MOD 6 – INFRAESTRUCTURA BIG DATA

APARTADO B

**1. Accede a la página minhap.gov y descarga el fichero de deuda ayuntamientos del año 2016**

**Se ha descargado el archivo indicado: *deuda\_viva\_ayuntamientos\_2016\_\_ovel\_20170630.xlsx***

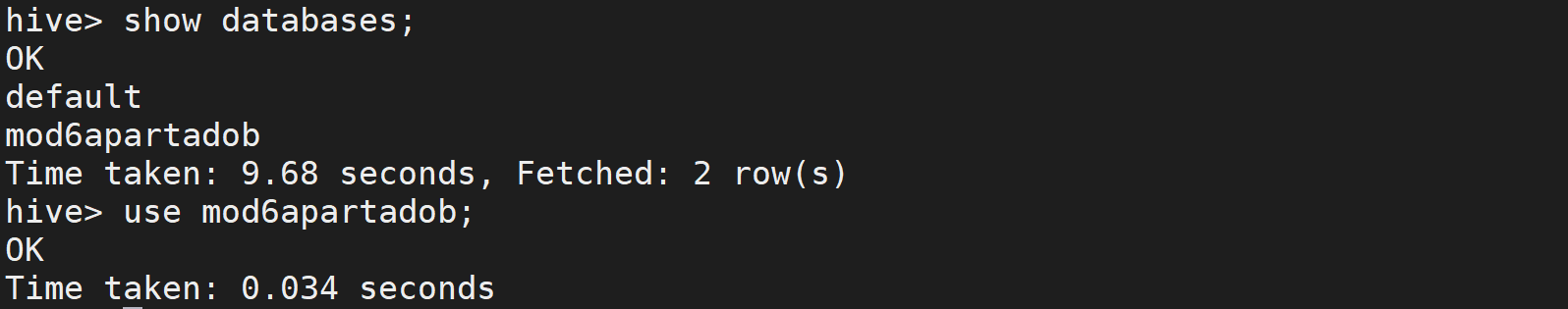
****

**El archivo se abrió y se analizó la información. El archivo contiene tres hojas de datos relativos a Deuda Viva (31/12/2016 – Miles €) por municipio**

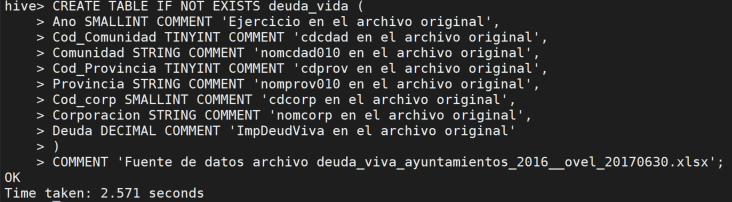
* **Datos: 8125 registros, con columnas:**
  + **Ejercicio (Ejercicio)**
  + **Código CC.AA. (cdcdad)**
  + **CC.AA (nomcdad010)**
  + **Código Provincia (cdprov)**
  + **Provincia (nomprov010)**
  + **Código Municipio (cdcorp)**
  + **Municipio (nomcorp)**
  + **Deuda viva 31/12/2016 (ImpDeudViva)**
* **Datos\_format: Deuda Viva entidades locales - 8124 registros (+1 relativo a “Diferencias de conciliación con el Banco de España), con columnas:**
  + **Código CC.AA. (cdcdad)**
  + **Código Provincia (cdprov)**
  + **Código Municipio (cdcorp)**
  + **Municipio (nomcorp)**
  + **Deuda viva 31/12/2016 (ImpDeudViva)**
* **Resumen\_CCAA\_Prov: datos agregados por comunidades autónomas y provincias – 71 registros, con columnas:**
  + **Codigo CA**
  + **CC.AA.**
  + **Codigo Provin**
  + **Provincia**
  + **Deuda 31/12/2016 (Miles de €)**

**2. Indicar el comando de creación de una tabla en Hive que contenga la información de la solapa Datos con las siguientes columnas: Año, Cod\_Comunidad, Comunidad, Cod\_Provincia, Provincia,** **Cod\_corp, Corporacion, Deuda**

