

## Assignment – V (SSIS)

### Objective:

This assignment aims to test your understanding and practical skills in using SQL Server Integration Services (SSIS) for advanced ETL processes, including integration with Data Warehouses (DWH), data warehouse migrations, the use of Script Task, pivot transformations, and implementing incremental loads and advanced ETL processes involving transformations, merge and fuzzy lookup operations, error handling, SSIS checkpoint and pivot operations, event handling, loops, and expressions.

Dataset: You can use **sales\_data\_sample.csv** or you are also open to generate new dataset using **mockaroo.com** (if required).

### Submit the following in a single zip file:

1. All SSIS package files for each task named as per task names.
2. Documentation with screenshots and explanations for each step of the tasks.
3. Any additional scripts or configuration files used in the tasks.

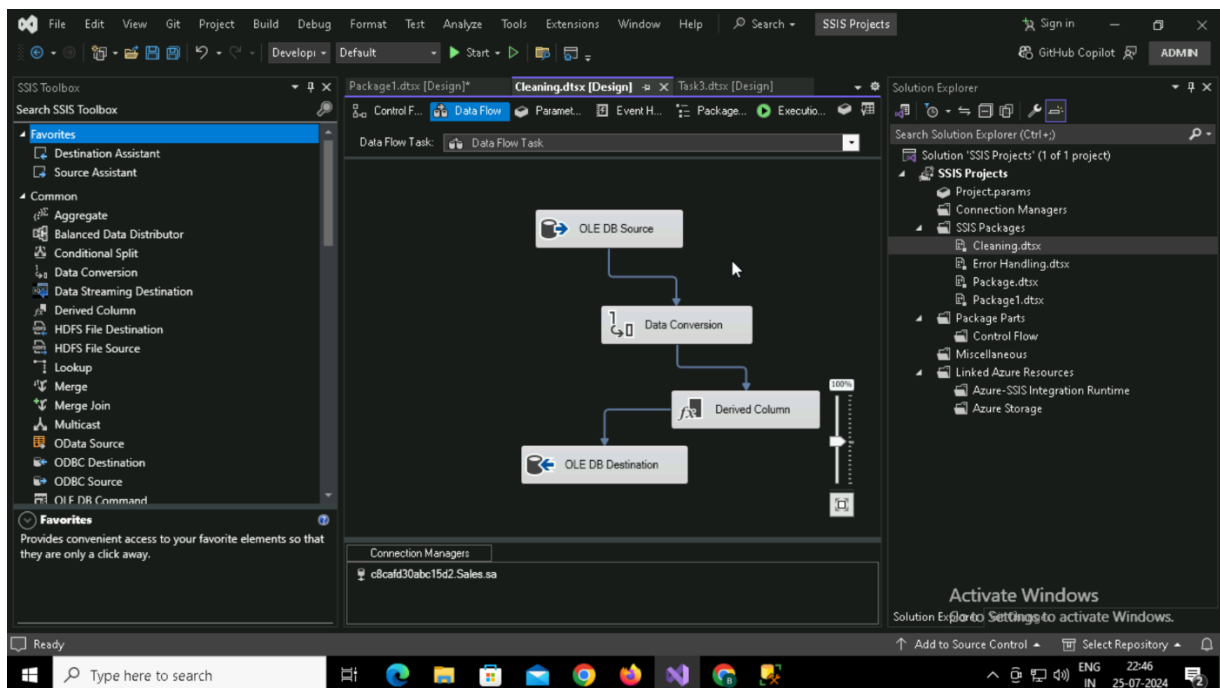
### Task 1: Integration with ETL Data Warehouse (DWH)

Scenario: Your company has a data warehouse designed to consolidate data from various sources for analytical purposes. You need to create an SSIS package that extracts data from a transactional database and loads it into the data warehouse.

