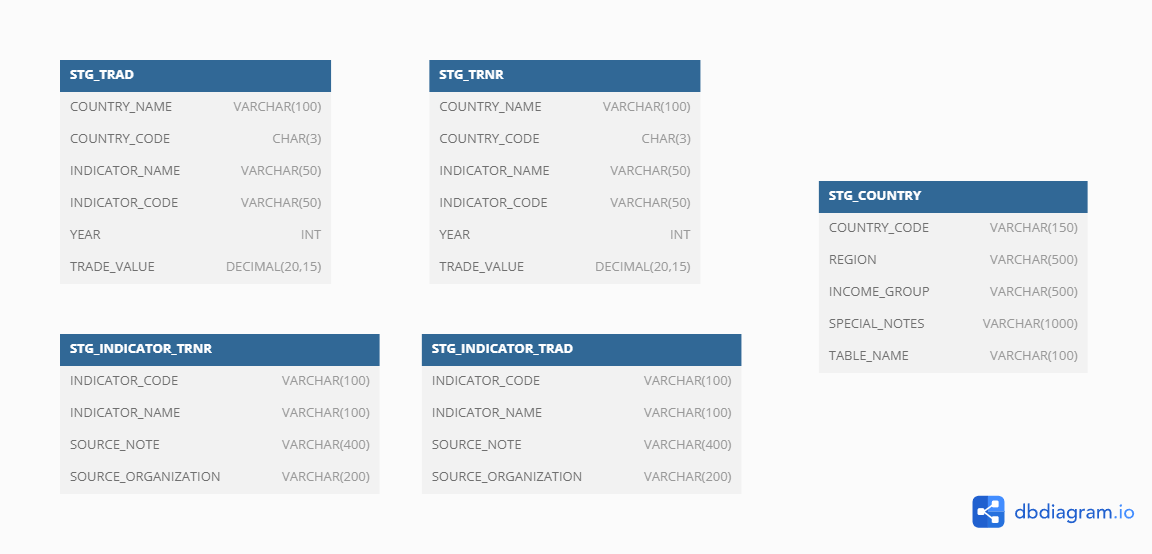
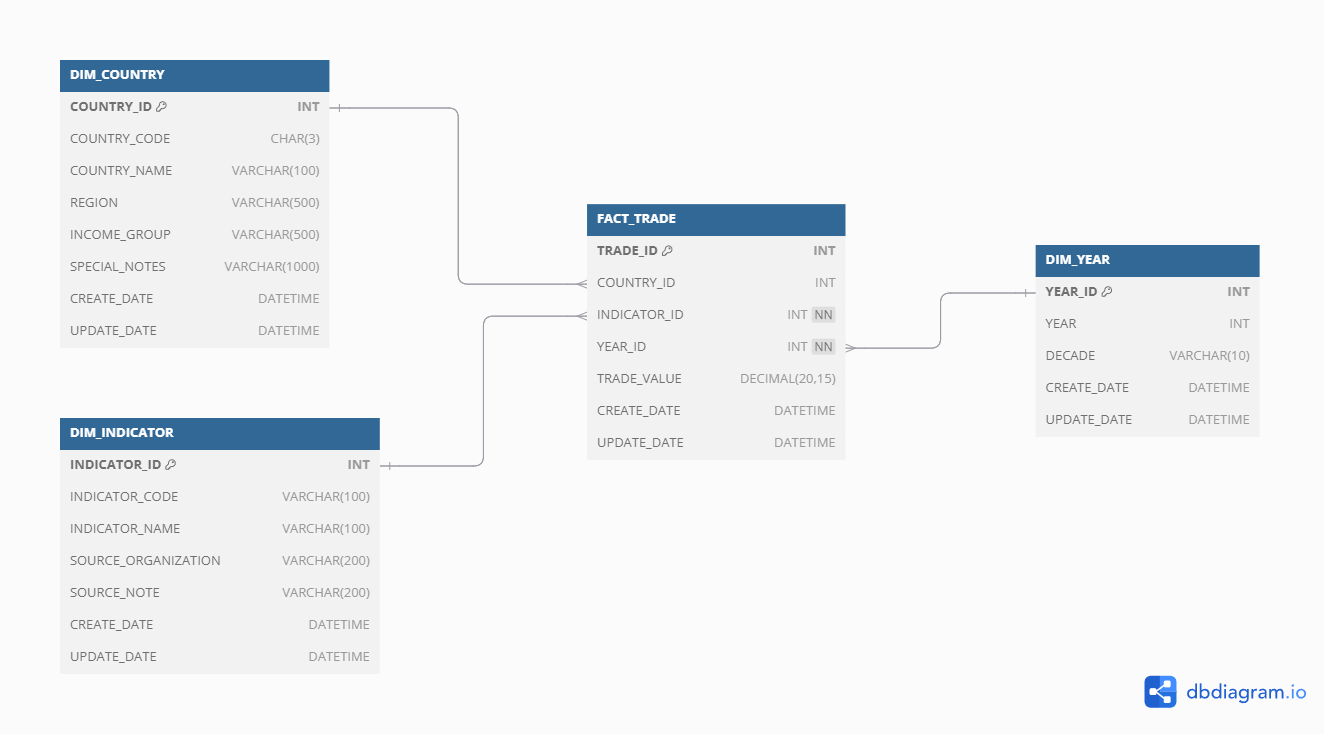
**Files and Description –**

