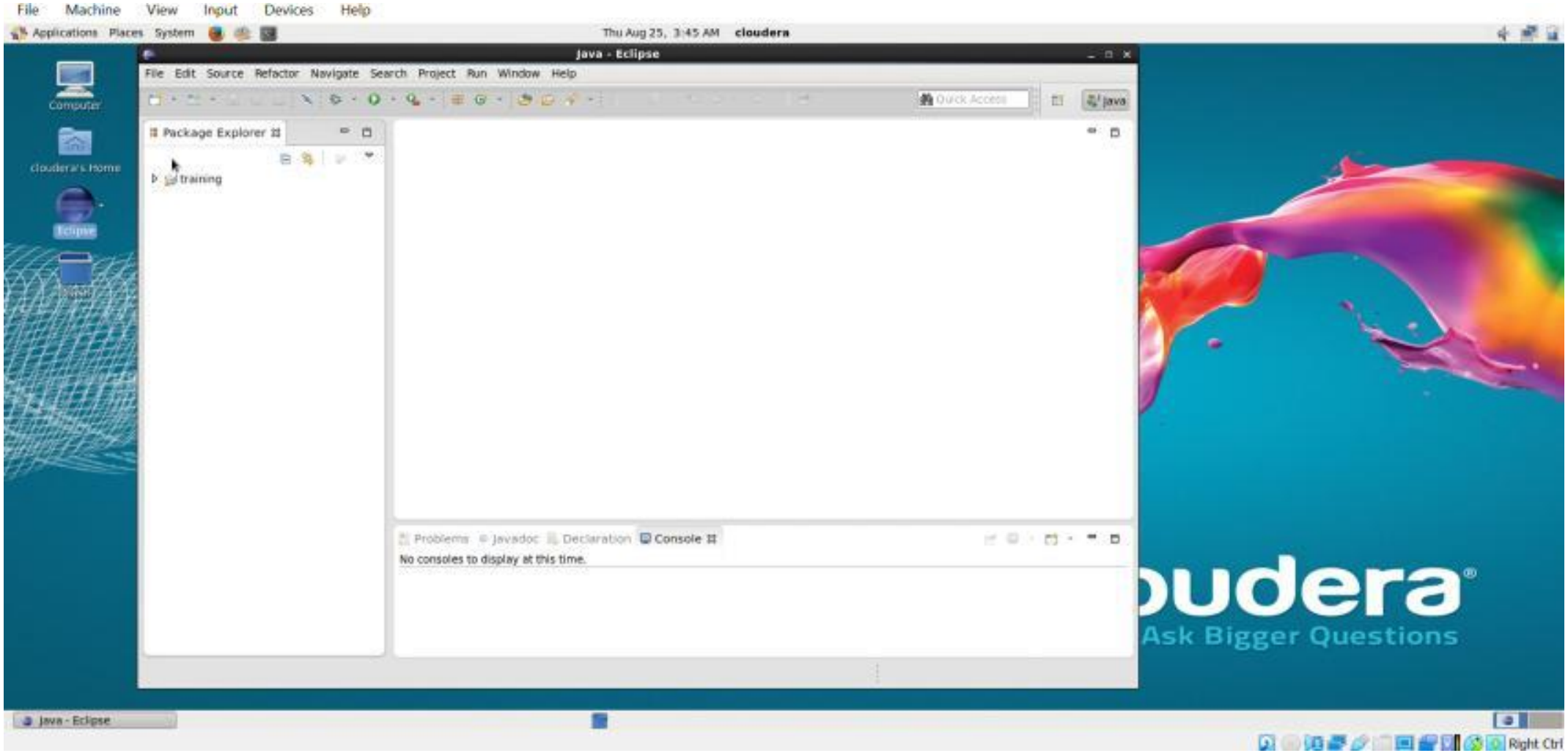
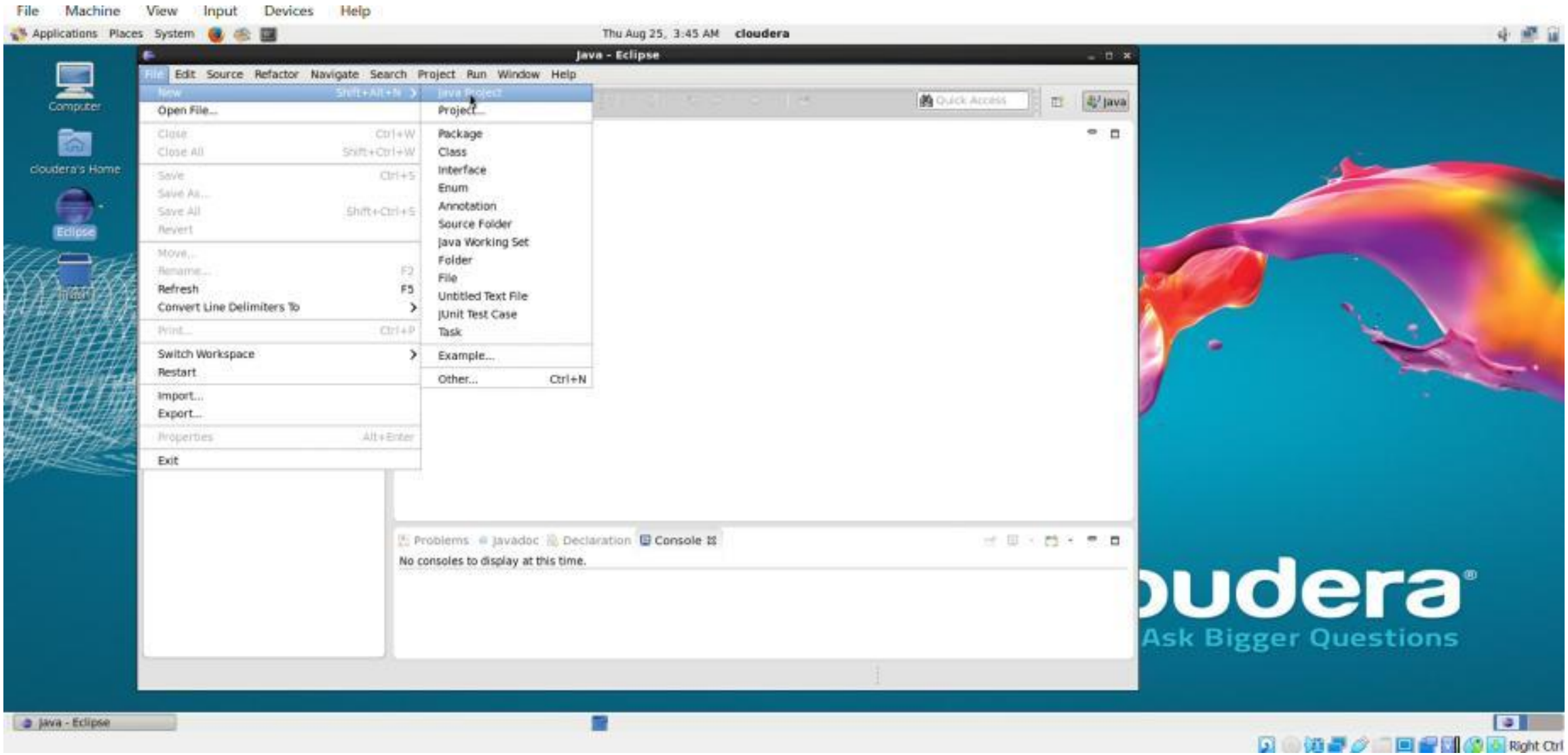
Aim:Run a basic Word Count MapReduce program to understand MapReduce Paradigm.

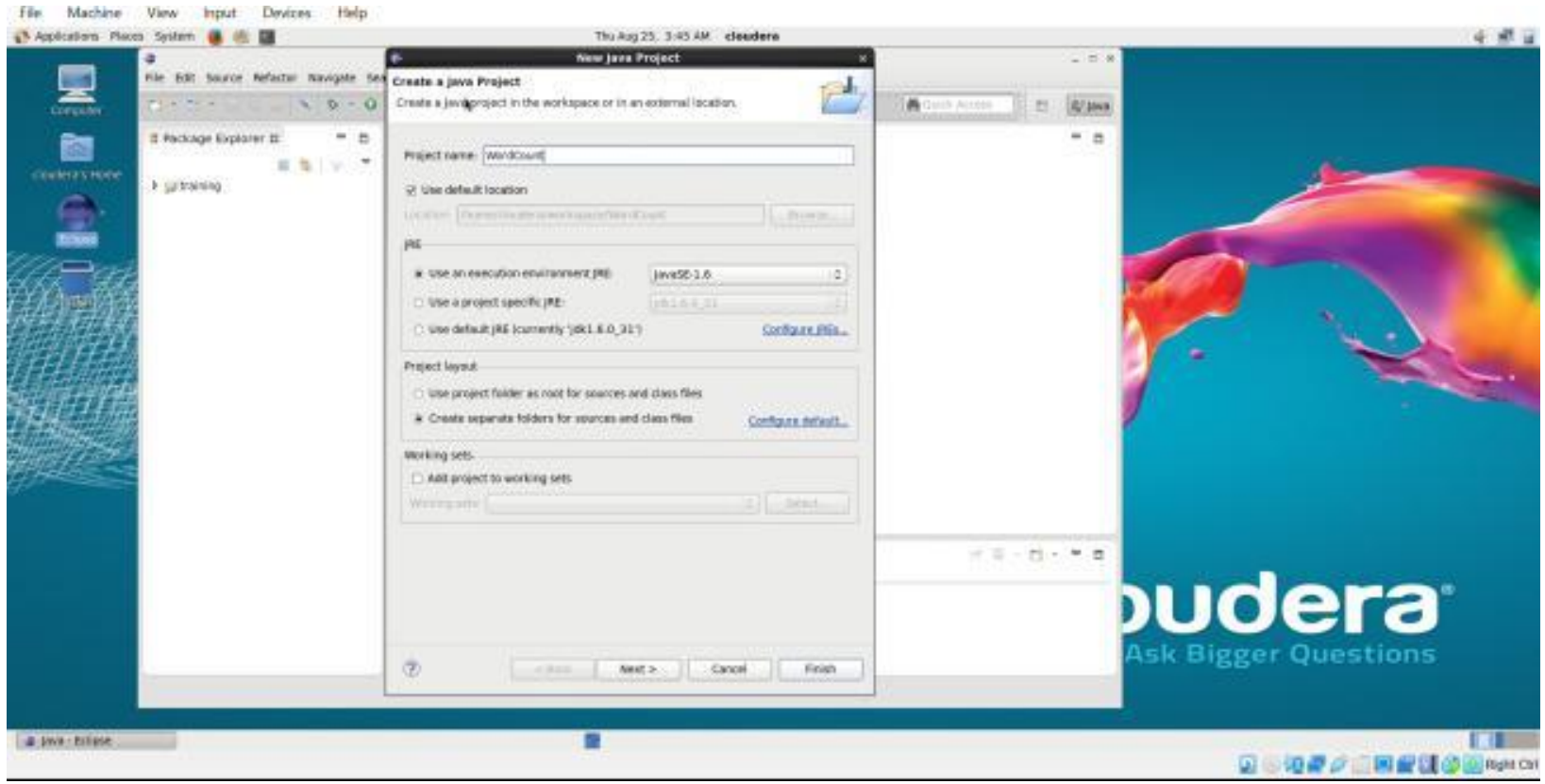
Open Oracle VM Virtual box -> click start -> open cloudera -> open eclipse.


File Machine View Input Devices Help
Applications Places System
Thu Aug 25, 3:45 AM cloudera
java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
java
Package Explorer
training
Computer
cloudera's Home
Eclipse
Problems Javadoc Declaration Console
No consoles to display at this time.
oudera®
Ask Bigger Questions
Java - Eclipse
Right Ctrl

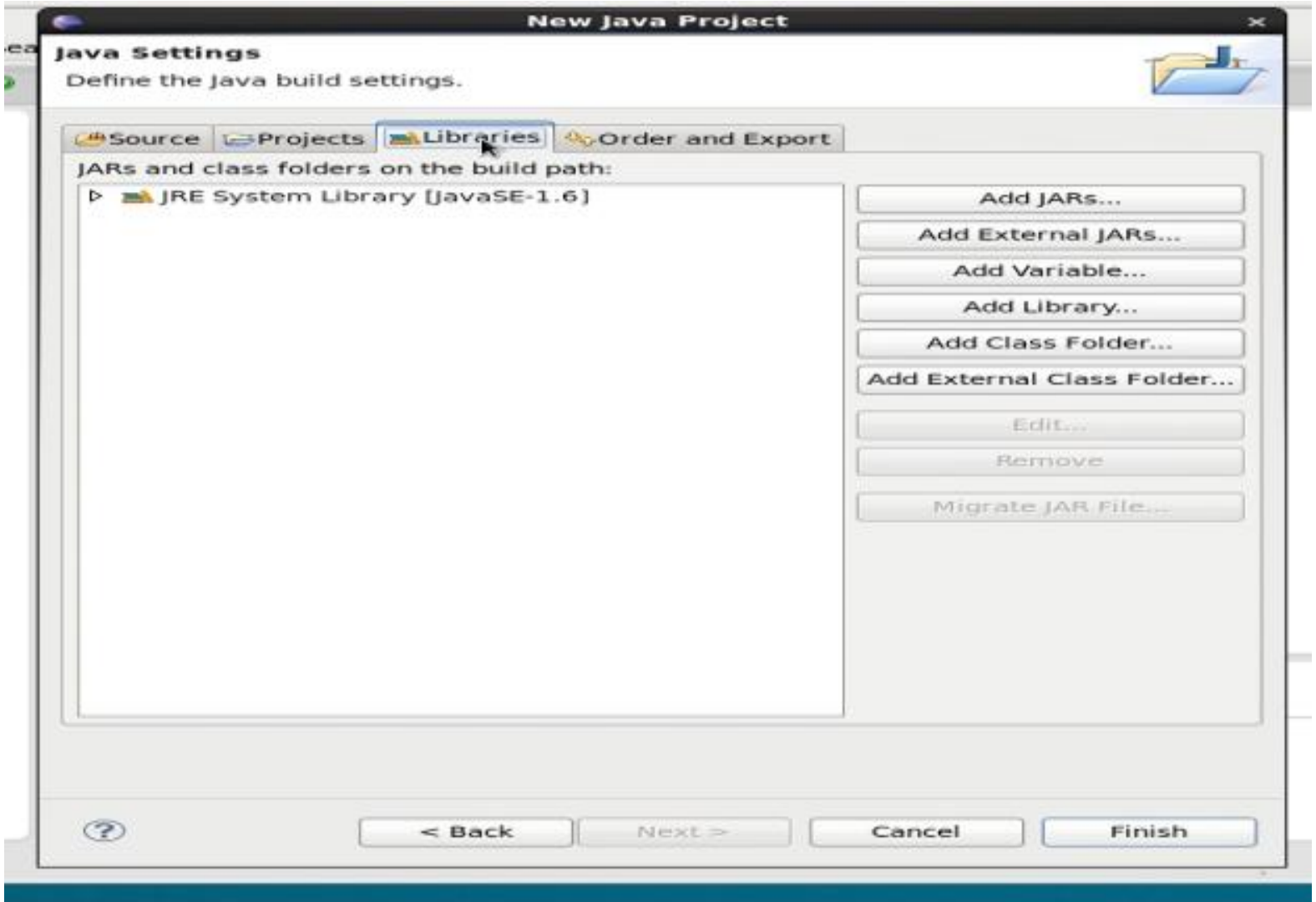
Click on file->New->Java Project



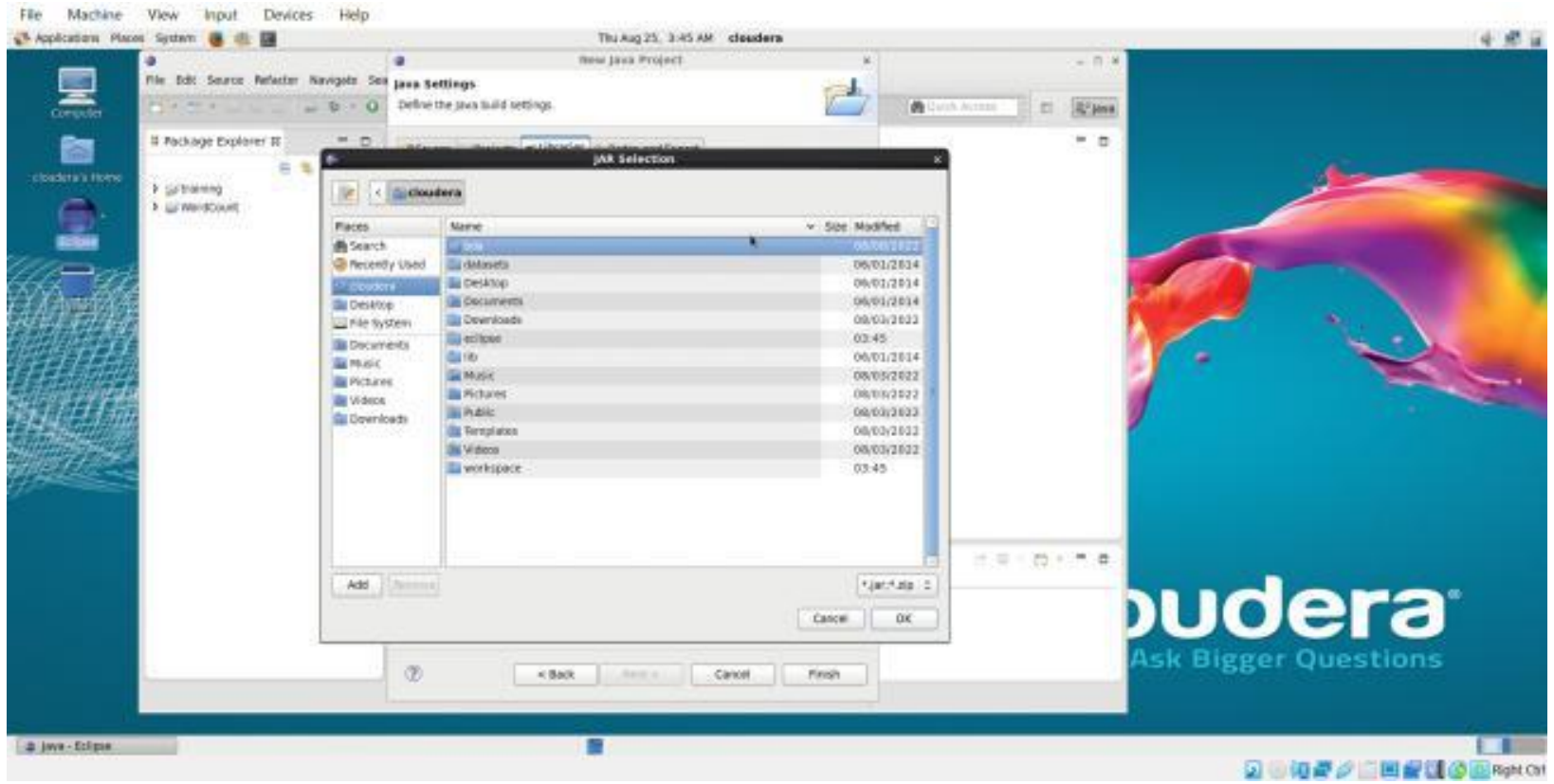
Give your project name as “**WordCount**” and click on next.



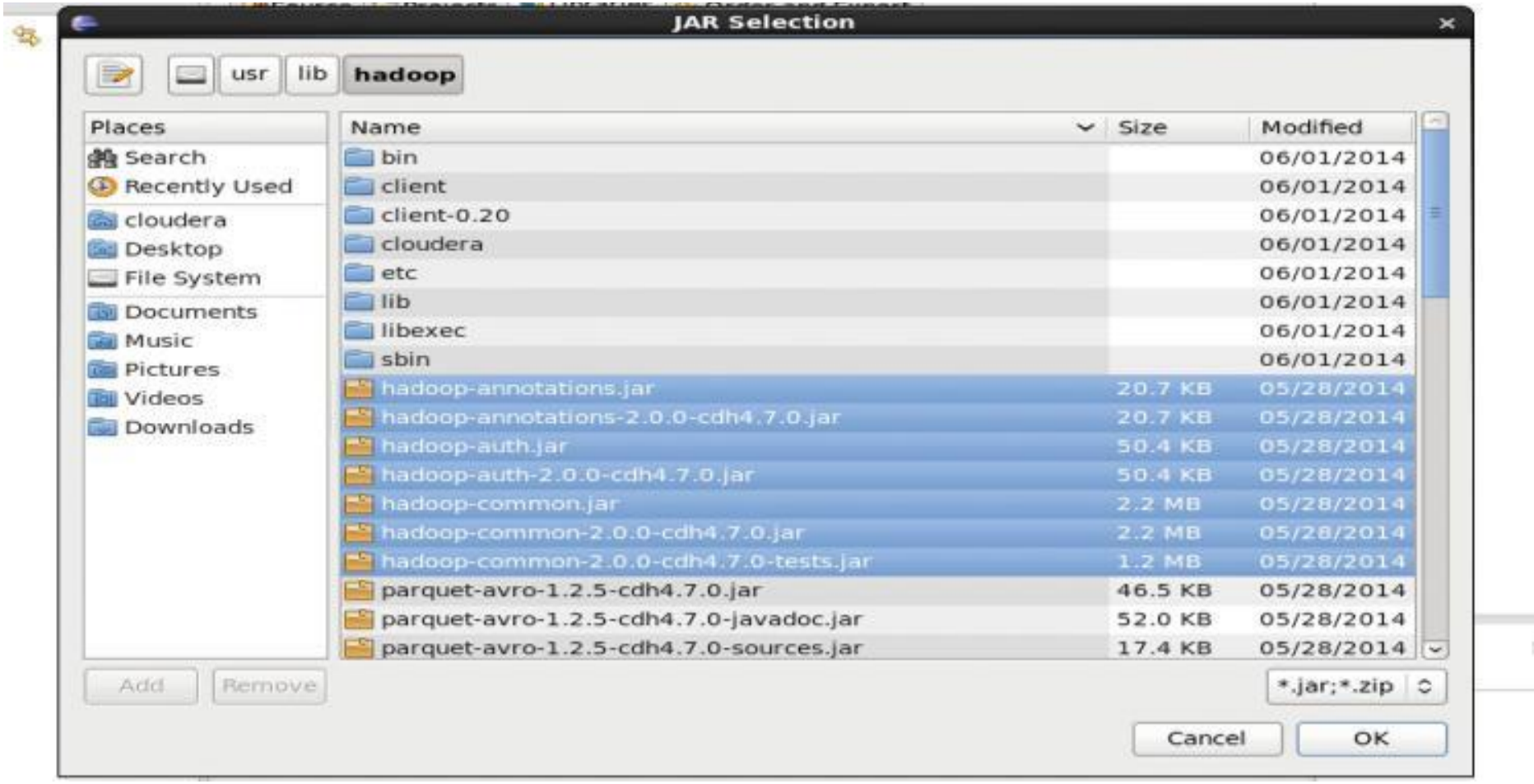
After that, click on libraries on the top and click on add external jars on the right side



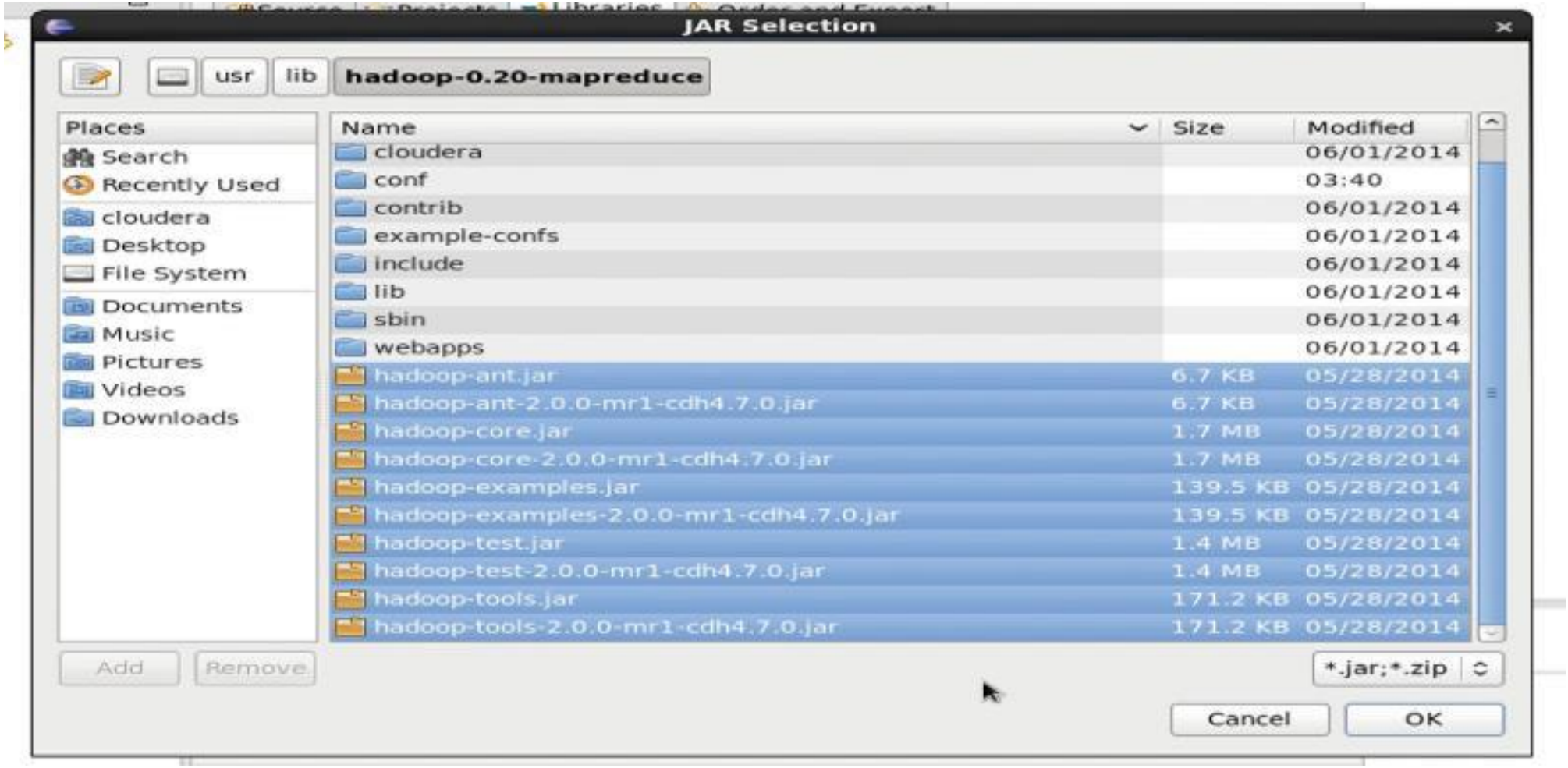
A dialog box will appear.



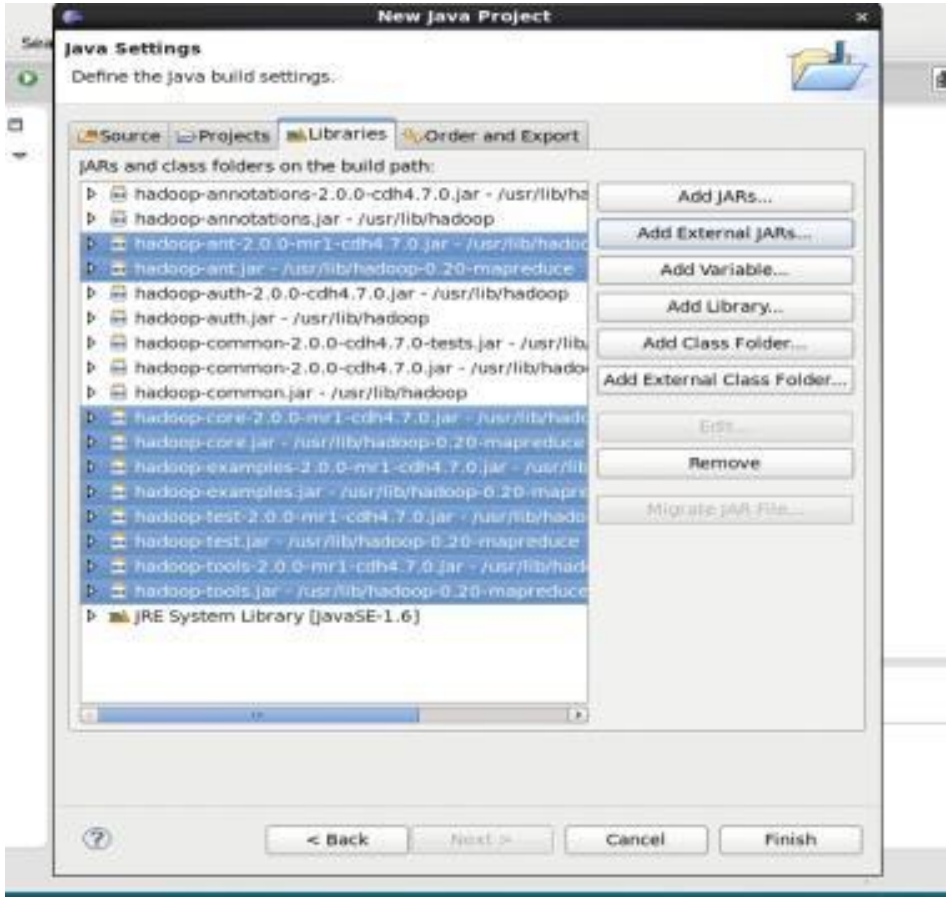
Click on file system on the left, and then usr->lib ->hadoop.
Select all the files named with hadoop and click ok.

