

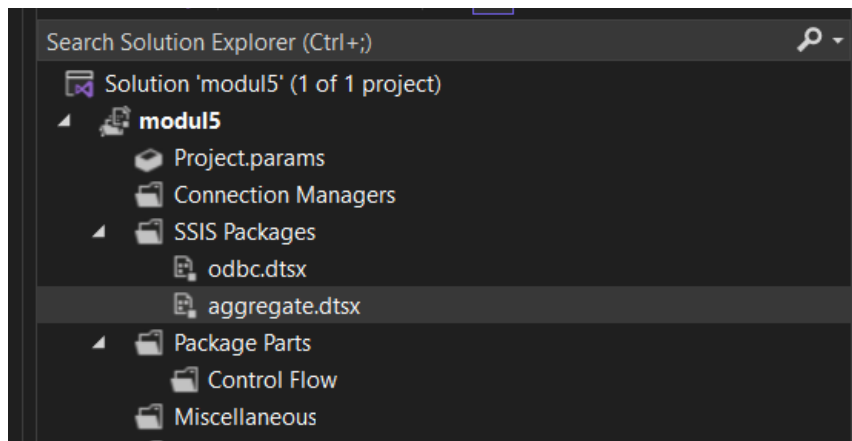
Praktikum Teknologi Perekayasaan Data

Tanggal Praktikum : Monday, March 6, 2023

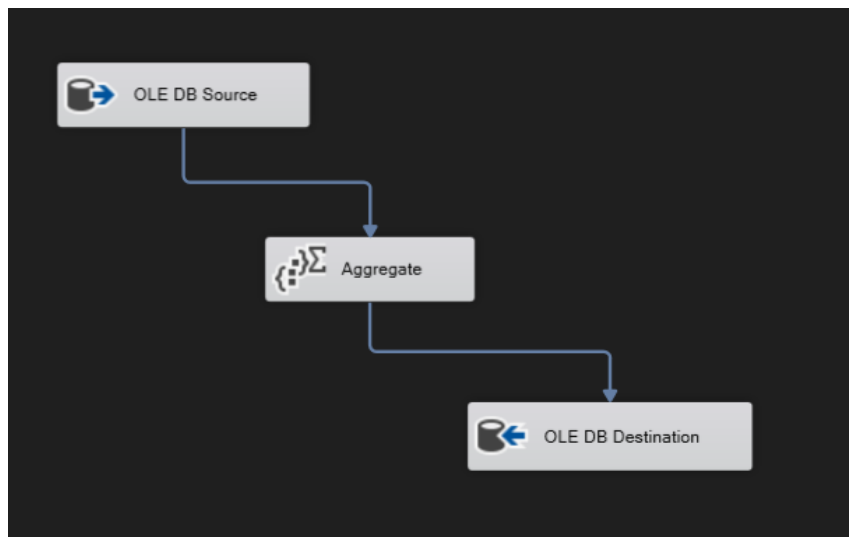
1. Lakukan langkah-langkah pada bagian ETL sesi 3 dan 4 dengan seksama dan screenshoot setiap package yang dibuat, proses run package, dan hasil tabel/file.

