La finalidad de esta práctica es familiarizarse con el entorno HIVE.

Vete haciendo capturas de pantalla de todos los pasos que vayas dando, acompañándolas de comentarios descriptivos de los mismos.

INTRODUCCIÓN

A.- Continuaremos trabajando con el conjunto de datos **Movielens** de **Kaggle:**

https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset

CONTENIDO

APARTADO A

Práctica con HIVE

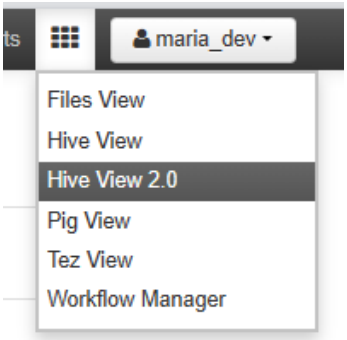
De modo similar a como hicimos con PIG, por cada uno de los *scripts* siguientes:

Prueba su ejecución tanto desde la consola de Hive

```
[maria_dev@sandbox-hdp ~]$ hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive>
```

como desde Ambari Hive View 2.0.



- Muestra el contenido de los *scripts*.
- Muestra un ejemplo de su ejecución con la salida de los datos por pantalla tanto en la consola como desde el entorno de Ambari.
- Posteriormente guarda las consultas en un archivo y ejecútalas todas juntas desde la línea de comandos.

1.- Crea una base de datos que llamaremos **movielens** para almacenar las tablas necesarias. Para cada una de las consultas deberás crear previamente las tablas y cargar los datos necesarios para poder realizarlas.

2.- Encontrar las 10 ocupaciones más frecuentes entre los votantes

3.- Y luego el número de hombres y mujeres

4.- Muestra la edad media por géneros.

5.- Muestra la edad media por ocupaciones.

6.- Encontrar las cinco películas (código, título y número de votos) más votadas (recuento de votos, no media).