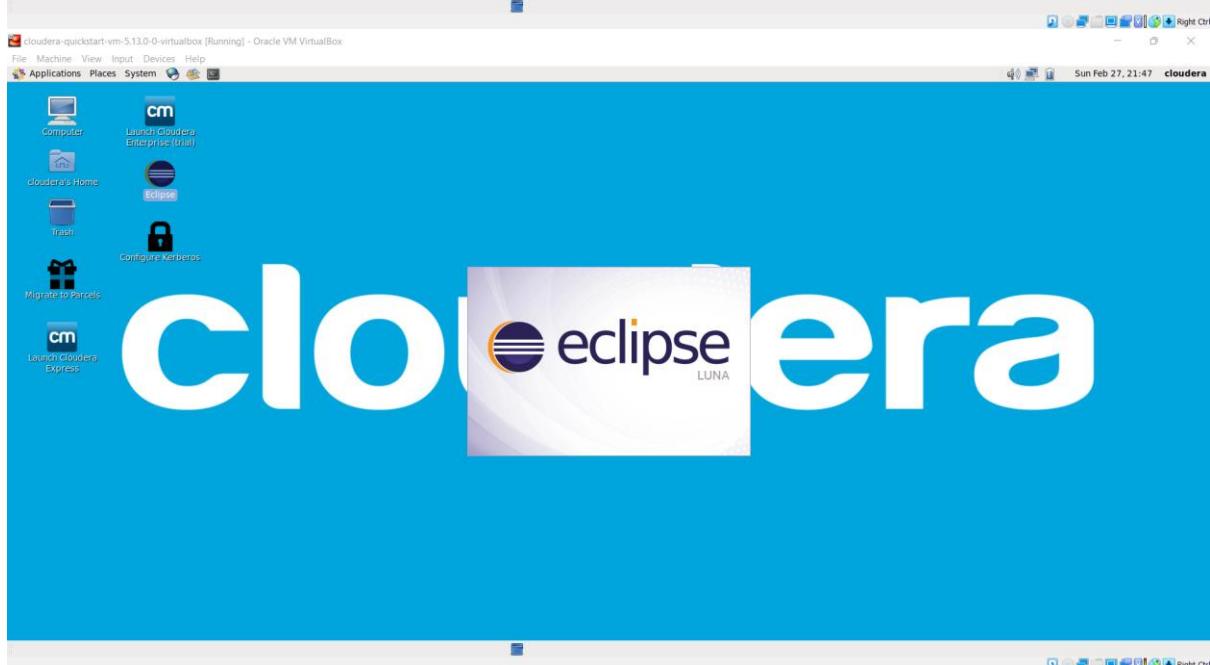
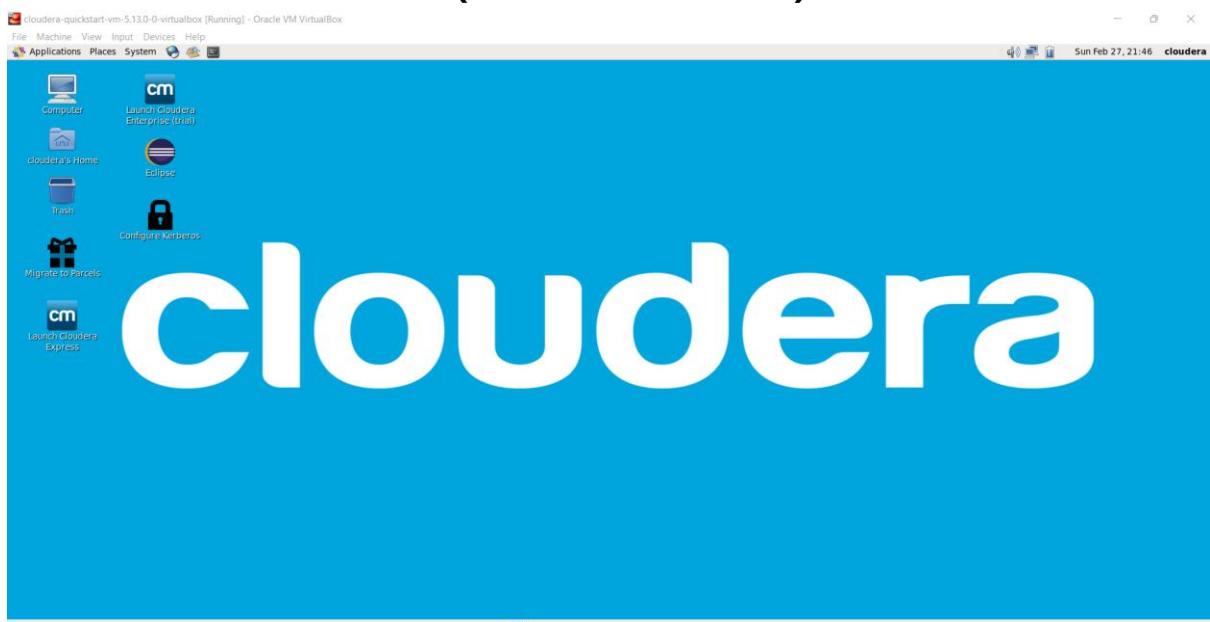
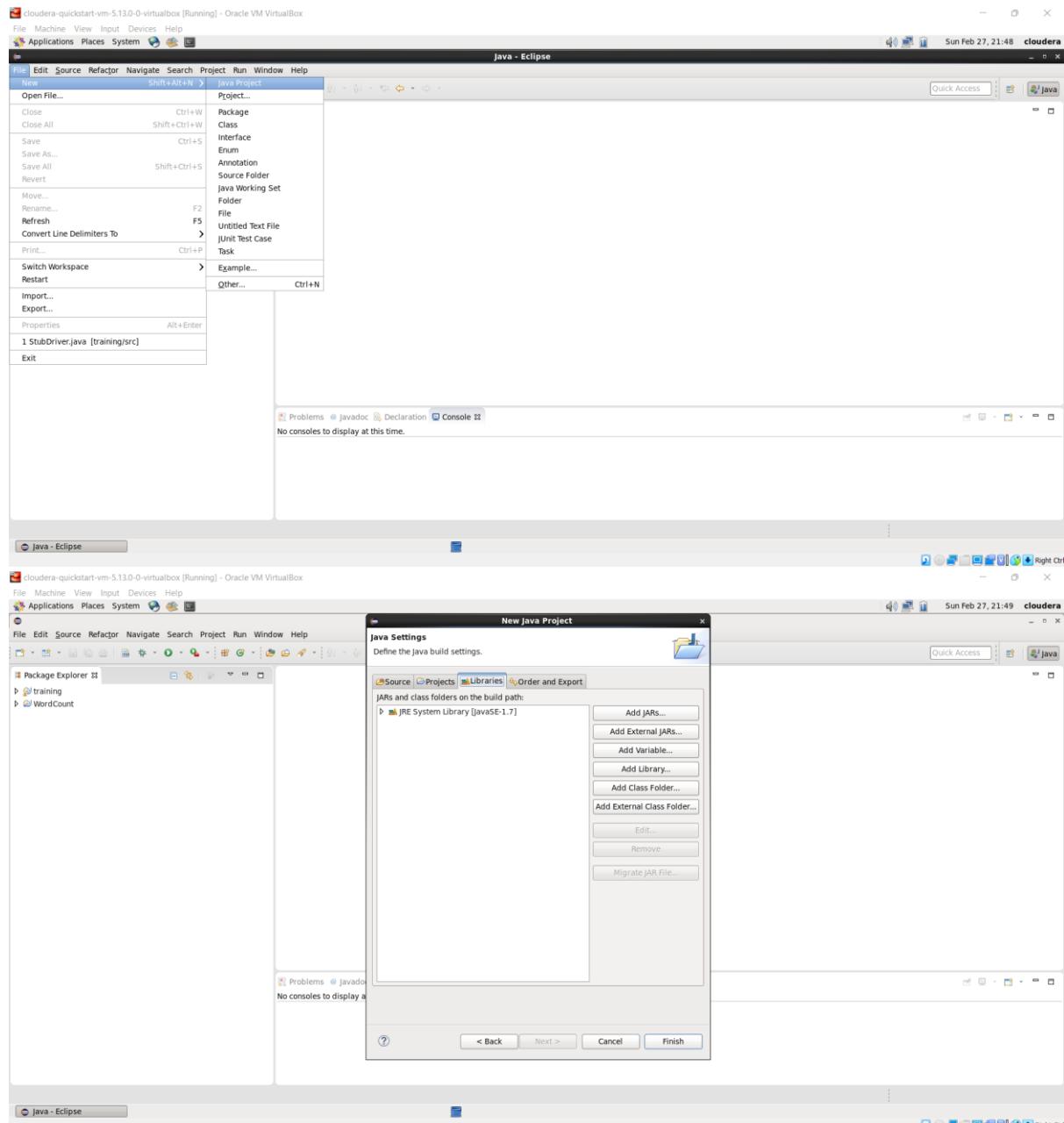


