ICE-2

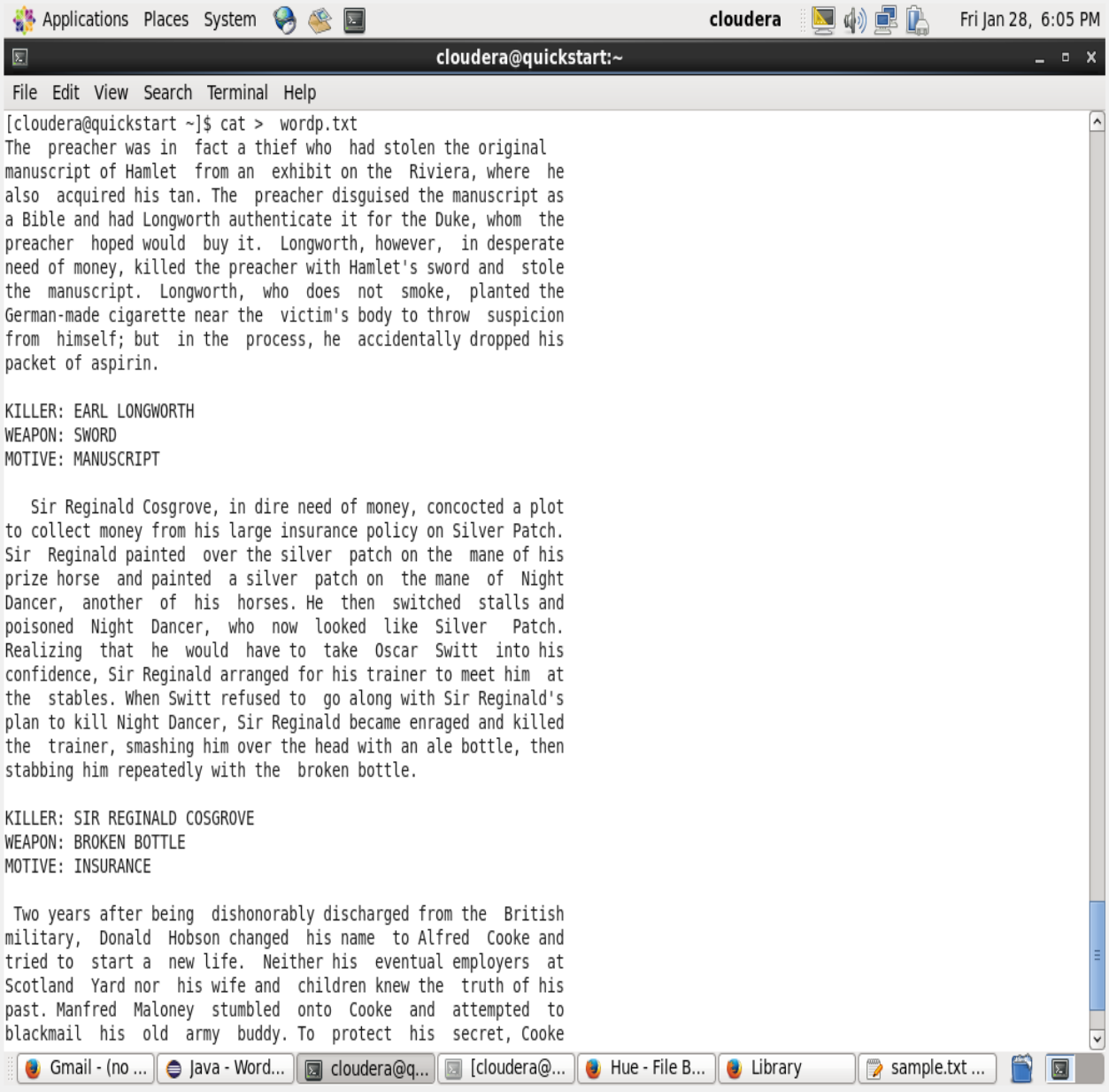
Krishna Reddy Chaduvu

UNT id:11551974

1.By following the step-by-step instructions given,implemented and successfully executed the wordcount program and explained in detail below.

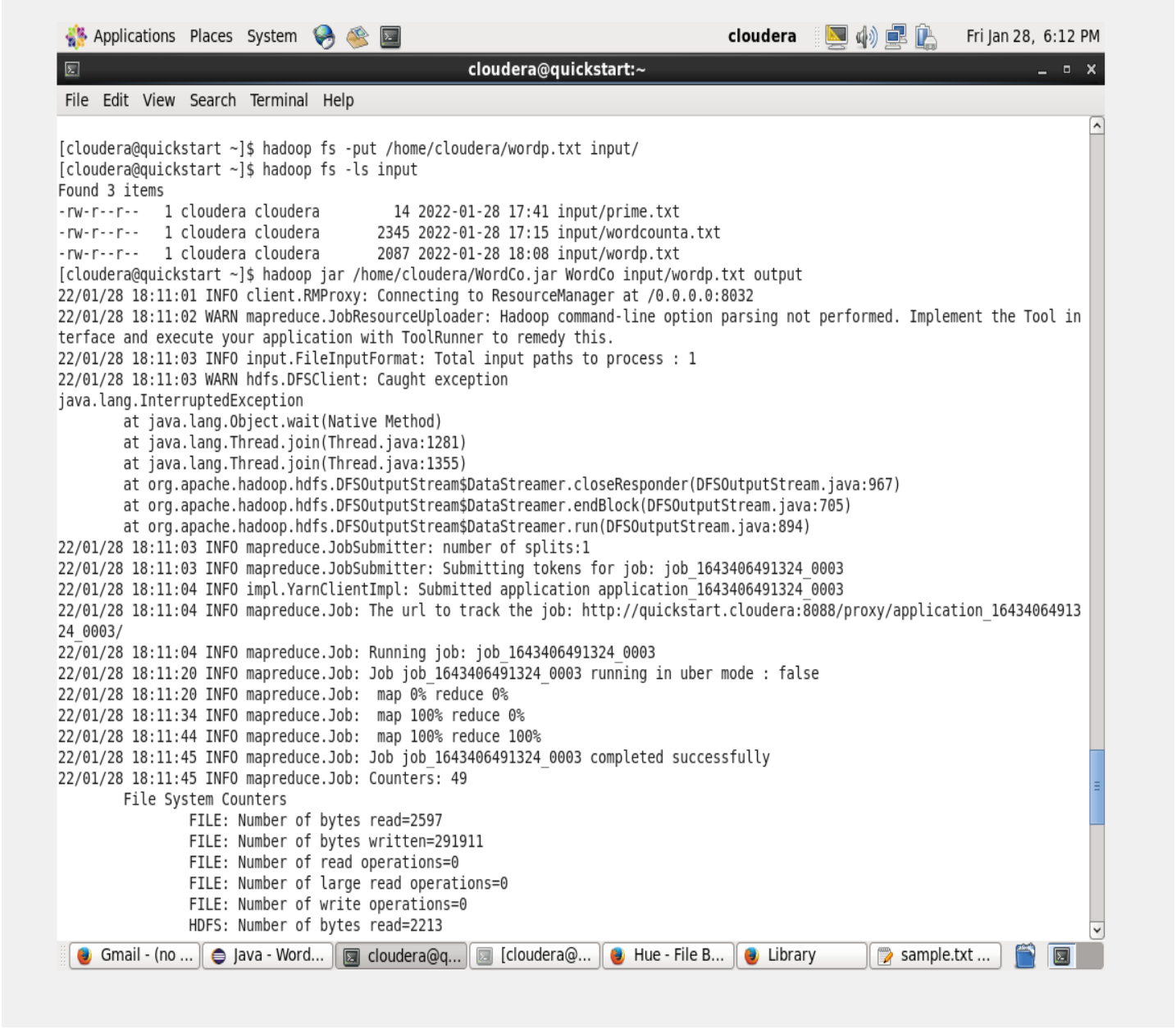
Step:1

Follwed by the instructions given,created a new file called wordp.txt using cat command and given the input data.

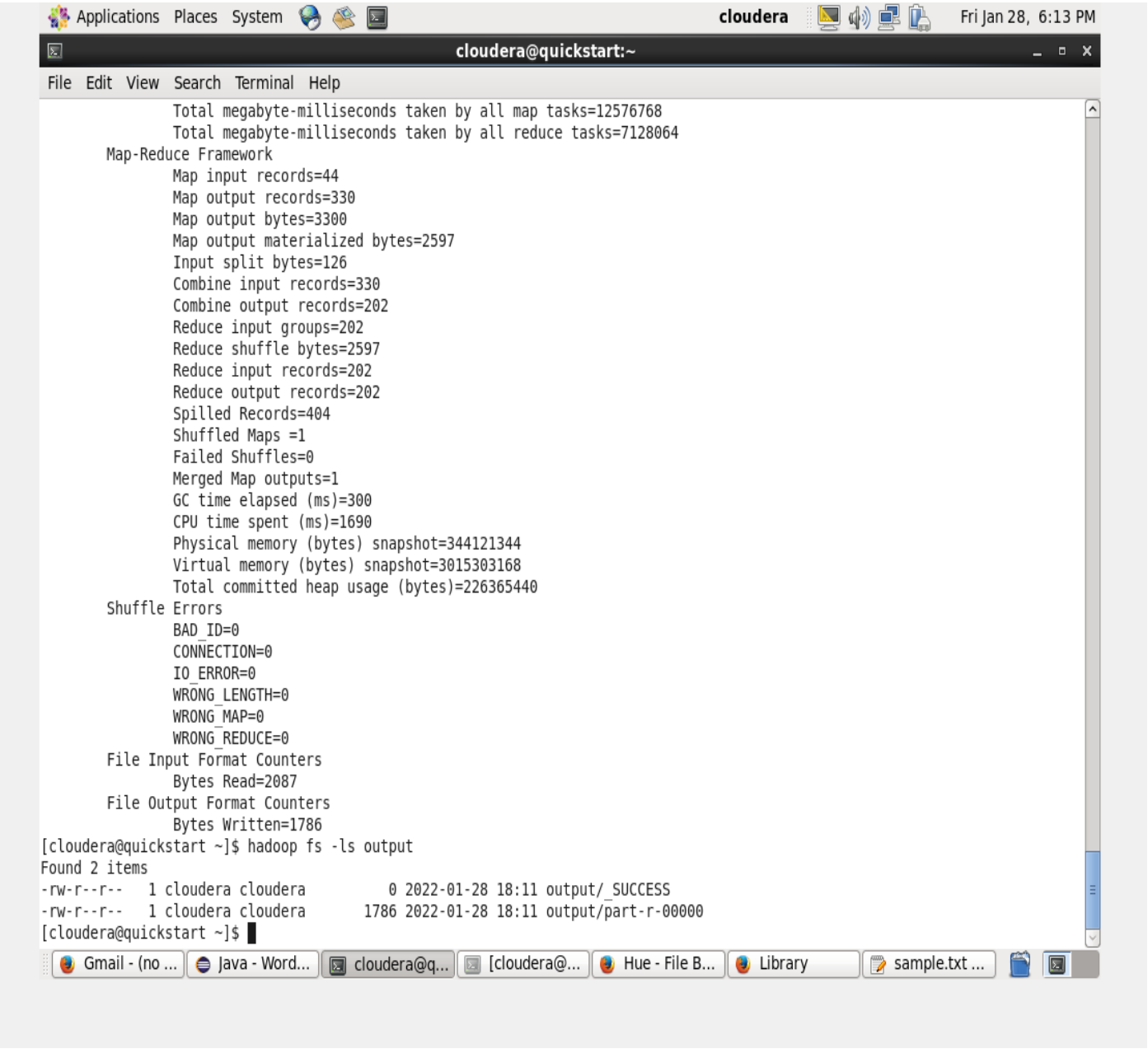

Applications Places System
cloudera
Fri Jan 28, 6:05 PM
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ cat > wordp.txt
The preacher was in fact a thief who had stolen the original manuscript of Hamlet from an exhibit on the Riviera, where he also acquired his tan. The preacher disguised the manuscript as a Bible and had Longworth authenticate it for the Duke, whom the preacher hoped would buy it. Longworth, however, in desperate need of money, killed the preacher with Hamlet's sword and stole the manuscript. Longworth, who does not smoke, planted the German-made cigarette near the victim's body to throw suspicion from himself; but in the process, he accidentally dropped his packet of aspirin.
KILLER: EARL LONGWORTH
WEAPON: SWORD
MOTIVE: MANUSCRIPT
Sir Reginald Cosgrove, in dire need of money, concocted a plot to collect money from his large insurance policy on Silver Patch. Sir Reginald painted over the silver patch on the mane of his prize horse and painted a silver patch on the mane of Night Dancer, another of his horses. He then switched stalls and poisoned Night Dancer, who now looked like Silver Patch. Realizing that he would have to take Oscar Switt into his confidence, Sir Reginald arranged for his trainer to meet him at the stables. When Switt refused to go along with Sir Reginald's plan to kill Night Dancer, Sir Reginald became enraged and killed the trainer, smashing him over the head with an ale bottle, then stabbing him repeatedly with the broken bottle.
KILLER: SIR REGINALD COSGROVE
WEAPON: BROKEN BOTTLE
MOTIVE: INSURANCE
Two years after being dishonorably discharged from the British military, Donald Hobson changed his name to Alfred Cooke and tried to start a new life. Neither his eventual employers at Scotland Yard nor his wife and children knew the truth of his past. Manfred Maloney stumbled onto Cooke and attempted to blackmail his old army buddy. To protect his secret, Cooke
Gmail - (no ...
Java - Word...
cloudera@q...
[cloudera@...
Hue - File B...
Library
sample.txt ...

Step:2

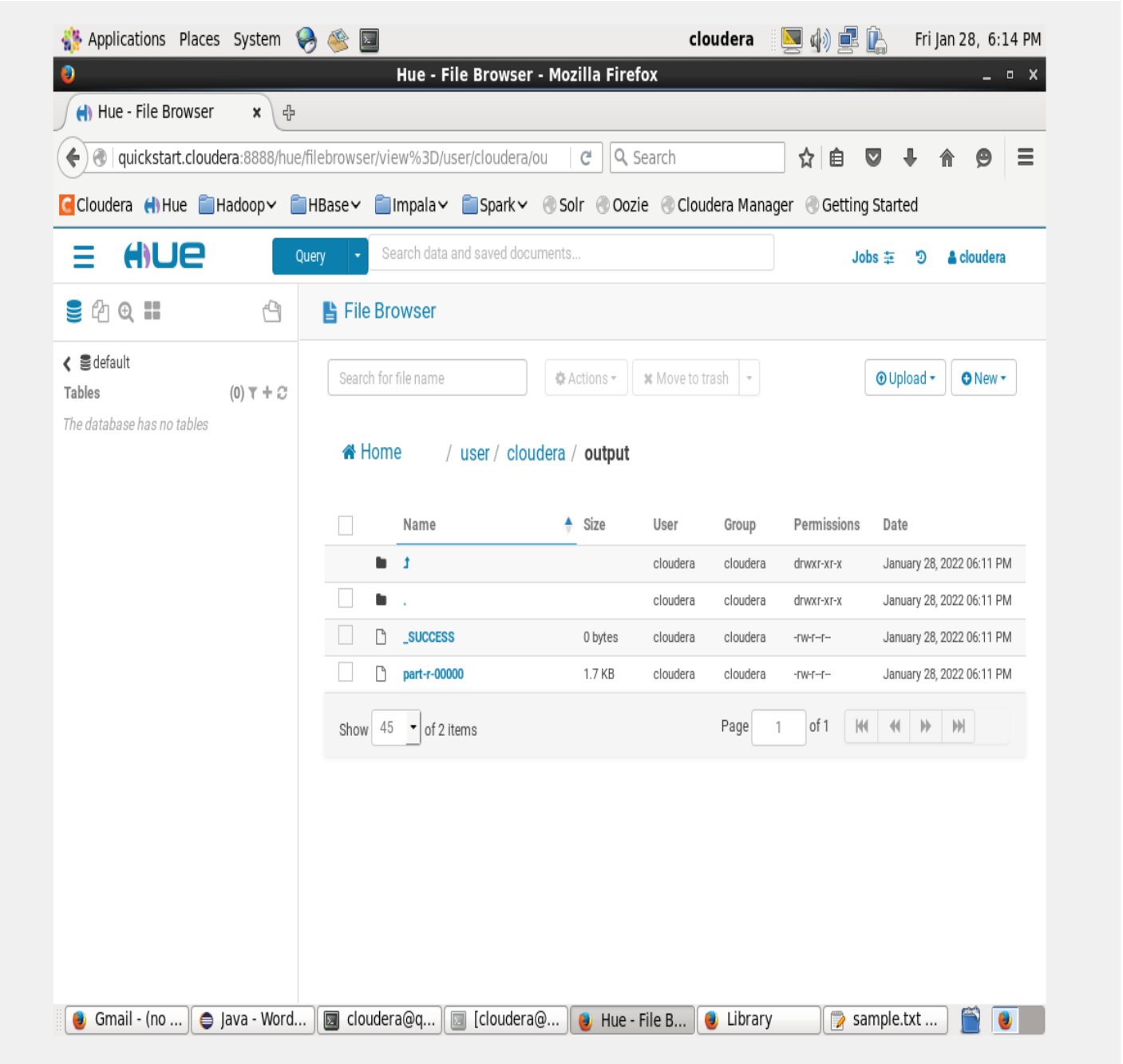
Loading the wordp.txt file into hadoop hdfs and implementing the mapreducing alogorithm.



Map Reducing is done and the given command is success and two items are found

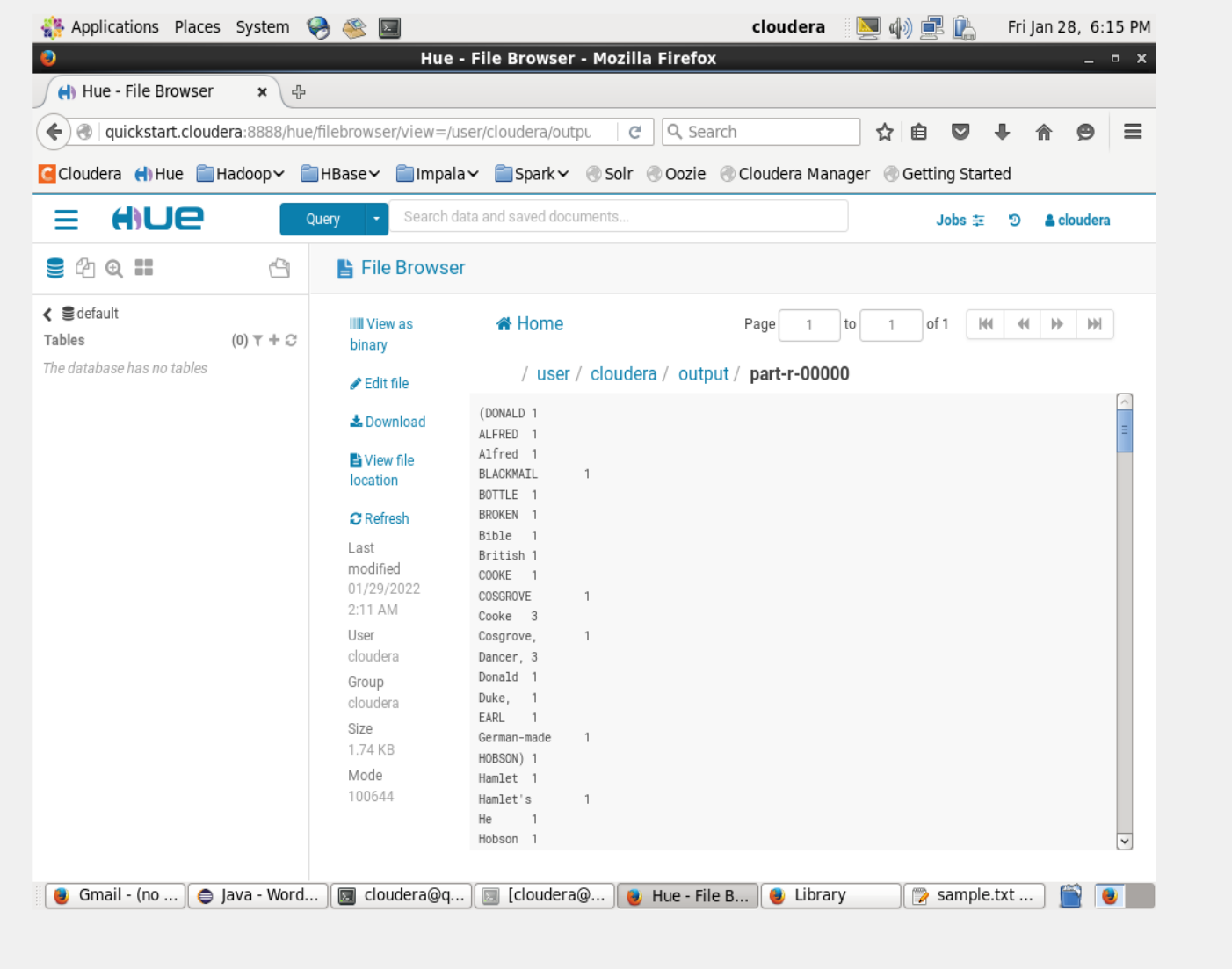


Visualization of the items found.



Step:3

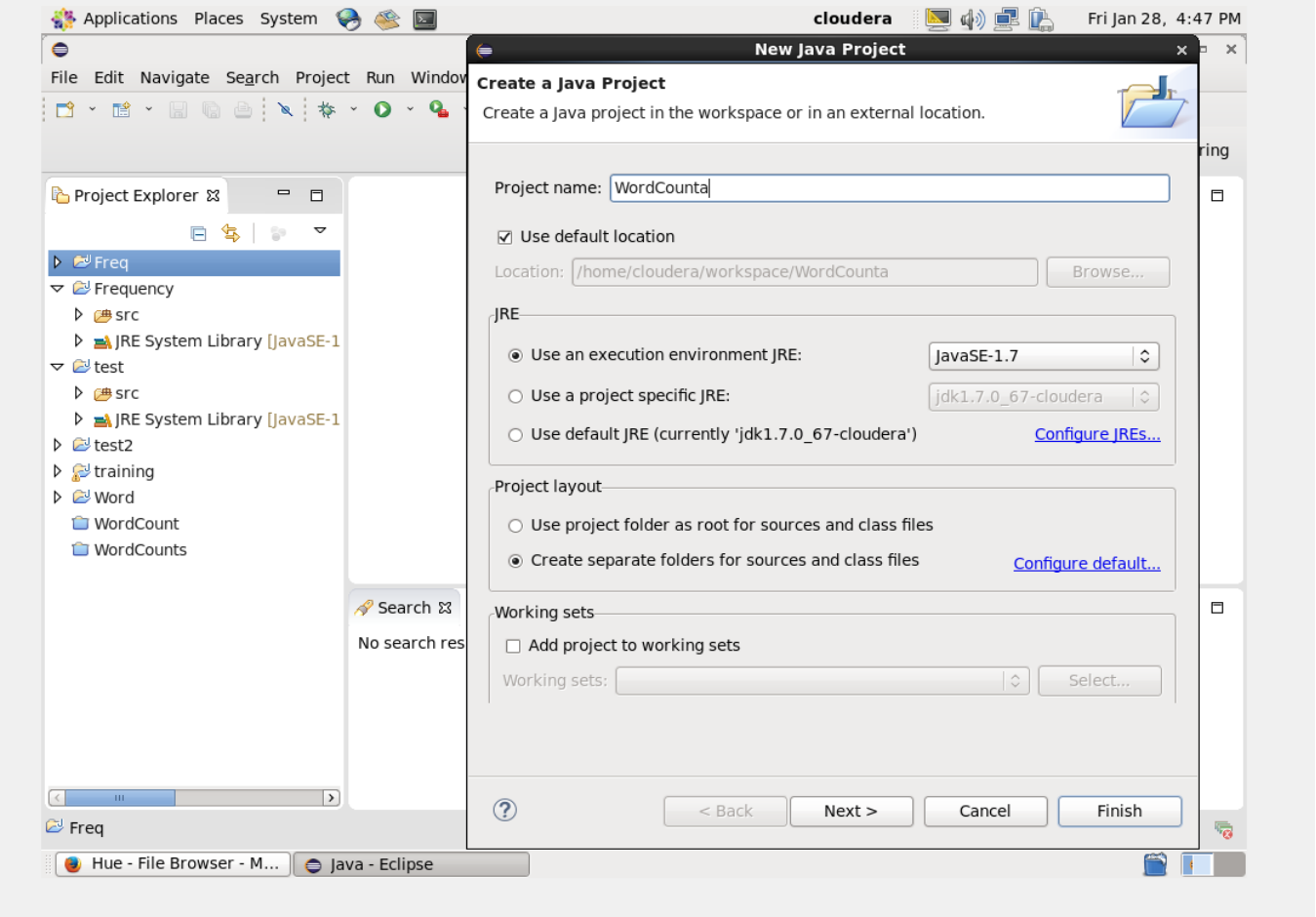
Frequency count of words, of the given input is done and the output is below.



2.Counting the frequency of words in given text file that starts with ‘a’.

