

Practical – 2

- 1) List all the files and directories under /user/hirw/input in HDFS.

Command - **hadoop fs -ls /user/hirw/input**

```
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -ls /user/hirw/input
Found 12 items
drwxr-xr-x - hirw hirw 0 2017-01-07 19:59 /user/hirw/input/companies
drwxr-xr-x - hirw hirw 0 2017-01-07 19:59 /user/hirw/input/dividends
drwxr-xr-x - hirw hirw 0 2017-02-15 02:16 /user/hirw/input/employee-pig
drwxr-xr-x - hirw hirw 0 2017-01-07 19:58 /user/hirw/input/facebook
drwxr-xr-x - hirw hirw 0 2017-03-25 12:38 /user/hirw/input/hive
drwxr-xr-x - hirw hirw 0 2020-07-12 16:00 /user/hirw/input/kickstarter
drwxr-xr-x - hirw hirw 0 2020-07-12 16:02 /user/hirw/input/nyse
drwxr-xr-x - hirw hirw 0 2017-01-07 19:58 /user/hirw/input/pagerank
drwxr-xr-x - hirw hirw 0 2017-01-07 19:58 /user/hirw/input/songs
drwxr-xr-x - hirw hirw 0 2017-04-09 13:46 /user/hirw/input/stocks
drwxr-xr-x - hirw hirw 0 2017-01-07 19:58 /user/hirw/input/text-to-pdf
drwxr-xr-x - hirw hirw 0 2020-05-25 17:05 /user/hirw/input/twitter
hirwuser150430@ip-172-31-45-217:~$ ^C
hirwuser150430@ip-172-31-45-217:~$
```

- 2) Create a directory in HDFS named ilovebigdata on HDFS.

Command - **hadoop fs -mkdir ILOVEBIGDATA**

```
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -ls
Found 97 items
drwx----- - hirwuser150430 hirwuser150430 0 2021-06-09 09:31 .staging
drwxr-xr-x - hirwuser150430 hirwuser150430 0 2021-05-25 09:29 25052021training
-rw-r--r-- 3 hirwuser150430 hirwuser150430 3326129 2021-05-19 02:35 AlfaroA
drwxr-xr-x - hirwuser150430 hirwuser150430 0 2021-06-11 12:45 BIGDATA
drwxr-xr-x - hirwuser150430 hirwuser150430 0 2021-06-11 12:58 Big
drwxr-xr-x - hirwuser150430 hirwuser150430 0 2021-05-19 05:57 FuentesJ
drwxr-xr-x - hirwuser150430 hirwuser150430 0 2021-06-12 06:45 ILOVEBIGDATA
```

- 3) Copy the file dwp-payments-april10.csv from /hirw-workshop/input/hdfs to ilovebigdata directory in HDFS.

Command - **hadoop fs -copyFromLocal /hirw-starterkit/hdfs/commands/dwp-payments-april10.csv ILOVEBIGDATA**

```
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -mkdir ILOVEBIGDATA
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -copyFromLocal /hirw-starterkit/hdfs/commands/dwp-payments-april10.csv ILOVEBIGDATA
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -ls ILOVEBIGDATA
Found 1 items
-rw-r--r-- 3 hirwuser150430 hirwuser150430 3326129 2021-06-12 06:45 ILOVEBIGDATA/dwp-payments-april10.csv
```

- 4) Display the content of file dwp-payments-april10.csv under ilovebigdata in hdfs.

Command - **hadoop fs -cat /ILOVEBIGDATA/dwp-payments-april10.csv**

```
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -cat ILOVEBIGDATA/dwp-payments-april10.csv
Department for Work and Pensions ,Jobcentre Plus,30/04/2010,PRINTING STATIONERY IT & CONSUMABLES,JOBCENTRE PLUS,XEROX UK LIMITED,2015
125301,91.38,,
Department for Work and Pensions ,Jobcentre Plus,30/04/2010,PRINTING STATIONERY IT & CONSUMABLES,JOBCENTRE PLUS,XEROX UK LIMITED,2015
125301,14.17,,
Department for Work and Pensions ,Jobcentre Plus,30/04/2010,PRINTING STATIONERY IT & CONSUMABLES,JOBCENTRE PLUS,XEROX UK LIMITED,2015
125301,91.66,,
Department for Work and Pensions ,Jobcentre Plus,30/04/2010,PRINTING STATIONERY IT & CONSUMABLES,JOBCENTRE PLUS,XEROX UK LIMITED,2015
125301,15.07,,
Department for Work and Pensions ,Jobcentre Plus,30/04/2010,PRINTING STATIONERY IT & CONSUMABLES,JOBCENTRE PLUS,XEROX UK LIMITED,2015
125301,10.18,,
```

- 5) Create directory with name mynewdirectory in HDFS.

Command - **hadoop fs -mkdir /mynewdirectory**

```

hirwuser150430@ip-172-31-45-217:~$ hadoop fs -mkdir MYNEWDIRECTORY
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -ls
Found 98 items

```

```

drwxr-xr-x  - hirwuser150430 hirwuser150430      0 2021-06-12 06:53 MYNEWDIRECTORY

```

- 6) Copy dwp-payments-april10.csv from ilovebigdata to mynewdirectory in HDFS.

Command - **`hadoop fs -cp ILOVEBIGDATA/dwp-payments-april10.csv MYNEWDIRECTORY`**

```

hirwuser150430@ip-172-31-45-217:~$ hadoop fs -cp ILOVEBIGDATA/dwp-payments-april10.csv MYNEWDIRECTORY
hirwuser150430@ip-172-31-45-217:~$ hadoop fs -ls MYNEWDIRECTORY
Found 1 items
-rw-r--r--  3 hirwuser150430 hirwuser150430    3326129 2021-06-12 06:55 MYNEWDIRECTORY/dwp-payments-april10.csv

```

- 7) Set the replication factor of dwp-payments-april10.csv under mynewdirectory directory in HDFS to 2.

Command - **`hadoop fs -setrep 2 MYNEWDIRECTORY/dwp-payments-april10.csv`**

```

hirwuser150430@ip-172-31-45-217:~$ hadoop fs -setrep 2 MYNEWDIRECTORY/dwp-payments-april10.csv
Replication 2 set: MYNEWDIRECTORY/dwp-payments-april10.csv

```

- 8) check health of mynewdirectory in HDFS (chmod)(fsck) - **`hdfs fsck /user/hirwuser150430/MYNEWDIRECTORY -files -blocks -locations`**

```

hirwuser150430@ip-172-31-45-217:~$ hdfs fsck /user/hirwuser150430/MYNEWDIRECTORY -files -blocks -locations
Connecting to namenode via http://ec2-54-92-244-237.compute-1.amazonaws.com:50070
FSCK started by hirwuser150430 (auth:SIMPLE) from /172.31.45.217 for path /user/hirwuser150430/MYNEWDIRECTORY at
Sat Jun 12 07:23:13 UTC 2021
/user/hirwuser150430/MYNEWDIRECTORY <dir>
/user/hirwuser150430/MYNEWDIRECTORY/dwp-payments-april10.csv 3326129 bytes, 1 block(s): OK
0. BP-2125152513-172.31.45.216-1410037307133:blk_1075416868_1676131 len=3326129 Live_repl=2 [DatanodeInfoWithStorage[172.31.46.124:50010,DS-fe24aecb-f56f-4c9c-8cf9-a3b1259bc0d0,DISK], DatanodeInfoWithStorage[172.31.45.216:50010,DS-d0d9eb5c-f35f-4a12-bfdf-544085d693a3,DISK]]

Status: HEALTHY
Total size:      3326129 B
Total dirs:      1
Total files:      1
Total symlinks:    0
Total blocks (validated): 1 (avg. block size 3326129 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: 3
Average block replication: 2.0
Corrupt blocks:    0
Missing replicas:  0 (0.0 %)
Number of data-nodes: 3
Number of racks:   1
FSCK ended at Sat Jun 12 07:23:13 UTC 2021 in 1 milliseconds

The filesystem under path '/user/hirwuser150430/MYNEWDIRECTORY' is HEALTHY

```

- 9) Bring the file dwp-payments-april10.csv under ilovebigdata directory to the local file system.

Command - **`hadoop fs -copyToLocal ILOVEBIGDATA/dwp-payments-april10.csv`**

```

hirwuser150430@ip-172-31-45-217:~$ hadoop fs -copyToLocal ILOVEBIGDATA/dwp-payments-april10.csv
copyToLocal: /home/hirwuser150430/dwp-payments-april10.csv._COPYING_ (Permission denied)

```

Note - Since I don't have rights to make changes in the local system, here is the command and its output.