



Integration Services

Mag. Thomas Griesmayer

ETL

- Extract, Transform, Load is a process, that is able to use data from different data sources, transform the data and store the result in any data destination.
- ETL
 - Extract
 - Transform
 - Load
- ETL is very well known in the dataware house area, where a huge amount of data from different database systems is transformed and stored in a huge dataware house database.

Mag. Thomas Griesmayer

1

SQL Server Integration Services

- Microsoft Integration Services is a platform for building enterprise-level data integration and data transformations solutions.
- You use Integration Services to solve complex business problems by copying or downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data.



<https://msdn.microsoft.com/en-us/library/ms141026.aspx> (12.2.2017)

Mag. Thomas Griesmayer

2

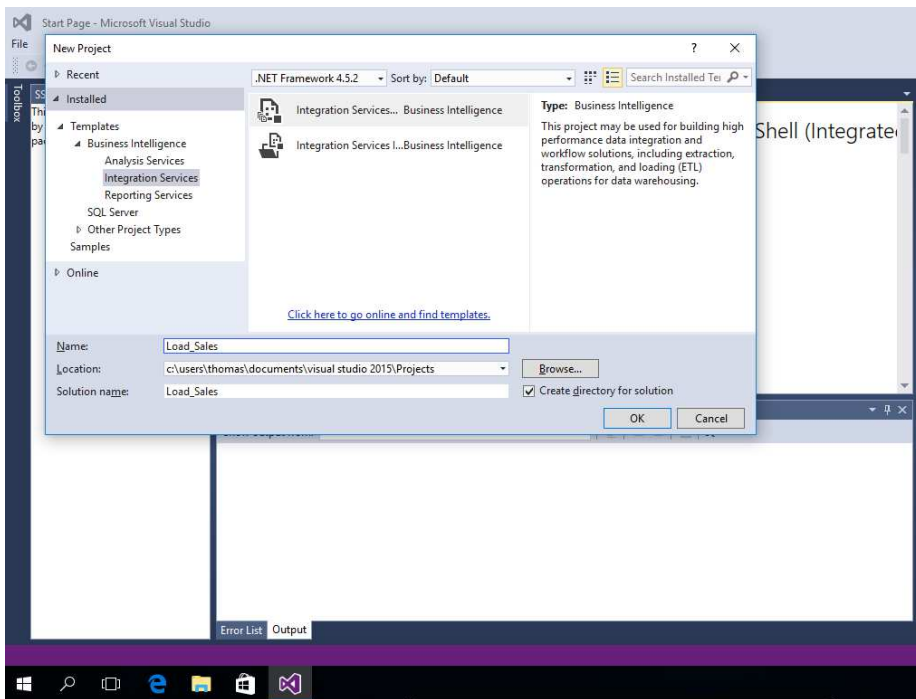
SQL Server Integration Services

- Integration Services can extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations.
- Integration Services includes a rich set of built-in tasks and transformations; tools for constructing packages; and the Integration Services service for running and managing packages.

<https://msdn.microsoft.com/en-us/library/ms141026.aspx> (12.2.2017)

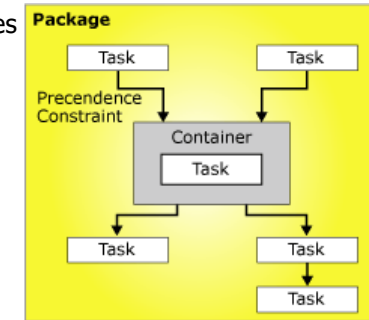
Mag. Thomas Griesmayer

3



Control Flow

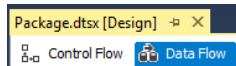
- A package consists of a control flow and, optionally, one or more data flows.
- SQL Server Integration Services provides three different types of control flow elements:
 - containers that provide structures in packages
 - tasks that provide functionality
 - precedence constraints that connect the executables, containers, and tasks into an ordered control flow.



