Hortonworks Data Platform

HDP-2.4.3 Release Notes

(September 2, 2016)

docs.cloudera.com

Hortonworks Data Platform: HDP-2.4.3 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 Software Foundation projects that focus on the storage and processing of Big Data, along with operations, security, and governance for the resulting system. This includes Apache Hadoop – which includes MapReduce, Hadoop Distributed File System (HDFS), and Yet Another Resource Negotiator (YARN) – along with Ambari, Falcon, Flume, HBase, Hive, Kafka, Knox, Oozie, Phoenix, Pig, Ranger, Slider, Spark, Sqoop, Storm, Tez, and ZooKeeper. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, training and partner-enablement services. All of our technology is, and will remain, free and open source.

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

Table of Contents

List of Tables

1. HDP 2.4.3 Release Notes

This document provides you with the latest information about Hortonworks Data Platform (HDP) 2.4.3 and its product documentation.

Component Versions

The official Apache versions of most HDP 2.4.3 components are unchanged from HDP 2.4.0.0, with the exception of Spark. Spark is upgraded from 1.6.1+ to 1.6.2. See more details in the New Features section of these notes. All HDP components listed here are official Apache releases of the most recent stable versions available.

The Hortonworks approach is to provide patches only when necessary, to ensure the interoperability of components. Unless you are explicitly directed by Hortonworks Support to take a patch update, each of the HDP components should remain at the following package version levels, to ensure a certified and supported copy of HDP 2.4.3:

Official Apache versions for HDP 2.4.3.

- Apache Accumulo 1.7.0
- Apache Atlas 0.5.0
- Apache Calcite 1.2.0
- Apache DataFu 1.2.0
- Apache Falcon 0.6.1
- Apache Flume 1.5.2
- Apache Hadoop 2.7.1
- Apache HBase 1.1.2
- Apache Hive 1.2.1
- Apache Kafka 0.9.0.1
- Apache Knox 0.6.0
- Apache Mahout 0.9.0+
- Apache Oozie 4.2.0
- Apache Phoenix 4.4.0
- Apache Pig 0.15.0
- Apache Ranger 0.5.2
- Apache Slider 0.80.0
- Apache Solr 5.5.0

- Apache Spark 1.6.2
- Apache Sqoop 1.4.6
- Apache Storm 0.10.0
- Apache Tez 0.7.0
- Apache ZooKeeper 3.4.6

Versions of non-Apache components:

- Cascading 3.0.1
- Hue 2.6.1

1.1. New Features

This section highlights new features in Hortonworks Data Platform (HDP) 2.4.3.

Table 1.1. New Features

Apache Component	Feature
Spark	General availability of Spark 1.6.2

1.2. Unsupported Features

Some features exist within HDP 2.4.3, but Hortonworks does not currently support these specific capabilities.

- Technical Preview Features [2]
- Community Features [3]

1.2.1. Technical Preview Features

The following features are available within HDP 2.4.3 but are not ready for production deployment. Hortonworks encourages you to explore these technical preview features in non-production environments and then provide feedback on your experiences through the Hortonworks Community Forums.

Table 1.2. Technical Previews

Apache Component	Feature
HBase and Phoenix	Introduced in a previous release: • Phoenix Query Server • Phoenix Query Server (PHOENIX-971) • Phoenix-Spark Integration • RPC Throttling • Support for `init.d` scripts

Apache Component	Feature
Hive	Introduced in a previous release: • Hive Streaming • ACID support
Slider	Introduced in a previous release: • Support for Docker-based application packaging (SLIDER-780)
Spark	Introduced in a previous release: • GraphX • SparkR • Spark-HBase connector
YARN	Introduced in a previous release: • Add support for network I/O isolation and scheduling for containers (YARN-2140) • Add cgroup support for disk I/O isolation in NodeManager (YARN-2619)
Zeppelin	Introduced in a previous release: • Zeppelin 0.6.0 • Install and Manage Zeppelin with Ambari 2.2.2 • Improvements to PySpark support in Zeppelin • One-way Zeppelin support for SSL • Zeppelin support for form-based LDAP authentication

1.2.2. Community Features

The following features are developed and tested by the Hortonworks Community but are not officially supported by Hortonworks due to insufficient reliability or incomplete test case coverage, declaration of non-production readiness by the community at large, feature deviation from Hortonworks best practices, and other reasons. Do not use them in your production environments.

Table 1.3. Community Features

Apache Component	Feature
Falcon	Introduced in a previous release: • Prism Server • User Recipes
HBase	Introduced in a previous release: • Use of HDFS data-at-rest encryption instead of HBase Column Family Encryption • Use of memcached as block cache is unsupported (HBASE-13170) • Region assignment without using ZooKeeper • Region size balancing (HBASE-13103)

Apache Component	Feature
HDFS	Introduced in a previous release: • NameNode Federation (HDFS-1052) • Disk Hot Swap (HDFS-1362) • Block-volume device choice (HDFS-1804) • viewFS (HADOOP-7257)
Kafka	Introduced in a previous release: • New Consumer API
Knox	Introduced in a previous release: • Storm REST APIs
Oozie	Introduced in a previous release: • Spark action (OOZIE-1983)
Slider	Introduced in a previous release: • Simplified Application Packaging
Spark	Introduced in a previous release: • Spark Standalone • Spark on Mesos • Jupyter/iPython Notebook • Tech note available for HDP customers who want to use Oozie Spark, which is not supported
YARN	Introduced in a previous release: • Fair Scheduler • MapReduce Eclipse Plug-in • MapReduce Uber AM

1.3. Upgrading to HDP 2.4.3

These Release Notes include abbreviated upgrade instructions; for full upgrade instructions, use the Non-Ambari Upgrade Guide. If you already have HDP 2.4.x installed, upgrading your cluster to HDP 2.4.3 means:

- Keeping the same configuration files you used for HDP 2.4.x
- Keeping the same data and metadata in the same location you used for HDP 2.4.x
- Installing any new components (added for the first time in HDP 2.4.3) side-by-side with existing components

Important

Upgrading from HDP 2.3.6 to 2.4.0 or 2.4.2 is not supported; you can only upgrade HDP 2.3.6 to 2.4.3+.

The following table summarizes HDP 2.4.x-to-2.4.3 upgrade options:

Cluster Management	Supporting Doc	
Cluster managed manually (HDP 2.1 and earlier)	Non-Ambari Upgrade Guide	
Cluster managed manually (HDP 2.3 and later)	Before you begin [6]	
Cluster managed using Apache Ambari 1.7.0	Before you begin [6]	
Cluster managed using Apache Ambari 2.0	Upgrading Ambari Guide	Ambari 2.0 supports a rolling upgrade between HDP 2.2.x and HDP 2.4.3.  **Note** Ambari does not support rolling upgrade between HDP 2.1 and 2.3 for Apache Falcon. Use Configure and Validate Falcon to upgrade this component. When upgrading to HDP 2.4.3 using Ambari, Apache Spark 1.6.x is automatically upgraded to 1.6.2. If you want to return to using your previous version of Spark, use the Spark Manual Downgrade Procedure.
Cluster managed using Apache Ambari 2.1	Upgrading Ambari Guide	Ambari 2.1 supports rolling upgrade between HDP 2.4.x and HDP 2.4.3.  **Note** Ambari does not support rolling upgrade between HDP 2.1 and 2.3 for Falcon. Use Configure and Validate Falcon to upgrade this component. When upgrading to HDP 2.4.3 using Ambari, Spark 1.6.x is automatically upgraded to 1.6.2. If you want to return to using your previous version of Spark, use the Spark Manual Downgrade Procedure.
Cluster managed using Apache Ambari 2.2	Upgrading Ambari Guide	Ambari 2.2 supports rolling upgrade between HDP 2.4.x and HDP 2.4.3.  **Note** Ambari does not support rolling upgrade between HDP 2.1 and

Cluster Management	Supporting Doc	
		2.3 for Falcon. Use Configure and Validate Falcon to upgrade this component. When upgrading to HDP 2.4.3 using Ambari, Spark 1.6.x is automatically upgraded to 1.6.2. If you want to return to using your previous version of Spark, use the Spark Manual Downgrade Procedure.

1.3.1. Before you begin

- Ensure that you know which HDP components need to be upgraded at your installation.
- Decide whether you are going to upgrade using a local repository or a remote repository.

Important

Upgrading from HDP 2.3.6 to 2.4.0 or 2.4.2 is not supported; you can only upgrade HDP 2.3.6 to 2.4.3+.

1.3.2. Upgrade Procedure

Important

Upgrading from HDP 2.3.6 to 2.4.0 or 2.4.2 is not supported; you can only upgrade HDP 2.3.6 to 2.4.3+.

To upgrade your cluster from HDP 2.3.6.0 or HDP 2.4.x to HDP 2.4.3:

1. Download either the hdp.repo or the RPM single repository tarball:

   a. **HDP 2.4.3 hdp.repo file:**

Operating System	Repository Location
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.3.0/hdp.list
Debian 7	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.3.0/hdp.list
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.3.0/hdp.repo
RHEL/CentOS/Oracle LINUX 7	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.3.0/hdp.repo
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.3.0/hdp.repo
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.3.0/hdp.list
Ubuntu 14	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.3.0/hdp.list

   b. **HDP 2.4.3 RPM single repository tarball:**

Operating System	Tarball Location
Debian 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.3.0/HDP-2.4.3.0-debian6-deb.tar.gz
Debian 7	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.3.0/HDP-2.4.3.0-debian7-deb.tar.gz
RHEL/CentOS/Oracle LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.3.0/HDP-2.4.3.0-centos6-rpm.tar.gz
RHEL/CentOS/Oracle LINUX 7	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.3.0/HDP-2.4.3.0-centos7-rpm.tar.gz
SLES 11 SP3/SP4	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.3.0/HDP-2.4.3.0-suse11sp3-rpm.tar.gz
Ubuntu 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.3.0/HDP-2.4.3.0-ubuntu12-deb.tar.gz
Ubuntu 14	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.3.0/HDP-2.4.3.0-ubuntu14-deb.tar.gz

For information about how to install the repositories, see the local repository instructions.

