

Hortonworks Data Platform

Release Notes

(May 26, 2015)

Hortonworks Data Platform: Release Notes

Copyright © 2012-2015 Hortonworks, Inc. All 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 Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, ZooKeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

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

For more information on Hortonworks technology, see [the Hortonworks Data Platform](#). For more information on Hortonworks services, see either the [Support](#) or [Training](#) page. Feel free to [Contact Us](#) directly to discuss your specific needs.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Table of Contents

1. Release Notes HDP-2.1.15	1
1.1. Product Version: HDP-2.1.15	1
1.2. Unsupported Apache Components	2
1.3. Fixed in This Release	2
1.4. Behavioral Changes	4
1.4.1. Hive behavioral changes	5
1.5. Patch Information	5
1.5.1. Patch information for Hadoop Common/HDFS	5
1.5.2. Patch information for ZooKeeper	7
1.5.3. Patch information for HBase	7
1.5.4. Patch information for Pig	11
1.5.5. Patch information for Tez	11
1.5.6. Patch information for Hive/HCatalog	12
1.5.7. Patch information for Oozie	15
1.5.8. Patch information for Sqoop	16
1.5.9. Patch information for Falcon	17
1.5.10. Patch information for Knox	18
1.5.11. Patch information for Flume	18
1.6. Minimum System Requirements	19
1.6.1. Hardware recommendations	19
1.6.2. Operating systems requirements	19
1.6.3. Software requirements	19
1.6.4. Database requirements	20
1.6.5. Virtualization and cloud platforms	20
1.6.6. Configuring the local repositories	20
1.7. Upgrading From HDP 2.1.7 to HDP 2.1.15	20
1.8. Improvements	23
1.9. Common Vulnerabilities and Exposures	23
1.10. Known Issues	24
1.10.1. Known Issues for Hive	24
1.10.2. Known Issues for Flume	24
1.10.3. Known Issues for the Hortonworks Connector for Teradata	24
1.10.4. Known Issues for Ambari	24
1.11. Deprecated Features	25
1.12. Third-party Licenses	25
2. Release Notes HDP-2.1.10	26
2.1. Product Version: HDP-2.1.10	26
2.2. Unsupported Apache components	27
2.3. Fixed in This Release	27
2.4. Behavioral Changes	29
2.4.1. Hive behavioral changes	30
2.5. Patch Information	30
2.5.1. Patch information for Hadoop Common/HDFS	30
2.5.2. Patch information for ZooKeeper	32
2.5.3. Patch information for HBase	32
2.5.4. Patch information for Pig	36
2.5.5. Patch information for Tez	36
2.5.6. Patch information for Hive/HCatalog	37

2.5.7. Patch information for Oozie	40
2.5.8. Patch information for Sqoop	41
2.5.9. Patch information for Falcon	43
2.5.10. Patch information for Knox	43
2.5.11. Patch information for Flume	43
2.6. Minimum System Requirements	44
2.6.1. Hardware recommendations	44
2.6.2. Operating systems requirements	45
2.6.3. Software requirements	45
2.6.4. Database requirements	45
2.6.5. Virtualization and cloud platforms	45
2.6.6. Configuring the local repositories	46
2.7. Upgrading From HDP 2.1.7 to HDP 2.1.10	46
2.8. Improvements	48
2.9. Common Vulnerabilities and Exposures	49
2.10. Known Issues	49
2.10.1. Known Issues for HBase	50
2.10.2. Known Issues for Hive	50
2.10.3. Known Issues for Tez	50
2.10.4. Known Issues for Oozie	50
2.10.5. Known Issues for Flume	51
2.10.6. Known Issues for the Hortonworks Connector for Teradata	51
2.10.7. Known Issues for Ambari	51
2.11. Deprecated Features	51
2.12. Third-party Licenses	51
3. Release Notes HDP-2.1.7	52
3.1. Product Version: HDP-2.1.7	52
3.2. Unsupported Apache components	53
3.3. Fixed in This Release	53
3.4. Behavioral Changes	55
3.4.1. Hive behavioral changes	55
3.5. Patch Information	55
3.5.1. Patch information for Hadoop Common/HDFS	56
3.5.2. Patch information for ZooKeeper	57
3.5.3. Patch information for HBase	58
3.5.4. Patch information for Pig	61
3.5.5. Patch information for Tez	61
3.5.6. Patch information for Hive/HCatalog	62
3.5.7. Patch information for Oozie	65
3.5.8. Patch information for Sqoop	66
3.5.9. Patch information for Falcon	67
3.5.10. Patch information for Knox	68
3.5.11. Patch information for Flume	68
3.6. Minimum System Requirements	69
3.6.1. Hardware recommendations	69
3.6.2. Operating systems requirements	69
3.6.3. Software requirements	69
3.6.4. Database requirements	70
3.6.5. Virtualization and cloud platforms	70
3.6.6. Configuring the local repositories	70
3.7. Upgrading From HDP 2.1.5 to HDP 2.1.7	70

3.8. Improvements	73
3.9. Common Vulnerabilities and Exposures	73
3.10. Known Issues	74
3.10.1. Known Issues for HBase	74
3.10.2. Known Issues for Hive	74
3.10.3. Known Issues for Tez	75
3.10.4. Known Issues for Oozie	75
3.10.5. Known Issues for Flume	75
3.10.6. Known Issues for the Hortonworks Connector for Teradata	75
3.10.7. Known Issues for Ambari	75
3.11. Deprecated Features	75
3.12. Third-party Licenses	75
4. Release Notes HDP-2.1.5	77
4.1. Product Version: HDP-2.1.5	77
4.2. Unsupported Apache components	78
4.3. Tech Previews in This Release	78
4.4. Fixed in This Release	79
4.5. Behavioral Changes	80
4.5.1. Hive behavioral changes	80
4.6. Patch Information	80
4.6.1. Patch information for Hadoop Common/HDFS	81
4.6.2. Patch information for ZooKeeper	82
4.6.3. Patch information for HBase	83
4.6.4. Patch information for Pig	86
4.6.5. Patch information for Tez	86
4.6.6. Patch information for Hive/HCatalog	87
4.6.7. Patch information for Oozie	90
4.6.8. Patch information for Sqoop	91
4.6.9. Patch information for Falcon	93
4.6.10. Patch information for Knox	93
4.7. Minimum System Requirements	93
4.7.1. Hardware recommendations	93
4.7.2. Operating systems requirements	93
4.7.3. Software requirements	94
4.7.4. Database requirements	94
4.7.5. Virtualization and cloud platforms	94
4.7.6. Configuring the local repositories	95
4.8. Upgrading From HDP 2.1.3 to HDP 2.1.5	95
4.9. Improvements	97
4.10. Common Vulnerabilities and Exposures	97
4.11. Known Issues	98
4.11.1. Known Issues for HBase	98
4.11.2. Known Issues for Hive	98
4.11.3. Known Issues for Tez	99
4.11.4. Known Issues for Oozie	99
4.11.5. Known Issues for the Hortonworks Connector for Teradata	99
4.11.6. Known Issues for Ambari	99
4.12. Deprecated Features	99
4.13. Third-party Licenses	99
5. Release Notes HDP-2.1.3	101
5.1. Product Version: HDP-2.1.3	101

5.2. Unsupported Apache components	102
5.3. Tech Previews in This Release	103
5.4. Fixed in This Release	103
5.5. Behavioral Changes	103
5.5.1. YARN behavioral changes	103
5.6. Patch Information	103
5.6.1. Patch information for Hadoop Common/HDFS	104
5.6.2. Patch information for ZooKeeper	105
5.6.3. Patch information for HBase	105
5.6.4. Patch information for Pig	109
5.6.5. Patch information for Tez	109
5.6.6. Patch information for Hive/HCatalog	110
5.6.7. Patch information for Oozie	112
5.6.8. Patch information for Sqoop	113
5.7. Minimum System Requirements	115
5.7.1. Hardware recommendations	115
5.7.2. Operating systems requirements	115
5.7.3. Software requirements	116
5.7.4. Database requirements	116
5.7.5. Virtualization and cloud platforms	116
5.7.6. Configuring the local repositories	117
5.8. Upgrading From HDP 2.1.2 to HDP 2.1.3	117
5.9. Improvements	120
5.10. Common Vulnerabilities and Exposures	120
5.11. Known Issues	120
5.11.1. Known Issues for HDP	121
5.11.2. Known Issues for YARN	122
5.11.3. Known Issues for HBase	122
5.11.4. Known Issues for Apache Pig	122
5.11.5. Known Issues for Hive	122
5.11.6. Known Issues for Hue	125
5.11.7. Known Issues for Storm	127
5.11.8. Known Issues for Knox	127
5.11.9. Known Issues for the Hortonworks Connector for Teradata	127
5.11.10. Known Issues for Ambari	127
5.12. Deprecated Features	127
5.13. Third-party Licenses	127
6. Release Notes HDP-2.1.2	129
6.1. Product Version: HDP-2.1.2	129
6.2. Unsupported Apache components	130
6.3. Fixed in This Release	131
6.4. Behavioral Changes	131
6.4.1. Mahout behavioral changes	131
6.4.2. HDP 2.1 clusters deployed via Ambari	132
6.4.3. Hue behavioral changes	132
6.4.4. HBase behavioral changes	132
6.4.5. Hive behavioral changes	132
6.4.6. Oozie behavioral changes	132
6.5. Patch Information	132
6.5.1. Patch information for Hadoop Common/HDFS	133
6.5.2. Patch information for ZooKeeper	133

6.5.3. Patch information for HBase	133
6.5.4. Patch information for Pig	137
6.5.5. Patch information for Tez	137
6.5.6. Patch information for Hive/HCatalog	138
6.5.7. Patch information for Oozie	139
6.5.8. Patch information for Sqoop	140
6.6. Minimum System Requirements	141
6.6.1. Hardware recommendations	142
6.6.2. Operating systems requirements	142
6.6.3. Software requirements	142
6.6.4. Database requirements	143
6.6.5. Virtualization and cloud platforms	143
6.6.6. Configuring the local repositories	143
6.7. Upgrading From HDP 2.1.1 to HDP 2.1.2	143
6.8. Improvements	146
6.9. Common Vulnerabilities and Exposures	146
6.10. Known Issues	146
6.10.1. Known Limitation for HDFS Upgrade	147
6.10.2. Known Issues for SLES 11	147
6.10.3. Known Issues for HDP	147
6.10.4. Known Issues for YARN	149
6.10.5. Known Issues for HBase	150
6.10.6. Known Issues for Phoenix	150
6.10.7. Known Issues for Hive	150
6.10.8. Known Issues for Tez	155
6.10.9. Known Issues for WebHCat	155
6.10.10. Known Issues for Oozie	156
6.10.11. Known Issues for Hue	157
6.10.12. Known Issues for Flume	157
6.10.13. Known Issues for Storm	157
6.10.14. Known Issues for Falcon	158
6.10.15. Known Issues for Knox	160
6.10.16. Known Issues for the Hortonworks Connector for Teradata	160
6.10.17. Known Issues for Ambari	160
6.11. Deprecated Features	161
6.12. Third-party Licenses	161
7. Release Notes HDP-2.1.1	162
7.1. Product Version: HDP-2.1.1	162
7.2. Unsupported Apache components	163
7.3. Behavioral Changes	164
7.3.1. Mahout behavioral changes	164
7.3.2. HDP 2.1 clusters deployed via Ambari	164
7.3.3. Hue behavioral changes	164
7.3.4. HBase behavioral changes	164
7.3.5. Hive behavioral changes	164
7.3.6. Oozie behavioral changes	165
7.4. Patch Information	165
7.4.1. Patch information for Hadoop Common/HDFS	165
7.4.2. Patch information for ZooKeeper	166
7.4.3. Patch information for HBase	166
7.4.4. Patch information for Pig	169

7.4.5. Patch information for Tez	169
7.4.6. Patch information for Hive/HCatalog	170
7.4.7. Patch information for Oozie	172
7.4.8. Patch information for Sqoop	173
7.5. Minimum System Requirements	174
7.5.1. Hardware recommendations	175
7.5.2. Operating systems requirements	175
7.5.3. Software requirements	175
7.5.4. Database requirements	176
7.5.5. Virtualization and cloud platforms	176
7.5.6. Configuring the local repositories	176
7.6. Improvements	176
7.7. Common Vulnerabilities and Exposures	177
7.8. Known Issues	177
7.8.1. Known Limitation for HDFS Upgrade	178
7.8.2. Known Limitation for Oracle DB Metastore	178
7.8.3. Known Issues for SLES 11	178
7.8.4. Known Issues for HDP	178
7.8.5. Known Issues for YARN	179
7.8.6. Known Issues for HBase	180
7.8.7. Known Issues for Phoenix	181
7.8.8. Known Issues for Hive	182
7.8.9. Known Issues for Tez	186
7.8.10. Known Issues for Oozie	186
7.8.11. Known Issues for Hue	188
7.8.12. Known Issues for Flume	188
7.8.13. Known Issues for Storm	189
7.8.14. Known Issues for Falcon	189
7.8.15. Known Issues for Knox	191
7.8.16. Known Issues for the Hortonworks Connector for Teradata	191
7.8.17. Known Issues for Ambari	191
7.9. Deprecated Features	192
7.10. Third-party Licenses	192

List of Tables

- 1.1. Operating Systems mapped to each OS Family 23
- 1.2. Third-party Licenses 25
- 2.1. Operating Systems mapped to each OS Family 48
- 2.2. Third-party Licenses 51
- 3.1. Operating Systems mapped to each OS Family 72
- 3.2. Third-party Licenses 75
- 4.1. Operating Systems mapped to each OS Family 97
- 4.2. Third-party Licenses 99
- 5.1. Operating Systems mapped to each OS Family 119
- 5.2. Third-party Licenses 127
- 6.1. Operating Systems mapped to each OS Family 145
- 6.2. Third-party Licenses 161
- 7.1. Third-party Licenses 192

1. Release Notes HDP-2.1.15

The HDP 2.1.15 Release Notes include the following sections:

- [Product Version: HDP-2.1.15](#)
- [Unsupported Apache Components](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading From HDP 2.1.7 to HDP 2.1.15](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

1.1. Product Version: HDP-2.1.15

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



Note

HDP 2.1 requires Apache Ambari v 1.5.1 or greater.

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

Apache Hadoop 2.4.0	Apache Falcon 0.5.0
Apache HBase 0.98.0	Apache Sqoop 1.4.4
Apache Pig 0.12.1	Apache Knox 0.4.0
Apache Hive 0.13.1	Apache Flume 1.4.0
Apache Tez 0.4.0	Apache Accumulo 1.5.1
Apache ZooKeeper 3.4.5	Apache Phoenix 4.0.0
Hue 2.5.1	Apache Avro 1.7.4

Storm 0.9.1	Apache Mahout 0.9.0
Apache Oozie 4.0.0	

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

1.2. Unsupported Apache Components

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

- Avro v 1.7.4

The following Hue Beeswax feature is shipped as part of HDP 2.1, but not supported:

- Beeswax Visualization, an experimental feature not intended for production use

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

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

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

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

1.3. Fixed in This Release

This release contains the following bug fixes:

Incorrect Results

Component	Apache JIRA	Key	Summary
HBase, Pig	HBASE-10848	BUG-34490	Filter SingleColumnValueFilter combined with

			NullComparator does not work
Hive	8746	BUG-33097	ORC timestamp columns are sensitive to daylight savings time

Potential Data Loss

Component	Apache JIRA	Key	Summary
HDFS	HDFS-7707	BUG-31599	Edit log corruption due to delayed block removal again
HDFS		BUG-34431	NameNode doesn't schedule replication when a DN storage fails (also addressed in BUG-30296)
HDFS	HDFS-7587	BUG-36017	Edit log corruption can happen if append fails with a quota violation
Hue		BUG-32177	Eliminate Django Vulnerabilities in Hue

Stability

Component	Apache JIRA	Key	Summary
Hadoop Common, HDFS	HADOOP-11333	BUG-30296	NameNode doesn't schedule replication when a DN storage fails
HBase		BUG-31416	hql/jsp: The character 0x19 is not valid
HBase	HBASE-12958	BUG-31485	SSH doing hbase:meta get but hbase:meta not assigned
HDFS	HDFS-3443	BUG-30813	NameNode startup in HA mode fails with java.lang.NullPointerException
HDFS	HDFS-7733	BUG-31096	NFS: readdir/readdirplus return null directory attribute on failure
HDFS	HDFS-7045	BUG-31148	Fix NameNode deadlock when opening file under /.reserved path
HDFS		BUG-31410	Simultaneous restart of HA NameNodes and DataNode can cause DataNode to register successfully with only one NameNode.
HDFS	HDFS7009	BUG-32105	Active NN and standby NN have different live nodes
HDFS	HDFS-8127	BUG-34627	NameNode Failover during HA upgrade can cause DataNode to finalize upgrade that will fail datanode -rollback
Hive	HIVE-7147	BUG-30611	ORC PPD should handle CHAR/VARCHAR types
Hive	HIVE-9593	BUG-31212	ORC reader should ignore new/unknown metadata streams

Hive	HIVE-9621	BUG-31335	HiveServer2 http mode - embedded jetty should use SynchronousQueue
Hive	HIVE-9684	BUG-31705	Incorrect disk range computation in ORC because of optional stream kind
Hive	HIVE-10428	BUG-35140	NPE in RegexSerDe using HCatalog
Hue		BUG-33322	Created tables don't available under Hcatalog
HDFS	HDFS-7611	BUG-31672	deleteSnapshot and delete of a file can leave orphaned blocks in the blocksMap on NameNode restart

Upgrade

Component	Apache JIRA	Key	Summary
HDFS	HDFS-8172	BUG-34627	NameNode Failover during HA upgrade can cause DataNode to finalize upgrade that will fail datanode --rollback
Oozie	OOZIE-1615	BUG-30806	Shell action cannot find script file and fails in uber mode

Other

Component	Apache JIRA	Key	Summary
Falcon		BUG-32846	Falcon dependency on inmoabi hadoop-distcp error
Flume		BUG-32860	JMS source setup fails with "org.apache.flume.FlumeException: Could not lookup ConnectionFactory"
HBase	HBASE-12098	BUG-33442	Unable to define permissions via namespaces in Hbase; cannot give users rights table by table.
Hive	HIVE-7175	BUG-31515	Provide password file option to Hive
Hive	HIVE-9831	BUG-32564	HiveServer2 should use ConcurrentHashMap in ThreadFactory
Oozie		BUG-32918	Oozie has a limit of 65535 characters for a particular action
Phoenix		BUG-33633	Unable to load tab delimited data via Phoenix

1.4. Behavioral Changes

The following Apache Components changed in HDP 2.1:

- [What's Changed in Hive](#)

1.4.1. Hive behavioral changes

In HDP 2.1 (Hive 0.13.0) the Decimal data type is now treated as the type Decimal(10,0): 10 digits of precision and 0 scale. This is a change from the variable precision and scale that was available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

In order to avoid unintended "rounding" of decimal data, sites that were previously running Hive 0.11.0 and Hive 0.12.0 may need to migrate tables with Decimal columns after upgrading to Hive 0.13.0. For details, see the [Apache Hive wiki](#). For assistance with upgrades that involve Decimal data, please contact Hortonworks Support.

1.5. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)
- [Patch Information for Sqoop](#)
- [Patch Information for Falcon](#)
- [Patch Information for Knox](#)

1.5.1. Patch information for Hadoop Common/HDFS

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

- [HADOOP-874](#): UT Failures TestNativeAzureFileSystemMocked
- [HADOOP-10342](#): Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken()
- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table

- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10710](#): hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster.
- [HDFS-7560](#): ACLs removed by removeDefaultAcl() will be back after NameNode restart/failover
- [HDFS-7385](#): ThreadLocal used in FSEditLog class causes FSImage permission mess up
- [HDFS-7503](#): Namenode restart after large deletions can cause slow processReport (due to logging)
- [HDFS-7005](#): DFS input streams do not timeout
- [HDFS-7470](#): SecondaryNameNode need twice memory when calling reloadFromImageFile'
- [HDFS-6621](#): Hadoop Balancer prematurely exits iterations
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE_NAME_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN
- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE
- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it

- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#), [HDFS-6618](#), [HDFS-6622](#), and [HDFS-6647](#): Fix potential editlog corruption
- [HDFS-6599](#): Land HDFS-6599 into 2.1-maint
- [HDFS-6616](#): bestNode shouldn't always return the first DataNode
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6979](#): Back-port java code only of HIVE-6979
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible
- [MAPREDUCE-6044](#): [onprem]: hadoopqa user does not have write permissions to /mapred/history causing jobs to fail
- [YARN-1994](#): Support multi-homing for YARN/MR service endpoints

1.5.2. Patch information for ZooKeeper

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

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

1.5.3. Patch information for HBase

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

- [HBASE-12319](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-12533](#): staging directories are not deleted after secure bulk load
- [HBASE-12791](#): HBase does not attempt to clean up an aborted split when the regionserver shutting down
- [HBASE-12467](#): Master joins cluster but never completes initialization
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port

- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluatiotool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up

- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-11118](#): non environment variable solution for ZeroCopyLiteralByteString

Windows Fixes:

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

- [HBASE-10875](#) Metas own location should be cached
- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool
- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#) TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer

- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) Null Pointer Exception in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas

- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

1.5.4. Patch information for Pig

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

- [PIG-4044](#), [PIG-4045](#): Pig job processing avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar
- [PIG-4080](#): Add Preprocessor commands and more to the black/whitelisting feature
- [PIG-3802](#): Fix TestBlackAndWhitelistValidator failures
- [PIG-3765](#): Ability to disable Pig commands and operators
- [PIG-4215](#): Fix unit test failure TestParamSubPreproc and TestMacroExpansion
- [PIG-3765](#): Ability to disable Pig commands and operators - minor refactor
- [PIG-3806](#): PigServer constructor throws NPE after [PIG-3765](#)
- [PIG-3916](#): isEmpty should not be early terminating.
- [PIG-3744](#): SequenceFileLoader does not support BytesWritable
- [PIG-3650](#): Fix for PIG-3100 breaks column pruning
- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

1.5.5. Patch information for Tez

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

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.
- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.
- [TEZ-1066](#): Generate events to integrate with YARN timeline server.

- [TEZ-1048](#): Fix an Null Pointer Exception which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

1.5.6. Patch information for Hive/HCatalog

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

- [HADOOP-10839](#), [HADOOP-10840](#), [HADOOP-7664](#): Need hotfix for [HADOOP-879](#) and [HADOOP-880](#)
- [HIVE-9168](#): Vectorized Coalesce for strings is broken
- [HIVE-6468](#): HS2 & Metastore using SASL out of memory error when curl sends a get request
- [HIVE-6679](#): HiveServer2 should support configurable the server side socket timeout and keepalive for various transports types where applicable
- [HIVE-8371](#): HCatStorer should fail by default when publishing to an existing partition

- [HIVE-5072](#): Enable directly invoke Sqoop job through WebHCat
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6338](#): Hive thrift metastore fails to start with error message .MetaException cannot be cast to RuntimeException
- [HIVE-6521](#): WebHCat cannot fetch correct percentComplete for Hive jobs
- [HIVE-6564](#): WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6602](#): Multi-user HiveServer2 throws error
- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6698](#): hcat.py script does not correctly load the hbase storage handler jars
- [HIVE-6726](#): Hcat cli does not close SessionState
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6931](#): Windows unit test fixes

- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-7006](#): Fix ql_rewrite_gbtoidx.q output file
- [HIVE-7009](#): HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator
- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges
- [HIVE-7062](#): Hive query using SUM() windowing function fails to complete and stays stuck on reduce task
- [HIVE-7065](#), [HIVE-7085](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing
- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7155](#): Need WebHCat fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
- [HIVE-7167](#): Hive Metastore fails to start with SQLServerException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization

- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers
- [HIVE-7210](#): Null Pointer Exception with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

1.5.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence
- [OOZIE-1447](#): Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException
- [OOZIE-615](#): Support Oozie HA.
- [OOZIE-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [OOZIE-1306](#): Bring cron syntax to coordinator frequency
- [OOZIE-1460](#): Implement and document oozie HA security
- [OOZIE-1486](#): cut down on number of small files to track a running action.
- [OOZIE-1491](#): Make sure oozie works with secure ZooKeeper
- [OOZIE-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [OOZIE-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [OOZIE-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname
- [OOZIE-1541](#): typo in oozie HA admin-server command line documentation.
- [OOZIE-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [OOZIE-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [OOZIE-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [OOZIE-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [OOZIE-1576](#): Add documentation for oozie sqoop CLI.
- [OOZIE-1587](#): Add "recovery" column to CoordJob table.

- [OOZIE-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.
- [OOZIE-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [OOZIE-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [OOZIE-1618](#): Dryrun should check variable substitution in workflow.xml.
- [OOZIE-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [OOZIE-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [OOZIE-1726](#): Oozie does not support _HOST when configuring kerberos security.
- [OOZIE-1733](#): Fix test failures by [OOZIE-1722](#).

1.5.8. Patch information for Sqoop

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

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): -password-file option triggers FileSystemClosed exception at end of Oozie action.
- [SQOOP-1260](#): HADOOP_MAPRED_HOME should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with -queries because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling autoCommit on created connections.

- [SQOOP-1246](#): HBaseImportJob should add job authtoken only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding *_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

1.5.9. Patch information for Falcon

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

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

- [FALCON-482](#): Concurrent requests made using Proxy Oozie client fail
- [FALCON-357](#): HCatalog Feed replication: Hive export job fails when table partition contains multiple dated columns

1.5.10. Patch information for Knox

Knox is based on Apache Knox 0.4.0 and includes the following patch:

- [KNOX-242](#): kxox needs to support basedn, search attribute based LDAP authentication

1.5.11. Patch information for Flume

Flume is based on Apache Flume 1.4.0 and includes the following patch:

- [FLUME-1227](#): Introduce some sort of SpillableChannel
- [FLUME-1734](#): Hive Sink based on the new Hive Streaming support (Linux only)
- [FLUME-2226](#): Refactor BlobHandler out of morphline sink and into HTTP source
- [FLUME-2227](#): Move BlobDeserializer from Morphline Sink to flume-ng-core
- [FLUME-2135](#): Add zip to the build distribution for Windows support
- [FLUME-1334](#): Flume start script for Windows
- [FLUME-2148](#): Add flume-env.ps1 for Windows
- [FLUME-1796](#): Upgrade Thrift due to race condition in TThreadSeverPool
- [FLUME-2255](#): Spooling Directory Source cannot handle channel exceptions
- [FLUME-2145](#): TestCheckpointRebuilder.testFastReplay fails on Windows
- [FLUME-2218](#): TestFileChannelIntegrityTool tests failing on Windows
- [FLUME-2219](#) : Windows: Flume tests need to know location of Hadoop native libraries (hadoop.dll)
- [FLUME-2151](#): Update TestExecSource to use native commands on Windows
- [FLUME-2137](#): Fix StagedInstall.java to invoke the correct startup script on Windows
- [FLUME-2058](#): TestFlumeEventQueue in FileChannel fails on Windows
- [FLUME-2068](#): File Channel issue - recovering from BadCheckpoint exception on Windows
- [FLUME-2136](#): Fix intermittent test failure in TestMonitoredCounterGroup on Windows
- [FLUME-2150](#): Fix TestFileChannelEncryption failure on Windows
- [FLUME-2358](#) - File Channel needs to close BackingStore & EventQueue before deleting files in checkpoint directory
- [FLUME-2359](#) - TestFileChannelIntegrityTool throws exception on class teardown on Windows

- [FLUME-2402](#) - Warning seen when overflow is disabled for Spillable Channel Stopping Windows Service leaves the Flume java process running

1.6. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

1.6.1. Hardware recommendations

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

1.6.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1 or SP3
- 64-bit Ubuntu Precise 12.04

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

1.6.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp

- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

1.6.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

1.6.5. Virtualization and cloud platforms

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

For the list of supported operating systems for HDP, see [Operating Systems Requirements](#).

1.6.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

1.7. Upgrading From HDP 2.1.7 to HDP 2.1.15

This section describes how to upgrade an existing HDP 2.1.7 installation to HDP 2.1.15.

If you are upgrading from a previous HDP version, such as HDP 2.0, follow the complete Stack upgrade instructions to HDP 2.1. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.1.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html

- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.15/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.
- Decide to a going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.7 to HDP 2.1.7, do the following:

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

RHEL/CENTOS/ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.15.0/hdp.repo
RHEL/CENTOS/ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.15.0/hdp.repo
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/2.x/updates/2.1.15.0/hdp.repo
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1.15.0/hdp.repo
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.1.15.0/hdp.list
DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.1.15.0/hdp.list

OR Download the HDP RPMs single repository tarball. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.15.0-centos5-tars-tarball.tar.gz
RHEL/CENTOS/ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.15.0-centos6-tars-tarball.tar.gz
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/HDP-2.1.15.0-suse11-tars-tarball.tar.gz
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/sles11sp3/HDP-2.1.15.0-suse11-tars-tarball.tar.gz
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/HDP-2.1.15.0-ubuntu12-tars-tarball.tar.gz
DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/HDP-2.1.15.0-debian6-tars-tarball.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.15 Reference Guide](#).



Note

If you are upgrading an HA NameNode configuration, keep your JournalNodes running while performing this upgrade procedure. Upgrade,

rollback and finalization operations on HA NameNodes must be performed with all JournalNodes running.

3. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/Oracle Linux		
	Upgrade the following components:	<pre>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "webhcat-tar-hive*" "webhcat-tar-pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. Enter:	<pre>yum list installed grep HDP-\$old-stack-version-number</pre>
SLES		
	Upgrade the following components:	<pre>zypper up "collectd*" "epel- release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "hive*" "webhcat-tar-hive*" "webhcat-tar-pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. For example:	<pre>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</pre>
	If components were not upgraded, upgrade them. For example:	<pre>yast --update hadoop hcatalog hive</pre>
Ubuntu or Debian		
	Upgrade the following components:	<pre>apt-get update "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive-hcatalog*" "oozie*" "webhcat-tar-hive*" "webhcat-tar-pig*" "hive*" hdp_mon_nagios_addons</pre>
		<pre>apt-get --force-depends bigtop- jsvcyum install bigtop-jsvc</pre>
	Verify that the components were upgraded. Enter:	<pre>dpkg -s HDP-2 grep HDP-\$old- stack-version-number</pre>

4. Complete the Stack upgrade.

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.7 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/
${$os}/2.x/updates/2.1.15.0 ${$os}
```

where {\$os} is the Operating System Family (OS Family). See the following table:

Table 1.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11 sp1

5. Start all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to start each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.15 Reference Guide](#).

**Note**

Remember to restart Hue as the root user:

```
/etc/init.d/hue restart
```

1.8. Improvements

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

- Support for Debian6
- Support for SLES11 SP3
- Pre-emption – In situations where a queue has a guaranteed level of cluster resources, but must wait to run applications because other queues are utilizing all of the available resources, pre-emption enables higher-priority applications to "bump" lower-priority applications in the queue, and take the resources they need first. For a full description of how pre-emption works, see the [Capacity Scheduler chapter](#) of the Hortonworks System Administration Guide.

1.9. Common Vulnerabilities and Exposures

Fixed in this release:

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

Severity: Critical

Vendor: The Apache Software Foundation

Versions Affected: Tomcat 6.0.37

Users Affected: CVS

Impact: See BUG-21622 and EAR-592

Recommended Action: Tomcat 6.0.37 users should upgrade to Tomcat 6.0.41

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

Severity: Major

Vendor: The Apache Software Foundation

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

Users Affected: CVS

Impact: See BUG-21623

1.10. Known Issues

In this section:

- [Known Issues for Hive and HCat](#)
- [Known Issues for Flume](#)
- [Known Issues for Hortonworks Teradata Connector](#)

1.10.1. Known Issues for Hive

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

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

1.10.2. Known Issues for Flume

- Async HBase Sink is not currently supported due to issues with AsyncHbase library
- [FLUME-2224](#) - Disable File channel dual checkpointing on Windows

1.10.3. Known Issues for the Hortonworks Connector for Teradata

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

1.10.4. Known Issues for Ambari

- See [Ambari Release Notes](#).

1.11. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

1.12. Third-party Licenses

Table 1.2. Third-party Licenses

HDP Component	Subcomponents	License
Phoenix		EPL
Storm	Logback	EPL
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

2. Release Notes HDP-2.1.10

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.10](#)
- [Unsupported Apache Components](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading From HDP 2.1.7 to HDP 2.1.10](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

2.1. Product Version: HDP-2.1.10

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



Note

HDP 2.1 requires Apache Ambari v 1.5.1 or greater.

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

Apache Hadoop 2.4.0	Apache Falcon 0.5.0
Apache HBase 0.98.0	Apache Sqoop 1.4.4
Apache Pig 0.12.1	Apache Knox 0.4.0
Apache Hive 0.13.1	Apache Flume 1.4.0
Apache Tez 0.4.0	Apache Accumulo 1.5.1
Apache ZooKeeper 3.4.5	Apache Phoenix 4.0.0
Hue 2.5.1	Apache Avro 1.7.4
Storm 0.9.1	Apache Mahout 0.9.0
Apache Oozie 4.0.0	

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

2.2. Unsupported Apache components

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

- Avro v 1.7.4

The following Hue Beeswax feature is shipped as part of HDP 2.1, but not supported:

- Beeswax Visualization, an experimental feature not intended for production use

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

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

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

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

2.3. Fixed in This Release

This release contains the following bug fixes:

Component	Apache JIRA	Bug No.	Description
Accumulo	ACCUMULO-3290	BUG-30625	ShellServerTest#listScans failed in 2.1-maint UT
Falcon	OOZIE-1741 FALCON-357	BUG-28213	Falcon needs to use coord:dataInPartitions('input', 'hive-export') instead of coord:dataInPartitionFilter for hive export with multiple partition keys
Falcon	FALCON-482	BUG-24967	Concurrently update/delete a process fails in Falcon
Flume	N/A, Security Fix	BUG-28516	Flume - Address Security Issues in Jetty 6.1.26
HBase	HBASE-12467	BUG-30275	HBase Master is in active state but still its is not handling any client operations

HBase	HBASE-12791	BUG-29998	HBase does not attempt to clean up an aborted SPLIT when the Regions Server is shutting down.
HBase	HBASE-12533	BUG-28135	Staging directories are not deleted after secure bulk load
HBase	HBASE-12319	BUG-30057	Inclusion of HBASE-12319 onto 2.1-maint
HBase		BUG-28682	HBase - Address Security Issues in Jetty 6.1.26
HDFS	HDFS-7470	BUG-30297	Inclusion of HDFS-7470 in HDP-2.1 Maint
HDFS	HDFS-7005	BUG-29743	Backport HDFS-7005 into 2.1-maint
HDFS	HDFS-7503	BUG-29088	NameNode restart after a large deletion could cause slow processReport.
HDFS	HDFS-7385	BUG-28668	Potential incorrect ACLs for clusters that enable Default ACL
HDFS	HDFS-6621	BUG-29911	Hadoop Balancer prematurely exits iterations
HDFS	HDFS-7560	BUG-29479	ACLs removed by removeDefaultAcl() will be back after NameNode restart/failover
Hive	HIVE-8371	BUG-27162	After upgrade from HDP 2.1.3 > 2.1.7 Hive jobs failing with NPE related to PTFs and null values - Tez & MR
Hive	HIVE-6679	BUG-21861	HiveServer2 should support TCP Keepalive & Server Socket Timeout on blocking calls
Hive	HIVE-6468	BUG-21861	Vectorization to_date(timestamp column) returns incorrect values ORC table
Hive	HIVE-6468	BUG-21946	HiveServer2 (tcp mode, with SASL layer) OOMs when getting a non-thrift message
Hive	HIVE-9168	BUG-28168	Vectorization coalesce gives wrong result, or ArrayOutOfIndex error
Hive	HIVE-8732	BUG-27248	ORC string statistics are not merged correctly
Hive	HIVE-7270	BUG-29571	Serde info is not shown in show create table statement, but shows in the desc table
Hive	HIVE-8295	BUG-27301	directSQL with Oracle metastore fails with error ORA-01795, Maximum number of expression can not be more than 1000
Hive	HIVE-7764	BUG-28870	Hive on Tez query output duplicate rows when there is explode in subqueries for joins
Hive	HIVE-8719	BUG-22746	HiveServer2 should support non-kerberos authentication (LDAP or LDAP+kerberos) on a secure cluster.
Hive	HIVE-9359	BUG-24087	load data or insert into partition changes partition location previously defined.
Hive	HIVE-9053	BUG-24531	Hive export OOM error when table is huge. (32TB data, 4800+ partitions)
Hive	HIVE-9359	BUG-29272	2.1 maint port HIVE-9053 (Data Correctness Problem)
Hive	HIVE-7203	BUG-28765	Optimize limit 0 - backport to 2.1-maint and 2.1.7.1 for Teradata request
Hive	HIVE-7943	BUG-28763	hive.security.authorization.createtable.owner.grants is ineffective with Default Authorization
Hive	HIVE-7473	BUG-25815	getPrimitiveJavaObject() throws NPE with NULL decimal values
Hive	HIVE-8560	BUG-25896	3rd party SerDes do not get table properties during initialize()

Hue	N/A	BUG-26462	Remove Visualizations from Hue/Hive query results
Hue	N/A	BUG-28340	Security scan – Hue Web application is vulnerable for XML External Entity Processing (XXE) attacks in the Oozie Workflow
Hue	N/A	BUG-28341	Security scan – Horizontal escalation (Can view others jobs through Hue interface)
Hue	N/A	BUG-24916	Hue job browser throws an error when clicking view all tasks for a job that has many tasks
Knox	KNOX-459	BUG-26393	Knox user hitting the ulimit and return 500 error code
Knox	KNOX-464	BUG-28352	Knox not returning proper redirection url when executing file creation/upload operation via load balancer.
Knox	KNOX-464	BUG-27211	Knox not returning proper redirection url when executing file creation/upload operation via load balancer.
Knox		BUG-25774	Knox vulnerable to Poodle Attack - Baikal Backport
Oozie	OOZIE-2034	BUG-26575	Oozie vulnerable to Poodle Attack
Pig	PIG-4080 PIG-3802 PIG-3765 PIG-4215 PIG-3765 PIG-3806	BUG-28642	Backport pig.blacklist feature to 2.1-maint
Ranger	N/A	BUG-30379	XA Secure does not allow to modify existing policies, gives, error "Policy already exists"
Ranger	N/A	BUG-27540	When Hive server is configured as doAs=false, then XA gets "hive" as the user for all queries
Ranger	N/A	BUG-29537	Disable EclipseLink logs in policy admin
Ranger	N/A	BUG-29177	security loophole: jdbc client can do hdfs commands with 'hive' user privileges when Hiveserver2 doAs=false in xasecure setup
Ranger	N/A	BUG-30060	Change the default audit-mode for ranger plugins to async for better audit log performance
Ranger	N/A	BUG-28606	Hive Component in Ranger doesn't completely reconnect to HiveServer2 after HS2 is restarted in the background
Ranger	N/A	BUG-28597	When you install uxugsync for Ranger v (3.5.001) the symlink /usr/lib/uxugsync is not created.
Tez	TEZ-1342	BUG-28767	tez.am.client.am.port-range not taking effect - backport to 2.1 per request
YARN	CVE-2014-3627	BUG-29634	Security fix for CVE-2014-3627 in HDP 2.1
YARN		BUG-28771	Jobs submitted that require 2 mappers and 32 reducers to a 16-node cluster causing a "deadlock" situation

2.4. Behavioral Changes

The following Apache Components changed in HDP 2.1:

- [What's Changed in Hive](#)

2.4.1. Hive behavioral changes

In HDP 2.1 (Hive 0.13.0) the Decimal data type is now treated as the type Decimal(10,0): 10 digits of precision and 0 scale. This is a change from the variable precision and scale that was available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

In order to avoid unintended "rounding" of decimal data, sites that were previously running Hive 0.11.0 and Hive 0.12.0 may need to migrate tables with Decimal columns after upgrading to Hive 0.13.0. For details, see the [Apache Hive wiki](#). For assistance with upgrades that involve Decimal data, please contact Hortonworks Support.

2.5. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)
- [Patch Information for Sqoop](#)
- [Patch Information for Falcon](#)
- [Patch Information for Knox](#)

2.5.1. Patch information for Hadoop Common/HDFS

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

- [HADOOP-874](#): UT Failures TestNativeAzureFileSystemMocked
- [HADOOP-10342](#): Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken()
- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager

- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10710](#): hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster.
- [HDFS-7560](#): ACLs removed by removeDefaultAcl() will be back after NameNode restart/failover
- [HDFS-7385](#): ThreadLocal used in FSEditLog class causes FSImage permission mess up
- [HDFS-7503](#): Namenode restart after large deletions can cause slow processReport (due to logging)
- [HDFS-7005](#): DFS input streams do not timeout
- [HDFS-7470](#): SecondaryNameNode need twice memory when calling reloadFromImageFile'
- [HDFS-6621](#): Hadoop Balancer prematurely exits iterations
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE_NAME_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN
- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE

- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#), [HDFS-6618](#), [HDFS-6622](#), and [HDFS-6647](#): Fix potential editlog corruption
- [HDFS-6599](#): Land HDFS-6599 into 2.1-maint
- [HDFS-6616](#): bestNode shouldn't always return the first DataNode
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6979](#): Back-port java code only of HIVE-6979
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible
- [MAPREDUCE-6044](#): [onprem]: hadoopqa user does not have write permissions to /mapred/history causing jobs to fail
- [YARN-1994](#): Support multi-homing for YARN/MR service endpoints

2.5.2. Patch information for ZooKeeper

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

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

2.5.3. Patch information for HBase

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

- [HBASE-12319](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-12533](#): staging directories are not deleted after secure bulk load
- [HBASE-12791](#): HBase does not attempt to clean up an aborted split when the regionserver shutting down

- [HBASE-12467](#): Master joins cluster but never completes initialization
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluationtool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state

- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-11118](#): non environment variable solution for ZeroCopyLiteralByteString

Windows Fixes:

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

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

- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HCK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) Null Pointer Exception in MetaCache.clearCache()

- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

2.5.4. Patch information for Pig

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

- [PIG-4044](#), [PIG-4045](#): Pig job processing avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar
- [PIG-4080](#): Add Preprocessor commands and more to the black/whitelisting feature
- [PIG-3802](#): Fix TestBlackAndWhitelistValidator failures
- [PIG-3765](#): Ability to disable Pig commands and operators
- [PIG-4215](#): Fix unit test failure TestParamSubPreproc and TestMacroExpansion
- [PIG-3765](#): Ability to disable Pig commands and operators - minor refactor
- [PIG-3806](#): PigServer constructor throws NPE after [PIG-3765](#)
- [PIG-3916](#): isEmpty should not be early terminating.
- [PIG-3744](#): SequenceFileLoader does not support BytesWritable
- [PIG-3650](#): Fix for PIG-3100 breaks column pruning
- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

2.5.5. Patch information for Tez

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

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.
- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.
- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an Null Pointer Exception which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

2.5.6. Patch information for Hive/HCatalog

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

- [HADOOP-10839](#), [HADOOP-10840](#), [HADOOP-7664](#): Need hotfix for [HADOOP-879](#) and [HADOOP-880](#)
- [HIVE-9168](#): Vectorized Coalesce for strings is broken
- [HIVE-6468](#): HS2 & Metastore using SASL out of memory error when curl sends a get request
- [HIVE-6679](#): HiveServer2 should support configurable the server side socket timeout and keepalive for various transports types where applicable
- [HIVE-8371](#): HCatStorer should fail by default when publishing to an existing partition
- [HIVE-5072](#): Enable directly invoke Sqoop job through WebHCat
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6338](#): Hive thrift metastore fails to start with error message .MetaException cannot be cast to RuntimeException
- [HIVE-6521](#): WebHCat cannot fetch correct percentComplete for Hive jobs
- [HIVE-6564](#): WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6602](#): Multi-user HiveServer2 throws error
- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6698](#): hcat.py script does not correctly load the hbase storage handler jars
- [HIVE-6726](#): Hcat cli does not close SessionState
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI

- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6931](#): Windows unit test fixes
- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-7006](#): Fix ql_rewrite_gbtoidx.q output file
- [HIVE-7009](#): HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator
- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges
- [HIVE-7062](#): Hive query using SUM() windowing function fails to complete and stays stuck on reduce task
- [HIVE-7065](#), [HIVE-7085](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing

- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7155](#): Need WebHCat fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
- [HIVE-7167](#): Hive Metastore fails to start with SQLException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization
- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers
- [HIVE-7210](#): Null Pointer Exception with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

2.5.7. Patch information for Oozie

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

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

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

2.5.8. Patch information for Sqoop

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

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

- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): `-password-file` option triggers `FileSystemClosed` exception at end of Oozie action.
- [SQOOP-1260](#): `HADOOP_MAPRED_HOME` should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with `-queries` because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling `autoCommit` on created connections.
- [SQOOP-1246](#): `HBaseImportJob` should add job `authtoken` only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method `Configuration#unset` is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against `commons-io` higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option `"-skip-dist-cache"` to allow Sqoop not copying jars in `%SQOOP_HOME%\lib` folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the `-bulk-load-dir` option to support the HBase `doBulkLoad` function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding `*_HOME` when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.

- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying `-hive-import` was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

2.5.9. Patch information for Falcon

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

- [FALCON-598](#): `org.apache.falcon.entity.ProcessHelper` throws `NullPointerException` if the process has no inputs OR no outputs defined.
- [FALCON-482](#): Concurrent requests made using Proxy Oozie client fail
- [FALCON-357](#): HCatalog Feed replication: Hive export job fails when table partition contains multiple dated columns

2.5.10. Patch information for Knox

Knox is based on Apache Knox 0.4.0 and includes the following patch:

- [KNOX-242](#): kinox needs to support basedn, search attribute based LDAP authentication

2.5.11. Patch information for Flume

Flume is based on Apache Flume 1.4.0 and includes the following patch:

- [FLUME-1227](#): Introduce some sort of SpillableChannel
- [FLUME-1734](#): Hive Sink based on the new Hive Streaming support (Linux only)
- [FLUME-2226](#): Refactor BlobHandler out of morphline sink and into HTTP source
- [FLUME-2227](#): Move BlobDeserializer from Morphline Sink to flume-ng-core
- [FLUME-2135](#): Add zip to the build distribution for Windows support
- [FLUME-1334](#): Flume start script for Windows

- [FLUME-2148](#): Add flume-env.ps1 for Windows
- [FLUME-1796](#): Upgrade Thrift due to race condition in TThreadSeverPool
- [FLUME-2255](#): Spooling Directory Source cannot handle channel exceptions
- [FLUME-2145](#): TestCheckpointRebuilder.testFastReplay fails on Windows
- [FLUME-2218](#): TestFileChannelIntegrityTool tests failing on Windows
- [FLUME-2219](#) : Windows: Flume tests need to know location of Hadoop native libraries (hadoop.dll)
- [FLUME-2151](#): Update TestExecSource to use native commands on Windows
- [FLUME-2137](#): Fix StagedInstall.java to invoke the correct startup script on Windows
- [FLUME-2058](#): TestFlumeEventQueue in FileChannel fails on Windows
- [FLUME-2068](#): File Channel issue - recovering from BadCheckpoint exception on Windows
- [FLUME-2136](#): Fix intermittent test failure in TestMonitoredCounterGroup on Windows
- [FLUME-2150](#): Fix TestFileChannelEncryption failure on Windows
- [FLUME-2358](#) - File Channel needs to close BackingStore & EventQueue before deleting files in checkpoint directory
- [FLUME-2359](#) - TestFileChannelIntegrityTool throws exception on class teardown on Windows
- [FLUME-2402](#) - Warning seen when overflow is disabled for Spillable Channel Stopping Windows Service leaves the Flume java process running

2.6. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

2.6.1. Hardware recommendations

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

2.6.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1 or SP3
- 64-bit Ubuntu Precise 12.04

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

2.6.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

2.6.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

2.6.5. Virtualization and cloud platforms

HDP is certified and supported when running on virtual or cloud platforms (for example, VMware vSphere or Amazon Web Services EC2) as long as the respective guest OS is

supported by HDP and any issues that are detected on these platforms are reproducible on the same supported OS installed on bare metal.

For the list of supported operating systems for HDP, see [Operating Systems Requirements](#) Improve.

2.6.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

2.7. Upgrading From HDP 2.1.7 to HDP 2.1.10

This section describes how to upgrade an existing HDP 2.1.7 installation to HDP 2.1.10.

If you are upgrading from a previous HDP version, such as HDP 2.0, follow the complete Stack upgrade instructions to HDP 2.1. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.1.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html
- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.15/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.
- Decide to a going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.7 to HDP 2.1.7, do the following:

1. Download the appropriate `hdp.repo` file for your OS:

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.10.0/hdp.repo
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.10.0/hdp.repo
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/2.x/updates/2.1.10.0/hdp.repo
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1.10.0/hdp.repo
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.1.10.0/hdp.list

DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.1.10.0/hdp.list
----------	---

OR Download the HDP RPMs single repository tarball. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.10.0-centos5-tars-tarball.tar.gz
RHEL/CENTOS/ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.10.0-centos6-tars-tarball.tar.gz
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/HDP-2.1.10.0-suse11-tars-tarball.tar.gz
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/sles11sp3/HDP-2.1.10.0-suse11-tars-tarball.tar.gz
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/HDP-2.1.10.0-ubuntu12-tars-tarball.tar.gz
DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/HDP-2.1.10.0-debian6-tars-tarball.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.10 Reference Guide](#).



Note

If you are upgrading an HA NameNode configuration, keep your JournalNodes running while performing this upgrade procedure. Upgrade, rollback and finalization operations on HA NameNodes must be performed with all JournalNodes running.

3. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux		
	Upgrade the following components:	<pre>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "webhcat-tar-hive*" "webhcat-tar- pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. Enter:	<pre>yum list installed grep HDP-$\\$old-stack-version-number$</pre>
SLES		
	Upgrade the following components:	<pre>zypper up "collectd*" "epel- release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "hive*" "webhcat-tar-hive*" "webhcat-tar- pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. For example:	<pre>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</pre>

Operating System	Instructions	Commands
	If components were not upgraded, upgrade them. For example:	<code>yast --update hadoop hcatalog hive</code>
Ubuntu or Debian		
	Upgrade the following components:	<code>apt-get update "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive-hcatalog*" "oozie*" "webhcat-tar-hive*" "webhcat-tar-pig" "hive*" hdp_mon_nagios_addons</code>
		<code>apt-get --force-depends bigtop-jsvcyum install bigtop-jsvc</code>
	Verify that the components were upgraded. Enter:	<code>dpkg -s HDP-2 grep HDP-\$old-stack-version-number</code>

4. Complete the Stack upgrade.

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.7 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/{$os}/2.x/updates/2.1.10.0 {$os}
```

