

SQL Server Integration Services (SSIS) Training Kit(Part 3)

Lesson 16: Export Column
Transformation

Export Column Transformation

- The Export Column transformation reads data in a data flow and inserts the data into a file.
- The data to be written must have a DT_TEXT, DT_NTEXT, or DT_IMAGE data type.
- You can configure the Export Column transformation in the following ways:
- Specify the data columns and the columns that contain the path of files to which to write the data.
- Specify whether the data-insertion operation appends or truncates existing files.
- Specify whether a byte-order mark (BOM) is written to the file.
- Note:A BOM is written only when the data is not appended to an existing file and the data has the DT_NTEXT data type.



Copyright © Capgemini 2015. All Rights Reserved 2

Export Column Transformation

- EXAMPLE: The Data below shown has POLICYNUMBER column(type should be dt_text,dt_ntext or dt_image) which has to be exported to the path mentioned in column called PLANNNAME.

POLICYNUMBER	CARRIERCODE	POLICYSTATUS	PLANNNAME
POL1	CAR3	3	D:\TUSHAR\45895 on pc-p41284\assignments_Files\new1.txt
POL1	CAR1	4	D:\TUSHAR\45895 on pc-p41284\assignments_Files\new.txt
POL2	CAR2	5	D:\TUSHAR\45895 on pc-p41284\assignments_Files\new2.txt

1st record will be exported to new1.txt

2nd record will be exported to new.txt

3rd record will be exported to new2.txt

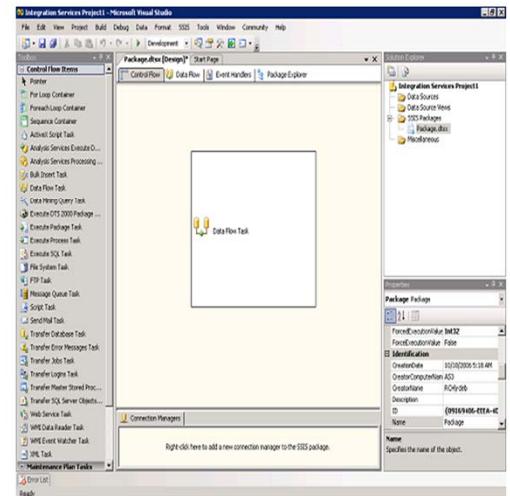


Export Column Transformation

- Drag a 'Data Flow Task' from Control Flow Items Navigator to Control Flow Task Window.

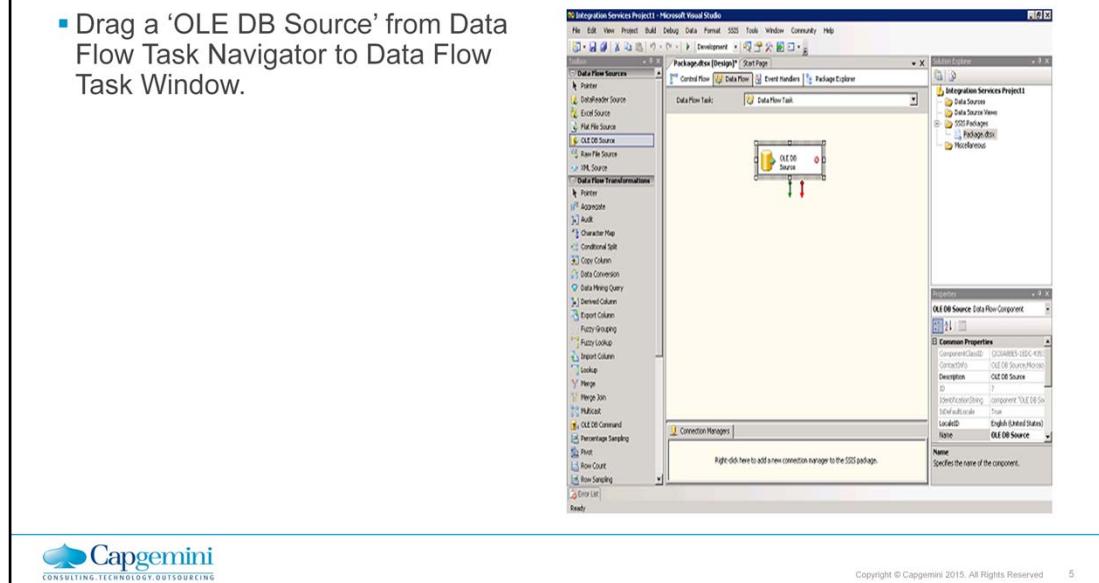


Copyright © Capgemini 2015. All Rights Reserved 4



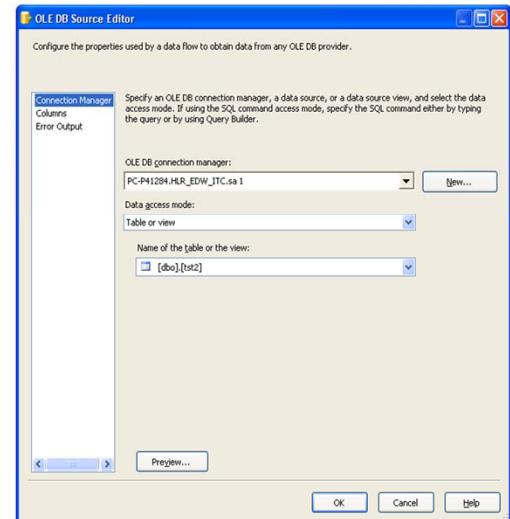
Export Column Transformation

- Drag a 'OLE DB Source' from Data Flow Task Navigator to Data Flow Task Window.



Export Column Transformation

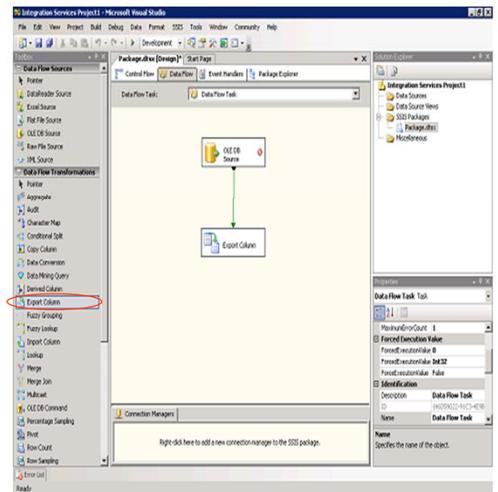
- Double click or Right click on 'OLE DB Source' to open 'OLE DB Source Editor'.
- Set the 'Connection Manager' by giving appropriate information, e.g. Connection Manager Name, Table or Query, Table Name.



Copyright © Capgemini 2015. All Rights Reserved 6

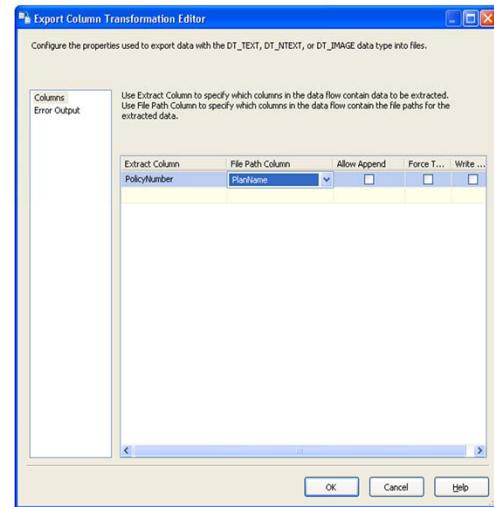
Export Column Transformation

- Drag ‘Export Column’ transformation from Data Flow Transformation Navigation Toolbox to the data flow task window.
- Link ‘Export Column Transformation’ with ‘OLE DB Source’.



Export Column Transformation

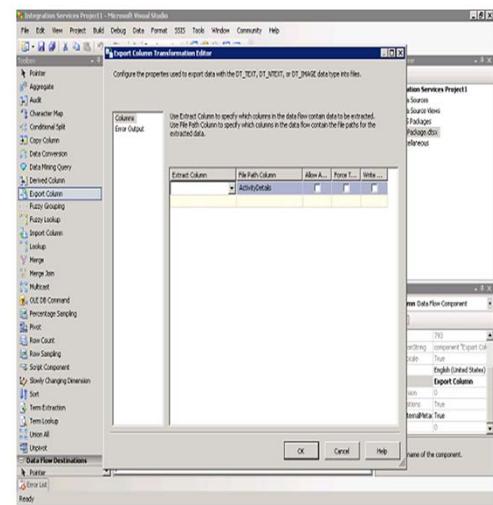
- Double Click on 'Export Column transformation' or Right click and click 'Edit'.
- Properly set the fields 'extract column' and 'file path column' and others.
- As per previous example set 'policynumber' as extract column and file path column as 'planname'



Copyright © Capgemini 2015. All Rights Reserved 8

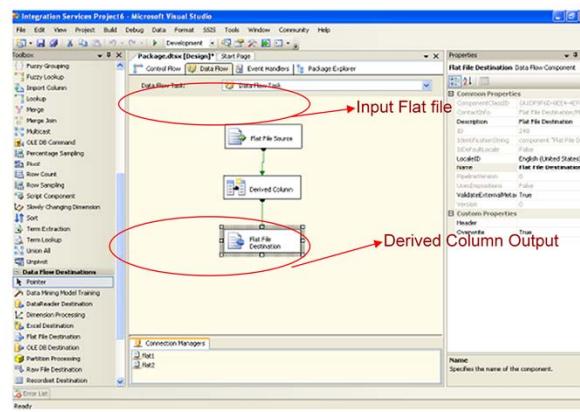
Export Column Transformation

- Allow Append property if checked will append the data to the same file on every load
- Force Truncate property will truncate the data for every load.
- Click Ok
- Execute the dataflow.



