

# Práctica 3 - PIG

## Ejecución de script .pig

En primer lugar, crearemos un directorio donde trabajar esta práctica:

- `cd workspace`
- `mkdir pig`
- `cd pig`

Descargamos un fichero csv de opiniones de productos obtenido desde Kaggle:

- `wget https://github.com/profemronda/archivosPDF/blob/main/Eopinions.csv`
- `ls`

Vemos el contenido del fichero, donde se aprecian 2 campos:

- `gedit Eopinions.csv`

Pasamos este fichero a hdfs, en una nueva carpeta que debemos crearnos:

- `hdfs dfs -mkdir pig`
- `hdfs dfs -put *.csv /user/cloudera/pig`
- `hdfs dfs -ls pig`

Descargamos un script para procesar las opiniones:

- `wget https://github.com/profemronda/archivosPDF/blob/main/proceso_opiniones.pig`

Vemos el contenido:

- `gedit proceso_opiniones.pig`

Ejecutamos el script:

- `pig proceso_opiniones.pig`

Mostramos el resultado de la ejecución del script en pig:

- `hdfs dfs -ls /user/cloudera/pig/out`
- `hdfs dfs -cat /user/cloudera/pig/out/* | head`
- `hdfs dfs -cat /user/cloudera/pig/out/* | tail`