

Big Data Frameworks CSE3120

Lab – 3 Hadoop introduction and HDFS commands

Name: Naveen Nidadavolu

Roll No: 22MIA1049

Aim: To understand and demonstrate the usage of various HDFS commands for file and directory operations in the Hadoop Distributed File System (HDFS).

Algorithm/Procedure

1. Start the Hadoop environment and ensure HDFS services are running.
2. Execute each HDFS command as described, to perform operations like directory creation, file uploads, and retrieval.
3. Verify results using appropriate HDFS commands (`ls`, `cat`, etc.).

Preliminary Commands

1.start-all.sh

- **Command:** `start-all.sh`
- **Usage:** Starts all essential Hadoop daemons, including the HDFS (NameNode and DataNode) and YARN (ResourceManager and NodeManager) services. It ensures that the distributed environment is fully operational.

```
ponny@ubuntu:~$ start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /home/ponny/hadoop/libexec/../logs/hadoop-ponny-namenode-ubuntu.out
localhost: starting datanode, logging to /home/ponny/hadoop/libexec/../logs/hadoop-ponny-datanode-ubuntu.out
localhost: starting secondarynamenode, logging to /home/ponny/hadoop/libexec/../logs/hadoop-ponny-secondarynamenode-ubuntu.out
starting jobtracker, logging to /home/ponny/hadoop/libexec/../logs/hadoop-ponny-jobtracker-ubuntu.out
localhost: starting tasktracker, logging to /home/ponny/hadoop/libexec/../logs/hadoop-ponny-tasktracker-ubuntu.out
```

2.jps

- **Command:** `jps`
- **Usage:** Displays the list of Java processes running on the system, such as NameNode, DataNode, ResourceManager, and NodeManager. Useful for verifying whether Hadoop daemons are running.

```
ponny@ubuntu:~$ jps
3704 Jps
3415 SecondaryNameNode
3664 TaskTracker
3025 NameNode
3246 DataNode
3503 JobTracker
```

HDFS Commands and Usage Descriptions

1. version

- **Command:**

hdfs version

- **Usage:** Displays the Hadoop version currently installed and running. It helps verify the installation and compatibility of Hadoop.

```
ponny@ubuntu:~$ hadoop version
Warning: $HADOOP_HOME is deprecated.

Hadoop 1.0.4
Subversion https://svn.apache.org/repos/asf/hadoop/common/branches/branch-1.0 -r
1393290
Compiled by hortonfo on Wed Oct 3 05:13:58 UTC 2012
From source with checksum fe2baea87c4c81a2c505767f3f9b71f4
ponny@ubuntu:~$
```

2. mkdir

- **Command:**

hdfs dfs -mkdir /directory_name

- **Usage:** Creates a new directory in HDFS. This is useful for organizing files in the distributed file system.

```
ponny@ubuntu:~$ hadoop fs -mkdir /user/ponny/naveen
Warning: $HADOOP_HOME is deprecated.
```

3. touchz

- **Command:**

hdfs dfs -touchz /path_to_file

- **Usage:** Creates an empty file at the specified path in HDFS. If the file already exists, the command does nothing. Useful for creating placeholder files in HDFS.

```
ponny@ubuntu:~$ hadoop fs -touchz /user/ponny/naveen/myfile.txt
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -ls /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

Found 1 items
-rw-r--r-- 1 ponny supergroup 0 2025-01-13 11:05 /user/ponny/naveen/m
yfile.txt
ponny@ubuntu:~$
```

4. ls

- **Command:**

hdfs dfs -ls /path

- **Usage:** Lists the files and directories at the specified path in HDFS. It provides details like file permissions, owner, group, size, and modification date.

```
ponny@ubuntu:~$ hadoop fs -ls /user/ponny/
Warning: $HADOOP_HOME is deprecated.

Found 7 items
-rw-r--r--  1 ponny supergroup      44 2021-05-19 12:13 /user/ponny/lf
drwxr-xr-x  - ponny supergroup      0 2025-01-13 10:38 /user/ponny/naveen
drwxr-xr-x  - ponny supergroup      0 2021-05-19 12:14 /user/ponny/newfile
-rw-r--r--  1 ponny supergroup     44 2018-10-24 16:40 /user/ponny/rin1
-rw-r--r--  1 ponny supergroup     78 2018-10-24 16:40 /user/ponny/rin2
drwxr-xr-x  - ponny supergroup      0 2018-10-24 16:42 /user/ponny/rout
drwxr-xr-x  - ponny supergroup      0 2024-12-20 13:06 /user/ponny/thameen
ponny@ubuntu:~$ hadoop fs -touchz /user/ponny/naveen/myfile.txt
Warning: $HADOOP_HOME is deprecated.
```

5. put

- **Command:**

hdfs dfs -put localfile /path_in_hdfs

- **Usage:** Uploads a file from the local filesystem to HDFS. It is used to populate HDFS with data for processing

```
ponny@ubuntu:~$ mkdir localfile.txt
ponny@ubuntu:~$ hadoop fs -put ~/localfile.txt /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -ls /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

Found 2 items
drwxr-xr-x  - ponny supergroup      0 2025-01-13 11:08 /user/ponny/naveen/l
ocalfile.txt
-rw-r--r--  1 ponny supergroup      0 2025-01-13 11:05 /user/ponny/naveen/m
yfile.txt
ponny@ubuntu:~$
```

