

Hadoop commands:

Hadoop commands and linux commands almost same.

before run this hdfs commands please download sample datasets from these links and save in work folder datasets folder

<https://www.briandunning.com/sample-data/us-500.zip> //extract this file and place in work/datasets folder

<https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/Titanic.csv>

<http://archive.ics.uci.edu/ml/machine-learning-databases/00222/bank.zip> inside bank-full.csv

resources:

<https://hadoop.apache.org/docs/2.4.1/hadoop-project-dist/hadoop-common/FileSystemShell.html>

1) hdfs dfs -**mkdir** <path>

Creates a directory named path in HDFS.

Creates any parent directories in path that are missing *e. g. , mkdir -pinLinux*.

```
hadoop@bigdataanalyst:~$ hdfs dfs -ls /
Found 5 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
-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:~$ hdfs dfs -mkdir /home
hadoop@bigdataanalyst:~$
```

hdfs dfs -mlkdir /home

Similarly create many folders

hdfs dfs -mkdir /datasets

hdfs dfs -mkdir /titanicdata

hdfs dfs -mkdir /otherdataset

2) **hdfs dfs -put <localSrc> <dest>**

```
put: /home/ah.121: No such file or directory
hadoop@bigdataanalyst:~$ hdfs dfs -put file:///home/hadoop/sh.txt /
hadoop@bigdataanalyst:~$ hdfs dfs -ls /
Found 4 items
-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
-rw-r--r-- 1 hadoop supergroup 180 2016-07-19 06:35 /sh.txt
drwx-wx-wx - hadoop supergroup 0 2016-07-17 17:20 /tmp
hadoop@bigdataanalyst:~$
```

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

`hdfs dfs -put file:///home/hadoop/sh.txt /`

`hdfs dfs -put file:///home/hadoop/work/datasets/Titanic.csv /titanicdata`

//Now you have stored data in HDFS. Now verify this data stored in hdfs or not

3) **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@bigdataanalyst:~$ hdfs dfs -ls /
Found 5 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
-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:~$
```

`hdfs dfs -ls /`

`hdfs dfs -ls /Titanicdata`

//If you want to display recursively inside that folder, use `-lsr` or `-ls -R`

4) `hdfs dfs -lsr <path>`

```
hadoop@bigdataanalyst:~$ hdfs dfs -lsr /
lsr: DEPRECATED: Please use 'ls -R' instead.
-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 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:~$ hdfs dfs -ls -R /
-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 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
```

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

`hdfs dfs -lsr /`

//now you have stored data in hdfs, if you want to read that data, uses `-cat`

5) `hdfs dfs -cat <file-name>`

Displays the contents of filename on stdout.

`hdfs dfs -cat /sh.txt`

```
hadoop@bigdataanalyst:~$ hdfs dfs -cat /sh.txt
am venu happy to saay that i have started new batch. I am also giving discount. my na
am in bangalore, my full name venu katragadda alias sudha
hadoop@bigdataanalyst:~$
```

6) `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:~$
```

7) 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
drwx-wx-wx - hadoop supergroup 0 2016-07-17 17:20 /tmp
```

8) 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:~$
```

9) **hdfs dfs -rmr <path>**

Removes the file or directory identified by path. Recursively deletes any child entries

i. e. , files or subdirectories of path.

`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: Deletion
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:~$
```

10) **hdfs dfs -copyFromLocal <localSrc> <dest>**

Identical to -put

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

11) **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 /`

```
hadoop@bigdataanalyst:~$ hdfs dfs -moveFromLocal /home/hadoop/process-bank-data /
hadoop@bigdataanalyst:~$ hdfs dfs -ls - /
Chadoop@bigdataanalyst:~$ hdfs dfs -ls /
Found 6 items
-rw-r--r-- 1 hadoop supergroup 42352072 2016-07-17 20:07 /crimesdata.csv
-rwxr-xr-x - hadoop supergroup 0 2016-07-19 06:38 /home
-rw-r--r-- 1 hadoop supergroup 1827 2016-07-19 06:42 /process-bank-data
-rw-r--r-- 1 hadoop supergroup 143000464 2016-07-17 17:14 /sfpddata.csv
-rw-r--r-- 1 hadoop supergroup 180 2016-07-19 06:35 /sh.txt
-rwx-wx-wx - hadoop supergroup 0 2016-07-17 17:20 /tmp
hadoop@bigdataanalyst:~$
```

Now local data is deleted

12) `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`

www.bigdataanalyst.in