# Master en Ciencia de Datos e Ingeniería de Computadores

## Almacenamiento de datos masivos para procesamiento y análisis: NoSQL (MongoDB)

#### Cargar en vuestra base de datos de la cuenta MongoDB disponible en Hadoop.ugr.es el archivo /var/tmp/restaurantes1.json conforme se indica en la transparencia 43 de la presentación sobre NoSQL. Elaborar el código MapReduce que resuelva la consulta:

#### "Obtener el par de restaurantes más próximos en cada barrio (borough), mostrando el nombre y tipo de cocina de esos restaurantes, la distancia entre ellos y la cantidad de restaurantes evaluados para cada barrio, para aquellos restaurantes que hayan tenido un grado= ‘A’ en alguna ocasión"

#### El resultado del ejercicio será un archivo con el código que resuelve la consulta (según el formato mostrado en la transparencia 17 del tutorial) y que podrá ser ejecutado directamente sobre su base de datos MongoDB de hadoop.ugr.es, la cual que debe incluir la colección restaurants importada y la colección rest\_resultado con los documentos resultado de la consulta propuesta.

La consulta final quedaría:

db.runCommand({

mapReduce: "restaurants",

map : function Map() {

var key = this.borough;

emit(key, {

"data":

[

{

"name" : this.name,

"lat" : this.address.coord[0],

"lon" : this.address.coord[1],

"cuisine" : this.cuisine

,"grades" : this.grades

}

]

});

},

Donde en esta etapa vamos a filtrar por barrios y nos quedaremos con los datos que nos interesan como el nombre, la latitud, longitud y el tipo de concina del restaurante. Si quisiéramos también ver los grados deberíamos añadir la línea en amarillo para poder ver los certificados que se les dio a cada restaurante.

reduce : function Reduce(key, values) {

var reduced = {"data":[]};

for (var i in values) {

var inter = values[i];

for (var j in inter.data){

reduced.data.push(inter.data[j]);

}

}

return reduced;

},

Donde en esta etapa agregamos los diferentes valores para cada clave dada.

finalize : function Finalize(key, reduced){

if (reduced.data.length == 1) {

return { "message" : "Este barrio solo tiene un restaurante con categoría 'A' " };

}

var min\_dist = 999999999999;

var restaurante1 = { "name": "" , "cuisine": "" , "grades": ""};

var restaurante2 = { "name": "" , "cuisine": "" , "grades": ""};

var c1;

var c2;

var d;

var evaluados = reduced.data.length;

for (var i in reduced.data) {

for (var j in reduced.data) {

if (i>=j) continue; //termina la iteración actual y continua con la siguiente j

c1 = reduced.data[i];

c2 = reduced.data[j];

d = (c1.lat-c2.lat)\*(c1.lat-c2.lat)+(c1.lon-c2.lon)\*(c1.lon-c2.lon);

//Actualizo la minima distantica

if (d < min\_dist && d > 0) {

min\_dist = d; restaurante1 = c1; restaurante2 = c2;

}

}

}

return {"restaurante1": restaurante1.name, "cuisine1": restaurante1.cuisine, "grades1": restaurante1.grades, "restaurante2": restaurante2.name, "cuisine2": restaurante2.cuisine, "grades2":restaurante2.grades, "dist": Math.sqrt(min\_dist), "evaluados":evaluados};

},

Donde en Finalize encontraremos los dos restaurantes más próximos en cada barrio. Si queremos mostrar los grados en el resultado final añadiremos las zonas en amarillo.

query : { "grades.grade" : {"$eq" : 'A'}},

out : { merge: "rest\_resultado"}

});

En query vamos a filtrar por restaurantes que han tenido al menos una vez un certificado de grado ‘A’.

El resultado sin los grados sería:

db.rest\_resultado.find().pretty()

