

Ejercicio 2: Serie temporal - calcular la media del consumo mensual, utilizando map-reduce

Características fichero de carga measures_hourly_test.csv:

- Cantidad de registros: 892.370
- Contiene las mediciones del 31/12/2014 al 02/02/2016.
- Cantidad de registros con medición nula: 8.017
- Cantidad de registros con medición en cero: 138.332
- Total a cargar: 884.353

Proceso de carga

- 1- Se cargó el fichero measure_hourly_test.csv en una base de datos Mongo DB bajo la siguiente estructura, en donde cada columna representa la sección obtenida luego de hacer el split de cada fila por ";". No fueron consideradas aquellas filas en donde la medición es nula.

Fila csv = columna1; columna2; columna3; columna4; columna5; columna6

```
{ "ref": columna1, "date": columna2, "measure": columna3, "period": columna4, "unit": int(columna5), "type": columna6 }
```

Ejemplo de documento insertado:

```
{
  "_id" : ObjectId<"5d64e56b384c769fb3ba7325">,
  "ref" : "ES00000000000000000001",
  "date" : "2015-01-01T19:00:00Z",
  "measure" : 67,
  "period" : "INSTANT",
  "unit" : "kWh",
  "type" : "electricityConsumption\n"
}
```

Código de carga entregado: load_file_mongoDB.py

Obteniendo en total 884.353 documentos.

```
> db.measures.count()
884353
```

- 2- Para la aplicación del map reduce, se desarrollaron las funciones map , reduce y finalize.

Map : A partir de los documentos cargados, genera las duplas con el siguiente formato:

(key, value) = (YYYY-MM, {measure: measure , count:1})

- YYYY-MM: Año y mes de la medición.
- measure:Medición registrada.
- count: Representa la propia medición y se utiliza para el cálculo de la media.

Se definió como key YYYY-MM, dado que se debe informar la media mensual y de este modo el framework realiza la agrupación de datos.

Reduce: Recibe las duplas generadas en el map y se encarga de realizar la reducción que consiste en sumar las measure y count asociados a una misma key.

Finalize: Se encarga de calcular la media con los valores totales de measure y count.

Código entregado: map_reduce_mongo.txt

A continuación se presentan como evidencia las capturas de los resultados obtenidos:

Ejecución mapReduce

```
> db.measures.mapReduce(  
...   function () {  
...     emit( this.date.substr(0,7).trim(),{"measure":this.measure,"count":1}  
...   },  
...   function(keys, values) {  
...     reducedVal = { "measure" : 0 , "count" : 0};  
...     values.forEach(function(value) {  
...       reducedVal.measure += value.measure;  
...       reducedVal.count += value.count;  
...     });  
...     return reducedVal;  
...   },  
...   {  
...     "out": "map_reduce_result" ,  
...     "finalize": function(key, reduceVal) {  
...       reduceVal.avg = reduceVal.measure / reduceVal.count;  
...       return reduceVal;  
...     }  
...   })  
...  
...   "result" : "map_reduce_result",  
...   "timeMillis" : 38545,  
...   "counts" : {  
...     "input" : 884353,  
...     "emit" : 884353,  
...     "reduce" : 10058,  
...     "output" : 24  
...   },  
...   "ok" : 1  
... }
```

Resultados contenidos en colección map_reduce_result (24 documentos):

```
> db.map_reduce_result.find().sort(<<id:1>>).pretty()  
{  
  "_id" : "2015-01",  
  "value" : {  
    "measure" : 1554117,  
    "count" : 35136,  
    "avg" : 44.23147199453552  
  }  
}  
{  
  "_id" : "2015-02",  
  "value" : {  
    "measure" : 1460717,  
    "count" : 31968,  
    "avg" : 45.693099349349346  
  }  
}  
{  
  "_id" : "2015-03",  
  "value" : {  
    "measure" : 1610787,  
    "count" : 34969,  
    "avg" : 46.06328462352369  
  }  
}  
{  
  "_id" : "2015-04",  
  "value" : {  
    "measure" : 1472716,  
    "count" : 34562,  
    "avg" : 42.61084427984492  
  }  
}
```

```

    "_id" : "2015-05",
    "value" : {
      "measure" : 1530309,
      "count" : 37200,
      "avg" : 41.13733870967742
    }
  }

  "_id" : "2015-06",
  "value" : {
    "measure" : 1596608,
    "count" : 36649,
    "avg" : 43.564844879805726
  }
}

  "_id" : "2015-07",
  "value" : {
    "measure" : 1921617,
    "count" : 37200,
    "avg" : 50.67555379746835
  }
}

  "_id" : "2015-08",
  "value" : {
    "measure" : 1710894,
    "count" : 37942,
    "avg" : 45.092351483843764
  }
}

  "_id" : "2015-09",
  "value" : {
    "measure" : 1458842,
    "count" : 36000,
    "avg" : 40.52338888888889
  }
}

  "_id" : "2015-10",
  "value" : {
    "measure" : 1518049,
    "count" : 37513,
    "avg" : 40.46727801028977
  }
}

  "_id" : "2015-11",
  "value" : {
    "measure" : 1402931,
    "count" : 36673,
    "avg" : 38.25514683827339
  }
}

```

```

    "_id" : "2015-12",
    "value" : {
      "measure" : 1586009,
      "count" : 38592,
      "avg" : 41.09683354063018
    }
  }

  "_id" : "2016-01",
  "value" : {
    "measure" : 1545295,
    "count" : 38664,
    "avg" : 39.96728222636044
  }
}

  "_id" : "2016-02",
  "value" : {
    "measure" : 1396822,
    "count" : 36144,
    "avg" : 38.646027003098716
  }
}

  "_id" : "2016-03",
  "value" : {
    "measure" : 1451135,
    "count" : 38832,
    "avg" : 37.36956633704162
  }
}

  "_id" : "2016-04",
  "value" : {
    "measure" : 1406083,
    "count" : 38040,
    "avg" : 36.963275499474236
  }
}

  "_id" : "2016-05",
  "value" : {
    "measure" : 1432478,
    "count" : 39358,
    "avg" : 36.39610752578891
  }
}

  "_id" : "2016-06",
  "value" : {
    "measure" : 1458693,
    "count" : 36720,
    "avg" : 39.724754901960786
  }
}

```

```

    "_id" : "2016-07",
    "value" : {
      "measure" : 1687609,
      "count" : 37871,
      "avg" : 44.562039555332575
    }
  }

  "_id" : "2016-08",
  "value" : {
    "measure" : 1722392,
    "count" : 37200,
    "avg" : 46.30086021505377
  }
}

Type "it" for more
> it
{
  "_id" : "2016-09",
  "value" : {
    "measure" : 1574995,
    "count" : 36000,
    "avg" : 43.74986111111111
  }
}

  "_id" : "2016-10",
  "value" : {
    "measure" : 1427998,
    "count" : 37200,
    "avg" : 38.38704301075269
  }
}

  "_id" : "2016-11",
  "value" : {
    "measure" : 1430048,
    "count" : 36000,
    "avg" : 39.723555555555556
  }
}

  "_id" : "2016-12",
  "value" : {
    "measure" : 1439270,
    "count" : 37200,
    "avg" : 38.69005376344086
  }
}

```