



**DICE**  
ANALYTICS

# HDFS Commands

---

By: MOBEEN AHMED & FARAN SHAHID

## UNIX Commands

Here are the basic UNIX commands we are going to use.

**NOTE: PLEASE USE YOUR HANDS INSTEAD OF CTRL+C & CTRL+V**

- Know your current directory

The pwd command writes the full pathname of the current working directory. It tells you where you are and in which directory, starting from the root (/).

```
[root@sandbox ~]# pwd
/root
```

- List local files or directories

To list all the files or directories in the current directory we use ls command:

ls to view files in a directory.

```
[root@sandbox ~]# ls
anaconda-ks.cfg  demos                pg20417.txt          start_hbase.sh  tmp
blueprint.json  install.log          sandbox.info          start_solr.sh
build.out        install.log.syslog  start_ambari.sh      stop_solr.sh
```

ls -l to view files and additional information about each file.

```
[root@sandbox ~]# ls -l
total 724
-rw----- 1 root root 2139 2016-03-11 09:50 anaconda-ks.cfg
-rw-r--r-- 1 root root 18737 2016-03-11 10:19 blueprint.json
-rw-r--r-- 1 root root 20 2016-03-11 10:53 build.out
drwxr-xr-x 3 root root 4096 2018-07-27 15:09 demos
-rw-r--r-- 1 root root 9447 2016-03-11 09:50 install.log
-rw-r--r-- 1 root root 3314 2016-03-11 09:49 install.log.syslog
-rw-r--r-- 1 root root 674570 2018-07-22 10:54 pg20417.txt
-rw-r--r-- 1 root root 283 2016-03-11 10:55 sandbox.info
lrwxrwxrwx 1 root root 48 2016-03-11 10:19 start_ambari.sh -> /usr/lib/hue/tools
/start_scripts/start_ambari.sh
lrwxrwxrwx 1 root root 47 2016-03-11 10:19 start_hbase.sh -> /usr/lib/hue/tools/
start_scripts/start_hbase.sh
```

- Make a new directory

**mkdir <directory\_name>** to make a new directory in the current directory

```
[root@sandbox ~]# mkdir new_dir
[root@sandbox ~]# ls -l
total 728
-rw-----. 1 root root 2139 2016-03-11 09:50 anaconda-ks.cfg
-rw-r--r-- 1 root root 18737 2016-03-11 10:19 blueprint.json
-rw-r--r-- 1 root root 20 2016-03-11 10:53 build.out
drwxr-xr-x 3 root root 4096 2018-07-27 15:09 demos
-rw-r--r-- 1 root root 9447 2016-03-11 09:50 install.log
-rw-r--r-- 1 root root 3314 2016-03-11 09:49 install.log.syslog
drwxr-xr-x 2 root root 4096 2018-08-30 20:10 new_dir
-rw-r--r-- 1 root root 674570 2018-07-22 10:54 pg20417.txt
-rw-r--r-- 1 root root 283 2016-03-11 10:55 sandbox.info
lrwxrwxrwx 1 root root 48 2016-03-11 10:19 start_ambari.sh -> /usr/lib/hue/tools/
/start_scripts/start_ambari.sh
lrwxrwxrwx 1 root root 47 2016-03-11 10:19 start_hbase.sh -> /usr/lib/hue/tools/
start_scripts/start_hbase.sh
-rwxrwxrwx 1 packer packer 92 2016-03-11 10:23 start_solr.sh
-rwxrwxrwx 1 packer packer 62 2016-03-11 10:25 stop_solr.sh
drwxr-xr-x 2 root root 4096 2018-07-22 10:55 tmp
```

- Traverse through directories

You can move between the directories using **cd** command like that:

**cd <path>** to change directory

```
[root@sandbox ~]# cd new_dir/
[root@sandbox new_dir]#
```

**cd ..** to go back to the parent directory

```
[root@sandbox new_dir]# cd ..
[root@sandbox ~]#
```

- Create a copy of an existing local file

**cp install.log my\_install.log** to create a copy of install.log with the name my\_install.log in the current directory

```
[root@sandbox ~]# cp install.log my_install.log
[root@sandbox ~]# ls -l
total 740
-rw-----. 1 root root 2139 2016-03-11 09:50 anaconda-ks.cfg
-rw-r--r-- 1 root root 18737 2016-03-11 10:19 blueprint.json
-rw-r--r-- 1 root root 20 2016-03-11 10:53 build.out
drwxr-xr-x 3 root root 4096 2018-07-27 15:09 demos
-rw-r--r-- 1 root root 9447 2016-03-11 09:50 install.log
-rw-r--r-- 1 root root 3314 2016-03-11 09:49 install.log.syslog
-rw-r--r-- 1 root root 9447 2018-08-30 20:23 my_install.log
```

- View the content of file

Use **less** command to view the contents of file. Press **q** when you are finished.

```
less my_install.log
```

```
[root@sandbox ~]# less my_install.log
```

- Using VIM text editor

We use VIM editor to edit file contents

```
vi my_install.log
```

Press “**Esc**” to enter command mode. In command mode:

- Press “**i**” to insert text in file
- Press “**:wq**” to save file with the changes and press **Enter**
- Press “**:!q**” to exit without saving changes and press **Enter**

```
:q!
```

## HDFS Commands

To access the file in HDFS you can also use **hadoop fs** instead of **hdfs dfs** in the following commands. These two commands are interchangeable.

