Hortonworks Data Platform

Release Notes

(August 29, 2016)

Hortonworks Data Platform: Release Notes

Copyright © 2012-2016 Hortonworks, Inc. Some rights reserved.

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 2.5.0 Release Notes

This document provides you with the latest information about the HDP 2.5.0 release and its product documentation.

Component Versions

The official Apache versions of all HDP 2.5.0 components are listed below. All 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 2.5.0.

Official Apache versions for HDP 2.5.0:

- Apache Accumulo 1.7.1

- Apache Atlas 0.7.0

- Apache Calcite 1.2.0

- Apache DataFu 1.3.0

- Apache Falcon 0.10.0

- Apache Flume 1.5.2

- Apache Hadoop 2.7.3

- Apache HBase 1.1.2

- Apache Hive 1.2.1

- Apache Kafka 0.10.0.1

- Apache Knox 0.9.0

- Apache Mahout 0.9.0+

- Apache Oozie 4.2.0

- Apache Phoenix 4.7.0

- Apache Pig 0.16.0

- Apache Ranger 0.6.0

- Apache Slider 0.91.0

- Apache Solr 5.5.2

- Apache Spark 1.6.2

- Apache Sqoop 1.4.6

- Apache Storm 1.0.1

- Apache TEZ 0.7.0

- Apache Zeppelin 0.6.0

- Apache ZooKeeper 3.4.6

Later versions of a few Apache components are sometimes bundled in the HDP distribution in addition to the versions listed above. In this case, these later versions are listed in the Technical Previews table and should not substitute for the Apache component versions of the above list in a production environment.

Additional component versions:

- Cascading 3.0.0

- Hue 2.6.0

Note

For information on open source software licensing and notices, please refer to the Licenses and Notices files included with the software install package.

1.1. New Features

This section highlights new features in HDP 2.5.0.

Table 1.1. New Features

Apache Component	Feature
Atlas	Core Capabilities:
	• Object Versioning & Audit (ATLAS-493)
	• Message Durability & Replay (ATLAS-912)
	• Ambari Service Integration (AMBARI-14853)
	• Embedded Solr for Search (ATLAS-490, AMBARI-15990)
	• Embedded HBase (ATLAS-498)
	• Soft Deletes (ATLAS-622)
	• HDFS Path Model (ATLAS-751)
	Enterprise Readiness:
	• Ranger Plugin for Authorization (ATLAS-495)
	• Kerberos support (ATLAS-785)
	• HA improvements (ATLAS-513)
	Ease of Use:
	• Dynamic Tag Based Access Policy (RANGER-274)
	• UI Authentication (ATLAS-494)
	• Rolling upgrade support (AMBARI-15830)
	• Expanded Sqoop Support (ATLAS-379, ATLAS-837)
	• Expanded Hive Support (ATLAS-492)

Apache Component	Feature
	• Storm / Kafka Support (ATLAS-183)
	• Falcon Support (ATLAS-835)
Falcon	Core Features:
	• Apache Spark Integration (FALCON-1763)
	• Apache Atlas Integration (FALCON-1570)
	• Provider integration with Microsoft Azure Data Factory pipelines (FALCON-1627)
	Security Features:
	• Falcon REST API trusted proxy support (FALCON-1027)
	• Certify HDFS encryption support with Falcon
	Manageability:
	• Capture data transfer details for various available replications (FALCON-1479)
	• Support for multiple lib paths in Falcon process (FALCON-887)
	• Supply user-defined properties to Oozie workflows during schedule (FALCON-1573)
	• Add ability to rerun process by ID (FALCON-1525)
	• Improved Metrics related to pipelines and instances (FALCON-1766)
	Usability:
	• Falcon UI to enable entity/process/feed edits and management (FALCON-790)
	• Pluggable filesystem support (FALCON-1012)
Flume	Developer Productivity:
	• Flume-Kafka Channel (FLUME-2500)
	• TailDir (FLUME-2498)
HBase	Developer Productivity:
	• HBase C+ Client
	• Upgrade jamon-runtime to version 2.4.1 (HBASE-14493)
	Core Features:
	• HBase Backup/Restore Based on HBaseSnapshot (HBASE-7912)
	• MOB (Moderate Object Storage) Support (HBASE-11339)
	• Date-Tiered Compaction Strategy (HBASE-15181)
	• FIFO Compaction Strategy (HBASE-14468)
	Streamlined Operations:
	• Revamped HBase Dashboard in Ambari (HBASE-12133, HBASE-13965, HBASE-14082, HBASE-14166, HBASE-14205, HBASE-14274, HBASE-14274, HBASE-14278, HBASE-14459, HBASE-14778, HBASE-14862, HBASE-14869, HBASE-14891, HBASE-14976, HBASE-14983, HBASE-15135, HBASE-15163, HBASE-15222, HBASE-15376, HBASE-15377, HBASE-15412, HBASE-15435, HBASE-15464, HBASE-15518, HBASE-15519, HBASE-15586, HBASE-15671)
Hive	EDW Offload:
	• Dynamically partitioned hash join : Query Execution (HIVE-4924)
	• Composable Stats from partition stats/ HLL (HIVE-13418)

Apache Component	Feature
	• Vectorization: NPE for added columns in ORC non-partitioned tables (aka Schema Evolution) (HIVE-9985) • Microstrategy Superapp: SQL Gaps: order by UDF in over clause (HIVE-13475) • Cognos OLAP asks (HIVE-4662) • General ORC schema evolution (HIVE-13178) • Don't create empty bucket files (HIVE-13040) • More scalable partition loads (HIVE-12988) Interactive SQL: • Sub-5-Second: Compile simple query in less than 300 milliseconds (includes metastore) (HIVE-7926) • Bucketed Table: prune buckets for filters (HIVE-10924) • Sub-5-Second: Parallelize Query compilation in HS2 (HIVE-11525) • Statement.setQueryTimeout() implemented in the Hive JDBC driver (HIVE-4239) • Plan generation : Improve performance of PartitionPruner and PredicatePushDown transformation optimizations (HIVE-12878) • Support rely/no-validate PK/FK relationship in Hive (HIVE-13076) • Optimize metadata-only HAVING query (HIVE-12640, HIVE-12645) Enterprise Readiness: • Hiveserver2 ThriftHttpServlet does not process X-Forwarded-For (HIVE-10673) • Make Beeline usage easier (HIVE-11141) • Log additional information in HiveServer2 (HIVE-11684) • Support lazy initiation for TEZ session when start hive CLI (HIVE-11969) • Extend Hive JDBC functionality to support Hive View use cases (HIVE-6535) • Hive - support Row level filtering and column masking (HIVE-13125)
Kafka	Streamlined Operations: • Kafka Rack Awareness (KIP-36) Enterprise Readiness: • New Consumer API • Improved Kafka SASL Support (KIP-43)
Knox	Core Security: • Knox Support for Hadoop/Yarn/HBase/Hive High Availability (KNOX-567) • Knox should support authentication using SPNEGO from browser (KNOX-25) Enterprise Readiness: • Knox SSO for Ambari and Ranger UI (RANGER-685, AMBARI-15479, AMBARI-15778, AMBARI-16638) • Knox support for SAML v2.0 based SSO system integration (KNOX-641)
Oozie	Core Features: • Spark action (OOZIE-1983)
Phoenix	Developer Productivity:

Apache Component	Feature
	• "OFFSET" feature of Phoenix for traversing the data in pages and retrieving only those many records from HBase. (PHOENIX-2722) • Phoenix .NET Driver • Phoenix ODBC Driver • Phoenix Query Server (PHOENIX-971) and HardeningWork for General Availability (CALCITE-1128, CALCITE-1159, CALCITE-1180, CALCITE-1190, PHOENIX-2792) • Support Parallel Order-By (PHOENIX-2126) • Improve efficiency of the Phoenix bulkloader (PHOENIX-1973) • Query Phoenix via Hive using PhoenixStorageHandler (PHOENIX-331, PHOENIX-2743) Core Features: • Multi-tenancy Improvements: Namespace (PHOENIX-1311) • Multi-tenancy Improvements: RegionServer Group based Assignment (HBASE-6721) • Multi-tenancy Improvements: Local Index in Apache Phoenix goes into default HBase namespace (PHOENIX-1311) • Stabilize Phoenix Local Secondary Indexes (PHOENIX-1734) • Option to Make Secondary Indexes Always Readable (PHOENIX-2221)
Ranger	Advanced Security: • Ranger policy model to support data masking (RANGER-873) • Ranger policy model to support row-filtering (RANGER-908) • Dynamic Tag Based Access Policy (RANGER-274) • Time and Prohibition-based Access Policy (RANGER-606) • Location-based Access Policy (RANGER-596) Enterprise Readiness: • Ranger KMS – Safenet Luna HSM Integration (RANGER-868) • Improvements on Reports page in Ranger Admin : Enhanced search by resource access, tags, etc. download report to a Excel, csv (RANGER-913) • Remove support for DB based auditing (RANGER-900) • Script and process to migrate existing audit from RDBMS to Solr (RANGER-271) • Support multiple OU in LDAP search for Ranger usersync (RANGER-803) • UserSync : limit the groups associated with users based on the group-search results (RANGER-869) • Provide support to delete Users and Groups from Ranger Admin UI (RANGER-888) • Add Ranger related rules to SmartSense
Spark	Performance: • HBase connector (HBASE-14789) Streamlined Operations: • YARN ATS v1.5 integration (SPARK-11373) Developer Productivity: • SparkR (SPARK-5654)

Apache Component	Feature
	Advanced Security: • Security Enhancements: Certification with HDFS TDE & wire encryption
Storm	Developer Productivity: • Storm Windowing and State Management(STORM-1167) • New Storm Connectors (STORM-1075, STORM-845, STORM-851) • Kafka Spout using new Client APIs (STORM-822) • Storm Distributed Log Search (STORM-902) • Storm Dynamic Log Levels (STORM-412) Performance: • Storm Performance Improvements (STORM-1151, STORM-1526, STORM-1537, STORM-1539) Streamlined Operations: • Storm Management View via Ambari (AMBARI-16095) • Storm Topology Event Inspector (STORM-954) • Storm Dynamic Worker Profiling (STORM-1157) • Storm Resource Aware Scheduling (STORM-894) Enterprise Readiness: • Automatic Back Pressure (STORM-886) • Distributed Cache (STORM-876) • Improved Nimbus High Availability (STORM-166) • Pacemaker Storm Daemon (STORM-885)
Zeppelin	Core Capability: • Zeppelin general availability Advanced Security: • Authentication – notes run as user (ZEPPELIN-548, ZEPPELIN-797, ZEPPELIN-769) • Authorization of Notebook (ZEPPELIN-549, ZEPPELIN-838, ZEPPELIN-946) • Kerberos Support (ZEPPELIN-446) • Ambari managed LDAP configuration (AMBARI-15758) • Zeppelin support for SSL (one way) Streamlined Operations: • Multi User support (ZEPPELIN-773) • Ambari managed install (AMBARI-15265, AMBARI-16779) Manageability: • Zeppelin Ambari View (AMBARI-16957)

1.2. Platform Support Matrices

This section outlines the platform support matrices for HDP 2.5.0.

1.2.1. Operating System Support Matrix

Unless otherwise noted, the following operating systems are validated and supported for HDP 2.5.0:

Table 1.2. HDP 2.5.0 Operating System Support Matrix

Operating System	Version
Debian	Debian 7.8
Oracle (64-bit)	Oracle 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8
	Oracle 7.0, 7.1, 7.2
CentOS (64-bit)	CentOS 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8
	CentOS 7.0, 7.1, 7.2
Red Hat (64-bit)	RHEL 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8
	RHEL 7.0, 7.1, 7.2
SUSE (64-bit)	(SLES) Enterprise Linux 11, SP3[†]
	(SLES) Enterprise Linux 11, SP4
	(SLES) Enterprise Linux 12, SP1
Ubuntu (64-bit)	Ubuntu 12.04 (Precise)[†]
	Ubuntu 14.04 (Trusty)

[†]Not validated, but supported.

1.2.2. JDK Support Matrix

Unless otherwise noted, the following Java Development Kits (JDKs) are validated and supported for HDP 2.5.0:

Table 1.3. HDP 2.5.0 JDK Support Matrix

JDK	Version
Open Source	JDK7[‡]
	JDK8[† ‡]
Oracle	JDK 7
	JDK 8

[†]Not validated, but supported.

[‡]Open JDK does not run on SLES 11.

1.2.3. Metadata Database Support Matrix

Unless otherwise noted, the following metadata databases are validated and supported for HDP 2.5.0:

Table 1.4. HDP 2.5.0 Metadata Database Support Matrix

Database	Version
MariaDB	MariaDB 10.1.16
MySQL	MySQL 5.6

Database	Version
Oracle	Oracle 11gR2
	Oracle 12c
PostgreSQL	PostgreSQL 9.4

1.3. Unsupported Features

Although the following features exist within HDP 2.5.0, Hortonworks does not currently support these specific capabilities:

- Technical Preview Features [8]

- Community Features [9]

1.3.1. Technical Preview Features

The following features are available within HDP 2.5.0 but are not ready for production deployment. Hortonworks encourages you to explore these technical preview features in non-production environments and provide feedback on your experiences through the Hortonworks Community Forums.

Table 1.5. Technical Previews

Component	Feature
Atlas	New in this release: • Business Taxonomy (Catalog) (ATLAS-491). See Behavioral Changes.
HBase and Phoenix	New in this release: • Backup and restore • Phoenix operability with Microsoft .NET Driver • Phoenix Storage Handler for Hive integration Introduced in a previous release: • Cell-level ACLs • Column Family Encryption • Phoenix-Spark Integration • RPC Throttling • Support for `init.d` scripts
Hive	New in this release: • Stored Procedures Client-Side • LLAP Daemons (HIVE-7193) • LLAP / Spark security interface • Hive 2.1 • Phoenix Storage Handler for Hive integration
Slider	Introduced in a previous release:

Component	Feature
	• Support for Docker-based application packaging (SLIDER-780)
Spark	New in this release: • DataSet API • GraphX • Spark 2.0
Storm	New in this release: • Automatic back pressure • Pacemaker daemon • Resource-aware scheduling • Connectivity Enhancements: Kinesis spout, openTSDB bolt, new Kafka Consumer spout
YARN	Introduced in a previous release: • Add support for network I/O isolation/scheduling for containers (YARN-2140) • NodeManager: add cgroup support for disk I/O isolation (YARN-2619)
Zeppelin	New in this release: • Support for Phoenix interpreter.

1.3.2. Community Features

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

Table 1.6. Community Features

Component	Feature
Falcon	Introduced in a previous release: • Prism Server • User Recipes
HBase	Introduced in a previous release: • 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) • HBase Bulk Loads now replicate cross-cluster (HBASE-13153)
HDFS	Introduced in a previous release:

Component	Feature
	• HTTPFS (HDFS-2179) HTTPFS is used with Hue, in combination with Knox, and standalone as an API gateway. We highly recommend that you replace HTTPFS with Knox where possible. • NameNode Federation (HDFS-1052) • Disk Hot Swap (HDFS-1362) • Block-volume device choosing (HDFS-1804) • viewFS (HADOOP-7257)
Knox	Introduced in a previous release: • Storm REST APIs
Slider	Introduced in a previous release: • Simplified Application Packaging
Spark	Introduced in a previous release: • Spark Standalone • Spark on Mesos • Jupyter/iPython Notebook
YARN	Introduced in a previous release: • Fair Scheduler • MapReduce Eclipse Plug-in • MapReduce Uber AM
Zeppelin	For R interpreter support (ZEPPELIN-156), only the following subset of the interpreters are enabled by default: • Spark • Hive • Shell • MarkDown • R • Livy • Angular The following interpreters are not enabled by default: • Phoenix Introduced in a previous release: • The ability to share a note by sharing its URL. This is due to a lack of proper access control over with whom and how a note can be shared.

1.4. Upgrading from HDP 2.4.x to HDP 2.5.0

1.4.1. Before Upgrading

The upgrade patch contains updates to database schemas. To avoid timeouts for databases managing a large amount of users, manually execute one of the following SQL statements on ranger db before starting the upgrade.

DB Flavor	SQL Statement
MySQL	`ALTER TABLE x_user MODIFY COLUMN user_name varchar(767) NOT NULL, ADD CONSTRAINT x_user_UK_user_name UNIQUE(user_name(767));`
Oracle	`ALTER TABLE x_user MODIFY(user_name VARCHAR(767)) ADD CONSTRAINT x_user_UK_user_name UNIQUE (user_name);`
Postgres	`ALTER TABLE x_user ALTER COLUMN user_name TYPE VARCHAR(767),ADD CONSTRAINT x_user_UK_user_name UNIQUE(user_name);`
SQLServer	`ALTER TABLE [dbo].[x_user] ALTER COLUMN [user_name] [varchar](767) NOT NULL; ALTER TABLE [dbo].[x_user] ADD CONSTRAINT [x_user$x_user_UK_user_name] UNIQUE ([user_name]);`
SQL Anywhere	`ALTER TABLE dbo.x_user ALTER user_name varchar(767) NOT NULL;ALTER TABLE dbo.x_user ADD CONSTRAINT x_user_UK_user_name UNIQUE NONCLUSTERED (user_name);`

Note

Execution of SQL statement on x_user table having one million users should not take more than ~30 seconds.

1.4.2. 2.4.x Upgrade Options

HDP 2.5.0 is a feature-bearing minor release of HDP. For full instructions on how to upgrade your cluster from HDP 2.4.x to HDP 2.5.0, use the Non-Ambari Cluster Upgrade Guide.Non-Ambari Cluster Upgrade Guide

If your upgrade includes Apache Phoenix tables, see Hortonworks BUG-58308 in Known Issues.

Note

Storm does not support rolling upgrade from a previous HDP version to HDP-2.5. (see Known Issues - BUG-63535).

The following table summarizes HDP 2.2.x-to-2.5.0 upgrade options:

Cluster Management	Supporting Doc	Notes
Cluster managed manually (HDP 2.1 and earlier)	Non-Ambari Cluster Upgrade Guide	

Cluster Management	Supporting Doc	Notes
Cluster managed manually (HDP 2.2 and later)	Non-Ambari Cluster Upgrade Guide	
Cluster managed via Ambari 1.7.0	N/A	Ambari 1.7 does not support upgrading directly to an HDP 2.5 cluster. Please contact Hortonworks Support for recommendations.
Cluster managed via Ambari 2.0, 2.1, or 2.2 running HDP 2.2	Upgrading Ambari Guide	Ambari 2.x does not support upgrading directly from HDP 2.2 to HDP 2.5. You must upgrade to Ambari 2.4 prior to upgrading HDP. • Upgrade Ambari 2.x -> Ambari 2.4 • Upgrade from HDP 2.2 -> HDP 2.4 • Upgrade from HDP 2.4 -> HDP 2.5
Cluster managed via Ambari 2.0, 2.1, or 2.2 running HDP 2.3 or 2.4	Upgrading Ambari Guide	You must upgrade to Ambari 2.4 prior to upgrading HDP. • Upgrade Ambari 2.x -> Ambari 2.4 • Upgrade from HDP 2.3/2.4 -> HDP 2.5
Cluster managed via Ambari 2.4 running HDP 2.2	Upgrading Ambari Guide	Ambari 2.4 does not support upgrading directly from HDP 2.2 to HDP 2.5. • Upgrade from HDP 2.2 -> HDP 2.4 • Upgrade from HDP 2.4 -> HDP 2.5
Cluster managed via Ambari 2.4 running HDP 2.3 or 2.4	Upgrading Ambari Guide	Ambari 2.4 supports Rolling or Express upgrade from HDP 2.3/2.4 to HDP 2.5.

1.5. HDP 2.5.0 Repositories

HDP 2.5.0 for Linux supports the following:

• Debian 7 [13]

• RHEL/CentOS/Oracle LINUX 6 [13]

• RHEL/CentOS/Oracle LINUX 7 [13]

• SLES 12 [14]

• SLES 11 SP3/SP4 [13]

• Ubuntu 12 [14]

• Ubuntu 14 [14]

You can also view the Helper Files.

Note

The package identifier for HDP 2.5.0 components is **1245**. For example:

```
dfs -mkdir -p /hdp/apps/2.5.0.0-$BUILD/hive/
```

would become:

```
dfs -mkdir -p /hdp/apps/2.5.0.0-1245/hive/
```

Table 1.7. Debian 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.5.0.0/HDP-2.5.0.0-debian7-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.5.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.5.0.0/HDP-2.5.0.0-debian7-deb.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.8. RHEL/CentOS/Oracle LINUX 6

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0/HDP-2.5.0.0-centos6-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0/HDP-2.5.0.0-centos6-rpm.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.9. RHEL/CentOS/Oracle LINUX 7

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.5.0.0/HDP-2.5.0.0-centos7-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.5.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.5.0.0/HDP-2.5.0.0-centos7-rpm.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.10. SLES 11 SP3/SP4

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.5.0.0/HDP-2.5.0.0-suse11sp3-tars-tarball.tar.gz

Description	Link
HDP repo	http://public-repo-1.hortonworks.com/HDP/ suse11sp3/2.x/updates/2.5.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ suse11sp3/2.x/updates/2.5.0.0/HDP-2.5.0.0-suse11sp3-rpm.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ suse11sp3/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.11. SLES 12

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/sles12/2.x/ updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/sles12/2.x/ updates/2.5.0.0/HDP-2.5.0.0-sles12-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/sles12/2.x/ updates/2.5.0.0/hdp.repo
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/sles12/2.x/ updates/2.5.0.0/HDP-2.5.0.0-sles12-rpm.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/sles12/2.x/ updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.12. Ubuntu 12

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ ubuntu12/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/ ubuntu12/2.x/updates/2.5.0.0/HDP-2.5.0.0-ubuntu12-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/ ubuntu12/2.x/updates/2.5.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ ubuntu12/2.x/updates/2.5.0.0/HDP-2.5.0.0-ubuntu12-deb.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ ubuntu12/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.13. Ubuntu 14

Description	Link
Component metadata	http://public-repo-1.hortonworks.com/HDP/ ubuntu14/2.x/updates/2.5.0.0/build_metadata.txt
HDP Tars tarball	http://public-repo-1.hortonworks.com/HDP/ ubuntu14/2.x/updates/2.5.0.0/HDP-2.5.0.0-ubuntu14-tars-tarball.tar.gz
HDP repo	http://public-repo-1.hortonworks.com/HDP/ ubuntu14/2.x/updates/2.5.0.0/hdp.list
HDP RPM tarball	http://public-repo-1.hortonworks.com/HDP/ ubuntu14/2.x/updates/2.5.0.0/HDP-2.5.0.0-ubuntu14-deb.tar.gz
Slider app packages	http://public-repo-1.hortonworks.com/HDP/ ubuntu14/2.x/updates/2.5.0.0/slider-app-packages/pkg-list.txt

Table 1.14. Helper Files

Description	Link
Manual install helper files TAR.GZ	http://public-repo-1.hortonworks.com/HDP/tools/2.5.0.0/hdp_manual_install_rpm_helper_files-2.5.0.0.1245.tar.gz
Manual install helper files MD5	http://public-repo-1.hortonworks.com/HDP/tools/2.5.0.0/hdp_manual_install_rpm_helper_files-2.5.0.0.1245.tar.gz.md5
Manual install helper files ASC	http://public-repo-1.hortonworks.com/HDP/tools/2.5.0.0/hdp_manual_install_rpm_helper_files-2.5.0.0.1245.tar.gz.asc

1.6. Behavioral Changes

Behavioral changes denote a marked change in behavior from the previously released version to this version of software. In HDP 2.5.0, behavioral changes affect the following Hadoop components.

Table 1.15. Behavioral Changes

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
N/A	Falcon	N/A	Berkeley DB JAR file must be downloaded	Starting with the HDP 2.5.0 release, customers must obtain the Berkeley DB JAR file (available under open source license from Oracle) as part of a Falcon installation or upgrade. See the Data Movement and Integration guide for more information.
BUG-59164	Hive	N/A	Data type conversions are different between Hive 1 and Hive 2	**Component Affected**: Hive **Scenario:** In hive2, the table column data type conversion with ALTER TABLE CHANGE COLUMNS and ALTER TABLE REPLACE COLUMNS was changed. **Previous Behavior:** Previously, hive1 was very permissive and allowed changing from any primitive data type to any other primitive data type with the **DDL ALTER** statement. **New Behavior:** It is now more restrictive. By default, it only allows a small set of safe conversions. In other words, the data type being converted into is larger – it can handle more values. For example, changing a table column data type from INT to BIGINT is safe because BIGINT can represent more values than INT. In hive2, when the environment variable `hive.metastore.disallow.incompatible.col.type.changes` is true, ALTER TABLE CHANGE COLUMNS and ALTER TABLE REPLACE COLUMNS is restricted to safe conversions. The default for `hive.metastore.disallow.incompatible.col.type.changes` is `true`. To permit the old very permissive data type conversion behavior, you need to set this environment variable to `false`. Here are the safe conversions:

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
| | | | | ```
STRING_GROUP to DOUBLE
STRING_GROUP to DECIMAL
DATE_GROUP to STRING_GROUP
NUMERIC_GROUP to STRING_GROUP
STRING_GROUP to STRING_GROUP
SMALLINT -> INT -> BIGINT -> FLOAT ->
 DOUBLE -> DECIMAL
```<br><br>The last line shows increasing value ranges of the numeric types. For example, INT to FLOAT is a valid progression. FLOAT to INT is not.<br><br>```
STRING_GROUP is STRING, CHAR(n), or
 VARCHAR(n)
NUMERIC_GROUP is SMALLINT, INT, BIGINT,
 FLOAT,  DOUBLE, and DECIMAL
DATE_GROUP is DATE and TIMESTAMP
``` |
BUG-60199	Ranger	RANGER-1025	Ranger API change of behavior for HDP 2.4.0	**Component Affected:** Ranger admin **Scenario:** Search filter is not working as expected. For example : If expected search result is after first record and if in search request page size is 1. http://localhost:6080/service/public/api/policy?repositoryName=Sandbox_hadoop&resourceName=test&pageSize=1 Above search policy URL not returning a policy as expected. It seems as if the filtering is happening after retrieving the first *n* policies where *n* is the pageSize, while it should fetch all matching results first and apply the pageSize limit later. **Previous Behavior:** Searches for a policy by name limits the search to the default page size, search returns no records if the policy is farther down the list. If pageSize is big enough to include the policies we are searching for then we get the results we expect. It seems as if the filtering was happening after retrieving the first n policies where n is the pageSize. **New Behavior:** Able to search across all policies after new implementation. New pagination implementation shall send results according to requested page size after filtering the result.
BUG-60495	Hive	HIVE-14022	Left semi join should throw SemanticException if where clause contains columnname from right table	**Scenario:** Specifying left semi joins provides predictable results but only exposes columns from one table. **Previous Behavior:** Previously, you could specify left semi joins while at the same time accessing columns from the "right" table. **New Behavior:** Specifying left semi joins is no longer possible. As a result, you might experience this as a regression, which will require a query change to move forward.
BUG-61629	Zeppelin	N/A	Interpreters are now available on the top right corner, as a dropdown from the user Login button	**Component Affected:** Zeppelin JDBC Interpreters **Scenario:** Configuration of interpreter, JDBC (Generic), Spark, Livy, Shell Interpreters **Previous Behavior:** Earlier the Interpreter configuration would be present under the Notebook drop-down, which has now moved

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
				to the top-right of the page. Earlier MySQL, PgSQL, Hive, Phoenix Interpreters needed to be configured separately. **New Behavior:** Now these above described JDBC interpreters have a Generic JDBC interpreter configuration, which requires the user to additionally provide the driver class. Example - org.apache.hadoop.hive.jdbc.HiveDriver as part of interpreter properties

BUG-62267	Storm	STORM-1202	Migrate APIs to org.apache.storm, but try to provide backwards compatibility as a bridge	**Component Affected:** Storm core / trident APIs
				
				Scenario: Package name changed from:
				
				backtype.storm
				
				to:
				
				org.apache.storm
				
				Previous Behavior: You need to create a dependency on storm-core to build topologies and also import relevant classes in their code. For example:
				
				`import backtype.storm.topology. BasicOutputCollector;`
				
				New Behavior: With Apache Storm 1.0, all of the core and trident classes are moved from backtype.storm to org.apache.storm. You can import the same storm-core and trident API classes by using org.apache instead of backtype.
				
				`import org.apache.storm.topology. BasicOutputCollector;`
				
				For existing topologies you can deploy without changing the code by using the following class.
				
