Hortonworks Data Platform

HDP-2.2.9 Release Notes

(November 18, 2015)

docs.cloudera.com

Hortonworks Data Platform: HDP-2.2.9 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

1. HDP 2.2.9 Release Notes

The official Apache versions of most HDP 2.2 components are unchanged from HDP 2.2.0.0, with the exception of Spark. Spark is upgraded from 1.2.1 to 1.3.1. See more details of Spark 1.3.1 in the New Features section. All HDP 2.2 components listed here are official Apache releases of the most recent stable versions available.

Hortonworks' philosophy is to provide patches only when absolutely necessary to assure the interoperability of the components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP 2.2 components needs to remain at the following package version levels to ensure a certified and supported copy of HDP 2.2.

- Apache Accumulo 1.6.1
- Apache DataFu 1.2.0
- Apache Falcon 0.6.0
- Apache Flume 1.5.2
- Apache Hadoop 2.6.0
- Apache HBase 0.98.4
- Apache Hive 0.14.0
- Apache Kafka 0.8.1.1
- Apache Knox 0.5.0
- Apache Mahout 0.9.0
- Apache Oozie 4.1.0
- Apache Phoenix 4.2.0
- Apache Pig 0.14.0
- Apache Ranger 0.4.0
- Apache Slider 0.61.0
- Apache Spark 1.3.1
- Apache Sqoop 1.4.5
- Apache Storm 0.9.3
- Apache Tez 0.5.2
- Apache ZooKeeper 3.4.6

- Hue 2.6.1

As of HDP 2.2, the following third-party tools are deprecated, and will be removed in a future release:

- Ganglia 3.5.0
- Nagios 3.5.0
- Ganglia Web 3.5.7

1.1. New Features

This section highlights the only new feature in HDP 2.2.9.

Spark 1.3.1 is the default Spark version with HDP 2.2.9. With a new HDP 2.2.9 cluster install, Spark 1.3.1 is installed. With the upgrade of an existing HDP cluster to 2.2.9 using Ambari, the Spark component is automatically upgraded from 1.2.1 to 1.3.1. For clusters not managed with Ambari, Spark can be manually upgraded from 1.2.1 to Spark 1.3.1; see the Optional Spark Manual Upgrade Procedure for instructions.

If you have upgraded to Spark 1.3.1 and want to downgrade to Spark 1.2.1, follow the Optional Spark Manual Downgrade Procedure. The downgrade to Spark 1.2.1 is only available as a manual step.

- In HDP 2.2.8, Spark 1.3.1 became available as a certified component.
  - Spark Core
  - Spark on YARN
  - Spark History Server
  - Spark on YARN on Kerberos-enabled clusters
  - Spark MLLib
  - Support for Hive 0.13.1, including the collect_list UDF
- SmartSense 1.1
  - HST Gateway
- YARN
  - ATS (Application Timeline Server) 1.5

1.2. Unsupported Features

Some features exist within HDP 2.2.9, but Hortonworks does not currently support these specific capabilities.

1.2.1. Technical Preview Features

The following features are available within HDP 2.2.9, but are not ready for production deployment. We encourage you to explore these technical preview features in non-production environments and provide feedback on your experiences through the Hortonworks Community Forums.

Component	Feature
HDFS	• HDFS Transparent Data at Rest Encryption (HDFS-6134)
Phoenix	• Phoenix support for local secondary indexes
YARN	• YARN support for Docker
Spark	• Spark SQL • Spark Streaming • Spark DataFrame API • ORC file support • SparkSQL thrift JDBC/ODBC Server • Dynamic Executor Allocation

1.2.2. Community Features

The following features are developed and tested by the community, but are not officially supported by Hortonworks. There are a variety of reasons that these features are excluded, including: insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, feature deviates from Hortonworks best practices, and more. Do not use them in your production environments.

Component	Feature
HDFS	• NameNode Federation (HDFS-1052) • viewFS (HADOOP-7257) • block-volume device choosing (HDFS-1804)
Falcon	• Prism Server
YARN	• Fair Scheduler • MapReduce Uber AM • MapReduce Eclipse Plug-in
Spark	• Spark Standalone • BlinkDB • GraphX

1.3. Upgrading from HDP 2.2.x to HDP 2.2.9

HDP 2.2.9 is a maintenance release of HDP 2.2.x.

If you already have HDP 2.2.x installed, upgrading your cluster to HDP 2.2.9 means:

- Keeping the same configuration files you used for HDP 2.2.x
- Keeping the same data and metadata in the same location you used for HDP 2.2.x
- Installing any new components (added for the first time in HDP 2.2.9) side-by-side with existing components

The following table summarizes HDP 2.2.x-to-2.2.9 upgrade options:

Cluster Management	Supporting Doc	Notes
Cluster managed manually	Before you begin [4]	If you have an earlier version of HDP (such as HDP 2.0 or HDP 2.1), see the HDP 2.2.9 Manual Upgrade Guide
Cluster managed via Ambari 1.7.0	Before you begin [4]	
Cluster managed via Ambari 2.0	Use the Upgrading Ambari Guide	Ambari 2.0 supports rolling upgrade between HDP 2.2.x and HDP 2.2.9 When upgrading to HDP 2.2.9 using Ambari, Spark 1.2.1 will be automatically upgraded to 1.3.1. If you wish to return to using 1.2.1, use the Spark Manual Downgrade Procedure.
Cluster managed via Ambari 2.1	Use the Upgrading Ambari Guide	Ambari 2.1 supports rolling upgrade between HDP 2.2.x and HDP 2.2.9 When upgrading to HDP 2.2.9 using Ambari, Spark 1.2.1 will be automatically upgraded to 1.3.1. If you wish to return to using 1.2.1, use the Spark Manual Downgrade Procedure.

1.3.1. Before you begin

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation
- Think about whether you are going to upgrade using a local repository or a remote repository

1.3.2. Optional: Spark Manual Upgrade Procedure

Spark 1.3.1 is newly available as of HDP 2.2.8. You can choose to upgrade Spark from 1.2.1 to 1.3.1, or continue using Spark 1.2.1.

If you wish to upgrade to Spark 1.3.1, perform the following steps as root:

1. Install Spark:

   a. Enter `wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.9.0/hdp.repo -O /etc/yum.repos.d/hdp229.repo`.

   b. Enter `yum search spark`.

The prompt returns:

```
spark.noarch : spark Distro virtual package
spark-master.noarch : spark-master Distro virtual package
spark-python.noarch : spark-python Distro virtual package
spark-worker.noarch : spark-worker Distro virtual package
spark_2_2_9_0_3393-master.noarch : Server for Spark master
spark_2_2_9_0_3393-python.noarch : Python client for Spark
spark_2_2_9_0_3393-worker.noarch : Server for Spark worker
spark_2_2_9_0_3393-master.noarch : Server for Spark master
spark_2_2_9_0_3393-python.noarch : Python client for Spark
spark_2_2_9_0_3393-worker.noarch : Server for Spark worker
spark_2_2_9_0_3393.noarch : Lightning-Fast Cluster Computing
spark_2_2_9_0_3393.noarch : Lightning-Fast Cluster Computing
```

c. To install Spark 1.3.1, select spark_2_2_9_0_3393: `yum install spark_2_2_9_0_3393-master`.

d. Enter `hdp-select set spark-client 2.2.9.0-3393`.

e. Enter `hdp-select set spark-historyserver 2.2.9.0-3393`.

2. Validate the Spark installation:

a. Enter `su spark`

b. Enter `cd /usr/hdp/current/spark-client`.

c. Enter `./bin/spark-submit --class org.apache.spark.examples.SparkPi --master yarn-client --num-executors 3 --driver-memory 512m --executor-memory 512m --executor-cores 1 lib/spark-examples*.jar 10`.

3. *Optional:* Run Spark History Server:

a. Enter `su spark`.

b. Enter `cd /usr/hdp/current/spark-historyserver/`.

c. Enter `./sbin/start-history-server.sh`.

1.3.3. Upgrade Procedure

To upgrade your cluster from HDP 2.2.x to HDP 2.2.9:

1. **Download the appropriate HDP 2.2.9 hdp.repo file for your OS:**

Operating System	Repository Location
RHEL/CentOS/Oracle LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.2.9.0/hdp.repo
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.9.0/hdp.repo
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/sles11sp1/2.x/updates/2.2.9.0/hdp.repo

Operating System	Repository Location
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.2.9.0/hdp.repo
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.9.0/hdp.list
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.2.9.0/hdp.list

or

Download the HDP RPM single repository tarball. (For information on how to install the repositories, see the local repository instructions.)

Operating System	Tarball Location
RHEL/CentOS/Oracle LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.2.9.0/HDP-2.2.9.0-centos5-rpm.tar.gz
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.9.0/HDP-2.2.9.0-centos6-rpm.tar.gz
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/sles11sp1/2.x/updates/2.2.9.0/HDP-2.2.9.0-sles11sp1-rpm.tar.gz
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.2.9.0/HDP-2.2.9.0-suse11sp3-rpm.tar.gz
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.9.0/HDP-2.2.9.0-ubuntu12-rpm.tar.gz
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.2.9.0/HDP-2.2.9.0-debian6-rpm.tar.gz

2. **Run an update:**

```
apt-get update
```

3. **Install the HDP 2.2.9 bits:**

Operating System	Commands
RHEL/CentOS/Oracle LINUX	Install HDP 2.2.9 components on relevant nodes, according to the services that run on those hosts: yum install "hadoop_2_2_9_0_3393*" "oozie_2_2_9_0_3393*" "pig_2_2_9_0_3393*" "sqoop_2_2_9_0_3393*" "zookeeper_2_2_9_0_3393*" "hbase_2_2_9_0_3393*" "hive_2_2_9_0_3393*" "tez_2_2_9_0_3393*" "storm_2_2_9_0_3393*" "falcon_2_2_9_0_3393*" "flume_2_2_9_0_3393*" "phoenix_2_2_9_0_3393*" "accumulo_2_2_9_0_3393*" "mahout_2_2_9_0_3393*"
SLES	Install HDP 2.2.9 components on relevant nodes, according to the services that run on those hosts: zypper install "hadoop_2_2_9_0_3393*" "oozie_2_2_9_0_3393*" "pig_2_2_9_0_3393*" "sqoop_2_2_9_0_3393*" "zookeeper_2_2_9_0_3393*" "hbase_2_2_9_0_3393*" "hive_2_2_9_0_3393*" "tez_2_2_9_0_3393*" "storm_2_2_9_0_3393*"

Operating System	Commands
	"falcon_2_2_9_0_3393*" "flume_2_2_9_0_3393*" "phoenix_2_2_9_0_3393*" "accumulo_2_2_9_0_3393*" "mahout_2_2_9_0_3393*"
Ubuntu/Debian	Install HDP 2.2.9 components on relevant nodes, according to the services that run on those hosts: apt-get install "hadoop_2_2_9_0_3393*" "oozie_2_2_9_0_3393*" "pig_2_2_9_0_3393*" "sqoop_2_2_9_0_3393*" "zookeeper_2_2_9_0_3393*" "hbase_2_2_9_0_3393*" "hive_2_2_9_0_3393*" "tez_2_2_9_0_3393*" "storm_2_2_9_0_3393*" "falcon_2_2_9_0_3393*" "flume_2_2_9_0_3393*" "phoenix_2_2_9_0_3393*" "accumulo_2_2_9_0_3393*" "mahout_2_2_9_0_3393*"

4. **Stop all HDP 2.2.x Services.**

   If you are managing your cluster manually, stop all HDP 2.2.x Services. See "Controlling HDP Services Manually" in the *HDP Reference Guide*.

   If you are managing your cluster with Ambari 1.7.0, do the following:

   a. Open Ambari Web

   b. Browse to **Services**

   c. Use **Service Actions** to stop each service

   For all services, switch the active version to HDP 2.2.9.

   On each host in the cluster, use hdp-select to switch all services to the HDP 2.2.9 version:

   ```
   hdp-select set all <hdp2.2.9 version>
   ```

5. **(Ambari 1.7.0-managed clusters only) Complete the Stack Upgrade.**

   If you are managing your cluster with Ambari 1.7.0, update the repository Base URLs to use the HDP 2.2.9 repositories for HDP and HDP-UTILS:

   a. Open Ambari Web

   b. Browse to **Admin > Repositories**

   c. Edit the Base URLs

6. **Start all HDP 2.2.9 services, in the following order:**

   a. **ZooKeeper**

   ```
   su - zookeeper -c "export ZOOCFGDIR=/usr/hdp/current/
   zookeeper-server/conf ; export ZOOCFG=zoo.cfg; source /usr/
   hdp/current/zookeeper-server/conf/zookeeper-env.sh ; /usr/hdp/
   current/zookeeper-server/bin/zkServer.sh start"
   ```

   b. **(HA NameNode upgrade only) ZooKeeper Failover Controller Daemons**

```
/usr/hdp/current/hadoop-hdfs-namenode/../hadoop/sbin/hadoop-
daemon.sh start zkfc
```

c. **(HA NameNode upgrade only) JournalNodes**

```
su - hdfs /usr/hdp/current/hadoop-hdfs-journalnode/../hadoop/
sbin/hadoop-daemon.sh start journalnode
```

d. **HDFS NameNode(s)**

Start the HDFS NameNode(s). Because there is no metadata schema update for this upgrade, start the NameNode(s) in normal mode:

```
su - hdfs /usr/hdp/current/hadoop-hdfs-namenode/../hadoop/
sbin/hadoop-daemon.sh start namenode
```

e. **Remaining Services**

Start the rest of the HDP services. On each host in the cluster, start the services that are relevant to that cluster. To identify the start commands for all services, see "Controlling HDP Services Manually" in the *HDP Reference Guide*.