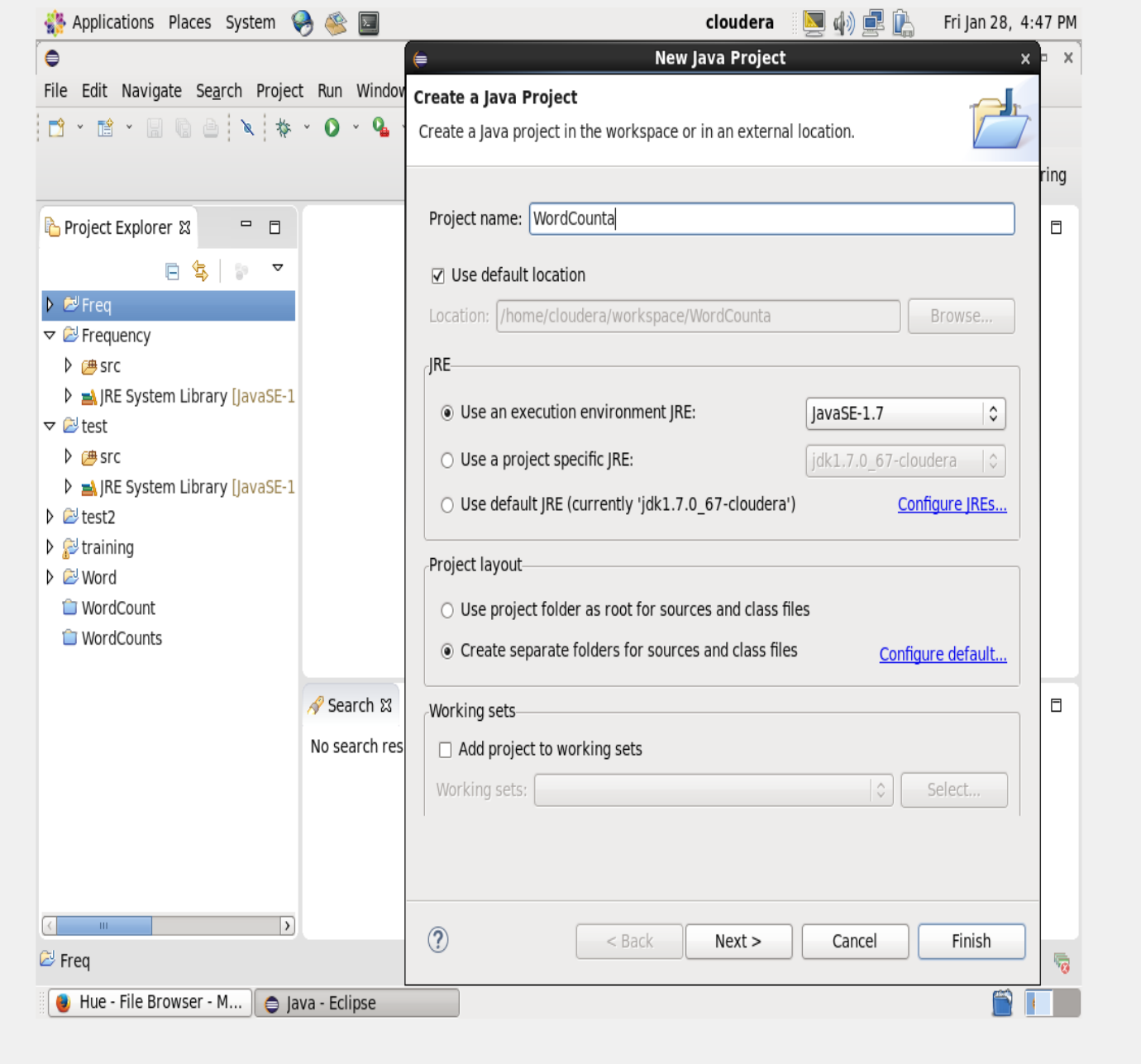
Step:1

Creating a new java project and naming it as WordCounta

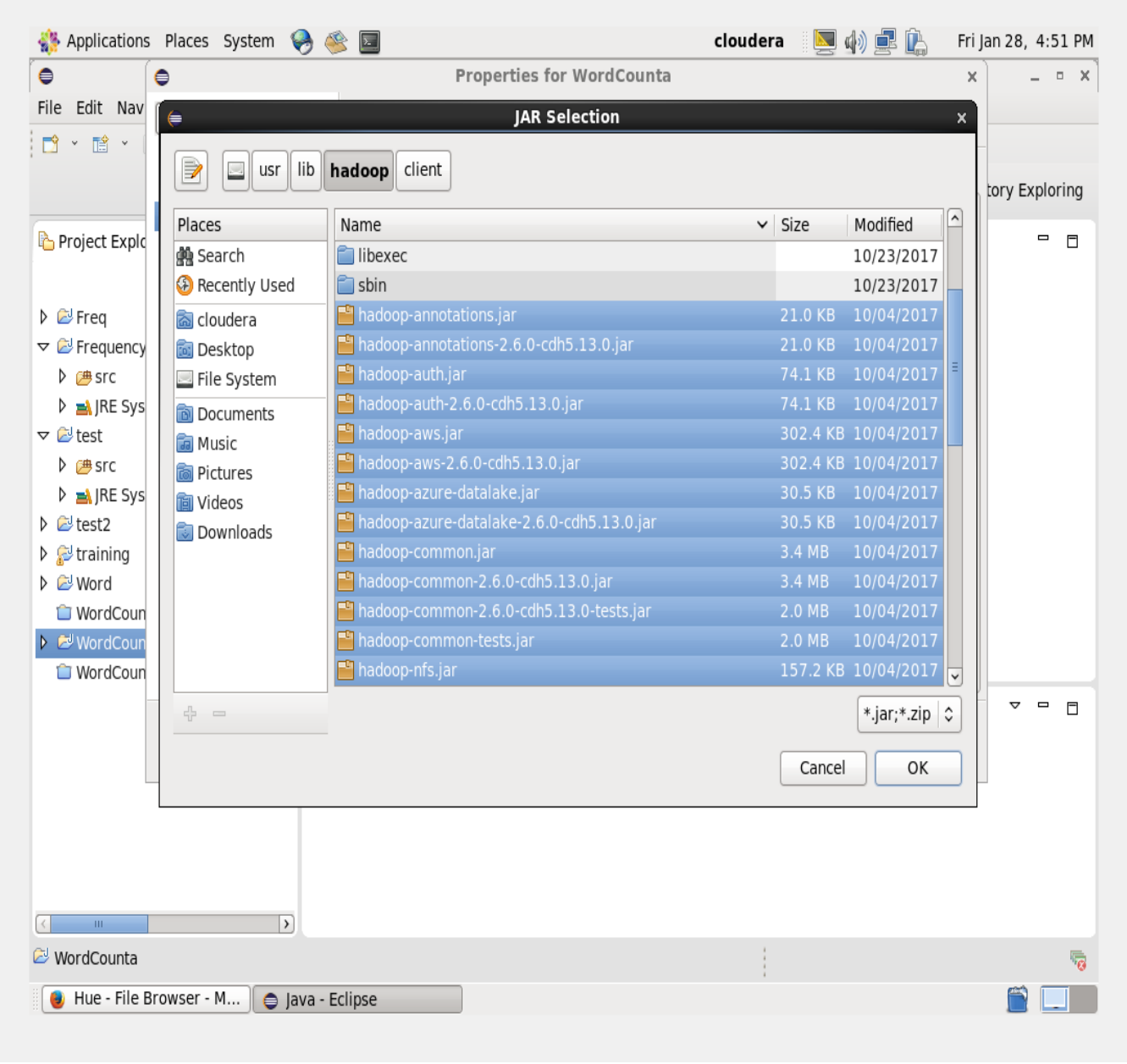


Step:2

Adding the external JARs files

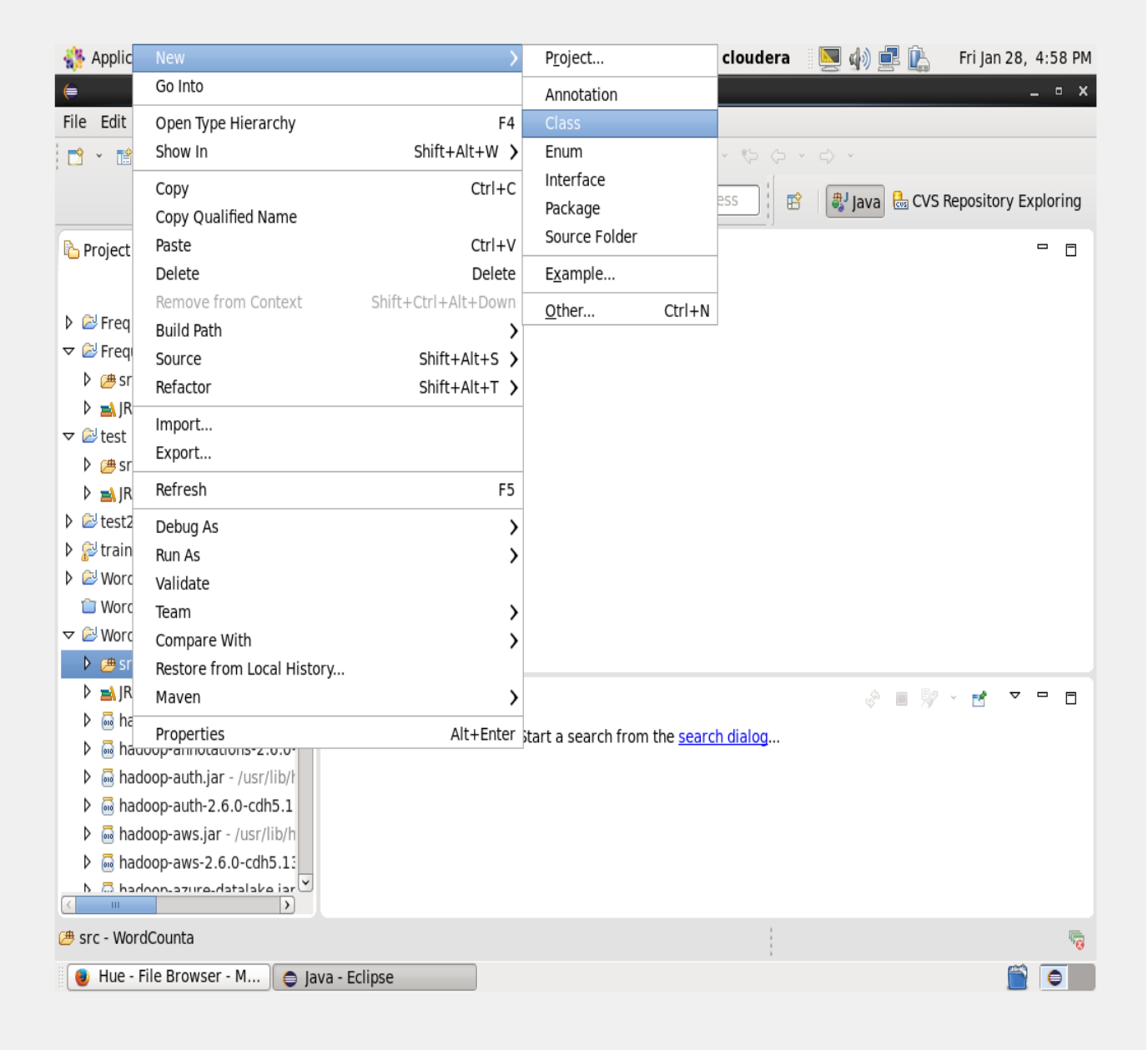


Adding the JARs files of hadoop and client

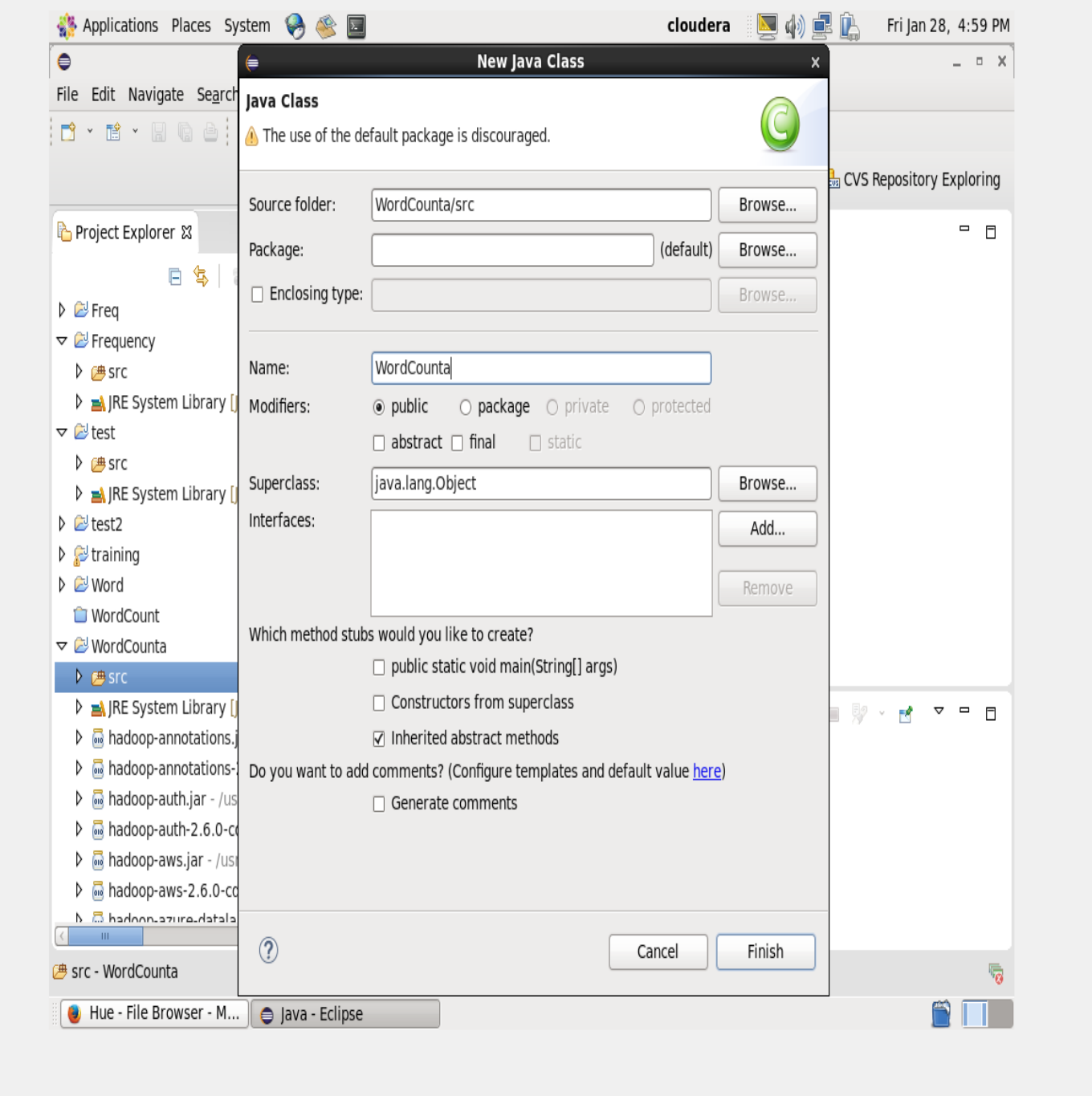


Step:3

Creating the class files. Right click on source,New>Class



Naming the file as WordCounta

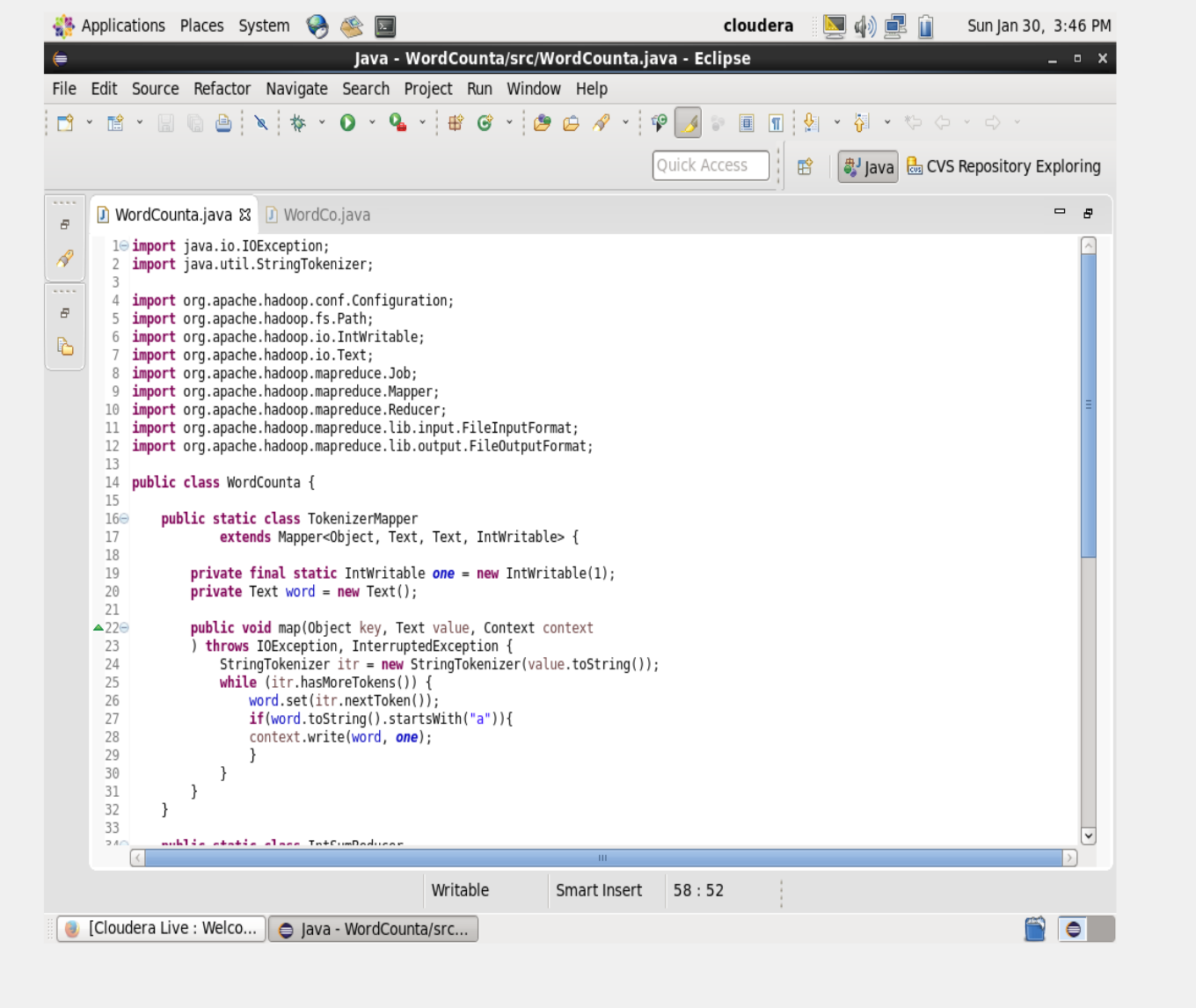


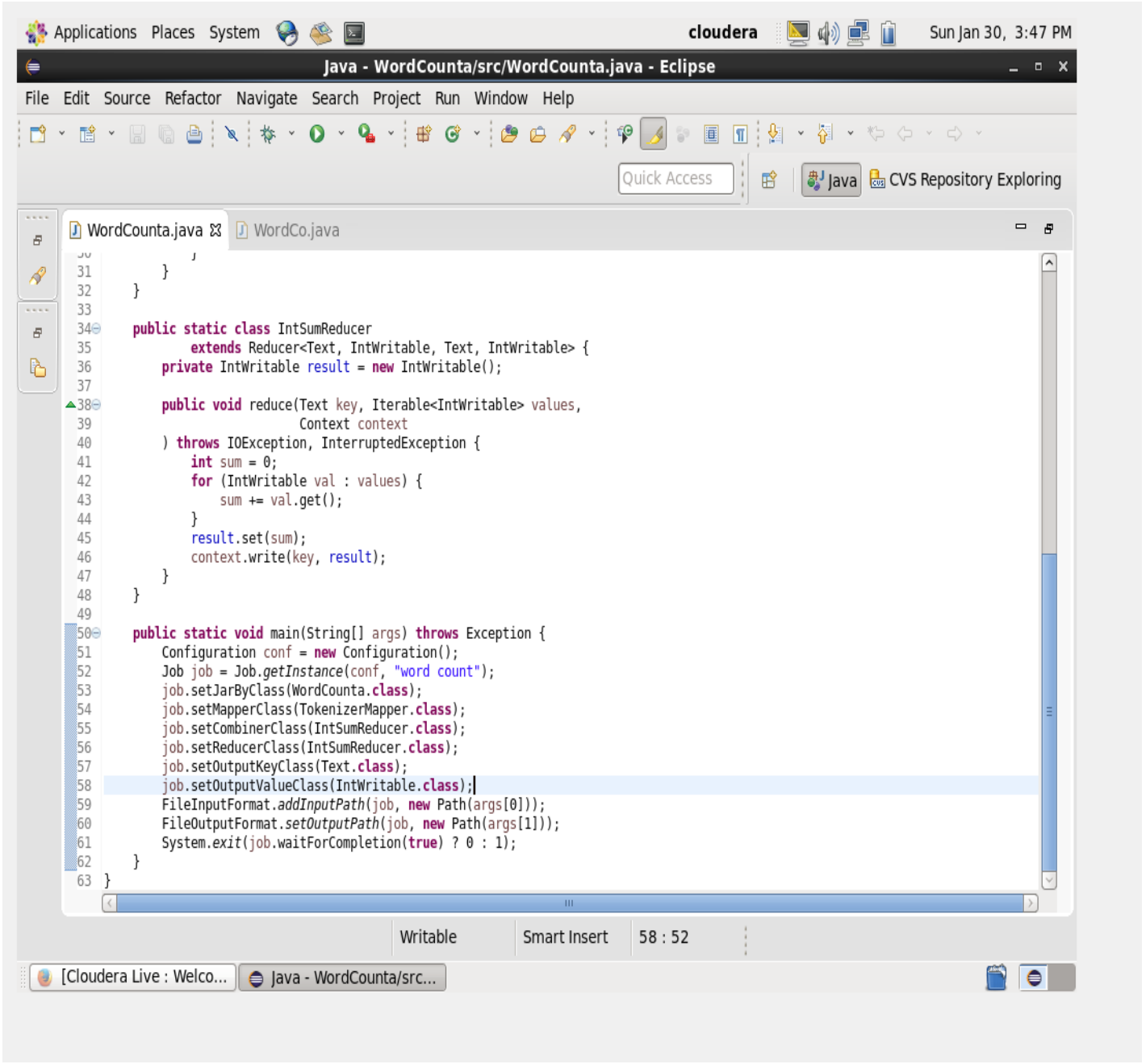
Step:4

Logic: if(word.toString().startsWith(“a”))

Logic explaination: The code is almost same as the wordcount program,but here only the frequency of words that starts with “a” are counted.