{

"\_id" : "Bronx",

"value" : {

"restaurante1" : "Subway",

"cuisine1" : "American ",

"restaurante2" : "Chen'S Foo Hai",

"cuisine2" : "Chinese",

"dist" : 5.656854265355972e-7,

"evaluados" : 2168

}

}

{

"\_id" : "Brooklyn",

"value" : {

"restaurante1" : "V J Bakery",

"cuisine1" : "Delicatessen",

"restaurante2" : "Dunkin' Donuts",

"cuisine2" : "Donuts",

"dist" : 0.000003862641583420379,

"evaluados" : 5610

}

}

{

"\_id" : "Manhattan",

"value" : {

"restaurante1" : "Rockwood Music Hall (Stage Zero)",

"cuisine1" : "American ",

"restaurante2" : "Cheese Grille",

"cuisine2" : "Sandwiches/Salads/Mixed Buffet",

"dist" : 0.000010016985567509196,

"evaluados" : 9579

}

}

{

"\_id" : "Missing",

"value" : {

"restaurante1" : "Lima 33",

"cuisine1" : "Peruvian",

"restaurante2" : "Little Neck Shanghai & Taiwanese Cuisine",

"cuisine2" : "Asian",

"dist" : 0.0011684476282611152,

"evaluados" : 50

}

}

{

"\_id" : "Queens",

"value" : {

"restaurante1" : "New Haw Hing Garden",

"cuisine1" : "Chinese",

"restaurante2" : "Domino'S Pizza",

"cuisine2" : "Pizza",

"dist" : 9.999997985232767e-8,

"evaluados" : 5157

}

}

{

"\_id" : "Staten Island",

"value" : {

"restaurante1" : "Egger'S Ice Cream Parlor",

"cuisine1" : "Ice Cream, Gelato, Yogurt, Ices",

"restaurante2" : "Call It A Wrap",

"cuisine2" : "American ",

"dist" : 0.000007071067808080774,

"evaluados" : 876

}

}

Donde podemos ver el nombre del barrio y para cada barrio los dos restaurantes, donde apreciamos su nombre, el tipo de concia, la distancia entre ellos y el número de restaurantes evaluados.

Si queremos ver los grados de estos restaurantes:

{

"\_id" : "Bronx",

"value" : {

"restaurante1" : "Subway",

"cuisine1" : "American ",

"grades1" : [

{

"date" : ISODate("2014-12-01T00:00:00Z"),

"grade" : "A",

"score" : 5

},

{

"date" : ISODate("2013-12-19T00:00:00Z"),

"grade" : "A",

"score" : 5

},

{

"date" : ISODate("2012-12-17T00:00:00Z"),

"grade" : "A",

"score" : 12

}

],

"restaurante2" : "Chen'S Foo Hai",

"cuisine2" : "Chinese",

"grades2" : [

{

"date" : ISODate("2014-07-15T00:00:00Z"),

"grade" : "A",

"score" : 7

},

{

"date" : ISODate("2013-07-30T00:00:00Z"),

"grade" : "A",

"score" : 11

},

{

"date" : ISODate("2012-12-17T00:00:00Z"),

"grade" : "A",

"score" : 12

},

{

"date" : ISODate("2012-06-11T00:00:00Z"),

"grade" : "A",

"score" : 10

},

{

"date" : ISODate("2011-12-28T00:00:00Z"),

"grade" : "A",

"score" : 13

}

],

"dist" : 5.656854265355972e-7,

"evaluados" : 2168

}

}

{

"\_id" : "Brooklyn",

"value" : {

"restaurante1" : "V J Bakery",

"cuisine1" : "Delicatessen",

"grades1" : [

{

"date" : ISODate("2014-10-20T00:00:00Z"),

"grade" : "B",

"score" : 15

},

{

"date" : ISODate("2014-06-17T00:00:00Z"),

"grade" : "C",

"score" : 21

},

{

"date" : ISODate("2013-12-31T00:00:00Z"),

"grade" : "A",

"score" : 9

},

{

"date" : ISODate("2012-12-04T00:00:00Z"),

"grade" : "A",

"score" : 10

},

{

"date" : ISODate("2011-11-22T00:00:00Z"),

"grade" : "A",

"score" : 12

}

],

"restaurante2" : "Dunkin' Donuts",

"cuisine2" : "Donuts",

"grades2" : [

{

"date" : ISODate("2015-01-06T00:00:00Z"),

"grade" : "A",

"score" : 9

},

{

"date" : ISODate("2014-01-21T00:00:00Z"),

"grade" : "A",

"score" : 6

},

{

"date" : ISODate("2013-01-28T00:00:00Z"),

"grade" : "A",

"score" : 8

},

{

"date" : ISODate("2012-02-01T00:00:00Z"),

"grade" : "A",

"score" : 2

}

],

"dist" : 0.000003862641583420379,

"evaluados" : 5610

}

}

{

"\_id" : "Manhattan",

"value" : {

"restaurante1" : "Rockwood Music Hall (Stage Zero)",

"cuisine1" : "American ",

"grades1" : [

{

"date" : ISODate("2014-04-29T00:00:00Z"),

"grade" : "A",

"score" : 12

},

{

"date" : ISODate("2013-04-12T00:00:00Z"),

"grade" : "A",

"score" : 7

},

{

"date" : ISODate("2012-03-02T00:00:00Z"),

"grade" : "A",

"score" : 12

}

],

"restaurante2" : "Cheese Grille",

"cuisine2" : "Sandwiches/Salads/Mixed Buffet",

"grades2" : [

{

"date" : ISODate("2014-10-31T00:00:00Z"),

"grade" : "A",

"score" : 7

}

],

"dist" : 0.000010016985567509196,

"evaluados" : 9579

}

}

{

"\_id" : "Missing",

"value" : {

"restaurante1" : "Little Neck Shanghai & Taiwanese Cuisine",

"cuisine1" : "Asian",

"grades1" : [

{

"date" : ISODate("2014-05-20T00:00:00Z"),

"grade" : "A",

"score" : 8

}

],

"restaurante2" : "Lima 33",

"cuisine2" : "Peruvian",

"grades2" : [

{

"date" : ISODate("2014-09-10T00:00:00Z"),

"grade" : "A",

"score" : 5

},

{

"date" : ISODate("2014-04-02T00:00:00Z"),

"grade" : "A",

"score" : 10

}

],

"dist" : 0.0011684476282611152,

"evaluados" : 50

}

}

{

"\_id" : "Queens",

"value" : {

"restaurante1" : "New Haw Hing Garden",

"cuisine1" : "Chinese",

"grades1" : [

{

"date" : ISODate("2014-09-15T00:00:00Z"),

"grade" : "A",

"score" : 7

},

{

"date" : ISODate("2014-03-27T00:00:00Z"),

"grade" : "A",

"score" : 11

},

{

"date" : ISODate("2013-02-22T00:00:00Z"),

"grade" : "A",

"score" : 3

},

{

"date" : ISODate("2012-06-29T00:00:00Z"),

"grade" : "B",

"score" : 21

},

{

"date" : ISODate("2012-01-24T00:00:00Z"),

"grade" : "B",

"score" : 27

}

],

"restaurante2" : "Domino'S Pizza",

"cuisine2" : "Pizza",

"grades2" : [

{

"date" : ISODate("2014-08-07T00:00:00Z"),

"grade" : "A",

"score" : 11

},

{

"date" : ISODate("2014-01-24T00:00:00Z"),

"grade" : "C",

"score" : 30

},

{

"date" : ISODate("2013-07-09T00:00:00Z"),

"grade" : "C",

"score" : 39

},

{

"date" : ISODate("2013-01-28T00:00:00Z"),

"grade" : "A",

"score" : 13

},

{

"date" : ISODate("2012-08-13T00:00:00Z"),

"grade" : "A",

"score" : 13

},

{

"date" : ISODate("2012-02-01T00:00:00Z"),

"grade" : "C",

"score" : 43

}

],

"dist" : 9.999997985232767e-8,

"evaluados" : 5157

}

}

{

"\_id" : "Staten Island",

"value" : {

"restaurante1" : "Egger'S Ice Cream Parlor",

"cuisine1" : "Ice Cream, Gelato, Yogurt, Ices",

"grades1" : [

{

"date" : ISODate("2014-11-24T00:00:00Z"),

"grade" : "A",

"score" : 13

},

{

"date" : ISODate("2014-05-05T00:00:00Z"),

"grade" : "A",

"score" : 8

},

{

"date" : ISODate("2013-03-14T00:00:00Z"),

"grade" : "A",

"score" : 3

},

{

"date" : ISODate("2012-01-28T00:00:00Z"),

"grade" : "A",

"score" : 9

}

],

"restaurante2" : "Call It A Wrap",

"cuisine2" : "American ",

"grades2" : [

{

"date" : ISODate("2014-12-23T00:00:00Z"),

"grade" : "A",

"score" : 11

},

{

"date" : ISODate("2014-06-26T00:00:00Z"),

"grade" : "A",

"score" : 5

},

{

"date" : ISODate("2013-12-04T00:00:00Z"),

"grade" : "A",

"score" : 9

},

{

"date" : ISODate("2012-10-23T00:00:00Z"),

"grade" : "A",

"score" : 10

}

],

"dist" : 0.000007071067808080774,

"evaluados" : 876

}

}