

- BUG-10301 AMBARI-3692 Add host wizard: on summary page of just added host, no host components are displayed.

- BUG-10296 AMBARI-3591 Services API errors after deleting host with HBase RegionServer (blocks UI)

- BUG-10284 AMBARI-3854 HostCleanup.py kills java processes not related to ambari

- BUG-10263 AMBARI-3593 Ambari ui works incorrect when we delete host which has last slave component on the cluster

- BUG-10247 AMBARI-3683 NameNode HA: Reconfigure dfs.namenode.https-address property in hdfs-site.xml

- BUG-10191 AMBARI-3570 Assign Slaves and Clients step perfomance defects

- BUG-10161 AMBARI-3582, AMBARI-3634 Cleanup old host-override functionality to use config-groups

- BUG-10134 AMABRI-3565 Error with Configs in the addHostWizard

- BUG-10117 AMBARI-3551 Reassign Master: option for the wizard should be hidden on single node cluster.

- BUG-10102 AMBARI-3670 Hosts should roll up host_components status into its status

- BUG-10100 AMBARI-3604 Host-components in API should have 'type' field

- BUG-10098 AMBARI-3677 Service state should accurately reflect state of host-components

- BUG-10096 AMBARI-3583 Requests API should rollup tasks information

- BUG-10084 AMBARI-3581 Js error occurs after deleting host

- BUG-10080 AMBARI-3555 Reassign Master (namenode HA): HDFS service check fails when standby namenode is stopped before wizard

- BUG-10039 AMBARI-3553 NameNode HA wizard: Refreshing the wizard displays incorrect manual commands.

- BUG-10038 AMBARI-3539 NameNode HA wizard: Refreshing on review page navigates to select host page

- BUG-10036 AMBARI-3623 LiveStatus of the component is not updated when username is changed.

- BUG-10021 AMBARI-3547 Can't delete host when heartbeating is lost

- BUG-9980 AMBARI-3635 Property yarn.log.server.url cannot be changed on UI

- BUG-9908 AMBARI-3643 Default Authorization Provider needs to be set in Hive configs

- BUG-9887 AMBARI-3721 Yarn Memory graph goes over 100%

- BUG-9826 AMBARI-3620 Ambari server setup-security kerberos jaas configuration accepts directory path as keytab path

- BUG-9797 AMBARI-3486 Enable Security fails when Ambari setup is rerun to set java_home to oracle jdk7

- BUG-9748 AMBARI-3502 Field 'hadoop.security.auth_to_local' changes tag type from Textarea to Input

- BUG-9739 AMBARI-3644AMBARI-3521 Incorrect status counters on cluster deploy

- BUG-9732 AMBARI-3644AMBARI-3493 HDFS service marked stopped when 1 journal node is taken down

- BUG-9730 AMBARI-3488 Status does not show up for newly added hosts

- BUG-9606 AMBARI-3595 Firewall issues on host checks after iptables disabling

- BUG-9566 AMBARI-3645 HA cluster: some dashboard's widgets contain "Null", "NaN" values after services stop

- BUG-9548 AMBARI-3452, AMBARI-3868 Add host fails after configuring NN HA with JavaScript error

- BUG-9542 AMBARI-3477 JavaScript errors during service tab changing

- BUG-9508 AMBARI-3716 Reset server does not indicate it has failed

- BUG-9380 AMBARI-3456 Text of installation stage doesn't correspond to reality

- BUG-9332 AMBARI-3398 The number of alerts in MapReduce item in menu gets out on the next line.

- BUG-8441 AMBARI-3534 Hadoop Core Health Check script needs to be included in Ambari HDP installations

- BUG-4654 AMBARI-3821 Make Ambari Add Host Wizard more clear on manual registration of new node.

3.3. Patch Information

In this section:

- Patch Information for Hadoop

- Patch Information for HBase

- Patch Information for ZooKeeper

- Patch Information for Pig

- Patch Information for Hive

- Patch Information for HCatalog

- Patch Information for Oozie

3.3.1. 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.

3.3.2. Patch information for HBase

HBase is based on Apache HBase 0.96.0.

3.3.3. 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.

3.3.4. 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.

3.3.5. 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.

3.3.6. Patch information for HCatalog

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

3.3.7. 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.

3.3.8. 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.

3.4. Minimum system requirements

In this section:

- Hardware Recommendations

- Operating Systems Requirements

- Software Requirements

- Database Requirements

- Virtualization and Cloud Platforms

- Configure the Local Repositories

3.4.1. Hardware Recommendations

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

3.4.2. Operating Systems Requirements

The following operating systems are supported:

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

- 64-bit CentOS v5.*, v6.*

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

- 64-bit Oracle Linux v5, v6

- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

- 64-bit Ubuntu Precise 12.04

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

3.4.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

- php-curl (Required for SLES installs.)

3.4.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.

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

- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

3.4.5. Virtualization and Cloud Platforms

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

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

3.4.6. 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.

3.5. Improvements

Apache Ambari update to version 1.4.2. This release of Apache Ambari includes the new features and improvements:

- Ability to move Master Components (NN, SNN, JT, RM)

- Simplified Local Repository Setup in Install Wizard

- Support for multiple HBase masters

- User customizable setting to disable "Background operations popup"

- Stop/start all host components

- Ability to add client components to a host

- Ability to delete a host from a cluster

3.6. Known Issues

In this section:

- Known Issues for Ubuntu 12.04

- Known Issues for SLES 11

- Known Issues for HDP

- Known Issues for HDFS

- Known Issues for Hive

- Known Issues for Oozie

- Known Issues for Hue

- Known Issues for Ambari

3.6.1. Known Issues for Ubuntu 12.04

- Hue does not support Ubuntu hosts at this time.

- Ambari does not support Ubuntu hosts at this time.

3.6.2. Known Issues for SLES 11

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

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

  Workaround: Install php_curl on your SLES host.

  ```
  zypper install php-curl
  ```

3.6.3. Known Issues for HDP

- **BUG-825:** EC2 m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

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

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

3.6.4. 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.

3.6.5. 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;
  ```

3.6.6. 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.

3.6.7. 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.

3.6.8. Known Issues for Ambari

- Ubuntu hosts not supported at this time.

- The component version information displayed by Ambari is based on the Ambari Stack definition. If you have applied patches to the Stack and to your software repository, that component version might differ from the actual version installed. There is no functional impact on Ambari if the patch versions mismatch. If you have any questions on component versions, refer to the rpm version installed on the actual host.

- **BUG-11634:** Upgraded single-node cluster to HDP-2.0.6, may be missing yarn job summary entries.

  Problem: After upgrade of a cluster to HDP 2 using Ambari, you may notice that the yarn job summary entries are missing. This typically happens if the YARN ResourceManager host is shared with MapReduce2 components.

  Workaround: To fix this issue modify the `log4j.properties` file at `/etc/hadoop/conf` on the ResourceManager host by adding the following lines:

  > **Note**
  >
  > Modify the value for `log4j.appender.RMSUMMARY.File` property to contain the actual value of `yarn_log_dir_prefix` and `yarn_user`. You can get the values from the latest global config type. Use `configs.sh` tool to read the global config type.

```
#
# Job Summary Appender
#
# Use following logger to send summary to separate file defined
 by
# hadoop.mapreduce.jobsummary.log.file rolled daily:
# hadoop.mapreduce.jobsummary.logger=INFO,JSA
#
hadoop.mapreduce.jobsummary.logger=${hadoop.root.logger}
hadoop.mapreduce.jobsummary.log.file=hadoop-mapreduce.jobsummary.
log
log4j.appender.JSA=org.apache.log4j.DailyRollingFileAppender
# Set the ResourceManager summary log filename
yarn.server.resourcemanager.appsummary.log.file=hadoop-mapreduce.
jobsummary.log
# Set the ResourceManager summary log level and appender
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.
logger}
#yarn.server.resourcemanager.appsummary.logger=INFO,RMSUMMARY

# To enable AppSummaryLogging for the RM,
# set yarn.server.resourcemanager.appsummary.logger to
# <LEVEL>,RMSUMMARY in hadoop-env.sh

