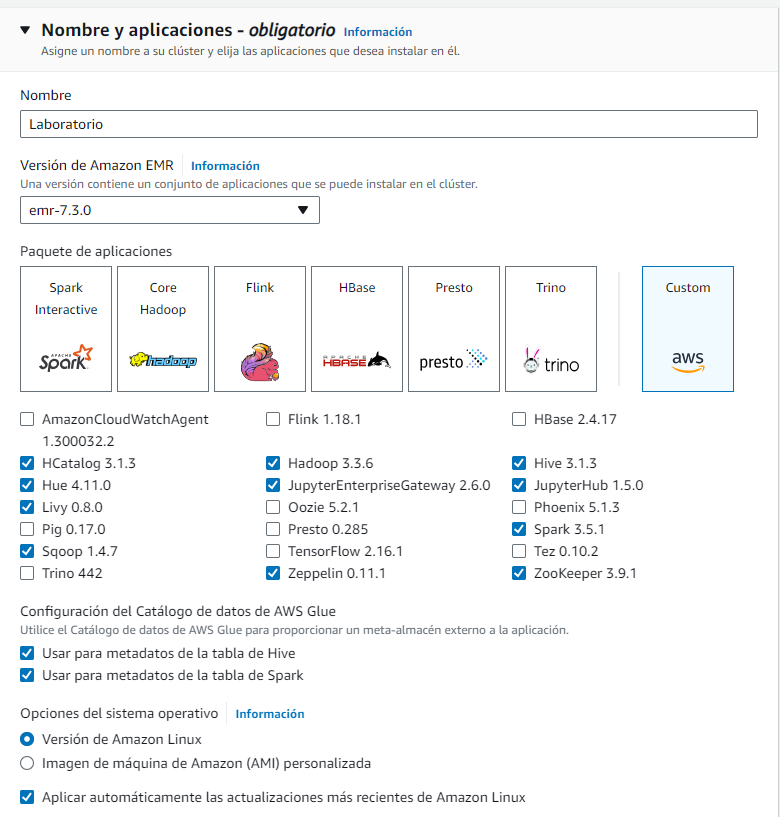
**Nombre: \_\_\_\_\_**Andrés Felipe Téllez Rodríguez**\_\_\_\_ - email: \_\_\_\_\_**aftellezr@eafit.edu.co**\_\_\_**

Laboratorios de clase:

Laboratorio 3.0 (EMR):





Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

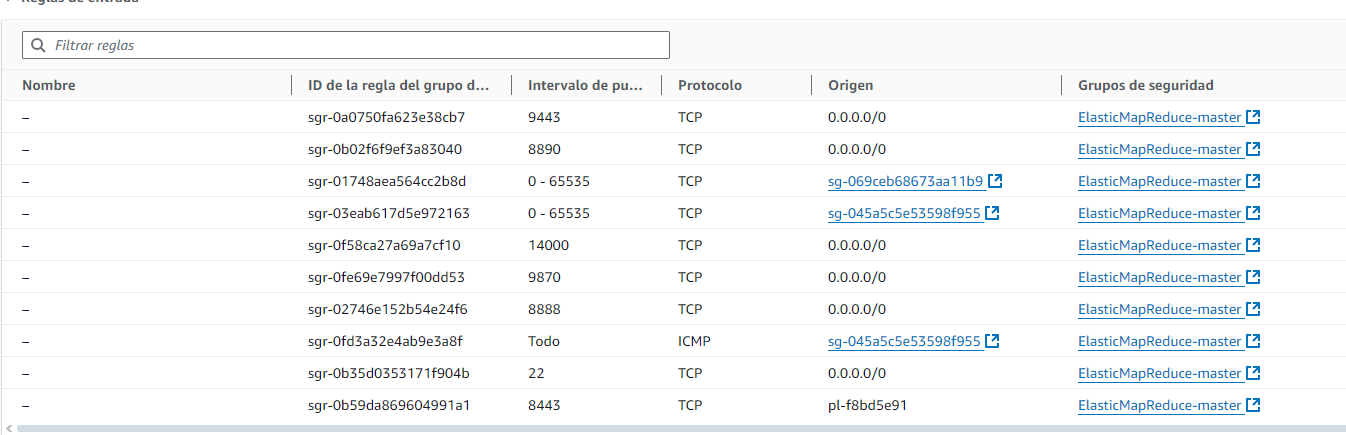
Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente



Laboratorio 3.1 (HDFS):



Texto

Descripción generada automáticamente

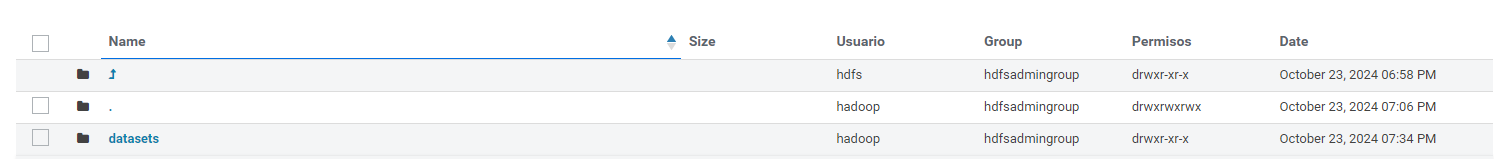
Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