where {\$os} is the Operating System Family (OS Family). See the following table:

Table 2.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11 sp1

5. Start all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to start each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.7 Reference Guide](#).



Note

Remember to restart Hue as the root user:

```
/etc/init.d/hue restart
```

2.8. Improvements

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

- Support for Debian6

- Support for SLES11 SP3
- Pre-emption – In situations where a queue has a guaranteed level of cluster resources, but must wait to run applications because other queues are utilizing all of the available resources, pre-emption enables higher-priority applications to "bump" lower-priority applications in the queue, and take the resources they need first. For a full description of how pre-emption works, see the [Capacity Scheduler chapter](#) of the Hortonworks System Administration Guide.

2.9. Common Vulnerabilities and Exposures

Fixed in this release:

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

Severity: Critical

Vendor: The Apache Software Foundation

Versions Affected: Tomcat 6.0.37

Users Affected: CVS

Impact: See BUG-21622 and EAR-592

Recommended Action: Tomcat 6.0.37 users should upgrade to Tomcat 6.0.41

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

Severity: Major

Vendor: The Apache Software Foundation

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

Users Affected: CVS

Impact: See BUG-21623

2.10. Known Issues

In this section:

- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for Oozie](#)
- [Known Issues for Flume](#)

- [Known Issues for Hortonworks Teradata Connector](#)

2.10.1. Known Issues for HBase

- **BUG-17850:** HBase Tests Intermittent Fail Due to Empty Region Qualifier Error
Problem: Region replicas are not always deleted/closed as expected, causing HBase tests to fail.
- **BUG-16257 (HBASE-10123):** Hbase master fails to start due to BindException
Problem: Apache defaults clash with the range LINUX assigns itself for creating come-and-go ephemeral ports.
- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.
Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the master logs.

2.10.2. Known Issues for Hive

- **BUG-17247:** In Hive Cli switching the hive.execution.engine from Tez to MapReduce does not also switch the YARN framework back to MapReduce
Problem: If we can't switch the YARN framework back to MR, Hive MR will still run on Tez.
- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.
Problem: The Hive session should shut down cleanly and not kill the app.
- **BUG-16393:** Bucketized Table feature fails in some cases.
Problem: Bucketized Table feature fails in some cases. If the source and destination are bucketed on the same key, and if the actual data in the source is not bucketed (because the data got loaded using LOAD DATA LOCAL INPATH) then the data won't be bucketed while writing to the destination.
- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.

2.10.3. Known Issues for Tez

- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.
Problem: The Hive session should shut down cleanly and not kill the app.

2.10.4. Known Issues for Oozie

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

2.10.5. Known Issues for Flume

- Async HBase Sink is not currently supported due to issues with AsyncHbase library
- [FLUME-2224](#) - Disable File channel dual checkpointing on Windows

2.10.6. Known Issues for the Hortonworks Connector for Teradata

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

2.10.7. Known Issues for Ambari

- See [Ambari Release Notes](#).

2.11. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

2.12. Third-party Licenses

Table 2.2. Third-party Licenses

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

3. Release Notes HDP-2.1.7

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.7](#)
- [Unsupported Apache Components](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading From HDP 2.1.5 to HDP 2.1.7](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

3.1. Product Version: HDP-2.1.7

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



Note

HDP 2.1 requires Apache Ambari v 1.5.1 or greater.

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

Apache Hadoop 2.4.0	Apache Falcon 0.5.0
Apache HBase 0.98.0	Apache Sqoop 1.4.4
Apache Pig 0.12.1	Apache Knox 0.4.0
Apache Hive 0.13.1	Apache Flume 1.4.0
Apache Tez 0.4.0	Apache Accumulo 1.5.1
Apache ZooKeeper 3.4.5	Apache Phoenix 4.0.0
Hue 2.5.1	Apache Avro 1.7.4

Storm 0.9.1	Apache Mahout 0.9.0
Apache Oozie 4.0.0	

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

3.2. Unsupported Apache components

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

- Avro v 1.7.4

The following Hue Beeswax feature is shipped as part of HDP 2.1, but not supported:

- Beeswax Visualization, an experimental feature not intended for production use

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

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

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

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

3.3. Fixed in This Release

This release contains the following bug fixes:

- BUG-23914 [[FALCON-357](#)] Falcon needs to use coord:dataInPartitions('input', 'hive-export') instead of coord:dataInPartitionFilter for Hive export with multiple partition keys
- BUG-24084 [[FALCON-357](#)] Falcon job fails with error: line 1:101 missing) at 'to' near '<EOF>' org.apache.hadoop.hive.ql.parse.ParseException: line 1:38 missing (at 'fraud_src_sys_name' near '<EOF>' line 1:101 missing) at 'to' near '<EOF>'

- BUG-24188 [[FALCON-357](#)] Falcon Hive import job fails with FAILED: ParseException line 2:49 missing (at 'fraud_src_sys_name' near "2014-10-02" in drop partition statement line 2:111 mismatched input '<EOF>' expecting) near "2014-10-02" in drop partition statement
- BUG-25100 [[FALCON-753](#)] When you try to delete a process there is a permissions denied error delaying the logs in the staging directory
- BUG-24053 [[FLUME-2397](#)] Flume v1.4.0.2.1.3.0-563 ships async-1.3.1.jar and asynchbase-1.4.1.jar which are not compatible with HBase v0.98.0.2.1.3.0-563-hadoop2
- BUG-24483 [[KNOX-439](#)] Knox fails with Null Pointer Exception when port is NOT specified
- BUG-22500 [[HIVE-6799](#)] Hiveserver2 kerberos http mode JDBC client fails with GSS exception when user principal not in user@LOCAL_HADOOP_REALM format
- BUG-24465 [[HDFS-6914](#)] (2.1) OIV tool at HDP 2.1.5 requires high heap to process fsimage.
- BUG-24441 [[HIVE-7027](#)] Hive query view using "lateral view explode" function on the base table. When you query on the view with a column that is in one of the struct field, it throws error "cannot find field error"
- BUG-19411 [[HIVE-7282](#)] Null map values in ORC table gives PIG error ERROR 6018: Error converting read value to tuple
- BUG-22738 [[HIVE-7353](#)] HiveServer2 leaks JDOPersistenceManager. While running multiple queries through Beeline, many JDOPersistenceManager objects get accumulated (and as reported in hive-user list, eventually causing an OOM). They linger on even after the parent session is closed. This happens specifically for describe/show/load queries.
- BUG-23976 [[HIVE-7426](#)] Backport [HIVE-7426](#) to HDP 2.1 to fix when a query returns the error "org.apache.hadoop.io.IntWritable cannot be cast to org.apache.hadoop.io.Text "
- BUG-21301 [[HIVE-7846](#)] When user created Hive security role, case is all converted to lower case.
- BUG-23818 [[HIVE-7971](#)] Not able to update decimal type on partitioned data in Hive 13
- BUG-22935 [[HIVE-8151](#)] Insert overwrite table dynamic partitioning generates wrong data when query has GROUP BY clause on partition column with hive.optimize.sort.dynamic.partition=true
- BUG-23850 [[HIVE-8298](#)] UDFToDouble cast called for int type data field in join queries, or predicate filtered wrong on multi-join queries, resulting in wrong results
- BUG-24407 [[HIVE-8386](#)] HCAT API call is case sensitive on fields in Array column, causing Falcon failure.
- BUG-24872 [[HIVE-8434](#)] Hive vectorization group by on date type partition of ORC table returns NULL for date type column value
- BUG-24203 [[HIVE-8475](#)] Query fails with cannot find index table when table name is prefixed with database.

- BUG-24409 [HUE] Coordinators actions do not appear because of [OOZIE-1481](#) which changed the Oozie coordinator API, need to add len=-1 parameter in order to obtain actions.
- BUG-24753 [HUE] After creating a workflow in Hue with FS action and running the workflow - receiving the IllegalArgumentException: Pathname is not a valid DFS filename.

BUG-21783 [HUE] Error when deleting to trash since other user cached in thread

BUG-24587 [HUE] User admin page on hue throwing Error 500 due to single left quote (u'\u2018') as unicode.
- BUG-23499 [[PIG-4195](#)] Add VARCHAR data type to be supported in Pig
- BUG-23687 [[OOZIE-2047](#)] Oozie coordinator job got killed with below error, when the input hive table has a column type timestamp.
- BUG-23842 [[OOZIE-1741](#)] Oozie job fails with error: FAILED: ParseException line 1:65 mismatched input 'AND' expecting) near "Zeus" in export statement
- BUG-24926 [[OOZIE-1925](#)] Oozie gets Error: AUTHENTICATION : Could not authenticate, Authentication failed, status: 404, message: Not Found
- BUG-21623 [[CVE-2014-0075](#), [CVE-2014-0096](#), [CVE-2014-0099](#), [CVE-2014-0119](#)] Tomcat Security Vulnerability in HttpFS

3.4. Behavioral Changes

The following Apache Components changed in HDP 2.1:

- [What's Changed in Hive](#)

3.4.1. Hive behavioral changes

In HDP 2.1 (Hive 0.13.0) the Decimal data type is now treated as the type Decimal(10,0): 10 digits of precision and 0 scale. This is a change from the variable precision and scale that was available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

In order to avoid unintended "rounding" of decimal data, sites that were previously running Hive 0.11.0 and Hive 0.12.0 may need to migrate tables with Decimal columns after upgrading to Hive 0.13.0. For details, see the [Apache Hive wiki](#). For assistance with upgrades that involve Decimal data, please contact Hortonworks Support.

3.5. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)

- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)
- [Patch Information for Sqoop](#)
- [Patch Information for Falcon](#)
- [Patch Information for Knox](#)

3.5.1. Patch information for Hadoop Common/HDFS

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

- [HADOOP-874](#): UT Failures TestNativeAzureFileSystemMocked
- [HADOOP-10342](#): Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken()
- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10710](#): hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster.
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE_NAME_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts

- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN
- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE
- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#), [HDFS-6618](#), [HDFS-6622](#), and [HDFS-6647](#): Fix potential editlog corruption
- [HDFS-6599](#): Land HDFS-6599 into 2.1-maint
- [HDFS-6616](#): bestNode shouldn't always return the first DataNode
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6979](#): Back-port java code only of HIVE-6979
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible
- [MAPREDUCE-6044](#): [onprem]: hadoopqa user does not have write permissions to /mapred/history causing jobs to fail
- [YARN-1994](#): Support multi-homing for YARN/MR service endpoints

3.5.2. Patch information for ZooKeeper

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

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

3.5.3. Patch information for HBase

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

- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluationtool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsock#connect() should use new connection
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-11118](#): non environment variable solution for ZeroCopyLiteralByteString

Windows Fixes:

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

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

- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on [HBASE-10070](#)
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HCK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation

- [HBASE-10517](#) Null Pointer Exception in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

3.5.4. Patch information for Pig

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

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

3.5.5. Patch information for Tez

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

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.
- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.

- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an Null Pointer Exception which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

3.5.6. Patch information for Hive/HCatalog

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

- [HADOOP-10839](#), [HADOOP-10840](#), [HADOOP-7664](#): Need hotfix for HADOOP-879 and HADOOP-880
- [HIVE-5072](#): Enable directly invoke Sqoop job through WebHCat
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6338](#): Hive thrift metastore fails to start with error message .MetaException cannot be cast to RuntimeException

- [HIVE-6521](#): WebHCat cannot fetch correct percentComplete for Hive jobs
- [HIVE-6564](#): WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6602](#): Multi-user HiveServer2 throws error
- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6698](#): hcat.py script does not correctly load the hbase storage handler jars
- [HIVE-6726](#): Hcat cli does not close SessionState
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6931](#): Windows unit test fixes
- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2

- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-7006](#): Fix ql_rewrite_gbtoidx.q output file
- [HIVE-7009](#): HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator
- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges
- [HIVE-7062](#): Hive query using SUM() windowing function fails to complete and stays stuck on reduce task
- [HIVE-7065](#), [HIVE-7085](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing
- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7155](#): Need WebHCat fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
- [HIVE-7167](#): Hive Metastore fails to start with SQLServerException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization
- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers

- [HIVE-7210](#): Null Pointer Exception with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

3.5.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence
- [OOZIE-1447](#): Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException
- [OOZIE-615](#): Support Oozie HA.
- [OOZIE-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [OOZIE-1306](#): Bring cron syntax to coordinator frequency
- [OOZIE-1460](#): Implement and document oozie HA security
- [OOZIE-1486](#): cut down on number of small files to track a running action.
- [OOZIE-1491](#): Make sure oozie works with secure ZooKeeper
- [OOZIE-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [OOZIE-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [OOZIE-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname
- [OOZIE-1541](#): typo in oozie HA admin-server command line documentation.
- [OOZIE-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [OOZIE-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [OOZIE-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [OOZIE-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [OOZIE-1576](#): Add documentation for oozie sqoop CLI.
- [OOZIE-1587](#): Add "recovery" column to CoordJob table.
- [OOZIE-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.

- [OOZIE-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [OOZIE-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [OOZIE-1618](#): Dryrun should check variable substitution in workflow.xml.
- [OOZIE-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [OOZIE-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [OOZIE-1726](#): Oozie does not support _HOST when configuring kerberos security.
- [OOZIE-1733](#): Fix test failures by [OOZIE-1722](#).

3.5.8. Patch information for Sqoop

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

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): -password-file option triggers FileSystemClosed exception at end of Oozie action.
- [SQOOP-1260](#): HADOOP_MAPRED_HOME should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with -queries because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling autoCommit on created connections.

- [SQOOP-1246](#): HBaseImportJob should add job authtoken only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding *_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

3.5.9. Patch information for Falcon

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

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

3.5.10. Patch information for Knox

Knox is based on Apache Knox 0.4.0 and includes the following patch:

- [KNOX-242](#): Knox needs to support basedn, search attribute based LDAP authentication

3.5.11. Patch information for Flume

Flume is based on Apache Flume 1.4.0 and includes the following patch:

- [FLUME-1227](#): Introduce some sort of SpillableChannel
- [FLUME-1734](#): Hive Sink based on the new Hive Streaming support (Linux only)
- [FLUME-2226](#): Refactor BlobHandler out of morphline sink and into HTTP source
- [FLUME-2227](#): Move BlobDeserializer from Morphline Sink to flume-ng-core
- [FLUME-2135](#): Add zip to the build distribution for Windows support
- [FLUME-1334](#): Flume start script for Windows
- [FLUME-2148](#): Add flume-env.ps1 for Windows
- [FLUME-1796](#): Upgrade Thrift due to race condition in TThreadSeverPool
- [FLUME-2255](#): Spooling Directory Source cannot handle channel exceptions
- [FLUME-2145](#): TestCheckpointRebuilder.testFastReplay fails on Windows
- [FLUME-2218](#): TestFileChannelIntegrityTool tests failing on Windows
- [FLUME-2219](#): Windows: Flume tests need to know location of Hadoop native libraries (hadoop.dll)
- [FLUME-2151](#): Update TestExecSource to use native commands on Windows
- [FLUME-2137](#): Fix StagedInstall.java to invoke the correct startup script on Windows
- [FLUME-2058](#): TestFlumeEventQueue in FileChannel fails on Windows
- [FLUME-2068](#): File Channel issue - recovering from BadCheckpoint exception on Windows
- [FLUME-2136](#): Fix intermittent test failure in TestMonitoredCounterGroup on Windows
- [FLUME-2150](#): Fix TestFileChannelEncryption failure on Windows
- [FLUME-2358](#) - File Channel needs to close BackingStore & EventQueue before deleting files in checkpoint directory
- [FLUME-2359](#) - TestFileChannelIntegrityTool throws exception on class teardown on Windows

- [FLUME-2402](#) - Warning seen when overflow is disabled for Spillable Channel Stopping Windows Service leaves the Flume java process running

3.6. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

3.6.1. Hardware recommendations

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

3.6.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1 or SP3
- 64-bit Ubuntu Precise 12.04

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

3.6.3. Software requirements

On each of your hosts:

- yum
- rpm

- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

3.6.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

3.6.5. Virtualization and cloud platforms

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

For the list of supported operating systems for HDP, see [Operating Systems Requirements](#) Improve.

3.6.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

3.7. Upgrading From HDP 2.1.5 to HDP 2.1.7

This section describes how to upgrade an existing HDP 2.1.5 installation to HDP 2.1.7.

If you are upgrading from a previous HDP version, such as HDP 2.0, follow the complete Stack upgrade instructions to HDP 2.1. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.1.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html
- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.15/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.
- Decide to a going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.5 to HDP 2.1.7, do the following:

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

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.7.0/hdp.repo
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.7.0/hdp.repo
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/2.x/updates/2.1.7.0/hdp.repo
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1.7.0/hdp.repo
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.1.7.0/hdp.list
DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/2.1.7.0/hdp.list

OR Download the HDP RPMs single repository tarball. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.7.0-centos5-tars-tarball.tar.gz
RHEL/CENTOS/ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.7.0-centos6-tars-tarball.tar.gz
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/HDP-2.1.7.0-suse11-tars-tarball.tar.gz
SLES 11 SP3	http://public-repo-1.hortonworks.com/HDP/sles11sp3/HDP-2.1.7.0-suse11-tars-tarball.tar.gz
UBUNTU 12	http://public-repo-1.hortonworks.com/HDP/ubuntu12/HDP-2.1.7.0-ubuntu12-tars-tarball.tar.gz
DEBIAN 6	http://public-repo-1.hortonworks.com/HDP/debian6/HDP-2.1.7.0-debian6-tars-tarball.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.7 Reference Guide](#).



Note

If you are upgrading an HA NameNode configuration, keep your JournalNodes running while performing this upgrade procedure. Upgrade, rollback and finalization operations on HA NameNodes must be performed with all JournalNodes running.

3. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux		
	Upgrade the following components:	<pre>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "webhcat-tar-hive*" "webhcat-tar- pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. Enter:	<pre>yum list installed grep HDP-\$old-stack-version-number</pre>
SLES		
	Upgrade the following components:	<pre>zypper up "collectd*" "epel- release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "oozie*" "hive*" "webhcat-tar-hive*" "webhcat-tar- pig*" "hive*" hdp_mon_nagios_addons</pre>
	Verify that the components were upgraded. For example:	<pre>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</pre>
	If components were not upgraded, upgrade them. For example:	<pre>yast --update hadoop hcatalog hive</pre>
Ubuntu or Debian		
	Upgrade the following components:	<pre>apt-get update "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive- hcatalog*" "oozie*" "webhcat-tar- hive*" "webhcat-tar-pig" "hive*" hdp_mon_nagios_addons</pre>
		<pre>apt-get --force-depends bigtop- jsvcyum install bigtop-jsvc</pre>
	Verify that the components were upgraded. Enter:	<pre>dpkg -s HDP-2 grep HDP-\$old- stack-version-number</pre>

4. Complete the Stack upgrade.

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.7 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/
{ $os }/2.x/updates/2.1.7.0 { $os }
```

where { \$os } is the Operating System Family (OS Family). See the following table:

Table 3.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11 sp1

5. Start all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to start each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.7 Reference Guide](#).



Note

Remember to restart Hue as the root user:

```
/etc/init.d/hue restart
```

3.8. Improvements

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

- Support for Debian6
- Support for SLES11 SP3
- Pre-emption – In situations where a queue has a guaranteed level of cluster resources, but must wait to run applications because other queues are utilizing all of the available resources, pre-emption enables higher-priority applications to "bump" lower-priority applications in the queue, and take the resources they need first. For a full description of how pre-emption works, see the [Capacity Scheduler chapter](#) of the Hortonworks System Administration Guide.

3.9. Common Vulnerabilities and Exposures

Fixed in this release:

- **CVE-2014-0075, -0096, -0099, -0119:** Tomcat Security Vulnerabilities in Oozie
Severity: Critical
Vendor: The Apache Software Foundation
Versions Affected: Tomcat 6.0.37
Users Affected: CVS
Impact: See BUG-21622 and EAR-592
Recommended Action: Tomcat 6.0.37 users should upgrade to Tomcat 6.0.41
- **CVE-2014-0075, -0096, -0099, -0119:** Tomcat Security Vulnerabilities in HttpFS
Severity: Major
Vendor: The Apache Software Foundation
Versions Affected: Tomcat versions prior to 6.0.41, 7.0.54, 8.0.8

Users Affected: CVS

Impact: See BUG-21623

3.10. Known Issues

In this section:

- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for Oozie](#)
- [Known Issues for Flume](#)
- [Known Issues for Hortonworks Teradata Connector](#)

3.10.1. Known Issues for HBase

- **BUG-17850:** HBase Tests Intermittent Fail Due to Empty Region Qualifier Error
Problem: Region replicas are not always deleted/closed as expected, causing HBase tests to fail.
- **BUG-16257 (HBASE-10123):** Hbase master fails to start due to BindException
Problem: Apache defaults clash with the range LINUX assigns itself for creating come-and-go ephemeral ports.
- **BUG-14986:** HBase HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.
Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the master logs.

3.10.2. Known Issues for Hive

- **BUG-17247:** In Hive Cli switching the hive.execution.engine from Tez to MapReduce does not also switch the YARN framework back to MapReduce
Problem: If we can't switch the YARN framework back to MR, Hive MR will still run on Tez.
- **BUG-16802:** Hive on Tez query passes, but the application is in the killed state.
Problem: The Hive session should shut down cleanly and not kill the app.
- **BUG-16393:** Bucketized Table feature fails in some cases.

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

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

3.10.3. Known Issues for Tez

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

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

3.10.4. Known Issues for Oozie

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

3.10.5. Known Issues for Flume

- Async HBase Sink is not currently supported due to issues with AsyncHbase library
- [FLUME-2224](#) - Disable File channel dual checkpointing on Windows

3.10.6. Known Issues for the Hortonworks Connector for Teradata

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

3.10.7. Known Issues for Ambari

- See [Ambari Release Notes](#).

3.11. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

3.12. Third-party Licenses

Table 3.2. Third-party Licenses

HDP Component	Subcomponents	License
Phoenix		EPL
Storm	Logback	EPL
Accumulo	JCommander	JCommander
Falcon	cern.colt*, cern.jet*, cern.clhep	CERN

HDP Component	Subcomponents	License
Knox	ApacheDS, Groovy	ANTLR
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle

4. Release Notes HDP-2.1.5

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.5](#)
- [Unsupported Apache Components](#)
- [Tech Previews in This Release](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum System Requirements](#)
- [Upgrading From HDP 2.1.3 to HDP 2.1.5](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

4.1. Product Version: HDP-2.1.5

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



Note

HDP 2.1 requires Apache Ambari v 1.5.1 or greater.

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

Apache Hadoop 2.4.0	Apache Falcon 0.5.0
Apache HBase 0.98.0	Apache Sqoop 1.4.4
Apache Pig 0.12.1	Apache Knox 0.4.0
Apache Hive 0.13.1	Apache Flume 1.4.0

Apache Tez 0.4.0	Apache Accumulo 1.5.1
Apache ZooKeeper 3.4.5	Apache Phoenix 4.0.0
Hue 2.5.1	Apache Avro 1.7.4
Storm 0.9.1	Apache Mahout 0.9.0
Apache Oozie 4.0.0	

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

4.2. Unsupported Apache components

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

- Avro v 1.7.4

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

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

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

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

4.3. Tech Previews in This Release

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

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

- Falcon – dependencies, viewed in Graph View

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

4.4. Fixed in This Release

This release contains the following bug fixes:

Project	Apache JIRA	Bug No.	Description
Hadoop Common	HADOOP-10342	BUG-19415	Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
Hadoop Common	HADOOP-10710	BUG-20468	hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster
WebHCat	HADOOP-10839 , -10840 , -7664	BUG-20182	Workload main process fails with an "out of memory" error on the head NameNode.
HDFS	HADOOP-874	BUG-21474	UT Failures TestNativeAzureFileSystemMocked
HBase	HADOOP-11118	BUG-19935	non environment variable solution for ZeroCopyLiteralByteString
HDFS	HDFS-6527 , HDFS-6618 , HDFS-6622 , HDFS-6647	BUG-19654	Fix potential editlog corruption
HDFS	HDFS-6604	BUG-20530	Major dependency change in hive causes wrong hadoop configurations to be loaded
HDFS	HDFS-6604	BUG-19944	The short-circuit cache doesn't correctly time out replicas that haven't been used in a while
Hive, Pig, WebHCat	HIVE-6835	BUG-19469	Pig returns only partitioned value when loading Avro partitioned table with HCatLoader()
Hive	HIVE-7062	BUG-17846	Hive query using SUM() windowing function fails to complete and stays stuck on REDUCE task
Hive, Tez	HIVE-7112	BUG-16162	Investigate failure in Hive
WebHCat	HIVE-7155	BUG-21065	Need fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
Pig	PIG-4044 , -4045	BUG-19608	Pig job processing in Avro fails because avro-mapred-h2.jar is not included in pig-withouthadoop.jar
HDFS	HDFS-6340	BUG-20095	Can't finalize upgrade
HDFS	HDFS-6616	BUG-20165	bestNode shouldn't always return the first DataNode
Hive	HIVE-6979	BUG-20189	Back-port Java code of HIVE-6979
MapReduce	MAPREDUCE-4951 and MAPREDUCE-5900	BUG-18039	Container preemption interpreted as task failures and eventually job failures
MapReduce	MAPREDUCE-5956	BUG-19735	MapReduce AM should not use maxAttempts to determine if this is the last retry
MapReduce	MAPREDUCE-6002	BUG-20419	MR task should prevent report error to AM when process is shutting down
Hive	HIVE-7366 , -7368	BUG-19937	Data nucleus returns empty results (table not found) when there is unexpected underlying exceptions
Pig	PIG-3744	BUG-21208	SequenceFileLoader does not support BytesWritable
YARN	YARN-2144	BUG-18468	Add logs when preemption occurs.
YARN	YARN-2181	BUG-19182	Add preemption info to RM Web UI and RM logs.

Project	Apache JIRA	Bug No.	Description
YARN	YARN-2125	BUG-18536	ProportionalCapacityPreemptionPolicy should only log CSV when debug enabled
YARN	YARN-2124	BUG-18535	ProportionalCapacityPreemptionPolicy cannot work because it's initialized before scheduler initialized
YARN	YARN-2022	BUG-19086	Preempting an Application Master container can be kept as least priority when multiple applications are marked for preemption by ProportionalCapacityPreemptionPolicy
YARN	YARN-2074	BUG-17995	Preemption of AM containers shouldn't count towards AM failures
YARN	YARN-1957	BUG-17996	ProportionalCapacitPreemptionPolicy handling of corner cases
YARN	YARN-1408	BUG-18120	Preemption caused Invalid State Event: ACQUIRED at KILLED and caused a task timeout for 30mins
Hue	n/a	BUG-18610	Receiving the message "Processing exception: Could not import about.views. Error was: cannot import name sudo" when trying to access Hue on a fresh install
Hue	n/a	BUG-9326	Error "unexpected keyword argument 'allow_duplicate'" while copying a file or directory in File Browser
Hue	n/a	BUG-19718	Traceback printed during Hue install on SLES
Hue		BUG-19530	User getting errors when trying to refresh and import new members of ldap groups
Hue	n/a	BUG-19762	beeswax server gets _message=' java.lang.OutOfMemoryError: unable to create new native thread', errorCode=0) and query hangs until Hue is restarted
Hue	n/a	BUG-20528	Add <code>force_username_uppercase</code> option to Hue for AD authorization.

4.5. Behavioral Changes

The following Apache Components changed in HDP 2.1:

- [What's Changed in Hive](#)

4.5.1. Hive behavioral changes

In HDP 2.1 (Hive 0.13.0) the Decimal data type is now treated as the type `Decimal(10,0)`: 10 digits of precision and 0 scale. This is a change from the variable precision and scale that was available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

In order to avoid unintended "rounding" of decimal data, sites that were previously running Hive 0.11.0 and Hive 0.12.0 may need to migrate tables with Decimal columns after upgrading to Hive 0.13.0. For details, see the [Apache Hive wiki](#). For assistance with upgrades that involve Decimal data, please contact Hortonworks Support.

4.6. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)

- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)
- [Patch Information for Falcon](#)
- [Patch Information for Knox](#)

4.6.1. Patch information for Hadoop Common/HDFS

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

- [HADOOP-874](#): UT Failures TestNativeAzureFileSystemMocked
- [HADOOP-10342](#): Extend UserGroupInformation to return a UGI given a preauthenticated kerberos Subject
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken()
- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HADOOP-10710](#): hadoop.auth cookie not in RFC2109 format. Oozie web console missing panel in secure cluster.
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE_NAME_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts

- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN
- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE
- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#), [HDFS-6618](#), [HDFS-6622](#), and [HDFS-6647](#): Fix potential editlog corruption
- [HDFS-6599](#): Land HDFS-6599 into 2.1-maint
- [HDFS-6616](#): bestNode shouldn't always return the first DataNode
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6979](#): Back-port java code only of HIVE-6979
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible
- [MAPREDUCE-6044](#): [onprem]: hadoopqa user does not have write permissions to /mapred/history causing jobs to fail
- [YARN-1994](#): Support multi-homing for YARN/MR service endpoints

4.6.2. Patch information for ZooKeeper

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

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

4.6.3. Patch information for HBase

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

- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluatiotool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin

- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-11118](#): non environment variable solution for ZeroCopyLiteralByteString

Windows Fixes:

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

- [HBASE-10875](#) Metas own location should be cached
- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool
- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#) TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky

- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.

- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

4.6.4. Patch information for Pig

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

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

4.6.5. Patch information for Tez

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

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.

- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.
- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

4.6.6. Patch information for Hive/HCatalog

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

- [HADOOP-10839](#), [HADOOP-10840](#), [HADOOP-7664](#): Need hotfix for HADOOP-879 and HADOOP-880
- [HIVE-5072](#): Enable directly invoke Sqoop job through WebHCat
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6338](#): Hive thrift metastore fails to start with error message .MetaException cannot be cast to RuntimeException
- [HIVE-6521](#): WebHCat cannot fetch correct percentComplete for Hive jobs
- [HIVE-6564](#): WebHCat E2E tests that launch MR jobs fail on check job completion timeout
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6602](#): Multi-user HiveServer2 throws error
- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6698](#): hcat.py script does not correctly load the hbase storage handler jars
- [HIVE-6726](#): Hcat cli does not close SessionState
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster

- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6931](#): Windows unit test fixes
- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-7006](#): Fix ql_rewrite_gbtoidx.q output file
- [HIVE-7009](#): HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator
- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges
- [HIVE-7062](#): Hive query using SUM() windowing function fails to complete and stays stuck on reduce task
- [HIVE-7065](#), [HIVE-7085](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing
- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables

- [HIVE-7155](#): Need WebHCat fix to add configuration parameter to override WebHCat configuration to overwrite mapreduce.map.memory.mb for the controller job
- [HIVE-7167](#): Hive Metastore fails to start with SQLServerException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization
- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers
- [HIVE-7210](#): NPE with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

4.6.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence
- [OOZIE-1447](#): Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException
- [OOZIE-615](#): Support Oozie HA.
- [OOZIE-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [OOZIE-1306](#): Bring cron syntax to coordinator frequency
- [OOZIE-1460](#): Implement and document oozie HA security
- [OOZIE-1486](#): cut down on number of small files to track a running action.
- [OOZIE-1491](#): Make sure oozie works with secure ZooKeeper
- [OOZIE-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [OOZIE-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [OOZIE-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname
- [OOZIE-1541](#): typo in oozie HA admin-server command line documentation.
- [OOZIE-1555](#): Launcher mapper to check for system properties before opening files for action data.

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

4.6.8. Patch information for Sqoop

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

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.

- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): `-password-file` option triggers `FileSystemClosed` exception at end of Oozie action.
- [SQOOP-1260](#): `HADOOP_MAPRED_HOME` should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with `-queries` because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling `autoCommit` on created connections.
- [SQOOP-1246](#): `HBaseImportJob` should add job `authToken` only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method `Configuration#unset` is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against `commons-io` higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option `"-skip-dist-cache"` to allow Sqoop not copying jars in `%SQOOP_HOME%\lib` folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the `-bulk-load-dir` option to support the HBase `doBulkLoad` function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding `*_HOME` when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class `HCatHadoopShims` will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.

- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying –hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

4.6.9. Patch information for Falcon

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

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

4.6.10. Patch information for Knox

Knox is based on Apache Knox 0.4.0 and includes the following patch:

- [KNOX-242](#): knox needs to support basedn, search attribute based LDAP authentication

4.7. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

4.7.1. Hardware recommendations

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

4.7.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1 or SP3
- 64-bit Ubuntu Precise 12.04

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

4.7.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

4.7.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

4.7.5. Virtualization and cloud platforms

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

For the list of supported operating systems for HDP, see [Operating Systems Requirements](#) Improve.

4.7.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

4.8. Upgrading From HDP 2.1.3 to HDP 2.1.5

This section describes how to upgrade an existing HDP 2.1.3 installation to HDP 2.1.5.

If you are upgrading from a previous HDP version, such as HDP 2.0, follow the complete Stack upgrade instructions to HDP 2.1. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.1.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html
- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.15/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.
- Decide to a going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.3 to HDP 2.1.5, do the following:

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

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.7.0/hdp.repo
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.7.0/hdp.repo
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/2.x/updates/2.1.7.0/hdp.repo

OR Download the HDP RPMs single repository tarball. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.5.0-centos5-tars-tarball.tar.gz
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.5.0-centos6-tars-tarball.tar.gz
SLES 11 SP1	http://public-repo-1.hortonworks.com/HDP/suse11sp1/HDP-2.1.5.0-suse11-tars-tarball.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.5 Reference Guide](#).

3. Back up Oozie files.

If you have Oozie installed and running, back up the files in the following directories on the Oozie server host, and make sure all files, including `*site.xml` files, are copied:

```
mkdir oozie-conf-bak
cp -R /etc/oozie/conf/* oozie-conf-bak
```

4. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux		
	Upgrade the following components:	<code>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" "oozie*" "webhcat*" hdp_mon_nagios_addons</code>
	Verify that the components were upgraded. Enter:	<code>yum list installed grep HDP-\$old-stack-version-number</code>
SLES		
	Upgrade the following components:	<code>zypper up "collectd*" "epel-release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" "oozie*" "webhcat*" hdp_mon_nagios_addons</code>
		<code>zypper up -r HDP-2.1.5.0</code>
	Verify that the components were upgraded:	<code>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</code>
	If components were not upgraded, upgrade them:	<code>yast --update hadoop hive</code>
Ubuntu or Debian		
	Upgrade the following components:	<code>apt-get update "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive-hcatalog*" "oozie*" "webhcat*" hdp_mon_nagios_addons</code>
		<code>apt-get --force-depends bigtop-jsvcyum install bigtop-jsvc</code>
	Verify that the components were upgraded. Enter:	<code>dpkg -s HDP-2 grep HDP-\$old-stack-version-number</code>

5. Complete the Stack upgrade.

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.5 repositories for HDP and HDP-UTILS. For Ambari 1.6.1 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/
${os}/2.x/updates/2.1.5.0 ${os}
```

where `{os}` is the Operating System Family (OS Family). See the following table:

Table 4.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11 sp1

4.9. Improvements

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

- Support for Debian6
- Support for SLES11 SP3
- Pre-emption – In situations where a queue has a guaranteed level of cluster resources, but must wait to run applications because other queues are utilizing all of the available resources, pre-emption enables higher-priority applications to "bump" lower-priority applications in the queue, and take the resources they need first. For a full description of how pre-emption works, see the [Capacity Scheduler chapter](#) of the Hortonworks System Administration Guide.

