**New Jersey Institute of Technology**

Fall 22 CS 643

Programming Assignment -II

Professor: Zunnurhain, Kazi (PhD)

Students:

Muhammad Imran Burhanullah (mib3@njit.edu)

**OBJECTIVE:**

The purpose of this individual assignment is to learn how to develop parallel machine learning (ML) applications in Amazon AWS cloud platform. Specifically, you will learn: (1) how to use Apache Spark to train an ML model in parallel on multiple EC2 instances; (2) how to use Spark’s MLlib to develop and use an ML model in the cloud; (3) How to use Docker to create a container for your ML model to simplify model deployment.

**STEP-BY-STEP HOW-TO SETUP THE CLOUD ENVIORNMENT AND RUNNING THE MODEL TRAINING AND THE APPLICATION PREDICTION:**

I used Amazon EMR (Amazon Elastic MapReduce) to create a Spark cluster with Hadoop for my application purpose. First of all, log into Amazon Web Services (AWS) Account

Graphical user interface, application

Description automatically generated

**Screenshot for Elastic Map Reduce > Opening EMR**

**Graphical user interface, text, application, email

Description automatically generated**

Creating Cluster with required details of Configurations: name, select, Spark as

Application, ec2 key pair, select number of instances > click create cluster

Graphical user interface, text, application, email

Description automatically generated

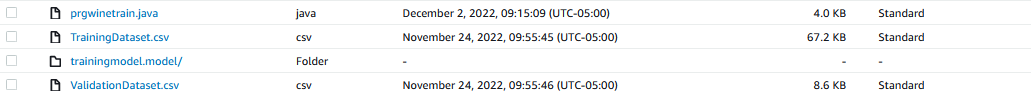
Note I have only as number of instance so it act as a EC2 instance with all configuration of

Spark required for running our prediction application.

Graphical user interface, text, application

Description automatically generated

I have also upload the 2 files in S3 bucket names [s3://[wineprgpredict](https://s3.console.aws.amazon.com/s3/buckets/wineprgpredict?region=us-east-1)](https://s3.console.aws.amazon.com/s3/buckets/myawss32022?region=us-east-1)



Cluster starting ….

Graphical user interface, text, application, email

Description automatically generated

Command : **hadoop@ec2-3-87-237-159.compute-1.amazonaws.com**

Graphical user interface, text

Description automatically generated

Text

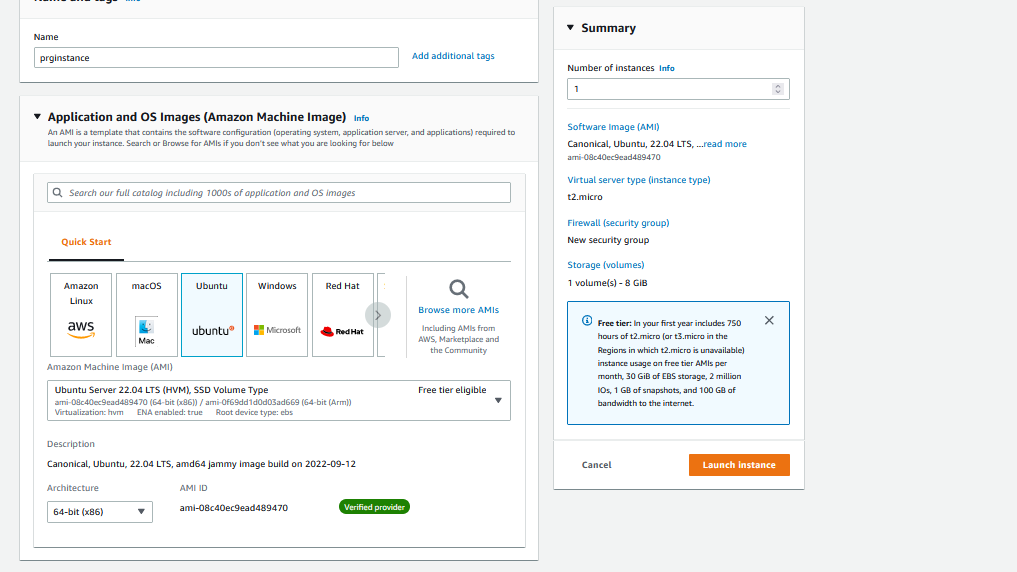
Description automatically generated with medium confidence

Table

Description automatically generated

**Task2: Running on single instance**

Launching an Ubuntu instance.

****

Graphical user interface, application

Description automatically generated

2.2 ssh to new instance from your local terminal

Text

Description automatically generated

2.3Installing scala

* wget http://downloads.typesafe.com/scala/2.11.6/scala-2.11.6.tgz
* tar -xzvf scala-2.11.6.tgz
* Update PATH environment variable:
  + - copy following lines into file and then save it
      * export SCALA\_HOME=/home/ec2-user/scala-2.11.6
      * export PATH=$PATH:/home/ec2-user/scala-2.11.6/bin
    - source ~/.bashrc

Text

Description automatically generated

* 1. Install Spark

1. wget <https://archive.apache.org/dist/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz>
2. sudo tar xvf spark-2.4.5-bin-hadoop2.7.tgz -C /opt
3. sudo chown -R ubuntu:ubuntu /opt/spark-2.4.5-bin-hadoop2.7
4. sudo ln -fs spark-2.4.5-bin-hadoop2.7 /opt/spark
5. Update PATH Environment

$ vim ~/.bash\_profile

1. copy following lines into file and then save it

export SPARK\_HOME=/opt/spark

PATH=$PATH:$SPARK\_HOME/bin

export PATH

1. Check java –version

If not install use cmd

Wget https://download.java.net/java/GA/jdk17/0d483333a00540d886896bac774ff48b/35/GPL/openjdk-17\_linux-x64\_bin.tar.gz

tar xvf openjdk-17\_linux-x64\_bin.tar.gz

sudo mv jdk-17 /opt/

sudo tee /etc/profile.d/jdk.sh <<EOF

export JAVA\_HOME=/opt/jdk-17

export PATH=\$PATH:\$JAVA\_HOME/bin

EOF

source /etc/profile.d/jdk.sh

Text

Description automatically generated

wget <https://mirrors.estointernet.in/apache/maven/maven-> 3/3.6.3/binaries/apache-maven-3.6.3-bin.tar.gz

tar -xvf apache-maven-3.6.3-bin.tar.gz

sudo mv apache-maven-3.6.3 /opt/

Run vim .profile

M2\_HOME='/opt/apache-maven-3.6.3'

PATH="$M2\_HOME/bin:$PATH"

export PATH

Run source .profile

Text

Description automatically generated

sudo apt install default-jdk scala git -y

java -version; javac -version; scala -version; git –version

wget <https://downloads.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-> hadoop3.tgz

Text

Description automatically generated

**Task 3: Prediction by using docker image.**

Install docker in ubuntu instance.

3.1 Install docker

1. sudo apt install apt-transport-https ca-certificates curl software-properties-common
2. sudo apt install docker-ce
3. apt-cache policy docker-ce
4. curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add -
5. $sudo apt-get install docker

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

#mkdir predict

#mkdir -p dataset

#nano prgwinetrain.java

#docker build -t predict