1. CM\_MKT\_TRAD\_GD\_ZS.csv – Contains the values for stocks traded as total value (% of GDP) for the given countries for the time period 1960 to 2022.
2. CM\_MKT\_TRNR.csv – Contains the values for stocks traded as turnover ratio of domestic shares (%) for the given countries for the time period 1960 to 2022.
3. Country\_CM\_MKT\_TRAD\_GD\_ZS.csv – Metadata about the given countries.
4. Country\_CM\_MKT\_TRNR.csv – Metadata about the given countries.
5. Indicator\_CM\_MKT\_TRAD\_GD\_ZS.csv – Metadata about the given indicator CM\_MKT\_TRAD\_GD\_ZS.
6. Indicator\_CM\_MKT\_TRNR.csv – Metadata about the given indicator CM\_MKT\_TRNR.

**Staging Design –**

****

**Warehouse Design –**

****

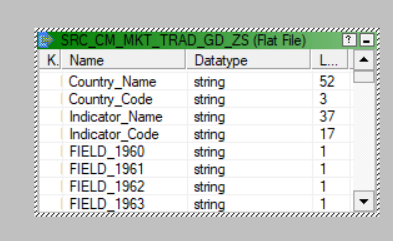
**Mappings Explained –**

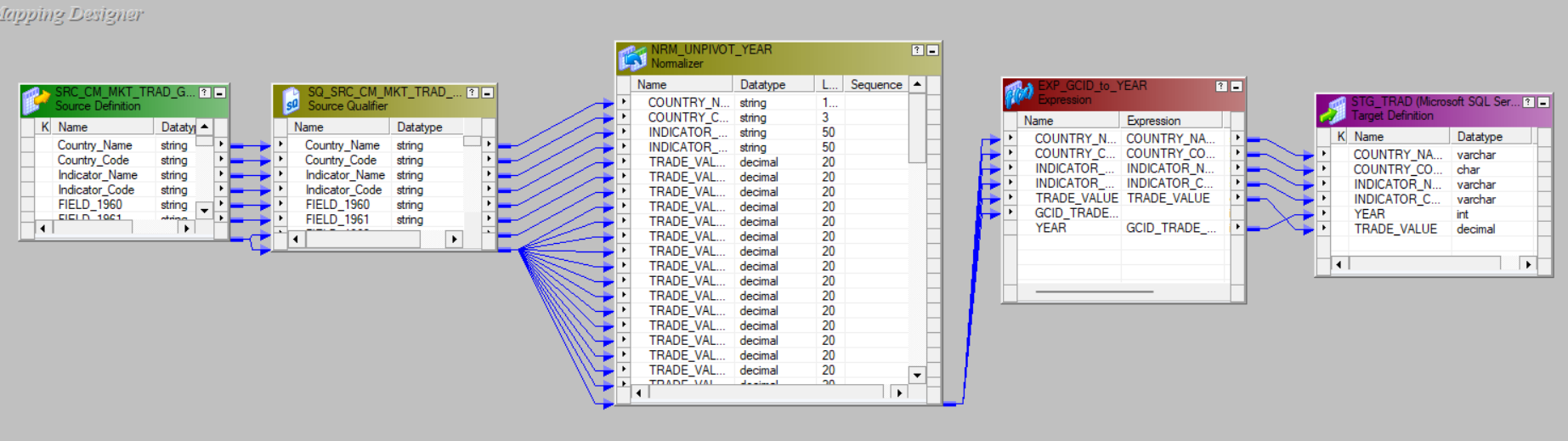
**1. CM\_MKT\_TRAD\_GD\_ZS.csv to STG\_TRAD:**

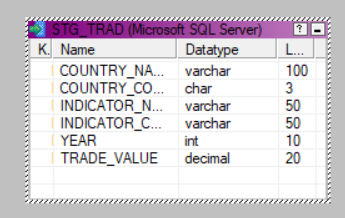
The CM\_MKT\_TRAD\_GD\_ZS.csv file contains year columns that need to be unpivoted. We achieve this by using the Normalizer transformation, which unpivots the year columns into a normalized format, making each row represent a single year’s trade data for each country.

We then use an Expression transformation to convert the GCID (Global Country Identifier) into the corresponding YEAR column.

This transformed data is then loaded into the STG\_TRAD staging table for further processing.