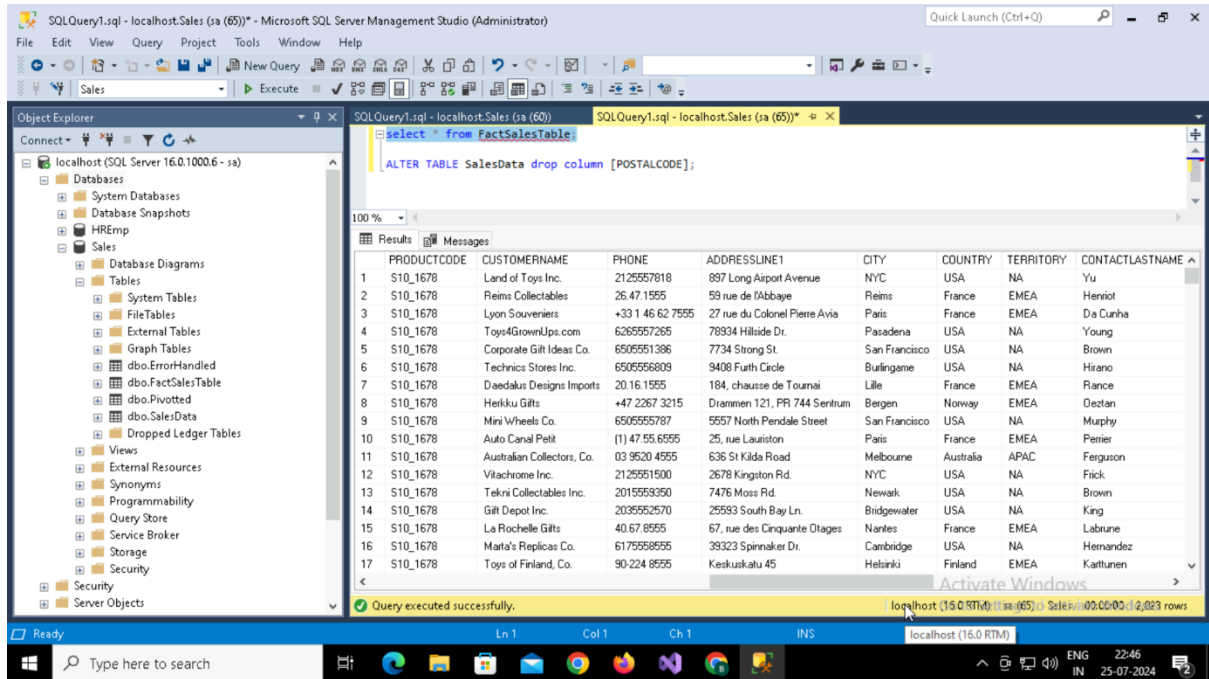
#### Requirements:

1. Create a Connection Manager to connect to the transactional database and the data warehouse.
2. Extract Data from a transactional table (e.g., SalesData) using an OLE DB Source.
3. Transform Data:  
Apply necessary transformations such as data type conversions, data cleansing, and calculations.
4. Load Data into the data warehouse (e.g., FactSales table).

