Projet Base de données évoluées

Intégration et entrepôts de données

Ibrahima DIALLO
Gurvan Le MEUR
Irving FALADE
Mamadou Saliou DIALLO

Analyse des plateformes de streaming vidéo

Plan

- ➔ Introduction
- ➔ Identification du processus d'entreprise à analyser
- ➔ Identification du grain de processus de l’entrepôt
- ➔ Présentation du modèle en étoile
- ➔ Quelques requêtes d'agrégats
- ➔ Méthodes d’intégration
- ➔ Nettoyage des données
- ➔ Difficultés rencontrées
- ➔ Conclusion
- ➔ Ressources

Introduction

- L’objectif de ce projet
- Pourquoi ce choix
- Choix des datasets

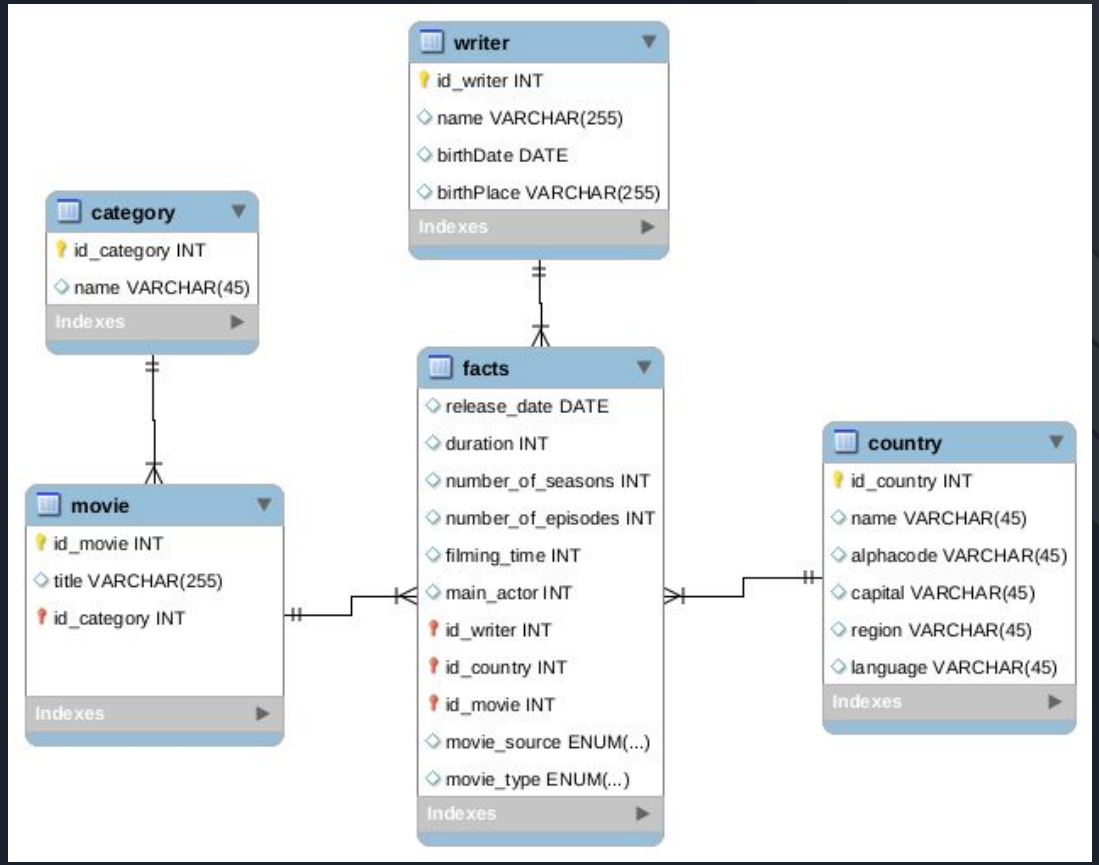
Processus d’entreprise à analyser

- Quels sont les films et séries que proposent les plateformes de streaming
- Dans quels pays ces films sont produits
- Les catégories associées à ces films
- Quels sont les réalisateurs