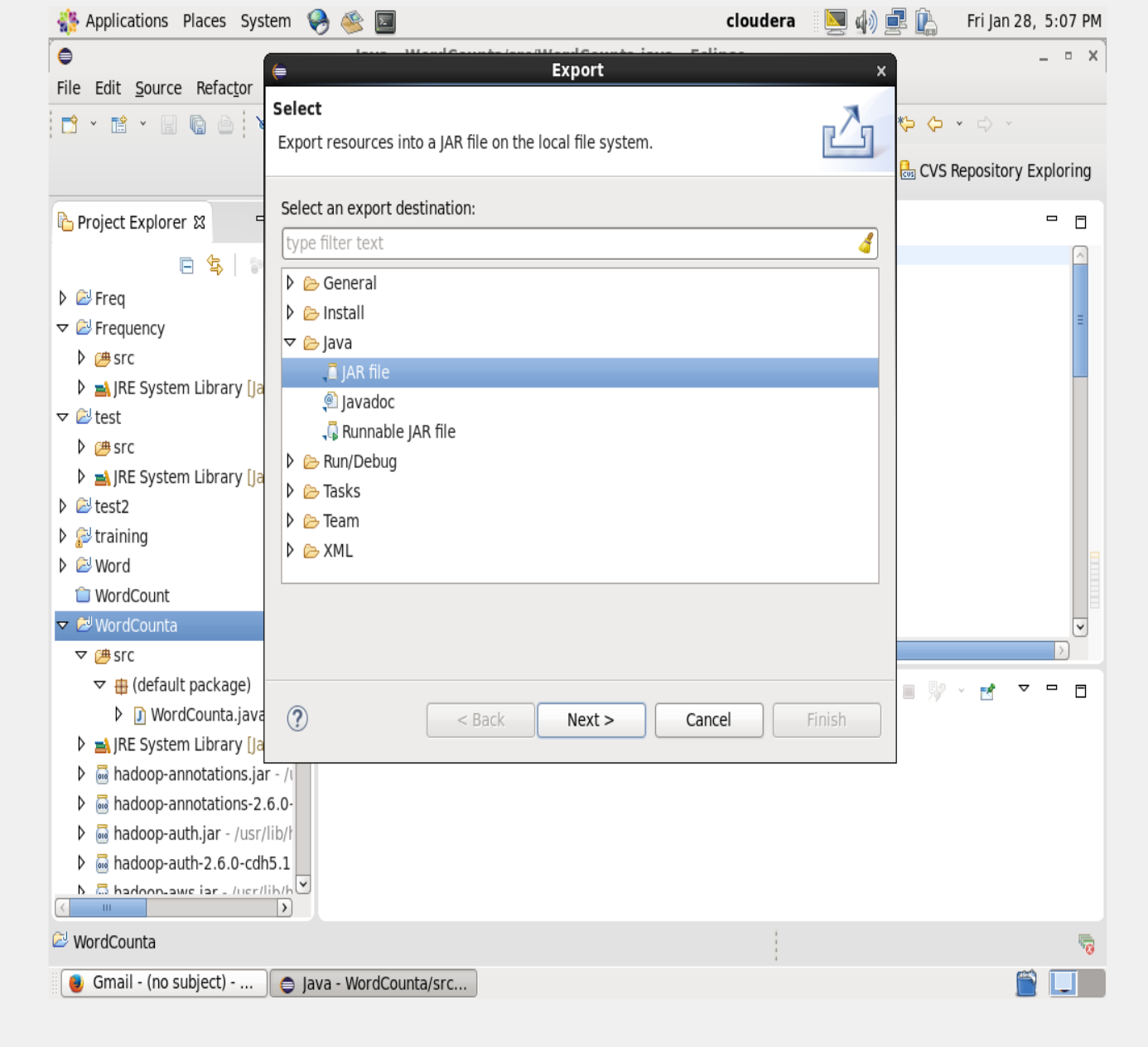
Program code:



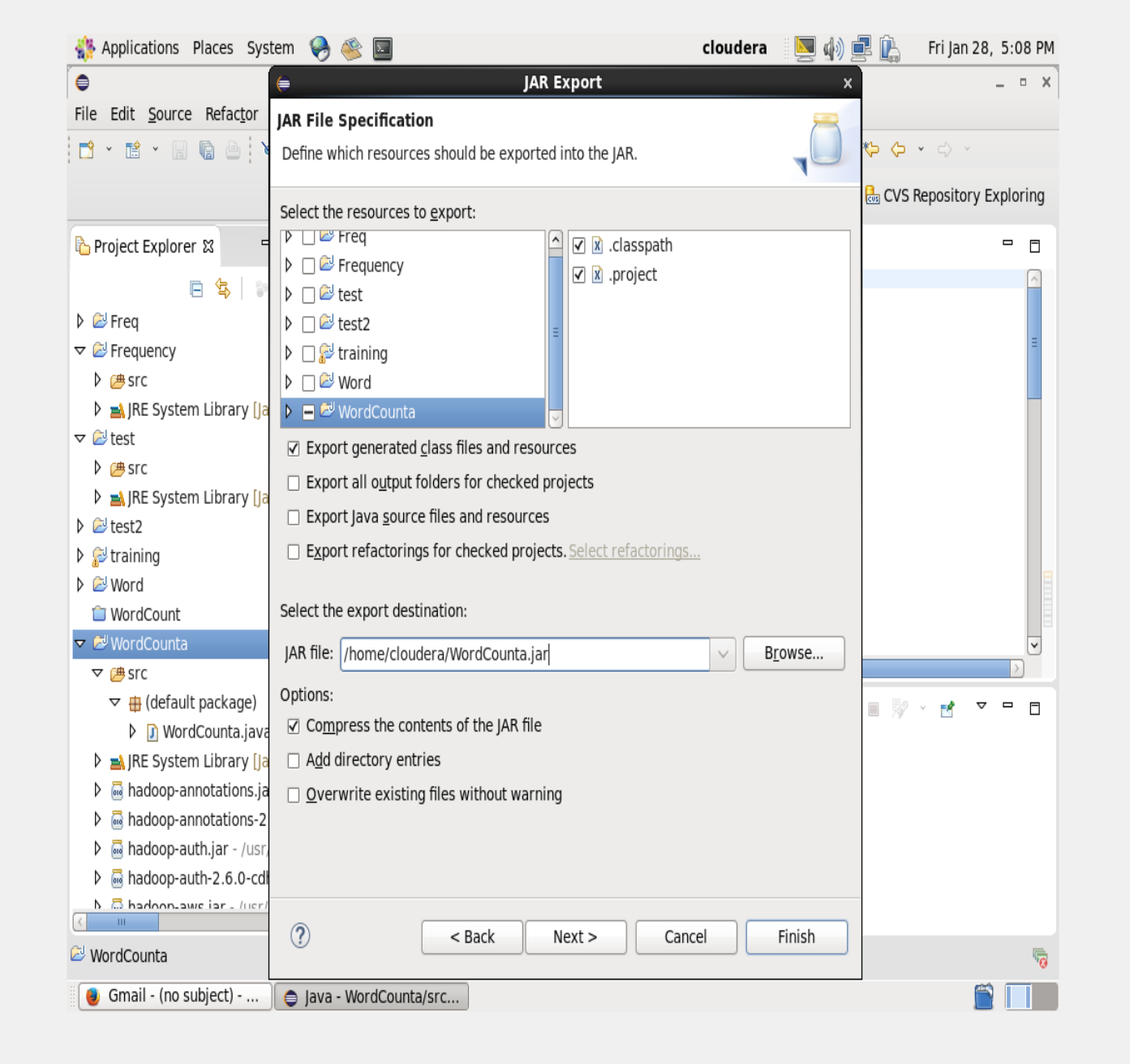

Applications Places System
cloudera
Sun Jan 30, 3:47 PM
Java - WordCounta/src/WordCounta.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
CVS Repository Exploring
WordCounta.java
WordCo.java
31 }
32 }
33
34 public static class IntSumReducer
35 extends Reducer<Text, IntWritable, Text, IntWritable> {
36 private IntWritable result = new IntWritable();
37
38 public void reduce(Text key, Iterable<IntWritable> values,
39 Context context
40) throws IOException, InterruptedException {
41 int sum = 0;
42 for (IntWritable val : values) {
43 sum += val.get();
44 }
45 result.set(sum);
46 context.write(key, result);
47 }
48 }
49
50 public static void main(String[] args) throws Exception {
51 Configuration conf = new Configuration();
52 Job job = Job.getInstance(conf, "word count");
53 job.setJarByClass(WordCounta.class);
54 job.setMapperClass(TokenizerMapper.class);
55 job.setCombinerClass(IntSumReducer.class);
56 job.setReducerClass(IntSumReducer.class);
57 job.setOutputKeyClass(Text.class);
58 job.setOutputValueClass(IntWritable.class);
59 FileInputFormat.addInputPath(job, new Path(args[0]));
60 FileOutputFormat.setOutputPath(job, new Path(args[1]));
61 System.exit(job.waitForCompletion(true) ? 0 : 1);
62 }
63 }
Writable
Smart Insert
58 : 52
[Cloudera Live : Welco...
Java - WordCounta/src...

Step5:

Exporting the JAR,right click on WordCounta and select export

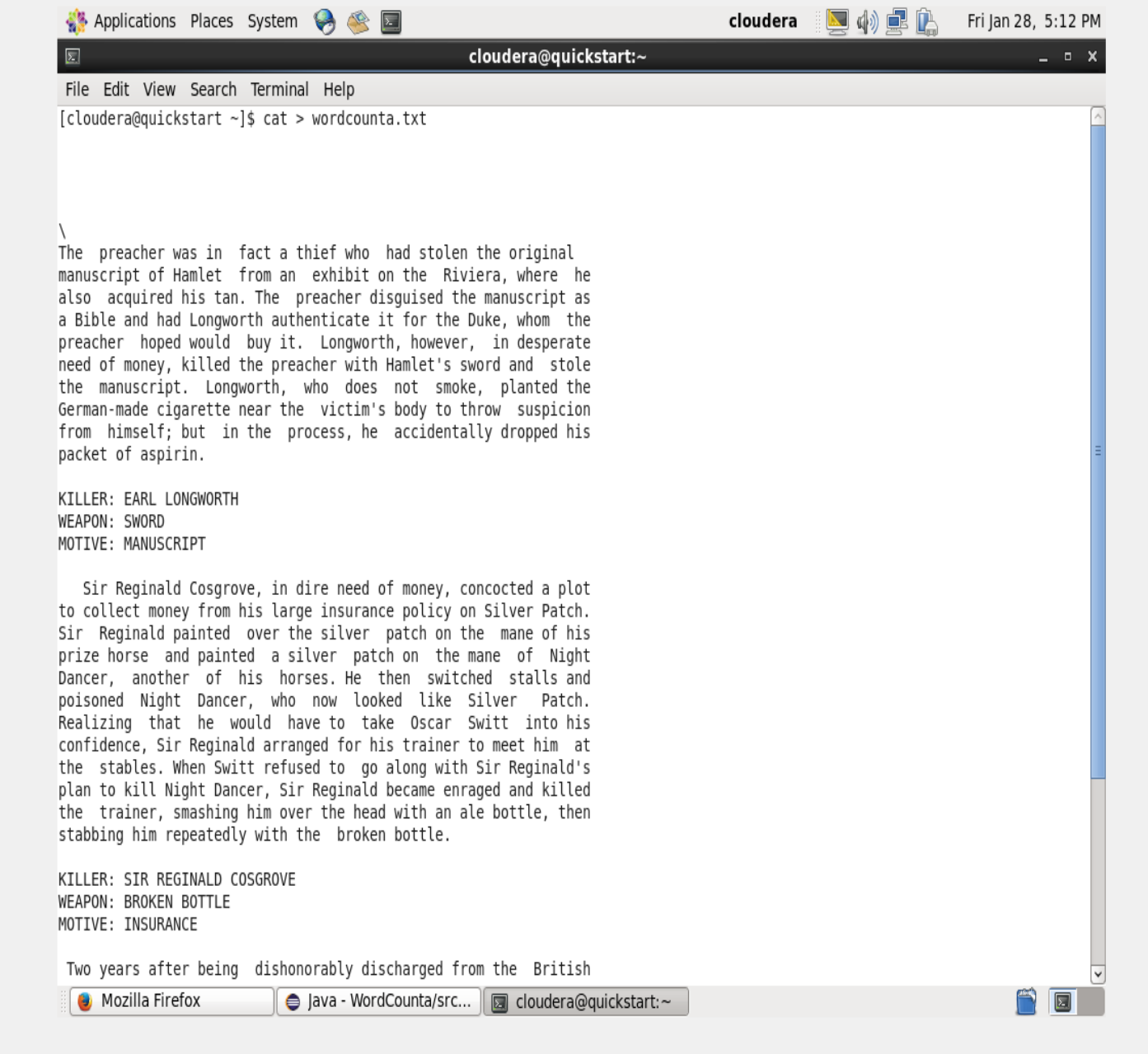


Select the export destination



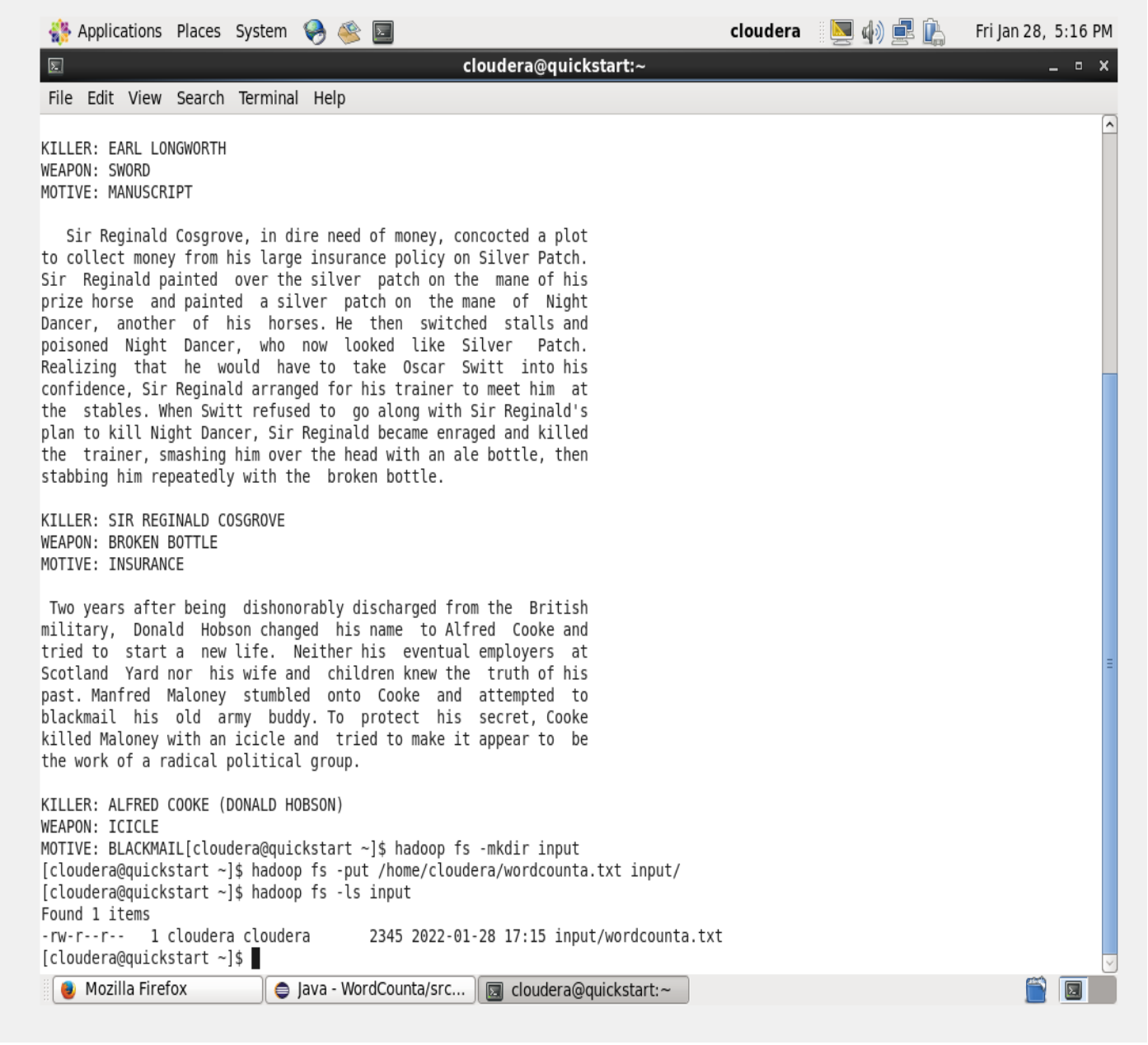
Step:6

Creating a new file called wordcounta.txt and giving the input



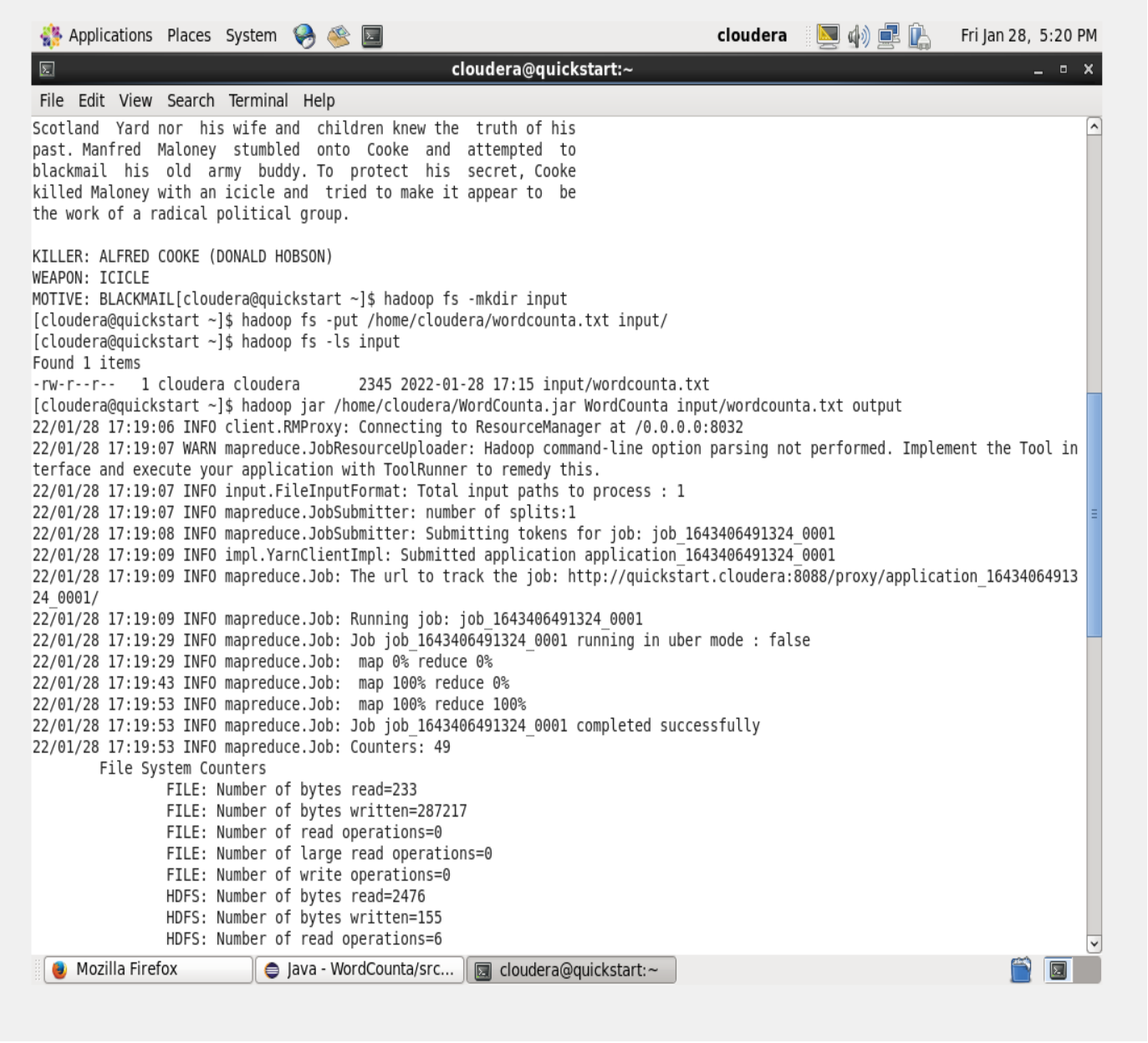
Step:7

Creating a directory called input and loading the wordcounta.txt file into hadoop hdfs

