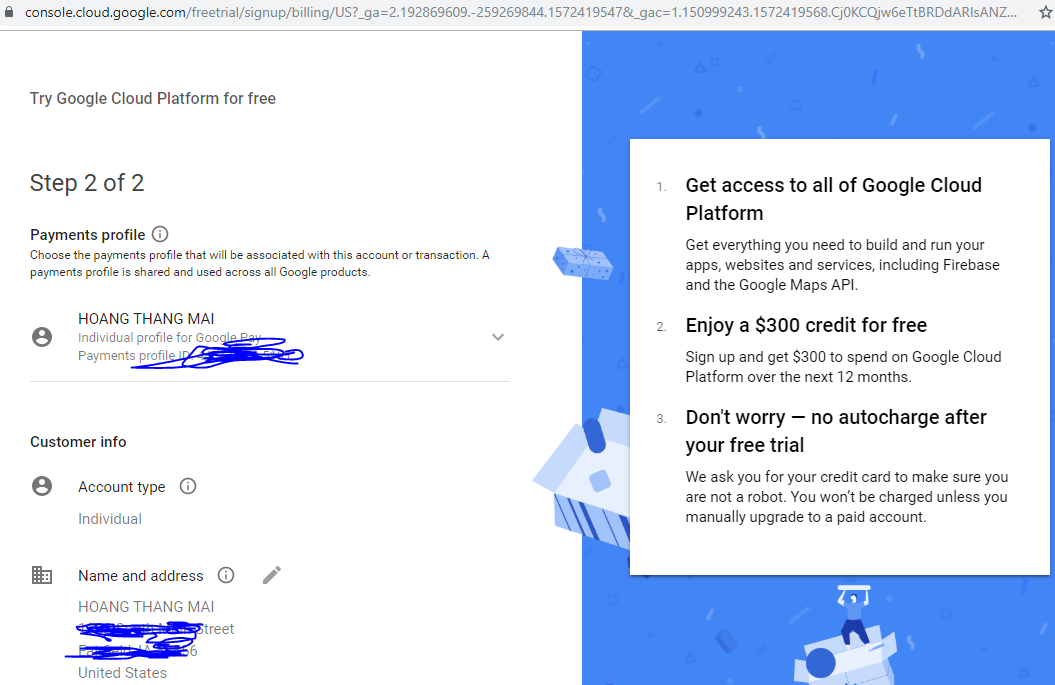
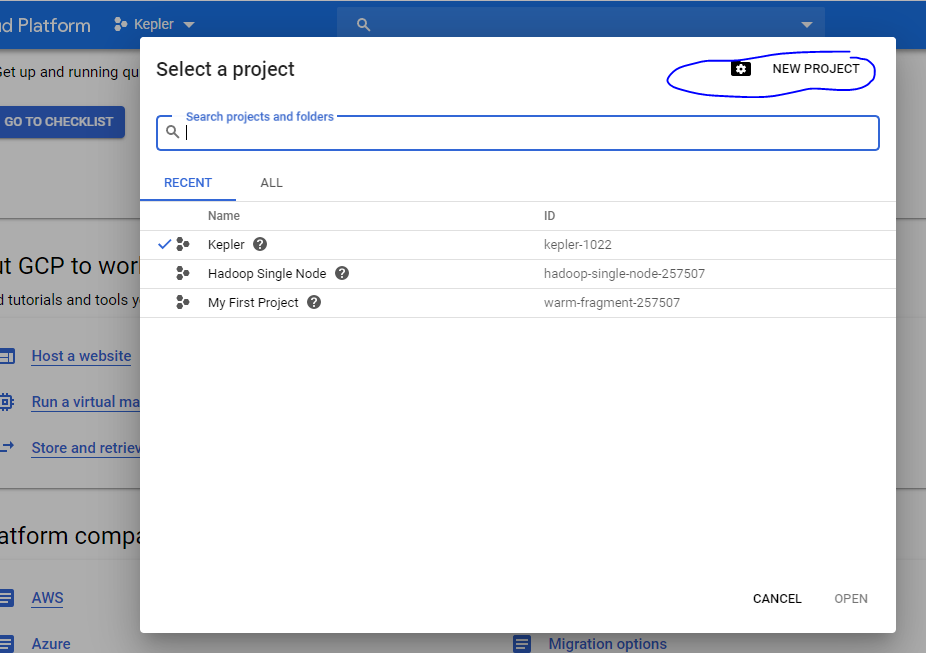
**Set Up a Single Node Cluster using Google Cloud Instance**

