

Bases de données Avancées

Atelier 4 : Utilisation d'une source externe avec MongoDB

Dans ce TP de révision, vous allez commencer par télécharger les deux sources JSON suivantes :

- http://www.lirmm.fr/~servajean/movielens_movies.json
- http://www.lirmm.fr/~servajean/movielens_users.json

Importez ces deux sources dans MongoDB :

- `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 des films : id, titre (contenant l'année de production) et le type du film.

La collection **users** contient des informations sur les utilisateurs et les évaluations qu'ils ont données aux différents films. Un utilisateur est caractérisé par son identifiant, le nom, l'âge, la profession et le sexe. Les évaluations sont données comme un tableau de documents intégrés, chacun ayant un identifiant de film (qui fait référence à l'un des films dans l'autre collection), une note et un horodatage qui indique quand la cote était comprise).

Travail à faire

1. Affichez les structures des deux collections : **movies** et **users**
2. Combien y-a-il d'utilisateurs dans la base de données ?
3. Combien y-a-il de films dans la base de données ?
4. Quelle est l'occupation de **Clifford Johnathan** ? affichez seulement son nom et son occupation
5. Combien d'utilisateurs agés entre 18 et 30. 18 et 30 compris ?
6. Combien d'utilisateurs sont des artistes ou des scientifiques ?
7. Affichez les 10 femmes écrivains (**writers**) les plus âgées ?
8. Affichez toutes les occupations de manière distinctes
9. Insérez une entrée dans la collection **users**, sans indiquer le film.
10. Modifiez cette dernière entrée en ajoutant un film, sous forme d'un tableau. Vous pouvez utiliser : **Math.round(new Date().getTime() / 1000)**, pour avoir le timestamp courant.
11. Supprimez cette entrée de la collection
12. Changez l'occupation **programmer** par **developer**
13. Combien de films ont été produits dans les années 80 ?
14. Combien de films ont été produits entre les années 1984 et 1992 ?
15. Combien y-a-t-il de films d'horreur ?

16. Combien de films de type **Musical** et **Romance**, y-a-t-il ?
17. Dans cette question nous voulons créer un nouveau champ, **year**. Modifier la collection **movies** en supprimant l'année du titre et de l'ajouter dans un nouveau champ **year**.
18. Modifier la collection **movies** en remplaçant le champ contenant la chaîne genres par un tableau contenant tous les genres
19. Combien d'utilisateurs ont vu le film dont **_id** est **1196** (*Star Wars: Episode V - The Empire Strikes Back (1980)*) ?
20. Combien d'utilisateurs ont noté exactement 48 films ?
21. Pour chaque utilisateur créez un champ **num_ratings** affichant le nombre de films notés.
22. Combien d'utilisateurs ont noté plus que 90 films ?
23. **Barry Erin** vient de voir le film **Nixon** dont **l'id** est **14**. Ajoutez le film avec une note 4 dans le tableau **movies** de **Barry Erin**. N'oubliez pas que **num_ratings** doit refléter la taille du tableau **movies**.
24. **Marquis Billie** n'assume pas avoir vu le film **Santa with Muscles** (**_id 1311**). Supprimez ce film de la liste de ses films
25. Combien de films ont été produits chaque année des années 90, ordonnés par nombre de films de manière décroissante ?
26. Quelle est la moyenne des notes (**rating**) du film **Pulp Fiction** (dont **l'id** est **296**) ?
27. Dans une seule requête, retournez pour chaque utilisateur id, nom, maximum, minimum et la moyenne des notes données aux films. L'affichage doit se faire par moyenne croissante
28. Créez une nouvelle collection **Join** qui associe à chaque film son **_id**, son titre, son genre, l'année et toutes les notes attribuées par les différents utilisateurs