

## Ejercicios PIG

1. Copiar en local file sistema de la MV el fichero datos\_pig.txt en la ruta /home/cloudera/ejercicios/pig y abrir el fichero para revisar su contenido.
2. Arranca el Shell de Pig en modo local  
`"pig -x local"`
3. Carga los datos en pig en una variable llamada "data". Los nombres de las columnas deben ser (key, campana, fecha, tiempo, display, accion, cpc, pais, lugar). Los tipos de las columnas deben ser chararray excepto accion y cpc que son int.  

```
data = LOAD
'/home/cloudera/ejercicios/pig/datos_pig.txt' AS
(key:chararray, campana:chararray, fecha:chararray,
tiempo:chararray, display:chararray, accion:int,
cpc:int, pais:chararray, lugar:chararray);
```
4. Usa el comando DESCRIBE para ver el esquema de la variable "data"
5. Selecciona las filas de "data" que provengan de USA  
`result = FILTER data BY pais=='USA';`
6. Listar los datos que contengan en su key el sufijo surf:  
`result = FILTER data BY key MATCHES '^surf.*';`
7. Crear una variable llamada "ordenado" que contenga las columnas de data en el siguiente orden: (campana, fecha, tiempo, key, display, lugar, accion, cpc)  

```
ordenado = FOREACH result GENERATE campana, fecha,
tiempo, key, display, lugar, accion, cpc;
```
8. Guarda el contenido de la variable "ordenado" en una carpeta en el local file system de tu MV llamada resultado en la ruta /home/cloudera/ejercicios/pig  

```
STORE ordenado INTO
'/home/cloudera/ejercicios/pig/resultado';
```
9. Comprobar el contenido de la carpeta.