EXPERIMENT 1

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

**Theory:**

**Data Warehouse:**

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.

A Data Warehouse can be viewed as a data system with the following attributes:

1. It is a database designed for investigative tasks, using data from various applications.
2. It supports a relatively small number of clients with relatively long interactions.
3. It includes current and historical data to provide a historical perspective of information.
4. Its usage is read-intensive.
5. It contains a few large tables.

**ETL**:

ETL stands for Extract, Transform, Load and it is a process used in data warehousing to extract data from various sources, transform it into a format suitable for loading into a data warehouse, and then load it into the warehouse. The process of ETL can be broken down into the following three stages:

1. Extract: The first stage in the ETL process is to extract data from various sources such as transactional systems, spreadsheets, and flat files. This step involves reading data from the source systems and storing it in a staging area.
2. Transform: In this stage, the extracted data is transformed into a format that is suitable for loading into the data warehouse. This may involve cleaning and validating the data, converting data types, combining data from multiple sources, and creating new data fields.
3. Load: After the data is transformed, it is loaded into the data warehouse. This step involves creating the physical data structures and loading the data into the warehouse.

The ETL process is an iterative process that is repeated as new data is added to the warehouse. The process is important because it ensures that the data in the data warehouse is accurate, complete, and up-to-date. It also helps to ensure that the data is in the format required for data mining and reporting.

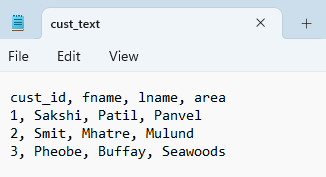
**OLAP:**

OLAP stands for On-Line Analytical Processing. OLAP is a classification of software technology which authorizes analysts, managers, and executives to gain insight into information through fast, consistent, interactive access in a wide variety of possible views of data that has been transformed from raw information to reflect the real dimensionality of the enterprise as understood by the clients.

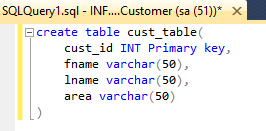
OLAP implements the multidimensional analysis of business information and support the capability for complex estimations, trend analysis, and sophisticated data modeling. It is rapidly enhancing the essential foundation for Intelligent Solutions containing Business Performance Management, Planning, Budgeting, Forecasting, Financial Documenting, Analysis, Simulation-Models, Knowledge Discovery, and Data Warehouses Reporting. OLAP enables end-clients to perform ad hoc analysis of records in multiple dimensions, providing the insight and understanding they require for better decision making.

**OUTPUTS:**

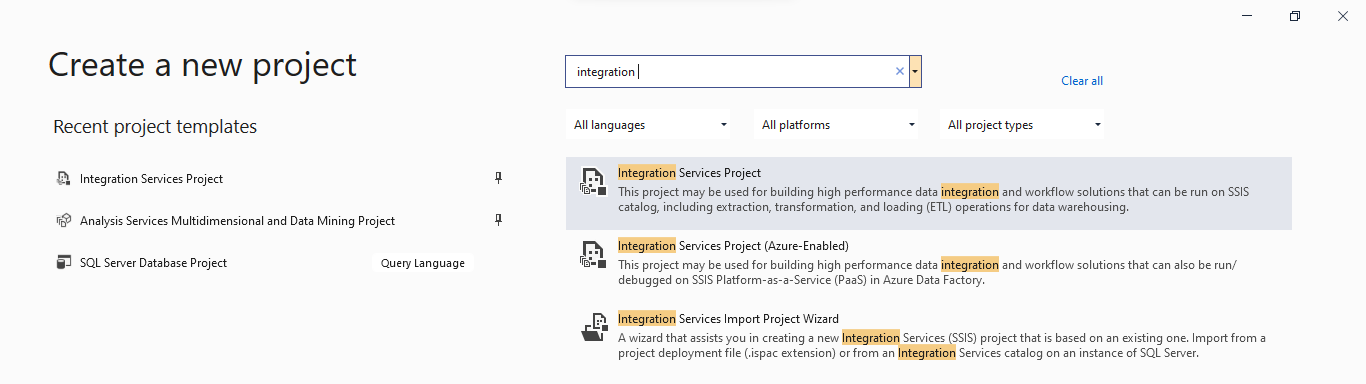
1. Create a text file with the first row filled with attributes and remaining rows filled with desired values.



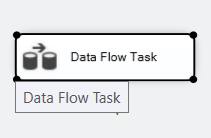
1. Create a table in your database with the same attributes which were filled in a text file in previous steps with appropriate data type.



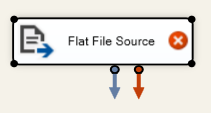
1. Open Visual Studio 2022, create a new project and search for an integration service Project.



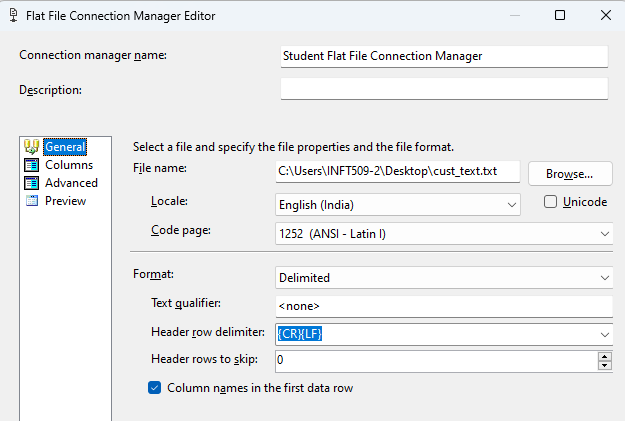
1. Drag Data Flow Task to workspace.

