**UNIVERSIDAD EAFIT**

**ST0263 Tópicos Especiales en Telemática**

**2023-2**

Documentación y Evidencias Laboratorios Unidad 3 – Big Data

**Nombre :** David González Tamayo **- email :** dgonzalez2@eafit.edu.co

LABORATORIO 3-1: Gestión de archivos en HDFS y s3 para BIG DATA

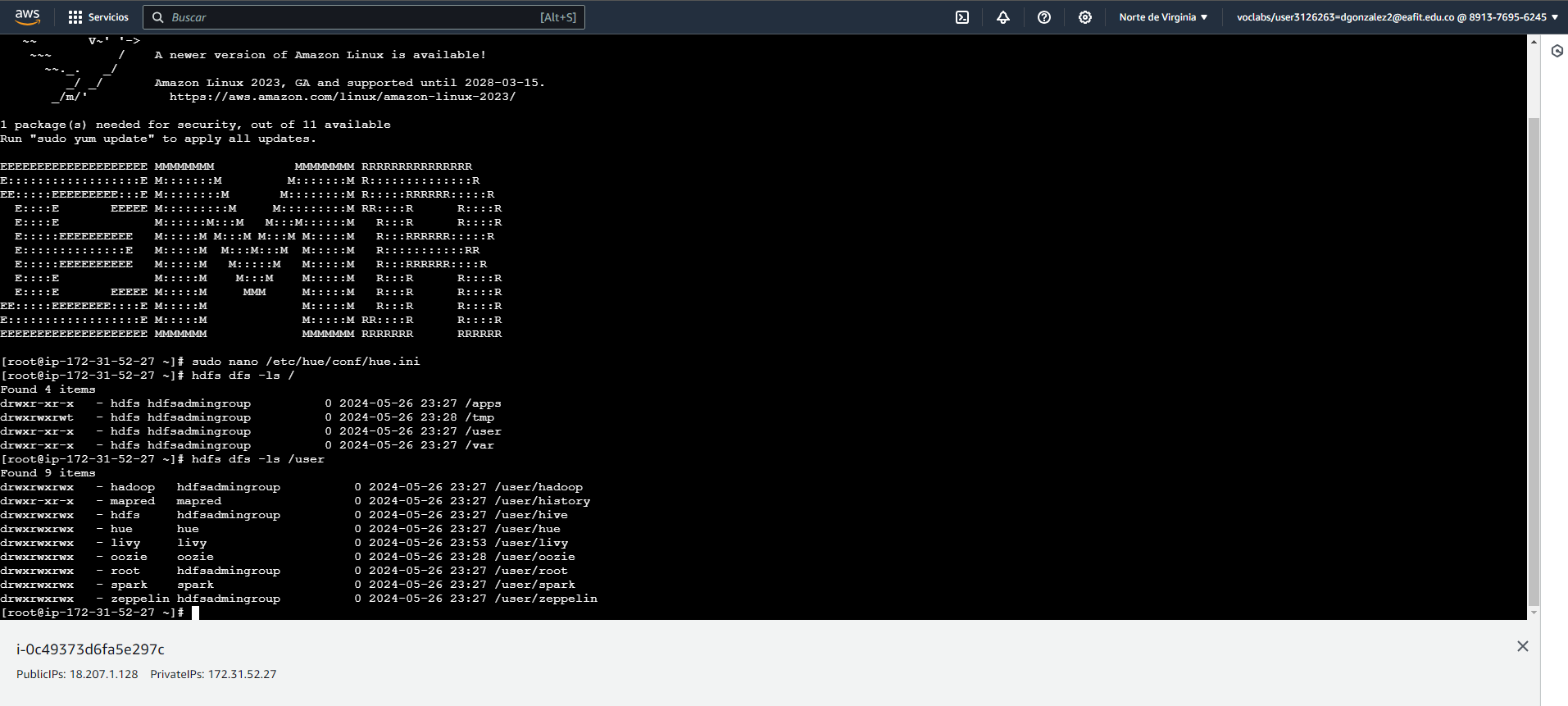
Para la creación de este laboratorio se basó con la guía dada por el profesor Edwin Nelson Montoya Munera. Para más detalles dar click en el siguiente link

* <https://github.com/st0263eafit/st0263-241/tree/main/bigdata/01-hdfs>

Para lograr este laboratorio nos debemos de conectar al nodo máster por medio de SSH:

Ahora, con los siguientes comandos, realizaremos diversas operaciones en el sistema de archivos HDFS.

Vamos a listar y crear carpetas desde la terminal al hdfs. Están se podrán ver desde Hue:



Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Ahora, con el siguiente comando subiremos desde la terminal al hdfs el datasets de gutenberg small



Verificamos, en Hue si podemos ver los archivos que se acabaron de montar:

Una captura de pantalla de una computadora

Descripción generada automáticamente

Los cuales fueron subidos exitosamente.

Una captura de pantalla de una computadora

Descripción generada automáticamente

**Copiar (gestión) de archivos hacia el HDFS vía HUE.**

Ahora, para subir archivos a través de Hue, se mostrará cómo cargar los archivos de la ONU. Primero, se crea una carpeta llamada "onu" para almacenar los archivos.