2. **Run an update:**

```
apt-get update
```

3. **Install the HDP 2.4.3 bits:**

Operating System	Commands
RHEL/CentOS/Oracle LINUX	Install HDP 2.4.3 components on relevant nodes, according to the services that run on those hosts: yum install "hadoop_2_4_3_227*" "oozie_2_4_3_227*" "pig_2_4_3_227*" "sqoop_2_4_3_227*" "zookeeper_2_4_3_227*" "hbase_2_4_3_227*" "hive_2_4_3_227*" "tez_2_4_3_227*" "storm_2_4_3_227*" "falcon_2_4_3_227*" "flume_2_4_3_227*" "phoenix_2_4_3_227*" "accumulo_2_4_3_227*" "mahout_2_4_3_227*"
SLES	Install HDP 2.4.3 components on relevant nodes, according to the services that run on those hosts: zypper install "hadoop_2_4_3_227*" "oozie_2_4_3_227*" "pig_2_4_3_227*" "sqoop_2_4_3_227*" "zookeeper_2_4_3_227*" "hbase_2_4_3_227*" "hive_2_4_3_227*" "tez_2_4_3_227*" "storm_2_4_3_227*" "falcon_2_4_3_227*" "flume_2_4_3_227*" "phoenix_2_4_3_227*" "accumulo_2_4_3_227*" "mahout_2_4_3_227*"
Ubuntu/Debian	Install HDP 2.4.3 components on relevant nodes, according to the services that run on those hosts: apt-get install "hadoop_2_4_3_227*" "oozie_2_4_3_227*" "pig_2_4_3_227*" "sqoop_2_4_3_227*" "zookeeper_2_4_3_227*"

Operating System	Commands
	"hbase_2_4_3_227*" "hive_2_4_3_227*" "tez_2_4_3_227*" "storm_2_4_3_227*" "falcon_2_4_3_227*" "flume_2_4_3_227*" "phoenix_2_4_3_227*" "accumulo_2_4_3_227*" "mahout_2_4_3_227*"

4. **Stop all HDP 2.4.x services for your scenario:**

   - Non-Ambari managed clusters

     a. Stop all HDP 2.4.x services using the Stopping HDP Services section of the *HDP Reference Guide*.

   - Ambari 1.7.0-managed clusters

     a. Open Ambari Web.

     b. Browse to **Services**.

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

5. **For all services, switch the active version to HDP 2.4.3.**

   On each host in the cluster, use `hdp-select` to switch all services to the HDP 2.4.3 version:

   ```
   hdp-select set all 2.4.3.0-227
   ```

6. **If you are running Ambari 1.7.0, update the repository base URLs to use the HDP 2.4.3 repositories for HDP and HDP-UTILS:**

   a. Open Ambari Web.

   b. Browse to **Admin > Repositories**.

   c. Edit the Base URLs.

