

1 INFORMATIONS GENERALES

Apprenti :	Nom:	Prénom :
Lieu de travail :	ETML / Vennes / 1004 LAUSANNE	
Client :	Nom:	Prénom: -
Dates de réalisation :	Semaines 5 à 11 Du 2.02.2024 au 15.03.2024	
Temps total	32 périodes	

2 PROCÉDURE

- Tous les apprentis réalisent le projet sur la base d'un cahier des charges.
- Le cahier des charges est présenté, commenté et discuté.
- Les apprentis sont entièrement responsables de la sécurité de leurs données.
- En cas de problèmes graves, les apprentis avertissent leur chef de projet au plus vite.
- Les apprentis ont la possibilité d'obtenir de l'aide externe, mais ils doivent le mentionner.
- Ce projet est individuel.
- Un mini-rapport est demandé à chaque apprenti de manière individuelle
- En cas d'imprévu, le temps total doit être géré par l'apprenti.

3 SUJET

Implémenter et utiliser une base de données MongoDB

4 MATÉRIEL ET LOGICIEL À DISPOSITION

- Un PC ETML
- Environnement docker avec container MongoDB (serveur et mongosh)
- Interface de gestion MongoDB compass + VS Code avec l'extension MongoDB
- Accès à Internet

5 PRÉREQUIS

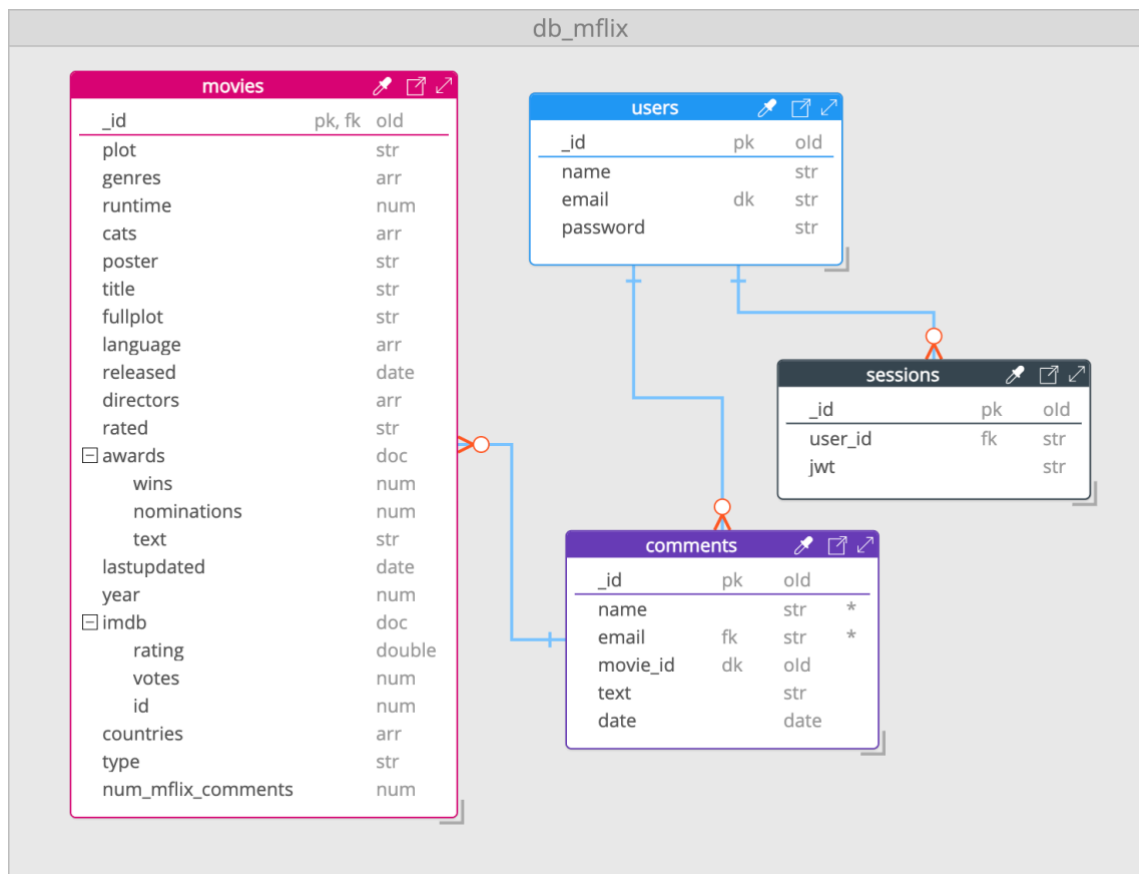
Module I165

6 LIVRABLES

- Un rapport de projet avec une mise en page soignée comprenant les explications et réponses du point A, F et G du cahier des charges.
- Pour chaque points B, C, D et E un fichier playground comprenant un commentaire par commande avec leur numéro d'identification respectif (4 fichiers).
- Vos fichiers devront respecter le nommage suivant :
X-165-P-DB-NoSQL-NomPrénom-<rapport | playground-B | ... >.(pdf | mongodb.js)
- Date de remise des livrables : jour de la dernière séance, 23h59 sur le « Devoir Teams ».

7 CAHIER DES CHARGES :