- List the contents of a directory

Just the UNIX command above we can list the content of directory using **-ls** but placing **hadoop fs** or **hdfs dfs** before that

```
hdfs dfs -ls <directory_path>
```

For example: **hdfs dfs -ls /Class-2**

```
[root@sandbox ~]# hdfs dfs -ls /Class-2
Found 1 items
drwxr-xr-x  - root hdfs          0 2018-07-13 14:55 /Class-2/hive
```

- Create a directory

**hdfs dfs -mkdir /<directory\_name>**

For example: hdfs dfs -mkdir /Installer

```
[root@sandbox ~]# hdfs dfs -mkdir /Installer
[root@sandbox ~]# hdfs dfs -ls /
Found 13 items
drwxr-xr-x  - root    hdfs          0 2018-07-13 14:52 /Class-2
drwxr-xr-x  - root    hdfs          0 2018-08-30 21:00 /Installer
drwxrwxrwx  - yarn    hadoop        0 2018-07-11 15:01 /app-logs
drwxr-xr-x  - hdfs    hdfs          0 2016-03-11 10:18 /apps
drwxr-xr-x  - yarn    hadoop        0 2016-03-11 10:12 /ats
drwxr-xr-x  - hdfs    hdfs          0 2016-03-11 10:41 /demo
drwxr-xr-x  - hdfs    hdfs          0 2016-03-11 10:12 /hdp
drwxr-xr-x  - mapred  hdfs          0 2016-03-11 10:12 /mapred
drwxrwxrwx  - mapred  hadoop        0 2016-03-11 10:12 /mr-history
drwxr-xr-x  - hdfs    hdfs          0 2016-03-11 10:33 /ranger
drwxrwxrwx  - spark   hadoop        0 2018-07-12 16:09 /spark-histor
```

- View file on HDFS

**hdfs dfs -cat /<file\_path>**

For example: hdfs dfs -cat /Class-2/full\_data.csv

```
[root@sandbox ~]# hdfs dfs -cat /Class-2/full_data.csv
id, name, province
1001,Owais,Punjab
1002,Haider,Blouchistan
1003,Fahad,KPK
1004,Abdullah,Punjab
1005,Qaiser,Sindh
1006,Khalid,KPK
1007,Raheel,Blouchistan
1008,Ahmed,KPK
1009,Faisal,Sindh
1010,Saad,Punjab[root@sandbox ~]#
```

- Change Permissions

**hdfs dfs -chmod <permissions> /<directory\_name>**

For example: hdfs dfs -chmod 777 /Installer

Before permissions

```
[root@sandbox ~]# hdfs dfs -ls /
Found 13 items
drwxr-xr-x  - root    hdfs          0 2018-07-13 14:52 /Class-2
drwxr-xr-x  - root    hdfs          0 2018-08-30 21:00 /Installer
drwxrwxrwx  - yarn    hadoop        0 2018-07-11 15:01 /app-logs
```

After applying permissions

```
[root@sandbox ~]# hdfs dfs -chmod 777 /Installer
[root@sandbox ~]# hdfs dfs -ls /
Found 13 items
drwxr-xr-x   - root    hdfs          0 2018-08-30 21:12 /Class-2
drwxrwxrwx   - root    hdfs          0 2018-08-30 21:00 /Installer
drwxrwxrwx   - yarn    hadoop        0 2018-07-11 15:01 /app-logs
```

- Upload file from NDFS to HDFS

**hdfs dfs -put ./<ndfs\_file\_path> /<hdfs\_directory\_path>**

For example: hdfs dfs -put ./my\_install.log /Installer

```
[root@sandbox ~]# hdfs dfs -put ./my_install.log /Installer
[root@sandbox ~]# hdfs dfs -ls /Installer
Found 1 items
-rw-r--r--   3 root    hdfs          9448 2018-08-30 21:19 /Installer/my_install.log
```

- Put the file with fixed block size

Try putting the file into HDFS with a block size of 30 bytes

**hdfs dfs -D dfs.blocksize=30 -put ./<ndfs\_file\_path> /<hdfs\_directory\_path>**

For example: hdfs dfs -D dfs.blocksize=1048576 -put ./my\_install.log /Installer

```
[root@sandbox ~]# hdfs dfs -D dfs.blocksize=1048576 -put ./my_install.log /Installer
[root@sandbox ~]# hdfs dfs -ls /Installer
Found 1 items
-rw-r--r--   3 root    hdfs          9447 2018-08-30 21:37 /Installer/my_install.log
```

The block size needs to be at least 1048576 according to the **dfs.namenode.fs-limits.min-block-size**. Add the block size that is multiple of 512 (the checksum size).

