École Nationale de la Statistique et de l'Analyse de l'Information

Technologies NoSQL

Projet NoSQL

Élève :
Théo Lavolé

Professeur :
Julien Balas

11 février 2017

Introduction

Tout d'abord, j'ai choisi de travailler sous MongoDB car j'ai commencé le projet après le tp mongo. L'avantage de cette base de données est qu'elle n'utilise pas les transactions et repose sur des réplications de maître-esclave. Toutes les requêtes sont soumises au maître et sont répliquées aux esclaves. En cas de crash du maître, un nouveau maître est élu parmi les esclaves : il y a donc une consistence de l'information présente dans la base de données. J'utilise la VM fournie pour le tp MongoDB et RoboMongo pour effectuer des requêtes et insérer mes données.

Vient ensuite ma définition d'un bon endroit où vivre. Le premier facteur qui m'intéresse est la **sécurité**, notamment dans les lieux publics dans lesquels je peux me balader. Pour rester dans la même lignée, et comme j'aime toujours me balader, je souhaite résider dans un quartier **historique**, ou du moins possédant de nombreux monuments que je puisse observer. Ensuite, je veux un quartier vivant, cela repose à la fois sur la présence de **marchés** et sur des **activités culturelles**. Enfin, je cherche un quartier avec des **hôpitaux** à proximité !

Une fois que j'aurais défini le quartier dans lequel je souhaite vivre, il m'appartient de choisir la rue. Pour cela, je m'intéresse uniquement aux **évènements** organisés par la ville dans ce quartier. C'est un peu la continuité de ce que je cherchais comme quartier : je veux quelque chose de vivant mais aussi avec des commodités et de la sécurité, qui seront supposées vraies à partir du choix du quartier.

Les scripts de mes données sont disponibles dans le dépôt Github **https://github.com/TheoLavole/NoSQL**. Ce sont tous les fichiers **import_*.js** qu'il faut importer dans un RoboMongo. Les requêtes sont aussi disponibles dans le fichier **ordres_exam.js** (il est d'ailleurs conseillé de les utiliser en copiant les requêtes du fichier plutôt que celles de ce rapport, des espaces sauvages apparaissent sans que je puisse faire quoi que ce soit).

Allons vivre à New-York

La première étape consiste à importer les données présentes dans **import_park_crimes.js**. J'ai ensuite trié les quartiers en fonction du nombre d'incidents dans les parcs avec la requête suivante :

```
use park_crimes
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$total'}}},
    {$sort :{total:1}}
])
```

On obtient les résultats suivants : Manhattan 59, Brooklyn 52, Bronx 42, Queens 29, Staten Island 3, Brooklyn/Queens 2.Ca signifie que les deux quartiers où il y a le plus d'accidents sont Manhattan et Brooklyn. Mais y a-t'il une différence si l'on se renseigne sur le détail des crimes ?

```
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$murder'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$rape'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$robbery'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$felony_assault'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$burglary'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',total:{$sum:'$grand_larceny'}}}
])
db.park_crimes.aggregate([
    {$group:{_id:'$borough',
    total:{$sum:'$grand_larceny_of_motor_vehicle'}}}
])
```

On optient peu d'informations de ces requêtes. Intéressons nous alors au nombre de monuments dans les parcs par quartier.

```
use park_monuments
db.park_monuments.aggregate([
    {$group:{_id:'$borough',total:{$sum:1}}},{$sort :{total:-1}}
])
```

Manhattan se détache du lot. En effet, Manhattan seul détient 946 monuments, soit plus de la moitié des monuments présents dans l'ensemble des parcs (Brooklyn 349, Queens 249, Bronx 222, Staten Island 96, inconnu 7).

