

Basic Commands of Hadoop (PART-1)

- To check whether all the daemons are working or not.

Command: jps

Description: When you run Hadoop on any machine, you can look at the specific processes of Hadoop through one of the utilities provided by Java called the JPS (Java Virtual Machine Process Status) tool.

Output:

```
hduser@ubuntu:~/hadoop-2.8.2/sbin$ jps
3218 ResourceManager
3662 SecondaryNameNode
2648 NameNode
3352 NodeManager
2762 DataNode
3676 Jps
hduser@ubuntu:~/hadoop-2.8.2/sbin$
```

- To know the contents or list of contents.

Command: `hadoop fs -ls` → to know list of contents in the hadoop home directory.

Command : `hadoop fs -ls /` → to know list of contents in the hadoop root directory.

Output 1:

```
cloudera@quickstart:~$ hadoop fs -ls
Found 18 items
drwxr-xr-x - cloudera cloudera      0 2022-09-06 22:29 csvfiles
drwxr-xr-x - cloudera cloudera      0 2022-09-07 01:50 dir1
drwxr-xr-x - cloudera cloudera      0 2022-09-07 01:38 external_files
-rwxr-xr-x 1 cloudera cloudera      0 2022-09-01 23:10 fl
drwxr-xr-x - cloudera cloudera      0 2022-09-01 00:29 folder1
drwxr-xr-x - cloudera cloudera      0 2022-09-01 09:02 folder
drwxr-xr-x - cloudera cloudera      0 2022-08-31 07:30 lineuram
-rwxr-xr-x 1 cloudera cloudera    38 2022-09-01 23:40 q3.txt
drwxr-xr-x - cloudera cloudera      0 2022-08-31 07:04 soumya
drwxr-xr-x - cloudera cloudera      0 2022-09-01 00:43 temp
cloudera@quickstart:~$
```

Output 2:

```
cloudera@quickstart:~$ hadoop fs -ls /
Found 15 items
drwxr-xr-x - cloudera supergroup      0 2022-09-22 02:43 /-ext-10000
-rwxr-xr-x 1 cloudera supergroup    53351 2022-09-20 08:39 /AgentLoggingReport.csv
drwxr-xr-x - hdfs supergroup          0 2017-10-23 09:15 /benchmarks
-rwxr-xr-x 1 cloudera supergroup      135 2022-09-06 00:55 /csv1.csv
-rwxr-xr-x 1 cloudera supergroup       42 2022-09-01 23:14 /dir1
-rwxr-xr-x 1 cloudera supergroup       42 2022-09-01 23:15 /f2.txt
-rwxr-xr-x 1 cloudera supergroup        5 2022-09-01 09:13 /folder1
-rwxr-xr-x 1 cloudera supergroup     119 2022-09-01 01:19 /folder
drwxr-xr-x - hbase supergroup          0 2022-09-27 09:02 /hbase
drwxr-xr-x - cloudera supergroup        0 2022-09-22 04:07 /home
drwxr-xr-x - cloudera supergroup        0 2022-09-07 02:53 /inputfile.csv
drwxr-xr-x - solr solr                  0 2017-10-23 09:18 /solr
drwxr-xr-x - hdfs supergroup          0 2022-08-01 01:26 /top
drwxr-xr-x - hdfs supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup          0 2017-10-23 09:17 /var
cloudera@quickstart:~$
```

- If we want to list the contents for a particular directory.

Command: `hadoop fs -ls /directoryname`

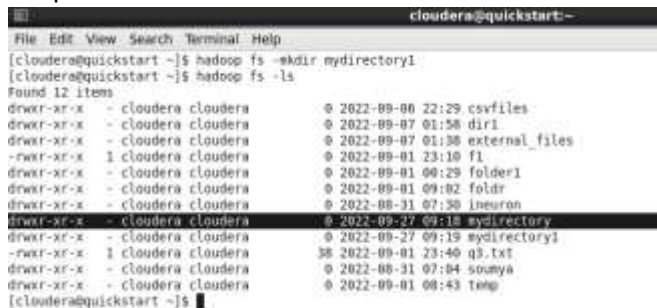
Output:

```
[cloudera@quickstart ~]$ hadoop fs -ls /dir1
-rwxr-xr-x  1 cloudera supergroup          42 2022-09-01 23:14 /dir1
[cloudera@quickstart ~]$
```

- If we want to create a directory .

Command: `hadoop fs -mkdir directoryname`

Output:

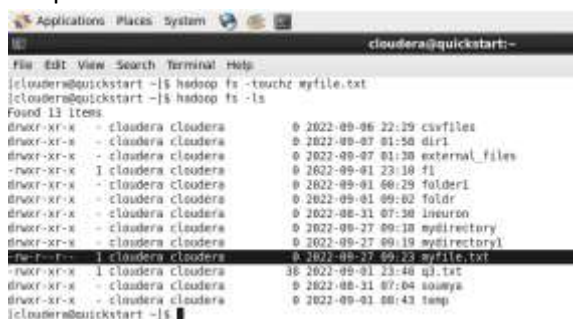


```
cloudera@quickstart:~$ hadoop fs -mkdir mydirectory1
[cloudera@quickstart ~]$ hadoop fs -ls
Found 12 items
drwxr-xr-x - cloudera cloudera @ 2022-09-06 22:29 csvfiles
drwxr-xr-x - cloudera cloudera @ 2022-09-07 01:58 dir1
drwxr-xr-x - cloudera cloudera @ 2022-09-07 01:38 external_files
-rwxr-xr-x 1 cloudera cloudera @ 2022-09-01 23:10 fl
drwxr-xr-x - cloudera cloudera @ 2022-09-01 00:29 folder1
drwxr-xr-x - cloudera cloudera @ 2022-09-01 09:02 folder
drwxr-xr-x - cloudera cloudera @ 2022-08-31 07:30 ineuron
drwxr-xr-x - cloudera cloudera @ 2022-09-27 09:18 mydirectory
drwxr-xr-x - cloudera cloudera @ 2022-09-27 09:19 mydirectory1
-rwxr-xr-x 1 cloudera cloudera 38 2022-09-01 23:40 q3.txt
drwxr-xr-x - cloudera cloudera @ 2022-08-31 07:04 soumya
drwxr-xr-x - cloudera cloudera @ 2022-09-01 00:43 temp
[cloudera@quickstart ~]$
```

- To create an empty file of any format.

Command: `hadoop fs -touchz filename`

Output:



```
cloudera@quickstart:~$ hadoop fs -touchz myfile.txt
[cloudera@quickstart ~]$ hadoop fs -ls
Found 13 items
drwxr-xr-x - cloudera cloudera @ 2022-09-06 22:29 csvfiles
drwxr-xr-x - cloudera cloudera @ 2022-09-07 01:58 dir1
drwxr-xr-x - cloudera cloudera @ 2022-09-07 01:38 external_files
-rwxr-xr-x 1 cloudera cloudera @ 2022-09-01 23:10 fl
drwxr-xr-x - cloudera cloudera @ 2022-09-01 00:29 folder1
drwxr-xr-x - cloudera cloudera @ 2022-09-01 09:02 folder
drwxr-xr-x - cloudera cloudera @ 2022-08-31 07:30 ineuron
drwxr-xr-x - cloudera cloudera @ 2022-09-27 09:18 mydirectory
drwxr-xr-x - cloudera cloudera @ 2022-09-27 09:19 mydirectory1
-rwxr-xr-x 1 cloudera cloudera @ 2022-09-27 09:23 myfile.txt
-rwxr-xr-x 1 cloudera cloudera 38 2022-09-01 23:40 q3.txt
drwxr-xr-x - cloudera cloudera @ 2022-08-31 07:04 soumya
drwxr-xr-x - cloudera cloudera @ 2022-09-01 00:43 temp
[cloudera@quickstart ~]$
```

- To Copy data from the local system

Command: `hadoop fs -copyFromLocal filelocation to hadoop system(location)` (OR)

`hadoop fs -put filelocation to hadoop system(location)`

Output:

```
Applications Places System cloudera cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ ls Desktop
AgentLoggingReport.csv data2.csv hive_join.csv join.csv mergesuit.txt pseudo_facebook.csv wine_data
AgentPerformance.csv data3.csv huge.csv input1.csv mergesuit.txt MOCK_DATA.csv sales_data.csv
country_wise_latest.csv demo.csv inputfile.csv input2.csv output.csv test_data.txt
csv1.csv Enterprise.desktop join.csv mergesuit.txt partition_data.csv TSRTC.pdf
data1.csv Express.desktop inputfile.csv mergesuit.txt partition_data.csv TSRTC.pdf
[cloudera@quickstart ~]$ hadoop fs -ls
Found 13 items
drwxr-xr-x - cloudera cloudera 0 2022-09-06 22:29 csvfiles
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:34 dir1
drwxr-xr-x - cloudera cloudera 0 2022-09-07 01:38 external_files
-rwxr-xr-x 1 cloudera cloudera 0 2022-09-01 23:10 fl
drwxr-xr-x - cloudera cloudera 0 2022-09-01 09:29 folder1
drwxr-xr-x - cloudera cloudera 0 2022-09-01 09:02 folder
drwxr-xr-x - cloudera cloudera 0 2022-08-31 07:38 ineuron
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:18 mydirectory
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:19 mydirectory1
-rw-r--r-- 1 cloudera cloudera 0 2022-09-27 09:23 myfile.txt
-rwxr-xr-x 1 cloudera cloudera 38 2022-09-01 23:40 q3.txt
drwxr-xr-x - cloudera cloudera 0 2022-08-31 07:04 soumya
drwxr-xr-x - cloudera cloudera 0 2022-09-01 08:43 temp
[cloudera@quickstart ~]$ hadoop fs -copyFromLocal Desktop/joininput1.csv dir1
[cloudera@quickstart ~]$
```

- To copy files from hadoop file system to local system.

Command: `hadoop fs -copyToLocal filelocation to required location.` (OR)
`hadoop fs -get filelocation to required location.`

Output:

```
[cloudera@quickstart ~]$ hadoop fs -copyToLocal myfile.txt Desktop
[cloudera@quickstart ~]$ ls Desktop
AgentLoggingReport.csv data3.csv hive_output.txt Kerberos.desktop Parcels.desktop TSRTC.pdf
AgentPerformance.csv demo.csv huge.csv mergesuit.txt partition_data.csv wine_data
country_wise_latest.csv Enterprise.desktop inputfile.csv mergesuit.txt pseudo_facebook.csv sales_data.csv
csv1.csv Express.desktop join.csv MOCK_DATA.csv test_data.txt
data1.csv myfile.txt output.csv
data2.csv hive_join.csv joininput2.csv
[cloudera@quickstart ~]$
```

- To display the contents of the file.

Command: `hadoop fs -cat filelocation`

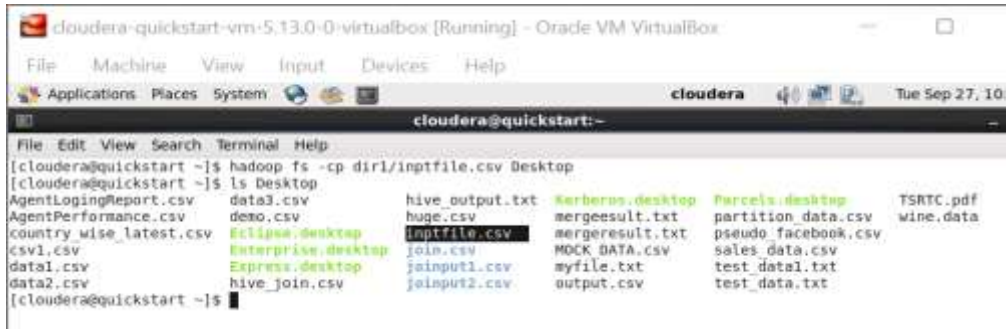
Output:

```
File Machine View Input Devices Help
Applications Places System cloudera cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -ls
Found 13 items
drwxr-xr-x - cloudera cloudera 0 2022-09-06 22:29 csvfiles
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:37 dir1
drwxr-xr-x - cloudera cloudera 0 2022-09-07 01:38 external_files
-rwxr-xr-x 1 cloudera cloudera 0 2022-09-01 23:10 fl
drwxr-xr-x - cloudera cloudera 0 2022-09-01 09:29 folder1
drwxr-xr-x - cloudera cloudera 0 2022-08-31 07:38 ineuron
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:18 mydirectory
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:19 mydirectory1
-rw-r--r-- 1 cloudera cloudera 0 2022-09-27 09:23 myfile.txt
-rwxr-xr-x 1 cloudera cloudera 38 2022-09-01 23:40 q3.txt
drwxr-xr-x - cloudera cloudera 0 2022-08-31 07:04 soumya
drwxr-xr-x - cloudera cloudera 0 2022-09-01 08:43 temp
[cloudera@quickstart ~]$ hadoop fs -cat q3.txt
abcdefghijklmnopqrstuvwxyz
abcdefghijklmnopqrstuvwxyz
unopgrstuvwxzy
[cloudera@quickstart ~]$ hadoop fs -ls dir1
Found 4 items
-rwxr-xr-x 1 cloudera cloudera 14784 2022-09-07 01:29 dir1/country_wise_latest.csv
-rwxr-xr-x 1 cloudera cloudera 56 2022-09-07 01:58 dir1/inputfile.csv
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:34 dir1/join.csv
drwxr-xr-x - cloudera cloudera 0 2022-09-27 09:37 dir1/joininput1.csv
[cloudera@quickstart ~]$ hadoop fs -cat dir1/inputfile.csv
1,sou,five
2,ari,six
3,honey,eight
[cloudera@quickstart ~]$
```

