

Workflow for Programming Assignment -2

28 Steps

Created by

Rama Chetan Atmudi

Creation Date

November 3, 2022

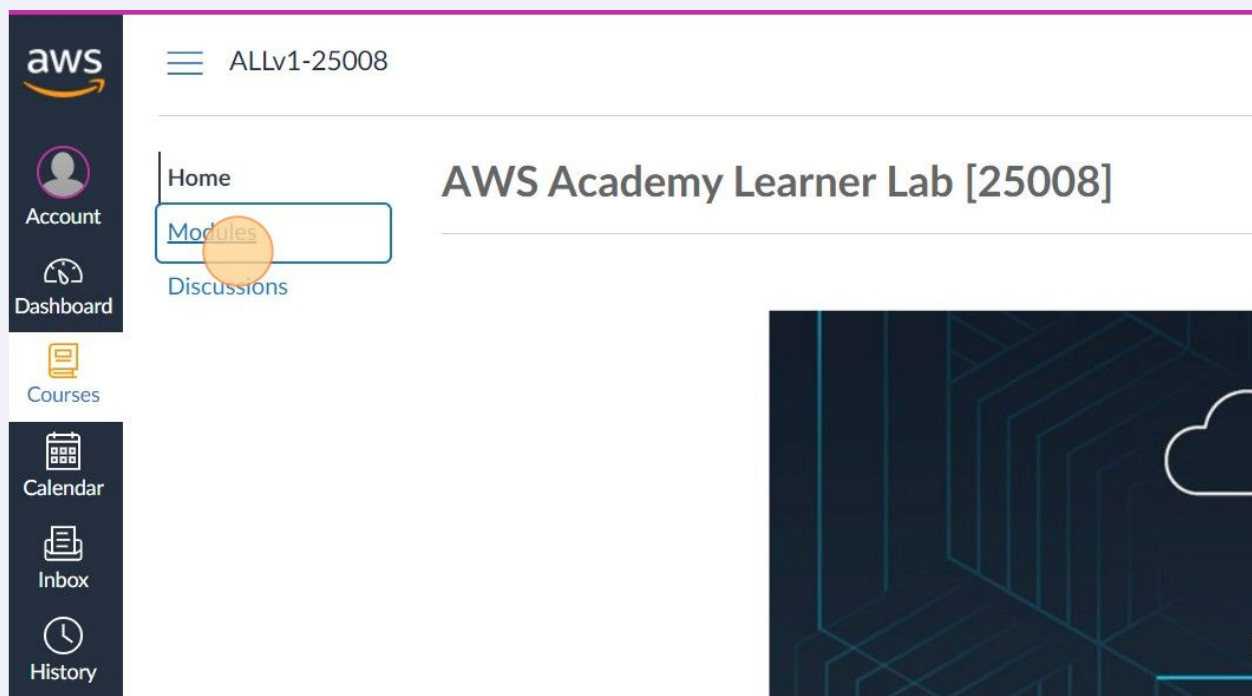
Last Updated

November 3, 2022