6. copyFromLocal

- **Command:**

hdfs dfs -copyFromLocal /local/path /hdfs/path

- **Usage:** Copies a file or directory from the local file system to HDFS. It works similarly to put but supports recursive copying for directories.

```
ponny@ubuntu:~$ hadoop fs -copyFromLocal ~/localfile1.txt /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -ls /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

Found 3 items
drwxr-xr-x - ponny supergroup      0 2025-01-13 11:08 /user/ponny/naveen/localfile.tx
t
-rw-r--r-- 1 ponny supergroup      0 2025-01-13 11:12 /user/ponny/naveen/localfile1.t
xt
-rw-r--r-- 1 ponny supergroup      0 2025-01-13 11:05 /user/ponny/naveen/myfile.txt
ponny@ubuntu:~$
```

7. get

- **Command:**

hdfs dfs -get /hdfs/path /local/path

- **Usage:** Downloads files or directories from HDFS to the local filesystem. It is useful for retrieving processed data or backups from HDFS.

```
ponny@ubuntu:~$ hadoop fs -touchz /user/ponny/naveen/hadoopfile.txt
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -get /user/ponny/naveen/hadoopfile.txt ~/copyfromHadoop
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ ls
1901                  hadoop                  pig_1712571308871.log
Check                 if2.txt                 pig_1712572265326.log
classes               if.txt                  Public
copyfromHadoop        if.txt~                 rc.jar
data_30l.txt          localfile1.txt          s1.hql
data_30l.txt~         localfile.txt           Templates
datagen_10.txt        mar                     TempStatsStore
derby.log             metastore_db            test
Desktop               Music                   thameem
Documents              pa.txt~                 Videos
Downloads              Pictures                 WhiteHouse-WAVES-Released-0827.csv
eclipse-java-juno-SR2-linux-gtk.tar.gz Pig                      workspace
examples.desktop      pig_1712569462040.log
fontconfig             pig_1712569713195.log
ponny@ubuntu:~$
```

8. copyToLocal

- **Command:**

hdfs dfs -copyToLocal /hdfs/path /local/path

- **Usage:** Copies files or directories from HDFS to the local file system, similar to get. It is often used for backups or local analysis.


```
ponny@ubuntu:~$ hadoop fs -copyToLocal /user/ponny/naveen/hadoopfile.txt ~/copyhadoopfile.txt
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ ls
1901                                fontconfig                        pig_1712569713195.log
Check                              hadoop                           pig_1712571308871.log
classes                           if2.txt                         pig_1712572265326.log
copyfromHadoop                   if.txt                          Public
copyhadoopfile.txt              if.txt~                        rc.jar
data_30l.txt                    localfile1.txt                 s1.hql
data_30l.txt~                   localfile.txt                  Templates
datagen_10.txt                  mar                             TempStatsStore
derby.log                       metastore_db                   test
Desktop                          Music                           thameem
Documents                       pa.txt~                        Videos
Downloads                       Pictures                        WhiteHouse-WAVES-Released-0827.csv
eclipse-java-juno-SR2-linux-gtk.tar.gz Pig                             workspace
examples.desktop                pig_1712569462040.log
```

9. cat

- **Command:**

hdfs dfs -cat /file_path_in_hdfs

- **Usage:** Displays the content of a file stored in HDFS. It is useful for verifying file content without downloading it.

```
ponny@ubuntu:~$ hadoop fs -cat /user/ponny/naveen/copyfile.txt
Warning: $HADOOP_HOME is deprecated.

Hello world!, This is NAVEEN exploring hadoop
ponny@ubuntu:~$
```

10. mv

- **Command:**

hdfs dfs -mv /source_path /destination_path

- **Usage:** Moves or renames files or directories within HDFS. This is helpful for organizing or restructuring data in HDFS.

```
ponny@ubuntu:~$ hadoop fs -mv /user/ponny/naveen/myfile.txt /user/ponny/
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -ls /user/ponny/
Warning: $HADOOP_HOME is deprecated.

Found 8 items
-rw-r--r-- 1 ponny supergroup 44 2021-05-19 12:13 /user/ponny/if
-rw-r--r-- 1 ponny supergroup 0 2025-01-13 11:05 /user/ponny/myfile.txt
drwxr-xr-x - ponny supergroup 0 2025-01-13 11:28 /user/ponny/naveen
drwxr-xr-x - ponny supergroup 0 2021-05-19 12:14 /user/ponny/newfile
-rw-r--r-- 1 ponny supergroup 44 2018-10-24 16:40 /user/ponny/rin1
-rw-r--r-- 1 ponny supergroup 78 2018-10-24 16:40 /user/ponny/rin2
drwxr-xr-x - ponny supergroup 0 2018-10-24 16:42 /user/ponny/rout
drwxr-xr-x - ponny supergroup 0 2024-12-20 13:06 /user/ponny/thameen
ponny@ubuntu:~$
```

11. cp

- **Command:**

`hdfs dfs -cp /source_path /destination_path`

- **Usage:** Copies files or directories within HDFS. It is useful for creating duplicates of data for backup or other purposes.

```
ponny@ubuntu:~$ hadoop fs -cp /user/ponny/newfile /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

ponny@ubuntu:~$ hadoop fs -ls /user/ponny/naveen/
Warning: $HADOOP_HOME is deprecated.

Found 6 items
-rw-r--r--  1 ponny supergroup      46 2025-01-13 11:25 /user/ponny/naveen/copyfile.txt
-rw-r--r--  1 ponny supergroup      0 2025-01-13 11:13 /user/ponny/naveen/hadoopfile.txt
drwxr-xr-x  - ponny supergroup      0 2025-01-13 11:08 /user/ponny/naveen/localfile.txt
-rw-r--r--  1 ponny supergroup      0 2025-01-13 11:12 /user/ponny/naveen/localfile1.txt
-rw-r--r--  1 ponny supergroup      0 2025-01-13 11:30 /user/ponny/naveen/myfile.txt
drwxr-xr-x  - ponny supergroup      0 2025-01-13 11:31 /user/ponny/naveen/newfile
ponny@ubuntu:~$
```