AIM: Experiment to design Data Warehouse for given case study and perform ETL and OLAP operations on it

THEORY:

A **Data Warehouse** (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources. A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modeling and analysis. A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users. It is not used for daily operations and transaction processing but used for making decisions.

The target of the design becomes how the record from multiple data sources should be extracted, transformed, and loaded (ETL) to be organized in a database as the data warehouse. There are two approaches

1. "top-down" approach
2. "bottom-up" approach

Dimensional Schema

Schema is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items and aggregates. Much like a database, a data warehouse also requires maintaining a schema. A database uses relational models, while a data warehouse uses **Star, Snowflake, and Fact Constellation schema**.

OLAP Operations in the Multidimensional Data Model

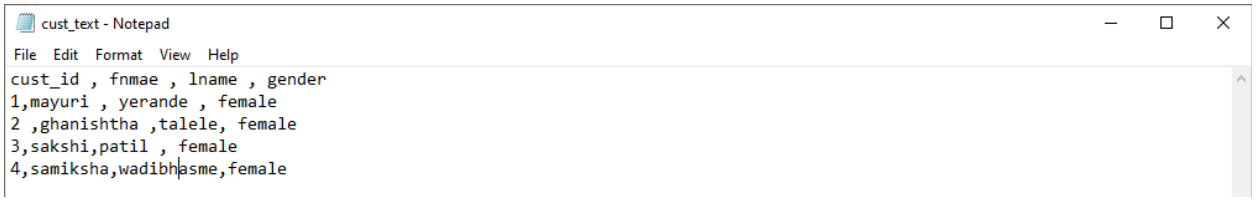
In the multidimensional model, the records are organized into various dimensions, and each dimension includes multiple levels of abstraction described by concept hierarchies. This organization supports users with the flexibility to view data from various perspectives. A number of OLAP data cube operations exist to demonstrate these different views, allowing interactive queries and search of the record at hand. Hence, OLAP supports a user-friendly environment for interactive data analysis.

Basic operations of OLAP Four types of analytical OLAP operations are:

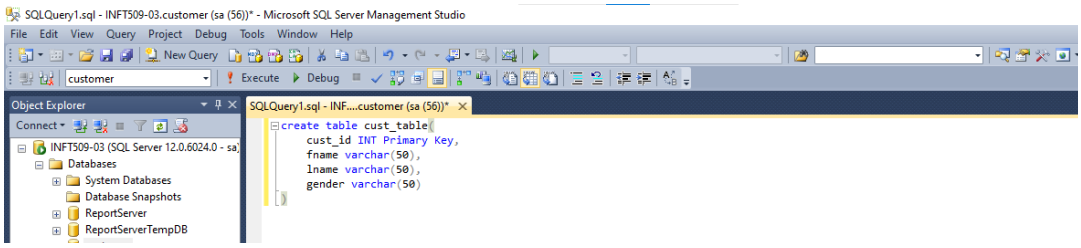
1. **Roll-up** - Roll-up is also known as “consolidation” or “aggregation.”
2. **Drill-down** - In drill-down data is fragmented into smaller parts. It is the opposite of the rollup process
3. **Slice and dice** - Here, one dimension is selected, and a new sub-cube is created.
4. **Pivot (rotate**) - This operation is similar to a slice. The difference in dice is you select 2 or more dimensions that result in the creation of a sub-cube

IMPLEMENTATION:

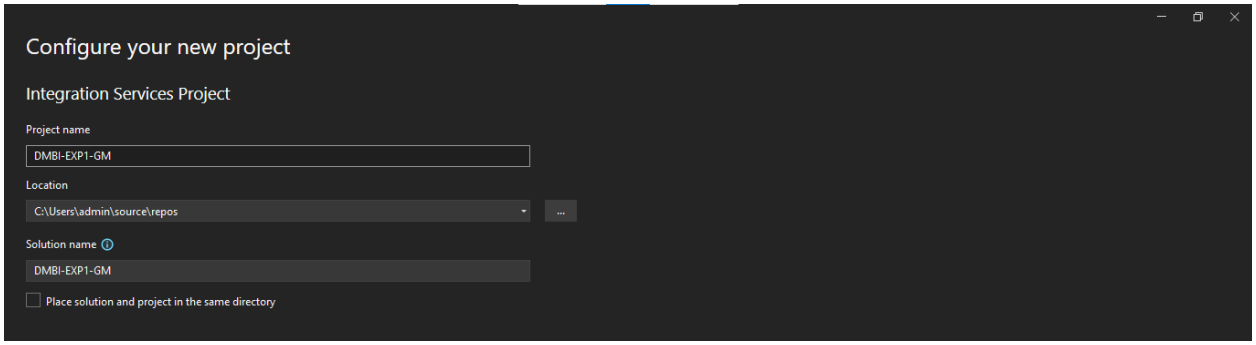
- Create database
- Create a text file with the first row filled with attributes and remaining rows filled with desired values.



