

TP BI : Concevoir un entrepôt de données

L'objectif de ce TP étant de travailler sur la base de données IMDB.

Il s'agit de construire une Data Warehouse permettant d'analyser plusieurs KPIs.

1. Choix des KPIs :

a. KPIs communs :

Nous comptons répondre aux requêtes suivantes

- Quel type le film est le mieux noté durant la dernière année du dataset ?
- Qui est l'acteur (actrice) qui a joué le plus de rôles dans les 5 dernières années du data set ? Quel est le pourcentage du genre trouvé dans la 1^{ère} requête parmi tous les genres qu'il assuré pendant cette période ?
- Donner les 10 films qui ont reporté le plus de chiffre d'affaire dans les Box Offices.

b. KPIs spécifiques :

Donner 3 KPIs de votre choix

2. Sources de données :

Inspirez-vous du travail de : <https://github.com/jayshilj/ETL-Pipelines-and-Business-Intelligence-on-IMDB-dataset>

Vous pouvez vous baser sur le répertoire « Data ».

Vous avez également le libre cours d'ajouter d'autres jeux de données. Sur Internet, plusieurs dataset IMDB existent gratuitement.

3. Technologies :

Vous avez le choix entre :

ETL	Stockage	Reporting
Talend ou SSIS	MySQL Ou Oracle Ou SQL Server Ou Postgres	Power BI* Tableau* Jasper*

4. Travail demandé :

a. Obligatoire

- Analyser les données sources
- Mettre en place les jobs ETL avec l'outil de votre choix
- Mettre en place la conception d'une DWH bien argumentée
- Alimenter votre DWH
- Écrire 2 requêtes répondant au moins à 2 KPIs

b. Obligatoire

- Monter un TDB avec l'outil de votre choix pour
 1. Démontrer les KPIs communs
 2. Démontrer les KPIs spécifiques