

ROLLNO:19UITE015

REGNO:920419205008

NAME:K.DEEPIKA

SUBJECT:IT8711-FOSS AND CLOUD COMPUTING LAB

EXERCISE NO:10

PROCEDURAL STEPS

Step 1: Create a text file “D:/data.txt”

Step 2: Create a directory in HDFS, where to kept text file.

hdfs dfs -mkdir /user



```
Administrator: Windows PowerShell
732 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup      26 2022-10-31 17:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
cloud
cloud
cloud
deebi
```

Step 3: Upload the data.txt file on HDFS in the specific directory

hdfs dfs -put D:/data.txt /user



```
Administrator: Windows PowerShell
732 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup      26 2022-10-31 17:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
cloud
cloud
cloud
deebi
```

Step 4: List the files or directories in hdfs

hdfs dfs -ls /user/



```
Administrator: Windows PowerShell
732 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup      26 2022-10-31 17:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
cloud
cloud
cloud
deebi
```

Step 5: To view the content of the file “/user/data.txt”

hdfs dfs -cat /user/data.txt



```
Administrator: Windows PowerShell
732 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup      26 2022-10-31 17:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
cloud
cloud
cloud
deebi
```

Step 6: Run the jar file

hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.2.jar wordcount /user /out

```
Administrator: Windows PowerShell
732 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:\data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup 26 2022-10-31 17:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
cloud
cloud
cloud
cloud
deep1
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.2.jar wordcount /user /out
Error: Could not find or load main class jarD:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.2.jar
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop jar D:\hadoop-env\hadoop-3.2.2\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.2.jar wordcount /user /out
2022-10-31 17:55:56,334 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-10-31 17:55:57,430 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Administrator/.staging/job_1667212332339_0001
2022-10-31 17:55:57,865 INFO input.FileInputFormat: Total input files to process : 1
2022-10-31 17:55:58,238 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-31 17:55:58,508 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667212332339_0001
2022-10-31 17:55:58,511 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-31 17:55:58,814 INFO conf.Configuration: resource-types.xml not found
2022-10-31 17:55:58,815 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-10-31 17:55:59,480 INFO impl.YarnClientImpl: Submitted application application_1667212332339_0001
2022-10-31 17:55:59,545 INFO mapreduce.Job: The url to track the job: http://adit66:8088/proxy/application_1667212332339_0001/
2022-10-31 17:55:59,547 INFO mapreduce.Job: Running job: job_1667212332339_0001
2022-10-31 17:56:17,930 INFO mapreduce.Job: Job job_1667212332339_0001 running in uber mode : false
2022-10-31 17:56:17,933 INFO mapreduce.Job: map 0% reduce 0%
2022-10-31 17:56:25,098 INFO mapreduce.Job: map 100% reduce 0%
2022-10-31 17:56:33,225 INFO mapreduce.Job: map 100% reduce 100%
2022-10-31 17:56:38,303 INFO mapreduce.Job: Job job_1667212332339_0001 completed successfully
2022-10-31 17:56:38,468 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=30
    FILE: Number of bytes written=471473
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=126
    HDFS: Number of bytes written=16
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=4790
    Total time spent by all reduces in occupied slots (ms)=4998
    Total time spent by all map tasks (ms)=4790
  Map-Reduce Framework
    Map input records=4
    Map output records=4
    Map output bytes=40
    Map output materialized bytes=30
    Input split bytes=100
    Combine input records=4
    Combine output records=2
    Reduce input groups=2
    Reduce shuffle bytes=30
    Reduce input records=2
    Reduce output records=2
    Spilled Records=4
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=146
    CPU time spent (ms)=1684
    Physical memory (bytes) snapshot=542150656
    Virtual memory (bytes) snapshot=778260480
    Total committed heap usage (bytes)=381157376
    Peak Map Physical memory (bytes)=340795392
    Peak Map Virtual memory (bytes)=482158544
    Peak Reduce Physical memory (bytes)=201355264
    Peak Reduce Virtual memory (bytes)=296103936
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=26
  File Output Format Counters
    Bytes Written=16
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop fs -cat /out/*
cloud 3
deep1 1
PS D:\hadoop-env\hadoop-3.2.2\sbin>
```