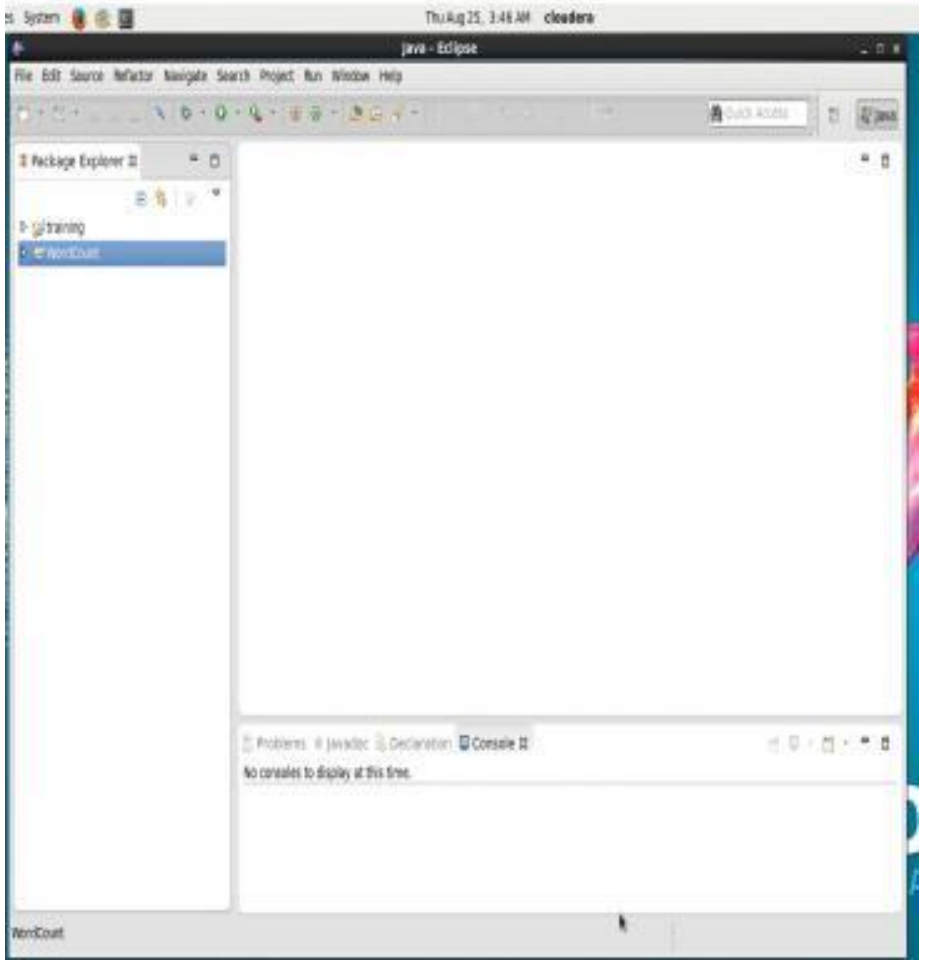


Go back to lib and click **hadoop-0.20-mapreduce** and select all the files named with hadoop and click ok.

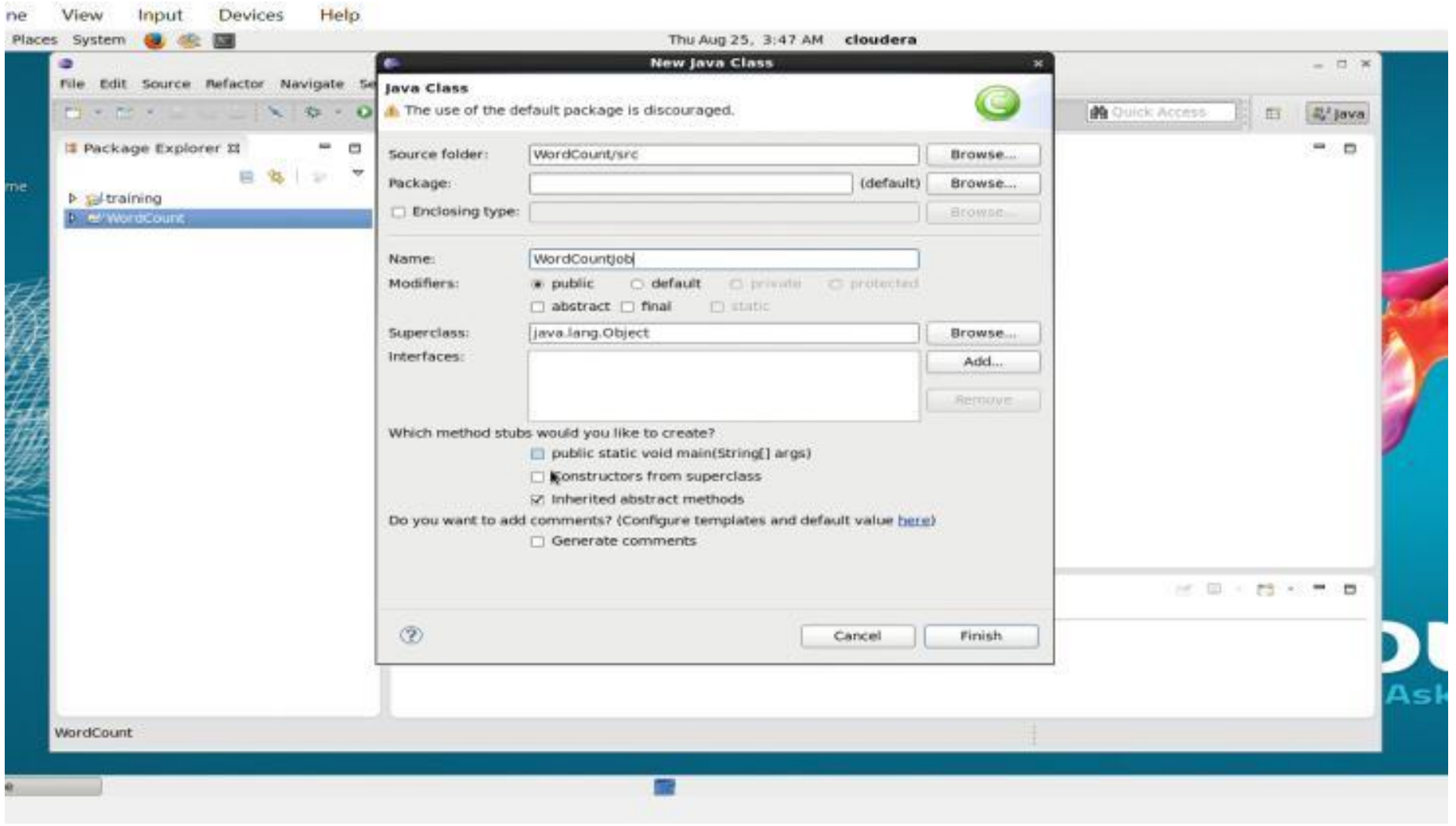


All the external files are added and now click on finish. The project is created successfully.

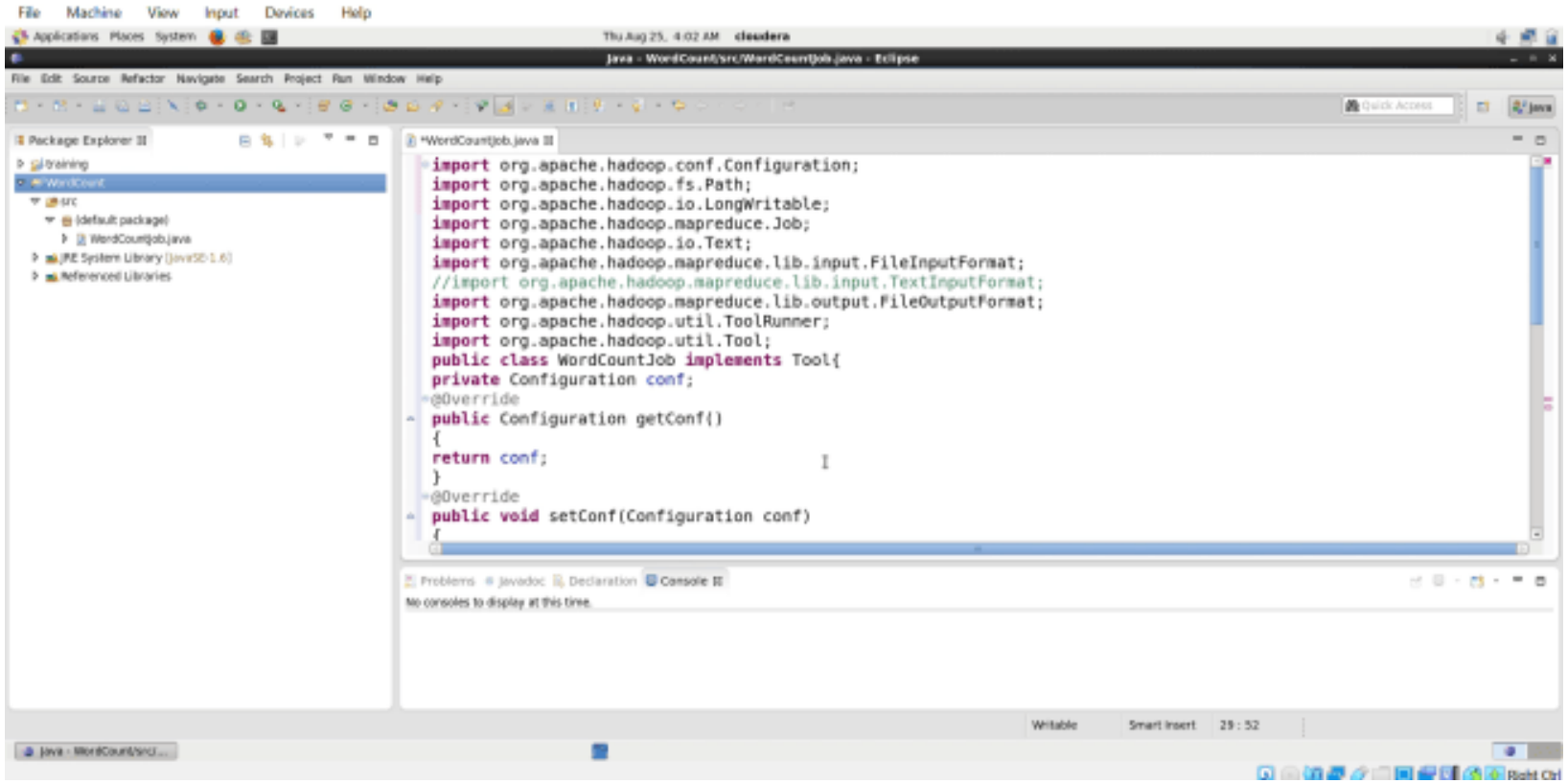




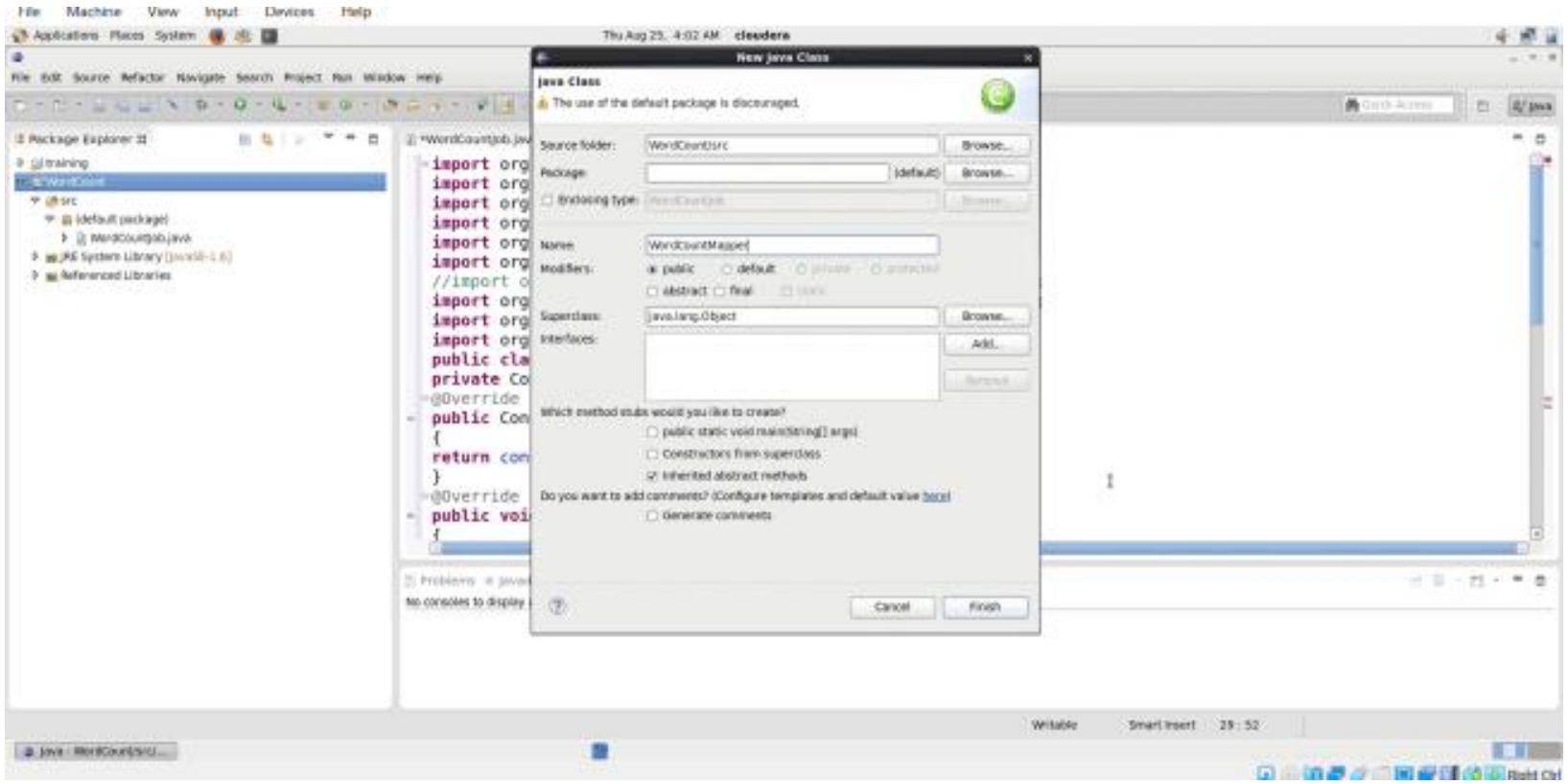
Create a class with name “**WordCountJob**”



