

# Actividad de refuerzo nº.7. HDFS

## Realizado con AWS

### 1. Introducción

La actividad se ha realizado en Amazon EMR para una mayor familiarización con un entorno empresarial real. El objetivo es practicar el uso de HDFS: subir archivos, gestionar carpetas y comprobar su estado mediante línea de comandos.

### 2. Preparación del entorno

Se ha utilizado Amazon EMR para montar un entorno Hadoop distribuido. Se ha creado clúster con hadoop 3.4.1 con nombre “act7-ivo”.

act7-ivo

Se ha actualizado hace menos de un minuto

Terminar

Clonar en AWS CLI

Clonar

▼ Resumen

Información del clúster

ID del clúster  
j-3D2INCXTU1RD9

ARN del clúster  
arn:aws:elasticmapreduce:us-east-1:025749687294:cluster/j-3D2INCXTU1RD9

Configuración del clúster  
Grupos de instancias

Capacidad  
1 Primary (Principal) | 1 Principal | 1 Tarea

Aplicaciones

Versión de Amazon EMR  
emr-7.8.0

Aplicaciones instaladas  
Hadoop 3.4.1, Hive 3.1.3, Hue 4.11.0, JupyterEnterpriseGateway 2.6.0, JupyterHub 1.5.0, Livy 0.8.0, Spark 3.5.4

Administración de clústeres

Destino del registro en Amazon S3  
aws-logs-025749687294-us-east-1/elasticmapreduce

IU de aplicación persistente  
Servidor de historial de Spark  
Servidor de línea de tiempo de YARN  
UI de Tez

DNS público del nodo principal  
ec2-44-202-71-18.compute-1.amazonaws.com  
Conectarse al nodo principal mediante SSH

Estado y hora

Estado  
Terminado

Hora de creación  
6 de mayo de 2025 13:21 (UTC+02:00)

Tiempo transcurrido  
1 hora, 27 minutos

Hora de finalización  
6 de mayo de 2025 14:48 (UTC+02:00)

1 En la imagen se muestra el clúster EMR creado y estado terminado

La clave se ha generado desde el panel de AWS y descargado localmente.

Pares de claves (1/2)

Información

Acciones

Crear par de claves

Q Buscar Par de claves por atributo o etiqueta

< 1 >

	Nombre	Tipo	Se ha creado	Huella digital	ID
<input type="checkbox"/>	vockey	rsa	2025/05/06 00:02 G...	be:d4:08:d8:29:25:ca:...	key-0f8e6e2b89b7808d1
<input checked="" type="checkbox"/>	ivonis2025	rsa	2025/05/06 12:56 G...	cb:b5:88:c9:65:25:d5:...	key-0497b21a7c77adc50

2 La clave ivonis2025.pem fue generada para habilitar el acceso SSH.


Para poder acceder al nodo principal del clúster mediante SSH, se añadió una regla de entrada en el grupo de seguridad asociado. Se abrió el puerto 22 permitiendo el tráfico desde la IP pública de la máquina cliente

## Actividad 7 – Alumno: Ivonis Florindo López


sg-0349ca295a9876b43 - ElasticMapReduce-master

Detalles


Nombre del grupo de seguridad

 ElasticMapReduce-master


Propietario

 025749687294


ID del grupo de seguridad

 sg-0349ca295a9876b43


Número de reglas de entrada

 8 Entradas de permisos



Descripción

 Master group for Elastic MapReduce created on 2025-05-05T21:27:26.03Z

Número de reglas de salida

 1 Entrada de permiso

ID de la VPC

 vpc-05001689310a844d4 

Reglas de entrada


Reglas de salida


Compartiendo: *novedad*

Asociaciones de VPC: *novedad*

Etiquetas



Reglas de entrada (8)


 Buscar



Administrar etiquetas

Editar reglas de entrada

<input type="checkbox"/>	Nombre	ID de la regla del gr...	Versión de IP	Tipo	Protocolo	Intervalo de puertos	Origen	Descripción
<input type="checkbox"/>	-	sg-0fda59879a2301d72	-	Todos los TCP	TCP	0 - 65535	<a href="#">sg-0c133084ed2cb5f09...</a>	-
<input type="checkbox"/>	-	sg-04ec3f52740b0d1cd	-	Todos los UDP	UDP	0 - 65535	<a href="#">sg-0349ca295a9876b4...</a>	-
<input type="checkbox"/>	-	sg-09d57c5967e91ac2	-	TCP personalizado	TCP	8443	<a href="#">pl-f8bd5e91</a> 	-
<input type="checkbox"/>	-	sg-05205a58f76cbd1e	-	Todos los TCP	TCP	0 - 65535	<a href="#">sg-0349ca295a9876b4...</a>	-
<input type="checkbox"/>	-	sg-068538eebc1deacc	-	Todos los ICMP IPv4	ICMP	Todo	<a href="#">sg-0c133084ed2cb5f09...</a>	-
<input type="checkbox"/>	-	sg-050c3473c4d007d6d	-	Todos los UDP	UDP	0 - 65535	<a href="#">sg-0c133084ed2cb5f09...</a>	-
<input type="checkbox"/>	-	sg-0daa167eab3ae3980	-	Todos los ICMP IPv4	ICMP	Todo	<a href="#">sg-0349ca295a9876b4...</a>	-
<input type="checkbox"/>	-	sg-033151a433341e42a	IPv4	SSH	TCP	22	80.31.78.40/32	-

3. Configuración del grupo de seguridad en AWS para permitir el acceso SSH al nodo maestro mediante la apertura del puerto 22 desde la IP pública del cliente.