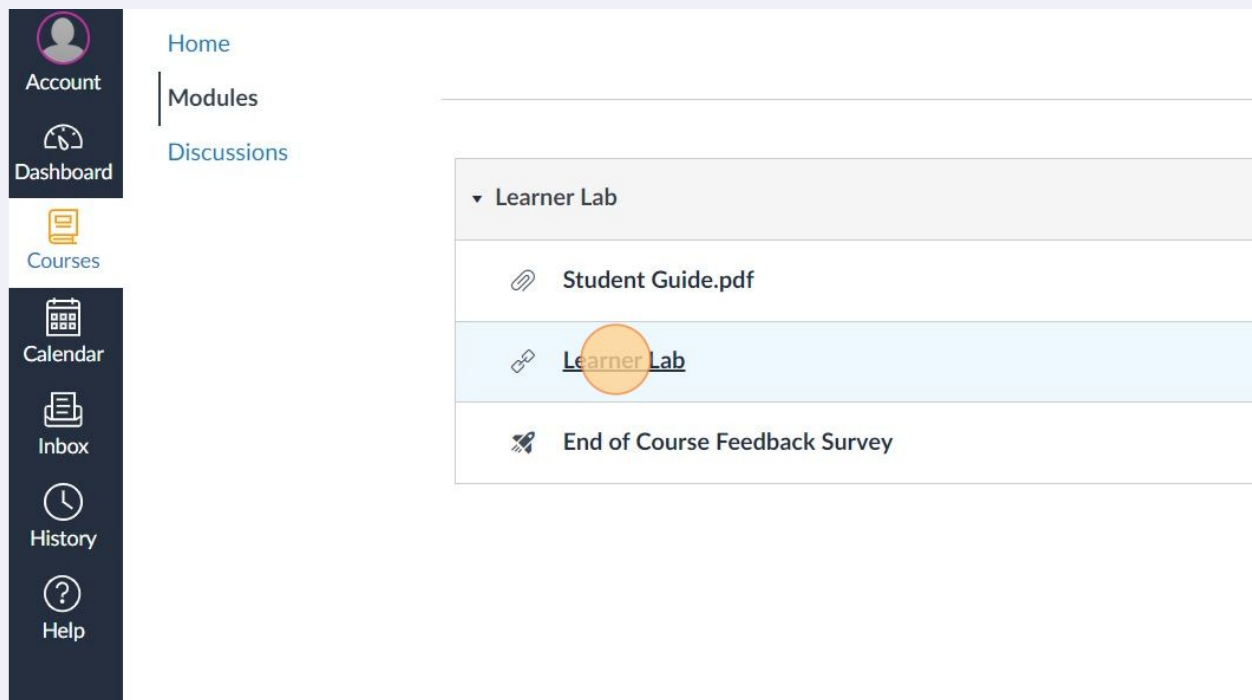
Programming Assignment -2 Workflow

1 Navigate to <https://awsacademy.instructure.com/courses/25008>

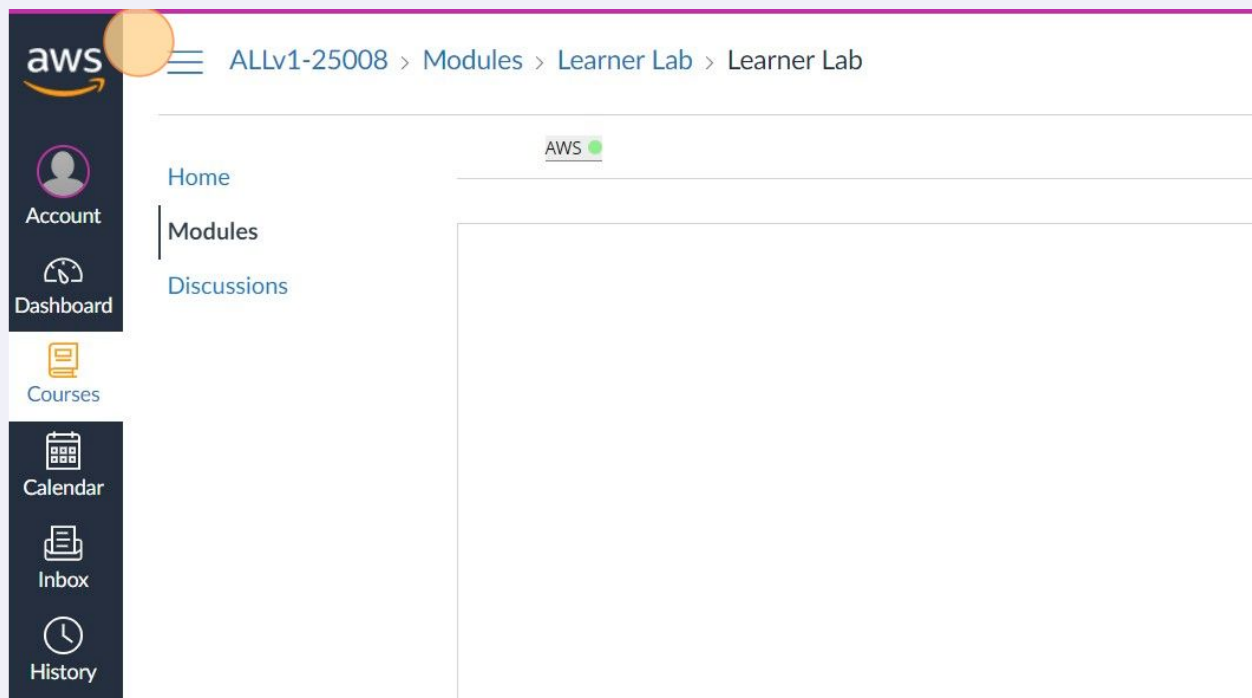
2 Click "Modules"



