# Actividad de refuerzo nº.7. HDFS

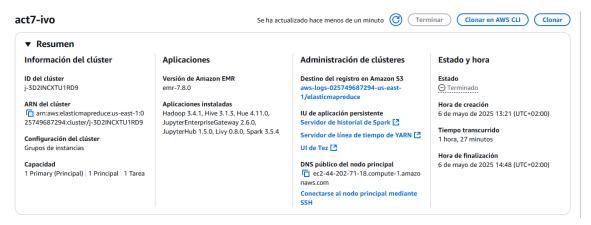
## Realizado con AWS

### 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.

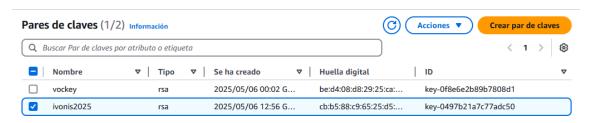
## 2. Preparación del entorno

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



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.

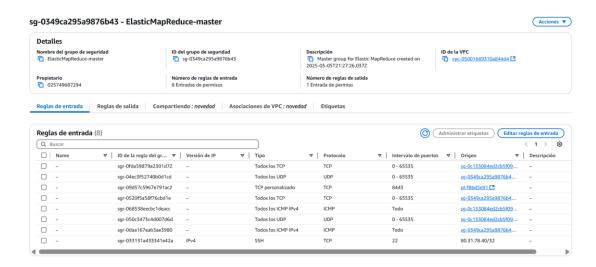


2 La clave ivonis2025.pem fue generada para habilitar el acceso SSH.

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



 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.

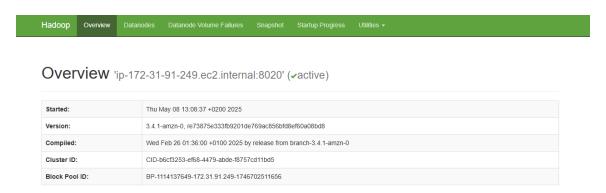
4. Captura de la ejecución de comandos en la terminal: conexión SSH, creación de directorios, subida de archivos a HDFS.

Se realiza visualización de la primera página.

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.

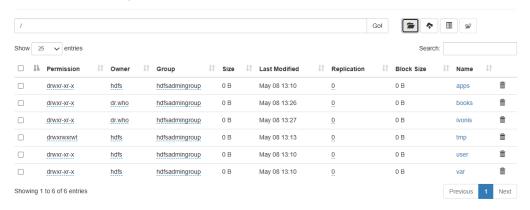
En este caso se opta por usar la interfaz del navegador.



6. Interfaz del navegador.

## Se crea en raíz /books y /ivonis

## **Browse Directory**



7. Contenidos del directorio raíz.

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



8. Verificación de la carga.

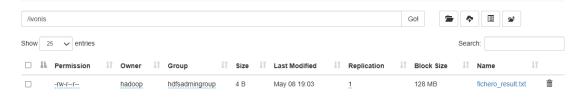
## Cargamos "fichero\_results.zip" y lo descomprimimos

```
[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**



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

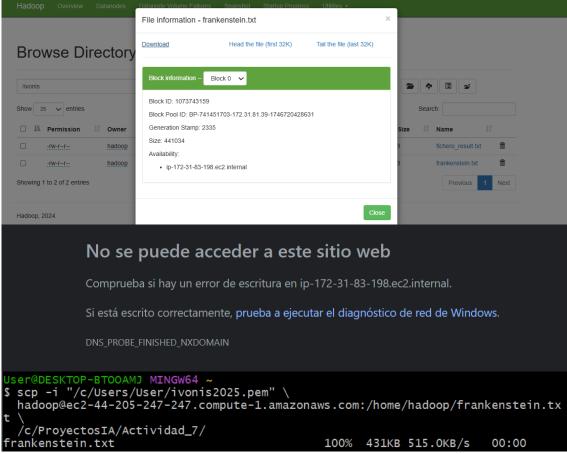
## Actividad 7 - Alumno: Ivonis Florindo López

Se procede a cortar y pegar "frankenstein.txt" a /ivonis

#### **Browse Directory** Show 25 v entries ↓↑ Size 11 Last Modified ↓↑ Block Size ☐ ♣ Permission ♣ Owner ♣ Group Name hdfsadmingroup 4 B -гw-г--г-hadoop -rw-r--r-hadoop hdfsadmingroup Showing 1 to 2 of 2 entries Previous 1

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.



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.