

Laporan Praktikum 3

Nama : Ernike Nelsi Manurung

NIM : 12S17007

Daftar Isi

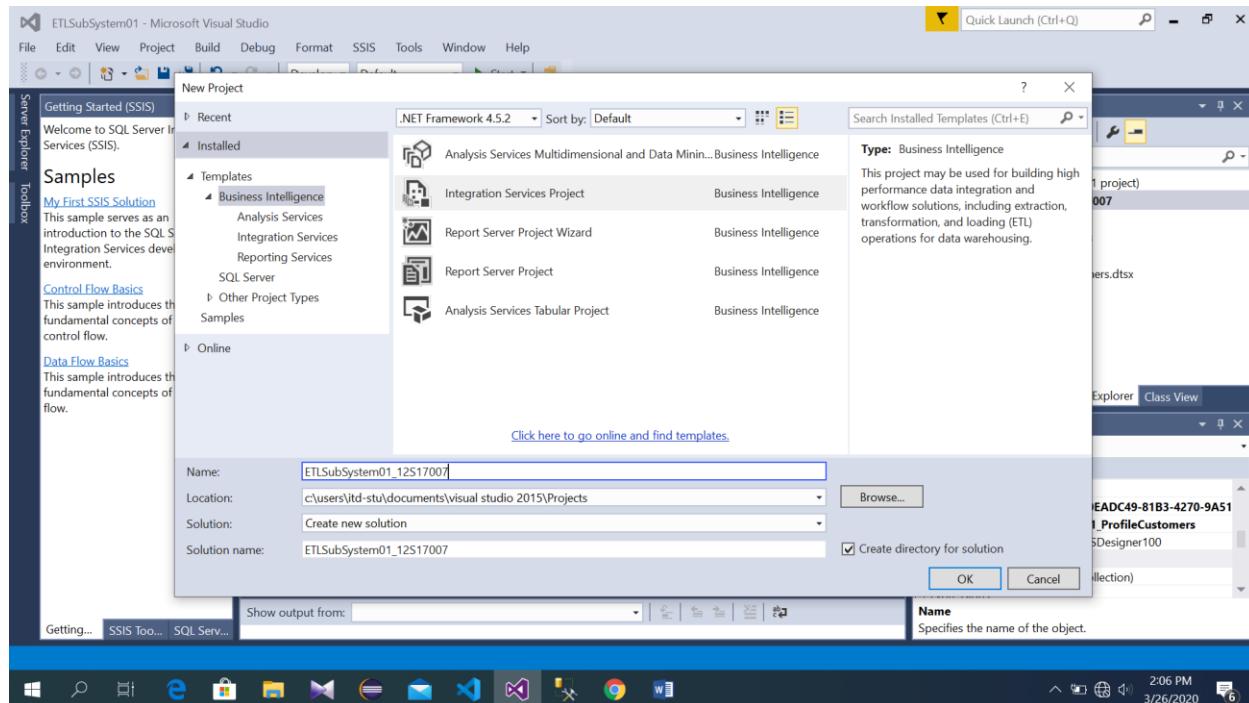
A.	Subsystem 1: Data Profiling.....	2
B.	Subsystem 2: Change Data Capture.....	8
C.	Data Cleansing System	15
D.	Deduplication System	24
E.	Slowly Changing Dimension Type 2	35
F.	Fact Table Builder	46

Kebutuhan:

- SQL Server 2014
- SQL Server Data Tools 2015
- AdventureWorks2014
- AdventureWorks2012
- Northwind

A. Subsystem 1: Data Profiling

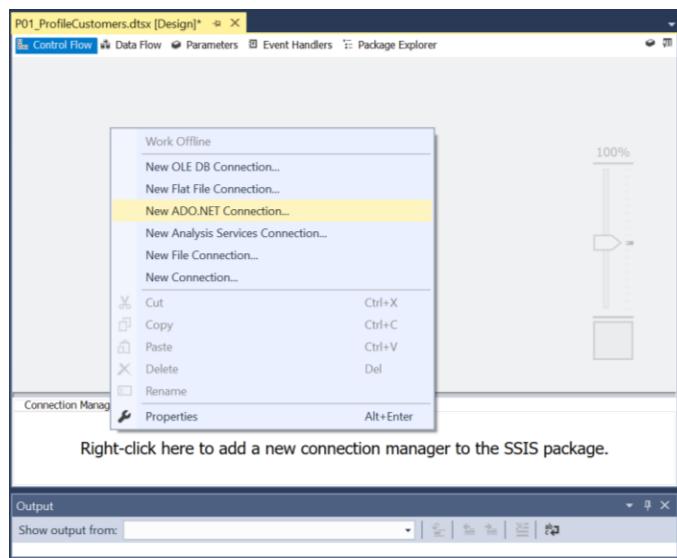
- 1) Buka aplikasi SQL Server Data Tools 2015 dan buatlah sebuah project integration baru dengan nama ETLSubSystem01_12S17007.



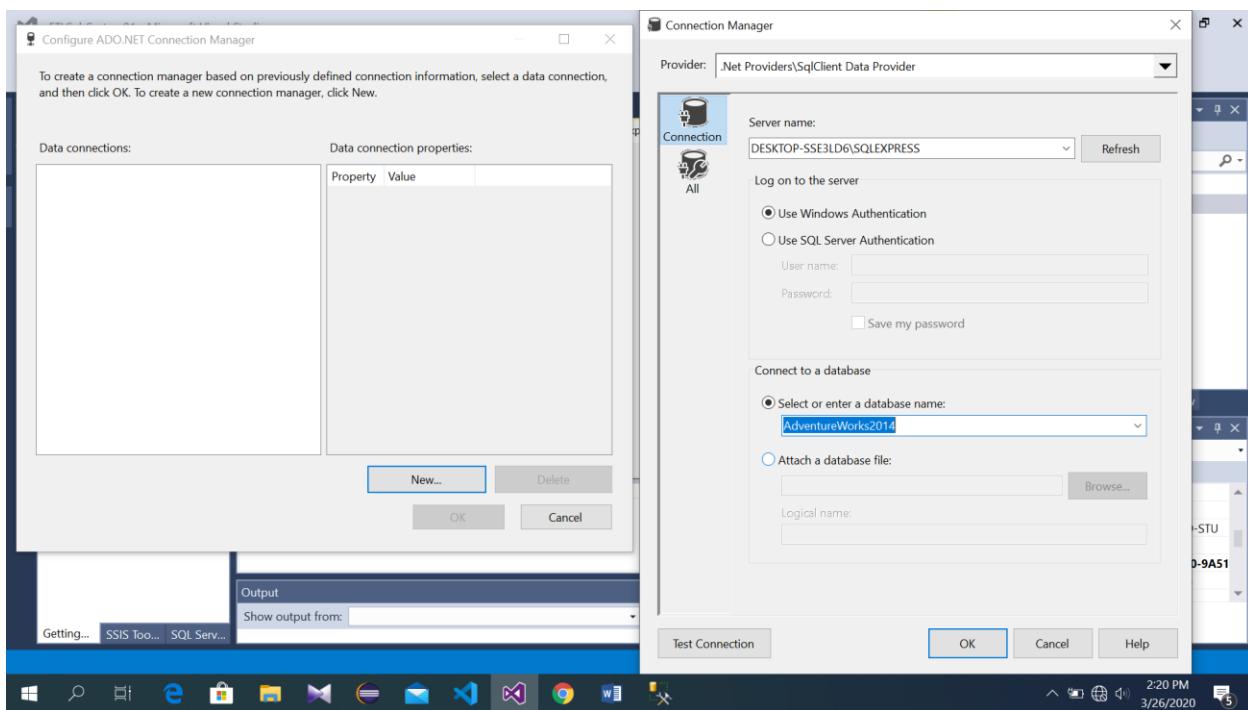
- 2) Rename package yang secara default dibuat waktu membuat project menjadi P01_ProfileCustomers.dtsx.



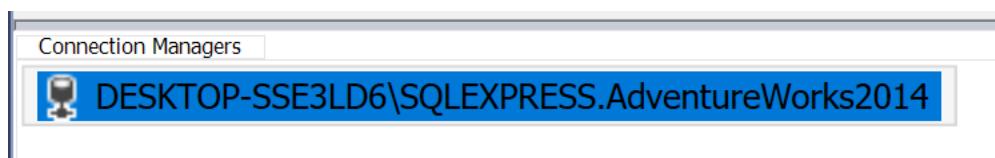
- 3) Dari connection manager panel, buatlah sebuah ADO.NET Connection yang merujuk basis data AdventureWorks2014. Klik kanan, kemudian pilih ADO.NET.



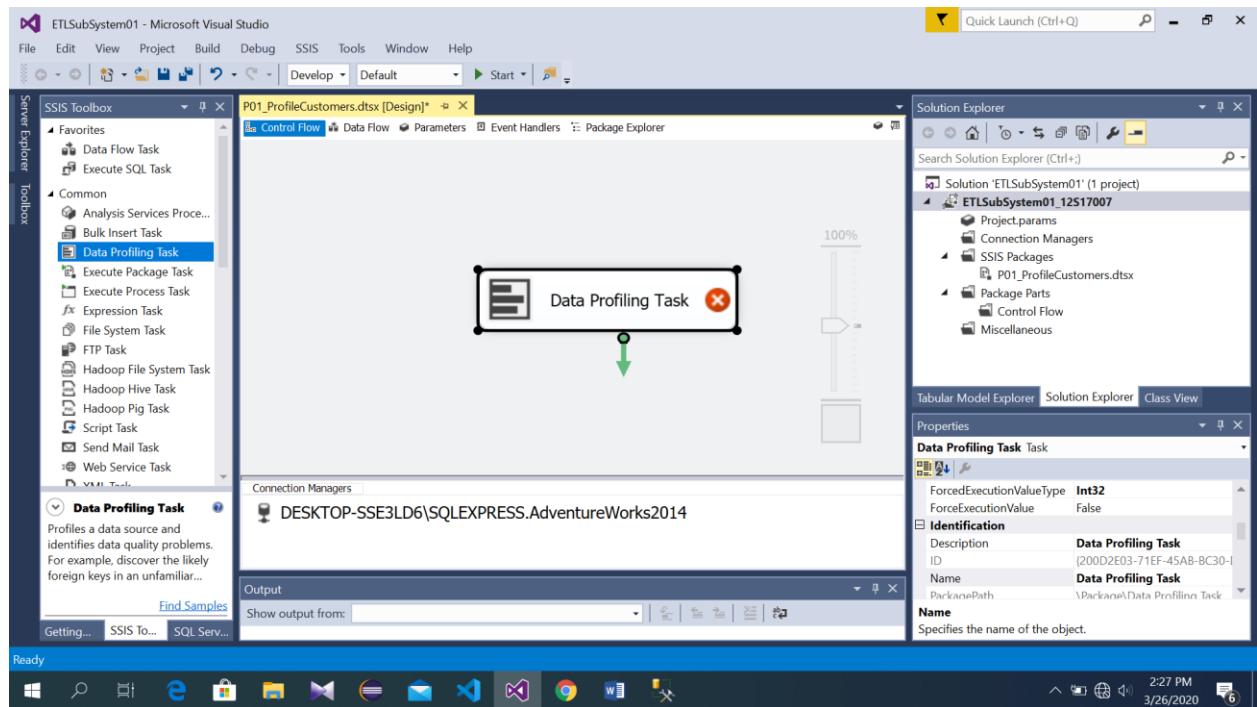
- 4) Setelah itu, pilih New, lalu masukkan Server name laptop dan pilih database AdventureWorks2014. Kemudian pilih OK.



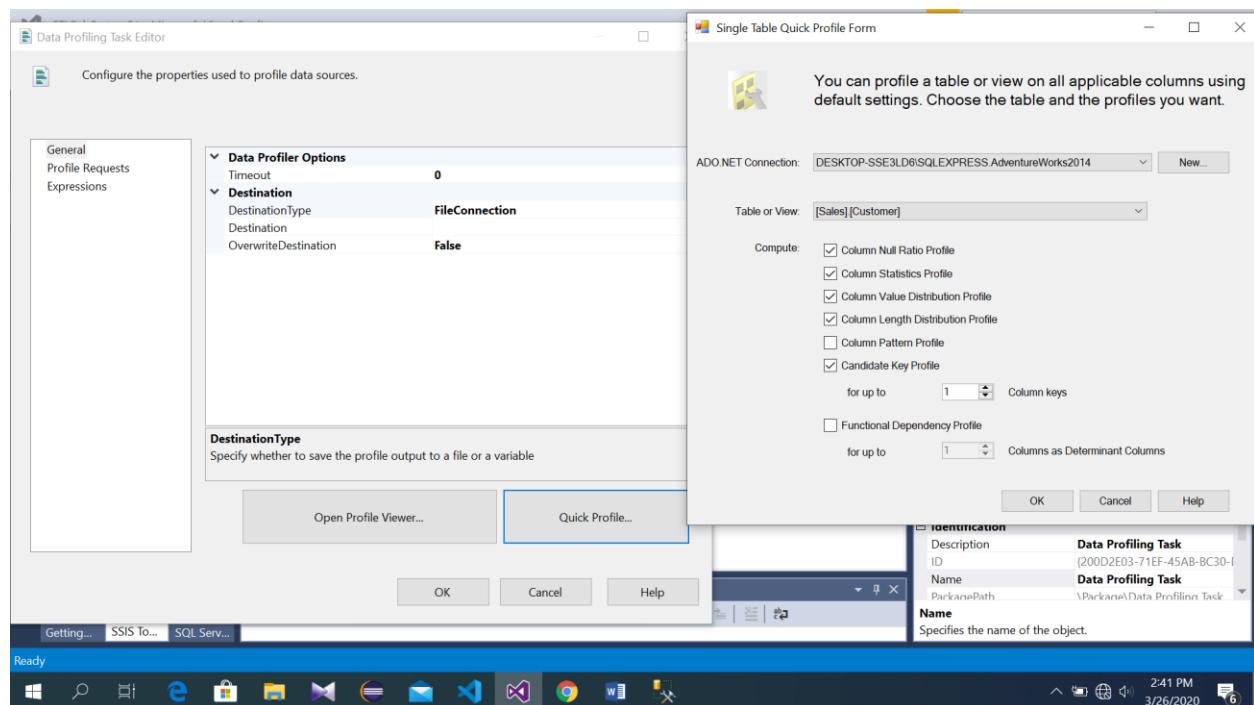
- 5) Jika telah berhasil, akan muncul seperti gambar di bawah ini.



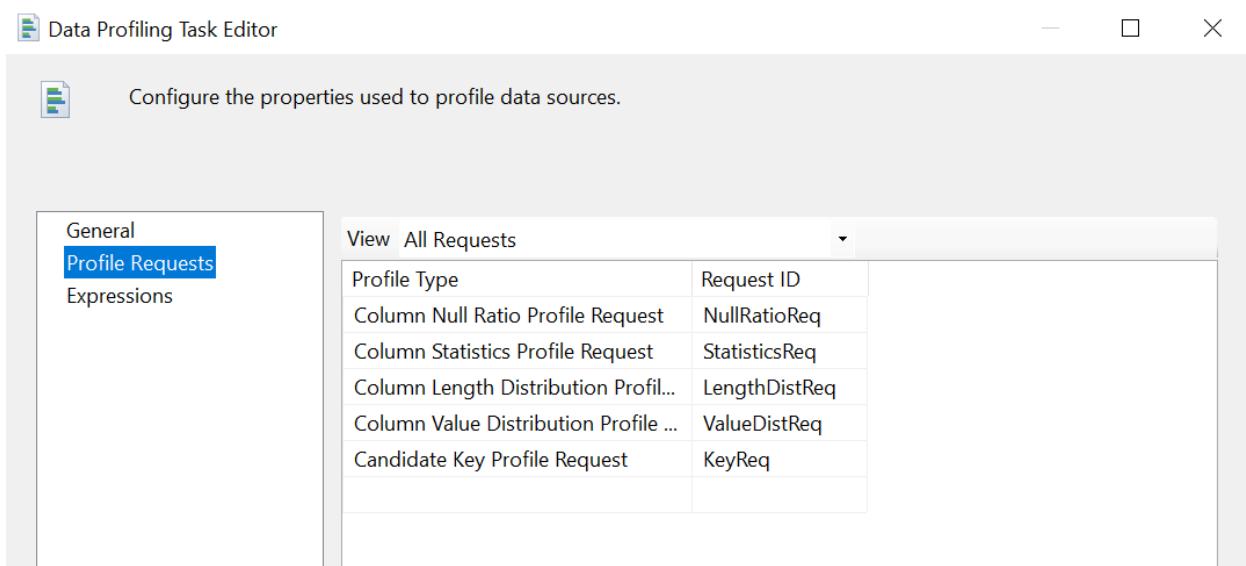
- 6) Kemudian pada Control Flow tambahkan Data Profiling Task. Dari SSIS Toolbox, drag and drop Data Profiling Task. Hal ini bertujuan untuk membuat profil data *customers* yang terdapat dalam basis data AdventureWorks2014.



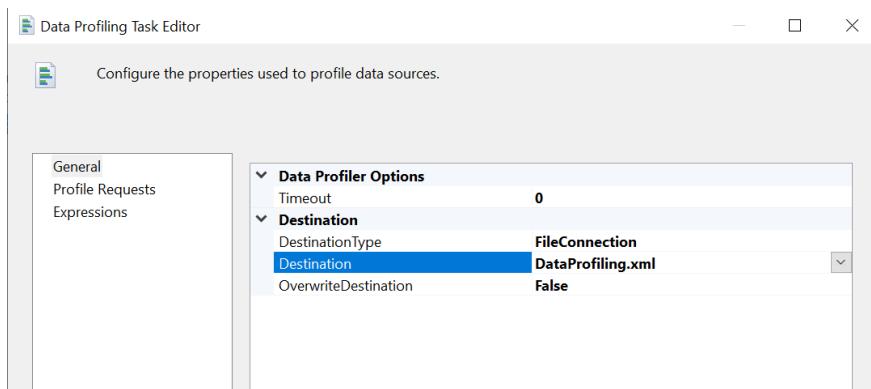
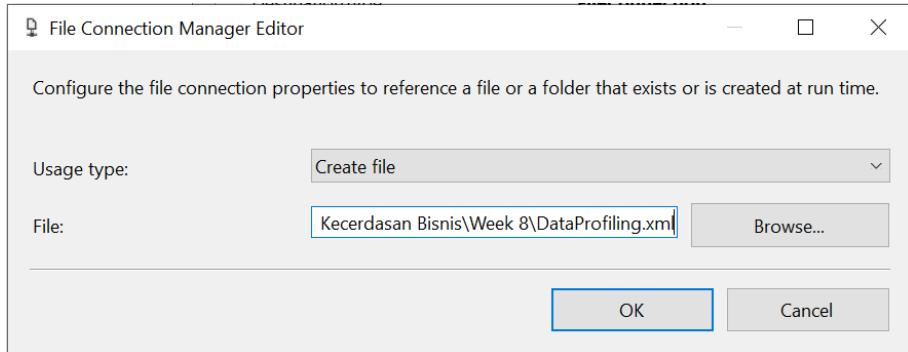
- 7) Double click Pada Data Profiling tersebut.
 8) Kemudian klik Quick Profile untuk menggunakan konfigurasi default.
 9) Dalam property ADO.NET Connection pilih koneksi device yang telah dibuat, kemudian pada Table or View, pilih [Sales].[Customer].



10) Klik OK dan lihat rincian permintaan pada Profile Request.



- 11) Setelah permintaan profil dibuat, kita perlu mengatur letak luaran yang akan dihasilkan. Pilih General, pada Destination Type pilih File Connection dan pada Destination pilih New File Connection.
- 12) Pada kotak dialog File Connection Manager Editor, pilih Create File pada property Usage Type dan pada property File, berikan nama berkas XML yang akan menyimpan luaran.



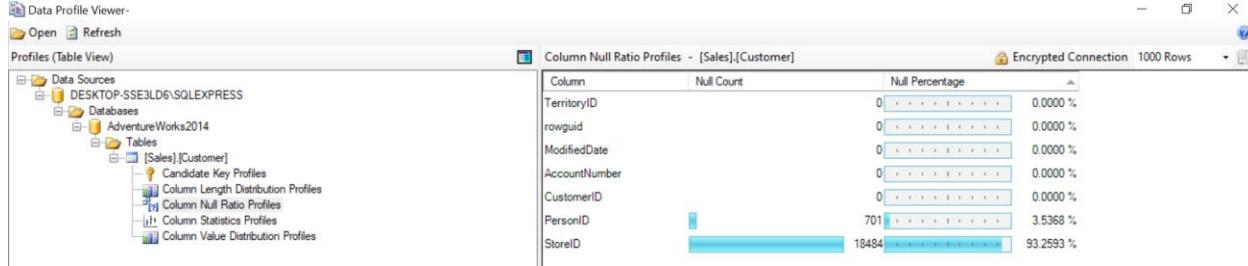
- 13) Jalankan package tersebut, tekan F5. Jika berhasil, laporannya akan dihasilkan dalam berkas XML pada Dataprofiling.xml dan pada Data Profiling Task akan terdapat centang hijau.



```
File Edit Format View Help
<?xml version="1.0"?>
<DataProfile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://</>
<ProfileVersion>1.0</ProfileVersion>
<DataSources>
  <DtsDataSource ID="{40C880AE-1480-4750-8F26-253C561FE946}" Name="DataProfiling.xml">
    <DtsConnectionManagerID>DataProfiling.xml</DtsConnectionManagerID>
  </DtsDataSource>
  <DtsDataSource ID="{D45EAB8D-4CDE-451F-A215-B8B67D67998E}" Name="DataProfiling.xml 1">
    <DtsConnectionManagerID>DataProfiling.xml 1</DtsConnectionManagerID>
  </DtsDataSource>
  <DtsDataSource ID="{31A1B853-479B-40C7-A381-13B0F1AEFF8C}" Name="DESKTOP-SSE3LD6\SQLEXPRESS.AdventureWorks2014">
    <DtsConnectionManagerID>DESKTOP-SSE3LD6\SQLEXPRESS.AdventureWorks2014</DtsConnectionManagerID>
  </DtsDataSource>
</DataSources>
```

- 14) Selain itu, untuk menampilkan laporan tersebut, double klik Data Profiling Task, kemudian pilih Open Profile Viewer.

- 15) Misalnya, kita tinjau poin Column Null Ratio Profiles.



Berdasarkan data tersebut, kita ketahui bahwa sebanyak 3,5368% dari Customer records tidak memiliki PersonID dan 93.2593% tidak memiliki StoreID.

- 16) Selain itu, pada data profile viewer kita dapat menggunakan fitur Drill Down. Misalnya kita klik baris PersonID kemudian klik ikon drill down. Hasilnya akan ditampilkan seperti pada gambar di bawah.

Data Profile Viewer -

Open Refresh

Profiles (Table View)

Column Null Ratio Profiles - [Sales].[Customer]

Column Null Ratio Profiles

Column Null Ratio Profiles

Column	Null Count	Null Percentage
AccountNumber	0	0.0000 %
CustomerID	0	0.0000 %
ModifiedDate	0	0.0000 %
PersonID	701	3.5368 %
rowguid	0	0.0000 %
StoreID	18484	93.2593 %
TerritoryID	0	0.0000 %

AccountNumber	CustomerID	ModifiedDate	PersonID	rowguid	StoreID	TerritoryID
AW00000001	1	9/12/2014 11:15...		3f5ae95e-b87d-4...	934	1
AW00000002	2	9/12/2014 11:15...		e552657-a9af-4...	1028	1
AW00000003	3	9/12/2014 11:15...		130774b1-db21...	642	4
AW00000004	4	9/12/2014 11:15...		#862851-1daa-4...	932	4
AW00000005	5	9/12/2014 11:15...		83905bdc-9f5e-4...	1026	4
AW00000006	6	9/12/2014 11:15...		1a52d88bfca24...	644	4
AW00000007	7	9/12/2014 11:15...		03e9273e-b193...	930	1
AW00000008	8	9/12/2014 11:15...		801368b1-4323...	1024	5
AW00000009	9	9/12/2014 11:15...		b900b7-23c34...	620	5
AW00000010	10	9/12/2014 11:15...		cdb6698d-2ff1-4...	928	6
AW00000011	11	9/12/2014 11:15...		750f3495-59c4-4...	1022	6
AW00000012	12	9/12/2014 11:15...		94bcacf1-ff32-4...	622	6
AW00000013	13	9/12/2014 11:15...		b0fa5854-2511-4...	434	7
AW00000014	14	9/12/2014 11:15...		2f9bbedc-723d-4...	1020	8

Message Column Null Ratio Profiles

3:26 PM 3/26/2020

B. Subsystem 2: Change Data Capture

- 1) Buatlah proyek baru pada SQL Server Data Tools 2015 dengan nama ETLSubsystem02_12S17007.
- 2) Rename package pada project tersebut dengan P01_IncrementalLoad.dtsx.
- 3) Aktifkan CDC pada basis data AdventureWorks2014 dengan mengeksekusi SQL Statement berikut.

```
EXEC sys.sp_cdc_enable_db
```

Pada saat pengeksekusian, kemungkinan akan terjadi error.

```
Msg 22830, Level 16, State 1, Procedure sp_cdc_enable_db_internal, Line 195  
Could not update the metadata that indicates database AdventureWorks2014 is enabled for Change Data Capture. The failure occurred when executi
```

Error tersebut dapat diatasi dengan mengeksekusi kueri berikut.

```
EXEC sp_changedbowner 'sa'
```

- 4) Setelah kueri tersebut berhasil dijalankan, basis data AdventureWorks2014 akan diaktifkan dan nilai is_cdc_enabled menjadi 1.

```
SELECT name AS Databases, is_cdc_enabled FROM sys.databases
```

	Databases	is_cdc_enabled
1	master	0
2	tempdb	0
3	model	0
4	msdb	0
5	ReportServer\$SQLENTERPRISE	0
6	ReportServer\$SQLENTERPRISETempDB	0
7	AdventureWorks2014	1

- 5) Kemudian aktifkan CDC Capture pada tabel Customer dengan mengeksekusi SQL Statement berikut.

```
declare @role_name varchar  
declare @supports_net_changes int  
EXEC sp_cdc_enable_table 'Sales', 'Customer', @role_name = NULL, @supports_net_changes = 1
```

- 6) Jika telah berhasil dieksekusi, maka akan muncul pada System Table seperti berikut.

□	System Tables
⊕	cdc.captured_columns
⊕	cdc.change_tables
⊕	cdc.ddl_history
⊕	cdc.index_columns
⊕	cdc.lsn_time_mapping
⊕	cdc.Sales_Customer_CT
⊕	dbo.systranschemas

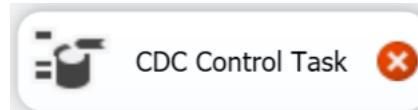
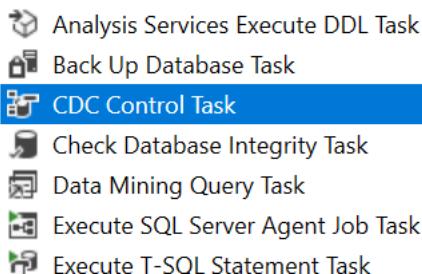
- 7) Setelah SQL statement tersebut berhasil dieksekusi, berikut akan ditampilkan Customer yang aktif dengan kueri berikut.

```
SELECT name, is_tracked_by_cdc FROM sys.tables
```

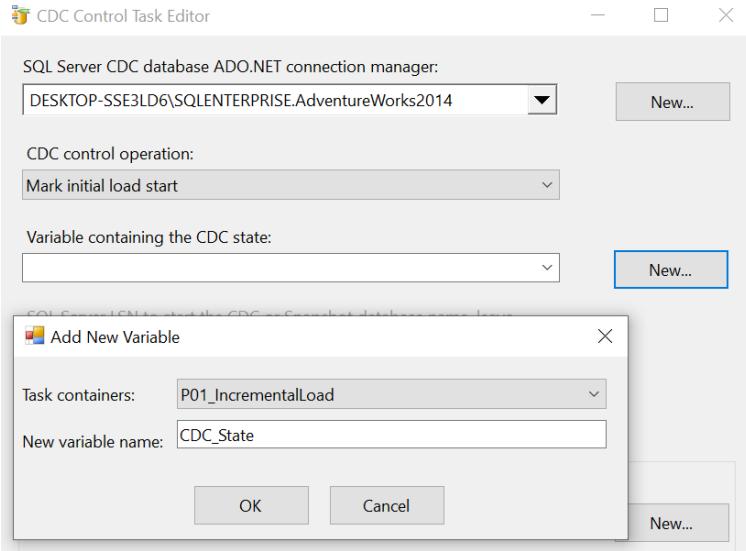
	name	is_tracked_by_cdc
1	ScrapReason	0
2	Shift	0
3	ProductCategory	0
4	ShipMethod	0
5	ProductCostHistory	0
6	ProductDescription	0
7	ShoppingCartItem	0
8	ProductDocument	0
9	DatabaseLog	0
10	ProductInventory	0
11	SpecialOffer	0
12	ErrorLog	0
13	ProductListPriceH...	0
14	Address	0
15	SpecialOfferProd...	0
16	ProductModel	0

- 8) Pada SQL Server Data Tools 2015, buatlah sebuah ADO.Net connection ke basis data AdventureWorks2014. Ikuti langkah yang sama seperti pada bagian A yang telah dibuat sebelumnya.
- 9) Kemudian dari SSIS Toolbox, pilih CDC Control Task, lalu drag and drop ke halaman control flow. Hal ini bertujuan untuk mulai mengisi log sequence number (LSN) terbaru dan konfigurasi tabel SQL untuk menyimpan CDC state.

▲ Other Tasks



- 10) Double click pada CDC Control Task. Kemudian pilih New untuk menambahkan Variable Containing the CDC State.



- 11) Kemudian pilih sebuah tabel untuk menyimpan CDC state untuk ekstraksi di masa yang akan datang. Pada Table to use for storing state pilih New.

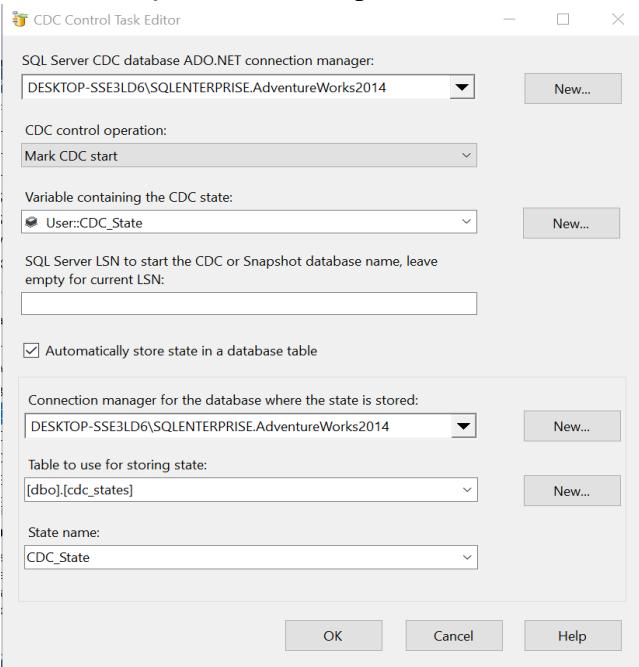
The screenshot shows the 'Create New State Table' dialog. It contains a SQL script for creating a table:

```

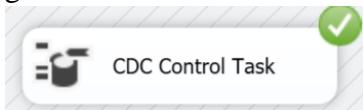
CREATE TABLE [dbo].[cdc_states]
([name] [nvarchar](256) NOT NULL,
[state] [nvarchar](256) NOT NULL) ON [PRIMARY]
GO
CREATE UNIQUE NONCLUSTERED INDEX [cdc_states_name] ON
[dbo].[cdc_states]
( [name] ASC )
WITH (PAD_INDEX = OFF) ON [PRIMARY]
GO

```

- 12) Klik Run untuk membuat tabel pada basis data AdventureWorks2014.
- 13) Pada CDC control operation, pilih Mark CDC start.
- 14) Jika semua *field* telah diisi pilih OK.



- 15) Klik F5 untuk mengeksekusi package dan CDC state terbaru yang telah dibuat.
- 16) Jika telah berhasil dieksekusi pada CDC Control Task terdapat centang hijau seperti pada gambar.



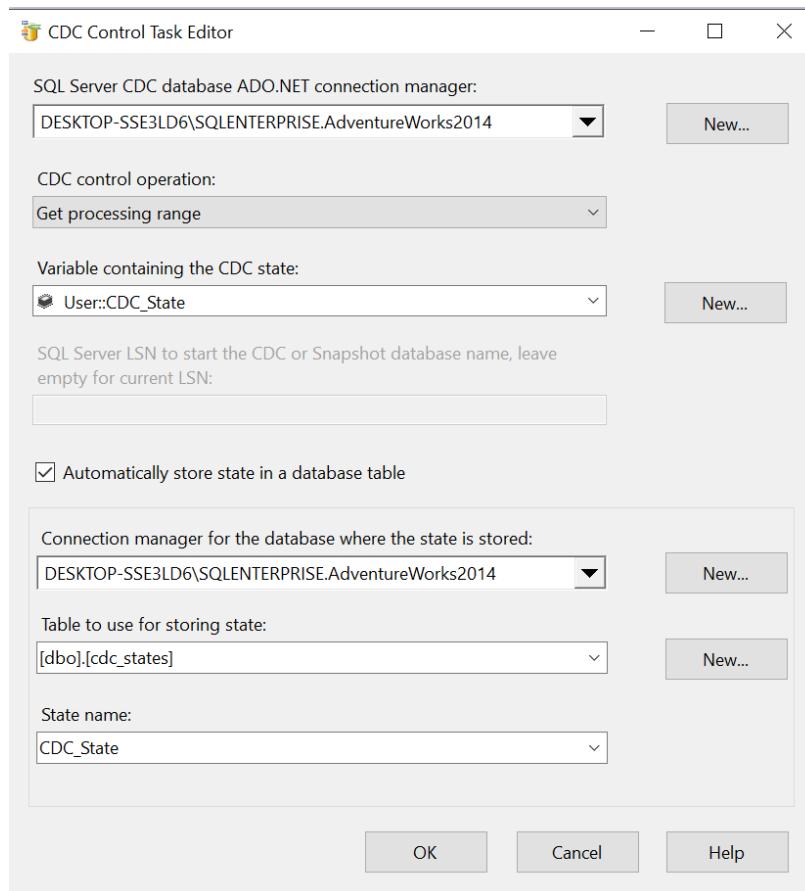
- 17) Pada SQL Server 2014, eksekusi kueri berikut untuk menampilkan *last access register* sebagai nomor LSN.

```
15 | SELECT * FROM [AdventureWorks2014].[dbo].[cdc_states]
```

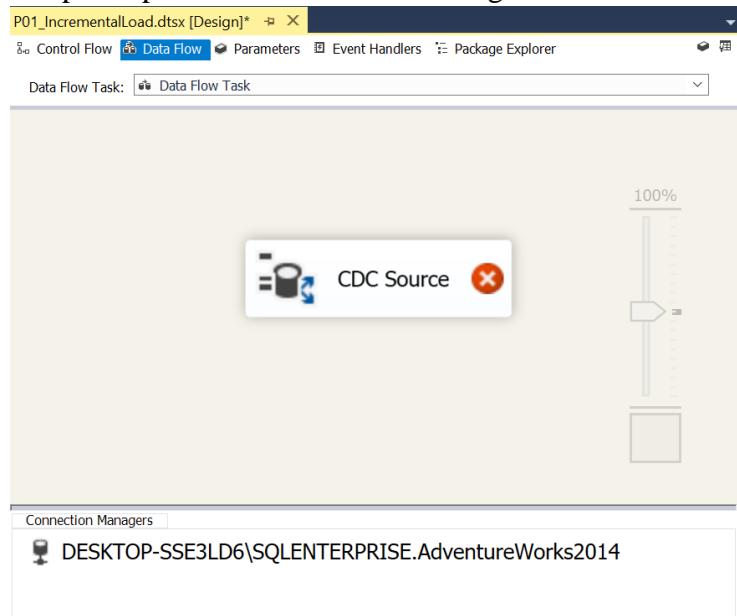
100 % <

name	state
CDC_State	TFEND/CS/0x000003000009400002/TS/2020-03-26T16...

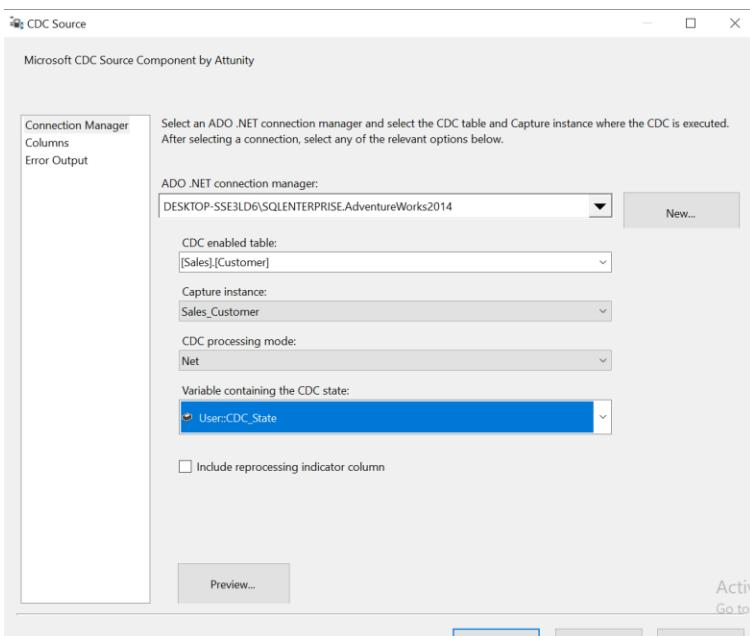
- 18) Kemudian dari SSIS Toolbox, lakukan drag and drop CDC Control Task untuk mendapatkan rentang pemrosesan data yang berubah sejak ekstraksi terakhir.
- 19) Kemudian isi data yang sama seperti data sebelumnya pada CDC Control Task Editor dengan mengganti property CDC control operation ke Get Processing Range yang akan memungkinkan perubahan dibuat antara LSN yang disimpan sebelumnya dan LSN yang dibuat oleh CDC task ini.



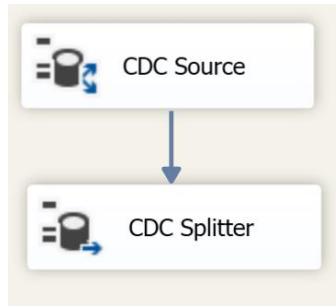
- 20) Setelah itu lakukan drag and drop Data Flow Task dari SSIS Toolbox ke design area control flow.
- 21) Kemudian double click pada komponen Data Flow.
- 22) Lakukan drag and drop komponen CDC Source ke design area data flow.



- 23) Double click komponen CDC Source, kemudian atur property ADO.Net connection manager ke connection yang telah dibuat sebelumnya.
- 24) Dalam properti CDC enabled table pilih [Sales].[Customer].
- 25) Properti CDC processing mode ke Net.
- 26) Atur properti Variable Containing the CDC state ke variabel SSIS CDC_state yang telah dibuat.

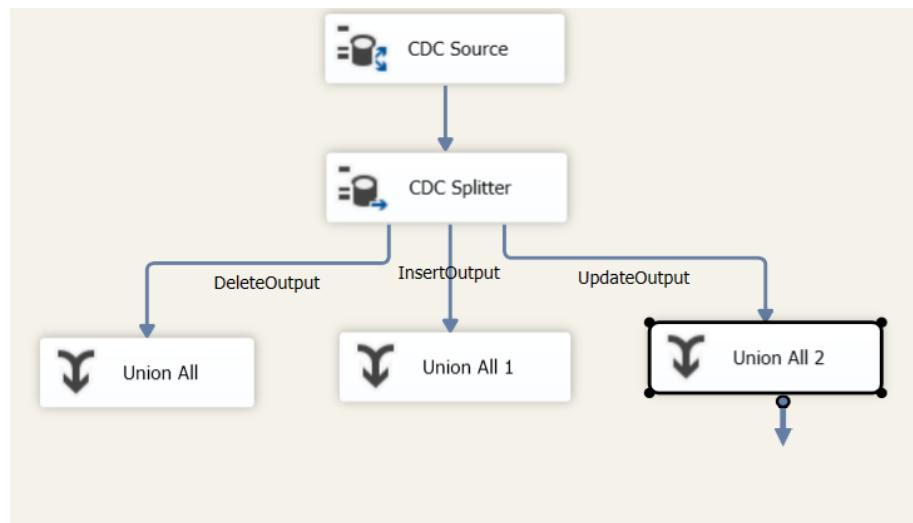


- 27) Drag and drop komponen CDC Splitter dari SSIS Toolbox ke Data Flow design area.
- 28) Kemudian hubungkan kedua komponen tersebut.



- 29) Drag-and-drop tiga komponen Union All dari SSIS Toolbox | Common ke setiap luaran yang disediakan oleh komponen CDC Splitter dan berdasarkan __\$operation:

- __\$operation = 1 Delete
- __\$operation = 2 Insert
- __\$operation = 3 Update before change
- __\$operation = 4 Update after change



- 30) Masukkan Customer yang baru dalam tabel Customer menggunakan SQL statement berikut ini:

```
INSERT INTO [Sales].[Customer] (PersonID, StoreID, TerritoryID, ModifiedDate) VALUES
('2521', '1400', '4', '2014-09-12 11:15:07.263')
```

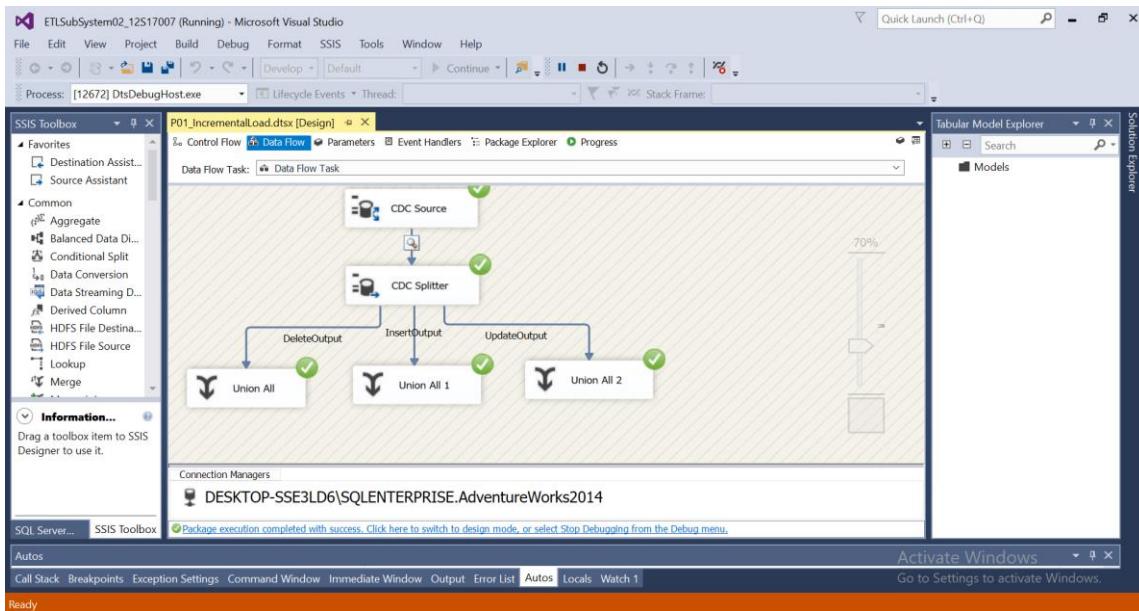
- 31) Perbarui customer yang telah ada dengan menggunakan statement berikut ini:

```
UPDATE [Sales].[Customer] SET StoreID='1994' WHERE CustomerID=3
```

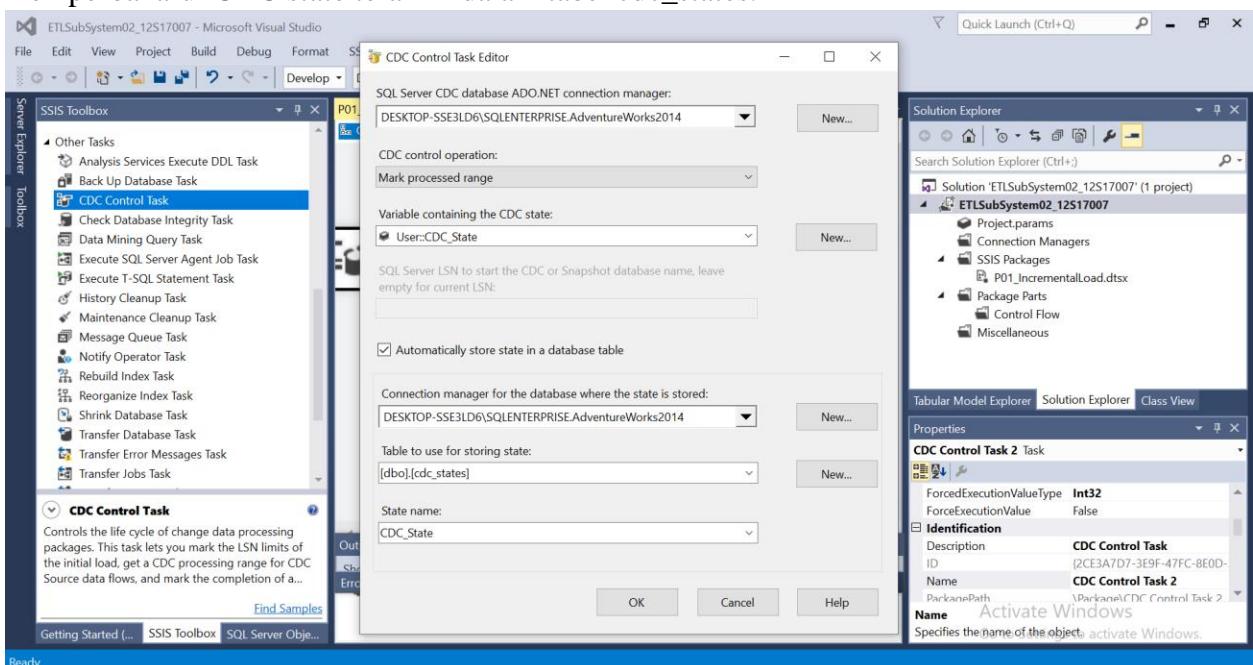
- 32) Kemudian execute package yang telah dibuat di SQL Server Data Tools dengan menekan F5.

- 33) Jika eksekusi berhasil maka akan ditampilkan seperti gambar berikut.

Output Data Viewer at Data Flow Task										
Detach Copy Data										
__\$start_lsn	__\$operation	__\$update_mask	CustomerID	PersonID	StoreID	TerritoryID	AccountNumber	rowguid	ModifiedDate	
0x00 0x00 0x00 0x3...	2	NULL	30120	2521	1400	4	NULL	efc58d31-1f6b-447...	2014-09-12 11:15:07.263	



- 34) Drag-and-drop CDC Control Task dari SSIS Toolbox | Other Tasks ke Control Flow design area.
- 35) Klik dua kali task tersebut atau klik kanan task tersebut kemudian pilih Edit... untuk membuka panel CDC Control Task Editor. Gunakan konfigurasi yang sama dengan task sebelumnya. Dalam property CDC control operations, gunakan nilai Mark processed range untuk memperbarui CDC state terakhir dalam tabel cdc_states.



- 36) Kemudian tekan F5 untuk mengeksekusi package tersebut.
- 37) Setelah dilakukan pengeksekusian, tidak ada yang ditampilkan. Hal ini berarti bahwa tidak ada perubahan yang dideteksi oleh CDC Source dalam Data Flow.

C. Data Cleansing System

Data Cleansing adalah proses mendeteksi dan memperbaiki (atau menghapus) data yang rusak atau tidak akurat dengan cara mengatur kembali data-data yang ada pada tabel atau basis data. Data Cleansing System dapat dilakukan menggunakan SSIS pada SQL Server Data Tools. Berikut adalah langkah-langkahnya:

Kebutuhan:

- AdventureWorks2012
- 1) Pada bagian Data Cleansing System ini, kita menggunakan AdventureWorks2012 karena pada database ini terdapat data yang perlu dibersihkan atau bersifat dirty.
 - 2) Tabel yang kita gunakan pada bagian ini adalah [Person].[Address]. Untuk melihat isi dari tabel tersebut, klik kanan pada tabel kemudian pilih Select Top 1000 Rows.

```
1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP 1000 [AddressID]
3      ,[AddressLine1]
4      ,[AddressLine2]
5      ,[City]
6      ,[StateProvinceID]
7      ,[PostalCode]
8      ,[SpatialLocation]
9      ,[rowguid]
10     ,[ModifiedDate]
11  FROM [AdventureWorks2012].[Person].[Address]
```

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	rowguid
1	1	1970 Napa Ct.	NULL	Bothell	79	98011	0xE6100000010CAE8BFC28BCE4474067A89189898A5EC0	9AADCB0D-36CF-4
2	2	9833 Mt. Dias Blv.	NULL	Bothell	79	98011	0xE6100000010CD6FA851AE6D74740BC262A0A03905EC0	32A54B9E-E034-4f
3	3	7484 Roundtree Drive	NULL	Bothell	79	98011	0xE6100000010C18E304C4ADE14740DA930CT7893915EC0	4C506923-6D1B-4f
4	4	9539 Glenside Dr	NULL	Bothell	79	98011	0xE6100000010C813A0D5F9FDE474011A5C28A7C955EC0	E5946C78-4BCC-4
5	5	1226 Shoo St.	NULL	Bothell	79	98011	0xE6100000010C61C64D8ABBD94740C460EA3FD8855EC0	FBAFF937-4A97-4f
6	6	1399 Firestone Drive	NULL	Bothell	79	98011	0xE6100000010CE0B4E50458DA47402F12A5F80C975EC0	FEBF8191-9804-44
7	7	5672 Hale Dr.	NULL	Bothell	79	98011	0xE6100000010C18E304C4ADE1474011A5C28A7C955EC0	0175A174-6C34-4f
8	8	6387 Scenic Avenue	NULL	Bothell	79	98011	0xE6100000010C0029A5D93BDF4740E248962FD5975EC0	3715E813-4DCA-4f
9	9	8713 Yosemite Ct.	NULL	Bothell	79	98011	0xE6100000010C6A80AD742DDC4740851574F7198C5EC0	268AF621-76D7-4C
10	10	250 Race Court	NULL	Bothell	79	98011	0xE6100000010C219D6AE1FE4474040862564B7825EC0	0B6B739D-8EB6-4
11	11	1318 Lasalle Street	NULL	Bothell	79	98011	0xE6100000010C219D6AE1FE4474040862564B7825EC0	981B3303-ACAA-4f
12	12	5415 San Gabriel Dr.	NULL	Bothell	79	98011	0xE6100000010C6A80AD742DDC474011A5C28A7C955EC0	1C2C9CFE-AB9F-4
13	13	9265 La Paz	NULL	Bothell	79	98011	0xE6100000010CC3228C0867E84740F3BC169BF835EC0	E0BA2F52-C907-4!
14	14	937 W. 1st	NULL	Bothell	79	98011	0xE6100000010C1020A44E15447407A320300A5EC0	A10050AE-0F55-4

Query executed successfully. DESKTOP-SSE3LD6\SQLENTERPRI... DESKTOP-SSE3LD6\ITD-ST... AdventureWorks2012 00:00:00 | 1000 rows

- 3) Pada tabel [Person].[Address] terdapat data yang dirty pada kolom City yaitu penulisan nama kota tidak diawali dengan huruf kapital sehingga hal ini menyebabkan perlunya dilakukan Data Cleansing.
- 4) Eksekusi kueri berikut untuk mendapatkan data yang dirty yaitu nama kota yang tidak diawali dengan huruf kapital.

```
13  SELECT * FROM [AdventureWorks2012].[Person].[Address]
14  WHERE SUBSTRING([City], 1, 1) <> UPPER (SUBSTRING([City], 1, 1)) COLLATE Latin1_General_CS_AS
```

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	rowguid
1	366	1619 Mills Dr.	NULL	albany	58	97321	0xE6100000010C128B69C2B55B4640910864B2FAC35EC0	A61B581D-A6DB-
2	376	9098 Story Lane	NULL	albany	54	12210	0xE6100000010C525A3CE36852454094FAB5CD486952C0	4DDA7288-7870-4
3	410	2565-175 Mitchell Road	NULL	alexandria	50	2015	0xE6100000010C24559C4BD2EC40C0AF79D20108ED6240	1154ED96-CF5C-4
4	441	Level 59	NULL	alexandria	50	2015	0xE6100000010C697AD82969F040C08D5B3B3E52E06240	B9DD0011-99D8-4
5	560	800 Interchange Blvd.	Suite 2501	austin	73	78701	0xE6100000010CB79C60EAEC413E40188A3E72166858C0	5790365C-DA3C-4
6	883	Heritage Mall	NULL	albany	58	97321	0xE6100000010C128B69C2B55B464008C4035A5FD15EC0	CF9C1921-D741-4
7	896	2255 254th Avenue Se	NULL	albany	58	97321	0xE6100000010C9C923D96B253464062C467118EC55EC0	77D18072-B9B8-4
8	11380	165 North Main	NULL	austin	73	78701	0xE6100000010C94212031DB3A3E4020A10CBA196658C0	412E3088-1FB8-4,

- 5) Buatlah sebuah proyek yang baru dan berikan nama ETLSubSytem04_12S17007.
- 6) Berikan sebuah nama dan lokasi untuk proyek SSIS tersebut.
- 7) Pilih Package yang telah dibuat secara *default* dan *rename* menjadi P01_DataCleansing.dtsx.
- 8) Dari SSIS Toolbox, drag and drop Data Flow Task ke design area Control Flow.
- 9) Untuk menyesuaikan dengan Data Cleansing yang akan kita lakukan, *rename* Data Flow Task tersebut menjadi Data Clean Task.