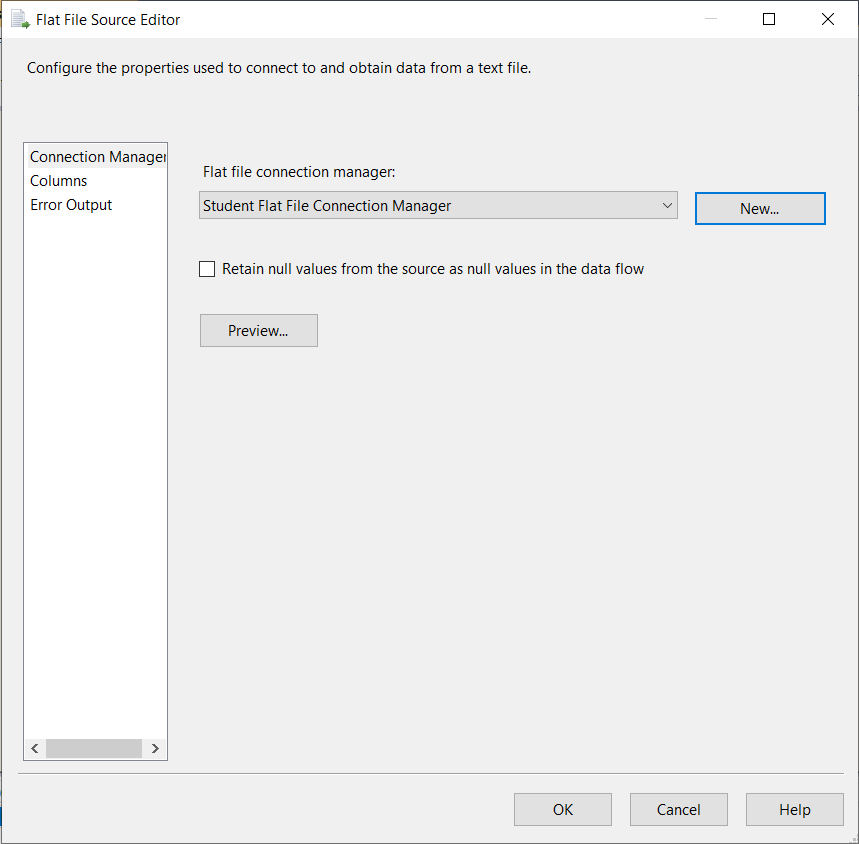


1. Click on Data Flow to go into its workspace and then drag Flat File Source to its Workspace.

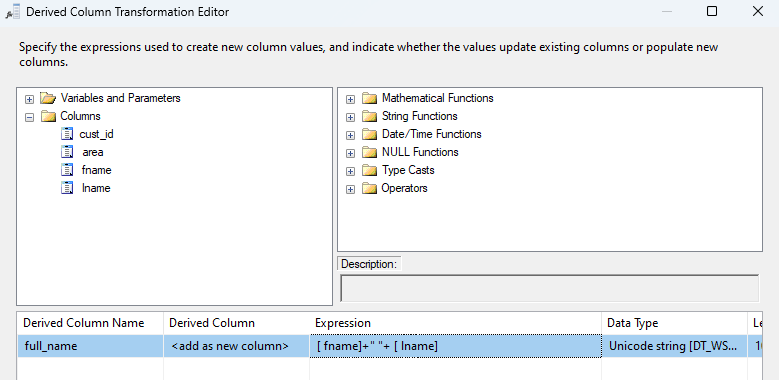


1. After Clicking on Flat File Source click on New to add the complaint.csv which was created in the first step.

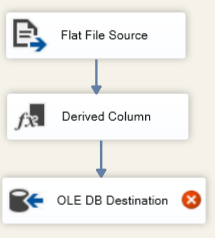


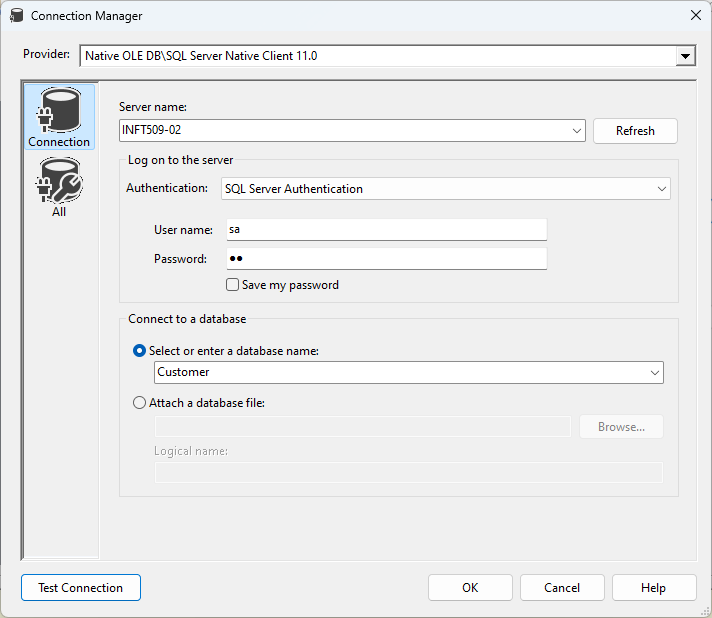


1. Drag Derived Column to workspace and then connect Flat File Source to Derived Column.
2. After Clicking on the Derived Column we see the following dialog box, add all the column names to be mapped and transformed with needful transformation in the expression column.

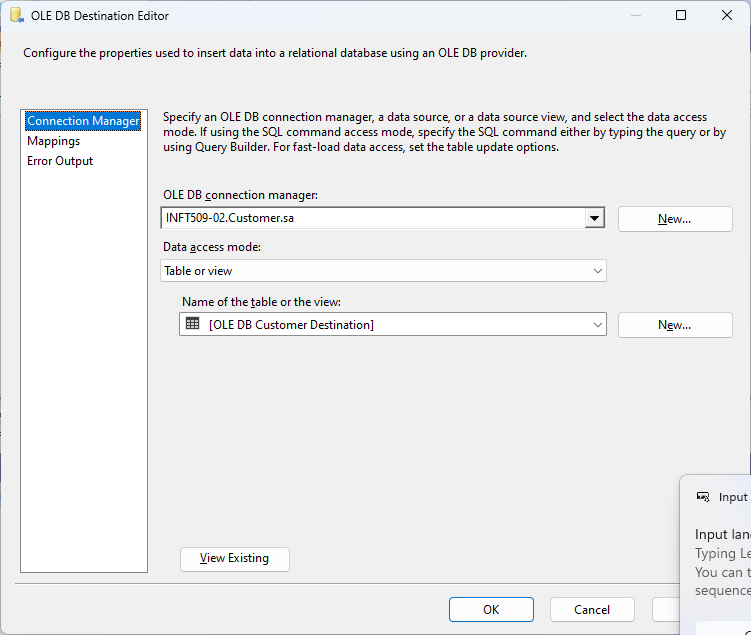


1. After Dragging OLE DB Destination connects it with Derived Column we have to define a database which contains the table to transform.