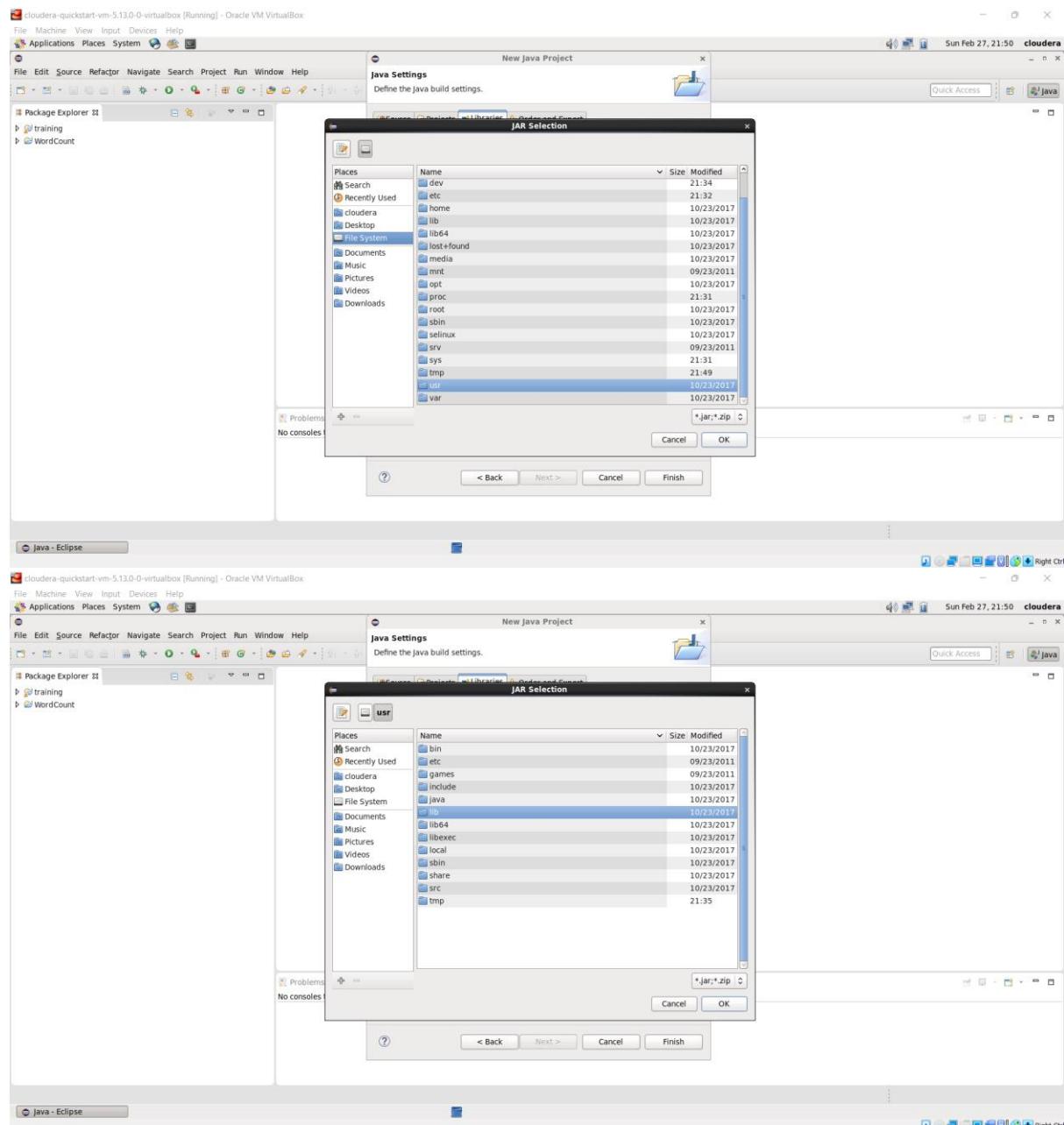
## Practical 3

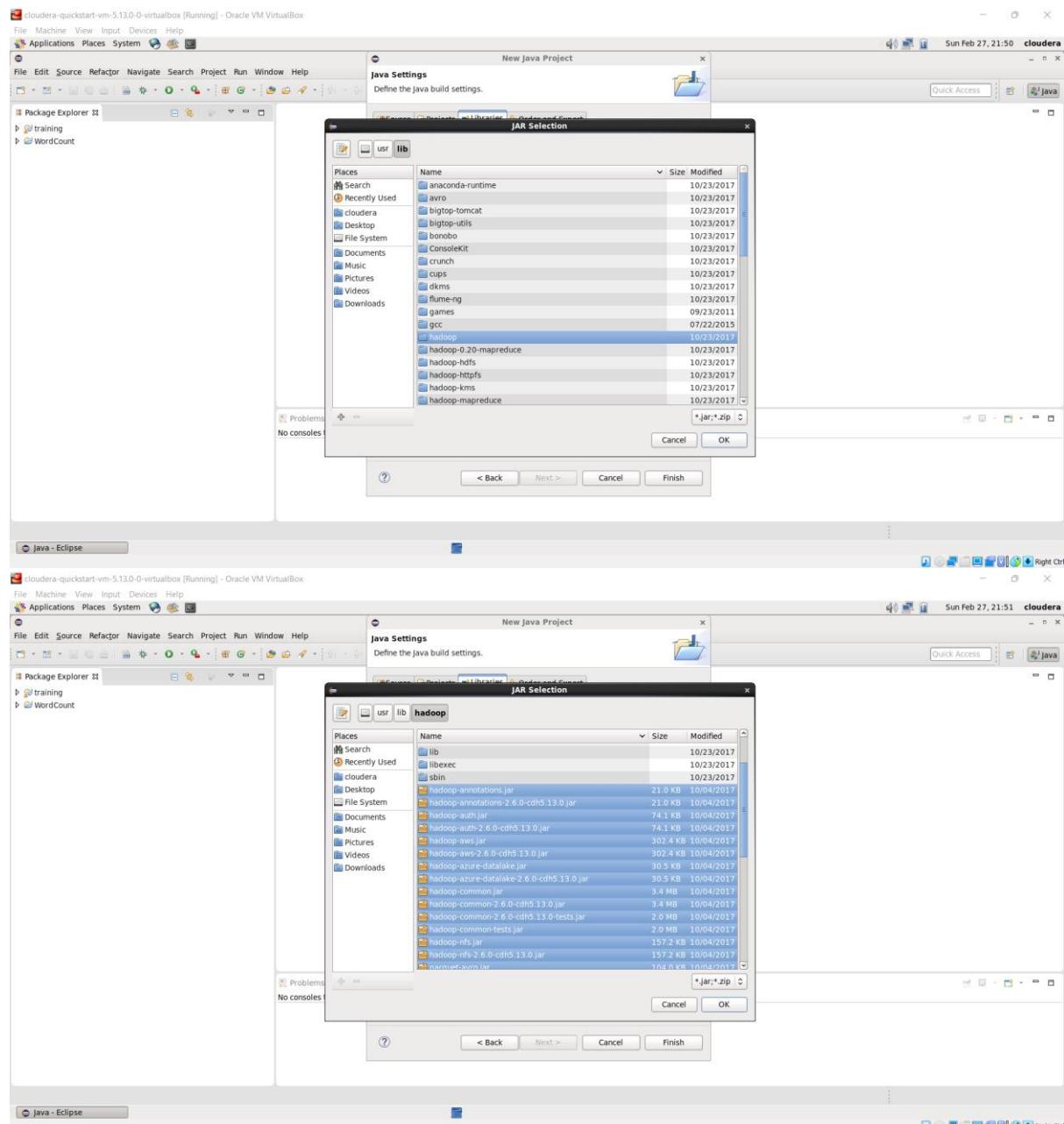
### To Implement WordCount problem using Hadoop MapReduce in Eclipse: (Without Combiner & With Combiner)

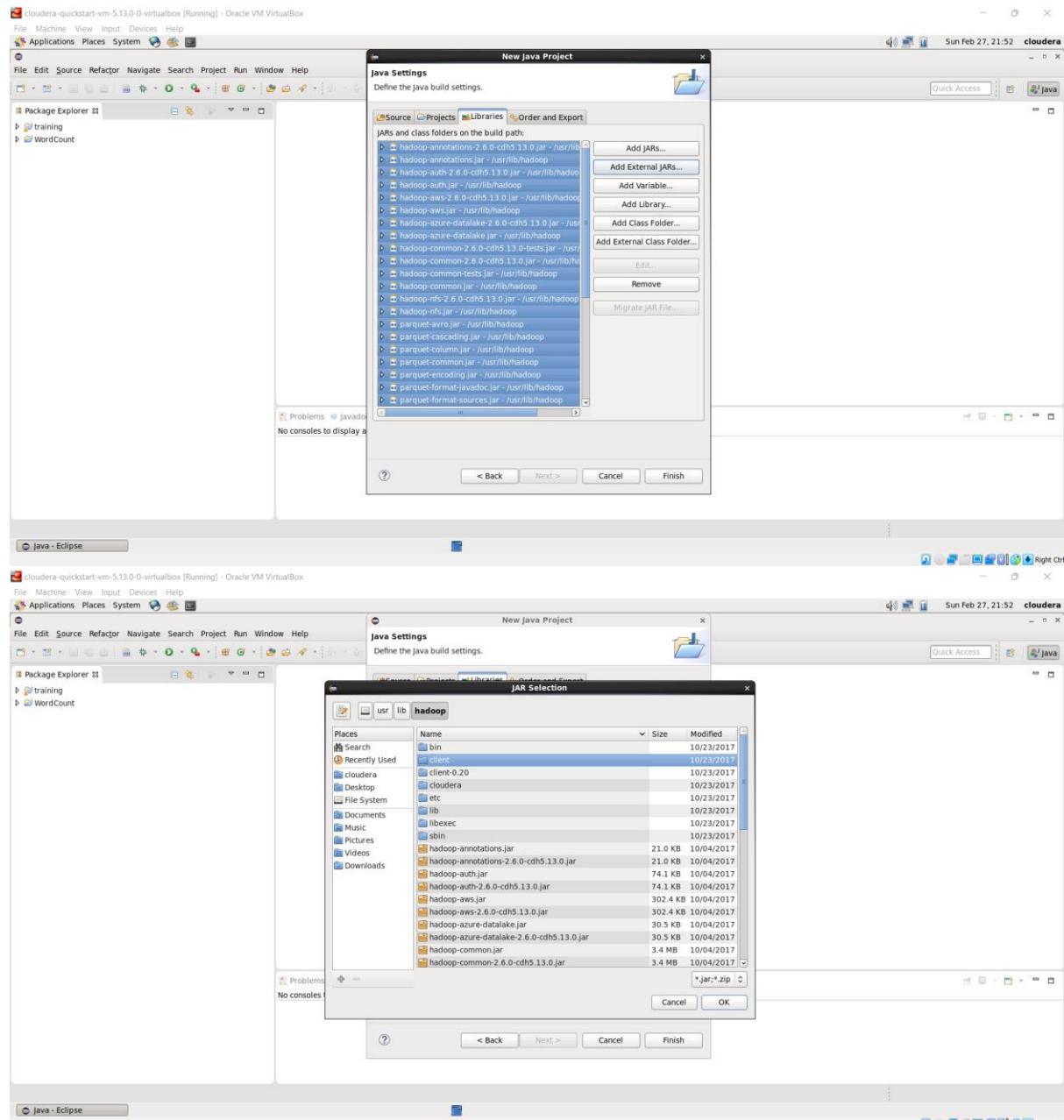
#### Word Count in Cloudera: (Without Combiner):

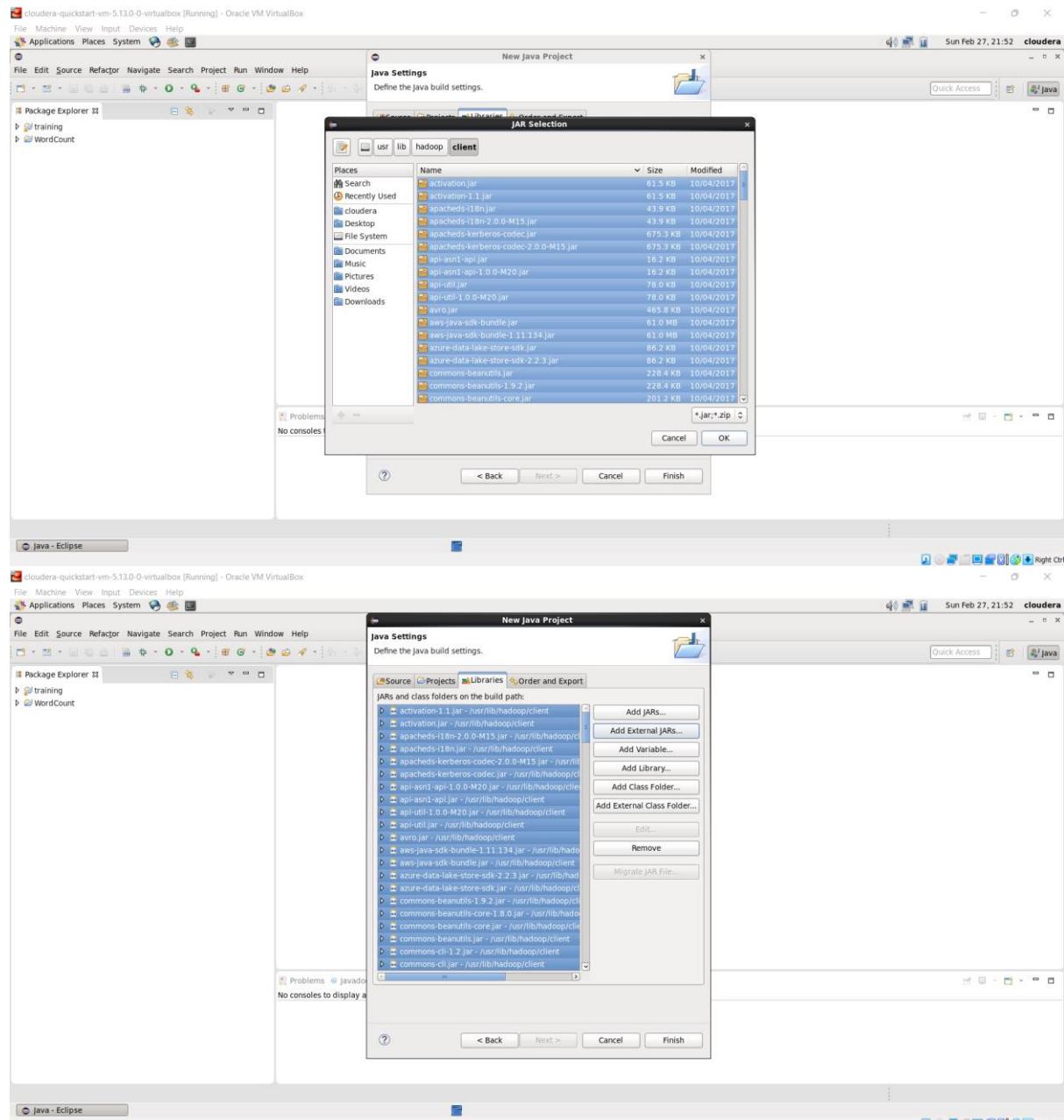


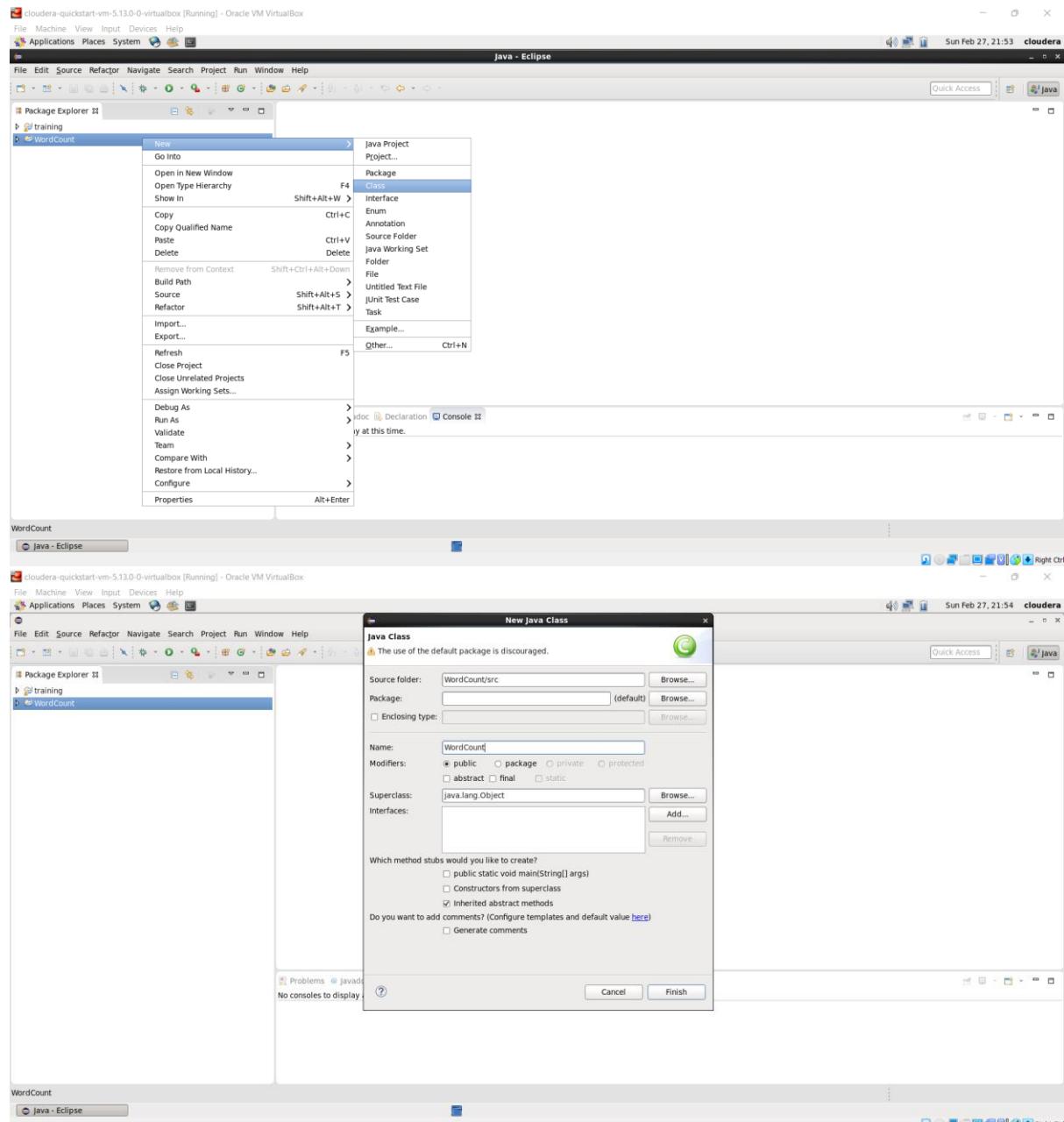


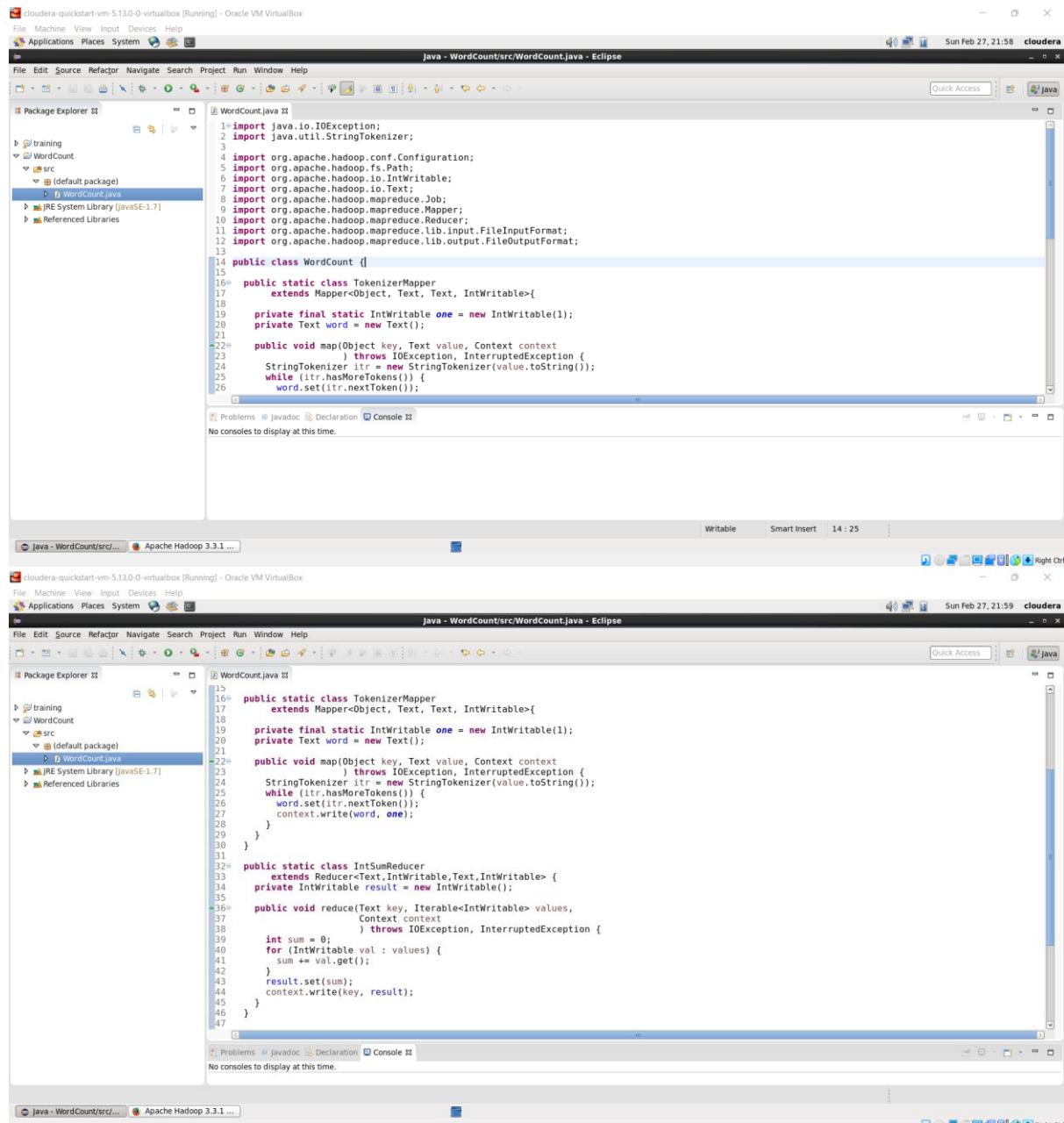












```

1 import java.io.IOException;
2 import java.util.StringTokenizer;
3 
4 import org.apache.hadoop.conf.Configuration;
5 import org.apache.hadoop.fs.Path;
6 import org.apache.hadoop.io.IntWritable;
7 import org.apache.hadoop.io.Text;
8 import org.apache.hadoop.mapreduce.Job;
9 import org.apache.hadoop.mapreduce.Mapper;
10 import org.apache.hadoop.mapreduce.Reducer;
11 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
12 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
13 
14 public class WordCount {
15 
16     public static class TokenizerMapper
17         extends Mapper<Object, Text, Text, IntWritable>{
18 
19         private final static IntWritable one = new IntWritable(1);
20         private Text word = new Text();
21 
22         public void map(Object key, Text value, Context context
23                         ) throws IOException, InterruptedException {
24             StringTokenizer itr = new StringTokenizer(value.toString());
25             while (itr.hasMoreTokens()) {
26                 word.set(itr.nextToken());
27                 word.set(itr.nextToken());
28             }
29         }
30     }
31 
32     public static class IntSumReducer
33         extends Reducer<Text,IntWritable,Text,IntWritable> {
34         private IntWritable result = new IntWritable();
35 
36         public void reduce(Text key, Iterable<IntWritable> values,
37                           Context context
38                           ) throws IOException, InterruptedException {
39             int sum = 0;
40             for (IntWritable val : values) {
41                 sum += val.get();
42             }
43             result.set(sum);
44             context.write(key, result);
45         }
46     }
47 }

