

**Contributor: Mohammed Hannan Desai(mmd76)**

## **Wine Quality Prediction App using Apache Spark over AWS EMR Cluster**

This project demonstrated the use of Apache Spark on AWS EMR Cluster to train machine learning model parallelly on multiple EC2 instance, 4 in our case and then save that model and use it to predict the quality of wine.

Data contains the information about wine quality, which is publicly available, so data has been available in format of two csv files TrainingDataSet.csv & ValidationDataSet.csv, both will be used to train and validate ML model.

This project contains two applications:

Both applications are developed using Java language and maven project management tool.

Link to download these two provided below:

- 1. Application for build and train the model (wineprediction-model-training):** This application is used to train machine learning model using Linear Regression Algorithm and once model is created & trained, we will save it on Amazon s3 bucket.

**GitHub Repo URL:**

<https://github.com/hannandesai/wineprediction-model-training.git>

This application will be run over Amazon EMR cluster using Apache Spark over 4 EC2 instance.

**Command to run this app:**

```
mvn clean package
```

```
mvn exec:java -Dexec.mainClass="org.example.wineprediction.App"
```

- 2. Prediction Application (winepredictionapp):** This application will use the model created by first application to predict quality of wines which will be provided as input from TestData.csv file.

**GitHub Repo URL:**

<https://github.com/hannandesai/winepredictionapp.git>

This application will be run over single EC2 instance and will run using docker created image.

**Docker Hub URL for image:**

<https://hub.docker.com/repository/docker/hannan9900/winepredictionapp/general>

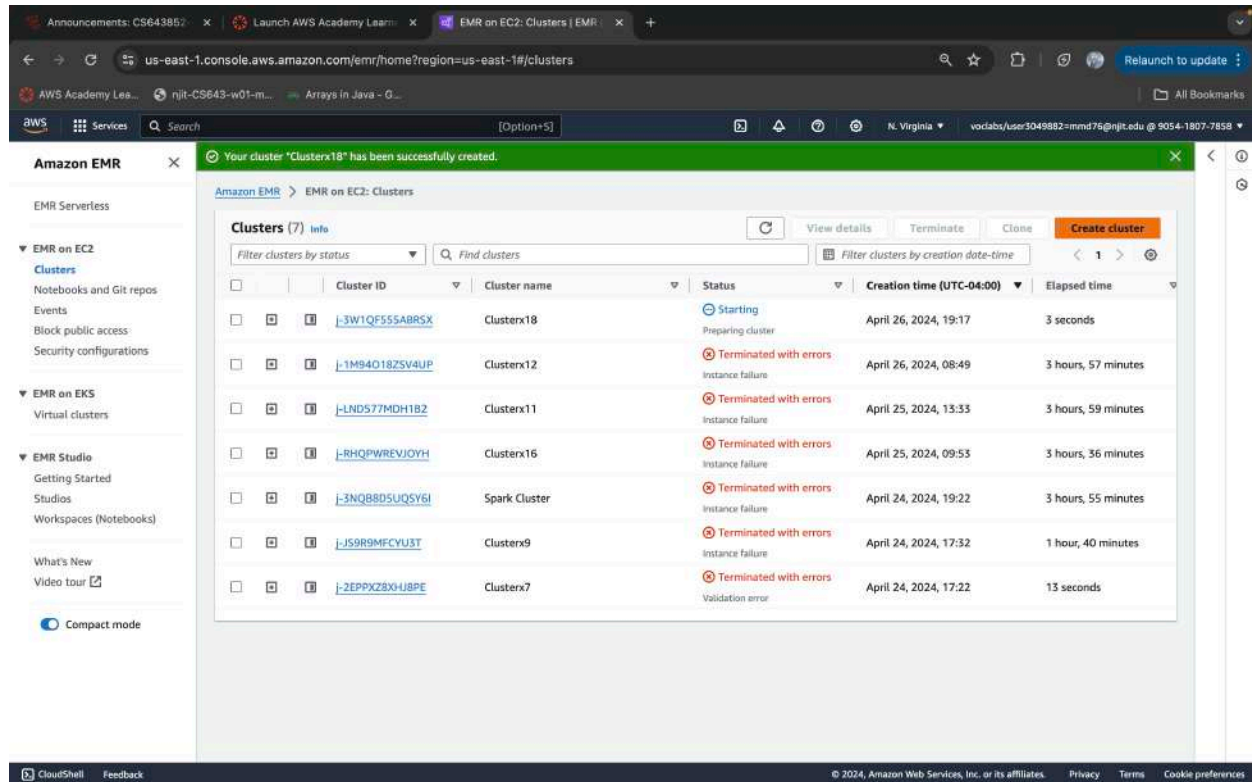
**Command to run this app:**

1. mvn clean package
2. mvn exec:java

`-Dexec.mainClass="org.example.winepredictionapp.App"`

## Cloud Setup:

### Setup EMR Cluster



Announcements: CS643852 x Launch AWS Academy Learn x EMR on EC2: Clusters | EMR x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusters

AWS Academy Lea... njit-CS643-w01-m... Arrays in Java - G...

Services Search [Option+S]

N. Virginia voclabs/user3049882:mmd76@njit.edu @ 9054-1807-7858

**Amazon EMR** x Your cluster "Clusterx18" has been successfully created.

Amazon EMR > EMR on EC2: Clusters

Clusters (7) info View details Terminate Clone Create cluster

Filter clusters by status Find clusters Filter clusters by creation date-time

	Cluster ID	Cluster name	Status	Creation time (UTC-04:00)	Elapsed time
<input type="checkbox"/>	<a href="#">j-3W1QF555ABRSX</a>	Clusterx18	Starting Preparing cluster	April 26, 2024, 19:17	3 seconds
<input type="checkbox"/>	<a href="#">j-1M84O18ZSV4UP</a>	Clusterx12	Terminated with errors Instance failure	April 26, 2024, 08:49	3 hours, 57 minutes
<input type="checkbox"/>	<a href="#">j-LND577MDH1B2</a>	Clusterx11	Terminated with errors Instance failure	April 25, 2024, 13:33	3 hours, 59 minutes
<input type="checkbox"/>	<a href="#">j-RHQPWREVJOYH</a>	Clusterx16	Terminated with errors Instance failure	April 25, 2024, 09:53	3 hours, 36 minutes
<input type="checkbox"/>	<a href="#">j-3NQ88D5UQSY6I</a>	Spark Cluster	Terminated with errors Instance failure	April 24, 2024, 19:22	3 hours, 55 minutes
<input type="checkbox"/>	<a href="#">j-J59R9MFCYU3T</a>	Clusterx9	Terminated with errors Instance failure	April 24, 2024, 17:32	1 hour, 40 minutes
<input type="checkbox"/>	<a href="#">j-2EPPXZ8X-U8PE</a>	Clusterx7	Terminated with errors Validation error	April 24, 2024, 17:22	13 seconds

Compact mode

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Announcements: CS643852...Launch AWS Academy Learn...Create cluster > EMR on EC2...+

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/cloneCluster/ji-1M94O18ZSV4UP

Relaunch to update

AWS Academy Lea...njit-CS643-w01-m...Arrays in Java - G...

ServicesSearch[Option+S]

N. Virginiavoclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

Amazon EMR > EMR on EC2: Clusters > Create cluster

Clone "Clusterx12" Info

Name

Clusterx18

Amazon EMR release Info

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

emr-7.1.0

Application bundle

Spark Interactive

Core Hadoop

Flink

HBase

Presto

Trino

Custom

☐ AmazonCloudWatchAgent 1.300032.2

☐ HCatalog 3.1.3

☐ Hue 4.11.0

☒ Livy 0.8.0

☐ Phoenix 5.1.3

☒ Spark 3.5.0

☐ Tez 0.10.2

☐ ZooKeeper 3.9.1

☐ Flink 1.18.1

☒ Hadoop 3.3.6

☒ JupyterEnterpriseGateway 2.6.0

☐ MXNet 1.9.1

☐ Pig 0.17.0

☐ Sqoop 1.4.7

☐ Trino 435

☐ HBase 2.4.17

☒ Hive 3.1.3

☐ JupyterHub 1.5.0

☐ Oozie 5.2.1

☐ Presto 0.284

☐ TensorFlow 2.11.0

☐ Zeppelin 0.10.1

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

Summary Info

Name and applications - required

Name

Clusterx18

Amazon EMR release

emr-7.1.0

Application bundle

Spark Interactive (Hadoop 3.3.6, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5...)

Cluster configuration - required

Uniform instance groups

Primary (m4.xlarge), Core (m4.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration

Core size: 4 instances

CancelClone cluster

CloudShellFeedback

© 2024, Amazon Web Services, Inc. or its affiliates. PrivacyTermsCookie preferences

Announcements: CS643852...Launch AWS Academy Learn...Create cluster > EMR on EC2...+

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/cloneCluster/ji-1M94O18ZSV4UP

Relaunch to update

AWS Academy Lea...njit-CS643-w01-m...Arrays in Java - G...

ServicesSearch[Option+S]

N. Virginiovoclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

Amazon EMR > EMR on EC2: Clusters > Create cluster

EBS root volume

EBS root volume applies to the operating systems and applications that you install on the cluster. [EBS root volume ratio constraints](#)

Size (GiB)

15

IOPS

3000

Throughput (MiB/s)

125

15 - 100 GiB per volume. General Purpose SSD (gp3)

3000 - 16000 IOPS per volume. Choose a maximum ratio of 500:1 between IOPS and volume size.

125 - 1000 MiB/s per volume. Choose a maximum ratio of 0.25:1 between throughput and IOPS.

Cluster scaling and provisioning - required Info

Choose how Amazon EMR should size your cluster.

Choose an option

☒ Set cluster size manually

☐ Use EMR-managed scaling

☐ Use custom automatic scaling

Use this option if you know your workload patterns in advance.

Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.

To programmatically scale core and task nodes, create custom automatic scaling policies.

Provisioning configuration

Set the size of your core instance group. Amazon EMR attempts to provision this capacity when you launch your cluster.

Name	Instance type	Instance(s) size	Use Spot purchasing option
Core	m4.xlarge	4	<input type="checkbox"/>

Networking - required Info

Choose the network settings that determine how you and other entities communicate with your cluster.

Summary Info

Name and applications - required

Name

Clusterx18

Amazon EMR release

emr-7.1.0

Application bundle

Spark Interactive (Hadoop 3.3.6, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5...)

Cluster configuration - required

Uniform instance groups

Primary (m4.xlarge), Core (m4.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration

Core size: 4 instances

CancelClone cluster

CloudShellFeedback

© 2024, Amazon Web Services, Inc. or its affiliates. PrivacyTermsCookie preferences

Announcements: CS643852 x Launch AWS Academy Learn x Create cluster > EMR on EC2 x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/cloneCluster/fj-1M94O18ZSV4UP

Relaunch to update

us-east-1 console

Services Search [Option+S]

N. Virginia voclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

EC2 security groups (firewall)

Steps (0) Info Remove Edit Add

Cluster termination and node replacement Info

Bootstrap actions (0) Info Remove Edit Add

Cluster logs Info

Summary Info

Name and applications - required

Cluster configuration - required

Cluster scaling and provisioning - required

Cancel Clone cluster

Announcements: CS643852 x Launch AWS Academy Learn x Create cluster > EMR on EC2 x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/cloneCluster/fj-1M94O18ZSV4UP

Relaunch to update

us-east-1 console

Services Search [Option+S]

N. Virginia voclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

Security configuration and EC2 key pair Info

Identity and Access Management (IAM) roles - required Info

EC2 instance profile for Amazon EMR

Summary Info

Name and applications - required

Cluster configuration - required

Cluster scaling and provisioning - required

Cancel Clone cluster

Announcements: CS643852 x Launch AWS Academy Learn x EMR on EC2: Clusters | EMR x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusters

AWS Academy Lea... njit-CS643-w01-m... Arrays in Java - G...

Services Search [Option+S]

N. Virginia voclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

### 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 (7) info

Filter clusters by status Find clusters Filter clusters by creation date-time

	Cluster ID	Cluster name	Status	Creation time (UTC-04:00)	Elapsed time
<input type="checkbox"/>	<a href="#">j-3W1QF555ABRSX</a>	Clusterx18	Waiting Ready to run steps	April 26, 2024, 19:17	7 minutes, 1 second
<input type="checkbox"/>	<a href="#">j-1M9401825VAUP</a>	Clusterx12	Terminated with errors Instance failure	April 26, 2024, 08:49	3 hours, 57 minutes
<input type="checkbox"/>	<a href="#">j-LND577MDH1B2</a>	Clusterx11	Terminated with errors Instance failure	April 25, 2024, 13:33	3 hours, 59 minutes
<input type="checkbox"/>	<a href="#">j-RHQPWREJJOYH</a>	Clusterx16	Terminated with errors Instance failure	April 25, 2024, 09:53	3 hours, 36 minutes
<input type="checkbox"/>	<a href="#">j-3NQ88D5UQSY6I</a>	Spark Cluster	Terminated with errors Instance failure	April 24, 2024, 19:22	3 hours, 55 minutes
<input type="checkbox"/>	<a href="#">j-J59RSMFCVU3T</a>	Clusterx9	Terminated with errors Instance failure	April 24, 2024, 17:32	1 hour, 40 minutes
<input type="checkbox"/>	<a href="#">j-2EPPX28XHJ8PE</a>	Clusterx7	Terminated with errors Validation error	April 24, 2024, 17:22	13 seconds

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Announcements: CS643852 x Launch AWS Academy Learn x Instances | EC2 | us-east-1 x +

us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#instances:v=3,\$case=tags:true%5C,client:false,\$regex=tags:false%5...

AWS Academy Lea... njit-CS643-w01-m... Arrays in Java - G...

Services Search [Option+S]

N. Virginia voclabs/user3049882=mmd76@njit.edu @ 9054-1807-7858

### EC2 Dashboard

EC2 Global View

Events

Console-to-Code [Preview](#)

▼ Instances

- Instances
- Instance Types
- Launch Templates
- Spot Requests
- Savings Plans
- Reserved Instances
- Dedicated Hosts
- Capacity Reservations [New](#)

▼ Images

- AMIs
- AMI Catalog

▼ Elastic Block Store

- Volumes
- Snapshots
- Lifecycle Manager

▼ Network & Security

- Security Groups
- Elastic IPs
- Placement Groups

### Instances (8) info

Find instance by attribute or tag (case-sensitive)

All states

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS
<input type="checkbox"/>	Java Server	i-05cd49fe103233bc0	Running	t4g.nano	2/2 checks passed	View alarms +	us-east-1d	ec2-54-88-93-165.com...
<input type="checkbox"/>	EC-B	i-00ec4f15808f98a96	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1c	ec2-44-208-34-251.co...
<input type="checkbox"/>	EC-A	i-0b24410448adebb29	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1c	ec2-52-87-182-243.co...
<input type="checkbox"/>		i-055c9b00ced715	Running	m4.xlarge	2/2 checks passed	View alarms +	us-east-1e	ec2-54-90-22-1.comput...
<input type="checkbox"/>		i-02f94c0fec253c513	Running	m4.xlarge	2/2 checks passed	View alarms +	us-east-1e	ec2-35-153-33-29.com...
<input type="checkbox"/>		i-03632d9e7a0234d19	Running	m4.xlarge	2/2 checks passed	View alarms +	us-east-1e	ec2-100-26-10-230.co...
<input type="checkbox"/>		i-00807dae3f8c35897	Running	m4.xlarge	2/2 checks passed	View alarms +	us-east-1e	ec2-100-24-244-252.co...
<input type="checkbox"/>		i-09761729bd9532370	Running	m4.xlarge	2/2 checks passed	View alarms +	us-east-1e	ec2-34-232-78-9.comp...

