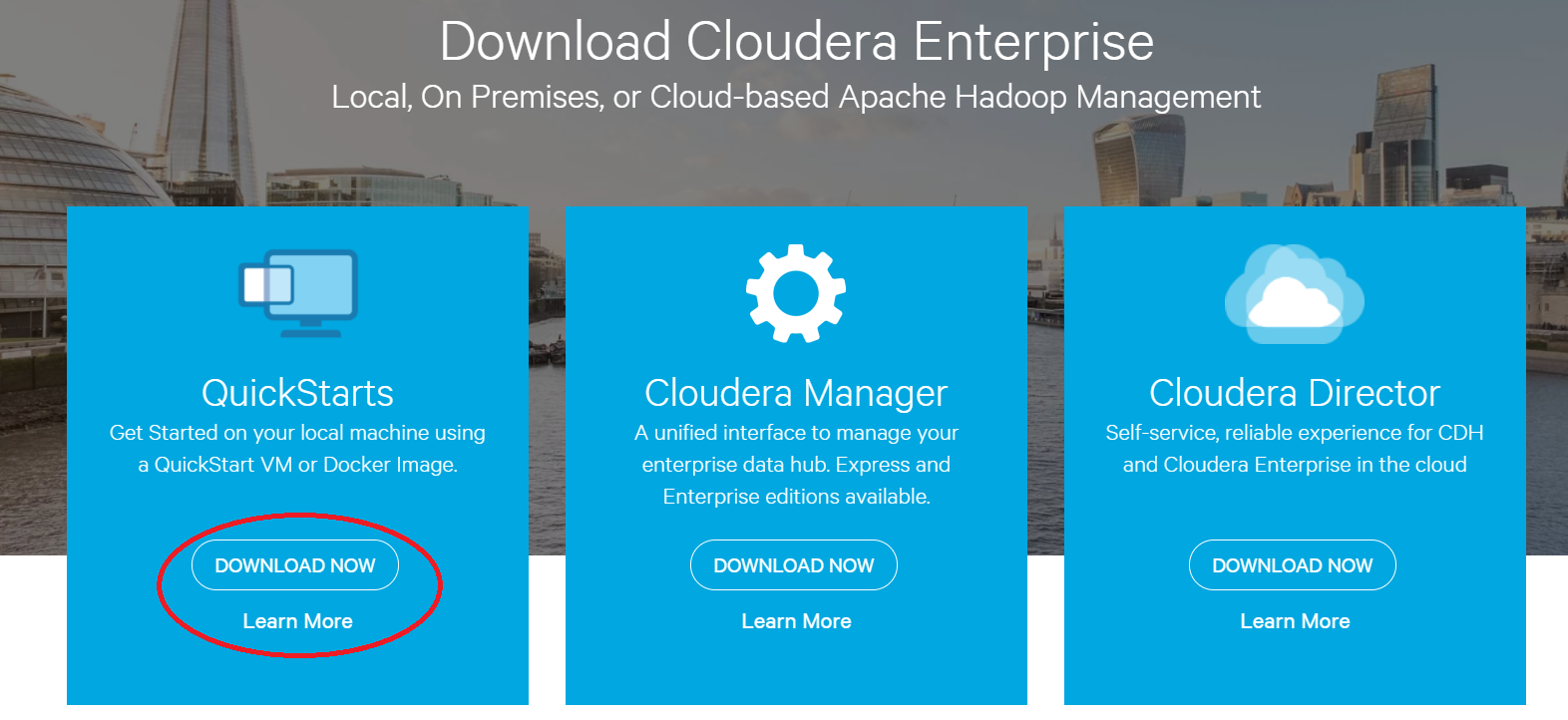
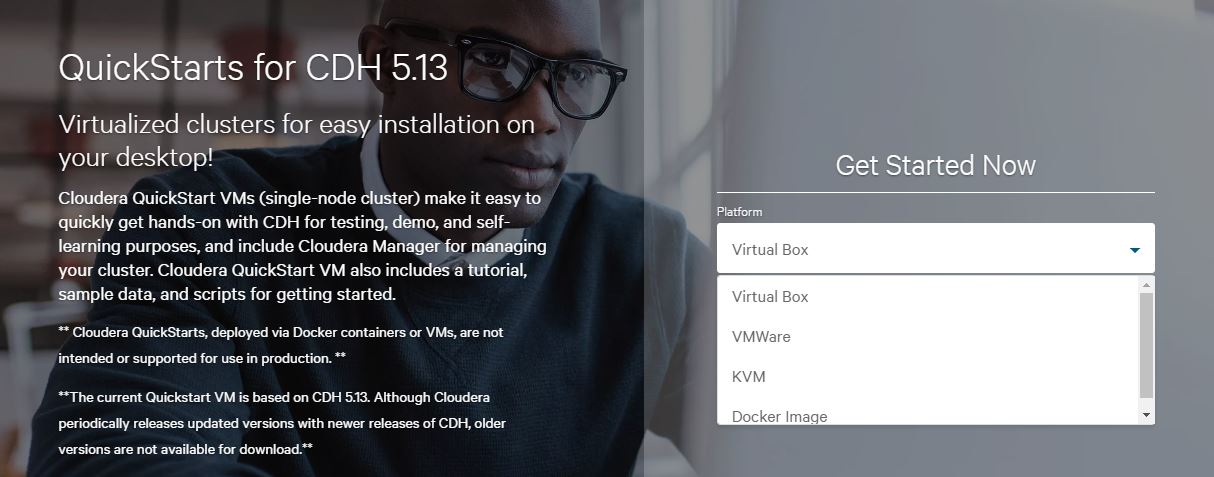
**Part 1.(a, b, c) Documentation Details**

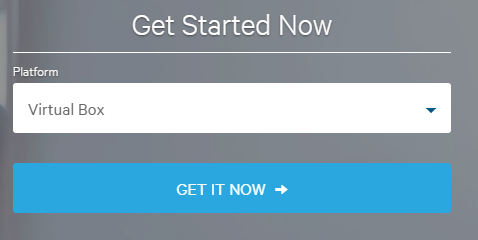
1. Download Cloudera Enterprise QuickStarts for CDH 5.13 from the following link: http://www.cloudera.com/downloads.html



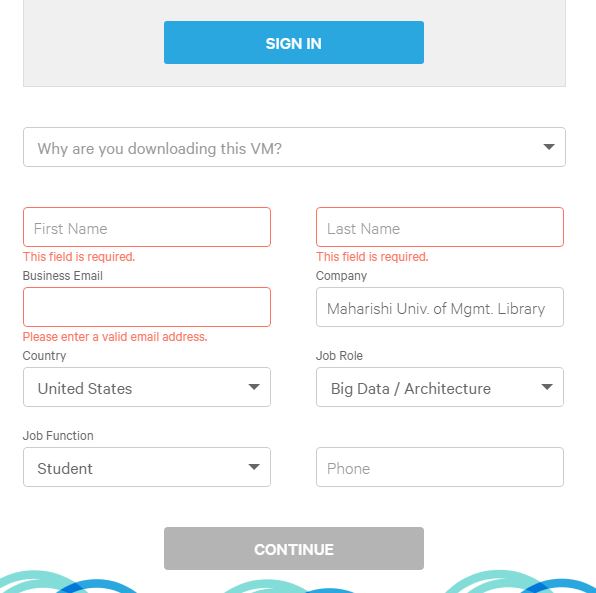
1. Select Virtual Box in “Get stated Now”