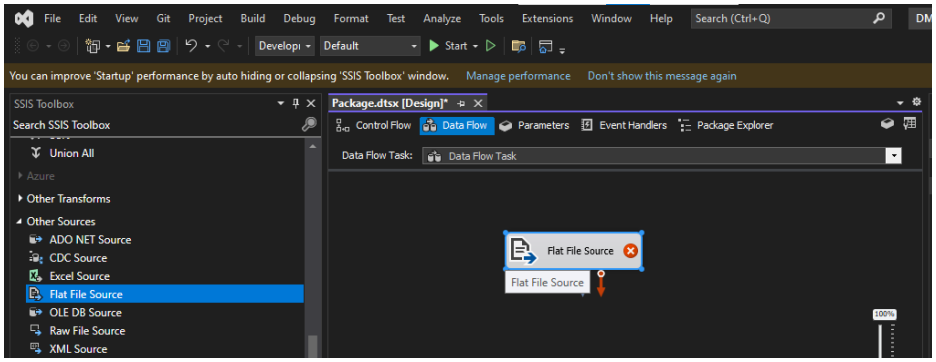
- Create a table in your database with the same attributes which were filled in a text file in previous steps with appropriate data type.



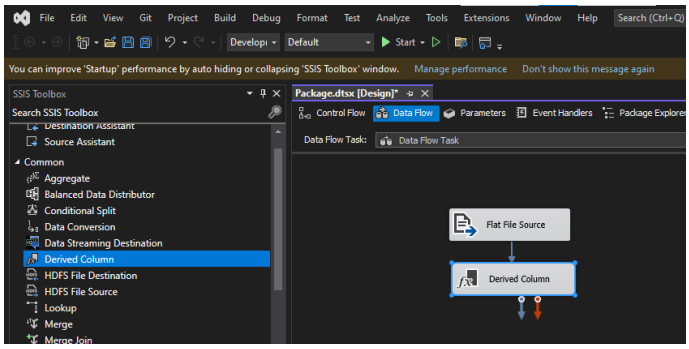
- Open Visual Studio 2022, create a new project and search for an integration service Project.



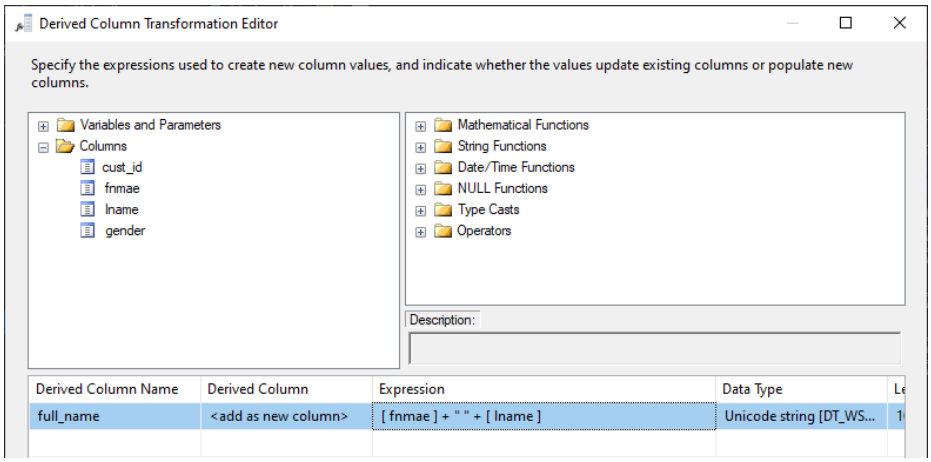
- Drag Data Flow Task to workspace.
- Click on Data Flow to go into its workspace and then drag Flat File Source to its Workspace.