# Step 1: Access into https://cloud.google.com to register account

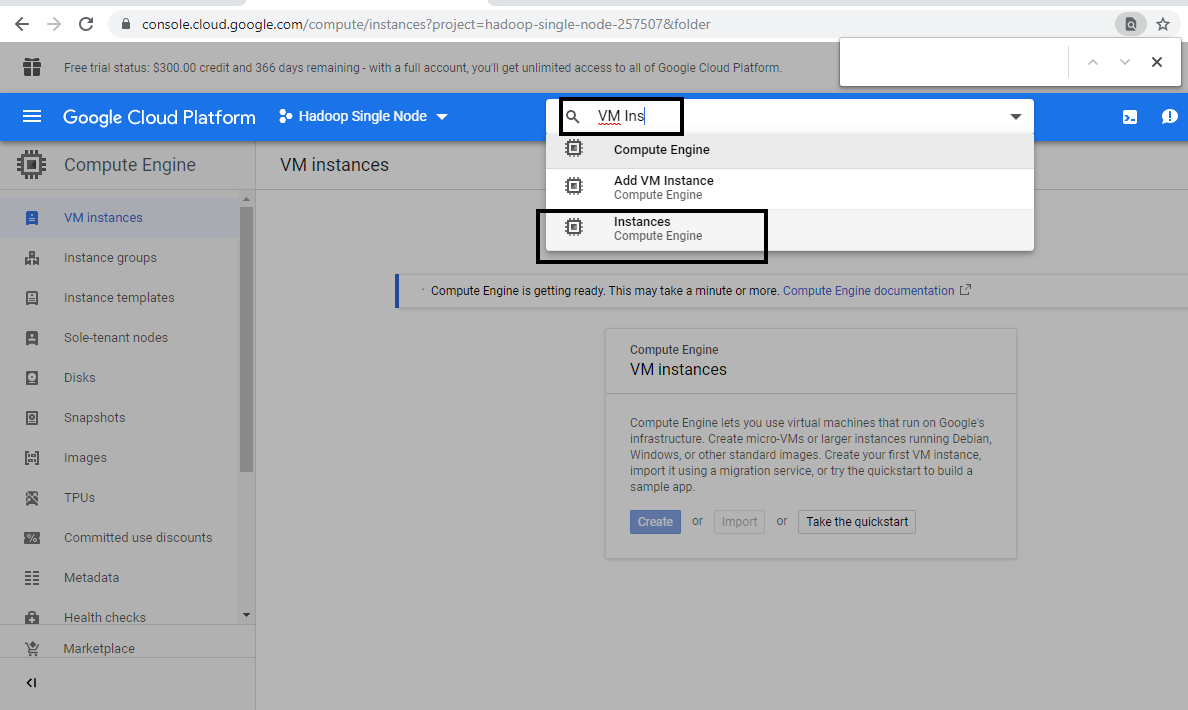
You will get $300 credit for 1 year

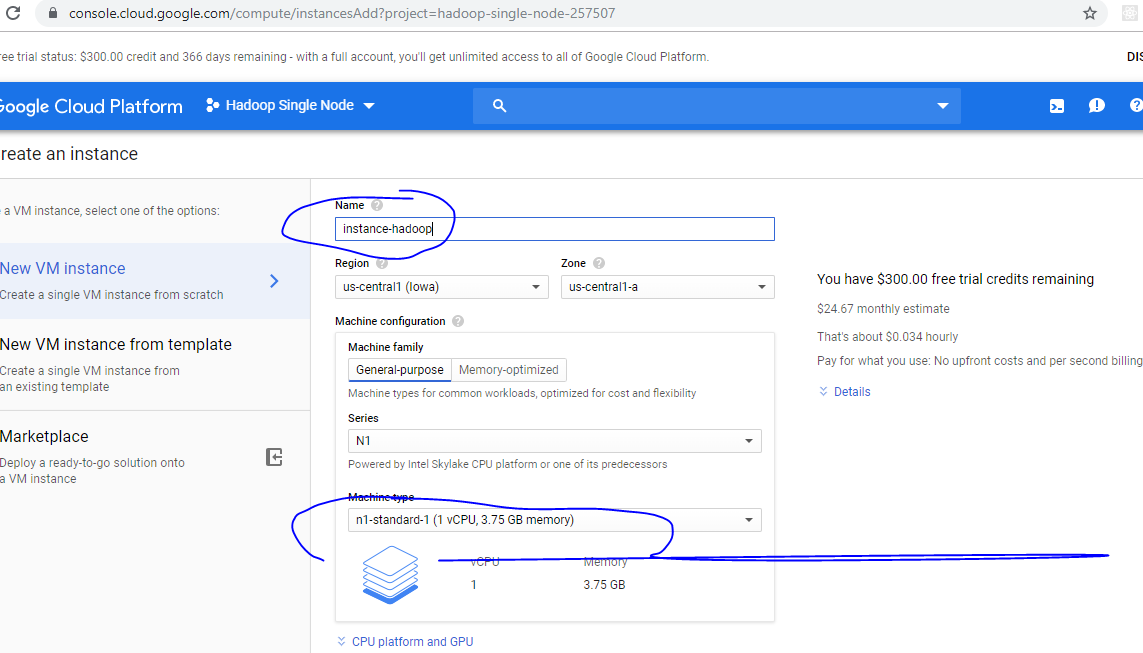