<https://msdn.microsoft.com/en-us/library/ms137681.aspx> (12.2.2017)

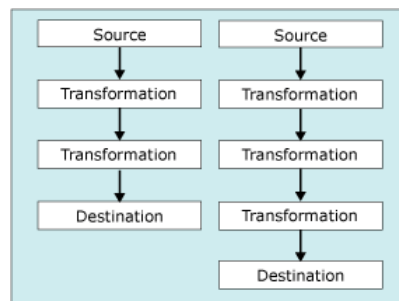
Mag. Thomas Griesmayer

5



Data Flow

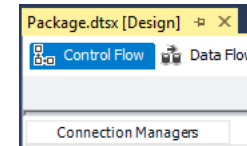
- The Data Flow task encapsulates the data flow engine that moves data between sources and destinations, and lets the user transform, clean, and modify data as it is moved.
- A data flow consists of at least one data flow component.
- typical components:
 - sources that extract data
 - transformations that modify, route, or summarize data
 - destinations that load data



<https://msdn.microsoft.com/en-us/library/ms141122.aspx> (12.2.2017)

Mag. Thomas Griesmayer

6



Connection Manager

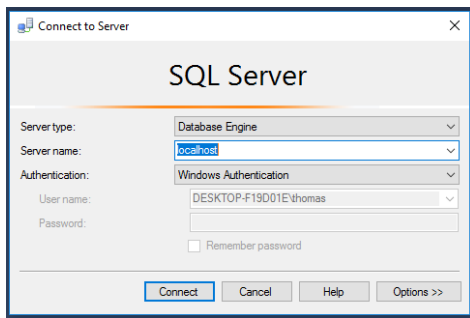
- Integration Services includes a variety of connection managers for connecting to different data sources, such as relational databases, Analysis Services databases, and files in CSV and XML formats.

ADO Connection	ADO.NET Connection
Excel Connection	Flat File Connection
FTP Connection	HTTP Connection
ODBC Connection	OLE DB Connection
SMTP Connection	...

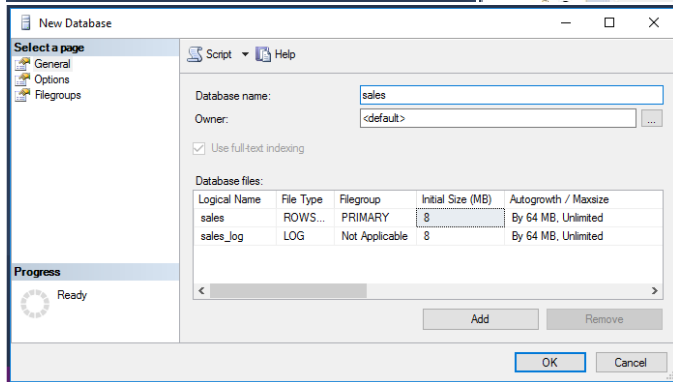
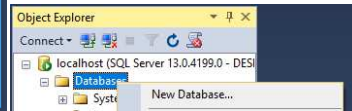
<https://msdn.microsoft.com/en-us/library/ms140237.aspx> (12.2.2017)

Mag. Thomas Griesmayer

7



Database



8

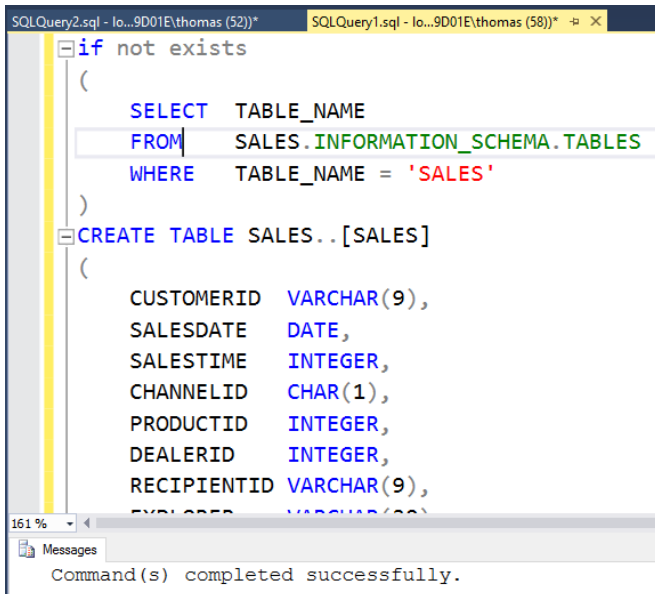
Table

```

if not exists
(
    SELECT TABLE_NAME
    FROM SALES.INFORMATION_SCHEMA.TABLES
    WHERE TABLE_NAME = 'SALES'
)
CREATE TABLE SALES..[SALES]
(
    CUSTOMERID VARCHAR(9),
    SALESDATE DATE,
    SALESTIME INTEGER,
    ...
)
go
  
