Sri Lanka Institute of Information Technology

Data Warehousing and Business Intelligence IT3021

Assignment 1

2025

Assignment 1 Report

Student Name – LAYATHMA B M A S

IT Number –IT22171542

Contents

•

1 Dataset Selection

1.1 Description

Dataset -Contoso(link to original dataset)

Description of the Data Set

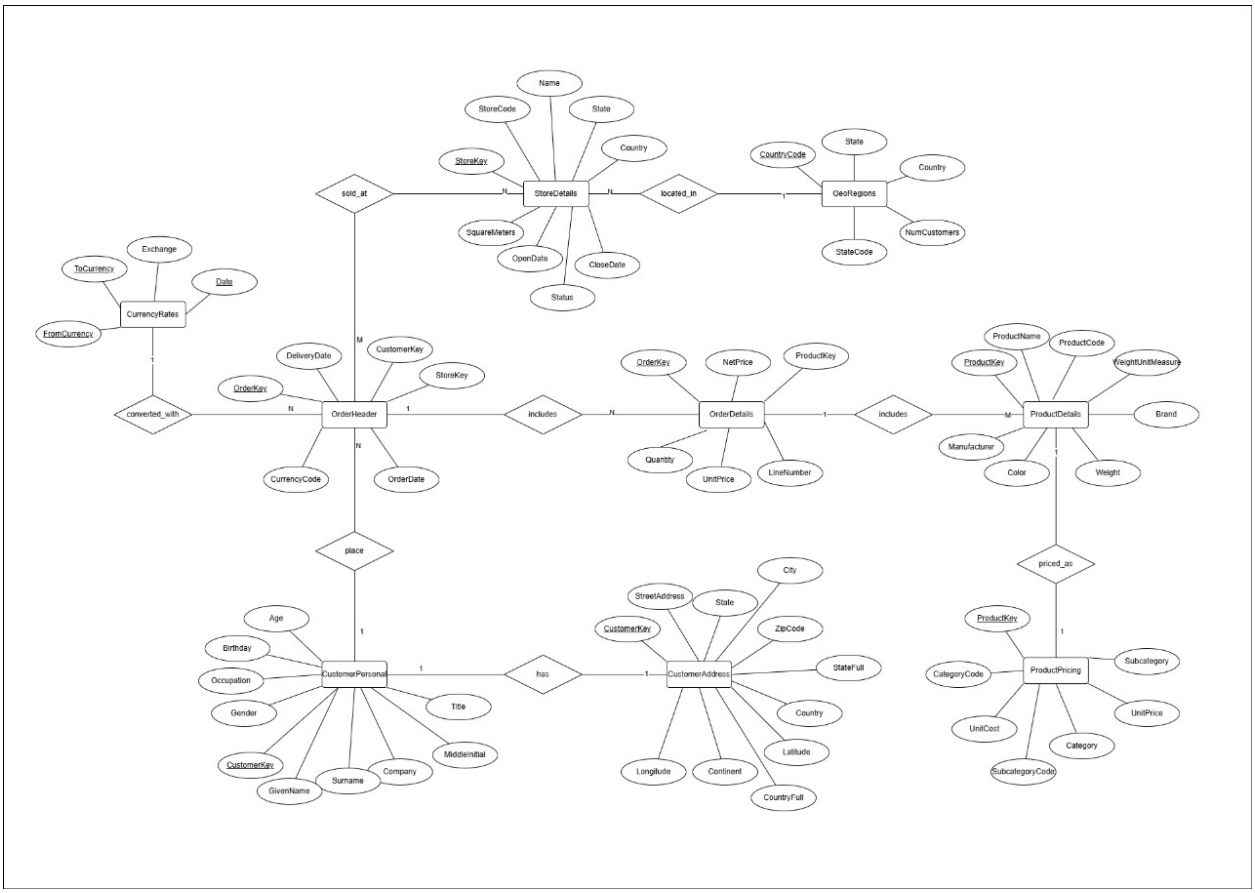
The **Retail Sales dataset** is designed to support the analysis and monitoring of activities related to an online retail business, focusing on sales transactions, customer behavior, product performance, and regional operations. This dataset provides a comprehensive view of the retail company's operations, capturing details about customers, products, orders, store locations, geographic regions, and currency exchange rates to facilitate international sales analysis.

The dataset spans over multiple years of sales transactions(2010-2020).The original dataset was sourced from a single transactional database, which has been edited, configured, and rearranged to suit the requirements of the project. To meet the assignment's need for multiple source types, the data has been split into three distinct sources: a SQL Server database (ContosoSourceDB), a text file (GeoRegions.txt), and a CSV file (CurrencyRates.csv).

The restructured dataset consists of the following seven data tables and two external files:

1. **CustomerPersonal** (in ContosoSourceDB): Stores unique information about each customer, such as ID, name, gender, age, and occupation.
2. **CustomerAddress** (in ContosoSourceDB): Records the address details of customers, including street, city, state, country, and geographic coordinates (latitude, longitude).
3. **ProductDetails** (in ContosoSourceDB): Contains details about products, such as ID, name, manufacturer, brand, color, and weight.
4. **ProductPricing** (in ContosoSourceDB): Records pricing and categorization details for products, including unit cost, unit price, category, and subcategory.
5. **StoreDetails** (in ContosoSourceDB): Provides information about store locations, including ID, name, country, state, size (square meters), opening/closing dates, and status.
6. **OrderHeader** (in ContosoSourceDB): Captures order metadata, such as order ID, customer ID, store ID, order date, delivery date, and currency used.
7. **OrderDetails** (in ContosoSourceDB): Records individual order line items, including order ID, sales order detail ID, product ID, quantity, and unit price.
8. **GeoRegions** (in GeoRegions.txt): A text file containing geographic region data, including country, state, and the number of customers in each region, used for regional sales analysis.
9. **CurrencyRates** (in CurrencyRates.csv): A CSV file storing historical currency exchange rates, including the date, from/to currency, and exchange rate, to support international sales calculations.

1.2 ER Diagram



2 Preparation of the Data Sources

Description of Data Sources

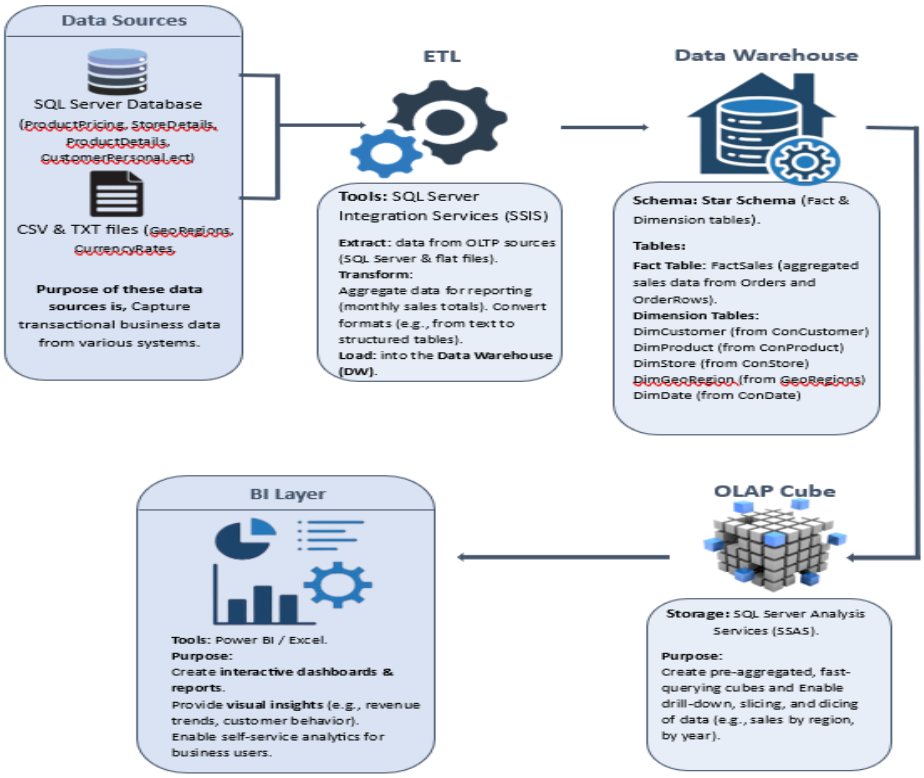
The dataset was restructured to include three source types (database, text, CSV), meeting the requirement of at least two types.

1. **ContosoSourceDB (SQL Server Database)**:
    - **CustomerPersonal**: Stores customer details (CustomerID, FirstName, LastName, Gender, Age, Occupation).
    - **CustomerAddress**: Contains address details (CustomerID, AddressLine1, City, State, Country, Latitude, Longitude).
    - **ProductDetails**: Holds product attributes (ProductID, ProductName, Manufacturer, Brand, Color, Weight).
    - **ProductPricing**: Includes pricing and category details (ProductID, UnitCost, UnitPrice, Category, Subcategory).
    - **StoreDetails**: Provides store information (StoreID, StoreName, Country, State, SquareMeters, Status).
    - **StoreDetails**: Stores order metadata (SalesOrderID, OrderDate, CustomerID, StoreID, DeliveryDate, Currency).
    - **OrderDetails**: Contains order line items (SalesOrderID, SalesOrderDetailID, ProductID, Quantity, UnitPrice).
2. **GeoRegions.txt (Text File)**:
    - **Format**: Tab-delimited.
    - **Columns**: Country (NVARCHAR(50)), State (NVARCHAR(50)), CustomerCount (INT).
    - **Purpose**: Provides geographic data for regional analysis.
3. **CurrencyRates.csv (CSV File)**:
    - **Format**: Comma-separated.
    - **Columns**: RateDate (DATE), FromCurrency (NVARCHAR(3)), ToCurrency (NVARCHAR(3)), ExchangeRate (FLOAT).
    - **Purpose**: Supports currency conversion for financial reporting.

3 Restructuring Process

The original dataset was a single database. To meet the requirement of multiple source types, geographic data (Country, State, CustomerCount) was extracted into GeoRegions.txt, and currency exchange rates were moved to CurrencyRates.csv. The remaining data was split into logical tables within ContosoSourceDB to separate customer, product, store, and order information. This restructuring enables complex ETL tasks (e.g., joins, SCD) and supports hierarchies (e.g., product Category → Subcategory).

4 Solution Architecture



5 Dataset Selection

1. **Source Systems**:
    - **ContosoSourceDB**: SQL Server database with customer, product, store, and order data.
    - **GeoRegions.txt**: Text file with geographic data (Country, State, CustomerCount).
    - **CurrencyRates.csv**: CSV file with currency exchange rates (RateDate, FromCurrency, ToCurrency, ExchangeRate).
    - **Role**: Provide raw OLTP data for the ETL process.