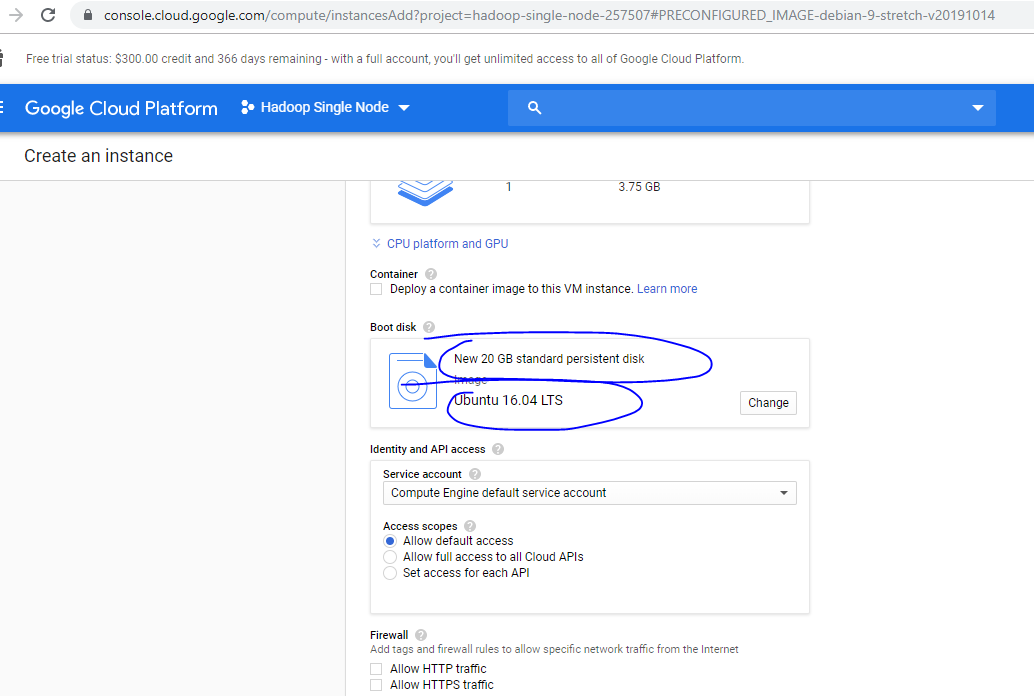
# Step 1: Create a new project



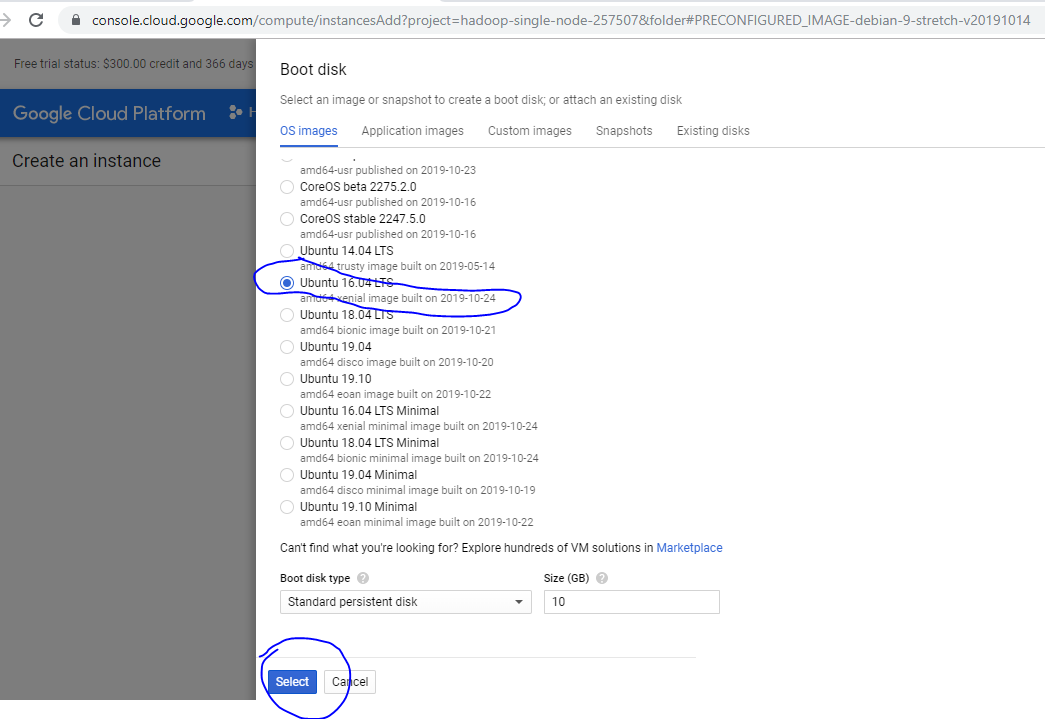
# Step 3: Init an instance

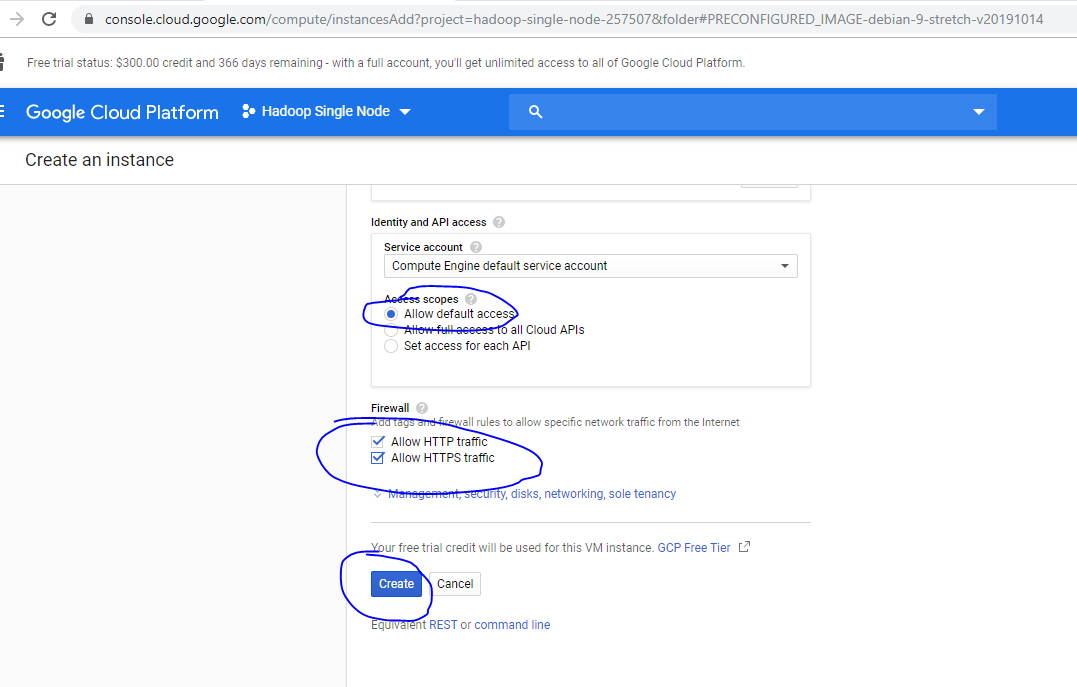




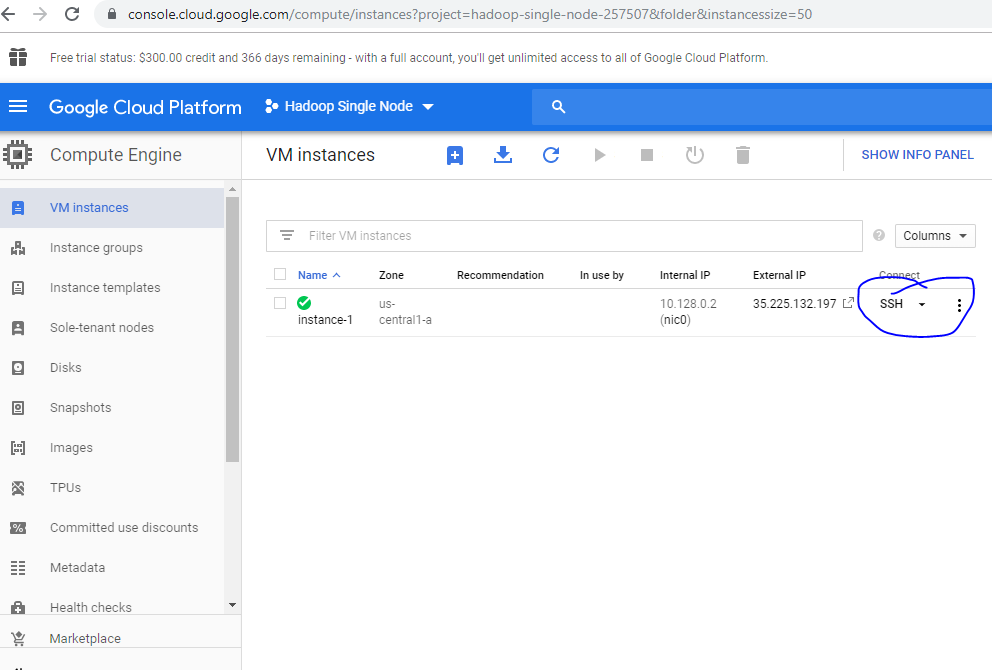


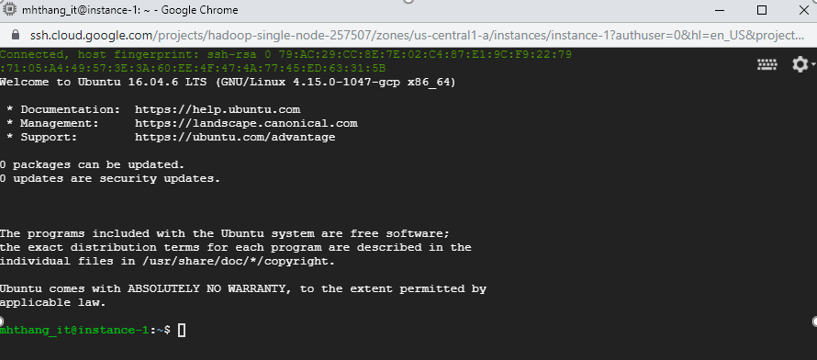
Recommend 20Gb hard-disk because we will have to download cloudera 4.7Gb then extract.





