

En EMR se crea un clúster

**Nombre y aplicaciones** [Información](#)

Nombre


Mi clúster 1


Versión de Amazon EMR [Información](#)


Una versión contiene un conjunto de aplicaciones que se puede instalar en el clúster.


emr-6.10.0 ▼


Paquete de aplicaciones


Spark  


Core  
Hadoop  


HBase  


Presto  


Trino  


Custom  


▼ Personalizar el paquete de aplicaciones

Aplicaciones incluidas en el paquete

<input type="checkbox"/> Flink 1.16.0	<input type="checkbox"/> Ganglia 3.7.2
<input type="checkbox"/> HBase 2.4.15	<input checked="" type="checkbox"/> HCatalog 3.1.3
<input checked="" type="checkbox"/> Hadoop 3.3.3	<input checked="" type="checkbox"/> Hive 3.1.3
<input checked="" type="checkbox"/> Hue 4.10.0	<input checked="" type="checkbox"/> JupyterEnterpriseGateway 2.6.0
<input checked="" type="checkbox"/> JupyterHub 1.5.0	<input checked="" type="checkbox"/> Livy 0.7.1
<input type="checkbox"/> MXNet 1.9.1	<input type="checkbox"/> Oozie 5.2.1
<input type="checkbox"/> Phoenix 5.1.2	<input type="checkbox"/> Pig 0.17.0
<input type="checkbox"/> Presto 0.278	<input checked="" type="checkbox"/> Spark 3.3.1
<input checked="" type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> TensorFlow 2.11.0
<input type="checkbox"/> Tez 0.10.2	<input type="checkbox"/> Trino 403
<input checked="" type="checkbox"/> Zeppelin 0.10.1	<input type="checkbox"/> ZooKeeper 3.5.10

Opciones del sistema operativo [Información](#)

☒ Versión de Amazon Linux

☐ Imagen de máquina de Amazon (AMI) personalizada

## Principal

Elegir tipo de instancia de EC2

m4.xlarge

4 vCore 16 GiB memoria

Únicamente EBS almacenamiento

Precio bajo demanda: - Precio de spot más bajo: -

Acciones ▼

☐ Usar varios nodos principales

Para mejorar la disponibilidad del clúster, utilice 3 nodos principales con la misma configuración y acciones de arranque. No puede utilizar varios nodos principales con flotas de instancias.

► Configuración de nodo - *opcional*

## Central

Elegir tipo de instancia de EC2

m4.xlarge

4 vCore 16 GiB memoria

Únicamente EBS almacenamiento

Precio bajo demanda: - Precio de spot más bajo: -

Acciones ▼

► Configuración de nodo - *opcional*

## Tarea 1 de 1

Eliminar grupo de instancias

Nombre

Tarea - 1

Elegir tipo de instancia de EC2

m4.xlarge

4 vCore 16 GiB memoria

Únicamente EBS almacenamiento

Precio bajo demanda: - Precio de spot más bajo: -

Acciones ▼

▼ Editar la configuración de software: *opcional* [Información](#)

☒ Ingresar la configuración

☐ Cargar JSON desde Simple Storage Service  
(Amazon S3)

```
1 ▼ [  
2 ▼  {  
3     "Classification": "jupyter-s3-conf",  
4 ▼   "Properties": {  
5       "s3.persistence.enabled": "true",  
6     "s3.persistence.bucket": "st0263jasanchez"  
7   }  
8 }  
9 ]
```

JSON Ln 6, Col 48



Par de claves de Amazon EC2 para el protocolo SSH al clúster - *opcional* [Información](#)

🔍 vockey

Examinar

Crear par de claves [↗](#)

## Roles de Identity and Access Management (IAM) [Información](#)

Elija o cree un rol de servicio y un perfil de instancia para las instancias de EC2 del clúster.

### Rol de servicio de Amazon EMR [Información](#)

El rol de servicio es un rol de IAM que Amazon EMR asume para aprovisionar recursos y realizar acciones de nivel de servicio con otros servicios de AWS.

☒ Elegir un rol de servicio existente

Seleccione un rol de servicio predeterminado o un rol personalizado con políticas de IAM asociadas para que el clúster pueda interactuar con otros servicios de AWS.

☐ Crear un rol de servicio

Deje que Amazon EMR cree un nuevo rol de servicio para que pueda conceder y restringir el acceso a los recursos de otros servicios de AWS.

Rol de servicio

EMR\_DefaultRole



### Perfil de instancia de EC2 para Amazon EMR

El perfil de instancia asigna un rol a cada instancia de EC2 de un clúster. El perfil de instancia debe especificar un rol que pueda acceder a los recursos de los pasos y las acciones de arranque.

☒ Elegir un perfil de instancia existente

Seleccione un rol predeterminado o un perfil de instancia personalizado con políticas de IAM asociadas para que el clúster pueda interactuar con sus recursos de Amazon S3.

☐ Crear un perfil de instancia

Deje que Amazon EMR cree un nuevo perfil de instancia para que pueda especificar un conjunto personalizado de recursos a los que tendrá acceso en Amazon S3.

Perfil de instancia

EMR\_EC2\_DefaultRole



Clústeres (1) <a href="#">Información</a>							
Filter clusters by status		Find clusters by ID or name, or search for text within loaded results			Filtrar clústeres por fecha y hora de creación		
<input type="checkbox"/>	ID del clúster	Nombre del clúster	Estado	Detalles del estado	Hora de creación (UTC-05:00)	Tiempo transcurrido	Horas de instancia normalizadas
<input type="checkbox"/>	j-1HJ38ZJRCYXIP	Mi clúster 1	Comenzando	Preparación de clúster	7 de mayo de 2023 11:44	1 minuto, 12 segundos	0

En S3 se crea un bucket

Buckets (1) <a href="#">Info</a>			
Los buckets son contenedores de datos almacenados en S3. <a href="#">Más información</a>			
🔍 Buscar buckets por nombre			
	Nombre	Región de AWS	Acceso
<input type="radio"/>	st0263jasanchez	EE. UU. Este (Norte de Virginia) us-east-1	Bucket y objetos que no son públicos

Elegir una clave de Amazon EC2 para el protocolo SSH al clúster

Pares de claves (1)

Buscar pares de claves

ID	Name	Fingerprint
key-0f7ec998ac016d8b0	vockey	dd:83:de:b7:0e:3d:d7:13:b4:6b: 54:89:ff:b7:76:0a:d6:c0:b0:38

Grupo de seguridad master:

Reglas de entrada (22)

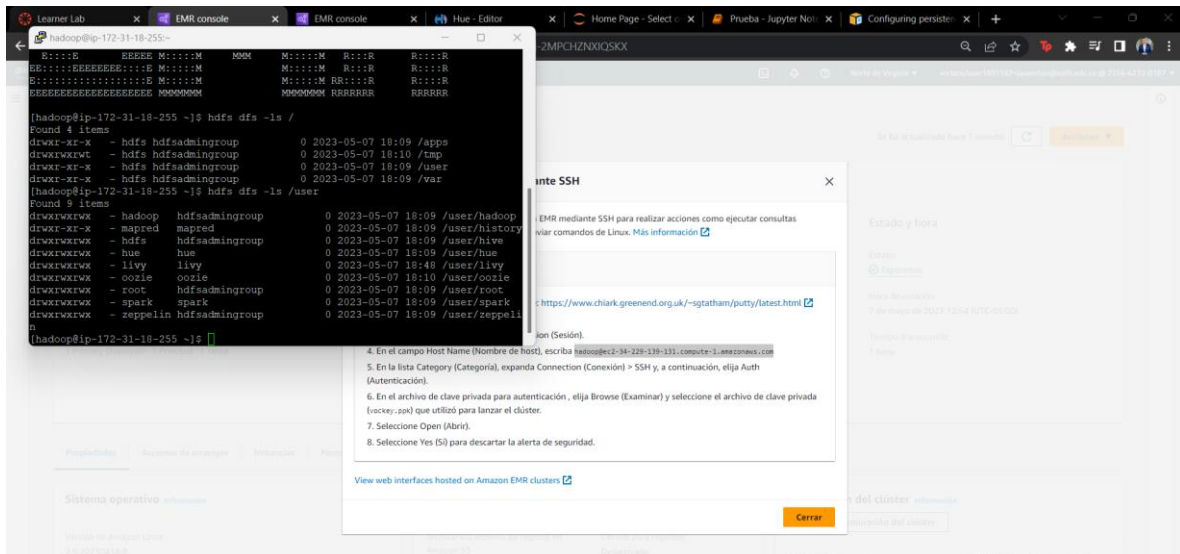
Administrar etiquetas

Editar reglas de entrada

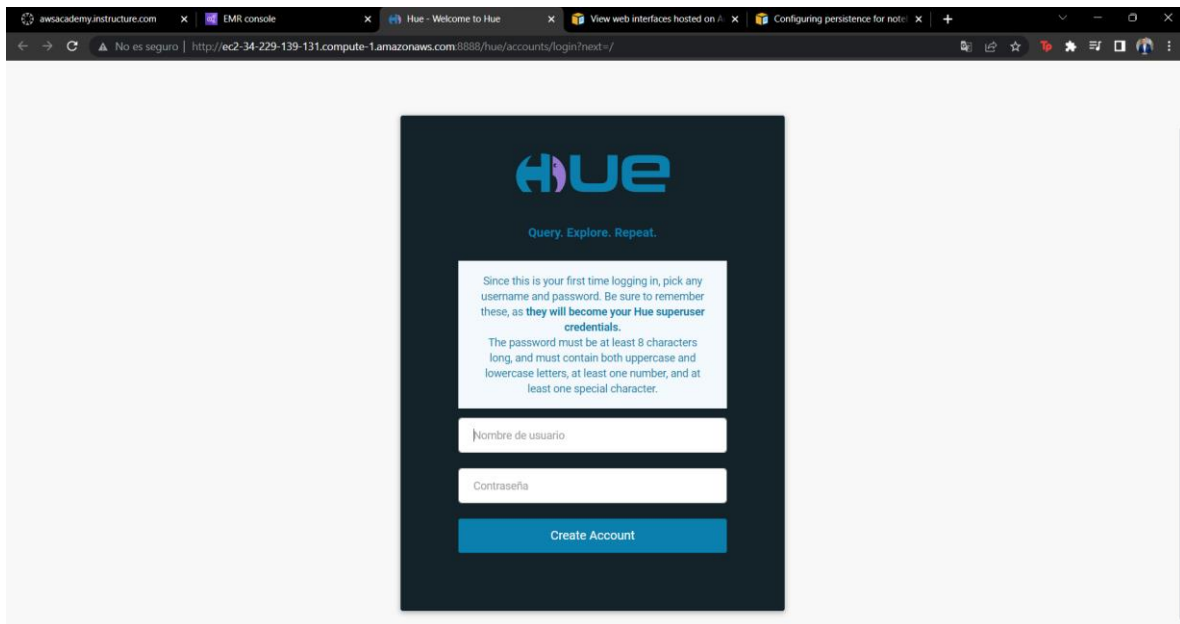
Filtrar reglas de grupo de seguridad

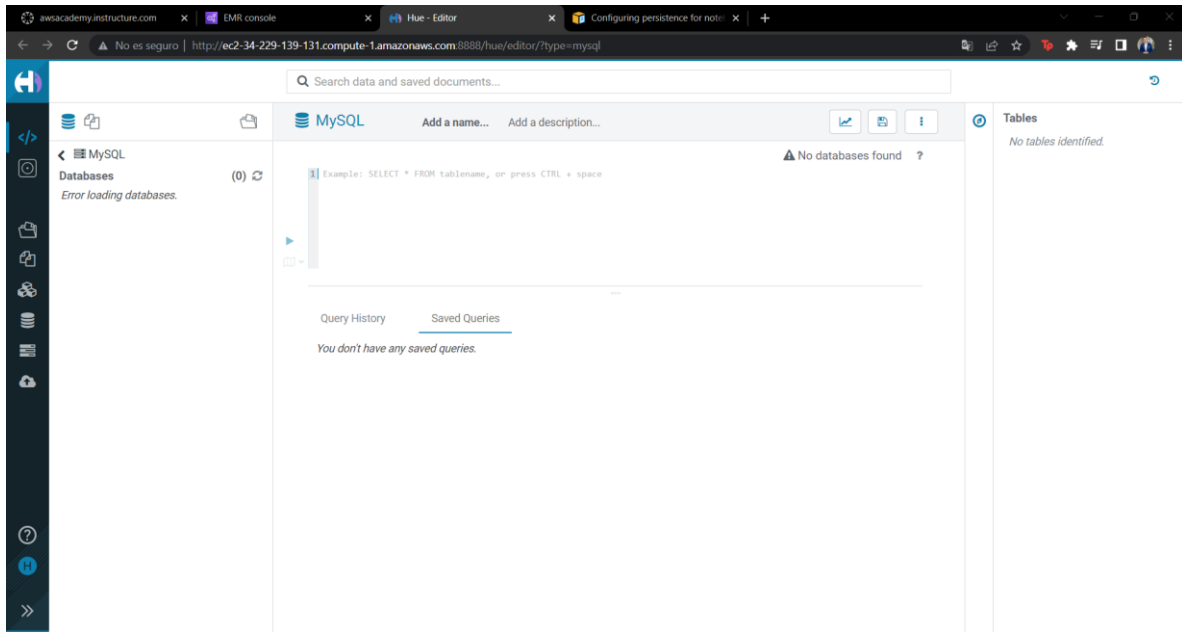
<input type="checkbox"/>	Name	ID de la regla del g...	Versión de IP	Tipo	Protocolo	Intervalo de puertos	Origen	Descripción
<input type="checkbox"/>	-	sgr-00f9b75c886a85391	IPv4	TCP personalizado	TCP	8443	54.240.217.8/29	-
<input type="checkbox"/>	-	sgr-060dd1d01f61c38c9	IPv4	TCP personalizado	TCP	8443	54.240.217.16/29	-
<input type="checkbox"/>	-	sgr-06598937e1d878...	IPv4	TCP personalizado	TCP	8443	54.239.98.0/24	-
<input type="checkbox"/>	-	sgr-098730617f25bc84a	IPv4	TCP personalizado	TCP	8443	54.240.217.80/29	-
<input type="checkbox"/>	-	sgr-09c3dae611dd6d88c	IPv4	TCP personalizado	TCP	8890	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-0a30841e19a195cf5	-	Todos los UDP	UDP	0 - 65535	sg-08d2ffc56a2c3492...	-
<input type="checkbox"/>	-	sgr-0ee26fdc8846bf414	-	Todos los TCP	TCP	0 - 65535	sg-08d2ffc56a2c3492...	-
<input type="checkbox"/>	-	sgr-02d95abf2382edfc5	IPv4	TCP personalizado	TCP	8443	72.21.196.64/29	-
<input type="checkbox"/>	-	sgr-03aff2d32e280436d	IPv4	TCP personalizado	TCP	8443	72.21.198.64/29	-
