

AG1 - Hadoop en EC2

Contenido

1 Map-Reduce	1
1.1 Creación de las máquinas virtuales EC2	1
1.2 Configuración de la seguridad	2
1.3 Instalación de Hadoop	3

Datos del estudiante

Nombre:

Ernesto Vladimir Palacios Merino

Asignatura:

Cloud Computing - Actividad 1

Fuentes: Build a Multi-node apache-hadoop cluster

1 Map-Reduce

El objetivo de la presente actividad se centra en la creación, configuración y ejecución de Hadoop:

Instalar un cluster hadoop multinodos usando AWS EC2

Se ha elegido utilizar como imagen base el sistema operativo **Ubuntu 22.04** por la familiaridad con el sistema y los comandos.

1.1 Creación de las máquinas virtuales EC2

En esta sección:

- Creación de cuatro máquinas virtuales
- Configurar el acceso remoto por ssh desde afuera en el puerto 22
- Configurar el acceso remoto completo interno desde el mismo grupo de seguridad
- Tomar nota de las direcciones IP Privadas de cada máquina creada

Creación de las instancias:

Name and tags

Name

Hadoop

Add additional tags

Application and OS Images (Amazon Machine Image)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

Search our full catalog including 1000s of application and OS images

Quick Start

Amazon Linux

macOS

Ubuntu

Windows

Red Hat

SUSE LI

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Ubuntu Server 24.04 LTS (HVM), SSD Volume Type

Free tier eligible

ami-04b70fa74e45c3917 (64-bit (x86)) / ami-0eac975a54dfee8cb (64-bit (Arm))

Virtualization: hvm ENA enabled: true Root device type: ebs

Description

Canonical, Ubuntu, 24.04 LTS, amd64 noble image build on 2024-04-23

Architecture

AMI ID

64-bit (x86)

ami-04b70fa74e45c3917

Verified provider

Summary

Number of instances

4

When launching more than 1 instance, consider EC2 Auto Scaling

Software Image (AMI)

Canonical, Ubuntu, 24.04 LTS, ...read more

ami-04b70fa74e45c3917

Virtual server type (instance type)

t2.micro

Firewall (security group)

New security group

Storage (volumes)

1 volume(s) - 8 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the Internet.

Cancel

Launch instance

Review commands

Figure 1: Creación de instancias EC2

Se crean la vez cuatro máquinas con Sistema Operativo: ubuntu 64-bits de tipo t2.micro con 8GB SSD

Se habilita el puerto 22 con la llave SSH vokey

Instances (1/4)											
Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs	Monitoring
Hadoop_1	i-01b1182a16d42642d	Running	t2.micro	2/2 checks passed	View alarms	us-east-1a	ec2-18-208-216-241.co...	18.208.216.241	-	-	disabled
Hadoop_2	i-025cb46273bf07085	Running	t2.micro	2/2 checks passed	View alarms	us-east-1a	ec2-107-23-20-249.co...	107.23.20.249	-	-	disabled
Hadoop_3	i-0e0b15a880c244740	Running	t2.micro	2/2 checks passed	View alarms	us-east-1a	ec2-54-221-184-240.co...	54.221.184.240	-	-	disabled
Hadoop_4	i-0b0eb99e87093d26d	Running	t2.micro	2/2 checks passed	View alarms	us-east-1a	ec2-54-89-197-241.co...	54.89.197.241	-	-	disabled

Details

Status and alarms

Monitoring

Security

Networking

Storage

Tags

Instance summary

Instance ID

i-0b0eb99e87093d26d (Hadoop_4)

Public IPv4 address

54.89.197.241 | open address

Private IPv4 addresses

172.31.31.56

Public IPv4 DNS

ec2-54-89-197-241.compute-1.amazonaws.com | open address

Elastic IP addresses

-

AWS Compute Optimizer finding

Opt-in to AWS Compute Optimizer for recommendations. | Learn more

Auto Scaling group name

-

Figure 2: Instancias creadas para Map-Reduce

Más adelante cambiaré los nombres a las instancias para identificarlas adecuadamente.

1.2 Configuración de la seguridad

Luego configuro las reglas Inbound del grupo de seguridad al que pertenecen todas las máquinas virtuales.

2

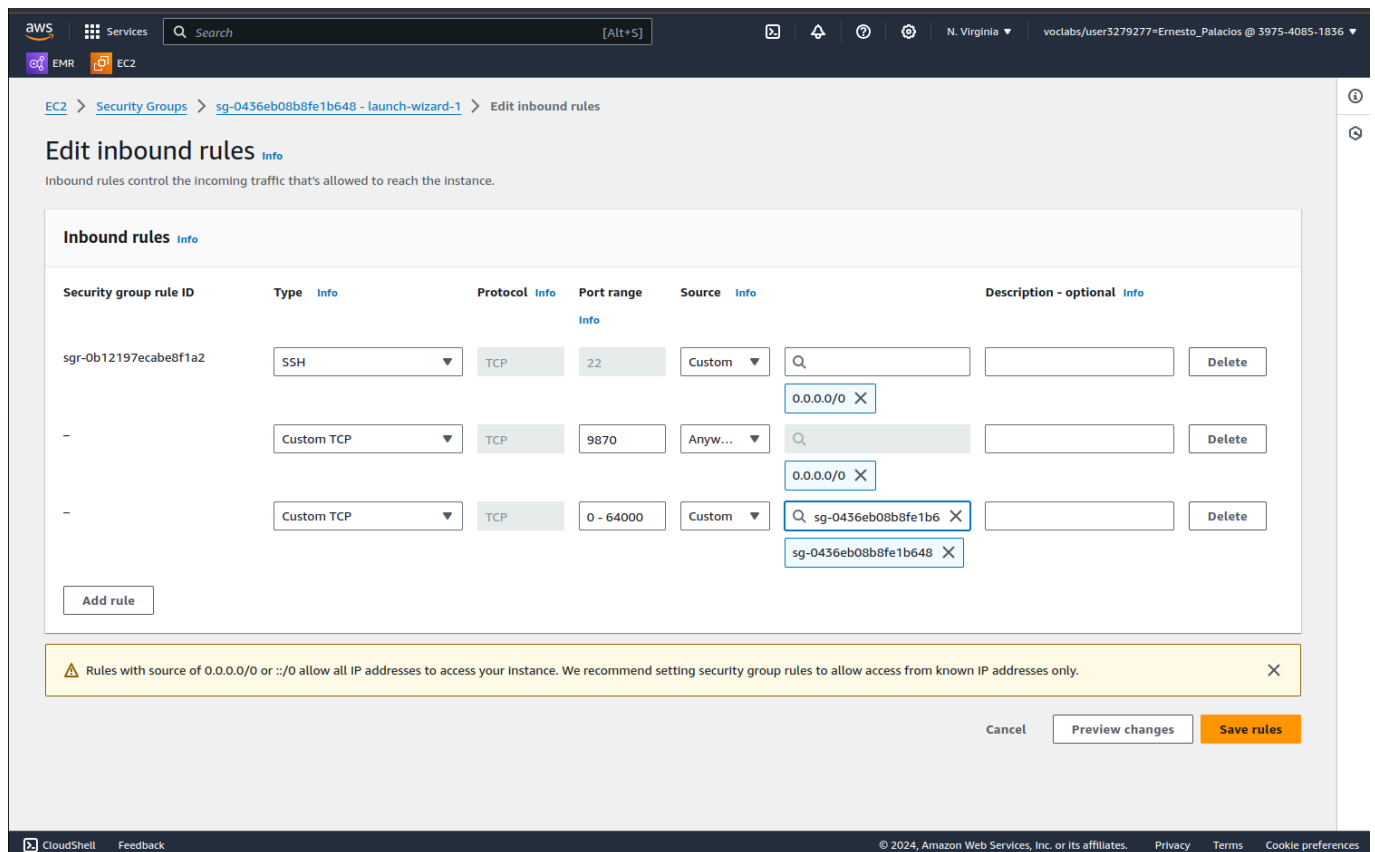


Figure 3: Configuración de seguridad

- El puerto 22 (SSH) puede ser accesado desde el internet, todas las IP, permite conexión remota.
- El puerto 9870 (Web) accesado desde el internet, todas las IP, permite visualización remota del clúster.
- Todos los puertos abiertos solamente a las máquinas que se encuentran en el mismo grupo de seguridad. En mi caso es la sg-0436eb08...

1.3 Instalación de Hadoop

Para la instalación de **Hadoop** he seguido la guía: Tema3_InstalacionHadoopStandalone.pdf

```
{bash}

# instalar JAVA-8 en todos los nodos.

sudo apt -y install openjdk-8-jdk-headless

# Descargo Hadoop versión 3.4 solo en el master.

wget https://archive.apache.org/dist/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

# Descomprimir Hadoop:

tar xzf hadoop-3.4.0.tar.gz

# Lo reubico hasta en local/

sudo mv hadoop-3.4.0 /usr/local/

# Cambio el dueño de la carpeta al usuario por defecto: ubuntu
```

```
sudo chown -R ubuntu:ubuntu /usr/local/hadoop-3.4.0

# Verifico la ubicación donde se ha instalado JAVA

readlink -f $(which java)

# en mi caso uso el siguiente directorio: /usr/lib/jvm/java-8-openjdk-amd64

# Edito el fichero: bash.rc

nano ~/.bashrc

# copio al final las siguientes líneas y guardo el archivo desde nano:

export HADOOP_HOME=/usr/local/hadoop-3.4.0
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:$HADOOP_HOME/bin

# finalmente reinicio BASH para que surta efecto los cambios

source ~/.bashrc
```