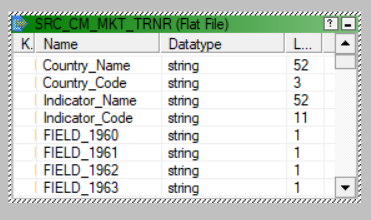


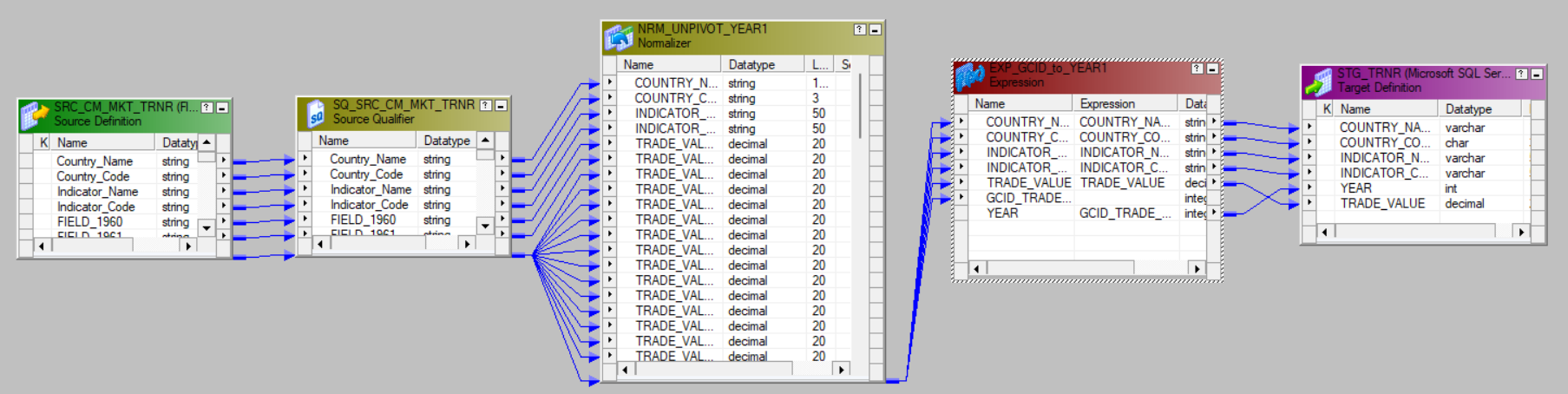


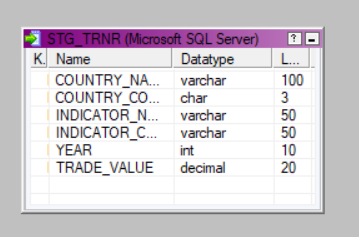


**2. CM\_MKT\_TRNR.csv to STG\_TRNR:**

The process for the CM\_MKT\_TRNR.csv file is identical to the CM\_MKT\_TRAD\_GD\_ZS.csv process. The year columns are unpivoted using the Normalizer, GCID is converted into the YEAR column using an Expression, and the data is loaded into STG\_TRNR.







**3. Country\_CM\_MKT\_TRAD\_GD\_ZS.csv to STG\_COUNTRY:**

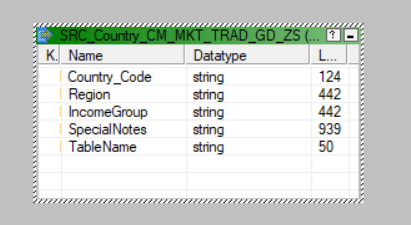
Upon analysis, we observe that the Country\_CM\_MKT\_TRAD\_GD\_ZS.csv and Country\_CM\_MKT\_TRNR.csv files are identical, so we choose to use Country\_CM\_MKT\_TRAD\_GD\_ZS.csv exclusively for creating the STG\_COUNTRY table.

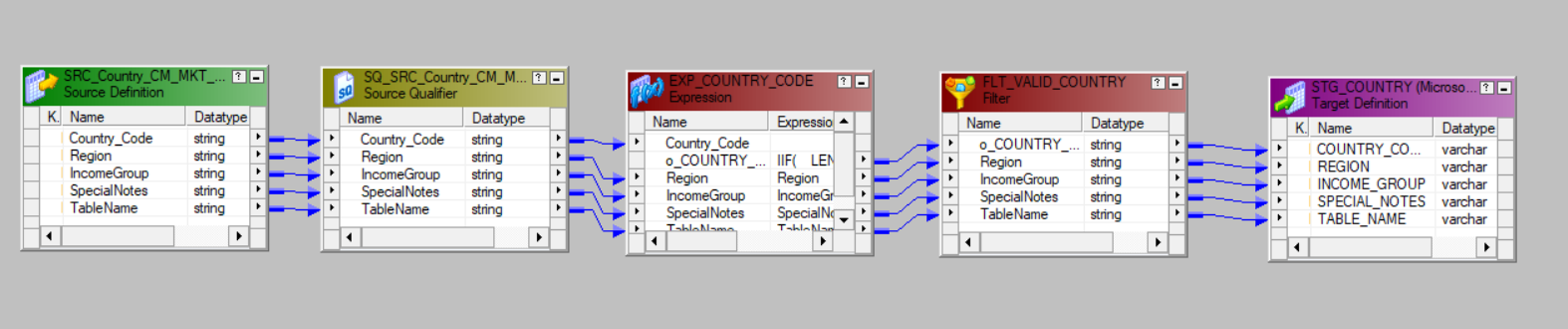
Due to the presence of multiline records in the Country\_CM\_MKT\_TRAD\_GD\_ZS.csv file, we configure the row delimiter as "\015 CR (\r)" in the source definition.