Copyright © Capgemini 2015. All Rights Reserved 9

Derived Column Transformation



Fuzzy Grouping Transformation

Lesson 17:

Fuzzy Grouping Transformation

- The transformation output includes all input columns, one or more columns with standardized data, and a column that contains the similarity score. The score is a decimal value between 0 and 1. The canonical row has a score of 1. Other rows in the fuzzy group have scores that indicate how well the row matches the canonical row. The closer the score is to 1, the more closely the row matches the canonical row. If the fuzzy group includes rows that are exact duplicates of the canonical row, these rows also have a score of 1. The transformation does not remove duplicate rows; it groups them by creating a key that relates the canonical row to similar rows.



Copyright © Capgemini 2015. All Rights Reserved 12

Fuzzy Grouping Transformation

The transformation produces one output row for each input row, with the following additional columns:

- `_key_in` : a column that uniquely identifies each row.
- `_key_out` : a column that identifies a group of duplicate rows. The `_key_out` column has the value of the `_key_in` column in the canonical data row. Rows with the same value in `_key_out` are part of the same group. The `_key_out` value for a group corresponds to the value of `_key_in` in the canonical data row.
- `_score` : a value between 0 and 1 that indicates the similarity of the input row to the canonical row.

These are the default column names and you can configure the Fuzzy Grouping transformation to use other names. The output also provides a similarity score for each column that participates in a fuzzy grouping



Copyright © Capgemini 2015. All Rights Reserved 13

Fuzzy Grouping Transformation

“The Fuzzy Grouping transformation performs data cleaning tasks by identifying rows of data that are likely to be duplicates and selecting a canonical row of data to use in standardizing the data.”

Types of match

1. Exact Match: An exact match guarantees that only rows that have identical values in that column will be grouped. Exact matching can be applied to columns of any Integration Services data type except DT_TEXT, DT_NTEXT, and DT_IMAGE.
2. Fuzzy match: A fuzzy match groups rows that have approximately the same values. The method for approximate matching of data is based on a user-specified similarity score. Only columns with the DT_WSTR and DT_STR data types can be used in fuzzy matching.



Copyright © Capgemini 2015. All Rights Reserved 14

Fuzzy Grouping Transformation

EXAMPLE

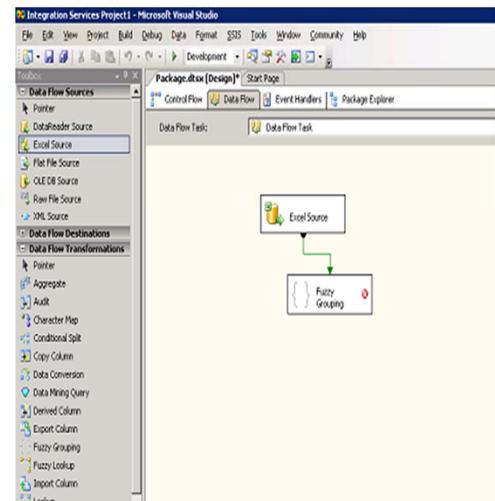
Suppose that the entries into the system are done manually across many countries and the CITY 'AUSTRALIA' is saved with many similes example 'AUSTRILIA', 'USSTRAILIA' or 'AUSTERILYA' etc. these are need to grouped to correct one. The transformation generate the value accordance with the similarity between the wrong one and right one, Based on acceptance criteria like .70 and above is treated as OR grouped to 'AUSTRALIA'



Copyright © Capgemini 2015. All Rights Reserved 15

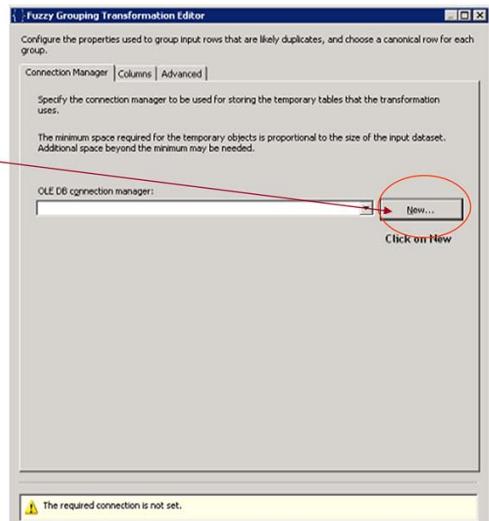
Fuzzy Grouping Transformation

- From the Toolbox, drag the Fuzzy Grouping transformation to the design surface.
- Connect the Fuzzy Grouping transformation to the data flow by dragging the connector—the green or red arrow—from the data source or a previous transformation to the Fuzzy Grouping transformation.
- Double-click the Fuzzy Grouping transformation.



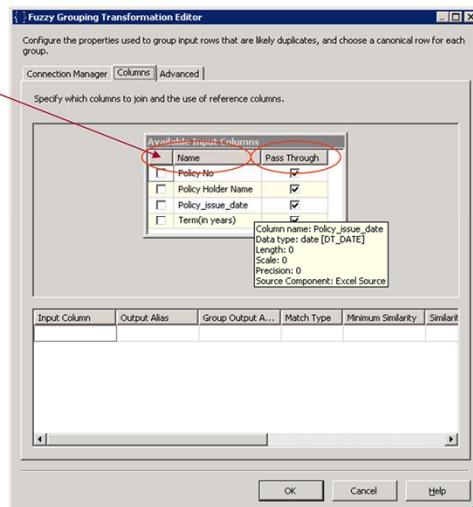
Fuzzy Grouping Transformation

- In the Fuzzy Grouping Transformation Editor dialog box, on the Connection Manager tab, select an OLE DB connection manager that connects to a SQL Server 2008 database. This will be used by SSIS for storing temporary tables that the transformation uses.
- Select a OLE DB connection



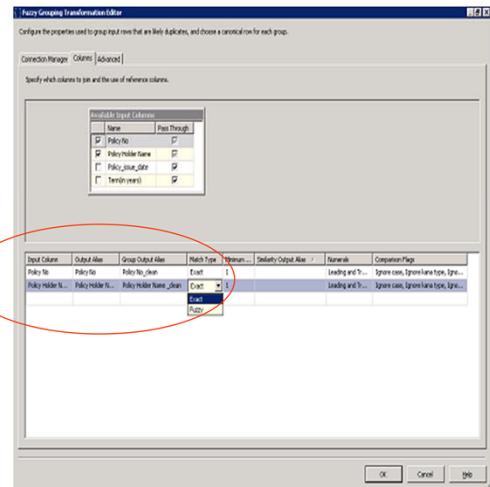
Fuzzy Grouping Transformation

- Click the Columns tab and, in the Available Input Columns list, select the check box of the input columns to use to identify similar rows in the dataset.
- Select the check box in the Pass Through column to identify the input columns to pass through to the transformation output. Pass-through columns are not included in the process of identification of duplicate rows.



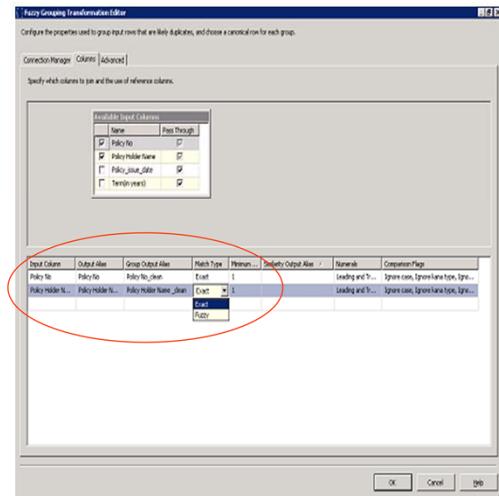
Fuzzy Grouping Transformation

- Optionally, update the names of output columns in the Output Alias column.
- Optionally, update the names of cleaned columns in the Group Output Alias column.
- Optionally, update the type of match to use in the Match Type column.
- Note: At least one column must use fuzzy matching.