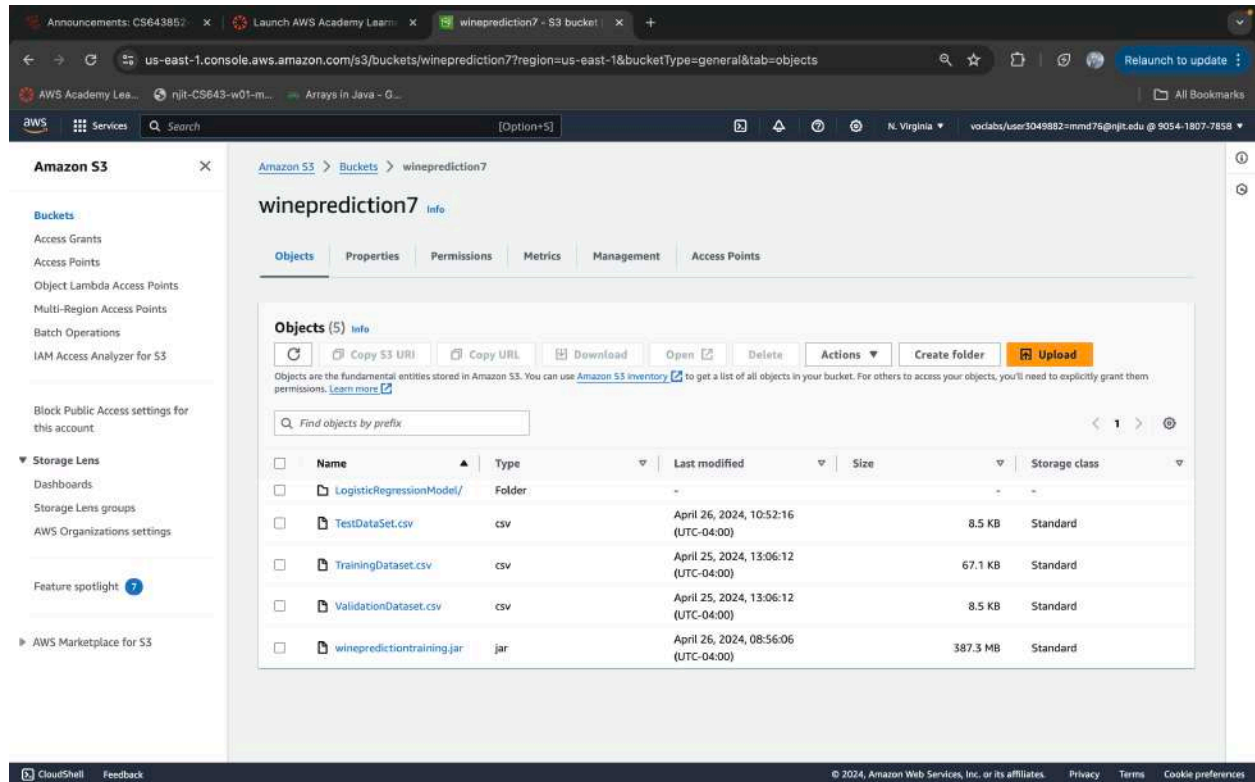
Select an instance

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences



## Create S3 bucket for storing input data csv files and trained ML Model:



The screenshot displays the Amazon S3 console interface for a bucket named 'wineprediction7'. The left sidebar shows the 'Amazon S3' navigation menu with options like Buckets, Access Grants, and Storage Lens. The main content area shows the 'Objects (5)' tab, which lists the following items:

Name	Type	Last modified	Size	Storage class
LogisticRegressionModel/	Folder	-	-	-
TestDataSet.csv	csv	April 26, 2024, 10:52:16 (UTC-04:00)	8.5 KB	Standard
TrainingDataset.csv	csv	April 25, 2024, 13:06:12 (UTC-04:00)	67.1 KB	Standard
ValidationDataset.csv	csv	April 25, 2024, 13:06:12 (UTC-04:00)	8.5 KB	Standard
winepredictiontraining.jar	jar	April 26, 2024, 08:56:06 (UTC-04:00)	387.3 MB	Standard

## Run “wineprediction-model-training” application on created EMR Cluster:

### 1. Connect to cluster:

Announcements: CS643852 x Launch AWS Academy Learn x Properties > Clusterx18 > EM x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/j-3W1QF555ABRSX

AWS Academy Lea... njit-CS643-w01-m... Arrays in Java - G...

AWS Services Search [Option+S] N. Virginia voclabs/user3049882-mmnd76@njit.edu @ 9054-1807-7858

## Clusterx18

Updated less than a minute ago

Terminate Clone in AWS CLI Clone

### Summary

<b>Cluster info</b> Cluster ID j-3W1QF555ABRSX Cluster configuration Instance groups Capacity 1 Primary 4 Core 0 Task	<b>Applications</b> Amazon EMR version emr-7.1.0 Installed applications Hadoop 3.3.6, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5.0	<b>Cluster management</b> Log destination in Amazon S3 aws-logs-905418077858-us-east-1/elasticmapreduce Persistent application UIs Spark History Server YARN timeline server Tez UI Primary node public DNS ec2-34-232-78-9.compute-1.amazonaws.com Connect to the Primary node using SSH Connect to the Primary node using SSM	<b>Status and time</b> Status Waiting Creation time April 26, 2024, 19:17 (UTC-04:00) Elapsed time 25 minutes, 59 seconds
---	---	---	---

Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events Tags (0)

<b>Operating system</b> Amazon Linux release 2023.4.20240416.0	<b>Cluster logs</b> Archive log files to Amazon S3 Turned on Amazon S3 location s3://aws-logs-905418077858-us-east-1/elasticmapreduce/ Encryption for logs Turned off	<b>Cluster termination and node replacement</b> Edit Termination option Manually terminate cluster Idle time - Termination protection On Unhealthy node replacement On
--	---	---

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Announcements: CS643852 x Launch AWS Academy Learn x Properties > Clusterx18 > EM x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/j-3W1QF555ABRSX

AWS Academy Lea... njit-CS643-w01-m... Arrays in Java - G...