7. **Start all HDP 2.4.3 services, in the following order:**

   a. **Apache ZooKeeper**

   ```
   su - zookeeper 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*.

Your cluster is upgraded. Ensure that your workloads run correctly on this upgraded cluster.

1.3.3. Optional: Spark Manual Upgrade Procedure

(Optional) Upgrade Spark from 1.6.1+ to 1.6.2:

1. As the root user, stop Spark 1.6.0: `su - spark -c "/usr/hdp/current/spark-client/sbin/stop-history-server.sh"`.
2. Remove Spark 1.6.1+: `yum erase "spark*"`.
3. Add the node where you want Spark 1.6.2 History Server to run:
   a. `su - root`
   b. `wget -nv http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos6/2.x/BUILDS/2.4.3.0-227/hdpbn.repo -O /etc/yum.repos.d/Spark141TP.repo`
   c. `yum install spark_2_4_3_0_227-master -y`
   d. To use Python: `yum install spark_2_4_3_0_227-python`
   e. `conf-select create-conf-dir --package spark --stack-version 2.4.3.0-227 --conf-version 0`
   f. `cp /etc/spark/2.4.3.0-227/0/* /etc/spark/2.4.3.0-227/0/`
   g. `conf-select set-conf-dir --package spark --stack-version 2.4.3.0-227 --conf-version 0`
   h. `hdp-select set spark-client 2.4.3.0-227`
   i. `hdp-select set spark-historyserver 2.4.3.0-227`

4. Validate the Spark installation by running SparkPI as the spark user, as in the following example:

   a. `su - spark -c "cd /usr/hdp/current/spark-client"`

   b. `./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`

5. Restart Spark on YARN in either yarn-cluster mode or yarn-client mode:

   - yarn-cluster mode: `./usr/hdp/current/spark-client/bin/spark-submit --class path.to.your.Class --master yarn-cluster [options] <app jar> [app options]`
   - yarn-client mode: `./usr/hdp/current/spark-client/bin/spark-shell --master yarn-client`

1.3.4. Optional: Spark Manual Downgrade Procedure

When upgrading to HDP 2.4.3 using Ambari, Spark 1.6.1 is automatically upgraded to 1.6.2. However, if you want to return to using 1.6.1, complete the following steps:

1. Remove Spark 1.6.2 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`

2. Manually install Spark 1.6.x by following the directions in HDP 2.3.0 Installing HDP Manually: Installing and Configuring Apache Spark.

1.4. Behavioral Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.4.3, behavioral changes affect the following Hadoop components:

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
BUG-52381	Apache Spark		Spark Kafka streaming jars dependency	Developers using Spark Streaming with Apache Kafka on a Kerberos-enabled cluster should use the HDP spark-streaming-kafka jar file and associated jar files. For more information, see Using Spark Streaming with Kafka on a Kerberos-Enabled Cluster in the *Spark Guide*.

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
BUG-55056	Apache YARN	YARN-4414	Nodemanager connection errors are retried at multiple levels.	**Scenario:** During NodeManager restart, all YARN clients connected to NodeManager need to retry the connection to NodeManager. We formerly had a two layer retry for YARN Client connection with NodeManager: one was at the RPC layer, and the other was at the NodeManager proxy layer. This caused the log of retry activities to expand logarithmically. After this fix, we consolidated the RPC layer retry and NodeManager proxy retry; the retry log no longer has duplicate entries.

1.5. Apache Patch Information

The following subsections list patches in each HDP 2.4.3 component beyond what was fixed in the base version of the Apache component.

See the Fixed Issues section for customer-reported issues.

- Hadoop [12]
- Accumulo [19]
- Atlas [19]
- Calcite [19]
- Falcon [19]
- Flume [20]
- HBase [20]
- Hive [22]
- Kafka [27]
- Knox [27]
- Mahout [28]

1.5.1. Hadoop

HDP 2.4.3 provides the following Apache patches:

- HADOOP-7817: RawLocalFileSystem.append() should give FSDataOutputStream with accurate .getPos().
- HADOOP-9477: Add posixGroups support for LDAP groups mapping service.
- HADOOP-10965: Print fully qualified path in CommandWithDestination error messages.
- HADOOP-11361: Fix a race condition in MetricsSourceAdapter.updateJmxCache.
- HADOOP-11404: Clarify the "expected client Kerberos principal is null" authorization message.
- HADOOP-11491: HarFs incorrectly declared as requiring an authority.
- HADOOP-12001: Fixed LdapGroupsMapping to include configurable Posix UID and GID attributes during the search.
- HADOOP-12076: Incomplete Cache Mechanism in CredentialProvider AP.
- HADOOP-12189: Improve CallQueueManager#swapQueue to make queue elements drop nearly impossible.
- HADOOP-12291: Add support for nested groups in LdapGroupsMapping.
- HADOOP-12296: When setnetgrent returns 0 in Linux, exception should be thrown.
- HADOOP-12345: Pad hostname correctly in CredentialsSys.java.
- HADOOP-12469: distcp should not ignore the ignoreFailures option.

- HADOOP-12472: Make GenericTestUtils.assertExceptionContains robust.
- HADOOP-12568: Update core-default.xml to describe posixGroups support.
- HADOOP-12604: Exception may be swallowed in KMSClientProvide.
- HADOOP-12622: Improve the loggings in RetryPolicies and RetryInvocationHandler.
- HADOOP-12636: Prevent ServiceLoader failure init for unused FileSystems.
- HADOOP-12659: Incorrect usage of config parameters in token manager of KMS.
- HADOOP-12672: RPC timeout should not override IPC ping interval.
- HADOOP-12716: KerberosAuthenticator#doSpnegoSequence use incorrect class to determine isKeyTab in JDK8.
- HADOOP-12751: While using kerberos Hadoop incorrectly assumes names with '@' to be non-simple.
- HADOOP-12772: NetworkTopologyWithNodeGroup.getNodeGroup() can loop infinitely for invalid 'loc' values.
- HADOOP-12782: Faster LDAP group name resolution with ActiveDirectory.
- HADOOP-12793: Write a new group mapping service guide.
- HADOOP-12810: FileSystem#listLocatedStatus causes unnecessary RPC call.
- HADOOP-12828: Print user when services are started.
- HADOOP-12831: LocalFS/FSOutputSummer NPEs in constructor if bytes per checksum set to 0.
- HADOOP-12847: Hadoop daemonlog should support https and SPNEGO for Kerberized cluster.
- HADOOP-12893: Update LICENSE.txt and NOTICE.txt.
- HADOOP-12895: SSLFactory#createSSLSocketFactory exception message is wrong.
- HADOOP-12901: Add warning log when KMSClientProvider cannot create a connection to the KMS server.
- HADOOP-12906: AuthenticatedURL should convert a 404/Not Found into an FileNotFoundException.
- HADOOP-12916: Allow RPC scheduler/callqueue backoff using response times.
- HADOOP-12962: KMS key names are incorrectly encoded when creating key.
- HADOOP-12985: Support MetricsSource interface for DecayRpcScheduler Metrics.
- HADOOP-12994: Specify PositionedReadable, add contract tests, fix problems.

- HADOOP-13030: Handle special characters in passwords in KMS startup script.
- HADOOP-13042: Restore lost leveldbjni LICENSE and NOTICE changes.
- HADOOP-13052: ChecksumFileSystem mishandles crc file permissions.
- HADOOP-13105: Support timeouts in LDAP queries in LdapGroupsMapping.
- HADOOP-13155: Implement TokenRenewer to renew and cancel delegation tokens in KMS.
- HADOOP-13159: Fix potential NPE in Metrics2 source for DecayRpcScheduler.
- HADOOP-13179: GenericOptionsParser is not thread-safe because commons-cli OptionBuilder is not thread-safe.
- HADOOP-13197: Add non-decayed call metrics for DecayRpcScheduler.
- HADOOP-13251: Authenticate with Kerberos credentials when renewing KMS delegation token.
- HADOOP-13255: KMSClientProvider should check and renew tgt when doing delegation token operations.
- HADOOP-13263: Reload cached groups in background after expiry.
- HADOOP-13270: BZip2CompressionInputStream finds the same compression marker twice in corner case, causing duplicate data blocks.
- HADOOP-13285: DecayRpcScheduler MXBean should only report decayed CallVolumeSummary.
- HADOOP-13350: Additional fix to LICENSE and NOTICE.
- HADOOP-13351: TestDFSClientSocketSize buffer size tests are flaky.
- HADOOP-13362: DefaultMetricsSystem leaks the source name when a source unregisters.
- HADOOP-13403: AzureNativeFileSystem rename/delete performance improvements.
- HADOOP-13459: Hadoop-Azure runs several test cases repeatedly, causing unnecessarily long running time.
- HADOOP-13513: Java 1.7 support for org.apache.hadoop.fs.azure test cases.
- HDFS-2043: TestHFlush failing intermittently.
- HDFS-6407: Add sorting and pagination in the datanode tab of the NN Web UI.
- HDFS-7166: SbNN Web UI shows #Under replicated blocks and #pending deletion blocks.
- HDFS-7314: When the DFSClient lease cannot be renewed, abort open-for-write files rather than the entire DFSClien.
- HDFS-7452: Skip StandbyException log for getCorruptFiles().

- HDFS-7597: DelegationTokenIdentifier should cache the TokenIdentifier to UGI mapping.
- HDFS-7978: Add LOG.isDebugEnabled() guard for some LOG.debug().
- HDFS-8546: Use try with resources in DataStorage and Storage.
- HDFS-8578: On upgrade, Datanode should process all storage/data dirs in parallel.
- HDFS-8581: ContentSummary on / skips further counts on yielding loc.
- HDFS-8758: Implement the continuation library in libhdfspp.
- HDFS-8816: Improve visualization for the Datanode tab in the NN UI.
- HDFS-8831: Trash Support for deletion in HDFS encryption zone.
- HDFS-8844: TestHDFSCLI does not cleanup the test directory.
- HDFS-8845: DiskChecker should not traverse the entire tree.
- HDFS-8964: When validating the edit log, do not read at or beyond the file offset that is being written.
- HDFS-9259: Make SO_SNDBUF size configurable at DFSClient side for hdfs write scenario.
- HDFS-9317: Document fsck -blockId and -storagepolicy options.
- HDFS-9365: Balancer does not work with the HDFS-6376 HA setup.
- HDFS-9395: Make HDFS audit logging consistent.
- HDFS-9412: getBlocks occupies FSLock and takes too long to complete.
- HDFS-9415: Document dfs.cluster.administrators and dfs.permissions.superusergroup.
- HDFS-9470: Encryption zone on root not loaded from fsimage after NN restart.
- HDFS-9516: Truncate file fails with data dirs on multiple disks.
- HDFS-9521: TransferFsImage.receiveFile should account and log separate times for image download and fsync to disk.
- HDFS-9530: ReservedSpace is not cleared for abandoned Block.
- HDFS-9533: seen_txid in the shared edits directory is modified during bootstrapping.
- HDFS-9555: LazyPersistFileScrubber should still sleep if there are errors in the clear progress.
- HDFS-9566: Remove expensive 'BlocksMap#getStorages(Block b, final DatanodeStorage.State state)' method.
- HDFS-9569: Log the name of the fsimage being loaded for better supportability.
- HDFS-9584: NPE in distcp when ssl configuration file does not exist in class path.

- HDFS-9608: Disk IO imbalance in HDFS with heterogeneous storages.
- HDFS-9624: DataNode start slowly due to the initial DU command operations.
- HDFS-9629: Update the footer of Web UI to show year 201.
- HDFS-9634: WebHDFS client side exceptions don't provide enough details.
- HDFS-9648: TestStartup.testImageChecksum is broken by HDFS-9569s message change.
- HDFS-9654: Code refactoring for HDFS-8578.
- HDFS-9670: DistCp throws NPE when source is root.
- HDFS-9690: ClientProtocol.addBlock is not idempotent after HDFS-8071.
- HDFS-9713: DataXceiver#copyBlock should return if block is pinned.
- HDFS-9715: Check storage ID uniqueness on datanode startup.
- HDFS-9721: Allow Delimited PB OIV tool to run upon fsimage that contains INodeReference.
- HDFS-9730: Storage ID update does not happen when there is a layout change.
- HDFS-9748: Avoid duplication in pendingReplications when addExpectedReplicasToPending is called twice.
- HDFS-9799: Reimplement getCurrentTrashDir to remove incompatibility.
- HDFS-9812: Streamer threads leak if failure happens when closing DFSOutputStream.
- HDFS-9844: Correct path creation in getTrashRoot to handle root dir.
- HDFS-9871: "Bytes Being Moved" -ve(-1 B) when cluster was already balanced.
- HDFS-9874: Long living DataXceiver threads cause volume shutdown to block.
- HDFS-9881: DistributedFileSystem#getTrashRoot returns incorrect path for encryption zones.
- HDFS-9882: Add heartbeatsTotal in Datanode metrics.
- HDFS-9902: Support different values of dfs.datanode.du.reserved per storage type.
- HDFS-9905: WebHdfsFileSystem#runWithRetry should display original stack trace on error.
- HDFS-9941: Do not log StandbyException on NN, other minor logging fixed.
- HDFS-10182: Hedged read might overwrite user's buf.
- HDFS-10186: DirectoryScanner Improve logs by adding full path of both actual and expected block directories. .

- HDFS-10189: PacketResponder#toString should include the downstreams for PacketResponderType.HAS_DOWNSTREAM_IN_PIPELINE.
- HDFS-10217: Show 'blockScheduled' tooltip in datanodes table.
- HDFS-10228: TestHDFSCLI fails.
- HDFS-10235: Last contact for Live Nodes should be relative time.
- HDFS-10253: Fix TestRefreshCallQueue failure.
- HDFS-10264: Logging improvements in FSImageFormatProtobuf.Saver.
- HDFS-10271: Extra bytes are getting released from reservedSpace for append.
- HDFS-10277: PositionedReadable test testReadFullyZeroByteFile failing in HDFS.
- HDFS-10291: TestShortCircuitLocalRead failing.
- HDFS-10319: Balancer should not try to pair storages with different types.
- HDFS-10324: Trash directory in an encryption zone should be pre-created with correct permissions.
- HDFS-10329: Bad initialisation of StringBuffer in RequestHedgingProxyProvider.
- HDFS-10344: DistributedFileSystem#getTrashRoots should skip encryption zone that does not have .Trash.
- HDFS-10360: DataNode may format directory and lose blocks if current/VERSION is missing.
- HDFS-10417: Improve error message from checkBlockLocalPathAccess.
- HDFS-10440: Improve DataNode web UI.
- HDFS-10458: getFileEncryptionInfo should return quickly for non-encrypted cluster.
- HDFS-10468: HDFS read ends up ignoring an interrupt.
- HDFS-10469: Add number of active xceivers to datanode metrics.
- HDFS-10481: HTTPFS server should correctly impersonate as end user to open file.
- HDFS-10493: Add links to datanode web UI in namenode datanodes page.
- HDFS-10508: DFSInputStream should set thread's interrupt status after catching InterruptException from sleep.
- HDFS-10556: DistCpOptions should be validated automatically.
- HDFS-10617: PendingReconstructionBlocks.size() should be synchronized.
- HDFS-10642: TestLazyPersistReplicaRecovery#testDnRestartWithSavedReplicas fails intermittently.

- HDFS-10688: BPServiceActor may run into a tight loop for sending block report when hitting IOException.
- MAPREDUCE-6543: Migrate MR Client test cases part 2.
- MAPREDUCE-6637: Testcase Failure TestFileInputFormat.testSplitLocationInfo.
- YARN-2104: Scheduler queue filter failed to work because index of queue column changed.
- YARN-4325: Nodemanager log handlers fail to send finished/failed events in some cases.
- YARN-4983: JVM and UGI metrics disappear after RM transitioned to standby mod.
- YARN-5076: YARN web interfaces lack XFS (Cross-Frame Script) protection.
- YARN-5112: Excessive log warnings for directory permission issue on NM recovery.

HDP 2.4.0 provided the following Apache patches:

- HADOOP-10406: TestIPC.testIpcWithReaderQueuing may fail.
- HADOOP-12551: Introduce FileNotFoundException for WASB FileSystem API.
- HADOOP-12608: Fix exception message in WASB when connecting with anonymous credential.
- HADOOP-12678: Handle empty rename pending metadata file during atomic rename in redo path.
- HDFS-8729: Fix TestFileTruncate#testTruncateWithDataNodesRestartImmediately which occasionally failed.
- HDFS-9358: TestNodeCount#testNodeCount timed out.
- HDFS-9406: FSImage may get corrupted after deleting snapshot.
- HDFS-9672: o.a.h.hdfs.TestLeaseRecovery2 fails intermittently.
- MAPREDUCE-6566: Add retry support to mapreduce CLI tool.
- MAPREDUCE-6618: YarnClientProtocolProvider leaking the YarnClient thread.
- MAPREDUCE-6621: Memory Link in JobClient#submitJobInternal().
- YARN-3480: Remove attempts that are beyond max-attempt limit from state store.
- YARN-4309: Add container launch related debug information to container logs when a container fails.
- YARN-4497: RM might fail to restart when recovering apps whose attempts are missing.
- YARN-4565: Sometimes when sizeBasedWeight FairOrderingPolicy is enabled, under stress appears that cluster is virtually in deadlock.

- YARN-4584: RM startup failure when AM attempts greater than max-attempts.
- YARN-4625: ApplicationSubmissionContext and ApplicationSubmissionContextInfo more consistent.

1.5.2. Accumulo

HDP 2.4.3 provides Accumulo 1.7.0 with no additional Apache patches included.

HDP 2.4.2 provided Accumulo 1.7.0 and the following Apache patches:

- ACCUMULO-4080 TabletServers should be less aggressively "monitoring RO filesystems".
- ACCUMULO-4185 Use SYNC durability and expect the update to be durable in ProxyDurabilityIT.

HDP 2.4.0 provided Accumulo 1.7.0 and the following Apache patches:

- ACCUMULO-4135 Add impersonation configuration keys which don't put the principal in the key.

1.5.3. Atlas

HDP 2.4.3 provides Atlas 0.5.0 with no additional Apache patches included.

HDP 2.4.2 provided Atlas 0.5.0 with no additional Apache patches included.

HDP 2.4.0 provided Atlas 0.5.0 and the following Apache patches:

- ATLAS-448: Hive IllegalArgumentException with Atlas hook enabled on SHOW TRANSACTIONS AND SHOW COMPACTIONS.

1.5.4. Calcite

HDP 2.4.3 provides Calcite 1.2.0 with no additional Apache patches included.

HDP 2.4.2 provides Calcite 1.2.0 and the following Apache patches:

- CALCITE-1128: Support addBatch()/executeBatch() in remote driver.
- CALCITE-1180: Support clearBatch() in remote driver.

HDP 2.4.0 provided Calcite 1.2.0 and the following Apache patch:

- CALCITE-984: NPE in ErrorResponse construction.

1.5.5. Falcon

HDP 2.4.3 provides Falcon 0.6.1 and the following Apache patches:

- FALCON-1880: Add –skipcrccheck to distcp options for HiveDR when using Distcp to copy the file from Encryption to Encryption.

- FALCON-1894: HDFS Data replication cannot be initiated independent of Oozie server location.
- FALCON-2007: Fix Hive DR Replication failing with "Can not create a Path from a null string."
- FALCON-2072: Hive2 URLs in Falcon should allow additional configuration elements in the URL.
- FALCON-2075: Falcon HiveDR tasks do not report progress and can get killed.

HDP 2.4.2 provided Falcon 0.6.1 with no additional Apache patches included.

HDP 2.4.0 provided Falcon 0.6.1 with no additional Apache patches included.

1.5.6. Flume

HDP 2.4.3 provides Flume 1.5.2 with no additional Apache patches included.

HDP 2.4.2 provided Flume 1.5.2 and the following Apache patch:

- FLUME-2889 Fixes to DateTime computations.

HDP 2.4.0 provided Flume 1.5.2 and the following Apache patch:

- FLUME-2865 Upgrade thrift version (0.9.2).

1.5.7. HBase

HDP 2.4.3 provides HBase 1.1.2 and the following Apache patch:

- HBASE-13153: Bulk Loaded HFile Replication.
- HBASE-14252: RegionServers fail to start when setting hbase.ipc.server.callqueue.scan.ratio to 0.
- HBASE-14460: [Perf Regression] Merge of MVCC and SequenceId (HBASE-8763) slowed Increments, CheckAndPuts, batch operations.
- HBASE-14535: Integration test for rpc connection concurrency / deadlock testing.
- HBASE-14650: Re-enable TestNamespaceAuditor.
- HBASE-14798: NPE reporting server load causes regionserver abort; causes TestAcidGuarantee to fail.
- HBASE-15295: MutateTableAccess.multiMutate() does not get high priority causing a deadlock.
- HBASE-15325: ResultScanner allowing partial result will miss the rest of the row if the region is moved between two rpc requests.
- HBASE-15471: Added in Priority, General, and Replication queue sizes for RegionServers to the Web UI Changed UI labels so that queue "size" refers to size in bytes and queue "length" refers to number of items in queue.

