

BLOC STOCKAGE ET BIG DATA

MONGODB

CORBEILLE D'EXERCICES - correction

1 INTRODUCTION	2
Pré-requis	2
2 DESCRIPTION DES DONNEES	2
3 INSTALLATION DE MONGODB	3
4 IMPORT DES DONNEES	4
5 MANIPULATION DES DOCUMENTS	5
6 RESSOURCES	10



1 INTRODUCTION

Cette corbeille d'exercices a pour objectif de vous faire découvrir une autre base de données de type NoSQL. MongoDB est une base de données NoSQL orientée documents. Au travers de cette corbeille, vous allez manipuler MongoDB en ligne de commande à travers quelques requêtes.

Pré-requis

- Dernière version en *current release* MongoDB Community Server (actuellement 5.0.6) : <https://www.mongodb.com/download-center/community>
- La suite « MongoDB Database Tools » nécessaire pour bénéficier des commandes de base avec MongoDB : <https://docs.mongodb.com/database-tools/installation/installation/>
- Les fichiers de données CSV disponibles sur Moodle.

2 DESCRIPTION DES DONNEES

Restaurants

Cet ensemble de données au format JSON contient 25 357 documents relatifs à des établissements de restauration. Nous avons les informations suivantes : nom du restaurant, adresse (bâtiment, coordonnées, rue, code postal), quartier, type de cuisine, note.

Netflix

Cet ensemble de données contient 8807 films et séries. Pour chaque entrée, nous pouvons avoir les informations suivantes : ID, type, titre, directeur, casting, pays, date d'ajout sur la plateforme, date de sortie, note, durée, catégorie, description.

Height per country

Cet ensemble de données au format CSV contient 199 documents répertoriant les tailles moyennes d'hommes et de femmes par pays.

3 INSTALLATION DE MONGODB

Dans cette partie, vous allez installer MongoDB sur votre machine.

1. Si ce n'est pas déjà fait, téléchargez la version demandée dans les pré-requis
2. Procédez à l'installation :
 - Lancez le .msi téléchargé
 - Choisissez de procéder à l'installation complète
 - Cochez l'installation de MongoD en tant que service
 - *Laissez cochée la checkbox « installation de MongoDB Compass » si vous souhaitez l'utiliser dans le futur*
 - Lancez l'installation

Il se peut que durant l'installation votre PC ait besoin de redémarrer pour effectuer des mises à jour.

3. Une fois l'installation complétée l'interface de MongoDB Compass s'ouvre directement. Vous pouvez la fermer, nous n'utiliserons que l'invite de commande dans la suite de la corbeille.
4. Ajoutez les chemins vers **MongoDB\Server\5.0\bin** et **MongoDB\Tools\100\bin** aux variables d'environnement pour permettre l'accès au MongoDB shell directement depuis l'invite de commande

Vous pourriez avoir besoin de redémarrer la session ou la machine pour que le PATH soit mis à jour.

Alternative pratique pour de futures installations :

Utiliser une image Docker disponible sur le Docker Hub <https://hub.docker.com/>.

4 IMPORT DES DONNEES

Vous trouverez dans les ressources de la corbeille le fichier CSV appelé :

- netflix_dataset.csv
- restaurants.json

Importez le fichier CSV en utilisant la commande *mongoimport* dans l'invite de commande.

```
>mongoimport --db=corbeille --collection=restaurants --type=json --file=.\Desktop\dataset\restaurants.json
5.569+0100    connected to: mongodb://localhost/
6.551+0100    25357 document(s) imported successfully. 0 document(s) failed to import.

>mongoimport --db=corbeille --collection=netflix --type=csv --headerline --file=.\Desktop\dataset\netflix_dataset.csv
5.841+0100    connected to: mongodb://localhost/
6.173+0100    8807 document(s) imported successfully. 0 document(s) failed to import.

e>mongoimport --db=netflix --collection=titles --type=csv --headerline --file=.\Desktop\dataset\netflix_dataset.csv
```

Ressource utile : <https://docs.mongodb.com/manual/reference/program/mongoimport/>

Note : si vous n'êtes pas dans le répertoire contenant le fichier CSV à importer pour lancer la commande, n'oubliez pas de spécifier le chemin vers le fichier.