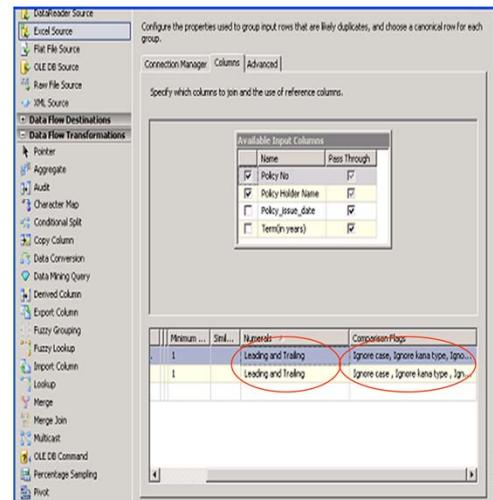
Fuzzy Grouping Transformation

- Specify the minimum similarity level columns in the Minimum Similarity column. The value must be between 0 and 1. The closer the value is to 1, the more similar the values in the input columns must be to form a group. A minimum similarity of 1 indicates an exact match.
- Optionally, update the names of similarity columns in the Similarity Output Alias column



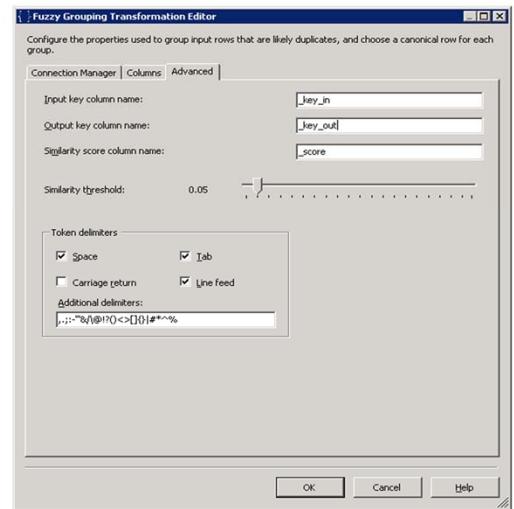
Fuzzy Grouping Transformation

- To specify the handling of numbers in data values, update the values in the Numerals column.
- To specify how the transformation compares the string data in a column, modify the default selection of comparison options in the Comparison Flags column.
- Click 'Advanced tab'.



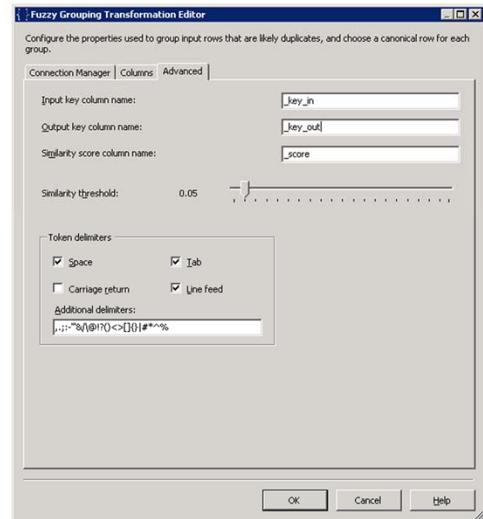
Fuzzy Grouping Transformation

- Click the Advanced tab to modify the names of the columns that the transformation adds to the output for the unique row identifier (_key_in), the duplicate row identifier (_key_out), and the similarity value (_score).
- Optionally, adjust the similarity threshold by moving the slider bar. The similarity threshold indicates how strictly the transformation identifies duplicates. The similarity thresholds can be set at the component and the column levels. The column-level similarity threshold is only available to columns that perform a fuzzy match.



Fuzzy Grouping Transformation

- The similarity range is 0 to 1. The closer to 1 the threshold is, the more similar the rows and columns must be to qualify as duplicates. You specify the similarity threshold among rows and columns by setting the Min Similarity property at the component and column levels. To satisfy the similarity that is specified at the component level, all rows must have a similarity across all columns that is greater than or equal to the similarity threshold that is specified at the component level.
- Optionally, clear the token delimiter check boxes to ignore delimiters in the data, and Click OK.



Fuzzy Lookup Transformation

Lesson 18:

Fuzzy Lookup Transformation

- The Fuzzy Lookup transformation performs data cleaning tasks such as standardizing data, correcting data, and providing missing values.
- This transformation differs from the Lookup transformation in its use of fuzzy matching. The Lookup transformation uses an equi-join to locate matching records in the reference table. It returns either an exact match or nothing from the reference table. In contrast, the Fuzzy Lookup transformation uses fuzzy matching to return one or more close matches from the reference table.
- This transformation has one input and one output.
- Only input columns with the DT_WSTR and DT_STR data types can be used in fuzzy matching. Exact matching can use any DTS data type except DT_TEXT, DT_NTEXT, and DT_IMAGE.



Copyright © Capgemini 2015. All Rights Reserved 25

Fuzzy Lookup Transformation

- Columns that participate in the join between the input and the reference table must have compatible data types. For example, it is valid to join a column with the DT_WSTR data type to a column with the SQL Server nvarchar data type, but invalid to join a column with the DT_WSTR data type to a column with the int data type.
- EXAMPLE
- Suppose that the entries into the system are done manually across many countries and the CITY 'AUSTRALIA' is saved with many similes example 'AUSTRILIA', 'USSTRAILIA' or 'AUSTERILYA' etc. when lookup on these wrong one they are need to be treated as correct one i.e. 'AUSTRALIA'. The transformation generate the value accordance with the similarity between the wrong one and right one, Based on acceptance criteria like .70 and above is treated 'AUSTRALIA'.



Copyright © Capgemini 2015. All Rights Reserved 26

Fuzzy Lookup Transformation

Steps to configure

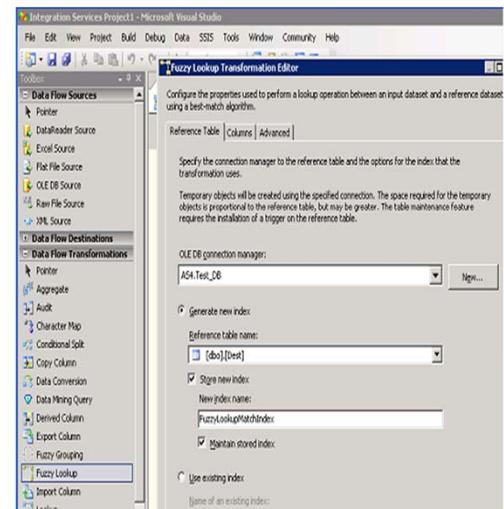
- In Business Intelligence Development Studio, open the Integration Services project that contains the package you want.
- In Solution Explorer, double-click the package to open it.
- Click the Data Flow tab, and then, from the Toolbox, drag the Fuzzy lookup transformation to the design surface.
- Connect the Fuzzy lookup transformation to the data source or a previous transformation to the Fuzzy lookup transformation.
- Double-click the Fuzzy lookup transformation to edit the pr.



Copyright © Capgemini 2015. All Rights Reserved 27

Fuzzy Lookup Transformation

- On Clicking Reference tab will be open
- Create a new connection by using the Configure OLE DB Connection Manager dialog box.



Copyright © Capgemini 2015. All Rights Reserved 28

Fuzzy Lookup Transformation

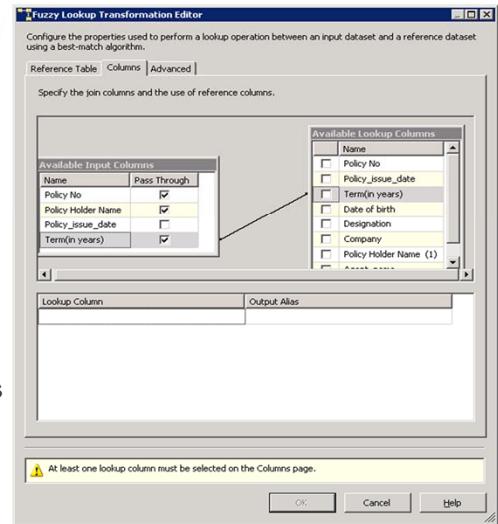
- Refer the Picture in previous slide
 - Generate new index :Specify that the transformation should create a new index to use for the lookup.
 - Reference table name :Select the existing table to use as the reference (lookup) table.
 - Store new index :Select this option if you want to save the new lookup index.
 - New index name :If you have chosen to save the new lookup index, type a descriptive name for it.
 - Maintain stored index :If you have chosen to save the new lookup index, specify whether you also want SQL Server to maintain the index.
 - Use existing index :Specify that the transformation should use an existing index for the lookup.
 - Name of an existing index :Select a previously created lookup index from the list.