- HBASE-15925: Provide default values for hadoop compatibility module related properties that match default hadoop profile.
- HBASE-15946: Eliminate possible security concerns in RS web UI's store file metrics.
- HBASE-15957: RpcClientImpl.close never ends in some circumstances.
- HBASE-16056: Procedure v2 - fix master crash for FileNotFound.

HDP 2.4.2 provided HBase 1.1.2 and the following Apache patches:

- HBASE-15592: Print Procedure WAL content.
- HBASE-7171: Initial web UI for region/memstore/storefiles details.
- HBASE-15144: Procedure v2 - Web UI displaying Store state.
- HBASE-15622: Superusers does not consider the keytab credentials.
- HBASE-14976: Add RPC call queues to the web ui.
- HBASE-15811: Batch Get after batch Put does not fetch all Cells We were not waiting on all executors in a batch to complete. The test for no-more-executors was damaged by the 0.99/0.98.4 fix "HBASE-11403 Fix race conditions around Object#notify".
- HBASE-15325: ResultScanner allowing partial result will miss the rest of the row if the region is moved between two rpc requests.
- HBASE-14674: Rpc handler / task monitoring seems to be broken after 0.98.
- HBASE-13829: Add more ThrottleType.
- HBASE-15471: Added in Priority, General, and Replication queue sizes for RegionServers to the Web UI Changed UI labels so that queue "size" refers to size in bytes and queue "length" refers to number of items in queue.
- HBASE-14867: SimpleRegionNormalizer needs to have better heuristics to trigger merge operation.
- HBASE-14805: status should show the master in shell.
- HBASE-15720: Print row locks at the debug dump page.
- HBASE-14941: locate_region shell command.
- HBASE-14283: Reverse scan doesn’t work with HFile inline index/bloom blocks.
- HBASE-14307: Incorrect use of positional read API in HFileBlock.
- HBASE-14812: Fix ResultBoundedCompletionService deadlock.
- HBASE-14883: TestSplitTransactionOnCluster#testFailedSplit flakey.
- HBASE-15220: Change two logs in SimpleRegionNormalizer to INFO level.

- HBASE-15221: Reload the cache on re-tried puts in HTableMultiplexer and adds a close() method to HTableMultiplexer.
- HBASE-15232: Handle region location cache management in AsyncProcess for multi()'s.
- HBASE-15621: Suppress Hbase SnapshotHFile cleaner error messages when a snapshot is going on.

HDP 2.4.0 provided HBase 1.1.2 and the following Apache patches:

- HBASE-14107: Administrative Task: Provide an API to List all procedures.
- HBASE-14108: Administrative Task: provide an API to abort a procedure.
- HBASE-14432: Enforce ACL on procedure admin tasks.
- HBASE-14468: Compaction improvements: FIFO compaction policy.
- HBASE-14471: Thrift - HTTP Error 413 full HEAD if using kerberos authentication.
- HBASE-14487: Shell command to list all procedures.
- HBASE-14488: Shell command to abort a procedure.
- HBASE-14575: Relax region read lock for compactions.
- HBASE-14761: Deletes with and without visibility expression do not delete the matching mutation.
- HBASE-14866: VerifyReplication should use peer configuration in peer connection.
- HBASE-14928: Start row should be set for query through HBase REST gateway involving globbing option.
- HBASE-14963: Remove Guava dependency from HBase client code.
- HBASE-14987: Compaction marker whose region name doesn't match current region's needs to be handled.
- HBASE-15014: Fix filterCellByStore in WALsplitter is awful for performance.
- HBASE-15022: replication_admin.rb throws undefined method 'getZooKeeperClusterKey' for ZKUtil.
- HBASE-15035: bulkloading hfiles with tags that require splits do not preserve tags.

1.5.8. Hive

HDP 2.4.3 provides Hive 1.2.1 and the following Apache patches:

- HIVE-10233: Hive on tez memory manager for grace hash join.
- HIVE-10631: create_table_core method has invalid update for Fast Stats.

- HIVE-10632: Make sure TXN_COMPONENTS gets cleaned up if table is dropped before compaction.
- HIVE-11037: HiveOnTez make explain user level = true as default.
- HIVE-11118: Load data query should validate file formats with destination tables.
- HIVE-11164: WebHCat should log contents of HiveConf on startup.
- HIVE-11498: HIVE Authorization v2 should not check permission for dummy entity.
- HIVE-11716: Reading ACID table from non-acid session should raise an error.
- HIVE-11891: Add basic performance logging to metastore calls.
- HIVE-12619: Switching the field order within an array of structs causes the query to fail.
- HIVE-12768: Thread safety binary sortable serde decimal deserialization.
- HIVE-13043: Reload function has no impact to function registry.
- HIVE-13051: Deadline class has numerous issues.
- HIVE-13175: Disallow making external tables transactional.
- HIVE-13191: DummyTable map joins mix up columns between tables.
- HIVE-13294: AvroSerde leaks the connection in a case when reading schema from a url.
- HIVE-13299: Column Names trimmed of leading and trailing spaces.
- HIVE-13303: Spill to YARN directories, not tmp, when available.
- HIVE-13313: TABLESAMPLE ROWS feature broken for Vectorization.
- HIVE-13338: Differences in vectorized_casts.q output for vectorized and non-vectorized runs.
- HIVE-13358: Stats state is not captured correctly turn off stats optimizer for sampled table.
- HIVE-13373: Use most specific type for numerical constants.
- HIVE-13390: Add more test to ZK service discovery using MiniHS2 (SSL Certificate Hostname vulnerability fix).
- HIVE-13392: Disable speculative execution for ACID Compactor.
- HIVE-13394: Analyze table fails in tez on empty partitions/files/tables.
- HIVE-13405: Fix Connection Leak in OrcRawRecordMerger.
- HIVE-13561: HiveServer2 is leaking ClassLoaders when add jar / temporary functions are used.

- HIVE-13592: Metastore calls map is not thread safe.
- HIVE-13596: HS2 should be able to get UDFs on demand from metastore.
- HIVE-13609: Fix UDTFs to allow local fetch task to fetch rows forwarded by GenericUDTF.close().
- HIVE-13618: Trailing spaces in partition column will be treated differently.
- HIVE-13619: Bucket map join plan is incorrect.
- HIVE-13632: Hive failing on insert empty array into parquet table.
- HIVE-13660: Vectorizing IN expression with list of columns throws java.lang.ClassCastException ExprNodeColumnDesc cannot be cast to ExprNodeConstantDesc.
- HIVE-13691: No record with CQ_ID=0 found in COMPACTION_QUEUE.
- HIVE-13693: Multi-insert query drops Filter before file output when there is a.val <> b.val.
- HIVE-13729: FileSystem$Cache leaks in FileUtils.checkFileAccessWithImpersonation.
- HIVE-13730: Avoid double spilling the same partition when memory threshold is set very low.
- HIVE-13948: Incorrect timezone handling in Writable results in wrong dates in queries.
- HIVE-13991: Union All on view fail with no valid permission on underneath table.
- HIVE-14027: NULL values produced by left outer join do not behave as NULL.
- HIVE-14045: (Vectorization) Add missing case for BINARY in VectorizationContext.getNormalizedName method.
- HIVE-14132: Don't fail config validation for removed configs.
- HIVE-14147: Hive PPD might remove predicates when they are defined as a simple expression e.g. WHERE 'a'.
- HIVE-14167: Use work directories provided by Tez instead of directly using YARN local dirs.
- HIVE-14178: Hive::needsToCopy should reuse FileUtils::equalsFileSystem.
- HIVE-14262: Inherit writetype from partition WriteEntity for table WriteEntity.
- HIVE-14277: Disable StatsOptimizer for all ACID tables.
- HIVE-14278: Migrate TestHadoop23SAuthBridge.java from Unit3 to Unit4.
- HIVE-14279: fix mvn test TestHiveMetaStore.testTransactionalValidation.
- HIVE-14282: HCatLoader ToDate() exception with hive partition table ,partitioned by column of DATE datatype.

- HIVE-14348: Alter Table Exchange Partition NullPointerException and authorization issues.
- HIVE-9605: Remove parquet nested objects from wrapper writable objects .

HDP 2.4.2 provided Hive 1.2.1 and the following Apache patches:

- HIVE-11097: HiveInputFormat uses String.startsWith to compare splitPath and PathToAliases.
- HIVE-11333: ColumnPruner prunes columns of UnionOperator that should be kept.
- HIVE-11470: NPE in DynamicPartFileRecordWriterContainer on null part-keys.
- HIVE-11816: Upgrade groovy to 2.4.4.
- HIVE-11841: KeyValuesInputMerger creates huge logs.
- HIVE-12022: NPE in SARG with timestamp cast.
- HIVE-12064: prevent transactional=false.
- HIVE-12165: wrong result when hive.optimize.sampling.orderby=true with some aggregate functions.
- HIVE-12189: The list in pushdownPreds of ppd.ExprWalkerInfo should not be allowed to grow very large.
- HIVE-12352: CompactionTxnHandler.markCleaned() may delete too much.
- HIVE-12684: NPE in stats annotation when all values in decimal column are NULLs.
- HIVE-12742: NULL table comparison within CASE does not work as previous hive versions.
- HIVE-12749: Constant propagate returns string values in incorrect format.
- HIVE-12784: Group by SemanticException: Invalid column reference.
- HIVE-12879: RowResolver of Semijoin not updated in CalcitePlanner.
- HIVE-12947: SMB join in tez has ClassCastException when container reuse is on.
- HIVE-13016: ORC FileDump recovery utility fails in Windows.
- HIVE-13092: Vectorized java.lang.ClassCastException: org.apache.hadoop.hive.serde2.typeinfo.ListTypeInfo cannot be cast to org.apache.hadoop.hive.serde2.typeinfo.PrimitiveTypeInfo.
- HIVE-13129: CliService leaks HMS connection.
- HIVE-13144: HS2 can leak ZK ACL objects when curator retries to create the persistent ephemeral node.
- HIVE-13151: Clean up UGI objects in FileSystem cache for transactions.

- HIVE-13174: Remove Vectorizer noise in logs.
- HIVE-13232: Aggressively drop compression buffers in ORC OutStreams.
- HIVE-13233: Use min and max values to estimate better stats for comparison operators.
- HIVE-13263: Vectorization: Unable to vectorize regexp_extract/regexp_replace " Udf: GenericUDFBridge, is not supported".