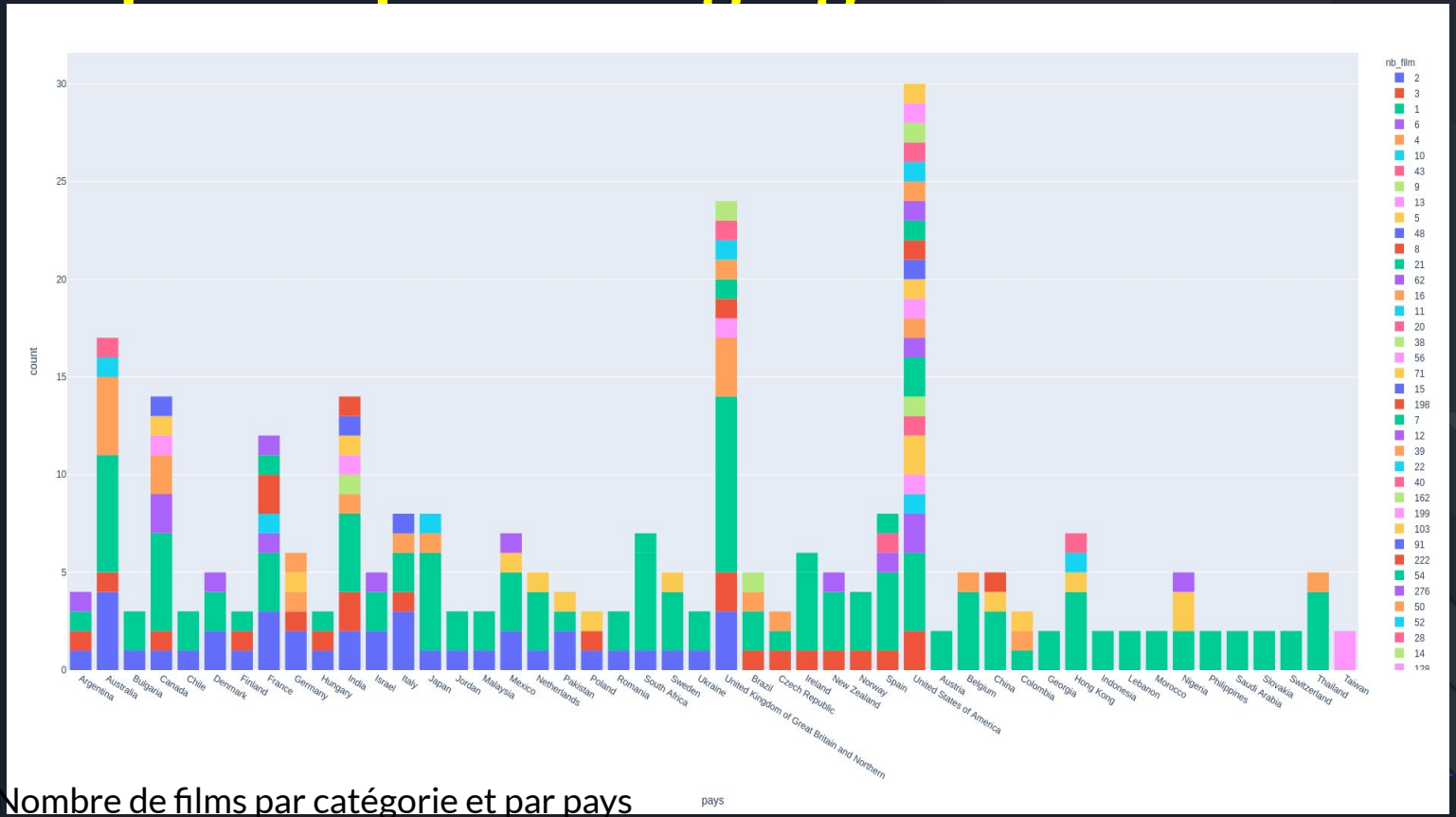
Grain de processus de l’entrepôt

- ★ Le titre
- ★ La date de sortie
- ★ La durée
- ★ Le nombre de saisons
- ★ Le nombre d’épisodes
- ★ Le temps de tournage
- ★ Le nombre d’acteurs principaux
- ★ Le type (film ou série)
- ★ La plateforme qui propose le film ou la série

Présentation du modèle en flocon



Quelques requêtes d'agrégats



Nombre de films par catégorie et par pays

Méthodes d’intégration

- Modélisation de la base de données
- Mapping entre les champs des datasets et ceux de la base de données
- Intégration automatique des données avec un script python
- Utilisation de requêtes SPARQL pour compléter les informations incomplètes dans les datasets

Nettoyage des données

- Suppression des tuples ne possédant pas de catégories
- Seul la première catégorie est prise en compte (cas des catégories multiples)
- Suppression des films dupliqués

Difficultés rencontrées

- Indisponibilité de CUBE et GROUPING SETS avec mysql
- Datasets un peu hétérogènes

Conclusion

- Projet très intéressant
- Libre, Autonome
- Possibilité d’ajouter des dimensions supplémentaires

Ressources

https://www.kaggle.com/unanimad/disney-plus-shows

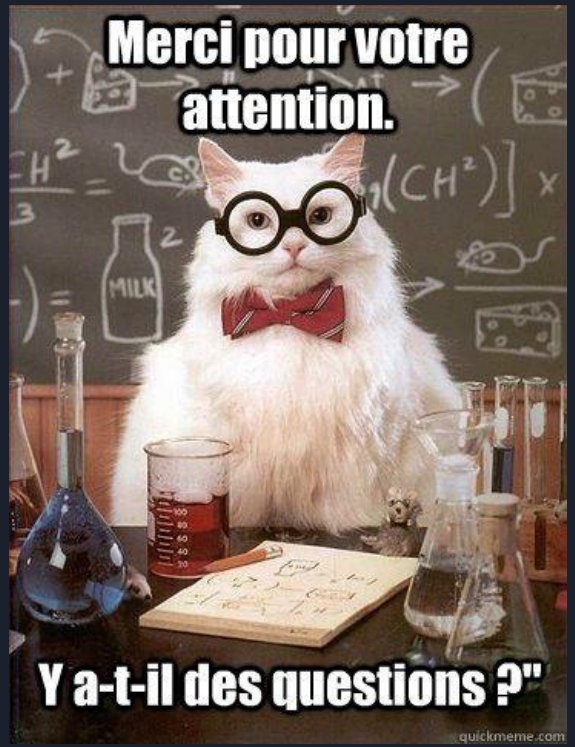
https://www.kaggle.com/shivamb/netflix-shows

https://www.kaggle.com/nilimajauhari/amazon-prime-tv-shows

https://restcountries.eu/rest/v2/all

Dépôt Git

https://github.com/saliou673/movie-datawarehouse


Merci pour votre attention.
Y a-t-il des questions ?"
MILK
quickmeme.com