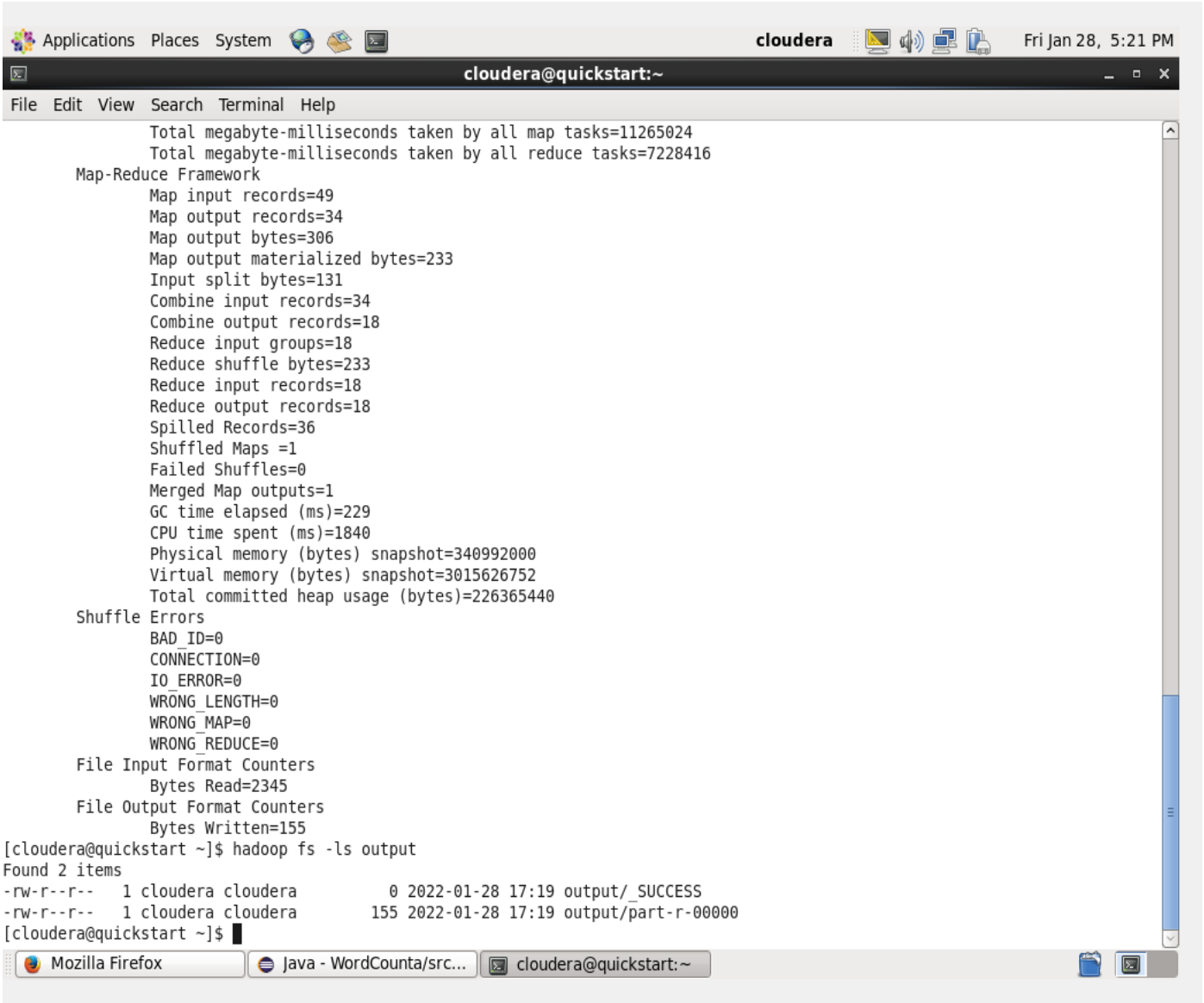


Step:8

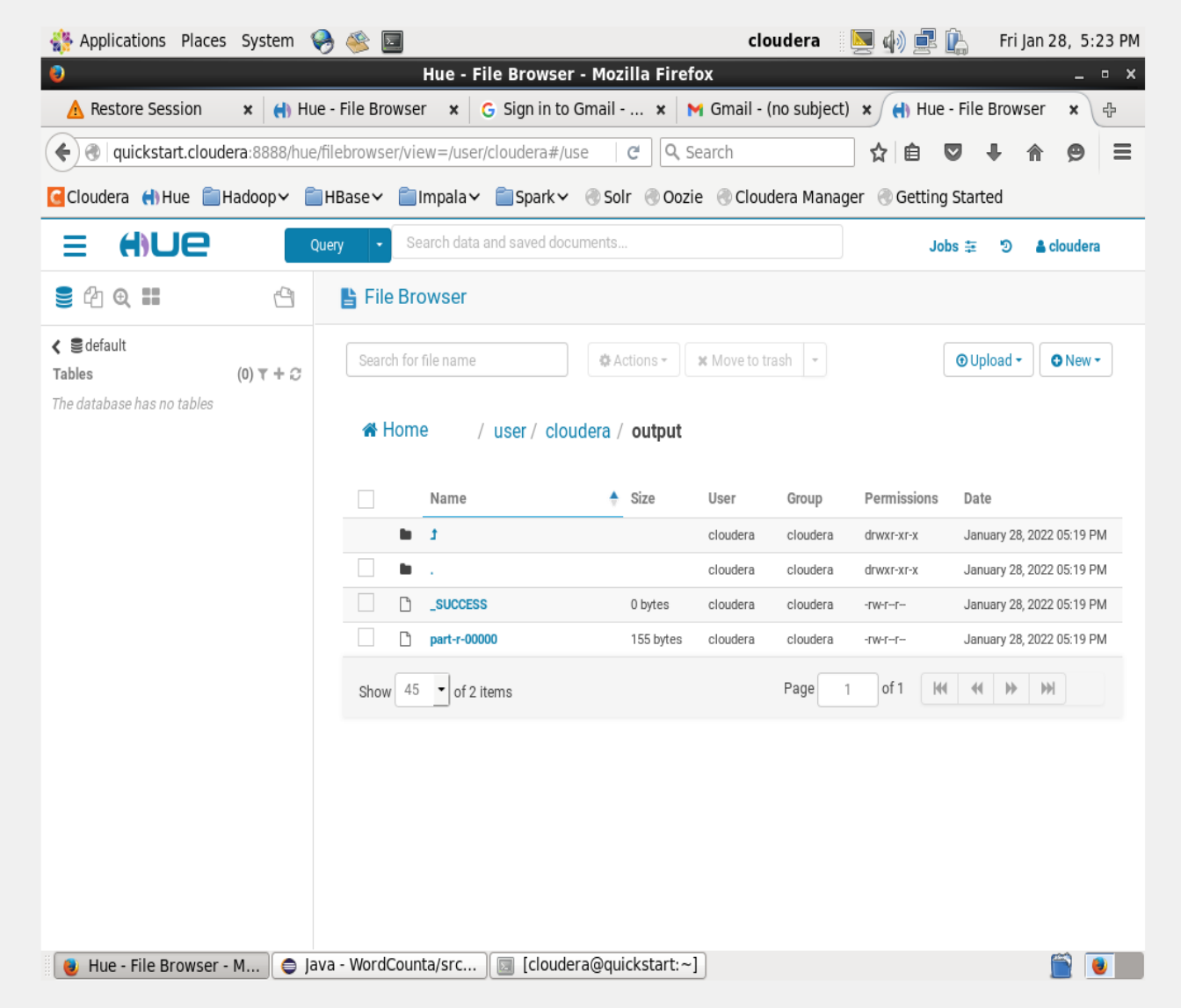
Performing the mapreducing



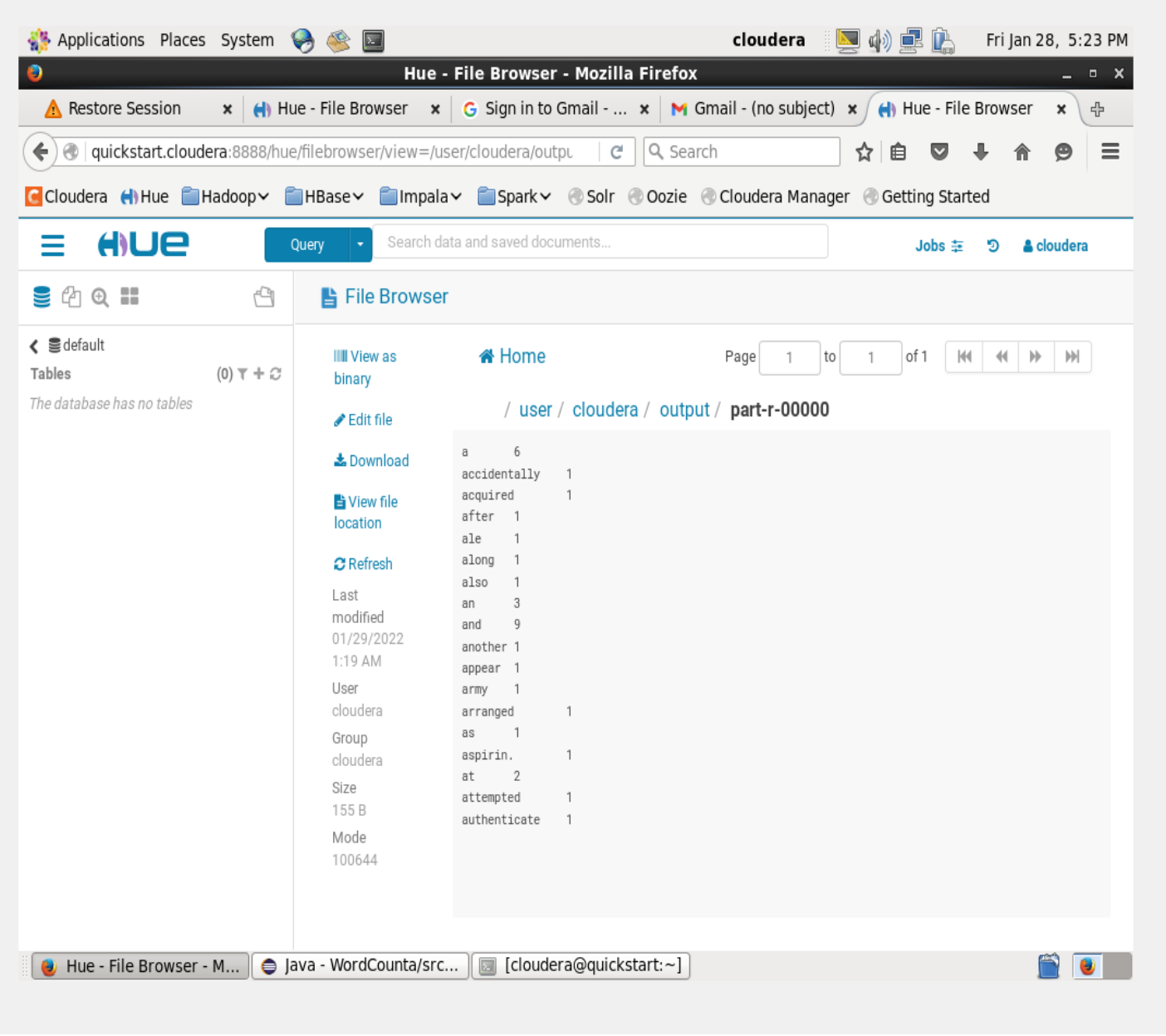
Map reducing is success and two items are found



Visualization of the found items



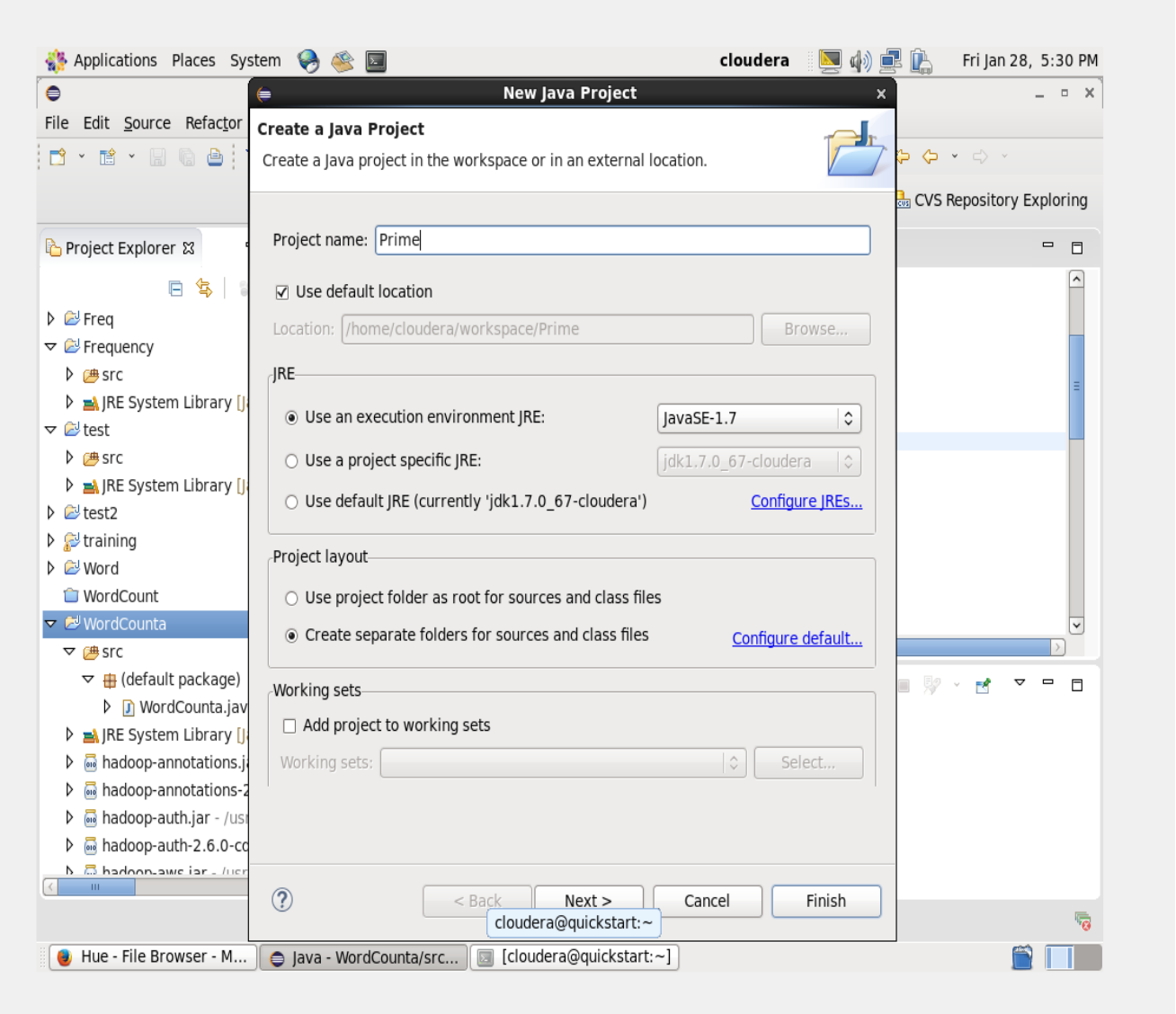
Frequency count of the words that starts with ‘a’ in the given dataset is found and output is below