4.10. Common Vulnerabilities and Exposures

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

Severity: Critical

Vendor: The Apache Software Foundation

Versions Affected: Tomcat 6.0.37

Users Affected: CVS

Impact: See BUG-21622 and EAR-592

Recommended Action: Tomcat 6.0.37 users should upgrade to Tomcat 6.0.41

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

Severity: Major

Vendor: The Apache Software Foundation

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

Users Affected: CVS

Impact: See BUG-21623

4.11. Known Issues

In this section:

- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for Oozie](#)
- [Known Issues for Hortonworks Teradata Connector](#)

4.11.1. Known Issues for HBase

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

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

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

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

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

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

4.11.2. Known Issues for Hive

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

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

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

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

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

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

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

4.11.3. Known Issues for Tez

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

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

4.11.4. Known Issues for Oozie

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

Problem:

Workaround:

4.11.5. Known Issues for the Hortonworks Connector for Teradata

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

4.11.6. Known Issues for Ambari

- See [Ambari Release Notes](#).

4.12. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

4.13. Third-party Licenses

Table 4.2. Third-party Licenses

HDP Component	Subcomponents	License
Phoenix		EPL
Storm	Logback	EPL
Accumulo	JCommander	JCommander

HDP Component	Subcomponents	License
Falcon	cern.colc* , cern.jet* , cern.clhep	CERN
Knox	ApacheDS, Groovy	ANTLR
Knox	SL4J	MIT
Knox	Jetty and Jerico	EPL
Knox	ApacheDS	Bouncy Castle

5. Release Notes HDP-2.1.3

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.3](#)
- [Tech Previews in This Release](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum system requirements](#)
- [Upgrading From HDP 2.1.2 to HDP 2.1.3](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

5.1. Product Version: HDP-2.1.3

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



Note

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

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

- Apache Hadoop 2.4.0
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.1
- Apache Tez 0.4.0

- Apache ZooKeeper 3.4.5
- Hue 2.3.4
- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5.0
- Apache Sqoop 1.4.4
- Apache Knox 0.4.0
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0

Third party components:

- Ganglia 3.5.0
- Ganglia Web 3.5.7
- Nagios 3.5.0

5.2. Unsupported Apache components

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

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

- NameNode Federation ([HDFS-1052](#))
- viewFS ([HADOOP-7257](#))

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

- MapReduce v1 Fair Scheduler ([HADOOP-3746](#))

YARN Fair Scheduler ([MAPREDUCE-3451](#))

MapReduce Uber AM ([MAPREDUCE-2405](#))

MapReduce Eclipse Plugin (for Non-Kerberos and Kerberos cluster)

Cgroup resource isolation ([YARN-3](#))

CPU Scheduling ([YARN-2](#))

5.3. Tech Previews in This Release

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

- Preemption – this feature has been found to contain defects that make it unsuitable for a production cluster at this time. For information about Preemption, see http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk_system-admin-guide/content/preemption.html.
- Falcon – lag time in displaying lineage information if graph contains many vertices; old graph persists until updated graph is loaded
- Falcon – dependencies, viewed in Graph View

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

5.4. Fixed in This Release

HDP 2.1.3 is a maintenance release that includes scores of bug fixes and optimizations. It is notable for resuming support for Ubuntu 12.0.4.

5.5. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- [What's Changed in YARN](#)

5.5.1. YARN behavioral changes

In HDP 2.1.3, when installing with Ambari 1.6.0, go to the Ambari **Services > YARN > Configs** menus and change the value of the YARN `yarn.timeline-service.store-class` property to `org.apache.hadoop.yarn.server.timeline.LeveldbTimelineStore`.

This change is required to ensure that YARN starts during cluster installation.

5.6. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)

- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)



Note

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

5.6.1. Patch information for Hadoop Common/HDFS

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

- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken().
- [HADOOP-10508](#): RefreshCallQueue fails when authorization is enabled
- [HADOOP-10562](#): Namenode exits on exception without printing stack trace in AbstractDelegationTokenSecretManager
- [HADOOP-10612](#): NFS failed to refresh the user group id mapping table
- [HADOOP-10630](#): Possible race condition in RetryInvocationHandler
- [HDFS-4052](#): BlockManager#invalidateWork should print logs outside the lock
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support FSIMAGE_NAME_OPTIMIZATION.
- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6227](#): ShortCircuitCache#unref should purge ShortCircuitReplicas whose streams have been closed by java interrupts
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.
- [HDFS-6245](#): datanode fails to start with a bad disk even when failed volumes is set
- [HDFS-6273](#): Config options to allow wildcard endpoints for namenode HTTP and HTTPS servers
- [HDFS-6278](#): Create HTML5-based UI for SNN

- [HDFS-6279](#): Create new index page for JN / DN
- [HDFS-6362](#): InvalidateBlocks is inconsistent in usage of DatanodeUuid and StorageID
- [HDFS-6364](#): Incorrect check for unknown datanode in Balancer
- [HDFS-6370](#): Web UI fails to display in intranet under IE
- [HDFS-6411](#): nfs-hdfs-gateway mount raises I/O error and hangs when a unauthorized user attempts to access it
- [HDFS-6423](#): Diskspace quota usage should be updated when appending data to partial block
- [HDFS-6432](#): Add snapshot related APIs to webhdfs
- [HDFS-6438](#): DeleteSnapshot should be a DELETE request in WebHdfs
- [HDFS-6458](#): NFS: stale NFS file handle Error for previous mount point
- [HDFS-6462](#): NFS: fsstat request fails with the secure hdfs
- [HDFS-6527](#): Edit log corruption due to deferred INode removal
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [MAPREDUCE-5014](#): Extending DistCp through a custom CopyListing is not possible

5.6.2. Patch information for ZooKeeper

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

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

5.6.3. Patch information for HBase

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

- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp

- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10592](#): Refactor PerformanceEvaluationTool
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10850](#): essential column family optimization is broken

- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList

Windows Fixes:

- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows

Changes related to HBASE-10070:

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

- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess

- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

5.6.4. Patch information for Pig

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

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

5.6.5. Patch information for Tez

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

- [TEZ-1159](#): Fix handling of corrupt or empty files in recovery data.
- [TEZ-1158](#): Disable multiple AM attempts if recovery is disabled.
- [TEZ-1135](#):: Fix ShuffledUnorderedKVInput handling of empty partitions.
- [TEZ-1125](#): Pre-warm broken.
- [TEZ-1097](#): Tez assumes that the scratch directory has to be same as the default filesystem.
- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.

- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

5.6.6. Patch information for Hive/HCatalog

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

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

- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6792](#): hive.warehouse.subdir.inherit.perms doesn't work correctly in CTAS
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6861](#): more hadoop2 only golden files to fix
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6869](#): Golden file updates for tez tests
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6915](#): Hive HBase queries fail on secure Tez cluster
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6931](#): Windows unit test fixes
- [HIVE-6936](#): Provide table properties to InputFormats
- [HIVE-6946](#): Make it easier to run WebHCat e2e tests
- [HIVE-6947](#): More fixes for tests on hadoop-2
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6967](#): Hive transaction manager fails when SQLServer is used as an RDBMS
- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-7006](#): Fix ql_rewrite_gbtoidx.q output file
- [HIVE-7009](#): HIVE_USER_INSTALL_DIR could not be set to non-HDFS filesystem
- [HIVE-7011](#): HiveInputFormat's split generation isn't thread safe
- [HIVE-7031](#): Utilites.createEmptyFile uses File.Separator instead of Path.Separator to create an empty file in HDFS
- [HIVE-7043](#): When using the tez session pool via hive, once sessions time out, all queries go to the default queue
- [HIVE-7055](#): config not propagating for PTFOperator

- [HIVE-7061](#): sql std auth - insert queries without overwrite should not require delete privileges
- [HIVE-7065](#): Hive jobs in webhcat run in default mr mode even in Hive on Tez setup
- [HIVE-7072](#): HCatLoader only loads first region of hbase table
- [HIVE-7076](#): Plugin (exec hook) to log to application timeline data to Yarn
- [HIVE-7099](#): Add Decimal datatype support for Windowing
- [HIVE-7112](#): Tez processor swallows errors
- [HIVE-7114](#): Extra Tez session is started during HiveServer2 startup
- [HIVE-7118](#): Oracle upgrade schema scripts do not map Java long datatype columns correctly for transaction related tables
- [HIVE-7166](#): Vectorization with UDFs returns incorrect results
- [HIVE-7167](#): Hive Metastore fails to start with SQLServerException
- [HIVE-7188](#): sum(if()) returns wrong results with vectorization
- [HIVE-7190](#): WebHCat launcher task failure can cause two concurrent user jobs to run
- [HIVE-7209](#): allow metastore authorization api calls to be restricted to certain invokers
- [HIVE-7210](#): NPE with "No plan file found" when running Driver instances on multiple threads
- [HIVE-7268](#): On Windows Hive jobs in Webhcat always run on default MR mode

5.6.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [OOZIE-1447](#): Sqoop actions that don't launch a map reduce job fail with an IllegalArgumentException.
- [OOZIE-615](#): Support Oozie HA.
- [OOZIE-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [OOZIE-1306](#): Bring cron syntax to coordinator frequency.
- [OOZIE-1460](#): Implement and document oozie HA security.

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

5.6.8. Patch information for Sqoop

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

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

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

- [SQOOP-1203](#): Add another default case for finding *_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

5.7. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

5.7.1. Hardware recommendations

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

5.7.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1
- 64-bit Ubuntu Precise 12.04

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

5.7.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

5.7.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

5.7.5. Virtualization and cloud platforms

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

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

5.7.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

5.8. Upgrading From HDP 2.1.2 to HDP 2.1.3

This section describes how to upgrade an existing HDP 2.1.2 installation to HDP 2.1.3.

If you are upgrading from a previous HDP version, such as HDP 2.0, please follow the complete Stack upgrade instruction to HDP 2.1. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html
- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.3/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.
- Think about whether you're going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.2 to HDP 2.1.3, do the following:

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

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.3.0/hdp.repo
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.3.0/hdp.repo
SLES 11	http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.3.0/hdp.repo

OR **Download the HDP RPMs single repository tarball.** (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.3.0-centos5-tars-tarball.tar.gz
--------------------------------	---

RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.3.0-centos6-tars-tarball.tar.gz
SLES 11	http://public-repo-1.hortonworks.com/HDP/suse11/HDP-2.1.3.0-suse11-tars-tarball.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.3 Reference Guide](#).



Note

If you are upgrading an HA NameNode configuration, keep your JournalNodes running while performing this upgrade procedure. Upgrade, rollback and finalization operations on HA NameNodes must be performed with all JournalNodes running.

3. Back up Oozie files.

If you have Oozie installed and running, back up the files in the following directories on the Oozie server host, and make sure all files, including `*site.xml` files, are copied:

```
mkdir oozie-conf-bak
cp -R /etc/oozie/conf/* oozie-conf-bak
```

4. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux	Remove WebHCat components:	<code>yum erase "webhcat"</code>
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<code>yum upgrade hive</code> <code>yum erase hcatalog</code> <code>yum install</code> <code>hive-hcatalog</code>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<code>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType <\$databaseType></code> The value for <code>\$databaseType</code> can be <code>derby</code> , <code>mysql</code> , <code>oracle</code> , or <code>postgres</code> .
	Upgrade the following components:	<code>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</code>
		<code>yum install webhcat-tar-hive webhcat-tar-pig</code> <code>yum install "hive-*</code> <code>yum install oozie oozie-client</code> <code>rpm -e --nodeps bigtop-jsvcyum install bigtop-jsvc</code>
	Verify that the components were upgraded. Enter:	<code>yum list installed grep</code> <code>HDP-\$old-stack-version-number</code>
SLES	Remove WebHCat, components:	<code>zypper remove webhcat*</code>

Operating System	Instructions	Commands
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<code>zypper install hive-hcatalog</code>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<code>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType < \$databaseType></code>
	Upgrade the following components:	<code>zypper up "collectd*" "epel-release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</code>
		<code>zypper install webhcat-tar-hive webhcat-tar-pig zypper up -r HDP-2.1.3.0 zypper install hive*zypper install oozie oozie-client</code>
	Verify that the components were upgraded:	<code>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</code>
	If components were not upgraded, upgrade them:	<code>yast --update hadoop hcatalog hive</code>

5. **If you are upgrading from an HA NameNode configuration, restart all JournalNodes.**
On each JournalNode host, enter the following command:

```
su -l {HDFS_USER} -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh start journalnode"
```

6. **Complete the Stack upgrade.**

For HDP 2.1, use the version of Hue shipped with HDP 2.1. If you have a previous version of Hue, follow the instructions for upgrading Hue in [Installing HDP Manually](#).

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.3 repositories for HDP and HDP-UTILS. For Ambari 1.6.0 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/{ $os }/2.x/updates/2.1.3.0 { $os }
```

where { \$os } is the Operating System Family (OS Family). See the following table:

Table 5.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11

7. **Finalize the upgrade.**

If you are not yet ready to discard your backup, you can start the upgraded HDFS without finalizing the upgrade. (At this stage, you can still roll back if need be.)

Verify your filesystem health. When you are ready to commit to this upgrade (are certain that you will not want to roll back), discard your backup and finalize the upgrade.

As \$HDFS_USER, execute the following command:

```
hdfs dfsadmin -finalizeUpgrade
```


5.9. Improvements

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

- HDP 2.1 supports the use of Oracle DB as a metastore.
- The Storm LogViewer service now allows only worker log files from the Storm log directory to be viewed from the UI, regardless of the privileges of the person running the service.

5.10. Common Vulnerabilities and Exposures

- **CVE-2014-0229:** Three HDFS admin commands lack proper privilege checks

Vendor: The Apache Software Foundation

Versions Affected: Apache Hadoop 0.23 prior to 0.23.11, Apache 2.x prior to 2.4.1

Impact: Three HDFS admin commands, refreshNamenodes, deleteBlockPool and shutdownDatanode, lack proper privilege checks in Apache Hadoop 0.23.x prior to 0.23.11, and in 2.x prior to 2.4.1. This allows arbitrary users to create data nodes unnecessarily, refresh its federated NameNode configuration in an untimely manner, delete inactive block pools, or shut itself down.

(The shutdownDatanode command was first introduced in 2.4.0; refreshNamenodes and deleteBlockPool were added in 0.23.0.)

Recommended Action: Hadoop 0.23.x users should upgrade to 0.23.11, Hadoop 2.x users should upgrade to 2.4.1.

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability

Severity: Major

Vendor: The Apache Software Foundation

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

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

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

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

Credit: This issue was discovered by Koji Noguchi of Yahoo

5.11. Known Issues

In this section:

- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Apache Pig](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Hue](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

5.11.1. Known Issues for HDP

- **BUG-19593:** Hive user does not have shell access on Deb

Problem: When attempting to access Deb as the user `hive`, the following displays:

```
hive:x:113:121:Hive User,,,:/var/lib/hive:/bin/false
```

- **BUG-19531:** ZooKeeper user does not have shell access on Deb

Problem: The `/etc/passwd` file has no shell access for ZooKeeper.

- **BUG-16003:** Knox PID directory does not exist on Ubuntu after reboot

Problem: Unable to start Knox after Ubuntu reboot

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

Problem: RHEL 6.2 and RHEL 6.3 contain known bugs in `nslcd`. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by `nslcd` times out.

Workaround: If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.
- Increase the configured `nslcd` search timeout in `/etc/nslcd.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGroupsMapping` in `core-site.xml`.
- **BUG-15360:**HADOOP-10519: In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error

5.11.2. Known Issues for YARN

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

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

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

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

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

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

```
mapred-site.xml
mapreduce.map.memory.mb 1024
mapreduce.map.java.opts -Xmx692m
mapreduce.task.io.sort.mb 200
```

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

5.11.3. Known Issues for HBase

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

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

5.11.4. Known Issues for Apache Pig

- **BUG-19608:** ([PIG-4044](#), [PIG-4045](#)) Pig job processing Avro fails because `avro-mapred-h2.jar` is not included in `pig-withouthadoop.jar`

Problem: If Pig was installed from `pig-withouthadoop.jar` (which is missing `avro-mapred-h2.jar`), trying to run Pig with Avro data returns:

```
Error running child : java.lang.IncompatibleClassChangeError: Found
interface org.apache.hadoop.mapreduce.TaskAttemptContext, but class was
expected
```

5.11.5. Known Issues for Hive

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

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

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

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

- **BUG-17850:** HBCK test fails intermittently due to Empty Region Qualifier error
- **BUG-17846:** Hive query using SUM() windowing function fails to complete and stays stuck on reduce task

Problem: The query never completes.

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

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

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

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

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

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

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

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

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

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

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

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

Problem: HBase on Suse 11 64 bit, Smoke test fails Intermittent with ERROR main client.`ConnectionManager$HConnectionImplementation`: The node `/hbase` is not in ZooKeeper. Basically HBase default ports clash with the range Linux assigns itself for

creating come-and-go ephemeral ports. Therefore, once in awhile we'll see HBase master can't start due to port binding issue.

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

Problem: The error returned on the Hive console is:

```
Here is the error in the Hive console log:
Vertex failed, vertexName=Map 1, vertexId=vertex_1395920136483_7733_1_00,
diagnostics=[Task failed, taskId=task_1395920136483_7733_1_00_000000,
diagnostics=[AttemptID:attempt_1395920136483_7733_1_00_000000_0 Info:Error:
java.io.IOException: java.lang.ClassCastException: org.apache.hadoop.
io.Text cannot be cast to org.apache.hadoop.hive.serde2.columnar.
BytesRefArrayWritable
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.
handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.
handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:344)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:79)
at org.apache.hadoop.hive ql.io.HiveRecordReader.doNext(HiveRecordReader.
java:33)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
next(HiveContextAwareRecordReader.java:122)
at org.apache.hadoop.mapred.split.TezGroupedSplitsInputFormat
$TezGroupedSplitsRecordReader.next(TezGroupedSplitsInputFormat.java:122)
at org.apache.tez.mapreduce.input.MRInput$MRInputKVReader.next(MRInput.
java:510)
at org.apache.hadoop.hive ql.exec.tez.MapRecordProcessor.
run(MapRecordProcessor.java:158)
at org.apache.hadoop.hive ql.exec.tez.TezProcessor.run(TezProcessor.
java:160)
at org.apache.tez.runtime.LogicalIOProcessorRuntimeTask.
run(LogicalIOProcessorRuntimeTask.java:306)
at org.apache.hadoop.mapred.YarnTezDagChild$4.run(YarnTezDagChild.java:549)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:396)
at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more
```

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

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

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

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

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

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

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

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

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

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-13551:** Oozie does not understand _HOST in the kerberos principal name

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

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

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

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

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

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

5.11.6. Known Issues for Hue

- **BUG-19718:** Traceback error shown when installing Hue on CentOS 5 or SLES 11

Problem: During installation of Hue on CentOS 5 and SLES 11, warning messages similar to the following display:

```
warning: user jenkins does not exist - using root (multiple messages)
warning: group jenkins does not exist - using root (multiple messages)
make: Entering directory `/usr/lib/hue/desktop'
--- Regenerating database at /usr/lib/hue/desktop/desktop.db
```

```

--- Syncing/updating database at /usr/lib/hue/desktop/desktop.db
Traceback (most recent call last):
  File "/usr/lib/hue/build/env/bin/hue", line 9, in module
    load_entry_point('desktop==2.3.4', 'console_scripts', 'hue')()
  File "/usr/lib/hue/desktop/core/src/desktop/manage_entry.py", line 60, in
  entry
    execute_manager(settings)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/__init__.py", line 438, in execute_manager
    utility.execute()
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/__init__.py", line 379, in execute
    self.fetch_command(subcommand).run_from_argv(self.argv)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 191, in run_from_argv
    self.execute(*args, **options.__dict__)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 219, in execute
    self.validate()
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/core/management/base.py", line 249, in validate
    num_errors = get_validation_errors(s, app)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.
  2.3-py2.6.egg/django/core/management/validation.py", line 28, in
  get_validation_errors
    for (app_name, error) in get_app_errors().items():
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/db/models/loading.py", line 146, in get_app_errors
    self._populate()
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/db/models/loading.py", line 64, in _populate
    self.load_app(app_name)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/db/models/loading.py", line 78, in load_app
    models = import_module('.models', app_name)
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/Django-1.2.3-py2.
  6.egg/django/utils/importlib.py", line 35, in import_module
    __import__(name)
  File "/usr/lib/hue/apps/jobbrowser/src/jobbrowser/models.py", line 20, in
  module
    import lxml.html
  File "/usr/lib/hue/build/env/lib/python2.6/site-packages/lxml-3.3.5-py2.6-
  linux-x86_64.egg/lxml/html/__init__.py", line 42, in module
    from lxml import etree
ImportError: /usr/lib/hue/build/env/lib/python2.6/site-packages/
lxml-3.3.5-py2.6-linux-x86_64.egg/lxml/etree.so: undefined symbol:
_PyByteArray_empty_string
make: *** [/usr/lib/hue/desktop/desktop.db] Error 1

```

Workaround: No action is required. The error message above can be ignored. Hue installation proceeds.

- **BUG-19511:** Error message during Hue start after upgrading Hue

Problem: After upgrading Hue from Hue 2.3.0 to Hue 2.3.2 and restarting the service, Hue displays