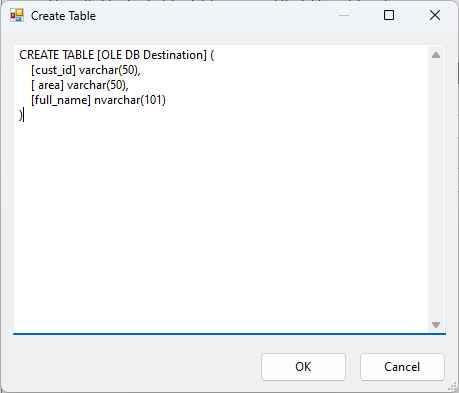


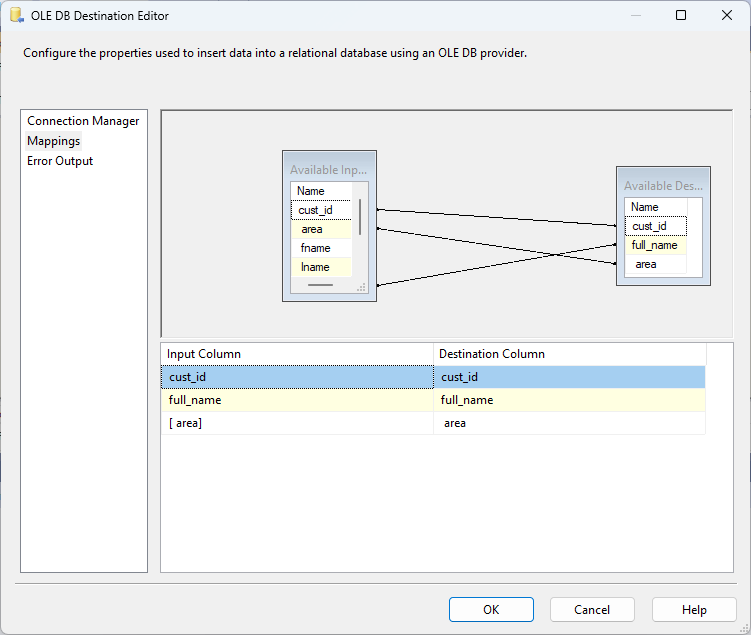


1. Make a new destination table.

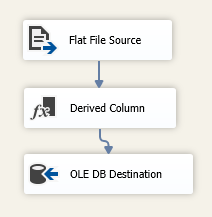
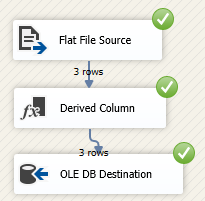


1. Check if there are any errors in mappings.

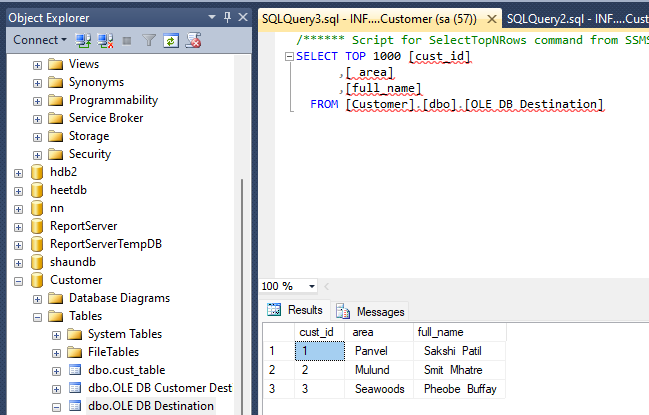




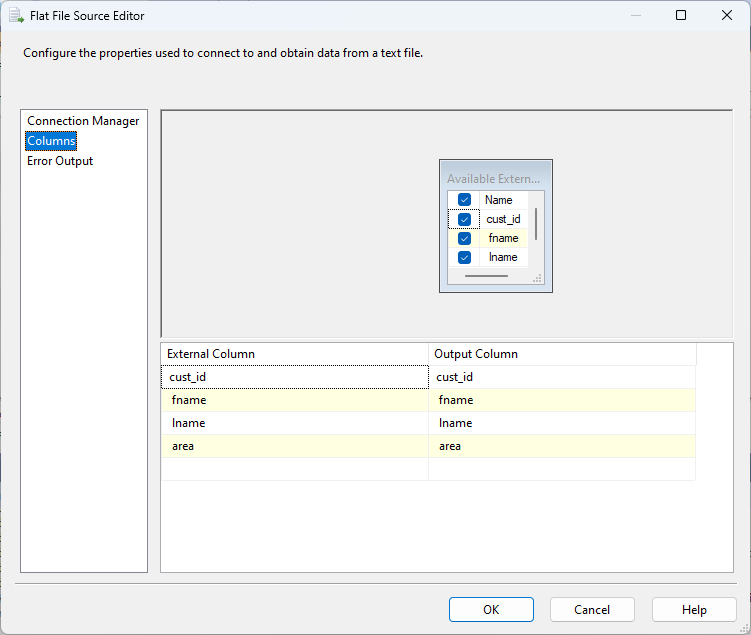
1. Check if there are any errors, if not then click on the run icon.

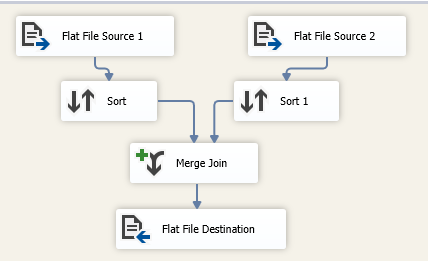
1. After execution Check if there are any errors and warnings and check if the state is ‘ready’.
2. In dbo.test37 Table we can see new entries are mapped as per Derived Columns expressions.



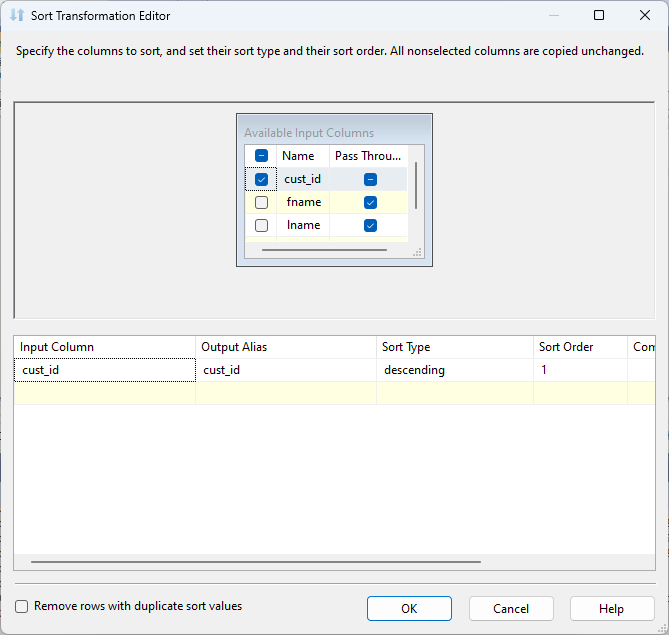
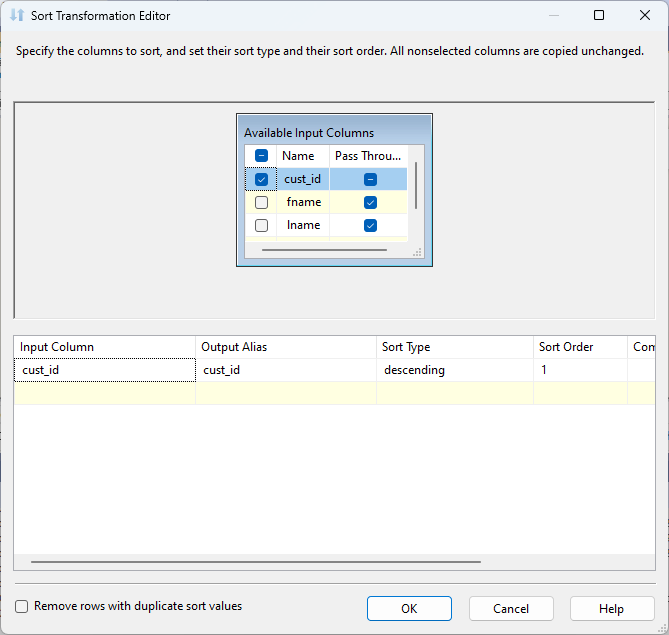
1. We are then taking two flat file sources, adding the same txt file to it and checking for correct mappings



