Name- Jaidev rawat

SEC-DS

ROLL no – 34

Uni roll no- 2015170

Problem Statement 1:

You must calculate the frequency of the following words in the input file through MapReduce program:

Apple, Banana and Grapes.

WordCountDriver

```
import java.io.IOException;
import java.net.URI;
import java.net.URISyntaxException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class WordCountDriver {
        public static void main(String[] args) throws IOException,
ClassNotFoundException, InterruptedException, URISyntaxException
        {
```

```
        Configuration conf = new Configuration();



        Job j = new Job();
        j.setJobName("My First Job");
        j.setJarByClass(WordCountDriver.class );
        j.setMapperClass(WordCountMapper.class );
        j.setReducerClass(WordCountReducer.class);
        j.setOutputKeyClass(Text.class);
        j.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(j, new Path(args[0]));
        FileOutputFormat.setOutputPath(j, new Path(args[1]));
        URI uri = new URI(args[1].toString());
        FileSystem fs =  FileSystem.get(uri, conf);
        boolean x = fs.delete(new Path(uri),true);
        int xxx =  j.waitForCompletion(true) ? 0 : 1;
    }
}
```

WordCountMapper

```
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.io.DoubleWritable;
```

```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class WordCountMapper extends Mapper<LongWritable,  Text, Text, IntWritable>

{
	protected void map(LongWritable key, Text value,
			org.apache.hadoop.mapreduce.Mapper.Context context)
			throws IOException, InterruptedException {
			String data[]=value.toString().split(" ");
			for(String word:data)
			{
			if(word.equalsIgnoreCase("Apple"))
			{
				context.write(new Text("Apple"), new IntWritable(1));
			}
			if(word.equalsIgnoreCase("Banana"))
			{
				context.write(new Text("Banana"), new IntWritable(1));
			}
			if(word.equalsIgnoreCase("Grapes"))
			{
				context.write(new Text("Grapes"), new IntWritable(1));
			}

		}
```

```
}
}
```

