

Hortonworks Data Platform Win

Release Notes

(Aug 13, 2013)

Hortonworks Data Platform Win 1.3 GA : Release Notes

Copyright © 2013 Hortonworks, Inc. All rights reserved.

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner. The Hortonworks Data Platform consists of the essential set of Apache Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, Zookeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, [training](#) and partner-enablement services. All of our technology is, and will remain free and open source. Please visit the [Hortonworks Data Platform](#) page for more information on Hortonworks technology. For more information on Hortonworks services, please visit either the [Support](#) or [Training](#) page. Feel free to [Contact Us](#) directly to discuss your specific needs.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Table of Contents

1. Release Notes HDP-Win-1.3.0	1
1.1. Product Version: HDP-1.3.0	1
1.2. Patch Information	2
1.2.1. Patch information for Hadoop	2
1.2.2. Patch information for HBase	3
1.2.3. Patch information for Hive	4
1.2.4. Patch information for HCatalog	5
1.2.5. Patch information for Pig	5
1.2.6. Patch information for ZooKeeper	5
1.2.7. Patch information for Oozie	5
1.2.8. Patch information for Sqoop	6
1.2.9. Patch information for Mahout	6
1.2.10. Patch information for Flume	6
1.3. Minimum System Requirements	8
1.3.1. Operating Systems Requirements	8
1.3.2. Software Requirements	8
1.3.3. Database Requirements	8
1.3.4. Virtualization and Cloud Platforms	9
1.4. Improvements	9
1.5. Known Issues	10
1.5.1. Known Issues for Hadoop	10
1.5.2. Known Issues for Hive	11
1.5.3. Known Issues for WebHCatalog	13
1.5.4. Known Issues for HBase	13
2. Release Notes HDP-Win-1.1.0 (GA)	14
2.1. Product Version: HDP-Win-1.1.0 (GA)	14
2.2. Patch Information	14
2.2.1. Patch Information for Hadoop	14
2.2.2. Patch Information for Pig	20
2.2.3. Patch Information for Hive	21
2.2.4. Patch Information for HCatalog	21
2.3. Improvements	21
2.4. Known Issues	22
3. Release Notes HDP-Win-1.1.0	23
3.1. Product Version: HDP-Win-1.1.0	23
3.2. Patch Information	23
3.2.1. Patch Information for Hadoop	23
3.2.2. Patch Information for Pig	29
3.2.3. Patch Information for Hive	30
3.2.4. Patch Information for HCatalog	30
3.3. Known Issues	30

1. Release Notes HDP-Win-1.3.0

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

This document contains:

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

1.1. Product Version: HDP-1.3.0

All HDP 1.3.0 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 1.3.0 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 1.3.0.

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

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



Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.11.0
- Apache Oozie 3.3.2
- Apache Sqoop 1.4.3
- Apache Flume 1.3.1
- Apache Mahout 0.7.0

1.2. Patch Information

In this section:

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

1.2.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 1.2.0 and includes the following additional patches, organized by project and patch number:

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

1.2.2. Patch information for HBase

HBase is based on Apache HBase 0.94.6 and includes the following patches organized by patch number:

- [HBASE-8550](#): 0.94 ChaosMonkey grep for master is too broad.
- [HBASE-8547](#): Fix `java.lang.RuntimeException: Cached an already cached block` (Patch file: `hbase-8547_v2-0.94-reduced.patch` and `addendum2+3`).
- [HBASE-8530](#): Refine error message from `ExportSnapshot` when there is leftover snapshot in target cluster.
- [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-8465](#): Added support for auto-drop rollback snapshot for snapshot restore.
- [HBASE-8455](#): Updated `ExportSnapshot` to reflect changes in [HBASE-7419](#).
- [HBASE-8413](#): Fixed `Snapshot verify region` will always fail if the HFile has been archived.
- [HBASE-8405](#): Added new custom options to how `ClusterManager` runs commands.
- [HBASE-8377](#): Fixed `IntegrationTestBigLinkedList` calculates wrap for linked list size incorrectly.
- [HBASE-8352](#): Rename `.snapshot` directory to `.hbase-snapshot`.
- [HBASE-8350](#): Added support to enable `ChaosMonkey` to run commands as different users.
- [HBASE-8326](#): `mapreduce.TestTableInputFormatScan` times out frequently (and addendum).
- [HBASE-8276](#): Backport [HBASE-6738](#) to 0.94. (Too aggressive task resubmission from the distributed log manager.)
- [HBASE-8274](#): Backport [HBASE-7488](#). (Implement `HConnectionManager.locateRegions` which is currently returning null.)
- [HBASE-8270](#): Backport [HBASE-8097](#) to 0.94. (`MetaServerShutdownHandler` may potentially keep bumping up `DeadServer.numProcessing`.)
- [HBASE-8260](#): Added support to create deterministic, longer running, and less aggressive generic integration test for HBase trunk and HBase branch 94.
- [HBASE-8259](#): Snapshot backport in 0.94.6 breaks rolling restarts.
- [HBASE-8246](#): Backport [HBASE-6318](#) to 0.94 where `SplitLogWorker` exits due to `ConcurrentModificationException`.
- [HBASE-8213](#): Fixed global authorization may lose efficacy.

- [HBASE-8207](#): Fixed replication could have data loss when machine name contains hyphen "_".
- [HBASE-8179](#): Fixed JSON formatting for cluster status.
- [HBASE-8158](#): Backport [HBASE-8140](#). (Added support to use `JarFinder` aggressively when resolving MR dependencies.)
- [HBASE-8146](#): Fixed `IntegrationTestBigLinkedList` for distributed setup.
- [HBASE-8106](#): Test to check replication log znodes move is done correctly.
- [HBASE-8081](#): Backport [HBASE-7213](#). (Separate `hlog` for meta tables.)
- [HBASE-7820](#): Added support for multi-realm authentication.
- [HBASE-7410](#): [snapshots] Add snapshot/clone/restore/export docs to reference guide. For more details, see [User Guide - HBase Snapshots](#).
- [HBASE-6508](#): Filter out edits at log split time.
- [HBASE-6466](#): Enabled multi-thread support for memstore flush.
- [HBASE-6338](#): Cache method in RPC handler.
- [HBASE-6134](#): Improvement for `split-worker` to improve distributed log splitting time.

1.2.3. Patch information for Hive

Hive is based on Apache Hive 0.11.0 and includes the following patches listed by patch number:



Note

Apache HCatalog is now merged with Apache Hive.

- [HIVE-5542](#): Fixed `TestJdbcDriver2.testMetaDataGetSchemas` failures.
- [HIVE-4611](#): Fixed SMB join failures because of conflicts in bigtable selection policy.
- [HIVE-4540](#): Fixed failures for `GROUPBY/DISTINCT` operations when `mapjoin.mapred=true`.
- [HIVE-4521](#): Fixed auto join conversion failures
- [HIVE-4513](#): Added support to disable Hive history logs by default.
- [HIVE-4510](#): Fixed `HiveServer2` nested exceptions.
- [HIVE-4551](#): Fixed `HCatLoader` failures caused when loading ORC table External apache (4551.patch).
- [HIVE-4524](#): Added support for Hive `HBaseStorageHandler` to work with HCatalog.
- [HIVE-4486](#): Fixed `FetchOperator` that was causing the SMB joins to slow down 50% when there are large number of partitions.