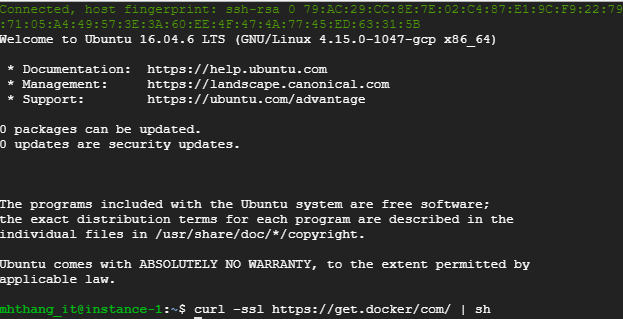
# Step 4: SSH to the instance



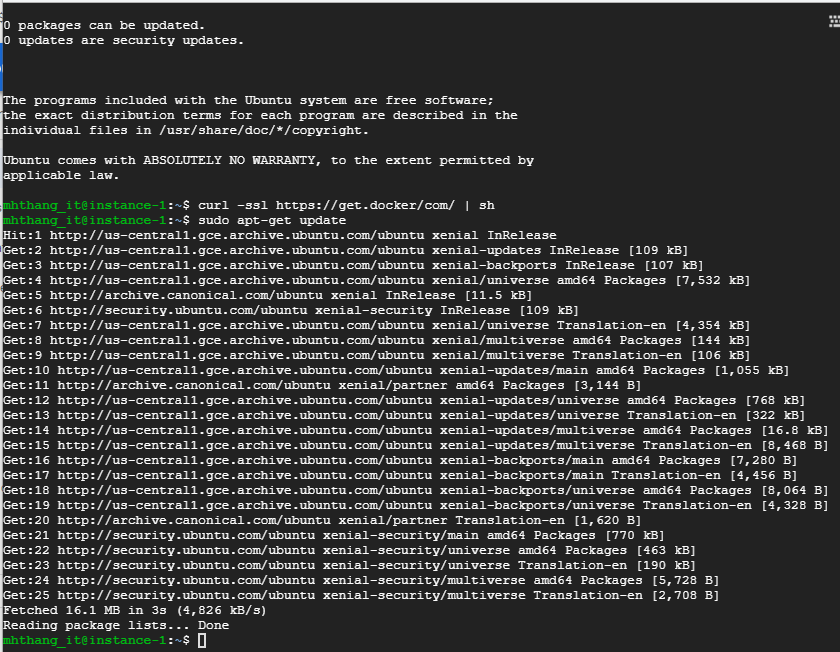


# Step 5: get Docker and update

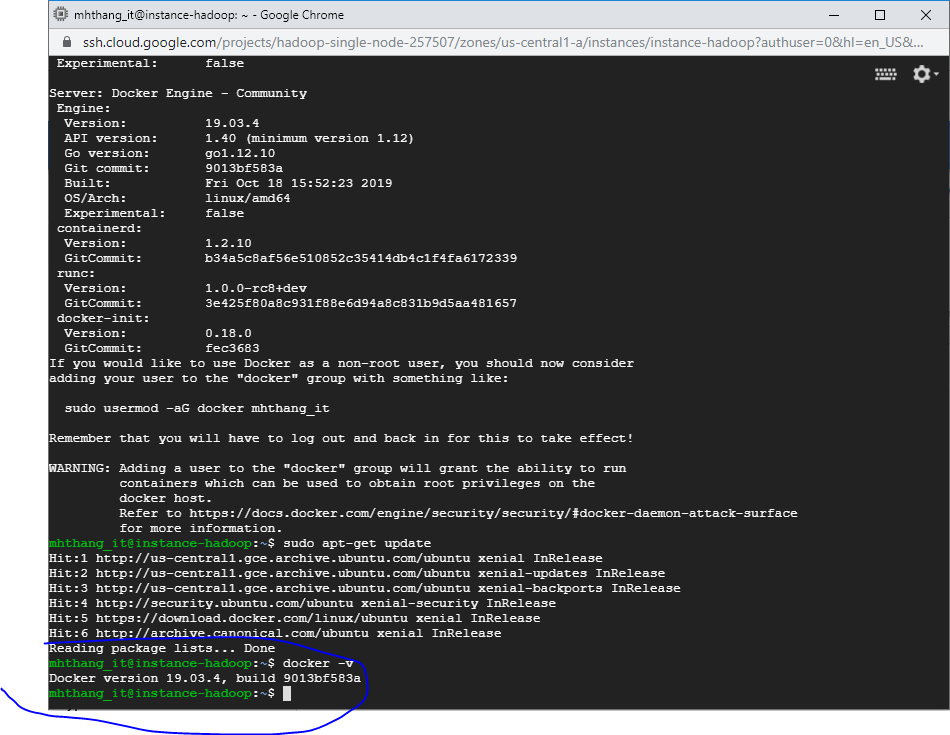
curl -ssl https://get.docker.com/ | sh



sudo apt-get update

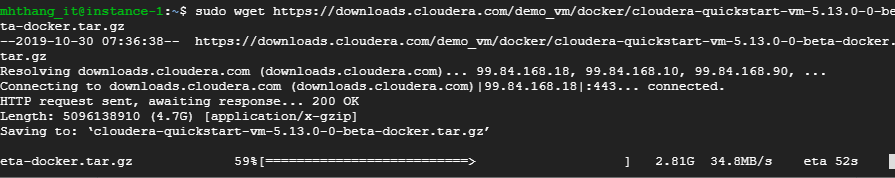


