

Integracja systemów

Laboratorium 5
Prowadzący: Marek Kowal
(M.Kowal@issi.uz.zgora.pl)

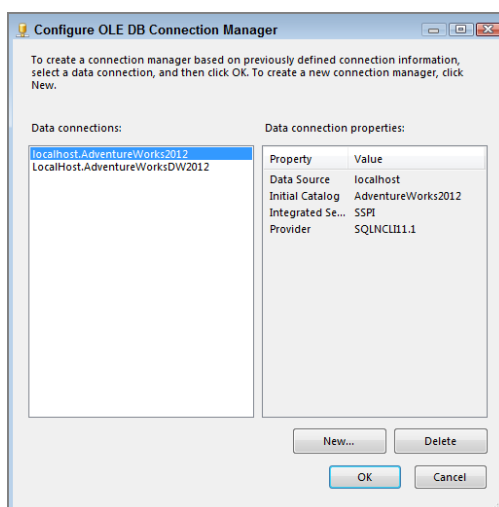
Data Flow Task

1. Cel ćwiczenia

Celem ćwiczenia jest utworzenie pakietu wykorzystującego zadanie Data Flow Task. Zadaniem pakietu będzie skopiowanie danych z tabeli pochodzącej z bazy SQL do pliku płaskiego.

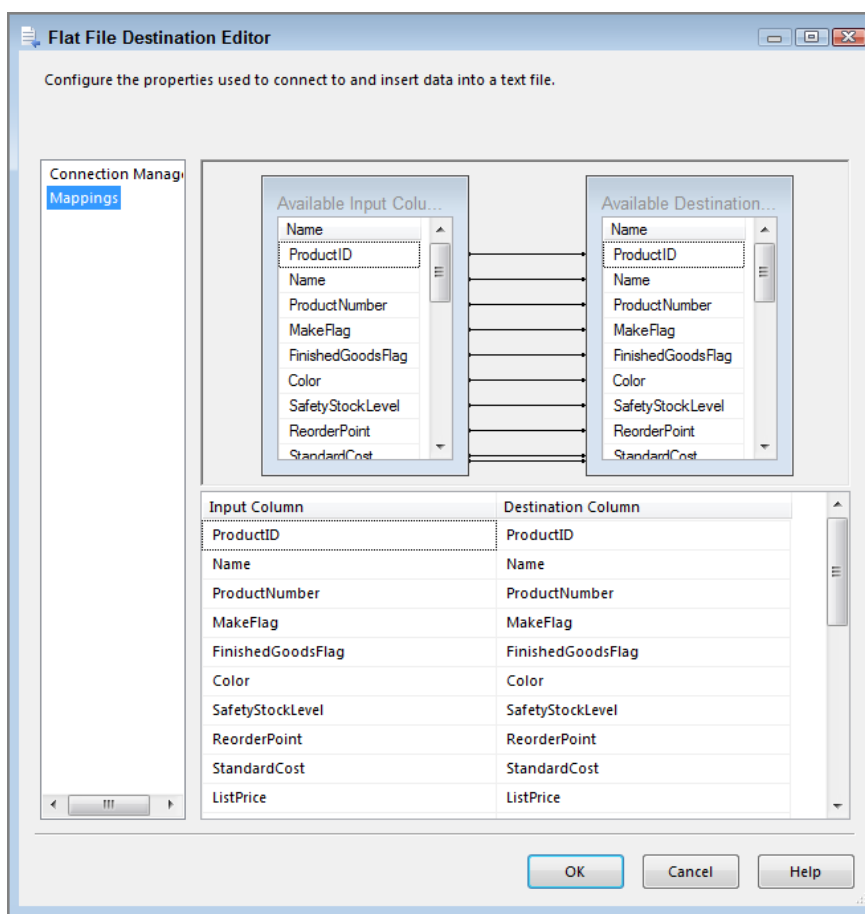
2. Przebieg ćwiczenia

- 1) Otwórz projekt używany podczas poprzednich zajęć i utwórz w nim nowy pakiet o nazwie Lesson17.dtsx
- 2) Przeciągnij na pusty panel pakietu Data Flow Task
- 3) Kliknij dwukrotnie na Data Flow Task aby przejść do panelu Data Flow
- 4) Przeciągnij na panel obiekt OLE DB Source
- 5) Kliknij dwukrotnie na OLE DB Source, aby otworzyć edytor
- 6) Kliknij na przycisk New, aby otworzyć Connection manager-a (Rysunek 1). Wybierz, jako źródło połączenie do bazy danych AdventureWorks2012
- 7) Jeśli wspomniane źródło nie istnieje kliknij na przycisk New, aby przejść do okna tworzenia nowego połączenia



Rys. 1

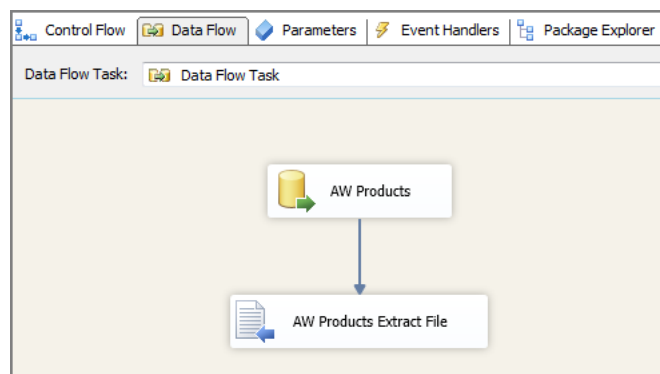
- 8) Ustaw właściwość Server Name na localhost
- 9) Wybierz z listy rozwijanej lub wpisz nazwę bazy danych AdventureWorks2012 i kliknij dwukrotnie OK
- 10) W edytorze OLE DB Source kliknij na menu rozwijane z nazwami tabel i wybierz tabelę Production.Product
- 11) Kliknij na węźle Columns w lewym panelu, aby zobaczyć kolumny tabeli Products
- 12) Kliknij OK, aby zamknąć edytor
- 13) Kliknij prawym przyciskiem myszy na obiekcie OLE DB Source i wybierz polecenie Rename
- 14) Zmień nazwę na AW Products
- 15) Przeciągnij na panel obiekt Flat File Destination
- 16) Połącz za pomocą niebieskiej linii AW Products z Flat File Destination
- 17) Kliknij dwukrotnie na Flat File Destination, aby otworzyć edytor
- 18) Kliknij na przycisk New
- 19) Wybierz format pliku jako Delimited i kliknij OK.
- 20) w Flat File Connection manager Editor zmień nazwę na AW Product Extract
- 21) Zmień opis (ang. description) na All AW product data
- 22) Wprowadź ścieżkę na dysku do pliku AWProducts.txt w polu File Name
- 23) Kliknij OK, aby zamknąć Connection Manager-a
- 24) Kliknij na węźle Mappings w lewym panelu, aby uzyskać ekran pokazany na rysunku 2



Rys. 2

- 25) Kliknij OK, aby zamknąć edytor Flat File Destination

- 26) Kliknij prawym przyciskiem myszy na obiekcie Flat File Destination i wybierz polecenie Rename. Zmień nazwę obiektu na AW Products Extract File. Ostatecznie Data Flow powinien wyglądać tak jak na rysunku 3.



Rys. 3

- 27) Uruchom pakiet.
28) Sprawdź czy w wybranym przez siebie miejscu docelowym powstał plik tekstowy. Sprawdź jego zawartość
29) Zatrzymaj pakiet