AWS Services Search [Option+S] N. Virginia voclabs/user3049882-mmnd76@njit.edu @ 9054-1807-7858

## Clusterx18

Updated less than a minute ago

Terminate Clone in AWS CLI Clone

### Summary

<b>Cluster info</b> Cluster ID j-3W1QF555ABRSX Cluster configuration Instance groups Capacity 1 Primary 4 Core 0 Task	<b>Applications</b> Amazon EMR version emr-7.1.0 Installed applications Hadoop 3.3.6, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5.0	<b>Cluster management</b> Log destination in Amazon S3 aws-logs-905418077858-us-east-1/elasticmapreduce Persistent application UIs Spark History Server YARN timeline server Tez UI Primary node public DNS ec2-34-232-78-9.compute-1.amazonaws.com Connect to the Primary node using SSH Connect to the Primary node using SSM	<b>Status and time</b> Status Waiting Creation time April 26, 2024, 19:17 (UTC-04:00) Elapsed time 25 minutes, 59 seconds
---	---	---	---

Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events Tags (0)

<b>Operating system</b> Amazon Linux release 2023.4.20240416.0	<b>Cluster logs</b> Archive log files to Amazon S3 Turned on Amazon S3 location s3://aws-logs-905418077858-us-east-1/elasticmapreduce/ Encryption for logs Turned off	<b>Cluster termination and node replacement</b> Edit Termination option Manually terminate cluster Idle time - Termination protection On Unhealthy node replacement On
--	---	---

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

### Connect to the primary node using SSH

You can connect to the Amazon EMR primary node using SSH to perform actions like running interactive queries, examining log files, submit Linux commands, and view web interfaces hosted on Amazon EMR clusters. [Learn more](#)

Windows Mac/Linux

1. Open a terminal window. On Mac OS X, choose Applications > Utilities > Terminal. On other Linux distributions, terminal is typically found at Applications > Accessories > Terminal.
2. To establish a connection to the primary node, enter the following command. Replace ~/cluster7-cert.pem with the location and filename of the private key file (.pem) that you used to launch the cluster.  

```
ssh -i ~/cluster7-cert.pem hadoop@ec2-34-232-78-9.compute-1.amazonaws.com
```
3. Enter yes to dismiss the security warning.

[View web interfaces hosted on Amazon EMR clusters](#)

Close

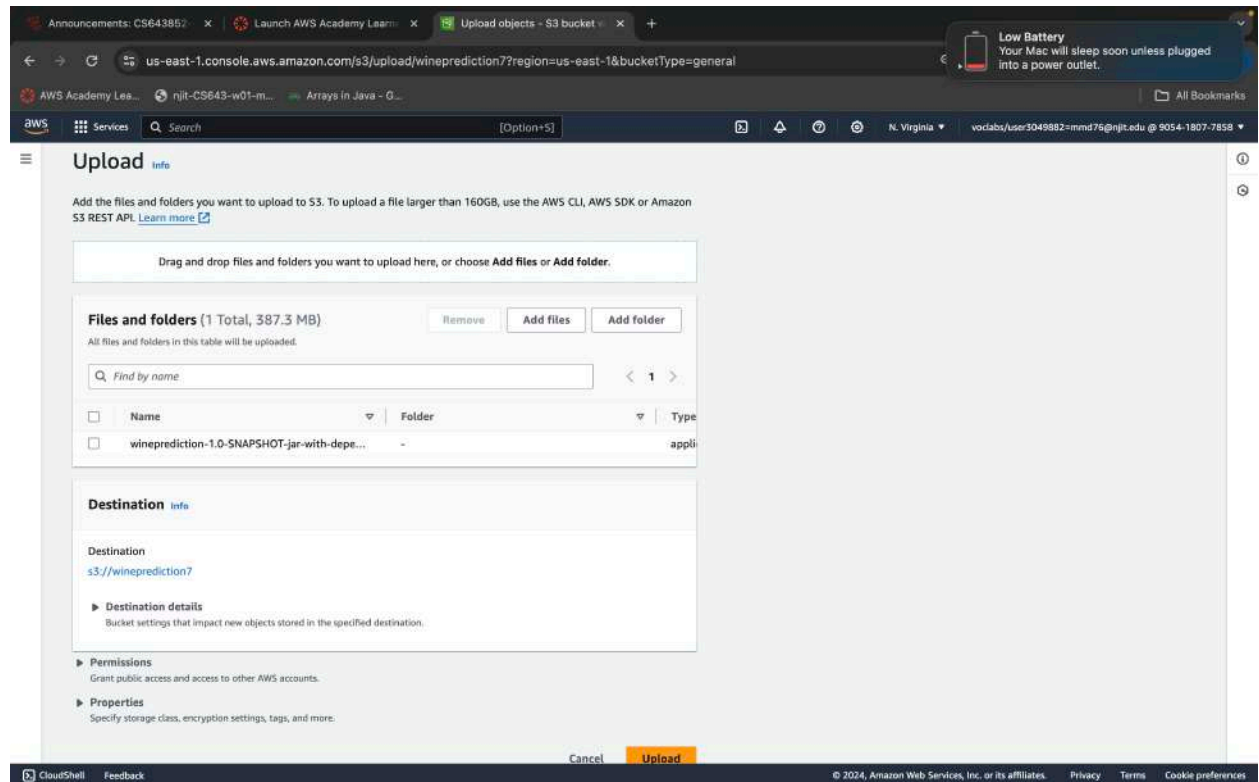


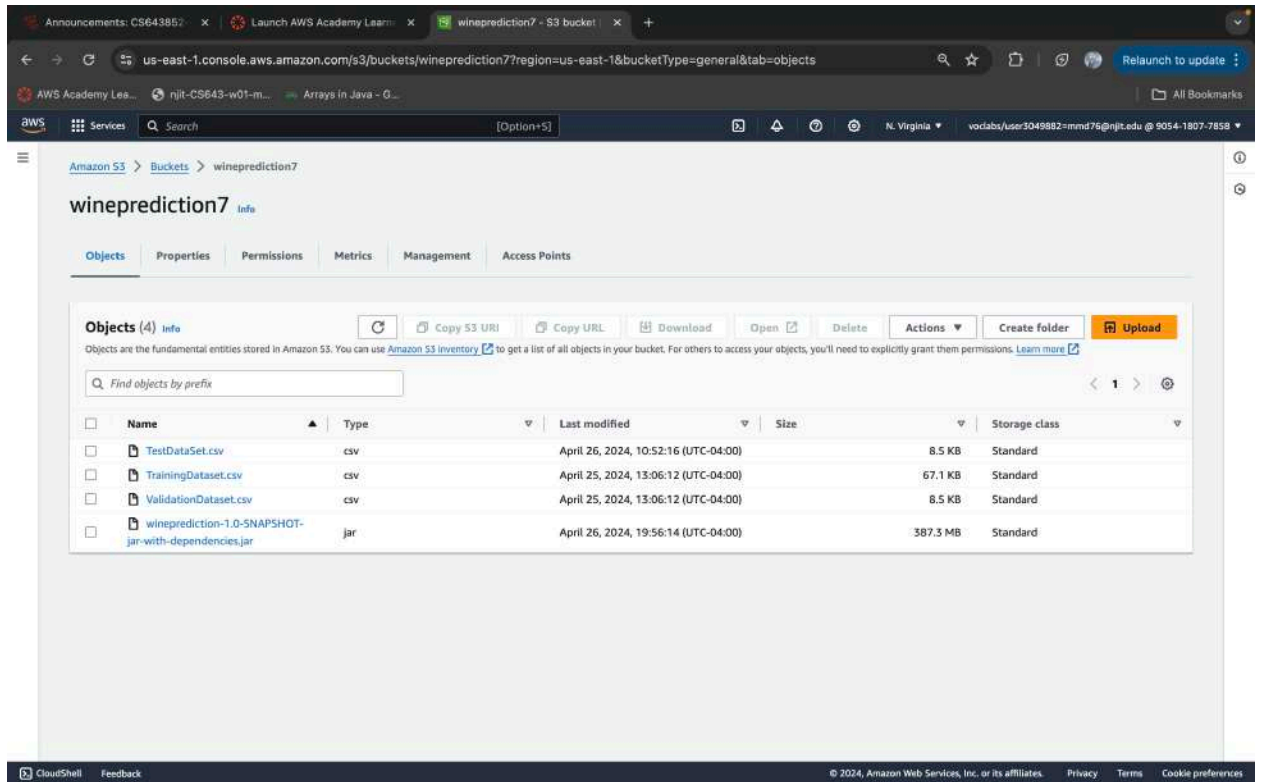


## 2. Create jar file of project:

Run “mvn clean package” in project directory, it will create jar file named “wineprediction-1.0-SNAPSHOT-jar-with-dependencies” in target folder.