- [HIVE-4485](#): Fixed beeline prints null as empty strings.
- Removed `npath` windowing function.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `mult` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-4392](#): Fixed `Illogical InvalidObjectException` when using `mult` aggregate functions with star columns.
- [HIVE-4343](#): Fixed HiveServer2 with Kerberos - local task for map join fails.
- [HIVE-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-3255](#): Added `DBTokenStore` to store Delegation Tokens in database.
- [HIVE-2084](#): Upgraded DataNuclues from v2.0.3 to v3.0.1.

1.2.4. Patch information for HCatalog

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

1.2.5. Patch information for Pig

Pig is based on Apache Pig 0.11 and includes the following patches listed by patch number:

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

1.2.6. Patch information for ZooKeeper

ZooKeeper is based on Apache ZooKeeper 3.4.5 and includes the following patches listed by patch number:

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

1.2.7. Patch information for Oozie

Oozie is based on Apache Oozie 3.2.0 and includes the following patches listed by patch number:

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

- [OOZIE-1351](#): Fixed issue for Oozie jobs in `PAUSEDWITHERROR` state that fail to change to `SUSPENDEDWITHERROR` state when suspended.
- [OOZIE-1349](#): Fixed issues for `oozieCLI -Doozie.auth.token.cache`.

1.2.8. Patch information for Sqoop

Sqoop is based on Apache Sqoop 1.4.3 and includes the following patches listed by patch number:

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

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

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

1.2.9. Patch information for Mahout

Mahout is based on Apache Mahout 0.7.0 and includes the following patches listed by patch number:

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

1.2.10. Patch information for Flume

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

- [FLUME-924](#): Implement a JMS source for Flume NG.
- [FLUME-1784](#): `JMSsource` fix minor documentation problem and parameter name.
- [FLUME-1804](#): JMS source not included in binary distribution.
- [FLUME-1777](#): `AbstractSource` does not provide enough implementation for sub-classes
- [FLUME-1886](#): Add a JMS enum type to `SourceType` so that users do not need to enter FQCN for `JMSsource`.
- [FLUME-1976](#): JMS Source document should provide instruction on JMS implementation JAR files. For more details, see [Flume User Guide - JMS Source](#).
- [FLUME-2043](#): JMS Source removed on failure to create configuration

- [FLUME-1227](#): Introduce some sort of SpillableChannel ([Spillable Channel - Experimental]).
- Spillable Channel dependencies:
 - [FLUME-1630](#): Improved Flume configuration code.
 - [FLUME-1502](#): Support for running simple configurations embedded in host process.
 - [FLUME-1772](#): `AbstractConfigurationProvider` should remove component which throws exception from configure method.
 - [FLUME-1852](#): Fixed issues with `EmbeddedAgentConfiguration`.
 - [FLUME-1849](#): Embedded Agent doesn't shutdown supervisor
- [FLUME-1878](#): `FileChannel` replay should print status every 10000 events.
- [FLUME-1891](#): Fast replay runs even when checkpoint exists.
- [FLUME-1762](#): File Channel should recover automatically if the checkpoint is incomplete or bad by deleting the contents of the checkpoint directory.
- [FLUME-1334](#): Write a startscript for flume agents on Windows.
- [FLUME-2127](#): JMX Shutdown Command for Flume.
- [FLUME-2148](#): Windows: Add flume-env.ps1
- [FLUME-2151](#): Windows: Update `TestExecSource` to use native commands on Windows
- [FLUME-1870](#): Flume sends non-numeric values with type as float to Ganglia causing it to crash.
- [FLUME-1918](#): File Channel cannot handle capacity of more than 500 Million events.
- [FLUME-1262](#): Move doc generation to a different profile.
- [FLUME-2054](#): Support Version Info on Windows and fix failure of `TestVersionInfo`.
- [FLUME-2057](#): Failures in `FileChannel's TestEventQueueBackingStoreFactory` on Windows
- [FLUME-2058](#): `TestFlumeEventQueue` in `FileChannel` fails on Windows
- [FLUME-2060](#): Failure in `TestLog.testReplaySucceedsWithUnusedEmptyLogMetaDataFastReplay` test on Windows
- [FLUME-2068](#): File Channel issue - recovering from `BadCheckpoint` exception on Windows
- [FLUME-2134](#): AsyncHbase Sink bugfix plus tests errors on Windows
- [FLUME-2135](#): Add zip to the build distribution for Windows support
- [FLUME-2136](#): Windows - Fix intermittent test failure in `TestMonitoredCounterGroup.java`
- [FLUME-2137](#): Fix `StagedInstall.java` to invoke the correct startup script on Windows

- [FLUME-2143](#): Flume build occasionally fails with OutOfMemoryError on Windows.
- [FLUME-2145](#): TestCheckpointRebuilder.testFastReplay fails on Windows due to checkpoint file being memory mapped
- [FLUME-2146](#): Windows: Tmp file creation in TestBodyTextEventSerializer.java needs fixing
- [FLUME-2149](#): Windows: Temporary files generated by File Channel unit tests are not cleaned up properly
- [FLUME-2150](#): Windows: TestFileChannelEncryption needs to be a bit more conservative when examining the exception message
- [FLUME-2094](#): Remove the deprecated - Recoverable Memory Channel

1.3. Minimum System Requirements

In this section:

- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)

1.3.1. Operating Systems Requirements

The following operating systems (OS) are supported:

- Windows Server 2008 R2 (64 bit)
- Windows 2012 (64 bit)

1.3.2. Software Requirements

On each of your hosts you must have the following software installed:

- Microsoft Visual C++ 2010 Redistributable Package (64 bit)
- Python 2.7.5
- Microsoft.NET framework 4.0
- JDK 6u31 or higher
- Python 2.7

1.3.3. Database Requirements

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

database types. For more information on providing access to an existing database, see [Supported Database Matrix for Hortonworks Data Platform](#).

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

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

1.3.4. Virtualization and Cloud Platforms

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

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

1.4. Improvements

- Apache Hadoop updated to version 1.2.0.
- Apache HBase updated to version 0.94.6.1.
- Apache Pig updated to version 0.11.
- Apache Hive updated to version 0.11.
- Apache Oozie updated to version 3.3.2.
- Apache Sqoop updated to version 1.4.3.
- Added support for PostgreSQL v.8.x for Hive Metastore, Oozie, and Sqoop.
- Added the following to Apache Hadoop:
 - [HDFS-2802](#): Added support for RW/RO snapshots in HDFS.

Snapshots are point in time images of parts of the filesystem or the entire filesystem. Snapshots can be a read-only or a read-write point in time copy of the filesystem. There are several use cases for snapshots in HDFS. For details, see [User Guide - HDFS Snapshots](#).

- [HDFS-4750](#): Added support for NFSv3 interface to HDFS. NFS interface support provides the ability for HDFS to have seamless integration with client's file system. For details, see [User Guide - HDFS NFS Gateway](#).
- Added the following to Apache Flume NG:
 - Implemented a JMS source for Apache Flume NG. See [FLUME-924](#), [FLUME-1784](#), [FLUME-1804](#), [FLUME-1777](#), [FLUME-1886](#), [FLUME-1976](#), and [FLUME-2043](#). Also see [Apache Flume Documentation](#).

- Added SpillableChannel (experimental) to Apache Flume NG. See [FLUME-1227](#) for more details.

Also see [FLUME-1630](#), [FLUME-1502](#), [FLUME-1772](#), [FLUME-1852](#), and [FLUME-1849](#) for SpillableChannel dependencies.
- Improvements to Flume NG: [FLUME-1878](#), [FLUME-1891](#), [FLUME-1762](#), [FLUME-1334](#), [FLUME-2127](#), [FLUME-2148](#), and [FLUME-2151](#)
- Bug fixes: [FLUME-1870](#), [FLUME-1918](#), [FLUME-1262](#), [FLUME-2054](#), [FLUME-2057](#), [FLUME-2058](#), [FLUME-2060](#), [FLUME-2068](#), [FLUME-2134](#), [FLUME-2135](#), [FLUME-2136](#), [FLUME-2137](#), [FLUME-2143](#), [FLUME-2145](#), [FLUME-2146](#), [FLUME-2149](#), and [FLUME-2150](#)
- Dropped Feature: [FLUME-2094: Removed the deprecated Recoverable Memory Channel](#).
- Added support to integrate Apache HCatalog with Apache Sqoop.

This Sqoop-HCatalog connector supports storage formats abstracted by HCatalog. For more information, see [SQOOP-931](#).

1.5. Known Issues

In this section:

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

1.5.1. Known Issues for Hadoop

- Old logs getting deleted instead of being copied to a new file.

Problem: Log4J's DailyRollingFileAppender on Windows can truncate log files instead of creating backups during the daily rollover.

Workaround: Reconfigure log4j.properties to use a different appender.

- File upload fails to upload in NFS-MountDir.

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

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

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

- Do not use init.d scripts for starting or stopping Hadoop services. It is not recommended.

1.5.2. Known Issues for Hive

- Mapreduce task from Hive dynamic partitioning query is killed.

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

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

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

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

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

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

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

the following error is reported in the Hive log file:

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

Workaround: The workaround is to use the following query:

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

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

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

the following error is reported in the Hive log file:

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

```
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.  
selectList(HiveParser_SelectClauseParser.java:1042)  
at org.apache.hadoop.hive.ql.parse.HiveParser_SelectClauseParser.  
selectClause(HiveParser_SelectClauseParser.java:779)  
at org.apache.hadoop.hive.ql.parse.HiveParser.selectClause(HiveParser.  
java:30649)  
at org.apache.hadoop.hive.ql.parse.HiveParser.selectStatement(HiveParser.  
java:28851)  
at org.apache.hadoop.hive.ql.parse.HiveParser.regular_body(HiveParser.  
java:28766)  
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatement(HiveParser.  
java:28306)  
at org.apache.hadoop.hive.ql.parse.HiveParser.  
queryStatementExpression(HiveParser.java:28100)  
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.  
java:1213)  
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.  
java:928)  
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:190)  
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:418)  
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:337)  
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:902)  
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:259)  
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:216)  
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:413)  
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:348)  
at org.apache.hadoop.hive.cli.CliDriver.processReader(CliDriver.java:446)  
at org.apache.hadoop.hive.cli.CliDriver.processFile(CliDriver.java:456)  
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:712)  
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:614)  
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)  
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.  
java:39)  
at sun.reflect.DelegatingMethodAccessorImpl.  
invoke(DelegatingMethodAccessorImpl.java:25)  
at java.lang.reflect.Method.invoke(Method.java:597)  
at org.apache.hadoop.util.RunJar.main(RunJar.java:160)  
FAILED: ParseException line 1:53 cannot recognize input near '/' '10.0'  
'from' in selection target
```

Workaround: The workaround is to use the following query:

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

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

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

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

```
CREATE TABLE 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: Avoid adding string to int fields.

1.5.3. Known Issues for WebHCatalog

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

The following error is reported in the WebHCatalog log file:

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

1.5.4. Known Issues for HBase

- HBase RegionServers fails to shutdown.

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

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

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

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


2. Release Notes HDP-Win-1.1.0 (GA)

RELEASE NOTES: Hortonworks Data Platform 1.1.0 (GA) for Windows powered by Apache Hadoop In this document:

- [Product Version: HDP-Win-1.1.0 \(GA\)](#)
- [Patch Information](#)
- [Improvements](#)
- [Known Issues](#)

2.1. Product Version: HDP-Win-1.1.0 (GA)

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

- Apache Hadoop 1.0.3
- Apache Pig 0.9.3
- Apache HCatalog 0.4.1
- Apache Templeton 0.1.4
- Apache Hive 0.9.0
- Apache Sqoop 1.4.2
- Apache Oozie 3.2.0

2.2. Patch Information

In this section:

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

2.2.1. Patch Information for Hadoop

The following list provides information on the **Apache JIRAs** for this release:

- [HDFS-6527](#): Backport HADOOP-7389: Use of TestingGroups by tests causes subsequent tests to fail.
- [HADOOP-8836](#): UGI should throw exception in case `winutils.exe` cannot be loaded.

- [MAPREDUCE-4657](#): WindowsResourceCalculatorPlugin has Null Pointer Exception.
- [HADOOP-8903](#): Added support HADOOP_USER_CLASSPATH_FIRST in Windows Hadoop cmd.
- [HADOOP-8908](#): Refactor winutil.exe related code.
- [HADOOP-8911](#): CRLF characters in source and text files.
- [HADOOP-8912](#): Add .gitattributes file to prevent CRLF and LF mismatches for source and text files.
- [HADOOP-8868](#): FileUtil#chmod should normalize the path before calling into shell APIs.
- [HDFS-4065](#): TestDFSShell.testGet sporadically fails attempting to corrupt block files due to race condition.
- [HADOOP-8902](#): Enabled Gridmix v1 and v2 benchmarks on the Windows platform.
- [HADOOP-8564](#): Port and extend Hadoop native libraries for Windows to address DataNode concurrent reading and writing issue.
- [HADOOP-4093](#): Fix a bug that AzureBlockPlacementPolicy#chooseTarget only returns one DataNode when replication factor is greater than three.
- [HADOOP-8907](#): Provide means to look for zlib1.dll next to hadoop.dll on Windows.
- [HADOOP-8420](#): Hadoop Common creating package-info.java must not depend on sh.
- [HADOOP-8972](#): Move winutils tests from bat to Java.
- [HADOOP-9026](#): Hadoop.cmd fails to initialize if user's %path% variable has parenthesis.
- [HADOOP-9027](#): Build fails on Windows without sh/sed/echo in the path.
- [HADOOP-8847](#): Change untar to use Java API on Windows instead of spawning tar process.
- [HADOOP-9062](#): hadoop-env.cmd overwrites the value of *_OPTS set before install.
- [HADOOP-9074](#): Hadoop install scripts for Windows.
- [HADOOP-9110](#): winutils ls off-by-one error indexing MONTHS array can cause access violation.
- [HADOOP-9102](#): winutils task isAlive does not return a non-zero exit code if the requested task is not alive.
- [HADOOP-9061](#): Java6+Windows does not work well with symlinks.
- [HADOOP-8645](#): HADOOP_HOME and -Dhadoop.home (from hadoop wrapper script) are not uniformly handled.
- [HADOOP-9185](#): TestFileCreation.testFsClose should clean up on exit.

- [HADOOP-9191](#): `TestAccessControlList` and `TestJobHistoryConfig` fail with JDK7.
- [HADOOP-9177](#): Address issues that reported by static code analysis on `winutils`.
- [HADOOP-9250](#): Windows installer bugfixes
- [MAPREDUCE-4915](#): `TestShuffleExceptionCount` fails with open JDK7
- [MAPREDUCE-4914](#): `TestMiniMRDFSsort` fails with openJDK7.
- [HADOOP-9179](#): `TestFileSystem` fails with open JDK7.
- [HDFS-4358](#): `TestCheckpoint` failure with JDK7.
- [HDFS-4355](#): `TestNameNodeMetrics.testCorruptBlock` fails with open JDK7
- [MAPREDUCE-4909](#): `TestKeyValueTextInputFormat` fails with Open JDK 7 on Windows
- [HADOOP-9175](#): `TestWritableName` fails with Open JDK 7
- [HADOOP-9174](#): `TestSecurityUtil` fails with Open JDK 7
- [HDFS-4337](#): Backport HDFS-4240: For nodegroup-aware block placement, when a node is excluded, the nodes in the same nodegroup should also be excluded.
- [HDFS-4341](#): Set default data dir permission in `MiniDFSclusterWithNodeGroup`.
- [HDFS-4320](#): Add a separate configuration for NameNode RPC address instead of using `fs.default.name`.
- [HDFS-3942](#): Backport HDFS-3495 and HDFS-4234: Update Balancer to support new `NetworkTopology` with `NodeGroup` and use generic code for choosing `DataNode` in Balancer.
- [MAPREDUCE-782](#): Use `PureJavaCrc32` in MapReduce spills.
- [HDFS-496](#): Backport: Use `PureJavaCrc32` in HDFS.
- [HADOOP-7096](#): Allow setting of end-of-record delimiter for `TextInputFormat`.
- [HADOOP-8617](#): Backport HADOOP-6148, HADOOP-6166 and HADOOP-7333 for a pure Java CRC32 calculator implementation.
- [HADOOP-9111](#): Change some JUnit 3 tests to JUnit 4 so that `@Ignore` tests can be run with Ant v1.8.x.
- [HADOOP-9090](#): Support on-demand publish of metrics.
- [HADOOP-9099](#): `TestNetUtils` fails if "UnknownHost" is resolved as a valid hostname.
- [HADOOP-9095](#): Backport HADOOP-8372: `NetUtils.normalizeHostName()` incorrectly handles hostname starting with a numeric character.
- [HADOOP-9036](#): Fix racy test case `TestSinkQueue`

- [HADOOP-8900](#): `BuiltInGzipDecompressor` throws `IOException` - stored gzip size doesn't match decompressed size.
- [HADOOP-6496](#): `HttpServer` sends wrong content-type for CSS files.
- [HADOOP-7868](#): Hadoop native fails to compile when default linker option is `-Wl,-as-needed`.
- [HDFS-3941](#): Backport [HDFS-3498](#) and [HDFS-3601](#): Support replica removal in `BlockPlacementPolicy` and make `BlockPlacementPolicyDefault` extensible for reusing code in subclasses, and add `BlockPlacementPolicyWithNodeGroup` to support block placement with 4-layer network topology.
- [HADOOP-8820](#): Backport [HADOOP-8469](#) and [HADOOP-8470](#): Make `NetworkTopology` class pluggable and add `NetworkTopologyWithNodeGroup`, a 4-layer implementation of `NetworkTopology`.
- [HADOOP-8645](#): `HADOOP_HOME` and `-Dhadoop.home` (from hadoop wrapper script) are not uniformly handled.
- [MAPREDUCE-1806 \(HADOOP-136\)](#): Fixed issues in `CombineFileInputFormat` that caused failure while using Sqoop to export files in ASV.
- [HADOOP-8874](#): Added an API to retrieve valid `HADOOP_HOME` and `bin` path. This patch adds a consistency layer for `HADOOP_HOME` lookups and provides abstractions to qualify bin paths of hadoop binary components.
- [HADOOP-8731](#): Added Public distributed cache support for Windows.
- [HADOOP-9040](#): Added fixes for `TaskController`.
- [HADOOP-9071](#): Improved Ivy log levels.
- [HADOOP-8872](#): Fixed issue caused while invoking `FileSystem.length()` method on a Windows machine using JDK 6.x.
- [MAPREDUCE-4564](#): Fixed shell timeout mechanism. This fix enables successful termination of those processes that are spawned by `Winutils`.
- [MAPREDUCE-4561](#): Added support for node health scripts on Windows
- [MAPREDUCE-4597](#): Fixed intermittent failures for `TestKillSubProcesses`.
- [HADOOP-8739](#): Fixed command line parsing on Windows.
- [HADOOP-8664](#): Fixed the Hadoop streaming job issue that required the user to provide full path to commands.
- [HDFS-3766](#): Fixed `TestStorageRestore` on Windows.
- [HADOOP-8634](#): Fixed the errors caused when `FileSystem.deleteOnExit` method is invoked.
- [HDFS-3763](#): Fixed the `TestNameNodeMXBean` failures on Windows.
- [MAPREDUCE-4510](#): Fixed redundant checks and logging of `getconf` on Windows.

