PROGRAMMING ASSIGNMENT-2

Wine Quality Prediction AWS Application:

Pa2Winepred: This project involves the development of a python application utilizing the PySpark interface.

The application is deployed on an AWS Elastic MapReduce cluster. The objective is to parallelly train a machine learning model on EC2 instances for predicting wine quality using publicly available data. The trained model is employed to predict the quality of wine. Docker is utilized to create a container image for the trained machine learning model, streamlining the deployment process.  
  
GITHUB: https://github.com/DineshMadiraju/winepredqulity

DOCKER: https://hub.docker.com/repository/docker/dm765/predwiequ/general

How to execute Wine Quality Prediction AWS Application:

1. Create a Key-pair for the EMR Cluster.
2. Go to EC2/Network/Key-pairs
3. We will be using the format of .pem and download the keypair.
   1. Keypair created: winepress.pem
4. Create an S3 bucket
   1. S3 bucket: pa2wine
5. Go to EMR console and create EMR cluster.
6. Create the Spark in AWS instance using EMR Console:
   1. Cerate spark cluster by using the EMR console and create 4 instances
   2. Name and Application: Dineshcluster
   3. Application bundle: Hadoop 2.10.1, Spark 2.4.7, Zippeline 0.9.0, and Yarn
   4. Cluster configuration Instance Group
   5. Cluster Scaling and proivisioning: Set the cluster size for the instances
   6. Networking and Cluster Termination: mark on Manually Termination
   7. IAM Role: Set default IAM roles.
   8. And click on create cluster.
7. ML model into spark cluster with ec2 instances in parallel:
   1. We will be connecting master instance in the terminal
   2. Ssh -I “winepress.pem” ec2-52-207-146-147.compute-1.amazonaws.com
   3. After logging in to Master instance change the root using sudo su
8. Submit the task by the command:

spark-submit s3://pa2winebucket1/ winequilityprediction.py

1. Submit the task by the command:

spark-submit s3://pa2wine/ winequilityprediction.py

1. Then you can find the trace status for the above tasks, The status is succeed then there is a creation of test.model in the s3 bucket s3://pa2wine.
2. We have to run the ML model using the Docker:
   1. Create docker account, download and sign in
   2. After logging in we will buld the image:

docker build -t predwiequ

1. The push and pull into the docker hub repository:

PUSH:

docker tag predwiequ dm765/ predwiequ

docker push dm765/ predwiequ

PULL:

docker pull dm765/ predwiequ

1. Store the test data file in the container
   1. docker run -v C:\Winepredqulity\data\csv predwiequ testdata.csv

A screenshot of a computer screen

Description automatically generated

And we have the Accuracy.