

Programming Assignment-2

CS643-Cloud Computing

Submitted By:
Rohit Dsouza
UCID: rgd3

Section-1

GitHub URLs:

GitHub link:https://github.com/RohitD007/CS643_CloudProgramming/Prediction.py

Docker hub URL:

Docker container for prediction application-

Docker

Link: https://hub.docker.com/repository/docker/dsouzarohit/cs643_rgd3_programming_assignment2

Goal:

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.

Description:

Using Flintrock I created Spark Cluster for my application. I used the Python as my programming language. The input of my model is Training Dataset and Validation dataset. I used WinSCP to give Dataset input for all m slave nodes and only master node contains Python File. Login Into AWS Account Initially created single instance and

downloaded New Key Pair. Converted Key Pair using PuTTY gen and Started created instance using PuTTY using Public IPv4 DNS.

Credentials:

Using PuTTY Set the credentials for our instances. Using nano credentials command. Set the Access key, Secret Key along with the aws session token

Flintrock

Installed the Flintrock on the instance using pip3 install flintrock command. After installing I configured the yml file with the instance credentials along with that I specified the number of nodes that I wanted to create.

inbound Rules

After Creating the Cluster add the SSH inbound rule to the master node for port 22

Login into Master node

Using command Flintrock start . I started the cluster and Using the command Flintrock login

WinSCP

Start WinSCP and login into master node and all the slave nodes using the Public IPv4 DNS. I upload the Python code file in master node along with the two dataset file. and slave also contains the datafile.

Run Code

Run the code using command: `/home/ec2-user/spark/bin/spark-submit --master spark://ec2.internal:7077 /home/ec2-user/Prediction.py`

Dockers

Install Docker on EC2 Create a simple EC2 instance and fire these commands after you SSH into it. `$ sudo yum update -y`

```
$ sudo amazon-linux-extras install docker
```

```
$ sudo service docker start
```

```
$ sudo usermod -a -G docker ec2-user
```