

## Programming Assignment 2

**Name :** Himaniben Ashokbhai Patel

**UCID :** Hp437

**Email Id :** [hp437@njit.edu](mailto:hp437@njit.edu)

**Goal:** The purpose of this individual assignment is to learn how to develop parallel machine learning (ML) applications in Amazon AWS cloud platform.

A wine quality prediction ML model in Spark over AWS. Here spark cluster is created by EMR. Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark. Here **Decision Tree Model** is used to predict the model. Here I couldn't implement docker. This describes step by step guide to run without docker.

- Create spark cluster having **4 instances** in that and with KeyPair named **mykeyp2.pem**.

aws Services Resource Groups

vocstartsoft/user487191=hp43... N. Virginia Support

### Create Cluster - Quick Options [Go to advanced options](#)

#### General Configuration

Cluster name

☒ Logging ⓘ

S3 folder

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

#### Software configuration

Release  ⓘ

Applications

- ☐ Core Hadoop: Hadoop 2.8.5 with Ganglia 3.7.2, Hive 2.3.6, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.2
- ☐ HBase: HBase 1.4.10 with Ganglia 3.7.2, Hadoop 2.8.5, Hive 2.3.6, Hue 4.4.0, Phoenix 4.14.3, and ZooKeeper 3.4.14
- ☐ Presto: Presto 0.227 with Hadoop 2.8.5 HDFS and Hive 2.3.6 Metastore
- ☐ Spark: Spark 2.4.0 with Hadoop 2.8.5 EMRFS with

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

aws Services Resource Groups

vocstartsoft/user487191=hp43... N. Virginia Support

Ganglia 3.7.2 and Zeppelin 0.8.2

☐ Use AWS Glue Data Catalog for table metadata ⓘ

#### Hardware configuration

Instance type  ⓘ The selected instance type adds 64 GiB of GP2 EBS storage per instance by default. [Learn more](#)

Number of instances  (1 master and 3 core nodes)

#### Security and access

EC2 key pair  ⓘ [Learn how to create an EC2 key pair.](#)

Permissions ☒ Default ☐ Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

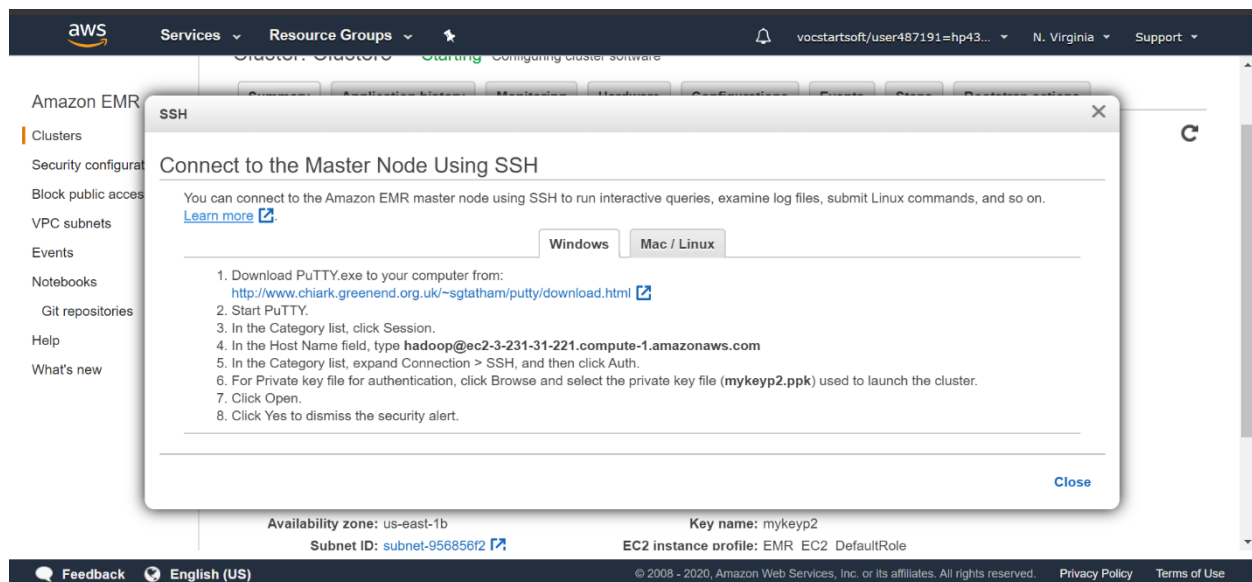
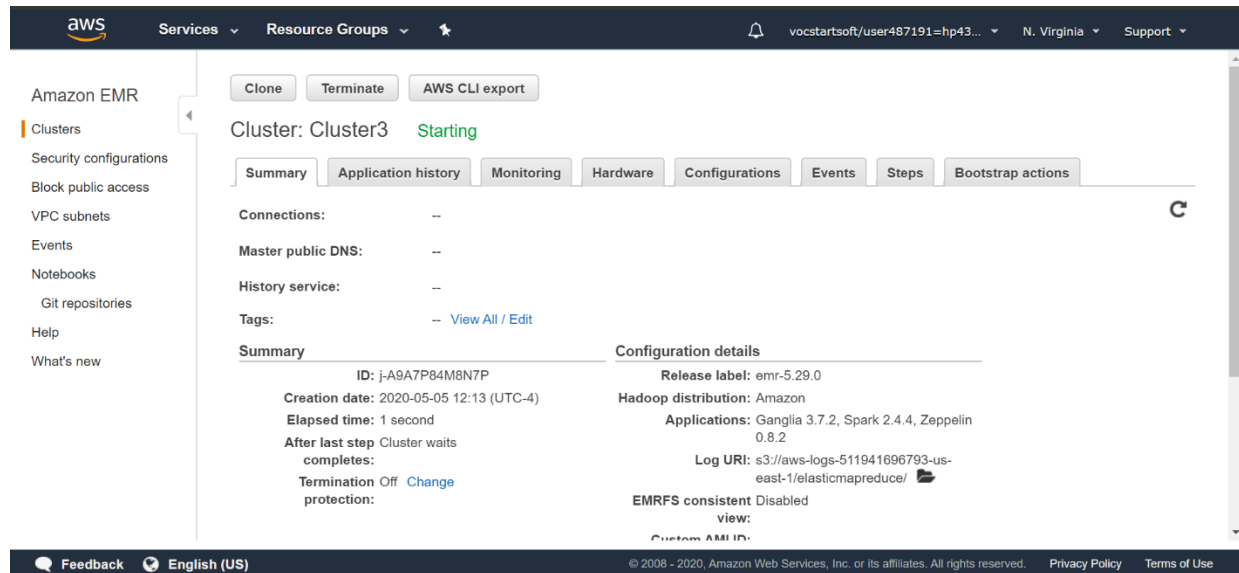
EMR role [EMR\\_DefaultRole](#) ⓘ

EC2 instance profile [EMR\\_EC2\\_DefaultRole](#) ⓘ

[Cancel](#) [Create cluster](#)

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

- After creating cluster it will look like as shown below. It has host name with that we can log in.



aws

Services

Resource Groups

🔔

vocstartsoft/user487191=hp43...

N. Virginia

Support

Amazon EMR

Clusters

Security configurations

Block public access

VPC subnets

Events

Notebooks

Git repositories

Help

What's new

Create cluster

View details

Clone

Terminate

Filter: Active clusters

Filter clusters ...

No clusters found

	Name	ID	Status	Creation time (UTC-4)	Elapsed time
--	------	----	--------	-----------------------	--------------

aws

Services

Resource Groups

🔔

vocstartsoft/user487191=hp43...

N. Virginia

Support

Amazon EMR

Clusters

Security configurations

Block public access

VPC subnets

Events

Notebooks

Git repositories

Help

What's new

Create cluster

View details

Clone

Terminate

Filter: Active clusters

Filter clusters ...

1 cluster (all loaded)

	Name	ID	Status	Creation time (UTC-4)	Elapsed time
<input type="checkbox"/>	Cluster3	j-A9A7P84M8N7P	Starting	2020-05-05 12:13 (UTC-4)	5 minutes

Feedback

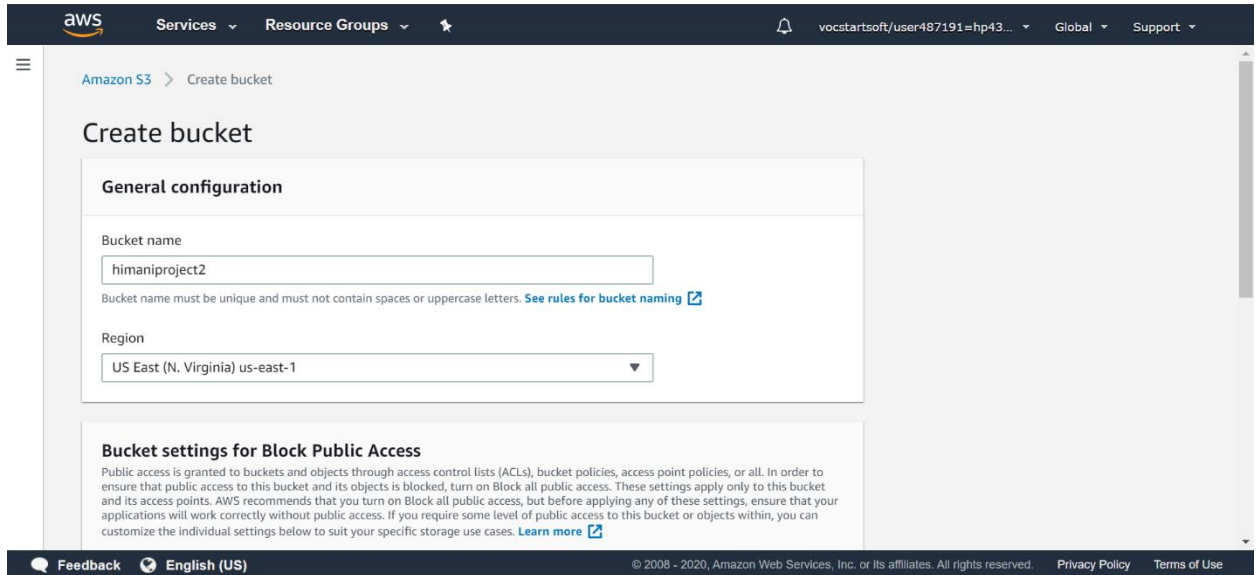
English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

- Create bucket named **himaniProject2** for storing dataset and python program.



**Create bucket**

**General configuration**

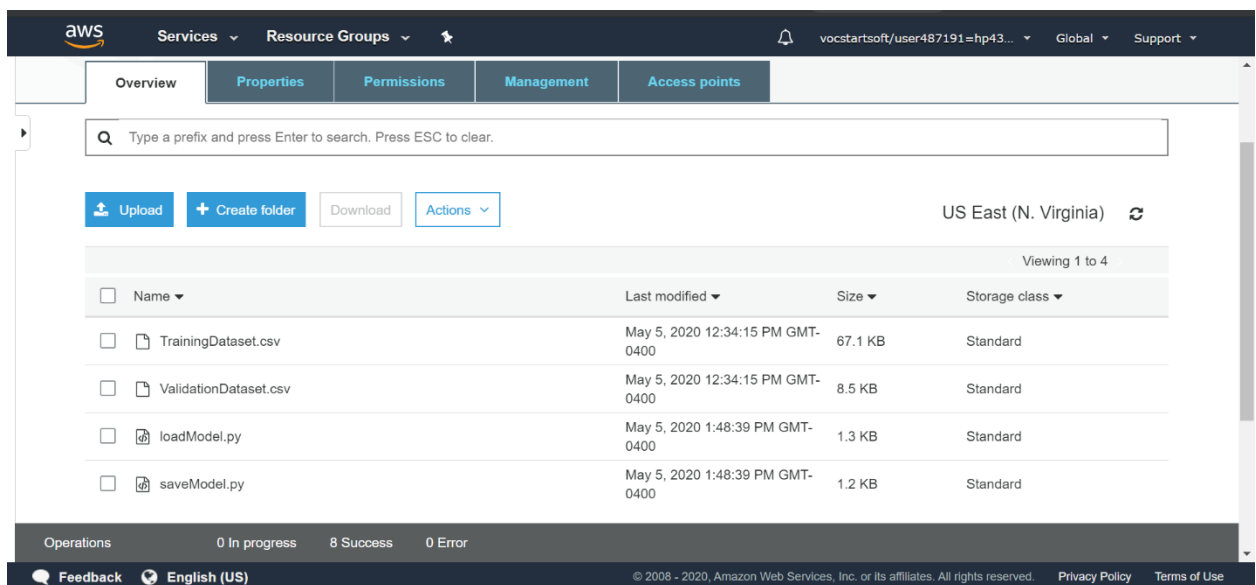
Bucket name

Bucket name must be unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

Region

**Bucket settings for Block Public Access**

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](#)



**Overview** | Properties | Permissions | Management | Access points

Q Type a prefix and press Enter to search. Press ESC to clear.

[Upload](#) [+ Create folder](#) [Download](#) [Actions](#)

US East (N. Virginia) [Refresh](#)

Viewing 1 to 4

<input type="checkbox"/>	Name	Last modified	Size	Storage class
<input type="checkbox"/>	TrainingDataset.csv	May 5, 2020 12:34:15 PM GMT-0400	67.1 KB	Standard
<input type="checkbox"/>	ValidationDataset.csv	May 5, 2020 12:34:15 PM GMT-0400	8.5 KB	Standard
<input type="checkbox"/>	loadModel.py	May 5, 2020 1:48:39 PM GMT-0400	1.3 KB	Standard
<input type="checkbox"/>	saveModel.py	May 5, 2020 1:48:39 PM GMT-0400	1.2 KB	Standard

Operations 0 In progress 8 Success 0 Error

- Now login with cluster host name

**ssh -i path\_to\_keypair host\_name**

**ssh -i C:\Users\Patel\Downloads\mykeyp2.pem hadoop@ec2-34-239-157-134.compute-1.amazonaws.com**

```
hadoop@ip-172-31-5-201:~
C:\Users\Patel>ssh -i C:\Users\Patel\Downloads\mykeyp2.pem hadoop@ec2-3-231-31-221.compute-1.amazonaws.com
The authenticity of host 'ec2-3-231-31-221.compute-1.amazonaws.com (3.231.31.221)' can't be established.
ECDSA key fingerprint is SHA256:uC5ykIZsitpql/ArkzqErUag@pcdQnaTtm77vB2Piz8.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-231-31-221.compute-1.amazonaws.com,3.231.31.221' (ECDSA) to the list of known hosts.
Last login: Tue May 5 16:19:23 2020

 _ | _ | _ )
 _ | ( _ /
 _ | \ | _ |
                Amazon Linux AMI

https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
21 package(s) needed for security, out of 40 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E:::EEEEEEEEEEEEEEEE M:::M M:::M R:::R
EE:::EEEEEEEEEEEEEEEE M:::M M:::M R:::RRRRRRRRRRRRRRR
E:::E EEEEE M:::M M:::M RR:::R R:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::EEEEEEEEEEEE M:::M M:::M M:::M R:::RRRRRRRRRRRRRRR
E:::E M:::M M:::M M:::M R:::RRRRRRRRRRRRRRR
E:::E M:::M M:::M M:::M R:::R R:::R
E:::E EEEEE M:::M MMM M:::M R:::R R:::R
EE:::EEEEEEEEEEEE M:::M M:::M R:::R R:::R
E:::EEEEEEEEEEEE M:::M M:::M RR:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRR RRRRRR

[hadoop@ip-172-31-5-201 ~]$
```

- Save **saveModel.py** from s3 bucket named **himaniproject2** to aws cluster.  
And the run that with spark-submit command

**aws s3 cp s3://himaniproject2/saveModel.py .**

**spark-submit ./saveModel.py**

```
hadoop@ip-172-31-5-201:~
Connection to ec2-3-231-31-221.compute-1.amazonaws.com closed.

C:\Users\Patel>ssh -i C:\Users\Patel\Downloads\mykeyp2.pem hadoop@ec2-3-231-31-221.compute-1.amazonaws.com
Last login: Tue May  5 18:20:38 2020

  _|  _|  _|
 _|  _|  _| /
 _|  _|  _|

Amazon Linux AMI

https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
21 package(s) needed for security, out of 40 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMM      MMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M      M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::::E M::::::::M      M::::::::M R::::::::RRRRRR::::R
E::::E      EEEEE M::::::::M      M::::::::M RR::::R      R::::R
E::::E      M::::::::M M::::::::M M::::::::M R::::R      R::::R
E::::EEEEEEEE M::::::::M M::::::::M M::::::::M R::::RRRRR::::R
E::::::::::::E M::::::::M M::::::::M M::::::::M R::::::::::::RR
E::::EEEEEEEE M::::::::M M::::::::M M::::::::M R::::::::RRRRR::::R
E::::E      M::::::::M M::::::::M M::::::::M R::::R      R::::R
E::::E      EEEEE M::::::::M      MMM      M::::::::M R::::R      R::::R
EE::::EEEEEEEE::::E M::::::::M      M::::::::M R::::R      R::::R
E::::::::::::E M::::::::M      M::::::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM      MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-5-201 ~]$ aws s3 cp s3://himaniproject2/saveModel.py .
download: s3://himaniproject2/saveModel.py to ./saveModel.py
[hadoop@ip-172-31-5-201 ~]$ spark-submit ./saveModel.py
```

```

hadoop@ip-172-31-5-201:~
[hadoop@ip-172-31-5-201 ~]$ spark-submit ./saveModel.py
20/05/05 18:27:10 INFO SparkContext: Running Spark version 2.4.4
20/05/05 18:27:10 INFO SparkContext: Submitted application: saveModel.py
20/05/05 18:27:10 INFO SecurityManager: Changing view acls to: hadoop
20/05/05 18:27:10 INFO SecurityManager: Changing modify acls to: hadoop
20/05/05 18:27:10 INFO SecurityManager: Changing view acls groups to:
20/05/05 18:27:10 INFO SecurityManager: Changing modify acls groups to:
20/05/05 18:27:10 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view per
missions: Set(hadoop); groups with view permissions: Set(); users with modify permissions: Set(hadoop); groups with mod
ify permissions: Set()
20/05/05 18:27:10 INFO Utils: Successfully started service 'sparkDriver' on port 35963.
20/05/05 18:27:10 INFO SparkEnv: Registering MapOutputTracker
20/05/05 18:27:10 INFO SparkEnv: Registering BlockManagerMaster
20/05/05 18:27:10 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topo
logy information
20/05/05 18:27:10 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
20/05/05 18:27:10 INFO DiskBlockManager: Created local directory at /mnt/tmp/blockmgr-b266552f-baf9-488b-982d-fead1506c8
20
20/05/05 18:27:10 INFO MemoryStore: MemoryStore started with capacity 1028.8 MB
20/05/05 18:27:10 INFO SparkEnv: Registering OutputCommitCoordinator
20/05/05 18:27:10 INFO Utils: Successfully started service 'SparkUI' on port 4040.
20/05/05 18:27:10 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://ip-172-31-5-201.ec2.internal:4040
20/05/05 18:27:11 INFO Utils: Using initial executors = 50, max of spark.dynamicAllocation.initialExecutors, spark.dynam
icAllocation.minExecutors and spark.executor.instances
20/05/05 18:27:11 INFO RMPProxy: Connecting to ResourceManager at ip-172-31-5-201.ec2.internal/172.31.5.201:8032
20/05/05 18:27:11 INFO Client: Requesting a new application from cluster with 3 NodeManagers
20/05/05 18:27:11 INFO Client: Verifying our application has not requested more than the maximum memory capability of th
e cluster (12288 MB per container)
20/05/05 18:27:11 INFO Client: Will allocate AM container, with 896 MB memory including 384 MB overhead
20/05/05 18:27:11 INFO Client: Setting up container launch context for our AM

```

```

hadoop@ip-172-31-5-201:~
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1219
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1237
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1234
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1242
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1247
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1222
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1244
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1227
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1243
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1241
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1229
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1246
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Interrupting monitor thread
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Shutting down all executors
20/05/05 18:27:40 INFO YarnSchedulerBackend$YarnDriverEndpoint: Asking each executor to shut down
20/05/05 18:27:40 INFO SchedulerExtensionServices: Stopping SchedulerExtensionServices
(serviceOption=None,
services=List(),
started=false)
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Stopped
20/05/05 18:27:40 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/05/05 18:27:40 INFO MemoryStore: MemoryStore cleared
20/05/05 18:27:40 INFO BlockManager: BlockManager stopped
20/05/05 18:27:40 INFO BlockManagerMaster: BlockManagerMaster stopped
20/05/05 18:27:40 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/05/05 18:27:40 INFO SparkContext: Successfully stopped SparkContext
20/05/05 18:27:40 INFO ShutdownHookManager: Shutdown hook called
20/05/05 18:27:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-b9ca3188-89e6-4c5b-92fc-ed37730bdaff
20/05/05 18:27:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-feb47605-b172-421d-b938-95112bf19b05/pyspa
rk-d473c48e-2253-4501-b498-afd8891ee612

```



- Create another cluster with only one instance and load prediction model.

```

[hadoop@ip-172-31-36-46 ~]$ ssh -i C:\Users\Patel\Downloads\mykeyp2.pem hadoop@ec2-52-206-154-34.compute-1.amazonaws.com
C:\Users\Patel>ssh -i C:\Users\Patel\Downloads\mykeyp2.pem hadoop@ec2-52-206-154-34.compute-1.amazonaws.com
The authenticity of host 'ec2-52-206-154-34.compute-1.amazonaws.com (52.206.154.34)' can't be established.
ECDSA key fingerprint is SHA256:C18QmXtyRbiW4G6o7Q3SV1t2mAimkSTdar8NLxrqB8.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-52-206-154-34.compute-1.amazonaws.com,52.206.154.34' (ECDSA) to the list of known hosts.

  _ _ _ _ _
 _| ( _| /
 _| \_|_|_|

Amazon Linux AMI

https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
22 package(s) needed for security, out of 41 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::E M:::::M M:::::M R:::::R
EE:::::EEEEEEEE::E M:::::M M:::::M R:::::RRRRRR:::R
 E::E EEEEE M:::::M M:::::M RR:::R R:::R
E::E M:::::M::M M:::::M R:::R R:::R
E:::::EEEEEEEE M:::M M:::M M:::M R:::RRRRR:::R
E:::::E M:::::M M:::M M:::M R:::::RR
E:::::EEEEEEEE M:::::M M:::::M R:::RRRRR:::R
 E::E M:::::M M:::M M:::M R:::R R:::R
E::E EEEEE M:::::M MMM M:::::M R:::R R:::R
EE:::::EEEEEEEE::E M:::::M M:::::M R:::R R:::R
E:::::E M:::::M M:::::M RR:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-36-46 ~]$ aws s3 cp s3://himaniproject2/loadModel.py

```

```
hadoop@ip-172-31-5-201:~  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1247  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1222  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1244  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1227  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1243  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1241  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1229  
20/05/05 18:27:40 INFO ContextCleaner: Cleaned accumulator 1246  
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Interrupting monitor thread  
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Shutting down all executors  
20/05/05 18:27:40 INFO YarnSchedulerBackend$YarnDriverEndpoint: Asking each executor to shut down  
20/05/05 18:27:40 INFO SchedulerExtensionServices: Stopping SchedulerExtensionServices  
(serviceOption=None,  
  services=List(),  
  started=false)  
20/05/05 18:27:40 INFO YarnClientSchedulerBackend: Stopped  
20/05/05 18:27:40 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!  
20/05/05 18:27:40 INFO MemoryStore: MemoryStore cleared  
20/05/05 18:27:40 INFO BlockManager: BlockManager stopped  
20/05/05 18:27:40 INFO BlockManagerMaster: BlockManagerMaster stopped  
20/05/05 18:27:40 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!  
20/05/05 18:27:40 INFO SparkContext: Successfully stopped SparkContext  
20/05/05 18:27:40 INFO ShutdownHookManager: Shutdown hook called  
20/05/05 18:27:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-b9ca3188-89e6-4c5b-92fc-ed37730bda4f  
20/05/05 18:27:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-feb47605-b172-421d-b938-95112bf19b05/pyspark-  
rk-d473c48e-2253-4501-b498-afd8891ee612  
20/05/05 18:27:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-feb47605-b172-421d-b938-95112bf19b05  
[hadoop@ip-172-31-5-201 ~]$ aws s3 cp s3://himaniproject2/loadModel.py .  
download: s3://himaniproject2/loadModel.py to ./loadModel.py  
[hadoop@ip-172-31-5-201 ~]$ spark-submit ./loadModel.py
```

```
hadoop@ip-172-31-5-201:~
20/05/05 18:46:15 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 7 (PythonRDD[33] at count at /home/hadoop/.loadModel.py:27) (first 15 tasks are for partitions Vector(0))
20/05/05 18:46:15 INFO YarnScheduler: Adding task set 7.0 with 1 tasks
20/05/05 18:46:15 INFO TaskSetManager: Starting task 0.0 in stage 7.0 (TID 29, ip-172-31-13-182.ec2.internal, executor 2, partition 0, RACK_LOCAL, 8268 bytes)
20/05/05 18:46:15 INFO BlockManagerInfo: Added broadcast_12_piece0 in memory on ip-172-31-13-182.ec2.internal:46617 (size: 8.7 KB, free: 5.4 GB)
20/05/05 18:46:15 INFO BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-172-31-13-182.ec2.internal:46617 (size: 26.9 KB, free: 5.4 GB)
20/05/05 18:46:15 INFO TaskSetManager: Finished task 0.0 in stage 7.0 (TID 29) in 874 ms on ip-172-31-13-182.ec2.internal (executor 2) (1/1)
20/05/05 18:46:15 INFO YarnScheduler: Removed TaskSet 7.0, whose tasks have all completed, from pool
20/05/05 18:46:15 INFO DAGScheduler: ResultStage 7 (count at /home/hadoop/.loadModel.py:27) finished in 0.880 s
20/05/05 18:46:15 INFO DAGScheduler: Job 6 finished: count at /home/hadoop/.loadModel.py:27, took 0.882179 s
Model accuracy.....: 46.250%
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 176
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 184
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 219
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 177
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 206
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 218
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 182
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 217
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 174
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 209
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 173
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 172
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 207
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 189
20/05/05 18:46:16 INFO ContextCleaner: Cleaned accumulator 178

hadoop@ip-172-31-5-201:~
20/05/05 18:46:17 INFO BlockManagerInfo: Added broadcast_16_piece0 in memory on ip-172-31-11-44.ec2.internal:38665 (size: 2013.0 B, free: 5.4 GB)
20/05/05 18:46:17 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 2 to 172.31.11.44:46776
20/05/05 18:46:17 INFO TaskSetManager: Finished task 0.0 in stage 11.0 (TID 33) in 35 ms on ip-172-31-11-44.ec2.internal (executor 1) (1/1)
20/05/05 18:46:17 INFO YarnScheduler: Removed TaskSet 11.0, whose tasks have all completed, from pool
20/05/05 18:46:17 INFO DAGScheduler: ResultStage 11 (countByValue at MulticlassMetrics.scala:42) finished in 0.042 s
20/05/05 18:46:17 INFO DAGScheduler: Job 8 finished: countByValue at MulticlassMetrics.scala:42, took 1.034711 s
F1 Score..... = 0.45
20/05/05 18:46:17 INFO SparkContext: Invoking stop() from shutdown hook
20/05/05 18:46:17 INFO SparkUI: Stopped Spark web UI at http://ip-172-31-5-201.ec2.internal:4040
20/05/05 18:46:17 INFO YarnClientSchedulerBackend: Interrupting monitor thread
20/05/05 18:46:17 INFO YarnClientSchedulerBackend: Shutting down all executors
20/05/05 18:46:17 INFO YarnSchedulerBackend$YarnDriverEndpoint: Asking each executor to shut down
20/05/05 18:46:17 INFO SchedulerExtensionServices: Stopping SchedulerExtensionServices
(serviceOption=None,
services=List(),
started=false)
20/05/05 18:46:17 INFO YarnClientSchedulerBackend: Stopped
20/05/05 18:46:17 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/05/05 18:46:17 INFO MemoryStore: MemoryStore cleared
20/05/05 18:46:17 INFO BlockManager: BlockManager stopped
20/05/05 18:46:17 INFO BlockManagerMaster: BlockManagerMaster stopped
20/05/05 18:46:17 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/05/05 18:46:17 INFO SparkContext: Successfully stopped SparkContext
20/05/05 18:46:17 INFO ShutdownHookManager: Shutdown hook called
20/05/05 18:46:17 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-d095f55c-665e-403f-b923-b1ecadebb020
20/05/05 18:46:17 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-af738c58-728b-4988-a6a4-2fe19db95013/pyspark-be07a621-5d0e-4c01-81b9-cd5a67bcb6ec
```

Here decision tree model is giving 46.25% accuracy and 0.45 f1 score.