5 MANIPULATION DES DOCUMENTS

Vos bases de données sont maintenant créées.

Dans l'invite de commande, tapez « mongo » pour lancer le MongoDB shell dans lequel toutes les étapes suivantes seront réalisées.

Note : Si le shell ne se lance pas en tapant « mongo », vérifiez que vous avez correctement réalisé l'étape 4 de la partie « Installation de MongoDB ».

Quelques commandes à taper :

- ➔ *show dbs* : vous permet de voir vos bases de données.
- ➔ *db* : vous permet de voir dans quelle base vous êtes.
- ➔ *use nomBDD* : vous permet d'utiliser (ou créer) la-dite base.

1. Affichez la liste des collections contenues dans la base *corbeille*.

```
e(snelincip2).1.1  
> show collections  
netflix  
restaurants  
>
```

2. Quelle est la requête permettant de connaître le nombre de documents contenus dans la collection *netflix* et la collection *restaurants* ?
Combien de documents ces collections comptent-elles ?

```
Type it for more  
> db.netflix.find().count()  
8807  
> db.restaurants.find().count()  
25357  
> █
```

3. Ecrivez la commande vous permettant de récupérer un seul document dans chacune des deux collections et notez les différents champs existants.

```
> db.restaurants.findOne()
{
  "_id" : ObjectId("622666756fdca4f7383bd4eb"),
  "address" : {
    "building" : "97-22",
    "coord" : {
      "type" : "Point",
      "coordinates" : [
        -73.8601152,
        40.7311739
      ]
    },
    "street" : "63 Road",
    "zipcode" : "11374"
  },
  "borough" : "Queens",
  "cuisine" : "Jewish/Kosher",
  "grades" : [
    {
      "date" : ISODate("2014-11-24T00:00:00Z"),
      "grade" : "Z",
      "score" : 20
    },
    {
      "date" : ISODate("2013-01-17T00:00:00Z"),
      "grade" : "A",
      "score" : 13
    },
    {
      "date" : ISODate("2012-08-02T00:00:00Z"),
      "grade" : "A",
      "score" : 13
    },
    {
      "date" : ISODate("2011-12-15T00:00:00Z"),
      "grade" : "B",
      "score" : 25
    }
  ],
  "name" : "Tov Kosher Kitchen",
  "restaurant_id" : "40356068"
}
```

```
> db.netflix.findOne()
{
  "id" : ObjectId("62266689e78941980423f9f2"),
  "show_id" : "s1",
  "type" : "Movie",
  "title" : "Dick Johnson Is Dead",
  "director" : "Kirsten Johnson",
  "cast" : "",
  "country" : "United States",
  "date_added" : "September 25, 2021",
  "release_year" : 2020,
  "rating" : "PG-13",
  "duration" : "90 min",
  "listed_in" : "Documentaries",
  "description" : "As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help them both face the inevitable."
}
```

Quelques requêtes à réaliser :

4. Le nombre de films sortis en 2016.

```
> db.netflix.find({"release_year":2016}).count()  
902
```

5. Le nombre de films.

```
> db.netflix.find({"type":"Movie"}).count()  
6131
```

6. Le nombre de séries TV.

```
> db.netflix.find({"type":"TV Show"}).count()  
2676
```

7. Le nombre de films français dont la durée est égale à 90min.

```
> db.netflix.find({"country":"France"}, {"duration":"90 min"}).count()  
124
```

8. Le nombre de films américains sortis depuis 2020.

```
> db.netflix.find({"country":"United States","release_year":{"$gte:2020}}).count()  
474
```