You now have an upgraded cluster. Ensure that your workloads run correctly on this upgraded cluster.

1.3.4. Optional: Spark Manual Downgrade Procedure

When upgrading to HDP 2.2.9 using Ambari, Spark 1.2.1 is automatically upgraded to 1.3.1.

Note

- Spark downgrade is only available manually, and not from Ambari.
- Downgraded Spark is not managed from Ambari.

However, if you wish to return to using 1.2.1:

1. Stop all services related to Spark.
2. Remove Spark 1.3.1 from your HDP cluster using Ambari:

   curl -u admin:admin -H "X-Requested-By: ambari" -X DELETE

   http://<AMBARI_HOST>:8080/api/v1/clusters/<CLUSTER_NAME>/services/SPARK

3. Manually install Spark 1.2.1 with HDP 2.2.6 Installing HDP Manually: Installing and Configuring Apache Spark.

1.4. Behavior Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.2.9, behavioral changes affect the following Hadoop components.

Hortonworks Bug ID	Component	Summary	Details
BUG-40345	Hue	Hue now shows log messages for create and insert statements.	Fixed data fetch mechanism to get logs from HiveServer2.
BUG-43069	Hue	Hue File browser does not respect dfs.umaskmode, fs.permissions.umask-mode when creating files or folders.	Hue File browser previously used umask 022 when creating files and folders. In this release, the octal value of umask in hue.ini specifies the umask value to apply when creating files and folders through the Hue File Browser.
BUG-44184	HDFS	When HDFS metadata is replaced, namenode enters safe mode.	**Scenario**: A user loads an old Fsimage and starts Namenode. This Fsimage is missing the latest transactions that recently occurred on the cluster. **Previous Behavior**: When a user starts the Namenode with this old image, the Namenode will come out of safemode as long as the amount of blocks reported reaches the set threshold (0.999 is default for `dfs.namenode.safemode.threshold-pct`). When the Namenode learns about a block without a filename, it instructs the Datanode to delete those blocks, resulting in lost data. **New Behavior**: The new behavior is that when the Namenode detects such a scenario where there are blocks in the cluster which are tagged with future generation stamps, the Namenode will remain in the safemode state to prevent the blocks from being deleted. If the user manually executes `hdfs dfsadmin –safemode leave`, the Namenode will refuse to leave safemode. At this point, the user has an opportunity to shut down the Namenode and load the most current FSimage. If the user for some reason is not able to find the latest Fsimage, the last resort to get the cluster up and running with minimal data loss is to force the Namenode out of safemode by running the command `hdfs dfsadmin –safemode forceExit`

Hortonworks Bug ID	Component	Summary	Details
			after the Namenode has been started.
BUG-44387	YARN	Disable container stats publish to ATS by default	Changed the value of `yarn.timeline-service.generic-application-history.save-non-am-container-meta-info` to false. This config defines if YARN will save container meta-info in the application history store.
BUG-44447	HDFS	The feature introduces a new setting `hadoop.caller.context.enabled`. When set to additional fields are written into namenode audit log records to help identify the job or query that introduced each NameNode operation. This feature is **enabled** by default starting with this release of HDP.	This feature brings a new key-value pair at the end of each audit log record. The newly added key at is `callerContext`, value `context:signature`. The overall format would be `callerContext=context:signature`. If the signature is *null* or *empty*, the value will be context only, in the format of `callerContext=context`. If the `hadoop.caller.context.enabled` config key is false, the key-value pair will not be showing. The audit log format is not changed in this case. It is also possible to limit the maximum length of context and signature. Consider the `hadoop.caller.context.max.size` config key (default 128 bytes) and `hadoop.caller.context.signature.max.s` (default 40 bytes) config key respectively. There is a chance that the new information in the audit log may break existing scripts/automation that was being used to analyze the audit log. In this case the scripts may need to be fixed. We do not recommend disabling this feature as it can be a useful troubleshooting aid.
BUG-44523	Hue	Hue now displays Hive tables in alphabetical order	In Hue Beeswax/Hive UI tables were previously displayed in chronological order (the order in which tables were added). The change in this release is that tables are sorted by alphabetical order by default.

1.5. Apache Patch Information

The following sections list patches in each HDP 2.2.9 component beyond what was fixed in the base version of the Apache component.

1.5.1. Accumulo 1.6.1

HDP 2.2.9 provides Accumulo 1.6.1 and the following Apache patches:

- ACCUMULO-3967: Bulk import loses records when loading pre-split table.

HDP 2.2.8 provided Accumulo 1.6.1 and the following Apache patches:

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

1.5.2. Falcon 0.6.0

HDP 2.2.8 provided Apache Falcon 0.6.0 and the following additional Apache patches:

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

1.5.3. Hadoop Common/HDFS 2.6.0

HDP 2.2.9 provides Apache Hadoop Core 2.6.0 and the following additional Apache patches:

- HADOOP-10728: Metrics system for Windows Azure Storage Filesystem.
- HADOOP-10809: HADOOP azure: page blob support.
- HADOOP-10839: Add unregisterSource() to MetricsSystem API.
- HADOOP-11000: HAServiceProtocol's health state is incorrectly transitioned to SERVICE_NOT_RESPONDING.
- HADOOP-11032: Replace use of Guava's Stopwatch with Hadoop's StopWatch.
- HADOOP-11188: HADOOP azure: automatically expand page blobs when they become full.
- HADOOP-11291: Log the cause of SASL connection failures.
- HADOOP-11321: copyToLocal cannot save a file to an SMB share unless the user has Full Control permissions.
- HADOOP-11333: Fix deadlock in DomainSocketWatcher when the notification pipe is full.

- HADOOP-11390: Metrics 2 ganglia provider to include hostname in unresolved address problems.
- HADOOP-11441: HADOOP azure: Change few methods scope to public.
- HADOOP-11442: HADOOP azure: Create test jar.
- HADOOP-11490: Expose truncate API via FileSystem and shell command.
- HADOOP-11509: change parsing sequence in GenericOptionsParser to parse -D parameters before -files.
- HADOOP-11595: Add default implementation for AbstractFileSystem#truncate.
- HADOOP-11642: Upgrade azure SDK version from 0.6.0 to 2.0.0.
- HADOOP-11648: Set DomainSocketWatcher thread name explicitly.
- HADOOP-11685: StorageException complaining " no lease ID" during HBase distributed log splitting.
- HADOOP-11730: Regression: s3n read failure recovery broken.
- HADOOP-11918: Listing an empty s3a root directory throws FileNotFound.
- HADOOP-11960: Enable Azure-Storage Client Side logging.
- HADOOP-12089: StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- HADOOP-12186: ActiveStandbyElector shouldn't call monitorLockNodeAsync multiple times.
- HADOOP-12239: StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- HADOOP-12318: Expose underlying LDAP exceptions in SaslPlainServer.
- HADOOP-12324: Better exception reporting in SaslPlainServer.
- HADOOP-12325: RPC Metrics : Add the ability track and log slow RPCs.
- HADOOP-12334: Change Mode Of Copy Operation of HBase WAL Archiving to bypass Azure Storage Throttling after retries.
- HADOOP-12350: WASB Logging: Improve WASB Logging around deletes, reads and writes.
- HADOOP-12358: Add -safely flag to rm to prompt when deleting many files.
- HADOOP-12407: Test failing: hadoop.ipc.TestSaslRPC.
- HADOOP-12413: AccessControlList should avoid calling getGroupNames in isUserInList with empty groups.
- HADOOP-12437: Allow SecurityUtil to lookup alternate hostnames.

- HADOOP-12438: TestLocalFileSystem tests can fail on Windows after HDFS-8767 fix for handling pipe.
- HADOOP-12542: TestDNS fails on Windows after HADOOP-12437.
- HADOOP-7984: Add hadoop –loglevel option to change log level.
- HADOOP-9629: Support Windows Azure Storage - Blob as a file system in Hadoop.
- HDFS-4015: Safemode should count and report orphaned blocks.
- HDFS-4366: Block Replication Policy Implementation May Skip Higher-Priority Blocks for Lower-Priority Blocks.
- HDFS-4396: Add START_MSG/SHUTDOWN_MSG for ZKFC.
- HDFS-4660: Block corruption can happen during pipeline recovery.
- HDFS-4937: ReplicationMonitor can infinite-loop in BlockPlacementPolicyDefault#chooseRandom.
- HDFS-5782: BlockListAsLongs should take lists of Replicas rather than concrete class.
- HDFS-6481: DatanodeManager#getDatanodeStorageInfos() should check the length of storageIDs.
- HDFS-6581: Support for writing to single replica in RAM.
- HDFS-6663: Admin command to track file and locations from block id.
- HDFS-6860: BlockStateChange logs are too noisy.
- HDFS-6917: Add an hdfs debug command to validate blocks, call recoverlease, etc.
- HDFS-6945: BlockManager should remove a block from excessReplicateMap and decrement ExcessBlocks metric when the block is removed.
- HDFS-6982: nntop: top–like tool for name node users.
- HDFS-7009: Active NN and standby NN have different live nodes.
- HDFS-7097: Allow block reports to be processed during checkpointing on standby name node.
- HDFS-7153: Add storagePolicy to NN edit log during file creation.
- HDFS-7182: JMX metrics aren't accessible when NN is busy.
- HDFS-7213: processIncrementalBlockReport performance degradation.
- HDFS-7222: Expose DataNode network errors as a metric.
- HDFS-7263: Snapshot read can reveal future bytes for appended files.
- HDFS-7278: Add a command that allows sysadmins to manually trigger full block reports from a DN .

- HDFS-7435: PB encoding of block reports is very inefficient.
- HDFS-7448: TestBookKeeperHACheckpoints fails in trunk build.
- HDFS-7491: Add incremental blockreport latency to DN metrics.
- HDFS-7575: Upgrade should generate a unique storage ID for each volume.
- HDFS-7579: Improve log reporting during block report rpc failure.
- HDFS-7596: NameNode should prune dead storages from storageMap.
- HDFS-7604: Track and display failed DataNode storage locations in NameNode.
- HDFS-7608: hdfs dfsclient newConnectedPeer has no write timeout.
- HDFS-7611: deleteSnapshot and delete of a file can leave orphaned blocks in the blocksMap on NameNode restart.
- HDFS-7634: Disallow truncation of Lazy persist files.
- HDFS-7643: Test case to ensure lazy persist files cannot be truncated.
- HDFS-7659: truncate should check negative value of the new length.
- HDFS-7704: DN heartbeat to Active NN may be blocked and expire if connection to Standby NN continues to time out.
- HDFS-7707: Edit log corruption due to delayed block removal again.
- HDFS-7738: Revise the exception message for recover lease; add more truncate tests such as truncate with HA setup, negative tests, truncate with other operations and multiple truncates.
- HDFS-7858: Improve HA Namenode Failover detection on the client.
- HDFS-7928: Scanning blocks from disk during rolling upgrade startup takes a lot of time if disks are busy.
- HDFS-7931: DistributedFileSystem should not look for keyProvider in cache if Encryption is disable.
- HDFS-7933: fsck should also report decommissioning replicas.
- HDFS-7980: Incremental BlockReport will dramatically slow down the startup of a namenode.
- HDFS-8055: NullPointerException when topology script is missing.
- HDFS-8070: Pre-HDFS-7915 DFSClient cannot use short circuit on post-HDFS-7915 DataNode.
- HDFS-8127: NameNode Failover during HA upgrade can cause DataNode to finalize upgrade.
- HDFS-8155: Support OAuth2 in WebHDFS.

- HDFS-8163: Using monotonicNow for block report scheduling causes test failures on recently restarted systems.
- HDFS-8180: AbstractFileSystem Implementation for WebHdfs.
- HDFS-8219: setStoragePolicy with folder behavior is different after cluster restart.
- HDFS-8270: create() always retried with hardcoded timeout when file already exists with open lease.
- HDFS-8384: Allow NN to startup if there are files having a lease but are not under construction.
- HDFS-8435: Support CreateFlag in WebHdfs.
- HDFS-8542: WebHDFS getHomeDirectory behavior does not match specification.
- HDFS-8554: TestDatanodeLayoutUpgrade fails on Windows.
- HDFS-8576: Lease recovery should return true if the lease can be released and the file can be closed.
- HDFS-8722: Optimize datanode writes for small writes and flushes.
- HDFS-8785: TestDistributedFileSystem is failing in trunk.
- HDFS-8797: WebHdfsFileSystem creates too many connections for pread.
- HDFS-8826: In Balancer, add an option to specify the source node list so that balancer only selects blocks to move from those nodes.
- HDFS-8883: NameNode Metrics : Add FSNameSystem lock Queue Length.
- HDFS-8885: ByteRangeInputStream used in webhdfs does not override available().
- HDFS-8939: Test(S)WebHdfsFileContextMainOperations failing on branch-2.
- HDFS-8950: NameNode refresh doesn't remove DataNodes that are no longer in the allowed list.
- HDFS-8965: Harden edit log reading code against out of memory errors.
- HDFS-8995: Flaw in registration bookkeeping can make DN die on reconnect.
- HDFS-9009: Send metrics logs to NullAppender by default.
- HDFS-9082: Change the log level in WebHdfsFileSystem.initialize() from INFO to DEBUG.
- HDFS-9083: Replication violates block placement policy.
- HDFS-9106: Transfer failure during pipeline recovery causes permanent write failures.
- HDFS-9107: Prevent NN's unrecoverable death spiral after full GC.
- HDFS-9109: Adding informative message to sticky bit permission denied exception.