Step 7: To view the output in “/out/*”

hadoop fs -cat /out/*

```
Administrator: Windows PowerShell
Total vcore-milliseconds taken by all map tasks=4790
Total vcore-milliseconds taken by all reduce tasks=4998
Total megabyte-milliseconds taken by all map tasks=4904960
Total megabyte-milliseconds taken by all reduce tasks=5117952
  Map-Reduce Framework
    Map input records=4
    Map output records=4
    Map output bytes=40
    Map output materialized bytes=30
    Input split bytes=100
    Combine input records=4
    Combine output records=2
    Reduce input groups=2
    Reduce shuffle bytes=30
    Reduce input records=2
    Reduce output records=2
    Spilled Records=4
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=146
    CPU time spent (ms)=1684
    Physical memory (bytes) snapshot=542150656
    Virtual memory (bytes) snapshot=778260480
    Total committed heap usage (bytes)=381157376
    Peak Map Physical memory (bytes)=340795392
    Peak Map Virtual memory (bytes)=482158544
    Peak Reduce Physical memory (bytes)=201355264
    Peak Reduce Virtual memory (bytes)=296103936
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=26
  File Output Format Counters
    Bytes Written=16
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop fs -cat /out/*
cloud 3
deep1 1
PS D:\hadoop-env\hadoop-3.2.2\sbin>
```

Output

The screenshot shows a web browser window displaying the Hadoop DFS Health Overview page. The browser's address bar shows the URL `localhost:9870/dfshealth.html#tab-overview`. The page has a green header with navigation tabs: **Hadoop**, **Overview**, **Datanodes**, **Datanode Volume Failures**, **Snapshot**, and **Startup Progress**. The **Utilities** dropdown menu is open, showing options: **Browse the file system**, **Logs**, **Log Level**, **Metrics**, **Configuration**, and **Process Thread Dump**.

Overview 'localhost:9820' (active)

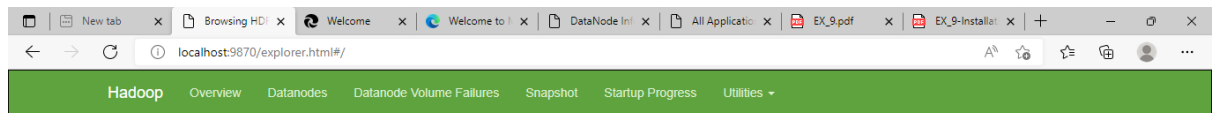
| | |
|-----------------------|--|
| Started: | Mon Oct 31 16:00:19 +0530 2022 |
| Version: | 3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932 |
| Compiled: | Sun Jan 03 14:56:00 +0530 2021 by hexiaoqiao from branch-3.2.2 |
| Cluster ID: | CID-9d314296-acea-4485-bbec-89b4e4407716 |
| Block Pool ID: | BP-1107299340-192.168.56.1-1667212058215 |

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 80.65 MB of 189.5 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 49.06 MB of 50.21 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Activate Windows
Go to Settings to activate Windows.

The Windows taskbar at the bottom shows the search bar, task view button, and several application icons including Edge, File Explorer, Mail, and various utilities. The system tray on the right shows the date and time as 17:59 on 31-10-2022.



Browse Directory

/

Show entries Search:

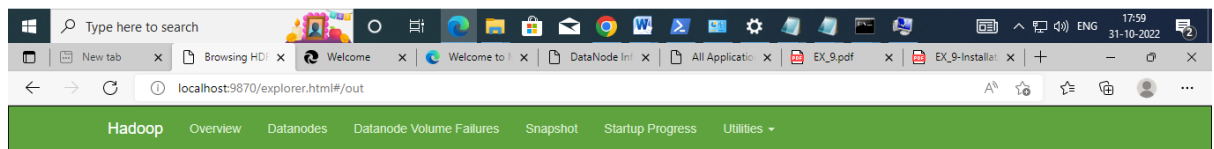
| <input type="checkbox"/> | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name | |
|--------------------------|------------|---------------|------------|------|---------------|-------------|------------|------|--|
| <input type="checkbox"/> | drwxr-xr-x | Administrator | supergroup | 0 B | Oct 31 17:56 | 0 | 0 B | out | |
| <input type="checkbox"/> | drwx----- | Administrator | supergroup | 0 B | Oct 31 17:55 | 0 | 0 B | tmp | |
| <input type="checkbox"/> | drwxr-xr-x | Administrator | supergroup | 0 B | Oct 31 17:53 | 0 | 0 B | user | |

Showing 1 to 3 of 3 entries

Previous **1** Next

Hadoop, 2021.

Activate Windows
Go to Settings to activate Windows.



Browse Directory

/out

Show entries Search:

| <input type="checkbox"/> | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name | |
|--------------------------|------------|---------------|------------|------|---------------|-------------|------------|--------------|--|
| <input type="checkbox"/> | -rw-r--r-- | Administrator | supergroup | 0 B | Oct 31 17:56 | 1 | 128 MB | _SUCCESS | |
| <input type="checkbox"/> | -rw-r--r-- | Administrator | supergroup | 16 B | Oct 31 17:56 | 1 | 128 MB | part-r-00000 | |

Showing 1 to 2 of 2 entries

Previous **1** Next

Hadoop, 2021.

Activate Windows
Go to Settings to activate Windows.

