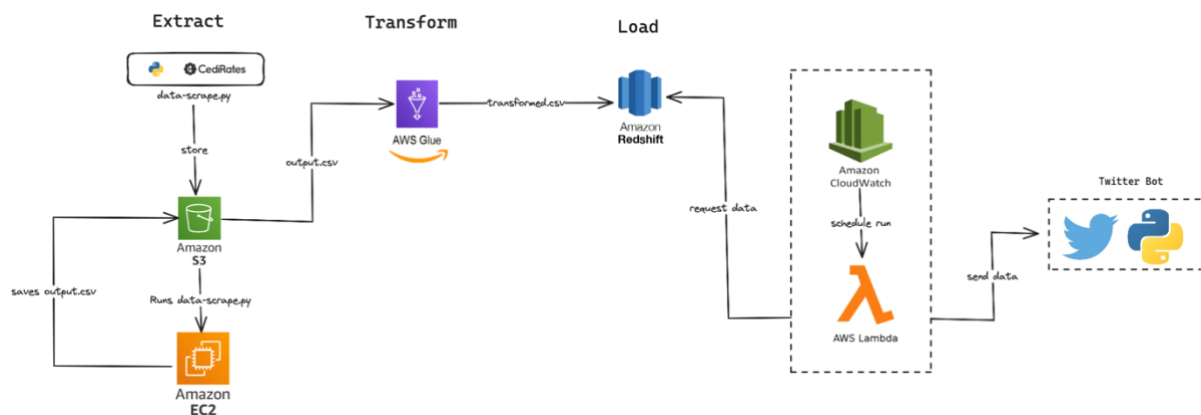# Fuel Prices Pipeline

OBJECTIVE:
The objective of this project is to build a data pipeline that gets data regarding the prices of fuel on a daily basis, saves it to a data warehouse and creates an automated tweet daily announcing the prices.

ARCHITECTURE DIAGRAM

Primary Data Source:
www.cedirates.com

Primary Data Scraped:
Daily Diesel and Petrol Prices

AWS Services Used
- AWS S3: to persist data
- Amazon EC2: to run scripts on the cloud
- AWS Glue: To transform data and store it in Redshift
- Amazon Redshift: Data warehouse service, stores all daily prices from inception till termination
- Amazon CloudWatch: For monitoring, logging and scheduling the pipeline.
- AWS Lambda: serverless execution of the automated tweet logic.

CURRENT LIMITATIONS:
1. The pipeline implementation is incomplete
2. The pipeline has not yet been orchestrated to run at specific intervals (daily)
3. Twitter bot implementation is not complete
4. Minimal transformation ideas of the raw scraped data