

# *Metody przetwarzania dużej ilości danych - projekt*

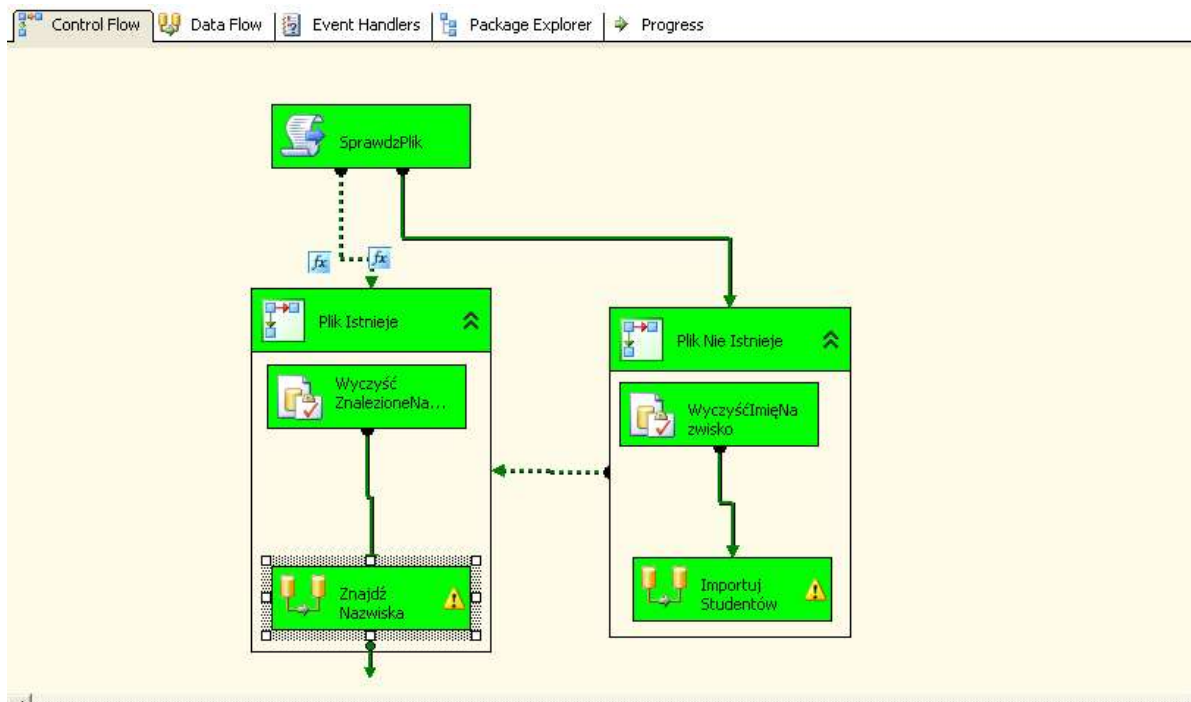
## *Sprawozdanie nr 2*

Adam Hyjek 234987, Bartosz Rodziewicz 226105

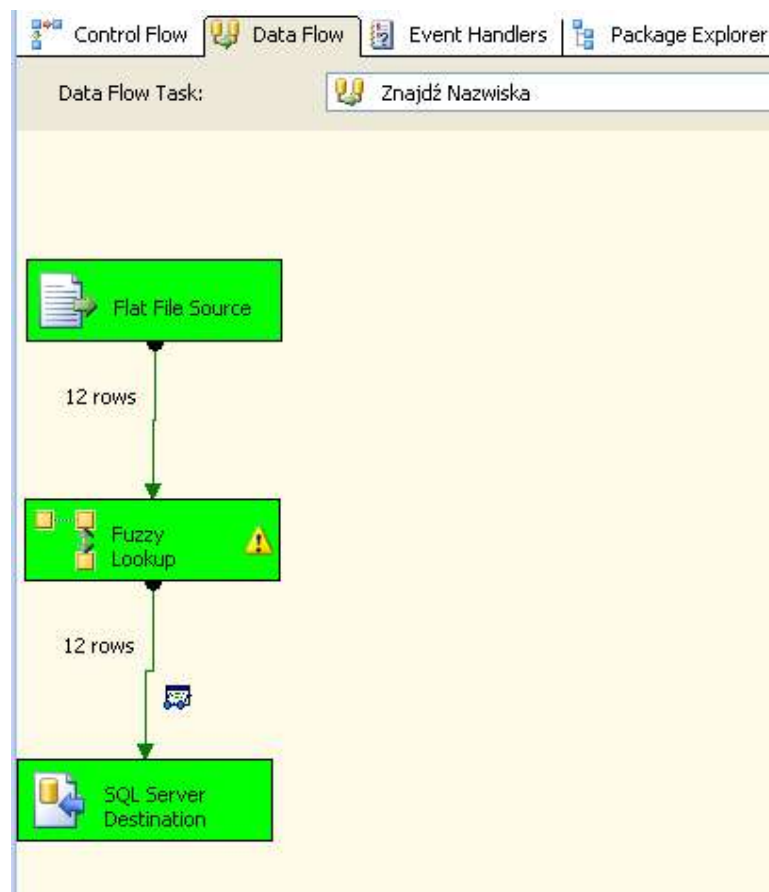
3 listopada 2020

Ścieżka 1 (4.5)

W ramach drugiej instrukcji kontynuowano projekt rozpoczęty podczas poprzedniej instrukcji. Do diagramu Control Flow dodano obiekt Script Task SprawdzPlik wraz ze skryptem w języku Visual Basic sprawdzającym czy plik StudenciDestination istnieje oraz zapisującym wynik operacji do zmiennej PlikIstnieje. Na podstawie zmiennej PlikIstnieje wykonuje się jeden z kontenerów sekwencji: Plik Istnieje lub Plik Nie Istnieje. Plik Istnieje dokonuje przeszukiwania rozmytego, natomiast Plik Nie Istnieje tworzy plik StudenciDestination (instrukcja nr 1). Przeszukiwanie rozmyte zostało zaimplementowane za pomocą Data Flow Task zawierającego Flat File Source, Fuzzy Lookup i SQL Server Destination. Flat File Source wczytuje dane z pliku InputA za pomocą menadżera połączeń Wprowadź nazwiska. SQL Server Destination służy do utworzenia w bazie danych Studenci tabeli ZnalezioneNazwiska, w której zapisano wyniki wyszukiwania rozmytego, do którego wykorzystano komponent Fuzzy Lookup i nowo utworzony indeks na tabeli ImieNazwisko. Kolejne kroki i rezultaty przedstawiono na zrzutach ekranowych:



Rysunek 1: Control Flow



Rysunek 2: Data Flow Znajdź Nazwiska

**Flat File Connection Manager Editor**

Connection manager name:

Description:

**General**  
Columns  
Advanced  
Preview

Select a file and specify the file properties and the file format.

File name:

Locale:  ☐ Unicode

Code page:

Format:

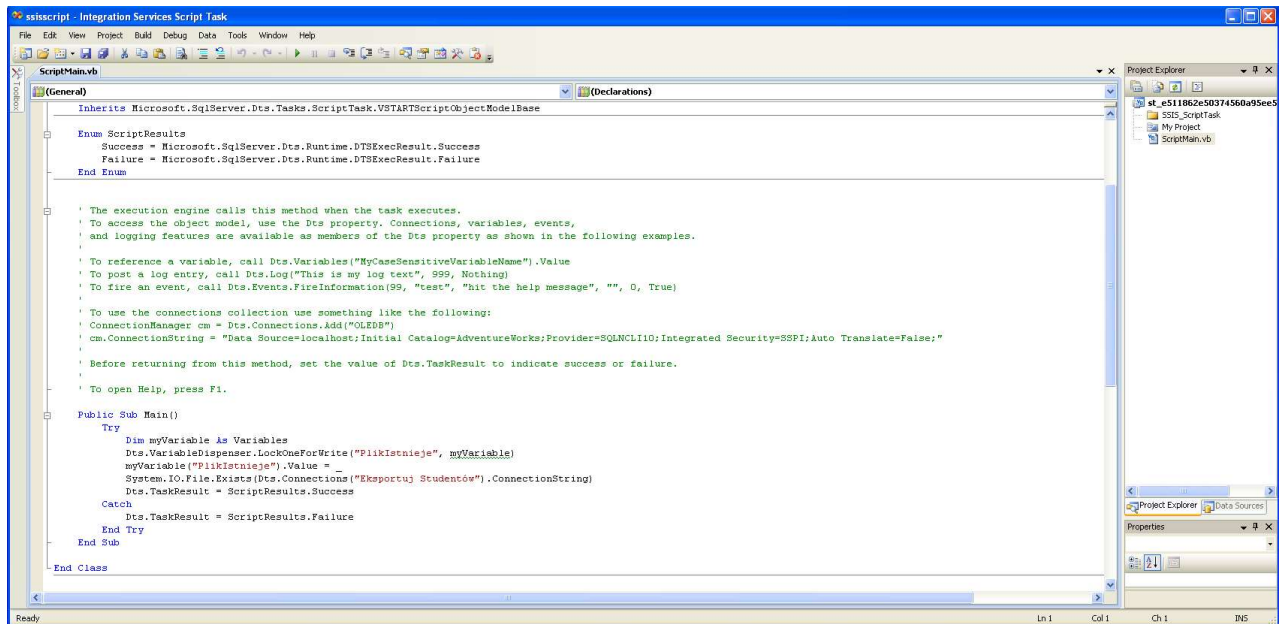
Text qualifier:

Header row delimiter:

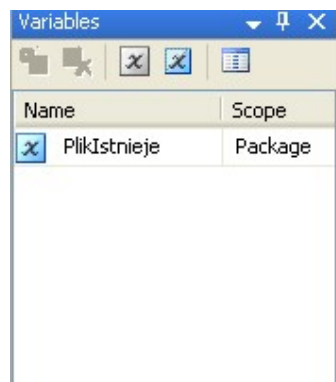
Header rows to skip:

☐ Column names in the first data row

Rysunek 3: Menadżer Połączeń Wprowadź Nazwiska



Rysunek 4: Skrypt z SprawdzPlik



Rysunek 5: Zmienna PlikIstnieje

**Precedence Constraint Editor**

A precedence constraint defines the workflow between two executables. The precedence constraint can be based on a combination of the execution results and the evaluation of expressions.

**Constraint options**

Evaluation operation: Expression and Constraint

Value: Success

Expression: @PlikIstnieje == False Test

**Multiple constraints**

If the constrained task has multiple constraints, you can choose how the constraints interoperate to control the execution of the constrained task.

☒ Logical AND. All constraints must evaluate to True

☐ Logical OR. One constraint must evaluate to True

OK Cancel Help

Rysunek 6: Połączenie pomiędzy obiektami SprawdźPlik i Plik Nie Istnieje

**Precedence Constraint Editor**

A precedence constraint defines the workflow between two executables. The precedence constraint can be based on a combination of the execution results and the evaluation of expressions.

**Constraint options**

Evaluation operation: Expression and Constraint

Value: Success

Expression: @PlikIstnieje == True Test

**Multiple constraints**

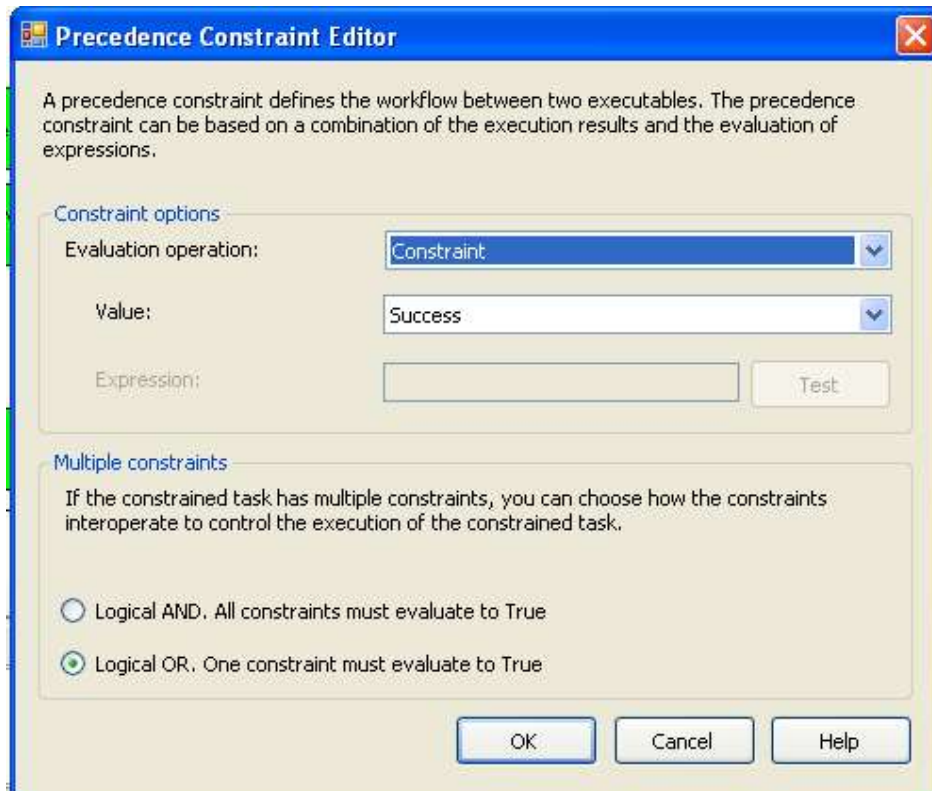
If the constrained task has multiple constraints, you can choose how the constraints interoperate to control the execution of the constrained task.

☐ Logical AND. All constraints must evaluate to True

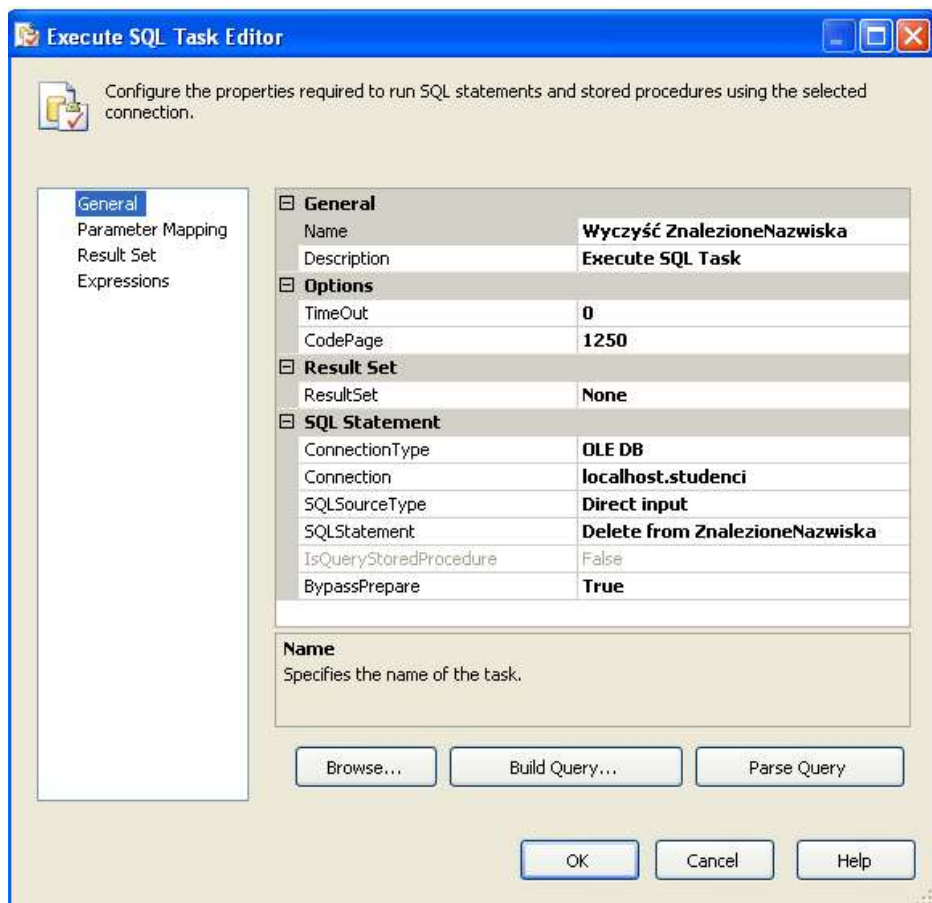
☒ Logical OR. One constraint must evaluate to True

OK Cancel Help

Rysunek 7: Połączenie pomiędzy obiektami SprawdźPlik i Plik Istnieje



Rysunek 8: Połączenie pomiędzy obiektami Plik Nie Istnieje i Plik Istnieje



Rysunek 9: SQL Task Wyczyść ZnalezioneNazwiska

**Fuzzy Lookup Transformation Editor**

Configure the properties used to perform a lookup operation between an input dataset and a reference dataset using a best-match algorithm.

Reference Table Columns Advanced

Specify the connection manager to the reference table and the options for the index that the transformation uses.

Temporary objects will be created using the specified connection. The space required for the temporary objects is proportional to the reference table, but may be greater. The table maintenance feature requires the installation of a trigger on the reference table.

OLE DB connection manager:

localhost.studenci New...

☒ Generate new index

Reference table name:

[dbo].[ImieNazwisko]

☐ Store new index

New index name:

☐ Maintain stored index

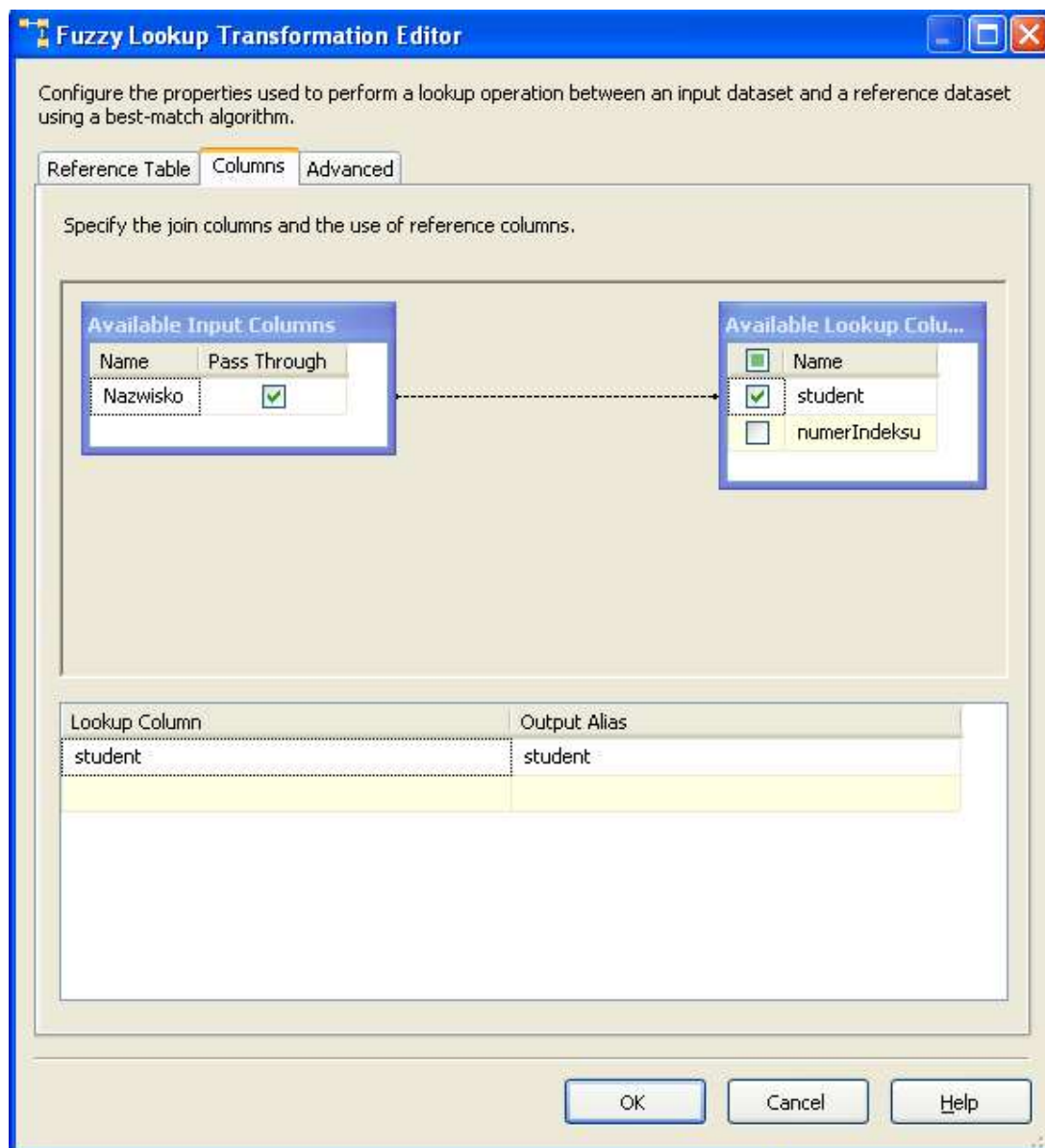
☐ Use existing index

Name of an existing index:

OK Cancel Help

Rysunek 10: Wyszukiwanie rozmyte: Reference Table

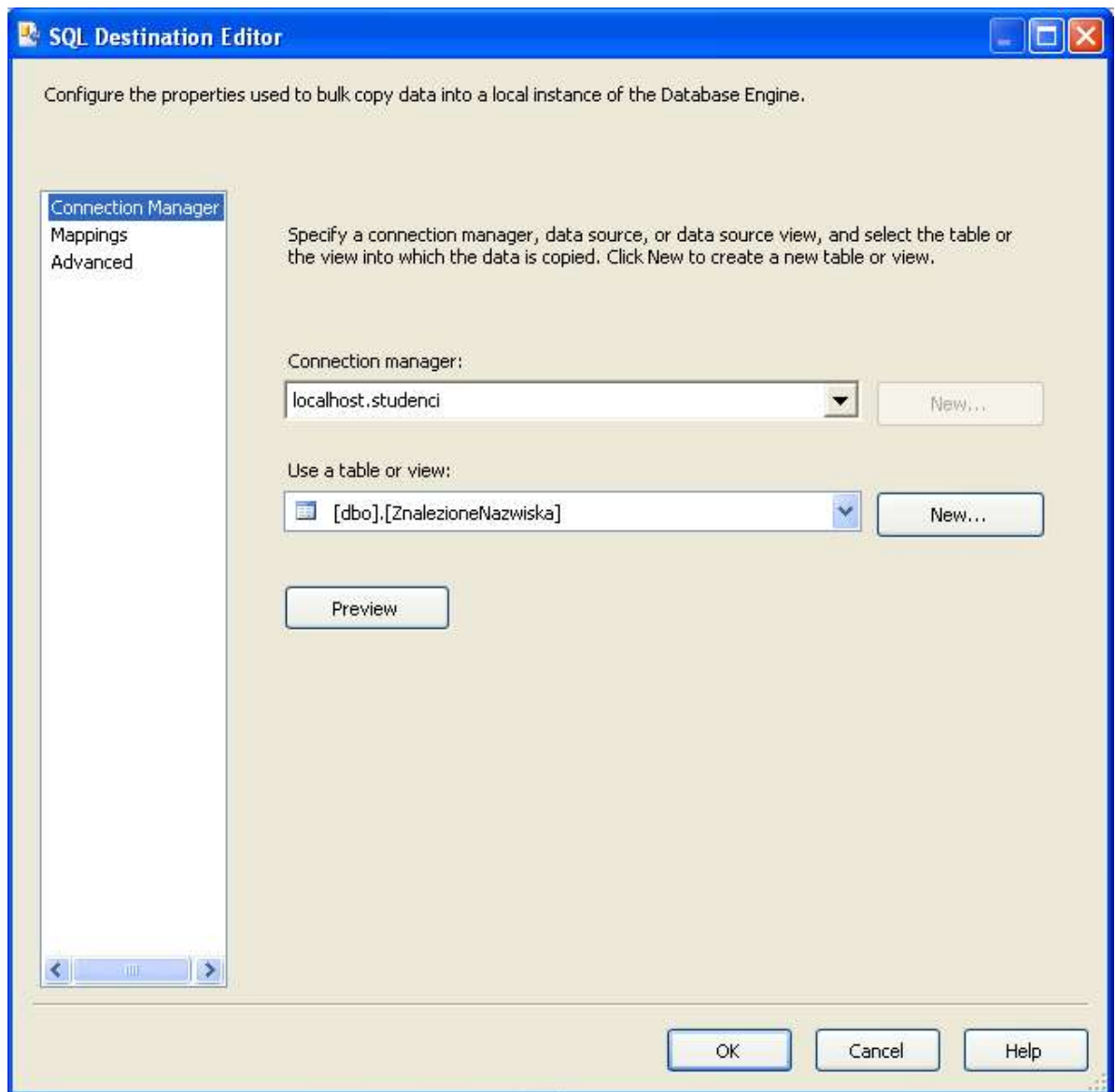




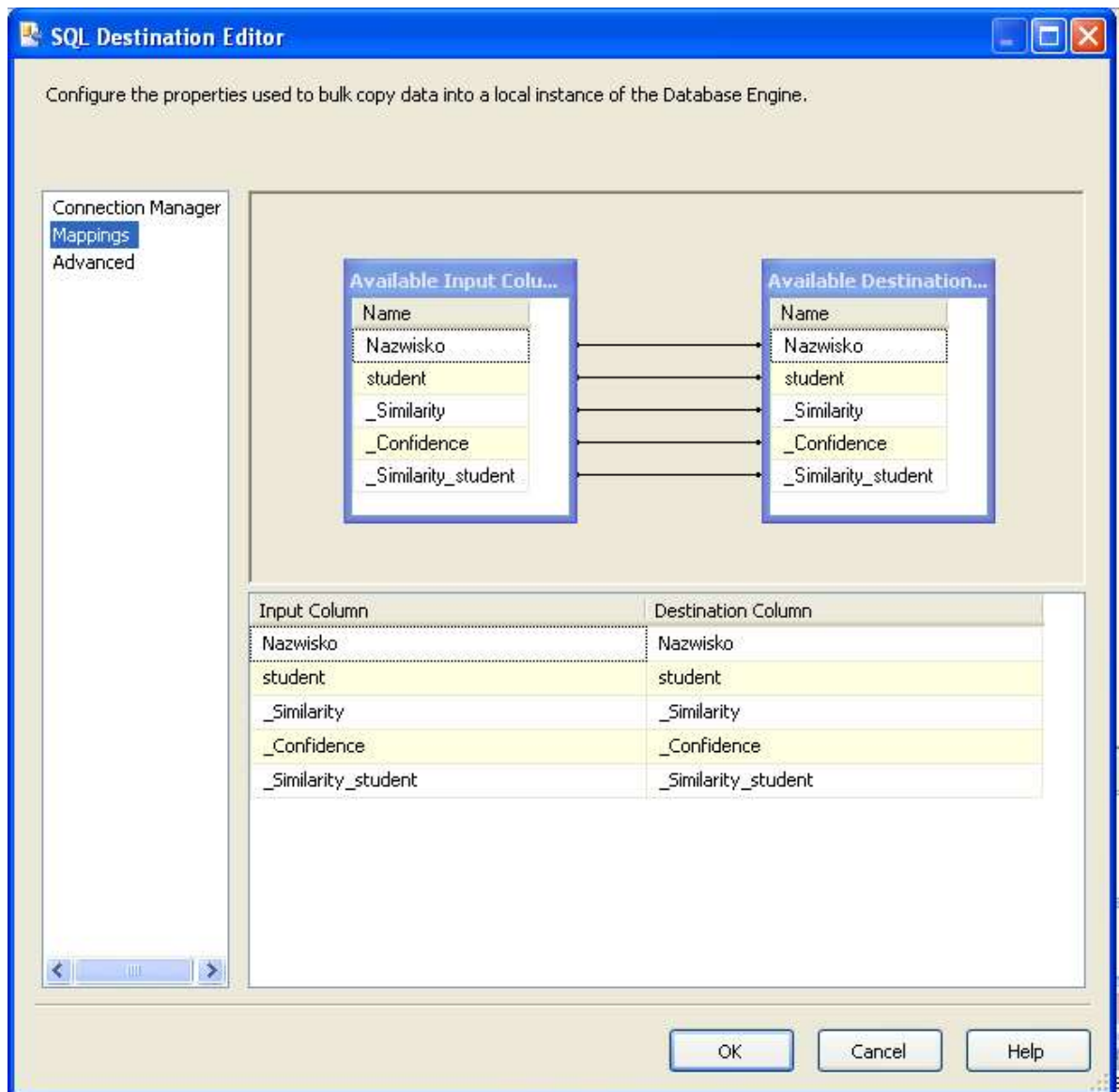
Lookup Column	Output Alias
student	student

Rysunek 11: Wyszukiwanie rozmyte: Columns

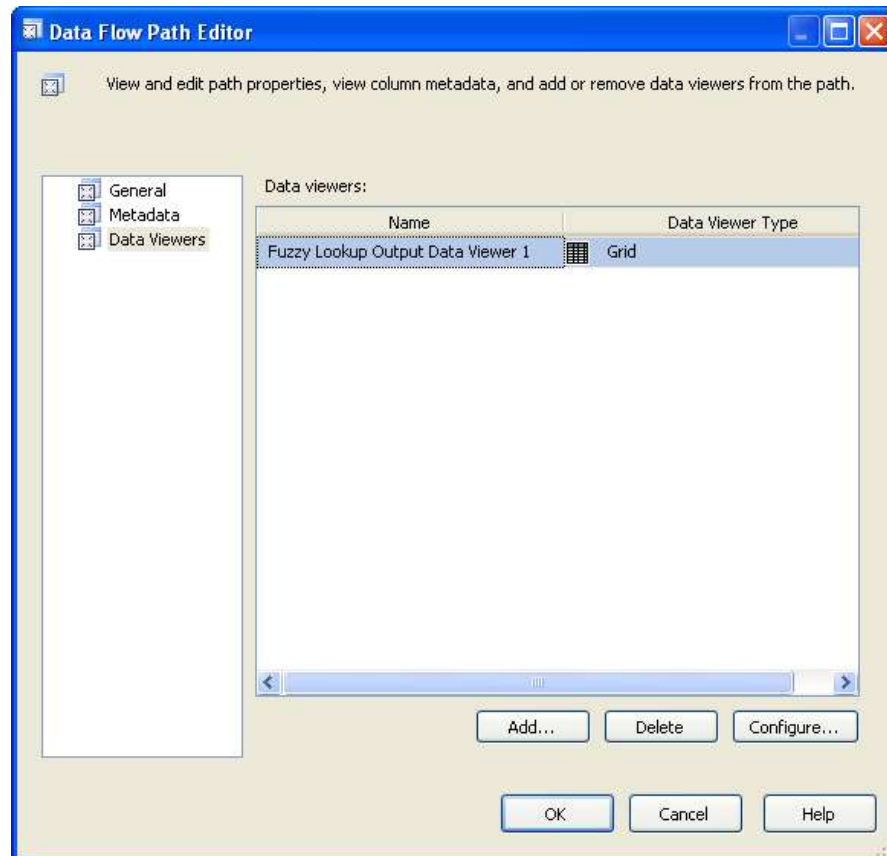




Rysunek 12: SQL Server Destination: Connection Manager



Rysunek 13: SQL Server Destination: Mappings



Rysunek 14: Ustawienie Data Viewera

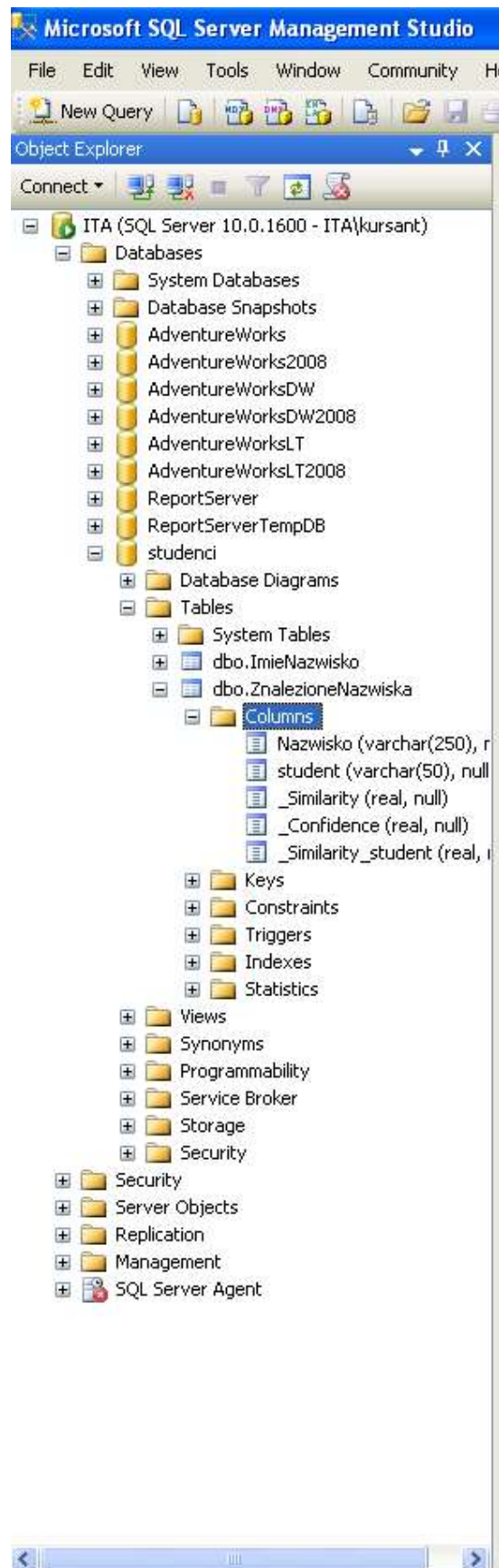
Fuzzy Lookup Output Data Viewer 1 at Fuzzy Lookup.Fuzzy Lookup Output

Detach Copy Data

Nazwisko	student	_Si...	_Co...	_Si...
Adam Radomski	Adam Radomski	1	1	1
A Radomski	Adam Radomski	0,...	0,...	0,...
A Raaomski	Adam Radomski	0,...	0,...	0,...
Aleksander Wukara	Aleksander Wukara	1	1	1
Alek Wukara	Aleksander Wukara	0,...	0,...	0,...
A Wukara	Aleksander Wukara	0,...	0,...	0,...
Artur Trulik	Artur Trulik	1	1	1
A Tulik	Artur Trulik	0,...	0,...	0,...
Art Trulik	Artur Trulik	0,...	0,...	0,...
Arkadiusz Dorzyński	Arkadiusz Dorzyński	1	1	1
Arek Dorzyński	Arkadiusz Dorzyński	0,...	0,...	0,...
A Dobrzyński	Arkadiusz Dorzyński	0,...	0,...	0,...

Attached Total rows: 12, buffers: 1 Rows displayed = 12

Rysunek 15: Wynik wyszukiwania rozmytego: Data Viewer



Rysunek 16: Nowa tabela w bazie danych