3 Click "Learner Lab"

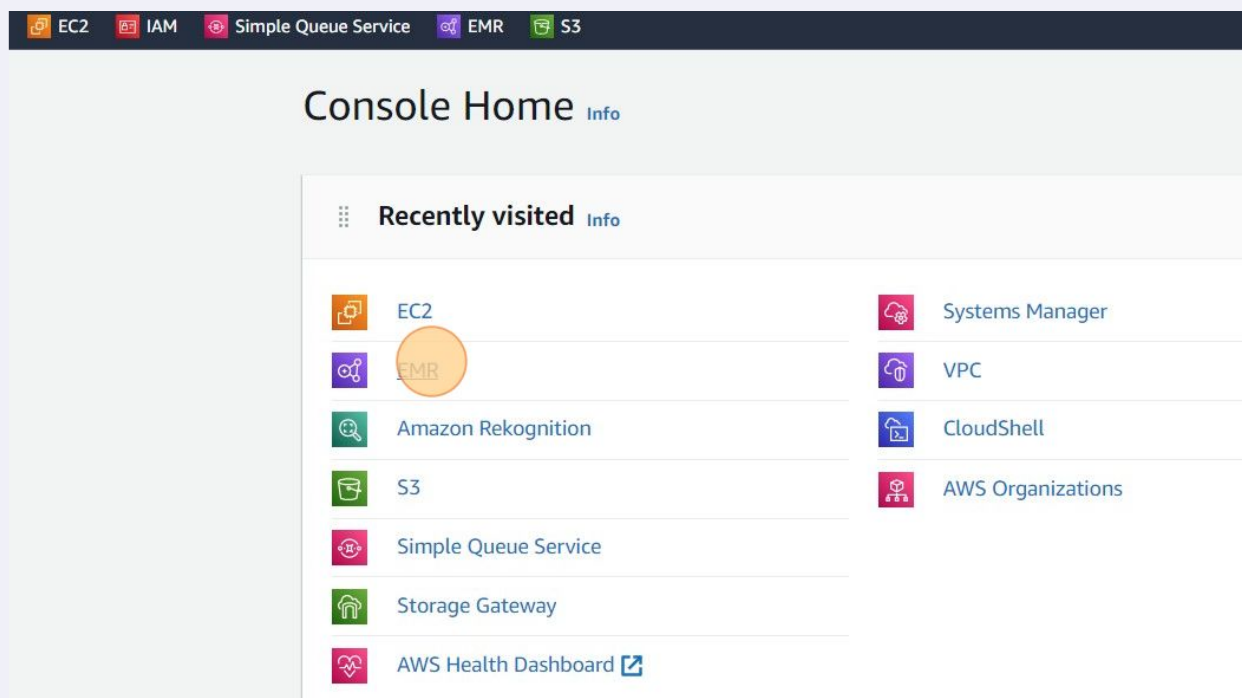


4 Start your lab session and Click "AWS"



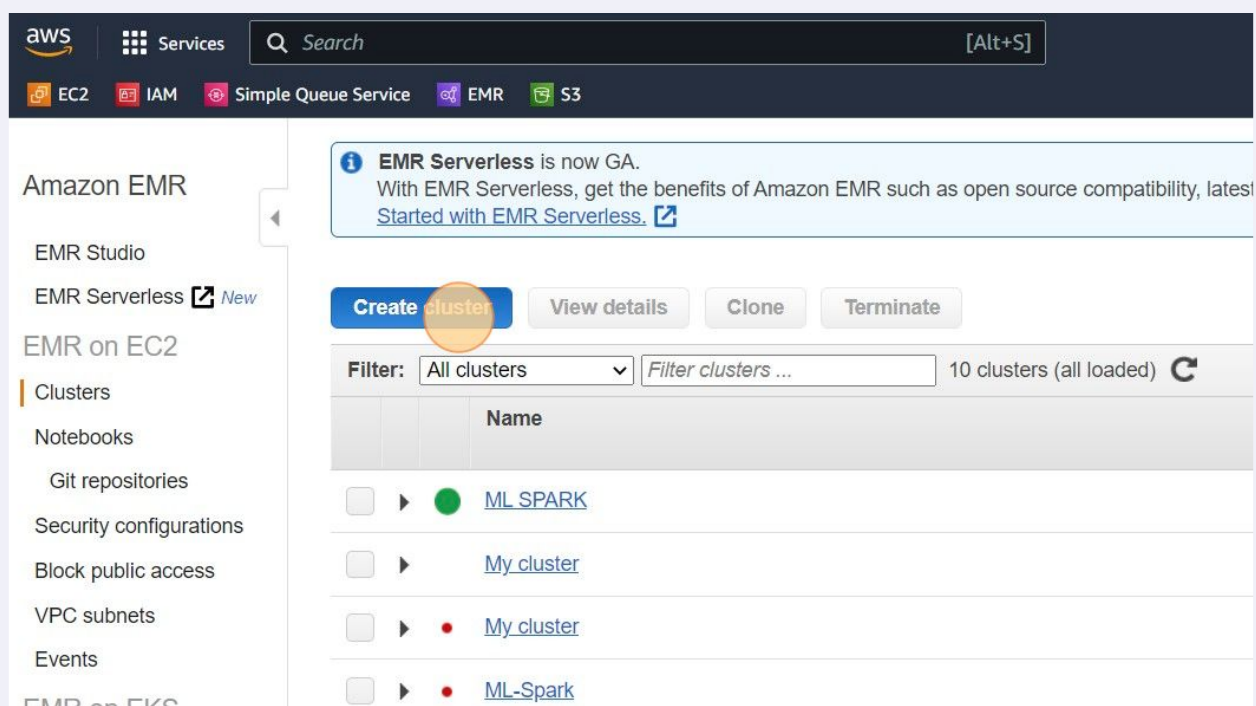
5

Search for "EMR" in the Services section of your AWS Management Console and then click "EMR." In the screenshot shown below I accessed EMR from "Recently Visited" tab.



6

Click "Create cluster"



7 Name your Cluster

[REDACTED]

as open source compatibility, latest versions and performance optimized runtime for popular frameworks along with easy provisioning, qu

[to advanced options](#)

General Configuration

Cluster name

☒ Logging ⓘ

S3 folder ⓘ

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release ⓘ

Applications ☒ Core Hadoop: Hadoop 2.10.1, Hive 2.3.9, Hue ⓘ

8 I named my Cluster as "ML Spark"

- 9 Click on "Spark" in Applications as we are dealing with Spark in this assignment.

Software configuration

Release ⓘ

- Applications
- ☒ Core Hadoop: Hadoop 2.10.1, Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.2
 - ☐ HBase: HBase 1.4.13, Hadoop 2.10.1, Hive 2.3.9, Hue 4.10.0, Phoenix 4.14.3, and ZooKeeper 3.4.14
 - ☐ Presto: Presto 0.267 with Hadoop 2.10.1 HDFS and Hive 2.3.9 Metastore
 - ☒ Spark: Spark 2.4.8 on Hadoop 2.10.1 YARN and Zeppelin 0.10.0

☐ Use AWS Glue Data Catalog for table metadata ⓘ

Hardware configuration

Instance type ⓘ The selected storage per instance

