Experiment No.4
Experiment on Hadoop Map-Reduce
Date of Performance: 04/09/23
Date of Submission: 11/09/23

AIM: -To write a program to implement a word count program using MapReduce.

THEORY:

WordCount is a simple program which counts the number of occurrences of each word in a given text input data set. WordCount fits very well with the MapReduce programming model making it a great example to understand the Hadoop Map/Reduce programming style. The implementation consists of three main parts:

1. Mapper

2. Reducer

3. Driver

Step-1. Write a Mapper

A Mapper overrides the ―map‖ function from the Class "org.apache.hadoop.mapreduce.Mapper" which provides <key, value> pairs as the input. A Mapper implementation may output <key,value> pairs using the provided Context .

Input value of the WordCount Map task will be a line of text from the input data file and the key would be the line number <line_number, line_of_text> . Map task outputs <word, one> for each word in the line of text.

Pseudo-code

void Map (key, value){

for each word x in

value:

output.collect(x,1);

}

Step-2. Write a Reducer

A Reducer collects the intermediate <key,value> output from multiple map tasks and assemble a single result. Here, the WordCount program will sum up the occurrence of each word to pairs as <word, occurrence>.

Pseudo-code

void Reduce (keyword, <list of value>){ for

each x in <list of value>:

```
sum+=x;
final_output.collect(keyword, sum);
}
```

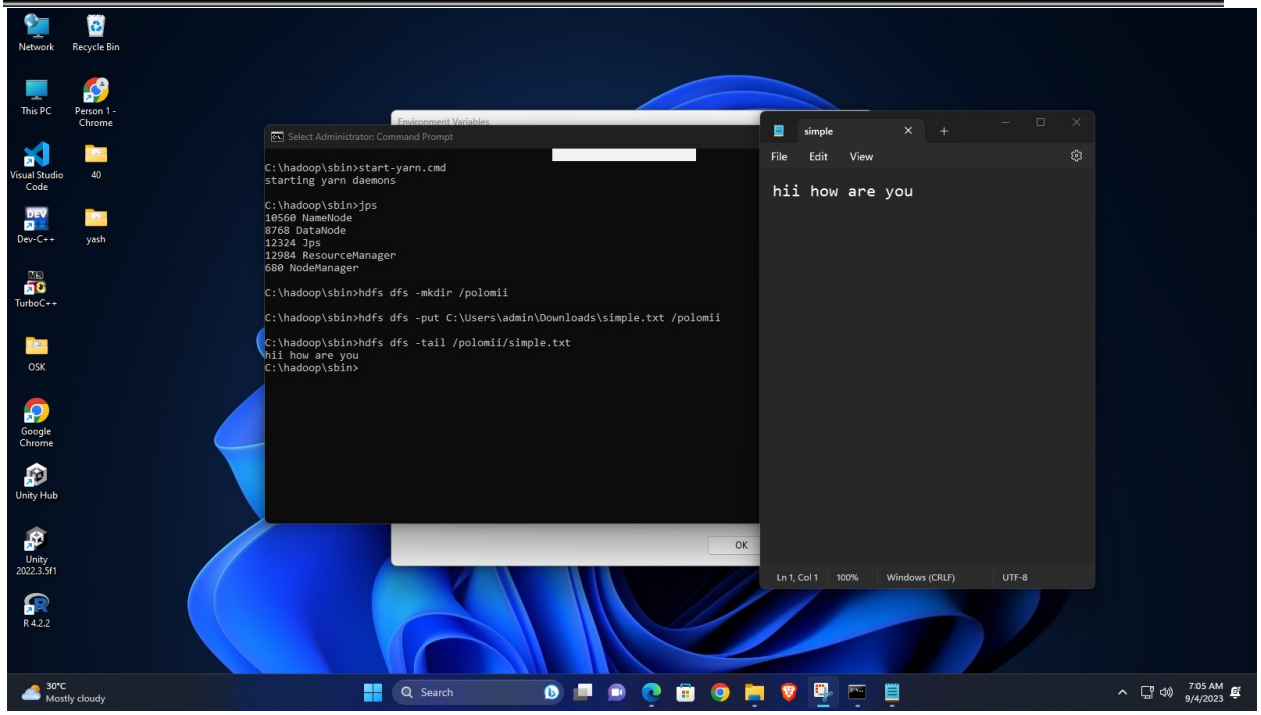
Code:

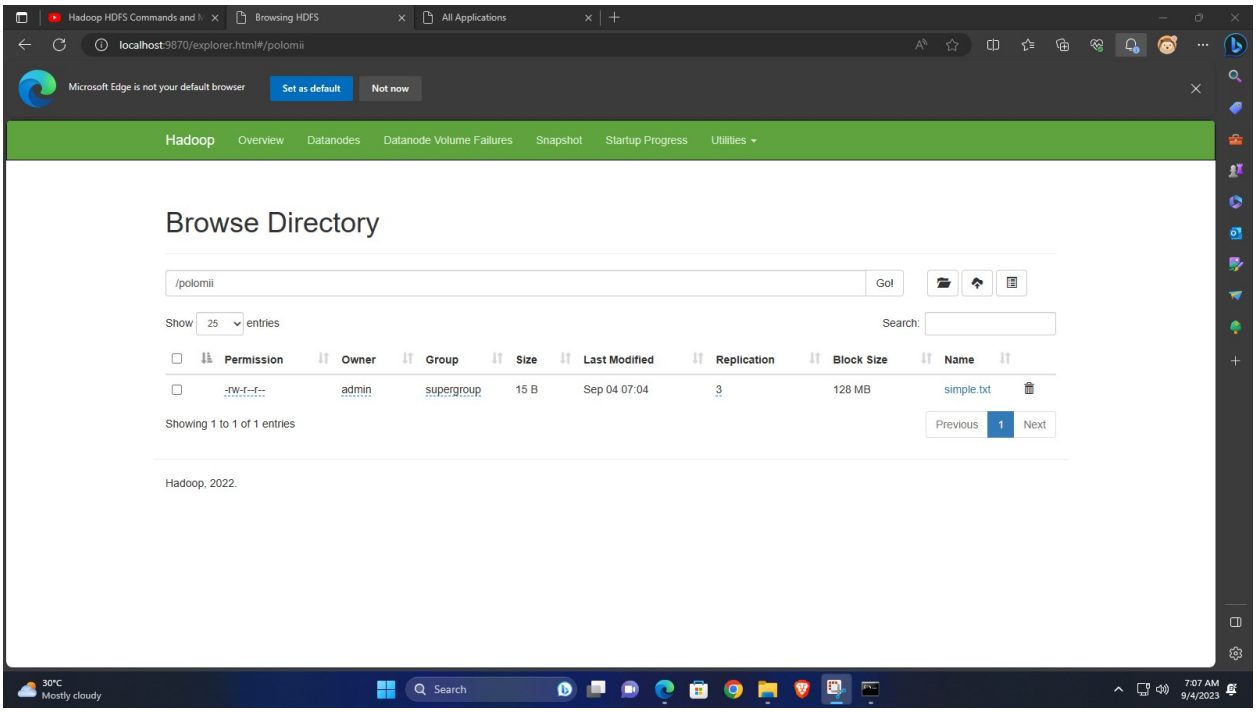
```
import java.io.IOException;
import
java.util.StringTokenizer;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.fs.Path;
public class WordCount
{
public static class Map extends Mapper<LongWritable,Text,Text,IntWritable> {
public void map(LongWritable key, Text value,Context context) throws
IOException,InterruptedException{
String line = value.toString();
StringTokenizer tokenizer = new
StringTokenizer(line); while
```