### 3. Subida de archivos y gestión en HDFS

Se accedió al nodo principal mediante SSH con el siguiente comando:

```
ssh -i "~/ivonis2025.pem" ec2-user@ec2-44-202-71-18.compute-1.amazonaws.com
```

Una vez en la instancia, se cargaron los archivos al sistema HDFS en la ruta /tmp/books y se verificó que los archivos estaban correctamente almacenados.

```
MINGW64/c:/ProyectosIA/Actividad_7
User@DESKTOP-BTOOAMJ MINGW64 /c:/ProyectosIA/Actividad_7
$ scp -i "/c/Users/User/ivonis2025.pem" frankenstein.txt ec2-user@ec2-44-202-71-18.compute-1.amazonaws.com:/home/ec2-user/libros/frankenstein.txt
100% 431KB 608.4KB/s 00:00
User@DESKTOP-BTOOAMJ MINGW64 /c:/ProyectosIA/Actividad_7
$ scp -i "/c/Users/User/ivonis2025.pem" quijote.txt ec2-user@ec2-44-202-71-18.compute-1.amazonaws.com:/home/ec2-user/libros/quijote.txt
100% 2091KB 2.0MB/s 00:01
User@DESKTOP-BTOOAMJ MINGW64 /c:/ProyectosIA/Actividad_7
$ ssh -i "/c/Users/User/ivonis2025.pem" ec2-user@ec2-44-202-71-18.compute-1.amazonaws.com
#####
~\#####
~~~\#####|
~~~\###|
~~~\#/
~~~~V' ~-> https://aws.amazon.com/linux/amazon-linux-2023
~~~~m'
Last login: Tue May 6 12:41:00 2025 from 80.31.78.40
```

#### 4. Captura de la ejecución de comandos en la terminal: conexión SSH, creación de directorios, subida de archivos a HDFS.

Actividad 7 – Alumno: Ivonis Florindo López

Se realiza visualización de la primera página.

```
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRR
E:::EEEEEEEEEEEEEE M:::M M:::M R:::R
EE::EEEEEEEEEEEEEE M:::M M:::M R:::RRRRRR:::R
E:::E EEEEE M:::M M:::M RR:::R R:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R:::RRRRRR:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R:::RRRRRR:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R:::RRRRRR:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::E EEEEE M:::M MMM M:::M R:::R R:::R
EE::EEEEEEEEEE::E M:::M M:::M R:::R R:::R
E:::EEEEEEEEEE::E M:::M M:::M RR:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[ec2-user@ip-172-31-89-91 ~]$ hdfs dfs -mkdir -p /tmp/books
[ec2-user@ip-172-31-89-91 ~]$ hdfs dfs -put ~/libros/frankenstein.txt /tmp/books/
[ec2-user@ip-172-31-89-91 ~]$ hdfs dfs -put ~/libros/quijote.txt /tmp/books/
[ec2-user@ip-172-31-89-91 ~]$ hdfs dfs -ls /tmp/books/
Found 2 items
-rw-r--r-- 1 ec2-user hdfsadmin group 441034 2025-05-06 12:43 /tmp/books/frankenstein.txt
-rw-r--r-- 1 ec2-user hdfsadmin group 2141519 2025-05-06 12:43 /tmp/books/quijote.txt
[ec2-user@ip-172-31-89-91 ~]$ hdfs dfs -cat /tmp/books/frankenstein.txt | head -n 10
Project Gutenberg's Frankenstein, by Mary Wollstonecraft (Godwin) Shelley

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
re-use it under the terms of the Project Gutenberg License included
with this eBook or online at www.gutenberg.net

Title: Frankenstein
      or The Modern Prometheus
cat: Unable to write to output stream.
[ec2-user@ip-172-31-89-91 ~]$
Broadcast message from root@localhost (Tue 2025-05-06 12:48:49 UTC):

The system will power off now!