Number of instances (1 master and 2 core nodes)

Cluster scaling ☐ scale cluster nodes based on workload

- 10 We need to run Spark in 4 EC2 Instances so, create 5 Instances 1 master and 4 core nodes

and Hive 2.3.9 Metastore

- ☒ Spark: Spark 2.4.8 on Hadoop 2.10.1 YARN and Zeppelin 0.10.0

☐ Use AWS Glue Data Catalog for table metadata ⓘ

Hardware configuration

Instance type ⓘ The selected instance storage per instance

Number of instances (1 master and 2 core nodes)

Cluster scaling ☐ scale cluster nodes based on workload

Auto-termination ☒ Enable auto-termination [Learn more](#) ⓘ

Terminate cluster when it is idle after hours minutes


Security and access

EC2 key pair ⓘ [Learn how](#)


Permissions ☒ Default ☐ Custom

11 Enter "5" in Number of Instances

12 Disable "Auto-Termination"


☐ Use AWS Glue Data Catalog for table metadata 

Hardware configuration

Instance type m5.xlarge  The selected instance type has 16 vCPUs and 32 GiB of memory.



Number of instances 5 (1 master and 4 core nodes)

Cluster scaling ☐ scale cluster nodes based on workload

Auto-termination ☒ Enable auto-termination [Learn more](#) 

Terminate cluster when it is idle after 1 hours 0 minutes

Security and access

EC2 key pair Choose an option   [Learn how to choose a 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.