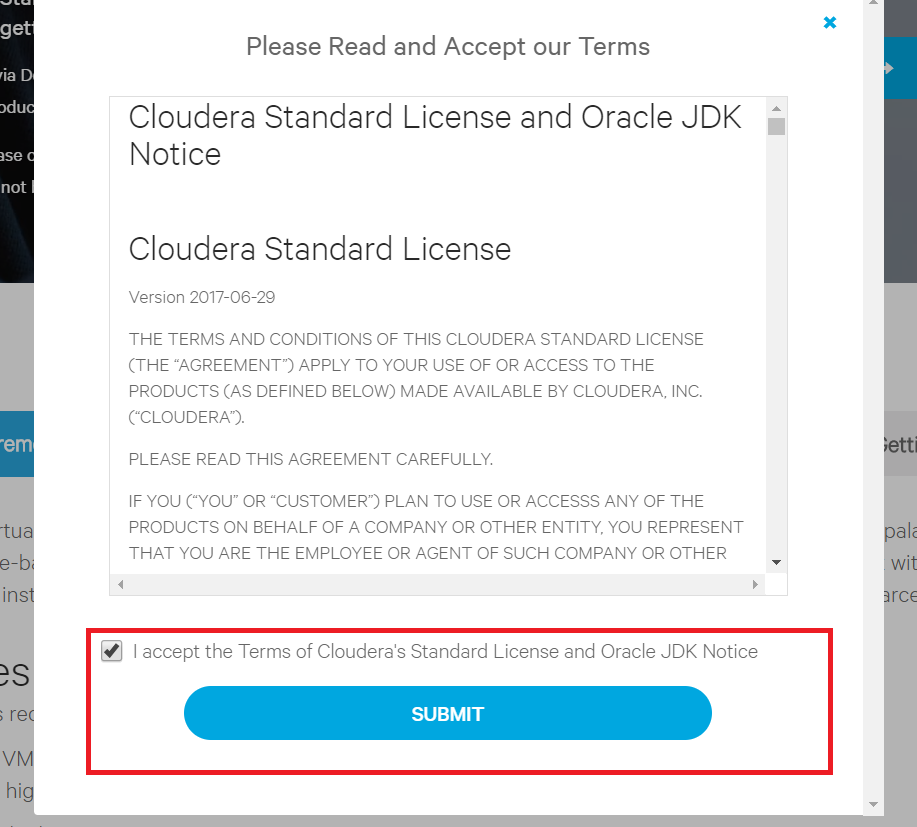
1. Click on “Get It Now”



1. Fill out all the required fields in the form



1. Click on “Continue” button
2. Then, check on accept box and click on submit button.

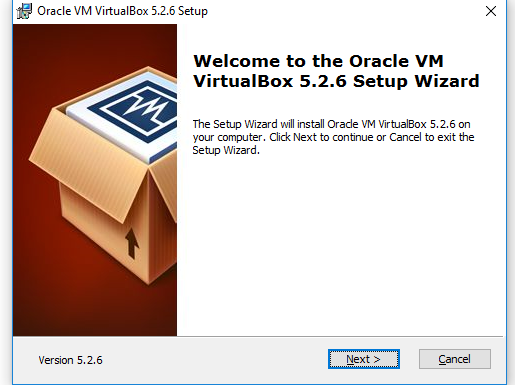


1. Download Virtual Box from the following link:

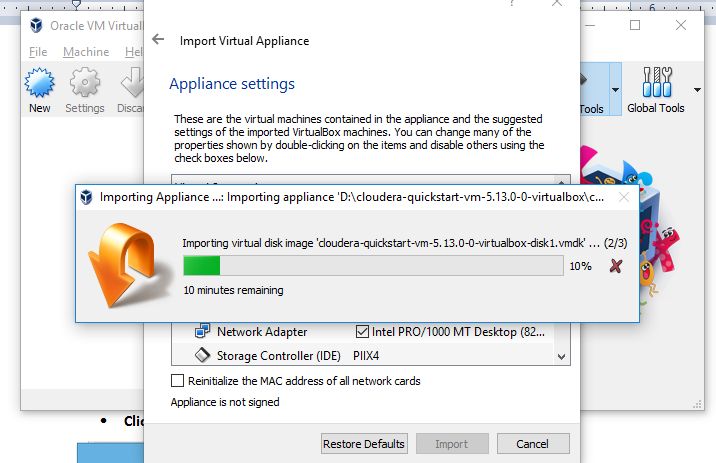
<https://www.virtualbox.org/wiki/Downloads>