- To copy file between local system and hadoop file system.

Command: `hadoop fs -cp filelocation(to be transfered) location(file to be saved)`

Output:

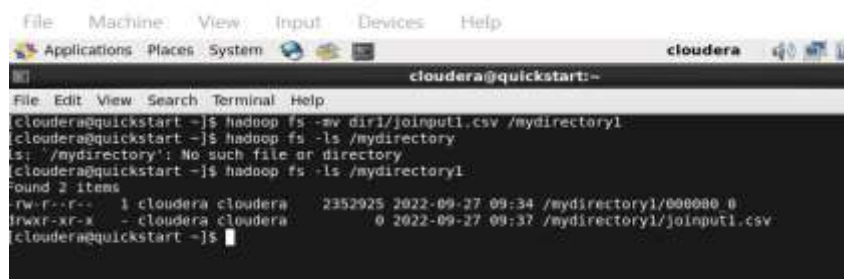


```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera Tue Sep 27, 10
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -cp dir1/inptfile.csv Desktop
[cloudera@quickstart ~]$ ls Desktop
AgentLoggingReport.csv      data3.csv      hive_output.txt      Kerberos.desktop      Parcels.desktop      TSRTC.pdf
AgentPerformance.csv       demo.csv       huge.csv             mergeesult.txt        partition_data.csv    wine.data
country_wise_latest.csv    Eclipse.desktop inptfile.csv         mergeresult.txt       pseudo_facebook.csv
csv1.csv                   Enterprise.desktop join.csv             MOCK_DATA.csv         sales_data.csv
data1.csv                   Express.desktop joininput1.csv        myfile.txt            test_data1.txt
data2.csv                   hive_join.csv  joininput2.csv        output.csv            test_data.txt
```

- To move file from one directory to another directory or between local and hadoop systems.

Command: `hadoop fs -mv filelocation(to be transferred) location(file to be saved)`

Output:

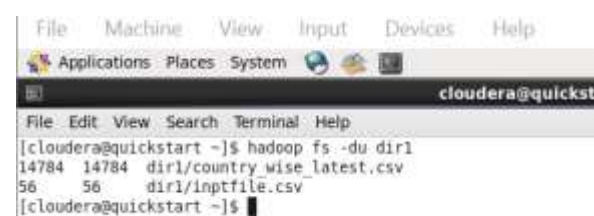


```
File Machine View Input Devices Help
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -mv dir1/joinput1.csv /mydirectory1
[cloudera@quickstart ~]$ hadoop fs -ls /mydirectory1
ls: '/mydirectory1': No such file or directory
[cloudera@quickstart ~]$ hadoop fs -ls /mydirectory1
Found 2 items
-rw-r--r-- 1 cloudera cloudera 2352925 2022-09-27 09:34 /mydirectory1/0000000_0
-rwxr-xr-x - cloudera cloudera 0 2022-09-27 09:37 /mydirectory1/joinput1.csv
[cloudera@quickstart ~]$
```

- To check the disk usage of any directory.

Command: `hadoop fs -du directoryname`

Output :

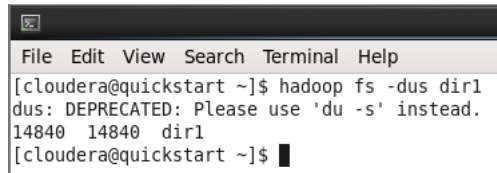


```
File Machine View Input Devices Help
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -du dir1
14784 14784 dir1/country_wise_latest.csv
56 56 dir1/inptfile.csv
[cloudera@quickstart ~]$
```

- **To check the disk usage of the entire directories at one place.**

Command: `hadoop fs -dus directoryname`

Output:



```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -dus dir1
dus: DEPRECATED: Please use 'du -s' instead.
14840 14840 dir1
[cloudera@quickstart ~]$
```