**Rapport Projet noSQL :**

Groupe : Tanguy Vuillemin et David Sidoun – Classe 47

Lien GitHub projet :

<https://github.com/MTGBV/ProjetnoSQL.git>

[git@github.com:MTGBV/ProjetnoSQL.git](mailto:git@github.com:MTGBV/ProjetnoSQL.git)

Introduction :

Ce projet a pour objectif de travailler sur l’intégration de données MongoDB puis de les transférer vers Neo4j, une base de données permettant de voire les données sous forme de graphes. Également, nous serons amené à manipuler un dataset .JSON pour stocker et interroger une grande liste de films et acteurs.

Première partie sur les requêtes mongo :

1. Trouver l’année avec le plus de films qui sont sortis

Fonction python pour la requête :

1. #question 1 : l'année avec le plus grand nombre de films sortis
2. def get\_year\_with\_most\_films():
3. db = get\_mongo\_connection()  # Connexion à la base MongoDB
4. collection = db["films"]  # Accède à la collection 'films'
6. pipeline = [
7. {"$group": {"\_id": "$year", "count": {"$sum": 1}}},
8. {"$sort": {"count": -1}},  # Trie par nombre de films, du plus grand au plus petit
9. {"$limit": 1}  # Limite à 1 résultat
10. ]
12. result = collection.aggregate(pipeline)
13. # Récupérer l'année et le nombre de films
14. for item in result:
15. return item["\_id"], item["count"]

Résultat de la requête sur streamlit :

Une image contenant texte, Police, capture d’écran, Graphique

Le contenu généré par l’IA peut être incorrect.

2. Trouver le nombre de films sortis après 1999 :

Fonction Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Résultat sur streamlit :

Une image contenant texte, Police, capture d’écran

Le contenu généré par l’IA peut être incorrect.

3. Question 3 : trouver la moyenne des votes des films sortis en 2007 :

Fonction Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage sur streamlit :

Une image contenant texte, Police, capture d’écran, Graphique

Le contenu généré par l’IA peut être incorrect.

4. Question 4

Requête Python :

Une image contenant texte, capture d’écran, Police, affichage

Le contenu généré par l’IA peut être incorrect.

Les modifications sur cette question sont surtout différentes au niveau d l’affichage sur streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, logiciel, Rectangle

Le contenu généré par l’IA peut être incorrect.

5. Question 5

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

6. Question 6

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

(mauvais film si je puis me permettre)

7. Question 7

Requête Python :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police, Graphique

Le contenu généré par l’IA peut être incorrect.

8. Question 8

Requête python :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

9. Question 9

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage Streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

10. Question 10

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

11. Question 11 :

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

12. Question 12 :

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police, conception

Le contenu généré par l’IA peut être incorrect.

13. Question 13

Requête Python :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

**Durée moyenne des films par genre (en minutes)**

* **War** : 153.0 min
* **History** : 134.67 min
* **Crime** : 133.67 min
* **Western** : 132.0 min
* **Biography** : 130.86 min
* **Sci-Fi** : 130.56 min
* **Action** : 128.9 min
* **Music** : 128.0 min
* **Adventure** : 127.85 min
* **Mystery** : 125.2 min
* **Drama** : 125.14 min
* **Fantasy** : 124.2 min
* **Family** : 120.5 min
* **Thriller** : 116.0 min
* **Romance** : 114.17 min
* **Comedy** : 110.91 min
* **Animation** : 99.57 min
* **Horror** : 96.0 min

**Partie 2 : Base de données Neo4j**

Tout d’abord, le premier objectif de cette partie est d’exporter les données implémentées sur MongoDB (fichiers movies.json dans la base Entertainment et dans la collection films) vers notre instance Neo4j aura ainsi que d’ajouter des nœuds pour voire les films et certaines de leurs caractéristiques, les acteurs et de créer des liens avec nos noms.

Nous avons écrit un fichier mongo\_to\_neo4j.py afin de se connecter à MongoDB puis d’extraire le fichier JSON cité précédemment afin de créer un fichier JSON films\_export prêt à l’utilisation pour neo4j.

Nous mettons ensuite en place le fichier neo4j\_load.py nous permettant de lire le fichier films\_export et surtout pour créer les différents nœuds et donc pour répondre aux différentes consignes demandées.

Ce fichier est ensuite compilé afin de bien envoyer les données vers l’instance neo4j aura :

Une image contenant logiciel, Logiciel multimédia, multimédia, Logiciel de graphisme

Le contenu généré par l’IA peut être incorrect.

Également, pour la partie sur la création de nœuds avec les membres du groupe, nous avons écrit séparément un fichier neo4j\_membresLink.py pour nous lier au film Rogue One comme vous pouvez le voire sur l’instance neo4j (ce fichier sera également runné):

Une image contenant capture d’écran, texte, cercle, horloge

Le contenu généré par l’IA peut être incorrect.

On centralise maintenant les requêtes cypher dans le fichier python neo4j\_queries.py afin de répondre aux questions du projet.

14. Question 14 :

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Résultat affiché sur streamlit :

Une image contenant texte, capture d’écran, Police, ligne

Le contenu généré par l’IA peut être incorrect.