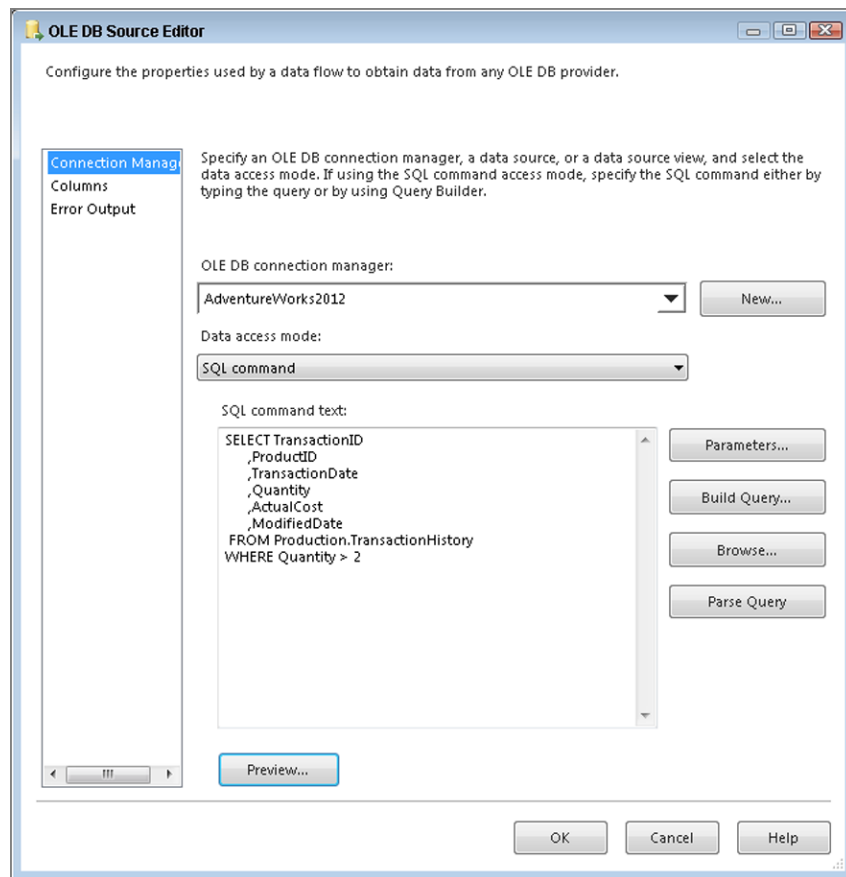
Ekstrakcja danych ze źródła OLE DB

1. Cel ćwiczenia

W trakcie ćwiczenia zostanie utworzone źródło danych OLE DB, aby pozyskać historię transakcji z bazy AdventureWorks2012.

2. Przebieg ćwiczenia

- 1) Otwórz projekt realizowany podczas poprzedniego ćwiczenia i utwórz nowy pakiet o nazwie lesson18.dtsx. Dodaj Data Flow Task na panel Control Flow i nazwij je OLE DB Extract
- 2) W panelu Data Flow utwórz źródło typu OLE DB. Kliknij na nim, aby otworzyć edytor
- 3) Kliknij przycisk New w OLE DB Connection Manager, aby utworzyć nowe połączenie do bazy AdventureWork2012 na serwerze SQL. Kliknij OK po wypełnieniu konfiguracji połączenia

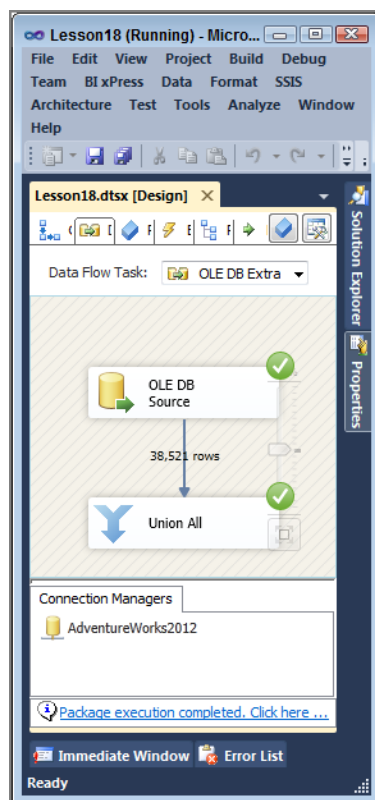


Rys. 1

- 4) W edytorze OLE DB Source wybierz metodę dostępu danych, jako SQL Command (rysunek 1). W oknie SQL command text wpisz następujące zapytanie SQL:

```
SELECT TransactionID
      ,ProductID
      ,TransactionDate
      ,Quantity
      ,ActualCost
      ,ModifiedDate
FROM Production.TransactionHistory
WHERE Quantity > 2
```

- 5) Przeciągnij obiekt Union All na panel Data Flow (Union All pełni obecnie rolę obiektu zastępczego, w następnych lekcjach ustawisz odpowiednie miejsca docelowe). Podłącz źródło danych do obiektu Union All i uruchom pakiet (rysunek 2).



Rys. 2

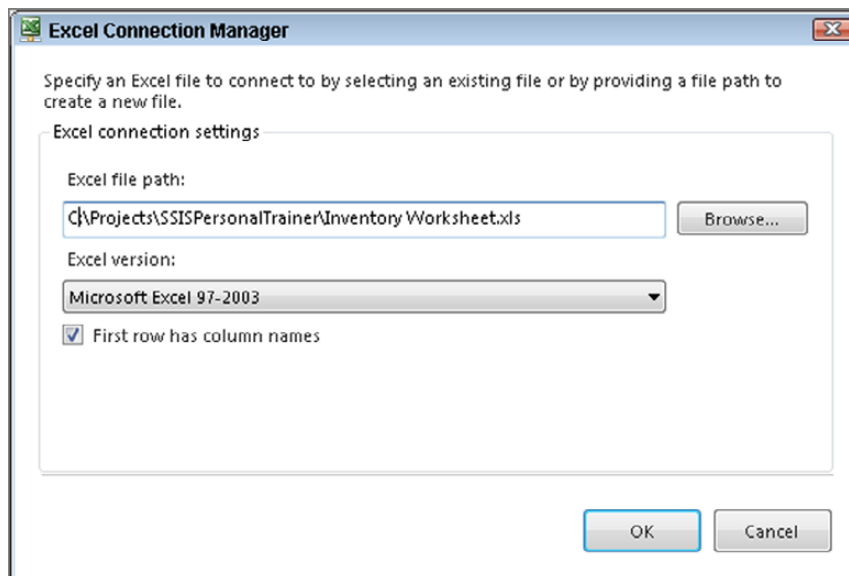
Ekstrakcja danych ze źródła plik Excel-a

1. Cel ćwiczenia

Celem ćwiczenia jest uzyskanie dostępu do danych inwentarzowych, które przechowywane są w pliku excela.

