Task No: 4 Date: 13/08/2025	Implement Map Reduce concept using apache Hadoop **Tools: Docker, Windows**	CO2

Task 4.1: Apache Hadoop Installation
Aim:
To download, install and configure the Apache Hadoop in windows operating system,

Procedure:

1. Download the software docker desktop from the url https://www.docker.com/products/docker-desktop/
2. Download the software GIT for windows from the url https://git-scm.com/download/win
3. Install the docker desktop and Git windows
4. Start the docker desktop engine and then open the command prompt
5. Clone the docker-hadoop using the git command

   git clone https://github.com/big-data-europe/docker-hadoop.git

6. Execute the following command to start the Hadoop sever using docker

   cd docker-hadoop/
   docker compose up

7. Enter into the bash mode and execute the hdfs commands
8. Execute the commands to check the containers, ip address and port number of the Hadoop server

   docker container ls

   ipconfig

9. Shut down the Hadoop server

   docker compose down

HDFS COMMANDS

Enter into the bash mode and execute the following commands

1. List files - ls

hdfs dfs -ls /

2. Make dir
hdfs dfs -mkdir /techcoreeasy

3. create empty file
hdfs dfs -touchz /techcoreeasy/test1.txt

4. cp file from local file system to hdfs
hdfs dfs -put tech.txt /techcoreeasy/

5. see contents - cat
hdfs dfs -cat /techcoreeasy/tech.txt

6. copy - cp
hdfs dfs -cp /techcoreeasy /test_cp

7. get file to local
hdfs dfs -get /techcoreeasy/test.txt .

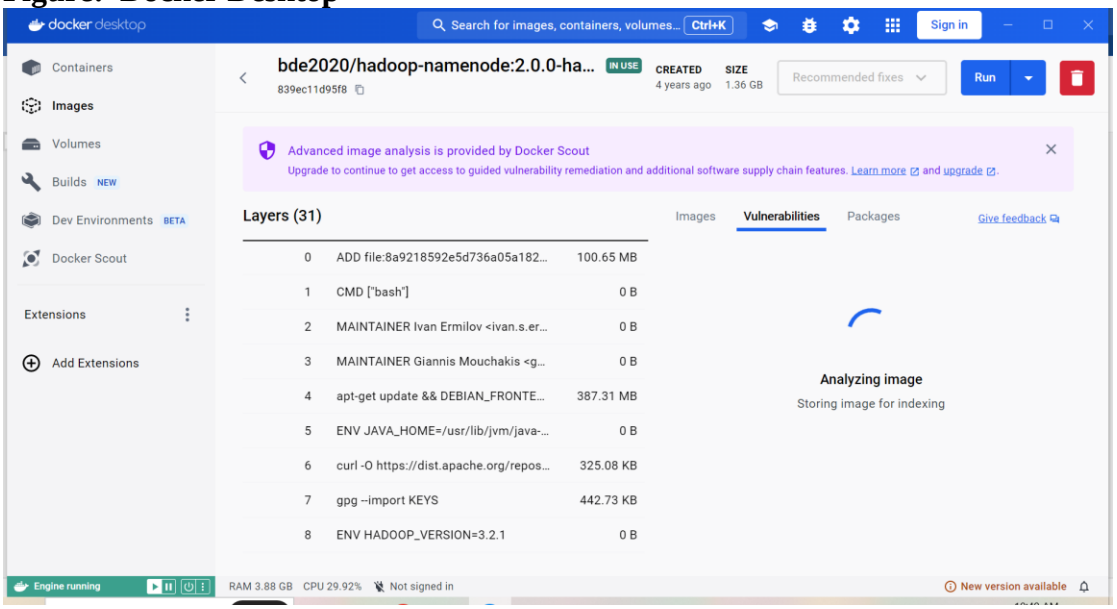
8. Remove file
hdfs dfs -rmr /techcoreeasy/test.txt

9. stst of a file
hdfs dfs -stat /path

10. exit command, switch from bash mode to command prompt

Output:

Figure: Docker Desktop



For docker compose up command

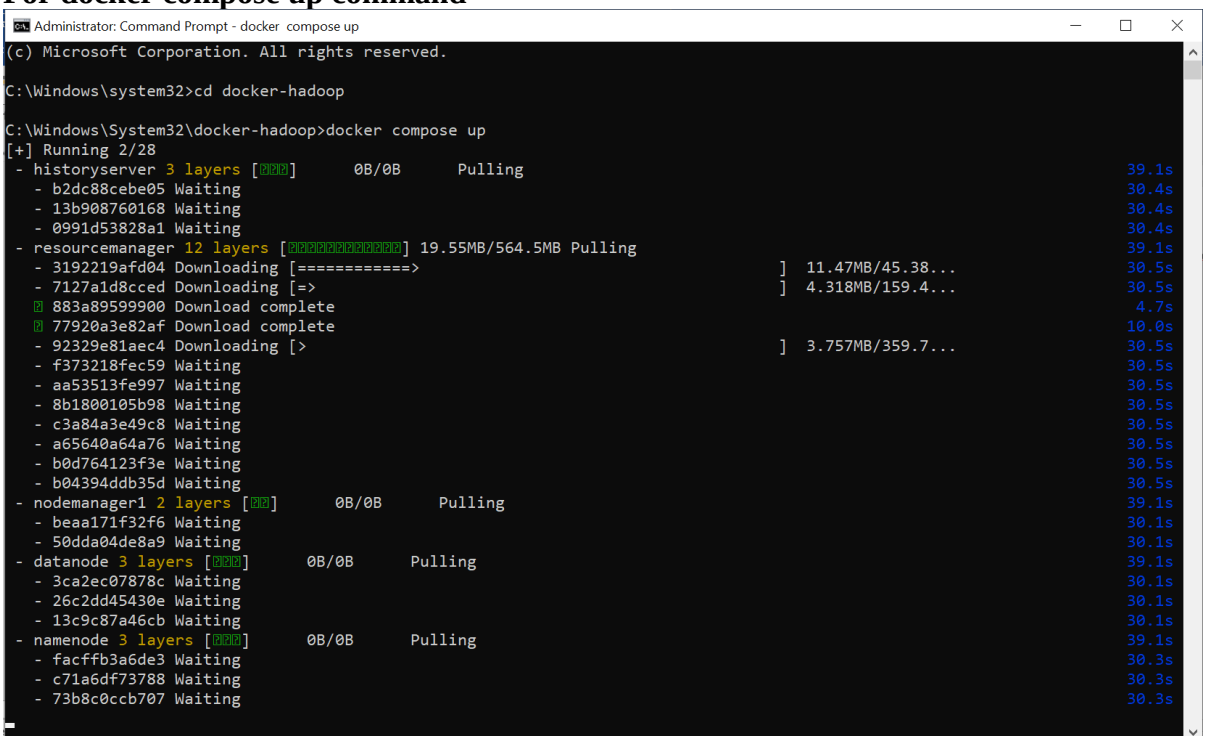