2. **ETL Pipeline**:
    - Implemented using SSIS packages (Contoso_Load_Staging.dtsx and Contoso_Load_DW.dtsx).
    - Extracts data from sources, applies transformations (e.g., joins, SCD Type-2, lookups), and loads into staging and data warehouse.
    - **Role**: Ensures data is cleaned, transformed, and loaded correctly.
3. **Staging Database (Contoso_Staging)**:
    - Temporary storage for raw data from sources.
    - Tables: StgCustomerPersonal, StgCustomerAddress, StgProductDetails, StgProductPricing, StgStoreDetails, StgOrderHeader, StgOrderDetails, StgGeoRegions, StgCurrencyRates.
    - **Role**: Simplifies transformations by consolidating data.
4. **Data Warehouse (Contoso_DW)**:
    - Dimensional model with dimensions (DimCustomer, DimProduct, DimStore, DimGeoRegion, DimCurrencyRate, DimDate) and fact table (FactSales).
    - **Role**: Supports analytical queries for reporting.
5. **BI Reporting**:
    - Tools: Power BI or Excel.
    - **Role**: Generates dashboards and reports (e.g., sales trends, customer demographics).

6 Data Warehouse Design and Development

Dimensional Model Description

The data warehouse uses a **star schema** to organize data into dimension and fact tables, optimized for analytical queries. The schema includes five dimension tables (including one SCD) and one fact table, meeting the assignment requirements.

Dimension Tables

1. **DimDate**:
    - **Source**: dbo.StgDate.

- **Columns**: DateKey (PK), Date, Year, YearQuarter, YearQuarterNumber, Quarter, YearMonth, YearMonthShort, YearMonthNumber, Month, MonthShort, MonthNumber, DayOfWeek, DayOfWeekShort, DayOfWeekNumber, WorkingDay, WorkingDayNumber, InsertDate, ModifiedDate.
- **Purpose**: Provides temporal context for sales transactions (e.g., OrderDate, DeliveryDate).
- **Key**: DateKey (surrogate key).

2. **DimCustomer (SCD Type 2)**:
   - **Source**: dbo.StgCustomerPersonal, dbo.StgCustomerAddress.
   - **Columns**: CustomerSK (PK, identity), AlternateCustomerID, Gender, Title, GivenName, MiddleInitial, Surname, StreetAddress, City, State, StateFull, ZipCode, Country, CountryFull, Birthday, Age, Occupation, Company, Vehicle, Latitude, Longitude, Continent, StartDate, EndDate, IsCurrent, InsertDate, ModifiedDate.
   - **Purpose**: Tracks customer information, with SCD Type 2 for address changes (e.g., StreetAddress, City).
   - **Key**: CustomerSK (surrogate key), AlternateCustomerID (business key).
   - **SCD Attributes**:
     - Historical (Type 2): StreetAddress, City, State, ZipCode, Country.
     - Changing (Type 1): Gender, Title, Occupation.
     - Fixed: GivenName, Surname, Birthday, etc.
3. **DimProduct**:
   - **Source**: dbo.StgProduct, dbo.StgProductPricing.
   - **Columns**: ProductSK (PK, identity), AlternateProductID, ProductCode, ProductName, Manufacturer, Brand, Color, WeightUnitMeasure, Weight, UnitCost, UnitPrice, SubcategoryCode, Subcategory, CategoryCode, Category, InsertDate, ModifiedDate.
   - **Purpose**: Stores product details and pricing, supporting hierarchies (Category > Subcategory).
   - **Key**: ProductSK (surrogate key), AlternateProductID (business key).
4. **DimStore**:
   - **Source**: dbo.StgStoreDetails.
   - **Columns**: StoreSK (PK, identity), AlternateStoreID, StoreCode, Country, State, Name, SquareMeters, OpenDate, CloseDate, Status, InsertDate, ModifiedDate.
   - **Purpose**: Represents retail stores where sales occur.
   - **Key**: StoreSK (surrogate key), AlternateStoreID (business key).
5. **DimGeoRegion**:
   - **Source**: dbo.StgGEORegions.
   - **Columns**: GeoRegionSK (PK, identity), AlternateGeoLocationID, CountryCode, Country, StateCode, State, NumCustomers, InsertDate, ModifiedDate.
   - **Purpose**: Provides geographic context for sales analysis.
   - **Key**: GeoRegionSK (surrogate key), AlternateGeoLocationID (business key).

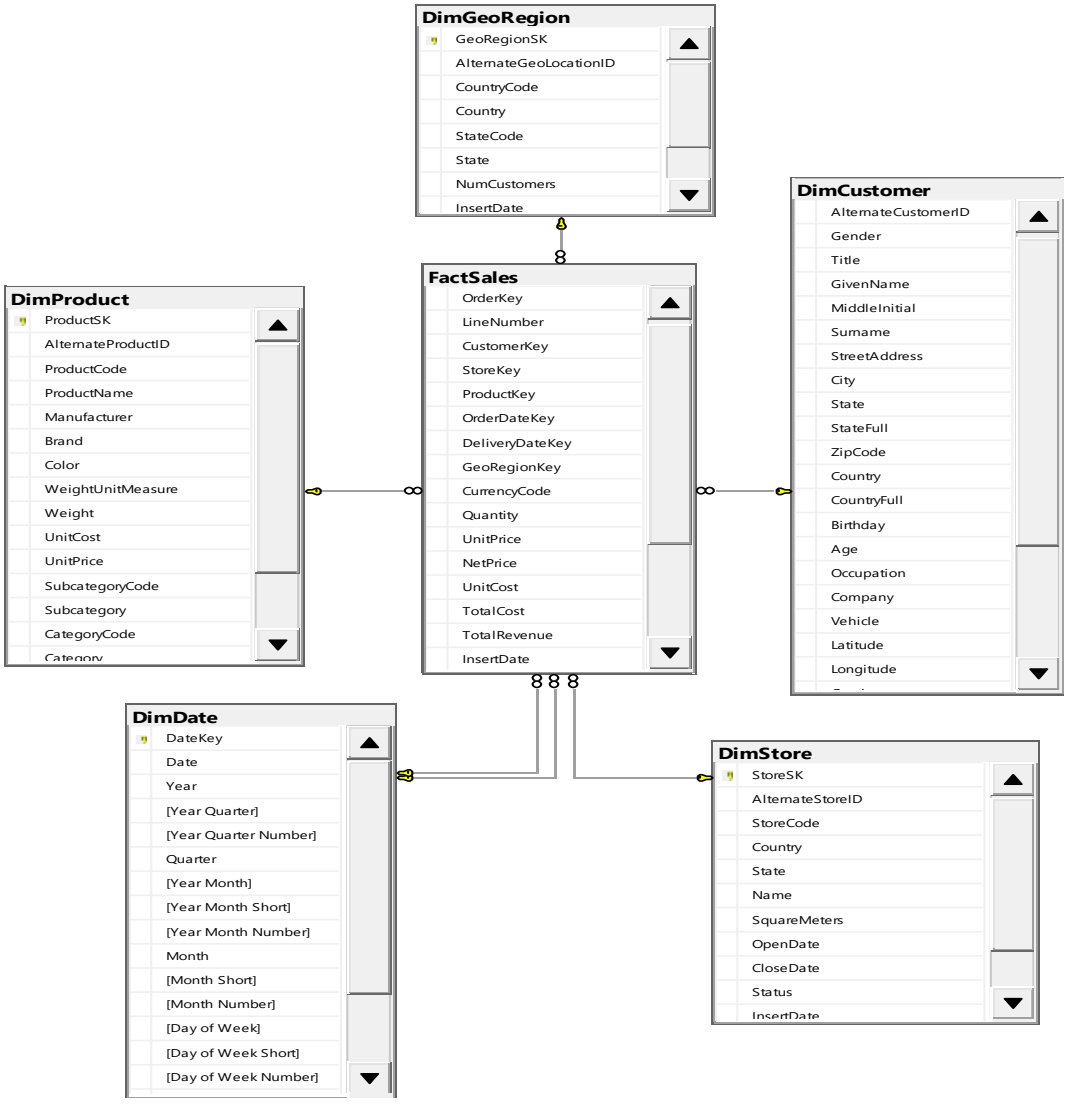
Fact Table

- **FactSales**:
  - **Source**: dbo.StgOrderHeader, dbo.StgOrderDetails.
  - **Columns**: SalesSK (PK, identity), OrderKey, LineNumber, CustomerKey (FK), StoreKey (FK), ProductKey (FK), OrderDateKey (FK), DeliveryDateKey (FK), GeoRegionKey (FK), CurrencyCode, Quantity, UnitPrice, NetPrice, UnitCost, TotalCost (computed), TotalRevenue (computed), InsertDate, ModifiedDate, accm_txn_create_time, accm_txn_complete_time, txn_process_time.
  - **Purpose**: Stores transactional sales data with measures (Quantity, UnitPrice, NetPrice, UnitCost, TotalCost, TotalRevenue).
  - **Keys**:
    - SalesSK (surrogate key).
    - Foreign Keys: CustomerKey (DimCustomer), StoreKey (DimStore), ProductKey (DimProduct), OrderDateKey (DimDate), DeliveryDateKey (DimDate), GeoRegionKey (DimGeoRegion).
  - **Measures**:
    - Quantity: Number of units sold.
    - UnitPrice, NetPrice, UnitCost: Pricing details.
    - TotalCost: Quantity * UnitCost.
    - TotalRevenue: Quantity * NetPrice.
    - txn_process_time: Difference in hours between accm_txn_create_time and accm_txn_complete_time.

The schema used : Star Schema

The **Star Schema** has been utilized in the dimensional modeling for the Contoso 100K_Staging dataset to optimize query performance and simplify data analysis. Unlike a snowflake schema, the dimension tables are denormalized, maintaining a direct connection to the central fact table, FactSales, to facilitate efficient querying. The dimension tables, including DimCustomer, DimProduct, DimStore, DimGeoRegion, and DimDate, are designed to store comprehensive descriptive attributes without further normalization, ensuring a straightforward structure.
It was assumed that the retail sales data, particularly attributes related to customers, products, stores, and geographic regions, would benefit from a denormalized structure to support rapid analytical queries. Therefore, the dimension tables contain hierarchical attributes to enable flexible categorization and analysis:

- **Country > State > City** for DimCustomer and DimGeoRegion, allowing geographic analysis of sales.
- **Category > Subcategory** for DimProduct, supporting product hierarchy-based reporting.
- **Year > Quarter > Month** for DimDate, enabling time-based analysis of sales trends.

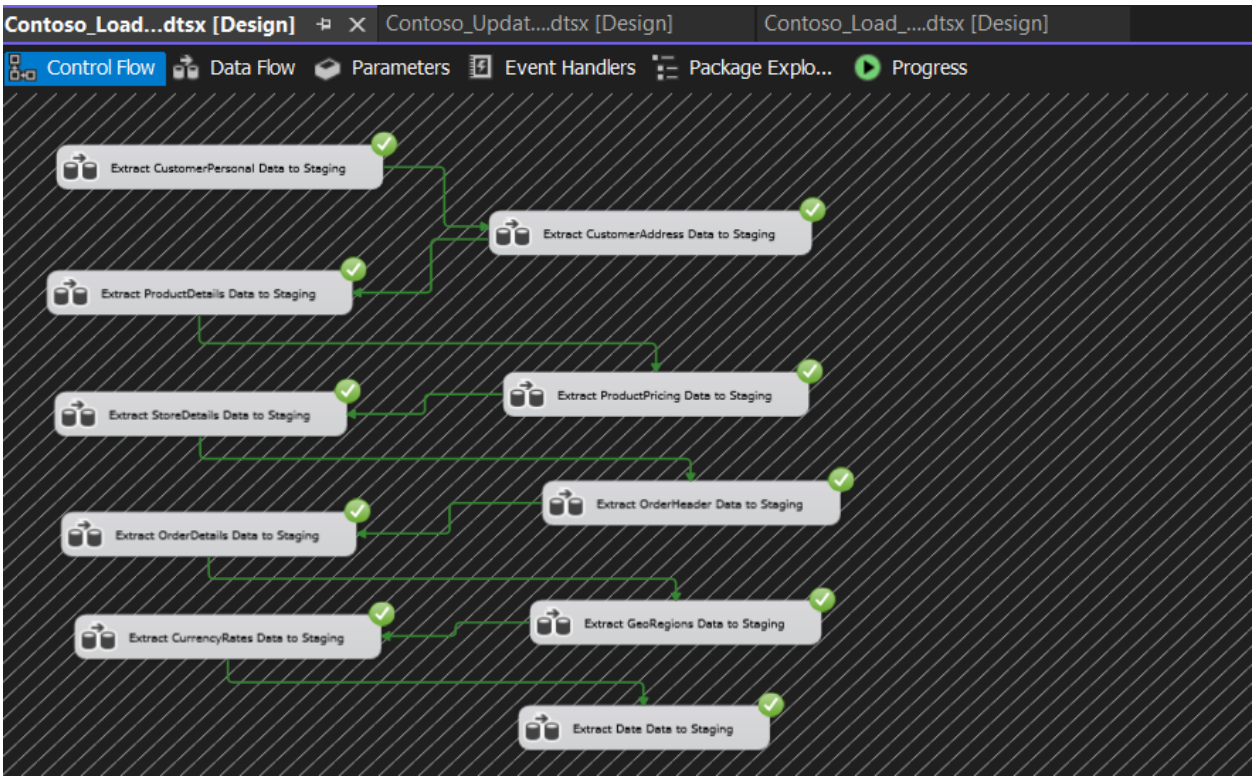

DimGeoRegion
GeoRegionSK
AlternateGeoLocationID
CountryCode
Country
StateCode
State
NumCustomers
InsertDate
DimProduct
ProductSK
AlternateProductID
ProductCode
ProductName
Manufacturer
Brand
Color
WeightUnitMeasure
Weight
UnitCost
UnitPrice
SubcategoryCode
Subcategory
CategoryCode
Category
FactSales
OrderKey
LineNumber
CustomerKey
StoreKey
ProductKey
OrderDateKey
DeliveryDateKey
GeoRegionKey
CurrencyCode
Quantity
UnitPrice
NetPrice
UnitCost
TotalCost
TotalRevenue
InsertDate
DimCustomer
AlternateCustomerID
Gender
Title
GivenName
MiddleInitial
Surname
StreetAddress
City
State
StateFull
ZipCode
Country
CountryFull
Birthday
Age
Occupation
Company
Vehicle
Latitude
Longitude
DimDate
DateKey
Date
Year
[Year Quarter]
[Year Quarter Number]
Quarter
[Year Month]
[Year Month Short]
[Year Month Number]
Month
[Month Short]
[Month Number]
[Day of Week]
[Day of Week Short]
[Day of Week Number]
DimStore
StoreSK
AlternateStoreID
StoreCode
Country
State
Name
SquareMeters
OpenDate
CloseDate
Status
InsertDate

7 ETL Development

7.1 Extract Data from Source to Staging

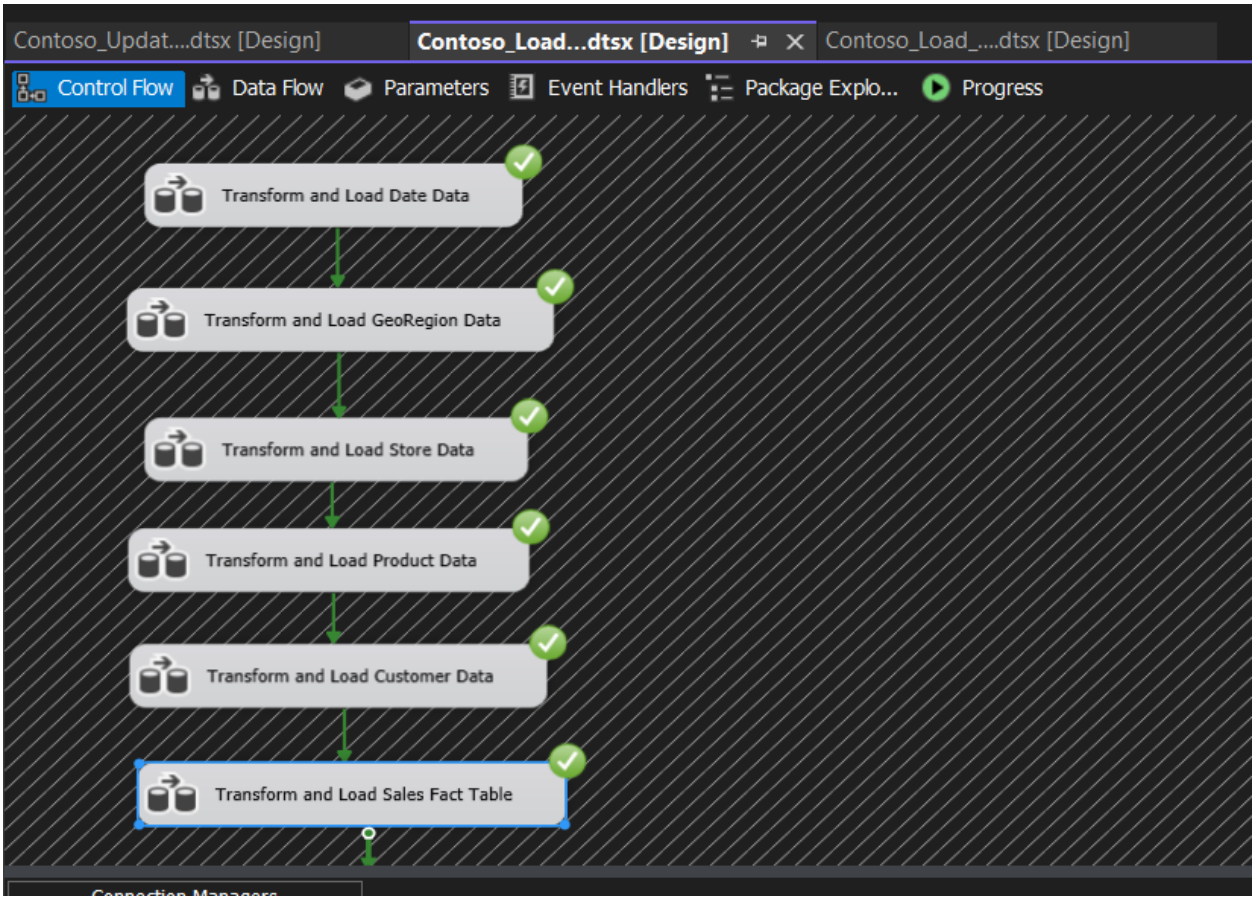
Execution order

.

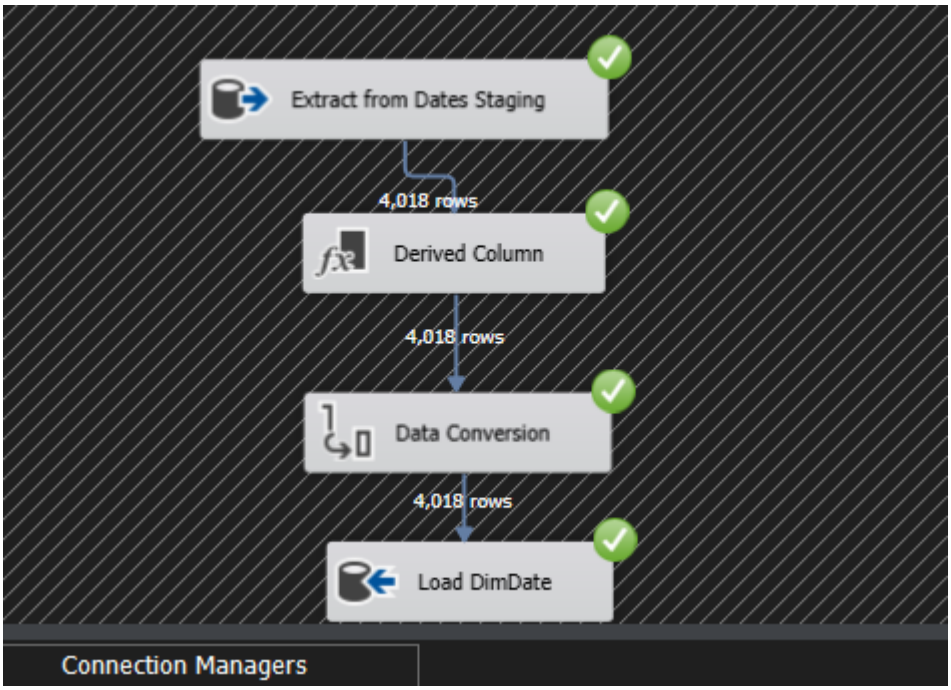


7.2 Loading the Transformed Data into the Data Warehouse

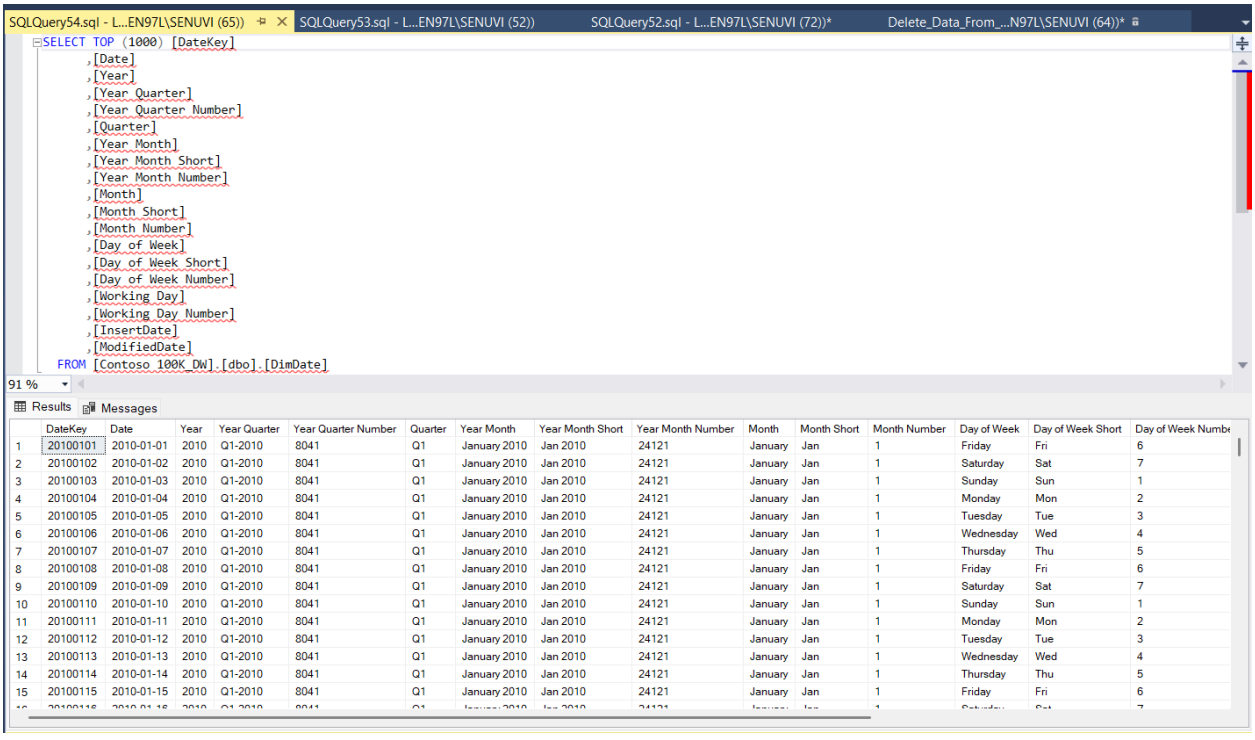
- Order of Execution



- Loading Date staging data in to data warehouse DimDate table.



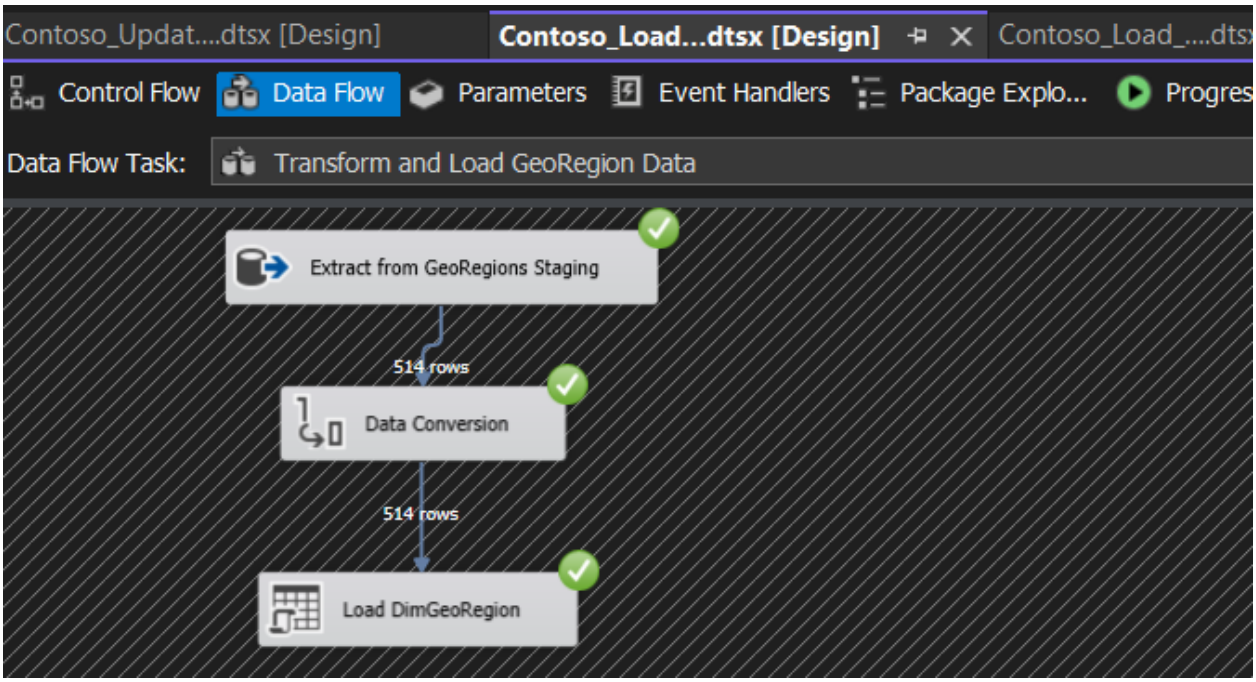
DimDate table



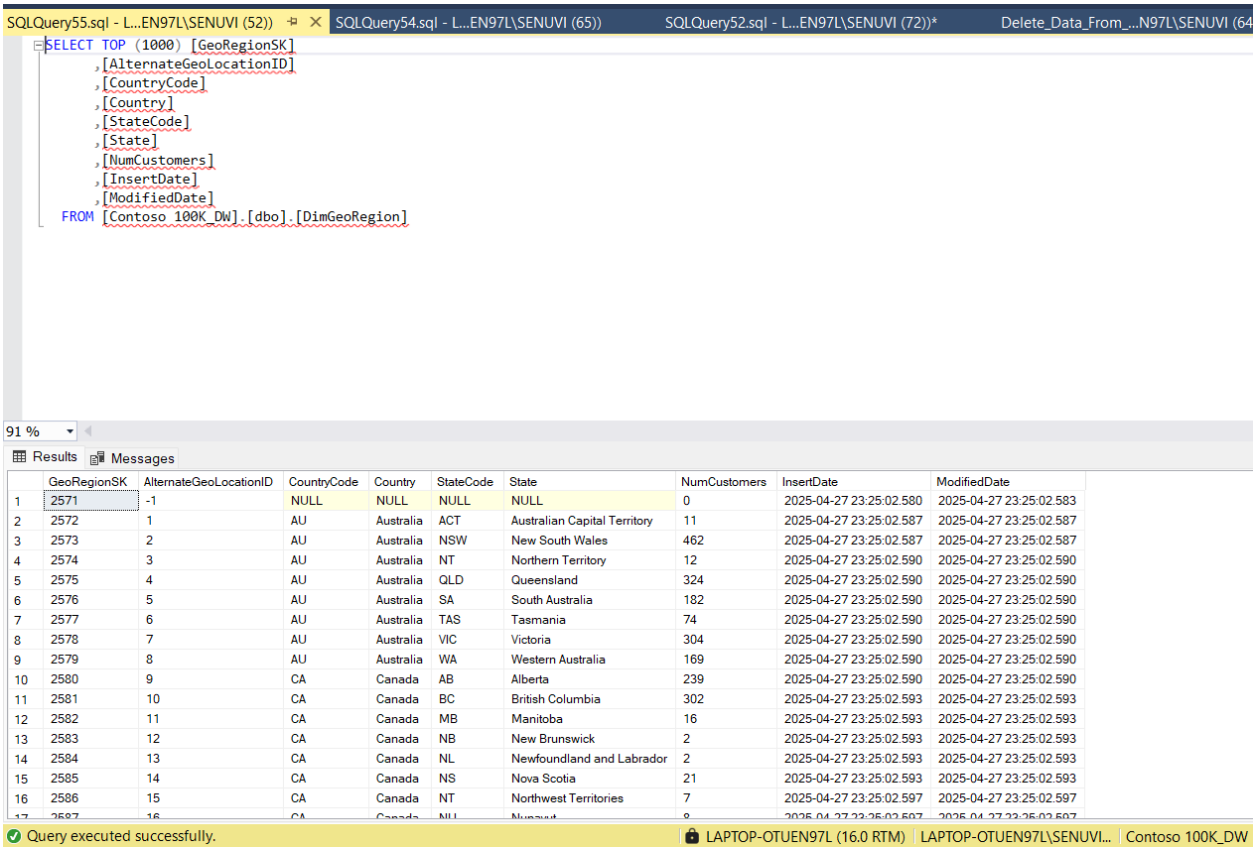
```
SELECT TOP (1000) [DateKey]
      ,[Date]
      ,[Year]
      ,[Year Quarter]
      ,[Year Quarter Number]
      ,[Quarter]
      ,[Year Month]
      ,[Year Month Short]
      ,[Year Month Number]
      ,[Month]
      ,[Month Short]
      ,[Month Number]
      ,[Day of Week]
      ,[Day of Week Short]
      ,[Day of Week Number]
      ,[Working Day]
      ,[Working Day Number]
      ,[InsertDate]
      ,[ModifiedDate]
  FROM [Contoso 100K DW].[dbo].[DimDate]
```

	DateKey	Date	Year	Year Quarter	Year Quarter Number	Quarter	Year Month	Year Month Short	Year Month Number	Month	Month Short	Month Number	Day of Week	Day of Week Short	Day of Week Numbe
1	20100101	2010-01-01	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Friday	Fri	6
2	20100102	2010-01-02	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Saturday	Sat	7
3	20100103	2010-01-03	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Sunday	Sun	1
4	20100104	2010-01-04	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Monday	Mon	2
5	20100105	2010-01-05	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Tuesday	Tue	3
6	20100106	2010-01-06	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Wednesday	Wed	4
7	20100107	2010-01-07	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Thursday	Thu	5
8	20100108	2010-01-08	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Friday	Fri	6
9	20100109	2010-01-09	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Saturday	Sat	7
10	20100110	2010-01-10	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Sunday	Sun	1
11	20100111	2010-01-11	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Monday	Mon	2
12	20100112	2010-01-12	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Tuesday	Tue	3
13	20100113	2010-01-13	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Wednesday	Wed	4
14	20100114	2010-01-14	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Thursday	Thu	5
15	20100115	2010-01-15	2010	Q1-2010	8041	Q1	January 2010	Jan 2010	24121	January	Jan	1	Friday	Fri	6

- Loading GeoRegions staging data in to data warehouse DimGeoRegions table.

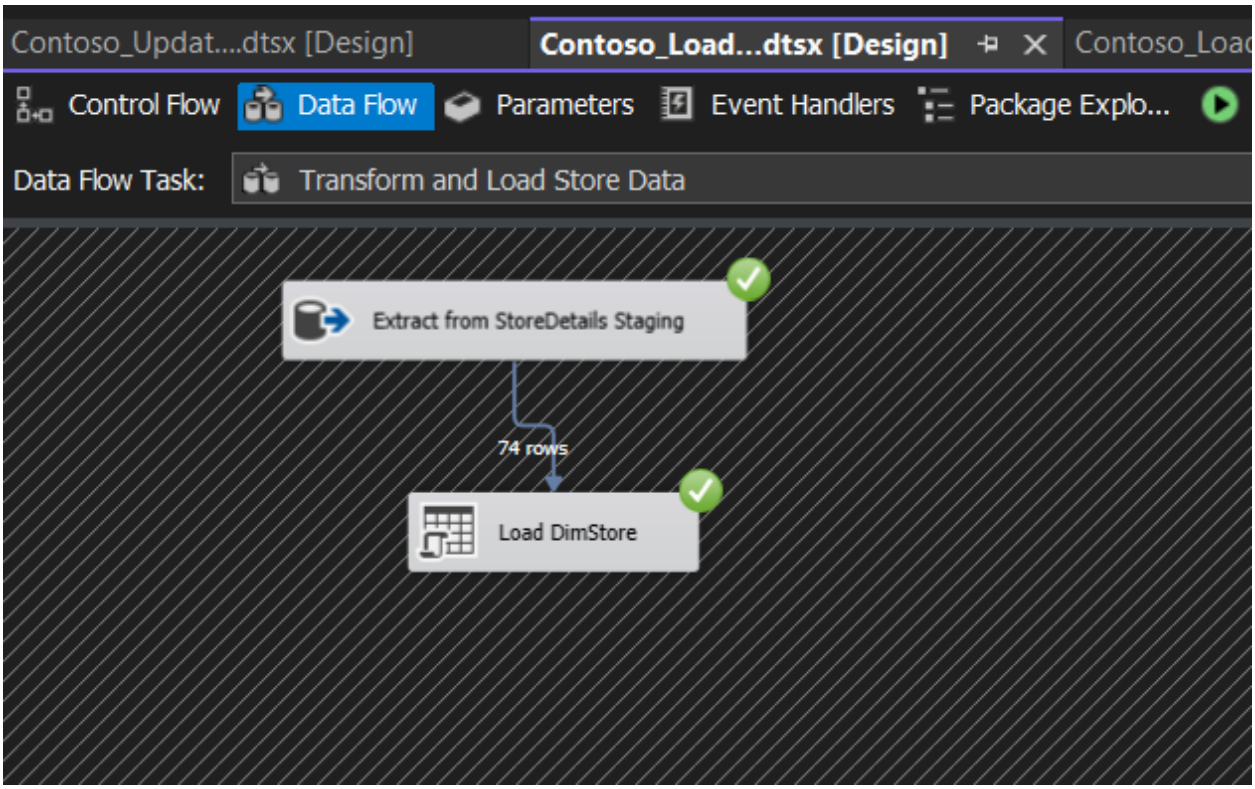


GeoRegions table

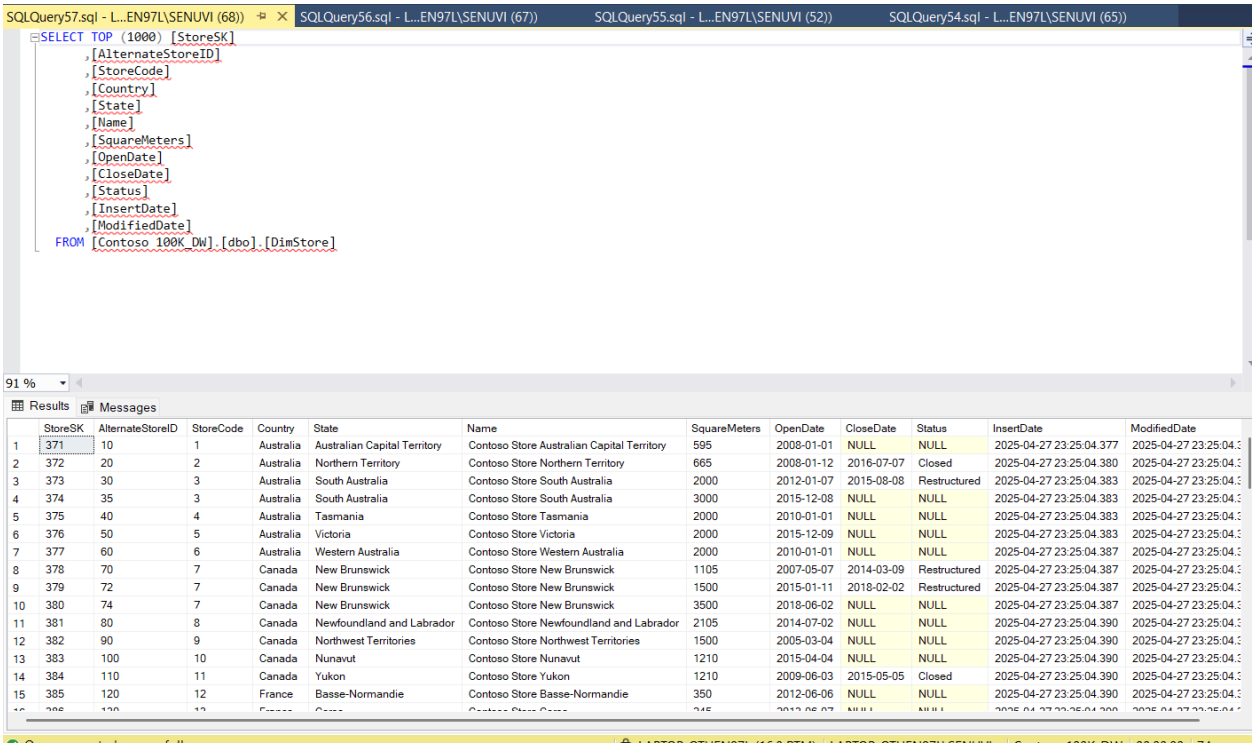


	GeoRegionSK	AlternateGeoLocationID	CountryCode	Country	StateCode	State	NumCustomers	InsertDate	ModifiedDate
1	2571	-1	NULL	NULL	NULL	NULL	0	2025-04-27 23:25:02.580	2025-04-27 23:25:02.583
2	2572	1	AU	Australia	ACT	Australian Capital Territory	11	2025-04-27 23:25:02.587	2025-04-27 23:25:02.587
3	2573	2	AU	Australia	NSW	New South Wales	462	2025-04-27 23:25:02.587	2025-04-27 23:25:02.587
4	2574	3	AU	Australia	NT	Northern Territory	12	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
5	2575	4	AU	Australia	QLD	Queensland	324	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
6	2576	5	AU	Australia	SA	South Australia	182	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
7	2577	6	AU	Australia	TAS	Tasmania	74	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
8	2578	7	AU	Australia	VIC	Victoria	304	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
9	2579	8	AU	Australia	WA	Western Australia	169	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
10	2580	9	CA	Canada	AB	Alberta	239	2025-04-27 23:25:02.590	2025-04-27 23:25:02.590
11	2581	10	CA	Canada	BC	British Columbia	302	2025-04-27 23:25:02.593	2025-04-27 23:25:02.593
12	2582	11	CA	Canada	MB	Manitoba	16	2025-04-27 23:25:02.593	2025-04-27 23:25:02.593
13	2583	12	CA	Canada	NB	New Brunswick	2	2025-04-27 23:25:02.593	2025-04-27 23:25:02.593
14	2584	13	CA	Canada	NL	Newfoundland and Labrador	2	2025-04-27 23:25:02.593	2025-04-27 23:25:02.593
15	2585	14	CA	Canada	NS	Nova Scotia	21	2025-04-27 23:25:02.593	2025-04-27 23:25:02.593
16	2586	15	CA	Canada	NT	Northwest Territories	7	2025-04-27 23:25:02.597	2025-04-27 23:25:02.597

- Loading StoreDetails staging data in to data warehouse DimStore table.



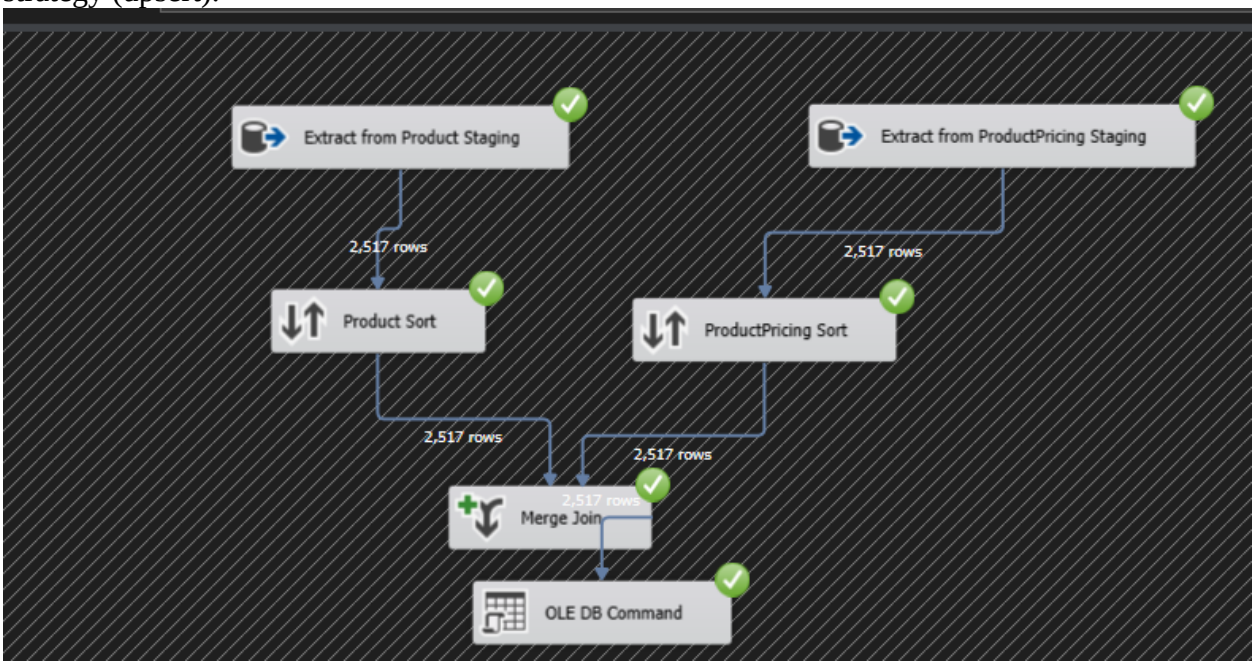
DimStore table



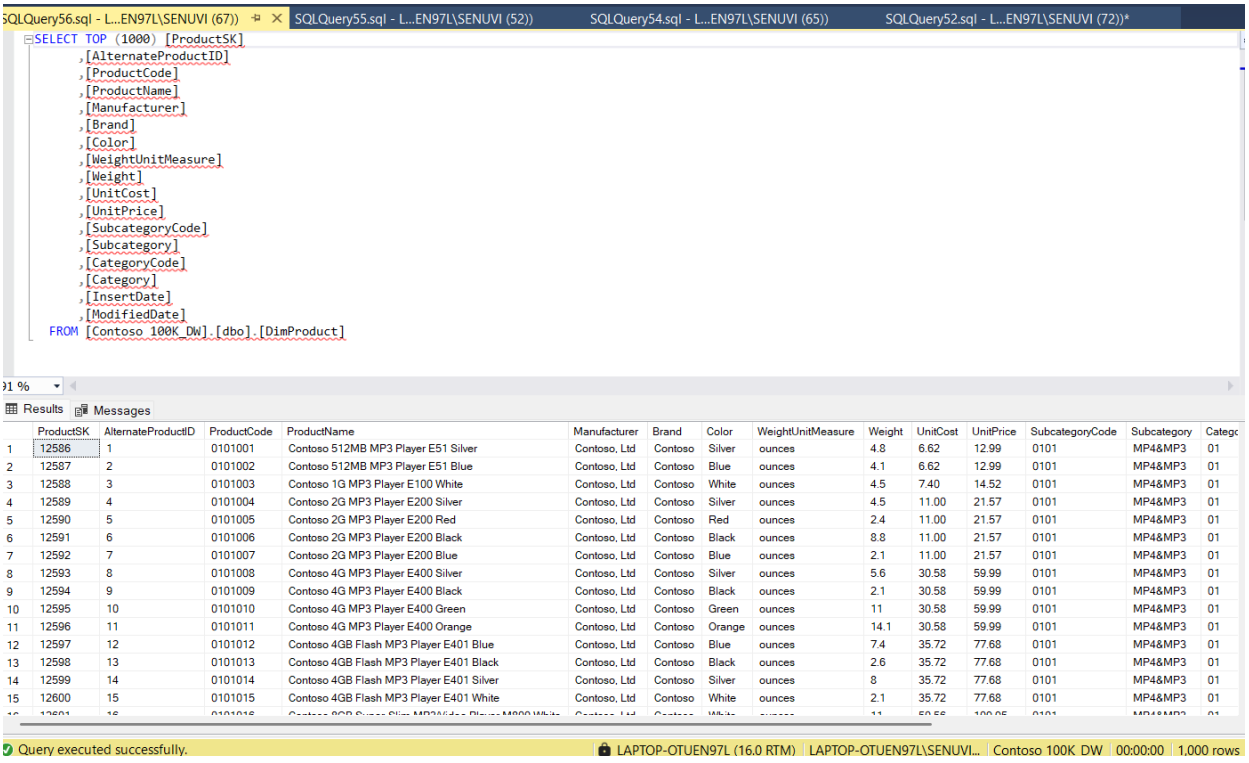
	StoreSK	AlternateStoreID	StoreCode	Country	State	Name	SquareMeters	OpenDate	CloseDate	Status	InsertDate	ModifiedDate
1	371	10	1	Australia	Australian Capital Territory	Contoso Store Australian Capital Territory	595	2008-01-01	NULL	NULL	2025-04-27 23:25:04.377	2025-04-27 23:25:04.3
2	372	20	2	Australia	Northern Territory	Contoso Store Northern Territory	665	2008-01-12	2016-07-07	Closed	2025-04-27 23:25:04.380	2025-04-27 23:25:04.3
3	373	30	3	Australia	South Australia	Contoso Store South Australia	2000	2012-01-07	2015-08-08	Restructured	2025-04-27 23:25:04.383	2025-04-27 23:25:04.3
4	374	35	3	Australia	South Australia	Contoso Store South Australia	3000	2015-12-08	NULL	NULL	2025-04-27 23:25:04.383	2025-04-27 23:25:04.3
5	375	40	4	Australia	Tasmania	Contoso Store Tasmania	2000	2010-01-01	NULL	NULL	2025-04-27 23:25:04.383	2025-04-27 23:25:04.3
6	376	50	5	Australia	Victoria	Contoso Store Victoria	2000	2015-12-09	NULL	NULL	2025-04-27 23:25:04.383	2025-04-27 23:25:04.3
7	377	60	6	Australia	Western Australia	Contoso Store Western Australia	2000	2010-01-01	NULL	NULL	2025-04-27 23:25:04.387	2025-04-27 23:25:04.3
8	378	70	7	Canada	New Brunswick	Contoso Store New Brunswick	1105	2007-05-07	2014-03-09	Restructured	2025-04-27 23:25:04.387	2025-04-27 23:25:04.3
9	379	72	7	Canada	New Brunswick	Contoso Store New Brunswick	1500	2015-01-11	2018-02-02	Restructured	2025-04-27 23:25:04.387	2025-04-27 23:25:04.3
10	380	74	7	Canada	New Brunswick	Contoso Store New Brunswick	3500	2018-06-02	NULL	NULL	2025-04-27 23:25:04.387	2025-04-27 23:25:04.3
11	381	80	8	Canada	Newfoundland and Labrador	Contoso Store Newfoundland and Labrador	2105	2014-07-02	NULL	NULL	2025-04-27 23:25:04.390	2025-04-27 23:25:04.3
12	382	90	9	Canada	Northwest Territories	Contoso Store Northwest Territories	1500	2005-03-04	NULL	NULL	2025-04-27 23:25:04.390	2025-04-27 23:25:04.3
13	383	100	10	Canada	Nunavut	Contoso Store Nunavut	1210	2015-04-04	NULL	NULL	2025-04-27 23:25:04.390	2025-04-27 23:25:04.3
14	384	110	11	Canada	Yukon	Contoso Store Yukon	1210	2009-06-03	2015-05-05	Closed	2025-04-27 23:25:04.390	2025-04-27 23:25:04.3
15	385	120	12	France	Basse-Normandie	Contoso Store Basse-Normandie	350	2012-06-06	NULL	NULL	2025-04-27 23:25:04.390	2025-04-27 23:25:04.3

- Loading Product and ProductPricing Staging Data into DimProduct

This process loads data from dbo.StgProduct and dbo.StgProductPricing (SQL sources) into dbo.DimProduct, combining product details and pricing information. It uses a Type 1 update strategy (upsert).



DimProduct table

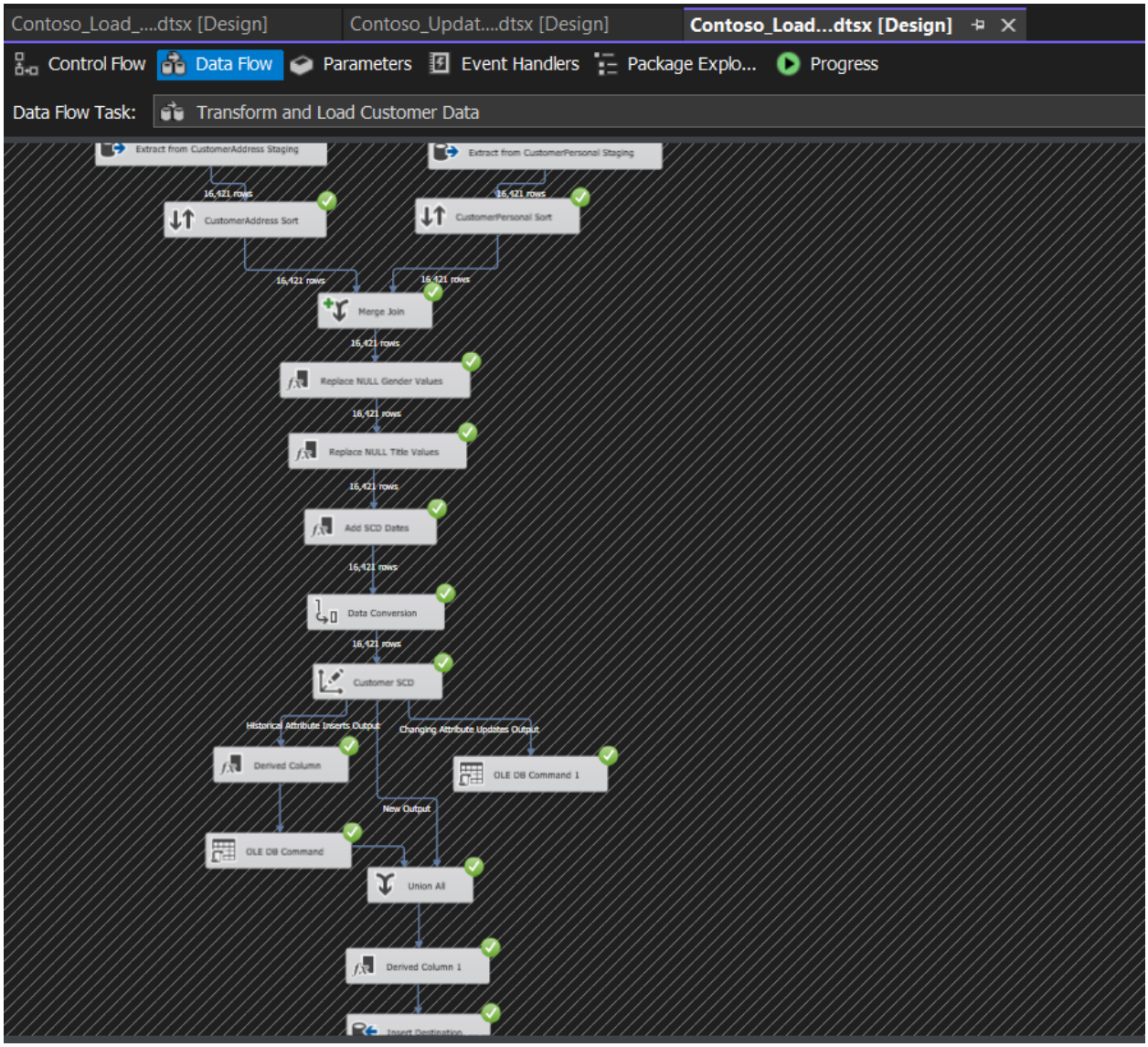


```
SELECT TOP (1000) [ProductSK]
      ,[AlternateProductID]
      ,[ProductCode]
      ,[ProductName]
      ,[Manufacturer]
      ,[Brand]
      ,[Color]
      ,[WeightUnitMeasure]
      ,[Weight]
      ,[UnitCost]
      ,[UnitPrice]
      ,[SubcategoryCode]
      ,[Subcategory]
      ,[CategoryCode]
      ,[Category]
      ,[InsertDate]
      ,[ModifiedDate]
  FROM [Contoso 100K_DW].[dbo].[DimProduct]
```

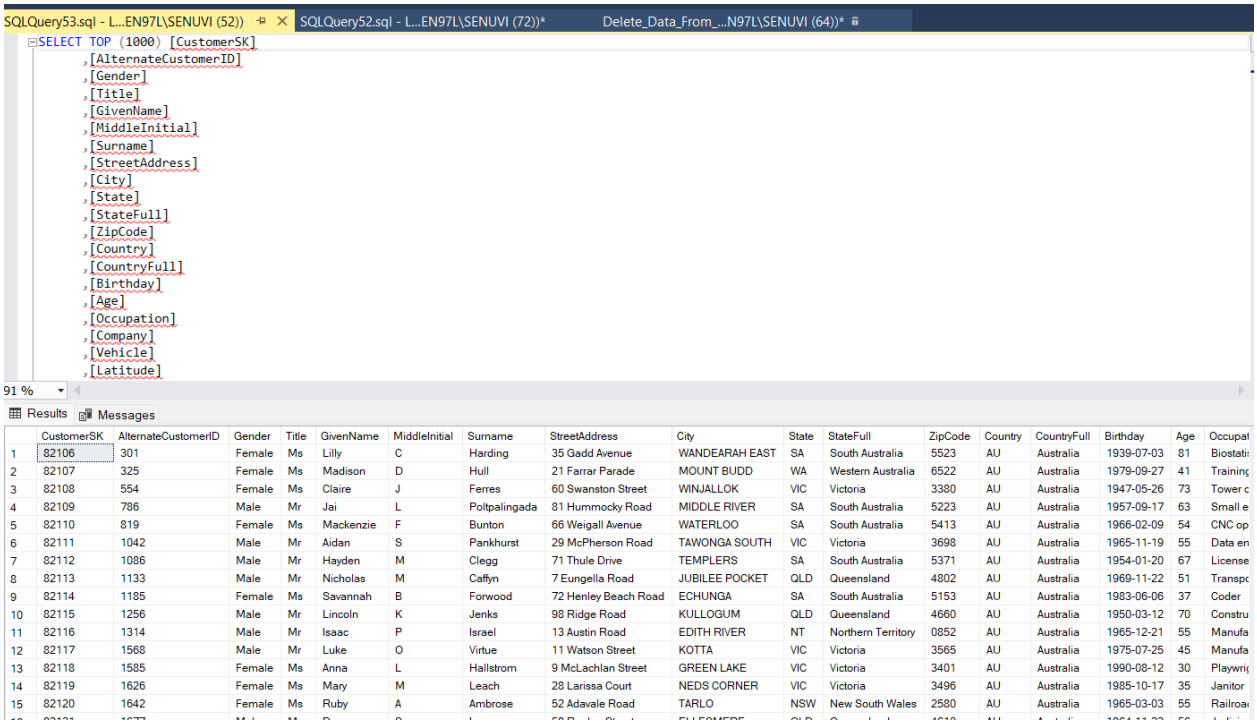
	ProductSK	AlternateProductID	ProductCode	ProductName	Manufacturer	Brand	Color	WeightUnitMeasure	Weight	UnitCost	UnitPrice	SubcategoryCode	Subcategory	Categc
1	12586	1	0101001	Contoso 512MB MP3 Player E51 Silver	Contoso, Ltd	Contoso	Silver	ounces	4.8	6.62	12.99	0101	MP4&MP3	01
2	12587	2	0101002	Contoso 512MB MP3 Player E51 Blue	Contoso, Ltd	Contoso	Blue	ounces	4.1	6.62	12.99	0101	MP4&MP3	01
3	12588	3	0101003	Contoso 1G MP3 Player E100 White	Contoso, Ltd	Contoso	White	ounces	4.5	7.40	14.52	0101	MP4&MP3	01
4	12589	4	0101004	Contoso 2G MP3 Player E200 Silver	Contoso, Ltd	Contoso	Silver	ounces	4.5	11.00	21.57	0101	MP4&MP3	01
5	12590	5	0101005	Contoso 2G MP3 Player E200 Red	Contoso, Ltd	Contoso	Red	ounces	2.4	11.00	21.57	0101	MP4&MP3	01
6	12591	6	0101006	Contoso 2G MP3 Player E200 Black	Contoso, Ltd	Contoso	Black	ounces	8.8	11.00	21.57	0101	MP4&MP3	01
7	12592	7	0101007	Contoso 2G MP3 Player E200 Blue	Contoso, Ltd	Contoso	Blue	ounces	2.1	11.00	21.57	0101	MP4&MP3	01
8	12593	8	0101008	Contoso 4G MP3 Player E400 Silver	Contoso, Ltd	Contoso	Silver	ounces	5.6	30.58	59.99	0101	MP4&MP3	01
9	12594	9	0101009	Contoso 4G MP3 Player E400 Black	Contoso, Ltd	Contoso	Black	ounces	2.1	30.58	59.99	0101	MP4&MP3	01
10	12595	10	0101010	Contoso 4G MP3 Player E400 Green	Contoso, Ltd	Contoso	Green	ounces	11	30.58	59.99	0101	MP4&MP3	01
11	12596	11	0101011	Contoso 4G MP3 Player E400 Orange	Contoso, Ltd	Contoso	Orange	ounces	14.1	30.58	59.99	0101	MP4&MP3	01
12	12597	12	0101012	Contoso 4GB Flash MP3 Player E401 Blue	Contoso, Ltd	Contoso	Blue	ounces	7.4	35.72	77.68	0101	MP4&MP3	01
13	12598	13	0101013	Contoso 4GB Flash MP3 Player E401 Black	Contoso, Ltd	Contoso	Black	ounces	2.6	35.72	77.68	0101	MP4&MP3	01
14	12599	14	0101014	Contoso 4GB Flash MP3 Player E401 Silver	Contoso, Ltd	Contoso	Silver	ounces	8	35.72	77.68	0101	MP4&MP3	01
15	12600	15	0101015	Contoso 4GB Flash MP3 Player E401 White	Contoso, Ltd	Contoso	White	ounces	2.1	35.72	77.68	0101	MP4&MP3	01

- Loading CustomerPersonal and CustomerAddress Staging Data into DimCustomer

This process loads data from dbo.StgCustomerPersonal (SQL source) and dbo.StgCustomerAddress (text source) into dbo.DimCustomer, implementing an SCD Type 2 for address-related fields and Type 1 for other attributes (e.g., Gender, Title).



DimCustomer Table



SQLQuery53.sql - L...EN97L\SENUVI (52)) | SQLQuery52.sql - L...EN97L\SENUVI (72))* | Delete_Data_From_...N97L\SENUVI (64))*

```
SELECT TOP (1000) [CustomerSK]
      ,[AlternateCustomerID]
      ,[Gender]
      ,[Title]
      ,[GivenName]
      ,[MiddleInitial]
      ,[Surname]
      ,[StreetAddress]
      ,[City]
      ,[State]
      ,[StateFull]
      ,[ZipCode]
      ,[Country]
      ,[CountryFull]
      ,[Birthday]
      ,[Age]
      ,[Occupation]
      ,[Company]
      ,[Vehicle]
      ,[Latitude]
```

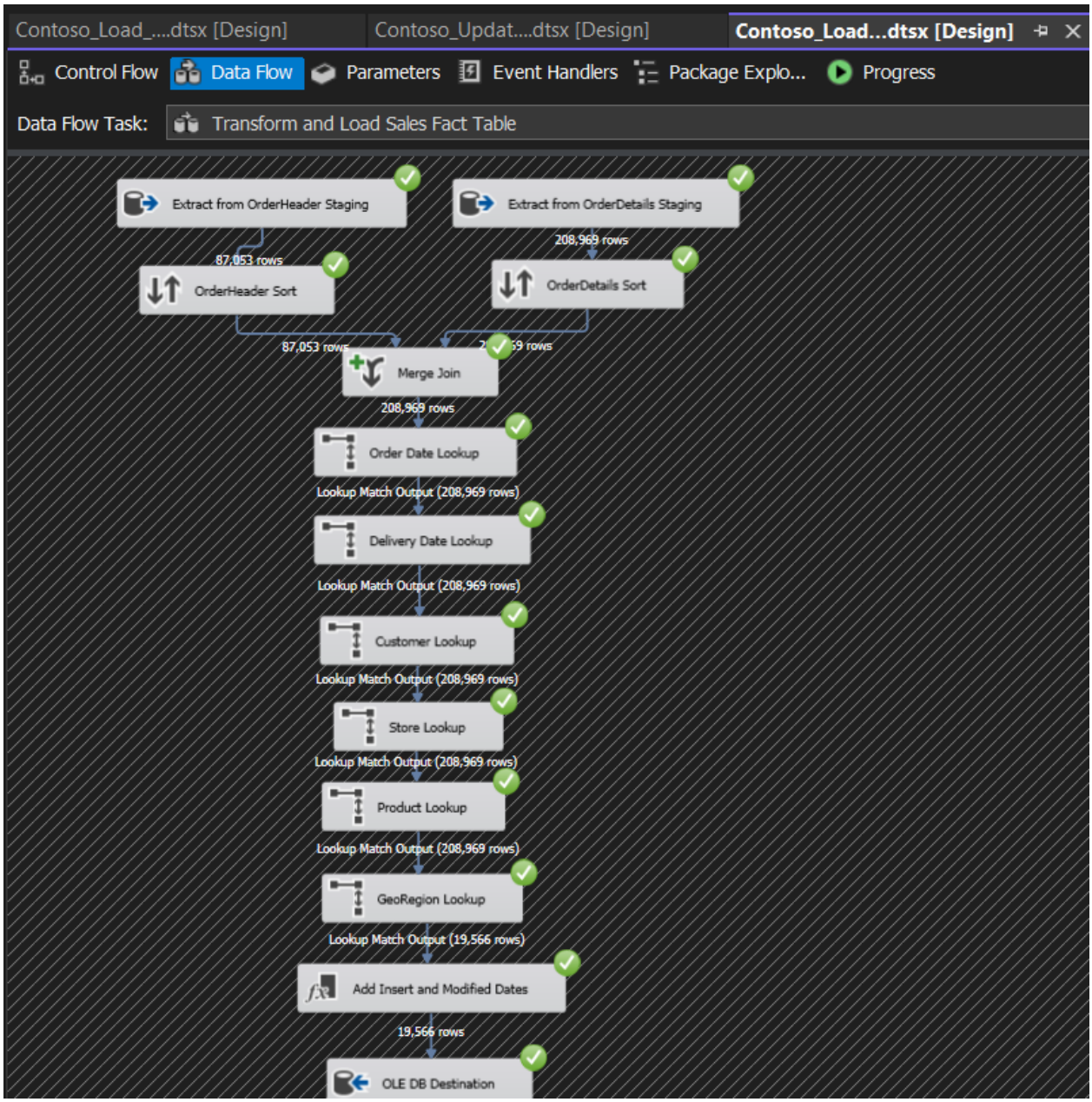
91 %

Results | Messages

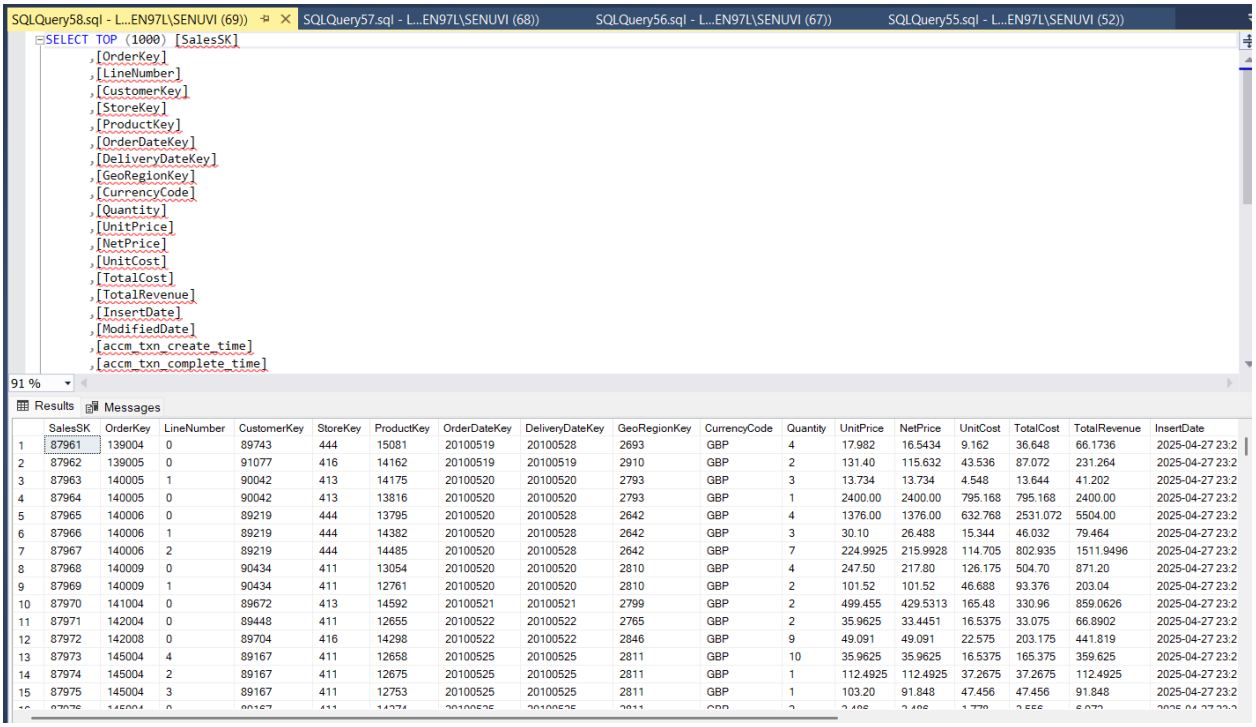
	CustomerSK	AlternateCustomerID	Gender	Title	GivenName	MiddleInitial	Surname	StreetAddress	City	State	StateFull	ZipCode	Country	CountryFull	Birthday	Age	Occupat
1	82106	301	Female	Ms	Lilly	C	Harding	35 Gadd Avenue	WANDEARAH EAST	SA	South Australia	5523	AU	Australia	1939-07-03	81	Biostati
2	82107	325	Female	Ms	Madison	D	Hull	21 Farrar Parade	MOUNT BUDD	WA	Western Australia	6522	AU	Australia	1979-09-27	41	Training
3	82108	554	Female	Ms	Claire	J	Ferres	60 Swanston Street	WINJALLOK	VIC	Victoria	3380	AU	Australia	1947-05-26	73	Tower c
4	82109	786	Male	Mr	Jai	L	Poltpalingada	81 Hummocky Road	MIDDLE RIVER	SA	South Australia	5223	AU	Australia	1957-09-17	63	Small e
5	82110	819	Female	Ms	Mackenzie	F	Bunton	66 Weigall Avenue	WATERLOO	SA	South Australia	5413	AU	Australia	1966-02-09	54	CNC op
6	82111	1042	Male	Mr	Aidan	S	Pankhurst	29 McPherson Road	TAWONGA SOUTH	VIC	Victoria	3698	AU	Australia	1965-11-19	55	Data en
7	82112	1086	Male	Mr	Hayden	M	Clegg	71 Thule Drive	TEMPLERS	SA	South Australia	5371	AU	Australia	1954-01-20	67	License
8	82113	1133	Male	Mr	Nicholas	M	Caffyn	7 Eungella Road	JUBILEE POCKET	QLD	Queensland	4802	AU	Australia	1969-11-22	51	Transpo
9	82114	1185	Female	Ms	Savannah	B	Forwood	72 Henley Beach Road	ECHUNGA	SA	South Australia	5153	AU	Australia	1983-06-06	37	Coder
10	82115	1256	Male	Mr	Lincoln	K	Jenks	98 Ridge Road	KULLOGUM	QLD	Queensland	4660	AU	Australia	1950-03-12	70	Constru
11	82116	1314	Male	Mr	Isaac	P	Israel	13 Austin Road	EDITH RIVER	NT	Northern Territory	0852	AU	Australia	1965-12-21	55	Manufa
12	82117	1568	Male	Mr	Luke	O	Virtue	11 Watson Street	KOTTA	VIC	Victoria	3565	AU	Australia	1975-07-25	45	Manufa
13	82118	1585	Female	Ms	Anna	L	Hallstrom	9 McLachlan Street	GREEN LAKE	VIC	Victoria	3401	AU	Australia	1990-08-12	30	Playwri
14	82119	1626	Female	Ms	Mary	M	Leach	28 Larissa Court	NEDS CORNER	VIC	Victoria	3496	AU	Australia	1985-10-17	35	Janitor
15	82120	1642	Female	Ms	Ruby	A	Ambrose	52 Adavale Road	TARLO	NSW	New South Wales	2580	AU	Australia	1965-03-03	55	Railroa

- Loading OrderHeader and OrderDetails Staging Data into FactSales

This process loads data from dbo.StgOrderHeader and dbo.StgOrderDetails (SQL sources) into dbo.FactSales, a transactional fact table. It includes lookups to map natural keys to surrogate keys and adds accumulating columns.



FactSales table



```
SELECT TOP (1000) [SalesSK]
      ,[OrderKey]
      ,[LineNumber]
      ,[CustomerKey]
      ,[StoreKey]
      ,[ProductKey]
      ,[OrderDateKey]
      ,[DeliveryDateKey]
      ,[GeoRegionKey]
      ,[CurrencyCode]
      ,[Quantity]
      ,[UnitPrice]
      ,[NetPrice]
      ,[UnitCost]
      ,[TotalCost]
      ,[TotalRevenue]
      ,[InsertDate]
      ,[ModifiedDate]
      ,[accm_txn_create_time]
      ,[accm_txn_complete_time]
```

Results Messages

	SalesSK	OrderKey	LineNumber	CustomerKey	StoreKey	ProductKey	OrderDateKey	DeliveryDateKey	GeoRegionKey	CurrencyCode	Quantity	UnitPrice	NetPrice	UnitCost	TotalCost	TotalRevenue	InsertDate
1	87961	139004	0	89743	444	15081	20100519	20100528	2693	GBP	4	17.982	16.5434	9.162	36.648	66.1736	2025-04-27 23:2
2	87962	139005	0	91077	416	14162	20100519	20100519	2910	GBP	2	131.40	115.632	43.536	87.072	231.264	2025-04-27 23:2
3	87963	140005	1	90042	413	14175	20100520	20100520	2793	GBP	3	13.734	13.734	4.548	13.644	41.202	2025-04-27 23:2
4	87964	140005	0	90042	413	13816	20100520	20100520	2793	GBP	1	2400.00	2400.00	795.168	795.168	2400.00	2025-04-27 23:2
5	87965	140006	0	89219	444	13795	20100520	20100528	2642	GBP	4	1376.00	1376.00	632.768	2531.072	5504.00	2025-04-27 23:2
6	87966	140006	1	89219	444	14382	20100520	20100528	2642	GBP	3	30.10	26.488	15.344	46.032	79.464	2025-04-27 23:2
7	87967	140006	2	89219	444	14485	20100520	20100528	2642	GBP	7	224.9925	215.9928	114.705	802.935	1511.9496	2025-04-27 23:2
8	87968	140009	0	90434	411	13054	20100520	20100520	2810	GBP	4	247.50	217.80	126.175	504.70	871.20	2025-04-27 23:2
9	87969	140009	1	90434	411	12761	20100520	20100520	2810	GBP	2	101.52	101.52	46.688	93.376	203.04	2025-04-27 23:2
10	87970	141004	0	89672	413	14592	20100521	20100521	2799	GBP	2	499.455	429.5313	165.48	330.96	859.0626	2025-04-27 23:2
11	87971	142004	0	89448	411	12655	20100522	20100522	2765	GBP	2	35.9625	33.4451	16.5375	33.075	66.8902	2025-04-27 23:2
12	87972	142008	0	89704	416	14298	20100522	20100522	2846	GBP	9	49.091	49.091	22.575	203.175	441.819	2025-04-27 23:2
13	87973	145004	4	89167	411	12658	20100525	20100525	2811	GBP	10	35.9625	35.9625	16.5375	165.375	359.625	2025-04-27 23:2
14	87974	145004	2	89167	411	12675	20100525	20100525	2811	GBP	1	112.4925	112.4925	37.2675	37.2675	112.4925	2025-04-27 23:2
15	87975	145004	3	89167	411	12753	20100525	20100525	2811	GBP	1	103.20	91.848	47.456	47.456	91.848	2025-04-27 23:2

8 ETL Development – Accumulating Fact Tables

The FactSales table was extended to include accumulating columns (accm_txn_create_time, accm_txn_complete_time, txn_process_time) to track transaction creation and completion times. A separate SSIS package, Contoso_Update_FactSales.dtsx, updates these columns using a CSV file.

Created a new SSIS package, Contoso_Update_FactSales.dtsx.

- Extended FactSales with Required Columns in SSMS

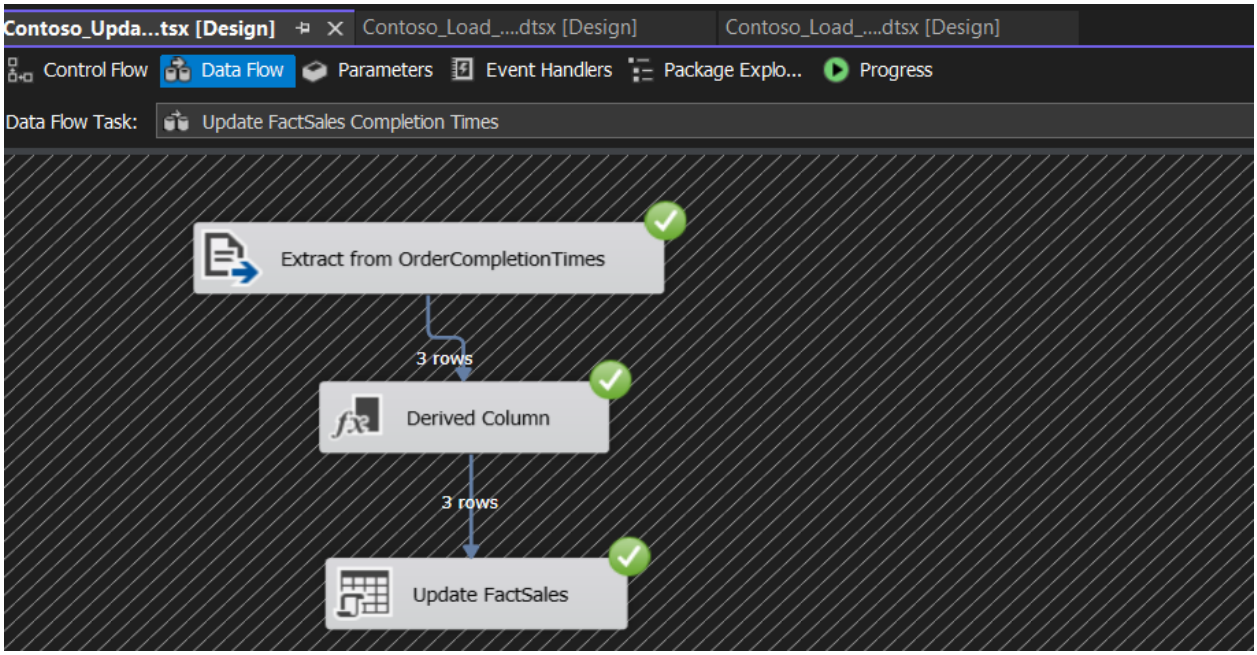
```
ALTER TABLE FactSales
ADD
    accm_txn_create_time DATETIME NOT NULL DEFAULT GETDATE(),
    accm_txn_complete_time DATETIME NULL,
    txn_process_time_hours INT NULL;
```

- Set “accm_txn_create_time” on Data Load

In **contoso_100k_DW SSIS package**, located the **Derived Column** transformation right before loading into FactSales.

Derived Column Name	Derived Column	Expression	Data Type
accm_txn_create_time	<add as new column>	GETDATE()	database timestamp ...

```
UPDATE dbo.FactSales
SET accm_txn_complete_time = ?,
    txn_process_time = DATEDIFF(HOUR, accm_txn_create_time, ?)
WHERE OrderKey = ?
```



Results Messages

	OrderKey	accm_txn_create_time	accm_txn_complete_time	txn_process_time
1	139004	2025-04-27 23:07:08.303	2025-04-28 12:00:00.000	13
2	139005	2025-04-27 23:07:08.303	2025-04-30 10:00:00.000	59
3	140006	2025-04-27 23:07:08.303	2025-04-29 12:00:00.000	37
4	140006	2025-04-27 23:07:08.303	2025-04-29 12:00:00.000	37
5	140006	2025-04-27 23:07:08.303	2025-04-29 12:00:00.000	37

9 Overall Execution Flow of the Total Solution

The following is an overall execution of the staging data flow tasks, followed by the execution of the data loading tasks to the data warehouse.

After the completion of the data loading, the final data flow task in the data loading package is set as a package execution task of the time update package, where it will be executed to update the fact table with the transaction complete times and the process times in hours.

This enables the proper flow execution of the data staging and then the execution of loading the staged data into the data warehouse and the updating of the fact table with the proper data in proper order.