

Les données du Cinéma

1. Présentation générale du laboratoire

Ce second laboratoire, d'une durée de deux semaines, vise à mettre en pratique vos apprentissages en termes d'interrogation de données, en utilisant le langage de manipulation de données de SQL (LMD).

Vous disposez d'une base de données - en lecture seule – comprenant plus d'un siècle de données du cinéma mondial. Voici une description de la base de données en chiffres :

- Plus de 150 000 films, dont certains remontent aux frères Lumière et Méliès;
- Près de 200 000 acteurs et réalisateurs ;
- Une répartition des films, des acteurs et des réalisateurs sur plus de 200 pays ;
- Des informations sur les œuvres dont les prix remportés et l'appréciation des internautes selon les chiffres de IMDB.

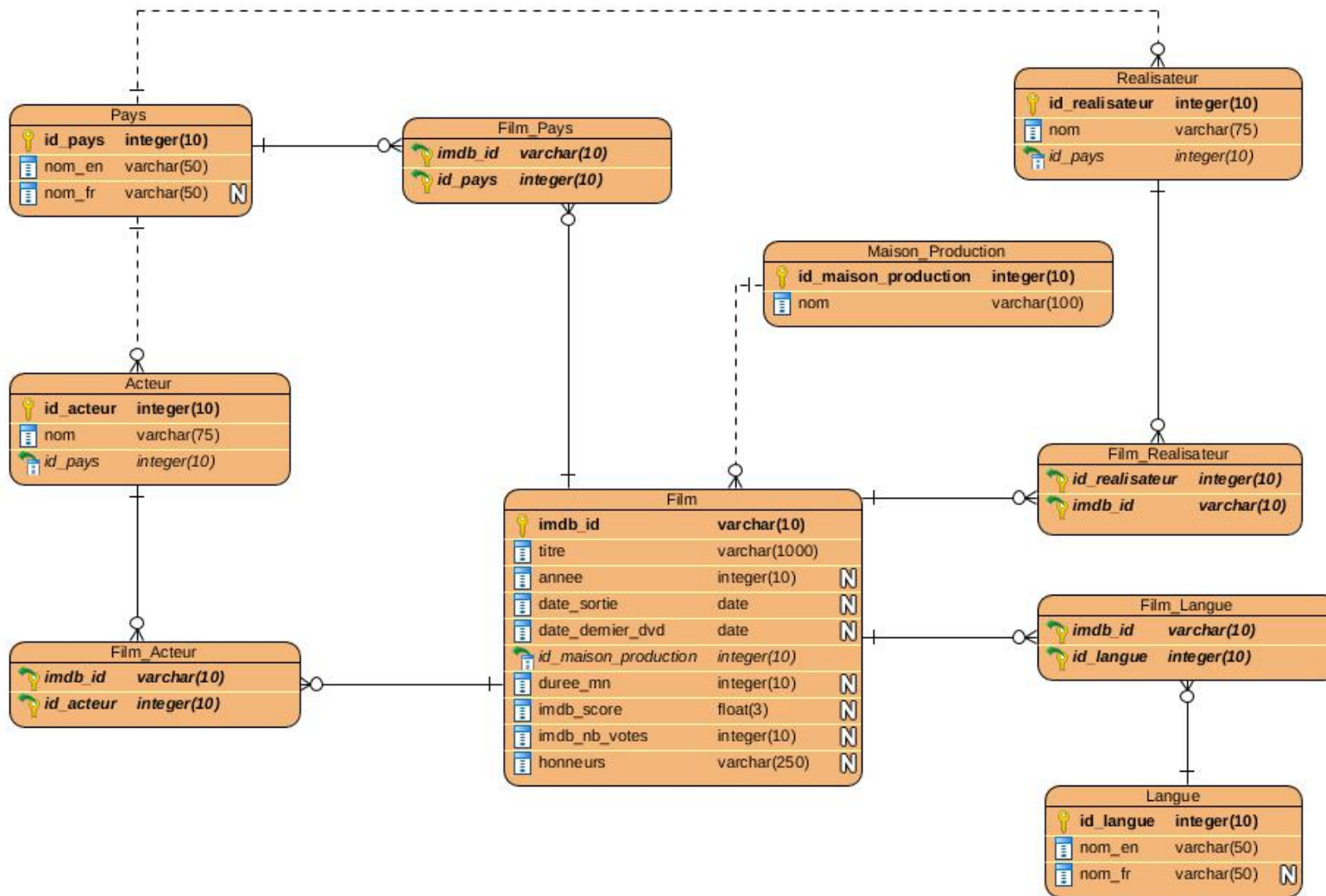
La base de données est fournie gracieusement par le site imdb.com par l'entremise du site kaggle.com (site qui fournit des ensembles de données aux fins expérimentales). La version dont vous disposerez a été modifiée par vos enseignants pour qu'elle puisse être exploitée dans un SGBD relationnel¹ et pour y ajouter des liens servant des objectifs académiques. C'est ainsi qu'il se peut que certaines informations soient fausses car elles ont été forgées de toute pièce, tel le pays d'origine des acteurs (mais elle ont été forgée avec un algorithme qui rend la chose plausible).