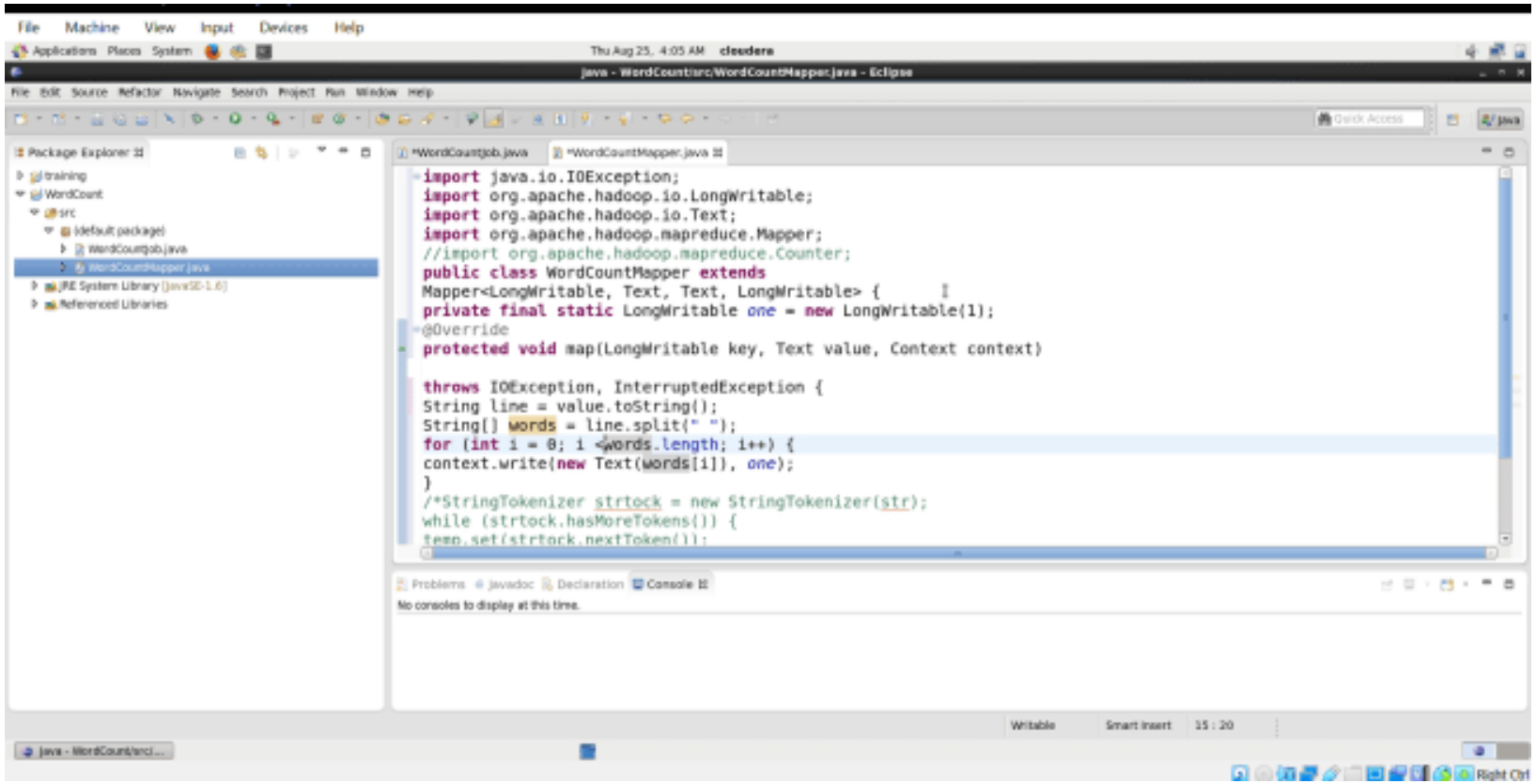
Copy and paste the program 4 lines at a time-
If copy and paste doesn't work then goto device in cloudera and set both shared clipboard and drag and drop as **bidirectional.**



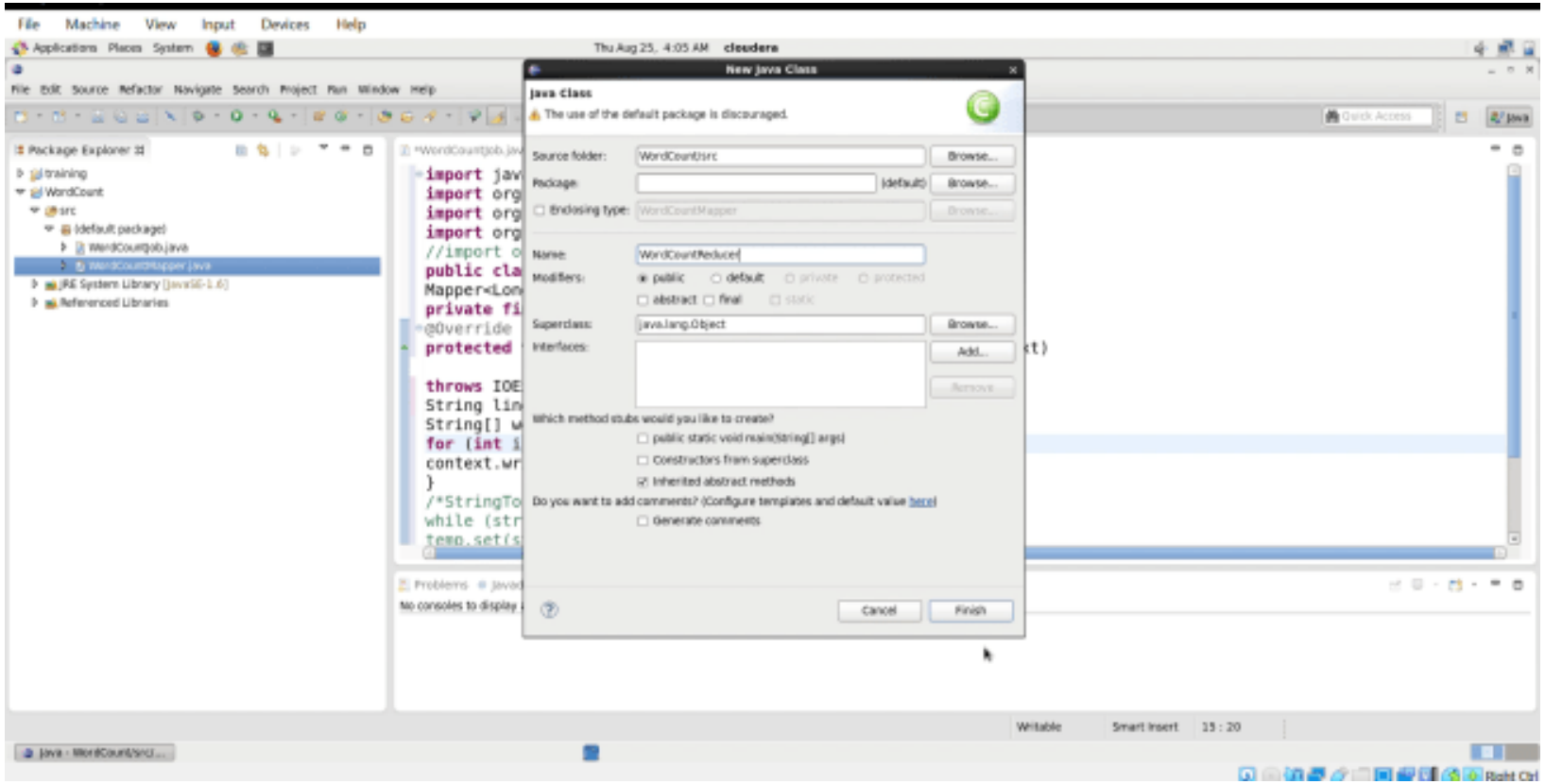
Create another class “**WordCountMapper**”



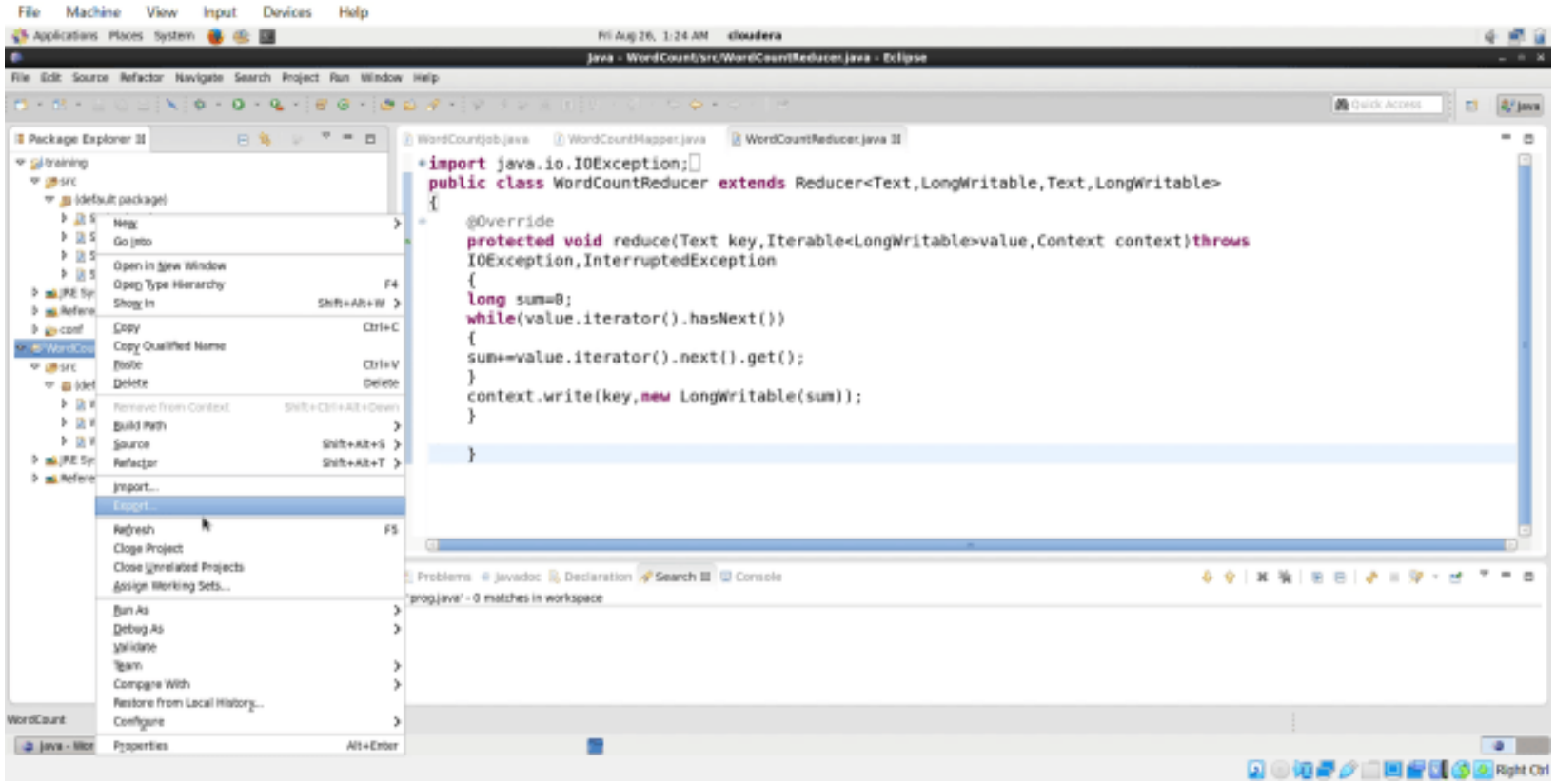
Copy and paste the program



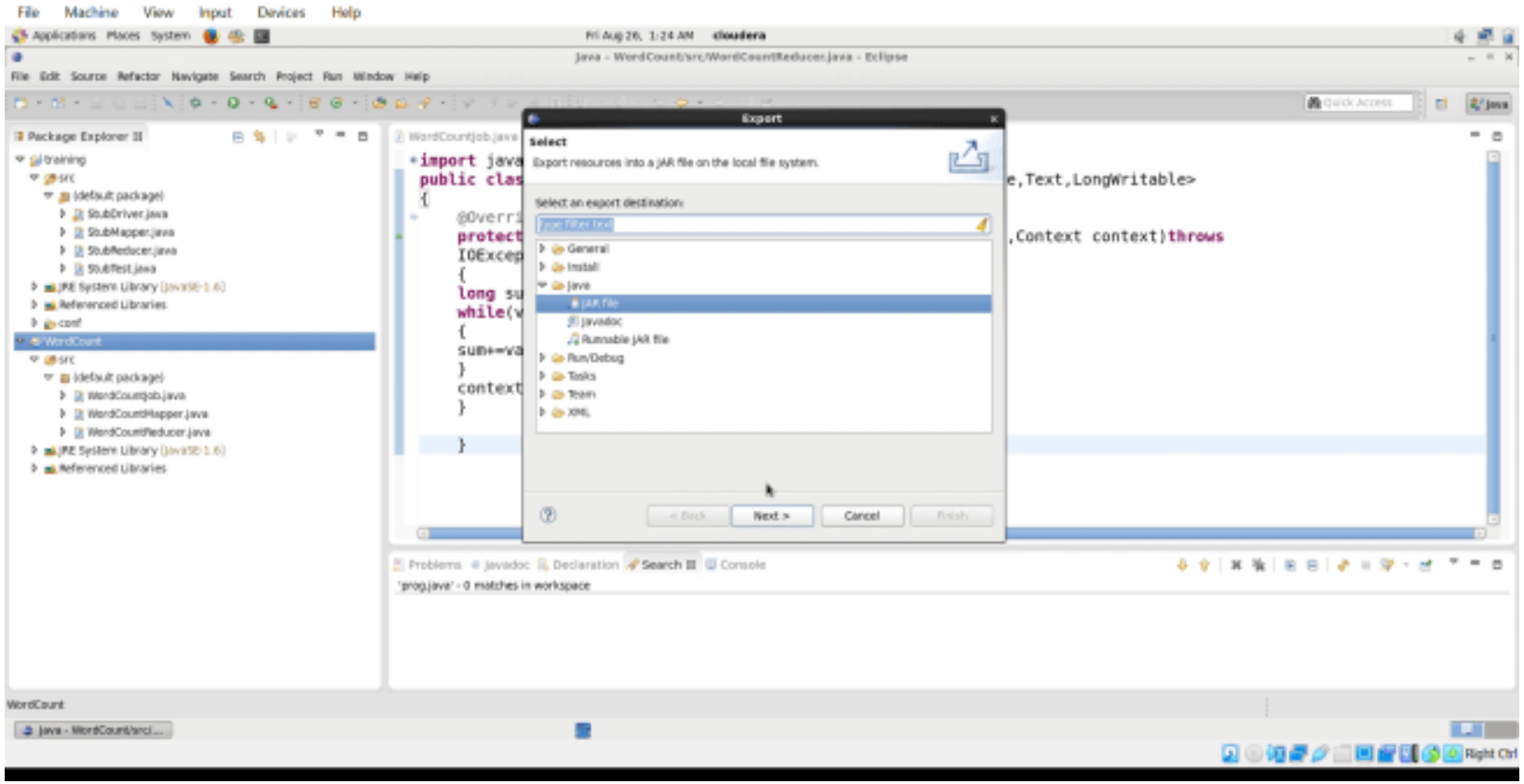
Create another class “**WordCountReducer**”



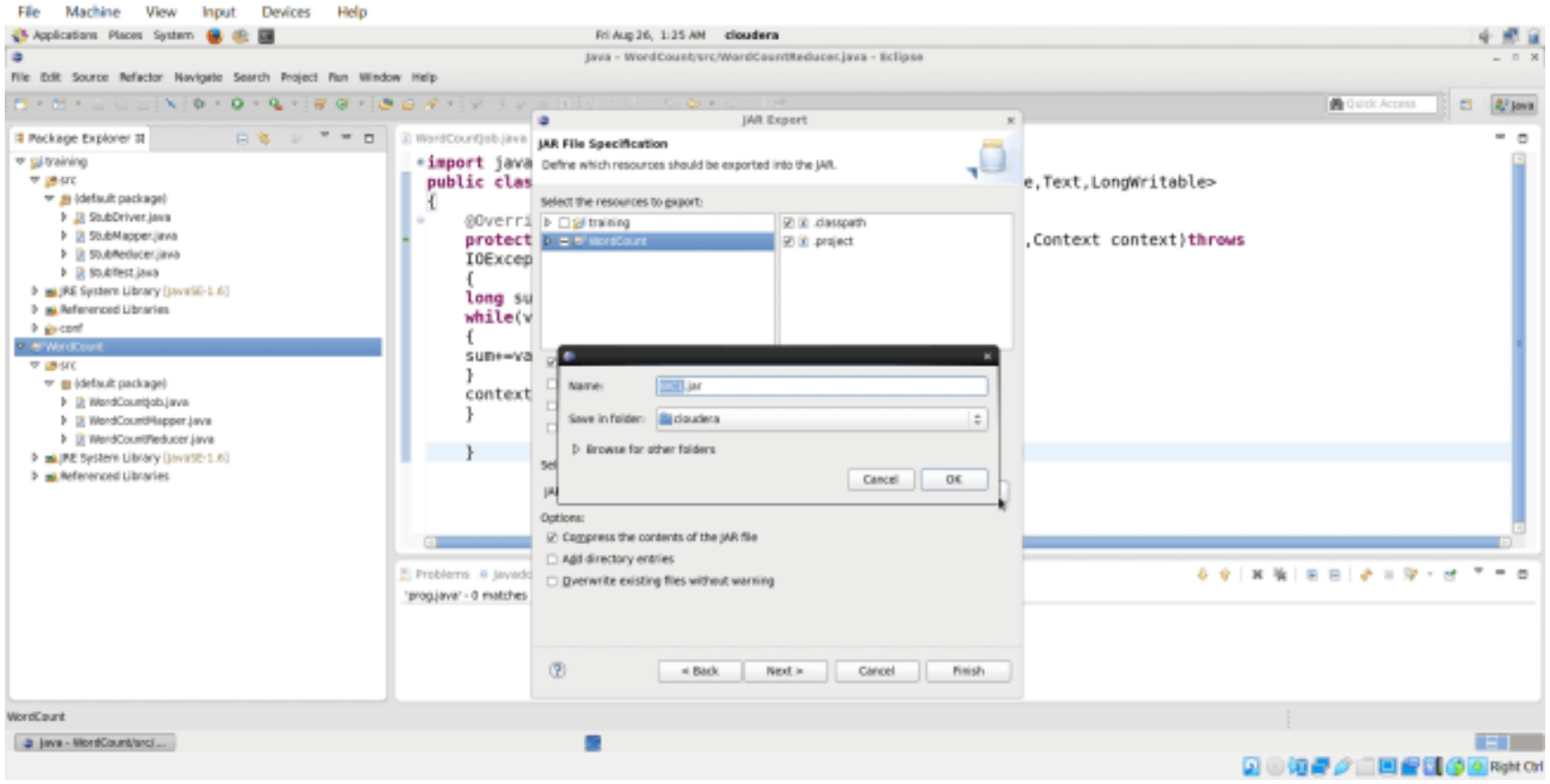
After finishing all the programs, right click on your project and select export.



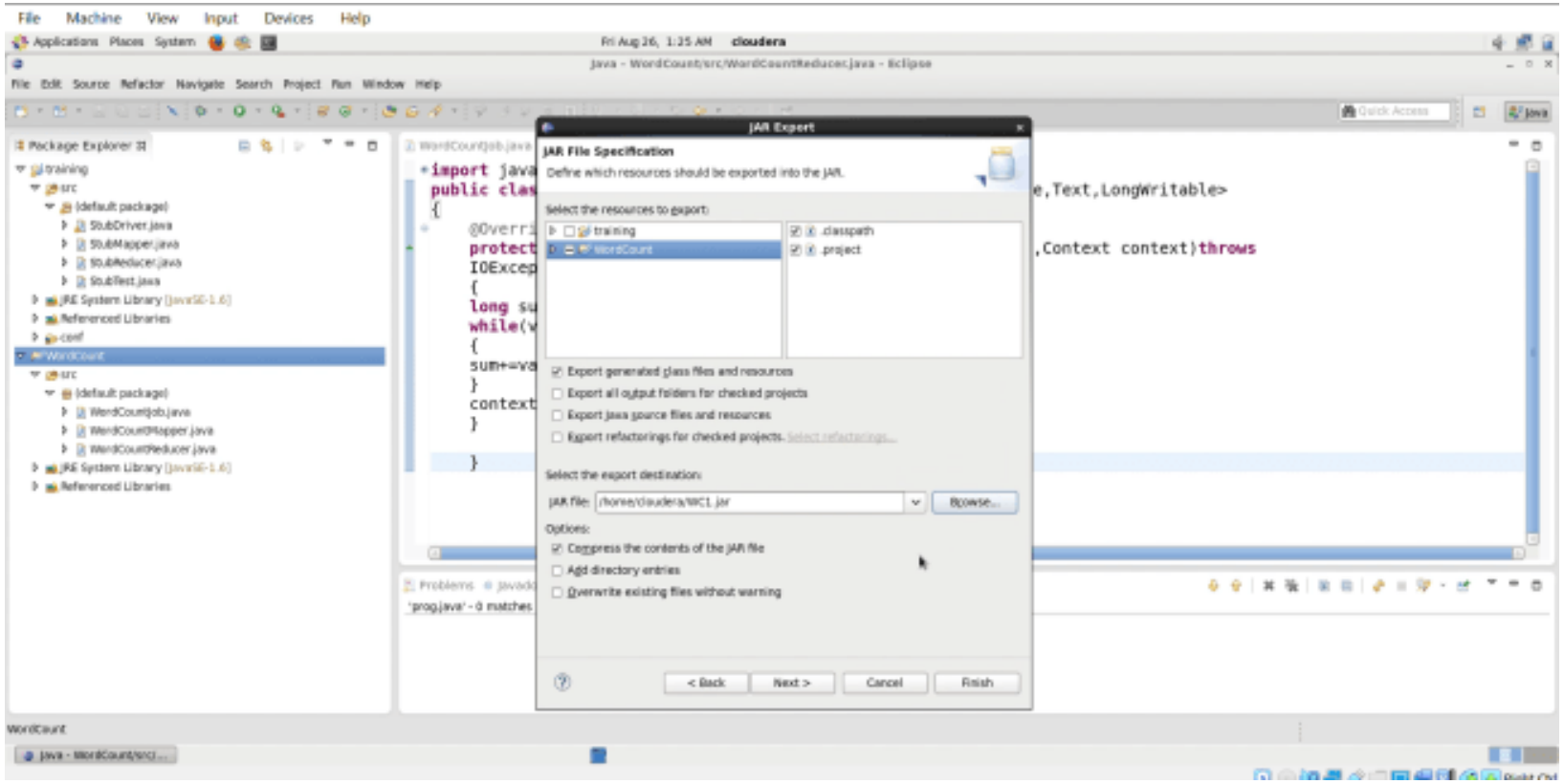
Select java-> JAR file->next



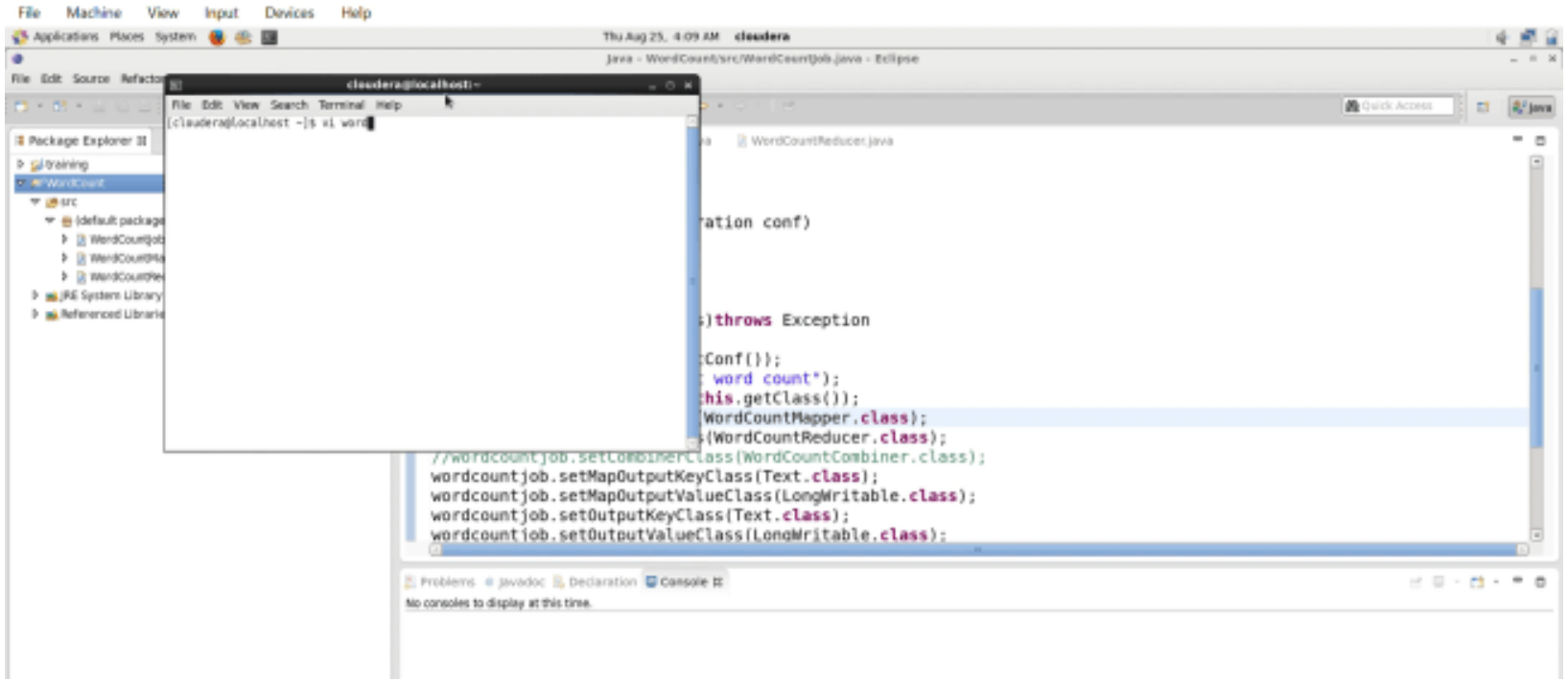
Click on browse and a dialog box will appear, give any name for the jar file and click ok.



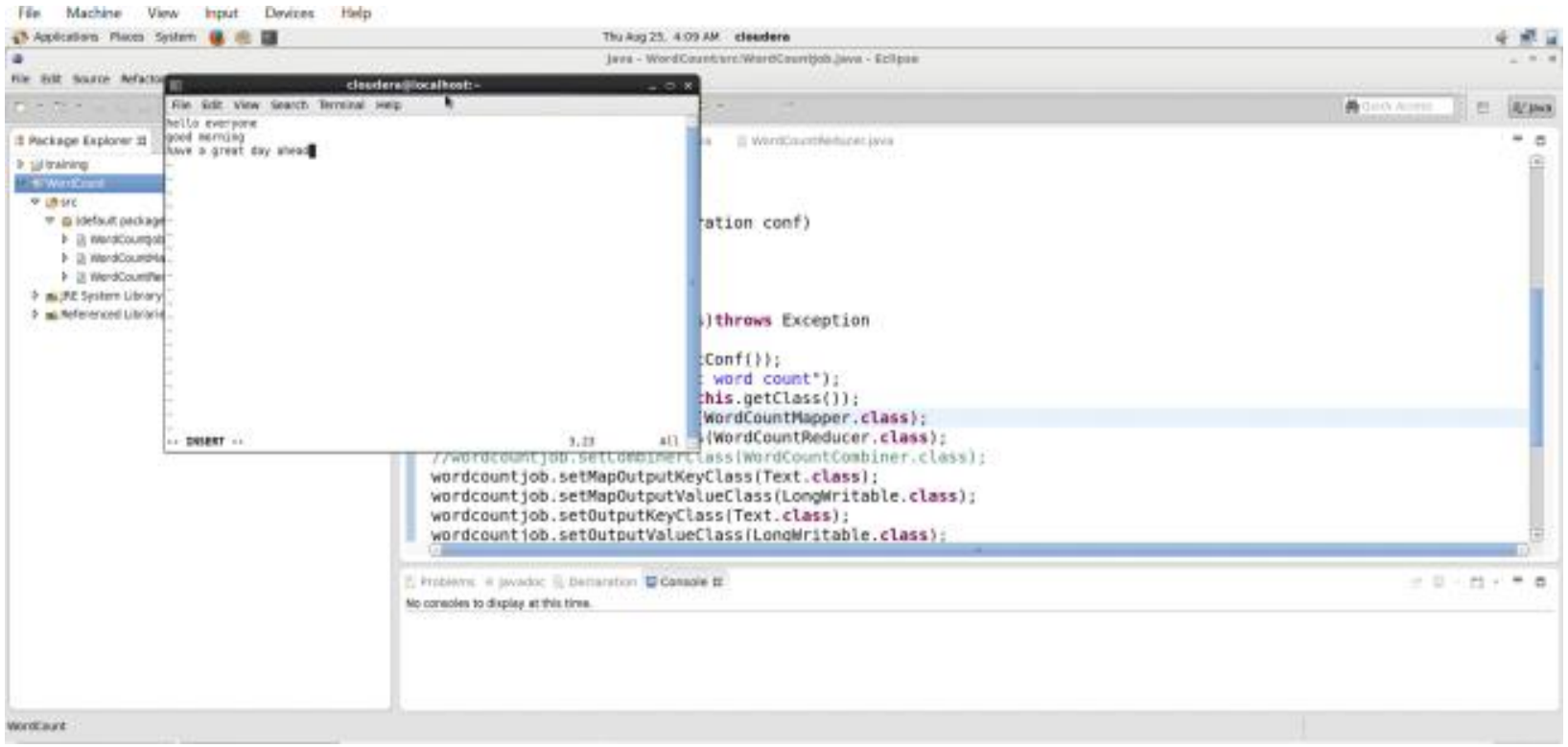
Click on finish.

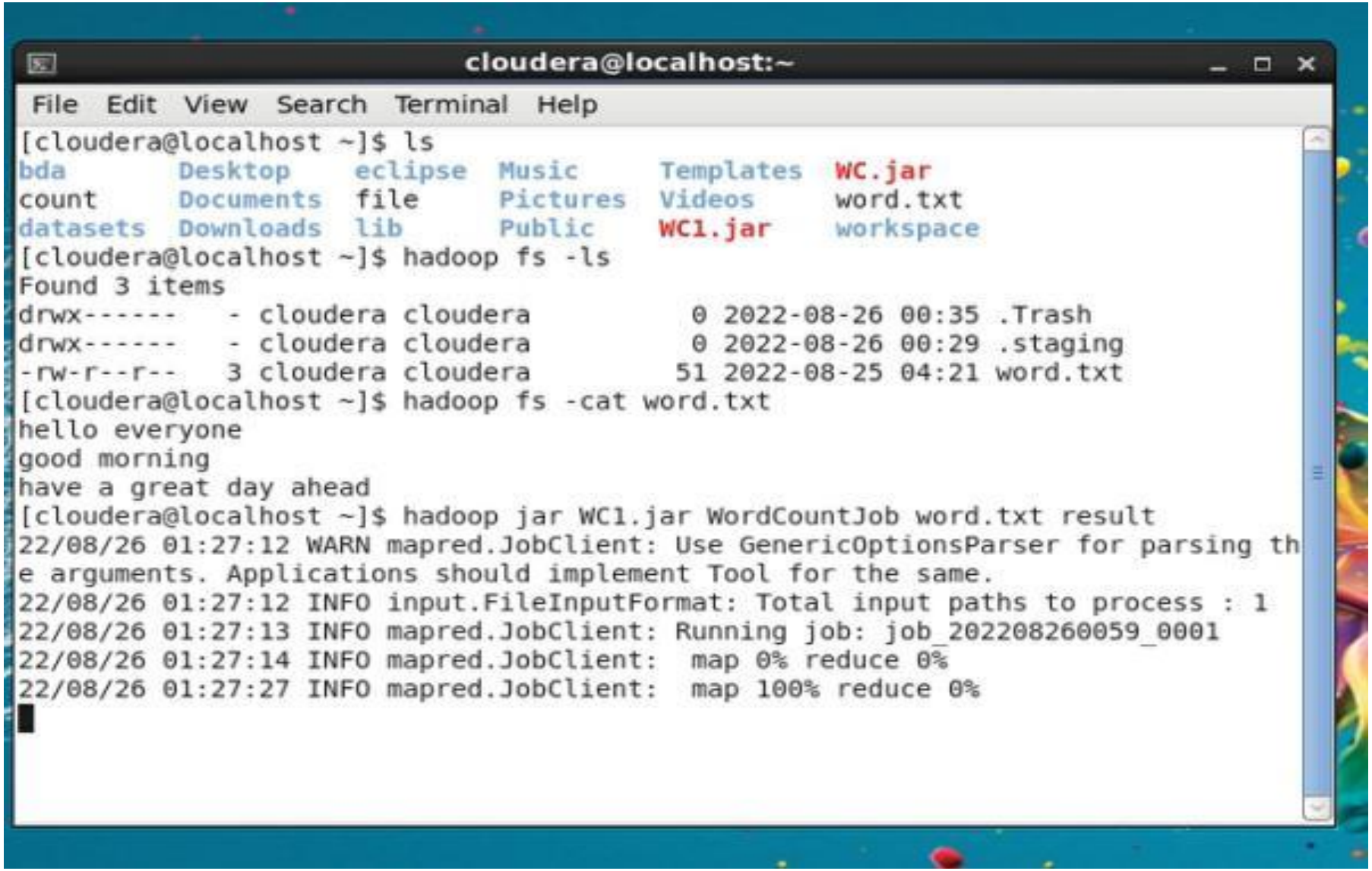


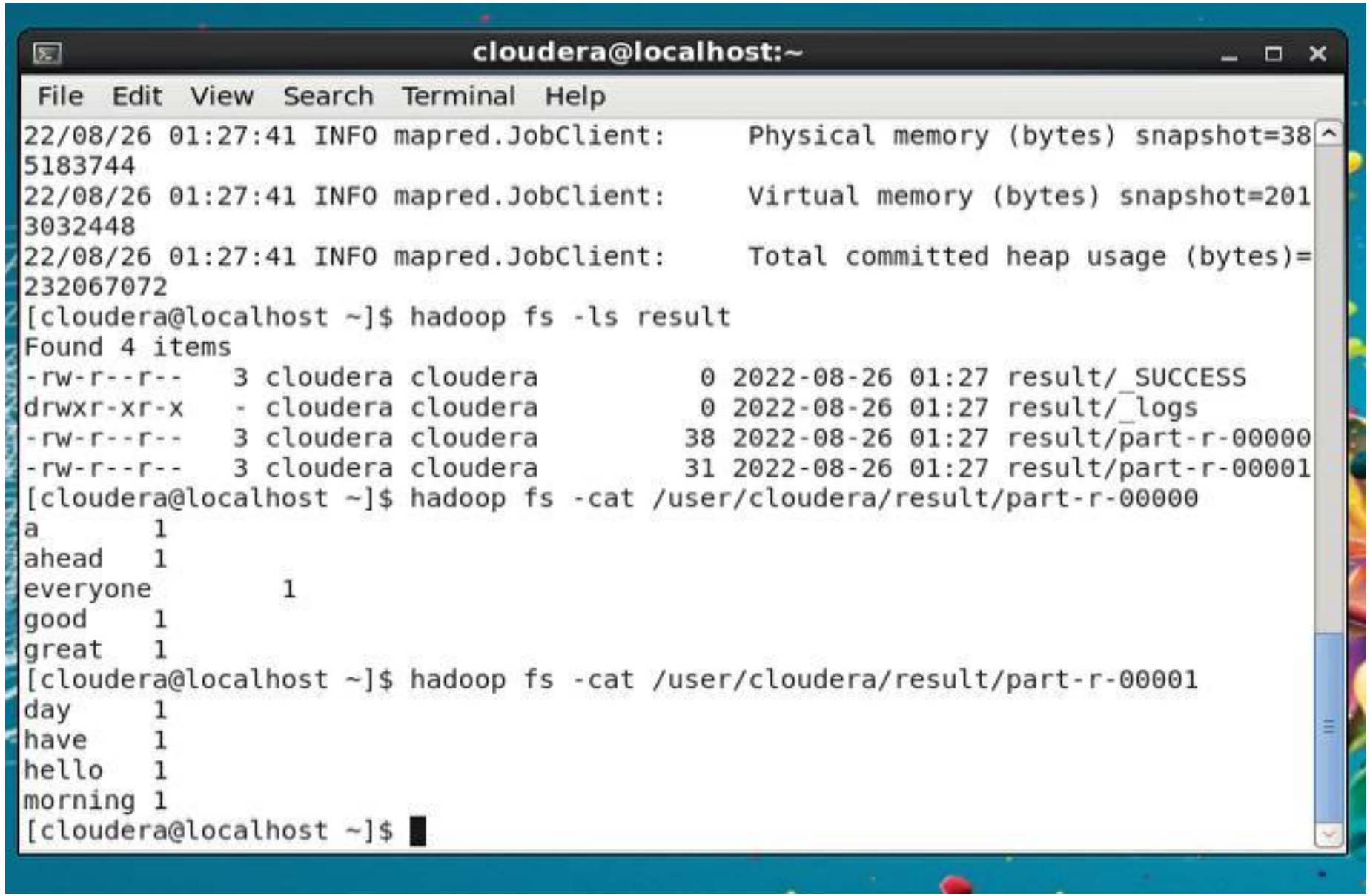
Now, open the terminal and create a text file “word”.



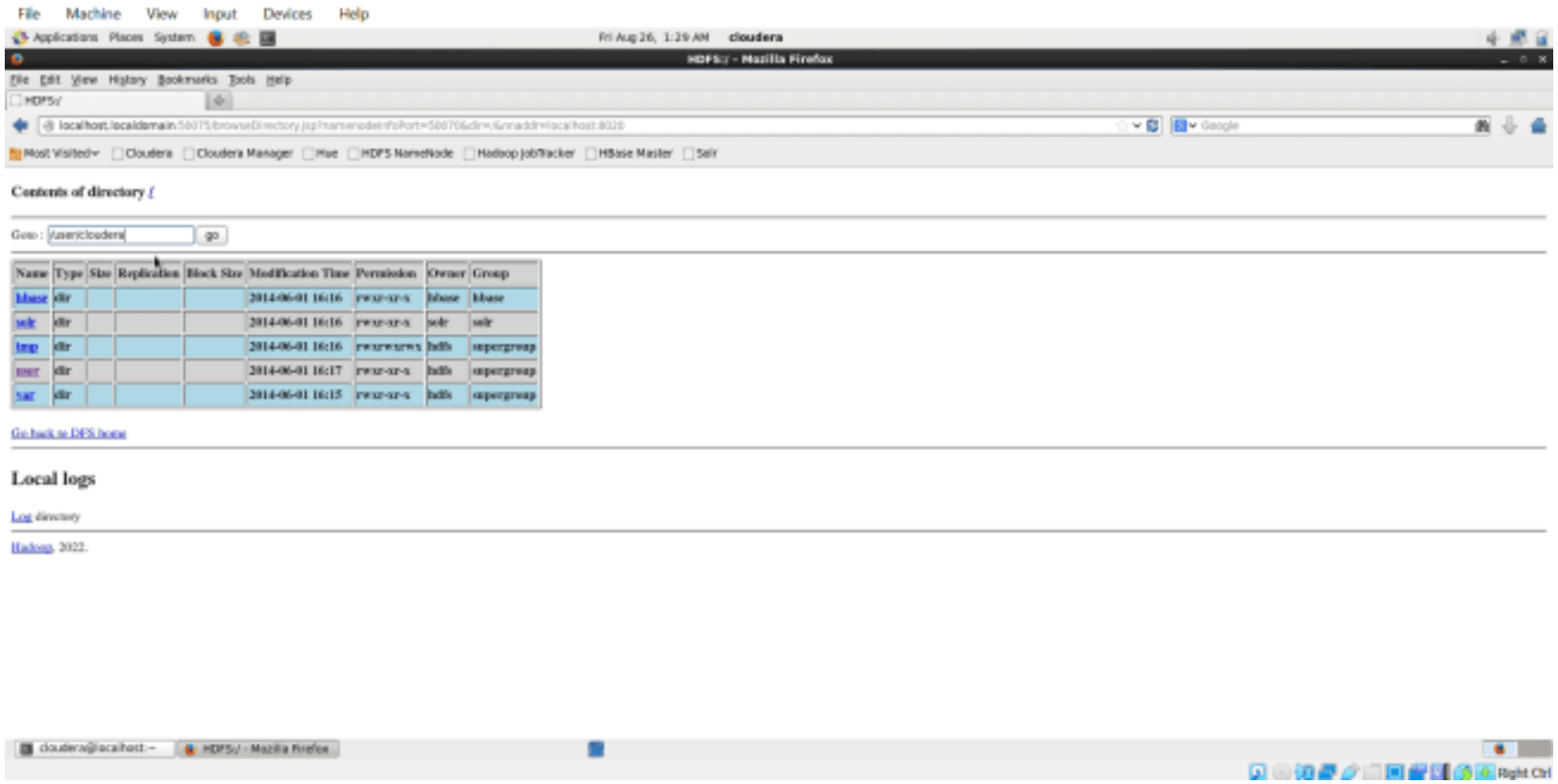
Insert few sentences in the file.