				`client.jartransformer.class: org.apache. storm.hack.StormShadeTransformer`
				
				Make sure you add the following configuration to `storm.yaml`.

BUG-63146	Storm	N/A	Parameter type change in org.apache.storm.spout.Scheme	**Component Affected:** Storm **Scenario:** Any user who is implementing the Scheme interface from Storm. **Previous Behavior:** Pass the byte[] parameter to the Scheme interface. **New Behavior:** Instead of passing byte[], pass ByteBuffer to the Scheme interface. See the following link for a code example: https://github.com/apache/storm/blob/1.x-branch/external/storm-kafka/src/jvm/org/apache/storm/kafka/StringScheme.java
RMP-4106	Falcon	FALCON-1107	Server Side extension infrastructure	**Scenario:** Mirroring jobs executed from the command line **Previous Behavior:** Falcon recipe tool was the client interface to execute mirroring jobs

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
				New Behavior: The extension support in Falcon CLI is used to execute mirroring jobs. Please check the upgrade documentation.
RMP-4486	Atlas	N/A	HBase Integration	**Component Affected:** HBase **Scenario:** HBase is now available as the default ATLAS storage backend via Ambari **Previous Behavior:** Previously an undocumented feature **New Behavior:** HBase can now be used to configure an HBase instance managed via Ambari or a custom HBase instance
RMP-5035	Ranger	AMBARI-15914 AMBARI-15916 RANGER-271 RANGER-900	Ranger: Remove option to store audit in DB In Ranger Audits, Audit to DB is no longer available. Users using Audit to DB must migrate to Solr. Use the HDP Security Guide - Migrating Audit Logs from DB to Solr in Ambari Clusters.	**Scenario:** Ranger Audits users who are currently using Audit to DB must migrate to Audit to Solr. **Previous Behavior:** Ranger Audit can be configured to go with any of the following destinations: DB, SOLR, and HDFS. **New Behavior:** Ranger Audit can no longer be configured to the destination DB. Ranger Audit can only be configured to go with the following destinations: SOLR and HDFS. During upgrade to HDP 2.5, If you have not enabled ranger-audit to SOLR, then you will have to configure audit to Solr post-upgrade. Otherwise, you will not see audit activities in Ranger UI. You can either use an externally managed Solr or Ambari managed Solr. For details on configuring these, refer to the Solr Audit configuration section in installation guide.
RMP-5421	Atlas	ATLAS-492	Hive Hook Phase II	**Component Affected:** Hive **Scenario:** Added support for capturing metadata changes for a table/database/column. **Previous Behavior:** Metadata changes were ignored or were known to have issues. **New Behavior:** These commands will successfully preserve the noted metadata changes: • rename table -https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-RenameTable • alter table column related commands like alter table add column, replace column, change column type, position comment etc - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-AlterColumn • alter table set TBLPROPERTIES - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-AlterTableProperties • alter database - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Create/Drop/Alter/UseDatabase

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
				• alter table storage properties - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-AlterTableStorageProperties • alter table set SERDEPROPERTIES - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-AddSerDeProperties Added support for Capturing dataset lineage in the following cases: • INSERT INTO/OVERWRITE - lineage between HDFS paths and tables - • https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries • https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Writingdataintothefilesystemfromqueries • lineage between newly created EXTERNAL tables and the HDFS paths that they are sourced from • LOAD - lineage between HDFS paths and table - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables • IMPORT/EXPORT - lineage between HDFS path and table - https://cwiki.apache.org/confluence/display/Hive/LanguageManual+ImportExport Added support for metadata capture when tables and database are dropped* - DROP TABLE, DROP DATABASE Deprecated support for hive_partition entity and no lineage for partitions are captured for hive tables. A few data model changes in the hive metadata to deprecate unused hive types and normalize the data types for consistent metadata capture
RMP-5498	Atlas	ATLAS-491	Business Taxonomy (Catalog)	**Component Affected:** Atlas **Scenario:** Enhanced search and data management **New Behavior:** • Browse business taxonomy hierarchically through graphical interface • Create child taxonomy terms • Search by taxonomy term • Search by tags

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary	Details
				• Search by combination of keyword, tag, free text in search field • Assign tags to assets (Hive, Falcon, HDFS, HBase, Storm, Kafka, Sqoop) • Assign terms to assets (Hive, Falcon, HDFS, HBase, Storm, Kafka, Sqoop) • Show audit of Atlas activity # Show current state of object (deleted or active)

1.7. Apache Patch Information

The following sections list patches in each HDP 2.5.0 component beyond what was fixed in the base version of the Apache component.

- Hadoop [21]

- Accumulo [58]

- Atlas [58]

- Calcite [61]

- Falcon [63]

- Flume [63]

- HBase [64]

- Hive [68]

- Kafka [79]

- Knox [80]

- Mahout [80]

- Oozie [81]

- Phoenix [81]

- Pig [83]

- Ranger [83]

- Slider [84]

- Spark [85]

- Sqoop [87]

- Storm [90]

- Tez [92]

- Zeppelin [100]

- ZooKeeper [101]

1.7.1. Hadoop

HDP 2.5.0 provides the following Apache patches:

- HADOOP-1540: Support file exclusion list in distcp.

- HADOOP-3733: "s3x:" URLs break when Secret Key contains a slash, even if encoded.

- HADOOP-5323: Trash documentation should describe its directory structure and configurations.

- HADOOP-5470: RunJar.unJar() should write the last modified time found in the jar entry to the uncompressed file.

- HADOOP-7139: Allow appending to existing SequenceFiles

- HADOOP-7713: dfs -count -q should label output column.

- HADOOP-7817: RawLocalFileSystem.append() should give FSDataOutputStream with accurate .getPos().

- HADOOP-8151: Error handling in snappy decompressor throws invalid exceptions.

- HADOOP-8818: Use equals instead == in MD5MD5CRC32FileChecksum and TFileDumper.

- HADOOP-8934: Shell command ls should include sort options.

- HADOOP-9121: InodeTree.java has redundant check for vName while throwing exception.

- HADOOP-9242: Duplicate surefire plugin config in HADOOP-common.

- HADOOP-9477: Add posixGroups support for LDAP groups mapping service.

- HADOOP-10035: Cleanup TestFilterFileSystem.

- HADOOP-1023: Add more logging in WASB page blob read function.

- HADOOP-1026: Logging bug in WASB causing out of index exception.

- HADOOP-10300: Allowed deferred sending of call responses.

- HADOOP-10365: BufferedOutputStream in FileUtil#unpackEntries() should be closed in finally block.

- HADOOP-10406: TestIPC.testIpcWithReaderQueuing may fail.

- HADOOP-10597: RPC Server signals backoff to clients when all request queues are full.

- HADOOP-10775: Shell operations to fail with meaningful errors on windows if winutils.exe not found.

- HADOOP-10823: TestReloadingX509TrustManager is flaky.

- HADOOP-10965: Print fully qualified path in CommandWithDestination error messages.

- HADOOP-10971: Add -C flag to make `hadoop fs -ls` print filenames only.

- HADOOP-11031: Design document for credential provider API.

- HADOOP-11098: [JDK8] Max Non Heap Memory default changed between JDK7 and 8.

- HADOOP-11212: NetUtils.wrapException to handle SocketException explicitly.

- HADOOP-11218: Add TLSv1.1,TLSv1.2 to KMS, HttpFS, SSLFactory.

- HADOOP-11252: RPC client does not time out by default.

- HADOOP-11262: Enable YARN to use S3A.

- 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-11581: Multithreaded correctness Warnings #org.apache.hadoop.fs.shell.Ls.

- HADOOP-11613: Remove commons-httpclient dependency from HADOOP-azure.

- HADOOP-11628: SPNEGO auth does not work with CNAMEs in JDK8.

- HADOOP-11685: StorageException complaining "no lease ID" during HBase distributed log splitting.

- HADOOP-11687: Ignore x-* and response headers when copying an Amazon S3 object.

- HADOOP-11692: Improve authentication failure WARN message to avoid user confusion.

- HADOOP-11713: ViewFileSystem should support snapshot methods.

- HADOOP-11717: Add Redirecting WebSSO behavior with JWT Token in Hadoop Auth.

- HADOOP-11772: RPC Invoker relies on static ClientCache which has synchronized(this) blocks.

- HADOOP-11785: Reduce the number of listStatus operation in distcp buildListing.

- HADOOP-11802: DomainSocketWatcher thread terminates sometimes after there is an I/O error during requestShortCircuitShm.

- HADOOP-11812: Implement listLocatedStatus for ViewFileSystem to speed up split calculation.

- HADOOP-11827: Speed-up distcp buildListing() using threadpool.

- HADOOP-11859: PseudoAuthenticationHandler fails with httpcomponents v4.4.

- HADOOP-11864: JWTRedirectAuthenticationHandler breaks java8 javadocs.

- HADOOP-11876: Refactor code to make it more readable, minor maybePrintStats bug.

- HADOOP-11901: BytesWritable fails to support 2G chunks due to integer overflow.

- HADOOP-11918: Listing an empty s3a root directory throws FileNotFound.

- HADOOP-11924: Tolerate JDK-8047340-related exceptions in Shell#isSetSidAvailable preventing class init.

- HADOOP-11932: MetricsSinkAdapter may hang when being stopped.

- HADOOP-11960: Enable Azure-Storage Client Side logging.

- HADOOP-12001: Fixed LdapGroupsMapping to include configurable Posix UID and GID attributes during the search.

- HADOOP-12006: Remove unimplemented option for from document.

- HADOOP-12043: Display warning if defaultFs is not set when running fs commands.

- HADOOP-12045: Enable LocalFileSystem#setTimes to change atime.

- HADOOP-12049: Control http authentication cookie persistence via configuration.

- HADOOP-12052: IPC client downgrades all exception types to IOE, breaks callers trying to use them.

- HADOOP-12058: Fix dead links to DistCp and Hadoop Archives pages.

- HADOOP-12059: S3Credentials should support use of CredentialProvider.

- HADOOP-12061: Incorrect command in single cluster setup document.

- HADOOP-12073: Azure FileSystem PageBlobInputStream does not return -1 on EOF.

- HADOOP-12074: In Shell.java#runCommand() rethrow InterruptedException as InterruptedIOException.

- HADOOP-12076: Incomplete cache mechanism in CredentialProvider API.

- HADOOP-12089: StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.

- HADOOP-12095: org.apache.hadoop.fs.shell.TestCount fails.

- HADOOP-12100: ImmutableFsPermission should not override applyUmask since that method doesn't modify the FsPermission.

- HADOOP-12103: Small refactoring of DelegationTokenAuthenticationFilter to allow code sharing.

- HADOOP-12107: Long running apps may have a huge number of StatisticsData instances under FileSystem.

- HADOOP-12119: Hadoop fs -expunge does not work for federated namespace.

- HADOOP-12159: Move DistCpUtils#compareFs() to org.apache.hadoop.fs.FileUtil and fix for HA namespaces.

- HADOOP-12161: Add getStoragePolicy API to the FileSystem interface.

- HADOOP-12169: ListStatus on empty dir in S3A lists itself instead of returning an empty list.

- HADOOP-12186: ActiveStandbyElector shouldn't call monitorLockNodeAsync multiple times.

- HADOOP-12189: Improve CallQueueManager#swapQueue to make queue elements drop nearly impossible.

- HADOOP-12191: Bzip2Factory is not thread safe.

- HADOOP-12213: Interrupted exception can occur when Client#stop is called.

- HADOOP-12230: HADOOP-project declares duplicate, conflicting curator dependencies.

- HADOOP-12232: Upgrade Tomcat dependency to 6.0.44.

- HADOOP-12239: StorageException complaining " no lease ID" when updating FolderLastModifiedTime in WASB.

- HADOOP-12258: Need translate java.nio.file.NoSuchFileException to FileNotFoundException to avoid regression.

- HADOOP-12259: Utility to Dynamic port allocation.

- HADOOP-12269: Update aws-sdk dependency to 1.10.6.

- HADOOP-12280: Skip unit tests based on maven profile rather than NativeCodeLoader.isNativeCodeLoaded.

- HADOOP-12284: UserGroupInformation doAs can throw misleading exception.

- HADOOP-12291: Add support for nested groups in LdapGroupsMapping.

- HADOOP-12292: Make use of DeleteObjects optional.

- HADOOP-12296: When setnetgrent returns 0 in Linux, exception should be thrown.

- HADOOP-12304: Applications using FileContext fail with the default file system configured to be wasb/s3/etc.

- HADOOP-12317: Applications fail on NM restart on some Linux distro because NM container recovery declares AM container as LOST.

- HADOOP-12324: Better exception reporting in SaslPlainServer.

- HADOOP-12325: RPC Metrics : Add the ability track and log slow RPCs.

- HADOOP-12334: Change Mode Of Copy Operation of HBase WAL Archiving to bypass Azure Storage Throttling after retries.

- HADOOP-12345: Pad hostname correctly in CredentialsSys.java.

- HADOOP-12346: Increase some default timeouts / retries for S3a connector.

- HADOOP-12348: MetricsSystemImpl creates MetricsSourceAdapter with wrong time unit parameter.

- HADOOP-12350: WASB Logging: Improve WASB Logging around deletes, reads and writes.

- HADOOP-12358: Add -safely flag to rm to prompt when deleting many files.

- HADOOP-12359: hadoop fs -getmerge doc is wrong.

- HADOOP-12366: Expose calculated paths (aw).

- HADOOP-12374: Description of hdfs expunge command is confusing.

- HADOOP-12406: Fixed AbstractMapWritable.readFields to use the thread's ClassLoader to load class instead of System ClassLoader.

- HADOOP-12407: Test failing: hadoop.ipc.TestSaslRPC.

- HADOOP-12413: AccessControlList should avoid calling getGroupNames in isUserInList with empty groups.

- HADOOP-12415: Add io.netty dependency to HADOOP-nfs and to HADOOP-hdfs, needed to build Bigtop successfully, see BIGTOP-2049.

- HADOOP-12416: Trash messages should be handled by Logger instead of being delivered on System.out.

- HADOOP-12423: Handle failure of registering shutdownhook by ShutdownHookManager in static block

- HADOOP-12426: Add Entry point for Kerberos health check.

- HADOOP-12437: Allow SecurityUtil to lookup alternate hostnames.

- HADOOP-12438: TestLocalFileSystem tests can fail on Windows after HDFS-8767 fix for handling pipe.

- HADOOP-12440: TestRPC#testRPCServerShutdown did not produce the desired thread states before shutting down.

- HADOOP-12441: Fixed kill-command behavior to work correctly across OSes by using bash shell built-in.

- HADOOP-12444: Support lazy seek in S3AInputStream.

- HADOOP-12463: Fix TestShell.testGetSignalKillCommand failure on windows.

- HADOOP-12464: Interrupted client may try to fail-over and retry.

- HADOOP-12465: Incorrect javadoc in WritableUtils.java.

- HADOOP-12469: distcp should not ignore the ignoreFailures option.

- HADOOP-12472: Make GenericTestUtils.assertExceptionContains robust.

- HADOOP-12478: Shell.getWinUtilsPath() has been renamed Shell.getWinutilsPath().

- HADOOP-12481: JWTRedirectAuthenticationHandler doesn't Retain Original Query String.

- HADOOP-12482: Race condition in JMX cache update.

- HADOOP-12484: Single File Rename Throws Incorrectly In Potential Race Condition Scenarios.

- HADOOP-12508: Delete fails with exception when lease is held on blob.

- HADOOP-12520: Use XInclude in HADOOP-azure test configuration to isolate Azure Storage account keys for service integration tests.

- HADOOP-12526: There are duplicate dependency definitions in poms.

- HADOOP-12533: Introduce FileNotFoundException in WASB for read and seek API.

- HADOOP-12535: Run FileSystem contract tests with HADOOP-azure.

- HADOOP-12537: S3A to support Amazon STS temporary credentials.

- HADOOP-12540: TestAzureFileSystemInstrumentation#testClientErrorMetrics fails intermittently due to assumption that a lease error will be thrown.

- HADOOP-12542: TestDNS fails on Windows after HADOOP-12437.

- HADOOP-12545: Hadoop javadoc has broken links for AccessControlList, ImpersonationProvider, DefaultImpersonationProvider, and DistCp.

- HADOOP-12548: Read s3a creds from a Credential Provider.

- HADOOP-12551: Introduce FileNotFoundException for WASB FileSystem API.

- HADOOP-12555: WASB to read credentials from a credential provider.

- HADOOP-12559: KMS connection failures should trigger TGT renewal.

- HADOOP-12564: Upgrade JUnit3 TestCase to JUnit 4 in org.apache.hadoop.io package.

- HADOOP-12565: Replace DSA with RSA for SSH key type in SingleCluster.md.

- HADOOP-12568: Update core-default.xml to describe posixGroups support.

- HADOOP-12570: HDFS Secure Mode Documentation updates.

- HADOOP-12573: TestRPC.testClientBackOff failing.

- HADOOP-12577: Bump up commons-collections version to 3.2.2 to address a security flaw.

- HADOOP-12588: Fix intermittent test failure of TestGangliaMetrics.

- HADOOP-12589: Fix intermittent test failure of TestCopyPreserveFlag.

- HADOOP-12590: TestCompressorDecompressor failing without stack traces.

- HADOOP-12602: TestMetricsSystemImpl#testQSize occasionally fails.

- HADOOP-12603: TestSymlinkLocalFSFileContext#testSetTimesSymlinkToDir occasionally fail.

- HADOOP-12604: Exception may be swallowed in KMSClientProvider.

- HADOOP-12605: Fix intermittent failure of TestIPC.testIpcWithReaderQueuingi.

- HADOOP-12608: Fix exception message in WASB when connecting with anonymous credential.

- HADOOP-12609: Fix intermittent failure of TestDecayRpcScheduler.

- HADOOP-12613: TestFind.processArguments occasionally fails.

- HADOOP-12622: Improve the loggings in RetryPolicies and RetryInvocationHandler.

- HADOOP-12635: Adding Append API support for WASB.

- HADOOP-12636: Prevent ServiceLoader failure init for unused FileSystems.

- HADOOP-12653: Use SO_REUSEADDR to avoid getting "Address already in use" when using kerberos and attempting to bind to any port on the local IP address.

- HADOOP-12659: Incorrect usage of config parameters in token manager of KMS.

- HADOOP-12663: Remove Hard-Coded Values From FileSystem.java.

- HADOOP-12668: Support excluding weak Ciphers in HttpServer2 through ssl- server.conf.

- HADOOP-12672: RPC timeout should not override IPC ping interval.

- HADOOP-12675: Fix description about retention period in usage of expunge command.

- HADOOP-12678: Handle empty rename pending metadata file during atomic rename in redo path.

- HADOOP-12682: Fix TestKMS#testKMSRestart* failure.

- HADOOP-12688: Fix dead links in Compatibility.md.

- HADOOP-12689: S3 filesystem operations stopped working correctly.

- HADOOP-12691: Add CSRF Filter for REST APIs to Hadoop Common.

- HADOOP-12691: Move files to correct location.

- HADOOP-12696: Add tests for S3Filesystem Contract.

- HADOOP-12699: TestKMS#testKMSProvider intermittently fails during 'test rollover draining'.

- HADOOP-12700: Remove unused import in TestCompressorDecompressor.java.

- HADOOP-12706: TestLocalFsFCStatistics#testStatisticsThreadLocalDataCleanUp times out occasionally.

- HADOOP-12711: Remove dependency on commons-httpclient for ServletUtil.

- HADOOP-12715: TestValueQueue#testgetAtMostPolicyALL fails intermittently.

- HADOOP-12716: KerberosAuthenticator#doSpnegoSequence use incorrect class to determine isKeyTab in JDK8.

- HADOOP-12717: NPE when trying to rename a directory in Windows Azure Storage FileSystem.

- HADOOP-12723: S3A: Add ability to plug in any AWSCredentialsProvider.

- HADOOP-12731: Remove useless boxing/unboxing code.

- HADOOP-12736: TestTimedOutTestsListener#testThreadDumpAndDeadlocks sometimes times out.

- HADOOP-12751: While using kerberos Hadoop incorrectly assumes names with '@' to be non-simple.

- HADOOP-12752: Improve diagnostics/use of envvar/sysprop credential propagation.

- HADOOP-12758: Extend CSRF Filter with UserAgent Checks.

- HADOOP-12761: Incremental maven build is not really incremental.

- HADOOP-12767: Update apache httpclient version to 4.5.2; httpcore to 4.4.4.

- HADOOP-12772: NetworkTopologyWithNodeGroup.getNodeGroup() can loop infinitely for invalid 'loc' values.

- HADOOP-12773: HBase classes fail to load with client/job classloader enabled.

- HADOOP-12776: Remove getaclstatus call for non-acl commands in getfacl.

- HADOOP-12780: During WASB atomic rename handle crash when one directory has been renamed but not file under it.

- HADOOP-12782: Faster LDAP group name resolution with ActiveDirectory.

- HADOOP-12786: "hadoop key" command usage is not documented.

- HADOOP-12787: KMS SPNEGO sequence does not work with WEBHDFS.

- HADOOP-12788: OpensslAesCtrCryptoCodec should log which random number generator is used.

- HADOOP-12789: Log classpath of ApplicationClassLoader at INFO level.

- HADOOP-12792: TestUserGroupInformation#testGetServerSideGroups fails in chroot.

- HADOOP-12793: Write a new group mapping service guide.

- HADOOP-12794: Support additional compression levels for GzipCodec.

- HADOOP-12795: KMS does not log detailed stack trace for unexpected errors.

- HADOOP-12800: Copy docker directory from 2.8 to 2.7/2.6 repos to enable pre-commit Jenkins runs.

- HADOOP-12801: Suppress obsolete S3FileSystem tests.

- HADOOP-12805: Annotate CanUnbuffer with @InterfaceAudience.Public

- HADOOP-12807: S3AFileSystem should read AWS credentials from environment variables.

- HADOOP-12810: FileSystem#listLocatedStatus causes unnecessary RPC calls

- HADOOP-12817: Enable TLS v1.1 and 1.2.

- HADOOP-12825: Log slow name resolutions.

- HADOOP-12828: Print user when services are started.

- HADOOP-12829: StatisticsDataReferenceCleaner swallows interrupt exceptions.

- HADOOP-12831: LocalFS/FSOutputSummer NPEs in constructor if bytes per checksum set to 0.

- HADOOP-12844: Recover when S3A fails on IOException in read().

- HADOOP-12846: Credential Provider Recursive Dependencies.

- HADOOP-12847: hadoop daemonlog should support https and SPNEGO for Kerberized cluster.

- HADOOP-12849: TestSymlinkLocalFSFileSystem fails intermittently.

- HADOOP-12851: S3AFileSystem Uptake of ProviderUtils.excludeIncompatibleCredentialProviders.

- HADOOP-12853: Change WASB documentation regarding page blob support.

- HADOOP-12870: Fix typo admininistration in CommandsManual.md.

- HADOOP-12871: Fix dead link to NativeLibraries.html in CommandsManual.md.

- HADOOP-12872: Fix formatting in ServiceLevelAuth.md.

- HADOOP-12878: KMS SPNEGO sequence does not work with WEBHDFS.

- HADOOP-12886: Exclude weak ciphers in SSLFactory through ssl-server.xml.

- HADOOP-12891: S3AFileSystem should configure Multipart Copy threshold and chunk size.

- 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-12902: JavaDocs for SignerSecretProvider are out-of-date in AuthenticationFilter.

- HADOOP-12903: IPC Server should allow suppressing exception logging by type.

- HADOOP-12903: IPC Server should allow suppressing exception logging by type, not log 'server too busy' messages.

- HADOOP-12906: AuthenticatedURL should convert a 404/Not Found into an FileNotFoundException.

- HADOOP-12909: Change ipc.Client to support asynchronous calls.

- HADOOP-12916: Allow RPC scheduler/callqueue backoff using response times.

- HADOOP-12926: lz4.c does not detect 64-bit mode properly.

- HADOOP-12929: JWTRedirectAuthenticationHandler must accommodate null expiration time.

- HADOOP-12943: Add -w -r options in dfs -test command.

- HADOOP-12947: Update documentation Hadoop Groups Mapping to add static group mapping, negative cache.

- HADOOP-12950: ShutdownHookManager should have a timeout for each of the Registered shutdown hook.

- HADOOP-12957: Limit the number of outstanding async calls.

- HADOOP-12958: PhantomReference for filesystem statistics can trigger OOM.

- HADOOP-12962: KMS key names are incorrectly encoded when creating key.

- HADOOP-12963: Allow using path style addressing for accessing the s3 endpoint.

- HADOOP-12964: HTTP server vulnerable to clickjacking.

- HADOOP-12973: Make DU pluggable.

- HADOOP-12975: Add jitter to CachingGetSpaceUsed's thread.

- HADOOP-12982: Document missing S3A and S3 properties.

- HADOOP-12984: Add GenericTestUtils.getTestDir method and use it for temporary directory in tests.

- HADOOP-12985: Support MetricsSource interface for DecayRpcScheduler Metrics.

- HADOOP-12989: Some tests in org.apache.hadoop.fs.shell.find occasionally time out.

- HADOOP-12993: Change ShutdownHookManger complete shutdown log from INFO to DEBUG.

- HADOOP-12994: Specify PositionedReadable, add contract tests, fix problems.

- HADOOP-13008: Add XFS Filter for UIs to Hadoop Common.

- HADOOP-13011: Clearly Document the Password Details for Keystore-based Credential Providers.

- HADOOP-13026: Should not wrap IOExceptions into a AuthenticationException in KerberosAuthenticator.

- HADOOP-13028: Add low level counter metrics for S3A; use in read performance tests.

- HADOOP-13030: Handle special characters in passwords in KMS startup script.

- HADOOP-13039: Add documentation for configuration property ipc.maximum.data.length

- HADOOP-13042: Restore lost leveldbjni LICENSE and NOTICE changes.

- HADOOP-13047: S3a Forward seek in stream length to be configurable.

- HADOOP-13052: ChecksumFileSystem mishandles crc file permissions.

- HADOOP-13058: S3A FS fails during init against a read-only FS if multipart purge.

- HADOOP-13065: Add a new interface for retrieving FS and FC Statistics.

- HADOOP-13072: WindowsGetSpaceUsed constructor should be public.

- HADOOP-13079: Add -q option to Ls to print ? instead of non-printable characters.

- HADOOP-13084: Fix ASF License warnings.

- HADOOP-13098: Dynamic LogLevel setting page should accept case-insensitive log level string.

- HADOOP-13103: Group resolution from LDAP may fail on javax.naming.ServiceUnavailableException.

- HADOOP-13105: Support timeouts in LDAP queries in LdapGroupsMapping.

- HADOOP-13113: Enable parallel test execution for HADOOP-aws.

- HADOOP-13116: Jets3tNativeS3FileSystemContractTest does not run.

- HADOOP-13122: Customize User-Agent header sent in HTTP requests by S3A.

- HADOOP-13130: s3a failures can surface as RTEs, not IOEs.

- HADOOP-13131: Add tests to verify that S3A supports SSE-S3 encryption.

- HADOOP-13138: Unable to append to a SequenceFile with Compression.NONE.

- HADOOP-13140: FileSystem#initialize must not attempt to create StorageStatistics objects with null or empty schemes.

- HADOOP-13145: In DistCp, prevent unnecessary getFileStatus call when not preserving metadata.

- HADOOP-13146: Refactor RetryInvocationHandler.

- HADOOP-13155: Implement TokenRenewer to renew and cancel delegation tokens in KMS.

- HADOOP-13158: S3AFileSystem#toString might throw NullPointerException due to null cannedACL.

- HADOOP-13159: Fix potential NPE in Metrics2 source for DecayRpcScheduler.

- HADOOP-13162: Consider reducing number of getFileStatus calls in S3AFileSystem.mkdirs.

- HADOOP-13163: Reuse pre-computed filestatus in Distcp-CopyMapper.

- HADOOP-13168: Support Future.get with timeout in ipc async calls.

- HADOOP-13171: Add StorageStatistics to S3A; instrument some more operations.

- HADOOP-13179: GenericOptionsParser is not thread-safe because commons-cli OptionBuilder is not thread-safe.

- HADOOP-13183: S3A proxy tests fail after httpclient/httpcore upgrade.

- HADOOP-13192: org.apache.hadoop.util.LineReader cannot handle multibyte delimiters correctly.

- HADOOP-13197: Add non-decayed call metrics for DecayRpcScheduler.

- HADOOP-13203: S3a: Consider reducing the number of connection aborts by setting correct length in s3 request.

- HADOOP-13213: Small Documentation bug with AuthenticatedURL in HADOOP-auth.

- HADOOP-13237: s3a initialization against public bucket fails if caller lacks any credentials.

- HADOOP-13241: Document s3a better.

- HADOOP-13244: o.a.h.ipc.Server#Server should honor handlerCount when queueSizePerHandler is specified in constructor.

- HADOOP-13249: RetryInvocationHandler need wrap InterruptedException in IOException when call Thread.sleep.

- 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-13280: FileSystemStorageStatistics#getLong("readOps") should return readOps + largeReadOps.

- HADOOP-13283: Support reset operation for new global storage statistics and per FS storage stats.

- HADOOP-13285: DecayRpcScheduler MXBean should only report decayed CallVolumeSummary.

- HADOOP-13287: TestS3ACredentials#testInstantiateFromURL fails if AWS secret key contains +.

- HADOOP-13288: Guard null stats key in FileSystemStorageStatistics.

- HADOOP-13290: Appropriate use of generics in FairCallQueue.

- HADOOP-13291: Probing stats in DFSOpsCountStatistics/S3AStorageStatistics should be correctly implemented.

- HADOOP-13305: Define common statistics names across schemes.

- HADOOP-13315: FileContext#umask is not initialized properly.

- HADOOP-13316: Enforce Kerberos authentication for required ops in DelegationTokenAuthenticator.

- HADOOP-13320: Fix arguments check in documentation for WordCount v2.0.

- HADOOP-13350: Additional fix to LICENSE and NOTICE.

- HADOOP-13351: TestDFSClientSocketSize buffer size tests are flaky.

- HADOOP-13352: Make X-FRAME-OPTIONS configurable in HttpServer2.

- HADOOP-13368: DFSOpsCountStatistics$OpType#fromSymbol and s3a.Statistic#fromSymbol should be O(1) operation.

- HADOOP-13382: Remove unneeded commons-httpclient dependencies from POM files in Hadoop and sub-projects, while adding it in to the only place it is still needed, HADOOP-openstack.

- HADOOP-13422: ZKDelegationTokenSecretManager JaasConfig does not work well with other ZK users in process.

- HADOOP-13434: Add quoting to Shell class.

- HADOOP-13467: Shell#getSignalKillCommand should use the bash built-in on Linux.

- HDFS-1477: Support reconfiguring dfs.heartbeat.interval and dfs.namenode.heartbeat.recheck-interval without NN restart.

- HDFS-2043: TestHFlush failing intermittently.

- HDFS-2580: NameNode#main(.) can make use of GenericOptionsParser.

- HDFS-27: HDFS CLI with –config set to default config complains log file not found error.

- HDFS-3677: dfs.namenode.edits.dir.required is missing from HDFS- default.xml.

- HDFS-4015: Safemode should count and report orphaned blocks.

- HDFS-4185: Add a metric for number of active leases.

- HDFS-4366: Block Replication Policy Implementation May Skip Higher-Priority Blocks for Lower-Priority Blocks.

- HDFS-4396: Add START_MSG/SHUTDOWN_MSG for ZKFC.

- HDFS-4937: ReplicationMonitor can infinite-loop in BlockPlacementPolicyDefault#chooseRandom.

- HDFS-4946: Allow preferLocalNode in BlockPlacementPolicyDefault to be configurable.

- HDFS-5640: Add snapshot methods to FileContext.

- HDFS-6054: MiniQJMHACluster should not use static port to avoid binding failure in unit test.

- HDFS-6101: TestReplaceDatanodeOnFailure fails occasionally.

- HDFS-6407: Add sorting and pagination in the datanode tab of the NN Web UI.

- HDFS-6481: DatanodeManager#getDatanodeStorageInfos() should check the length of storageIDs.

- HDFS-6533: TestBPOfferService#testBasicFunctionalitytest fails intermittently.

- HDFS-6581: Support for writing to single replica in RAM. Applied HDFS-6581 merge.15.patch to Champlain. This is a consolidated patch containing all feature changes and was applied to trunk.

- HDFS-6666: Abort NameNode and DataNode startup if security is enabled but block access token is not enabled.

- HDFS-6832: Fix the usage of 'hdfs namenode' command.

- HDFS-6860: BlockStateChange logs are too noisy.

- HDFS-6945: BlockManager should remove a block from excessReplicateMap and decrement ExcessBlocks metric when the block is removed.

- HDFS-7163: WebHdfsFileSystem should retry reads according to the configured retry policy.

- 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 DFSClient.

- HDFS-7390: Provide JMX metrics per storage type.

- HDFS-7452: skip StandbyException log for getCorruptFiles()

- HDFS-7483: Display information per tier on the Namenode UI.

- HDFS-7597: DelegationTokenIdentifier should cache the TokenIdentifier to UGI mapping.

- HDFS-7608: HDFS dfsclient newConnectedPeer has no write timeout.

- HDFS-7609: Avoid retry cache collision when Standby NameNode loading edits.

- HDFS-7645: Fix CHANGES.txt.

- HDFS-7645: Rolling upgrade is restoring blocks from trash multiple times.

- HDFS-7701: Support reporting per storage type quota and usage with hadoop/hdfs shell.

- HDFS-7725: Incorrect "nodes in service" metrics caused all writes to fail.

- HDFS-7833: DataNode reconfiguration does not recalculate valid volumes required, based on configured failed volumes tolerated.

- HDFS-7858: Improve HA Namenode Failover detection on the client.

- HDFS-7890: Improve information on Top users for metrics in RollingWindowsManager and lower log level.

- HDFS-7923: The DataNodes should rate-limit their full block reports by asking the NN on heartbeat messages.

- HDFS-7928: Scanning blocks from disk during rolling upgrade startup takes a lot of time if disks are busy.

- HDFS-7933: fsck should also report decommissioning replicas.

- HDFS-7978: Add LOG.isDebugEnabled() guard for some LOG.debug.

- HDFS-7990: IBR delete ack should not be delayed.

- HDFS-8008: Support client-side back off when the datanodes are congested.

- HDFS-8009: Signal congestion on the DataNode.

- HDFS-8046: Allow better control of getContentSummary.

- HDFS-8055: NullPointerException when topology script is missing.

- HDFS-8071: Redundant checkFileProgress() in PART II of getAdditionalBlock().

- HDFS-8099: Change "DFSInputStream has been closed already" message to debug log level.

- HDFS-8101: DFSClient use of non-constant DFSConfigKeys pulls in WebHDFS classes at runtime.

- HDFS-8113: Add check for null BlockCollection pointers in BlockInfoContiguous structures.

- HDFS-8131: Implement a space balanced block placement policy.

- HDFS-8144: Split TestLazyPersistFiles into multiple tests.

- HDFS-8152: Refactoring of lazy persist storage cases.

- HDFS-8155: Support OAuth2 in WebHDFS.

- HDFS-8180: AbstractFileSystem Implementation for WebHdfs.

- HDFS-8205: CommandFormat#parse() should not parse option as value of option.

- HDFS-8209: Support different number of datanode directories in MiniDFSCluster.

- HDFS-8211: DataNode UUID is always null in the JMX counter.

- HDFS-8219: setStoragePolicy with folder behavior is different after cluster restart.

- HDFS-8229: LAZY_PERSIST file gets deleted after NameNode restart.

- HDFS-8232: Missing datanode counters when using Metrics2 sink interface.

- HDFS-8276: LazyPersistFileScrubber should be disabled if scrubber interval configured zero.

- HDFS-8278: When computing max-size-to-move in Balancer, count only the storage with remaining >= default block size.

- HDFS-8309: Skip unit test using DataNodeTestUtils#injectDataDirFailure() on Windows.

- HDFS-8310: Fix TestCLI.testAll "help: help for find" on Windows.

- HDFS-8311: DataStreamer.transfer() should timeout the socket InputStream.

- HDFS-8337: Accessing httpfs via webhdfs doesn't work from a jar with Kerberos.

- HDFS-8345: Storage policy APIs must be exposed via the FileSystem interface.

- HDFS-8384: Allow NN to startup if there are files having a lease but are not under construction.

- HDFS-8431: HDFS crypto class not found in Windows.

- HDFS-8432: Introduce a minimum compatible layout version to allow downgrade in more rolling upgrade use cases.

- HDFS-8435: Support CreateFlag in WebHDFS.

- HDFS-8539: Hdfs doesn't have class 'debug' in windows.

- HDFS-8542: WebHDFS getHomeDirectory behavior does not match specification.

- HDFS-8546: Use try with resources in DataStorage and Storage.

- HDFS-8548: Minicluster throws NPE on shutdown.

- HDFS-8549: Abort the balancer if an upgrade is in progress.

- HDFS-8554: TestDatanodeLayoutUpgrade fails on Windows.

- HDFS-8578: On upgrade, Datanode should process all storage/data dirs in parallel.

- HDFS-8581: ContentSummary on / skips further counts on yielding lock.

- HDFS-8582: Support getting a list of reconfigurable config properties and do not generate spurious reconfig warnings.

- HDFS-8593: Calculation of effective layout version mishandles comparison to current layout version in storage.

- HDFS-8615: Correct HTTP method in WebHDFS document.

- HDFS-8633: Fix setting of dfs.datanode.readahead.bytes in HDFS-default.xml to match DFSConfigKeys.

- HDFS-8642: Make TestFileTruncate more reliable.

- HDFS-8647: Abstract BlockManager's rack policy into BlockPlacementPolicy.

- HDFS-8656: Preserve compatibility of ClientProtocol#rollingUpgrade after finalization.

- HDFS-8659: Block scanner INFO message is spamming logs.

- HDFS-8676: Delayed rolling upgrade finalization can cause heartbeat expiration.

- HDFS-8696: Make the lower and higher watermark in the DN Netty server configurable.

- HDFS-8709: Clarify automatic sync in FSEditLog#logEdit.

- HDFS-8722: Optimize datanode writes for small writes and flushes.

- HDFS-8729: Fix TestFileTruncate#testTruncateWithDataNodesRestartImmediately which occasionally failed.

- HDFS-8767: RawLocalFileSystem.listStatus() returns null for UNIX pipefile.

- HDFS-8772: Fix TestStandbyIsHot#testDatanodeRestarts which occasionally fails.

- HDFS-8778: TestBlockReportRateLimiting#testLeaseExpiration can deadlock.

- HDFS-8785: TestDistributedFileSystem is failing in trunk.

- HDFS-8797: WebHdfsFileSystem creates too many connections for pread.

- HDFS-8806: Inconsistent metrics: number of missing blocks with replication factor 1 not properly cleared.

- HDFS-8809: HDFS fsck reports under construction blocks as CORRUPT.

- HDFS-8815: DFS getStoragePolicy implementation using single RPC call.

- HDFS-8816: Improve visualization for the Datanode tab in the NN UI.

- HDFS-8818: Changes the global moveExecutor to per datanode executors and changes MAX_SIZE_TO_MOVE to be configurable.

- HDFS-8824: Do not use small blocks for balancing the cluster.

- HDFS-8826: In Balancer, add an option to specify the source node list so that balancer only selects blocks to move from those nodes.

- HDFS-8828: Utilize Snapshot diff report to build diff copy list in distcp.

- HDFS-8829: Make SO_RCVBUF and SO_SNDBUF size configurable for DataTransferProtocol sockets and allow configuring auto-tuning.

- 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-8846: Add a unit test for INotify functionality across a layout version upgrade.

- HDFS-8850: VolumeScanner thread exits with exception if there is no block pool to be scanned but there are suspicious blocks.

- HDFS-8852: HDFS architecture documentation of version 2.x is outdated about append write support.

- HDFS-8855: Webhdfs client leaks active NameNode connections.

- HDFS-8856: Make LeaseManager#countPath O(1).

- HDFS-8863: The remaining space check in BlockPlacementPolicyDefault is flawed.

- HDFS-8867: Enable optimized block reports.

- HDFS-8879: Quota by storage type usage incorrectly initialized upon namenode restart.

- HDFS-8880: NameNode metrics logging.

- HDFS-8883: NameNode Metrics : Add FSNameSystem lock Queue Length.

- HDFS-8885: ByteRangeInputStream used in webhdfs does not override available().

- HDFS-8887: Expose storage type and storage ID in BlockLocation.

- HDFS-8891: HDFS concat should keep srcs order.

- HDFS-8896: DataNode object isn't GCed when shutdown, because it has GC root in ShutdownHookManager.

- HDFS-8911: NameNode Metric : Add Editlog counters as a JMX metric.

- HDFS-8914: Document HA support in the HDFS HdfsDesign.md.

- HDFS-8923: Add -source flag to balancer usage message.

- HDFS-8930: Block report lease may leak if the 2nd full block report comes when NN is still in safemode.

- HDFS-8932: NPE thrown in NameNode when try to get TotalSyncCount metric before editLogStream initialization.

- HDFS-8939: Test(S)WebHdfsFileContextMainOperations failing.

- HDFS-8950: NameNode refresh doesn't remove DataNodes that are no longer in the allowed list.

- HDFS-8964: When validating the edit log, do not read at or beyond the file offset that is being written.

- HDFS-8965: Harden edit log reading code against out of memory errors.

- HDFS-8969: Clean up findbugs warnings for HDFS-8823 and HDFS-8932.

- HDFS-8983: NameNode support for protected directories.

- HDFS-8995: Flaw in registration bookkeeping can make DN die on reconnect.

- HDFS-9008: Balancer#Parameters class could use a builder pattern.

- HDFS-9009: Send metrics logs to NullAppender by default.

- HDFS-9019: Adding informative message to sticky bit permission denied exception.

- HDFS-9033: dfsadmin -metasave prints "NaN" for cache used%.

- HDFS-9042: Update document for the Storage policy name.

- HDFS-9043: Doc update for commands in HDFS Federation.

- HDFS-9047: Retire libwebhdfs.

- HDFS-9048: DistCp documentation is out-of-dated.

- HDFS-9063: Correctly handle snapshot path for getContentSummary.

- HDFS-9072: Fix random failures in TestJMXGet.

- HDFS-9082: Change the log level in WebHdfsFileSystem.initialize() from INFO to DEBUG.

- HDFS-9083: Replication violates block placement policy.

- HDFS-9094: Add command line option to ask NameNode reload configuration.

- HDFS-9106: Transfer failure during pipeline recovery causes permanent write failures.

- HDFS-9107: Prevent NNs unrecoverable death spiral after full GC.

- HDFS-9112: Improve error message for Haadmin when multiple name service IDs are configured.

- HDFS-9128: TestWebHdfsFileContextMainOperations and TestSWebHdfsFileContextMainOperations fail due to invalid HDFS path on Windows.

- HDFS-9130: Use GenericTestUtils#setLogLevel to the logging level.

- HDFS-9142: Separating Configuration object for namenode(s) in MiniDFSCluster.

- HDFS-9175: Change scope of 'AccessTokenProvider.getAccessToken()' and 'CredentialBasedAccessTokenProvider.getCredential()' abstract methods to public.

- HDFS-9178: Slow datanode I/O can cause a wrong node to be marked bad.

- HDFS-9184: Logging HDFS operation's caller context into audit logs.

- HDFS-9198: Coalesce IBR processing in the NN.

- HDFS-9205: Do not schedule corrupt blocks for replication.

- HDFS-9210: Fix some misuse of %n in VolumeScanner#printStats.

- HDFS-9214: Support reconfiguring dfs.datanode.balance.max.concurrent.moves without DN restart.

- HDFS-9220: Reading small file (< 512 bytes) that is open for append fails due to incorrect checksum.

- HDFS-9221: HdfsServerConstants#ReplicaState#getState should avoid calling values() since it creates a temporary array.

- HDFS-9226: Unit Tests failing due to mockito version issue.

- HDFS-9239: DataNode Lifeline Protocol: an alternative protocol for reporting DataNode liveness.

- HDFS-9259: Make SO_SNDBUF size configurable at DFSClient side for hdfs write scenario.

- HDFS-9273: ACLs on root directory may be lost after NN restart.

- HDFS-9289: Make DataStreamer#block thread safe and verify genStamp in commitBlock.

- HDFS-9290: DFSClient#callAppend() is not backward compatible for slightly older NameNodes.

- HDFS-9294: DFSClient deadlock when close file and failed to renew lease.

- HDFS-9305: Delayed heartbeat processing causes storm of subsequent heartbeats.

- HDFS-9311: Support optional offload of NameNode HA service health checks to a separate RPC server.

- HDFS-9313: Possible NullPointerException in BlockManager if no excess replica can be chosen.

- HDFS-9314: Improve BlockPlacementPolicyDefault's picking of excess replicas.

- HDFS-9317: Document fsck -blockId and -storagepolicy options.

- HDFS-9343: Empty caller context considered invalid.

- HDFS-9347: Invariant assumption in TestQuorumJournalManager.shutdown() is wrong.

- HDFS-9349: Support reconfiguring fs.protected.directories without NN restart.

- HDFS-9354: Fix TestBalancer#testBalancerWithZeroThreadsForMove on Windows.

- HDFS-9358: TestNodeCount#testNodeCount timed out.

- HDFS-9362: TestAuditLogger#testAuditLoggerWithCallContext assumes Unix line endings, fails on Windows.

- HDFS-9364: Unnecessary DNS resolution attempts when creating NameNodeProxies.

- HDFS-9383: TestByteArrayManager#testByteArrayManager fails.

- HDFS-9384: TestWebHdfsContentLength intermittently hangs and fails due to TCP conversation mismatch between client and server.

- HDFS-9395: Make HDFS audit logging consistent.

- HDFS-9397: Fix typo for readChecksum() LOG.warn in BlockSender.java.

- HDFS-9402: Switch DataNode.LOG to use slf4j.

- HDFS-9406: FSImage may get corrupted after deleting snapshot.

- HDFS-9412: getBlocks occupies FSLock and takes too long to complete.

- HDFS-9413: getContentSummary() on standby should throw StandbyException.

- HDFS-9414: Refactor reconfiguration of ClientDatanodeProtocol for reusability.

- HDFS-9415: Document dfs.cluster.administrators and dfs.permissions.superusergroup.

- HDFS-9426: Rollingupgrade finalization is not backward compatible.

- HDFS-9428: Fix intermittent failure of TestDNFencing.testQueueingWithAppend.

- HDFS-9431: DistributedFileSystem#concat fails if the target path is relative.

- HDFS-9434: Recommission a datanode with 500k blocks may pause NN for 30 seconds for printing info log messages.

- HDFS-9445: Datanode may deadlock while handling a bad volume.

- HDFS-9466: TestShortCircuitCache#testDataXceiverCleansUpSlotsOnFailure is flaky.

- HDFS-9470: Encryption zone on root not loaded from fsimage after NN restart.

- HDFS-9476: TestDFSUpgradeFromImage#testUpgradeFromRel1BBWImage occasionally fail.

- HDFS-9478: Reason for failing ipc.FairCallQueue construction should be thrown.

- HDFS-9493: Test o.a.h.hdfs.server.namenode.TestMetaSave fails in trunk.

- HDFS-9505: HDFS Architecture documentation needs to be refreshed.

- HDFS-9516: Truncate file fails with data dirs on multiple disks.

- HDFS-9517: Fix missing @Test annotation on TestDistCpUtils.testUnpackAttributes.

- 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 Blocks.

- HDFS-9533: seen_txid in the shared edits directory is modified during bootstrapping.

- HDFS-9534: Add CLI command to clear storage policy from a path.

- HDFS-9549: TestCacheDirectives#testExceedsCapacity is flaky.

- HDFS-9555: LazyPersistFileScrubber should still sleep if there are errors in the clear progress.

- HDFS-9557: Reduce object allocation in PB conversion.

- 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-9572: Prevent DataNode log spam if a client connects on the data transfer port but sends no data.

- HDFS-9574: Reduce client failures during datanode restart.

- HDFS-9584: NPE in distcp when ssl configuration file does not exist in class path.

- HDFS-9600: Do not check replication if the block is under construction.

- HDFS-9605: Add links to failed volumes to explorer.html in HDFS Web UI.

- HDFS-9608: Disk IO imbalance in HDFS with heterogeneous storages.

- HDFS-9612: DistCp worker threads are not terminated after jobs are done.

- HDFS-9619: SimulatedFSDataset sometimes can not find blockpool for the correct namenode.

- HDFS-9623: Update example configuration of block state change log in log4j.properties.

- HDFS-9624: DataNode start slowly due to the initial DU command operations.

- HDFS-9625: Set replication for empty file failed when set storage policy.

- HDFS-9626: TestBlockReplacement#testBlockReplacement fails occasionally.

- HDFS-9629: Update the footer of Web UI to show year 2016.

- HDFS-9630: DistCp minor refactoring and clean up.

- HDFS-9634: WebHDFS client side exceptions don't provide enough details.

- HDFS-9638: Improve DistCp Help and documentation.

- HDFS-9644: Update encryption documentation to reflect nested EZs.

- HDFS-9655: NN should start JVM pause monitor before loading fsimage.

- HDFS-9661: Deadlock in DN.FsDatasetImpl between moveBlockAcrossStorage and createRbw.

- HDFS-9669: TcpPeerServer should respect ipc.server.listen.queue.size.

- HDFS-9670: DistCp throws NPE when source is root.

- HDFS-9672: o.a.h.hdfs.TestLeaseRecovery2 fails intermittently.

- HDFS-9688: Test the effect of nested encryption zones in HDFS downgrade.

- HDFS-9706: Log more details in debug logs in BlockReceiver's constructor.

- HDFS-9710: DN can be configured to send block receipt IBRs in batches.

- HDFS-9711: Integrate CSRF prevention filter in WebHDFS.

- 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-9724: Degraded performance in WebHDFS listing as it does not reuse ObjectMapper.

- HDFS-9726: Refactor IBR code to a new class.

- HDFS-9730: Storage ID update does not happen when there is a layout change.

- HDFS-9739: DatanodeStorage.isValidStorageId() is broken.

- HDFS-9740: Use a reasonable limit in DFSTestUtil.waitForMetric().

- HDFS-9743: Fix TestLazyPersistFiles#testFallbackToDiskFull.

- HDFS-9748: Avoid duplication in pendingReplications when addExpectedReplicasToPending is called twice.

- HDFS-9752: Permanent write failures may happen to slow writers during datanode rolling upgrades.

- HDFS-9760: WebHDFS AuthFilter cannot be configured with custom AltKerberos auth handler.

- HDFS-9764: DistCp doesn't print value for several arguments including -numListstatusThreads.

- HDFS-9765: TestBlockScanner#testVolumeIteratorWithCaching fails intermittently.

- HDFS-9766: TestDataNodeMetrics#testDataNodeTimeSpend fails intermittently.

- HDFS-9768: Reuse ObjectMapper instance in HDFS to improve the performance.

- HDFS-9772: TestBlockReplacement#testThrottler doesn't work as expected.

- HDFS-9777: Fix typos in DFSAdmin command line and documentation.

- HDFS-9779: TestReplicationPolicyWithNodeGroup NODE variable picks wrong rack value.

- HDFS-9784: Example usage is not correct in Transparent Encryption document.

- HDFS-9790: HDFS Balancer should exit with a proper message if upgrade is not finalized.

- HDFS-9797: Log Standby exceptions thrown by RequestHedgingProxyProvider at DEBUG Level.

- HDFS-9799: Reimplement getCurrentTrashDir to remove incompatibility.

- HDFS-9801: ReconfigurableBase should update the cached configuration.

- HDFS-9812: Streamer threads leak if failure happens when closing DFSOutputStream.

- HDFS-9839: Reduce verbosity of processReport logging.

- HDFS-9842: dfs.datanode.balance.bandwidthPerSec should accept friendly size units.

- HDFS-9844: Correct path creation in getTrashRoot to handle root dir.

- HDFS-9851: NameNode throws NPE when setPermission is called on a path that does not exist.

- HDFS-9854: Log cipher suite negotiation more verbosely.

- HDFS-9864: Correct reference for RENEWDELEGATIONTOKEN and CANCELDELEGATIONTOKEN in webhdfs doc.

- HDFS-9865: TestBlockReplacement fails intermittently in trunk.

- 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-9880: TestDatanodeRegistration fails occasionally.

- 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-9904: testCheckpointCancellationDuringUpload occasionally fails.

- HDFS-9905: WebHdfsFileSystem#runWithRetry should display original stack trace on error.

- HDFS-9906: Remove spammy log spew when a datanode is restarted.

- HDFS-9917: IBR accumulate more objects when SNN was down for sometime.

- HDFS-9926: MiniDFSCluster leaks dependency Mockito via DataNodeTestUtils.

- HDFS-9941: Do not log StandbyException on NN, other minor logging fixes.

- HDFS-9949: Add a test case to ensure that the DataNode does not regenerate its UUID when a storage directory is cleared.

- HDFS-9958: BlockManager#createLocatedBlocks can throw NPE for corruptBlocks on failed storages.

- HDFS-10178: Permanent write failures can happen if pipeline recoveries occur for the first packet.

- 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-10199: Unit tests TestCopyFiles, TestDistCh, TestLogalyzer under org.apache.hadoop.tools are failing.

- HDFS-10207: Support enable Hadoop IPC backoff without namenode restart.

- HDFS-10209: Support enable caller context in HDFS namenode audit log without restart namenode.

- HDFS-10216: Distcp -diff throws exception when handling relative path.

- HDFS-10217: Show 'blockScheduled' tooltip in datanodes table.

- HDFS-10223: peerFromSocketAndKey performs SASL exchange before setting connection timeouts.

- HDFS-10224: Implement asynchronous rename for DistributedFileSystem.

- HDFS-10228: TestHDFSCLI fails.

- HDFS-10235: Last contact for Live Nodes should be relative time.

- HDFS-10239: Fsshell mv fails if port usage doesn't match in src and destination paths.

- HDFS-10245: Fix the findbugs warnings.

- HDFS-10253: Fix TestRefreshCallQueue failure.

- HDFS-10261: TestBookKeeperHACheckpoints doesn't handle ephemeral HTTP ports.

- HDFS-10264: Logging improvements in FSImageFormatProtobuf.Saver.

- HDFS-10270: TestJMXGet:testNameNode() fails.

- HDFS-10271: Extra bytes are getting released from reservedSpace for append.

- HDFS-10275: TestDataNodeMetrics failing intermittently due to TotalWriteTime counted incorrectly.

- HDFS-10277: PositionedReadable test testReadFullyZeroByteFile failing in HDFS.

- HDFS-10281: TestPendingCorruptDnMessages fails intermittently.

- HDFS-10291: TestShortCircuitLocalRead failing.

- HDFS-10309: Balancer doesn't honor dfs.blocksize value defined with suffix k(kilo), m(mega), g(giga).

- HDFS-10312: Large block reports may fail to decode at NameNode due to 64 MB protobuf maximum length restriction.

- HDFS-10313: Distcp need to enforce the order of snapshot names passed to -diff.

- 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-10335: Mover$Processor#chooseTarget() always chooses the first matching target storage group.

- HDFS-10341: Add a metric to expose the timeout number of pending replication blocks.

- HDFS-10344: DistributedFileSystem#getTrashRoots should skip encryption zone that does not have .Trash.

- HDFS-10346: Implement asynchronous setPermission/setOwner for DistributedFileSystem.

- HDFS-10347: Namenode report bad block method doesn't log the bad block or datanode.

- HDFS-10360: DataNode may format directory and lose blocks if current/VERSION is missing.

- HDFS-10367: TestDFSShell.testMoveWithTargetPortEmpty fails with Address bind exception.

- HDFS-10372: Fix for failing TestFsDatasetImpl#testCleanShutdownOfVolume.

- HDFS-10377: CacheReplicationMonitor shutdown log message should use INFO level.

- HDFS-10390: Implement asynchronous setAcl/getAclStatus for DistributedFileSystem.

- HDFS-10396: Using -diff option with DistCp may get "Comparison method violates its general contract" exception.

- HDFS-10397: Distcp should ignore -delete option if -diff option is provided instead of exiting.

- HDFS-10410: RedundantEditLogInputStream.LOG is set to wrong class.

- HDFS-10415: TestDistributedFileSystem#MyDistributedFileSystem attempts to set up statistics before initialize() is called.

- HDFS-10417: Improve error message from checkBlockLocalPathAccess.

- HDFS-10424: DatanodeLifelineProtocol not able to use under security cluster.

- HDFS-10431: Refactor and speedup TestAsyncDFSRename.

- HDFS-10437: ReconfigurationProtocol not covered by HDFSPolicyProvider.

- HDFS-10438: When NameNode HA is configured to use the lifeline RPC server, it should log the address of that server.

- HDFS-10440: Improve DataNode web UI.

