732\_Saksham Shetkar

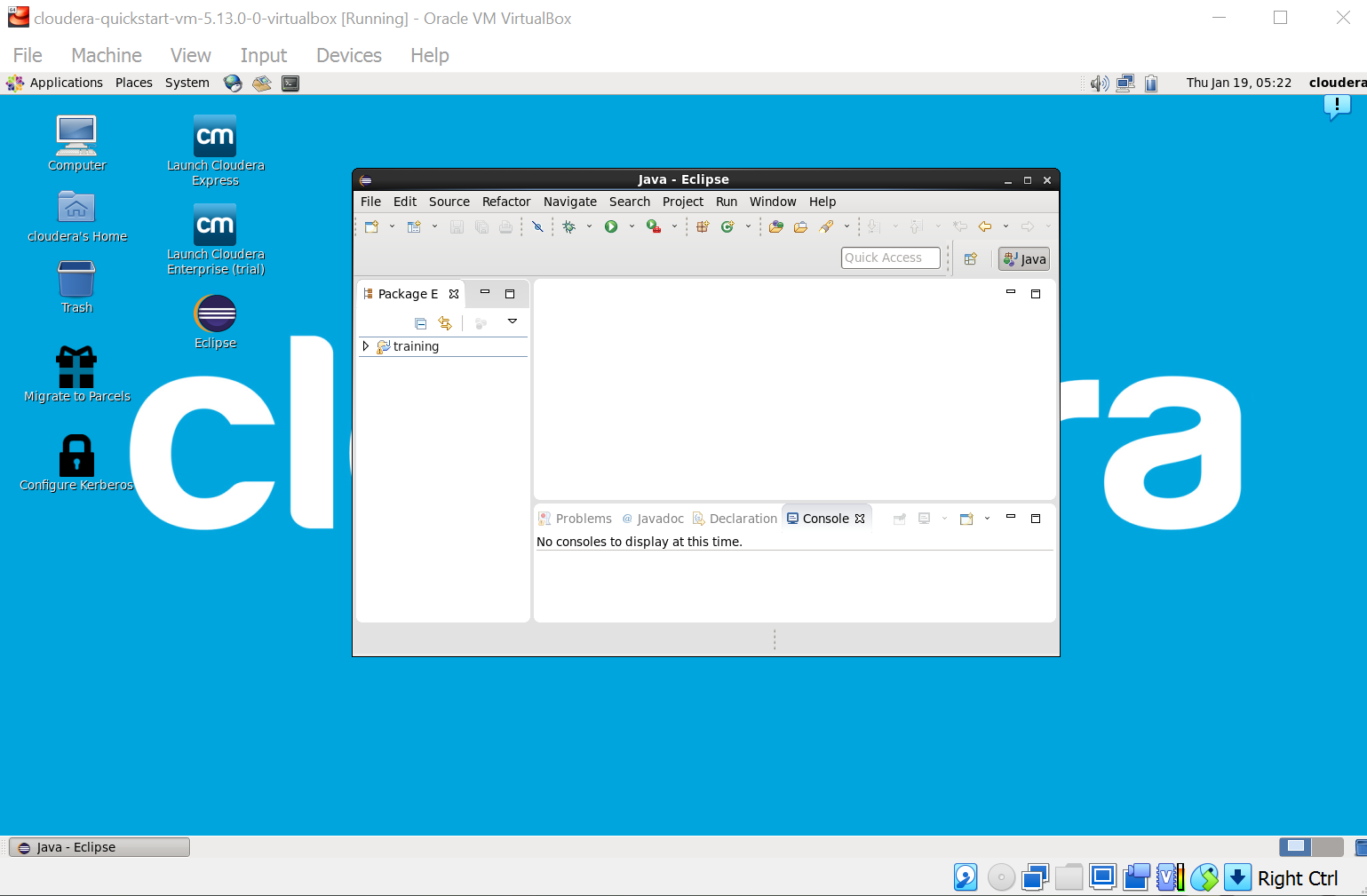
Big Data technology

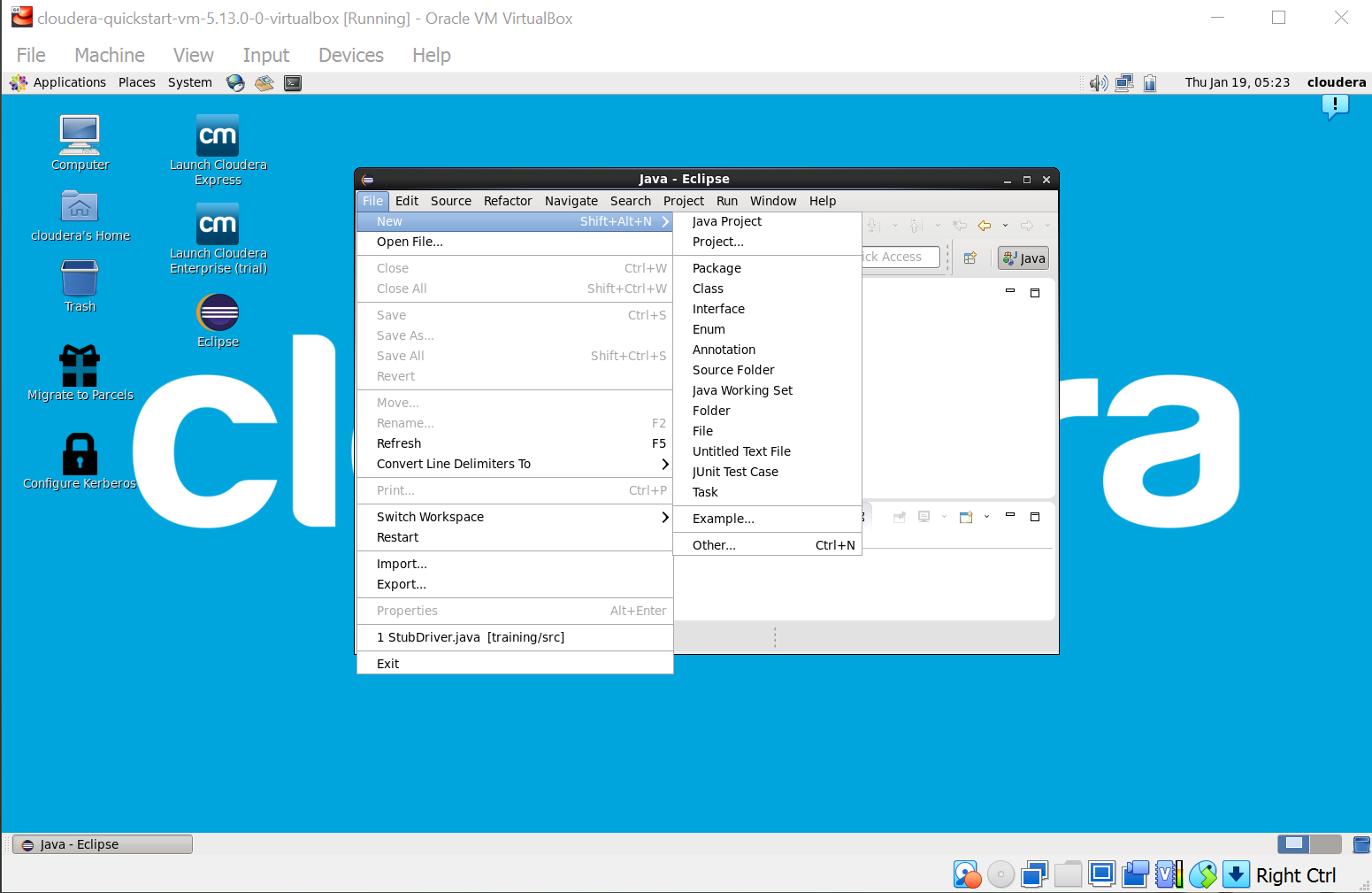
MapReduce

**Practical : 01**

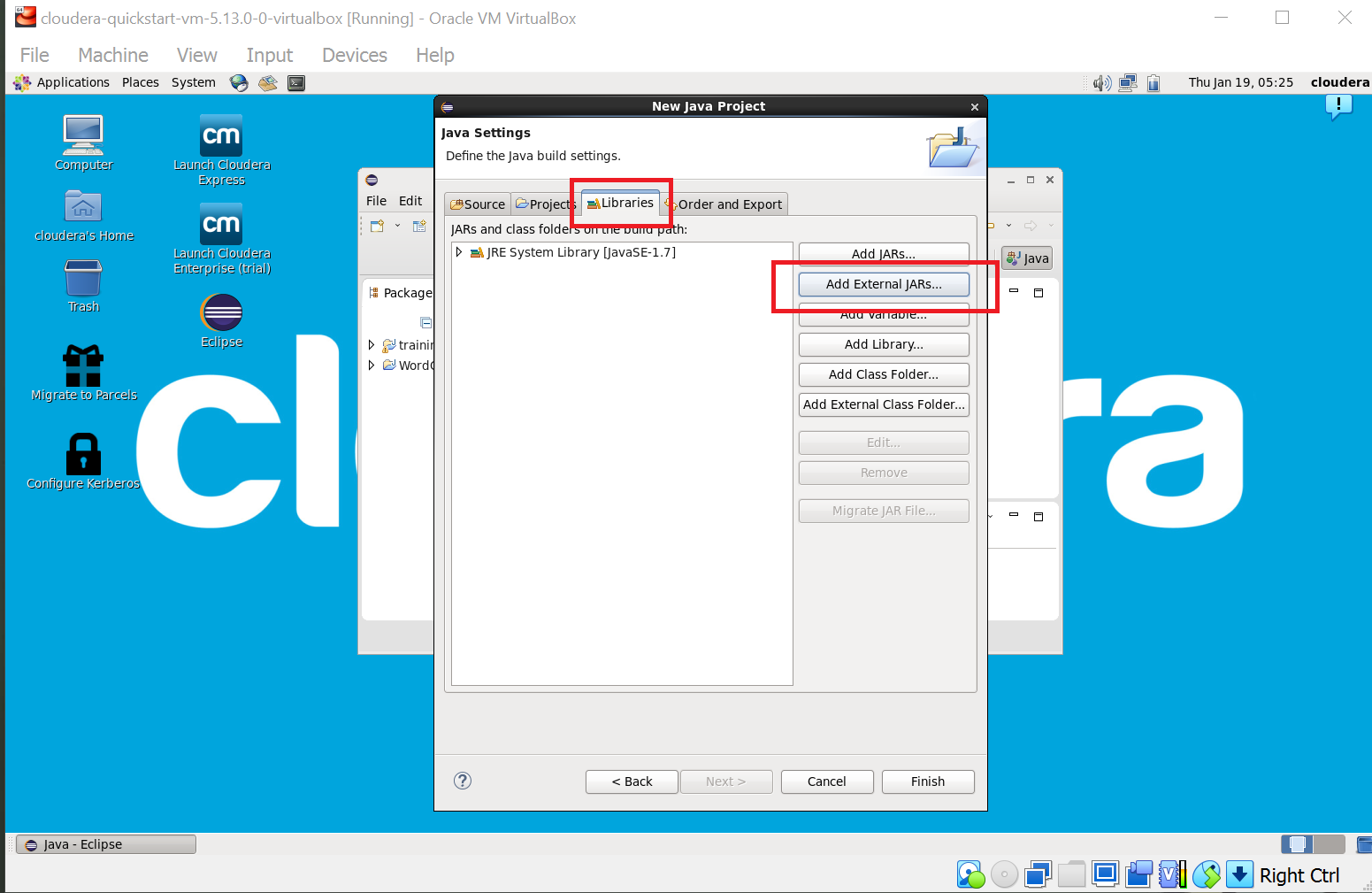
Aim : To Implement WordCount problem using Hadoop MapReduce in Eclips.

Step 1 : Run your cloudera system. Open Eclips.



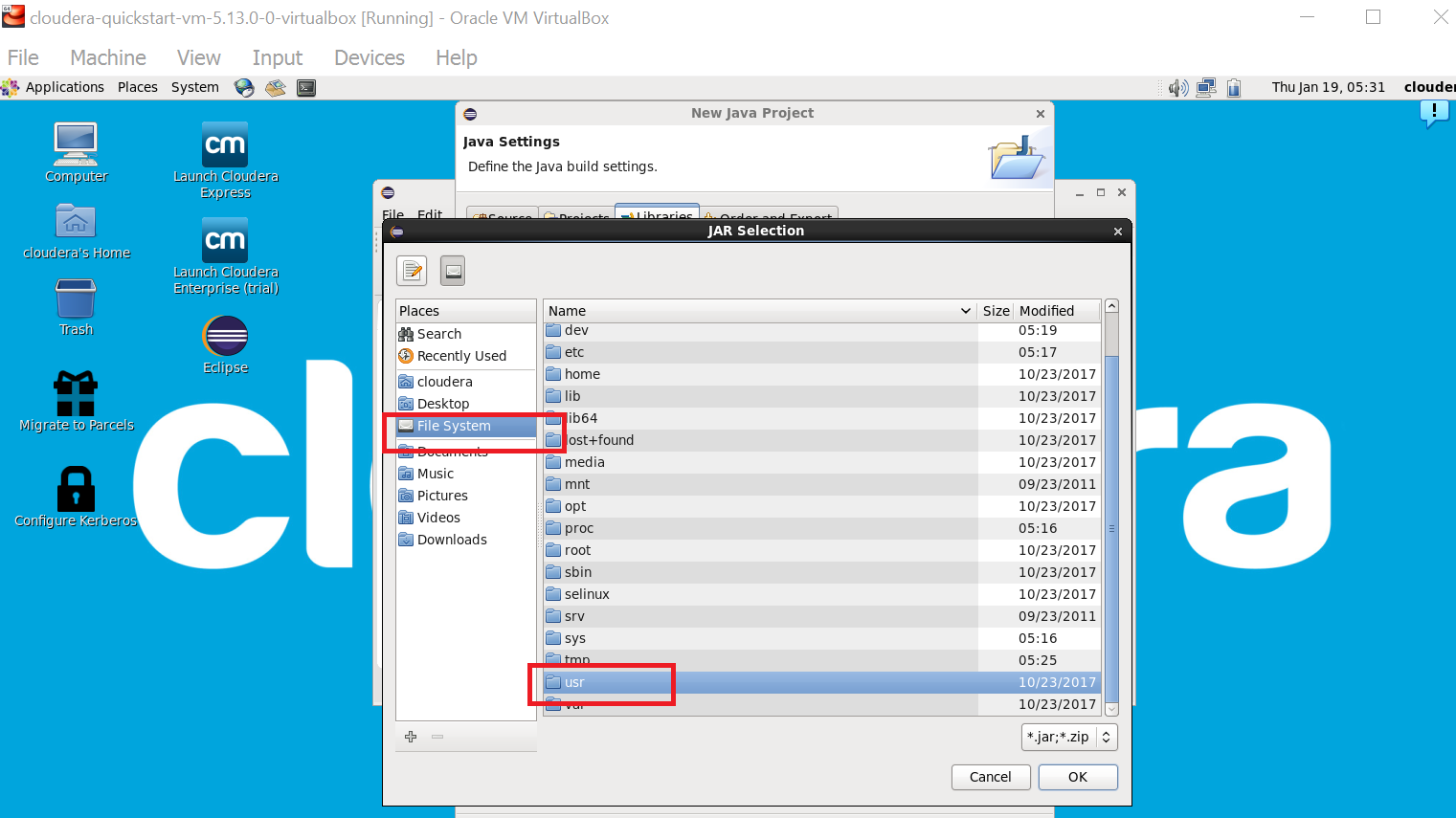
Step 2 : Click on File > New > java project. Give Project Name (“WordCount”). 

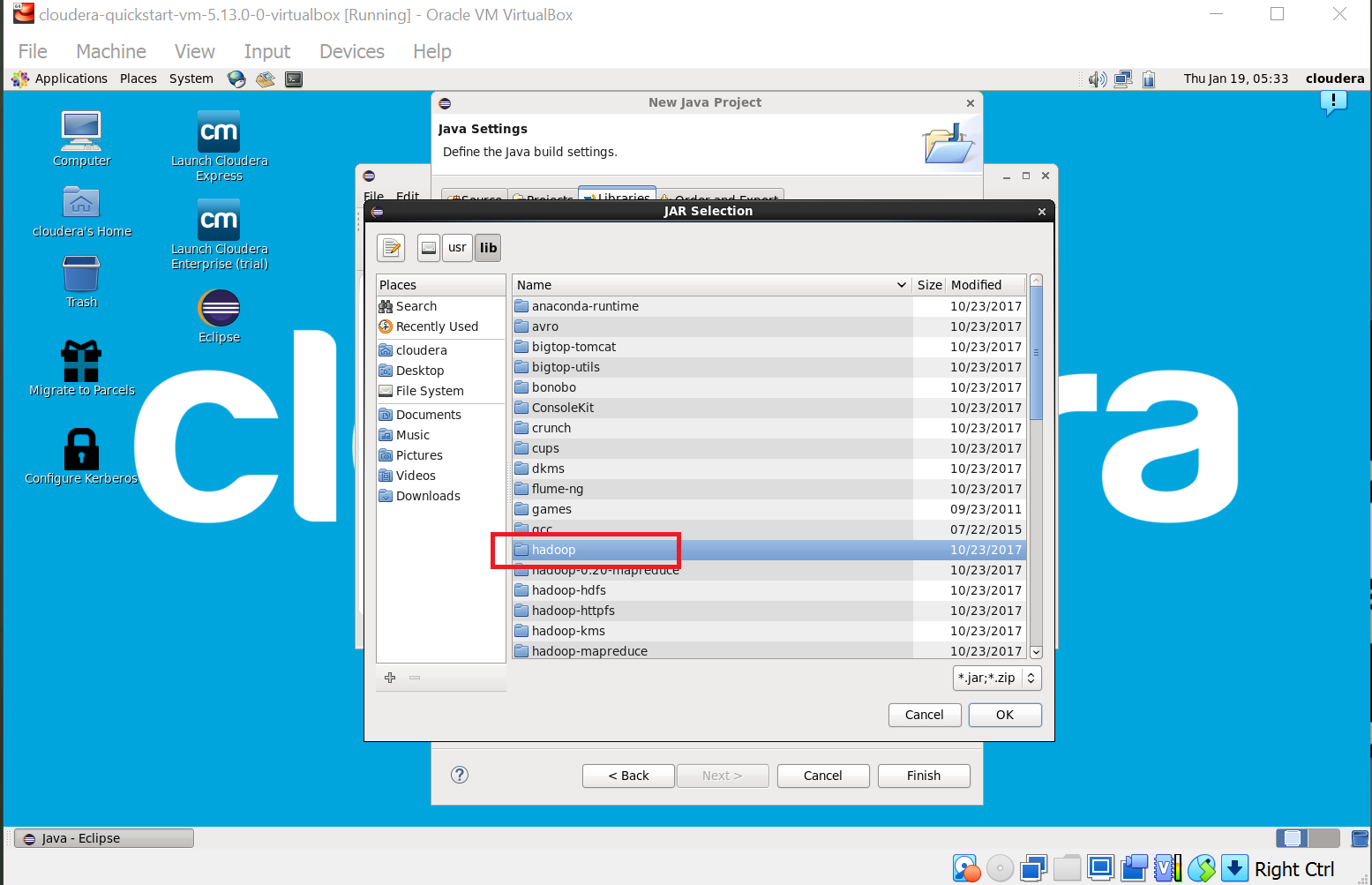
Step 3 : Click on Libraries tab. Then click on Add External JARs… Tab To add Hadoop Libraries.



