Hortonworks Data Platform

HDP-2.4.0 Release Notes

(March 1, 2016)

docs.cloudera.com

Hortonworks Data Platform: HDP-2.4.0 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 2.4.0 Release Notes

This document provides you with the latest information about the HDP 2.4.0 release and its product documentation.

Component Versions

The official Apache versions of most HDP 2.4.0 components are unchanged from HDP 2.3.4.0, with the exception of Spark. Spark is upgraded from 1.5.2 to 1.6.0. See more details of Spark 1.6.0 in the New Features section. All HDP 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 components should remain at the following package version levels to ensure a certified and supported copy of HDP 2.4.0.

Official Apache versions for HDP 2.4.0.

- Apache Accumulo 1.7.0

- Apache Atlas 0.5.0

- Apache Calcite 1.2.0

- Apache DataFu 1.3.0

- 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

- 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 0.5.0

- Apache Slider 0.80.0

- Apache Solr 5.5.0

- Apache Spark 1.6.0

- Apache Sqoop 1.4.6

- Apache Storm 0.10.0

- Apache Tez 0.7.0

- Apache ZooKeeper 3.4.6

Additional component versions:

- Cascading 3.0.1

- Hue 2.6.1

1.1. New Features

This section highlights the new feature in HDP 2.4.0.

Table 1.1. New Features

Component	Feature
Spark 1.6.0	General availability of Spark 1.6.0
	Dynamic executor allocation

1.2. Unsupported Features

Some features exist within HDP 2.4.0, 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 2.4.0, 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.

Table 1.2. Technical Previews

Component	Feature
HBase and Phoenix	• Phoenix Query Server
	• Phoenix Query Server (PHOENIX-971)
	• Phoenix-Spark Integration
	• RPC Throttling
	• Support for `init.d` scripts
Hive	• Hive Streaming
	• ACID support

Component	Feature
Slider	• Support for Docker-based application packaging (SLIDER-780)
Spark	• GraphX • SparkR • Zeppelin
YARN	• Add support for network I/O isolation/scheduling for containers (YARN-2140) • NodeManager: add cgroup support for disk I/O isolation (YARN-2619) • Spark-ATS integration (BUG-49244)
Zeppelin	• Zeppelin 0.6.0 • Install & Manage Zeppelin with Ambari 2.2.1 • PySpark support in Zeppelin - improvements • Zeppelin support for SSL (one way) • Zeppelin support for form-based LDAP authentication

1.2.2. Community Features

The following features are developed and tested by the community, but are not officially supported by Hortonworks. There are 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.

Table 1.3. Community Features

Component	Feature
Falcon	• Prism Server • User Recipes
HBase	• HBase Column Family Encryption: use HDFS data at rest encryption instead • Use of memcached as block cache is unsupported (HBASE-13170) • ZooKeeper-less region assignment • Region size balancing (HBASE-13103)
HDFS	• block-volume device choosing (HDFS-1804) • NameNode Federation (HDFS-1052) • viewFS (HADOOP-7257)
Kafka	• Mirror Maker (not supported when Kafka security is active) • New Consumer API
Knox	• Storm REST APIs
Oozie	• Spark action (OOZIE-1983)
Slider	• Simplified Application Packaging
Spark	• Spark Standalone

Component	Feature
	• Spark on Mesos
	• Jupyter/iPython Notebook
	• Oozie Spark action is not supported, but there is a tech note available for HDP customers
YARN	• Fair Scheduler
	• MapReduce Eclipse Plug-in
	• MapReduce Uber AM

1.3. Upgrading from HDP 2.3.x to HDP 2.4.0

HDP 2.4.0 is a feature-bearing minor release of HDP. For full instructions on how to upgrade your cluster from HDP 2.3.x to HDP 2.4.0, use the Non-Ambari Cluster Upgrade Guide.

> ⚠ **Important**
>
> Because HDP 2.3.6 contains features and fixes that are not applicable to HDP 2.4.0 or HDP 2.4.2, do not upgrade from HDP 2.3.6 to HDP 2.4.0 or HDP 2.4.2.

The following table summarizes HDP 2.2.x-to-2.4.0 upgrade options:

Cluster Management	Supporting Doc	Notes
Cluster managed manually (HDP 2.1 and earlier)	Non-Ambari Cluster Upgrade Guide	
Cluster managed manually (HDP 2.2 and later)	Non-Ambari Cluster Upgrade Guide	
Cluster managed via Ambari 1.7.0	Upgrading Ambari Guide	
Cluster managed via Ambari 2.0	Upgrading Ambari Guide	Ambari 2.0 supports rolling upgrade between HDP 2.2.x and HDP 2.4.0. **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.
Cluster managed via Ambari 2.1	Upgrading Ambari Guide	Ambari 2.1 supports rolling upgrade between HDP 2.3.x and HDP 2.4.0. **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.
Cluster managed via Ambari 2.2	Upgrading Ambari Guide	Ambari 2.2 supports rolling upgrade between HDP 2.3.x and HDP 2.4.0.

Cluster Management	Supporting Doc	Notes
		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.

