

Exercice MongoDB Stackoverflow

Objectif :

- Créer une base de données non structurée en MongoDB
- Faire une requête complexe dans cette base de données

Recupérer la bdd au format xml :

Telecharger une base les fichier xml du lien suivant :

<https://ia800107.us.archive.org/27/items/stackexchange/ai.stackexchange.com.7z>

décompresser le dossier dans « sources/data/mongodb_xml/ »

Nous nous intéresserons à deux dossiers :

Posts.xml

Users.xml

chacun contienne des données sous forme de ligne <row> (n'hésitez pas à les afficher)

Q1 :

À l'aide du script : « template_read_xml_mongodb.py » charger ces données dans une nouvelle base de données « stackoverflow » sous deux collections : « Posts » et « Users »

Q2 :

Lister les différentes clefs disponibles pour les deux collections.

Q3 :

Quelles sont les valeurs possibles pour la clef « PostTypeId » dans la collection Posts

la valeur « 1 » signifie que le post correspond au post initial, la valeur « 2 » correspond à un post réponse.

Q4 :

On souhaite à présent remplir une nouvelle collection Posts2 où chaque objet correspond aux posts et à leurs réponses. Supposant que la base précédente n'existe pas écrivez un script python remplissant cette nouvelle collection.

Q5 :

Trouver l'utilisateur avec la plus grande réputation sur la collection Users.

Q6 :

Compter le nombre de posts sur lesquels cet utilisateur apparaît dans la base Posts2

Q7 :

Créer une nouvelle variable dans la collection User comptant le nombre de posts sur lesquels l'utilisateur apparaît.

(une boucle for dans un nouveau script python avec une fonction update me semble le plus efficace, mais libre à vous de m'impressionner !)

