

The background features several abstract, organic shapes in shades of teal and grey. A large, light teal shape is in the top right, a dark teal circle is in the top center, and a grey circle is in the top left. In the bottom left, there are overlapping dark teal, grey, and light teal shapes. A grey circle is also in the bottom right.

# Catalogue De données

# Description des données :

L'ensemble de données sur les évaluations de films comprend une collection complète de 100 000 évaluations de films fournies par 943 utilisateurs pour 1 682 éléments distincts. Chaque utilisateur a consciencieusement évalué au moins 20 films, et les utilisateurs comme les éléments sont identifiés de manière unique par des identifiants numériques attribués consécutivement. Les données sont organisées dans un format séparé par des tabulations.

## 1. Movies (u\_item.txt) :

Contient des informations sur chaque film, notamment son titre, sa date de sortie et ses genres.

### Fields :

- Movie ID
- Title
- Release Date
- Genres (Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western)

## 2. Users (u\_user.txt) :

Contient des informations démographiques sur chaque utilisateur, telles que l'âge, le sexe, la profession et le code postal.

### Fields :

- User
- IDAge
- Gender
- Occupation
- Zip Code

## 3. Ratings (u\_data.txt) :

Contient les notes attribuées par les utilisateurs aux films, ainsi que les horodatages.

### Fields :

- User ID
- Movie ID
- Rating
- Timestamp

#### **4. Genres (u\_genre.txt) :**

Contient une liste de genres.

##### **Fields :**

- Genre ID
- Genre Name

#### **5. Occupations (u\_occupation.txt) :**

Contient une liste de professions.

##### **Fields :**

- Occupation ID
- Occupation Name

### **Données utilisées dans le projet :**

#### **u\_data :**

- Un ensemble de données fondamentales contenant 100 000 évaluations par 943 utilisateurs pour 1 682 films. Les évaluations des utilisateurs, accompagnées d'horodatages, fournissent une vue dynamique des préférences des utilisateurs au fil du temps.

#### **u\_item :**

- Informations sur les films, y compris les titres, les dates de sortie et les classifications par genre. Les genres sont représentés sous forme d'indicateurs binaires, ce qui permet une exploration détaillée des divers éléments thématiques présents dans chaque film.

#### **u\_genre :**

- Cet ensemble de données sert de référence pour les genres, en fournissant une liste de genres avec des identifiants uniques. Il constitue un lien essentiel pour comprendre et classer les films en fonction de leur contenu thématique.

## Comment on a utilisé les données :

- **Exploration des évaluations de films et des préférences des utilisateurs :**
- **Dataset:** u\_data
- **Objectif :** Comprendre les préférences des utilisateurs et les modèles d'engagement.
- **Approche :**
  - Utilisez l'ensemble de données u\_data pour explorer les évaluations de films fournies par les utilisateurs.
  - Analyser les préférences des utilisateurs en fonction des classements, des genres et des modèles temporels.
  - Identifier les films et les genres populaires pour informer les stratégies de recommandation.

### La curation de contenu basée sur le genre :

- **Datasets:** u\_genre, u\_item
- **Objectif :** Créer du contenu basé sur des genres spécifiques.
- **Approche :**
  - Exploiter u\_genre et u\_item pour comprendre le paysage des genres.
  - Identifier les genres en vogue et adapter les recommandations de contenu en conséquence.
  - Mettre en œuvre des stratégies de promotion de contenus alignés sur les préférences des utilisateurs.

### Traitement de données en temps réel avec Kafka et Memgraph :

- **Datasets:** u\_data
- **Objectif :** Permettre le traitement des données en temps réel pour des recommandations dynamiques.
- **Approche :**
  - Configurer Kafka pour la diffusion en continu des données d'évaluation des films.
  - Développer un module de transformation Python pour adapter les messages Kafka à Memgraph.
  - Configurer Memgraph pour traiter des flux de données en temps réel et mettre à jour la base de données de graphes.

## Analyse du comportement des utilisateurs à l'aide de Memgraph et de Cypher Queries :

- **Datasets:** u\_data
- **Objectif :** Développer des connaissances sur le comportement des utilisateurs pour des recommandations personnalisées.
- **Approche :**
  - Utiliser Memgraph pour l'analyse en temps réel des interactions des utilisateurs avec la base de données graphique.
  - Développer et optimiser les requêtes Cypher pour extraire des modèles significatifs.
  - Identifier les groupes de préférences des utilisateurs et améliorer l'algorithme de recommandation.