```
Failed to register some apps: Details in /var/log/hue/hue_re_register.log
```

The message is benign, and can be ignored.

Workaround: To eliminate this message, enter:

```
chown -R hue:hue /usr/lib/hue
```

- **BUG-19502:** Hue does not ship hive.aux.jars.path

Problem: Hue does not ship hive.aux.jars.path. This can cause problems with hbasestorage tables.

Workaround: To place the HBase jars on the Hue server host, enter:

```
ln -s /usr/lib/hbase/lib*.jar /usr/lib/hive/lib/
ln -s /etc/hbase/conf/*.xml /etc/hive/conf/
```

5.11.7. Known Issues for Storm

- **BUG-19709:** Storm /etc directory not created on Ubuntu after manual installation

Problem: After doing the manual install of Storm on ubuntu (apt-get install storm), as per the [HDP 2.1.3 manual installation instructions](#), no configuration folder is created for Storm under /etc; neither /etc/storm/conf nor /etc/storm/conf.dist are present.

Workaround: Create /etc/storm/conf on all the nodes where Storm is deployed, and create the storm.yaml file with contents as noted in [Installing HDP Manually](#).

5.11.8. Known Issues for Knox

- **BUG-16003:** Knox PID directory does not exist on Ubuntu after reboot

Problem: Unable to start Knox after Ubuntu reboot.

5.11.9. Known Issues for the Hortonworks Connector for Teradata

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

5.11.10. Known Issues for Ambari

- See [Ambari Release Notes](#).

5.12. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

5.13. Third-party Licenses

Table 5.2. Third-party Licenses

HDP Component	License
Phoenix	EPL

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

6. Release Notes HDP-2.1.2

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.2](#)
- [Fixed in This Release](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum system requirements](#)
- [Upgrading From HDP 2.1.1 to HDP 2.1.2](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

6.1. Product Version: HDP-2.1.2

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



Note

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

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

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Hive 0.13.0
- Apache Tez 0.4

- Apache ZooKeeper 3.4.5
- Hue 2.3.1
- Storm 0.9.1
- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0
- Third party components:
 - Ganglia 3.5.0
 - Ganglia Web 3.5.7
 - Nagios 3.5.0

6.2. Unsupported Apache components

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

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

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

- Application Timeline Server (Hive-on-Tez metrics)
- AM failure/restart resiliency
- MapReduce Uber AM
- YARN CGroup resource isolation

- Admin Node labels
- CPU Scheduling
- Fair Scheduler
- MapReduce Eclipse Plug-in

6.3. Fixed in This Release

HDP 2.1.2 includes the following fixes:

BUG-17210	STORM-269	Any readable file is exposed via the log viewer UI
BUG-17186	HADOOP-813	Allow wildcard bind for additional YARN/MR endpoints
BUG-17060	HIVE-6952	Hive Record Writer is not backward-compatible on HDP-2.1 release
BUG-16976	HIVE-6945	Dropping ranges of partitions returns an error after succeeding on Oracle
BUG-16890	HIVE-6957	Hive SQL standard auth calls accessing local or HDFS urls fails in Kerberos secure cluster with binary HS2 transport
BUG-16809, BUG-16008	HDFS-6273	Provide service binding address lists for HA services currently lacking them (HDFS/Namenode in 16808, MR/Yarn in 16809)
BUG-16797	HIVE-4576	templeton.hive.properties does not allow values with commas

6.4. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- [What's Changed in Mahout](#)
- [HDP 2.1 Clusters Deployed via Ambari](#)
- [What's Changed in Hue](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in Oozie](#)

6.4.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

Deprecated algorithms were removed ([MAHOUT-1296](#)) without Frequent Pattern Mining. Multilayer Perceptron was added ([MAHOUT-1265](#)).

6.4.2. HDP 2.1 clusters deployed via Ambari

Freshly-installed HDP 2.1 clusters deployed via Ambari will have the new Hive authorization system turned on by default. Manual installs and upgrades are not affected by this, only fresh HDP 2.1 installs via Ambari. This default behavior will be modified in Ambari 1.6.0.

6.4.3. Hue behavioral changes

Hue now supports the optional ability to input Unix usernames in lowercase letters and have Active Directory return usernames in upper case, in cases where LDAP/Active directory is being used as the back end.

6.4.4. HBase behavioral changes

In HDP 2.1, if the user does not have read privileges to a table and scans that table he will get an empty result set back.

6.4.5. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable `hive.server2.enable.doAs` is set to true, before the user starts the HiveServer2 process, they should create a scratch directory, `/tmp/hive-<username>`, on HDFS, where `<username>` is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

6.4.6. Oozie behavioral changes

When Oozie has been installed manually, before a site can execute any Oozie actions, the Oozie shared libraries must be made explicitly available. Add the following information to the file `oozie-site.xml`:

```
<property>
<name>oozie.service.WorkflowAppService.system.libpath</name>
<value>/user/$username/share/lib</value>
</property>
```

6.5. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)

- [Patch Information for Hive/HCat](#)
- [Patch Information for Oozie](#)



Note

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

6.5.1. Patch information for Hadoop Common/HDFS

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

- [HDFS-5257](#): addBlock() retry should return LocatedBlock with locations else client will get AIOBE
- [HDFS-5089](#): When a LayoutVersion supports SNAPSHOT, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- [HADOOP-10475](#): ConcurrentModificationException in AbstractDelegationTokenSelector.selectToken().
- [HDFS-6160](#): TestSafeMode occasionally fails.
- [HDFS-6233](#): Datanode throws HardLink exception during upgrade from 1.3 to 2.1 in Windows.

6.5.2. Patch information for ZooKeeper

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

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

6.5.3. Patch information for HBase

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

- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10592](#): Refactor PerformanceEvaluationtool
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10548](#): Correct commons-math dependency version

- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10621](#): Unable to grant user permission to namespace
- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table

- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

Windows Fixes:

- [HBASE-10799](#) [WINDOWS]
TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows

Changes related to HBASE-10070:

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

- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files
- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage

- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

6.5.4. Patch information for Pig

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

- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

6.5.5. Patch information for Tez

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

- [TEZ-1066](#): Generate events to integrate with YARN timeline server.
- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.

- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitonException: Invalid event: V_INIT at INITED.

6.5.6. Patch information for Hive/HCatalog

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

- [HIVE-6976](#): Show query id only when there's jobs on the cluster
- [HIVE-6966](#): More fixes for TestCliDriver on Windows
- [HIVE-6927](#): Add support for MSSQL in schematool
- [HIVE-6919](#): hive sql std auth select query fails on partitioned tables
- [HIVE-6915](#): Hive Hbase queries fail on secure Tez cluster
- [HIVE-6898](#): Functions in hive are failing with java.lang.ClassNotFoundException on Tez
- [HIVE-6888](#): Hive leaks MapWork objects via Utilities::gWorkMap
- [HIVE-6883](#): Dynamic partitioning optimization does not honor sort order or order by
- [HIVE-6880](#): TestHWISessionManager fails with -Phadoop-2
- [HIVE-6871](#): Build fixes to allow Windows to run TestCliDriver
- [HIVE-6868](#): Create table in HCatalog sets different SerDe defaults than what is set through the CLI
- [HIVE-6828](#): Hive tez bucket map join conversion interferes with map join conversion
- [HIVE-6826](#): Hive-tez has issues when different partitions work off of different input types
- [HIVE-6824](#): Hive HBase query fails on Tez due to missing jars - part 2
- [HIVE-6817](#): Some hadoop2-only tests need diffs to be updated
- [HIVE-6799](#): HiveServer2 needs to map kerberos name to local name before proxy check
- [HIVE-6788](#): Abandoned opened transactions not being timed out
- [HIVE-6745](#): HCat MultiOutputFormat hardcodes DistributedCache keynames
- [HIVE-6741](#): HiveServer2 startup fails in secure (kerberos) mode due to backward incompatible hadoop change
- [HIVE-6726](#): Hcat cli does not close SessionState

- [HIVE-6695](#): bin/hcat should include hbase jar and dependencies in the classpath [followup/clone of HCATALOG-621]
- [HIVE-6571](#): query id should be available for logging during query compilation
- [HIVE-6569](#): HCatalog still has references to deprecated property hive.metastore.local
- [HIVE-5775](#): Introduce Cost Based Optimizer to Hive

6.5.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [Oozie-615](#): Support Oozie HA.
- [Oozie-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [Oozie-1306](#): Bring cron syntax to coordinator frequency.
- [Oozie-1460](#): Implement and document oozie HA security.
- [Oozie-1486](#): cut down on number of small files to track a running action.
- [Oozie-1491](#): Make sure oozie works with secure ZooKeeper.
- [Oozie-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [Oozie-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [Oozie-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- [Oozie-1541](#): typo in oozie HA admin-server command line documentation.
- [Oozie-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [Oozie-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [Oozie-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [Oozie-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [Oozie-1576](#): Add documentation for oozie sqoop CLI.
- [Oozie-1587](#): Add "recovery" column to CoordJob table.
- [Oozie-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.

- [Oozie-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [Oozie-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [Oozie-1618](#): Dryrun should check variable substitution in workflow.xml.
- [Oozie-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [Oozie-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [Oozie-1726](#): Oozie does not support _HOST when configuring kerberos security.
- [Oozie-1733](#): Fix test failures by oozie-1722.

6.5.8. Patch information for Sqoop

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

- [SQOOP-1617](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1209](#): DirectNetezzaManager fails to find tables from older Netezza system catalogs.
- [SQOOP-1298](#): Cannot export to VARBINARY with null value.
- [SQOOP-1297](#): Parameterize the Accumulo version in the build files.
- [SQOOP-1282](#): Consider Avro files even if they carry no extension.
- [SQOOP-1278](#): Allow use of uncommitted isolation for databases that support it as an import option.
- [SQOOP-1273](#): Multiple append jobs can easily end up sharing directories.
- [SQOOP-1268](#): Sqoop tarballs do not contain .gitignore and .gitattribute files.
- [SQOOP-1056](#): Implement connection resiliency in Sqoop using pluggable failure handlers.
- [SQOOP-1057](#): Introduce fault injection framework to test connection resiliency.
- [SQOOP-1271](#): Sqoop hcatalog location should support older bigtop default location also.
- [SQOOP-1226](#): -password-file option triggers FileSystemClosed exception at end of Oozie action.
- [SQOOP-1260](#): HADOOP_MAPRED_HOME should be defaulted correctly.
- [SQOOP-1259](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1261](#): Sqoop on Windows can't run HCatalog/HBase multinode jobs.
- [SQOOP-1249](#): Sqoop HCatalog Import fails with -queries because of validation issues.
- [SQOOP-1250](#): Oracle connector is not disabling autoCommit on created connections.

- [SQOOP-1246](#): HBaseImportJob should add job authtoken only if HBase is secured.
- [SQOOP-767](#): Add support for Accumulo.
- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding *_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

6.6. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

6.6.1. Hardware recommendations

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

6.6.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

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

6.6.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

6.6.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

6.6.5. Virtualization and cloud platforms

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

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

6.6.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

6.7. Upgrading From HDP 2.1.1 to HDP 2.1.2

This section describes how to upgrade an existing HDP 2.1.1 installation to HDP 2.1.2.

If you are upgrading from a previous HDP version, such as HDP 2.0, please upgrade to HDP 2.1.1 before beginning this process. See:

- **Ambari:** http://docs.hortonworks.com/HDPDocuments/Ambari-1.6.0.0/bk_upgrading_Ambari/content/ambari-upgrade-stack-21.html
- **Manual:** http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.1/bk_installing_manually_book/content/upgrade_2X.html

Before You Begin

- Make sure you know what HDP components need to be upgraded at your installation.

- Think about whether you're going to upgrade using a [local repository](#) or a [remote repository](#).

To upgrade from HDP 2.1.1 to HDP 2.1.2, do the following:

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

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.1.2.0/hdp.repo
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.2.0/hdp.repo
SLES 11	http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.1.2.0/hdp.repo

OR Download the HDP RPMs, and hdp.repo as a single tar file. (For further information, see the [local repository instructions](#).)

RHEL/CENTOS/ ORACLE LINUX 5	http://public-repo-1.hortonworks.com/HDP/centos5/2.x/GA/2.1-latest/HDP-2.1-latest-centos5-rpm.tar.gz
RHEL/CENTOS/ ORACLE LINUX 6	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/GA/2.1-latest/HDP-2.1-latest-centos6-rpm.tar.gz
SLES 11	http://public-repo-1.hortonworks.com/HDP/suse11/2.x/GA/2.1-latest/HDP-2.1-latest-suse11-rpm.tar.gz

2. Stop all services.

If you are managing your deployment via Ambari, open Ambari Web, browse to **Services** and use the **Service Actions** command to stop each service.

If you are upgrading manually, follow the instructions in the [HDP 2.1.2 Reference Guide](#).



Note

If you are upgrading an HA NameNode configuration, make sure you only stop the services and do not stop any of the JournalNodes. If any of your JournalNodes are not running while performing any upgrade, rollback or finalization operation, or else the operation will fail.

3. If you have Oozie installed/running:

Back up the files in the following directories on the Oozie server host, and make sure all files, including *site.xml files, are copied	<pre>mkdir oozie-conf-bak cp -R /etc/oozie/conf/* oozie-conf-bak</pre>
Remove the old Oozie directories on all Oozie server and client hosts	<pre>rm -R /etc/oozie/conf/* oozie-conf-bak</pre>

4. Upgrade the stack on all Agent hosts.

The following instructions include all possible components that can be upgraded. If your installation does not use a particular component, skip those installation instructions.

Operating System	Instructions	Commands
RHEL/CentOs/ Oracle Linux	Remove WebHCat components:	<pre>yum erase "webhcat*"</pre>
	If you haven't done so already, stop the Hive Metastore.	

Operating System	Instructions	Commands
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<pre>yum upgrade hive yum erase hcatalog yum install hive-hcatalog</pre>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType < \$databaseType></pre> <p>The value for \$databaseType can be derby, mysql, oracle, or postgres.</p>
	Upgrade the following components:	<pre>yum upgrade "collectd*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "tez*" "falcon*" "storm*" "squoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</pre>
		<pre>yum install webhcat-tar-hive webhcat-tar-pig yum install "hive-*" yum install oozie oozie-client rpm -e --nodeps bigtop-jsvcyum install bigtop-jsvc</pre>
	Verify that the components were upgraded. Enter:	<pre>yum list installed grep HDP-\$old-stack-version-number</pre>
SLES	Remove WebHCat, components:	<pre>zypper remove webhcat*</pre>
	If you haven't done so already, stop the Hive Metastore.	
	Upgrade Hive and HCatalog. On the Hive and HCatalog host machines, enter:	<pre>zypper install hive-hcatalog</pre>
	Upgrade the Hive Metastore database schema. On the Hive host machine, enter:	<pre>\$HIVE_HOME/bin/schematool -upgradeSchema -dbType < \$databaseType></pre>
	Upgrade the following components:	<pre>zypper up "collectd*" "epel-release*" "gccxml*" "pig*" "hadoop*" "phoenix*" "knox*" "falcon*" "tez*" "storm*" "sqoop*" "zookeeper*" "hbase*" "hive*" hdp_mon_nagios_addons</pre>
		<pre>zypper install webhcat-tar-hive webhcat-tar-pig zypper up -r HDP-2.1.2.0 zypper install hive*zypper install oozie oozie-client</pre>
	Verify that the components were upgraded:	<pre>rpm -qa grep hadoop, rpm -qa grep hive and rpm -qa grep hcatalog</pre>
	If components were not upgraded, upgrade them:	<pre>yast --update hadoop hcatalog hive</pre>

5. If you are upgrading from an HA NameNode configuration, restart all JournalNodes. On each JournalNode host, enter the following command:

```
su -l {HDFS_USER} -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh start
journalnode"
```

6. Complete the Stack upgrade.

If this is an Ambari-managed cluster, update the Repository Base URLs to use the HDP 2.1.3 repositories for HDP and HDP-UTILS. For Ambari 1.6.0 or earlier, enter:

```
ambari-server upgradestack HDP-2.1 http://public-repo-1.hortonworks.com/HDP/
{$os}/2.x/updates/2.1.3.0 {$os}
```

where {\$os} is the Operating System Family (OS Family). See the following table:

Table 6.1. Operating Systems mapped to each OS Family

OS Family	Operating System
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5

OS Family	Operating System
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11

7. Finalize the upgrade.

If you are not yet ready to discard your backup, you can start the upgraded HDFS without finalizing the upgrade. (At this stage, you can still roll back if need be.)

Verify your filesystem health. When you are ready to commit to this upgrade (are certain that you will not want to roll back), discard your backup and finalize the upgrade.

As \$HDFS_USER, execute the following command:

```
hdfs dfsadmin -finalizeUpgrade
```

6.8. Improvements

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

- HDP 2.1 supports the use of Oracle DB as a metastore.
- The Storm LogViewer service now allows only worker log files from the Storm log directory to be viewed from the UI, regardless of the privileges of the person running the service.

6.9. Common Vulnerabilities and Exposures

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability
- **Severity:** Major
- **Vendor:** The Apache Software Foundation
- **Versions Affected:** Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0
- **Users Affected:** Users who have enabled Hadoop's MapReduce security features
- **Impact:** Vulnerability allows an unauthorized user to retrieve job details from the job history server
- **Mitigation:** Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to 2.3.0
- **Credit:** This issue was discovered by Koji Noguchi of Yahoo

6.10. Known Issues

In this section:

- [Known Limitation for HDFS Upgrade](#)
- [Known Issues for SLES 11](#)

- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for WebHCat](#)
- [Known Issues for Oozie](#)
- [Known Issues for Hue](#)
- [Known Issues for Flume](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

6.10.1. Known Limitation for HDFS Upgrade

- **BUG-17431:** ([HDFS-5526](#)) If you upgrade from HDP 2.0.9.1 to HDP 2.1, the DataNode cannot be rolled back to HDP 2.0.9.1.

Contact Support for an update to fix this HDP 2.0.9.1 issue.

6.10.2. Known Issues for SLES 11

- **BUG-9904:** php_curl Required for SLES 11 Sp1

Problem: Several alerts return with `Return code of 255 is out of bounds` while trying to install a cluster on SLES because `php_curl` is not installed.

Workaround: Install `php_curl` on your SLES host.

```
zypper install php-curl
```

6.10.3. Known Issues for HDP

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

Problem: RHEL 6.2 and RHEL 6.3 contain known bugs in `nslcd`. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by `nslcd` times out.

Workaround: If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.

- Increase the configured nslcd search timeout in `/etc/nslcd.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGro` in `core-site.xml`.
- **BUG-15360:HADOOP-10519:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error

Problem:

- **BUG-825:ED2** m1.large cluster root partition is only 5GB and fills up quickly by HDP logs

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

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

6.10.3.1. Known issues for HDFS

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

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

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_00000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msec
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
9 more
```

6.10.3.2. Known issues for MapReduce

- **BUG-15360: (HADOOP-10519)** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.
- **BUG-14749:** CombineFileInputFormat.getSplits() including directories in its results.

Problem: This is causing Hive test `root_dir_external_table.q` to fail when running against hadoop-2. Opened Apache Jira [MAPREDUCE-5756](#) Created in Monarch as <https://hwxmonarch.atlassian.net/browse/HADOOP-801>, creating equivalent bug for HDP-2.1.

- **BUG-12005:** ([MAPREDUCE-2306](#)) `Mapreduce.task.io.sort.mb` is capped at 2047.

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

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

6.10.4. Known Issues for YARN

- **BUG-15834:** YARN and/or mapred client should add tokens for default filesystem.

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

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

Problem: Seeing about 1 GB of logs per hour.

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

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

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

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

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

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

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

```
mapred-site.xml
mapreduce.map.memory.mb 1024
mapreduce.map.java.opts -Xmx692m
mapreduce.task.io.sort.mb 200
```

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

- **BUG-7531:** ([YARN-991](#)) Hadoop metrics link does not contain correct content.

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

6.10.5. Known Issues for HBase

- **BUG-16900: (HBASE-11036)** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.

Problem: The Big Linked List Test with Chaos Monkey Test run shows Not Serving Region Exceptions in the Yarn logs, causing failure.

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

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

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

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

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

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

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

6.10.6. Known Issues for Phoenix

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

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

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

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

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

6.10.7. Known Issues for Hive

- **BUG-17773:** Hive-on-Tez does not work with encrypted shuffle.

Problem: The Encrypted Shuffle capability enables the MapReduce shuffle to be encrypted using HTTPS, with optional client authentication. (This is also known as bi-directional HTTPS, or HTTPS with client certificates.) Hive-on-Tez does not work in an environment where encrypted shuffle is enabled.

- **BUG-17603:** ([HIVE-6985](#)) Grant privileges on PUBLIC role are not being honored.

Problem: When a privilege is granted to public role, the privilege is supposed to be applicable for all users. However, the privilege check fails for users, even if they have public role in the list of current roles. (This issue only exists with the Public role; it does not affect the granting of privileges to other roles.)

- **BUG-16945:** HiveSysTest-testTableCreationAndQuery fails with error "Permission denied inode="/tmp/hive-hadoop"".

Problem: HiveSysTest-testTableCreationAndQuery fails with error "Permission denied inode="/tmp/hive-hadoop"". Hadoop user is not cleaned up as guest user and thus the test fails. Verified hiveserver2.doas property is set to true in hive-site.xml.