```

Mag. Thomas Griesmayer

9

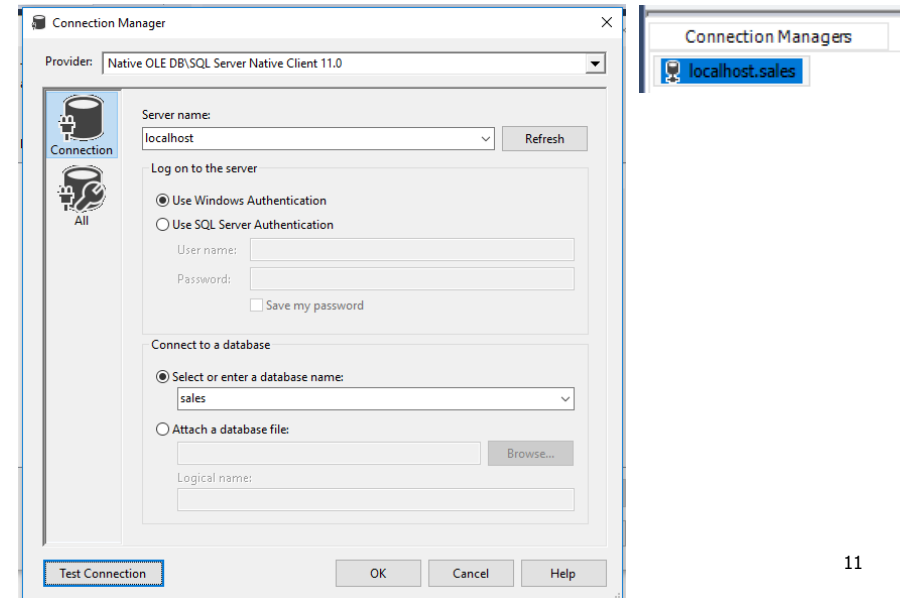


Table

Mag. Thomas Griesmayer

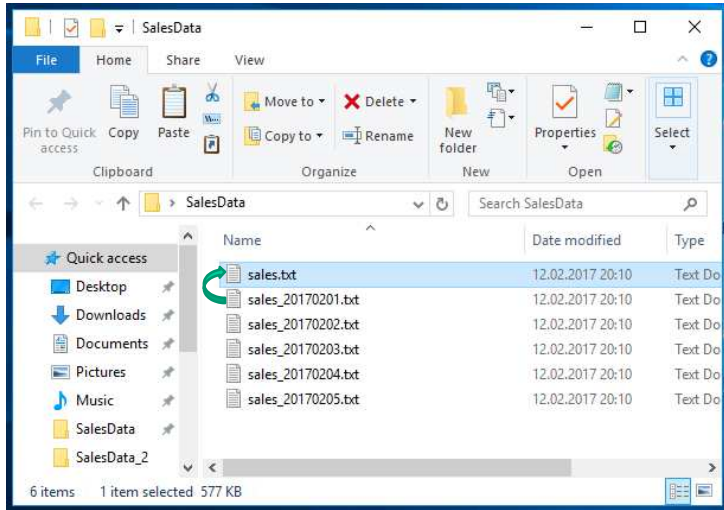
10

Connection



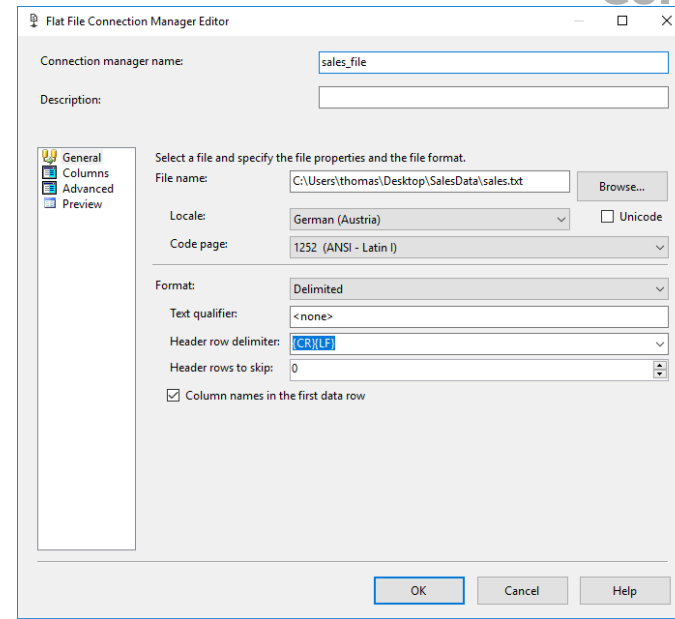
11

File

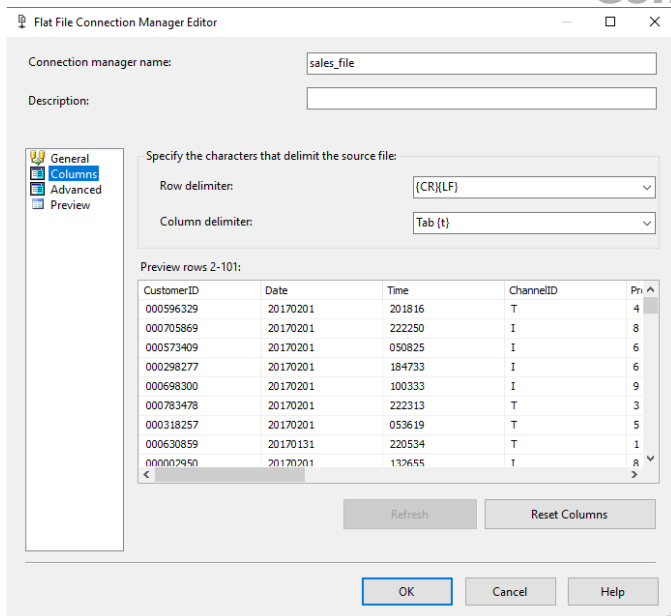


Mag. Thomas Griesmayer

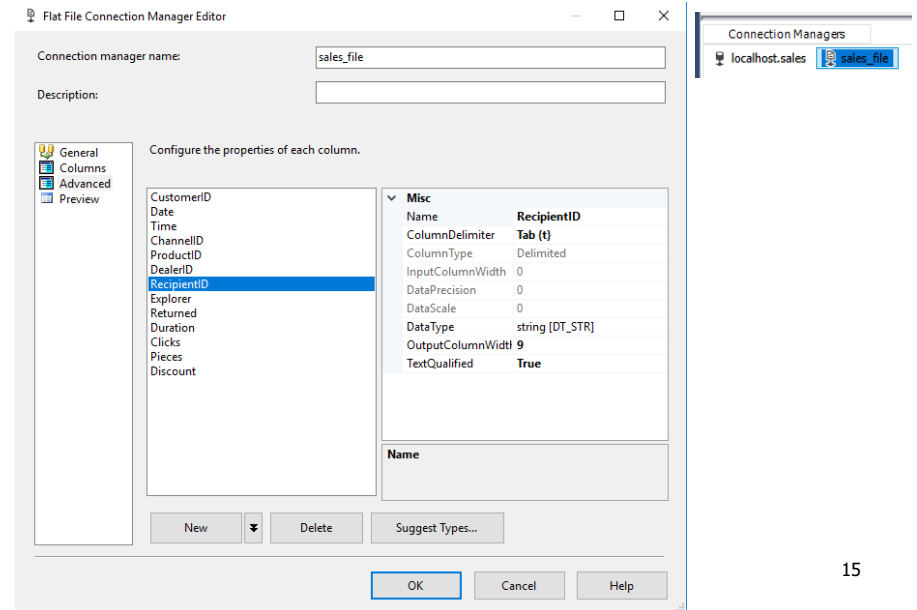
Connection

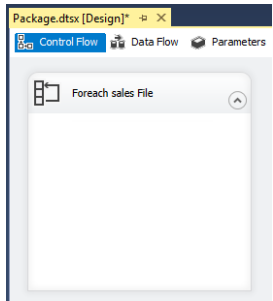


Connection

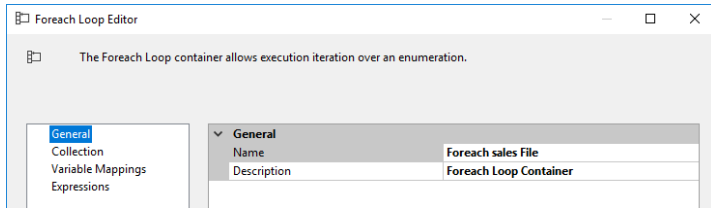


Connection