- HDFS-9112: Improve error message for Haadmin when multiple name service IDs are configured.
- HDFS-9128: TestWebHdfsFileContextMainOperations and TestSWebHdfsFileContextMainOperations fail due to invalid HDFS path on Windows.
- HDFS-9142: Separating Configuration object for namenode(s) in MiniDFSCluster.
- HDFS-9175: Change scope of 'AccessTokenProvider.getAccessToken()' and 'CredentialBasedAccessTokenProvider.getCredential()' abstract methods to public.
- HDFS-9178: Slow datanode I/O can cause a wrong node to be marked bad.
- HDFS-9184: Logging HDFS operation's caller context into audit log.
- HDFS-9205: Do not schedule corrupt blocks for replication.
- HDFS-9220: Reading small file (greater than 512 bytes) that is open for append fails due to incorrect checksum.
- HDFS-9273: ACLs on root directory may be lost after NN restart.
- HDFS-9305: Delayed heartbeat processing causes rapid storm of subsequent heartbeat messages.
- HDFS-9311: Support optional offload of NameNode HA service health checks to a separate RPC server.
- HDFS-9343: Empty caller context considered invalid.
- HDFS-9354: Fix TestBalancer#testBalancerWithZeroThreadsForMove on Windows.
- HDFS-9362: TestAuditLogger#testAuditLoggerWithCallContext assumes Unix line endings, fails on Windows.
- HDFS-9384: TestWebHdfsContentLength intermittently hangs and fails due to TCP conversation mismatch between client and server.
- MAPREDUCE-4815: Speed up FileOutputCommitter#commitJob for many output files.
- MAPREDUCE-6238: MR2 can't run local jobs with -libjars command options which is a regression from MR1.
- MAPREDUCE-6442: Stack trace is missing when error occurs in client protocol provider's constructor.
- YARN-1979: TestDirectoryCollection fails when the umask is unusual.
- YARN-1984: LeveldbTimelineStore does not handle db exceptions properly.
- YARN-2165: Timeline server should validate the numeric configuration values.
- YARN-2246: Made the proxy tracking URL always be http(s)://proxy addr:port/proxy/appId to avoid duplicate sections.
- YARN-2254: TestRMWebServicesAppsModification should run against both CS and FS.

- YARN-2513: Host framework UIs in YARN for use with the ATS.
- YARN-2605: RM HA Rest api endpoints doing redirect incorrectly.
- YARN-2816: NM fail to start with NPE during container recovery.
- YARN-2821: Distributed shell app master becomes unresponsive sometimes.
- YARN-2906: CapacitySchedulerPage shows HTML tags for a queue's Active Users.
- YARN-2922: ConcurrentModificationException in CapacityScheduler's LeafQueue.
- YARN-3197: Confusing log generated by CapacitySchedule.
- YARN-3207: secondary filter matches entities which do not have the key being filtered for.
- YARN-3227: Timeline renew delegation token fails when RM user's TGT is expire.
- YARN-3238: Connection timeouts to nodemanagers are retried at multiple levels.
- YARN-3267: Timelineserver applies the ACL rules after applying the limit on the number of records.
- YARN-3351: AppMaster tracking URL is broken in HA.
- YARN-3448: Add Rolling Time To Lives Level DB Plugin Capabilities.
- YARN-3469: ZKRMStateStore: Avoid setting watches that are not required.
- YARN-3487: CapacityScheduler scheduler lock obtained unnecessarily when calling getQueue.
- YARN-3526: ApplicationMaster tracking URL is incorrectly redirected on a QJM cluster.
- YARN-3530: ATS throws exception on trying to filter results without otherinfo.
- YARN-3654: ContainerLogsPage web UI should not have meta-refresh.
- YARN-3700: Made generic history service load a number of latest applications according to the parameter or the configuration.
- YARN-3766: Fixed the apps table column error of generic history web UI.
- YARN-3787: Allowed generic history service to load a number of applications whose started time is within the given range.
- YARN-3804: Both RM are on standBy state when kerberos user not in yarn.admin.acl.
- YARN-3978: Configurably turn off the saving of container info in Generic AHS.
- YARN-4087: Followup fixes after YARN-2019 regarding RM behavior when state- store error occurs.
- YARN-4105: Capacity Scheduler headroom for DRF is wrong.

- YARN-4243: Add retry on establishing ZooKeeper connection in EmbeddedElectorService#serviceInit.
- YARN-4313: Race condition in MiniMRYarnCluster when getting history server address.

HDP 2.2.8 provided Apache Hadoop Core 2.6.0 and the following additional Apache patches:

- HADOOP-11321: copyToLocal cannot save a file to an SMB share unless the user has Full Control permissions.
- HADOOP-11368: Fix SSLFactory truststore reloader thread leak in KMSClientProvider.
- HADOOP-11381: Fix findbugs warnings in hadoop-distcp, hadoop-aws, HADOOP azure, and hadoop-openstack
- HADOOP-11412: POMs mention "The Apache Software License" rather than "Apache License".
- HADOOP-11490: Expose truncate API via FileSystem and shell command.
- HADOOP-11509: change parsing sequence in GenericOptionsParser to parse -D parameters before -files.
- HADOOP-11510: Expose truncate API via FileContext.
- HADOOP-11523: StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.
- HADOOP-11579: Documentation for truncate.
- HADOOP-11595: Add default implementation for AbstractFileSystem#truncate.
- HADOOP-926: Do not fail job history iteration when encountering missing directories.
- HADOOP-941: Addendum patch.
- HDFS-3107: Introduce truncate.
- HDFS-7009: Active NN and standby NN have different live nodes.
- HDFS-7056: Snapshot support for truncate.
- HDFS-7058: Tests for truncate CLI
- HDFS-7263: Snapshot read can reveal future bytes for appended files.
- HDFS-7425: NameNode block deletion logging uses incorrect appender.
- HDFS-7443: Datanode upgrade to BLOCKID_BASED_LAYOUT fails if duplicate block files are present in the same volume
- HDFS-7470: SecondaryNameNode need twice memory when calling reloadFromImageFile.
- HDFS-7489: Incorrect locking in FsVolumeList#checkDirs can hang datanodes.

- HDFS-7503: Namenode restart after large deletions can cause slow processReport
- HDFS-7606: Fix potential NPE in INodeFile.getBlocks().
- HDFS-7634: Disallow truncation of Lazy persist files.
- HDFS-7638: Small fix and few refinements for FSN#truncate.
- HDFS-7643: Test case to ensure lazy persist files cannot be truncated.
- HDFS-7655: Expose truncate API for Web HDFS.
- HDFS-7659: Truncate should check negative value of the new length
- HDFS-7676: Fix TestFileTruncate to avoid bug of HDFS-7611.
- HDFS-7677: DistributedFileSystem#truncate should resolve symlinks.
- HDFS-7707: Edit log corruption due to delayed block removal again.
- HDFS-7714: Simultaneous restart of HA NameNodes and DataNode can cause DataNode to register successfully with only one NameNode.
- HDFS-7733: NFS: readdir/readdirplus return null directory attribute on failure.
- HDFS-7738: Revise the exception message for recover lease; add more truncate tests such as truncate with HA setup, negative tests, truncate with other operations and multiple truncates.
- HDFS-7760: Document truncate for WebHDFS.
- HDFS-7831 Fix the starting index and end condition of the loop in FileDiffList.findEarlierSnapshotBlocks().
- HDFS-7843: A truncated file is corrupted after rollback from a rolling upgrade.
- HDFS-7885: Datanode should not trust the generation stamp provided by client.
- MAPREDUCE-6230: Fixed RMContainerAllocator to update the new AMRMToken service name properly.
- YARN-2246: Made the proxy tracking URL always be http(s)://proxy addr:port/proxy/<appId> to avoid duplicate sections.
- YARN-2571: RM to support YARN registry
- YARN-2683: registry config options: document and move to core-default
- YARN-2837: Support TimeLine server to recover delegation token when restarting.
- YARN-2917: Fixed potential deadlock when system.exit is called in AsyncDispatcher
- YARN-2964: RM prematurely cancels tokens for jobs that submit jobs (oozie).
- YARN-3103: AMRMClientImpl does not update AMRM token properly.

- YARN-3207: Secondary filter matches entities which do not have the key being filtered for.
- YARN-3227: Timeline renew delegation token fails when RM user's TGT is expired.
- YARN-3239: WebAppProxy does not support a final tracking URL which has query fragments and params.
- YARN-3251: Fixed a deadlock in CapacityScheduler when computing absoluteMaxAvailableCapacity in LeafQueue.
- YARN-3269: Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.
- YARN-570: Time strings are formatted in different timezone.

1.5.4. HBase 0.98.4

HDP 2.2.9 provides HBase 0.98.4 and the following additional Apache patches:

- HBASE-12128: Cache configuration and RpcController selection for Table in Connection
- HBASE-12366: Add login code as AuthUtil
- HBASE-12464: Meta table region assignment stuck in the FAILED_OPEN state due to region server not fully ready to serve
- HBASE-13825: Use ProtobufUtil#mergeFrom and ProtobufUtil#mergeDelimitedFrom in place of builder methods of same name
- HBASE-14207: Region was hijacked and remained in transition when RS failed to open a region and later regionplan changed to new RS on retry
- HBASE-14280: Bulk Upload from HA cluster to remote HA HBase cluster fails
- HBASE-14309: Allow load balancer to operate when there is region in transition by adding force flag
- HBASE-14361: ReplicationSink should create Connection instances lazily
- HBASE-14475: Region split requests are always audited with "HBase" user rather than request user
- HBASE-14497: Reverse Scan threw StackOverflow caused by readPt checking
- HBASE-14581: Znode cleanup throws auth exception in secure mode
- HBASE-14594: Use new DNS API introduced in HADOOP-12437
- HBASE-14621: ReplicationLogCleaner stuck on RS crash
- HBASE-14680: Two configs for snapshot timeout and better defaults

HDP 2.2.8 provided HBase 0.98.4 and the following additional Apache patches:

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

1.5.5. Hive 0.14.0

HDP 2.2.9 provides Hive 0.14.0 and the following additional Apache patches:

- HIVE-10083: SMBJoin fails in case one table is uninitialized.
- HIVE-10178: DateWritable incorrectly calculates daysSinceEpoch for negative Unix time.
- HIVE-10209: FetchTask with VC may fail because ExecMapper.done is true.
- HIVE-10213: MapReduce jobs using dynamic-partitioning fail on commit.
- HIVE-10244: Vectorization : TPC-DS Q80 fails with java.lang.ClassCastException when hive.vectorized.execution.reduce.enabled is enabled.
- HIVE-10303: HIVE-9471 broke forward compatibility of ORC files.
- HIVE-10451: PTF deserializer fails if values are not used in reducer.
- HIVE-10528: Hiveserver2 in HTTP mode is not applying auth_to_local rules.
- HIVE-10569: Hive CLI gets stuck when hive.exec.parallel=true; and some exception happens during SessionState.start.
- HIVE-10571: HiveMetaStoreClient should close existing thrift connection before its reconnect.
- HIVE-10646: ColumnValue does not handle NULL_TYPE.

- HIVE-10651: ORC file footer cache should be bounded.
- HIVE-10732: Hive JDBC driver does not close operation for metadata queries.
- HIVE-10746: Hive 1.2.0+Tez produces 1-byte FileSplits from mapred.TextInputFormat.
- HIVE-10790: orc write on viewFS throws exception.
- HIVE-10835: Concurrency issues in JDBC driver.
- HIVE-10907: Hive on Tez: Classcast exception in some cases with SMB joins.
- HIVE-10925: Non-static threadlocals in metastore code can potentially cause memory leak.
- HIVE-10965: direct SQL for stats fails in 0-column case.
- HIVE-10966: direct SQL for stats has a cast exception on some databases.
- HIVE-10968: Windows: analyze json table via beeline failed throwing Class org.apache.hive.hcatalog.data.JsonSerDe not found.
- HIVE-10971: count(*) with count(distinct) gives wrong results when hive.groupby.skewindata=true.
- HIVE-11013: MiniTez tez_join_hash test on the branch fails with NPE (initializeOp not called?).
- HIVE-11024: Error inserting a date value via parameter marker (PreparedStatement.setDate).
- HIVE-11031: ORC concatenation of old files can fail while merging column statistics.
- HIVE-11054: Read error : Partition Varchar column cannot be cast to string.
- HIVE-11123: Fix how to confirm the RDBMS product name at Metastore.
- HIVE-11132: Queries using join and group by produce incorrect output when hive.auto.convert.join=false and hive.optimize.reducededuplication=true.
- HIVE-11149: Fix issue with sometimes HashMap in PerfLogger.java hangs.
- HIVE-11172: Vectorization wrong results for aggregate query with where clause without group by.
- HIVE-11176: Caused by: java.lang.ClassCastException: org.apache.hadoop.hive.serde2.lazybinary.LazyBinaryStruct cannot be cast to [Ljava.lang.Object;.
- HIVE-11193: ConstantPropagateProcCtx should use a Set instead of a List to hold operators to be deleted.
- HIVE-11301: thrift metastore issue when getting stats results in disconnect.
- HIVE-11407: JDBC DatabaseMetaData.getTables with large no of tables call leads to HS2 OOM.

