

Hortonworks Data Platform 2.0

Release Notes

(Jan 20, 2014)

Hortonworks Data Platform 2.0 for Windows : Release Notes

Copyright © 2014 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-2.0.6.0	1
1.1. Product Version: HDP-Win-2.0.6.0	1
1.2. New Features	1
1.3. Patch Information	2
1.3.1. Patch Information for Hadoop	2
1.3.2. Patch Information for HBase	8
1.3.3. Patch Information for Zookeeper	9
1.3.4. Patch Information for Pig	10
1.3.5. Patch Information for Hive	11
1.3.6. Patch Information for HCatalog	14
1.3.7. Patch Information for Oozie	14
1.3.8. Patch Information for Sqoop	15
1.3.9. Patch Information for Mahout	15
1.4. Improvements	15
1.5. Limitations	16
1.6. Known Issues	16

1. Release Notes HDP-Win-2.0.6.0

RELEASE NOTES: Hortonworks Data Platform 2.0 for Windows powered by Apache Hadoop

1.1. Product Version: HDP-Win-2.0.6.0

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

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

- Apache Hadoop 2.2.0
- Apache HBase 0.96.0
- Apache Zookeeper 3.4.5
- Apache Pig 0.12.0
- Apache HCatalog 0.12.0



Note

Apache HCatalog is now merged with Apache Hive.

- Apache Hive 0.12.0
- Apache Sqoop 1.4.4
- Apache Oozie 4.0.0
- Apache Flume 1.4.0
- Apache Mahout 0.8.0

1.2. New Features

This release of Hortonworks Data Platform (HDP) includes the following new features:

- Web Console for Oozie
- Added a graphical user interface to define the cluster properties configuration for an HDP Windows MSI install
- Added Namenode HA support (Hadoop 2)

- Enhanced HCatalog support to allow implementations of direct mode connection manager

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](#)
- [Patch information for Sqoop](#)
- [Patch information for Mahout](#)

1.3.1. Patch Information for Hadoop

Hadoop is based on Apache Hadoop 2.2.0 and includes the following additional **Apache JIRAs** for this release:

- [HADOOP-682](#): Hadoop namenode -format doesn't work any more if target directory doesn't exist
- [HADOOP-717](#): When there are few reducers, sorting should be done by mappers
- [HADOOP-722](#): native-hadoop deficiencies
- [HADOOP-4093](#): Fix a bug that `AzureBlockPlacementPolicy#chooseTarget` only returns one `DataNode` when replication factor is greater than three
- [HADOOP-6496](#): Fixed the `HTTPServer` issue that caused incorrect rendering of the web interface for HBase; `HttpServer` sends wrong content-type for CSS files
- [HADOOP-7096](#): Allow setting of end-of-record delimiter for `TextInputFormat`
- [HADOOP-7389](#): Fixed test failures caused when tests use the `TestingGroups`
- [HADOOP-7827](#): Fixed issue with JSP pages for web interfaces
- [HADOOP-7868](#): Hadoop native fails to compile when default linker option is `-Wl,-as-needed`
- [HADOOP-8223](#): Applied initial patch for branch-1-win

