

TP MongoDB

Ce TP a pour objectif de vous familiariser avec MongoDB et son shell. Vous allez dans la suite charger un jeu de données et écrire les requêtes permettant chacune de répondre à un besoin en information précis.

Introduction

MongoDB est une base de données NoSQL orientée documents. Une base de données MongoDB consiste en un ensemble de collections, elles-mêmes formées de documents, construits selon un schéma dynamique et non prédéfini.

Un document est un ensemble ordonné de paires clé-valeurs, où les clés sont des chaînes de caractères et les valeurs peuvent être une instance de n'importe quel type de donnée parmi ceux prédéfinis (*null*, *boolean*, *number*, *string*, *date*, *regular expression*, *array*, *object id*, *binary data* et *code*), ou bien un autre document.

Préparation

1. Télécharger, installer et lancer un serveur MongoDB.
 - <https://www.mongodb.com/download-center#community>
 - <https://docs.mongodb.com/manual/installation/#mongodb-community-edition>
2. Télécharger les fichiers suivants :
 - [movielens_movies.json](#)
 - [movielens_users.json](#)
3. Importer ces deux fichiers dans MongoDB :

```
# $MONGO_HOME/bin/mongoimport --db MovieLens --collection movies --file
movielens_movies.json
connected to: localhost
imported 3883 objects

# $MONGO_HOME/bin/mongoimport --db MovieLens --collection users --file
movielens_users.json
connected to: localhost
[...]
imported 6040 objects
```

Ces commandes permettent de créer les collections movies et users dans la base de données MovieLens et de les peupler avec les données issues des fichiers json.

4. Connectez-vous à la base de données MovieLens :

```
# $MONGO_HOME/bin/mongo MovieLens
MongoDB shell version: 3.2.10
connecting to: MovieLens
>
```

Shell MongoDB

Cette dernière commande lance le shell MongoDB qui est en fait un interpréteur javascript complet. Vous pouvez donc y exécuter n'importe quel code javascript. Le shell propose en plus un certain

nombre de commandes spécifiques courantes dans les interfaces avec une base de données, par exemple :

- `use <db name>` permet de connecter la session à la base de donnée *db name*.
- `show collections` affiche les collections de la base de données courante.
- `help` donne un aperçu des commandes les plus importantes et de leur usage.

Schéma de la base de données

Comme dit plus haut, les documents d'une collection ne sont pas soumis à un schéma fixe. Cependant, les documents de chaque collection ont une structure similaire. Nous donnons dans cette partie un exemple de chaque collection.

La collection `movies` contient des informations sur les films, c'est-à-dire leur id, titre et genre.

```
> db.movies.findOne()  
{  
  "_id" : 1,  
  "title" : "Toy Story (1995)",  
  "genres" : "Animation|Children's|Comedy"  
}
```

La collection `users` contient des informations portant sur les utilisateurs et les notes qu'ils ont données aux films. Parmi les informations sur les utilisateurs, on trouve leur id, nom, âge, occupation et sexe. Les notes données par chaque utilisateur sont représentées dans un tableau de document, chaque document contenant l'id d'un film (faisant référence aux id de la collection `movies`), la note attribuée et la date à laquelle l'utilisateur a laissé la note.

```
> db.users.findOne({}, {movies : {$slice : 2}});  
{  
  "_id" : 6038,  
  "name" : "Yaeko Hassan",  
  "gender" : "F",  
  "age" : 95,  
  "occupation" : "academic/educator",  
  "movies" : [  
    {  
      "movieid" : 1419,  
      "rating" : 4,  
      "timestamp" : 956714815  
    },  
    {  
      "movieid" : 920,  
      "rating" : 3,  
      "timestamp" : 956706827  
    }  
  ]  
}
```

Requêtes simples

Question 1. Combien y a-t-il d'utilisateurs dans la base de données ?

- <https://docs.mongodb.com/manual/reference/command/count/>

Question 2. Combien y a-t-il de films dans la base de données ?

Question 3. Quelle est l'occupation de Clifford Johnathan ? Ecrivez une requête dont la réponse affiche uniquement son nom et son occupation.

- <http://docs.mongodb.org/manual/reference/method/db.collection.find/>

Question 4. Combien d'utilisateurs ont entre 18 et 30 ans (inclus) ?

Question 5. Combien d'utilisateurs sont artistes (*artist*) ou scientifiques (*scientist*) ?

Question 6. Quelles sont les dix femmes auteurs (*writer*) les plus âgées ?

Question 7. Quelles sont toutes les occupations présentes dans la base de données ?

Insertions, mises-à-jour et suppressions

Question 8. Insérer un nouvel utilisateur dans la base de données (vous, par exemple).

Question 9. Supprimer l'entrée de la base de données.

Question 10. Pour tous les utilisateurs qui ont pour occupation "programmer", changer cette occupation en "developer".

Question 11. Les genres du film "Cinderella" devraient être Animation, Children's et Musical. Modifier en une seule requête le document correspondant pour qu'il contienne ces trois genres sans doublon.

Expressions régulières

- <http://docs.mongodb.org/manual/reference/operator/query/regex/>

Question 12. Combien de films sont sortis dans les années quatre-vingt ? (l'année de sortie est indiquée entre parenthèses à la fin du titre de chaque film)

Question 13. Combien y a-t-il de films d'horreur ?

Question 14. Combien de films ont pour type à la fois "Musical" et "Romance" ?

Requêtes sur des tableaux

Question 15. Combien d'utilisateurs ont noté le film qui a pour id 1196 (Star Wars: Episode V - The Empire Strikes Back (1980)) ?

Question 16. Combien d'utilisateurs ont noté tous les films de la première trilogie Star Wars (id 260, 1196, 1210) ?

- <http://docs.mongodb.org/manual/reference/operator/query/all/>

Question 17. Combien d'utilisateurs ont notés exactement 48 films ?

- <http://docs.mongodb.org/manual/reference/operator/query/size/>

Question 18. Pour chaque utilisateur, créer un champ num_ratings qui indique le nombre de films

qu'il a notés.

Question 19. Combien d'utilisateurs ont noté plus de 90 films ?

Question 20. Quels sont les trois derniers films notés par Jayson Brad ?

Agrégats

Question 21. Montrer combien de films ont été produits durant chaque année des années 90 ; ordonner les résultats de l'année la plus à la moins fructueuse.

- <http://docs.mongodb.org/manual/core/aggregation-pipeline>
- <http://docs.mongodb.org/manual/reference/operator/aggregation-pipeline>

Question 22. Quelle est la note moyenne du film Pulp Fiction, qui a pour id 296 ?

Question 23. En une seule requête, retourner pour chaque utilisateur son id, son nom, les notes maximale, minimale et moyenne qu'il a données, et ordonner le résultat par note moyenne croissante.

Question 24. Quel est le genre le plus populaire en termes de nombre de notes ?

Question 25. Quel est le genre le mieux noté (celui dont la moyenne de toutes les notes est la plus élevée) ?