Votre travail, dans le cadre de ce laboratoire, sera donc d'écrire les requêtes qui permettent d'obtenir les informations qui vous sont demandées depuis la base de données. La section suivante vous fournit le modèle relationnel de la base de données. La section 3 énumère les requêtes auxquelles vous devez répondre dans le cadre de ce laboratoire.

Ce laboratoire doit être réalisé en groupes, en préservant les mêmes groupes que le premier laboratoire (à moins qu'un changement ait été autorisé par votre enseignant-e).

¹ Les données disponibles sur Kaggle.com le sont sous forme aplatie au format *CSV* ou *JSON*.

2. Modèle relationnel



3. Liste des requêtes

Votre travail consiste à fournir 36 requêtes permettant de répondre aux questions métier ci-dessous. Les questions sont réparties en six blocs, chacun portant sur un type de requêtes spécifique.

* : Signifie que vous pouvez répondre à la requête dès la semaine 1 du laboratoire.

Bloc 1 (10 pts)

Requête 1.1 *:

Listez tous les titres des films. Le résultat serait semblable à celui-ci:

```
Titre
-----
The City Beautiful
Cursed by His Beauty
Dance of the Vampire
...
```

Requête 1.2 *:

Identifiez tous les films (titre, et date de sortie) ayant remporté un Oscar depuis l'an 2000.

Indice: La colonne "honneurs" contient des valeurs telles que :

"Won 3 Oscars. Another 213 wins & 260 nominations."

Requête 1.3 *:

Trouvez tous les films ayant une durée de plus de 4 heures et dont le score sur le site IMDB est supérieur 7. Votre requête doit afficher le titre du film (réduit à 75 caractères), l'année de sortie, sa durée et son score. Le résultat devra apparaître ainsi :

TITRE_PARTIEL	DUREE_MN	SCORE	DATE_SOR
Kill Bill: The Whole Bloody Affair	247	8.8	11-03-27
Happy Hour	317	7.3	15-12-12
A Lullaby to the Sorrowful Mystery	485	7.8	16-03-26
Beautiful Violin Guitar Flute Harp Instrumental Music	263	7.8	17-08-10

81 rows selected.

Requête 1.4 :

Trouvez le nombre de titre de film différents dans la base de données. Veuillez renommer la sortie pour ne pas laisser apparaître la formule.

Requête 1.5 *:

Retrouvez l'id_pays du/des réalisateur(s) dont le nom de famille est 'Carruth'.

Requête 1.6 *:

Sélectionnez les différents noms des acteurs de manière unique

Bloc 2 (10 pts)

Requête 2.1 *:

Trouvez la liste des maisons de production (présentée en ordre alphabétique) ayant produit au moins un film dans les 5 dernières années. Note: Votre requête doit s'ajuster à la date à laquelle elle est exécutée (i.e. 2019 ne doit pas y figurer explicitement).

Requête 2.2 *:

Vous voulez regarder un bon film français récent (pas plus vieux que 2015) ayant obtenu un score d'au moins 8. Comment feriez-vous pour obtenir la liste des titres de ces films et l'année? La liste obtenue doit être classée par année de sortie (en commençant par les films les plus récents), puis par score IMDB.

Requête 2.3 *:

Pour le film ayant pour imdb_id 'tt0019514', veuillez lister les noms des acteurs y ayant participé.

Requête 2.4 :

Pour affiner quelques statistiques sur notre SGBD, nous souhaitons connaître le nombre d'acteurs participant par films.

Requête 2.5 *:

Veuillez lister les titres des films où l'acteur 'Seth Green' a tourné.

Requête 2.6 *:

Trouvez les noms et l'année des films de réalisateur canadiens.

Bloc 3 (20 pts)

Requête 3.1 *:

Trouvez la liste des acteurs ayant joué dans au moins un film de leur propre pays.

