

Travail Dirigé et Projet

Intégration et entrepôts de données

Description

1

L'objectif de ce projet est d'intégrer des données dans un entrepôt de données. Chaque **quadrinôme** doit définir un sujet d'entrepôt différent. Le choix du sujet est libre, cependant, le rapprochement avec un exemple réel contera pour la note du projet.

Ce projet sera noté par une soutenance de projet (date à définir) et par un compte rendu (document pdf).

Voici quelques exemples de sujets :

- Analyse du marché de la location chez Airbnb.
- Analyse du réseau du transport public et leur relation avec le covoiturage.
- Trafic maritime dans le canal de La Manche.
- Analyse de stations de vélo publiques dans le monde.

N'hésitez pas à chercher des datasets sur les données ouvertes disponibles sur internet.

Travail demandé

- Identifier le sujet d'analyse et les datasets à réutiliser. Vérifier que les licences de ces datasets vous permettent leur réutilisation. Prévoyez une licence pour le dataset que vous intégrerez.
- Identifier les "faits numériques" qui permettront de faire des analyses.
- Réalisez la conception de l'entrepôt de données avec un **schéma en étoile** ou avec un (ou plusieurs) **agrégat(s)** si vous utilisez de datastores NoSQL (Cassandra, MonetDB, MongoDB, etc.).
- Les datasets utilisés peuvent être dans un format hétérogène comme CSV, JSON, EXCEL, XML, etc., vous devez intégrer les données dans un format commun selon votre schéma ou agrégats.
- Donnez une dizaine de requêtes intéressantes en utilisant les opérateurs GROUP BY, GROUP BY CUBE, GROUP BY ROLLUP et GROUP BY GROUPING SETS, GROUPING, GROUPING_ID, fenêtres mobiles, RANK, PARTITION BY, top n, NTILE, etc. Si vous utilisez de datastores NoSQL cherchez les commandes nécessaires pour ces types d'analyse.

Barème

Pour noter le projet on prendra en compte :

- Le rapprochement avec un exemple réel, la pertinence et originalité du sujet d'analyse.
- La clarté de votre démarche technique.
- La facilité de réutilisation de votre travail.
- La présentation orale de votre projet (une démonstration sera un plus).
- La présentation et la clarté du compte rendu (distribution et organisation des tâches, démarche, mise en valeur du travail effectué, difficultés rencontrées, etc.).