CS643 Cloud Programming Assignment 2

READme file:

Git Hub link:

https://github.com/sg2658/cloud_programming_assignment_2

DOCKER HUB link:

sg2658/wine_quality_prediction Tags | Docker Hub

Steps for the Execution:

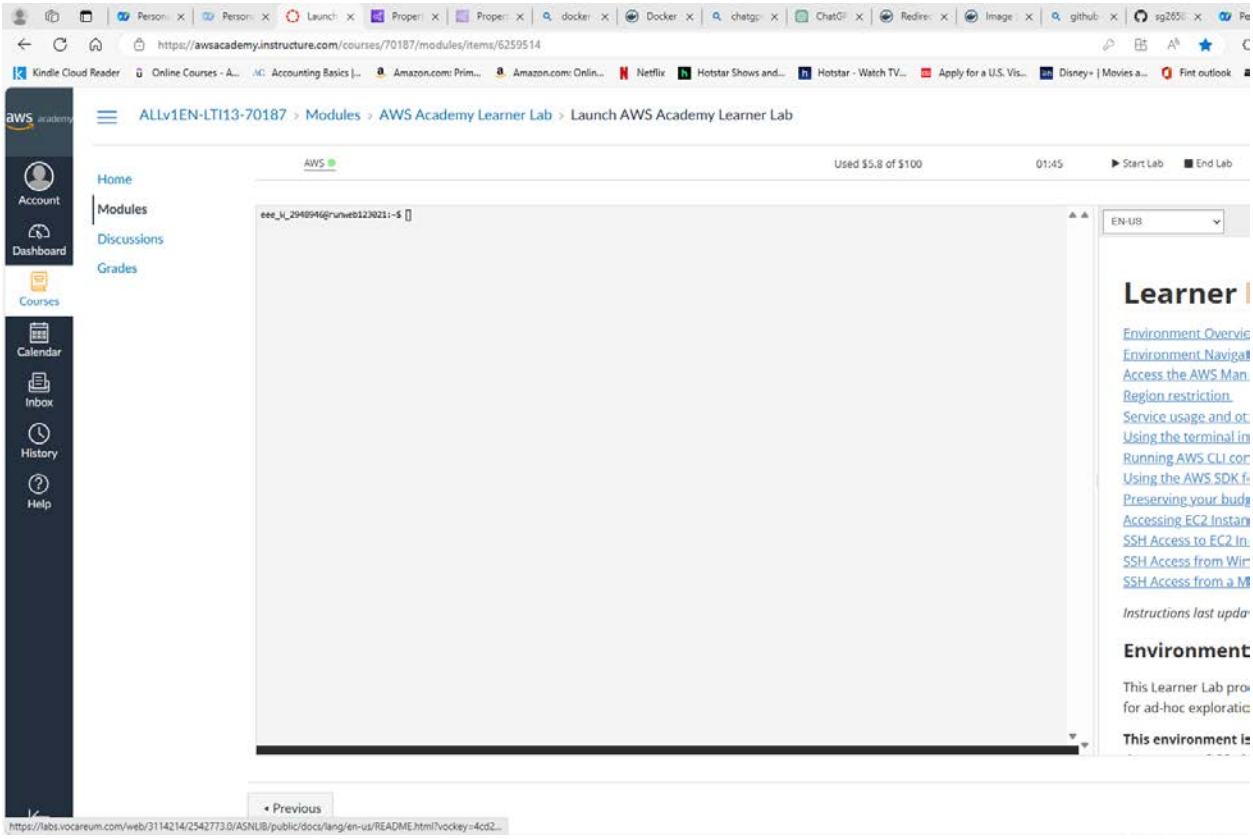