```

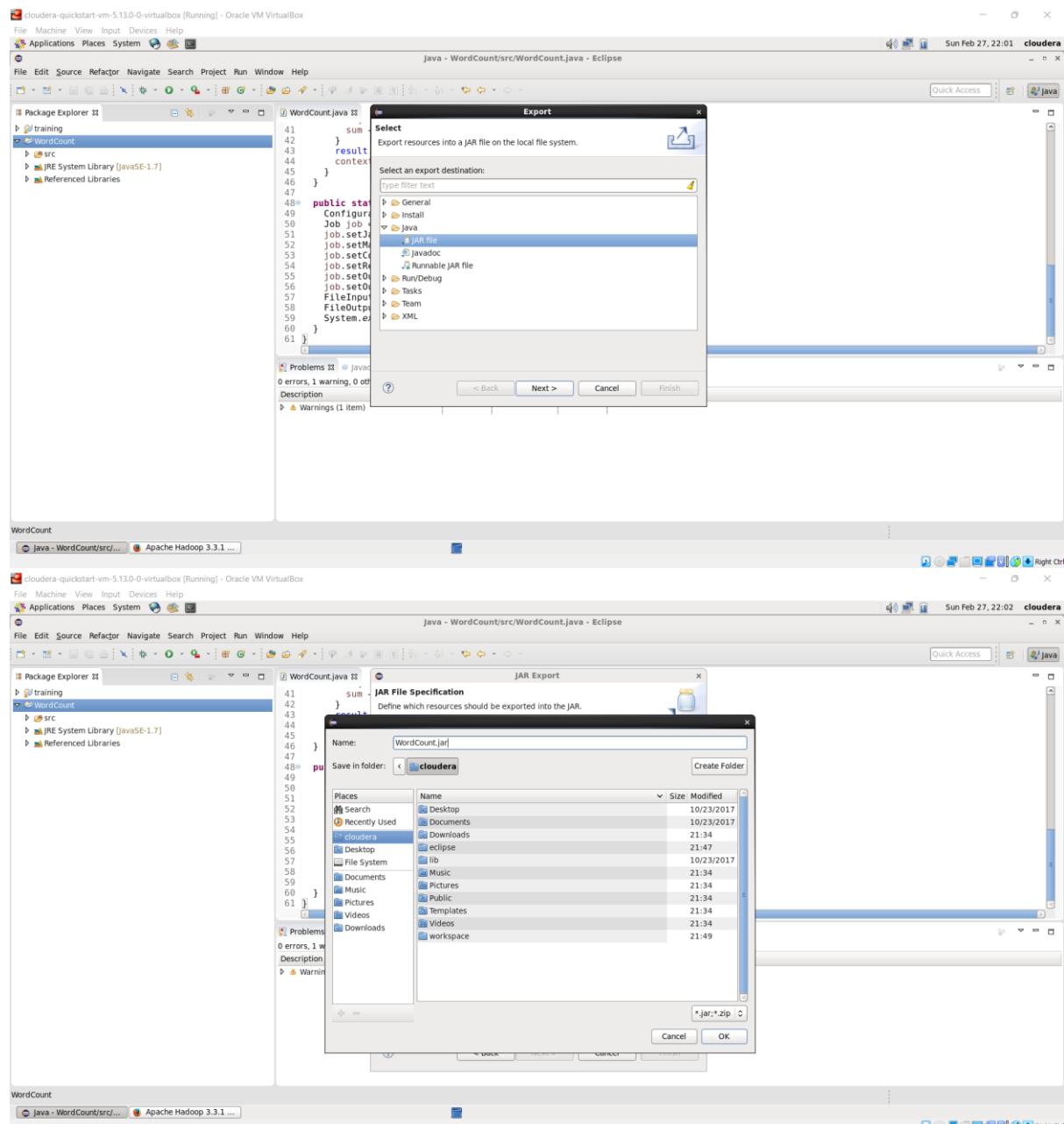
```

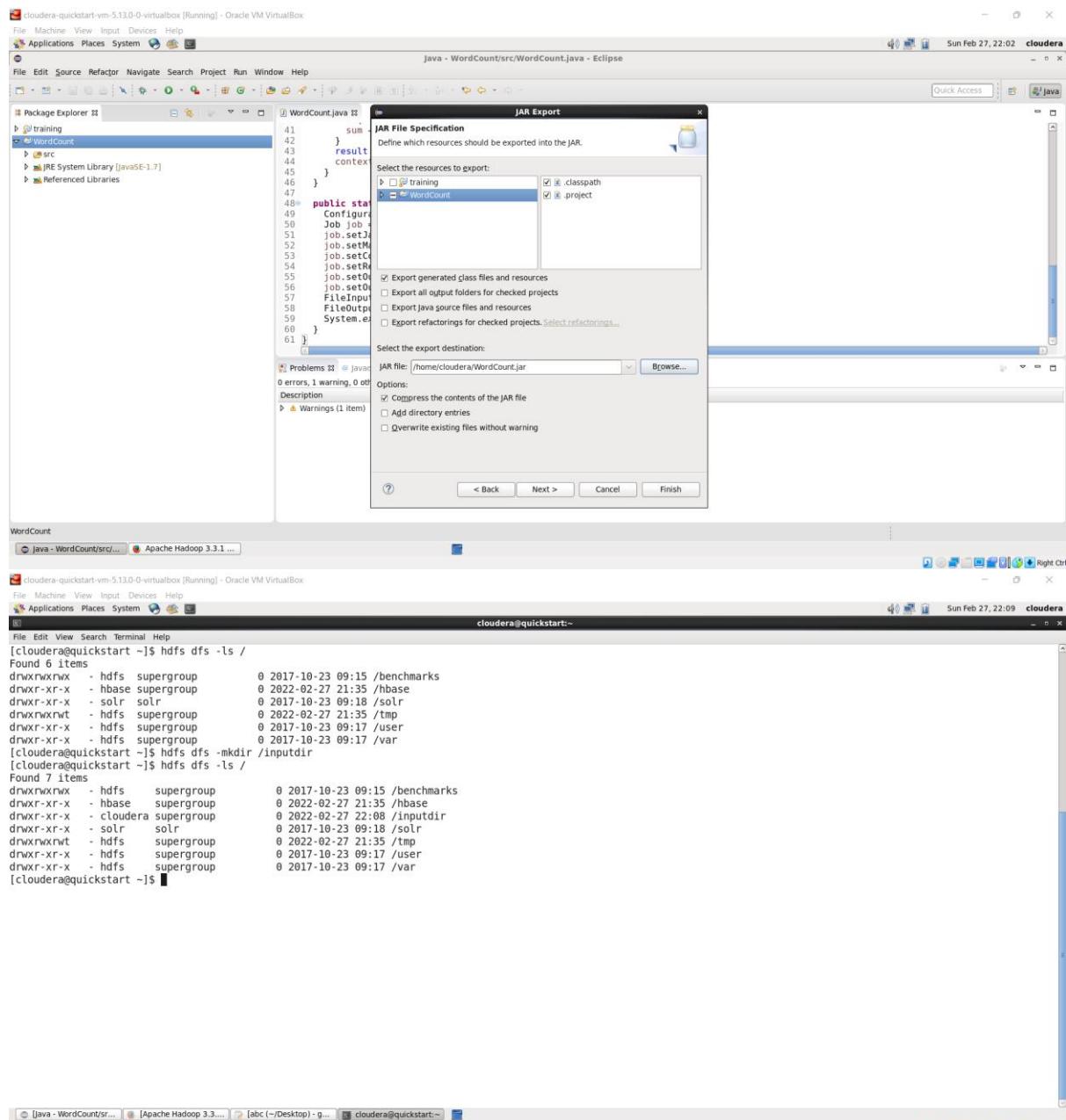
30 }
31
32 public static class IntSumReducer
33   extends Reducer<Text, IntWritable, Text, IntWritable> {
34   private IntWritable result = new IntWritable();
35
36   public void reduce(Text key, Iterable<IntWritable> values,
37                      Context context
38                      ) throws IOException, InterruptedException {
39     int sum = 0;
40     for (IntWritable val : values) {
41       sum += val.get();
42     }
43     result.set(sum);
44     context.write(key, result);
45   }
46 }
47
48 public static void main(String[] args) throws Exception {
49   Configuration conf = new Configuration();
50   Job job = Job.getInstance(conf, "word count");
51   job.setJarByClass(WordCount.class);
52   job.setMapperClass(TokenizerMapper.class);
53   job.setCombinerClass(IntSumReducer.class);
54   job.setReducerClass(IntSumReducer.class);
55   job.setOutputKeyClass(Text.class);
56   job.setOutputValueClass(IntWritable.class);
57   FileInputFormat.addInputPath(job, new Path(args[0]));
58   FileOutputFormat.setOutputPath(job, new Path(args[1]));
59   System.exit(job.waitForCompletion(true) ? 0 : 1);
60 }
61 }

```

Problems: 0 errors, 1 warning, 0 others

Java - WordCount/src... Apache Hadoop 3.3.1...





```

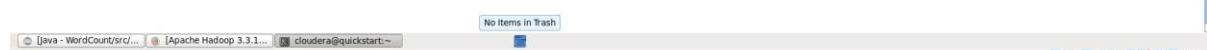
cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /inputdir
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:08 /inputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/abc /inputdir
22/02/27 22:10:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/abc /inputdir/
put: '/inputdir/abc': File exists
[cloudera@quickstart ~]$ 

```

```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /inputdir
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:08 /inputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/abc /inputdir/
22/02/27 22:10:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/abc /inputdir/
put: '/inputdir/abc': File exists
[cloudera@quickstart ~]$ hdfs dfs -cat /inputdir/abc
Hello Hrishikesh. This is Map Reduce practical.
[cloudera@quickstart ~]$ 

```



```

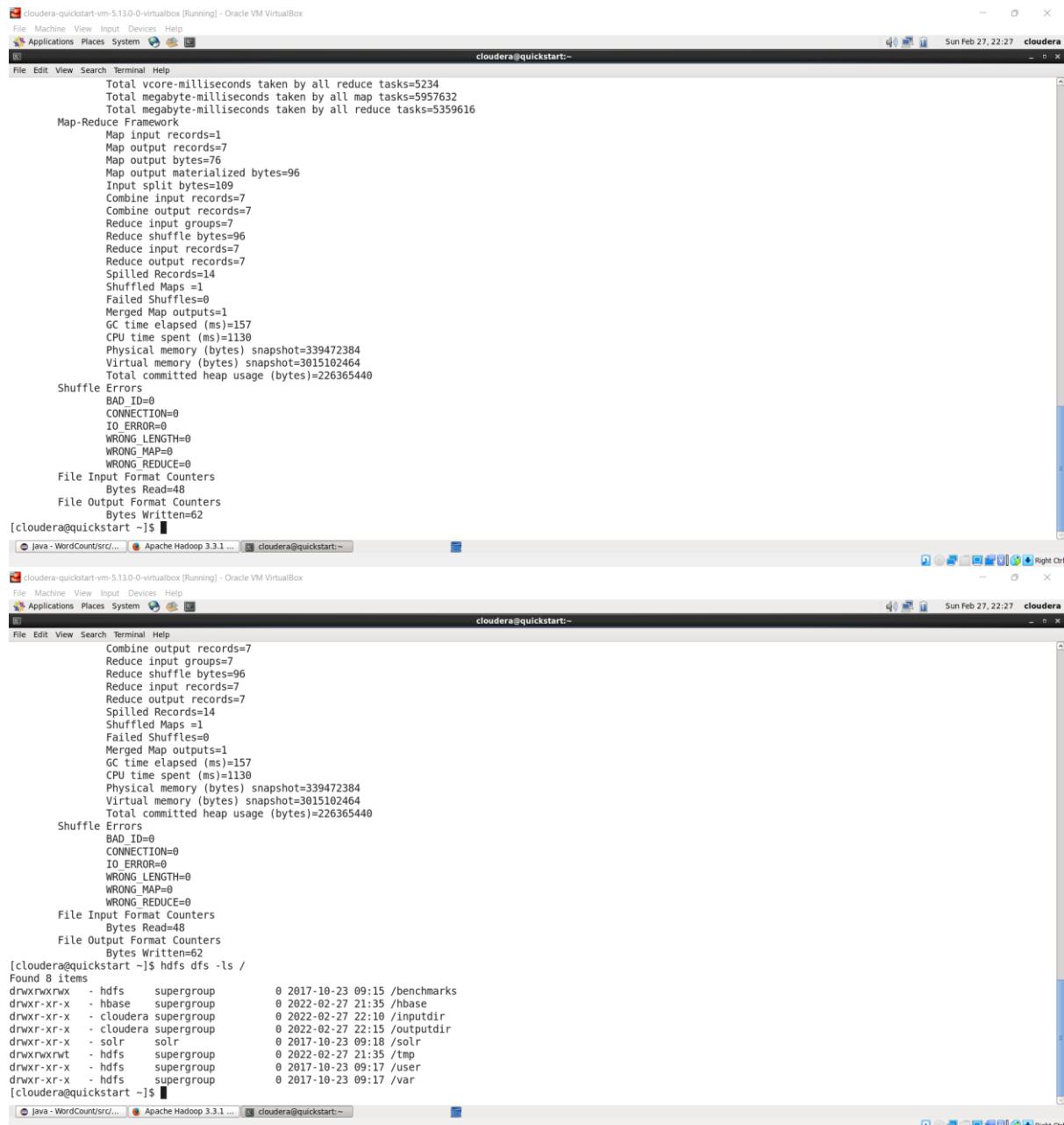
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~ Sun Feb 27, 22:19 cloudera
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /outputdir
22/02/27 22:14:57 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 22:14:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 22:14:58 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 22:14:58 INFO mapreduce.JobsSubmitter: number of splits:1
22/02/27 22:14:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646026409740_0001
22/02/27 22:14:59 INFO impl.YarnClientImpl: Submitted application application_1646026409740_0001
22/02/27 22:14:59 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646026409740_0001/
22/02/27 22:14:59 INFO mapreduce.Job: Running job: job_1646026409740_0001
22/02/27 22:15:07 INFO mapreduce.Job: Job job_1646026409740_0001 running in uber mode : false
22/02/27 22:15:07 INFO mapreduce.Job: map 0% reduce 0%
22/02/27 22:15:15 INFO mapreduce.Job: map 100% reduce 0%
22/02/27 22:15:24 INFO mapreduce.Job: map 100% reduce 100%
22/02/27 22:15:24 INFO mapreduce.Job: Job job_1646026409740_0001 completed successfully
22/02/27 22:15:25 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=96
FILE: Number of bytes written=286871
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=157
HDFS: Number of bytes written=62
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)=5818
Total time spent by all reduces in occupied slots (ms)=5234
Total time spent by all map tasks (ms)=5818
Total time spent by all reduce tasks (ms)=5234
Total vcore-milliseconds taken by all map tasks=5818

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~ Sun Feb 27, 22:27 cloudera
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /outputdir
22/02/27 22:14:57 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 22:14:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 22:14:58 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 22:14:58 INFO mapreduce.JobsSubmitter: number of splits:1
22/02/27 22:14:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646026409740_0001
22/02/27 22:14:59 INFO impl.YarnClientImpl: Submitted application application_1646026409740_0001
22/02/27 22:14:59 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646026409740_0001/
22/02/27 22:14:59 INFO mapreduce.Job: Running job: job_1646026409740_0001
22/02/27 22:15:07 INFO mapreduce.Job: map 0% reduce 0%
22/02/27 22:15:15 INFO mapreduce.Job: map 100% reduce 0%
22/02/27 22:15:24 INFO mapreduce.Job: map 100% reduce 100%
22/02/27 22:15:24 INFO mapreduce.Job: Job job_1646026409740_0001 completed successfully
22/02/27 22:15:25 INFO mapreduce.Job: Counters: 49
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)=5818
Total time spent by all reduces in occupied slots (ms)=5234
Total time spent by all map tasks (ms)=5818
Total time spent by all reduce tasks (ms)=5234
Total vcore-milliseconds taken by all map tasks=5818
Total vcore-milliseconds taken by all reduce tasks=5234
Total megabyte-milliseconds taken by all map tasks=5957632
Total megabyte-milliseconds taken by all reduce tasks=5359616

Map-Reduce Framework
Map input records=1
Map output records=7
Map output bytes=76
Map output materialized bytes=96
Input split bytes=109
Combine input records=7
Combine output records=7
Reduce input groups=7
Reduce shuffle bytes=96
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=157
CPU time spent (ms)=1130
Physical memory (bytes) snapshot=339472384
Virtual memory (bytes) snapshot=3015102464
Total committed heap usage (bytes)=226365440

Shuffle Errors
BAD_ID=0
CONNECTION=0

```



```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all reduce tasks=5234
Total megabyte-milliseconds taken by all map tasks=5957632
Total megabyte-milliseconds taken by all reduce tasks=5359616
Map-Reduce Framework
  Map Input records=1
  Map output records=7
  Map output bytes=76
  Map output materialized bytes=96
  Input split bytes=109
  Combine input records=7
  Combine output records=7
  Reduce input groups=7
  Reduce shuffle bytes=96
  Reduce input records=7
  Reduce output records=7
  Spilled Records=14
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=157
  CPU time spent (ms)=1130
  Physical memory (bytes) snapshot=339472384
  Virtual memory (bytes) snapshot=3015102464
  Total committed heap usage (bytes)=226365440
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=48
File Output Format Counters
  Bytes Written=62
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs  supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase  supergroup          0 2022-02-27 21:35 /base
drwxr-xr-x - cloudera supergroup        0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup        0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr   supergroup          0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs  supergroup          0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs  supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs  supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ 

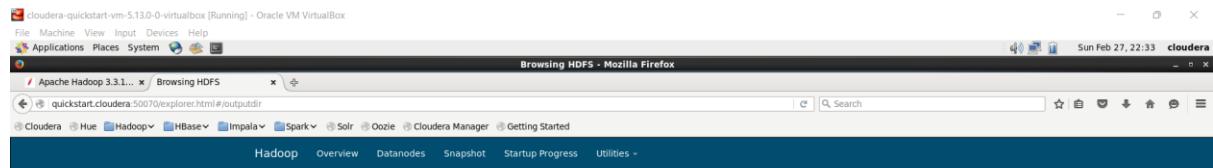
```

