

## HDFS commands

### Task 1: Create a directory on HDFS

Started services and created 'assignment\_data' directory

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.22621.2283]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd..

C:\Windows>cd..

C:\>cd hadoop-3.2.4

C:\hadoop-3.2.4>cd sbin

C:\hadoop-3.2.4\sbin>start-dfs.cmd

C:\hadoop-3.2.4\sbin>start-yarn.cmd
starting yarn daemons

C:\hadoop-3.2.4\sbin>hdfs dfs -mkdir /assignment_data

C:\hadoop-3.2.4\sbin>
```

Created one more directory named bda\_assignment\_4

```
C:\hadoop-3.2.4\sbin>hdfs dfs -mkdir /bda_assignment_4

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /
Found 2 items
drwxr-xr-x  - HP supergroup          0 2023-10-08 15:43 /assignment_data
drwxr-xr-x  - HP supergroup          0 2023-10-08 15:50 /bda_assignment_4
```

### Task 2: Upload files to HDFS

Uploading file.txt, file.csv, file3.json from local machine to directory assignment\_data

```
C:\hadoop-3.2.4\sbin>hdfs dfs -mkdir /bda_assignment_4

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /
Found 2 items
drwxr-xr-x  - HP supergroup          0 2023-10-08 15:43 /assignment_data
drwxr-xr-x  - HP supergroup          0 2023-10-08 15:50 /bda_assignment_4

C:\hadoop-3.2.4\sbin>hdfs dfs -copyFromLocal "D:\MSCS Course\1.3\BDA\Assignments\Assignment 4\files\file1.txt" /assignment_data/

C:\hadoop-3.2.4\sbin>hdfs dfs -copyFromLocal "D:\MSCS Course\1.3\BDA\Assignments\Assignment 4\files\file2.csv" /assignment_data/

C:\hadoop-3.2.4\sbin>
C:\hadoop-3.2.4\sbin>hdfs dfs -copyFromLocal "D:\MSCS Course\1.3\BDA\Assignments\Assignment 4\files\file3.json" /assignment_data/
```

### Task 3: List files in HDFS

Listing files in assignment\_data directory

```
C:\hadoop-3.2.4\sbin>hdfs dfs -ls /assignment_data
Found 3 items
-rw-r--r--   1 HP supergroup      21 2023-10-08 15:56 /assignment_data/file1.txt
-rw-r--r--   1 HP supergroup      32 2023-10-08 15:56 /assignment_data/file2.csv
-rw-r--r--   1 HP supergroup      46 2023-10-08 15:56 /assignment_data/file3.json
C:\hadoop-3.2.4\sbin>
```

#### Task 4: View file content in HDFS

Viewing contents of file1.txt using cat

```
C:\hadoop-3.2.4\sbin>hdfs dfs -cat /assignment_data/file1.txt
Big data applications
C:\hadoop-3.2.4\sbin>
```

#### Task 5: Create a new directory in HDFS

Creating a sub directory 'docs' in assignment\_data directory and checking if it is created

```
C:\hadoop-3.2.4\sbin>hdfs dfs -mkdir /assignment_data/docs

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /assignment_data
Found 4 items
drwxr-xr-x   - HP supergroup      0 2023-10-08 15:59 /assignment_data/docs
-rw-r--r--   1 HP supergroup      21 2023-10-08 15:56 /assignment_data/file1.txt
-rw-r--r--   1 HP supergroup      32 2023-10-08 15:56 /assignment_data/file2.csv
-rw-r--r--   1 HP supergroup      46 2023-10-08 15:56 /assignment_data/file3.json
C:\hadoop-3.2.4\sbin>
```

#### Task 6: Move files to a different directory in HDFS

Moving file2.csv and file3.json to a docs directory and checking if those files are moved.

```
C:\hadoop-3.2.4\sbin>hdfs dfs -mv /assignment_data/file2.csv /assignment_data/docs/

C:\hadoop-3.2.4\sbin>hdfs dfs -mv /assignment_data/file3.json /assignment_data/docs/

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /assignment_data
Found 2 items
drwxr-xr-x   - HP supergroup      0 2023-10-08 16:01 /assignment_data/docs
-rw-r--r--   1 HP supergroup      21 2023-10-08 15:56 /assignment_data/file1.txt

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /assignment_data/docs
Found 2 items
-rw-r--r--   1 HP supergroup      32 2023-10-08 15:56 /assignment_data/docs/file2.csv
-rw-r--r--   1 HP supergroup      46 2023-10-08 15:56 /assignment_data/docs/file3.json
```

#### Task 7: Delete files from HDFS

Deleting file1.txt from assignment\_data directory

```
C:\hadoop-3.2.4\sbin>hdfs dfs -rm /assignment_data/file1.txt
Deleted /assignment_data/file1.txt

C:\hadoop-3.2.4\sbin>hdfs dfs -ls /assignment_data
Found 1 items
drwxr-xr-x   - HP supergroup      0 2023-10-08 16:01 /assignment_data/docs

C:\hadoop-3.2.4\sbin>
```

## Task 8: Check HDFS file status

### HDFS status of file2.csv

```
Administrator: Command Prompt

C:\hadoop-3.2.4\hdfs fsck /assignment_data/docs/file2.csv -files -blocks -locations
Connecting to namenode via http://localhost:9870/fsck?ugi=HP&files=1&blocks=1&locations=1&path=%2Fassignment_data%2Fdocs%2Ffile2.csv
FSCK started by HP (auth:SIMPLE) from /127.0.0.1 for path /assignment_data/docs/file2.csv at Sun Oct 08 16:26:01 EDT 2023

/assignment_data/docs/file2.csv 32 bytes, replicated: replication=1, 1 block(s): OK
0. BP-1536519246-192.168.56.1-1696790261588:blk_1073741826_1002 len=32 Live_repl=1 [DatanodeInfoWithStorage[127.0.0.1:9866,DS-3716e37f-7b95-4c26-8b1e-3afa29acd61b,DISK]]

Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 0
Total symlinks: 0

Replicated Blocks:
Total size: 32 B
Total files: 1
Total blocks (validated): 1 (avg. block size 32 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0

14°C Partly sunny 16:26 08-10-2023
```

```
Replicated Blocks:
Total size: 32 B
Total files: 1
Total blocks (validated): 1 (avg. block size 32 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0

FSCK ended at Sun Oct 08 16:26:01 EDT 2023 in 10 milliseconds

The filesystem under path '/assignment_data/docs/file2.csv' is HEALTHY
```

## Task 9: Delete a directory from HDFS

Deleting directory from hdfs and checking if it is deleted.

```
C:\hadoop-3.2.4>
C:\hadoop-3.2.4>hdfs dfs -rm -r /assignment_data/docs
Deleted /assignment_data/docs

C:\hadoop-3.2.4>hdfs dfs -ls /assignment_data

C:\hadoop-3.2.4>
```

Now, we don't have docs directory. Also we can use below interface via localhost:9870 to view status.

The screenshot shows the Hadoop web interface at localhost:9870. The top navigation bar includes links for Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Browse Directory' and shows the root path '/'. A search bar and a 'Go!' button are present. Below the search bar, there is a table of files and directories. The table has columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. Two entries are listed: 'assignment\_data' and 'bda\_assignment\_4', both with a size of 0 B and a replication factor of 0. The interface also includes a 'Showing 1 to 2 of 2 entries' message and pagination controls.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	HP	supergroup	0 B	Oct 08 16:28	0	0 B	assignment_data
drwxr-xr-x	HP	supergroup	0 B	Oct 08 15:50	0	0 B	bda_assignment_4