- [HADOOP-8731](#): Added public distributed cache support for Windows. This fixes the failures for `TestTrackerDistributedCacheManager`.
- [HADOOP-8763](#): Fixed issues caused when setting group owner on Windows.
- [HADOOP-8694](#): Added symlink support to Windows platform.
- [HADOOP-8732](#): Fixed test failures caused due to incorrect process serialization on Windows.
- [HDFS-3564](#): Added enhancements to the block placement policy. This enhancement enables a pluggable placement policy and provides a new API that enables moving blocks for balancing. It also enables the placement policy to decide the number of racks and provides ability to extend the policy.
- [HDFS-3566](#): Add `AzureBlockPlacementPolicy` to handle fault and upgrade domains in Azure. This policy distributes replicas across both the fault and the upgrade domains to ensure zero data loss.
- [HADOOP-8457](#): Fixed the file ownership issue for users in the Administrators groups on Windows.
- [MAPREDUCE-4374](#): Added support for configurable environment for child map/reduce tasks on Windows.
- [HADOOP-8453](#): Added unit tests for Winutils program. Winutils is the Windows console program that emulates the Linux command line utilities used by Hadoop.
- [HDFS-385](#): Added new experimental API `BlockPlacementPolicy` allows investigating alternate rules for locating block replicas.
- [HADOOP-8899](#): Fixed issues caused because of Classpath exceeding the maximum operating system (OS) limit.
- [MAPREDUCE-1806](#): Fixed issues with `CombineFileInputFormat`.
- [HADOOP-8935](#): Improved Winutils to handle the failures caused for the `winutils ls` command.
- [HADOOP-6496](#): Fixed the `HTTPServer` issue that caused incorrect rendering of the web interface for HBase.
- [HADOOP-7827](#): Fixed issue with JSP pages for web interfaces.
- [HADOOP-8903](#): Added support for `HADOOP_USER_CLASSPATH_FIRST` environment variable in the `hadoop.cmd` file.
- [HADOOP-8880](#): Fixed Hive test failures caused because of missing Jersey JAR files in the POM template.
- [HADOOP-8733](#): Fixed the failures caused by the dependencies in the `test.sh` script file.
- [MAPREDUCE-4400](#): Fixed performance regression for small jobs and workflows.
- [HADOOP-8734](#): Fixed `LocalJobRunner` to support private distributed cache.

- [HDFS-3833](#): Fixed TestDFSShell failures on Windows caused due to concurrent file read/write.
- [MAPREDUCE-4598](#): Added support for node health scripts on Windows.
- [HADOOP-8657](#): Fixed TestCLI to remove hardcoded value for the file length.
- [HDFS-3163](#): Fixed failures for TestHDFSCLI.testAll which occurred when the user name was not provided in lowercase.
- [HADOOP-8618](#): Fixed build failures caused due to merging of the Hadoop v1.0.3 branch.
- [HADOOP-8544](#): Moved an assertion location in **winutils chmod** command.
- [HADOOP-7389](#): Fixed test failures caused when tests use the TestingGroups.
- [HADOOP-8414](#): Fixed issues caused by localhost resolving to incorrect address on Windows.
- [MAPREDUCE-4368](#): Fixed TaskRunner to handle the event when `java.library.path` contains a quoted path with embedded spaces on Windows platform.
- [MAPREDUCE-4369](#): Fixed streaming job failures with WindowsResourceCalculatorPlugin.
- [MAPREDUCE-4332](#): Fixed command length abort issues on Windows.
- [HADOOP-8487](#): Fixed the HDFS tests to use correct test paths.
- [HADOOP-8534](#): Fixed failures for those tests that left configuration files open.
- [HADOOP-8486](#): Fixed the resource leak caused because of open file handles for SequenceFile.
- [MAPREDUCE-4203](#): Added an implementation of the process tree for Windows.
- [HADOOP-8454](#): Fixed bugs for the **chmod** command in Winutils program.
- [HADOOP-8409](#): Fixed TestCommandLineJobSubmission and TestGenericOptionsParser to work for Windows.
- [MAPREDUCE-4260](#): Added support to use JobObject for spawning tasks on Windows platform.
- [MAPREDUCE-4321](#): Fixed failures for DefaultTaskController on Windows.
- [HADOOP-8424](#): Fixed Classpath issues that caused intermittent failures for web user interface on Windows.
- [HDFS-3424](#): Fixed TestDatanodeBlockScanner and TestReplication failures on Windows.
- [HADOOP-8374](#): Improved support for hard link manipulation on Windows.
- [HADOOP-8440](#): Fixed failures for HarFileSystem.decodeHarURI.
- [HADOOP-8411](#): Fixed TestStorageDirectorFailure, TestTaskLogsTruncater, TestWebHdfsUrl and TestSecurityUtil failures on Windows.

- [HADOOP-8235](#): Added support file permissions and ownership on Windows for RawLocalFileSystem.
- [MAPREDUCE-4204](#): Improved ProcfsBasedProcessTree to enable the resource collection object to be pluggable.
- [MAPREDUCE-4201](#): Fixed issues related to obtaining PIDs on Windows.
- [HADOOP-8234](#): Enabled user group mappings on Windows platform.
- [HADOOP-8223](#): Applied initial patch for branch-1-win.
- [HDFS-4413](#): Secondary namenode won't start if HDFS isn't the default file system

2.2.2. Patch Information for Pig

The following list provides information on the **Apache JIRAs** for this release:

- [PIG-2958](#): Fixed failure issues caused when Pig tests have no associated logger.
- [PIG-2957](#): Fixed failures for TetsScriptUDF test.
- [PIG-2960](#): Increased the timeout for unit tests on Windows.
- [PIG-2943](#): Improved DevTests and Windows checks to include the **Util.Windows** method.
- [PIG-2942](#): Fixed failures for DevTests and TestLoad tests.
- [PIG-2941](#): Added consistency in chaining the Ivy resolvers. This fix also adds the fallback mechanism.
- [PIG-2953](#): Added support for **which** utility on Windows
- [PIG-2956](#): Fixed issues with invalid cache specification for some streaming statement.
- [PIG-2954](#): Fixed test failures caused due to the dependency on **bash**.
- [PIG-2959](#): Fixes for `pig.cmd` to run on Windows.
- [PIG-2801](#): Fixes for grunt **sh** command.
- [PIG-2800](#): Fixed issues with `pig.additional.jars` path separator.
- [PIG-2799](#): Updated Pig streaming interface to run correctly on Windows without Cygwin support.
- [PIG-2798](#): Fixed issues with Pig streaming tests on Windows.
- [PIG-2797](#): Fixed the Pig tests to use **Util.generateURI** method.
- [PIG-2796](#): Fixed issues with invalid path names on HDFS caused because the Pig tests use local temporary paths.
- [PIG-2795](#): Added support to handle paths generated on Windows.

- [PIG-2794](#): Added utilities to facilitate testing on Windows platform.
- [PIG-2793](#): Improved Pig to work on Windows platform without Cygwin support.

2.2.3. Patch Information for Hive

The following list provides information on the **Apache JIRAs** for this release:

- [HIVE-3025](#): Fixed Hive archive command for Hadoop v 0.22 and 0.23.
- [HIVE-3448](#): Fixed failures for the skewjoin.q testcase on Windows.
- [HIVE-3441](#): Fixed failures caused due to the partition column strings in Windows file names.
- [HIVE-3436](#): Fixed the script_pipe.q failures on Windows.
- [HIVE-3483](#): Fixed issues with joins that use partitioned table on Windows.
- [HIVE-3317](#): Fixed TestDocToUnix unit tests on Windows.
- [HIVE-3320](#): Fixed test case failures caused by incorrect handling of CRLF line endings on Windows.
- [HIVE-3319](#): Fixed path related issues that caused the unit test failures for Windows.
- [HIVE-3327](#): Fixed failures caused while execution of the `/bin/cat` script files on Windows.
- [HIVE-3479](#): Fixed issues with negative unit tests.
- [HIVE-3494](#): Fixed JDBC test case failures on Windows.
- [HIVE-3480](#): Fixed file handle leaks in Symbolic and symlink related input formats.

2.2.4. Patch Information for HCatalog

The following list provides information on the **Apache JIRAs** for this release:

- [HCATALOG-512](#): Fixed HCatalog unit tests on Windows.
- [HCATALOG-514](#): Fixed HCatalog python scripts in the package build for Windows.

2.3. Improvements

- Added ability to perform push install from a central computer.
- Added ability to uninstall and reinstall over existing data.

Use the following instructions to upgrade from HDP-Win-1.1:

1. Uninstall HDP using `DESTROY_DATA=no`. On each cluster host, execute the following command from the command shell:


```
msiexec /x "$MSI_PATH" /lv "$PATH_to_Installer_Log_File" DESTROY_DATA="no"
```

2. Reinstall HDP using DESTROY_DATA=no. On each cluster host, execute the following command from the command shell:

```
msiexec /x "$MSI_PATH" /lv "$PATH_to_Installer_Log_File" DESTROY_DATA="no"
```

- Improved error feedback during deployments.

2.4. Known Issues

- One of the regression test for Oozie fails with MS SQL because `java.io.IOException` does not finish within the `600000timeout`.
- One of the regression test for Oozie fails with MS SQL because `java.io.UnsupportedEncodingException` does not finish within the `600000timeout`.
- Running MapReduce jobs using pipes is currently not supported on Windows.
- Non-default compression codecs that are based on zlib or snappy are currently not supported on Windows
- It is possible to encounter the following exception while starting the Hive command line interface (CLI).

```
FAILED: Error in metadata: javax.jdo.JDOFatalDataStoreException: DERBY SQL
error: SQLCODE: -1, SQLSTATE: XJ041, SQLERRMC:
    Failed to create database 'metastore_db', see the next exception for
    details.
    :SQLSTATE: XBM0JDirectory C:\Hadoop\hive-0.9.0\bin\metastore_db already
    exists.
```

This typically happens when the user attempts to install and uninstall HDP repeatedly. In such cases, the directories for Hive might not get deleted properly on uninstall. You can use any of the following workaround:

- Manually delete the `metastore_db` directory (`C:\Hadoop\hive-0.9.0\bin\metastore_db`).
- Uninstall HDP, delete all the files in the Hadoop directory (`C:\Hadoop`), and install HDP again
- HDP creates some files in the HDFS directory (`C:\hdfs`) that are not deleted on uninstall. This issue is observed when `hadoop.tmp.dir` is not defined to point to the `C:\hadoop` location. There is no impact on the deployment of your cluster. However, it is recommended that you manually delete the files in the HDFS directory (`C:\hdfs`) after you uninstall HDP.

3. Release Notes HDP-Win-1.1.0

RELEASE NOTES: Hortonworks Data Platform 1.1 for Windows (Beta) powered by Apache Hadoop

3.1. Product Version: HDP-Win-1.1.0

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

- Apache Hadoop 1.0.3
- Apache Pig 0.9.3
- Apache HCatalog 0.4.1
- Apache Templeton 0.1.4
- Apache Hive 0.9.0
- Apache Sqoop 1.4.2
- Apache Oozie 3.2.0

3.2. Patch Information

In this section:

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

3.2.1. Patch Information for Hadoop

The following list provides information on the **Apache JIRAs** for this release:

- [HDFS-6527](#): Backport HADOOP-7389: Use of TestingGroups by tests causes subsequent tests to fail.
- [HADOOP-8836](#): UGI should throw exception in case `winutils.exe` cannot be loaded.
- [MAPREDUCE-4657](#): WindowsResourceCalculatorPlugin has Null Pointer Exception.
- [HADOOP-8903](#): Added support `HADOOP_USER_CLASSPATH_FIRST` in Windows Hadoop cmd.
- [HADOOP-8908](#): Refactor `winutils.exe` related code.

- [HADOOP-8911](#): CRLF characters in source and text files.
- [HADOOP-8912](#): Add .gitattributes file to prevent CRLF and LF mismatches for source and text files.
- [HADOOP-8868](#): FileUtil#chmod should normalize the path before calling into shell APIs.
- [HDFS-4065](#): TestDFSShell.testGet sporadically fails attempting to corrupt block files due to race condition.
- [HADOOP-8902](#): Enabled Gridmix v1 and v2 benchmarks on the Windows platform.
- [HADOOP-8564](#): Port and extend Hadoop native libraries for Windows to address DataNode concurrent reading and writing issue.
- [HADOOP-4093](#): Fix a bug that AzureBlockPlacementPolicy#chooseTarget only returns one DataNode when replication factor is greater than three.
- [HADOOP-8907](#): Provide means to look for zlib1.dll next to hadoop.dll on Windows.
- [HADOOP-8420](#): Hadoop Common creating package-info.java must not depend on sh.
- [HADOOP-8972](#): Move winutils tests from bat to Java.
- [HADOOP-9026](#): Hadoop.cmd fails to initialize if user's %path% variable has parenthesis.
- [HADOOP-9027](#): Build fails on Windows without sh/sed/echo in the path.
- [HADOOP-8847](#): Change untar to use Java API on Windows instead of spawning tar process.
- [HADOOP-9062](#): hadoop-env.cmd overwrites the value of *_OPTS set before install.
- [HADOOP-9074](#): Hadoop install scripts for Windows.
- [HADOOP-9110](#): winutils ls off-by-one error indexing MONTHS array can cause access violation.
- [HADOOP-9102](#): winutils task isAlive does not return a non-zero exit code if the requested task is not alive.
- [HADOOP-9061](#): Java6+Windows does not work well with symlinks.
- [HADOOP-8645](#): HADOOP_HOME and -Dhadoop.home (from hadoop wrapper script) are not uniformly handled.
- [HADOOP-9185](#): TestFileCreation.testFsClose should clean up on exit.
- [HADOOP-9191](#): TestAccessControlList and TestJobHistoryConfig fail with JDK7.
- [HADOOP-9177](#): Address issues that reported by static code analysis on winutils.
- [HADOOP-9250](#): Windows installer bugfixes

- [MAPREDUCE-4915](#): `TestShuffleExceptionCount` fails with open JDK7
- [MAPREDUCE-4914](#): `TestMiniMRDFSsort` fails with openJDK7.
- [HADOOP-9179](#): `TestFileSystem` fails with open JDK7.
- [HDFS-4358](#): `TestCheckpoint` failure with JDK7.
- [HDFS-4355](#): `TestNameNodeMetrics.testCorruptBlock` fails with open JDK7
- [MAPREDUCE-4909](#): `TestKeyValueTextInputFormat` fails with Open JDK 7 on Windows
- [HADOOP-9175](#): `TestWritableName` fails with Open JDK 7
- [HADOOP-9174](#): `TestSecurityUtil` fails with Open JDK 7
- [HDFS-4337](#): Backport HDFS-4240: For nodegroup-aware block placement, when a node is excluded, the nodes in the same nodegroup should also be excluded.
- [HDFS-4341](#): Set default data dir permission in `MiniDFSclusterWithNodeGroup`.
- [HDFS-4320](#): Add a separate configuration for NameNode RPC address instead of using `fs.default.name`.
- [HDFS-3942](#): Backport HDFS-3495 and HDFS-4234: Update Balancer to support new `NetworkTopology` with `NodeGroup` and use generic code for choosing `DataNode` in Balancer.
- [MAPREDUCE-782](#): Use `PureJavaCrc32` in MapReduce spills.
- [HDFS-496](#): Backport: Use `PureJavaCrc32` in HDFS.
- [HADOOP-7096](#): Allow setting of end-of-record delimiter for `TextInputFormat`.
- [HADOOP-8617](#): Backport HADOOP-6148, HADOOP-6166 and HADOOP-7333 for a pure Java CRC32 calculator implementation.
- [HADOOP-9111](#): Change some JUnit 3 tests to JUnit 4 so that `@Ignore` tests can be run with Ant v1.8.x.
- [HADOOP-9090](#): Support on-demand publish of metrics.
- [HADOOP-9099](#): `TestNetUtils` fails if "UnknownHost" is resolved as a valid hostname.
- [HADOOP-9095](#): Backport HADOOP-8372: `NetUtils.normalizeHostName()` incorrectly handles hostname starting with a numeric character.
- [HADOOP-9036](#): Fix racy test case `TestSinkQueue`
- [HADOOP-8900](#): `BuiltInGzipDecompressor` throws `IOException` - stored gzip size doesn't match decompressed size.
- [HADOOP-6496](#): `HttpServer` sends wrong content-type for CSS files.

- [HADOOP-7868](#): Hadoop native fails to compile when default linker option is `-Wl,-as-needed`.
- [HDFS-3941](#): Backport HDFS-3498 and HDFS-3601: Support replica removal in `BlockPlacementPolicy` and make `BlockPlacementPolicyDefault` extensible for reusing code in subclasses, and add `BlockPlacementPolicyWithNodeGroup` to support block placement with 4-layer network topology.
- [HADOOP-8820](#): Backport HADOOP-8469 and HADOOP-8470: Make `NetworkTopology` class pluggable and add `NetworkTopologyWithNodeGroup`, a 4-layer implementation of `NetworkTopology`.
- [HADOOP-8645](#): `HADOOP_HOME` and `-Dhadoop.home` (from hadoop wrapper script) are not uniformly handled.
- [MAPREDUCE-1806 \(HADOOP-136\)](#): Fixed issues in `CombineFileInputFormat` that caused failure while using Sqoop to export files in ASV.
- [HADOOP-8874](#): Added an API to retrieve valid `HADOOP_HOME` and `bin` path. This patch adds a consistency layer for `HADOOP_HOME` lookups and provides abstractions to qualify `bin` paths of hadoop binary components.
- [HADOOP-8731](#): Added Public distributed cache support for Windows.
- [HADOOP-9040](#): Added fixes for `TaskController`.
- [HADOOP-9071](#): Improved Ivy log levels.
- [HADOOP-8872](#): Fixed issue caused while invoking `FileSystem.length()` method on a Windows machine using JDK 6.x.
- [MAPREDUCE-4564](#): Fixed shell timeout mechanism. This fix enables successful termination of those processes that are spawned by **Winutils**.
- [MAPREDUCE-4561](#): Added support for node health scripts on Windows
- [MAPREDUCE-4597](#): Fixed intermittent failures for `TestKillSubProcesses`.
- [HADOOP-8739](#): Fixed command line parsing on Windows.
- [HADOOP-8664](#): Fixed the Hadoop streaming job issue that required the user to provide full path to commands.
- [HDFS-3766](#): Fixed `TestStorageRestore` on Windows.
- [HADOOP-8634](#): Fixed the errors caused when `FileSystem.deleteOnExit` method is invoked.
- [HDFS-3763](#): Fixed the `TestNameNodeMXBean` failures on Windows.
- [MAPREDUCE-4510](#): Fixed redundant checks and logging of **getconf** on Windows.
- [HADOOP-8731](#): Added public distributed cache support for Windows. This fixes the failures for `TestTrackerDistributedCacheManager`.
- [HADOOP-8763](#): Fixed issues caused when setting group owner on Windows.

- [HADOOP-8694](#): Added symlink support to Windows platform.
- [HADOOP-8732](#): Fixed test failures caused due to incorrect process serialization on Windows.
- [HDFS-3564](#): Added enhancements to the block placement policy. This enhancement enables a pluggable placement policy and provides a new API that enables moving blocks for balancing. It also enables the placement policy to decide the number of racks and provides ability to extend the policy.
- [HDFS-3566](#): Add `AzureBlockPlacementPolicy` to handle fault and upgrade domains in Azure. This policy distributes replicas across both the fault and the upgrade domains to ensure zero data loss.
- [HADOOP-8457](#): Fixed the file ownership issue for users in the Administrators groups on Windows.
- [MAPREDUCE-4374](#): Added support for configurable environment for child map/reduce tasks on Windows.
- [HADOOP-8453](#): Added unit tests for Winutils program. Winutils is the Windows console program that emulates the Linux command line utilities used by Hadoop.
- [HDFS-385](#): Added new experimental API `BlockPlacementPolicy` allows investigating alternate rules for locating block replicas.
- [HADOOP-8899](#): Fixed issues caused because of Classpath exceeding the maximum operating system (OS) limit.
- [MAPREDUCE-1806](#): Fixed issues with CombineFileInputFormat.
- [HADOOP-8935](#): Improved Winutils to handle the failures caused for the `winutils ls` command.
- [HADOOP-6496](#): Fixed the HTTPServer issue that caused incorrect rendering of the web interface for HBase.
- [HADOOP-7827](#): Fixed issue with JSP pages for web interfaces.
- [HADOOP-8903](#): Added support for `HADOOP_USER_CLASSPATH_FIRST` environment variable in the `hadoop.cmd` file.
- [HADOOP-8880](#): Fixed Hive test failures caused because of missing Jersey JAR files in the POM template.
- [HADOOP-8733](#): Fixed the failures caused by the dependencies in the `test.sh` script file.
- [MAPREDUCE-4400](#): Fixed performance regression for small jobs and workflows.
- [HADOOP-8734](#): Fixed LocalJobRunner to support private distributed cache.
- [HDFS-3833](#): Fixed TestDFSShell failures on Windows caused due to concurrent file read/write.
- [MAPREDUCE-4598](#): Added support for node health scripts on Windows.

- [HADOOP-8657](#): Fixed TestCLI to remove hardcoded value for the file length.
- [HDFS-3163](#): Fixed failures for TestHDFSCLI.testAll which occurred when the user name was not provided in lowercase.
- [HADOOP-8618](#): Fixed build failures caused due to merging of the Hadoop v1.0.3 branch.
- [HADOOP-8544](#): Moved an assertion location in **winutils chmod** command.
- [HADOOP-7389](#): Fixed test failures caused when tests use the TestingGroups.
- [HADOOP-8414](#): Fixed issues caused by localhost resolving to incorrect address on Windows.
- [MAPREDUCE-4368](#): Fixed TaskRunner to handle the event when `java.library.path` contains a quoted path with embedded spaces on Windows platform.
- [MAPREDUCE-4369](#): Fixed streaming job failures with WindowsResourceCalculatorPlugin.
- [MAPREDUCE-4332](#): Fixed command length abort issues on Windows.
