Programming Assignment-2
Cloud Computing

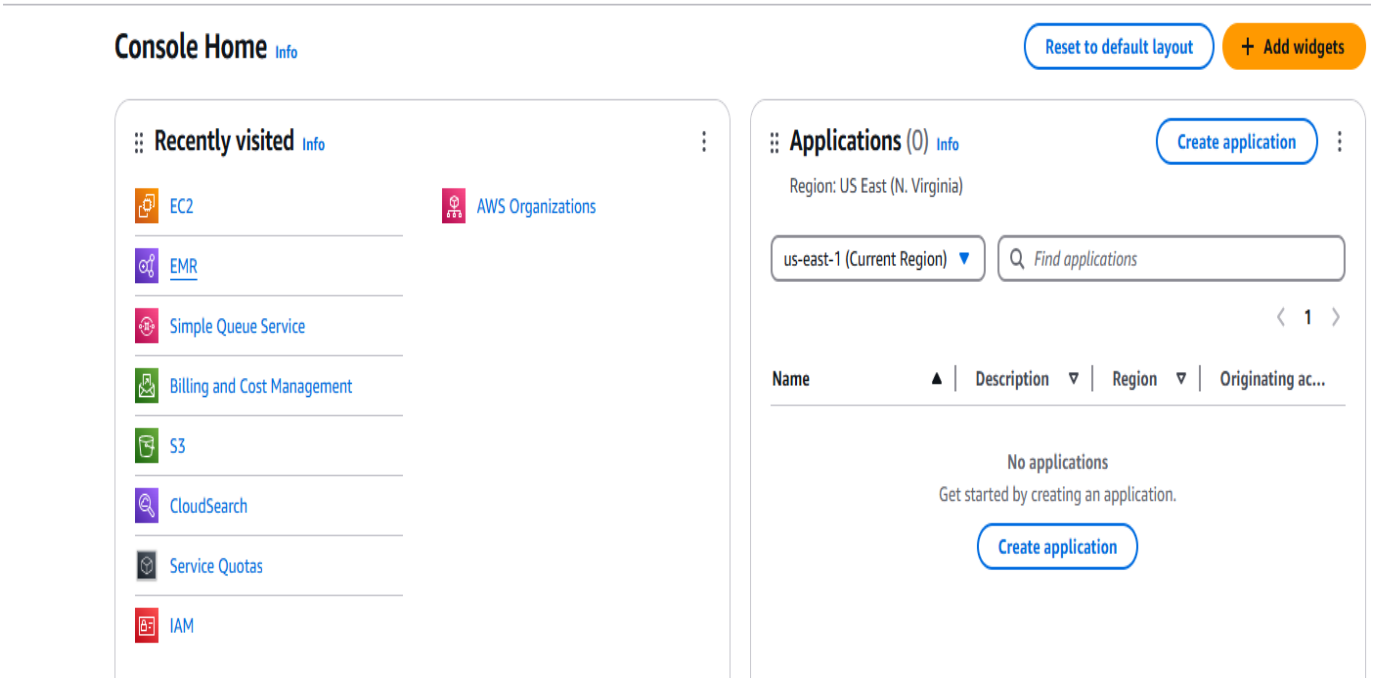
Name – Radhika Rayala
UCID – rr745@njit.edu

GitHub – https://github.com/rr745/wine_quality_predictor

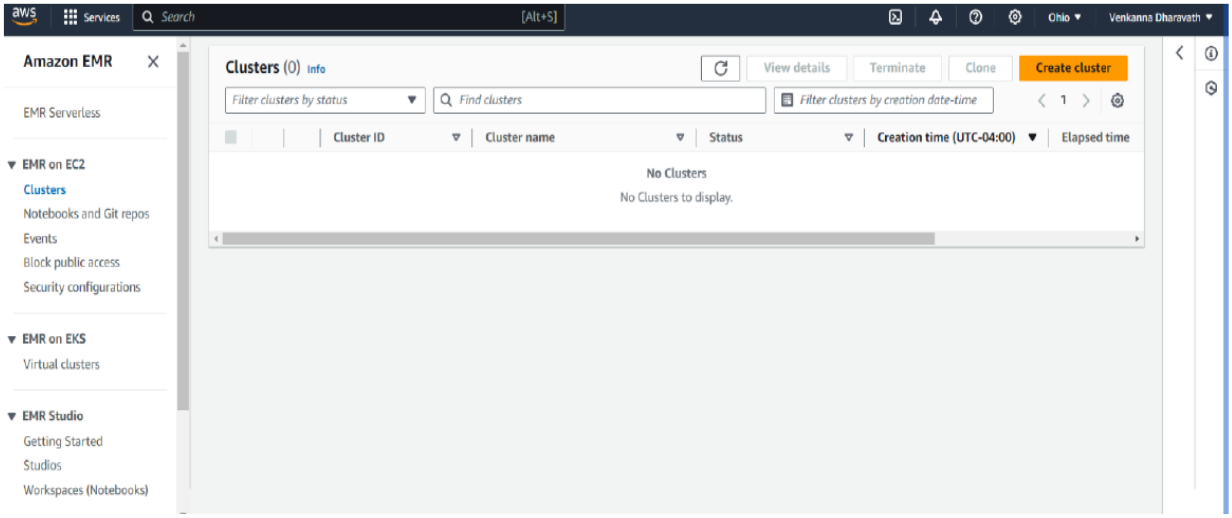
DockerHub - https://hub.docker.com/repository/docker/rr745/cs643-programming-assignment-2/general

Log into the AWS Management Console.

The AWS EMR service from the list of available services. Then, choose EMR on EC2 clusters.



On the Clusters page, you'll see that no clusters are currently in use. To start a new cluster, click the "Create Cluster" button.



You can name the cluster whatever you prefer.

Be sure to select **"Spark Interactive"** from the Application Bundle options and choose the latest available version of EMR.

The cluster is named wine_prediction_1112 with Amazon EMR release emr-7.5.0. **Spark Interactive** is selected as the application bundle, including Spark 3.5.2, Livy 0.8.0, and Jupyter Enterprise Gateway 2.6.0. Additional components like Hadoop and Hive are optional and can be enabled if needed. Review the configuration summary and click **Create Cluster** to finalize the setup.

Clone "My cluster" Info

▼ Name and applications - *required* Info

Name your cluster and choose the applications that you want to install to your cluster.

Name

wine_prediction_1112

Amazon EMR release Info

A release contains a set of applications which can be installed on your cluster.

emr-7.5.0

Application bundle

Spark Interactive | Core Hadoop | Flink | HBase | Presto | Trino | Custom

- [ ] AmazonCloudWatchAgent 1.300032.2
- [ ] Flink 1.19.1
- [ ] HBase 2.5.10
- [ ] HCatalog 3.1.3
- [x] Hadoop 3.4.0
- [x] Hive 3.1.3
- [ ] Hue 4.11.0
- [x] JupyterEnterpriseGateway 2.6.0
- [ ] JupyterHub 1.5.0
- [x] Livy 0.8.0
- [ ] Oozie 5.2.1
- [ ] Phoenix 5.2.0
- [ ] Pig 0.17.0
- [ ] Presto 0.287
- [x] Spark 3.5.2
- [ ] TensorFlow 2.16.1
- [ ] Tez 0.10.2
- [ ] Trino 446
- [ ] Zeppelin 0.11.1
- [ ] ZooKeeper 3.9.2

AWS Glue Data Catalog settings

Use the AWS Glue Data Catalog to provide an external metastore for your application.

- [ ] Use for Hive table metadata
- [ ] Use for Spark table metadata

Operating system options Info

- (•) Amazon Linux release
- ( ) Custom Amazon Machine Image (AMI)

Summary Info

Name and applications - *required*

Name
wine_prediction_1112

Amazon EMR release
emr-7.5.0

Application bundle
Spark Interactive (Hadoop 3.4.0, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5....)

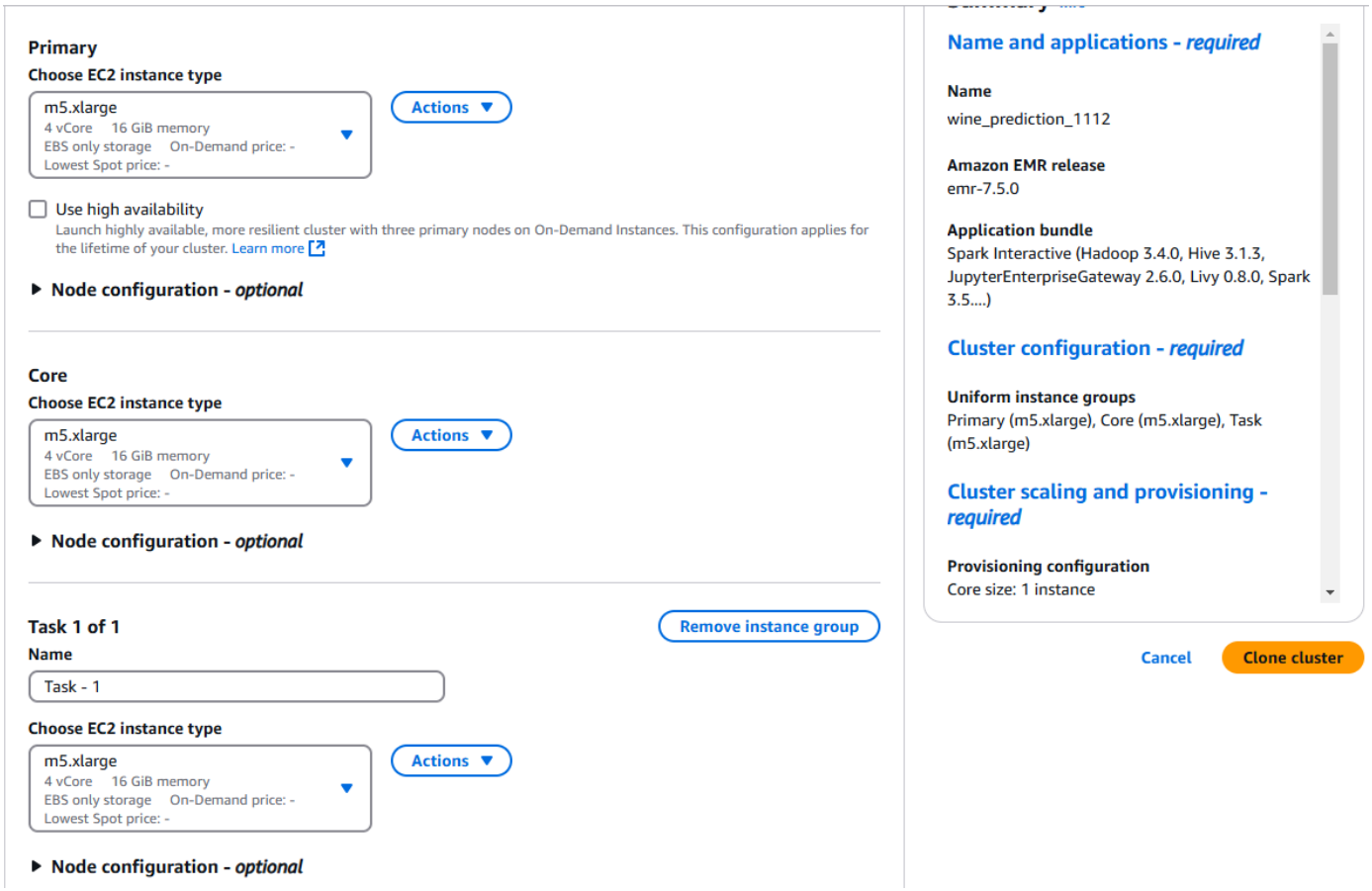
Cluster configuration - *required*

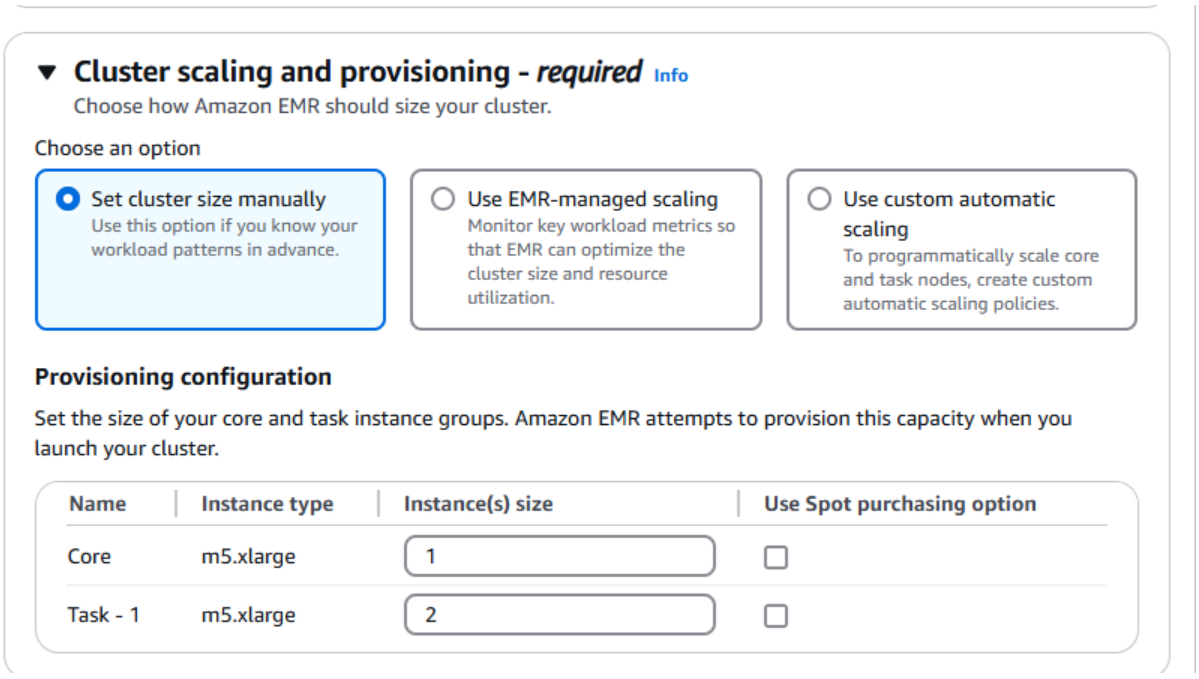
Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - *required*

Provisioning configuration
Core size: 1 instance

Cancel | Clone cluster


Primary
Choose EC2 instance type
m5.xlarge
4 vCore 16 GiB memory
EBS only storage On-Demand price: -
Lowest Spot price: -
Actions
Use high availability
Launch highly available, more resilient cluster with three primary nodes on On-Demand Instances. This configuration applies for the lifetime of your cluster. Learn more
Node configuration - optional
Core
Choose EC2 instance type
m5.xlarge
4 vCore 16 GiB memory
EBS only storage On-Demand price: -
Lowest Spot price: -
Actions
Node configuration - optional
Task 1 of 1
Remove instance group
Name
Task - 1
Choose EC2 instance type
m5.xlarge
4 vCore 16 GiB memory
EBS only storage On-Demand price: -
Lowest Spot price: -
Actions
Node configuration - optional
Name and applications - required
Name
wine_prediction_1112
Amazon EMR release
emr-7.5.0
Application bundle
Spark Interactive (Hadoop 3.4.0, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5....)
Cluster configuration - required
Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)
Cluster scaling and provisioning - required
Provisioning configuration
Core size: 1 instance
Cancel
Clone cluster


Cluster scaling and provisioning - required Info
Choose how Amazon EMR should size your cluster.
Choose an option
Set cluster size manually
Use this option if you know your workload patterns in advance.
Use EMR-managed scaling
Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.
Use custom automatic scaling
To programmatically scale core and task nodes, create custom automatic scaling policies.
Provisioning configuration
Set the size of your core and task instance groups. Amazon EMR attempts to provision this capacity when you launch your cluster.
Name
Instance type
Instance(s) size
Use Spot purchasing option
Core
m5.xlarge
1
Task - 1
m5.xlarge
2

Cluster termination and node replacement Info

Choose termination settings and protect your cluster from accidental shutdown.

Termination option

- (•) Manually terminate cluster
- ( ) Automatically terminate cluster after last step ends
- ( ) Automatically terminate cluster after idle time (Recommended)

[x] Use termination protection

Protects your cluster from accidental termination. If on, you must first turn off protection to terminate the cluster. We recommend turning on termination protection for your long running clusters.

Unhealthy node replacement - *new* | Info

- (•) Turn on

  Amazon EMR gracefully stops processes on unhealthy nodes to minimize data loss and job interruptions. It quickly replaces unhealthy nodes with new EC2 instances to keep your jobs running smoothly.

- ( ) Turn off

  Amazon EMR adds unhealthy nodes to a denylist while keeping them in the cluster, allowing you continued access for troubleshooting.

EC2 security groups (firewall)

Change notice

We've updated the names of some security groups to use more inclusive language. For example, groups that included terms like "master" and "slave" now use the terms "primary" and "core" instead.

Primary node

EMR-managed security group

EMR will automatically update the selected group.

ElasticMapReduce-Primary
sg-049e6f8a21365ae26

Additional security groups - *optional*

Select up to 4 additional security groups.

Choose additional security groups

Core and task nodes

EMR-managed security group

EMR will automatically update the selected group.

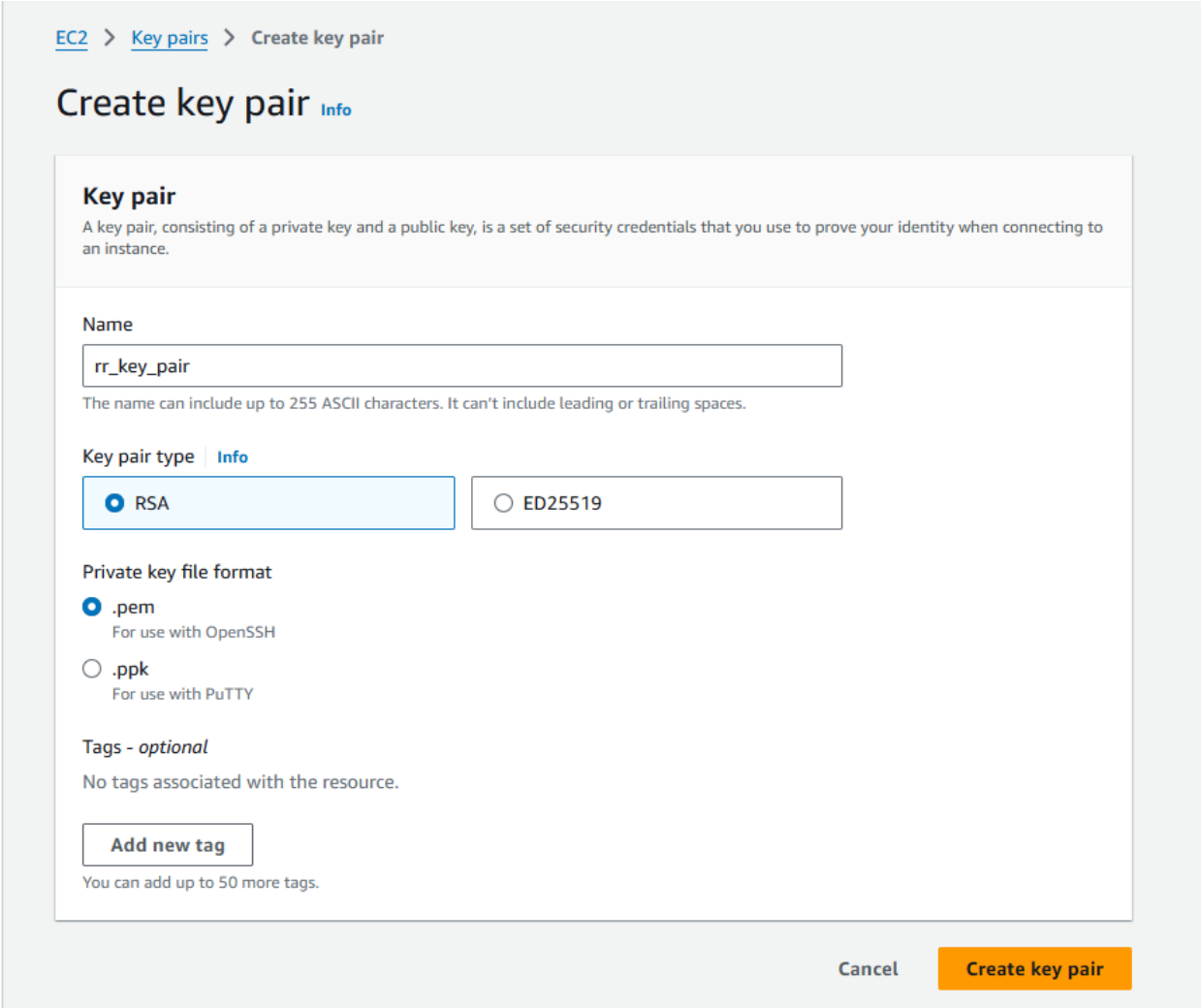
ElasticMapReduce-Core
sg-0246a15d4114d42f8

Additional security groups - *optional*

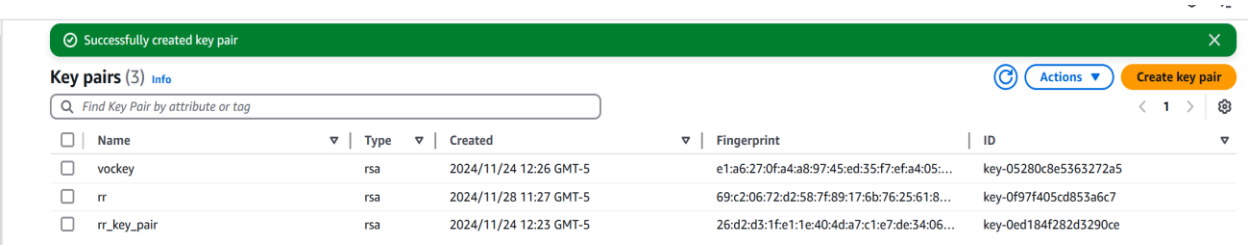
Select up to 4 additional security groups.

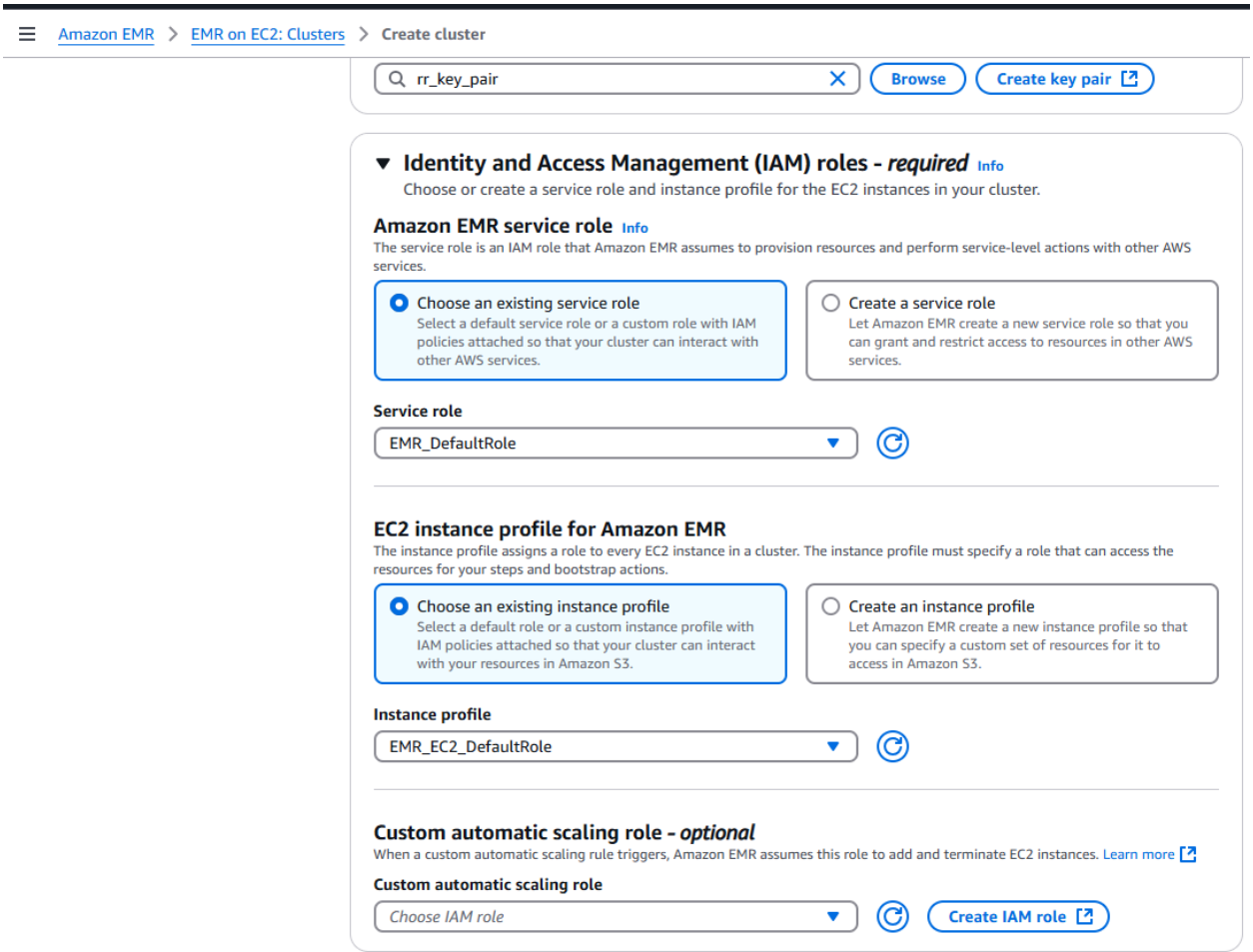
Choose additional security groups

Create a new key pair and provide a custom name for the key pair and select 'pem' as the file format when creating the key pair.

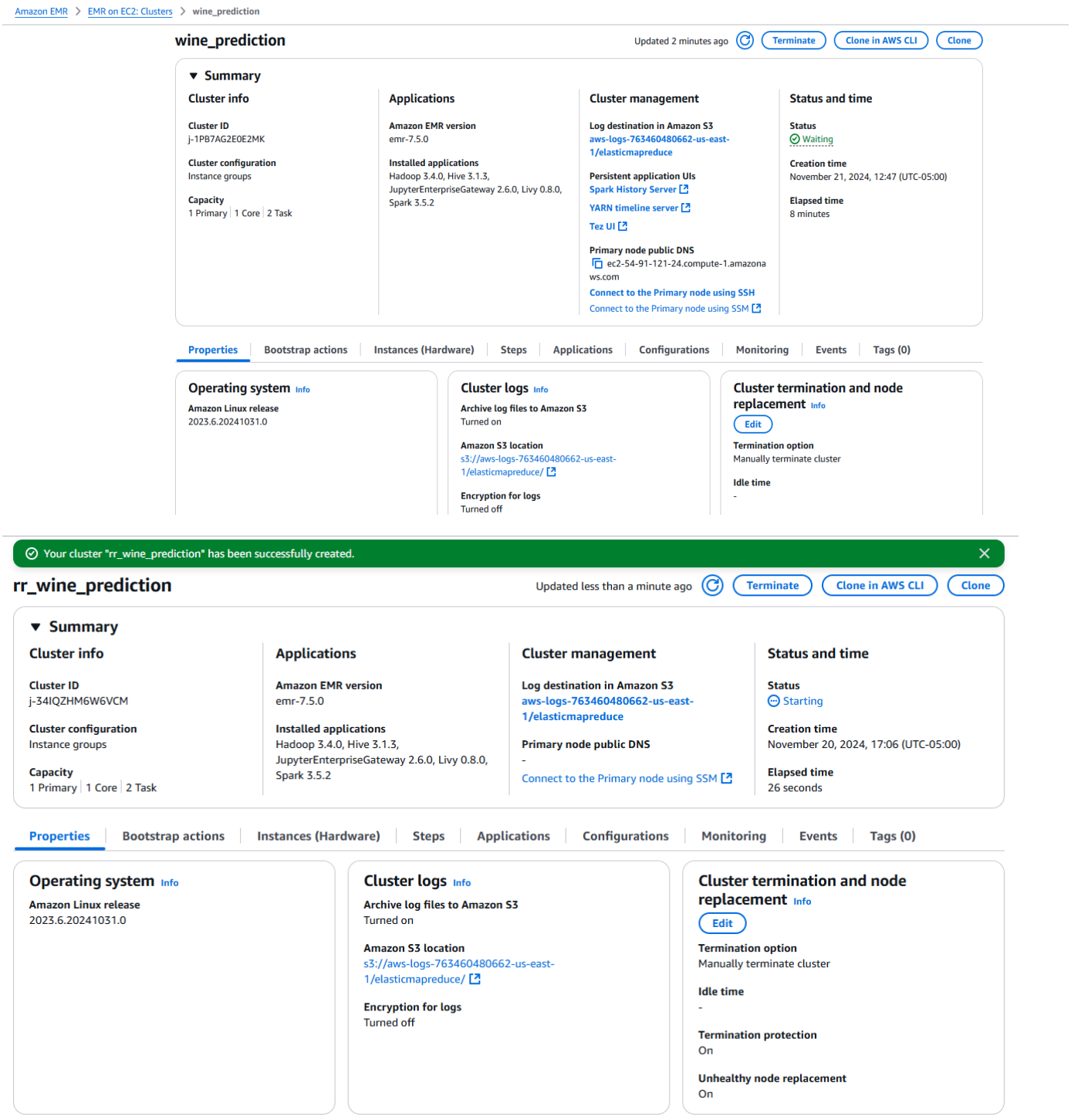


The key pair named 'rr_key_pair" has been created, and the key has been downloaded and stored on the local system for connecting to the cluster using PuTTY as the SSH.

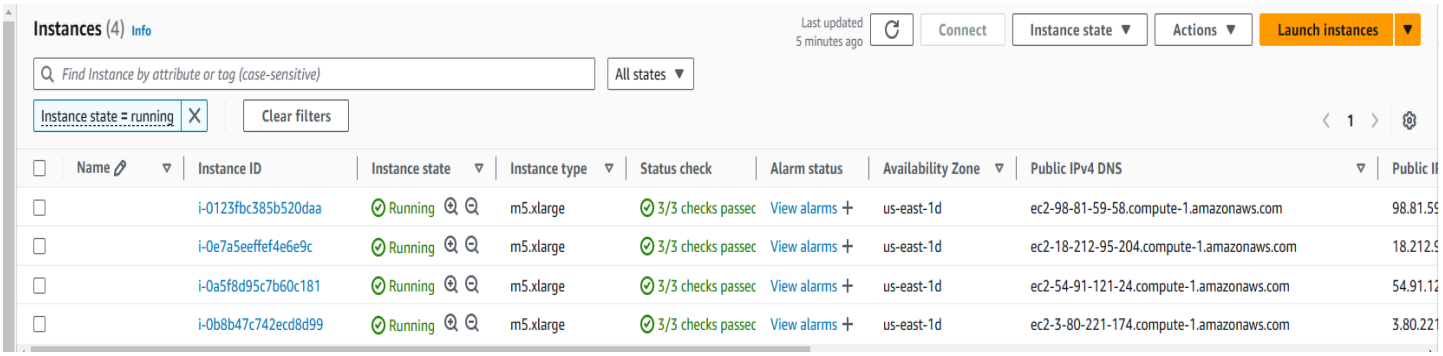




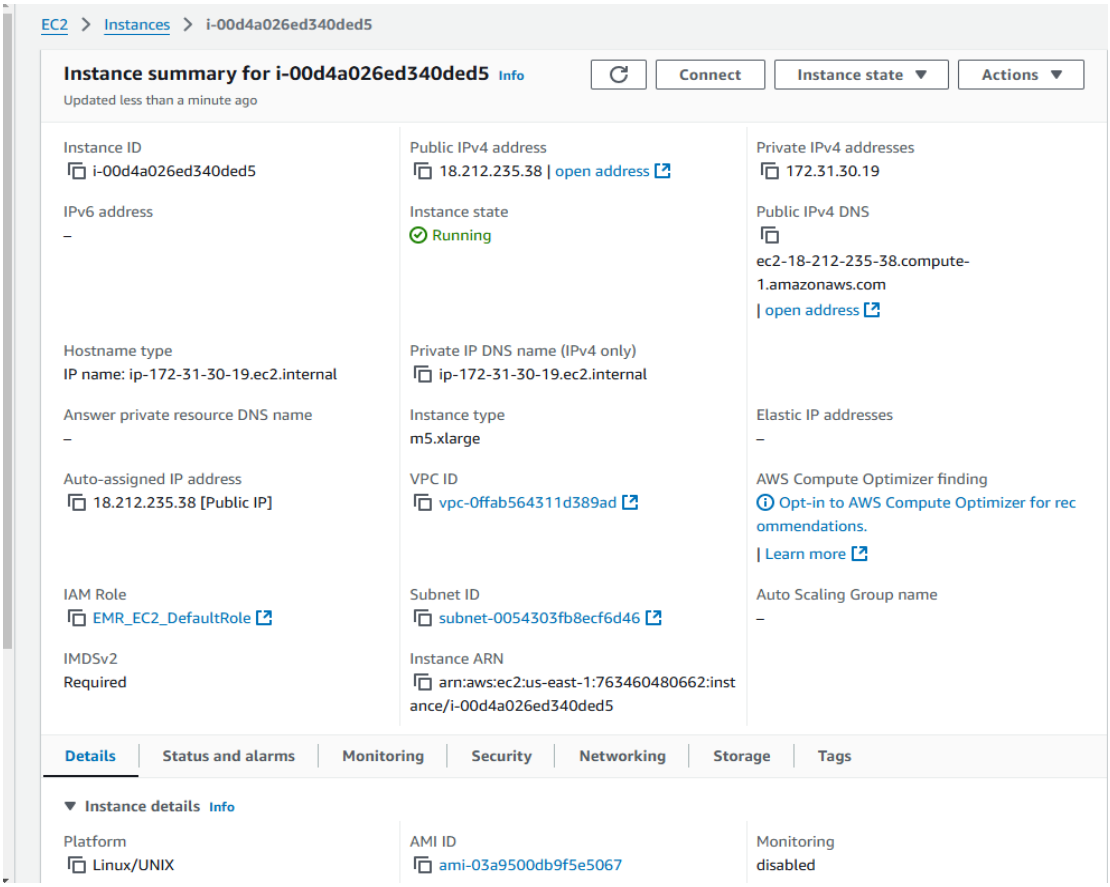
Click the button labeled "Create Cluster" to initiate the process of creating the cluster. Cluster has been created successfully.



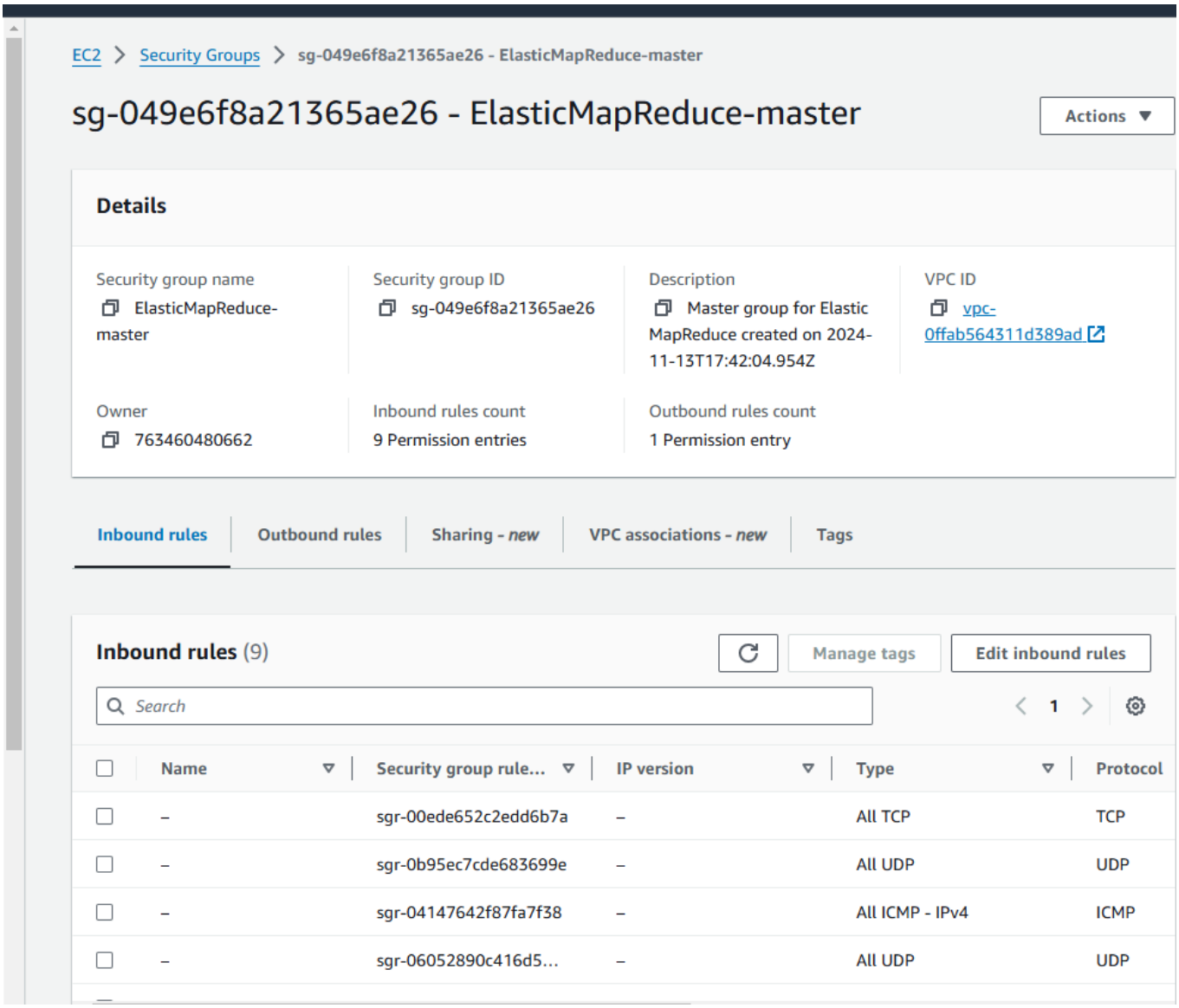
Go to the EC2 Instances page. As shown below, there are four EC2 instances running, one designated as the Master node and the other three as Slave nodes.

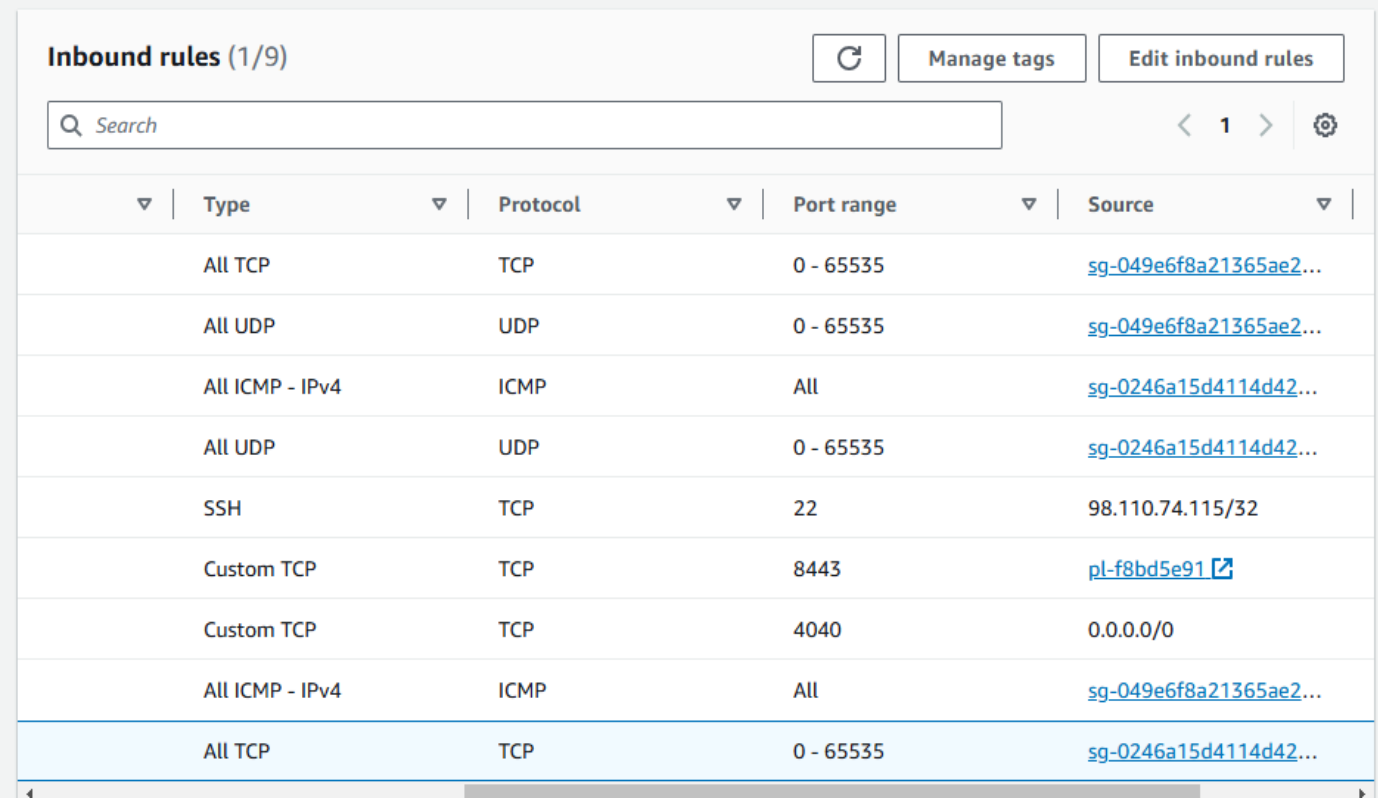


To access the "ElasticMapReduce-Master" security group, go to the EC2 service and select the corresponding Security Group ID. Click on the security group to edit its inbound rules.



Inbound rules section, select ‘Edit inbound rules’ and enter the port numbers 22 and 4040 with the settings indicated below, then click the add rule button and save the rules.





Next step is to create a S3 bucket in AWS services to store the training & validation datasets.

Choose "Create Bucket" - Labeled the bucket as "rr-programming-assignment-2". Scroll down and click on the "Create bucket" button.

Create bucket Info

Buckets are containers for data stored in S3.

General configuration

AWS Region

US East (N. Virginia) us-east-1

Bucket type Info

- (•) General purpose
  Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.
- ( ) Directory
  Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name Info

rr-programming-assignment-2

Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Choose bucket

Format: s3://bucket/prefix

Object Ownership Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

- (•) ACLs disabled (recommended)
  All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.
- ( ) ACLs enabled
  Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

- [x] **Block *all* public access**
  Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.
  - [x] **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
    S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
  - [x] **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
    S3 will ignore all ACLs that grant public access to buckets and objects.
  - [x] **Block public access to buckets and objects granted through *new* public bucket or access point policies**
    S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
  - [x] **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**
    S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. Learn more

Bucket Versioning

- (•) Disable
- ( ) Enable

Tags - *optional* (0)

You can use bucket tags to track storage costs and organize buckets. Learn more

No tags associated with this bucket.

Tags - *optional* (0)

You can use bucket tags to track storage costs and organize buckets. Learn more

No tags associated with this bucket.

Add tag

Default encryption Info

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type Info

- (•) Server-side encryption with Amazon S3 managed keys (SSE-S3)
- ( ) Server-side encryption with AWS Key Management Service keys (SSE-KMS)
- ( ) Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)
  Secure your objects with two separate layers of encryption. For details on pricing, see **DSSE-KMS pricing** on the **Storage** tab of the Amazon S3 pricing page.

Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. Learn more

- ( ) Disable
- (•) Enable

▸ Advanced settings

ⓘ After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

Cancel Create bucket

☰ Amazon S3 › Buckets

▸ Account snapshot - *updated every 24 hours* All AWS Regions

Storage lens provides visibility into storage usage and activity trends. Learn more

View Storage Lens dashboard

General purpose buckets | Directory buckets

General purpose buckets (2) Info All AWS Regions

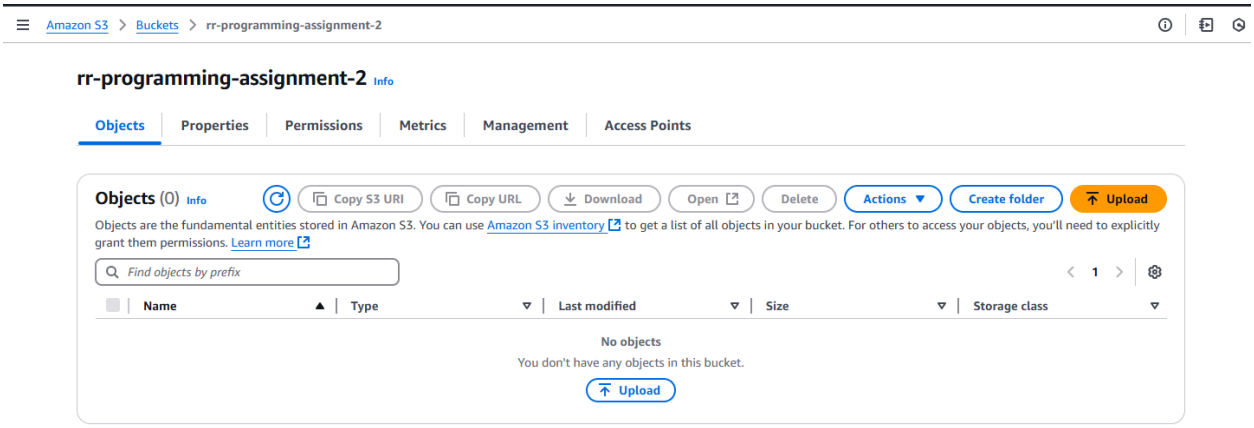
Buckets are containers for data stored in S3.

Copy ARN | Empty | Delete | Create bucket

Find buckets by name

‹ 1 ›

	Name	AWS Region	IAM Access Analyzer	Creation date
( )	aws-logs-763460480662-us-east-1	US East (N. Virginia) us-east-1	View analyzer for us-east-1	November 14, 2024, 13:58:43 (UTC-05:00)
( )	rr-programming-assignment-2	US East (N. Virginia) us-east-1	View analyzer for us-east-1	November 20, 2024, 18:07:45 (UTC-05:00)



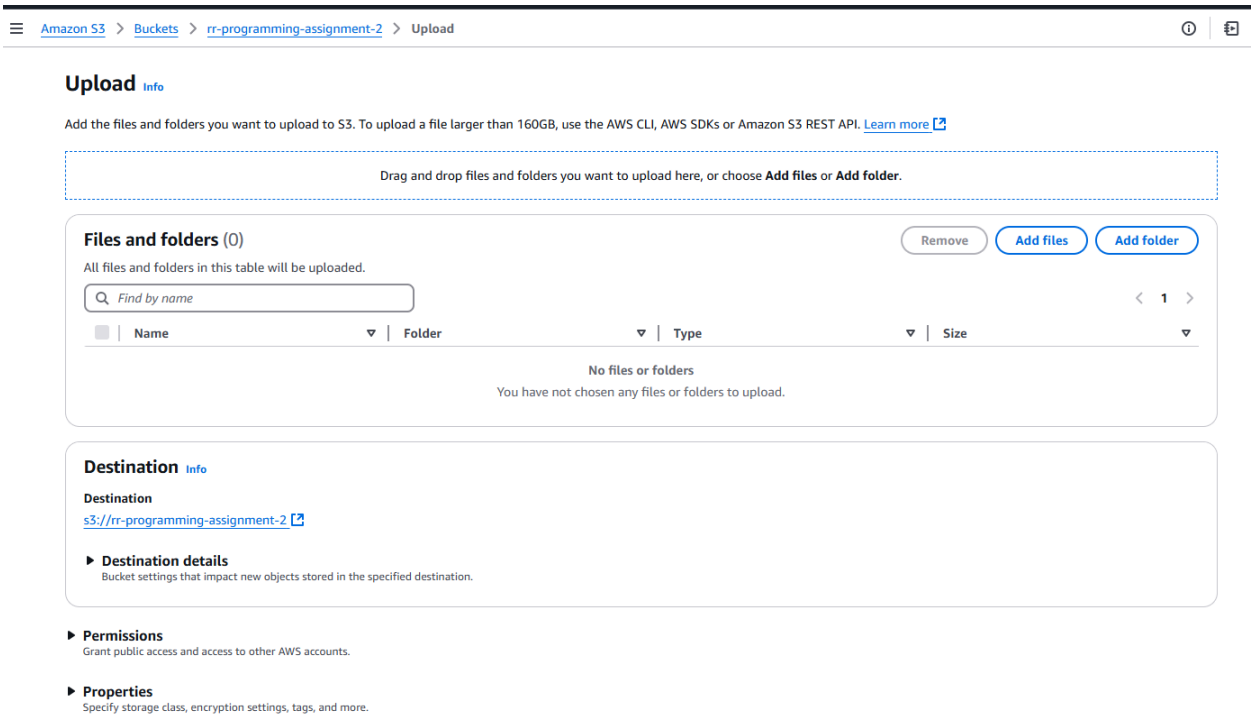
Details | Status and alarms | Monitoring | Security | Networking | Storage | Tags

Instance details Info

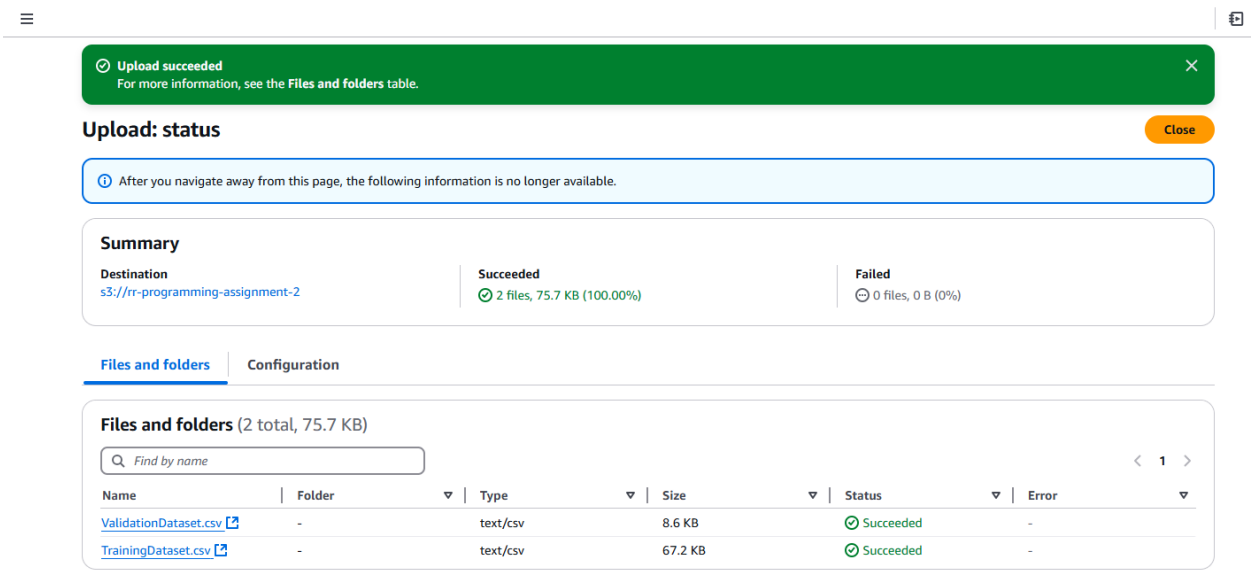
Platform Linux/UNIX	AMI ID ami-03a9500db9f5e5067	Monitoring disabled
Platform details Linux/UNIX	AMI name emr-7_5_0-x86_64-2023_6_20241031_0-Hadoop_Hive_Spark-2024-11-14T05-56-27.240Z	Termination protection Enabled
Stop protection Disabled	Launch time Wed Nov 20 2024 17:06:50 GMT-0500 (Eastern Standard Time) (38 minutes)	AMI location amazon/emr-7_5_0-x86_64-2023_6_20241031_0-Hadoop_Hive_Spark-2024-11-14T05-56-27.240Z
Instance auto-recovery Default	Lifecycle normal	Stop-hibernate behavior Disabled
AMI Launch index 0	Key pair assigned at launch rr_key_pair	State transition reason –
Credit specification Not supported by instance type	Kernel ID –	State transition message –
Usage operation RunInstances	RAM disk ID –	Owner 763460480662
Enclaves Support Disabled	Boot mode uefi-preferred	Current instance boot mode uefi
Allow tags in instance metadata Disabled	Use RBN as guest OS hostname Disabled	Answer RBN DNS hostname IPv4 Disabled

Host and placement group Info

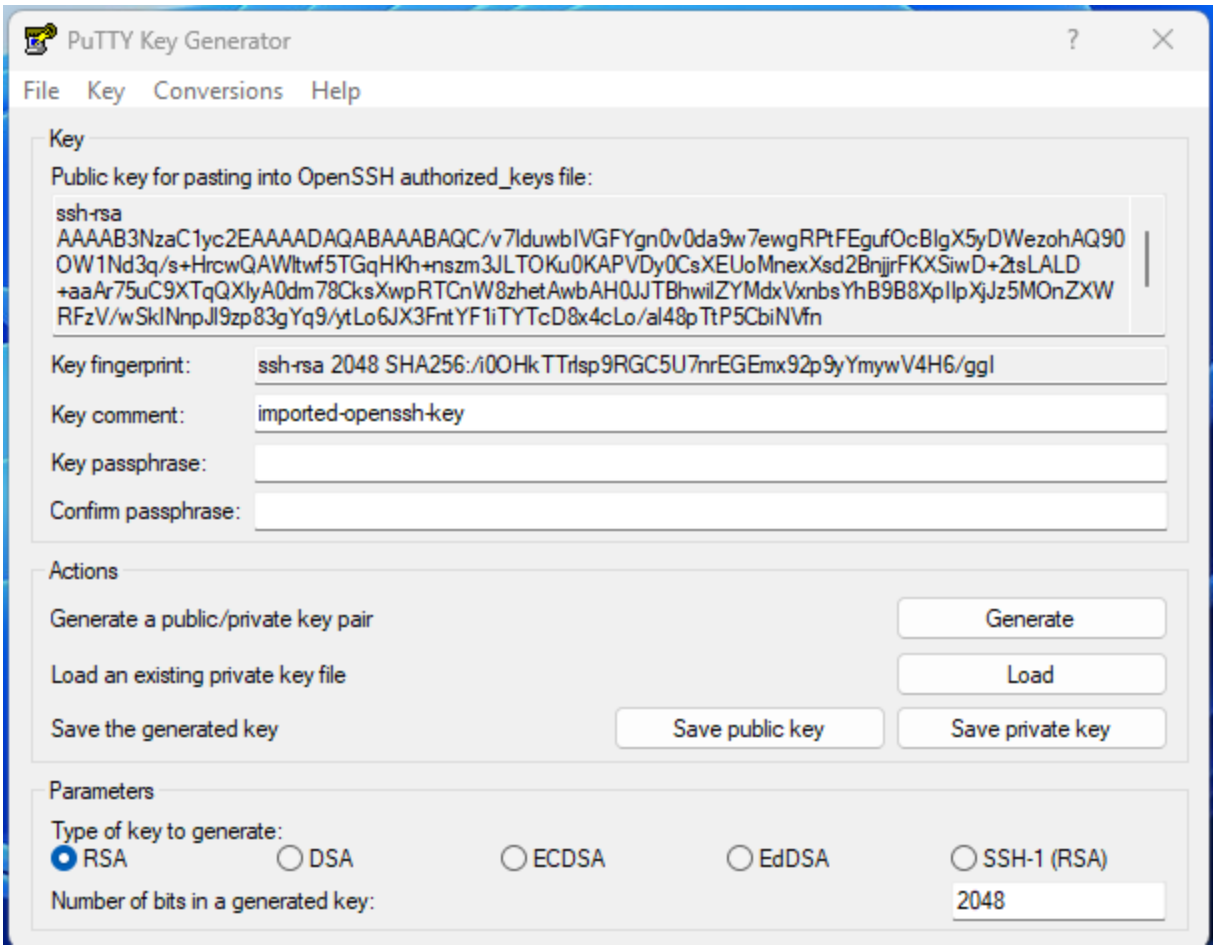
Host ID –	Affinity –	Placement group –



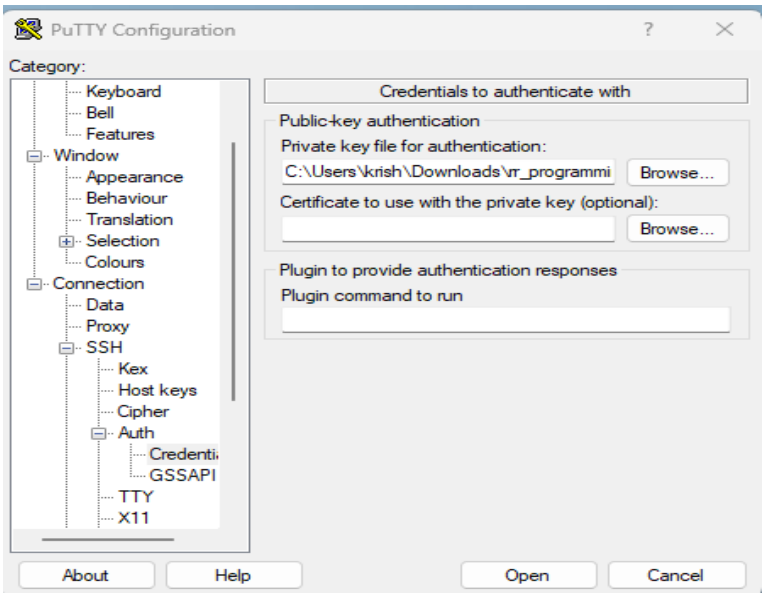
Click on **"Add Files"** to select the .csv files from your local system. Then, click the **"Upload"** button to upload the datasets to the S3 bucket. Once uploaded, your S3 bucket will contain two .csv files: ValidationDataset.csv and TrainingDataset.csv.

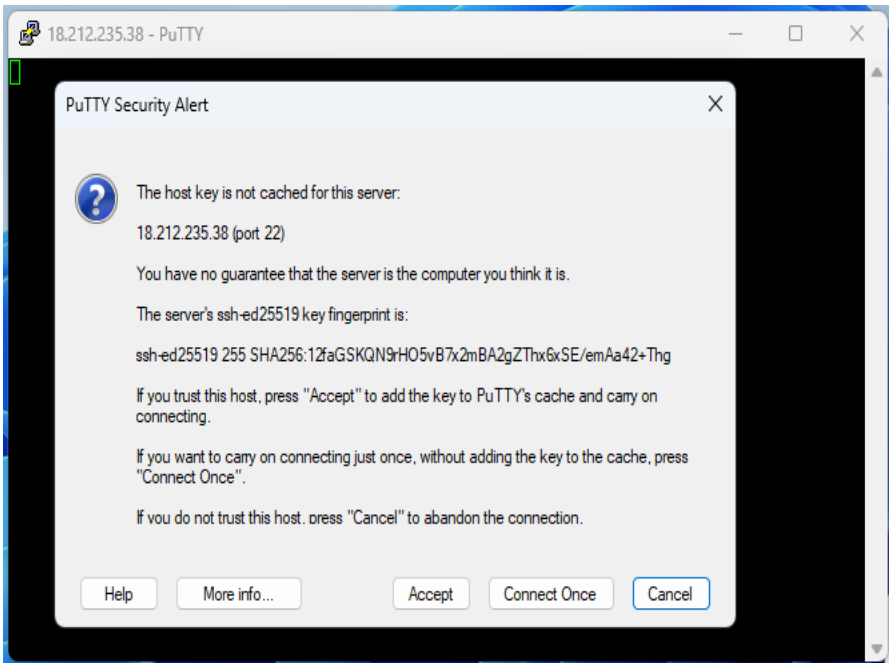


Convert pem to ppk file using puttygen



Click on "SSH" under "Auth", select "Credentials", provide the path to the location of the PPK file, and then click "Open".


PuTTY Configuration
Category:
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
SSH
Kex
Host keys
Cipher
Auth
Credenti
GSSAPI
TTY
X11
Credentials to authenticate with
Public-key authentication
Private key file for authentication:
C:\Users\krish\Downloads\rr_programmi
Browse...
Certificate to use with the private key (optional):
Browse...
Plugin to provide authentication responses
Plugin command to run
About
Help
Open
Cancel

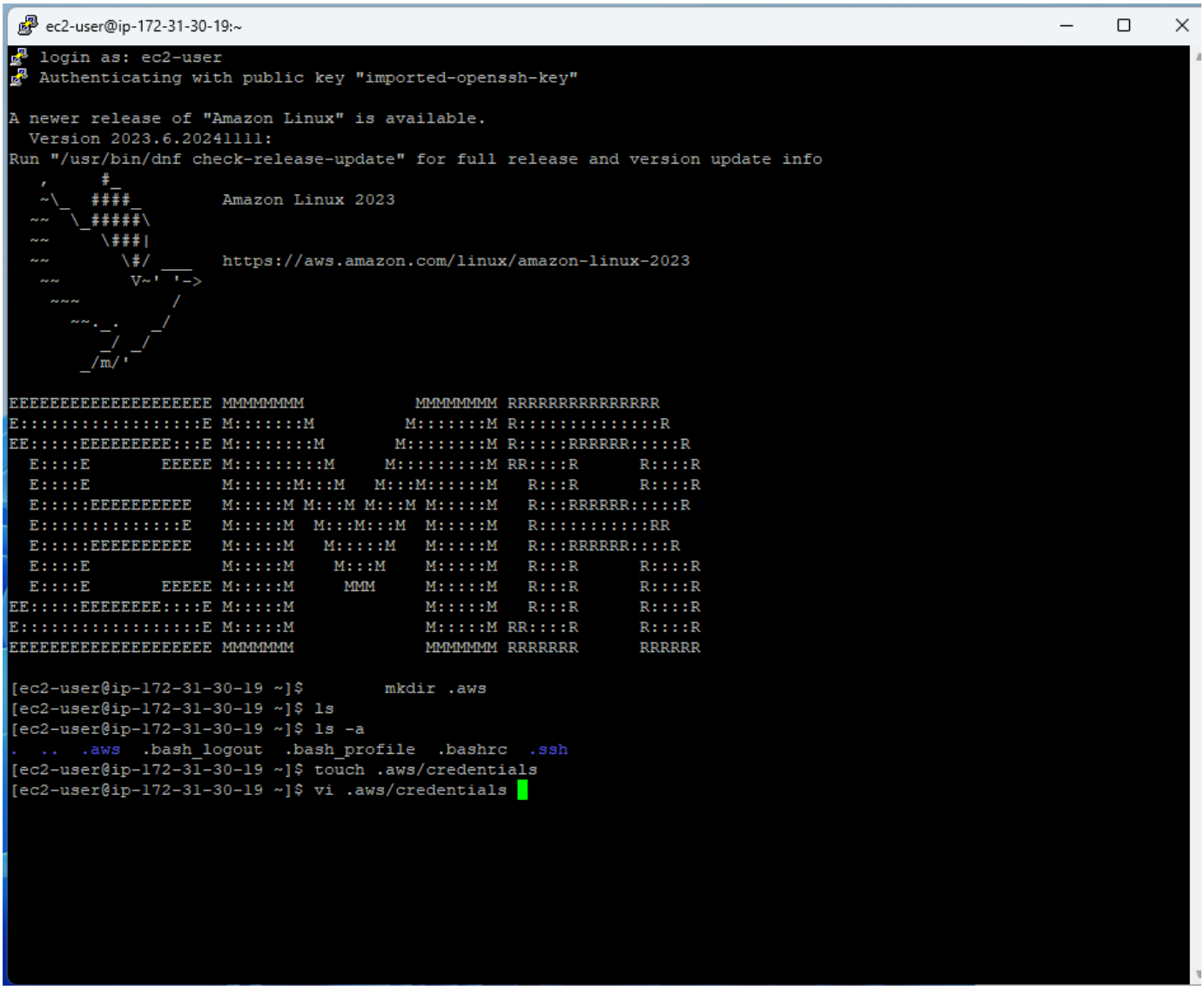

18.212.235.38 - PuTTY
PuTTY Security Alert
The host key is not cached for this server:
18.212.235.38 (port 22)
You have no guarantee that the server is the computer you think it is.
The server's ssh-ed25519 key fingerprint is:
ssh-ed25519 255 SHA256:12faGSKQN9rHO5vB7x2mBA2gZThx6xSE/emAa42+Thg
If you trust this host, press "Accept" to add the key to PuTTY's cache and carry on connecting.
If you want to carry on connecting just once, without adding the key to the cache, press "Connect Once".
If you do not trust this host, press "Cancel" to abandon the connection.
Help
More info...
Accept
Connect Once
Cancel

Set up the credentials for the Master node in the EC2 instance by running the following commands in the terminal:

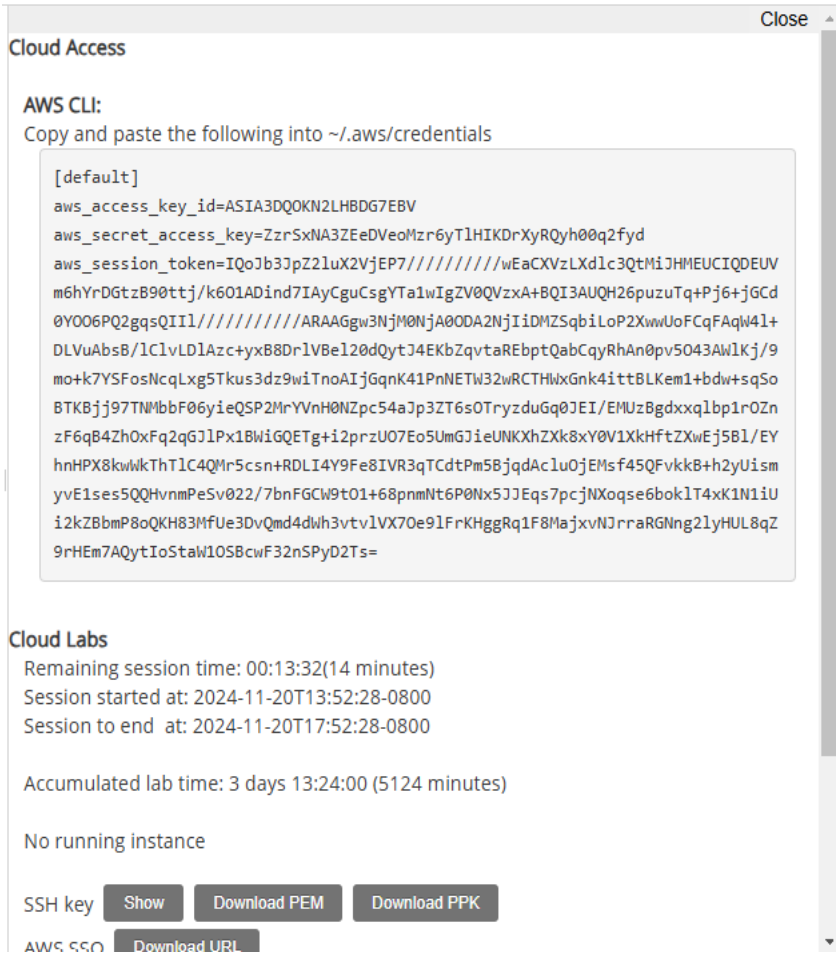
mkdir .aws

touch .aws/credentials

vi .aws/credentials



Copy and paste the credentials from the AWS Academy page



Install Required Packages. Run the following commands to update the ec2 instance with necessary packages:

sudo yum update

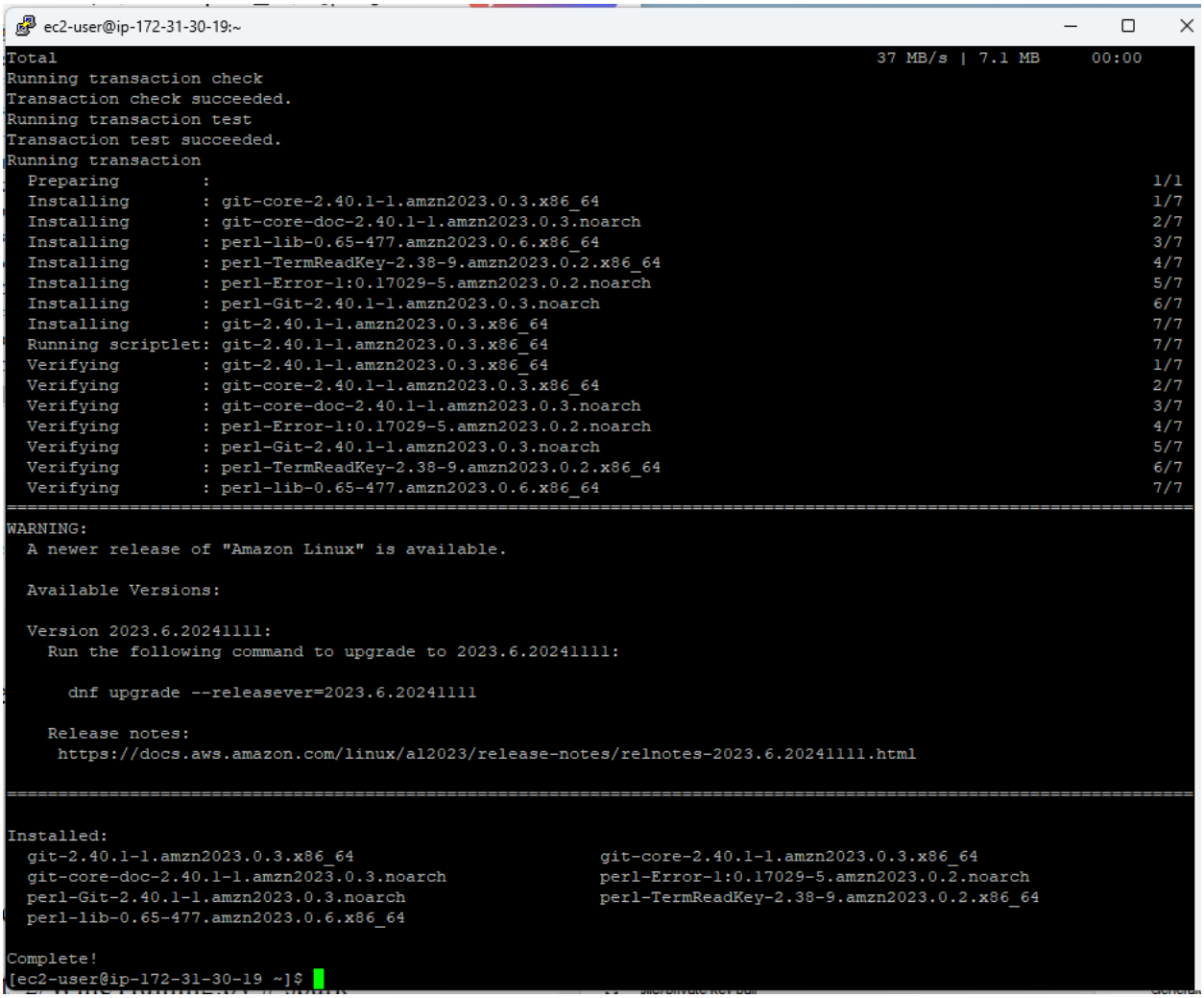
sudo yum install git

pip install pyspark findspark boto3 numpy pandas scikit-learn datetime

To clone the GitHub repository, enter the following command:

#git clone https://github.com/rr745/wine_quality_predictor

```
spark-submit \
  --master yarn \
  --deploy-mode cluster \
  WinePredictor_Training.py
```



```
|"""""fixed acidity""""|"""""volatile acidity""""|""""citric acid""""|""""residual sugar""""|""""chlorides""""|""""free sulfur dioxide""""|""""total
sulfur dioxide""""|""""density""""|""""pH""""|""""sulphates""""|""""alcohol""""|""""quality""""|
+---------------------+----------------------+------------------+--------------------+----------------+--------------------------+----------
----------------+--------------+----------+----------------+--------------+---------------+
|                  8.9|                  0.22|              0.48|                 1.8|           0.077|                        29|
               60|        0.9968|      3.39|            0.53|           9.4|              6|
|                  7.6|                  0.39|              0.31|                 2.3|           0.082|                        23|
               71|        0.9982|      3.52|            0.65|           9.7|              5|
|                  7.9|                  0.43|              0.21|                 1.6|           0.106|                        10|
               37|        0.9966|      3.17|            0.91|           9.5|              5|
|                  8.5|                  0.49|              0.11|                 2.3|           0.084|                         9|
               67|        0.9968|      3.17|            0.53|           9.4|              5|
|                  6.9|                   0.4|              0.14|                 2.4|           0.085|                        21|
               40|        0.9968|      3.43|            0.63|           9.7|              6|
|                  6.3|                  0.39|              0.16|                 1.4|            0.08|                        11|
               23|        0.9955|      3.34|            0.56|           9.3|              5|
|                  7.6|                  0.41|              0.24|                 1.8|            0.08|                         4|
               11|        0.9962|      3.28|            0.59|           9.5|              5|
|                  7.9|                  0.43|              0.21|                 1.6|           0.106|                        10|
               37|        0.9966|      3.17|            0.91|           9.5|              5|
|                  7.1|                  0.71|                 0|                 1.9|            0.08|                        14|
               35|        0.9972|      3.47|            0.55|           9.4|              5|
|                  7.8|                 0.645|                 0|                   2|           0.082|                         8|
               16|        0.9964|      3.38|            0.59|           9.8|              6|
|                  6.7|                 0.675|              0.07|                 2.4|           0.089|                        17|
               82|        0.9958|      3.35|            0.54|          10.1|              5|
|                  6.9|                 0.685|                 0|                 2.5|           0.105|                        22|
               37|        0.9966|      3.46|            0.57|          10.6|              6|
|                  8.3|                 0.655|              0.12|                 2.3|           0.083|                        15|
              113|        0.9966|      3.17|            0.66|           9.8|              5|
|                  6.9|                 0.605|              0.12|                10.7|           0.073|                        40|
               83|        0.9993|      3.45|            0.52|           9.4|              6|
|                  5.2|                  0.32|              0.25|                 1.8|           0.103|                        13|
               50|        0.9957|      3.38|            0.55|           9.2|              5|
|                  7.8|                 0.645|                 0|                 5.5|           0.086|                         5|
               18|        0.9986|       3.4|            0.55|           9.6|              6|
|                  7.8|                   0.6|              0.14|                 2.4|           0.086|                         3|
               15|        0.9975|      3.42|             0.6|          10.8|              6|
|                  8.1|                  0.38|              0.28|                 2.1|           0.066|                        13|
               30|        0.9968|      3.23|            0.73|           9.7|              7|
|                  5.7|                  1.13|              0.09|                 1.5|           0.172|                         7|
               19|         0.994|       3.5|            0.48|           9.8|              4|
|                  7.3|                  0.45|              0.36|                 5.9|           0.074|                        12|
               87|        0.9978|      3.33|            0.83|          10.5|              5|
+---------------------+----------------------+------------------+--------------------+----------------+--------------------------+----------
----------------+--------------+----------+----------------+--------------+---------------+
only showing top 20 rows

Training DecisionTree model...
Model - DecisionTree Created
s3a://rr-programming-assignment-2/models/model_dt.model
model_dt.model
False
>>>>> DecisionTree model saved
Training RandomForest model...
Model - Randomforest Created
s3a://rr-programming-assignment-2/models/model_rf.model
model_rf.model
False
Model for Random Forest algorithm
```

