Lab 2: HDFS

-ls

List the content of the root directory

```
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - ec2-user supergroup          0 2022-02-05 03:46 /user
```

-mkdir

Create a new directory called “lab2” under the root directory

```
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -mkdir /lab2
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - ec2-user supergroup          0 2022-02-12 00:30 /lab2
drwxr-xr-x   - ec2-user supergroup          0 2022-02-05 03:46 /user
```

Create a new directory called “test” under the “lab2” directory

```
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -mkdir /lab2/test
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -ls /lab2
Found 1 items
drwxr-xr-x   - ec2-user supergroup          0 2022-02-12 00:31 /lab2/test
```

-rmdir

Remove the “test” directory from the “lab2” directory

```
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -rmdir /lab2/test
[ec2-user@ip-172-31-1-43 hadoop-3.3.1]$ hdfs dfs -ls /lab2
```

-rm (done after -put)

Remove the “capacity-scheduler.xml” file from the “lab2” directory

```
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -rm /lab2/capacity-scheduler.xml
Deleted /lab2/capacity-scheduler.xml
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -ls /lab2
Found 1 items
-rw-r--r--   1 ec2-user supergroup         12 2022-02-12 00:33 /lab2/hello.txt
```

-put

Upload the file called “hello.txt” (created in Lab 1 and stored under the EC2 /dsci551 directory) to the “lab2” directory

```
[ec2-user@ip-172-31-1-43 ~]$ ls
dsci551  hadoop-3.3.1
[ec2-user@ip-172-31-1-43 ~]$ cd dsci551
[ec2-user@ip-172-31-1-43 dsci551]$ ls
hello.txt
[ec2-user@ip-172-31-1-43 dsci551]$ hdfs dfs -put hello.txt /lab2
[ec2-user@ip-172-31-1-43 dsci551]$ hdfs dfs -ls /lab2
Found 1 items
-rw-r--r--   1 ec2-user supergroup         12 2022-02-12 00:33 /lab2/hello.txt
```

Upload the file called “capacity-scheduler.xml” (stored under the EC2 /hadoop-3.3.1/etc/hadoop directory) to the “lab2” directory

```
[ec2-user@ip-172-31-1-43 ~]$ cd hadoop-3.3.1/etc/hadoop
[ec2-user@ip-172-31-1-43 hadoop]$ ls
capacity-scheduler.xml  hadoop-env.sh                    hdfs-site.xml           kms-env.s
configuration.xsl       hadoop-metrics2.properties       httpfs-env.sh           kms-log4j
container-executor.cfg  hadoop-policy.xml                httpfs-log4j.properties kms-site.
core-site.xml           hadoop-user-functions.sh.example httpfs-site.xml         log4j.pro
hadoop-env.cmd          hdfs-rbf-site.xml                kms-acls.xml            mapred-en
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -put capacity-scheduler.xml /lab2
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -ls /lab2
Found 2 items
-rw-r--r--   1 ec2-user supergroup       9213 2022-02-12 00:34 /lab2/capacity-scheduler.xml
-rw-r--r--   1 ec2-user supergroup         12 2022-02-12 00:33 /lab2/hello.txt
```

-get

Download the “hello.txt” file from the “lab2” directory to the EC2 /hadoop-3.3.1/etc/hadoop directory

```
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -get /lab2/hello.txt
[ec2-user@ip-172-31-1-43 hadoop]$ ls
capacity-scheduler.xml  hadoop-env.sh                    hdfs-site.xml
configuration.xsl       hadoop-metrics2.properties       hello.txt
container-executor.cfg  hadoop-policy.xml                httpfs-env.sh
core-site.xml           hadoop-user-functions.sh.example httpfs-log4j.properties
hadoop-env.cmd          hdfs-rbf-site.xml                httpfs-site.xml
```

-cp

Copy the “hello.txt” file from the “lab2” directory to the “user” directory as “copied_hello.txt”

```
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -cp /lab2/hello.txt /user/copied_hello.txt
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -ls /user
Found 1 items
-rw-r--r--   1 ec2-user supergroup         12 2022-02-12 00:40 /user/copied_hello.txt
```

-cat

Print the content of the “copied_hello.txt” file

```
[ec2-user@ip-172-31-1-43 hadoop]$ hdfs dfs -cat /user/copied_hello.txt
hello world
```