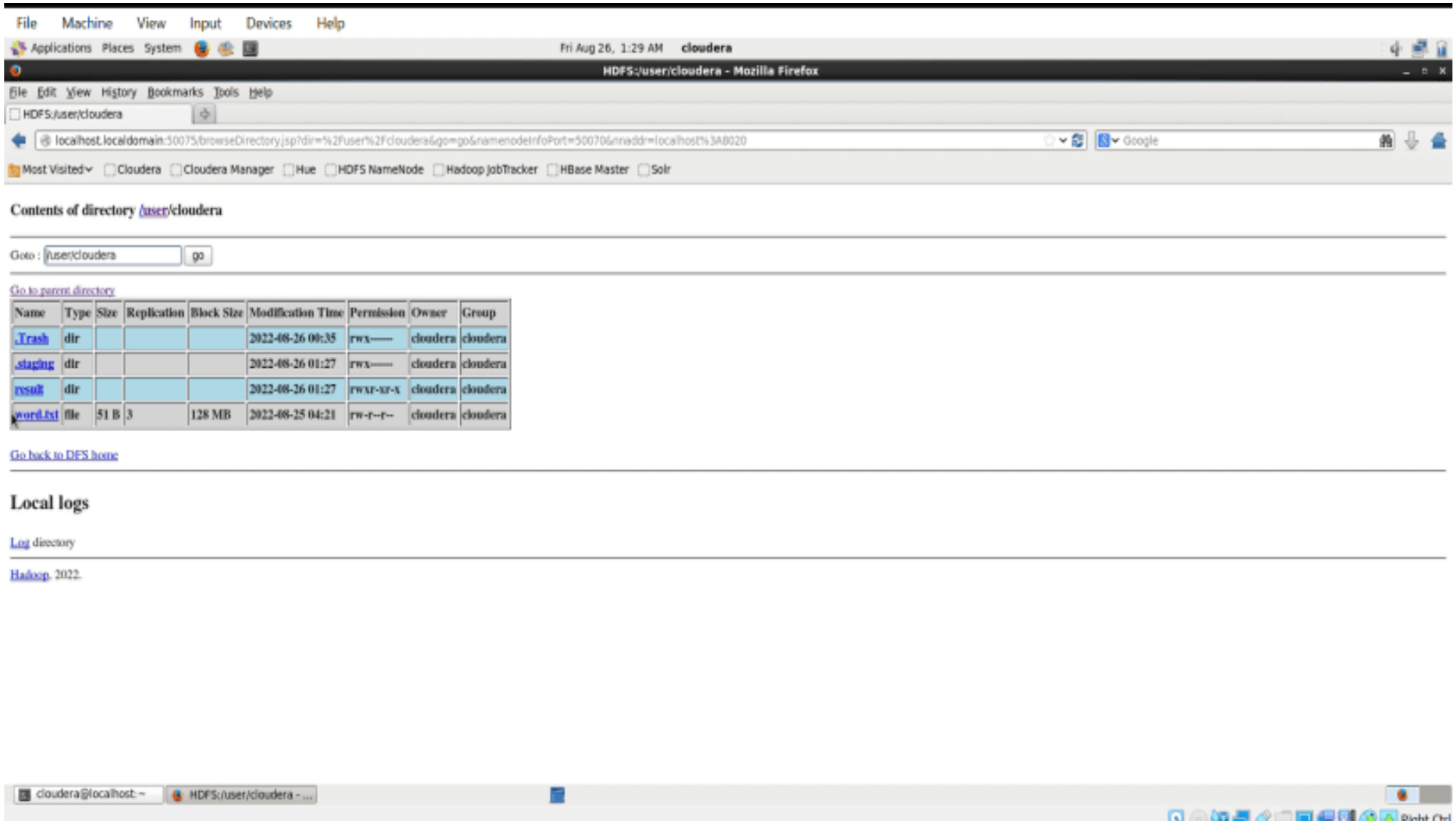


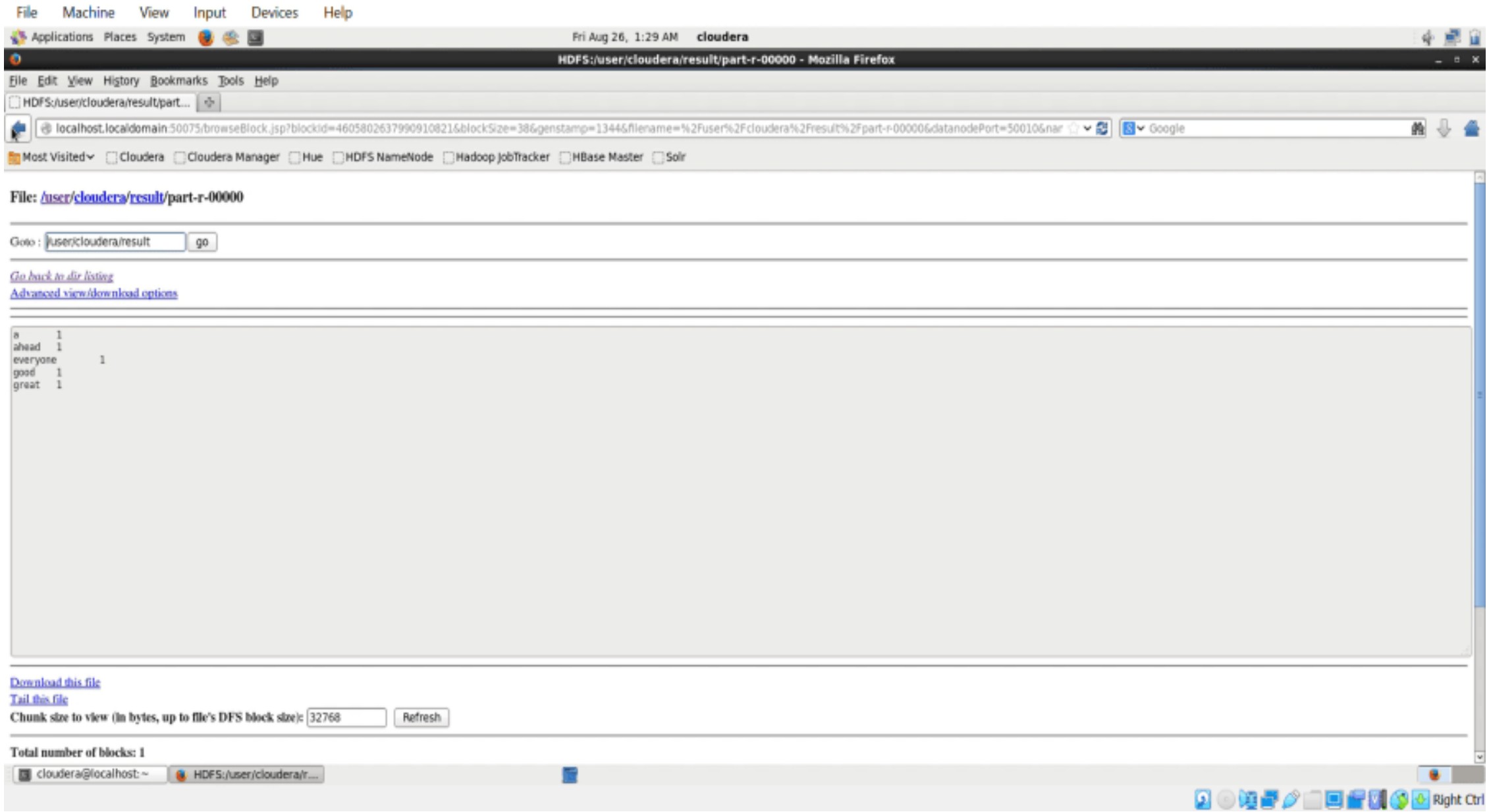

cloudera@localhost:~
File Edit View Search Terminal Help
[cloudera@localhost ~]$ ls
bda Desktop eclipse Music Templates WC.jar
count Documents file Pictures Videos word.txt
datasets Downloads lib Public WC1.jar workspace
[cloudera@localhost ~]$ hadoop fs -ls
Found 3 items
drwx------ - cloudera cloudera 0 2022-08-26 00:35 .Trash
drwx------ - cloudera cloudera 0 2022-08-26 00:29 .staging
-rw-r--r-- 3 cloudera cloudera 51 2022-08-25 04:21 word.txt
[cloudera@localhost ~]$ hadoop fs -cat word.txt
hello everyone
good morning
have a great day ahead
[cloudera@localhost ~]$ hadoop jar WC1.jar WordCountJob word.txt result
22/08/26 01:27:12 WARN mapred.JobClient: Use GenericOptionsParser for parsing th
e arguments. Applications should implement Tool for the same.
22/08/26 01:27:12 INFO input.FileInputFormat: Total input paths to process : 1
22/08/26 01:27:13 INFO mapred.JobClient: Running job: job_202208260059_0001
22/08/26 01:27:14 INFO mapred.JobClient: map 0% reduce 0%
22/08/26 01:27:27 INFO mapred.JobClient: map 100% reduce 0%

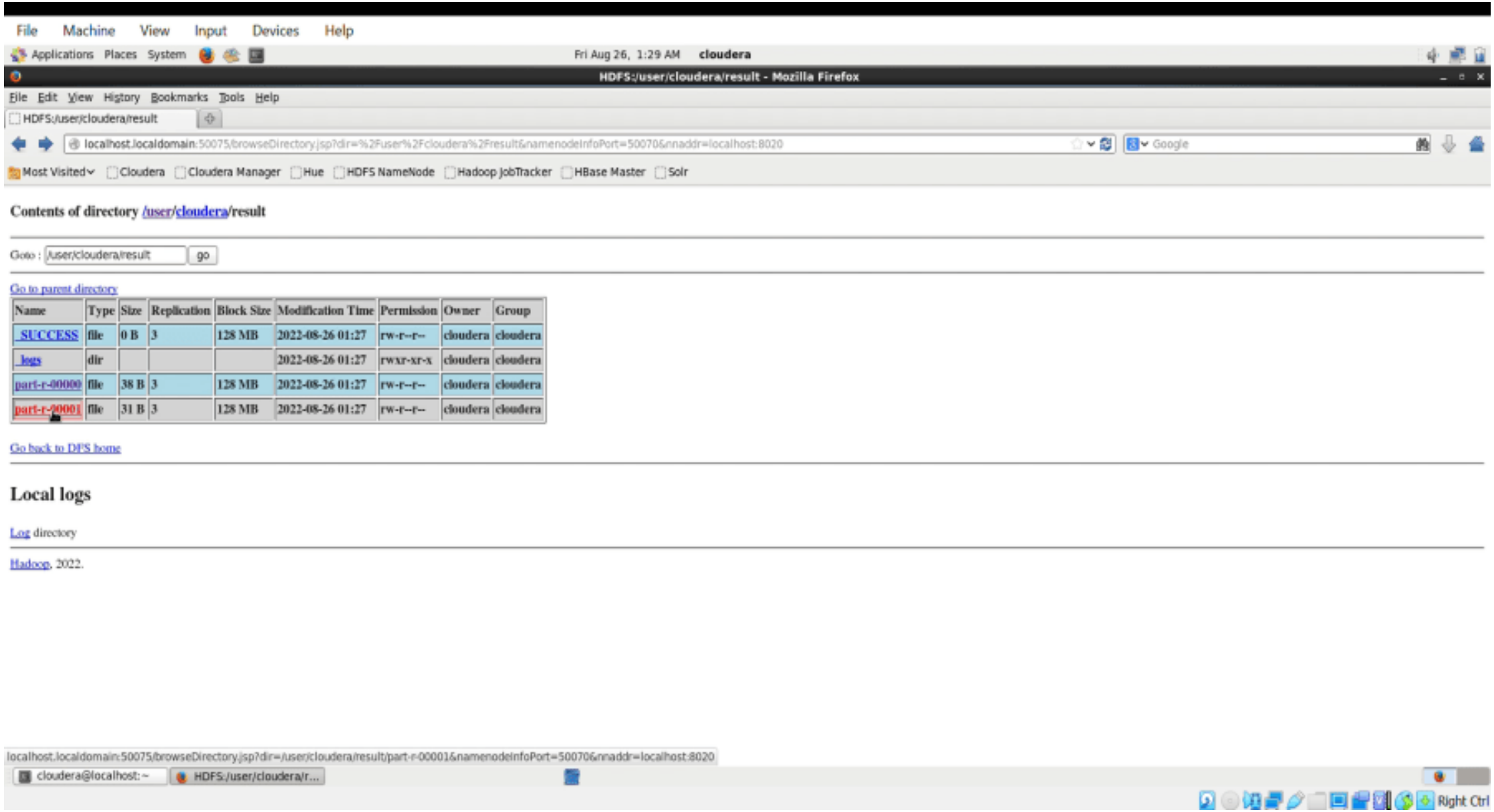

cloudera@localhost:~
File Edit View Search Terminal Help
22/08/26 01:27:41 INFO mapred.JobClient: Physical memory (bytes) snapshot=38
5183744
22/08/26 01:27:41 INFO mapred.JobClient: Virtual memory (bytes) snapshot=201
3032448
22/08/26 01:27:41 INFO mapred.JobClient: Total committed heap usage (bytes)=
232067072
[cloudera@localhost ~]$ hadoop fs -ls result
Found 4 items
-rw-r--r-- 3 cloudera cloudera 0 2022-08-26 01:27 result/_SUCCESS
drwxr-xr-x - cloudera cloudera 0 2022-08-26 01:27 result/_logs
-rw-r--r-- 3 cloudera cloudera 38 2022-08-26 01:27 result/part-r-00000
-rw-r--r-- 3 cloudera cloudera 31 2022-08-26 01:27 result/part-r-00001
[cloudera@localhost ~]$ hadoop fs -cat /user/cloudera/result/part-r-00000
a 1
ahead 1
everyone 1
good 1
great 1
[cloudera@localhost ~]$ hadoop fs -cat /user/cloudera/result/part-r-00001
day 1
have 1
hello 1
morning 1
[cloudera@localhost ~]$

