Hortonworks Data Platform

HDP-2.2.4 Release Notes for Windows

(Apr 13, 2015)

docs.hortonworks.com

Hortonworks Data Platform: HDP-2.2.4 Release Notes for Windows

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.

For more information about Hortonworks technology, please visit the Hortonworks Data Platform page. For more information on Hortonworks services, please visit either the Support or Training page. Feel free to contact us directly to discuss your specific needs.

Table of Contents

1. HDP-2.2.4 Release Notes for Windows

Official Apache Versions in HDP 2.2.4 for Windows

All HDP 2.2 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 2.2 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.2.

Apache Version	Apache Version
Apache Hadoop 2.6.0	Apache DataFu 1.2.0
Apache Falcon 0.6.0	Apache Flume 1.5.2
Apache HBase 0.98.4	Apache Hive 0.14.0
Apache Knox 0.5.0	Apache Mahout 1.0.0
Apache Oozie 4.1.0	Apache Pig 0.14.0
Apache Phoenix 4.2.0	Apache Ranger 0.4.0
Apache Slider 0.60	Apache Sqoop 1.4.5
Apache Storm 0.9.3	Apache Tez 0.5.2
Apache Zookeeper 3.4.6	

Included third-party tools:

- Ganglia 3.5.0
- Nagios 3.5.0
- Ganglia Web 3.5.7

The following components are supported on HDP 2.2 for Linux, and are not currently supported on HDP 2.2 for Windows:

- Apache Accumulo 1.6.1
- Hue 2.6.1
- Apache Kafka 0.8.1.1

1.1. Behavioral Changes

The following subsections describe behavioral changes for HDP 2.2.

1.1.1. HDFS Behavioral Changes

(HDFS-6376) distcp can copy data between HA clusters. You can use the new configuration property dfs.internal.nameservices to explicitly specify the name services belonging to the

local cluster, while continue using the configuration property dfs.nameservices to specify all the name services in the local and remote clusters.

For example, if you perform distcp between two clusters (c1 and c2), both of which run in the HA mode, make the following configuration changes on the cluster that runs the distcp job:

1. Set dfs.nameservices to both c1 and c2.

2. Modify the configuration to include the settings for both clusters. See the comments in HDFS-6376 for more details.

3. Set dfs.internal.nameservice to the nameservice id that corresponds to the cluster that runs the distcp job.

1.1.2. Hive Behavioral Changes

HIVE-6432: Hive 0.14 removes all org.apache.hcatalog.* classes in favor of org.apache.hive.hcatalog.* classes.

The following previously deprecated features were also removed in this release as part of this change:

- HdfsAuthorizationProvider - the Hive StorageBasedAuthorizationProvider replaces it.

- HBaseHCatStorageHandler - Hive's native HBaseStorageHandler is now directly usable from HCatalog.

HIVE-6455 If you store Hive data in a columnar format, such as ORC, verify that the hive.optimize.sort.dynamic.partition property is set to true. It was the default value for Hive 0.13.x, but is now disabled by default in Hive 0.14.x.

Enabling dynamic partitions improves Hive performance when using columnar formats.

```
<property>
    <name>hive.optimize.sort.dynamic.partition</name>
    <value>true</value>
 </property>
```

(HIVE-5940, HIVE-6252) Added SHOW GRANT and SHOW ROLE GRANT. Users will be able to list only their own privileges on a table/role unless they are are in an admin role.

For grant/revoke statements, if a user is granted a privilege WITH GRANT OPTION on a table or view, then the user can also grant that privilege other users.

HIVE-6972 The JDBC URI for HiveServer2 has the form:

```
jdbc:hive2://<host1>:<port1>,<host2>:<port2>/dbName;sess_var_list?
hive_conf_list#hive_var_list
```

Ideally, sess_var_list is supposed to be used to pass parameters that will be used within the JDBC Driver. However, some of the http mode parameters were being passed in the

hive_conf_list which is fixed now (in a backward-compatible way). Additionally, JDBC Driver has a naming convention for the parameters in the JDBC URI which were not followed to the sasl.qop parameter. It is renamed it in a backward compatible way.

- HTTP Mode parameters have changed. As you can see in the following examples, hive.server2.transport.mode has been renamed to transportMode, hive.server2.thrift.http.path has been renamed to httpPath and both have been moved from hive_conf_list to sess_var_list.
  - Old example URI:

    ```
    jdbc:hive2://<host>:<port>/<db>;user=<username>;password=<password>?hive.
    server2.transport.mode=http;hive.server2.thrift.http.path=<http_endpoint>
    ```

  - New example URI:

    ```
    jdbc:hive2://<host>:<port>/<db>;user=<username>;password=
    <password>;transportMode=http;httpPath=<http_endpoint>?hive.server2.
    logging.operation.enabled=false
    ```

- Sasl qop renamed to saslQop:
  - Old example URI:

    ```
    jdbc:hive2://<host>:<port>/<db>;principal=
    <hiveserver2_kerberos_principal>;sasl.qop=<qop_value>
    ```

  - New example URI:

    ```
    jdbc:hive2://<host>:<port>/<db>;principal=
    <hiveserver2_kerberos_principal>;saslQop=<qop_value>
    ```

    All changes are backward-compatible and print a deprecation warning message similar to the following:

    ```
    0: jdbc:hive2://localhost:10007> !connect jdbc:hive2://localhost:10007/
    ;sasl.qop=auth?hive.server2.transport.mode=http;hive.server2.thrift.http.
    path=cliservice
     user user org.apache.hive.jdbc.HiveDriver
     Connecting to jdbc:hive2://localhost:10007/;sasl.qop=auth?hive.server2.
    transport.mode=http;hive.server2.thrift.http.path=cliservice
     14/10/07 16:22:24 INFO jdbc.Utils: Supplied authorities: localhost:10007
     14/10/07 16:22:24 WARN jdbc.Utils: ***** JDBC param deprecation *****
     14/10/07 16:22:24 WARN jdbc.Utils: The use of sasl.qop is deprecated.
     14/10/07 16:22:24 WARN jdbc.Utils: Please use saslQop like so:
     jdbc:hive2://<host>:<port>/dbName;saslQop=<qop_value>
     14/10/07 16:22:24 WARN jdbc.Utils: ***** JDBC param deprecation *****
     14/10/07 16:22:24 WARN jdbc.Utils: The use of hive.server2.transport.mode
     is deprecated.
     14/10/07 16:22:24 WARN jdbc.Utils: Please use transportMode like so:
     jdbc:hive2://<host>:<port>/dbName;transportMode=<transport_mode_value>
     14/10/07 16:22:24 WARN jdbc.Utils: ***** JDBC param deprecation *****
     14/10/07 16:22:24 WARN jdbc.Utils: The use of hive.server2.thrift.http.
    path is deprecated.
     14/10/07 16:22:24 WARN jdbc.Utils: Please use httpPath like so:
     jdbc:hive2://<host>:<port>/dbName;httpPath=<http_path_value>
     Connected to: Apache Hive (version 0.14.0-SNAPSHOT)
     Driver: Hive JDBC (version 0.14.0-SNAPSHOT)
     Transaction isolation: TRANSACTION_REPEATABLE_READ
    ```

Be aware that org.apache.hcatalog.pig.HCatLoader was deprecated in Hive 0.13.x and removed in Hive 0.14.x. You must now use org.apache.hive.hcatalog.pig.HCatLoader.

1.1.3. Oozie Behavioral Changes

The Oozie system environment requires JDK 1.6 or higher. On the machine where you the run oozie command line, set $JAVA_HOME to a java version of 1.6 or higher.

Calling logj4 has changed in Oozie for HDP 2.2. If you are using custom configurations or Ambari to install, add the following line to oozie-log4j.properties:

```
log4j.appender.oozie.layout.ConversionPattern=%d{ISO8601} %5p
%c{1}:%L - SERVER[${oozie.instance.id}] %m%n
```

where ${oozie.instance.id} is determined by Oozie automatically.

1.1.4. Pig Behavioral Changes

Be aware that org.apache.hcatalog.pig.HCatLoader was deprecated in Hive 0.13.x and removed in Hive 0.14.x. You must now use org.apache.hive.hcatalog.pig.HCatLoader.

1.2. Tech Previews

All HDP 2.2 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 2.2 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.2.

- HDFS Encryption
- HDFS Read/Write Storage Tier
- HDFS-6932 HDFS Blocking Mover/Balancer from moving files/blocks to/from transient storage.
- Falcon Data Lineage
- Falcon Graph View of Dependencies
- Phoenix support for local secondary indexes
- Spark 1.2.0 over YARN

1.3. Improvements

The following components are now available on Windows:

- Apache Ranger

- HBase and Storm on YARN via Slider

The following improvements are now available on Windows:

- HDFS DataNode caching
- C libHDFS API

1.4. Apache Patch Information

This section describes Apache JIRAs that were addressed in HDP 2.2.4.

1.4.1. Accumulo 1.6.1

HDP 2.2.4 provides Accumulo 1.6.1 and the following Apache patches:

- ACCUMULO-3312 Fix incorrect path generation during clone table.
- ACCUMULO-3600 Fix deleterows to properly compute file counts.

1.4.2. Falcon 0.6.0

HDP 2.2.4 provides Apache Falcon 0.6.0 and the following additional Apache patches:

- FALCON-1005: In ProcessInstanceSuspendTest clean test dir after each method execution
- FALCON-1007: Improve test output and variable names.
- FALCON-1072: Dumping of Oozie info should use os specific newline

1.4.3. Hadoop Common/HDFS 2.6.0

HDP 2.2.4 provides Apache Hadoop Core 2.6.0 and the following additional Apache patches:

- HDFS-3107. Introduce truncate.
- HDFS-7009. Active NN and standby NN have different live nodes.
- HDFS-7056. Snapshot support for truncate.
- HDFS-7058. Tests for truncate CLI
- HDFS-7263. Snapshot read can reveal future bytes for appended files.
- HDFS-7425. NameNode block deletion logging uses incorrect appender.
- HDFS-7443. Datanode upgrade to BLOCKID_BASED_LAYOUT fails if duplicate block files are present in the same volume
- HDFS-7470. SecondaryNameNode need twice memory when calling reloadFromImageFile.