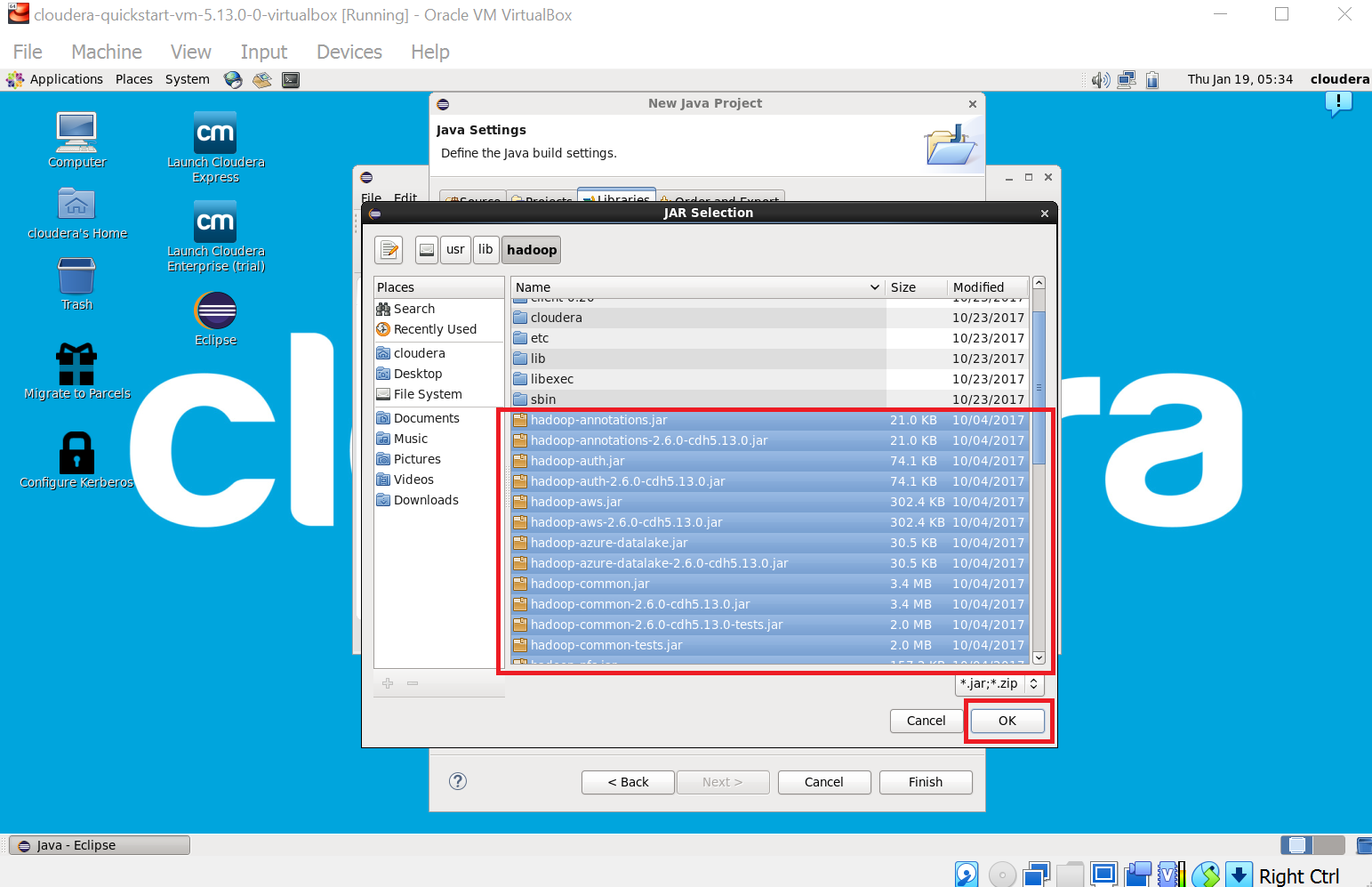
Step 4 : Follow this steps : Click on File System -> usr -> lib -> hadoop (Select all the libraries (JAR files) -> Click OK.

Again Click on Add External JARs… -> client -> select all jar files -> ok -> Finish.

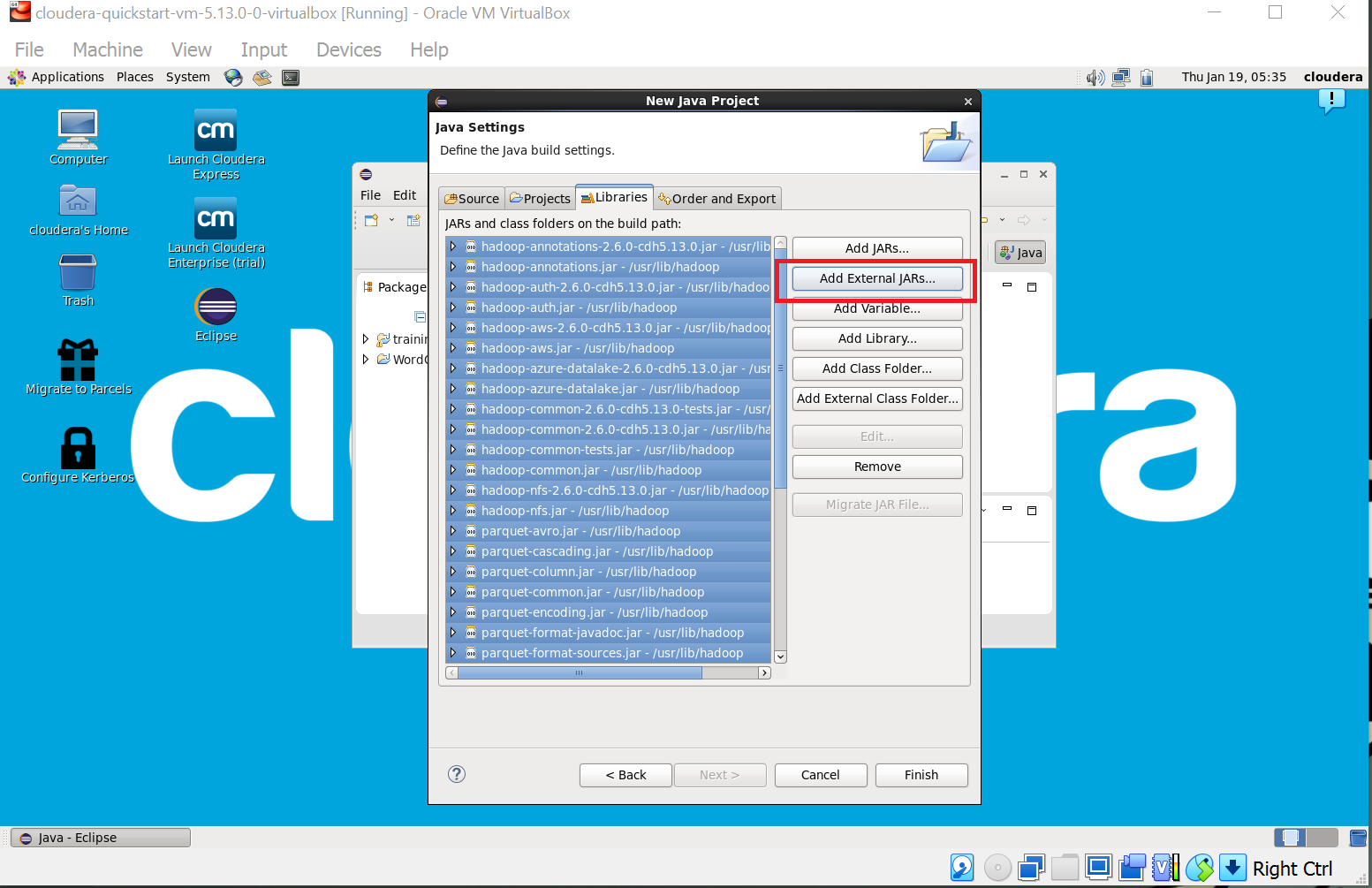


Step : 5

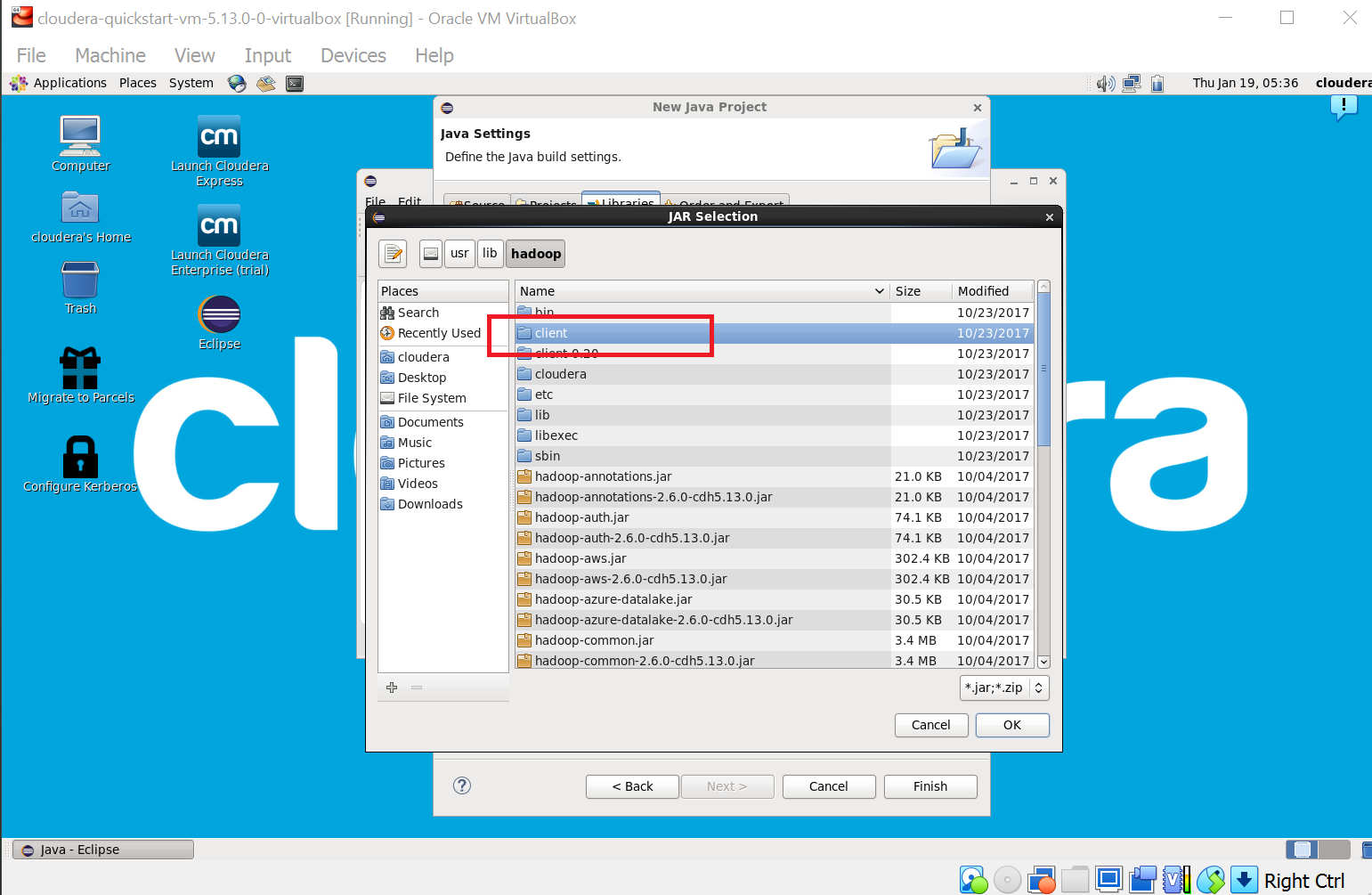
Step : 6



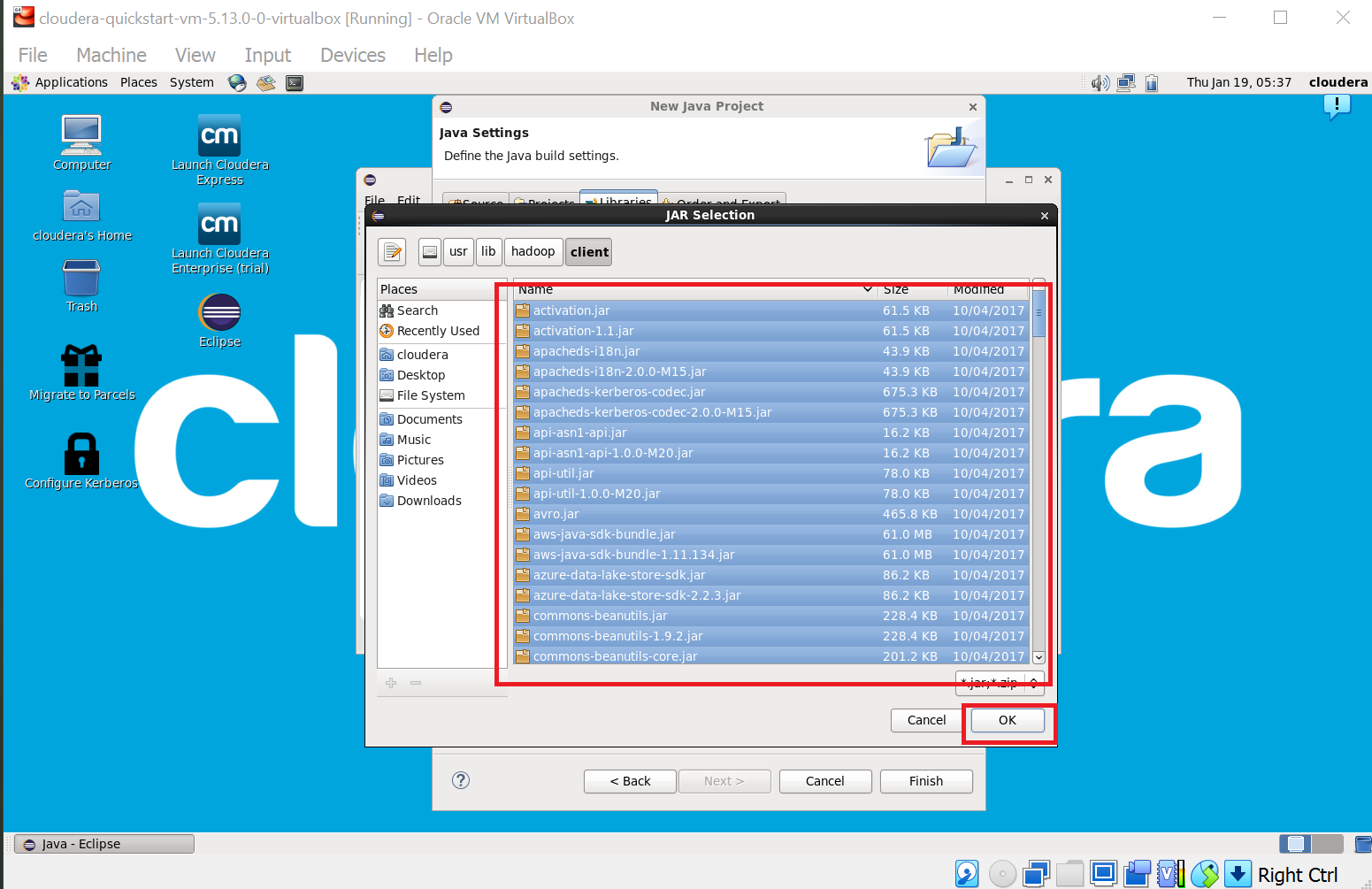
Step : 7



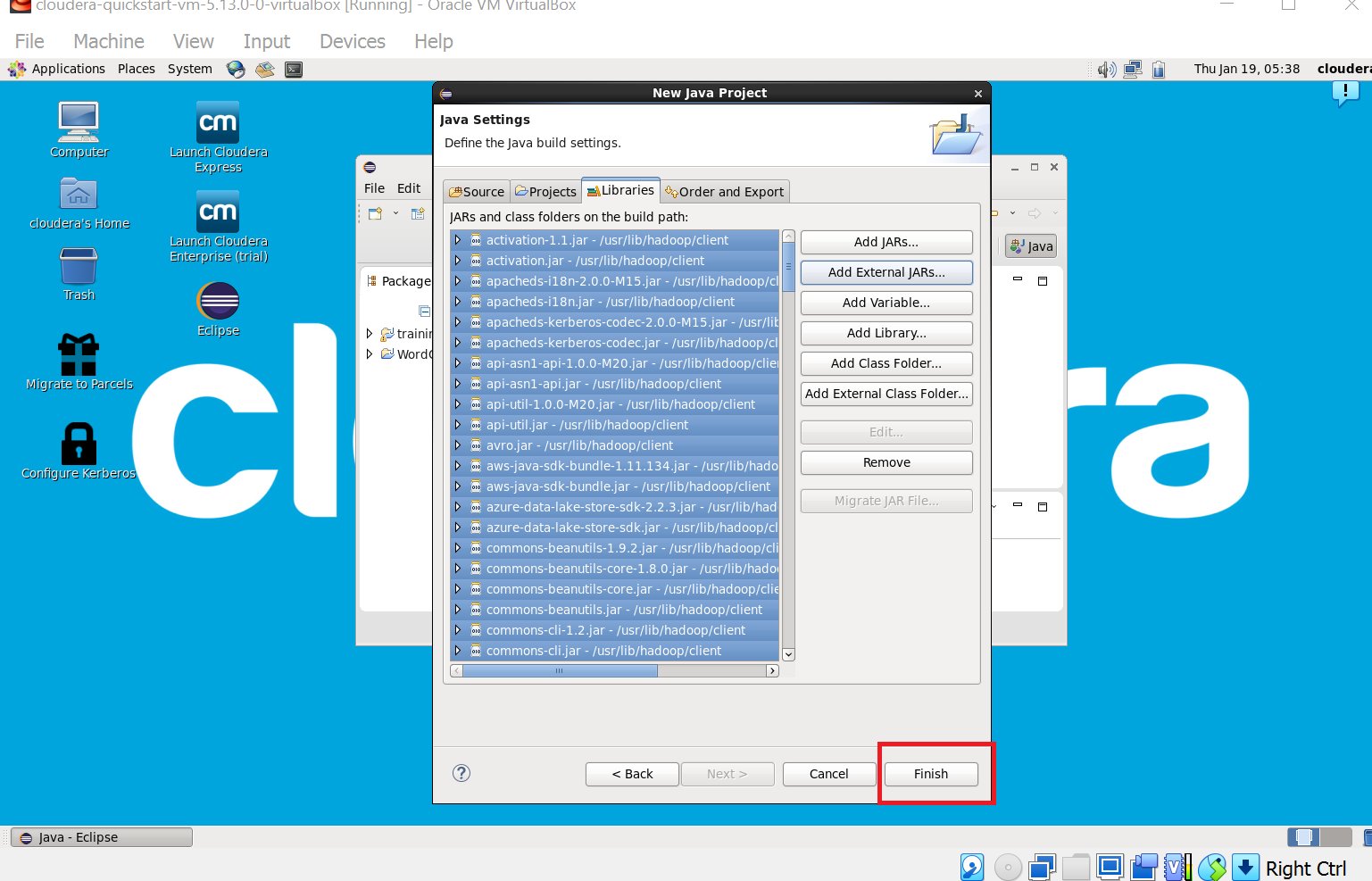
Step : 8



Step : 9