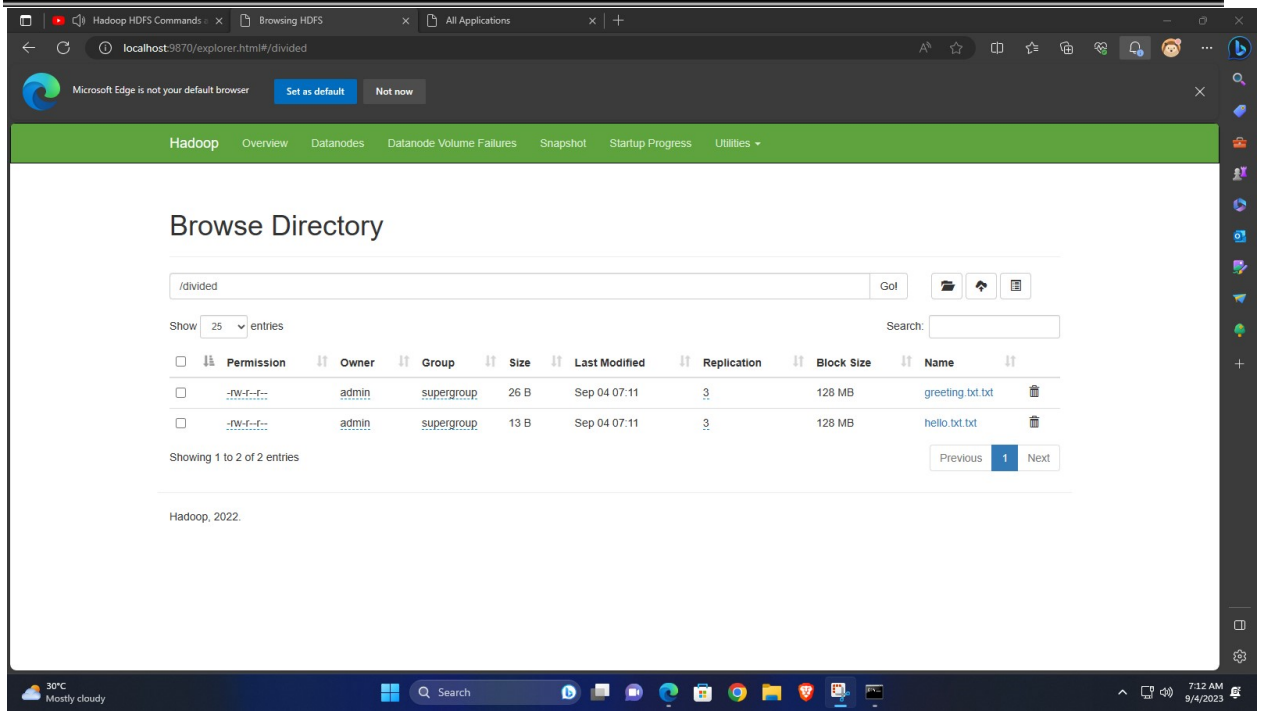
```
(tokenizer.hasMoreTokens()) {
value.set(tokenizer.nextToken());
context.write(value, new IntWritable(1));

}
}
}
public static class Reduce extends Reducer<Text,IntWritable,Text,IntWritable> {
public void reduce(Text key, Iterable<IntWritable> values,Context context)
throws IOException,InterruptedException {
int sum=0;
for(IntWritable x:
values)
{
sum+=x.get();
}
context.write(key, new IntWritable(sum));
}
}
public static void main(String[] args) throws Exception {
Configuration conf= new Configuration();
Job job = new Job(conf,"My Word Count Program");
job.setJarByClass(WordCount.class);
job.setMapperClass(Map.class);
job.setReducerClass(Reduce.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
```

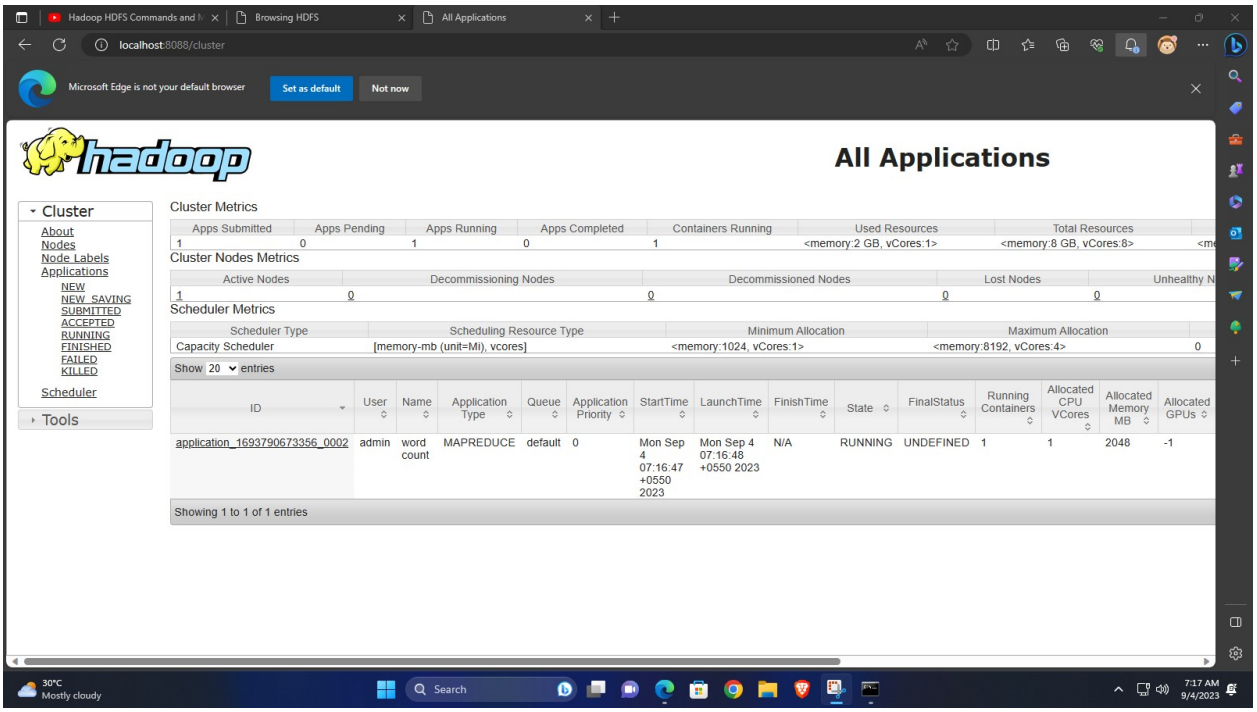
```
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
Path outputPath = new Path(args[1]);
//Configuring the input/output path from the filesystem into the job
FileInputFormat.addInputPath(job, new Path(args[0]));

FileOutputFormat.setOutputPath(job, new Path(args[1]));
//deleting the output path automatically from hdfs so that we don't have to
delete it explicitly
outputPath.getFileSystem(conf).delete(outputPath);
//exiting the job only if the flag value becomes
false System.exit(job.waitForCompletion(true) ? 0 :
1);
}
}
```