- [HADOOP-8487](#): Fixed the HDFS tests to use correct test paths.
- [HADOOP-8534](#): Fixed failures for those tests that left configuration files open.
- [HADOOP-8486](#): Fixed the resource leak caused because of open file handles for SequenceFile.
- [MAPREDUCE-4203](#): Added an implementation of the process tree for Windows.
- [HADOOP-8454](#): Fixed bugs for the **chmod** command in Winutils program.
- [HADOOP-8409](#): Fixed TestCommandLineJobSubmission and TestGenericOptionsParser to work for Windows.
- [MAPREDUCE-4260](#): Added support to use JobObject for spawning tasks on Windows platform.
- [MAPREDUCE-4321](#): Fixed failures for DefaultTaskController on Windows.
- [HADOOP-8424](#): Fixed Classpath issues that caused intermittent failures for web user interface on Windows.
- [HDFS-3424](#): Fixed TestDatanodeBlockScanner and TestReplication failures on Windows.
- [HADOOP-8374](#): Improved support for hard link manipulation on Windows.
- [HADOOP-8440](#): Fixed failures for HarFileSystem.decodeHarURI.
- [HADOOP-8411](#): Fixed TestStorageDirectorFailure, TestTaskLogsTruncater, TestWebHdfsUrl and TestSecurityUtil failures on Windows.
- [HADOOP-8235](#): Added support file permissions and ownership on Windows for RawLocalFileSystem.

- [MAPREDUCE-4204](#): Improved ProcfsBasedProcessTree to enable the resource collection object to be pluggable.
- [MAPREDUCE-4201](#): Fixed issues related to obtaining PIDs on Windows.
- [HADOOP-8234](#): Enabled user group mappings on Windows platform.
- [HADOOP-8223](#): Applied initial patch for branch-1-win.
- [HDFS-4413](#): Secondary namenode won't start if HDFS isn't the default file system

3.2.2. Patch Information for Pig

The following list provides information on the **Apache JIRAs** for this release:

- [PIG-2958](#): Fixed failure issues caused when Pig tests have no associated logger.
- [PIG-2957](#): Fixed failures for TetsScriptUDF test.
- [PIG-2960](#): Increased the timeout for unit tests on Windows.
- [PIG-2943](#): Improved DevTests and Windows checks to include the **Util.Windows** method.
- [PIG-2942](#): Fixed failures for DevTests and TestLoad tests.
- [PIG-2941](#): Added consistency in chaining the Ivy resolvers. This fix also adds the fallback mechanism.
- [PIG-2953](#): Added support for **which** utility on Windows
- [PIG-2956](#): Fixed issues with invalid cache specification for some streaming statement.
- [PIG-2954](#): Fixed test failures caused due to the dependency on **bash**.
- [PIG-2959](#): Fixes for `pig.cmd` to run on Windows.
- [PIG-2801](#): Fixes for grunt **sh** command.
- [PIG-2800](#): Fixed issues with `pig.additional.jars` path separator.
- [PIG-2799](#): Updated Pig streaming interface to run correctly on Windows without Cygwin support.
- [PIG-2798](#): Fixed issues with Pig streaming tests on Windows.
- [PIG-2797](#): Fixed the Pig tests to use **Util.generateURI** method.
- [PIG-2796](#): Fixed issues with invalid path names on HDFS caused because the Pig tests use local temporary paths.
- [PIG-2795](#): Added support to handle paths generated on Windows.
- [PIG-2794](#): Added utilities to facilitate testing on Windows platform.
- [PIG-2793](#): Improved Pig to work on Windows platform without Cygwin support.

3.2.3. Patch Information for Hive

The following list provides information on the **Apache JIRAs** for this release:

- [HIVE-3025](#): Fixed Hive archive command for Hadoop v 0.22 and 0.23.
- [HIVE-3448](#): Fixed failures for the skewjoin.q testcase on Windows.
- [HIVE-3441](#): Fixed failures caused due to the partition column strings in Windows file names.
- [HIVE-3436](#): Fixed the script_pipe.q failures on Windows.
- [HIVE-3483](#): Fixed issues with joins that use partitioned table on Windows.
- [HIVE-3317](#): Fixed TestDocToUnix unit tests on Windows.
- [HIVE-3320](#): Fixed test case failures caused by incorrect handling of CRLF line endings on Windows.
- [HIVE-3319](#): Fixed path related issues that caused the unit test failures for Windows.
- [HIVE-3327](#): Fixed failures caused while execution of the `/bin/cat` script files on Windows.
- [HIVE-3479](#): Fixed issues with negative unit tests.
- [HIVE-3494](#): Fixed JDBC test case failures on Windows.
- [HIVE-3480](#): Fixed file handle leaks in Symbolic and symlink related input formats.

3.2.4. Patch Information for HCatalog

The following list provides information on the **Apache JIRAs** for this release:

- [HCATALOG-512](#): Fixed HCatalog unit tests on Windows.
- [HCATALOG-514](#): Fixed HCatalog python scripts in the package build for Windows.

3.3. Known Issues

- One of the regression test for Oozie fails with MS SQL because `java.io.IOException` does not finish within the `600000timeout`.
- One of the regression test for Oozie fails with MS SQL because `java.io.UnsupportedEncodingException` does not finish within the `600000timeout`.
- Running MapReduce jobs using pipes is currently not supported on Windows.
- Non-default compression codecs that are based on zlib or snappy are currently not supported on Windows

- It is possible to encounter the following exception while starting the Hive command line interface (CLI).

```
FAILED: Error in metadata: javax.jdo.JDOFatalDataStoreException: DERBY SQL
error: SQLCODE: -1, SQLSTATE: XJ041, SQLERRMC:
    Failed to create database 'metastore_db', see the next exception for
    details.
    ::SQLSTATE: XBM0JDirectory C:\Hadoop\hive-0.9.0\bin\metastore_db already
    exists.
```

This typically happens when the user attempts to install and uninstall HDP repeatedly. In such cases, the directories for Hive might not get deleted properly on uninstall. You can use any of the following workaround:

- Manually delete the `metastore_db` directory (`C:\Hadoop\hive-0.9.0\bin\metastore_db`).
- Uninstall HDP, delete all the files in the Hadoop directory (`C:\Hadoop`), and install HDP again
- HDP creates some files in the HDFS directory (`C:\hdfs`) that are not deleted on uninstall. This issue is observed when `hadoop.tmp.dir` is not defined to point to the `C:\hadoop` location. There is no impact on the deployment of your cluster. However, it is recommended that you manually delete the files in the HDFS directory (`C:\hdfs`) after you uninstall HDP.