Hadoop file system most frequently used commands.

(After start-dfs.sh and start-yarn.sh run these commands)

First step:

start-dfs.sh && start-yarn.sh

Testing purpose please run jps to check all daemons running or not.

```
hadoop@bigdata:~$ start-dfs.sh && start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/h
oop-hadoop-namenode-bigdata.out
localhost: starting datanode, logging to /home/h
oop-hadoop-datanode-bigdata.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to
ogs/hadoop-hadoop-secondarynamenode-bigdata.out
starting yarn daemons
starting resourcemanager, logging to /home/hadoo
doop-resourcemanager-bigdata.out
localhost: starting nodemanager, logging to /hom
yarn-hadoop-nodemanager-bigdata.out
hadoop@bigdata:~$ jps
4484 NodeManager
3877 DataNode
4325 ResourceManager
4522 Jps
4108 SecondaryNameNode
3711 NameNode
```

http://archive.ics.uci.edu/ml/machine-learning-databases/00222/bank.zip

http://stat-computing.org/dataexpo/2009/2008.csv.bz2

https://raw.github.com/vincentarelbundock/Rdatasets/master/csv/datasets/Titanic.csv

https://www.briandunning.com/sample-data/us-500.zip

Unzip these files as well

Place above data in work datasets folde

To learn & demonstration purpose please download few files from these links: and place in work folder

0) hdfs dfs -mkdir <path>

Its used to create a folders, based on usecases create any number of folders. Makesure / (root) must be exists. If you are not mention its stored in default location.

```
hadoop@bigdata:~$ hdfs dfs -mkdir /datasets
hadoop@bigdata:~$ 
```

1 hdfs dfs -ls <path>

Lists the contents of the directory specified by path, showing the names, permissions, owner, size and modification date for each entry.

```
hadoop@bigdata:~$ hdfs dfs -mkdir /datasets
hadoop@bigdata:~$ hdfs dfs -ls /
Found 11 items
drwxr-xr-x   - hadoop supergroup          0 2017-11-24 07:17 /asl
drwxr-xr-x   - hadoop supergroup          0 2017-11-24 06:55 /asldata
drwxr-xr-x   - hadoop supergroup          0 2017-11-18 09:13 /asltab
drwxr-xr-x   - hadoop supergroup          0 2017-11-25 17:34 /datasets
drwxr-xr-x   - hadoop supergroup          0 2017-11-18 08:27 /expdir
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 07:58 /home
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 08:04 /sqoopimportdata
-rwxrwxrwx   1 hadoop supergroup        139 2017-11-18 09:22 /sqoopscript.txt
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 08:09 /tmp
drwxr-xr-x   - hadoop supergroup          0 2017-11-18 08:56 /usdata
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 08:14 /user
hadoop@bigdata:~$ 
```

hdfs dfs -ls /

2. hdfs dfs **-lsr <path>**

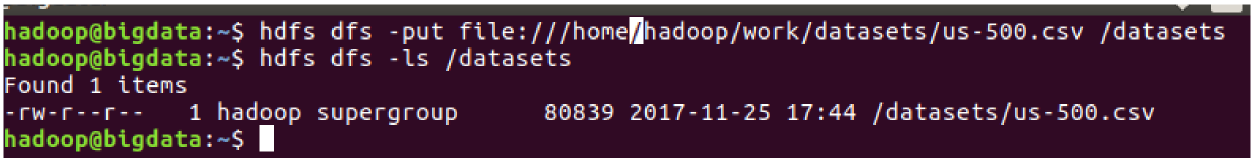
Behaves like -ls, but recursively displays entries in all subdirectories of path.
hdfs dfs -ls -R /

```
hadoop@bigdata:~$ hdfs dfs -ls -R /
drwxr-xr-x   - hadoop supergroup          0 2017-11-24 07:17 /asl
-rw-r--r--   1 hadoop supergroup        180 2017-11-24 07:17 /asl/asl1.txt
drwxr-xr-x   - hadoop supergroup          0 2017-11-24 06:55 /asldata
-rw-r--r--   1 hadoop supergroup        180 2017-11-24 06:55 /asldata/asl1.txt
drwxr-xr-x   - hadoop supergroup          0 2017-11-18 09:13 /asltab
-rw-r--r--   1 hadoop supergroup         51 2017-11-18 09:13 /asltab/us-500.csv
drwxr-xr-x   - hadoop supergroup          0 2017-11-25 17:34 /datasets
drwxr-xr-x   - hadoop supergroup          0 2017-11-18 08:27 /expdir
-rw-r--r--   1 hadoop supergroup         51 2017-11-18 08:27 /expdir/us-500.csv
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 07:58 /home
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 07:58 /home/hadoop
drwxr-xr-x   - hadoop supergroup          0 2017-11-17 08:15 /home/hadoop/work
```