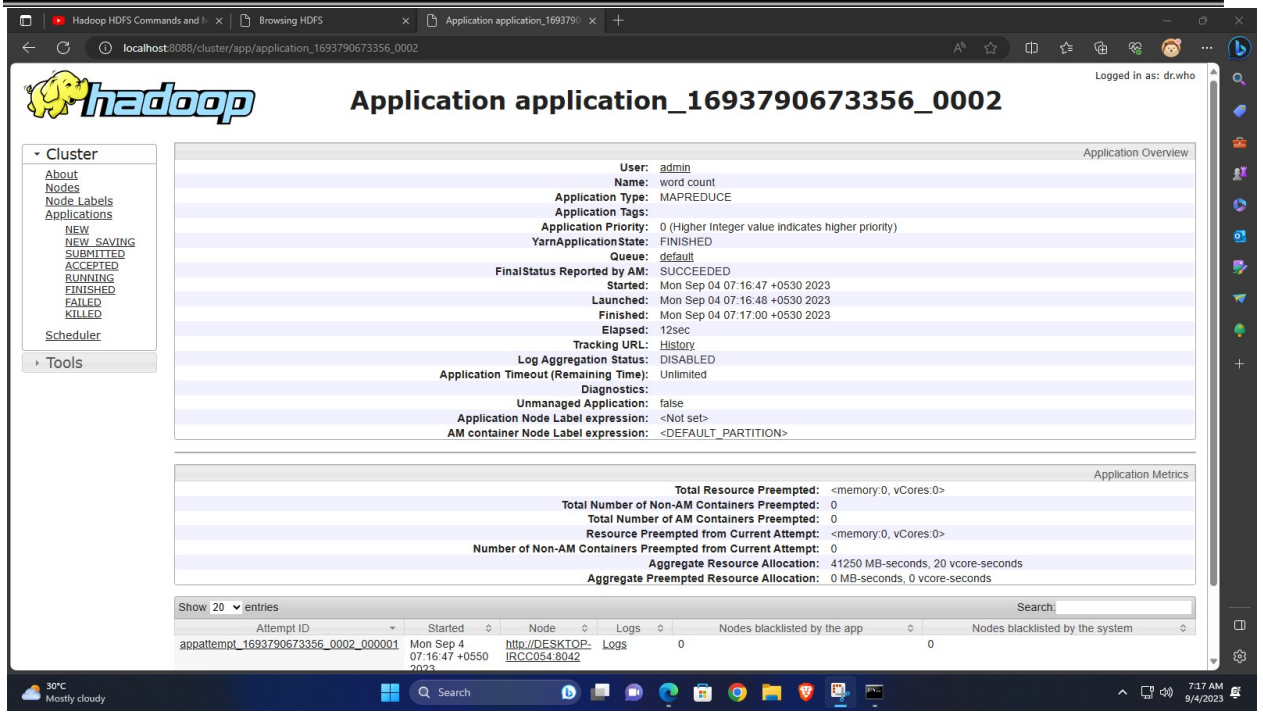
OUTPUT:

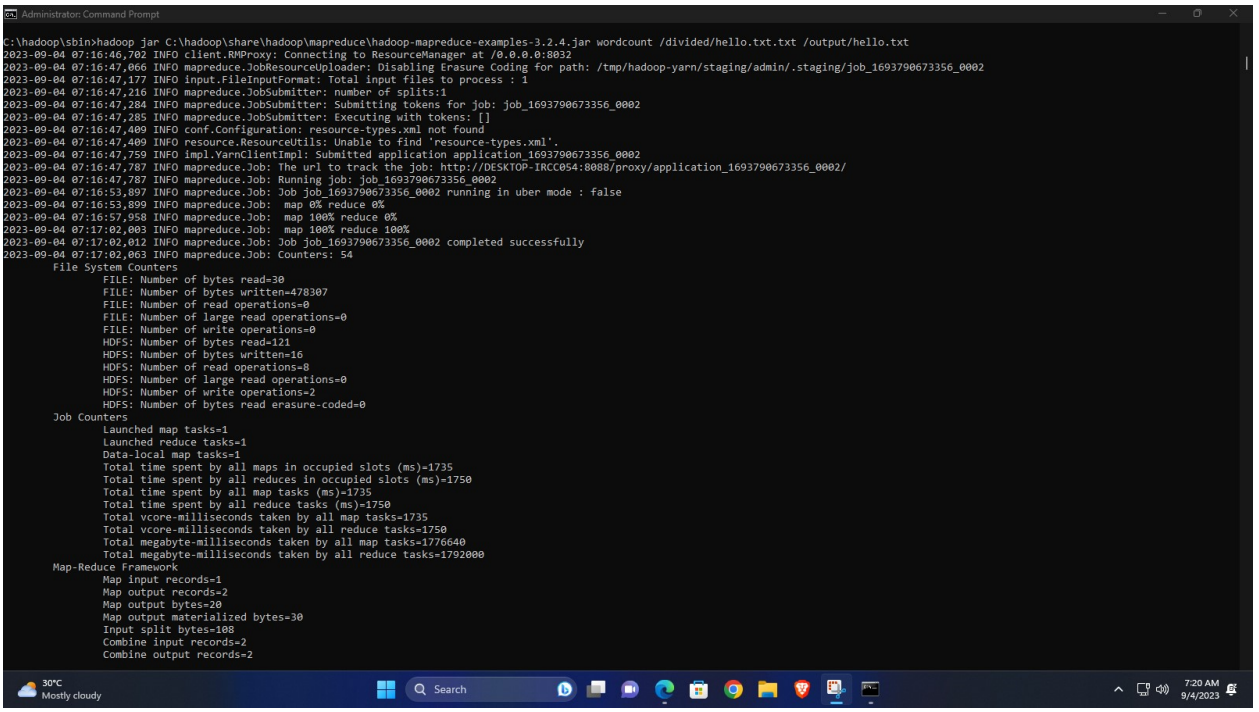

Select Administrator: Command Prompt
C:\hadoop\sbin>start-yarn.cmd
starting yarn daemons
C:\hadoop\sbin>jps
10560 NameNode
8768 DataNode
12324 Jps
12984 ResourceManager
680 NodeManager
C:\hadoop\sbin>hdfs dfs -mkdir /polomii
C:\hadoop\sbin>hdfs dfs -put C:\Users\admin\Downloads\simple.txt /polomii
C:\hadoop\sbin>hdfs dfs -tail /polomii/simple.txt
hii how are you
C:\hadoop\sbin>
simple
File Edit View
hii how are you
Ln 1, Col 1 100% Windows (CRLF) UTF-8
OK

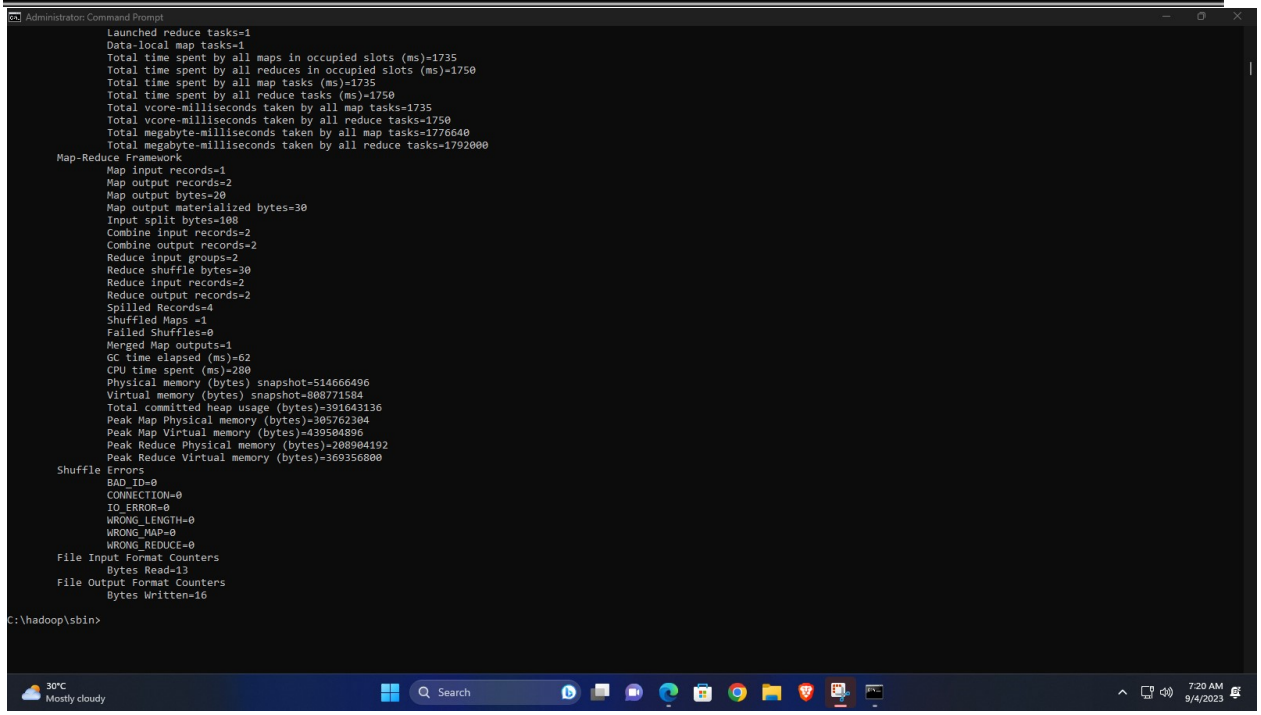

localhost:9870/explorer.html#/polomii
Microsoft Edge is not your default browser
Set as default
Not now
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/polomii
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
admin
supergroup
15 B
Sep 04 07:04
3
128 MB
simple.txt
Showing 1 to 1 of 1 entries
Previous
1
Next
Hadoop, 2022.

