



## EJERCICIO DEL TEMA 4

### MODELOS DE INFORMACIÓN

El objetivo de esta práctica es trabajar con algunas de las herramientas de modelado de la información revisadas en el material teórico

#### *APARTADO 1*

Hay muchos ejemplos en la web. Uno muy completo puede encontrarse en:

<http://hadoopontheroad.blogspot.ch/2013/02/mapreduce-ejemplo-teorico.html>

Imaginemos que tenemos un fichero con las puntuaciones de un casino. La estructura es sencilla: “fecha del juego, nombre jugador y puntuación”. Por ejemplo:

01-11-2012 Maria 11  
01-11-2012 Pablo 9  
01-11-2012 Angel 3  
23-11-2012 Pablo 22  
23-11-2012 Maria 15  
15-12-2012 Pablo 32  
15-12-2012 Maria 47  
15-12-2012 Angel 13  
01-01-2013 Pablo 2  
01-01-2013 Maria 3  
01-01-2013 Angel 32

A partir de estos datos queremos sacar la suma total de puntuaciones de cada usuario.



Las siguientes líneas van a representar los pares (key, value) que recibe la función map teniendo en cuenta que las key son el offset de la línea (en bytes), y los value son la línea de texto entera:

(0, 01-11-2012 Maria 11)  
(22, 01-11-2012 Pablo 9)  
(45, 01-11-2012 Angel 3)  
(65, 23-11-2012 Pablo 22)  
(86, 23-11-2012 Maria 15)  
(110, 15-12-2012 Pablo 32)  
(134, 15-12-2012 Maria 47)  
(158, 15-12-2012 Angel 13)  
(182, 01-01-2013 Pablo 2)  
(205, 01-01-2013 Maria 3)  
(225, 01-01-2013 Angel 32)

La función map lo que hará entonces será tratar cada línea recogiendo sólo la información que nos interesa y omitiendo el resto, para finalmente emitir para cada línea un par key/value con el nombre del jugador como key y la puntuación como value:



(Maria, 11)

(Pablo, 9)

(Angel, 3)

(Pablo, 22)

(Maria, 15)

(Pablo, 32)

(Maria, 47)

(Angel, 13)

(Pablo, 2)

(Maria, 3)

(Angel, 32)

Esta salida de la función map va a ser tratada a continuación por el Shuffle and Sort antes de enviar los datos a la función reduce. En esta fase las key se van a ordenar y los value se van a agrupar por cada key, y el resultado será la entrada de la función reduce, quedando de la siguiente forma:

(Angel, [3, 13, 32])

(Maria, [11, 15, 47, 3])

(Pablo, [9, 22, 32, 2])

Así que la función reduce va a recibir como entrada cada una de estas líneas, siendo una key y por cada key una lista de value. Por cada línea el reduce emitirá la suma de las puntuaciones de cada usuario:

(Angel, 48)

(Maria, 76)

(Pablo, 65)



## APARTADO 2

Describamos por ejemplo, un alojamiento hotelero

```
<rdf:RDF
  xmlns:j.0="http://www.w3.org/ns/locn#"
  xmlns:j.1="http://vocab.linkeddata.es/datosabiertos/def/turismo/lugar#"
  xmlns:j.2="http://www.zaragoza.es/api/def/turismo/lugar#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:j.3="http://www.w3.org/2006/vcard/ns#"
  xmlns:j.4="http://schema.org/"
  xmlns:j.5="http://vocab.linkeddata.es/kos/turismo/"
  xmlns:j.6="http://vocab.linkeddata.es/datosabiertos/def/turismo/alojamiento#"
  xmlns:j.7="http://purl.org/dc/terms/"
  xmlns:j.8="http://www.opengis.net/ont/geosparql#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:j.9="http://www.opengis.net/ont/sf#"
  xmlns:j.10="http://www.w3.org/2003/01/geo/wgs84_pos#" >
<rdf:Description rdf:about="http://www.zaragoza.es/api/recurso/turismo/alojamiento/1091">
  <j.3:postal-code rdf:datatype="http://www.w3.org/2001/XMLSchema#string">50004</j.3:postal-code>
  <j.10:geometry rdf:resource="http://www.zaragoza.es/api/recurso/geometry/WGS84/41.650616_-0.892465"/>
  <j.3:url>http://www.pensionmiramar.es</j.3:url>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PENSIÓN MIRAMAR</rdfs:label>
  <j.4:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">C/ Capitán Casado, 17, 1ºdcha
  &lt;strong&gt;&lt;abbr title="Teléfono"&gt;Tel.:&lt;/abbr&gt;&lt;/strong&gt; 976 28 10 94</j.4:description>
  <j.6:numCamas>17</j.6:numCamas>
  <j.3:street-addr>C/ Capitán Casado, 17, 1ºdcha</j.3:street-addr>
  <j.4:postalCode rdf:datatype="http://www.w3.org/2001/XMLSchema#string">50004</j.4:postalCode>
  <j.2:socioHoreca>>true</j.2:socioHoreca>
  <j.6:numHabitaciones>11</j.6:numHabitaciones>
```



```
<j.7:description
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">C/ Capitán Casado, 17, 1ºdcha &lt;strong&gt;&lt;abbr
title="Teléfono"&gt;Tel.:&lt;/abbr&gt;&lt;/strong&gt; 976 28 10 94</j.7:description>
<j.4:telephone rdf:datatype="http://www.w3.org/2001/XMLSchema#string">976 28 10 94</j.4:telephone>
<j.0:postCode rdf:datatype="http://www.w3.org/2001/XMLSchema#string">50004</j.0:postCode>
<j.7:modified>2015-04-09</j.7:modified>
<j.0:addressArea rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Zaragoza</j.0:addressArea>
<rdf:type rdf:resource="http://vocab.linkeddata.es/kos/turismo/alojamiento"/>
<j.4:addressLocality rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Zaragoza</j.4:addressLocality>
<j.7:identifíer>1091</j.7:identifíer>
<j.3:hasTelephone rdf:datatype="http://www.w3.org/2001/XMLSchema#string">976 28 10 94</j.3:hasTelephone>
<j.3:organization-name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PENSIÓN
MIRAMAR</j.3:organization-name>
<j.3:locality rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Zaragoza</j.3:locality>
<j.6:categoria>Pensión</j.6:categoria>
</rdf:Description>
</rdf:RDF>
```