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
cloudera@quickstart:~$ 
File Edit View Search Terminal Help
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=157
CPU time spent (ms)=1130
Physical memory (bytes) snapshot=339472384
Virtual memory (bytes) snapshot=3015102464
Total committed heap usage (bytes)=226365440
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=48
File Output Format Counters
Bytes Written=62
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs    supergroup      0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase   supergroup      0 2022-02-27 21:35 /base
drwxr-xr-x - cloudera supergroup      0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup      0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr    solr           0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs   supergroup      0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs   supergroup      0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs   supergroup      0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -ls /outputdir
Found 2 items
-rw-r--r-- 1 cloudera supergroup      0 2022-02-27 22:15 /outputdir/_SUCCESS
-rw-r--r-- 1 cloudera supergroup 62 2022-02-27 22:15 /outputdir/part-r-00000
[cloudera@quickstart ~]$ 
Java - WordCount/src... Apache Hadoop 3.3.1 ... cloudera@quickstart:~ 

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
cloudera@quickstart:~$ 
File Edit View Search Terminal Help
Virtual memory (bytes) snapshot=3015102464
Total committed heap usage (bytes)=226365440
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=48
File Output Format Counters
Bytes Written=62
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs    supergroup      0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase   supergroup      0 2022-02-27 21:35 /base
drwxr-xr-x - cloudera supergroup      0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup      0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr    solr           0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs   supergroup      0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs   supergroup      0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs   supergroup      0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -ls /outputdir
Found 2 items
-rw-r--r-- 1 cloudera supergroup      0 2022-02-27 22:15 /outputdir/_SUCCESS
-rw-r--r-- 1 cloudera supergroup 62 2022-02-27 22:15 /outputdir/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /outputdir/part-r-00000
Hello 1
Hrishikesh. 1
Map 1
Reduce 1
This 1
is 1
practical. 1
[cloudera@quickstart ~]$ 
Java - WordCount/src... Apache Hadoop 3.3.1 ... cloudera@quickstart:~ 

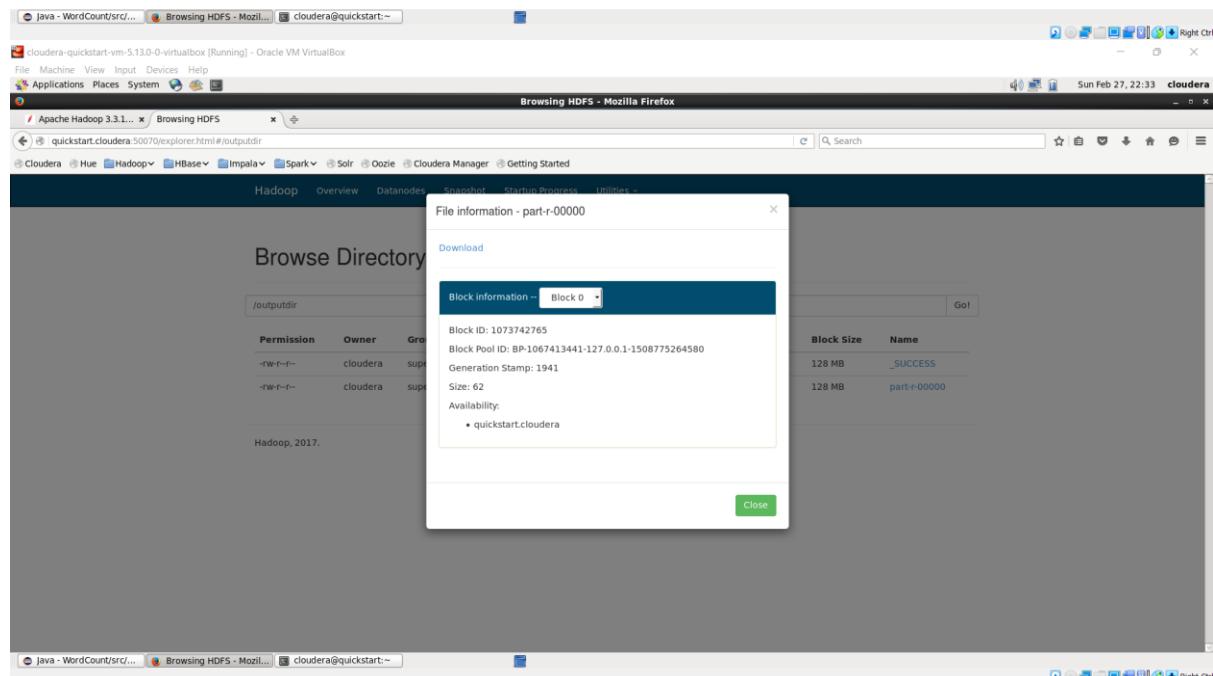
```

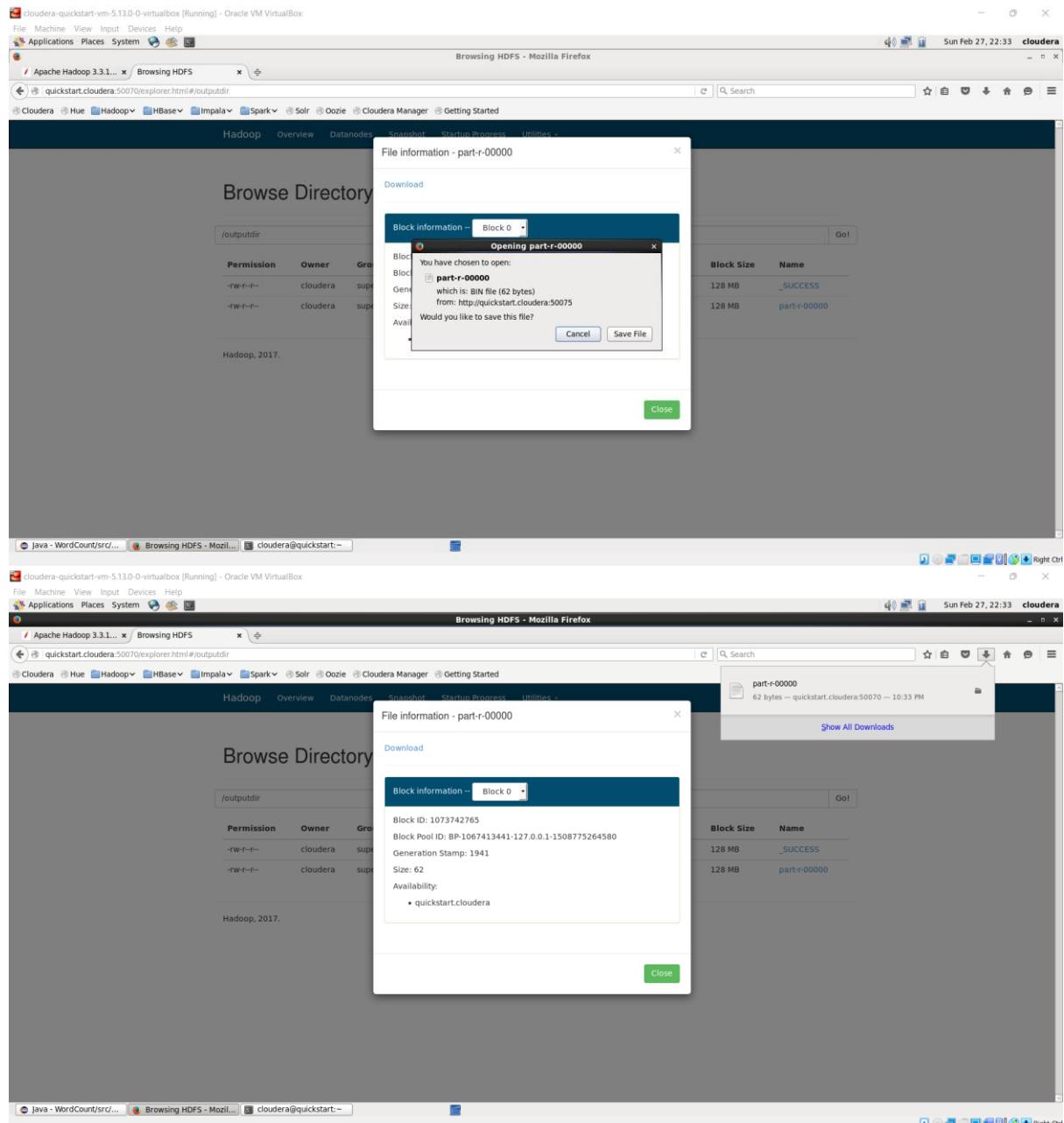


## Browse Directory

Browse Directory							
/outputdir							
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	supergroup	0 B	Sun Feb 27 22:15:23 -0800 2022	1	128 MB	_SUCCESS
-rw-r--r--	cloudera	supergroup	62 B	Sun Feb 27 22:15:22 -0800 2022	1	128 MB	part-r-00000

Hadoop, 2017.





