Pipeline Documentation

# 1/ Installation & Run

* Requirements
* Python 3.9+
* SQL Server (using ODBC Driver 17)
* File:
  + pipeline\_mssql.py
  + Data Engineer - Round 2 Task.xlsx
  + schema.sql – Script to create database
  + query.sql – Script to create store procedure to query and execute report

## Create virtual environment

Open Command Prompt at the project folder location

python -m venv .venv

## Activate Virtual Environment

.venv\Scripts\Activate

## Install usage libraries

pip install pandas numpy sqlalchemy pyodbc openpyxl

## Run SQL Script

## Run pipeline

python pipeline\_mssql.py

## Execute report

### Execute report - DESTINATION\_device\_report\_template

EXEC dbo.sp\_Load\_Device\_Report @MonthYear = '2025/06';  
GO  
SELECT \* FROM DESTINATION\_device\_report\_template  
GO

### Execute report - DESTINATION\_daily\_report\_template

EXEC dbo.sp\_Load\_Daily\_Report @StartDate = '2025-06-01', @EndDate = '2025-06-30';  
GO  
SELECT \* FROM DESTINATION\_device\_report\_template  
GO

# 2/ Mapping between raw data and tables's columns

## First Step:

Data is pulled from Excel sheet into raw data table (No transformation)

* SOURCE\_GGAds data → raw\_ggads
* SOURCE\_DV360 data → raw\_dv360
* SOURCE\_One data → raw\_one

## Second Step:

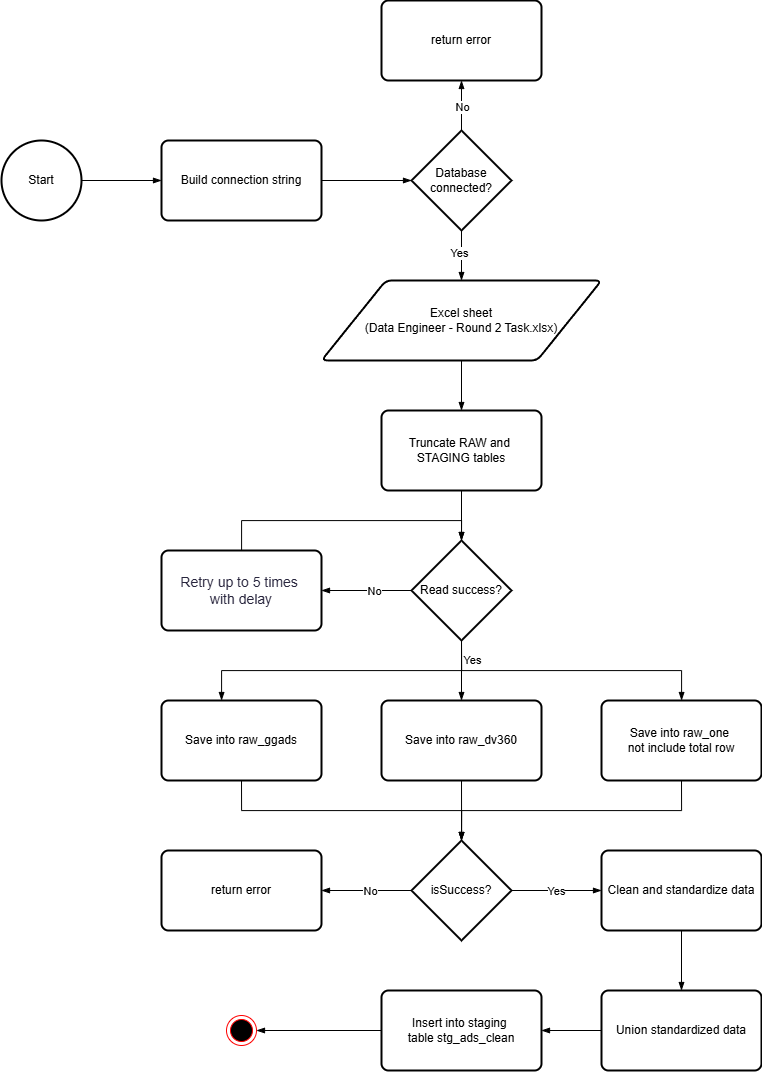
Data is cleaned and standardized before loading into staging

|  |  |  |
| --- | --- | --- |
| Column Name | Raw | Transformation |
| date | * GGAds: Day * DV360: Date * ONE: Date | Parse date to YYYY-MM-DD |
| platform | Fixed per sheet: GGADS, DV360, ONE | GGADS/ DV360/ ONE |
| advertiser\_account | * GGAds: Account name * DV360: Advertiser * ONE: Advertiser | Trim whitespace, allow NULL |
| campaign | * GGAds: Campaign * DV360: Campaign * ONE: Campaign | Trim whitespace |
| ad\_group | * GGAds: Ad group * DV360: NULL * ONE: NULL | Only available for GGAds |
| insertion\_order | * GGAds: NULL * DV360: Insertion Order * ONE: NULL | Only available for DV360 |
| line\_item | * GGAds: NULL * DV360: Line Item * ONE: NULL | Only available for DV360 |
| flight | * GGAds: NULL * DV360: NULL * ONE: Flight | Only available for ONE |
| creative | * GGAds: Ad name * DV360: Creative * ONE: Creative | Trim whitespace |
| device\_standardize | * GGAds: Device * DV360: Device Type * ONE: Device | Map values to Desktop, Phone, Tablet, Phone/Tablet, CTV or N/A if not existed |
| impressions | * GGAds: Impr. * DV360: Impressions * ONE: Impression | Convert string → int, remove commas, handle NULL as 0 |
| clicks | * GGAds: Clicks * DV360: Clicks * ONE: Click | Convert string → int, remove commas, handle NULL as 0 |

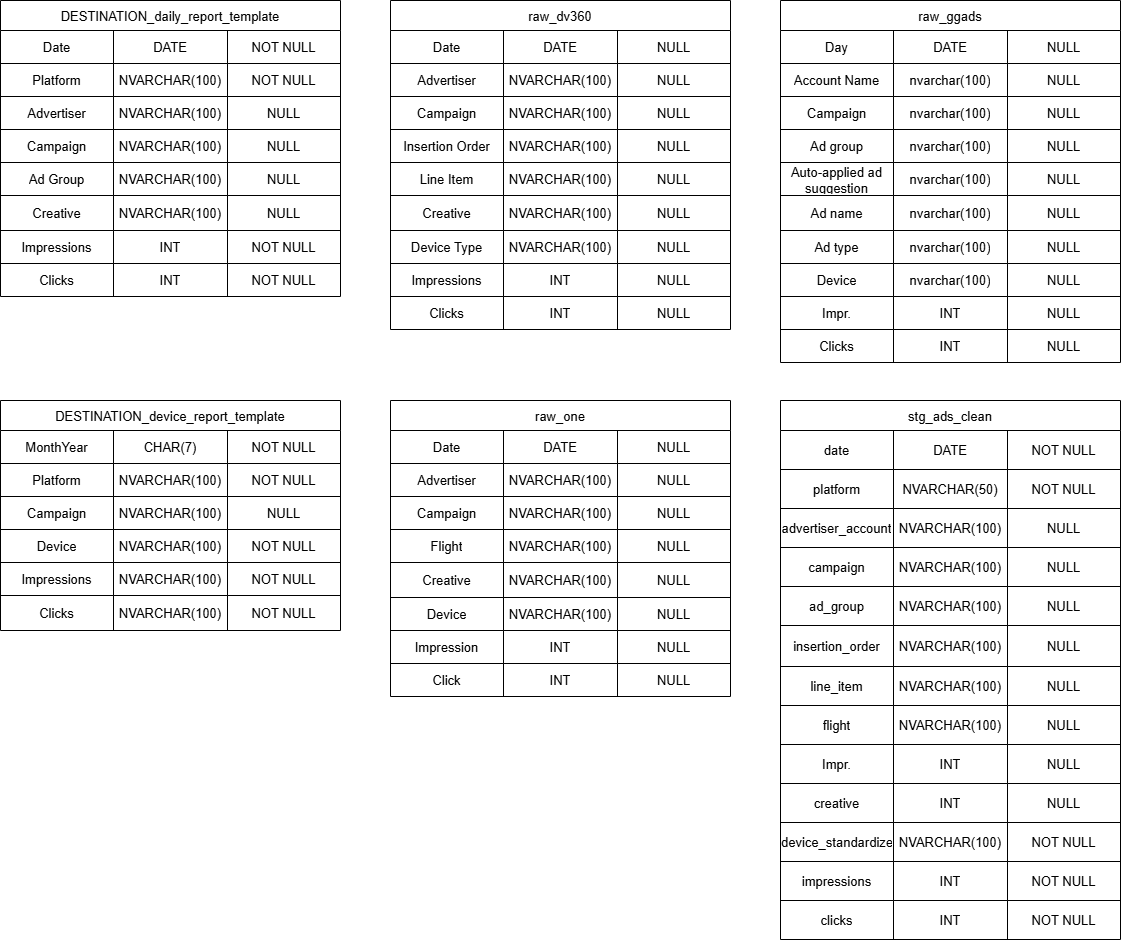
# 3/ Source code

*Attached*

# 4/ Flow chart



# 5/ Diagram chart



# 6/ Schema.sql

*Attached*

# 7/ Query.sql

*Attached*

# 8/ Final exported database

*Attached*