- HDFS-7489. Incorrect locking in FsVolumeList#checkDirs can hang datanodes.
- HDFS-7503. Namenode restart after large deletions can cause slow processReport
- HDFS-7606. Fix potential NPE in INodeFile.getBlocks().
- HDFS-7634. Disallow truncation of Lazy persist files.
- HDFS-7638. Small fix and few refinements for FSN#truncate.
- HDFS-7643. Test case to ensure lazy persist files cannot be truncated.
- HDFS-7655. Expose truncate API for Web HDFS.
- HDFS-7659. Truncate should check negative value of the new length
- HDFS-7676. Fix TestFileTruncate to avoid bug of HDFS-7611.
- HDFS-7677. DistributedFileSystem#truncate should resolve symlinks.
- HDFS-7707. Edit log corruption due to delayed block removal again.
- HDFS-7714. Simultaneous restart of HA NameNodes and DataNode can cause DataNode to register successfully with only one NameNode.
- HDFS-7733. NFS: readdir/readdirplus return null directory attribute on failure.
- HDFS-7738. Revise the exception message for recover lease; add more truncate tests such as truncate with HA setup, negative tests, truncate with other operations and multiple truncates.
- HDFS-7760. Document truncate for WebHDFS.
- HDFS-7831. Fix the starting index and end condition of the loop in FileDiffList.findEarlierSnapshotBlocks().
- HDFS-7843. A truncated file is corrupted after rollback from a rolling upgrade.
- HDFS-7885. Datanode should not trust the generation stamp provided by client.
- HADOOP-926. Do not fail job history iteration when encounting missing directories.
- HADOOP-941: Addendum patch.
- HADOOP-11321. copyToLocal cannot save a file to an SMB share unless the user has Full Control permissions.
- HADOOP-11368. Fix SSLFactory truststore reloader thread leak in KMSClientProvider.
- HADOOP-11381. Fix findbugs warnings in hadoop-distcp, hadoop-aws, hadoop-azure, and hadoop-openstack
- HADOOP-11381. Revert. (Fix findbugs warnings in hadoop-distcp, hadoop-aws, hadoop-azure, and hadoop-openstack.)

- HADOOP-11412. POMs mention "The Apache Software License" rather than "Apache License".
- HADOOP-11412. Revert. (POMs mention "The Apache Software License" rather than "Apache License"
- HADOOP-11490. Expose truncate API via FileSystem and shell command.
- HADOOP-11509. change parsing sequence in GenericOptionsParser to parse -D parameters before -files.
- HADOOP-11510. Expose truncate API via FileContext.
- HADOOP-11523. StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- HADOOP-11579. Documentation for truncate.
- HADOOP-11595. Add default implementation for AbstractFileSystem#truncate.
- MAPREDUCE-6230. Fixed RMContainerAllocator to update the new AMRMToken service name properly.
- YARN-570. Time strings are formated in different timezone.
- YARN-2246. Made the proxy tracking URL always be http(s)://proxy addr:port/proxy/<appId> to avoid duplicate sections.
- YARN-2571. RM to support YARN registry
- YARN-2683. registry config options: document and move to core-default
- YARN-2837. Support TimeLine server to recover delegation token when restarting.
- YARN-2917. Fixed potential deadlock when system.exit is called in AsyncDispatcher
- YARN-2964. RM prematurely cancels tokens for jobs that submit jobs (oozie).
- YARN-3103. AMRMClientImpl does not update AMRM token properly.
- YARN-3207. Secondary filter matches entites which do not have the key being filtered for.
- YARN-3227. Timeline renew delegation token fails when RM user's TGT is expired.
- YARN-3239. WebAppProxy does not support a final tracking url which has query fragments and params.
- YARN-3251. Fixed a deadlock in CapacityScheduler when computing absoluteMaxAvailableCapacity in LeafQueue.
- YARN-3269. Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.

1.4.4. HBase 0.98.4

HDP 2.2.4 provides HBase 0.98.4 and the following addtional Apache patches:

- HBASE-212. Addendum. Fixes a unit test
- HBASE-10499 In write-heavy scenario one of the regions does not get flushed causing RegionTooBusyException
- HBASE-11569 Addendum for not skipping replayed edits for primary region replica
- HBASE-12238 A few exceptions on startup - PARTIAL BACKPORT
- HBASE-12533 staging directories are not deleted after secure bulk load
- HBASE-12575 Sanity check table coprocessor classes are loadable
- HBASE-12536. Reduce the effective scope of GLOBAL CREATE and ADMIN permission
- HBASE-12562 Handling memory pressure for secondary region replicas - ADDENDUM for fixing findbug reported issues
- HBASE-12791 HBase does not attempt to clean up an aborted split when the regionserver shutting down
- HBASE-12714 RegionReplicaReplicationEndpoint should not set the RPC Codec
- HBASE-12958 SSH doing hbase:meta get but hbase:meta not assigned
- HBASE-13120 Allow disabling hadoop classpath and native library lookup

1.4.5. Hive 0.14.0

HDP 2.2.4 provides Hive 0.14.0 and the following additional Apache patches:

- HIVE-480: HDFSCleanup thread holds reference to FileSystem
- HIVE-6468: HiveServer2 (tcp mode, with SASL layer) OOMs when getting a non-thrift message
- HIVE-6679: HiveServer2 should support TCP Keepalive & Server Socket Timeout on blocking calls
- HIVE-7175: Provide password file option to beeline
- HIVE-7270: Serde info is not shown in show create table statement, but shows in the desc table
- HIVE-8295: Add batch retrieve partition objects for metastore direct sql
- HIVE-8485: SUMMARY-Hive metastore NPE with Oracle DB when there is empty value for string for tblproperties/serdeproperties/etc, table not usable after creation

