

Projet BI (Partie Théorique)

1) Sujet d'analyse principal

Le sujet principal de l'entrepôt est l'analyse des tweets pour identifier les tendances, les sentiments associés, les hashtags les plus utilisés et les interactions comme les retweets.

2) Modélisation multidimensionnelle

Tables et nature des champs :

1. Table des faits : **Faits_Tweets**

- Clés primaires :

- ID_Tweet : Identifiant unique du tweet.

- Attributs :

- Date_ID : Identifiant de la date du tweet.
- ID_Utilisateur : Identifiant de l'utilisateur ayant posté le tweet.
- ID_Sentiment : Sentiment associé au tweet.
- ID_Hashtag : Identifiant du hashtag (peut contenir plusieurs hashtags).
- Est_Retweet : Booléen indiquant si c'est un retweet.

2. Tables de dimensions :

- Date

- Date_ID (PK)
- Jour, Mois, Année

- Utilisateur

- ID_Utilisateur (PK)
- Nom_Complet
- Sexe
- Tranche_Age
- Nationalité

- Sentiment

- ID_Sentiment (PK)
- Sentiment_Description (euphorique, heureux, neutre, triste, en colère, scandaleux)

- Hashtag

- ID_Hashtag (PK)
- Hashtag_Texte

3) Schéma de données obtenu :

Le modèle est un **schéma en étoile** avec une table de faits centrale connectée à plusieurs dimensions.

Schéma :

