

Analyse de logs avec Python et SQL

Ce projet, réalisé dans le cadre de la formation 1SIO en janvier 2025, vise à structurer et analyser des fichiers de logs générés par un proxy utilisé par la société STESIO. L'objectif principal est de faciliter l'analyse des accès web des employés pour répondre à des obligations légales et produire des statistiques utiles.

Principales étapes du projet :

1. Génération de pseudonymes : Création d'un script Python (`pseudo.py`) pour transformer des données d'entrée en pseudonymes selon un format spécifique.
2. Mise en place de la base de données : Configuration d'une base MariaDB (`logs_web`) avec des tables pour stocker les données des logs et des employés, incluant la gestion des utilisateurs et des droits d'accès.
3. Analyse des logs : Développement de scripts Python pour extraire et analyser les données des fichiers logs, notamment les URL les plus consultées par utilisateur (`url_par_utilisateur.py`).
4. Conversion et insertion des données : Transformation des logs en fichiers CSV (`logproxy.py`) et génération de scripts SQL (`generate_sql_insert.py`) pour insérer les données dans la base.

Environnement et outils :

1. Langages : Python, SQL.
2. Base de données : MariaDB.
3. Outils : Spyder, Visual Studio Code, phpMyAdmin, bibliothèques Python (Pandas, Matplotlib, Seaborn, etc.).
4. Machine virtuelle : Rocky Linux 9 avec Apache, MariaDB et phpMyAdmin.

Livrables :

1. Scripts Python commentés.
2. Documentation technique (compte-rendus, schémas, captures d'écran).
3. Fichiers CSV et SQL générés.

Lien vers drive du projet:

https://drive.google.com/file/d/1Gpn4RmphEaDqTSQJk0VdEWWCg3ywKA7x/view?usp=drive_link