WordCountReducer

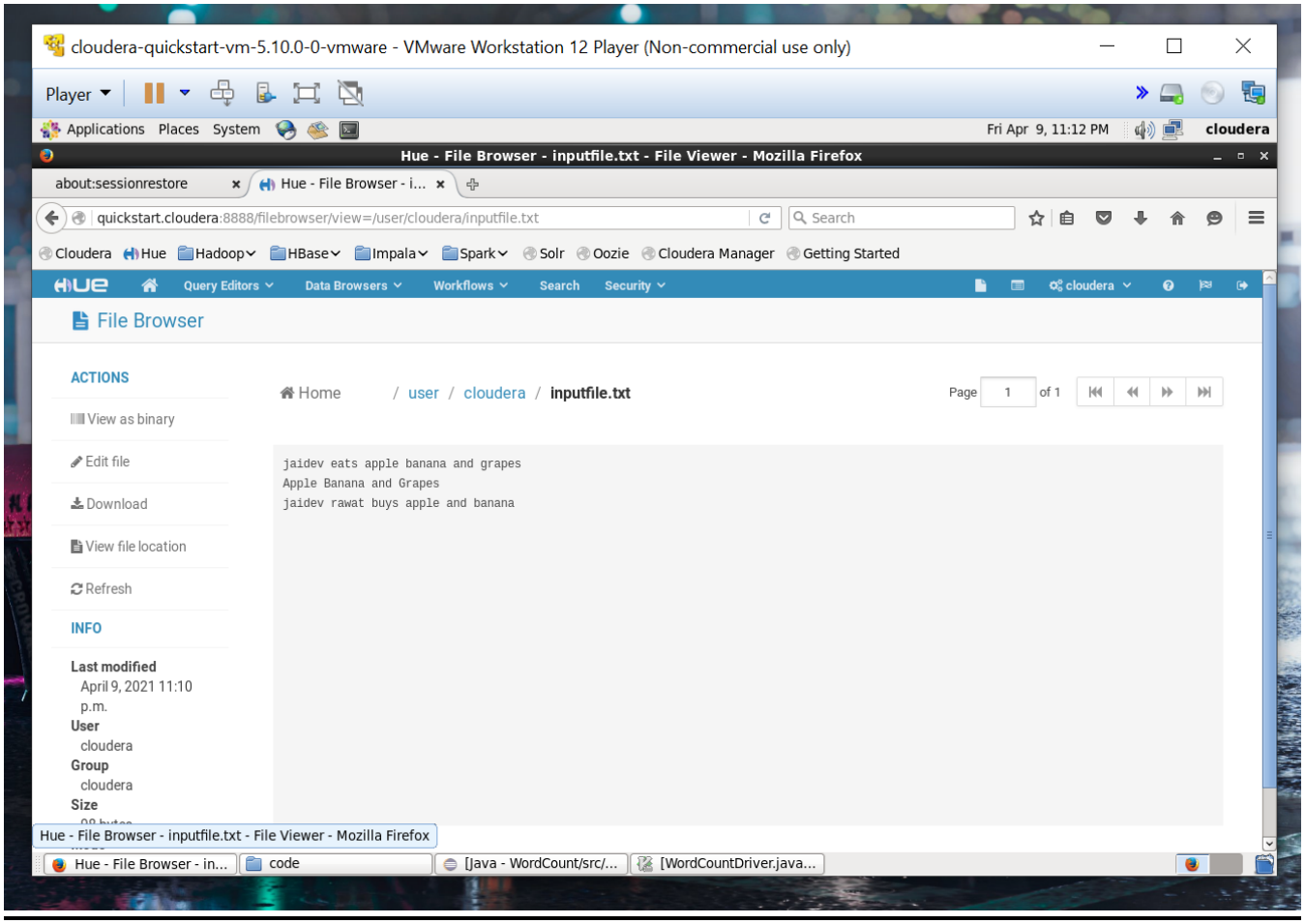
```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.Reducer.Context;
public class WordCountReducer extends Reducer<Text, IntWritable, Text, IntWritable>{
	public void reduce(Text key, Iterable<IntWritable> values, Context context)
	throws IOException, InterruptedException {

		int sum = 0;

		for(IntWritable x : values)
		{
			sum ++;
		}
		context.write(key, new IntWritable(sum) );

	}
}
```