- HIVE-11408: HiveServer2 is leaking ClassLoaders when add jar / temporary functions are used due to constructor caching in Hadoop ReflectionUtils.
- HIVE-11432: Hive macro give same result for different arguments.
- HIVE-11499: Datanucleus leaks classloaders when used using embedded metastore with HiveServer2 with UDFs.
- HIVE-11502: Map side aggregation is extremely slow.
- HIVE-11510: Metatool updateLocation warning on views.
- HIVE-11517: Vectorized auto_smb_mapjoin_14.q produces different results.
- HIVE-11583: When PTF is used over a large partitions result could be corrupted.
- HIVE-11603: IndexOutOfBoundsException thrown when accessing a union all subquery and filtering on a column which does not exist in all underlying tables.
- HIVE-11605: Incorrect results with bucket map join in tez.
- HIVE-11606: Bucket map joins fail at hash table construction time.
- HIVE-11628: DB type detection code is failing on Oracle 12.
- HIVE-11670: Strip out password information from TezSessionState configuration.
- HIVE-11712: Duplicate groupby keys cause ClassCastException.
- HIVE-11720: Allow HiveServer2 to set custom http request/response header size.
- HIVE-11724: WebHcat get jobs to order jobs on time order with latest at top.
- HIVE-11727: Hive on Tez through Oozie: Some queries fail with fnf exception.
- HIVE-11737: IndexOutOfBounds compiling query with duplicated groupby keys.
- HIVE-11807: Set ORC buffer size in relation to set stripe size.
- HIVE-11835: Type decimal(1,1) reads 0.0, 0.00, etc from text file as NULL.
- HIVE-11839: Vectorization wrong results with filter of (CAST AS CHAR).
- HIVE-11849: NPE in HiveHBaseTableShapshotInputFormat in query with just count(*).
- HIVE-11901: StorageBasedAuthorizationProvider requires write permission on table for SELECT statements.
- HIVE-11950: WebHCat status file doesn't show UTF8 character.
- HIVE-11960: braces in join conditions are not supported.
- HIVE-11964: RelOptHiveTable.hiveColStatsMap might contain mismatched column stats.
- HIVE-11977: Hive should handle an external avro table with zero length files present.

- HIVE-11988: [hive] security issue with hive and ranger for import table command.
- HIVE-12012: select query on json table with map containing numeric values fails.
- HIVE-12057: ORC sarg is logged too much.
- HIVE-12076: WebHCat listing jobs after the given JobId even when templeton.jobs.listorder is set to lexicographicaldesc.
- HIVE-12083: HIVE-10965 introduces thrift error if partNames or colNames are empty.
- HIVE-12084: Hive queries with ORDER BY and large LIMIT fails with OutOfMemoryError Java heap space.
- HIVE-12201: Tez settings need to be shown in set -v output when execution engine is tez.
- HIVE-12204: Tez queries stopped running with ApplicationNotRunningException.
- HIVE-12206: ClassNotFound Exception during query compilation with Tez and Union query and GenericUDFs.
- HIVE-12223: Filter on Grouping__ID does not work properly.
- HIVE-12249: Improve logging with tez.
- HIVE-12254: Improve logging with yarn/hdfs.
- HIVE-12261: schematool version info exit status should depend on compatibility, not equality.
- HIVE-12262: Session log dir cannot be created in some cases.
- HIVE-12277: Hive macro results on macro_duplicate.q different after adding ORDER BY.
- HIVE-12340: ExecDriver.execute() unnecessarily sets METASTOREPWD to HIVE.
- HIVE-12344: Wrong types inferred for SemiJoin generation in CBO.
- HIVE-12354: MapJoin with double keys is slow on MR.
- HIVE-5277: HBase handler skips rows with null valued first cells when only row key is selected.
- HIVE-5545: HCatRecord getInteger method returns String when used on Partition columns of type INT.
- HIVE-5631: Index creation on a skew table fails.
- HIVE-6099: Multi insert does not work properly with distinct count.
- HIVE-6308: COLUMNS_V2 Metastore table not populated for tables created without an explicit column list.
- HIVE-7049: Unable to deserialize AVRO data when file schema and record schema are different and nullable.

- HIVE-8330: HiveResultSet.findColumn() parameters are case sensitive.
- HIVE-8512: queries with star and gby produce incorrect results.
- HIVE-8610: Compile time skew join optimization doesn't work with auto map join.
- HIVE-8666: hive.metastore.server.max.threads default is too high.
- HIVE-8848: data loading from text files or text file processing doesn't handle nulls correctly.
- HIVE-8875: hive.optimize.sort.dynamic.partition should be turned off for ACID.
- HIVE-8895: bugs in mergejoin.
- HIVE-8914: HDFSCleanup thread holds reference to FileSystem.
- HIVE-8917: HIVE-5679 adds two thread safety problems.
- HIVE-9013: Hive set command exposes metastore db password.
- HIVE-9073: NPE when using custom windowing UDAFs.
- HIVE-9151: Checking s against null in TezJobMonitor#getNameWithProgress() should be done earlier.
- HIVE-9390: Enhance retry logic with regard to DB access in TxnHandler.
- HIVE-9404: NPE in org.apache.hadoop.hive.metastore.txn.TxnHandler.determineDatabaseProduct().
- HIVE-9474: truncate table changes permissions on the target.
- HIVE-9484: ThriftCLIService#getDelegationToken does case sensitive comparison.
- HIVE-9496: Slf4j warning in hive command.
- HIVE-9566: HiveServer2 fails to start with NullPointerException.
- HIVE-9580: Server returns incorrect result from JOIN ON VARCHAR columns.
- HIVE-9617: UDF from_utc_timestamp throws NPE if the second argument is null.
- HIVE-9686: HiveMetastore.logAuditEvent can be used before SASL server is started.
- HIVE-9695: Redundant filter operator in reducer Vertex when CBO is disabled.
- HIVE-9711: ORC Vectorization DoubleColumnVector.isRepeating=false if all entries are NaN.
- HIVE-9720: Metastore does not properly migrate column stats when renaming a table across databases.
- HIVE-9721: Hadoop23Shims.setFullFileStatus should check for null.
- HIVE-9811: Hive on Tez leaks WorkMap objects.

- HIVE-9839: HiveServer2 leaks OperationHandle on async queries which fail at compile phase.
- HIVE-9855: Runtime skew join doesn't work when skewed data only exists in big table.
- HIVE-9920: DROP DATABASE IF EXISTS throws exception if database does not exist.

HDP 2.2.8 provides Hive 0.14.0 and the following additional Apache patches:

- HIVE-10062: : HiveOnTez Union followed by Multi-GB followed by Multi-insert loses data
- HIVE-10122: Hive metastore filter-by-expression is broken for non-partition expressions
- HIVE-10140: Window boundary is not compared correctly
- HIVE-10183: [CBO] self-join failing in a test case
- HIVE-10273: Union with partition tables which have no data fails with NPE
- HIVE-10303: HIVE-9471 broke forward compatibility of ORC files
- HIVE-10384: RetryingMetaStoreClient does not retry wrapped TTransportExceptions
- HIVE-10450: More than one TableScan in MapWork not supported in Vectorization – causes query to fail during vectorization
- HIVE-10538: Fix NPE in FileSinkOperator from hashcode mismatch
- HIVE-10578: update SQL standard authorization configuration whitelist
- HIVE-10595: Dropping a table can cause NPEs in the compactor
- HIVE-10607: Combination of ReducesinkDedup + TopN optimization yields incorrect result if there are multiple GBY in reducer
- HIVE-10685: Alter table concatenate operator will cause duplicate data
- HIVE-10698: query on view results fails with table not found error if view is created with subquery alias (CTE).
- HIVE-10719: Hive metastore failure when alter table rename is attempted.
- HIVE-10722: external table creation with msck in Hive can create unusable partition
- HIVE-10808: Inner join on Null throwing Cast Exception
- HIVE-10816: NPE in ExecDriver handleSampling when submitted via child JVM
- HIVE-10841: [WHERE col is not null] does not work sometimes for queries with many JOIN statements
- HIVE-10929: In Tez mode,dynamic partitioning query with union all fails at moveTask,Invalid partition key and values
- HIVE-10963: Hive throws NPE rather than meaningful error message when window is missing

- HIVE-10992: WebHCat should not create delegation tokens when Kerberos is not enabled
- HIVE-11008: WebHCat GET /jobs retries on getting job details from history server is too aggressive
- HIVE-11023: Disable directSQL if datanucleus.identifierFactory = datanucleus2
- HIVE-11027: Hive on tez Bucket map joins fail when hashcode goes negative
- HIVE-11028: Tez: table self join and join with another table fails with IndexOutOfBoundsException
- HIVE-11031: ORC concatenation of old files can fail while merging column statistics
- HIVE-11035: PPD Orc Split elimination fails because filterColumns=[-1]
- HIVE-11095: SerDeUtils another bug ,when Text is reused
- HIVE-11104: Select operator doesn't propagate constants appearing in expressions
- HIVE-11112: ISO-8859-1 text output has fragments of previous longer rows appended
- HIVE-11147: MetaTool doesn't update FS root location for partitions with space in name
- HIVE-11172: Vectorization wrong results for aggregate query with where clause without group by
- HIVE-11271: java.lang.IndexOutOfBoundsException when union all with if function
- HIVE-11285: ObjectInspector for partition columns in FetchOperator in SMBJoin causes exception
- HIVE-11303: Getting Tez LimitExceededException after dag execution on large query
- HIVE-11356: SMB join on tez fails when one of the tables is empty
- HIVE-11429: Increase default JDBC result set fetch size (# rows it fetches in one RPC call) to 1000 from 50
- HIVE-11438: Join a ACID table with non-ACID table fail with MR on 1.0.0
- HIVE-11442: Remove commons-configuration.jar from Hive distribution
- HIVE-11493: Predicate with integer column equals double evaluates to false
- HIVE-11613: schematool should return non zero exit status for info command, if state is inconsistent
- HIVE-5664: Drop cascade database fails when the db has any tables with indexes
- HIVE-6308: COLUMNS_V2 Metastore table not populated for tables created without an explicit column list.
- HIVE-6727: Table level stats for external tables are set incorrectly

- HIVE-7351: ANALYZE TABLE statement fails on postgres metastore
- HIVE-7997: Potential null pointer reference in ObjectInspectorUtils#compareTypes()
- HIVE-8099: IN operator for partition column fails when the partition column type is DATE
- HIVE-8295: Add batch retrieve partition objects for metastore direct SQL
- HIVE-8326: Using DbTxnManager with concurrency off results in run time error
- HIVE-8330: HiveResultSet.findColumn() parameters are case sensitive
- HIVE-8374: schematool fails on Postgres versions less than 9.2
- HIVE-8448: Union All might not work due to the type conversion issue
- HIVE-8450: Create table like does not copy over table properties
- HIVE-8518: Compile time skew join optimization returns duplicated results
- HIVE-8594: Wrong condition in SettableConfigUpdater#setHiveConfWhiteList()
- HIVE-8627: Compute stats on a table from impala caused the table to be corrupted
- HIVE-8680: Set Max Message for Binary Thrift endpoints
- HIVE-8706: Table statistic collection on counter failed due to table name character case.
- HIVE-8784: Querying partition does not work with JDO enabled against PostgreSQL
- HIVE-8863: Cannot drop table with uppercase name after compute statistics for columns
- HIVE-8869: RowSchema not updated for some ops when columns are pruned
- HIVE-8874: Error Accessing HBase from Hive via Oozie on Kerberos 5.0.1 cluster
- HIVE-8889: JDBC Driver ResultSet.getXXXXXX(String columnLabel) methods Broken
- HIVE-8926: Projections that only swap input columns are identified incorrectly as identity projections
- HIVE-9024: NullPointerException when starting WebHCat server if templeton.hive.properties is not set
- HIVE-9060: Fix child operator references after NonBlockingOpDeDupProc
- HIVE-9113: Explain on query failed with NPE
- HIVE-9155: HIVE_LOCKS uses int instead of bigint hive-txn-schema-0.14.0.mssql.sql
- HIVE-9177: Fix child operator references after NonBlockingOpDeDupProc
- HIVE-9217: UnionProcessor misses results for multi-insert when hive.optimize.union.remove=true
- HIVE-9228: Problem with subquery using windowing functions