# Appender for ResourceManager Application Summary Log
# Requires the following properties to be set
```

```
#     - hadoop.log.dir (Hadoop Log directory)
#     - yarn.server.resourcemanager.appsummary.log.file (resource
 manager app summary log filename)
#     - yarn.server.resourcemanager.appsummary.logger (resource
 manager app summary log level and appender)
log4j.appender.RMSUMMARY=org.apache.log4j.RollingFileAppender
log4j.appender.RMSUMMARY.File=[yarn_log_dir_prefix]/[yarn_user]/
${yarn.server.resourcemanager.appsummary.log.file}
log4j.appender.RMSUMMARY.MaxFileSize=256MB
log4j.appender.RMSUMMARY.MaxBackupIndex=20
log4j.appender.RMSUMMARY.layout=org.apache.log4j.PatternLayout
log4j.appender.RMSUMMARY.layout.ConversionPattern=%d{ISO8601} %p
 %c{2}: %m%n
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.appender.JSA.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss}
 %p %c{2}: %m%n
log4j.appender.JSA.DatePattern=.yyyy-MM-dd
log4j.appender.JSA.layout=org.apache.log4j.PatternLayout
log4j.logger.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=${yarn.server.resourcemanager.
appsummary.logger}
log4j.additivity.org.apache.hadoop.yarn.server.resourcemanager.
RMAppManager$ApplicationSummary=false
```

- **BUG-11607:** Add `dfs.journalnode.edits.dir` property on upgrading to HDP 2.0.6.0 stack.

  Problem: If you are upgrading from a stack version prior to HDP 2.0.6.0, the Enable NameNode HA Wizard fails due to a missing property expected in hdfs-site.xml.

  Workaround: From the HDFS service page add the following property in the Custom hdfs-site.xml section:

```
<property>
                          <name>dfs.journalnode.edits.dir</name>
                          <value>/grid/0/hdfs/journal</value>
                          </property>
```

- **BUG-11600:** Hive check execute fails after upgrading from BWGA with Oracle.

  Problem: When Ambari is upgraded to 1.4.2 and security is enabled, the Hive service check can fail due to a conflicting combination of authorization properties.

  Workaround: Disable authorization. Using Ambari UI, set `hive.security.authorization.enabled` to false. Or, verify that the correct combination of authorization properties are used. For example:

```
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-11571:** HA wizard freezes after adding host to upgraded cluster.

  Problem: Deployed 3-node cluster using Ambari 1.4.2 version with HDP 1.3.3 stack on SUSE with MySQL DB for Hive and Oozie.

1. Upgraded Ambari to Ambari 1.4.2.

2. Upgraded HDP from 1.3.3 to 2.0.6.

3. Configured hdfs-site.xml on all hosts.

4. Added 4th host.
The HA wizard holds on 2nd step and outputs the following JS errors:

```
Uncaught TypeError: Cannot set property 'addNNHost' of undefined db.js:428
App.db.setRollBackHighAvailabilityWizardAddNNHost db.js:428
module.exports.Em.Route.extend.step2.Em.Route.extend.next
 high_availability_routes.js:145
Ember.StateManager.Ember.State.extend.sendRecursively ember-latest.js:15579
Ember.StateManager.Ember.State.extend.send ember-latest.js:15564
App.WizardStep5Controller.Em.Controller.extend.submit step5_controller.
js:646
ActionHelper.registeredActions.(anonymous function).handler ember-latest.
js:19458
(anonymous function) ember-latest.js:11250
f.event.dispatch jquery-1.7.2.min.js:3
h.handle.i jquery-1.7.2.min.js:3
```

Workaround: Close the other open windows and login again from the current window.

• **BUG-11553:** Unable to start gmond process after upgrade to HDP 2.0.6 Stack from HDP 1.3.2 Stack.

