**Análisis sentimientos**

1. APARTADO A
   1. Siguiendo los pasos que se explican en el tema, crea un clúster Hadoop EMR con 1 master y dos nodos. Selecciona la opción Core Hadoop (versión 7.0.0). No te olvides de seleccionar Sqoop, ya que lo utilizaremos en las prácticas siguientes.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Conexión

Texto

El contenido generado por IA puede ser incorrecto.

1. APARTADO B
   1. Viene en formato Excel. Desde el propio Excel puedes convertirlo a formato ’csv’.

Exportamos

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

* 1. Crea una carpeta con tu nombre en el directorio user del HDFS de EMR.

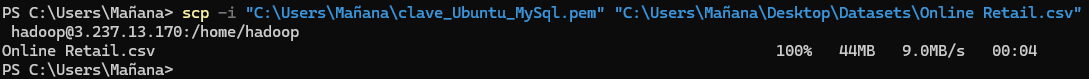
Texto

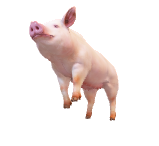
El contenido generado por IA puede ser incorrecto.

* 1. Crea dentro de él una carpeta llamada ventas y sube a ella el ’csv’ que obtuviste anteriormente.



Lo subimos a la machine



1. APARTADO C
   1. Usando Pig
      1. Cargar los datos del dataset en PIG
   2. Consultas:
      1. ¿Cuántos registros tiene la tabla?
      2. Mostrar registros con cantidades mayores o iguales a cero.
      3. A partir de la consulta anterior, mostrar registros precio mayor a cero
      4. A partir de la consulta anterior, mostrar solamente los registros con algún valor en el campo CustomerID.
      5. ¿Cuántos registros tiene la última consulta?
      6. Almacena la consulta final del punto 4 en un fichero llamado ventas.csv dentro de la carpeta de apartado B.
2. APARTADO D
   1. Crear base de datos
   2. Crear tabla externa sobre los datos RAW (CSV)
   3. Hive no maneja muy bien el formato de fecha original, conviértelo a d/M/yyyy H:mm
   4. Crea la misma estructura de tabla, pero particionada por año y mes.
   5. Inserta los registros del punto 1.2 en la tabla particionada.
3. APARTADO E

Análisis de clientes

* 1. 10 clientes con mayor gasto total
  2. Clientes con más compras (cantidad de facturas)

Análisis de productos

* 1. 10 productos más vendidos
  2. 10 productos más rentables (suma de precio unitario por cantidad)

Análisis geográfico

* 1. Países con mayor volumen de ventas

Análisis temporal

* 1. Ventas totales por mes (suma de precio unitario por cantidad)
  2. Hora del día con más actividad

1. APARTADO F

SQOOP

* 1. En tu servidor MySQL en la máquina EC2 crea una base de datos y una tabla para almacenar los datos del fichero ventas.csv.
  2. Exporta con SQOOP los datos de ventas.csv a la tabla que creaste en el punto anterior