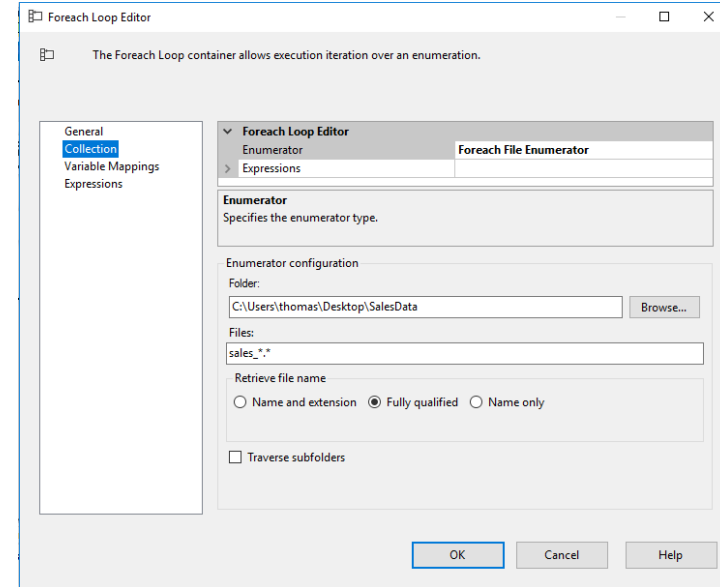




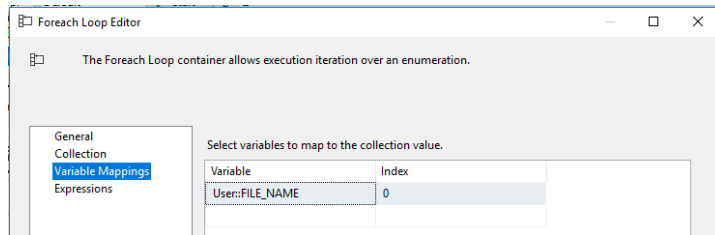
ForEach



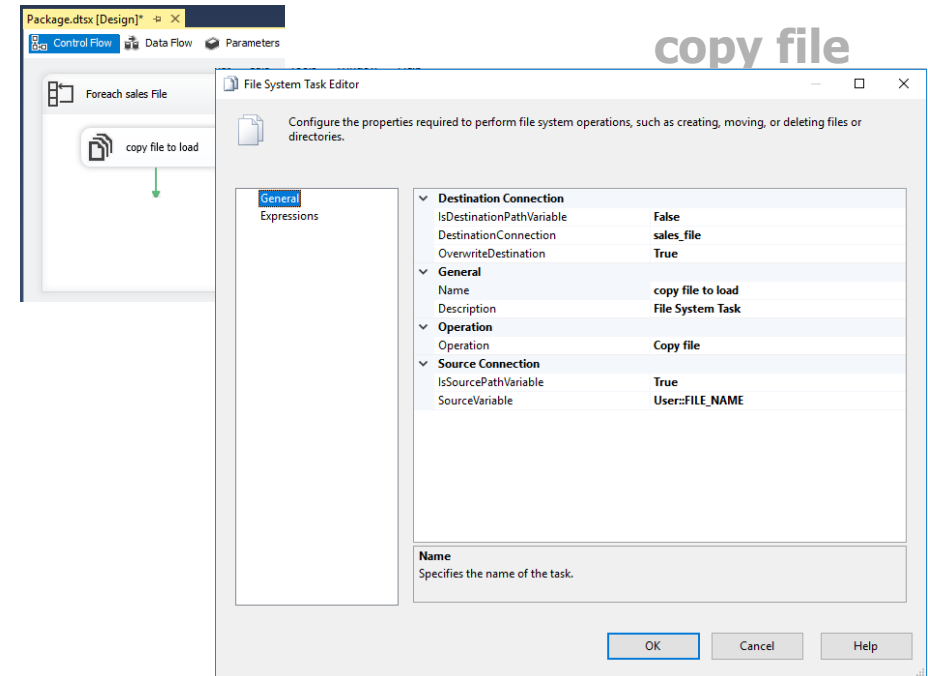
ForEach

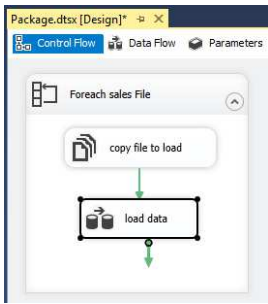


ForEach

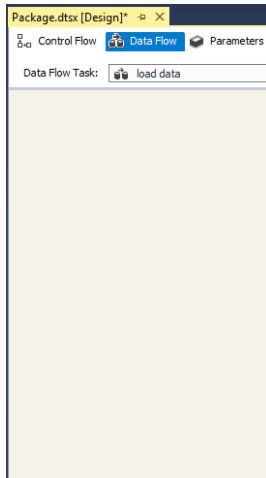


copy file

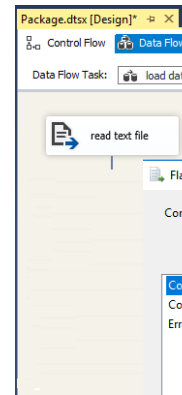




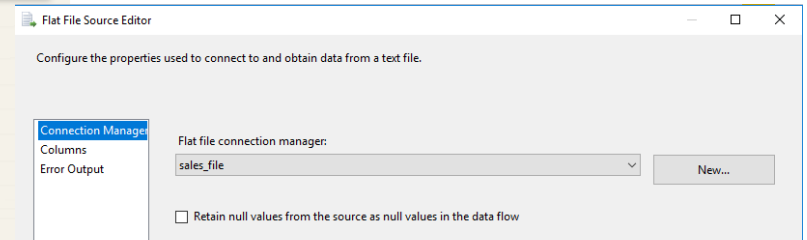
Data Flow Task



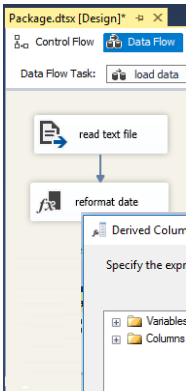
Mag. Thomas Griesmayer



Data Flow Task