- View number of blocks

**hdfs dfs fsck /<hdfs\_file\_path> -files -blocks**

For example: `hdfs fsck /Installer/my_install.log -files -blocks`

```
[root@sandbox ~]# hdfs fsck /Installer/my_install.log -files -blocks
Connecting to namenode via http://sandbox.hortonworks.com:50070/fsck?ugi=root&files=1&blocks=1&path=%2FInstaller%2Fmy_install.log
FSCK started by root (auth:SIMPLE) from /192.168.183.153 for path /Installer/my_install.log at Thu Aug 30 21:50:04 UTC 2018
/Installer/my_install.log 9447 bytes, 1 block(s): Under replicated BP-267552868-172.16.137.143-1457691099567:blk_1073742645_1829. Target Replicas is 3 but found 1 live replica(s), 0 decommissioned replica(s) and 0 decommissioning replica(s).
0. BP-267552868-172.16.137.143-1457691099567:blk_1073742645_1829 len=9447 repl=1

Status: HEALTHY
Total size:      9447 B
Total dirs:      0
Total files:      1
Total symlinks:      0
Total blocks (validated):      1 (avg. block size 9447 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks:      0 (0.0 %)
Under-replicated blocks:      1 (100.0 %)
Mis-replicated blocks:      0 (0.0 %)
Default replication factor:      3
Average block replication:      1.0
Corrupt blocks:      0
Missing replicas:      2 (66.666664 %)
Number of data-nodes:      1
Number of racks:      1
FSCK ended at Thu Aug 30 21:50:04 UTC 2018 in 4 milliseconds

The filesystem under path '/Installer/my_install.log' is HEALTHY
```

- Block Location of file

**hdfs dfs fsck /<hdfs\_file\_path> -files -blocks**

For example: hdfs fsck /Installer/my\_install.log -files -blocks -locations

```
[root@sandbox ~]# hdfs fsck /Installer/my_install.log -files -blocks -locations
Connecting to namenode via http://sandbox.hortonworks.com:50070/fsck?ugi=root&files=1&blocks=1&locations=1&path=%2FInstaller%2Fmy_install.log
FSCK started by root (auth:SIMPLE) from /192.168.183.153 for path /Installer/my_install.log at Fri Aug 31 07:36:00 UTC 2018
/Installer/my_install.log 9447 bytes, 1 block(s): Under replicated BP-267552868-172.16.137.143-1457691099567:blk_1073742645_1829. Target Replicas is 3 but found 1 live replica(s), 0 decommissioned replica(s) and 0 decommissioning replica(s).
0. BP-267552868-172.16.137.143-1457691099567:blk_1073742645_1829 len=9447 repl=1 [DatanodeInfoWithStorage[192.168.183.153:50010,DS-9267f0a7-e4cf-40e2-b113-c12e16fc4aa7,DISK]]
```

- Move file from NDFS to HDFS

**hdfs dfs -moveFromLocal <source\_ndfs\_file> <destination\_hdfs\_directory>**

For example: hdfs dfs -moveFromLocal ./my\_install.log /Installer

```
[root@sandbox ~]# hdfs dfs -moveFromLocal ./my_install.log /Installer
[root@sandbox ~]#
[root@sandbox ~]# hdfs dfs -ls /Installer
Found 1 items
-rw-r--r--  3 root hdfs      9448 2018-08-30 21:23 /Installer/my_install.log
```

- Copy file from NDFS to HDFS

**hdfs dfs -copyFromLocal <source\_ndfs\_file> <destination\_hdfs\_directory>**

For example: hdfs dfs -copyFromLocal ./my\_install.log /Installer

**NOTE:** Find the difference between -put, -movefromlocal, -copyfromlocal.

- Copy files within HDFS

**hdfs dfs -cp /<source\_path> /<destination\_path>**

For example: hdfs dfs -cp /Installer/my\_install.log /Class-2

```
[root@sandbox ~]# hdfs dfs -cp /Installer/my_install.log /Class-2
[root@sandbox ~]# hdfs dfs -ls /Class-2
Found 3 items
-rw-r--r--  3 admin hdfs      214 2018-07-13 14:55 /Class-2/full_data.csv
drwxr-xr-x  - root  hdfs        0 2018-08-30 21:12 /Class-2/hive
-rw-r--r--  3 root  hdfs     9447 2018-08-30 21:43 /Class-2/my_install.log
```