Problem: gmond process fails to start on a host during an upgrade

Workaround: Use the following steps to work around the issue:

1. Log onto the host where gmond fails to start.

2. For the gmond process that fails go to the corresponding directory. For example for HDPSlaves go to:

    ```
    /var/run/ganglia/hdp/HDPSlaves/
    ```

3. Remove the PID in the directory.

4. Stop gmond.

    ```
    service hdp-gmond stop
    ```

5. Start gmond.

    ```
    service hdp-gmond start
    ```

• **BUG-11374:** After performing a cluster install using local repositories, UI incorrectly says "no".

Problem: If you have entered your own repositories during **Install Wizard** > **Advanced Repository Options** and then you add hosts to the cluster, the Review page of the Add Hosts Wizard shows `Local Repository = No`.

• **BUG-11105:** During upgrade fs.checkpoint.size needs to be in proper units.

Problem: Previous versions of Ambari assumed this setting to be in GB. The setting is in bytes.

Workaround: In HDFS Services Configs General, enter an appropriate integer value in bytes to set the HDFS maximum edit log size for checkpointing. For example, 500000000.

• **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. When the client state became 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

3.7. Third-party Licenses

Table 3.1. 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

4. Release Notes HDP-2.0.6.0

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

In this document:

- Product Version: HDP-2.0.6.0

- Patch Information

- Minimum system requirements

- Improvements

- Known Issues

- Licenses

4.1. Product Version: 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

4.2. Patch Information

In this section:

- Patch Information for Hadoop

- Patch Information for HBase

- Patch Information for ZooKeeper

- Patch Information for Pig

- Patch Information for Hive

- Patch Information for HCatalog

- Patch Information for Oozie

4.2.1. 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.

4.2.2. Patch information for HBase

HBase is based on Apache HBase 0.96.0.

4.2.3. 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.

4.2.4. 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.

4.2.5. 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.

4.2.6. Patch information for HCatalog

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

4.2.7. 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.

4.2.8. 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.

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.

- 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.

4.3.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

- php-curl (Required for SLES installs.)

4.3.4. 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.

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 Ubuntu host support.

- Added support to upgrade Ambari.

- Added support for Oracle JDK 1.7 and Open JDK 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 Open JDK 7.

- Added Ambari Agent host-level Nagios alert.

- Added Host Checks and cleanup script during Cluster Install and Add Hosts wizards.

4.5. Known Issues

In this section:

- Known Issues for Ubuntu 12.04

- Known Issues for SLES 11

- Known Issues for HDFS

- Known Issues for Hive

- Known Issues for Oozie

- Known Issues for Hue

- Known Issues for Ambari

4.5.1. Known Issues for Ubuntu 12.04

- Hue does not support Ubuntu hosts at this time.

- Ambari does not support Ubuntu hosts at this time.

4.5.2. Known Issues for SLES 11

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

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

  Workaround: Install php_curl on your SLES host.

  ```
  zypper install php-curl
  ```

4.5.3. 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.