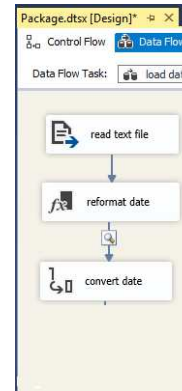
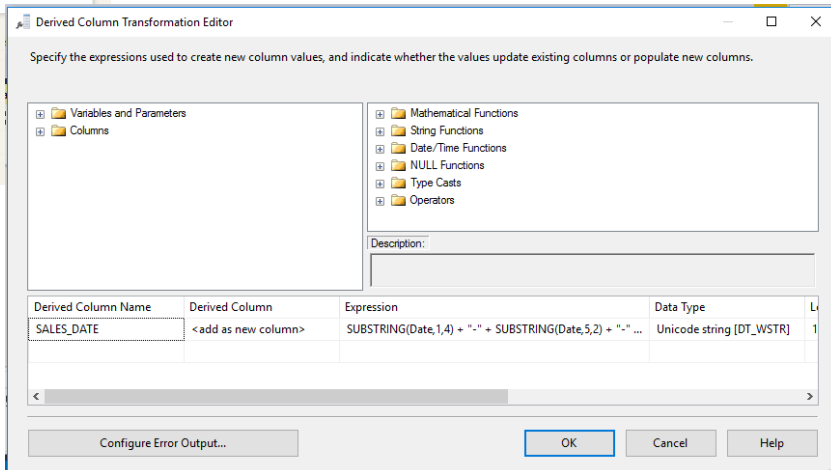


Input or Output	Column	Error	Truncation	Description
Flat File Source ...	CustomerID	Ignore failure	Ignore failure	Conversion
	Date	Ignore failure	Ignore failure	Conversion
	Time	Ignore failure	Ignore failure	Conversion
	ChannelID	Ignore failure	Ignore failure	Conversion
	ProductID	Ignore failure	Ignore failure	Conversion
	DealerID	Ignore failure	Ignore failure	Conversion

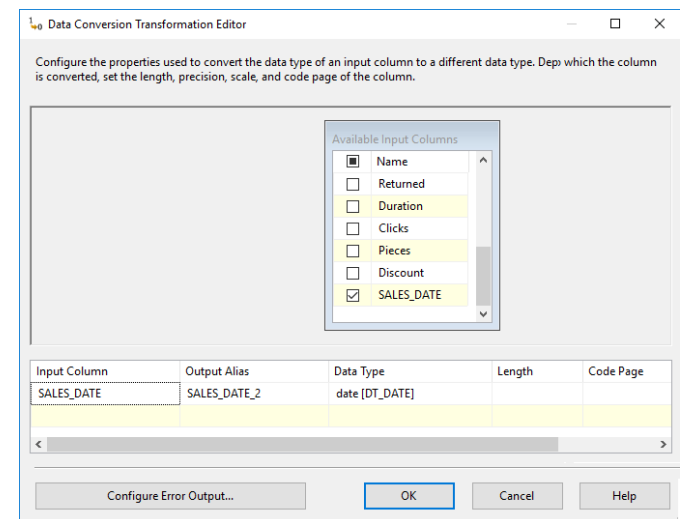


Data Flow Task

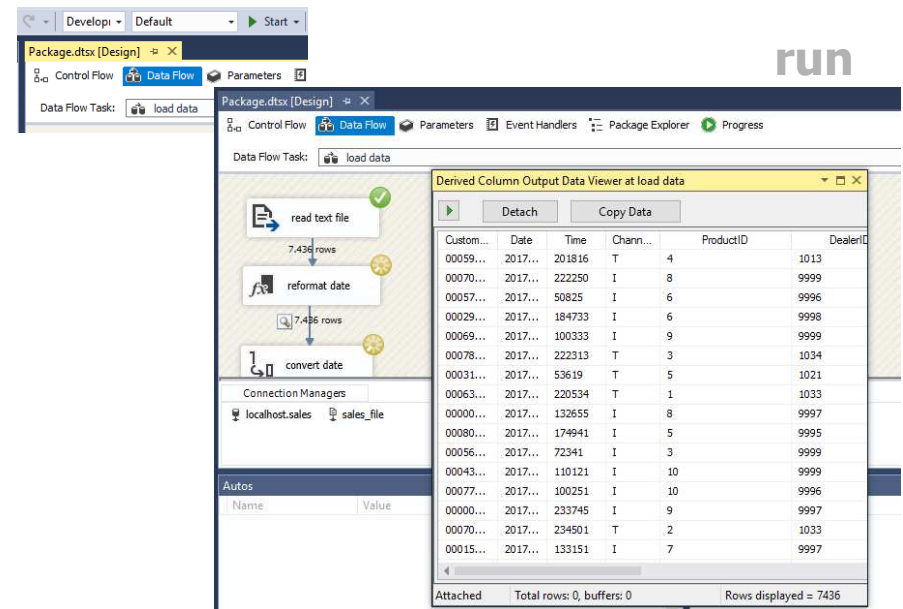
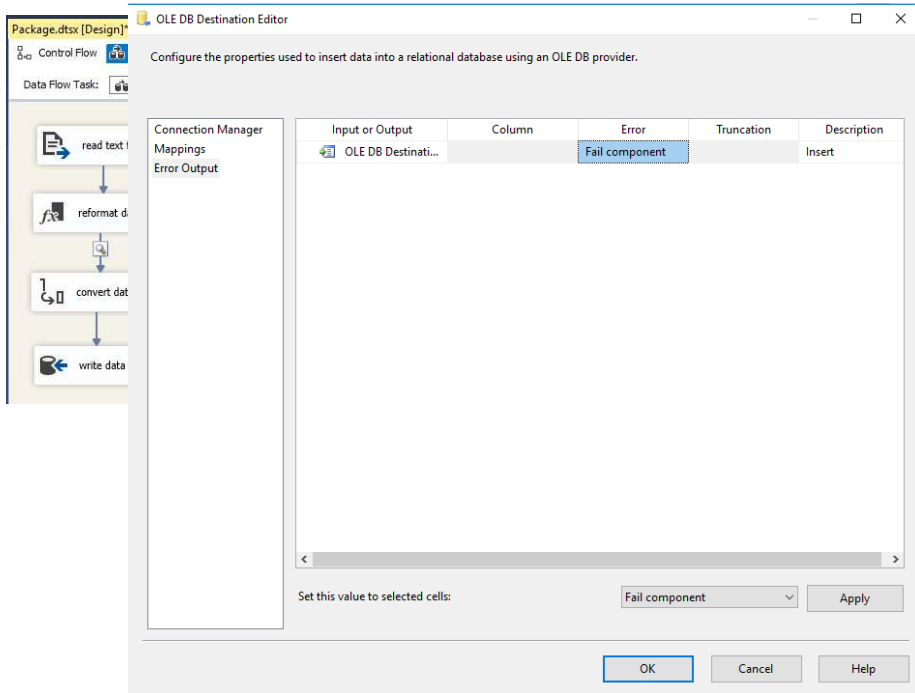
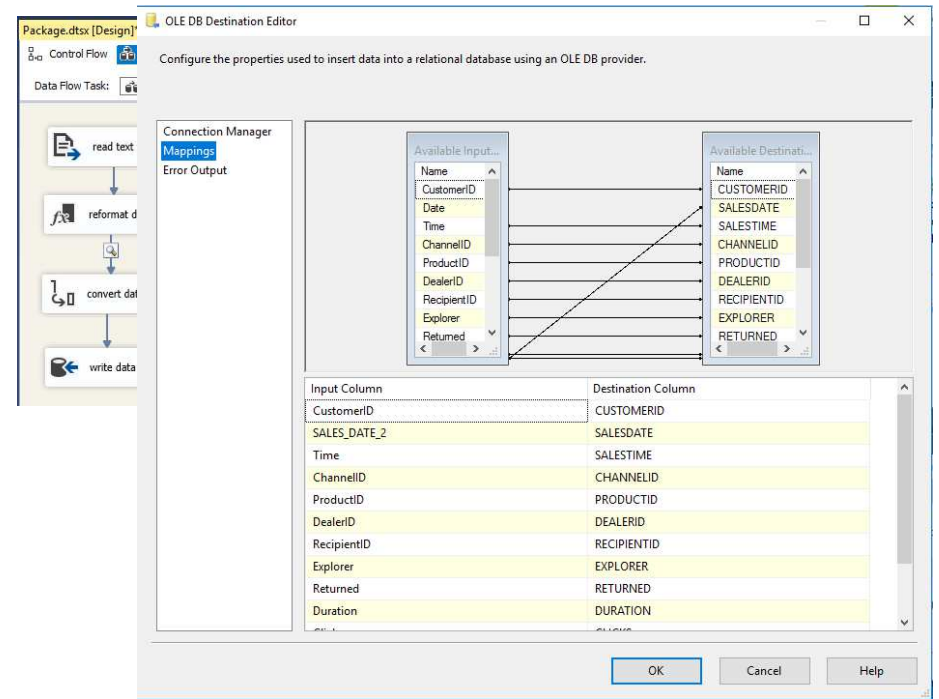
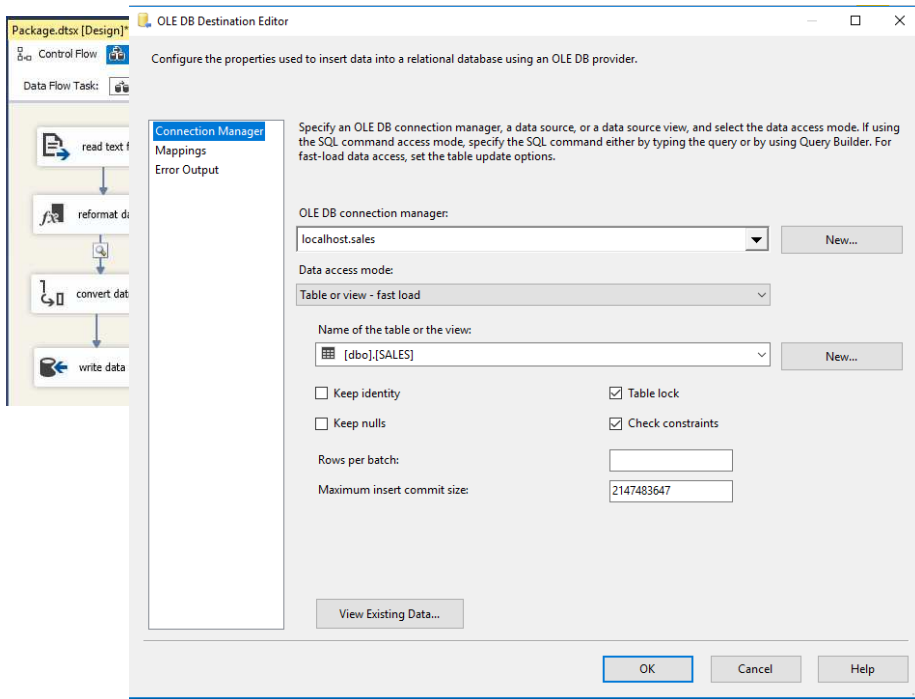
SUBSTRING(Date,1,4) + "-" +
 SUBSTRING(Date,5,2) + "-" +
 SUBSTRING(Date,7,2)



