Steps of building an environment of Hadoop:

1. download VM ware workstation and install

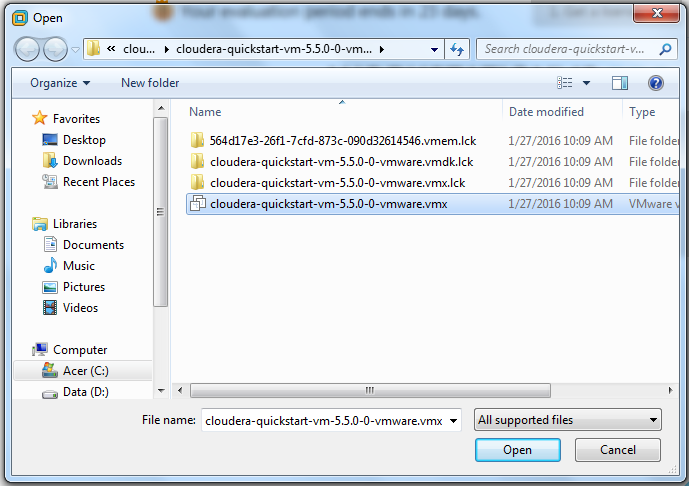
http://www.vmware.com/products/workstation/workstation-evaluation

2. download Cloudera QuickStarts

http://www.cloudera.com/downloads/quickstart\_vms/5-5.html#

3. Unzip the compressed file and open the file .vmx inside the folder. It has already make a operation system with eclipse , hadoop libs & other tools.

File -> Open , and then it will show following window:



4. Create a java file WordCount.java in the training project

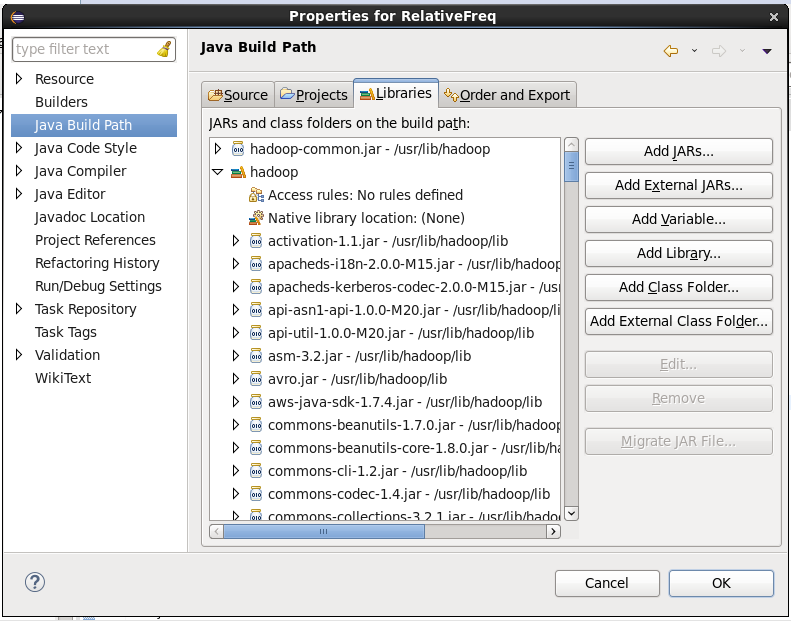
5. Coding and debugging

6. Compile the project and find there is no some lib file of hadoop, so include those in the properties for training project in Java Build Path:

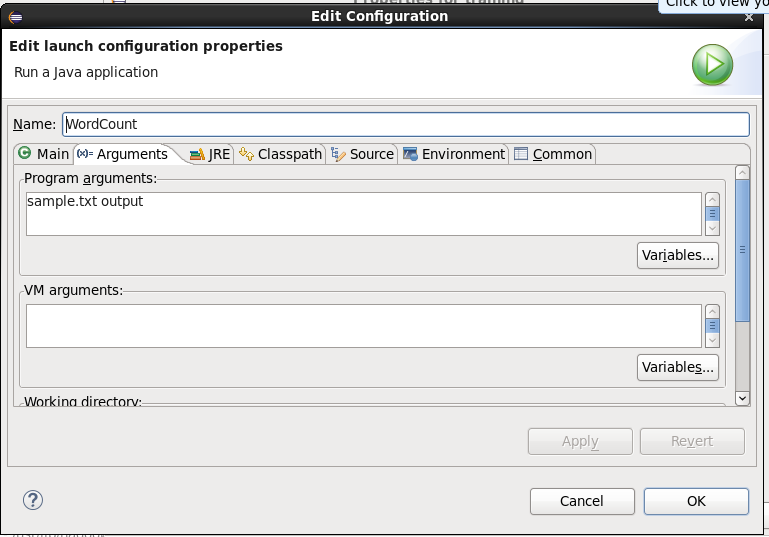
/usr/lib/hadoop/lib (i add all jar in this folder except sub folder native)

/usr/lib/hadoop (i add all jar in this folder except sub folder)

/usr/lib/hadoop/client (i add all jar in this folder except sub folder)



7. add argument in eclipse for debugging locally.



After run well locally, then try the jar, following step of running a jar in hadoop:

1. hdfs dfs -mkdir /user/cloudera/input

2. hdfs dfs -ls /user/cloudera/input

3. echo "this is a good good student , this is a bad bad engineer" > sample.txt

4. hdfs dfs -put /home/cloudera/temp/sample.txt /user/cloudera/input/

5. hadoop jar WordCount.jar /user/cloudera/input/sample.txt /user/cloudera/output

8. hdfs dfs -ls /user/cloudera/output

hdfs dfs -cat /user/cloudera/output/part-r-00000

, 1

a 2

bad 2

engineer 1

good 2

is 2

student 1

this 2