#### Initial cleaning of data after loading



## Output

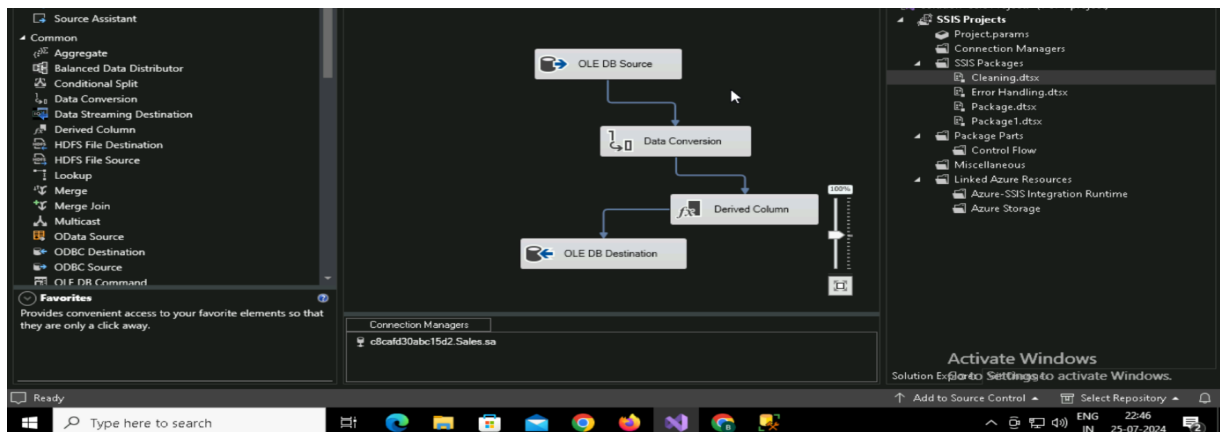


## Task 2: Data Warehouse Migrations

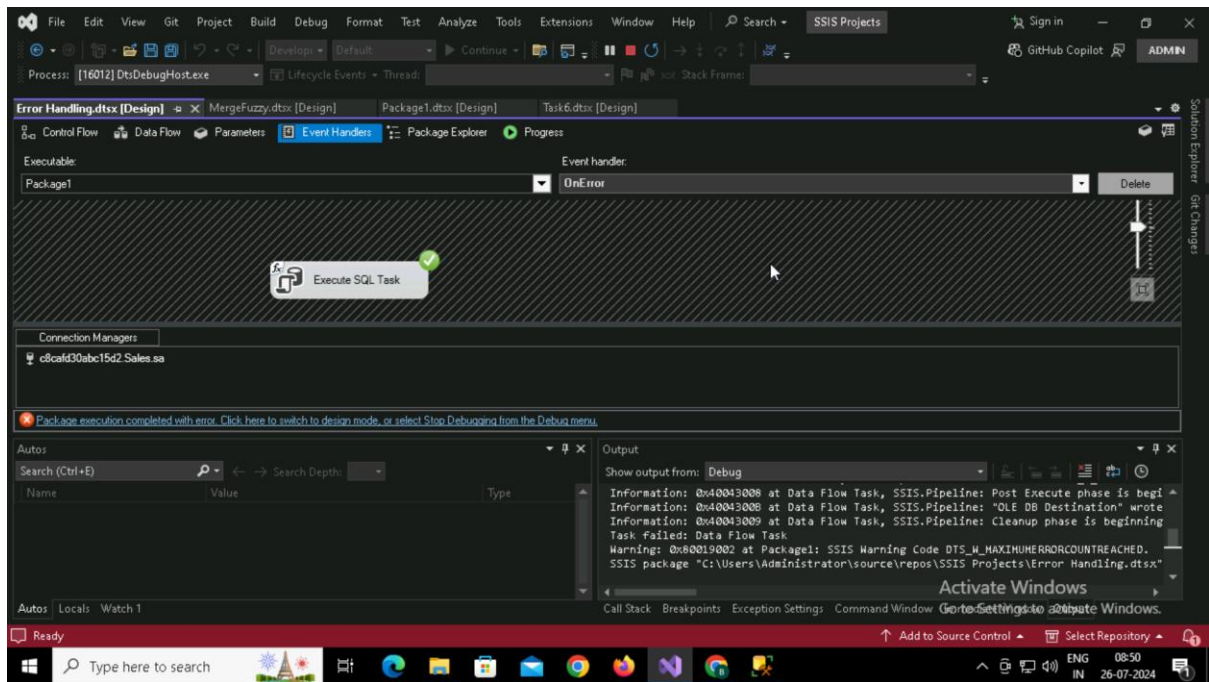
Scenario: Your organization is migrating its data warehouse from one server to another. You need to create an SSIS package that facilitates this migration.

Requirements:

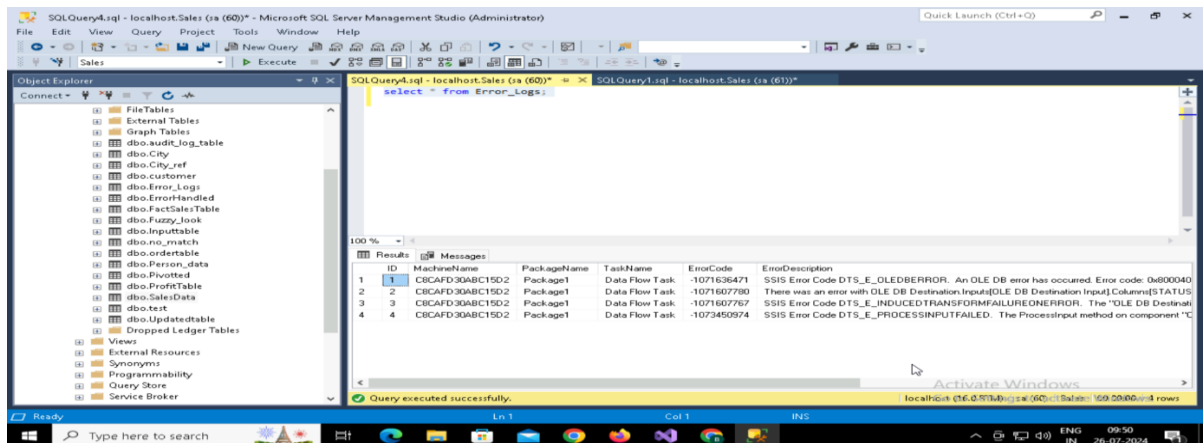
1. Create Connection Managers for both the source and destination data warehouses.
2. Transfer Data from the source data warehouse to the destination using the Data Flow Task.
3. Ensure Data Integrity:
  - a) Include checks and balances to ensure data is correctly migrated.
  - b) Log the success or failure of the migration process.

