

Extract and Transform Data



Thomas LeBlanc

DATA WAREHOUSE ARCHITECT

@TheSmilingDBA www.Thomas-LeBlanc.com



Agenda



Implement data flow

Connections and sources

Transforming data in a data flow

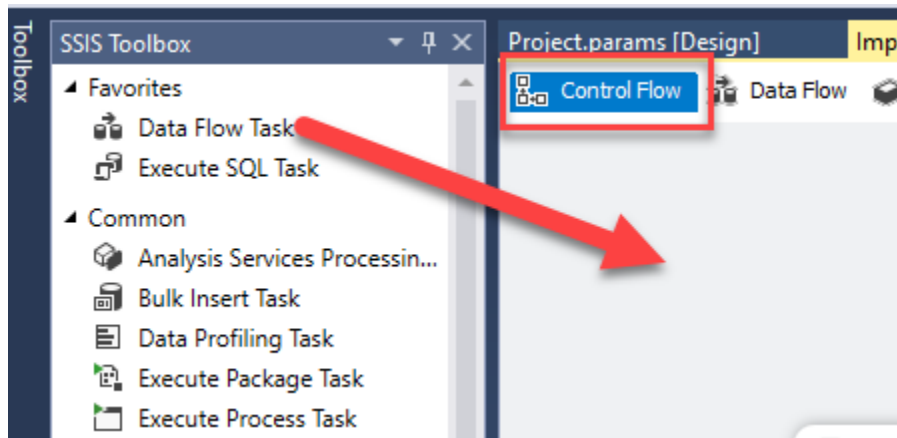
Checking for existing data

Using a data viewer



Implementing a Data Flow





Added to control flow

Favorites can be customized

Multiple data flows on control flow

Main source for transformations

- Source
- Derived columns
- Conditional Split
- Data type conversion
- Lookups

Source in a Data Flow

Required

No data, no data flow

Connection

Must have connection
created

Various types

Flexibility of using SSIS



Demo



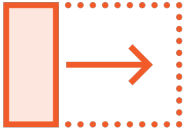
Navigate a data flow



Connections and Sources



Connections and Sources



Discover various sources – text files, tables, Excel, web data, etc.



Source components provide extraction



Control items in the data stream

Connection to extract
Source component
Columns selectable

Flat File Source Editor

Configure the properties used to connect to and obtain data from a text file.

Connection Manager

Columns

Error Output

Available External Columns

<input type="checkbox"/>	Name
<input checked="" type="checkbox"/>	ProductID
<input checked="" type="checkbox"/>	ProductName
<input checked="" type="checkbox"/>	SupplierID
<input checked="" type="checkbox"/>	CategoryID
<input type="checkbox"/>	QuantityPerUnit
<input checked="" type="checkbox"/>	UnitPrice
<input checked="" type="checkbox"/>	UnitsInStock
<input checked="" type="checkbox"/>	UnitsOnOrder
<input checked="" type="checkbox"/>	ReorderLevel

External Column	Output Column
ProductID	ProductID
ProductName	ProductName
SupplierID	SupplierID
CategoryID	CategoryID
UnitPrice	UnitPrice
UnitsInStock	UnitsInStock
UnitsOnOrder	UnitsOnOrder
ReorderLevel	ReorderLevel
Discontinued	Discontinued

OK

Cancel

Help



Demo



Data Flow



Transforming Data in a Data Flow



Convert data types

Common transformation

Data Conversion component

Derived Column

Select statement

- Advanced properties
- Easier to re-create
- Not apparent to developer
- Needs documentation



Rules to split data

Uses logic

Function (Math,
String, Date or Null)

Type casts

Operators

Order and Names

Conditional Split



Other Transformations

Component	Task
Merge	Bring together 2 sorted data flows
Union	Combine data after split
Sort	Reorder data in flow
Multicast	Split flow of data
Copy Column	Makes a copy of a column (new name)
Pivot or Unpivot	Pivot or unpivot data structure