Copyright © Capgemini 2015. All Rights Reserved 29

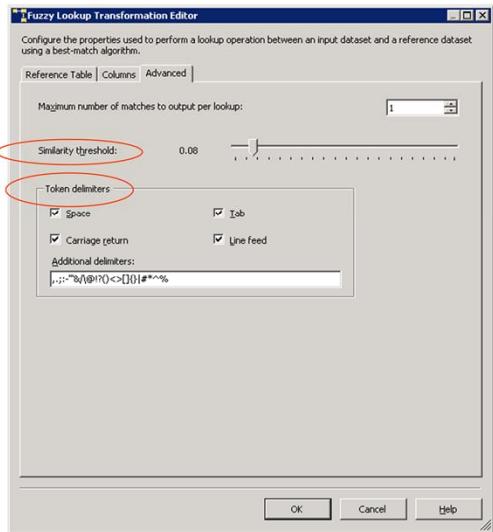
Fuzzy Lookup Transformation

- In COLUMNS tab join the ports on basis of which lookup will be performed. The both input and lookup port should have same Integration Services Data Types
- Pass Through : Specify whether to include the input columns in the output of the transformation.
- Lookup Column : Select lookup columns from the list of available reference table columns. Your selections are reflected in the check box selections in the Available Lookup Columns table. Selecting a column in the Available Lookup Columns table creates an output column that contains the reference table column value for each matching row returned.



Fuzzy Lookup Transformation

- Output Alias : Type an alias for the output for each lookup column. The default is the name of the lookup column with a numeric index value appended; however, you can choose any unique, descriptive name.
 - Maximum number of matches to output per lookup
 - Specify the maximum number of matches the transformation can return for each input row. The default is 1.



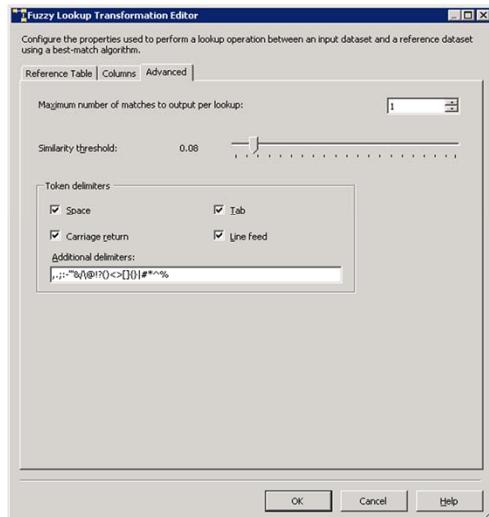
Fuzzy Lookup Transformation

■ Similarity threshold

- Set the similarity threshold at the component level by using the slider. The closer the value is to 1, the closer the resemblance of the lookup value to the source value must be to qualify as a match. Increasing the threshold can improve the speed of matching since fewer candidate records need to be considered.

■ Token delimiters

- Specify the delimiters that the transformation uses to tokenize column values. Press OK



Fuzzy Lookup Transformation

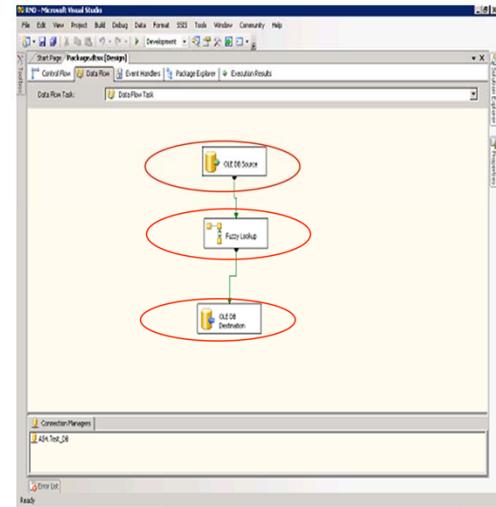
- The closer the value of Similarity Threshold is to 1, the closer the resemblance of the lookup value to the source value, & It must be to qualify as a match. Increasing the threshold can improve the speed of matching since fewer candidate records left for the consideration.



Copyright © Capgemini 2015. All Rights Reserved 33

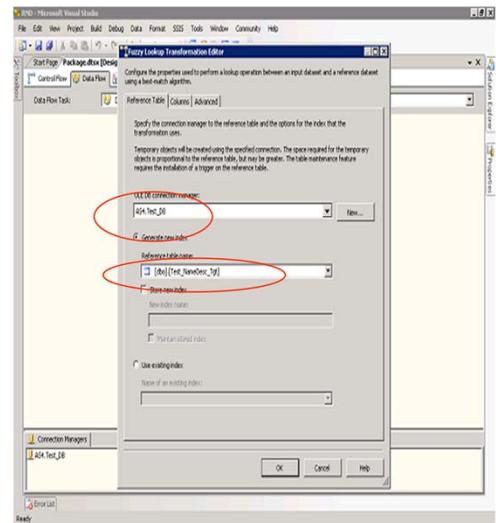
Fuzzy Lookup Transformation

- In the Example there is Source Table "Test_NameDesc". It has got 2 columns "fname" and "desc"
"Test_NameDesc_Tgt" is the Lookup Table having 2 columns "fname" and "desc" "Test_NameDesc_123" is the Target Table having 2 columns "fname" and "desc" From the Source Tables, Values from the Column "fname" needs to be Looked up with Column "fname" from Lookup Table and the "fname and desc" needs to be populated in the Target Table
- There are 3 transformation
 1. Source Transformation
 2. Fuzzy Lookup Transformation
 3. Destination Transformation



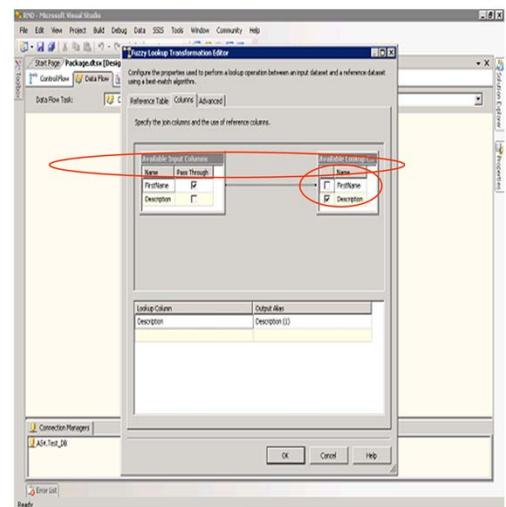
Fuzzy Lookup Transformation

- Select the Database Connection
- Select the Lookup Table



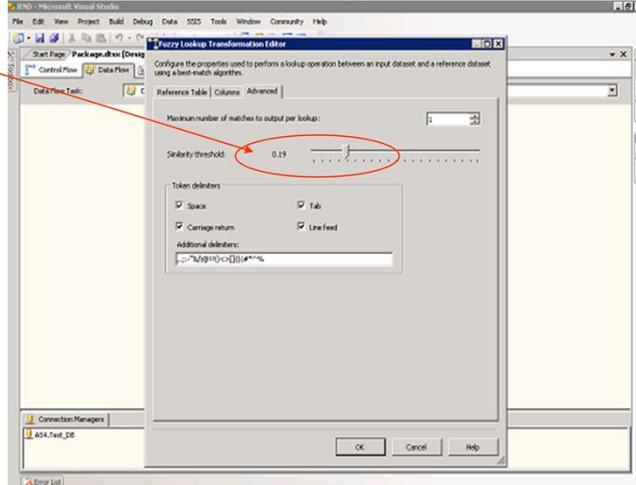
Fuzzy Lookup Transformation

- Map the Lookup Columns from Source Table with the Lookup Table
- Select the Description.



Fuzzy Lookup Transformation

➤ Select the Similarity Threshold.
Here it is kept close to 0



Similarity threshold: 0.19

Token delimiters: Space, Tab, Carriage return, Line feed

Additional delimiters: \r\n\t\b\f\v

OK Cancel Help

Capgemini CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 37

Fuzzy Lookup Transformation

All the Columns with the Similarity found is been loaded in Target Table with the "fname and Desc"

	fname	desc
1	amit	b123
2	kaushal	NULL
3	amita	b123
4	abhinav	c123
5	aditya	d213
6	kailash	NULL

```
AS4\test_DB - SQLQuery1.adt [Summary]
truncate table test_namedesc_123
select * from test_namedesc_123
```

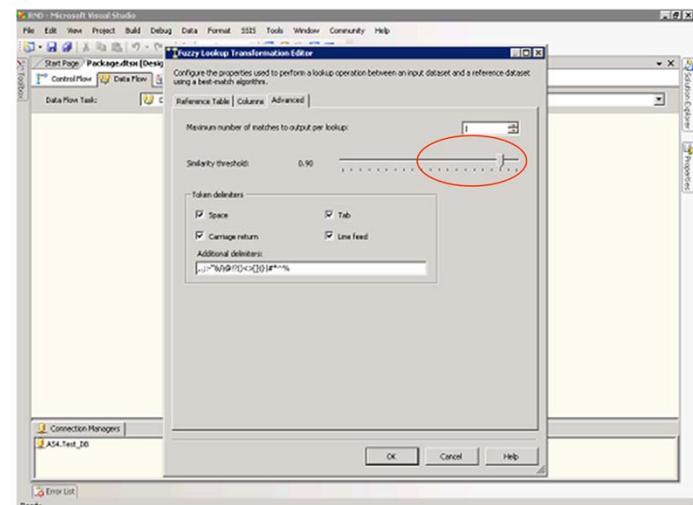
Results | Messages

name	desc
1 amit	b123
2 kaushal	NULL
3 amita	b123
4 abhinav	c123
5 aditya	d213
6 kailash	NULL

Copyright © Capgemini 2015. All Rights Reserved 38

Fuzzy Lookup Transformation

- Now the Similarity Threshold is kept close to 1



Fuzzy Lookup Transformation

- Only 1 Columns with the Similarity found is been loaded in Target Table with the “fname and Desc”. Rest Columns are populated with NULL.

	fname	desc
1	amit	NULL
2	kaushal	NULL
3	amita	b123
4	abhinav	NULL
5	aditya	NULL
6	kailash	NULL

```
AS4Test_DB>SELECT * FROM test_needsdes_123
```

fname	desc
amit	NULL
NULL	NULL
NULL	b123
NULL	NULL
NULL	NULL
NULL	NULL

Import Column Transformation

Lesson 19:

Import Column Transformation

- Import Column Transformation reads the data from files and adds the data to columns in a data flow. Using this Transformation, a package can add text and images stored in separate files to a data flow.
- For example, a data flow loading data into table that stores product information can include Import Column Transformation to import customer reviews of each product from files and add the reviews to data flow.



Copyright © Capgemini 2015. All Rights Reserved 42

Import Column Transformation

Example::

- Load the records for the field 'PLANNAME' from the following path mentioned in the output data set by using Import Column Transformation.

	POLICYNUMBER	CARRIERCODE	POLICYSTATUS	PAYMENTMETH...	CARRI...	PLANNAME
	POL1	CAR3	3	4	56	D:\dipankar\new.txt
	POL1	CAR1	4	889	22	D:\dipankar\new1.txt
	POL2	CAR2	5	66	NULL	D:\dipankar\new2.txt
	NULL	NULL	NULL	NULL	NULL	NULL



Import Column Transformation

Output::

- Records from the field 'PLANNNAME' is successfully loaded in the output dataset
- after Import Column Transformation.

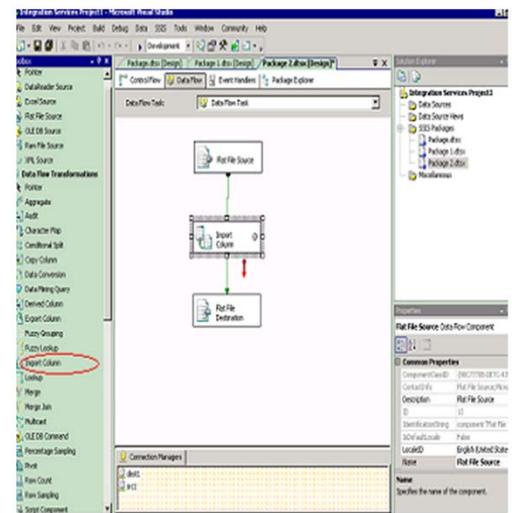
POLICYNUMBER	CARRIERCODE	POLICYSTATUS	PLANNNAME	IMPORTED CO...
POL1	CAR3	3	D:\dipankar\ne...	POL4
POL1	CAR1	4	D:\dipankar\ne...	POL5
POL2	CAR2	5	D:\dipankar\ne...	POL6



Copyright © Capgemini 2015. All Rights Reserved 44

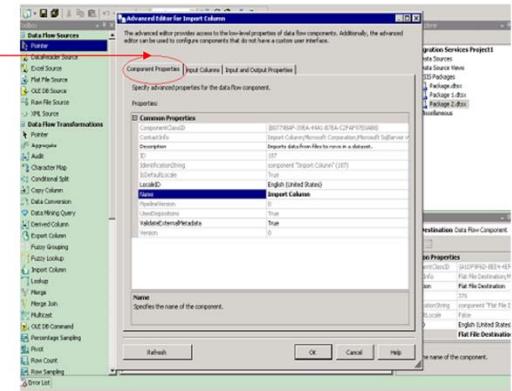
Import Column Transformation

- Click the Data Flow tab, and then, from the Toolbox, drag the Import Column transformation to the design surface
- Link it with FlatFile source.
- Double click on Import Column Transformation to open Advanced Editor for Import Column.



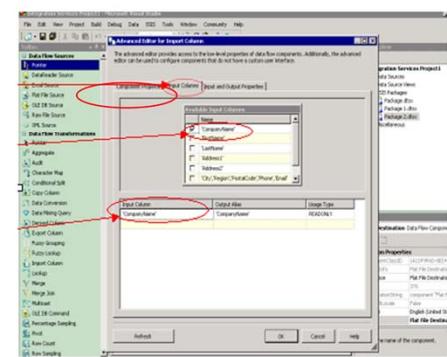
Import Column Transformation

- Component Properties:
- This tab is used for specifying advanced properties of Data Flow Component such as Name of component, Description, Locale ID.



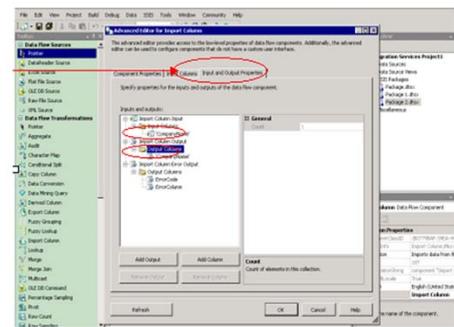
Import Column Transformation

- Input Columns:
 - One of the Columns of source contains the path of a file (from which data is to be Imported).
 - This column is to be selected under Input Columns tab.
 - When the selected column is displayed, one may alter its usage type (Read Only / Read-Write).



Import Column Transformation

- Input & Output Properties:
- Firstly, create a new column (under Import Column Output section).
- Select the datatype of this column as Text/Image.
- The ID of this column is to be mentioned against "FileDataColumnID" property for the selected Input Column (in the previous tab) under Import Columns Input.



Lookup Transformation

Lesson 20:

Lookup Transformation

“The Lookup transformation performs lookups by joining data in input columns with columns in a reference dataset”.

NOTE :

- The reference dataset can be an existing table or view or a new table, or the result of an SQL statement. The Lookup transformation uses an OLE DB connection manager to connect to the database that contains the data that is the source of the reference dataset.
- The Lookup transformation performs an equi-join between values in the transformation input and values in the reference dataset. If there is no matching entry in the reference dataset, no join occurs and no values are returned from the reference dataset. This is an error, and the transformation fails, unless it is configured to ignore errors or redirect error rows to the error output.
- The lookups performed by the Lookup transformation are case sensitive.



Copyright © Capgemini 2015. All Rights Reserved 50

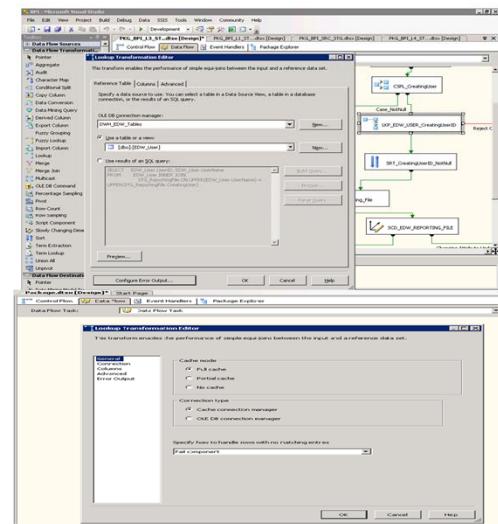
Lookup Transformation

- Typically, values from the reference dataset are added to the transformation output. The values from the reference table can replace column values or can be added to new columns.
- Caching SQL Statement:
“The Lookup transformation can be configured to use a caching SQL statement to select a subset of the reference dataset, thus, limiting its size”.
- EXAMPLE
 - Suppose, Our source consist of CITY_CODE and in target we want CITY_NAME instead. We have a master table consisting of code and name of cities, This transformation take city code as input, And lookup on master table, matches the code and return the corresponding Name of the city.



Lookup Transformation

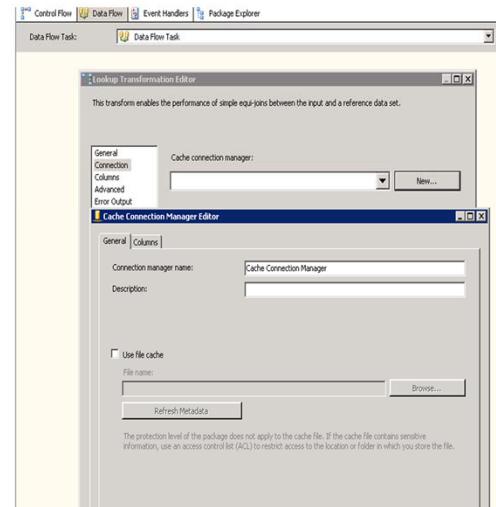
- Click the Data Flow tab, and then, from the Toolbox, drag the Lookup transformation to the workspace area.
- Connect the Lookup transformation to the data flow by dragging a connector—the green or red arrow—from a source or a previous transformation to the Lookup transformation.
- Double-click the Lookup transformation component to open the Editor.
- In the General tab, select the Cache Mode and the Connection Type.



Copyright © Capgemini 2015. All Rights Reserved 52

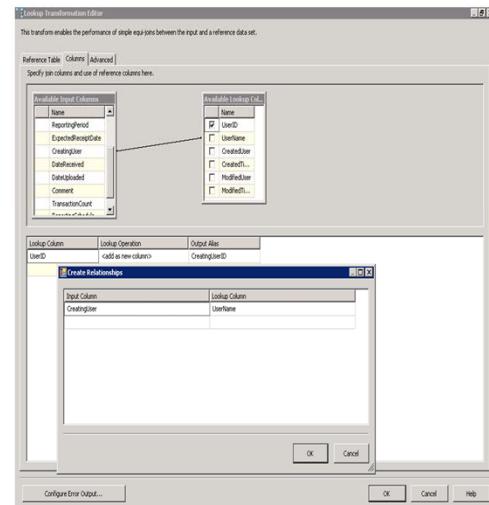
Lookup Transformation

- Select Cache Connection Manager to select the required Connection Manager Name and the file name.



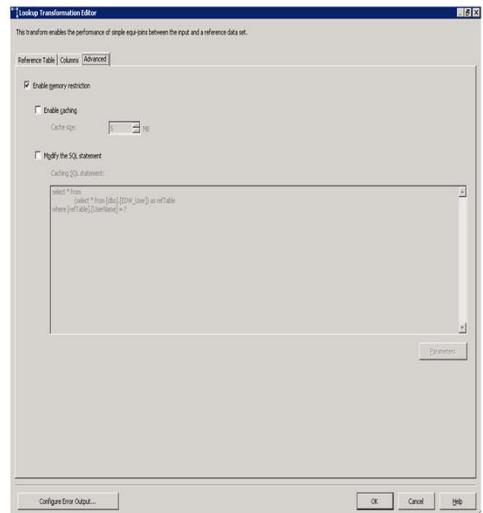
Lookup Transformation

- Select Columns tab to create relationships between input fields for Lookup transformation and Reference table fields.
- Right-click in the space between “Available Input columns” & “Available lookup columns” to open the dialog box for creating the relationships.
- Once relationships are created, close the dialog box (Click OK) and then include the list of columns to be fetched incase a matching entry is found from the reference table.



Lookup Transformation

- Advanced tab has to be selected if one need to mention Caching SQL Statement by enabling Memory restriction and / or Caching
- Click OK to complete lookup transformation



Lookup Transformation

- Source Input

EMPLOYEE		
First Name	Last Name	City Code
saurabh	ganguly	452001
sachin	tandulkar	400045
rahul	dravid	300016

- Lookup Table

CITY MASTER	
City Code	City Name
452001	Kalkattha
400045	Bombay
300016	Pune

- Lookup Output

First Name	Last Name	City Code	Lkp_City Name
saurabh	ganguly	452001	Kalkattha
sachin	tandulkar	400045	Bombay
rahul	dravid	300016	Pune

Merge Transformation

Lesson 21:

Merge Transformation

“The Merge transformation combines two sorted datasets into a single dataset. The row from each datasets are inserted into the output based on values in their key columns.”

By including Merge transformation into the dataflow following tasks can be performed :

- Merge data from two data sources, such as tables and files.
- Create complex datasets by nesting Merge transformations.
- Remerge rows after correcting errors in the data.
- The Union All transformation is used instead of Merge transformation in the following situation :
 - The transformation inputs are not sorted.
 - The combined output does not need to be sorted.
 - The transformation has more than two inputs.



Copyright © Capgemini 2015. All Rights Reserved 58

Merge Transformation

- EXAMPLE: The data coming from two different databases is as shown below.

Input Data Source 1

POLICYNUMBER	CARRIERCODE	POLICYST...	PAYMENTM...	CARRIERPA...	PLANNAME	LINEOFBU...
POL2	CAR2	5	NULL	NULL	NULL	1
POL1	CAR2	567	25	NULL	NULL	1
POL1	CAR1	1	NULL	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL	NULL

Input Data Source 2

POLICNU...	CARRIERC...	POLICYST...	PAYMENT...	CARRIERP...	PLANNAME	LINEOFBU...
POL1	CAR1	8	NULL	NULL	NULL	6
POL1	CAR1	1	NULL	NULL	NULL	NULL
POL2	CAR2	5	NULL	NULL	NULL	1
POL7	CAR2	567	25	NULL	NULL	1
POL1	CAR1	1	NULL	NULL	NULL	NULL

- The data is sorted based on PolicyNumber & CarrierCode Columns using Sort transformation. This Sorted data is given as a input to merge transformation.



Copyright © Capgemini 2015. All Rights Reserved 59

Merge Transformation

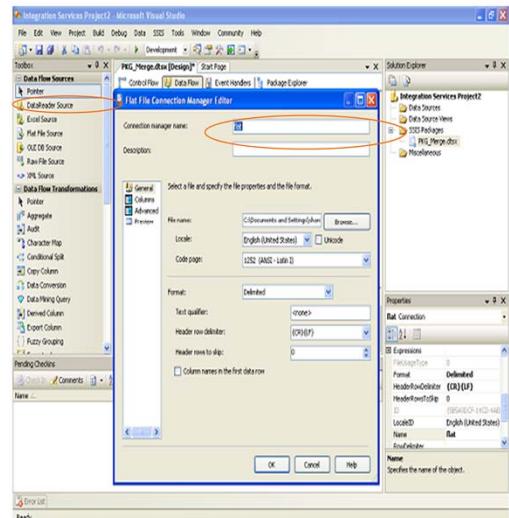
- The output of the Merge Transformation is as shown. The Merging of records is done on the basis of 'POLICYNUMBER' & 'CARRIERCODE' which are used as key columns.

POLICYNUMBER	CARRIERCODE	POLICYSTATUS	PAYMENTMETH...	CARRIERPART...	LINEOFBUSINESS
POL1	CAR1	8	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL
POL1	CAR1	1	NULL	NULL	NULL
POL1	CAR2	567	25	NULL	NULL
POL2	CAR2	5	NULL	NULL	1
POL2	CAR2	5	NULL	NULL	NULL
POL7	CAR2	567	25	NULL	1



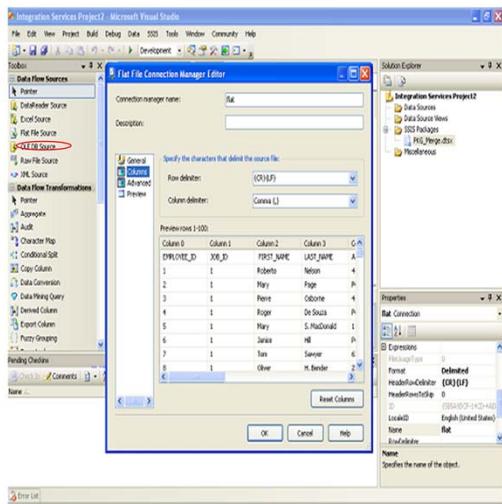
Merge Transformation

- Drag the 'Flat File Source' from Data Flow Task Navigator from Data Flow Task Window.
- Double click or Right click on 'Flat File Source' to open 'Flat File Source Editor'
- Set the 'Connection Manager' by giving appropriate information, e.g. Connection Manager Name, File Name.



Merge Transformation

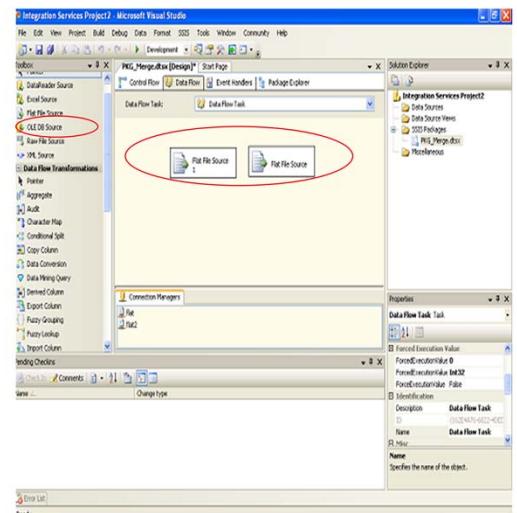
- Select Columns and check required 'Column'.
- Click 'OK'



Copyright © Capgemini 2015. All Rights Reserved 62

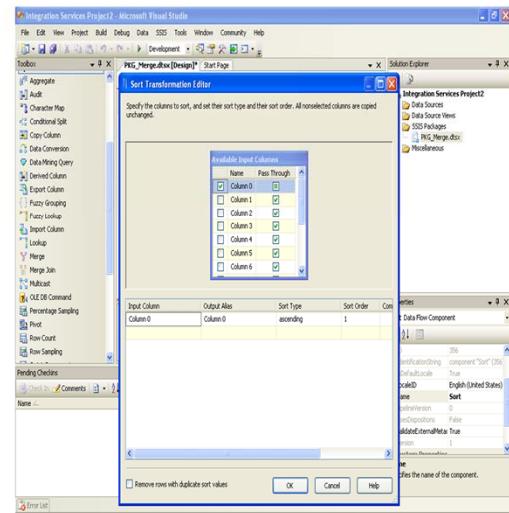
Merge Transformation

- Drag another 'Flat File Source' from Data Flow Task Navigator into Data Flow Task Window.
- Create a Flat File Connection as explained in the previous slide.



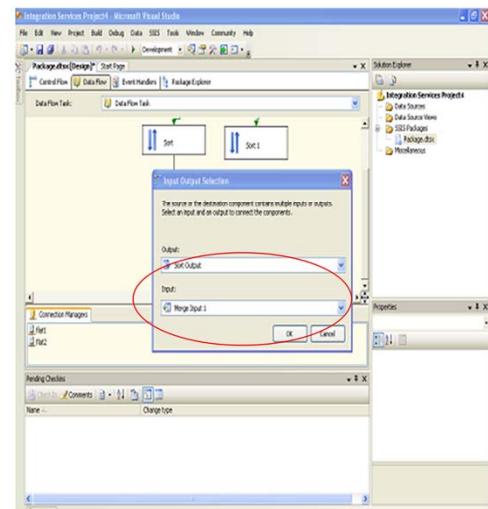
Merge Transformation

- Drag a ‘Sort Transformation’ from the Data Flow Transformation into the Data Flow Task Window.
- Connect the output of one Flat File Source to the Sort Transformation.
- Double click or Right click the Sort Transformation to enter in Sort Transformation Editor.
- Put a check on the right check box of the field according to which sorting is to be done.
- Click OK.



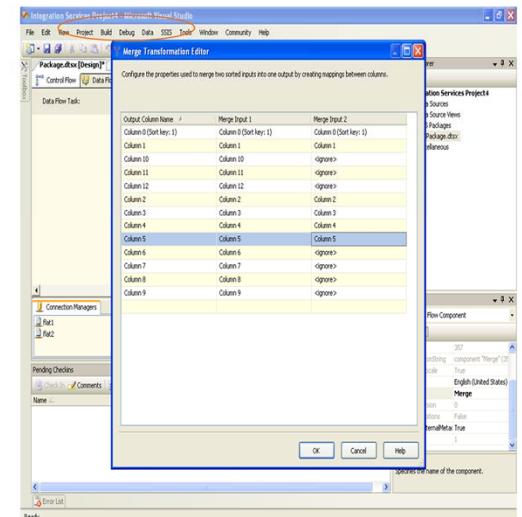
Merge Transformation

- Drag a ‘Merge Transformation’ from the Data Flow Transformation into the Data Flow Task Window.
- Connect the Output of one Sort Transformation to the Merge Transformation, an Input Output Selection Window will be opened. Select an Input and an Output to connect the components.
- Similarly connect the output of the other Sort Transformation to the Merge Transformation.



Merge Transformation

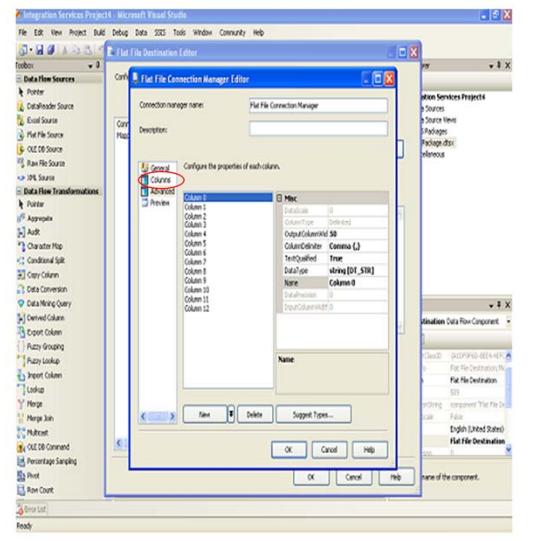
- Double Click or Right Click on the Merge Transformation to enter into the 'Merge Transformation Editor'.
- In Merge Transformation Editor one can configure the properties to merge two sorted inputs into one output by creating mappings between the columns.
- Click OK.



Copyright © Capgemini 2015. All Rights Reserved 66

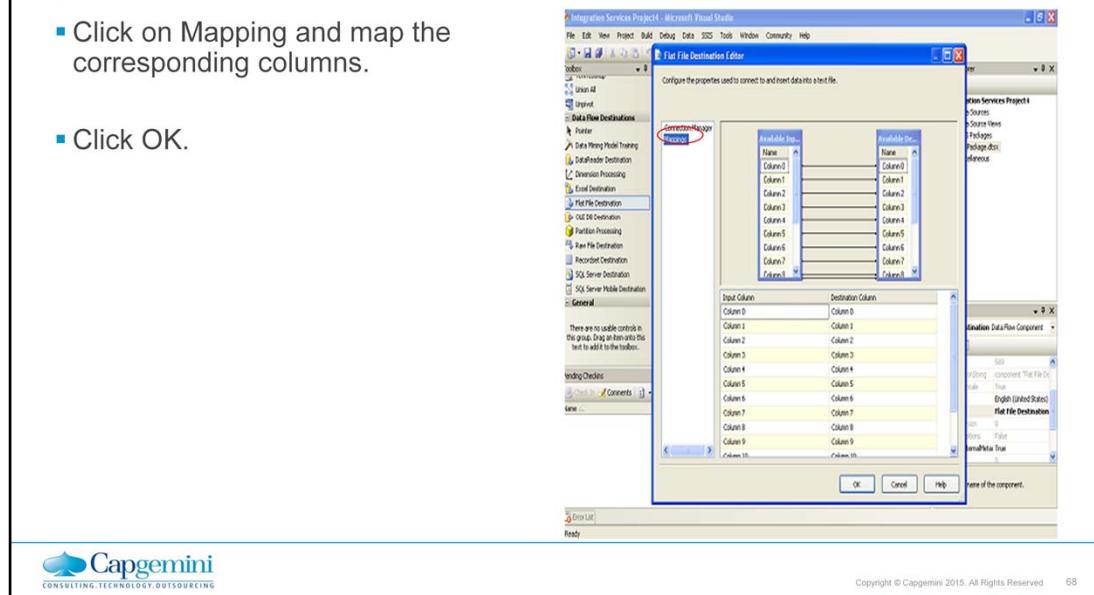
Merge Transformation

- Drag the 'Flat File Destination' from the Data Flow Destination into Data Flow task Window.
- Double Click or Right Click the Flat File Destination to enter into the 'Flat File Destination Editor'.
- Click on 'Connection Manager' and create connection as described before. Click on Columns to specify the number of columns.
- Click OK.

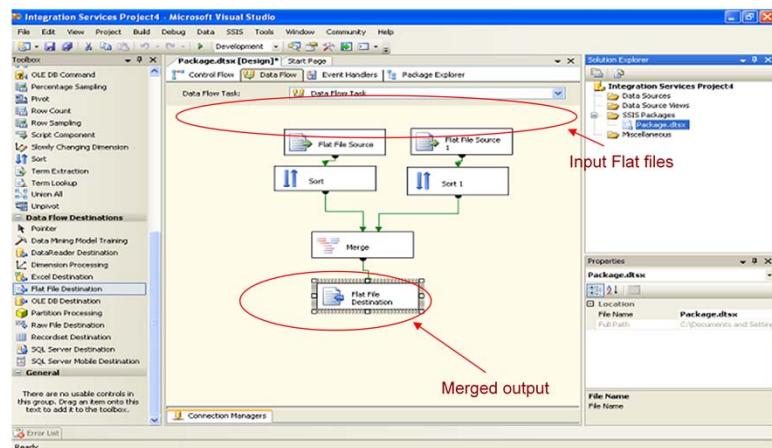


Merge Transformation

- Click on Mapping and map the corresponding columns.
- Click OK.