- HIVE-13285: Orc concatenation may drop old files from moving to final path.
- HIVE-13298: nested join support causes undecipherable errors in SemanticAnalyzer.
- HIVE-13325: Excessive logging when ORC PPD fails type conversions.

HDP 2.4.0 provided Hive 1.2.1 and the following Apache patches:

- HIVE-11141: Improve RuleRegExp when the Expression node stack gets huge.
- HIVE-11291: Avoid allocation storm while doing rule matching on operator/expression trees.
- HIVE-11310: Avoid expensive AST tree conversion to String for expressions in WHERE clause.
- HIVE-11311: Avoid dumping AST tree String in Explain unless necessary.
- HIVE-11328: Avoid String representation of expression nodes in ConstantPropagateProcFactory unless necessary.
- HIVE-11330: Add early termination for recursion in StatsRulesProcFactory.evaluateExpression.
- HIVE-11397: Parse Hive OR clauses as they are written into the AST.
- HIVE-11405: Add early termination for recursion in StatsRulesProcFactory.evaluateExpression for OR expression.
- HIVE-11406: Vectorization: StringExpr::compare() == 0 is bad for performance.
- HIVE-11981: ORC Schema Evolution Issues (Vectorized, ACID, and Non-Vectorized).
- HIVE-12625: Backport to branch-1 HIVE-11981 ORC Schema Evolution Issues (Vectorized, ACID, and Non-Vectorized).
- HIVE-12660: HS2 memory leak with .hiverc file use.
- HIVE-12706: Incorrect output from from_utc_timestamp()/to_utc_timestamp when local timezone has DST.
- HIVE-12728: Apply DDL restrictions for ORC schema evolution.
- HIVE-12766: TezTask does not close DagClient after execution.
- HIVE-12799: Always use Schema Evolution for ACID.

1.5.9. Kafka

HDP 2.4.3 provides Kafka 0.9.0.1 with no additional Apache patches included.

HDP 2.4.2 provided Kafka 0.9.0.1 and the following Apache patches:

- KAFKA-2803: Add hard bounce system test for Kafka Connect.
- KAFKA-2812: Improve consumer integration tests.
- KAFKA-2862: Fix MirrorMaker's message.handler.args description.
- KAFKA-2872: Unite sink nodes with parent nodes in addSink.
- KAFKA-2877: Handle request timeout in sync group.
- KAFKA-2878: Guard against OutOfMemory in Kafka broker.
- KAFKA-2879: Make MiniKDC test service slightly more generic.
- KAFKA-2880: Consumer should handle disconnect/timeout for metadata requests.
- KAFKA-2881: Improve Consumer Configs and API Documentation.
- KAFKA-2882: Add constructor cache for Snappy and LZ4 Output/Input streams in Compressor.java.
- KAFKA-2892: Consumer Docs Use Wrong Method.
- KAFKA-2899: Improve logging when unexpected exceptions thrown in reading local log.
- KAFKA-2906: Fix Connect javadocs, restrict only to api subproject, and clean up javadoc warnings.
- KAFKA-2913: Missing partition check when removing groups from cache.
- KAFKA-2942: Inadvertent auto-commit when pre-fetching can cause message loss.
- KAFKA-2950: Fix performance regression in the producer.

HDP 2.4.0 provided Kafka 0.9.0 with no additional Apache patches included.

1.5.10. Knox

HDP 2.4.3 provided Knox 0.6.0 and the following Apache patches:

- KNOX-556: Knox needs better diagnostics for Keystore failures.

HDP 2.4.2 provided Knox 0.6.0 and the following Apache patches:

- KNOX-677: Upgrade to latest Groovy.
- KNOX-695: Expose configuration of HttpClient's connection and socket timeout settings.

HDP 2.4.0 provided Knox 0.6.0 and the following Apache patch:

- KNOX-647: Rename LDAP artifacts from test to demo.

1.5.11. Mahout

In HDP-2.3.x and 2.4.x, instead of shipping a specific Apache release of Mahout, we synchronized to a particular revision point on Apache Mahout trunk. This revision point is after the 0.9.0 release, but before the 0.10.0 release. This provides a large number of bug fixes and functional enhancements over the 0.9.0 release, but provides a stable release of the Mahout functionality before the complete conversion to new Spark-based Mahout in 0.10.0.

The revision point chosen for Mahout in HDP 2.3.x and 2.4.x is from the "mahout-0.10.x" branch of Apache Mahout, as of 19 December 2014, revision 0f037cb03e77c096 in GitHub.

In addition, we have provided the following patches:

- MAHOUT-1493 Port Naive Bayes to Scala DSL.
- MAHOUT-1589 mahout.cmd has duplicated content.

1.5.12. Oozie

HDP 2.4.3 provides Oozie 4.2.0 and the following Apache patches:

- OOZIE-1837: LauncherMainHadoopUtils sensitive to clock skew.
- OOZIE-2350: Package changes for release.
- OOZIE-2614: Add XFrame Filter for REST APIs.

HDP 2.4.2 provided Oozie 4.2.0 with no additional Apache patches included.

HDP 2.4.0 provided Oozie 4.2.0 with no additional Apache patches included.

1.5.13. Phoenix

HDP 2.4.3 provides Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- PHOENIX-1478: Can't upsert value of 127 into column of type unsigned tinyint.
- PHOENIX-2125: ORDER BY on full PK on salted table does not work.
- PHOENIX-2568: phoenix-spark client-spark JAR conflicts with some distributions.
- PHOENIX-2876: Using aggregation function in ORDER BY.
- PHOENIX-2878: CURRENT_TIME fails to provide correct time when projected table is using (#6).
- PHOENIX-2892: Scan for pre-warming the block cache for 2ndary index should be removed

- PHOENIX-2940: Query the stats table and cache stats in the client.
- PHOENIX-3037: Setup proper security context in compaction/split coprocessor hooks.

HDP 2.4.2 provided Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

- PHOENIX-2601: Query result is incorrect when both index hint and limit are used.
- PHOENIX-2817: Phoenix-Spark plugin doesn't work in secured environment.

HDP 2.4.0 provided Phoenix 4.4.0-HBase-1.1 and the following Apache patch:

- PHOENIX-2608: Incompatibility between Jackson1 version shipped with Phoenix, YARN.

1.5.14. Pig

HDP 2.4.3 provides Pig 0.15.0 and the following Apache patches:

- PIG-4587: Applying isFirstReduceOfKey for Skewed left outer join skips records.
- PIG-4714: Improve logging across multiple components with callerId (Pig launch throws ATS errors).
- PIG-4873: InputSplit.getLocations return null and result a NPE in Pig.

HDP 2.4.2 provided Pig 0.15.0 and the following Apache patches:

- PIG-4690: Union with self replicate join will fail in Tez.
- PIG-4760: TezDAGStats.convertToHadoopCounters is not used, but impose MR counter limit.
- PIG-4790: Join after union fail due to UnionOptimizer.
- PIG-4814: AvroStorage does not take namenode HA as part of schema file URL.
- PIG-4816: Read a null scalar causing a Tez failure.

HDP 2.4.0 provided Pig 0.15.0 with no additional Apache patches included.

1.5.15. Ranger

HDP 2.4.3 provides Ranger 0.5.2 and the following Apache patches:

- RANGER-995: Session tokens created or regenerated by the information system after successful authentication.
- RANGER-1025: Fix not returning a policy as expected if pageSize is big enough to include the policies we are searching for.
- RANGER-1039: User and groups having special characters are not populating in edit policy page.

HDP 2.4.2 provided Ranger 0.5.2 and the following Apache patches:

- RANGER-218: LDAP Groups incorrectly labelled internal.
- RANGER-746: Addressing suggestions from Review - Add wildcard, multiple CN & SAN support when validating plugins' SSL certs.
- RANGER-771: Fix 4+ Log entries upon login in X_AUTH_SESS.
- RANGER-777: Remove the maven profile that was added for building Kafka plugin.
- RANGER-789: Fix incorrect policy list paging for non-admin users.
- RANGER-794: Ranger policy engine performance measurement.
- RANGER-798: Handle different timezone issue while saving audit logs to Solr.
- RANGER-799: Ranger UI fixes - partial search not working on Policy listing page.
- RANGER-804: Delete groups associated with User causes Exception in UserSync.
- RANGER-809: Audit framework need to cache the getHostName() values to reuse for successive calls.
- RANGER-821: Ranger shutdown hook should not only do its processing asynchronously but also terminate itself if it runs more than a configurable amount of time.
- RANGER-825: groupId in ranger's child poms should be consistent and follow convention.
- RANGER-831: policy version incremented twice when resources are updated.
- RANGER-833: In Ranger UI add support for usernames containing a plus + symbol.
- RANGER-834: Correct the excludes flag's treatment when resource value denotes everything.
- RANGER-836: Optimize policy retrieval.
- RANGER-844: Optimize policy retrieval for non-admin users.
- RANGER-848: Policy Listing page fix: users column is empty for non-admin users.
- RANGER-857: Unify (and update) Tomcat versions.
- RANGER-863: Make parameters like maxHttpHeaderSize configurable for EmbeddedServer.
- RANGER-882: Scrub received policies before policy engine uses it to guard against inadvertent data corruption: remove null policy resource values.
- RANGER-889: Policy engine API to find list of users/groups having access to a resource.
- RANGER-894: Fixing few issues in the ldap tool when user search base or group search base is configured as part of input.

HDP 2.4.0 provided Ranger 0.5.0 and the following Apache patches:

- RANGER-173: Utility scripts to create HDFS audit folders and policies.
- RANGER-725: Add the right .gitignore file to the newly added projects so that directory listing is clean after a build.
- RANGER-767: Refactor UserGroupSink implementation and consolidate performance improvements.
- RANGER-772: Hive plugin Update Ranger authorizer to mimic changes made by Hive standard authorizer for the case when IMPORT can end up creating a table.
- RANGER-773: Fix newly found Coverity scan issues for Ranger KMS.
- RANGER-778: Fix user update issue.
- RANGER-809: Audit framework need to cache the getHostName() values to reuse for successive calls.

1.5.16. Slider

HDP 2.4.3 provides Slider 0.80.0 with no additional Apache patches included.

HDP 2.4.2 provided Slider 0.80.0 and the following Apache patches:

- SLIDER-657 Introduce –force switch for slider destroy command.
- SLIDER-1079 Cache metainfo object in AgentClientProvider.