Upload this jar file on S3 bucket that we created in previous step:





### 3. Submit job to apache-spark cluster:

Run below command in terminal in which have connected our cluster:

**spark-submit s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar**

```
..wineprediction-model-training --zsh ..pace/cloud-roadmap/wineapp --zsh --zsh ... ..doop@ip-172-31-63-158:~ --zsh ... -78-9.compute-1.amazonaws.com
[hadoop@ip-172-31-68-46 ~]$ spark-submit s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar
Apr 27, 2024 12:01:38 AM org.apache.spark.launcher.Log4jHotPatchOption staticJavaAgentOption
WARNING: spark.log4jHotPatch.enabled is set to true, but /usr/share/log4j-cve-2021-44228-hotpatch/jdk17/Log4jHotPatchFat.jar does not exist at the configured location

24/04/27 00:01:46 INFO SparkContext: Running Spark version 3.5.0-amzn-1
24/04/27 00:01:46 INFO SparkContext: OS info Linux, 6.1.84-99.169.amzn2023.x86_64, amd64
24/04/27 00:01:46 INFO SparkContext: Java version 17.0.8
24/04/27 00:01:46 INFO ResourceUtils:
=====
24/04/27 00:01:46 INFO ResourceUtils: No custom resources configured for spark.driver.
24/04/27 00:01:46 INFO ResourceUtils:
=====
24/04/27 00:01:46 INFO SparkContext: Submitted application: Wine Quality Prediction App
24/04/27 00:01:46 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 4, script: , vendor: , memory -> name: memory, amount: 3872, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpu -> name: cpu, amount: 1.0)
24/04/27 00:01:46 INFO ResourceProfile: Limiting resource is cpu at 4 tasks per executor
24/04/27 00:01:47 INFO ResourceProfileManager: Added ResourceProfile id: 0
24/04/27 00:01:47 INFO SecurityManager: Changing view acls to: hadoop
24/04/27 00:01:47 INFO SecurityManager: Changing modify acls to: hadoop
24/04/27 00:01:47 INFO SecurityManager: Changing view acls groups to:
24/04/27 00:01:47 INFO SecurityManager: Changing modify acls groups to:
24/04/27 00:01:47 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: hadoop; groups with view permissions: EMPTY; users with modify permissions: hadoop; groups with modify permissions: EMPTY
24/04/27 00:01:47 INFO Utils: Successfully started service 'sparkDriver' on port 38711.
24/04/27 00:01:47 INFO SparkEnv: Registering MapOutputTracker
24/04/27 00:01:47 INFO SparkEnv: Registering BlockManagerMaster
24/04/27 00:01:47 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
24/04/27 00:01:47 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
24/04/27 00:01:47 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
24/04/27 00:01:47 INFO LocalDiskBlockManager: Created local directory at /mnt/tmp/blockmgr-32d71d73-dabb-4e99-a226-d388df901de
24/04/27 00:01:47 INFO MemoryStore: MemoryStore started with capacity 1848.8 MiB
24/04/27 00:01:47 INFO SparkEnv: Registering OutputCommitCoordinator
24/04/27 00:01:47 INFO SubResultCacheManager: Sub-result caches are disabled.
24/04/27 00:01:48 INFO JettyUtils: Start Jetty 0.0.0.0:4040 for SparkUI
24/04/27 00:01:48 INFO SparkUI: Successfully started service 'SparkUI' on port 4040.
24/04/27 00:01:48 INFO SparkContext: Added JAR s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar at s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar with timestamp 1724176186904
24/04/27 00:01:48 INFO Executor: Starting executor ID driver on host ip-172-31-68-46.ec2.internal
24/04/27 00:01:48 INFO Executor: OS info Linux, 6.1.84-99.169.amzn2023.x86_64, amd64
24/04/27 00:01:48 INFO Executor: Java version 17.0.8
24/04/27 00:01:48 INFO Executor: Starting executor with user classpath (userClassPathFirst = false): 'file:/usr/lib/hadoop-lzo/lib/*, file:/usr/lib/hadoop/hadoop-aws.jar, file:/usr/share/aws/aws-java-sdk/*, file:/usr/share/aws/emr/emr-security/conf, file:/usr/share/aws/emr/security/lib/*, file:/usr/share/aws/redshift/jdbc/*, file:/usr/share/aws/redshift/lib/*, file:/usr/share/aws/emr/kinesis/lib/*, file:/usr/share/aws/hmclient/lib/aws-glue-datalake-spark-client.jar, file:/usr/share/java/Hive-JSON-Serde/hive-openx-serde.jar, file:/usr/share/aws/emr/s3select/lib/emr-s3-select-spark-connector.jar, file:/docker/usr/lib/hadoop/hadoop-aws.jar, file:/docker/usr/share/aw
re/aws-java-sdk/*, file:/docker/usr/share/aws/aws-java-sdk/*, file:/docker/usr/share/aws/emr/goodies/lib/emr-spark-goodies.jar, file:/docker/usr/share/aws/emr/security/conf, file:/docker/usr/share/aws/emr/security/lib/*, file:/docker/usr/share/aws/redshift/jdbc/*, file:/docker/usr/share/aws/redshift/lib/*, file:/docker/usr/share/aws/kinesis/lib/*, file:/docker/usr/share/aws/hmclient/lib/aws-glue-datalake-spark-client.jar, file:/docker/usr/share/java/Hive-JSON-Serde/hive-openx-serde.jar, file:/docker/usr/share/aws/emr/s3select/lib/emr-s3-select-spark-connector.jar, file:/home/hadoop/conf, file:/home/hadoop/emr-s3-select-spark-connector.jar, file:/home/hadoop/*, file:/home/hadoop/aws-glue-datalake-spark-client.jar, file:/home/hadoop/hive-openx-serde.jar, file:/home/hadoop/hadoop-aws.jar, file:/home/hadoop/emr-s3-select-spark-connector.jar'
24/04/27 00:01:48 INFO Executor: Created or updated repl class loader org.apache.spark.util.MutableURLClassLoader@8324dda2 for default.
24/04/27 00:01:48 INFO Executor: Fetching s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar with timestamp 1724176186904
24/04/27 00:01:48 INFO S3NativeFileSystem: Opening 's3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar' for reading
24/04/27 00:01:48 INFO Utils: Fetching s3://wineprediction7/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar to /mnt/tmp/spark-87e5385c-c5ad-4863-b83e-1f8df4484f4/userFiles-F9bf911-3ad3-463c-b9ab-285becdecffc/fetchFileTemp1451333549833433517.tmp
24/04/27 00:01:52 INFO Executor: Adding file:/mnt/tmp/spark-87e5385c-c5ad-4863-b83e-1f8df4484f4/userFiles-F9bf911-3ad3-463c-b9ab-285becdecffc/wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar to class loader default
24/04/27 00:01:52 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 37847.
24/04/27 00:01:52 INFO NettyBlockTransferService: Server created on ip-172-31-68-46.ec2.internal:37847
24/04/27 00:01:52 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
```

We can see out submitted job using “Spark Histroy Server UI” link:

Announcements: CS643852 x Launch AWS Academy Learn x Applications > Clusterx18 > i x History Server x +

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/3W1QF555ABRSX

AWS Academy Lea njit-CS643-w01-m... Arrays in Java - G...

