Hortonworks Data Platform

Upgrading HDP Manually

(Jun 9, 2015)

docs.hortonworks.com

Hortonworks Data Platform: Upgrading HDP Manually

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner. The Hortonworks Data Platform consists of the essential set of Apache Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, Zookeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

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

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

Table of Contents

List of Tables

1. Upgrade from HDP 2.1 to HDP 2.2 Manually

HDP 2.2 supports side-by-side installation of HDP 2.2 and subsequent releases, which lets you perform rolling upgrades on your cluster and improve execution times for in-place upgrade. To support side-by-side installation, the HDP package version naming convention for both RPMs and Debs has changed to include the HDP 2.2 product version. For example, hadoop-hdfs is now hadoop-2.2.6.0-hdfs. HDP 2.2 marks the first release where HDP rpms, debs, and directories contain versions in the names to permit side-by-side installations of later HDP releases.

To select from the releases you have installed side-by-side, Hortonworks provides hdp-select, a command that lets you select the active version of HDP from the versions you have selected.

The HDP packages for a complete installation of HDP 2.2 will take about 2.5 GB of disk space.

Warning

Until the upgrade is finalized, no HDFS data is deleted from the cluster. Be sure to review your capacity and ensure that you have extra space available during the upgrade window.

Important

You cannot perform a rolling upgrade from 2.1 to 2.2; only upgrade in place is supported. Rolling upgrade support starts with releases HDP 2.2 and above.

This document provides instructions on how to upgrade to HDP 2.2 from the HDP 2.1 release.

Note

These instructions cover the upgrade between two minor releases. If you need to upgrade between two maintenance releases, follow the upgrade instructions in the HDP Release Notes.

Here is an overview of steps for upgrading to the latest release of HDP 2.2 from HDP 2.1:

1. Get ready to upgrade
2. Upgrade HDP 2.1 Components
3. Symlink Directories with hdp-select
4. Configure and Start Apache ZooKeeper
5. Configure and Start Hadoop
6. Start HDFS

7. Configure and Start Apache HBase
8. Configure and Start Apache Phoenix
9. Configure and Start Apache Accumulo
10. Configure and Start Apache Tez
11. Configure and Start Apache Hive and Apache HCatalog
12. Configure and Start Apache Oozie
13. Configure and Start Apache WebHCat (Templeton)
14. Configure and Start Apache Pig
15. Configure and Start Apache Sqoop
16. Configure and Start Apache Flume
17. Configure and Start Apache Mahout
18. Configure and Start Hue
19. Configure and Start Apache Knox
20. Configure and Start Apache Falcon
21. Finalize the Upgrade
22. Install new HDP 2.2 services

1.1. Getting Ready to Upgrade

HDP Stack upgrade involves upgrading from HDP 2.1 to HDP 2.2 versions and adding the new HDP 2.2 services. These instructions change your configurations.

Hardware recommendations

Although there is no single hardware requirement for installing HDP, there are some basic guidelines. The HDP packages for a complete installation of HDP 2.2 will take up about 2.5 GB of disk space.

The first step is to ensure you keep a backup copy of your HDP 2.1 configurations.

1. Back up the following HDP directories:
   - `/etc/hadoop/conf`
   - `/etc/hbase/conf`
   - `/etc/hive-hcatalog/conf`
   - `/etc/hive-webhcat/conf`

- /etc/accumulo/conf
- /etc/hive/conf
- /etc/pig/conf
- /etc/sqoop/conf
- /etc/flume/conf
- /etc/mahout/conf
- /etc/oozie/conf
- /etc/hue/conf
- /etc/zookeeper/conf
- /etc/tez/conf
- /etc/storm/conf
- Optional - Back up your userlogs directories, ${mapred.local.dir}/userlogs.

2. Run the fsck command as the HDFS Service user and fix any errors. (The resulting file contains a complete block map of the file system.)

   ```
   su -l $HDFS_USER

   hdfs fsck / -files -blocks -locations > dfs-old-fsck-1.log
   ```

   Where $HDFS_USER is the HDFS Service user. For example, hdfs.

3. Use the following instructions to compare status before and after the upgrade:

   The following commands must be executed by the user running the HDFS service (by default, the user is hdfs).

   a. Capture the complete namespace of the file system. (The following command does a recursive listing of the root file system.)

      ```
      su -l $HDFS_USER

      hdfs dfs -ls -R / > dfs-old-lsr-1.log
      ```

      where $HDFS_USER is the HDFS Service user. For example, hdfs.

      

      Note

      In secure mode you must have kerberos credentials for the hdfs user.

   b. Run the report command to create a list of DataNodes in the cluster.

      ```
      su -l $HDFS_USER
      ```

```
hdfs dfsadmin -report > dfs-old-report-1.log
```

where *$HDFS_USER* is the HDFS Service user. For example, hdfs.

c. **Optional:** You can copy all or unrecoverable only data storelibext-customer directory in HDFS to a local file system or to a backup instance of HDFS.

d. **Optional:** You can also repeat the steps 3 (a) through 3 (c) and compare the results with the previous run to ensure the state of the file system remained unchanged.

4. Save the namespace by executing the following commands:

```
su -l $HDFS_USER

hdfs dfsadmin -safemode enter

hdfs dfsadmin -saveNamespace
```

5. Backup your NameNode metadata.

a. Copy the following checkpoint files into a backup directory:

- `dfs.name.dir/edits`
- `dfs.namenode.name.dir/image/fsimage`
- `dfs.namenode.name.dir/current/fsimage`

b. Store the layoutVersion of the namenode.

```
${dfs.namenode.name.dir}/current/VERSION
```

6. Finalize any prior HDFS upgrade, if you have not done so already.

```
su -l $HDFS_USER

hdfs dfsadmin -finalizeUpgrade
```

7. **Optional:** Back up the Hive Metastore database.

The following instructions are provided for your convenience. For the latest backup instructions, please see your database documentation.

Table 1.1. Hive Metastore Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sqlsbr For example: mysqldump hive > /tmp/mydir/backup_hive.sql	mysql $dbname < $inputfilename.sqlsbr For example: mysql hive < /tmp/mydir/backup_hive.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql sbr	sudo -u $username psql $databasename < $inputfilename.sqlsbr

Database Type	Backup	Restore
	For example: sudo -u postgres pg_dump hive > /tmp/mydir/backup_hive.sql	For example: sudo -u postgres psql hive < /tmp/mydir/backup_hive.sql
Oracle	Export the database: exp username/password@database full=yes file=output_file.dmp	Export the database: exp username/password@database full=yes file=output_file.dmp

8. Optional: Back up the Oozie metastore database.

   These instructions are provided for your convenience. Please check your database documentation for the latest backup instructions.

Table 1.2. Oozie Metastore Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sql For example: mysqldump oozie > /tmp/mydir/backup_hive.sql	mysql $dbname < $inputfilename.sql For example: mysql oozie < /tmp/mydir/backup_oozie.sql

Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql
	
	For example:
	
	sudo -u
	
	postgres pg_dump oozie > /tmp/mydir/backup_oozie.sql
		sudo -u $username psql $databasename < $inputfilename.sql
		
		For example:
		
		`sudo -u`
		
		`postgres psql oozie < /tmp/mydir/backup_oozie.sql`

Oracle	Export the database: exp username/password@database full=yes file=output_file.dmp	Export the database: exp username/password@database full=yes file=output_file.dmp

9. **Optional**: Back up the Hue database.

   The following instructions are provided for your convenience. For the latest backup instructions, please see your database documentation. For database types that are not listed below, follow your vendor-specific instructions.

Table 1.3. Hue Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sqlsbr For example: mysqldump hue > /tmp/mydir/backup_hue.sql	mysql $dbname < $inputfilename.sqlsbr For example: mysql hue < /tmp/mydir/backup_hue.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql sbr For example:	sudo -u $username psql $databasename < $inputfilename.sqlsbr For example:

Database Type	Backup	Restore
	sudo -u postgres pg_dump hue > /tmp/mydir/backup_hue.sql	sudo -u postgres psql hue < /tmp/mydir/backup_hue.sql
Oracle	Connect to the Oracle database using sqlplus. Export the database. For example: exp username/password@database full=yes file=output_file.dmp mysql $dbname < $inputfilename.sqlsbr	Import the database: For example: imp username/password@database file=input_file.dmp
SQLite	/etc/init.d/hue stop su $HUE_USER mkdir ~/hue_backup sqlite3 desktop.db .dump > ~/hue_backup/desktop.bak /etc/init.d/hue start	/etc/init.d/hue stop cd /var/lib/hue mv desktop.db desktop.db.old sqlite3 desktop.db < ~/hue_backup/desktop.bak /etc/init.d/hue start

10. Stop all services (including MapReduce) and client applications deployed on HDFS:

Component	Command
Knox	`cd $GATEWAY_HOME su -l knox -c "bin/gateway.sh stop"`

Oozie	`su $OOZIE_USER`
	
	`/usr/lib/oozie/bin/oozied.sh stop`

WebHCat	`su -l hcat -c "/usr/lib/hcatalog/sbin/webhcat_server.sh stop"`
Hive	Run this command on the Hive Metastore and Hive Server2 host machine: `ps aux | awk '{print $1,$2}' | grep hive | awk '{print $2}' | xargs kill >/dev/null 2>&1`
HBase RegionServers	`su -l hbase -c "/usr/lib/hbase/bin/hbase-daemon.sh --config /etc/hbase/conf stop regionserver"`
HBase Master host machine	`su -l hbase -c "/usr/lib/hbase/bin/hbase-daemon.sh --config /etc/hbase/conf stop master"`

YARN	Run this command on all NodeManagers:
	
	`su -l yarn -c "export HADOOP_LIBEXEC_DIR=/usr/lib/hadoop/libexec && /usr/lib/hadoop-yarn/sbin/yarn-daemon.sh --config /etc/hadoop/conf stop nodemanager"`
	
	Run this command on the History Server host machine:
	
	`su -l mapred -c "export HADOOP_LIBEXEC_DIR=/usr/lib/hadoop/libexec && /usr/lib/hadoop-mapreduce/sbin/mr-jobhistory-daemon.sh --config /etc/hadoop/conf stop historyserver"`
	
	Run this command on the ResourceManager host machine(s):

Component	Command

	`su -l yarn -c "export HADOOP_LIBEXEC_DIR=/usr/lib/hadoop/libexec && /usr/lib/hadoop-yarn/sbin/yarn-daemon.sh --config /etc/hadoop/conf stop resourcemanager"`
	
	Run this command on the YARN Timeline Server node:
	
	`su -l yarn -c "export HADOOP_LIBEXEC_DIR=/usr/lib/hadoop/libexec && /usr/lib/hadoop-yarn/sbin/yarn-daemon.sh --config /etc/hadoop/conf stop timelineserver"`

HDFS	On all DataNodes:
	
	If you are running secure cluster, run following command as root:
	
	`/usr/hdp/current/hadoop-client/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop datanode`
	
	Else:
	
	`su -l hdfs -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop datanode"`
	
	If you are not running NameNode HA (High Availability), stop the Secondary NameNode by executing this command on the Secondary NameNode host machine:
	
	`su -l hdfs -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop secondarynamenode”`
	
	On the NameNode host machine(s):
	
	`su -l hdfs -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop namenode"`
	
	If you are running NameNode HA, stop the ZooKeeper Failover Controllers (ZKFC) by executing this command on the NameNode host machine:
	
	`su -l hdfs -c "/usr/lib/hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop zkfc"`
	
	If you are running NameNode HA, stop the JournalNodes by executing these commands on the JournalNode host machines:
	
	`su hdfs /usr/lib/hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf stop journalnode`

ZooKeeper Host machines	`su - zookeeper -c "export ZOOCFGDIR=/etc/zookeeper/conf ; export ZOOCFG=zoo.cfg ;source /etc/zookeeper/conf/zookeeper-env.sh ; /usr/lib/zookeeper/bin/zkServer.sh stop"`

Ranger (XA Secure)	`service xapolicymgr stop`
	
	`service uxugsync stop`

11. Verify that edit logs in `${dfs.namenode.name.dir}/current/edits*` are empty.

    a. Run: `hdfs oev -i ${dfs.namenode.name.dir}/current/edits_inprogress_* -o edits.out`

    b. Verify edits.out file. It should only have OP_START_LOG_SEGMENT transaction. For example:

    ```
    <?xml version="1.0" encoding="UTF-8"?>
    <EDITS>
    <EDITS_VERSION>-56</EDITS_VERSION>
    <RECORD>
    <OPCODE>OP_START_LOG_SEGMENT</OPCODE>
    <DATA>
    <TXID>5749</TXID>
    </DATA>
    </RECORD>
    ```

    c. If edits.out has transactions other than OP_START_LOG_SEGMENT, run the following steps and then verify edit logs are empty.

    - Start the existing version NameNode.
    - Ensure there is a new FS image file.
    - Shut the NameNode down:

      `hdfs dfsadmin – saveNamespace`

12. Rename or delete any paths that are reserved in the new version of HDFS.

    When upgrading to a new version of HDFS, it is necessary to rename or delete any paths that are reserved in the new version of HDFS. If the NameNode encounters a reserved path during upgrade, it will print an error such as the following:

    ```
    /.reserved is a reserved path and .snapshot is a reserved path
    component in this version of HDFS.

    Please rollback and delete or rename this path, or upgrade with the
    -renameReserved key-value pairs option to automatically rename these
    paths during upgrade.
    ```

    Specifying `-upgrade -renameReserved` optional key-value pairs causes the NameNode to automatically rename any reserved paths found during startup.

    For example, to rename all paths named `.snapshot` to `.my-snapshot` and change paths named `.reserved` to `.my-reserved`, specify `-upgrade -renameReserved .snapshot=.my-snapshot,.reserved=.my-reserved`.

    If no key-value pairs are specified with `-renameReserved`, the NameNode will then suffix reserved paths with:

    `.<LAYOUT-VERSION>.UPGRADE_RENAMED`

    For example: `.snapshot.-51.UPGRADE_RENAMED`.

Note

We recommend that you perform a `-saveNamespace` before renaming paths (running `-saveNamespace` appears in a previous step in this procedure). This is because a data inconsistency can result if an edit log operation refers to the destination of an automatically renamed file.

Also note that running `-renameReserved` will rename all applicable existing files in the cluster. This may impact cluster applications.

1.2. Upgrade HDP 2.1 Components

The upgrade process to HDP 2.2 involves the following steps. Select your OS:

RHEL/CentOS/Oracle 6

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 2.1 components. This command uninstalls the HDP 2.1 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   yum erase "hadoop*" "webhcat*" "oozie*" "collectd*" "gccxml*" "pig*" "hdfs*"
    "sqoop*" "zookeeper*" "hbase*" "hive*" "tez*" "storm*" "falcon*" "flume*"
    "phoenix*" "accumulo*" "mahout*" "hue" "hue-common" "hue-shell" "knox*"
    "hdp_mon_nagios_addons"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/yum.repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.
     6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
     ```

   - Confirm the HDP repository is configured.

     ```
     yum repolist
     ```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

     ```
     Loaded plugins: fastestmirror, security
     Loading mirror speeds from cached hostfile
     * base: mirrors.cat.pdx.edu
     * extras: linux.mirrors.es.net
     * updates: mirrors.usc.edu
     repo id repo namestatus
     HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

```
yum install "hadoop" "hadoop-hdfs" "hadoop-libhdfs" "hadoop-yarn"
 "hadoop-mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie"
 "collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hue" "hive"
 "tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout" "knox"
 "hdp_mon_nagios_addons"
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

6. Verify that the components were upgraded.

   ```
   yum list installed | grep HDP-<old.stack.version.number>
   ```

   No component file names should appear in the returned list.

RHEL/CentOS/Oracle 5 (Deprecated)

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 2.1 components. This command uninstalls the HDP 2.1 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   yum erase "hadoop*" "webhcat*" "oozie*" "collectd*" "gccxml*" "pig*" "hdfs*"
    "sqoop*" "zookeeper*" "hbase*" "hive*" "tez*" "storm*" "falcon*" "flume*"
    "phoenix*" "accumulo*" "mahout*" "hue" "hue-common" "hue-shell" "knox*"
    "hdp_mon_nagios_addons"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/yum.repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/2.x/updates/2.2.
     6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
     ```

   - Confirm the HDP repository is configured.

     ```
     yum repolist
     ```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

     ```
     Loaded plugins: fastestmirror, security
     Loading mirror speeds from cached hostfile
     * base: mirrors.cat.pdx.edu
     * extras: linux.mirrors.es.net
     * updates: mirrors.usc.edu
     repo id repo namestatus
     HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

```
yum install "hadoop" "hadoop-hdfs" "hadoop-libhdfs" "hadoop-yarn"
 "hadoop-mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie"
 "collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "hue"
 "tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout" "knox"
 "hdp_mon_nagios_addons"
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

6. Verify that the components were upgraded.

```
yum list installed | grep HDP-<old.stack.version.number>
```

No component file names should appear in the returned list.

SLES 11 SP 1

1. On all hosts, clean the yum repository.

```
zypper clean -all
```

2. Remove your old HDP 2.1 components. This command uninstalls the HDP 2.1 components. It leaves the user data, and metadata, but removes your configurations:

```
zypper rm "hadoop*" "webhcat*" "oozie*" "collectd*" "gccxml*" "pig*" "hdfs*"
 "sqoop*" "zookeeper*" "hbase*" "hive*" "tez*" "storm*" "falcon*" "flume*"
 "phoenix*" "accumulo*" "mahout*" "hue" "hue-common" "hue-shell" "knox*"
 "hdp_mon_nagios_addons"
```

3. Remove your old hdp.repo file:

```
rm /etc/zypp/repos.d/hdp.repo
```

4. Download the HDP 2.2 hdp.repo file:

```
wget -nv http://public-repo-1.hortonworks.com/HDP/sles11sp1/2.x/updates/2.2.
6.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo
```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

```
zypper install "hadoop" "hadoop-hdfs" "hadoop-libhdfs" "hadoop-yarn"
 "hadoop-mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie"
 "collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "hue"
 "tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout" "knox"
 "hdp_mon_nagios_addons"
```

```
zypper install webhcat-tar-hive webhcat-tar-pig
```

```
zypper up -r HDP-2.2.6.0
```

```
zypper install oozie-client
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

6. Verify that the components were upgraded. For example, to verify hdfs, hive, and hcatlog:

   ```
   rpm -qa | grep hdfs, && rpm -qa | grep hive && rpm -qa | grep
   hcatalog
   ```

   No component files names should appear in the returned list.

SLES 11 SP3/SP4

1. On all hosts, clean the yum repository.

   ```
   zypper clean -all
   ```

2. Remove your old HDP 2.1 components.

   ```
   zypper rm "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*" "gccxml*"
    "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*" "tez*" "storm*"
    "falcon*" "flume*" "phoenix*" "accumulo*" "mahout*" "hue" "hue-common"
    "hue-shell" "knox*" "hdp_mon_nagios_addons"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/zypp/repos.d/hdp.list
   ```

4. Download the HDP 2.2 hdp.repo file:

   ```
   http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/
   updates/2.2.6.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo
   ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

   ```
   zypper install "hadoop" "hadoop-hdfs" "hadoop-libhdfs" "hadoop-yarn"
    "hadoop-mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie"
    "collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hue" "hive"
    "tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout" "knox"
    "hdp_mon_nagios_addons"
   ```

   ```
   zypper install webhcat-tar-hive webhcat-tar-pig
   ```

   ```
   zypper up -r HDP-2.2.6.0
   ```

   ```
   zypper install oozie-clientt
   ```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

6. Verify that the components were upgraded. For example, to verify hdfs, hive, and hcatalog:

   ```
   rpm -qa | grep hdfs, && rpm -qa | grep hive && rpm -qa | grep
   hcatalog
   ```

   No component files names should appear in the returned list.

Ubuntu 12

1. On all hosts, clean the apt-get repository.

   ```
   apt-get clean --all
   ```

2. Remove your old HDP 2.1 components. This command uninstalls the HDP 2.1 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   apt-get remove "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
    "gccxml*" "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*" "tez.
   *" "storm*" "falcon*" "flume*" "phoenix*" "accumulo*" "mahout*" "hue.*"
    "knox*" "hdp_mon_nagios_addons" --purge
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/apt/sources.list.d/hdp.list
   ```

4. Download the HDP 2.2 hdp.repo file:

   ```
   wget http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/
   updates/2.2.6.0/hdp.list - O /etc/apt/sources.list.d/hdp.list
   ```

5. Run an update:

   ```
   apt-get update
   ```

6. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

   ```
   apt-get install "hadoop" "hadoop-hdfs" "libhdfs0" "hadoop-yarn" "hadoop-
   mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie" "collectd"
    "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "tez" "storm" "falcon"
    "flume" "phoenix" "accumulo" "mahout" "knox" "hdp_mon_nagios_addons"
   ```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

Debian 6

1. On all hosts, clean the apt-get repository.

   ```
   apt-get clean --all
   ```

2. Remove your old HDP 2.1 components. This command uninstalls the HDP 2.1 components. It leaves the user data, and metadata, but removes your configurations:

```
apt-get remove "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
 "gccxml*" "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*" "tez*"
 "storm*" "falcon*" "flume*" "phoenix*" "accumulo*" "mahout*" "hue*" "knox*"
 "hdp_mon_nagios_addons"
```

3. Remove your old hdp.repo file:

   ```
   rm /etc/apt/sources.list.d/hdp.list
   ```

4. Download the HDP 2.2 hdp.repo file:

   ```
   wget -nv http://public-repo-1.hortonworks.com/HDP/debian6/2.x/
   updates/2.2.6.0/hdp.list - O /etc/apt/sources.list.d/hdp.list
   ```

5. Run an update:

   ```
   apt-get update
   ```

6. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.1 components:

   ```
   apt-get install "hadoop" "hadoop-hdfs" "libhdfs0" "hadoop-yarn" "hadoop-
   mapreduce" "hadoop-client" "openssl" "webhcat" "hcatalog" "oozie" "collectd"
    "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "tez" "storm" "falcon"
    "flume" "phoenix" "accumulo" "mahout" "knox" "hdp_mon_nagios_addons"
   ```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path.

1.3. Symlink Directories with hdp-select

Warning

HDP 2.2 installs hdp-select automatically with the installation or upgrade of the first HDP component. If you have not already upgraded ZooKeeper, hdp-select has not been installed.

To prevent version-specific directory issues for your scripts and updates, Hortonworks provides hdp-select, a script that symlinks directories to hdp-current and modifies paths for configuration directories.

1. Before you run hdp-select, remove one link:

   ```
   rm /usr/bin/oozie
   ```

2. Run hdp-select set all on your NameNode and all your DataNodes:

   ```
   hdp-select set all 2.2.6.0-<$version>
   ```

   For example:

   ```
   /usr/bin/hdp-select set all 2.2.6.0-2800
   ```

1.4. Configure and Start Apache ZooKeeper

Tip

If you are running a highly available cluster, configure and restart Apache ZooKeeper **before** you upgrade HDFS. This best practice lets the upgraded ZKFC work with your primary NameNode and your Standby NameNode.

1. Replace your configuration after upgrading on all the ZooKeeper nodes. Replace the ZooKeeper template configuration in `/etc/zookeeper/conf`.
2. Start ZooKeeper.

   On all ZooKeeper server host machines, run the following command to start ZooKeeper and the ZKFC:

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

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

1.5. Configure Hadoop

RHEL/CentOS/Oracle Linux

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/CPU settings in yarn-site.xml and mapred-site.xml
2. Paths have changed in HDP 2.2. Make sure you remove old path specifications from hadoop-env.sh, such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load, as this is not where lzo is installed.

SLES

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/CPU settings in yarn-site.xml and mapred-site.xml
2. Paths have changed in HDP 2.2. make sure you remove old path specifications from hadoop-env.sh, such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load, as this is not where lzo is installed.

Ubuntu/Debian

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/CPU settings in yarn-site.xml and mapred-site.xml

2. Paths have changed in HDP 2.2. make sure you remove old path specifications from hadoop-env.sh, such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load, as this is not where lzo is installed.

1.6. Start Hadoop Core

Warning

Before you start HDFS on an HA system you must start the ZooKeeper service. If you do not start the ZKFC, there can be failures.

To start HDFS, run commands as the $HDFS_USER.

1. If you are upgrading from an HA NameNode configuration, start all JournalNodes. On each JournalNode host, run the following commands:

   ```
   su - hdfs

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

   

   ### Important

   All JournalNodes must be running when performing the upgrade, rollback, or finalization operations. If any JournalNodes are down when running any such operation, the operation fails.

2. Start the NameNode.

   Because the file system version has now changed you must start the NameNode manually. On the active NameNode host, run the following commands:

   ```
   su - hdfs

   /usr/hdp/current/hadoop-client/sbin/hadoop-daemon.sh start namenode -upgrade
   ```

   On a large system, this can take a long time to complete.

   

   ### Note

   Run this command with the -upgrade option only once. After you have completed this step, you can bring up the NameNode using this command without including the -upgrade option.

To check if the Upgrade is in progress, check that the "\previous" directory has been created in the \NameNode and \JournalNode directories. The "\previous" directory contains a snapshot of the data before upgrade.

In a NameNode HA configuration, this NameNode will not enter the standby state as usual. Rather, this NameNode will immediately enter the active state, perform an upgrade of its local storage directories, and also perform an upgrade of the shared edit log. At this point, the standby NameNode in the HA pair is still down. It will be out of sync with the upgraded active NameNode.

To synchronize the active and standby NameNode, re-establishing HA, re-bootstrap the standby NameNode by running the NameNode with the '-bootstrapStandby' flag. Do NOT start this standby NameNode with the '-upgrade' flag.

```
su -l <HDFS_USER> -c "hdfs namenode -bootstrapStandby -force"
```

where <HDFS_USER> is the HDFS service user. For example, hdfs.

The bootstrapStandby command will download the most recent fsimage from the active NameNode into the $dfs.name.dir directory of the standby NameNode. You can enter that directory to make sure the fsimage has been successfully downloaded. After verifying, start the ZKFailoverController, then start the standby NameNode. You can check the status of both NameNodes using the Web UI.

