

## Ejercicio 1 Pig:

### Parte 1

Datos de entrada (fichero separado por comas):

Index	Field	Data Type	Description	Example
0	keyword	chararray	Keyword that triggered ad	tablet
1	campaign_id	chararray	Uniquely identifies our ad	A3
2	date	chararray	Date of ad display	05/29/2013
3	time	chararray	Time of ad display	15:49:21
4	display_site	chararray	Domain where ad shown	www.example.com
5	was_clicked	int	Whether ad was clicked	1
6	cpc	int	Cost per click, in cents	106
7	country	chararray	Name of country in which ad ran	USA
8	placement	chararray	Where on page was ad displayed	TOP

Datos de Salida:

Index	Field	Description
0	campaign_id	Uniquely identifies our ad
1	date	Date of ad display
2	time	Time of ad display
3	keyword	Keyword that triggered ad
4	display_site	Domain where ad shown
5	placement	Where on page was ad displayed
6	was_clicked	Whether ad was clicked
7	cpc	Cost per click, in cents

### Operaciones a realizar:

- 1) Cargar datos con el formato correcto
- 2) Filtrar los datos seleccionando únicamente los registros que provengan de USA
- 3) Generar el nuevo formato de salida (Foreach...Generate)
- 4) Convertir el campo keyword a mayúsculas y eliminar los que vengan en blanco

## Parte 2:

Dado el fichero del paso anterior (fichero separado por comas):

Index	Field	Data Type	Description	Example
0	campaign_id	chararray	Uniquely identifies our ad	A3
1	date	chararray	Date of ad display	05/29/2013
2	time	chararray	Time of ad display	15:49:21
3	display_site	chararray	Domain where ad shown	www.example.com
4	placement	chararray	Where on page was ad displayed	TOP
5	was_clicked	int	Whether ad was clicked	Y
6	cpc	int	Cost per click, in cents	106
7	keyword	chararray	Keyword that triggered ad	tablet

Realizar:

1. Carga del fichero
2. Eliminar registros duplicados
3. Generar el siguiente formato

Index	Field	Description
0	campaign_id	Uniquely identifies our ad
1	date	Date of ad display
2	time	Time of ad display
3	keyword	Keyword that triggered ad
4	display_site	Domain where ad shown
5	placement	Where on page was ad displayed
6	was_clicked	Whether ad was clicked
7	cpc	Cost per click, in cents

4. Modificar el formato de fecha:

-Formato de fecha actual: MM-DD-YYYY

-Formato deseado: MM/DD/YYYY

## **Ejercicio 2 Pig:**

### **Parte1:**

- 1) Filtrar los registros que únicamente tengan el campo was\_clicket = 1
- 2) Agrupar por el campo "display\_site"
- 3) Generar una nueva relación que incluya 2 campos: display\_site y el coste del número total de clicks
- 4) Ordenar por coste en orden ascendente
- 5) Mostrar únicamente los 3 primeros registros

### **Parte2:**

- 1) Modificar el anterior script para agrupar por el campo "keyword"
- 2) Ordenar de forma descendente
- 3) Mostrar los primeros 5 resultados