Services Search [Option+S]

Cluster configuration

Instance groups	Hadoop 3.3.6, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5.0	Persistent application UIs Spark History Server YARN timeline server Tez UI	Creation time April 26, 2024, 19:17 (UTC-04:00)
Capacity	1 Primary 4 Core 0 Task	Primary node public DNS ec2-34-232-78-9.compute-1.amazonaws.com	Elapsed time 47 minutes, 30 seconds

Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events Tags (0)

### Application user interfaces

Applications installed on your Amazon EMR cluster publish user interfaces (UI) as websites. You can use these to monitor cluster activity.

☒ On-cluster application UIs

On-cluster UIs are available only while your cluster is running. Use the following links to get started. To access all the application UIs, set up SSH tunneling.

☐ Persistent application UIs

Persistent UIs don't require SSH tunneling. They are hosted off of the cluster and are available for 30 days after an application ends.

### Live Application UIs

These on-cluster application UIs are available without SSH tunneling.

Application UIs

Spark History Server UI

### Application UIs on the primary node

These require SSH tunneling to be enabled.

Application	UI URL
HDFS Name Node	<a href="http://ec2-34-232-78-9.compute-1.amazonaws.com:9870/">http://ec2-34-232-78-9.compute-1.amazonaws.com:9870/</a>
Livy	<a href="http://ec2-34-232-78-9.compute-1.amazonaws.com:8998/">http://ec2-34-232-78-9.compute-1.amazonaws.com:8998/</a>
Resource Manager	<a href="http://ec2-34-232-78-9.compute-1.amazonaws.com:8088/">http://ec2-34-232-78-9.compute-1.amazonaws.com:8088/</a>

Enable an SSH connection

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Announcements: CS643852 x Launch AWS Academy Learn x Applications > Clusterx18 > x History Server x +

j-3w1qf556brsx.emrappui-prod.us-east-1.amazonaws.com/shs/

Relaunch to update

History Server

Event log directory: hdfs:///var/log/spark/apps

Last updated: 2024-04-26 20:03:49

Client local time zone: America/New\_York

Search:

Version	App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.5.0-amzn-1	local-1714176108197	Wine Quality Prediction App	2024-04-26 20:01:46	2024-04-26 20:02:32	46 s	hadoop	2024-04-26 20:02:32	<a href="#">Download</a>

Showing 1 to 1 of 1 entries

[Show incomplete applications](#)

So once this job is finished it will save the Trained model named “LinerRegressionModel” in S3 bucket:

Announcements: CS643852 x Launch AWS Academy Learn x wineprediction7 - S3 bucket x History Server x +

us-east-1.console.aws.amazon.com/s3/buckets/wineprediction7?region=us-east-1&bucketType=general&tab=objects

Relaunch to update

Amazon S3

Buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

wineprediction7

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (5)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
LogisticRegressionModel/	Folder	-	-	-
TestDataSet.csv	csv	April 26, 2024, 10:52:16 (UTC-04:00)	8.5 KB	Standard
TrainingDataset.csv	csv	April 25, 2024, 13:06:12 (UTC-04:00)	67.1 KB	Standard
ValidationDataset.csv	csv	April 25, 2024, 13:06:12 (UTC-04:00)	8.5 KB	Standard
wineprediction-1.0-SNAPSHOT-jar-with-dependencies.jar	jar	April 26, 2024, 19:56:14 (UTC-04:00)	387.3 MB	Standard

https://us-east-1.console.aws.amazon.com/s3/buckets/wineprediction7?region=us-east-1&bucketType=general&prefix=LogisticRegressionModel/&showversions=false

2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences



**Create DOCKER image for prediction app:**

- 1. Create docker file in winepredictionapp root folder:**

### 1. Create docker file in winepredictionapp root folder:

The screenshot displays an IDE interface with three main panels:

- EXPLORER (Left):** Shows the project structure for 'WINEPREDICTIONAPP'. The 'dockerfile' file is selected under the 'resources' directory.
- EDITOR (Center):** Displays the content of the 'dockerfile' file. The Dockerfile is a multi-stage build:
 

```

1 # Use an official Maven image as the base image
2 FROM maven:3.8.4-openjdk-11-slim AS build
3 # Set the working directory in the container
4 WORKDIR /winepredictionapp
5 # Copy the pom.xml and the project files to the container
6 COPY pom.xml .
7 COPY src ./src
8 # Build the application using Maven
9 RUN mvn clean package -DskipTests
10 # Use an official OpenJDK image as the base image
11 FROM openjdk:11-jre-slim
12 # Set the working directory in the container
13 WORKDIR /winepredictionapp
14 # Copy the built JAR file from the previous stage to the container
15 COPY --from=build /winepredictionapp/target/winepredictionapp-1.0-SNAPSHOT-jar-with-dependencies.jar .
16 # Set the command to run the application
17 CMD ["java", "-jar", "winepredictionapp-1.0-SNAPSHOT-jar-with-dependencies.jar"]
```
- TERMINAL (Bottom):** Shows the output of running the application. It indicates that the tests passed successfully:
 

```

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0
[INFO] Results:
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0
```

## 2. Build the docker image:

```

...prediction-model-training --zsh ... --roadmap/winepredictionapp --zsh ... --zsh ... ..doop@ip-172-31-63-188 ~ --zsh ... --78-9.compute-1.amazonaws.com +
hannandesai@hannans-mbp winepredictionapp % docker build --platform linux/amd64 -t winepredictionapp .
[*] Building 92.6s (16/16) FINISHED
docker:desktop-linux
=> [internal] load build definition from Dockerfile
=> => transferring Dockerfile: 297B
=> [internal] load metadata for docker.io/library/openjdk:11-jre-slim
=> [internal] load metadata for docker.io/library/maven:3.8.4-openjdk-11-slim
=> [auth] library/openjdk:pull token for registry-1.docker.io
=> [auth] library/maven:pull token for registry-1.docker.io
=> [internal] load Dockerignore
=> => transferring context! 2B
=> CACHED [build 1/5] FROM docker.io/library/maven:3.8.4-openjdk-11-slim@sha256:8478e58a4a67a7b7f797948bc75ac734852872bd17f651acc5f9e2de9ba28b48
=> [internal] load build context
=> => transferring context! 11.38kB
=> CACHED [stage-1 1/3] FROM docker.io/library/openjdk:11-jre-slim@sha256:93af7df2380c831a751ca830a86c8717db182b3c31fb12aa79db842dc472a82
=> [stage-1 2/3] WORKDIR /winepredictionapp
=> [build 2/5] WORKDIR /winepredictionapp
=> [build 3/5] COPY pom.xml .
=> [build 4/5] COPY src /src
=> [build 4/5] RUN mvn clean package -DskipTests
=> [stage-1 3/3] COPY --from=build /winepredictionapp/target/winepredictionapp-1.0-WINAPSHOT-jar-with-dependencies.jar .
=> exporting to image
=> => exporting layers
=> => writing image sha256:c8dc836f546c257faaf49c5673ac649e32ea7eaf78234448c5cd8d1b9e91e64
=> => naming to docker.io/library/winepredictionapp
Build multi-platform images faster with Docker Build Cloud: https://docs.docker.com/go/docker-build-cloud
hannandesai@hannans-mbp winepredictionapp % docker image ls
REPOSITORY          TAG             IMAGE ID        CREATED         SIZE
winepredictionapp    latest          c8ac8a36f545    19 seconds ago  429MB
hannan9988/wineapp  latest         2cb9222dd9d4    8 hours ago     429MB
wineapp              latest         2cb9222dd9d4    8 hours ago     429MB
hannandesai/wineapp latest         36dda86e9e5e    9 hours ago     422MB
hannan9988/wineapp  <none>         36dda86e9e5e    9 hours ago     422MB
wineprediction       latest         5fc1b374e93e    28 hours ago    422MB
<none>              <none>         421a885a8cfc    28 hours ago    422MB
hannandesai@hannans-mbp winepredictionapp %

```



### 3. Push to created image to public docker hub:

```
hannandesai@hannans-mbp winepredictionapp % docker push hannan9900/winepredictionapp
Using default tag: latest
The push refers to repository [docker.io/hannan9900/winepredictionapp]
7d848e7fdebb: Pushed
a584edf95513: Pushed
d7882b858aef: Mounted from hannan9900/wineapp
a3abdc2e9252: Mounted from hannan9900/wineapp
ea7ee832dbd: Mounted from hannan9900/wineapp
92a4e8a3140f: Mounted from hannan9900/wineapp
latest: digest: sha256:18f2ff4fbc91589ee85ab94c3cfa2ee9db9eae5948628c36bf9971561d72b67 size: 1579
hannandesai@hannans-mbp winepredictionapp %
```

Announcements: CS643... Launch AWS Academy Le... wineprediction7 - S3 bucl... History Server... hannan9900/winepredic... hannan9900/winepredic... +

hub.docker.com/repository/docker/hannan9900/winepredictionapp/general

dockerhub Explore Repositories Organizations Search Docker Hub

hannan9900 / Repositories / winepredictionapp / General Using 0 of 1 private repositories. [Get more](#)

General Tags Builds Collaborators Webhooks Settings

### hannan9900/winepredictionapp

Updated 1 minute ago

This repository does not have a description [✎](#) INCOMPLETE

This repository does not have a category [✎](#) INCOMPLETE

### Docker commands

To push a new tag to this repository:

```
docker push hannan9900/winepredictionapp:tagname
```

[Public View](#)

### Tags

This repository contains 1 tag(s).

Tag	OS	Type	Pulled	Pushed
latest	linux	Image	a minute ago	a minute ago

[See all](#)

### Automated Builds

Manually pushing images to Hub? Connect your account to GitHub or Bitbucket to automatically build and tag new images whenever your code is updated, so you can focus your time on creating.

Available with Pro, Team and Business subscriptions. [Read more about automated builds](#).

[Upgrade](#)

### Repository overview

INCOMPLETE

By clicking "Accept All Cookies", you agree to the storing of cookies on your device to enhance site navigation, analyze site usage, and assist in our marketing efforts.

[Cookies Settings](#) [Reject All](#) [Accept All Cookies](#) [X](#)

### Run this docker image in AWS EC2 instance:

## 1. Connect to an EC2 instance:

[illegible]

## 2. Install docker on instance:

```
--reproduction-model-training -- ~$  
--roadmap/winepredictionapp -- ~$  
--78-9.compute-1.amazonaws.com  
--78-9.compute-1.amazonaws.com
```

```
[ec2-user@ip-172-31-68-46 ~]$ docker  
AC  
[ec2-user@ip-172-31-68-46 ~]$ sudo yum install docker  
Last metadata expiration check: 1:45:24 ago on Fri Apr 26 23:18:22 2024.  
Package docker-25.0.3-1.eln2823.0.i.x86_64 is already installed.  
Dependencies resolved.  
Nothing to do.  
Complete!  
[ec2-user@ip-172-31-68-46 ~]$ sudo service docker start  
Redirecting to /bin/systemctl start docker.service  
[ec2-user@ip-172-31-68-46 ~]
```

### 3. Fetch docker image from docker hub to instance:

```
...ineprediction-model-training -- -zsh | ...-roadmap/winepredictionapp -- -zsh | ... -zsh | ... | ...-78-9.compute-1.amazonaws.com | ...-78-9.compute-1.amazonaws.com |
[ec2-user@ip-172-31-60-46 ~]$ sudo docker pull hannan9908/winepredictionapp
Using default tag: latest
latest: Pulling from hannan9908/winepredictionapp
1efc276f4ff9: Pull complete
5f2f293dc482: Pull complete
12cc092b13c: Pull complete
d73cf48caaac: Pull complete
6d3aa8ab05ee: Pull complete
69aec5b28f3c: Pull complete
Digest: sha256:10f2ffa1be91589e8e6b994c3cfa3ee9ddb9eae5948628c36bf9971561d72b87
Status: Downloaded newer image for hannan9908/winepredictionapp:latest
docker.io/hannan9908/winepredictionapp:latest
[ec2-user@ip-172-31-60-46 ~]$ sudo docker image ls
REPOSITORY          TAG             IMAGE ID        CREATED         SIZE
hannan9908/winepredictionapp   latest         c8ac8436f545   22 minutes ago  629MB
[ec2-user@ip-172-31-60-46 ~]$
```