<input type="checkbox"/>	-	sgr-0bb8bce1f47ebc88c	-	Todos los UDP	UDP	0 - 65535	sg-0d8c5a21a9b5bb7...	-
<input type="checkbox"/>	-	sgr-0ab815e38fd326119	IPv4	SSH	TCP	22	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-0c24d14a52b49e2...	IPv4	TCP personalizado	TCP	8443	207.171.172.6/32	-
<input type="checkbox"/>	-	sgr-072236d417a0d2f...	IPv4	TCP personalizado	TCP	8443	207.171.167.101/32	-
<input type="checkbox"/>	-	sgr-0679727199f8edae14	IPv4	TCP personalizado	TCP	9443	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-090e7fbb7444414cf	IPv4	TCP personalizado	TCP	8443	72.21.217.0/24	-
<input type="checkbox"/>	-	sgr-00769d787883d3...	IPv4	TCP personalizado	TCP	8443	207.171.167.25/32	-
<input type="checkbox"/>	-	sgr-010218615f0ac6b03	IPv4	TCP personalizado	TCP	8888	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-08e8856c5bc4a130c	IPv4	TCP personalizado	TCP	8443	207.171.167.26/32	-
<input type="checkbox"/>	-	sgr-0b3dfb5e0ce50099c	IPv4	TCP personalizado	TCP	8443	54.240.217.64/28	-
<input type="checkbox"/>	-	sgr-0610ff7b1eae1c046	-	Todos los ICMP IPv4	ICMP	Todo	sg-0d8c5a21a9b5bb7...	-
<input type="checkbox"/>	-	sgr-012faafb01b95ccd	-	Todos los ICMP IPv4	ICMP	Todo	sg-08d2ffc56a2c3492...	-
<input type="checkbox"/>	-	sgr-003120f2fa5a003eb	-	Todos los TCP	TCP	0 - 65535	sg-0d8c5a21a9b5bb7...	-

