

# Guía reto número 1

Por: Sebastián Urrego García

## Prerrequisitos

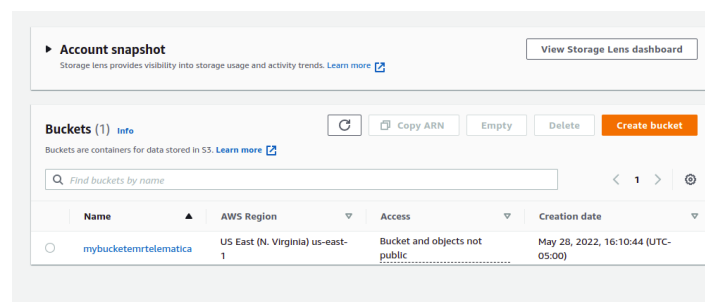
Antes de comenzar a configurar el clúster de Amazon EMR, usted debe completar los requisitos que a continuación se indican:

1. Tener una cuenta de AWS

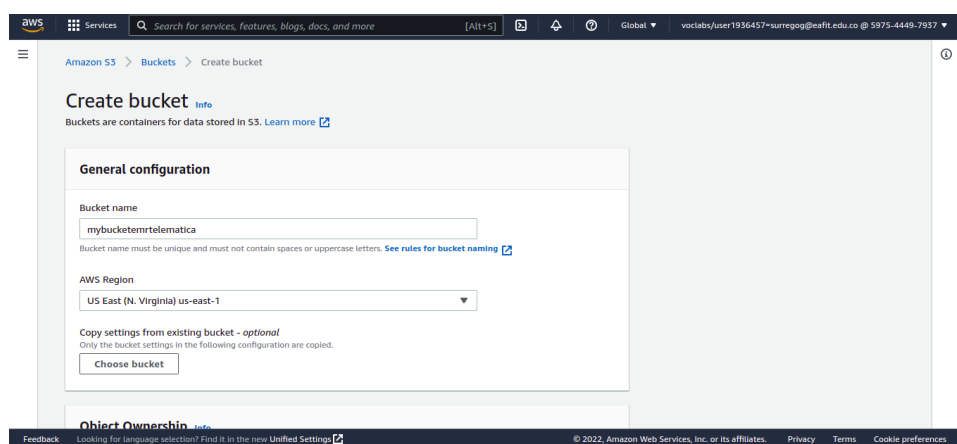
## Paso a paso

### Creación bucket de S3

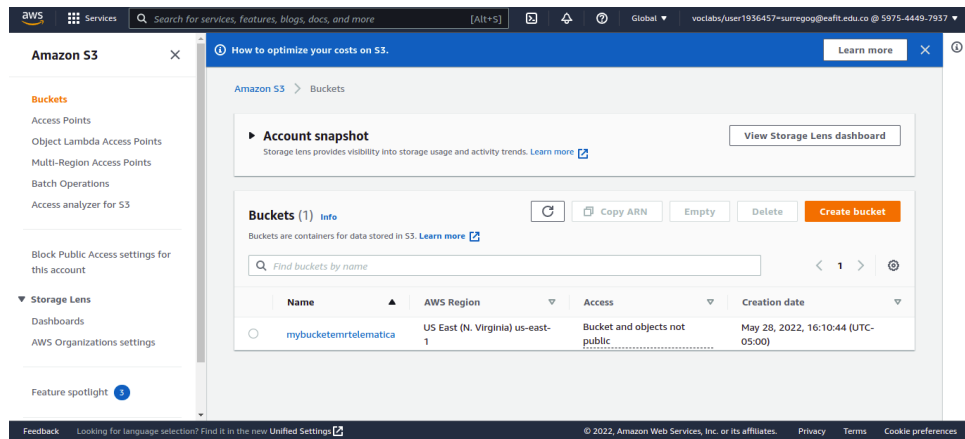
En la consola de AWS, se debe de dirigir al módulo de S3, allí debe de selecciona la opción de create bucket:



Ya dentro del panel de configuración le ponemos el nombre que nos guste, recuerde que debe ser un nombre unico y no acepta caracteres especiales, en nuestro caso le vamos a poner “mybucketmrtelematica” y vamos a dejar todas las configuraciones por defecto

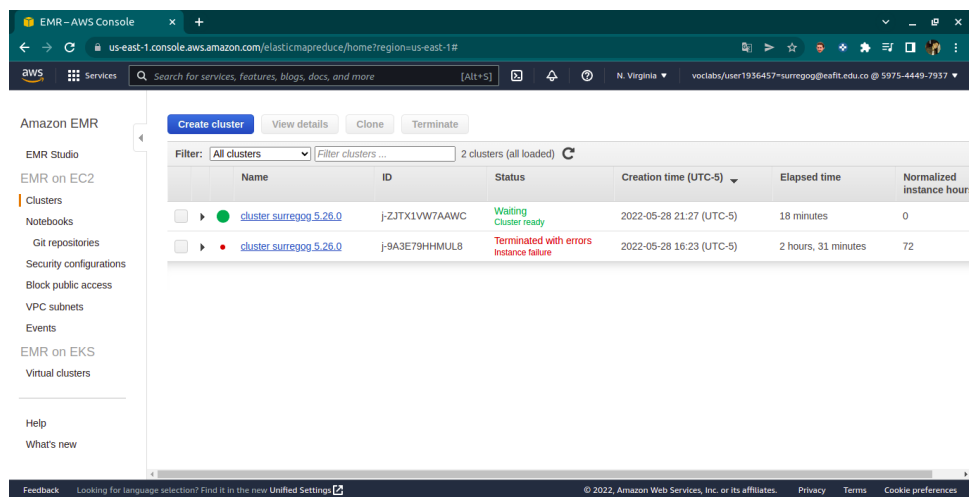


Ya con el bucket de S3 creado podemos proceder al siguiente paso, se debe de ver así:

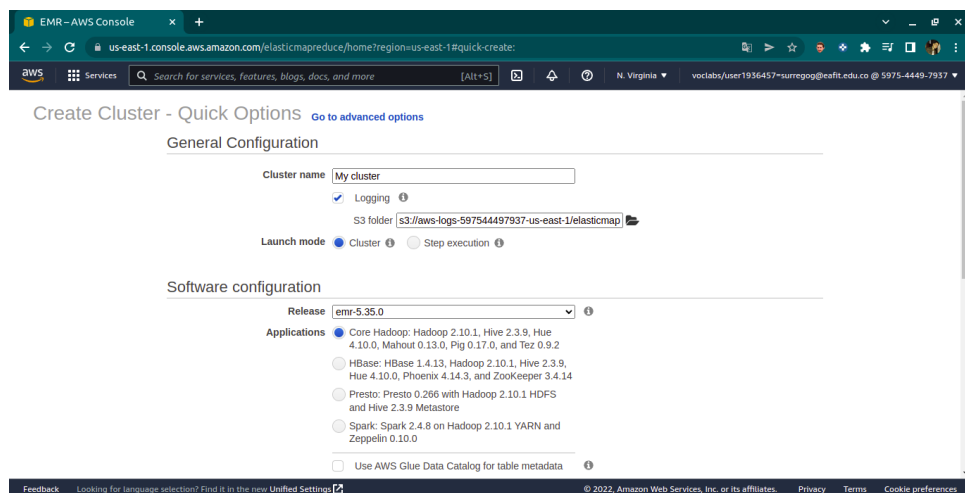


## Creación del EMR

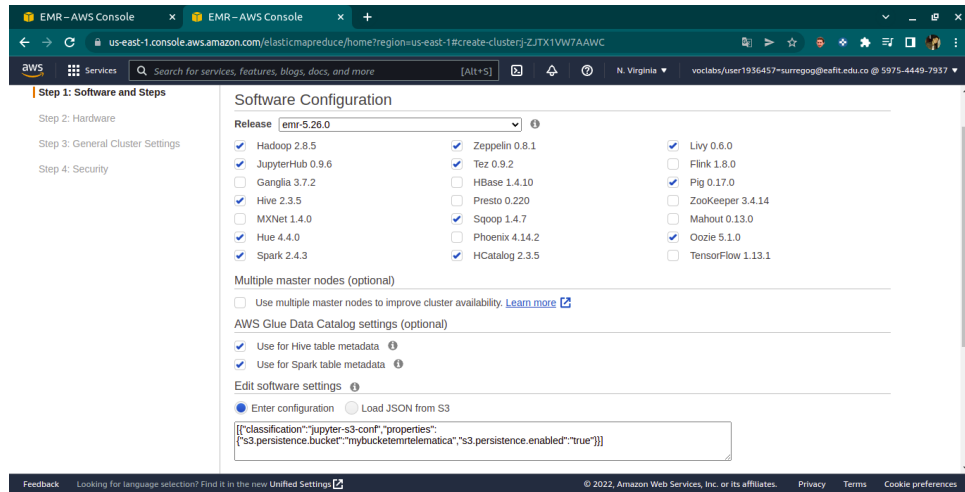
Nos dirigimos al módulo de EMR de AWS, allí seleccionamos en Create cluster y seguimos los pasos a continuación



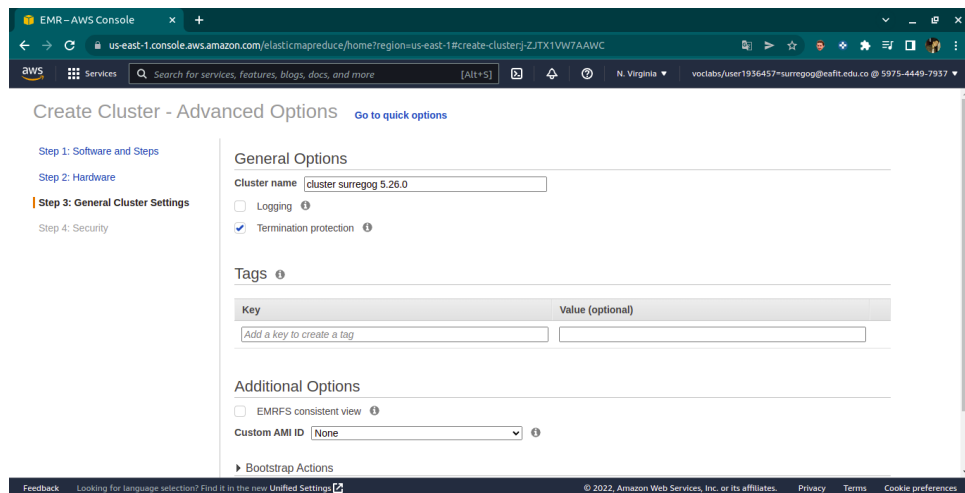
Ya dentro de la vista de creación nos dirigimos al link de Go to advanced options:



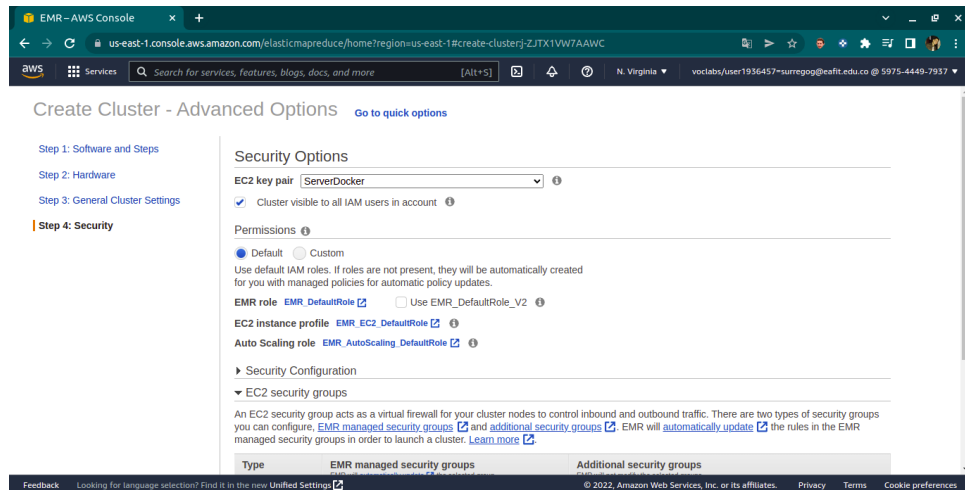
En software and Steps llenamos las siguientes configuraciones y escogemos la versión 5.26 del EMR, además adicionamos los elementos que se pueden ver a continuación en la captura, mucho cuidado en la parte de configuración del EMR para colocar el bucket de S3 que acabamos de crear.