1.4. HDP 2.4.0 Repositories

HDP 2.4.0 for Linux supports the following:

- Debian 6 (Deprecated) [5]

- Debian 7 [6]

- RHEL/CentOS/Oracle LINUX 6 [6]

- RHEL/CentOS/Oracle LINUX 7 [6]

- SLES 11 SP3/SP4 [7]

- Ubuntu 12 [7]

- Ubuntu 14 [7]

Important

Because HDP 2.3.6 contains features and fixes that are not applicable to HDP 2.4.0 or HDP 2.4.2, do not upgrade from HDP 2.3.6 to HDP 2.4.0 or HDP 2.4.2.

You can also view the Helper Files.

Note

The package identifier for HDP 2.4.0 components is **169**. For example:

```
dfs -mkdir -p /hdp/apps/2.4.0.0-$BUILD/hive/
```

would become:

```
dfs -mkdir -p /hdp/apps/2.4.0.0-169/hive/
```

Table 1.4. Debian 6 (Deprecated)

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian6-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian6-deb.tar.gz

Description	Link
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/debian6/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.5. Debian 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian7-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/HDP-2.4.0.0-debian7-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.6. RHEL/CentOS/Oracle LINUX 6

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos6-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos6-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.7. RHEL/CentOS/Oracle LINUX 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos7-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/HDP-2.4.0.0-centos7-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos7/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm

Description	Link
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.8. SLES 11 SP3/SP4

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/HDP-2.4.0.0-suse11sp3-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/HDP-2.4.0.0-suse11sp3-rpm.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.9. Ubuntu 12

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu12-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu12-deb.tar.gz
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.10. Ubuntu 14

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/build_metadata.txt
HDP M2 Artifacts	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu14-m2-artifacts.tar
HDP repo	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/HDP-2.4.0.0-ubuntu14-deb.tar.gz

Description	Link
HDP Search package	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/lucid/lucidworks-hdpsearch-2.3-4.noarch.rpm
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.4.0.0/slider-app-packages/pkg-list.txt

Table 1.11. Helper Files

Description	Link
Companion files	http://public-repo-1.hortonworks.com/HDP/tools/2.4.0.0/hdp_manual_install_rpm_helper_files-2.4.0.0.169.tar.gz

1.5. Behavioral Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.4.0, behavioral changes affect the following Hadoop components.

Table 1.12. Behavioral Changes

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
			Debian 6 support deprecated.	HDP support for Debian 6 is deprecated with HDP 2.4.0. Support for Debian 6 will be removed in a future HDP release.
BUG-43020	Hue		CherryPy is not the default webserver	**Scenario:** In previous releases of Hue, CherryPy was not the default web-server, but spawning webserver, which is generally not advisable as CherryPy is multithreaded and scales better than spawning webserver. **Previous Behavior:** By not commenting out `use_cherrpy_server=false` in `hue.ini`, Hue used spawning webserver instead of Hue. **New Behavior:** CherryPy is the default webserver for Hue. **Workaround/ Expected Customer Action**: No action required, unless spawning web server is the desired webserver for Hue. In this case, modify

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
				use_cherrypy_server accordingly.
BUG-50292	Hue		HUE DataTables Warning in partition columns tab.	**Scenario:** In previous releases of Hue, an error message appeared when **Partition Columns** tab was clicked in the Beeswax describe table view. **Previous Behavior:** When clicking the **Partition Columns** tab in the Beeswax describe table view, an error message appeared. **New Behavior:** No error message appears. **Workaround:** None.
BUG-50775	YARN		Enable collection of debug information by default	**Scenario:** Container failures during job execution. **Previous Behavior:** New feature - no previous behaviour. **New Behavior:** This feature(enabled by default) enables generation of additional information which is useful for debugging container failures. The extra information generated is: • A copy of the `launch_container.sh` script. • A listing of the contents of the container work directory. This information is automatically collected if log-aggregation is enabled, or ignored and discarded if log-aggregation is disabled. The feature can be disabled by setting `yarn.nodemanager.log-container-debug-info.enabled` to

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
				`false` in `yarn-site.xml`.
BUG-50786	Tez		Increase the default value for MAX_COUNTERS	The default value of the max counter limit has been increased from 2000 to: • tez.counters.max: 10000 • tez.counters.max.groups: 3000
BUG-51480	Spark		We should recommend to run spark history server on top of HDFS	In HDP 2.4, the Spark History Service reads data from HDFS. Prior to HDP 2.4, the data was read from the YARN ATS (Application Timeline Service). Refer to the Spark Guide for more info.
BUG-51481	Spark		Behavior Change for hdp.version for spark configuration from HDP 2.3.4.	**Previous Behavior:** Before HDP 2.3.4, user need to configure hdp.version variable in spark-defaults.conf: `spark.driver. extraJavaOptions -Dhdp.version= $VERSION$BUILD spark.yarn.am. extraJavaOptions -Dhdp.version= $VERSION$BUILD` **New Behavior:** As of HDP 2.4.0: when running HDP, Spark users no longer need to specify hdp.version through the Spark configuration files; it is inferred from the version of HDP installed on the client machine. There are a few exceptions to this rule: • If you are programmatically submitting jobs in a way that spark-env.sh is not executed during the submit step. • If you want to specify a different cluster version than what is installed on the client.

Hortonworks Bug ID	Component	Apache JIRA	Summary	Details
	Ambari, Spark		The Spark shuffle service is enabled by default with this release. It is necessary for Spark Dynamic Executor Allocation to work correctly.	