- HDFS-10448: CacheManager#addInternal tracks bytesNeeded incorrectly when dealing with replication factors other than 1.

- 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-10471: DFSAdmin#SetQuotaCommand's help msg is not correct.

- HDFS-10474: hftp copy fails when file name with Chinese+special char.

- HDFS-10481: HTTPFS server should correctly impersonate as end user to open file.

- HDFS-10485: Fix findbugs warning in FSEditLog.java.

- HDFS-10488: Update WebHDFS documentation regarding CREATE and MKDIR default permissions.

- 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-10525: Fix NPE in CacheReplicationMonitor#rescanCachedBlockMap.

- HDFS-10556: DistCpOptions should be validated automatically.

- HDFS-10579: HDFS web interfaces lack configs for X-FRAME-OPTIONS protection.

- HDFS-10589: Javadoc for HAState#HAState and HAState#setStateInternal contains non-existent params.

- HDFS-10617: PendingReconstructionBlocks.size() should be synchronized.

- HDFS-10623: Remove unused import of httpclient.HttpConnection from TestWebHdfsTokens.

- HDFS-10642: TestLazyPersistReplicaRecovery#testDnRestartWithSavedReplicas fails intermittently.

- HDFS-10643: HDFS namenode should always use service user (hdfs) to generateEncryptedKey.

- HDFS-10660: Expose storage policy APIs via HDFSAdmin interface.

- HDFS-10688: BPServiceActor may run into a tight loop for sending block report when hitting IOException.

- MAPREDUCE-5485: Allow repeating job commit by extending OutputCommitter API.

- MAPREDUCE-6197: Cache MapOutputLocations in ShuffleHandler.

- MAPREDUCE-6273: HistoryFileManager should check whether summaryFile exists to avoid FileNotFoundException causing HistoryFileInfo into MOVE_FAILED state.

- MAPREDUCE-6302: Backport preempt reducers after a configurable timeout irrespective of headroom.

- MAPREDUCE-6436: JobHistory cache issue.

- MAPREDUCE-6460: TestRMContainerAllocator.testAttemptNotFoundCausesRMCommunicatorException fails.

- MAPREDUCE-6492: AsyncDispatcher exit with NPE on TaskAttemptImpl#sendJHStartEventForAssignedFailTask.

- MAPREDUCE-6514: Job hangs as ask is not updated after ramping down of all reducers.

- MAPREDUCE-6543: Migrate MR Client test cases part 2.

- MAPREDUCE-6549: multibyte delimiters with LineRecordReader cause duplicate records.

- MAPREDUCE-6566: Add retry support to mapreduce CLI tool.

- MAPREDUCE-6577: MR AM unable to load native library without MR_AM_ADMIN_USER_ENV set.

- MAPREDUCE-6579: TestNetworkedJob.testNetworkedJob failed due to YARN-3946.

- MAPREDUCE-6593: TestJobHistoryEventHandler.testTimelineEventHandling fails because of NPE.

- MAPREDUCE-6618: YarnClientProtocolProvider leaking the YarnClient thread.

- MAPREDUCE-6621: Memory Leak in JobClient#submitJobInternal().

- MAPREDUCE-6635: Unsafe long to int conversion in UncompressedSplitLineReader and IndexOutOfBoundsException.

- MAPREDUCE-6639: Process hangs in LocatedFileStatusFetcher if FileSystem.get throws.

- MAPREDUCE-6672: TestTeraSort fails on Windows.

- MAPREDUCE-6689: MapReduce job can infinitely increase number of reducer resource requests.

- YARN-998: Keep NM resource updated through dynamic resource config for RM/NM restart.

- YARN-1376: NM need to notify the log aggregation status to RM through heartbeat.

- YARN-1402: Update related Web UI and CLI with exposing client API to check log aggregation status.

- YARN-1462: AHS API and other AHS changes to handle tags for completed MR jobs.

- YARN-1556: NPE getting application report with a null appId.

- YARN-1880: Cleanup TestApplicationClientProtocolOnHA.

- YARN-2019: EAR-1682. Retrospect on decision of making RM crashed if any exception throw in ZKRMStateStore.

- YARN-2046: Out of band heartbeats are sent only on container kill and possibly too early.

- YARN-2194: Fix bug causing CGroups functionality to fail on RHEL7.

- YARN-2392: Add more diags about app retry limits on AM failures.

- YARN-2495: Allow admin specify labels from each NM (Distributed configuration for node label).

- YARN-2498: Respect labels in preemption policy of capacity scheduler for inter-queue preemption.

- YARN-2513: Host framework UIs in YARN for use with the ATS Note: Patch is in progress.

- YARN-2571: RM to support YARN registry.

- YARN-2619: Added NodeManager support for disk IO isolation through cgroups.

- YARN-2666: TestFairScheduler.testContinuousScheduling fails intermittently.

- YARN-2696: Queue sorting in CapacityScheduler should consider node label.

- YARN-2740: Fix NodeLabelsManager to properly handle node label modifications when distributed node label configuration enabled.

- YARN-2821: Fixed a problem that DistributedShell AM may hang if restarted.

- YARN-2868: FairScheduler: Metric for latency to allocate first container for an application.

- YARN-2901: Add errors and warning metrics page to RM, NM web UI.

- YARN-3021: YARN's delegation-token handling disallows certain trust setups to operate properly over DistCp.

- YARN-3026: Move application-specific container allocation logic from LeafQueue to FiCaSchedulerApp.

- YARN-3102: Decommissioned Nodes not listed in Web UI.

- YARN-3110: Few issues in ApplicationHistory web UI.

- YARN-3136: Fixed a synchronization problem of AbstractYarnScheduler#getTransferredContainers.

- YARN-3215: Respect labels in CapacityScheduler when computing headroom.

- YARN-3216: Max-AM-Resource-Percentage should respect node labels.

- YARN-3225: New parameter of CLI for decommissioning node gracefully in RMAdmin CLI.

- YARN-3243: CapacityScheduler should pass headroom from parent to children to make sure ParentQueue obey its capacity limits.

- YARN-3248: Display count of nodes blacklisted by apps in the web UI.

- YARN-3266: RMContext#inactiveNodes should have NodeId as map key.

- YARN-3269: Yarn.nodemanager.remote-app-log-dir could not be configured to fully qualified path.

- YARN-3293: Track and display capacity scheduler health metrics in web UI.

- YARN-3294: Allow dumping of Capacity Scheduler debug logs via web UI for a fixed time period.

- YARN-3305: Normalize AM resource request on app submission.

- YARN-3318: Create Initial OrderingPolicy Framework and FifoOrderingPolicy.

- YARN-3319: Implement a FairOrderingPolicy.

- YARN-3326: Support RESTful API for getLabelsToNodes.

- YARN-3343: Increased TestCapacitySchedulerNodeLabelUpdate#testNodeUpdate timeout.

- YARN-3345: Add non-exclusive node label API.

- YARN-3347: Improve YARN log command to get AMContainer logs as well as running containers logs.

- YARN-3348: Add a 'yarn top' tool to help understand cluster usage.

- YARN-3354: Add node label expression in ContainerTokenIdentifier to support RM recovery.

- YARN-3356: Capacity Scheduler FiCaSchedulerApp should use ResourceUsage to track used-resources-by-label.

- YARN-3357: Move TestFifoScheduler to FIFO package.

- YARN-3360: Add JMX metrics to TimelineDataManager.

- YARN-3361: CapacityScheduler side changes to support non-exclusive node labels.

- YARN-3362: Add node label usage in RM CapacityScheduler web UI.

- YARN-3365: Enhanced NodeManager to support using the 'tc' tool via container-executor for outbound network traffic control.

- YARN-3366: Enhanced NodeManager to support classifying/shaping outgoing network bandwidth traffic originating from YARN containers.

- YARN-3383: AdminService should use "warn" instead of "info" to log exception when operation fails.

- YARN-3387: Previous AM's container completed status couldn't pass to current AM if AM and RM restarted during the same time.

- YARN-3394: Enrich WebApplication proxy documentation.

- YARN-3397: YARN rmadmin should skip -failover.

- YARN-3404: Display queue name on application page.

- YARN-3406: Display count of running containers in the RM's Web UI.

- YARN-3410: YARN admin should be able to remove individual application records from RMStateStore.

- YARN-3413: Changed Nodelabel attributes (like exclusivity) to be settable only via addToClusterNodeLabels but not changeable at runtime.

- YARN-3425: NPE from RMNodeLabelsManager.serviceStop when NodeLabelsManager.serviceInit failed.

- YARN-3433: Jersey tests failing with Port in Use -again. .

- YARN-3435: AM container to be allocated Appattempt AM container shown as null.

- YARN-3443: Create a 'ResourceHandler' subsystem to ease addition of support for new resource types on the NM.

- YARN-3448: Added a rolling time-to-live LevelDB timeline store implementation.

- YARN-3451: Display attempt start time and elapsed time on the web UI.

- YARN-3459: Fix failure of TestLog4jWarningErrorMetricsAppender.

- YARN-3463: Integrate OrderingPolicy Framework with CapacityScheduler.

- YARN-3480: Remove attempts that are beyond max-attempt limit from state store.

- YARN-3494: Expose AM resource limit and usage in CS QueueMetrics.

- YARN-3503: Expose disk utilization percentage and bad local and log dir counts in NM metrics.

- YARN-3505: Node's Log Aggregation Report with SUCCEED should not cached in RMApps.

- YARN-3511: Add errors and warnings page to ATS.

- YARN-3517: RM web UI for dumping scheduler logs should be for admins only.

- YARN-3521: Support return structured NodeLabel objects in REST API.

- YARN-3530: ATS throws exception on trying to filter results without otherinfo.

- YARN-3541: Add version info on timeline service / generic history web UI and REST API.

- YARN-3543: ApplicationReport should be able to tell whether the Application is AM managed or not.

- YARN-3552: RM Web UI shows -1 running containers for completed apps.

- YARN-3565: NodeHeartbeatRequest/RegisterNodeManagerRequest should use NodeLabel object instead of String.

- YARN-3579: CommonNodeLabelsManager should support NodeLabel instead of string label name when getting node-to-label/label-to-label mappings.

- YARN-3580: [JDK8] TestClientRMService.testGetLabelsToNodes fails.

- YARN-3581: Deprecate -directlyAccessNodeLabelStore in RMAdminCLI.

- YARN-3583: Support of NodeLabel object instead of plain String in YarnClient side.

- YARN-3593: Add label-type and Improve "DEFAULT_PARTITION" in Node Labels Page.

- YARN-3600: AM container link is broken (on a killed application, at least).

- YARN-3604: removeApplication in ZKRMStateStore should also disable watch.

- YARN-3632: Ordering policy should be allowed to reorder an application when demand changes.

- **YARN-3635**: Refactored current queue mapping implementation in CapacityScheduler to use a generic PlacementManager framework.

- **YARN-3654**: ContainerLogsPage web UI should not have meta-refresh.

- **YARN-3695**: ServerProxy (NMProxy, etc.) shouldn't retry forever for non network exception.

- **YARN-3700**: Made generic history service load a number of latest applications according to the parameter or the configuration.

- **YARN-3707**: RM Web UI queue filter doesn't work.

- **YARN-3717**: Expose app/am/queue's node-label-expression to RM web UI / CLI / REST-API.

- **YARN-3727**: For better error recovery, check if the directory exists before using it for localization.

- **YARN-3740**: Fixed the typo in the configuration name: APPLICATION_HISTORY_PREFIX_MAX_APPS.

- **YARN-3751**: Fixed AppInfo to check if used resources are null.

- **YARN-3766**: Fixed the apps table column error of generic history web UI.

- **YARN-3769**: Consider user limit when calculating total pending resource for preemption policy in Capacity Scheduler.

- **YARN-3785**: Support for Resource as an argument during submitApp call in MockRM test class.

- **YARN-3846**: RM Web UI queue filter is not working for sub queue.

- **YARN-3849**: Too much of preemption activity causing continuous killing of containers across queues.

- **YARN-3873**: PendingApplications in LeafQueue should also use OrderingPolicy.

- **YARN-3885**: ProportionalCapacityPreemptionPolicy doesn't preempt if queue is more than 2 level.

- **YARN-3894**: RM startup should fail for wrong CS xml NodeLabel capacity configuration.

- **YARN-3896**: RMNode transitioned from RUNNING to REBOOTED because its response id has not been reset synchronously.

- **YARN-3932**: SchedulerApplicationAttempt#getResourceUsageReport and UserInfo should based on total-used-resources.

- **YARN-3941**: Proportional Preemption policy should try to avoid sending duplicate PREEMPT_CONTAINER event to scheduler.

- **YARN-3946**: Update exact reason as to why a submitted app is in ACCEPTED state to app's diagnostic message.

- YARN-3948: Display Application Priority in RM Web UI.

- YARN-3961: Expose pending, running and reserved containers of a queue in REST API and YARN top.

- YARN-3966: Fix excessive loggings in CapacityScheduler.

- YARN-3971: Skip RMNodeLabelsManager#checkRemoveFromClusterNodeLabelsOfQueue on nodelabel recovery.

- YARN-3978: Configurably turn off the saving of container info in Generic AHS.

- YARN-3983: Refactored CapacityScheduleri#FiCaSchedulerApp to easier extend container allocation logic.

- YARN-3986: getTransferredContainers in AbstractYarnScheduler should be present in YarnScheduler interface.

- YARN-3987: AM container complete msg ack to NM once RM receive it.

- YARN-4000: RM crashes with NPE if leaf queue becomes parent queue during restart.

- YARN-4023: Publish Application Priority to TimelineServer.

- YARN-4026: Refactored ContainerAllocator to accept a list of priories rather than a single priority.

- YARN-4082: Container shouldn't be killed when node's label updated.

- YARN-4087: Followup fixes after YARN-2019 regarding RM behavior when state-store error occurs.

- YARN-4092: Fixed UI redirection to print useful messages when both RMs are in standby mode.

- YARN-4101: RM should print alert messages if ZooKeeper and Resourcemanager gets connection issue.

- YARN-4108: CapacityScheduler: Improve preemption to only kill containers that would satisfy the incoming request.

- YARN-4140: RM container allocation delayed in case of app submitted to Nodelabel partition.

- YARN-4149: YARN logs -am should provide an option to fetch all the log files.

- YARN-4155: TestLogAggregationService.testLogAggregationServiceWithInterval failing.

- YARN-4162: CapacityScheduler: Add resource usage by partition and queue capacity by partition to REST API.

- YARN-4215: RMNodeLabels Manager Need to verify and replace node labels for the only modified Node Label Mappings in the request.

- YARN-4225: Add preemption status to yarn queue -status for capacity scheduler.

- YARN-4233: YARN Timeline Service plugin: ATS v1.5.

- YARN-4285: Display resource usage as percentage of queue and cluster in the RM UI.

- YARN-4287: Capacity Scheduler: Rack Locality improvement.

- YARN-4288: Fixed RMProxy to retry on IOException from local host.

- YARN-4290: Add -showDetails option to YARN Nodes CLI to print all nodes reports information.

- YARN-4304: AM max resource configuration per partition to be displayed/updated correctly in UI and in various partition related metrics.

- YARN-4309: Add container launch related debug information to container logs when a container fails.

- YARN-4313: Race condition in MiniMRYarnCluster when getting history server address.

- YARN-4315: NaN in Queue percentage for cluster apps page.

- YARN-4345: YARN rmadmin -updateNodeResource doesn't work.

- YARN-4347: Resource manager fails with Null pointer exception.

- YARN-4349: YARN_APPLICATION call to ATS does not have YARN_APPLICATION_CALLER_CONTEXT.

- YARN-4365: FileSystemNodeLabelStore should check for root dir existence on startup.

- YARN-4384: updateNodeResource CLI should not accept negative values for resource.

- YARN-4390: Do surgical preemption based on reserved container in CapacityScheduler.

- YARN-4405: Support node label store in non-appendable file system.

- YARN-4414: Nodemanager connection errors are retried at multiple levels.

- YARN-4416: Deadlock due to synchronised get Methods in AbstractCSQueue.

- YARN-4417: Make RM and Timeline-server REST APIs more consistent.

- YARN-4418: AM Resource Limit per partition can be updated to ResourceUsage as well.

- YARN-4422: Generic AHS sometimes doesn't show started, node, or logs on App page.

- YARN-4428: Redirect RM page to AHS page when AHS turned on and RM page is not available.

- YARN-4439: Clarify NMContainerStatus#toString method.

- YARN-4454: NM to nodelabel mapping going wrong after RM restart.

- YARN-4479: Change CS LeafQueue pendingOrderingPolicy to honor recovered apps.

- YARN-4497: RM might fail to restart when recovering apps whose attempts are missing.

- YARN-4502: Fix two AM containers get allocated when AM restart.

- YARN-4522: Queue acl can be checked at app submission.

- YARN-4535: Fix checkstyle error in CapacityScheduler.java.

- YARN-4546: ResourceManager crash due to scheduling opportunity overflow.

- YARN-4557: Fix improper Queues sorting in PartitionedQueueComparator when accessible-node-labels=*.

- YARN-4565: Sometimes when sizeBasedWeight FairOrderingPolicy is enabled, under stress appears that cluster is virtually in deadlock.

- YARN-4573: Fix test failure in TestRMAppTransitions#testAppRunningKill and testAppKilledKilled.

- YARN-4584: RM startup failure when AM attempts greater than max-attempts.

- YARN-4598: Invalid event: RESOURCE_FAILED at CONTAINER_CLEANEDUP_AFTER_KILL.

- YARN-4607: Pagination support for AppAttempt page TotalOutstandingResource Requests table.

- YARN-4610: Reservations continue looking for one app causes other apps to starve.

- YARN-4617: LeafQueue#pendingOrderingPolicy should always use fixed ordering policy instead of using same as active applications ordering policy.

- YARN-4623: TestSystemMetricsPublisher#testPublishAppAttemptMetricsForUnmanagedAM fails with NPE. 7.

- YARN-4625: Make ApplicationSubmissionContext and ApplicationSubmissionContextInfo more consistent.

- YARN-4633: TestRMRestart.testRMRestartAfterPreemption fails intermittently.

- YARN-4634: Scheduler UI/Metrics need to consider cases like non-queue label mappings.

- YARN-4667: RM Admin CLI for refreshNodesResources throws NPE when nothing is configured.

- YARN-4696: TimelineClient to add flush operation for deterministic writes (including testing) and Changes to EntityGroupFSTimelineStore for testability.

- YARN-4699: Scheduler UI and REST o/p is not in sync when -replaceLabelsOnNode is used to change label of a node.

- YARN-4709: NMWebServices produces incorrect JSON for containers.

- YARN-4723: NodesListManager$UnknownNodeId ClassCastException.

- YARN-4737: Add CSRF filter support in YARN.

- YARN-4746: YARN web services should convert parse failures of appId, appAttemptId and containerId to 400.

- YARN-4769: Add support for CSRF header in the dump capacity scheduler logs and kill app buttons in RM web UI.

- YARN-4785: Inconsistent value type of the "type" field for LeafQueueInfo in response of RM REST API.

- YARN-4811: Generate histograms in ContainerMetrics for actual container resource usage.

- YARN-4814: ATS 1.5 timelineclient impl call flush after every event write.

- YARN-4815: ATS 1.5 timelineclinet impl try to create attempt directory for every event call.

- YARN-4822: Refactor existing Preemption Policy of CS for easier adding new approach to select preemption candidates.

- YARN-4832: NM side resource value should get updated if change applied in RM side.

- YARN-4844: Add getMemorySize/getVirtualCoresSize to o.a.h.y.api.records.Resource.

- YARN-4846: Fix random failures for TestCapacitySchedulerPreemption #testPreemptionPolicyShouldRespectAlreadyMarkedKillableContainers.

- YARN-4865: Track Reserved resources in ResourceUsage and QueueCapacities.

- YARN-4890: Unit test intermittent failure: TestNodeLabelContainerAllocation#testQueueUsedCapacitiesUpdate.

- YARN-4893: Fix intermittent failures in TestRMAdminService.

- YARN-4916: TestNMProxy.tesNMProxyRPCRetry fails.

- YARN-4934: Reserved Resource for QueueMetrics needs to be handled correctly in few cases.

- YARN-4955: Add retry for SocketTimeoutException in TimelineClient.

- YARN-4984: LogAggregationService shouldn't swallow exception in handling createAppDir() which cause thread leak.

- YARN-5023: TestAMRestart#testShouldNotCountFailureToMaxAttemptRetry random failure.

- YARN-5048: DelegationTokenRenewer#skipTokenRenewal may throw NPE.

- YARN-5076: YARN web interfaces lack XFS (Cross-Frame Script) protection.

- YARN-5112: Excessive log warnings for directory permission issue on NM recovery.

- YARN-5131: Distributed shell AM fails when extra container arrives during finishing.

- YARN-5190: Registering/unregistering container metrics triggered by ContainerEvent and ContainersMonitorEvent are conflict which cause uncaught exception in ContainerMonitorImpl.

- YARN-5214: Fixed locking in DirectoryCollection to avoid hanging NMs when various code-paths hit slow disks.

- YARN-5270: Solve miscellaneous issues caused by YARN-4844.

- YARN-5294: Pass remote ip address down to YarnAuthorizationProvider.

- YARN-5296: NMs going OutOfMemory because ContainerMetrics leak in ContainerMonitorImpl.

- YARN-5297: Avoid printing a stack trace when recovering an app after the RM restarts.

- YARN-5337: Fix Dshell OOM issues.

1.7.2. Accumulo

HDP 2.5.0 provides Accumulo 1.7.1 and the following Apache patches:

- ACCUMULO-4361: Configure ShellServerIT to run with one tserver.

- ACCUMULO-4368: Offline and online the table when constraints are altered.

- ACCUMULO-4369: Restructure try-catch to ensure a logical exception is raised.

- ACCUMULO-4370: Replace bounded checks for replication table state with unbounded checks.

1.7.3. Atlas

HDP 2.5.0 provides Atlas 0.7.0 and the following Apache patches:

- ATLAS-347: Atlas search APIs should allow pagination of results.

- ATLAS-584: Integrate CSRF prevention filter.

- ATLAS-639: Exception for lineage request.

- ATLAS-655: Please delete old releases from mirroring system.

- ATLAS-762: Assertion in NegativeSSLAndKerberosTest.testUnsecuredClient needs to be hardened.

- ATLAS-847: UI Audit versioning does not paginate details from Atlas server.

- ATLAS-861: 1 table out of 50,000 tables is left unimported throwing exception during deserialization.

- ATLAS-902: Atlas throws exception due to null definition in Hive create table statement.

- ATLAS-917: Add hdfs paths to process qualified name for non-partition based queries.

- ATLAS-936: Update atlas website for 0.7 release.

- ATLAS-949: UI improvement for modal and tag styling in table.

- ATLAS-957: Atlas is not capturing topologies that have $ in the data payload.

- ATLAS-963: UI Entity details is not display String array attribute values correctly.

- ATLAS-965: Old lineage still exists after dropping tables and re-creating tables with same name.

- ATLAS-966: Exit execution of import_hive.sh if HIVE_HOME is not set.

- ATLAS-968: Set group information from UGI for Ldap authentication.

- ATLAS-970: Remove glyphicon from login.jsp.

- ATLAS-971: UI not displaying results for this query - Eg "hive_table as t where qualifiedName = 'default.input@cl1' select t".

- ATLAS-986: Ability to differentiate business catalog terms from traits.

- ATLAS-987: Atlas hooks should avoid adding dependent libraries to component CLASSPATH.

- ATLAS-988: HiveHookIT.testInsertIntoTable is broken.

- ATLAS-990: Hive Import metadata script fails with auth exception.

- ATLAS-993: If condition in DSL order by clause is not defined then DSL query fails.

- ATLAS-995: Atlas to setup ldap authentication type as either LDAP / AD or None.

- ATLAS-996: DSL queries with comparisons of many primitive types fail.

- ATLAS-998: Determine HA mode from property atlas.server.ids, instead of atlas.server.ha.enabled.

- ATLAS-1001: UI Paginate search APIs.

- ATLAS-1002: Create default user rangertagsync in atlas file authentication for Ranger tag sync module.

- ATLAS-1003: DataSetLineageServiceTest, GraphBackedDiscoveryServiceTest, and GraphRepoMapperScaleTest failing in some environments.

- ATLAS-1004: Option to enable taxonomy feature.

- ATLAS-1006: Paginate full text search results.

- ATLAS-1009: Source HIVE_HOME and HIVE_CONF_DIR from hive_env.sh.

- ATLAS-1010: Atlas allows recreation of tags with same name.

- ATLAS-1021: Update Atlas architecture wiki.

- ATLAS-1022: Update typesystem wiki with details.

- ATLAS-1025: Set HIVE_HOME if hive is available in relative path to import hive script.

- ATLAS-1026: StoreBackedTypeCache issues.

- ATLAS-1027: Atlas hooks should use properties from ATLAS-application.properties, instead of component's configuration.

- ATLAS-1030: Add instrumentation to measure performance REST API.

- ATLAS-1032: Atlas hook package should not include libraries already present in host component - like log4j.

- ATLAS-1033: Fix for issues flagged by Coverity scan.

- ATLAS-1034: Incorrect Falcon hook impl class name in Falcon hook shim.

- ATLAS-1036: Compilation error on java 1.8 - GraphBackedDiscoveryService.

- ATLAS-1038: Multiple instances of AtlasPluginClassloader getting initialized.

- ATLAS-1042: Performance improvement changes for propertykey+typeName based queries.

- ATLAS-1046: Search pagination refinements.

- ATLAS-1048: TestMetadata.py test in distro project fails on Windows.

- ATLAS-1049: Fix validation while filtering by supertypes.

- ATLAS-1049: List types by supertype.

- ATLAS-1051: Sqoop Hook does not package HDFS model jars which is required.

- ATLAS-1052: Fix NPE in HiveHook due to null Session State.

- ATLAS-1053: Fix issues flagged by Coverity scan - potential NPE.

- ATLAS-1056: Differentiate between tag and term using attribute "taxonomy.namespace".

- ATLAS-1059: Change log level to debug for search APIs.

- ATLAS-1060: Add composite indexes for exact match performance improvements for all attributes.

- ATLAS-1064: Pagination for full text search results.

- ATLAS-1065: Full text search view same as DSL's.

- ATLAS-1066: Falcon fails to post entity to Atlas due to Kafka exception.

- ATLAS-1071: Regression - UI - Details Button under Audits Tab is not working.

- ATLAS-1080: Regression - UI - hive_storagedesc is shown as "undefined" in UI.

- ATLAS-1086: Build failure in hive-bridge after security fixes in ATLAS-762.

- ATLAS-1087: Provide an option to turn off persisting entity definition in audits.

- ATLAS-1088: Fix /search api to default to full text on DSL failure.

- ATLAS-1090: Multi-Select Tagging.

- ATLAS-1091: Improvement in DSL search functionality.

- ATLAS-1092: Add Table.CreateTime to process qualified Name for all hive_process for lineage to work on dropped/recreated tables.

- ATLAS-1096: Modify HiveMetaStoreBridge.import to use getEntity instead of DSL.

- ATLAS-1099: Multiple tag assign button hides wrongly.

- ATLAS-1103: Search type list is not refreshed.

- ATLAS-1104: Get outgoing edges by label doesn't work in some cases.

- ATLAS-1105: Disable HiveLiteralRewriterTest since its not used currently.

- ATLAS-1108: In Atlas HA mode , import-hive.sh in Passive instance fails.

- ATLAS-1111: Data loss is observed when atlas is restarted while hive_table metadata ingestion into Kafka topic is in-progress.

- ATLAS-1112: Hive table GET response from atlas server had duplicate column entries.

- ATLAS-1115: Show Tag / Taxonomy Listing in sorted order.

- ATLAS-1119: Add retries for edge label creation.

- ATLAS-1121: NPE while submitting topology in StormHook.

1.7.4. Calcite

HDP 2.5.0 provides Calcite 1.2.0 and the following Apache patches:

- CALCITE-259: Using sub-queries in CASE statement against JDBC tables generates invalid Oracle SQL.

- CALCITE-429: Add statistics SPI for lattic optimization algorithm.

- CALCITE-522: In remote JDBC driver, transmit static database properties as a map.

- CALCITE-585: Avatica JDBC methods should throw SQLFeatureNotSupportedException.

- CALCITE-645: Pass server-side exceptions back to the client.

- CALCITE-661: Remote fetch in Calcite JDBC driver.

- CALCITE-671: ByteString does not deserialize properly as a FetchRequest parameterValue.

- CALCITE-677: RemoteDriverTest.testTypeHandling fails east of Greenwich.

