

How to reformat HDFS and Hive

If you are getting unexplained errors, and you have verified that Hadoop is running and that your input file is downloaded (to local directory, such as /home/ec2-user/vehicles.csv for Hive or to HDFS, such as /user/ec2-user/vehicles.csv that can be checked with `hadoop fs -ls /user/ec2-user`), you can opt for a quick solution of reformatting HDFS. Note that you do not need to create a new instance, or install anything, just reformat. Reformat process will wipe out all HDFS files.

NOTE: If you are running a micro or nano instance, then you must change instance type (stop instance, change instance type to at least small and then start it again)

Step1: Stop Hadoop (**`stop-dfs.sh: stop-yarn.sh: mr-jobhistory-daemon.sh stop historyserver`**). `jps` command will indicate whether any processes are running. If stop command did not work, you can go for the quick solution by running **`killall java`** (that kills every Java process on the system).

Step2: Delete the local storage that holds HDFS files. Please do not skip that step.

```
rm -rf /tmp/hadoop-ec2-user/dfs/
```

Step3: Reformat

`hdfs namenode -format` (it is the same as **`hadoop namenode -format`** that you used before)

(start Hadoop, multiple commands can be separated by semi-colon)

```
start-dfs.sh: start-yarn.sh: mr-jobhistory-daemon.sh start historyserver
```

(create an HDFS directory)

```
hadoop fs -mkdir -p /user/ec2-user/
```

You are done reformatting a blank HDFS storage.

To setup Hive on that new storage, you need to run the following:

(to create the hive storage folder in new hdfs)

```
$HADOOP_HOME/bin/hadoop fs -mkdir -p /user/hive/warehouse
```

```
$HADOOP_HOME/bin/hadoop fs -chmod g+w /user/hive/warehouse
```

(to create the hive storage folder in new hdfs)

```
$HADOOP_HOME/bin/hadoop fs -mkdir -p /user/hive/warehouse
```

(delete the meta-store)

```
rm -rf ~/apache-hive-2.0.1-bin/metastore_db/
```

(re-initialize Hive)

```
$HIVE_HOME/bin/schematool -initSchema -dbType derby
```

You can now go to Hive (`cd $HIVE_HOME`) or Pig (`cd $PIG_HOME`) to use it. For pig, you have to use the `put` command to place vehicles.csv file in HDFS (i.e. **`hadoop fs -ls /user/ec2-user`** shows that file in HDFS)