8. hdfs dfs **-put <localSrc> <dest>**

Copies the file or directory from the local file system identified by localSrc to dest within the DFS.

hdfs dfs -put file:///home/hadoop/work/datasets/us-500.csv /datasets

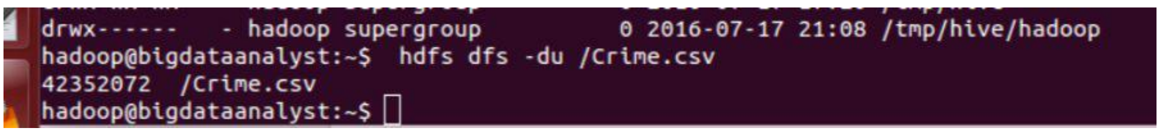


```
hadoop@bigdata:~$ hdfs dfs -put file:///home/hadoop/work/datasets/us-500.csv /datasets
hadoop@bigdata:~$ hdfs dfs -ls /datasets
Found 1 items
-rw-r--r--   1 hadoop supergroup      80839 2017-11-25 17:44 /datasets/us-500.csv
hadoop@bigdata:~$
```

3. hdfs dfs **-du <path>**

Shows disk usage, in bytes, for all the files which match path; filenames are reported with the full HDFS protocol prefix.

hdfs dfs -du /

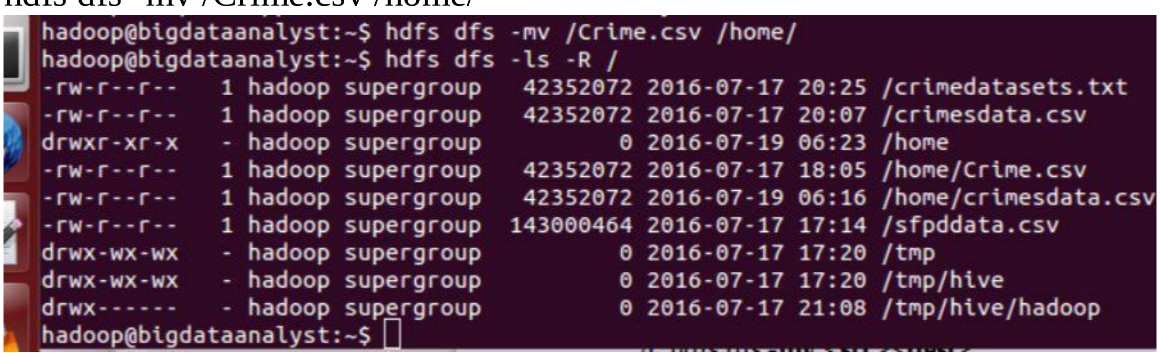


```
drwx------   - hadoop supergroup          0 2016-07-17 21:08 /tmp/hive/hadoop
hadoop@bigdataanalyst:~$  hdfs dfs -du /Crime.csv
42352072  /Crime.csv
hadoop@bigdataanalyst:~$
```

4. hdfs dfs **-mv <src><dest>**

Moves the file or directory indicated by src to dest, within HDFS.