3. Verify that the NameNode is up and running:

   ```
   ps -ef|grep -i NameNode
   ```

4. Start the Secondary NameNode. On the Secondary NameNode host machine, run the following commands:

   ```
   su - hdfs

   /usr/hdp/current/hadoop-client/sbin/hadoop-daemon.sh start
   secondarynamenode
   ```

5. Verify that the Secondary NameNode is up and running:

   ```
   ps -ef|grep SecondaryNameNode
   ```

6. Start DataNodes.

   On each of the DataNodes, enter the following command. Note: If you are working on a `non-secure` DataNode, use $HDFS_USER. For a secure DataNode, use root.

   ```
   su - hdfs

   /usr/hdp/current/hadoop-client/sbin/hadoop-daemon.sh start
   datanode
   ```

7. Verify that the DataNode process is up and running:

   ```
   ps -ef|grep DataNode
   ```

8. Verify that NameNode can go out of safe mode.

   ```
   >hdfs dfsadmin -safemode wait

   Safemode is OFF
   ```

   In general, it takes 5-10 minutes to get out of safemode. For thousands of nodes with millions of data blocks, getting out of safemode can take up to 45 minutes.

1.7. Verify HDFS filesystem health

Analyze if the filesystem is healthy.

1. Run the fsck command on namenode as $HDFS_USER:

   ```
   hdfs fsck / -files -blocks -locations > dfs-new-fsck-1.log
   ```

2. Run `hdfs` namespace and report.

   a. List directories.

      ```
      hdfs dfs -ls -R / > dfs-new-lsr-1.log
      ```

   b. Run report command to create a list of DataNodes in the cluster.

      ```
      hdfs dfsadmin -report > dfs-new-report-1.log
      ```

3. Compare the namespace report before the upgrade and after the upgrade. Verify that user files exist after upgrade.

   ```
   dfs-old-fsck-1.log < -- > dfs-new-fsck-1.log

   dfs-old-lsr-1.log < -- > dfs-new-lsr-1.log
   ```

   

   Note

   You must do this comparison manually to catch all errors.

4. From the NameNode WebUI, see if all DataNodes are up and running.

   ```
   http://<namenode>:50070
   ```

1.8. Configure YARN and MapReduce

After you upgrade Hadoop, complete the following steps to update your configs.

1. Upload the MapReduce tarball to HDFS. As the HDFS user, for example 'hdfs':

   ```
   hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/mapreduce/

   hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hadoop/
   mapreduce.tar.gz /hdp/apps/2.2.6.0-<$version>/mapreduce/

   hdfs dfs -chown -R hdfs:hadoop /hdp
   ```

```
hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/mapreduce

hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/mapreduce/
mapreduce.tar.gz
```

2. Make the following changes to mapred-site.xml:

   - Add:

```
<property>
 <name>mapreduce.admin.map.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true
   -Dhdp.version=${hdp.version}
   </value>
 <final>true</final>
</property>

<property>
<name>yarn.app.mapreduce.am.admin-command-opts</name>
<value>-Dhdp.version=$
{hdp.version}

</value>
</property>
```

Note

You do not need to modify ${hdp.version}.

   - Modify the following existing properties to include ${hdp.version}:

```
<property>
 <name>mapreduce.admin.user.env</name>
 <value>LD_LIBRARY_PATH=/usr/hdp/${hdp.version}
   /hadoop/lib/native:/usr/hdp/${hdp.version}/hadoop/
     lib/native/Linux-amd64-64
   </value>
</property>

</property>
 <name>mapreduce.admin.map.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true
   -Dhdp.version=${hdp.version}
   </value>
 <final>true</final>
</property>

</property>
 <name>mapreduce.admin.reduce.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
 <final>true</final>
</property>

</property>
 <name>mapreduce.application.framework.path</name>
 <value>/hdp/apps/${hdp.version}
   /mapreduce/mapreduce.tar.gz#mr-framework
```

```
 </value>
</property>

</property>
 <name>mapreduce.application.classpath</name>
 <value>$PWD/mr-framework/hadoop/share/hadoop/mapreduce/*:
   $PWD/mr-framework/hadoop/share/hadoop/mapreduce/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/common/*:
   $PWD/mr-framework/hadoop/share/hadoop/common/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/yarn/*:
   $PWD/mr-framework/hadoop/share/hadoop/yarn/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/hdfs/*:
   $PWD/mr-framework/hadoop/share/hadoop/hdfs/lib/*:
  /usr/hdp/${hdp.version}/hadoop/lib/hadoop-lzo-0.6.0.${hdp.version}.jar:
  /etc/hadoop/conf/secure</value>
</property>
```

Note

You do not need to modify ${hdp.version}.

3. Add the following property to yarn-site.xml:

```
<property>
 <name>yarn.application.classpath</name>
 <value>$HADOOP_CONF_DIR,/usr/hdp/${hdp.version}/hadoop-client/*,
  /usr/hdp/${hdp.version}/hadoop-client/lib/*,
  /usr/hdp/${hdp.version}/hadoop-hdfs-client/*,
  /usr/hdp/${hdp.version}/hadoop-hdfs-client/lib/*,
  /usr/hdp/${hdp.version}/hadoop-yarn-client/*,
  /usr/hdp/${hdp.version}/hadoop-yarn-client/lib/*</value>
</property>
```

4. For secure clusters, you must create and configure the container-executor.cfg configuration file:

   - Create the container-executor.cfg file in /etc/hadoop/conf/.
   - Insert the following properties:

     ```
     yarn.nodemanager.linux-container-executor.group=hadoop
     banned.users=hdfs,yarn,mapred
     min.user.id=1000
     ```

     - yarn.nodemanager.linux-container-executor.group - Configured value of yarn.nodemanager.linux-container-executor.group. This must match the value of yarn.nodemanager.linux-container-executor.group in yarn-site.xml.
     - banned.users - Comma-separated list of users who can not run container-executor.
     - min.user.id - Minimum value of user id. This prevents system users from running container-executor.
     - allowed.system.users - Comma-separated list of allowed system users.

   - Set the file /etc/hadoop/conf/container-executor.cfg file permissions to only be readable by root:

```
chown root:hadoop /etc/hadoop/conf/container-executor.cfg
chmod 400 /etc/hadoop/conf/container-executor.cfg
```

- Set the container-executor program so that only root or hadoop group users can run it:

```
chown root:hadoop /usr/hdp/${hdp.version}/hadoop-yarn-server-nodemanager/
bin/container-executor

chmod 6050 /usr/hdp/${hdp.version}/hadoop-yarn-server-nodemanager/bin/
container-executor
```

1.9. Start YARN/MapReduce Services

To start YARN, run commands as a YARN user. To start MapReduce, run commands as a MapReduce user.

1. Start the ResourceManager on all your ResourceManager hosts.

```
su -l yarn -c "/usr/hdp/current/hadoop-yarn-resourcemanager/sbin/yarn-
daemon.sh start resourcemanager"

ps -ef | grep -i resourcemanager
```

2. Start the NodeManager on all your NodeManager hosts.

```
su - l yarn -c "/usr/hdp/current/hadoop-yarn-nodemanager/sbin/yarn-daemon.sh
 start nodemanager"

ps -ef | grep -i nodemanager
```

3. To start MapReduce, run the following commands:

```
su -l yarn -c "/usr/hdp/current/hadoop-mapreduce-historyserver/sbin/mr-
jobhistory-daemon.sh start historyserver"

ps -ef | grep -i jobhistoryserver
```

1.10. Run Hadoop Smoke Tests

To smoke test your Hadoop upgrade, you can run the following MapReduce job as a regular user.

The job uses MapReduce to write 100MB of data into HDFS with RandomWriter

```
hadoop jar/usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.
jar

randomwriter -Dtest.randomwrite.total_bytes=10000000 test-after-upgrade.
```

You should see messages similar to:

```
map 0% reduce 0%
…map 100% reduce 100%
Job …. completed successfully
```

MapReduce upgraded successfully. You can now upgrade your other components.

Basic Troubleshooting

To find the number of active nodes and NodeManagers, access the ResourceManager web UI:

```
http://<resource manager host>:8088/cluster/nodes
```

The number of active nodes should be equal to the number of nodemanagers.

Accessing error messages:

1. Access the ApplicationMaster WebUI to view the container logs.
2. At your console logs for MapReduce job, look for a line with this format:

   13/10/02 17:57:21 INFO mapreduce.Job: The url to track the job: http://<resource manager host>:8088/proxy/application_1380673658357_0007/
3. Select the logs link under ApplicationMaster table. It will redirect you to the container logs. Error messages display here.

1.11. Configure and Start Apache HBase

1. Replace your configuration after upgrading. Replace the HBase template configuration in `/etc/hbase/conf`.
2. Start services. From root, assuming that $HBASE_USER=hbase:

   ```
   su -l hbase -c "/usr/hdp/current/hbase-master/bin/hbase-daemon.sh start master; sleep 25"

   su -l hbase -c "/usr/hdp/current/hbase-regionserver/bin/hbase-daemon.sh start regionserver"
   ```
3. Check processes.

   ```
   ps -ef | grep -i hmaster

   ps -ef | grep -i hregion
   ```

1.12. Configure Apache Phoenix

To enable global indexing and (optionally) local indexing in Phoenix, complete the following steps.

1. Add the following property to the hbase-site.xml file on all HBase nodes, the Master Server, and all Region servers, to prevent deadlocks from occurring during maintenance on global indexes:

   ```
   </property>
    <name>hbase.region.server.rpc.scheduler.factory.class</name>
   ```

```
 <value>org.apache.phoenix.hbase.index.ipc.
    PhoenixIndexRpcSchedulerFactory</value>
 <description>Factory to create the Phoenix RPC Scheduler that knows to put
 index
 updates into index queues</description>
</property>
```

2. **(Optional)** To use local indexing, set the following two properties. These properties ensure collocation of data table and local index regions:

Note

The local indexing feature is a technical preview and considered under development. Do not use this feature in your production systems. If you have questions regarding this feature, contact Support by logging a case on our Hortonworks Support Portal at http://hortonworks.com/support/.

```
</property>
 <name>hbase.master.loadbalancer.class</name>
 <value>org.apache.phoenix.hbase.index.balancer.IndexLoadBalancer</value>
</property>

</property>
 <name>hbase.coprocessor.master.classes</name>
 <value>org.apache.phoenix.hbase.index.master.IndexMasterObserver</value>
</property>
```

3. Restart the HBase Master and Region Servers.

Configuring Phoenix to Run in a Secure Cluster

Perform the following additional steps to configure Phoenix to run in a secure Hadoop cluster:

1. To link the HBase configuration file with the Phoenix libraries:

   ```
   ln -sf HBASE_CONFIG_DIR/hbase-site.xml PHOENIX_HOME/bin/hbase-
   site.xml
   ```

2. To link the Hadoop configuration file with the Phoenix libraries:

   ```
   ln -sf HADOOP_CONFIG_DIR/core-site.xml PHOENIX_HOME/bin/core-
   site.xml
   ```

Note

When running the pssql.py and sqlline.py Phoenix scripts in secure mode, you can safely ignore the following warnings.

```
14/04/19 00:56:24 WARN util.NativeCodeLoader:
Unable to load native-hadoop library for your platform...
  using builtin-java classes where applicable

14/04/19 00:56:24 WARN util.DynamicClassLoader: Failed to identify the fs of
dir hdfs://<HOSTNAME>:8020/apps/hbase/data/lib, ignored java.io.IOException:
No FileSystem for scheme: hdfs
```

1.13. Configure and Start Apache Accumulo

1. You must replace your configuration after upgrading. Copy /etc/accumulo/conf from the template to the conf directory in Apache Accumulo hosts.
2. Start the services:

```
sudo su - l accumulo -c "/usr/hdp/current/accumulo-master/bin/start-server.
sh `hostname` master"

sudo su - l accumulo -c "/usr/hdp/current/accumulo-master/bin/start-server.
sh `hostname` tserver"

sudo su - l accumulo -c "/usr/hdp/current/accumulo-master/bin/start-server.
sh `hostname` gc"

sudo su - l accumulo -c "/usr/hdp/current/accumulo-master/bin/start-server.
sh `hostname` tracer"

sudo su - l accumulo -c "/usr/hdp/current/accumulo-master/bin/start-server.
sh `hostname` monitor"
```

3. Check that the processes are running

   ```
   ps -ef | grep accumulo
   ```

   or visit http://<hostname>:50095 in your browser

1.14. Configure and Start Apache Tez

To upgrade Apache Tez:

1. Copy your previously backed-up copy of tez-site.xml into the /etc/tez/conf directory.
2. Upload the Tez tarball to HDFS.

```
su $HDFS_USER
hdfs dfs -mkdir -p /hdp/apps/<hdp_version>/tez/
hdfs dfs -put /usr/hdp/<hdp_version>/tez/lib/tez.tar.gz /hdp/apps/
<hdp_version>/tez/
hdfs dfs -chown -R hdfs:hadoop /hdp
hdfs dfs -chmod -R 555 /hdp/apps/<hdp_version>/tez
hdfs dfs -chmod -R 444 /hdp/apps/<hdp_version>/tez/tez.tar.gz
```

   Where <hdp_version> is the current HDP version, for example 2.2.6.0-2800.

3. Set the $TEZ_CONF_DIR environment variable to the location of the tez-site.xml file.

```
export TEZ_CONF_DIR=/etc/tez/conf
```

4. Set the $TEZ_JARS environment variable to the location of the Tez .jar files and their dependencies.

```
export TEZ_JARS=/usr/hdp/current/tez-client/*:/usr/hdp/current/tez-client/
lib/*
```

5. Edit the tez.lib.uris property in the tez-site.xml file to point to /hdp/apps/<hdp_version>/tez/tez.tar.gz

```
...
<property>
 <name>tez.lib.uris</name>
 <value>/hdp/apps/<hdp_version>/tez/tez.tar.gz</value>
</property>
...
```

Where <hdp_version> is the current HDP version, for example 2.2.6.0-2800.

6. Remove the following properties from the tez-site.xml configuration file:

```
tez.runtime.intermediate-input.key.comparator.class
tez.runtime.intermediate-input.key.class
tez.runtime.intermediate-input.value.class
tez.runtime.intermediate-input.is-compressed
tez.runtime.intermediate-input.compress.codec
tez.profile.container.list
tez.profile.jvm.opts
tez.am.slowstart-dag-scheduler.min-resource-fraction
```

7. Update the following tez-site.xml properties to their new names.

Old Property Name	New Property Name
tez.am.java.opts	tez.am.launch.cmd-opts
tez.am.env	tez.am.launch.env
tez.am.shuffle-vertex-manager.min-src-fraction	tez.shuffle-vertex-manager.min-src-fraction
tez.am.shuffle-vertex-manager.max-src-fraction	tez.shuffle-vertex-manager.max-src-fraction
tez.am.shuffle-vertex-manager.enable.auto-parallel	tez.shuffle-vertex-manager.enable.auto-parallel
tez.am.shuffle-vertex-manager.desired-task-input-size	tez.shuffle-vertex-manager.desired-task-input-size
tez.am.shuffle-vertex-manager.min-task-parallelism	tez.shuffle-vertex-manager.min-task-parallelism
tez.am.grouping.split-count	tez.grouping.split-count
tez.am.grouping.by-length	tez.grouping.by-length
tez.am.grouping.by-count	tez.grouping.by-count
tez.am.grouping.max-size	tez.grouping.max-size
tez.am.grouping.min-size	tez.grouping.min-size
tez.am.grouping.rack-split-reduction	tez.grouping.rack-split-reduction
tez.am.am.complete.cancel.delegation.tokens	tez.cancel.delegation.tokens.on.completion
tez.am.max.task.attempts	tez.am.task.max.failed.attempts
tez.generate.dag.viz	tez.generate.debug.artifacts
tez.runtime.intermediate-output.key.comparator.class	tez.runtime.key.comparator.class
tez.runtime.intermediate-output.key.class	tez.runtime.key.class
tez.runtime.intermediate-output.value.class	tez.runtime.value.class
tez.runtime.intermediate-output.should-compress	tez.runtime.compress
tez.runtime.intermediate-output.compress.codec	tez.runtime.compress.codec
tez.runtime.intermediate-input.key.secondary.comparator.class	tez.runtime.key.secondary.comparator.class
tez.runtime.broadcast.data-via-events.enabled	tez.runtime.transfer.data-via-events.enabled
tez.runtime.broadcast.data-via-events.max-size	tez.runtime.transfer.data-via-events.max-size
tez.runtime.shuffle.input.buffer.percent	tez.runtime.shuffle.fetch.buffer.percent
tez.runtime.task.input.buffer.percent	tez.runtime.task.input.post-merge.buffer.percent
tez.runtime.job.counters.max	tez.am.counters.max.keys

Old Property Name	New Property Name
tez.runtime.job.counters.group.name.max	tez.am.counters.group-name.max.keys
tez.runtime.job.counters.counter.name.max	tez.am.counters.name.max.keys
tez.runtime.job.counters.groups.max	tez.am.counters.groups.max.keys
tez.task.merge.progress.records	tez.runtime.merge.progress.records
tez.runtime.metrics.session.id	tez.runtime.framework.metrics.session.id
tez.task.scale.memory.additional.reservation.fraction.per-io	tez.task.scale.memory.additional-reservation.fraction.per-io
tez.task.scale.memory.additional.reservation.fraction.max	tez.task.scale.memory.additional-reservation.fraction.max
tez.task.initial.memory.scale.ratios	tez.task.scale.memory.ratios
tez.resource.calculator.process-tree.class	tez.task.resource.calculator.process-tree.class

For more information on setting Tez configuration parameters in HDP-2.2, see "Installing and Configuring Tez" in the HDP Manual Installation Guide.

1.15. Configure and Start Apache Hive and Apache HCatalog

1. Copy the jdbc connector jar from OLD_HIVE_HOME/lib to CURRENT_HIVE_HOME/lib.
2. Upgrade the Apache Hive Metastore database schema. Restart the Hive Metastore database and run:

```
/usr/hdp/current/hive-metastore/bin/schematool -upgradeSchema -dbType <
$databaseType>
```

The value for $databaseType can be derby, mysql, oracle or postgres.

Note

If you are using Postgres 8 and Postgres 9, you should reset the Hive Metastore database owner to <HIVE_USER>:

```
sudo <POSTGRES_USER>
ALTER DATABASE <HIVE-METASTORE-DB-NAME> OWNER TO <HIVE_USER>
```

Note

If you are using Oracle 11, you may see the following error message:

```
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
optimize.mapjoin.mapreduce does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
heapsize does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
server2.enable.impersonation does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
semantic.analyzer.factory.impl does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.auto.
convert.sortmerge.join.noconditionaltask does not exist
Metastore connection URL: jdbc:oracle:thin:@//ip-172-31-42-1.ec2.
internal:1521/XE
```

```
Metastore Connection Driver : oracle.jdbc.driver.OracleDriver
Metastore connection User: hiveuser
Starting upgrade metastore schema from version 0.13.0 to 0.14.0
Upgrade script upgrade-0.13.0-to-0.14.0.oracle.sql
Error: ORA-00955: name is already used by an existing object
 (state=42000,code=955)
Warning in pre-upgrade script pre-0-upgrade-0.13.0-to-0.14.0.
oracle.sql: Schema script failed, errorcode 2
Completed upgrade-0.13.0-to-0.14.0.oracle.sql
schemaTool completed
```

You can safely ignore this message. The error is in the pre-upgrade script and can be ignored; the schematool succeeded.

Note

Copy only the necessary configuration files. Do not copy the env.sh files, for example, hadoop-env.sh, hive-env.sh, and so forth. Additionally, all env.sh files must be properly configured.

3. Download and extract the HDP companion files (see "Download the Companion Files" in Chapter 1 of the Manual Install Guide).

   Copy the hive-site.xml file in the configuration_files/hive directory of the extracted companion files to the etc/hive/conf directory on your Hive host machine. This new version of the hive-site.xml file contains new properties for HDP-2.2 features.

4. Edit the hive-site.xml file and modify the properties based on your environment. Search for TODO in the file for the properties to replace.

   a. Edit the connection properties for your Hive metastore database in hive-site.xml:

```
<property>
 <name>javax.jdo.option.ConnectionURL</name>
 <value>jdbc:mysql://TODO-HIVE-METASTORE-DB-SERVER:
   TODO-HIVE-METASTORE-DB-PORT/TODO-HIVE-METASTORE-DB-NAME
   ?createDatabaseIfNotExist=true</value>
 <description>Enter your Hive Metastore Connection URL,
   for example if MySQL: jdbc:mysql://localhost:3306/mysql
   ?createDatabaseIfNotExist=true</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionUserName</name>
 <value>TODO-HIVE-METASTORE-DB-USER-NAME</value>
 <description>Enter your Hive Metastore database user name.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionPassword</name>
 <value>TODO-HIVE-METASTORE-DB-PASSWORD</value>
 <description>Enter your Hive Metastore database password.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionDriverName</name>
 <value>TODO-HIVE-METASTORE-DB-CONNECTION-DRIVER-NAME</value>
 <description>Enter your Hive Metastore Connection Driver Name,
```

```
   for example if MySQL: com.mysql.jdbc.Driver</description>
</property>
```

b. Edit the following properties in the hive-site.xml file:

```
<property>
 <name>fs.file.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.file.impl.disable.cache</
description>
</property>

<property>
 <name>fs.hdfs.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.hdfs.impl.disable.cache
 </description>
<property>
```

c. **Optional:** If you want Hive Authorization, set the following Hive authorization parameters in the hive-site.xml file:

```
<property>
 <name>hive.security.authorization.enabled</name>
 <value>true</value>
</property>

<property>
 <name>hive.security.authorization.manager</name>
 <value>org.apache.hadoop.hive.ql.security.authorization.
   StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.metastore.authorization.manager</name>
 <value>org.apache.hadoop.hive.ql.security.authorization.
   StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.authenticator.manager</name>
 <value>org.apache.hadoop.hive.ql.security.ProxyUserAuthenticator</value>
</property>
```

d. For a remote Hive metastore database, set the IP address (or fully-qualified domain name) and port of the metastore host using the following hive-site.xml property value.

   (To enable HiveServer2, please this property value empty.)

```
<property>
 <name>hive.metastore.uris</name>
 <value>thrift://$metastore.server.full.hostname:9083</value>
 <description>URI for client to contact metastore server.
   To enable HiveServer2, leave the property value empty.
   </description>
</property>
```

You can further fine-tune your configuration settings based on node hardware specifications, using the HDP utility script.

5. Start Hive.

   On the Hive Metastore host machine, run the following command:

   ```
   su - hive

   nohup /usr/hdp/current/hive-metastore/bin/hive --service
   metastore>/var/log/hive/hive.out 2>/var/log/hive/hive.log &
   ```

6. Start Hive Server2.

   On the Hive Server2 host machine, run the following command:

   ```
   su - hive

   nohup /usr/hdp/current/hive-server2/bin/hiveserver2 -hiveconf
   hive.metastore.uris=" " -hiveconf hive.log.file=hiveserver2.log
   >/var/log/hive/hiveserver2.out 2> /var/log/hive/
   hiveserver2err.log &
   ```

   where $HIVE_USER is the Hive Service user. For example, hive.

1.16. Configure and Start Apache Oozie

Upgrading Apache Oozie is a complex process. Although the instructions are straightforward, set aside a dedicated block of time to upgrade oozie clients and servers.

Perform the following preparation steps on each oozie server host:

1. You must restore oozie-site.xml from your backup to the conf directory on each oozie server and client.

2. Copy the JDBC jar to libext-customer:

   a. Create the /usr/hdp/current/oozie-client/libext-customer directory.

      ```
      cd /usr/hdp/current/oozie-client mkdir libext-customer
      ```

   b. Grant read/write/execute access to all users for the libext-customer directory.

      ```
      chmod -R 777 /usr/hdp/current/oozie-client/libext-customer
      ```

3. Copy these files to the libext-customer directory

   ```
   cp /usr/hdp/current/hadoop/lib/hadooplzo*.jar /usr/hdp/current/
   oozie/libext-customer

   cp /usr/share/HDP-oozie/ext-2.2.zip /usr/hdp/current/oozie/
   libext-customer/
   ```

Note

If you do not have an LZO JAR file, you must enable LZO compression first. See Install Compression Libraries.

4. Extract share-lib.

   ```
   /usr/hdp/current/oozie/bin/oozie-setup.sh sharelib create -fs
   hdfs://<namenode>:8020
   ```

   To verify that the sharelibs extracted correctly, run the following command:

   ```
   oozie admin -oozie http://<oozie server host address>:11000/
   oozie -shareliblist
   ```

   There should be:

   - Available ShareLib
   - oozie
   - hive
   - distcp
   - hcatalog
   - sqoop
   - mapreduce-streaming
   - pig

   Change the ownershop and permissions of the oozie directory:

   ```
   su -l hdfs -c "hdfs dfs -chown oozie:hadoop /user/oozie"
   ```

   ```
   su -l hdfs -c "hdfs dfs -chmod -R 755 /user/oozie"
   ```

5. If a previous version of Oozie was created using auto schema creation, you must run an SQL query.

   Use the `oozie-site.xml` properties:

   - `oozie.service.JPAService.jdbc.username`
   - `oozie.service.JPAService.jdbc.username`
   - `oozie.service.JPAService.jdbc.url`

   to obtain the password, username and db to run the query.

   Run the SQL query:

   ```
   insert into oozie_sys (name, data) values ('db.version', '2.5');
   ```

6. As the Oozie user (not root), run the upgrade.

   ```
   su $OOZIE_USER

   /usr/hdp/current/oozie-server/bin/ooziedb.sh upgrade -run
   ```

7. As root, prepare the Oozie WAR file.

   ```
   sudo su -l oozie -c "/usr/hdp/current/oozie/bin/oozie-setup.sh
   prepare-war -d /usr/hdp/current/oozie/libext-customer"
   ```

   Look for console output to indicate success. For example, if you are using MySQL you should see something similar to:

   ```
   INFO: Adding extension: libext-customer/mysql-connector-java.jar
   New Oozie WAR file with added 'JARs' at /var/lib/oozie/oozie-server/webapps/
   oozie.war
   ```

8. Replace the content of /user/oozie-server/share in HDFS. On the Oozie server host:

   ```
   sudo su -l oozie -c "/usr/hdp/current/oozie-server/bin/oozie-
   setup.sh prepare-war -d /usr/hdp/current/oozie/libext-customer"
   ```

9. Add the following property to oozie-log4j.properties:

   ```
   log4j.appender.oozie.layout.ConversionPattern=%d{ISO8601} %5p
   %c{1}:%L - SERVER[${oozie.instance.id}] %m%n
   ```

   where ${oozie.instance.id} is determined by oozie, automatically.

10. Start Oozie as the Oozie user:

    ```
    sudo su -l <OOZIE_User> -c "export OOZIE_CATALINA_HOME='usr/lib/bigtop-
    tomcat'; $OOZIE_CATALINA_HOME/catalina.sh $OOZIE_CATALINA_HOME/setclasspath.
    sh
    /usr/hdp/current/oozie-server/bin/oozied.sh start”
    ```

11. Check processes.

    ```
    ps -ef | grep -i oozie
    ```

1.17. Configure and Start Apache WebHCat

1. You must replace your configuration after upgrading. Copy `/etc/webhcat/conf` from the template to the conf directory in webhcat hosts.

2. Modify the Apache WebHCat configuration files.

   a. Upload Pig, Hive and Sqoop tarballs to HDFS as the $HDFS_User (in this example, hdfs):

      ```
      hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/pig/

      hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/hive/

      hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/sqoop/
      ```

```
hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/pig/pig.tar.gz /hdp/apps/2.2.6.
0-<$version>/pig/

hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hive/hive.tar.gz /hdp/apps/2.2.
6.0-<$version>/hive/

hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/sqoop/sqoop.tar.gz /hdp/apps/2.
2.6.0-<$version>/sqoop/

hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/pig

hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/pig/pig.tar.gz

hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/hive

hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/hive/hive.tar.gz

hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/sqoop

hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/sqoop/sqoop.tar.gz

hdfs dfs -chown -R hdfs:hadoop /hdp
```

b. Update the following properties in the webhcat-site.xml configuration file, as their values have changed:

```
<property>
 <name>templeton.pig.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/pig/pig.tar.gz</value>
</property>

<property>
 <name>templeton.hive.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/hive/hive.tar.gz</value>
</property>

<property>
 <name>templeton.streaming.jar</name>
 <value>hdfs:///hdp/apps/${hdp.version}/mapreduce/
   hadoop-streaming.jar</value>
 <description>The hdfs path to the Hadoop streaming jar file.</
description>
</property>

<property>
 <name>templeton.sqoop.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/sqoop/sqoop.tar.gz</value>
 <description>The path to the Sqoop archive.</description>
</property>

<property>
 <name>templeton.sqoop.path</name>
 <value>sqoop.tar.gz/sqoop/bin/sqoop</value>
 <description>The path to the Sqoop executable.</description>
</property>

<property>
 <name>templeton.sqoop.home</name>
 <value>sqoop.tar.gz/sqoop</value>
 <description>The path to the Sqoop home in the exploded archive.
```

```
   </description>
</property>
```

Note

You do not need to modify ${hdp.version}.

c. Remove the following obsolete properties from webhcat-site.xml:

```
<property>
 <name>templeton.controller.map.mem</name>
 <value>1600</value>
 <description>Total virtual memory available to map tasks.</description>
</property>

<property>
 <name>hive.metastore.warehouse.dir</name>
 <value>/path/to/warehouse/dir</value>
</property>
```

d. Add new proxy users, if needed. In core-site.xml, make sure the following properties are also set to allow WebHcat to impersonate your additional HDP 2.2 groups and hosts:

```
<property>
 <name>hadoop.proxyuser.hcat.groups</name>
 <value>*</value>
</property>

<property>
 <name>hadoop.proxyuser.hcat.hosts</name>
 <value>*</value>
</property>
```

Where:

`hadoop.proxyuser.hcat.group`

Is a comma-separated list of the Unix groups whose users may be impersonated by 'hcat'.

`hadoop.proxyuser.hcat.hosts`

A comma-separated list of the hosts which are allowed to submit requests by 'hcat'.

3. Start WebHCat:

```
sudo su -l $WEBHCAT_USER -c "/usr/hdp/current/hive-hcatalog/
sbin/webhcat_server.sh start"
```

4. Smoke test WebHCat.

a. At the WebHCat host machine, run the following command:

```
http://$WEBHCAT_HOST_MACHINE:50111/templeton/v1/status
```

b. If you are using a secure cluster, run the following command:

```
curl --negotiate -u: http://cluster.$PRINCIPAL.$REALM:50111/
templeton/v1/status {"status":"ok","version":"v1"}
[machine@acme]$
```

1.18. Configure Apache Pig

1. Replace your configuration after upgrading. Copy `/etc/pig/conf` from the template to the conf directory in pig hosts.

2. To validate the Apache Pig upgrade, complete the following steps:

   a. On the host machine where Pig is installed, run the following commands:

   ```
   su - $HDFS_USER/usr/hdp/current/hadoop/bin/hadoop
   fs -copyFromLocal /etc/passwd passwd
   ```

   b. Create a Pig script file named `/tmp/id.pig` that contains the following Pig Latin commands:

   ```
   A = load 'passwd' using PigStorage(':');B = foreach A generate $0 as id;
    store B into '/tmp/id.out';
   ```

   c. Run the Pig script:

   ```
   su - $HDFS_USER
   pig -l /tmp/pig.log /tmp/id.pig
   ```

1.19. Configure and Start Apache Sqoop

1. Replace your configuration after upgrading. Copy `/etc/sqoop/conf` from the template to the conf directory in sqoop hosts.

2. As the HDFS Service user, upload the Apache Sqoop tarball to HDFS.

   ```
   su -c "hdfs dfs -mkdir -p /hdp/apps/2.2.0.0-<$version>/sqoop"

   hdfs su -c "hdfs dfs -chmod -R 555 /hdp/apps/2.2.0.0-<$version>/sqoop"

   hdfs su -c "hdfs dfs -chown -R hdfs:hadoop /hdp/apps/2.2.0.0-<$version>/
   sqoop"

   hdfs su -c "hdfs dfs -put /usr/hdp/2.2.0.0-<$version>/sqoop/sqoop.tar.gz /
   hdp/apps/2.2.0.0-<$version>/sqoop/sqoop.tar.gz"

   hdfs su -c "hdfs dfs -chmod 444 /hdp/apps/2.2.0.0-<$version>/sqoop/sqoop.
   tar.gz"
   ```

3. If you are using the MySQL database as a source or target, then the MySQL connector jar must be updated to 5.1.29 or later.

   Refer to the MySQL web site for information on updating the MySQL connector jar.

4. Because Sqoop is a client tool with no server component, you will need to run your own jobs to validate the upgrade.

1.20. Configure, Start, and Validate Apache Flume

1. If you have not already done so, upgrade Apache Flume. On the Flume host machine, run the following command:

   - For **RHEL/CentOS/Oracle Linux**:

     ```
     yum upgrade flume
     ```

   - For **SLES**:

     ```
     zypper update flume
     ```

     ```
     zypper remove flume
     ```

     ```
     zypper se -s flume
     ```

     You should see Flume in the output.

     Install Flume:

     ```
     zypper install flume
     ```

   - For **Ubuntu/Debian**:

     ```
     apt-get install flume
     ```

2. To confirm that Flume is working correctly, create an example configuration file. The following snippet is a sample configuration that can be set using the properties file. For more detailed information, see the "Flume User Guide."

   ```
   agent.sources = pstream
   agent.channels = memoryChannel
   agent.channels.memoryChannel.type = memory

   agent.sources.pstream.channels = memoryChannel
   agent.sources.pstream.type = exec
   agent.sources.pstream.command = tail -f /etc/passwd

   agent.sinks = hdfsSink
   agent.sinks.hdfsSink.type = hdfs
   agent.sinks.hdfsSink.channel = memoryChannel
   agent.sinks.hdfsSink.hdfs.path = hdfs://tmp/flumetest
   agent.sinks.hdfsSink.hdfs.fileType = SequenceFile
   agent.sinks.hdfsSink.hdfs.writeFormat = Text
   ```

   The source here is defined as an exec source. The agent runs a given command on startup, which streams data to stdout, where the source gets it. The channel is defined as an in-memory channel and the sink is an HDFS sink.

3. Given this configuration, you can start Flume as follows:

   ```
   $ bin/flume-ng agent --conf ./conf --conf-file example.conf --
   name a1 -Dflume.root.logger=INFO,console
   ```

Note

The directory specified for `--conf agrument` would include a shell script flume-env.sh and potentially a log4j properties file. In this example, we pass a Java option to force Flume to log to the console and we go without a custom environment script.

4. After validating data in `hdfs://tmp/flumetest`, stop Flume and restore any backup files. Copy `/etc/flume/conf` to the conf directory in Flume hosts.

1.21. Configure, Start, and Validate Apache Mahout

Replace your configuration after upgrading. Copy `/etc/mahout/conf` from the template to the conf directory in mahout hosts.

To validate Apache Mahout:

1. Create a test user named "testuser" in the Linux cluster and in HDFS, and log in as that user.

   ```
   hdfs dfs -put /tmp/sample-test.txt /user/testuser
   ```

2. Export the required environment variables for Mahout:

   ```
   export JAVA_HOME="your_jdk_home_install_location_here
   export HADOOP_HOME=/usr/hdp/current/hadoop-client
   export MAHOUT_HOME=/usr/hdp.current/mahout-client
   export PATH="$PATH":$HADOOP_HOME/bin:$MAHOUT_HOME/bin
   export CLASSPATH="$CLASSPATH":$MAHOUT_HOME
   ```

3. Upload a few megabytes of natural-language plain text to the Linux server as `/tmp/sample-test.txt`.

4. Transfer the sample-test.txt file to a subdirectory of the testusers's HDFS home directory.

   ```
   hdfs dfs -mkdir /user/testuser/testdata
   hdfs dfs -put /tmp/sample-test.txt /user/testuser/testdata
   ```

5. Enter the mahout command to convert the plain text file sample-test.txt into a sequence file stored in the output directory mahouttest:

   ```
   mahout seqdirectory --input /user/testuser/testdata --output /user/testuser/
   mahouttest -ow --charset utf-8
   ```

1.22. Configure and Start Hue

For HDP 2.2, use the Hue version shipped with HDP 2.2. If you have a previous version of Hue, use the following steps to upgrade Hue.

1. Migrate hue.ini setting from your old hue.ini configuration file to new hue.ini configuration file.

2. If you are using the embedded SQLite database, remember to restore your database after upgrade.

   To restore the database from a backup, make sure the destination database is empty before copying (if necessary, rename or remove the current destination database), then copy your backup to the destination database. For example:

   ```
   su $HUE_USER
   cd /var/lib/hue
   mv desktop.db desktop.db.old
   sqlite3 desktop.db < ~/hue_backup/desktop.bak
   exit
   ```

3. Synchronize Database

   ```
   cd /usr/lib/hue
   source ./build/env/bin/activate
   hue syncdb
   deactivate
   ```

4. Start Hue. As a root user, run the following command on the Hue Server:

   ```
   /etc/init.d/hue start
   ```

1.23. Configure and Start Apache Knox

For HDP 2.2, the default paths for Apache Knox change. Upgrade Knox in order to update these paths.

1. Review your topologies. If you prefer to use the new topologies, copy the topology templates to the active topologies directory:

   ```
   cp /usr/hdp/current/knox/templates/topology.xml /usr/hdp/
   current/knox/conf/topologies/{clustername}.xml
   ```

   Modify the topology to match the cluster resources in your deployment.

2. Create the Master Secret:

   ```
   su - knox -c “/usr/hdp/current/knox/bin/knoxcli.sh create-
   master”
   ```

3. Start the Gateway:

   ```
   su -l knox -c "/usr/hdp/current/knox-server/bin/gateway.sh
   start"
   ```

4. Check /var/log/knox/gateway.log for errors or successful start.

1.24. Configure and Validate Apache Falcon

Upgrade Apache Falcon after you have upgraded HDFS, Hive, Oozie, and Pig. Stop Oozie jobs while running Falcon.

1. Replace your configuration after upgrading. Copy /etc/falcon/conf from the template to the conf directory in falcon hosts.

2. Check your Falcon entities. There should be no changes, but in some cases you may need to update your entities post-upgrade.

1.25. Configure and Start Apache Storm

Storm is fairly independent of changes to the HDP cluster, but you must upgrade Storm for rolling upgrade support in HDP 2.2 and to be on the latest version of Apache Storm.

1. Deactivate all running topologies.

2. Stop Storm Services on the storm node.

3. Stop ZooKeeper Services on the storm node.

4. Remove Storm and zookeeper from the storm node and install the HDP 2.2 version:

   - For **RHEL/CentOS/Oracle Linux**:

     ```
     yum erase storm

     yum erase zookeeper

     yum install storm

     yum install zookeeper
     ```

   - For **SLES**:

     ```
     zypper rm storm

     zypper rm zookeeper

     zypper install storm

     zypper install zookeeper
     ```

   - For **Ubuntu/Debian**:

     ```
     apt-get remove storm --purge

     apt-get remove zookeeper --purge

     apt-get install storm

     apt-get install zookeeper
     ```

5. Replace your configuration after upgrading. Copy /etc/storm/conf from the template to the conf directory .

6. Replace your ZooKeeper configuration after upgrading. Replace the ZooKeeper template configuration in `/etc/zookeeper/conf`.

7. Start ZooKeeper. On the storm node, run the following command:

```
sudo su -l $ZOOKEEPER_USER -c "source /etc/zookeeper/conf/zookeeper-env.sh;
 export ZOOCFGDIR=/etc/zookeeper/conf; /usr/hdp/current/zookeeper-server/
bin/zkServer.sh start >> $ZOO_LOG_DIR/zoo.out\"
```

   where

   - $ZOOKEEPER_USER is the ZooKeeper Service user. For example, zookeeper.
   - $ZOO_LOG_DIR is the directory where ZooKeeper server logs are stored. For example, /var/log/zookeeper.

8. Start Storm using a process controller, such as supervisord:

   ```
   su storm /usr/hdp/current/storm-supervisor/bin/storm supervisor
   ```

   where $STORM_USER is the operating system user that installed Storm. For example, storm.

1.26. Upgrade Apache Ranger

Ranger is the new name for XASecure, an HDP 2.1 add-on product.

1. Backup the following config directories:

```
xapolicymgr config (from /usr/lib/xapolicymgr)
uxugsync config (from /usr/lib/uxugsync)
agents config (/etc/hadoop/conf, /etc/hive/conf, /etc/hbase/conf)
```

   Backup Apache Ranger policy/audit databases. For example:

   ```
   mysqldump xasecure > /tmp/xasecure/backup_xasecure.sql
   ```

   

   Note

   The DB host and xaadmin and xalogger user details are important to note because you must specify the same database details when installing the HDP 2.2 version of Ranger.

2. Stop services.

   - Stop Policy Admin

     ```
     service xapolicymgr stop
     ```

   - Stop user sync.

     ```
     service uxugsync stop
     ```

   - Stop other HDP Services.

3. Cleanup startup files/symbolic links

```
rm -f /etc/init.d/xapolicymgr
```

```
rm -rf /etc/rc*.d/*xapolicymgr
rm -f /etc/init.d/uxugsync
rm -rf /etc/rc*.d/*uxugsync
rm -f /usr/lib/xapolicymgr
rm -f /usr/lib/uxugsync
```

4. Follow the Ranger Install instructions for HDP 2.2 to install the following Ranger components: Ranger-Admin Ranger-UserSync (Plugins for HDFS, HBase, Hive, Knox and Storm as required.)

5. Restart Ranger Services

   ```
   service ranger-admin start

   service ranger-usersync start
   ```

1.27. Finalize the Upgrade

You can start HDFS without finalizing the upgrade. When you are ready to discard your backup, you can finalize the upgrade.

1. Verify your filesystem health before finalizing the upgrade. (After you finalize an upgrade, your backup will be discarded!)

2. As the $HDFS_USER, enter:

   ```
   hdfs dfsadmin -finalizeUpgrade
   ```

1.28. Install New HDP 2.2 Services

Install new HDP 2.2 Services (see the Manual Install Guide):

- Slider – a YARN application to deploy existing distributed applications on YARN, monitor them, and make them larger or smaller as desired even while the application is running.
- Ranger – a comprehensive approach to central security policy administration across the core enterprise security requirements of authentication, authorization, accounting and data protection
- Kafka – an open-source message broker project written in Scala.
- Spark – an open-source cluster computing engine.

2. Upgrade from HDP 2.0 to HDP 2.2 Manually

HDP 2.2 supports side-by-side installation of HDP 2.2 and above releases, which lets you perform rolling upgrades on your cluster and improve execution times for in-place upgrade. To support side-by-side installation, the HDP package version naming convention for both RPMs and Debs has changed to include the HDP 2.2 product version. For example, hadoop-hdfs is now hadoop-2.2.6.0-hdfs. HDP 2.2 marks the first release where HDP rpms, debs, and directories contain versions in the names to permit side-by-side installations of later HDP releases.

To select from the releases you have installed side-by-side, Hortonworks provides hdp-select, a command that lets you select the active verison of HDP from the versions you have selected.

HDP packages for a complete installation of HDP 2.2 will take about 2.5 GB of disk space.

Warning

Until the upgrade is finalized, no HDFS data is deleted from the cluster. Be sure to review your capacity and ensure that you have extra space available during the upgrade window.

This document provides instructions on how to upgrade to HDP 2.2 from the HDP 2.0 release.

Note

These instructions cover the upgrade between two minor releases. If you need to upgrade between two maintenance releases, follow the upgrade instructions in the HDP Release Notes.

1. Getting Ready to Upgrade
2. Upgrade HDP 2.0 Components
3. Symlink Directories with hdp-select
4. Configure and Start Apache ZooKeeper
5. Configure Hadoop
6. Start Hadoop Core
7. Verify HDFS Filesystem Health
8. Configure YARN and MapReduce
9. Start YARN/MapReduce Services
10. Run Hadoop Smoke Tests
11. Configure and Start Apache HBase

12. Configure and Start Apache Hive and Apache HCatalog
13. Configure and Start Apache Oozie
14. Configure and Start Apache WebHCat (Templeton)
15. Configure and Start Apache Pig
16. Configure and Start Apache Sqoop
17. Configure, Start, and Validate Apache Flume
18. Configure, Start, and Validate Apache Mahout
19. Configure and Start Hue
20. Finalize the Upgrade
21. Install New HDP 2.2 Services

2.1. Getting Ready to Upgrade

HDP Stack upgrade involves upgrading from HDP 2.0 to HDP 2.2 versions and adding the new HDP 2.2 services.

The first step is to ensure you keep a backup copy of your HDP 2.0 configurations.

1. Hardware recommendations

   Although there is no single hardware requirement for installing HDP, there are some basic guidelines. The HDP packages for a complete installation of HDP 2.2 will take up about 2.5 GB of disk space.

2. Back up the following HDP directories:

   - /etc/hadoop/conf
   - /etc/hbase/conf
   - /etc/hive-hcatalog/conf
   - /etc/hive/conf
   - /etc/pig/conf
   - /etc/sqoop/conf
   - /etc/flume/conf
   - /etc/mahout/conf
   - /etc/oozie/conf
   - /etc/hue/conf
   - /etc/zookeeper/conf

- **Optional:** Back up your userlogs directories, ${mapred.local.dir}/userlogs.

3. Run the fsck command as the HDFS Service user and fix any errors. (The resulting file contains a complete block map of the file system.)

   ```
   su -l <HDFS_USER>

   hdfs fsck / -files -blocks -locations > dfs-old-fsck-1.log
   ```

   where $HDFS_USER is the HDFS Service user. For example, hdfs.

4. Use the following instructions to compare status before and after the upgrade:

   The following commands must be executed by the user running the HDFS service (by default, the user is hdfs).

   a. Capture the complete namespace of the file system. (The second command does a recursive listing of the root file system.)

      ```
      su -l <$HDFS_USER>

      hdfs dfs -ls -R / > dfs-old-lsr-1.log
      ```

      

      Note

      In secure mode you must have kerberos credentials for the hdfs user.

   b. Run the report command to create a list of DataNodes in the cluster.

      ```
      su -l <HDFS_USER>

      hdfs dfsadmin -report > dfs-old-report-1.log
      ```

   c. **Optional:** You can copy all or unrecoverable only data storelibext-customer directory in HDFS to a local file system or to a backup instance of HDFS.

   d. **Optional:** You can also repeat the steps 3 (a) through 3 (c) and compare the results with the previous run to ensure the state of the file system remained unchanged.

5. As the HDFS user, save the namespace by executing the following command:

   ```
   su -l <HDFS_USER>

   hdfs dfsadmin -safemode enter

   hdfs dfsadmin -saveNamespace
   ```

6. Backup your NameNode metadata.

   - Copy the following checkpoint files into a backup directory:
     - `dfs.namenode.name.dir/current/edits_*`
     - `dfs.namenode.name.dir/current/fsimage`

- `dfs.namenode.name.dir/current/fsimage_*`

- Store the layoutVersion of the namenode.

  `${dfs.namenode.name.dir}/current/VERSION`

7. Finalize any **PRIOR** HDFS upgrade, if you have not done so already.

   ```
   su -l <HDFS_USER>
   ```

   ```
   hdfs dfsadmin -finalizeUpgrade
   ```

8. **Optional:** Back up the Hive Metastore database.

   The following instructions are provided for your convenience. For the latest backup instructions, please see your database documentation.

   Table 2.1. Hive Metastore Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sqlsbr For example: mysqldump hive > /tmp/mydir/backup_hive.sql	mysql $dbname < $inputfilename.sqlsbr For example: mysql hive < /tmp/mydir/backup_hive.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql sbr For example: sudo -u postgres pg_dump hive > /tmp/mydir/backup_hive.sql	sudo -u $username psql $databasename < $inputfilename.sqlsbr For example: sudo -u postgres psql hive < /tmp/mydir/backup_hive.sql
Oracle	Connect to the Oracle database using sqlplus. Export the database. exp username/password@database full=yes file=output_file.dmp mysql $dbname < $inputfilename.sqlsbr For example: mysql hive < /tmp/mydir/backup_hive.sql	Import the database: imp username/password@database file=input_file.dmp

9. Optional: Back up the Oozie metastore database.

   These instructions are provided for your convenience. Please check your database documentation for the latest back up instructions.

   Table 2.2. Oozie Metastore Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sql For example:	mysql $dbname < $inputfilename.sql For example:

Database Type	Backup	Restore
	mysqldump oozie > /tmp/mydir/backup_hive.sql	mysql oozie < /tmp/mydir/backup_oozie.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql For example: sudo -u postgres pg_dump oozie > /tmp/mydir/backup_oozie.sql	sudo -u $username psql $databasename < $inputfilename.sql For example: sudo -u postgres psql oozie < /tmp/mydir/backup_oozie.sql

10. **Optional**: Back up the Hue database.

    The following instructions are provided for your convenience. For the latest backup instructions, please see your database documentation. For database types that are not listed below, follow your vendor-specific instructions.

Table 2.3. Hue Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sqlsbr For example: mysqldump hue > /tmp/mydir/backup_hue.sql	mysql $dbname < $inputfilename.sqlsbr For example: mysql hue < /tmp/mydir/backup_hue.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql sbr For example: sudo -u postgres pg_dump hue > /tmp/mydir/backup_hue.sql	sudo -u $username psql $databasename < $inputfilename.sqlsbr For example: sudo -u postgres psql hue < /tmp/mydir/backup_hue.sql
Oracle	Connect to the Oracle database using sqlplus export the database: For example: exp username/password@database full=yes file=output_file.dmp mysql $dbname < $inputfilename.sqlsbr	Import the database: For example: imp username/password@database file=input_file.dmp
SQLite	/etc/init.d/hue stop su $HUE_USER mkdir ~/hue_backup sqlite3 desktop.db .dump > ~/hue_backup/desktop.bak /etc/init.d/hue start	/etc/init.d/hue stop cd /var/lib/hue mv desktop.db desktop.db.old sqlite3 desktop.db < ~/hue_backup/desktop.bak /etc/init.d/hue start

11. Stop all services (including MapReduce) and client applications deployed on HDFS using the instructions provided in the Stopping HDP services.

12. Verify that edit logs in ${dfs.namenode.name.dir}/current/edits* are empty.

    a. Run: `hdfs oev -i ${dfs.namenode.name.dir}/current/edits_inprogress_* -o edits.out`

b. Verify edits.out file. It should only have OP_START_LOG_SEGMENT transaction. For example:

```
<?xml version="1.0" encoding="UTF-8"?>
<EDITS>
<EDITS_VERSION>-56</EDITS_VERSION>
<RECORD>
<OPCODE>OP_START_LOG_SEGMENT</OPCODE>
<DATA>
<TXID>5749</TXID>
</DATA>
</RECORD>
```

c. If edits.out has transactions other than OP_START_LOG_SEGMENT run the following steps and then verify edit logs are empty.

- Start the existing version NameNode.
- Ensure there is a new FS image file.
- Shut the NameNode down.

  `hdfs dfsadmin – saveNamespace`

13. Rename or delete any paths that are reserved in the new version of HDFS.

When upgrading to a new version of HDFS, it is necessary to rename or delete any paths that are reserved in the new version of HDFS. If the NameNode encounters a reserved path during upgrade, it will print an error such as the following:

```
/.reserved is a reserved path and .snapshot is a reserved path component in
 this version of HDFS.
Please rollback and delete or rename this path, or upgrade with the
-renameReserved key-value pairs option to automatically rename these paths
 during upgrade.
```

Specifying `-upgrade -renameReserved` optional key-value pairs causes the NameNode to automatically rename any reserved paths found during startup.

For example, to rename all paths named .snapshot to .my-snapshot and change paths named .reserved to .my-reserved, specify:

`-upgrade -renameReserved .snapshot=.my-snapshot,.reserved=.my-reserved`.

If no key-value pairs are specified with -renameReserved, the NameNode will suffix reserved paths with .<LAYOUT-VERSION>.UPGRADE_RENAMED. For example:

`.snapshot.-51.UPGRADE_RENAMED`.