3.Determining the prime number in input and print number only once

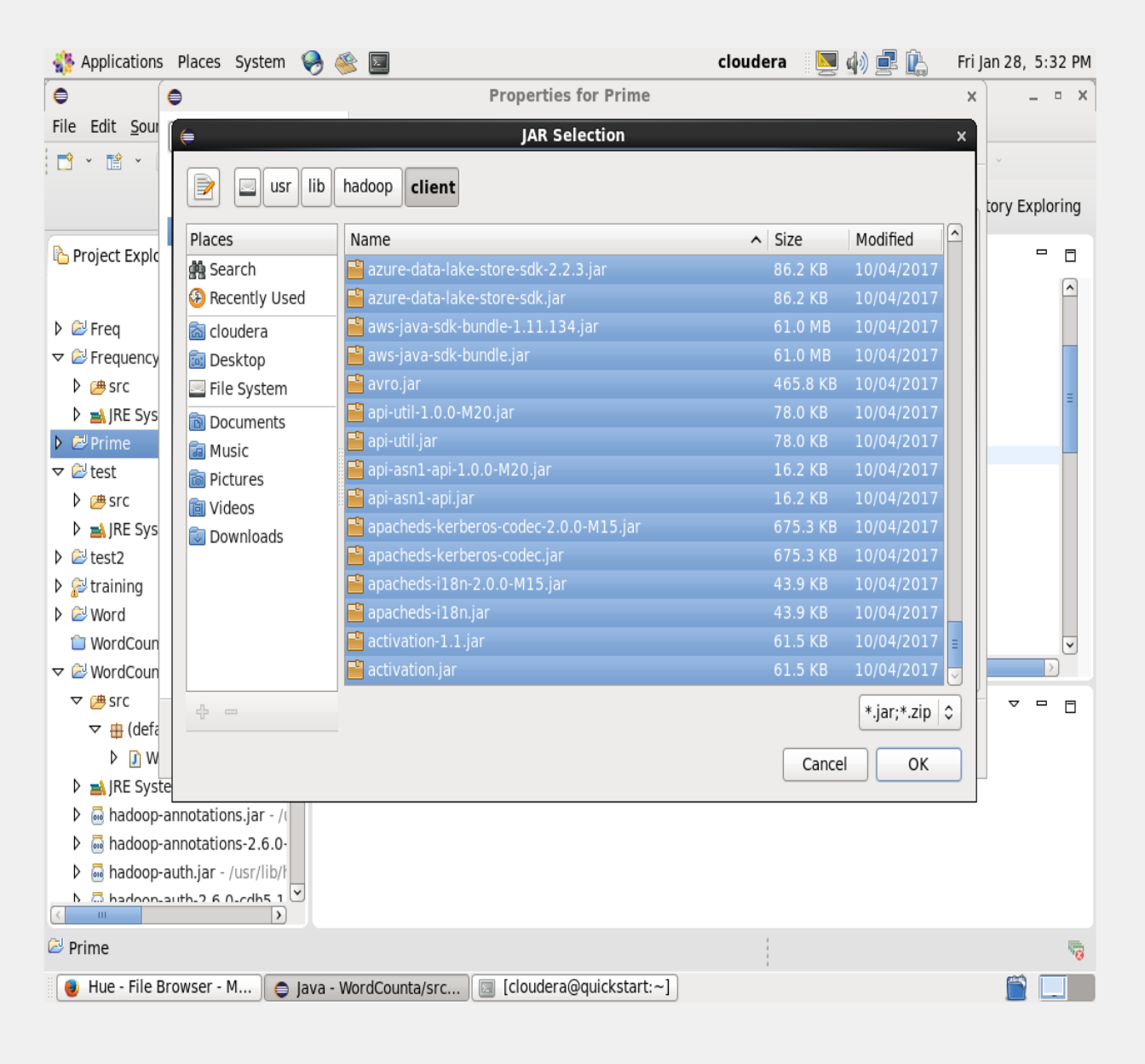
Step:1

Creating a new java project and naming it as Prime



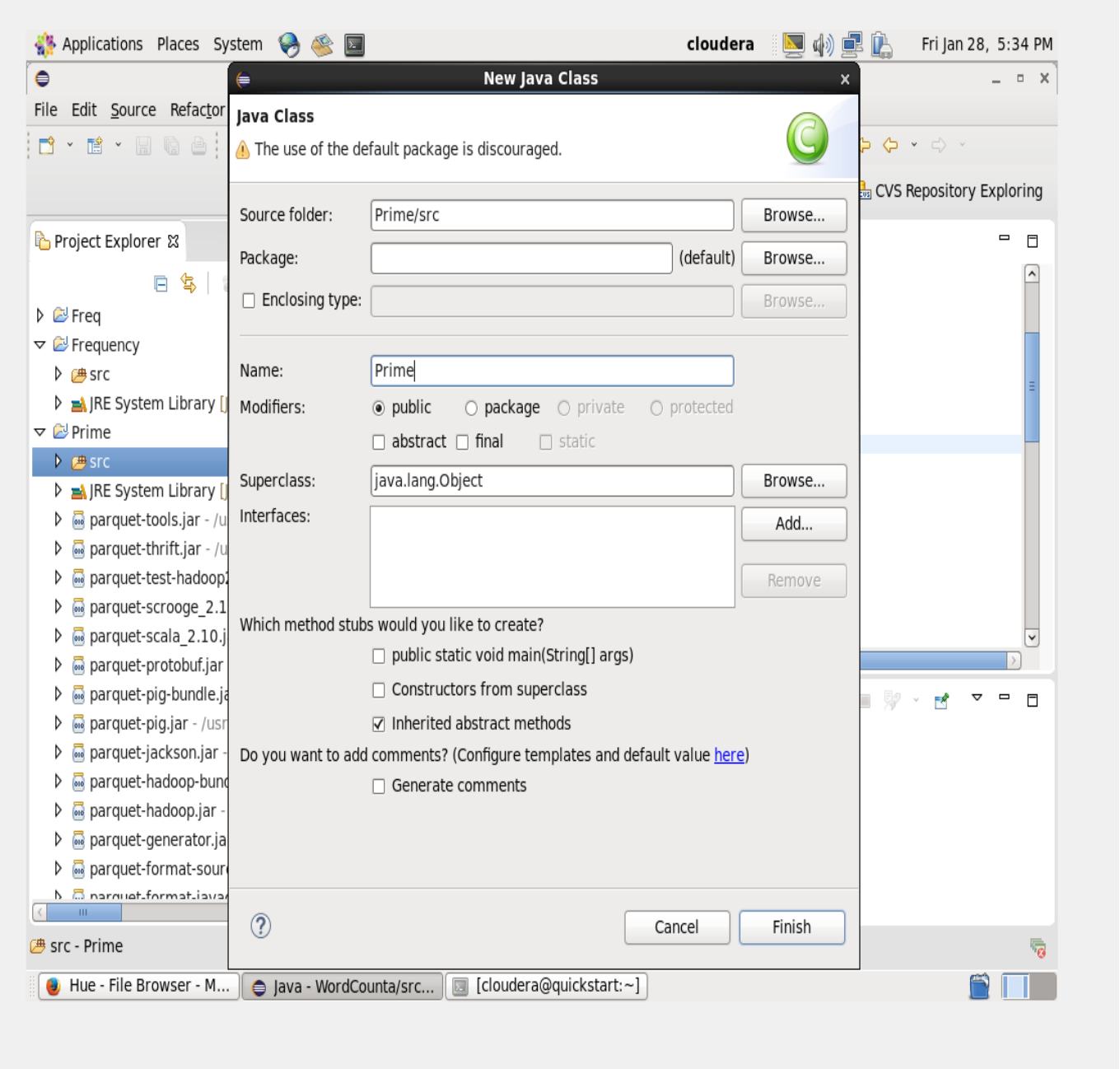
Step:2

Adding the JARs files of hadoop and client



Step:3

Creating the class file

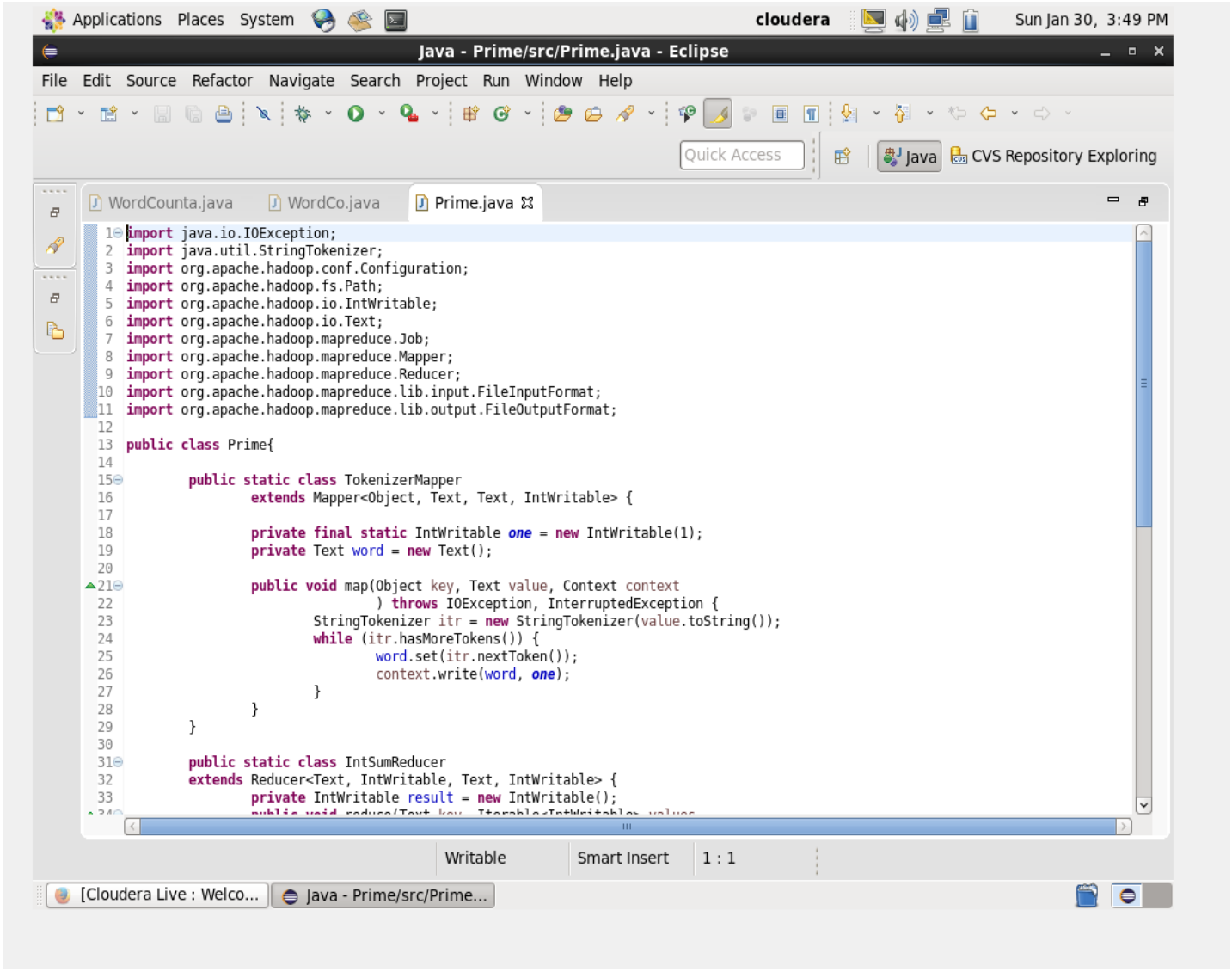


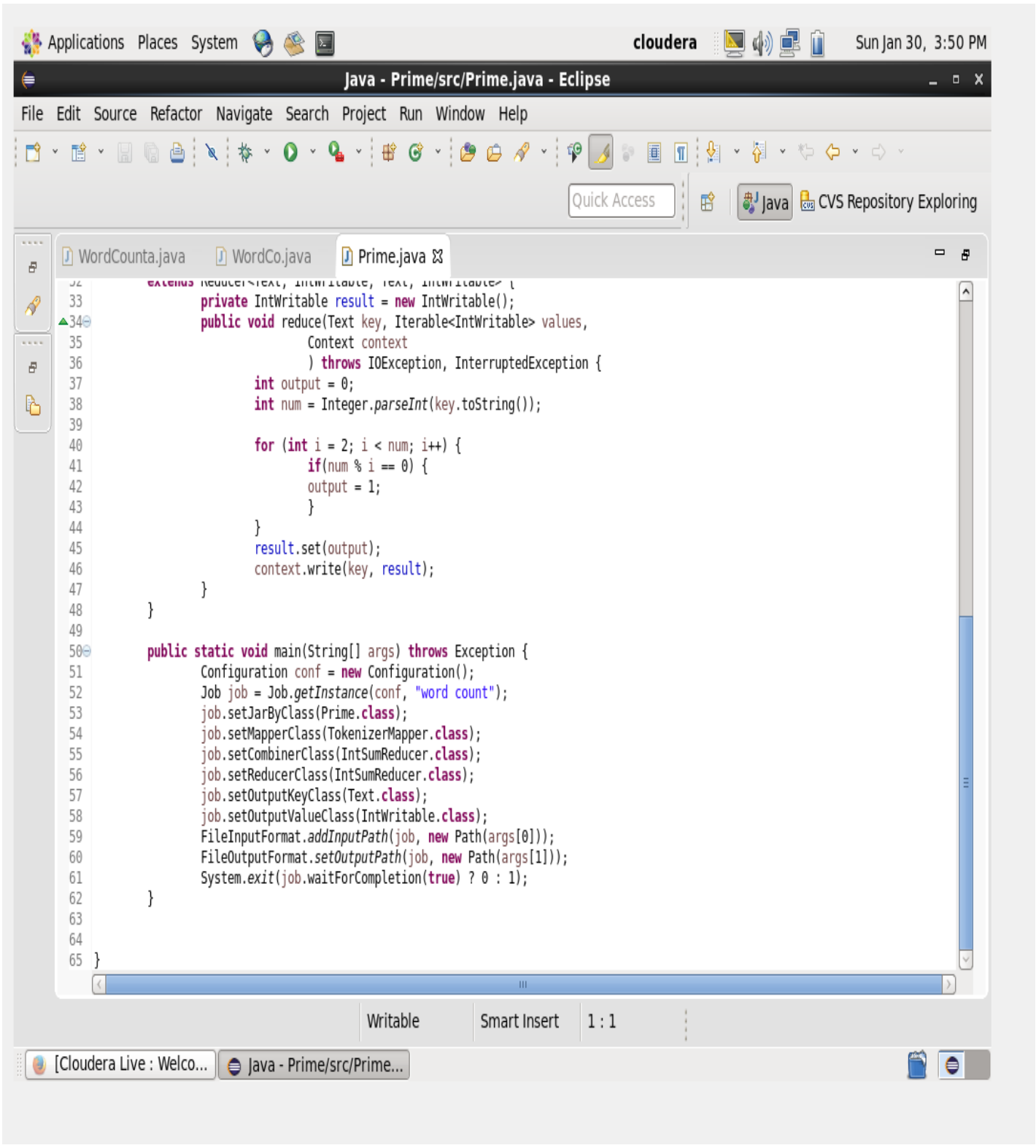
Step:4

Logic:

```
int output =0;
int num = Integer.parseInt(Key.toString());
for(int i=2;i<num;i++) {
        if(num%i==0) {
        Output=1; }
}
```

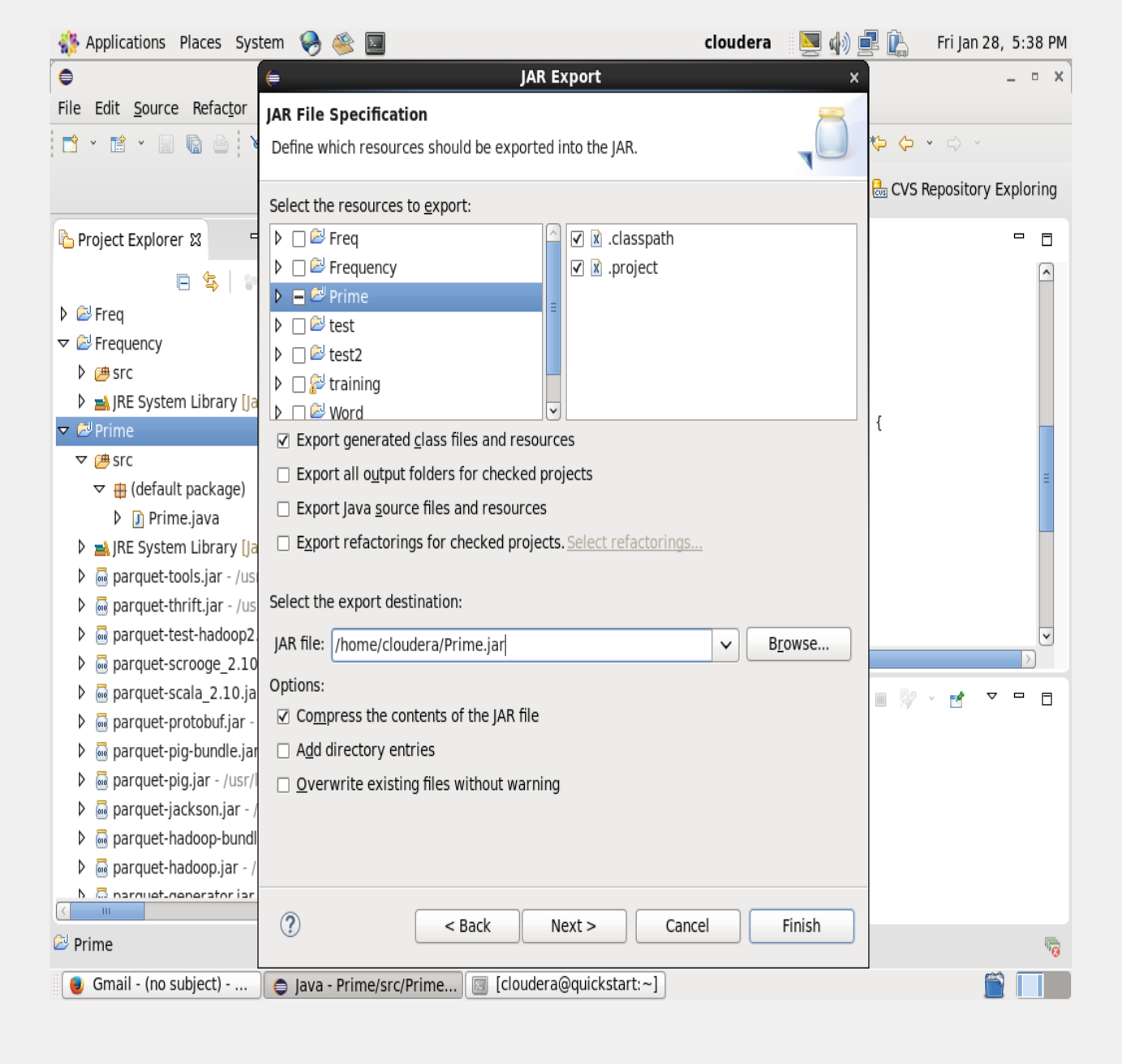
Program code:




Applications Places System
cloudera
Sun Jan 30, 3:50 PM
Java - Prime/src/Prime.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
CVS Repository Exploring
WordCounta.java
WordCo.java
Prime.java
33 private IntWritable result = new IntWritable();
34 public void reduce(Text key, Iterable<IntWritable> values,
35 Context context
36) throws IOException, InterruptedException {
37 int output = 0;
38 int num = Integer.parseInt(key.toString());
39
40 for (int i = 2; i < num; i++) {
41 if(num % i == 0) {
42 output = 1;
43 }
44 }
45 result.set(output);
46 context.write(key, result);
47 }
48 }
49
50 public static void main(String[] args) throws Exception {
51 Configuration conf = new Configuration();
52 Job job = Job.getInstance(conf, "word count");
53 job.setJarByClass(Prime.class);
54 job.setMapperClass(TokenizerMapper.class);
55 job.setCombinerClass(IntSumReducer.class);
56 job.setReducerClass(IntSumReducer.class);
57 job.setOutputKeyClass(Text.class);
58 job.setOutputValueClass(IntWritable.class);
59 FileInputFormat.addInputPath(job, new Path(args[0]));
60 FileOutputFormat.setOutputPath(job, new Path(args[1]));
61 System.exit(job.waitForCompletion(true) ? 0 : 1);
62 }
63
64
65 }
Writable
Smart Insert
1 : 1
[Cloudera Live : Welco...
Java - Prime/src/Prime...

