Automatiser : Création d’un espace Amazone S3 Charger et Scrapper

En tant qu’analyste BI nous travaillons sur les données, dans notre cas nous recevons régulièrement les données en provenant de plusieurs sources.

Nous avons pensé à créer un espace dans AWS S3 ou les données seront chargées, pour nos analyses et création des dashboard nous avons mis en place un flux ssis qui recupèrent automatiquement les données se trouvant dans un dossier en locale pour ensuite les charger dans les rapports et nous faisons les refresh manuellement.

L’objectif de ce projet est d’automatiser le chargement des données en provenance de plusieurs sources dans les rapports.

Etapes :

* Créer un compte dans amazone S3, Créer un Bucket
* Charger des fichiers dans le Bucket

Objectif : Télécharger ces fichiers en locale automatiquement tous les N-temps

* Ecrire un script
* import boto3
* import os
* from dotenv import load\_dotenv
* load\_dotenv()
* bucket\_name = os.environ['BUCKET\_NAME']
* local\_directory = 'C:\\\\\\\\\\'
* session = boto3.Session(
* aws\_access\_key\_id=os.environ['AWS\_ACCESS\_KEY\_ID'],
* aws\_secret\_access\_key=os.environ['AWS\_SECRET\_ACCESS\_KEY'],
* region\_name=os.environ['AWS\_REGION']
* )
* s3 = session.client('s3')
* objects = s3.list\_objects(Bucket=bucket\_name)['Contents']
* #ignored\_objects=["IVC.xlsx"]
* for obj in objects:
* key = obj['Key']
* #  if(key not in ignored\_objects):
* if 'retail' not in key.lower():
* continue
* local\_file\_path = os.path.join(local\_directory, key)
* os.makedirs(os.path.dirname(local\_file\_path), exist\_ok=True)
* s3.download\_file(bucket\_name, key, local\_file\_path)
* print(f'Downloaded: {key} to {local\_file\_path}')
* Créer le fichier .env ou mettre les credentiels
* Créer un environnement virtuelle dans vscode

Python -m venv venv

Venv\script\activates

* Installer les prérequis

Créer un fichier .txt et configurer les prérequis

boto3==1.33.4

botocore==1.33.4

jmespath==1.0.1

python-dateutil==2.8.2

python-dotenv==1.0.0

s3transfer==0.8.2

six==1.16.0

urllib3==2.0.7

Installer les prérequis

Pip install -r requirement.txt

Exécuter le code

Mahamat taher Ali Medillaye

Consultant Junior Analyste Business Intelligence