Hortonworks Data Platform 2.1.5

Release Notes

(Aug 29, 2014)

Hortonworks Data Platform 2.1.5 for Windows : Release Notes

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

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

Table of Contents

List of Tables

1. Release Notes HDP-2.1.5

The HDP 2.1 Release Notes include the following sections:

1.1. Product Version: HDP-2.1.5

All HDP 2.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.

Note

The minimum level of Apache Ambari to use with HDP 2.1, is version 1.5.1.

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

Apache Hadoop 2.4.0	Apache Falcon 0.5.0
Apache HBase 0.98.0	Apache Sqoop 1.4.4
Apache Pig 0.12.1	Apache Knox 0.4.0
Apache Hive 0.13.1	Apache Flume 1.4.0
Apache Tez 0.4.0	Apache Accumulo 1.5.1
Apache ZooKeeper 3.4.5	Apache Phoenix 4.0.0
Storm 0.9.1	Apache Mahout 0.9.0
Apache Oozie 4.0.0	

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

1.2. Unsupported Apache components

The following Apache component is shipped with HDP 2.1, but not supported:

- Avro v 1.7.4

The following Apache components are shipped as part of HDP 2.1 HDFS, but not supported:

- NameNode Federation (HDFS-1052)
- viewFS (HADOOP-7257)

The following Apache components are shipped as part of HDP 2.1 YARN, but not supported:

- MapReduce v1 Fair Scheduler (HADOOP-3746)
- YARN Fair Scheduler (MAPREDUCE-3451)
- MapReduce Uber AM (MAPREDUCE-2405)
- MapReduce Eclipse Plugin (for Non-Kerberos and Kerberos cluster)
- Cgroup resource isolation (YARN-3)
- CPU Scheduling (YARN-2)

Note

Hue is not supported on Windows.

1.3. Tech Previews in This Release

The following features are provided with HDP 2.1.5 as technical previews; that is, they are considered to be still under development, and as such not supported. Do not use the following features in your production systems.

- Preemption – this feature has been found to contain defects that make it unsuitable for a production cluster at this time. For information about Preemption, see http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_system-admin-guide/content/preemption.html.
- Falcon – lag time in displaying lineage information if graph contains many vertices; old graph persists until updated graph is loaded

- Falcon – dependencies, viewed in Graph View

If you are an official Hortonworks Customer or Partner, contact Hortonworks Technical Support, by logging a case at https://support.hortonworks.com/. If you are not currently an official Hortonworks Customer or Partner, please seek assistance at our Hortonworks Forums at http://hortonworks.com/community/forums/.

1.4. Fixed in This Release

This release contains the following bug fixes:

Project	Apache JIRA	Bug No.	Description
Hadoop Common	HADOOP-10342	BUG-19415	Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
Hadoop Common	HADOOP-10710	BUG-20468	hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster
WebHCat	HADOOP-10839, -10840, -7664	BUG-20182	Need hotfix for HADOOP-879 and HADOOP-880
HDFS	HADOOP-874	BUG-21474	UT Failures TestNativeAzureFileSystemMocked
HBase	HADOOP-11118	BUG-19935	non environment variable solution for ZeroCopyLiteralByteString
HDFS	HDFS-6527, HDFS-6618, HDFS-6622, HDFS-6647	BUG-19654	Fix potential editlog corruption
HDFS	HDFS-6604	BUG-20530	Major dependency change in hive causes wrong hadoop configurations to be loaded
HDFS	HDFS-6604	BUG-19944	The short-circuit cache doesn't correctly time out replicas that haven't been used in a while
Hive, Pig, WebHCat	HIVE-6835	BUG-19469	Pig returns only partitioned value when loading Avro partitioned table with HCatLoader()
Hive	HIVE-7062	BUG-17846	Hive query using SUM() windowing function fails to complete and stays stuck on REDUCE task
Hive, Tez	HIVE-7112	BUG-16162	Investigate failure in Hive
WebHCat	HIVE-7155	BUG-21065	Need fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
Pig	PIG-4044, -4045	BUG-19608	Pig job processing in Avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar
HDFS	HDFS-6340	BUG-20095	Can't finalize upgrade
HDFS	HDFS-6616	BUG-20165	bestNode shouldn't always return the first DataNode
Hive	HIVE-6979	BUG-20189	Back-port Java code of HIVE-6979
MapReduce	MAPREDUCE-4951 and MAPREDUCE-5900	BUG-18039	Container preemption interpreted as task failures and eventually job failures
MapReduce	MAPREDUCE-5956	BUG-19735	MapReduce AM should not use maxAttempts to determine if this is the last retry
MapReduce	MAPREDUCE-6002	BUG-20419	MR task should prevent report error to AM when process is shutting down

Project	Apache JIRA	Bug No.	Description
Hive	HIVE-7366, -7368	BUG-19937	Data nucleus returns empty results (table not found) when there is unexpected underlying exceptions
Pig	PIG-3744	BUG-21208	SequenceFileLoader does not support BytesWritable
YARN	YARN-2144	BUG-18468	Add logs when preemption occurs.
YARN	YARN-2181	BUG-19182	Add preemption info to RM Web UI and RM logs.
YARN	YARN-2125	BUG-18536	ProportionalCapacityPreemptionPolicy should only log CSV when debug enabled
YARN	YARN-2124	BUG-18535	ProportionalCapacityPreemptionPolicy cannot work because it's initialized before scheduler initialized
YARN	YARN-2022	BUG-19086	Preempting an Application Master container can be kept as least priority when multiple applications are marked for preemption by ProportionalCapacityPreemptionPolicy
YARN	YARN-2074	BUG-17995	Preemption of AM containers shouldn't count towards AM failures
YARN	YARN-1957	BUG-17996	ProportionalCapacitPreemptionPolicy handling of corner cases
YARN	YARN-1408	BUG-18120	Preemption caused Invalid State Event: ACQUIRED at KILLED and caused a task timeout for 30mins

1.5. Patch Information

In this section:

- Patch Information for Hadoop Common/HDFS
- Patch Information for ZooKeeper
- Patch Information for HBase
- Patch Information for Pig
- Patch Information for Tez
- Patch Information for Hive/HCat
- Patch Information for Oozie
- Patch Information for Falcon

Note

Apache Knox requires no additional patches.

1.5.1. Patch information for Hadoop Common/HDFS

Hadoop is based on Apache Hadoop 2.4.0 and includes the following additional patches:

- HADOOP-874: UT Failures TestNativeAzureFileSystemMocked

- HADOOP-10342: Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
- HADOOP-10475: ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken()
- HADOOP-10508: RefreshCallQueue fails when authorization is enabled
- HADOOP-10562: Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- HADOOP-10612: NFS failed to refresh the user group id mapping table
- HADOOP-10630: Possible race condition in RetryInvocationHandler
- HADOOP-10630: Possible race condition in RetryInvocationHandler
- HADOOP-10710: hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster.
- HDFS-4052: BlockManager#invalidateWork should print logs outside the lock
- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- HDFS-5257: addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- HDFS-6160: TestSafeMode occasionally fails.
- HDFS-6227: ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- HDFS-6233: Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- HDFS-6245: datanode fails to start with a bad disk even when failed volumes is set
- HDFS-6273: Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- HDFS-6278: Create HTML5-based UI for SNN
- HDFS-6279: Create new index page for JN / DN
- HDFS-6362: InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- HDFS-6364: Incorrect check for unknown datanode in Balancer
- HDFS-6370: Web UI fails to display in intranet under IE
- HDFS-6411: nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- HDFS-6423: Diskspace quota usage should be updated when appending data to partial block

- HDFS-6432: Add snapshot related APIs to webhdfs
- HDFS-6438: DeleteSnapshot should be a DELETE request in WebHdfs
- HDFS-6458: NFS: stale NFS file handle Error for previous mount point
- HDFS-6462: NFS: fsstat request fails with the secure hdfs
- HDFS-6527, HDFS-6618, HDFS-6622, and HDFS-6647: Fix potential editlog corruption
- HDFS-6599: Land HDFS-6599 into 2.1-maint
- HDFS-6616: bestNode shouldn't always return the first DataNode
- HIVE-5775: Introduce Cost Based Optimizer to Hive
- HIVE-6979: Back-port java code only of HIVE-6979
- HIVE-6799: HiveServer2 needs to map kerberos name to local name before proxy check
- HIVE-6915: Hive HBase queries fail on secure Tez cluster
- MAPREDUCE-5014: Extending DistCp through a custom CopyListing is not possible
- MAPREDUCE-6044: [onprem]: hadoopqa user does not have write permissions to /mapred/history causing jobs to fail
- YARN-1994: Support multi-homing for YARN/MR service endpoints

1.5.2. Patch information for ZooKeeper

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

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

1.5.3. Patch information for HBase

HBase is based on Apache HBase 0.98.0. It includes the following patches:

- HBASE-8304: Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- HBASE-9721: RegionServer should not accept regionOpen RPC intended for another(previous) server
- HBASE-10419: Add multiget support to PerformanceEvaluation
- HBASE-10486: ProtobufUtil Append and Increment deserialization lost cell level timestamp
- HBASE-10500: Some tools OOM when BucketCache is enabled
- HBASE-10514: Forward port HBASE-10466, possible data loss when failed flushes.

- HBASE-10548: Correct commons-math dependency version
- HBASE-10581: ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- HBASE-10582: 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- HBASE-10591: Sanity check table configuration in createTable
- HBASE-10592: Refactor PerformanceEvaluatiotool
- HBASE-10618: User should not be allowed to disable/drop visibility labels table
- HBASE-10621: Unable to grant user permission to namespace
- HBASE-10632: Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- HBASE-10635: thrift#TestThriftServer fails due to TTL validity check
- HBASE-10638: Improve error message when there is no region server available for move
- HBASE-10660: MR over snapshots can OOM when alternative blockcache is enabled
- HBASE-10670: HBaseFsck#connect() should use new connection
- HBASE-10688: Add a draining_node script to manage nodes in draining mode
- HBASE-10700: IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- HBASE-10751: TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- HBASE-10767: Load balancer may interfere with tests in TestHBaseFsck
- HBASE-10793: AuthFailed as a valid zookeeper state
- HBASE-10809: HBaseAdmin#deleteTable fails when META region happen to move around same time
- HBASE-10829: Flush is skipped after log replay if the last recovered edits file is skipped
- HBASE-10833: Region assignment may fail during cluster start up
- HBASE-10844: Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- HBASE-10850: essential column family optimization is broken
- HBASE-10852: TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment

- HBASE-10895: unassign a region fails due to the hosting region server is in FailedServerList
- HBASE-11118: non environment variable solution for ZeroCopyLiteralByteString

Windows Fixes:

- HBASE-10685 [WINDOWS] TestKeyStoreKeyProvider fails on windows

  HBASE-10686 [WINDOWS] TestStripeStoreFileManager fails on windows
- HBASE-10735 [WINDOWS] Set -XX:MaxPermSize for unit tests
- HBASE-10799 [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

- HBASE-10875 Metas own location should be cached
- HBASE-10791 Add integration test to demonstrate performance improvement
- HBASE-10810 LoadTestTool should share the connection and connection pool
- HBASE-10794 multi-get should handle missing replica location from cache
- HBASE-10634 Multiget doesn't fully work.
- HBASE-10661 TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- HBASE-10701 Cache invalidation improvements from client side
- HBASE-10778 Unique keys accounting in MultiThreadedReader is incorrect
- HBASE-10743 Replica map update is problematic in RegionStates
- HBASE-10616 Integration test for multi-get calls
- HBASE-10734 Fix RegionStates.getRegionAssignments to not add duplicate regions
- HBASE-10729 Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- HBASE-10726 Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- HBASE-10720 rpcClient: Wrong log level when closing the connection
- HBASE-10704 BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- HBASE-10633 StoreFileRefresherChore throws ConcurrentModificationException sometimes
- HBASE-10572 Create an IntegrationTest for region replicas.

- HBASE-10703 TestAsyncProcess does not pass on HBASE-10070
- HBASE-10637 rpcClient: Setup the iostreams when writing
- HBASE-10620 LoadBalancer.needsBalance() should check for co-located region replicas as well
- HBASE-10672 Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- HBASE-10630 NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- HBASE-10356 Failover RPC's for multi-get.
- HBASE-10525 Allow the client to use a different thread for writing to ease interrupt.
- HBASE-10355 Failover RPC's from client using region replicas.
- HBASE-10352 Region and RegionServer changes for opening region replicas, and refreshing store files
- HBASE-10351 LoadBalancer changes for supporting region replicas
- HBASE-10359 Master/RS WebUI changes for region replicas.
- HBASE-10362 HBCK changes for supporting region replicas.
- HBASE-10361 Enable/AlterTable support for region replicas.
- HBASE-10350 Master/AM/RegionStates changes to create and assign region replicas.
- HBASE-10490 Simplify RpcClient code (Nicolas Liochon)
- HBASE-10511 Add latency percentiles on PerformanceEvaluation
- HBASE-10517 NPE in MetaCache.clearCache()
- HBASE-10479 HConnection interface is public but is used internally, and contains a bunch of methods
- HBASE-10348 HTableDescriptor changes for region replicas
- HBASE-10354 Add an API for defining consistency per request
- HBASE-10347 HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- HBASE-10277 refactor AsyncProcess
- HBASE-10427 clean up HRegionLocation/ServerName usage
- HBASE-10472 Manage the interruption in ZKUtil#getData
- HBASE-10859 HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().

- HBASE-10858 TestRegionRebalancing is failing

1.5.4. Patch information for Pig

Pig is based on Apache Pig 0.12.1. It includes the following patches:

- PIG-4044, PIG-4045: Pig job processing avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar
- PIG-3916: isEmpty should not be early terminating.
- PIG-3744: SequenceFileLoader does not support BytesWritable
- PIG-3650: Fix for PIG-3100 breaks column pruning
- PIG-3573: Provide StoreFunc and LoadFunc for Accumulo.
- PIG-3558: ORC support for Pig.
- PIG-3257: Add a UUID function to Pig.

1.5.5. Patch information for Tez

Tez is based on Apache Tez 0.4.0 incubating release. It includes the following patches:

- TEZ-1159: Fix handling of corrupt or empty files in recovery data.
- TEZ-1158: Disable multiple AM attempts if recovery is disabled.
- TEZ-1135:: Fix ShuffledUnorderedKVInput handling of empty partitions.
- TEZ-1125: Pre-warm broken.
- TEZ-1097: Tez assumes that the scratch directory has to be same as the default filesystem.
- TEZ-1066: Generate events to integrate with YARN timeline server.
- TEZ-1048: Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- TEZ-1045: TezMiniCluster tests can fail intermittently.
- TEZ-1040: Fix a bug which could cause the Merger to hang.
- TEZ-1034: Shuffling can sometimes hang with duplicate inputs for the same index.
- TEZ-1033: AM hangs during recovery with Tasks awaiting init event.
- TEZ-1030: Address intermittent errors created due to race condition in YARN-1915.
- TEZ-1028: Handle killed tasks and attempts when handling recovery data.
- TEZ-1029: Fetcher can fail to report input failed event upon connection error.
- TEZ-1021: TezClient cannot connect to AM in a secure cluster when launched via Oozie.

- TEZ-1020: VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- TEZ-1015: Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- TEZ-1014: Add a log message to indicate last AM attempt.
- TEZ-1004: AM relocalization doesn't handle conflicting resources correctly.
- TEZ-1005: AM relocalization adds resources to the wrong classloader.
- TEZ-1011: TestDAGRecovery timing out on jenkins builds.
- TEZ-1010: TestAMNodeMap.testSelfBlacklist fails intermittently
- TEZ-997: Internal Errror in am logs during dag shutdown.
- TEZ-1009: Fixes in log file roll-over
- TEZ-998: InvalidStateTransitonException: Invalid event: V_INIT at INITED.

1.5.6. Patch information for Hive/HCatalog

Hive is based on Apache Hive 0.13.1. Apache HCatalog is now merged with Apache Hive. Hive/HCatalog includes the following patches:

- HADOOP-10839, HADOOP-10840, HADOOP-7664: Need hotfix for HADOOP-879 and HADOOP-880
- HIVE-5072: Enable directly invoke Sqoop job through WebHCat
- HIVE-5775: Introduce Cost Based Optimizer to Hive
- HIVE-6338: Hive thrift metastore fails to start with error message .MetaException cannot be cast to RuntimeException
- HIVE-6521: WebHCat cannot fetch correct percentComplete for Hive jobs
- HIVE-6564: WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- HIVE-6569: HCatalog still has references to deprecated property hive.metastore.local
- HIVE-6571: query id should be available for logging during query compilation
- HIVE-6602: Multi-user HiveServer2 throws error
- HIVE-6695: bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- HIVE-6698: hcat.py script does not correctly load the hbase storage handler jars
- HIVE-6726: Hcat cli does not close SessionState
- HIVE-6741: HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- HIVE-6745: HCat MultiOutputFormat hardcodes DistributedCache keynames

- HIVE-6788: Abandoned opened transactions not being timed out
- HIVE-6792: hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- HIVE-6799: HiveServer2 needs to map kerberos name to local name before proxy check
- HIVE-6817: Some hadoop2-only tests need diffs to be updated
- HIVE-6824: Hive HBase query fails on Tez due to missing jars - part 2
- HIVE-6861: more hadoop2 only golden files to fix
- HIVE-6868: Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- HIVE-6869: Golden file updates for tez tests
- HIVE-6871: Build fixes to allow Windows to run TestCliDriver
- HIVE-6880: TestHWISessionManager fails with -Phadoop-2
- HIVE-6888: Hive leaks MapWork objects via Utilities::gWorkMap
- HIVE-6915: Hive HBase queries fail on secure Tez cluster
- HIVE-6927: Add support for MSSQL in schematool
- HIVE-6931: Windows unit test fixes
- HIVE-6936: Provide table properties to InputFormats
- HIVE-6946: Make it easier to run WebHCat e2e tests
- HIVE-6947: More fixes for tests on hadoop-2
- HIVE-6966: More fixes for TestCliDriver on Windows
- HIVE-6967: Hive transaction manager fails when SQLServer is used as an RDBMS
- HIVE-6976: Show query id only when there's jobs on the cluster
- HIVE-7006: Fix ql_rewrite_gbtoidx.q output file
- HIVE-7009: HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- HIVE-7011: HiveInputFormat's split generation isn't thread safe
- HIVE-7031: Utiltites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- HIVE-7043: When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- HIVE-7055: config not propagating for PTFOperator
- HIVE-7061: sql std auth - insert queries without overwrite should not require delete privileges

- HIVE-7062: Hive query using SUM() windowing function fails to complete and stays stuck on reduce task
- HIVE-7065, HIVE-7085: Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- HIVE-7072: HCatLoader only loads first region of hbase table
- HIVE-7076: Plugin (exec hook) to log to application timeline data to Yarn
- HIVE-7099: Add Decimal datatype support for Windowing
- HIVE-7112: Tez processor swallows errors
- HIVE-7114: Extra Tez session is started during HiveServer2 startup
- HIVE-7118: Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- HIVE-7118: Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- HIVE-7155: Need WebHCat fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
- HIVE-7167: Hive Metastore fails to start with SQLServerException
- HIVE-7188: sum(if()) returns wrong results with vectorization
- HIVE-7190: WebHCat launcher task failure can cause two concurent user jobs to run
- HIVE-7209: allow metastore authorization api calls to be restricted to certain invokers
- HIVE-7210: NPE with "No plan file found" when running Driver instances on multiple threads
- HIVE-7268: On Windows Hive jobs in Webhcat always run on default MR mode

1.5.7. Patch information for Oozie

Oozie is based on Apache Oozie 4.0.0. It includes the following patches:

- OOZIE-1593: Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters
- OOZIE-1563: Fixed Colt jar includes GPL licence
- OOZIE-1447: Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException
- OOZIE-615: Support Oozie HA.
- OOZIE-1305: Coordinator job should have an option to recover "none" of the actions after downtime.
- OOZIE-1306: Bring cron syntax to coordinator frequency

- OOZIE-1460: Implement and document oozie HA security
- OOZIE-1486: cut down on number of small files to track a running action.
- OOZIE-1491: Make sure oozie works with secure ZooKeeper
- OOZIE-1520: SequenceFile reader fails to use doas for reading action data file.
- OOZIE-1525: Oozie workflow does not update status sometimes and is stuck in Running state.
- OOZIE-1540: When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname
- OOZIE-1541: typo in oozie HA admin-server command line documentation.
- OOZIE-1555: Launcher mapper to check for system properties before opening files for action data.
- OOZIE-1560: Log messages should have a way to identify when server it comes from when using HA.
- OOZIE-1569: Maintain backward compatibility for running jobs before upgrade.
- OOZIE-1575: Add functionality to submit sqoop jobs through http from oozie server side.
- OOZIE-1576: Add documentation for oozie sqoop CLI.
- OOZIE-1587: Add "recovery" column to CoordJob table.
- OOZIE-1580: EL variables don't get resolved in configurations imported from <job-xml>.
- OOZIE-1600: Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- OOZIE-1608: update curator to 2.4.0 when it's available to fix security hole.
- OOZIE-1618: Dryrun should check variable substitution in workflow.xml.
- OOZIE-1691: StackOverflowError in TimestampedMessageParser.parseNextLine().
- OOZIE-1722: When an ApplicationMaster restarts, it restarts the launcher job.
- OOZIE-1726: Oozie does not support _HOST when configuring kerberos security.
- OOZIE-1733: Fix test failures by oozie-1722.

1.5.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1209: DirectNetezzaManager fails to find tables from older Netezza system catalogs.

- SQOOP-1298: Cannot export to VARBINARY with null value.
- SQOOP-1297: Parameterize the Accumulo version in the build files.
- SQOOP-1282: Consider Avro files even if they carry no extension.
- SQOOP-1278: Allow use of uncommitted isolation for databases that support it as an import option.
- SQOOP-1273: Multiple append jobs can easily end up sharing directories.
- SQOOP-1268: Sqoop tarballs do not contain .gitignore and .gitattribute files.
- SQOOP-1056: Implement connection resiliency in Sqoop using pluggable failure handlers.
- SQOOP-1057: Introduce fault injection framework to test connection resiliency.
- SQOOP-1271: Sqoop hcatalog location should support older bigtop default location also.
- SQOOP-1226: –password-file option triggers FileSystemClosed exception at end of Oozie action.
- SQOOP-1260: HADOOP_MAPRED_HOME should be defaulted correctly.
- SQOOP-1259: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1261: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1249: Sqoop HCatalog Import fails with -queries because of validation issues.
- SQOOP-1250: Oracle connector is not disabling autoCommit on created connections.
- SQOOP-1246: HBaseImportJob should add job authtoken only if HBase is secured.
- SQOOP-767: Add support for Accumulo.
- SQOOP-1228: Method Configuration#unset is not available on Hadoop 1.2.0.
- SQOOP-1224: Enable use of Oracle Wallets with Oracle Manager.
- SQOOP-1227: Sqoop fails to compile against commons-io higher than 1.4.
- SQOOP-1223: Enhance the password file capability to enable plugging-in custom loaders.
- SQOOP-1216: Improve error message on corrupted input while doing export.
- SQOOP-435: Avro import should write the Schema to a file.
- SQOOP-1192: Add option "–skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- SQOOP-1032: Add the –bulk-load-dir option to support the HBase doBulkLoad function.
- SQOOP-1213: Support reading password files from Amazon S3.
- SQOOP-1203: Add another default case for finding *_HOME when not explicitly defined.
- SQOOP-1197: Enable Sqoop to build against Hadoop-2.1.0-beta jar files.

- SQOOP-1194: Make changes to Sqoop build file to enable Netezza third party tests.
- SQOOP-1167: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1190: Class HCatHadoopShims will be removed in HCatalog 0.12.
- 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-1107: Further improve error reporting when exporting malformed data.
- SQOOP-1185: LobAvroImportTestCase is sensitive to test method order execution.
- SQOOP-1170: Can't import columns with name "public".
- SQOOP-1179: Incorrect warning saying –hive-import was not specified when it was specified.
- SQOOP-1161: Generated Delimiter Set Field Should be Static.
- SQOOP-1172: Make Sqoop compatible with HBase 0.95+.

1.5.9. Patch information for Falcon

Falcon is based on Apache Falcon 0.5.0 and includes the following patch:

- FALCON-598: org.apache.falcon.entity.ProcessHelper throws NullPointerException if the process has no inputs OR no outputs defined.

1.6. Minimum System Requirements

In this section:

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

1.6.1. Hardware recommendations

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

1.6.2. Operating systems requirements

The following operating systems are supported:

- Microsoft Windows 2008 R2 x64 Server
- Microsoft Windows 2012 Server

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

1.6.3. Software requirements

On each of your hosts:

- Java version: 1.7.0_51
- Python version 2.7.X
- Visual C++ 2010
- .Net 4.0 Framework

1.6.4. Database requirements

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

1.6.5. Virtualization and cloud platforms

HDP is certified and supported when running on virtual or cloud platforms (for example, VMware vSphere or Microsoft Azure) 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.

1.7. Improvements

No improvements for Windows HDP in this release.

1.8. Common Vulnerabilities and Exposures

- **CVE-2014-0075, -0096, -0099, -0119:** Tomcat Security Vulnerabilities in Oozie

  Severity: Critical

  Vendor: The Apache Software Foundation

  Versions Affected: Tomcat 6.0.37

  Users Affected: CVS

Impact: See BUG-21622 and EAR-592

Recommended Action: Tomcat 6.0.37 users should upgrade to Tomcat 6.0.41

- **CVE-2014-0075, -0096, -0099, -0119:** Tomcat Security Vulnerabilities in HttpFS

  Severity: Major

  Vendor: The Apache Software Foundation

  Versions Affected: Tomcat versions prior to 6.0.41, 7.0.54, 8.0.8

  Users Affected: CVS

  Impact: See BUG-21623

1.9. Known Issues

In this section:

1.9.1. Known Issues for the 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: With HTD Setup Window, click continue to install even though the hostname lookup fails. For silent installs, you will have to 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…
  ```

  Put the truncated hostname in the clusterproperties file and then update the hostname in the `hdfs-site.xml` after the installation completes.

- **BUG-16051:** lzo2.dll not provided with the HDP for Windows Installer

  Description: If you enable LZO by sepcifying `ENALBLE_LZO=yes` in `cluster.properties` file, you must provide your own version of `lzo2.dll`.

Workaround: Download the LZO open source from here, and build lzo2.dll (use Visual C++ 2010 in your build environment because the dll uses the corresponding VC runtime and HDP for Windows requires VC 2010 runtime).

When deploying HDP with the LZO compression enabled, put the following three files in the same directory as the HDP for Windows Installer and the `cluster.properties` file:

- `hadoop-lzo-0.4.19.2.1.1.0-1621.jar` (from the HDP for Windows Installation zip)
- `gplcompression.dll` (from the HDP for Windows Installation zip)
- `lz2.dll` (that you built must be named lz2.dll)

If you are using the push installer script, include all three of these files in the *`$filelist`* field.

- **BUG-19468**: Random password not generated when password left blank

  Description: When installing HDP using Windows MSI from the command line and pass the HDP_USER_PASSWORD parameter without a value, a random password is not generated.

  Workaround: Specify a password, when installing from the command line.

```
cmd.exe /C start /wait msiexec /i $package /l* $log_path HDP_LAYOUT=
$cluster_layout_path HDP_DIR=$root_drive\hdp\hadoop DESTROY_DATA=`"yes`"
 HDP_USER_PASSWORD="@!S0meP@ssw0rd" KNOX_MASTER_SECRET=master_secret /qn
```

1.9.2. Known Issues for HBase

- **BUG-17850:** HBCK Tests Intermittent Fail Due to Empty Region Qualifier Error

  Problem: Region replicas are not always deleted/closed as expected, causing HBCK tests to fail.

- **BUG-16257 (HBASE-10123):** Hbase master fails to start due to BindException

  Problem: Apache defaults clash with the range LINUX assigns itself for creating come-and-go ephemeral ports.

- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

  Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the master logs.

1.9.3. Known Issues for Hive

- **BUG-17247:** In Hive Cli switching the hive.execution.engine from Tez to MapReduce does not also switch the YARN framework back to MapReduce

Problem: If we can't switch the YARN framework back to MR, Hive MR will still run on Tez.

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

  Problem: The Hive session should shut down cleanly and not kill the app.

- **BUG-16393:** Bucketized Table feature fails in some cases.

  Problem: Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination.

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.

1.9.4. Known Issues for Tez

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

  Problem: The Hive session should shut down cleanly and not kill the app.

1.9.5. Known Issues for Oozie

- **BUG-13551**: Oozie does not understand _HOST in the kerberos principal name

  Problem:

  Workaround:

1.9.6. Known Issues for the Hortonworks Connector for Teradata

- The Hortonworks Connector for Teradata is not supported at this time. Please check http://www.hortonworks.com often for updates, which will be coming soon.

1.10. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

1.11. Third-party Licenses

Table 1.1. Third-party Licenses

HDP Component	License
Phoenix	EPL
Storm	EPL
Accumulo	JCommander

HDP Component	License
Falcon	CERN
Falcon	Tinkerpop
Knox	ANTLR
Knox	MIT
Knox	EPL
Knox	Bouncy Castle
Knox	OWS

2. Release Notes HDP-2.1.3-Win

RELEASE NOTES: Hortonworks Data Platform 2.1 for Windows powered by Apache Hadoop

The HDP 2.1 Release Notes include the following sections:

2.1. Product Version: HDP-2.1.3

All HDP 2.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.

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

- Apache Hadoop 2.4.0
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.0
- Apache Tez 0.4.0
- Apache ZooKeeper 3.4.5
- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5.0
- Apache Sqoop 1.4.4

- Apache Knox 0.4.0
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0
- Quest-oraoop 1.7.0

2.1.1. Unsupported Apache components

Arvo v 1.7.4 is shipped with HDP 2.1, but is not supported.

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation (HDFS-1052)
- viewFS (HADOOP-7257)

The following Apache Components are shipped as part of HDP 2.1 YARN, but are not supported:

- MapReduce v1 Fair Scheduler (HADOOP-3746)
- YARN Fair Scheduler (MAPREDUCE-3451)
- MapReduce Uber AM (MAPREDUCE-2405)
- MapReduce Eclipse Plugin (for Non-Kerberos and Kerberos cluster)
- Cgroup resource isolation (YARN-3)
- CPU Scheduling (YARN-2)

Note

Hue is not supported on Windows.

2.1.2. Tech Previews in This Release

The following features are provided with HDP 2.1.3 as technical previews; that is, they are considered to be still under development, and as such not supported. Do not use the following features in your production systems.

- Preemption – this feature has been found to contain defects that make it unsuitable for a production cluster at this time. For information about Preemption, see http://docs.hortonworks.com/HDPDocuments/HDP2/hdp-2.1.3/bk_system-admin-guide/content/preemption.html.

- Falcon – lag time in displaying lineage information if graph contains many vertices; old graph persists until updated graph is loaded
- Falcon – dependencies, viewed in Graph View

If you have questions regarding this feature, contact Support by logging a case at the Hortonworks Support Portal. If you are currently not an official Hortonworks Customer or Partner, then please seek assistance on our Hortonworks Forums at:http://hortonworks.com/community/forums/.

2.1.3. Behavioral Changes

There are no changes in the HDP 2.1.3 for Windows release.

2.2. Patch Information

In this section:

- Patch Information for Hadoop Common/HDFS
- Patch Information for ZooKeeper
- Patch Information for HBase
- Patch Information for Pig
- Patch Information for Tez
- Patch Information for Hive/HCat
- Patch Information for Oozie

Note

Apache YARN, Apache MapReduce and Apache Knox require no additional patches.

2.2.1. Patch information for Hadoop Common/HDFS

Hadoop is based on Apache Hadoop 2.4.0 and includes the following additional patches:

- HADOOP-10630: Possible race condition in RetryInvocationHandler
- HADOOP-10612: NFS failed to refresh the user group id mapping table
- HADOOP-10562: Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- HADOOP-10508: RefreshCallQueue fails when authorization is enabled
- HADOOP-10475: ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken().
- HDFS-4052: BlockManager#invalidateWork should print logs outside the lock

- HDFS-6227: ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- HDFS-6245: datanode fails to start with a bad disk even when failed volumes is set
- HDFS-6273: Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- HDFS-6278: Create HTML5-based UI for SNN
- HDFS-6279: Create new index page for JN / DN
- HDFS-6362: InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- HDFS-6364: Incorrect check for unknown datanode in Balancer
- HDFS-6370: Web UI fails to display in intranet under IE
- HDFS-6411: nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- HDFS-6423: Diskspace quota usage should be updated when appending data to partial block
- HDFS-6432: Add snapshot related APIs to webhdfs
- HDFS-6438: DeleteSnapshot should be a DELETE request in WebHdfs
- HDFS-6458: NFS: stale NFS file handle Error for previous mount point
- HDFS-6462: NFS: fsstat request fails with the secure hdfs
- HDFS-6527: Edit log corruption due to defered INode removal
- HDFS-5257: addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- HDFS-6160: TestSafeMode occasionally fails.
- HDFS-6233: Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- MAPREDUCE-5014: Extending DistCp through a custom CopyListing is not possible

2.2.2. Patch information for ZooKeeper

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

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

2.2.3. Patch information for HBase

HBase is based on Apache HBase 0.98.0. It includes the following patches:

- HBASE-10833: Region assignment may fail during cluster start up
- HBASE-10829: Flush is skipped after log replay if the last recovered edits file is skipped
- HBASE-10514: Forward port HBASE-10466, possible data loss when failed flushes.
- HBASE-10700: IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- HBASE-10592: Refactor PerformanceEvaluatiotool
- HBASE-10419: Add multiget support to PerformanceEvaluation
- HBASE-10548: Correct commons-math dependency version
- HBASE-10809: HBaseAdmin#deleteTable fails when META region happen to move around same time
- HBASE-10793: AuthFailed as a valid zookeeper state
- HBASE-10767: Load balancer may interfere with tests in TestHBaseFsck
- HBASE-9721: RegionServer should not accept regionOpen RPC intended for another(previous) server
- HBASE-10688: Add a draining_node script to manage nodes in draining mode
- HBASE-8304: Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- HBASE-10660: MR over snapshots can OOM when alternative blockcache is enabled
- HBASE-10635: thrift#TestThriftServer fails due to TTL validity check
- HBASE-10591: Sanity check table configuration in createTable
- HBASE-10670: HBaseFsck#connect() should use new connection
- HBASE-10632: Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- HBASE-10621: Unable to grant user permission to namespace
- HBASE-10638: Improve error message when there is no region server available for move
- HBASE-10582: 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- HBASE-10581: ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- HBASE-10500: Some tools OOM when BucketCache is enabled
- HBASE-10486: ProtobufUtil Append and Increment deserialization lost cell level timestamp

- HBASE-10844: Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10852: TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10618: User should not be allowed to disable/drop visibility labels table
- HBASE-10895: unassign a region fails due to the hosting region server is in FailedServerList
- HBASE-10850: essential column family optimization is broken
- HBASE-10751: TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

Windows Fixes:

- HBASE-10799 [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- HBASE-10735 [WINDOWS] Set -XX:MaxPermSize for unit tests
- HBASE-10685 [WINDOWS] TestKeyStoreKeyProvider fails on windows

  HBASE-10686 [WINDOWS] TestStripeStoreFileManager fails on windows

Changes related to HBASE-10070:

- HBASE-10875 Metas own location should be cached
- HBASE-10791 Add integration test to demonstrate performance improvement
- HBASE-10810 LoadTestTool should share the connection and connection pool
- HBASE-10794 multi-get should handle missing replica location from cache
- HBASE-10634 Multiget doesn't fully work.
- HBASE-10661 TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- HBASE-10701 Cache invalidation improvements from client side
- HBASE-10778 Unique keys accounting in MultiThreadedReader is incorrect
- HBASE-10743 Replica map update is problematic in RegionStates
- HBASE-10616 Integration test for multi-get calls

- HBASE-10734 Fix RegionStates.getRegionAssignments to not add duplicate regions
- HBASE-10729 Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- HBASE-10726 Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- HBASE-10720 rpcClient: Wrong log level when closing the connection
- HBASE-10704 BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- HBASE-10633 StoreFileRefresherChore throws ConcurrentModificationException sometimes
- HBASE-10572 Create an IntegrationTest for region replicas.
- HBASE-10703 TestAsyncProcess does not pass on HBASE-10070
- HBASE-10637 rpcClient: Setup the iostreams when writing
- HBASE-10620 LoadBalancer.needsBalance() should check for co-located region replicas as well
- HBASE-10672 Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- HBASE-10630 NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- HBASE-10356 Failover RPC's for multi-get.
- HBASE-10525 Allow the client to use a different thread for writing to ease interrupt.
- HBASE-10355 Failover RPC's from client using region replicas.
- HBASE-10352 Region and RegionServer changes for opening region replicas, and refreshing store files
- HBASE-10351 LoadBalancer changes for supporting region replicas
- HBASE-10359 Master/RS WebUI changes for region replicas.
- HBASE-10362 HBCK changes for supporting region replicas.
- HBASE-10361 Enable/AlterTable support for region replicas.
- HBASE-10350 Master/AM/RegionStates changes to create and assign region replicas.
- HBASE-10490 Simplify RpcClient code (Nicolas Liochon)
- HBASE-10511 Add latency percentiles on PerformanceEvaluation
- HBASE-10517 NPE in MetaCache.clearCache()
- HBASE-10479 HConnection interface is public but is used internally, and contains a bunch of methods

- HBASE-10348 HTableDescriptor changes for region replicas
- HBASE-10354 Add an API for defining consistency per request
- HBASE-10347 HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- HBASE-10277 refactor AsyncProcess
- HBASE-10427 clean up HRegionLocation/ServerName usage
- HBASE-10472 Manage the interruption in ZKUtil#getData
- HBASE-10859 HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- HBASE-10858 TestRegionRebalancing is failing

2.2.4. Patch information for Pig

Pig is based on Apache Pig 0.12.1. It includes the following patches:

- PIG-3916: isEmpty should not be early terminating.
- PIG-3650: Fix for PIG-3100 breaks column pruning
- PIG-3573: Provide StoreFunc and LoadFunc for Accumulo.
- PIG-3558: ORC support for Pig.
- PIG-3257: Add a UUID function to Pig.

2.2.5. Patch information for Tez

Tez is based on Apache Tez 0.4.0 incubating release. It includes the following patches:

- TEZ-1159: Fix handling of corrupt or empty files in recovery data.
- TEZ-1158: Disable multiple AM attempts if recovery is disabled.
- TEZ-1135:: Fix ShuffledUnorderedKVInput handling of empty partitions.
- TEZ-1125: Pre-warm broken.
- TEZ-1097: Tez assumes that the scratch directory has to be same as the default filesystem.
- TEZ-1066: Generate events to integrate with YARN timeline server.
- TEZ-1048: Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- TEZ-1045: TezMiniCluster tests can fail intermittently.
- TEZ-1040: Fix a bug which could cause the Merger to hang.
- TEZ-1034: Shuffling can sometimes hang with duplicate inputs for the same index.

- TEZ-1033: AM hangs during recovery with Tasks awaiting init event.
- TEZ-1030: Address intermittent errors created due to race condition in YARN-1915.
- TEZ-1028: Handle killed tasks and attempts when handling recovery data.
- TEZ-1029: Fetcher can fail to report input failed event upon connection error.
- TEZ-1021: TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- TEZ-1020: VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- TEZ-1015: Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- TEZ-1014: Add a log message to indicate last AM attempt.
- TEZ-1004: AM relocalization doesn't handle conflicting resources correctly.
- TEZ-1005: AM relocalization adds resources to the wrong classloader.
- TEZ-1011: TestDAGRecovery timing out on jenkins builds.
- TEZ-1010: TestAMNodeMap.testSelfBlacklist fails intermittently
- TEZ-997: Internal Errror in am logs during dag shutdown.
- TEZ-1009: Fixes in log file roll-over
- TEZ-998: InvalidStateTransitonException: Invalid event: V_INIT at INITED.

2.2.6. Patch information for Hive/HCatalog

Hive is based on Apache Hive 0.13.0. Apache HCatalog is now merged with Apache Hive. Hive/HCatalog includes the following patches:

- HIVE-6976: Show query id only when there's jobs on the cluster
- HIVE-6966: More fixes for TestCliDriver on Windows
- HIVE-6927: Add support for MSSQL in schematool
- HIVE-6919: hive sql std auth select query fails on partitioned tables
- HIVE-6915: Hive Hbase queries fail on secure Tez cluster
- HIVE-6898: Functions in hive are failing with java.lang.ClassNotFoundException on Tez
- HIVE-6888: Hive leaks MapWork objects via Utilities::gWorkMap
- HIVE-6883: Dynamic partitioning optimization does not honor sort order or order by
- HIVE-6880: TestHWISessionManager fails with -Phadoop-2
- HIVE-6871: Build fixes to allow Windows to run TestCliDriver
- HIVE-6868: Create table in HCatalog sets different SerDe defaults than what is set through the CLI

- HIVE-6828: Hive tez bucket map join conversion interferes with map join conversion
- HIVE-6826: Hive-tez has issues when different partitions work off of different input types
- HIVE-6824: Hive HBase query fails on Tez due to missing jars - part 2
- HIVE-6817: Some hadoop2-only tests need diffs to be updated
- HIVE-6799: HiveServer2 needs to map kerberos name to local name before proxy check
- HIVE-6788: Abandoned opened transactions not being timed out
- HIVE-6745: HCat MultiOutputFormat hardcodes DistributedCache keynames
- HIVE-6741: HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- HIVE-6726: Hcat cli does not close SessionState
- HIVE-6695: bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- HIVE-6571: query id should be available for logging during query compilation
- HIVE-6569: HCatalog still has references to deprecated property hive.metastore.local
- HIVE-5775: Introduce Cost Based Optimizer to Hive

2.2.7. Patch information for Oozie

Oozie is based on Apache Oozie 4.0.0. It includes the following patches:

- OOZIE-1593: Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- OOZIE-1563: Fixed Colt jar includes GPL licence.
- OOZIE-1447: Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException.
- Oozie-615: Support Oozie HA.
- Oozie-1305: Coordinator job should have an option to recover "none" of the actions after downtime.
- Oozie-1306: Bring cron syntax to coordinator frequency.
- Oozie-1460: Implement and document oozie HA security.
- Oozie-1486: cut down on number of small files to track a running action.
- Oozie-1491: Make sure oozie works with secure ZooKeeper.
- Oozie-1520: SequenceFile reader fails to use doas for reading action data file.
- Oozie-1525: Oozie workflow does not update status sometimes and is stuck in Running state.

- Oozie-1540: When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- Oozie-1541: typo in oozie HA admin-server command line documentation.
- Oozie-1555: Launcher mapper to check for system properties before opening files for action data.
- Oozie-1560: Log messages should have a way to identify when server it comes from when using HA.
- Oozie-1569: Maintain backward compatibility for running jobs before upgrade.
- Oozie-1575: Add functionality to submit sqoop jobs through http from oozie server side.
- Oozie-1576: Add documentation for oozie sqoop CLI.
- Oozie-1587: Add "recovery" column to CoordJob table.
- Oozie-1580: EL variables don't get resolved in configurations imported from <job-xml>.
- Oozie-1600: Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- Oozie-1608: update curator to 2.4.0 when it's available to fix security hole.
- Oozie-1618: Dryrun should check variable substitution in workflow.xml.
- Oozie-1691: StackOverflowError in TimestampedMessageParser.parseNextLine().
- Oozie-1722: When an ApplicationMaster restarts, it restarts the launcher job.
- Oozie-1726: Oozie does not support _HOST when configuring kerberos security.
- Oozie-1733: Fix test failures by oozie-1722.

2.2.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1209: DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- SQOOP-1298: Cannot export to VARBINARY with null value.
- SQOOP-1297: Parameterize the Accumulo version in the build files.
- SQOOP-1282: Consider Avro files even if they carry no extension.
- SQOOP-1278: Allow use of uncommitted isolation for databases that support it as an import option.
- SQOOP-1273: Multiple append jobs can easily end up sharing directories.

- SQOOP-1268: Sqoop tarballs do not contain .gitignore and .gitattribute files.
- SQOOP-1056: Implement connection resiliency in Sqoop using pluggable failure handlers.
- SQOOP-1057: Introduce fault injection framework to test connection resiliency.
- SQOOP-1271: Sqoop hcatalog location should support older bigtop default location also.
- SQOOP-1226: –password-file option triggers FileSystemClosed exception at end of Oozie action.
- SQOOP-1260: HADOOP_MAPRED_HOME should be defaulted correctly.
- SQOOP-1259: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1261: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1249: Sqoop HCatalog Import fails with -queries because of validation issues.
- SQOOP-1250: Oracle connector is not disabling autoCommit on created connections.
- SQOOP-1246: HBaseImportJob should add job authtoken only if HBase is secured.
- SQOOP-767: Add support for Accumulo.
- SQOOP-1228: Method Configuration#unset is not available on Hadoop 1.2.0.
- SQOOP-1224: Enable use of Oracle Wallets with Oracle Manager.
- SQOOP-1227: Sqoop fails to compile against commons-io higher than 1.4.
- SQOOP-1223: Enhance the password file capability to enable plugging-in custom loaders.
- SQOOP-1216: Improve error message on corrupted input while doing export.
- SQOOP-435: Avro import should write the Schema to a file.
- SQOOP-1192: Add option "–skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- SQOOP-1032: Add the –bulk-load-dir option to support the HBase doBulkLoad function.
- SQOOP-1213: Support reading password files from Amazon S3.
- SQOOP-1203: Add another default case for finding *_HOME when not explicitly defined.
- SQOOP-1197: Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- SQOOP-1194: Make changes to Sqoop build file to enable Netezza third party tests.
- SQOOP-1167: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1190: Class HCatHadoopShims will be removed in HCatalog 0.12.
- 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-1107: Further improve error reporting when exporting malformed data.
- SQOOP-1185: LobAvroImportTestCase is sensitive to test method order execution.
- SQOOP-1170: Can't import columns with name "public".
- SQOOP-1179: Incorrect warning saying –hive-import was not specified when it was specified.
- SQOOP-1161: Generated Delimiter Set Field Should be Static.
- SQOOP-1172: Make Sqoop compatible with HBase 0.95+.

2.3. Minimum System Requirements

In this section:

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

2.3.1. Hardware recommendations

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

2.3.2. Operating systems requirements

The following operating systems are supported:

- Microsoft Windows 2008 R2 x64 Server
- Microsoft Windows 2012 Server

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

2.3.3. Software requirements

On each of your hosts:

- Java version: 1.7.0_51
- Python version 2.7.X
- Visual C++ 2010
- .Net 4.0 Framework

2.3.4. Database requirements

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

2.3.5. Virtualization and cloud platforms

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

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

2.4. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- The Storm LogViewer service now allows only worker log files from the Storm log directory to be viewed from the UI, regardless of the privileges of the person running the service.

2.5. Common Vulnerabilities and Exposures

- **CVE-2014-0229:** Several HDFS admin commands lack proper privilege checks
- **CVE-2013-6446:** Apache Hadoop job history server vulnerability

  Severity: Major

  Vendor: The Apache Software Foundation

  Versions Affected: Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0

  Users Affected: Users who have enabled Hadoop's MapReduce security features

  Impact: Vulnerability allows an unauthorized user to retrieve job details from the job history server

  Mitigation: Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to 2.3.0

  Credit: This issue was discovered by Koji Noguchi of Yahoo

2.6. Known Issues

In this section:

2.6.1. Known Issues for the 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: With HTD Setup Window, click continue to install even though the hostname lookup fails. For silent installs, you will have to 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…
  ```

  Put the truncated hostname in the clusterproperties file and then update the hostname in the `hdfs-site.xml` after the installation completes.

- **BUG-16051:** lzo2.dll not provided with the HDP for Windows Installer

  Description: If you enable LZO by sepcifying ENALBLE_LZO=`yes` in `cluster.properties` file, you must provide your own version of `lzo2.dll`.

  Workaround: Download the LZO open source from here, and build lzo2.dll (use Visual C++ 2010 in your build environment because the dll uses the corresponding VC runtime and HDP for Windows requires VC 2010 runtime).

  When deploying HDP with the LZO compression enabled, put the following three files in the same directory as the HDP for Windows Installer and the `cluster.properties` file:

  - `hadoop-lzo-0.4.19.2.1.1.0-1621.jar` (from the HDP for Windows Installation zip)
  - `gplcompression.dll` (from the HDP for Windows Installation zip)
  - `lz2.dll` (that you built must be named lz2.dll)

  If you are using the push installer script, include all three of these files in the *`$filelist`* field.

- **BUG-19468**: Random password not generated when password left blank

Description: When installing HDP using Windows MSI from the command line and pass the HDP_USER_PASSWORD parameter without a value, a random password is not generated.

Workaround: Specify a password, when installing from the command line.

```
cmd.exe /C start /wait msiexec /i $package /l* $log_path HDP_LAYOUT=
$cluster_layout_path HDP_DIR=$root_drive\hdp\hadoop DESTROY_DATA=`"yes`"
 HDP_USER_PASSWORD="@!S0meP@ssw0rd" KNOX_MASTER_SECRET=master_secret /qn
```

2.6.2. Known Issues for HDP

- **BUG-15796:** Sigsegv in mapred history server due to segfault in JniBasedUnixGroupsMapping.

  Problem: Got sigsegv in mapred-mapred-historyserver-hor3n25.gq1.ygridcore.net.out:

```
14/03/28 15:48:28 INFO hs.HistoryFileManager: Deleting JobSummary file:
 [hdfs://hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525.summary]
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://
hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist to hdfs:/
/hor3n25.gq1.ygridcore.net:8020/mapred/history/done/2014/03/28/
000000/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://hor3n25.
gq1.ygridcore.net:8020/mapred/history/done_intermediate/hrt_qa/
job_1396000455255_0525_conf.xml to hdfs://hor3n25.gq1.ygridcore.net:8020/
mapred/history/done/2014/03/28/000000/job_1396000455255_0525_conf.xml
#
# A fatal error has been detected by the Java Runtime Environment:
#
# SIGSEGV (0xb) at pc=0x00007f6ef9cdf984, pid=5986, tid=140112490882816
#
# JRE version: 7.0_09
# Java VM: OpenJDK 64-Bit Server VM (23.2-b09 mixed mode linux-amd64
 compressed oops)
# Problematic frame:
# C 0x00007f6ef9cdf984
#
# Failed to write core dump. Core dumps have been disabled. To enable core
 dumping, try ""ulimit -c unlimited"" before starting Java again
#
 # An error report file with more information is saved as:
# /tmp/hs_err_pid5986.log
#
# If you would like to submit a bug report, please include
# instructions on how to reproduce the bug and visit:
# http://icedtea.classpath.org/bugzilla
#
```

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

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

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

2.6.3. Known issues for HDFS

- **BUG-14542:** HDP 2.1 exception during namenode service work.

  Problem: After the start of the NameNode service, the following exception occurred:

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_0000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msecs
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
 callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
     9 more
```

2.6.4. Known Issues for YARN

- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.

  Problem: YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.

- **BUG-12327:** (YARN-90) NM cannot detect when bad disks become healthy again.

  Problem: If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.

- **BUG-9919:** Shuffle test fails with mapreduce.map.java.opts and mapreduce.reduce.java.opts -Xmx692m

  Problem: Shuffle tests fails in ambari with the following configuration:

mapred-site.xml mapreduce.map.memory.mb 1024

```
mapreduce.map.java.opts -Xmx692m

mapreduce.task.io.sort.mb 200
```

Shuffle test is failing with "java.lang.OutOfMemoryError: Java heap space".

- **BUG-12705:** Encryption for YARN and MR shuffle on Windows

  Problem: Encryption for YARN and MR shuffle not supported on HDP for Windows clusters.

2.6.5. Known Issues for HBase

- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

  Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.

- **BUG-14986:** HBase Region Server Crash After Large Table Creation on Windows with Load Balancing

  Problem: When using a Load Balancer, the HBase Region Server may crash after creating a large table.

2.6.6. Known Issues for Hive

- **BUG-19152:** Upon start, HiveServer2 doesn't know about the admin users in hive.users.in.admin.role for a while

  Problem: When HiveServer2 is started it takes a while for it to initialize the users set in the hive.users.in.admin.role property. This causes the first few tests in Hive SQL Standard Auth test suite to fail.

- **BUG-18002:** NullPointerException in OrcInputFormat when Vectorization is turned on

  Problem: We are unable to verify if the NPE occurs with the normal expected settings.

- **BUG-17850:** HBCK test fails intermittently due to Empty Region Qualifier error

- **BUG-17846:** Hive query using SUM() windowing function fails to complete and stays stuck on reduce task

  Problem: The query never completes.

- **BUG-17247:** In Hive Cli switching the hive.execution.engine from Tez to MapReduce does not also switch the YARN framework back to MapReduce

  Problem: If we can't switch the YARN framework back to MR, Hive MR will still run on Tez.

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

Problem: The Hive session should shut down cleanly and not kill the app.

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

  Problem: We force create_index to run in MR mode when we have a TEZ run, but it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16476:** Oozie-hive tests run as hadoopqa creates/accesses /tmp/hive-hadoop folder

  Problem: The issue occurs because Oozie launches the Hive client as the mapreduce user (hadoop in this case). However, the ugi information is that of the user using Oozie (hadoopqa in this case), so Hive always creates the /tmp/hive-hadoop directory for use as a scratch directory with hadoopqa as the owner. The right fix for this would be to create user specific directories in the first place and should be addressed in HIVE-6782.

  Workaround: Either wipe out the directory or to set permissions of 777 on the directory.

- **BUG-16393:** Bucketized Table feature fails in some cases.

  Problem: Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination.

- **BUG-16257:** HBase master fails to start due to BindException

  Problem: HBase on Suse 11 64 bit, Smoke test fails Intermittent with ERROR main client.ConnectionManager$HConnectionImplementation: The node /hbase is not in ZooKeeper. Basically HBase default ports clash with the range Linux assigns itself for creating come-and-go ephemeral ports. Therefore, once in awhile we'll see HBase master can't start due to port binding issue.

- **BUG-15733:** (HIVE-7071) Schema evolution is broken on Tez.

  Problem: The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
 diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
 diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
 java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
```

```
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive.ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive.ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
 cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive.ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more
```

- **BUG-16476:** Oozie-hive tests run as hadoopqa creates/accesses /tmp/hive-hadoop folder

  Problem: The issue occurs because Oozie launches the Hive client as the mapreduce user (hadoop in this case). However, the ugi information is that of the user using Oozie (hadoopqa in this case), so Hive always creates the /tmp/hive-hadoop directory for use as a scratch directory with hadoopqa as the owner. The right fix for this would be to create user specific directories in the first place and should be addressed in HIVE-6782.

  Workaround: Either wipe out the directory or to set permissions of 777 on the directory.

- **BUG-15003:** Hive sink throws exception on shutdown

  Problem: When using the Hive sink in Flume, you are likely to see the below warning in the logs followed by a stack trace when shutting down the Flume agent:

  ```
  14/03/16 17:39:07 WARN hive.HiveSink: Exception
  while closing HiveEndPoint ...
  ```

  There is no current evidence that this exception indicates data loss.

- **BUG-14986:** Region assignments for large number of regions may cause timeouts on windows

  Problem: On the Windows env, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the master logs.

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.

- **BUG-13551:** Oozie does not understand _HOST in the kerberos principal name

Problem: Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack where we can just send _HOST and the service at run time will replace _HOST with machine hostname. This is important so that in a HA setup we can push the same configs to all oozie servers.

- **BUG-10512:** Streaming / SELECT TRANSFORM doesn't work with Tez

  Problem: SELECT TRANSFORM doesn't work with Tez enabled, works in same build with Tez disabled.

- **BUG-8227:** (HIVE-6638) Hive needs to implement recovery or extend FileOutputComitter.

  Problem: When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

```
public boolean isRecoverySupported() {
        return false;
```

- **BUG-16391:** Streaming transactions fail on MSSQL.

  Problem: After creating tables using the MSSQL composite script provided by BUG-15827 running Flume, Hive Sink tests failed because no data made it into Hive tables.

2.6.7. Known Issues for Knox

- **BUG-14461:** Knox has Kerberos config at the global level rather than cluster topology level.

  Problem: The gateway.hadoop.kerberos.secured=false Boolean flag indicates whether the Hadoop cluster protected by Gateway is secured with Kerberos in gateway-site.xml.

- **BUG-16592:** Oozie only supported on HDP for Linux multi-node clusters

  Problem: When accessing a Hadoop cluster through an Apache Knox Gateway, Oozie is not supported for HDP for Windows or HDP for Linux single-node clusters.

  Workaround: Only use Oozie when access a multinode HDP for Linux cluster.

2.7. Deprecated Features

- Oracle JDK 6 is Deprecated in this release.

2.8. Third-party Licenses

Table 2.1. Third-party Licenses

HDP Component	License
Phoenix	EPL
Storm	EPL

HDP Component	License
Accumulo	JCommander
Falcon	CERN
Falcon	Tinkerpop
Knox	ANTLR
Knox	MIT
Knox	EPL
Knox	Bouncy Castle
Knox	OWS

3. Release Notes HDP-2.1.2-Win

RELEASE NOTES: Hortonworks Data Platform 2.1 for Windows powered by Apache Hadoop

The HDP 2.1 Release Notes include the following sections:

3.1. Product Version: HDP-2.1.2

All HDP 2.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.