Demo



Convert data types

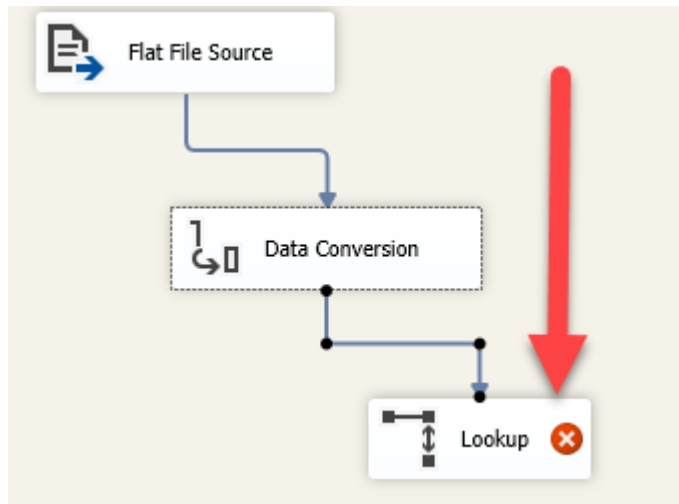
Conditional Split



Checking for Existing Data



Check for Existing Data



No duplicates

Connect to destination

Compare columns

Gotchas

- Data types
- Nulls
- Additional columns (rename)

Demo



Lookup

Handle problems



Using a Data Stream Viewer



Debugging Data Flow

Write Data

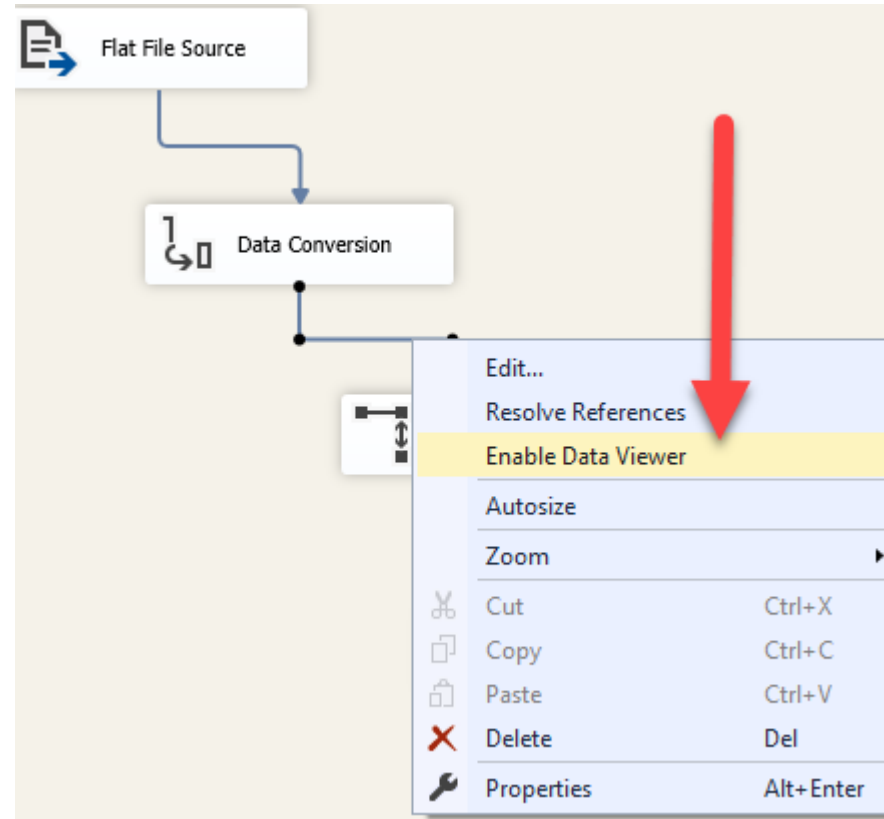
Very time consuming

Enable Data Viewer

Inline debugging with stop options



Data Flow option
Not in Control Flow
Selectable columns
Pausing while
executing
Stop processing
Compare data in
columns



Demo



Debug data



Summary



Connection options

Detailed a data flow

Cool transformations

Previewing data in flow

Stepping up to next task

