

WPROWADZENIE DO INTEGRATION SERVICES 2008

Maciej Kiewra
mkiewra@qbico.pl

Quality Business Intelligence Consulting
<http://www.qbico.pl>



Wstęp

- ⦿ Integration Services – narzędzie do integracji danych
- ⦿ Pomyślane do implementacji procesów ETL
 - Extract – ekstrakcja danych ze źród(e)ł(a)
 - Transformation – transformacja
 - Load – ładowanie
- ⦿ Inne zastosowania:
 - Import/eksport danych
 - Migracja danych
 - Czyszczenie danych
- ⦿ Wchodzi w skład platformy SQL Server 2005/2008

Integration Services – podstawowe pojęcia

- ⦿ Zadanie (ang. task) – pewna czynność, którą należy wykonać w procesie ETL i która posiada swoją implementację w IS
- ⦿ Przykłady zadań:
 - Wykonaj zapytanie SQL
 - Wyślij maila
 - Skopiuj dane
 - Pobierz plik przez FTP
 - Wykonaj program
- ⦿ Pakiet – grupa zadań, których kolejność wykonywania określona jest za pomocą odpowiednich więzów

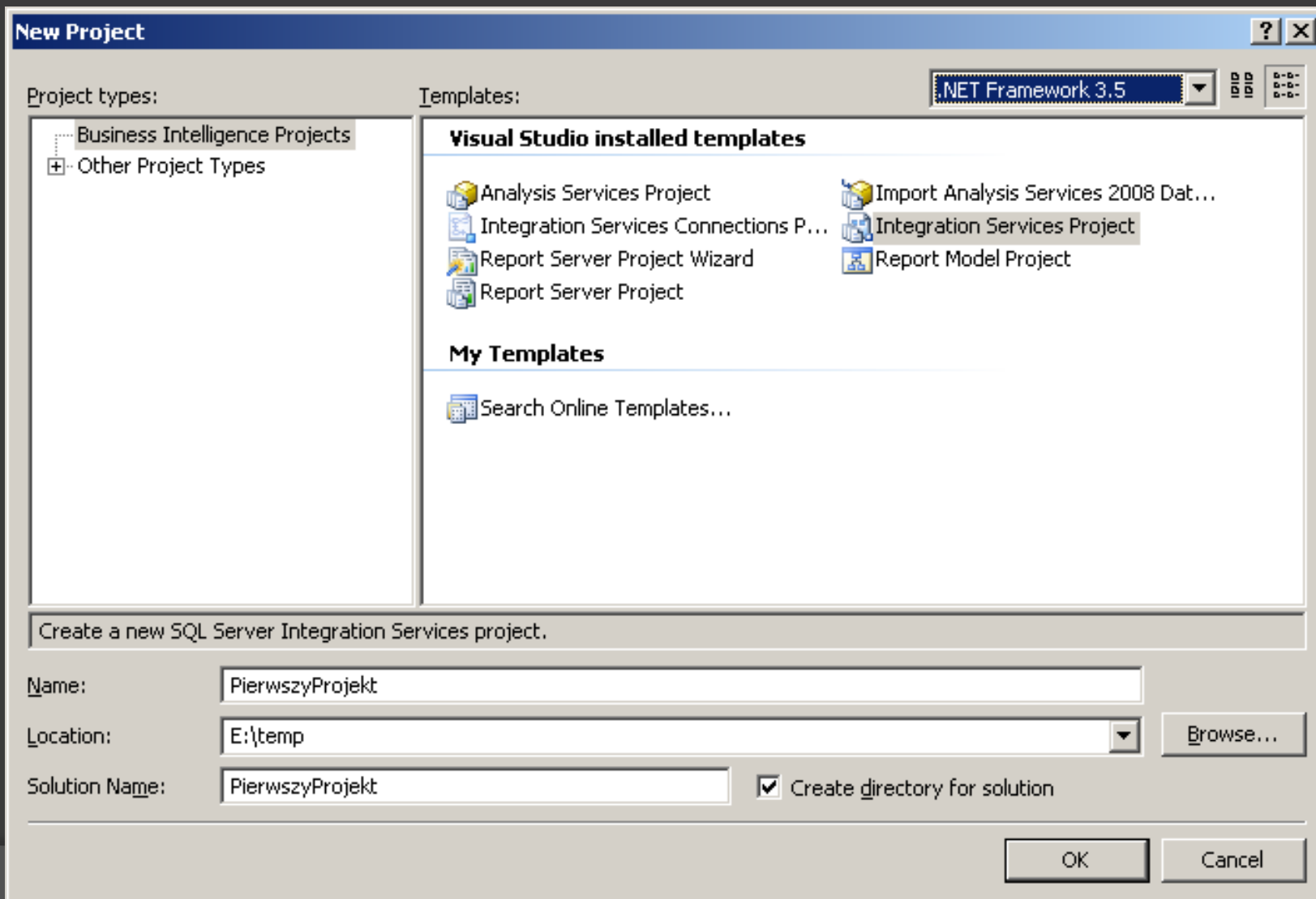
Tworzenie pierwszego pakietu Integration Services (1/2)