INPUT FILE

jaidev eats apple banana and grapes

Apple Banana and Grapes

jaidev rawat buys apple and banana

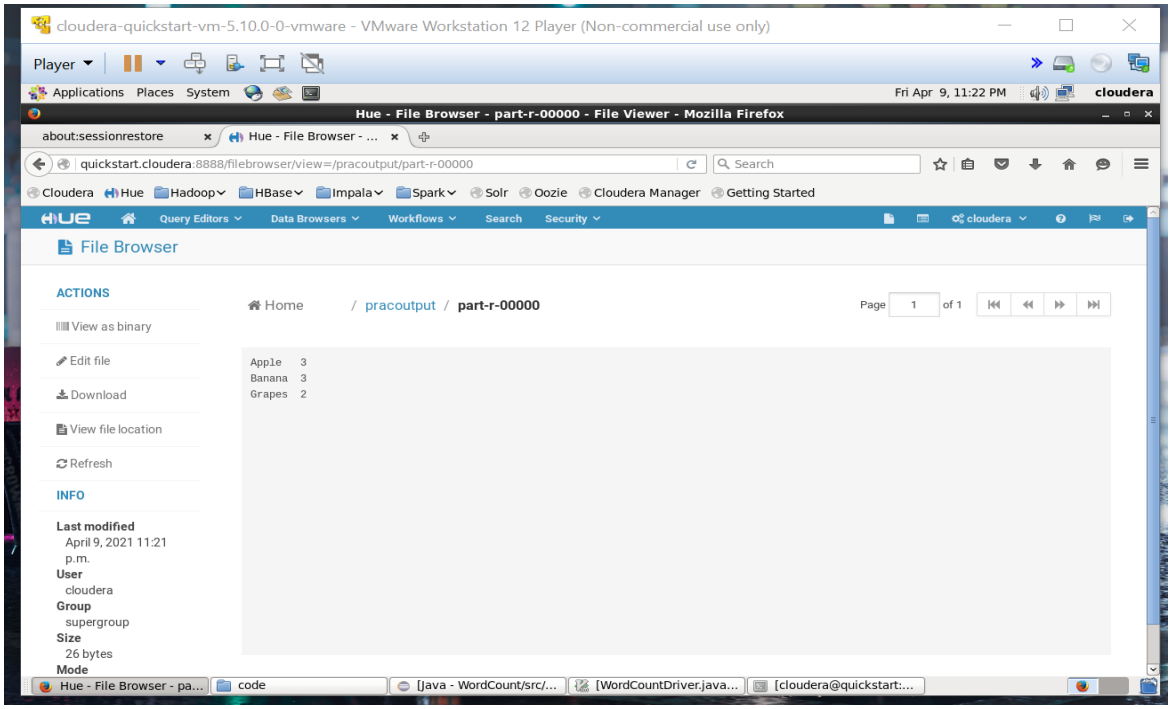


OUTPUT FILE

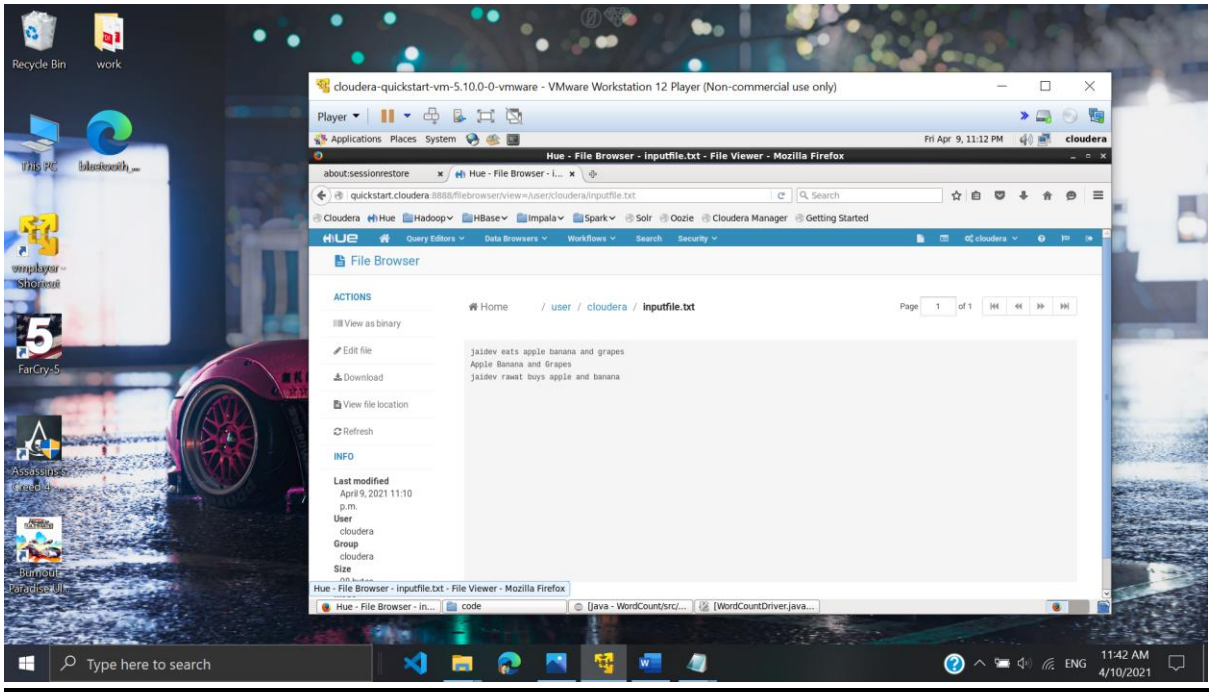
Apple 3

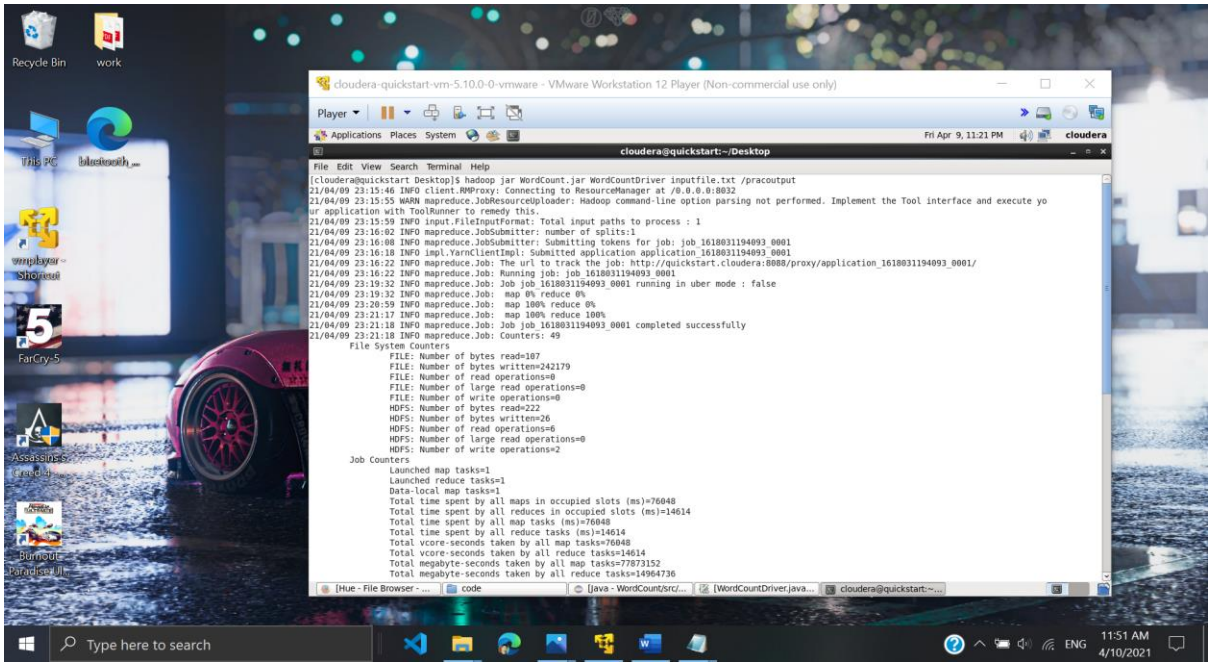
Banana 3

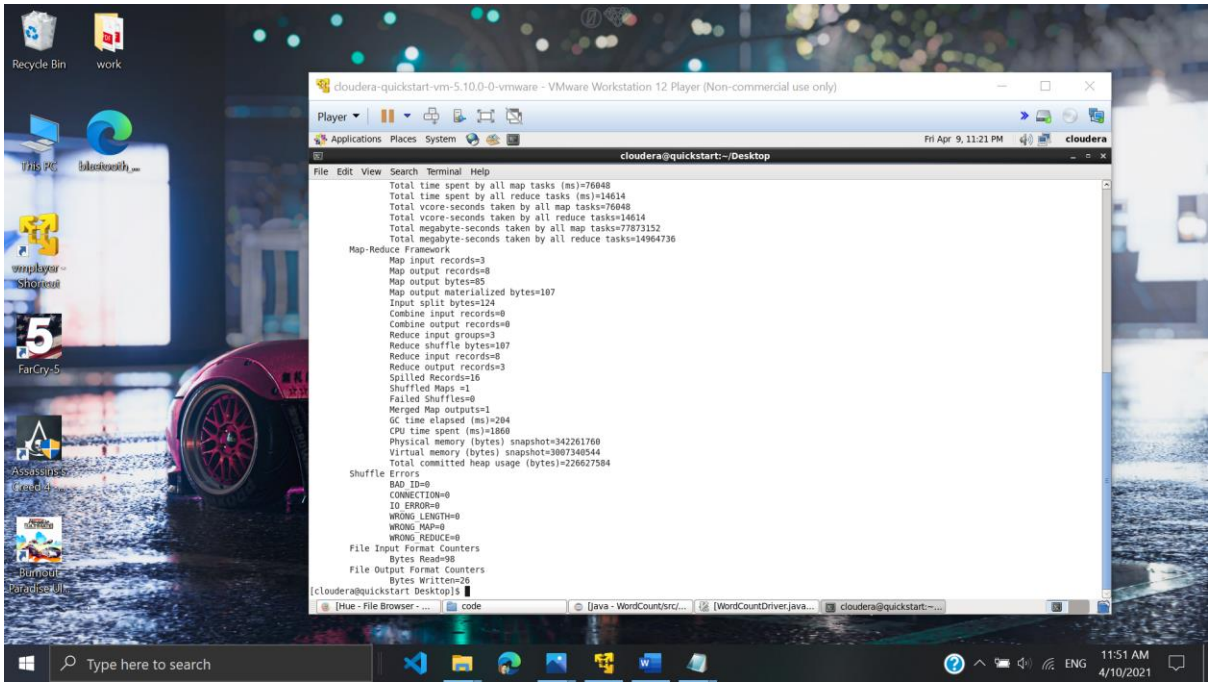
Grapes 2

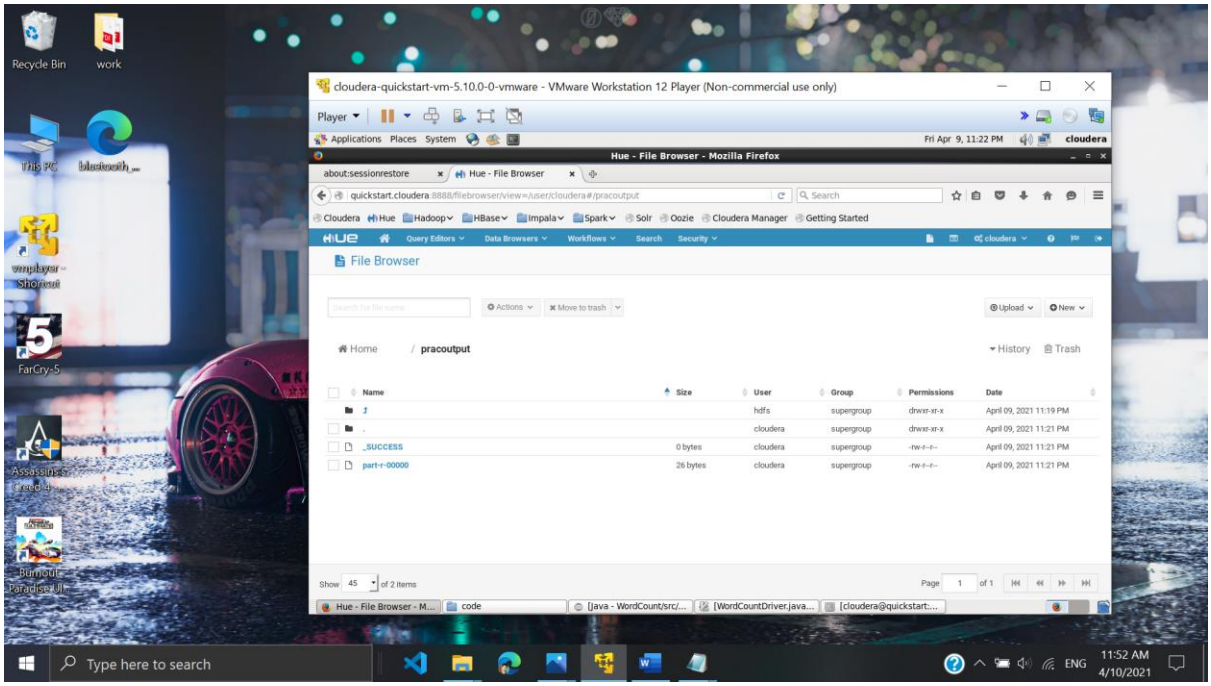


