

Usługi analityczne – podstawy budowy kostki analitycznej

Część druga - zarządzanie

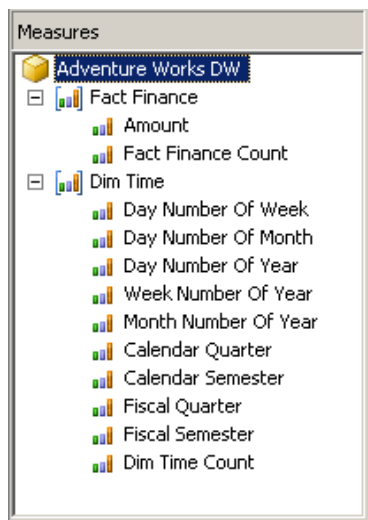
Nasz definicja kostki analitycznej nie zawiera jeszcze danych. Aby zbudować kostkę funkcjonalnie działającą musimy, dokonać wygenerowania odpowiednich struktur danych, będącymi istotą mechanizmów analitycznych. Są to wszelkie agregacje związane z wymiarami naszej kostki. Proces ten pobiera dane i dokonuje budowy struktury bazy danych OLAP (ang. On-Line Analytical Processing). Charakterystyczną cechą tego typu struktur bazodanowych, jest ich zoptymalizowanie w zakresie odczytu danych. Budowę kostki możemy wymusić na kilka sposobów:

- wybieramy opcję „Deploy Project...” z menu „Build”
- Z menu kontekstowego na nazwie naszego projektu wybieramy opcję „Deploy”

Zarządzanie kostką analityczną

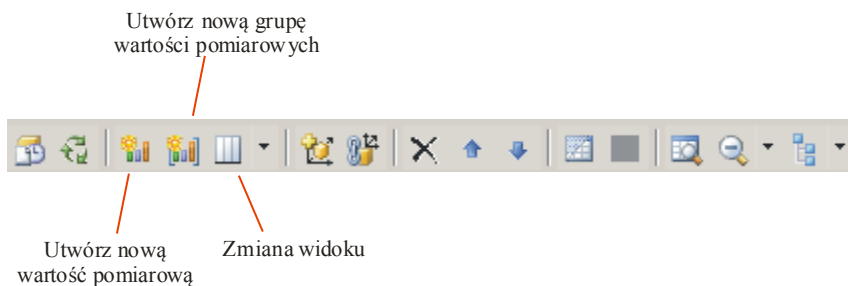
- **wartości pomiarowe**

W zależności od potrzeb, możemy teraz przystąpić do przeglądu i ewentualnie przebudowy kostki analitycznej. W zakładce „Cube Structure”, która jest domyślnie aktywna, możemy zmodyfikować ustawienia kostki. Możemy także prześledzić w jaki sposób została utworzona kostka przez kreator. Zgodnie z zasadami analizy, jako tabela faktów została wybrana tabela „FactFinance” zawierająca wartości obliczeniowe. Ponadto, ze względu na powiązania pomiędzy danymi, tabela „DimTime” zawierająca oś czasową, również została wybrana jako druga tabela faktów. Pozostałe tabel zostały ustalone jako tabele wymiarowe. Informacje te możemy prześledzić w części „Measures”.



Rys 1. Struktura wartości pomiarowych

Jak można również zauważyć, w wyniku pogrupowania danych według dni, miesięcy, kwartałów i lat, w kostce analitycznej zostały dokonane agregacje związane w/w okresami czasu. Bardzo elastyczna budowa struktur danych, pozwala w każdej chwili modyfikować wartości pomiarowe jaki i dodawać kolejne.



Rys 2. Pasek narzędziowy budowy kostki analitycznej

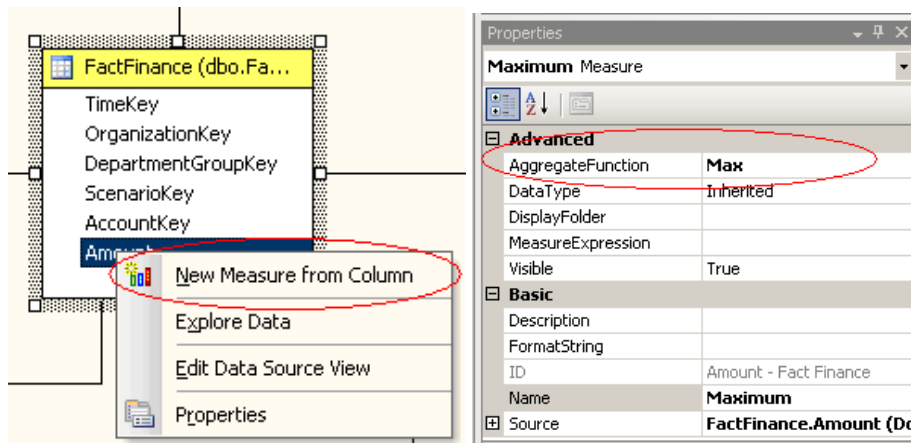
Bardzo przydatny w zakresie definiowania wartości pomiarowych, jest możliwość przełączania widoków ze struktury hierarchicznej na strukturę tabelaryczną – „Grid”. Mamy tutaj możliwość zdefiniowania w szybki sposób określonych funkcji agregacyjnych oraz zdefiniować typy danych.

Measures				
	Name	Measure Group	Data Type	Aggregation
	Amount	Fact Finance	Double	Sum
	Fact Finance Count	Fact Finance	Integer	Count
	Maximum	Fact Finance	Double	Max
	Minimum	Fact Finance	Double	Min
	Day Number Of W...	Dim Time	UnsignedInt	Sum
	Day Number Of Mo...	Dim Time	UnsignedInt	Sum
	Day Number Of Year	Dim Time	Integer	Sum
	Week Number Of Y...	Dim Time	UnsignedInt	Sum
	Month Number Of ...	Dim Time	UnsignedInt	Sum
	Calendar Quarter	Dim Time	UnsignedInt	Sum
	Calendar Semester	Dim Time	UnsignedInt	Sum
	Fiscal Quarter	Dim Time	UnsignedInt	Sum
	Fiscal Semester	Dim Time	UnsignedInt	Sum
	Dim Time Count	Dim Time	Integer	Count
	Add new measure...			

Rys 3. Tryb pracy „Grid”

W naszym przykładzie dodamy dwie nowe wartości pomiarowe: wartość minimalną i maksymalną sprzedaży.

Kliknięcie prawym przyciskiem na danej kolumnie w tabeli, na bazie której zamierzamy utworzyć nową wartość pomiarową, i wybranie opcji „New measure from column”. Po utworzeniu nowej kolumny pomiarowej zmieniamy jej nazwę na „Maksimum” a następnie we właściwościach definiujemy z listy dostępnych funkcji „Max”. Podobnie postępujemy w przypadku kolumny „Minimum”. Oczywiście w/w operacje można również wykonywać wywołując opcje z menu czy też z paska narzędziowego.

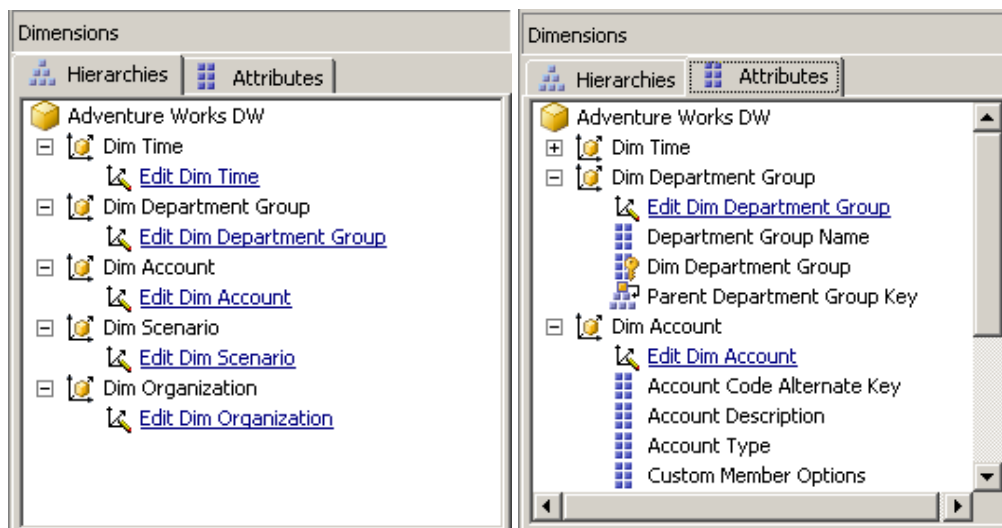


Rys 4. Definicja nowych wartości pomiarowych

W SQL2005 wprowadzono nowe mechanizmy wykorzystania wartości pomiarowych. Możliwe jest zdefiniowanie grup pomiarowych „Measure Group” i powiązanie ich z określonymi wymiarami w określony sposób, który zostanie przedstawiony poniżej.

- **Wymiary (Dimensions)**

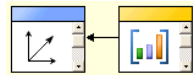
W części opisanej jako „Dimensions” możemy zweryfikować budowę wymiarów i przejść do ewentualnych modyfikacji tychże wymiarów. W usługach analitycznych SQL2005 wprowadzono znaczne modyfikacje w stosunku do poprzednich wersji w zakresie organizacji wymiarów. Przede wszystkim wymiary nie są oparte tylko i wyłącznie na strukturze hierarchicznej ale także na atrybutach wymiarów takich jak np. kolor, rozmiar czy też typ odbiorcy. W takim razie nie jest już wymagane jawne tworzenie atrybutów i na ich podstawie definiowanie kolejnych wymiarów.



Rys 5. Struktura wymiarów i atrybutów

Powiązanie wymiarów z wartościami pomiarowymi możemy określić w zakładce „Dimension”. W SQL2005 wprowadzono bardzo elastyczne techniki odwzorowania wartości pomiarowych w wymiarach. Możemy zdefiniować tutaj następujące ich typy:

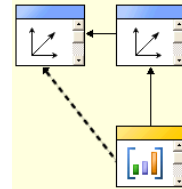
- Brak powiązania – tabela wymiarów nie jest powiązana z tabelą faktów
- Regularny – tabela wymiarowa jest bezpośrednio powiązana z tabelą faktów



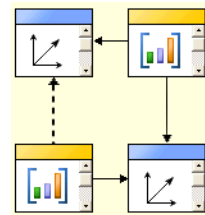
- Tabela fakt – wymiar jest elementem tabeli fakt



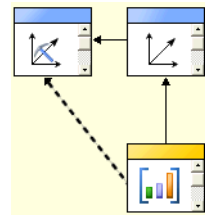
- Odniesienie – tabela wymiarowa jest powiązana z tabelą pośrednią, która następnie jest powiązana z tabelą faktów



- Wiele-do-wielu – tabela wymiarowa jest powiązana z pośrednią tabelą faktów, która jest powiązana z tabelą wymiarową, która z kolei z wyjściową tabelą faktów

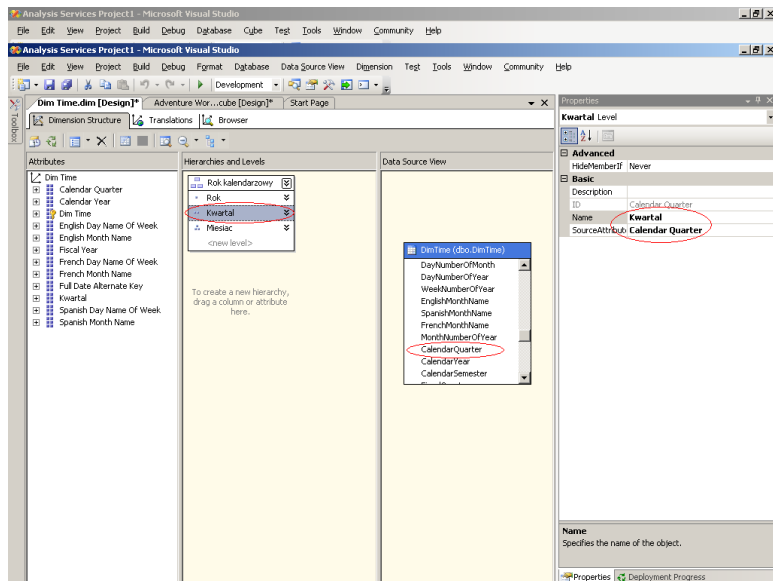


- Data mining – powiązanie tabeli wymiarowej opartej o model data minig (projekt musi być zawarty w kostce) z tabelą faktów.



W przypadku, gdy kreator nie zdefiniowałby wymaganych wymiarów bądź grupami pomiarowymi, w tej zakładce możemy zdefiniować nowy wymiar oraz powiązać z tabelą faktów.

Oczywiście kreator nie zawsze zdefiniuje wszystkie niezbędne hierarchie w danym wymiarze. Dodajmy, w naszym projekcie hierarchię opartą o wymiar czasu zawierającą lata kalendarzowe, kwartały oraz miesiące. Procedurę rozpoczniemy od wywołania edytora wymiaru „Dim time”, klikając na odpowiedni link w wymiarze.



Rys 6. Tworzenie własnej hierarchii

Tworząc hierarchię mamy możliwość definiowania własnej nazwy zarówno hierarchii jak i każdego jej poziomu. Ponadto, w przypadku, gdy poszczególne poziomy nie są w pełni symetryczne, definiujemy, jak mają być one wyświetlane np. pomijanie powielających się elementów.

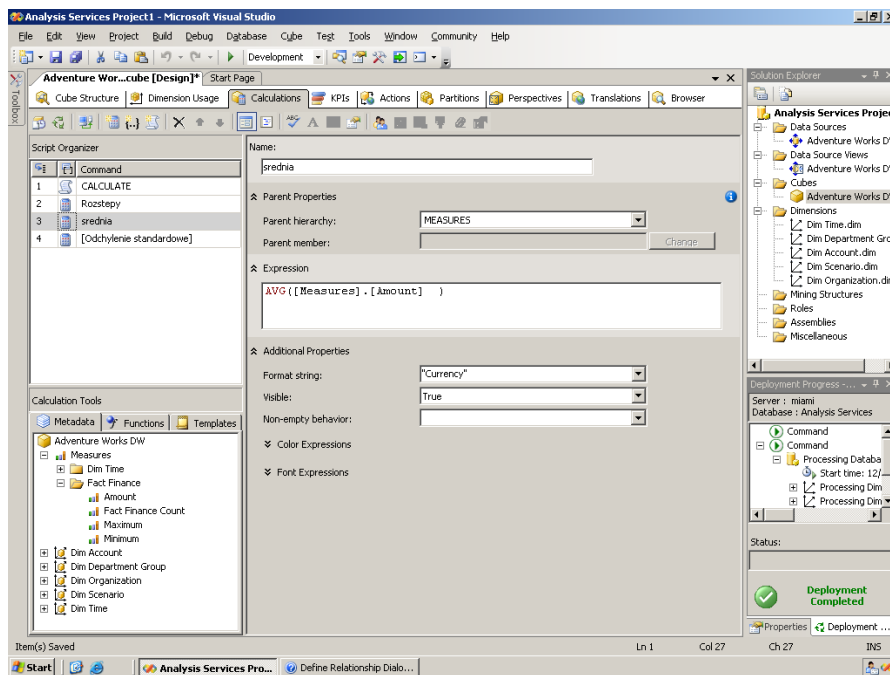
W tym miejscu mamy też możliwość usunięcia niepotrzebnych hierarchii.

- **Wartości obliczeniowe (Calculations)**

Domyślne procesy analizy zbiorów danych przez kreatora, zakłada realizację obliczeń, jedynie w zakresie ich podsumowania. W przypadku bardziej skomplikowanych mechanizmów obliczeniowych staje się konieczne zdefiniowanie swoich własnych metod obliczeniowych. Usługi analityczne udostępniają bardzo przejrzysty interfejs, umożliwiający zbudowanie dowolnych wyrażeń obliczeniowych.

W naszym projekcie postaramy się zdefiniować, mechanizm wyliczenia różnicy pomiędzy wartością sprzedaży minimum a maksimum (rozstęp), wartość średnią oraz odchylenie standardowe populacji. (średnie odchylenie od średniej).

W tym celu przełączamy się na zakładkę „calculations” i z paska menu wybieramy opcję „add calculate member”. Następnie definiujemy mechanizmy obliczeniowe. Bardzo przydatne w tym celu jest okno „Calculation tools”, w którym są dostępne wszystkie elementy naszej kostki analitycznej – wymiary oraz wartości obliczeniowe. Ponadto, w zakładce „Functions”, są pogrupowane wszystkie dostępne funkcje w usługach analitycznych. Natomiast zakładka „template” zawiera przygotowane szablony mechanizmów obliczeniowych. Wykorzystanie tych informacji polega na przeniesieniu, przy pomocy myszki, określonej wartości lub funkcji na pole obliczeniowe.



Rys 7. Tworzenie własnych elementów obliczeniowych

Po zdefiniowaniu mechanizmów obliczeniowych musimy również określić w jakim kontekście będą zdefiniowane obliczenia. Służą do tego opcje „Parent hierarchy” oraz „Parent member”. W naszym przypadku wartość obliczeniowa jest związana ze wszystkimi wartościami pomiarowymi. Ponieważ, wartości te nie mają struktury hierarchicznej, pole „Parent member” jest niedostępne.

Ponadto, w opcji „Additional Properties”, możemy zdefiniować atrybuty fontów, tła oraz formatowanie wyświetlanego ciągu znaków – w naszym wypadku wartości liczbowej.

Wartości obliczeniowe możemy zdefiniować następująco:

- „rozstęp” – [Measures].[Maximum]-[Measures].[Minimum]
- „srednia” – AVG([Measures].[Amount])
- „odchylenie standardowe – STDDEVP([Measures].[Amount])

Dla wartości obliczeniowej „rozstęp” oraz „średnia” możemy zdefiniować format walutowy a dla „odchylenia standardowego” zdefiniujemy format standard.

- **Wskaźniki wydajności (key performance indicator)**



W SQL2005 wprowadzono nowe mechanizmy sygnalizujące przekroczenie określonych wskaźników – wskaźników wydajności. Pojęciem to ma szersze znaczenie i oznacza możliwość zdefiniowania funkcji celu oraz trendu. KPI pozwalają w każdej chwili określić, czy badana wartość spełnia nasze oczekiwania oraz jak kształtuje się tendencja jej zmian.

W naszym projekcie zdefiniujemy KPI sprawdzający, czy wartość sprzedaży jest większa niż jej wartość średnia. Oczywiście w tym celu wykorzystamy wcześniej zdefiniowaną wartość obliczeniową – „średnia”. Dla tego wskaźnika zdefiniujemy „value expressions” oraz „status expressions” w następujący sposób:

- „value expressions” – [Measures].[Amount]-[Measures].[srednia]
- „status expressions” – [Measures].[Amount]-[Measures].[srednia]>0

A jako "status indicator" wykorzystamy ikonę „faces”.

KPI nie jest prezentowany w standardowej przeglądarce kostki analitycznej (omówionej w dalszej części). W pierwszym etapie należy uzupełnić kostkę analityczną o zdefiniowane elementy. W tym celu wykorzystamy opcję „process” – ikonę znajdującą się na pasku narzędziowym. Indykator możemy podejrzeć przełączając się od razu na widok KPI – ikona na pasku narzędziowym: „browser view”.

Display Structure	Value	Goal	Status	Trend	Weight
 Status sprzedaży	\$0.00				

Rys 8. Przykładowy podgląd KPI

- **Akcje (Actions)**

Kostka analityczna może być również uzupełniona o możliwość przypisania różnych rodzajów akcji w określonych wymiarach. Akcje mogą obejmować różne zadania: tworzenia łączy do zasobów Internetu, mechanizmów drażenia danych czy też realizacji raportów. Zadania możemy definiować dla określonych elementów: poczynając od całej kostki, danego wymiaru, hierarchii a kończąc na pojedynczej komórce. Ponadto, można również zdefiniować określone kryteria związane z wywołaniem danej akcji.

- **Partycje (Partitions)**

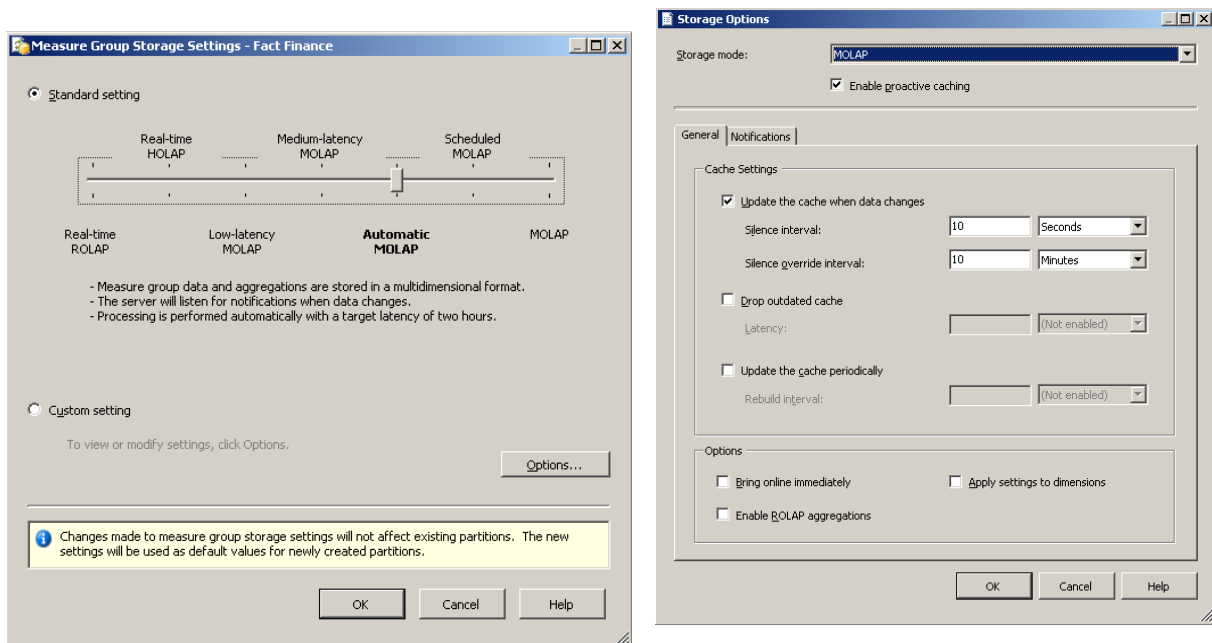
Sposób przechowywania informacji, wykorzystywanych w kostkach analitycznych może przybierać różną postać. W SQL 2005 wprowadzono zasadnicze różnice w budowie mechanizmu komunikacji ze źródłami danymi. Jednym z najważniejszych zmian, jest wprowadzenie aktywnych mechanizmów buforowania (proactive caching). Mechanizm ten pozwala budować aplikacje analityczne pracujące w trybie on-line.

W przypadku kostek analitycznych możemy zdefiniować kilka lokalizacji przechowywania danych – partycji. Każda z partycji może przechowywać dane w różnych strukturach. Wykorzystuje się w sytuacji, gdy dane archiwalne nie muszą być już aktualizowane. Wobec tego można wykorzystać struktury danych MOLAP (patrz poniżej). Dane bieżące przechowujemy w partycji zdefiniowanej jako np. HOLAP Real-time, dzięki temu możemy analizować zawsze aktualne informacje.

Wspomniane powyżej struktury przechowywania danych obejmuje następujące opcje:

- MOLAP – architektura przechowywania kostki analitycznej, w której wszystkie wartości pomiarowe i agregacje przechowywane są w bazie wielowymiarowej (OLAP). W tym trybie kostka nie reaguje na zmiany danych źródłowych a proces przebudowy kostki musi być wykonywany ręcznie.
- MOLAP sheduled – tryb pracy w którym wszystkie wartości pomiarowe i agregacje przechowywane są w bazie wielowymiarowej. Kostka nie reaguje na zmiany danych źródłowych. Proces budowy kostki jest inicjowany co 24 godziny.

- MOLAP automatic – wartości pomiarowe i agregacje przechowywane są w bazie MOLAP. Serwer kontroluje zmiany danych źródłowych i następuje pobranie danych do kostki co dwie godziny.
- MOLAP Medium-latency – wartości pomiarowe i agregacje przechowywane są w bazie MOLAP. Serwer kontroluje zmiany danych źródłowych i następuje pobranie danych do kostki co cztery godziny.
- MOLAP Low-latency – wartości pomiarowe i agregacje przechowywane są w bazie MOLAP. Serwer kontroluje zmiany danych źródłowych i następuje pobranie danych do kostki co trzydzieści minut.
- HOLAP Real-time – wartości pomiarowe przechowywane są w formacie baz relacyjnych (typowo: hurtownia danych). Wartości agregowane przechowywane są w bazie wielowymiarowej MOLAP. Kosta analityczna prezentuje aktualne dane pochodzące z baz relacyjnych.
- ROLAP Real-time – wartości pomiarowe oraz agregacje przechowywane są w bazach relacyjnych. Kostka analityczna prezentuje aktualne dane a serwer natychmiast reaguje na zmiany danych.



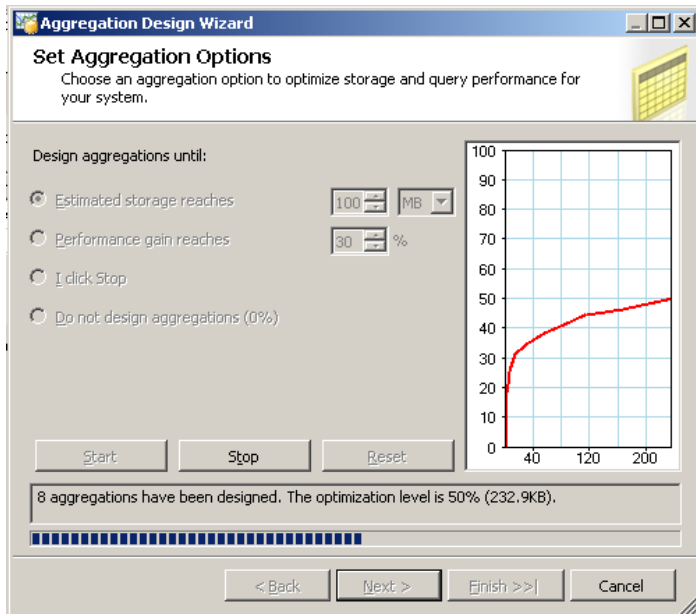
Rys 9. Definicja mechanizmów przechowywania danych

Ponadto, opcja „custom settings” umożliwia zdefiniowanie swoich własnych parametrów przechowywania danych w zakresie opóźnień związanych z prezentacją i wykorzystaniem proactive caching.

Ostatnim elementem konfiguracji przechowywania kostki analitycznej, jest możliwość zdefiniowania parametrów budowy agregacji. Poprzez uruchomienie kratora możemy precyzyjnie określić takie parametry jak: wielkość bazy z agregacjami czy też procent zagregowanych wartości (Uwaga: w kostce analityczne niekoniecznie muszą być wyliczone wszystkie agregacje. Jeżeli posiadamy agregacje na różnych poziomach hierarchii, to wartość globalna może być podsumowaniem wartości pośrednich, a nie wyliczeniem tej wartości atomowych).

Generowanie agregacji możemy realizować w oparciu o następujące opcje:

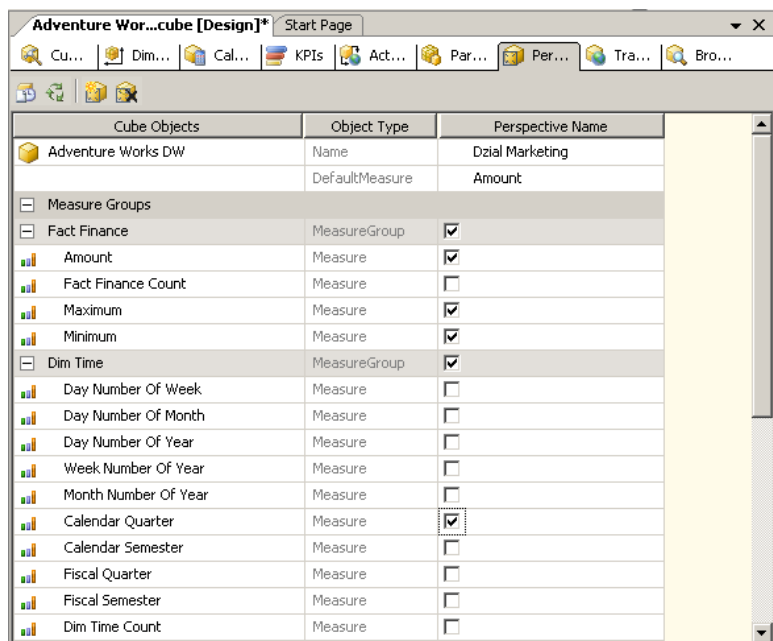
- Maksymalna wielkość bazy
- Poziom wydajności
- Ręczne zatrzymanie realizacji wydajności



Rys 10. Opcje definiowania agregacji

- **Perspektywa (Perspective)**

Zdefiniowana przez nas kostka analityczna może zawierać dziesiątki wartości pomiarowych oraz wymiarów. Jednak w wielu zastosowaniach użytkownicy mogą potrzebować tylko określone informacje związane z wartościami pomiarowymi czy też wymiarami opartymi na danej kostce. W celu ułatwienia korzystania z tak dużej kostki, możliwe jest zdefiniowanie, perspektywy – czyli widoku – zawierającego wymagane przez nas dane.

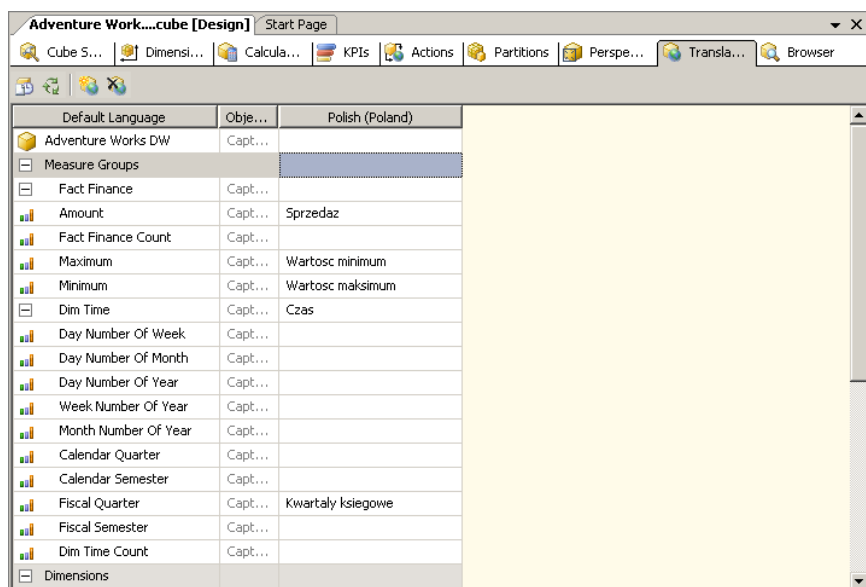


Rys 11. Tworzenie perspektywy

Utworzona perspektywa może zawierać wybrane przez nas wartości pomiarowe lub całe grupy pomiarowe, wymiary bądź poszczególne elementy w hierarchii wymiarów, wskaźniki wydajności oraz zdefiniowane mechanizmy obliczeniowe. Tak zdefiniowana perspektywa posiada swoją własną nazwę. Ponadto, istnieje możliwość nadawania użytkownikom uprawnień do poszczególnych perspektyw.

- **Translacja (Translation)**

W usługach analitycznych serwera SQL2005 wprowadzono bardzo proste techniki translacji opisów wymiarów i wartości pomiarowych na dowolny język. Procedura tworzenia mechanizmu translacji przebiega następująco: po wybraniu określonego języka, samodzielnie definiujemy tłumaczenia dla poszczególnych elementów.

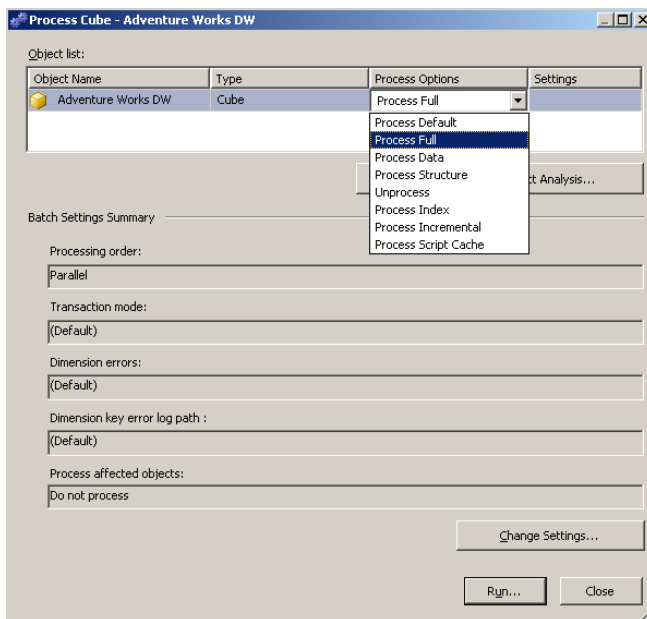


Rys 12. Translacja wymiarów i wartości pomiarowych na język polski

Wykorzystanie tak zdefiniowanych nazw u klienta wymaga określenia domyślnego języka wykorzystywanego przez klienta.

- **Przeglądarka kostek analitycznych (Browser)**

Mając przygotowane wszystkie elementy związane z kostką analityczną, możemy dokonać procesu budowy kostki oraz wyliczenia agregacji. W zależności od potrzeb, proces budowy kostki może obejmować przebudowę całej struktury kostki, realizację agregacji czy też przyrostową przebudowę kostki.



Rys 12. Generowanie kostki analitycznej

BI Development Studio ma wbudowaną przeglądarkę kostek analitycznych – browser. W lewej części okna znajdują się wszystkie dostępne wymiary oraz wartości pomiarowe. W centralnej części znajduje się okno organizacyjne, umożliwiające zbudowanie widoku do analizy danych. W oknie znajdują się pola umożliwiające umieszczenie wybranych wymiarów, czy też hierarchii. Dodatkowy obszar na górze przeglądarki umożliwia zbudowanie wyrażenia, dodatkowo filtrującego wybrane dane. Przeglądarka umożliwia dołączenie się do wybranej kostki lub perspektywy oraz wybranie domyślnego języka. Wybranie w naszym przykładzie języka polskiego, powoduje, że przetłumaczone pola wyświetlone są w języku polskim.

Oprócz podstawowych zadań, browser umożliwia wykonywanie zadań dodatkowych takich jak sortowanie danych, filtrowanie czy też formatowanie określonych komórek lub hierarchii.

- Efektem naszego projektu jest kostka analityczna, dla której dobraliśmy parametry, związane z wcześniej definiowanymi elementami:
- hierarchia – „rok kalendarzowy”
- perspektywa – „dział marketingu”
- język – „polski”

- wartości obliczeniowe – „rozstęp” oraz „średnia”
- dodatkowo – zdefiniowano formatowanie podsumowania

Dimension	Hierarchy	Operator	Filter Expression
<Select dimension>			
Drop Filter Fields Here			
	Scenario Name		
	Actual	Budget	Grand Total
Rok	Kwartal Miesiac	Rozstęp	Średnia
2001	3	\$2,533,485.50	\$49,874,474.31
	4	\$3,512,162.00	\$76,871,487.09
	Total	\$4,309,807.50	\$126,745,961.40
2002	1	\$2,607,159.00	\$70,058,876.95
	2	\$3,314,125.00	\$73,418,437.25
	3	\$5,259,479.00	\$136,436,173.26
	4	\$3,693,636.00	\$120,463,291.40
	Total	\$5,259,479.00	\$400,376,778.86
2003	1	\$2,941,272.00	\$92,774,845.17
	2	\$4,031,996.00	\$108,487,845.61
	3	\$5,894,222.00	\$169,108,038.92
	4	\$4,078,438.00	\$155,069,803.74
	Total	\$5,894,222.00	\$525,440,533.44
2004	1	\$3,195,055.00	\$122,763,103.13
	2	\$4,351,926.00	\$140,902,185.87
	Total	\$4,351,926.00	\$263,665,289.00
Grand Total		\$5,942,906.00	\$1,316,228,562.70
		\$893,300.00	\$42,411,850.00
		\$5,942,906.00	\$1,358,640,412.70

Rys 13. Efekt tworzenia kostki analitycznej

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