1.6. Apache Patch Information

The following sections list patches in each HDP 2.4.0 component beyond what was fixed in the base version of the Apache component.

- Hadoop [12]

- Accumulo [12]

- Atlas [12]

- Calcite [13]

- Falcon [13]

- Flume [13]

- HBase [15]

- Hive [16]

- Kafka [17]

- Knox [17]

- Mahout [17]

- Oozie [17]

- Phoenix [17]

- Pig [17]

- Ranger [17]

- Slider [18]

- Spark [18]

- Sqoop [18]

- Storm [18]

- Tez [19]

- ZooKeeper [19]

1.6.1. Hadoop

HDP 2.4.0 provides 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.

1.6.2. Accumulo

HDP 2.4.0 provides Accumulo 1.7.0 and the following Apache patch:

- ACCUMULO-4080 TabletServers should be less aggressively "monitoring RO filesystems".

1.6.3. Atlas

HDP 2.4.0 provides Atlas 0.5.0 and the following Apache patch:

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

1.6.4. Calcite

HDP 2.4.0 provides Calcite 1.2.0 and the following Apache patch:

- CALCITE-984: NPE in ErrorResponse construction.

1.6.5. Falcon

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

1.6.6. Flume

HDP 2.4.0 provides Flume 1.5.2 and the following Apache patch:

- FLUME-1334: Flume startup script for Windows.

- FLUME-1734: Hive Sink based on the new Hive Streaming support.

- FLUME-2058: TestFlumeEventQueue in FileChannel fails on Windows.

- FLUME-2068: File Channel issue - recovering from BadCheckpoint exception on Windows.

- FLUME-2095: JMS source with TIBCO (patch-1).

- FLUME-2122: Minor cleanups of User guide.

- FLUME-2123: Morphline Solr sink missing short type name.

- FLUME-2136: Fix intermittent test failure in TestMonitoredCounterGroup on Windows.

- FLUME-2137: Fix StagedInstall.java to invoke the correct startup script on Windows.

- FLUME-2145: TestCheckpointRebuilder.testFastReplay fails on Windows.

- FLUME-2146: Windows Tmp file creation in TestBodyTextEventSerializer.java needs fixing.

- FLUME-2150: Fix TestFileChannelEncryption failure on Windows.

- FLUME-2151: Windows Update TestExecSource to use native commands on Windows.

- FLUME-2162: TestHDFSEventSinkOnMiniCluster.maxUnderReplicationTest fails on hadoop2.

- FLUME-2175: Update Developer Guide with notes on how to upgrade Protocol Buffer version.

- FLUME-2218: TestFileChannelIntegrityTool tests failing on Windows.

- FLUME-2219: Windows Flume tests need to know location of Hadoop native libraries (hadoop.dll).

- FLUME-2224: Disable File channel dual checkpointing on Windows.

- FLUME-2224: Disable File channel dual checkpointing on Windows.

- FLUME-2226: Refactor BlobHandler out of morphline sink and into HTTP source.

- FLUME-2227: Move BlobDeserializer from Morphline Sink to flume-ng-core.

- FLUME-2250: Add support for Kafka Source.

- FLUME-2251: Add support for Kafka Sink.

- FLUME-2337: export JAVA_HOME in flume-env.sh.template and increase heap size.

- FLUME-2358: File Channel needs to close BackingStore and EventQueue before deleting files in checkpoint directory.

- FLUME-2359: TestFileChannelIntegrityTool throws exception on class teardown on Windows.

- FLUME-2402: Warning seen when overflow is disabled for Spillable Channel.

- FLUME-2407: Spillable Channel sometimes fails on reconfigure.

- FLUME-2412: Improve Logging in Spillable Channel.

- FLUME-2441: Unit test TestHTTPSource.java failed with IBM JDK 1.7.

- FLUME-2442: Need an alternative to providing clear text passwords in flume config.

- FLUME-2450: Improve replay index insertion speed.

- FLUME-2451: HDFS Sink Cannot Reconnect After NameNode Restart.

- FLUME-2454: Support batchSize to allow multiple events per transaction to the Kafka Sink.

- FLUME-2455: Documentation update for Kafka Sink.

- FLUME-2467: Disable Morphline Sink and Kite Dataset sinks on Windows Builds.

- FLUME-2470: Kafka Sink and Source must use camel case for all configs.

- FLUME-2499: Include Kafka Message Key in Event Header, Updated Comments.

- FLUME-2501: Updating HttpClient lib version to ensure compat with Solr.

- FLUME-2508: LineDeserializer causes incorrect behavior in SpoolDir Source on Windows.

- FLUME-2511: Allow configuration of enabled protocols in Avro source and Rpc client.

- FLUME-2520: HTTP Source should be able to block a prefixed set of protocols.

- FLUME-2530: Resource leaks found by Coverity tool.

- FLUME-2532: Windows TestReliableSpoolingFileEventReader - need to close filereader.

- FLUME-2533: HTTPS tests fail on Java 6.

- FLUME-2534: Windows : Windows TestFlumeEventQueue has many failures.

- FLUME-2541: Bug in TestBucketWriter.testSequenceFileCloseRetries.

- FLUME-2586: HDFS Sink should have an option to try rename even if close fails.

- FLUME-2595: Add option to checkpoint on file channel shutdown.