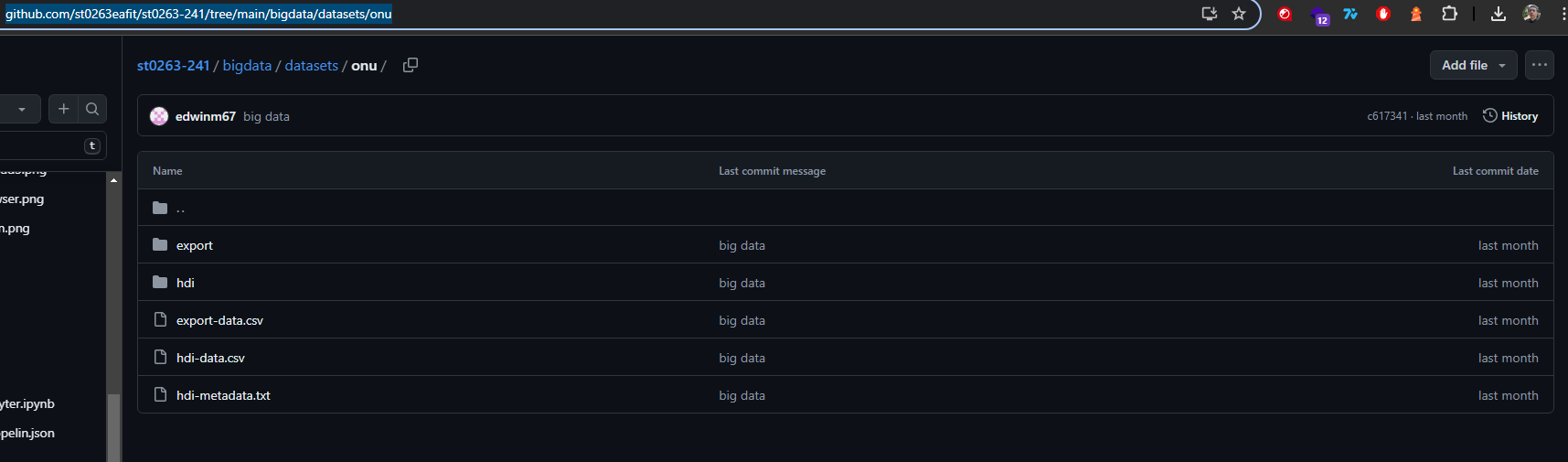
Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente

Luego, entramos a la carpeta **onu** y subimos los archivos ubicados en el github del profesor:

<https://github.com/st0263eafit/st0263-241/tree/main/bigdata/datasets/onu>

Los cuales son ‘export-data.csv’ y ‘hdi-data.csv’.



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Si entramos a export-data.csv observamos lo siguiente:

Interfaz de usuario gráfica, Texto

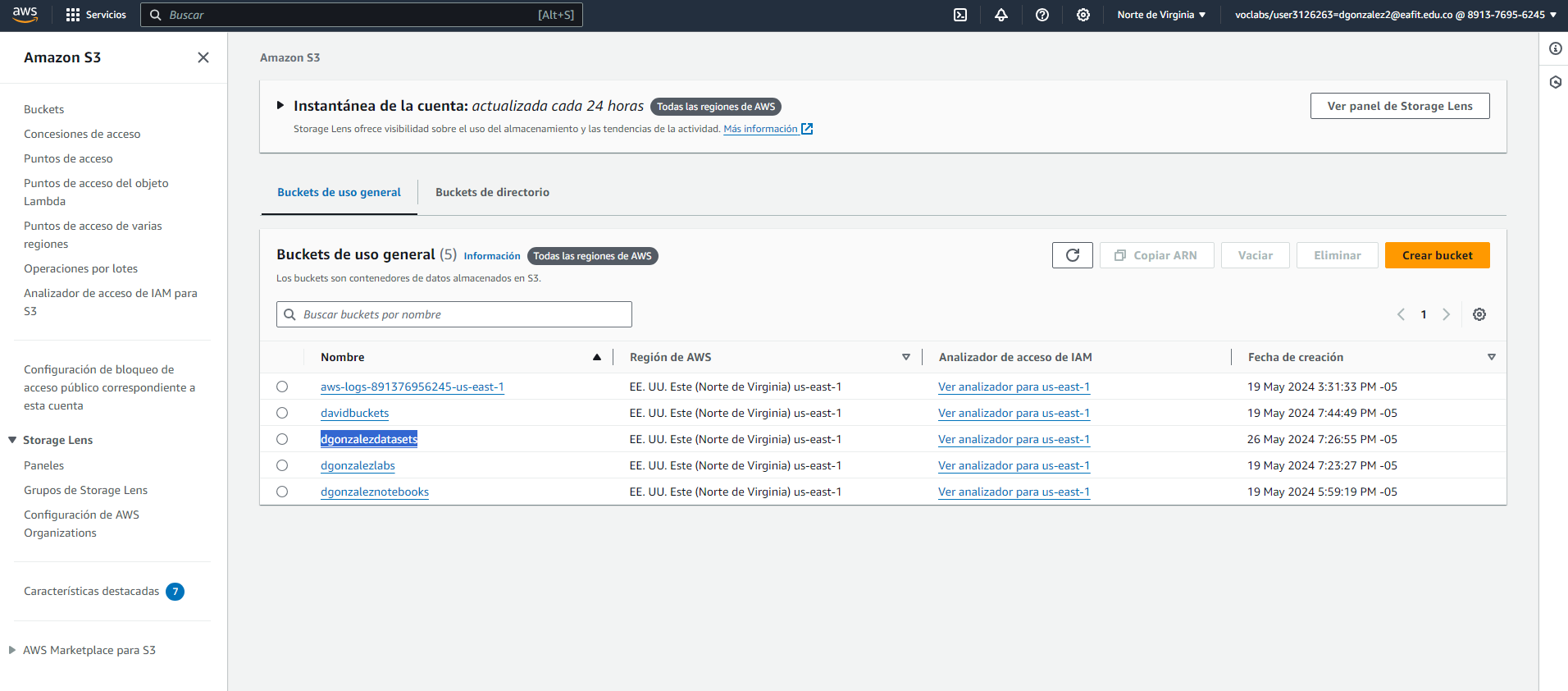
Descripción generada automáticamente

Creamos un bucket llamado ‘dgonzalezdatasets’.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Si nos ubicamos en el S3 de AWS, veremos el bucket creado:



Luego nos ubicamos de nuevo en el HUE y creamos la carpeta onu para subir todos los archivos:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Una vez subido los archivos se mostrará lo siguiente:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Entramos al ‘Buckets’ de AWS, al apartado de ‘onu’ y verificamos si se guardaron los archivos subidos anteriormente.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico, Sitio web

Descripción generada automáticamente

**Copiar (gestión) de archivos hacia el AWS S3 vía SSH**

Ahora subiremos todos los conjuntos de datos del repositorio utilizando la consola. Esto se hará con el siguiente comando.

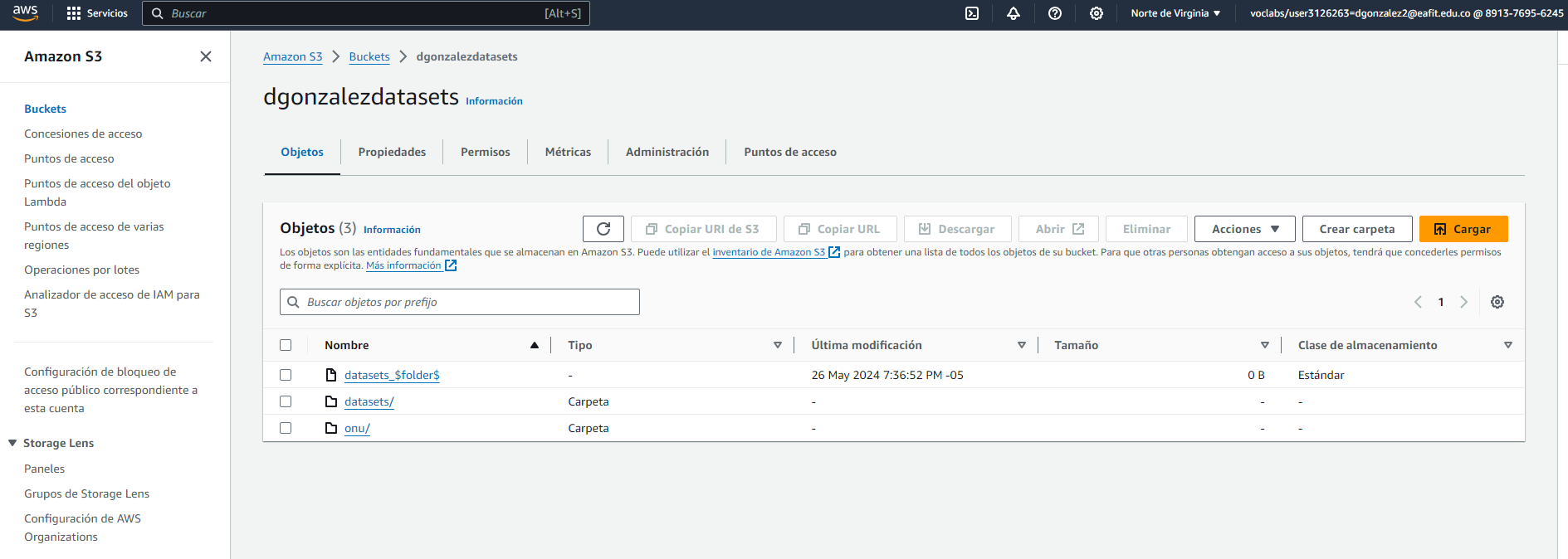
Texto

Descripción generada automáticamente con confianza media

Imagen de la pantalla de un computador

Descripción generada automáticamente con confianza baja

Ahora podemos ver en S3 de AWS todos los datasets que hay:



Abrimos el ‘datasets’ y nos mostrara lo siguiente:

Una captura de pantalla de una computadora

Descripción generada automáticamente