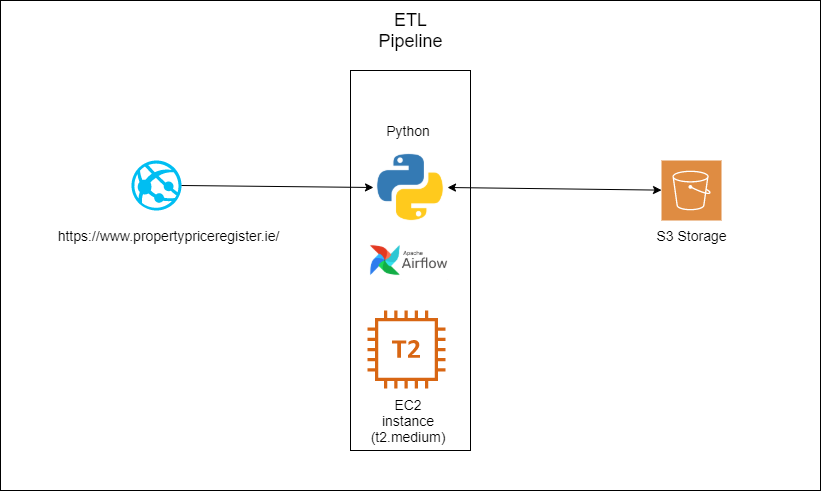
**Architecture for ETL Pipeline**

The below architecture diagram explains the ETL pipeline for geowox data processing, cleaning and transformation. The whole architecture is set up in the AWS cloud service. EC2 (t2.medium) instance has python and airflow installed. Python is used for the ETL process and the pipeline is orchestrated using apache airflow. The data lake is set up in AWS S3 for storing raw, historical and transformed data.

****

This architecture is cost-effective and the instance (t2.medium) has 2 VCPU’s and 4GB RAM can process huge volume and variety of data in memory.