- SLIDER-1088 Slider client getting UnknownHost exception trying to RPC to AM in different subdomain.

HDP 2.4.0 provided Slider 0.80.0 with no additional Apache patches included.

1.5.17. Spark

HDP 2.4.3 provides Spark 1.6.2 with no additional Apache patches included.

HDP 2.4.2 provided Spark 1.6.1 and the following Apache patches:

- SPARK-10582: Using dynamic-executor-allocation, if AM failed, the new AM will be started. But the new AM does not allocate executors to driver.
- SPARK-11137: Make StreamingContext.stop() exception-safe.
- SPARK-11182: HDFS Delegation Token will be expired when calling "UserGroupInformation.getCurrentUser.addCredentials" in HA mode.
- SPARK-11314: Add service API and test service for Yarn Cluster schedulers.
- SPARK-11627: Spark Streaming backpressure mechanism has no initial rate limit, receivers receive data at the maximum speed, it might cause OOM exception.

- SPARK-11969: SQL UI does not work with PySpark.
- SPARK-12001: StreamingContext cannot be completely stopped if the stop() is interrupted.
- SPARK-12009: Avoid re-allocate yarn container while driver want to stop all executors.
- SPARK-12142: Can't request executor when container allocator is not ready.
- SPARK-12241: Improve failure reporting in Yarn client obtainTokenForHBase().
- SPARK-12353: Wrong output for countByValue and countByValueAndWindow.
- SPARK-12513: SocketReceiver hang in Netcat example.
- SPARK-12523: Support long-running of the Spark on HBase and hive metastore.
- SPARK-12920: Fix high CPU usage in Spark thrift server with concurrent users.
- SPARK-12925: Improve HiveInspectors.unwrap for StringObjectInspector.getPrimitiveWritableObject.
- SPARK-12948: OrcRelation uses HadoopRDD which can broadcast conf objects frequently.
- SPARK-12998: Enable OrcRelation when connecting via Spark thrift server.
- SPARK-13021: Fail fast when custom RDD's violate RDD.partition's API contract.
- SPARK-13117: WebUI should use the local IP not 0.0.0.0.
- SPARK-13308: ManagedBuffers passed to OneToOneStreamManager need to be freed in non error cases.
- SPARK-13360: Pyspark related environment variable is not propagated to driver in yarn-cluster mode.
- SPARK-13468: Fix a corner case where the page UI should show DAG but it doesn't show.
- SPARK-13478: Use real user when fetching delegation tokens.
- SPARK-13599: Groovy-all ends up in spark-assembly if Hive profile set.
- SPARK-13642: Properly handle signal kill in ApplicationMaster.
- SPARK-13885: Fix attempt ID regression for Spark running on Yarn.
- SPARK-14062: Fix log4j and upload metrics.properties automatically with distributed cache.
- SPARK-6847: Stack overflow on updateStateByKey which followed by a stream with checkpoint set.

HDP 2.4.0 provided Spark 1.6.0 and the following Apache patches:

- SPARK-11315: Add YARN extension service to publish Spark events to YARN timeline service.
- SPARK-11323: Add History Service Provider to service application histories from YARN timeline server.
- SPARK-12417: Support to have ORC bloom filters during write code path.
- SPARK-12898: Consider having dummyCallSite for HiveTableScan.

1.5.18. Sqoop

HDP 2.4.3 provides Sqoop 1.4.6 and the following Apache patches:

- SQOOP-2470: Incremental Hive import with append not working after validation check for –hive-import and –import.

HDP 2.4.2 provided Sqoop 1.4.6 with no additional Apache patches included

HDP 2.4.0 provided Sqoop 1.4.6 with no additional Apache patches included.

1.5.19. Storm

HDP 2.4.3 provides Storm 0.10.0-beta and the following Apache patches:

- STORM-721: Storm UI server should support SSL.
- STORM-994: Close download clients and channels to avoid resource leaks.
- STORM-1968: Storm logviewer does not work for nimbus.log in secure cluster.
- STORM-1989: X-Frame-Options support for Storm UI.

HDP 2.4.2 provided Storm 0.10.0-beta and the following Apache patches:

- STORM-584:Fix logging for LoggingMetricsConsumer metrics.log file.
- STORM-793:Made change to logviewer.clj in order to remove the invalid http 500 response.
- STORM-810:PartitionManager in STORM-kafka should commit latest offset before close.
- STORM-837:HdfsState ignores commits.
- STORM-866:Use storm.log.dir instead of storm.home in log4j2 config.
- STORM-966:ConfigValidation.DoubleValidator doesn't really validate whether the type of the object is a double.
- STORM-977:Incorrect signal (-9) when as-user is true.
- STORM-992:A bug in the timer.clj might cause unexpected delay to schedule new event.

- STORM-1001:Undefined STORM_EXT_CLASSPATH adds '::' to classpath of workers.
- STORM-1005:Supervisor do not get running workers after restart.
- STORM-1024:log4j changes leaving ${sys:storm.log.dir} under STORM_HOME dir.
- STORM-1027:Use overflow buffer for emitting metrics.
- STORM-1037:Do not remove STORM-code in supervisor until kill job.
- STORM-1044:Setting dop to zero does not raise an error.
- STORM-1096:Fix some issues with impersonation on the UI.
- STORM-1108:Fix NPE in simulated time.
- STORM-1121:Deprecate test only configuration nimbus.reassign.
- STORM-139:hashCode does not work for byte[].
- STORM-1481:Avoid Math.abs(Integer) get a negative value.
- STORM-1482:Add missing 'break' for RedisStoreBolt.
- STORM-1521:When using Kerberos login from keytab with multiple bolts/executors ticket is not renewed.

HDP 2.4.0 provided Storm 0.10.0-beta and the following Apache patches:

- STORM-422: Allow more arguments to be passed to storm jar.
- STORM-745: Fix storm.cmd to evaluate 'shift' correctly with 'storm jar'.
- STORM-1476: Filter -c options from args and add them as part of storm.options.

1.5.20. Tez

HDP 2.4.3 provides Tez 0.7.0 with no additional Apache patches included.

HDP 2.4.2 provided Tez 0.7.0 and the following Apache patches:

- TEZ-1961: Remove misleading exception "No running dag" from AM logs.
- TEZ-2863: Container, node, and logs not available in UI for tasks that fail to launch.
- TEZ-2963: RecoveryService#handleSummaryEvent exception with HDFS transparent encryption & Kerberos authentication.
- TEZ-3066: TaskAttemptFinishedEvent ConcurrentModificationException in recovery or history logging services.
- TEZ-3101: Tez UI: Task attempt log link doesn't have the correct protocol.
- TEZ-3117: Deadlock in Edge and Vertex code.

- TEZ-3123: Containers can get re-used even with conflicting local resources.
- TEZ-3128: Avoid stopping containers on the AM shutdown thread.
- TEZ-3137: Tez task failed with illegal state exception.
- TEZ-3156: Tez client keeps trying to talk to RM even if RM does not know about the application.
- TEZ-3177: Non-DAG events should use the session domain or no domain if the data does not need protection.
- TEZ-3189: Pre-warm dags should not be counted in submitted dags count by DAGAppMaster.

HDP 2.4.0 provided Tez 0.7.0 and the following Apache patches:

- TEZ-2307: Possible wrong error message when submitting new DAG.
- TEZ-2886: Ability to merge AM credentials with DAG credentials.
- TEZ-2898: Tez tools: swimlanes.py is broken.
- TEZ-2900: Ignore V_INPUT_DATA_INFORMATION when vertex is in Failed/Killed/Error.
- TEZ-3017: HistoryACLManager does not have a close method for cleanup.
- TEZ-3025: InputInitializer creation should use the DAG GUI.
- TEZ-3032: Incorrect start time in different events for DAG history events.
- TEZ-3037: History URL should be set regardless of which history logging service is enabled.
- TEZ-3126: Log reason for not reducing parallelism.

1.5.21. ZooKeeper

HDP 2.4.3 provides ZooKeeper 3.4.6 with no additional Apache patches included.

HDP 2.4.2 provided ZooKeeper 3.4.6 with no additional Apache patches included.

HDP 2.4.0 provided ZooKeeper 3.4.6 with no additional Apache patches included.

1.6. Third-Party Licenses

Global: Apache 2.0

Apache Component	Subcomponents	License
Accumulo	JCommander	JCommander
Falcon	cern.colt* , cern.jet*, cern.clhep	CERN
Knox	ApacheDS, Groovy	ANTLR

Apache Component	Subcomponents	License
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle
Oozie	JDOM Oro	
Phoenix		EPL
Storm	Logback	EPL

1.7. Fixed Issues in HDP 2.4.3

The following tables list and describe issues reported by customers through Hortonworks Support or Hortonworks Quality Engineering, and their resolutions in HDP 2.4.3.

Incorrect Results

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-62579	HADOOP-12291	Hadoop Common, HDFS	Add support for nested groups in LdapGroupsMapping.
BUG-59399	HBASE-15325	HBase	ResultScanner allowing partial result will miss the rest of the row if the region is moved between two rpc requests.
BUG-59881	HIVE-13693, HIVE-13831	Hive	Wrong results Multi-insert query drops filter before file output when there is a.val <> b.val filter.
BUG-57984	HIVE-13299, HIVE-13618	Hive	Column Names trimmed of leading and trailing spaces.
BUG-60335	HIVE-14147	Hive, Hive2	Hive PPD might remove predicates when they are defined as a simple expression e.g. WHERE 'a'.
BUG-59881	HIVE-13693, HIVE-13831	Hive	Wrong results Multi-insert query drops filter before file output when there is a.val <> b.val filter.

Security

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-62567	HADOOP-12901, HADOOP-13155, HADOOP-13251, HADOOP-12962	HDFS	Apply Hadoop KMS patch.
BUG-57460	HBASE-15837, PHOENIX-2883	HBase, Phoenix	Invalid memstore state during close of region for Phoenix table.
BUG-55917	STORM-1711	Storm	HiveUtils.authenticate reuses current UGI, mixes up principles in use.
BUG-57857	STORM-721	Storm	Storm UI server should support SSL.

Usability

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-49428	HBASE-13153	HBase	Bulk Loaded HFile Replication.

Query Failure

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-52478	N/A	Phoenix, Spark	Phoenix-Spark integration inoperative.

Performance

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-59466	HBASE-14867	HBase	Improved HBase Normalizer.
BUG-46240	HBASE-14473	HBase	Compute region locality in parallel.