Note

We recommend that you perform a `-saveNamespace` before renaming paths (running `-saveNamespace` appears in a previous step in this

procedure). This is because a data inconsistency can result if an edit log operation refers to the destination of an automatically renamed file.

Also note that running `-renameReserved` will rename all applicable existing files in the cluster. This may impact cluster applications.

14. Get public keys for GPG.

2.2. Upgrade HDP 2.0 Components

The upgrade process to HDP 2.2 involves the following steps. Select your OS:

RHEL/CentOS/Oracle 6

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 2.0 components. This command uninstalls the HDP 2.0 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
   "flume*" "mahout*" "hue*"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/yum.repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/2.x/
     updates/2.2.6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
     ```

   - Confirm that the HDP repository is configured.

     ```
     yum repolist
     ```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

     ```
     Loaded plugins: fastestmirror, security
     Loading mirror speeds from cached hostfile
     * base: mirrors.cat.pdx.edu
     * extras: linux.mirrors.es.net
     * updates: mirrors.usc.edu
     repo id repo namestatus
     HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

   ```
   yum install "hadoop" "hadooplzo" "webhcat" "oozie" "collectd"
   "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hue" "hive" "flume"
   "accumulo" "mahout" "hdp_mon_nagios_addons"
   ```

```
yum install install hive-webhcat webhcat-tar-hive webhcat-tar-
pig
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded.

   ```
   yum list installed | grep HDP-<old.stack.version.number>
   ```

   No component file names should appear in the returned list.

RHEL/CentOS/Oracle 5 (Deprecated)

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 2.0 components. This command uninstalls the HDP 2.0 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
   "flume*" "mahout*" "hue*"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/yum.repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/2.x/
     updates/2.2.6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
     ```

   - Confirm the HDP repository is configured.

     ```
     yum repolist
     ```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

     ```
     Loaded plugins: fastestmirror, security
     Loading mirror speeds from cached hostfile
     * base: mirrors.cat.pdx.edu
     * extras: linux.mirrors.es.net
     * updates: mirrors.usc.edu
     repo id repo namestatus
     HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

```
yum install "hadoop" "hadooplzo" "webhcat" "hcatalog" "oozie"
"collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive"
"flume" "accumulo" "mahout" "hue" "hdp_mon_nagios_addons"

yum install install hive-webhcat webhcat-tar-hive webhcat-tar-
pig
```

6. Verify that the components were upgraded.

   ```
   yum list installed | grep HDP-<old.stack.version.number>
   ```

   No component file names should appear in the returned list.

SLES 11 SP 1

1. On all hosts, clean the yum repository.

   ```
   zypper clean all
   ```

2. Remove your old HDP 2.0 components.

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
   "flume*" "mahout*" "hue*"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/zypp/repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/
     sles11sp1/2.x/updates/2.2.6.0/hdp.repo -O /etc/zypp/repos.d/
     hdp.repo
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

   ```
   zypper install "hadoop" "hadooplzo" "webhcat" "oozie" "collectd"
   "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "tez" "storm"
   "falcon" "flume" "phoenix" "accumulo" "mahout" "hue" "knox"
   "hdp_mon_nagios_addons"

   zypper install hive-webhcat webhcat-tar-hive webhcat-tar-pig

   zypper up -r HDP-2.2.6.0

   zypper install oozie-client
   ```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded. For example, to verify hdfs, hive, and hcatlog:

```
rpm -qa | grep hdfs

rpm -qa | grep hive

rpm -qa | grep hcatalog
```

No component files names should appear in the returned list.

SLES 11 SP3/SP4

1. On all hosts, clean the yum repository.

```
zypper clean all
```

2. Remove your old HDP 2.0 components.

```
yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
"gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
"flume*" "mahout*" "hue*"
```

3. Remove your old hdp.repo file:

```
rm /etc/zypp/repos.d/hdp.list
```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

```
http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/
updates/2.2.6.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo
```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

```
zypper install "hadoop" "hadooplzo" "webhcat" "oozie" "collectd"
"gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "tez" "storm"
"falcon" "flume" "phoenix" "accumulo" "mahout" "hue" "knox"
"hdp_mon_nagios_addons"

zypper install hive-webhcat webhcat-tar-hive webhcat-tar-pig

zypper up -r HDP-2.2.6.0

zypper install oozie-client
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded. For example, to verify hdfs, hive, and hcatlog:

```
rpm -qa | grep hdfs, && rpm -qa | grep hive && rpm -qa | grep
hcatalog
```

No component files names should appear in the returned list.

Ubuntu 12

1. On all hosts, clean the apt-get repository.

```
apt-get clean all
```

2. Remove your old HDP 2.0 components.

```
yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
"gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
"flume*" "mahout*" "hue*"
```

3. Remove your old hdp.repo file:

```
rm /etc/apt/sources.list.d/hdp.list
```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

```
wget http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/
updates/2.2.6.0/hdp.list -O /etc/apt/sources.list.d/hdp.list
```

5. Run an update:

```
apt-get update
```

6. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

```
apt-get install "hadoop" "hadooplzo" "webhcat" "oozie"
"collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive"
"tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout"
"knox" "hdp_mon_nagios_addons"
```

```
apt-get install hive-webhcat webhcat-tar-hive webhcat-tar-pig
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

Debian 6

1. On all hosts, clean the apt-get repository.

   ```
   apt-get clean all
   ```

2. Remove your old HDP 2.0 components. This command uninstalls the HDP 2.0 components. It leaves the user data, and metadata, but removes your configurations:

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig" "hdfs*" "sqoop*" "zookeeper*" "hbase*" "hive*"
   "flume*" "mahout*" "hue*"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/apt/sources.list.d/hdp.list
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget http://public-repo-1.hortonworks.com/HDP/debian6/2.x/
     updates/2.2.6.0/hdp.list -O /etc/apt/sources.list.d/hdp.list
     ```

5. Run an update:

   ```
   apt-get update
   ```

6. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 2.2 components:

   ```
   apt-get install "hadoop" "hadooplzo" "webhcat" "oozie"
   "collectd" "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive"
   "tez" "storm" "falcon" "flume" "phoenix" "accumulo" "mahout"
   "knox" "hdp_mon_nagios_addons"

   apt-get install hive-webhcat webhcat-tar-hive webhcat-tar-pig
   ```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

2.3. Symlink Directories with hdp-select

Warning

HDP 2.2 installs hdp-select automatically with the installation or upgrade of the first HDP component. If you have not already upgraded ZooKeeper, hdp-select has not been installed.

To prevent version-specific directory issues for your scripts and updates, Hortonworks provides hdp-select, a script that symlinks directories to hdp-current and modifies paths for configuration directories.

- Before you run hdp-select, remove one link:

  ```
  rm /usr/bin/oozie
  ```

- Run hdp-select set all on your NameNode and all your DataNodes:

  ```
  hdp-select set all 2.2.6.0-<$version>
  ```

  For example:

  ```
  /usr/bin/hdp-select set all 2.2.6.0-2
  ```

2.4. Configure and Start Apache ZooKeeper

Tip

If you are running a highly available cluster, upgrade Apache ZooKeeper **before** you upgrade HDFS. This best practice lets the upgraded ZKFC work with your primary NameNode and your Standby NameNode.

1. Replace your configuration after upgrading. Replace the ZooKeeper template configuration in /etc/zookeeper/conf.

2. Start ZooKeeper.

   On all the ZooKeeper server host machines, run the following command to start ZooKeeper and the ZKFC:

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

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

2.5. Configure Hadoop

RHEL/CentOS/Oracle Linux

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/cpu settings in yarn-site.xml and mapred-site.xml

2. Paths have changed in HDP 2.2. make sure you remove old path specifications from hadoop-env.sh, such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-
   amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load, as this is not where lzo is installed.

SLES

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/cpu settings in yarn-site.xml and mapred-site.xml
2. Paths have changed in HDP 2.2. make sure you remove old path specifications from hadoop-env.sh, such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load, as this is not where lzo is installed.

Ubuntu/Debian

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/cpu settings in yarn-site.xml and mapred-site.xml
2. Paths have changed in HDP 2.2. make sure you remove lines from hadoop-env.sh such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load as this is not where lzo is installed.

2.6. Start Hadoop Core

Warning

Before you start HDFS on an HA system you must start the ZooKeeper service. If you do not start the ZKFC, there can be failures.

To start HDFS, run commands as the $HDFS_USER.

1. Change the value of `hadoop.rpc.protection` to authentication in `core-site.xml`:

   ```
   <property>
    <name>hadoop.rpc.protection</name>
    <value>authentication</value>
   </property>
   ```

2. If you are upgrading from an HA NameNode configuration, start all JournalNodes. On each JournalNode host, run the following command:

   ```
   su -l <HDFS_USER> -c "/usr/hdp/current/hadoop-hdfs-journalnode/../hadoop/sbin/hadoop-daemon.sh start journalnode"
   ```

Important

All JournalNodes must be running when performing the upgrade, rollback, or finalization operations. If any JournalNodes are down when running any such operation, the operation fails.

3. To successfully upgrade to HDP 2.2, you must change the value of hadoop.rpc.protection to authentication in core-site.xml:

```
<property>
 <name>hadoop.rpc.protection</name>
 <value>authentication</value>
</property>
```

4. Start the NameNode.

   Because the file system version has now changed you must start the NameNode manually. On the active NameNode host, run the following command:

```
su -l <HDFS_USER> -c "/usr/hdp/current/hadoop-hdfs-namenode/../hadoop/sbin/
hadoop-daemon.sh start namenode -upgrade"
```

   On a large system, this can take a long time to complete.

   

   Note

   Run this command with the -upgrade option only once. After you have completed this step, you can bring up the NameNode using this command without including the `-upgrade` option.

   To check if the Upgrade is in progress, check that the `\previous` directory has been created in \NameNode and \JournalNode directories. The `\previous` directory contains a snapshot of the data before upgrade.

   In a NameNode HA configuration, this NameNode will not enter the standby state as usual. Rather, this NameNode will immediately enter the active state, perform an upgrade of its local storage directories, and also perform an upgrade of the shared edit log. At this point, the standby NameNode in the HA pair is still down. It will be out of sync with the upgraded active NameNode. To synchronize the active and standby NameNode, re-establishing HA, re-bootstrap the standby NameNode by running the NameNode with the '-bootstrapStandby' flag. Do NOT start this standby NameNode with the '-upgrade' flag.

```
su -l <HDFS_USER> -c "hdfs namenode -bootstrapStandby -force"
```

   where <HDFS_USER> is the HDFS service user. For example, hdfs.

   The `bootstrapStandby` command will download the most recent fsimage from the active NameNode into the `$dfs.name.dir` directory of the standby NameNode. You can enter that directory to make sure the fsimage has been successfully downloaded. After verifying, start the ZKFailoverController, then start the standby NameNode. You can check the status of both NameNodes using the Web UI.

5. Verify that the NameNode is up and running:

```
ps -ef|grep -i NameNode
```

6. Start the Secondary NameNode. On the Secondary NameNode host machine, run the following command:

```
su -l <HDFS_USER> -c "/usr/hdp.current/hadoop-client/sbin/
hadoop-daemon.sh start secondarynamenode
```

7. Verify that the Secondary NameNode is up and running:

   ```
   ps -ef|grep SecondaryNameNode
   ```

8. Start DataNodes.

   On each of the DataNodes, enter the following command. If you are working on a non-secure DataNode, use $HDFS_USER. For a secure DataNode, use root.

   ```
   su -l <HDFS_USER> -c "/usr/hdp/current/hadoop-hdfs-datanode/../
   hadoop/sbin/hadoop-daemon.sh --config /etc/hadoop/conf start
   datanode"
   ```

9. Verify that the DataNode process is up and running:

   ```
   ps -ef|grep DataNode
   ```

10. Verify that NameNode can go out of safe mode.

    ```
    <hdfs dfsadmin -safemode wait

    Safemode is OFF
    ```

    In general, it takes 5-10 minutes to get out of safemode. For thousands of nodes with millions of data blocks, getting out of safemode can take up to 45 minutes.

2.7. Verify HDFS Filesystem Health

Analyze if the filesystem is healthy.

1. Run the fsck command on namenode as $HDFS_USER:

   ```
   hdfs fsck / -files -blocks -locations > dfs-new-fsck-1.log
   ```

2. Run hdfs namespace and report.

   a. List directories.

      ```
      hdfs dfs -ls -R / > dfs-new-lsr-1.log
      ```

   b. Run report command to create a list of DataNodes in the cluster.

      ```
      hdfs dfsadmin -report > dfs-new-report-1.log
      ```

3. Compare the namespace report from before the upgrade with the report after the upgrade. Verify that user files exist after upgrade:

   ```
   dfs-old-fsck-1.log < -- > dfs-new-fsck-1.log dfs-old-lsr-1.log <
   -- > dfs-new-lsr-1.log
   ```

Note

You must do this comparison manually to catch all errors.

4. From the NameNode WebUI, see if all DataNodes are up and running.

   `http://<namenode>:50070`

2.8. Configure YARN and MapReduce

After you upgrade hadoop, complete the following steps to update your configs.

1. Upload the MapReduce tarball to HDFS. As the HDFS user, for example 'hdfs':

```
hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/mapreduce/
hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hadoop/mapreduce.tar.gz /hdp/apps/
2.2.6.0-<$version>/mapreduce/
hdfs dfs -chown -R hdfs:hadoop /hdp
hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/mapreduce
hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/mapreduce/mapreduce.tar.
gz
```

2. Make the following changes to mapred-site.xml:

   - Add:

```
<property>
 <name>mapreduce.admin.map.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
 <final>true</final>
</property>

 <property>
 <name>yarn.app.mapreduce.am.admin-command-opts</name>
 <value>-Dhdp.version=${hdp.version}</value>
</property>

<property>
 <name>mapreduce.application.framework.path</name>
 <value>/hdp/apps/${hdp.version}/mapreduce/mapreduce.tar.gz#mr-framework</
value>
</property>
```

   - Modify the following existing properties to include ${hdp.version}:

```
<property>
 <name>mapreduce.admin.user.env</name>
 <value>LD_LIBRARY_PATH=/usr/hdp/${hdp.version}/hadoop/lib/native:/usr/
hdp/${hdp.version}/hadoop/lib/native/Linux-amd64-64</value>
</property>

<property>
 <name>mapreduce.admin.reduce.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
</property>
```

```
<property>
 <name>mapreduce.application.classpath</name>
 <value>$PWD/mr-framework/hadoop/share/hadoop/mapreduce/*:$PWD/mr-
 framework/hadoop/share/hadoop/mapreduce/lib/*:$PWD/mr-framework/hadoop/
share/hadoop/common/*:$PWD/mr-
 framework/hadoop/share/hadoop/common/lib/*:$PWD/mr-framework/hadoop/
share/hadoop/yarn/*:$PWD/mr-
 framework/hadoop/share/hadoop/yarn/lib/*:$PWD/mr-framework/hadoop/share/
hadoop/hdfs/*:$PWD/mr-
 framework/hadoop/share/hadoop/hdfs/lib/*:/usr/hdp/${hdp.version}/hadoop/
lib/hadoop-lzo-0.6.0.${hdp.version}.
 jar:/etc/hadoop/conf/secure</value>
</property>
```

3. Add the following property to yarn-site.xml:

```
<property>
 <name>yarn.application.classpath</name>
 <value>$HADOOP_CONF_DIR,/usr/hdp/${hdp.version}/hadoop-client/*,/usr/hdp/
${hdp.version}/hadoop-
 client/lib/*,/usr/hdp/${hdp.version}/hadoop-hdfs-client/*,/usr/hdp/${hdp.
version}/hadoop-hdfs-
 client/lib/*,/usr/hdp/${hdp.version}/hadoop-yarn-client/*,/usr/hdp/${hdp.
version}/hadoop-yarn-client/lib/*<
 /value>
</property>
```

4. For secure clusters, you must create and configure the container-executor.cfg configuration file:

   - Create the container-executor.cfg file in `/etc/hadoop/conf/`.
   - Insert the following properties:

     ```
     yarn.nodemanager.linux-container-executor.group=hadoop
     banned.users=hdfs,yarn,mapred
     min.user.id=1000
     ```

     - yarn.nodemanager.linux-container-executor.group - Configured value of yarn.nodemanager.linux-container-executor.group. This must match the value of yarn.nodemanager.linux-container-executor.group in yarn-site.xml.
     - banned.users - Comma-separated list of users who can not run container-executor.
     - min.user.id - Minimum value of user id. This prevents system users from running container-executor.
     - allowed.system.users - Comma-separated list of allowed system users.
   - Set the file /etc/hadoop/conf/container-executor.cfg file permissions to only be readable by root:

     ```
     chown root:hadoop /etc/hadoop/conf/container-executor.cfg
     chmod 400 /etc/hadoop/conf/container-executor.cfg
     ```

   - Set the container-executor program so that only root or hadoop group users can run it:

```
chown root:hadoop /usr/hdp/${hdp.version}/hadoop-yarn-server-nodemanager/
bin/container-executor
chmod 6050 /usr/hdp/${hdp.version}/hadoop-yarn-server-nodemanager/bin/
container-executor
```

2.9. Start YARN/MapReduce Services

Application Timeline Server has changed. If you installed ATS with HDP 2.0, you need to add the following properties to yarn-site.xml to continue using ATS:

```
yarn.timeline-service.leveldb-timeline-store.path=/var/log/hadoop-yarn/
timeline
yarn.timeline-service.leveldb-timeline-store.ttl-interval-ms=300000

** If you are upgrading to HDP 2.1.3 or higher, use the following setting: **
yarn.timeline-service.store-class=org.apache.hadoop.yarn.server.timeline.
LeveldbTimelineStore

** If you are upgrading to HDP 2.1.2, use the following setting: **
yarn.timeline-service.store-class=org.apache.hadoop.yarn.server.
applicationhistoryservice.timeline.LeveldbTimelineStore

yarn.timeline-service.ttl-enable=true
yarn.timeline-service.ttl-ms=2678400000
yarn.timeline-service.generic-application-history.store-class=org.apache.
hadoop.yarn.server.applicationhistoryservice.NullApplicationHistoryStore
yarn.timeline-service.webapp.address={PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:8188
yarn.timeline-service.webapp.https.address=
{PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:8190
yarn.timeline-service.address={PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:10200
```

Add the following properties to hive-site.xml:

```
hive.execution.engine=mr
hive.exec.failure.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.exec.post.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.exec.pre.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.tez.container.size={map-container-size}

** If mapreduce.map.memory.mb > 2GB then set it equal to mapreduce.map.memory.

Otherwise, set it equal to mapreduce.reduce.memory.mb **

hive.tez.java.opts=-server -Xmx800m -Djava.net.preferIPv4Stack=true -
XX:NewRatio=8
-XX:+UseNUMA -XX:+UseParallelGC
```

Use configuration values appropriate for your environment. For example, the value "800" in the preceding example is an example, not a requirement.

To start YARN, run commands as a YARN user. To start MapReduce, run commands as a MapReduce user.

1. If you are using an HA enabled system, you must upgrade the ZooKeeper service and it must be running.

2. Start the ResourceManager on all your ResourceManager hosts.

```
su -l yarn -c "/usr/hdp/current/hadoop-yarn-resourcemanager/
sbin/yarn-daemon.sh start resourcemanager"

ps -ef | grep -i resourcemanager
```

3. Start the NodeManager on all your NodeManager hosts.

```
su - l yarn -c "/usr/hdp/current/hadoop-yarn-nodemanager/sbin/
yarn-daemon.sh start nodemanager"

ps -ef | grep -i nodemanager
```

4. To start MapReduce, run the following commands:

```
su -l yarn -c "/usr/hdp/current/hadoop-mapreduce-historyserver/
sbin/mr-jobhistory-daemon.sh

start historyserver"

ps -ef | grep -i jobhistoryserver
```

2.10. Run Hadoop Smoke Tests

To smoke test your Hadoop upgrade, you can run the following MapReduce job as a regular user.

The job uses MapReduce to write 100MB of data into HDFS with RandomWriter.

```
hadoop jar /usr/hdp/current/hadoop-mapreduce-
client/hadoop-mapreduce-examples.jar randomwriter -
Dtest.randomwrite.total_bytes=10000000 test-after-upgrade.
```

You should see messages similar to:

```
map 0% reduce 0%
…map 100% reduce 100%
Job …. completed successfully
```

MapReduce works successfully. You can now upgrade your other components.

Basic Troubleshooting

To find the number of active nodes and NodeManagers, access the ResourceManager web UI:

```
http://<resource manager host>:8088/cluster/nodes
```

The number of active nodes should be equal to the number of nodemanagers.

Accessing error messages:

1. Access the ApplicationMaster WebUI to view the container logs.

2. In your console log for the MapReduce job, look for a line in this format:

```
13/10/02 17:57:21 INFO mapreduce.Job: The url to track
the job: http://<resource manager host>:8088/proxy/
application_1380673658357_0007/
```

3. Select the logs link under ApplicationMaster table. It will redirect you to the container logs. Error messages display here.

2.11. Configure and Start Apache HBase

1. Start the ZooKeeper zkServer and NameNode services.
2. Replace your configuration after upgrading. Replace the Apache HBase template configuration in /etc/hbase/conf.
3. As the HBase user, run an upgrade:

   ```
   sudo su -l $HBASE_USER -c "hbase upgrade -execute"
   ```

   You should see a completed Znode upgrade with no errors.
4. Start services. Run as root and suppose $HBASE_USER = hbase:

   ```
   su -l hbase -c "/usr/hdp/current/hbase-master/bin/hbase-
   daemon.sh start master; sleep 25"
   ```

   ```
   su -l hbase -c "/usr/hdp/current/hbase-regionserver/bin/hbase-
   daemon.sh start regionserver"
   ```
5. Check processes.

   ```
   ps -ef | grep -i hmaster
   ```

   ```
   ps -ef | grep -i hregion
   ```

2.12. Configure and Start Apache Hive and Apache HCatalog

Warning

In HDP 2.1.3 (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 Apache Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

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. Upgrade the Hive Metastore database schema. Restart the Hive Metastore database and run:

```
/usr/hdp/current/hive-metastore/bin/schematool -upgradeSchema -
dbType <$databaseType>
```

The value for $databaseType can be derby, mysql, oracle or postgres.

Note

If you are using Postgres 8 and Postgres 9, you should reset the Hive Metastore database owner to <HIVE_USER>, run the following commands:

```
sudo <POSTGRES_USER>

ALTER DATABASE <HIVE-METASTORE-DB-NAME> OWNER
TO<HIVE_USER>
```

Note

If you are using Oracle 11, you may see the following error message:

```
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
optimize.mapjoin.mapreduce does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
heapsize does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
server2.enable.impersonation does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
semantic.analyzer.factory.impl does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.auto.
convert.sortmerge.join.noconditionaltask does not exist
Metastore connection URL: jdbc:oracle:thin:@//ip-172-31-42-1.ec2.
internal:1521/XE
Metastore Connection Driver : oracle.jdbc.driver.OracleDriver
Metastore connection User: hiveuser
Starting upgrade metastore schema from version 0.13.0 to 0.14.0
Upgrade script upgrade-0.13.0-to-0.14.0.oracle.sql
Error: ORA-00955: name is already used by an existing object
 (state=42000,code=955)
Warning in pre-upgrade script pre-0-upgrade-0.13.0-to-0.14.0.
oracle.sql: Schema script failed, errorcode 2
Completed upgrade-0.13.0-to-0.14.0.oracle.sql
schemaTool completed
```

You can safely ignore this message. The error is in the pre-upgrade script and can be ignored; the schematool succeeded.

2. Download and extract the HDP companion files (see "Download the Companion Files" in Chapter 1 of the Manual Install Guide).

   Copy the hive-site.xml file in the configuration_files/hive directory of the extracted companion files to the etc/hive/conf directory on your Hive host machine. This new version of hive-site.xml contains new properties for HDP-2.2 features.

3. Edit hive-site.xml and modify the properties based on your environment. Search for TODO in the file for the properties to replace.

- Edit the connection properities for your Hive metastore database in hive- site.xml:

```
<property>
 <name>javax.jdo.option.ConnectionURL</name>
 <value>jdbc:mysql://TODO-HIVE-METASTORE-DB-SERVER:TODO-HIVE-METASTORE-DB-
PORT/
 TODO-HIVE-METASTORE-DB-NAME?createDatabaseIfNotExist=true</value>
 <description>Enter your Hive Metastore Connection URL, for example if
 MySQL: jdbc:mysql://localhost:3306/mysql?createDatabaseIfNotExist=true</
description>
</property>

<property>
 <name>javax.jdo.option.ConnectionUserName</name>
 <value>TODO-HIVE-METASTORE-DB-USER-NAME</value>
 <description>Enter your Hive Metastore database user name.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionPassword</name>
 <value>TODO-HIVE-METASTORE-DB-PASSWORD</value>
 <description>Enter your Hive Metastore database password.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionDriverName</name>
 <value>TODO-HIVE-METASTORE-DB-CONNECTION-DRIVER-NAME</value>
 <description>Enter your Hive Metastore Connection Driver Name, for
 example if MySQL:
 com.mysql.jdbc.Driver</description>
</property>
```

- Edit the following properties in the hive-site.xml file:

```
<property>
 <name>fs.file.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.file.impl.disable.cache</
description>
</property>

<property>
 <name>fs.hdfs.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.hdfs.impl.disable.cache</
description>
</property>
```

- **Optional:** If you want Hive Authorization, set the following Hive authorization parameters in the hive-site.xml file:

```
<property>
 <name>hive.security.authorization.enabled</name>
 <value>true</value>
</property>

<property>
 <name>hive.security.authorization.manager</name>
```

```
 <value>org.apache.hadoop.hive.ql.security.authorization.
StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.metastore.authorization.manager</name>
 <value>org.apache.hadoop.hive.ql.security.authorization.
StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.authenticator.manager</name>
 <value>org.apache.hadoop.hive.ql.security.ProxyUserAuthenticator</value>
</property>
```

- For a remote Hive metastore database, set the IP address (or fully-qualified domain name) and port of the metastore host using the following hive-site.xml property value.

  (To enable HiveServer2, leave this property value empty.)

