

# **Projet Final : Data Warehouse et ETL**

L'objectif de ce projet est de créer une solution complète pour l'intégration, la transformation, et le stockage des données de la population mondiale pour la période 1960-2023. Ce projet met en œuvre des concepts d'ETL (**Extract, Transform, Load**) en utilisant **Apache NiFi** pour l'intégration des données, et un entrepôt de données (Data Warehouse) pour l'analyse et le reporting, avec **SQL Server** comme base de données de destination.

## **Données sources :**

- Fichier **population\_mondiale.csv** contient les données de la population mondiale pour la période de 1960-2023.

## **Partie 1 : Processus ETL**

1. Créer un groupe de processeur nommé **INGESTION\_POPULATION\_MONDIALE**
2. Récupérer le fichier concerné puis effectuer les transformations nécessaires pour bien nettoyer les données.
3. Charger les données dans la table **population\_mondiale\_1960\_2023** (penser à gérer les dédoublemnages des données dans la table en créant un trigger).
4. Configurer le pipeline pour qu'il s'exécute automatiquement chaque le 5 janvier.

## **Partie 2 : Data Warehouse**

A partir du flowfile ci-dessus (question 3), alimenter les tables suivantes en utilisant les processeurs adéquats (créer au préalable les tables dans SQL Server) :

- population\_afrique\_1960\_2023
- population\_amerique\_du\_nord\_1960\_2023
- population\_amerique\_du\_sud\_1960\_2023
- population\_europe\_1960\_2023
- population\_asie\_1960\_2023
- population\_oceanie\_1960\_2023