- CALCITE-687: Make RemoteDriverTest thread-safe.

- CALCITE-699: In Avatica, synchronize access to Calendar.

- CALCITE-705: DML in Avatica, and split Execute out from Fetch request.

- CALCITE-708: Avatica and Calcite to support DatabaseMetaData getTypeInfo.

- CALCITE-712: Avatica statement execute return all resultset instead of MaxRows from setMaxRows.

- CALCITE-717: Compare BINARY and VARBINARY on unsigned byte values.

- CALCITE-718: Enable fetch to work for Statement.execute().

- CALCITE-728: Test suite hangs on Windows.

- CALCITE-730: ClassCastException in table from CloneSchema.

- CALCITE-741: Dependencies should not be empty.

- CALCITE-765: RPC server returns JSON data with Content-Type set to text/html.

- CALCITE-767: Commit functionality not exposed by the RPC server.

- CALCITE-780: HTTP error 413 when sending a long string to the Avatica server.

- CALCITE-789: MetaImpl.MetaCatalog should expose TABLE_CAT instead of TABLE_CATALOG.

- CALCITE-795: Loss of precision when sending a decimal number via the remote JSON service.

- CALCITE-813: Upgrade updateCount, maxRows from int to long.

- CALCITE-825: Allow user to specify sort order of an ArrayTable.

- CALCITE-840: Protobuf transport for Avatica.

- CALCITE-843: AvaticaConnection.getAutoCommit throws NullPointerException.

- CALCITE-865: NullPointerException in getTables with PostgreSQL.

- CALCITE-866: Create documentation for RPC message format(s).

- CALCITE-871: JdbcResultSet returns incomplete Frame with "default" statement ID.

- CALCITE-903: Enable client to recover from missing server-side state.

- CALCITE-905: getTables returns empty result in JdbcMeta.

- CALCITE-906: Avatica JDbcMeta statement IDs are not unique.

- CALCITE-908: Bump protobuf dependency to protobuf-3.0.0-beta-1.

- CALCITE-910: Improve handling of ARRAY, MULTISET, STRUCT types.

- CALCITE-912: Pass URL connection properties to avatica server.

- CALCITE-913: Avatica transport of array fields fails.

- CALCITE-914: Missing JsonSubType for ExecuteResponse.

- CALCITE-919: ArithmeticException when querying against decimal field.

- CALCITE-921: ClassCastException byte[] to String with protobuf.

- CALCITE-927: ColumnsRequest Service call doesn't "fix" ResultSetResponse.

- CALCITE-951: Print the server-side stack in the local exception.

- CALCITE-962: Server-side exception (stack trace) not propagated in JdbcMeta.propagate.

- CALCITE-984: NPE in ErrorResponse construction.

- CALCITE-989: Provide generic server metadata in responses.

- CALCITE-1156: Increase jetty version to 9.2.15.v20160210.

- CALCITE-1239: Upgrade to avatica-1.8.0.

1.7.5. Falcon

HDP 2.5.0 provides Falcon 0.10.0 and the following Apache patches:

- FALCON-1944: Ability to provide additional DistCP options.

- FALCON-2066: Extend support for Import and Export of various database types.

- FALCON-2072: Hive2 URLs in Falcon should allow additional configuration elements in the URL.

- FALCON-2082: Add CSRF filter for REST APIs.

- FALCON-2083: Handle connector specific direct mode and verbose options for Database Import and Export.

- FALCON-2093: Database import and export to support TD connector.

- FALCON-2095: Hive Replication jobs are failing with UnknownHostException in NN HA.

- FALCON-2117: Implement X-Frame-Options header for Falcon UI.

- FALCON-2118: Proposal for new UI changes for FALCON.

1.7.6. Flume

HDP 2.5.0 provides Flume 1.5.2 and the following Apache patch:

- FLUME-2498: Implement Taildir Source.

- FLUME-2500: Add a channel that uses Kafka.

- FLUME-2523: Document Kafka channel.

- FLUME-2560: Add Kafka Channel to dist tarballs.

- FLUME-2562: Add metrics for Kafka Source, Kafka Sink and Kafka Channel.

- FLUME-2729: Allow pollableSource backoff times to be configurable.

- FLUME-2734: Kafka Channel timeout property is overridden by default value.

- FLUME-2781: Kafka Channel with parseAsFLUMEEvent=true should write data as is, not as FLUME events.

- FLUME-2801: Performance improvement on TailDir source.

- FLUME-2823: FLUME-Kafka-Channel with new APIs.

- FLUME-2901: Document Kerberos setup for Kafka channel.

- FLUME-2947: Upgrade Hive and thrift dependencies.

1.7.7. HBase

HDP 2.5.0 provides HBase 1.1.2 and the following Apache patches:

- HBASE-5291: Add Kerberos HTTP SPNEGO authentication support toHBASE web consoles.

- HBASE-13318: RpcServer.getListenerAddress should handle when the accept channel is closed.

- HBASE-13829: Add more ThrottleType.

- HBASE-14123: HBASE Backup/Restore Phase 2.

- HBASE-14140: HBASE Backup Admin API.

- HBASE-14172: Upgrade existing thrift binding using thrift 0.9.3 compiler.

- HBASE-14258: Make region_mover.rb script case insensitive with regard to hostname.

- HBASE-14269: FuzzyRowFilter omits certain rows when multiple fuzzy key exist.

- HBASE-14283: Reverse scan doesn't work with HFile inline index/bloom blocks.

- HBASE-14302: TableSnapshotInputFormat should not create back references when restoring snapshot.

- HBASE-14307: Incorrect use of positional read api in HFileBlock.

- HBASE-14313: After a Connection sees ConnectionClosingException it never recovers.

- HBASE-14407: NotServingRegion:HBASE region closed forever.

- HBASE-14449: Rewrite deadlock prevention for concurrent connection close.

- HBASE-14535: Integration test for rpc connection concurrency / deadlock testing.

- HBASE-14536: Balancer & SSH interfering with each other leading to unavailability.

- HBASE-14598: ByteBufferOutputStream grows its HeapByteBuffer beyond JVM limitations.

- HBASE-14655: Narrow the scope of doAs() calls to region observer notifications for compaction.

- HBASE-14667: HBASEFsck constructors have diverged.

- HBASE-14680: Two configs for snapshot timeout and better defaults.

- HBASE-14712: Increase MasterProcWALs clean up granularity.

- HBASE-14805: Status should show the master in shell.

- HBASE-14818: user_permission does not list namespace permissions.

- HBASE-14843: TestWALProcedureStore.testLoad is flakey.

- HBASE-14850: C++ client implementation (TECH PREVIEW. DO NOT DEPLOY IN PROD).

- HBASE-14906: Improvements on FlushLargeStoresPolicy.

- HBASE-14947: WALProcedureStore improvements.

- HBASE-14963: Remove Guava dependency fromHBASE client code.

- HBASE-15019: Replication stuck when HDFS is restarted.

- HBASE-15100: Master WALProcs still never clean up.

- HBASE-15125: HBASEFsck's adoptHdfsOrphan function creates region with wrong end key boundary.

- HBASE-15142: Procedure v2 - Basic WebUI listing the procedures.

- HBASE-15144: Procedure v2 - Web UI displaying Store state.

- 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-15292: Refined ZooKeeperWatcher to prevent ZooKeeper's callback while construction.

- HBASE-15377: Per-RS Get metric is time based, per-region metric is size-based.

- HBASE-15378: Scanner cannot handle heartbeat message with no results.

- HBASE-15465: userPermission returned by getUserPermission() for the selected namespace does not have namespace set.

- HBASE-15518: Add per-table metrics.

- HBASE-15519: Add per-user metrics.

- HBASE-15636: hard coded wait time out value inHBASETestingUtility#waitUntilAllRegionsAssigned might cause test failure.

- HBASE-15752: ClassNotFoundException is encountered when custom WAL codec is not found in WALPlayer job.

- HBASE-15767: Upgrade httpclient to 4.3.6.

- HBASE-15808: Reduce potential bulk load intermediate space usage and waste.

- HBASE-15817: Backup history should mention the type (full or incremental) of the backup.

- HBASE-15842: SnapshotInfo should display ownership information.

- HBASE-15850: Localize the configuration change in testCheckTableLocks to reduce flakiness of TestHBASEFsck test suite.

- HBASE-15858: Some region server group shell commands don't work.

- HBASE-15861: Add support for table sets in restore operation.

- HBASE-15862: Backup - Delete- Restore does not restore deleted data.

- HBASE-15873: ACL for snapshot restore / clone is not enforced.

- HBASE-15880: RpcClientImpl#tracedWriteRequest incorrectly closes HTrace span.

- HBASE-15884: NPE in StoreFileScanner#skipKVsNewerThanReadpoint during reverse scan.

- HBASE-15899: HBASE incremental restore should handle namespaces properly.

- HBASE-15904: Use comma as separator for list of tables in BackupInfo.

- HBASE-15925: compat-module maven variable not evaluated.

- HBASE-15928: HBASE backup delete command does not remove backup root dir from hdfs.

- HBASE-15931: Add log for long-running tasks in AsyncProcess.

- HBASE-15933: NullPointerException may be thrown from SimpleRegionNormalizer#getRegionSize().

- HBASE-15953: HBASE backup set list command does not list any existing sets.

- HBASE-15971: Regression: Random Read/WorkloadC slower in 1.x than 0.98 (Stack).

- HBASE-15972: HBASE backup set command should not accept non-existing table.

- HBASE-16006: FileSystem should be obtained from specified path in WALInputFormat#getSplits().

- HBASE-16007: Job's Configuration should be passed to TableMapReduceUtil#addDependencyJars() in WALPlayer.

- HBASE-16012: Major compaction can't work due to obsolete scanner read point in RegionServer.

- HBASE-16016: Fix overflow deadline value in AssignmentManager#waitForAssignment.

- HBASE-16037: Backup/Restore: Make automatic mode default one.

- HBASE-16045: Endtime argument for VerifyReplication was incorrectly specified in usage.

- HBASE-16048: Tag InternalScanner with LimitedPrivate(HBASEInterfaceAudience.COPROC).

- HBASE-16052: ImproveHBASEFsck Scalability.

- HBASE-16056: Procedure v2 - fix master crash for FileNotFound.

- HBASE-16059: Region normalizer fails to trigger merge action where one of the regions is empty.

- HBASE-16062: Improper error handling in WAL Reader/Writer creation.

- HBASE-16064: Delete backup command shows HDFS permission error when deleting the intended backup.

- HBASE-16065: HBASE backup set describe command does not inform if the set does not exist.

- HBASE-16088: HBASE restore gives a misleading message if the table doesn't exist.

- HBASE-16132: Scan does not return all the result when regionserver is busy.

- HBASE-16267: Remove commons-httpclient dependency from HBASE-rest module.

- HBASE-16288: HFile intermediate block level indexes might recurse forever creating multi TB files.

- HBASE-16314: Retry on table snapshot failure.

- HBASE-16315: RegionSizeCalculator prints region names as binary without escapes.

- HBASE-16319: Fix TestCacheOnWrite after HBASE-16288.

- HBASE-16350: Undo server abort from HBASE-14968.

- HBASE-16358: Backup create command log success/failure with backup id.

- HBASE-16359: NullPointerException in RSRpcServices.openRegion().

- HBASE-16367: Race between master and region server initialization may lead to premature server abort.

- HBASE-16377: ServerName check is ineffective in region_mover.rb.

1.7.8. Hive

HDP 2.5.0 provides Hive 1.2.1 as part of the General Availablility release. In addtion, HDP 2.5.0 provides Hive 2.1.0 as a Technical Preview.

The following Apache patches are included for Hive 1.2.1:

- HIVE-4924: JDBC Support query timeout for jdbc.

- HIVE-6113: Upgrade Datanucleus to 4.x.

- HIVE-6535: JDBC provide an async API to execute query and fetch results.

- HIVE-7193: Hive should support additional LDAP authentication parameters.

- HIVE-9365: The Metastore should take port configuration from HIVE-site.xml.

- HIVE-9605: Remove parquet nested objects from wrapper writable objects.

- HIVE-9862: Vectorized execution corrupts timestamp values.

- HIVE-10233: Hive on tez memory manager for grace hash join.

- HIVE-10249: ACID show locks should show who the lock is waiting for.

- HIVE-10485: Create md5 UDF.

- 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-10639: Create SHA1 UDF.

- HIVE-10641: Create CRC32 UDF.

- HIVE-10644: Create SHA2 UDF.

- HIVE-10729: Query failed when select complex columns from joined table (tez map join only).

- HIVE-10761: Create codahale-based metrics system for Hive.

- HIVE-10815: Let HiveMetaStoreClient Choose MetaStore Randomly.

- HIVE-10927: Add number of HMS/HS2 connection metrics.

- HIVE-10944: Fix HS2 for Metrics.

- HIVE-10975: Bump the parquet version up to 1.8.1.

- HIVE-11037: HiveOnTez make explain user level = true as default.

- HIVE-11043: ORC split strategies should adapt based on number of files.

- HIVE-11096: Bump the parquet version to 1.7.0.

- HIVE-11097: MR mode query fails if one table path string starts with another's.

- HIVE-11118: Load data query should validate file formats with destination tables.

- HIVE-11164: WebHCat should log contents of HiveConf on startup.

- HIVE-11388: Allow ACID Compactor components to run in multiple metastores.

- HIVE-11401: Query on Partitioned-parquet table with where clause on a partitioned column fails as Column was not found in schema.

- HIVE-11427: Location of temporary table for CREATE TABLE SELECT broken by HIVE-7079.

- HIVE-11498: HIVE Authorization v2 should not check permission for dummy entity.

- HIVE-11512: Hive LDAP Authenticator should also support full DN in Authenticate().

- HIVE-11550: ACID queries pollute HiveConf.

- HIVE-11582: Remove conf variable hive.mapred.supports.subdirectories.

- HIVE-11593: Add aes_encrypt and aes_decrypt UDFs.

- HIVE-11695: If user have no permission to create LOCAL DIRECTORY, the HQL does not throw any exception and fail silently.

- HIVE-11716: Reading ACID table from non-acid session should raise an error.

- HIVE-11768: java.io.DeleteOnExitHook leaks memory on long running Hive Server2 Instances.

- HIVE-11793: SHOW LOCKS with DbTxnManager ignores filter options.

- HIVE-11802: Float-point numbers are displayed with different precision in Beeline/JDBC.

- HIVE-11815: Correct the column/table names in subquery expression when creating a view.

- HIVE-11824: Insert to local directory causes staging directory to be copied.

- HIVE-11832: HIVE-11802 breaks compilation in JDK 8.

- HIVE-11848: Tables in subqueries don't get locked.

- HIVE-11866: Add framework to enable testing using LDAPServer using LDAP protocol.

- HIVE-11891: Add basic performance logging to metastore calls.

- HIVE-11903: Add lock metrics to HS2.

- HIVE-11945: ORC with non-local reads may not be reusing connection to DN.

- HIVE-11956: HOW LOCKS should indicate what acquired the lock.

- HIVE-11984: Add HS2 open operation metrics.

- HIVE-12007: Hive LDAP Authenticator should allow just Domain without baseDN (for AD).

- HIVE-12008: Hive queries failing when using count(*) on column in view.

- HIVE-12213: Investigating the test failure TestHCatClient.testTableSchemaPropagation.

- HIVE-12224: Remove HOLD_DDLTIME.

- HIVE-12239: Constants in hive.common.metrics.common.MetricsConstant are not final.

- HIVE-12271: Add metrics around HS2 query execution and job submission for Hive.

- HIVE-12279: Testcase to verify session temporary files are removed after HIVE-11768.

- HIVE-12353: When Compactor fails it calls CompactionTxnHandler.markedCleaned(). it should not.

- HIVE-12366: Refactor Heartbeater logic for transaction.

- HIVE-12435: SELECT COUNT(CASE WHEN.) GROUPBY returns 1 for 'NULL' in a case of ORC and vectorization is enabled.

- HIVE-12439: CompactionTxnHandler.markCleaned() and TxnHandler.openTxns() misc improvements.

- HIVE-12499: Add HMS metrics for number of tables and partitions.

- HIVE-12505: Insert overwrite in same encrypted zone silently fails to remove some existing files.

- HIVE-12525: Cleanup unused metrics in HMS.

- HIVE-12584: Vectorized join with partition column of type char does not trim spaces.

- HIVE-12619: Switching the field order within an array of structures causes the query to fail.

- HIVE-12620: Misc improvement to Acid module.

- HIVE-12628: Eliminate flakiness in TestMetrics.

- HIVE-12634: Add command to kill an ACID transaction.

- HIVE-12637: Make retryable SQLExceptions in TxnHandler configurable.

- HIVE-12724: CID Major compaction fails to include the original bucket files into MR job.

- HIVE-12733: UX improvements for HIVE-12499.

- HIVE-12741: HS2 ShutdownHookManager holds extra of Driver instance.

- HIVE-12751: Fix NVL explain syntax.

- HIVE-12827: Vectorization VectorCopyRow/VectorAssignRow/VectorDeserializeRow assign needs explicit isNull[offset] modification.

- HIVE-12837: Better memory estimation/allocation for hybrid grace hash join during hash table loading.

- HIVE-12868: Fix empty operation-pool metrics.

- HIVE-12875: SUMMARY-[Verify sem.getInputs() and sem.getOutputs()].

- HIVE-12885: LDAP Authenticator improvements.

- HIVE-12887: Handle ORC schema on read with fewer columns than file schema (after Schema Evolution changes).

- HIVE-12893: Hive optimize.sort.dynamic.partition not enabled in case of combination of bucketing and constant propagation if subset of partition column value set.

- HIVE-12894: Detect whether ORC is reading from ACID table correctly for Schema Evolution.

- HIVE-12897: Improve dynamic partition loading.

- HIVE-12907: Improve dynamic partition loading - II.

- HIVE-12908: Improve dynamic partition loading III.

- HIVE-12937: DbNotificationListener unable to clean up old notification events.

- HIVE-12965: Insert overwrite local directory should preserve the overwritten directory permission.

- HIVE-12988: Improve dynamic partition loading IV.

- HIVE-12992: Hive on tez Bucket map join plan is incorrect.

- HIVE-12996: Temp tables shouldn't be locked.

- HIVE-13008: WebHcat DDL commands in secure mode NPE when default FileSystem doesn't support delegation tokens.

- HIVE-13013: Further Improve concurrency in TxnHandler.

- HIVE-13017: Child process of HiveServer2 fails to get delegation token from non default FileSystem.

- HIVE-13018: "RuntimeException: Vectorization is not supported for datatype:LIST".

- HIVE-13033: SPDO unnecessarily duplicates columns in key & value of mapper output.

- HIVE-13040: Handle empty bucket creations more efficiently.

- HIVE-13043: Reload function has no impact to function registry.

- HIVE-13051: Deadline class has numerous issues.

- HIVE-13056: Delegation tokens do not work with HS2 when used with http transport and Kerberos.

- HIVE-13093: Hive metastore does not exit on start failure.

- HIVE-13095: Support view column authorization.

- HIVE-13101: NullPointerException in HiveLexer.g.

- HIVE-13108: Operators SORT BY randomness is not safe with network partitions.

- HIVE-13120: Propagate doAs when generating ORC splits.

- HIVE-13125: Support masking and filtering of rows/columns.

- HIVE-13126: Clean up MapJoinOperator properly to avoid object cache reuse with unintentional states.

- HIVE-13169: HiveServer2 Support delegation token based connection when using http transport.

- HIVE-13175: Disallow making external tables transactional.

- HIVE-13178: Enhance ORC Schema Evolution to handle more standard data type conversions.

- HIVE-13198: Authorization issues with cascading views.

- HIVE-13200: Aggregation functions returning empty rows on partitioned columns.

- HIVE-13201: Compaction shouldn't be allowed on non-ACID table.

- HIVE-13209: metastore get_delegation_token fails with null ip address.

- HIVE-13213: Make DbLockManger work for non-acid resources.

- HIVE-13216: ORC Reader will leave file open until GC when opening a malformed ORC file.

- HIVE-13240: Drop the hash aggregates when closing operator .

- HIVE-13242: DISTINCT keyword is dropped by the parser for windowing.

- HIVE-13249: Hard upper bound on number of open transactions.

- HIVE-13261: Can not compute column stats for partition when schema evolves.

- HIVE-13264: JDBC driver makes 2 Open Session Calls for every open session.

- HIVE-13267: Vectorization Add SelectLikeStringColScalar for non-filter operations.

- HIVE-13287: Add logic to estimate stats for IN operator.

- HIVE-13291: ORC BI Split strategy should consider block size instead of file size.

- HIVE-13294: AvroSerde leaks the connection in a case when reading schema from a URL.

- HIVE-13295: Improvement to LDAP search queries in HS2 LDAP Authenticator.

- HIVE-13296: Add vectorized Q test with complex types showing count(*) etc work correctly.

- HIVE-13299: Column Names trimmed of leading and trailing spaces.

- HIVE-13302: Direct SQL cast to date doesn't work on Oracle.

- HIVE-13313: TABLESAMPLE ROWS feature broken for vectorization.

- HIVE-13318: Cache the result of getTable from metastore.

- HIVE-13326: HiveServer2 Make ZK config publishing configurable.

- HIVE-13330: Vectorization returns NULL for empty values for varchar/string data type.

- HIVE-13338: Differences in vectorized_casts.q output for vectorized and non-vectorized runs.

- HIVE-13344: Port HIVE-12902 to 1.x line.

- HIVE-13354: Add ability to specify Compaction options per table and per request.

- HIVE-13358: Stats state is not captured correctly turn off stats optimizer for sampled table.

- HIVE-13360: Refactoring Hive Authorization.

- HIVE-13361: Orc concatenation should enforce the compression buffer size.

- HIVE-13362: Commit binary file required for HIVE-13361.

- HIVE-13373: Use most specific type for numerical constants.

- HIVE-13381: Timestamp & date should have precedence in type hierarchy than string group.

- HIVE-13390: HiveServer2 Add more test to ZK service discovery using MiniHS2.

- HIVE-13392: Disable speculative execution for ACID Compactor.

- HIVE-13393: Beeline Print help message for the –incremental option.

- HIVE-13394: Analyze table call fails with ArrayIndexOutOfBounds exception.

- HIVE-13394: Analyze table fails in tez on empty partitions.

- HIVE-13395: Lost Update problem in ACID.

- HIVE-13405: Fix Connection Leak in OrcRawRecordMerger.

- HIVE-13418: HiveServer2 HTTP mode should support X-Forwarded-Host header for authorization/audits.

- HIVE-13434: BaseSemanticAnalyzer.unescapeSQLString doesn't unescape \u0000 style character literals.

- HIVE-13439: JDBC: provide a way to retrieve GUID to query Yarn ATS.

- HIVE-13458: Heartbeater doesn't fail query when heartbeat fails.

- HIVE-13462: HiveResultSetMetaData.getPrecision() fails for NULL columns.

- HIVE-13463: Fix ImportSemanticAnalyzer to allow for different src/dst filesystems.

- HIVE-13476: HS2 ShutdownHookManager holds extra of Driver instance in nested compile.

- HIVE-13480: Add hadoop2 metrics reporter for Codahale metrics.

- HIVE-13486: Cast the column type for column masking.

- HIVE-13493: TransactionBatchImpl.getCurrentTxnId() ArrayIndexOutOfBounds.

- HIVE-13510: Dynamic partitioning doesn't work when remote metastore is used.

- HIVE-13541: Pass view's ColumnAccessInfo to HiveAuthorizer.

- HIVE-13553: CTE with upperCase alias throws exception.

- HIVE-13561: HiveServer2 is leaking ClassLoaders when add jar / temporary functions are used.

- HIVE-13562: Enable vector bridge for all non-vectorized udfs.

- HIVE-13563: Hive Streaming does not honor orc.compress.size and orc.stripe.size table properties.

- HIVE-13568: UDFs for use in column-masking - includes updates for review comments.

- HIVE-13570: Some queries with Union all fail when CBO is off.

- HIVE-13572: Redundant setting full file status in Hive copyFiles.

- HIVE-13592: Metastore calls map is not thread safe.

- HIVE-13596: HS2 should be able to get UDFs on demand from metastore.

- HIVE-13602: TPCH q16 return wrong result when CBO is on.

- 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-13621: Compute stats in certain cases fails with NPE.

- HIVE-13622: WriteSet tracking optimizations.

- [HIVE-13632](): Hive failing on insert empty array into parquet table.

- [HIVE-13645](): SUMMARY-[Beeline needs null-guard around hiveVars and hiveConfVars read.

- [HIVE-13646](): Make hive.optimize.sort.dynamic.partition compatible with ACID tables.

- [HIVE-13660](): Vectorizing IN expression with list of columns throws java.lang.ClassCastException ExprNodeColumnDesc cannot be cast to ExprNodeConstantDesc.

- [HIVE-13670](): Improve Beeline connect/reconnect semantics.

- [HIVE-13677](): org.apache.hive.com.esotericsoftware.kryo.KryoException java.lang.NullPointerException when folding CASE expression.

- [HIVE-13682](): EOFException with fast hashtable.

- [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-13705](): Insert into table removes existing data.

- [HIVE-13716](): Improve dynamic partition loading V.

- [HIVE-13725](): ACID Streaming API should synchronize calls when multiple threads use the same endpoint.

- [HIVE-13726](): Improve dynamic partition loading VI.

- [HIVE-13729](): FileSystem leaks in FileUtils.checkFileAccessWithImpersonation.

- [HIVE-13730](): Avoid double spilling the same partition when memory threshold is set very low.

- [HIVE-13743](): Data move codepath is broken with hive.

- [HIVE-13750](): ReduceSinkDeDuplication not working with hive.optimize.sort.dynamic.partition and ACID.

- [HIVE-13753](): Make metastore client thread safe in DbTxnManager.

- [HIVE-13767](): Wrong type inferred in Semijoin condition leads to AssertionError to.

- [HIVE-13788](): hive msck listpartitions need to make use of directSQL instead of datanucleus.

- [HIVE-13796](): Fix some tests.

- [HIVE-13799](): Optimize TableScanRule:checkBucketedTable.

- [HIVE-13809](): Hybrid Grace Hash Join memory usage estimation didn't take into account the bloom filter size.

- HIVE-13821: Delete statement launches 1 Mappers that takes forever to finish for 1 TB TPC-H database(transactional).

- HIVE-13823: Remove unnecessary log line in common join operator.

- HIVE-13831: Error pushing predicates to HBase storage handler.

- HIVE-13833: Add an initial delay when starting the heartbeat.

- HIVE-13837: current_timestamp() output format is different in some cases.

- HIVE-13840: Orc split generation is reading file footers twice.

- HIVE-13841: Orc split generation returns different strategies with cache enabled vs disabled.

- HIVE-13849: Wrong plan for hive.optimize.sort.dynamic.partition=true.

- HIVE-13853: Add X-XSRF-Header filter to HS2 HTTP mode and WebHCat].

- HIVE-13856: Fetching transaction batches during ACID streaming against Hive Metastore using Oracle DB fails.

- HIVE-13857: Insert overwrite select from some table fails throwing org.apache.hadoop.security.AccessControlException - II.

- HIVE-13859: mask() UDF not retaining day and month field values.

- HIVE-13867: Restore HiveAuthorizer interface changes.

- HIVE-13901: Hive Metastore add partitions can be slow depending on filesystems.

- HIVE-13902: [Refactor] Minimize metastore jar dependencies on task nodes.

- HIVE-13904: Ignore case when retrieving ColumnInfo from RowResolver.

- HIVE-13905: Optimize ColumnStatsTask constructColumnStatsFromPackedRows to have lesser number of getTable calls.

- HIVE-13910: Select from a table is not working if used as <dbname.tablename>.

- HIVE-13911: Load inpath fails throwing org.apache.hadoop.security.AccessControlException.

- HIVE-13912: bTxnManager.commitTxn(): ORA-00918 column ambiguously defined.

- HIVE-13922: Optimize the code path that analyzes/updates col stats.

- HIVE-13929: org.apache.hadoop.hive.metastore.api.DataOperationType class not found error when a job is submitted by hive.

- HIVE-13931: Add support for HikariCP and replace BoneCP usage with HikariCP.

- HIVE-13932: Hive SMB Map Join with small set of LIMIT failed with NPE.

- HIVE-13933: Add an option to turn off parallel file moves.

- HIVE-13941: Return better error messages from SchemaTool].

- HIVE-13948: Incorrect timezone handling in Writable results in wrong dates in queries.

- HIVE-13954: Any query on Parquet table throws SLF4J:Failed to load class org.slf4j.impl.StaticLoggerBinder - resulting unwanted WARN and INFO messages on stdout.