Merge Transformation



Merge Join Transformation

Lesson 22:

Merge Join Transformation

- EXAMPLE: Following are the two different data sources named as Source1 & Source2 respectively. Here we are using Inner Join with the help of Merge Join Transformation.

POLICYNUMBER	CARRIERCODE	POLICYSSTATUS	PAYMENTMETHOD
POL1	CAR3	3	4
POL1	CAR1	4	889
POL2	CAR2	5	66

POLICYNUMBER	CARRIERCODE	POLICYSSTATUS	PAYMENTMETHOD
POL1	CAR1	8	545
POL3	CAR4	44	555
POL2	CAR2	5	NULL

- The Table shown below represents output when Inner Join is adopted for POLICYNUMBER & CARRIERCODE fields.

POLICYNUMBER	CARRIERCODE	POLICYSSTATUS	POLICYSSTATUS (1)	PAYMENTMETH...	PAYMENTMETHOD (1)
POL1	CAR1	4	8	889	545
POL2	CAR2	5	5	66	

Merge Join Transformation

- This transformation has two inputs and one output. It does not support an error output.

You can configure the Merge Join transformation in the following ways:

- Specify the join is a FULL, LEFT, or INNER join.
- Specify the columns the join uses.
- Specify whether the transformation handles null values as equal to other nulls



Copyright © Capgemini 2015. All Rights Reserved 72

Merge Join Transformation

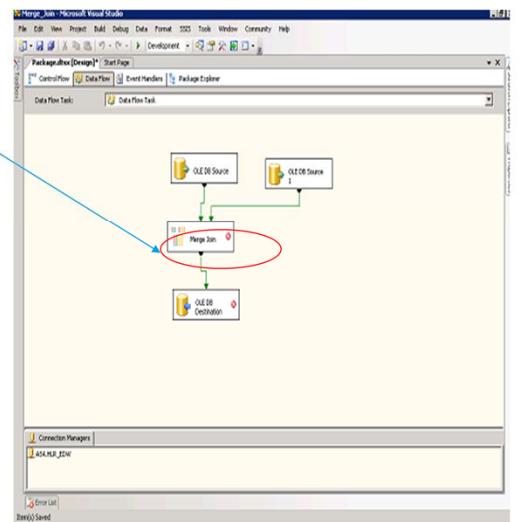
- In Business Intelligence Development Studio, open the Integration Services project that contains the package you want.
- In Solution Explorer, double-click the package to open it.
- Click the Data Flow tab, and then, from the Toolbox, drag the Merge Join transformation to the design surface.
- Connect the Merge Join transformation to the data flow by dragging the connector—the green or red arrow—from a data source or a previous transformation to the Merge Join transformation.
- Double-click the Merge Join transformation.
- In the Merge Join Transformation Editor dialog box, select the type of join to use in the Join type list.



Copyright © Capgemini 2015. All Rights Reserved 73

Merge Join Transformation

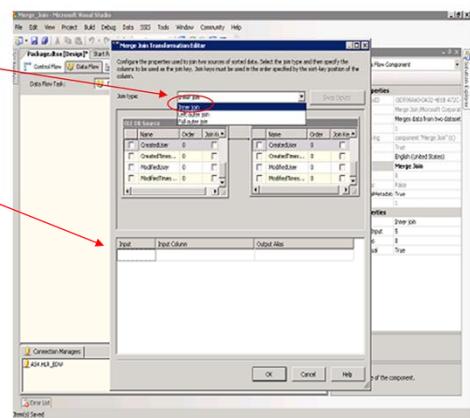
- Configure the Is Sorted Property of Source Transformation to TRUE.
- Double Click on the Merge Join Transformation to Open.



Copyright © Capgemini 2015. All Rights Reserved 74

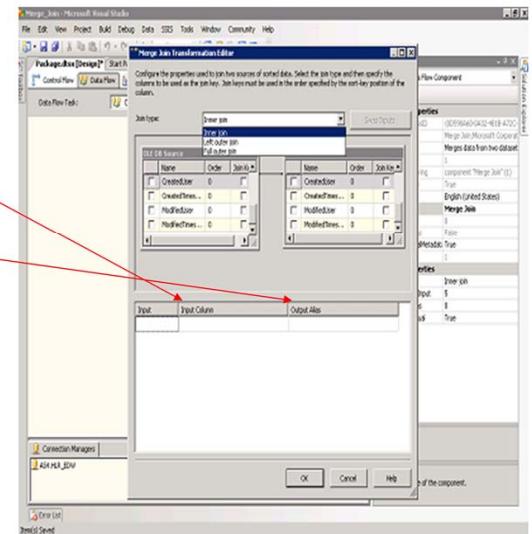
Merge Join Transformation

- Join type
- Specify whether you want to use an inner join, left outer join, or full join.
- Swap Inputs
- Switch the order between inputs by using the Swap Inputs button. This selection may be useful with the Left outer join option.
- Input
- For each column that you want in the merged output, first select from the list of available inputs.
- Inputs are displayed above in two separate tables. Select columns to include in the output. Drag columns to create a join between the tables.



Merge Join Transformation

- Input Column
- Select a column to include in the merged output from the list of available columns on the selected input.
- Output Alias
- Type an alias for each output column. The default is the name of the input column; however, you can choose any unique, descriptive name.



Merge Join Transformation

- Drag columns in the left input to columns in the right input to specify the join columns. If the columns have the same name, you can select the Join Key check box and the Merge Join transformation automatically creates the join.
- In the left and right inputs, select the check boxes of additional columns to include in the output. Join columns are included by default.
- Optionally, update the names of output columns in the Output Alias column.
- Click OK.
- To save the updated package, click Save Selected Items on the File menu.



Copyright © Capgemini 2015. All Rights Reserved 77

Merge Join Transformation

- The Merge Join transformation provides an output that is generated by joining two sorted datasets using a FULL, LEFT, or INNER join.
- The Merge Join transformation requires that both inputs be sorted and that the joined columns have matching metadata. For example, you cannot join a column that has a numeric data type with a column that has a character data type. If the data has a string data type, the length of the column in the second input must be less than or equal to the length of the column in the first input with which it is merged.
- The data can be sorted by setting sort options on the outputs of the source or other upstream data flow components, or by inserting a Sort transformation into the data flow before the Merge Join transformation. If the sort options indicate that the data is sorted, but the data is not actually sorted, the merge join operation may have unpredictable results.



Copyright © Capgemini 2015. All Rights Reserved 78

Multicast Transformation

Lesson 23:

Multicast Transformation

“The Multicast transformation creates copies of dataset. This transformation distributes its input to one or more output”

- Using the Multicast transformation, a package can create logical copies of data.
- This capability is useful when the package needs to apply multiple sets of transformations to the same data.
- This transformation has one input and multiple outputs. It does not support an error output.

EXAMPLE

- One copy of the data is aggregated and only the summary information is loaded into its destination, While another copy of the data is extended with lookup values and derived columns before it is loaded into its destination.



Copyright © Capgemini 2015. All Rights Reserved 80

Multicast Transformation

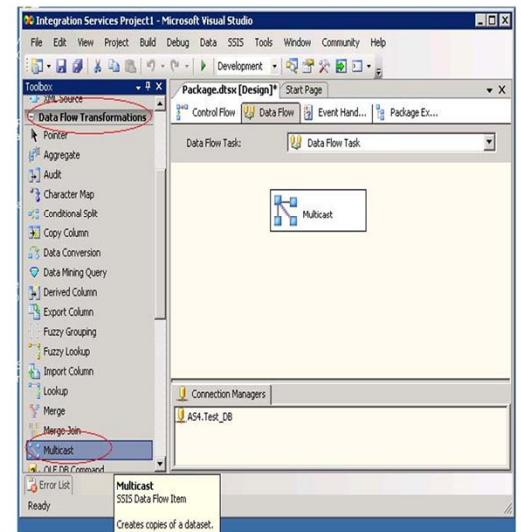
- To configure Multicast Transformation, you can set properties through SSIS Designer or programmatically.
- You configure the Multicast transformation by adding outputs.
- The Multicast transformation distributes its input to one or more outputs.
- This transformation is similar to the Conditional Split transformation. Both transformations direct an input to multiple outputs.
- The difference between the two is that the Multicast transformation directs every row to every output, and the Conditional Split directs a row to a single output. For more information, see Conditional Split Transformation.



Copyright © Capgemini 2015. All Rights Reserved 81

Multicast Transformation

- Drag a 'Multicast' transformation from Data Flow Transformations Navigator to Data Flow Task Window.



Multicast Transformation

- You can now add as many destinations as you want by dragging each from the Data Flow Destinations Pane.