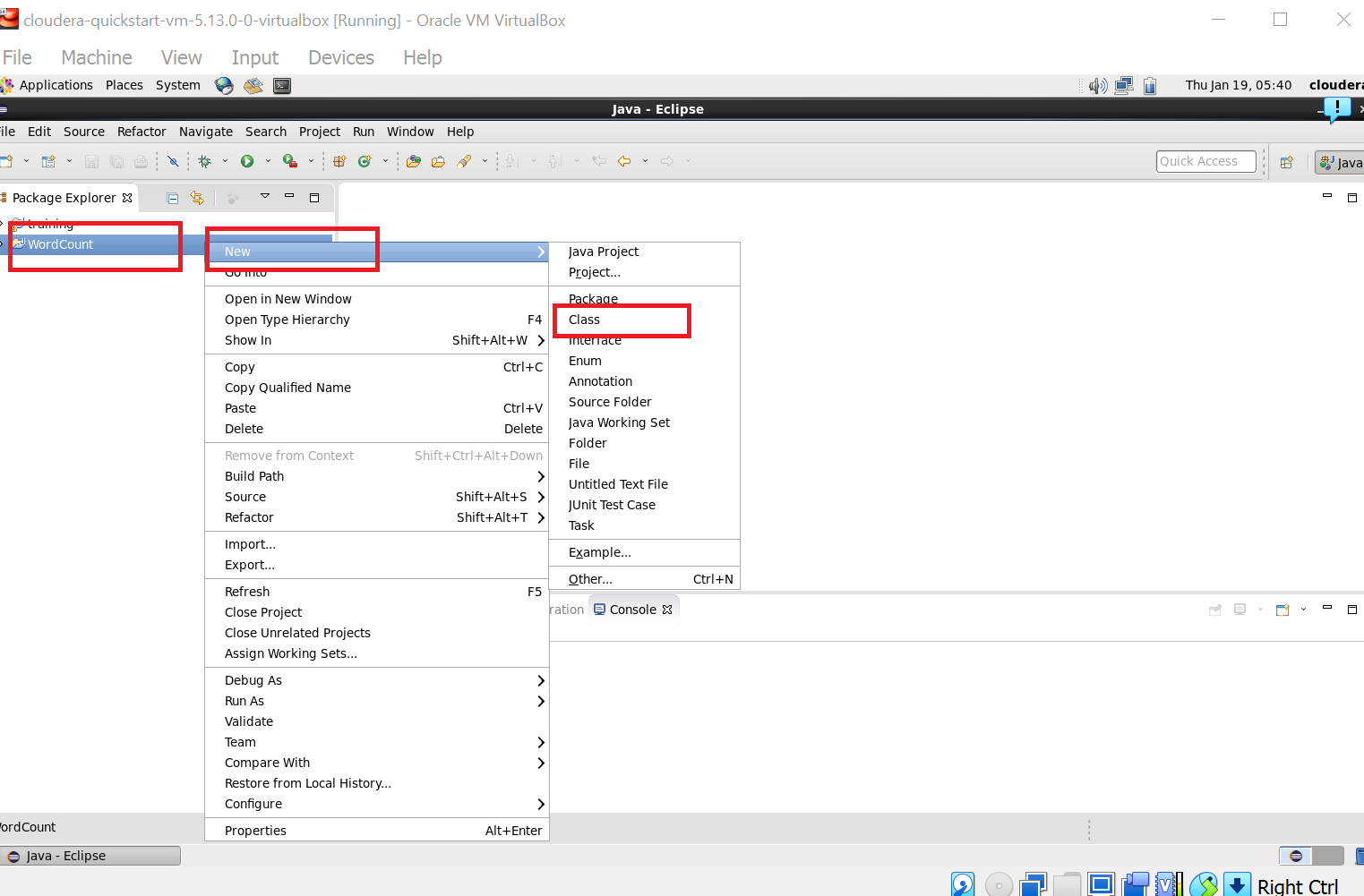
Step : 10



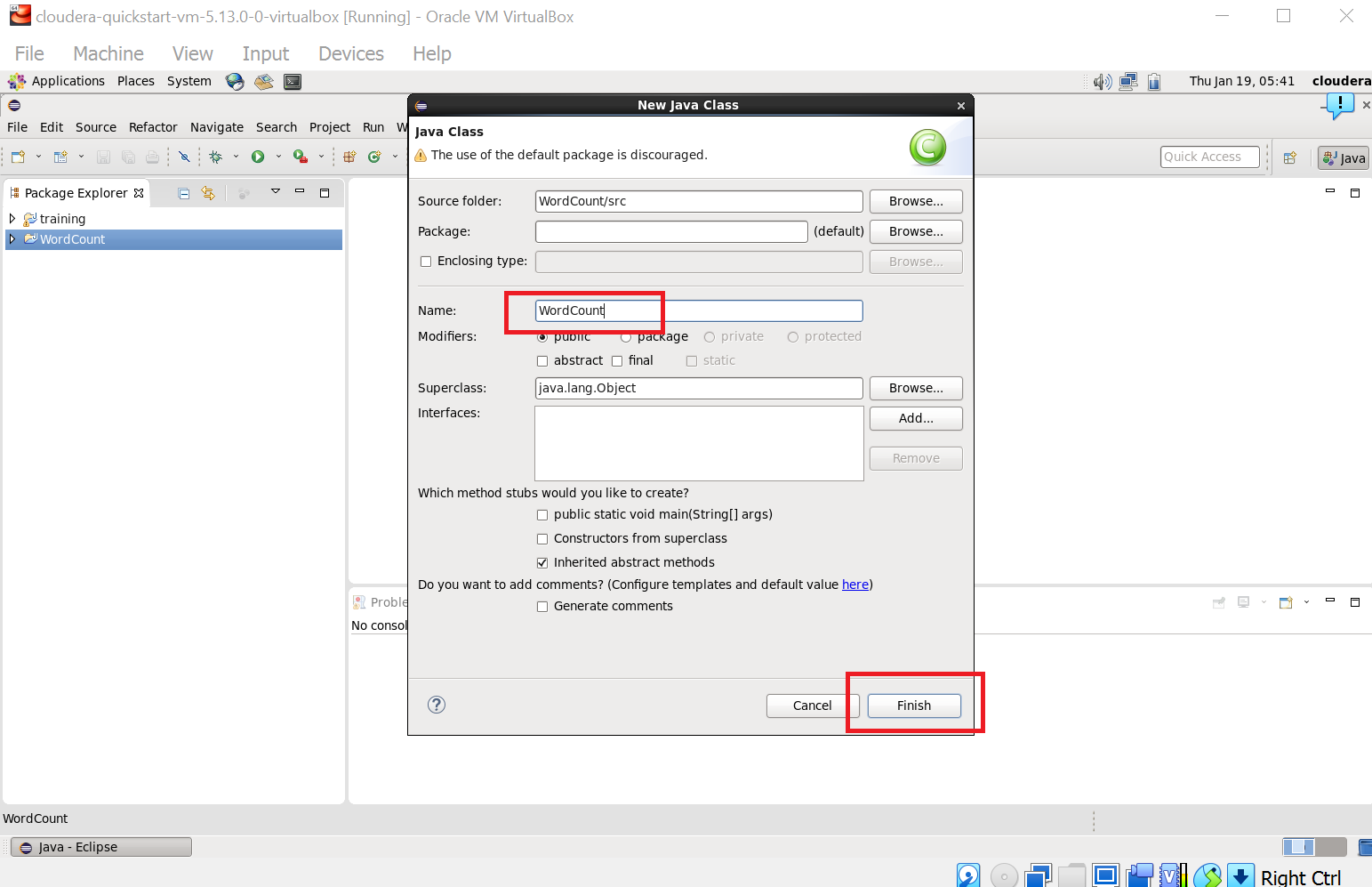
Step 11 : Now you will see project name “WordCount” on sidebar. Right Click on Project name “WordCount” -> New -> class.

For package write name “WordCloud” -> Finish.