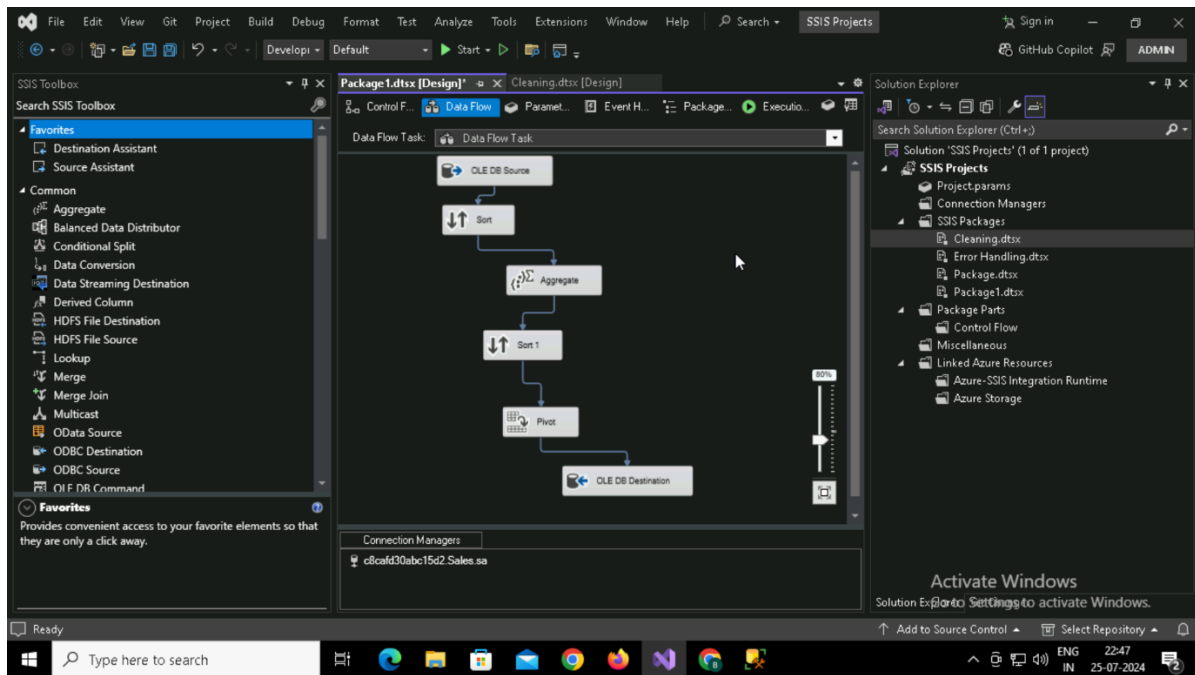


## Output after error handling

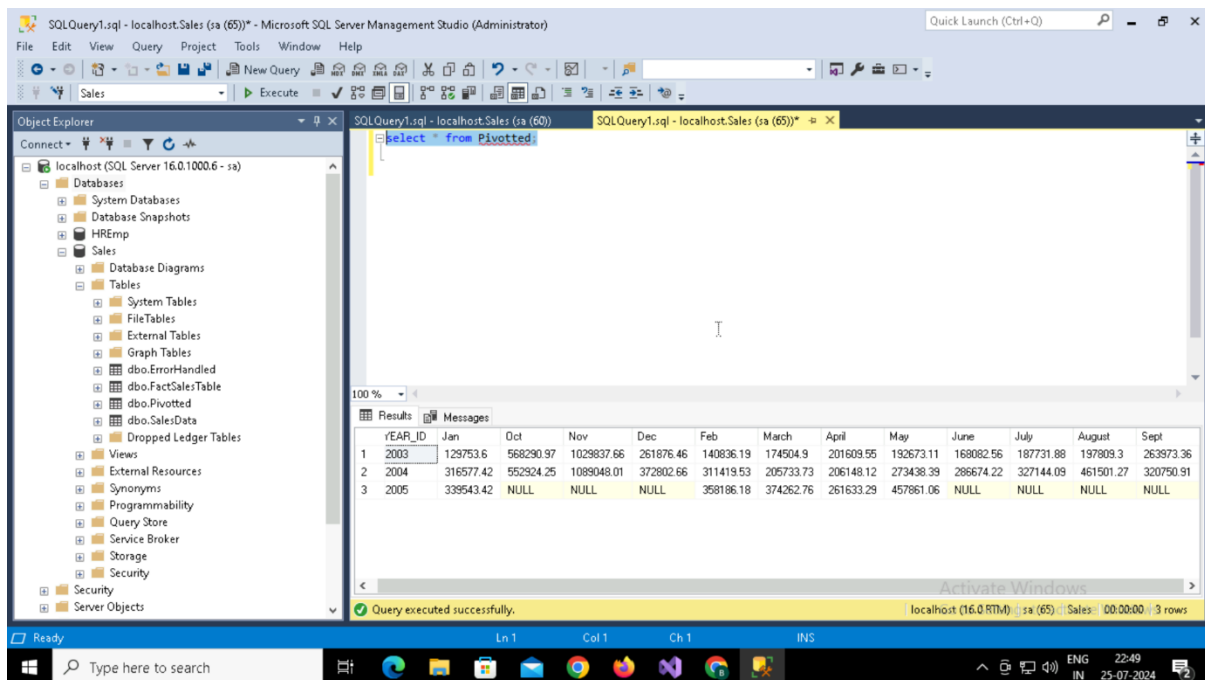


## Error Logs





## Pivot Table



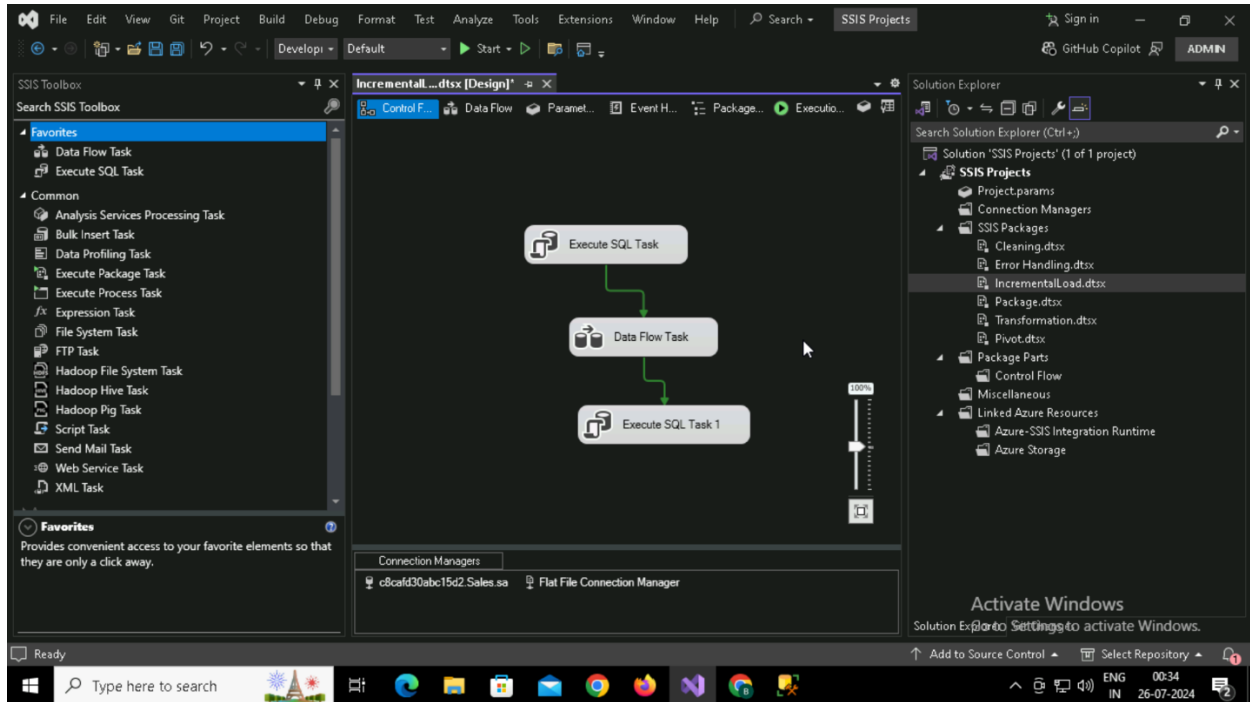