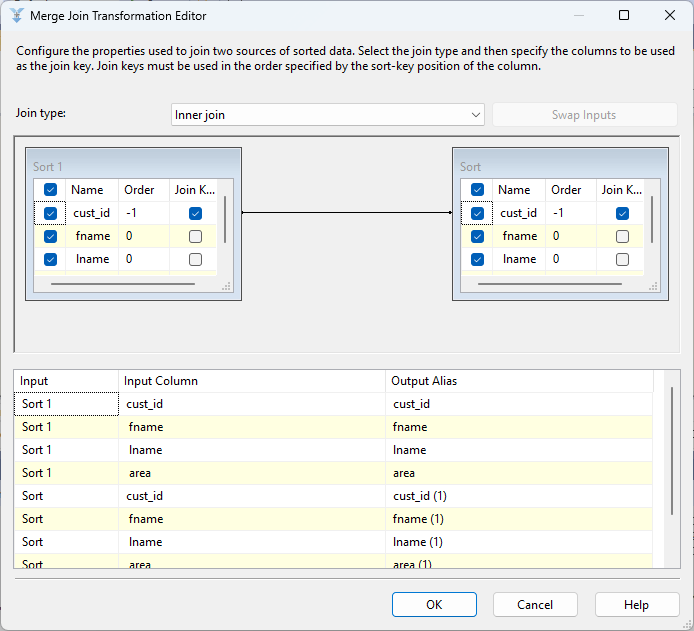
1. Create the following Structure for Performing Merge Join.



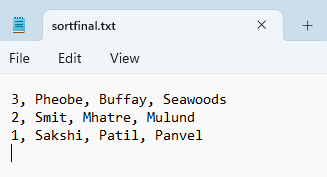
1. In Sort we will be using cust\_id as a parameter to rearrange the file contents in descending order.

1. So we get Merge Join Output of above Operation
2. We are then taking a dataset file and checking for the columns being used for sorting.

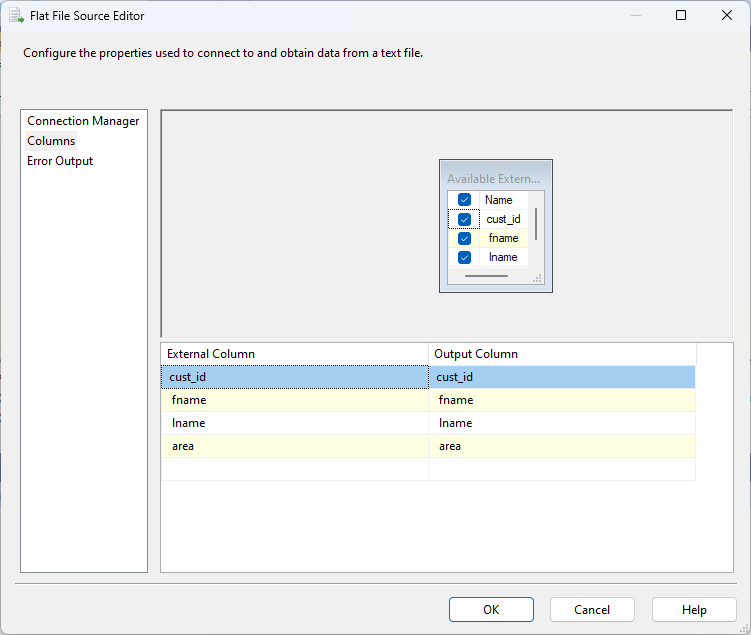


1. After performing necessary operations we have to look for the errors before debugging and then click on ‘start’.
2. We are choosing where the output will be stored i.e sort.txt in Flat File Destination.
3. We will get the image below after clicking on the start of the error free setup
4. This is the output of Sorted File.

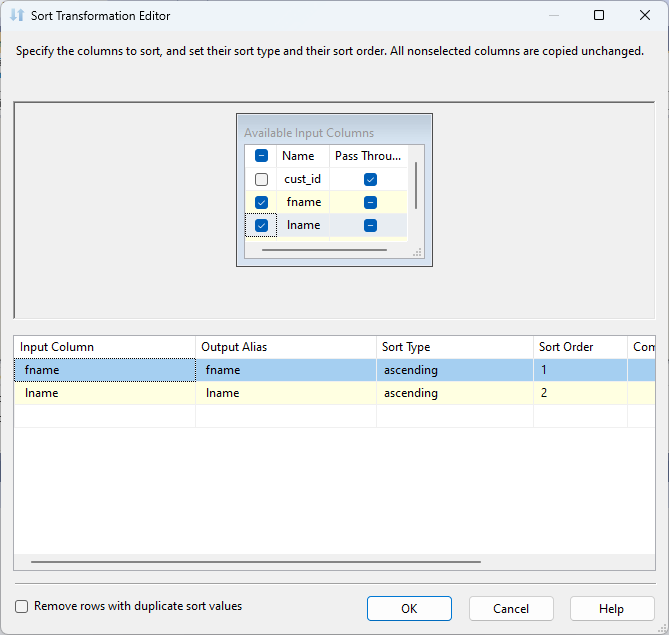


SORTING

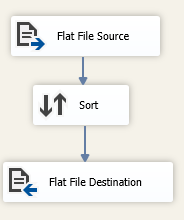
1. We are then taking a dataset file and checking for the columns being used for sorting.



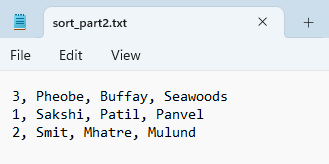
1. In Sort we will be using fname and name with priority 1 and 2 respectively as a parameter to rearrange the file contents.



1. We are choosing where the output will be stored i.e sort\_part2.txt in Flat File Destination.
2. Structure for the Sorting operation:

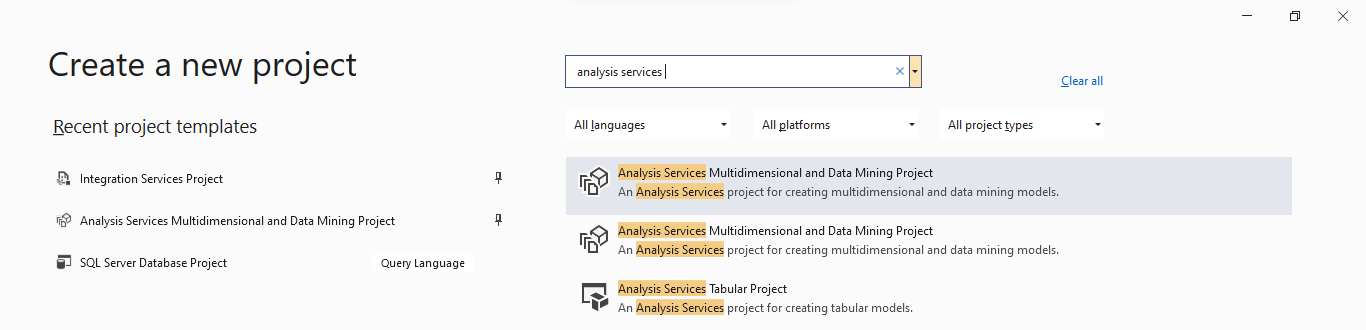


1. This is the output of Sorted File

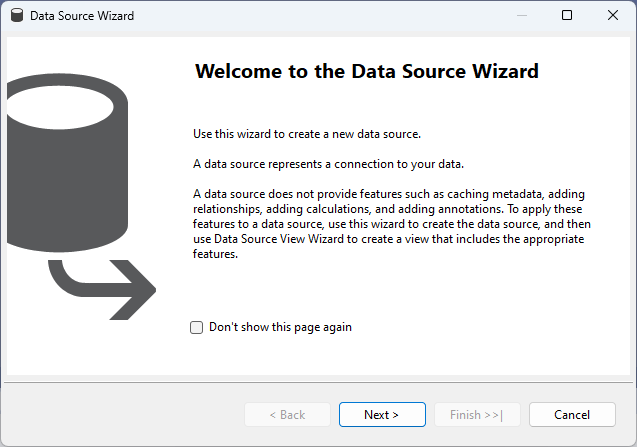


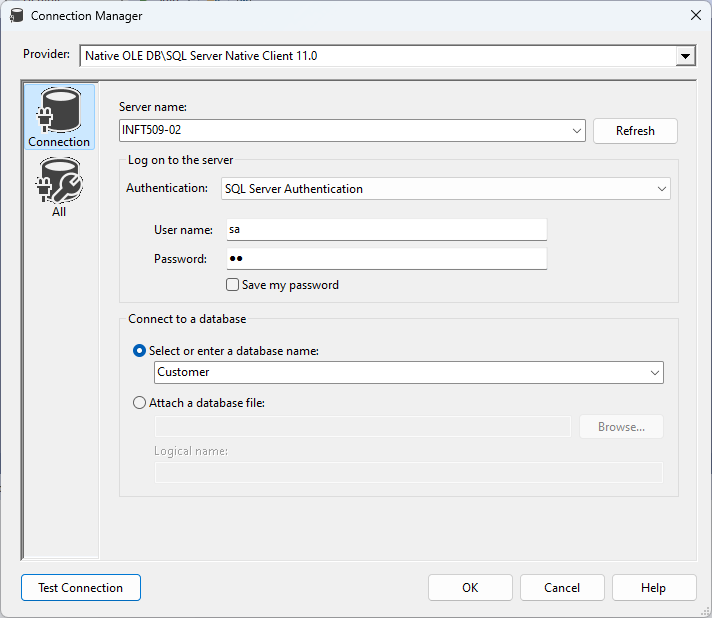
**ANALYSIS PROJECT**

1. After creating the Analysis Project on Visual Studio we click on the new data source wizard.

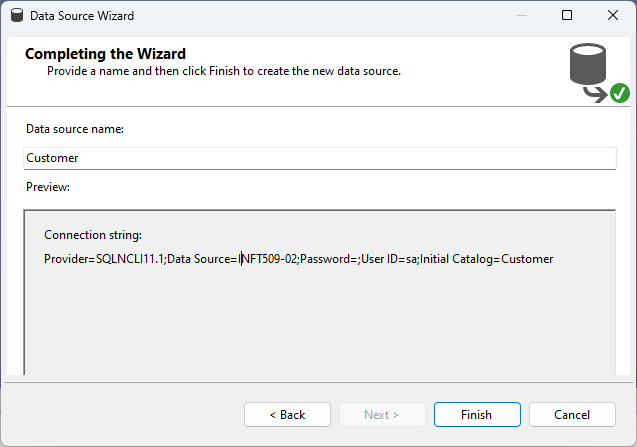


1. Then we select the required Database.

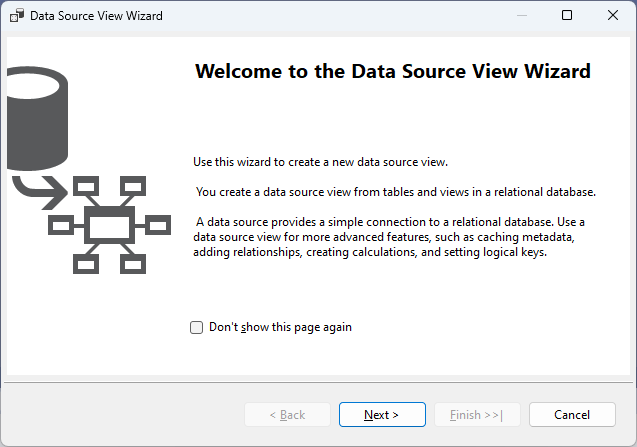




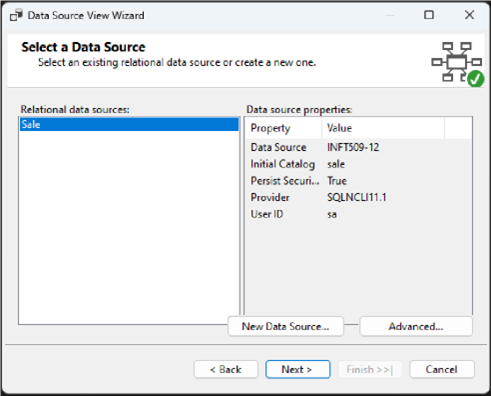
1. Then we give the data source name



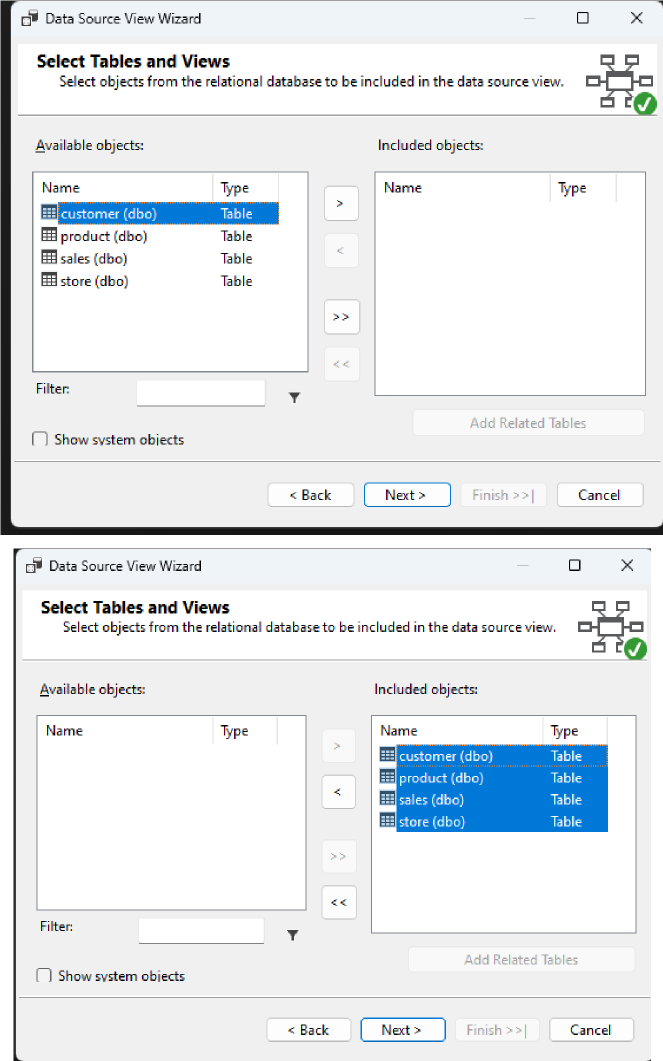
1. Then we click on new data source views

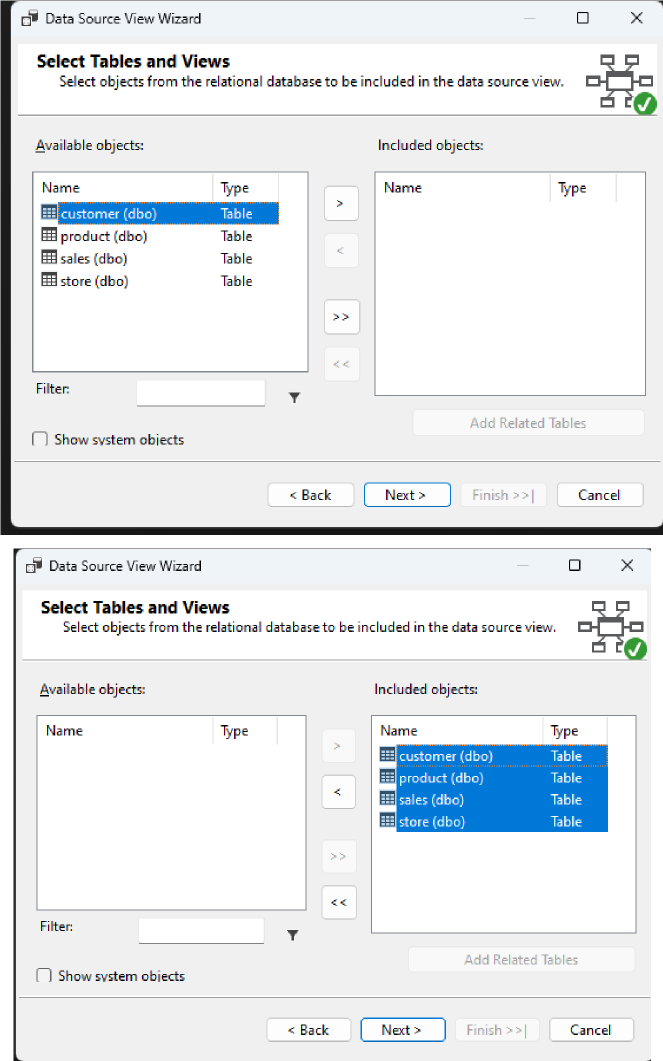


1. In that we choose data source

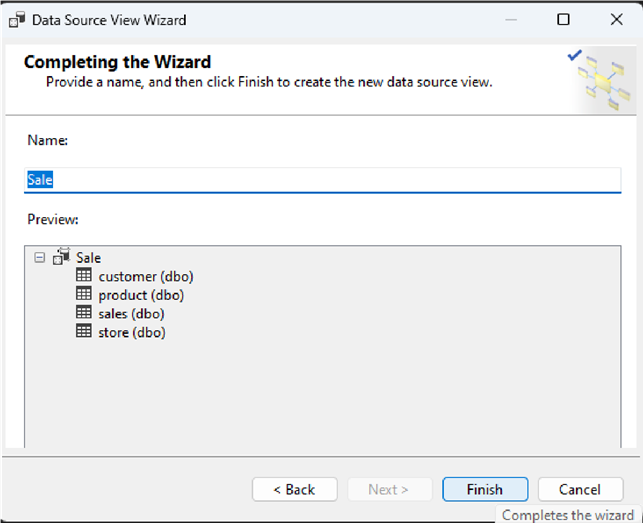


1. In this we add the tables we will be using for creating star schema.

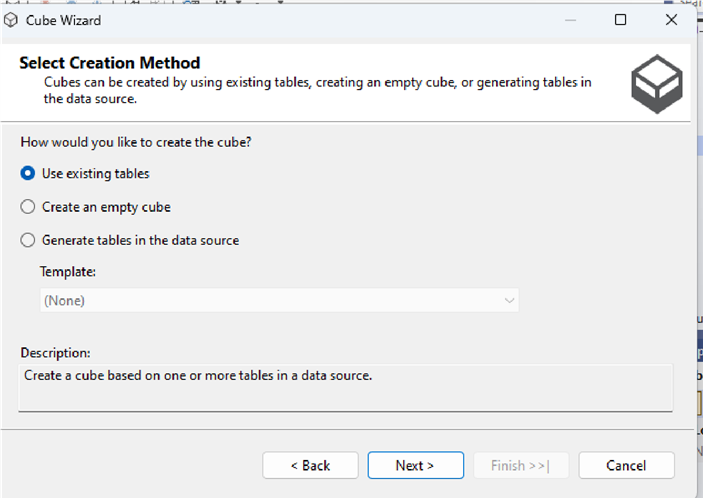




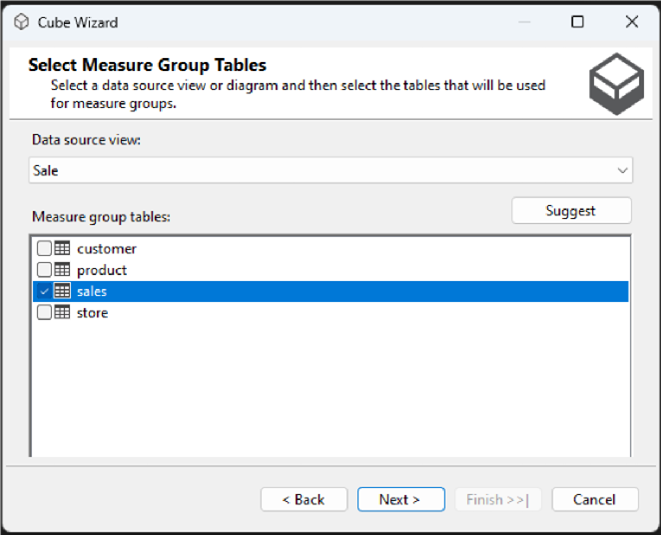
1. We give name to our data source view