Check version of docker



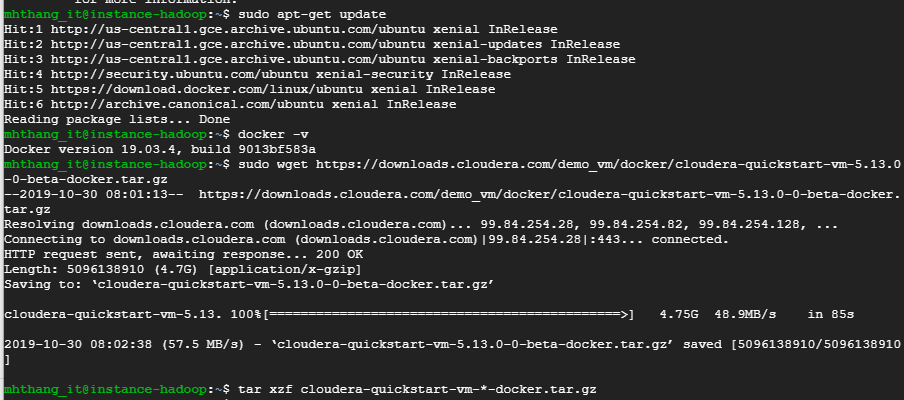
# Step 6: Get the Cloudera Quickstart Image

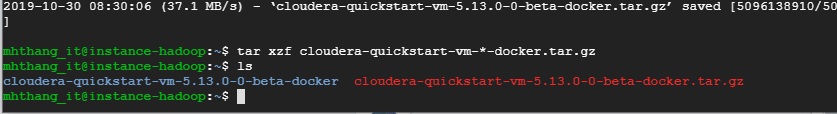
sudo wget <https://downloads.cloudera.com/demo_vm/docker/cloudera-quickstart-vm-5.13.0-0-beta-docker.tar.gz>