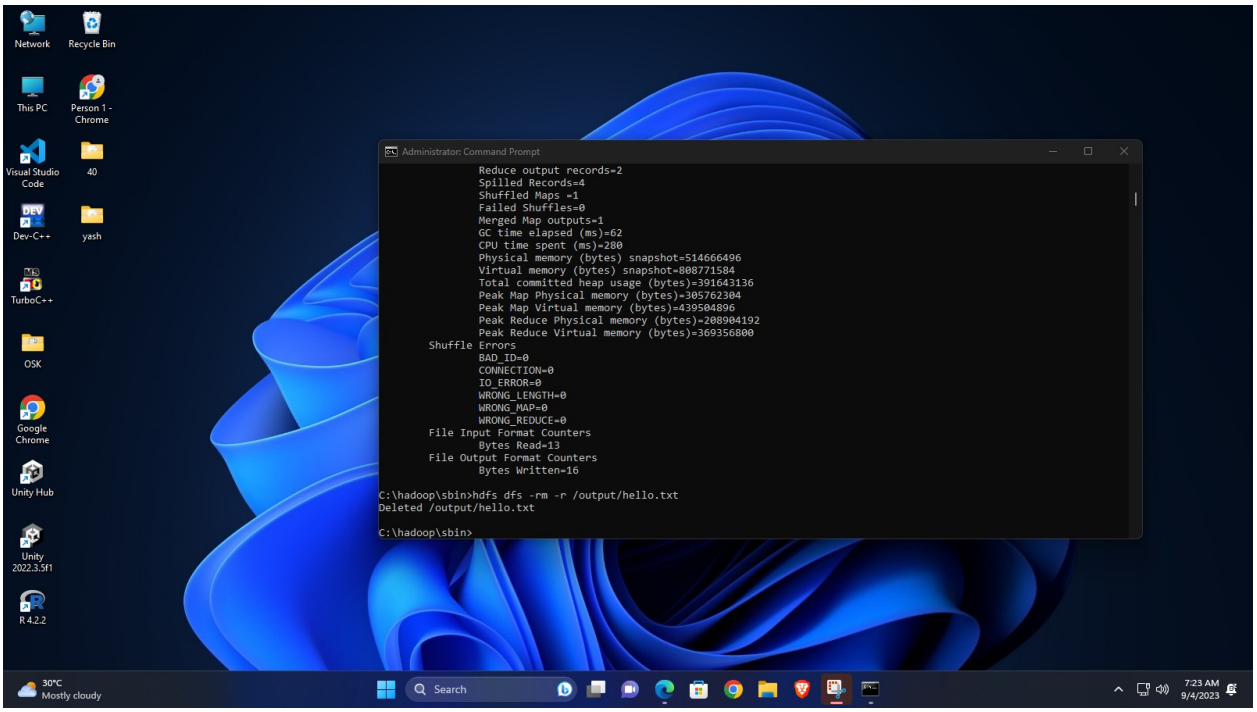

Hadoop HDFS Commands
Browsing HDFS
All Applications
localhost:9870/explorer.html#/divided
Microsoft Edge is not your default browser
Set as default
Not now
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/divided
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
admin
supergroup
26 B
Sep 04 07:11
3
128 MB
greeting.txt.txt
-rw-r--r--
admin
supergroup
13 B
Sep 04 07:11
3
128 MB
hello.txt.txt
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2022.
30°C
Mostly cloudy
Search
7:12 AM
9/4/2023

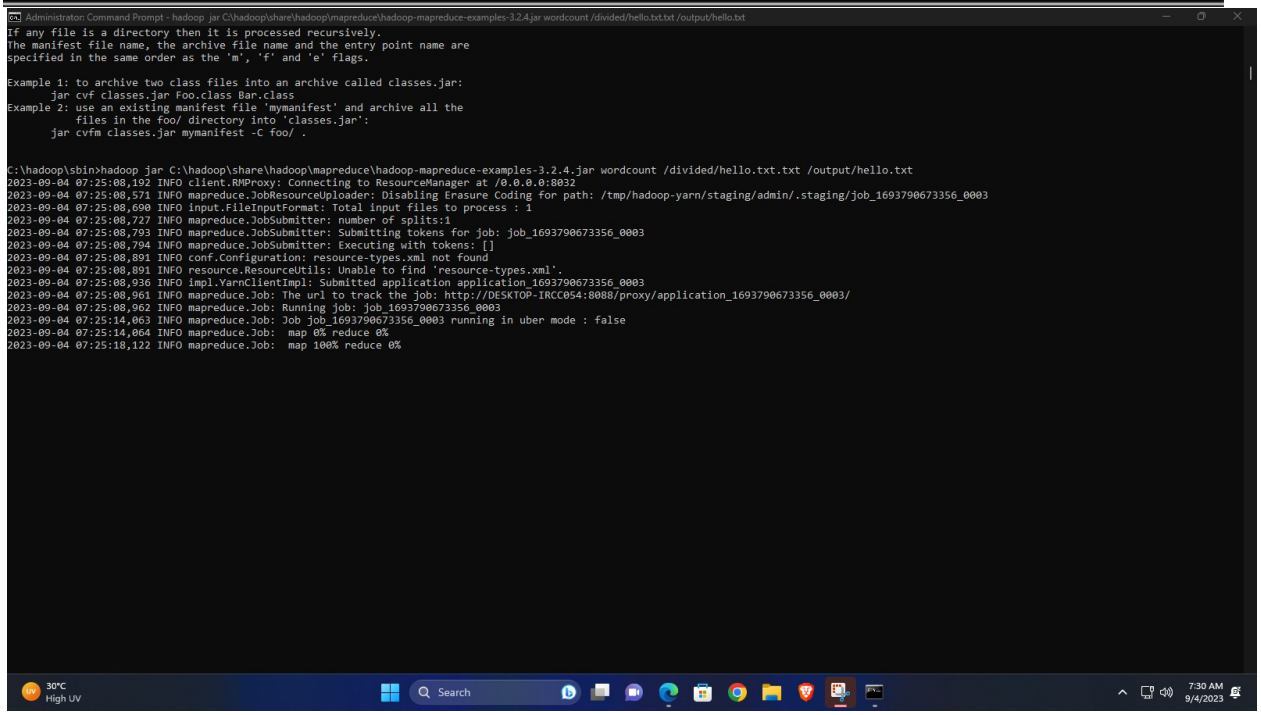

Hadoop HDFS Commands and
Browsing HDFS
All Applications
localhost:8088/cluster
Microsoft Edge is not your default browser
Set as default
Not now
hadoop
All Applications
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Used Resources
Total Resources
1
0
1
0
1
<memory:2 GB, vCores:1>
<memory:8 GB, vCores:8>
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy N
1
0
0
0
0
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
0
Show 20 entries
ID
User
Name
Application Type
Queue
Application Priority
StartTime
LaunchTime
FinishTime
State
FinalStatus
Running Containers
Allocated CPU VCores
Allocated Memory MB
Allocated GPUs
application_1693790673356_0002
admin
word count
MAPREDUCE
default
0
Mon Sep 4 07:16:47 +0550 2023
Mon Sep 4 07:16:48 +0550 2023
N/A
RUNNING
UNDEFINED
1
1
2048
-1
Showing 1 to 1 of 1 entries
30°C
Mostly cloudy
Search
7:17 AM
9/4/2023

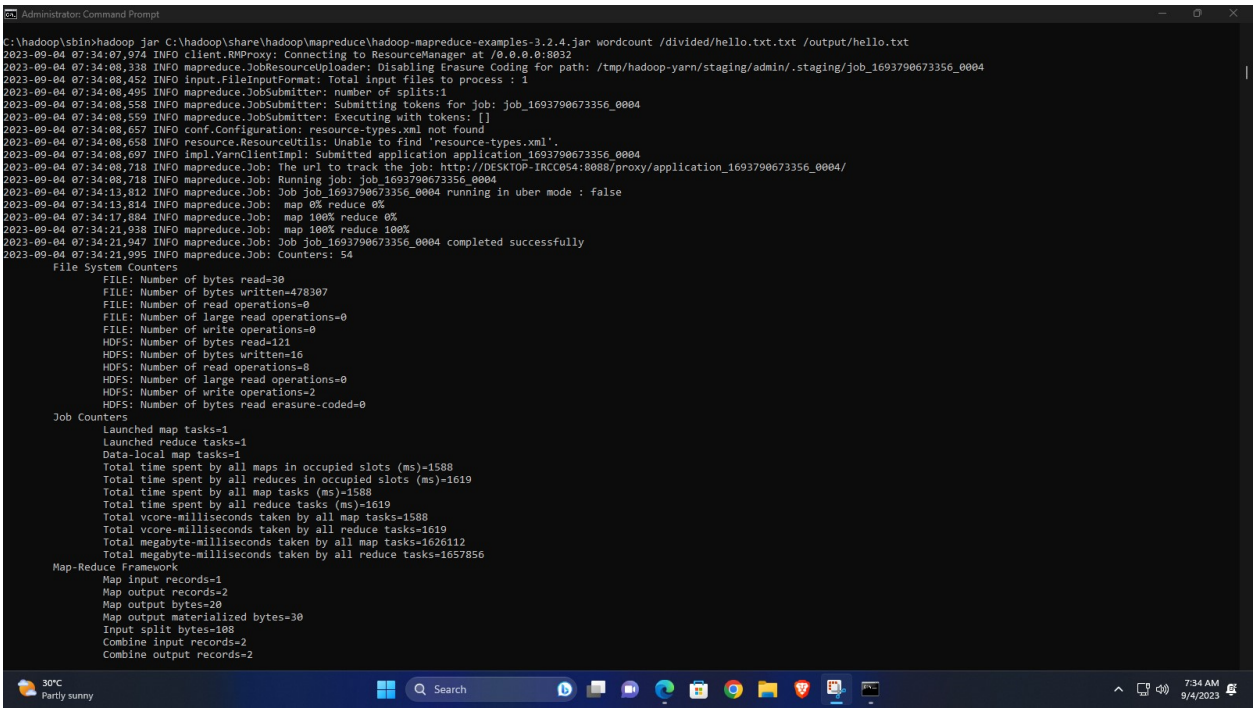

Hadoop HDFS Commands and
Browsing HDFS
Application application_1693790
localhost:8088/cluster/app/application_1693790673356_0002
Logged in as: dr.who
hadoop
Application application_1693790673356_0002
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Application Overview
User: admin
Name: word count
Application Type: MAPREDUCE
Application Tags:
Application Priority: 0 (Higher Integer value indicates higher priority)
YarnApplicationState: FINISHED
Queue: default
FinalStatus Reported by AM: SUCCEEDED
Started: Mon Sep 04 07:16:47 +0530 2023
Launched: Mon Sep 04 07:16:48 +0530 2023
Finished: Mon Sep 04 07:17:00 +0530 2023
Elapsed: 12sec
Tracking URL: History
Log Aggregation Status: DISABLED
Application Timeout (Remaining Time): Unlimited
Diagnostics:
Unmanaged Application: false
Application Node Label expression: <Not set>
AM container Node Label expression: <DEFAULT_PARTITION>
Application Metrics
Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 41250 MB-seconds, 20 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds
Show 20 entries
Search:
Attempt ID
Started
Node
Logs
Nodes blacklisted by the app
Nodes blacklisted by the system
appattempt_1693790673356_0002_000001
Mon Sep 4 07:16:47 +0550 2023
http://DESKTOP-IRCC054:8042
Logs
0
0
30°C
Mostly cloudy
Search
7:17 AM
9/4/2023

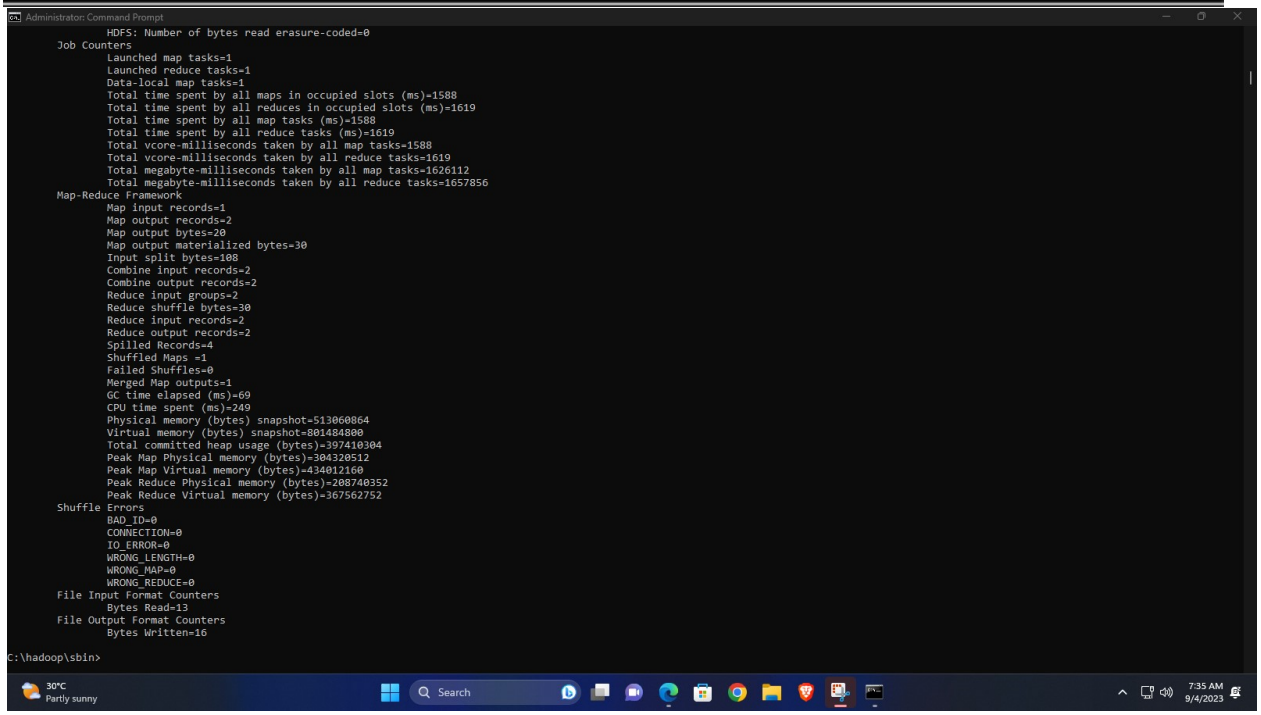

Administrator: Command Prompt
C:\hadoop\sbin>hadoop jar C:\hadoop\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.4.jar wordcount /divided/hello.txt.txt /output/hello.txt
2023-09-04 07:16:46,702 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-09-04 07:16:47,066 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/admin/.staging/job_1693790673356_0002
2023-09-04 07:16:47,177 INFO input.FileInputFormat: Total input files to process : 1
2023-09-04 07:16:47,216 INFO mapreduce.JobSubmitter: number of splits:1
2023-09-04 07:16:47,284 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1693790673356_0002
2023-09-04 07:16:47,285 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-09-04 07:16:47,409 INFO conf.Configuration: resource-types.xml not found
2023-09-04 07:16:47,409 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-09-04 07:16:47,759 INFO impl.YarnClientImpl: Submitted application application_1693790673356_0002
2023-09-04 07:16:47,787 INFO mapreduce.Job: The url to track the job: http://DESKTOP-IRCC054:8088/proxy/application_1693790673356_0002/
2023-09-04 07:16:47,787 INFO mapreduce.Job: Running job: job_1693790673356_0002
2023-09-04 07:16:53,897 INFO mapreduce.Job: Job job_1693790673356_0002 running in uber mode : false
2023-09-04 07:16:53,899 INFO mapreduce.Job: map 0% reduce 0%
2023-09-04 07:16:57,958 INFO mapreduce.Job: map 100% reduce 0%
2023-09-04 07:17:02,003 INFO mapreduce.Job: map 100% reduce 100%
2023-09-04 07:17:02,012 INFO mapreduce.Job: Job job_1693790673356_0002 completed successfully
2023-09-04 07:17:02,063 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=30
FILE: Number of bytes written=478307
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=121
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)=1735
Total time spent by all reduces in occupied slots (ms)=1750
Total time spent by all map tasks (ms)=1735
Total time spent by all reduce tasks (ms)=1750
Total vcore-milliseconds taken by all map tasks=1735
Total vcore-milliseconds taken by all reduce tasks=1750
Total megabyte-milliseconds taken by all map tasks=1776640
Total megabyte-milliseconds taken by all reduce tasks=1792000
Map-Reduce Framework
Map input records=1
Map output records=2
Map output bytes=20
Map output materialized bytes=30
Input split bytes=108
Combine input records=2
Combine output records=2
30°C
Mostly cloudy
Search
7:20 AM
9/4/2023

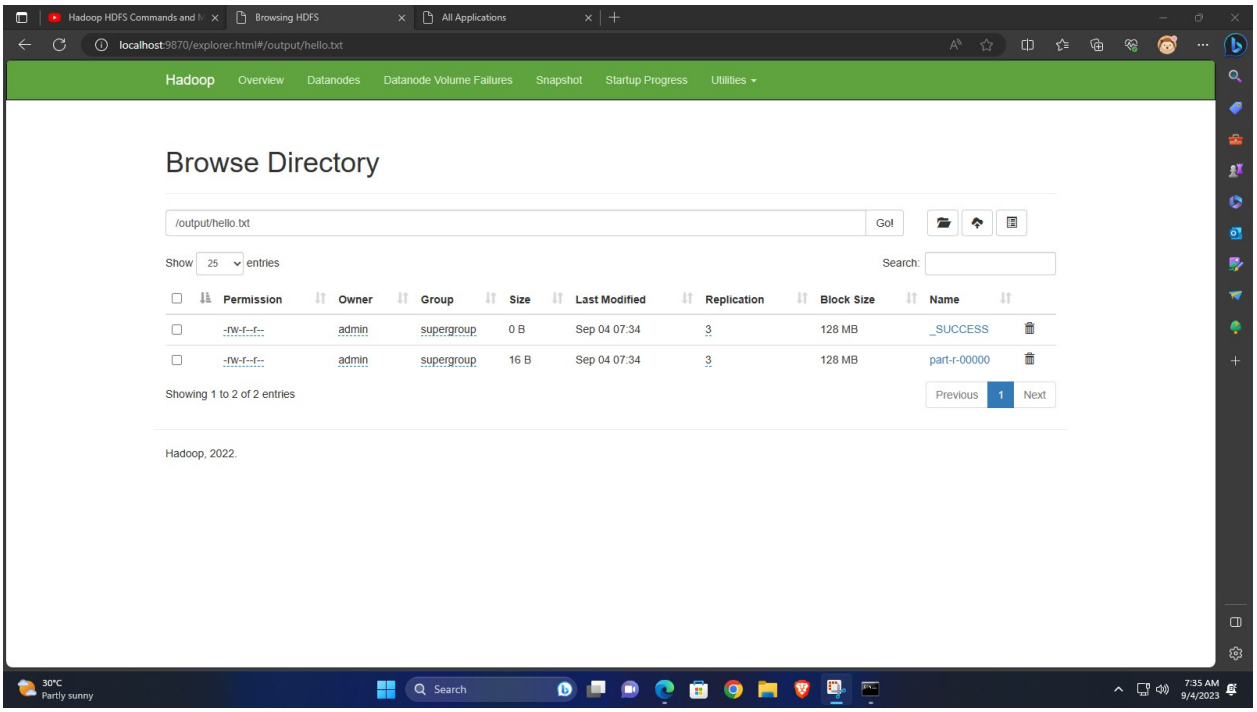

Administrator: Command Prompt
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=1735
Total time spent by all reduces in occupied slots (ms)=1750
Total time spent by all map tasks (ms)=1735
Total time spent by all reduce tasks (ms)=1750
Total vcore-milliseconds taken by all map tasks=1735
Total vcore-milliseconds taken by all reduce tasks=1750
Total megabyte-milliseconds taken by all map tasks=1776640
Total megabyte-milliseconds taken by all reduce tasks=1792000
Map-Reduce Framework
Map input records=1
Map output records=2
Map output bytes=20
Map output materialized bytes=30
Input split bytes=108
Combine input records=2
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)=62
CPU time spent (ms)=280
Physical memory (bytes) snapshot=514666496
Virtual memory (bytes) snapshot=808771584
Total committed heap usage (bytes)=391643136
Peak Map Physical memory (bytes)=305762304
Peak Map Virtual memory (bytes)=439504896
Peak Reduce Physical memory (bytes)=208904192
Peak Reduce Virtual memory (bytes)=369356800
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=13
File Output Format Counters
Bytes Written=16
C:\hadoop\sbin>
30°C
Mostly cloudy
Search
7:20 AM
9/4/2023