A. Transformation Aggregation Basic

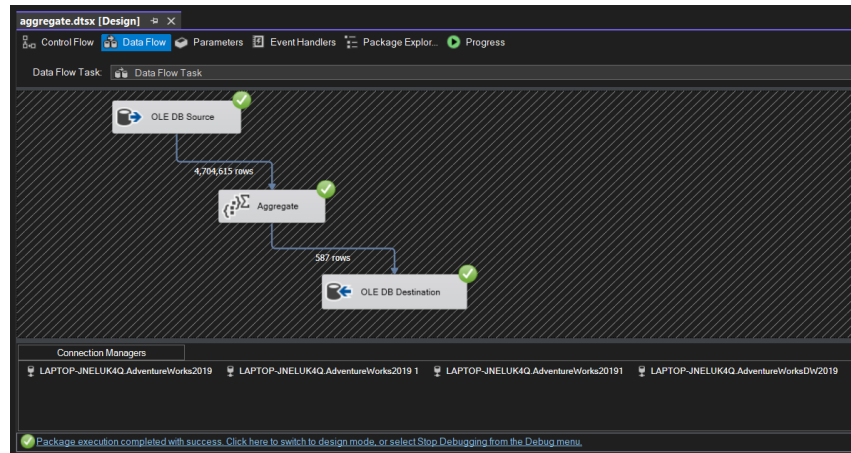
- Package Aggregate



- Data Flow Task



- Proses Run

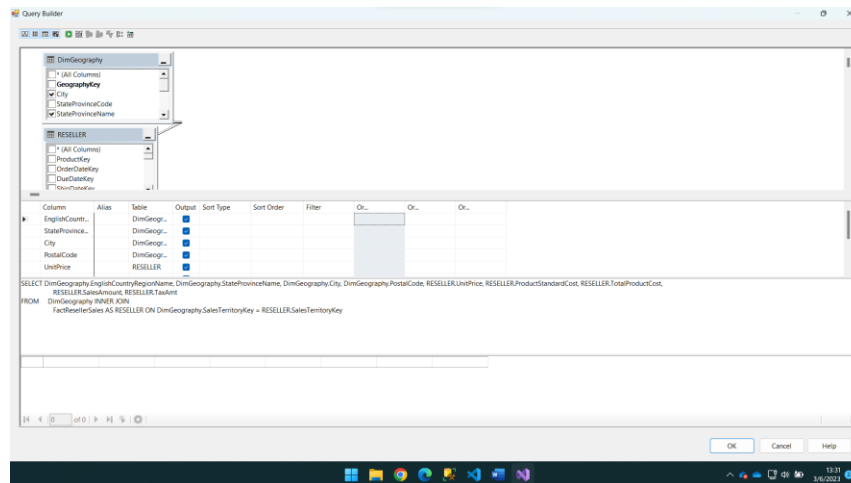


Output

Show output from: Debug

SSIS package "C:\Users\Vendra Kusuma\source\repos\modul5\modul5\aggregate.dtsx" starting.
 Information: 0x4004300A at Data Flow Task, SSIS.Pipeline: Validation phase is beginning.
 Information: 0x4004300A at Data Flow Task, SSIS.Pipeline: Validation phase is beginning.
 Warning: 0x80049304 at Data Flow Task, SSIS.Pipeline: Warning: Could not open global shared memory to communicate with performance DLI; data flow performance counters are not available. To resolve, run this package as an administrator, or on the system's console.
 Warning: 0x80047076 at Data Flow Task, SSIS.Pipeline: The output column "EnglishCountryRegionName" (97) on output "OLE DB Source Output" (96) and component "OLE DB Source" (85) is not subsequently used in the Data Flow task. Removing this unused output column can increase Data Flow task performance.
 Warning: 0x80047076 at Data Flow Task, SSIS.Pipeline: The output column "PostalCode" (100) on output "OLE DB Source Output" (96) and component "OLE DB Source" (85) is not subsequently used in the Data Flow task. Removing this unused output column can increase Data Flow task performance.
 Information: 0x40043006 at Data Flow Task, SSIS.Pipeline: Prepare for Execute phase is beginning.
 Information: 0x40043007 at Data Flow Task, SSIS.Pipeline: Pre-Execute phase is beginning.
 Information: 0x4004300C at Data Flow Task, SSIS.Pipeline: Execute phase is beginning.
 Information: 0x4020900F at Data Flow Task, OLE DB Destination [54]: The final commit for the data insertion in "OLE DB Destination" has started.
 Information: 0x402090E9 at Data Flow Task, OLE DB Destination [54]: The final commit for the data insertion in "OLE DB Destination" has ended.
 Information: 0x40043008 at Data Flow Task, SSIS.Pipeline: Post Execute phase is beginning.
 Information: 0x4004300B at Data Flow Task, SSIS.Pipeline: "OLE DB Destination" wrote 587 rows.
 Information: 0x40043009 at Data Flow Task, SSIS.Pipeline: Cleanup phase is beginning.
 SSIS package "C:\Users\Vendra Kusuma\source\repos\modul5\modul5\aggregate.dtsx" finished: Success.
 The program [13088] DtsDebugHost.exe: DTS has exited with code 0 (0x0).

- Query Builder on Data Source



- Database

SQL Query Editor - LAPTOP-INELUK4Q\AdventureWorks2019 (LAPTOP-INELUK4Q\Hendra Kusuma 870) - Microsoft SQL Server Management Studio

```

SELECT TOP (1000) [City],
               [UnitPrice],
               [StateProvinceName],
               [ProductStandardCost],
               [TotalProductCost],
               [SalesAmount],
               [TaxAmt]
FROM [AdventureWorks2019].[dbo].[AggregationTransformation]

```

City	UnitPrice	StateProvinceName	ProductStandardCost	TotalProductCost	SalesAmount	TaxAmt
Atlanta	602144.6497	Georgia	2841382.8457	0.0000	2700.7604	100.0000
Boston	2071790.1088	Massachusetts	2030445.2271	0.0000	2700.7604	100.0000
Chicago	137211.0618	Illinois	1301001.0624	0.0000	2242.1704	87.2802
Denver	2071790.1088	Colorado	2030445.2271	0.0000	2700.7604	100.0000
London	137211.0618	England	1301001.0624	0.0000	2242.1704	87.2802
Los Angeles	2071790.1088	California	2030445.2271	0.0000	2700.7604	100.0000
Manchester	137211.0618	England	1301001.0624	0.0000	2242.1704	87.2802
Minneapolis	2071790.1088	Minnesota	2030445.2271	0.0000	2700.7604	100.0000
New York	2071790.1088	New York	2030445.2271	0.0000	2700.7604	100.0000
Phoenix	137211.0618	Arizona	1301001.0624	0.0000	2242.1704	87.2802
Portland	2071790.1088	Oregon	2030445.2271	0.0000	2700.7604	100.0000
San Francisco	2071790.1088	California	2030445.2271	0.0000	2700.7604	100.0000
Seattle	2071790.1088	Washington	2030445.2271	0.0000	2700.7604	100.0000
Toronto	137211.0618	Ontario	1301001.0624	0.0000	2242.1704	87.2802
Winnipeg	2071790.1088	Manitoba	2030445.2271	0.0000	2700.7604	100.0000
Yonkers	2071790.1088	New York	2030445.2271	0.0000	2700.7604	100.0000

B. Transformation Aggregation Advance

- Database Source

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager: Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

Columns

Error Output

OLE DB connection manager:
LAPTOP-INELUK4Q\AdventureWorksDW2019

Data access mode:
SQL command

SQL command text:
SELECT [EnglishCountryRegionName]
, [StateProvinceName]
, [City]
, [PostalCode]
, RESELLER.[UnitPrice]
, RESELLER.[ProductStandardCost]
, RESELLER.[TotalProductCost]
, RESELLER.[SalesAmount]
, RESELLER.[TaxAmt]

FROM [DimGeography]
INNER JOIN
[FactResellerSales] AS RESELLER ON
[DimGeography].[SalesTerritoryKey] = RESELLER.

Parameters...

Build Query...

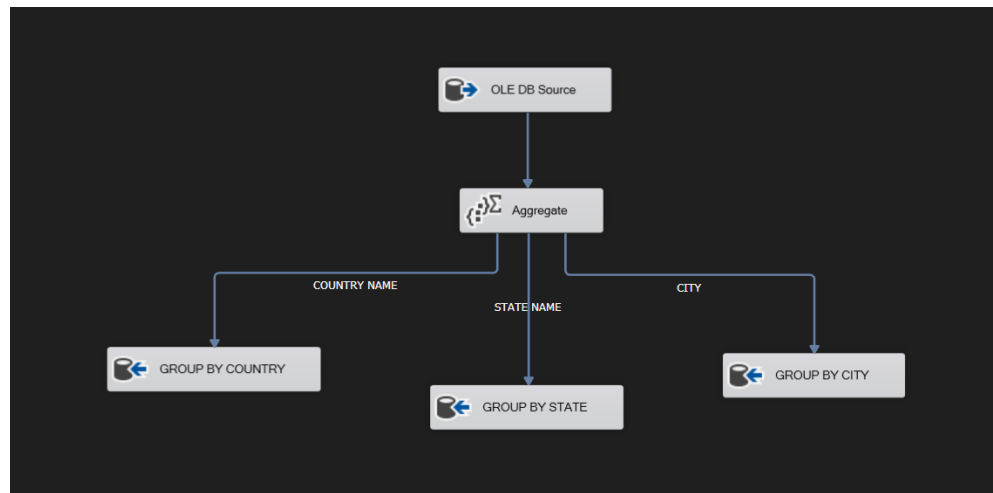
Browse...

Parse Query

Preview...

OK Cancel Help

- Data Flow



- Aggregation

The screenshot shows the 'Aggregate Transformation Editor' window. The 'Aggregations' tab is selected. The window contains a table for defining group by operations and a list of available input columns.

Aggregation Name	Group By Columns	
COUNTRY NAME	EnglishCountryRegionName	k
STATE NAME	StateProvinceName	l

Below the table, there is a 'Basic' tab and a list of 'Available Input Columns'.

Input Column	Output Alias	Operation	Con
EnglishCountryRegionName	EnglishCountryRegionName	Group by	
UnitPrice	UnitPrice	Sum	
ProductStandardCost	ProductStandardCost	Sum	
TotalProductCost	TotalProductCost	Sum	

The 'Available Input Columns' list includes: Name, (*), EnglishCountryRegionName, StateProvinceName, City, PostalCode, and UnitPrice. The 'UnitPrice' column is selected.

- Running

The screenshot displays the SSDT interface for a package named 'aggregate-adv.dtsx'. The Data Flow Task is designed with an 'OLE DB Source' connected to an 'Aggregate' task, which then feeds into three 'GROUP BY' tasks: 'GROUP BY COUNTRY', 'GROUP BY STATE', and 'GROUP BY CITY'. The 'Aggregate' task shows row counts: 4,704,615 rows from the source, and 6 rows for COUNTRY NAME, 71 rows for STATE NAME, and 562 rows for CITY. The bottom status bar indicates 'Package execution completed with success'.

The Output window shows the following log messages:

```

SSIS package "C:\Users\Hendra Kusuma\source\repos\modul5\modul5\aggregate-adv.dtsx" starting.
Information: 0x40004300A at Data Flow Task, SSIS.Pipeline: Validation phase is beginning.
Information: 0x40004300A at Data Flow Task, SSIS.Pipeline: Validation phase is beginning.
Warning: 0x80049304 at Data Flow Task, SSIS.Pipeline: Warning: Could not open global shared memory to communicate with performance DLL; data flow performance counters are not available. To resolve, run this package as an administrator, or on the system's console.
Warning: 0x80047076 at Data Flow Task, SSIS.Pipeline: The output column "PostalCode" (148) on output "OLE DB Source Output" (144) and component "OLE DB Source" (133) is not subsequently used in the Data Flow Task. Removing this unused output column can increase Data Flow task performance.
Information: 0x400043006 at Data Flow Task, SSIS.Pipeline: Prepare for execute phase is beginning.
Information: 0x400043007 at Data Flow Task, SSIS.Pipeline: Pre-Execute phase is beginning.
Information: 0x40004300C at Data Flow Task, SSIS.Pipeline: Execute phase is beginning.
Information: 0x402090DF at Data Flow Task, GROUP BY COUNTRY [85]: The final commit for the data insertion in "GROUP BY COUNTRY" has started.
Information: 0x402090DF at Data Flow Task, GROUP BY STATE [114]: The final commit for the data insertion in "GROUP BY STATE" has started.
Information: 0x402090DF at Data Flow Task, GROUP BY CITY [66]: The final commit for the data insertion in "GROUP BY CITY" has started.
Information: 0x402090E0 at Data Flow Task, GROUP BY COUNTRY [85]: The final commit for the data insertion in "GROUP BY COUNTRY" has ended.
Information: 0x402090E0 at Data Flow Task, GROUP BY STATE [114]: The final commit for the data insertion in "GROUP BY STATE" has ended.
Information: 0x402090E0 at Data Flow Task, GROUP BY CITY [66]: The final commit for the data insertion in "GROUP BY CITY" has ended.
Information: 0x400043008 at Data Flow Task, SSIS.Pipeline: Post Execute phase is beginning.
Information: 0x400043008 at Data Flow Task, SSIS.Pipeline: "GROUP BY CITY" wrote 562 rows.
Information: 0x400043008 at Data Flow Task, SSIS.Pipeline: "GROUP BY COUNTRY" wrote 6 rows.
Information: 0x400043008 at Data Flow Task, SSIS.Pipeline: "GROUP BY STATE" wrote 71 rows.
Information: 0x400043009 at Data Flow Task, SSIS.Pipeline: Cleanup phase is beginning.
SSIS package "C:\Users\Hendra Kusuma\source\repos\modul5\modul5\aggregate-adv.dtsx" finished: Success.
The program '12968 DtsDebugHost.exe: DTS' has exited with code 0 (0x0).

```

- Database

The screenshot shows a SQL query window with the following query:

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [EnglishCountryRegionName]
, [UnitPrice]
, [ProductStandardCost]
, [TotalProductCost]
, [SalesAmount]
, [TaxAmt]
FROM [AdventureWorksDW2019].[dbo].[GROUPB_COUNTRY]

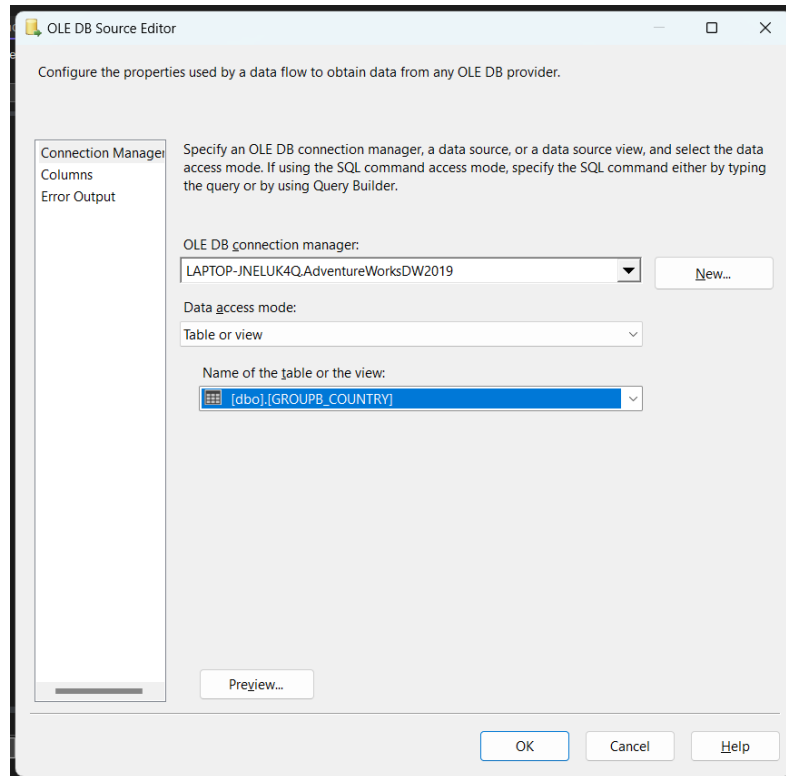
```

The Results tab displays the following data:

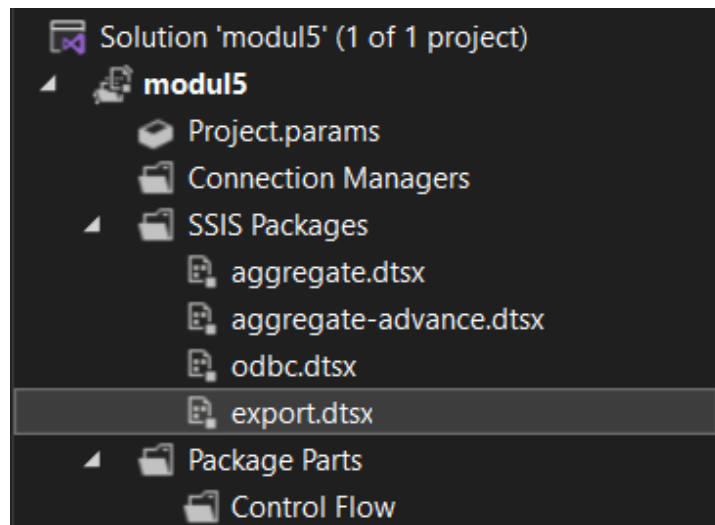
	EnglishCountryRegionName	UnitPrice	ProductStandardCost	TotalProductCost	SalesAmount	TaxAmt
1	Australia	30709185.608	32268055.172	68122250.352	63773415.068	5101873.328
2	Canada	355437442.4311	348796647.4618	1036577990.8723	1049588568.5445	83967083.241
3	France	63163624.6416	62414778.696	222952681.8624	221161820.88	17692945.4352
4	Germany	44378402.2305	45910017.66	136190712.8725	128959222.4245	10316737.613
5	United States	1543196376.8455	1519351920.9312	4602504006.3798	4637647572.8797	371011797.0967
6	United Kingdom	72992233.9754	72159534.5772	226449226.7229	226787467.8098	18142996.9516

C. Export Data

- Data Source

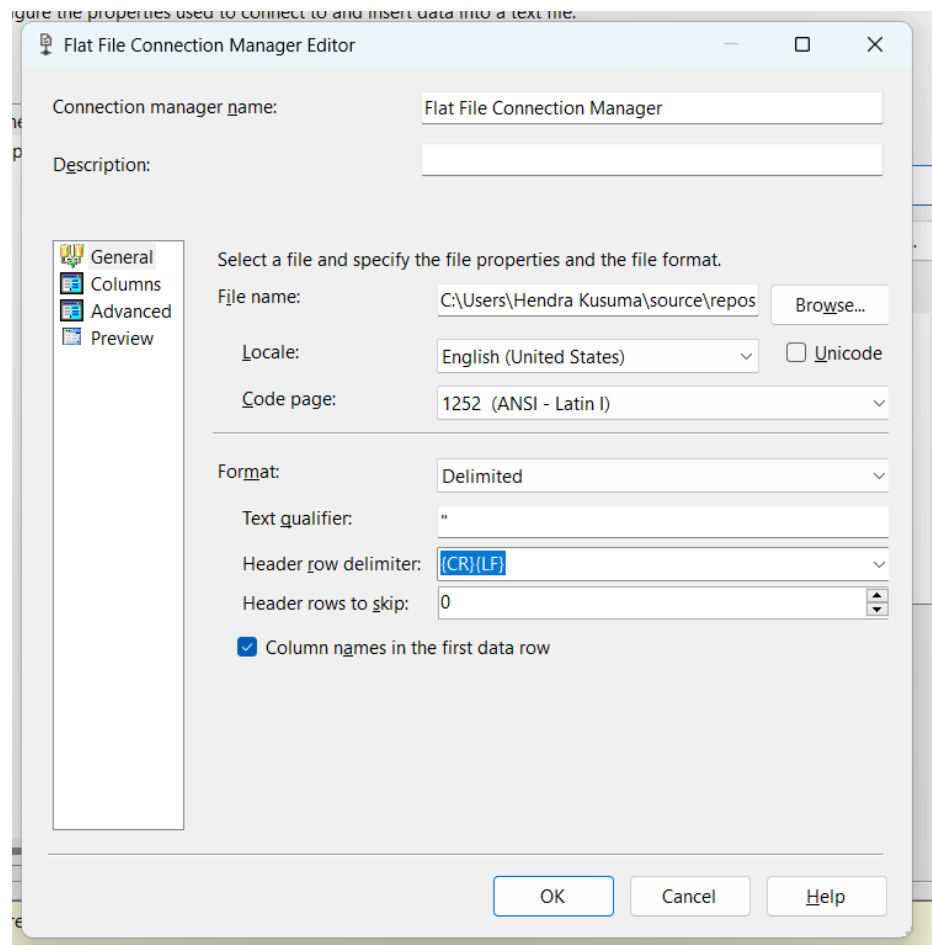


- Package

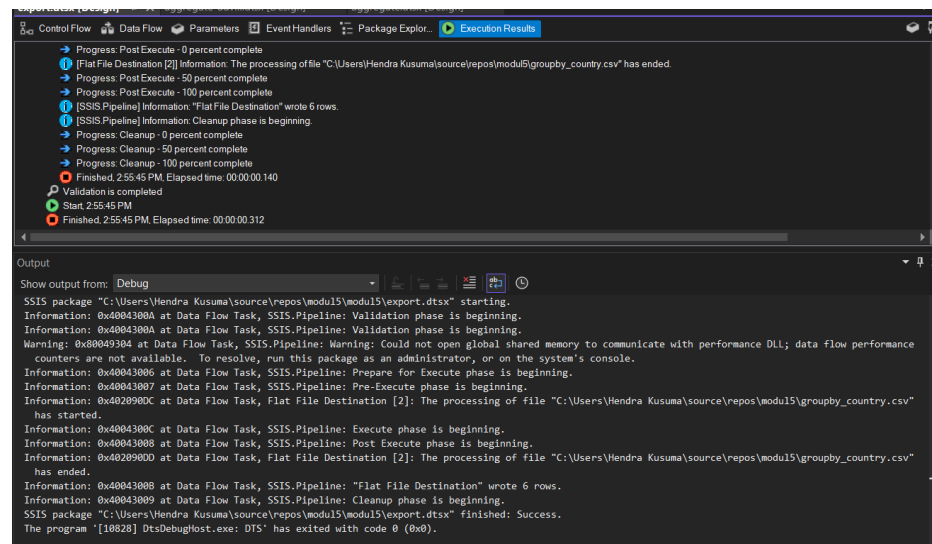


- Konfigurasi Flat File

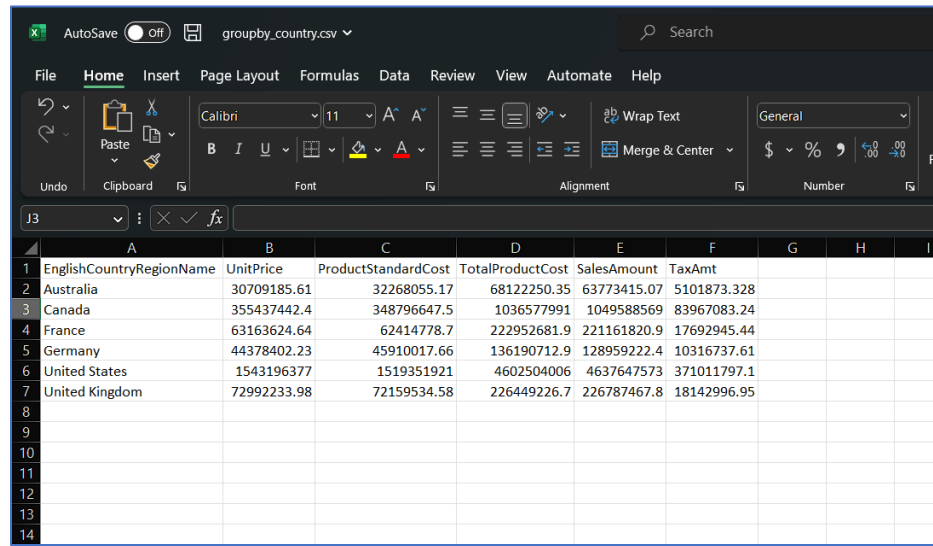
Figure the properties used to connect to and insert data into a text file.



- Running Result



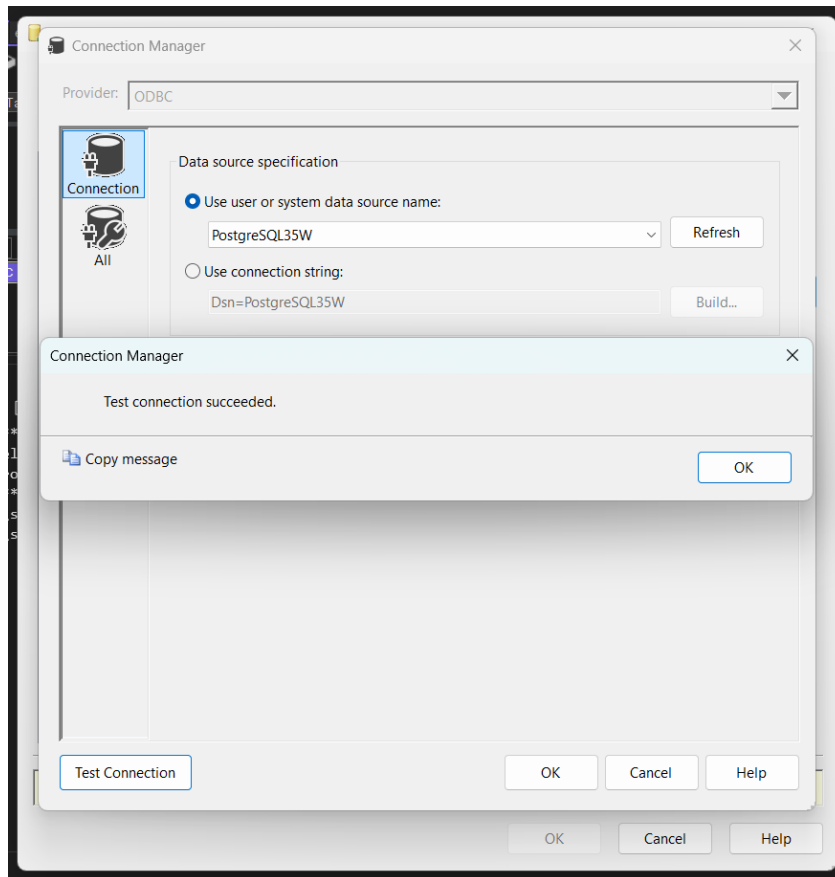
- CSV Output



The screenshot shows the Microsoft Excel interface with a file named 'groupby_country.csv' open. The data is organized into columns: EnglishCountryRegionName, UnitPrice, ProductStandardCost, TotalProductCost, SalesAmount, and TaxAmt. The rows represent different countries: Australia, Canada, France, Germany, United States, and United Kingdom. The data is grouped by country, with each country having multiple rows of product data.

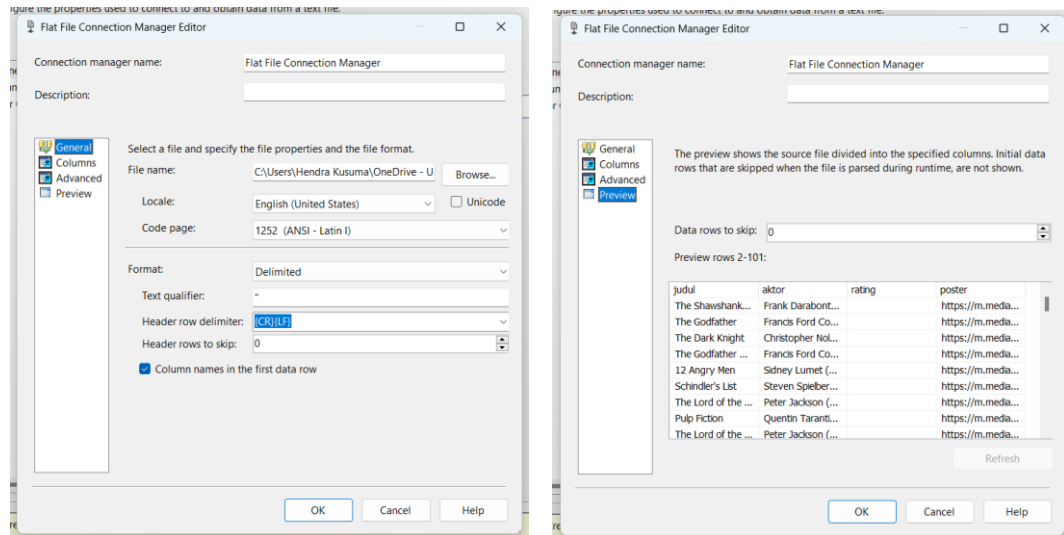
EnglishCountryRegionName	UnitPrice	ProductStandardCost	TotalProductCost	SalesAmount	TaxAmt
Australia	30709185.61	32268055.17	68122250.35	63773415.07	5101873.328
Canada	355437442.4	348796647.5	1036577991	1049588569	83967083.24
France	63163624.64	62414778.7	222952681.9	221161820.9	17692945.44
Germany	44378402.23	45910017.66	136190712.9	128959222.4	10316737.61
United States	1543196377	1519351921	4602504006	4637647573	371011797.1
United Kingdom	72992233.98	72159534.58	226449226.7	226787467.8	18142996.95

2. Buatlah koneksi odbc ke database selain SQL server, misal **PosgreSQL/MySQL**.

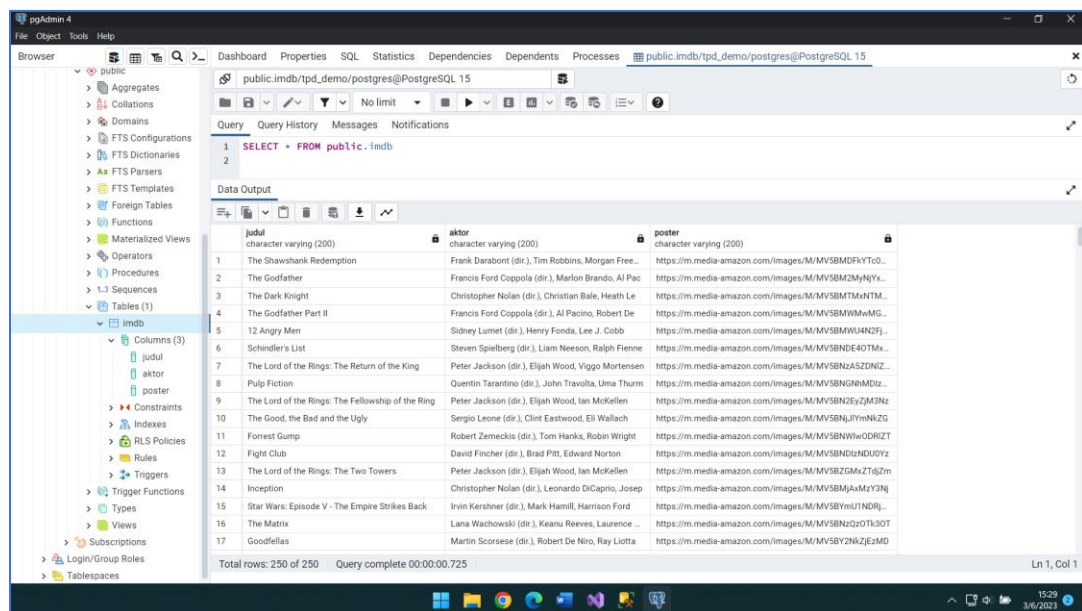


3. Import file imdb_data.csv ke database yang dipilih pada poin 2 dengan menggunakan SSIS.

- Konfigurasi Import CSV



- Database output in PostgreSQL



4. Buatlah package dengan memanfaatkan database AdventureWorks2019/ AdventureWorksDW2019 dengan minimal dua fitur transformasi yang terdapat pada SSIS. Cantumkan dokumentasi pada file laporan praktikum. Cara penggunaan fitur transformasi dapat diperoleh dari berbagai referensi yang ada.

- Data Flow

1. OLE DB Source

Pada tahapan ini, terdapat task untuk mengambil data dari AdventureWorksDW2019 dengan operasi join

2. Aggregate

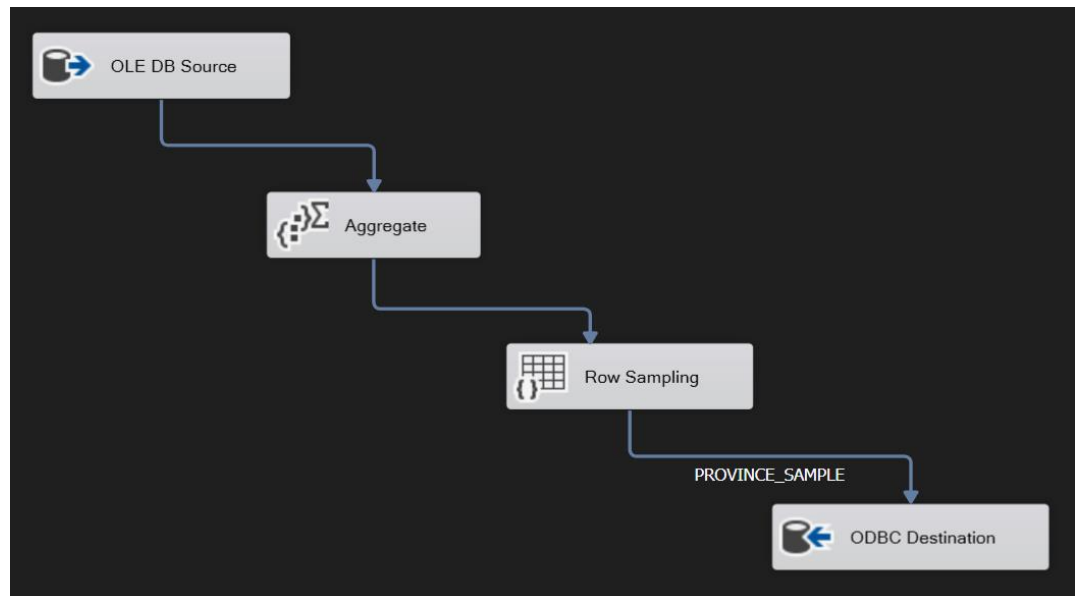
Pada tahapan ini dilakukan agregasi dari data yang tersedia berdasarkan state

3. Row Sampling

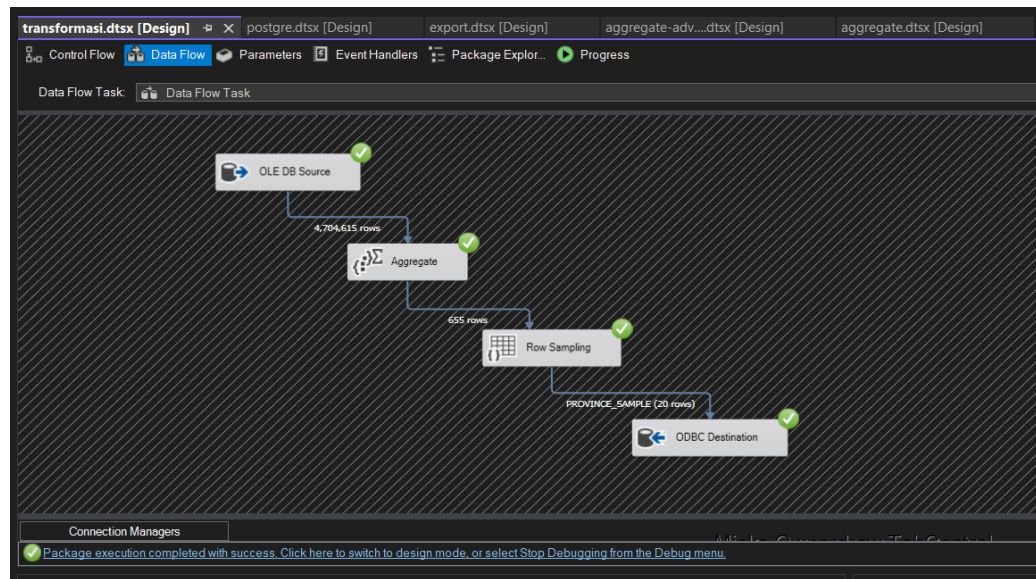
Pada row sampling dilakukan pengambilan sample berjumlah 20

4. ODBC Destination

Terdapat operasi write ke database PostgreSQL



- Running Task



- Database

The screenshot shows the pgAdmin 4 interface with a query result for the `public.sample_state` table. The query is:

```
SELECT * FROM public.sample_state
```

The Data Output shows 20 rows of data. The columns are:

- state**: character varying
- unitPrice**: double precision
- productStandardCost**: double precision
- totalProductCost**: double precision
- salesAmount**: double precision
- taxAmount**: double precision

state	unitPrice	productStandardCost	totalProductCost	salesAmount	taxAmount
1 Ohio	2475750.3051	2443589.4188	6855591.6208	6932842.0148	554627.3442
2 Seine et Marne	1315908.8467	1300307.8895	4644847.5388	4607537.935	368603.0299
3 Seine (Paris)	1315908.8467	1300307.8895	4644847.5388	4607537.935	368603.0299
4 Washington	3929388.8607	3834616.6399	12214179.2314	12435076.0001	994806.059
5 Charente-Maritime	1315908.8467	1300307.8895	4644847.5388	4607537.935	368603.0299
6 Utah	3929388.8607	3834616.6399	12214179.2314	12435076.0001	994806.059
7 British Columbia	4869006.0607	4778036.2666	14199698.5051	14377925.5965	1150234.017
8 New South Wales	767729.6402	806701.3793	1703056.2588	1594335.3767	127546.8332
9 England	1377211.9618	1361500.6524	4272626.9193	4279008.8266	342320.6972
10 Quebec	4869006.0607	4778036.2666	14199698.5051	14377925.5965	1150234.017
11 Wisconsin	2679790.1088	2635645.2271	7785449.5799	7906008.1777	632480.6347
12 Massachusetts	2475750.3051	2443589.4188	6855591.6208	6932842.0148	554627.3442
13 Seine (Paris)	1315908.8467	1300307.8895	4644847.5388	4607537.935	368603.0299
14 England	1377211.9618	1361500.6524	4272626.9193	4279008.8266	342320.6972
15 California	6054471.7463	5985742.2894	18482078.6921	18466458.7925	1477316.673
16 England	1377211.9618	1361500.6524	4272626.9193	4279008.8266	342320.6972
17 New South Wales	767729.6402	806701.3793	1703056.2588	1594335.3767	127546.8332

Total rows: 20 of 20 Query complete 00:00:00.366 Ln 1, Col 1