1. Install the downloaded Virtual Box Application in your machine



1. Open the installed Virtual Box and open unzipped cloudera-quickstart-vm-5.13.0



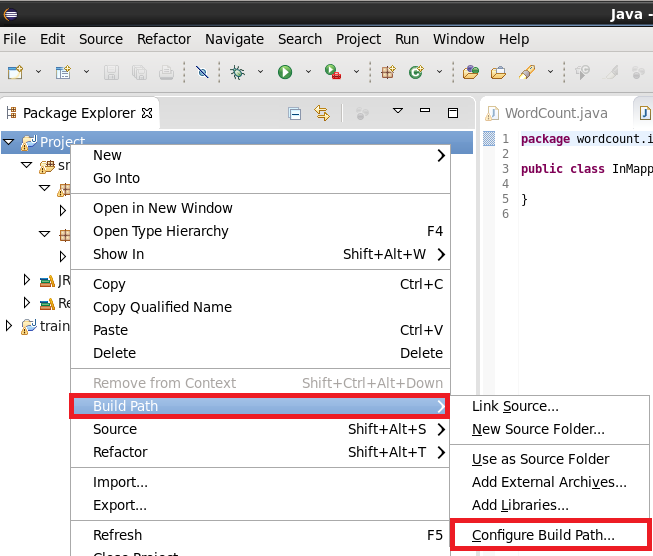
1. Desktop will be look like below:



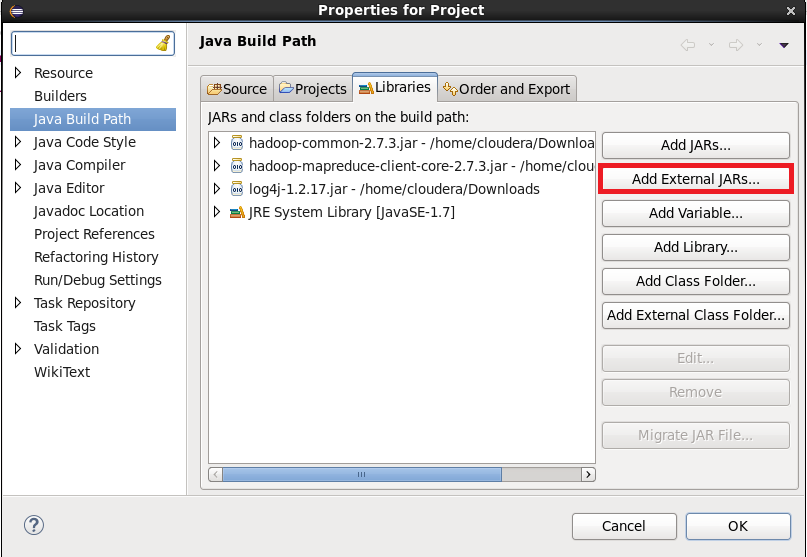
1. Download jar files for hadoop common and hadoop mapreduce client core from below link: <http://mvnrepository.com/artifact/org.apache.hadoop/hadoop-common/2.7.3> and <http://mvnrepository.com/artifact/org.apache.hadoop/hadoop-mapreduce-client-core/2.7.3>
2. Create a java class named, WordCount under a new java project in Eclipse IDE. Get WordCount example from <https://wiki.apache.org/hadoop/WordCount>. Then add one more line

job.setJarByClass(WordCount.class); after the following line: [Job job = new Job(conf, "wordcount");]

1. Add downloaded jar files to reference library . Right click on java project and select “Build Path” -> Configure Build Path



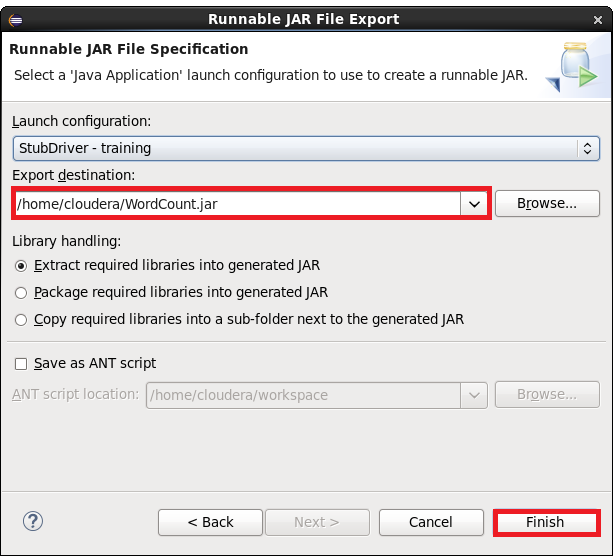
1. Add External Jars..