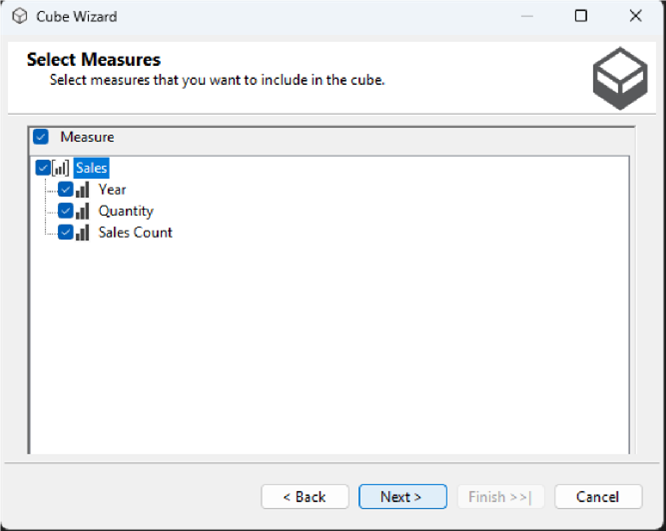
1. Now click on New Cube

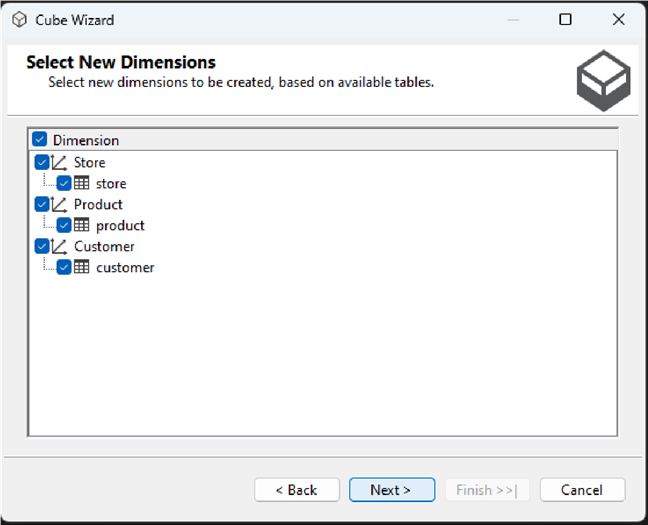


1. Check on All the tables

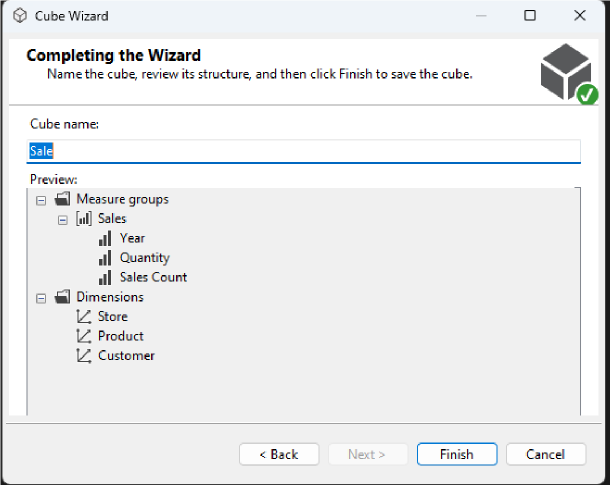


1. Click on All the measures

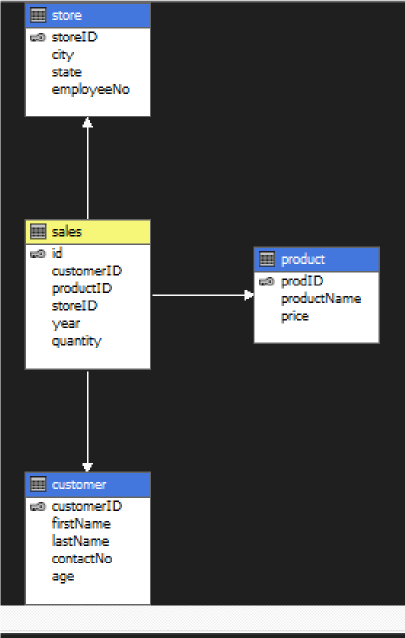




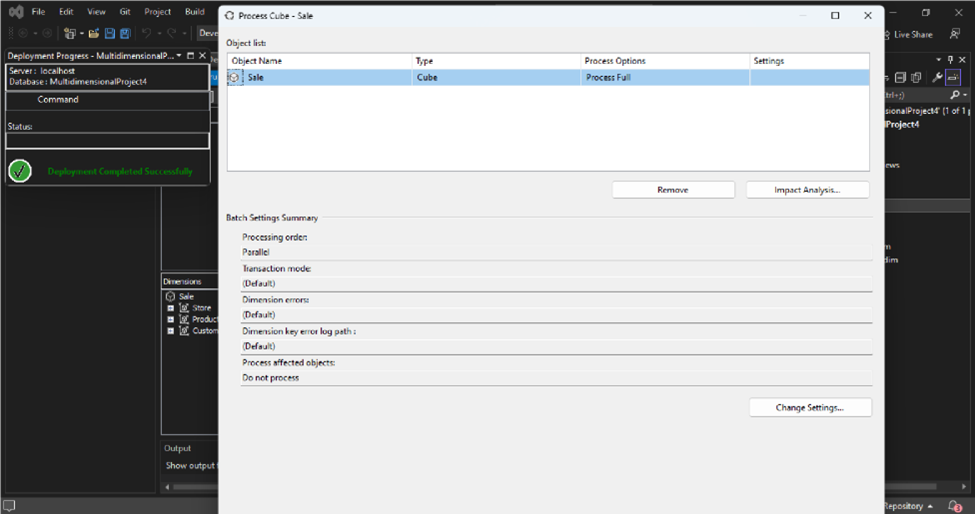
1. Give name to the cube

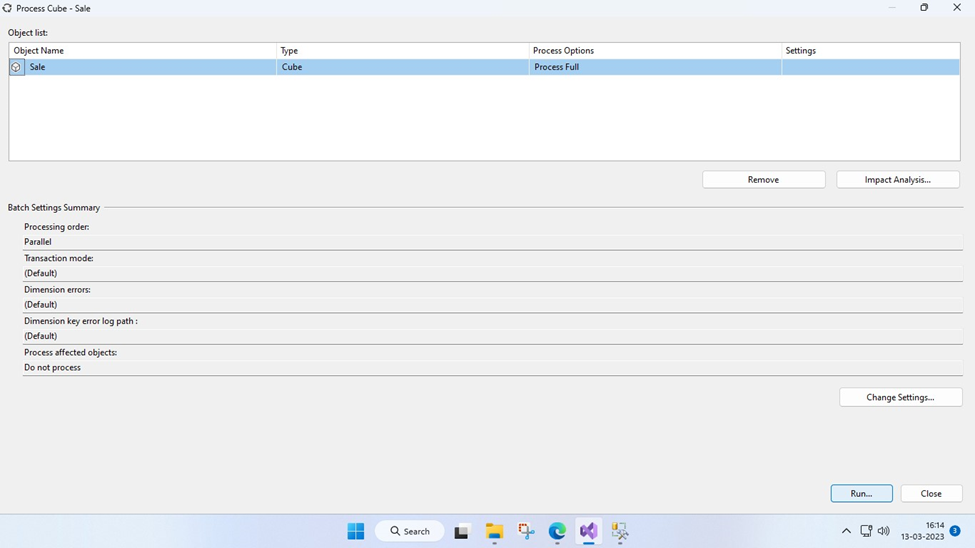


1. We can see the Data Source View.

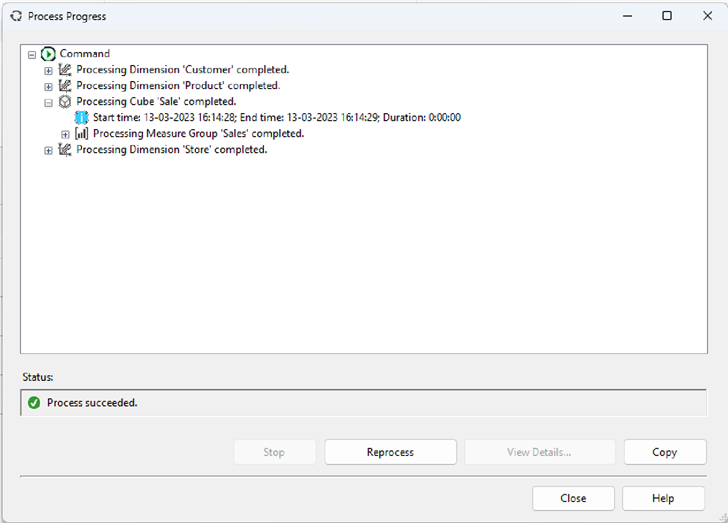


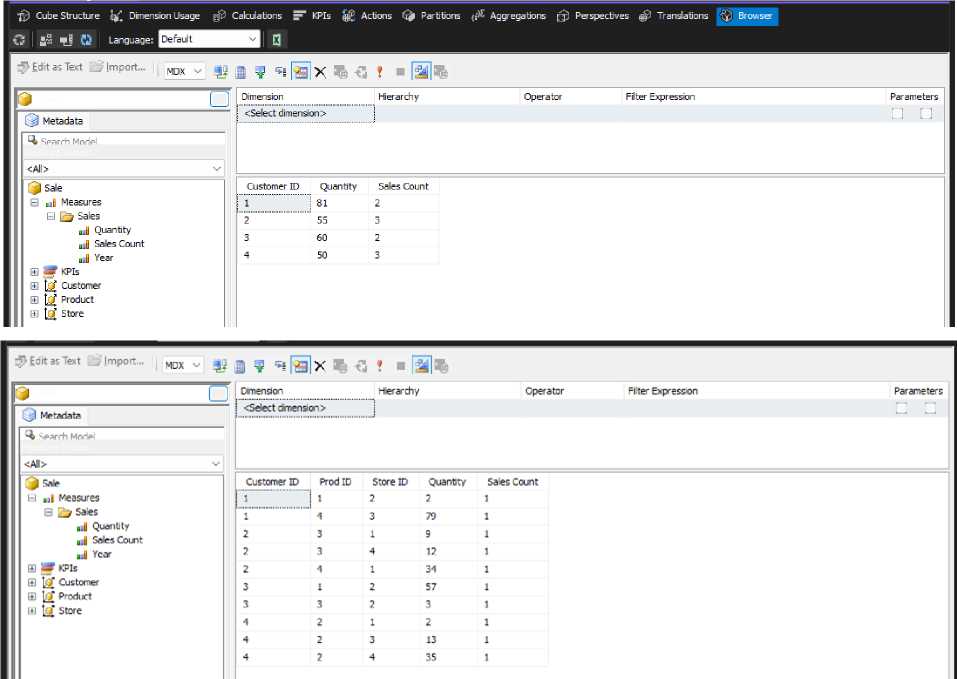
1. We now create a new process and click on Run.





1. 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.