

Lab 2: HDFS
DSCI 551 – Fall 2024

Q1.

set to the root of our Java installation in HW1

```
export JAVA_HOME=/usr/lib/jvm/default-java
```

specify the default file system as our local file system

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

Set the number of copies of data as 1

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

Generate key-pair and set ssh to localhost

```
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
```

Q2. Use screenshots showing that you can successfully execute the following HDFS shell commands:

ls & mkdir:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x  - ubuntu supergroup          0 2024-09-28 06:30 /user
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

put:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -put README.txt /user
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x  - ubuntu supergroup          0 2024-09-28 06:36 /user
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /user/
Found 1 items
-rw-r--r--  1 ubuntu supergroup        175 2024-09-28 06:36 /user/README.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

cat:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -cat /user/README.txt
For the latest information about Hadoop, please visit our website at:

  http://hadoop.apache.org/

and our wiki, at:

  https://cwiki.apache.org/confluence/display/HADOOP/
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

get:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -get /user/README.txt README_getF
romHDFS.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ ls
LICENSE-binary  NOTICE.txt      bin             lib             logs
LICENSE.txt     README.txt       etc             libexec         sbin
NOTICE-binary   README_getFromHDFS.txt  include         licenses-binary share
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

stat:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -stat /user/
2024-09-28 06:36:44
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -stat "type:%F perm:%a %u:%g size
:%b mtime:%y atime:%x name:%n" /user/
type:directory perm:755 ubuntu:supergroup size:0 mtime:2024-09-28 06:36:44 atime
:1970-01-01 00:00:00 name:user
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

cp:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -cp /user/README.txt /user/README
_Copied.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /user/
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 175 2024-09-28 06:36 /user/README.txt
-rw-r--r-- 1 ubuntu supergroup 175 2024-09-28 06:43 /user/README_Copied
.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

mv:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -mv /user/README.txt /user/README
_Moved.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /user/
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 175 2024-09-28 06:43 /user/README_Copied
.txt
-rw-r--r-- 1 ubuntu supergroup 175 2024-09-28 06:36 /user/README_Moved.
txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

rm:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -rm /user/README*
Deleted /user/README_Copied.txt
Deleted /user/README_Moved.txt
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /user/
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```

rmdir:

```
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -rmdir /user
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$ hdfs dfs -ls /
ubuntu@ip-172-31-7-22:~/hadoop-3.4.0$
```