

Part I: Query Mechanisms in MongoDB

Convert the following SQL queries to JavaScript (the main query language in MongoDB and used in the shell client) and describe what is returned by each query:

1. SELECT * FROM world_bank WHERE regionname = "Africa"

```
db.world_bank.find({ "regionname" : "Africa" });
```

The above query can be run and gives the result from world_bank collection where regionname is "Africa"

This will give the result in batch. After each batch mongo shell asks to type "it" to see more.

By default mongo shell prints only first 20 documents, we can get the next batch by typing "it" in the shell. (ignore double quotes)

This can be changed by DBQuery.shellBatchSize = preferable number

2. SELECT COUNT(*) FROM world_bank WHERE regionname = "South Asia"

```
db.world_bank.find( { "regionname" : "South Asia" }).count();
```

```
> 65
```

This query gives the total number of South Asia region.

3. SELECT project_name FROM world_bank WHERE lendprojectcost > 1000000

```
db.world_bank.find({ "lendprojectcost" : { "$gt" : 1000000 } , { "project_name" : 1, -id : 0 } });
```

OR

```
db.world_bank.aggregate([ { $group: { _id: "$regionname", sumQuantity: { $sum: "$lendprojectcost" } } }, { $sort: { sumQuantity: -1 } }, { $limit: 1 } ])
```

By running any of these queries, we will get all the project names whose lendprojectcost is greater than 1000000.

Part II: Querying the World Bank Collection [3 points]

4. Which are the three countries in the MENA region ("Middle East and North Africa") with the most projects?

```
db.world_bank.find({ "$and" : [ { "regionname" : "Middle East and North Africa" }, { "projectstatusdisplay" : "Active" } ], { "countryshortname" : 1, "_id" : 0 } });
```

yemen, Republic of - 5

Morocco - 5

Jordan - 3

OR

```
db.world_bank.aggregate([ { "$match": { "regionname": "Middle East and North Africa" } }, { "$group": { "_id": "$countryshortname", "count": { "$sum": 1 } } }, { "$sort": { "count": -1 } }, { "$limit": 3 } ])
```

5. Which region has the largest total amount of lend project cost?

```
db.world_bank.find({}, { "lendprojectcost" : 1, "regionname" : 1, "-id" :
0}).sort({"lendprojectcost" : -1 }).limit(1);
South Asia – 5170000000
```

Another Query :

```
db.world_bank.aggregate({ $group : { _id: null, max: { $max : "$lendprojectcost" }}});
```

In this query we will get the number but not the name of the region.

- Which active projects have a lend project cost of over 1 billion USD?

```
db.world_bank.find({ "$and" : [{"lendprojectcost" : { "$gt" : 1000000000 }},
{"projectstatusdisplay" : "Active" }]}, {"project_name" :1, _id : 0});
```

There are 14 projects which has lend project cost more than 1 billion USD.

```
> db.world_bank.find({"$and" : [{"lendprojectcost" : { "$gt" : 1000000000 }}, {"projectstatusdisplay" : "Active"}]}, {"project
{ "project_name" : "National Highways Interconnectivity Improvement Project" }
{ "project_name" : "Quito Metro Line One" }
{ "project_name" : "EG - Helwan South Power Project" }
{ "project_name" : "Bangladesh Safety Net Systems for the Poorest Project" }
{ "project_name" : "China: Nanchang Urban Rail Project" }
{ "project_name" : "Punjab Health Sector Reform Project" }
{ "project_name" : "Nepal: School Sector Reform Program Additional Financing" }
{ "project_name" : "Kyrgyz Second Health and Social Protection Project" }
{ "project_name" : "Sri Lanka - Second Health Sector Development Project" }
{ "project_name" : "Nigeria - State Education Program Investment Project" }
{ "project_name" : "Pakistan: Second Sindh Education Sector Project" }
{ "project_name" : "NATIONAL PROGRAM FOR COMMUNITY EMPOWERMENT IN RURAL AREA (PNPM RURAL 2012-2015" }
{ "project_name" : "Road Rehabilitation and Maintenance Program" }
{ "project_name" : "Ethiopia Promoting Basic Services Program Phase III Project" }
>
```

Part III: The NoSQL Paradigm

- How many collections are there in the mdb database and how many documents (objects) does each collection contain?

By using the query below we found that there are 500 documents(objects) in 1 collection that is world_bank.json

```
db.stats();
: {
  "db" : "mdb",
  "collections" : 1,
  "views" : 0,
  "objects" : 500,
  "avgObjSize" : 5742.092,
  "dataSize" : 2871046,
  "storageSize" : 1327104,
  "numExtents" : 0,
  "indexes" : 1,
  "indexSize" : 16384,
  "ok" : 1
}
```

Collections : 1
Objects : 500

SQL Databases are relational while NoSQL databases are non-relational. SQL databases are table based but as we see in world_bank collection there are no tables, instead there are documents.

8. If the World Bank wants to start recording more information about its projects, how would the existing database (w.r.t. to its schema) need to be redesigned? Motivate your answer.

As we know that World Bank database is Document Stores database. We don't need to redesign the database as it is schema free. We can say it has flexible schema. In this type of database we can store any data under documents with no structural restrictions.

It allows for easily changing data storage as we learn more about the project like here. We can easily add new things.

In this particular data scaling out can be easily done. Aggregate orientation fits well with scaling out which means we can easily add more data horizontally.

So in other words we can say it's more convenient to save more information in NoSQL without redesigning it.

Submitted By Group 27-

Name : Augustin Ionut Ionascu, Muthupriya Shankaran & Suchi Siwach