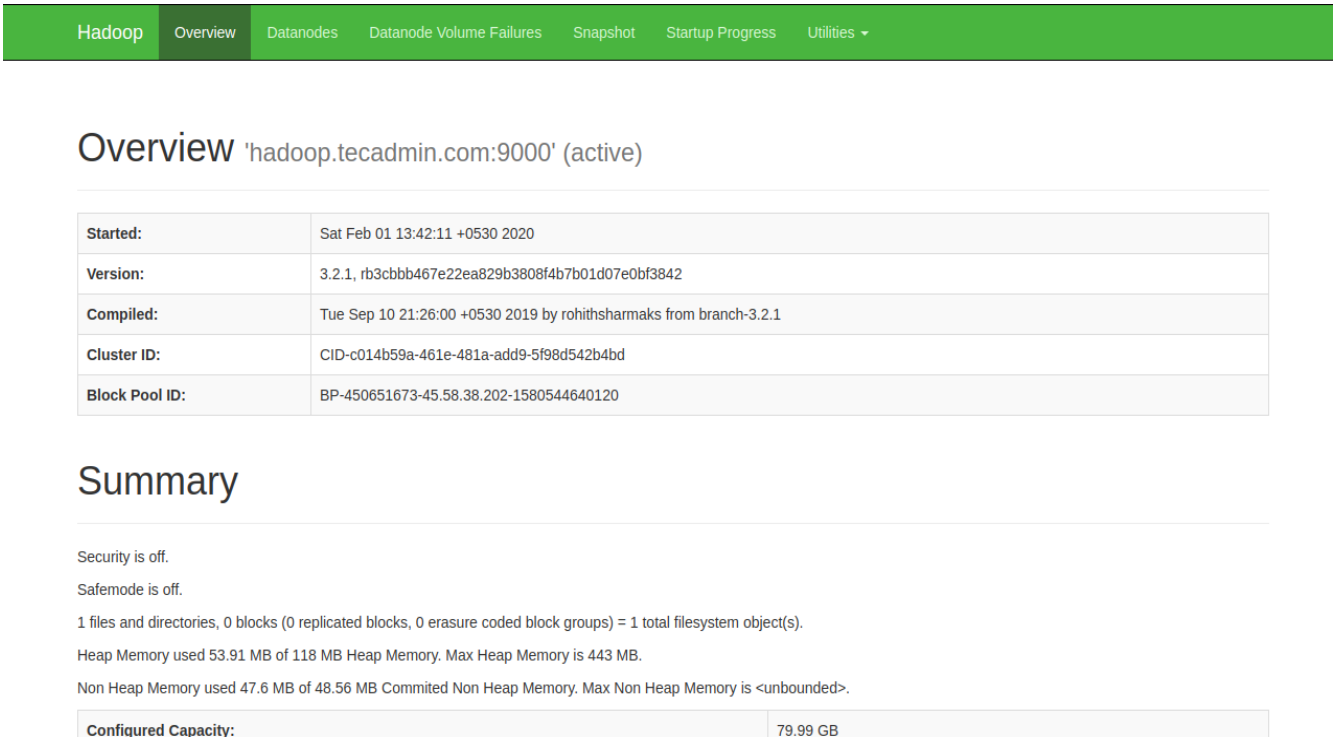
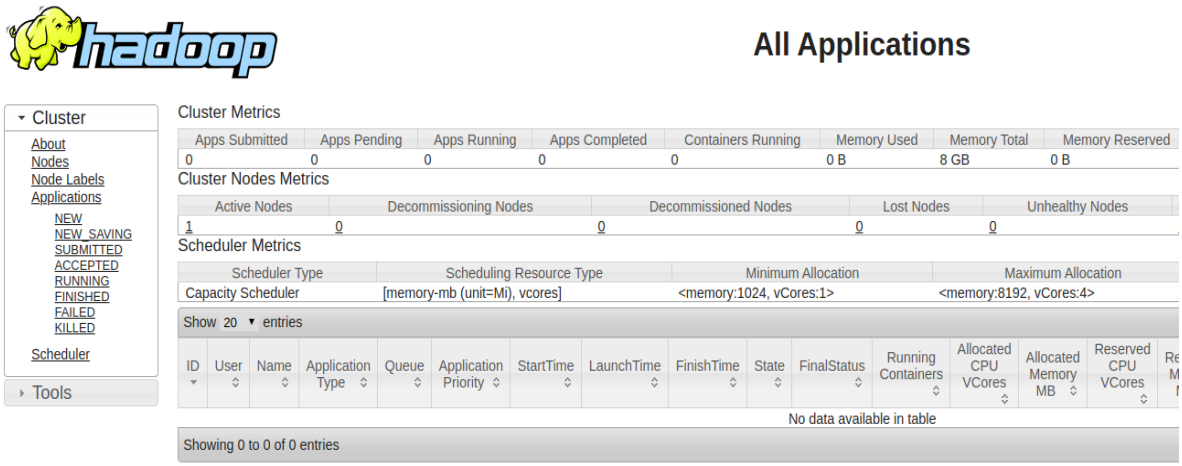