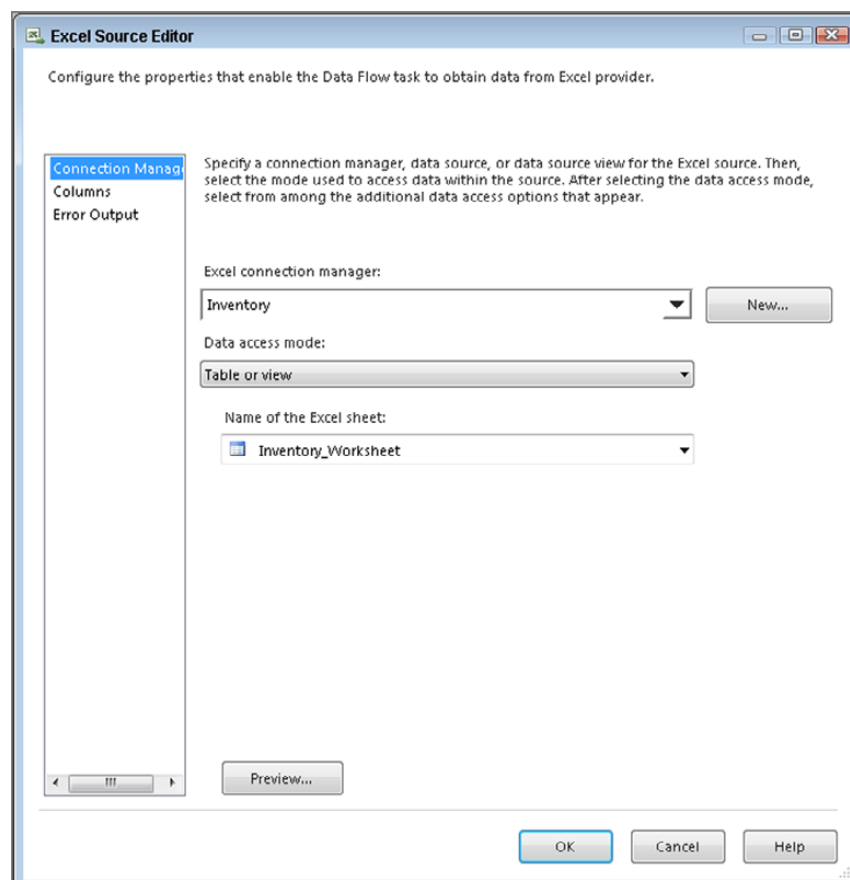
2. Przebieg ćwiczenia

- 1) Otwórz projekt realizowany podczas poprzedniego ćwiczenia i wybierz pakiet o nazwie lesson18.dtsx. Dodaj kolejny Data Flow Task na panel Control Flow i nazwij go Excel Extract
- 2) Przeciągnij obiekt Excel Source na panel Data Flow i otwórz jego edytor
- 3) Kliknij przycisk New, aby otworzyć connection manager-a
- 4) W oknie Excel Connection Manager wskaż ścieżkę do pliku Inventory Worksheet.xls. Wybierz Excel version jako Microsoft Excel 97-2003. Zaznacz opcję First row has column (rysunek 1)



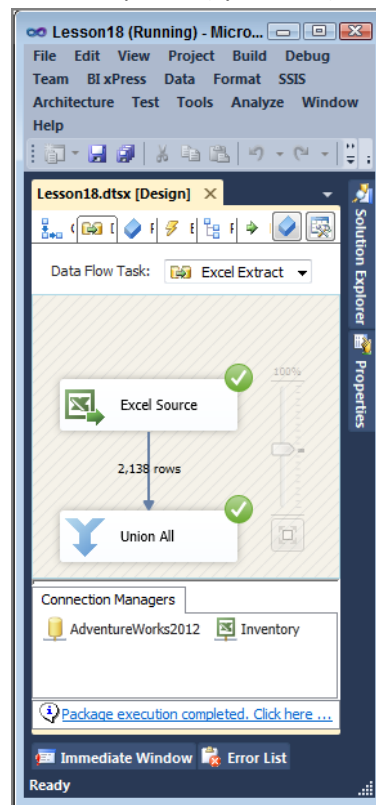
Rys. 1

- 5) Po utworzeniu connection manager-a w edytorze Excel Source wybierz nazwę arkusza Inventory_Worksheet z rozwijanego pola o nazwie Name of the Excel sheet i kliknij OK. (rysunek 2)



Rys. 2

- 6) Przeciągnij obiekt Union All na panel Data Flow (Union All pełni obecnie rolę obiektu zastępczego, w następnych lekcjach ustawisz odpowiednie miejsca docelowe). Podłącz źródło danych do obiektu Union All i uruchom pakiet (rysunek 3).



