Programming Assignment-2

Quality prediction:

The aim of this project is to create a Python application using the PySpark framework, running on an Amazon Web Services (AWS) Elastic MapReduce (EMR) cluster. Its main goal is to train a machine learning model concurrently on EC2 instances to predict wine quality using publicly accessible data. After training, the model is used to make predictions on wine quality. Docker is utilized to build a container image for the trained model, streamlining deployment.

Github:

Docker: https://hub.docker.com/repository/docker/sp3294/qulpred

EMR Cluster Steps:

1. Begin by accessing the AWS Management Console: Visit https://aws.amazon.com/ and log in with your credentials to access the AWS Management Console.

2. Proceed to the EC2 Dashboard: After logging in, locate the EC2 Dashboard by typing "EC2" into the AWS services search bar and selecting it.

3. Select "Key Pairs": Within the EC2 Dashboard, navigate to the "Network & Security" section on the left-hand side, then choose "Key Pairs".

4. Generate a Key Pair: Click on the "Create Key Pair" button.

5. Name Your Key Pair: Assign a name to your key pair.

6. Download Your Key Pair: Save the key pair as "predkey.pem".

7. Next, navigate to the EMR console and create an EMR cluster.

8. Creating Spark on the AWS instance using the EMR console: Use the EMR console to set up the Spark cluster and create four instances.:

Name and application:

Amazon EMR release: emr-5.33.0

Application bundle: Hadoop 2.10.1 Spark 2.4.7, Zippeline 0.9.0, and Yarn

By following images, we need to create EMR Cluster:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Running AWS in CMD:

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer error

Description automatically generated

Docker:

A screen shot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated