

Usługi analityczne – budowa kostki analitycznej

Część pierwsza.

Wprowadzenie

W wielu dziedzinach działalności człowieka analiza zebranych danych jest jednym z najważniejszych mechanizmów podejmowania decyzji. Od wyboru właściwego rozwiązania zależy bardzo wiele, wręcz istnienie firmy. Posiadanie i wykorzystywanie odpowiednich narzędzi jest w takich sytuacjach niezbędne.

Platforma SQL 2005 a usługi analityczne

Microsoft SQL Server 2005 jest kompletnym rozwiązaniem, zawierającym narzędzia jak i funkcje do tworzenia mechanizmów analizy danych. Nowoczesna technologia serwera umożliwia tworzenie zaawansowanych modeli analitycznych a rozbudowane procesy automatyzacji pozwalają na zarządzanie złożonymi systemami Business Intelligence (BI).

SQL Server 2005 zawiera szereg nowych komponentów BI:

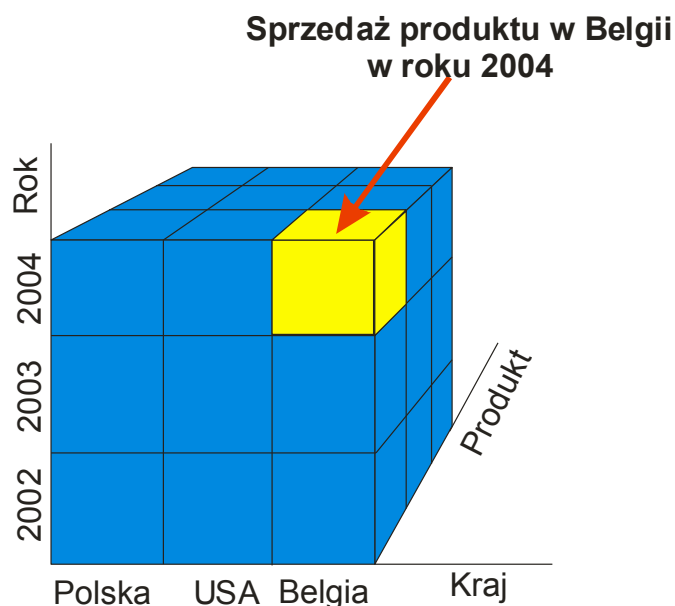
- Usługi Analysis Services – OLAP
- Usługi Analysis Services – Data mining
- Usługi Reporting Services
- SQL Server Integration services (SSIS)

Nowoczesne narzędzia zarządzające i projektowe: SQL Server Management Studio oraz SQL Server Business Intelligence Development Studio pozwalają w pełni realizować cykl projektowy oraz wdrożeniowy aplikacji analitycznej.

BI Development Studio, zawierający Visual studio 2005, jest zintegrowanym środowiskiem dla programistów, pozwalającym zaprojektować aplikację, przetestować oraz ją uruchomić. Usługi analityczne (OLAP) oraz data mining znacznie rozszerzają techniki analizy danych. Zastosowane rozwiązania techniczne pozwalają dostarczać usługi analityczne praktycznie bez opóźnień – analityka „on-line”. Usługi SSIS umożliwiają transformację dużych ilości danych. Nowoczesne przekształcenia, pozwalają w bardzo szybki sposób integrować i zarządzać ogromną ilością danych. Usługi raportowania, zintegrowane z całością, zapewniają ujednoczone mechanizmy prezentacji danych.

Budowa kostki analitycznej

Jedną z bardzo szeroko stosowanych metod analizy danych jest wykorzystanie tzw. kostek analitycznych (ang. Cube). Pojęcie to bardzo obrazowo pokazuje w jaki sposób mamy zamiar analizować dane. Oczywiście w prezentowanym przykładzie kostka posiada tylko trzy wymiary. W rzeczywistość, w zastosowaniach praktycznych, liczba wymiarów sięga kilkunastu lub kilkudziesięciu.

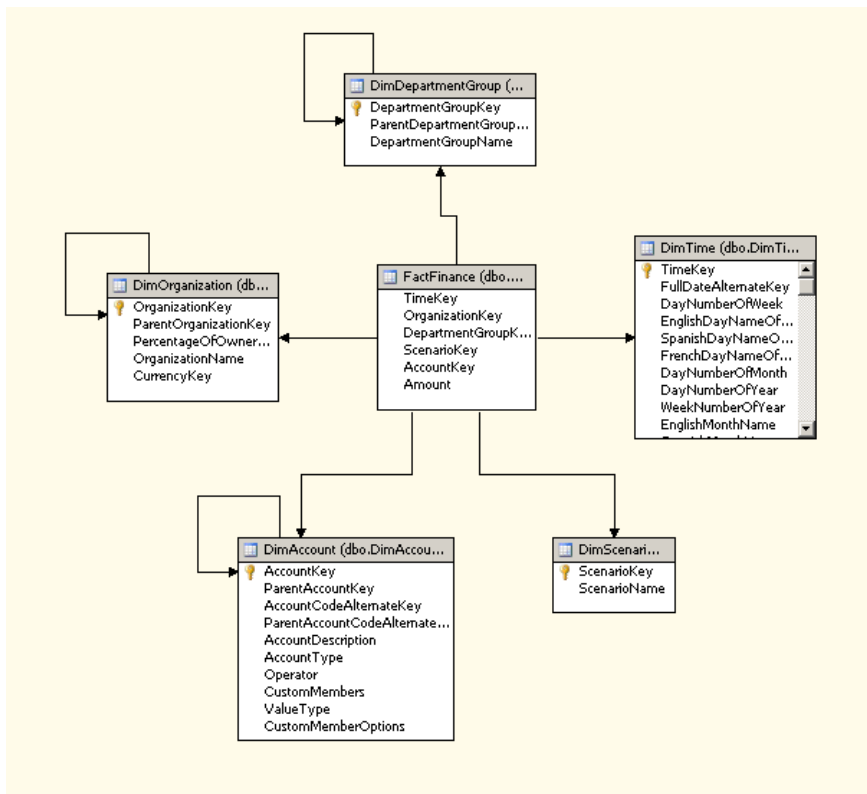


Rys 1. Obrazowa prezentacja danych w procesie analizy danych – kostka analityczna

Proces przygotowanie struktur danych dla mechanizmów analitycznych obejmuje kilka kroków:

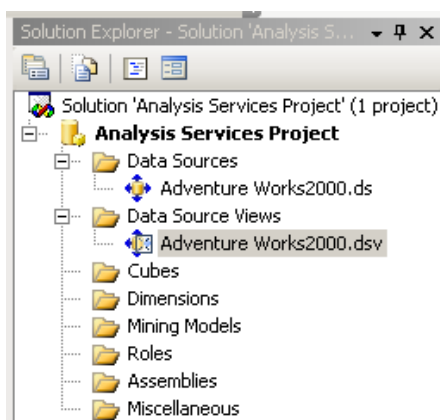
- zdefiniowanie źródła danych
- zdefiniowanie widoku źródła danych
- zdefiniowanie kostki analitycznej
- zarządzanie kostką analityczną

Jako źródło danych, dla mechanizmów analitycznych można wykorzystać dowolne struktury relacyjnych baz danych. Jednak ze względów wydajnościowych zaleca się wykorzystanie baz danych opartych o serwery SQL. Usługi analityczne serwera SQL 2005 mogą tworzyć kostki analityczne w oparciu o dowolną strukturę tabel bądź widoków. Jednak, również ze względów wydajnościowych, architektura źródła danych powinna być oparta o hurtownię danych (ang. Data warehouse). Struktura tabel hurtowni danych opiera się o gwiazdę, gdzie centralna tabela, zwana tabelą faktów (ang. Fact), przechowuje informacje, zwane wartościami pomiarowymi (ang. Measure), obejmujące wszystkie wartości liczbowe. Z tabelą faktów, poprzez klucze zewnętrzne, powiązane są tabele wymiarowe (ang. Dimension) zawierające opisy wartości pomiarowych np. nazwa towaru, odbiorca towaru czy też datę sprzedaży towaru. Tak zbudowana struktury danych uzyskuje się poprzez przekształcenia danych pochodzących z baz operacyjnych. Dane zawarte w hurtowni danych aktualizuje się w zależności od potrzeb, związanych z dostępnością danych analitycznych. Hurtownie danych są bazami danych typu OLTP.



Rys 2. Architektura hurtownia danych

Mając źródło danych możemy przystąpić do budowy kostki analitycznej. Podstawowym narzędziem budowy kostek analitycznych jest wykorzystanie narzędzia Business Intelligence Development Studio. Przygotowane szablony projektów pozwalają bardzo szybko zdefiniować projekt. W naszym przypadku wybierzemy szablon „Analysis Services”.



Rys 3. Definicja źródeł danych i widoków źródeł danych

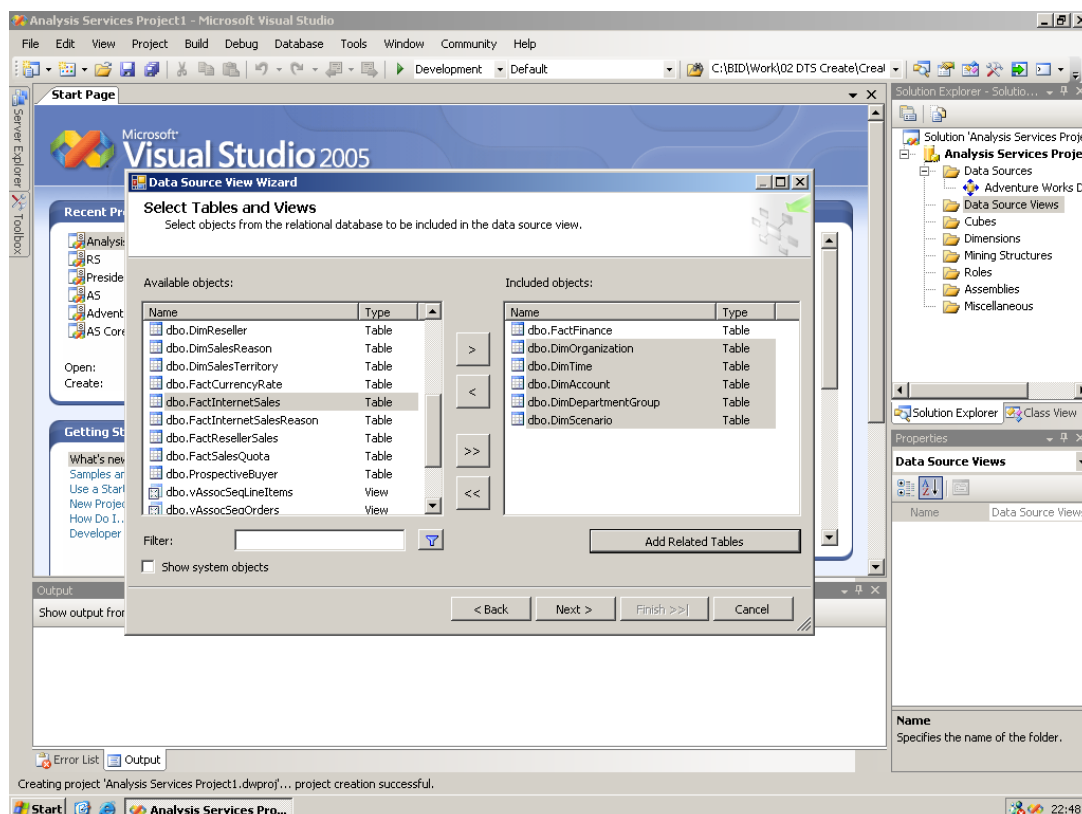
Gdy zostanie utworzony pusty projekt, należy utworzyć źródło danych (ang. Data Source) do podłączenia do źródłowej bazy danych, która może się znajdować w dowolnym obsługiwany systemie zarządzania relacyjną bazą danych. Wykorzystamy dostarczone wraz z serwerem MS SQL 2005 przykładową hurtownię danych „AdventureWorksDW”. Źródło

danych przechowuje informacje niezbędne do połączenia się ze źródłem danych, natomiast widoki źródeł danych (ang. Data Source Views) zawierają informacje na temat zbiorów tabel w źródłowej bazie danych. Te informacje nie są ograniczone do fizycznej struktury tabel w źródłowej bazie danych. Istnieje możliwość modyfikacji struktur tabel jak i można dodać informacje takie jak relacje, modyfikować nazwy tabel i kolumn, dodawać kolumny obliczane i zapytania.

Widoki źródeł danych mają szerokie zastosowanie zwłaszcza gdy:

- Ilość tabel w źródle danych jest bardzo duża. Wybór tylko określonych z nich pozwala w szybki sposób budować struktury analityczne
- Osoba tworząca projekt nie posiada uprawnień do wszystkich tabel w bazie
- Widoki mogą być wykorzystywane w wielu projektach
- Twórca projektu nie ma bezpośredniego kontaktu ze źródłem danych (projekt off-line)
- Ponadto odpowiednie zaprojektowanie widoków źródeł danych pozwala znacznie zaoszczędzić czas w procesie projektowania samej kostki analitycznej.

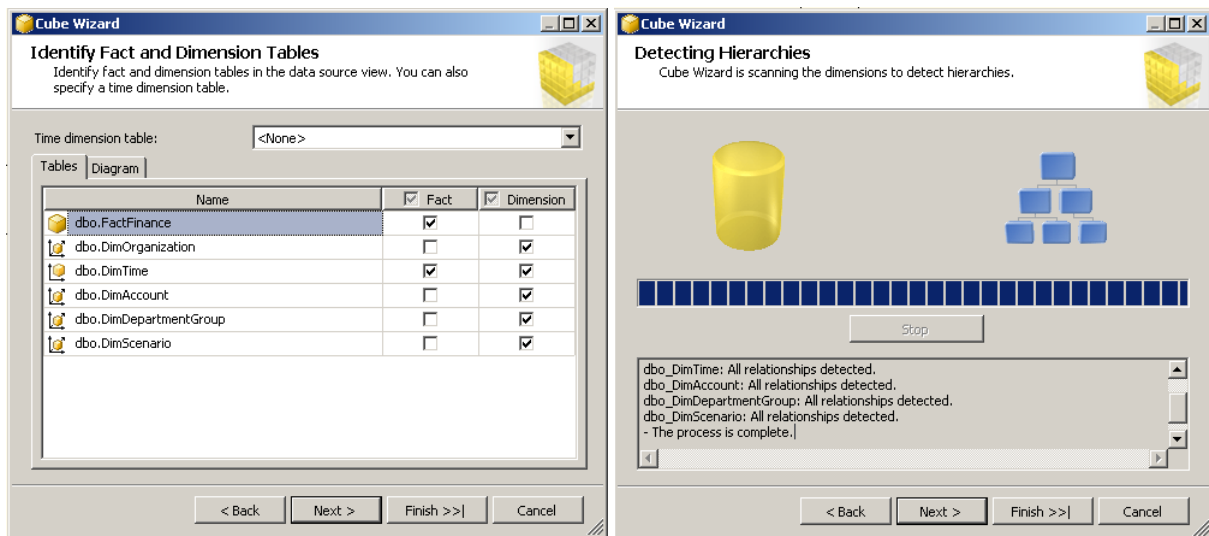
W naszym projekcie, wykorzystamy strukturę hurtowni danych o nazwie AdventureWorksDW zawierającą tabelę faktów o nazwie dbo.FactFinance oraz powiązane z nią tabele wymiarowe: dbo.dimOrganization, dbo.dimTime, dbo.dimAccount, dbo.dimDepartmentGroup oraz dbo.dimScenario.



Rys 4. Definicja widoków źródeł danych dla bazy AdventureWorksDW

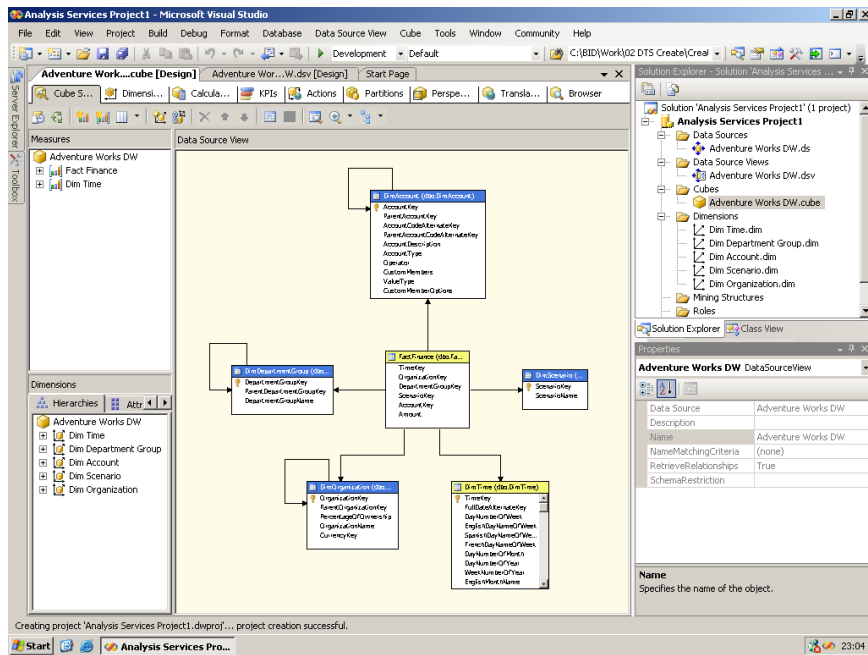
Mając zdefiniowane mechanizmy dostępu do danych możemy przystąpić do budowy kostki analitycznej. Mamy możliwości wyboru pomiędzy wykorzystaniem inteligentnego kreatora

„IntelliCube„, a budową kostki samodzielnie. W naszym projekcie wykorzystamy kreatora do stworzenia podstawowej kostki, a następnie dokonamy modyfikacji jej parametrów. Po kliknięciu prawym przyciskiem myszy na folder „cube” w oknie eksploratora projektu, wywołujemy operację „New cube” przechodzimy do tworzenia kostki analitycznej. W pierwszych krokach definiujemy parametry naszej kostki czy utworzyć kostkę zoptymalizowaną dla tabel przestawnych czy raportów. Po wybraniu opcji dotyczącej tabel przestawnych, kreator przystępuje do badania struktury powiązań w widoku źródeł danych. W pierwszej kolejności identyfikuje tabelę faktów oraz tabele wymiarowe. Następnie przystępuje do analizy struktur danych w wymiarach, tworząc odpowiednie hierarchie oraz powiązania z danymi w tabeli faktów.



Rys 5. Identyfikacja tabel w źródle danych oraz wykrywanie hierarchii w wymiarach

W efekcie po zakończeniu działania kreatora uzyskujemy funkcjonalną definicję kostki analitycznej. Zostanie utworzonych szereg zakładek pozwalających zarządzać zdefiniowaną kostką: wymiary, tabela faktów, powiązania, KPI, perspektywy. Zostaje też uruchomiona przeglądarka kostek.



Rys 6. Efekt działania kreatora IntelliCube

Koniec części pierwszej

NetSystem Tomasz Skurniak
 CNI, CNE, MCT, MCSE, MCDBA, MCTS, MCITP
 Ul. J. Burszty 25, 61-422 Poznań
 E: Tomasz@Skurniak.pl
 W: www.protis.pl
 T: +48 601761013
 F: +48 618308249