- [HADOOP-8234](#): Enabled user group mappings on Windows platform
- [HADOOP-8235](#): Added support file permissions and ownership on Windows for RawLocalFileSystem
- [HADOOP-8374](#): Improved support for hard link manipulation on Windows
- [HADOOP-8409](#): Fixed TestCommandLineJobSubmission and TestGenericOptionsParser to work for Windows
- [HADOOP-8411](#): Fixed TestStorageDirectorFailure, TestTaskLogsTruncater, TestWebHdfsUrl and TestSecurityUtil failures on Windows
- [HADOOP-8414](#): Fixed issues caused by localhost resolving to incorrect address on Windows
- [HADOOP-8420](#): Hadoop Common creating package-info.java must not depend on sh
- [HADOOP-8424](#): Fixed Classpath issues that caused intermittent failures for web user interface on Windows
- [HADOOP-8440](#): Fixed failures for `HarFileSystem.decodeHarURI`
- [HADOOP-8453](#): Added unit tests for Winutils program. Winutils is the Windows console program that emulates the Linux command line utilities used by Hadoop
- [HADOOP-8454](#): Fixed bugs for the **chmod** command in Winutils program
- [HADOOP-8457](#): Fixed the file ownership issue for users in the Administrators groups on Windows
- [HADOOP-8486](#): Fixed the resource leak caused because of open file handles for SequenceFile
- [HADOOP-8487](#): Fixed the HDFS tests to use correct test paths
- [HADOOP-8534](#): Fixed failures for those tests that left configuration files open
- [HADOOP-8544](#): Moved an assertion location in **winutils chmod** command
- [HADOOP-8564](#): Port and extend Hadoop native libraries for Windows to address DataNode concurrent reading and writing issue
- [HADOOP-8617](#): Backport HADOOP-6148, HADOOP-6166 and HADOOP-7333 for a pure Java CRC32 calculator implementation
- [HADOOP-8618](#): Fixed build failures caused due to merging of the Hadoop v1.0.3 branch
- [HADOOP-8657](#): Fixed TestCLI to remove hardcoded value for the file length
- [HADOOP-8664](#): Fixed the Hadoop streaming job issue that required the user to provide full path to commands
- [HADOOP-8731](#): Added public distributed cache support for Windows. This fixes the failures for `TestTrackerDistributedCacheManager`

- [HADOOP-8732](#): Fixed test failures caused due to incorrect process serialization on Windows
- [HADOOP-8733](#): Fixed the failures caused by the dependencies in the `test.sh` script file
- [HADOOP-8734](#): Fixed LocalJobRunner to support private distributed cache
- [HADOOP-8739](#): Fixed command line parsing on Windows
- [HADOOP-8763](#): Fixed issues caused when setting group owner on Windows
- [HADOOP-8820](#): Backport [HADOOP-8469](#) and [HADOOP-8470](#): Make `NetworkTopology` class pluggable and add `NetworkTopologyWithNodeGroup`, a 4-layer implementation of `NetworkTopology`
- [HADOOP-8634](#): Fixed the errors caused when `FileSystem.deleteonExit` method is invoked
- [HADOOP-8836](#): UGI should throw exception in case `winutils.exe` cannot be loaded
- [HADOOP-8645](#): `HADOOP_HOME` and `-Dhadoop.home` (from `hadoop` wrapper script) are not uniformly handled
- [HADOOP-8694](#): Added symlink support to Windows platform
- [HADOOP-8847](#): Change `untar` to use Java API on Windows instead of spawning `tar` process
- [HADOOP-8868](#): `FileUtil#chmod` should normalize the path before calling into shell APIs
- [HADOOP-8872](#): Fixed issue caused while invoking `FileSystem.length()` method on a Windows machine using JDK 6.x
- [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-8880](#): Fixed Hive test failures caused because of missing Jersey JAR files in the POM template
- [HADOOP-8899](#): Fixed issues caused because of Classpath exceeding the maximum operating system (OS) limit
- [HADOOP-8900](#): `BuiltInGzipDecompressor` throws `IOException` - stored `gzip` size doesn't match decompressed size
- [HADOOP-8902](#): Enabled Gridmix v1 and v2 benchmarks on the Windows platform
- [HADOOP-8903](#): Added support for `HADOOP_USER_CLASSPATH_FIRST` environment variable in the `hadoop.cmd` file
- [HADOOP-8907](#): Provide means to look for `zlib1.dll` next to `hadoop.dll` on Windows
- [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-8935](#): Improved Winutils to handle the failures caused for the `winutils ls` command
- [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-9036](#): Fix racy test case `TestSinkQueue`
- [HADOOP-9040](#): Added fixes for `TaskController`
- [HADOOP-9061](#): Java6+Windows does not work well with symlinks
- [HADOOP-9062](#): `hadoop-env.cmd` overwrites the value of `*_OPTS` set before install
- [HADOOP-9071](#): Improved Ivy log levels
- [HADOOP-9074](#): Hadoop install scripts for Windows
- [HADOOP-9090](#): Support on-demand publish of metrics
- [HADOOP-9095](#): Backport [HADOOP-8372](#): `NetUtils.normalizeHostName()` incorrectly handles hostname starting with a numeric character
- [HADOOP-9099](#): `TestNetUtils` fails if "UnknownHost" is resolved as a valid hostname
- [HADOOP-9102](#): winutils task `isAlive` does not return a non-zero exit code if the requested task is not alive
- [HADOOP-9110](#): winutils `ls` off-by-one error indexing `MONTHS` array can cause access violation
- [HADOOP-9174](#): `TestSecurityUtil` fails with Open JDK 7
- [HADOOP-9175](#): `TestWritableName` fails with Open JDK 7
- [HADOOP-9177](#): Address issues that reported by static code analysis on winutils
- [HADOOP-9179](#): `TestFileSystem` fails with open JDK7
- [HADOOP-9185](#): `TestFileCreation.testFsClose` should clean up on exit
- [HADOOP-9191](#): `TestAccessControlList` and `TestJobHistoryConfig` fail with JDK7
- [HADOOP-9111](#): Change some JUnit 3 tests to JUnit 4 so that `@Ignore` tests can be run with Ant v1.8.x
- [HADOOP-9250](#): Windows installer bugfixes