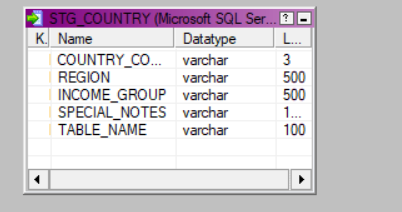
We utilize an Expression transformation to ensure the COUNTRY\_CODE is in the correct format, removing any unwanted characters such as leading/trailing spaces or quotes.

A Filter transformation is then used to ensure that no records with a NULL COUNTRY\_CODE are inserted into STG\_COUNTRY.

Finally, the cleaned data is loaded into the STG\_COUNTRY staging table.

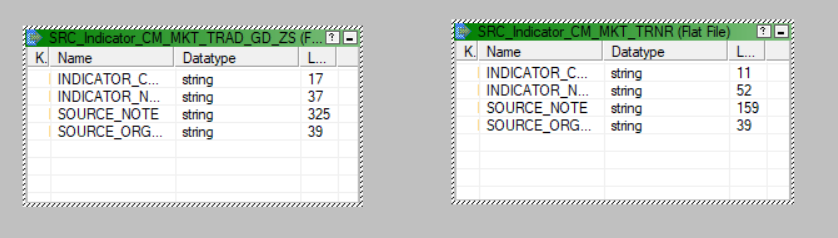


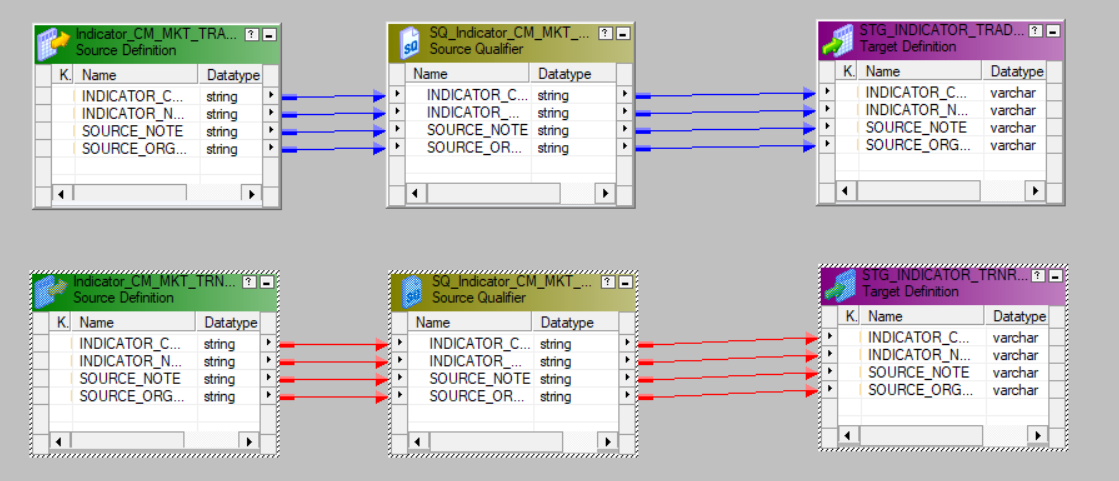


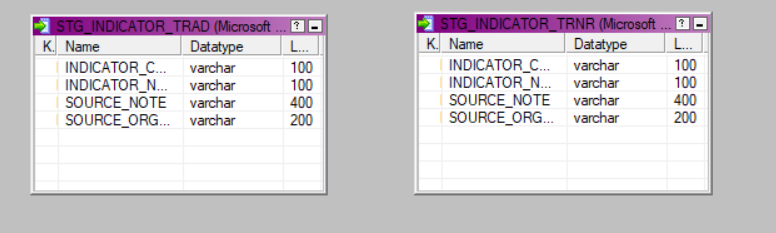


**4. Indicator\_CM\_MKT\_TRAD\_GD\_ZS.csv to STG\_INDICATOR\_TRAD and Indicator\_CM\_MKT\_TRNR.csv to STG\_INDICATOR\_TRNR:**

Indicator\_CM\_MKT\_TRAD\_GD\_ZS.csv is loaded into the STG\_INDICATOR\_TRAD table, while Indicator\_CM\_MKT\_TRNR.csv is loaded into STG\_INDICATOR\_TRNR.