## Task 4: Incremental Load

Scenario: To optimize ETL processes, you need to implement an incremental load to update only the changed data in the data warehouse.

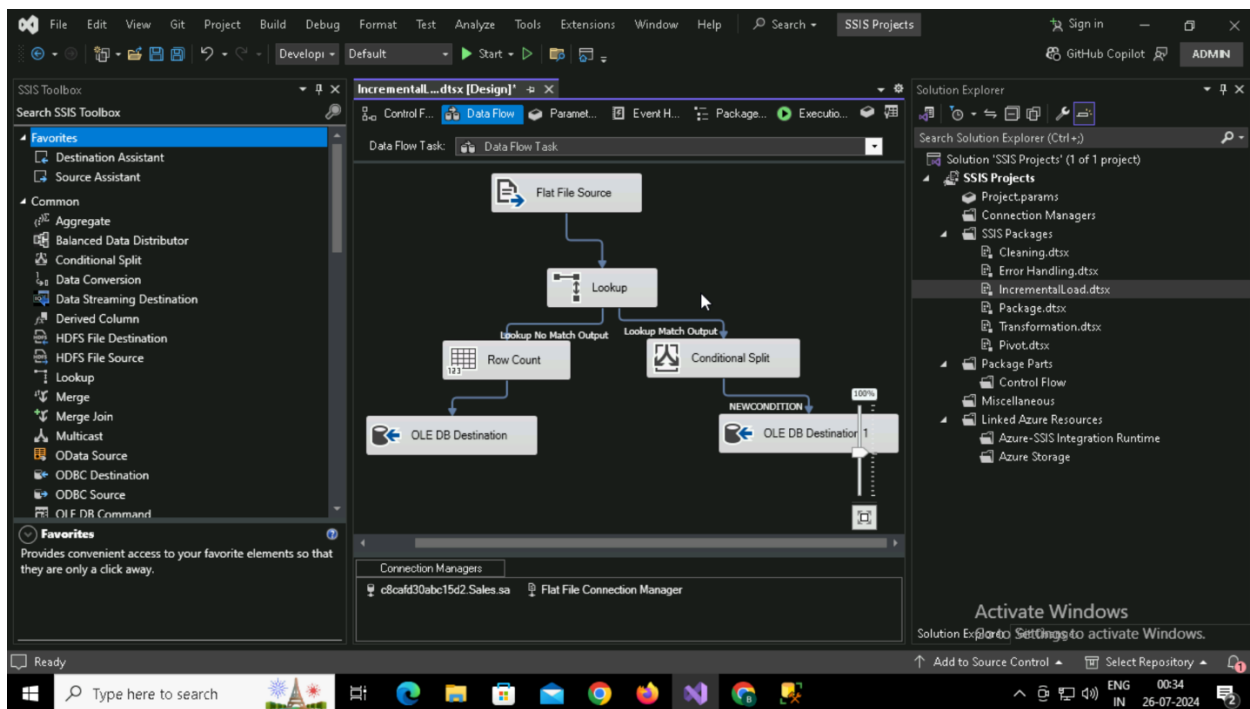
Requirements:

1. Identify Changed Data: Use methods such as timestamps, change data capture using lookup, or checksums.
2. Extract Only the Changed Data from the source.
3. Update the Data Warehouse with the new and changed data only.

## Control flow of incremental load



## Data Flow of Incremental Load



## Task 5: Transformations

Scenario: Your company needs to transform raw data into a format suitable for reporting. You need to perform multiple transformations within an SSIS package.

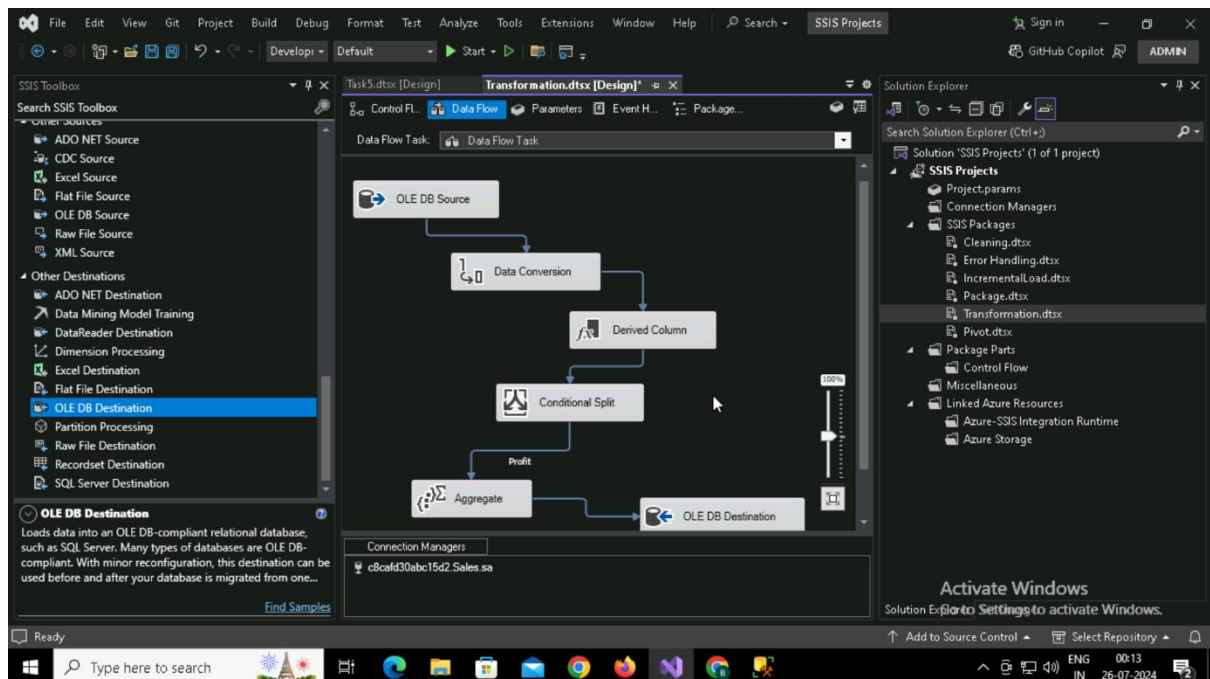
Requirements:

1. Extract Data from a source table using an OLE DB Source.
2. Apply Transformations such as:

- Data Conversion
- Derived Column
- Conditional Split
- Aggregate

3. Load Transformed Data into a destination table.

## Data Flow of Transformation



## Output – Sum of profit in each productline

The screenshot shows the query result in SQL Server Enterprise Manager. The query is:

```
USE [Sales]
GO
SELECT [PRODUCTLINE]
FROM [dbo].[ProfitTable]
GO
```

The result set is as follows:

| PRODUCTLINE      | sum_SALES         |
|------------------|-------------------|
| Ships            | 307687.94         |
| Trains           | 111555.38         |
| Planes           | 477445.55         |
| Classic Cars     | 3113137.65        |
| Trucks and Buses | 879914.4          |
| Vintage Cars     | 1146286.32        |
| Motorcycles      | 758931.8900000001 |



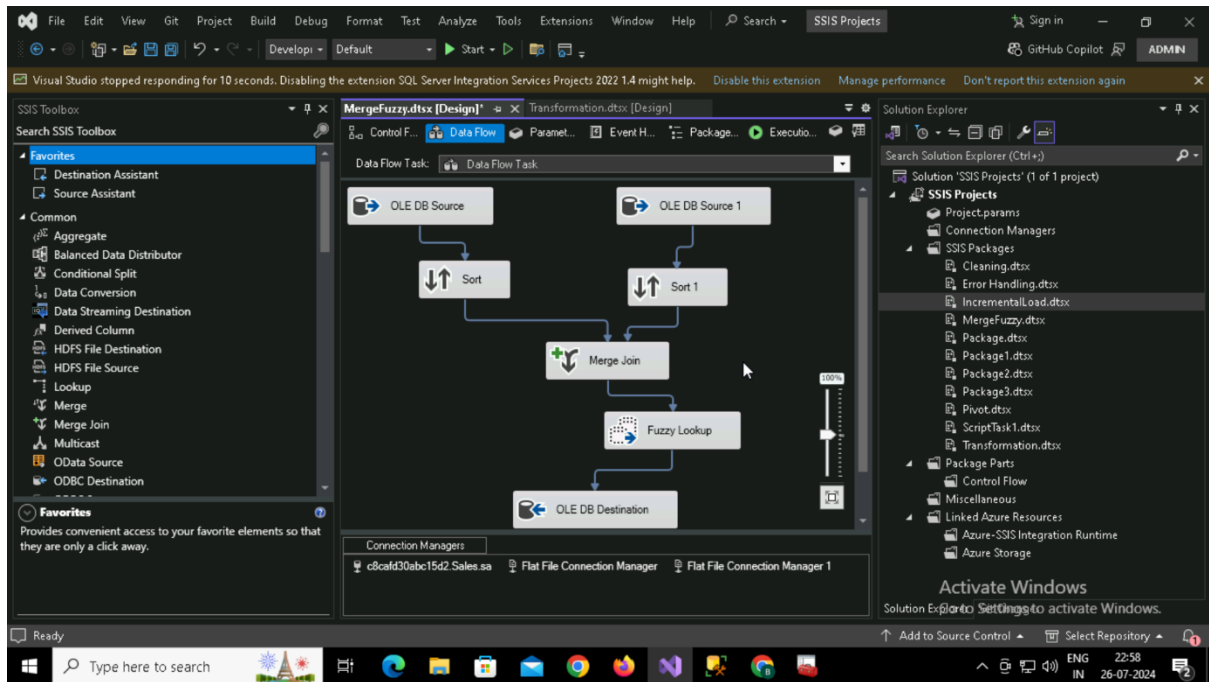
## Task 6: MERGE & FUZZY LOOKUP

Scenario: You need to merge two datasets and use fuzzy matching to handle potential duplicates.

Requirements:

1. Extract Data from two source tables using OLE DB Sources.
2. Apply a Merge Join to combine the datasets based on a common key.
3. Use Fuzzy Lookup to identify and resolve duplicates in the merged data.
4. Load the Cleaned Data into a destination table.

### Fuzzy LookUp Dataflow



### Fuzzy LookUp Output

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure, including tables like System Tables, FileTables, External Tables, Graph Tables, and various user-defined tables such as dbo.audit\_log\_table, dbo.City\_ref, dbo.City\_lookup, dbo.Customer, dbo.Customers, dbo.Error\_Logs, dbo.ErrorHandled, dbo.FactSalesTable, dbo.Fuzzy\_lookup, dbo.Inputtable, dbo.no\_match, dbo.orderable, dbo.Person\_data, dbo.Pivoted, dbo.ProfitTable, dbo.SalesData, dbo.test, dbo.UpdatedTable, and Dropped Ledger Tables. Views, External Resources, and Synonyms are also listed.

The SQL Query window in the center displays the following query:

```
UPDATE tb1
SET tb1.QUANTITYORDERED = tb2.QUANTITYORDERED
FROM [dbo].[transact_table] tb1 inner join [dbo].[no_match] as tb2
ON tb1.ORDERNUMBER = tb2.ORDERNUMBER
SET @updated = @ROWCOUNT
INSERT INTO [dbo].[audit_log_table]
SELECT 'IncrementalLoad.dtsx',@updated,@updated,getdate()

select * from Fuzzy_lookup
```

The Results window at the bottom shows the output of the query, which is a table with the following columns: first\_name, last\_name, email, gender, City, City (1), \_Similarity, \_Confidence, \_Similarity\_id, and \_Similarity\_City. The table contains 8 rows of data.

