

Exercice 1 : Prise en main de MongoDB

Nous utiliserons une base de données contenant des informations sur des films. Les données sont enregistrées dans deux fichiers :

- ***moviesEmbedded.json*** : Ce fichier contient un document pour chaque film avec les attributs suivants:
 - *_id*, identifiant du film.
 - *title*, titre du film.
 - *year*, année de sortie du film.
 - *genre*, le genre du film.
 - *summary*, le résumé du film.
 - *country*, le code du pays du film (FR pour France, IT pour Italie...).
 - *director*, document imbriqué décrivant un réalisateur contenant les attributs suivants :
 - *_id*, identifiant du réalisateur.
 - *first_name*, prénom du réalisateur.
 - *last_name*, nom du réalisateur
 - *birth_date*, date de naissance du réalisateur.
 - *actors*, collection de documents. Chaque document décrit un acteur avec les mêmes attributs que pour les réalisateurs + l'attribut :
 - *role*, le rôle de l'acteur dans le film.

- ***moviesBoxOffice.json*** : Ce fichier contient le nombre d'entrée de chaque film, avec les attributs suivants :
 - *_id*, identifiant du film.
 - *box_office*, nombre d'entrée du film.

Question 1.1 : Création d'un répertoire de stockage et lancement du serveur MongoDB

- Lancez un premier terminal pour créer un répertoire de stockage de votre base MongoDB ayant le nom de votre login :

```
mkdir /data/mdb/ppsbdj/ppsbdj_i (remplacez j par votre groupe de TP et i par votre numéro de compte dans le groupe)
```

- Utilisez ce terminal pour lancer le serveur « mongod » en précisant le répertoire de stockage :

```
Smongod --dbpath /data/mdb/ppsbdj/ppsbdj_i --port 10000
```

- **Laissez ce terminal actif** (iconifiez-le).

Question 1.2 : Téléchargement des données dans MongoDB

Nous créons dans MongoDB une base de données nommée *cinema* (n'utilisez pas d'accents) et deux collections nommées respectivement *movies* et *movies_boffice* et nous y téléchargerons les données.

- a. Recopiez deux fichiers de données au format JSON sur votre compte :

```
cp ~cpu_vialle/PPS-BD/movies* ~/
```

- b. Puis créer la collection *movies*, dans la base de données *cinema*, qui contiendra les documents du fichier *movies.json* :

```
mongoimport --port 10000 --db cinema --collection movies  
--jsonArray --file moviesEmbedded.json
```

Vous devriez lire un message disant : *imported 88 documents*.

- c. Puis créer la collection *movies_boffice*, qui contiendra les documents du fichier *moviesBoxOffice.json* :

```
mongoimport --port 10000 ...
```

Rmq : si le port 10000 est déjà pris et que *mongod* refuse de se lancer : choisissez le port 10001, 10002, ... jusqu'à en trouver un de libre !

Question 1.3 : Lancement du *shell* MongoDB pour interroger la BdD

- a. Afin de lancer le *shell* MongoDB que vous utiliserez pour interagir avec le serveur MongoDB, entrez la commande suivante :

```
mongosh --port 10000
```

- b. Spécifiez au serveur MongoDB que vous voulez utiliser la base de données *cinema* par la commande suivante :

```
use cinema
```

- c. Vérifiez que les documents sont bien dans la collection *movies* en tapant la requête suivante :

```
db.movies.findOne()
```

- d. Vérifiez que les documents sont bien dans la collection *movies_boffice* en tapant la requête suivante :

```
db.movies_boffice.findOne()
```

Exercice 2 : Requêtes, Agrégations, Join et mapReduce

Question 2.1 : Requêtes MongoDB

En utilisant le *shell* MongoDB, écrivez les requêtes suivantes en MongoDB pour afficher :

- a. L'année d'apparition du film "Le parrain III".
- b. Le titre des films sortis entre 1980 et 1990.
- c. Même requête que b. avec un tri alphabétique des titres.
- d. Le titre des films français.
- e. Le titre des films du genre "crime" ou "drama".
- f. Les noms et dates de naissance des réalisateurs des films français.
- g. Le titre des films dont Sofia Coppola est l'un des acteurs.
- h. Le titre et les genres des films dont le réalisateur est Hitchcock.

Question 2.2 : Agrégation en MongoDB

En utilisant des agrégations (*avec aggregate([...])*), écrivez les requêtes suivantes en MongoDB pour afficher :

- a. Le nombre de films par pays. Afficher par nombre décroissant.
- b. Le nom de l'acteur dans le rôle de "Mary Corleone" dans le film "Le parrain III".
- c. Le nombre d'acteurs par film, triés par nombre décroissant.
- d. Le nombre moyen d'acteurs par film.

Question 2.3 : Join en MongoDB

Dans une agrégation, en utilisant l'opérateur *\$lookup* (qui réalise un "Join") sur les collections *movies* et *movies_boffice*, retrouvez le nombre d'entrée du film "Le parrain III".

Le résultat devrait être : 5061739.

Question 2.4 : Join et mapReduce

- a. Reprenez tout d'abord l'exercice précédent pour générer une collection "movies2" contenant le résultat du *\$lookup* pour tous les films de la base.
- b. Ecrivez maintenant un mapReduce en MongoDB pour calculer le nombre d'entrées au box-office auxquelles a contribué chaque acteur :
 - Décrivez votre algorithme.
 - Définissez une fonction *map1* et une fonction *reduce1* en JavaScript dans le Mongo *shell*.
 - Réalisez un mapReduce en appelant les fonctions *map1* et *reduce1*, et en stockant le résultat dans une collection "actorsbo".
 - Affichez les nombres d'entrées de chaque acteur, par ordre décroissant.

Le résultat devrait être :

Bruce Willis : 34090927

Morgan Freeman : 26362352

....

Rutger Hauer : 2123163

Christian Slater : 2118341

Consignes pour le CR simplifié (pour bilan en fin de TP)

- **CR en monôme ou bien en binôme pour le xx/11/22 à 23h59**
 - **Indiquez clairement vos nom, prénoms, date et titre de TP**
 - **8 pages maximum (« tout inclus »)**
 - Mettez les extraits de vos requêtes et codes MongoDB
 - Mettez les réponses de MongoDB à vos requêtes (ou des extraits de ses réponses)
 - Vous ne rendrez QUE votre CR (pas de fichiers de code en plus)
 - **Envoyez votre CR par email à :**
 - Gianluca.Quercini@centralesupelec.fr
- ET**
- Stephane.Vialle@centralesupelec.fr