- **BUG-16890:** ([HIVE-6957](#)) Hive SQL standard auth calls accessing local or HDFS URLs fail in Kerberos secure cluster with binary HS2 transport.

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

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

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

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

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

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

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.
- **BUG-16608: (FALCON-390)** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

Problem: The job fails with the following permission error:

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

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

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

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

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

Problem: The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
(SessionState.java:printError(546)) - FAILED: HiveAccessControlException
Permission denied. Principal [name=hrt_qa, type=USER] does not have
following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index_] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
name=default__missing_ddl_3_missing_ddl_3_index_] : [OBJECT
OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
```

```

at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Problem: The error returned on the Hive console is:

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

```

at org.apache.hadoop.security.UserGroupInformation.
doAs(UserGroupInformation.java:1548)
at org.apache.hadoop.mapred.YarnTezDagChild.main(YarnTezDagChild.java:538)
Caused by: java.lang.ClassCastException: org.apache.hadoop.io.Text cannot be
cast to org.apache.hadoop.hive.serde2.columnar.BytesRefArrayWritable
at org.apache.hadoop.hive ql.io.RCFileRecordReader.next(RCFileRecordReader.
java:44)
at org.apache.hadoop.hive ql.io.HiveContextAwareRecordReader.
doNext(HiveContextAwareRecordReader.java:339)
... 13 more

```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-8227: (HIVE-6638)** Hive needs to implement recovery or extend FileOutputComitter.

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

```

public boolean isRecoverySupported() {
    return false;
}

```

6.10.8. Known Issues for Tez

- **BUG-17773:** Hive-on-Tez does not work with encrypted shuffle.
- **BUG-15376: (YARN-1892)** CS fast scheduling patch ends up causing excessive logging.

Problem: Seeing about 1 GB of logs per hour.

6.10.9. Known Issues for WebHCat

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

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

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

When the Hive command is run it builds the hiveconf from the templeton.hive.properties. To enable Tez we would need to atleast add

"hive.execution.engine=tez" to templeton.hive.properties. On Windows this is not a problem because we use the local Hive installation.

6.10.10. Known Issues for Oozie

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

Problem: The job fails with the following permission error:

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

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

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

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

- **BUG-16397:** Documentation does not describe how to renew an expired Oozie authorization token.

Problem: Due to the introduction of [HADOOP-10416](#), users may encounter situations where they are not authorized to perform certain actions because they are in an unsecured environment and their Oozie authorization token has expired.

Workaround: Remove the oozie auth token cache file at `System.getProperty("user.home")/.oozie-auth-token`, then re-run the Oozie command to renew the token.

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

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

- **BUG-10400:** Oozie workflows that contain Hive queries which run mapreduce jobs fail on secure clusters.

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

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

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

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

6.10.11. Known Issues for Hue

- **BUG-9734:** Data loss during Migration of Hue DB from default (sqlite) to Oracle DB:

Problem: Migration of data and tables from SQLite to Oracle does not work and needs to be performed manually.

1. Install Hue and start Hue (Hue creates table in sqlite db).
2. Do NOT perform any tasks (such as uploading files, pig jobs, or hcat jobs) on the HDP stack from Hue UI.
3. Stop Hue, configure Oracle.
4. Start Hue.

Result: Hue starts fine and continues working, but there is loss of data.

5. Some tables are lost in HCatalog.
6. Some pig scripts do not show up on UI.

Workaround: Manually migrate the data and tables from SQLite to Oracle.

6.10.12. Known Issues for Flume

- **BUG-16771:** ([HIVE-6867](#)) Hive table returns multiple copies of streaming data when testing the Hive Server restart scenario.

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

6.10.13. Known Issues for Storm

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

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

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

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

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

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
initialization")
```

6.10.14. Known Issues for Falcon

- **BUG-16608: (FALCON-390)** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

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

Workaround: Add a Hive default configuration to Oozie.

Stop the Oozie service.



Warning

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

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

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

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

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

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

Workaround: Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.
2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to `/etc/hadoop/conf-1` and the other to `/etc/hadoop/conf-2`.
3. For each `oozie-site.xml` file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

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


```

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

```

4. Restart the Oozie service on all clusters.

6.10.15. Known Issues for Knox

- **BUG-16592:** When accessing through Knox Gateway, Oozie not supported for HDP for Windows or Linux single-node clusters.

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

Workaround: Either use Oozie only when accessing a multinode HDP cluster in Linux, or access Oozie without going through Knox.

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

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

6.10.16. Known Issues for the Hortonworks Connector for Teradata

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

6.10.17. Known Issues for Ambari

- See [Ambari Release Notes](#).

6.11. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

6.12. Third-party Licenses

Table 6.2. Third-party Licenses

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

7. Release Notes HDP-2.1.1

The HDP 2.1 Release Notes include the following sections:

- [Product Version: HDP-2.1.1](#)
- [Behavioral Changes](#)
- [Patch Information](#)
- [Minimum system requirements](#)
- [Improvements](#)
- [Common Vulnerabilities and Exposures](#)
- [Known Issues](#)
- [Deprecated Features](#)
- [Third-party Licenses](#)

7.1. Product Version: HDP-2.1.1

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



Note

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

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

- Apache Hadoop 2.4
- Apache HBase 0.98.0
- Apache Pig 0.12.1
- Apache Tez 0.4
- Apache ZooKeeper 3.4.5
- Hue 2.3.1
- Storm 0.9.1

- Apache Oozie 4.0.0
- Apache Falcon 0.5
- Apache Sqoop 1.4.4
- Apache Knox 0.4
- Apache Flume 1.4.0
- Apache Accumulo 1.5.1
- Apache Phoenix 4.0.0
- Apache Avro 1.7.4
- Apache Mahout 0.9.0
- Third party components:
 - Ganglia 3.5.0
 - Ganglia Web 3.5.7
 - Nagios 3.5.0

7.2. Unsupported Apache components

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

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

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

- Application Timeline Server (Hive-on-Tez metrics)
- AM failure/restart resiliency
- MapReduce Uper AM
- YARN CGroup resource isolation
- Admin Node labels
- CPU Scheduling
- Fair Scheduler

- MapReduce Eclipse Plug-in

7.3. Behavioral Changes

The following Apache Components Changed in HDP 2.1:

- [What's Changed in Mahout](#)
- [HDP 2.1 Clusters Deployed via Ambari](#)
- [What's Changed in Hue](#)
- [What's Changed in HBase](#)
- [What's Changed in Hive](#)
- [What's Changed in Oozie](#)

7.3.1. Mahout behavioral changes

Mahout is now Mahout 0.9.

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

7.3.2. HDP 2.1 clusters deployed via Ambari

Freshly-installed HDP 2.1 clusters deployed via Ambari will have the new Hive authorization system turned on by default. Manual installs and upgrades are not affected by this, only fresh HDP 2.1 installs via Ambari. This default behavior will be modified in Ambari 1.6.0.

7.3.3. Hue behavioral changes

Hue now supports the optional ability to input Unix usernames in lowercase letters and have Active Directory return usernames in upper case, in cases where LDAP/Active directory is being used as the back end.

7.3.4. HBase behavioral changes

In HDP 2.1, if the user does not have read privileges to a table and scans that table he will get an empty result set back.

7.3.5. Hive behavioral changes

When using Tez as the Hive execution engine, if the variable `hive.server2.enable.doAs` is set to true, before the user starts the HiveServer2 process, they should create a scratch directory, `/tmp/hive-<username>`, on HDFS, where `<username>` is the user who will be running the HiveServer2 process. The directory should have read-write-execute (777) permission.

7.3.6. Oozie behavioral changes

When Oozie has been installed manually, before a site can execute any Oozie actions, the Oozie shared libraries must be made explicitly available. Add the following information to the file `oozie-site.xml`:

```
<property>
<name>oozie.service.WorkflowAppService.system.libpath</name>
<value>/user/$username/share/lib</value>
</property>
```

7.4. Patch Information

In this section:

- [Patch Information for Hadoop Common/HDFS](#)
- [Patch Information for ZooKeeper](#)
- [Patch Information for HBase](#)
- [Patch Information for Pig](#)
- [Patch Information for Tez](#)
- [Patch Information for Hive//HCat](#)
- [Patch Information for Oozie](#)



Note

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

7.4.1. Patch information for Hadoop Common/HDFS

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

- [HDFS-5257](#): `addBlock()` retry should return `LocatedBlock` with locations else client will get AIOBE
- [HDFS-5089](#): When a `LayoutVersion` supports `SNAPSHOT`, it must support `FSIMAGE_NAME_OPTIMIZATION`.
- [HADOOP-10475](#): `ConcurrentModificationException` in `AbstractDelegationTokenSelector.selectToken()`.
- [HDFS-6160](#): `TestSafeMode` occasionally fails.
- [HDFS-6233](#): Datanode throws `HardLink` exception during upgrade from 1.3 to 2.1 in Windows.

7.4.2. Patch information for ZooKeeper

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

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

7.4.3. Patch information for HBase

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

- [HBASE-10833](#): Region assignment may fail during cluster start up
- [HBASE-10829](#): Flush is skipped after log replay if the last recovered edits file is skipped
- [HBASE-10514](#): Forward port HBASE-10466, possible data loss when failed flushes.
- [HBASE-10700](#): IntegrationTestWithCellVisibilityLoadAndVerify should allow current user to be the admin
- [HBASE-10592](#): Refactor PerformanceEvaluationtool
- [HBASE-10419](#): Add multiget support to PerformanceEvaluation
- [HBASE-10548](#): Correct commons-math dependency version
- [HBASE-10809](#): HBaseAdmin#deleteTable fails when META region happen to move around same time
- [HBASE-10793](#): AuthFailed as a valid zookeeper state
- [HBASE-10767](#): Load balancer may interfere with tests in TestHBaseFsck
- [HBASE-9721](#): RegionServer should not accept regionOpen RPC intended for another(previous) server
- [HBASE-10688](#): Add a draining_node script to manage nodes in draining mode
- [HBASE-8304](#): Bulkload fails to remove files if fs.default.name / fs.defaultFS is configured without default port
- [HBASE-10660](#): MR over snapshots can OOM when alternative blockcache is enabled
- [HBASE-10635](#): thrift#TestThriftServer fails due to TTL validity check
- [HBASE-10591](#): Sanity check table configuration in createTable
- [HBASE-10670](#): HBaseFsck#connect() should use new connection
- [HBASE-10632](#): Region lost in limbo after ArrayIndexOutOfBoundsException during assignment
- [HBASE-10621](#): Unable to grant user permission to namespace

- [HBASE-10638](#): Improve error message when there is no region server available for move
- [HBASE-10582](#): 0.94->0.96 Upgrade: ACL can't be repopulated when ACL table contains row for table '-ROOT' or '.META.'
- [HBASE-10581](#): ACL znode are left without PBed during upgrading hbase0.94* to hbase0.96+
- [HBASE-10500](#): Some tools OOM when BucketCache is enabled
- [HBASE-10486](#): ProtobufUtil Append and Increment deserialization lost cell level timestamp
- [HBASE-10844](#): Coprocessor failure during batchmutation leaves the memstore datastructs in an inconsistent state (Note: the committed fix only improves logging)
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10852](#): TestDistributedLogSplitting#testDisallowWritesInRecovering occasionally fails
- [HBASE-10863](#): Scan doesn't return rows for user who has authorization by visibility label in secure deployment
- [HBASE-10618](#): User should not be allowed to disable/drop visibility labels table
- [HBASE-10895](#): unassign a region fails due to the hosting region server is in FailedServerList
- [HBASE-10850](#): essential column family optimization is broken
- [HBASE-10751](#): TestHRegion testWritesWhileScanning occasional fail since HBASE-10514 went in

Windows Fixes:

- [HBASE-10799](#) [WINDOWS] TestImportTSVWithVisibilityLabels.testBulkOutputWithTsvImporterTextMapper fails on windows
- [HBASE-10735](#) [WINDOWS] Set -XX:MaxPermSize for unit tests
- [HBASE-10685](#) [WINDOWS] TestKeyStoreKeyProvider fails on windows
- [HBASE-10686](#) [WINDOWS] TestStripeStoreFileManager fails on windows

Changes related to HBASE-10070:

- [HBASE-10875](#) Metas own location should be cached
- [HBASE-10791](#) Add integration test to demonstrate performance improvement
- [HBASE-10810](#) LoadTestTool should share the connection and connection pool

- [HBASE-10794](#) multi-get should handle missing replica location from cache
- [HBASE-10634](#) Multiget doesn't fully work.
- [HBASE-10661](#)
TestStochasticLoadBalancer.testRegionReplicationOnMidClusterWithRacks() is flaky
- [HBASE-10701](#) Cache invalidation improvements from client side
- [HBASE-10778](#) Unique keys accounting in MultiThreadedReader is incorrect
- [HBASE-10743](#) Replica map update is problematic in RegionStates
- [HBASE-10616](#) Integration test for multi-get calls
- [HBASE-10734](#) Fix RegionStates.getRegionAssignments to not add duplicate regions
- [HBASE-10729](#) Enable table doesn't balance out replicas evenly if the replicas were unassigned earlier
- [HBASE-10726](#) Fix java.lang.ArrayIndexOutOfBoundsException in StochasticLoadBalancer
- [HBASE-10720](#) rpcClient: Wrong log level when closing the connection
- [HBASE-10704](#) BaseLoadBalancer#roundRobinAssignment() may add same region to assignment plan multiple times
- [HBASE-10633](#) StoreFileRefresherChore throws ConcurrentModificationException sometimes
- [HBASE-10572](#) Create an IntegrationTest for region replicas.
- [HBASE-10703](#) TestAsyncProcess does not pass on HBASE-10070
- [HBASE-10637](#) rpcClient: Setup the iostreams when writing
- [HBASE-10620](#) LoadBalancer.needsBalance() should check for co-located region replicas as well
- [HBASE-10672](#) Table snapshot should handle tables whose REGION_REPLICATION is greater than one.
- [HBASE-10630](#) NullPointerException in ConnectionManager.locateRegionInMeta() due to missing region info
- [HBASE-10356](#) Failover RPC's for multi-get.
- [HBASE-10525](#) Allow the client to use a different thread for writing to ease interrupt.
- [HBASE-10355](#) Failover RPC's from client using region replicas.
- [HBASE-10352](#) Region and RegionServer changes for opening region replicas, and refreshing store files

- [HBASE-10351](#) LoadBalancer changes for supporting region replicas
- [HBASE-10359](#) Master/RS WebUI changes for region replicas.
- [HBASE-10362](#) HBACK changes for supporting region replicas.
- [HBASE-10361](#) Enable/AlterTable support for region replicas.
- [HBASE-10350](#) Master/AM/RegionStates changes to create and assign region replicas.
- [HBASE-10490](#) Simplify RpcClient code (Nicolas Liochon)
- [HBASE-10511](#) Add latency percentiles on PerformanceEvaluation
- [HBASE-10517](#) NPE in MetaCache.clearCache()
- [HBASE-10479](#) HConnection interface is public but is used internally, and contains a bunch of methods
- [HBASE-10348](#) HTableDescriptor changes for region replicas
- [HBASE-10354](#) Add an API for defining consistency per request
- [HBASE-10347](#) HRegionInfo changes for adding replicaId and MetaEditor/MetaReader changes for region replicas
- [HBASE-10277](#) refactor AsyncProcess
- [HBASE-10427](#) clean up HRegionLocation/ServerName usage
- [HBASE-10472](#) Manage the interruption in ZKUtil#getData
- [HBASE-10859](#) HStore.openStoreFiles() should pass the StoreFileInfo object to createStoreFileAndReader().
- [HBASE-10858](#) TestRegionRebalancing is failing

7.4.4. Patch information for Pig

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

- [PIG-3573](#): Provide StoreFunc and LoadFunc for Accumulo.
- [PIG-3558](#): ORC support for Pig.
- [PIG-3257](#): Add a UUID function to Pig.

7.4.5. Patch information for Tez

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

- [TEZ-1066](#): Generate events to integrate with YARN timeline server.

- [TEZ-1048](#): Fix an NPE which can occur when the source task generates no data for a partition, and runs multiple attempts.
- [TEZ-1045](#): TezMiniCluster tests can fail intermittently.
- [TEZ-1040](#): Fix a bug which could cause the Merger to hang.
- [TEZ-1034](#): Shuffling can sometimes hang with duplicate inputs for the same index.
- [TEZ-1033](#): AM hangs during recovery with Tasks awaiting init event.
- [TEZ-1030](#): Address intermittent errors created due to race condition in [YARN-1915](#).
- [TEZ-1028](#): Handle killed tasks and attempts when handling recovery data.
- [TEZ-1029](#): Fetcher can fail to report input failed event upon connection error.
- [TEZ-1021](#): TezClient cannot connect to AM in a secure cluster when launched via Oozie.
- [TEZ-1020](#): VertexImpl handling of task failed in SUCCEEDED state is incorrect.
- [TEZ-1015](#): Dag failed with Invalid event: V_ROUTE_EVENT at RECOVERING.
- [TEZ-1014](#): Add a log message to indicate last AM attempt.
- [TEZ-1004](#): AM relocation doesn't handle conflicting resources correctly.
- [TEZ-1005](#): AM relocation adds resources to the wrong classloader.
- [TEZ-1011](#): TestDAGRecovery timing out on jenkins builds.
- [TEZ-1010](#): TestAMNodeMap.testSelfBlacklist fails intermittently
- [TEZ-997](#): Internal Error in am logs during dag shutdown.
- [TEZ-1009](#): Fixes in log file roll-over
- [TEZ-998](#): InvalidStateTransitionException: Invalid event: V_INIT at INITED.

7.4.6. Patch information for Hive/HCatalog

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

- [HIVE-6117](#): HBase_1 and HBase_2 tests are failing
- [HIVE-5601](#): NPE in ORC's PPD when using SELECT * from table with WHERE predicate
- [HIVE-5542](#): WebHCat is failing to run ddl command on a secure cluster
- [HIVE-5515](#): Writing to an HBase table throws IllegalArgumentException, failing job submission
- [HIVE-5511](#): percentComplete returned by job status from WebHCat is null

- [HIVE-5496](#): `hcat -e drop database if exists` fails on authorizing non-existent null db
- [HIVE-5485](#): SBAP errors on null partition being passed into partition level authorization
- [HIVE-5484](#): TestSchemaTool failures when Hive version has more than 3 revision numbers
- [HIVE-5480](#): WebHCat e2e tests for doAs feature are failing
- [HIVE-5479](#): SBAP restricts `hcat -e 'show databases'`
- [HIVE-5478](#): WebHCat e2e testsuite for hcat authorization tests needs some fixes
- [HIVE-5474](#): drop table hangs when `concurrency=true`
- [HIVE-5453](#): `jobsubmission2.conf` should use 'timeout' property
- [HIVE-5448](#): WebHCat duplicate test `TestMapReduce_2` should be removed
- [HIVE-5425](#): Provide a configuration option to control the default stripe size for ORC
- [HIVE-5422](#): Upgrade Kryo to 2.22 now that it is released
- [HIVE-5411](#): Migrate expression serialization to Kryo
- [HIVE-5379](#): `NoClassDefFoundError` is thrown when using lead/lag with kryo serialization
- [HIVE-5353](#): job submission that requires access to metastore should not require additional jars to be shipped to target node
- [HIVE-5290](#): Some HCatalog tests have been behaving flaky
- [HIVE-5279](#): Kryo cannot instantiate `GenericUDAFEvaluator` in `GroupByDesc`
- [HIVE-5263](#): Query Plan cloning time could be improved by using Kryo
- [HIVE-5133](#): webhcat jobs that need to access metastore fails in secure mode
- [HIVE-5112](#): Upgrade protobuf to 2.5 from 2.4
- [HIVE-5070](#): Need to implement `listLocatedStatus()` in `ProxyFileSystem` for 0.23 shim
- [HIVE-4910](#): Hadoop 2 archives broken
- [HIVE-4545](#): HS2 should return describe table results without space padding
- [HIVE-4485](#): beeline prints null as empty strings
- [HIVE-4388](#): HBase tests fail against Hadoop 2
- [HIVE-3815](#): hive table rename fails if filesystem cache is disabled
- [HIVE-1511](#): Hive plan serialization is slow.

7.4.7. Patch information for Oozie

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

- [OOZIE-1593](#): Fixed Oozie HCatCredential provider needs to include hadoop rpc protection to work with encrypted secure clusters.
- [OOZIE-1563](#): Fixed Colt jar includes GPL licence.
- [Oozie-615](#): Support Oozie HA.
- [Oozie-1305](#): Coordinator job should have an option to recover "none" of the actions after downtime.
- [Oozie-1306](#): Bring cron syntax to coordinator frequency.
- [Oozie-1460](#): Implement and document oozie HA security.
- [Oozie-1486](#): cut down on number of small files to track a running action.
- [Oozie-1491](#): Make sure oozie works with secure ZooKeeper.
- [Oozie-1520](#): SequenceFile reader fails to use doas for reading action data file.
- [Oozie-1525](#): Oozie workflow does not update status sometimes and is stuck in Running state.
- [Oozie-1540](#): When oozie.zookeeper.oozie.id is not specified, it's using space instead of a hostname.
- [Oozie-1541](#): typo in oozie HA admin-server command line documentation.
- [Oozie-1555](#): Launcher mapper to check for system properties before opening files for action data.
- [Oozie-1560](#): Log messages should have a way to identify when server it comes from when using HA.
- [Oozie-1569](#): Maintain backward compatibility for running jobs before upgrade.
- [Oozie-1575](#): Add functionality to submit sqoop jobs through http from oozie server side.
- [Oozie-1576](#): Add documentation for oozie sqoop CLI.
- [Oozie-1587](#): Add "recovery" column to CoordJob table.
- [Oozie-1580](#): EL variables don't get resolved in configurations imported from <job-xml>.
- [Oozie-1600](#): Mapreduce actions without configuration section in workflow.xml throws "IllegalArgumentException: element cannot be null".
- [Oozie-1608](#): update curator to 2.4.0 when it's available to fix security hole.
- [Oozie-1618](#): Dryrun should check variable substitution in workflow.xml.

- [Oozie-1691](#): StackOverflowError in TimestampedMessageParser.parseNextLine().
- [Oozie-1722](#): When an ApplicationMaster restarts, it restarts the launcher job.
- [Oozie-1726](#): Oozie does not support _HOST when configuring kerberos security.
- [Oozie-1733](#): Fix test failures by oozie-1722.

7.4.8. Patch information for Sqoop

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

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

- [SQOOP-1228](#): Method Configuration#unset is not available on Hadoop 1.2.0.
- [SQOOP-1224](#): Enable use of Oracle Wallets with Oracle Manager.
- [SQOOP-1227](#): Sqoop fails to compile against commons-io higher than 1.4.
- [SQOOP-1223](#): Enhance the password file capability to enable plugging-in custom loaders.
- [SQOOP-1216](#): Improve error message on corrupted input while doing export.
- [SQOOP-435](#): Avro import should write the Schema to a file.
- [SQOOP-1192](#): Add option "--skip-dist-cache" to allow Sqoop not copying jars in %SQOOP_HOME%\lib folder when launched by Oozie and use Oozie share lib.
- [SQOOP-1032](#): Add the --bulk-load-dir option to support the HBase doBulkLoad function.
- [SQOOP-1213](#): Support reading password files from Amazon S3.
- [SQOOP-1203](#): Add another default case for finding *_HOME when not explicitly defined.
- [SQOOP-1197](#): Enable Sqoop to build against Hadoop-2.1.0-beta jar files.
- [SQOOP-1194](#): Make changes to Sqoop build file to enable Netezza third party tests.
- [SQOOP-1167](#): Enhance HCatalog support to allow direct mode connection manager implementations.
- [SQOOP-1190](#): Class HCatHadoopShims will be removed in HCatalog 0.12.
- [SQOOP-1132](#): Print out Sqoop version into log during execution.
- [SQOOP-1137](#): Put a stress in the user guide that eval tool is meant for evaluation purpose only.
- [SQOOP-1107](#): Further improve error reporting when exporting malformed data.
- [SQOOP-1185](#): LobAvroImportTestCase is sensitive to test method order execution.
- [SQOOP-1170](#): Can't import columns with name "public".
- [SQOOP-1179](#): Incorrect warning saying --hive-import was not specified when it was specified.
- [SQOOP-1161](#): Generated Delimiter Set Field Should be Static.
- [SQOOP-1172](#): Make Sqoop compatible with HBase 0.95+.

7.5. Minimum System Requirements

In this section:

- [Hardware Recommendations](#)

- [Operating Systems Requirements](#)
- [Software Requirements](#)
- [Database Requirements](#)
- [Virtualization and Cloud Platforms](#)
- [Configuring the Local Repositories](#)

7.5.1. Hardware recommendations

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

7.5.2. Operating systems requirements

The following operating systems are supported:

- 64-bit Red Hat Enterprise Linux (RHEL) v5.*, v6.*
- 64-bit CentOS v5.*, v6.*



Important

All hosts in the cluster must run the same OS, version and patch sets.

- 64-bit Oracle Linux v5, v6
- 64-bit SUSE Linux Enterprise Server (SLES) 11 SP1

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

7.5.3. Software requirements

On each of your hosts:

- yum
- rpm
- scp
- curl
- wget
- pdsh
- php-curl (Required for SLES installs.)

7.5.4. Database requirements

- Hive and HCatalog require a database to use as a metadata store and come with an embedded Derby database by default.
- Oozie requires a database to use as a metadata store and comes with an embedded Derby database by default.
- Ambari requires a database to use for storing cluster configuration information and comes with an embedded PostgreSQL database by default.

7.5.5. Virtualization and cloud platforms

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

See [Operating Systems Requirements](#) for the list of supported operating systems for HDP.

7.5.6. Configuring the local repositories

If your cluster does not have access to the Internet, or you are creating a large cluster and you want to conserve bandwidth, you need to provide access to the HDP installation packages using an alternative method. For more information, see [Deploying HDP In Production Data Centers](#).



Important

The installer pulls many packages from the base OS repositories. If you do not have a complete base OS available to all your machines at the time of installation, you may run into issues. If you encounter problems due to the unavailability of base OS repositories, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored.

7.6. Improvements

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

- Storm
- Falcon
- Tez
- YARN
- Phoenix
- MapReduce2

- Improved SQL compliance (GRANT, REVOKE, WHERE clause subqueries, common table expressions, data type support)

7.7. Common Vulnerabilities and Exposures

- **CVE-2013-6446:** Apache Hadoop job history server vulnerability
- **Severity:** Major
- **Vendor:** The Apache Software Foundation
- **Versions Affected:** Hadoop 0.23.1 to 0.23.9, Hadoop 2.0.0 to 2.2.0
- **Users Affected:** Users who have enabled Hadoop's MapReduce security features
- **Impact:** Vulnerability allows an unauthorized user to retrieve job details from the job history server
- **Mitigation:** Hadoop 0.23.x users should upgrade to 0.23.10, Hadoop 2.x users should upgrade to 2.3.0
- **Credit:** This issue was discovered by Koji Noguchi of Yahoo

7.8. Known Issues

In this section:

- [Known Limitation for Oracle DB Metastore](#)
- [Known Issues for SLES 11](#)
- [Known Issues for HDP](#)
- [Known Issues for YARN](#)
- [Known Issues for HBase](#)
- [Known Issues for Hive and HCat](#)
- [Known Issues for Tez](#)
- [Known Issues for Oozie](#)
- [Known Issues for Hue](#)
- [Known Issues for Flume](#)
- [Known Issues for Storm](#)
- [Known Issues for Knox](#)
- [Known Issues for Hortonworks Teradata Connector](#)

7.8.1. Known Limitation for HDFS Upgrade

BUG-17431: ([HDFS-5526](#)) If you upgrade from HDP 2.0.9.1 to HDP 2.1, the DataNode cannot be rolled back to HDP 2.0.9.1. Contact Support for an update to fix this HDP 2.0.9.1 issue.

7.8.2. Known Limitation for Oracle DB Metastore

HDP 2.1.1 does not support the use of Oracle DB as a metastore. Upgrade to HDP 2.1.2 or higher.

7.8.3. Known Issues for SLES 11

- **BUG-9904:** php_curl Required for SLES 11 Sp1

Problem: Several alerts return with `Return code of 255 is out of bounds` while trying to install a cluster on SLES because `php_curl` is not installed.

Workaround: Install `php_curl` on your SLES host.