1.Create a Key-pair for the EMR Cluster

2.Create an S3 bucket

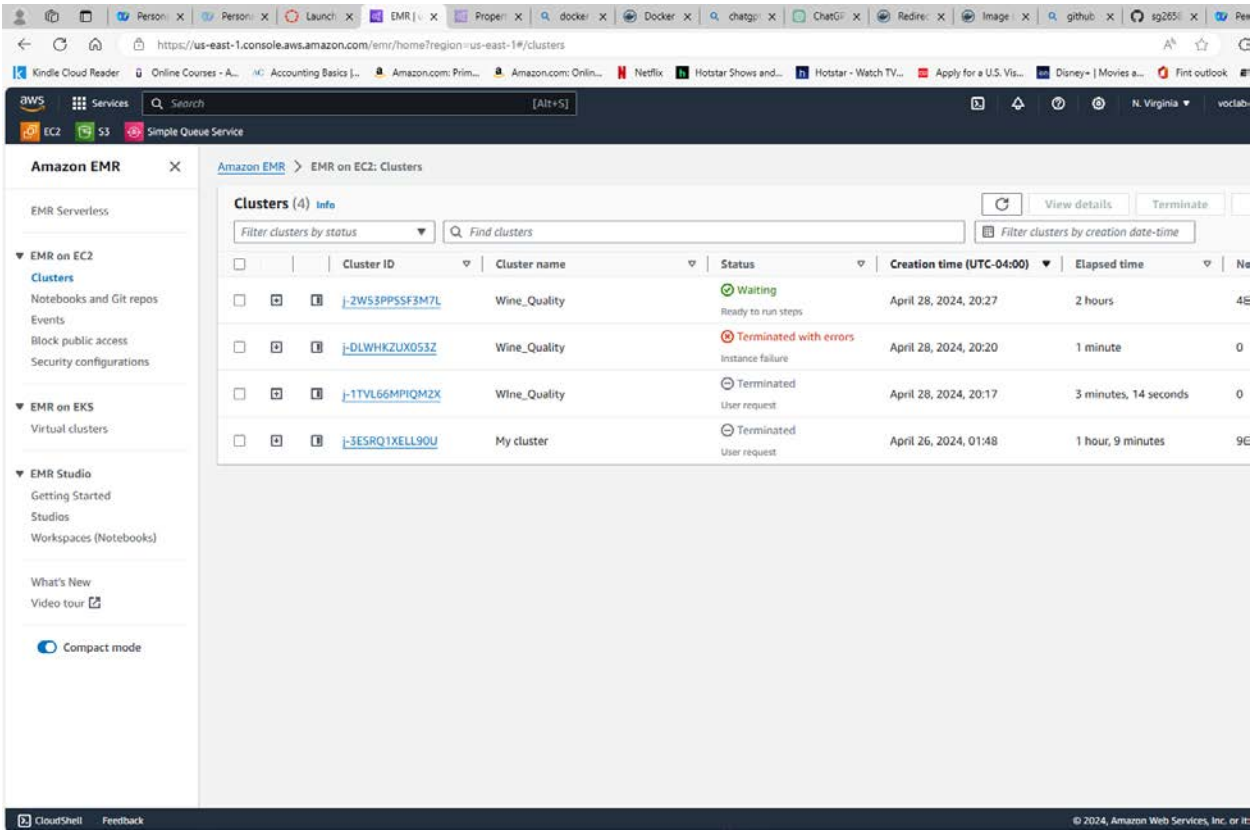
Created an S3 bucket in aws: emrbucketone

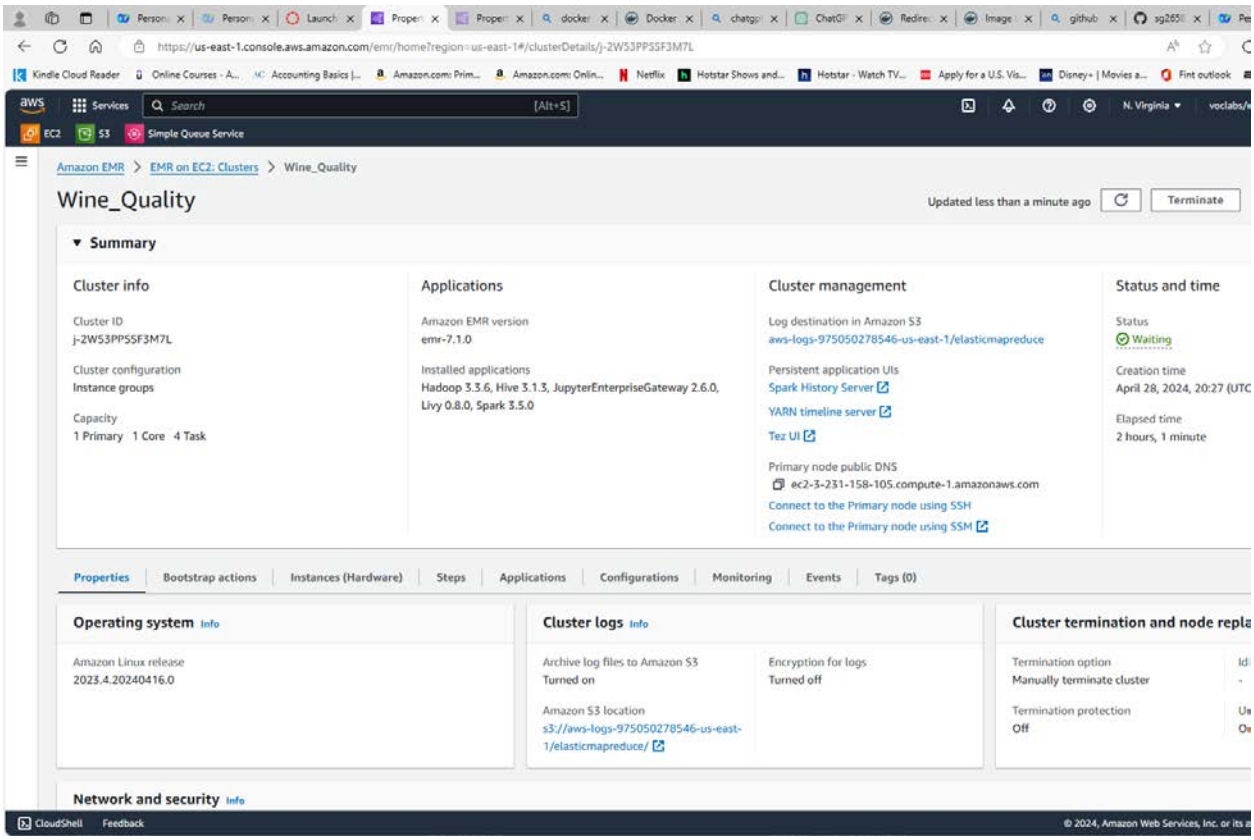
3.Then go to EMR console and create EMR cluster: EMRCluster

4. Creating the spark in the AWS instance by using EMR console



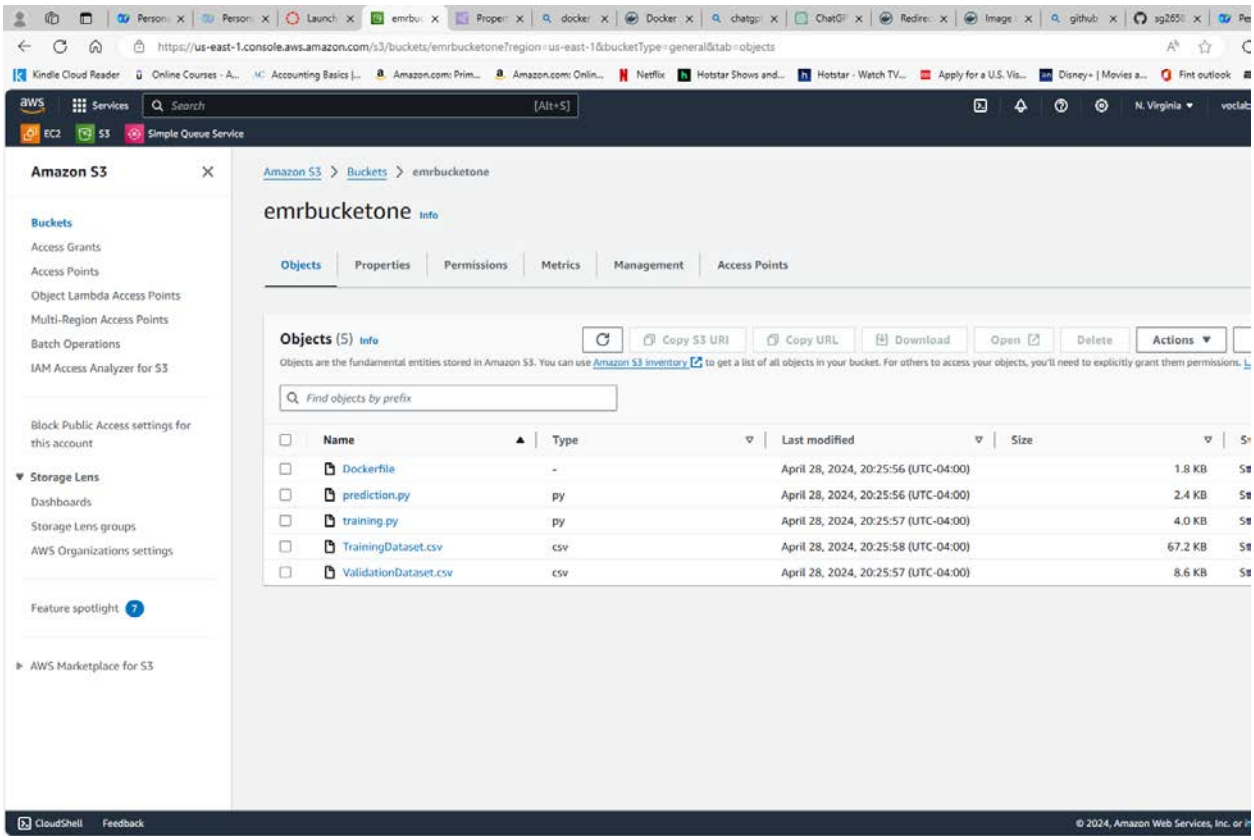
EMR cluster:


https://us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusters
Amazon EMR
EMR Serverless
EMR on EC2
Clusters
Notebooks and Git repos
Events
Block public access
Security configurations
EMR on EKS
Virtual clusters
EMR Studio
Getting Started
Studios
Workspaces (Notebooks)
What's New
Video tour
Compact mode
Amazon EMR > EMR on EC2: Clusters
Clusters (4) Info
View details
Terminate
Filter clusters by status
Find clusters
Filter clusters by creation date-time
Cluster ID
Cluster name
Status
Creation time (UTC-04:00)
Elapsed time
j-2WS3PPSSF3M7L
Wine_Quality
Waiting
Ready to run steps
April 28, 2024, 20:27
2 hours
j-DLWHKZUX053Z
Wine_Quality
Terminated with errors
Instance failure
April 28, 2024, 20:20
1 minute
j-1TVL66MPIQM2X
WIne_Quality
Terminated
User request
April 28, 2024, 20:17
3 minutes, 14 seconds
j-3ESRQ1XELL90U
My cluster
Terminated
User request
April 26, 2024, 01:48
1 hour, 9 minutes
CloudShell
Feedback

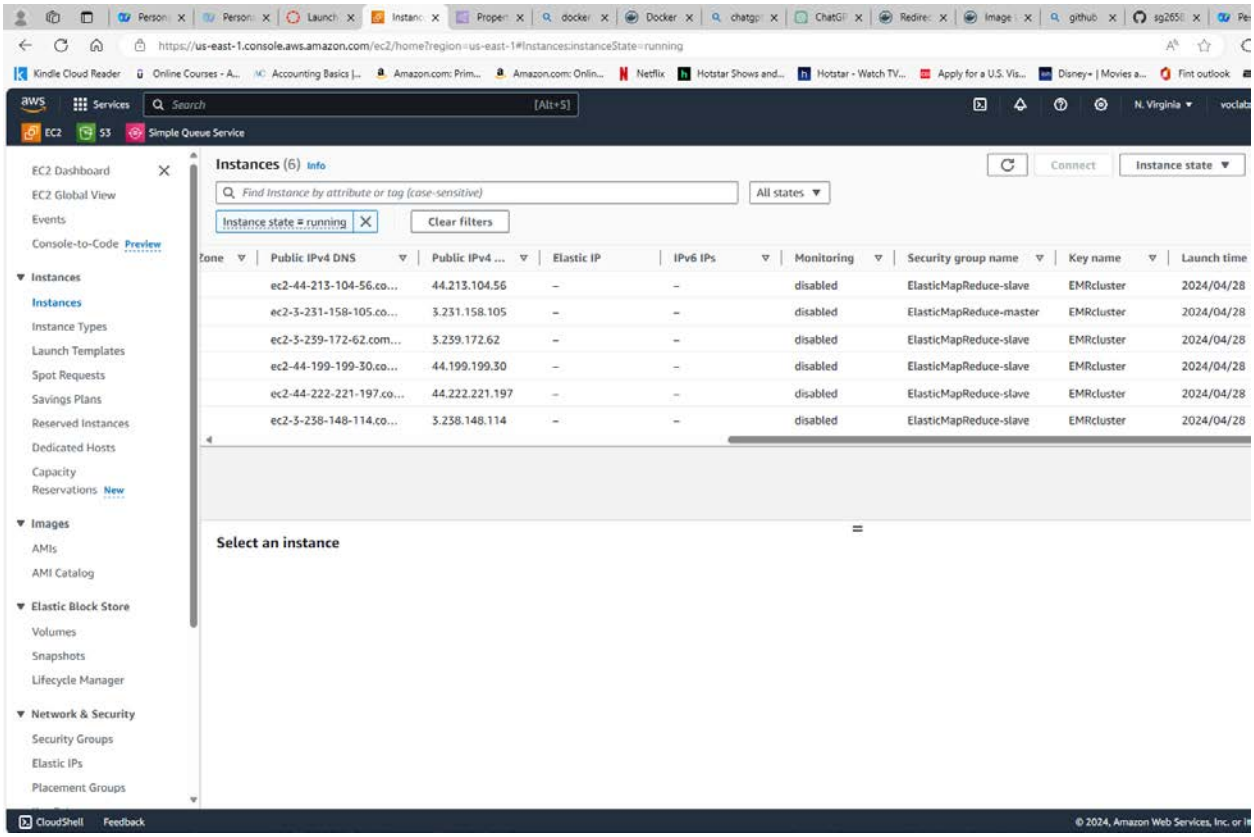


Bucket creation:

Bucket name: emrbucketone

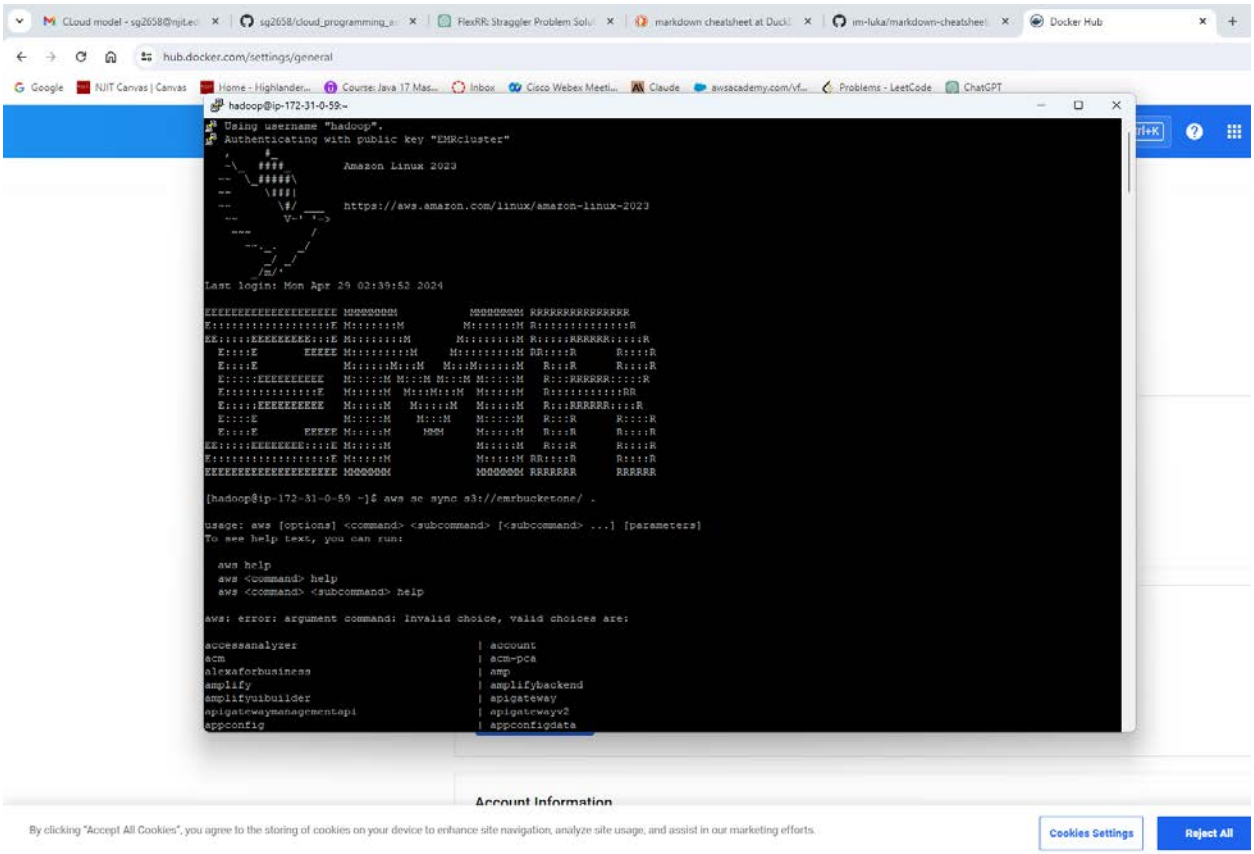


EC2 instance-master:

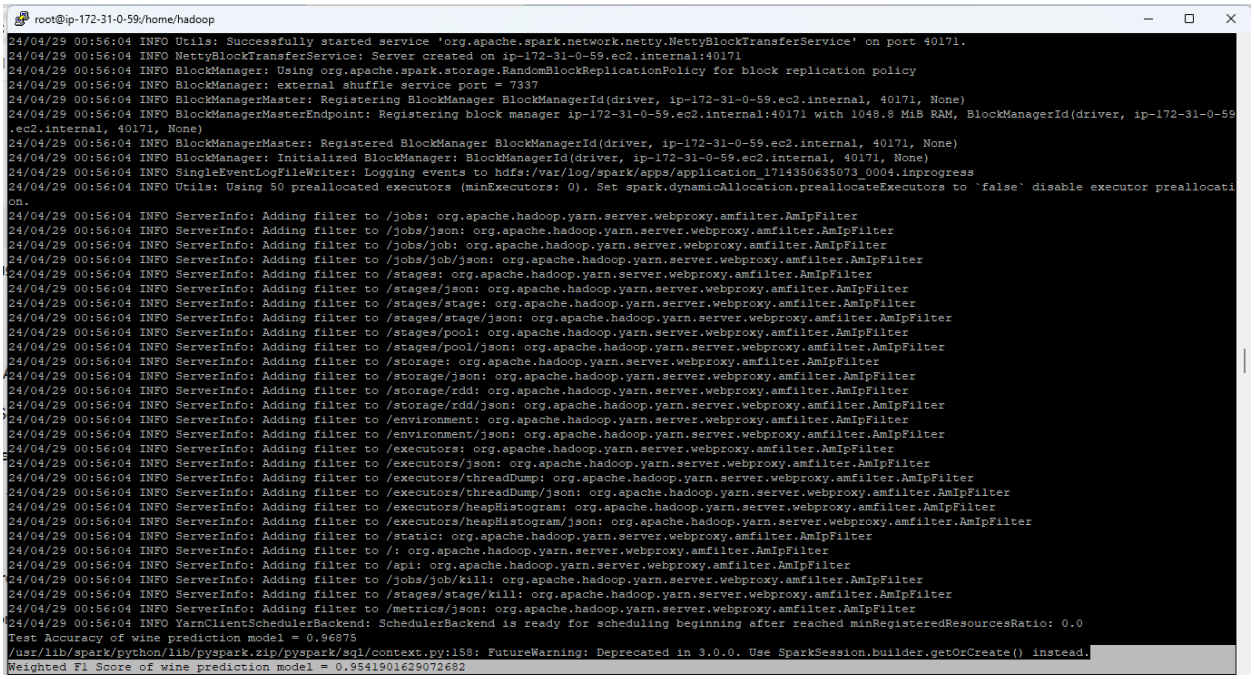


We are training an ML model on an EC2 instance in a spark cluster without Docker:

1.Now the cluster will accept the tasks to run the ML model



WITHOUT Docker:



Now we are running ML model using the Docker:

1.Create a docker account and sign up.

2. After the successful login, download and setup the docker in your local system

3.Install the docker

4. Login the docker in the power shell by the command

docker login

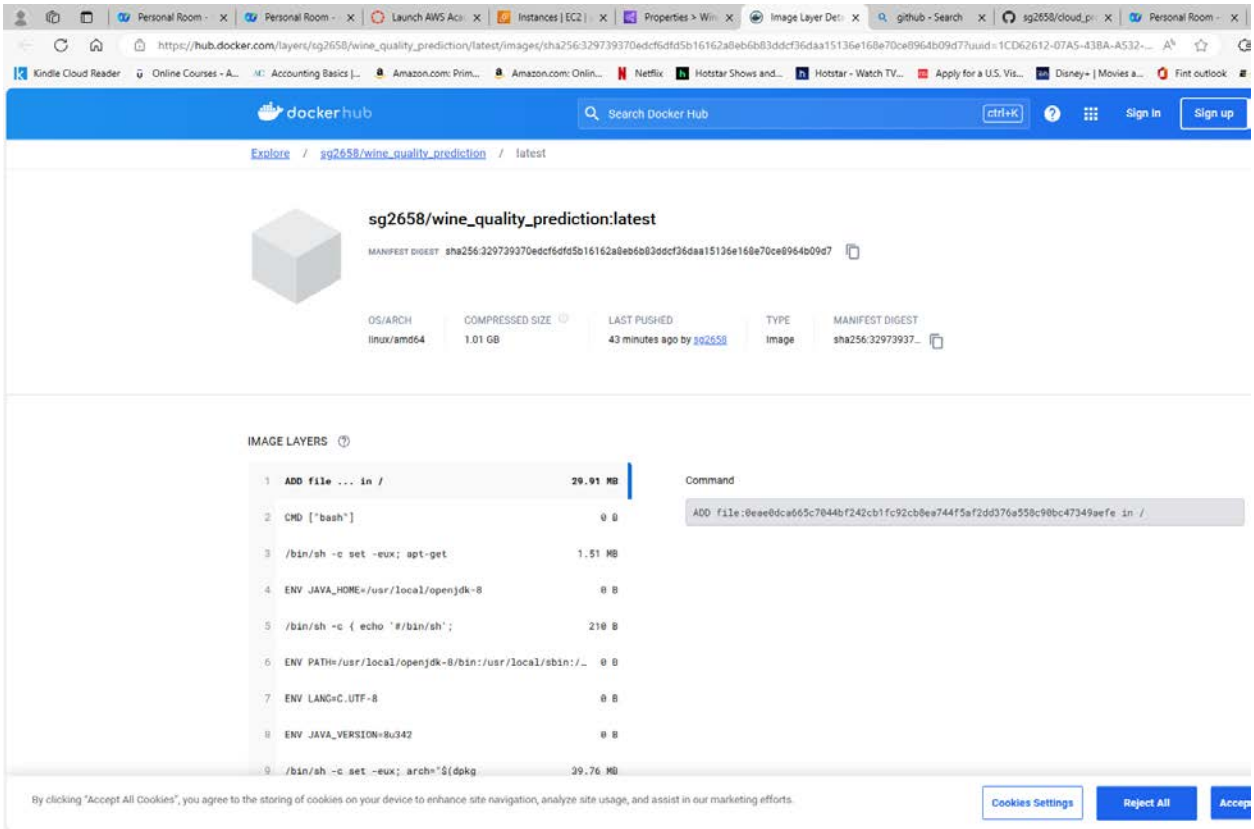
username:

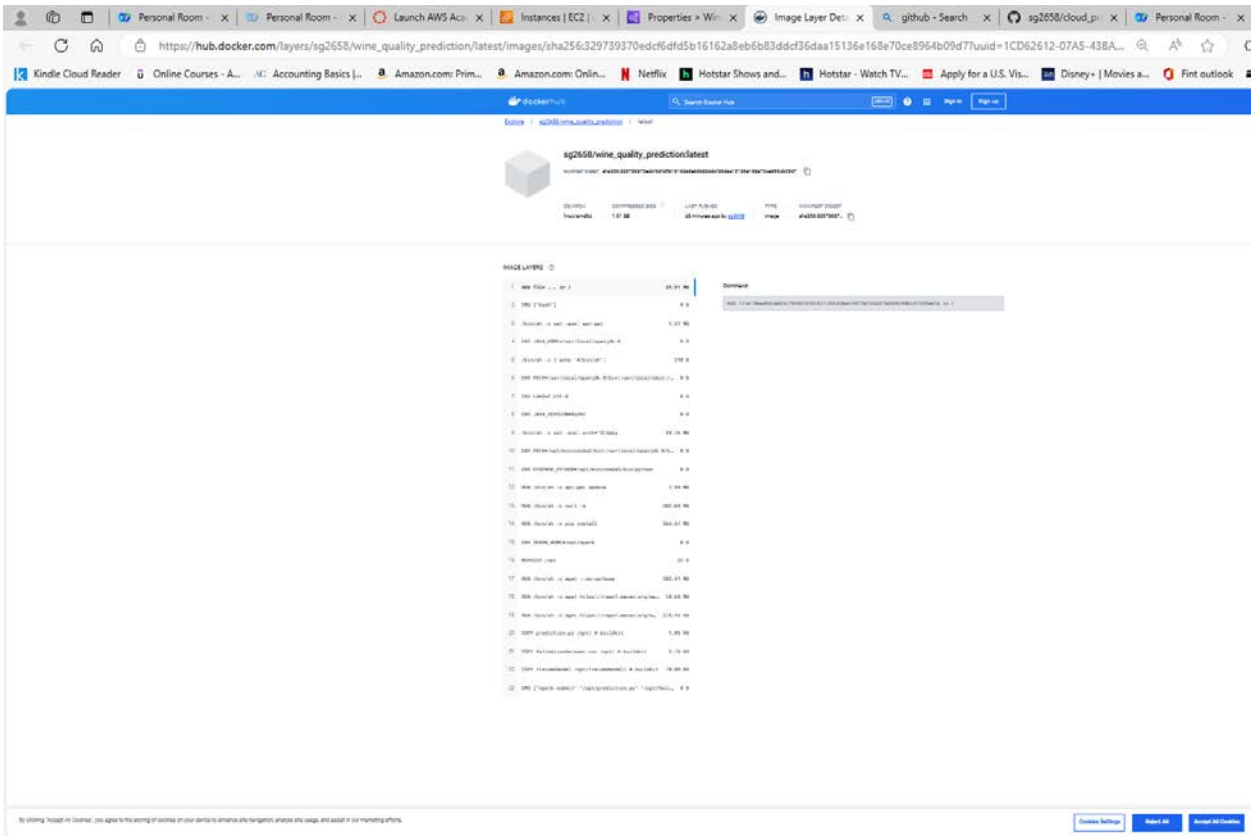
Pwd:

With DOCKER:

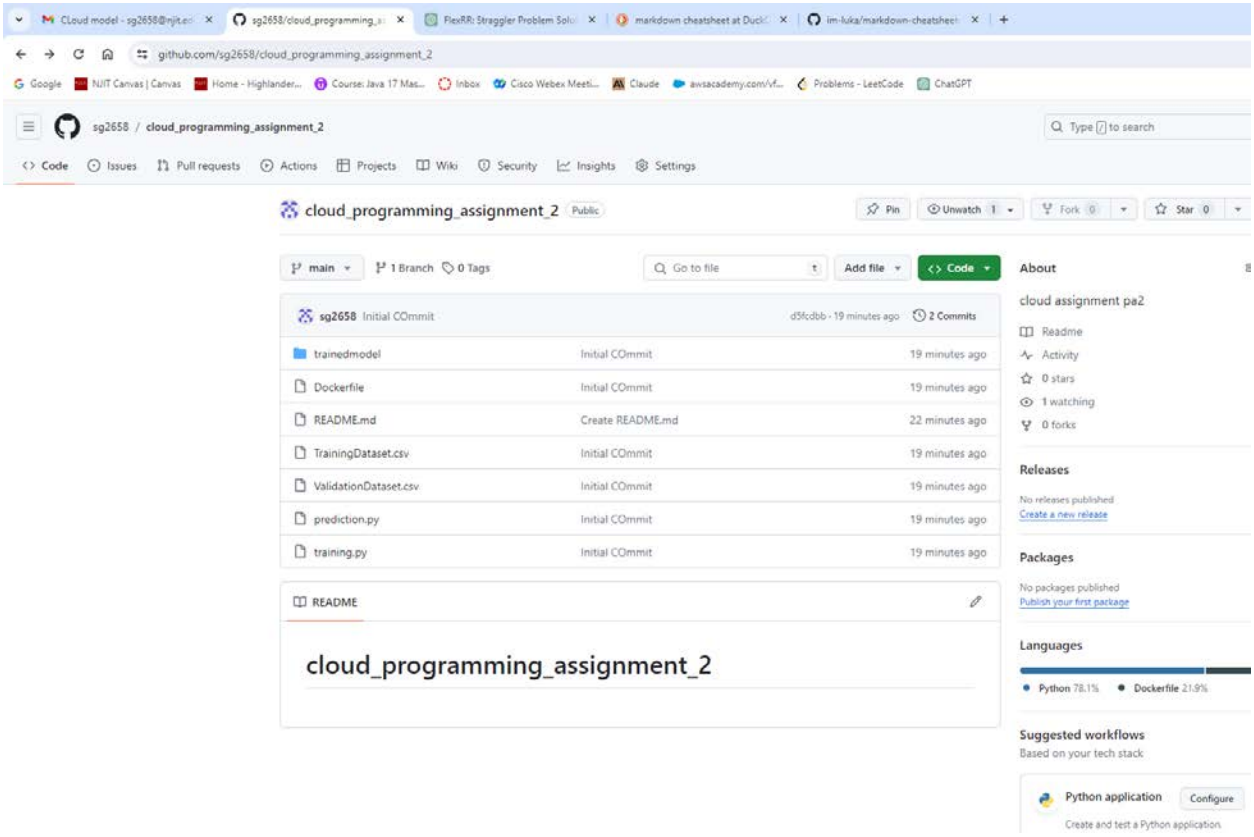
```
24/04/29 01:17:20 INFO Utils: Successfully started service 'SparkUI' on port 4040.
24/04/29 01:17:20 INFO Executor: Starting executor ID driver on host 2605a4a95061
24/04/29 01:17:20 INFO Executor: OS info Linux, 6.1.84-99.169.amzn2023.x86_64, amd64
24/04/29 01:17:20 INFO Executor: Java version 1.8.0_342
24/04/29 01:17:20 INFO Executor: Starting executor with user classpath (userClassPathFirst = false): ''
24/04/29 01:17:20 INFO Executor: Created or updated repl class loader org.apache.spark.util.MutableURLClassLoader@2153aa7c for default.
24/04/29 01:17:20 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 41057.
24/04/29 01:17:20 INFO NettyBlockTransferService: Server created on 2605a4a95061:41057
24/04/29 01:17:20 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
24/04/29 01:17:20 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 2605a4a95061, 41057, None)
24/04/29 01:17:20 INFO BlockManagerMasterEndpoint: Registering block manager 2605a4a95061:41057 with 366.3 MiB RAM, BlockManagerId(driver, 2605a4a95061, 41057, None)
24/04/29 01:17:20 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 2605a4a95061, 41057, None)
24/04/29 01:17:20 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 2605a4a95061, 41057, None)
Test Accuracy of wine prediction model = 0.96875
/opt/spark/python/lib/pyspark.zip/pyspark/sql/context.py:158: FutureWarning: Deprecated in 3.0.0. Use SparkSession.builder.getOrCreate() instead.
Weighted F1 Score of wine prediction model = 0.9541901629072682
Exiting Spark Application
```

Docker hub:




https://hub.docker.com/layers/sg2658/wine_quality_prediction/latest/images/sha256:329739370edcf6dfd5b16162a8eb6b83ddcf36daa15136e168e70ce8964b09d77uuid=1CD62612-07A5-43BA...
sg2658/wine_quality_prediction:latest

GIT hub:


github.com/sg2658/cloud_programming_assignment_2
sg2658 / cloud_programming_assignment_2
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
cloud_programming_assignment_2
Public
Pin
Unwatch 1
Fork 0
Star 0
main
1 Branch
0 Tags
Go to file
Add file
Code
sg2658 Initial COmmit
d5fcdbb · 19 minutes ago
2 Commits
trainedmodel Initial COmmit 19 minutes ago
Dockerfile Initial COmmit 19 minutes ago
README.md Create README.md 22 minutes ago
TrainingDataset.csv Initial COmmit 19 minutes ago
ValidationDataset.csv Initial COmmit 19 minutes ago
prediction.py Initial COmmit 19 minutes ago
training.py Initial COmmit 19 minutes ago
README
cloud_programming_assignment_2
About
cloud assignment pa2
Readme
Activity
0 stars
1 watching
0 forks
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
Languages
Python 78.1%
Dockerfile 21.9%
Suggested workflows
Based on your tech stack
Python application
Configure
Create and test a Python application.