- FLUME-2624: Streaming ingest performance improvement.

- FLUME-2655: Update documentation for hdfs.closeTries based on FLUME-2586.

- FLUME-2662: Upgrade to Commons-IO 2.4.

- FLUME-2663: Address Build warnings of duplicate dependencies listed.

- FLUME-2722: Windows TestKafkaSourceUtil Unit Tests fail on Windows.

- FLUME-2754: Hive Sink skipping first transaction in each Batch of Hive Transactions.

- FLUME-2761: Move Hive sink out of preview mode.

- FLUME-2804: Hive sink should try to clean up transactions when flume exits.

- FLUME-2841: Upgrade commons-collections to 3.2.2.

- FLUME-2854: parameterizing jetty version.

- FLUME-2865: Upgrade thrift version (0.9.2).

- FULME-2544: Windows Incorrect Path Separator used in HDFS path (HDFS Sink).

1.6.7. HBase

HDP 2.4.0 provides 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.

1.6.8. Hive

HDP 2.4.0 provides 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.

1.6.9. Kafka

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

1.6.10. Knox

HDP 2.4.0 provides Knox 0.6.0 and the following Apache patch:

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

1.6.11. 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. In the future, after the Spark-based Mahout functionality has stabilized, HDP plans to ship with it.

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.

1.6.12. Oozie

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

1.6.13. Phoenix

HDP 2.4.0 provides Phoenix 4.4.0-HBase-1.1 and the following Apache patches:

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

1.6.14. Pig

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

1.6.15. Ranger

HDP 2.4.0 provides 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.

1.6.16. Slider

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

1.6.17. Spark

HDP 2.4.0 provides Spark 1.6.0 and the following Apache patches:

- SPARK-11314: Add service API and test service for Yarn Cluster schedulers.

- 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.

- SPARK-12925: Improve HiveInspectors.unwrap for StringObjectInspector.getPrimitiveWritableObject.

1.6.18. Sqoop

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

1.6.19. Storm

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

- STORM-1476: Filter -c options from args and add them as part of storm.options.

- STORM-422: Allow more arguments to be passed to storm jar.

- STORM-745: Fix storm.cmd to evaluate 'shift' correctly with 'storm jar'.

1.6.20. Tez

HDP 2.4.0 provides Tez 0.7.0 and the following Apache patches:

- TEZ-2307: Possible wrong error message when submitting new dag.

- 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 UGI.

- 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.

1.6.21. ZooKeeper

HDP 2.4.0 provides ZooKeeper 3.4.6 with the patches specified below. No new additional Apache patches have been included in this release.

1.7. Common Vulnerabilities and Exposures

- **CVE-2015-7521:** Apache Hive authorization bug disclosure

  Severity: Important

  Vendor: The Apache Software Foundation

  Versions Affected: HDP versions 2.1.x, 2.2.x and 2.3.x versions before HDP 2.3.6

  Description: Some partition-level operations exist that do not explicitly also authorize privileges of the parent table. This can lead to issues when the parent table would have denied the operation, but no denial occurs because the partition-level privilege is not checked by the authorization framework, which defines authorization entities only from the table level upwards. This issue is known to affect Hive clusters protected by both Ranger as well as SqlStdHiveAuthorization.

  Mitigation: For Hive 0.13.x, 0.14.x, 1.0, 1.1 and 1.2, a separate jar is being made available, which users can put in their ${HIVE_HOME}/lib/, and this provides a hook for administrators to add to their hive-site.xml, by setting hive.semantic.analyzer.hook=org.apache.hadoop.hive.ql.parse.ParentTableAuthorizationHook . This parameter is a comma-separated-list and this hook can be appended to an existing list if one already exists in the setup. You will then want to make sure that you protect the hive.semantic.analyzer.hook parameter from being changed at runtime by adding it to hive.conf.restricted.list. This jar and associated source tarball are available for download over at : https://hive.apache.org/downloads.html along with their gpg-signed .asc signatures, as well as the md5sums for verification in the hive-parent-auth-hook/ directory. This issue has already been patched in all Hive branches that are affected, and any future release will not need these mitigation steps.

Hortonworks Bug ID: BUG-50827

- **CVE-2015-5167:** Restrict REST API data access for non-admin users

Severity: Important

Vendor: Hortonworks

Versions Affected: All HDP 2.3.x releases prior to 2.3.4

Users Affected: All users of ranger policy admin tool.

Impact: See BUG-41604 and RANGER-630. Data access restrictions via REST API are not consistent with restrictions in policy admin UI.

Recommended Action: Upgrade to 2.3.4.x+ or HDP 2.4.0+.

- **CVE-2015-7521:** Apache Hive authorization bug disclosure

Severity: Important

Vendor: The Apache Software Foundation

Versions Affected: Apache Hive 1.0.0 - 1.0.1, Apache Hive 1.1.0 - 1.1.1, and Apache Hive 1.2.0 - 1.2.1

Description: Some partition-level operations exist that do not explicitly also authorize privileges of the parent table. This can lead to issues when the parent table would have denied the operation, but no denial occurs because the partition-level privilege is not checked by the authorization framework, which defines authorization entities only from the table level upwards. This issue is known to affect Hive clusters protected by both Ranger as well as SqlStdHiveAuthorization.

Mitigation: For Hive 1.0, 1.1 and 1.2, a separate jar is being made available, which users can put in their ${HIVE_HOME}/lib/, and this provides a hook for administrators to add to their hive-site.xml, by setting hive.semantic.analyzer.hook=org.apache.hadoop.hive.ql.parse.ParentTableAuthorizationHook . This parameter is a comma-separated-list and this hook can be appended to an existing list if one already exists in the setup. You will then want to make sure that you protect the hive.semantic.analyzer.hook parameter from being changed at runtime by adding it to hive.conf.restricted.list. This jar and associated source tarball are available for download over at : https://hive.apache.org/downloads.html along with their gpg-signed .asc signatures, as well as the md5sums for verification in the hive-parent-auth-hook/ directory. This issue has already been patched in all Hive branches that are affected, and any future release will not need these mitigation steps.

Hortonworks Bug ID: BUG-50827

- **CVE-2016-0733:** Ranger Admin authentication issue

Severity: Important

Vendor: Hortonworks

Versions Affected: All HDP 2.3.x releases prior to 2.3.4

Users Affected: All users of ranger policy admin tool.

Impact: See BUG-50669 and RANGER-835. Malicious Users can gain access to ranger admin UI without proper authentication.

Recommended Action: Upgrade to 2.3.4.x+ or HDP 2.4.0+.

- **CVE-2016-0735:** In some cases, presence of an exclude policy at a level can give the user access at its parent level.

  Severity: Critical

  Vendor: Hortonworks

  Versions Affected: All HDP 2.3.0+.

  Users Affected: All users that use Ranger to authorize HBase, Hive, and Knox.

  Impact: See BUG-50558. In some cases, presence of an exclude policy at a level can give the user access at its parent level. For example, if a hive policy excludes access for a user to a particular column, then such a user would be able to alter the name of that table. Only a user who has access at the table level should be able to do so. Due to this bug however, the user is able to do the operation which is caused by presence of an exclude policy at the column-level for that table. Recommended Action: Upgrade to HDP 2.4.0.0+ or contact Hortonworks support team.

  Recommended Action: Upgrade to 2.3.7+ or HDP 2.4.0+.

1.8. Third-party Licenses

Global: Apache 2.0

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
Oozie	JDOM Oro	
Phoenix		EPL
Storm	Logback	EPL

1.9. 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.

Potential Data Loss

None.

Security

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-46910		Hue	Request to add HUE-1663 to fix LDAP referral handling for authentication.
BUG-50350	HBASE-14471	HBase	Thrift - HTTP Error 413 full HEAD if using kerberos authentication.
BUG-51013	TEZ-3025	Tez	InputInitializer creation should use the dag UGI.

Incorrect Results

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-36030	HIVE-12706	Hive	Hive from_utc_timestamp() function doesn't return the right result.

Stability

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-41224		Hue	Workaround Hiveserver2 memory Leak.
BUG-50215	FALCON-1595	Falcon	In secure cluster, Falcon server loses ability to communicate with HDFS over time.
BUG-50700	HIVE-12660	Hive	HS2 memory leak with .hiverc file use.

Query Failure

None.

Upgrade

None.

Usability

None.

Performance

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-42479	HBASE-14575	HBase	Relax region read lock for compactions.

Other

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-50101		Hadoop Common	Remove ivy jar from hadoop-lzo source code.

1.10. Known Issues

Hortonworks Bug ID	Apache JIRA	Component	Summary
BUG-23260	HIVE-11421, HIVE-11981	Hive	Insert after alter table on Acid table causes ArrayIndexOutOfBoundsException **Issue**: Schema evolution is not supported for Acid tables. Alter Table commands on Acid table may make the table unreadable. There is not w/ a after that.
BUG-55196	HIVE-12937	Hive	DbNotificationListener unable to clean up old notification events
BUG-30022	HIVE-7693	Hive	Invalid column ref error in order by when using column alias in select clause and using having
BUG-30556		Hive	Killed job is not display correctly by ATS hook failure
BUG-38054	RANGER-577	Ranger	Ranger performs authorization check, even if Hive authorization is disabled When Ranger is Enabled for authorization, it would do the authorization check even if the hive authorization is disabled (via `hive.security.authorization.enabled = false`).
BUG-38928	HIVE-11599	Hive	Alter table transactional=false on ACID table misleading user. create transactional table would be better **Issue:** Once an ACID table is created (with transactional=true) and data is written to it, changing the property to transactional=false may cause wrong data to be returned by queries over this table.
BUG-39663		Hive	Hive compactor fails when 'hive' user does not have x permissions on the table directory
BUG-39988	HIVE-11110	Hive	CBO : Default partition filter is from MetaStore query causing TPC-DS to regress by 3x
BUG-43128	HIVE-12724	Hive	Query wrong results: non-ACID to ACID, major compaction does not

