

Travaux pratiques MongoDB

Dans ce TP, vous allez prendre en main le langage NoSQL MongoDB et aller de l'installation jusqu'à la réalisation des commandes CRUD de base sur le langage.

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.

Mongodb Shell

```
$ mongosh
```

Cette 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 données 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

1. Installation de MongoDB

En fonction de votre environnement de travail (Windows, Mac, linux), suivez la documentation sur le site de MongoDB pour installer mongodb-community server sur votre PC ainsi que l'outil en ligne de commande Mongosh et l'application GUI de visualisation de vos bases Compass.

<https://www.mongodb.com/docs/manual/installation/>

Lancez votre service mongodb-community (ex sur mac : `$ brew services start mongodb-community`)

Puis accédez à l'utilitaire en ligne de commande (ex sur mac : `$ mongosh`)

2. Prise en main de MongoDB

Récupérez sur Boostcamp, dans la page du TP, le fichier zip : `dblp.json.zip`

1. Décompresser le fichier `dblp.json.zip`
2. Dans le même répertoire, lancer l'importation du fichier :
3. `$ mongoimport --host localhost:27017 --db DBLP --collection publis --jsonArray --type json --file dblp.json`

Dans la console mongo vérifier que les données ont été insérées : `db.publis.count()` ;
Il devrait y avoir d'affiché ~118000 items d'inséré.

Exercice #1 :

Interrogations simples sur la base DBPL

Exprimez des requêtes simples (pas de MapReduce) pour les recherches suivantes :

1. Liste de tous les livres (type « Book ») ;
2. Liste des publications depuis 2011 ;
3. Liste des livres depuis 2014 ;
4. Liste des publications de l'auteur « Toru Ishida » ;
5. Liste de tous les éditeurs (type « publisher »), distincts ;
6. Liste de tous les auteurs distincts ;
7. Trier les publications de « Toru Ishida » par titre de livre et par page de début ;
8. Projeter le résultat sur le titre de la publication, et les pages ;
9. Compter le nombre de ses publications ;
10. Compter le nombre de publications depuis 2011 et par type ;
11. Compter le nombre de publications par auteur et trier le résultat par ordre croissant ;

Pour vous aider, appuyez-vous sur toutes les ressources en ligne que vous pouvez trouver, notamment la doc MongoDB disponible sur : <https://www.mongodb.com/docs/manual/crud/>

EXERCICE #2

Pour l'exercice 2, importez les deux fichiers dans MongoDB : movielens_movies.json et movielens_users.json présent dans l'archive zip du TP :

- `$ mongoimport --db MovieLens --collection movies --file movielens_movies.json`
- `$ mongoimport --db MovieLens --collection users --file movielens_users.json`

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

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.

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 ? Écrivez 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). Ne pas inclure pour l'instant le champ movies.

Question 9. Choisir un film de la collection movies et mettre à jour l'entrée insérée précédemment en ajoutant le champ movies respectant le schéma adopté par les autres entrées. Pour le champ timestamp, utiliser l'heure courante :

```
Math.round(new Date().getTime() / 1000)
```

Question 10. Supprimer l'entrée de la base de données.

Question 11. Pour tous les utilisateurs qui ont pour occupation "programmer", changer cette occupation en "developer".

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 de films sont sortis entre 1984 et 1992 ?

Question 14. Combien y a-t-il de films d'horreur ?

Question 15. Combien de films ont pour type à la fois "Musical" et "Romance" ?

ForEach

Question 16. Comme vous avez pu le constater, stocker l'année de sortie du film dans son titre n'est pas très pratique. Modifier la collection movies en ajoutant à chaque film un champ year contenant l'année et en supprimant cette information du titre. Ne nombreuses méthodes peuvent répondre à ce besoin ; privilégier au maximum les approches exploitant les fonctionnalités de MongoDB (il est par exemple déconseillé, pour des raisons évidentes de performances, de demander l'intégralité des films à la base de données, de les stocker dans une liste javascript, puis d'itérer sur cette liste pour calculer les nouvelles valeurs de champs et mettre à jour les éléments, toujours en javascript).

<http://docs.mongodb.org/manual/reference/method/cursor.forEach/>

<http://docs.mongodb.org/manual/reference/method/cursor.snapshot/>

Question 17. Modifier la collection movies en remplaçant pour chaque film la valeur du champ genres par un tableau de chaînes de caractères.

Requêtes sur des tableaux

Lecture

Question 18. Combien d'utilisateurs ont noté le film qui a pour id 1196 (Star Wars: Episode V – The Empire Strikes Back (1980)) ?

Question 19. 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 20. Combien d'utilisateurs ont notés exactement 48 films ?

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

Notez que `$size` ne peut être apparié qu'à des nombres exacts. La sélection des utilisateurs qui ont vu plus d'un certain nombre de films doit être effectuée en deux étapes ; c'est le sujet des questions suivantes.

Question 21. Pour chaque utilisateur, créer un champ `num_ratings` qui indique le nombre de films qu'il a notés.

Question 22. Combien d'utilisateurs ont noté plus de 90 films ?

Question 23. Combien de notes ont été soumises après le 1er janvier 2001 ?

Question 24. Quels sont les trois derniers films notés par Jayson Brad ?

Question 25. Obtenez les informations portant uniquement sur Tracy Edward et sa note du film Star Wars: Episode VI - Return of the Jedi, qui a pour id 1210.

Question 26. Combien d'utilisateurs ont donné au film "Untouchables, The" la note de 5.

Ecriture

Question 27. L'utilisateur Barry Erin vient juste de voir le film Nixon, qui a pour id 14 ; il lui attribue la note de 4. Mettre à jour la base de données pour prendre en compte cette note. N'oubliez pas que le champ `num_ratings` doit représenter le nombre de films notés par un utilisateur.

Question 28. L'utilisatrice Marquis Billie n'a en fait pas vu le film "Santa with Muscles", qui a pour id 1311. Supprimer la note entrée par mégarde dans la base de données.

Question 29. 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.