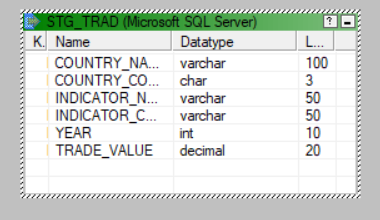
**5. From STG\_TRAD to DIM\_YEAR:**

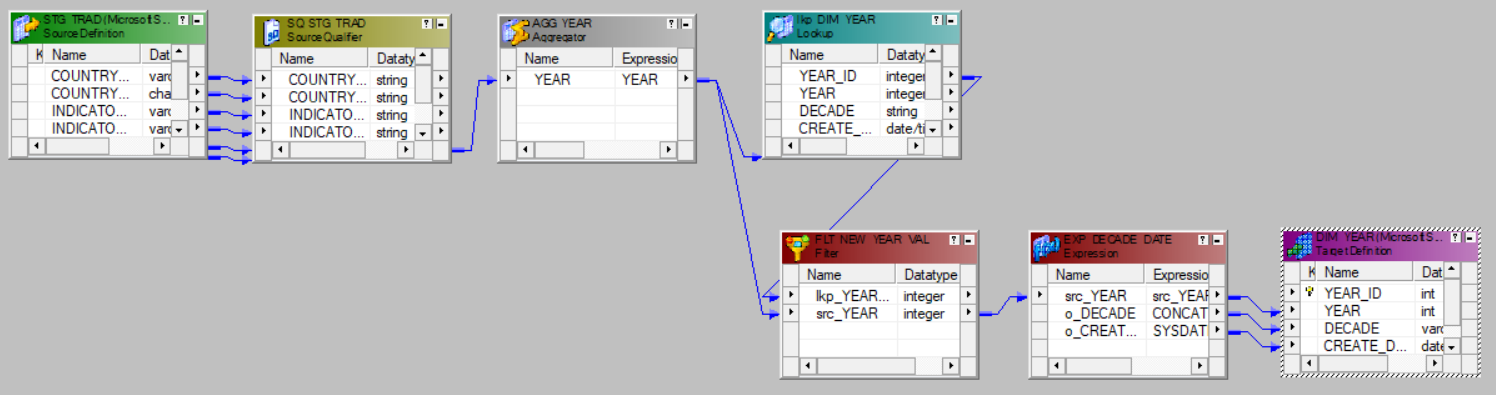
We extract the YEAR column from STG\_TRAD and use an Aggregator transformation to get unique year values. This ensures we do not have duplicate years in the system.

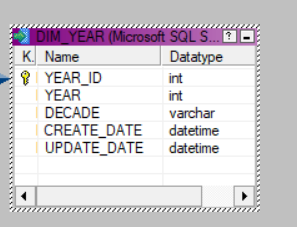
A Lookup transformation is performed on the DIM\_YEAR table to check if the year already exists.

If the year does not exist in DIM\_YEAR, an Expression transformation is used to derive the DECADE based on the YEAR.

The record is then loaded into the DIM\_YEAR table, along with the CREATE\_DATE and UPDATE\_DATE.







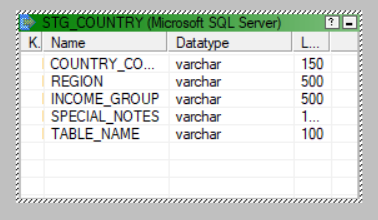
**6. From STG\_COUNTRY to DIM\_COUNTRY:**

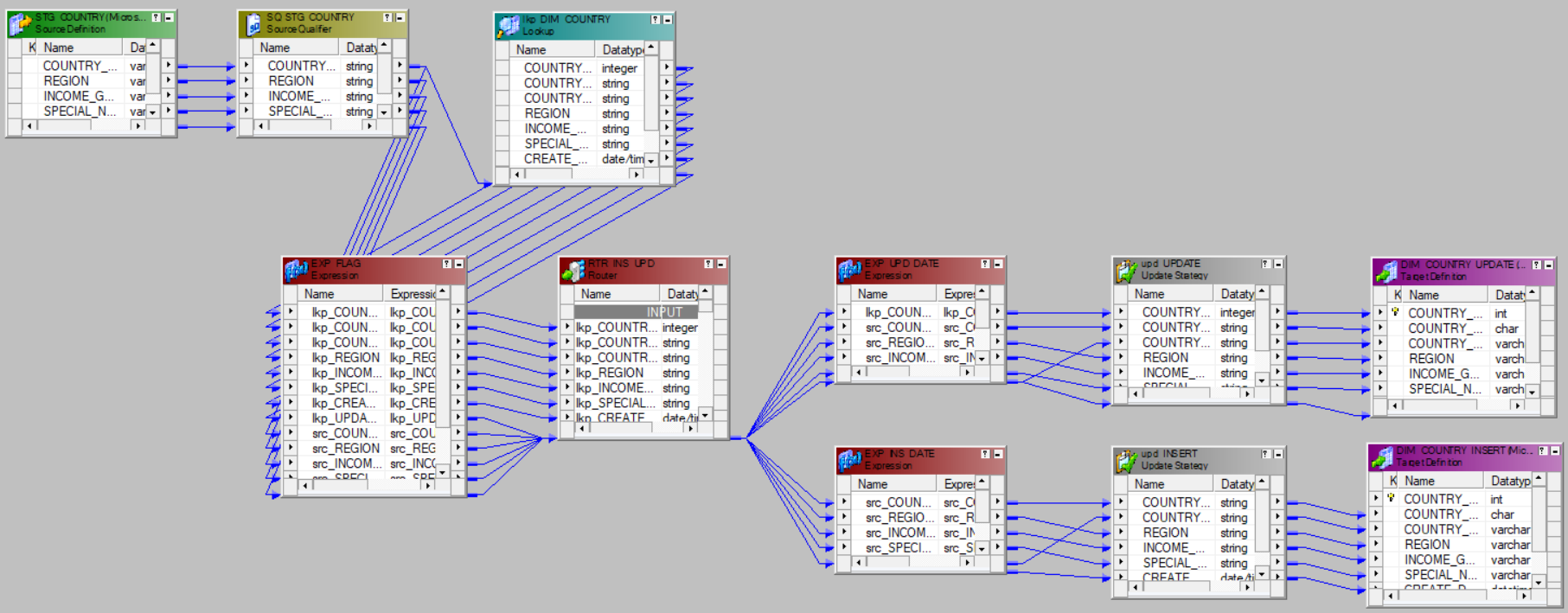
We perform a Lookup transformation on DIM\_COUNTRY to check if the record already exists.