Rys. 3

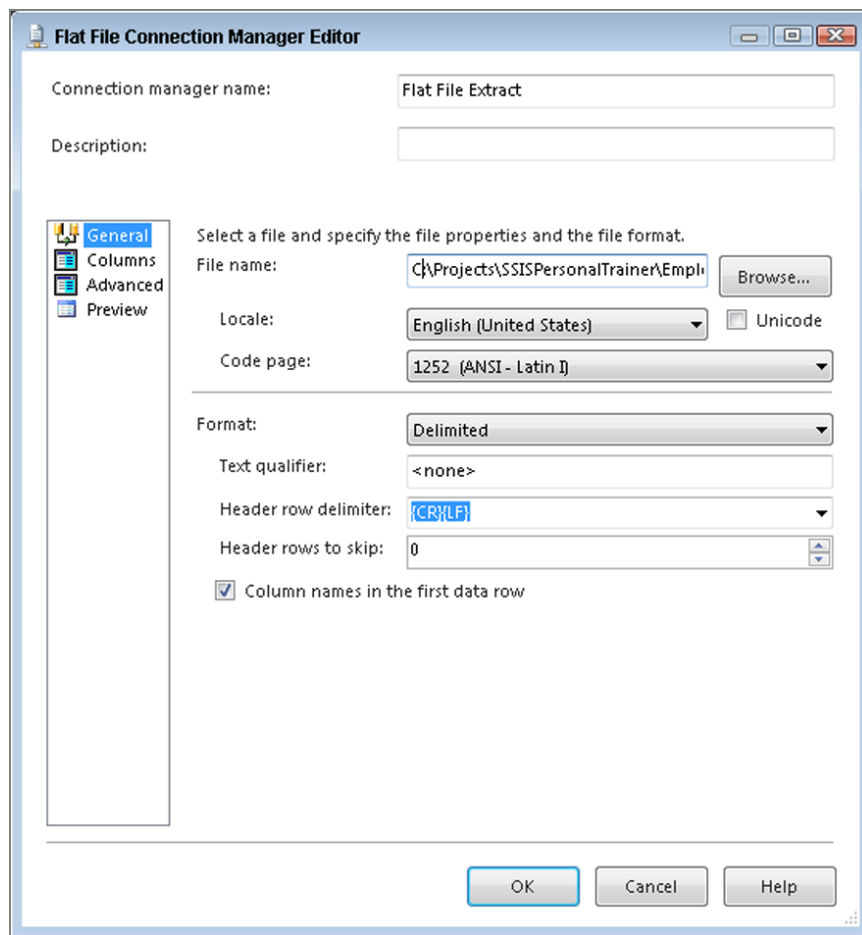
Ekstrakcja danych ze źródła Flat File

1. Cel ćwiczenia

Celem ćwiczenia jest uzyskanie dostępu do pliku płaskiego z danymi pracowników opisującymi liczbę wykorzystanych przez nich dni urlopowych i chorobowych.

2. Przebieg ćwiczenia

- 1) Otwórz projekt realizowany podczas poprzedniego ćwiczenia i wybierz pakiet o nazwie lesson18.dtsx. Dodaj kolejny Data Flow Task na panel Control Flow i nazwij go Flat File Extract
- 2) Przeciągnij Flat File Source na panel Data Flow i kliknij na nim dwukrotnie, aby otworzyć edytor
- 3) Kliknij przycisk New, aby otworzyć edytor connection manager-a
- 4) W zakładce General zmień nazwę connection managera na Flat File Extract i wskaż umiejscowienie pliku źródłowego o nazwie EmployeeList.txt. Zaznacz opcję Column names in the first data row (rysunek 1).



Flat File Connection Manager Editor

Connection manager name: Flat File Extract

Description:

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

File name: C:\Projects\SSISPersonalTrainer\Empl... Browse...

Locale: English (United States) Unicode

Code page: 1252 (ANSI - Latin I)

Format: Delimited

Text qualifier: <none>

Header row delimiter: {CR}{LF}

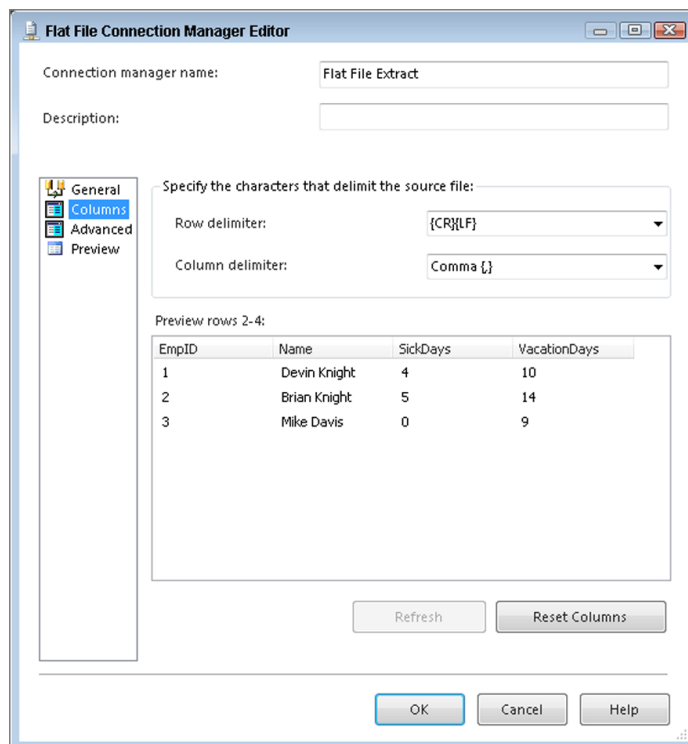
Header rows to skip: 0

☒ Column names in the first data row

OK Cancel Help

Rys. 1

- 5) Wybierz węzeł Columns i upewnij się, że Column delimiter ustawiono, jako przecinek (rysunek 2)



Flat File Connection Manager Editor

Connection manager name: Flat File Extract

Description:

Specify the characters that delimit the source file:

Row delimiter: {CR}{LF}

Column delimiter: Comma {,}

Preview rows 2-4:

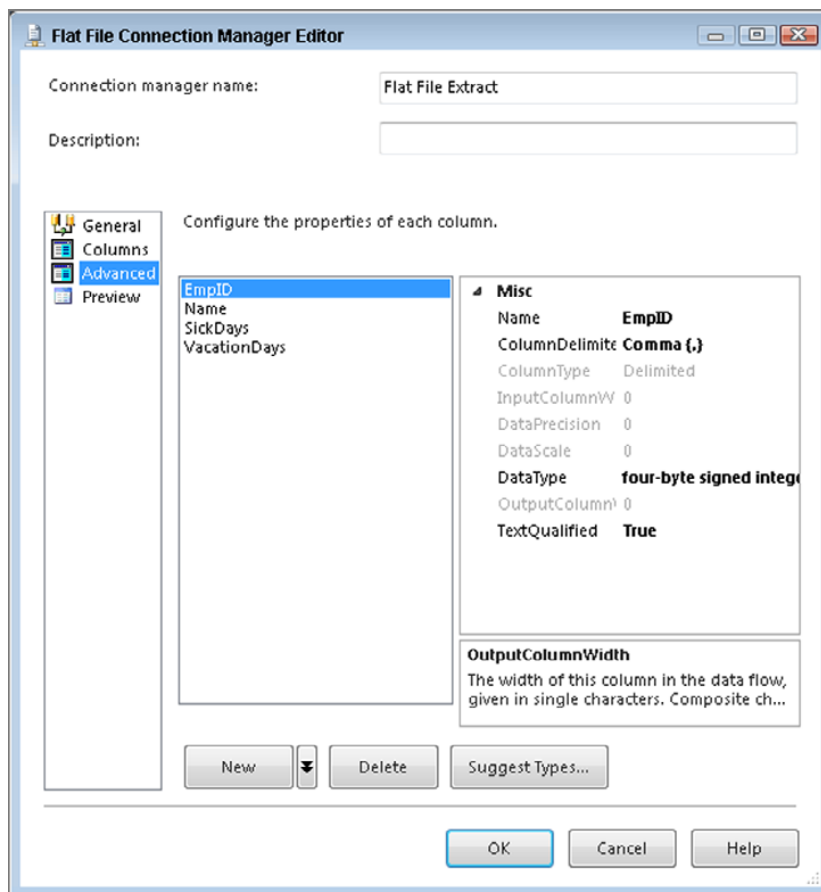
EmpID	Name	SickDays	VacationDays
1	Devin Knight	4	10
2	Brian Knight	5	14
3	Mike Davis	0	9