- Move files within HDFS

**hdfs dfs -mv /<source\_path> /<destination\_path>**

for example: hdfs dfs -mv /Installer/my\_install.log /Class-2

```
[root@sandbox ~]# hdfs dfs -mv /Installer/my_install.log /Class-2
[root@sandbox ~]# hdfs dfs -ls /Class-2
Found 3 items
-rw-r--r--   3 admin hdfs      214 2018-07-13 14:55 /Class-2/full_data.csv
drwxr-xr-x   - root  hdfs        0 2018-08-30 21:12 /Class-2/hive
-rw-r--r--   3 root  hdfs    9447 2018-08-30 21:37 /Class-2/my_install.log
```

- Check Space Utilization of Directory

**hdfs dfs -du /<directory\_path>**

For example: hdfs dfs -du /Class-2

```
[root@sandbox ~]# hdfs dfs -du /Class-2
214   /Class-2/full_data.csv
0     /Class-2/hive
9447  /Class-2/my_install.log
```

- Copy file from HDFS to NDFS

**hdfs dfs -copyToLocal <source\_hdfs\_file> <destination\_ndfs\_directory>**

For example: hdfs dfs -copyToLocal /Installer/my\_install.log

~/new\_dir/my\_copied\_installed.log

or

**hdfs dfs -get <source\_hdfs\_file> <destination\_ndfs\_directory>**

For example: hdfs dfs -get /Install/install.log /Install

```
[root@sandbox ~]# hdfs dfs -copyToLocal /Installer/my_install.log ~/new_dir/my_copied_installed.log
[root@sandbox ~]# cd new_dir/
[root@sandbox new_dir]# ls
my_copied_installed.log
[root@sandbox new_dir]# cd my_copied_installed.log
-bash: cd: my_copied_installed.log: Not a directory
[root@sandbox new_dir]# vi my_copied_installed.log
```

- Delete file from HDFS

**hdfs dfs -rm <file\_path>**

For example: hdfs dfs -rm /install.log

```
[root@sandbox ~]# hdfs dfs -rm /Class-2/my_install.log
18/08/31 07:28:24 INFO fs.TrashPolicyDefault: Namenode trash configura
tion: Deletion interval = 360 minutes, Emptier interval = 0 minutes.
Moved: 'hdfs://sandbox.hortonworks.com:8020/Class-2/my_install.log' to
trash at: hdfs://sandbox.hortonworks.com:8020/user/root/.Trash/Curren
t
```

- View last content of file from the end

**hdfs dfs -tail <file\_path>**

For example: hdfs dfs -tail /Install/install.log

```
[root@sandbox ~]# hdfs dfs -tail /Class-2/my_install.log
device-mapper-multipath-0.4.9-87.el6.x86_64
Installing cryptsetup-luks-1.2.0-11.el6.x86_64
Installing lvm2-2.02.118-2.el6.x86_64
Installing openssh-clients-5.3p1-111.el6.x86_64
Installing openssh-server-5.3p1-111.el6.x86_64
Installing mdadm-3.3.2-5.el6.x86_64
Installing dhclient-4.1.1-49.P1.el6.centos.x86_64
Installing iscsi-initiator-utils-6.2.0.873-14.el6.x86_64
Installing passwd-0.77-4.el6_2.2.x86_64
Installing authconfig-6.1.12-23.el6.x86_64
Installing grub-0.97-94.el6.x86_64
install-info: No such file or directory for /usr/share/info/grub.info.
gz
install-info: No such file or directory for /usr/share/info/multiboot.
info.gz
Installing efibootmgr-0.5.4-13.el6.x86_64
Installing sudo-1.8.6p3-19.el6.x86_64
Installing audit-2.3.7-5.el6.x86_64
Installing e2fsprogs-1.41.12-22.el6.x86_64
Installing xfsprogs-3.1.1-16.el6.x86_64
Installing acl-2.2.49-6.el6.x86_64
Installing attr-2.4.44-7.el6.x86_64
Installing bridge-utils-1.2-10.el6.x86_64
Installing rootfiles-8.1-6.1.el6.noarch
*** FINISHED INSTALLING PACKAGES ***[root@sandbox ~]#
```

- Create Zero Size File

**hdfs dfs -touchz <file\_path>**

For example: hdfs dfs -touchz /install/install2.log

- Check the number of directories and files as well as the size of the directory

**hdfs dfs -count <directory\_path>**

For example: hdfs dfs -count /install

```
[root@sandbox ~]# hdfs dfs -count /Class-2
      2          2          9661 /Class-2
```

The output will show the number of sub-directories in that directory, the number of files in that directory, size of the directory and name of the directory in that order