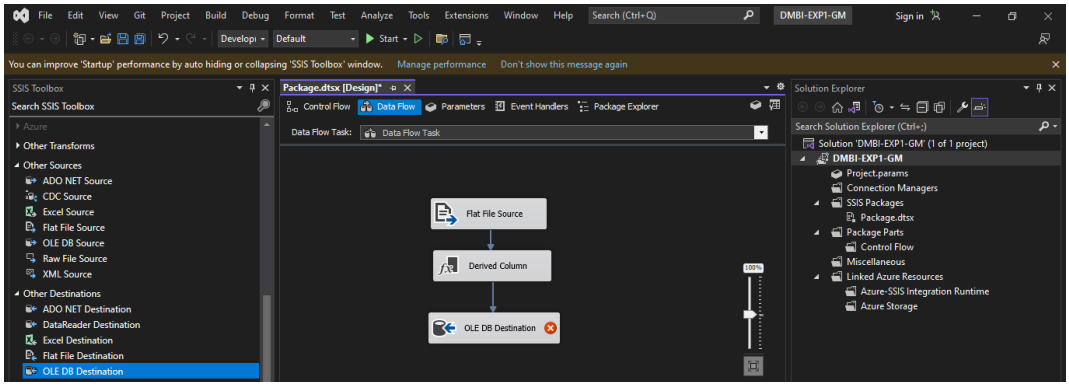
- After Clicking on Flat File Source click on New to add the cust.txt which was created in the first step.
- After Clicking on the Derived Column we see the following dialog box, add all the column names to be mapped and transformed with needful transformation in the expression column.



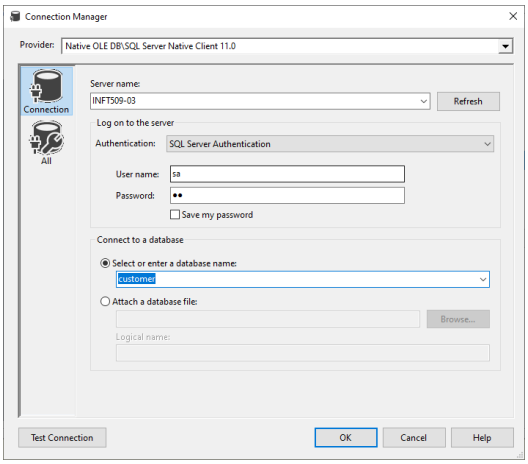
- We will merge fname and lname as full name

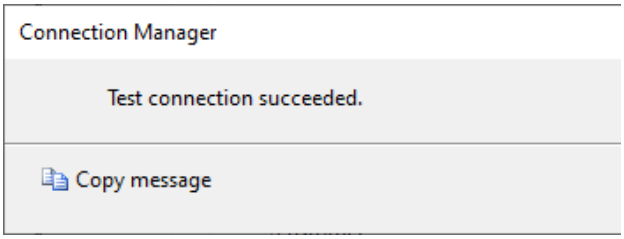


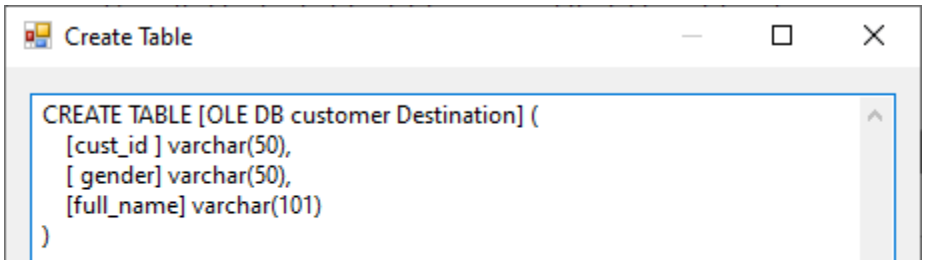
- Connect OLE DB Connection



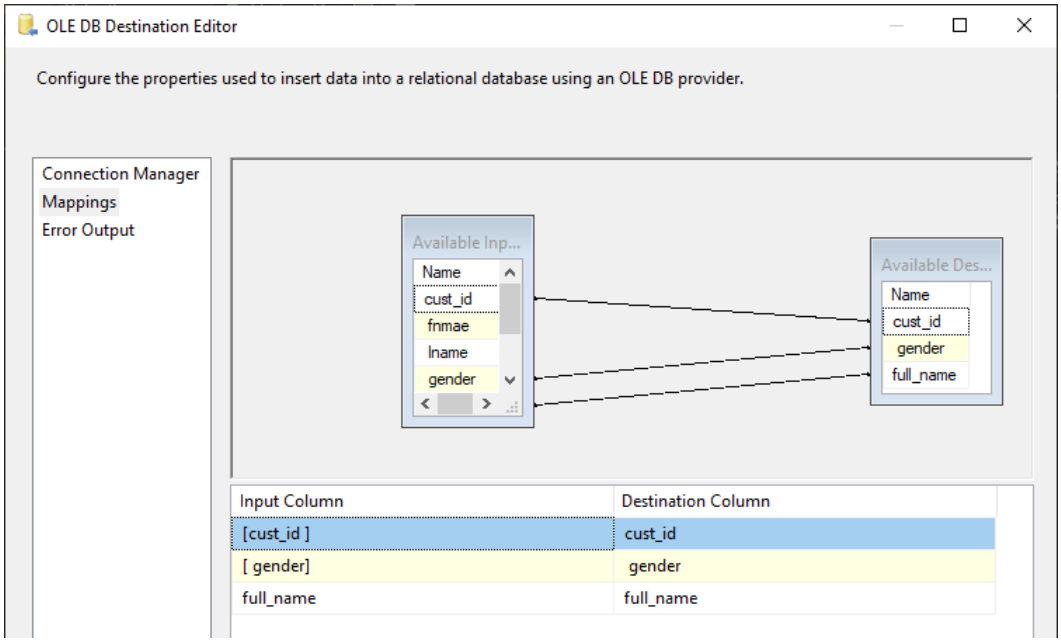
- Creating OLE manager and test connection



