**1. Preparation of Database Schema**

The ETL process begins with the preparation of the database schema. This step ensures that the structure of the data warehouse is defined and optimized to accommodate the incoming datasets. Schemas, Source Table, Processing Table, database schema, and History Tables are created along with the necessary triggers.

A screenshot of a computer

Description automatically generated

**(DB setup.dtsx )**

**2. Data Merging**

A Python script is executed in this step, which is responsible for merging two datasets into a single cohesive dataset. This merged dataset is then prepared for loading into the source table. **(main.py file)**

**3. Data Loading into Source Table**

Initially, data is loaded into the source table. This step supports incremental loading to handle new and updated records efficiently. A dedicated package manages this process, ensuring that only the new data is ingested during each run.

**(Insert new records in source table.dtsx)**

A screenshot of a computer

Description automatically generated

**4. Staging Process**

When the data is ready to be loaded into the data warehouse schema, another package is executed. This package identifies and processes the newly added and modified records since the last execution. The process involves the following steps:

* **Control Table and Triggers:** A control table and triggers on the source table track the last update time of each record. This mechanism ensures that only records with recent changes are considered.
* **Staging Table:** The identified records are then loaded into the StagingTable. This staging area serves as temporary storage for data that will undergo further processing before being moved to the final schema.

A screenshot of a computer

Description automatically generated

**(Load New Records from source.dtsx)**

**5. Data Cleaning and Transformation**

Once the data is in the StagingTable, it undergoes a series of cleaning and transformation steps:

* **Handling Missing and Null Values:** Missing and null values are addressed to ensure data completeness.
* **Dropping Incorrect Records:** Records that do not meet predefined criteria are removed.
* **Units Transformation:** Data units are standardized to ensure consistency.

**(Data Preprocessing.dtsx)**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**6. Loading Data into the Schema**

After the data is cleaned and transformed, it is loaded from the StagingTable into the final schema of the data warehouse. This step integrates the processed data into the main analytical framework.

**7. Control Table Update**

Post-loading, the control table is updated to reflect the latest data load timestamp. This update prevents the reprocessing of the same records in subsequent ETL runs, ensuring efficiency and data integrity.

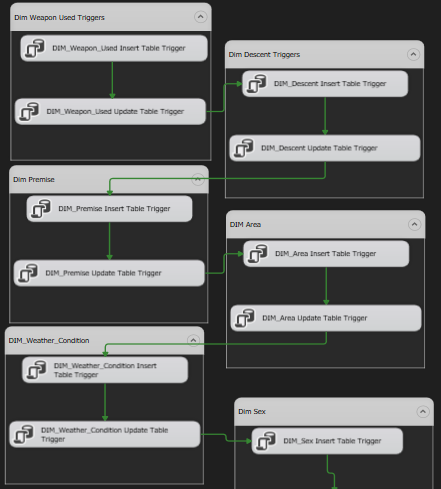
**8. Implementation of Slowly Changing Dimensions (SCD)**

The data warehouse schema incorporates Slowly Changing Dimensions (SCD) Type 4 to maintain a comprehensive history of changes:

* **SCD Type 4:** This type was chosen to preserve the historical data without influencing the current data analysis results, a challenge often associated with SCD Type 2.

**(Prepare History Tables Triggers.dtsx)**

**(History Tables Creation.dtsx)**



P.S The entire ETL process structure is encapsulated within a Visual Studio project,