  ```
  <property>
   <name>hive.metastore.uris</name>
   <value>thrift://$metastore.server.full.hostname:9083</value>
   <description>URI for client to contact metastore server. sbrTo enable
   HiveServer2,
   leave the property value empty. </description>
  </property>
  ```

  You can further fine-tune your configuration settings based on node hardware specifications, using the HDP utility script.

- Disable autocreation of schemas to match HDP 2.1+ configurations. Edit `hive-site.xml` to set the value of `datanucleus.autoCreateSchema` to false.

  ```
  <property>
   <name>datanucleus.autoCreateSchema</name>
   <value>false</value>
   <description>Creates necessary schema on a startup if one doesn't exist.
  </description>
  </property>
  ```

4. Start Hive.

   On the Hive Metastore host machine, run the following commands:

   ```
   u - hive

   nohup /usr/hdp/current/hive-metastore/bin/hive --service
   metastore>/var/log/hive/hive.out 2>/var/log/hive/hive.log &
   ```

5. Start Hive Server2.

   On the Hive Server2 host machine, run the following commands:

   ```
   su - hive

   /usr/hdp/current/hive-server2/bin/hiveserver2 >/var/log/hive/
   hiveserver2.out 2> /var/log/hive/hiveserver2.log &
   ```

where $HIVE_USER is the Hive Service user. For example, hive.

2.13. Configure and Start Apache Oozie

Upgrading Apache Oozie is a complex process. Although the instructions are straightforward, set aside a dedicated block of time to upgrade oozie clients and servers.

Perform the following preparation steps on each oozie server host:

1. You must restore oozie-site.xml from your backup to the conf directory on each oozie server and client.
2. Copy the JDBC jar to libext-customer.
   a. Create the /usr/hdp/current/oozie-client/libext-customer directory.

      ```
      cd /usr/hdp/current/oozie-client mkdir libext-customer
      ```

   b. Grant read/write/execute access to all users for the libext-customer directory.

      ```
      chmod -R 777 /usr/hdp/current/oozie-client/libext-customer
      ```

3. Copy these files to the libext-customer directory

   ```
   cp /usr/hdp/current/hadoop/lib/hadooplzo*.jar /usr/hdp/current/
   oozie/libext-customer
   ```

   ```
   cp /usr/share/HDP-oozie/ext-2.2.zip /usr/hdp/current/oozie/
   libext-customer/
   ```

   If you do not have an LZO JAR file, you must enable LZO compression first. See Install Compression Libraries.

4. Extract share-lib.

   ```
   cd /usr/hdp/current/oozie
   ```

   ```
   tar xzvf /usr/hdp/current/oozie/oozie-sharelib.tar.gz
   ```

   ```
   su -l hdfs -c "hdfs dfs -mkdir -p /user/oozie"
   ```

   ```
   su -l hdfs -c "hdfs dfs -copyFromLocal /usr/hdp/current/oozie/
   share /user/oozie/."
   ```

   You can expect to see complaints that some files already exist. Delete any exisitng /oozie/share and replace it with the newly-extracted files.

   ```
   su -l hdfs -c "hdfs dfs -chown oozie:hadoop /user/oozie" su -l
   hdfs -c "hdfs dfs -chmod -R 755 /user/oozie"
   ```

5. if a previous version of Oozie was created using auto schema creation, run the following SQL query:

   ```
   insert into oozie_sys (name, data) values ('db.version', '2.5');
   ```

6. As the Oozie user (not root), run the upgrade.

   ```
   su $OOZIE_USER

   /usr/hdp/current/oozie-server/bin/ooziedb.sh upgrade -run
   ```

7. As root, prepare the Oozie WAR file.

   ```
   sudo su -l oozie -c "/usr/hdp/current/oozie/bin/oozie-setup.sh
   prepare-war -d /usr/hdp/current/oozie/libext-customer"
   ```

   Look for console output to indicate success. For example, if you are using MySQL you should see something similar to:

   ```
   INFO: Adding extension: libext-customer/mysql-connector-java.jar
   New Oozie WAR file with added 'JARs' at /var/lib/oozie/oozie-server/webapps/
   oozie.war
   ```

8. Add the following property to oozie-log4j.properties:

   ```
   log4j.appender.oozie.layout.ConversionPattern=%d{ISO8601} %5p
   %c{1}:%L - SERVER[${oozie.instance.id}] %m%n
   ```

   where ${oozie.instance.id} is determined by oozie, automatically.

9. Start Oozie as the Oozie user:

   ```
   sudo su -l <OOZIE_User> -c "export OOZIE_CATALINA_HOME='usr/lib/bigtop-
   tomcat'; $OOZIE_CATALINA_HOME/catalina.sh $OOZIE_CATALINA_HOME/setclasspath.
   sh
   /usr/hdp/current/oozie-server/bin/oozied.sh start”
   ```

10. Check processes.

    ```
    ps -ef | grep -i oozie
    ```

2.14. Configure and Start Apache WebHCat (Templeton)

1. You must replace your configuration after upgrading. Copy /etc/webhcat/conf from the template to the conf directory in webhcat hosts.

2. Modify the Apache WebHCat configuration files.

   - Upload Pig, Hive and Sqoop tarballs to HDFS as the $HDFS_User. In this example, hdfs:

     ```
     hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/pig/
     hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/hive/
     hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/sqoop/
     hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/pig/pig.tar.gz /hdp/apps/2.2.6.
     0-<$version>/pig/
     hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hive/hive.tar.gz /hdp/apps/2.2.
     6.0-<$version>/hive/
     hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/sqoop/sqoop.tar.gz /hdp/apps/2.
     2.6.0-<$version>/sqoop/
     hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/pig
     ```

```
hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/pig/pig.tar.gz
hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/hive
hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/hive/hive.tar.gz
hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/sqoop
hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/sqoop/sqoop.tar.gz
hdfs dfs -chown -R hdfs:hadoop /hdp
```

- Update the following properties in the webhcat-site.xml configuration file, as their values have changed:

```
<property>
 <name>templeton.pig.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/pig/pig.tar.gz</value>
</property>

<property>
 <name>templeton.hive.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/hive/hive.tar.gz</value>
</property>

<property>
 <name>templeton.streaming.jar</name>
 <value>hdfs:///hdp/apps/${hdp.version}/mapreduce/hadoop-streaming.jar</
value>
 <description>The hdfs path to the Hadoop streaming jar file.</
description>
</property>

<property>
 <name>templeton.sqoop.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/sqoop/sqoop.tar.gz</value>
 <description>The path to the Sqoop archive.</description>
</property>

<property>
 <name>templeton.sqoop.path</name>
 <value>sqoop.tar.gz/sqoop/bin/sqoop</value>
 <description>The path to the Sqoop executable.</description>
</property>

<property>
 <name>templeton.sqoop.home</name>
 <value>sqoop.tar.gz/sqoop</value>
 <description>The path to the Sqoop home in the exploded archive.</
description>
</property>
```

Note

You do not need to modify ${hdp.version}.

- Remove the following obsolete properties from webhcat-site.xml:

```
<property>
 <name>templeton.controller.map.mem</name>
 <value>1600</value>
 <description>Total virtual memory available to map tasks.</description>
</property>
```

```
<property>
 <name>hive.metastore.warehouse.dir</name>
 <value>/path/to/warehouse/dir</value>
</property>
```

- Add new proxy users, if needed. In core-site.xml, make sure the following properties are also set to allow WebHcat to impersonate your additional HDP 2.2 groups and hosts:

```
<property>
 <name>hadoop.proxyuser.hcat.groups</name>
 <value>*</value>
</property>

<property>
 <name>hadoop.proxyuser.hcat.hosts</name>
 <value>*</value>
</property>
```

Where:

`hadoop.proxyuser.hcat.group`

is a comma-separated list of the Unix groups whose users may be impersonated by 'hcat'.

`hadoop.proxyuser.hcat.hosts`

A comma-separated list of the hosts that are allowed to submit requests by 'hcat'.

3. Start WebHCat:

```
sudo su -l $WEBHCAT_USER -c "/usr/hdp/current/hive-hcatalog/
sbin/webhcat_server.sh start"
```

4. Smoke test WebHCat.

- At the WebHCat host machine, run the following command:

```
http://$WEBHCAT_HOST_MACHINE:50111/templeton/v1/status
```

- If you are using a secure cluster, run the following command:

```
curl --negotiate -u: http://cluster.$PRINCIPAL.$REALM:50111/
templeton/v1/status {"status":"ok","version":"v1"}
[machine@acme]$
```

2.15. Configure and Start Apache Pig

1. Replace your configuration after upgrading. Copy `/etc/pig/conf` from the template to the conf directory in pig hosts.

2. To validate the Apache Pig upgrade, complete the following steps:

   a. On the host machine where Pig is installed, run the following commands:

```
su - $HDFS_USER/usr/hdp/current/hadoop/bin/hadoop
fs -copyFromLocal /etc/passwd passwd
```

b. Create a Pig script file named `/tmp/id.pig` that contains the following Pig Latin commands:

```
A = load 'passwd' using PigStorage(':');B = foreach A generate $0 as id;
 store B into '/tmp/id.out';
```

c. Run the Pig script:

```
su - $HDFS_USER
pig -l /tmp/pig.log /tmp/id.pig
```

2.16. Configure and Start Apache Sqoop

1. Replace your configuration after upgrading. Copy `/etc/sqoop/conf` from the template to the conf directory in sqoop hosts.

2. As the HDFS Service user, upload the Apache Sqoop tarball to HDFS.

```
su -c "hdfs dfs -mkdir -p /hdp/apps/2.2.0.0-<$version>/sqoop"

hdfs su -c "hdfs dfs -chmod -R 555 /hdp/apps/2.2.0.0-<$version>/sqoop"

hdfs su -c "hdfs dfs -chown -R hdfs:hadoop /hdp/apps/2.2.0.0-<$version>/
sqoop"

hdfs su -c "hdfs dfs -put /usr/hdp/2.2.0.0-<$version>/sqoop/sqoop.tar.gz /
hdp/apps/2.2.0.0-<$version>/sqoop/sqoop.tar.gz"

hdfs su -c "hdfs dfs -chmod 444 /hdp/apps/2.2.0.0-<$version>/sqoop/sqoop.
tar.gz"
```

3. If you are using the MySQL database as a source or target, then the MySQL connector jar must be updated to 5.1.29 or later.

   Refer to the MySQL web site for information on updating the MySQL connector jar.

4. Because Sqoop is a client tool with no server component, you will need to run your own jobs to validate the upgrade.

2.17. Configure, Start, and Validate Apache Flume

1. If you have not already done so, upgrade Apache Flume. On the Flume host machine, run the following command:

   - For **RHEL/CentOS/Oracle Linux**:

     ```
     yum upgrade flume
     ```

   - For **SLES**:

     ```
     zypper update flume
     ```

```
zypper remove flume

zypper se -s flume
```

You should see Flume in the output.

Install Flume:

```
zypper install flume
```

- For **Ubuntu/Debian**:

```
apt-get install flume
```

2. To confirm that Flume is working correctly, create an example configuration file. The following snippet is a sample configuration that can be set using the properties file. For more detailed information, see the "Flume User Guide."

```
agent.sources = pstream
agent.channels = memoryChannel
agent.channels.memoryChannel.type = memory

agent.sources.pstream.channels = memoryChannel
agent.sources.pstream.type = exec
agent.sources.pstream.command = tail -f /etc/passwd

agent.sinks = hdfsSink
agent.sinks.hdfsSink.type = hdfs
agent.sinks.hdfsSink.channel = memoryChannel
agent.sinks.hdfsSink.hdfs.path = hdfs://tmp/flumetest
agent.sinks.hdfsSink.hdfs.fileType = SequenceFile
agent.sinks.hdfsSink.hdfs.writeFormat = Text
```

The source here is defined as an exec source. The agent runs a given command on startup, which streams data to stdout, where the source gets it. The channel is defined as an in-memory channel and the sink is an HDFS sink.

3. Given this configuration, you can start Flume as follows:

```
$ bin/flume-ng agent --conf ./conf --conf-file example.conf --name a1 -
Dflume.root.logger=INFO,console
```

Note

The directory specified for `--conf agrument` would include a shell script flume-env.sh and potentially a log4j properties file. In this example, we pass a Java option to force Flume to log to the console and we go without a custom environment script.

4. After validating data in `hdfs://tmp/flumetest`, stop Flume and resore any backup files. Copy `/etc/flume/conf` to the conf directory in Flume hosts.

2.18. Configure, Start, and Validate Apache Mahout

Replace your configuration after upgrading. Copy `/etc/mahout/conf` from the template to the `conf` directory in mahout hosts.

To validate Apache Mahout:

1. Create a test user named "testuser" in the Linux cluster and in HDFS, and log in as that user.

   ```
   hdfs dfs -put /tmp/sample-test.txt /user/testuser
   ```

2. Export the required environment variables for Mahout:

   ```
   export JAVA_HOME="your_jdk_home_install_location_here
   export HADOOP_HOME=/usr/hdp/current/hadoop-client
   export MAHOUT_HOME=/usr/hdp.current/mahout-client
   export PATH="$PATH":$HADOOP_HOME/bin:$MAHOUT_HOME/bin
   export CLASSPATH="$CLASSPATH":$MAHOUT_HOME
   ```

3. Upload a few megabytes of natural-language plain text to the Linux server as `/tmp/sample-test.txt`.

4. Transfer the `sample-test file` to a subdirectory of the testuser's HDFS home directory.

   ```
   hdfs dfs -mkdir /user/testuser/testdata
   hdfs dfs -put /tmp/sample-test.txt /user/testuser/testdata
   ```

5. Enter the mahout command to convert the plain text file `sample-test.txt` into a sequence file stored in the output directory mahouttest:

   ```
   mahout seqdirectory --input /user/testuser/testdata --output /user/testuser/
   mahouttest -ow --charset utf-8
   ```

2.19. Configure and Start Hue

For HDP 2.2, use the Hue version shipped with HDP 2.2. If you have a previous version of Hue, use the following steps to upgrade Hue.

1. Migrate the hue.ini setting from your old hue.ini configuration file to new hue.ini configuration file.

2. If you are using the embedded SQLite database, remember to restore your database after upgrade.

   To restore the database from a backup, make sure the destination database is empty before copying (if necessary, rename or remove the current destination database), then copy your backup to the destination database.

   For example:

   ```
   su $HUE_USER
   ```

```
cd /var/lib/hue

mv desktop.db desktop.db.old

sqlite3 desktop.db < ~/hue_backup/desktop.bak

exit
```

3. Synchronize the database.

```
cd /usr/lib/hue

source ./build/env/bin/activate

hue syncdb

deactivate
```

4. Start Hue. As a root user, run the following command on the Hue Server:

```
/etc/init.d/hue start
```

2.20. Finalize the Upgrade

You can start HDFS without finalizing the upgrade. When you are ready to discard your backup, you can finalize the upgrade.

1. Verify your filesystem health before finalizing the upgrade and removing the old config.
2. As the $HDFS_USER, enter:

```
hdfs dfsadmin -finalizeUpgrade
```

2.21. Install New HDP 2.2 Services

Install new HDP 2.2 Services:

1. Tez – YARN-based processing framework that greatly improves query response times for Hive and other YARN applications.
2. Phoenix – SQL skin over HBase that makes it easier and faster to build HBase applications.
3. Accumulo – a high-performance data storage and retrieval system with cell-level access control. It is a scalable implementation of Google's Big Table design that works on top of Hadoop and ZooKeeper.
4. Storm – Real-time event stream processing platform that provides fixed, continuous, & low latency processing for very high frequency streaming data.
5. Falcon – Framework for simplifying the development of data management applications in Apache Hadoop.

6. Knox – Apache Knox is the Web/REST API Gateway solution for Hadoop. It provides a single access point for all of Hadoop resources over REST.

7. Slider – YARN application to deploy existing distributed applications on YARN, monitor them, and make them larger or smaller as desired even while the application is running.

8. Ranger – Comprehensive approach to central security policy administration across the core enterprise security requirements of authentication, authorization, accounting and data protection

9. Kafka – an open-source message broker project written in Scala.

10. Spark – an open-source cluster computing engine.

3. Upgrade from HDP 1.3 to HDP 2.2 Manually

HDP 2.2 supports side-by-side installation of HDP 2.2 and above releases, which lets you perform rolling upgrades on your cluster and improve execution times for in-place upgrade. To support side-by-side installation, the HDP package version naming convention for both RPMs and Debs has changed to include the HDP 2.2 product version. For example, hadoop-hdfs is now hadoop-2.2.6.0-hdfs. HDP 2.2 marks the first release where HDP rpms, debs, and directories contain versions in the names to permit side-by-side installations of later HDP releases.

To select from the releases you have installed side-by-side, Hortonworks provides hdp-select, a command that lets you select the active version of HDP from the versions you have selected.

The HDP packages for a complete installation of HDP 2.2 will take about 2.5 GB of disk space.

Warning

Until the upgrade is finalized, no HDFS data is deleted from the cluster. Be sure to review your capacity and ensure that you have extra space available during the upgrade window.

Note

These instructions cover the upgrade between two major releases. If you need to upgrade between two maintenance releases, follow the upgrade instructions in the HDP Release Notes.

Use the following instructions to upgrade to the latest release of HDP from HDP 1.3:

1. Download HDP 2.2
2. Getting Ready to Upgrade
3. Upgrade Hadoop
4. Migrate the HDP Configurations
5. Create Local Directories
6. Start HDFS
7. Upgrade Apache ZooKeeper
8. Upgrade Apache HBase
9. Upgrade Apache Hive and Apache HCatalog
10. Upgrade Apache Oozie

11. Upgrade Apache WebHCat (Templeton)
12. Upgrade Apache Pig
13. Upgrade Apache Sqoop
14. Upgrade Apache Flume
15. Upgrade Apache Mahout
16. Upgrade Hue
17. Finalize the Upgrade
18. Install new HDP 2.2 Services

3.1. Getting Ready to Upgrade

HDP Stack upgrade involves removing HDP 1.x MapReduce and replacing it with HDP 2.x YARN and MapReduce2. Before you begin, review the upgrade process and complete the backup steps.

1. Hardware recommendations

   Although there is no single hardware requirement for installing HDP, there are some basic guidelines. The HDP packages for a complete installation of HDP 2.2 will take up about 2.5 GB of disk space.

2. Back up the following HDP 1.x directories:

   - /etc/hadoop/conf
   - /etc/hbase/conf
   - /etc/hcatalog/conf (Note: With HDP 2.2, /etc/hcatalog/conf is divided into /etc/hive-hcatalog/conf and /etc/hive-webhcat/conf. You cannot use /etc/hcatalog/conf in HDP 2.2.)
   - /etc/hive/conf
   - /etc/pig/conf
   - /etc/sqoop/conf
   - /etc/flume/conf
   - /etc/mahout/conf
   - /etc/oozie/conf
   - /etc/hue/conf
   - /etc/zookeeper/conf

3. **Optional:** Back up your userlogs directories, ${mapred.local.dir}/userlogs.

   Run the fsck command as the HDFS Service user and fix any errors. (The resulting file contains a complete block map of the file system.) For example, in a situation where clusters are unsecure and Kerberos credentials are not required for the HDFS user:

   ```
   su -l <HDFS_USER>
   ```

   ```
   hadoop fsck / -files -blocks -locations > dfs-old-fsck-1.log
   ```

   where $HDFS_USER is the HDFS Service user. For example, hdfs.

4. As the user running the HDFS service (by default, the user is hdfs), run the following commands:

   - Capture the complete namespace of the file system. (The following command does a recursive listing of the root file system.)

     ```
     su -l <HDFS_USER>
     hadoop dfs -lsr / > dfs-old-lsr-1.log
     ```

     where $HDFS_USER is the HDFS Service user. For example, hdfs.

   - Run the report command to create a list of DataNodes in the cluster.

     ```
     su -l <HDFS_USER>
     hadoop dfsadmin -report > dfs-old-report-1.log
     ```

     where $HDFS_USER is the HDFS Service user. For example, hdfs

   - **Optional:** You can copy all or unrecoverable only data stored in HDFS to a local file system or to a backup instance of HDFS.

   - **Optional:** You can also repeat the steps 3 (a) through 3 (c) and compare the results with the previous run to ensure the state of the file system remained unchanged.

5. HBase 0.96.0 and subsequent releases discontinue support for the HFileV1 file format, a common format prior to HBase 0.94. Before you upgrade, check for V1-format files as follows:

   - Download the Apache 0.94.24+HBase tarball in a machine. Run the binaries.

   - On the machine running the HBase 0.94 binaries, point the hbase-site.xml configuration file to a 0.94 cluster.

   - Check for HFiles in V1 format as follows:

     ```
     ./bin/hbase org.apache.hadoop.hbase.util.HFileV1Detector -p
     <hbase root data path>
     ```

   When you run the upgrade check, if “Count of HFileV1” returns any files, start the HBase shell to use major compaction for regions that have HFileV1 format. For example, the following sample output indicates that you need to compact two regions, fa02dac1f38d03577bd0f7e666f12812 and ecdd3eaee2d2fcf8184ac025555bb2af:

   ```
   Tables Processed:
   ```

```
hdfs://localhost:41020/myHBase/.META.
hdfs://localhost:41020/myHBase/usertable
hdfs://localhost:41020/myHBase/TestTable
hdfs://localhost:41020/myHBase/tCount of HFileV1: 2 HFileV1:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812/
 family/249450144068442524
hdfs://localhost:41020/myHBase/usertable/ecdd3eaee2d2fcf8184ac025555bb2af/
 family/249450144068442512

Count of corrupted files: 1
Corrupted Files:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812/
 family/1
Count of Regions with HFileV1: 2
Regions to Major Compact:
hdfs://localhost:41020/myHBase/usertable/fa02dac1f38d03577bd0f7e666f12812
hdfs://localhost:41020/myHBase/usertable/ecdd3eaee2d2fcf8184ac025555bb2af
```

6. **Optional:** If you are upgrading HBase on a secure cluster, flush the ACL table by running the following HBase shell command as the $HBase_User.

   ```
   flush '_acl_'
   ```

7. Stop all HDP 1.3 services (including MapReduce) except HDFS:

   - Stop Nagios. On the Nagios host machine, run the following command:

     ```
     service nagios stop
     ```

   - Stop Ganglia.

     - Run this command on the Ganglia server host machine:

       ```
       /etc/init.d/hdp-gmetad stop
       ```

     - Run this command on all the nodes in your Hadoop cluster:

       ```
       /etc/init.d/hdp-gmond stop
       ```

   - Stop Oozie. On the Oozie server host machine, run the following command:

     ```
     sudo su -l $OOZIE_USER -c "cd $OOZIE_LOG_DIR; /usr/lib/oozie/
     bin/oozie-stop.sh"
     ```

     where:

     $OOZIE_USER is the Oozie Service user. For example, oozie

     $OOZIE_LOG_DIR is the directory where Oozie log files are stored (for example: /var/log/oozie).

   - Stop WebHCat. On the WebHCat host machine, run the following command:

     ```
     su -l $WEBHCAT_USER -c "/usr/lib/hcatalog/sbin/
     webhcat_server.sh stop"
     ```

     where $WEBHCAT_USER is the WebHCat Service user. For example, hcat.

- Stop Hive. On the Hive Metastore host machine and Hive Server2 host machine, run the following command:

  ```
  ps aux | awk '{print $1,$2}' | grep hive | awk '{print $2}' |
  xargs kill >/dev/null 2>&1
  ```

  This stops the Hive Metastore and HCatalog services.

- Stop ZooKeeper. On the ZooKeeper host machine, run the following command:

  ```
  su - $ZOOKEEPER_USER -c "export ZOOCFGDIR=/etc/zookeeper/conf ;
  export ZOOCFG=zoo.cfg ;source /etc/zookeeper/conf/zookeeper-
  env.sh ; /usr/lib/zookeeper-server/bin/zkServer.sh stop"
  ```

  where *$ZOOKEEPER_USER* is the ZooKeeper Service user. For example, zookeeper.

- Stop HBase.

  - Run these commands on all RegionServers:

    ```
    su -l $HBASE_USER -c "/usr/lib/hbase/bin/hbase-daemon.sh --
    config /etc/ hbase/conf stop regionserver"
    ```

  - Run these commands on the HBase Master host machine:

    ```
    su -l $HBASE_USER -c "/usr/lib/hbase/bin/hbase-daemon.sh --
    config /etc/ hbase/conf stop master"
    ```

    where *$HBASE_USER* is the HBase Service user. For example, hbase.

- Stop MapReduce

  - Run these commands on all TaskTrackers slaves:

    ```
    su -l $MAPRED_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh
    --config / etc/hadoop/conf stop tasktracker"
    ```

  - Run these commands on the HistoryServer host machine:

    ```
    su -l $MAPRED_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh
    --config / etc/hadoop/conf stop historyserver"
    ```

  - Run these commands on the node running the JobTracker host machine:

    ```
    su -l $MAPRED_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh
    --config / etc/hadoop/conf stop jobtracker"
    ```

  where $MAPRED_USER is the MapReduce Service user. For example, mapred.

8. As the HDFS user, save the namespace by executing the following command:

   ```
   su -l <HDFS_USER>

   hadoop dfsadmin -safemode enter
   ```

```
hadoop dfsadmin -saveNamespace
```

9. Backup your NameNode metadata.
   - Copy the following checkpoint files into a backup directory:
     - dfs.name.dir/edits
     - dfs.name.dir/image/fsimage
     - dfs.name.dir/current/fsimage
   - Store the layoutVersion of the namenode.

     ```
     ${dfs.name.dir}/current/VERSION
     ```

10. If you have a prior HDFS upgrade in progress, finalize it if you have not done so already.

    ```
    su -l <HDFS_USER>
    ```

    ```
    hadoop dfsadmin -finalizeUpgrade
    ```

11. **Optional:** Back up the Hive Metastore database.

    These instructions are provided for your convenience. Please check your database documentation for the latest backup instructions.

    The following instructions are provided for your convenience. For the latest backup instructions, please check your database documentation.

Table 3.1. Hive Metastore Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sql For example: mysqldump hive > / tmp/mydir/ backup_hive.sql	mysql $dbname < $inputfilename.sql For example: mysql hive </tmp/mydir/ backup_hive.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql For example: sudo -u postgres pg_dump hive > / tmp/mydir/ backup_hive.sql	sudo -u $username psql $databasename < $inputfilename.sql For example: sudo -u postgres psql hive </tmp/ mydir/ backup_hive.sql
Oracle	Connect to the Oracle database using sqlplus. Export the database: exp username/ password@database full=yes file=output_file.dmp	Import the database: imp username/password@database ile=input_file.dmp

12. **Optional**: Back up the Oozie Metastore database.

The following instructions are provided for your convenience. For the latest backup instructions, please check your database documentation.

Table 3.2. Oozie Metastore Database Backup and Restore

Database Type	Backup	Restore

MySQL	`mysqldump $dbname > $outputfilename.sql`
	
	For example:
	
	`mysqldump oozie > / tmp/ mydir/backup_oozie.sql`
		`mysql $dbname < $inputfilename.sql`
		
		For example:
		
		`mysql oozie </tmp/mydir/ backup_oozie.sql`

Postgres	`sudo -u $username pg_dump $databasename > $outputfilename.sql`
	
	For example:
	
	`sudo -u postgres pg_dump oozie > /tmp/mydir/ backup_oozie.sql`
		`sudo -u $username psql $databasename < $inputfilename.sql`
		
		For example:
		
		`sudo -u postgres psql oozie </tmp/mydir/ backup_oozie.sql`

13. **Optional**: Back up the Hue database.

The following instructions are provided for your convenience. For the latest backup instructions, please see your database documentation. For database types that are not listed below, follow your vendor-specific instructions.

Table 3.3. Hue Database Backup and Restore

Database Type	Backup	Restore
MySQL	mysqldump $dbname > $outputfilename.sqlsbr For example: mysqldump hue > /tmp/mydir/ backup_hue.sql	mysql $dbname < $inputfilename.sqlsbr For example: mysql hue < /tmp/mydir/ backup_hue.sql
Postgres	sudo -u $username pg_dump $databasename > $outputfilename.sql sbr For example: sudo -u postgres pg_dump hue > / tmp/mydir/backup_hue.sql	sudo -u $username psql $databasename < $inputfilename.sqlsbr For example: sudo -u postgres psql hue < /tmp/ mydir/backup_hue.sql
Oracle	Connect to the Oracle database using sqlplus. Export the database. For example: exp username/password@database full=yes file=output_file.dmp mysql $dbname < $inputfilename.sqlsbr	Import the database: For example: imp username/password@database file=input_file.dmp
SQLite	/etc/init.d/hue stop su $HUE_USER mkdir ~/hue_backup	/etc/init.d/hue stop cd /var/lib/hue mv desktop.db desktop.db.old

Database Type	Backup	Restore
	sqlite3 desktop.db .dump > ~/hue_backup/desktop.bak /etc/init.d/hue start	sqlite3 desktop.db < ~/hue_backup/desktop.bak /etc/init.d/hue start

14. Stop HDFS.

    a. Run these commands on all DataNodes:

    ```
    su -l $HDFS_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh --config /etc/ hadoop/conf stop datanode"

    su -c "/usr/lib/hadoop/bin/hadoop-daemon.sh --config /etc/ hadoop/conf stop datanode"
    ```

    b. Run these commands on the Secondary NameNode host machine:

    ```
    su -l $HDFS_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh --config /etc/ hadoop/conf stop secondarynamenode”
    ```

    c. Run these commands on the NameNode host machine:

    ```
    su -l $HDFS_USER -c "/usr/lib/hadoop/bin/hadoop-daemon.sh --config /etc/hadoop/conf stop namenode"
    ```

    where $HDFS_USER is the HDFS Service user. For example, hdfs.

15. Verify that edit logs in ${dfs.namenode.name.dir}/current/edits* are empty.

    a. Run the following command:

    ```
    hdfs oev -i ${dfs.namenode.name.dir}/current/ edits_inprogress_* -o edits.out
    ```

    b. Verify edits.out file. It should only have OP_START_LOG_SEGMENT transaction. For example:

    ```
    <?xml version="1.0" encoding="UTF-8"?>
     <EDITS>
     <EDITS_VERSION>-56</EDITS_VERSION>
     <RECORD>
     <OPCODE>OP_START_LOG_SEGMENT</OPCODE>
     <DATA>
     <TXID>5749</TXID>
     </DATA>
     </RECORD>
    ```

    c. If edits.out has transactions other than OP_START_LOG_SEGMENT, run the following steps and then verify edit logs are empty.

    - Start the existing version NameNode.
    - Ensure there is a new FS image file.
    - Shut the NameNode down.

16. Rename or delete any paths that are reserved in the new version of HDFS.

    When upgrading to a new version of HDFS, it is necessary to rename or delete any paths that are reserved in the new version of HDFS. If the NameNode encounters a reserved path during upgrade, it will print an error such as the following:

    ```
    /.reserved is a reserved path and .snapshot is a reserved path
    component in this version of HDFS. Please rollback and delete or
    rename this path, or upgrade with the -renameReserved key-value
    pairs option to automatically rename these paths during upgrade.
    ```

    Specifying -upgrade -renameReserved optional key-value pairs causes the NameNode to automatically rename any reserved paths found during startup. For example, to rename all paths named .snapshot to .my-snapshot and change paths named .reserved to .my-reserved, a user would specify:

    ```
    -upgrade -renameReserved .snapshot=.my-snapshot,.reserved=.my-
    reserved.
    ```

    If no key-value pairs are specified with -renameReserved, the NameNode will then suffix reserved paths with `.<LAYOUT-VERSION>.UPGRADE_RENAMED`, for example:

    ```
    .snapshot.-51.UPGRADE_RENAMED.
    ```

    

    Note

    We recommend that you perform a `-saveNamespace` before renaming paths (running `-saveNamespace` appears in a previous step in this procedure). This is because a data inconsistency can result if an edit log operation refers to the destination of an automatically renamed file.

    Also note that running `-renameReserved` will rename all applicable existing files in the cluster. This may impact cluster applications.

3.2. Upgrade HDP 1.3 Components

The upgrade process to HDP 2.2 involves the following steps. Select your OS:

RHEL/CentOS/Oracle 6

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 1.3 components. This command uninstalls the HDP 1.3 components. It leaves the user data, metadata and modified configurations in place and does not delete them:

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*"
   "hive*" "flume*" "mahout*" "hue" "hue-common" "hue-shell"
   "hdp_mon_nagios_addons"
   ```

3. Remove your old hdp.repo file:

   ```
   rm /etc/yum.repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/2.x/
     updates/2.2.6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
     ```

   - Confirm the HDP repository is configured.

     ```
     yum repolist
     ```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

     ```
     Loaded plugins: fastestmirror, security
     Loading mirror speeds from cached hostfile
     * base: mirrors.cat.pdx.edu
     * extras: linux.mirrors.es.net
     * updates: mirrors.usc.edu
     repo id repo namestatus
     HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 1.3 components:

   ```
   yum install "hadoop" "hadooplzo" "webhcat" "oozie" "collectd"
   "gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "flume"
   "mahout" "hue" "knox" "hdp_mon_nagios_addons"
   ```

   

   Note

   If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded.

   ```
   yum list installed | grep HDP-<old.stack.version.number>
   ```

   No component file names should appear in the returned list.

RHEL/CentOS/Oracle 5 (Deprecated)

1. On all hosts, clean the yum repository.

   ```
   yum clean all
   ```

2. Remove your old HDP 1.3 components. This command uninstalls the HDP 1.3 components. It leaves the user data and metadata but deletes your modified configurations:

   ```
   yum erase "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
   "gccxml*" "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*"
   ```

```
"hive*" "flume*" "mahout*" "hue" "hue-common" "hue-shell"
"hdp_mon_nagios_addons"
```

3. Remove your old hdp.repo file:

```
rm /etc/yum.repos.d/hdp.repo
```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/2.x/
updates/2.2.6.0/hdp.repo -O /etc/yum.repos.d/hdp.repo
```

   - Confirm the HDP repository is configured.

```
yum repolist
```

     You should see something like this. Verify that you have the HDP-2.2.6.0 directory:

```
Loaded plugins: fastestmirror, security
Loading mirror speeds from cached hostfile
* base: mirrors.cat.pdx.edu
* extras: linux.mirrors.es.net
* updates: mirrors.usc.edu
repo id repo namestatus
HDP-2.2.6.0 Hortonworks Data Platform Version - HDP-2.2.6.0
```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 1.3 components:

```
yum install "hadoop" "hadooplzo" "webhcat" "oozie" "collectd"
"gccxml" "pig" "sqoop" "zookeeper" "hbase" "hive" "flume"
"mahout" "hue" "knox" "hdp_mon_nagios_addons"
```

Note

If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded.

```
yum list installed | grep HDP-<old.stack.version.number>
```

   No component file names should appear in the returned list.

SLES 11 SP 1

1. On all hosts, clean the yum repository.

```
zypper clean -all
```

2. Remove your old HDP 1.3 components. This command uninstalls the HDP 1.3 components. It leaves the user data and metadata but deletes your modified configurations:

```
zypper rm "hadoop*" "webhcat*" "hcatalog*" "oozie*" "collectd*"
"gccxml*" "pig*" "hdfs*" "sqoop*" "zookeeper*" "hbase*"
"hive*" "flume*" "mahout*" "hue" "hue-common" "hue-shell"
"hdp_mon_nagios_addons"
```

3. Remove your old hdp.repo file:

   ```
   rm /etc/zypp/repos.d/hdp.repo
   ```

4. Install the HDP 2.2 repo:

   - Download the hdp.repo file:

     ```
     wget -nv http://public-repo-1.hortonworks.com/HDP/
     sles11sp1/2.x/updates/2.2.6.0/hdp.repo -O /etc/zypp/repos.d/
     hdp.repo
     ```

5. Install the HDP 2.2 versions of the components that you want to upgrade. For example, if you installed and want to upgrade all HDP 1.3 components:

   ```
   zypper install "hadoop" "webhcat" "oozie" "collectd" "gccxml"
   "pig" "hdfs" "sqoop"

   "zookeeper" "hbase" "hive" "flume" "mahout" "hue"
   "hdp_mon_nagios_addons"

   zypper install webhcat-tar-hive webhcat-tar-pig

   zypper up -r HDP-2.2.6.0

   zypper install oozie-client
   ```

   

   Note

   If you installed Apache Argus, it is now Apache Ranger. See Upgrade Ranger for more information on the upgrade path

6. Verify that the components were upgraded. For example, to verify hdfs, hive, and hcatlog:

   ```
   rpm -qa | grep hdfs, && rpm -qa | grep hive && rpm -qa | grep
   hcatalog
   ```

   No component files names should appear in the returned list.

3.3. Symlink Directories with hdp-select

Warning

HDP 2.2 installs hdp-select automatically with the installation or upgrade of the first HDP component. If you have not already upgraded ZooKeeper, hdp-select has not been installed.

To prevent version-specific directory issues for your scripts and updates, Hortonworks provides hdp-select, a script that symlinks directories to hdp-current and modifies paths for configuration directories.

- Before you run hdp-select, you must remove one link:

  ```
  rm /usr/bin/oozie
  ```

- Run hdp-select set all on your NameNode and all your DataNodes:

  ```
  hdp-select set all 2.2.6.0-<$version>
  ```

  For example:

  ```
  /usr/bin/hdp-select set all 2.2.6.0-2
  ```

3.4. Configure and Start Apache ZooKeeper

Tip

If you are running a highly available cluster, upgrade Apache ZooKeeper **before** you upgrade HDFS.This best practice lets the upgraded ZKFC work with your primary NameNode and your Standby NameNode.

RHEL/CentOS/Oracle Linux

1. Run the following commands on all the ZooKeeper nodes:

   ```
   yum erase zookeeper
   ```

   ```
   yum install zookeeper
   ```

2. Replace your configuration after upgrading. Replace the ZooKeeper template configuration in /etc/zookeeper/conf.

3. Start ZooKeeper.

   On all the ZooKeeper server host machines, run the following command to start ZooKeeper and the ZKFC:

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

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

SLES

1. Run the following command on all the ZooKeeper nodes:

   ```
   zypper rm zookeeper
   ```

   ```
   zypper install zookeeper
   ```

2. Replace your configuration after upgrading. Replace the ZooKeeper template configuration in /etc/zookeeper/conf.

3. Start ZooKeeper.

   On all the ZooKeeper server host machines, run the following command to start ZooKeeper and the ZKFC:

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

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

Ubuntu/Debian

1. Run the following command on all the ZooKeeper nodes:

   ```
   apt-get remove zookeeper --purge

   apt-get install zookeeper
   ```

2. Replace your configuration after upgrading. Replace the ZooKeeper template configuration in /etc/zookeeper/conf.

3. Start ZooKeeper.

   On all the ZooKeeper server host machines, run the following command to start ZooKeeper and the ZKFC:

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

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

3.5. Configure and Start Hadoop

RHEL/CentOS/Oracle Linux

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/CPU settings in yarn-site.xml and mapred-site.xml

2. Paths have changed in HDP 2.2. make sure you remove lines from hadoop-env.sh such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load as this is not where lzo is installed.

SLES

1. Use the HDP Utility script to calculate memory configuration settings. You must update the memory/CPU settings in yarn-site.xml and mapred-site.xml

2. Paths have changed in HDP 2.2. make sure you remove lines from hadoop-env.sh such as:

   ```
   export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-amd64-64
   ```

   If you leave these paths in your hadoop-env.sh file, the lzo compression code will not load as this is not where lzo is installed.

3.6. Migrate the HDP Configurations

Configurations and configuration file names have changed between HDP 1.3.2 (Hadoop 1.2.x) and HDP 2.1 (Hadoop 2.4). To upgrade to HDP 2.x, back up your current configuration files, download the new HDP 2.1 files, and compare. The following tables provide mapping information to make the comparison between releases easier.

To migrate the HDP Configurations

1. Back up the following HDP 1.x configurations on all nodes in your clusters.

   - /etc/hadoop/conf
   - /etc/hbase/conf
   - /etc/hcatalog/conf (Note: With HDP 2.1, /etc/hcatalog/conf is divided into /etc/hive-hcatalog/conf and /etc/hive-webhcat.You cannot use /etc/ hcatalog/conf in HDP 2.1.)
   - /etc/hive/conf
   - /etc/pig/conf
   - /etc/sqoop/conf
   - /etc/flume/conf
   - /etc/mahout/conf
   - /etc/oozie/conf

2. Edit /etc/hadoop/conf/core-site.xml and set hadoop.rpc.protection from none to authentication.

Note

Hadoop lets cluster administrators control the quality of protection in the configuration parameter “hadoop.rpc.protection” in core-site.xml. It is an optional parameter in HDP 2.2. If not present, the default QOP setting of “auth” is used, which implies “authentication only”.

Valid values for this parameter are: “authentication” : Corresponds to “auth” “integrity” : Corresponds to “auth-int” “privacy” : Corresponds to “auth-conf”

The default setting is authentication-only because integrity checks and encryption are a performance cost.

3. Copy your /etc/hcatalog/conf configurations to /etc/hive-hcatalog/conf and /etc/hive-webhcat as appropriate.
4. Copy log4j.properties from the hadoop config directory of the companion files to /etc/hadoop/conf. The file should have owners and permissions similar to other files in /etc/hadoop/conf.
5. Download the your HDP 2.x companion files (see "Download the Companion Files" in Chapter 1 of the Manual Install Guide) and migrate your HDP 1.x configuration.
6. Copy these configurations to all nodes in your clusters.
   - /etc/hadoop/conf
   - /etc/hbase/conf
   - /etc/hcatalog/conf
   - /etc/hive/conf
   - /etc/pig/conf
   - /etc/sqoop/conf
   - /etc/flume/conf
   - /etc/mahout/conf
   - /etc/oozie/conf
   - /etc/zookeeper/conf

Note

Upgrading the repo using `yum` or `zypper` resets all configurations. Prepare to replace these configuration directories each time you perform a yum or zypper rmgrade.

7. Review the following HDP 1.3.2 Hadoop Core configurations and the new configurations or locations in HDP 2.x.

Table 3.4. HDP 1.3.2 Hadoop Core Site (core-site.xml)

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
`fs.default.name`	core-site.xml	`fs.defaultFS`	core-site.xml
`fs.checkpoint.dir`	core-site.xml	`dfs.namenode.checkpoint.dir`	hdfs-site.xml

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
fs.checkpoint.edits.dir	core-site.xml	dfs.namenode.checkpoint.edits.dir	hdfs-site.xml
fs.checkpoint.period	core-site.xml	dfs.namenode.checkpoint.period	hdfs-site.xml
io.bytes.per.checksum	core-site.xml	dfs.bytes-per-checksum	hdfs-site.xml
dfs.df.interval	hdfs-site	fs.df.interval	core-site.xml
hadoop.native.lib	core-site.xml	io.native.lib.available	core-site.xml
hadoop.configured.node.mapping	–	net.topology.configured.node.mapping	core-site.xml
topology.node.switch.mapping.impl	core-site-xml	net.topology.node.switch.mapping.impl	core-site.xml
topology-script.file.name	core-site.xml	net.topology.script.file.name	core-site.xml
topology.script.number.args	core-site.xml	net.topology.script.number.args	core-site.xml

Note

The hadoop.rpc.protection configuration property in core- site.xml needs to specify authentication, integrity and/or privacy. No value defaults to authentication, but an invalid value such as "none" causes an error.

8. Review the following 1.3.2 HDFS site configurations and their new configurations and files in HDP 2.x.

Table 3.5. HDP 1.3.2 Hadoop Core Site (hdfs-site.xml)

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
dfs.block.size	hdfs-site.xml	dfs.blocksize	hdfs-site.xml
dfs.write.packet.size	hdfs-site.xml	dfs.client-write-packet-size	hdfs-site.xml
dfs.https.client.keystore.resource	hdfs-site.xml	dfs.client.https.keystore.resource	hdfs-site.xml
dfs.https.need.client.auth	hdfs-site.xml	dfs.client.https.need-auth	hdfs-site.xml
dfs.read.prefetch.size	hdfs-site.xml	dfs.bytes-per-checksum	hdfs-site.xml
dfs.socket.timeout	hdfs-site.xml	dfs.client.socket-timeout	hdfs-site.xml
dfs.balance.bandwidthPerSec	hdfs-site.xml	dfs.datanode.balance.bandwidthPerSec	hdfs-site.xml
dfs.data.dir	hdfs-site.xml	dfs.datanode.data.dir	hdfs-site.xml
dfs.datanode.max.xcievers	hdfs-site.xml	dfs.datanode.max.transfer.threads	hdfs-site.xml
session.id	hdfs-site.xml	dfs.metrics.session-id	hdfs-site.xml
dfs.access.time.precision	hdfs-site.xml	dfs.namenode.accesstime.precision	hdfs-site.xml

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
dfs.backup.address	hdfs-site.xml	dfs.namenode.backup.address	hdfs-site.xml
dfs.backup.http.address	hdfs-site.xml	dfs.namenode.backup.http-address	hdfs-site.xml
fs.checkpoint.dir	hdfs-site.xml	dfs.namenode.checkpoint.dir	hdfs-site.xml
fs.checkpoint.edits.dir	hdfs-site.xml	dfs.namenode.checkpoint.edits.dir	hdfs-site.xml
fs.checkpoint.period	hdfs-site.xml	dfs.namenode.checkpoint.period	hdfs-site.xml
dfs.name.edits.dir	hdfs-site.xml	dfs.namenode.edits.dir	hdfs-site.xml
heartbeat.recheck.interval	hdfs-site.xml	dfs.namenode.heartbeat.recheck-interval	hdfs-site.xml
dfs.http.address	hdfs-site.xml	dfs.namenode.http-address	hdfs-site.xml
dfs.https.address	hdfs-site.xml	dfs.namenode.https-address	hdfs-site.xml
dfs.max.objects	hdfs-site.xml	dfs.namenode.max.objects	hdfs-site.xml
dfs.name.dir	hdfs-site.xml	dfs.namenode.name.dir	hdfs-site.xml
dfs.name.dir.restore	hdfs-site.xml	dfs.namenode.name.dir.restore	hdfs-site.xml
dfs.replication.considerLoad	hdfs-site.xml	dfs.namenode.replication.considerLoad	hdfs-site.xml
dfs.replication.interval	hdfs-site.xml	dfs.namenode.replication.interval	hdfs-site.xml
dfs.max-repl-streams	hdfs-site.xml	dfs.namenode.replication. max-streams	hdfs-site.xml
dfs.replication.min	hdfs-site.xml	dfs.namenode.replication. min	hdfs-site.xml
dfs.replication.pending.timeout.sec	hdfs-site.xml	dfs.namenode.replication.pending.timeout-sec	hdfs-site.xml
dfs.safemode.extension	hdfs-site.xml	dfs.namenode.safemode. extension	hdfs-site.xml
dfs.safemode.threshold.pct	hdfs-site.xml	dfs.namenode.secondary.threshold-pct	
dfs.secondary.http.address	hdfs-site.xml	dfs.namenode.secondary.http-address	hdfs-site.xml
dfs.permissions	hdfs-site.xml	dfs.permissions.enabled	hdfs-site.xml
dfs.permissions.supergroup	hdfs-site.xml	dfs.permissions.superusergroup	hdfs-site.xml
dfs.df.interval	hdfs-site.xml	fs.df.interval	core-site.xml
dfs.umaskmode	hdfs-site.xml	fs.permissions.umask-mode	hdfs-site.xml

9. Review the following HDP 1.3.2 MapReduce Configs and their new HDP 2.x mappings.

Table 3.6. HDP 1.3.2 Configs now in Capacity Scheduler for HDP 2.x (mapred-site.xml)

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
mapred.map.child.java.opts	mapred-site.xml	mapreduce.map.java.opts	mapred-site.xml
mapred.job.map.memory.mb	mapred-site.xml	mapred.job.map.memory.mb	mapred-site.xml
mapred.reduce.child.java.opts	mapred-site.xml	mapreduce.reduce.java.opts	mapred-site.xml
mapreduce.job.reduce.memory.mb	mapred-site.xml	mapreduce.reduce.memory.mb	mapred-site.xml
security.task.umbilical.protocol.acl	mapred-site.xml	security.job.task.protocol.acl	mapred-site.xml

10. Review the following HDP 1.3.2 Configs and their new HDP 2.x Capacity Scheduler mappings.

Table 3.7. HDP 1.3.2 Configs now in capacity scheduler for HDP 2.x (capacity-scheduler.xml)

HDP 1.3.2 config	HDP 1.3.2 config file	HDP 2.2 config	HDP 2.2 config file
mapred.queue.names	mapred-site.xml	yarn.scheduler.capacity.root.queues	capacity-scheduler.xml
mapred.queue.default.acl-submit.job	mapred-queue-acls.xml	yarn.scheduler.capacity.root.default.acl_submit_jobs	capacity-scheduler.xml
mapred.queue.default.acl.administer-jobs	mapred-queue-acls.xml	yarn.scheduler.capacity.root.default acl_administer_jobs	capacity-scheduler.xml
mapred.capacity-scheduler.queue.default.capacity	capacity-scheduler.xml	yarn-scheduler.capacity.root.default.capacity	capacity-scheduler.xml
mapred.capacity-scheduler.queue.default.user-limit-factor	capacity-scheduler.xml	yarn.scheduler.capacity.root.default user-limit-factor	capacity-scheduler.xml
mapred.capacity-scheduler.queue.default.maximum-capacity	capacity-scheduler.xml	yarn.scheduler.capacity.root.default maximum-capacity	capacity-scheduler.xml
mapred.queue.default.state	capacity-scheduler.xml	yarn.scheduler.capacity.root.default.state	capacity-scheduler.xml

11. Compare the following HDP 1.3.2 configs in hadoop-env.sh with the new configs in HDP 2.x.

    Paths have changed in HDP 2.2 to /usr/hdp/current. You must remove lines such as:

```
export JAVA_LIBRARY_PATH=/usr/lib/hadoop/lib/native/Linux-
amd64-64
```

Table 3.8. HDP 1.3.2 Configs and HDP 2.x for hadoop-env.sh

HDP 1.3.2 config	HDP 2.2 config	Description
JAVA_HOME	JAVA_HOME	Java implementation to use
HADOOP_HOME_WARN_SUPPRESS	HADOOP_HOME_WARN_SUPPRESS	–
HADOOP_CONF_DIR	HADOOP_CONF_DIR	Hadoop configuration directory
not in hadoop-env.sh.	HADOOP_HOME	–
not in hadoop-env.sh.	HADOOP_LIBEXEC_DIR	–
HADOOP_NAMENODE_INIT_ HEAPSIZE	HADOOP_NAMENODE_INIT_ HEAPSIZE	–
HADOOP_OPTS	HADOOP_OPTS	Extra Java runtime options; empty by default
HADOOP_NAMENODE_OPTS	HADOOP_NAMENODE_OPTS	Command-specific options appended to HADOOP-OPTS
HADOOP_JOBTRACKER_OPTS	not in hadoop-env.sh.	Command-specific options appended to HADOOP-OPTS
HADOOP_TASKTRACKER_OPTS	not in hadoop-env.sh.	Command-specific options appended to HADOOP-OPTS
HADOOP_DATANODE_OPTS	HADOOP_DATANODE_OPTS	Command-specific options appended to HADOOP-OPTS
HADOOP_BALANCER_OPTS	HADOOP_BALANCER_OPTS	Command-specific options appended to HADOOP-OPTS
HADOOP_SECONDARYNAMENODE_ OPTS	HADOOP_SECONDARYNAMENODE_ OPTS	Command-specific options appended to HADOOP-OPTS
HADOOP_CLIENT_OPTS	HADOOP_CLIENT_OPTS	Applies to multiple commands (fs, dfs, fsck, distcp, etc.)
HADOOP_SECURE_DN_USER	not in hadoop-env.sh.	Secure datanodes, user to run the datanode as
HADOOP_SSH_OPTS	HADOOP_SSH_OPTS	Extra ssh options.
HADOOP_LOG_DIR	HADOOP_LOG_DIR	Directory where log files are stored in the secure data environment.
HADOOP_SECURE_DN_LOG_DIR	HADOOP_SECURE_DN_LOG_DIR	Directory where pid files are stored; / tmp by default.
HADOOP_PID_DIR	HADOOP_PID_DIR	Directory where pid files are stored, / tmp by default.
HADOOP_SECURE_DN_PID_DIR	HADOOP_SECURE_DN_PID_DIR	Directory where pid files are stored, / tmp by default.
HADOOP_IDENT_STRING	HADOOP_IDENT_STRING	String representing this instance of hadoop. $USER by default
not in hadoop-env.sh.	HADOOP_MAPRED_LOG_DIR	–
not in hadoop-env.sh.	HADOOP_MAPRED_PID_DIR	–
not in hadoop-env.sh.	JAVA_LIBRARY_PATH	–
not in hadoop-env.sh.	JSVC_HOME	For starting the datanode on a secure cluster

Note

Some of the configuration settings refer to the variable HADOOP_HOME. The value of HADOOP_HOME is automatically inferred from the location of the startup scripts. HADOOP_HOME is the parent directory of the bin directory that holds the Hadoop scripts. In many instances this is $HADOOP_INSTALL/hadoop.