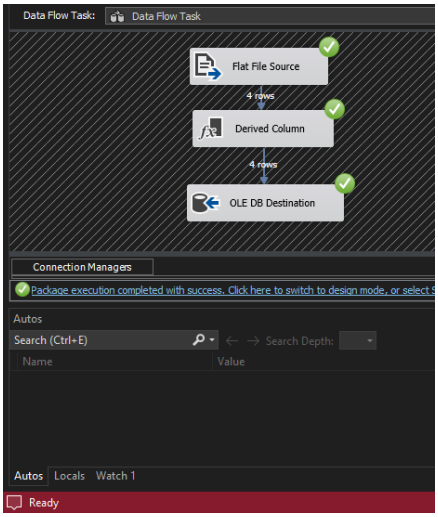


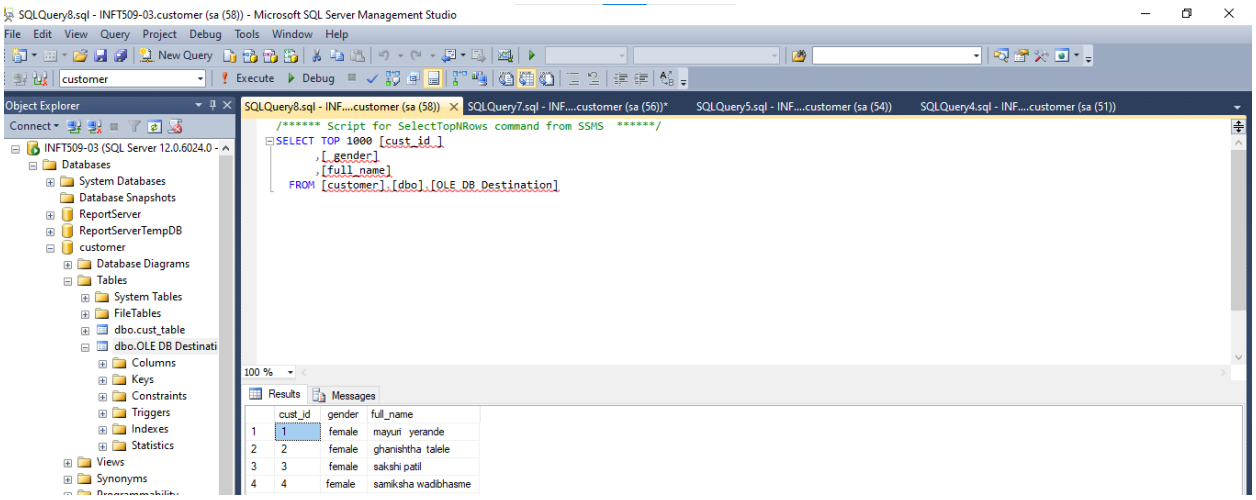


- Mapping done



- Connection successful
- Now run project with “start”
- No we enter the ready state
- Output is seen in sql server