15. Question 15

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

16. Question 16

Requête Cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

On a rajouté le champ revenue dans le load.py afin d’avoir les champs des revenues des films.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

17.Question 17

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

18.Question 18

Ajout du champ genre dans neo4j

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage du résultat streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

19.Question 19

Requête Cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

20.Question 20

Requête Cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

21.Question 21 :

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage sur streamlit du résultat :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

22.Question 22

Requête Cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Résultat :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

23.Question 23 (trop bien celle-là)

Requête cypher :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit quand on entre Ben Affleck sur streamlit :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

24.Question 24

Requête cypher :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

25.Question 25

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit avec les acteurs de l’exemple (Tom Hanks le goat et Scarlett Johansson):

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Ce chemin signifie que la route la plus courte pour aller de Tom Hanks à Scarlett Johansson est en passant par le fait que Tom Hanks a joué dans Sully ainsi qu’une autre personne (Unknown) et cet acteur inconnu a joué dans The Prestige, un film où Scarlett Johansson a joué.

26.Question 26

Je n’ai pas réussi à réaliser cette question car je crois que gds.graph.project n’est pas disponible sur neo4j aura mais je ne suis pas sur que mes erreurs soient du à ca.

**Questions transversales**

27.Question 27

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Partie du résultat sur streamlit :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

28.Question 28

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage streamlit :

Une image contenant texte, Appareils électroniques, capture d’écran, logiciel

Le contenu généré par l’IA peut être incorrect.

29.Question 29

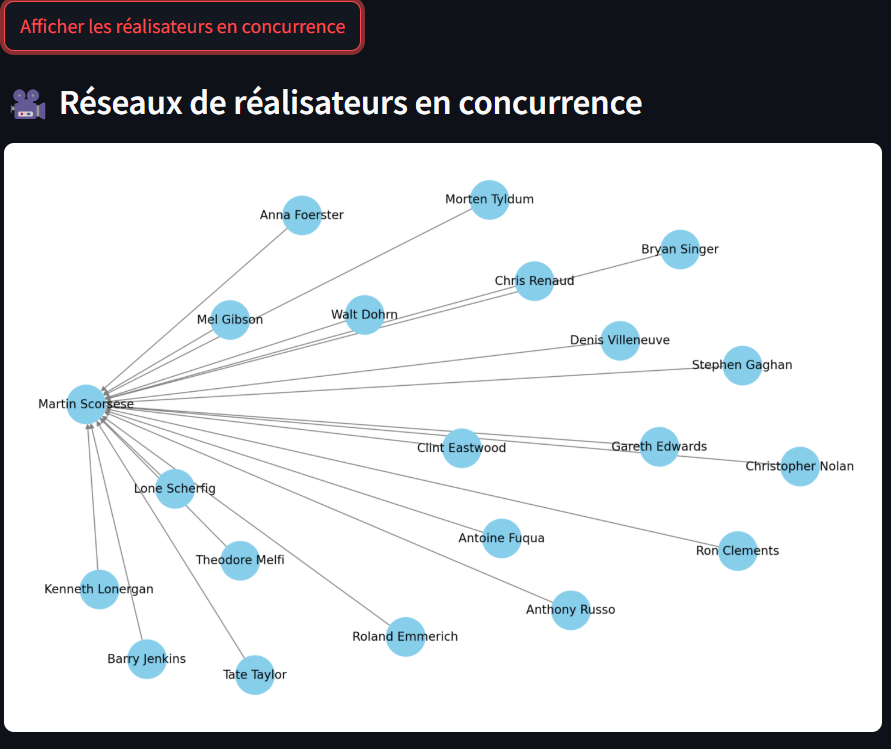
Installation de networkx pour cette question.

Requête cypher :

Une image contenant texte, capture d’écran

Le contenu généré par l’IA peut être incorrect.

Affichage du résultat sur streamlit :



On peut en conclure que Martin Scorsese est un réalisateur ayant un lien avec beaucoup d’autres réalisateurs, ce qui signifie qu’il a réalisé beaucoup de films.

30.Question 30

Requête cypher :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Affichage du résultat sur streamlit :

Une image contenant texte, capture d’écran, logiciel, Police

Le contenu généré par l’IA peut être incorrect.

Difficultés rencontrées :

Tout d’abord, une des difficultés majeures que nous avons rencontré a été l’exportation des données de MongoDB vers Neo4j. En effet, certains champs comme les genres étaient stockés sous forme de texte sur MongoDB ce qui pouvait rendre difficile leur traitement de MongoDB vers Neo4j.

Pour résoudre ces problèmes avec certains champs, nous avons créé une étape secondaire avant de traiter les données vers Neo4j. En effet, nous avons converti certaines informations du fichier .JSON dans un autre fichier de façon à ce que les données soient adaptés pour la réalisation d’une structure de graphes sur Neo4j.

Également, une autre difficulté a été la question sur le graph project car même lors de l’exécution d’un code basique de graph project, nous rencontrions des erreurs sur Neo4j et nous n’avons pas réussi à résoudre ces erreurs.