

DSCI551- Lab 2

1. Command “-mkdir” & “-ls”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ip-172-31-9-171.us-west-1.compute.internal]
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -mkdir /user
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -mkdir /user/ec2-user
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /
Found 1 items
drwxr-xr-x  - ec2-user supergroup          0 2022-02-08 01:42 /user
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user
Found 1 items
drwxr-xr-x  - ec2-user supergroup          0 2022-02-08 01:42 /user/ec2-user
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

Start hdfs

Create /user and /user/ec2-user

- Brief Explanation:

First, I start running hdfs. Then, I run “-ls” to check there is not any folder under root fold. Next, I run “-mkdir” to create two folders “/user” and “user/ec2-user”. Finally, I run “-ls” to check every folder has been created successfully.

2. Command “-put”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -put ~/ /user/ec2-user/test
.bash_history      Chaoyu_Li_churn.py  file1.txt          .ssh/
.bash_logout       Chaoyu_Li_load.py   hadoop-3.3.1/      .viminfo
.bash_profile      chaoyuli-resume.pages  hadoop-3.3.1.tar.gz
.bashrc            Chaoyu_Li_tenure.py  .local/
.cache/            dsci551/             .python_history
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -put ~/file1.txt /user/ec2-user/test
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test
Found 1 items
-rw-r--r--  1 ec2-user supergroup          0 2022-02-08 01:50 /user/ec2-user/test/file1.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

I press “tab” to check files under ~

- Brief Explanation:

First, I press “tab” to show the files under the folder “~”. Then, I run “-put” to copy “file1.txt” from the local system to hdfs. Finally, I run “-ls” to check there is “file1.txt” under “/user/ec2-user/test”.

3. Command “-get”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ ls
bin etc include lib libexec LICENSE-binary licenses-binary LICENSE.txt logs NOTICE-binary NOTICE.txt README.txt sbin share
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -get /user/ec2-user/test test1
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test
Found 4 items
-rw-r--r-- 1 ec2-user supergroup      697 2022-02-08 01:52 /user/ec2-user/test/Chaoyu_Li_churn.py
-rw-r--r-- 1 ec2-user supergroup    1923 2022-02-08 01:52 /user/ec2-user/test/Chaoyu_Li_load.py
-rw-r--r-- 1 ec2-user supergroup     461 2022-02-08 01:52 /user/ec2-user/test/Chaoyu_Li_tenure.py
-rw-r--r-- 1 ec2-user supergroup       0 2022-02-08 01:50 /user/ec2-user/test/file1.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ ls
bin etc include lib libexec LICENSE-binary licenses-binary LICENSE.txt logs NOTICE-binary NOTICE.txt README.txt sbin share test1
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ ls test1
Chaoyu_Li_churn.py Chaoyu_Li_load.py Chaoyu_Li_tenure.py file1.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

The files under folder "/user/ec2-user/test"

The files under folder "~/hadoop-3.3.1/test1"

- Brief Explanation:

First, I run “ls” to show the files under “~/hadoop-3.3.1/”. Then, I run “-get” to copy folder “test” from hdfs to “~/hadoop-3.3.1/”. Finally, I run “ls” to show there is a folder named “test1” under “~/hadoop-3.3.1” and it contains all files in “test”.

4. Command “-cat”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -cat /user/ec2-user/test/file1.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

- Brief Explanation:

There is not any output means that it is on success.

5. Command “-cp”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test2
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -cp /user/ec2-user/test/file1.txt /user/ec2-user/test2/file.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test2
Found 1 items
-rw-r--r-- 1 ec2-user supergroup       0 2022-02-08 02:06 /user/ec2-user/test2/file.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

- Brief Explanation:

First, I run “-ls” to check that “test2” is empty. Then, I run “-cp” to copy file “file1.txt” from “/test/test1” to “/test/test2”. Finally, I run “-ls” again to show there is a file named “file.txt” under folder “test2”.

6. Command “-rm”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test2
Found 1 items
-rw-r--r-- 1 ec2-user supergroup 0 2022-02-08 02:06 /user/ec2-user/test2/file.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -rm /user/ec2-user/test2/file.txt
Deleted /user/ec2-user/test2/file.txt
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user/test2
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

- Brief Explanation:

First, I run “-ls” to check that “test2” has a file named “file.txt”. Then, I run “-rm” to delete “file.txt”. Finally, I run “-ls” again to show “test2” become empty.

7. Command “-rmdir”

- Screenshot:

```
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user
Found 2 items
drwxr-xr-x - ec2-user supergroup 0 2022-02-08 01:52 /user/ec2-user/test
drwxr-xr-x - ec2-user supergroup 0 2022-02-08 02:08 /user/ec2-user/test2
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -rmdir /user/ec2-user/test2
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$ bin/hdfs dfs -ls /user/ec2-user
Found 1 items
drwxr-xr-x - ec2-user supergroup 0 2022-02-08 01:52 /user/ec2-user/test
[ec2-user@ip-172-31-9-171 hadoop-3.3.1]$
```

- Brief Explanation:

First, I run “-ls” to check that “/user/ec2-user” has a sub-folder named “test2”. Then, I run “-rmdir” to delete “/user/ec2-user/test2”. Finally, I run “-ls” again to show “test2” has been deleted.