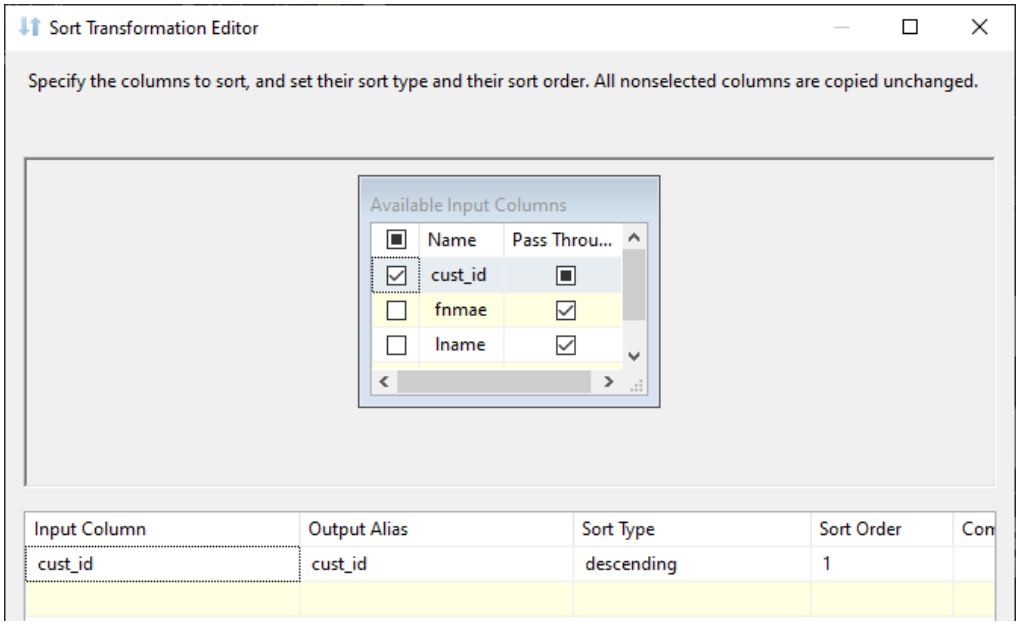




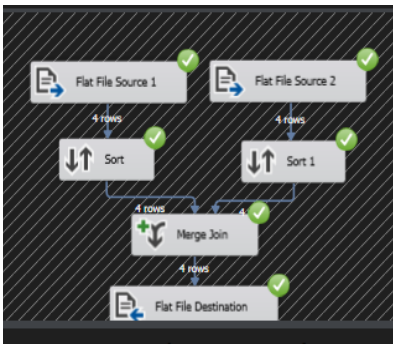
MERGE

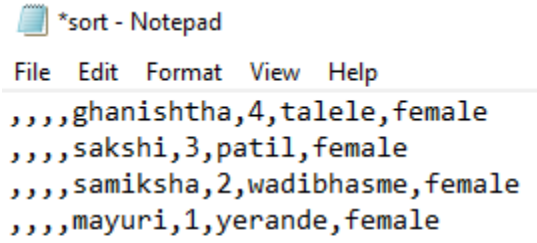
- Create 2 flat source files

- Sort the files according to cust id



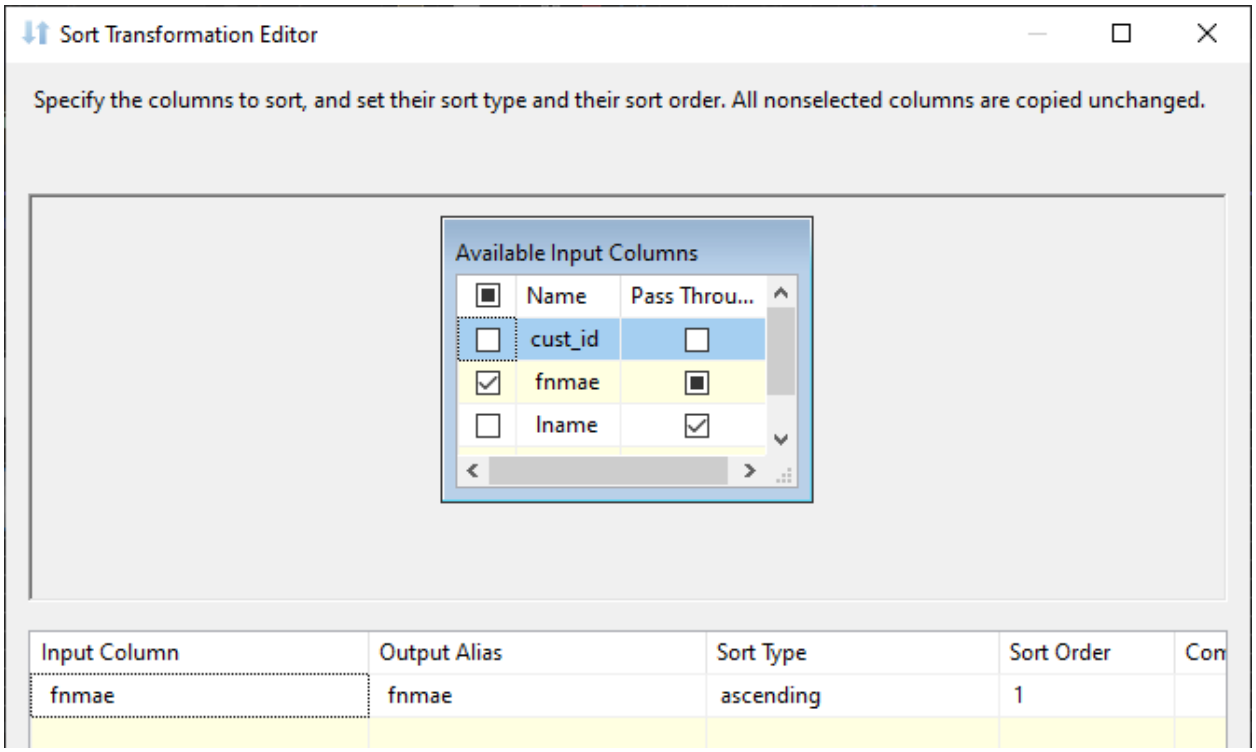
- Select all the columns of the files
- Connection successful
- Output



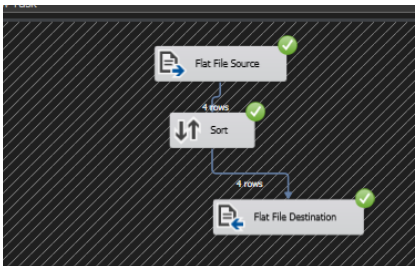


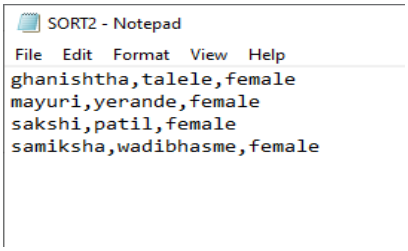
SORTING

- Create flat file source manager and add the same cust file
- Sorting according to fname in ascending order



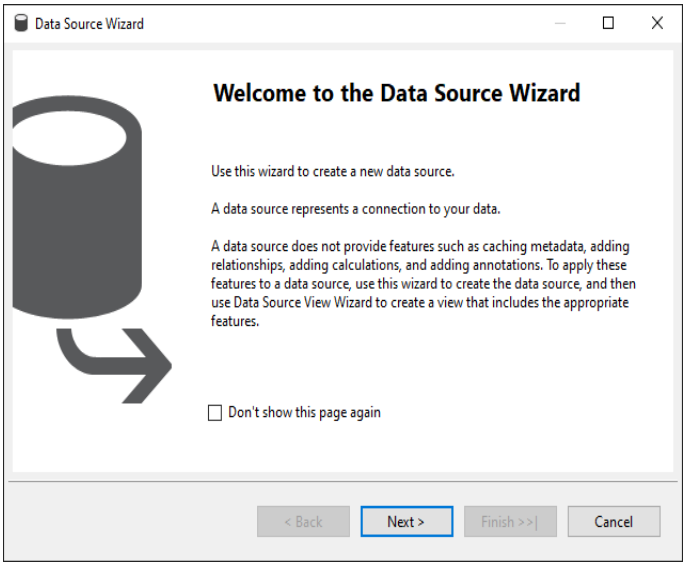
- Connecting flat file destination with sort2.txt file where our output will be seen
- Connection successful

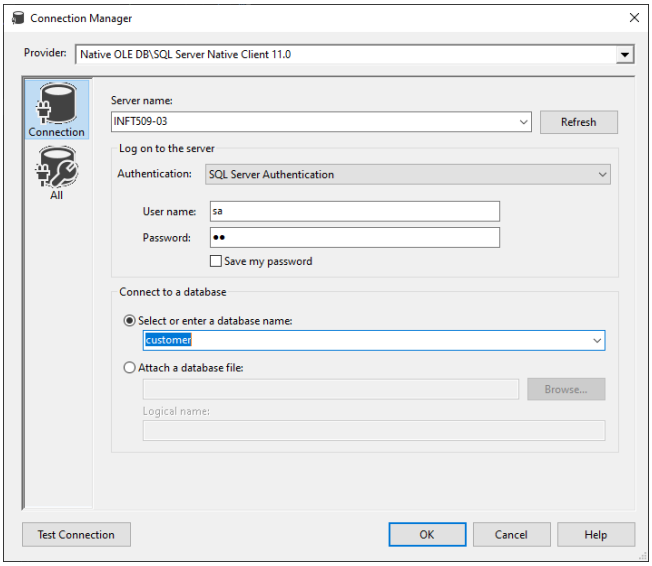


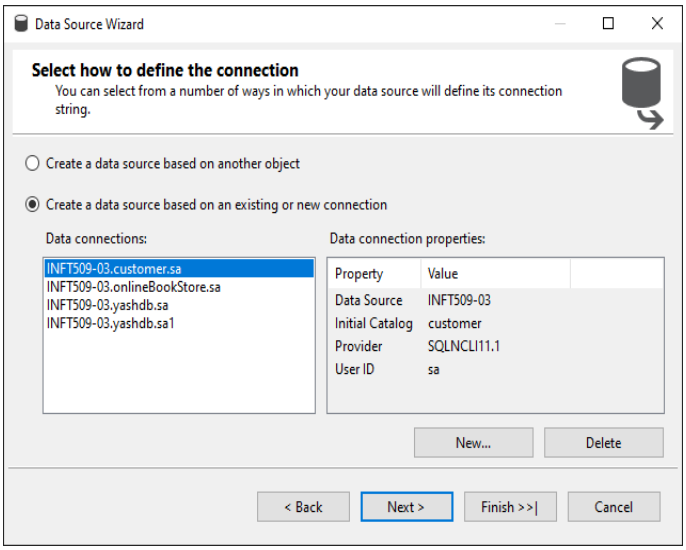


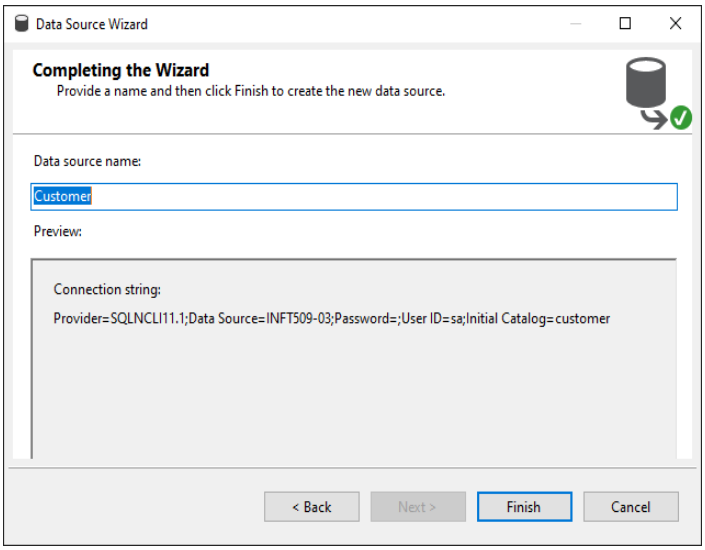
ANALYSIS OF PROJECT

- After creating the Analysis Project on Visual Studio we click on the new data source wizard.
- Then we select the required Database and select the service account.
- Then we give the data source name.