Requête 3.2 *:

Trouvez la liste des acteurs qui jouent dans un film qu'ils réalisent.

Requête 3.3 *:

Trouvez les films qui sont réalisés par des maisons de productions qui ont le même nom que le réalisateur du film.

Requête 3.4 *:

Trouvez les films dont au moins un des acteurs est du même pays que le réalisateur.

Requête 3.5 *:

Trouver les films dont la langue commence par la même suite de 4 lettres que le pays du réalisateur (i.e. « Canada », « Canadien »).

Requête 3.6 *:

Trouvez tous les films, sortis depuis l'an 2000, ayant gagné un Oscar et qui ont été réalisés par au moins un réalisateur canadien ou ceux où dans lesquels moins un acteur canadien joue. Voici le résultat:

TITRE	ANNEE
-----	-----
Blade Runner 2049	2017
Brokeback Mountain	2005
Chicago	2002

Bloc 4 (15 pts)

Requête 4.1 *:

Listez toutes les informations des maisons de productions n'ayant jamais produit de films.

Requête 4.2 *:

Listez les identifiants des acteurs, leurs noms et les identifiants des pays pour les acteurs n'ayant pas joué dans des films.

Requête 4.3 *:

Sélectionnez les pays qui n'ont aucun acteur.

Bloc 5 (20 pts)

Requête 5.1 *:

Combien y-a-t-il de films dans la base de données?

Requête 5.2 :

Trouvez la liste des réalisateurs ayant réalisé au moins 40 films.

Requête 5.3 :

Listez les titres et le nombre de langues différentes pour les films ayant plus de deux langues.

Requête 5.4 :

Trouver la durée moyenne (en mn) des films pour chaque maison de production.

Requête 5.5 :

Évaluez le score moyen, le score minimal et le score maximal des films pour chaque maison de production.

Requête 5.6 :

Veillez indiquer le nombre d'acteurs par pays, en affichant l'id_pays et le nombre d'acteurs correspondants.

Bloc 6 (25 pts)

Requête 6.1 :

Donnez la liste de noms des acteurs et des réalisateurs canadiens, en spécifiant pour chaque personne si c'est un acteur ou un réalisateur. Voici le résultat attendu:

NOM	TYPE_PERSON
Aaron Au	Réalisateur
Aaron Chan	Acteur
Aaron Chan	Réalisateur
Aaron Davidson	Réalisateur
Aaron Edelstein	Acteur
Aaron Ellis	Réalisateur
Aaron Floresco	Réalisateur
Aaron Forward	Acteur
Aaron Hagey-MacKay	Acteur
Aaron Hall	Réalisateur
Aaron Huggett	Réalisateur

La liste n'a pas besoin d'être triée et les doublons sont permis (si une personne est à la fois actrice et réalisatrice).

Requête 6.2 :

Trouvez les noms des réalisateurs, leur pays et le nombre de films réalisés des réalisateurs ayant réalisé au moins 40 films.

Requête 6.3 :

Listez l'id_maison_production des maisons de production, ayant une moyenne des scores supérieure à la moyenne totale des scores de l'ensemble des maisons de productions.

Requête 6.4 :

Trouver les années de sortie des films ayant le plus petit nombre de votes.

Requête 6.5 :

En moyenne, dans combien de films jouent chaque acteur de notre base de données?

Requête 6.6 :

Listez les acteurs qui participent à plus de films que la moyenne des acteurs.

Connexion à la base de données

Vous devez utiliser les mêmes paramètres de connexion et utiliser le code d'accès suivant:

- **Nom d'utilisateur:** equipe100
- **Mot de passe:** jNwp2xue

Livrable et remise finale

Vous devez remettre un fichier à l'extension **.sql** comprenant l'ensemble de vos requêtes en utilisant le gabarit qui vous est fourni (vous n'aurez qu'à insérer votre requête dans l'endroit prévu).

Vous devez porter une attention particulière à la forme de vos requêtes : bien distinguer les diverses sections de la requête (ne pas tout écrire sur une ligne) et commenter les parties que vous jugez les plus délicates pour y expliquer votre raisonnement.

La remise doit être effectuée sur Moodle à remettre durant la semaine du 11 mars 2019 (voir date précise selon votre groupe sur Moodle). En respect des règles du Service des Enseignements Généraux, toute remise tardive se verra attribuer la note de 0 (à moins d'entente justifiée et préalable avec l'enseignant-e).

Bon travail 😊