Refresh Reset Columns

OK Cancel Help

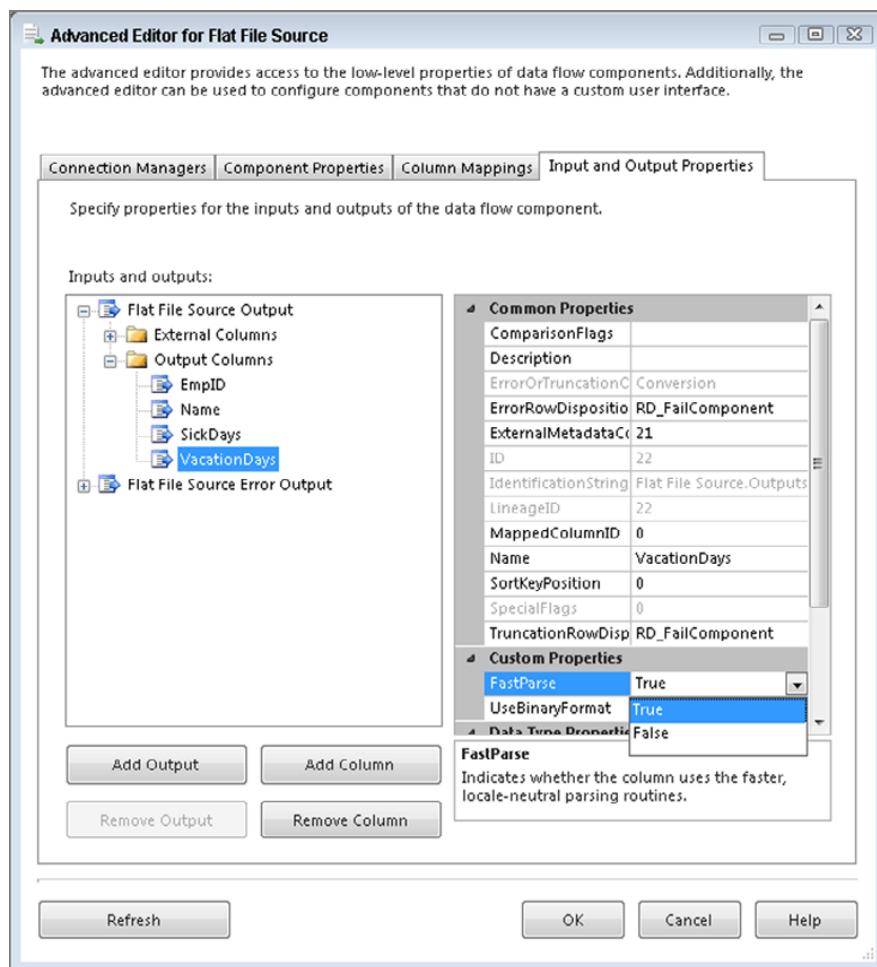
Rys. 2

- 6) Domyślnie wszystkie kolumny z danymi mają przypisany typ danych string. Aby zmienić typ danych należy wybrać węzeł **Advanced**. Typ danych można wybrać manualnie lub dobrać automatycznie przez kliknięcie na przycisk **Suggest Types**. Zwykle SSIS wykrywa poprawne typ danych jednak prawidłowość propozycji należy zawsze zweryfikować. Przykładowo, dla rozważanego pliku dla kolumn **EmpID**, **SickDays** i **VacationDays** SSIS sugeruje typ danych single-byte signed integer (**tinyint**) jednak poprawny typ danych to four-byte signed integer (**int**) (rysunek 3). Ustaw ręcznie poprawny typ danych dla poszczególnych kolumn i kliknij **OK**.



Rys. 3

- 7) Przeciągnij obiekt **Union All** na panel **Data Flow** (**Union All** pełni obecnie rolę obiektu zastępczego, w następnych lekcjach ustawisz odpowiednie miejsca docelowe). Podłącz źródło danych do obiektu **Union All** i uruchom pakiet
- 8) Dobrą praktyką przy korzystaniu ze źródeł typu **Flat File** jest wykorzystanie opcji **Fast Parse**, która drastycznie przyspiesza proces przetwarzania danych z plików płaskich. Domyślnie SSIS dokonuje walidacji w pliku płaskim wszystkich kolumn typu numerycznego lub dat. Jednak przy opcji **Fast Parse** ustawionej na wartość **true** operacja ta jest pomijana, co przyspiesza wykonanie pakietu. Aby ustawić opcję **Fast Parse** należy kliknąć prawym przyciskiem myszy na obiekt **Flat File Source** i wybrać polecenie **Show Advanced Editor**. Następnie w oknie edytora wybrać zakładkę **Input and Output Properties**. Na panelu **Input and Outputs** wybrać kolumnę, dla której będą zmieniane właściwości. W oknie właściwości rozwinąć sekcję **Custom Properties** i ustawić właściwość **FastParse** na **True**. Kliknąć **OK** (rysunek 4).
- 9) Uruchom pakiet ponownie po wprowadzonych zmianach



Rys. 4

Bibliografia

- 1) Knight B., Knight D., Davis M, Snyder W. (2013): Knight's Microsoft® SQL Server® 2012 Integration Services 24-Hour Trainer, John Wiley & Sons.
- 2) Knight B., Veerman E., Moss J.M., Davis M., Rock C. (2012): PROFESSIONAL Microsoft® SQL Server® 2012 Integration Services, John Wiley & Sons.
- 3) <http://www.wrox.com/WileyCDA/Section/id-814197.html>
- 4) [https://msdn.microsoft.com/library/ms169917\(SQL.120\).aspx](https://msdn.microsoft.com/library/ms169917(SQL.120).aspx)
- 5) Tok W-H., Parida R. Masson M. Ding X. Sivashanmugam (2012): Microsoft SQL Server 2012 Integration Services, Promise (tłumaczenie j. polski).
- 6) Kimball R. (2004): The Data Warehouse ETL Toolkit. John Wiley & Sons