- HIVE-9301: Potential null dereference in MoveTask#createTargetPath
- HIVE-9309: schematool fails on Postgres 8.1
- HIVE-9322: Make null-checks consistent for MapObjectInspector subclasses.
- HIVE-9347: Bug with max() together with rank() and grouping sets
- HIVE-9377: UDF in_file() in WHERE predicate causes NPE.
- HIVE-9397: SELECT max(bar) FROM foo is broken after ANALYZE ... FOR COLUMNS
- HIVE-9437: Beeline does not add any existing HADOOP_CLASSPATH
- HIVE-9445: Revert HIVE-5700 - enforce single date format for partition column storage
- HIVE-9471: Bad seek in uncompressed ORC, at row-group boundary.
- HIVE-9507: Make "LATERAL VIEW inline(expression) mytable" tolerant to nulls
- HIVE-9509: Restore partition spec validation removed by HIVE-9445
- HIVE-9513: NULL POINTER EXCEPTION
- HIVE-9526: ClassCastException thrown by HiveStatement
- HIVE-9560: When hive.stats.collect.rawdatasize=true, 'rawDataSize' for an ORC table will result in value '0' after running 'analyze table TABLE_NAME compute statistics;'
- HIVE-9567: JSON SerDe not escaping special chars when writing char/varchar data
- HIVE-9592: fix ArrayIndexOutOfBoundsException in date_add and date_sub initialize
- HIVE-9619: Uninitialized read of numBitVectors in NumDistinctValueEstimator
- HIVE-9622: Getting NPE when trying to restart HS2 when metastore is configured to use org.apache.hadoop.hive.thrift.DBTokenStore
- HIVE-9648: Null check key provider before doing set
- HIVE-9655: Dynamic partition table insertion error
- HIVE-9716: Map job fails when table's LOCATION does not have scheme
- HIVE-9720: Metastore does not properly migrate column stats when renaming a table across databases
- HIVE-9755: Hive built-in "ngram" UDAF fails when a mapper has no matches.
- HIVE-9936: fix potential NPE in DefaultUDAFEvaluatorResolver
- HIVE-9950: fix rehash in CuckooSetBytes and CuckooSetLong
- HIVE-9952: fix NPE in CorrelationUtilities
- HIVE-9953: fix NPE in WindowingTableFunction

- HIVE-9975: Renaming a non-existent partition should not throw out NullPointerException
- HIVE-9984: JoinReorder's getOutputSize is exponential

HDP 2.2.6 provided Hive 0.14.0 and the following additional Apache patches:

- HIVE-480: HDFSCleanup thread holds reference to FileSystem
- HIVE-6468: HiveServer2 (tcp mode, with SASL layer) OOMs when getting a non-thrift message
- HIVE-6679: HiveServer2 should support TCP Keepalive & Server Socket Timeout on blocking calls
- HIVE-7175: Provide password file option to beeline
- HIVE-7270: Serde info is not shown in show create table statement, but shows in the description table
- HIVE-8295: Add batch retrieve partition objects for metastore direct SQL
- HIVE-8485: SUMMARY-Hive metastore NPE with Oracle DB when there is empty value for string for tblproperties/serdeproperties/etc, table not usable after creation
- HIVE-8762: HiveMetaStore.BooleanPointer should be replaced with an AtomicBoolean
- HIVE-8791 : Hive permission inheritance throws exception S3
- HIVE-8850: SUMMARY-[ObjectStore:: rollbackTransaction() needs to be looked into further
- HIVE-8881: Receiving json "could not find job" error when web client tries to fetch all jobs from WebHCat but HDFS does not have the data.
- HIVE-8888: Hive on Tez query output duplicate rows when there is explode in subqueries for joins
- HIVE-8891: SUMMARY-[Another possible cause to NucleusObjectNotFoundException from drops/rollback
- HIVE-8893: Implement whitelist for built-in UDFs to avoid untrusted code execution in multiuser mode
- HIVE-8966: Delta files created by hive hcatalog streaming cannot be compacted
- HIVE-9025: join38.q (without map join) produces incorrect result when testing with multiple reducers
- HIVE-9038: Join tests fail on Tez
- HIVE-9055: Tez: union all followed by group by followed by another union all gives error
- HIVE-9106: improve the performance of null scan optimizer when several table scans share a physical path

- HIVE-9112: Query may generate different results depending on the number of reducers
- HIVE-9141: HiveOnTez: mix of union all, distinct, group by generates error
- HIVE-9155: HIVE_LOCKS uses int instead of bigint hive-txn-schema-0.14.0.mssql.sql
- HIVE-9205: Change default Tez install directory to use /tmp instead of /user and create the directory if it does not exist
- HIVE-9234: HiveServer2 leaks FileSystem objects in FileSystem.CACHE
- HIVE-9235: Turn off Parquet Vectorization until all data types work: DECIMAL, DATE, TIMESTAMP, CHAR, and VARCHAR
- HIVE-9249: java.lang.ClassCastException: org.apache.hadoop.hive.serde2.io.HiveVarcharWritable cannot be cast to org.apache.hadoop.hive.common.type.HiveVarchar when joining tables
- HIVE-9278: Cached expression feature broken in one case
- HIVE-9316: TestSqoop tests in WebHCat test suite hardcode libdir path to hdfs
- HIVE-9351: Running Hive Jobs with Tez cause Templeton to never report percent complete
- HIVE-9359: SUMMARY-[Hive export OOM error when table is huge. (32TB data, 4800+ partitions)]
- HIVE-9382 : Query got rerun with Global Limit optimization on and Fetch optimization off
- HIVE-9390: Enhance retry logic with regard to DB access in TxnHandler
- HIVE-9401: SimpleFetchOptimizer for limited fetches without filters
- HIVE-9404 NPE in org.apache.hadoop.hive.metastore.txn.TxnHandler.determineDatabaseProduct()
- HIVE-9436: SUMMARY-[RetryingMetaStoreClient does not retry JDOExceptions]
- HIVE-9446: JDBC DatabaseMetadata.getColumns() does not work for temporary tables
- HIVE-9473 : SQL std auth should disallow built-in udfs that allow any java methods to be called
- HIVE-9593: ORC reader should ignore new/unknown metadata streams.
- HIVE-9652: STDERR redirection should not use in place updates TEZ UI
- HIVE-9665: Parallel move task optimization causes race condition
- HIVE-9673 : Set operation handle in ATS entities for lookups
- HIVE-9683: Client TCP keep-alive for Thrift as a workaround for THRIFT-2788

- HIVE-9684: Incorrect disk range computation in ORC because of optional stream kind
- HIVE-9743 Incorrect result set for vectorized left outer join
- HIVE-9779 : ATSHook does not log the end user if doAs=false (it logs the hs2 server user)
- HIVE-9832: Merge join followed by union and a map join in hive on tez fails.
- HIVE-9836: Hive on Tez: fails when virtual columns are present in the join conditions (for e.g. partition columns)
- HIVE-9841: IOException thrown by ORC should include the path of processing file
- HIVE-9886: Tez NPE error when initialize reducer from 2 row_number function on each join side
- HIVE-9892: HIVE schema failed to upgrade from HDP 2.0/2.1 to HDP 2.2 with schematool

1.5.6. Knox 0.5.0

HDP 2.2.9 provides Knox 0.5.0 and the following additional Apache patches:

- KNOX-598: Concurrent JDBC clients via KNOX to Kerberized HiveServer2 causes HTTP 401 error.
- KNOX-599: Template with {**} in queries are expanded with =null for query params without a value

HDP 2.2.8 provided Knox 0.5.0 and the following additional Apache patches:

- KNOX-459: Apache Knox appears to be leaking sockets.
- KNOX-464: Location headers have wrong hostname when used behind load balancer.
- KNOX-465: Initial audit record can contain leftover principal name.
- KNOX-466: Log exception stack traces at INFO level when they reach gateway servlet.
- KNOX-467: Unit tests failing on windows.
- KNOX-468: update group lookup topologies to configure cache manager.
- KNOX-469: Bring CHANGES and ISSUES up to date.
- KNOX-470: add README and site docs for samples.
- KNOX-473: Configurable front end URL for simplified load balancer configuration.
- KNOX-491: Increase default replay buffer size to 8K.
- KNOX-492: Support service level replayBufferLimit for Ozzie, Hive and HBase.
- KNOX-504: Enable SSL Mutual Authentication.
- KNOX-521: Enhance Principal Mapping to Handle Dynamic Mappings.

- KNOX-529: Fix wildcard based principal group mapping.

1.5.7. Oozie 4.1.0

HDP 2.2.8 provided Apache Oozie 4.1.0 and the following additional Apache patch:

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

1.5.8. Phoenix 4.2.0

HDP 2.2.9 provides Apache Phoenix 4.2.0 and the following additional Apache patches:

- PHOENIX-1560: Join between global index and data table if INDEX hint used.
- PHOENIX-1980: CsvBulkLoad cannot load hbase-site.xml from classpath.
- PHOENIX-2005: Connection utilities omit zk client port, parent znode.
- PHOENIX-2007: java.sql.SQLException: Encountered exception in sub plan [0] execution.
- PHOENIX-2030: CsvBulkLoadTool should use logical name of the table for output directory suffix.
- PHOENIX-2031: Unable to process timestamp/Date data loaded via Phoenix org.apache.phoenix.pig.PhoenixHBaseLoader:w.
- PHOENIX-2313: TypeMismatchException thrown while querying a table that has an index with a Boolean.
- PHOENIX-2372: Check for null before passing value to Calendar.

HDP 2.2.8 provided Apache Phoenix 4.2.0 and the following additional Apache patches:

- PHOENIX-1248: CsvBulkLoadTool is failing with IAE when local index specified for –index-table parameter
- PHOENIX-1336: Exception when select from local index:Cache of region boundaries are out of date
- PHOENIX-1346: the choice of index by phoenix is not correct where do querying
- PHOENIX-1362: Min/max aggregate query on CHAR and BINARY types always return null
- PHOENIX-1427: Reduce work in StatsCollector
- PHOENIX-1428: Queries with limit against tenant views is broken
- PHOENIX-1429: Cancel queued threads when limit reached
- PHOENIX-1431: DELETE using Subqueries
- PHOENIX-1436 : The choice of index by phoenix is not correct where do querying

- PHOENIX-1446: Add/fix logging for LIMIT optimization
- PHOENIX-1447: Increase guidepost width default to take into account FAST_DIFF compression
- PHOENIX-1448: Fix resource leak when work rejected by thread executor
- PHOENIX-1449: Fix PropertiesUtil.deepCopy()
- PHOENIX-1455: Replace org.xerial.snappy with org.iq80.snappy pure Java snappy implementation
- PHOENIX-1456: Incorrect query results caused by reusing buffers in SpoolingResultIterator
- PHOENIX-1462: Create unit test for COUNT DISTINCT using compression (Ram)
- PHOENIX-1463: phoenix.query.timeoutMs doesn't work as expected
- PHOENIX-1466: Prevent multiple scans when query run serially
- PHOENIX-1467: Upgrade to 4.12 Junit and update tests by removing @Category annotation
- PHOENIX-1468: Add org.iq80.snappy to client and server assemblies
- PHOENIX-1469: Binary columns do not work correctly for indexing
- PHOENIX-1472: Fix salt bucket for Integer.MIN_VALUE
- PHOENIX-1473: Connecting with Phoenix client when Phoenix is not deployed on region server(s) takes down region server(s)
- PHOENIX-1474: NPE when RVC between combined with key part comparison
- PHOENIX-1476: Immediately open scanner for parallel queries
- PHOENIX-1477: Phoenix 4.0 -> 4.2 upgrade doesn't work
- PHOENIX-1480: Incorrect query results may occur when VIEW uses indexes from physical table
- PHOENIX-1484: Index creation failed due to specifying DEFAULT_COLUMN_FAMILY option
- PHOENIX-1485: Date columns should be displayed without timezone offset.
- PHOENIX-1486: Fix SpooledTmpFileDeleteIT by having it use a different directory for spooling files
- PHOENIX-1498: Turn KEEP_DELETED_CELLS off by default
- PHOENIX-1500: Disallow mutations and queries using PhoenixPreparedStatement.executeQuery() and executeUpdate() respectively

- PHOENIX-1503: - Compilation on Mac OS fails on maven-assembly-plugin
- PHOENIX-1532: Phoenix tarball assembly does not include server jar
- PHOENIX-1533: Last key part not taken into child/parent optimization
- PHOENIX-1537: Set reuseForks to false for integration tests
- PHOENIX-1551: Upgrading from Phoenix 4.0.x to 4.2.2 and throw can't find SYSTEM.STATS
- PHOENIX-1571: Replace hard-coded date
- PHOENIX-1575: Identical SQL type values for UNSIGNED_TIMESTAMP and UNSIGNED_DATE
- PHOENIX-1592: ORDER BY not closing ResultIterator properly
- PHOENIX-1596: Turning tracing on causes region servers to crash
- PHOENIX-1601: Performance.py script not working in binary distribution
- PHOENIX-18: Several tests are failing with SequenceNotFoundException error
- PHOENIX-32: Some phoenix tests are failing with different assertion errors

1.5.9. Pig 0.14.0

HDP 2.2.9 provides Apache Pig 0.14.0 and the following additional Apache patches:

- PIG-4624: Error on ORC empty file without schema.
- PIG-4628: Pig 0.14 job with order by fails in mapreduce mode with Oozie.
- PIG-4679: Performance degradation due to InputSizeReducerEstimator since PIG-3754.
- PIG-4683: Nested order is broken after PIG-3591 in some cases.
- PIG-4688: Limit followed by POPartialAgg can give empty or partial results in Tez.
- PIG-4696: Empty map returned by a streaming_python udf wrongly contains a null key.
- PIG-4714: Improve logging across multiple components with callerId.
- PIG-4715: Pig 0.14 does not work with Tez 0.5.4.