```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
Sun Feb 27, 22:36 cloudera

[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /op1
22/02/27 22:35:41 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 22:35:41 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 22:35:42 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 22:35:42 INFO mapreduce.JobSubmitter: number of splits:1
22/02/27 22:35:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646026409740_0002
22/02/27 22:35:42 INFO impl.YarnClientImpl: Submitted application application_1646026409740_0002/
22/02/27 22:35:42 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646026409740_0002/
22/02/27 22:35:42 INFO mapreduce.Job: Running job: job_1646026409740_0002
22/02/27 22:35:50 INFO mapreduce.Job: Job job_1646026409740_0002 running in uber mode : false
22/02/27 22:35:50 INFO mapreduce.Job: map 0% reduce 0%
22/02/27 22:36:06 INFO mapreduce.Job: map 100% reduce 100%
22/02/27 22:36:07 INFO mapreduce.Job: Job job_1646026409740_0002 completed successfully
22/02/27 22:36:07 INFO mapreduce.Job: Counters: 49

File System Counters
FILE: Number of bytes read=96
FILE: Number of bytes written=286859
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=157
HDFS: Number of bytes written=62
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)=6002
Total time spent by all reduces in occupied slots (ms)=4780
Total time spent by all map tasks (ms)=6002
Total time spent by all reduce tasks (ms)=4780
Total vcore-milliseconds taken by all map tasks=6002

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
Sun Feb 27, 22:36 cloudera

[cloudera@quickstart ~]$ Total vcore-milliseconds taken by all reduce tasks=4780
Total megabyte-milliseconds taken by all map tasks=6146048
Total megabyte-milliseconds taken by all reduce tasks=4894720

Map-Reduce Framework
Map input records=1
Map output records=7
Map output bytes=76
Map output materialized bytes=96
Input split bytes=109
Combine input records=7
Combine output records=7
Reduce input groups=7
Reduce shuffle bytes=96
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=165
CPU time spent (ms)=160
Physical memory (bytes) snapshot=341397504
Virtual memory (bytes) snapshot=3015102464
Total committed heap usage (bytes)=226365440

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=48
File Output Format Counters
Bytes Written=62

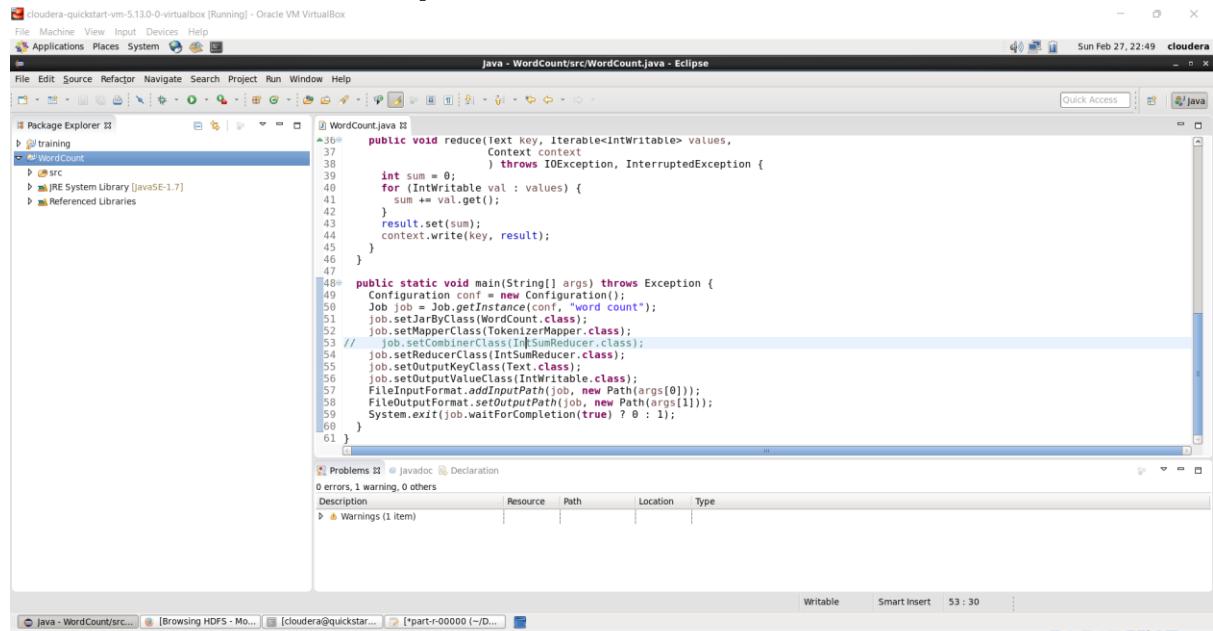
[cloudera@quickstart ~]$
```

The screenshot shows a Linux desktop environment with several windows open. In the foreground, a terminal window titled "gedit" displays the contents of a file named "part-r-00000". The file contains the following text:

```
Hello  
Hrishikesh, 1  
Map 1  
Reduce 1  
This 1  
is 1  
practical. 1$
```

Below the terminal, the taskbar shows other open applications: "Java - WordCount/sr...", "Browsing HDFS - Mo...", "[cloudera@quickstar...]", and "[part-r-00000 (~/Do...)]". The status bar at the bottom indicates "Plain Text" mode, a tab width of 8, and the current position as Ln 7, Col 19.

## Implementation of WordCount problem using Hadoop MapReduce (With Combiner) in Eclipse:

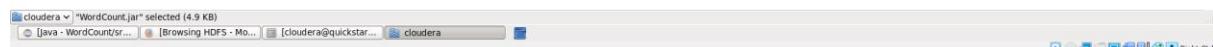


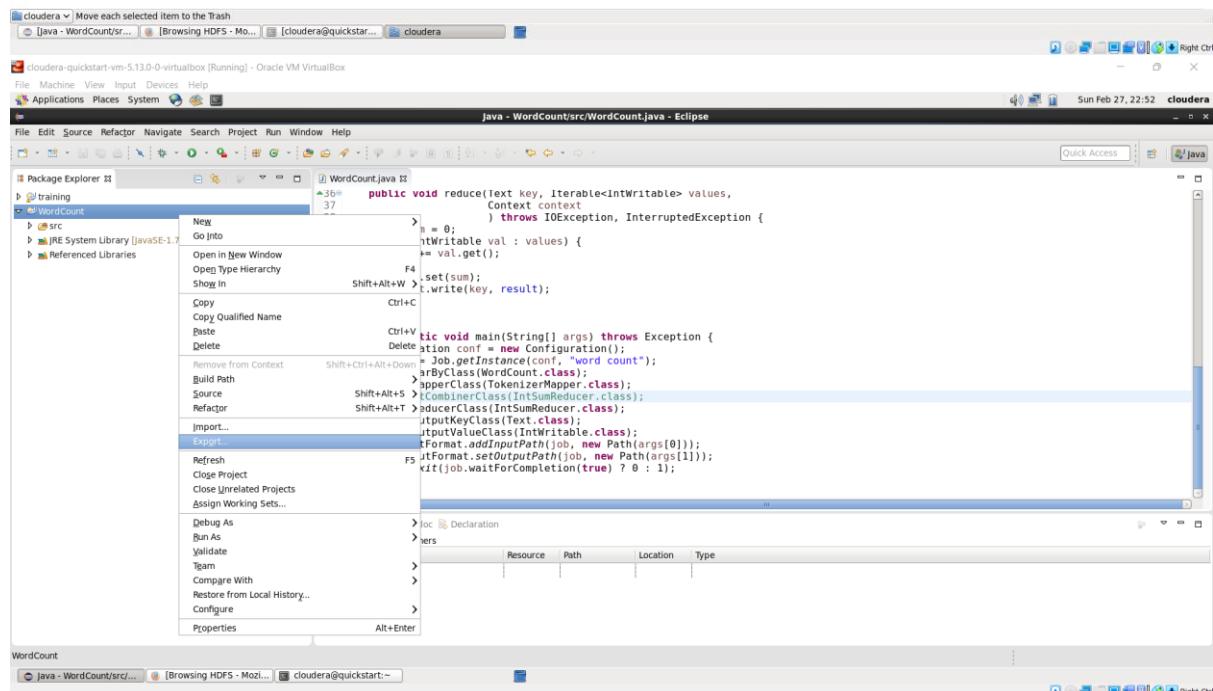
```

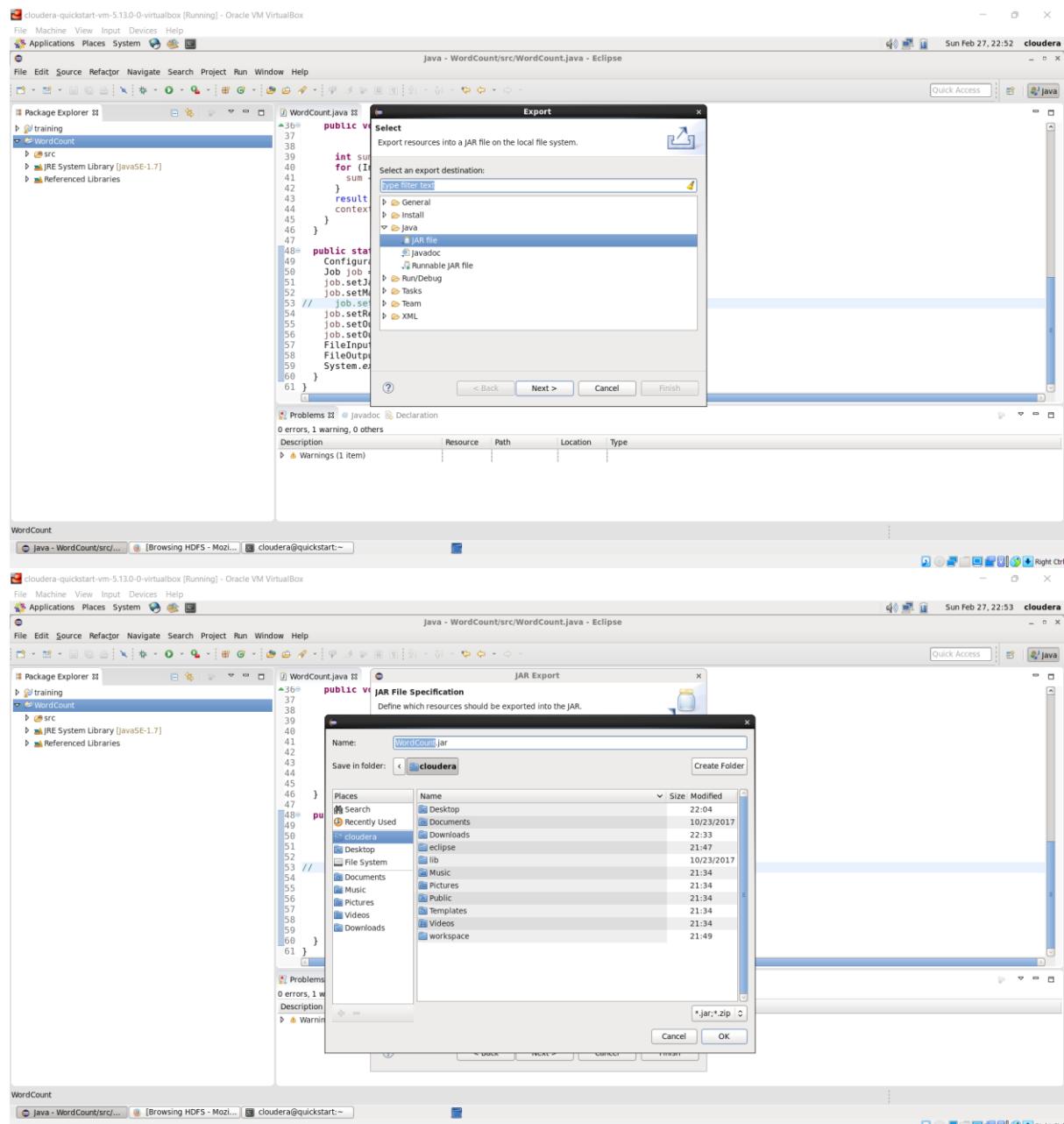
public void reduce(Text key, Iterable<IntWritable> values,
                    Context context)
                    throws IOException, InterruptedException {
    int sum = 0;
    for (IntWritable val : values) {
        sum += val.get();
    }
    result.set(sum);
    context.write(key, result);
}

