

Realtor

The project name is **Realtor.** It is a Data Engineering project which uses ETL data integration. Tools used in the project:

* Python3 (NumPy, Pandas), Jupyter notebook
* Web Scraping
* PostgreSQL
* AWS
* RDS
* S3
* Lambda
* CloudWatch
* EventBridge
* Cloud9
* Redshift

Operational database consists of 17 tables, from which 15 tables are static and 2 tables are transaction. OLTP process is done in RDS environment with PostgreSQL (pgadmin4). ETL is automated with AWS Lambda and Python3. Scheduling is performed by CloudWatch and EventBridge. Data Warehouse is built and maintained in Redshift database.

**Realtor** operations:

1. With web scraping of realtor.ca 95635 estates are scraped.
2. With generation process of Python are generated **users**, **owners**, **buyers** tables
3. Data in 17 tables is migrated into AWS S3
4. Data pipelines are created for ETL process
5. Redshift database has star schema