I. ABOUT

Author name: Maciej Tomaszewski

Create date: 02.12.2023

Description: my first Python & SQL project, this file describes its general functionality

Tech-stack: Python, Microsoft SQL Server, draw.io, Git, GitHub, Microsoft Word

Attachments:

1. SQL scripts: MT_PythonSQL_P1_sql1.sql, MT_PythonSQL_P1_sql2.sql, MT_PythonSQL_P1_sql3.sql

2. Python script: MT_PythonSQL_P1_py1.py

3. Flat files: etl1_customers.txt, etl1_sales.csv

II. INTRO

The purpose of the project is to process data:

1. Extract from 2 external flat files saved on local PC

2. Transform – combine with data available in databases

3. Load into separate database tables

This is the first project I prepared using Python and SQL. Its main purpose was to learn how it can interact with data and process them. Few dozens of hours were spent for completing it. I faced plenty of problems and errors when doing so. It definitely boosted my knowledge in field of using Python to work with data.

III. HOW PROJECT WORKS

A flow chart with detailed description of each project step is attached on the next pages.

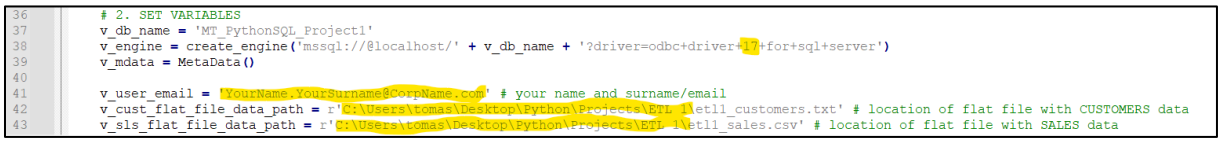
The projects is divided into two main sections:

1. User activities before execution of Python script

2. Execution of Python script

First section includes activities to perform before testing the solution:

- Installing necessary software
- Downloading all project components
- Creating test data set via executing database scripts
- Changing following variables in Python script where highlighted below:
    - **v_engine** – to 16 if MS SQL Server version is older than 2022
    - **v_user_email** – to email of person testing the solution, it will work with default email as well
    - **v_cust_flat_file_data_path** and **v_sls_flat_file_data_path** – for location where flat files are saved



```
# 2. SET VARIABLES
v_db_name = 'MT_PythonSQL_Project1'
v_engine = create_engine('mssql://@localhost/' + v_db_name + '?driver=odbc+driver+17+for+sql+server')
v_mdata = MetaData()

v_user_email = 'YourName.YourSurname@CorpName.com' # your name and surname/email
v_cust_flat_file_data_path = r'C:\Users\tomas\Desktop\Python\Projects\ETL 1\etl1_customers.txt' # location of flat file with CUSTOMERS data
v_sls_flat_file_data_path = r'C:\Users\tomas\Desktop\Python\Projects\ETL 1\etl1_sales.csv' # location of flat file with SALES data
```

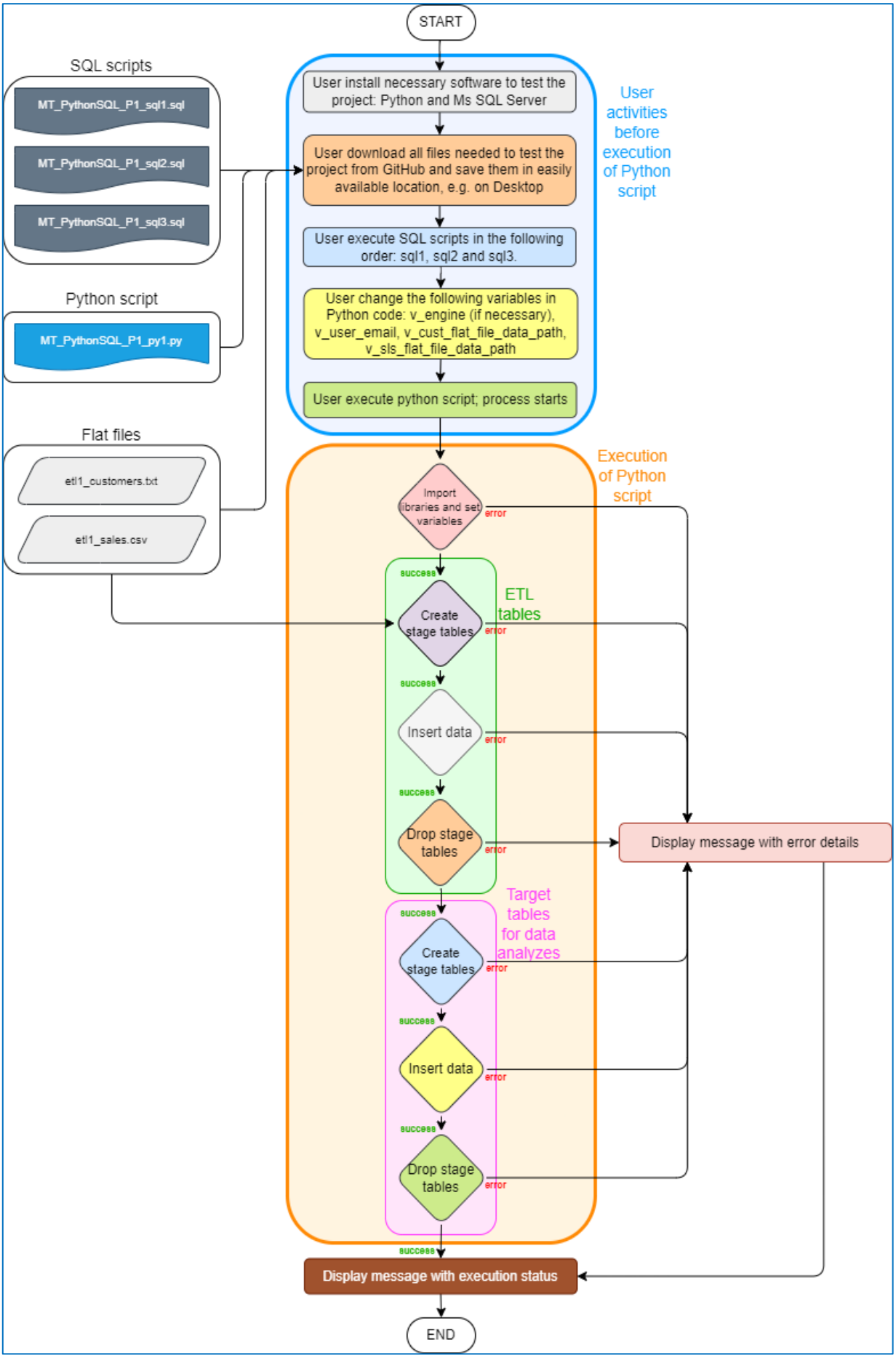
- Performing ETL process via executing Python script

Second section include primary part of a project – execution of Python script. It is divided into two main sub-sections:

2.1. ETL tables: tables used to combine data uploaded from flat files with data available in databases

2.2. Target tables for data analysis: tables used to store transformed, final data

Detailed description of each step is presented on a flow chart available on the next page.


START
SQL scripts
MT_PythonSQL_P1_sql1.sql
MT_PythonSQL_P1_sql2.sql
MT_PythonSQL_P1_sql3.sql
Python script
MT_PythonSQL_P1_py1.py
Flat files
etl1_customers.txt
etl1_sales.csv
User install necessary software to test the project: Python and Ms SQL Server
User download all files needed to test the project from GitHub and save them in easily available location, e.g. on Desktop
User execute SQL scripts in the following order: sql1, sql2 and sql3.
User change the following variables in Python code: v_engine (if necessary), v_user_email, v_cust_flat_file_data_path, v_sls_flat_file_data_path
User execute python script; process starts
User activities before execution of Python script
Execution of Python script
Import libraries and set variables
error
success
ETL tables
Create stage tables
Insert data
Drop stage tables
Display message with error details
Target tables for data analyzes
Create stage tables
Insert data
Drop stage tables
Display message with execution status
END