Connection to ec2-44-202-71-18.compute-1.amazonaws.com closed by remote host.
Connection to ec2-44-202-71-18.compute-1.amazonaws.com closed.
```

5. Captura con la verificación, visualización de contenido y cierre del servidor.

La sesión se cerró automáticamente por lo que volvemos a realizar el proceso anterior clonando el clúster, copiando de nuevo la DNS.

En este caso se opta por usar la interfaz del navegador.

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'ip-172-31-91-249.ec2.internal:8020' (✓active)

Started:	Thu May 08 13:08:37 +0200 2025
Version:	3.4.1-amzn-0, re73875e333fb9201de769ac856bfd8ef60a08bd8
Compiled:	Wed Feb 26 01:36:00 +0100 2025 by release from branch-3.4.1-amzn-0
Cluster ID:	CID-b6cf3253-ef68-4479-abde-fb757cd11bd5
Block Pool ID:	BP-1114137649-172.31.91.249-1746702511656

6. Interfaz del navegador.

Actividad 7 – Alumno: Ivonis Florindo López

Se crea en raíz /books y /ivonis

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	hdfs	hdfsadmin	0 B	May 08 13:10	0	0 B	apps	
<input type="checkbox"/>	drwxr-xr-x	dr.who	hdfsadmin	0 B	May 08 13:26	0	0 B	books	
<input type="checkbox"/>	drwxr-xr-x	dr.who	hdfsadmin	0 B	May 08 13:27	0	0 B	ivonis	
<input type="checkbox"/>	drwxrwxrwt	hdfs	hdfsadmin	0 B	May 08 13:13	0	0 B	tmp	
<input type="checkbox"/>	drwxr-xr-x	hdfs	hdfsadmin	0 B	May 08 13:10	0	0 B	user	
<input type="checkbox"/>	drwxr-xr-x	hdfs	hdfsadmin	0 B	May 08 13:10	0	0 B	var	

Showing 1 to 6 of 6 entries

7. Contenidos del directorio raíz.

Cargamos de nuevo los dos libros y los movemos a /books:

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	hdfsadmin	2.04 MB	May 08 13:53	1	128 MB	Quijote.txt	
<input type="checkbox"/>	-rw-r--r--	hadoop	hdfsadmin	430.7 KB	May 08 13:53	1	128 MB	frankenstein.txt	

Showing 1 to 2 of 2 entries

8. Verificación de la carga.

Cargamos “fichero\_results.zip” y lo descomprimos

```
[hadoop@ip-172-31-81-39 ~]$ cd /home/hadoop
[hadoop@ip-172-31-81-39 ~]$ unzip fichero_result.zip -d fichero_result
Archive:  fichero_result.zip
inflating: fichero_result/fichero_result.txt
```

9. Captura de la línea de comando para descomprimir.

A continuación se mueve a /ivonis

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	hdfsadmin	4 B	May 08 19:03	1	128 MB	fichero_result.txt	

9. Archivo “fichero\_result.txt” correctamente cargado en /ivonis.

Actividad 7 – Alumno: Ivonis Florindo López

Se procede a cortar y pegar “frankenstein.txt” a /ivonis

Browse Directory

/ivonis

Gol

Search

25 entries

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	hadoop	hdfsadmin	4 B	May 08 19:03	1	128 MB	fichero_result.txt
-rw-r--r--	hadoop	hdfsadmin	430.7 KB	May 08 18:54	1	128 MB	frankenstein.txt

Showing 1 to 2 of 2 entries

Previous1Next

10. Captura que muestra que se ha movido correctamente “frankenstein.txt”.

Por último intentamos descargar “frankenstein.txt” pero por navegador no es posible, por lo que lo hacemos por línea de comando.

Hadoop Overview Datanodes

Block information - frankenstein.txt

Download Head the file (first 32K) Tail the file (last 32K)

Block information Block 0

Block ID: 1073743159  
Block Pool ID: BP-741451703-172.31.81.39-1746720428631  
Generation Stamp: 2335  
Size: 441034  
Availability:

- ip-172-31-83-198.ec2.internal

Close

Browse Directory

/ivonis

25 entries

Permission	Owner
-rw-r--r--	hadoop
-rw-r--r--	hadoop

Showing 1 to 2 of 2 entries

Hadoop, 2024.

No se puede acceder a este sitio web

Comprueba si hay un error de escritura en ip-172-31-83-198.ec2.internal.

Si está escrito correctamente, prueba a ejecutar el diagnóstico de red de Windows.

DNS\_PROBE\_FINISHED\_NXDOMAIN

User@DESKTOP-BTOOAMJ MINGW64 ~

\$ scp -i "/c/Users/User/ivonis2025.pem" \

hadoop@ec2-44-205-247-247.compute-1.amazonaws.com:/home/hadoop/frankenstein.tx

t \

/c/ProyectosIA/Actividad\_7/

frankenstein.txt

100% 431KB 515.0KB/s 00:00

11. Capturas de la interfaz de descarga, el resultado de la misma y la línea de comando para la descarga

#### **4. Conclusión**

Esta actividad ha permitido familiarizarme con HDFS en un entorno empresarial real usando Amazon EMR.

He aprendido a subir datos al clúster, transferirlos a HDFS y verificar su correcto almacenamiento mediante línea de comandos básicos.