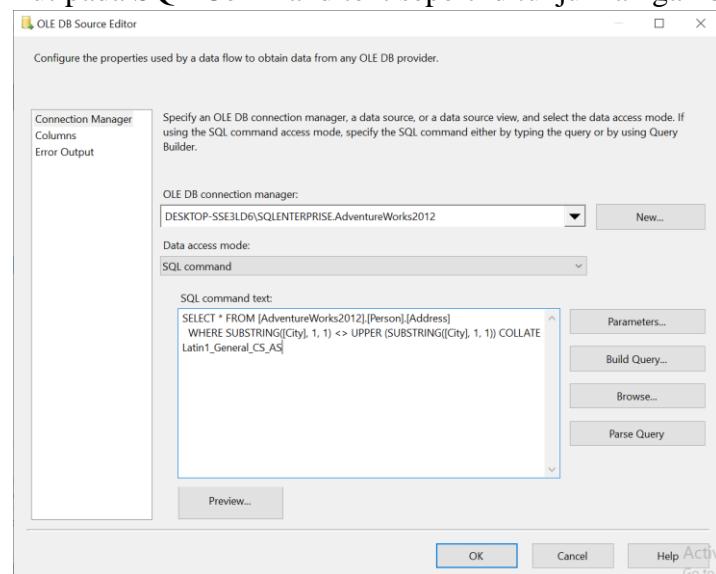
menjadi

- 10) Kemudian Double Click pada Data Cleanse Task.
- 11) Drag and Drop OLE DB Source ke design area Data Flow, dan *rename* menjadi Person_Address sesuai dengan tabel yang akan kita Clean.



menjadi

- 12) Double Click pada Person_Address.
- 13) Kemudian pilih OLE DB Connection Manager sesuai dengan device kita dan database serta tabel yang kita gunakan.
- 14) Pada Data Access Mode pilih SQL Command.
- 15) Ketikkan kueri berikut pada SQL Command text seperti ditunjukkan gambar berikut.



16) Pilih Preview untuk melihat data yang akan kita gunakan.

The screenshot shows a Windows application window titled "Preview Query Results". Inside, there is a table with 10 columns and approximately 200 rows of data. The columns are labeled: AddressID, AddressLine1, AddressLine2, City, StateProvinceID, PostalCode, SpatialLocation, Rowguid, and ModifiedDate. The data includes various city names like Albany, Austin, and Baldwin Park, along with their respective state/province IDs and postal codes.

AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	Rowguid	ModifiedDate
366	1619...	NULL	albany	58	97321	Syste...	a61b5...	3/1/2...
376	9098...	NULL	albany	54	12210	Syste...	4dd47...	3/5/2...
410	2565...	NULL	alexa...	50	2015	Syste...	1154e...	9/1/2...
441	Level 59	NULL	alexa...	50	2015	Syste...	b9dd...	8/1/2...
560	800...	Suite ...	austin	73	78701	Syste...	5790...	7/1/2...
883	Herita...	NULL	albany	58	97321	Syste...	d9c1...	7/1/2...
896	2255...	NULL	albany	58	97321	Syste...	77d1...	9/1/2...
11380	165 N...	NULL	austin	73	78701	Syste...	412e3...	7/1/2...

17) Kemudian pilih Parse Query.

18) Pilih OK.

19) Pada SQL Server, buat sebuah tabel yang menyimpan daftar City. Eksekusi kueri berikut.

```
1 CREATE TABLE [Person].[City]
2     City_Name NVARCHAR (255)
3 )
```

20) Isi tabel tersebut sesuai data City yang terdapat pada tabel [Person].[Address].

```
5 INSERT INTO Person.City (City_Name) VALUES
6 ('Albany')
7
8 INSERT INTO Person.City (City_Name) VALUES
9 ('Alexandria')
10
11 INSERT INTO Person.City (City_Name) VALUES
12 ('Alhambra')
13
14 INSERT INTO Person.City (City_Name) VALUES
15 ('Austin')
16
17 INSERT INTO Person.City (City_Name) VALUES
18 ('Bad Soden')
19
20 INSERT INTO Person.City (City_Name) VALUES
21 ('Baldwin Park')
22
23 INSERT INTO Person.City (City_Name) VALUES
24 ('Balard')
25
26 INSERT INTO Person.City (City_Name) VALUES
27 ('Baltimor')
```

21) Kemudian eksekusi kueri berikut untuk melihat isi dari tabel yang telah kita buat.

```
29 SELECT * FROM [Person].[City]
```

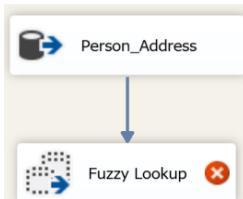
The screenshot shows the "Results" tab of the SQL Server Management Studio interface. It displays a table with one column, "City_Name", containing eight rows of data: Albany, Alexandria, Alhambra, Austin, Bad Soden, Baldwin Park, Balard, and Baltimor.

	City_Name
1	Albany
2	Alexandria
3	Alhambra
4	Austin
5	Bad Soden
6	Baldwin Park
7	Balard
8	Baltimor

22) Dari SSIS Toolbox, pilih Fuzzy Lookup. Kemudian drag and drop ke design area Data Flow. Fuzzy Lookup digunakan untuk mencocokkan dan melakukan standardisasi terhadap data.

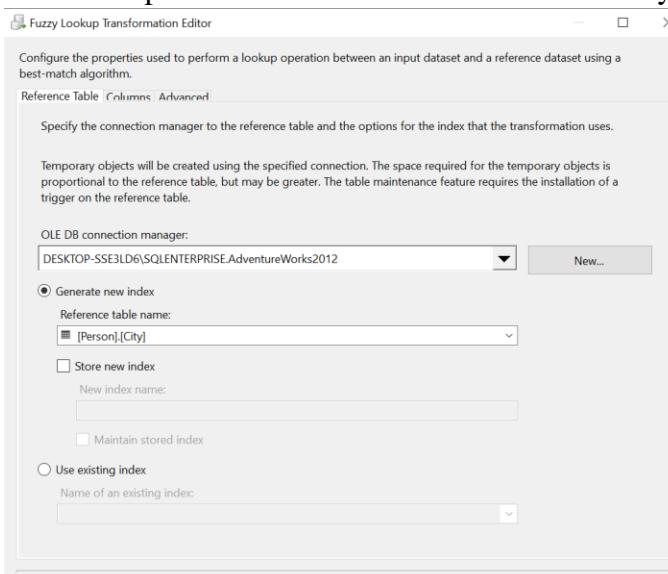


23) Kemudian hubungkan garis dari Person_Address ke Fuzzy Lookup.

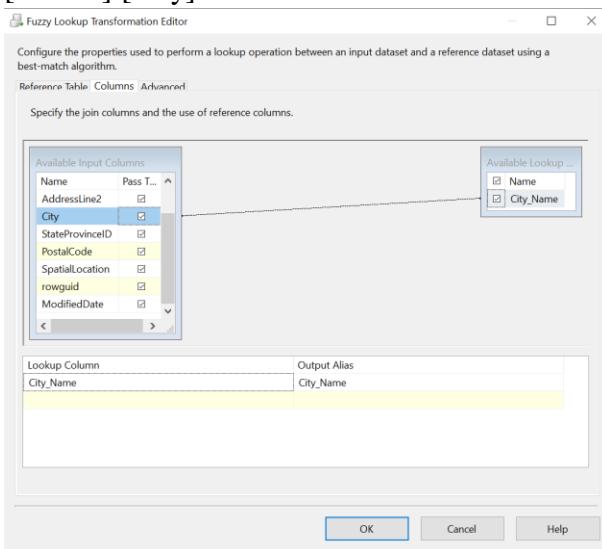


24) Double click pada Fuzzy Lookup.

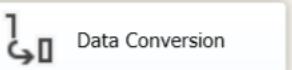
25) Kemudian pilih Reference Table Name: Person.City.



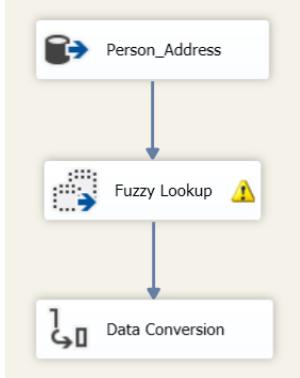
26) Setelah itu pilih Column, kita akan melakukan join antara [Person].[Address] dan [Person].[City].



27) Kemudian Drag and Drop Data Conversion dari SSIS Toolbox ke design area Data Flow.

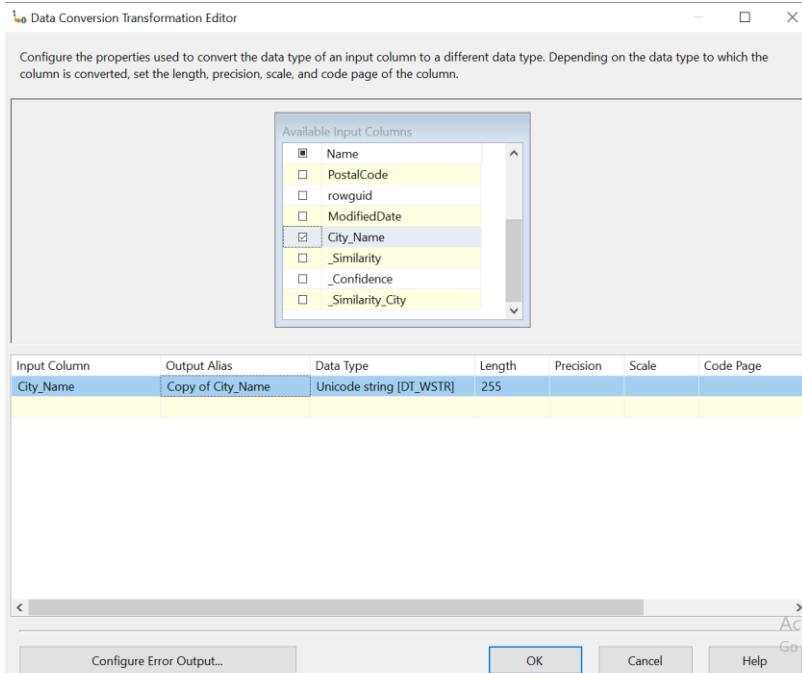


28) Hubungkan Fuzzy Lookup dengan Data Conversion.



29) Double click pada Data Conversion tersebut.

30) Kemudian pilih atribut atau data yang akan kita ubah, yaitu City_Name.



31) Rename Output Alias menjadi Formatted_City_Name.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
City_Name	Formatted_City_Name	Unicode string [DT_WSTR]	255			

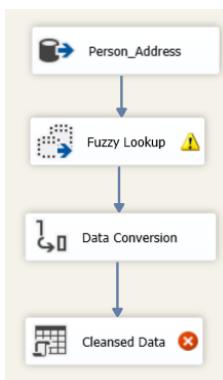
32) Kemudian kita ingin memasukkan data tersebut ke target kita. Pilih OLE DB Command dari SSIS Toolbox. Drag and Drop ke design area Data Flow.



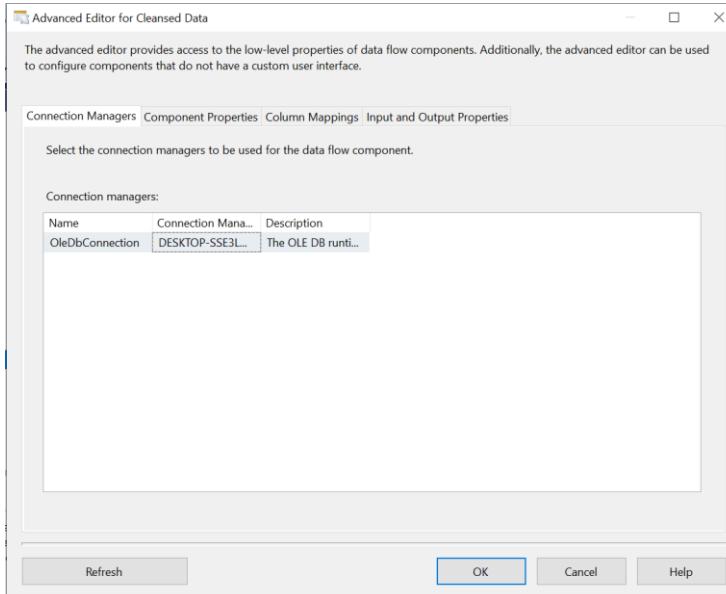
33) Kemudian hubungkan dari Data Conversion ke OLE DB Command.

34) Pada OLE DB Command kita dapat melakukan *update* pada tabel misalnya.

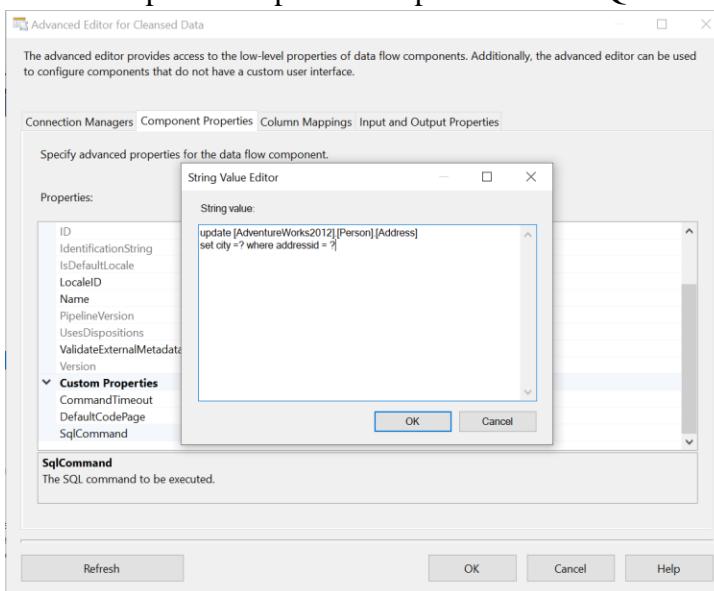
35) Untuk memudahkan kita mengenalinya, *rename* OLE DB Command menjadi Cleansed_Data.



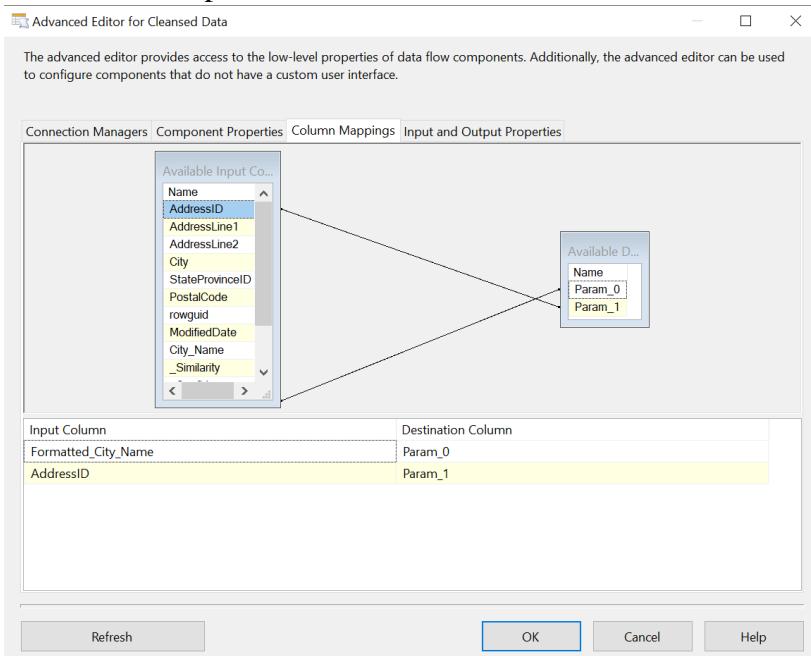
- 36) Kemudian double click pada Cleansed_Data.
 37) Pilih Connection Manager yaitu device kita.



- 38) Kemudian pilih Component Properties. Pada SQL Command isilah dengan kueri berikut.



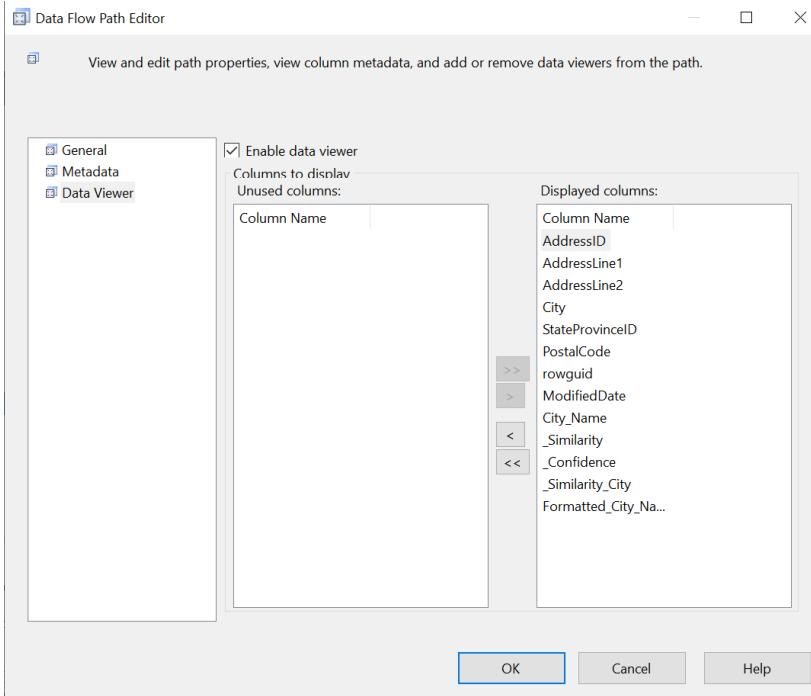
39) Pada Column Mappings, lakukan pemetaan Formatted_City_Name ke Param 0 dan Address_ID ke param 1.



40) Kemudian pilih OK.

41) Setelah itu, double click garis penghubung antara Data Conversion dan Cleansed_Data.

42) Pada Data Viewer, pilih enable data viewer.



43) Kemudian pada SQL Server, eksekusi kueri berikut untuk melihat kembali data yang dirty, yaitu City dengan nama austin, alexandria, albany.

```

2 |SELECT TOP 1000 [AddressID]
3 |    ,[AddressLine1]
4 |    ,[AddressLine2]
5 |    ,[City]
6 |    ,[StateProvinceID]
7 |    ,[PostalCode]
8 |    ,[SpatialLocation]
9 |    ,[rowguid]
10 |   ,[ModifiedDate]
11 |FROM [AdventureWorks2012].[Person].[Address]
12 |WHERE city in ('austin', 'alexandria', 'albany')

```

Results Spatial results Messages

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	rowguid
1	366	1619 Mills Dr.	NULL	albany	58	97321	0xE6100000010C128B69C2B55B4640910864B2FAC35EC0	A61B581D-A6DB-
2	11380	165 North Main	NULL	austin	73	78701	0xE6100000010C94212031DB3A3E4020A10CBA196658C0	412E3088-1FB8-4,
3	896	2255 254th Avenue Se	NULL	albany	58	97321	0xE6100000010C9C923D96B253464062C467118EC55EC0	77D18072-B99B-4
4	410	2565-175 Mitchell Road	NULL	alexandria	50	2015	0xE6100000010C24559C4BD2EC40C0AF79D20108ED6240	1154ED96-CF5C-4
5	560	800 Interchange Blvd.	Suite 2501	austin	73	78701	0xE6100000010CB79C60EAEC413E40188A3E72166858C0	5790365C-DA3C-4
6	376	9098 Story Lane	NULL	albany	54	12210	0xE6100000010C525A3CE36852454094FAB5CD486952C0	4DDA7288-7870-4
7	883	Heritage Mall	NULL	albany	58	97321	0xE6100000010C128B69C2B55B464008C4035A5FD15EC0	CF9C1921-D741-4
8	441	Level 59	NULL	alexandria	50	2015	0xE6100000010C697AD82969F040C08D5B3B3E52E06240	B9DD0011-99D8-4

44) Kemudian tekan F5 pada package di SQL Server Data Tools.

Data Conversion Output Data Viewer at Data Cleanse Task

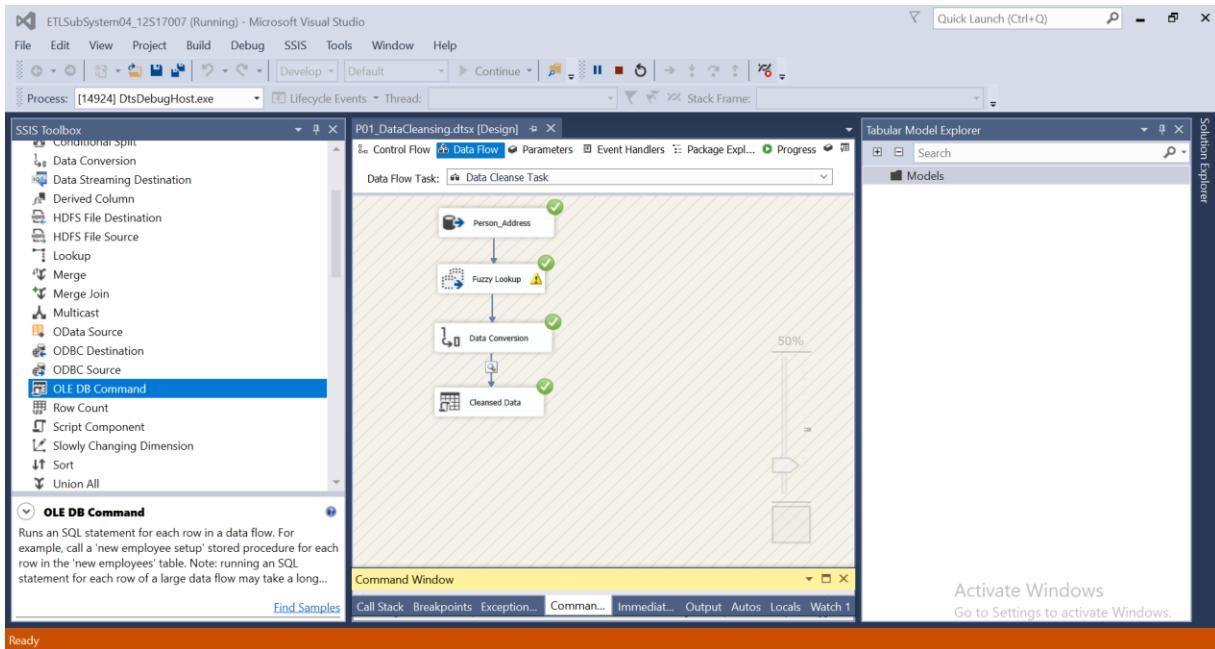
Detach Copy Data

Addr...	AddressLine1	AddressLine2	City	State...	PostalCode	rowguid	Modified...	City_Name	_Similarity	_Confidence	_Similarity_City	Formatted_City_
366	1619 Mills Dr.	NULL	albany	58	97321	a61b581d-a6db-4c...	2006-03...	Albany	1	1	1	Albany
376	9098 Story Lane	NULL	albany	54	12210	4dda7288-7870-43...	2006-03...	Albany	1	1	1	Albany
410	2565-175 Mitchell Road	NULL	alexandria	50	2015	1154ed96-cf5c-4ca...	2007-09...	Alexandria	1	1	1	Alexandria
441	Level 59	NULL	alexandria	50	2015	b9dd0011-99d8-4e...	2007-08...	Alexandria	1	1	1	Alexandria
560	800 Interchange Blvd.	Suite 2501	austin	73	78701	5790365c-d3c-43...	2006-07...	Austin	1	1	1	Austin
883	Heritage Mall	NULL	albany	58	97321	c9c1921-d741-45...	2007-07...	Albany	1	1	1	Albany
896	2255 254th Avenue Se	NULL	albany	58	97321	77d18072-b99b-4...	2007-09...	Albany	1	1	1	Albany
11380	165 North Main	NULL	austin	73	78701	412e3088-1fb8-4a...	2006-07...	Austin	1	1	1	Austin

Attached Total rows: 0, buffers: 0 Rows displayed = 8

45) Pada gambar di atas, kita dapat melihat Similarity dan Confidence yang bernilai 1 atau menandakan 100%. Hal ini berarti bahwa antara City dan City_Name terdapat kesamaan.

46) Pada design area Data Flow kita dapat melihat terdapat 8 rows yang terpengaruh atau ter-update.



- 47) Kemudian eksekusi kueri berikut untuk melihat apakah nama City yang dirty tersebut telah berubah menjadi diawali dengan huruf kapital.

```

2 |SELECT TOP 1000 [AddressID]
3 |    ,[AddressLine1]
4 |    ,[AddressLine2]
5 |    ,[City]
6 |    ,[StateProvinceID]
7 |    ,[PostalCode]
8 |    ,[SpatialLocation]
9 |    ,[rowguid]
10 |   ,[ModifiedDate]
11 |FROM [AdventureWorks2012].[Person].[Address]
12 |WHERE city in ('austin', 'alexandria', 'albany')

```

- 48) Hasilnya akan ditampilkan seperti berikut.

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	rowguid
1	366	1619 Mills Dr.	NULL	Albany	58	97321	0xE6100000010C128B69C2B55B4640910864B2FAC35EC0	A61B581D-A6DB-
2	11380	165 North Main	NULL	Austin	73	78701	0xE6100000010C94212031DB3A3E4020A10CBA196658C0	412E3088-1FB8-4
3	896	2255 254th Avenue Se	NULL	Albany	58	97321	0xE6100000010C9C923D96B253464062C467118EC55EC0	77D18072-B99B-4
4	410	2565-175 Mitchell Road	NULL	Alexandria	50	2015	0xE6100000010C24559C4BD2EC40C0AF79D20108ED6240	1154ED96-CF5C-4
5	560	800 Interchange Blvd.	Suite 2501	Austin	73	78701	0xE6100000010CB79C60EAEC413E40188A3E72166858C0	5790365C-DA3C-4
6	376	9098 Story Lane	NULL	Albany	54	12210	0xE6100000010C525A3CE36852454094FAB5CD486952C0	4DDA7288-7870-4
7	883	Heritage Mall	NULL	Albany	58	97321	0xE6100000010C128B69C2B55B464008C4035A5FD15EC0	CF9C1921-D741-4
8	441	Level 59	NULL	Alexandria	50	2015	0xE6100000010C697AD82969F040C08D5B3B3E52E06240	B9DD0011-99D8-

Query executed successfully. | DESKTOP-SSE3LD6\SQLENTERPRI... | DESKTOP-SSE3LD6\ITD-ST... | AdventureWorks2012 | 00:00:00 | 8 rows

- 49) Pada data tersebut, dapat kita lihat bahwa City telah berubah. Dengan kata lain, tidak ada lagi data yang dirty karena semua nama City telah diawali dengan huruf kapital.

D. Deduplication System

Deduplication System merupakan bagian dari ETL Component yang berperan untuk mengenali data yang duplikat dan menghilangkan data yang duplikat. Kita dapat melakukan deduplication system menggunakan SQL Server Data Tools dan SQL Server 2014. Berikut adalah tahapan pengerjaannya.

- 1) Pada bagian ini, kita akan menggunakan database AdventureWorks2014 dan tabel [Person].[Person].
- 2) Untuk melihat isi dari tabel tersebut, klik kanan pada tabel kemudian pilih Select Top 1000 Rows.

The screenshot shows the SQL Server Management Studio interface. In the Object Explorer, the AdventureWorks2014 database is selected. In the main pane, a query window titled 'SQLQuery1.sql' contains the following T-SQL code:

```
2  SELECT TOP 1000 [BusinessEntityID]
3      ,[PersonType]
4      ,[NameStyle]
5      ,[Title]
6      ,[FirstName]
7      ,[MiddleName]
8      ,[LastName]
9      ,[Suffix]
10     ,[EmailPromotion]
11     ,[AdditionalContactInfo]
12     ,[Demographics]
13     ,[rowguid]
14     ,[ModifiedDate]
15  FROM [AdventureWorks2014].[Person].[Person]
```

Below the code, the 'Results' tab is active, displaying the output of the query. The results show 12 rows of data from the Person table, including columns such as BusinessEntityID, PersonType, NameStyle, Title, FirstName, MiddleName, LastName, Suffix, EmailPromotion, AdditionalContactInfo, and Demographics. A message at the bottom indicates the query was executed successfully.

Query executed successfully. DESKTOP-SSE3LD6\SQLENTERPRISE... | DESKTOP-SSE3LD6\JTD-ST... | AdventureWorks2014 | 00:00:00 | 1000 rows

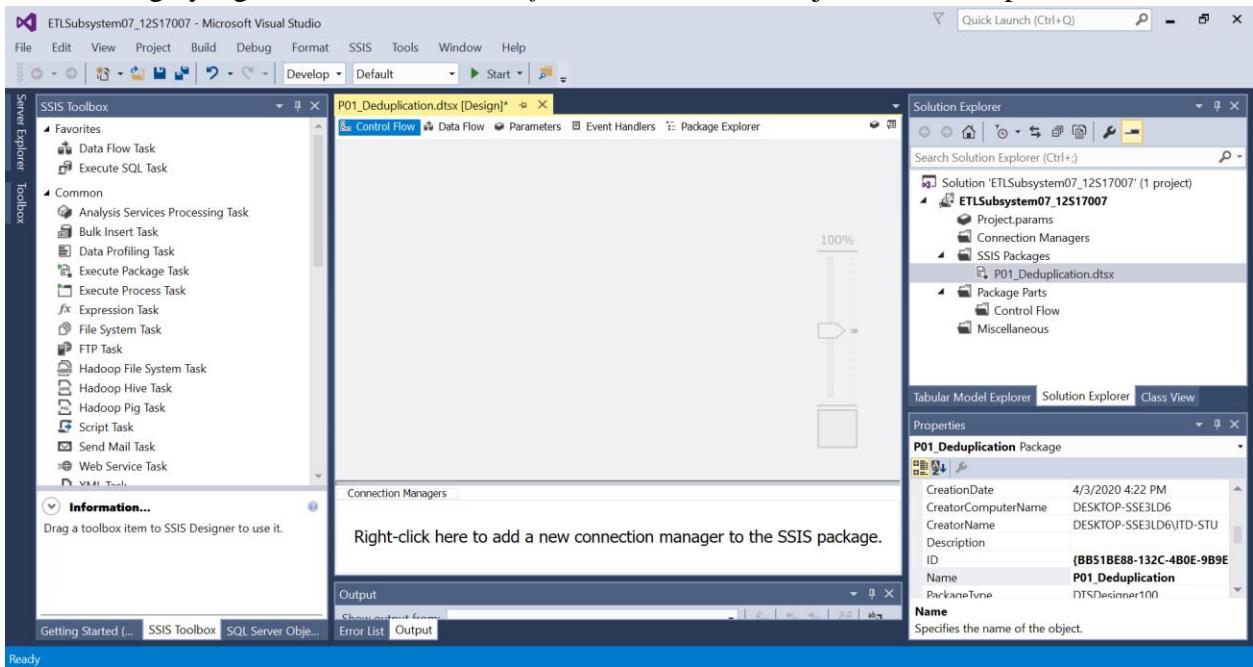
- 3) Pada tabel tersebut, terdapat data yang duplikat yaitu pada kolom FirstName dan LastName. Eksekusi kueri berikut untuk melihat FirstName dan LastName yang duplikat.

```
17  SELECT
18      FirstName, COUNT(FirstName),
19      LastName, COUNT(LastName)
20  FROM [AdventureWorks2014].[Person].[Person]
21  GROUP BY
22      FirstName,
23      LastName
24  HAVING COUNT(FirstName) > 1
25      AND COUNT(LastName) > 1;
```

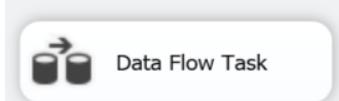
	FirstName	(No column name)	LastName	(No column name)
1	Kim	3	Abercrombie	3
2	Pilar	2	Ackerman	2
3	Jay	2	Adams	2
4	Jordan	2	Adams	2
5	Kim	3	Akers	3
6	Amy	3	Alberts	3
7	Sean	2	Alexander	2
8	Jordan	2	Allen	2
9	Michael	3	Allen	3
10	Gary	2	Altman	2
11	Chris	2	Ashton	2
12	Dan	2	Bacon	2
13	Mary	2	Baker	2
14	Angela	2	Barbariol	2
15	David	2	Barber	2
16	Abigail	2	Barnes	2
17	Josh	4	Barnhill	4
18	Shaun	2	Beasley	2
19	Bradley	1	Book	1

✓ Query executed successfully.

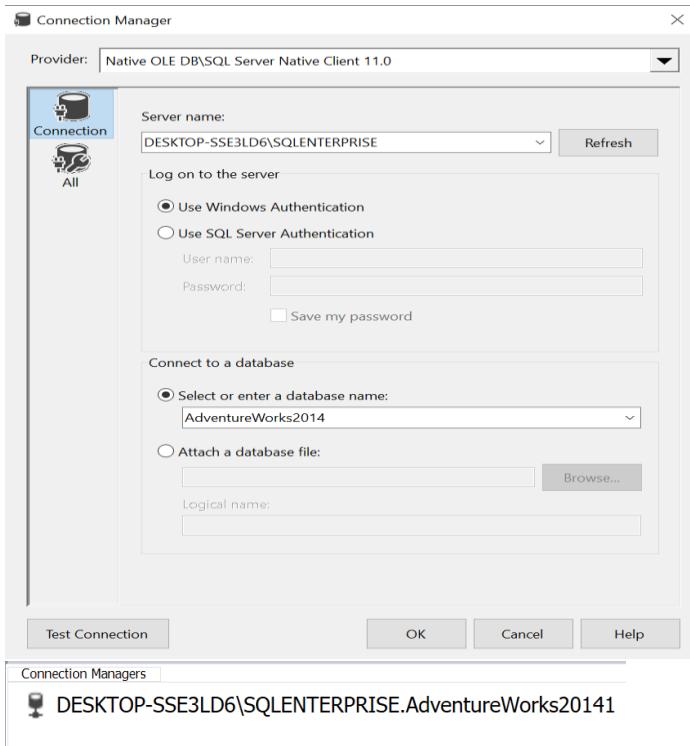
- 4) Pada data tersebut terdapat FirstName dan LastName yang sama.
- 5) Kita akan mengolah data yang duplikat tersebut menggunakan SQL Server Data Tools.
- 6) Buat sebuah proyek baru dan beri nama ETLSubsystem07_12S17007.
- 7) Pilih Package yang telah dibuat secara *default* dan *rename* menjadi P01_Deduplication.dtsx.



- 8) Drag and drop Data Flow Task dari SSIS Toolbox ke design area Control Flow.

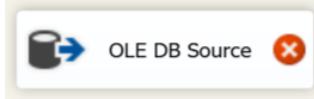


- 9) Kemudian double click pada Data Flow Task tersebut.
- 10) Pada Connection Manager, klik kanan kemudian pilih New OLE DB Connection. Pilih nama server dari device yang digunakan.
- 11) Jika belum ada Connection yang digunakan sebelumnya. Pilih New untuk membuatnya.



DESKTOP-SSE3LD6\SQLENTERPRISE.AdventureWorks20141

12) Dari SSIS Toolbox, Drag and Drop OLE DB Source ke design area Data Flow.

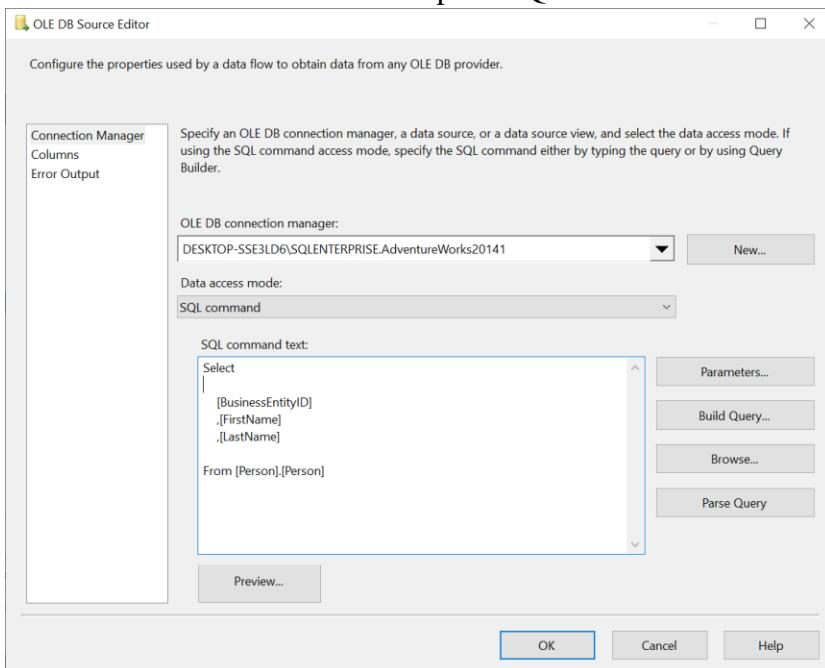


13) Double click pada OLE DB Source.

14) Pilih Connection yang telah dibuat sebelumnya.

15) Pada Data Access Mode pilih SQL Command.

16) Kemudian ketikkan kueri berikut pada SQL Command Text.



17) Pilih Preview untuk melihat isi tabel dari [Person].[Person].

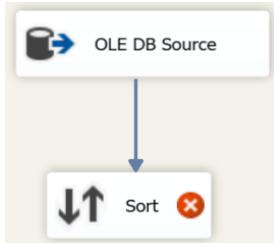
The screenshot shows a Windows application window titled "Preview Query Results". Inside, a table displays the results of a query. The columns are labeled "BusinessEntityID", "FirstName", and "LastName". The data includes rows such as (285, Syed, Abbas), (293, Catherine, Abel), (295, Kim, Abercrombie), etc., up to row 10292, Adams.

BusinessEntityID	FirstName	LastName
285	Syed	Abbas
293	Catherine	Abel
295	Kim	Abercrombie
2170	Kim	Abercrombie
38	Kim	Abercrombie
211	Hazem	Abolrous
2357	Sam	Abolrous
297	Humberto	Acevedo
291	Gustavo	Achong
299	Pilar	Ackerman
121	Pilar	Ackerman
16867	Aaron	Adams
16901	Adam	Adams
16724	Alex	Adams
10263	Alexandra	Adams
10312	Allison	Adams
10274	Amanda	Adams
10292	Amber	Adams

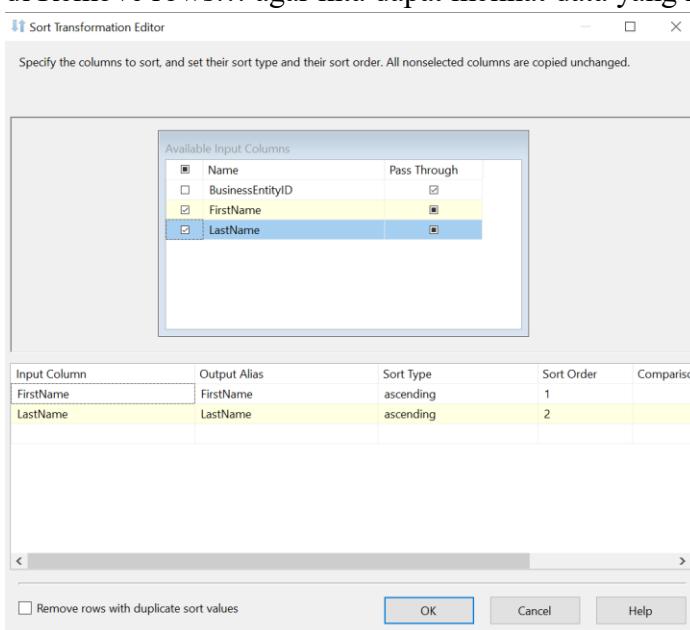
18) Dari SSIS Toolbox, pilih Sort kemudian drag and drop ke design area Data Flow.



19) Hubungkan OLE DB Source dengan Sort. Sort ini digunakan untuk mensort FirstName.



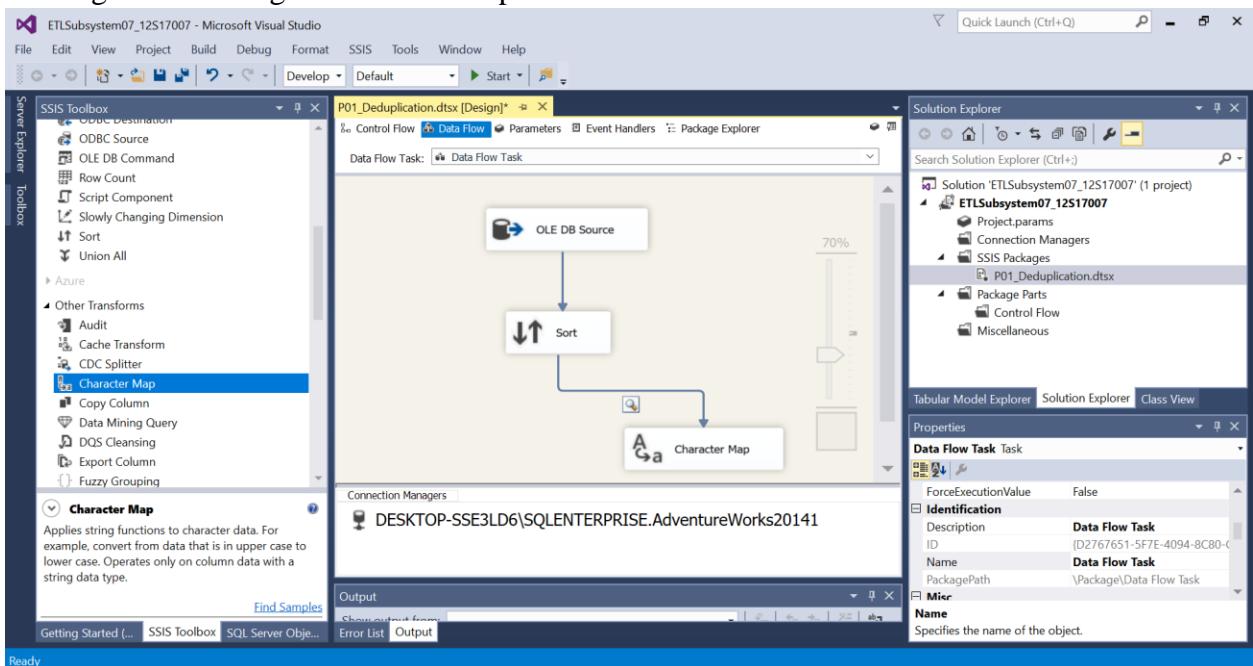
20) Double Click pada Sort, kemudian pilih column apa saja yang ingin disort. Dalam hal ini terdapat 2 kolom yang akan disort yaitu FirstName dan LastName. Jangan beri tanda centang di Remove rows... agar kita dapat melihat data yang redundan terlebih dahulu.



21) Kita memerlukan sebuah penampung data pada SSIS. Dari SSIS Toolbox, drag and drop Character Map ke design area Data Flow.

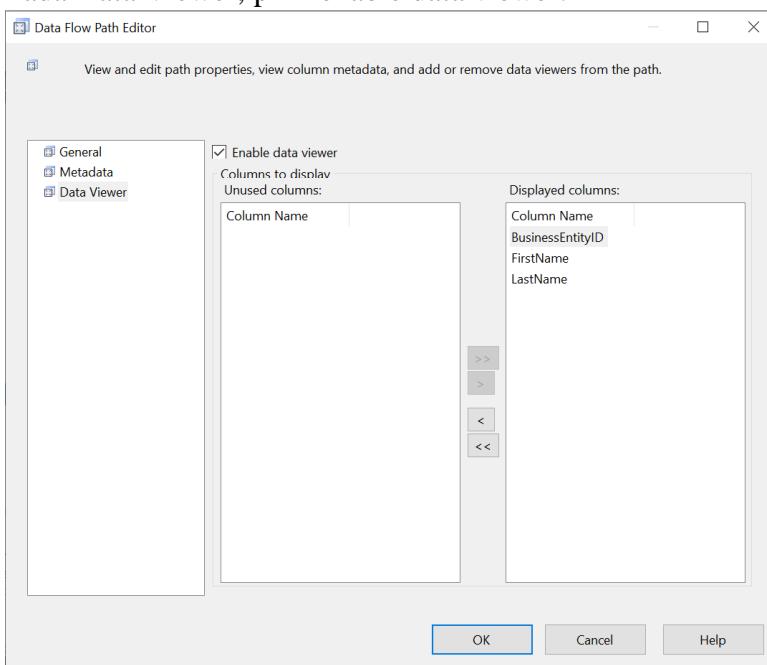


22) Hubungkan Sort dengan Character Map.

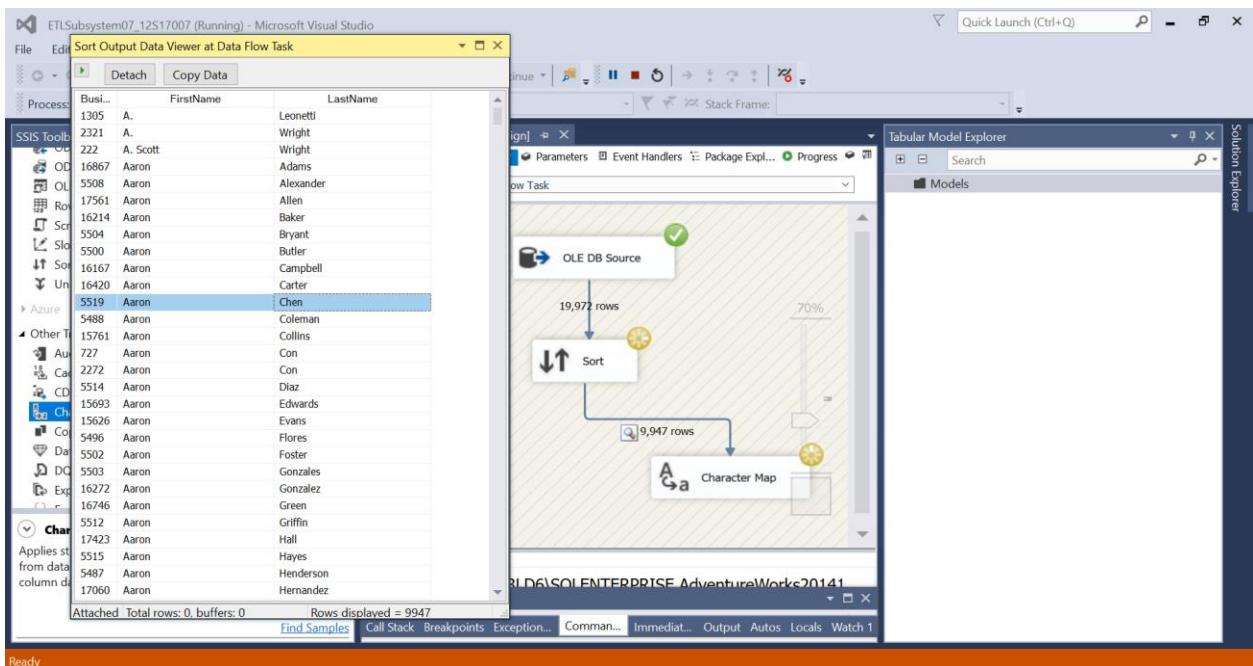


23) Double click pada garis yang menghubungkan Sort dengan Character Map.

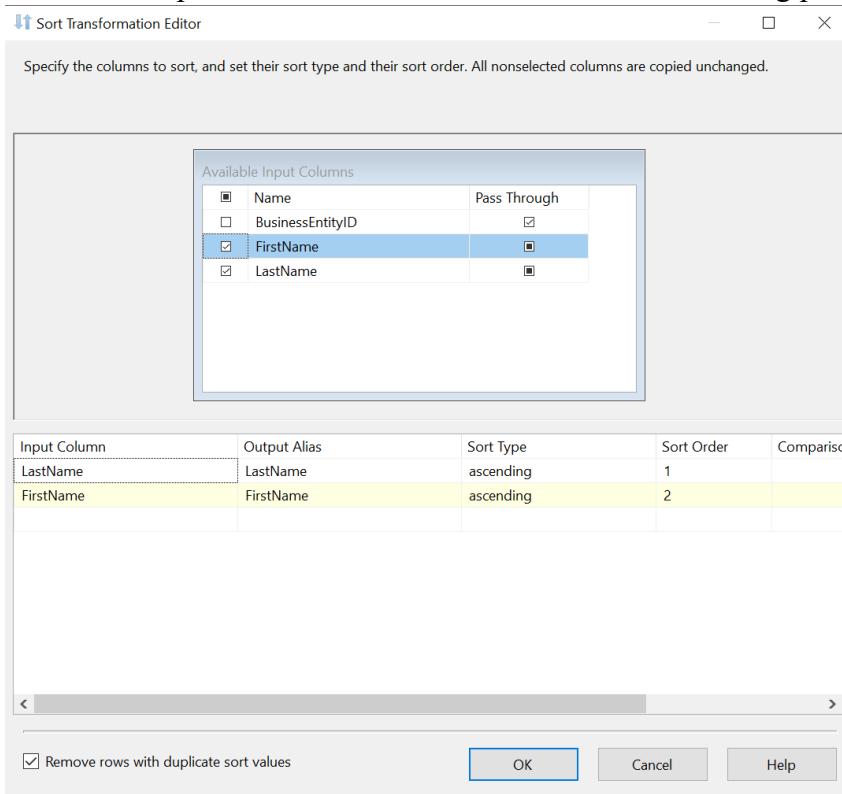
24) Pada Data Viewer, pilih enable data viewer.



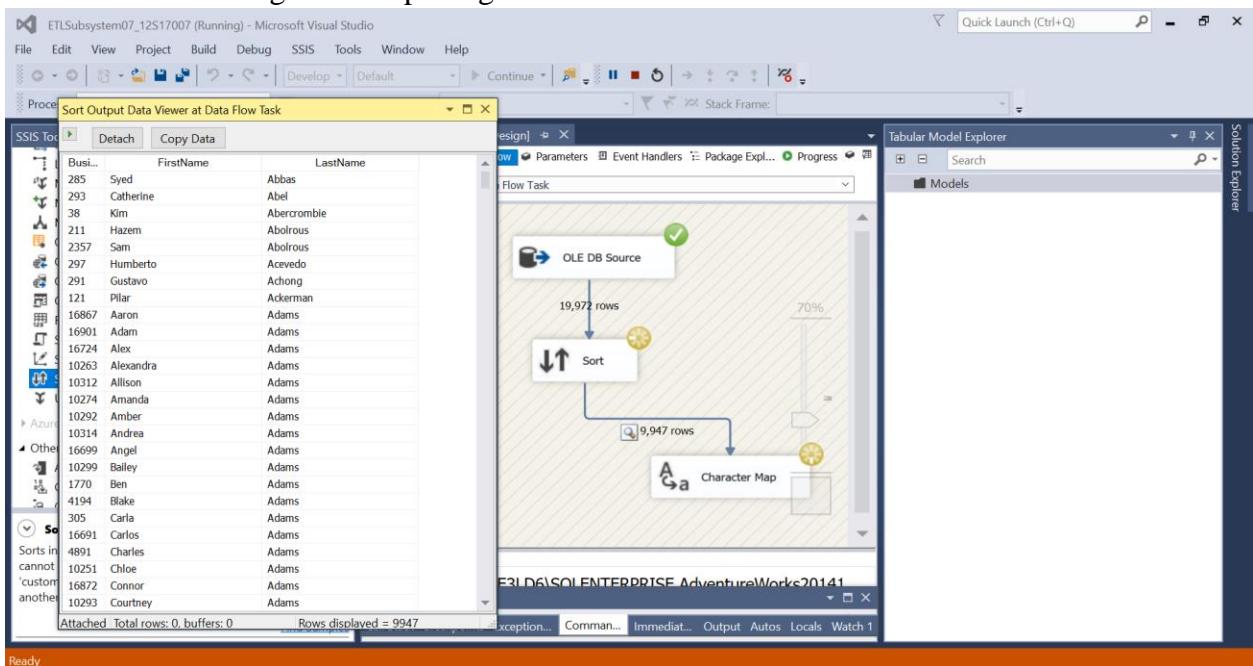
25) Eksekusi package tersebut, tekan F5.



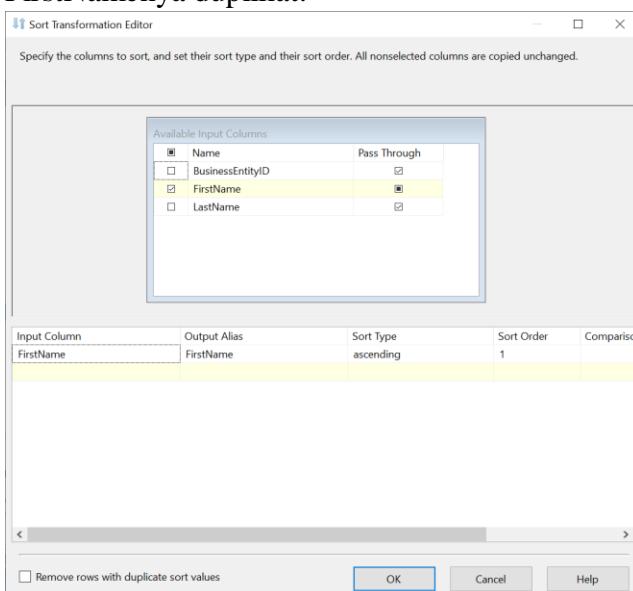
- 26) Pada data tersebut, kita lihat bahwa data input sebanyak 19,972 rows sementara data output sebanyak 9,947 rows. Hal ini berarti terdapat data yang redundan sebanyak 9,947 rows.
- 27) Selain itu kita juga dapat melihat terdapat data yang redundan, baik pada FirstName maupun LastName.
- 28) Kemudian stop pada debugging tersebut.
- 29) Double click pada Sort tersebut, kemudian beri tanda centang pada Remove rows...



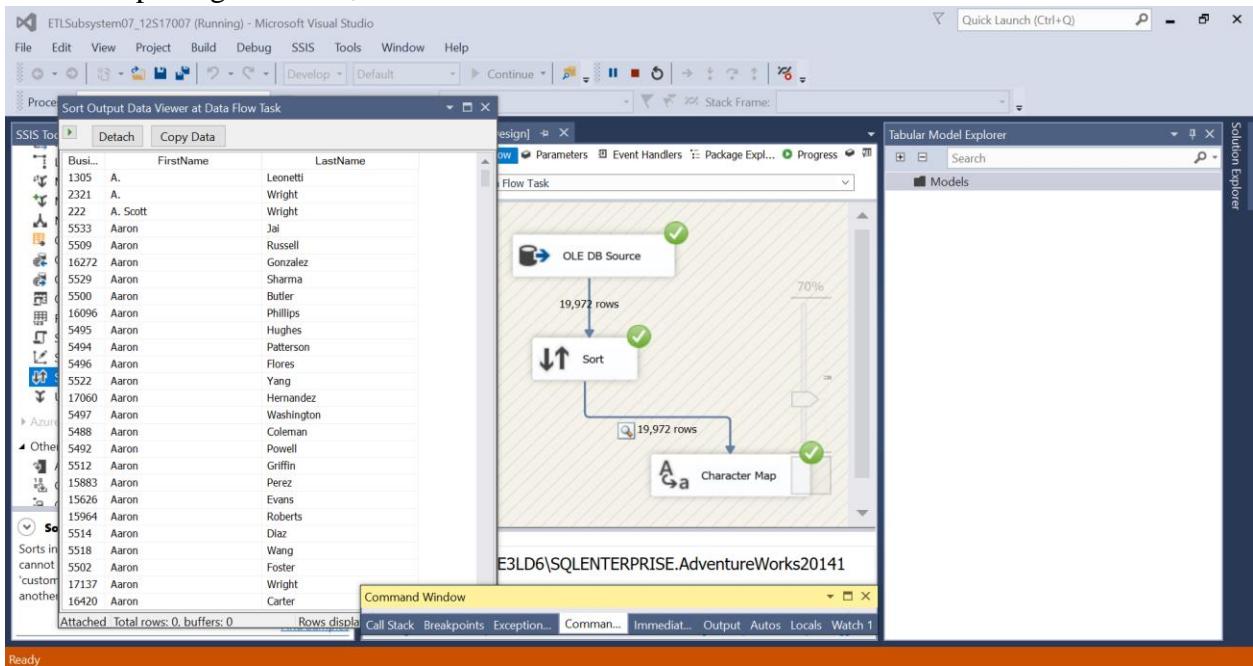
30) Tekan F5 untuk mengeksekusi package tersebut.



- 31) Data yang ditampilkan memiliki input sebanyak 19,972 rows dan output sebanyak 9,947 rows. Hal ini sama seperti sebelum kita memberikan tanda centang pada Remove rows.... Hal ini dikarenakan terdapat 2 kategori yang memiliki data duplikat yaitu FirstName dan LastName. Untuk itu, jika ingin melihat secara detail data pada FirstName maupun LastName. Kita lakukan satu per satu.
- 32) Pilih Stop Debugging.
- 33) Kita akan melakukan sort dengan memilih double click pada sort.
- 34) Pertama kita melakukan sort pada FirstName terlebih dahulu agar datanya terlihat lebih jelas dan tidak digabungkan dengan data redundan dari kolom LastName.
- 35) Kemudian jangan beri tanda centang pada Remove rows... untuk mengetahui jumlah data yang FirstNamenya duplikat.



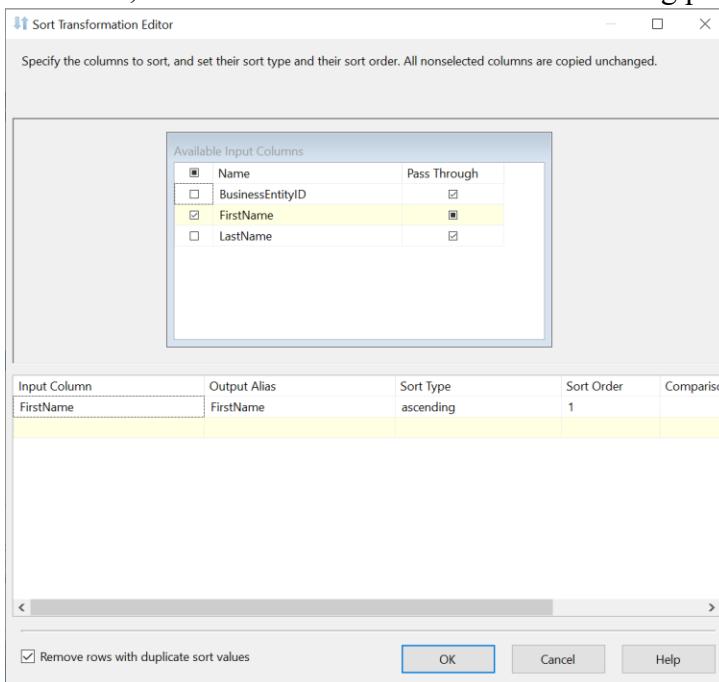
36) Eksekusi package tersebut, tekan F5.



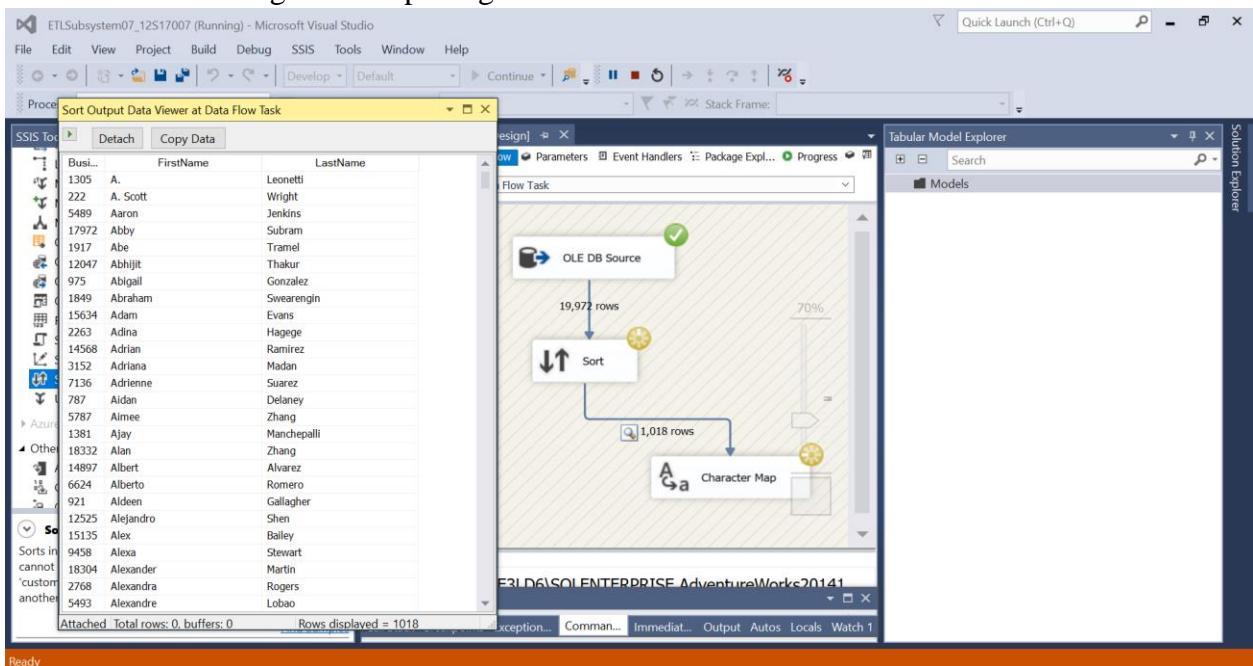
37) Pada gambar tersebut, kita dapat melihat data duplikat pada FirstName. Data input sebanyak 19,972 rows sementara data output sebanyak 19,972 rows. Hal ini menandakan bahwa data yang duplikat belum di-remove.

38) Kemudian pilih stop debugging.

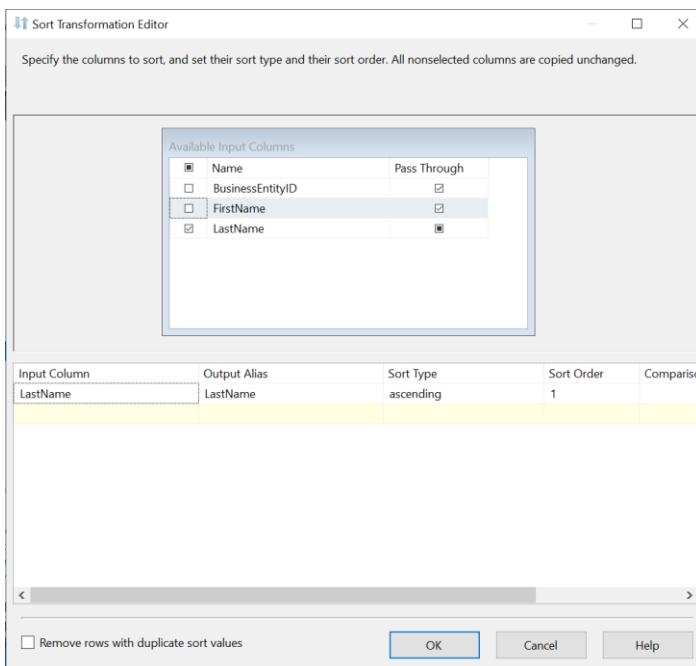
39) Pada Sort, double click kemudian beri tanda centang pada Remove rows...



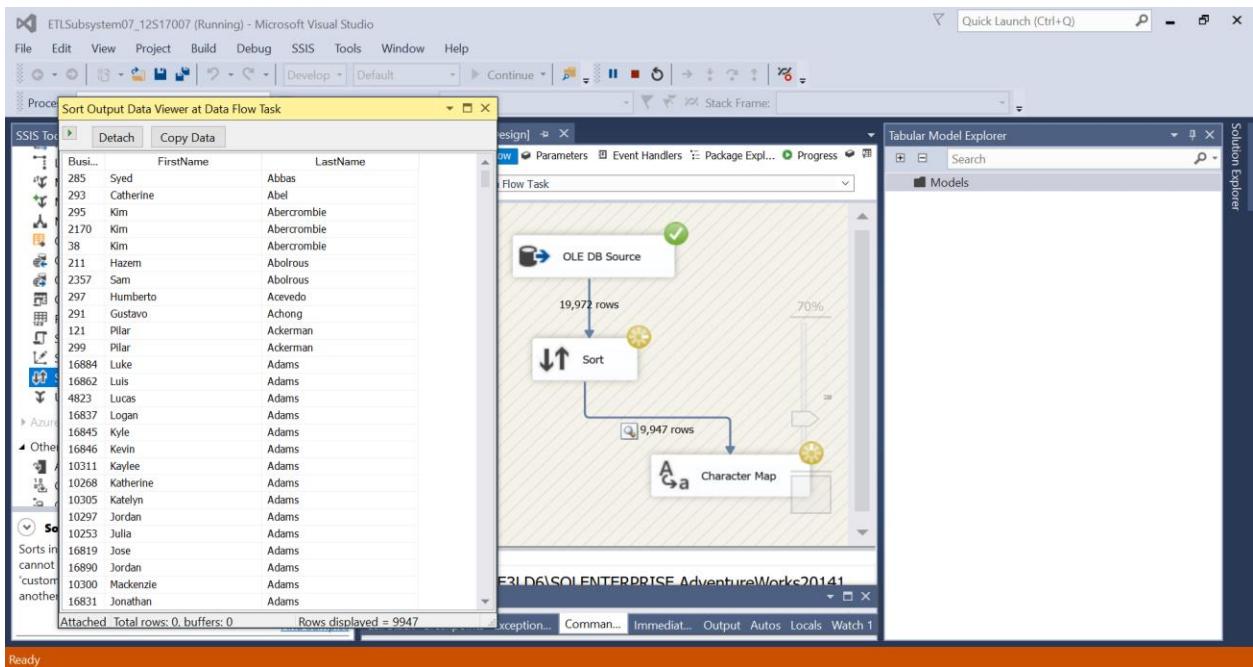
40) Tekan F5 untuk mengeksekusi package tersebut.



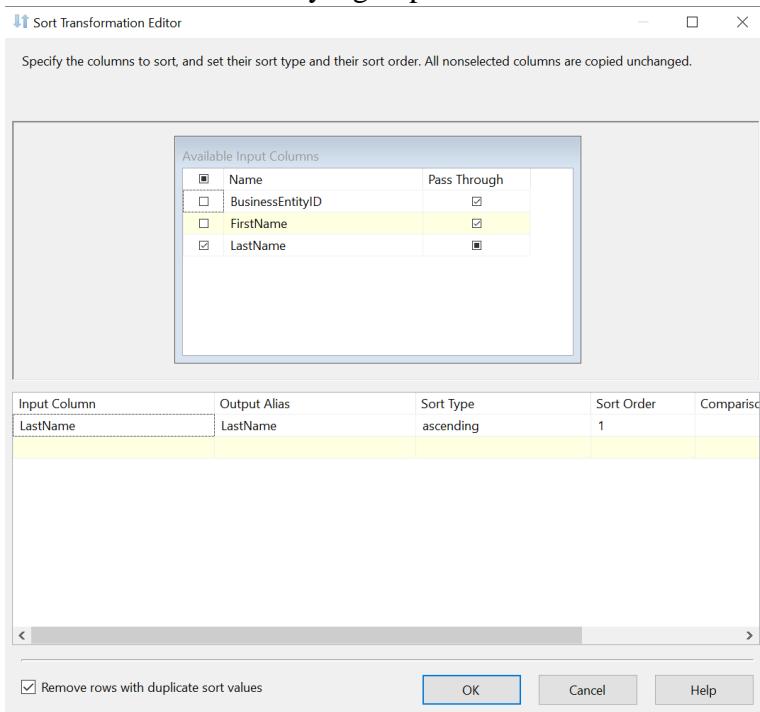
- 41) Pada data tersebut, data input sebanyak 19,972 dan ketika diberi tanda centang pada Remove rows... data output menjadi sebanyak 1,018 rows. Hal ini menandakan bahwa data yang duplikat tidak ditampilkan lagi atau telah di-remove.
- 42) Kemudian pilih Stop Debugging.
- 43) Kita akan melakukan remove duplication pada LastName.
- 44) Double click Sort. Kemudian pilih atau beri tanda centang pada LastName.
- 45) Jangan centang Remove Rows.. agar kita dapat melihat jumlah data sebelum data yang duplikat di-remove.



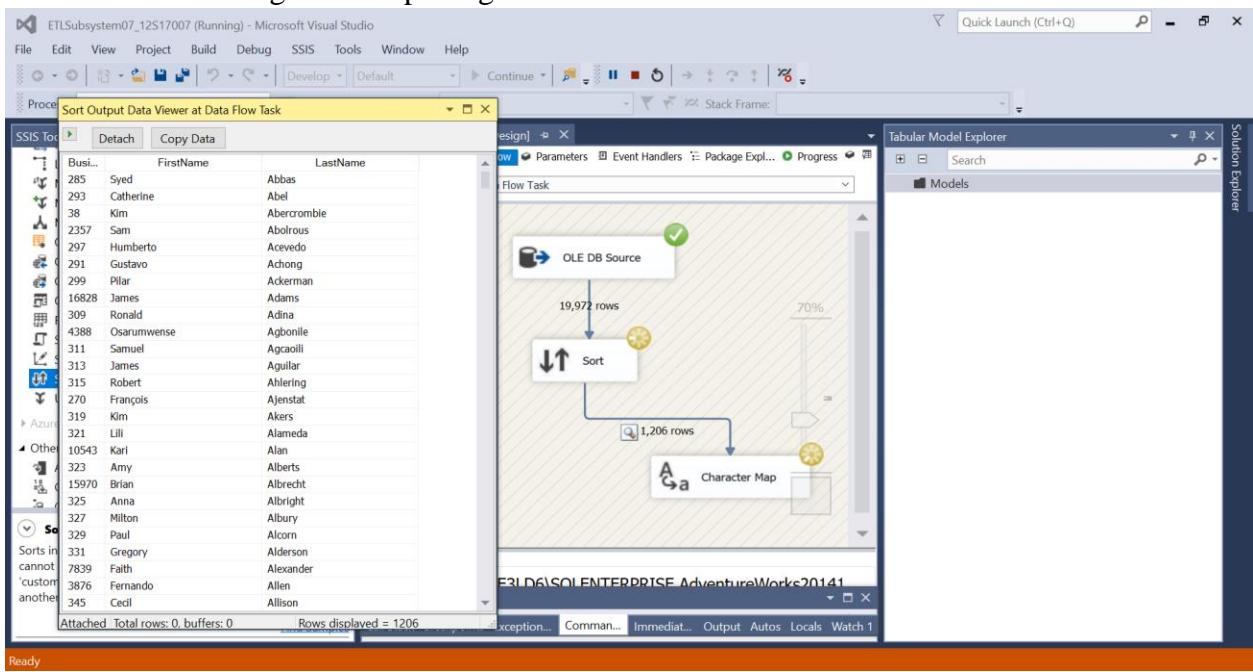
46) Tekan F5 untuk mengeksekusi package tersebut.



- 47) Pada gambar tersebut, kita dapat melihat data duplikat pada LastName. Data input sebanyak 19,972 rows sementara data output sebanyak 9,947 rows. Hal ini menandakan bahwa data yang duplikat pada LastName sebanyak 9,947 rows.
48) Pilih Stop Debugging.
49) Double click pada sort, kemudian berikan tanda centang pada Remove rows... untuk me-remove data LastName yang duplikat.



50) Tekan F5 untuk mengeksekusi package tersebut.

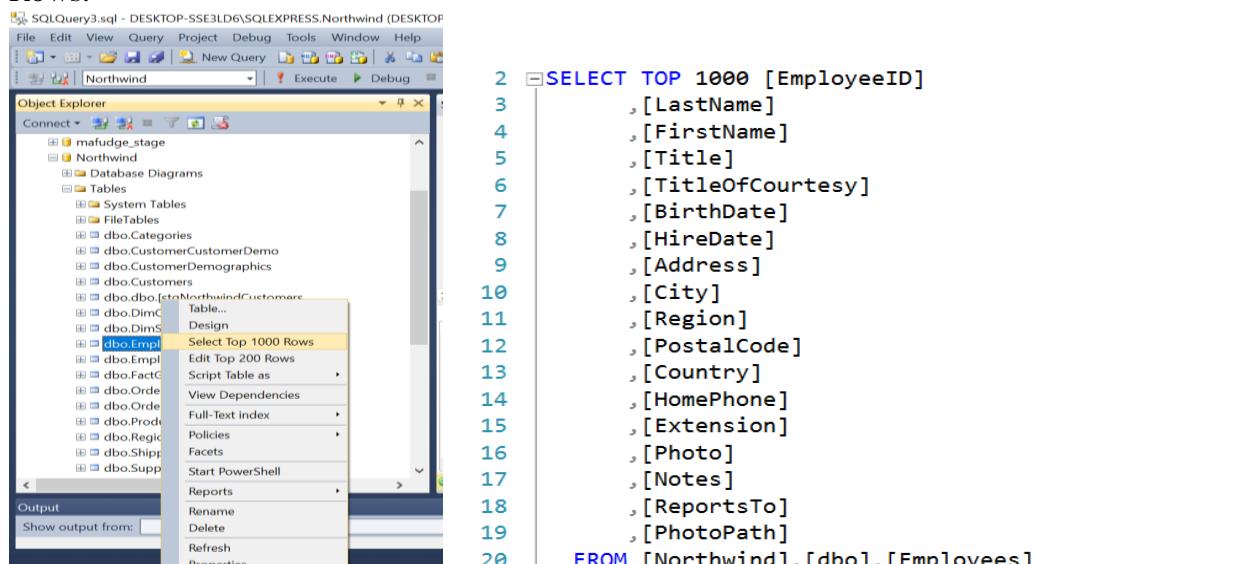


- 51) Berdasarkan gambar tersebut, dapat kita lihat bahwa data input sebanyak 19,972 rows, dan setelah dilakukan *remove* terhadap data yang duplikat, data output menjadi sebanyak 1,206 rows. Sehingga pada Data Viewer yang ditampilkan kita tidak lagi menemukan data yang duplikat.
- 52) Dengan SSIS kita dapat melihat data yang duplikat serta melakukan *remove* terhadap data yang duplikat agar selanjutnya data tersebut dapat kita gunakan.

E. Slowly Changing Dimension Type 2

Slowly Changing Dimension adalah teknik menyimpan dimensi dalam datawarehouse dalam hubungannya terhadap perubahan dimensi dan atribut data. Slowly Changing Dimension memiliki beberapa tipe, yaitu tipe 0, 1, 2, 3, dan 4. Pada laporan ini, kita akan membahas tentang Slowly Changing Dimension Type 2. Type 2 dapat menyimpan histori perubahan data dengan menambahkan row baru dengan natural key yang sama namun surrogate key yang berbeda. Pada laporan ini, kita akan mengerjakan Slowly Changing Dimension Type 2 menggunakan SQL Server Data Tools 2015 dan SQL Server 2014. Berikut langkah-langkahnya.

- 1) Pada bagian ini, kita akan menggunakan database Northwind dan tabel [dbo].[employee].
- 2) Untuk melihat isi dari tabel tersebut, klik kanan pada tabel kemudian pilih Select Top 1000 Rows.



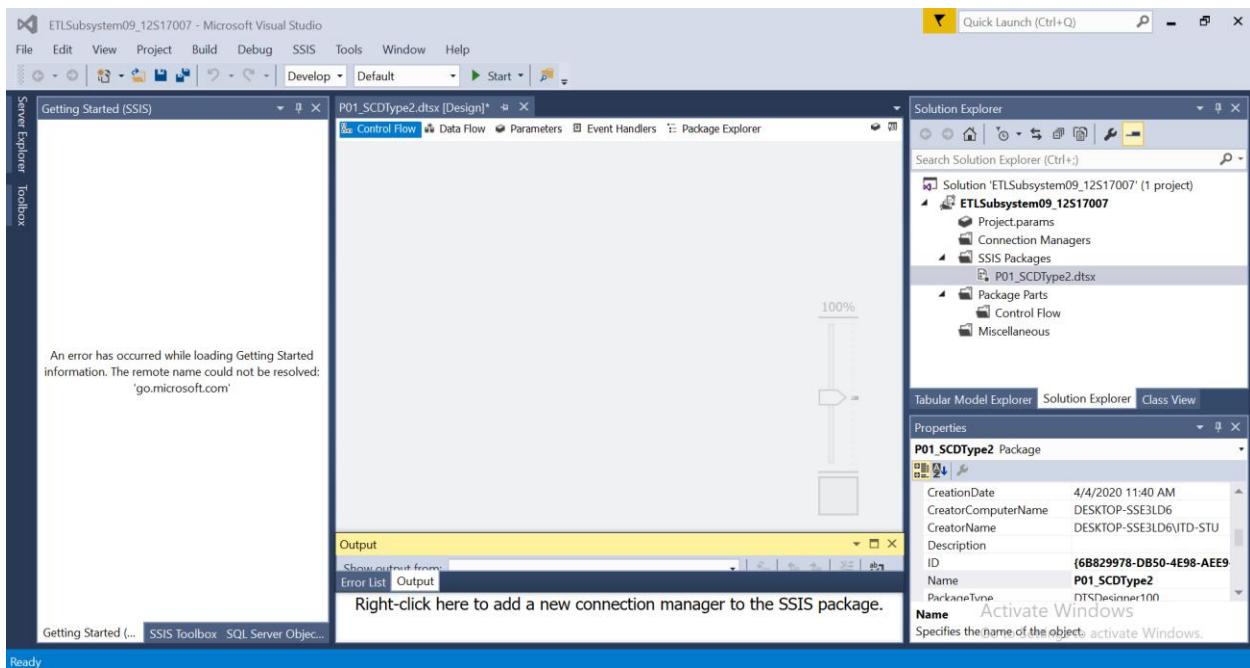
```

2  SELECT TOP 1000 [EmployeeID]
3      ,[LastName]
4      ,[FirstName]
5      ,[Title]
6      ,[TitleOfCourtesy]
7      ,[BirthDate]
8      ,[HireDate]
9      ,[Address]
10     ,[City]
11     ,[Region]
12     ,[PostalCode]
13     ,[Country]
14     ,[HomePhone]
15     ,[Extension]
16     ,[Photo]
17     ,[Notes]
18     ,[ReportsTo]
19     ,[PhotoPath]
20  FROM [Northwind].[dbo].[Employees]

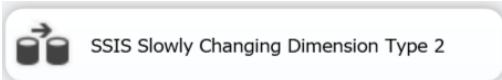
```

	EmployeeID	LastName	FirstName	Title	TitleOfCourtesy	BirthDate	HireDate	Address
1	1	Davolio	Nancy	Sales Representative	Ms.	1948-12-08 00:00:00.000	1992-05-01 00:00:00.000	507 - 20th Ave. E. Apt
2	2	Fuller	Andrew	Vice President, Sales	Dr.	1952-02-19 00:00:00.000	1992-08-14 00:00:00.000	908 W. Capital Way
3	3	Leverling	Janet	Sales Representative	Ms.	1963-08-30 00:00:00.000	1992-04-01 00:00:00.000	722 Moss Bay Blvd.
4	4	Peacock	Margaret	Sales Representative	Mrs.	1937-09-19 00:00:00.000	1993-05-03 00:00:00.000	4110 Old Redmond Rd
5	5	Buchanan	Steven	Sales Manager	Mr.	1955-03-04 00:00:00.000	1993-10-17 00:00:00.000	14 Garrett Hill
6	6	Suyama	Michael	Sales Representative	Mr.	1963-07-02 00:00:00.000	1993-10-17 00:00:00.000	Coventry House Miner
7	7	King	Robert	Sales Representative	Mr.	1960-05-29 00:00:00.000	1994-01-02 00:00:00.000	Edgeham Hollow Wind
8	8	Callahan	Laura	Inside Sales Coordinator	Ms.	1958-01-09 00:00:00.000	1994-03-05 00:00:00.000	4726 - 11th Ave. N.E.
9	9	Dodsworth	Anne	Sales Representative	Ms.	1966-01-27 00:00:00.000	1994-11-15 00:00:00.000	7 Houndstooth Rd.

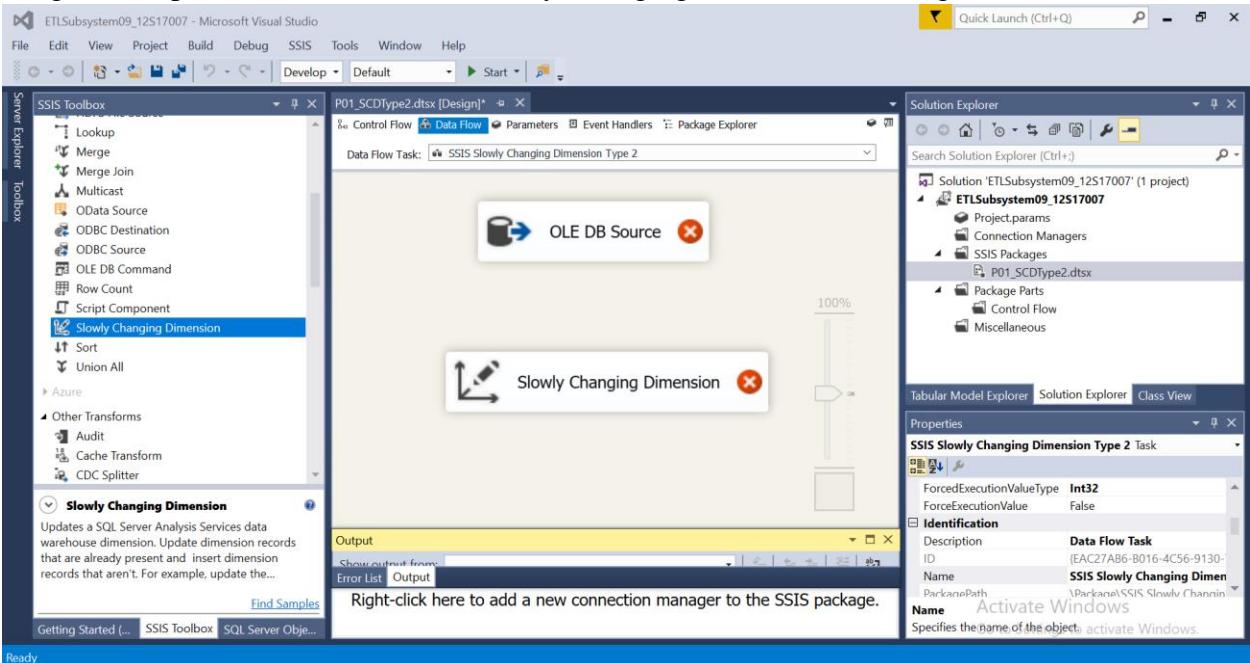
- 3) Jika kita amati, pada gambar di atas terdapat beberapa Title untuk setiap Employee. Misalnya FirstName dari seorang employee berubah. (Hal ini hanya contoh, meskipun sebenarnya tidak mungkin FirstName berubah). Dengan Slowly Changing Dimension kita akan menambahkan baris yang baru untuk menyimpan perubahan terhadap data yang lama.
- 4) Buatlah sebuah proyek yang baru dan berikan nama ETLSubSytem09_12S17007.
- 5) Berikan sebuah nama dan lokasi untuk proyek SSIS tersebut.
- 6) Pilih Package yang telah dibuat secara *default* dan *rename* menjadi P01_SCDType2.dtsx.



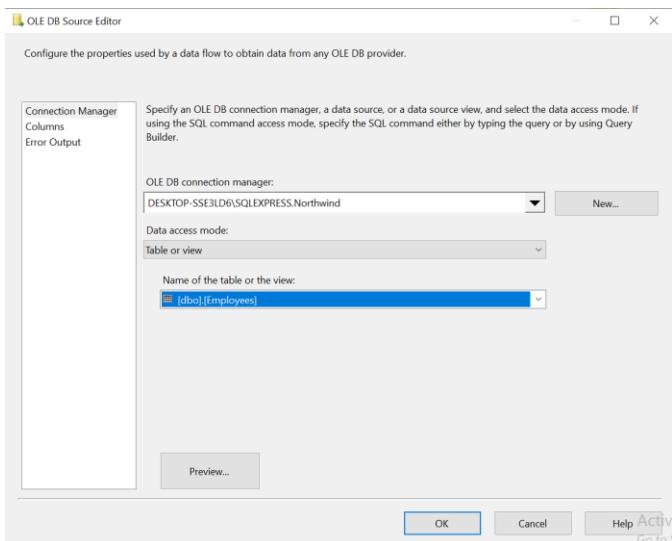
- 7) Dari SSIS Toolbox, drag and drop Data Flow Task ke design area Control Flow. Kemudian rename menjadi SSIS Slowly Changing Dimension Type 2.



- 8) Double click pada SSIS Slowly Changing Dimension Type 2.
9) Drag and drop OLE DB Source dan Slowly Changing Dimension ke design area Data Flow.



- 10) Double click pada OLE DB Source dan lakukan konfigurasi terhadap Connection Manager. Jika Connection sebelumnya belum ada, pilih New untuk membuat connection sesuai dengan device dan database yang kita gunakan.



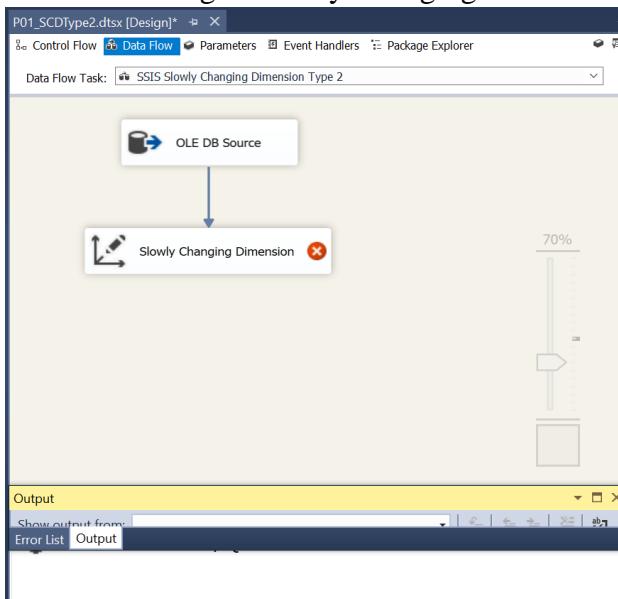
11) Untuk melihat isi tabel [dbo].[Customer] pilih Preview.

Preview Query Results									
Query result (up to the first 200 rows):									
1	Davolio	Nancy	Sales ...	Ms.	12/8/...	5/1/1...	507 - ...	Seattle	
2	Fuller	Andrew	Vice P...	Dr.	2/19/...	8/14/...	908 ...	Tacoma	
3	Leverl...	Janet	Sales ...	Ms.	8/30/...	4/1/1...	722 M...	Kirkland	
4	Peacock	Marga...	Sales ...	Mrs.	9/19/...	5/3/1...	4110 ...	Redm...	
5	Bucha...	Steven	Sales ...	Mr.	3/4/1...	10/17...	14 Ga...	London	
6	Suyama	Michael	Sales ...	Mr.	7/2/1...	10/17...	Coven...	London	
7	King	Robert	Sales ...	Mr.	5/29/...	1/2/1...	Edgeh...	London	
8	Callah...	Laura	Inside...	Ms.	1/9/1...	3/5/1...	4726 ...	Seattle	
9	Dods...	Anne	Sales ...	Ms.	1/27/...	11/15/...	7 Hou...	London	

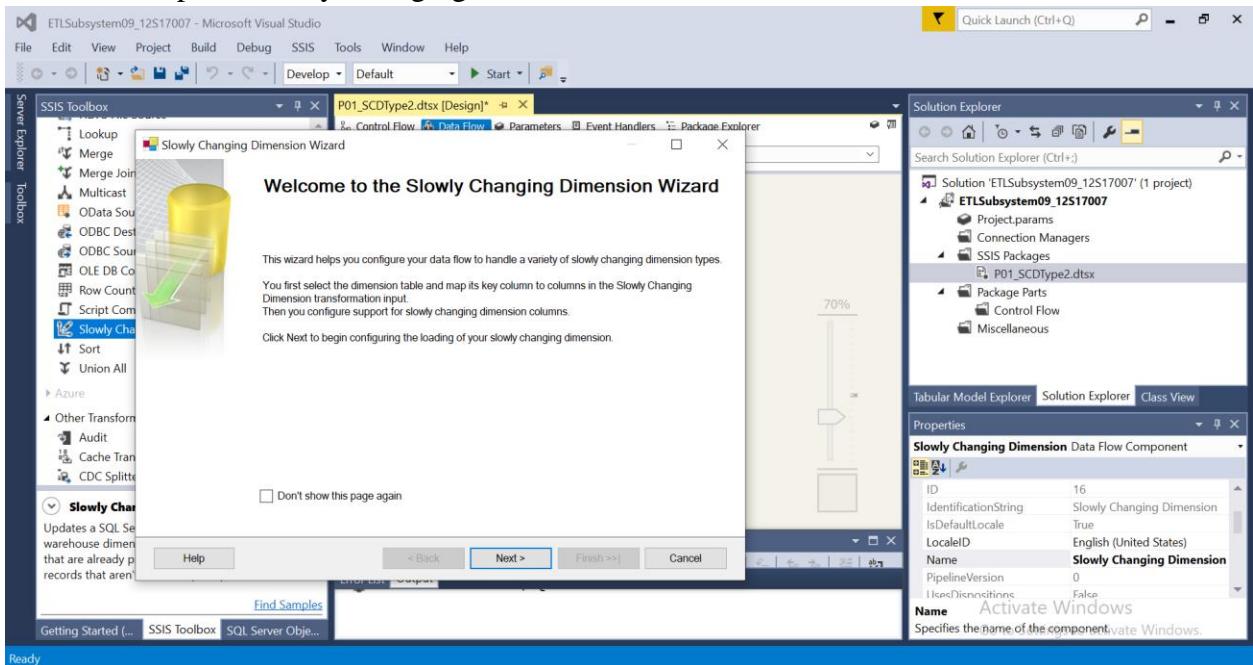
12) Pilih Columns untuk melihat kolom yang terdapat pada tabel [dbo].[employee].

External Column	Output Column
EmployeeID	EmployeeID
LastName	LastName
FirstName	FirstName
Title	Title
TitleOfCourtesy	TitleOfCourtesy
BirthDate	BirthDate
HireDate	HireDate
Address	Address
City	City
Region	Region

13) Karena kita akan menggunakan data yang disimpan dalam OLE DB Source, hubungkan OLE DB Source dengan Slowly Changing Dimension.

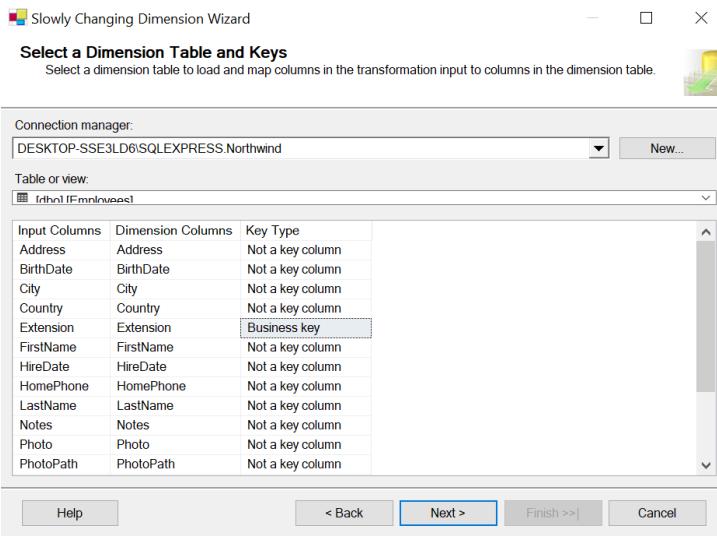


14) Double click pada Slowly Changing Dimension



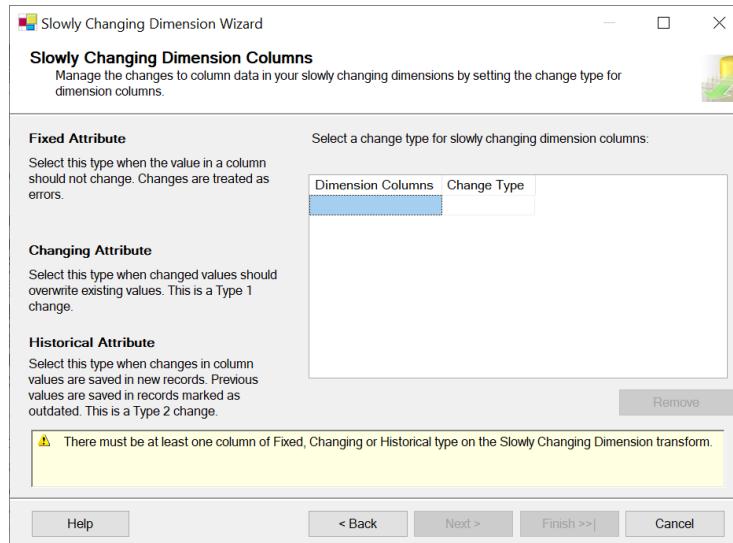
15) Kemudian pilih Next.

16) Pilih tabel yang digunakan dan ubah Key Type Extension menjadi Business Key.

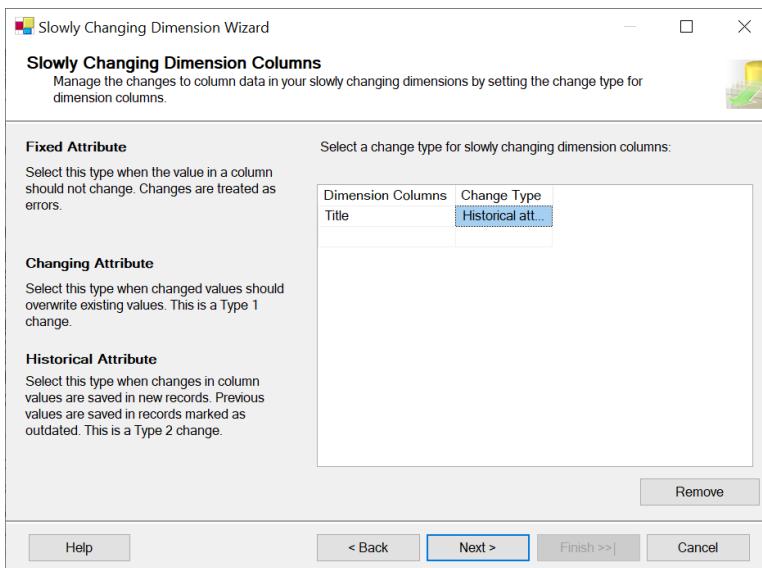


- Input Columns** : Kolom yang berasal dari OLE DB Source.
Dimension Columns : Kolom yang tersedia pada tabel [dbo].[Employee]
Key Type : Untuk menunjukkan Slowly Changing Dimension Type 2, kita memerlukan paling tidak satu Business Key. Pada gambar di atas kita menggunakan Extension sebagai Business Key dari tabel [dbo].[Customer].

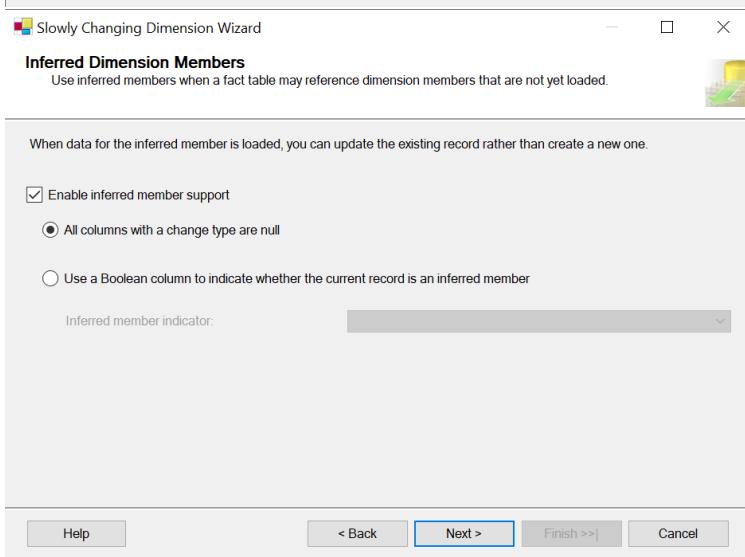
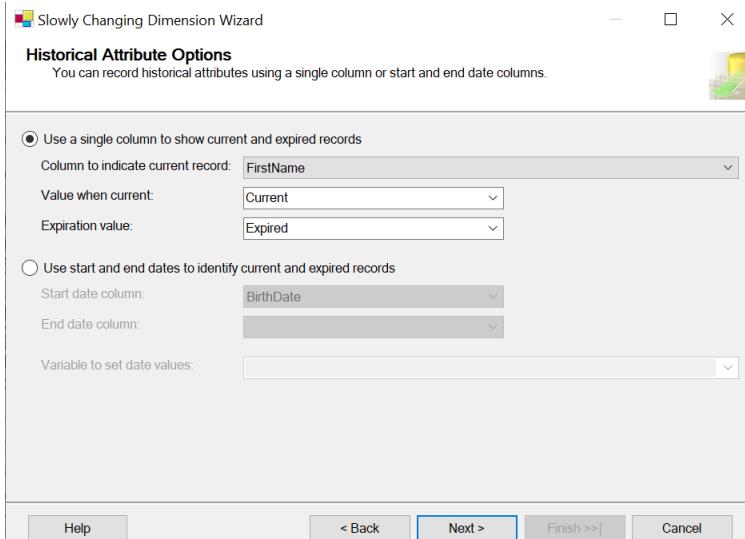
17) Pilih Next.

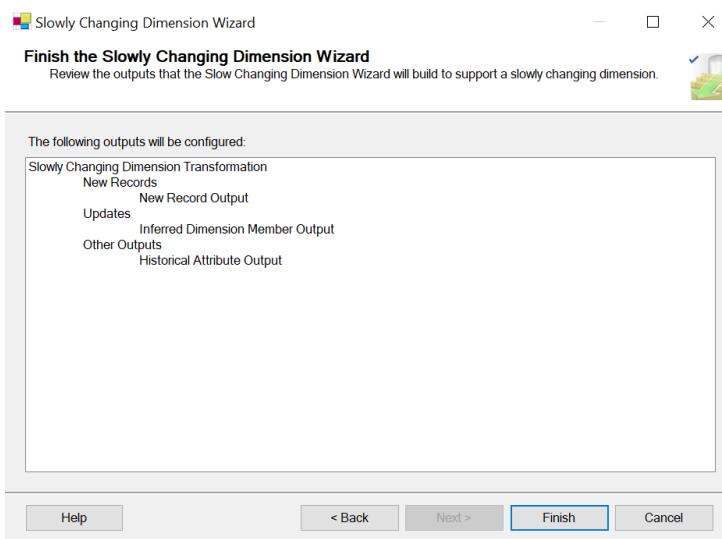


- Dimension Columns** : Semua kolom yang tersedia pada tabel dimensi. Pada bagian ini kita memilih kolom yang ingin dipilih untuk Change Type. Pada bagian ini kita pilih Title karena kita ingin mengubah Title dari Employee.
Change Type : Pilih Change Type. Change Type yang kita pilih adalah Historical attribute.

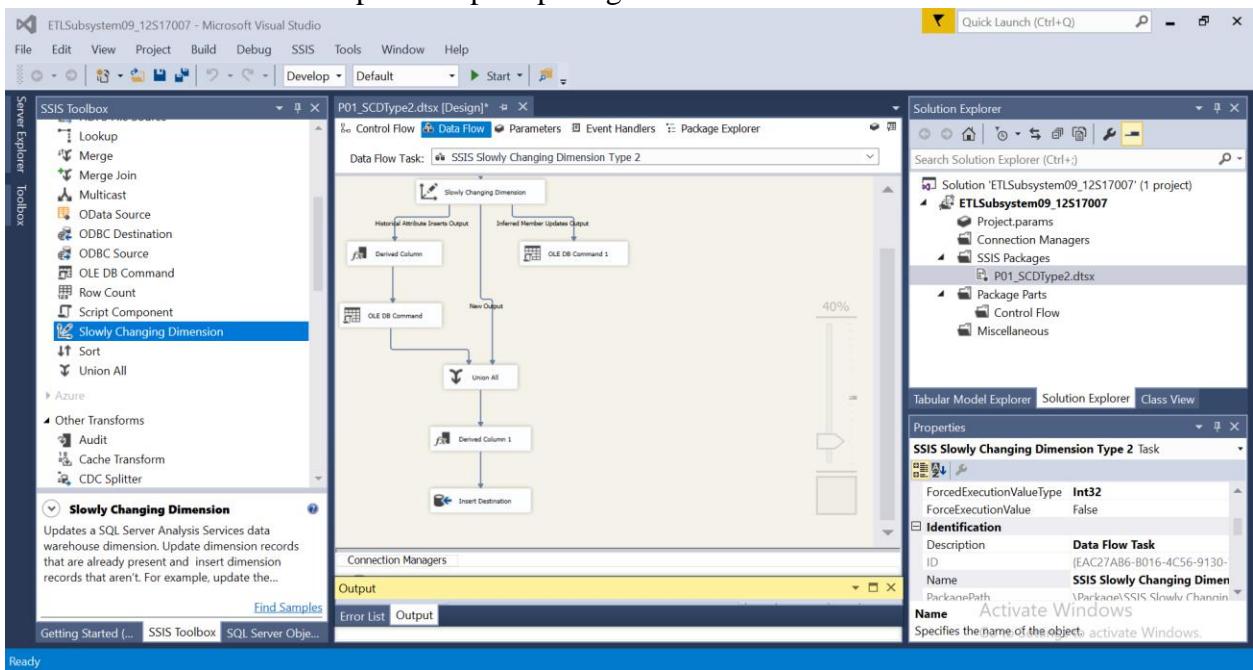


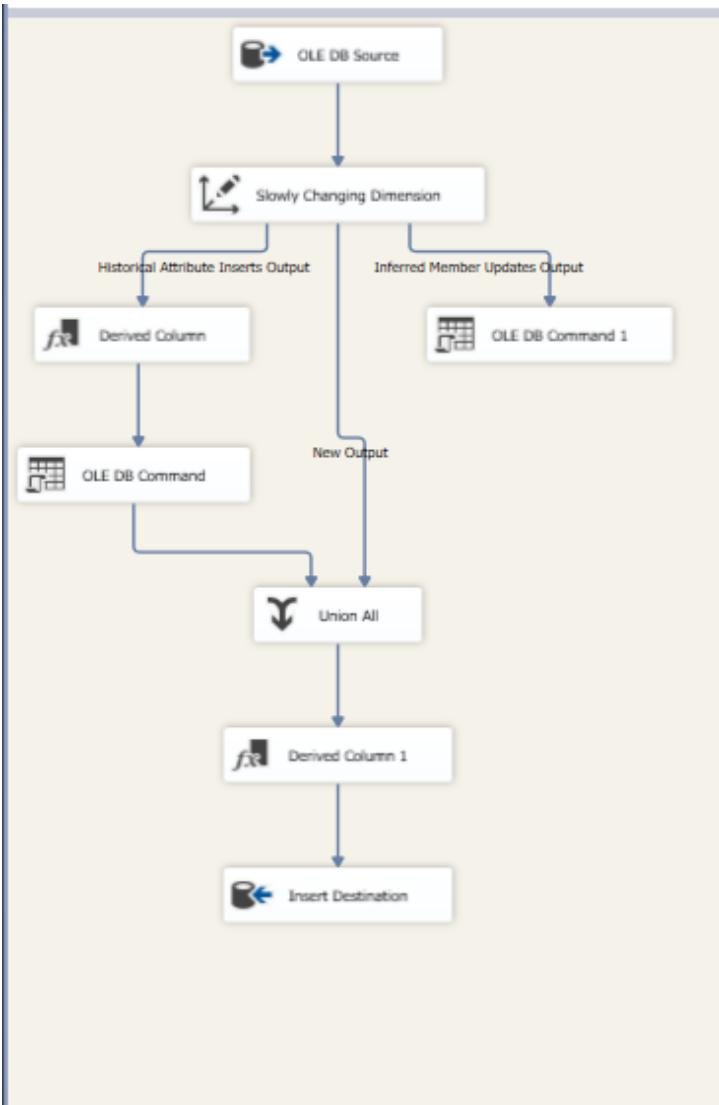
18) Pilih Next dan lengkapi isian selanjutnya.



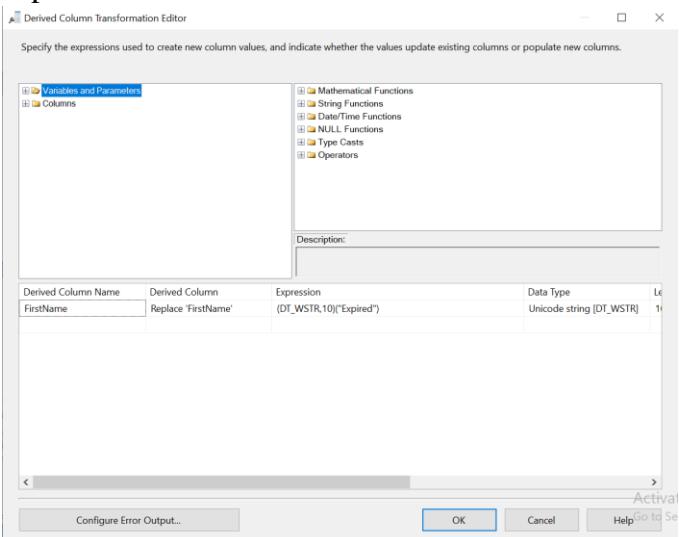


19) Pilih Finish dan akan ditampilkan seperti pada gambar berikut.

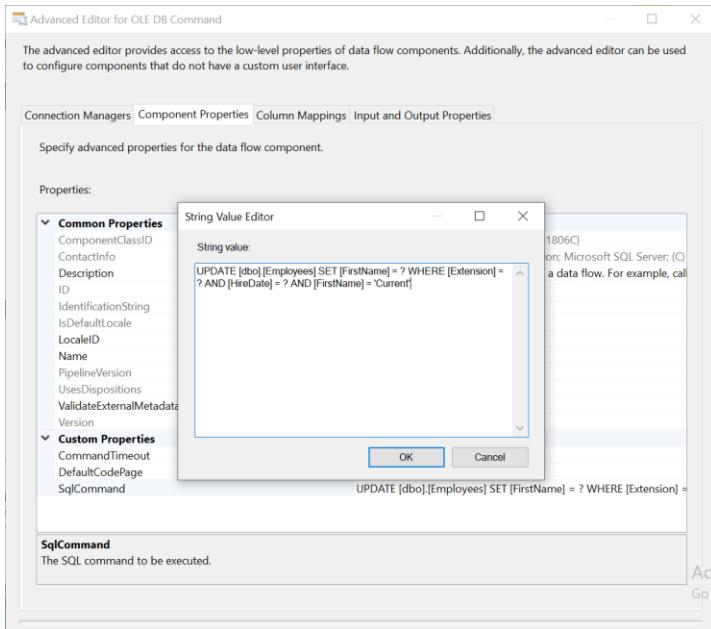




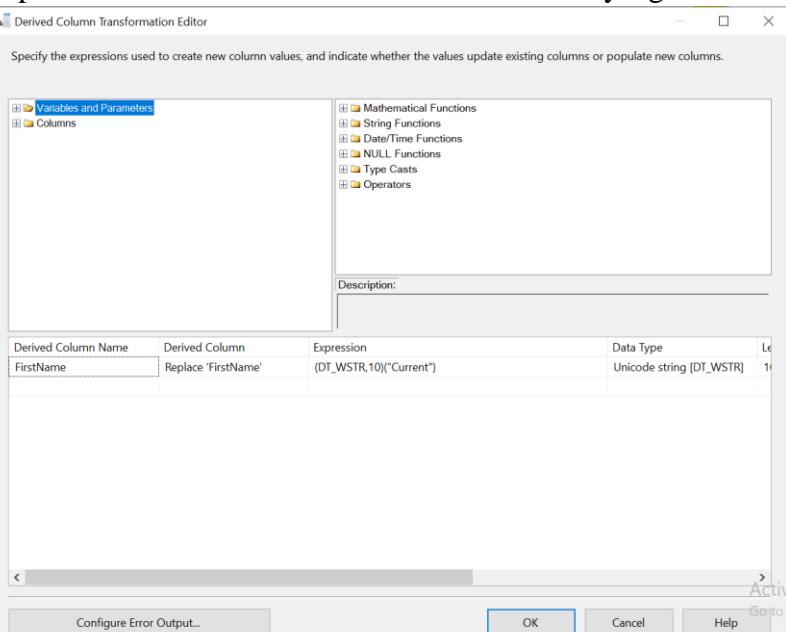
- 20) Kemudian double click pada Derived Column untuk mengecek automatically generated expression to set FirstName.



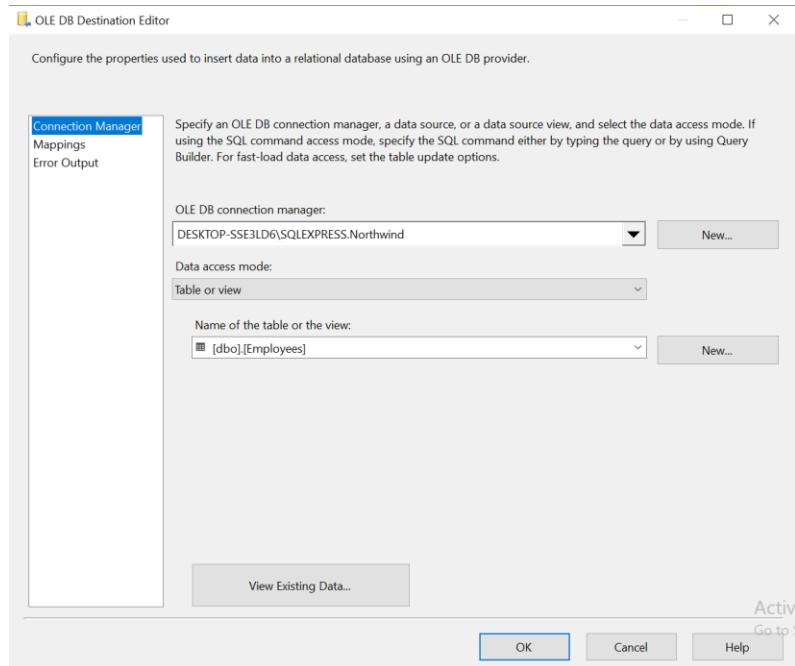
- 21) Double click pada OLE DB Command Transformation untuk mencek Update Statement yang dibuat secara otomatis.



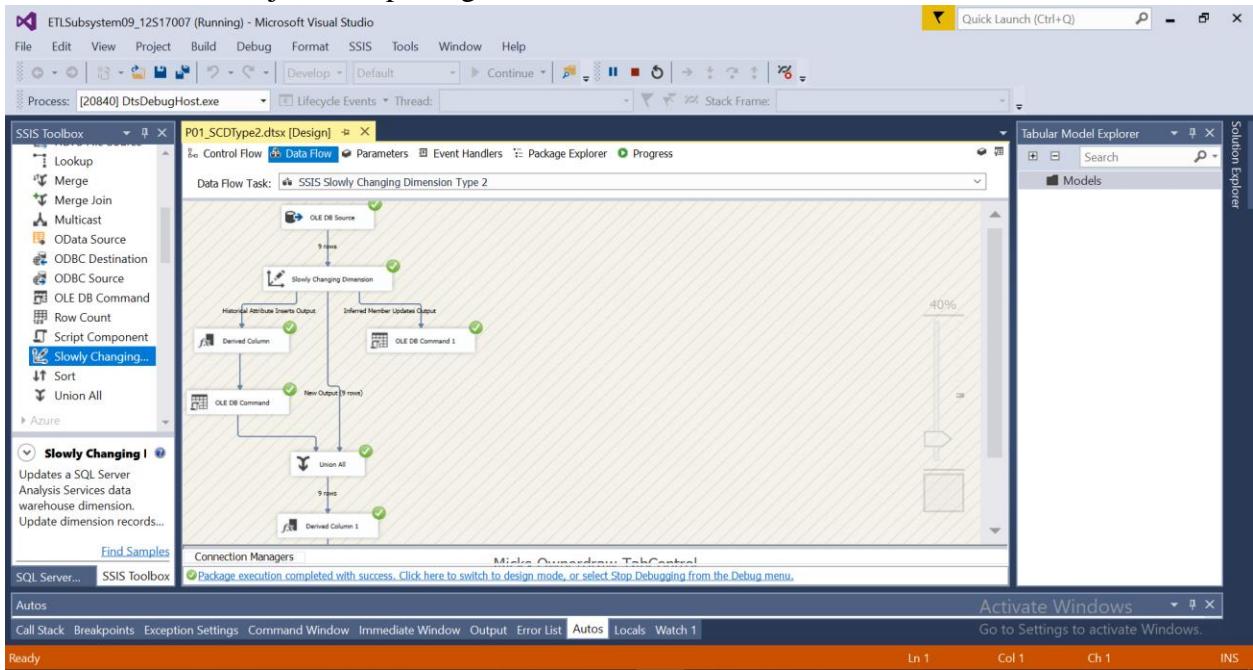
- 22) Double click pada Derived Column Transformation untuk mengecek automatically generated expression untuk menset FirstName dari record yang baru.

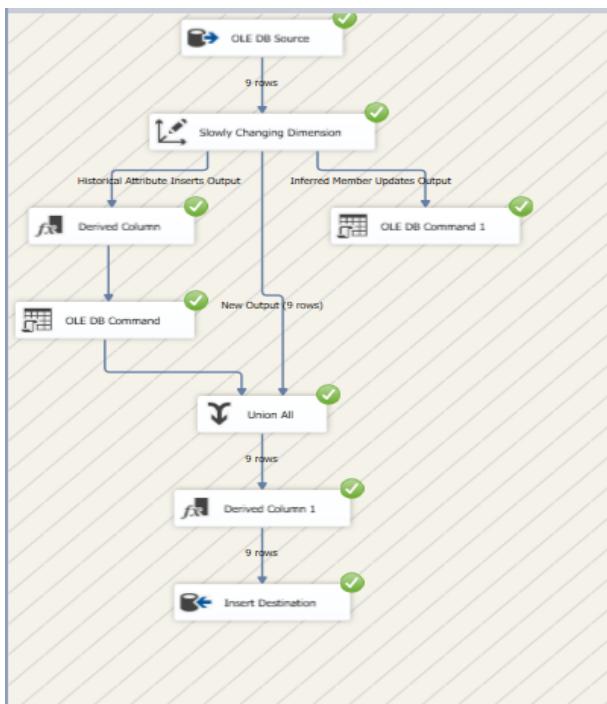


- 23) Double click pada Insert Destination dan cek tabel yang secara otomatis dibuat.



24) Tekan F5 untuk menjalankan package tersebut.





- 25) Dari gambar tersebut terlihat bahwa terdapat 9 rows yang dimiliki dari OLE DB Source. 9 rows tersebut dimasukkan secara langsung dan me-maintain historical information kemudian memasukkannya.
- 26) Buka SQL Server 2014 dan cek hasilnya.
- 27) Eksekusi Query berikut untuk melihat hasilnya.

22 | `SELECT * FROM [Northwind].[dbo].[Employees]`

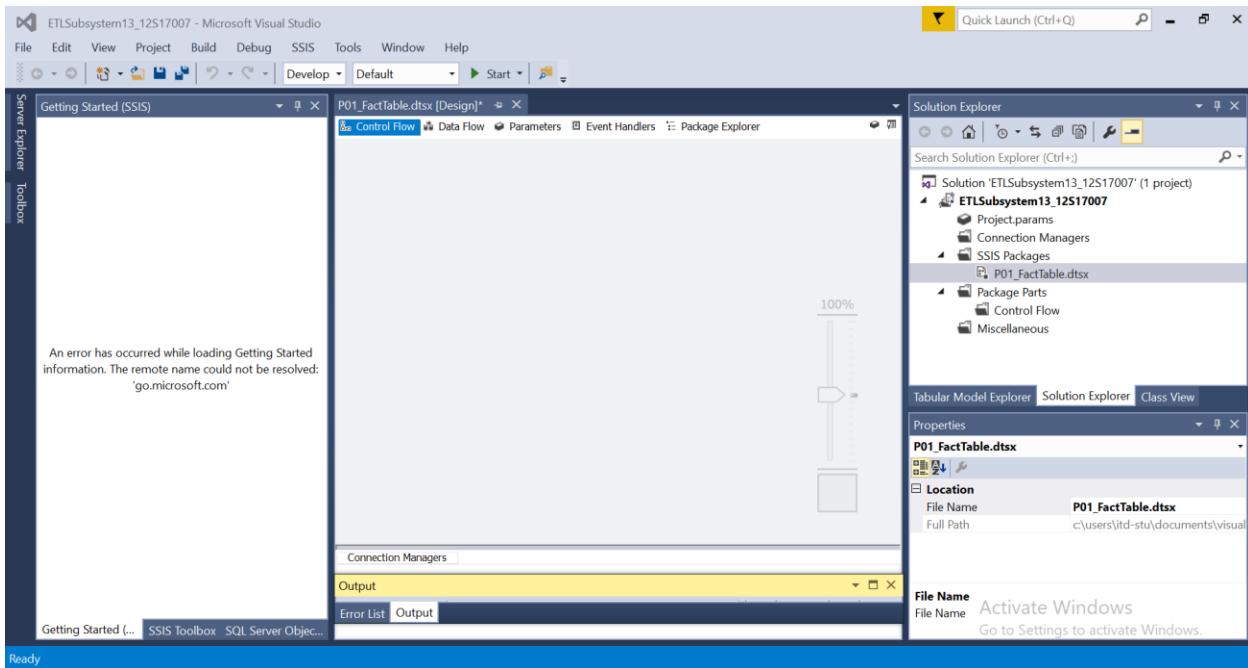
	EmployeeID	LastName	FirstName	Title	TitleOfCourtesy	BirthDate	HireDate	Address
1	1	Davolio	Nancy	Sales Representative	Ms.	1948-12-08 00:00:00.000	1992-05-01 00:00:00.000	507 - 20th Ave. E. Apt
2	2	Fuller	Andrew	Vice President, Sales	Dr.	1952-02-19 00:00:00.000	1992-08-14 00:00:00.000	908 W. Capital Way
3	3	Leverling	Janet	Sales Representative	Ms.	1963-08-30 00:00:00.000	1992-04-01 00:00:00.000	722 Moss Bay Blvd.
4	4	Peacock	Margaret	Sales Representative	Mrs.	1937-09-19 00:00:00.000	1993-05-03 00:00:00.000	4110 Old Redmond Rd
5	5	Buchanan	Steven	Sales Manager	Mr.	1955-03-04 00:00:00.000	1993-10-17 00:00:00.000	14 Garrett Hill
6	6	Suyama	Michael	Sales Representative	Mr.	1963-07-02 00:00:00.000	1993-10-17 00:00:00.000	Coventry House Miner
7	7	King	Robert	Sales Representative	Mr.	1960-05-29 00:00:00.000	1994-01-02 00:00:00.000	Edgeham Hollow Wind
8	8	Callahan	Laura	Inside Sales Coordinator	Ms.	1958-01-09 00:00:00.000	1994-03-05 00:00:00.000	4726 - 11th Ave. N.E.
9	9	Dodsworth	Anne	Sales Representative	Ms.	1966-01-27 00:00:00.000	1994-11-15 00:00:00.000	7 Houndstooth Rd.
10	10	Davolio	Current	Sales Representative	Ms.	1948-12-08 00:00:00.000	1992-05-01 00:00:00.000	507 - 20th Ave. E. Apt
11	11	Fuller	Current	Vice President, Sales	Dr.	1952-02-19 00:00:00.000	1992-08-14 00:00:00.000	908 W. Capital Way
12	12	Leverling	Current	Sales Representative	Ms.	1963-08-30 00:00:00.000	1992-04-01 00:00:00.000	722 Moss Bay Blvd.
13	13	Peacock	Current	Sales Representative	Mrs.	1937-09-19 00:00:00.000	1993-05-03 00:00:00.000	4110 Old Redmond Rd
14	14	Buchanan	Current	Sales Manager	Mr.	1955-03-04 00:00:00.000	1993-10-17 00:00:00.000	14 Garrett Hill
15	15	Suyama	Current	Sales Representative	Mr.	1963-07-02 00:00:00.000	1993-10-17 00:00:00.000	Coventry House Miner
16	16	King	Current	Sales Representative	Mr.	1960-05-29 00:00:00.000	1994-01-02 00:00:00.000	Edgeham Hollow Wind
17	17	Callahan	Current	Inside Sales Coordinator	Ms.	1958-01-09 00:00:00.000	1994-03-05 00:00:00.000	4726 - 11th Ave. N.E.
18	18	Dodsworth	Current	Sales Representative	Ms.	1966-01-27 00:00:00.000	1994-11-15 00:00:00.000	7 Houndstooth Rd.

- 28) Dari gambar tersebut dapat kita lihat jumlah rows pada tabel [dbo].[Employees] bertambah. Hal ini menandakan bahwa Slowly Changing Dimension berhasil dibuat, yaitu slowly changing dimension yang dilakukan pada FirstName. Dengan kata lain, pada baris 10 hingga 18 pada gambar di atas, FirstNamenya berubah menjadi Current sementara data yang lain tetap hanya FirstNamenya yang berubah. Penerapan SCD Type 2 ini dilakukan dengan membuat baris yang baru dengan data yang sama dan hanya mengubah data yang dilakukan update selain itu datanya tetap.

F. Fact Table Builder

Fact table adalah tabel yang umumnya mengandung sesuatu yang dapat diukur (*measure*), seperti harga, jumlah barang, dll. Fact table juga merupakan jumpalan foreign key yang terdapat pada masing-masing dimension table. Fact table juga mengandung data yang bersifat historis.

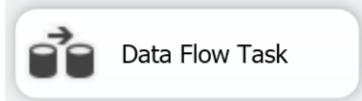
- 1) Buatlah sebuah proyek yang baru dan berikan nama ETLSubSytem13_12S17007.
- 2) Berikan sebuah nama dan lokasi untuk proyek SSIS tersebut.
- 3) Pilih Package yang telah dibuat secara *default* dan *rename* menjadi P01_FactTable.dtsx.



- 4) Pada Connection Manager, klik kanan kemudian pilih New OLE DB Connection.
- 5) Pilih connection dari device dan database yang akan digunakan yaitu AdventureWorks2014.



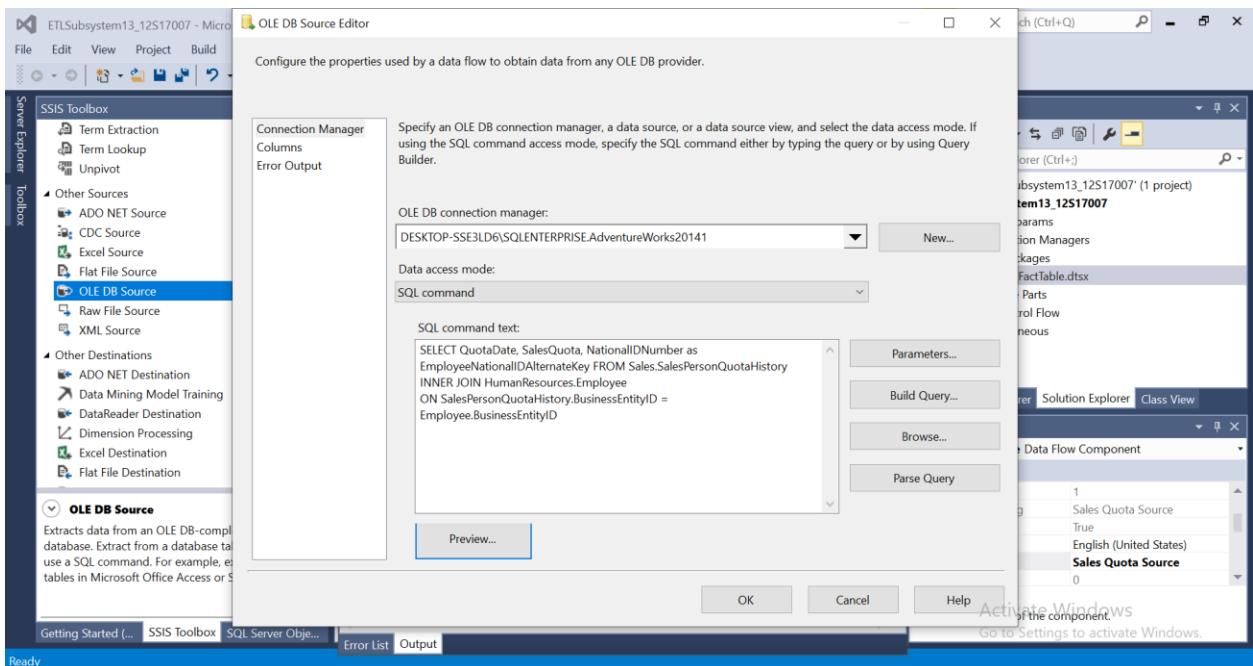
- 6) Dari SSIS Toolbox, drag and drop Data Flow Task ke design area Control Flow.



- 7) Double click pada Data Flow Task tersebut.
- 8) Drag and drop OLE DB Source ke design area Data Flow dan *rename* menjadi Sales Quota Source.



- 9) Lakukan konfigurasi pada Sales Quota Source dengan double click pada Sales Quota Source dan sesuaikan dengan connection dari device dan database yang digunakan.
- 10) Pilih Data Access Mode: SQL Command dan isi SQL Command text dengan kueri berikut.



11) Pilih Preview untuk melihat hasil eksekusi dari tabel tersebut.

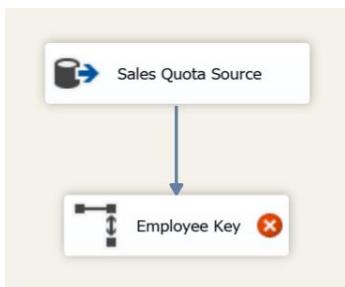
Preview Query Results			
Query result (up to the first 200 rows):			
QuotaDate	SalesQuota	EmployeeNationalIDAlternateKey	
5/31/2011 12:00:00 AM	28000	502097814	
8/31/2011 12:00:00 AM	7000	502097814	
12/1/2011 12:00:00 AM	91000	502097814	
2/29/2012 12:00:00 AM	140000	502097814	
5/30/2012 12:00:00 AM	70000	502097814	
8/30/2012 12:00:00 AM	154000	502097814	
11/30/2012 12:00:00 AM	107000	502097814	
2/28/2013 12:00:00 AM	58000	502097814	
5/30/2013 12:00:00 AM	263000	502097814	
8/30/2013 12:00:00 AM	116000	502097814	
11/30/2013 12:00:00 AM	84000	502097814	
3/1/2014 12:00:00 AM	187000	502097814	
5/31/2011 12:00:00 AM	367000	841560125	
8/31/2011 12:00:00 AM	556000	841560125	
12/1/2011 12:00:00 AM	502000	841560125	
2/29/2012 12:00:00 AM	550000	841560125	
5/30/2012 12:00:00 AM	1429000	841560125	
8/30/2012 12:00:00 AM	1324000	841560125	

12) Kemudian klik Parse Query dan Pilih OK.

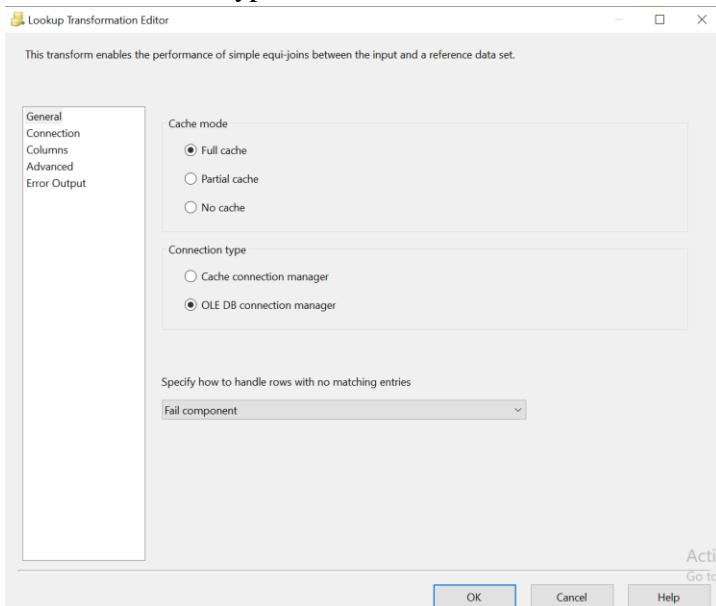
- 13) Untuk memperoleh Surrogate Key dari Dimension Tables, kita akan menggunakan Lookup Transformation. Drag and drop Lookup ke design area Data Flow.
- 14) Hubungkan OLE DB Source yaitu Sales Quota Source ke Lookup.
- 15) Kemudian rename Lookup menjadi Employee Key.



menjadi

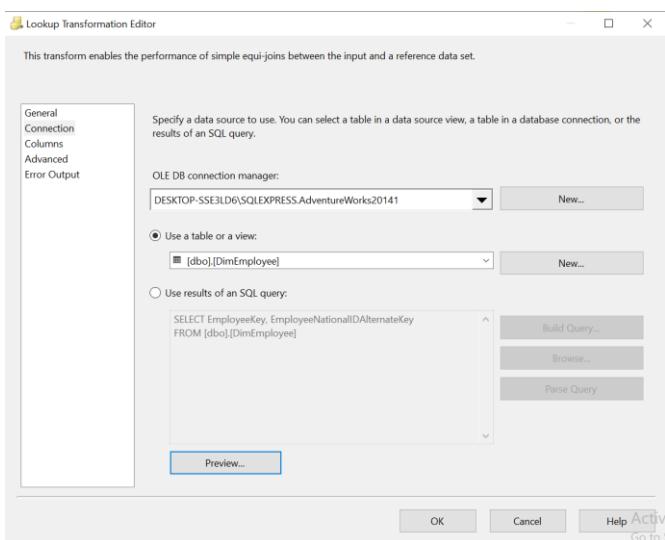


- 16) Double click pada Employee Key. Pada General Properti pilih modanya menjadi Full Cache dan Connection Type set ke OLE DB Connection manager.



- 17) Kemudian pilih Connection dan pilih Connection Manager sesuai device dan database yang digunakan.

- 18) Pilih Use atable or view.



- 19) Pilih Preview untuk melihat hasil dari kueri tersebut.

Preview Query Results

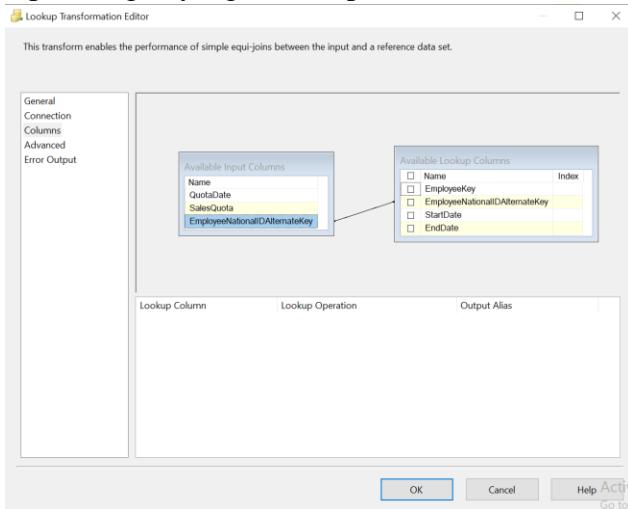
Query result (up to the first 200 rows):

EmployeeKey	EmployeeNational...	StartDate	EndDate
1	502097814	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
2	841560125	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
3	191644724	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
4	615389812	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
5	234474252	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
6	716374314	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
7	61161660	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM
8	139397894	1/1/1900 12:00:00 AM	1/1/1900 12:00:00 AM

< >

Close

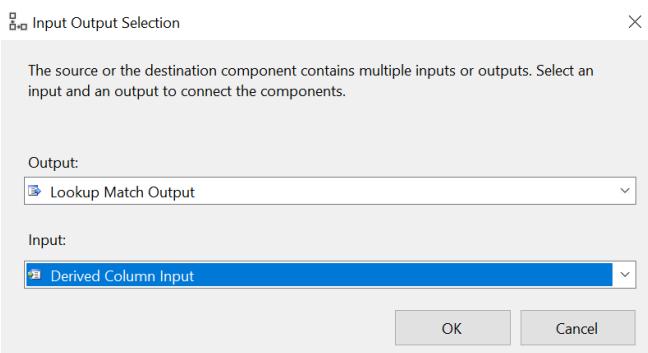
- 20) Kemudian pada bagian Column, lakukan mapping pada EmployeeNationalIDAlternate pada input dengan yang dilookup.



- 21) Dari SSIS Toolbox pilih Derived Column kemudian drag and drop ke design area Data Flow. Rename menjadi Date Keys.

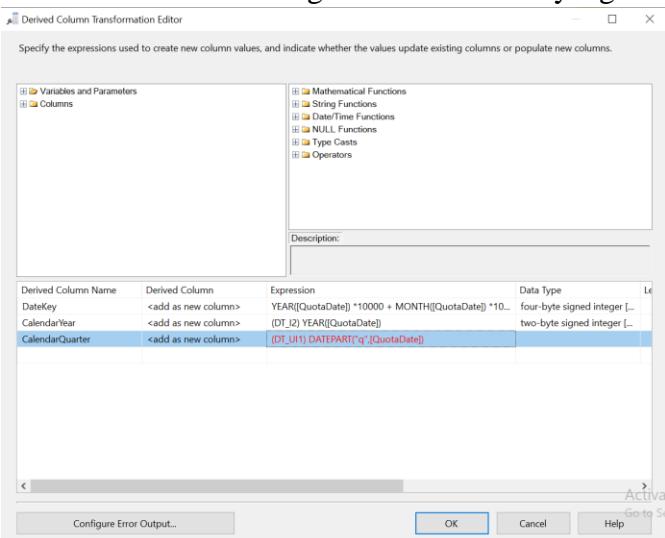


- 23) Lengkapi data berikut.

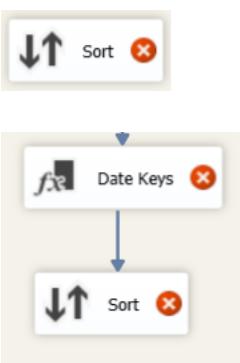


24) Double click pada Derived Column.

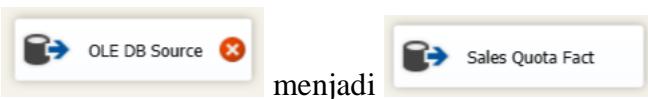
25) Kemudian tambahkan tiga derived column yang baru.



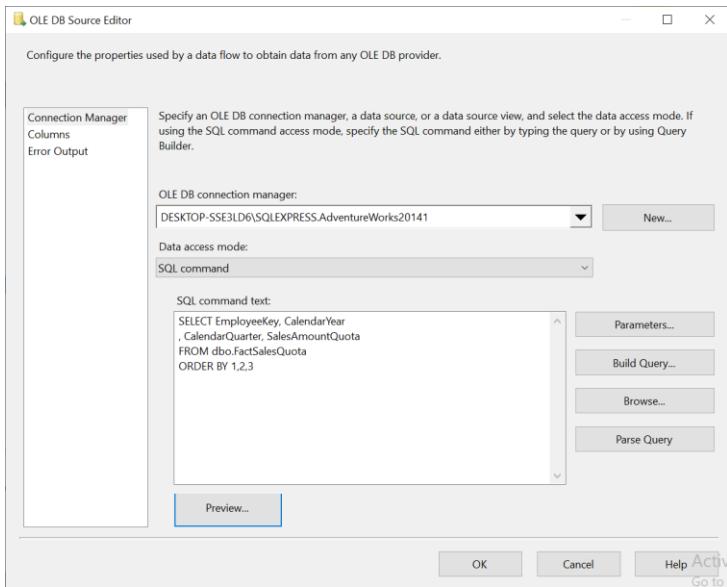
26) Kemudian dari SSIS Toolbox drag and drop Sort ke design area Data Flow. Hubungkan Derived Column dengan Sort.



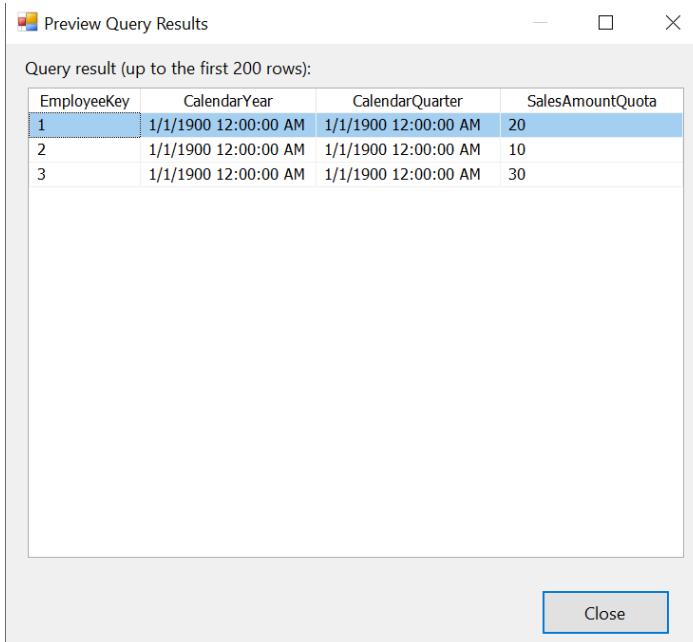
27) Drag and drop OLE DB Source ke design area Data Flow *rename* menjadi Sales Quota Fact.



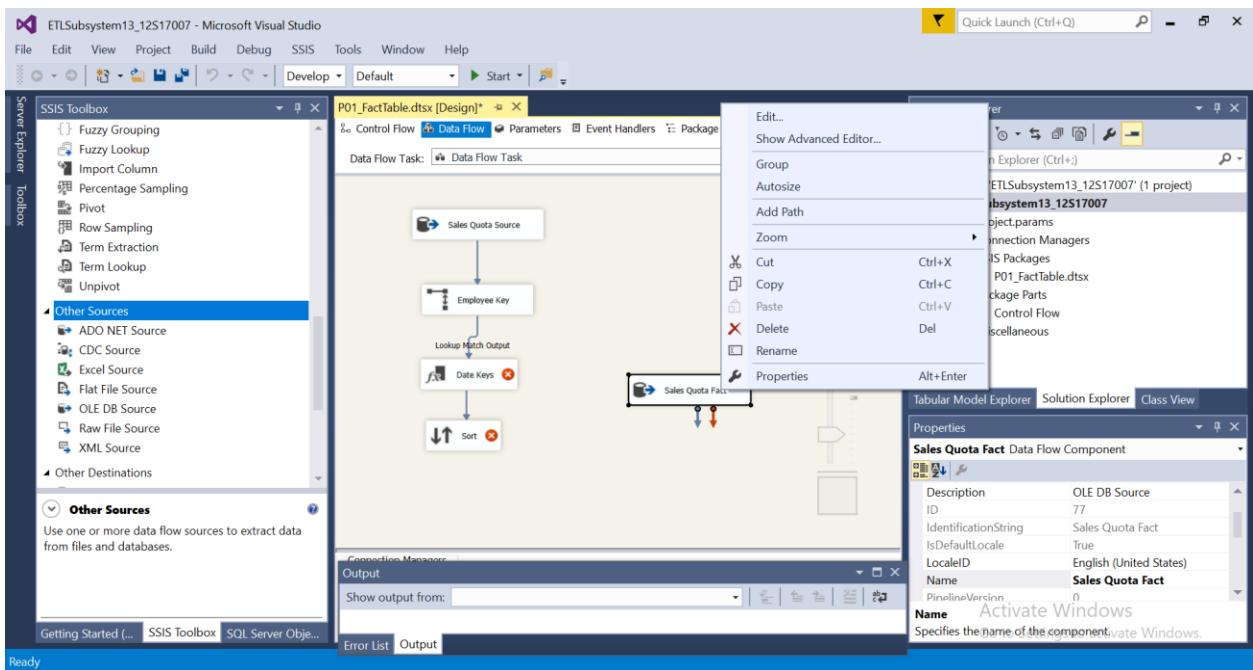
- 28) Double click pada OLE DB Source kemudian lakukan konfigurasi sesuai dengan connection dari device yang kita miliki.
- 29) Pilih SQL Data Access: SQL Command.
- 30) Berikan query berikut pada SQL Command Text.



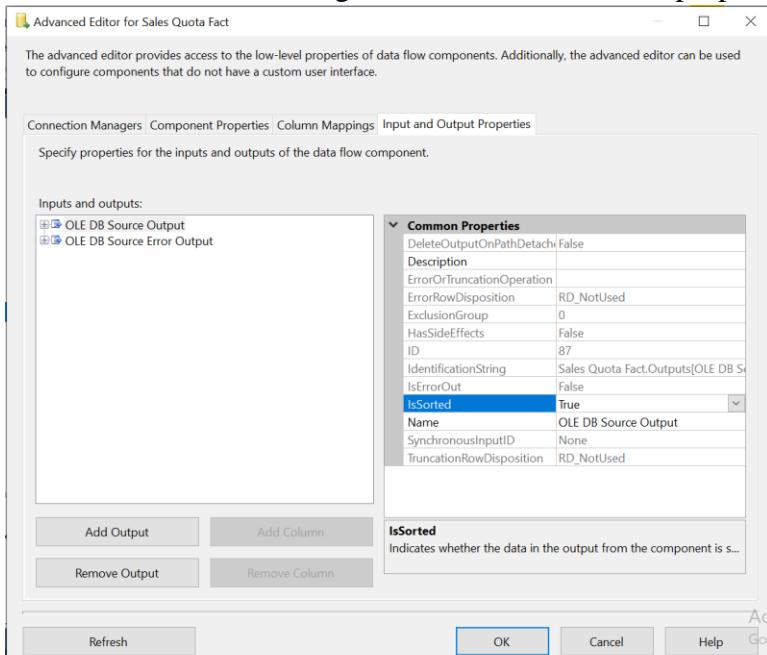
- 31) Pilih Preview untuk melihat hasil dari query tersebut.



- 32) Klik kanan pada Sales Quota Fact, kemudian pilih Show Advanced Editor.

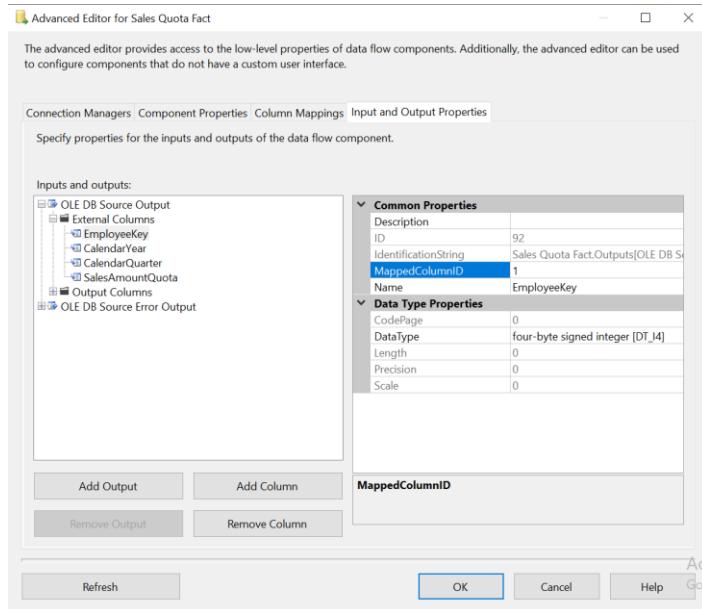


- 33) Pada bagian Input and Output Properties tab, klik OLE DB Source Output pada sisi kiri window. Pada window bagian kanan, ubah IsSorted property menjadi True.

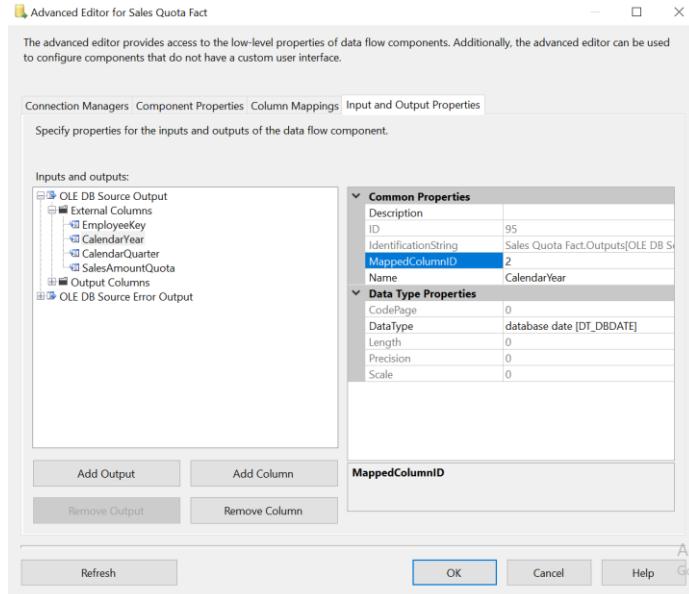


- 34) Expand OLE DB Source Output pada bagian kiri Window. Lakukan perubahan.

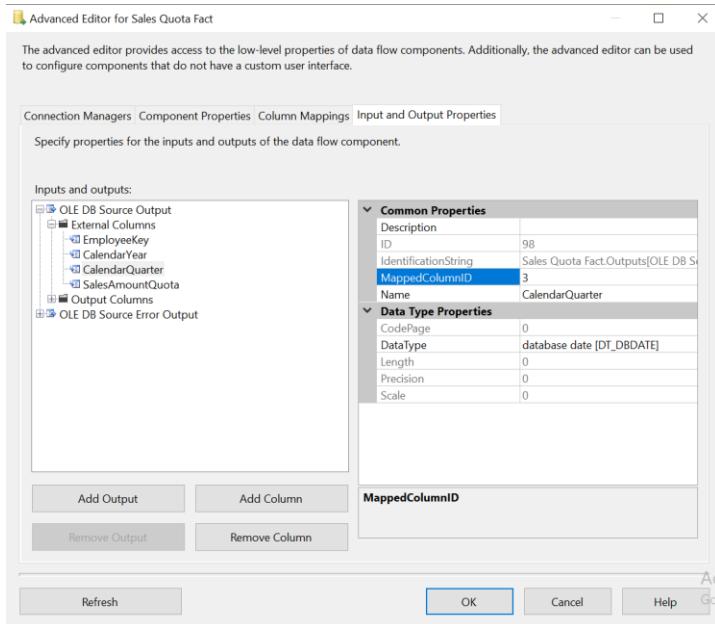
- Pilih EmployeeKey Column dan ubah SortKeyPosition ke 1.



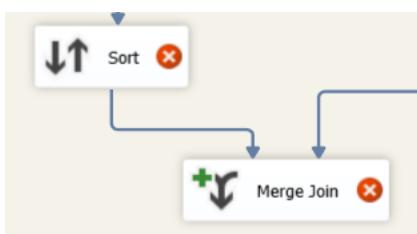
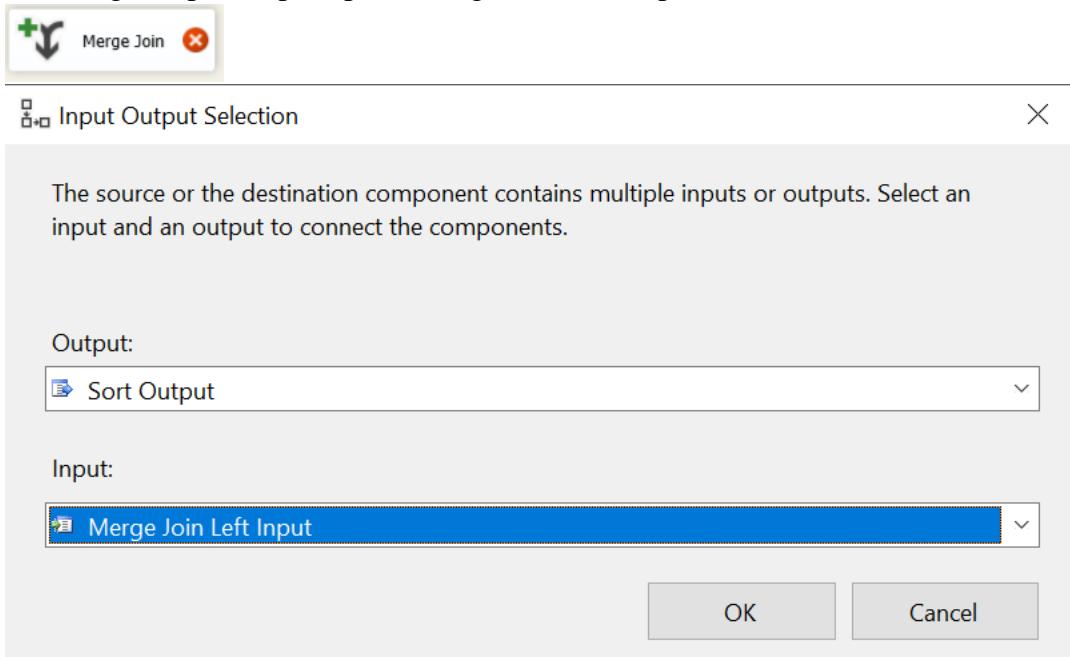
- Pilih CalendarYear dan ubah SortKeyPosition ke 2.



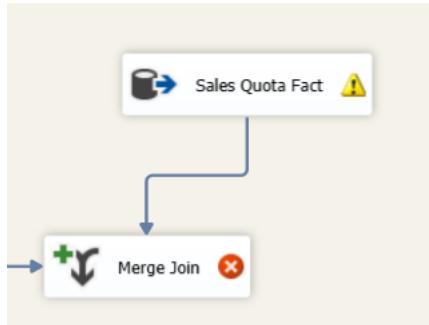
- Pilih CalendarQuarter dan ubah SortKeyPosition ke 3.



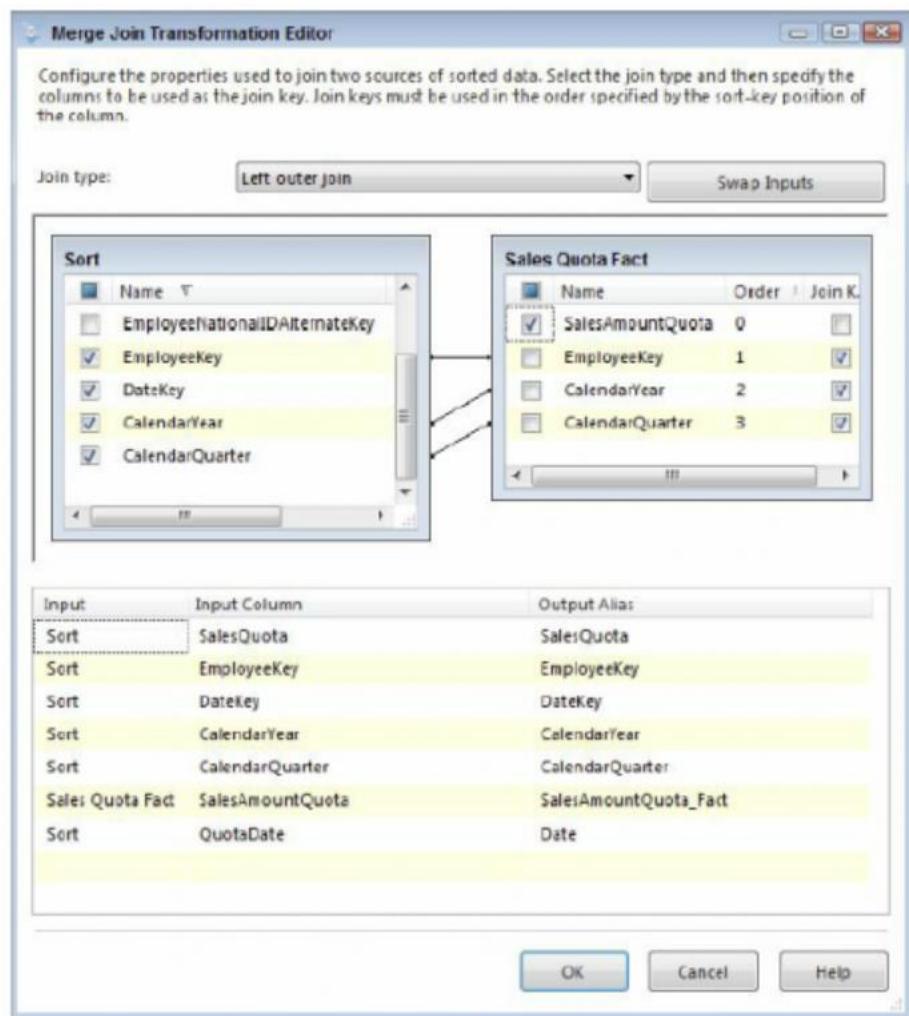
- 35) Klik OK untuk menyimpan perubahan yang dilakukan.
 36) Dari SSIS Toolbox, drag and drop Merge Join. Kemudian hubungkan ke Sort. Ketika dihubungkan, pilih input option Merge Join Left Input.



37) Hubungkan Sales Quota Fact ke Merge Join.

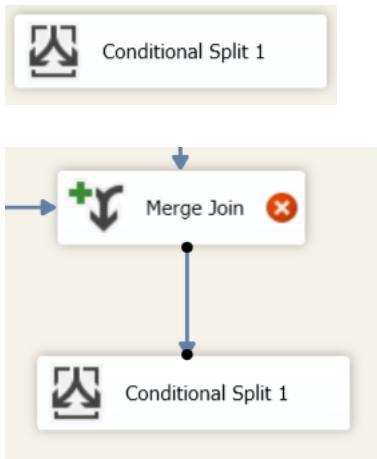


38) Double click pada Merge Join. Kita akan melihat bahwa EmployeeKey, CalendarYear, and CalendarQuarter telah dijoin antara input.

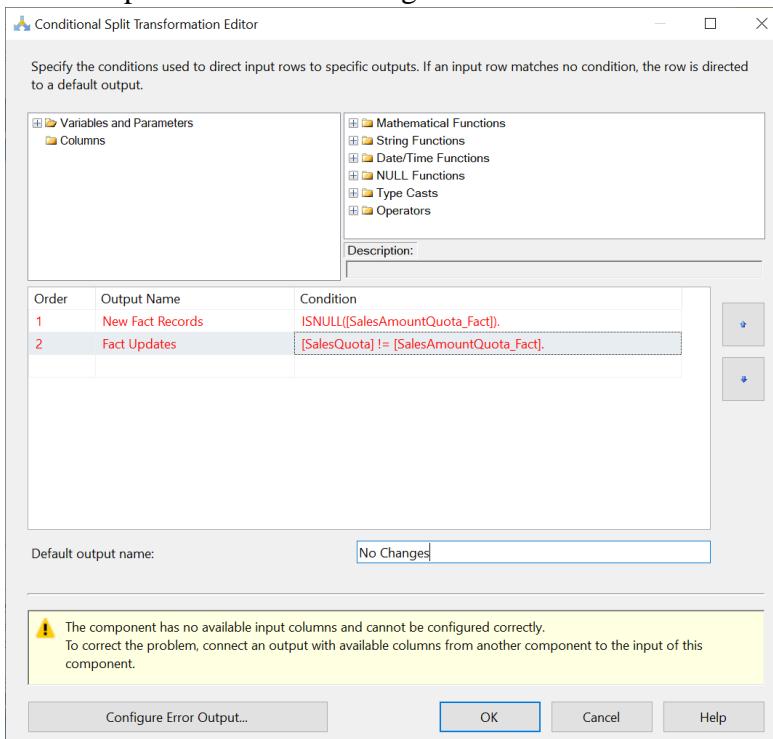


39) Pilih OK.

40) Kemudian kita akan melakukan identifikasi terhadap quotas yang baru dan yang mana Sales quotas yang berubah. Drag and drop Conditional Split ke design area Data Flow dan hubungkan antara Merge Join dengan Conditional Split.

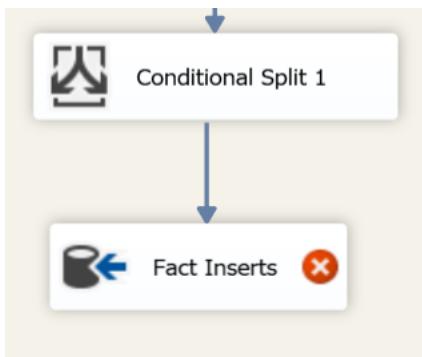


- 41) Double click pada Conditional split.
- 42) Tambahkan kondisi baru New Fact Records dengan kondisi ISNULL ([SalesAmountQuota_Fact]).
- 43) Tambahkan kondisi yang kedua dengan nama Fact Updates dengan kondisi [SalesQuota] != [SalesAmountQuota_Fact].
- 44) Ubah output name ke No Changes.

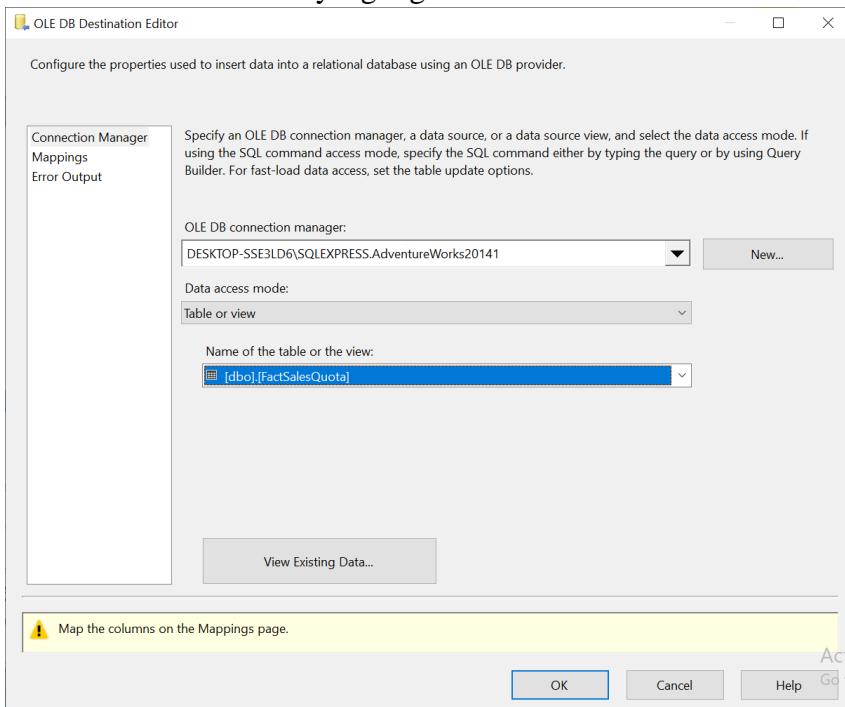


- 45) Tambahkan OLE DB Destination ke design area Data Flow dan *rename* menjadi Fact Inserts.
- 46) Hubungkan Conditional Split dengan OLE DB Destination atau Fact Inserts.





- 47) Double click pada Fact Insert, kemudian sesuaikan dengan Connection Manager dari device dan database serta tabel yang digunakan.



- 48) Untuk menghandle fact table updates, drag and drop OLE DB Command ke design area Data Flow dan *rename* menjadi Fact Updates. Hubungkan dari Condition Split ke Fact Updates.