- HIVE-8762 : HiveMetaStore.BooleanPointer should be replaced with an AtomicBoolean
- HIVE-8791 : Hive permission inheritance throws exception S3
- HIVE-8850: SUMMARY-[ObjectStore:: rollbackTransaction() needs to be looked into further
- HIVE-8881: Receiving json "could not find job" error when web client tries to fetch all jobs from WebHCat but HDFS does not have the data.
- HIVE-8888: Hive on Tez query output duplicate rows when there is explode in subqueries for joins
- HIVE-8891: SUMMARY-[Another possible cause to NucleusObjectNotFoundException from drops/rollback
- HIVE-8893 : Implement whitelist for builtin UDFs to avoid untrused code execution in multiuser mode
- HIVE-8966 : Delta files created by hive hcatalog streaming cannot be compacted
- HIVE-9025 : join38.q (without map join) produces incorrect result when testing with multiple reducers
- HIVE-9038: Join tests fail on Tez
- HIVE-9055: Tez: union all followed by group by followed by another union all gives error
- HIVE-9106 : improve the performance of null scan optimizer when several table scans share a physical path
- HIVE-9112 : Query may generate different results depending on the number of reducers
- HIVE-9141: HiveOnTez: mix of union all, distinct, group by generates error
- HIVE-9155: HIVE_LOCKS uses int instead of bigint hive-txn-schema-0.14.0.mssql.sql
- HIVE-9205: Change default Tez install directory to use /tmp instead of /user and create the directory if it does not exist
- HIVE-9234: HiveServer2 leaks FileSystem objects in FileSystem.CACHE
- HIVE-9235: Turn off Parquet Vectorization until all data types work: DECIMAL, DATE, TIMESTAMP, CHAR, and VARCHAR
- HIVE-9249: java.lang.ClassCastException: org.apache.hadoop.hive.serde2.io.HiveVarcharWritable cannot be cast to org.apache.hadoop.hive.common.type.HiveVarchar when joining tables
- HIVE-9278: Cached expression feature broken in one case
- HIVE-9316: TestSqoop tests in WebHCat testsuite hardcode libdir path to hdfs
- HIVE-9351: Running Hive Jobs with Tez cause templeton to never report percent complete

- HIVE-9359: SUMMARY-[Hive export OOM error when table is huge. (32TB data, 4800+ partitions)]
- HIVE-9382 : Query got rerun with Global Limit optimization on and Fetch optimization off
- HIVE-9390: Enhance retry logic wrt DB access in TxnHandler
- HIVE-9401: SimpleFetchOptimizer for limited fetches without filters
- HIVE-9404 NPE in org.apache.hadoop.hive.metastore.txn.TxnHandler.determineDatabaseProduct()
- HIVE-9436: SUMMARY-[RetryingMetaStoreClient does not retry JDOExceptions]
- HIVE-9446: JDBC DatabaseMetadata.getColumns() does not work for temporary tables
- HIVE-9473 : sql std auth should disallow built-in udfs that allow any java methods to be called
- HIVE-9593. ORC reader should ignore new/unknown metadata streams.
- HIVE-9652: STDERR redirection should not use in place updates TEZ UI
- HIVE-9665: Parallel move task optimization causes race condition
- HIVE-9673: Set operationhandle in ATS entities for lookups
- HIVE-9683: Client TCP keep-alive for Thrift as a workaround for THRIFT-2788
- HIVE-9684: Incorrect disk range computation in ORC because of optional stream kind
- HIVE-9743: Incorrect result set for vectorized left outer join
- HIVE-9779: ATSHook does not log the end user if doAs=false (it logs the hs2 server user)
- HIVE-9836: Hive on Tez: fails when virtual columns are present in the join conditions (for e.g. partition columns)
- HIVE-9841: IOException thrown by ORC should include the path of processing file
- HIVE-9886: Tez NPE error when initialize reducer from 2 row_number function on each join side
- HIVE-9892: HIVE schema failed to upgrade from HDP 2.0/2.1 to HDP 2.2 with schematool
- HIVE-9832: Merge join followed by union and a map join in hive on tez fails.

1.4.6. Knox 0.5.0

HDP 2.2.4 provides Knox 0.5.0 and the following additional Apache patch:

- KNOX-492: Support service level replayBufferLimit for Ozzie, Hive and HBase.

1.4.7. Oozie 4.1.0

HDP 2.2.4 provides Apache Oozie 4.1.0 and the following additional Apache patch:

- OOZIE-208. Adding missing oozie property oozie.service.HadoopAccessorService.hadoop.configurations to oozie install script.

1.4.8. Phoenix 4.2.0

HDP 2.2.4 includes Apache Phoenix 4.2.0 and the following additional Apache patches:

- PHOENIX-18 Several tests are failing with SequenceNotFoundException error
- PHOENIX-32 Some phoenix tests are failing with different assertion errors
- PHOENIX-1248 CsvBulkLoadTool is failing with IAE when local index specified for –index-table parameter
- PHOENIX-1336 Exception when select from local index:Cache of region boundaries are out of date
- PHOENIX-1346 the choice of index by phoneix is not correct where do querying
- PHOENIX-1362 Min/max aggregate query on CHAR and BINARY types always return null
- PHOENIX-1427 Reduce work in StatsCollector
- PHOENIX-1428 Queries with limit against tenant views is broken
- PHOENIX-1429 Cancel queued threads when limit reached
- PHOENIX-1431 DELETE using Subqueries
- PHOENIX-1436 the choice of index by phoneix is not correct where do querying
- PHOENIX-1446 Add/fix logging for LIMIT optimization
- PHOENIX-1447 Increase guidepost width default to take into account FAST_DIFF compression
- PHOENIX-1448 Fix resource leak when work rejected by thread executor
- PHOENIX-1449 Fix PropertiesUtil.deepCopy()
- PHOENIX-1455 Replace org.xerial.snappy with org.iq80.snappy pure Java snappy implementation
- PHOENIX-1456 Incorrect query results caused by reusing buffers in SpoolingResultIterator
- PHOENIX-1462 Create unit test for COUNT DISTINCT using compression (Ram)
- PHOENIX-1463 phoenix.query.timeoutMs doesn't work as expected
- PHOENIX-1466 Prevent multiple scans when query run serially