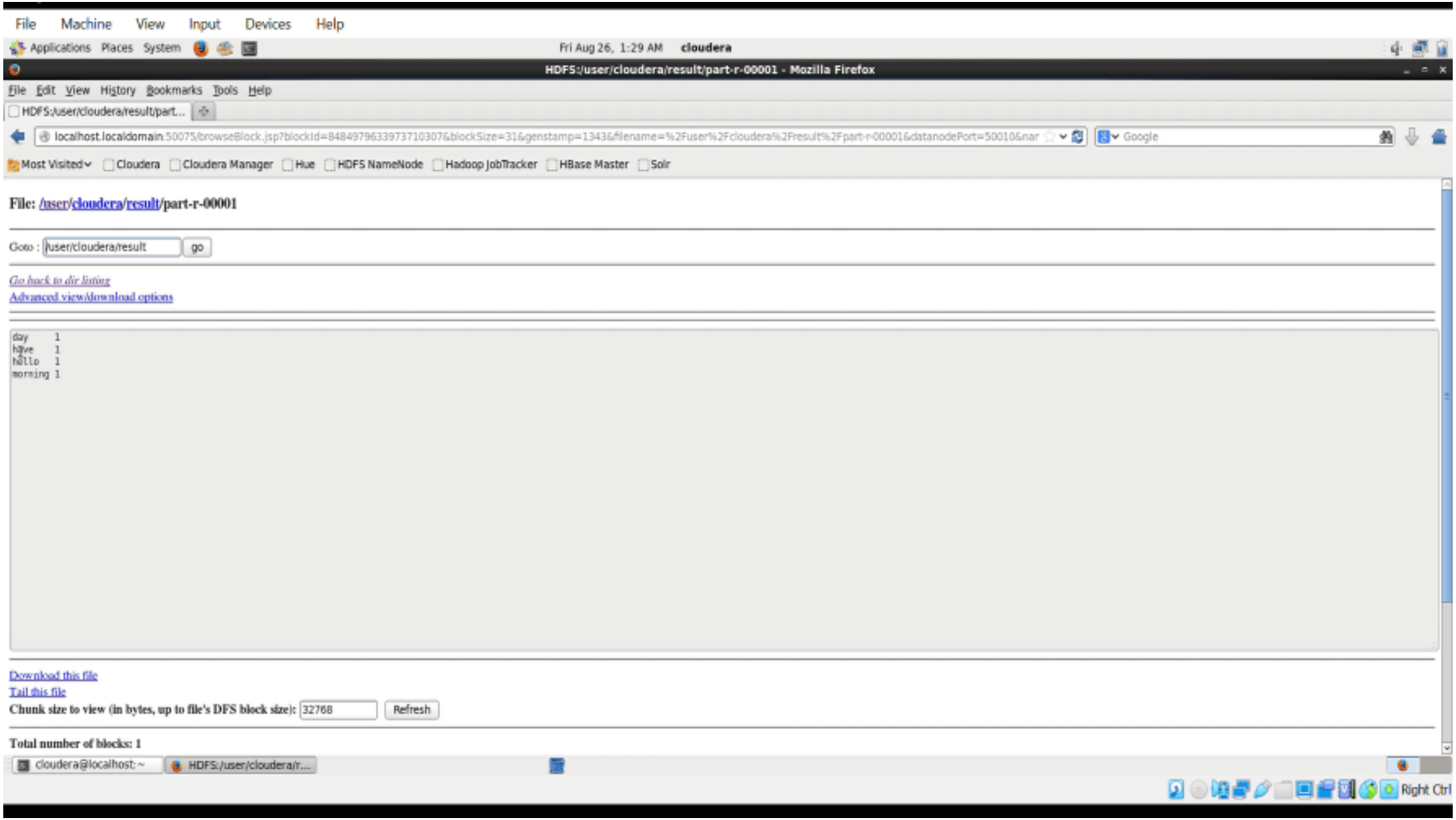
OUTPUT:


File Machine View Input Devices Help
Applications Places System
Fri Aug 26, 1:29 AM cloudera
HDFS:/ - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr
Contents of directory /
Goto : /user/cloudera go
Name Type Size Replication Block Size Modification Time Permission Owner Group
hbase dir 2014-06-01 16:16 rwxr-xr-x hbase hbase
solr dir 2014-06-01 16:16 rwxr-xr-x solr solr
tmp dir 2014-06-01 16:16 rwxrwxrwx hdfs supergroup
user dir 2014-06-01 16:17 rwxr-xr-x hdfs supergroup
var dir 2014-06-01 16:15 rwxr-xr-x hdfs supergroup
Go back to DFS home
Local logs
Log directory
Hadoop, 2022.
cloudera@localhost:~
HDFS:/ - Mozilla Firefox
Right Ctrl

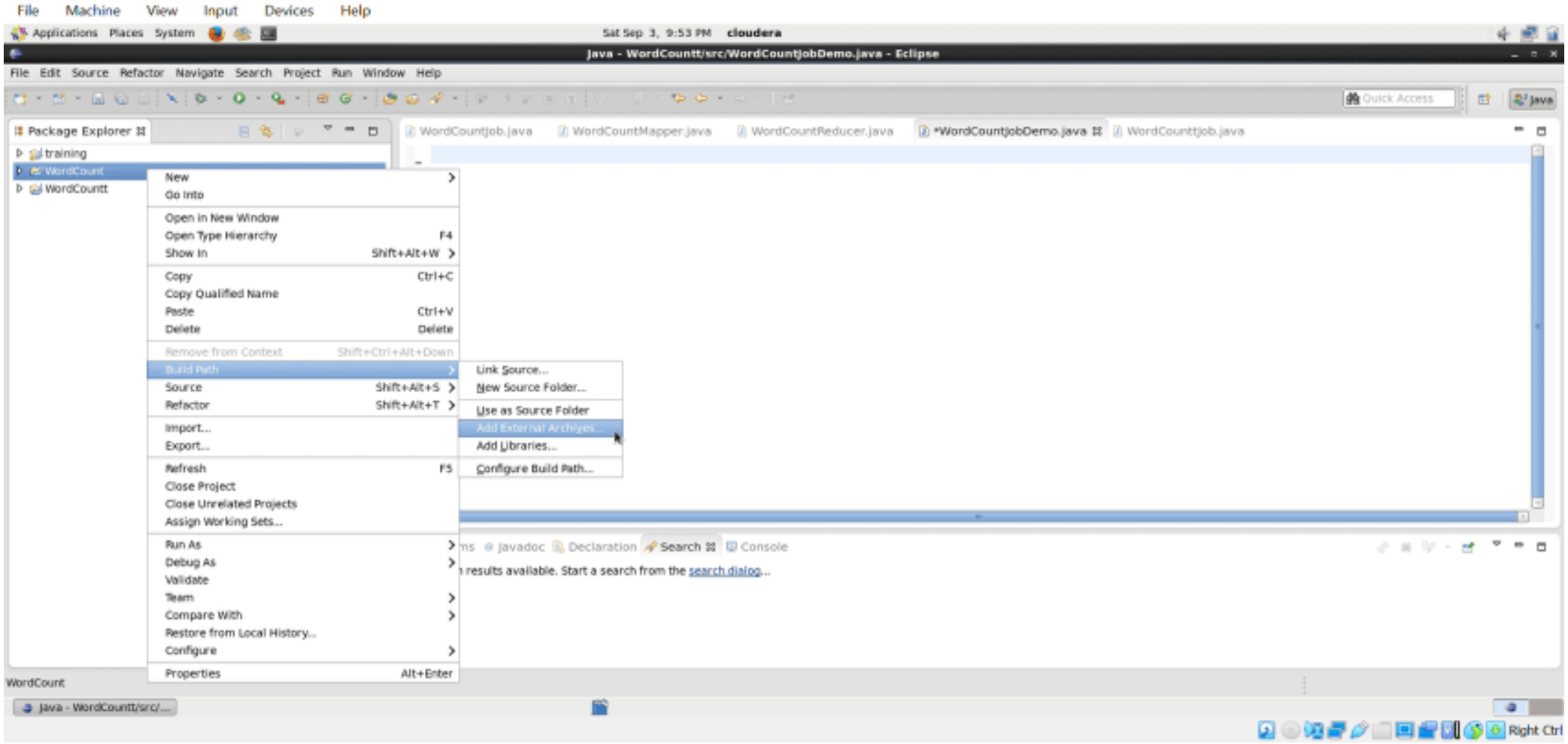
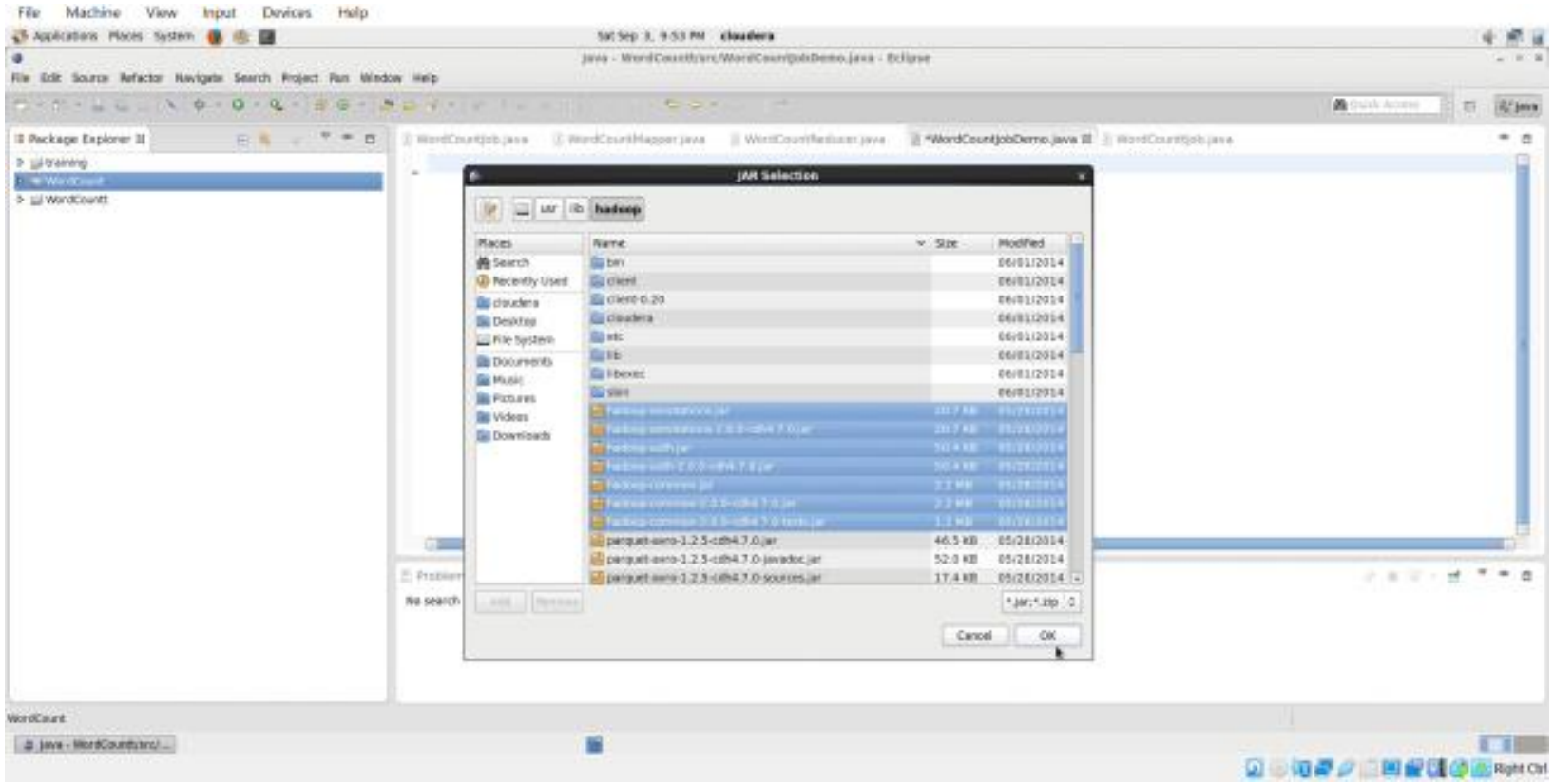

File Machine View Input Devices Help
Applications Places System
Fri Aug 26, 1:29 AM cloudera
HDFS:/user/cloudera - Mozilla Firefox
File Edit View History Bookmarks Tools Help
HDFS:/user/cloudera
localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2Fcloudera&go=go&namenodeInfoPort=50070&nnaddr=localhost%3A8020
Google
Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr
Contents of directory /user/cloudera
Goto : /user/cloudera go
Go to parent directory
Name Type Size Replication Block Size Modification Time Permission Owner Group
.Trash dir 2022-08-26 00:35 rwx------ cloudera cloudera
.staging dir 2022-08-26 01:27 rwx------ cloudera cloudera
result dir 2022-08-26 01:27 rwxr-xr-x cloudera cloudera
word.txt file 51 B 3 128 MB 2022-08-25 04:21 rw-r--r-- cloudera cloudera
Go back to DFS home
Local logs
Log directory
Hadoop, 2022.
cloudera@localhost:~
HDFS:/user/cloudera - ...


File Machine View Input Devices Help
Applications Places System
Fri Aug 26, 1:29 AM cloudera
HDFS:/user/cloudera/result/part-r-00000 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
HDFS:/user/cloudera/result/part...
localhost.localdomain:50075/browseBlock.jsp?blockId=4605802637990910821&blockSize=38&genstamp=1344&filename=%2Fuser%2Fcloudera%2Fresult%2Fpart-r-00000&datanodePort=50010&nar
Google
Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr
File: /user/cloudera/result/part-r-00000
Goto : /user/cloudera/result go
Go back to dir listing
Advanced view/download options
a 1
ahead 1
everyone 1
good 1
great 1
Download this file
Tail this file
Chunk size to view (in bytes, up to file's DFS block size): 32768 Refresh
Total number of blocks: 1
cloudera@localhost:~
HDFS:/user/cloudera/r...
Right Ctrl


File Machine View Input Devices Help
Applications Places System
Fri Aug 26, 1:29 AM cloudera
HDFS:/user/cloudera/result - Mozilla Firefox
File Edit View History Bookmarks Tools Help
HDFS:/user/cloudera/result
localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2Fcloudera%2Fresult&namenodeInfoPort=50070&nnaddr=localhost:8020
Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr
Contents of directory /user/cloudera/result
Goto : /user/cloudera/result go
Go to parent directory
Name Type Size Replication Block Size Modification Time Permission Owner Group
_SUCCESS file 0 B 3 128 MB 2022-08-26 01:27 rw-r--r-- cloudera cloudera
_logs dir 2022-08-26 01:27 rwxr-xr-x cloudera cloudera
part-r-00000 file 38 B 3 128 MB 2022-08-26 01:27 rw-r--r-- cloudera cloudera
part-r-00001 file 31 B 3 128 MB 2022-08-26 01:27 rw-r--r-- cloudera cloudera
Go back to DFS home
Local logs
Log directory
Hadoop, 2022.
localhost.localdomain:50075/browseDirectory.jsp?dir=/user/cloudera/result/part-r-00001&namenodeInfoPort=50706&nnaddr=localhost:8020
cloudera@localhost:~
HDFS:/user/cloudera/r...
Right Ctrl


File Machine View Input Devices Help
Applications Places System
Fri Aug 26, 1:29 AM cloudera
HDFS:/user/cloudera/result/part-r-00001 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
HDFS:/user/cloudera/result/part...
localhost.localdomain:50075/browseBlock.jsp?blockId=8484979633973710307&blockSize=31&genstamp=1343&filename=%2Fuser%2Fcloudera%2Fresult%2Fpart-r-00001&datanodePort=50010&nar
Google
Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr
File: /user/cloudera/result/part-r-00001
Goto : /user/cloudera/result go
Go back to dir listing
Advanced view/download options
day 1
have 1
hello 1
morning 1
Download this file
Tail this file
Chunk size to view (in bytes, up to file's DFS block size): 32768 Refresh
Total number of blocks: 1
cloudera@localhost:~
HDFS:/user/cloudera/r...
Right Ctrl

If there is any error in executing then set the path- right click on your project->build path->add external archives



Add all the hadoop and mapreduce external jar files and click ok.