**En primer lugar, se creó una base de datos ‘mod6apartadob’ para el ejercicio:******



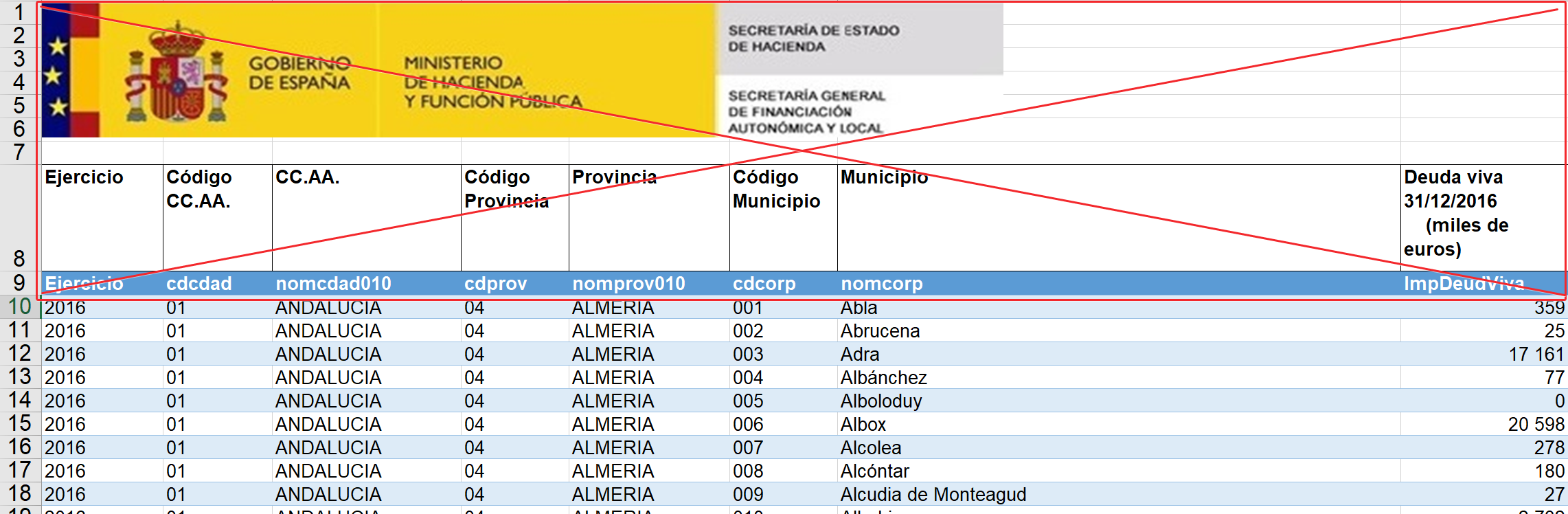
De seguida se creó la tabla ‘datos’:



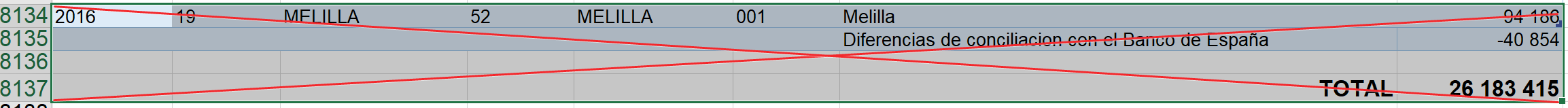
**3. Cargar el fichero anterior descargado en esa tabla e indicar el comando empleado**

Primero, se preparó el archivo descargado para su procesamiento en Hive:

* Eliminar las 9 primeras líneas para quedarnos con las filas que solo poseen datos para análisis



* También eliminar las 3 ultimas líneas



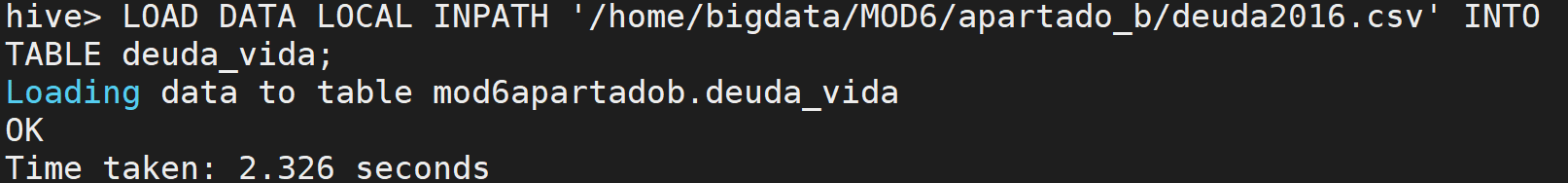
* Eliminar el exceso de espacios en blanco en la columna nomcorp (usando la función EXCEL TRIM)
* Formatee la columna ImpDeudaViva a un número con dos decimales