ec2-user@ip-172-31-24-82:~/wine_quality_predictor

```
ulfur dioxide"""|"""density"""|"""pH"""|"""sulphates"""|"""alcohol"""|"""quality"""|
+--------------------+----------------------+------------------+--------------------+----------------+--------------------------+-----------
----------------+--------------+----------+----------------+--------------+---------------+
|                8.9|                 0.22|              0.48|                 1.8|           0.077|                        29|
            60|        0.9968|      3.39|            0.53|           9.4|              6|
|                7.6|                 0.39|              0.31|                 2.3|           0.082|                        23|
            71|        0.9982|      3.52|            0.65|           9.7|              5|
|                7.9|                 0.43|              0.21|                 1.6|           0.106|                        10|
            37|        0.9966|      3.17|            0.91|           9.5|              5|
|                8.5|                 0.49|              0.11|                 2.3|           0.084|                         9|
            67|        0.9968|      3.17|            0.53|           9.4|              5|
|                6.9|                  0.4|              0.14|                 2.4|           0.085|                        21|
            40|        0.9968|      3.43|            0.63|           9.7|              6|
|                6.3|                 0.39|              0.16|                 1.4|            0.08|                        11|
            23|        0.9955|      3.34|            0.56|           9.3|              5|
|                7.6|                 0.41|              0.24|                 1.8|            0.08|                         4|
            11|        0.9962|      3.28|            0.59|           9.5|              5|
|                7.9|                 0.43|              0.21|                 1.6|           0.106|                        10|
            37|        0.9966|      3.17|            0.91|           9.5|              5|
|                7.1|                 0.71|                 0|                 1.9|            0.08|                        14|
            35|        0.9972|      3.47|            0.55|           9.4|              5|
|                7.8|                0.645|                 0|                   2|           0.082|                         8|
            16|        0.9964|      3.38|            0.59|           9.8|              6|
|                6.7|                0.675|              0.07|                 2.4|           0.089|                        17|
            82|        0.9958|      3.35|            0.54|          10.1|              5|
|                6.9|                0.685|                 0|                 2.5|           0.105|                        22|
            37|        0.9966|      3.46|            0.57|          10.6|              6|
|                8.3|                0.655|              0.12|                 2.3|           0.083|                        15|
           113|        0.9966|      3.17|            0.66|           9.8|              5|
|                6.9|                0.605|              0.12|                10.7|           0.073|                        40|
            83|        0.9993|      3.45|            0.52|           9.4|              6|
|                5.2|                 0.32|              0.25|                 1.8|           0.103|                        13|
            50|        0.9957|      3.38|            0.55|           9.2|              5|
|                7.8|                0.645|                 0|                 5.5|           0.086|                         5|
            18|        0.9986|       3.4|            0.55|           9.6|              6|
|                7.8|                  0.6|              0.14|                 2.4|           0.086|                         3|
            15|        0.9975|      3.42|             0.6|          10.8|              6|
|                8.1|                 0.38|              0.28|                 2.1|           0.066|                        13|
            30|        0.9968|      3.23|            0.73|           9.7|              7|
|                5.7|                 1.13|              0.09|                 1.5|           0.172|                         7|
            19|         0.994|       3.5|            0.48|           9.8|              4|
|                7.3|                 0.45|              0.36|                 5.9|           0.074|                        12|
            87|        0.9978|      3.33|            0.83|          10.5|              5|
+--------------------+----------------------+------------------+--------------------+----------------+--------------------------+-----------
----------------+--------------+----------+----------------+--------------+---------------+
only showing top 20 rows

Training DecisionTree model...
Model - DecisionTree Created
s3a://rr-programming-assignment-2/models/model_dt.model
model_dt.model
False
>>>>> DecisionTree model saved
Training RandomForest model...
Model - Randomforest Created
s3a://rr-programming-assignment-2/models/model_rf.model
model_rf.model
True
>>>> Prexisting Folder Deleted: models/model_rf.model
Model for Random Forest algorithm
```

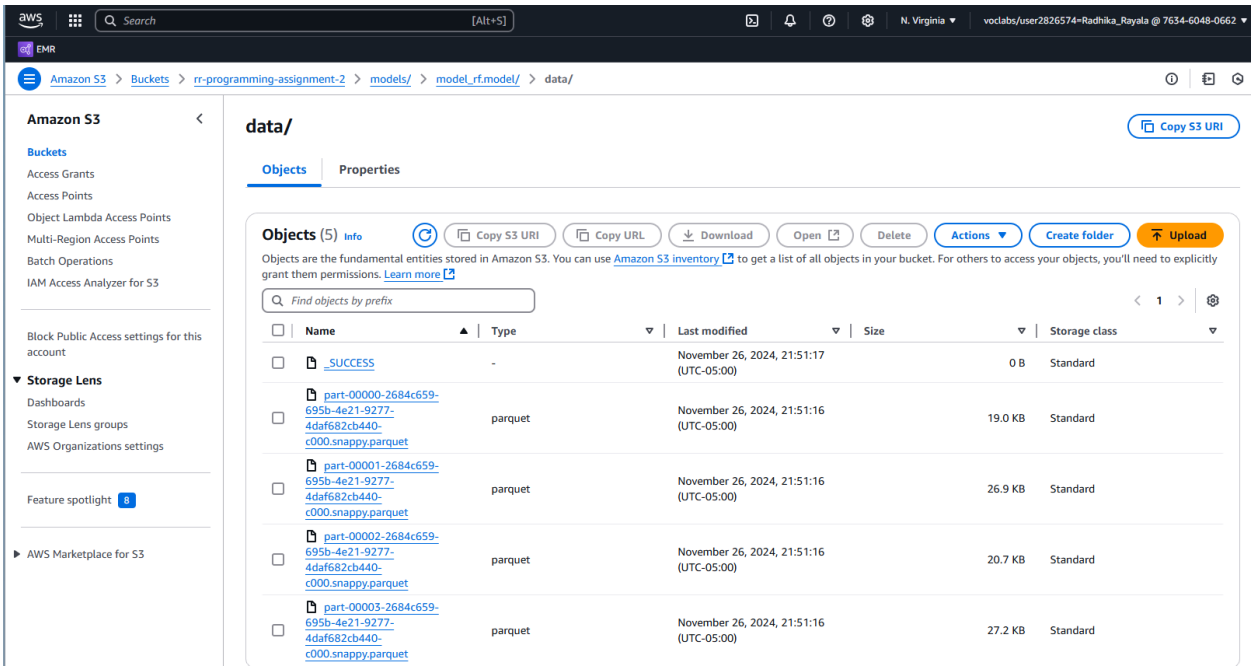
The results will be displayed here, along with the Accuracy and F1 scores of the Machine Learning methods used.

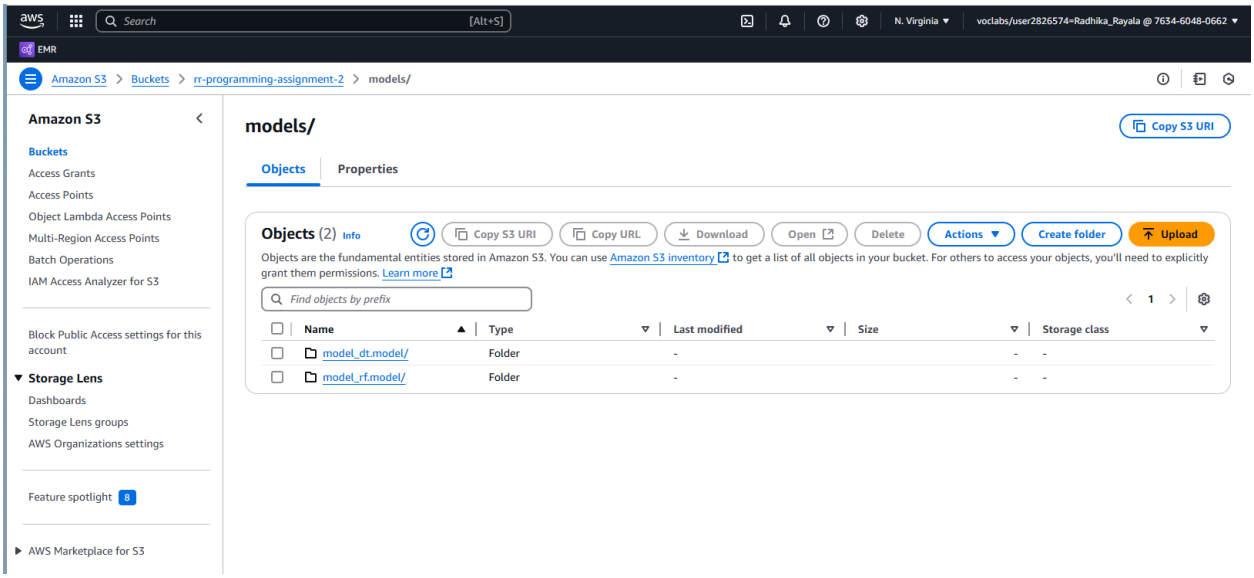
```
Training DataSet Metrics
Accuracy: 0.6114151681000782
F-measure: 0.5904050519731796
+-------------+----------------+-----------+-------------+---------+--------------------+---------------------+-------+----+---------+-------+-----+
|fixed_acidity|volatile_acidity|citric_acid|residual_sugar|chlorides|free_sulfur_dioxide|total_sulfur_dioxide|density|  pH|sulphates|alcohol|label|
+-------------+----------------+-----------+-------------+---------+--------------------+---------------------+-------+----+---------+-------+-----+
|         7.4|             0.7|        0.0|          1.9|    0.076|                11|                   34| 0.9978|3.51|     0.56|    9.4|  5.0|
|         7.8|            0.88|        0.0|          2.6|    0.098|                25|                   67| 0.9968| 3.2|     0.68|    9.8|  5.0|
|         7.8|            0.76|       0.04|          2.3|    0.092|                15|                   54|  0.997|3.26|     0.65|    9.8|  5.0|
|        11.2|            0.28|       0.56|          1.9|    0.075|                17|                   60|  0.998|3.16|     0.58|    9.8|  6.0|
|         7.4|             0.7|        0.0|          1.9|    0.076|                11|                   34| 0.9978|3.51|     0.56|    9.4|  5.0|
+-------------+----------------+-----------+-------------+---------+--------------------+---------------------+-------+----+---------+-------+-----+
only showing top 5 rows


 Validation Training Set Metrics
+--------------------------------------------------+-----+----------+
|features                                          |label|prediction|
+--------------------------------------------------+-----+----------+
|[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]  |5.0  |5.0       |
|[9.8,0.68,3.2,0.9968,25.0,67.0,0.098,2.6,0.0,0.88,7.8]  |5.0  |5.0       |
|[9.8,0.65,3.26,0.997,15.0,54.0,0.092,2.3,0.04,0.76,7.8] |5.0  |5.0       |
|[9.8,0.58,3.16,0.998,17.0,60.0,0.075,1.9,0.56,0.28,11.2]|6.0  |5.0       |
|[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]  |5.0  |5.0       |
+--------------------------------------------------+-----+----------+
only showing top 5 rows


The accuracy of the model is 0.575
F1: 0.5619407071339173
```

Saved models





<u>DOCKER IMPLEMENTATION –</u>

1.Update System Packages:

sudo yum update -y

2.Install Docker:

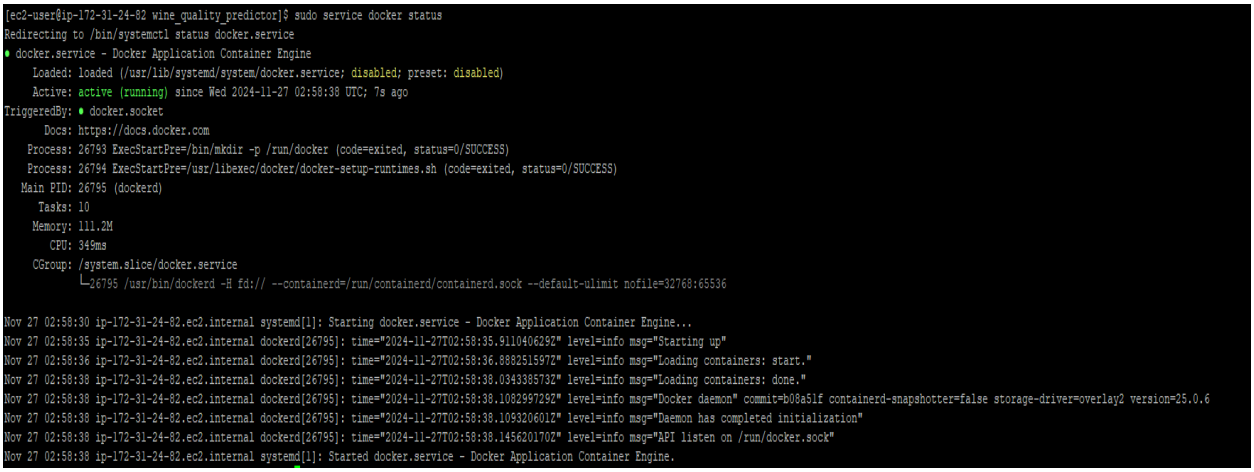
sudo yum install -y docker

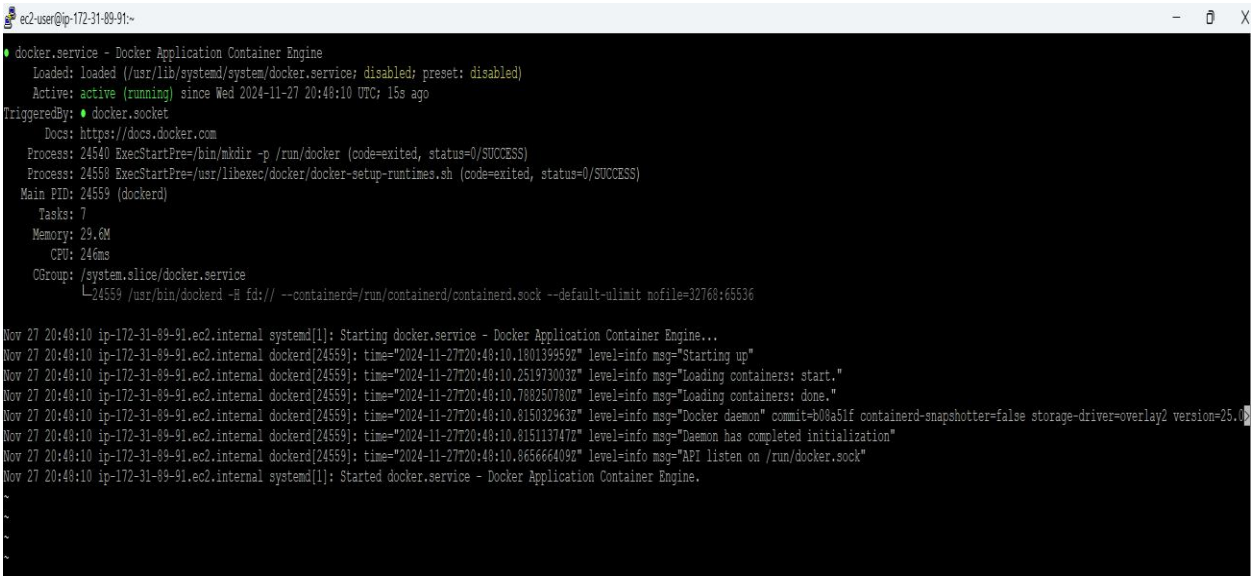
3.Start Docker Service:

sudo service docker start

4.Check Docker Service Status:

sudo service docker status



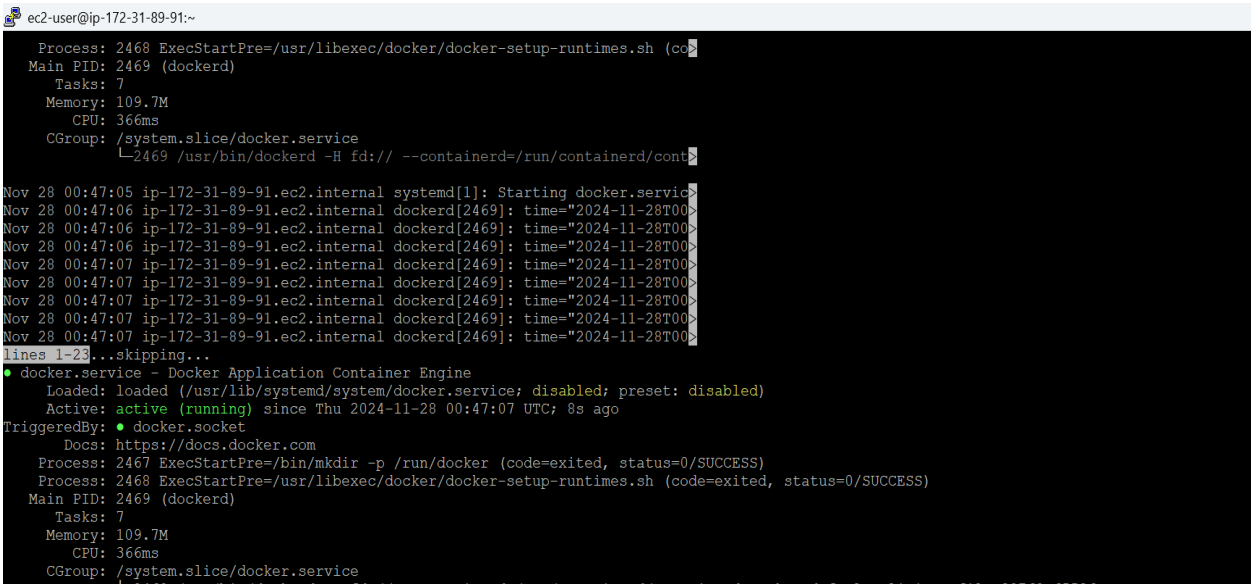


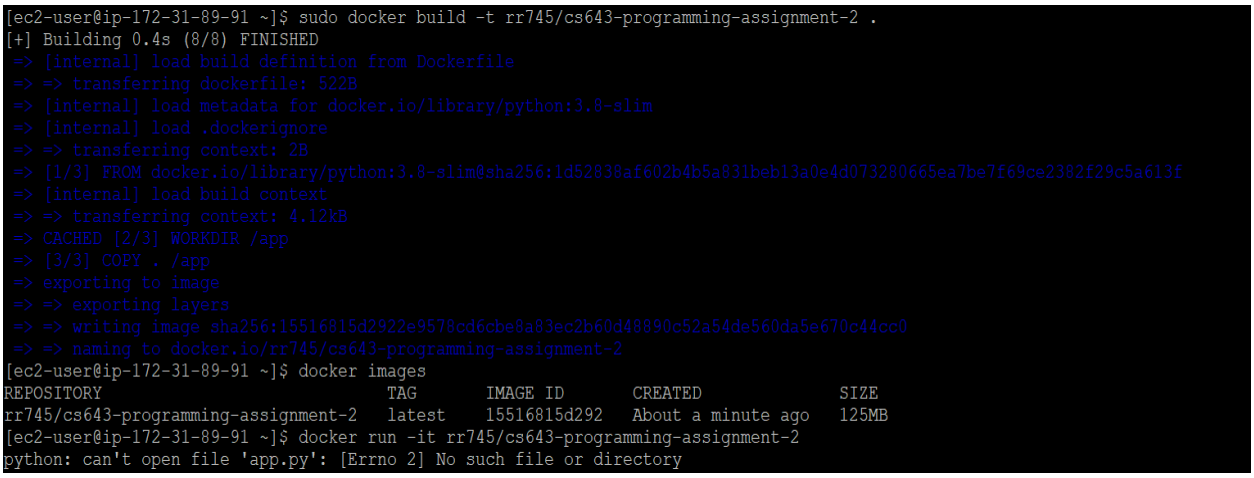
Create a dockerfile and build an image with the docker build command.

sudo docker build -t rr745/cs643-píogíamming-assignment-2 .

To check if a Docker image was built, use the following command:

sudo docker image ls



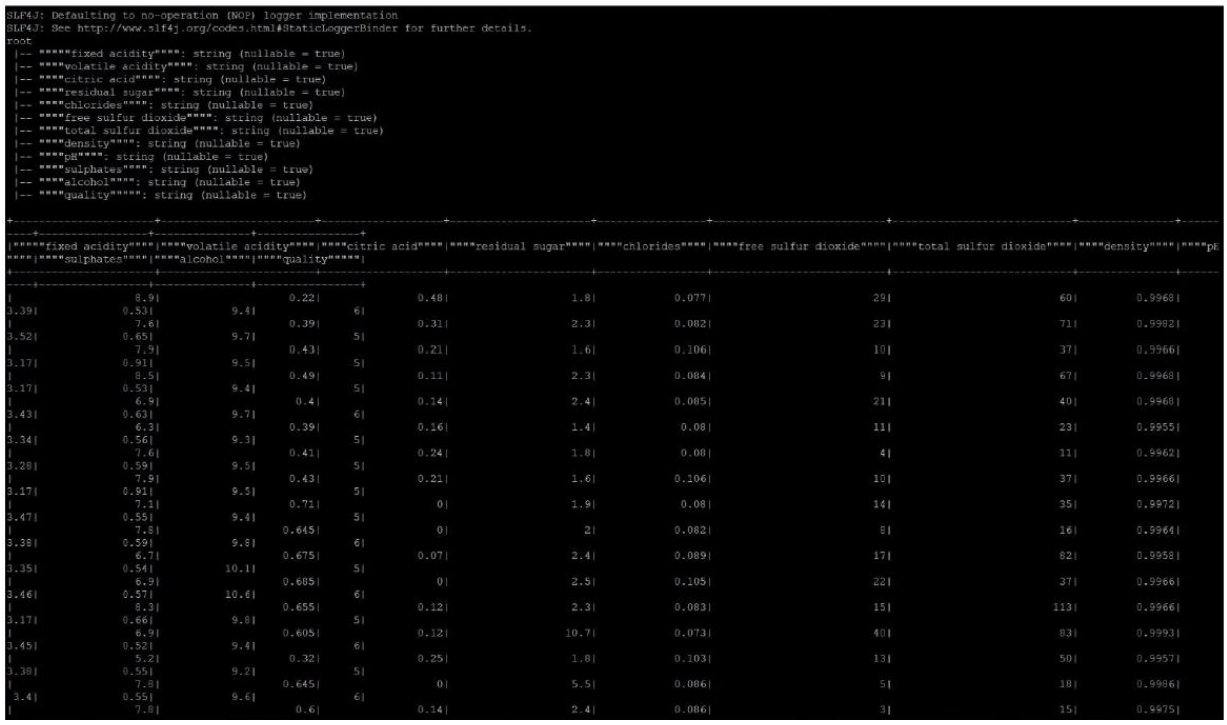


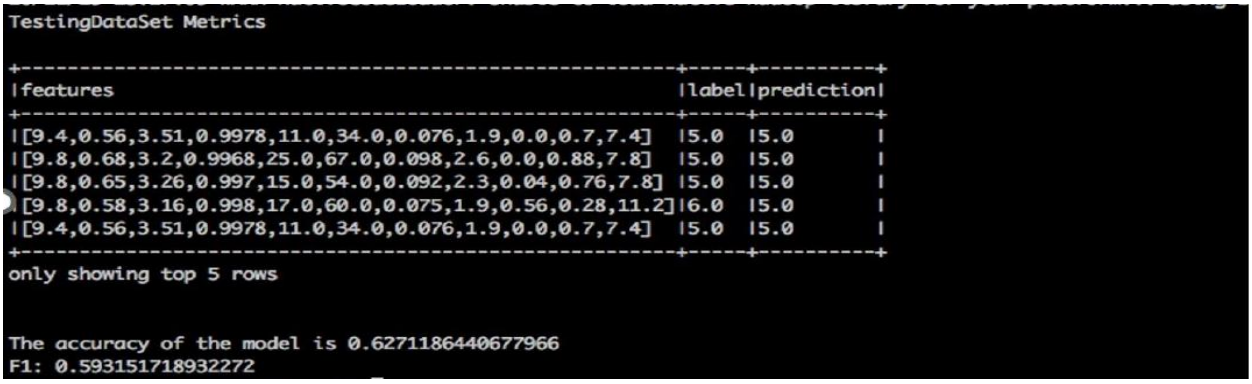
To run the docker image, use the following command:

sudo docker run -it rr745/cs643-programming-assignment-2

Instead of using the image name, you can use the image ID: # sudo docker run -it

This results in the same return for Accuracy and F1 scores.





Run the following command to submit the produced Docker image to the DockerHub repository:

sudo docker push rr745/643-programming-assignment-2

```
[ec2-user@ip-172-31-89-91 ~]$ docker login
Authenticating with existing credentials...
WARNING! Your password will be stored unencrypted in /home/ec2-user/.docker/config.json.
Configure a credential helper to remove this warning. See
https://docs.docker.com/engine/reference/commandline/login/#credentials-store

Login Succeeded
[ec2-user@ip-172-31-89-91 ~]$ sudo docker push rr745/cs643-programming-assignment-2
Using default tag: latest
The push refers to repository [docker.io/rr745/cs643-programming-assignment-2]
b0470c1d8c15: Preparing
88f82b0a637e: Preparing
d2a2207b52a4: Preparing
5d2d143f3d7f: Preparing
c3772b569c3a: Preparing
8d853c8add5d: Waiting
```

```
[ec2-user@ip-172-31-89-91 ~]$ docker login -u rr745
Password:
WARNING! Your password will be stored unencrypted in /home/ec2-user/.docker/config.json.
Configure a credential helper to remove this warning. See
https://docs.docker.com/engine/reference/commandline/login/#credentials-store

Login Succeeded
[ec2-user@ip-172-31-89-91 ~]$ docker push rr745/cs643-programming-assignment-2:latest
The push refers to repository [docker.io/rr745/cs643-programming-assignment-2]
fc4ddb5d5303: Pushed
88f82b0a637e: Pushed
d2a2207b52a4: Pushed
5d2d143f3d7f: Pushed
c3772b569c3a: Pushed
8d853c8add5d: Pushed
latest: digest: sha256:58e7a5a91cf558182c18635d034edd9d94fa95e2bab2ca7af63689e1652670c2 size: 1574
[ec2-user@ip-172-31-89-91 ~]$ docker pull rr745/cs643-programming-assignment-2:latest
latest: Pulling from rr745/cs643-programming-assignment-2
```

Download and execute the Docker image from the DockerHub repository, following the instructions provided on the website.

Git Bash:

the steps for moving code from my local system to a GitHub repository using Git Bash.

1. git init
2. git add
3. git status
4. git commit -m “Updates”
5. git remote add origin https://github.com/rr745/wine_quality_predictor.git
6. git push -u origin main