9. La liste des types de restaurants (cuisine).

BLOC BDD BIG DATA / NO SQL

```
> db.restaurants.distinct('cuisine')
[
  "Afghan",
  "African",
  "American ",
  "Armenian",
  "Asian",
  "Australian",
  "Bagels/Pretzels",
  "Bakery",
  "Bangladeshi",
  "Barbecue",
  "Bottled beverages, including water, sodas, juices, etc.",
  "Brazilian",
  "Café/Coffee/Tea",
  "Café/Coffee/Tea",
  "Cajun",
  "Californian",
  "Caribbean",
  "Chicken",
  "Chilean",
  "Chinese",
  "Chinese/Cuban",
  "Chinese/Japanese",
  "Continental",
  "Creole",
  "Creole/Cajun",
  "Czech",
  "Delicatessen",
  "Donuts",
  "Eastern European",
  "Egyptian",
  "English",
  "Ethiopian",
  "Filipino",
  "French",
  "Fruits/Vegetables",
  "German",
  "Greek",
  "Hamburgers",
  "Hawaiian",
  "Hotdogs",
  "Hotdogs/Pretzels",
  "Ice Cream, Gelato, Yogurt, Ices",
  "Indian",
  "Indonesian",
  "Iranian",
  "Irish",
  "Italian",
  "Japanese",
  "Jewish/Kosher",
  "Juice, Smoothies, Fruit Salads",
  "Korean",
  "Latin (Cuban, Dominican, Puerto Rican, South & Central American)",
  "Mediterranean",
  "Mexican",
  "Middle Eastern",
  "Moroccan",
  "Not Listed/Not Applicable",
  "Nuts/Confectionary",
  "Other",
  "Pakistani",
  "Pancakes/Waffles",
  "Peruvian",
  "Pizza",
  "Pizza/Italian",
  "Polish",
  "Polynesian",
  "Portuguese",
  "Russian",
  "Salads",
]
```

10. Le nombre de restaurants dans le Queens.

```
> db.restaurants.find({"borough":"Queens"}).count()
5656
```

11. La liste de quartiers pour lesquels la collection contient au moins un restaurant.

```
> db.restaurants.distinct('borough')
[
  "Bronx",
  "Brooklyn",
  "Manhattan",
  "Missing",
  "Queens",
  "Staten Island"
]
```

12. Le nombre de pays pour lesquels la collection contient au moins un film.

```
> db.netflix.distinct('country').length
749
```

13. La taille moyenne des hommes, tous pays confondus ?

```
> db.height.aggregate({"$group":{"_id":null, "avg_maleheigh":{"$avg":"$Male Height in Cm"}}})
{ "_id" : null, "avg_maleheigh" : 173.08904522613065 }
```

14. La taille moyenne des femmes en cm, tous pays confondus ?


```
> db.height.aggregate({"$group":{"_id":null, "avg_femaleheight":{"$avg":"$Female Height in Cm"}}})  
{ "_id" : null, "avg_femaleheight" : 160.9429145728643 }
```

15. Le pays ayant les hommes les plus petits ? Quelle est la taille ?

```
> db.height.aggregate({"$group":{"_id":null, "minmaleheight":{"$min":"$Male Height in Cm"}}})  
{ "_id" : null, "minmaleheight" : 160.13 }
```

⇒ Timor-Leste, 160.16cm

16. Le pays ayant les femmes les plus grandes ? Quelle est la taille ?

```
> db.height.aggregate({"$group":{"_id":"$Country name", "maxfemaleheight":{"$max":"$Female Height in Cm"}}})  
{ "_id" : null, "maxfemaleheight" : 170.36 }
```

⇒ Pays-bas, 170.36cm

6 RESSOURCES

- <https://www.mongodb.com/fr>
- <https://docs.mongodb.com/guides/server/import/>
- <https://www.kaggle.com/shivamb/netflix-shows>
- <https://s3-eu-west-1.amazonaws.com/course.oc-static.com/courses/4462426/restaurants.json.zip>
- <https://www.kaggle.com/majyhain/height-of-male-and-female-by-country-2022>
- <https://ourworldindata.org/>

Cheatsheet des commandes MongoDB

<https://gist.github.com/bradtraversy/f407d642bdc3b31681bc7e56d95485b6>