

PRÁCTICA Nº 2.6

Pig



- Nombre y apellidos: Alvaro Lucio-Villegas de Cea



Índice

Tarea.	3
Script	3
Salida	4

Tarea.

A partir del csv de los aparcamientos de bicicletas:

Desarrolla un script en Pig Latin que almacene en un archivo el conteo de los aparcamientos agrupándolos por número de anclajes.

Script

Comandos:

```
#Cargamos los datos en la variable "bici"

bicis = LOAD 'aparcamiento-bicicleta.csv' USING PigStorage(',') as
(id:chararray,title:chararray,tipo:chararray,plazas:int,anclajes:int,lastUpdated:chararray,geometry:chararray,icon:chararray);

#Agrupamos por la columna anclajes

grupo_anclajes = GROUP bicis BY anclajes;

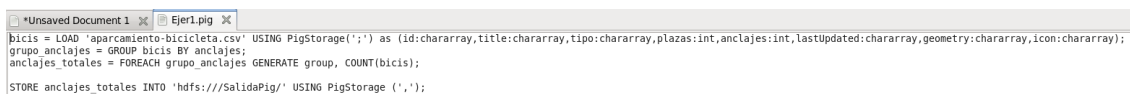
#Generamos los grupos donde se almacenaran el conteo de los anclajes segun su tipo.

anclajes_totales = FOREACH grupo_anclajes GENERATE group, COUNT(bicis);

#Se almacena la variable en un fichero en HDFS

STORE anclajes_totales INTO 'hdfs:///SalidaPig/' USING PigStorage(',');
```

Fichero:



```
*Unsaved Document 1 x Ejer1.pig x
bicis = LOAD 'aparcamiento-bicicleta.csv' USING PigStorage(',') as (id:chararray,title:chararray,tipo:chararray,plazas:int,anclajes:int,lastUpdated:chararray,geometry:chararray,icon:chararray);
grupo_anclajes = GROUP bicis BY anclajes;
anclajes_totales = FOREACH grupo_anclajes GENERATE group, COUNT(bicis);
STORE anclajes_totales INTO 'hdfs:///SalidaPig/' USING PigStorage(',');
```



Salida

Salida del Script, generación del fichero con los datos solicitados:

