03 HDFS Practice

$ hadoop fs -ls / // issue this command after Hadoop setup

Column: **1** **2** **3** **4** **5** **6** **7** **8**

drwxr-xr-x 1 itadmin supergroup 0 2017-05-30 21:34 /input
-rw-r--r-- 3 itadmin supergroup 23170 2017-05-31 14:54 /largedeck.txt

Column 1: file mode -> file permissions

- `-` is set if not applicable
- d = directory
- `-` file
- Next three characters are permissions to the user.
- next three specify permission to the users in the same group
- the last three letters signify the access permission of non-group users.
- Three characters denote the access permission: read (r), write (w), executable (x).
- If anything is not applicable, hyphen (`-`) symbol is used

Column 2: indicates RF

- Not applicable for dirs

Column 3: username (owner).

Column 4: group name.

Column 5: file size in bytes, which is zero for directories.

Column 6 and 7: last modified date and time.

Column 8: file/directory name.

1. **List all the files and directories under the /user/nick/input directory in HDFS.**

```
hdfs dfs -mkdir /user/nick
hdfs dfs -mkdir /user/nick/input
hdfs dfs -ls /user/nick/input
```

2. **Create a directory named "hdfs_assignment" in HDFS under the /user/nick directory.**

```
hdfs dfs -mkdir /user/nick/hdfs_assignment
```

3. **Copy the file "data.txt" from the local file system to the "hdfs_assignment" directory in HDFS.**

```
 curl https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBM-BD0225EN-SkillsNetwork/labs/data/data.txt --
output data.txt

hdfs dfs -put /data.txt /user/nick/hdfs_assignment
hdfs dfs -put /data.txt /user/nick/hdfs_assignment
```

4. **Display the contents of the "data.txt" file located in the "hdfs_assignment" directory in HDFS.**

```
hdfs dfs -cat /user/nick/hdfs_assignment/data.txt
```

5. **Create a directory named "hdfs_backup" in HDFS under the /user/nick directory.**

```
hdfs dfs -mkdir /user/nick/hdfs_backup
```

6. **Copy the entire "hdfs_assignment" directory to the "hdfs_backup" directory in HDFS.**

```
hdfs dfs -cp /user/nick/hdfs_assignment /user/nick/hdfs_backup
```

7. **Check the replication factor of the "data.txt" file in the "hdfs_assignment" directory.**

```
hdfs dfs -stat %r /user/nick/hdfs_assignment/data.txt
```

8. **Set the replication factor of the "data.txt" file in the "hdfs_assignment" directory to 2.**

```
hdfs dfs -setrep 2 /user/nick/hdfs_assignment/data.txt
```

9. **Retrieve the "data.txt" file from the HDFS "hdfs_assignment" directory to the local file system.**

```
hdfs dfs -get /user/nick/hdfs_assignment/data.txt /nik
```

10. **Check the health status of the "hdfs_assignment" directory in HDFS.**

```
hdfs fsck /user/nick/hdfs_assignment -files -blocks
```

`fsck` = file system check

11. **Set the recursive permission of the "hdfs_assignment" directory to 777 (give rwx access to all users).**

```
hdfs dfs -chmod -R 777 /user/nick/hdfs_assignment
```

12. **Delete the "hdfs_assignment" and "hdfs_backup" directories from HDFS.**

```
hdfs dfs -rm -r /user/nick/hdfs_assignment /user/nick/hdfs_backup
```