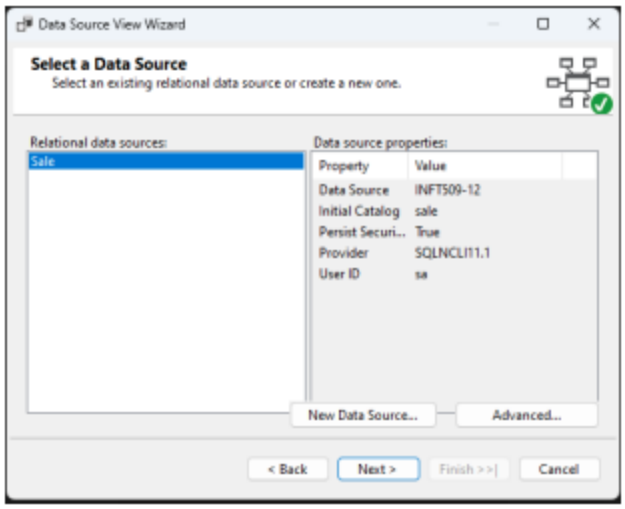


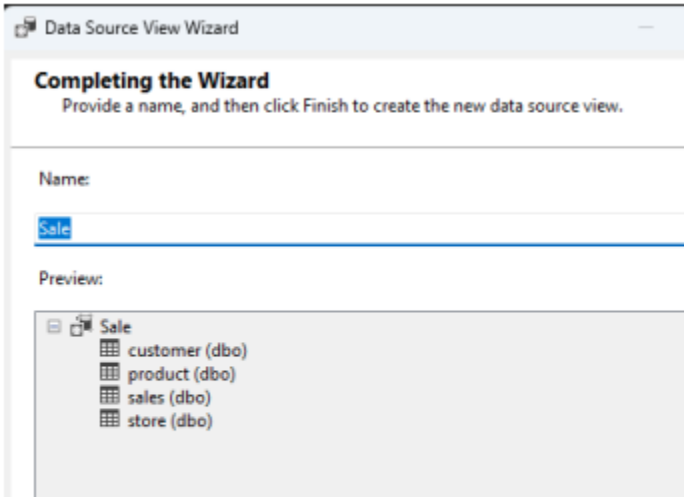






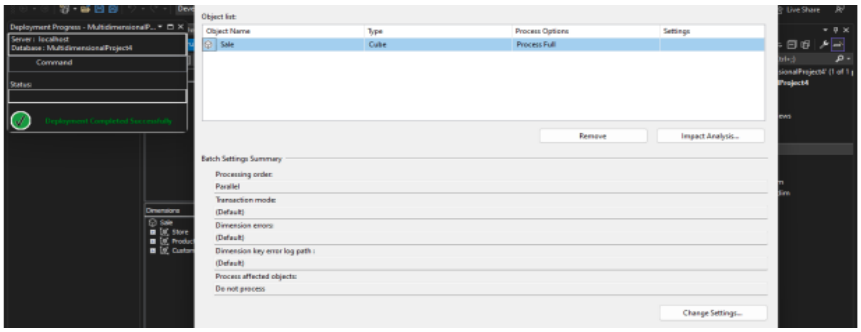
- Then we click on new data source views.
- In that we choose a data source.
- In this we add the tables we will be using for creating star schema.
- We give names to our data source view.



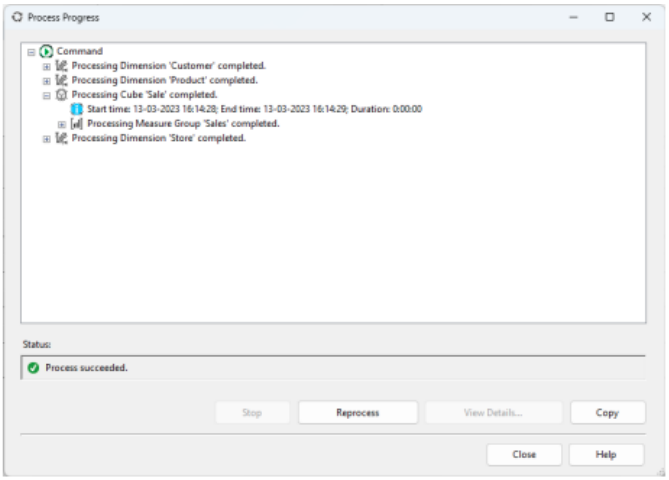


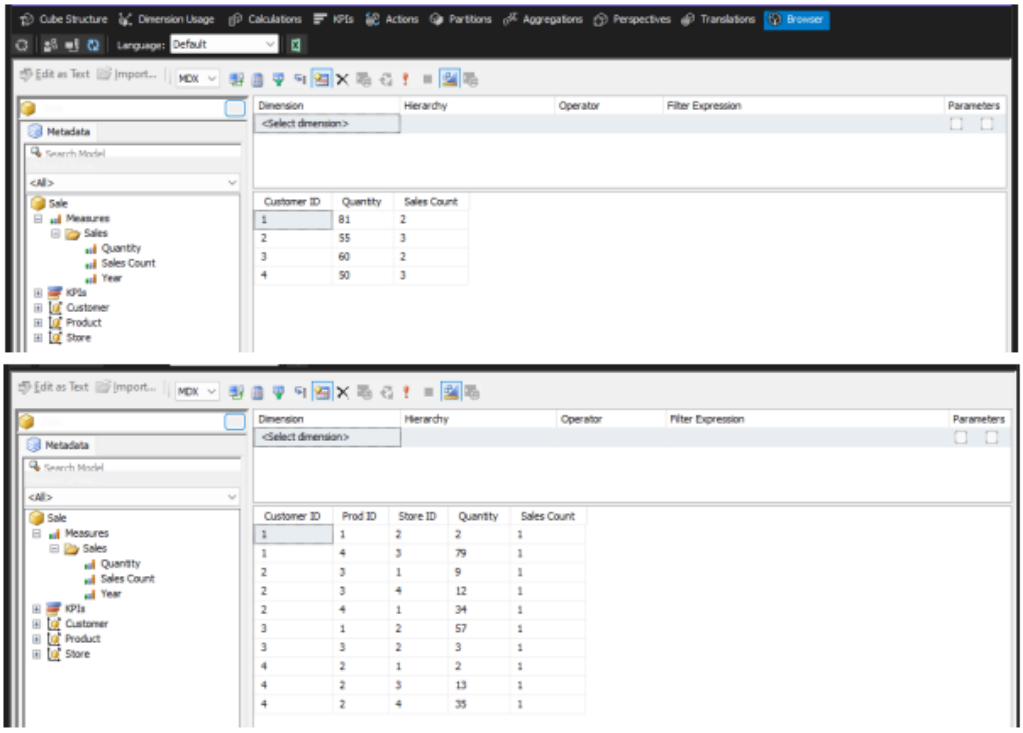
- Now click on New Cube
- Check on All the tables.
- Click on All the measures.
- We can see the Data Source View.

- We now create a new process and click on Run.



- We can see that it was successful and now we drag different parameters to see the analysis.





CONCLUSION: Thus we have successfully designed a Data Warehouse for a given case study and performed ETL and OLAP operations on it.