Ici, seul le suartier de Manhattan se détache un peu du lot, mais revenons-en aux marchés de producteurs !

```
use farmers_market
db.farmers_market.mapReduce(
    function(){
        emit(this.borough, 1);
    },
    function(key,values){
        return Array.sum(values)
    },
    {out:{inline:1}}
)
```

Et voila, Brooklyn est en tête avec le plus grand nombre de marchés (48), suivi de Manhattan (39) puis du Bronx (32). Continuons nos recherches en s'intéressant aux infrastructures culturelles.

```
use cultural
db.cultural.aggregate([
    {$group:
        _id:{borough:'$borough',discipline:'$discipline'},
        total:{$sum:1}
    }},
    {$sort :{total:-1}}
])
```

Ici le résultat est sans appel, Manhattan fait la course seul en tête (277 théatres, 206 espaces de musique, 136 espaces de danse, 58 musées...) alors que les principales activités de Brooklyn sont des complexes multi-activités (64) puis des théatres (54). Ca sent plutôt bon pour Manhattan, mais regardons maintenant les hôpitaux par quartiers.

```
use hospital_health
db.hospital_health.aggregate([
    {$group:{_id:'$borough',total:{$sum:1}}},{$sort :{total:-1}}
])
```

Pas de franche différence entre Brooklyn (26 hôpitaux) et Manhattan (24). Néanmoins, on peut à partir de maintenant décider de ne s'intéresser qu'à Brooklyn et Manhattan. Dans cette optique, quel est le loyer moyen dans ces deux quartiers ?

```
use development
db.development.aggregate([
    {$match:{$or:[{borough:'MANHATTAN'},{borough:'BROOKLYN'}]}},
    {$group:{_id:'$borough',moyenne:{$avg:'$avg_rent'}}},
    {$sort :{total:-1}}
])
```

Manhattan est un peu moins cher que Brooklyn ($482.9027 contre $483.5938). Quel est le quartier le plus peuplé ?

```
db.development.aggregate([
    {$match:{$or:[{borough:'MANHATTAN'},{borough:'BROOKLYN'}]}},
    {$group:{_id:'$borough',total:{$sum:'$total_pop'}}},
    {$sort :{total:-1}}
])
```

Brooklyn est plus peuplé que Manhattan, alors que les loyers sont plus chers. Et quel est le quartier avec le plus d'appartements ?

```
db.development.aggregate([
    {$match:{$or:[{borough:'MANHATTAN'},{borough:'BROOKLYN'}]}},
    {$group:{_id:'$borough',total:{$sum:'$total_number_apartments'}}},
    {$sort :{total:-1}}
])
```

Brooklyn propose plus d'appartements que Manhattan (51745 contre 50303). Le choix ici parait évident et je choisi d'aller vivre à Manhattan. Mais dans ce cas, où aller vivre ? Pour faire ce dernier choix, je choisi d'aller voir les évènements organisés dans les différentes rues de New-York, en me concentrant sur Manhattan.

```
use events
db.events.aggregate([
    {$match:{borough:'Manhattan'}},
    {$group:{_id:{address:'$address'},total:{$sum:1}}},
    {$sort :{total:-1}}
])
```

Cette dernière requête nous donne les différentes adresses auxquelles ont lieu le plus d'évènements culturels. Le premier résultat retourné est une chaîne de caractères vide, car l'adresse n'était pas renseignée. La deuxième plus importante est situé au 250 W. 127th Street. On peut donc se renseigner pour vivre dans une des rues retournées par cette dernière requête.

Pour conclure, on va dire que je souhaite vivre à **Manhattan**, dans la **127ème rue**, **William Street** ou **Grand Street** par exemple.

Contenu du Github

Le Github **https://github.com/TheoLavole/NoSQL** doit contenir 11 fichiers :

import_cultural.js	importe des données
import_development.js	importe des données
import_events.js	importe des données
import_farmers_market.js	importe des données
import_hospitals_and_health.js	importe des données
import_park_crimes.js	importe des données
import_park_monuments.js	importe des données
liens_data.txt	liens vers les données utilisées sur le site de NYC
ordres_exam.js	ordres mongo soumis et à copier
projet.pdf	pdf de ce qui est attendu
rapport.pdf	ce rapport