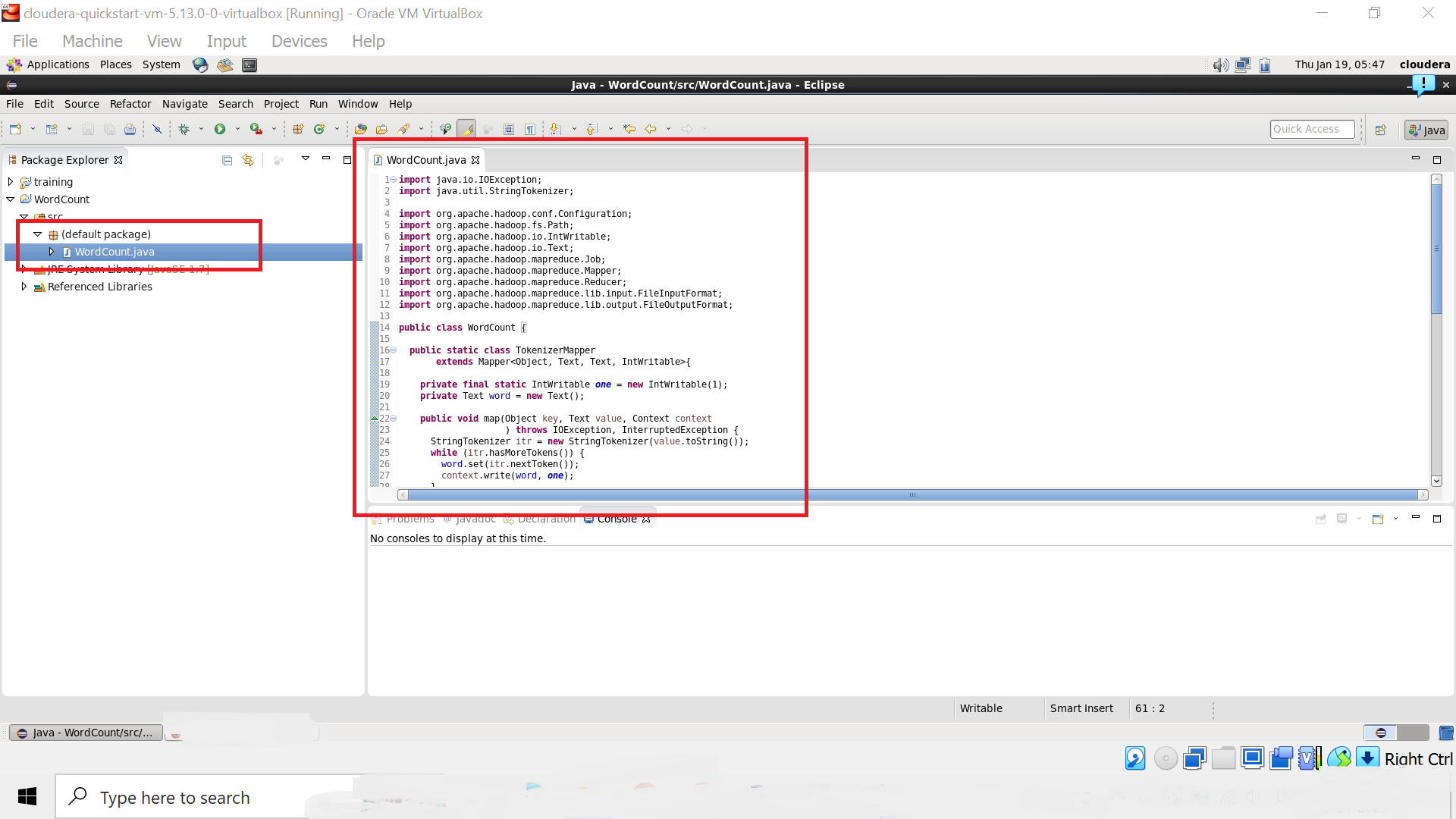
WordCount.java window will pop up.



Step : 12



Step 13 : Write your code in window(WordCount.java).



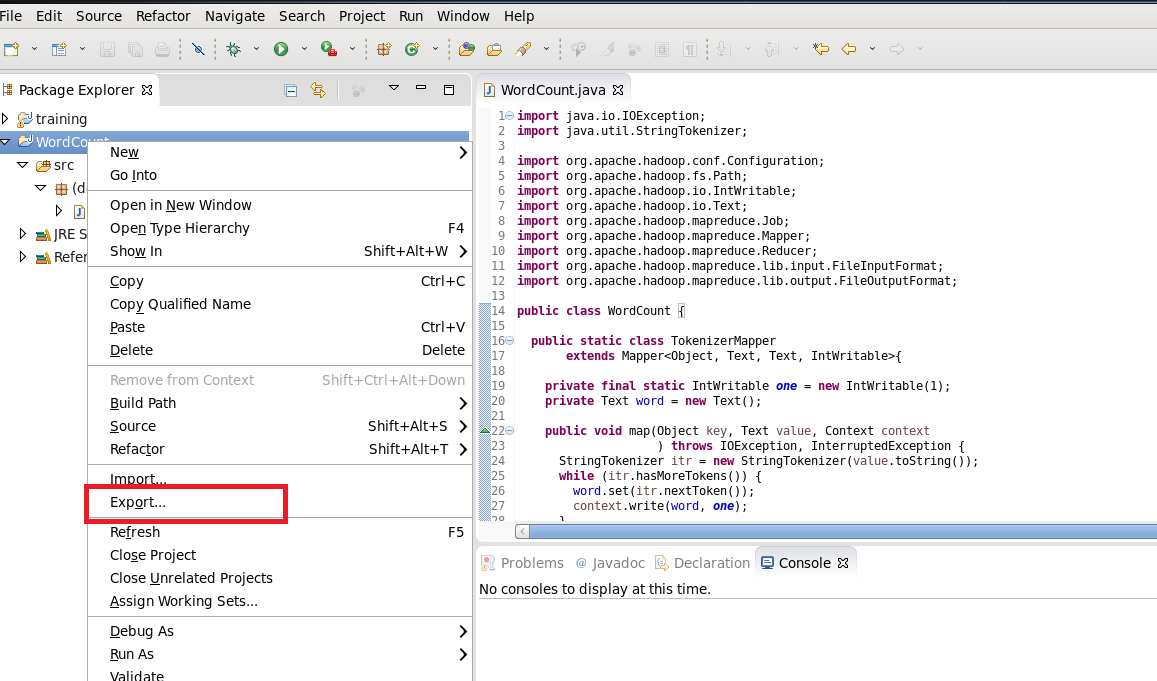
Step 14 : Right Click on the project name