☐ Use EMR_DefaultRole as default role

13 Create a New Key Pair for this cluster to access it.

Hardware configuration



Instance type m5.xlarge  The selected instance type adds 64 GiB of GP storage per instance by default. [Learn more](#)

Number of instances 5 (1 master and 4 core nodes)



Cluster scaling ☐ scale cluster nodes based on workload



Auto-termination ☐ Enable auto-termination [Learn more](#)

Security and access

EC2 key pair Choose an option   [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](#)  ☐ Use EMR_DefaultRole_V2 

EC2 instance profile [EMR_EC2_DefaultRole](#)  


Cancel Create cluster

14 After you setup everything you are ready to "Create cluster"



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

ent, they will be automatically created
c policy updates.

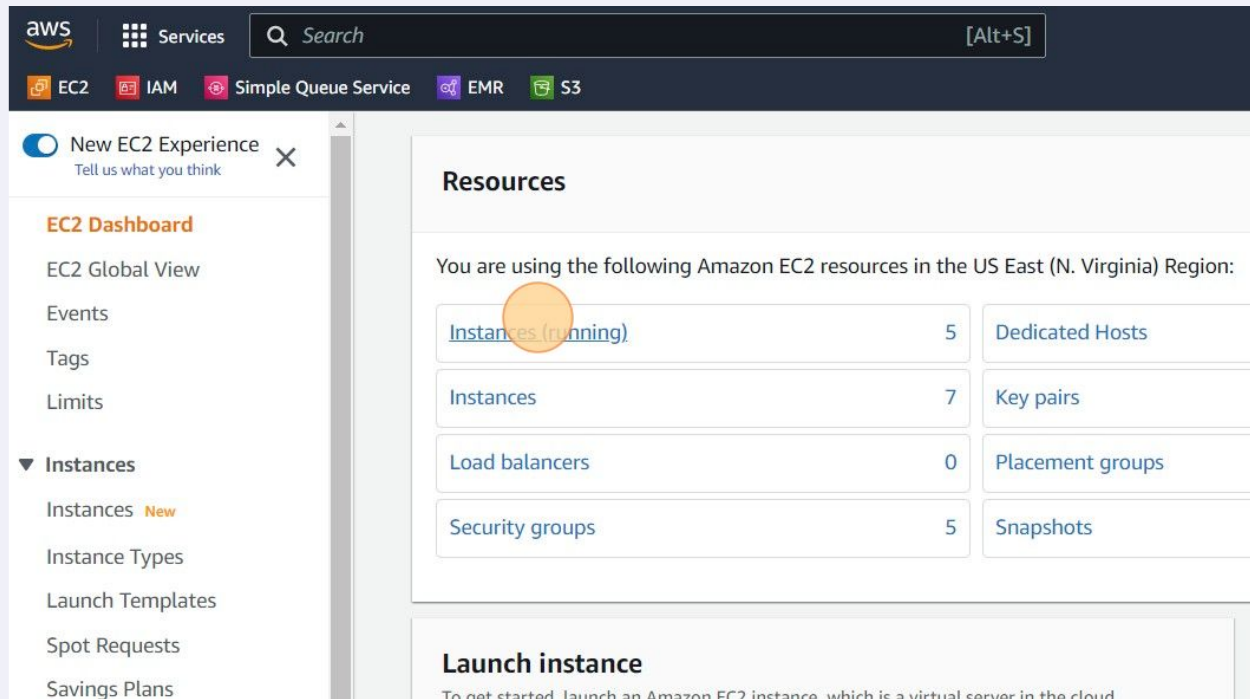
efaultRole_V2 

Cancel Create cluster

© 2022, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#)

15

After you have configured your cluster, navigate to the EC2 Dashboard and you will notice 5 new instances created.