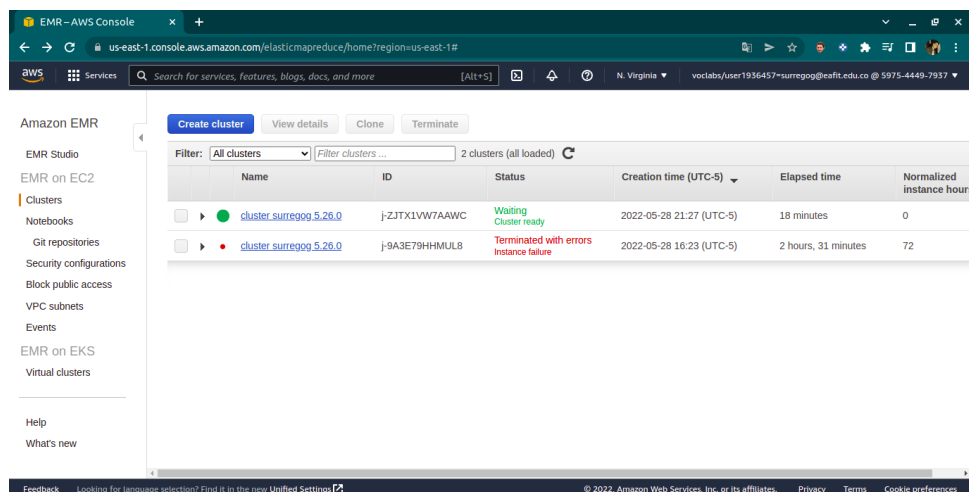
Le damos en Next y en la siguiente pestaña dejamos todo por defecto y contiuamos, entonces volvemos a apretar Next, en la pestaña que estamos le damos un nombre al cluster en este caso le coloqe así:



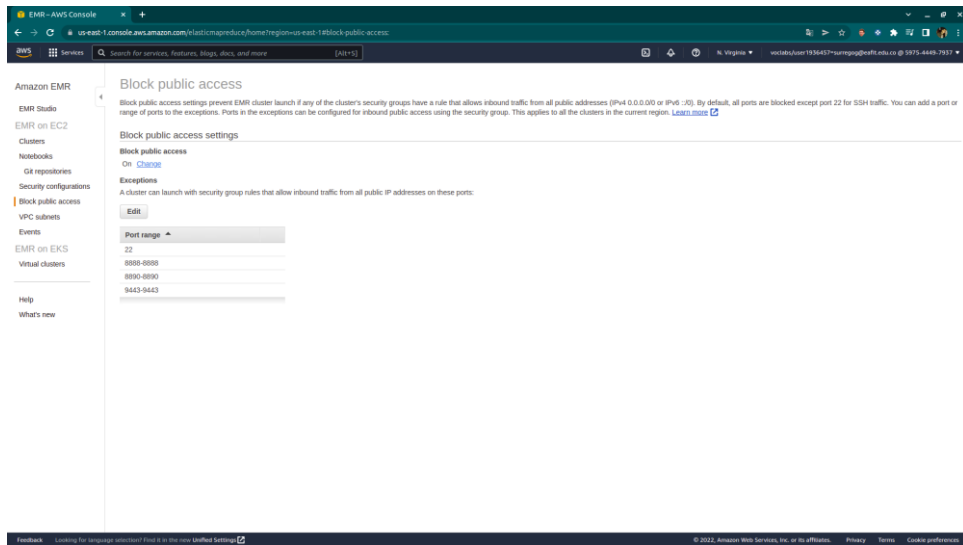
Por último, en la pestaña de seguridad le damos un .Pem que tengamos guardados y le damos a crear cluster, esto se puede tardar un momento:



Ahora cuando el EMR esté completamente funcional nos aparecerá una bolita verde rellena de que configuramos y creamos bien el cluster así:

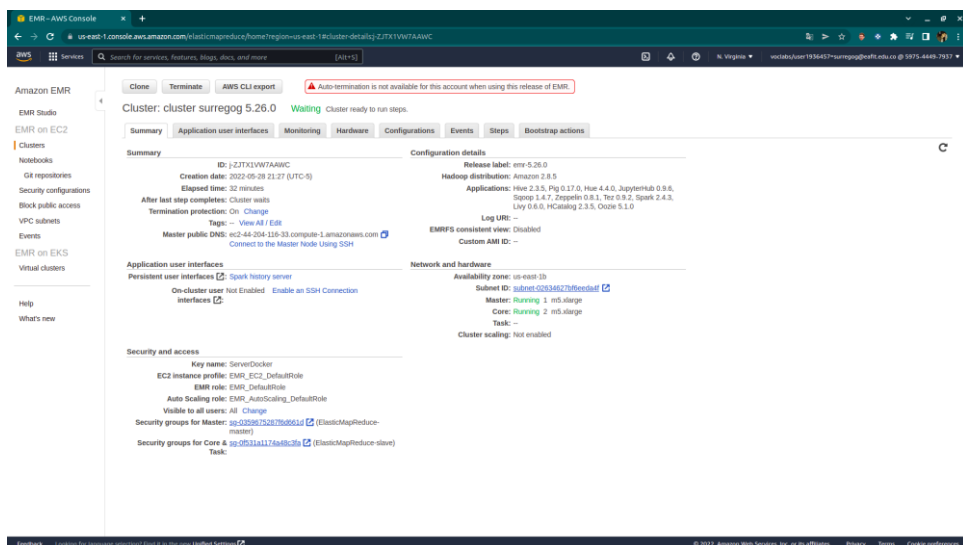


No nos olvidemos en esta misma sección dirigirnos a block public access en la consola del EMR y aquí vamos activarla y activar los siguiente puertos:

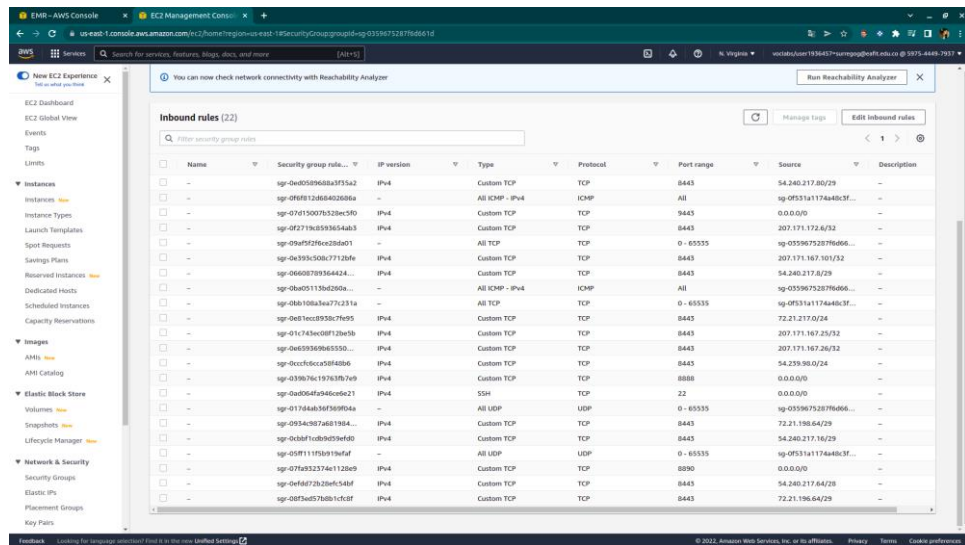


## Configuración de los puertos

Ahora para poder usar el EMR y poder conectarnos a él por SSH y utilizar otras funciones que este nos ofrece, en la consola del EMR, entramos a nuestro cluster y allí buscamos las configuraciones de Security and access y nos entramos al security group for masters:



Después de entrar al security group vamos a seleccionar el security group de masters y allí, vamos a añadir una regla adicional para conexiones por ssh adicional a eso para el uso de más herramientas tenemos que activar los puertos 8888, 8890 y el 9443 así debería quedar para poder utilizar todos los servicios que creamos en el EMR:



## Creación usuario en HUE

En la consola de EMR de AWS, nos dirigimos al cluster que ya creamos y allí entramos a la opción de Application user interfaces y aquí buscamos la URL que diga HUE, copiamos esa URL y la pegamos en nuestro navegador luego, se nos pedirá que creamos un usuario y contraseña en este caso el usuario que voy a crear es hadoop así:

Cluster: cluster-surreog 5.26.0 **Waiting** Cluster ready to run steps.

**Persistent application user interfaces**

Applications installed on the Amazon EMR cluster publish user interfaces (UI) as web sites to monitor cluster activity. Persistent UI logs are available for 30 days after an application ends. Persistent UI don't require SSH tunneling. They are hosted off of the cluster.

**Application user interface**

[Spark history server](#)

**On-cluster application user interfaces**

On-cluster UI are available only while clusters are running. Because they are hosted on the master node, on-cluster UI require a connection via SSH tunneling. Set up SSH tunneling before accessing these application UI. [Learn more](#)

Application	User interface URL	Status
HDFS Name Node	http://ec2-44-204-116-33.compute-1.amazonaws.com:50070/	SSH tunnel not enabled
Hue	http://ec2-44-204-116-33.compute-1.amazonaws.com:8888/	SSH tunnel not enabled
JupyterHub	https://ec2-44-204-116-33.compute-1.amazonaws.com:8443/	SSH tunnel not enabled
Zeppelin	http://ec2-44-204-116-33.compute-1.amazonaws.com:8890/	SSH tunnel not enabled
Tez UI	http://ec2-44-204-116-33.compute-1.amazonaws.com:8080/tez-ui	SSH tunnel not enabled
Spark History Server	http://ec2-44-204-116-33.compute-1.amazonaws.com:18080/	SSH tunnel not enabled
Livy	http://ec2-44-204-116-33.compute-1.amazonaws.com:8998/	SSH tunnel not enabled
Resource Manager	http://ec2-44-204-116-33.compute-1.amazonaws.com:8080/	SSH tunnel not enabled

The following table lists web interfaces you can view on the task nodes:

Application	User interface URL
HDFS Data Node	http://ec2-44-204-116-33.compute-1.amazonaws.com:50070/
Node Manager	http://ec2-44-204-116-33.compute-1.amazonaws.com:8042/

**High-level application history**

Amazon EMR collects information from YARN applications on your cluster and keeps a summary of historical information for seven days after applications have completed. [Learn more](#)

**YARN applications (8)**

Filter: All applications 0 applications (all loaded)

Application ID	Type	Action	Status	Start time (UTC-5)	Duration	Finish time (UTC-5)	User
----------------	------	--------	--------	--------------------	----------	---------------------	------

**Hue**

Query. Explore. Repeat.

Since this is your first time logging in, pick any username and password. Be sure to remember these, as they will become your Hue superuser credentials.

The password must be at least 8 characters long, and must contain both uppercase and lowercase letters, at least one number, and at least one special character.

Username:

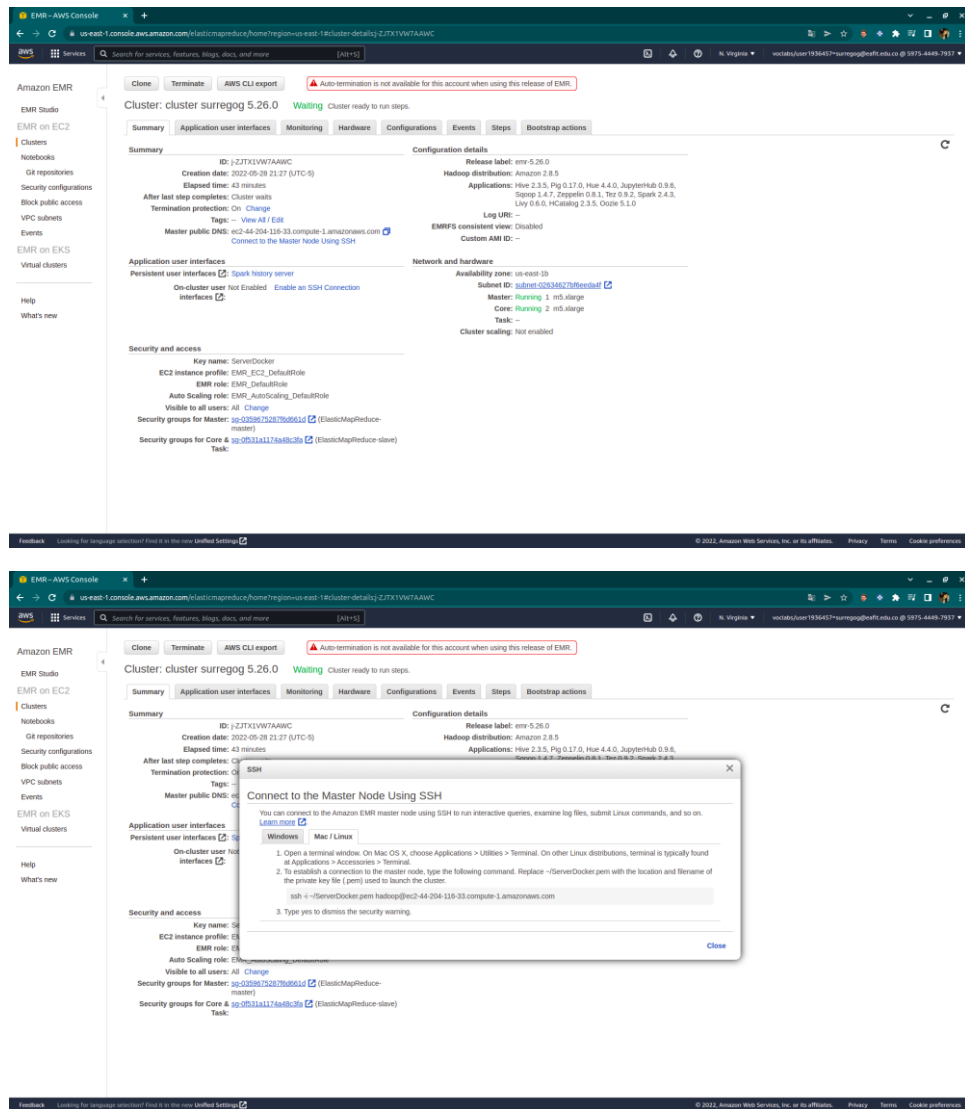
Password:

[Create Account](#)

Hue and the Hue logo are trademarks of Cloudera, Inc.

## Conexión por SSH

Para conectarnos a nuestro EMR vía ssh, debemos de dirigirnos a nuestro cluster en AWS y allí en la sección de Summary, buscar la URL que sale de conectarse a SSH, le damos click y nos da las siguientes instrucciones



Entonces como nos dice el instructivo debemos de abrir una terminal (En este caso en linux) y colocamos la siguiente línea de comando, vale aclarar que recuerde que debe de encontrar su llave .Pem con la que creo el cluster para que funcione, entonces después de poner el comando debe de tener algo así:



