**Set up, initialize, and interact with PostgreSQL database by Python**

**Setups**

Tools for project:

**VScode** as Python IDE, **PostgreSQL** as RDBMS and **pgAdmin** is the management tool.

RDBMS (relational database management system): MySQL, PostgreSQL, SQL Server, etc.

Data Warehouse, Data Lake: Snowflake, Databricks, AWS Redshift, Azure Synapse, etc.

RDBMS is typically used for transaction recording or OLTP (online transaction processing).

DW/DL is usually used for storing insights of an organization or OLAP (online analytical processing), DSS (decision support systems).

**1. Initialize and load data into PostgreSQL RDBMS**

Initialize and load data: *Khanh\_user*, *retail\_db*

“*Initialize DB.sql*”

“*Create DB tables.sql*”

“*Load DB tables.sql*”

Make a few queries:

“*Cumulative aggregation and ranking.sql*”

“*Performance tuning for SQL queries.sql*”

**2. Process some data using Python**

“*dataTypesInPython.ipynb*”

“*dataProcessingUsingPandas.ipynb*”

**3. File format converter using Python**

To improve the efficiency of data engineering pipelines, we need to convert CSV files into JSON files, since JSON is better to use in downstream applications than CSV files.

“*fileFormatConverter.ipynb*”

“*app-fileFormatConverter.py*”

**4. Loading files to database**

“*loadingFilesToDatabase.ipynb*”

“*truncateTables.ipynb*”

“*app-loadingFilesToDatabase.py*”

***The End!***