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

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.0

Note

Apache HCatalog is merged with Apache Hive 0.13.0

- Apache Tez 0.4
- Apache ZooKeeper 3.4.5
- Hue 2.3.0 (not supported on Windows)

- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Mahout 0.9.0
- Third party components:
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

3.1.1. Unsupported Apache components

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation (Apache JIRA HDFS-1052)
- viewFS (Apache JIRA HADOOP-7257)
- viewFS (Apache JIRA HADOOP-7257)

The following Apache Components are shipped as part of HDP 2.1 YARN, but are not supported:

- Resource Manager HA
- Application Timeline Server (Hive-on-Tez metrics)
- Capacity Schedule Pre-emption
- AM failure/restart resiliency
- MapReduce Uper AM
- YARN CGrouop resource isolation
- Admin Node labels
- CPU Scheduling

- Anti-Affinity for container requests
- Secure token renewal for long-running clusters
- Fair Scheduler
- MapReduce Eclipse Plug-in

3.2. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- What's Changed in Mahout
- What's Changed in Hive

3.2.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

Deprecated algorithms were removed (https://issues.apache.org/jira/browse/MAHOUT-1296) without Frequent Pattern Mining. And was added Multilayer Perceptron (https://issues.apache.org/jira/browse/MAHOUT-1265).

3.2.2. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable hive.server2.enable.doAs is set to true, before the user starts the HiveServer2 process, they should create a scratch directory,

```
/tmp/hive-$username
```

on HDFS, where *$username* is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

3.3. Patch Information

In this section:

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

3.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 2.4 and includes the following additional patches:

- HDFS-5257: `addBlock()` retry should return LocatedBlock with locations else client will get AIOBE
- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- **BUG-8178**: Datanodes fail to register with namenode due to minimum version check.

3.3.2. Patch information for HBase

HBase is based on Apache HBase 0.98.0.

- HBASE-10188: Deprecate ServerName constructors, but make it public
- HBASE-10210: During master startup, RS can be you-are-dead-ed by master in error
- HBASE-10833: Region assignment may fail during cluster start up
- HBASE-10829: Flush is skipped after log replay if the last recovered edits file is skipped
- HBASE-10514: Forward port HBASE-10466, possible data loss when failed flushes.
- HBASE-10700: IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- HBASE-10592: Refactor PerformanceEvaluatiotool
- HBASE-10419: Add multiget support to PerformanceEvaluation
- HBASE-10548: Correct commons-math dependency version
- HBASE-10809: HBaseAdmin#deleteTable fails when META region happen to move around same time
- HBASE-10793: AuthFailed as a valid zookeeper state
- HBASE-10767: Load balancer may interfere with tests in TestHBaseFsck
- HBASE-9721: RegionServer should not accept regionOpen RPC intended for another(previous) server
- HBASE-10688: Add a draining_node script to manage nodes in draining mode
- HBASE-8304: Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- HBASE-10660: MR over snapshots can OOM when alternative blockcache is enabled
- HBASE-10635: thrift#TestThriftServer fails due to TTL validity check
- HBASE-10591: Sanity check table configuration in createTable

- HBASE-10670: HBaseFsck#connect() should use new connection
- HBASE-10632: Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- HBASE-10621: Unable to grant user permission to namespace
- HBASE-10638: Improve error message when there is no region server available for move
- HBASE-10582: 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- HBASE-10581: ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- HBASE-10500: Some tools OOM when BucketCache is enabled
- HBASE-10486: ProtobufUtil Append and Increment deserialization lost cell level timestamp
- HBASE-10844: Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10852: TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10618: User should not be allowed to disable/drop visibility labels table
- HBASE-10895: unassign a region fails due to the hosting region server is in FailedServerList
- HBASE-10850: essential column family optimization is broken
- HBASE-10751: TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- HBASE-10799: TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- HBASE-10735: `Set -XX:MaxPermSize` for unit tests
- HBASE-10685: `TestKeyStoreKeyProvider` fails on windows
- HBASE-10686: `TestStripeStoreFileManager` fails on windows

3.3.3. Patch information for ZooKeeper

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

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

3.3.4. Patch information for Pig

Pig is based on Apache Pig 0.13.0 and includes the following patches:

- PIG-3522: Remove JSCH jar from PIG
- PIG-3518: Need to ship jruby.jar in the release.
- PIG-3517: Fix PermGen error in Pig Unit test on Hadoop 2.
- PIG-3516: Pig does not bring in joda-time as dependency in its pig-template.xml.
- PIG-3512: Reducer estimater is broken by PIG-3497.
- PIG-3257: Add a UUID function to Pig.

3.3.5. Patch information for Hive

Hive is based on Apache Hive 0.13.0

Note

Apache HCatalog is now merged with Apache Hive.

- HIVE-6117: HBase_1 and HBase_2 tests are failing
- HIVE-5601: NPE in ORC's PPD when using select * from table with where predicate
- HIVE-5542: Webhcat is failing to run ddl command on a secure cluster
- HIVE-5515: Writing to an HBase table throws IllegalArgumentException, failing job submission
- HIVE-5511: `percentComplete` returned by job status from WebHCat is null
- HIVE-5496: hcat -e `drop database if exists` fails on authorizing non-existent null db
- HIVE-5485: SBAP errors on null partition being passed into partition level authorization
- HIVE-5484: TestSchemaTool failures when Hive version has more than 3 revision numbers
- HIVE-5480: WebHCat e2e tests for doAs feature are failing
- HIVE-5479:SBAP restricts `hcat -e 'show databases'`
- HIVE-5478: WebHCat e2e testsuite for hcat authorization tests needs some fixes
- HIVE-5474: drop table hangs when `concurrency=true`

- HIVE-5453: `jobsubmission2.conf` should use 'timeout' property
- HIVE-5448:webhcat duplicate test TestMapReduce_2 should be removed
- HIVE-5425: Provide a configuration option to control the default stripe size for ORC
- HIVE-5422: Upgrade Kyro to 2.22 now that it is released
- HIVE-5411: Migrate expression serialization to Kryo
- HIVE-5379: `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- HIVE-5353: job submission that requires access to metastore should not require additional jars to be shipped to target node
- HIVE-5290: Some HCatalog tests have been behaving flaky
- HIVE-5279: Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`
- HIVE-5263: Query Plan cloning time could be improved by using Kryo
- HIVE-5133: webhcat jobs that need to access metastore fails in secure mode
- HIVE-5112: Upgrade protobuf to 2.5 from 2.4
- HIVE-5070: Need to implement `listLocatedStatus()` in ProxyFileSystem for 0.23 shim
- HIVE-4910: Hadoop 2 archives broken
- HIVE-4545: HS2 should return describe table results without space padding
- HIVE-4485: beeline prints null as empty strings
- HIVE-4388: HBase tests fail against Hadoop 2
- HIVE-3815: hive table rename fails if filesystem cache is disabled
- HIVE-1511: Hive plan serialization is slow.

3.3.6. Patch information for HCatalog

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

3.3.7. Patch information for Oozie

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

- OOZIE-1593: Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- OOZIE-1563: Fixed colt jar includes GPL licence.

3.3.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1209: DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- SQOOP-1298: Cannot export to VARBINARY with null value.
- SQOOP-1297: Parameterize the Accumulo version in the build files.
- SQOOP-1282: Consider Avro files even if they carry no extension.
- SQOOP-1278: Allow use of uncommitted isolation for databases that support it as an import option.
- SQOOP-1273: Multiple append jobs can easily end up sharing directories.
- SQOOP-1268: Sqoop tarballs do not contain .gitignore and .gitattribute files.
- SQOOP-1056: Implement connection resiliency in Sqoop using pluggable failure handlers.
- SQOOP-1057: Introduce fault injection framework to test connection resiliency.
- SQOOP-1271: Sqoop hcatalog location should support older bigtop default location also.
- SQOOP-1226: –password-file option triggers FileSystemClosed exception at end of Oozie action.
- SQOOP-1260: HADOOP_MAPRED_HOME should be defaulted correctly.
- SQOOP-1259: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1261: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1249: Sqoop HCatalog Import fails with -queries because of validation issues.
- SQOOP-1250: Oracle connector is not disabling autoCommit on created connections.
- SQOOP-1246: HBaseImportJob should add job authtoken only if HBase is secured.
- SQOOP-767: Add support for Accumulo.
- SQOOP-1228: Method Configuration#unset is not available on Hadoop 1.2.0.
- SQOOP-1224: Enable use of Oracle Wallets with Oracle Manager.
- SQOOP-1227: Sqoop fails to compile against commons-io higher than 1.4.
- SQOOP-1223: Enhance the password file capability to enable plugging-in custom loaders.

- SQOOP-1216: Improve error message on corrupted input while doing export.
- SQOOP-435: Avro import should write the Schema to a file.
- SQOOP-1192: Add option "–skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- SQOOP-1032: Add the –bulk-load-dir option to support the HBase doBulkLoad function.
- SQOOP-1213: Support reading password files from Amazon S3.
- SQOOP-1203: Add another default case for finding *_HOME when not explicitly defined.
- SQOOP-1197: Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- SQOOP-1194: Make changes to Sqoop build file to enable Netezza third party tests.
- SQOOP-1167: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1190: Class HCatHadoopShims will be removed in HCatalog 0.12.
- 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-1107: Further improve error reporting when exporting malformed data.
- SQOOP-1185: LobAvroImportTestCase is sensitive to test method order execution.
- SQOOP-1170: Can't import columns with name "public".
- SQOOP-1179: Incorrect warning saying –hive-import was not specified when it was specified.
- SQOOP-1161: Generated Delimiter Set Field Should be Static.
- SQOOP-1172: Make Sqoop compatible with HBase 0.95+.

3.4. Minimum System Requirements

In this section:

3.4.1. Hardware recommendations

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

3.4.2. Operating systems requirements

The following operating systems are supported:

- Microsoft Windows 2008 R2 x64 Server
- Microsoft Windows 2012 Server

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

3.4.3. Software requirements

On each of your hosts:

- Java version: 1.7.0_51
- Python version 2.7.X
- Visual C++ 2010
- .Net 4.0 Framework

3.4.4. Database requirements

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

3.4.5. Virtualization and cloud platforms

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

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

3.5. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- Storm

- Falcon
- Hive on Tez
- YARN
- MapReduce2
- ATS
- SQL Compliance

3.6. Known Issues

In this section:

3.6.1. Known Issues for the 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: With HTD Setup Window, click continue to install even though the hostname lookup fails. For silent installs, you will have to 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…
  ```

  Put the truncated hostname in the clusterproperties file and then update the hostname in the `hdfs-site.xml` after the installation completes.

- **BUG-15899:** lzo2.dll not provided with the HDP for Windows Installer

Description: If you enable LZO by sepcifying ENALBLE_LZO=yes in `cluster.properties` file, you must provide your own version of `lzo2.dll`.

Workaround: Download the LZO open source from here, and build lzo2.dll (use Visual C++ 2010 in your build environment because the dll uses the corresponding VC runtime and HDP for Windows requires VC 2010 runtime).

When deploying HDP with the LZO compression enabled, put the following three files in the same directory as the HDP for Windows Installer and the `cluster.properties` file:

- `hadoop-lzo-0.4.19.2.1.1.0-1621.jar` (from the HDP for Windows Installation zip)
- `gplcompression.dll` (from the HDP for Windows Installation zip)
- `lz2.dll` (that you built must be named lz2.dll)

If you are using the push installer script, include all three of these files in the *`$filelist`* field.

3.6.2. Known Issues for HDP

- **BUG-15796:** Sigsegv in mapred history server due to segfault in JniBasedUnixGroupsMapping.

  Problem: Got sigsegv in mapred-mapred-historyserver-hor3n25.gq1.ygridcore.net.out:

```
14/03/28 15:48:28 INFO hs.HistoryFileManager: Deleting JobSummary file:
 [hdfs://hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525.summary]
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://
hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist to hdfs:/
/hor3n25.gq1.ygridcore.net:8020/mapred/history/done/2014/03/28/
000000/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://hor3n25.
gq1.ygridcore.net:8020/mapred/history/done_intermediate/hrt_qa/
job_1396000455255_0525_conf.xml to hdfs://hor3n25.gq1.ygridcore.net:8020/
mapred/history/done/2014/03/28/000000/job_1396000455255_0525_conf.xml
#
# A fatal error has been detected by the Java Runtime Environment:
#
# SIGSEGV (0xb) at pc=0x00007f6ef9cdf984, pid=5986, tid=140112490882816
#
# JRE version: 7.0_09
# Java VM: OpenJDK 64-Bit Server VM (23.2-b09 mixed mode linux-amd64
 compressed oops)
# Problematic frame:
# C 0x00007f6ef9cdf984
#
# Failed to write core dump. Core dumps have been disabled. To enable core
 dumping, try ""ulimit -c unlimited"" before starting Java again
#
```

```
 # An error report file with more information is saved as:
# /tmp/hs_err_pid5986.log
#
# If you would like to submit a bug report, please include
# instructions on how to reproduce the bug and visit:
# http://icedtea.classpath.org/bugzilla
#
```

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

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

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

3.6.2.1. Known issues for HDFS

- **BUG-14542:** HDP 2.1 exception during namenode service work.

  Problem: After the start of the NameNode service, the following exception occurred:

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_0000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msecs
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
 callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
     9 more
```

3.6.3. Known issues for MapReduce

- **BUG-12005:** Mapreduce.task.io.sort.mb is capped at 2047.

  Problem: mapreduce.task.io.sort.mb is hardcoded to not allow values larger than 2047. If you enter a value larger then this the map tasks will always crash at this line:

```
https://github.com/apache/hadoop-mapreduce/blob/HDFS-641/src/java/org/
apache/hadoop/mapred/MapTask.java?source=cc#L746
```

- **BUG-14749:** CombineFileInputFormat.getSplits() including directories in its results.

Problem: This is causing Hive test root_dir_external_table.q to fail when running against hadoop-2. Opened Apache Jira MAPREDUCE-5756 Created in Monarch as https://hwxmonarch.atlassian.net/browse/HADOOP-801, creating equivalent bug for Baikal.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

3.6.4. Known Issues for YARN

- **BUG-158341:** YARN and/or mapred client should add tokens for default filesystem.

  Problem: As noticed in BUG-15360 if jobs are using webhdfs then they could run into an issue where the job only has webhdfs tokens and yarn jobs would fail as it tries to talk to hdfs over default fs since tokens for default fs as not part of the job.

- **BUG-15376:** [YARN-1892] CS fast scheduling patch ends up causing excessive logging.

  Problem: Seeing about 1 GB of logs per hour.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.

  Problem: YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.

- **BUG-12327:** [Yarn-90] NM cannot detect when bad disks become healthy again.

  Problem: If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.

- **BUG-7531:** Hadoop metrics link does not contain correct content.

  Problem: In the Resource Manager UI Tools section, clicking on Logs and Metrics opens pages that do not contain correct information.

- **BUG-16811:** YARN and MR configuration not optimized for production environments or specific hardware

  Problem: The YARN and MR default configuration is adequate for single-node and some multi-node Hadoop clusters. However, Hortonworks recommends optimizing the configuration for your production environnment.

  Workaround: For Hadoop clusters in production environments, see Determine YARN and MapReduce Memory Configuration Settings to determine the proper configuration and after installation update `yarn-site.xml` and `mapred-site.xml` properties accordingly.

- **BUG-12705:** Encryption for YARN and MR shuffle on Windows

Problem: Encryption for YARN and MR shuffle not supported on HDP for Windows clusters.

- **BUG-12702:** Pipes are not support for YARN, Tez, or MapReduce

  Problem: This release does not support pipes for YARN, Tez, or MapReduce.

3.6.5. Known Issues for HBase

- **BUG-16900:** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.

  Problem: The Big Linked List Test with Chaos Monkey Test run fails with a Not Serving Region exception in the YARN logs.

- **BUG-16513:** HBCK Tests Fail Intermittently Due to NotServingRegionException.

  Problem: The HBCK Tool tests fails intermittently due to a NotServingRegionException, noted in the Master Logs

- **BUG-16257:** HBase master fails to start due to BindException.

  Problem: HBase on Suse 11 64 bit, smoke test fails intermittently with:

```
 ERROR [main] client.ConnectionManager$HConnectionImplementation: The node /
hbase is not in ZooKeeper.
```

- **BUG-14986:** HBase Bloomberg HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

  Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.

- **BUG-12167,**HBASE-10304: Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString

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

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

```
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
convertScanToString(TableMapReduceUtil.java:433)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:186)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:147)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:270)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:100)
...
```

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

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

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

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

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

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

- **BUG-14986:** HBase Region Server Crash After Large Table Creation on Windows with Load Balancing

  Problem: When using a Load Balancer, the HBase Region Server may crash after creating a large table.

3.6.6. Known Issues for Phoenix

- **BUG-16484:** Phoenix ZooKeeper quorum string cannot contain the port number.

  Problem: HDP 2.1 defines port numbers in hbase.zookeeper.quorum in hbase-site.xml, which causes conflicts when you use Phoenix on HBase. This results in an error message similar to the following:

  ```
  java.sql.SQLException: ERROR 102 (08001): Malformed connection url.
  ```

- **Workaround:** Remove the port number from hbase.zookeeper.quorum in hbase-site.xml, and include the port number in the JDBC connector string

```
jdbc:phoenix [ :<zookeeper quorum> [:<port number> ] [ :/hbase ]]
```

3.6.7. Known Issues for Hive

- **BUG-16890:** Hive SQL standard auth calls accessing local or HDFS URLs fail in Kerberos secure cluster with binary HS2 transport.

  Problem: This is blocking all CREATE table calls where we access LOCAL or HDFS uri.

```
>>> create external table studenttab10k(
name string,
age int,
gpa double)
row format delimited
fields terminated by '\t'
stored as textfile
location '/user/hcat/tests/data/studenttab10k';
2014-04-17 00:12:13,627 DEBUG [main] transport.TSaslTransport: writing data
 length: 297
2014-04-17 00:12:13,657 DEBUG [main] transport.TSaslTransport: CLIENT:
 reading data length: 351
Error: Error while compiling statement: FAILED: HiveAccessControlException
 Permission denied.
Principal [name=hrt_qa@HORTON.YGRIDCORE.NET, type=USER] does not have
 following privileges on Object
[type=DFS_URI, name=/user/hcat/tests/data/studenttab10k] : [INSERT, DELETE,
 OBJECT OWNERSHIP] (state=42000,code=40000)
```

- **BUG-16660:** On Tez setup, Hive jobs in webhcat run in default mr mode even in Hive.

  Problem: Currently when we run Hive jobs through Webhcat we always run in MR mode even though we are running them in a cluster where Hive queries would have run in Tez mode. This is only on Linux installs. The problem here is that we run hive queries using hive.tar.gz on HDFS and specifying explicit hive configurations, here are the properties that we use in webhcat-site.xml:

templeton.hive.archive	hdfs:///apps/webhcat/hive.tar.gz
templeton.hive.path	value: hive.tar.gz/hive/bin/hive
templeton.hive.home	value: hive.tar.gz/hive
templeton.hive.properties	hive.metastore.local=false, hive.metastore. uris=thrift://hivehost:9083, hive.metastore.sasl.enabled=false, hive.metastore.execute.setugi=true

  When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add "hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.
ql.exec.DDLTask. MetaException(message:Got exception: org.apache.
hadoop.security.AccessControlException Permission denied: user=
hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/
HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.
checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

  Problem: Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HFDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16864:** When Hive standard authorization is enabled, the owner of the table backing index is missing.

  Problem: The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
 (SessionState.java:printError(546)) - FAILED: HiveAccessControlException
 Permission denied. Principal [name=hrt_qa, type=USER] does not have
 following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
 name=default.default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT
 OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
```

```
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...
```

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

  Problem: The Hive session should shut down cleanly and not kill the app.

- **BUG-16771:** (Apache Bug: HIVE-6867) Hive table has multiple copies of streaming data when testing the Hive Server restart scenario.

  Problem: When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

  Problem: We force create_index to run in MR mode when we have a TEZ run. But it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16393:** Bucketized Table feature fails in some cases.

  Problem: Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination. Example follows:

```
CREATE TABLE P1(key STRING, val STRING)
CLUSTERED BY (key) SORTED BY (key) INTO 2 BUCKETS STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/Users/jpullokkaran/apache-hive1/data/files/P1.txt'
 INTO TABLE P1;

-- perform an insert to make sure there are 2 files
INSERT OVERWRITE TABLE P1 select key, val from P1;
```

- **Workaround:** Avoid loading data for bucketed table.

- **BUG-16391:** Streaming transactions fail on MSSQL.

  Problem: After creating tables using the MSSQL composite script provided by BUG-15827 running Flume, Hive Sink tests failed because no data made it into Hive tables.

- **BUG-15733:** Schema evolution is broken on Tez.

  Problem: The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
 diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
 diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
 java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive.ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive.ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
 cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive.ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
```

```
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more
```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.

- **BUG-8227:** Hive needs to implement recovery or extend FileOutputComitter.

  Problem: When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

  ```
  public boolean isRecoverySupported() {
          return false;
  ```

- **BUG-14965:** HCatalog log located at `%HIVE_HOME%/log/hcat.log`

  Problem: The HCatalog log file is in the wrong location. HCatalog logs are not written to `%HDP_LOG_DIR%`.

  Workaround: View the log at `%HIVE_HOME%/log/hcat.log`.

- **BUG-16391:** Streaming Transaction not supported for Microsoft SQL Server metastores

  Problem: HCatalog's new streaming ingest feature does not work when Microsoft SQL Server as the metadata store database.

  Workaround: When using the streaming transactions feature, use Derby for the metastore.

3.6.8. Known Issues for Tez

- **BUG-15376:** {YARN-1892] CS fast scheduling patch ends up causing excessive logging.

  Problem: Seeing about 1 GB of logs per hour.

3.6.9. Known Issues for Oozie

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.
ql.exec.DDLTask. MetaException(message:Got exception: org.apache.
hadoop.security.AccessControlException Permission denied: user=
hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/
HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.
checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

  Problem: Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HFDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

  ```
  D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
  drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
  drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
  ```

- **BUG-13551:** Oozie does not understand _HOST in the Kerberos principal name.

  Problem: Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack, where we can just send _HOST and at run time the service replaces _HOST with machine hostname. This is important so that in a HA setup we can push the same configs to all Oozie servers.

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

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

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

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

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

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

  Problem: From the Oozie log:

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

But this job and the application complete successfully.

3.6.10. Known Issues for Flume

- **BUG-16771:** Hive table returns multiple copies of streaming data when testing the Hive Server restart scenario.

  Problem: When running the Hive restart test where the Hive metastore is bounced while Flume is treaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

3.6.11. Known Issues for Storm

- **BUG-16232:** Storm python support can use wrong version of python if supervisor host has more than one version of python installed.

  Problem: Storm requires the default system python interpreter to be version 2.6 or higher. Earlier versions of python can see this conflict.

  Workaround: Ensure that the default system python interpreter is version 2.6 or higher.

- **BUG-15960:** Worker node gets 'FileNotFoundException : stormconf.ser'.

  Problem: While running Storm-HDFS topologies in a secure environment, the following error was observed in the worker node:

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
 CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
 ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
 mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
 ~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
 ~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
 ~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
```

```
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$_main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
 initialization")
```

3.6.12. Known Issues for Falcon

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: Falcon generated hive-action does not pass the hive-site.xml with the right configuration parameters. One manifestation of the problem will be the failure in table import job where user "hive" will be used to write to a directory owned by the table owner. This is because hive.metastore.execute.setugi parameter is not being passed as part of the hive action.

  Workaround: Add a Hive default configuration to Oozie.

  Stop the Oozie service.

  

  ### Warning

  This change allows you to work with Hive tables and Oozie workflows, but will impact all Hive actions, including non-Falcon Oozie workflows.

  Under the oozie configuration directory (typically `/etc/oozie/conf`), there will be a subdirectory called `action-conf`. Under that directory, either create or modify the file hive-site.xml and add the following:

  ```
  <property>
    <name>hive.metastore.execute.setugi</name>
    <value>true</value>
</property>
  ```

  After making this change restart the Oozie service. If Oozie is configured for HA, perform this configuration change on all Oozie server nodes.

- **BUG-16290:** Strange delegation token issues in secure clusters

  Problem: Inconsistencies in rules for `hadoop.security.auth_to_local` can lead to issues with delgation token renewals in secure clusters.

  Workaround: Verify that `hadoop.security.auth_to_local` in core-site.xml is consistent across all clusters.

- **BUG-16290, FALCON-389:** Oozie config changes needed to support HCat replication in Falcon

Problem: Oozie config changes are needed before Falcon can handle HCat replication.

Workaround: Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.

2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to /etc/hadoop/conf-1 and the other to /etc/hadoop/conf-2.

3. For each oozie-site.xml file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

   For example, if Falcon connects to 3 clusters, specify:

```
<property>
      <name>oozie.service.HadoopAccessorService.hadoop.configurations</
name>
      <value>*=/etc/hadoop/
conf,$NameNode:$rpcPortNN=$hadoopConfDir1,$ResourceManager1:$rpcPortRM=$hadoopConfDir1,$
 :$rpcPortNN =$hadoopConfDir3,$ResourceManager3 :$rpcPortRM
 =$hadoopConfDir3</value>
      <description>
          Comma separated AUTHORITY=HADOOP_CONF_DIR, where AUTHORITY is
 the HOST:PORT of
          the Hadoop service (JobTracker, HDFS). The wildcard '*'
 configuration is
          used when there is no exact match for an authority. The
 HADOOP_CONF_DIR contains
          the relevant Hadoop *-site.xml files. If the path is relative is
 looked within
          the Oozie configuration directory; though the path can be
 absolute (i.e. to point
          to Hadoop client conf/ directories in the local filesystem.
      </description>
   </property>
```

4. Restart the Oozie service on all clusters.

3.6.13. Known Issues for Knox

- **BUG-14461:** Knox has Kerberos config at the global level rather than cluster topology level.

  Problem: The gateway.hadoop.kerberos.secured=false Boolean flag indicates whether the Hadoop cluster protected by Gateway is secured with Kerberos in gateway-site.xml.

- **BUG-16592:** Oozie only supported on HDP for Linux multi-node clusters

  Problem: When accessing a Hadoop cluster through an Apache Knox Gateway, Oozie is not supported for HDP for Windows or HDP for Linux single-node clusters.

  Workaround: Only use Oozie when access a multinode HDP for Linux cluster.

3.7. Deprecated Features

- Oracle JDK 6 is Deprecated in this release.

3.8. Third-party Licenses

Table 3.1. Third-party Licenses

HDP Component	License
Phoenix	EPL
Storm	MIT License
Storm	LGPL (org.jgrapht-core 0.9.0 jar)
Storm	EPL
Falcon	CERN
Falcon	Tinkerpop
Knox	ANTLR
Knox	MIT
Knox	EPL
Knox	Bouncy Castle
Knox	OWS

4. Release Notes HDP-2.1.1-Win

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

The HDP 2.1 Release Notes include the following sections:

4.1. Product Version: HDP-2.1.1

All HDP 2.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.

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

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.0

Note

Apache HCatalog is merged with Apache Hive 0.13.0

- Apache Tez 0.4
- Apache ZooKeeper 3.4.5
- Hue 2.3.0 (not supported on Windows)

- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Mahout 0.9.0
- Third party components:
  - Ganglia 3.5.0
  - Ganglia Web 3.5.7
  - Nagios 3.5.0

4.1.1. Unsupported Apache components

The following Apache Components are shipped as part of HDP 2.1 HDFS, but are not supported:

- NameNode Federation (Apache JIRA HDFS-1052)
- viewFS (Apache JIRA HADOOP-7257)
- viewFS (Apache JIRA HADOOP-7257)

The following Apache Components are shipped as part of HDP 2.1 YARN, but are not supported:

- Resource Manager HA
- Application Timeline Server (Hive-on-Tez metrics)
- Capacity Schedule Pre-emption
- AM failure/restart resiliency
- MapReduce Uper AM
- YARN CGrouop resource isolation
- Admin Node labels
- CPU Scheduling

- Anti-Affinity for container requests
- Secure token renewal for long-running clusters
- Fair Scheduler
- MapReduce Eclipse Plug-in

4.2. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- What's Changed in Mahout
- What's Changed in Hive

4.2.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

Deprecated algorithms were removed (https://issues.apache.org/jira/browse/MAHOUT-1296) without Frequent Pattern Mining. And was added Multilayer Perceptron (https://issues.apache.org/jira/browse/MAHOUT-1265).

4.2.2. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable hive.server2.enable.doAs is set to true, before the user starts the HiveServer2 process, they should create a scratch directory,