Hortonworks Bug ID	Apache JIRA	Component	Summary
			compact all pre-ACID data into base dir
BUG-43129		Hive	ACID update/delete query fails with ArrayIndexOutOfBoundsException when there is null struct
BUG-45846	HIVE-11605	Hive	Incorrect results with bucket map join in tez
BUG-45857	HIVE-5623	Hive	ORC accessing array column that's empty will fail with java out of bound exception
BUG-46092	HIVE-7193	Hive	We need to filter users who can connect on our HiveServer2 to those in a particular LDAP group (with nested groups)
BUG-46128		Storm	Kafka spout showing errors in the log streaming workflow
BUG-46771	HIVE-11716	Hive	Reading ACID table from non-acid session should raise an error
BUG-47626		HBase, Hive, Phoenix	Creating Hive table on HBase fails in HDP 2.3.x with Phoenix Dependencies
BUG-48217		YARN	After restarting Yarn services, applications stays in ACCEPTED state
BUG-48416		Falcon	Unable to save a feed via Falcon UI
BUG-48979		YARN	**Issue**: Jobs fail with javax.security.sasl.SaslException **Cause**: OpenJDK 7u91 is stricter about validating ticket caches; if there is any ticket in the cache without a default realm, SASL will fail to work. **Workaround**: To avoid this problem, the /etc/krb5.conf file must contain a domain_realm section.
BUG-48990	HIVE-12837	Hive	MicroStrategy query 117 fails with out of memory error
BUG-49726	HIVE-10632	Hive	Hive compactor processing partitions of table that does not exist
BUG-49784	HDFS-9445	HDFS	Datanode may deadlock while handling a bad volume
BUG-49949		Hive	Query with duplicate columns in GROUP BY clause fails
BUG-49955	PHOENIX-2531	Phoenix	Phoenix thin-client jar has incorrect META-INF/services/java.sql.Driver file

Hortonworks Bug ID	Apache JIRA	Component	Summary

			Issue: The Phoenix Thin Client Driver, org.apache.phoenix.queryserver.client.Driver, is not automatically registered in the JDBC DriverManager.
			
			Workaround:
			
			1. Add the following to your Java application before trying to obtain the driver:
			
			<pre>Properties props = new
			 Properties();
			String URL =
			 "jdbc:phoenix:thin:URL=
			http://
			localhost:8765";
			try {
			  Class.forName("org.
			apache.phoenix.
			queryserver.client.
			Driver");
			} catch (Exception e)
			 {
			  throw new
			RuntimeException(e);
			}
			DriverManager.
			getConnection(URL,
			 props);</pre>
			2. Manually register the driver.

BUG-50073	HDFS-9557	HDFS	Reduce object allocation in PB conversion
BUG-50089	HDFS-9572	HDFS	Prevent DataNode log spam if a client connects on the data transfer port but sends no data.
BUG-50302	HBASE-15039	HBase	HMaster and RegionServers should try to refresh token keys from zk when facing InvalidToken **Issue**: Many customers use importTsv to load data into HBase, but the cluster network may not always be good. Sometimes they face many map tasks inside importTsv throwing InvalidToken exception. **Workaround**: Restart the region server.
BUG-50479	HADOOP-11252	Hadoop Common	RPC client does not time out by default
BUG-50531		Kafka	Kafka file system support **Issue:** Encrypted file systems such as SafenetFS are not supported for Kafka. Index file corruption can occur.

Hortonworks Bug ID	Apache JIRA	Component	Summary
			Workaround: N/A For more information, see: Install Kafka.
BUG-50576	SQOOP-2779	Sqoop	Sqoop metastore doesn't seem to recognize –schema option
BUG-50599	HDFS-9600	HDFS	Do not check replication if the block is under construction
BUG-50652	HDFS-9574	HDFS	Reduce client failures during datanode restart
BUG-51239		Ranger	NameNode shutdown hangs due to Ranger HDFS plugin
BUG-51292		Kafka, Spark	Spark Streaming Kafka direct API unit test failure in HDP release **Workaround**: None.
BUG-51293	HDFS-9655	HDFS	NN should start JVM pause monitor before loading fsimage
BUG-51423	HDFS-9313	HDFS	Possible NullPointerException in BlockManager if no excess replica can be chosen
BUG-51424	HDFS-9314	HDFS	Improve BlockPlacementPolicyDefault's picking of excess replicas
BUG-51426	HDFS-9574	HDFS	Reduce client failures during datanode restart
BUG-51427	HDFS-8647, HDFS-9600	HDFS	Do not check replication if the block is under construction
BUG-51428	HADOOP-11252	Hadoop Common	RPC client does not time out by default
BUG-51433	HDFS-9572	HDFS	Prevent DataNode log spam if a client connects on the data transfer port but sends no data.
BUG-51436	HDFS-9557	HDFS	Reduce object allocation in PB conversion
BUG-51439	HDFS-9445	HDFS	Datanode may deadlock while handling a bad volume
BUG-51442	HDFS-9655	HDFS	NN should start JVM pause monitor before loading fsimage
BUG-51463	HDFS-8898	HDFS	Create API and command-line argument to get quota and quota usage without detailed content summary
BUG-51657		Ranger	Cyclic symlink for ranger usersync configuration after upgrade **Issue:** When manually upgrading from 2.3.4 to

Hortonworks Bug ID	Apache JIRA	Component	Summary

			2.4.0, a cyclic symlink gets created for usersync conf.
			
			Cause: This issue arises due to wrongly created symlinks which make the conf links cyclic and invalid.
			