# Step 7: Extract the cloudera quickstart tar file

tar xzf cloudera-quickstart-vm-\*-docker.tar.gz



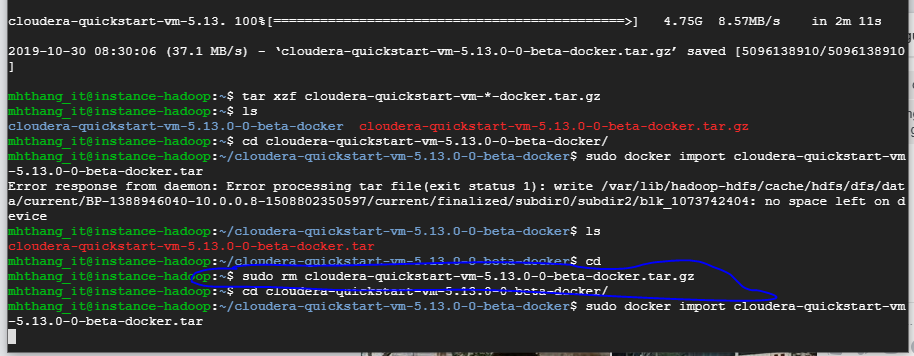


# Step 8: Import Cloudera quickstart docker image

cd cloudera-quickstart-vm-5.13.0-0-beta-docker

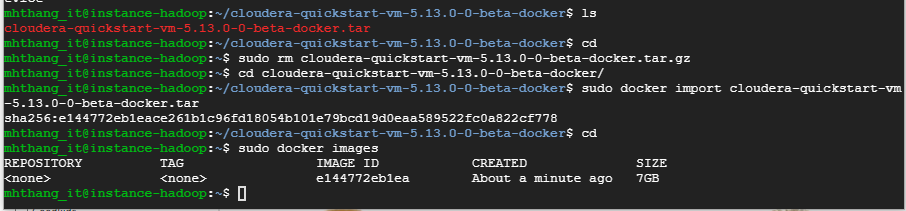
sudo docker import cloudera-quickstart-vm-5.13.0–0-beta-docker.tar

Remove downloaded file if short of disk space



# Step 9: Check the container image ID

sudo docker images



# Step 10: Step Run docker image

sudo docker run --hostname=quickstart.cloudera --privileged=true -t -i -p 8777:8888 -p 7190:7180 -p 90:80 cd751c781bf8 /usr/bin/docker-quickstart

***Now, we are ready to run Hadoop, example with wordcount map reduce. Step to create and export Hadoop project, ref to Set Up a Single Node Cluster using VM or Docker document.***

# Step 11: Create folder

[root@quickstart ~]# hadoop fs -mkdir /user/cloudera/wordcount /user/cloudera/wordcount/input



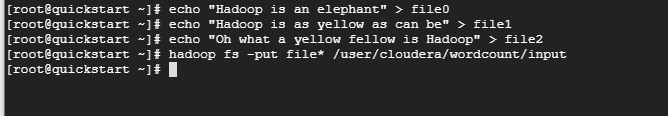
# Step 11: Create documents

echo "Hadoop is an elephant" > file0

echo "Hadoop is as yellow as can be" > file1

echo "Oh what a yellow fellow is Hadoop" > file2

hadoop fs -put file\* /user/cloudera/wordcount/input



# Step 12: Load \*.jar file into cloudera

Jar ~ julius~ hadoop jar wordcount1.jar hadoop.inmapperwordcount.WordCount /user/cloudera/wordcount/input /user/cloudera/wordcount/output

Jar~ mine~ hadoop jar wordcount2.jar AverageComputation /user/cloudera/wordcount/input /user/cloudera/wordcount/output

In this step, the wordcount.jar is downloaded from an hyperlink. You can find another way to copy file from your machine into google cloud instance.

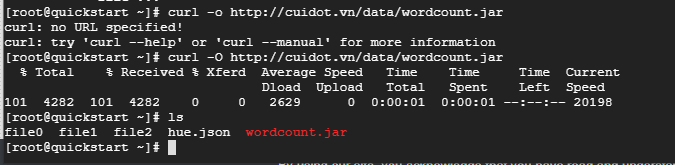
//inmapper

docker container ls

sudo docker cp wordcount2.jar 3ecda7cb598e:/wordcount2.jar

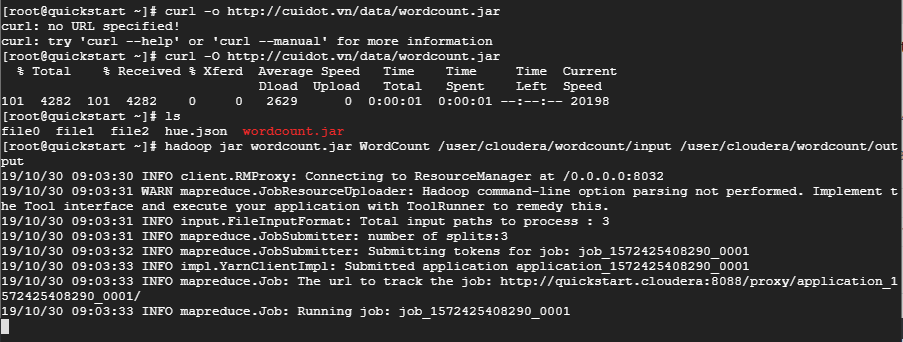
curl -O <https://drive.google.com/uc?export=download&id=1bj-Hi3fPZiz4Zs1KuvYX8Ph8xpXxTgrf>

curl -O <https://github.com/reza5630/DataFrame/raw/master/DataFrame/wordcount.jar>

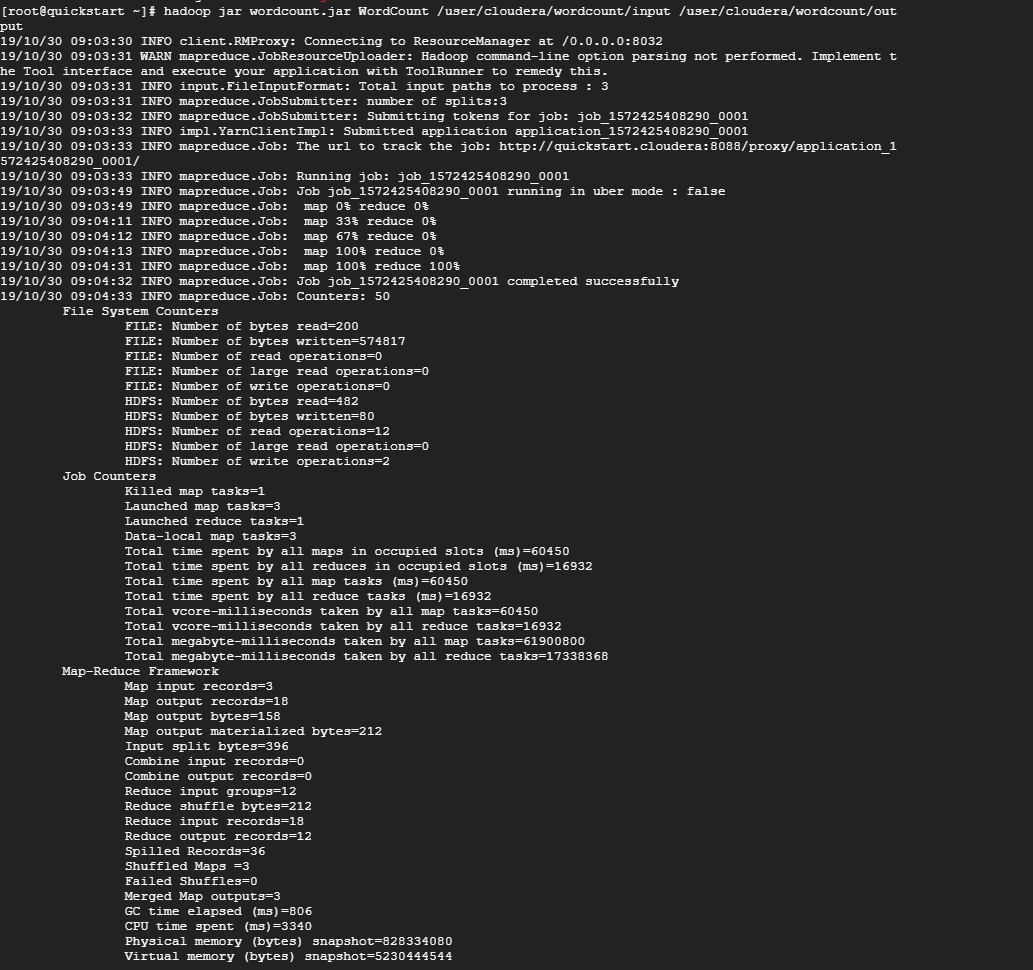


# Step 13: Run the WordCount application from the JAR file we created on Eclipse, giving the paths to the input and output directories in HDFS.

hadoop jar wordcount.jar org.myorg.WordCount /user/cloudera/wordcount/input /user/cloudera/wordcount/output



Result:



Good Luck

To delete dir

hadoop fs -rm -r -f /user/cloudera/wordcount/output

To view the contents of output folder

hdfs dfs -ls /user/cloudera/wordcount/output

To open the output file using cat

hadoop fs -cat /user/cloudera/wordcount/output/part-r-00000