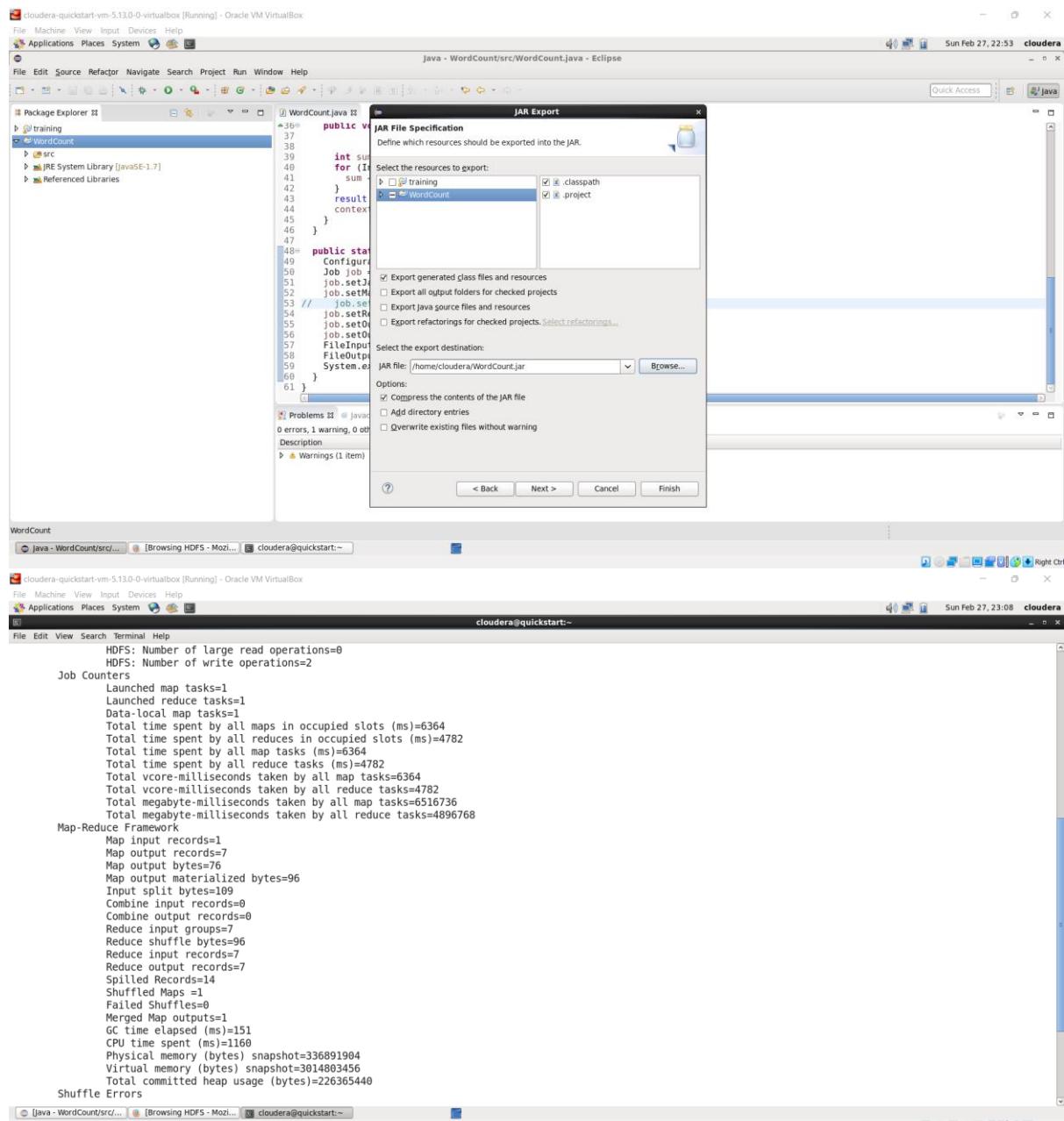
public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    // job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(TextOutputMapper.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}

```







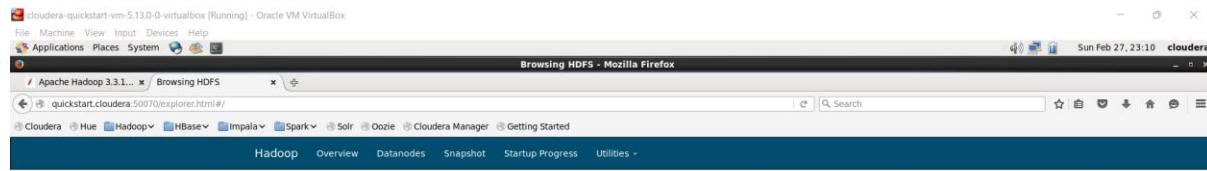


```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~ Sun Feb 27, 23:08 cloudera
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all reduce tasks=4782
Total megabyte-milliseconds taken by all map tasks=6516736
Total megabyte-milliseconds taken by all reduce tasks=4896768
Map-Reduce Framework
Map Input records=1
Map output records=7
Map output bytes=76
Map output materialized bytes=96
Input split bytes=109
Combine input records=0
Combine output records=0
Reduce input groups=7
Reduce shuffle bytes=96
Reduce input records=7
Reduce output records=7
Spilled Records=14
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=151
CPU time spent (ms)=1160
Physical memory (bytes) snapshot=336891904
Virtual memory (bytes) snapshot=3014803456
Total committed heap usage (bytes)=226365440
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=48
File Output Format Counters
Bytes Written=62
[cloudera@quickstart ~]$ 

Java - WordCount/src... [Browsing HDFS - Mozilla Firefox cloudera@quickstart:~]
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~ Sun Feb 27, 23:08 cloudera
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ ls
cloudera-manager Desktop Downloads enterprise-deployment.json kerberos Music Pictures Templates WordCount.jar
cm_api.py Documents eclipse express-deployment.json lib parcels Public Videos workspace
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /op2
22/02/27 23:07:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 23:07:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 23:07:14 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 23:07:14 INFO mapreduce.JobSubmitter: number of splits:1
22/02/27 23:07:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646026409740_0003
22/02/27 23:07:15 INFO impl.YarnClientImpl: Submitted application application_1646026409740_0003
22/02/27 23:07:15 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646026409740_0003/
22/02/27 23:07:15 INFO mapreduce.Job: Running job: job_1646026409740_0003
22/02/27 23:07:25 INFO mapreduce.Job: Job job_1646026409740_0003 running in uber mode : false
22/02/27 23:07:25 INFO mapreduce.Job: map 0% reduce 0%
22/02/27 23:07:34 INFO mapreduce.Job: map 100% reduce 0%
22/02/27 23:07:41 INFO mapreduce.Job: map 100% reduce 100%
22/02/27 23:07:41 INFO mapreduce.Job: Job job_1646026409740_0003 completed successfully
22/02/27 23:07:41 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=96
FILE: Number of bytes written=286507
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=157
HDFS: Number of bytes written=62
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)=6364
Total time spent by all reduces in occupied slots (ms)=4782
[cloudera@quickstart ~]$ 
Java - WordCount/src... [Browsing HDFS - Mozilla Firefox cloudera@quickstart:~]

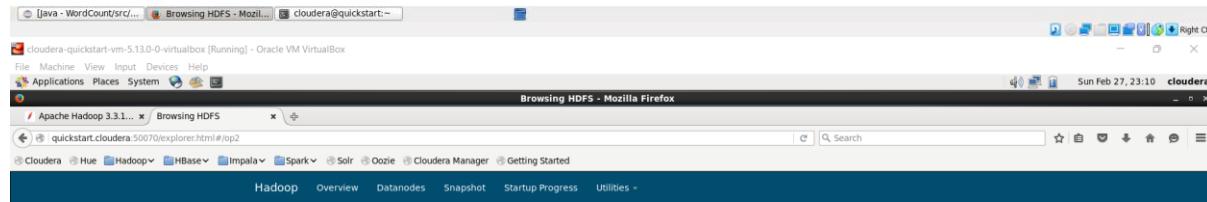
```



## Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	hdfs	supergroup	0 B	Mon Oct 23 09:15:43 -0700 2017	0	0 B	benchmarks
drwxr-xr-x	hbase	supergroup	0 B	Sun Feb 27 21:35:19 -0800 2022	0	0 B	hbase
drwxr-xr-x	cloudera	supergroup	0 B	Sun Feb 27 22:10:16 -0800 2022	0	0 B	inputdir
drwxr-xr-x	cloudera	supergroup	0 B	Sun Feb 27 22:36:05 -0800 2022	0	0 B	op1
drwxr-xr-x	cloudera	supergroup	0 B	Sun Feb 27 23:07:39 -0800 2022	0	0 B	op2
drwxr-xr-x	cloudera	supergroup	0 B	Sun Feb 27 22:15:23 -0800 2022	0	0 B	outputdir
drwxr-xr-x	solr	solr	0 B	Mon Oct 23 09:18:01 -0700 2017	0	0 B	solr
drwxrwxrwt	hdfs	supergroup	0 B	Sun Feb 27 21:35:18 -0800 2022	0	0 B	tmp
drwxr-xr-x	hdfs	supergroup	0 B	Mon Oct 23 09:17:33 -0700 2017	0	0 B	user
drwxr-xr-x	hdfs	supergroup	0 B	Mon Oct 23 09:17:24 -0700 2017	0	0 B	var

Hadoop, 2017.



## Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	supergroup	0 B	Sun Feb 27 23:07:39 -0800 2022	1	128 MB	_SUCCESS
-rw-r--r--	cloudera	supergroup	62 B	Sun Feb 27 23:07:39 -0800 2022	1	128 MB	part-r-00000

Hadoop, 2017.