			Workaround:
			
			• For Ranger-Admin
			
			  1. Delete the conf links from below paths:
			
			    `/usr/hdp/current/`
			    `ranger-admin/`
			    `/usr/hdp/current/`
			    `ranger-admin/ews/`
			    `webapp/WEB-INF/`
			    `classes/`
			    `/etc/ranger/admin/`
			
			  2. Run the `set_globals.sh` script present under `/usr/hdp/current/ranger-admin/` to create appropriate symlinks.
			
			  3. Re-Install the application from Ambari.
			
			• For Ranger-Usersync:
			
			  1. Delete the conf links from below paths:
			
			    `/usr/hdp/current/`
			    `ranger-usersync/`
			    `/etc/ranger/`
			    `usersync/`
			
			  2. Run the `set_globals.sh` script present under `/usr/hdp/current/ranger-usersync/` to create appropriate symlinks.
			
			  3. Re-Install the application from Ambari.

BUG-51712		Ranger	After upgrade getting java.lang.NoClassDefFoundError **Issue**: java.lang.NoClassDefFoundError error found in ranger usersync log after upgrade. **Workaround**: Copy the below jar files into ranger usersync lib folder and restart the service.

Hortonworks Bug ID	Apache JIRA	Component	Summary
			``cp /usr/hdp/current/
hadoop-client/lib/
commons-httpclient-3.
1.jar /usr/hdp/current/
ranger-usersync/lib/
cp /usr/hdp/current/
hadoop-client/lib/
commons-codec-1.4.jar /
usr/hdp/current/ranger-
usersync/lib/```
BUG-67482	RANGER-1136	Ranger	**Description of Problem:** Ranger audit to HDFS fails with TGT errors **Workaround:** Currently, there is no known workaround for this issue.

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**

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
			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

1.11. Documentation Errata

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

- Flume: Kafka Sink [29]

- Hive Sink [30]

1.11.1. Flume: Kafka Sink

This is a Flume Sink implementation that can publish data to a Kafka topic. One of the objectives is to integrate Flume with Kafka so that pull-based processing systems can process the data coming through various Flume sources. This currently supports Kafka 0.8.x series of releases.

Property Name	Default	Description
type	-	Must be set to org.apache.flume.sink.kafka.KafkaSink.
brokerList	-	List of brokers Kafka-Sink will connect to, to get the list of topic partitions. This can be a partial list of brokers, but we recommend at least two for HA. The format is a comma separated list of hostname:port.
topic	default-flume-topic	The topic in Kafka to which the messages will be published. If this parameter is configured, messages will be published to this topic. If the event header contains a "topic" field, the event will be published to that topic overriding the topic configured here.
batchSize	100	How many messages to process in one batch. Larger batches improve throughput while adding latency.
requiredAcks	1	How many replicas must acknowledge a message before it is considered successfully written. Accepted values are 0 (Never wait for

Property Name	Default	Description
		acknowledgement), 1 (wait for leader only), -1 (wait for all replicas) Set this to -1 to avoid data loss in some cases of leader failure.
Other Kafka Producer Properties	-	These properties are used to configure the Kafka Producer. Any producer property supported by Kafka can be used. The only requirement is to prepend the property name with the prefix "Kafka.". For example: kafka.producer.type.

Note: Kafka Sink uses the topic and key properties from the FlumeEvent headers to send events to Kafka. If the topic exists in the headers, the event will be sent to that specific topic, overriding the topic configured for the Sink. If key exists in the headers, the key will used by Kafka to partition the data between the topic partitions. Events with same key will be sent to the same partition. If the key is null, events will be sent to random partitions.

An example configuration of a Kafka sink is given below. Properties starting with the prefix Kafka (the last 3 properties) are used when instantiating the Kafka producer. The properties that are passed when creating the Kafka producer are not limited to the properties given in this example. It is also possible include your custom properties here and access them inside the preprocessor through the Flume Context object passed in as a method argument.

```
a1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink a1.sinks.k1.topic =
 mytopic
a1.sinks.k1.brokerList = localhost:9092
a1.sinks.k1.requiredAcks = 1
a1.sinks.k1.batchSize = 20
a1.sinks.k1.channel = c1
```

1.11.2. Hive Sink

This sink streams events containing delimited text or JSON data directly into a Hive table or partition. Events are written using Hive transactions. As soon as a set of events are committed to Hive, they become immediately visible to Hive queries. Partitions to which flume will stream to can either be pre-created or, optionally, Flume can create them if they are missing. Fields from incoming event data are mapped to corresponding columns in the Hive table.

Property Name	Default	Description
channel	–	
type	–	The component type name, needs to be hive.
hive.metastore	–	Hive metastore URI (E.g., thrift://a.b.com:9083).
hive.database	–	Hive database name .
hive.table	–	Hive table name.
hive.partition	–	Comma separated list of partition values identifying the partition to write to. May contain escape sequences. E.g.: If the table is partitioned by (continent: string, country :string, time : string)