| first_name | last_name | email       | gender                  | City        | City (1)         | _Similarity      | _Confidence | _Similarity_id | _Similarity_City |           |
|------------|-----------|-------------|-------------------------|-------------|------------------|------------------|-------------|----------------|------------------|-----------|
| 1          | Ashley    | Lockey      | alockey00@gov.uk        | Genderqueer | Baicun           | Baicun           | 1           | 1              | 1                |           |
| 2          | Skipton   | Collaton    | scollaton9@mtv.com      | Male        | Belogorsk        | Belogorsk        | 1           | 1              | 1                |           |
| 3          | Gilberto  | Enrico      | gennico2@desdev.cn      | Male        | Buenavista       | Buenavista       | 1           | 1              | 1                |           |
| 4          | Karly     | Castagnasso | kcastagnasso2@home.pl   | Female      | Erpeldange       | Erpeldange       | 0.9396087   | 0.8627352      | 1                | 0.9090909 |
| 5          | Ledley    | Colax       | lcolax2@yahoo.com       | Male        | Lukunor          | Lukunor          | 1           | 1              | 1                | 1         |
| 6          | Adi       | Barnaby     | abarnaby2u@lavars.me    | Female      | Rumat Heib       | Rumat Heib       | 1           | 1              | 1                | 1         |
| 7          | Milissent | Bernath     | mbernath2w@auds.org.au  | Female      | Francistown      | Francistown      | 1           | 1              | 1                | 1         |
| 8          | Simone    | Lumbers     | slumbers2w@nehvibes.com | Male        | Lautaro          | Lautaro          | 1           | 1              | 1                | 1         |
| 9          | Berkly    | Leele       | bleele2u@oaic.gov.au    | Bigender    | Staraya Stanitsa | Staraya Stanitsa | 1           | 1              | 1                | 1         |

The status bar at the bottom indicates that the query was executed successfully, returning 500 rows.

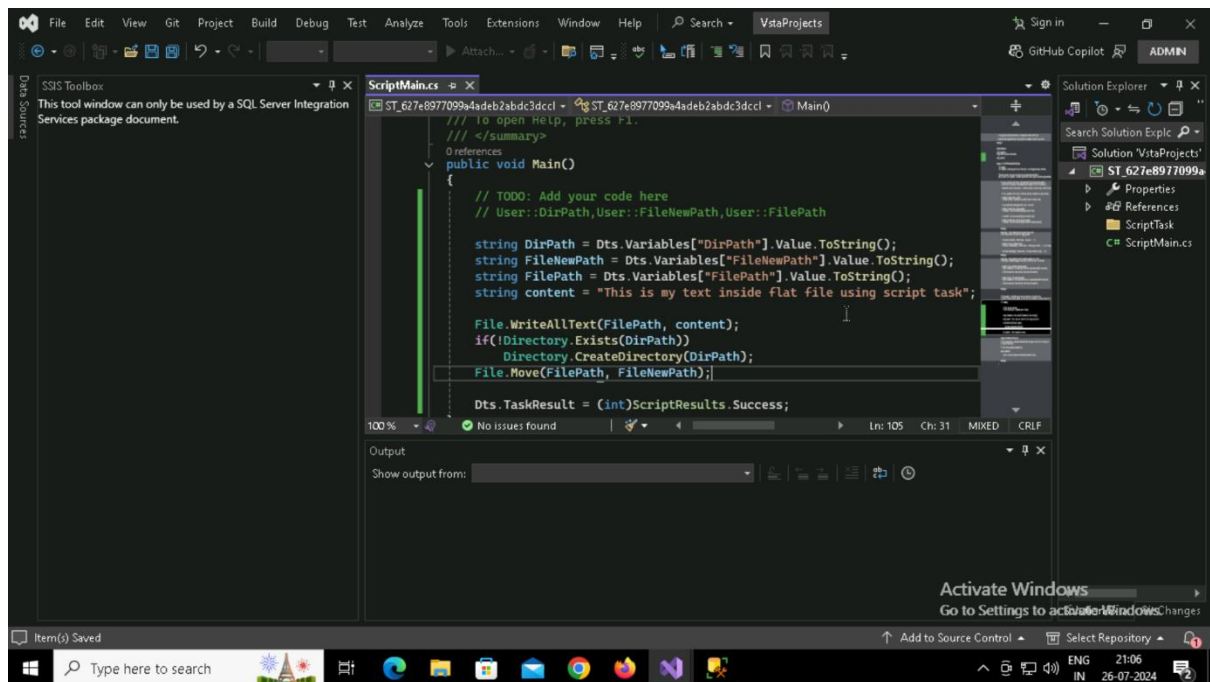
## Task 7: Using Script Task

**Scenario:** You need to perform a complex data transformation that is not supported by the standard SSIS components. A Script Task can be used to achieve this.

Requirements:

1. Add a Script Task to the Control Flow.
2. Write a Script: that performs the required transformation. e.g. Reading data from a file, processing it, and writing the results to a database table.
3. Execute the Script Task within an SSIS package.

### Script code in c# using ssis scripts



### Output of script Task : read a csv file into sql database

