Cloud Computing

Programming Assignment-2

Name: Olive Satya Priya Gundla

Ucid : og84

GITHUB LINK: https://github.com/Olivesatyapriya/cloudprogrammingassignment2

DOCKER LINK: https://hub.docker.com/u/olivesatyapriya

Steps to Launch Instances:

The following procedure is used to launch instances:

- Go to Services, then EC2, and select Launch Instances.
- Enter the number of instances you want to launch (5 in this case).
- Choose the Amazon Machine Image "Ubuntu Server 20.04 LTS... ami-04505e74c0741db8d".
- Select the t2.large type, which is ideal for Docker as it minimizes memory-related problems.
- Create a new key pair and name it "olive". Download the key pair.
- Under Network Settings, in the Security Groups (Firewall) section, enable "Allow SSH traffic from [Anywhere 0.0.0.0/0]".
- For Configure Storage, increase it from 8 GiB to 16 GiB to install and configure modules/packages on the EC2 instances.
- Keep all other options as default and click "Launch Instance".
- Once your EC2 instance is up and running, you can connect to it through your terminal using the following command.

ssh -i olive.pem ec2-user@Public IPv4 DNS

Links to Training and Validation Datasets:

https://github.com/Olivesatyapriya/cloudprogrammingassignment2/blob/abce7c4e3eb2533f0796f40432e96717834b33ee/TrainingDataset.csv

https://github.com/Olivesatyapriya/cloudprogrammingassignment2/blob/abce7c4e3eb2533f0796f40432e96717834b33ee/ValidationDataset.csv

Parallel Implementation:

Use the following commands:

hadoop/sin/start-dfs.sh

spark/sbin/start-all.sh

spark-submit --master spark://ip- 172-31-26-255.ec2.internal: 7077 train.py hdfs:// ip-172-31-26-255.ec2.internal: 9000/data/Training Dataset cv hdfs:// ip-172-31-26-255.ec2.internal:9000/model

Optional command to show model exists hafs dfs -get /model

spark-submit --master local[*] test.py file:///home/ec2-user/ValidationDataset.csv file:///home/ ec2-user/model/

Single Implementation:

To run Jupyter on SSH, you must first install Anaconda using the condaforge command. Then, use the following commands:

Set a password for Jupyter Notebook:

$ jupyter notebook password

Start Jupyter Notebook with the following command:

$ jupyter notebook –no-browser

To create a new untitled notebook, paste the code from the "prediction.py" file (which can be found in the link provided) into the first command line

https://github.com/Olivesatyapriya/cloudprogrammingassignment2/blob/abce7c4e3eb2533f0796f40432e96717834b33ee/prediction.py

Docker:

Log in to Docker using the following command:

$ docker login

Enter your credentials as prompted.

Navigate to the folder where the Dockerfile is saved. This folder should also contain the model, the Prediction.py application, and the dataset to be used for prediction.

$ cd /home/ubuntu/AWS- cloudprogrammingassignment2/

Build the Docker image with the following command:

$ docker build -t cc362/aws - cloudprogrammingassignment2.

To run the Docker image, use the following command:

docker run -v /home/ec2-user/:/data olivesatyapriya/cloudprogrammingassignment2:latest