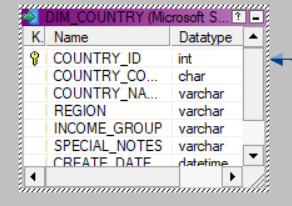
The Expression transformation is used to check whether the record should be inserted, updated, or deleted.

A Router transformation routes records based on whether they are new, require updates, or need to be deleted.

The Update Strategy transformation is used to specify whether the record should be inserted or updated, and the record is then processed into DIM\_COUNTRY.



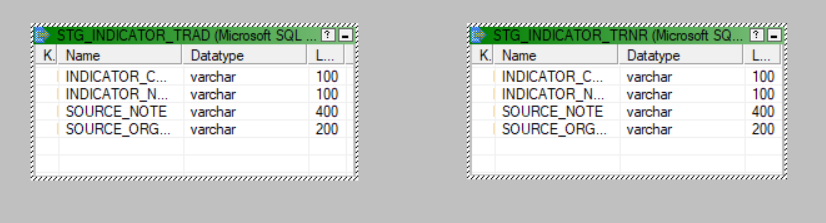


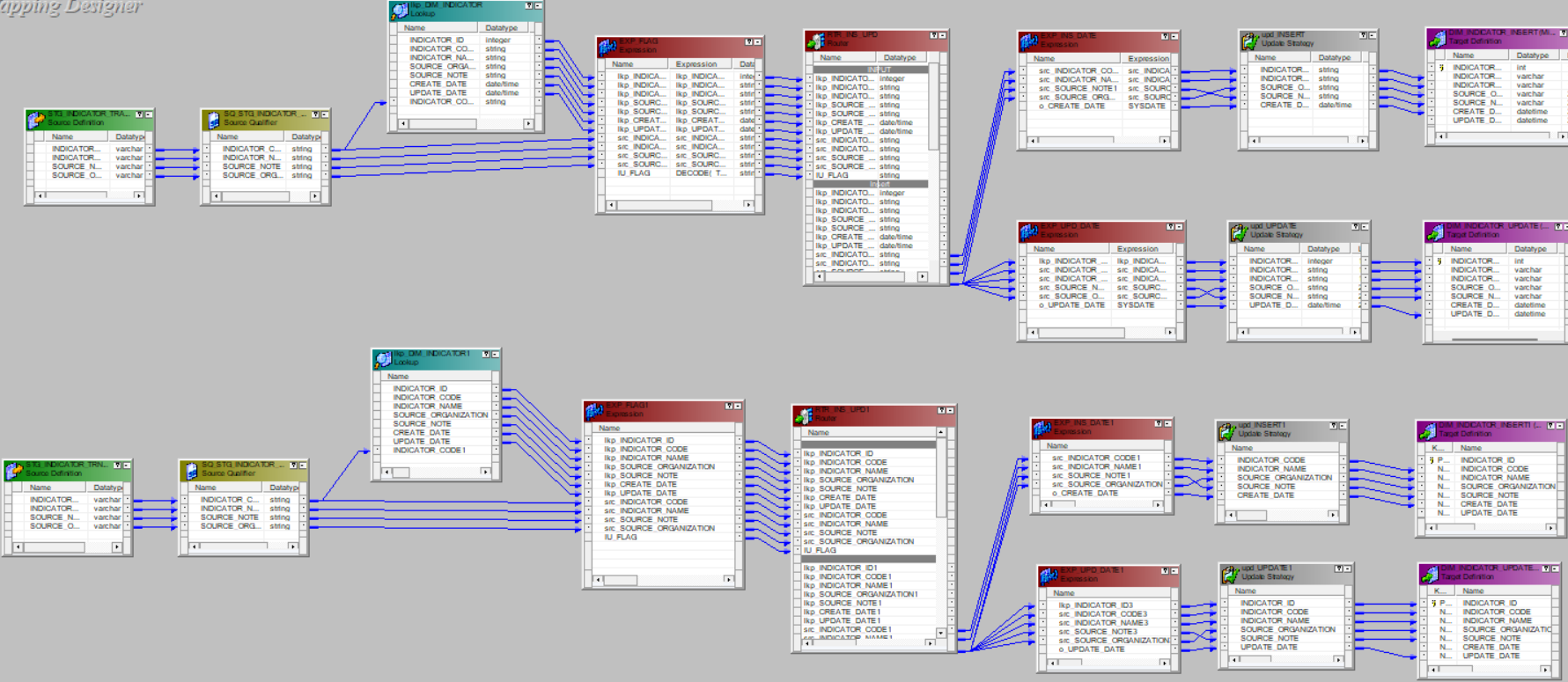


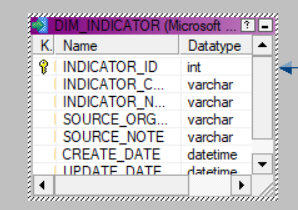
**7. From STG\_INDICATOR\_TRAD to DIM\_INDICATOR:**

The process for STG\_INDICATOR\_TRAD is similar to that of STG\_COUNTRY. We use Lookup to check for existing records in DIM\_INDICATOR, followed by Expression and Router transformations to manage new, updated, and deleted records, and then load the data into DIM\_INDICATOR.