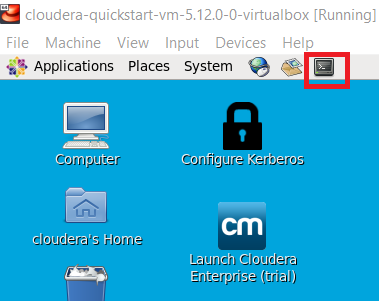
1. Once you build the java class, export to a jar file. Right click on java file and select JAR file under Java folder. In another way, we can also type “jar” in text box to search JAR file.

|  |  |
| --- | --- |
|  |  |

1. Browse the desire path to save the jar file and click on “Finish”.



1. Open a terminal



1. Create input and output locations in HDFS through command line. [Reference: Professor Nair’s Note]. Load input file in input folder in HDFS.

$ sudo su hdfs

$ hadoop fs -mkdir /user/cloudera

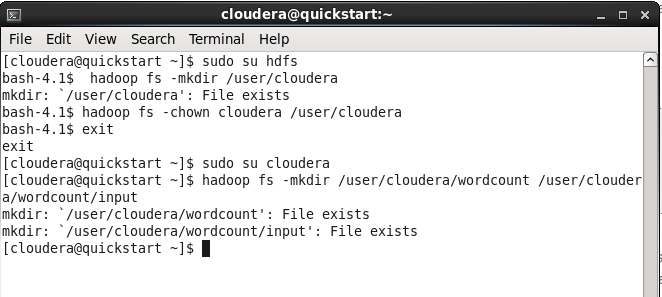
$ hadoop fs -chown cloudera /user/cloudera

$ exit

$ sudo su cloudera

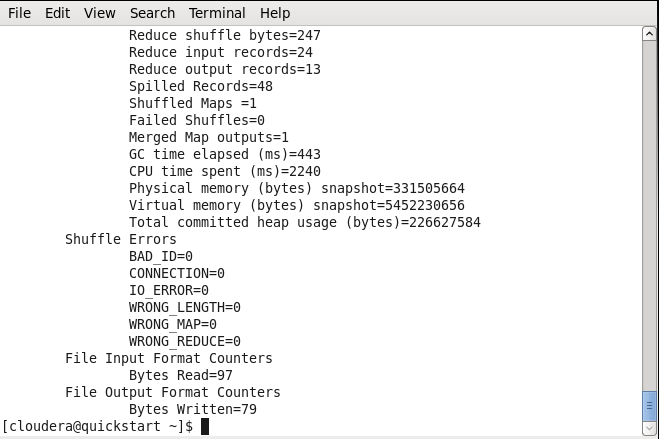
$ hadoop fs -mkdir /user/cloudera/wordcount /user/cloudera/wordcount/input

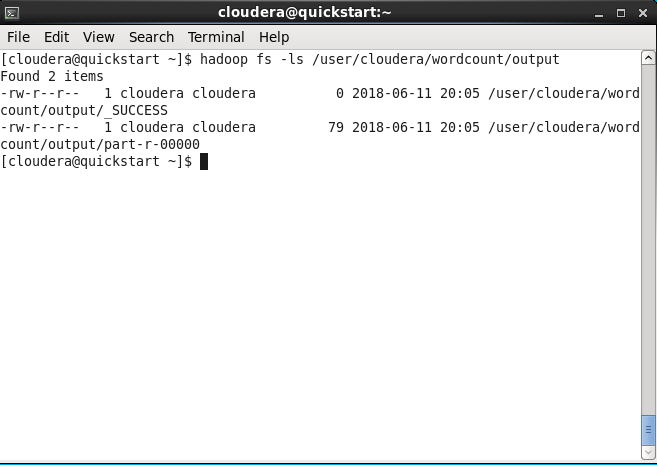
$hadoop fs -put text.txt /user/cloudera/wordcount/input



1. Run the WordCount application in Hadoop pseudo-distributed mode using the following command lines:

$ hadoop jar /home/cloudera/Desktop/WordCount.jar WordCount /user/cloudera/wordcount/input/input.txt /user/cloudera/wordcount/output





1. Output can be viewed using the command: hadoop fs -cat /user/cloudera/wordcount/output/\*

