CS 643 101

Cloud Computing - Fall 2020

Programming Assignment 2 - Wine Quality Prediction

Name: Richa Shrivastava

UCID: rs2353

Description:

The purpose of this individual project 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.

You have to build a wine quality prediction ML model in Spark over AWS. The model must be trained in parallel on multiple EC2 instances. Then, you need to save and load the model in an application that will perform wine quality prediction; this application will run on one EC2 instance. The project must be implemented in Java, Scala, or Python on Ubuntu Linux.

Docker Link : <https://hub.docker.com/repository/docker/richashrivastava/winequalityprediction>

Github Link : https://github.com/Richa2312/Cloud-pa2-wine-quality-prediction

Step by Step Installation process and commands for setup on cloud:

\*\*Creating EMR

Create EMR Cluster on AWS using key pair and 4 instances

Enable SSH inbound rule for port 22 in security group of EMR Master

Connect to Master instance using PuTTY

Upload the folder containing csv and python files using WinSCP

\*\* Update master & check python versions:

sudo yum update -y

python --version

sudo nano /etc/sudoers

/\*find below line with Defaults secure\_path = /sbin:/bin:/usr/sbin:/usr/bin

Just add /usr/local/bin to this PATH \*/

\*\* Install findspark, pyspark, prettytable and upgrade pip:

sudo pip install --upgrade pip

sudo pip install findspark

sudo pip install pyspark

sudo pip install prettytable

\*\* Run the training program:

python train.py

\*\* Run the Test program for prediction: (Add TestDataset before running this program)

python test.py

\*\* Docker Installation and Run the program:

sudo yum install -y docker

sudo service docker start

sudo docker build . –t richashrivastava/winequalityprediction

sudo docker run -t richashrivastava/winequalityprediction

\*\* Please find below set of commands for installation errors and checks:

sudo pip install --upgrade pip

sudo pip install findspark

sudo pip install prettytable

sudo pip install numpy

sudo docker rmi -f image image\_name

sudo docker images

sudo docker run username/image\_tag\_name

sudo docker build . -t username/image\_tag\_name

sudo docker start container\_name

sudo docker stop container\_name

sudo docker login -u richashrivastava