

Przykładowe wykorzystanie komponentu Lookup w procesie ETL w celu uniknięcia duplikatów w docelowej bazie danych

Film (krok po kroku)

<http://www.youtube.com/watch?v=rDDlIFYZgX4>

Lista poszczególnych kroków:

1. Tworzymy bazę SQL oraz wypełniamy ją danymi przy pomocy plików "stworzenieBazy.sql" oraz "dane.sql"
2. Otwieramy plik "TutorialLookup.sln"
3. Przechodzimy do zakładki "Data Flow"
4. W obecnej sytuacji dane z pliku tekstowego są ładowane do bazy danych dla hurtowni bezpośrednio. W przykładzie mamy do czynienia z wypełnianiem tablicy "Smieci", której elementy nie powinny się powtarzać.. Niestety, taki sposób nam tego nie zapewnia. Aby pozbyć się tej niedogodności użyjemy komponentu "Lookup"
5. Kasujemy połączenie pomiędzy "Flat File Source" a "OLE DB Destination" (prawe kliknięcie na nim oraz wybranie delete)
6. Z toolboxa przeciągamy element "Lookup"
7. Przeciągamy wyjście (zielone) z elementu "Flat File Source" do elementu "Lookup"
8. Klikamy dwukrotnie na el. "Lookup"
9. W polu "Specify how to handle rows with no matching entries" wybieramy opcję "Redirect rows to no match output"
10. Przechodzimy do zakładki "Connection"
11. Wybieramy tabelę, która ma służyć do porównania ([dbo].[Smieci])
12. Przechodzimy do zakładki "Columns"
13. Tutaj specyfikujemy jakie elementy mają być porównywane. W naszym przypadku nie chcemy w bazie mieć żadnych duplikatów, dlatego łączymy kolumny wejściowe ("Available Inputs Columns") Stanowisko oraz FormaPlatnosc z ich odpowiednikami w dostępnych do porównania ("Avalanche Lookup Columns")
14. Klikamy OK
15. Tworzymy połączenie (zielone) od elementu "Lookup" do elementu "OLE DB Destination"
16. W okienku które wyskoczyło w polu "Output" wybieramy "Lookup No Match Output"
17. W celu sprawdzenia jakie wartości "przeszły" przez komponent Lookup klikamy dwukrotnie na połączenie, które przed chwilą stworzyliśmy i tworzymy podgląd danych (zakładka "Data Viewers" -> "Add" -> "OK")
18. Naciskamy F5 i oglądamy wyniki :-). Przy pierwszym uruchomieniu ukażą się nam dwa rekody, których w bazie jeszcze nie ma. Przy każdym ponownym już nie, gdyż zostały dodane wcześniej do bazy.