12. Add the following properties to the yarn-site.xml file:

```
<property>
 <name>yarn.resourcemanager.scheduler.class</name>
 <value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.capacity.
 CapacityScheduler</value>
</property>

<property>
 <name>yarn.resourcemanager.resource-tracker.address</name>
 <value>$resourcemanager.full.hostname:8025</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.scheduler.address</name>
 <value>$resourcemanager.full.hostname:8030</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.address</name>
 <value>$resourcemanager.full.hostname:8050</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.admin.address</name>
 <value>$resourcemanager.full.hostname:8141</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.nodemanager.local-dirs</name>
 <value>/grid/hadoop/yarn/local,/grid1/hadoop/yarn/local</value>
 <description>Comma-separated list of paths. Use the list of directories
 from $YARN_LOCAL_DIR.For example, /grid/hadoop/yarn/local,/grid1/hadoop/
yarn/local.</description>
</property>

<property>
 <name>yarn.nodemanager.log-dirs</name>
 <value>/grid/hadoop/yarn/log</value>
 <description>Use the list of directories from $YARN_LOCAL_LOG_DIR.For
 example, /grid/hadoop/yarn/log,/grid1/hadoop/yarn/log,/grid2/hadoop/yarn/
log</description>
</property>

<property>
 <name>yarn.log.server.url</name>
```

```
 <value>http://$jobhistoryserver.full.hostname:19888/jobhistory/logs/</
 value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.resourcemanager.webapp.address</name>
 <value>$resourcemanager.full.hostname:8088</value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.nodemanager.admin-env</name>
 <value>MALLOC_ARENA_MAX=$MALLOC_ARENA_MAX</value>
 <description>Restrict the number of memory arenas to prevent
    excessive VMEM use by the glib arena allocator.
    For example, MALLOC_ARENA_MAX=4</description>
</property>
```

13. Add the following properties to the yarn-site.xml file:

```
<property>
 <name>yarn.resourcemanager.scheduler.class</name>
 <value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.capacity.
 CapacityScheduler</value>
</property>

<property>
 <name>yarn.resourcemanager.resource-tracker.address</name>
 <value>$resourcemanager.full.hostname:8025</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.scheduler.address</name>
 <value>$resourcemanager.full.hostname:8030</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.address</name>
 <value>$resourcemanager.full.hostname:8050
 </value><description>Enter your ResourceManager hostname.
 </description></property>

<property>
 <name>yarn.resourcemanager.admin.address</name>
 <value>$resourcemanager.full.hostname:8141</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.nodemanager.local-dirs</name>
 <value>/grid/hadoop/yarn/local,/grid1/hadoop/yarn/local</value>
 <description>Comma separated list of paths. Use the list of directories
  from $YARN_LOCAL_DIR. For example,
  /grid/hadoop/yarn/local,/grid1/hadoop/yarn/local.
</description>
</property>
```

```
<property>
 <name>yarn.nodemanager.log-dirs</name>
 <value>/grid/hadoop/yarn/log</value>
 <description>Use the list of directories from $YARN_LOCAL_LOG_DIR.
  For example, /grid/hadoop/yarn/log,/grid1/hadoop/yarn/log,/
  grid2/hadoop/yarn/log
</description>
</property>

<property>
 <name>yarn.log.server.url</name>
 <value>http://$jobhistoryserver.full.hostname:19888/jobhistory/logs/</
 value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.resourcemanager.webapp.address</name>
 <value>$resourcemanager.full.hostname:8088</value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.nodemanager.admin-env</name>
 <value>MALLOC_ARENA_MAX=$MALLOC_ARENA_MAX</value>
 <description>Restrict the number of memory arenas to prevent excessive VMEM
 use by
 the glib arena allocator. For example, MALLOC_ARENA_MAX=4</description>
</property>
```

14. Add the following properties to the yarn-site.xml file:

```
<property>
 <name>yarn.resourcemanager.scheduler.class</name>
 <value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.capacity.
 CapacityScheduler</value>
</property>

<property>
 <name>yarn.resourcemanager.resource-tracker.address</name>
 <value>$resourcemanager.full.hostname:8025</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.scheduler.address</name>
 <value>$resourcemanager.full.hostname:8030</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.address</name>
 <value>$resourcemanager.full.hostname:8050</value>
 <description>Enter your ResourceManager hostname.</description>
</property>

<property>
 <name>yarn.resourcemanager.admin.address</name>
 <value>$resourcemanager.full.hostname:8141</value>
 <description>Enter your ResourceManager hostname.</description>
```

```
</property>

<property>
 <name>yarn.nodemanager.local-dirs</name>
 <value>/grid/hadoop/yarn/local,/grid1/hadoop/yarn/local</value>
 <description>Comma separated list of paths. Use the list of directories
  from $YARN_LOCAL_DIR. For example,
  /grid/hadoop/yarn/local,/grid1/hadoop/yarn/local.
 </description>
</property>

<property>
 <name>yarn.nodemanager.log-dirs</name>
 <value>/grid/hadoop/yarn/log</value>
 <description>Use the list of directories from $YARN_LOCAL_LOG_DIR.
  For example, /grid/hadoop/yarn/log,
  /grid1/hadoop/yarn/log,/grid2/hadoop/yarn/log
 </description>
</property>

<property>
 <name>yarn.log.server.url</name>
 <value>http://$jobhistoryserver.full.hostname:19888/jobhistory/logs/
 </value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.resourcemanager.webapp.address</name>
 <value>$resourcemanager.full.hostname:8088</value>
 <description>URL for job history server</description>
</property>

<property>
 <name>yarn.nodemanager.admin-env</name>
 <value>MALLOC_ARENA_MAX=$MALLOC_ARENA_MAX</value>
 <description>Restrict the number of memory arenas to prevent excessive
 VMEM use by the glib arena allocator. For example,
 MALLOC_ARENA_MAX=4</description>
</property>
```

15. Adding the following properties to the mapred-site.xml file:

```
<property>
 <name>mapreduce.jobhistory.address</name>
 <value>$jobhistoryserver.full.hostname:10020</value>
 <description>Enter your JobHistoryServer hostname.</description>
</property>

<property>
 <name>mapreduce.jobhistory.webapp.address</name>
 <value>$jobhistoryserver.full.hostname:19888</value>
 <description>Enter your JobHistoryServer hostname.</description>
</property>

<property>
 <name>mapreduce.shuffle.port</name>
 <value>13562</value>
</property>
```

```
<property>
 <name>mapreduce.framework.name</name>
 <value>yarn</value>
</property>
```

16. For a secure cluster, add the following properties to mapred-site.xml:

```
<property>
 <name>mapreduce.jobhistory.principal</name>
 <value>jhs/_PRINCIPAL@$REALM.ACME.COM</value>
 <description>Kerberos principal name for the MapReduce JobHistory Server.
 </description>
</property>

</property>
 <name>mapreduce.jobhistory.keytab</name>
 <value>/etc/security/keytabs/jhs.service.keytab</value>
 <description>Kerberos keytab file for the MapReduce JobHistory Server.</
description>
</property>
```

17. For a secure cluster, you must also update hadoop.security.auth_to_local in core- site.xml to include a rule regarding the mapreduce.jobhistory.principal value you set in the previous step:

    RULE:[2:$1@$0](PRINCIPAL@$REALM.ACME.COM)s/.*/mapred/

    where PRINCIPAL and REALM are the kerberos principal and realm you specified in mapreduce.jobhistory.principal.

18. Delete any remaining HDP1 properties in the mapred-site.xml file.

19. Replace the default memory configuration settings in yarn-site.xml and mapred-site.xml with the YARN and MapReduce memory configuration settings you calculated previously.

3.7. Create Local Directories

You must create local directories for YARN on each NodeManager host in your cluster (in HDP-2, the NodeManager replaces the TaskTracker) and set the appropriate permissions for the YARN log directories.

1. Set the permissions in the yarn.nodemanager.local-dirs directories. Run these commands on all DataNodes in your cluster.

   ```
   chown -R yarn:hadoop ${yarn.nodemanager.local-dirs}

   chmod 755 ${yarn.nodemanager.local-dirs}
   ```

   where ${yarn.nodemanager.local-dirs} is your local directory.

2. Change the permissions of the directories in yarn.nodemanager.log-dirs. If these directories do not exist, you can create them using the instructions in the Create Directories section of the HDP Manual Install Guide. Run these commands on all DataNodes in your cluster.

```
chown -R yarn:hadoop ${yarn.nodemanager.log-dirs}

chmod 755 ${yarn.nodemanager.log-dirs}
```

where ${yarn.nodemanager.log-dirs} is your log directory.

3. Create directories for YARN_LOG_DIR and YARN_PID_DIR.

   - Open /etc/hadoop/conf/yarn-env.sh
   - Write down your values for YARN_LOG_DIR and YARN_PID_DIR, as the following instructions require values for the ${YARN_LOG_DIR} and ${YARN_PID_DIR}.

   For example, in yarn-env.sh:

   ```
   YARN_LOG_DIR=/grid/0/var/log/hadoop/yarn

   YARN_PID_DIR=/grid/0/var/run/hadoop/yarn
   ```

4. Make directories for ${YARN_LOG_DIR} and ${YARN_PID_DIR} and set the appropriate permissions for them.

   ```
   mkdir ${YARN_LOG_DIR}

   chown yarn:hadoop ${YARN_LOG_DIR}

   chmod 755 ${YARN_LOG_DIR}

   mkdir ${YARN_PID_DIR}

   chown yarn:hadoop ${YARN_PID_DIR}

   chmod 755 ${YARN_PID_DIR}
   ```

3.8. Start Hadoop Core

Warning

Before you start HDFS on an HA cluster you must start the ZooKeeper service. If you do not start the ZKFC, there can be failures.

Start HDFS, executing commands as $HDFS_USER.

1. If you are upgrading from an HA NameNode configuration, start all JournalNodes. On each JournalNode host, run the following command:

   ```
   su -l $HDFS_USER -c "/usr/hdp/current/hadoop-hdfs-
   journalnode/../hadoop/sbin/hadoop-daemon.sh start journalnode"
   ```

Important

All JournalNodes must be running when performing the upgrade, rollback, or finalization operations. If any JournalNodes are down when running any such operation, the operation fails.

2. Start the NameNode. On the active NameNode host machine, run the following command:

   ```
   su <HDFS_USER> -c "export HADOOP_LIBEXEC_DIR=/usr/hdp/current/
   hadoop-hdfs-namenode/../hadoop/sbin/hadoop-daemon.sh start
   namenode -upgrade"
   ```

   On a large system, this can take a long time to complete.

   

   Note

   Run this command with the -upgrade option only once. After you have completed this step, you can bring up the NameNode using this command without including the -upgrade option.

3. Verify that the NameNode is up and running:

   ```
   ps -ef|grep -i NameNode
   ```

4. Start the Secondary NameNode.

   On the Secondary NameNode host machine, run the following commands:

   ```
   su -l <HDFS_USER>
   ```

   ```
   export HADOOP_LIBEXEC_DIR=/usr/hdp/current/hadoop-hdfs-
   secondarynamenode/../hadoop/libexec/usr/hdp/current/hadoop-
   hdfs-secondarynamenode/../hadoop/sbin/hadoop-daemon.sh start
   secondarynamenode
   ```

5. Verify that the Secondary NameNode is up and running:

   ```
   ps -ef|grep SecondaryNameNode
   ```

   If you are working on a non-secure DataNode, use $HDFS_USER. For a secure DataNode, use root.

6. Start DataNodes.

   On all the DataNodes, run the following command:

   ```
   export HADOOP_LIBEXEC_DIR=/usr/hdp/current/hadoop-hdfs-
   datanode/../hadoop/sbin/hadoop-daemon.sh start datanode
   ```

7. Verify that the DataNode process is up and running:

   ```
   ps -ef|grep DataNode
   ```

8. Verify that Namenode can go out of safe mode.

   ```
   su <HDFS_USER>
   ```

   ```
   hdfs dfsadmin -safemode wait Safemode is OFF
   ```

   In general, it takes 5-10 minutes to get out of safemode. For thousands of nodes with millions of data blocks, getting out of safemode could take up to 45 minutes.

3.9. Verify HDFS filesystem health

Analyze if the filesystem is healthy.

1. Run the fsck command on namenode as $HDFS_USER:

   ```
   su $HDFS_USER
   ```

   ```
   hdfs fsck / -files -blocks -locations > dfs-new-fsck-1.log
   ```

2. Run hdfs namespace and report.

   a. List directories.

      ```
      su -l <HDFS_USER>
      ```

      ```
      hdfs dfs -ls -R / > dfs-new-lsr-1.log
      ```

   b. Run report command to create a list of DataNodes in the cluster.

      ```
      su -l <HDFS_USER>
      ```

      ```
      hdfs dfsadmin -report > dfs-new-report-1.log
      ```

   c. Compare the namespace report before the upgrade and after the upgrade. Verify that user files exist after upgrade.

      ```
      su $HDFS_USER
      ```

      ```
      dfs-old-fsck-1.log < -- > dfs-new-fsck-1.log dfs-old-lsr-1.log
      < -- > dfs-new-lsr-1.log
      ```

      

      Note

      You must do this comparison manually to catch all errors.

   d. From the Namenode WebUI, see if all DataNodes are up and running.

      ```
      http://<namenode>:50070
      ```

3.10. Configure YARN and MapReduce

After you install Hadoop, modify your configs.

1. Upload the MapReduce tarball to HDFS. As the HDFS user, for example hdfs:

```
hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/mapreduce/

hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hadoop/
mapreduce.tar.gz /hdp/apps/2.2.6.0-<$version>/mapreduce/

hdfs dfs -chown -R hdfs:hadoop /hdp

hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/mapreduce

hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/mapreduce/
mapreduce.tar.gz
```

2. Make the following changes to mapred-site.xml:

   - Add the following property:

```
 <property>
 <name>mapreduce.admin.map.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
 <final>true</final>
</property>
```

Note

You do not need to modify ${hdp.version}.

   - Modify the following existing properties to include ${hdp.version}:

```
<property>
 <name>mapreduce.admin.user.env</name>
 <value>LD_LIBRARY_PATH=/usr/hdp/${hdp.version}/hadoop/lib/native:/usr/
hdp/${hdp.version}/hadoop/lib/native/Linux-amd64-64</value>
</property>

</property>
 <name>mapreduce.admin.map.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
 <final>true</final>
</property>

</property>
 <name>mapreduce.admin.reduce.child.java.opts</name>
 <value>-server -Djava.net.preferIPv4Stack=true -Dhdp.version=${hdp.
version}</value>
 <final>true</final>
</property>

</property>
 <name>mapreduce.application.framework.path</name>
 <value>/hdp/apps/${hdp.version}/mapreduce/mapreduce.tar.gz#mr-framework</
value>
</property>

</property>
 <name>mapreduce.application.classpath</name>
```

```
 <value>$PWD/mr-framework/hadoop/share/hadoop/mapreduce/*:
   $PWD/mr-framework/hadoop/share/hadoop/mapreduce/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/common/*:
   $PWD/mr-framework/hadoop/share/hadoop/common/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/yarn/*:
   $PWD/mr-framework/hadoop/share/hadoop/yarn/lib/*:
   $PWD/mr-framework/hadoop/share/hadoop/hdfs/*:
   $PWD/mr-framework/hadoop/share/hadoop/hdfs/lib/*:
  /usr/hdp/${hdp.version}/hadoop/lib/hadoop-lzo-0.6.0.${hdp.version}.jar:
  /etc/hadoop/conf/secure</value>
</property>
```

Note

You do not need to modify ${hdp.version}.

3. Add the following property to yarn-site.xml:

```
<property>
 <name>yarn.application.classpath</name>
 <value>$HADOOP_CONF_DIR,/usr/hdp/${hdp.version}/hadoop-client/*,
   /usr/hdp/${hdp.version}/hadoop-client/lib/*,
   /usr/hdp/${hdp.version}/hadoop-hdfs-client/*,
   /usr/hdp/${hdp.version}/hadoop-hdfs-client/lib/*,
   /usr/hdp/${hdp.version}/hadoop-yarn-client/*,
   /usr/hdp/${hdp.version}/hadoop-yarn-client/lib/*</value>
</property>
```

4. For secure clusters, you must create and configure the container-executor.cfg configuration file:

   - Create the container-executor.cfg file in /etc/hadoop/conf/.
   - Insert the following properties:

     ```
     yarn.nodemanager.linux-container-executor.group=hadoop
     banned.users=hdfs,yarn,mapred
     min.user.id=1000
     ```

     - yarn.nodemanager.linux-container-executor.group - Configured value of yarn.nodemanager.linux-container-executor.group. This must match the value of yarn.nodemanager.linux-container-executor.group in yarn-site.xml.
     - banned.users - Comma-separated list of users who can not run container-executor.
     - min.user.id - Minimum value of user id. This prevents system users from running container-executor.
     - allowed.system.users - Comma-separated list of allowed system users.

   - Set the file /etc/hadoop/conf/container-executor.cfg file permissions to only be readable by root:

     ```
     chown root:hadoop /etc/hadoop/conf/container-executor.cfg

     chmod 400 /etc/hadoop/conf/container-executor.cfg
     ```

- Set the container-executor program so that only root or hadoop group users can run it:

  ```
  chown root:hadoop /usr/hdp/${hdp.version}/hadoop-yarn-server-
  nodemanager/bin /container-executor

  chmod 6050 /usr/hdp/${hdp.version}/hadoop-yarn-server-
  nodemanager/bin /container-executor
  ```

3.11. Start YARN/MapReduce Services

Add the following properties to yarn-site.xml to configure Application Timeline Server (ATS):

```
yarn.timeline-service.leveldb-timeline-store.path=/var/log/hadoop-yarn/
timeline
yarn.timeline-service.leveldb-timeline-store.ttl-interval-ms=300000

** If you are upgrading to HDP 2.1.3 or higher, use the following setting:
 yarn.timeline-service.store-class=org.apache.hadoop.yarn.server.timeline.
LeveldbTimelineStore**

** If you are upgrading to HDP 2.1.2, use the following setting:
 yarn.timeline-service.store-class=org.apache.hadoop.yarn.server.
applicationhistoryservice.timeline.LeveldbTimelineStore **

yarn.timeline-service.ttl-enable=true
yarn.timeline-service.ttl-ms=2678400000
yarn.timeline-service.generic-application-history.store-class=org.apache.
hadoop.yarn.server.applicationhistoryservice.NullApplicationHistoryStore
yarn.timeline-service.webapp.address={PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:8188
yarn.timeline-service.webapp.https.address=
{PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:8190
yarn.timeline-service.address={PUT_THE_FQDN_OF_ATS_HOST_NAME_HERE}:10200

HIVE (hive-site.xml)
hive.execution.engine=mr
hive.exec.failure.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.exec.post.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.exec.pre.hooks=org.apache.hadoop.hive.ql.hooks.ATSHook
hive.tez.container.size={map-container-size}

*If mapreduce.map.memory.mb > 2GB then set it equal to mapreduce.map.memory.
 Otherwise, set it equal to mapreduce.reduce.memory.mb*

hive.tez.java.opts=-server -Xmx800m -Djava.net.preferIPv4Stack=true -
XX:NewRatio=8 -XX:+UseNUMA -XX:+UseParallelGC
```

Use configuration values appropriate for your environment. For example, the value "800" in the preceding example is an example, not a requirement.

To start YARN, run commands as a YARN user. To start MapReduce, run commands as a MapReduce user.

1. If you have a secure cluster, create the following principals and keytabs for YARN before you start the YARN service:

2. Start the ResourceManager on your previous JobTracker host.

   ```
   su -l yarn -c "/usr/hdp/current/hadoop-yarn-resourcemanager/
   sbin/yarn-daemon.sh start resourcemanager"

   ps -ef | grep -i resourcemanager
   ```

3. Prepare the NodeManager hosts.

   a. Change permissions for /usr/hdp/current/hadoop-yarn/bin/container- executor.cfg:

      ```
      chown yarn:hadoop /usr/hdp/current/hadoop-yarn/bin/container-
      executor

      chmod -R 650 /usr/hdp/current/hadoop-yarn/bin/container-
      executor
      ```

   b. On all NodeManager hosts, add the yarn user to the hadoop group.

      For example, if you are using CentOS6:

      ```
      usermod -a -G hadoop yarn
      ```

4. Start the NodeManager on your previous TaskTracker hosts.

   ```
   su - l yarn -c "/usr/hdp/current/hadoop-yarn-nodemanager/sbin/
   yarn-daemon.sh start nodemanager"

   ps -ef | grep -i nodemanager
   ```

5. To start MapReduce, run the following commands:

   ```
   su -l yarn -c "/usr/hdp/current/hadoop-mapreduce-historyserver/
   sbin/mr-jobhistory-daemon.sh start historyserver"

   ps -ef | grep -i jobhistoryserver
   ```

3.12. Run Hadoop Smoke Tests

To smoke test your Hadoop upgrade, you can run the following MapReduce job as a regular user. The job uses MapReduce to write 100MB of data into HDFS with RandomWriter.

```
hadoop jar /usr/hdp/current/hadoop-mapreduce-
client/hadoop-mapreduce-examples.jar randomwriter -
Dtest.randomwrite.total_bytes=100000000 test-after-upgrade
```

You should see messages similar to:

```
map 0% reduce 0%
…map 100% reduce 100%
Job …. completed successfully
```

You have successfully submitted your first MapReduce job in HDP 2.x. The next steps will upgrade your other components to 2.x.

Basic troubleshooting:

1. To find the number of active nodes and NodeManagers, access the ResourceManager web UI:

   ```
   http://<resource manager host>:8088/cluster/nodes
   ```

   The number of active nodes should be equal to the number of nodemanagers.

2. Error messages. Access the ApplicationMaster WebUI to view the container logs.

   a. Looking at your console logs for MapReduce job, there is a line in this format:

   ```
   13/10/02 17:57:21 INFO mapreduce.Job: The url to track
   the job: http://<resource manager host>:8088/proxy/
   application_1380673658357_0007/
   ```

   b. Go to the URL and select the job link.

   c. Select the logs link under ApplicationMaster table. It will redirect you to the container logs. Error messages display here.

3.13. Configure and Start Apache HBase

RHEL/CentOS/Oracle Linux

1. Replace your configuration after upgrading. Replace the HBase template configuration in `/etc/hbase/conf`.

2. Start services as the <HBase_User>:

   ```
   su -l hbase -c "/usr/hdp/current/hbase-master/bin/hbase-
   daemon.sh start master; sleep 25"

   su -l hbase -c "/usr/hdp/current/hbase-regionserver/bin/hbase-
   daemon.sh start regionserver"
   ```

3. Check processes.

   ```
   ps -ef | grep -i hmaster ps -ef | grep -i hregion
   ```

SLES

1. Replace your configuration after upgrading. Replace the HBase template configuration in /etc/hbase/conf.

2. Start services as the <HBase_User>:

   ```
   su -l hbase -c "/usr/hdp/current/hbase-master/bin/hbase-
   daemon.sh start master; sleep 25"

   su -l hbase -c "/usr/hdp/current/hbase-regionserver/bin/hbase-
   daemon.sh start regionserver"
   ```

3. Check processes.

```
ps -ef | grep -i hmaster ps -ef | grep -i hregion
```

Ubuntu/Debian

1. Replace your configuration after upgrading. Replace the HBase template configuration in /etc/hbase/conf.
2. Start services as the <HBase_User>:

```
su -l hbase -c "/usr/hdp/current/hbase-master/bin/hbase-
daemon.sh start master; sleep 25

su -l hbase -c "/usr/hdp/current/hbase-regionserver/bin/hbase-
daemon.sh start regionserver"
```

3. Check processes.

```
ps -ef | grep -i hmaster ps -ef | grep -i hregion
```

3.14. Configure and Start Apache Hive and Apache HCatalog

Warning

In HDP 2.1.3 (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 available in Hive 0.11.0 and Hive 0.12.0, which allowed up to 38 digits of precision and unlimited scale.

