

Mise en place d'un Pipeline ETL dans le cloud sur les données météorologiques en Afrique

Description

Ce projet a pour but de collecter journalièrement sur une API de la NASA des données sur la météorologie en Afrique, de les transformer et de les stocker dans le cloud pour permettre aux utilisateurs d'avoir des bases finies pour leurs analyses.

Introduction

La compréhension des données météorologiques est un sujet clé pour la mise en place de politiques contre le réchauffement climatique. L'Afrique, bien que produisant moins de gaz à effet de serre reste tant bien que mal concernée par ce phénomène. Il est donc particulièrement exposé aux variations climatiques, avec des phénomènes météorologiques extrêmes tels que la sécheresse, les inondations et les vagues de chaleur qui affectent les populations, l'agriculture et l'économie. Malgré l'importance de ces enjeux, les données météorologiques restent souvent fragmentées, difficiles d'accès et insuffisamment exploitées. La mise en place d'un pipeline dans le cloud permet de collecter, nettoyer, stocker de vastes ensembles de données météorologiques en temps quasi réel (journalier), facilitant ainsi l'accès à des informations fiables

Objectif de l'Étude

L'objectif de l'étude est d'extraire les données climatiques de toute l'Afrique, de les transformer et de les stocker dans Amazon S3.

Méthodologie

➤ **Extraction**

Les données sont extraites tous les jours de l'API <https://power.larc.nasa.gov/api> de la NASA. Celle-ci est faite à travers un algorithme python intégrée dans une fonction de Amazon Lambda.

➤ **Transformation**

Dans cette partie, il est question d'identifier les problèmes liés à la base et de les corriger. Ici, il s'agit ici des types des variables et du formatage des caractères spéciaux. Cette partie a également été réalisée à partir d'un script python implémenté sur Amazon lambda.

➤ Chargement

Le chargement a été fait sur Amazon S3.