The same process applies to STG\_INDICATOR\_TRNR, but it is inserted into the DIM\_INDICATOR table as well.







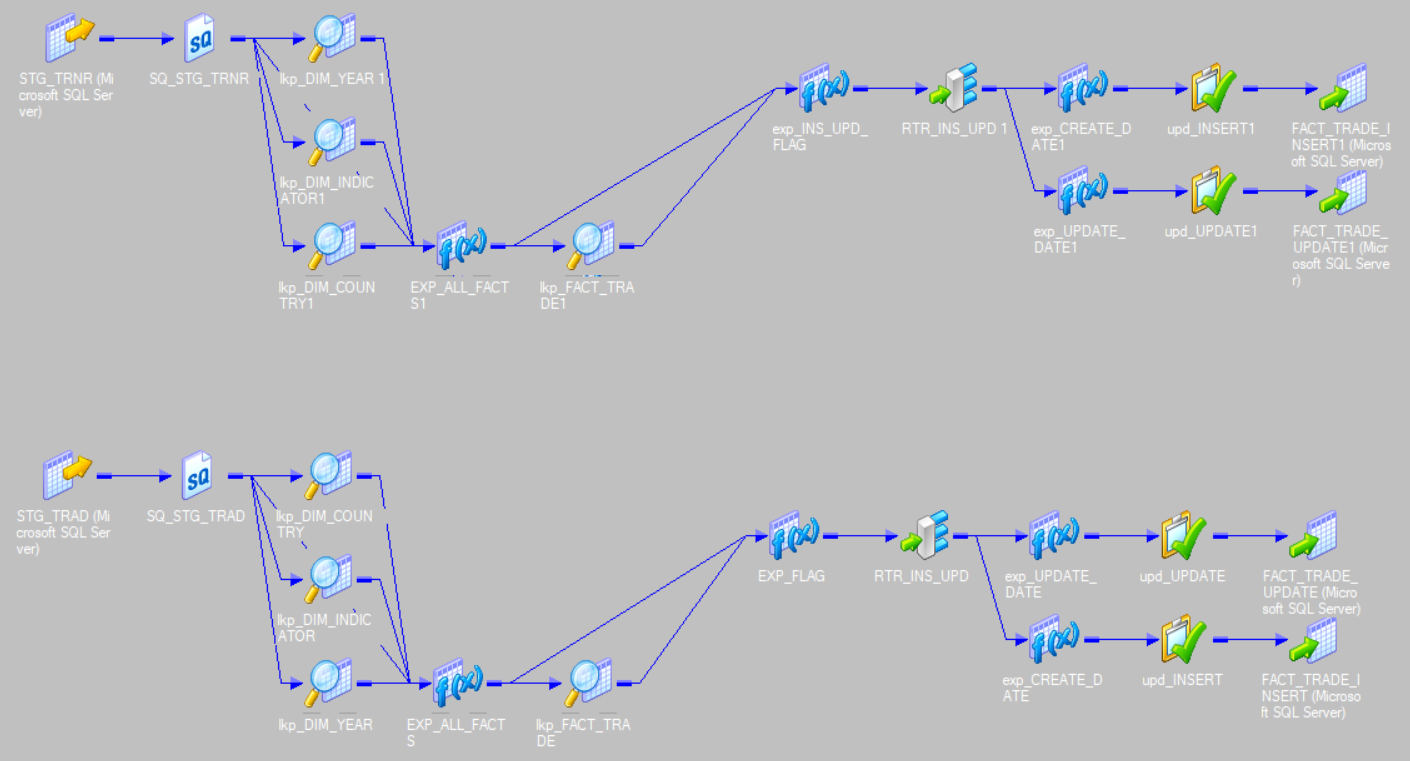
**8. From Staging to Fact Table (FACT\_TRADE):**