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. Stop the Hive Metastore, if you have not done so already.
2. Upgrade the Hive Metastore database schema by running the upgrade scripts included in HDP for your metastore database and then running the schematool to upgrade to Hive 14:

Database	Run as user...	This metastore upgrade script
MySQL	root	>cd /usr/hdp/current/hive/scripts/metastore/upgrade/mysql > mysql hivemysql > source upgrade-0.11.0-to-0.12.0.mysql.sql mysql > source upgrade-0.12.0-to-0.13.0.mysql.sql > source upgrade-0.12.0-to-0.14.0.mysql.sql

Database	Run as user...	This metastore upgrade script
Postgres	root	cd /usr/hdp/current/hive/scripts/metastore/upgrade/postgres psql -d hive -a -f upgrade-0.11.0-to-0.12.0.postgres.sql psql -d hive -a -f upgrade-0.12.0-to-0.13.0.postgres.sql psql -d hive -a -f upgrade-0.13.0-to-0.14.0.postgres.sql
Oracle	root	cd /usr/hdp/2.2.6.0-<version>/hive/scripts/metastore/upgrade/oracle sqlplus SQL> @upgrade-0.11.0-to-0.12.0.oracle.sql SQL> @upgrade-0.12.0-to-0.13.0.oracle.sql SQL> @upgrade-0.13.0-to-0.14.0.oracle.sql

```
/usr/hdp/current/hive-metastore/bin/schematool -upgradeSchema -
dbType <$databaseType>
```

Note

If you are using Postgres 8 and Postgres 9, reset the Hive Metastore database owner to the <HIVE_USER>:

```
sudo <POSTGRES_USER>

ALTER DATABASE <HIVE-METASTORE-DB-NAME> OWNER TO
<HIVE_USER>
```

Note

If you are using Oracle 11, you may see the following error message:

```
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
optimize.mapjoin.mapreduce does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
heapsize does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
server2.enable.impersonation does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.
semantic.analyzer.factory.impl does not exist
14/11/17 14:11:38 WARN conf.HiveConf: HiveConf of name hive.auto.
convert.sortmerge.join.noconditionaltask does not exist
Metastore connection URL: jdbc:oracle:thin:@//ip-172-31-42-1.ec2.
internal:1521/XE
Metastore Connection Driver : oracle.jdbc.driver.OracleDriver
Metastore connection User: hiveuser
Starting upgrade metastore schema from version 0.13.0 to 0.14.0
Upgrade script upgrade-0.13.0-to-0.14.0.oracle.sql
Error: ORA-00955: name is already used by an existing object
 (state=42000,code=955)
```

```
Warning in pre-upgrade script pre-0-upgrade-0.13.0-to-0.14.0.
oracle.sql: Schema script failed, errorcode 2
Completed upgrade-0.13.0-to-0.14.0.oracle.sql
schemaTool completed
```

You can safely ignore this message. The error is in the pre-upgrade script and can be ignored; the schematool succeeded.

3. Edit hive-site.xml to update the value of hive.metadata.export.location to the new, joint hive-hcatalog jar (previously hcatalog-core.jar):

```
<property>
 <name>hive.metadata.export.location</name>
 <value>export HIVE_AUX_JARS_PATH=/usr/hdp/2.2.6.0-<version>/hive-hcatalog/
share/hcatalog/hive-hcatalog-core.jar</value>
</property>
```

4. Edit hive-env.sh to point to the new hive-hcatalog.jar:

```
if [ "${HIVE_AUX_JARS_PATH}" != " " ]; then
export HIVE_AUX_JARS_PATH=/usr/hdp/current/hive-hcatalog/share/hcatalog/
hive- hcatalog-core.jar:${HIVE_AUX_JARS_PATH}
else
export HIVE_AUX_JARS_PATH=/usr/hdp/current/hive-hcatalog/share/hcatalog/
hive- hcatalog-core.jar
fi
```

5. Edit the hive-site.xml file and modify the properties based on your environment. Search for TODO in the file for the properties to replace.

   - Edit the connection properties for your Hive metastore database in hive- site.xml:

```
<property>
 <name>javax.jdo.option.ConnectionURL</name>
 <value>jdbc:mysql://TODO-HIVE-METASTORE-DB-SERVER:TODO-HIVE-METASTORE-DB-
PORT/
 TODO-HIVE-METASTORE-DB-NAME?createDatabaseIfNotExist=true</value>
 <description>Enter your Hive Metastore Connection URL,
   for example if MySQL:
   jdbc:mysql://localhost:3306/mysql?createDatabaseIfNotExist=true
 </description>
</property>

<property>
 <name>javax.jdo.option.ConnectionUserName</name>
 <value>TODO-HIVE-METASTORE-DB-USER-NAME</value>
 <description>Enter your Hive Metastore database user name.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionPassword</name>
 <value>TODO-HIVE-METASTORE-DB-PASSWORD</value>
 <description>Enter your Hive Metastore database password.</description>
</property>

<property>
 <name>javax.jdo.option.ConnectionDriverName</name>
 <value>TODO-HIVE-METASTORE-DB-CONNECTION-DRIVER-NAME</value>
 <description>Enter your Hive Metastore Connection Driver Name, for
 example if
```

```
 MySQL: com.mysql.jdbc.Driver</description>
</property>
```

- Edit the following properties in the hive-site.xml file:

```
<property>
 <name>fs.file.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.file.impl.disable.cache</
description>
</property>

<property>
 <name>fs.hdfs.impl.disable.cache</name>
 <value>false</value>
 <description>Set to false or remove fs.hdfs.impl.disable.cache</
description>
</property>
```

- **Optional:** If you want Hive Authorization, set the following Hive authorization parameters in the hive-site.xml file:

```
<property>
 <name>hive.security.authorization.enabled</name>
 <value>true</value>
</property>

<property>
 <name>hive.security.authorization.manager</name>
 <value>org.apache.hadoop.hive.ql.security.authorization.
StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.metastore.authorization.manager</name>
 <value>org.apache.hadoop.hive.ql.security.authorization.
   StorageBasedAuthorizationProvider</value>
</property>

<property>
 <name>hive.security.authenticator.manager</name>
 <value>org.apache.hadoop.hive.ql.security.ProxyUserAuthenticator</ value>
</property>
```

- For a remote Hive metastore database, use the following hive-site.xml property value to set the IP address (or fully-qualified domain name) and port of the metastore host.

  To enable HiveServer2, leave this property value empty.

```
<property>
  <name>hive.metastore.uris</name>
  <value>thrift://$metastore.server.full.hostname:9083</value>
  <description>URI for client to contact metastore server.
    To enable HiveServer2, leave the property value empty.
  </description>
</property>
```

Note

You can also use the HDP utility script to fine-tune your configuration settings based on node hardware specifications.

6. Disable autocreation of schemas to match HDP 2.1+ configurations. Edit hive-site.xml to set the value of `datanucleus.autoCreateSchema` to false.

```
<property>
 <name>datanucleus.autoCreateSchema</name>
 <value>false</value>
 <description>Creates necessary schema on a startup if one doesn't exist.
 </description>
</property>
```

7. Start Hive. On the Hive Metastore host machine, run the following commands:

```
su - hive

nohup /usr/hdp/current/hive-metastore/bin/hive --service
metastore>/var/log/hive/hive.out 2>/var/log/hive/hive.log &
```

8. Start Hive Server2.

   On the Hive Server2 host machine, run the following commands:

```
su - hive

/usr/hdp/current/hive-server2/bin/hiveserver2 >/var/log/hive/
hiveserver2.out 2> /var/log/hive/hiveserver2.log&
```

   where $HIVE_USER is the Hive Service user. For example, hive.

3.15. Configure and Start Apache Oozie

Upgrading Apache Oozie is a complex process. Although the instructions are straightforward, set aside a dedicated block of time to upgrade oozie clients and servers.

Perform the following preparation steps on each oozie server host:

1. You must restore oozie-site.xml from your backup to the conf directory on each oozie server and client.

2. Copy the JDBC jar to libext-customer.

   - Create the /usr/hdp/current/oozie/libext-customer directory.

     ```
     cd /usr/hdp/current/oozie mkdir libext-customer
     ```

   - Grant read/write/execute access to all users for the libext-customer directory.

     ```
     chmod -R 777 /usr/hdp/current/oozie/libext-customer
     ```

3. Copy these files to the libext-customer directory:

```
cp /usr/hdp/current/hadoop/lib/hadooplzo*.jar /usr/hdp/current/
oozie/libext-customer

cp /usr/share/HDP-oozie/ext-2.2.zip /usr/hdp/current/oozie/
libext-customer/
```

4. Extract share-lib.

   ```
   cd /usr/hdp/current/oozie

   tar xzvf /usr/hdp/current/oozie/oozie-sharelib.tar.gz

   su -l hdfs -c "hdfs dfs -mkdir -p /user/oozie"su -l hdfs -c
   "hdfs dfs -copyFromLocal

   /usr/hdp/current/oozie/share /user/oozie/."
   ```

   You can expect to see messages stating that some files already exist. Delete any existing /oozie/share and replace it with the newly-extracted files.

   ```
   su -l hdfs -c "hdfs dfs -chown oozie:hadoop /user/oozie"

   su -l hdfs -c "hdfs dfs -chmod -R 755 /user/oozie"
   ```

5. if a previous version of Oozie was created using auto schema creation, run the following SQL query:

   ```
   insert into oozie_sys (name, data) values ('db.version', '2.5');
   ```

6. As the Oozie user (not root), run the upgrade.

   ```
   su $OOZIE_USER

   /usr/hdp/current/oozie-server/bin/ooziedb.sh upgrade -run
   ```

7. As root, prepare the Oozie WAR file.

   ```
   sudo su -l oozie -c "/usr/hdp/current/oozie/bin/oozie-setup.sh
   prepare-war -d /usr/hdp/current/oozie/libext-customer"
   ```

   Look for console output to indicate success. For example, if you are using MySQL you should see something similar to:

   ```
   INFO: Adding extension: libext-customer/mysql-connector-java.jar
   New Oozie WAR file with added 'JARs' at /var/lib/oozie/oozie-server/webapps/
   oozie.war
   ```

8. Replace the content of /user/oozie/share in HDFS. On the Oozie server host:

   ```
   sudo su -l oozie -c "/usr/hdp/current/oozie/bin/oozie-setup.sh
   prepare-war -d /usr/hdp/current/oozie/libext-customer"
   ```

9. Add the following property to oozie-log4j.properties:

```
log4j.appender.oozie.layout.ConversionPattern=%d{ISO8601} %5p
%c{1}:%L - SERVER[${oozie.instance.id}] %m%n
```

where ${oozie.instance.id} is determined by oozie, automatically.

10. As the oozie user, start Oozie:

```
sudo su -l <OOZIE_User> -c "cd /grid/0/var/log/oozie; /usr/hdp/
current/oozie/oozie-server/bin/catalina.sh /usr/hdp/current/
oozie/oozie-server/bin/setclasspath.sh /usr/hdp/current/oozie-
server/bin/oozied.sh start"
```

11. Check processes.

```
ps -ef | grep -i oozie
```

3.16. Configure and Start Apache WebHCat (Templeton)

RHEL/CentOS/Oracle Linux

1. Copy the appropriate configurations from /etc/hcatalog/conf to /etc/hive- webhcat/conf/.

2. Copy the new Pig, Hive and Hadoop-streaming jars to HDFS using the path you specified in ./etc/hive-webhcat/conf/ and change ownership to the hcat user with 755 permissions. For example:

```
hdfs dfs -copyFromLocal /usr/share/HDP-webhcat/hive.tar.gz /usr/
share/HDP-webhcat/pig.tar.gz/usr/hdp/version/hadoop-mapreduce/
hadoop-streaming.jar hdfs:///apps/webhcat/.

hdfs dfs -chmod -R 755 hdfs:///apps/webhcat/*

hdfs dfs -chown -R hcat hdfs:///apps/webhcat/*
```

3. Replace your WebHCat configuration after upgrading. Copy your modified/etc/webhcat/conf from the template to the configuration directory in all your WebHCat hosts.

4. Start WebHCat:

```
sudo su -l $WEBHCAT_USER -c "//hive-hcatalog/sbin/
webhcat_server.sh start"
```

5. Smoke test WebHCat.

   On the WebHCat host machine, run the following command:

```
http://$WEBHCAT_HOST_MACHINE:50111/templeton/v1/status
```

   If you are using a secure cluster, run the following command:

```
curl --negotiate -u:http://cluster.$PRINCIPAL.$REALM:50111/
templeton/v1/ status{"status":"ok","version":"v1"}
[machine@acme]$
```

6. Remove shared libraries from old Templeton installation.

   On the WebHCat host machine, run the following command:

   ```
   sudo su -l $HDFS_USER -c "hdfs dfs -rmr -skipTrash /apps/
   templeton" rm -rf /usr/share/HDP-templeton
   ```

   where

   - *$WEBHCAT_USER*is the WebHCat Service user. For example, hcat.
   - *$HDFS_USER*is the HDFS Service user. For example, hdfs.

SLES

1. Copy the appropriate configurations from /etc/hcatalog/conf to /etc/hive-webhcat/conf/.
2. Copy the new Pig, Hive and Hadoop-streaming jars to HDFS using the path you specified in ./etc/hive-webhcat/conf/ and change ownership to the hcat user with 755 permissions. For example:

   ```
   hdfs dfs -copyFromLocal /usr/share/HDP-webhcat/hive.tar.gz /usr/
   share/HDP-webhcat/pig.tar.gz/usr/hdp/version/hadoop-mapreduce/
   hadoop-streaming.jar hdfs:///apps/webhcat/.

   hdfs dfs -chmod -R 755 hdfs:///apps/webhcat/*

   hdfs dfs -chown -R hcat hdfs:///apps/webhcat/*
   ```

3. Replace your WebHCat configuration after upgrading. Copy your modified /etc/webhcat/conf from the template to the configuration directory in all your WebHCat hosts.
4. Modify the WebHCat configuration files.

   - Upload Pig, Hive and Sqoop tarballs to HDFS as the $HDFS_User. In this example, hdfs:

   ```
   hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/pig/
   hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/hive/
   hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/sqoop/
   hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/pig/pig.tar.gz /hdp/apps/2.2.6.
   0-<$version>/pig/
   hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/hive/hive.tar.gz /hdp/apps/2.2.
   6.0-<$version>/hive/
   hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/sqoop/sqoop.tar.gz /hdp/apps/2.
   2.6.0-<$version>/sqoop/
   hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/pig
   hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/pig/pig.tar.gz
   hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/hive
   hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/hive/hive.tar.gz
   hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/sqoop
   ```

```
hdfs dfs -chmod -R 444 /hdp/apps/2.2.6.0-<$version>/sqoop/sqoop.tar.gz
hdfs dfs -chown -R hdfs:hadoop /hdp
```

- Update the following properties in the webhcat-site.xml configuration file, as their values have changed:

```
<property>
 <name>templeton.pig.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/pig/pig.tar.gz</value>
</property>

</property>
 <name>templeton.hive.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/hive/hive.tar.gz</value>
</property>

</property>
 <name>templeton.streaming.jar</name>
 <value>hdfs:///hdp/apps/${hdp.version}/mapreduce/hadoop-streaming.jar</
value>
 <description>The hdfs path to the Hadoop streaming jar file.</
description>
</property>

</property>
 <name>templeton.sqoop.archive</name>
 <value>hdfs:///hdp/apps/${hdp.version}/sqoop/sqoop.tar.gz</value>
 <description>The path to the Sqoop archive.</description>
</property>

</property>
 <name>templeton.sqoop.path</name>
 <value>sqoop.tar.gz/sqoop/bin/sqoop</value>
 <description>The path to the Sqoop executable.</description>
</property>

</property>
 <name>templeton.sqoop.home</name>
 <value>sqoop.tar.gz/sqoop</value>
 <description>The path to the Sqoop home in the exploded archive.</
description>
</property>
```

Note

You do not need to modify ${hdp.version}.

- Remove the following obsolete properties from webhcat-site.xml:

```
<property>
 <name>templeton.controller.map.mem</name>
 <value>1600</value>
 <description>Total virtual memory available to map tasks.</description>
</property>

</property>
 <name>hive.metastore.warehouse.dir</name>
 <value>/path/to/warehouse/dir</value>
</property>
```

- Add new proxy users, if needed. In core-site.xml, make sure the following properties are also set to allow WebHCat to impersonate your additional HDP 2.2 groups and hosts:

```
<property>
 <name>hadoop.proxyuser.hcat.groups</name>
 <value>*</value>
</property>

</property>
 <name>hadoop.proxyuser.hcat.hosts</name>
 <value>*</value>
</property>
```

Where:

`hadoop.proxyuser.hcat.group`

Is a comma-separated list of the Unix groups whose users may be impersonated by 'hcat'.

`hadoop.proxyuser.hcat.hosts`

A comma-separated list of the hosts which are allowed to submit requests by 'hcat'.

5. Start WebHCat:

```
su -l hcat -c "/usr/hdp/current/hive-webhcat/sbin/
webhcat_server.sh start"
```

6. Smoke test WebHCat.

On the WebHCat host machine, run the following command:

```
http://$WEBHCAT_HOST_MACHINE:50111/templeton/v1/status
```

If you are using a secure cluster, run the following command:

```
curl --negotiate -u:http://cluster.$PRINCIPAL.$REALM:50111/
templeton/v1/ status{"status":"ok","version":"v1"}
[machine@acme]$
```

7. Remove shared libraries from old Templeton installation.

On the WebHCat host machine, run the following command:

```
sudo su -l $HDFS_USER -c "hdfs dfs -rmr -skipTrash /apps/
templeton" rm -rf /usr/share/HDP-templeton
```

where

- *$WEBHCAT_USER* is the WebHCat Service user. For example, hcat.
- *$HDFS_USER* is the HDFS Service user. For example, hdfs.

Ubuntu/Debian

1. Copy the appropriate configurations from /etc/hcatalog/conf to /etc/hive- webhcat/conf/.
2. Copy the new Pig, Hive and Hadoop-streaming jars to HDFS using the path you specified in ./etc/hive-webhcat/conf/ and change ownership to the hcat user with 755 permissions. For example:

   ```
   hdfs dfs -copyFromLocal /usr/share/HDP-webhcat/hive.tar.gz /usr/
   share/HDP- webhcat/pig.tar.gz/usr/hdp/version/hadoop-mapreduce/
   hadoop-streaming.jar hdfs:///apps/webhcat/.

   hdfs dfs -chmod -R 755 hdfs:///apps/webhcat/*

   hdfs dfs -chown -R hcat hdfs:///apps/webhcat/*
   ```

3. Replace your WebHCat configuration after upgrading. Copy your modified/etc/webhcat/conf from the template to the configuration directory in all your WebHCat hosts.
4. Start WebHCat:

   ```
   sudo su -l $WEBHCAT_USER -c "/usr/lib/hive-hcatalog/sbin/
   webhcat_server.sh start"
   ```

5. Smoke test WebHCat.

   On the WebHCat host machine, run the following command:

   ```
   http://$WEBHCAT_HOST_MACHINE:50111/templeton/v1/status
   ```

   If you are using a secure cluster, run the following command:

   ```
   curl --negotiate -u:http://cluster.$PRINCIPAL.$REALM:50111/
   templeton/v1/ status{"status":"ok","version":"v1"}
   [machine@acme]$
   ```

6. Remove shared libraries from old Templeton installation.

   On the WebHCat host machine, run the following command:

   ```
   sudo su -l $HDFS_USER -c "hdfs dfs -rmr -skipTrash /apps/
   templeton" rm -rf /usr/share/HDP-templeton
   ```

   where

   - *$WEBHCAT_USER* is the WebHCat Service user. For example, hcat.
   - *$HDFS_USER* is the HDFS Service user. For example, hdfs.

3.17. Configure and Start Apache Pig

1. Replace your configuration after upgrading. Copy `/etc/pig/conf` from the template to the conf directory in pig hosts.
2. To validate the Apache Pig upgrade, complete the following steps:

a. On the host machine where Pig is installed, run the following commands:

```
su - $HDFS_USER/usr/hdp/current/hadoop/bin/hadoop
fs -copyFromLocal /etc/passwd passwd
```

b. Create a Pig script file named `/tmp/id.pig` that contains the following Pig Latin commands:

```
A = load 'passwd' using PigStorage(':');B = foreach A generate $0 as id;
 store B into '/tmp/id.out';
```

c. Run the Pig script:

```
su - $HDFS_USER
pig -l /tmp/pig.log /tmp/id.pig
```

3.18. Configure and Start Apache Sqoop

1. Replace your configuration after upgrading.

   Copy /etc/sqoop/conf from the template to the conf directory in Apache Sqoop hosts.

2. Upload the Sqoop tarball to HDFS. As the <HDFS_User>, for example 'hdfs':

   ```
   su -c "hdfs dfs -mkdir -p /hdp/apps/2.2.6.0-<$version>/sqoop"

   hdfs su -c "hdfs dfs -chmod -R 555 /hdp/apps/2.2.6.0-<$version>/
   sqoop"

   hdfs su -c "hdfs dfs -chown -R hdfs:hadoop /hdp/apps/2.2.6.0-<
   $version>/sqoop"

   hdfs su -c "hdfs dfs -put /usr/hdp/2.2.6.0-<$version>/sqoop/
   sqoop.tar.gz /hdp/apps/2.2.6.0-<$version>/sqoop/sqoop.tar.gz"

   hdfs su -c "hdfs dfs -chmod 444 /hdp/apps/2.2.6.0-<$version>/
   sqoop/sqoop.tar.gz" hdfs
   ```

3.19. Configure, Start, and Validate Apache Flume

1. If you have not already done so, upgrade Apache Flume. On the Flume host machine, run the following command:

   - For **RHEL/CentOS/Oracle Linux**:

     ```
     yum upgrade flume
     ```

   - For **SLES**:

     ```
     zypper update flume

     zypper remove flume

     zypper se -s flume
     ```

You should see Flume in the output.

Install Flume:

```
zypper install flume
```

- For **Ubuntu/Debian**:

```
apt-get install flume
```

2. To confirm that Flume is working correctly, create an example configuration file. The following snippet is a sample configuration that can be set using the properties file. For more detailed information, see the "Flume User Guide."

```
agent.sources = pstream
agent.channels = memoryChannel
agent.channels.memoryChannel.type = memory

agent.sources.pstream.channels = memoryChannel
agent.sources.pstream.type = exec
agent.sources.pstream.command = tail -f /etc/passwd

agent.sinks = hdfsSink
agent.sinks.hdfsSink.type = hdfs
agent.sinks.hdfsSink.channel = memoryChannel
agent.sinks.hdfsSink.hdfs.path = hdfs://tmp/flumetest
agent.sinks.hdfsSink.hdfs.fileType = SequenceFile
agent.sinks.hdfsSink.hdfs.writeFormat = Text
```

The source here is defined as an exec source. The agent runs a given command on startup, which streams data to stdout, where the source gets it. The channel is defined as an in-memory channel and the sink is an HDFS sink.

3. Given this configuration, you can start Flume as follows:

```
$ bin/flume-ng agent --conf ./conf --conf-file example.conf --
name a1 -Dflume.root.logger=INFO,console
```

Note

The directory specified for `--conf agrument` would include a shell script flume-env.sh and potentially a log4j properties file. In this example, we pass a Java option to force Flume to log to the console and we go without a custom environment script.

4. After validating data in `hdfs://tmp/flumetest`, stop Flume and restore any backup files. Copy `/etc/flume/conf` to the conf directory in Flume hosts.

3.20. Configure, Start, and Validate Apache Mahout

Replace your configuration after upgrading. Copy `/etc/mahout/conf` from the template to the conf directory in mahout hosts.

To validate Apache Mahout:

1. Create a test user named "testuser" in the Linux cluster and in HDFS, and log in as that user.

   ```
   hdfs dfs -put /tmp/sample-test.txt /user/testuser
   ```

2. Export the required environment variables for Mahout:

   ```
   export JAVA_HOME="your_jdk_home_install_location_here
   export HADOOP_HOME=/usr/hdp/current/hadoop-client
   export MAHOUT_HOME=/usr/hdp.current/mahout-client
   export PATH="$PATH":$HADOOP_HOME/bin:$MAHOUT_HOME/bin
   export CLASSPATH="$CLASSPATH":$MAHOUT_HOME
   ```

3. Upload a few megabytes of natural-language plain text to the Linux server as `/tmp/sample-test.txt`.

4. Transfer the sample-test.txt file to a subdirectory of the testusers's HDFS home directory.

   ```
   hdfs dfs -mkdir /user/testuser/testdata
   hdfs dfs -put /tmp/sample-test.txt /user/testuser/testdata
   ```

5. Enter the mahout command to convert the plain text file sample-test.txt into a sequence file stored in the output directory mahouttest:

   ```
   mahout seqdirectory --input /user/testuser/testdata --output /user/testuser/
   mahouttest -ow --charset utf-8
   ```

3.21. Configure and Start Hue

For HDP 2.2, use the Hue version shipped with HDP 2.2. If you have a previous version of Hue, use the following steps to upgrade Hue.

1. Migrate the hue.ini setting from your old hue.ini configuration file to new hue.ini configuration file.

2. If you are using the embedded SQLite database, remember to restore your database after upgrade.

   To restore the database from a backup, make sure the destination database is empty before copying (if necessary, rename or remove the current destination database), then copy your backup to the destination database.

   For example:

   ```
   su $HUE_USER

   cd /var/lib/hue

   mv desktop.db desktop.db.old

   sqlite3 desktop.db < ~/hue_backup/desktop.bak

   exit
   ```

3. Synchronize the database:

```
cd /usr/lib/hue

source ./build/env/bin/activate

hue syncdb

deactivate
```

4. Start Hue. As a root user, run the following command on the Hue Server:

```
/etc/init.d/hue start
```

3.22. Finalize the Upgrade

You can start HDFS without finalizing the upgrade. When you are ready to discard your backup, you can finalize the upgrade.

Warning

You must verify your filesystem health before finalizing the upgrade. After you finalize an upgrade, you cannot roll back.

To finalize the upgrade process, run the following command as the $HDFS_USER:

```
hadoop dfsadmin -finalizeUpgrade
```

3.23. Install New HDP 2.2 Services

Install new HDP 2.2 Services:

- Tez – YARN-based processing framework that greatly improves query response times for Hive and other YARN applications.
- Phoenix – SQL skin over HBase that makes it easier and faster to build HBase applications.
- Accumulo – a high-performance data storage and retrieval system with cell-level access control. It is a scalable implementation of Google's Big Table design that works on top of Hadoop and ZooKeeper.
- Storm – Real-time event stream processing platform that provides fixed, continuous, & low latency processing for very high frequency streaming data.
- Falcon – Framework for simplifying the development of data management applications in Apache Hadoop.
- Knox – Apache Knox is the Web/REST API Gateway solution for Hadoop. It provides a single access point for all of Hadoop resources over REST.
- Slider – YARN application to deploy existing distributed applications on YARN, monitor them, and make them larger or smaller as desired even while the application is running.

- Ranger – Comprehensive approach to central security policy administration across the core enterprise security requirements of authentication, authorization, accounting and data protection
- Kafka – an open-source message broker project written in Scala.
- Spark – an open-source cluster computing engine.