- PHOENIX-1467 Upgrade to 4.12 Junit and update tests by removing @Category annotation
- PHOENIX-1468 Add org.iq80.snappy to client and server assemblies
- PHOENIX-1469 Binary columns do not work correctly for indexing
- PHOENIX-1472 Fix salt bucket for Integer.MIN_VALUE
- PHOENIX-1473 Connecting with Phoenix client when Phoenix is not deployed on region server(s) takes down region server(s)
- PHOENIX-1474 NPE when RVC between combined with key part comparison
- PHOENIX-1476 Immediately open scanner for parallel queries
- PHOENIX-1477: Phoenix 4.0 -> 4.2 upgrade doesn't work
- PHOENIX-1480 Incorrect query results may occur when VIEW uses indexes from physical table
- PHOENIX-1484 Index creation failed due to specifying DEFAULT_COLUMN_FAMILY option
- PHOENIX-1485 Correct arg count in FunctionParseNode
- PHOENIX-1485 Add timezone awareness
- PHOENIX-1486 Fix SpooledTmpFileDeleteIT by having it use a different directory for spooling files
- PHOENIX-1498 Turn KEEP_DELETED_CELLS off by default
- PHOENIX-1500 Disallow mutations and queries using using PhoenixPreparedStatement.executeQuery() and executeUpdate() respectively
- PHOENIX-1503 - Compilation on Mac OS fails on maven-assembly-plugin
- PHOENIX-1532 Phoenix tarball assembly does not include server jar
- PHOENIX-1533 Last key part not taken into child/parent optimization
- PHOENIX-1537 Set reuseForks to false for integration tests
- PHOENIX-1551 Upgrading from Phoenix 4.0.x to 4.2.2 and throw can't find SYSTEM.STATS
- PHOENIX-1571 Replace hard-coded date
- PHOENIX-1575 Identical sql type values for UNSIGNED_TIMESTAMP and UNSIGNED_DATE
- PHOENIX-1592 ORDER BY not closing ResultIterator properly
- PHOENIX-1596 Turning tracing on causes region servers to crash

- PHOENIX-1601 - Performance.py script not working in binary distribution

1.4.9. Pig 0.14.0

HDP 2.2.4 provides Apache Pig 0.14.0 and the following additional Apache patches:

- PIG-156: Pig command fails because the input line is too long
- PIG-4334: PigProcessor does not set pig.datetime.default.tz
- PIG-4342: Pig 0.14 cannot identify the uppercase of DECLARE and DEFAULT
- PIG-4377: Skewed outer join produce wrong result in some cases (PIG-4377-2.patch)
- PIG-4377: Skewed outer join produce wrong result in some cases

1.4.10. Ranger 0.4.0

HDP 2.2.4 provides Ranger 0.4.0 and the following additional Apache patch:

- RANGER-188: Added LSB headers to Ranger Admin/Usersync init.d scripts

1.4.11. Slider 0.61.0

HDP 2.2.4 provides Slider 0.61.0 and the following additional Apache patch:

- SLIDER-769: Modify Slider app packages for seamless integration with AMS provider (Ambari Metrics Service)

1.4.12. Storm 0.9.3

HDP 2.2.4 provides Apache Storm 0.9.3 and the following additional patches:

- STORM-586: TridentKafkaEmitter should catch updateOffsetException.
- STORM-682: supervisor should handle worker state corruption gracefully.

1.4.13. Tez 0.5.2

HDP 2.2.4 provides Apache Tez 0.5.2 and the following Apache patches:

- TEZ-1642. TestAMRecovery sometimes fails.
- TEZ-1775. Allow setting log level per logger.
- TEZ-1800. Integer overflow in ExternalSorter.getInitialMemoryRequirement()
- TEZ-1836. Provide better error messages when tez.runtime.io.sort.mb, spill percentage is incorrectly configured.
- TEZ-1851. FileSystem counters do not differentiate between different FileSystems

- TEZ-1852. Get examples to work in Local Mode.
- TEZ-1861. Fix failing test: TestOnFileSortedOutput.
- TEZ-1878. Task-specific log level override not working in certain conditions
- TEZ-1924. Tez AM does not register with AM with full FQDN causing jobs to fail in some environments.
- TEZ-1931. Publish tez version info to Timeline.
- TEZ-1934. TestAMRecovery may fail due to the execution order is not determined.
- TEZ-1942. Number of tasks show in Tez UI with auto-reduce parallelism is misleading.
- TEZ-1949: Whitelist TEZ_RUNTIME_OPTIMIZE_SHARED_FETCH for broadcast edges
- TEZ-1962. Fix a thread leak in LocalMode.
- TEZ-2024. Compiliation error due to conflict.
- TEZ-2024. TaskFinishedEvent may not be logged in recovery.
- TEZ-2037. Should log TaskAttemptFinishedEvent if taskattempt is recovered to KILLED
- TEZ-2135. ACL checks handled incorrectly in AMWebController.

As part of HPD 2.2.4 Hortonworks is probiving a Tez Debugging User Interface. This interface does not impact the behavior or function of jobs that leverage Tez, and its use is optional. Patches added to facilitate the Tez Debugging User Interface include:

- TEZ-1990. Tez UI: DAG details page shows Nan for end time when a DAG is running.
- TEZ-2031. Tez UI: horizontal scrollbars do not appear in tables, causing them to look truncated.
- TEZ-2038. TEZ-UI DAG is always running in tez-ui when the app is failed but no DAGFinishedEvent is logged.
- TEZ-2043. Tez UI: add progress info from am webservice to dag and vertex views.
- TEZ-2052. Tez UI: log view fixes, show version from build, better handling of ats url config.
- TEZ-2056. Tez UI: fix VertexID filter,show only tez configs by default,fix appattemptid.
- TEZ-2063. Tez UI: Flaky log url in tasks table.
- TEZ-2065. Setting up tez.tez-ui.history-url.base with a trailing slash can result in failures to redirect correctly.
- TEZ-2068. Tez UI: Dag view should use full window height, disable webuiservice in localmode.
- TEZ-2069. Tez UI: appId should link to application in dag details view.

- TEZ-2077. Tez UI: No diagnostics on Task Attempt Details page if task attempt failed
- TEZ-2078. Tez UI: Task logs url use in-progress url causing various errors.
- TEZ-2079. Tez UI: trailing slash in timelineBaseUrl in ui should be handled.
- TEZ-2092. Tez UI history url handler injects spurious trailing slash.
- TEZ-2098. Tez UI: Dag details should be the default page for dag, fix invalid time entries for failed Vertices.
- TEZ-2101. Tez UI: Issues on displaying a table.
- TEZ-2102. Tez UI: DAG view has hidden edges, dragging DAG by holding vertex causes unintended click.
- TEZ-2106. TEZ UI: Display data load time, and add a refresh button for items that can be refreshed.
- TEZ-2112. Tez UI: fix offset calculation, add home button to breadcrumbs.
- TEZ-2114. Tez UI: task/task attempt status is not available when its running.
- TEZ-2116. Tez UI: dags page filter does not work if more than one filter is specified.
- TEZ-2134. TEZ UI: On request failure, display request URL and server name in error bar.
- TEZ-2136. Some enhancements to the new Tez UI.
- TEZ-2142. TEZ UI: Breadcrumb border color looks out of place in wrapped mode.
- TEZ-2158. TEZ UI: Display dag/vertex names, and task/attempt index in breadcrumb
- TEZ-2160. Tez UI: App tracking URL should support navigation back.
- TEZ-2165. Tez UI: DAG shows running status if killed by RM in some cases.

1.4.14. ZooKeeper 3.4.6

HDP 2.2.4 provides ZooKeeper 3.4.6 and the following additional Apache patch:

- ZOOKEEPER-1506: Re-try DNS hostname -> IP resolution if node connection fails

1.5. Fixed in This Release

The following features and fixes were contributed back to Apache with the release of HDP 2.2.4 for Windows:

Key	Apache JIRA	Summary
BUG-31413	n/a	HBASE-212 needs to be backported to HDP 2.2
BUG-30180	n/a	HiveServer2 loses ephemeral node on a long GC pause or SIGHUP

Key	Apache JIRA	Summary
BUG-29423	HADOOP-799	Hotfix for MAPREDUCE-32
BUG-28620	HADOOP-9629	Synchronize WASB code from upstream Apache
BUG-28115	HADOOP-11321	CopyToLocal cannot save a file to an SMB share unless the user has Full Control permissions.
BUG-26072	n/a	Receiving error "ERROR [main]: bonecp.ConnectionHandle (ConnectionHandle.java:markPossiblyBroken(388)) - Database access problem. Killing off this connection and all remaining connections in the connection pool. SQL State = 08S01 "
BUG-23339	n/a	Sqoop jobs cannot run on Windows Secure Cluster because of issues with the classpath manifests

1.6. Known Issues for This Release

The following HDP areas have known issues for HDP 2.2:

- HDP for Windows Installer
- Falcon
- HBase
- Hive
- HDFS
- Knox
- Oozie
- Phoenix
- Pig
- Ranger
- Sqoop
- YARN
- ZooKeeper

1.6.1. HDP for Windows Installer

- **BUG-14957**: Installer fails to install HDP on systems that have hostnames longer than 15 characters

  Problem: The installer truncates hostnames that are longer than 15 characters during the installation process and then fails system check.

  Workaround: From the HDP Setup Window, click continue to install even though the hostname lookup fails. For silent installs, add the truncated hostname and IP address to the hosts file of the host as follows:

```
IP truncatedhost-1 hostname // A line with the ip of the machine
and full and truncated hostname…
```

1.6.2. Falcon

- **BUG-26944:** E0803: IO error, null during scheduling feed on a cluster with wire encryption on.

  Problem: Falcon has difficulty writing to target clusters with wire encryption on. The ssl-client on the CP must be valid for doing SSL negotiation with the target cluster.

  Workaround: Have Falcon write to target clusters with wire encryption disabled.

- **BUG-25179:** Falcon's default port is 15443 if TLS is enabled, 15000 otherwise.

  Problem: Falcon server is on port 15443 or 15000 based on setting for TLS. In Falcon startup.properties:

  If falcon.enableTLS is set to true explicitly or not set at all, falcon starts at port 15443 on https:// by default. If falcon.enableTLS is set to false, falcon starts at port 15000 on http://.

  To change the port, use -port option. If falcon.enableTLS is not set explicitly but port is set explicitly, port that ends with 443 will automatically put falcon on https://. Any other port will put falcon on http://.

- **BUG-25179:** You must add your own keystore to `/etc/falcon/conf/` directory as prism.keystore, when falcon.enableTLS is set to true.

1.6.3. HBase

- **BUG-26592:** Multiple tests are failing because of System time goes back frequently on SLES

  Problem: SUSE 11 SP3 clusters running on Amazon EC2 have had the system clock frequently going back (hundreds of ms), which causes multiple issues in HBase. The HBase data model assumes that timestamps are monotonically increasing, and failing that may result in updates being eclipsed by previous mutations.

  Workaround: Set up NTP and ensure that system clock time does not drift when running HBase. For more details, see https://hbase.apache.org/book.html#basic.prerequisites.

- **BUG-23981 [Windows]:** IOException: Invalid HFile block magic in Windows unit tests

  If you use HBase on Windows on a local filesystem (not HDFS) for single-node non-distributed clusters, concurrent hfile readers might return data belonging to a different offset than what was requested, resulting in read failures.