Resources

You are using the following Amazon EC2 resources in the US East (N. Virginia) Region:

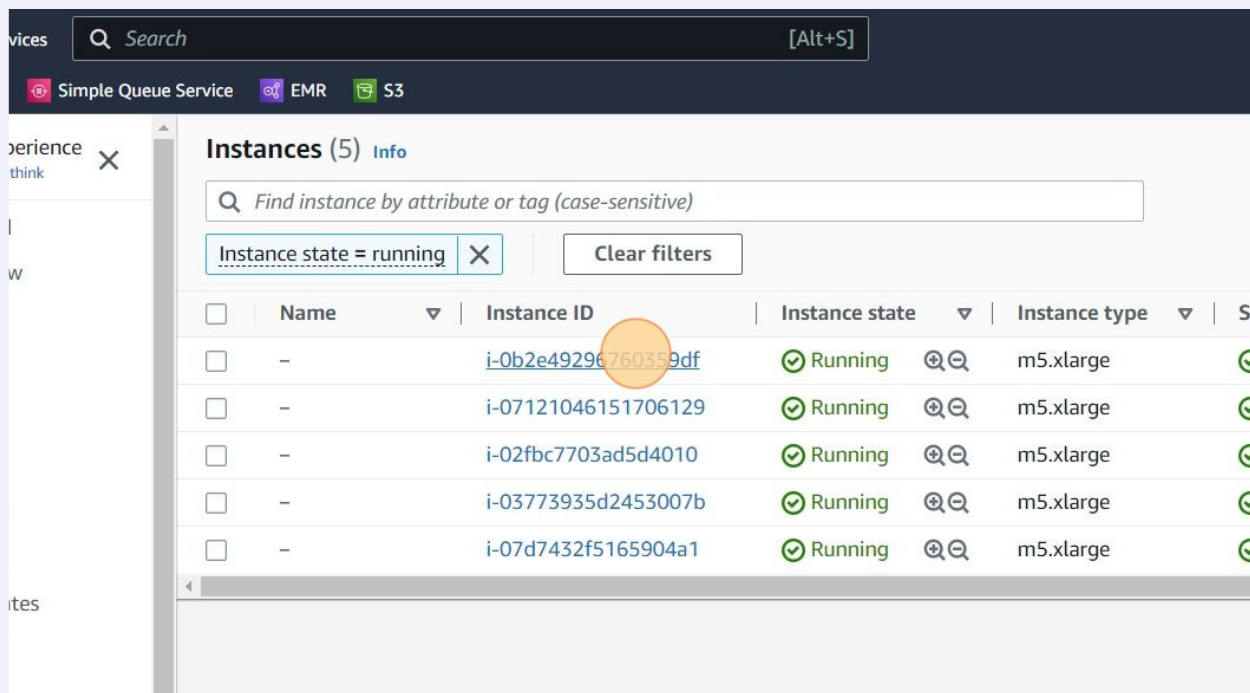
Instances (running)	5	Dedicated Hosts
Instances	7	Key pairs
Load balancers	0	Placement groups
Security groups	5	Snapshots

Launch instance

To get started, launch an Amazon EC2 instance, which is a virtual server in the cloud.

16

Select the first instance you created.



Instances (5) Info

Find instance by attribute or tag (case-sensitive)

Instance state = running X Clear filters

	Name	Instance ID	Instance state	Instance type	Size
<input type="checkbox"/>	-	i-0b2e49296760359df	Running	m5.xlarge	✓
<input type="checkbox"/>	-	i-07121046151706129	Running	m5.xlarge	✓
<input type="checkbox"/>	-	i-02fbc7703ad5d4010	Running	m5.xlarge	✓
<input type="checkbox"/>	-	i-03773935d2453007b	Running	m5.xlarge	✓
<input type="checkbox"/>	-	i-07d7432f5165904a1	Running	m5.xlarge	✓

17 Click "Security"

The screenshot shows the AWS Management Console interface. On the left, a navigation menu lists various services: Instance Types, Launch Templates, Spot Requests, Savings Plans, Reserved Instances (marked 'New'), Dedicated Hosts, Scheduled Instances, Capacity Reservations, Images (with a dropdown arrow), AMIs (marked 'New'), AMI Catalog, Elastic Block Store (with a dropdown arrow), Volumes, and Snapshots. The main content area displays details for an EC2 instance. At the top, it shows the IP name 'ip-172-31-34-126.ec2.internal', the answer private resource DNS name as '-', and the auto-assigned IP address '54.87.146.62 [Public IP]'. Below this, the IAM Role is listed as 'EMR_EC2_DefaultRole' with an external link icon. A horizontal tab bar at the bottom of the main content area includes 'Details', 'Security' (highlighted with an orange circle), 'Networking', 'Storage', 'Status checks', and 'Monitoring'. Under the 'Security' tab, the 'Instance details' section is expanded, showing the platform as 'Linux/UNIX (Inferred)' and platform details as 'Linux/UNIX'. It also indicates that 'Stop protection' is 'Disabled'.

Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances **New**
Dedicated Hosts
Scheduled Instances
Capacity Reservations

▼ Images
AMIs **New**
AMI Catalog

▼ Elastic Block Store
Volumes
Snapshots

IP name: ip-172-31-34-126.ec2.internal
Answer private resource DNS name
-
Auto-assigned IP address
54.87.146.62 [Public IP]
IAM Role
EMR_EC2_DefaultRole

Details | **Security** | Networking | Storage | Status checks | Monitoring

▼ Instance details Info
Platform
Linux/UNIX (Inferred)
Platform details
Linux/UNIX
Stop protection
Disabled

18 Click the link below Security Groups as shown in the screenshot.

This screenshot shows the 'Security' tab of the AWS Management Console for an EC2 instance. The left navigation menu is similar to the previous screenshot but includes 'Events', 'Tags', and 'Limits' under the 'Instances' section. The main content area shows the 'Security details' section, which lists the IAM Role as 'EMR_EC2_DefaultRole' and the security groups as 'sg-043e4865523061cbe (ElasticMapReduce-master)'. A red circle highlights the link to this security group. Below this, the 'Inbound rules' section is expanded, showing a table of rules. The table has columns for 'Security group rule ID', 'Port range', 'Protocol', and 'Source'. Three rules are listed: one for UDP on port range 0-65535, and two for TCP on port 8443.

Events
Tags
Limits

▼ Instances
Instances **New**
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances **New**
Dedicated Hosts
Scheduled Instances
Capacity Reservations

▼ Images
AMIs **New**

IAM Role
EMR_EC2_DefaultRole

Details | **Security** | Networking | Storage | Status checks | Monitoring

▼ Security details
IAM Role
EMR_EC2_DefaultRole
Security groups
sg-043e4865523061cbe (ElasticMapReduce-master)

▼ Inbound rules

Security group rule ID	Port range	Protocol	Source
sgr-0d8dc1b00bbadcbfb	0 - 65535	UDP	sg-0...
sgr-0a55c0b9674a3b300	8443	TCP	72.2...
sgr-05474d165e1dad124	8443	TCP	72.2...

19 Click "Edit inbound rules"

Run Reachability Analyzer X

Manage tags Edit inbound rules

< 1 > ⚙

▼	Port range	▼	Source	▼	Description	▼
	0 - 65535		sg-043e4865523061c...		-	
	8443		72.21.217.0/24		-	
	8443		72.21.196.64/29		-	
	8443		207.171.167.26/32		-	

20 Click "Add rule"

Custom TCP TCP

sgr-0ac51fdb09d145b1f Custom TCP TCP

sgr-08faab197f378e657 Custom TCP TCP

sgr-0af68990f09bd53e0 All UDP UDP

Add rule



21 Click "Custom TCP"

The screenshot shows a firewall rule configuration interface. The 'Protocol' dropdown menu is open, showing options like Custom TCP, TCP, UDP, and All traffic. An orange circle highlights the 'Custom TCP' option. The interface also shows a list of rules with their IDs, protocols, and ports.

Rule ID	Protocol	Port
sgr-0ac51fdb09d145b1f	Custom TCP	8443
sgr-08faab197f378e657	Custom TCP	8443
sgr-0af68990f09bd53e0	All UDP	0 - 65535
-	Custom TCP	0

[Add rule](#)

[back](#) Looking for language selection? Find it in the new [Unified Settings](#)

22 Search for "SSH"

The screenshot shows a firewall rule configuration interface. The 'Protocol' dropdown menu is open, showing a search results list. An orange circle highlights the 'SSH' option. The interface also shows a list of rules with their IDs, protocols, and ports.

Rule ID	Protocol	Port
sgr-0bac442f23e78e1b6	Custom UDP	TCP
sgr-0634f17e928c6391b	Custom ICMP - IPv4	TCP
sgr-061544937ef41e5de	Custom Protocol	TCP
sgr-0186e524a643553e8	All TCP	TCP
sgr-05bc0b71f47096130	All UDP	TCP
sgr-0ac51fdb09d145b1f	All ICMP - IPv4	TCP
	All ICMP - IPv6	ICMP
	All traffic	TCP
	SSH	TCP
	SMTP	TCP
	DNS (UDP)	TCP
	DNS (TCP)	TCP
	HTTP	TCP
	POP3	TCP
	IMAP	TCP
	LDAP	TCP
	HTTPS	TCP

23 Click "Custom"

TCP	8443	Custom	72.21.198.64/29
TCP	8443	Custom	54.240.217.80/29
TCP	8443	Custom	207.171.167.101/32
UDP	0 - 65535	Custom	sg-00bad27f201b4f285
TCP	22	Custom	

24 Change it to "My IP"

TCP	8443	Custom	72.21.198.64/29
TCP	8443	Custom	54.240.217.80/29
TCP	8443	Custom	207.171.167.101/32
UDP	0 - 65535	Custom	sg-00bad27f201b4f285
TCP	22	Custom	

Custom

Anywhere-IPv4

Anywhere-IPv6

My IP

Custom

25 Click "Save rules"

240.217.80/29 X

171.167.101/32 X

00bad27f201b4f285 X

0.35.98.17/32 X

Cancel Preview changes **Save rules**

© 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

26 To connect to your Cluster go back to "EMR Dashboard"

aws Services Search [Alt+S]

EC2 IAM Simple Queue Service **EMR** S3

Amazon EMR

EMR Studio

EMR Serverless **New**

EMR on EC2

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

EMR on EKS

EMR Serverless is now GA.
With EMR Serverless, get the benefits of Amazon EMR such as open source compatibility, latest
[Started with EMR Serverless.](#)

Clone Terminate AWS CLI export

Cluster: ML SPARK **Waiting** Cluster ready after last step completed.

Summary Application user interfaces Monitoring Hardware Configurations

Summary **Config**

ID: j-3NC0YPPN26OMG

Creation date: 2022-11-03 08:01 (UTC-4)

Elapsed time: 3 hours, 30 minutes

After last step completes: Cluster waits

Termination protection: On [Change](#)

Tags: -- [View All / Edit](#)

27 Click on your Cluster

Started with EMR Serverless. [↗](#)

EMR Studio
EMR Serverless [↗](#) *New*

EMR on EC2

- Clusters
- Notebooks
- Git repositories
- Security configurations
- Block public access
- VPC subnets
- Events

EMR on EKS

- Virtual clusters

Help
What's new

Create cluster **View details** **Clone** **Terminate**

Filter: All clusters 11 clusters (all loaded) [↻](#)

		Name
<input type="checkbox"/>	▶	ML Spark
<input type="checkbox"/>	▶	ML SPARK
<input type="checkbox"/>	▶	My cluster
<input type="checkbox"/>	▶	My cluster
<input type="checkbox"/>	▶	ML-Spark
<input type="checkbox"/>	▶	ML-Spark
<input type="checkbox"/>	▶	ML-Spark

28 Click "Connect to the Master Node Using SSH"

Summary **Application user interfaces** Monitoring Hardware Configurations Events Steps Bootstrap

Summary

ID: j-3NC0YPPN26OMG
Creation date: 2022-11-03 08:01 (UTC-4)
Elapsed time: 3 hours, 34 minutes
After last step completes: Cluster waits
Termination protection: On [Change](#)
Tags: -- [View All / Edit](#)
Master public DNS: [ec2-54-87-146-62.compute-1.amazonaws.com](#) [↗](#)
[Connect to the Master Node Using SSH](#)

Configuration details

Release label: emr-6.8.0
Hadoop distribution: Amazon
Applications: Spark 3.3.0
Log URI: s3://aws-log-1/elasticmapreduce-logs-2022-11-03-08-01-00-000000000000
EMRFS consistent view: Disabled
Custom AMI ID: --
Amazon Linux Release: 2.0.20221001

Network and hardware

Availability zone: us-east-1c
Subnet ID: [subnet-0f674a9502b3ea8dc](#) [↗](#)
Master: **Running** 1 m5.xlarge
Core: **Running** 4 m5.xlarge
Task: --
Cluster scaling: Not enabled
Auto-termination: Not enabled

Security and access

Key name: ProgAssgn2
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Visible to all users: All [Change](#)
Security groups for Master: [sg-043e4861](#) (master)
Security groups for Core & Task: [sg-00bad271](#)

You can access your Cluster after following the procedures provided there in accordance with your OS. It should look something like this.

[illegible]

End