WordCount -> Export -> Java -> JAR File -> Next ->

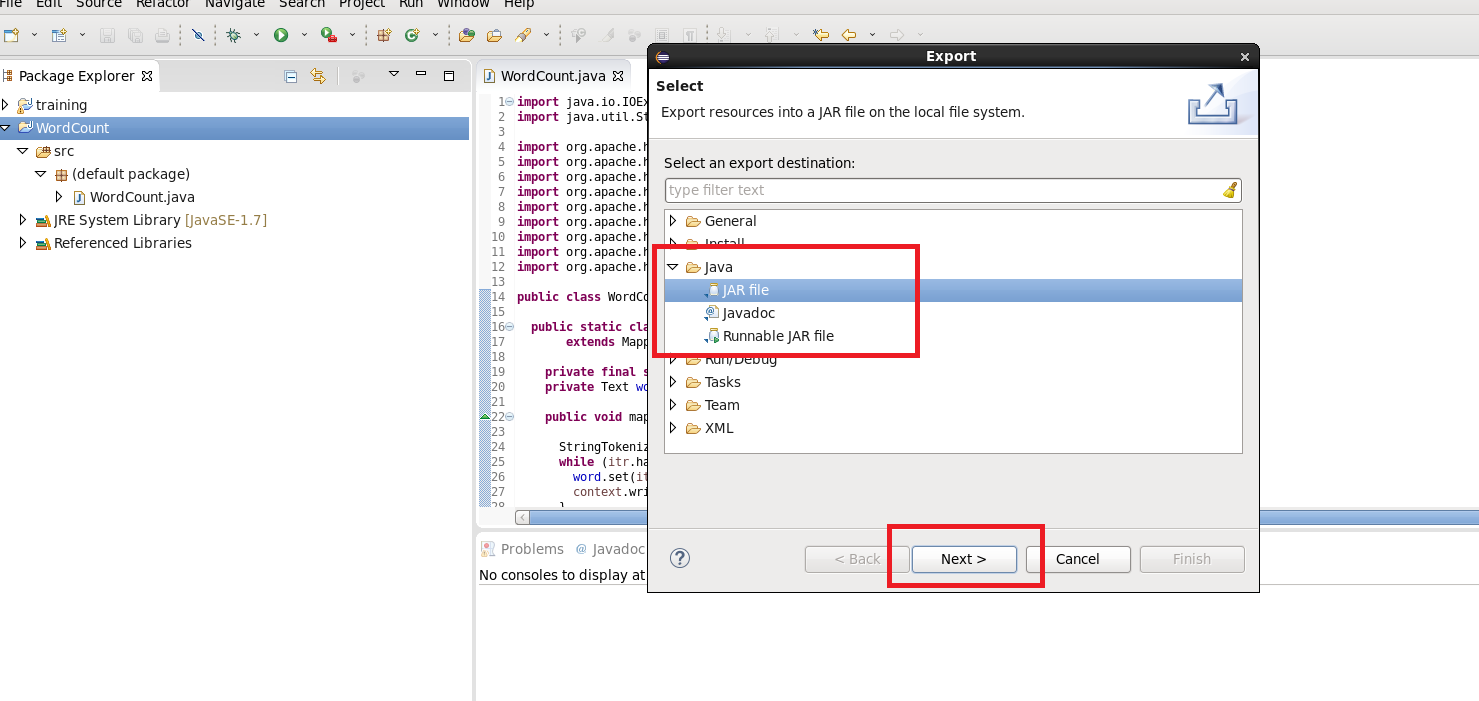
For select the export destination for JAR file:

browse -> Name : WordCount.jar -> save in folder -> cloudera -> Finish -> OK

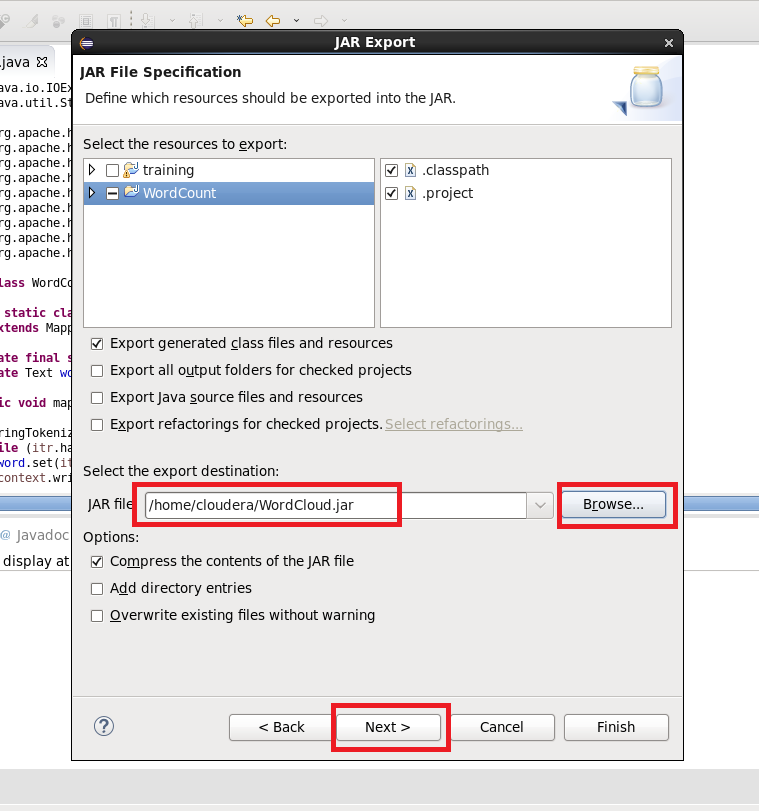
(\*\* I used “WordCloud.jar” here)



Step : 15



Step : 16



Step 17 : Now open terminal . follow this commands.

[cloudera@quickstart ~]$ ls

cloudera-manager cm\_api.py Desktop Documents Downloads eclipse

enterprise-deployment.json express-deployment.json kerberos lib

Music parcels Pictures Public Templates Videos WordCloud.jar

Workspace

[cloudera@quickstart ~]$ pwd

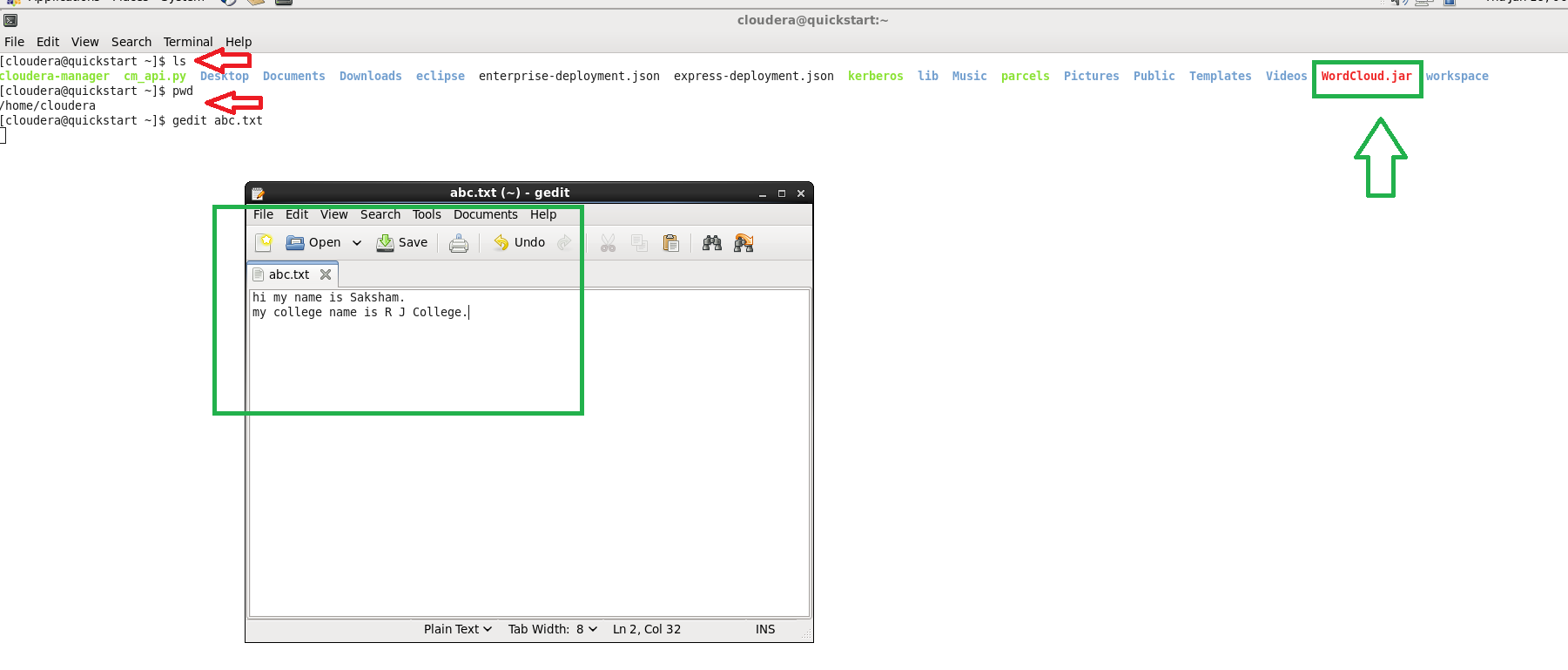
/home/cloudera

[cloudera@quickstart ~]$ gedit abc.txt

[cloudera@quickstart ~]$ cat abc.txt

hi my name is Saksham.

my college name is R J College.