Nodo master:

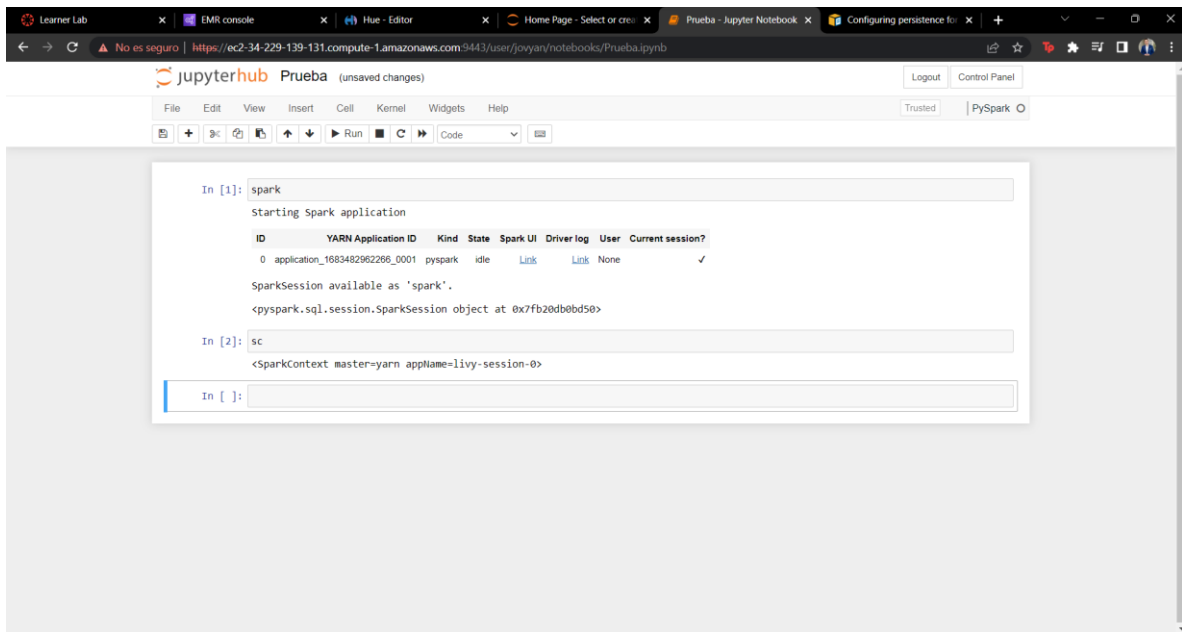


HUE: <http://ec2-34-229-139-131.compute-1.amazonaws.com:8888/hue/accounts/login?next=/>





JupyterHUB: <https://ec2-34-229-139-131.compute-1.amazonaws.com:9443/user/jovyan/notebooks/Prueba.ipynb>



Zeppelin: <http://ec2-34-229-139-131.compute-1.amazonaws.com:8890/#/notebook/2J1H1VBAX>

Zeppelin Notebook Job

PruebaZep

Finished

```
%spark.pyspark
spark

<pyspark.sql.session.SparkSession object at 0x7f065af743d0>
```

Finished

```
%spark.pyspark
sc

<SparkContext master=yarn appName=Zeppelin>
```

Finished

```
%sql
show databases
```

namespace

default