Finalmente lo guardamos como fichero CSV (UTF-8) – delimitador ‘;’ en la ruta /home/bigdata/MOD6/apartado\_b/

Después, en Hive cargar el fichero deuda2016.csv en la tabla *deuda\_vida:*

Use mod6apartadob;

LOAD DATA LOCAL INPATH '/home/bigdata/MOD6/apartado\_b/deuda2016.csv' INTO TABLE deuda\_vida;

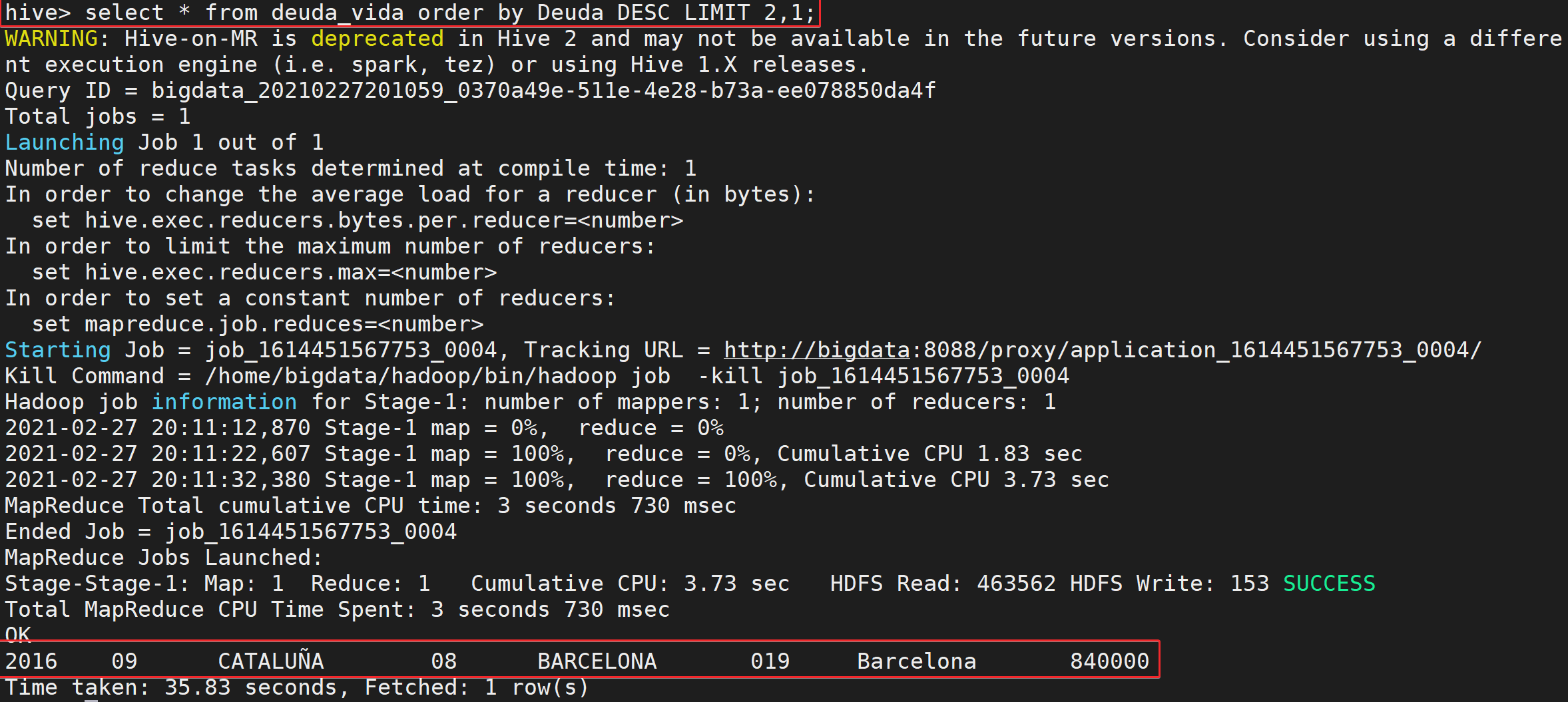


**4. Obtener la corporación que tiene la tercera deuda más grande**

Para encontrar la corporación con la n deuda más grande, se puede utilizar la siguiente consulta:

SELECT \* FROM deuda\_vida ORDER BY Deuda DESC LIMIT (n-1),1;

Luego se ejecutó la siguiente consulta:



La tercera mayor deuda es entonces la de Barcelona = 840 000 miles de euros.