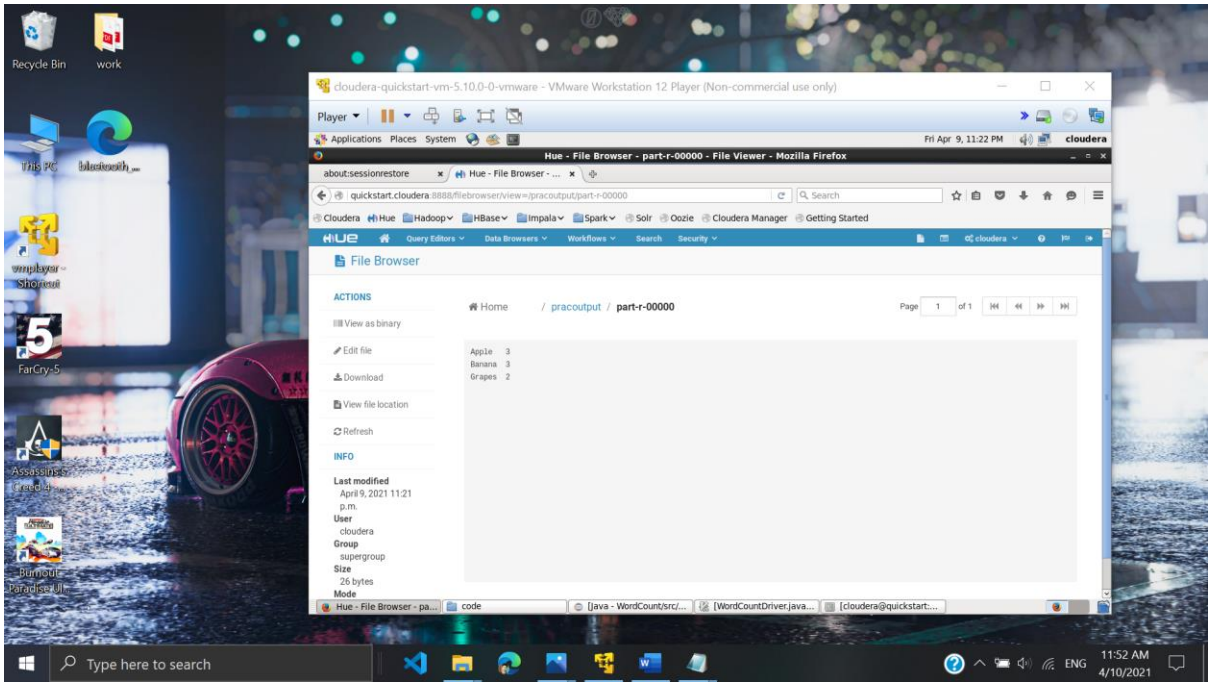
Process


Recycle Bin
work
This PC
cloudera-quickstart-vm-5.10.0-0-vmware - VMware Workstation 12 Player (Non-commercial use only)
Player
Applications Places System
Fri Apr 9, 11:12 PM
cloudera
Hue - File Browser - inputfile.txt - File Viewer - Mozilla Firefox
about:sessionrestore
Hue - File Browser - i...
quickstart.cloudera:8888/filebrowser/view=/user/cloudera/inputfile.txt
Search
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started
Query Editors Data Browsers Workflows Search Security
File Browser
ACTIONS
View as binary
Edit file
Download
View file location
Refresh
INFO
Last modified
April 9, 2021 11:10 p.m.
User
cloudera
Group
cloudera
Size
Home / user / cloudera / inputfile.txt
Page 1 of 1
jaidev eats apple banana and grapes
Apple Banana and Grapes
jaidev rawat buys apple and banana
Hue - File Browser - inputfile.txt - File Viewer - Mozilla Firefox
Hue - File Browser - in...
code
[Java - WordCount/src/...
[WordCountDriver.java...
Type here to search
ENG
11:42 AM
4/10/2021