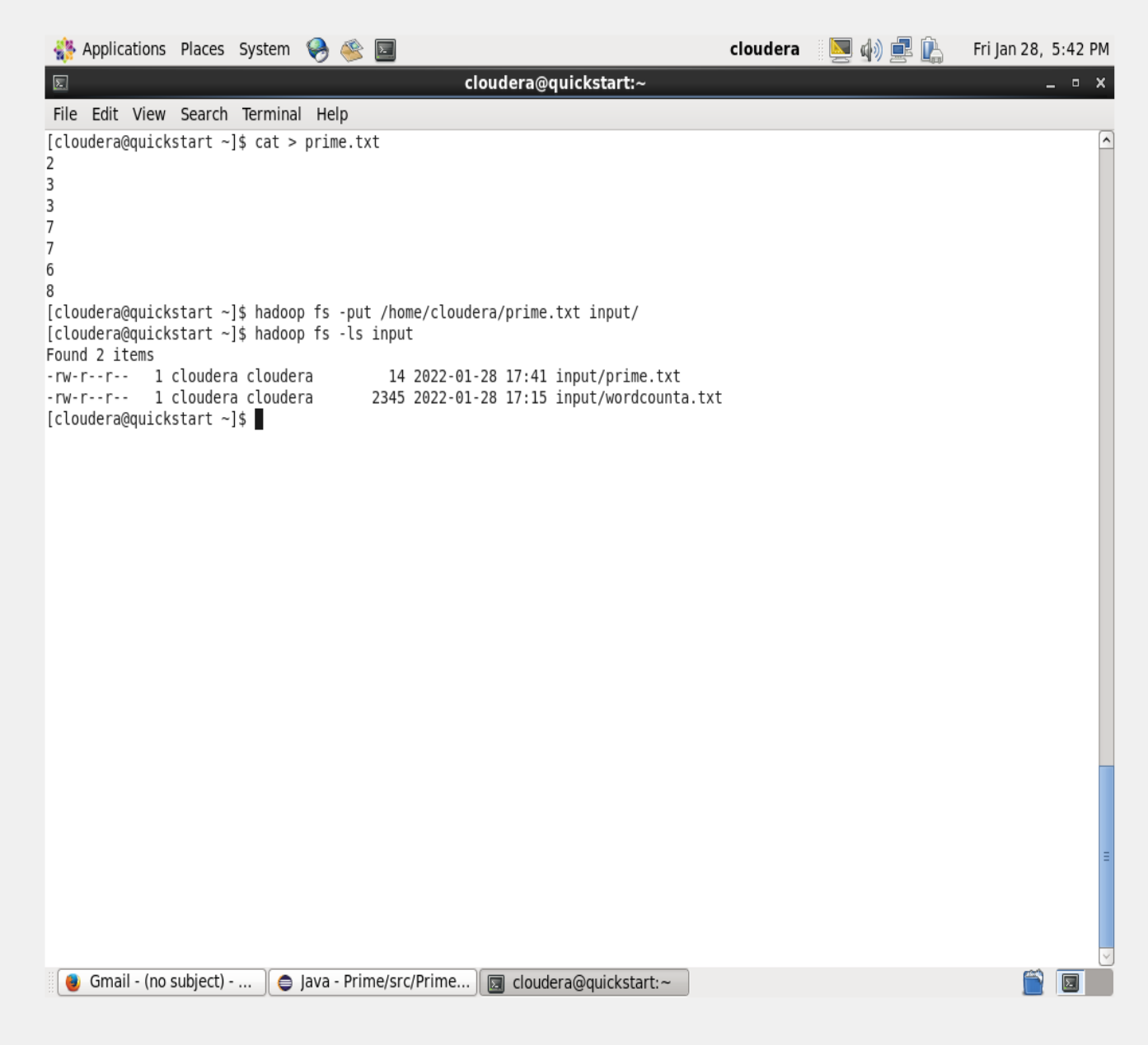
Step:5

Exporting the JARs file



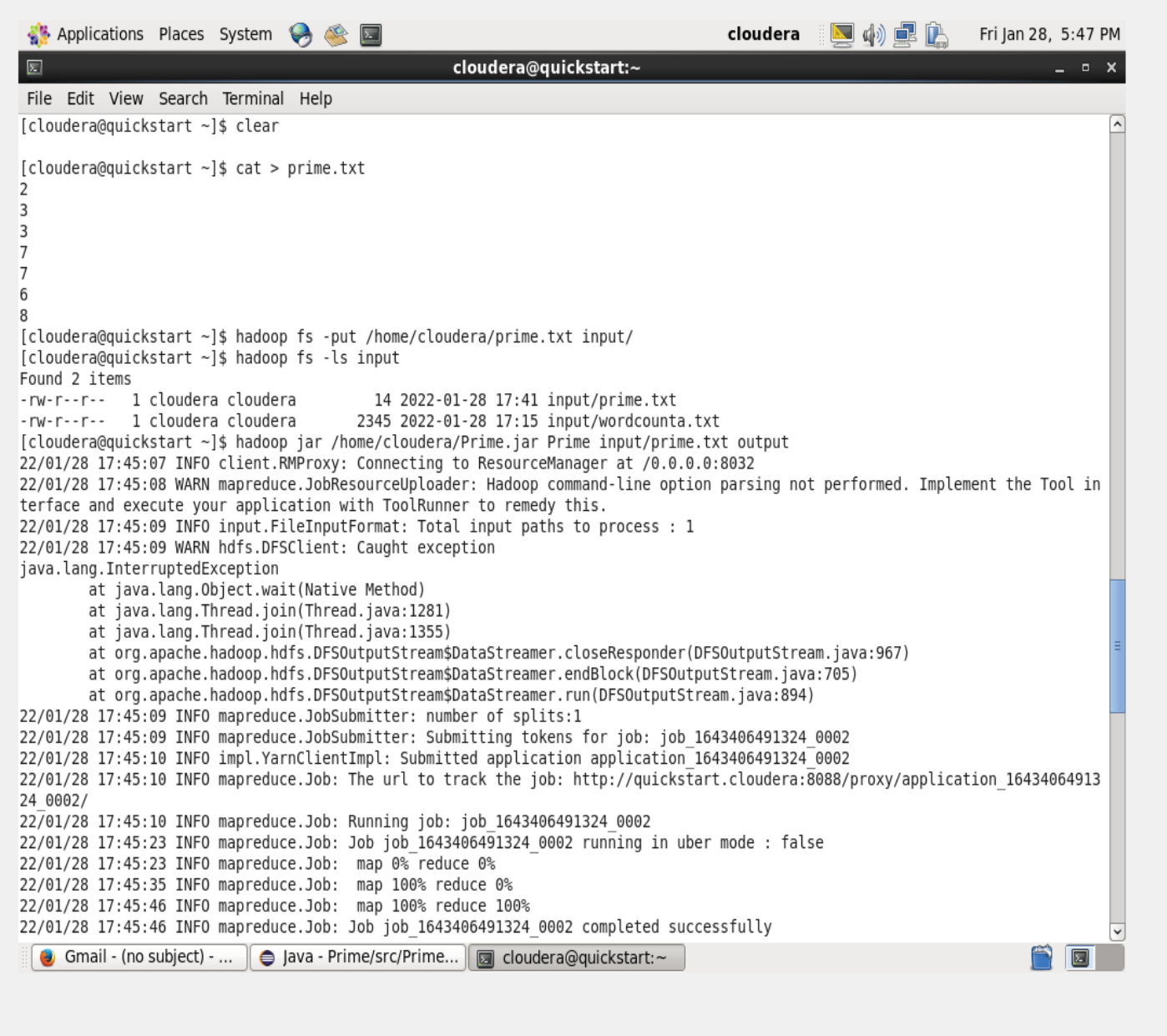
Step:6

Creating a newfile called prime.txt and giving it the input data,then loading it into the hadoop hdfs

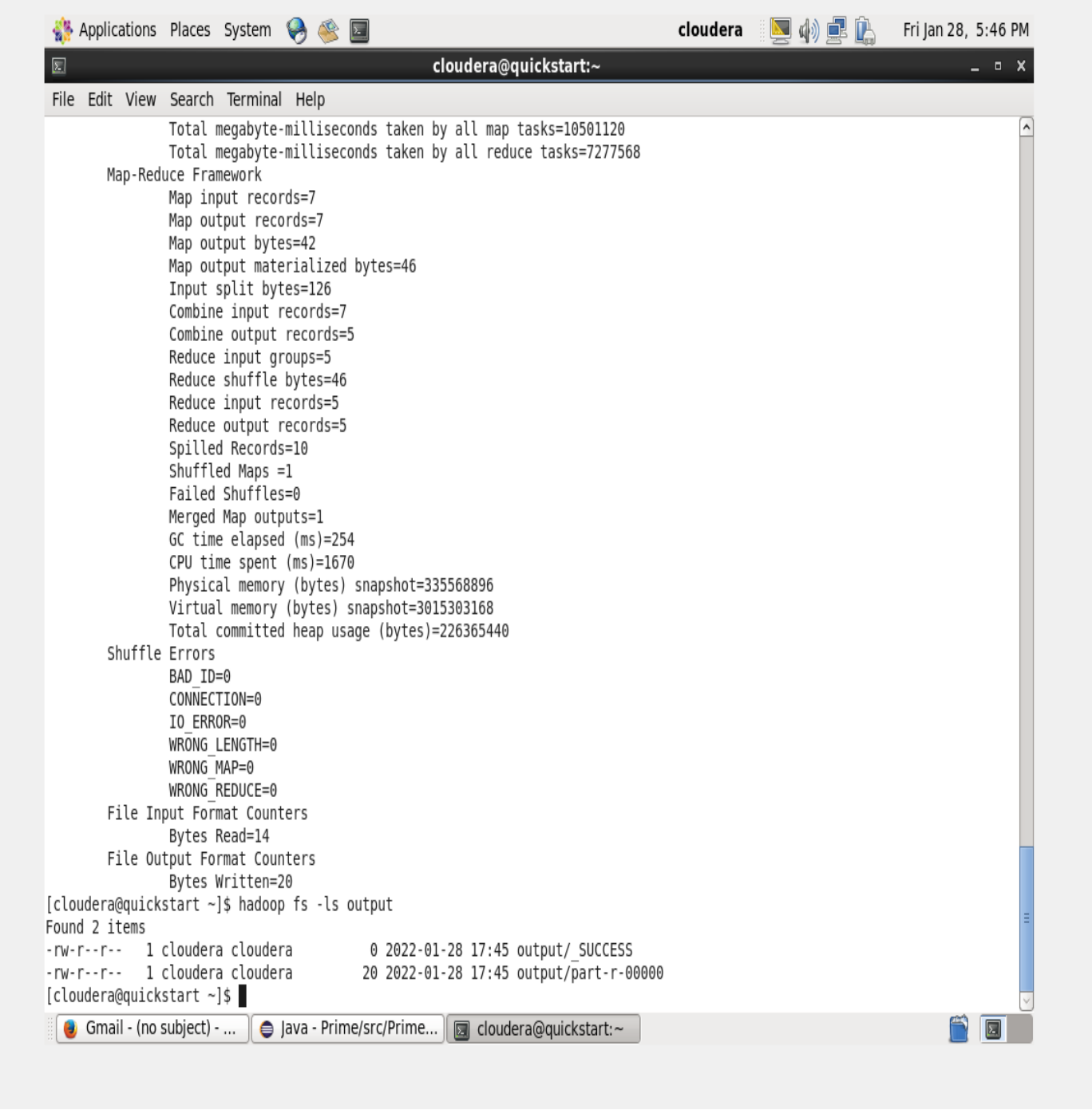


Step:7

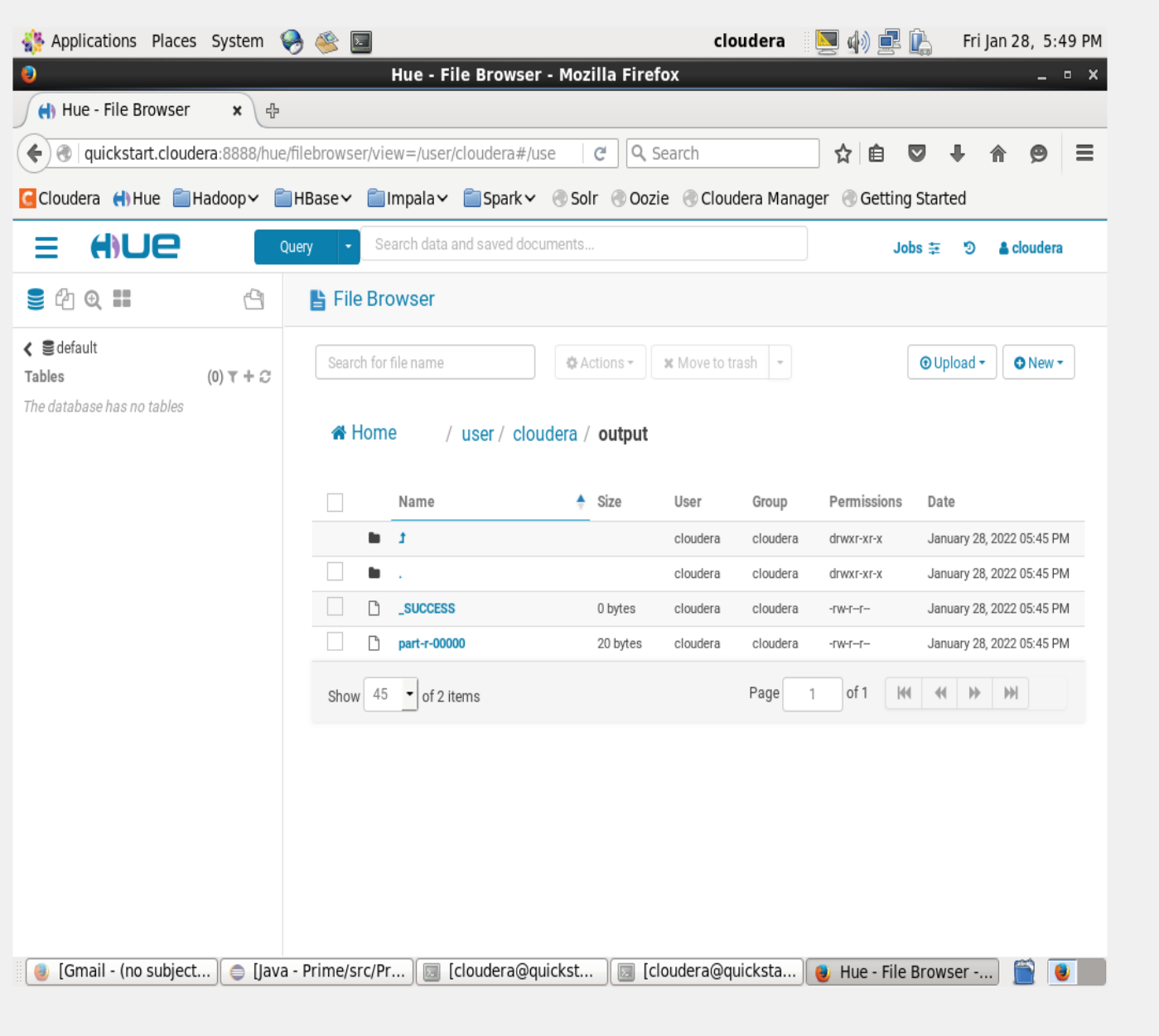
Map reducing



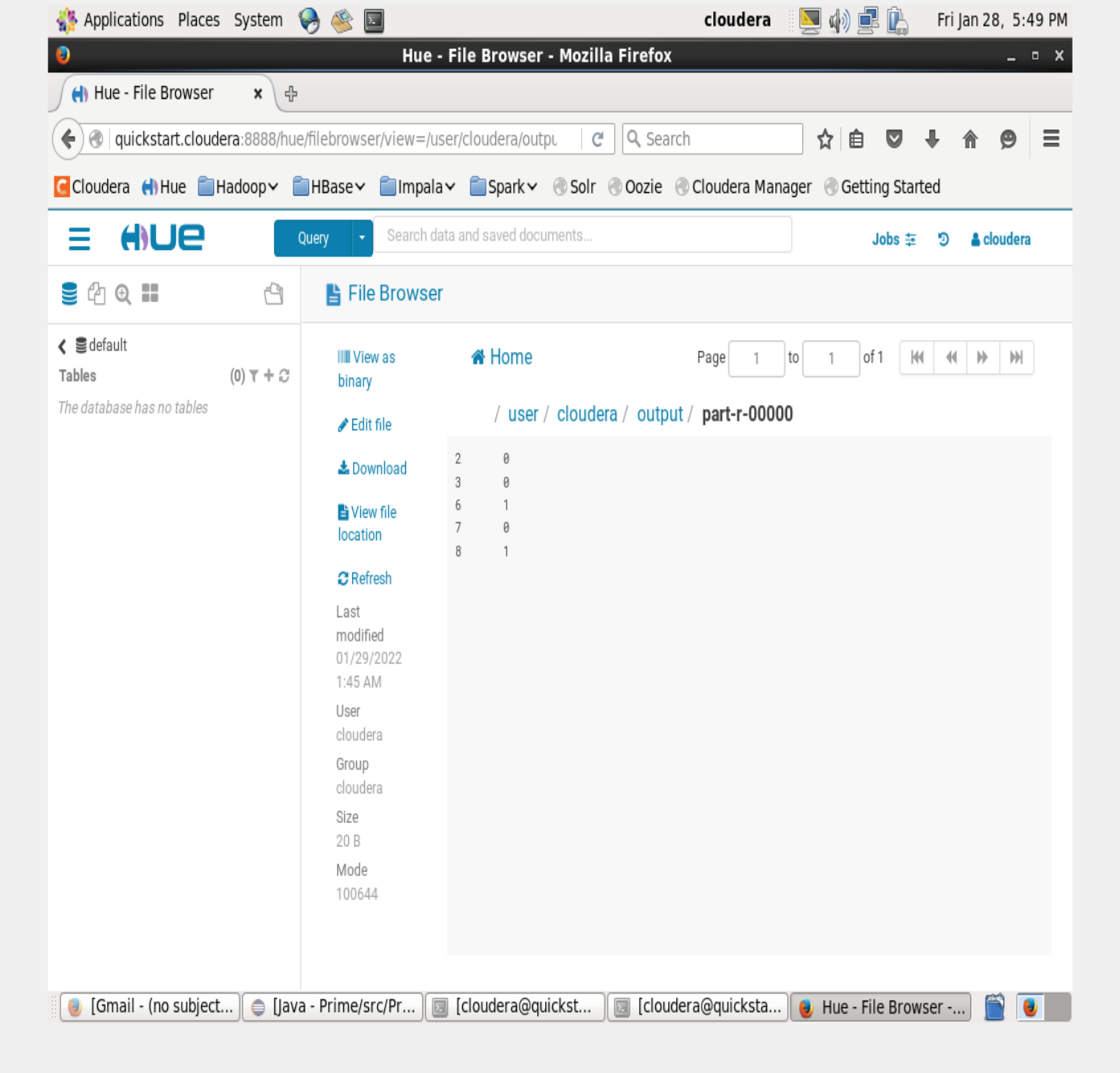
Map reducing is done and two items are found



Visualizing the items found



Prime numbers are determined in the given input and the output is shown in the below figure



Overall Explaination of the ICE-2

Using the Hadoop MapReducing technique we found the required output.MapReducing technique bascially has two functions 1.Mapper 2.Reducer.

Mapper takes the input data given and forms the key/value pairs for that data.

Reducer inturn,receives the key/value pairs as input from the mapper and perfoms the functions of shuffling,sorting and reducing.

Thus large datasets are reduced and required data/output is obtained by using hadoop mapreducing algorthim.