Data Flow Task



Mag. Thomas Griesmayer



result

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the server structure, including the 'sales' database. The query window contains two SQL queries:

```
SELECT count(*)  
FROM sales.sales;  
  
SELECT *  
FROM sales.sales;
```

The Results pane shows a grid of data with the following columns: CUSTOMERID, SALESDATE, SALESTIME, CHANNELID, PRODUCTID, DEALERID, and REC. The first row is highlighted.

	CUSTOMERID	SALESDATE	SALESTIME	CHANNELID	PRODUCTID	DEALERID	REC
1	000596329	2017-02-01	NULL	T	4	1013	000
2	000705869	2017-02-01	NULL	I	8	9999	000
3	000573409	2017-02-01	50825	I	6	9996	000
4	000298277	2017-02-01	NULL	I	6	9998	000
5	000698300	2017-02-01	NULL	I	9	9999	000
6	000783478	2017-02-01	NULL	T	3	1034	000
7	000318257	2017-02-01	53619	T	5	1021	000
8	000630859	2017-01-31	NULL	T	1	1033	000
9	000002950	2017-02-01	NULL	I	8	9997	000
10	000809083	2017-02-01	NULL	I	5	9995	000
11	000560254	2017-02-01	72341	I	3	9999	000
12	000430618	2017-02-01	NULL	I	10	9999	000
13	000778531	2017-02-01	NULL	I	10	9996	000