Step 18 :

[cloudera@quickstart ~]$ hdfs dfs -ls /

Found 6 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - hbase supergroup 0 2023-01-19 05:20 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-01-19 05:20 /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 2023-01-19 05:20 /hbase

drwxr-xr-x - cloudera supergroup 0 2023-01-19 06:04 /inputdir

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-01-19 05:20 /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/abc.txt /inputdir/

[cloudera@quickstart ~]$ hdfs dfs -ls /inputdir

Found 1 items

-rw-r--r-- 1 cloudera supergroup 55 2023-01-19 06:05 /inputdir/abc.txt

[cloudera@quickstart ~]$ hdfs dfs -cat /inputdir/abc.txt

hi my name is Saksham.

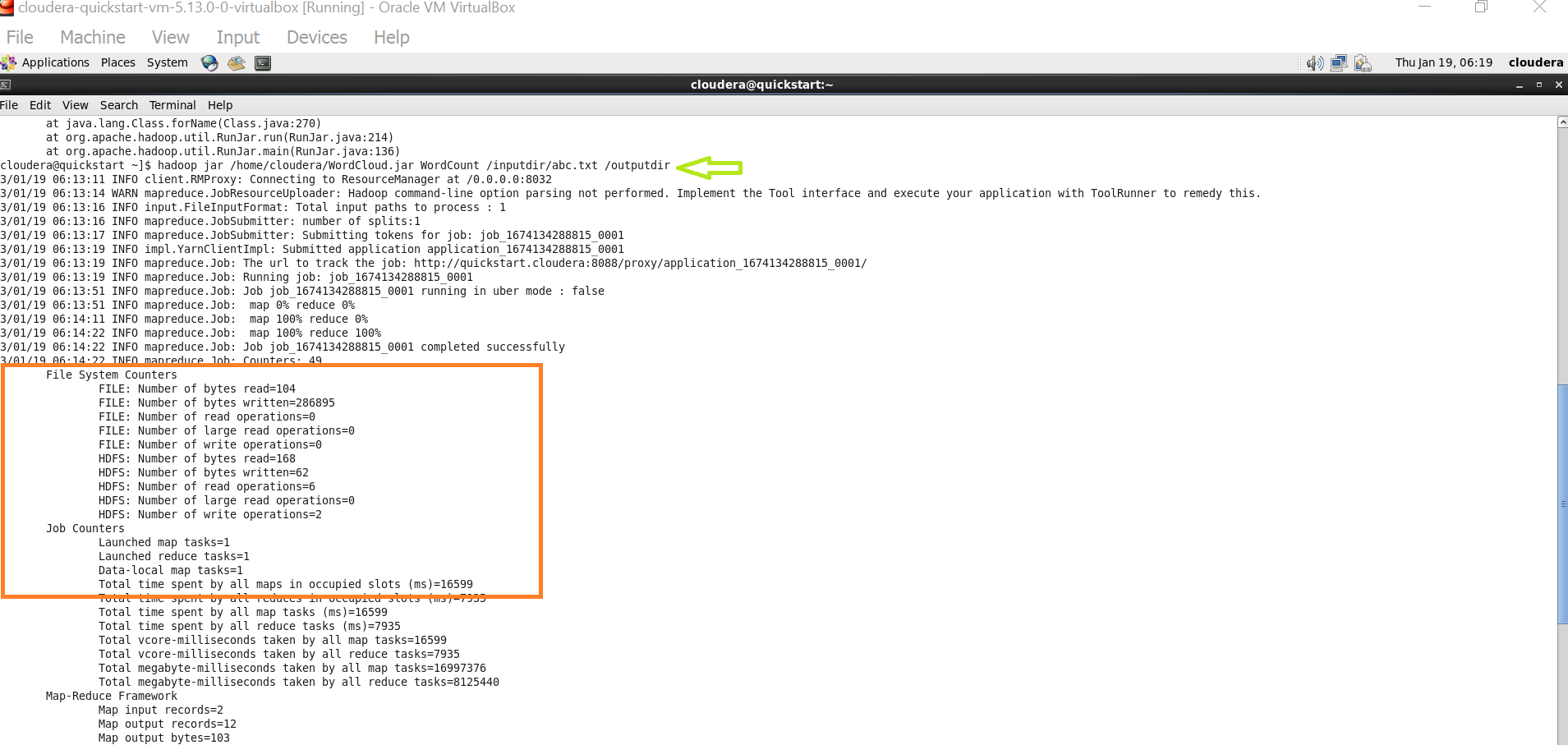
my college name is R J College.



Step 19 :

[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCloud.jar WordCount /inputdir/abc.txt /outputdir

It will Start executing the program.



Step 20 :

[cloudera@quickstart ~]$ hdfs dfs -ls /

Found 8 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - hbase supergroup 0 2023-01-19 05:20 /hbase

drwxr-xr-x - cloudera supergroup 0 2023-01-19 06:05 /inputdir

drwxr-xr-x - cloudera supergroup 0 2023-01-19 06:14 /outputdir

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-01-19 05:20 /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 2023-01-19 06:14 /outputdir/\_SUCCESS

-rw-r--r-- 1 cloudera supergroup 62 2023-01-19 06:14 /outputdir/part-r-00000

**Final Output :**

[cloudera@quickstart ~]$ hdfs dfs -cat /outputdir/part-r-00000

College. 1

J 1

R 1

Saksham. 1

college 1

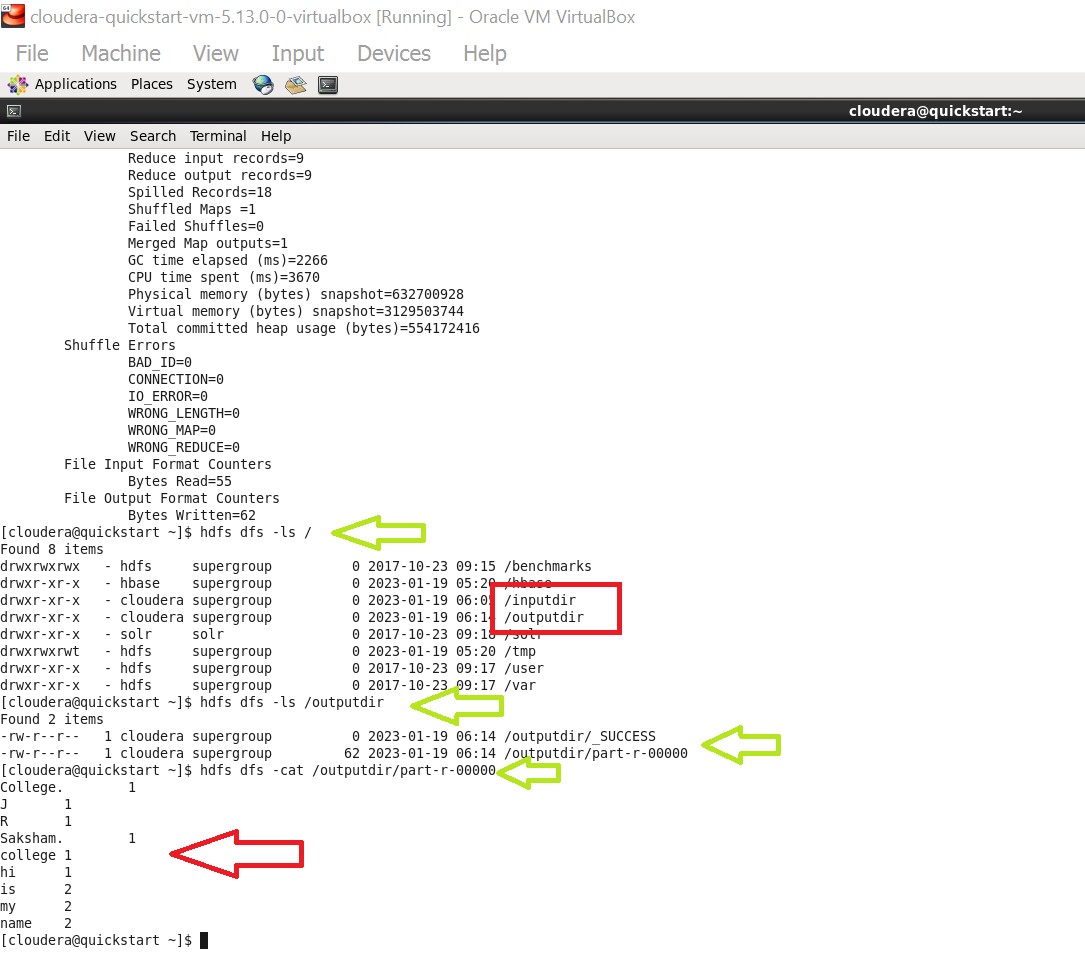
hi 1

is 2

my 2

name 2

[cloudera@quickstart ~]$



*Thank you.*