- HIVE-13957: Vectorized IN is inconsistent with non-vectorized (at least for decimal in (string)).

- HIVE-13961: Major compaction fails to include the original bucket files if there's no delta directory.

- HIVE-13972: Resolve class dependency issue introduced by HIVE-13354.

- HIVE-13984: Use multi-threaded approach to listing files for msck.

- HIVE-13985: ORC improvements for reducing the file system calls in task side.

- HIVE-13997: Insert overwrite directory doesn't overwrite existing files.

- HIVE-14006: Hive query with UNION ALL fails with ArrayIndexOutOfBoundsException.

- HIVE-14010: parquet-logging.properties from HIVE_CONF_DIR should be used when available.

- HIVE-14014: Zero length file is being created for empty bucket in tez mode (II).

- HIVE-14018: Make IN clause row selectivity estimation customizable.

- HIVE-14022: Left semi join should throw SemanticException if where clause contains columnname from right table.

- HIVE-14038: Miscellaneous acid improvements.

- HIVE-14045: (Vectorization) Add missing case for BINARY in VectorizationContext.getNormalizedName method.

- HIVE-14054: TestHiveMetaStoreChecker fails on master.

- HIVE-14070: hive.tez.exec.print.summary=true returns wrong performance numbers on HS2.

- HIVE-14073: Update config whiltelist for SQL std authorization.

- HIVE-14080: hive.metastore.schema.verification should check for schema compatibility.

- HIVE-14084: ORC table data load failure when metadata size is very large.

- HIVE-14114: Ensure RecordWriter in streaming API is using the same UserGroupInformation as StreamingConnection.

- HIVE-14126: With ranger enabled, partitioned columns is returned first when you execute select star.

- 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-14192: False positive error due to thrift.

The following Apache patches are included for Hive 2.1.0:

- HIVE-10815: Let HiveMetaStoreClient Choose MetaStore Randomly.

- HIVE-13391: add an option to LLAP to use keytab to authenticate to read data.

- HIVE-13443: LLAP: signing for the second state of submit (the event).

- HIVE-13617: LLAP: support non-vectorized execution in IO.

- HIVE-13675: LLAP: add HMAC signatures to LLAPIF splits.

- HIVE-13731: LLAP: return LLAP token with the splits.

- HIVE-13759: LlapTaskUmbilicalExternalClient should be closed by the record reader.

- HIVE-13771: LLAPIF: generate app ID.

- HIVE-13827: LLAPIF: authentication on the output channel.

- HIVE-13909: upgrade ACLs in LLAP registry when the cluster is upgraded to secure.

- HIVE-13930: upgrade Hive to latest Hadoop version.

- HIVE-13931: Update connection pool usage from bonecp to hikaricp.

- HIVE-13956: LLAP: external client output is writing to channel before it is writable again.

- HIVE-13986: LLAP: kill Tez AM on token errors from plugin.

- HIVE-14023: LLAP: Make the Hive query id available in ContainerRunner.

- HIVE-14045: (Vectorization) Add missing case for BINARY in VectorizationContext.getNormalizedName method.

- HIVE-14072: QueryIds reused across different queries.

- HIVE-14078: LLAP input split should get task attempt number from conf if available.

- HIVE-14080: hive.metastore.schema.verification should check for schema compatiblity.

- HIVE-14091: some errors are not propagated to LLAP external clients.

- HIVE-14093: LLAP output format connection should wait for all writes to finish before closing channel.

- HIVE-14119: LLAP external recordreader not returning non-ascii string properly.

- HIVE-14136: LLAP ZK SecretManager should resolve _HOST in principal.

- HIVE-14180: Disable LlapZooKeeperRegistry ZK auth setup for external clients.

- HIVE-14182: Revert HIVE-13084/HIVE-13924/HIVE-14034.

- HIVE-14219: LLAP external client on secure cluster: Protocol interface org.apache.hadoop.hive.llap.protocol.LlapTaskUmbilicalProtocol is not known.

- HIVE-14224: LLAP: Rename query specific log files when a query completes execution.

- HIVE-14230: Hadoop23Shims.cloneUgi() doesn't add credentials from original UGI.

- HIVE-14245: NoClassDefFoundError when starting LLAP daemon.

- HIVE-14349: Vectorization: LIKE should anchor the regexes.

- HIVE-14364: Update timeouts for llap comparator tests.

- HIVE-14392: llap daemons should try using YARN local dirs, if available.

- HIVE-14403: LLAP node specific preemption will only preempt once on a node per AM.

- HIVE-14421: FS.deleteOnExit holds references to _tmp_space.db files.

- HIVE-14436: Hive 1.2.1/Hitting "ql.Driver: FAILED: IllegalArgumentException Error: , expected at the end of 'decimal(9'" after enabling hive.optimize.skewjoin and with MR engine.

- HIVE-14439: LlapTaskScheduler should try scheduling tasks when a node is disabled.

1.7.9. Kafka

HDP 2.5.0 provides Kafka 0.10.0.1 and the following Apache patches:

- KAFKA-3258: Delete broker topic metrics of deleted topics.

- KAFKA-3393: Updated the docs to reflect the deprecation of block.on.buffer.full and usage of max.block.ms.

- KAFKA-3660: Log exception message in ControllerBrokerRequestBatch.

- KAFKA-3683: Add file descriptor recommendation to ops guide.

- KAFKA-3704: Revert "Remove hard-coded block size in KafkaProducer."

- KAFKA-3717: Support building aggregate javadoc for all project modules.

- KAFKA-3718: Propagate all KafkaConfig __consumer_offsets configs to OffsetConfig instantiation.

- KAFKA-3721: Put UpdateMetadataRequest V2 in 0.10.0-IV1.

- KAFKA-3728: EndToEndAuthorizationTest offsets_topic misconfigured.

- KAFKA-3747: Close `RecordBatch.records` when append to batch fails.

- KAFKA-3784: TimeWindows#windowsFor calculation is incorrect.

- KAFKA-3785: Fetcher spending unnecessary time during metrics recording.

- KAFKA-3787: Preserve the message timestamp in mirror maker.

- KAFKA-3789: Upgrade Snappy to fix snappy decompression errors.

1.7.10. Knox

HDP 2.5.0 provides Knox 0.9.0 and the following Apache patches:

- KNOX-705: Improvement on 404 error messages with ambari views.

- KNOX-711: Added the ability to scope rewrite rules.

- KNOX-712: submitPig does not allow multiple argument via arg().

- KNOX-713: Knox Shell HDFS.get.Request is Package Private.

- KNOX-714: Remove Permissive MIT License from NOTICE File.

- KNOX-715: submitJava should allow multiple argument via arg().

- KNOX-716: replayBufferSize is kept in bytes.

- KNOX-717: Changed the location header in RANGERUI rewrite to be more specific.

- KNOX-721: Added rewrite rules to take care of no slash url.

- KNOX-722: Changed KnoxSessionStore to not base64 empty cookie value.

- KNOX-723: Qualifying the login page resources with the front end path.

- KNOX-724: Ambari view fixes.

- KNOX-725: Defining policy chain for ambari and ranger services.

- KNOX-726: Clean up Skipped Unit Tests.

- KNOX-731: Added ranger and ambari services to topology.

1.7.11. 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 HDP-2.5.x, we removed the "commons-httpclient" library from Mahout because we view it as an obsolete library with possible security issues., and upgraded the Hadoop-Client in Mahout to version 2.7.3, the same version used in HDP-2.5. As a result:

- Previously compiled Mahout jobs will need to be recompiled in the HDP-2.5 environment.

- There is a small possibility that some Mahout jobs may encounter "ClassNotFoundException" or "could not load class" errors related to "org.apache.commons.httpclient", "net.java.dev.jets3t", or related class name prefixes. If these errors happen, you may consider whether to manually install the needed jars in your classpath for the job, if the risk of security issues in the obsolete library is acceptable in your environment.

- There is an even smaller possibility that some Mahout jobs may encounter crashes in Mahout's hbase-client code calls to the hadoop-common libraries, due to binary compatibility problems. Regrettably, there is no way to resolve this issue except revert to the HDP-2.4.2 version of Mahout, which may have security issues. Again, this should be very unusual, and is unlikely to occur in any given Mahout job suite.

1.7.12. Oozie

HDP 2.5.0 provides Oozie 4.2.0 and the following Apache patches:

- OOZIE-2289: hive-jdbc dependency in core/pom.xml should be compile.

- OOZIE-2290: Oozie db version update should happen after all DDL tweak.

- OOZIE-2291: Hive2 workflow.xml.security should have "cred" in action tag instead of "hive2" tag.

- OOZIE-2492: JSON security issue in js code.

1.7.13. Phoenix

HDP 2.5.0 provides Phoenix 4.7.0 and the following Apache patches:

- PHOENIX-1523: Make it easy to provide a tab literal as separator for CSV imports.

- PHOENIX-2276: Addednum2 to fix test failures.

- PHOENIX-2743: HivePhoenixHandler for big-big join with predicate push down.

- PHOENIX-2748: Disable auto-commit during bulk load.

- PHOENIX-2758: Ordered GROUP BY not occurring with leading PK equality expression.

- PHOENIX-2803: Initializing ToCharFunction with the List<Expression< constructor fails.

- PHOENIX-2817: PHOENIX-Spark plugin doesn't work in secured env.

- PHOENIX-2820: Investigate why SortMergeJoinIT has a sort in the explain plan.

- PHOENIX-2828: Ordinality should be 1-based in UNNEST WITH ORDINALITY.

- PHOENIX-2845: Timestamp ignores nanos in literal expressions.

- PHOENIX-2856: HBaseManagedTimeTableReuseTest category isn't getting run.

- PHOENIX-2871: Workaround sqlline issue.

- PHOENIX-2878: CURRENT_TIME fails to provide correct time when projected table is using.

- PHOENIX-2886: Union ALL with Char column not present in the table in Query 1 but in Query 2 throw exception.

- PHOENIX-2894: Sort-merge join works incorrectly with DESC columns.

- PHOENIX-2898: HTable not closed in ConnectionQueryServicesImpl.

- PHOENIX-2905: hadoop-2.5.1 artifacts are in the dependency tree.

- PHOENIX-2908: PHOENIX-core depends on both antlr 3.5 and antlr 2.7.7.

- PHOENIX-2912: Broken IT tests after PHOENIX-2905.

- PHOENIX-2919: PreparedStatement Returns Incorrect Number of Deleted Records.

- PHOENIX-2920: Incorrect Queries on Multi-tenant tables with WHERE clause containing Row Value Constructor.

- PHOENIX-2934: Checking a coerce expression at top level should not be necessary for Union All query.

- PHOENIX-2936: Missing antlr runtime on server side after PHOENIX-2908.

- PHOENIX-2940: Query the stats table and cache stats in the client.

- PHOENIX-2942: Order by incorrect for RVC.

- PHOENIX-2952: Addendum to fix test case failure.

- PHOENIX-2952: array_length return negative value.

- PHOENIX-2961: Select on a composite primary key fails when DESC is set in the table.

- PHOENIX-2983: ClassCastException on auto coerce of BIGINT to DECIMAL.

- PHOENIX-2994: Coalesce function may return null.

- PHOENIX-3004: Allow configuration in hbase-site to define realms other than the server's.

- PHOENIX-3006: Fix all ScalarFunctions to implement clone(List) or <init>(List).

- PHOENIX-3008: Prevent upgrade of existing multi-tenant table to map to namespace until we support it correctly.

- PHOENIX-3011: Fix missing apache licenses.

- PHOENIX-3013: TO_CHAR fails to handle indexed null value.

- PHOENIX-3014: SELECT DISTINCT pk ORDER BY pk DESC gives the wrong results with salted tables.

- PHOENIX-3015: Any metadata changes may cause unpredictable result when local indexes are using.

- PHOENIX-3021: Using local index during compaction is producing NPE.

- PHOENIX-3057: Set incremental=false for sqlline-thin.

- PHOENIX-3136: Do not relocate org.apache.calcite in queryserver/thin-client.

- PHOENIX-3149: Local index got corrupted if intermediate compaction happen during the split.

- PHOENIX-3152: Incorrect comparator in QueryOptimizer may cause IllegalArgumentException.

1.7.14. Pig

HDP 2.5.0 provides Pig 0.16.0 and the following Apache patches:

- PIG-4916: Pig on Tez fail to remove temporary HDFS files in some cases.

- PIG-4921: Kill running jobs on InterruptedException (partial).

- PIG-4947: LOAD with HBaseStorage using a mix of pure wildcards and prefixed wildcards results in empty maps for the pure wildcarded column families.

- PIG-4948: Pig on Tez AM use too much memory on a small cluster.

1.7.15. Ranger

HDP 2.5.0 provides Ranger 0.6.0 and the following Apache patches:

- RANGER-1090: Revoke command with grant option does not disable delegated admin permission for users/groups in the corresponding policy.

- RANGER-1094: One way SSL (when Kerberos is enabled) for Ranger and its plugins.

- RANGER-1096: Revert to jceks scheme for credential store related operations.

- RANGER-1097: Ranger KMS Plugin should not fails to download policy when UGI ticket expires.

- RANGER-1099: Keyadmin user is not able to create service/repo using public APIs.

- RANGER-1100: Hive authorizer does not block update when row-filter/column-mask is specified on the table for the user.

- RANGER-1101: JCEKS keystore is not created successfully after enabling SSL for Atlas Ranger plugin.

- RANGER-1103: Added maven version enforcer and moved the plugin to be run as part of maven compile.

- RANGER-1104: Catching and Logging DB transaction exceptions during Ranger startup.

- RANGER-1105: Ranger should provide configuration to do hdfs audit file rollover at absolute time.

- RANGER-1106: Issue after upgrade on ranger hive policy page.

- RANGER-1111: Enhancements to the db admin setup scripts.

- RANGER-1113: Ranger Hive authorizer updated to get query string from HiveConf.

- RANGER-1114: Nimbus, Storm UI server stopped after disabling ranger plugins.

- RANGER-1116: Ranger HivePluginUnitTest fails due to Hive Metastore version check.

- RANGER-1119: Exclude test jars from RANGER-admin plugin folders as dependency.

- RANGER-1120: Need a java patch to handle upgrade of hive servicedef.

- RANGER-1121: Resolving circular dependency of spring beans by enabling lazy initialization of the beans.

- RANGER-1123: Keyadmin user is not able to make getservice call using rest v2 public api.

- RANGER-1124: Good coding practices in Ranger recommended by static code analysis - UI .

- RANGER-1126: Authorization checks for non existent file/directory should not be recursive in Ranger Hive authorizer.

- RANGER-1127: Ranger HA Handle scenarios for request with X-Forwarded-Server.

- RANGER-1128: Data Masking label changes for ranger policies.

- RANGER-1129: Ability to specify 'audit all accesses' via Ranger admin configuration.

- RANGER-1132: Ranger Storm Plugin should include commons-codec jar as a dependency.

- RANGER-1134: Audit to Secure solr fails in case of Ranger Knox Plugin due to MDC context issue.

- RANGER-1135: Knox and Storm plugins should use secure policy download endpoint in kerberos mode.

- RANGER-1135: Modified InMemory JAAS configuration to use parent config - if exists.

- RANGER-1136: Ranger audit to HDFS fails with TGT errors in Ranger HiveServer2 plugin when UGI -TGT expires in audit thread.

- RANGER-1141: Null pointer exception while retrieving the key during copy file.

- RANGER-1143: Added RANGER-plugins-cred lib for tagsync deployment.

1.7.16. Slider

HDP 2.5.0 provides Slider 0.91.0 and the following Apache patches:

- SLIDER-942: Slider agent fails with SSL validation errors with python 2.7.9.

- SLIDER-1129: Completely remove dependency on commons-httpclient library from test code as well due to HADOOP-13382

- SLIDER-1145: Work around hadoop 2.8 changes in DefaultResourceCalculator.normalize.

- SLIDER-1150: Slider app fails to start on AmazonLinux.

- SLIDER-1152: Resource leaks found in code.

- SLIDER-1153: Code issues - 14 null pointer deferences found.

- SLIDER-1154: Fix concurrent data access violations.

- SLIDER-1156: OutstandingRequestTracker edits map while iterating over its entries.

- SLIDER-1157: Code issues - integer handling issues.

1.7.17. Spark

HDP 2.5.0 provides Spark 1.6.2 and the following Apache patches:

- SPARK-16214: Fix the denominator of SparkP.

- SPARK-6005: Flaky test o.a.s.streaming.kafka.DirectKafkaStreamSuite.offset recover.

- SPARK-6717: Clear shuffle files after checkpointing in AL.

- SPARK-6735: Add window based executor failure tracking mechanism for long running service.

- SPARK-6847: Stack overflow on updateStateByKey which followed by a stream with checkpoint set.

- SPARK-7481: Add SPARK-cloud module to pull in AWS+azure object store FS accessors; test integration.

- SPARK-7889: Jobs progress of apps on complete page of HistoryServer shows incomplete.

- SPARK-10582: Using dynamic-executor-allocation, if AM failed, the new AM will be started. But the new AM does not allocate executors to drive.

- SPARK-11137: Make StreamingContext.stop() exception-saf.

- SPARK-11182: HDFS Delegation Token will be expired when calling "UserGroupInformation.getCurrentUser.addCredentials" in HA mod.

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

- SPARK-11315: Add YARN extension service to publish Spark events to YARN timeline service (part of SPARK-1537.

- SPARK-11323: Add History Service Provider to service application histories from YARN timeline server (part of SPARK-1537.)

- SPARK-11627: Spark Streaming backpressure mechanism has no initial rate limit, receivers receive data at the maximum speed , it might cause OOM exception.

- 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 executor.

- SPARK-12142: Can't request executor when container allocator us bit read.

- 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-12948: OrcRelation uses HadoopRDD which can broadcast conf objects frequently.

- SPARK-12967: NettyRPC races with SparkContext.stop() and throws exception.

- SPARK-12998: Enable OrcRelation even 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.

- SPARK-13278: Launcher fails to start with JDK 9 E.

- SPARK-13308: ManagedBuffers passed to OneToOneStreamManager need to be freed in non error case.

- SPARK-13360: pyspark related environment variable is not propagated to driver in yarn-cluster mod.

- 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 token.

- SPARK-13885: Fix attempt id regression for Spark running on Yarn.

- SPARK-13902: Make DAGScheduler not to create duplicate stage.

- SPARK-14062: Fix log4j and upload metrics.properties automatically with distributed cache.

- SPARK-14091: Consider improving performance of SparkContext.getCallSite().

- SPARK-15067: YARN executors are launched with fixed perm gen size.

- SPARK-1537: Add integration with Yarn's Application Timeline Serve.

- SPARK-15606: Driver hang in o.a.s.DistributedSuite on 2 core machine.

- SPARK-15844: HistoryServer doesn't come up if spark.authenticate = true.

- SPARK-15990: Add rolling log aggregation support for Spark on yarn.

- SPARK-15990: Add rolling log aggregation support for Spark on yarn.

- SPARK-16077: Python UDF may fail because of six.

- SPARK-16110: Can't set Python via spark-submit for YARN cluster mode when PYSPARK_PYTHON & PYSPARK_DRIVER_PYTHON are set (JIRA in Apache Spark is IN PROGRESS).

- SPARK-16193: Address flaky ExternalAppendOnlyMapSuite spilling test.

1.7.18. Sqoop

HDP 2.5.0 provides Sqoop 1.4.6 and the following Apache patches:

- SQOOP-1094: Add Avro support to merge tool.

- SQOOP-1281: Support of glob paths during export.

- SQOOP-1369: Avro export ignores –columns option.

- SQOOP-1493: Add ability to import/export true decimal in Avro instead of serializing it to String.

- SQOOP-2103: Not able define Decimal(n,p) data type in map-column-hive option.

- SQOOP-2264: Exclude and remove SqoopUserGuide.xml from git repository.

- SQOOP-2283: Support usage of –exec and –password-alias.

- SQOOP-2286: Ensure Sqoop generates valid avro column names.

- SQOOP-2295: Hive import with Parquet should append automatically.

- SQOOP-2296: Merge statement incorrectly includes table hints.

- SQOOP-2297: Explicitly add zookeeper as a dependency in ivy.xml.

- SQOOP-2298: TestAvroImport test case error.

- SQOOP-2326: Fix Netezza trunc-string option handling and unnecessary log directory during imports.

- SQOOP-2333: Sqoop to support Custom options for User Defined Plugins(Tool).

- SQOOP-2334: Sqoop Volume Per Mapper.

- SQOOP-2339: Move sub-directory might fail in append mode.

- SQOOP-2343: AsyncSqlRecordWriter stuck if any exception is thrown out in its close method.

- SQOOP-2362: Add oracle direct mode in list of supported databases.

- SQOOP-2363: wrong option for escape character with mysqlimport.

- SQOOP-2367: Tests from PostgresqlImportTest can fail on various versions of Hadoop.

- SQOOP-2370: Netezza - need to support additional options for full control character handling.

- SQOOP-2371: Tests from *LobAvroImportTest might fail.

- SQOOP-2372: Imports all tables as parquet will meet a NPE.

- SQOOP-2381: Add test for mysql export with –escape-by option.

- SQOOP-2387: Sqoop should support importing from table with column names containing some special character.

- SQOOP-2399: Sqoop import into Hbase - BigDecimalSplitter java.lang.ArrayIndexOutOfBoundsException.

- SQOOP-2400: hive.metastore.sasl.enabled should be set to true for Oozie integration.

- SQOOP-2406: Add support for secure mode when importing Parquet files into Hive.

- SQOOP-2437: Use hive configuration to connect to secure metastore.

- SQOOP-2454: Drop JDK6 support.

- SQOOP-2457: Add option to automatically compute statistics after loading date into a hive table.

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

- SQOOP-2531: readlink -f not supported on OS X.

- SQOOP-2535: Add error handling to HiveConf.

- SQOOP-2561: Special Character removal from Column name as avro data results in duplicate column and fails the import.

- SQOOP-2582: Query import won't work for parquet.

- SQOOP-2597: Missing method AvroSchemaGenerator.generate().

- SQOOP-2607: Direct import from Netezza and encoding.

- SQOOP-2609: Provide Apache Atlas integration for hive and hcatalog based imports.

- SQOOP-2617: Log file that is being used for LOB files.

- SQOOP-2642: Document ability to specify commas in –map-column-hive option.

- SQOOP-2647: Add option for drop-if-exists when using sqoop hcat import.

- SQOOP-2651: Do not dump data on error in TextExportMapper by default.

- SQOOP-2707: Upgrade commons-collections to 3.2.2.

- SQOOP-2712: Run only one map task attempt during export (second edition).

- SQOOP-2723: Oracle connector not working with lowercase columns.

- SQOOP-2737: Cannot import table from Oracle with column with spaces in name.

- SQOOP-2745: Using datetime column as a splitter for Oracle no longer works.

- SQOOP-2746: Add test case for Oracle incremental import using Timestamp.

- SQOOP-2747: Allow customizing test username and password for Oracle tests.

- SQOOP-2753: TestSqoopJsonUtil.testGetJsonStringFromMap is depending on Map ordering in JDK.

- SQOOP-2767: Test is failing SystemImportTest.

- SQOOP-2779: Sqoop metastore doesn't seem to recognize –schema option.

- SQOOP-2780: Sqoop 1 unit tests fail with TestIncrementalImport test.

- SQOOP-2783: Query import with parquet fails on incompatible schema.

- SQOOP-2787: MySql import and export fails with 5.1 server and 5.1.17+ drivers.

- SQOOP-2795: Clean up useless code in class TestSqoopJsonUtil.

- SQOOP-2810: Upgrade to non-snapshot dependency on Avro 1.8.0 as soon as it gets released.

- SQOOP-2839: Sqoop import failure due to data member conflict in ORM code for table.

- SQOOP-2846: Sqoop Export with update-key failing for avro data file.

- SQOOP-2847: Sqoop –incremental + missing parent –target-dir reports success with no data.

- SQOOP-2863: Properly escape column names for generated INSERT statements.

- SQOOP-2864: ClassWriter chokes on column names containing double quotes.

- SQOOP-2880: Provide argument for overriding temporary directory.

- SQOOP-2896: Sqoop exec job fails with SQLException Access denied for user.

- SQOOP-2906: Optimization of AvroUtil.toAvroIdentifier.

- SQOOP-2909: Oracle related ImportTest fails after SQOOP-2737.

- SQOOP-2910: Add capability to Sqoop to require an explicit option to be specified with – split-by for a String column.

- SQOOP-2911: Fix failing HCatalogExportTest caused by SQOOP-2863.

- SQOOP-2915: Fixing Oracle related unit tests.

- SQOOP-2920: Sqoop performance deteriorates significantly on wide datasets; sqoop 100% on cpu.

- SQOOP-2939: Extend mainframe module to support GDG, sequential data sets, and data sets stored on tape.

- SQOOP-2971: OraOop does not close connections properly.

1.7.19. Storm

HDP 2.5.0 provides Storm 1.0.1 and the following Apache patches:

- STORM-1136: Command line module to return Kafka spout offsets lag and display in storm UI.

- STORM-1575: Fix TwitterSampleSpout NPE on close.

- STORM-1674: Idle KafkaSpout consumes more bandwidth than needed.

- STORM-1694: Kafka Spout Trident Implementation Using New Kafka Consumer API.

- STORM-1698: Asynchronous MetricsConsumerBolt.

- STORM-1700: Introduce 'whitelist' / 'blacklist' option to MetricsConsumer.

- STORM-1705: Cap number of retries for a failed message.

- STORM-1709: Added group by support in storm SQL standalone mode.

- STORM-1719: Introduce REST API: Topology metric stats for stream.

- STORM-1720: Support GEO in storm-redis.

- STORM-1723: Introduce ClusterMetricsConsumer.

- STORM-1728: TransactionalTridentKafkaSpout error.

- STORM-1730: LocalCluster#shutdown() does not terminate all storm threads/thread pools.

- STORM-1742: More accurate 'complete latency'.

- STORM-1771: HiveState should flushAndClose before closing old or idle Hive connections.

- STORM-1833: Simple equi-join in storm-sql standalone mode.

- STORM-1839: Kinesis Spout.

- STORM-1841: Address a few minor issues in windowing and doc.

- STORM-1842: Forward references in storm.thrift cause tooling issues.

- STORM-1848: Make KafkaMessageId and Partition serializable to support.

- STORM-1849: HDFSFileTopology should use the 3rd argument as topologyName.

- STORM-1850: State Checkpointing Documentation update.

- STORM-1851: Fix default nimbus impersonation authorizer config.

- STORM-1859: Late tuples in windowed mode.

- STORM-1862: Flux ShellSpout and ShellBolt can't emit to named streams.

- STORM-1864: StormSubmitter should throw respective exceptions and log respective errors for registered submitter hook invocation.

- STORM-1865: Update command line client document.

- STORM-1866: Update Resource Aware Scheduler Documentation.

- STORM-1868: Modify TridentKafkaWordCount to run in distributed mode.

- STORM-1873: Implement alternative behaviour for late tuples.

- STORM-1874: Update logger private permissions.

- STORM-1878: Flux can now handle IStatefulBolts.

- STORM-1882: Expose TextFileReader public.

- STORM-1884: Prioritize pendingPrepare over pendingCommit.

- STORM-1887: Fixed help message for set_log_level command.

- STORM-1888: Add description for shell command.

- STORM-1893: Support OpenTSDB for storing timeseries data.

- STORM-1902: Add a simple & flexible FileNameFormat for storm-hdfs.

- STORM-1906: Window count/length of zero should be disallowed.

- STORM-1907: PartitionedTridentSpoutExecutor has incompatible types that cause ClassCastException.

- STORM-1909: Update HDFS spout documentation.

- STORM-1911: IClusterMetricsConsumer should use seconds to timestamp unit.

- STORM-1914: Storm Kafka Field Topic Selector.

- STORM-1919: Introduce FilterBolt on storm-redis.

- STORM-1924: Adding conf options for Persistent Word Count Topology.

- STORM-1925: Remove Nimbus thrift call from Nimbus itself.

- STORM-1930: Kafka New Client API - Support for Topic Wildcards.

- STORM-1934: Fix race condition between sync-supervisor and sync-processes.

- STORM-1945: Fix NPE bugs on topology spout lag for storm-kafka-monitor.

- STORM-1950: Change response json of "Topology Lag" REST API to keyed by spoutId, topic, partition.

- STORM-1956: Disabling Backpressure by default.

- STORM-1959: Add missing license header to KafkaPartitionOffsetLag.