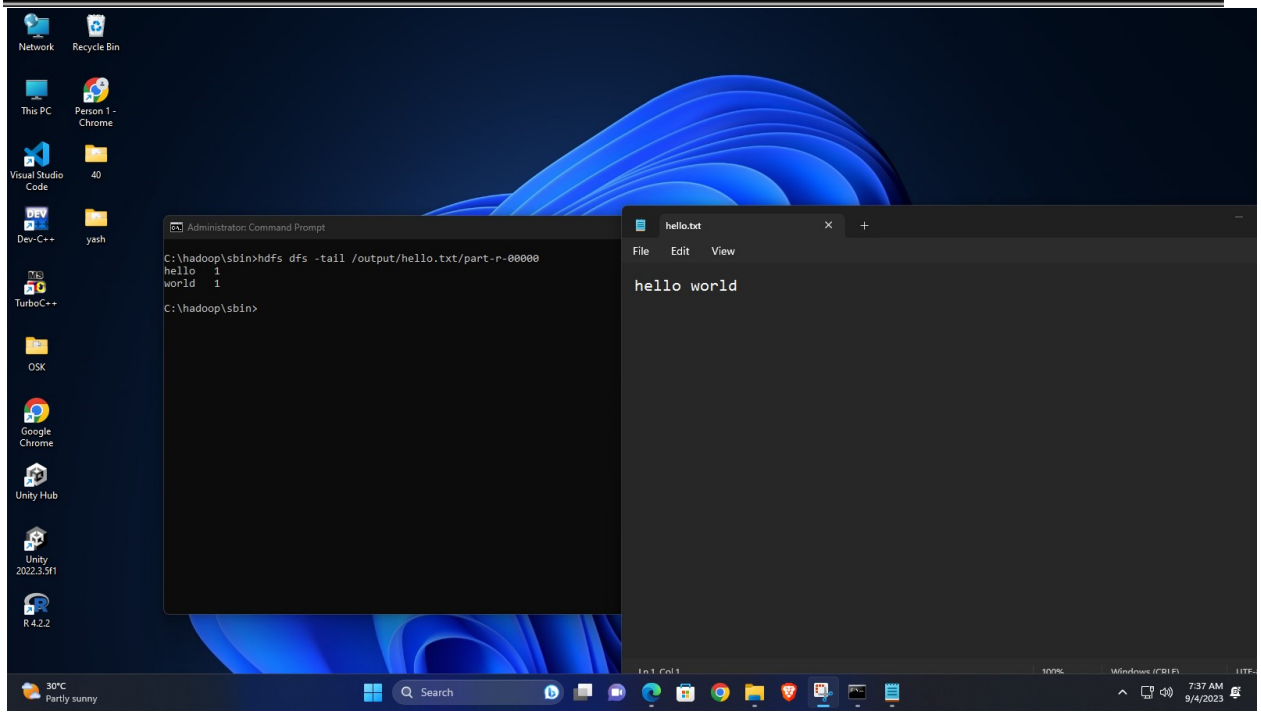

Network
Recycle Bin
This PC
Person 1 - Chrome
Visual Studio Code
40
Dev-C++
yash
TurboC++
OSK
Google Chrome
Unity Hub
Unity 2022.3.5f1
R 4.2.2
Administrator: Command Prompt
Reduce output records=2
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=62
CPU time spent (ms)=280
Physical memory (bytes) snapshot=514666496
Virtual memory (bytes) snapshot=808771584
Total committed heap usage (bytes)=391643136
Peak Map Physical memory (bytes)=305762304
Peak Map Virtual memory (bytes)=439504896
Peak Reduce Physical memory (bytes)=208904192
Peak Reduce Virtual memory (bytes)=369356800
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=13
File Output Format Counters
Bytes Written=16
C:\hadoop\sbin>hdfs dfs -rm -r /output/hello.txt
Deleted /output/hello.txt
C:\hadoop\sbin>
30°C
Mostly cloudy
Search
7:23 AM
9/4/2023

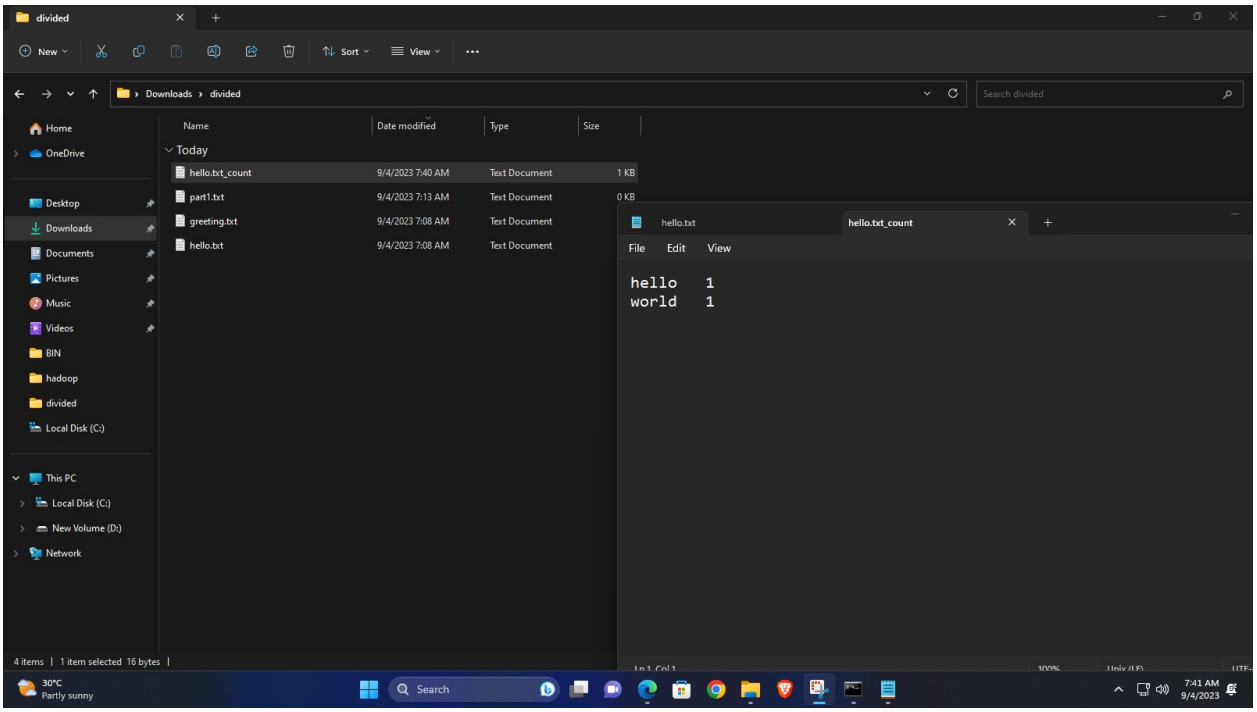

Administrator: Command Prompt - hadoop jar C:\hadoop\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.4.jar wordcount /divided/hello.txt.txt /output/hello.txt
If any file is a directory then it is processed recursively.
The manifest file name, the archive file name and the entry point name are
specified in the same order as the 'm', 'f' and 'e' flags.
Example 1: to archive two class files into an archive called classes.jar:
jar cvf classes.jar Foo.class Bar.class
Example 2: use an existing manifest file 'mymanifest' and archive all the
files in the foo/ directory into 'classes.jar':
jar cvfm classes.jar mymanifest -C foo/ .
C:\hadoop\sbin>hadoop jar C:\hadoop\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.4.jar wordcount /divided/hello.txt.txt /output/hello.txt
2023-09-04 07:25:08,192 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-09-04 07:25:08,571 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/admin/.staging/job_1693790673356_0003
2023-09-04 07:25:08,690 INFO input.FileInputFormat: Total input files to process : 1
2023-09-04 07:25:08,727 INFO mapreduce.JobSubmitter: number of splits:1
2023-09-04 07:25:08,793 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1693790673356_0003
2023-09-04 07:25:08,794 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-09-04 07:25:08,891 INFO conf.Configuration: resource-types.xml not found
2023-09-04 07:25:08,891 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-09-04 07:25:08,936 INFO impl.YarnClientImpl: Submitted application application_1693790673356_0003
2023-09-04 07:25:08,961 INFO mapreduce.Job: The url to track the job: http://DESKTOP-IRCC054:8088/proxy/application_1693790673356_0003/
2023-09-04 07:25:08,962 INFO mapreduce.Job: Running job: job_1693790673356_0003
2023-09-04 07:25:14,063 INFO mapreduce.Job: Job job_1693790673356_0003 running in uber mode : false
2023-09-04 07:25:14,064 INFO mapreduce.Job: map 0% reduce 0%
2023-09-04 07:25:18,122 INFO mapreduce.Job: map 100% reduce 0%
30°C
High UV
Search
7:30 AM
9/4/2023

