Brittany Hancock
IT420: Lenore Montalbano
Project 2: ETL Package Deliverables
October 5, 2025

I created Data Flow Tasks for each of the file subjects: Customers, Products, and Sales. Inside each, I started by adding a Flat File Source to connect to the downloaded CSV files, which contain the data to be transformed. I adjusted the data types to match the tasks to be performed, such as converting the start dates from strings to dates and converting prices from strings to currency.

Once the sources were configured, I added Derived Column transformations to modify the data according to client requirements. For the Customers' data flow, I standardized state abbreviations to two letters (WA, CA, OR) by using the first letter of each state name. In the Products data flow, I identified empty or null price inputs and replaced them with 0.00. I also standardized categories to correct misspellings. In the Sales data flow, I replaced missing or invalid sale dates with the default value of 1900-01-01, turned negative quantities to zero, and ensured foreign key consistency with the Customer and Product tables.

I added a Sort transformation to the Customers data flow to remove duplicate rows and sort CustomerID in ascending order. The same approach was applied to the Products data flow, sorting ProductID in ascending order and removing duplicates. In the Sales data flow, I sorted by ascending Sale ID, starting at 1. Finally, I added a Multicast transformation to all three data flows to split the output into two destinations: one to export clean CSV files and the other to load cleaned data into the SQL database. Figure 1. Control Flow displaying three Data Flow Tasks (Customers, Products, Sales) connected and executed successfully. Figure 2. Customers Data Flow showing Flat File Source, Derived Column, Sort, and Multicast components connected to both CSV and SQL destinations. Figure 3. Products Data Flow showing transformations for standardizing categories, fixing null prices, sorting ProductID, and writing to dual destinations. Figure 4. Sales Data Flow showing Derived Columns for date and quantity corrections, sorting by SaleID, and output to both CSV and SQL destinations.

Figure 1. Control Flow overview

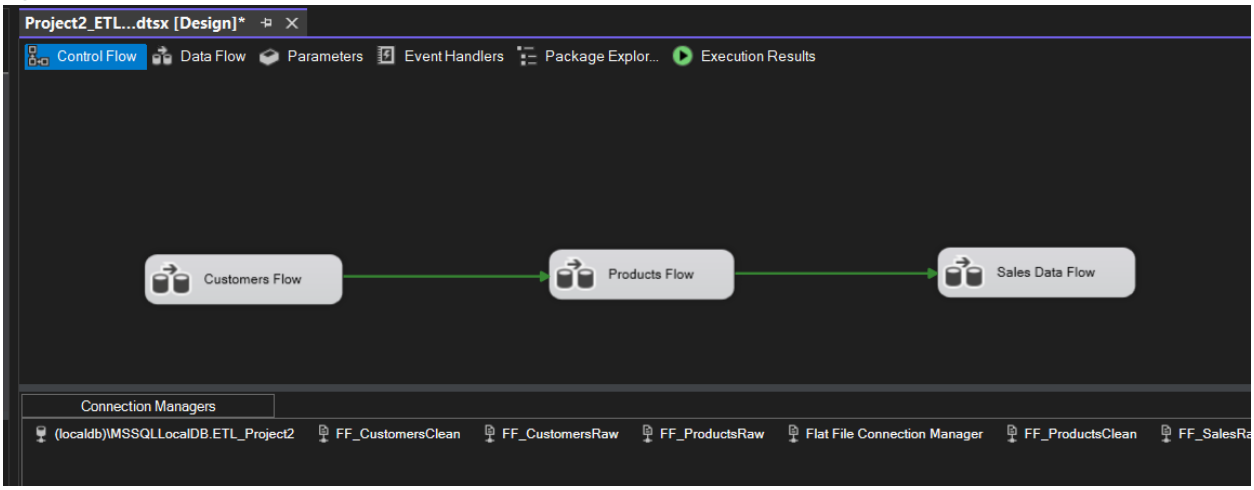


Figure 2. Customer Data Flow

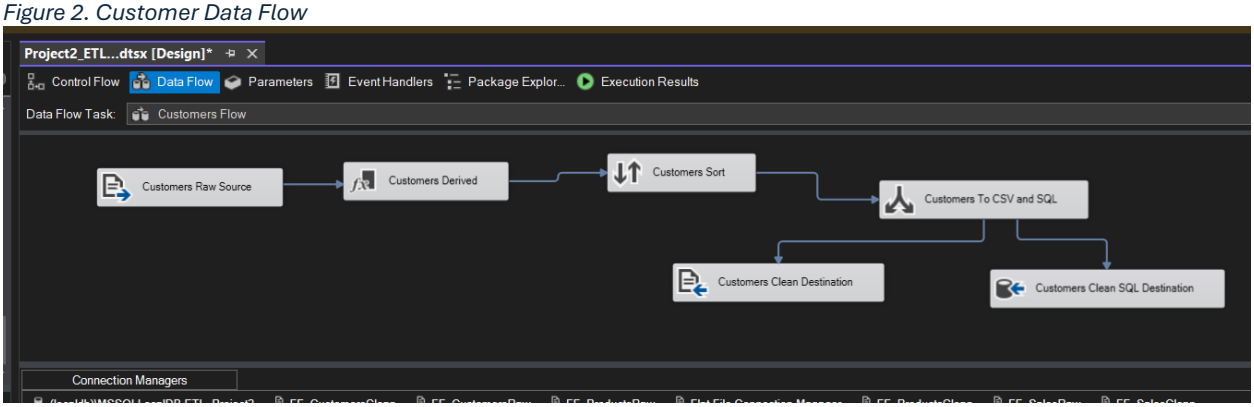


Figure 3. Products Data Flow

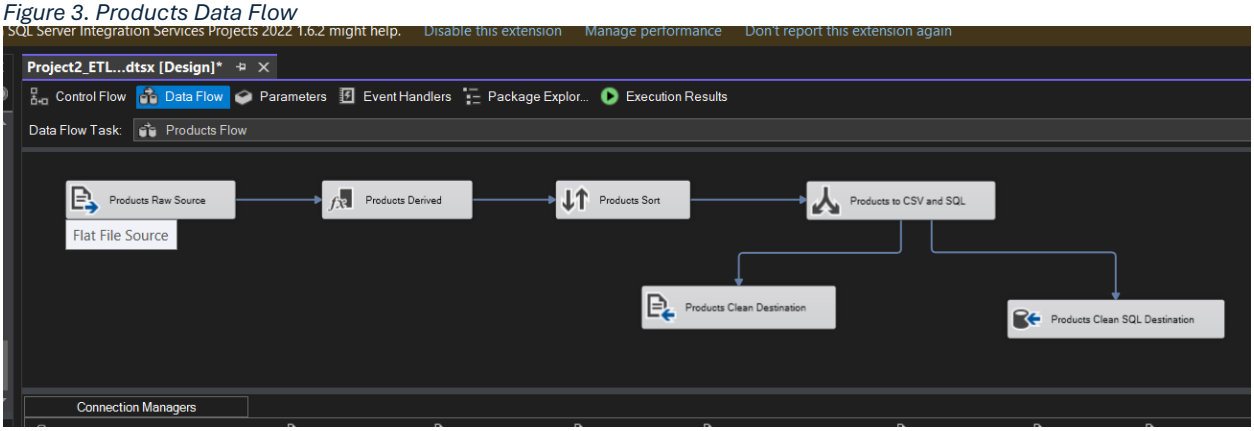


Figure 4. Sales Data Flow

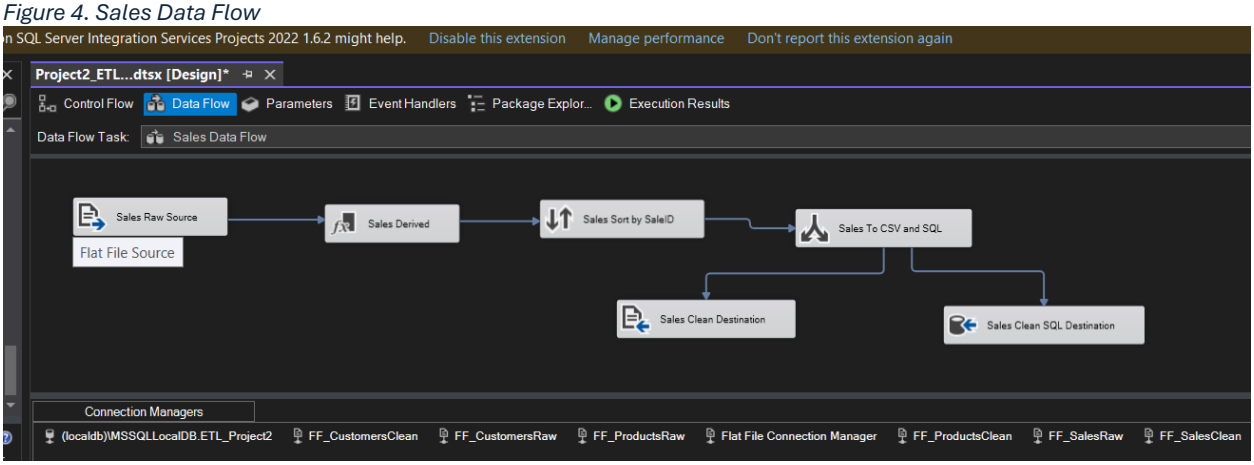
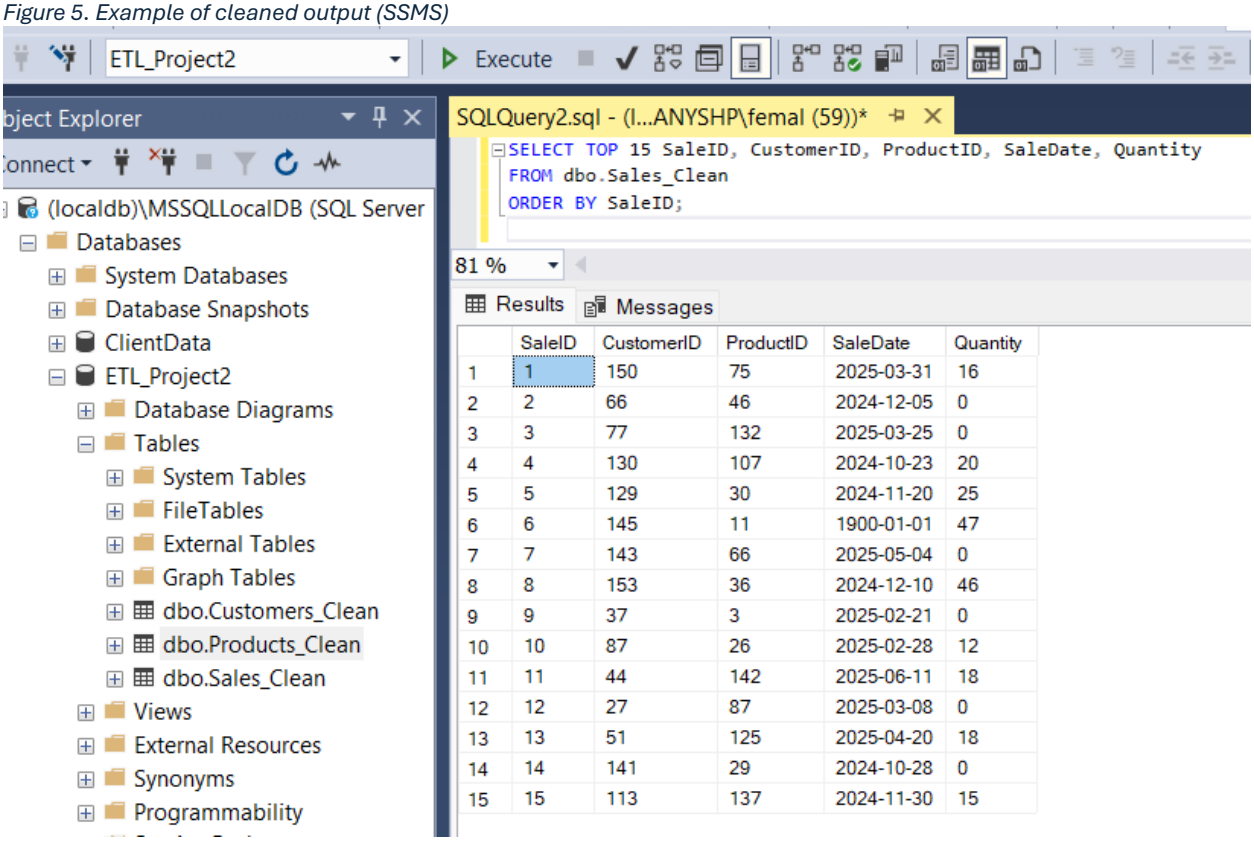


Figure 5. Example of cleaned output (SSMS)



	SaleID	CustomerID	ProductID	SaleDate	Quantity
1	1	150	75	2025-03-31	16
2	2	66	46	2024-12-05	0
3	3	77	132	2025-03-25	0
4	4	130	107	2024-10-23	20
5	5	129	30	2024-11-20	25
6	6	145	11	1900-01-01	47
7	7	143	66	2025-05-04	0
8	8	153	36	2024-12-10	46
9	9	37	3	2025-02-21	0
10	10	87	26	2025-02-28	12
11	11	44	142	2025-06-11	18
12	12	27	87	2025-03-08	0
13	13	51	125	2025-04-20	18
14	14	141	29	2024-10-28	0
15	15	113	137	2024-11-30	15