- [HADOOP-9660](#): [WINDOWS] Powershell / cmd parses `-Dkey=value` from command line as `[-Dkey, value]` which breaks `GenericOptionsParser`
- [HADOOP-10093](#): `hadoop-env.cmd` sets `HADOOP_CLIENT_OPTS` with a max heap size that is too small
- [HADOOP-10094](#): NPE in `GenericOptionsParser#preProcessForWindows()`
- [MAPREDUCE-782](#): Use `PureJavaCrc32` in `MapReduce` spills
- [MAPREDUCE-1806 \(HADOOP-136\)](#): Fixed issues in `CombineFileInputFormat` that caused failure while using `Sqoop` to export files in `ASV`
- [MAPREDUCE-4201](#): Fixed issues related to obtaining PIDs on Windows
- [MAPREDUCE-4203](#): Added an implementation of the process tree for Windows
- [MAPREDUCE-4204](#): Improved `ProcfsBasedProcessTree` to enable the resource collection object to be pluggable
- [MAPREDUCE-4260](#): Added support to use `JobObject` for spawning tasks on Windows platform
- [MAPREDUCE-4321](#): Fixed failures for `DefaultTaskController` on Windows
- [MAPREDUCE-4332](#): Fixed command length abort issues 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-4374](#): Added support for configurable environment for child map/reduce tasks on Windows
- [MAPREDUCE-4400](#): Fixed performance regression for small jobs and workflows
- [MAPREDUCE-4510](#): Fixed redundant checks and logging of `getconf` on Windows
- [MAPREDUCE-4561](#): Added support for node health scripts on Windows
- [MAPREDUCE-4564](#): Fixed shell timeout mechanism. This fix enables successful termination of those processes that are spawned by **Winutils**
- [MAPREDUCE-4597](#): Fixed intermittent failures for `TestKillSubProcesses`
- [MAPREDUCE-4598](#): Added support for node health scripts on Windows
- [MAPREDUCE-4657](#): `WindowsResourceCalculatorPlugin` has Null Pointer Exception.
- [MAPREDUCE-4909](#): `TestKeyValueTextInputFormat` fails with Open JDK 7 on Windows
- [MAPREDUCE-4914](#): `TestMiniMRDFSsort` fails with openJDK7
- [MAPREDUCE-4915](#): `TestShuffleExceptionCount` fails with open JDK7

- [MAPREDUCE-5451](#): MR uses LD_LIBRARY_PATH which doesn't mean anything in Windows
- [MAPREDUCE-5604](#): TestMRAMWithNonNormalizedCapabilities fails on Windows due to exceeding max path length
- [MAPREDUCE-5616](#): MR Client-AppMaster RPC max retries on socket timeout is too high
- [HDFS-385](#): Added new experimental API BlockPlacementPolicy allows investigating alternate rules for locating block replicas
- [HDFS-496](#): Backport: Use PureJavaCrc32 in HDFS
- [HDFS-3163](#): Fixed failures for TestHDFSCLI.testAll which occurred when the user name was not provided in lowercase
- [HDFS-3424](#): Fixed TestDatanodeBlockScanner and TestReplication failures 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
- [HDFS-3763](#): Fixed the TestNameNodeMXBean failures on Windows
- [HDFS-3766](#): Fixed TestStorageRestore on Windows
- [HDFS-3833](#): Fixed TestDFSShell failures on Windows caused due to concurrent file read/write
- [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
- [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
- [HDFS-4065](#): TestDFSShell.testGet sporadically fails attempting to corrupt block files due to race condition
- [HDFS-4320](#): Add a separate configuration for NameNode RPC address instead of using fs.default.name
- [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-4355](#): TestNameNodeMetrics.testCorruptBlock fails with open JDK7

- [HDFS-4358](#): `TestCheckpoint` failure with JDK7
- [HDFS-4413](#): Secondary namenode won't start if HDFS isn't the default file system
- [HDFS-4633](#): `TestDFSClientExcludedNodes` fails sporadically if excluded nodes cache expires too quickly
- [HDFS-5065](#): `TestSymlinkHdfsDisable` fails on Windows
- [HDFS-5089](#): When a `LayoutVersion` supports `SNAPSHOT`, it must support `FSIMAGE_NAME_OPTIMIZATION`
- [HDFS-5338](#): Add a conf to disable hostname check in DN registration
- [HDFS-5375](#): `hdfs . cmd` does not expose several snapshot commands
- [HDFS-5413](#): `hdfs . cmd` does not support passthrough to any arbitrary class
- [HDFS-5432](#): `TestDatanodeJsp` fails on Windows due to assumption that loopback address resolves to host name
- [HDFS-5456](#): NameNode startup progress creates new steps if caller attempts to create a counter for a step that doesn't already exist
- [HDFS-6527](#): Backport HADOOP-7389: Use of `TestingGroups` by tests causes subsequent tests to fail
- [YARN-1331](#): `yarn . cmd` exits with `NoClassDefFoundError` trying to run `rmadmin` or `logs`
- [YARN-1357](#): `TestContainerLaunch . testContainerEnvVariables` fails on Windows
- [YARN-1358](#): `TestYarnCLI` fails on Windows due to line endings
- [YARN-1349](#): `yarn . cmd` does not support passthrough to any arbitrary class
- [YARN-1395](#): Distributed shell application master launched with debug flag can hang waiting for external `ls` process
- [BUG-8178](#): Datanodes fail to register with namenode due to minimum version check

1.3.2. Patch Information for HBase

HBase is based on Apache HBase 0.96.0 and includes the following **Apache JIRAs** for this release:

- [HBASE-9562](#): Make `HLogPE` run against configured FS
- [HBASE-9639](#): `SecureBulkLoad` dispatches file load requests to all Regions
- [HBASE-9685](#): `IntegrationTestMTTR` should stop on `RetriesExhaustedException`
- [HBASE-9768](#): Two issues in `AsyncProcess`
- [HBASE-9772](#): Normalize new client default values

- [HBASE-9784](#): Switch to Hadoop-2.2
- [HBASE-9786](#): hbck -metaonly incorrectly reports inconsistent regions after HBASE-9698 fix
- [HBASE-9788](#): [WINDOWS] Update rest server class name
- [HBASE-9796](#): npe in RegionServerCallable
- [HBASE-9798](#): Include dependency hamcrest-core
- [HBASE-9813](#): Log splitting doesn't prevent RS creating new hlog file
- [HBASE-9827](#): Intermittent
`TestLogRollingNoCluster#testContendedLogRolling` failure
- [HBASE-9843](#): Various fixes in client code
- [HBASE-9862](#): Manage error per server and per region in the protobuffed client
- [HBASE-9867](#): Save on array copies with a subclass of LiteralByteString
- [HBASE-9868](#): Remove some array copy, especially around protobuf
- [HBASE-9880](#): `client.TestAsyncProcess.testWithNoClearOnFail` broke on 0.96 by HBASE-9867
- [HBASE-9885](#): Avoid some Result creation in protobuf conversions
- [HBASE-9886](#): Optimize `ServerName#compareTo`
- [HBASE-9906](#): Restore snapshot fails to restore the meta edits sporadically
- [HBASE-9908](#): Fix filesystem / classloader-related unit tests
- [HBASE-9909](#): `TestFilePerformance` should not be a unit test, but a tool
- [HBASE-9952](#): Snapshot restore may fail due to `NullPointerException`
- [HBASE-9957](#): [WINDOWS] `TestNamespaceUpgrade` fails on Windows
- [HBASE-9961](#): [WINDOWS] Multicast bind to local address
- [HBASE-10084](#): [WINDOWS] `bin\hbase.cmd` should allow whitespaces in values for `java.library.path` and `classpath`
- [HBASE-10098](#): [WINDOWS] pass in native library directory from hadoop for unit tests
- [HBASE-10163](#): Example Thrift `DemoClient` is flaky

1.3.3. Patch Information for Zookeeper

Zookeeper is based on Apache Zookeeper 3.4.5 and includes the following **Apache JIRAs** for this release:

- [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

Hadoop is based on Apache Pig 0.12.0 and includes the following **Apache JIRAs** for this release:

- [PIG-94](#): Valid Pig queries with a slash (back or forward) fail
- [PIG-95](#): Pig + HCat e2e tests fail with Not a Valid Application error
- [PIG-96](#): Pig e2e tests fail with perl errors
- [PIG-97](#): Jython_Check_2 test fails with sort check failure
- [PIG-98](#): Native_3 is trying to run `/bin/cat` instead it should just run `cat` as that would work on linux and windows
- [PIG-99](#): RubyUDFs_10 checksum match fails
- [PIG-100](#): Pig Unit Test Failures
- [PIG-104](#): piggybank.jar should be available under `PIG_HOME`
- [PIG-106](#): Pig tests fail with module stringutil not found
- [PIG-107](#): Jython_CompileBindRun_11 fails with file not found error
- [PIG-114](#): StreamingPerformance_4 fails as tasks time out after 5 minutes and get killed
- [PIG-2793](#): Improved Pig to work on Windows platform without Cygwin support
- [PIG-2794](#): Added utilities to facilitate testing on Windows platform
- [PIG-2795](#): Added support to handle paths generated on Windows
- [PIG-2796](#): Fixed issues with invalid path names on HDFS caused because the Pig tests use local temporary paths
- [PIG-2797](#): Fixed the Pig tests to use `Util.generateURI` method
- [PIG-2798](#): Fixed issues with Pig streaming tests on Windows
- [PIG-2799](#): Updated Pig streaming interface to run correctly on Windows without Cygwin support
- [PIG-2801](#): Fixes for grunt `sh` command
- [PIG-2800](#): Fixed issues with `pig.additional.jars` path separator
- [PIG-2941](#): Added consistency in chaining the Ivy resolvers. This fix also adds the fallback mechanism
- [PIG-2942](#): Fixed failures for DevTests and TestLoad tests
- [PIG-2943](#): Improved DevTests and Windows checks to include the `Util.Windows` method.
- [PIG-2953](#): Added support for `which` utility on Windows

- [PIG-2954](#): Fixed test failures caused due to the dependency on **bash**
- [PIG-2956](#): Fixed issues with invalid cache specification for some streaming statement
- [PIG-2957](#): Fixed failures for TetsScriptUDF test
- [PIG-2958](#): Fixed failure issues caused when Pig tests have no associated logger
- [PIG-2959](#): Fixes for `pig.cmd` to run on Windows
- [PIG-2960](#): Increased the timeout for unit tests on Windows
- [PIG-3257](#): Add unique identifier UDF
- [PIG-3512](#): Reducer estimate is broken by PIG-3497
- [PIG-3516](#): Pig does not bring in joda-time as dependency in its pig-template.xml
- [PIG-3517](#): Fix PermGen error in Pig Unit test on Hadoop 2
- [PIG-3518](#): Need to ship jruby.jar in the release
- [PIG-3572](#): Fix all unit test for during build pig with Hadoop 2.x on Windows
- [PIG-3635](#): Fix e2e tests for Hadoop 2.x on Windows

1.3.5. Patch Information for Hive

Hive is based on Apache Hive 0.12.0 and includes the following **Apache JIRAs** for this release:

- [HIVE-1511](#): Hive plan serialization is slow
- [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
- [HIVE-3815](#): Hive table rename fails if filesystem cache is disabled
- [HIVE-3846](#): alter view rename NPEs with authorization on
- [HIVE-4388](#): HBase tests fail against Hadoop 2
- [HIVE-4417](#): Make unit tests pass on Windows which use Linux script in query files
- [HIVE-4485](#): beeline prints null as empty strings
- [HIVE-4545](#): HS2 should return describe table results without space padding
- [HIVE-4679](#): WebHCat can deadlock Hadoop if the number of concurrently running tasks is higher or equal than the number of mappers
- [HIVE-4763](#): Add support for thrift over http transport in HS2
- [HIVE-4831](#): QTestUtil based test exiting abnormally on windows fails startup of other QTestUtil tests
- [HIVE-4844](#): Add varchar data type
- [HIVE-4910](#): Hadoop 2 archives broken
- [HIVE-5070](#): Need to implement `listLocatedStatus()` in ProxyFileSystem for 0.23 shim
- [HIVE-5072](#): [WebHCat] Enable directly invoke Sqoop job through Templeton
- [HIVE-5098](#): Fix metastore for SQL Server
- [HIVE-5099](#): Some partition publish operation cause OOM in metastore backed by SQL Server
- [HIVE-5112](#): Upgrade protobuf to 2.5 from 2.4
- [HIVE-5115](#): WebHCat e2e tests TestMapReduce_1 and TestHeartbeat_2 require changes for Hadoop 2
- [HIVE-5127](#): Upgrade xerces and xalan for WebHCat
- [HIVE-5129](#): Multiple table insert fails on count(distinct)
- [HIVE-5133](#): WebHCat jobs that need to access metastore fails in secure mode
- [HIVE-5150](#): UnsatisfiedLinkError when running hive unit test
- [HIVE-5156](#): HiveServer2 jdbc ResultSet.close should free up resources on server side
- [HIVE-5176](#): Wincompat: Changes for allowing various path compatibilities with Windows
- [HIVE-5218](#): datanucleus does not work with MS SQLServer in Hive metastore

- [HIVE-5229](#): Better thread management for Iveserver2 async threads
- [HIVE-5241](#): Default log4j level for WebHcat should be INFO not DEBUG
- [HIVE-5260](#): Introduce HivePassThroughOutputFormat that allows Hive to use general purpose OutputFormats instead of HiveOutputFormats in StorageHandlers
- [HIVE-5261](#): Make the Hive HBase storage handler work from HCatalog, and use HiveStorageHandlers instead of HCatStorageHandlers
- [HIVE-5263](#): Query Plan cloning time could be improved by using Kryo
- [HIVE-5274](#): HCatalog package renaming backward compatibility follow-up
- [HIVE-5278](#): Move some string UDFs to GenericUDFs, for better varchar support
- [HIVE-5279](#): Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`
- [HIVE-5290](#): Some HCatalog tests have been behaving flaky
- [HIVE-5353](#): Job submission that requires access to metastore should not require additional jars to be shipped to target node
- [HIVE-5364](#): NPE on some queries from partitioned orc table
- [HIVE-5379](#): `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- [HIVE-5410](#): Hive command line option `-auxpath` still does not work post HIVE-5363
- [HIVE-5411](#): Migrate expression serialization to Kryo
- [HIVE-5413](#): `StorageDelegationAuthorizationProvider` uses non-existent `org.apache.hive.hcatalog.hbase.HBaseHCatStorageHandler`
- [HIVE-5422](#): Upgrade Kryo to 2.22 now that it is released
- [HIVE-5425](#): Provide a configuration option to control the default stripe size for ORC
- [HIVE-5431](#): PassthroughOutputFormat SH changes causes `IllegalArgumentExpection`
- [HIVE-5433](#): Fix varchar unit test to work with hadoop-2.1.1
- [HIVE-5448](#): WebHcat duplicate test `TestMapReduce_2` should be removed
- [HIVE-5453](#): `jobsubmission2.conf` should use 'timeout' property
- [HIVE-5474](#): drop table hands when `concurrency=true`
- [HIVE-5478](#): WebHcat e2e testsuite for hcat authorization tests needs some fixes
- [HIVE-5479](#): SBAP restricts `hcat -e'show databases'`
- [HIVE-5480](#): WebHcat e2e tests or doAs feature are failing
- [HIVE-5484](#): TestSchema failures when Hive version has more than 3 revision numbers

- [HIVE-5485](#): SBAP errors on null partition being passed into partition level authorization
- [HIVE-5496](#): `hcat -e "drop database if exists"` fails on authorizing non-existent null db
- [HIVE-5507](#): [WebHCat] `test.other.user.name` parameter is missing from build.xml in e2e harness
- [HIVE-5508](#): [WebHCat] ignore log collector e2e tests for Hadoop 2
- [HIVE-5511](#): percentComplete returned by job status from WebHCat is null
- [HIVE-5115](#): WebHCat e2e tests TestMapReduce_1 and TestHeartbeat_2 require changes or Hadoop 2
- [HIVE-5542](#): WebHCat is failing to run ddl command on a secure cluster
- [HIVE-5789](#): WebHCat E2E tests do not launch on Windows
- [HIVE-5975](#): [WebHCat] templeton mapreduce job failed if provide "define" parameters
- [HIVE-6035](#): Windows: percentComplete returned by job status from WebHCat is null
- [HIVE-6064](#): Wincompat; wndows path substitutions overridden by `MiniMrShim.getConfiguration()` on hadoop 2
- [HIVE-6065](#): `orc/FileDump` and `testRCFile` should close on appropriate resources
- [HIVE-6066](#): Wincompat: `describe_comment_nonascii.q` failing on Windows
- [HIVE-6068](#): HiveServer2 beeline client on windows does not handle the non-ascii characters properly
- [HIVE-6071](#): Create hive command script for Windows, `hive.cmd`
- [HIVE-6087](#): Hive/Templeton winpkg build
- [HIVE-6088](#): Change StreamingDelegator to pass distributed cache files to jar delegator
- [HCATALOG-512](#): [HCATALOG] Fixed HCatalog unit tests on Windows
- [HCATALOG-514](#): [HCATALOG] Fixed HCatalog python scripts in the package build for Windows

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 **Apache JIRAs** for this release:

- [OOZIE-108](#): Port `oozie distro/src/main/bin/oozie-setup.ps1` script to work with hadoop 2

- [OOZIE-112](#): `oozie.war` has two copies of `WEB-INF\web.xml` file
- [OOZIE-113](#): oozie is still referring to old hadoop config files in the installer code
- [OOZIE-121](#): Oozie fails to start
- [OOZIE-122](#): Oozie configs should be picked up from `HADOOP_CONF_DIR`
- [OOZIE-127](#): Enable HCatalog confis in Oozie
- [OOZIE-129](#): `TestCoordRerun7` fails with `DONEWITHERROR` rather than the expected `RUNNING` state
- [OOZIE-137](#): `ooziedb` command fails in HDP 2.0 clusters on Azure
- [OOZIE-145](#): UnitTests fail because of wrong path
- [OOZIE-148](#): Wrong tests in `TestCoordActionInputCheckXCommand`
- [OOZIE-163](#): Oozie DB upgrade command fails
- [OOZIE-164](#): Oozie upgrade DB command fails due to missing dependencies
- [OOZIE-167](#): remove `col.jar` from oozie windows
- [OOZIE-168](#): cron-schedule job does not work after upgrade

1.3.8. Patch Information for Sqoop

Sqoop is based on Apache Sqoop 1.4.4 and contains the following **Apache JIRAs** for this release:

- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static
- [SQOOP-1170](#): Can't import columns with name "public"
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+
- [SQOOP-1179](#): Incorrect warning saying `-hive-import` was not specified when it was specified
- [SQOOP-1185](#): `LobAvrollImportTestCase` is sensitive to test method order execution

1.3.9. Patch Information for Mahout

Mahout is based on Apache Mahout 0.8.0 and contains the following **Apache JIRAs** for this release:

- [MAHOUT-14](#): Mahout clustering tests fail with `java.lang.IllegalStateException`

1.4. Improvements

This section describes improvements introduced in this release:

- [SQOOP-1107](#): Further improve error reporting when exporting malformed data
- [SQOOP-1132](#): Print out Sqoop version into log during execution
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only
- [SQOOP-1190](#): Class `HCatHadoopShims` will be removed in HCatalog 0.12

1.5. Limitations

- HDP only supports the Hortonworks provided compression codecs for zlib and snappy.

1.6. Known Issues

- HDP smoke tests fail when running as a non-hadoop user.

Problem: MSI installer does not set permissions to allow users other than the hadoop user to run jobs on the HDP 2.0 cluster for Windows.

Workaround: As the hadoop user, set permissions for the `/mapred` directory in Hadoop:

```
C:\hdp\hadoop-2.2.0.2.0.6.0-0009>hadoop fs -chown 757 /mapred
```

You should now be able to run smoke tests as Administrator or another non-hadoop user that has been added to the Hadoop group.

- Windows Firewall Stealth Mode interferes with an RPC interaction between the MapReduce client, the MapReduce YARN Application Master, and the MapReduce History Server. The symptom of this problem is that the submitter of a MapReduce job appears to stall right after job completion. During the stall, the console logs show repeated re-attempts of an RPC connection. Under default configuration, the period of this stall can last up to 15 minutes.

Problem: Windows Firewall Stealth Mode intentionally drops outbound TCP RST packets if a client initiates a connection to a destination port that has no server listening. You can read more from the Microsoft Technet site [here](#). This behavior interferes with the design of the MapReduce client, which initiates a failover to the History Server after detecting that the connection to the Application Master is no longer available. By dropping outbound TCP RST packets under Stealth Mode, the client does not observe an immediate connection refused. Instead, it must wait for a longer connection timeout period before initiating failover to the History Server.

Workaround: Disable Stealth Mode by setting registry key `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\SharedAccess\Parameters\FirewallPolicy\PublicProfile` with the `DisableStealthMode` value. For details, see [Disable Stealth Mode](#). Restart the Windows Firewall service to make this change take effect.

- Running MapReduce jobs using pipes is 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 workarounds:

- 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.
- Migrating data and tables from SqlLite to Oracle must be performed manually to prevent data loss.
- HADOOP-717: NameNode Upgrade Error Messges convert to Warnings
- HADOOP-722: YARN client returns with application not found error if RM restarted within 5s of job submission
- HADOOP-682: A lot of retries seen while running Pig tests causing tests to be slow
- ONPREM-397: Cancelling installation midway hangs
- ONPREM-400: Installer sometimes hangs during unzip phase
- ONPREM-402: Services being configured to restart immediately might cause issues with HA