```
/tmp/hive-[username]
```

, on HDFS, where [username] is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

4.3. 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

4.3.1. Patch information for Hadoop

Hadoop is based on Apache Hadoop 2.4 and includes the following additional patches:

- HDFS-5257: `addBlock()` retry should return LocatedBlock with locations else client will get AIOBE
- HDFS-5089: When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- **BUG-8178**: Datanodes fail to register with namenode due to minimum version check.

4.3.2. Patch information for HBase

HBase is based on Apache HBase 0.98.0.

- HBASE-10188: Deprecate ServerName constructors, but make it public
- HBASE-10210: During master startup, RS can be you-are-dead-ed by master in error
- HBASE-10833: Region assignment may fail during cluster start up
- HBASE-10829: Flush is skipped after log replay if the last recovered edits file is skipped
- HBASE-10514: Forward port HBASE-10466, possible data loss when failed flushes.
- HBASE-10700: IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- HBASE-10592: Refactor PerformanceEvaluatiotool
- HBASE-10419: Add multiget support to PerformanceEvaluation
- HBASE-10548: Correct commons-math dependency version
- HBASE-10809: HBaseAdmin#deleteTable fails when META region happen to move around same time
- HBASE-10793: AuthFailed as a valid zookeeper state
- HBASE-10767: Load balancer may interfere with tests in TestHBaseFsck
- HBASE-9721: RegionServer should not accept regionOpen RPC intended for another(previous) server
- HBASE-10688: Add a draining_node script to manage nodes in draining mode
- HBASE-8304: Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- HBASE-10660: MR over snapshots can OOM when alternative blockcache is enabled
- HBASE-10635: thrift#TestThriftServer fails due to TTL validity check
- HBASE-10591: Sanity check table configuration in createTable

- HBASE-10670: HBaseFsck#connect() should use new connection
- HBASE-10632: Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- HBASE-10621: Unable to grant user permission to namespace
- HBASE-10638: Improve error message when there is no region server available for move
- HBASE-10582: 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- HBASE-10581: ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- HBASE-10500: Some tools OOM when BucketCache is enabled
- HBASE-10486: ProtobufUtil Append and Increment deserialization lost cell level timestamp
- HBASE-10844: Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10852: TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- HBASE-10863: Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- HBASE-10618: User should not be allowed to disable/drop visibility labels table
- HBASE-10895: unassign a region fails due to the hosting region server is in FailedServerList
- HBASE-10850: essential column family optimization is broken
- HBASE-10751: TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- HBASE-10799: `TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper` fails on windows
- HBASE-10735: `Set -XX:MaxPermSize` for unit tests
- HBASE-10685: `TestKeyStoreKeyProvider` fails on windows
- HBASE-10686: `TestStripeStoreFileManager` fails on windows

4.3.3. Patch information for ZooKeeper

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

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

4.3.4. Patch information for Pig

Pig is based on Apache Pig 0.13.0 and includes the following patches:

- PIG-3522: Remove JSCH jar from PIG
- PIG-3518: Need to ship jruby.jar in the release.
- PIG-3517: Fix PermGen error in Pig Unit test on Hadoop 2.
- PIG-3516: Pig does not bring in joda-time as dependency in its pig-template.xml.
- PIG-3512: Reducer estimater is broken by PIG-3497.
- PIG-3257: Add a UUID function to Pig.

4.3.5. Patch information for Hive

Hive is based on Apache Hive 0.13.0

Note

Apache HCatalog is now merged with Apache Hive.

- HIVE-6117: HBase_1 and HBase_2 tests are failing
- HIVE-5601: NPE in ORC's PPD when using select * from table with where predicate
- HIVE-5542: Webhcat is failing to run ddl command on a secure cluster
- HIVE-5515: Writing to an HBase table throws IllegalArgumentException, failing job submission
- HIVE-5511: `percentComplete` returned by job status from WebHCat is null
- HIVE-5496: hcat -e `drop database if exists` fails on authorizing non-existent null db
- HIVE-5485: SBAP errors on null partition being passed into partition level authorization
- HIVE-5484: TestSchemaTool failures when Hive version has more than 3 revision numbers
- HIVE-5480: WebHCat e2e tests for doAs feature are failing
- HIVE-5479:SBAP restricts `hcat -e 'show databases'`
- HIVE-5478: WebHCat e2e testsuite for hcat authorization tests needs some fixes
- HIVE-5474: drop table hangs when `concurrency=true`

- HIVE-5453: `jobsubmission2.conf` should use 'timeout' property
- HIVE-5448:webhcat duplicate test TestMapReduce_2 should be removed
- HIVE-5425: Provide a configuration option to control the default stripe size for ORC
- HIVE-5422: Upgrade Kyro to 2.22 now that it is released
- HIVE-5411: Migrate expression serialization to Kryo
- HIVE-5379: `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- HIVE-5353: job submission that requires access to metastore should not require additional jars to be shipped to target node
- HIVE-5290: Some HCatalog tests have been behaving flaky
- HIVE-5279: Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`
- HIVE-5263: Query Plan cloning time could be improved by using Kryo
- HIVE-5133: webhcat jobs that need to access metastore fails in secure mode
- HIVE-5112: Upgrade protobuf to 2.5 from 2.4
- HIVE-5070: Need to implement `listLocatedStatus()` in ProxyFileSystem for 0.23 shim
- HIVE-4910: Hadoop 2 archives broken
- HIVE-4545: HS2 should return describe table results without space padding
- HIVE-4485: beeline prints null as empty strings
- HIVE-4388: HBase tests fail against Hadoop 2
- HIVE-3815: hive table rename fails if filesystem cache is disabled
- HIVE-1511: Hive plan serialization is slow.

4.3.6. Patch information for HCatalog

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

4.3.7. Patch information for Oozie

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

- OOZIE-1593: Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- OOZIE-1563: Fixed colt jar includes GPL licence.

4.3.8. Patch information for Sqoop

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

- SQOOP-1617: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1209: DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- SQOOP-1298: Cannot export to VARBINARY with null value.
- SQOOP-1297: Parameterize the Accumulo version in the build files.
- SQOOP-1282: Consider Avro files even if they carry no extension.
- SQOOP-1278: Allow use of uncommitted isolation for databases that support it as an import option.
- SQOOP-1273: Multiple append jobs can easily end up sharing directories.
- SQOOP-1268: Sqoop tarballs do not contain .gitignore and .gitattribute files.
- SQOOP-1056: Implement connection resiliency in Sqoop using pluggable failure handlers.
- SQOOP-1057: Introduce fault injection framework to test connection resiliency.
- SQOOP-1271: Sqoop hcatalog location should support older bigtop default location also.
- SQOOP-1226: –password-file option triggers FileSystemClosed exception at end of Oozie action.
- SQOOP-1260: HADOOP_MAPRED_HOME should be defaulted correctly.
- SQOOP-1259: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1261: Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- SQOOP-1249: Sqoop HCatalog Import fails with -queries because of validation issues.
- SQOOP-1250: Oracle connector is not disabling autoCommit on created connections.
- SQOOP-1246: HBaseImportJob should add job authtoken only if HBase is secured.
- SQOOP-767: Add support for Accumulo.
- SQOOP-1228: Method Configuration#unset is not available on Hadoop 1.2.0.
- SQOOP-1224: Enable use of Oracle Wallets with Oracle Manager.
- SQOOP-1227: Sqoop fails to compile against commons-io higher than 1.4.
- SQOOP-1223: Enhance the password file capability to enable plugging-in custom loaders.

- SQOOP-1216: Improve error message on corrupted input while doing export.
- SQOOP-435: Avro import should write the Schema to a file.
- SQOOP-1192: Add option "–skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- SQOOP-1032: Add the –bulk-load-dir option to support the HBase doBulkLoad function.
- SQOOP-1213: Support reading password files from Amazon S3.
- SQOOP-1203: Add another default case for finding *_HOME when not explicitly defined.
- SQOOP-1197: Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- SQOOP-1194: Make changes to Sqoop build file to enable Netezza third party tests.
- SQOOP-1167: Enhance HCatalog support to allow direct mode connection manager implementations.
- SQOOP-1190: Class HCatHadoopShims will be removed in HCatalog 0.12.
- 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-1107: Further improve error reporting when exporting malformed data.
- SQOOP-1185: LobAvroImportTestCase is sensitive to test method order execution.
- SQOOP-1170: Can't import columns with name "public".
- SQOOP-1179: Incorrect warning saying –hive-import was not specified when it was specified.
- SQOOP-1161: Generated Delimiter Set Field Should be Static.
- SQOOP-1172: Make Sqoop compatible with HBase 0.95+.

4.4. Minimum System Requirements

In this section:

4.4.1. Hardware recommendations

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

4.4.2. Operating systems requirements

The following operating systems are supported:

- Microsoft Windows 2008 R2 x64 Server
- Microsoft Windows 2012 Server

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

4.4.3. Software requirements

On each of your hosts:

- Java version: 1.7.0_51
- Python version 2.7.X
- Visual C++ 2010
- .Net 4.0 Framework

4.4.4. Database requirements

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

4.4.5. Virtualization and cloud platforms

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

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

4.5. Improvements

In addition to improvements of existing features, this release of HDP 2.1 includes the following new features and improvements:

- Storm

- Falcon
- Hive on Tez
- YARN
- MapReduce2
- ATS
- SQL Compliance

4.6. Known Issues

In this section:

4.6.1. Known Issues for the 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: With HTD Setup Window, click continue to install even though the hostname lookup fails. For silent installs, you will have to 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…
  ```

  Put the truncated hostname in the clusterproperties file and then update the hostname in the `hdfs-site.xml` after the installation completes.

- **BUG-15899:** lzo2.dll not provided with the HDP for Windows Installer

Description: If you enable LZO by sepcifying ENALBLE_LZO=yes in `cluster.properties` file, you must provide your own version of `lzo2.dll`.

Workaround: Download the LZO open source from here, and build lzo2.dll (use Visual C++ 2010 in your build environment because the dll uses the corresponding VC runtime and HDP for Windows requires VC 2010 runtime).

When deploying HDP with the LZO compression enabled, put the following three files in the same directory as the HDP for Windows Installer and the `cluster.properties` file:

- `hadoop-lzo-0.4.19.2.1.1.0-1621.jar` (from the HDP for Windows Installation zip)
- `gplcompression.dll` (from the HDP for Windows Installation zip)
- `lz2.dll` (that you built must be named lz2.dll)

If you are using the push installer script, include all three of these files in the *`$filelist`* field.

4.6.2. Known Issues for HDP

- **BUG-15796:** Sigsegv in mapred history server due to segfault in JniBasedUnixGroupsMapping.

  Problem: Got sigsegv in mapred-mapred-historyserver-hor3n25.gq1.ygridcore.net.out:

```
14/03/28 15:48:28 INFO hs.HistoryFileManager: Deleting JobSummary file:
 [hdfs://hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525.summary]
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://
hor3n25.gq1.ygridcore.net:8020/mapred/history/done_intermediate/
hrt_qa/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist to hdfs:/
/hor3n25.gq1.ygridcore.net:8020/mapred/history/done/2014/03/28/
000000/job_1396000455255_0525-1396021694603-hrt_qa-PigLatin%3Aid.
pig-1396021706500-1-0-SUCCEEDED-default-1396021699875.jhist
14/03/28 15:48:28 INFO hs.HistoryFileManager: Moving hdfs://hor3n25.
gq1.ygridcore.net:8020/mapred/history/done_intermediate/hrt_qa/
job_1396000455255_0525_conf.xml to hdfs://hor3n25.gq1.ygridcore.net:8020/
mapred/history/done/2014/03/28/000000/job_1396000455255_0525_conf.xml
#
# A fatal error has been detected by the Java Runtime Environment:
#
# SIGSEGV (0xb) at pc=0x00007f6ef9cdf984, pid=5986, tid=140112490882816
#
# JRE version: 7.0_09
# Java VM: OpenJDK 64-Bit Server VM (23.2-b09 mixed mode linux-amd64
 compressed oops)
# Problematic frame:
# C 0x00007f6ef9cdf984
#
# Failed to write core dump. Core dumps have been disabled. To enable core
 dumping, try ""ulimit -c unlimited"" before starting Java again
#
```

```
 # An error report file with more information is saved as:
# /tmp/hs_err_pid5986.log
#
# If you would like to submit a bug report, please include
# instructions on how to reproduce the bug and visit:
# http://icedtea.classpath.org/bugzilla
#
```

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

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

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

4.6.2.1. Known issues for HDFS

- **BUG-14542:** HDP 2.1 exception during namenode service work.

  Problem: After the start of the NameNode service, the following exception occurred:

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_0000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msecs
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
 callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
     9 more
```

4.6.3. Known issues for MapReduce

- **BUG-12005:** Mapreduce.task.io.sort.mb is capped at 2047.

  Problem: mapreduce.task.io.sort.mb is hardcoded to not allow values larger than 2047. If you enter a value larger then this the map tasks will always crash at this line:

```
https://github.com/apache/hadoop-mapreduce/blob/HDFS-641/src/java/org/
apache/hadoop/mapred/MapTask.java?source=cc#L746
```

- **BUG-14749:** CombineFileInputFormat.getSplits() including directories in its results.

Problem: This is causing Hive test root_dir_external_table.q to fail when running against hadoop-2. Opened Apache Jira MAPREDUCE-5756 Created in Monarch as https://hwxmonarch.atlassian.net/browse/HADOOP-801, creating equivalent bug for Baikal.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

4.6.4. Known Issues for YARN

- **BUG-158341:** YARN and/or mapred client should add tokens for default filesystem.

  Problem: As noticed in BUG-15360 if jobs are using webhdfs then they could run into an issue where the job only has webhdfs tokens and yarn jobs would fail as it tries to talk to hdfs over default fs since tokens for default fs as not part of the job.

- **BUG-15376:** [YARN-1892] CS fast scheduling patch ends up causing excessive logging.

  Problem: Seeing about 1 GB of logs per hour.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.

- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.

  Problem: YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.

- **BUG-12327:** [Yarn-90] NM cannot detect when bad disks become healthy again.

  Problem: If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.

- **BUG-7531:** Hadoop metrics link does not contain correct content.

  Problem: In the Resource Manager UI Tools section, clicking on Logs and Metrics opens pages that do not contain correct information.

- **BUG-16811:** YARN and MR configuration not optimized for production environments or specific hardware

  Problem: The YARN and MR default configuration is adequate for single-node and some multi-node Hadoop clusters. However, Hortonworks recommends optimizing the configuration for your production environnment.

  Workaround: For Hadoop clusters in production environments, see Determine YARN and MapReduce Memory Configuration Settings to determine the proper configuration and after installation update `yarn-site.xml` and `mapred-site.xml` properties accordingly.

- **BUG-12705:** Encryption for YARN and MR shuffle on Windows

Problem: Encryption for YARN and MR shuffle not supported on HDP for Windows clusters.

- **BUG-12702:** Pipes are not support for YARN, Tez, or MapReduce

  Problem: This release does not support pipes for YARN, Tez, or MapReduce.

4.6.5. Known Issues for HBase

- **BUG-16900:** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.

  Problem: The Big Linked List Test with Chaos Monkey Test run fails with a Not Serving Region exception in the YARN logs.

- **BUG-16513:** HBCK Tests Fail Intermittently Due to NotServingRegionException.

  Problem: The HBCK Tool tests fails intermittently due to a NotServingRegionException, noted in the Master Logs

- **BUG-16257:** HBase master fails to start due to BindException.

  Problem: HBase on Suse 11 64 bit, smoke test fails intermittently with:

```
 ERROR [main] client.ConnectionManager$HConnectionImplementation: The node /
hbase is not in ZooKeeper.
```

- **BUG-14986:** HBase Bloomberg HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.

  Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.

- **BUG-12167,**HBASE-10304: Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString

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

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

```
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
convertScanToString(TableMapReduceUtil.java:433)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:186)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:147)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:270)
 at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.
initTableMapperJob(TableMapReduceUtil.java:100)
...
```

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

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

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

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

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

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

- **BUG-14986:** HBase Region Server Crash After Large Table Creation on Windows with Load Balancing

  Problem: When using a Load Balancer, the HBase Region Server may crash after creating a large table.

4.6.6. Known Issues for Phoenix

- **BUG-16484:** Phoenix ZooKeeper quorum string cannot contain the port number.

  Problem: HDP 2.1 defines port numbers in hbase.zookeeper.quorum in hbase-site.xml, which causes conflicts when you use Phoenix on HBase. This results in an error message similar to the following:

  ```
  java.sql.SQLException: ERROR 102 (08001): Malformed connection url.
  ```

- **Workaround:** Remove the port number from hbase.zookeeper.quorum in hbase-site.xml, and include the port number in the JDBC connector string

```
jdbc:phoenix [ :<zookeeper quorum> [:<port number> ] [ :/hbase ]]
```

4.6.7. Known Issues for Hive

- **BUG-16890:** Hive SQL standard auth calls accessing local or HDFS URLs fail in Kerberos secure cluster with binary HS2 transport.

  Problem: This is blocking all CREATE table calls where we access LOCAL or HDFS uri.

```
>>> create external table studenttab10k(
name string,
age int,
gpa double)
row format delimited
fields terminated by '\t'
stored as textfile
location '/user/hcat/tests/data/studenttab10k';
2014-04-17 00:12:13,627 DEBUG [main] transport.TSaslTransport: writing data
 length: 297
2014-04-17 00:12:13,657 DEBUG [main] transport.TSaslTransport: CLIENT:
 reading data length: 351
Error: Error while compiling statement: FAILED: HiveAccessControlException
 Permission denied.
Principal [name=hrt_qa@HORTON.YGRIDCORE.NET, type=USER] does not have
 following privileges on Object
[type=DFS_URI, name=/user/hcat/tests/data/studenttab10k] : [INSERT, DELETE,
 OBJECT OWNERSHIP] (state=42000,code=40000)
```

- **BUG-16660:** On Tez setup, Hive jobs in webhcat run in default mr mode even in Hive.

  Problem: Currently when we run Hive jobs through Webhcat we always run in MR mode even though we are running them in a cluster where Hive queries would have run in Tez mode. This is only on Linux installs. The problem here is that we run hive queries using hive.tar.gz on HDFS and specifying explicit hive configurations, here are the properties that we use in webhcat-site.xml:

templeton.hive.archive	hdfs:///apps/webhcat/hive.tar.gz
templeton.hive.path	value: hive.tar.gz/hive/bin/hive
templeton.hive.home	value: hive.tar.gz/hive
templeton.hive.properties	hive.metastore.local=false, hive.metastore. uris=thrift://hivehost:9083, hive.metastore.sasl.enabled=false, hive.metastore.execute.setugi=true

  When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add "hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.
ql.exec.DDLTask. MetaException(message:Got exception: org.apache.
hadoop.security.AccessControlException Permission denied: user=
hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/
HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.
checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

  Problem: Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HFDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

```
D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
```

- **BUG-16864:** When Hive standard authorization is enabled, the owner of the table backing index is missing.

  Problem: The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
 (SessionState.java:printError(546)) - FAILED: HiveAccessControlException
 Permission denied. Principal [name=hrt_qa, type=USER] does not have
 following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
 name=default.default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT
 OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
```

```
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...
```

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.

  Problem: The Hive session should shut down cleanly and not kill the app.

- **BUG-16771:** (Apache Bug: HIVE-6867) Hive table has multiple copies of streaming data when testing the Hive Server restart scenario.

  Problem: When running the Hive restart test where the Hive metastore is bounced while Flume is streaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

- **BUG-16667:** Alter index rebuild fails with FS-based stats gathering.

  Problem: We force create_index to run in MR mode when we have a TEZ run. But it is failing intermittently. (This problem is not seen on non-Tez runs.)

- **BUG-16393:** Bucketized Table feature fails in some cases.

  Problem: Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination. Example follows:

```
CREATE TABLE P1(key STRING, val STRING)
CLUSTERED BY (key) SORTED BY (key) INTO 2 BUCKETS STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/Users/jpullokkaran/apache-hive1/data/files/P1.txt'
 INTO TABLE P1;

-- perform an insert to make sure there are 2 files
INSERT OVERWRITE TABLE P1 select key, val from P1;
```

- **Workaround:** Avoid loading data for bucketed table.

- **BUG-16391:** Streaming transactions fail on MSSQL.

  Problem: After creating tables using the MSSQL composite script provided by BUG-15827 running Flume, Hive Sink tests failed because no data made it into Hive tables.

- **BUG-15733:** Schema evolution is broken on Tez.

  Problem: The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
 diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
 diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
 java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive.ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive.ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
 cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive.ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
```

```
at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more
```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.

- **BUG-8227:** Hive needs to implement recovery or extend FileOutputComitter.

  Problem: When running Hive jobs and restarting RM, Hive jobs start again from scratch, causing the job to fail after the maximum number of retries. OutputComitter defaults recovery to false (see below). Hive needs to implement recovery or move to extending FileOutputComitter.

  ```
  public boolean isRecoverySupported() {
          return false;
  ```

- **BUG-14965:** HCatalog log located at `%HIVE_HOME%/log/hcat.log`

  Problem: The HCatalog log file is in the wrong location. HCatalog logs are not written to `%HDP_LOG_DIR%`.

  Workaround: View the log at `%HIVE_HOME%/log/hcat.log`.

- **BUG-16391:** Streaming Transaction not supported for Microsoft SQL Server metastores

  Problem: HCatalog's new streaming ingest feature does not work when Microsoft SQL Server as the metadata store database.

  Workaround: When using the streaming transactions feature, use Derby for the metastore.

4.6.8. Known Issues for Tez

- **BUG-15376:** {YARN-1892] CS fast scheduling patch ends up causing excessive logging.

  Problem: Seeing about 1 GB of logs per hour.

4.6.9. Known Issues for Oozie

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: The job fails with the following permission error:

```
Copying data from hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20
Copying file: hdfs://arpit-falcon-2.cs1cloud.internal:8020/projects/
ivory/staging/FALCON_FEED_REPLICATION_raaw-logs16-a6acf050-a038-48d5-9867-
de63707291a8_corp-cdd34e35-86b6-45ae-a6cf-d6e879b7b7fb/default/
HCatReplication_oneSourceOneTarget_hyphen/dt=2010-01-01-20/2010-01-01-20-00/
data/dt=2010-01-01-20/data.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.
ql.exec.DDLTask. MetaException(message:Got exception: org.apache.
hadoop.security.AccessControlException Permission denied: user=
hive, access=WRITE, inode="/tmp/falcon-regression/HCatReplication/
HCatReplication_oneSourceOneTarget_hyphen":arpit:hdfs:drwxr-xr-x
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.
checkFsPermission(FSPermissionChecker.java:265)
```

- **BUG-16476:** Oozie-Hive tests run as hadoopqa creates/accesses the /tmp/hive-hadoop folder.

  Problem: Oozie-Hive tests were run as "hadoopqa" user, concurrently with hcatalog tests. When the tests failed, the HFDFS permissions were as shown below. It is unclear why /tmp/hive-hadoop folder was ever created.

  ```
  D:\hdp\hadoop-2.4.0.2.1.1.0-1533\bin>hadoop.cmd dfs -ls /tmp
  drwxr-xr-x - hadoop hdfs 0 2014-04-09 19:01 /tmp/hive-hadoop
  drwxr-xr-x - hadoopqa hdfs 0 2014-04-09 18:50 /tmp/hive-hadoopqa
  ```

- **BUG-13551:** Oozie does not understand _HOST in the Kerberos principal name.

  Problem: Oozie currently expects the actual hostname in the kerberos principal. This is unlike other services in the stack, where we can just send _HOST and at run time the service replaces _HOST with machine hostname. This is important so that in a HA setup we can push the same configs to all Oozie servers.

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

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

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

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

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

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

  Problem: From the Oozie log:

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

But this job and the application complete successfully.

4.6.10. Known Issues for Flume

- **BUG-16771:** Hive table returns multiple copies of streaming data when testing the Hive Server restart scenario.

  Problem: When running the Hive restart test where the Hive metastore is bounced while Flume is treaming data to Hive, 3 duplicate copies were observed for each row in the Hive table. (Expected: 200 rows; observed: 800 rows, or 3 complete copies of the expected set of 200.)

4.6.11. Known Issues for Storm

- **BUG-16232:** Storm python support can use wrong version of python if supervisor host has more than one version of python installed.

  Problem: Storm requires the default system python interpreter to be version 2.6 or higher. Earlier versions of python can see this conflict.

  Workaround: Ensure that the default system python interpreter is version 2.6 or higher.

- **BUG-15960:** Worker node gets 'FileNotFoundException : stormconf.ser'.

  Problem: While running Storm-HDFS topologies in a secure environment, the following error was observed in the worker node:

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
 CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
 ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
 mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
 ~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
 ~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
 ~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
```

```
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$_main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
 initialization")
```

4.6.12. Known Issues for Falcon

- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

  Problem: Falcon generated hive-action does not pass the hive-site.xml with the right configuration parameters. One manifestation of the problem will be the failure in table import job where user "hive" will be used to write to a directory owned by the table owner. This is because hive.metastore.execute.setugi parameter is not being passed as part of the hive action.

  Workaround: Add a Hive default configuration to Oozie.

  Stop the Oozie service.

  

  ### Warning

  This change allows you to work with Hive tables and Oozie workflows, but will impact all Hive actions, including non-Falcon Oozie workflows.

  Under the oozie configuration directory (typically `/etc/oozie/conf`), there will be a subdirectory called `action-conf`. Under that directory, either create or modify the file hive-site.xml and add the following:

  ```
  <property>
    <name>hive.metastore.execute.setugi</name>
    <value>true</value>
  </property>
  ```

  After making this change restart the Oozie service. If Oozie is configured for HA, perform this configuration change on all Oozie server nodes.

- **BUG-16290:** Strange delegation token issues in secure clusters

  Problem: Inconsistencies in rules for `hadoop.security.auth_to_local` can lead to issues with delgation token renewals in secure clusters.

  Workaround: Verify that `hadoop.security.auth_to_local` in core-site.xml is consistent across all clusters.

- **BUG-16290, FALCON-389:** Oozie config changes needed to support HCat replication in Falcon

Problem: Oozie config changes are needed before Falcon can handle HCat replication.

Workaround: Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.

2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to /etc/hadoop/conf-1 and the other to /etc/hadoop/conf-2.

3. For each oozie-site.xml file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

   For example, if Falcon connects to 3 clusters, specify:

```
<property>
      <name>oozie.service.HadoopAccessorService.hadoop.configurations</
name>
      <value>*=/etc/hadoop/
conf,$NameNode:$rpcPortNN=$hadoopConfDir1,$ResourceManager1:$rpcPortRM=$hadoopConfDir1,$
 :$rpcPortNN =$hadoopConfDir3,$ResourceManager3 :$rpcPortRM
 =$hadoopConfDir3</value>
      <description>
          Comma separated AUTHORITY=HADOOP_CONF_DIR, where AUTHORITY is
 the HOST:PORT of
          the Hadoop service (JobTracker, HDFS). The wildcard '*'
 configuration is
          used when there is no exact match for an authority. The
 HADOOP_CONF_DIR contains
          the relevant Hadoop *-site.xml files. If the path is relative is
 looked within
          the Oozie configuration directory; though the path can be
 absolute (i.e. to point
          to Hadoop client conf/ directories in the local filesystem.
      </description>
   </property>
```

4. Restart the Oozie service on all clusters.

4.6.13. Known Issues for Knox

- **BUG-14461:** Knox has Kerberos config at the global level rather than cluster topology level.

  Problem: The gateway.hadoop.kerberos.secured=false Boolean flag indicates whether the Hadoop cluster protected by Gateway is secured with Kerberos in gateway-site.xml.

- **BUG-16592:** Oozie only supported on HDP for Linux multi-node clusters

  Problem: When accessing a Hadoop cluster through an Apache Knox Gateway, Oozie is not supported for HDP for Windows or HDP for Linux single-node clusters.

  Workaround: Only use Oozie when access a multinode HDP for Linux cluster.

4.7. Deprecated Features

- Oracle JDK 6 is Deprecated in this release.

4.8. Third-party Licenses

Table 4.1. Third-party Licenses

HDP Component	License
Phoenix	EPL
Storm	MIT License
Storm	LGPL (org.jgrapht-core 0.9.0 jar)
Storm	EPL
Falcon	CERN
Falcon	Tinkerpop
Knox	ANTLR
Knox	MIT
Knox	EPL
Knox	Bouncy Castle
Knox	OWS

5. Release Notes HDP-2.1.0.0-Win

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

5.1. Product Version: HDP-2.1.0.0-Win

All HDP 2.1 components listed here are official Apache releases of the most recent stable versions available. Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.1 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.1.

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

- Apache Hadoop 2.4.0
- Apache HBase 0.98.0
- Apache Zookeeper 3.4.5
- Apache Pig 0.12.1
- Apache Hive 0.13.0

Note

Apache HCatalog is now merged with Apache Hive.

- Apache Sqoop 1.4.4
- Apache Oozie 4.0.0
- Apache Flume 1.4.0
- Apache Mahout 0.9.0

5.2. 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 propeties configuration for an HDP Windows MSI install
- Added Namenode HA support (Hadoop 2)
- Enhanced HCatalog support to allow implementations of direct mode connection manager

5.3. Patch Information

In this section:

5.3.1. 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 `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-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: TesMRAMWithNonNormalizedCapabilities 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 snapshop 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

5.3.2. 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 prvent 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

5.3.3. Patch Information for Zookeeper

Zookeper 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

5.3.4. 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

5.3.5. 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 Kyro 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 IllegalArguementException
- 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 overriden 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

5.3.6. Patch Information for HCatalog

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

5.3.7. 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

5.3.8. 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: LobAvrolImportTestCase is sensitive to test method order execution

5.3.9. 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`

5.4. 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

5.5. 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 Micosoft 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

- 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 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 wihin 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