1. Evidencias de HDFS y S3 desde EMR



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Evidencias de Hive desde EMR

Tabla

Descripción generada automáticamente

1. Evidencias de spark desde EMR y Google Colab

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Laboratorio 3.2 (MapReduce):

Texto

Descripción generada automáticamente

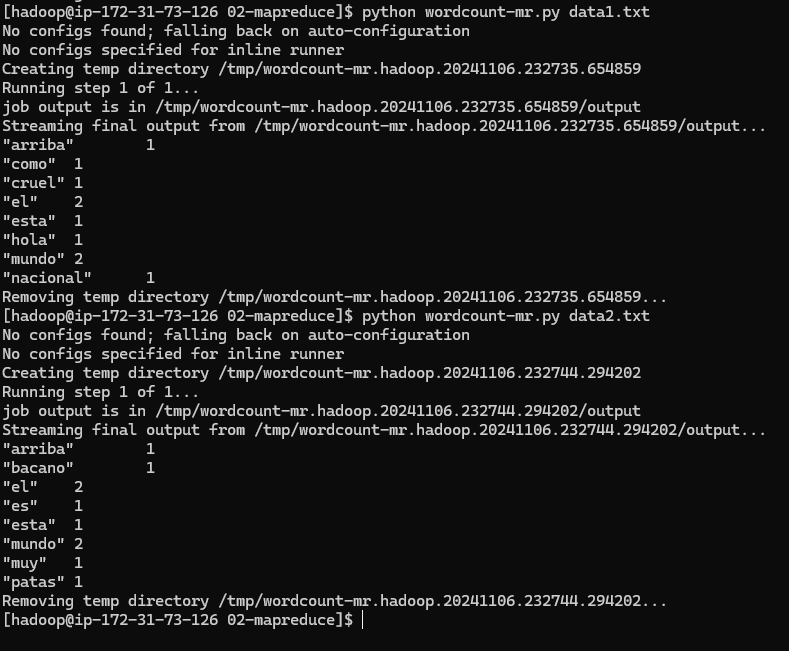
Texto

Descripción generada automáticamente



Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente



Laboratorio 3.3 (Hive-Sparksql):

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente**

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

****

Para poder llenar los resultados de la tabla con los resultados del query que se nos da el ejemplo, es decir del wordcount. Se deben de seguir algunos pasos.

El primer paso es crear la tabla de destino. Un ejemplo de esto es:

CREATE TABLE IF NOT EXISTS frecuencias\_palabras (

palabra STRING,

frecuencia INT

);

Después se ejecuta el query que se propone. Y de ahí se guarda los resultados en la tabla de destino

INSERT INTO TABLE frecuencias\_palabras

SELECT palabra, COUNT(\*) AS frecuencia

FROM (

SELECT explode(split(contenido, ' ')) AS palabra

FROM texto

) AS palabras

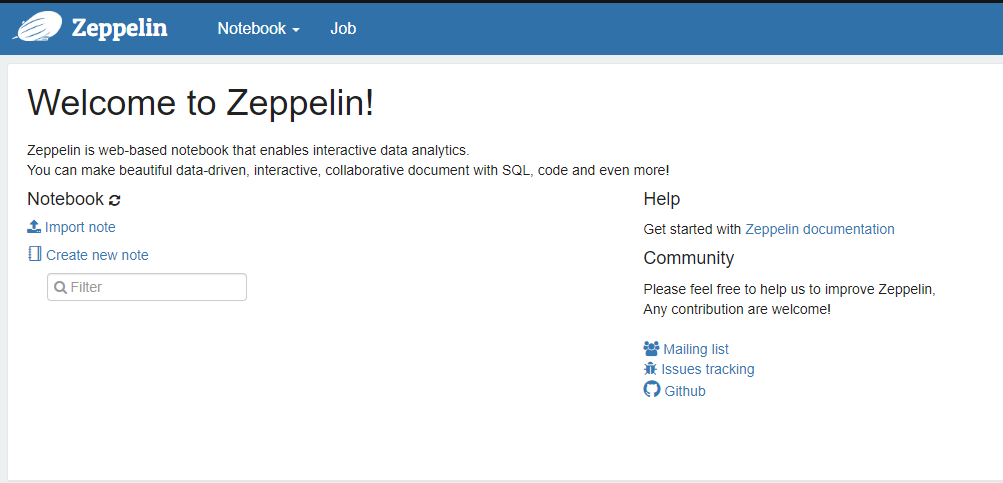
GROUP BY palabra;

Y por último después de ejecutar el query se revisa que los resultados que se esperan si estén dentro de la tabla de destino:

SELECT \* FROM frecuencias\_palabras;

También si es necesario se envía el resultado de la consulta a una carpeta en hue o S3 y subirlo a la base de datos de destino

Laboratorio 3.4 (Spark):



Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente