

## Exercice : Implémentation de Deux DAGs Airflow dans Cloud Composer pour la Gestion et Transformation de Données CSV

**Objectif Principal** : Le but est de créer deux DAGs dans Apache Airflow, déployés via Google Cloud Composer. Le premier DAG gèrera le processus d'extraction des données des fichiers CSV et leur chargement dans une table RAW. Le second DAG se déclenchera suite au succès du premier, pour charger les données RAW dans une table de l'entrepôt de données (DWH) après leur transformation.

### Instructions de l'Exercice

#### DAG 1: Extraction et Chargement des Données

##### 1. Conception du DAG :

- Planifiez le DAG pour qu'il se déclenche à 10h du matin chaque jour.
- Sélectionnez les opérateurs Airflow adéquats pour interagir avec Cloud Storage.

##### 2. Extraction et Chargement :

- Créez une table externe référençant l'emplacement des fichiers CSV dans Cloud Storage.  
Créer un dossier nominatif dans ce bucket **gcs-viseo-data-academy-22024**  
Dossier à créer dans le bucket NOM/in
- Chargez les données de la table externe vers une table RAW ayant la même structure dans la base de données : ( tous les champs doivent avoir le type string )
- Dataset BQ à utiliser **raw**  
Table à préfixer par RAW\_NOM\_xxx ( tous les champs doivent avoir le type string )

##### 3. Gestion des Fichiers :

- En cas de succès, déplacez le fichier CSV traité dans un dossier **archive** du même bucket/NOM
- En cas d'échec, déplacez le fichier dans un dossier **error**.

#### DAG 2: Transformation et Chargement dans l'Entrepôt de Données

##### 1. Déclenchement :

- Configurez ce DAG pour qu'il se déclenche automatiquement après le succès du premier DAG.

##### 2. Transformation des Données et Chargement dans l'Entrepôt de Données :

- Créer la table cible dans le dataset **datawarehouse**
- Utilisez une requête SQL pour transformer les données RAW\_NOM\_xx  
Transformation à faire :

1. Démarrer le script par un truncate de la table
2. Modifier le type de chaque champs cible ( à vous de voir , date , numeric etc.. )  
Calculer une nouvelle colonne TotalPrice =  
Quantity\*Price
3. Chargez les données transformées dans une table spécifique de l'entrepôt de données (DWH).