Property Name	Default	Description
		then 'Asia,India,2014-02-26-01-21' will indicate continent=Asia,country=India,time=2014-02-26-01-21.
hive.txnsPerBatchAsk	100	Hive grants a batch of transactions instead of single transactions to streaming clients like Flume. This setting configures the number of desired transactions per Transaction Batch. Data from all transactions in a single batch end up in a single file. Flume will write a maximum of batchSize events in each transaction in the batch. This setting in conjunction with batchSize provides control over the size of each file. Note that eventually Hive will transparently compact these files into larger files.
heartBeatInterval	240	(In seconds) Interval between consecutive heartbeats sent to Hive to keep unused transactions from expiring. Set this value to 0 to disable heartbeats .
autoCreatePartitions	true	Flume will automatically create the necessary Hive partitions to stream to.
batchSize	15000	Max number of events written to Hive in a single Hive transaction.
maxOpenConnections	500	Allow only this number of open connections. If this number is exceeded, the least recently used connection is closed.
callTimeout	10000	(In milliseconds) Timeout for Hive & HDFS I/O operations, such as openTxn, write, commit, abort.
serializer	–	Serializer is responsible for parsing out field from the event and mapping them to columns in the hive table. Choice of serializer depends upon the format of the data in the event. Supported serializers: DELIMITED and JSON.
roundUnit	minute	The unit of the round down value - second, minute or hour.
roundValue	1	Rounded down to the highest multiple of this (in the unit configured using hive.roundUnit), less than current time.
timeZone	Local	Name of the timezone that should be used for resolving the escape sequences in partition, e.g. Time America/Los_Angeles.
useLocalTimeStamp	false	Use the local time (instead of the timestamp from the event header) while replacing the escape sequences.

Following serializers are provided for Hive sink:

- **JSON:** Handles UTF8 encoded Json (strict syntax) events and requires no configuration. Object names in the JSON are mapped directly to columns with the same name in the Hive table. Internally uses org.apache.hive.hcatalog.data.JsonSerDe but is independent of the Serde of the Hive table. This serializer requires HCatalog to be installed.

- **DELIMITED:** Handles simple delimited textual events. Internally uses LazySimpleSerde but is independent of the Serde of the Hive table.

Property Name	Default	Description
serializer.delimiter	,	(Type: string) The field delimiter in the incoming data. To use special characters, surround them with double quotes like "\t".
serializer.fieldnames	–	The mapping from input fields to columns in hive table. Specified as a comma separated list (no spaces) of hive table columns names, identifying the input fields in order of their occurrence. To skip fields leave the column name unspecified. Eg. 'time,,IP,message' indicates the 1st, 3rd and 4th fields in input map to time, IP and message columns in the hive table.
serializer.serdeSeparator	Ctrl-A	(Type: character) Customizes the separator used by underlying serde. There can be a gain in efficiency if the fields in serializer.fieldnames are in same order as table columns, the serializer.delimiter is same as the serializer.serdeSeparator and number of fields in serializer.fieldnames is less than or equal to number of table columns, as the fields in incoming event body do not need to be reordered to match order of table columns. Use single quotes for special characters like '\t'. Ensure input fields do not contain this character. **Note:** If serializer.delimiter is a single character, preferably set this to the same character.

The following are the escape sequences supported:

Alias	Description
%{host}	Substitute value of event header named "host". Arbitrary header names are supported.
%t	Unix time in milliseconds .
%a	Locale's short weekday name (Mon, Tue, ...)
%A	Locale's full weekday name (Monday, Tuesday, ...)
%b	Locale's short month name (Jan, Feb, ...)
%B	Locale's long month name (January, February, ...)
%c	Locale's date and time (Thu Mar 3 23:05:25 2005)
%d	Day of month (01)
%D	Date; same as %m/%d/%y
%H	Hour (00..23)
%I	Hour (01..12)
%j	Day of year (001..366)
%k	Hour (0..23)
%m	Month (01..12)
%M	Minute (00..59)

Alias	Description
%p	Locale's equivalent of am or pm
%s	Seconds since 1970-01-01 00:00:00 UTC
%S	Second (00..59) %y last two digits of year (00..99)
%Y	Year (2015)
%z	+hhmm numeric timezone (for example, -0400)

Example Hive table:

```
create table weblogs ( id int , msg string )

partitioned by (continent string, country string, time string)
clustered by (id) into 5 buckets
stored as orc;
```

Example for agent named a1:

```
a1.channels = c1

a1.channels.c1.type = memory

a1.sinks = k1

a1.sinks.k1.type = hive

a1.sinks.k1.channel = c1

a1.sinks.k1.hive.metastore = thrift://127.0.0.1:9083
a1.sinks.k1.hive.database = logsdb

a1.sinks.k1.hive.table = weblogs
a1.sinks.k1.hive.partition = asia,%{country},%y-%m-%d-%H-%M
a1.sinks.k1.useLocalTimeStamp = false

a1.sinks.k1.round = true

a1.sinks.k1.roundValue = 10

a1.sinks.k1.roundUnit = minute

a1.sinks.k1.serializer = DELIMITED
a1.sinks.k1.serializer.delimiter = "\t"
a1.sinks.k1.serializer.serdeSeparator = '\t'
a1.sinks.k1.serializer.fieldnames =id,,msg
```

Note: For all of the time related escape sequences, a header with the key "timestamp" must exist among the headers of the event (unless useLocalTimeStampis set to true). One way to add this automatically is to use the TimestampInterceptor.

The above configuration will round down the timestamp to the last 10th minute. For example, an event with timestamp header set to 11:54:34 AM, June 12, 2012 and 'country' header set to 'india' will evaluate to the partition (continent='asia',country='india',time='2012-06-12-11-50'. The serializer is configured to accept tab separated input containing three fields and to skip the second field.