cloudera-quickstart-vm-5.10.0-0-vmware - VMware Workstation 12 Player (Non-commercial use only)
Player
Applications Places System
Fri Apr 9, 11:21 PM
cloudera
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hadoop jar WordCount.jar WordCountDriver inputfile.txt /pracoutput
21/04/09 23:15:46 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/04/09 23:15:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute yo
ur application with ToolRunner to remedy this.
21/04/09 23:15:59 INFO input.FileInputFormat: Total input paths to process : 1
21/04/09 23:16:02 INFO mapreduce.JobSubmitter: number of splits:1
21/04/09 23:16:08 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1618031194093_0001
21/04/09 23:16:18 INFO impl.YarnClientImpl: Submitted application application_1618031194093_0001
21/04/09 23:16:22 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1618031194093_0001/
21/04/09 23:16:22 INFO mapreduce.Job: Running job: job_1618031194093_0001
21/04/09 23:19:32 INFO mapreduce.Job: Job job_1618031194093_0001 running in uber mode : false
21/04/09 23:19:32 INFO mapreduce.Job: map 0% reduce 0%
21/04/09 23:20:59 INFO mapreduce.Job: map 100% reduce 0%
21/04/09 23:21:17 INFO mapreduce.Job: map 100% reduce 100%
21/04/09 23:21:18 INFO mapreduce.Job: Job job_1618031194093_0001 completed successfully
21/04/09 23:21:18 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=107
FILE: Number of bytes written=242179
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=222
HDFS: Number of bytes written=26
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
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)=76048
Total time spent by all reduces in occupied slots (ms)=14614
Total time spent by all map tasks (ms)=76048
Total time spent by all reduce tasks (ms)=14614
Total vcore-seconds taken by all map tasks=76048
Total vcore-seconds taken by all reduce tasks=14614
Total megabyte-seconds taken by all map tasks=77873152
Total megabyte-seconds taken by all reduce tasks=14964736
[Hue - File Browser - ...
code
[Java - WordCount/src/...
[WordCountDriver.java...
cloudera@quickstart:~...
Recycle Bin
work
This PC
blackearth_...
vmplayer - Shortcut
FarCry-5
Assassin's Creed 4 ...
Burnout Paradise Full...
Type here to search
ENG
11:51 AM
4/10/2021


cloudera-quickstart-vm-5.10.0-0-vmware - VMware Workstation 12 Player (Non-commercial use only)
Player
Applications Places System
Fri Apr 9, 11:21 PM
cloudera
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
Total time spent by all map tasks (ms)=76048
Total time spent by all reduce tasks (ms)=14614
Total vcore-seconds taken by all map tasks=76048
Total vcore-seconds taken by all reduce tasks=14614
Total megabyte-seconds taken by all map tasks=77873152
Total megabyte-seconds taken by all reduce tasks=14964736
Map-Reduce Framework
Map input records=3
Map output records=8
Map output bytes=85
Map output materialized bytes=107
Input split bytes=124
Combine input records=0
Combine output records=0
Reduce input groups=3
Reduce shuffle bytes=107
Reduce input records=8
Reduce output records=3
Spilled Records=16
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=204
CPU time spent (ms)=1860
Physical memory (bytes) snapshot=342261760
Virtual memory (bytes) snapshot=3007340544
Total committed heap usage (bytes)=226627584
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=98
File Output Format Counters
Bytes Written=26
[cloudera@quickstart Desktop]$
[Hue - File Browser - ...
code
[Java - WordCount/src/...
[WordCountDriver.java...
cloudera@quickstart:~...
Recycle Bin
work
This PC
blackearth_...
vmplayer - Shortcut
FarCry-5
Assassin's Creed 4 ...
Burnout Paradise Full...
Type here to search
ENG
11:51 AM
4/10/2021


cloudera-quickstart-vm-5.10.0-0-vmware - VMware Workstation 12 Player (Non-commercial use only)
Hue - File Browser - Mozilla Firefox
File Browser
Home / pracoutput
Name
Size
User
Group
Permissions
Date
_SUCCESS
0 bytes
part-r-00000
26 bytes
Type here to search
11:52 AM
4/10/2021


cloudera-quickstart-vm-5.10.0-0-vmware - VMware Workstation 12 Player (Non-commercial use only)
Hue - File Browser - part-r-00000 - File Viewer - Mozilla Firefox
File Browser
ACTIONS
View as binary
Edit file
Download
View file location
Refresh
INFO
Home / pracoutput / part-r-00000
Apple 3
Banana 3
Grapes 2
Type here to search
11:52 AM
4/10/2021