Vous avez à disposition un fichier ayant pour nom **db_mflix.gz**, contenant les collections de la base de données « db_mflix ». Ce fichier est un dump fait avec **mongodump** de type *archive* et *compressé*, dont voici le MPD :



Ce MPD exprime les collections ainsi que les liaisons entre les collections avec leurs cardinalités selon la notation dite « **Crow's Foot** », [patte d'oie].

A- Importer les données et le schéma de la base de données

Travail demandé :

- Restaurez les collections dans une base de données nommée « db_mflix » dans une instance de MongoDB.
- Expliquer en détail la commande utilisée.

B- Gestion des utilisateurs

Trois types d'utilisateurs interagiront avec la base de données MongoDB. Pour ce faire, chaque type d'utilisateur a des besoins d'accès et de manipulation de données différents :

1. Administrateur de **db_mflix**

- 1.1. Peut créer ou supprimer des collections ;
- 1.2. Peut créer, lire, mettre à jour et supprimer (CRUD) n'importe quels documents des collections ;
- 1.3. Gérer les indexes pour toutes les collections ;
- 1.4. Gérer les rôles (et donc les utilisateurs) et leurs privilèges de cette base de données.

2. Utilisateur
 - 2.1. Lire les informations sur les films et les commentaires ;
 - 2.2. Ajouter ou supprimer un ou des commentaires.
3. Gestionnaire
 - 3.1. Lire les informations sur tous les utilisateurs (pour savoir qui a fait un commentaire) ;
 - 3.2. Mettre à jour, lire et supprimer des films ou des commentaires ;
 - 3.3. Lire tous les commentaires.

Travail demandé :

- Expliquer comment créer des utilisateurs et attribuer des rôles spécifiques qui correspondent à leurs besoins ;
- Créer un exemple d'utilisateur pour chaque type.

C- Requêtes de sélection

Lister tous les films ou séries selon les critères suivants :