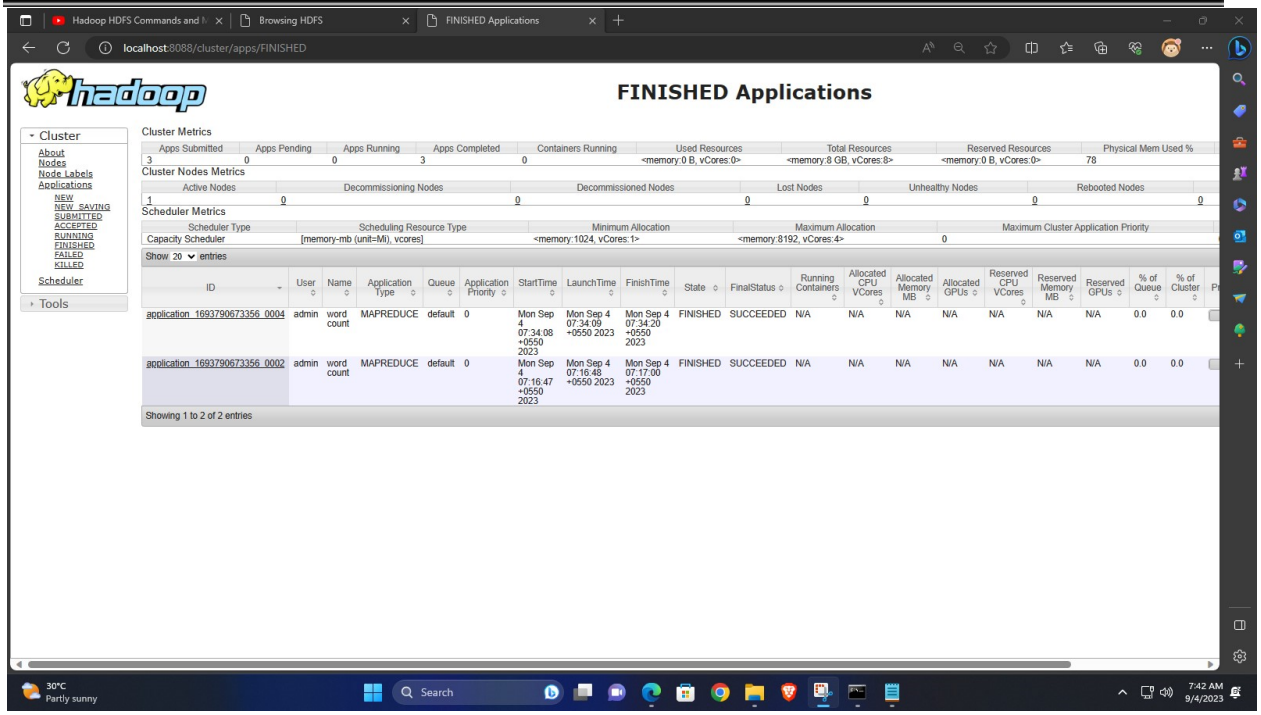

Administrator: Command Prompt
C:\hadoop\sbin>hadoop jar C:\hadoop\share\hadoop\mapreduce\hadoop-mapreduce-examples-3.2.4.jar wordcount /divided/hello.txt.txt /output/hello.txt
2023-09-04 07:34:07,974 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-09-04 07:34:08,338 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/admin/.staging/job_1693790673356_0004
2023-09-04 07:34:08,452 INFO input.FileInputFormat: Total input files to process : 1
2023-09-04 07:34:08,495 INFO mapreduce.JobSubmitter: number of splits:1
2023-09-04 07:34:08,558 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1693790673356_0004
2023-09-04 07:34:08,559 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-09-04 07:34:08,657 INFO conf.Configuration: resource-types.xml not found
2023-09-04 07:34:08,658 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-09-04 07:34:08,697 INFO impl.YarnClientImpl: Submitted application application_1693790673356_0004
2023-09-04 07:34:08,718 INFO mapreduce.Job: The url to track the job: http://DESKTOP-IRCC054:8088/proxy/application_1693790673356_0004/
2023-09-04 07:34:08,718 INFO mapreduce.Job: Running job: job_1693790673356_0004
2023-09-04 07:34:13,812 INFO mapreduce.Job: Job job_1693790673356_0004 running in uber mode : false
2023-09-04 07:34:13,814 INFO mapreduce.Job: map 0% reduce 0%
2023-09-04 07:34:17,884 INFO mapreduce.Job: map 100% reduce 0%
2023-09-04 07:34:21,938 INFO mapreduce.Job: map 100% reduce 100%
2023-09-04 07:34:21,947 INFO mapreduce.Job: Job job_1693790673356_0004 completed successfully
2023-09-04 07:34:21,995 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=30
FILE: Number of bytes written=478307
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=121
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)=1588
Total time spent by all reduces in occupied slots (ms)=1619
Total time spent by all map tasks (ms)=1588
Total time spent by all reduce tasks (ms)=1619
Total vcore-milliseconds taken by all map tasks=1588
Total vcore-milliseconds taken by all reduce tasks=1619
Total megabyte-milliseconds taken by all map tasks=1626112
Total megabyte-milliseconds taken by all reduce tasks=1657856
Map-Reduce Framework
Map input records=1
Map output records=2
Map output bytes=20
Map output materialized bytes=30
Input split bytes=108
Combine input records=2
Combine output records=2
30°C
Partly sunny
Search
7:34 AM
9/4/2023


Administrator: Command Prompt
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)=1588
Total time spent by all reduces in occupied slots (ms)=1619
Total time spent by all map tasks (ms)=1588
Total time spent by all reduce tasks (ms)=1619
Total vcore-milliseconds taken by all map tasks=1588
Total vcore-milliseconds taken by all reduce tasks=1619
Total megabyte-milliseconds taken by all map tasks=1626112
Total megabyte-milliseconds taken by all reduce tasks=1657856
Map-Reduce Framework
Map input records=1
Map output records=2
Map output bytes=20
Map output materialized bytes=30
Input split bytes=108
Combine input records=2
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)=69
CPU time spent (ms)=249
Physical memory (bytes) snapshot=513060864
Virtual memory (bytes) snapshot=801484800
Total committed heap usage (bytes)=397410304
Peak Map Physical memory (bytes)=304320512
Peak Map Virtual memory (bytes)=434012160
Peak Reduce Physical memory (bytes)=208740352
Peak Reduce Virtual memory (bytes)=367562752
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=13
File Output Format Counters
Bytes Written=16
C:\hadoop\sbin>
30°C
Partly sunny
Search
7:35 AM
9/4/2023


Hadoop HDFS Commands and
Browsing HDFS
All Applications
localhost:9870/explorer.html#/output/hello.txt
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/output/hello.txt
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
admin
supergroup
0 B
Sep 04 07:34
3
128 MB
_SUCCESS
-rw-r--r--
admin
supergroup
16 B
Sep 04 07:34
3
128 MB
part-r-00000
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2022.
30°C
Partly sunny
Search
7:35 AM
9/4/2023


Network
Recycle Bin
This PC
Person 1 - Chrome
Visual Studio Code
40
Dev-C++
yash
TurboC++
OSK
Google Chrome
Unity Hub
Unity 2022.3.5f1
R 4.2.2
Administrator: Command Prompt
C:\hadoop\sbin>hdfs dfs -tail /output/hello.txt/part-r-00000
hello 1
world 1
C:\hadoop\sbin>
hello.txt
File Edit View
hello world
30°C Partly sunny
Search
7:37 AM 9/4/2023


divided
New Sort View
Downloads › divided
Search divided
Home
OneDrive
Desktop
Downloads
Documents
Pictures
Music
Videos
BIN
hadoop
divided
Local Disk (C:)
This PC
Local Disk (C:)
New Volume (D:)
Network
Name Date modified Type Size
Today
hello.txt_count 9/4/2023 7:40 AM Text Document 1 KB
part1.txt 9/4/2023 7:13 AM Text Document 0 KB
greeting.txt 9/4/2023 7:08 AM Text Document
hello.txt 9/4/2023 7:08 AM Text Document
4 items | 1 item selected 16 bytes
hello.txt
hello.txt_count
File Edit View
hello 1
world 1
30°C Partly sunny
Search
7:41 AM 9/4/2023



CONCLUSION:

The implementation of a word count program using MapReduce and Hadoop demonstrates the power of distributed computing for handling massive datasets. This approach not only simplifies the word counting task but also lays the foundation for more complex data processing and analysis tasks in the world of big data, providing a scalable and efficient solution for extracting meaningful insights from vast amounts of textual information. With Hadoop's distributed architecture, it becomes possible to process data across a cluster of machines, making it an essential tool for modern data-driven applications and research.