HDP 2.2.8 provided Apache Pig 0.14.0 and the following additional Apache patches:

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

1.5.10. Ranger 0.4.0

HDP 2.2.9 provides Ranger 0.4.0 and the following additional Apache patch:

- RANGER-684: Added support for Ranger Usersync to transform AD usernames and/or group names to Linux compliant format

HDP 2.2.8 provided Ranger 0.4.0 and the following additional Apache patch:

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

1.5.11. Slider 0.61.0

HDP 2.2.8 provided Slider 0.61.0 and the following additional Apache patches:

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

1.5.12. Storm 0.9.3

HDP 2.2.9 provides Apache Storm 0.9.3 and the following additional patches:

- STORM-1017: If ignoreZkOffsets set true,KafkaSpout will reset zk offset when recover from failure.
- STORM-1037: Do not remove storm-code in supervisor until kill job.
- STORM-1054: Excessive logging ShellBasedGroupsMapping if the user doesn't have any groups.
- STORM-965: Excessive logging in Storm when non-kerberos client tries to connect.

HDP 2.2.8 provided Apache Storm 0.9.3 and the following additional patches:

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

1.5.13. Tez 0.5.2

HDP 2.2.9 provides Apache Tez 0.5.2 and the following Apache patches:

- TEZ-2896: Fix thread names used during Input/Output initialization.
- TEZ-2910: Tez should invoke HDFS Client API to set up caller context.
- TEZ-2936: Support HDFS-based Timeline writer.

HDP 2.2.8 provides Apache Tez 0.5.2 and the following Apache patches:

- TEZ-1521: VertexDataMovementEventsGeneratedEvent may be logged twice in recovery log.

- TEZ-1560: Invalid state machine handling for V_SOURCE_VERTEX_RECOVERED in recovery.
- TEZ-1642: TestAMRecovery sometimes fails.
- TEZ-1687: Use logIdentifier of Vertex for logging.
- TEZ-1697: DAG submission fails if a local resource added is already part of tez.lib.uris.
- TEZ-1737: Should add taskNum in VertexFinishedEvent.
- TEZ-1772: Failing tests post TEZ-1737.
- TEZ-1773: Add attempt failure cause enum to the attempt failed/killed history record.
- TEZ-1796: Use of DeprecationDelta broke build against 2.2 Hadoop.
- TEZ-1808: Job can fail since name of intermediate files can be too long in specific situation.
- TEZ-1816: It is possible to receive START event when DAG is failed.
- TEZ-1832: TestSecureShuffle fails with NoClassDefFoundError: org/bouncycastle/x509/X509V1CertificateGenerator.
- TEZ-1909: Remove need to copy over all events from attempt 1 to attempt 2 dir.
- TEZ-1929: pre-empted tasks should be marked as killed instead of failed.
- TEZ-1961: Remove misleading exception "No running dag" from AM logs.
- TEZ-2000: Source vertex exists error during DAG submission.
- TEZ-2011: InputReadyVertexManager not resilient to updates in parallelism.
- TEZ-2033: Update TestOrderedWordCount to add processor configs as history text and use MR configs correctly.
- TEZ-2036: OneToOneEdgeManager should enforce that source and destination tasks have same number.
- TEZ-2064: SessionNotRunning Exception not thrown is all cases.
- TEZ-2133: Secured Impersonation: Failed to delete tez scratch data dir.
- TEZ-2162: org.apache.hadoop.mapreduce.lib.output.SequenceFileOutputFormat is not recognized.
- TEZ-2179: Timeline related entries missing cause exaggerated warning.
- TEZ-2186: TEZ-2186 OOM with a simple scatter gather job with re-use.
- TEZ-2189: Tez UI live AM tracking URL only works for localhost addresses.
- TEZ-2193: Check returned value from EdgeManagerPlugin before using it.

- TEZ-2195: TestTezJobs :testInvalidQueueSubmission/ testInvalidQueueSubmissionToSession fail with hadoop branch-2.
- TEZ-2199: updateLocalResourcesForInputSplits assumes wrongly that split data is on same FS as the default FS.
- TEZ-2203: Intern strings in tez counters.
- TEZ-2205: Tez still tries to post to ATS when yarn.timeline-service.enabled=false.
- TEZ-2221: VertexGroup name should be unique.
- TEZ-2224: EventQueue empty doesn't mean events are consumed in RecoveryService.
- TEZ-2257: Fix potential NPEs in TaskReporter.
- TEZ-2289: ATSHistoryLoggingService can generate ArrayOutOfBoundsException.
- TEZ-2290: Scale memory for Default Sorter down to a max of 2047 MB if configured higher.
- TEZ-2293: When running in "mr" mode, always use MR config settings.
- TEZ-2299: Invalid dag creation in MRRSleepJob post TEZ-2293.
- TEZ-2303: ConcurrentModificationException while processing recovery.
- TEZ-2304: InvalidStateTransitonException TA_SCHEDULE at START_WAIT during recovery.
- TEZ-2305: MR compatibility sleep job fails with IOException: Undefined job output-path.
- TEZ-2311: AM can hang if kill received while recovering from previous attempt.
- TEZ-2317: Event processing backlog can result in task failures for short tasks.
- TEZ-2323: Fix TestOrderedWordcount to use MR memory configs.
- TEZ-2379: org.apache.hadoop.yarn.state.InvalidStateTransitonException: Invalid event: T_ATTEMPT_KILLED at KILLED.
- TEZ-2383: Cleanup input/output/processor contexts in LogicalIOProcessorRuntimeTask.
- TEZ-2398: Flaky test: TestFaultTolerance.
- TEZ-2436: Tez UI: Add cancel button in column selector.
- TEZ-2460: Temporary solution for issue due to YARN-2560.
- TEZ-2474: The old taskNum is logged incorrectly when parallelism is changed.
- TEZ-2483: Tez should close task if processor fail.
- TEZ-2488: Tez AM crashes if a submitted DAG is configured to use invalid resource sizes.
- TEZ-2509: YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.

- TEZ-2511: Add exitCode to diagnostics when container fails.
- TEZ-2533: AM deadlock when shutdown.
- TEZ-2534: Error handling summary event when shutting down AM.
- TEZ-2538: ADDITIONAL_SPILL_COUNT wrongly populated for DefaultSorter with multiple partitions.
- TEZ-2548: TezClient submitDAG can hang if the AM is in the process of shutting down.
- TEZ-2552: CRC errors can cause job to run for very long time in large jobs.
- TEZ-2561: Port for TaskAttemptListenerImpTezDag should be configurable.
- TEZ-2566: Allow TaskAttemptFinishedEvent without TaskAttemptStartedEvent when it is KILLED/FAILED.
- TEZ-2629: LimitExceededException in Tez client when DAG has exceeds the default max counters.
- TEZ-2635: Limit number of attempts being downloaded in unordered fetch.
- TEZ-2636: MRInput and MultiMRInput should work for cases when there are 0 physical inputs.
- TEZ-2662: Provide a way to check whether AM or task opts are valid and error if not.
- TEZ-2663: SessionNotRunning exceptions are wrapped in a ServiceException from a dying AM.
- TEZ-2687: Tez should release/kill all held containers before stopping services during the shutdown phase.
- TEZ-2716: DefaultSorter.isRleNeeded not thread safe.
- TEZ-2719: Consider reducing logs in unordered fetcher with shared-fetch option.
- TEZ-2731: Fix Tez GenericCounter performance bottleneck.
- TEZ-2732: DefaultSorter throws ArrayIndex exceptions on 2047 Mb size sort buffers.
- TEZ-2734: Add a test to verify the filename generated by OnDiskMerge.
- TEZ-2742: VertexImpl.finished() terminationCause hides member var of the same name.
- TEZ-2744: Improvements and cleanup of logging for the AM and parts of the runtime.
- TEZ-2745: ClassNotFoundException of user code should fail dag.
- TEZ-2752: logUnsuccessful completion in Attempt should write original finish time to ATS.
- TEZ-2758: Remove append API in RecoveryService after TEZ-1909.
- TEZ-2761: Tez UI: update the progress on the dag and vertices pages with info from AM.

- TEZ-2767: Make TezMxBeanResourceCalculator the default resource calculator.
- TEZ-2768: Log a useful error message when the summary stream cannot be closed when shutting down an AM.
- TEZ-2775: Improve and consolidate logging in Runtime components.
- TEZ-2781: Fallback to send only TaskAttemptFailedEvent if taskFailed heartbeat fails.
- TEZ-2787: Tez AM should have java.io.tmpdir=./tmp to be consistent with tasks.
- TEZ-2792: Add AM web service API for tasks.
- TEZ-2808: Race condition between preemption and container assignment.
- TEZ-2813: Tez UI: add counter data for rest api calls to AM Web Services v2.
- TEZ-2816: Preemption sometimes does not respect heartbeats between pre-emptions.
- TEZ-2825: Report progress in terms of completed tasks to reduce load on AM for Tez UI.
- TEZ-2829: Tez UI: minor fixes to in-progress update of UI from AM.
- TEZ-2834: Make Tez preemption resilient to incorrect free resource reported by YARN.
- TEZ-2850: Tez MergeManager OOM for small Map Outputs.
- TEZ-2851: Support a way for upstream applications to pass in a caller context to Tez.
- TEZ-2855: Fix a potential NPE while routing VertexManager events.
- TEZ-2857: Fix flakey tests in TestDAGImpl.
- TEZ-2858: Stop using System.currentTimeMillis in TestInputReadyTracker.
- TEZ-2860: NPE in DAGClientImpl.
- TEZ-2868: Fix setting Caller Context in Tez Examples.
- TEZ-2874: Improved logging for caller context.
- TEZ-2882: Consider improving fetch failure handling.
- TEZ-2885: Remove counter logs from AMWebController.
- TEZ-2887: Tez build failure due to missing dependency in pom files.
- TEZ-2896: Fix thread names used during Input/Output initialization.
- TEZ-2907: NPE in IFile.Reader.getLength during final merge operation.
- TEZ-814: Improve heuristic for determining a task has failed outputs.
- TEZ-900: Confusing message for incorrect queue for some tez examples.

HDP 2.2.6 provided Apache Tez 0.5.2 and the following Apache patches:

- TEZ-1642: TestAMRecovery sometimes fails.
- TEZ-1775: Allow setting log level per logger.
- TEZ-1800: Integer overflow in ExternalSorter.getInitialMemoryRequirement()
- TEZ-1836: Provide better error messages when tez.runtime.io.sort.mb, spill percentage is incorrectly configured.
- TEZ-1851: FileSystem counters do not differentiate between different FileSystems
- TEZ-1852: Get examples to work in Local Mode.
- TEZ-1861: Fix failing test: TestOnFileSortedOutput.
- TEZ-1878: Task-specific log level override not working in certain conditions
- TEZ-1924: Tez AM does not register with AM with full FQDN causing jobs to fail in some environments.
- TEZ-1931: Publish tez version info to Timeline.
- TEZ-1934: TestAMRecovery may fail due to the execution order is not determined.
- TEZ-1942: Number of tasks show in Tez UI with auto-reduce parallelism is misleading.
- TEZ-1949: Whitelist TEZ_RUNTIME_OPTIMIZE_SHARED_FETCH for broadcast edges
- TEZ-1962: Fix a thread leak in LocalMode.
- TEZ-2024: TaskFinishedEvent may not be logged in recovery.
- TEZ-2037: Should log TaskAttemptFinishedEvent if taskattempt is recovered to KILLED
- TEZ-2135: ACL checks handled incorrectly in AMWebController.

As part of HDP 2.2.6 Hortonworks is providing a Tez Debugging User Interface. This interface does not impact the behavior or function of jobs that leverage Tez, and its use is optional. Patches added to facilitate the Tez Debugging User Interface include:

- TEZ-1990: Tez UI: DAG details page shows Nan for end time when a DAG is running.
- TEZ-2031: Tez UI: horizontal scrollbars do not appear in tables, causing them to look truncated.
- TEZ-2038: TEZ-UI DAG is always running in Tez-UI when the app is failed but no DAGFinishedEvent is logged.
- TEZ-2043: Tez UI: add progress info from am webservice to dag and vertex views.
- TEZ-2052: Tez UI: log view fixes, show version from build, better handling of ATS URL config.
- TEZ-2056: Tez UI: fix VertexID filter,show only tez configs by default,fix app attempt ID.
- TEZ-2063: Tez UI: Flaky log URL in tasks table.

- TEZ-2065: Setting up tez.tez-ui.history-url.base with a trailing slash can result in failures to redirect correctly.
- TEZ-2068: Tez UI: Dag view should use full window height, disable web UI service in local mode.
- TEZ-2069: Tez UI: appId should link to application in dag details view.
- TEZ-2077: Tez UI: No diagnostics on Task Attempt Details page if task attempt failed
- TEZ-2078: Tez UI: Task logs URL use in-progress URL causing various errors.
- TEZ-2079: Tez UI: trailing slash in timelineBaseUrl in UI should be handled.
- TEZ-2092: Tez UI history URL handler injects spurious trailing slash.
- TEZ-2098: Tez UI: Dag details should be the default page for dag, fix invalid time entries for failed Vertices.
- TEZ-2101: Tez UI: Issues on displaying a table.
- TEZ-2102: Tez UI: DAG view has hidden edges, dragging DAG by holding vertex causes unintended click.
- TEZ-2106: TEZ UI: Display data load time, and add a refresh button for items that can be refreshed.
- TEZ-2112: Tez UI: fix offset calculation, add home button to breadcrumbs.
- TEZ-2114: Tez UI: task/task attempt status is not available when its running.
- TEZ-2116: Tez UI: dags page filter does not work if more than one filter is specified.
- TEZ-2134: TEZ UI: On request failure, display request URL and server name in error bar.
- TEZ-2136: Some enhancements to the new Tez UI.
- TEZ-2142: TEZ UI: Breadcrumb border color looks out of place in wrapped mode.
- TEZ-2158: TEZ UI: Display dag/vertex names, and task/attempt index in breadcrumb
- TEZ-2160: Tez UI: App tracking URL should support navigation back.
- TEZ-2165: Tez UI: DAG shows running status if killed by RM in some cases.