Figure: Start the Hadoop server

Open the browser and enter the url http://172.17.205.161:9870/ or http://your ip address:9870



Access the Resource Manager and Data Node



RESULT:
Thus the Apache Hadoop installation was successfully completed.

Task 4.2: Find the Word Count using Map reduce
Aim:
To implement the word count with Map Reduce Using Apache Hadoop

Procedure:

1. Start the Docker engine and Hadoop server
2. Check the docker containers using the command
   docker container ls
3. Download the jar file for the word count from the url
   https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-examples/2.7.1/hadoop-mapreduce-examples-2.7.1-sources.jar
4. Copy the jar file, input1.txt into the tmp folder and enter into the bash mode

   C:\Users\abc> docker cp hadoop-mapreduce-examples-2.7.1-sources.jar namenode:/tmp

   C:\Users\abc> docker cp input1.txt namenode:/tmp/

Commands

Execute the following commands in bash mode to execute the jar file

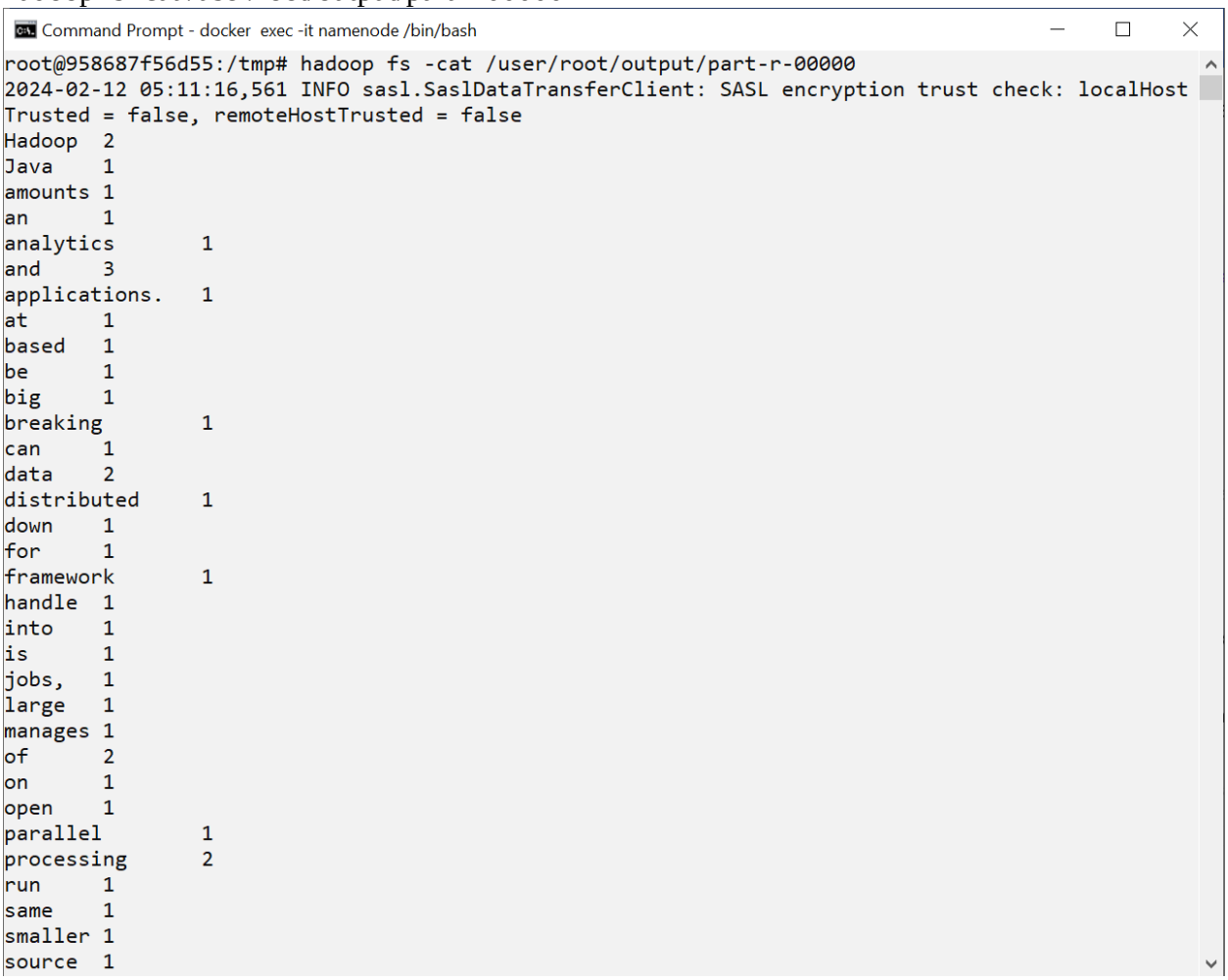
```
ls

cd /tmp/

cat input1.txt

hdfs dfs -mkdir /user

hdfs dfs -mkdir /user/root

hdfs dfs -mkdir /user/root/input

hdfs dfs -put input1.txt /user/root/input/

hdfs dfs -cat /user/root/input/input1.txt

hadoop jar hadoop-mapreduce-examples-2.7.1-sources.jar
org.apache.hadoop.examples.WordCount input output
```

Input File: input1.txt

Hadoop is an open source framework based on Java that manages the storage and processing of large amounts of data for applications. Hadoop uses distributed storage and parallel processing to handle big data and analytics jobs, breaking workloads down into smaller workloads that can be run at the same time.

Output:

Execute the command in the bash mode
hadoop fs -cat /user/root/output/part-r-00000



Result:

Thus the Map reduce concept was implement for word count using the Hadoop and Docker engine.