- STORM-1960: Add CORS support to STORM UI REST API.

1.7.20. Tez

HDP 2.5.0 provides Tez 0.7.0 and the following Apache patches:

- TEZ-814: Improve heuristic for determining a task has failed output.

- TEZ-1248: Reduce slow-start should special case 1 reducer runs.

- TEZ-1314: Port MAPREDUCE-5821 to Tez.

- TEZ-1529: ATS and TezClient integration in secure kerberos enabled cluster.

- TEZ-1911: MergeManager's unconditionalReserve() should check for memory limits before allocating.

- TEZ-1961: Remove misleading exception "No running dag" from AM log.

- TEZ-2076: Tez framework to extract/analyze data stored in ATS for specific dag.

- TEZ-2097: TEZ-UI Add dag logs backend support.

- TEZ-2198: Fix sorter spill count.

- TEZ-2211: Tez UI: Allow users to configure timezone.

- TEZ-2291: TEZ UI: Improper vertex name in tables.

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

- TEZ-2311: AM can hang if kill received while recovering from previous attempt.

- TEZ-2391: TestVertexImpl timing out at times on Jenkins builds.

- TEZ-2398: Flaky test: TestFaultTolerance.

- TEZ-2409: Allow different edges to have different routing plugin.

- TEZ-2436: Tez UI: Add cancel button in column selector.

- TEZ-2440: Sorter should check for indexCacheList.size() in flush(.

- TEZ-2447: Tez UI: Generic changes based on feedbacks.

- TEZ-2453: Tez UI: show the dagInfo is the application has set the same.

- TEZ-2455: Tez UI: Dag view caching, error handling and minor layout change.

- TEZ-2460: Temporary solution for issue due to YARN-256.

- TEZ-2461: tez-history-parser compile fails.

- TEZ-2468: Change the minimum Java version to Java 7.

- TEZ-2474: The old taskNum is logged incorrectly when parallelism is changed.

- TEZ-2475: Fix a potential hang in Tez local mode caused by incorrectly handled interrupts.

- TEZ-2478: Move OneToOne routing to store events in Tasks.

- TEZ-2481: Tez UI: graphical view does not render properly on IE1.

- TEZ-2482: Tez UI: Mouse events not working on IE1.

- TEZ-2483: TEZ-2843 Tez UI: Show error if in progress fails due to AM not reachable.

- TEZ-2489: Disable warn log for Timeline ACL error when tez.allow.disabled.timeline-domains set to true.

- TEZ-2504: Tez UI: tables - show status column without scrolling, numeric 0 shown as Not available.

- TEZ-2505: PipelinedSorter uses Comparator objects concurrently from multiple threads.

- TEZ-2509: YarnTaskSchedulerService should not try to allocate containers if AM is shutting down.

- TEZ-2513: Tez UI: Allow filtering by DAG ID on All dags table.

- TEZ-2523: Tez UI: derive applicationId from dag/vertex id instead of relying on json date.

- TEZ-2527: Tez UI: Application hangs on entering erroneous RegEx in counter table search bot.

- TEZ-2528: Tez UI: Column selector buttons gets clipped, and table scroll bar not visible in mac.

- TEZ-2535: Tez UI: Failed task attempts link in vertex details page is broken.

- TEZ-2538: ADDITIONAL_SPILL_COUNT wrongly populated for DefaultSorter with multiple partition.

- TEZ-2539: Tez UI: Pages are not updating in IE.

- TEZ-2541: DAGClientImpl enable TimelineClient check is wrong.

- TEZ-2545: It is not necessary to start the vertex group commit when DAG is in TERMINATING

- TEZ-2546: Tez UI: Fetch hive query text from timeline if dagInfo is not set.

- TEZ-2547: Tez UI: Download Data fails on secure, cross-origin cluster.

- TEZ-2548: TezClient submitDAG can hang if the AM is in the process of shutting down.

- TEZ-2549: Reduce Counter Load on the Timeline Server.

- TEZ-2552: CRC errors can cause job to run for very long time in large jobs.

- TEZ-2553: Tez UI: Tez UI Nits.

- TEZ-2554: Tez UI: View log link does not correctly propagate login credential to read log from yarn web.

- TEZ-2560: Fix Tez-ui build for maven 3.3+.

- TEZ-2561: Port for TaskAttemptListenerImpTezDag should be configurable.

- TEZ-2567: Tez UI: download dag data does not work within ambari.

- TEZ-2568: V_INPUT_DATA_INFORMATION may happen after vertex is initialized.

- TEZ-2575: Handle KeyValue pairs size which do not fit in a single block in PipelinedSorte.

- TEZ-2579: Incorrect comparison of TaskAttemptId.

- TEZ-2602: Throwing EOFException when launching MR job.

- TEZ-2629: LimitExceededException in Tez client when DAG has exceeds the default max counters.

- TEZ-2635: Limit number of attempts being downloaded in unordered fetch.

- TEZ-2636: MRInput and MultiMRInput should work for cases when there are 0 physical inputs.

- TEZ-2660: Tez UI: need to show application page even if system metrics publish is disabled.

- TEZ-2662: Provide a way to check whether AM or task opts are valid and error if not.

- TEZ-2663: SessionNotRunning exceptions are wrapped in a ServiceException from a dying AM.

- TEZ-2687: Tez should release/kill all held containers before stopping services during the shutdown phase.

- TEZ-2719: Consider reducing logs in unordered fetcher with shared-fetch option.

- TEZ-2730: tez-api missing dependency on org.codehaus.jettison for json.

- TEZ-2732: DefaultSorter throws ArrayIndex exceptions on 2047 Mb size sort buffer.

- TEZ-2734: Add a test to verify the filename generated by OnDiskMerge.

- TEZ-2742: VertexImpl.finished() terminationCause hides member var of the same name.

- TEZ-2745: ClassNotFound in InputInitializer causes AM to crash.

- TEZ-2752: logUnsuccessful completion in Attempt should write original finish time to ATS.

- TEZ-2754: Tez UI: StartTime & EndTime is not displayed with right format in Graphical View.

- TEZ-2756: MergeManager close should not try merging files on close if invoked after a shuffle exception.

- TEZ-2758: Remove append API in RecoveryService after TEZ-190.

- TEZ-2761: Addendum fix build failure for java.

- TEZ-2761: Tez UI: update the progress on the dag and vertices pages with info from AM.

- TEZ-2766: Tez UI: Add vertex in-progress info in DAG detail.

- TEZ-2767: Make TezMxBeanResourceCalculator the default resource calculator.

- TEZ-2768: Log a useful error message when the summary stream cannot be closed when shutting down an AM.

- TEZ-2780: Tez UI: Update All Tasks page while in progress.

- TEZ-2781: Fallback to send only TaskAttemptFailedEvent if taskFailed heartbeat fails.

- TEZ-2787: Tez AM should have java.io.tmpdir=./tmp to be consistent with tasks.

- TEZ-2789: Backport events added in TEZ-2612.

- TEZ-2792: Add AM web service API for task.

- TEZ-2792: Addendum fix build failure for java.

- TEZ-2807: Log data in the finish event instead of the start event.

- TEZ-2808: Race condition between preemption and container assignment.

- TEZ-2812: Tez UI: Update task & attempt tables while in progress.

- TEZ-2813: Tez UI: add counter data for rest api calls to AM Web Services v2.

- TEZ-2816: Preemption sometimes does not respect heartbeats between preemption.

- TEZ-2817: Tez UI: update in progress counter data for the dag vertices and tasks table.

- TEZ-2825: Report progress in terms of completed tasks to reduce load on AM for Tez U.

- TEZ-2829: Tez UI: minor fixes to in-progress update of UI from AM.

- TEZ-2830: Backport TEZ-2774. Improvements to logging in the AM and part of the runtime.

- TEZ-2834: Make Tez preemption resilient to incorrect free resource reported by YARN.

- TEZ-2842: Tez UI: Update Tez App details page while in-progress.

- TEZ-2844: Backport TEZ-2775. Improve and consolidate logging in Runtime components.

- TEZ-2846: Flaky test: TestCommit.testVertexCommit_OnDAGSuccess.

- TEZ-2847: Tez UI: Task details doesn't gets updated on manual refresh after job complete.

- TEZ-2850: Tez MergeManager OOM for small Map Outputs.

- TEZ-2851: Support a way for upstream applications to pass in a caller context to Tez.

- TEZ-2853: Tez UI: task attempt page is coming empty.

- TEZ-2857: Fix flakey tests in TestDAGImpl.

- TEZ-2863: Container, node, and logs not available in UI for tasks that fail to launch.

- TEZ-2866: Tez UI: Newly added columns wont be displayed by default in table.

- TEZ-2868: Fix setting Caller Context in Tez Examples.

- TEZ-2874: Improved logging for caller context.

- TEZ-2876: Tez UI: Update vertex, task & attempt details page while in progress.

- TEZ-2878: Tez UI: AM error handling - Make the UI handle cases in which AM returns unexpected/no date.

- TEZ-2882: Consider improving fetch failure handling.

- TEZ-2885: Remove counter logs from AMWebController.

- TEZ-2886: Ability to merge AM credentials with DAG credentials.

- TEZ-2887: Tez build failure due to missing dependency in pom files.

- TEZ-2893: Tez UI: Retain vertex info displayed in DAG details page even after completion.

- TEZ-2894: Tez UI: Disable sorting for few columns while in progress. Display an alert on trying to sort them.

- TEZ-2895: Tez UI: Add option to enable and disable in-progress.

- TEZ-2896: Fix thread names used during Input/Output initialization.

- TEZ-2898: Tez tools : swimlanes.py is broken.

- TEZ-2899: Backport graphical view fix from TEZ-2899.

- TEZ-2900: Ignore V_INPUT_DATA_INFORMATION when vertex is in Failed/Killed/Error.

- TEZ-2907: NPE in IFile.Reader.getLength during final merge operation.

- TEZ-2908: Tez UI: Errors are logged, but not displayed in the UI when AM fetch fails.

- TEZ-2909: Tez UI: Application link in All DAGs table is disable when applicationhistory is unavailable.

- TEZ-2910: Tez should invoke HDFS Client API to set up caller context.

- TEZ-2915: Tez UI: Getting back to the DAG details page is difficult.

- TEZ-2923: Tez Live UI counters view empty for vertices, tasks, attempt.

- TEZ-2927: Tez UI: Graciously fail when system-metrics-publisher is disabled.

- TEZ-2929: Tez UI: Dag details page displays vertices to be running even when dag have completed.

- TEZ-2930: Tez UI: Parent controller is not polling at time.

- TEZ-2933: Tez UI: Load application details from RM when available.

- TEZ-2936: Support HDFS-based Timeline writer.

- TEZ-2946: Tez UI: At times RM return a huge error message making the yellow error bar to fill the whole screen.

- TEZ-2947: Tez UI: Timeline, RM & AM requests gets into a consecutive loop in counters page without any delay.

- TEZ-2949: Allow duplicate dag names within session for Tez.

- TEZ-2960: Tez UI: Move hardcoded url namespace to the configuration file.

- TEZ-2963: RecoveryService#handleSummaryEvent exception with HDFS transparent encryption + kerberos authentication.

- TEZ-2968: Counter limits exception causes AM to crash.

- TEZ-2970: Re-localization in TezChild does not use correct UGI.

- TEZ-2975: Bump up apache commons dependency.

- TEZ-2988: DAGAppMaster:shutdownTezAM should return with a no-op if it has been invoked earlier.

- TEZ-2995: Timeline primary filter should only be on callerId and not type.

- TEZ-2997: Tez UI: Support searches by CallerContext ID for DAGs.

- 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-3036: Tez AM can hang on startup with no indication of error.

- TEZ-3037: History URL should be set regardless of which history logging service is enabled.

- TEZ-3052: Task internal error due to Invalid event: T_ATTEMPT_FAILED at FAILED.

- TEZ-3063: Tez UI : Display Input, Output, Processor, Source and Sink configurations under a vertex.

- TEZ-3066: TaskAttemptFinishedEvent ConcurrentModificationException in recovery or history logging services.

- TEZ-3086: Tez UI: Backward compatibility changes.

- TEZ-3101: Tez UI: Task attempt log link doesn't have the correct protocol.

- TEZ-3103: Shuffle can hang when memory to memory merging enabled.

- TEZ-3105: TezMxBeanResourceCalculator does not work on IBM JDK 7 or 8 causing Tez failures.

- TEZ-3107: tez-tools: Log warn messages in case ATS has wrong values (e.g. startTime > finishTime).

- TEZ-3114: Shuffle OOM due to EventMetaData flood.

- TEZ-3117: Deadlock in Edge and Vertex code.

- TEZ-3123: Containers can get re-used even with conflicting local resources.

- TEZ-3126: Log reason for not reducing parallelism.

- TEZ-3128: Avoid stopping containers on the AM shutdown thread.

- TEZ-3131: Support a way to override test_root_dir for FaultToleranceTestRunner.

- TEZ-3137: Tez task failed with illegal state exception.

- TEZ-3147: Intermediate mem-to-mem: Fix early exit when only one segment can fit into memory.

- TEZ-3155: Support a way to submit DAGs to a session where the DAG plan exceeds hadoop ipc limits.

- TEZ-3156: Tez client keeps trying to talk to RM even if RM does not know about the application.

- TEZ-3166: Counters aren't fully updated and sent for failed task.

- TEZ-3173: Update Tez AM REST APIs for more information for each vertex.

- TEZ-3175: Add tez client submit host.

- 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.

- TEZ-3192: IFile#checkState creating unnecessary objects though auto-boxing.

- TEZ-3193: Deadlock in AM during task commit request.

- TEZ-3196: java.lang.InternalError from decompression codec is fatal to a task during shuffle.

- TEZ-3202: Reduce the memory need for jobs with high number of segments.

- TEZ-3203: DAG hangs when one of the upstream vertices has zero tasks.

- TEZ-3213: Uncaught exception during vertex recovery leads to invalid state transition loop.

- TEZ-3223: Support a NullHistoryLogger to disable history logging if needed.

- TEZ-3224: User payload is not initialized before creating vertex manager plugin.

- TEZ-3233: Tez UI: Have LLAP information reflect in Tez UI.

- TEZ-3254: Tez UI: Consider downloading Hive/Pig explain plan.

- TEZ-3255: Tez UI: Hide swimlane while displaying running DAGs from old versions of Tez.

- TEZ-3256: [Backport HADOOP-11032] Remove Guava Stopwatch dependency.

- TEZ-3258: JVM Checker does not ignore DisableExplicitGC when checking JVM GC options.

- TEZ-3259: Tez UI: Build issue - File saver package is not working well with bower.

- TEZ-3262: Tez UI : zip.js is not having a bower friendly versioning system.

- TEZ-3264: Tez UI: UI discrepancies.

- TEZ-3276: Tez Example MRRSleep job fails when tez.staging-dir fs is not same as default FS.

- TEZ-3281: Tez UI: Swimlane improvement.

- TEZ-3286: Allow clients to set processor reserved memory per vertex (instead of per container).

- TEZ-3288: Tez UI: Display more details in the error bar.

- TEZ-3289: Tez Example MRRSleep job does not set Staging dir correctly on secure cluster.

- TEZ-3291: Optimize splits grouping when locality information is not available.

- TEZ-3292: Tez UI: UTs breaking with timezone change.

- TEZ-3293: Fetch failures can cause a shuffle hang waiting for memory merge that never starts.

- TEZ-3294: DAG.createDag() does not clear local state on repeat calls.

- TEZ-3295: TestOrderedWordCount should handle relative input/output path.

- TEZ-3297: Deadlock scenario in AM during ShuffleVertexManager auto reduce.

- TEZ-3304: TestHistoryParser fails with Hadoop 2.7.

- TEZ-3305: TestAnalyzer fails with Hadoop 2.7.

- TEZ-3308: Add counters to capture input split length.

- TEZ-3314: Double counting input bytes in MultiMRInput.

- TEZ-3318: Tez UI: Polling is not restarted after RM recovery.

- TEZ-3325: Flaky test in TestDAGImpl.testCounterLimits.

- TEZ-3327: ATS Parser: Populate config details available in dag.

- TEZ-3329: Tez ATS data is incomplete for a vertex which fails or gets killed before initialization.

- TEZ-3331: Add operation specific HDFS counters to ATS.

- TEZ-3333: Tez UI: Handle cases where Vertex/Task/Task Attempt data is missing.

- TEZ-3337: Do not log empty fields of TaskAttemptFinishedEvent to avoid confusion.

- TEZ-3357: Change TimelineCachePlugin to handle DAG grouping.

- TEZ-3370: Tez UI: Display the log link as N/A if the app does not provide a log line.

- TEZ-3374: Change TEZ_HISTORY_LOGGING_TIMELINE_NUM_DAGS_PER_GROUP conf key name.

- TEZ-3376: Fix groupId generation to account for dagId starting with 1.

- TEZ-3379: Tez analyzer: Move sysout to log4j.

- TEZ-3382: Tez analyzer: Should be resilient to new counters.

- TEZ-3398: Tez UI: Bread crumb link to Application from Application details dag/configuration tab is broken

1.7.21. Zeppelin

HDP 2.5.0 provides Zeppelin 0.6.0 and the following Apache patches:

- ZEPPELIN-1094: Run all paragraphs keeps appending empty paragraph.

- ZEPPELIN-1130: Make Livy create session retries configurable.

- ZEPPELIN-1143: Interpreter dependencies are not downloaded on zeppelin start.

- ZEPPELIN-1146: Zeppelin JDBC interpreter should work in a Kerberos environment.

- ZEPPELIN-1149: %sh interpreter kerberos support.

- ZEPPELIN-1159: Catch RestTemplate exception.

- ZEPPELIN-1166: Change the jquery-ui CSS file name.

- ZEPPELIN-1168: Add http header X-Requested-By in post request to Livy Interpreter.

- ZEPPELIN-1175: AM log is not available for yarn-client mode.

- ZEPPELIN-1199: Need to login using keytab and principal before creating SparkContext in secured cluster.

- ZEPPELIN-1212: User impersonation support in JDBC interpreter for Hive.

- ZEPPELIN-1242: Should set property SPARK_YARN_MODE and do login before creating any spark stuff.

- ZEPPELIN-1305: Fix bug of ZEPPELIN-1215.

- ZEPPELIN-1319: Use absolute path for ssl truststore and keystore when available.

- ZEPPELIN-1320: Run zeppelin interpreter process as web front end user.

1.7.22. ZooKeeper

HDP 2.5.0 provides ZooKeeper 3.4.6 and the following Apache patches:

- ZOOKEEPER-706: Large numbers of watches can cause session re-establishment to fail.

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

- ZOOKEEPER-1574: Set svn:eol-style property to native.

- ZOOKEEPER-1575: Adding .gitattributes to prevent CRLF and LF mismatches for source and text files.

- ZOOKEEPER-1576: ZooKeeper cluster - failed to connect to cluster if one of the provided IPs causes java.net.UnknownHostException.

- ZOOKEEPER-1791: ZooKeeper package includes unnecessary jars that are part of the package.

- ZOOKEEPER-1848: [WINDOWS] Java NIO socket channels does not work with Windows ipv6 on JD.

- ZOOKEEPER-1876: Add support for installing windows services in .cmd scripts.

- ZOOKEEPER-1888: ZkCli.cmd commands fail with "'java' is not recognized as an internal or external command".

- ZOOKEEPER-1897: ZK Shell/Cli not processing commands.

- ZOOKEEPER-1926: Unit tests should only use build/test/data for data.

- ZOOKEEPER-1952: Default log directory and file name can be changed.

- ZOOKEEPER-2053: ZooKeeper scripts should honor ZOOKEEPER_HOME

- ZOOKEEPER-2064: Prevent resource leak in various classes.

1.8. Common Vulnerabilities and Exposures

- **CVE-2016-5395:** Apache Ranger Stored Cross Site Scripting vulnerability

Severity: Moderate

Vendor: Hortonworks

Versions Affected: All HDP 2.3/2.4 versions including Apache Ranger versions 0.5.x

Users Affected: All users of ranger policy admin tool.

Impact: Apache Ranger was found to be vulnerable to a Stored Cross-Site Scripting in the create user functionality. Admin users can store some arbitrary javascript code to be executed when normal users login and access policies. See RANGER-1124.

Fix detail: Added logic to sanitize the user input.

Recommended Action: Users should upgrade to HDP 2.5+ (with Apache Ranger 0.6.1+)

1.9. 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
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle
Oozie	JDOM Oro	
Phoenix		EPL
Storm	Logback	EPL

1.10. 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

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-57918	HDFS	HDFS-10178	Permanent write failures can happen if pipeline recoveries occur for the first packet.
BUG-58254	HBase	HBASE-15811	Batch Get after batch Put does not fetch all Cells

Security

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-39696	Hive	HIVE-6026	LDAP Authenticator should be more generic with BindDN.
BUG-53072	Kafka	KAFKA-2854	The full Kerberos principal must be passed through for incoming requests to Kafka.
BUG-55917	Storm	STORM-1711	HiveUtils.authenticate reuses current UGI, mixes up principles in use.
BUG-56577	Accumulo, HDFS	HADOOP-12559, HADOOP-12682, HADOOP-12716	Port HOTFIX for HADOOP-12787 (KMS SPNEGO sequence does not work with WEBHDFS) and HADOOP-12559 (KMS connection failures should trigger TGT renewal).
BUG-57460	HBase, Phoenix	HBASE-15837, PHOENIX-2883	Invalid memstore state during close of region for Phoenix table.
BUG-58666	Hive	HIVE-13729	FileSystem$Cache leaks in FileUtils.checkFileAccessWithImpersonation.

Incorrect Results

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-35149	YARN	YARN-4392	ATS logs show start timestamp later than end timestamp.
BUG-49254	Ranger	RANGER-798	Ranger "Access > Audit" not showing anything with Source "Solr" due to time filter issue (GMT).
BUG-53364	HBase	HBASE-15325	ResultScanner allowing partial result will miss the rest of the row if the region is moved between two rpc requests.
BUG-55160	Phoenix	PHOENIX-1478	Can't upsert value of 127 into column of type unsigned tinyint.
BUG-56961	HBase	HBASE-15676	FuzzyRowFilter fails and matches all the rows in the table if the mask consists of all 0s.
BUG-57030	Phoenix	PHOENIX-2876	Phoenix ORDER BY clause does not work if the ORDER BY column is not present in the Select clause.
BUG-58764	HBase, Phoenix	PHOENIX-2915	Dropping of Index can still leave some non-replayed writes in WAL.
BUG-60197	Knox, Ranger	RANGER-1039	Knox permission details (User and Groups) are not listed from Ranger UI.
BUG-62125	Hive	HIVE-12768	Tez various wrong results with map <- (map , reduce) join on decimal query.

Stability

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-17827	Sqoop	SQOOP-2777	Sqoop Fails to Create Tables with Reserved Keywords.
BUG-18218	YARN	YARN-3846	Scheduler queue filter failed to work because index of queue column changed.
BUG-27069	Flume	FLUME-2118	Occasional multi-hour pauses in file channel replay.
BUG-41311	HDFS	HDFS-9917	Namenode quickly runs out of memory at start due to queued incremental block reports.
BUG-43539	HBase	HBASE-14229	Flushing canceled by coprocessor still leads to memstoreSize set down.
BUG-45878	Hive, Hive2	HIVE-13563	Hive Streaming does not honor orc.compress.size and orc.stripe.size table properties.
BUG-46789	Hive, Hive2	HIVE-13151	Hive metastore memory leak of fs cache with ACID compaction on.
BUG-48217	YARN	YARN-4325	After restarting YARN services, applications stays in ACCEPTED state.
BUG-49898	Hive	HIVE-12684	Autoquerygen tests failing in secure runs
BUG-51596	Hive	HIVE-12439	Hive Compactor cleaner thread fails to clean aborted txns due to ORA-01795 limit.
BUG-51816	Hive, Hive2	HIVE-12976	DirectSQL fails for large TPC-H runs.
BUG-52300	HBase	HBASE-15234	ReplicationLogCleaner can abort due to transient ZK issues.
BUG-53270	Hive		Query with duplicate columns in GROUP BY clause fails.
BUG-53536	Oozie	OOZIE-2185	Make oozie cli source conf/oozie-env.sh.
BUG-54166	HBase	HBASE-15441	Fix WAL splitting when region has moved multiple times.
BUG-54308	YARN	YARN-998	Dynamic resource persistent from NM/RM restart.
BUG-54498	Hive	HIVE-12165	Wrong result when hive.optimize.sampling.orderby=true with some aggregate functions.
BUG-54500	Hive	HIVE-12064	Prevent transactional=false.
BUG-54603	Tez	TEZ-3114	Shuffle OOM due to EventMetaData flood.
BUG-54671	Hive	HIVE-13233	Use min and max values to estimate better stats for comparison operators.

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-55447	Hadoop Common	HADOOP-11687	Ignore x-* and response headers when copying an Amazon S3 object.
BUG-55494	Hadoop Common	HADOOP-11901	BytesWritable fails to support 2G chunks due to integer overflow
BUG-55550	Hive	HIVE-13200	Aggregation functions returning empty rows on partitioned columns.
BUG-55810	Hive	HIVE-13242	DISTINCT keyword is dropped by the parser for windowing.
BUG-55816	Hive	HIVE-10555	Improve windowing spec of range based windowing to support additional range formats.
BUG-56083	Hive	HIVE-13394	Analyze table fails in tez on empty partitions/files/tables.
BUG-56420	HDFS	HDFS-10319	Balancer should not try to pair storages with different types.
BUG-56423	HDFS	HDFS-10312	NameNode is stuck at SafeMode due to not all data blocks validated
BUG-56528	Hive	HIVE-13596	Need JAR fix for UDF reloading.
BUG-56772	HDFS, Ranger	HADOOP-12423, HADOOP-12950, HADOOP-12993, RANGER-891	ShutdownHookManager should have a timeout for each of the Registered shutdown hook.
BUG-56869	HDFS	HDFS-9902	Support different values of dfs.datanode.du.reserved per storage-type.
BUG-56932	HDFS	HDFS-10335	Mover $Processor#chooseTarget() always chooses the first matching target storage group.
BUG-57007	HDFS	HDFS-9917	Namenode quickly runs out of memory at start due to queued incremental block reports.
BUG-57016	Hive, Hive2	HIVE-14083	ALTER INDEX via JDBC & HiveCLI causes NullPointerException.
BUG-57556	Hive, Hive2	HIVE-14027	NULL values produced by left outer join do not behave as null.
BUG-57732	HBase	HBASE-14712, HBASE-15100	MasterProcWALs never clean up.
BUG-57907	HBase	PHOENIX-2892	Scan for pre-warming the block cache for 2ndary index should be removed.
BUG-58450	Flume, Hive, Phoenix	PHOENIX-2568	Flume dependency of Phoenix is bringing in an old version of Thrift (0.7.0).

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-58850	Sqoop	SQOOP-2596	Sqoop import from sybase fails with varchars.
BUG-59089	Hive	HIVE-13592	Metastore calls map is not thread safe.
BUG-59268	Hive	HIVE-13632	Hive failing on insert empty array into parquet table.
BUG-60071	Hive	HIVE-10631, HIVE-12213	Create partitioned table takes longer time and then eventually fails.
BUG-60335	Hive, Hive2	HIVE-14147	Hive PPD might remove predicates when they are defined as a simple expression e.g. WHERE 'a'.

Upgrade

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-54675	Ambari, Kafka	AMBARI-16027	RU does not handle Kafka .8.* topics well (Changes broker IDS).
BUG-54742	Ambari, Ranger	RANGER-910	Error during EU while applying Ranger Java patches.
BUG-61600	Ambari, Ranger	AMBARI-16756	[EU] Ranger restart is not happening after upgrade.

Usability

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-41362	Sqoop	SQOOP-1361	Sqoop import with –as-avrofile does not work when DB table column names have '$' character in them.
BUG-42190	Ambari, Storm		Ambari start Storm UI reports failure, although UI starts.
BUG-46316	Sqoop	SQOOP-2471	Support for Importing of arrays and struct datatypes in Sqoop Hcatalog.
BUG-49832	Falcon	FALCON-678	Port doesn't change from 15000 to 15443 when change TLS to true.
BUG-50767	Falcon	FALCON-790	Can't edit Falcon process specifications on Falcon Web UI once process is saved.
BUG-52058	Ambari, Spark	AMBARI-16757	Spark History Server heap size is not exposed (History Server crashed with OOM).
BUG-52622	Falcon, Oozie, Spark	OOZIE-2170	Fix Issues in current Oozie Spark Integration.
BUG-56556	Workflow Designer	OOZIE-1983	Add spark action executor.
BUG-57755	HBase	HBASE-15808	Reduce potential bulk load intermediate space usage and waste.
BUG-60495	Hive	HIVE-14022	Left semi join should throw SemanticException if where

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
			clause contains columnname from right table.

Performance

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-50730	Hive	HIVE-12893	Hive optimize.sort.dynamic.partition not enabled in case of combination of bucketing and constant propagation if subset of partition column value set.
BUG-51723	Hadoop Common	HADOOP-12444	Consider implementing lazy seek in S3AInputStream.
BUG-52149	HBase	HBASE-15213	Fix increment performance regression.
BUG-52844	Ranger	RANGER-794, RANGER-836, RANGER-843, RANGER-844	Policy download optimizations.
BUG-53757	Hive, Hive2	HIVE-11424, HIVE-13233	Stats are not properly accounted for when multiple COUNT(DISTINCT) are in the query.
BUG-53830	Hive2	HIVE-13189, HIVE-13248	Performance improvements for date_add/date_sub/ to_date.
BUG-55291	HDFS	HDFS-10223	peerFromSocketAndKey performs SASL exchange before setting connection timeouts.
BUG-57513	HDFS	HDFS-9198, HDFS-9958	BlockManager#createLocatedBlocks can throw NPE for corruptBlocks on failed storages.
BUG-57906	HBase	HBASE-14460	Regression for CheckAndPut and Increment updates.

Other

Hortonworks Bug ID	Apache Component	Apache JIRA	Summary
BUG-56919	YARN	YARN-2910	FSLeafQueue can throw ConcurrentModificationException.
BUG-60199	Ranger	RANGER-1025	Policy search REST API implemented in public api V1 is not returning results as expected.

1.11. Known Issues

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
BUG-38148	ACCUMULO-4389	Accumulo	**Description of Problem:** Apache Accumulo has a feature called "Replication" which automatically propagates updates to one table to a list of other Accumulo cluster instances. This feature is used for disaster-recovery scenarios allowing data-center level failover. With this replication feature, there are a number of client API methods which support developer interactions with the feature.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			The ReplicationOperations#drain(String, Set) method is intended to serve as a blocking call which waits for all of the provided write-ahead log files that need to be replicated to other peers. Sometimes, the method reportedly does not actually wait for a sufficient amount of time. **Associated error message:** No direct error message is generated; the primary symptom is when the configured Accumulo replication peers do not have all of the expected data from the source Accumulo cluster. **Workaround:** None at this time. **Upstream fix:** https://issues.apache.org/jira/browse/ACCUMULO-4389 has been opened to track this issue.
BUG-40773	N/A	Kafka	**Description of Problem:** Kafka broker fails to start after disabling Kerberos security. **Workaround:** Before disabling Kerberos, you need to stop Kafka brokers. 1. Run the following command as the Kafka user: **./bin/kafka-run-class.sh kafka.admin.ZkSecurityMigrator –zookeeper.acl unsecure –zookeeper.connect 'hostname:2181'** 2. Follow the instructions for disabling Kerberos through Ambari. 3. Restart Kafka nodes.
BUG-42784	N/A	Phoenix	**sqlline** Shell for Phoenix Truncates Table Columns of Longer Rows in a Terminal If **sqlline** runs in a terminal that has insufficient width to display all of the columns in the SQL row, the additional columns are truncated to the maximum terminal width. In this case, resize the terminal width and restart **sqlline**. Alternatively, you can enable XML Output Format.
BUG-51100	SPARK-12516	Spark	**Component Affected:** Dynamic allocation enabled on YARN **Description of Problem:** Due to lack of blacklist mechanism, Spark can still schedule tasks on bad nodes where external shuffle has already occurred. This will lead to job failure and driver exit. **Workaround:** Avoid configuring `spark.yarn.max.executor.failures 3` to a static number. Spark itself will figure out a reasonable failure number whether dynamic allocation is enabled or not.
BUG-52223	HIVE-13014	Hive	**Component Affected:** Modules using Remote metastore such as Hive CLI and Streaming Ingest API in HCatalog. **Description of Problem:** In rare circumstances (due to network issues such as temporary partitions, lost messages, etc.), it is possible for a insert/update/delete operation on a transactional table to report a failure, when it actually committed successfully. **Workaround:** Embedded metastore (used by HiveServer2) is unaffected by this behavior.
BUG-55799	HIVE-12930	Hive	**Description of Problem:** SSL shuffle for LLAP is not supported

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Workaround: Currently, there is no workaround.
BUG-58308	PHOENIX-2067 and PHOENIX-2120	Phoenix	After an upgrade from an earlier HDP 2.x version to HDP 2.5, the rows of a table that has any of the following column types are not ordered correctly: • VARCHAR DESC columns • DECIMAL DESC columns • ARRAY DESC columns • Nullable DESC columns that are indexed (affects the index, but not the data table) • BINARY columns included in the primary key constraint This is a result of a column sort-order issue in Apache Phoenix. You can resolve this issue by upgrading the affected tables as described in Phoenix-4.5.0 Release Notes.
| BUG-59714 | HIVE-13974 | Hive | **Description of Problem:** ORC Schema Evolution does not support adding columns to a STRUCT type column unless the STRUCT column is the last column.<br><br>You can add column C to the last column last_struct:<br><br>```
CREATE TABLE orc_last_struct (
str STRING,
last_struct STRUCT<A:STRING,B:STRING>
) STORED AS ORC;

ALTER TABLE orc_last_struct REPLACE columns (str
 STRING, last_struct
STRUCT<A:STRING,B:STRING,C:BIGINT>);
```<br><br>You will be able to read the table.<br><br>However, in this table:<br><br>```
CREATE TABLE orc_inner_struct (
str STRING,
inner_struct STRUCT<A:STRING,B:STRING>,
last DATE
) STORED AS ORC;

ALTER TABLE orc_inner_struct REPLACE columns (str
 STRING, inner_struct
STRUCT<A:STRING,B:STRING,C:BIGINT>, last DATE);
```<br><br>You will not be able to read the table.<br>You will get execution errors like:<br>`java.lang.ArrayIndexOutOfBoundsException.`<br><br>**Workaround:** The workaround is not to use tables with Schema Evolution in inner STRUCT type columns. |
| BUG-60690 | KNOX-718 | Knox | **Description of Problem:** Unable to log in using Knox SSO even when providing correct credentials. This is because the whitelist is not correctly configured. The login page will not provide an error message to indicate a reason for the failed login.

Associated error message: Found in `<log_directory_for_knox>/gateway.log`

Workaround: In `knoxsso.xml`, modify the value of the `knoxsso.redirect.whitelist.regex` parameter to reflect the configuration of your environment. |
| BUG-61739 | N/A | Knox | **Description of Problem:** Defined configurations can exist in two places but cause problems if they do not match. |

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Typically when using Ambari, you do not need to define the same configuration in multiple places. However, in some cases, defining the configuration multiple times is necessary. **Workaround:** Typically, HiveServer2 is setup in HA mode, which uses ZooKeeper. You can now use Knox to access HiveServer2 in HA mode by using the following configuration in a Knox topology file. ``` <provider> <role>ha</role> <name>HaProvider</name> <enabled>true</enabled> <param> <name>HIVE</name> <value>maxFailoverAttempts=3;failoverSleep=1000;enabled=true;zookeeperEnsemble=machine1:2181,machine2:2181,machine3:2181; zookeeperNamespace=hiveserver2</value> </param> </provider> ``` In the above configuration, you must define the value of the `zookeeperEnsemble` property with the same value as the `hive.zookeeper.quorum` parameter in the Hive configuration file `hive-site.xml`. If the value of `hive.zookeeper.quorum` changes, you will need to manually change the value of `zookeeperEnsemble` to match the change.
BUG-62423	N/A	Falcon	For HDP 2.5.0, Falcon binaries ship with `zookeeper-3.4.6.2.4.3.0-126-tests.jar`.
BUG-62588	N/A	Storm	**Description of Problem:** Storm does not support rolling upgrade from a previous HDP version to HDP-2.5. **Solution:** If your cluster is managed by Ambari, the rolling upgrade process will ask you to stop Storm topologies, perform the upgrade, and redeploy your topologies. If your cluster is not managed by Ambari, perform the following manual upgrade steps for Storm before starting the rolling upgrade process: 1. Stop all topologies. 2. Stop all storm daemons. 3. Delete storm.local.dir contents on all nodes. 4. Delete storm.zookeeper.root node. Next, upgrade the cluster to HDP-2.5. To finish the Storm upgrade process: start the storm daemons, and then redeploy the topologies.
BUG-62662	HADOOP-13382	Hadoop Common	**Description of Problem:** This backward-incompatible change was done to remove a CVE vulnerability due to commons-httpclient-3.1. **Workaround:** Projects or Java jobs with undeclared transitive dependencies on commons-httpclient, previously provided via hadoop-common or hadoop-client, will have to either stop using commons-httpclient (recommended), or import it as an explicit dependency.
BUG-63132	N/A	Storm	**Summary:** Solr bolt does not run in a Kerberos environment.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Associated error message: The following is an example: [ERROR] Request to collection hadoop_logs failed due to (401) org.apache.solr.client.solrj.impl.HttpSolrClient $RemoteSolrException: Error from server at http:[...] Error 401 Authentication required **Workaround:** None at this time.
BUG-63165	PHOENIX-3126	Zeppelin	**Description of problem**: When Kerberos is enabled in the cluster, Kerberos-based user authentication in the Zeppelin UI is not correctly passed to Phoenix/HBase. The user credentials will be unavailable to Phoenix, resulting in standard HBase authentication/authorization schemes working as intended. **Associated error message**: Unexpected failed authentication and authorization messages from Zeppelin in talking to Phoenix/HBase. **Workaround**: There is no known workaround at this time. This issue will be addressed in a future maintenance release.
BUG-63226	OOZIE-1983	Oozie	**Component Affected:** Spark shared library **Description of Problem:** In HDP 2.5, after you convert an insecure cluster to a secure cluster, or vice versa, the Oozie Spark share library does not update with the required secure parameters. **Workaround:** To update, run the following command as "oozie" user. **/usr/hdp/current/oozie-client/bin/oozie-setup.sh sharelib create -fs hdfs://<HDFS FQDN>:8020 -locallib /usr/hdp/ current/oozie-client/oozie-sharelib.tar.gz** When the command completes, restart Oozie from the Ambari web UI. • In Ambari, select Oozie from the services list. • On the Oozie Summary page, click **Service Actions > Restart All**.
BUG-63833	AMBARI-18096	Spark and other components that process long-running jobs.	**Description of Problem:** Long running jobs (such as Spark Streaming jobs) do not work in secure clusters when transparent data encryption is enabled. **Workaround:** None at this time.
BUG-63885	HIVE-14446	Hive, Hive2	**Component Affected:** ACID **Description of Problem:** Small tables estimated to have about 300 million rows that broadcast to a Mapjoin will cause the BloomFilter to overflow. Typically, this is due to bad stats estimation. **Workaround:** It is possible to avoid this issue with the following: `set hive.mapjoin.hybridgrace.hashtable=false` However, if this is caused by bad stats estimation and Hybrid grace hash join does not work, the regular mapjoin also will not work.
BUG-64028	N/A	Ranger	**Component Affected:** Create Policy Audit **Description of Problem:** When attempting to view the details of a Audit record associated with a deleted ranger

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			repository, the admin UI shows Page Not Found Error Page (401). **Workaround:** Currently, there is no workaround for this. This will be addressed in a future release.
BUG-64033	RANGER-1143	Ranger	tagsync policy exist but still authorization is failing (see Workaround for BUG-64033).
BUG-64084	N/A	Atlas, Storm	**Description of Problem:** Hive topology fails when the `hive-site.xml` contains an Atlas hook that tries to register any new tables/partitions created through the hcatalog streaming API. Currently, the use case that will cause this failure is copying the `hive-site.xml` from the target cluster to your topology codebase and packaging/creating an uber jar. **Associated Error Message:** Since the Atlas hook and its configuration is not getting packaged with Storm Topology jar, the result is NoClassDefFoundError. **Workaround:** After copying the `hive-site.xml` to your topology code, delete the Atlas hook configuration references and than package the jar.
BUG-64098	N/A	Spark	**Description of Problem:** When installing Spark manually on Debian/Ubuntu, the **apt-get install spark command** does not install all Spark packages. **Workaround:** Use the `-t` option in your apt-get install command: **apt-get install -t HDP spark**
BUG-64385	N/A	Falcon, Hive2	**Description of Problem:** Using HSI for writes and HiveDR are incompatible. This affects all users of HSI (hive2 hiveserver2-interactive) and HiveDR(hive replication). The resolution of BUG-64385 introduces a side effect. If you use HSI to write to warehouse, your changes will not replicate, which can lead to missed updates. **Workaround:** If you want to use Replication, use HSI for reads only.
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 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-64965	N/A	Zeppelin	**Component Affected:** Zeppelin UI **Description of Problem:** When Zeppelin SSL is enabled, the Zeppelin UI is unavailable through Safari due to a WebSocket network error: `WebSocket network error: OSStatus Error -9807: Invalid certificate chain`

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Workaround: This occurs due to the use of self signed certificates. Self-signed certificates require OS or Browser specific steps that you must follow prior to use in production. In production, use Certificate Authority signed certificate to prevent this error from occurring.

BUG-65028	N/A	Zeppelin	**Description of Problem:** On secure clusters that run Zeppelin, configure settings to limit interpreter editing privileges to admin roles.
			
			Workaround: Add the following lines to the [urls] section of the Zeppelin `shiro.ini` configuration file. For a cluster not managed by Ambari, add the lines to `/etc/zeppelin/conf/shiro.ini`.
			
			`/api/interpreter/** = authc, roles[admin]`
			
			`/api/configurations/** = authc, roles[admin]`
			
			`/api/credential/** = authc, roles[admin]`

BUG-65043	N/A	Ambari, Atlas	**Description of Problem:** When upgrading a secure cluster from HDP-2.4.x to HDP-2.5.x, the Kafka service has incorrect security properties. **Workaround:** After upgrading from HDP 2.4x to 2.5x (and after removing and then replacing Atlas, as recommended); manually update the user-defined Kerberos descriptor to use the 2.5 stack default values (if acceptable). If default values are unacceptable, copy Kerbeos descriptor from the stack, make required changes, and then replace the descriptor. Then, use Regenerate Keytabs to create the missing configurations, principals, and keytab files.

BUG-65058	N/A	Ambari, Hive	**Description of Problem:** LLAP containers may end up getting killed due to insufficient memory available in the system.
			
			Assocaited Error Message: The following messages in the AM log of LLAP YARN Application.
			
			`# There is insufficient memory for the Java Runtime Environment to continue.`
			`# Native memory allocation (mmap) failed to map 194347270144 bytes for committing reserved memory.`
			`# An error report file with more information is saved as:`
			
			Workaround: Reduce the YARN NodeManager available memory. This is defind as the Memory allocated for all YARN containers on a node under the YARN Configuration tab.
			
			Desription of Problem: LLAP daemons can be killed by the YARN Memory Monitor
			
			Associated Error Message: The following messages in the AM log of LLAP YARN Application.
			
			`is running beyond physical memory limits. Current usage: <USED> of <ALLOCATED> GB physical memory used`
			
			Workaround: Lower the **LLAP heap size** under the **Advanced hive-interactive-env** section of the Advanced Hive config.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Note You will need to change this value each time any configs are changed under the Hive Interactive section on the main Hive Config page.
BUG-65080	N/A	Ambari, Atlas	**Description of Problem:** Atlas web UI produces inaccessible alert after adding Atlas service on upgraded cluster **Workaround:** 1. Stop Atlas Server. 2. Copy solr xml files to correct config folder and **chown** as $atlas_user:$hadoop_group: `cp -R /usr/hdp/2.5.0.0-1245/etc/atlas/conf.dist/solr/* /etc/atlas/conf/solr/` `cp: overwrite `/etc/atlas/conf/solr/solrconfig.xml'? n` `chown atlas:hadoop /etc/atlas/conf/solr/*`  `cp /usr/hdp/2.5.0.0-1245/etc/atlas/conf.dist/user-credentials.properties /etc/atlas/conf/` `cp /usr/hdp/2.5.0.0-1245/etc/atlas/conf.dist/policy-store.txt /etc/atlas/conf/`  `chown atlas:hadoop /etc/atlas/conf/users-credentials.properties` `chown atlas:hadoop /etc/atlas/conf/policy-store.txt`  3. Delete zookeeper znode:  `# kinit -kt /etc/security/keytabs/atlas.service.keytab atlas/<HOST>@<DOMAIN>` `# cd /usr/hdp/current/zookeeper-client/bin/` `# ./zkCli.sh -server` ` <zookeepernode>:<zookeeperport>` `[...... (CONNECTED)] rmr /infra-solr/configs/atlas_configs`  4. Ensure Atlas application-properties are present:  `atlas.jaas.KafkaClient.option.keyTab = /etc/security/keytabs/atlas.service.keytab` `atlas.jaas.KafkaClient.option.principal = atlas/_HOST@EXAMPLE.COM` 5. Start Atlas.
BUG-65286	ATLAS-1147	Atlas	**Component Affected:** Atlas UI **Description of Problem:** On the **Schema** tab, the **Name** column is missing. **Workaround:** Install the available patch: 1. Go to ATLAS-1147. 2. Download `SchemaLayoutView.js`. 3. Copy `SchemaLayoutView.js` in the hosts that run Atlas server to: `/usr/hdp/current/atlas-server/server/webapp/atlas/js/views/schema` 4. Refresh the browser to update the browser cache.
BUG-66325, BUG-66326	N/A	Zeppelin	**Description of Problem:** Zeppelin (with or without Livy) cannot access data on encrypted (TDE) clusters when the default user settings are in effect.

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			Workaround: 1. Add the following proxy users to the Ranger KMS configuration, replacing 'livy' and 'yarn' with the actual configured service user names for Livy and YARN, if they differ from the default service users livy and yarn on your cluster: hadoop.kms.proxyuser.livy.groups=* hadoop.kms.proxyuser.livy.hosts=* hadoop.kms.proxyuser.livy.users=* hadoop.kms.proxyuser.yarn.groups=* hadoop.kms.proxyuser.yarn.hosts=* hadoop.kms.proxyuser.yarn.users=* 2. Add the following property and setting to your `yarn-site.xml` file: yarn.resourcemanager.proxy-user-privileges.enabled=true 3. Restart KMS and YARN Resource Manager.
BUG-69158	N/A	Zeppelin, Spark	**Description of Problem:** By default, the Livy server times out after being idle for 60 minutes. **Associated error message:** Subsequent attempts to access Livy generate an error, Exception: Session not found, Livy server would have restarted, or lost session. **Workaround:** Set the timeout to a larger value through the property livy.server.session.timeout, and restart the Zeppelin Livy interpreter.
BUG-77311	N/A	Zeppelin	**Description of Problem:** When one user restarts the %livy interpreter from the Interpreters (admin) page, other users' sessions restart too. **Workaround:** Restart the %livy interpreter from within a notebook.
BUG-80901	N/A	Zeppelin	**Component Affected:** Zeppelin/Livy **Description of Problem:** This occurs when running applications through Zeppelin/Livy that requires some 3rd-party libraries. These libraries cannot be installed on all nodes in the cluster but they are installed on their edge nodes. Running in yarn-client mode this all works as the job is submitted on the edge node where the libraries are installed and runs there. In yarn-cluster mode, it fails because the libraries are missing. **Workaround:** Set either spark.jars in `spark-defaults.conf` or livy.spark.jars in livy interpreters conf. Both are globally applicable. The jars need to be present on the livy machine in both cases. Updating livy conf is preferable since it affects only the zeppelin users.
RMP-5613	N/A	Zeppelin	**Component Affected:** Zeppelin **Description of Problem:** Zeppelin is not supported on Internet Explorer 8/9 because there is no native support for websockets. **Workaround:** Use one of the following browsers: • Internet Explorer 10 or 11 • Google Chrome latest stable release

Hortonworks Bug ID	Apache JIRA	Apache Component	Summary
			• Firefox latest stable release • Safari latest stable release
RMP-7856	HBASE-14417	HBase	Incremental backups do not capture bulk-loaded data.
RMP-7858	HBASE-14141	HBase	All HBase WAL data is copied to the backup destination during an incremental backup, including data for tables that are not part of the backup. This is a limitation for the performance and security aspects of the HBase backup-and-restore feature. HBASE-14141 will introduce more granular copy WAL implementation.

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

Technical Service Bulletin	Apache JIRA	Apache Component	Summary
			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

1.12. Documentation Errata

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

1.12.1. Workaround for BUG-64033

Description of Problem: Tags are not getting synchronized from any source in SSL environment or when Ranger admin is SSL enabled (BUG-64033).

Error Message: Tagsync process is not coming up due to below error in file `/var/log/ranger/tagsync/tagsync.log`

```
10 Aug 2016 07:25:08  INFO TagSynchronizer [main] – 177
        at org.apache.hadoop.conf.Configuration.getProps(Configuration.
java:2418)
        at org.apache.hadoop.conf.Configuration.get(Configuration.java:981)
        at org.apache.ranger.plugin.util.RangerRESTClient.
init(RangerRESTClient.java:246)
        at org.apache.ranger.plugin.util.RangerRESTClient.
<init>(RangerRESTClient.java:116)
        at org.apache.ranger.tagsync.sink.tagadmin.TagAdminRESTSink.
initialize(TagAdminRESTSink.java:97)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
initializeTagSink(TagSynchronizer.java:202)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
initialize(TagSynchronizer.java:104)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
main(TagSynchronizer.java:60)
10 Aug 2016 07:25:10 ERROR TagSynchronizer [main] – 207 Failed to initialize
 TAG sink. Error details:
java.lang.RuntimeException: com.sun.org.apache.xerces.internal.impl.io.
MalformedByteSequenceException: Invalid byte 1 of 1–byte UTF–8 sequence.
        at org.apache.hadoop.conf.Configuration.loadResource(Configuration.
java:2673)
        at org.apache.hadoop.conf.Configuration.loadResources(Configuration.
java:2536)
        at org.apache.hadoop.conf.Configuration.getProps(Configuration.
java:2418)
        at org.apache.hadoop.conf.Configuration.get(Configuration.java:981)
        at org.apache.ranger.plugin.util.RangerRESTClient.
init(RangerRESTClient.java:246)
```

```
        at org.apache.ranger.plugin.util.RangerRESTClient.
<init>(RangerRESTClient.java:116)
        at org.apache.ranger.tagsync.sink.tagadmin.TagAdminRESTSink.
initialize(TagAdminRESTSink.java:97)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
initializeTagSink(TagSynchronizer.java:202)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
initialize(TagSynchronizer.java:104)
        at org.apache.ranger.tagsync.process.TagSynchronizer.
main(TagSynchronizer.java:60)
Caused by: com.sun.org.apache.xerces.internal.impl.io.
MalformedByteSequenceException: Invalid byte 1 of 1-byte UTF-8 sequence.
        at com.sun.org.apache.xerces.internal.impl.io.UTF8Reader.
invalidByte(UTF8Reader.java:687)
        at com.sun.org.apache.xerces.internal.impl.io.UTF8Reader.
read(UTF8Reader.java:557)
        at com.sun.org.apache.xerces.internal.impl.XMLEntityScanner.
load(XMLEntityScanner.java:1753)
        at com.sun.org.apache.xerces.internal.impl.XMLEntityScanner.
arrangeCapacity(XMLEntityScanner.java:1629)
        at com.sun.org.apache.xerces.internal.impl.XMLEntityScanner.
skipString(XMLEntityScanner.java:1667)
        at com.sun.org.apache.xerces.internal.impl.XMLVersionDetector.
determineDocVersion(XMLVersionDetector.java:196)
        at com.sun.org.apache.xerces.internal.parsers.XML11Configuration.
parse(XML11Configuration.java:812)
        at com.sun.org.apache.xerces.internal.parsers.XML11Configuration.
parse(XML11Configuration.java:777)
        at com.sun.org.apache.xerces.internal.parsers.XMLParser.
parse(XMLParser.java:141)
        at com.sun.org.apache.xerces.internal.parsers.DOMParser.
parse(DOMParser.java:243)
        at com.sun.org.apache.xerces.internal.jaxp.DocumentBuilderImpl.
parse(DocumentBuilderImpl.java:347)
        at javax.xml.parsers.DocumentBuilder.parse(DocumentBuilder.java:121)
        at org.apache.hadoop.conf.Configuration.parse(Configuration.java:2514)
        at org.apache.hadoop.conf.Configuration.loadResource(Configuration.
java:2587)
```

Workaround:

1. Create file `ranger-policymgr-ssl.xml` under /usr/hdp/current/ranger-tagsync/
   conf/ with below content

```
<configuration>
    <property>
      <name>xasecure.policymgr.clientssl.keystore</name>
      <value>/etc/security/serverKeys/ranger-tagsync-keystore.jks</value>
    </property>

    <property>
      <name>xasecure.policymgr.clientssl.keystore.credential.file</name>
      <value>jceks://file/etc/ranger/tagsync/cred.jceks</value>
    </property>

    <property>
      <name>xasecure.policymgr.clientssl.keystore.password</name>
      <value>myKeyFilePassword</value>
    </property>
```

```
    <property>
      <name>xasecure.policymgr.clientssl.truststore</name>
      <value>/etc/security/serverKeys/ranger-tagsync-mytruststore.jks</
value>
    </property>

    <property>
      <name>xasecure.policymgr.clientssl.truststore.credential.file</name>
      <value>jceks://file/etc/ranger/tagsync/cred.jceks</value>
    </property>

    <property>
      <name>xasecure.policymgr.clientssl.truststore.password</name>
      <value>changeit</value>
    </property>

  </configuration>
```

2. Created below directory path if not exist:

   a. `/etc/ranger/tagsync/`

   b. `/etc/security/serverKeys/`

3. Create `ranger-tagsync-keystore.jks`:

   > **Note**
   >
   > This is just provided as an example. How you implement this step depends on your deployment.

   a. **keytool -genkey -keyalg RSA -alias rangerTagsync -keystore /etc/security/serverKeys/ranger-tagsync-keystore.jks -storepass myKeyFilePassword -validity 360 -keysize 2048**

   b. **chmod 640 /etc/security/serverKeys/ranger-tagsync-keystore.jks**

   c. **chmod ranger:ranger /etc/security/serverKeys/ranger-tagsync-keystore.jks [1]**

4. Create truststore `ranger-tagsync-mytruststore.jks`:

   > **Note**
   >
   > This is just provided as an example. How you implement this step depends on your deployment.

   a. Export `ranger-admin-keystore.jks` into `ranger-admin-trust.cer` running the below command on ranger-admin host [2]:

      keytool -export -keystore /etc/ranger/admin/conf/ranger-admin-keystore.jks -alias rangeradmin -file ranger-admin-trust.cer

   b. Import `ranger-admin-trust.cer` into `ranger-tagsync-mytruststore.jks`.

      keytool -import -file ranger-admin-trust.cer -alias rangeradmintrust -keystore /etc/security/serverKeys/ranger-tagsync-mytruststore.jks -storepass changeit

    c. **chmod 640 /etc/security/serverKeys/ranger-tagsync-mytruststore.jks**

    d. **chmod ranger:ranger /etc/security/serverKeys/ranger-tagsync-mytruststore.jks [1]**

5. Create `cred.jceks`:

    a. **java -cp "/usr/hdp/current/ranger-tagsync/lib/*" org.apache.ranger.credentialapi.buildks create sslKeyStore -value myKeyFilePassword -provider jceks://file/etc/ranger/tagsync/cred.jceks**

    b. **java -cp "/usr/hdp/current/ranger-tagsync/lib/*" org.apache.ranger.credentialapi.buildks create sslTrustStore -value changeit -provider jceks://file/etc/ranger/tagsync/cred.jceks**

    c. **chmod 640 /etc/ranger/tagsync/cred.jceks**

    d. **chown ranger:ranger /etc/ranger/tagsync/cred.jceks [1]**

6. From Ambari:

    a. In **Ranger config -> Advanced -> Advanced ranger-tagsync-site**, update the ranger.tagsync.dest.ranger.ssl.config.filename property value to /usr/hdp/current/ranger-tagsync/conf/ranger-policymgr-ssl.xml

    b. Restart Ranger Tagsync.

Note:

1. Supply ownership to the ranger tagsync process user for .jks and .jceks files.

2. You will need to copy `ranger-admin-trust.cer` to ranger-tagsync host if ranger-admin and ranger-tagsync are not running on the same host.