1. d'action
2. sortis après l'an 2000
3. réalisés après 2010 mais avant 2020
4. sortis dans les années 2000 mais ayant un style des années 1980
5. réalisés par « Quentin Tarantino »
6. de la série « Star Wars »
7. avec un score « IMDb » supérieur à 8
8. qui ne sont pas de genre « Horror » ou « Sci-Fi »
9. ayant exactement 3 différents genres
10. dont le dernier genre est « Drama »
11. qui durent entre 1h30 et 2h
12. avec plus de 100 commentaires
13. qui ne sont pas classés « R »
14. dont le titre commence par « The »
15. ayant reçu un certain nombre d'awards
16. où le réalisateur et le premier acteur sont les mêmes
17. dans lesquels « Brad Pitt » et « Angelina Jolie » sont acteurs
18. dans lesquels « Brad Pitt » est acteur et dont le nombre de commentaires est au moins égal à 100
19. où l'acteur principal est une « femme »
20. où « Tom Hanks » est acteur, mais pas « réalisateur »
21. mais doivent apparaître uniquement le titre et l'année de sortie de chaque film
22. Dans le cadre d'une pagination qui renvoie à chaque fois une liste de 10 films par page, quel est la requête permettant de renvoyer uniquement la liste des films de la 3^e page ?
23. Rechercher les films qui ont au moins la langue « française » ou la langue « anglaise » disponible, qui sont sortis à partir de « 1980 » inclus, et qui ont une note « Rotten Tomatoe » de plus de 4 ou un score « IMDb » supérieur ou égal à 8. Nous voulons également que « Brad Pitt » joue dans le film. Nous souhaitons n'avoir que les titres pour pouvoir les afficher directement. Enfin, nous ne voulons pas que le « synopsis » du film parle de « nazis ».
24. Nous voulons maintenant trier notre résultat suivant les « notes » attribués au film. D'abord par la note « Rotten Tomatoe », puis par la note « IMDb ».

D- Agrégations

1. Compter le nombre de films par genre
2. Compter le nombre de films par classification (rated)
3. Calculer la durée moyenne des films par genre
4. Calculer la durée moyenne des films par décennie
5. Calculer la durée moyenne des films par acteur
6. Lister les 5 réalisateurs les plus fréquents
7. Lister les 5 acteurs les plus fréquents dans les films « PG-13 »
8. Quel est le nombre moyen de commentaires par film
9. Le genre le plus populaire par année
10. Lister les genres distincts des films
11. Lister les films par décennie avec le nombre total de films par décennie
12. Trouver le genre le plus courant dans chaque pays
13. Trouver le nombre de films par classification et par décennie
14. Calculer le nombre total de films et la durée moyenne des films par réalisateur
15. Notre objectif est de créer des groupes par pays. Pour chaque pays, nous voulons créer des groupes de chaque genre et obtenir le nombre de films, la note moyenne des films et la part de marché (nombre de films d'un genre pour un pays / total de films du pays).

E- Requêtes avec Recherche Textuelle

1. Recherche des films contenant le mot-clé « matrix »
2. Recherche tous les films contenant la phrase « best movie ever »
3. Recherche des films contenant le terme « war », tout en excluant ceux contenant le terme « world »
4. Trouver des films où la description contient « space » mais pas « mars »
5. Trouver des films avec le mot exact « time travel » dans la description
6. Rechercher des films avec des mots commençant par « inter » dans le titre

F- Création des index

Dans votre base de données **db_mflix**, identifiez les champs qui bénéficieraient le plus de la création d'index.

- Expliquez pourquoi la création d'index sur ces champs serait pertinente et comment cela pourrait améliorer les performances des requêtes et l'efficacité globale de la base de données.
- Démontrez minimum 3 exemples avec ou sans les indexes ajoutés (collection, indexe(s), requêtes, temps de réponse avant/après).

G- Backup/Restore

Nous souhaitons réaliser un backup complet de notre base de données **db_mflix**. Ensuite, nous aimerions procéder à sa restauration. Nous aimerions que la sauvegarde prenne le moins de place possible.

- Expliquez chaque étape de ces processus et les raisons pour lesquelles ces étapes sont importantes.

8 EVALUATION :

1. Une auto-évaluation challengée par le client basé sur des éléments observables inscrits dans la planification initiale à environ 80% du temps imparti pour le projet sera effectuée entre les semaines 9 et 10.
2. Une évaluation finale sera effectuée par le client après la livraison final du projet.
3. Le recours à des outils en ligne d'intelligence artificielle (ex. : Chat GPT) doit être mentionné et ne peut servir que d'inspiration à la réalisation. En cas d'abus, l'évaluation du projet en tiendra compte.