- ⦿ Pakiety Integration Services tworzymy przy pomocy:

Business **I**ntelligence **D**evelopment **S**tudio

- ⦿ BIDS = Platforma Visual Studio zaadaptowana do rozwiązań Business Intelligence
 - ETL – Integration Services
 - Raportowanie – Reporting Services
 - Kostki analityczne – Analysis Services

Tworzenie pierwszego pakietu Integration Services (2/2)



Projekt Integration Services – główne okna

- ⦿ Eksplorator rozwiązania (ang. Solution Explorer) – wszystkie pliki wchodzące w rozwiązanie podzielone na projekty
- ⦿ ToolBox – zadania i komponenty
- ⦿ Okno główne podzielone na zakładki:
 - Control Flow – determinuje kolejność wykonywanych zadań
 - Data Flow – szczegółowe informacje na temat przepływu danych
 - Event Handler – pozwala definiować zdarzenia
 - Package Explorer – dane dotyczące pakietu w postaci zagregowanej

Dane źródłowe wykorzystane w przykładach

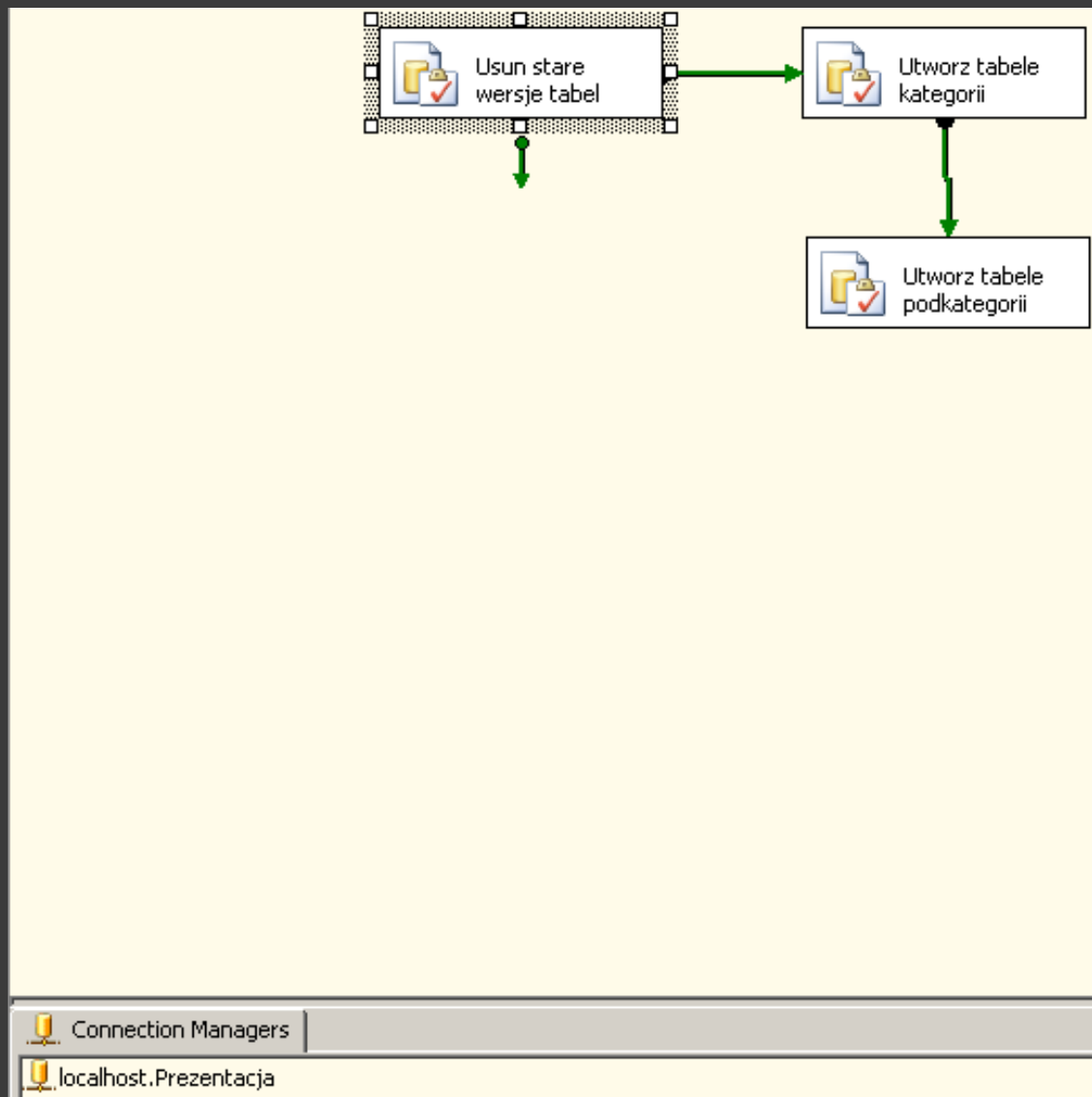
- Oceny książek dokonywane przez anonimowych użytkowników
- AdventureWorksDW
- Trzy pliki CSV
 - Users.csv – id użytkownika, wiek i miejsce zamieszkania
 - Books.csv – opisuje książki
 - Books-ratings.csv – oceny użytkowników
- Pochodzą ze strony

<http://www.informatik.uni-freiburg.de/~chiegler/BX/>

ControlFlow - ogólne informacje

- Strzałki między komponentami – więzy wymuszające kolejność (*precedence constraints*) oznaczają kolejność wykonywania operacji
- 4 rodzaje:
 - Użyj więzów (OnSuccess, OnFailure, OnComplete)
 - Użyj wyrażenia
 - Użyj więzów i wyrażenia
 - Użyj więzów lub wyrażenia

Pierwszy pakiet



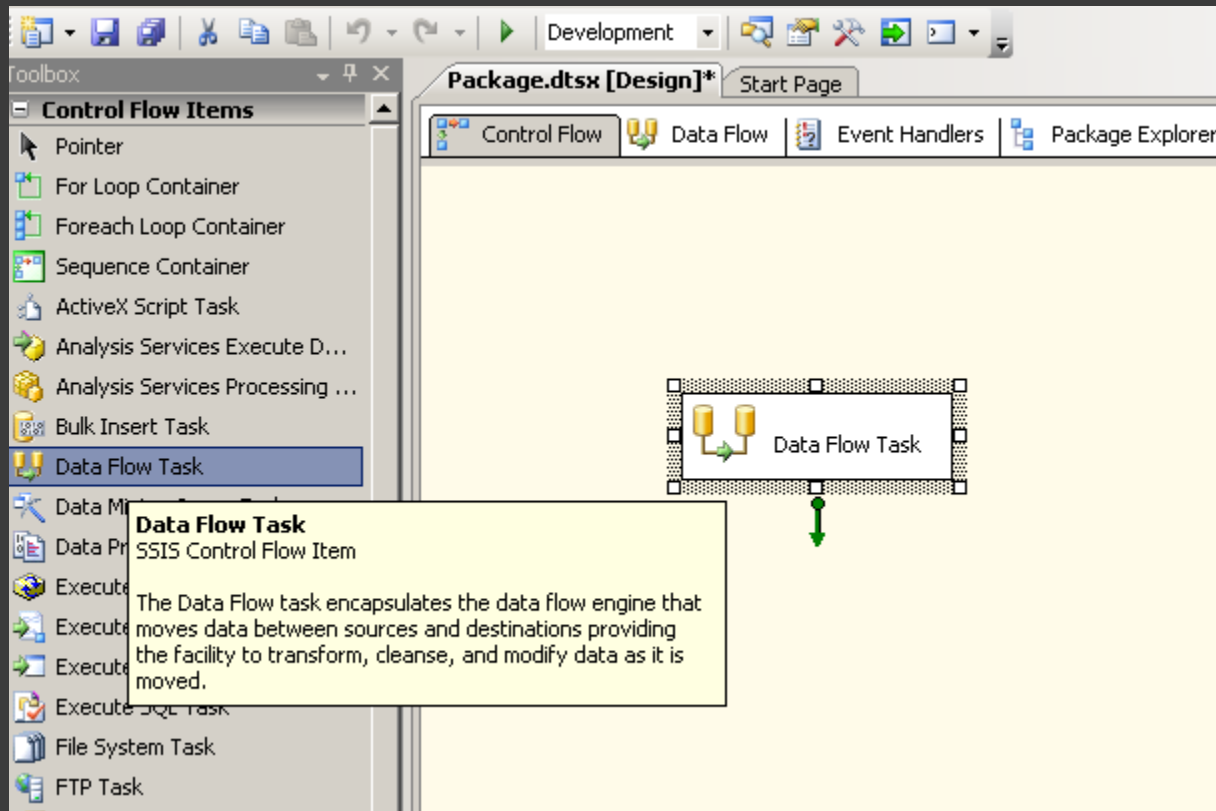
Pakiet 1 wnioski

- „Connection managers” – możliwość zapisania dostępu do zewnętrznych źródeł danych w jednym miejscu
- Bardzo często kolejność wykonywanych zadań ma znaczenie
- Narzędzie pozwala na budowę procesu ETL za pomocą interfejsu graficznego

Najważniejsze typy zadań

- ◉ Execute SQL Task – wykonuje zapytanie SQL (także procedurę składowalną)
- ◉ Execute Package Task – wykonuje inny pakiet integration services
- ◉ Execute Process Task – wykonuje zewnętrzny proces (program) zainstalowany w bieżącym systemie operacyjnym
- ◉ SendMailTask – pozwala wysłać wiadomość elektroniczną (e-mail)
- ◉ DataFlow – przepływ danych (do szczegółowego omówienia potrzebna osobna sesja)

Dodawanie DataFlow Task

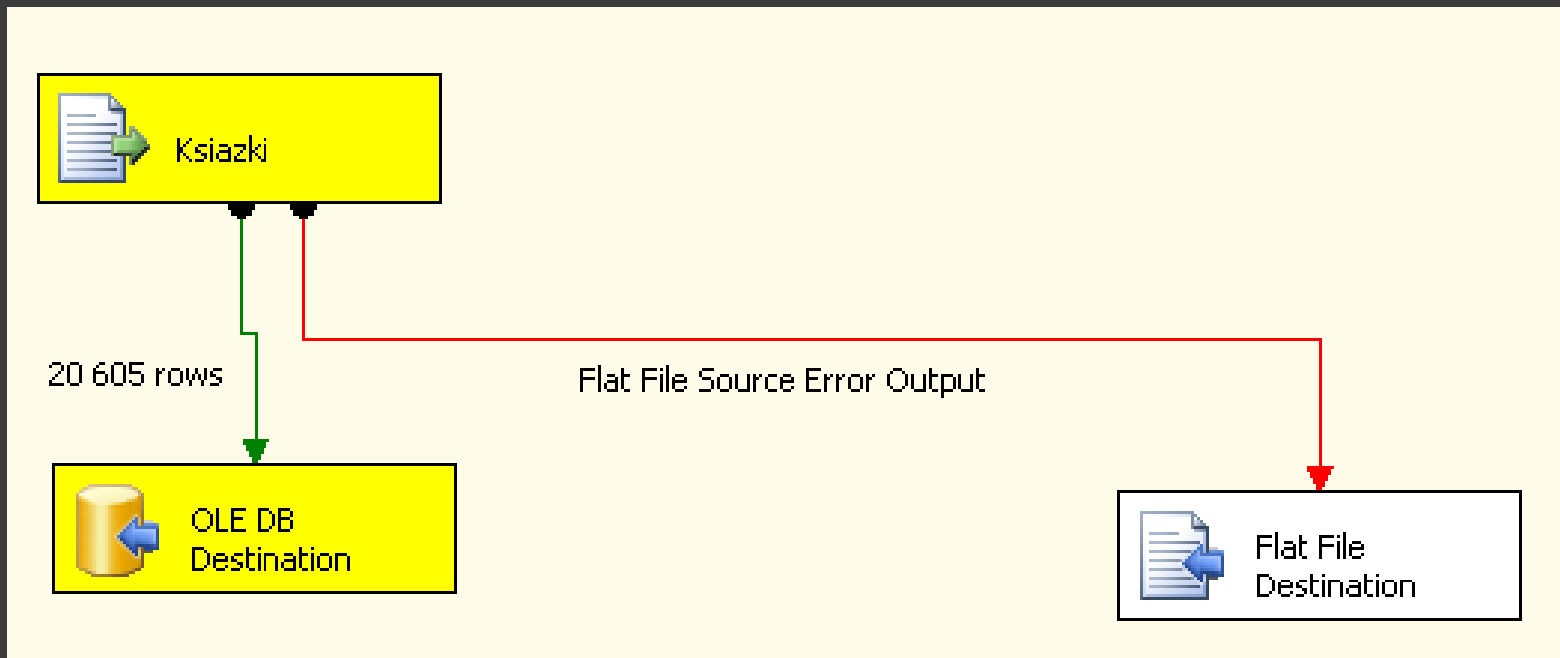


Data Flow podstawowe informacje

- ⦿ Osobna zakładka w Integration Services
- ⦿ 3 Rodzaje komponentów
 - Wejścia
 - Transformacji
 - Wyjścia
- ⦿ 3 tryby widoczności komponentów
- ⦿ Pojęcie Connection Manager
- ⦿ Pojęcie wejść i wyjść
- ⦿ „Metafora hydrauliczna”

Najprostszy „data flow”

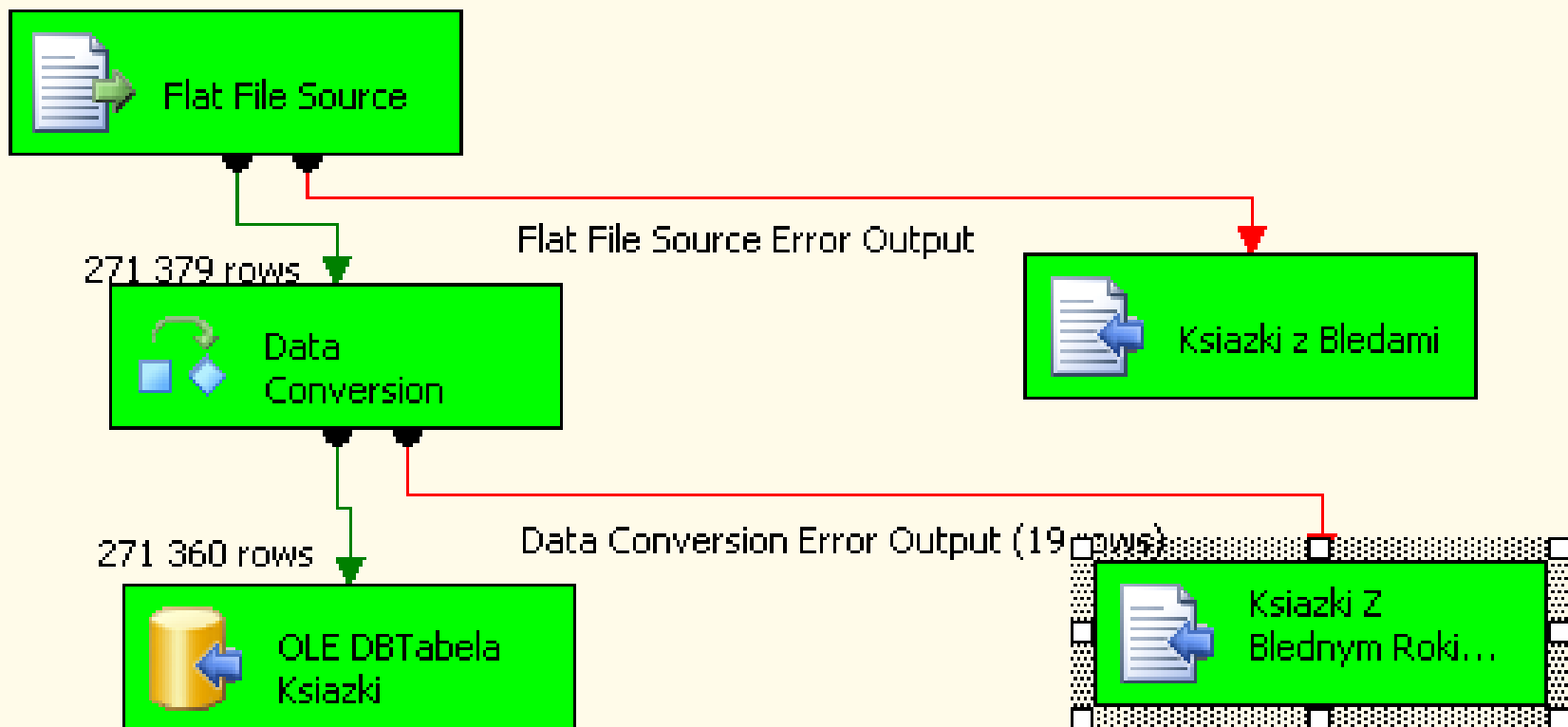
- Łączymy wejście z wyjściem (bez transformacji)



Wnioski z przykładu

- Strzałki w „data flow” i w „control flow” mają odmienne znaczenie!
- Możliwość przekierowywania błędnych wierszy
- Łatwa kontrola nad przebiegiem procesu

Przykład z transformacją



Wnioski z przykładu

- Skomplikowane operacje konwersji możliwe do zrealizowania wizualnie
- Możliwość „podglądania danych”

Dlaczego Integration Services?

- Graficzne środowisko pracy ułatwia zarządzanie skomplikowanymi procesami ETL
- Prostsze wytwarzanie procesów
- Łatwa parametryzacja i eksternalizacja wartości konfigurowalnych
- Zaawansowane opcje związane z diagnostyką i logowaniem błędów
- Potęga .NET w zasięgu ręki
- Bezproblemowa integracja z SQL Server

Dziękuję za uwagę



QBICO - podwykonawstwo projektów

◎ Business Intelligence

- Projektowanie hurtowni danych
- ETL (Integration Services)
- Kostki OLAP i KPI
- Reporting Services
- ProClarity
- Data-Mining

◎ Oprogramowanie .NET

- Systemy na zamówienie
- Aplikacje ASP.NET
- Audyt kodu źródłowego
- Badanie wydajności
- Middleware dla hurtowni danych