### 4. Run Docker image:

```
...ineprediction-model-training -- -zsh | ...-roadmap/winepredictionapp -- -zsh | ... -zsh | ... | ...-78-9.compute-1.amazonaws.com | ...-78-9.compute-1.amazonaws.com |
[ec2-user@ip-172-31-60-46 ~]$ docker run hannan9908/winepredictionapp
WARNING: sun.reflect.Reflection.getCallerClass is not supported. This will impact performance.
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
24/04/27 02:05:04 INFO SparkContext: Running Spark version 3.5.1
24/04/27 02:05:04 INFO SparkContext: OS info Linux, 6.1.84-99.169.amzn2023.x86_64, amd64
24/04/27 02:05:04 INFO SparkContext: Java version 11.0.16
24/04/27 02:05:04 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/04/27 02:05:04 INFO ResourceUtils:
=====
24/04/27 02:05:04 INFO ResourceUtils: No custom resources configured for spark.driver.
24/04/27 02:05:04 INFO ResourceUtils:
=====
24/04/27 02:05:04 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 3872, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
24/04/27 02:05:04 INFO ResourceProfile: Limiting resource is cpu
24/04/27 02:05:04 INFO ResourceProfileManager: Added ResourceProfile id: 0
24/04/27 02:05:04 INFO SecurityManager: Changing view acls to: root
24/04/27 02:05:04 INFO SecurityManager: Changing modify acls to: root
24/04/27 02:05:04 INFO SecurityManager: Changing view acls groups to:
24/04/27 02:05:04 INFO SecurityManager: Changing modify acls groups to:
24/04/27 02:05:04 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: root; groups with view permissions: EMPTY; users with modify permissions: r
oot; groups with modify permissions: EMPTY
24/04/27 02:05:04 INFO Utils: Successfully started service 'sparkDriver' on port 35733.
24/04/27 02:05:04 INFO SparkEnv: Registering MapOutputTracker
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file: /winepredictionapp/winepredictionapp-1.0-SNAPSHOT-jar-with-dependencies.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/04/27 02:05:04 INFO SparkEnv: Registering BlockManagerMaster
24/04/27 02:05:04 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
24/04/27 02:05:04 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
24/04/27 02:05:04 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
24/04/27 02:05:04 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-42cb4740-ef34-4e6f-b812-51217089a893
24/04/27 02:05:04 INFO MemoryStore: MemoryStore started with capacity 2.2 GiB
24/04/27 02:05:04 INFO SparkEnv: Registering OutputCommitCoordinator
24/04/27 02:05:05 INFO JettyUtils: Start Jetty 9.4.44.v20210824 for SparkUI
24/04/27 02:05:05 INFO Utils: Successfully started service 'SparkUI' on port 4040.
24/04/27 02:05:05 INFO Executor: Starting executor ID driver on host a60f5b01d74d
24/04/27 02:05:05 INFO Executor: OS info Linux, 6.1.84-99.169.amzn2023.x86_64, amd64
24/04/27 02:05:05 INFO Executor: Java version 11.0.16
24/04/27 02:05:05 INFO Executor: Starting executor with user classpath (userClassPathFirst = false): ''
24/04/27 02:05:05 INFO Executor: Created or updated repl class loader org.apache.spark.util.MutableURLClassLoader@1f11f6ae for default.
24/04/27 02:05:05 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 39513.
24/04/27 02:05:05 INFO NettyBlockTransferService: Server created on a60f5b01d74d:39513
24/04/27 02:05:05 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
24/04/27 02:05:05 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, a60f5b01d74d, 39513, None)
24/04/27 02:05:05 INFO BlockManagerMasterEndpoint: Registering block manager a60f5b01d74d:39513 with 2.2 GiB RAM, BlockManagerId(driver, a60f5b01d74d, 39513, None)
24/04/27 02:05:05 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, a60f5b01d74d, 39513, None)
24/04/27 02:05:05 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, a60f5b01d74d, 39513, None)
TestingDataSet Metrics
24/04/27 02:05:06 INFO MemoryStore: Block broadcast_0 stored as values in memory (estimated size 291.8 KiB, free 2.2 GiB)
24/04/27 02:05:06 INFO MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 47.6 KiB, free 2.2 GiB)
24/04/27 02:05:06 INFO BlockManagerInfo: Added broadcast_0_piece0 in memory on a60f5b01d74d:39513 (size: 47.6 KiB, free: 2.2 GiB)
24/04/27 02:05:06 INFO SparkContext: Created broadcast 0 from textFile at ReadWrite.scala:587
24/04/27 02:05:06 WARN MetricsConfig: Cannot locate configuration: tried hadoop-metrics2-s3a-file-system.properties,hadoop-metrics2.properties
24/04/27 02:05:06 INFO MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
```

To run this prediction app without docker:

Run following command in project root directory:

mvn clean package

mvn exec:java -

Dexec.mainClass="org.example.winepredictionapp.App";

```
...wineprediction-model-training --zsh ...roadmap/winepredictionapp --zsh --zsh ...2-user@ip-172-31-60-46:~ --zsh ...adoop@ip-172-31-60-46:~ --zsh
hannandesai@hannans-mbp winepredictionapp % mvn exec:java -Dexec.mainClass="org.example.winepredictionapp.App";
[INFO] Scanning for projects...
[INFO]
[INFO] ----- org.example.winepredictionapp:winepredictionapp -----
[INFO] Building winepredictionapp 1.0-SNAPSHOT
[INFO] from pom.xml
[INFO]
[INFO] [ jar ]
[INFO]
[INFO] ----- exec:3.1.1:java (default-cli) @ winepredictionapp -----
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@f47f286]
SLF4J(W): Found provider [org.slf4j.simple.SimpleServiceProvider@6ca3649f]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@f47f286]
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
24/04/26 23:49:02 INFO SparkContext: Running Spark version 3.5.1
24/04/26 23:49:02 INFO SparkContext: OS info Mac OS X, 14.4.1, aarch64
24/04/26 23:49:02 INFO SparkContext: Java version 21.0.1
24/04/26 23:49:02 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/04/26 23:49:02 INFO ResourceUtils:
=====
24/04/26 23:49:02 INFO ResourceUtils: No custom resources configured for spark.driver.
24/04/26 23:49:02 INFO ResourceUtils:
=====
24/04/26 23:49:02 INFO SparkContext: Submitted application: Wine Quality Prediction App
24/04/26 23:49:02 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 3872, script: , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
24/04/26 23:49:02 INFO ResourceProfile: Limiting resources is cpu
24/04/26 23:49:02 INFO ResourceProfileManager: Added ResourceProfile id: 0
24/04/26 23:49:02 INFO SecurityManager: Changing view acls to: hannandesai
24/04/26 23:49:02 INFO SecurityManager: Changing modify acls to: hannandesai
24/04/26 23:49:02 INFO SecurityManager: Changing view acls groups to:
24/04/26 23:49:02 INFO SecurityManager: Changing modify acls groups to:
24/04/26 23:49:02 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: hannandesai; groups with view permissions: EMPTY; users with modify permissions: hannandesai; groups with modify permissions: EMPTY
24/04/26 23:49:02 INFO Utils: Successfully started service 'sparkDriver' on port 51754.
24/04/26 23:49:02 INFO SparkEnv: Registering MapOutputTracker
24/04/26 23:49:02 INFO SparkEnv: Registering BlockManagerMaster
24/04/26 23:49:02 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
24/04/26 23:49:02 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
24/04/26 23:49:02 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
24/04/26 23:49:02 INFO DiskBlockManager: Created local directory at /private/var/folders/gc/y19vqzn2q58jx9w5td5w8c80000gn/T/blockmgr-b056d7f8-0007-4a74-8421-36e75c82c8a9
24/04/26 23:49:02 INFO MemoryStore: MemoryStore started with capacity 1848.8 MiB
24/04/26 23:49:02 INFO SparkEnv: Registering OutputCommitCoordinator
24/04/26 23:49:02 INFO JettyUtils: Start Jetty 8.0.0.0:4040 for SparkUI
24/04/26 23:49:02 INFO Utils: Successfully started service 'SparkUI' on port 4040.
24/04/26 23:49:02 INFO Executor: Starting executor ID driver on host hannans-mbp
24/04/26 23:49:02 INFO Executor: OS info Mac OS X, 14.4.1, aarch64
24/04/26 23:49:02 INFO Executor: Java version 21.0.1
24/04/26 23:49:02 INFO Executor: Starting executor with user classpath (userClassPathFirst = false): ''
24/04/26 23:49:02 INFO Executor: Created or updated repl class loader org.apache.spark.util.MutableURLClassLoader@fd77d2 for default.
24/04/26 23:49:03 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 51755.
24/04/26 23:49:03 INFO NettyBlockTransferService: Server created on hannans-mbp:51755
24/04/26 23:49:03 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
24/04/26 23:49:03 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, hannans-mbp, 51755, None)
24/04/26 23:49:03 INFO BlockManagerMasterEndpoint: Registering block manager hannans-mbp:51755 with 1848.8 MiB RAM, BlockManagerId(driver, hannans-mbp, 51755, None)
24/04/26 23:49:03 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, hannans-mbp, 51755, None)
24/04/26 23:49:03 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, hannans-mbp, 51755, None)
TestingDataSet Metrics
```