BUG-63752	HBASE-13829	HBase	Hbase throttling Issue.
BUG-62675	HADOOP-12916, HADOOP-12985, HADOOP-13159, HADOOP-13263, HDFS-10253, HDFS-10458, HDFS-7597, HDFS-8845, HDFS-9412	HDFS	Performance improvements.
BUG-62673	HDFS-8546, HDFS-8758, HDFS-9654, HDFS-9715, HDFS-9730	HDFS	Upgrade improvements.
BUG-61590	HDFS-9259, HDFS-9700	HDFS	Backport HDFS-9259 (Make SO_SNDBUF size configurable at DFSClient side for hdfs write scenario) & HDFS-9700 (DFSClient and DFSOutputStream should set TCP_NODELAY on sockets for DataTransferProtocol).
BUG-59585	HDFS-8828	HDFS	Utilize Snapshot diff report to build diff copy list in distcp
BUG-61590	HDFS-9259, HDFS-9700	HDFS	Backport HDFS-9259 (Make SO_SNDBUF size configurable at DFSClient side for hdfs write scenario) & HDFS-9700 (DFSClient and DFSOutputStream should set TCP_NODELAY on sockets for DataTransferProtocol).
BUG-55808	YARN-3021, YARN-5048	YARN	Backport YARN-3021 (YARN's delegation-token handling disallows certain trust setups to operate properly over DistCp) and YARN-5048 (DelegationTokenRenewer#skipTokenRenewal may throw NPE).
BUG-56919	YARN-2910	YARN	FSLeafQueue can throw ConcurrentModificationException .

Stability

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-59387	HADOOP-13105	Hadoop Common	Support timeouts in LDAP queries in LdapGroupsMapping.
BUG-57732	HBASE-14712, HBASE-15100	HBase	Failover takes a long time due to MasterProcWALs never clean up.
BUG-57814	HBASE-13318	HBase	RpcServer.getListenerAddress should handle when the accept channel is closed.
BUG-60724	HDFS-10271, HDFS-9530	HDFS	ReservedSpace is not cleared for abandoned Blocks
BUG-63102	HDFS-10688	HDFS	BPServiceActor may run into a tight loop for sending block report when hitting IOException.
BUG-56423	HDFS-10312	HDFS	NameNode is stuck at SafeMode due to not all data blocks validated.
BUG-60419	HIVE-14282	Hive	Pig ToDate() exception with hive partition table ,partitioned by column of DATE datatype.
BUG-59118	HIVE-12619	Hive	Switching the field order within an array of structs causes the query to fail.
BUG-55812	HIVE-13242	Hive	DISTINCT keyword is dropped by the parser for windowing.
BUG-55551	HIVE-13200	Hive	Aggregation functions returning empty rows on partitioned columns.
BUG-63343	HIVE-13191	Hive	DummyTable map joins mix up columns between tables.
BUG-63342	HIVE-14236	Hive	CTAS with UNION ALL puts the wrong stats in Tez.
BUG-63341	HIVE-14262	Hive	Inherit writetype from partition WriteEntity for table WriteEntity.
BUG-60071	HIVE-10631, HIVE-12213	Hive	Create partitioned table takes longer time and then eventually fails
BUG-61883	HIVE-14178	Hive	Hive::needsToCopy should reuse FileUtils::equalsFileSystem.
BUG-60538	HIVE-14045	Hive	(Vectorization) Add missing case for BINARY in VectorizationContext.getNormalizedName method.
BUG-63523	HIVE-13991	Hive	Union All on view fail with no valid permission on underneath table.
BUG-59089	HIVE-13592	Hive	Metastore calls map is not thread safe.
BUG-58666	HIVE-13729	Hive	FileSystem$Cache leaks in FileUtils.checkFileAccessWithImpersonation.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-56528	HIVE-13596	Hive	Need JAR fix for UDF reloading.
BUG-56083	HIVE-13394	Hive	Analyze table fails in tez on empty partitions/files/ tables.
BUG-56528	HIVE-13596	Hive	Need JAR fix for UDF reloading.
BUG-58666	HIVE-13729	Hive	FileSystem$Cache leaks in FileUtils.checkFileAccessWithImpersonation.
BUG-58984	HIVE-13602	Hive	TPCH q16 return wrong result when CBO is on.
BUG-59089	HIVE-13592	Hive	Metastore calls map is not thread safe.
BUG-61775	HIVE-14132	Hive	Don't fail config validation for removed configs.
BUG-56776	MAPREDUCE-6514, MAPREDUCE-6689	MapReduce	MR job can infinitely increase number of reducer resource requests and job hangs as ask is not updated after ramping down of all reducers.
BUG-51701	MAPREDUCE-6635	MapReduce, Tez	java.lang.IndexOutOfBoundsException running select count(*) query on table with skip.header.line.count.
BUG-55158	PHOENIX-1478	Phoenix	Can't upsert value of 127 into column of type unsigned tinyint.
BUG-61855	PIG-4587	Pig	Applying isFirstReduceOfKey for Skewed left outer join skips records.
BUG-55869	PIG-4714	Pig	Pig launch throws ATS errors.
BUG-55895	PIG-4873	Pig	Pig job fails with Backend error : java.lang.NullPointerException.
BUG-60205	SQOOP-2470	Sqoop	Incremental Hive import with append not working after validation check for –hive-import and –import.
BUG-60024	TEZ-2533	Tez	AM deadlock when shutdown.
BUG-59515	TEZ-3193	Tez	Deadlock in AM during task commit request.
BUG-59368	TEZ-3203	Tez	DAG hangs when one of the upstream vertices has zero tasks.
BUG-62978	HADOOP-13362	YARN	OOM issue due to container metrics.
BUG-18218	YARN-3846	YARN	Scheduler queue filter failed to work because index of queue column changed.
BUG-48217	YARN-4325	YARN	After restarting Yarn services, applications stays in ACCEPTED state.

1.8. Known Issues

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-36435	KNOX-554	Knox	**Description of Problem:** In `gateway-site.xml`, changing the value of gateway.path to something other than the default 'gateway' is causing errors for some requests, such as trying to access the admin API to get the deployed topologies. **Workaround:** Use the default setting (*`gateway.path=gateway`*).
BUG-53343	YARN-5274	YARN	**Description of Problem:** The disk health checker in the NodeManager cannot detect disks with bad sectors. This can lead to situations where application data writes may fail or applications may write data that cannot be read later on. Ultimately, the result is failing and/or slow containers. **Workaround:** Use a third party disk monitoring tool that can detect bad disks such as smartctl. Once detected, remove and replace the disk.
BUG-59204	HIVE-11427	Hive	**Description of Problem:** Location of temporary table for **CREATE TABLE SELECT** broken by HIVE-7079. **Workaround:** Currently, there is no workaround available.
BUG-61991	HADOOP-13155	HDFS	**Description of Problem:** YARN cannot renew/cancel KMS delegation token for jobs because token renewer class for KMS is missing. Without it, YARN defaults to TrivialRenewer for KMS delegation token, resulting in the token not being renewed. **Associated error message:** java.io.IOException: Failed to renew token: Kind: kms-dt **Workaround:** Rerun job with a new KMS delegation token using *`addDelegationToken`*.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary

BUG-63211	HADOOP-11711, HADOOP-12158	HDFS	**Description of Problem:** You can configure the desired crypto codec implementation class for a given codec through a property such as hadoop.security.crypto.codec.classes.aes.ctr.nopad in `core-site.xml`. However, when this property is not configured in `core-site.xml`, the default value configured in `core-default.xml` does not get loaded and used without the fix provided by HADOOP-11711.
			
			Associated Error: Deployments will fail to resolve any codecs even when a default value is configured in `core-default.xml`.
			
			Workaround: Ensure desired codec class is configured in `core-site.xml`. For example, configure hadoop.security.crypto.codec.classes.<aes.ctr.nopa in the following way:
			
			`<property>`
			`  <name>hadoop.security.`
			`crypto.codec.classes.`
			`aes.ctr.nopadding</name>`
			`  <value>org.apache.`
			`hadoop.crypto.`
			`OpensslAesCtrCryptoCodec,`
			` org.apache.`
			`hadoop.crypto.`
			`JceAesCtrCryptoCodec</`
			`value>`
			`  <description>`
			`  Comma-separated`
			` list of crypto codec`
			` implementations for`
			` AES/CTR/NoPadding.`
			`  The first`
			` implementation will`
			` be used if available,`
			` others are fallbacks.`
			`  </description>`
			`</property>`

BUG-63342	HIVE-14236	Hive	**Description of Problem:** CTAS with **UNION ALL** puts the wrong stats in Tez. **Workaround:** Currently, there is no workaround available.
BUG-64511	HDFS-9618	HDFS	**Component Affected:** HDFS log **Description of Problem:** When the log level is set to **INFO**, unnecessary DEBUG log messages are generated in namenode

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			and, as a result, **namenode** performance is degraded. **Workaround:** A workaround is to set log level to higher than INFO, such as WARN, so that the unnecessary DEBUG messages will not be generated. As a side effect, this workaround will prevent INFO log messages from printing to the log. The bug is soon to be addressed by HDFS-9618, which is a very simple log message fix that changes the log level check from INFO to DEBUG since the log message is printed in DEBUG.
BUG-65005	N/A	Hive	**Component Affected:** HiveMetastore **Description of Problem:** In releases prior to HDP 2.5, Hive metastore has limited scalability when ACID is enabled. This is manifested by messages in the metastore log indicating repeated deadlocks in the RDBMS backing the metastore. Eventually the retry limit is exceeded and the metastore operation fails. **Workaround:** The only way to mitigate this is to reduce the number of operations against the metastore by making transactions larger.
BUG-65033	N/A	Hive	**Description of Problem:** Expected results of some queries mismatch with actual result. **Workaround:** Ensure that hive.convert.join.bucket.mapjoin.tez is set to `false`: `set hive.convert.join.bucket.mapjoin.tez = false`
BUG-67012	HIVE-11427	Hive	**Description of Problem:** When Hive warehouse permission is set to 750, users jobs, that do not have permission to the default database fails with TOK_TMP_FILE error. **Workaround:** Currently, there is no available

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			workaround. However, this issue is addressed in HDP 2.5.0 and will be addressed in the next HDP 2.4 release.

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 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

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
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