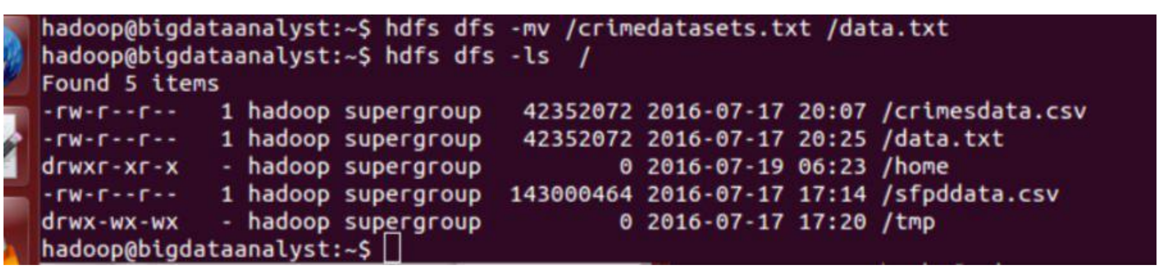
hdfs dfs -mv /Crime.csv /home/



```
hadoop@bigdataanalyst:~$ hdfs dfs -mv /Crime.csv /home/
hadoop@bigdataanalyst:~$ hdfs dfs -ls -R /
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:25 /crimedatasets.txt
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:07 /crimesdata.csv
drwxr-xr-x   - hadoop supergroup          0 2016-07-19 06:23 /home
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 18:05 /home/Crime.csv
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-19 06:16 /home/crimesdata.csv
-rw-r--r--   1 hadoop supergroup  143000464 2016-07-17 17:14 /sfpddata.csv
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp/hive
drwx------   - hadoop supergroup          0 2016-07-17 21:08 /tmp/hive/hadoop
hadoop@bigdataanalyst:~$
```

MV also useful to rename the dataset

let example



```
hadoop@bigdataanalyst:~$ hdfs dfs -mv /crimedatasets.txt /data.txt
hadoop@bigdataanalyst:~$ hdfs dfs -ls  /
Found 5 items
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:07 /crimesdata.csv
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:25 /data.txt
drwxr-xr-x   - hadoop supergroup          0 2016-07-19 06:23 /home
-rw-r--r--   1 hadoop supergroup  143000464 2016-07-17 17:14 /sfpddata.csv
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp
hadoop@bigdataanalyst:~$
```

5. hdfs dfs **-cp <src> <dest>**
Copies the file or directory identified by src to dest, within HDFS.
hdfs dfs -cp /data.txt /home/data/

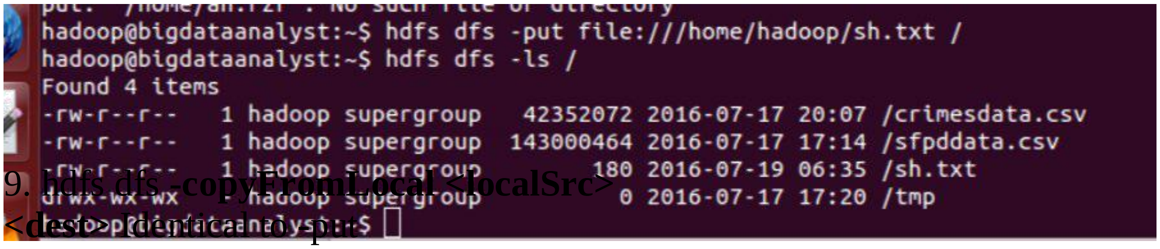
```
hadoop@bigdataanalyst:~$ hdfs dfs -cp /crimesdata.csv /home/
hadoop@bigdataanalyst:~$ hdfs dfs -ls /
Found 6 items
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 18:05 /Crime.csv
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:25 /crimedatasets.txt
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:07 /crimesdata.csv
drwxr-xr-x   - hadoop supergroup          0 2016-07-19 06:16 /home
-rw-r--r--   1 hadoop supergroup  143000464 2016-07-17 17:14 /sfpddata.csv
```

6 hdfs dfs **-rm <path>**
Removes the file or empty directory identified by path.
hdfs dfs -rm /data.txt

```
hadoop@bigdataanalyst:~$ hdfs dfs -rm /data.txt
16/07/19 06:28:13 INFO fs.TrashPolicyDefault: Namenode trash configuration: Delet
Deleted /data.txt
hadoop@bigdataanalyst:~$ hdfs dfs -ls  /
Found 4 items
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:07 /crimesdata.csv
drwxr-xr-x   - hadoop supergroup          0 2016-07-19 06:23 /home
-rw-r--r--   1 hadoop supergroup  143000464 2016-07-17 17:14 /sfpddata.csv
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp
hadoop@bigdataanalyst:~$
```

