

duEjercicios 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.

```
cp /mnt/Shared/datos_pig.txt pig/datos_pig.txt
```

2. Arranca el Shell de Pig en modo local
 - a. `"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.

- a. `Data= LOAD 'datos_pig.txt' USING PigStorage(',') as (key: chararray, campana: chararray, fecha: chararray, tiempo: chararray, display: chararray, accion: int, cpc: int, país: chararray, lugar: chararray)`

```
data= LOAD 'datos_pig.txt' USING PigStorage(',') as
(key:chararray, campana:chararray, fecha:chararray,
tiempo:chararray, display:chararray, accion:int, opc:int,
pais:chararray, lugar:chararray);
```

4. Usa el comando DESCRIBE para ver el esquema de la variable "data"

```
grunt> DESCRIBE data;
data: {key: chararray, campana: chararray, fecha: chararray, tiempo: chararray, disp
lay: chararray, accion: int, opc: int, pais: chararray, lugar: chararray}
```

5. Selecciona las filas de "data" que provengan de USA;
 - a. `FilasUSA=FILTER data by pais == USA;`

```
FilasUSA = FILTER data by pais == 'USA';
```

6. Listar los datos que contengan en su key el sufijo surf:

```
datosSurf= 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)
 - a. `ordenado = FOREACH data 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
 - a. `STORE ordenado INTO '/home/Cloudera/ejercicios/pig/resultado';`
9. Comprobar el contenido de la carpeta.

```
[root@quickstart resultado]# tail part-m-00000
```

	student	D1	05/31/2013	23:59:13	lawnmast
er.example.com	0	116	USA	INLINE	
	mobile	D3	05/31/2013	23:59:19	dvdrevie
w.example.com	0	120	USA	SIDE	
	train	A5	05/31/2013	23:59:24	bitpress
.example.com	0	120	USA	TOP	
	NEWS	D5	05/31/2013	23:59:27	datasnap
.example.com	0	102	USA	SIDE	
	bargain	A2	05/31/2013	23:59:32	masterba
ker.example.com	0	110	USA	BOTTOM	
	travel	B5	05/31/2013	23:59:45	pcexpert
.example.com	0	116	USA	INLINE	
	small	B1	05/31/2013	23:59:48	dealmonk
ey.example.com	0	120	NETHERLANDS	TOP	
	small	D8	05/31/2013	23:59:52	burritof
inder.example.com		0	118	USA	INLINE
	media	C4	05/31/2013	23:59:53	audioexp
ert.example.com	0	106	NETHERLANDS	BOTTOM	
	holiday	C2	05/31/2013	23:59:54	salestig
er.example.com	0	120	USA	TOP	