1.6.4. Hive

- **BUG-28161, BUG-28234:** Hive Server2 Long running test failed after 44 hours with OutOfMemory

Problem: Hive Server 2 out of memory due to FileSystem.CACHE leak

Workaround: To prevent memory leaks in unsecure mode, disable file system caches by setting the following parameters to true in hive configuration (hive-site.xml).

`fs.hdfs.impl.disable.cache` – Disable HDFS filesystem cache, default false.
`fs.file.impl.disable.cache` – Disable local filesystem cache, default false.

- **BUG-28102: Update to headers needed for** `/etc/init.d/ranger-admin` and `/etc/init.d/ranger-usersyncscripts` after Ranger install on Debian 6.

  Problem: The init scripts installed by Ranger for admin and usersync are missing a comment block at the beginning (see https://wiki.debian.org/LSBInitScripts/), resulting in error messages.

  Workaround: Add the following header to these scripts:

  >### BEGIN INIT INFO # Provides: scriptname # Required-Start: $remote_fs $syslog# Required-Stop: $remote_fs $syslog # Default-Start: 2 3 4 5 # Default-Stop: 0 1 6 # Short-Description: Start daemon at boot time # Description: Enable service provided by daemon. ### END INIT INFO

- **BUG-27913** Error while upgrading Hive Metastore schema for Oracle

  Problem: The pre-upgrade script causes an error when upgrading Hive Metastore using an Oracle DB. For example, after executing the command -upgradeSchema, an error message similar to the following displays:

```
               root@ip-172-31-42-3 ~# /usr/hdp/2.2.0.0-2036/hive/bin/
schematool -upgradeSchema -dbType oracle
               14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
optimize.mapjoin.mapreduce does not exist
               14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
heapsize does not exist
               14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
server2.enable.impersonation does not exist
               14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
semantic.analyzer.factory.impl does not exist
               14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
auto.convert.sortmerge.join.noconditionaltask does not exist
               Metastore connection URL: jdbc:oracle:thin:@//ip-172-31-42-1.
ec2.internal:1521/XE
               Metastore Connection Driver : oracle.jdbc.driver.
OracleDriver
               Metastore connection User: hiveuser Starting upgrade
 metastore schema from version 0.13.0 to 0.14.0
               Upgrade script upgrade-0.13.0-to-0.14.0.oracle.sql Error:
 ORA-00955: name is already used by an existing
               object (state=42000,code=955)
               Warning in pre-upgrade script pre-0-upgrade-0.13.0-to-0.14.0.
oracle.sql: Schema script failed, errorcode 2
               Completed upgrade-0.13.0-to-0.14.0.oracle.sql schemaTool
 completed
```

  Workaround: Ignore the error message. The schema tool has upgraded.

- **BUG-27637:** Postgres Metastore has intermittent test failure due to "ERROR: could not serialize access due to read/write dependencies among transactions"

  Problem: Sometimes the acid_concurrency test on Postgres Metastore test fails with a long error, beginning with the lines:

  ```
  ERROR: could not serialize access due to read/write dependencies
  among transactions Detail: Reason code: Canceled on
  identification as a pivot, during write. Hint: The transaction
  might succeed if retried.
  ```

- **BUG-27636** Oracle Metastore: intemittent acid_concurrency test failures due to NoSuchLockException

- **BUG-27197:** MatchPath not available in HDP 2.2.

  Intellectual property issues prevent Hortonworks from shipping MatchPath in Hive 0.14.x.

- **BUG-26418:** SSLv3 not supported in Hive 0.14.x

  Problem: SSLv3 is vulnerable to the Poodle Attack and is therefore disabled in Hive 0.14.x. Do not enable SSLv3 in Hive.

1.6.5. HDFS

- **BUG-26890:** Commands in HA mode behave differently compared to non-HA

  Problem: Commands in HA mode behave differently compared to non-HA.

  For instance,in HDA mode, executing `hdfs dfsadmin -safemode enter` will run without error if nn2 is active and nn1 is standby, but it will set standby to safemode and not active. If nn1 is actually off, the command will try nn2 and the command in effect will do what it needs to do.

  The same behavior is true for `hdfs dfsadmin -saveNamespace`

  Workaround: You can add the `fs` and specify `hdfs://hostname:8020` of the active namenode for instance `hdfs dfsadmin -fs hdfs://<namenodehostactive>:8020 -saveNamespace`

1.6.6. Knox

- **BUG-24945:** [KNOX-443] Support for active HS2 clients (through Knox) during rolling upgrades will be addressed in a later release.

1.6.7. Oozie

- **BUG-27982** Oozie client package installation from public HDP repo fails because of hdp-select

  Problem: Oozie-client is missing a dependency on hdp-select. If you install oozie-client prior to any other stack applications, installation fails with with an error message similar to the following:

```
Installing: oozie_2_2_0_0_1947-client-4.1.0.2.2.0.0-1947
                [error] Installation of oozie_2_2_0_0_1947-client-4.1.0.2.2.
0.0-1947 failed: (with --nodeps --force)
                Error: Subprocess failed.
                Error: RPM failed: warning: group jenkins does not exist -
 using root: group jenkins does not exist - using root
                .
                .
                .
                /var/tmp/rpm-tmp.48695: line 6: /usr/bin/hdp-select: No such
 file or directory
                error: %post(oozie_2_2_0_0_1947-client-4.1.0.2.2.0.0-1947.
noarch) scriptlet failed, exit status 127
                Abort, retry, ignore? [a/r/i] (a):
```

oozie-client didn't had a requirement on hep-select and that is causing the issue.

1.6.8. Phoenix

- **BUG-26661:** Windows runs failed with "region isn't co-located with its data" exception

  Problem: The Local Index feature is considered experimental and should not be enabled at this time except for preview purposes.

- **BUG-28226**: Error 504 after upgrading from HDP 2.1 to HDP 2.2

  Problem: After upgrading from HDP 2.1 to HDP 2.2, queries such as the following generate an ERROR 504:

  ```
  ./psql.py ip-172-31-43-221.ec2.internal:2181 /grid/0/hdp/2.2.0.0-2041/
  phoenix/doc/examples/WEB_STAT.sql /grid/0/hdp/2.2.0.0-2041/phoenix/doc/
  examples/WEB_STAT.csv /grid/0/hdp/2.2.0.0-2041/phoenix/doc/examples/
  WEB_STAT_QUERIES.sql
  ```

  Error detail:

  ```
  org.apache.phoenix.schema.ColumnNotFoundException: ERROR 504 (42703):
   Undefined column. columnName=INDEX_TYPE
  ```

  Workaround: Follow instructions in the "Manual Upgrade to HDP 2.2" document, "Configure Phoenix" section, step 4: replace the client jar file before connecting the new Phoenix client to the Phoenix cluster.

  (See also: Apache PHOENIX-1477)

1.6.9. Pig

- Nested fields will experience slower performance because they are not currently supported by predicate pushdown in Pig.

1.6.10. Ranger (formerly XA Secure)

- **BUG-28102: Update to headers needed for** `/etc/init.d/ranger-admin` and `/etc/init.d/ranger-usersyncscripts` after Ranger install on Debian 6.

Problem: The init scripts installed by Ranger for admin and usersync are missing a comment block at the beginning (details at https://wiki.debian.org/LSBInitScripts/), resulting in error messages.

Workaround: Add the following header to these scripts:

```
### BEGIN INIT INFO # Provides: scriptname # Required-Start:
$remote_fs $syslog # Required-Stop: $remote_fs $syslog # Default-
Start: 2 3 4 5 # Default-Stop: 0 1 6 # Short-Description: Start
daemon at boot time # Description: Enable service provided by
daemon. ### END INIT INFO
```

- **BUG-23867:** Documentation Update Ranger Advanced Security document for Hive configuration:

Problem: Documentation for Ranger will be updated shortly. The following information corrects Apache Ranger documentation at Section 4.2.3.1.

Where the documentation states:

Modify the Ambari Hive Startup Script OLD
==========

3. Comment out the following line by prepending a # at the beginning of the line as follows:

```
                HIVE_SERVER2_OPTS="$ {HIVE_SERVER2_OPTS} –hiveconf hive.
                security.authenticator.manager=org.apache.hadoop.hive.ql.
security.
                SessionStateUserAuthenticator -hiveconf hive.security.
authorization.
                manager=org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.
                SQLStdHiveAuthorizerFactory"
```

==========

The correction is:

==========

3. Comment out the following line by prepending a # at the beginning of the line as follows:

```
              # HIVE_SERVER2_OPTS="${HIVE_SERVER2_OPTS} –hiveconf hive.
              security.authenticator.manager=org.apache.hadoop.hive.ql.
security.
              SessionStateUserAuthenticator -hiveconf hive.security.
authorization.
              manager=org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.
              SQLStdHiveAuthorizerFactory”
```

... and add the following line (instead):

```
              HIVE_SERVER2_OPTS="$ {HIVE_SERVER2_OPTS} –hiveconf hive.
              security.authenticator.manager=org.apache.hadoop.hive.ql.
security.
              SessionStateUserAuthenticator”

============================================================
```

1.6.11. Sqoop

- **BUG-25282:** j.l.ArrayIndexOutOfBoundsException using –map-column-hive

  Problem: Running this command with the –map-column-hive statement:

  ```
  sqoop import --verbose --connect 'jdbc:mysql://sandbox.hortonworks.com/demo'
   --table customer_address --username carter --password password --create-
  hcatalog-table --hcatalog-table customer_address --hcatalog-storage-stanza
   "stored as orc" --map-column-hive fdsa -m 1
  ```

  returns an java.lang.ArrayIndexOutOfBoundsException.

  The command works without the –map-column-hive statement.

  Workaround: Sqoop expects a comma-separated list of mappings in the form

  `<name of column>=<new type>. For example: $ sqoop import ... --map-column-java id=String,value=Integer`

1.6.12. YARN

- **BUG-27604:** After upgrade from 2.1 of Highly Available system, ResourceManager fails to elect a master.

  Problem: After upgrading an HA system from HDP 2.1 to HDP 2.2, you must manually clear the ResourceManager state store as part of the upgrade or an "Exception transitioning to active PERMISSIONS" error will occur.

  Workaround: Manually clear the ResourceManager state store:

  - Run:

    ```
    yarn resourcemanager -format-state-store
    ```

  - Start the ResourceManager:

    ```
    su -l yarn -c "/usr/hdp/current/hadoop-yarn-resourcemanager/
    sbin/yarn-daemon.sh start resourcemanager"
    ```

1.6.13. ZooKeeper

- **BUG-19351:** Zookeeper does not have shell access on Ubuntu and Debian Hosts.

  Description: The `/etc/passwd` file has no shell access for Zookeeper:

```
zookeeper:x:108:115:ZooKeeper User,,,:/var/lib/zookeeper:/bin/
false
```

Workaround: Modify the zookeeper user to have shell access. For example:

```
usermod zookeeper -s /bin/bash
```

1.7. About Hortonworks Data Platform

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.**

For more information on Hortonworks technology, Please visit the Hortonworks Data Platform page. 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.