```
zypper install php-curl
```

7.8.4. Known Issues for HDP

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

Problem: RHEL 6.2 and RHEL 6.3 contain known bugs in `nslcd`. On these platforms, we have observed instability and crashes in Hadoop daemons after an LDAP query issued by `nslcd` times out.

Workaround: If your environment encounters this issue, then we recommend applying one of the following workarounds:

- Upgrade to RHEL 6.4 or later.
- Increase the configured `nslcd` search timeout in `/etc/nslcd.conf`.
- Disable Hadoop native code integration for obtaining users' groups by setting `hadoop.security.group.mappingtoorg.apache.hadoop.security.ShellBasedUnixGro` in `core-site.xml`.
- **BUG-825:** EC2 `m1.large` cluster root partition is only 5GB and fills up quickly by HDP logs

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

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

7.8.4.1. Known issues for HDFS

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

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

```
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImageFormatProtobuf: Loaded FSImage in 2 seconds.
2014-03-06 14:03:03,586 INFO org.apache.hadoop.hdfs.server.namenode.
FSImage: Loaded image for txid 0 from C:\hdpdata\hdfs\nn\current\
fsimage_00000000000000000000
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=
false, isRollingUpgrade=false)
2014-03-06 14:03:03,680 INFO org.apache.hadoop.hdfs.server.namenode.
FSEditLog: Starting log segment at 1
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
NameCache: initialized with 0 entries 0 lookups
2014-03-06 14:03:05,273 INFO org.apache.hadoop.hdfs.server.namenode.
FSNamesystem: Finished loading FSImage in 5703 msec
2014-03-06 14:03:08,883 INFO org.apache.hadoop.hdfs.server.namenode.
NameNode: RPC server is binding to VMG22:8020
2014-03-06 14:03:08,898 INFO org.apache.hadoop.ipc.CallQueueManager: Using
callQueue class java.util.concurrent.LinkedBlockingQueue
2014-03-06 14:03:08,930 FATAL org.apache.hadoop.hdfs.server.namenode.
NameNode: Exception in namenode join
java.lang.IllegalArgumentException: No enum const class org.apache.hadoop.
security.SaslRpcServer$QualityOfProtection.NONE
9 more
```

7.8.4.2. Known issues for MapReduce

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

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

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

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

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

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

7.8.5. Known Issues for YARN

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

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

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

Problem: Seeing about 1 GB of logs per hour.

- **BUG-15360:** In HDFS HA mode, Distcp/SLive with webhdfs on secure cluster fails with Client cannot authenticate via:[TOKEN, KERBEROS] error.
- **BUG-13231:** YARN RM won't failover if the RPC port is unreachable.
Problem: YARN does not have a service similar to HDFS where the zkfc process monitors the health of the NameNode. Thus, if the RPC port gets blocked the RM service will not failover.
- **BUG-12327:** [Yarn-90] NM cannot detect when bad disks become healthy again.
Problem: If you start NM with good log-dir, then rename the directory away, the NM will become unhealthy. If you then rename the directory away, NM will be unhealthy. If you rename the directory back, then wait for some period of time (120 sec), NM won't return to healthy state.
- **BUG-7531:** Hadoop metrics link does not contain correct content.
Problem: In the Resource Manager UI Tools section, clicking on Logs and Metrics opens pages that do not contain correct information.

7.8.6. Known Issues for HBase

- **BUG-16900 (HBASE-11036):** HBase Big Linked List with Chaos Monkey Not Serving Region Exception.
Problem: The Big Linked List Test with Chaos Monkey Test run fails with a Not Serving Region exception in the YARN logs.
- **BUG-16513:** HBCK Tests Fail Intermittently Due to NotServingRegionException.
Problem: The HBCK Tool tests fails intermittently due to a NotServingRegionException, noted in the Master Logs.
- **BUG-16257:** HBase master fails to start due to BindException.
Problem: HBase on Suse 11 64 bit, smoke test fails intermittently with:

```
ERROR [main] client.ConnectionManager$HConnectionImplementation: The node /
hbase is not in ZooKeeper.
```
- **BUG-14986:** HBase Bloomberg HA Load Balancer on Windows Env not Run due to Regions Stuck in Transition.
Problem: On the Windows environment, after creating a table with replicas and calling the Load Balancer, the Load Balancer does not run and throws RegionAlreadyInTransitionException in the in the master logs.
- **BUG-12167, HBASE-10304** Running an hbase job jar: IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString
Problem: Some MapReduce jobs fail to launch. An exception similar to the following displays:

```
Exception in thread "main" java.lang.IllegalAccessError: class com.google.protobuf.ZeroCopyLiteralByteString cannot access its superclass com.google.protobuf.LiteralByteString
    at java.lang.ClassLoader.defineClass1(Native Method)
    at java.lang.ClassLoader.defineClass(ClassLoader.java:792)
    at java.security.SecureClassLoader.defineClass(SecureClassLoader.java:142)
    at java.net.URLClassLoader.defineClass(URLClassLoader.java:449)
    at java.net.URLClassLoader.access$100(URLClassLoader.java:71)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:361)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
    at org.apache.hadoop.hbase.protobuf.ProtobufUtil.toScan(ProtobufUtil.java:818)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.convertScanToString(TableMapReduceUtil.java:433)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:186)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:147)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:270)
    at org.apache.hadoop.hbase.mapreduce.TableMapReduceUtil.initTableMapperJob(TableMapReduceUtil.java:100)
    ...
```

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

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

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

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

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

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

7.8.7. Known Issues for Phoenix

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

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

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

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

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

7.8.8. Known Issues for Hive

- **BUG-17603: (HIVE-6985)**(HIVE-6985) Grant privileges on PUBLIC role are not being honored.

Problem: When a privilege is granted to public role, the privilege is supposed to be applicable for all users. However, the privilege check fails for users, even if they have public role in the list of current roles. (This issue only exists with the Public role; it does not affect the granting of privileges to other roles.)

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

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

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

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

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

templeton.hive.archive	hdfs:///apps/webhcat/hive.tar.gz
------------------------	----------------------------------

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

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

- **Workaround:** The workaround for people who wants to run with Tez would be to add "hive.execution.engine=tez" to the templeton.hive.properties. The installer would need to change to accomodate this.
- **BUG-16608:** Oozie table import job fails with error where user hive wants to write to table dir owned by the table owner.

Problem: The job fails with the following permission error:

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

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

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

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

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

Problem: The query fails with the following error:

```
2014-04-16 16:50:13,312 ERROR [pool-7-thread-5]: ql.Driver
(SessionState.java:printError(546)) - FAILED: HiveAccessControlException
```



```

Permission denied. Principal [name=hrt_qa, type=USER] does not have
following privileges on Object [type=TABLE_OR_VIEW, name=default.
default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT OWNERSHIP]
org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAccessControlException: Permission denied. Principal [name=hrt_qa, type=
USER] does not have following privileges on Object [type=TABLE_OR_VIEW,
name=default__missing_ddl_3_missing_ddl_3_index__] : [OBJECT
OWNERSHIP]
at org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.
SQLAuthorizationUtils.assertNoMissingPrivilege(SQLAuthorizationUtils.
java:361)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:105)
at org.apache.hadoop.hive.ql.security.authorization.
plugin.sqlstd.SQLStdHiveAuthorizationValidator.
checkPrivileges(SQLStdHiveAuthorizationValidator.java:77)
at org.apache.hadoop.hive.ql.security.authorization.plugin.
HiveAuthorizerImpl.checkPrivileges(HiveAuthorizerImpl.java:84)
at org.apache.hadoop.hive.ql.Driver.doAuthorizationV2(Driver.java:695)
at org.apache.hadoop.hive.ql.Driver.doAuthorization(Driver.java:510)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:462)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:322)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:976)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:969)
at org.apache.hive.service.cli.operation.SQLOperation.prepare(SQLOperation.
java:99)
at org.apache.hive.service.cli.operation.SQLOperation.run(SQLOperation.
java:172)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementInternal(HiveSessionImpl.java:231)
at org.apache.hive.service.cli.session.HiveSessionImpl.
executeStatementAsync(HiveSessionImpl.java:218)
at org.apache.hive.service.cli.CLIService.executeStatementAsync(CLIService.
java:233)
at org.apache.hive.service.cli.thrift.ThriftCLIService.
ExecuteStatement(ThriftCLIService.java:346)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1313)
at org.apache.hive.service.cli.thrift.TCLIService$Processor
$ExecuteStatement.getResult(TCLIService.java:1298)
at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at org.apache.hive.service.auth.TSetIpAddressProcessor.
process(TSetIpAddressProcessor.java:55)
at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.
run(TThreadPoolServer.java:206)
at java.util.concurrent.ThreadPoolExecutor$Worker.
runTask(ThreadPoolExecutor.java:886)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.
java:908)
at java.lang.Thread.run(Thread.java:662)
...

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Problem:** The error returned on the Hive console is:

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

```

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

```

- **BUG-13796:** When running with correlation optimization enabled on Tez, TPCDS queries 1, 32, 94, 95 and 97 fail with ClassCastException.
- **BUG-8227:** Hive needs to implement recovery or extend FileOutputComitter.

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

```

public boolean isRecoverySupported() {
 return false;
}

```

### 7.8.9. Known Issues for Tez

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

**Problem:** Seeing about 1 GB of logs per hour.

### 7.8.10. Known Issues for Oozie

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

**Problem:** The job fails with the following permission error:

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

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

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

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

- **BUG-16397:** Documentation does not describe how to renew an expired Oozie authorization token.

**Problem:** Due to the introduction of [HADOOP-10416](#), users may encounter situations where they are not authorized to perform certain actions because they are in an unsecured environment and their Oozie authorization token has expired.

**Workaround:** Remove the oozie auth token cache file at `System.getProperty("user.home")/.oozie-auth-token`, then re-run the Oozie command to renew the token.

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

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

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

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

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

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

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

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

**Problem:** From the Oozie log:

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

But this job and the application complete successfully.

## 7.8.11. Known Issues for Hue

- **BUG-9734:** Data loss during Migration of Hue DB from default (sqlite) to Oracle DB:

**Problem:** Migration of data and tables from SQLite to Oracle does not work and needs to be performed manually.

1. Install Hue and start Hue, (Hue creates table in sqlite db).
2. Do NOT perform any tasks (such as uploading files, pig jobs, or hcat jobs) on the HDP stack from Hue UI.
3. Stop Hue, configure Oracle.
4. Start Hue.

Result: Hue starts fine and continues working, but there is loss of data.

5. Some tables are lost in HCatalog.
6. Some pig scripts do not show up on UI.

**Workaround:** Manually migrate the data and tables from SQLite to Oracle.

## 7.8.12. Known Issues for Flume

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

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

## 7.8.13. Known Issues for Storm

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

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

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

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

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

```
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [INFO] State change:
CONNECTED
2014-04-01 20:59:11 c.n.c.f.s.ConnectionStateManager [WARN] There are no
ConnectionStateListeners registered.
2014-04-01 20:59:11 b.s.d.worker [ERROR] Error on initialization of server
mk-worker
java.io.FileNotFoundException: File '/home/storm/supervisor/stormdist/
myPersistentWordCount-15-1396385521/stormconf.ser' does not exist
at org.apache.commons.io.FileUtils.openInputStream(FileUtils.java:299)
~[commons-io-2.4.jar:2.4]
at org.apache.commons.io.FileUtils.readFileToByteArray(FileUtils.java:1763)
~[commons-io-2.4.jar:2.4]
at backtype.storm.config$read_supervisor_storm_conf.invoke(config.clj:192)
~[storm-core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:170) ~[storm-
core-0.9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at backtype.storm.daemon.worker$eval4415$exec_fn__1103__auto____4416.
invoke(worker.clj:353) ~[na:na]
at clojure.lang.AFn.applyToHelper(AFn.java:185) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at clojure.core$apply.invoke(core.clj:601) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$eval4415$mk_worker__4471.doInvoke(worker.
clj:344) ~[na:na]
at clojure.lang.RestFn.invoke(RestFn.java:512) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker$_main.invoke(worker.clj:454) ~[storm-core-0.
9.1.2.1.1.0-290.jar:0.9.1.2.1.1.0-290]
at clojure.lang.AFn.applyToHelper(AFn.java:172) ~[clojure-1.4.0.jar:na]
at clojure.lang.AFn.applyTo(AFn.java:151) ~[clojure-1.4.0.jar:na]
at backtype.storm.daemon.worker.main(Unknown Source) ~[storm-core-0.9.1.2.1.
1.0-290.jar:0.9.1.2.1.1.0-290]
2014-04-01 20:59:11 b.s.util [INFO] Halting process: ("Error on
initialization")
```

## 7.8.14. Known Issues for Falcon

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

**Problem:** Falcon generated hive-action does not pass the hive-site.xml with the right configuration parameters. One manifestation of the problem will be the failure in table import job where user "hive" will be used to write to a directory owned by the table

owner. This is because `hive.metastore.execute.setugi` parameter is not being passed as part of the hive action.

**Workaround:** Add a Hive default configuration to Oozie.

Stop the Oozie service.



### Warning

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

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

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

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

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

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

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

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

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

**Workaround:** Modify Oozie on all clusters managed by Falcon:

1. Stop the Oozie service on all Falcon clusters.
2. Copy each cluster's hadoop conf directory to a different location. For example, if you have two clusters, copy one to `/etc/hadoop/conf-1` and the other to `/etc/hadoop/conf-2`.
3. For each `oozie-site.xml` file, modify the `oozie.service.HadoopAccessorService.hadoop.configurations` property, specifying clusters, the RPC ports of the NameNodes and HostManagers accordingly.

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

```

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

```

4. Restart the Oozie service on all clusters.

## 7.8.15. Known Issues for Knox

- **BUG-16592:** When accessing through Knox Gateway, Oozie not supported for HDP for Windows or Linux single-node clusters.

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

**Workaround:** Either use Oozie only when accessing a multinode HDP cluster in Linux, or access Oozie without going through Knox.

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

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

## 7.8.16. Known Issues for the Hortonworks Connector for Teradata

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

## 7.8.17. Known Issues for Ambari

- See [Ambari Release Notes](#).



## 7.9. Deprecated Features

- Oracle JDK 6 is deprecated in this release.

## 7.10. Third-party Licenses

Table 7.1. Third-party Licenses

HDP Component	License
Phoenix	<a href="#">EPL</a>
Storm	<a href="#">EPL</a>
Accumulo	<a href="#">JCommander</a>
Falcon	<a href="#">CERN</a>
Falcon	<a href="#">Tinkerpop</a>
Knox	<a href="#">ANTLR</a>
Knox	<a href="#">MIT</a>
Knox	<a href="#">EPL</a>
Knox	<a href="#">Bouncy Castle</a>
Knox	<a href="#">OWS</a>