1.5.14. ZooKeeper 3.4.6

HDP 2.2.8 provided ZooKeeper 3.4.6 and the following additional Apache patches:

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

1.6. Common Vulnerabilities and Exposures

No information-security vulnerabilities and exposures (CVEs) were fixed as part of HDP 2.2.9.

1.7. Third-party Licenses

Component	Subcomponents	License
Accumulo	JCommander	JCommander
Falcon	cern.colt* , cern.jet*, cern.clhep	CERN
Knox	ApacheDS, Groovy	ANTLR
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle
Phoenix		EPL
Storm	Logback	EPL

1.8. Fixed Issues

Fixed issues represents selected issues that were previously logged via Hortonworks Support, but are now addressed in the current release. These issues may have been reported in previous versions within the Known Issues section; meaning they were reported by customers or identified by Hortonworks Quality Engineering team.

Incorrect Results

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-41313		Hive	Hive temporary macro with variable inputs returning incorrect results
BUG-42916	HIVE-11606	Hive	Bucket mapjoin error "capacity must be power of two" when #rows < #buckets
BUG-43169		Hive	Hive + Tez: map side join giving different query results
BUG-44538	HBASE-13250 HBASE-14445	HBase	ExportSnapshot does not honor -chuser, -chgroup, -chmod options
BUG-44924		Hive	SMB Join query not returning proper results
BUG-45117	HIVE-11132	Hive	BackPort - wrong results: non-mapjoin with auto.convert.join.noconditionaltask=false and hive.optimize.reducededuplication=true wrong plan for join followed with group by
BUG-45581		Hue	Cluster running Hue 2.6.1 Null values are being returned as 'None' and they are converted to blank value while exporting as CSV file
BUG-45656		Hive	Correctness: Wrong results when using a not exists subquery with CBO off
BUG-45723		Phoenix	org.apache.phoenix.schema.TypeMismatchExcept Error while querying a table that has an index with a Boolean

Other

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-43228	HDFS-8885	HDFS	ByteRangeInputStream used in webhdfs does not override available()
BUG-44034	HDFS-9112	HDFS	After implementing distcp between two HA clusters, the HDFS haadmin commands do not work - Unable to determine the nameservice id.
BUG-44480	HDFS-9082	HDFS	Change the log level in WebHdfsFileSystem.initialize() from INFO to DEBUG
BUG-44523		Hue	Not displaying tables in alphabetical order

Performance

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-38575		Hive	Join on partition columns plan should know column stats without having to compute them, unreasonable estimate on partition stats may disable dynamic partition pruning
BUG-43027		HBase	Have metrics for blockcache take into account replicas
BUG-43071	HIVE-9223	Hive, Hue	HS2/hue/tez: multiple concurrent queries not allowed in single tez session
BUG-43268	HIVE-10651	Hive	OrcProto$StripeStatistics cache grows causing severe GC in Application master
BUG-44683	PIG-4679	Pig	Drastic performance degradation due to InputSizeReducerEstimator
BUG-45735		Phoenix	Index HINT not forcing the usage of secondary Index
BUG-46376		Hive	Vectorization Udf: GenericUDFOPGreaterThan, is not supported
BUG-46749	HDFS-7097 HDFS-7448	HDFS	Allow block reports to be processed during checkpointing on standby name node

Potential Data Loss

None.

Query Failure

None.

Security

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-28778	HIVE-9013	Hive	Beeline (hiveserver2 client) exposes sensitive metastore DB connection info (connection, password)
BUG-44484		Ranger	Ranger HBASE plugin doesn't audit close_region, split, and compact operations when action is authorized
BUG-36587	HBASE-12366	HBase	Add login code to HBase Canary tool
BUG-43072		Hue	Hue web application is vulnerable to stored Cross-Site Scripting (XSS)
BUG-43224	HDFS-8155	HDFS	Support OAuth2 in WebHDFS
BUG-43238	HIVE-10528	Hive	Hiveserver2 in HTTP mode is not applying auth_to_local rules
BUG-44889		Ranger	Implement X-Frame-Option header
BUG-44891		Ranger	Implement best practices for authentication cookies
BUG-45430	HDFS-9175	HDFS	Change scope of 'AccessTokenProvider.getAccessToken()' and 'CredentialBasedAccessTokenProvider.getCredenti abstract methods to public
BUG-46555		Ranger	Restrict group and role data for only ADMIN users

Stability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-28507		Ranger	Add timeout for connection settings
BUG-31129	YARN-3145	YARN	java.sql.SQLException: Error while processing statement: FAILED: Hive Internal Error: java.util.ConcurrentModificationException
BUG-37456		Ranger	Ambari's treatment of ranger-admin setup.sh can corrupt security-applicationContext.xml
BUG-40689	HIVE-11193	Hive	Query plan fails, tez dynamic partition pruning "fail to find child from parent " in this case (left outer join union all with constant on different data types)
BUG-41828	HIVE-11408	Hive	Hiveserver2 failing with OOM PermGen when using temporary functions due to Hadoop reflections utils caching classloaders

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-42203	HIVE-11499	Hive	Hiveserver2 failing with OOM PermGen when using temporary functions due to DataNucleus caching classloaders
BUG-43070		Hue	Select has a 1024 fd limit. If the FD number goes above 1024 for some reason (long running process), then all apps relying on thrift will fail
BUG-43221	TEZ-2745	Tez	ClassNotFound in InputInitializer causes AM to crash
BUG-43226	HDFS-8435	HDFS	Support CreateFlag in WebHDFS
BUG-43543	HBASE-14309	HBase	Allow load balancer to operate when there is region in transition by adding force flag
BUG-43604		Hive	Exception in FileSink when num_bucket > num_reducer
BUG-43658	HDFS-8995	HDFS	Flaw in registration bookkeeping can make DN die on reconnect
BUG-43895	HDFS-8797	HDFS	WebHdfsFileSystem creates too many connections for position read
BUG-43986	HBASE-12635	HBase	Delete ACL notify znode of table after the table is deleted
BUG-44099		Hive	Hiveserver2 failing with OOM PermGen when use temporary functions due to Hadoop ReflectionUtils caching classloaders
BUG-44102		Hue	When accessing a Table which has more than 4K columns, Beeswax and HCatalog fails to get the content and results in error
BUG-44212	HIVE-12084	Hive	Hive queries with ORDER BY & large LIMIT or PTF order by high cardinality fails with OutOfMemoryError Java heap space
BUG-44257	YARN-3896	YARN	Node Manager processes becomes unresponsive after RM restart
BUG-44260	HBASE-14412	HBase	Removing HBase cross cluster replication peer id containing hyphen "-" doesn't seem to remove all the related entries from ZooKeeper
BUG-44503	KNOX-598	Knox	Concurrent JDBC clients via KNOX to Kerberized HiveServer2 - causes HTTP 401 error (due to Kerberos Replay attack error)

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-44722	STORM-329 STORM-404	Storm	Worker on one machine crashes due to a failure of another worker on another machine
BUG-44937	HBASE-12464	HBase	Meta table region assignment stuck in the FAILED_OPEN state due to region server not fully ready to serve
BUG-44943		Hive	Tez queries stopped running - ApplicationNotFoundException
BUG-45049	HIVE-7049	Hive	Error: java.io.IOException: org.apache.avro.AvroTypeException: Found string, expecting union (state=,code=0)
BUG-45083	HIVE-10752	HCatalog, Pig	More than 66 columns breaks MR but not tez (Pig HCat)
BUG-45098	HIVE-11960	Hive	Hive query parse issue
BUG-45215	HADOOP-12089	Hadoop Common	StorageException complaining "no lease ID" when updating FolderLastModifiedTime in WASB
BUG-45216	HADOOP-12239	Hadoop Common	StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB
BUG-45465	HIVE-11977	Hive	Hive should handle an external avro table with zero length files present
BUG-45954	HBASE-12128 HBASE-12604	HBase	Cache configuration and RpcController selection for Table in Connection
BUG-46165	HBASE-14207	HBase	Region was hijacked and remained in transition when RS failed to open a region and later regionplan changed to new RS on retry
BUG-46166	HBASE-14361	HBase	ReplicationSink should create Connection instances lazily
BUG-46327	YARN-2910	YARN	FSLeafQueue can throw ConcurrentModificationException
BUG-46587		Ranger	Prevent thread exhaustion in Ranger due to badly configured service connection
BUG-47098	HDFS-8626	HDFS	Reserved RBW space is not released if creation of RBW File fails

Upgrade

None.

Usability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-30784	ZOOKEEPER-1952	ZooKeeper	zookeeper.log.file property is not respected; log output goes only to the zookeeper.out
BUG-40345		Hue	Log empty for create and insert statements
BUG-43069		Hue	Hue does not respect Support dfs.umaskmode, fs.permissions.umask-mode when creating files or folders
BUG-43350		Ranger	Ranger Installation fails due to missing ojdbc.jar in CLASSPATH for ambari 2.1.0
BUG-43403	HIVE-11581	Hive	HS2 URL parameters should be stored in zookeeper
BUG-43465		Hive	Hive "SLF4J: Class path contains multiple SLF4J bindings." error
BUG-43603	HIVE-12262	Hive	hive does not translate ${system:java.io.tmpdir}/ ${system:user.name}/ operation_logs properly
BUG-43766	HDFS-8180	HDFS	AbstractFileSystem Implementation for WebHDFS
BUG-43799	HDFS-8542	HDFS	WebHDFS getHomeDirectory behavior does not match specification
BUG-44528		Hue	Enable Parameterization doesn't work
BUG-45263	HADOOP-12437	HDFS	HDFS and YARN configs for Kerberized Dual Homing
BUG-45360	HBASE-14594	HBase	HBase configs for Kerberized Dual Homing
BUG-45675		Ranger	Allow special characters in passwords for manual install of Ranger
BUG-45721	HADOOP-12350	Hadoop Common	WASB Logging: Improve WASB Logging around deletes, reads and writes
RMP-3882	RANGER-684	Ranger	Mapping of Ranger AD usersync groups to Linux AD groups

1.9. Known Issues for HDP 2.2.9

HDP 2.2.9 has the following known issues, scheduled for resolution in a future release. Where available, a workaround has been provided.

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-36601		HBase	Scanner keyvalue size / partial results

Hortonworks Bug ID	Apache JIRA	Component	Summary
			Issue: In 2.2 clusters, setting "hbase.client.scanner.max.result.size" differently on the client side or server side might result in missed data from large scans because the scanner might wrongly assume region is exhausted before the region boundary is reached. Additionally, the result size calculation might be different if cell level tags are used even if "hbase.client.scanner.max.result.size" is the same between 2.2 and 2.3 clusters. This issue also affects rolling upgrades between 2.2 and 2.3 clusters, where a 2.2 client might miss data in a large scan when rolling upgrading or talking to 2.3 servers. As a protection, 2.3 does not set "hbase.client.scanner.max.result.size" by default. **Workaround**: Do not use partial scan results to be returned (by not setting "hbase.client.scanner.max.result.size") is recommended.
BUG-38980	HBASE-14223	HBase	Meta WALs are not split or cleared **Issue**: In case a meta region gets moved from one region server to another region server, the write ahead logs for the meta table might be left in the HDFS directory for the regionserver if that particular regionserver fails. This might leave the WAL files in that directory until they are removed manually, and also may result in the previous regionservers to be listed as "dead regionservers" in the master UI indefinitely. However, this is harmless and does not cause any data loss other operational issues in HBase.
BUG-40259		Ranger	**Issue:** Upgrade to HDP 2.2.6 breaks usersync. **Details:** Ranger in HDP 2.2.6 introduced authentication between usersync and the admin REST API. If you upgrade to HDP 2.2.6, there will be a previous version of `/etc/ranger/`

Hortonworks Bug ID	Apache JIRA	Component	Summary

			`usersync/conf/unixauthservice.properties` and `{{ /etc/ranger/admin/conf/security-applicationContext.xml}}` in place. Because setup.sh treats these files as templates, it expects certain base content to be there in order to generate correct configurations. It looks like there is new configuration in the default templates (the ones in `/usr/hdp/2.2.6.0-2800/etc/ranger/...`) that will not get pulled into the live configs. This results in usersync pulling 401 errors every time it tries to add/modify a user.
			
			Workaround:
			
			1. Apply to the copy files:
			
			`/usr/hdp/2.2.6.0-2800/etc/ranger/usersync/conf/unixauthservice.properties -> /etc/ranger/usersync/conf/unixauthservice.properties`
			`/usr/hdp/2.2.6.0-2800/etc/ranger/admin/conf/security-applicationContext.xml -> /etc/ranger/admin/conf/security-applicationContext.xml`
			
			2. Rerun setup.sh.
			
			3. Restart admin.
			
			4. Restart usersync.

BUG-42084		Hive	Prevent misconfiguration when StorageBasedAuthorization is set on hive.security.authorization.manager
BUG-42498	HDFS-8999	HDFS	Block received messages put heavy load on namenode
BUG-42569	HIVE-10022	Hive	Create database [db_name] location / tmp/[db_name].db' via beeline throws [hrt_qa] does not have [WRITE] privilege although hive.server2.enable.doAs=false when Ranger or StdAuth is on
BUG-42937	HDFS-8864	HDFS	Padding is needed in remaining space check

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-42938	HDFS-8871	HDFS	Decommissioning of a node with a failed volume may not start
BUG-42939	HDFS-8870	HDFS	Lease is leaked on write failure
BUG-43023	HDFS-5215 HDFS-9038	HDFS	dfs.datanode.du.reserved is not considered while computing available space
BUG-43583	SLIDER-931	Slider	test_runintegrationtestzkandfspermissions failed on HBase-Slider
BUG-43773		Spark	**Issue:** Spark: Fails to find table **Details:** Spark + Kerberos + yarn-cluster mode + remote hive meta store for SparkSQL does not work. **Workaround:** Try this with Spark + Kerberos + yarn-client mode for SparkSQL.
BUG-43979		YARN	RM UI prints ZK RM connection state="CONNECTED" , if >50% ZKs are disconnected from active RM
BUG-44485		Oozie	Oozie should invoke Hive/ MR/Pig client API to setup caller context
BUG-44538	HBASE-13250 HBASE-14445	HBase	ExportSnapshot does not honor -chuser, -chgroup, -chmod options
BUG-44732		Oozie	Oozie should push workflow id information to ATS when a new workflow is instantiated
BUG-45757		Phoenix	Rows get "locked" and block Phoenix operations
BUG-45759		Phoenix	Build index error on phoenix
BUG-46041	TEZ-2902	Tez	NPE if a container heartbeat fails before the task starts running
BUG-46313		YARN	RM UI should be smarter on when and when not to use authentication
BUG-46599		HBase	HBase's ExportSnapshotTool requires hdfs user to trigger ExportSnapshot, but application fails with 'Requested user hdfs is banned'
BUG-46660		Hive	Log HiveConf on startup for metastore
BUG-46661	HIVE-11891	Hive	Add basic performance logging to metastore calls
BUG-46663	HIVE-11892	Hive	Local fetch task does not work with UDTF

Hortonworks Bug ID	Apache JIRA	Component	Summary
			Workaround: Disable fetch task conversion if using a UDTF that forwards rows during GenericUDTF.close().
BUG-46664		Hive	Hive query execution failed complaining about absence of lock
BUG-46726		Hive	Select join tables fails in a case of there are ten thousand of orc files
BUG-46886	HADOOP-8830	Tez	Standalone Tez UI is not accessible in Secure cluster
BUG-46983		Hive	HDP 2.2.8 has hive.exec.parallel.* in confWhitelist, not including hive.exec.parallel
BUG-47062		HBase	ExportSnapshot with chown is not usable with current hdfs and yarn constraints
BUG-47099		Hive	Insert query on skewed table runs into MoveTask error in MR mode
BUG-47226		Hive	Join query returns different result with and without vectorization in Tez mode when all the entries are NaN
BUG-47317		Ranger, RelEng	Ranger Admin failed to start after the stack upgrade from HDP 2.2.6 to HDP 2.2.9 in AMBARI 2.1.2.1 **Issue**: If Ranger service is not stopped prior to upgrade (using the old version) and started after upgrade (using the new version), you will encounter issue starting ranger service. **Workaround**: Change the permission for stop-ranger-admin.sh file (from root to user running ranger service). The service should start up fine.
BUG-47472		Hive	Query fails on ORC ppd on timestamp datatype on stripes with all null on the column
BUG-47749	HBASE-11617	HBase	Cherry-pick HBASE-11617
BUG-48274		YARN	The RollingLevelDBTimelineStore in 2.2 does not work with JDK 1.6. Customers running JDK 1.6 should not run ATS (Application Timeline Server) with RollingLevelDBTimelineStore.
BUG-48600, BUG-28507		Ranger	**Issue**: When Ranger is upgraded from version

Hortonworks Bug ID	Apache JIRA	Component	Summary
			HDP 2.6.0.2.2.8.0-3150 to HDP-2.2.9.0-3350. After upgrade Ranger restart fails with error of missing property in xa_system.properties file: `2015-11-20 09:30:45,268 [localhost-startStop-1] ERROR org.springframework.web.context.Conte (ContextLoader.java:307) - Context initialization failed org.springframework.beans.factory.Bea Invalid bean definition with name 'defaultDataSource' defined in ServletContext resource [/META-INF/applicationContext.xml]: Could not resolve placeholder 'jdbc.maxIdleTimeExcessConnections' in string value "${jdbc.maxIdleTimeExcessConnections}`  **Workaround**: Add properties below given properties in xa_system.properties and Retry to start Ranger:  `jdbc.maxIdleTimeExcessConnections=300`  `jdbc.testConnectionOnCheckin=true` This should resolve the issue.
BUG-50531		Kafka	Kafka file system support **Issue**: Encrypted file systems such as SafenetFS are not supported for Kafka. Index file corruption can occur. For more information, see: Install Kafka.

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
TSB-405	N/A	N/A	**Impact of LDAP Channel Binding and LDAP signing changes in Microsoft Active Directory** Microsoft has introduced changes in LDAP Signing and LDAP Channel Binding to increase the security for communications between LDAP clients and Active Directory domain controllers. These optional changes will have an impact on how 3rd party products integrate with Active Directory using the LDAP protocol. **Workaround** Disable LDAP Signing and LDAP Channel Binding features in Microsoft Active Directory if they are enabled

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
			For more information on this issue, see the corresponding Knowledge article: TSB-2021 405: Impact of LDAP Channel Binding and LDAP signing changes in Microsoft Active Directory
TSB-406	N/A	HDFS	**CVE-2020-9492 Hadoop filesystem bindings (ie: webhdfs) allows credential stealing** WebHDFS clients might send SPNEGO authorization header to remote URL without proper verification. A maliciously crafted request can trigger services to send server credentials to a webhdfs path (ie: webhdfs://...) for capturing the service principal For more information on this issue, see the corresponding Knowledge article: TSB-2021 406: CVE-2020-9492 Hadoop filesystem bindings (ie: webhdfs) allows credential stealing
TSB-434	HADOOP-17208, HADOOP-17304	Hadoop	**KMS Load Balancing Provider Fails to invalidate Cache on Key Delete** For more information on this issue, see the corresponding Knowledge article: TSB 2020-434: KMS Load Balancing Provider Fails to invalidate Cache on Key Delete
TSB-465	N/A	HBase	**Corruption of HBase data stored with MOB feature** For more information on this issue, see the corresponding Knowledge article: TSB 2021-465: Corruption of HBase data stored with MOB feature on upgrade from CDH 5 and HDP 2
TSB-497	N/A	Solr	**CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler** The Apache Solr ReplicationHandler (normally registered at "/replication" under a Solr core) has a "masterUrl" (also "leaderUrl" alias) parameter. The “masterUrl” parameter is used to designate another ReplicationHandler on another Solr core to replicate index data into the local core. To help prevent the CVE-2021-27905 SSRF vulnerability, Solr should check these parameters against a similar configuration used for the "shards" parameter. For more information on this issue, see the corresponding Knowledge article: TSB 2021-497: CVE-2021-27905: Apache Solr SSRF vulnerability with the Replication handler
TSB-512	N/A	HBase	**HBase MOB data loss** HBase tables with the MOB feature enabled may encounter problems which result in data loss. For more information on this issue, see the corresponding Knowledge article: TSB 2021-512: HBase MOB data loss

1.10. Documentation Errata

The following section contains late additions or corrections to the product documentation.

- Configuring Pig Scripts to Use HCatalog in Oozie Workflows [55]
- Configuring a Sqoop Action to Use Tez to Load Data into a Hive Table [56]

1.10.1. Configuring Pig Scripts to Use HCatalog in Oozie Workflows

To access HCatalog with a Pig action in an Oozie workflow, you need to modify configuration information to point to the Hive metastore URIs.

There are two methods for providing this configuration information. Which method you use depends upon how often your Pig scripts access the HCatalog.

1.10.1.1. Configuring Individual Pig Actions to Access HCatalog

If only a few individual Pig actions access HCatalog, do the following:

1. Identify the URI (host and port) for the Thrift metastore server.
   a. In Ambari, click **Hive > Configs > Advanced**.
   b. Make note of the URI in the **hive.metastore.uris** field in the General section.

      This information is also stored in the `hive.default.xml` file.
2. Add the following two properties to the <configuration> elements in each Pig action.

Note

Replace *`[host:port(default:9083)]`* in the example below with the host and port for the Thrift metastore server.

```
<configuration>
    <property>
        <name>hive.metastore.uris</name>
        <value>thrift://[host:port(default:9083)]</value>
        <description>A comma separated list of metastore uris the client can
 use to contact the
        metastore server.</description>
    </property>
    <property>
        <name>oozie.action.sharelib.for.pig</name>
        <value>pig,hive,hcatalog</value>
        <description>A comma separated list of libraries to be used by the
 Pig action.</description>
    </property>
</configuration>
```

1.10.1.2. Configuring All Pig Actions to Access HCatalog

If all of your Pig actions access HCatalog, do the following:

1. Add the following line to the `job.properties` files, located in your working directory:

```
oozie.action.sharelib.for.pig=pig,hive,hcatalog
<!-- A comma separated list of libraries to be used by the Pig action.-->
```

2. Identify the URI (host and port) for the Thrift metastore server.

a. In Ambari, click **Hive > Configs > Advanced**.

b. Make note of the URI in the **hive.metastore.uris** field in the General section.

This information is also stored in the `hive.default.xml` file.

3. Add the following property to the <configuration> elements in each Pig action.

Note

Replace *[host:port(default:9083)]* in the example below with the host and port for the Thrift metastore server.

```
<configuration>
    <property>
        <name>hive.metastore.uris</name>
        <value>thrift://[host:port(default:9083)]</value>
        <description>A comma separated list of metastore uris the client can
 use to contact the
        metastore server.</description>
    </property>
    </configuration>
```

1.10.2. Configuring a Sqoop Action to Use Tez to Load Data into a Hive Table

You can use the Tez execution engine to load data into a Hive table using the `--hive-import` option,

In the code example in each step, replace the sample text in [square brackets] with the appropriate information for your configuration.

1. Create a workflow directory.

```
hdfs dfs -mkdir -p [/user/dummy/app]
```

2. Create a `lib` directory in the workflow directory.

```
hdfs dfs -mkdir -p [/user/dummy/app/lib]
```

3. Copy the database JDBC driver jar file to the `lib` directory.

```
hadoop fs -copyFromLocal [/usr/share/java/mysql-connector-java.jar]
        [/user/dummy/app/lib]
```

4. Copy the `hive-site.xml` and `tez-site.xml` files to a location accessible by the workflow. For example:

```
hadoop fs -copyFromLocal [/etc/oozie/conf/action-conf/hive/hive-site.xml /
user/dummy/app]
hadoop fs -copyFromLocal [/etc/oozie/conf/action-conf/hive/tez-site.xml /
user/dummy/app]
```

5. In the Sqoop action of the workflow, do the following:

- Add `hive-site` and `tez-site` resources in the <file> element of the Sqoop action in the workflow.

```
<file>/user/dummy/app/hive-site.xml#hive-site.xml</file>
<file>/user/dummy/app/tez-site.xml#tez-site.xml</file>
```

- Include the `--hive-import` option in the <command> element.

```
<command>import --connect [jdbc:mysql://db_host:port/database] --
username [user]
 --password [pwd] --driver c[om.mysql.jdbc.Driver] --table [table_name]
--hive-import -m 1 </command>
```

6. Add the following into the `job.properties` file.

```
oozie.use.system.libpath=true
oozie.action.sharelib.for.sqoop=sqoop,hive
```

More information regarding the Sqoop parameters can be found in the Apache documentation at https://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html#_importing_data_into_hive

Example Workflow Action

Replace all sample text in [square brackets] in the example below with the appropriate workflow name, URI, paths, file names, etc. for your configuration.

```
<action name="sqoop-node">
      <sqoop xmlns="uri:oozie:sqoop-action:0.2">
            <job-tracker>${jobTracker}</job-tracker>
            <name-node>${nameNode}</name-node>
            <configuration>
                 <property>
                      <name>mapred.job.queue.name</name>
                      <value>${queueName}</value>
                 </property>
            </configuration>
            <command>import --connect [jdbc:mysql://db_host:port/database] --
username [user]
--password [pwd] --driver [com.mysql.jdbc.Driver] --table [table_name] --hive-
import -m 1</command>
            <file>[/user/dummy/app/hive-site.xml#hive-site.xml]</file>
            <file>[/user/dummy/app/tez-site.xml#tez-site.xml]</file>
      </sqoop>
      <ok to="end"/>
      <error to="killJob"/>
</action>
```