4.5.4. 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;
  ```

4.5.5. 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.

4.5.6. 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.

4.5.7. 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

4.6. Third-party Licenses

Table 4.1. 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

5. Release Notes HDP-2.0.5.0 Beta

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

In this document:

- Product Version: HDP-2.0.5.0 Beta

- Patch Information

- Minimum system requirements

- Improvements

- Known Issues

5.1. 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

5.2. Patch Information

In this section:

- [Patch Information for Hadoop](#)

- [Patch Information for HBase](#)

- [Patch Information for ZooKeeper](#)

- [Patch Information for Pig](#)

- [Patch Information for Hive](#)

- [Patch Information for HCatalog](#)

- [Patch Information for Oozie](#)

5.2.1. 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`.

5.2.2. 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.

5.2.3. 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.

5.2.4. 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.

5.2.5. 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.

5.2.6. Patch information for HCatalog

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

5.2.7. 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`.

5.2.8. 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.

5.3. Minimum system requirements

In this section:

- Hardware Recommendations

- Operating Systems Requirements

- Software Requirements

- Database Requirements

- Virtualization and Cloud Platforms

- Configure the Local Repositories

5.3.1. Hardware Recommendations

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

5.3.2. Operating Systems Requirements

The following operating systems 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.

5.3.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

5.3.4. Database Requirements

- Hive and HCatalog require a database to use as a metadata store 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.

5.3.5. Virtualization and Cloud Platforms

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

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

5.3.6. 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

- 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.

5.5. Known Issues

In this section:

- Known Issues for Hosts

- Known Issues for Hadoop

- Known Issues for Pig

- Known Issues for Ambari

5.5.1. Known Issues for Hosts

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

5.5.2. Known Issues for Hadoop

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

5.5.3. 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.

5.5.4. 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.

6. 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:

- Product Version: HDP-2.0.4.0 (Community Preview)

- Patch Information

- Minimum system requirements

- Improvements

- Known Issues

6.1. 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

6.2. Patch Information

In this section:

- Patch Information for Hadoop

- Patch Information for HBase

- Patch Information for ZooKeeper

- Patch Information for Pig

- Patch Information for Hive

- Patch Information for HCatalog

- Patch Information for Oozie

6.2.1. Patch Information for Hadoop

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

- YARN-321: Generic Application History.

6.2.2. 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.

6.2.3. 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.

6.2.4. 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.

6.2.5. 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.

6.2.6. Patch information for HCatalog

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

6.2.7. 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`.

6.2.8. 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.

6.3. Minimum system requirements

In this section:

- Hardware Recommendations

- Operating Systems Requirements

- Software Requirements

- Database Requirements

- Virtualization and Cloud Platforms

- Configure the Local Repositories

6.3.1. Hardware Recommendations

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

6.3.2. Operating Systems Requirements

The following operating systems 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.

6.3.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

6.3.4. 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.

6.3.5. Virtualization and Cloud Platforms

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

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

6.3.6. 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.

6.4. 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.

6.5. Known Issues

In this section:

- Known Issues for Hadoop

- Known Issues for Hive

- Known Issues for HBase

6.5.1. Known Issues for Hadoop

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

6.5.2. 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.

6.5.3. 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.

7. Release Notes HDP-2.0.0.2 (Alpha)

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

7.1. 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

7.2. Patch Information

In this section:

• Patch Information for Hadoop

• Patch Information for HBase

• Patch Information for ZooKeeper

• Patch Information for Pig

• Patch Information for Hive

• Patch Information for HCatalog

7.2.1. 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.

7.2.2. 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.

7.2.3. 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.

7.2.4. 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.

7.2.5. 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.

7.2.6. 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.

7.3. Minimum system requirements

In this section:

- Hardware Recommendations

- Operating Systems Requirements

- Software Requirements

- Database Requirements

- Virtualization and Cloud Platforms

- Configure the Local Repositories

7.3.1. Hardware Recommendations

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

7.3.2. Operating Systems Requirements

The following operating systems 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.

7.3.3. Software Requirements

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

7.3.4. 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.

7.3.5. Virtualization and Cloud Platforms

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

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

7.3.6. 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.

7.4. 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

7.5. Known Issues

In this section:

- Known Issues for Hadoop

- Known Issues for Hive

- Known Issues for HBase

7.5.1. Known Issues for Hadoop

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

7.5.2. 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

7.5.3. 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.

8. Release Notes HDP-2.0.0.1 (Alpha)

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

8.1. 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

8.2. 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 (`longvarchar`to `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.

8.3. Minimum system requirements

Hardware Recommendations:

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

Operating Systems Requirements:

The following operating systems are supported:

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

- 64-bit CentOS v5.*, v6.*

Important

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

Graphics Requirements:

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

Software Requirements:

On each of your hosts:

- yum

- rpm

- scp

- curl

- wget

- pdsh

- On the machine from which you will run HMC:

  - Firefox v.12+

Database Requirements:

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

Optional: Configure the local repositories

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

> **Important**
>
> The installer pulls many packages from the base OS repos. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. For example, if you are using RHEL 6 your hosts must be able to access the "Red Hat Enterprise Linux Server 6 Optional (RPMs)" repo. If this repo is disabled, the installation is unable to access the `rubygems` package, which is necessary for HMC to operate. If you encounter problems with base OS repos being unavailable, please contact your system administrator to arrange for these additional repos to be proxied or mirrored.

8.4. Improvements

- 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.

8.5. Known Issues

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

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