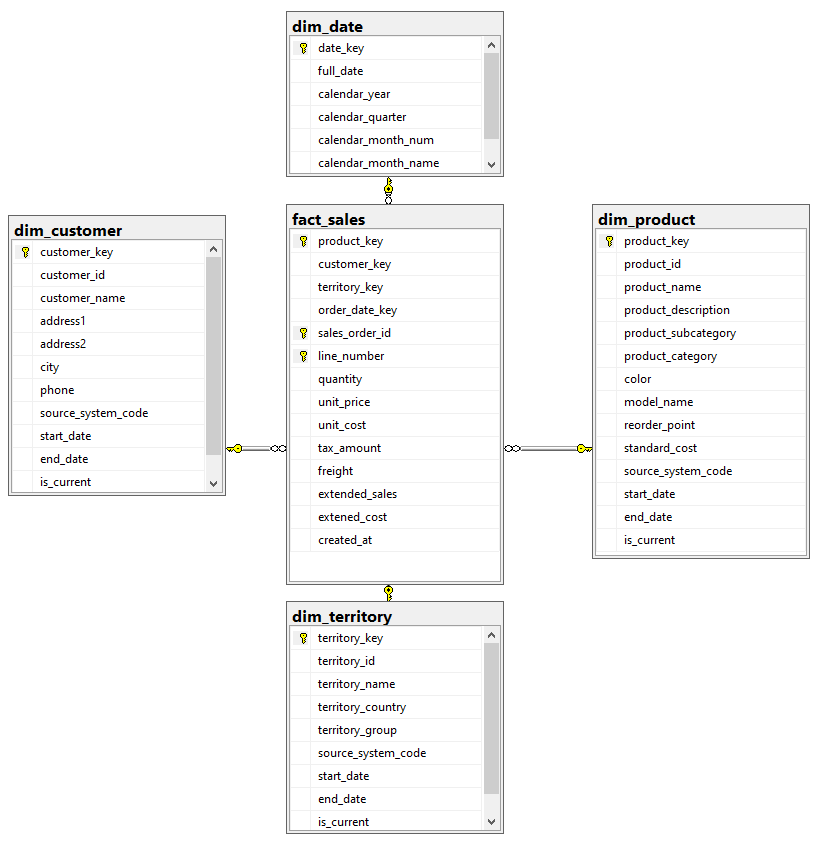
Kareem Adel Saad

https://github.com/KareemAdel10 || https://www.linkedin.com/in/kareem-adel-b76ab9201/

Sales Data Mart using SSis

# Introduction

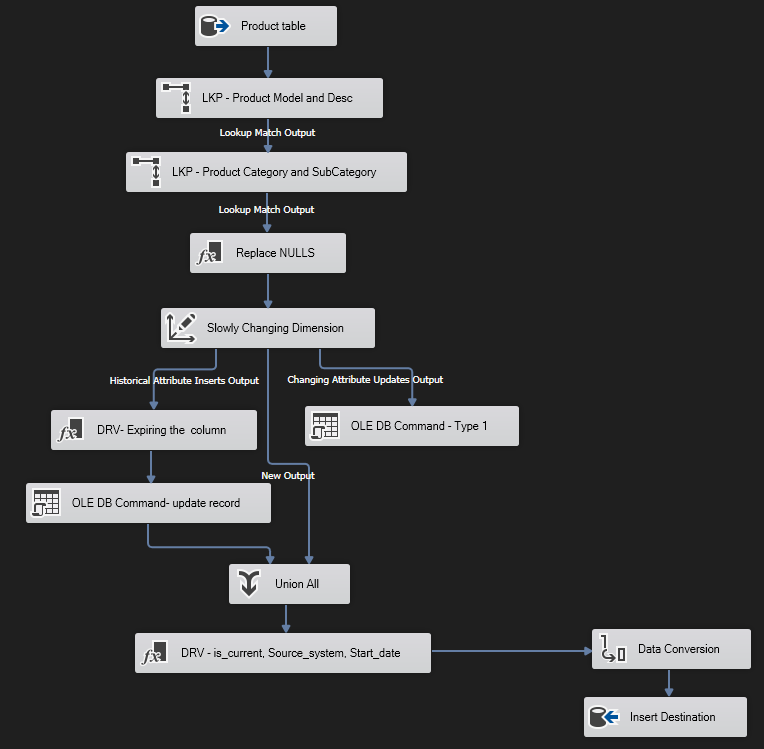
This document outlines a Sales Data Mart project developed with SQL Server Integration Services (SSIS) using the AdventureWorks2014 database. The objective was to create a scalable data mart for improved business intelligence and reporting.

The project involved extracting, transforming, and loading sales data into a star schema-based data mart. The central fact table, fact\_sales, contains key sales metrics, while the surrounding dimension tables categorize data by product, customer, territory, and time. This structure supports efficient, multidimensional analysis, enabling deeper insights into sales performance  
  


# Dimension Tables:

## Dim Product:

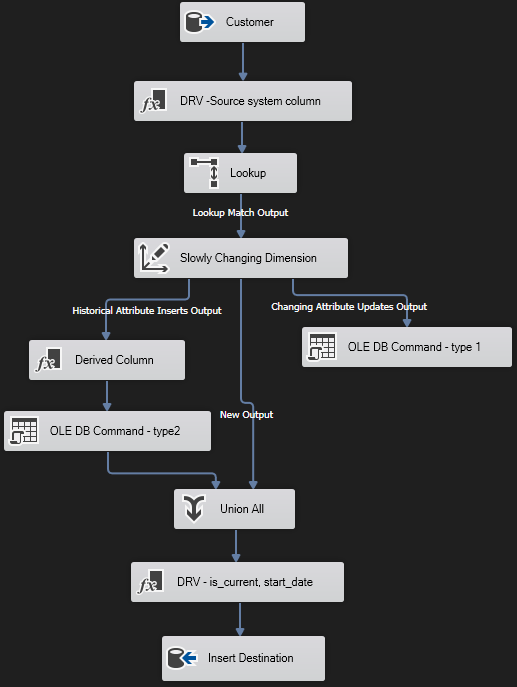
For the dim\_product dimension, I used the AdventureWorks2014 database as the source system. The ETL process began with multiple lookup steps to extract necessary columns, such as product details and categories. I then implemented a Slowly Changing Dimension (SCD) Type 2 mechanism to handle historical data, utilizing attributes like start\_date, end\_date, and is\_current to maintain versioned records of product information.

Following this, I applied transformations to align the input column data types with the destination schema requirements. Finally, the processed data was loaded into the dim\_product table, ensuring that it supports accurate historical tracking and reporting. This entire process is visualized in the attached SSIS data flow pipeline, highlighting each step from source extraction to data loading.

## Dim\_customer:

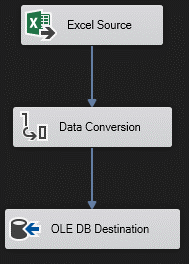
For the dim\_customer dimension, the ETL process started with the AdventureWorks2014 database as the source system. I first derived a source system column and performed a lookup to extract relevant customer data. The core of the process involved implementing a Slowly Changing Dimension (SCD) Type 2 mechanism. This ensured historical data accuracy by creating new records for changes using attributes like start\_date, end\_date, and is\_current.

The pipeline included managing updates through Type 1 commands and handling historical attribute inserts through Type 2. After deriving necessary columns and applying data transformations, I unified the outputs, performed final adjustments, and loaded the data into the dim\_customer table. The attached SSIS pipeline visualizes each step of this process, from data extraction to loading, highlighting the SCD implementation and its role in maintaining accurate customer data over time.



## Dim\_date:

For the dim\_date dimension, I extracted date-related data from an Excel source, applied necessary data type conversions, and loaded the transformed data into the dim\_date table within the data mart. This ensured proper integration of date dimensions for effective time-based analysis.



# Dim\_territory:

For the dim\_territory dimension, I loaded data from the AdventureWorks2014 database, applied necessary data conversions, and derived columns such as start\_date, is\_current, and source\_system. Additionally, I transformed territory\_country into its full name, performed further data conversion, and loaded the processed data into the destination table.



# Fact Table:

## Fact\_sales:

The ETL pipeline extracts sales data from the AdventureWorks2014 database using an OLE DB source, cleanses and enriches it with dimensions (product, customer, date, territory) through lookup transformations. The transformed data is loaded into an OLE DB destination. To optimize performance, an incremental load approach is employed, processing only new data since the last successful run. This is achieved by tracking a last\_loaded\_date variable.