7. **rmr <path>**
Removes the file or directory identified by path. Recursively deletes any child entries
i. e. , filesorsubdirectoriesofpath.
hdfs dfs -rmr /home/

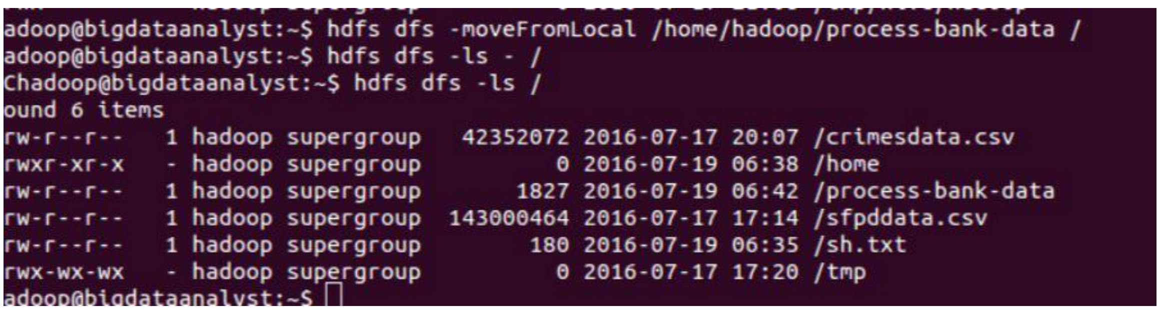
```
hadoop@bigdataanalyst:~$ hdfs dfs -rmr /home/
rmr: DEPRECATED: Please use 'rm -r' instead.
16/07/19 06:30:01 INFO fs.TrashPolicyDefault: Namenode trash configuration: Delet
Deleted /home
hadoop@bigdataanalyst:~$ hdfs dfs -ls -R /
-rw-r--r--   1 hadoop supergroup   42352072 2016-07-17 20:07 /crimesdata.csv
-rw-r--r--   1 hadoop supergroup  143000464 2016-07-17 17:14 /sfpddata.csv
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp
drwx-wx-wx   - hadoop supergroup          0 2016-07-17 17:20 /tmp/hive
drwx------   - hadoop supergroup          0 2016-07-17 21:08 /tmp/hive/hadoop
hadoop@bigdataanalyst:~$
```



9. hdfs dfs **-copyFromLocal <localSrc> <dest>** Identical to -put

hdfs dfs -copyFromLocal file:///home/hadoop/sh.txt /home/

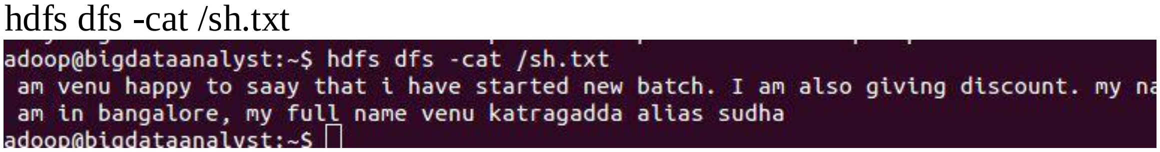
10. hdfs dfs **-moveFromLocal <localSrc> <dest>**
Copies the file or directory from the local file system identified by localSrc to dest within HDFS, and then deletes the local copy on success. hdfs dfs -moveFromLocal /home/hadoop/process-bank-data /



Now local data is deleted

11. hdfs dfs **-get [-crc] <src> <localDest>**
Copies the file or directory in HDFS identified by src to the local file system path identified by localDest.
hdfs dfs -get /process-bank-data /home/hadoop/

12. hdfs dfs **-cat <file-name>**
Displays the contents of filename on stdout.
hdfs dfs -cat /sh.txt



13. hdfs dfs **-copyToLocal <src> <localDest>** Identical to -get

hdfs dfs -copyToLocal /sfpddata.csv /home/hadoop/Desktop/

Resource: https://hadoop.apache.org/docs/r2.7.1/hadoop-project-dist/hadoop-common/FileSystemShell.html