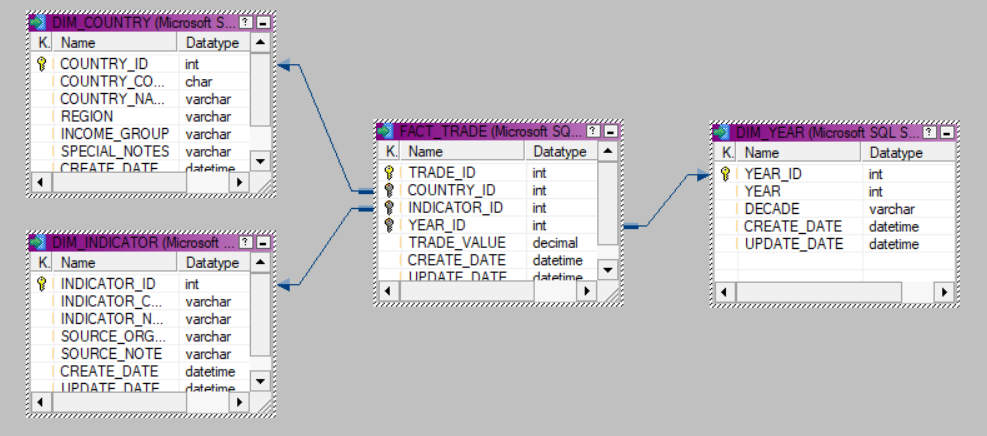
From STG\_TRAD to FACT\_TRADE:

* For STG\_TRAD, we perform Lookups on the DIM\_COUNTRY, DIM\_YEAR, and DIM\_INDICATOR tables to get the COUNTRY\_ID, YEAR\_ID, and INDICATOR\_ID, respectively.
* The TRADE\_VALUE from STG\_TRAD is then combined with the COUNTRY\_ID, YEAR\_ID, and INDICATOR\_ID to create a unique trade record.
* A Lookup transformation on the FACT\_TRADE table is performed to check if the record already exists based on the combination of COUNTRY\_ID, YEAR\_ID, and INDICATOR\_ID.
* An Expression transformation is used to determine whether the record should be inserted, updated, or deleted.
* The Router transformation is used to route records based on the action (insert, update, delete), and the Update Strategy transformation is used to specify the operation for each record.
* Finally, the transformed records are processed and inserted into the FACT\_TRADE table.

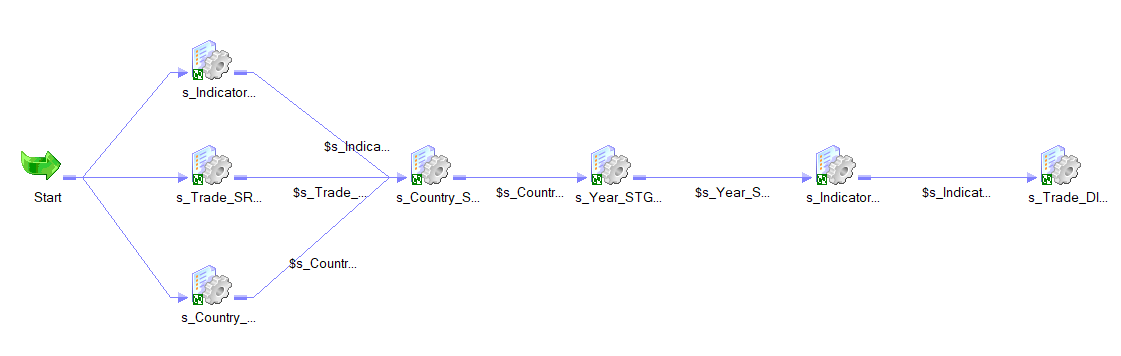
From STG\_TRNR to FACT\_TRADE:

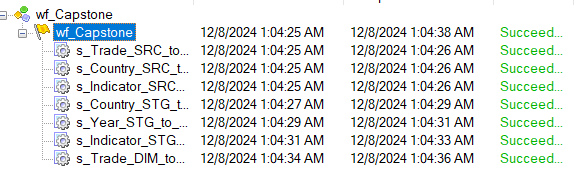
* The process for STG\_TRNR to FACT\_TRADE is similar to STG\_TRAD, with the same steps of performing Lookups on DIM\_COUNTRY, DIM\_YEAR, and DIM\_INDICATOR to get the relevant IDs.
* An additional Lookup on the FACT\_TRADE table is performed to check for existing records.
* Expression and Router transformations are used to handle inserts, updates, and deletes, followed by the Update Strategy transformation.
* Finally, records are inserted into FACT\_TRADE.

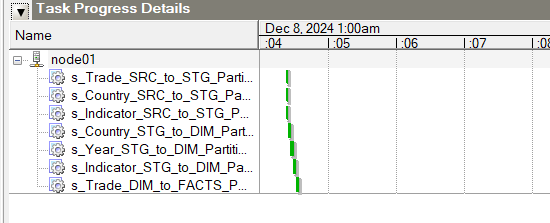


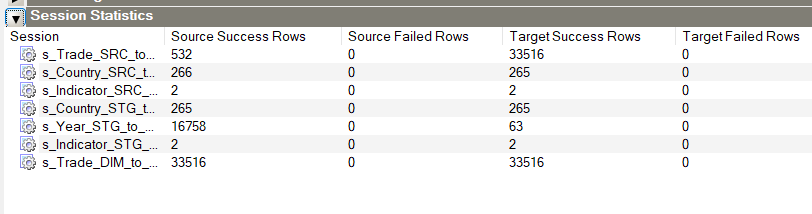


**Workflow Diagram –**









**Workflow Parameters –**

* Parameter File has been provided in the folder
* Parameter usage instances –

