

# ITA-102 Hurtownie Danych

Marcin Gorawski, Sławomir Bańkowski

## Moduł 3

Wersja 1.0

# Usługi analiz danych: zunifikowany model wymiarowy

## Spis treści

Usługi analiz danych: zunifikowany model wymiarowy.....	1
Informacje o module.....	3
Przygotowanie teoretyczne .....	4
Przykładowy problem .....	4
Podstawy teoretyczne.....	4
Identyfikacja wymiarów dla każdej z tablic faktów.....	5
Fakty.....	5
Definicja atrybutów.....	6
<i>Model logiczny</i> .....	6
Model danych w schemacie gwiazdy.....	8
Model danych w schemacie płotka śniegu.....	9
Przykładowe rozwiązanie.....	12
Porady praktyczne .....	12
Uwagi dla studenta .....	13
Dodatkowe źródła informacji.....	13
Laboratorium podstawowe.....	14
Tworzenie widoku danych, definiowanie kostki i wymiarów (czas realizacji 20min).....	14
Definiowanie i publikowanie kostki danych (czas realizacji 20 min) .....	17
Edycja wymiarów, poziomów i hierarchii (czas realizacji 15 min) .....	20
Budowanie raportów za pomocą tabel przestawnych (czas realizacji 10 min) .....	25
Przeglądanie opublikowanej kostki za pomocą Microsoft SQL Server Management Studio (czas realizacji 10 min) .....	26
Laboratorium rozszerzone .....	28
Zadanie 1 (czas realizacji 10 min).....	28

Zadanie 2 (czas realizacji 10 min).....	28
Zadanie 3 (czas realizacji 15 min).....	28
Zadanie 4 (czas realizacji 15 min).....	28
Zadanie 5 (czas realizacji 40 min).....	28

## Informacje o module

### Opis modułu

W module znajdziesz informacje na temat modelu wielowymiarowego, w którym przechowywane są dane użytkownika. Przedstawiamy budowę kostki zarówno z agregatami, jak i z współdzielonymi wymiarami i poziomami hierarchicznymi. Opisujemy edycję schematu wielowymiarowego, dostosowując schemat do modelu biznesowego. Przedstawiamy tworzenie prostych raportów za pomocą tabel przestawnych.

### Cel modułu

Celem modułu jest przedstawienie procesu projektowania modelu wymiarowego na podstawie schematu relacyjnego, a także późniejszą modyfikację kostek danych. Umożliwia to analizę schematu relacyjnego, a następnie utworzenie schematu wielowymiarowego.

### Uzyskane kompetencje

Po zrealizowaniu modułu będziesz:

- znał model wielowymiarowy,
- znał zasady konwersji modelu relacyjnego do modelu wymiarowego,
- potrafił zaprojektować kostkę, wymiary, hierarchie, poziomy i atrybuty,
- wiedział, jak szybko utworzyć prosty raport za pomocą tabel przestawnych.

### Wymagania wstępne

Przed przystąpieniem do pracy z tym modulem powinieneś:

- znać podstawowe pojęcia związane z modelem relacyjnym,
- rozumieć pojęcia tabeli, kolumny, wiersza i krotki,
- wiedzieć, na czym polegają agregacje, widoki, widoki zmaterializowane i partycje.

### Mapa zależności modułu

Aby rozpocząć ten moduł nie jest konieczne ukończenie innych modułów.

## Przygotowanie teoretyczne

### Przykładowy problem

Adventure Works Cycles jest dużym, międzynarodowym zakładem przemysłowym, który produkuje i sprzedaje ramy rowerowe i inne zbliżone wyroby metalowe na rynki Ameryki Północnej, Europy oraz Azji. Podstawowe procesy produkcyjne Adventure Works Cycles zachodzą w Bothel, w Waszyngtonie, gdzie zatrudnionych jest 290 pracowników, natomiast kilka regionalnych oddziałów sprzedaży zlokalizowanych jest na całym rynku, na którym działa firma. W 2000 roku Adventure Works Cycles kupiło mały oddział produkcyjny – Importadores Neptuno, z siedzibą w Meksyku. Jego zadaniem była produkcja kilku istotnych komponentów dla linii produkcyjnej Adventure Works Cycles, które były transportowane do Bothel, gdzie miało miejsce finalne złożenie produktu. W 2001 roku Importadores Neptuno zostało wyłącznym producentem oraz dystrybutorem rowerów trekkingowych.

Mijający, zdecydowanie pomyślny dla firmy rok podatkowy stworzył dogodne warunki dla zwiększenia udziału w rynku, na którym działa Adventure Works Cycles, poprzez ukierunkowanie sprzedaży na najlepszego klienta, ułatwienie dostępu do produktów dzięki stronie internetowej oraz redukcji kosztów sprzedaży uzyskanej poprzez obniżenie kosztów produkcji.

Do sprawnego zarządzania przedsiębiorstwem niezbędne są analizy danych i bardzo sprawny system raportów dla podejmowania szybkich, strategicznych decyzji. Dane powinny być w postaci biznesowej, łatwej do zrozumienia dla klienta podejmującego decyzje zarządcze.

### Podstawy teoretyczne

Budowa modelu HD jest poprzedzona jest przygotowaniem planu tego projektu oraz fazą zgromadzenia specyfikacji stawianych przed systemem HD. Przed analitykami i projektantami HD zostaje postawiony problem rozwiązania zagadnień budowy:

- konceptualnego (pojęciowego) modelu biznesowego systemu na potrzeby użytkownicy aplikacji DSS ( ang. *DSS – Decision Support Systems* ),
- logicznego modelu danych spełniających założenia konceptualnego modelu biznesowego
- zagadnienia implementacji modelu fizycznego zoptymalizowanego pod kątem używanego sprzętu i oprogramowania ( systemu bazodanowego ) – zagadnienie to rozwiązują specjaliści.

Dobrym rozwiązaniem wspomnianych zagadnień jest hurtownia, która:

- będzie pozwalała znajdować odpowiedzi na postawione przez użytkowników zapytania,
- będzie efektywnie realizowała zlecone zadania,
- będzie łatwa w utrzymaniu i prosta w rozbudowie.

### **Pojęciowy, wielowymiarowy model biznesowy**

Z dwunastu kluczowych decyzji, które należy podjąć przy projektowaniu hurtowni danych, aż pięć pierwszych ściśle związanych jest z tworzeniem wielowymiarowego modelu biznesowego systemu. Są to wybory:

- procesów biznesowych, które będą podlegały analizie
- poziomu granulacji tablic faktów,
- wymiarów, dla każdej z tablic faktów,
- faktów: wartości dostępnych bezpośrednio oraz wartości wyliczanych,
- atrybutów wymiarów i zależności między nimi.

Powyższe wybory i tworzony biznesowy model danych są ściśle zależne od wymagań końcowych użytkownika. Zwykle już na etapie projektowania modelu pojęciowego (konceptualnego),

projektanci HD mają na uwadze zapewnienie odpowiedniej wydajności przetwarzania zapytań i zminimalizowanego kosztu zarządzania hurtownią.

Budowę modelu biznesowego rozpoczyna się od zebrania wymagań funkcjonalnych, które powinien realizować tworzony system HD. Etap ten powinien być globalny w skali przedsiębiorstwa i obejmować definicję zintegrowanego modelu biznesowego danych przedsiębiorstwa oraz definicję zakresu hurtowni danych.

Przyjęte ustalenia powinny prowadzić do utworzenia i zapisu conceptualnego, wielowymiarowego modelu biznesowego hurtowni w postaci diagramu, gdzie obiektami, które występują na takim diagramie są tablice faktów, wymiary hurtowni danych, atrybuty wymiarów oraz relacje między atrybutami wymiarów – powiązania i hierarchie.

Bardziej szczegółowy diagram modelu może dodatkowo zawierać:

- wyszczególnienie faktów przechowywanych w tablicy faktów,
- dostarczenie szczegółowego opisu atrybutów oraz ustalenie przewidywanej liczby elementów unikalnych,
- uzupełnienie relacji o ich krotności ( 1:1, 1:N, N:1, M:N ) wyliczone na podstawie wspomnianej wyżej liczby dostępnych elementów.

#### **Identyfikacja wymiarów dla każdej z tablic faktów.**

Słowo *wymiar* jest pojęciem-koncepcją stosowanym w nazewnictwie związanym z hurtowniami danych i oznacza pewien obszar zainteresowań lub kierunek potencjalnych analiz. Wymiary dostarczają perspektyw, z których końcowy użytkownik będzie mógł analizować dane zawarte w tablicach faktów. Z reguły systemy posiadają od 5 do 25 wymiarów. Na etapie identyfikacji wymiarów należy uważać, aby występujące w projektowanym modelu wymiary nie były z sobą skorelowane. Generalnie można przyjąć, że pomiędzy wymiarami powinna zachodzić relacja M:N. W przypadku, gdy taka zależność nie zachodzi, to istnieje podejrzenie, że można zrezygnować z jednego z wymiarów. Minimalny poziom granulacji wymiaru wyznaczony jest przez ustalony wcześniej poziom granulacji tablic faktów, do których dane wymiary zostaną przypisane. Z racji istnienia w projektowanym modelu wielu wymiarów oraz z racji możliwości późniejszej analizy danych w wielu perspektywach (wymiarach) tworzony model nazywa się modelem *wielowymiarowym*. Wybór wymiarów jest decydującym krokiem w projekcie tego modelu. Pozostałe dwie kluczowe decyzje (czwarta i piąta) służą wypełnieniu przygotowanych wcześniej obiektów – tabel faktów i wymiarów odpowiednią zawartością: faktami oraz atrybutami. Ponieważ wymiary hurtowni są tylko pewną koncepcją, to nie muszą one mieć fizycznego odpowiednika w samej hurtowni danych.

#### **Fakty.**

Wielkości i informacje charakteryzujące obserwowane procesy przechowywane są w tabelach faktów jako fakty. Wielkości te mogą być mierzone bezpośrednio ( np. dostępne informacje o poziomach sprzedaży, cenach, stanach magazynowych ) lub też mogą być pewnymi wartościami wyliczanymi ( np. ilość klientów, która w danym dniu zakupiła dany towar - wielkość wyliczona na podstawie rodzaju i ilości zawartych transakcji ). Fakty są podstawą budowy metryk, które dalej podlegają wszechstronnej analizie biznesowej i mogą służyć m.in. do monitorowania ważnych dla przedsiębiorstwa wskaźników ekonomicznych, badania skuteczności przeprowadzanej akcji promocyjnej itp. Ilość i rodzaj faktów przechowywanych w danej tablicy faktów wyznacza złożoność procesu, który ma podlegać analizie. Spotykane są też systemy, w których dla scharakteryzowania procesu nie potrzeba przechowywać żadnych faktów. Do opisu takich procesów wystarczy występowanie odpowiedniej kombinacji wymiarów, a tablice faktów takich procesów nazywa się tablicami bez faktów ( ang. factless ). Przykładem takiej tablicy może być tabela sygnalizująca

zajście pewnego zdarzenia lub sytuacji – już samo istnienie wpisu w takiej tablicy wystarczająco zdarzenie reprezentuje. Przy definiowaniu faktów należy pamiętać, że informacje trafiające do hurtowni danych mogą pochodzić z różnych systemów źródłowych. Dane, które już znajdują się w hurtowni danych powinny być przechowywane w jednolitej postaci, tzn. posiadać ten sam stopień granulacji być wyrażone w tym samych jednostkach, w tym samym formacie, itd.

### Definicja atrybutów.

W ramach każdego z wymiarów specyfikuje się występujące w nim atrybuty., które są pewnymi pojęciami opisującymi proces od strony jakościowej. Końcowemu użytkownikowi aplikacji DSS atrybuty pozwalają na zdefiniowanie wyglądu i poziomu szczegółowości raportu oraz kwalifikowanie wyników raportów (narzucanie ograniczeń, filtrowanie według jednego lub większej liczby atrybutów). Można przyjąć, że wymiary kwalifikują obliczenia na pewnym uogólnionym poziomie, a atrybuty na poziomach odpowiednio niższych. W skład wymiaru może wchodzić jeden lub więcej atrybutów. Między atrybutami wchodzącymi w skład danego wymiaru zachodzą pewne relacje. Relacje wyznaczają logiczne powiązania atrybutów w ramach wymiaru. Definicja relacji opisuje hierarchiczną zależność jednego atrybutu, tzw. dziecka ( np. *regionu* ) od drugiego zwanego rodzicem ( dla *regionu* będzie to *menadżer* ) oraz zawiera ilościową krotność zależności ( 1:1, 1:N, N:1, N:M ). Relacje między atrybutami są podstawą do zdefiniowania wielopoziomowej, logicznie uporządkowanej hierarchii atrybutów w wymiarze. W danym wymiarze może istnieć więcej niż jedna hierarchia. Wyróżniamy:

hierarchię główną ( np. *menadżer* → *region* → *sklep* ),

hierarchie atrybutów dodatkowych, tzw. *hierarchie charakterystyczne* budowane obok hierarchii głównej ( np. *typ sklepu* → *sklep*, *województwo* → *sklep* ).

Hierarchię główną tworzą atrybuty, które leżą w centrum zainteresowań użytkowników. Hierarchie charakterystyczne - budowane w oparciu o atrybuty hierarchii głównej - dostarczają dodatkowych możliwości bardziej szczegółowego opisu procesów. Krotności relacji oraz zależności rodzic-dziecko odznaczają się własnością przechodniości w ramach hierarchii. Atrybuty należące do różnych wymiarów pozostają z sobą w relacji tylko dzięki łączącej wymiary tabeli faktów. W kontekście aplikacji DSS, relacje zachodzące między atrybutami definiują ścieżki przeglądania oraz analizy danych - ich rozwijania ( ang. *drill down* ) i zwijania ( ang. *drill up* ). Analizując dany atrybut należy zwrócić uwagę na sposób jego opisu. Atrybut może posiadać wiele cech charakterystycznych. Elementy opisu atrybutu przyjmują konkretne wartości występujące w rzeczywistości biznesowej i są atomowymi komponentami modelu danych, przechowywane w postaci rekordu w odpowiedniej tabeli wymiaru. Elementy atrybutu dostarczają wartości, według których można dokonać kwalifikacji ( filtrowania ) wyników raportów.

### Model logiczny

Mając dopracowany i uzgodniony z końcowym użytkownikiem pojęciowy model biznesowy można przystąpić do budowy systemu HD. Takie systemy bazując na hurtowniach danych są zdolne do **analitycznego przetwarzania danych na bieżąco** (ang. *On-Line Analytical Processing – OLAP*).

**OLAP** cechuje zdolność:

- wykonywania analiz wielowymiarowych wg złożonych kryteriów wyszukiwania,
- interaktywnego raportowania bez znajomości języków programowania,
- uzyskiwania odpowiedzi na skomplikowane i często niestandardowe (ang. *ad hoc*) zapytania w trybie bieżącym.

OLAP zdeterminowany jest zakresem czasowym i liczebnością przechowywanych danych. Analizy wielowymiarowe powinny być przeprowadzane na jak największej liczbie danych gromadzonych w organizacjach. Do analiz przekrojowych, wg złożonych kryteriów wyszukiwania, potrzebne są dane z całego okresu działalności organizacji i takie dane muszą być dostępne. Celem OLAP jest generowanie przekrojowych raportów ze wszystkich dziedzin działalności organizacji wg różnych kierunków analizy np. wymiarów, hierarchii. Dla pełniejszej analizy stosuje się różne techniki przetwarzania analitycznego np. rachunek prawdopodobieństwa, wnioskowanie statystyczne, analizy: regresji, korelacji, trendów, wrażliwości, czy analizy typu „co, jeśli”.

Architektura OLAP transformuje różnorodne zbiory danych w zintegrowaną informację decyzyjną przy pomocy komponentów tj.: motor bazy danych, motor ekstrakcji danych, motor OLAP, motor eksploracji danych, repozytorium metadanych - muszą stanowić dobrze zintegrowany system. Jak na ironię, ta architektura sama w sobie stała się tak złożona, że stanowi źródło coraz poważniejszych problemów integracyjnych.

Architektura OLAP pozwala na prezentację danych w postaci wielowymiarowej, odmiennej od tradycyjnie stosowanej w bazach danych znormalizowanej struktury relacyjnej. O ile przedmiotem relacyjnego modelu danych są pojedyncze rekordy danych i pojedyncze transakcje wykonywane w danym momencie czasu, to w modelu wielowymiarowym są analizowane skutki wielu transakcji dokonywanych w pewnym przedziale czasu lub pewnym obszarze analizy.

Wszystkie architektury OLAP wymagają tworzenia wielowymiarowej struktury danych, w której wymiary reprezentują „obiekty” istotne z punktu widzenia prowadzonej analizy. W takiej strukturze danych, określanej potocznie jako „sześcián” lub „kostka” danych użytkownicy prowadzą analizy korzystając ze specjalizowanego motoru OLAP.

Motory OLAP pozwalają użytkownikom na prezentowanie dowolnych widoków danych sumarycznych, a także rozwijanie tych danych do poziomu danych szczegółowych.

Struktury wielowymiarowe mogą być posadowione w dedykowanych, trwałych bazach wielowymiarowych, w tymczasowych „mikro-kostkach” rezydujących w pamięci lub w bazach relacyjnych (w postaci schematów gwiazdy lub płátka śniegu).

W architekturze OLAP wyróżniamy silniki:

- **MOLAP** (OLAP wielowymiarowy; ang. Multidimensional On-Line Analytical Processing Multidimensional,) – do prowadzenia analiz wykorzystuje serwer wielowymiarowej bazy danych MDDDB (ang. Multidimensional Database). W tej architekturze dane muszą być przechowywane wielowymiarowo, jeżeli mają być wielowymiarowo analizowane i prezentowane. Cechą charakterystyczną tych baz danych jest pełna agregacja danych.
- **DOLAP** (OLAP Desktop)– to silnik MOLAP operujący na plikach lub bazach danych zainstalowanych lokalnie, na stacjach klienta (brak serwera).
- **ROLAP** (OLAP relacyjny; ang. Relational On-Line Analytical Processing) – do prowadzenia analiz i wielowymiarowej prezentacji korzysta z danych przechowywanych w relacyjnej bazie danych (RDBMS) o specjalizowanej strukturze, najczęściej gwiazdy (ang. Star) lub płátka śniegu (ang. Snowflake). Dostęp do danych jest realizowany na poziomie zwykle złożonych zapytań SQL. Dane mogą być wstępnie agregowane lub potrzebne agregaty są wyliczane w trakcie wyszukiwania danych.
- **HOLAP** (OLAP hybrydowy; ang. Hybrid OLAP) – to architektura mieszana pozwalająca użytkownikom na korzystanie (nie zawsze jednocześnie) z dwóch baz danych: wielowymiarowej lub relacyjnej. HOLAP wzbogaca architekturę OLAP o mechanizm semafora serwerów ROLAP i MOLAP.

Ze względu na sposób przechowywania i przetwarzania danych najważniejsze są dwie architektury MOLAP i ROLAP, dwie pozostałe są ich pochodnymi.

### Model danych w schemacie gwiazdy.

W odróżnieniu od systemów transakcyjnych, które zazwyczaj starają się opisać rzeczywistość całościowo, podstawowe składniki systemu OLAP koncentrują się na pojedynczym procesie, którego przetwarzanie (analiza) powinna być maksymalnie efektywna. Podstawowym modelem realizującym ten postulat jest model zwany gwiazdą (ang. *star*). Struktura ta następująco odwzorowuje obiekty modelu biznesowego w model logiczny:

- obiekty przechowujące fakty są zapisywane w postaci tablic faktów
- wymiary i ich atrybuty są zapisywane w postaci tablic wymiarów, po jednej tablicy na każdy wymiar.

Cechami charakterystycznymi tej struktury jest to, że model nie jest jednorodny (posiada wyraźne centrum – tabelę bazową), na schemacie występuje minimalna liczba połączeń oraz tabele wymiarów nie są znormalizowane.

Zasadniczo można spotkać się z trzema typami struktur typu gwiazda, różniącymi się sposobem budowy tablic wymiarów. Każda z tych struktur może dodatkowo występować w dwóch odmianach zależnych od przyjętej strategii agregacji: z dodatkowymi tablicami faktów lub też z wykorzystaniem pojedynczej tablicy i dodatkowego pola wskazującego na poziom agregacji.

Zastosowanie modelu gwiazdy jako schematu hurtowni danych ma wiele zalet:

- model jest prosty i czytelny – w schemacie występuje niewielka liczba obiektów oraz minimalna liczba połączeń:
  - końcowemu użytkownikowi bardzo łatwo jest sobie przyswoić i zrozumieć ten model,
  - symetria układu powoduje, że każdy wymiar jest traktowany identycznie, podobnie wyglądają zapytania *SQL* kierowane do modelu i łatwo budować na nim aplikacje *DSS*,
  - ze względu na małą liczbę tabel uproszczona jest administracja systemem,
- zoptymalizowany jest dostęp do danych:
  - w zapytaniach pojawia się niewielka liczba złączeń – schemat odznacza się najwyższą wydajnością przy przeglądaniu atrybutów należących do tego samego wymiaru,
  - złączenia tablic wymiarów z tablicą faktów następują tylko z wykorzystaniem kluczy numerycznych,
  - istnieje prosty algorytm wyznaczania wynikowego zbioru rozwiązań zapytania *SQL* kierowanego do struktury gwiazdy,
  - istnieją specjalne techniki optymalizacji zapytań (indeksy bitmapowe, specjalna modyfikacja zapytań do schematu wykorzystująca tzw. transformację gwiazdzistą – ang. *star transformation*),
- model jest bardzo efektywny w zakresie przechowywania danych: w tablicy faktów oprócz informacji merytorycznych przechowywane są tylko identyfikatory numeryczne pochodzące z tabel wymiarów;
- rozbudowa modelu nie stanowi problemu:
  - model można rozbudowywać przez dodanie nowych faktów, wymiarów czy atrybutów wymiarów – co w żaden sposób nie zmienia podstawowej idei schematu,
  - rozbudowa modelu zasadniczo nie wpływa na ograniczenie funkcjonalności już wdrożonych systemów aplikacji *DSS* - nowy model jest pod względem schematu zgodny w dół ze starym. Modyfikacji natomiast będą musiały podlegać programy ładujące dane do hurtowni.
- model gwiazdy jest uniwersalny – w każdym modelu da się wyróżnić tabelę bazową z faktami oraz tabele wymiarów, stąd łatwo jest stworzyć narzędzia do projektowania systemów *DSS*



zapewniające możliwość współpracy z dowolnym modelem hurtowni przypominającym strukturę gwiazdy,

- model gwiazdy stanowi *de facto* standard wśród schematów hurtowni danych, jest on wykorzystywany od lat i znane są jego dobre i słabsze strony – dopracowano się też zbioru standardowych rozwiązań obsługujących sytuacje złożone oraz nietypowe.

Zalety modelu gwiazdy znacznie przewyższają jego wady i czynią go jednym ze standardów budowy schematu hurtowni danych.

### **Model danych w schemacie płatka śniegu.**

Wymienione wyżej wady modelu gwiazdy stara się wyeliminować model płatka śniegu ( ang. *snowflake* ) nazywany też modelem rozwinętej gwiazdy ( ang. *extended star* ). Zasadniczą zmianą, którą wprowadzono w tym modelu jest odejście od pełnej denormalizacji tablic wymiarów w kierunku ich normalizacji. W zależności od stopnia znormalizowania tablic wymiarów można wyróżnić cztery typy modeli płatka śniegu.

Model płatka śniegu wnosi następujące zalety, których brakowało w modelu gwiazdy:

- tablice wymiarów są znacznie lepiej skalowalne ( cecha ta nie dotyczy modelu płatka śniegu typu 3 ),
- łatwo obsłużyć relację M:N,
- w naturalny sposób można zrealizować agregację danych za pomocą dodatkowych tabel faktów zagregowanych, gdyż nie są konieczne żadne dodatkowe tablice wymiarów.

Dodatkowo model płatka śniegu wnosi następujące korzyści:

- zmniejszenie objętości bazy danych poprzez eliminację powtórzeń wartości atrybutów wymiarów – cecha ta dotyczy tylko modeli płatka śniegu typu 1 i 2, w modelu typu 3 może nastąpić zwiększenie ilości wymaganego miejsca.
- uzyskanie eleganckiego schematu jawnie ukazującego rzeczywistą, hierarchiczną strukturę wymiarów, np. czasu ( *rok* → *miesiąc* → *dzień* ), położenia geograficznego, asortymentów produktów ( podział wg *kategorii*, *grup*, *klas* ), czy też takich struktur jak struktury organizacyjne przedsiębiorstwa,
- uproszczenie niektórych zadań administracyjnych: tabele mają prostszą strukturę i są zazwyczaj mniejsze ( uwaga ta nie dotyczy modelu płatka śniegu typu 3 ):
  - tabele łatwiej się wypełnia danymi, składa i odtwarza,
  - w prostszy sposób można sprawdzić jakość informacji zawartej w tabeli,
- zwiększenie wydajności w przetwarzaniu zapytań typu: wyświetlenie listy dostępnych elementów atrybutów – w przypadku płatka śniegu możliwe jest użycie polecenia *select* w miejsce *select distinct* używanego w modelu gwiazdy [3],

Niestety normalizacja schematu powoduje utratę części zalet, które występowały w modelu gwiazdy:

- normalizacja może wywołać pewne problemy efektywnościowe, które nie występowały w modelu zdenormalizowanym – generalnie w planie wykonania zapytania pojawia się większa ilość złączeń [3]; uwaga dotyczy tylko modelu płatka śniegu typu 1 i 2, model płatka śniegu typu 3 jest w realizacji niektórych zapytań bardziej wydajny niż model gwiazdy,

- normalizacja zakłada maksymalną prostotę modelu gwarantowaną przez występowanie w nim minimalnej liczby obiektów i połączeń, przez co aplikacja DSS musi generować bardziej złożone frazy SQL,
- fakt zmniejszenia objętości bazy danych w przypadku modelu znormalizowanego jest bardzo niewielki; przy tabelach faktów wielkości gigabajtów oszczędności miejsca sięgają zwykle ułamków procenta, przy modelu płatka śniegu typu 3 - ilość potrzebnego może powiększyć się,
- pojawienie się większej ilości tabel w modelu płatka śniegu rozszerza zakres obowiązków administratora.

Model płatka śniegu wraz z modelem gwiazdy stanowią zestaw dwóch podstawowych modeli używanych przy tworzeniu schematu hurtowni danych. Modele te są sobie pokrewne i przez to nie wykluczają się wzajemnie – a wprost przeciwnie – ich podobieństwo pozwala na budowę modelu pośredniego posiadającego najlepsze cechy obu tych modeli. W zależności od specyfiki danego wymiaru: liczby jego atrybutów, liczby elementów każdego z atrybutów, liczby hierarchii i rodzaju występujących relacji można zdecydować się na jego implementację w schemacie gwiazdy lub płatka śniegu lub też w schemacie będącym kombinacją obu tych technik. Współczesne narzędzia do tworzenia aplikacji DSS bez problemu obsługują oba te modele oraz wszelkie modele pośrednie.

Powyższą wiedzę w wersji znacznie rozszerzonej zawierają następujące prace:

- **Frączek J., Gorawski M., Kozielski St.** Modelowanie struktur wielowymiarowych w hurtowniach danych. *Archiwum Informatyki Teoretycznej i Stosowanej*. 2000 t. 12 z. 3, 173-201, strony 173-201.
- **Gorawski, M.** Praktyczne aspekty projektowania hurtowni danych (Practical aspects of data warehouse design). *Studia Informatica*, 2003, vol. 24, nr 4(56). 2003, strony 189-210.
- **Gorawski, M.** Ocena efektywności architektur OLAP (Estimation of the efficiency of OLAP architecture). [red.] Z. Mazur. Z. Huzar. *Problemy i metody inżynierii oprogramowania*. Warszawa, Wydaw. Naukowo-Techniczne. 2003, strony 233-250.
- **Gorawski, M.** Definiowanie schematów rozszerzonej gwiazdy kaskadowej. [aut. książki] Praca zbiorowa. [red.] St. Kozielski [i in.]. *Bazy danych. Nowe technologie. Architektura, metody formalne i zaawansowana analiza danych*. Warszawa: Wydaw. Komunikacji i Łączności. 2007, strony 103-114.
- **Gorawski, M.** Laboratorium hurtowni danych poziomu MS SQL SEVER 2005. [red.] Z. Huzar. B. Hnatkowska. *Metody i narzędzia wytwarzania oprogramowania. Konferencja naukowa, Szklarska Poręba*. Wrocław: Oficyna Wydaw. Politechniki Wrocławskiej. 2007, strony 583-596.

**Microsoft SQL Server 2008 Analysis Services (SSAS)** ułatwia tworzenie złożonych procesów analitycznych w trybie on-line (OLAP, ang. *Online Analytical Processing*) i rozwiązań z dziedziny eksploracji danych. Narzędzia dostępne poprzez Analysis Services umożliwiają projektowanie, tworzenie i zarządzanie kostkami oraz modelami analitycznymi pochodzącymi z hurtowni danych. Dzięki odpowiednim narzędziom klienci mogą także uzyskać dostęp do danych OLAP oraz wyników analiz dataminingowych. Narzędzia Analysis Services łączą aspekty tradycyjnego przetwarzania OLAP oraz raportowania relacyjnego poprzez dostarczenie projektantom i użytkownikom możliwości zdefiniowania jednolitego modelu wymiarów i faktów korzystającego z jednego lub wielu źródeł danych. Powoduje to, że wszystkie zapytania użytkowników OLAP, usług raportujących i innych aplikacji BI (ang. *Business Intelligence*) mają zagwarantowany dostęp do różnych źródeł danych poprzez jeden wspólny widok biznesowy. Analysis Services dostarczają

bogaty zestaw algorytmów eksploracji danych umożliwiając użytkownikom biznesowym analizę danych pod kątem wykrywania specyficznych wzorców i trendów. Algorytmy eksploracji danych są stosowane do jednolitego modelu wymiarów i faktów lub bezpośrednio do danych z określonego źródła danych.

Struktura Analysis Services składa się z komponentów serwerowych i klienckich. Działanie tych komponentów jest zintegrowane w celu dostarczenia usług OLAP i funkcjonalności analitycznej dla aplikacji z dziedziny BI.

Komponent serwerowy Analysis Services jest zaimplementowany jako usługa Microsoft Windows. SQL Server 2008 Analysis Services wspiera współdziałanie wielu instancji Analysis Services na jednym komputerze (każdy serwis jest wtedy odrębną usługą systemu Windows).

Klienci komunikują się z serwisami Analysis Services używając standardu odmiany języka XML przeznaczonej do analiz, XML/A (ang. *XML for Analysis*) – protokołu opartego na SOAP służącego do wysyłania poleceń i odbierania odpowiedzi, działającego jako Usługa Web. Modele obiektów klienckich są także dostępne ponad XMLA i mogą być obsługiwane przy użyciu technologii określonego dostawcy (jak np. ADOMD.NET) lub przez natywnego dostawcę OLE DB.

Zapytania mogą być zapisywane za pomocą następujących języków: SQL, MDX (ang. *Multidimensional Expressions* – język zapytań zorientowany na analizę danych), DMX (ang. *Data Mining Extensions* – język zapytań zorientowany na eksplorację danych). Język skryptowy ASSL (ang. *Analysis Services Scripting Language*) jest używany do zarządzania obiektami bazy danych Analysis Services.

Analysis Services umożliwiają także pracę w aplikacjach klienckich odłączonych od serwera. Odpowiednie mechanizmy zapewniają możliwość przeglądania lokalnie przechowywanych danych wielowymiarowych.

#### *Komponenty Analysis Services, architektura serwera AS*

Komponent serwerowy Analysis Services działa jako serwis systemu Windows (jest uruchamiany za pomocą pliku **msmdsrv.exe**). Aplikacja ta składa się z komponentów bezpieczeństwa, komponentu do nasłuchu XML/A, komponentu procesora zapytań oraz licznych komponentów wewnętrznych, które wykonują następujące zadania:

- odbiór i analiza zapytań klienta
- zarządzanie zbiorem metadanych
- utrzymywanie transakcji
- wykonywanie obliczeń
- zapisywanie danych
- obliczanie i zapis danych zagregowanych
- kolejkowanie zapytań
- zachowywanie obiektów w pamięci podręcznej
- zarządzanie zasobami serwera

#### *Architektura klienta*

Analysis Services wspiera architekturę cienkiego klienta (klienta niezależnego od systemu operacyjnego, działającego w oparciu o przeglądarkę stron WWW). Silnik Analysis Services jest zlokalizowany całkowicie na serwerze, zaś wartości wszystkich zapytań są obliczane przez komputer, na którym pracuje serwer. Dzięki zastosowaniu takiej architektury możliwa jest pojedyncza komunikacja na linii klient-serwer podczas obliczania i prezentowania wyników zapytania. Pozwala to na osiągnięcie dużej skalowalności, a co za tym idzie, umożliwia wykonywanie i prezentację wyników skomplikowanych zapytań.

Głównym, rodzimym protokołem Analysis Services jest język XML rozszerzony w celu wspierania analizy danych (XMLA). Analysis Services dostarczają kilku interfejsów dostępowych dla aplikacji klienckich, jednak wszystkie te komponenty komunikują się z instancją Analysis Services za pomocą XMLA.

Analysis Services oferuje różnych dostawców, którzy zapewniają obsługę komunikacji dla różnych języków programowania. Dostawca komunikuje się z Analysis Services poprzez przesyłanie i odbieranie XMLA w pakietach SOAP za pomocą protokołu TCP/IP lub HTTP przy wykorzystaniu Internet Information Services (IIS). Połączenie HTTP wykorzystuje obiekt COM powołany do życia przez IIS, tzw. *pompę danych* (ang. *data pump*), która działa na zasadzie bezpośredniego połączenia z danymi Analysis Services. Pompa danych nie analizuje danych przesyłanych przez protokół HTTP.

### *Pojęcie kostki danych*

W dziedzinie analizy danych *kostka* (ang. *cube*) jest zbiorem powiązanych faktów i wymiarów używanym do analizy danych. *Fakt* (pomiar) jest wartością transakcyjną lub pomiarem, którą użytkownik może agregować. Fakty (pomiar) są zapisywane w kolumnach w jednej lub więcej tabel źródłowych i są grupowane w grupy faktów.

*Wymiar* jest grupą atrybutów, które reprezentują dziedzinę zainteresowania związaną z faktami zawartym w kostce. Atrybuty są wykorzystywane do analizowania i interpretowania wyników analiz faktów. Przykład: wymiar Klient może zawierać atrybuty takie jak: Imię i Nazwisko (Nazwa), Płeć oraz Miasto. Atrybuty są przechowywane w jednej lub większej liczbie tabel źródłowych. Atrybuty w każdym z wymiarów mogą być zorganizowane w hierarchie w celu dostarczenia ścieżek analiz.

Po zbudowaniu, kostka danych jest rozszerzana o wyniki obliczeń, perspektywy oraz translacje. W przypadku Analysis Services kostka danych jest utożsamiana z przedstawionym wcześniej jednolitym modelem wymiarów i faktów UDM (ang. *Unified Dimensional Model*). Analysis Services pozwalają na projektowanie i tworzenie kostki danych w oparciu o tabele i widoki modelowane w widoku danych źródłowych.

### **Przykładowe rozwiązanie**

Na podstawie bazy relacyjnej firmy Adventure Works Cycles należy zaprojektować schemat wielowymiarowy składający się z kostek danych. Najpierw należy przeanalizować schemat źródłowy, wybrać tabele faktów i potencjalne tabele wymiarów. Wybór tabeli faktów wiąże się z określeniem granulacji i agregatów. Kolejno, dla każdego wymiaru wybiera się poziomy, a na nich tworzy się hierarchie. W razie potrzeby można dowolnie modyfikować schemat wielowymiarowy. Na końcu za pomocą tabel przestawnych należy zaprojektować potrzebne raporty. Wygenerowane w ten sposób raporty mogą służyć do generowania zestawień, wykresów i przekrojowych analiz.

### **Porady praktyczne**

#### ***Uwagi ogólne***

- Podczas projektowania kostki danych zawsze ustal granulację każdej kostki: jakie dane są faktami i jakie agregaty można określić dla jednego faktu. Bardzo częstym błędem przy projektowaniu kostek danych jest niewłaściwa granulacja, przez co mimo braku błędów logicznych kostka jest całkowicie nieprzydatna.
- Określanie wymiarów dla kostki jest istotnym krokiem w projektowaniu kostki. W większości przypadków liczba wymiarów kostki pokrywa się z liczbą kluczy obcych.
- Wszystkie nazwy są widoczne z pozycji użytkownika końcowego. Z tego powodu, w odróżnieniu od schematu relacyjnego, używaj nazw znaczących ze spacjami jeżeli tylko potrzeba. Nazwę UnitPriceDctPct zmień na Unit Price Discount Percentage.
- Panele i okna Visual Studio posiadają bufor pamięciowe. Jeżeli informacja nie pokazuje się mimo zmiany, zamknij okno i otwórz je jeszcze raz, a zobaczysz zmianę.

## Uwagi dla studenta

Jesteś przygotowany do realizacji laboratorium jeśli:

- umiesz odczytać schemat relacyjny (ERD)
- rozumiesz, jak ze schematu relacyjnego powstaje schemat wielowymiarowy
- wiesz z jakich elementów zbudowany jest schemat wielowymiarowy (UDM)
- jesteś w stanie wymienić komponenty Analysis Services

Przed wykonywaniem każdego punktu dokładnie przeczytaj cały punkt, to spowoduje znaczne zmniejszenie liczby pomyłek podczas laboratorium i pozwoli na dokładniejsze poznanie materiału.

Pamiętaj o zapoznaniu się z uwagami i poradami zawartymi w tym module. Upewnij się, że rozumiesz omawiane w nich zagadnienia. Jeśli masz trudności ze zrozumieniem tematu zawartego w uwagach, przeczytaj ponownie informacje z tego rozdziału i zajrzyj do notatek z wykładów.

## Dodatkowe źródła informacji

1. Klaus Aschenbrenner, Beginning SQL Server 2008 for Developers: From Novice to Professional, Apress, 2008
2. Erik Thomsen, OLAP Solutions: Building Multidimensional Information Systems, Wiley, 2002
3. Ralph Kimball, Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, Wiley, 2002
4. Teo Lachev, Applied Microsoft Analysis Services 2005: And Microsoft Business Intelligence Platform, Prologika Press, 2005
5. Sivakumar Harinath, Robert Zare, Sethu Meenakshisundaram, Matt Carroll, Professional Microsoft SQL Server Analysis Services 2008 with MDX, Wrox, 2008
6. W. Lehner, J. Albrecht, H. Wedekind - Normal Forms for Multidimensional Databases, 10th International Conference on Scientific and Statistical Database Management, Proceedings, Capri, Italy, July, 1998. IEEE Computer Society 1-3, 1998.
7. P. Vassiliadis - Modeling Multidimensional Databases, Cubes and Cube Operations. Proceedings, Capri, Italy, July, 1998. IEEE Computer Society 1-3, 1998.
8. E. Franconi, U. Sattler - A Data Warehouse Conceptual Data Model for Multidimensional Aggregation. Proc. of the Int. Workshop on Design and Management of Data Warehouse, 1999.
9. M. Gyssens, L. Lakshmanan - A Foundation for Multi-Dimensional Databases. Proc. of the 23rd VLDB Conference, 1997.



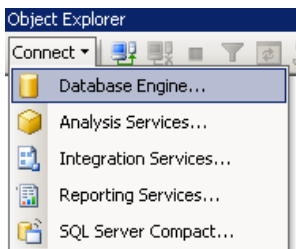
## Laboratorium podstawowe

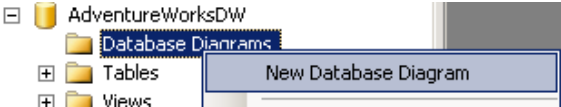

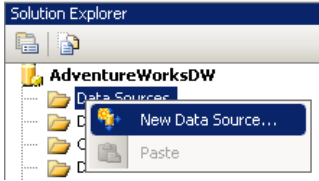
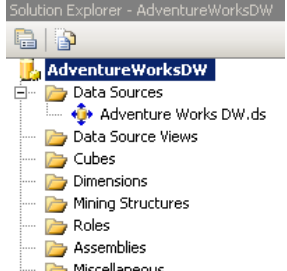
Zarząd Adventure Works Cycles zamierza usprawnić zarządzanie firmą. W tym celu potrzebna jest hurtownia danych pozwalająca na zamodelowanie relacyjnych danych w biznesową postać danych. Potrzebne jest szybkie generowanie raportów podsumowujących za pomocą tabel przestawnych. Utworzone kostki danych pomogą także w późniejszej analizie danych (zaawansowane raporty, eksploracja danych).

### Tworzenie widoku danych, definiowanie kostki i wymiarów (czas realizacji 20min)

W tej części zaprezentujemy schemat danych źródłowych, zdefiniujemy połączenie z bazą danych, wybierzemy interesujące tabele, które w kolejnym ćwiczeniu będą tworzyć wymiary kostki.

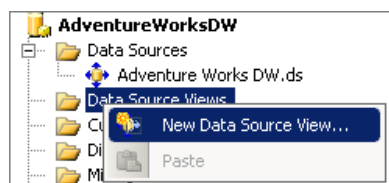
Dane wykorzystywane do budowy kostki danych mogą pochodzić z wielu różnego typu źródeł. Utworzenie widoku źródła danych pozwala na wybranie tych tabel, które będą potrzebne do zbudowania kostki, oraz odfiltrowanie pozostałych tabel, często niepotrzebnie komplikujących postać źródła danych. Po wybraniu tabel ze źródła danych istnieje możliwość przypisania im tzw. przyjaznych nazw (ang. friendly names). Przyjazne nazwy mogą się na początku wydać niepotrzebne, jednak jeśli ma się do czynienia z bazą danych, w której tabele są nazwane skrótami (często podobnymi do siebie), wtedy możliwość nadania friendly names jest bardzo przydatna.

Zadanie	Tok postępowania
1. Utworzenie projektu z szablonu Analysis Services	<ul style="list-style-type: none"><li>Uruchom aplikację SQL Server Business Intelligence Development Studio z grupy <b>Microsoft SQL Server 2005</b>.</li><li> Uruchamia się Microsoft Visual Studio z zainstalowanymi możliwościami tworzenia projektów BI dla SQL Server 2008.</li><li>Z menu głównego wybierz <b>File -&gt; New -&gt; Project</b>.</li><li>Z listy <b>Project Types</b> wybierz <b>Business Intelligence Projects</b>.</li><li> Jeżeli na komputerze jest już zainstalowany Microsoft Visual Studio w wersji 2003, 2005 lub 2008, to oprócz wspomnianego Business Intelligence Projects na liście może pojawić się więcej niż jeden typ projektu.</li><li>W <b>Templates</b> wskaż <b>Analysis Services Project</b>.</li><li>W polu <b>Name</b> podaj nazwę projektu (np. <b>AdventureWorksDW</b>).</li><li>Kliknij <b>OK</b>.</li></ul>
2. Przeglądanie źródła danych	<ul style="list-style-type: none"><li>Otwórz SQL Server Management Studio.</li><li>W <b>Object Explorer</b> wybierz <b>Connect</b>, a następnie <b>Database Engine</b>.</li></ul>  <p>Rys. 1 Połączenie do bazy danych w Object Explorer</p> <ul style="list-style-type: none"><li>Jako nazwę serwera wpisz <b>localhost</b> i kliknij <b>Connect</b>.</li><li>Rozwiń <b>Databases -&gt; Adventure Works DW</b>.</li><li>Rozwiń <b>Database Diagrams</b>.</li><li>Jeżeli pokaże się okienko o konieczności dodania nowych tabel do przechowywania diagramów, kliknij <b>Yes</b>.</li></ul>

	<ul style="list-style-type: none"> <li>Kliknij prawym przyciskiem myszy <b>Database Diagrams</b> i wybierz <b>New Database Diagram</b>.</li> </ul>  <p>Rys. 2 Tworzenie nowego diagramu dla bazy danych</p> <ul style="list-style-type: none"> <li>Zaznacz wszystkie dostępne tabele oprócz <b>AdventureWorksDWBuildVersion</b> i kliknij <b>Add</b>.</li> </ul> <p> Zaznaczenie wszystkich tabel można łatwo uzyskać poprzez zaznaczenie pierwszej tabeli, a następnie kliknięcie na ostatniej z przytrzymanym klawiszem <b>Shift</b>. Aby dodać lub usunąć tabelę z zaznaczenia należy kliknąć na tabelę z przytrzymanym klawiszem <b>Ctrl</b>.</p> <ul style="list-style-type: none"> <li>Kliknij <b>Close</b>.</li> <li>Na diagramie znajdź tabelę <b>FactInternetSales</b> i sprawdź, do jakich tabel prowadzą klucze obce.</li> <li>Zapisz diagram pod nazwą <b>WarehouseSchema</b> (<b>File -&gt; Save Diagram</b>).</li> <li>Zamknij SQL Server Management Studio.</li> </ul>
<p>3. Definiowanie źródła danych</p>	<ul style="list-style-type: none"> <li>Przejdź do Microsoft Visual Studio.</li> <li>W oknie <b>Solution Explorer</b> kliknij prawym przyciskiem myszy <b>Data Sources</b> i z kontekstowego menu wybierz <b>New Data Source</b>.</li> </ul>  <p>Rys. 3 Tworzenie nowego źródła danych</p> <ul style="list-style-type: none"> <li>W oknie dialogowym <b>Select how to define the connection</b> wybierz przycisk <b>New</b>.</li> <li>W dialogu <b>Connection Manager</b> podaj szczegóły połączenia: <ul style="list-style-type: none"> <li>– <b>Provider:</b> Native OLE DB\Microsoft OLE DB Provider for SQL Server.</li> <li>– <b>Server name:</b> localhost.</li> <li>– <b>Use Windows Authentication</b></li> <li>– <b>Select or enter a database name:</b> AdventureWorksDW.</li> </ul> </li> <li>Kliknij <b>OK</b>.</li> <li>Kliknij <b>Next</b>, wybierz <b>Use the service account</b> i kliknij <b>Next</b>. Kliknij <b>Finish</b>.</li> </ul>  <p>Rys. 4 Utworzone źródło danych</p>

#### 4. Definiowanie widoku źródła danych dla projektu Analysis Services

- Na panelu **Solution Explorer** kliknij prawym przyciskiem myszy **Data Source Views** i wybierz **New Data Source View**.



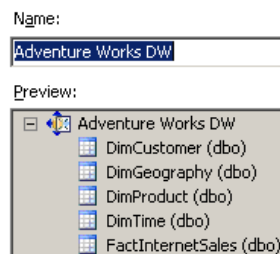
Rys. 5 Tworzenie nowego widoku źródła danych

- Z listy **Relational data sources** wybierz źródło **Adventure Works DW** i kliknij **Next**.
- Przytrzymując klawisz **Ctrl** wybierz interesujące tabele:
  - **DimCustomer**
  - **DimGeography**
  - **DimProduct**
  - **DimTime**
  - **FactInternetSales**



Należy dokładnie sprawdzić, jakie tabele zostały zaznaczone. Na liście jest kilka tabel o podobnych nazwach, więc nie trudno o pomyłkę. Aby odznaczyć tabelę, należy jeszcze raz ją kliknąć przytrzymując klawisz **Ctrl**.

- Za pomocą przycisku **>** przenieś wybrane tabele do listy **Included Objects**.
- Kliknij **Next** i później **Finish**.



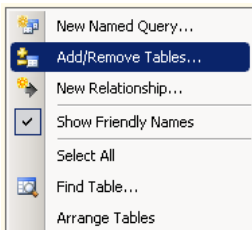
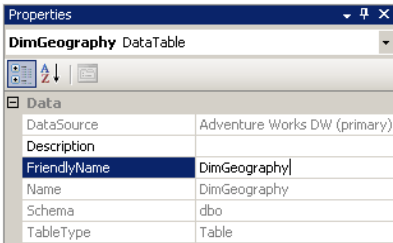
Rys. 6 Tabele wybrane do kostki danych



Po kliknięciu **Finish** został utworzony widok danych zawierające tabele potrzebne do budowy kostki. Panel edycyjny widoku został otwarty w głównym oknie Microsoft Visual Studio i dzieli się na trzy części:

- Diagram Organizer (po lewej na górze),
  - Tables (po lewej na dole),
  - Diagram (największa, centralna część).
- Przeanalizuj zawartość paneli **Diagram**, **Tables** oraz **Diagram Organizer**.
  - Kliknij prawym przyciskiem myszy puste pole **Diagram**, a następnie z kontekstowego menu wybierz **Add/Remove Tables**.

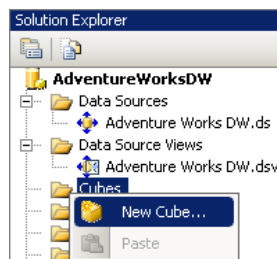


	 <p>Rys. 7 Dodawanie nowych tabel do źródła danych</p> <p>W otwartym oknie można dodawać lub usuwać tabele z widoku. Funkcja ta przydaje się, jeżeli zależy nam na dodatkowej tabeli, a zapomnieliśmy jej przy tworzeniu widoku.</p> <ul style="list-style-type: none"> <li>Kliknij <b>Anuluj</b>.</li> </ul>
<p>5. Zmiana domyślnych nazw tabel źródłowych</p>	<p>Każda z tabel oprócz nazwy, którą została jej nadana w bazie danych, posiada także nawę przyjazną dla użytkownika. Domyślnie nazwa ta jest taka, jak nazwa tabeli w bazie danych.</p> <ul style="list-style-type: none"> <li>Z menu głównego wybierz <b>View -&gt; Properties</b>, a następnie zaznacz tabelę <b>DimGeography</b>.</li> </ul>  <p>Rys. 8 Zmiana przyjaznej nazwy dla tabeli</p> <ul style="list-style-type: none"> <li>Za pomocą okna <b>Properties</b> zmień właściwość <b>FriendlyName</b> tabel zgodnie z poniższą listą: <ul style="list-style-type: none"> <li><b>FactInternetSales</b> -&gt; <b>InternetSales</b></li> <li><b>DimCustomer</b> -&gt; <b>Customer</b></li> <li><b>DimDate</b> -&gt; <b>Date</b></li> <li><b>DimGeography</b> -&gt; <b>Geography</b></li> <li><b>DimProduct</b> -&gt; <b>Product</b></li> </ul> </li> </ul>

### Definiowanie i publikowanie kostki danych (czas realizacji 20 min)

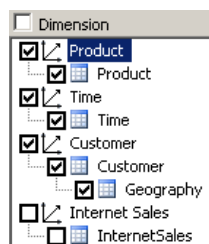
Kolejne ćwiczenie przeprowadzi Cię przez proces budowania kostki danych. Określisz hierarchie, atrybuty oraz fakty tworzące kostkę danych. Zweryfikujesz działanie kreatora pomagającego przy budowie kostki danych. Ostatnimi krokami ćwiczenia są: budowa i publikacja oraz późniejsze przeglądanie utworzonej kostki danych.

Zadanie	Tok postępowania
6. Definiowanie kostki danych	<ul style="list-style-type: none"> <li>W oknie <b>Solution Explorer</b> prawym przyciskiem myszy kliknij <b>Cubes</b> i wybierz <b>New Cube</b>.</li> </ul>



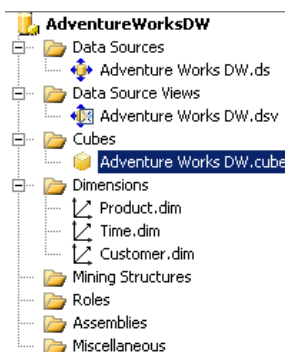
Rys. 9 Tworzenie nowej kostki danych

- Zaznacz **Use existing tables** i kliknij **Next**.
- Jako **Data source view** wybierz **Adventure Works DW** i zaznacz tabelę **InternetSales**. Kliknij **Next**.
- Odznacz następujące pola:
  - **Promotion Key**
  - **Currency Key**
  - **Sales Territory Key – Internet Sales**
  - **Revision Number**
- Kliknij **Next**.
- Wybierz wszystkie wymiary oprócz **Internet Sales**.



Rys. 10 Wybieranie wymiarów do utworzenia


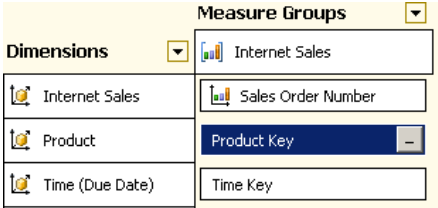



- Kliknij **Next**.
- Przejrzyj utworzone agregaty i wymiary. Kliknij **Finish**.



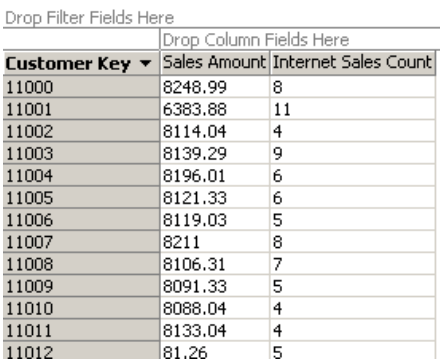



Rys. 11 Projekt ze źródłem danych, widokiem źródła, kostką danych i wymiarami




Została utworzona kostka **Adventure Works DW** i trzy wymiary: **Product**, **Date**, **Customer**. Domyślnie okno projektowania kostki otwiera się w panelu głównym Microsoft Visual Studio, nazwa zakładki to **Adventure Works DW.cube**. W widoku projektowania kostki mamy dostępnych kilka zakładek: **Cube Structure**, **Dimension Usage**, **Calculations**, **KPIs**, **Actions**, **Partitions**, **Aggregations**, **Perspectives**, **Translations**, **Browser**. Większość z tych zakładek omówimy w kolejnych punktach lub w następnych ćwiczeniach.

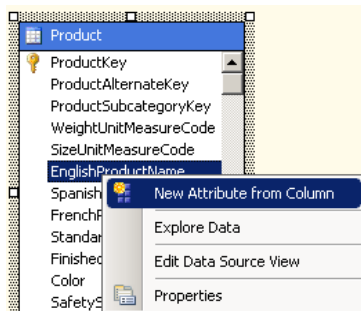
	<ul style="list-style-type: none"> <li>Przeanalizuj zawartość zakładki <b>Cube Structure</b>.</li> </ul>
7. Przeglądanie widoku edycji kostki	<ul style="list-style-type: none"> <li>Przełącz się na zakładkę <b>Dimension Usage</b>.</li> </ul> <p> Należy zwrócić szczególną uwagę na liczbę stworzonych wymiarów. Pomimo podania 3 tabel wymiarów i jednej tabeli wymiaru czasu, na panelu <b>Dimensions</b> jest widocznych 5 wymiarów. Jest to spowodowane zdefiniowaniem 3 różnych wymiarów czasu, a także złączeniem dwóch tablic wymiarów w jeden wymiar (tabele <b>Customer</b> i <b>Geography</b>).</p> <ul style="list-style-type: none"> <li>Kliknij przycisk w prawej części wymiaru <b>Product</b> w grupie agregatów <b>Internet Sales</b> (na przecięciu <b>Product</b> i <b>Internet Sales</b> znajduje się napis <b>Product Key</b>).</li> </ul>  <p style="text-align: center;"><i>Rys. 12 Wymiary dla grup pomiarów</i></p> <ul style="list-style-type: none"> <li>Sprawdź typ relacji (<b>relationship type</b>), granulację atrybutu (<b>Granularity attribute</b>), tabelę wymiaru (<b>Dimension Table</b>), tabelę grupy agregatów, czyli tabelę faktów (<b>Measure group table</b>), a także kolumnę klucza obcego do tabeli wymiaru (<b>Measure Group Columns</b>).</li> <li>Kliknij <b>OK</b>.</li> <li>Sprawdź pozostałe wymiary.</li> </ul>
8. Publikowanie kostki danych na AS2008	<p> W celu przeglądania danych zawartych w kostce danych konieczne jest opublikowanie stworzonej kostki danych. Idea procesu publikowania (deployowania) kostki danych jest w kilku aspektach zbliżona do procesu publikowania aplikacji webowych na serwerach aplikacji. Aby przygotować kostkę danych do opublikowania, należy odpowiednio ustawić jej właściwości.</p> <ul style="list-style-type: none"> <li>Kliknij prawym przyciskiem myszy nazwę projektu w oknie <b>Solution Explorer</b> i wybierz <b>Properties</b>.</li> <li>Z drzewka po lewej stronie okna dialogowego wybierz <b>Deployment</b>.</li> <li>Przeanalizuj zawartość pól po prawej stronie okna.</li> </ul> <p> Domyślną metodą publikacji jest publikowanie inkrementacyjne do domyślnej instancji <b>Analysis Services</b> na komputer lokalny.</p> <ul style="list-style-type: none"> <li>Kliknij <b>OK</b>, aby zamknąć okno właściwości.</li> <li>W celu opublikowania projektu w <b>Solution Explorer</b> kliknij prawym przyciskiem myszy jego nazwę i wybierz <b>Deploy</b>.</li> </ul> <p> Warto prześledzić zawartość okna <b>Output</b> oraz panelu <b>Deployment progress</b>. Oba okna zawierają szczegóły operacji budowania i publikowania projektu.</p> <ul style="list-style-type: none"> <li>Przeanalizuj zapytania używane do przetwarzania atrybutów wymiarów.</li> </ul>
9. Przeglądanie opublikowanej	<ul style="list-style-type: none"> <li>W widoku projektu kostki kliknij zakładkę <b>Browser</b>.</li> </ul> <p>Aby otworzyć widok projektu kostki w panelu <b>Solution Explorer</b>, kliknij</p>

kostki	<p>podwójnie kostkę <b>Adventure Works DW.cube</b>. Jeżeli wymiary i agregaty nie są widoczne, kliknij przycisk <b>Reconnect</b>. Niekiedy najlepiej jest zamknąć okno edycji kostki i otworzyć je ponownie, klikając kostkę w <b>Solution Explorer</b>.</p>  <p>Rys. 13 Odświeżenie danych poprzez ponowne połączenie</p> <p> Jakie są dostępne agregaty i wymiary?</p> <ul style="list-style-type: none"> <li>• Przenieś <b>Sales Amount</b> i <b>Internet Sales Count</b> na pole szczegółów (<b>Drop Total or Detail Fields Here</b>).</li> <li>• Przenieś <b>Customer Key</b> na pole wierszy (<b>Drop Row Fields Here</b>).</li> </ul>  <p>Rys. 14 Przykładowy raport zbudowany w oparciu o tabele przestawne</p> <p> Budowa raportu przestawnego została zakończona. Poziomy w wymiarach nie są jeszcze poprawnie zdefiniowane, dlatego nie ma sensu definiować innych raportów.</p>
--------	--

### Edycja wymiarów, poziomów i hierarchii (czas realizacji 15 min)

Po utworzeniu pierwszej wersji kostki danych należy przejrzeć dokładnie kostkę danych i dokonać niezbędne zmiany w odniesieniu do wymiarów, poziomów, hierarchii i atrybutów.

Zadanie	Tok postępowania
10. Edycja poziomu Product	<ul style="list-style-type: none"> <li>• W <b>Solution Explorer</b> rozwiń <b>Dimensions</b> i kliknij dwukrotnie <b>Product.dim</b>, aby otworzyć okno edycji wymiaru.</li> <li>•  W otworzonym oknie edycji wymiaru możemy zobaczyć trzy części: <b>Attributes</b>, <b>Hierarchies</b>, <b>Data Source View</b>. Aktualnie jest tylko jeden poziom nazwany <b>Product Key</b>.</li> <li>• Aby dodać kolejny poziom powiązany z konkretną kolumną w tabeli wymiaru kliknij prawym przyciskiem myszy kolumnę <b>EnglishProductName</b> i z kontekstowego menu wybierz <b>Add New Attribute from Column</b>.</li> </ul>



Rys. 15 Dodanie poziomu na podstawie kolumny



Został dodany poziom o nazwie **English Product Name**. Aplikacja sama utworzyła nazwę poziomu z nazwy kolumny poprzez wstawienie spacji przed duże litery (**EnglishProductName** -> **English Product Name**).

- Aby zmienić nazwę poziomu, kliknij prawym przyciskiem myszy poziom i z kontekstowego menu wybierz **Rename**. Zmień nazwę na **Product Name**.
- Dodaj pozostałe poziomy dla wymiaru **Product**:
  - **WeightUnitMeasureCode** -> **Weight Unit Measure Code**
  - **SizeUnitMeasureCode** -> **Size Unit Measure Code**
  - **Color** -> **Color**
  - **Size** -> **Size**
  - **Weight** -> **Weight**
  - **ProductLine** -> **Product Line**
  - **DealerPrice** -> **Dealer Price**
  - **Class** -> **Class**
  - **Style** -> **Style**
  - **ModelName** -> **Model Name**
  - **EnglishDescription** -> **Description**
  - **Status** -> **Status**
  - **DealerPrice** -> **Dealer Price**
  - **ListPrice** -> **List Price**
  - **DaysToManufacture** -> **Days To Manufacture**








Kolejnym krokiem jest modyfikacja poziomów dla wymiaru **Product**. Poziom **Product Key** jest kluczem i nie powinien być wyświetlany, z kolei inne poziomy posiadają pewien typ i powinny być wyświetlane w określony sposób.


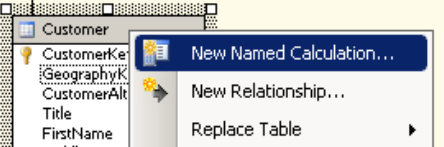

- Kliknij prawym przyciskiem myszy **Product Key** i z kontekstowego menu wybierz **Properties**.
- W oknie właściwości dla poziomu zmień **AttributeHierarchyVisible** z wartości **True** na wartość **False**.

11. Edycja poziomu Time

- Dla wymiaru **Date** dodaj następujące poziomy:
  - **CalendarYear** -> **Year**
  - **CalendarQuarter** -> **Quarter**
  - **MonthNumberOfYear** -> **Month Number**
  - **EnglishMonthName** -> **Month Name**
  - **WeekNumberOfYear** -> **Week Of Year**
  - **DayNumberOfYear** -> **Day Of Year**
  - **DayNumberOfMonth** -> **Day Of Month**

	<ul style="list-style-type: none"> <li>– <b>FullDateAlternateKey -&gt; Full Date</b></li> <li>• Kliknij prawym przyciskiem myszy poziom <b>Year</b> i z kontekstowego menu wybierz <b>Properties</b>. W oknie właściwości dla poziomu <b>Year</b> zmień <b>OrderBy</b> z <b>Name</b> na <b>Key</b>, a także <b>Type</b> z <b>Regular</b> na <b>Years</b> (z rozwijanego drzewa <b>Date \ Calendar \ Years</b>).</li> <li>• Dla pozostałych wymiarów zmień wartości <b>Type</b> na następujące: <ul style="list-style-type: none"> <li>– <b>Year -&gt; Date \ Calendar \ Years</b></li> <li>– <b>Quarter -&gt; Date \ Calendar \ Quarters</b></li> <li>– <b>Month Number -&gt; Date \ Calendar \ MonthOfYear</b></li> <li>– <b>Month Name -&gt; Date \ Calendar \ Months</b></li> <li>– <b>Week Of Year -&gt; Date \ Calendar \ WeekOfYear</b></li> <li>– <b>Day Of Year -&gt; Date \ Calendar \ DayOfYear</b></li> <li>– <b>Day Of Month -&gt; Date \ Calendar \ DayOfMonth</b></li> <li>– <b>Full Date -&gt; Date \ Calendar \ Date</b></li> </ul> </li> <li> Zmiana typów dla poziomów nie jest istotna z punktu widzenia użytkownika, tylko utworzonych później agregacji. Na podstawie typów można określić, jakie relacje są pomiędzy poszczególnymi poziomami. Poziomy zostały podświetlone na niebiesko, co oznacza pewne błędy w wymiarze. Powodem tego jest typ wymiaru <b>Time</b>, który jest typem <b>Regular</b>. Należy zmienić typ wymiaru na <b>Time</b>.</li> <li>• Kliknij prawym przyciskiem myszy wymiar <b>Time</b> i wybierz <b>Properties</b>.</li> </ul> <div data-bbox="858 1041 1098 1339" data-label="Image"> </div> <p style="text-align: center;"><i>Rys. 16 Wyświetlenie właściwości dla wymiaru</i></p> <ul style="list-style-type: none"> <li>• W okienku właściwości dla wymiaru <b>Time</b> zmień <b>Type</b> z <b>Regular</b> na <b>Time</b>.</li> <li> Kolejno chcemy ukryć ID dla tabeli daty i czasu. Poziom ten nie jest potrzebny w naszym wymiarze do widoku użytkownika. Nie możemy usunąć tego poziomu, ponieważ jest to klucz główny w tabeli wymiaru. Usunięcie <b>Time Key</b> spowoduje błąd w wymiarze uniemożliwiający publikację kostki.</li> <li>• Dla poziomu <b>Time Key</b> zmień właściwość <b>AttributeHierarchyVisible</b> na <b>False</b>.</li> </ul>
<p>12. Dodawanie hierarchii dla wymiaru Time</p>	<ul style="list-style-type: none"> <li>• W oknie projektowania wymiaru <b>Date</b> wybierz zakładkę <b>dimension Structure</b>.</li> <li>• Przenieś poziom <b>Year</b> z panelu <b>Attributes</b> do panelu <b>Hierarchies</b>.</li> <li> Została utworzona hierarchia o nazwie <b>Hierarchy</b>, która posiada tylko jeden poziom: <b>Year</b>.</li> <li>• Przenieś kolumnę <b>Quarter</b> z panelu <b>Attributes</b> do panelu <b>Hierarchies</b>,</li> </ul>


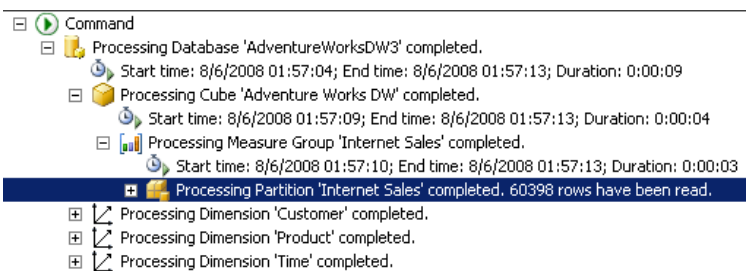

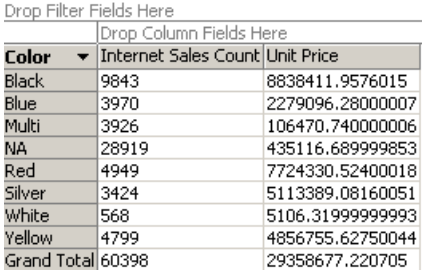
	<p>hierarchii <b>Hierarchy</b> poniżej poziomu <b>Year</b>.</p> <ul style="list-style-type: none"> <li>• Dodaj do hierarchii następujące kolumny: <b>Month Name, Day of Month, Full Date</b>.</li> <li>• We właściwościach hierarchii zmień nazwę na <b>Date by Year Quarter Month</b>.</li> </ul>
13. Edycja poziomu Customer	<ul style="list-style-type: none"> <li>• Dla wymiaru <b>Customer</b> dodaj poziomy z następujących kolumn: <ul style="list-style-type: none"> <li>– <b>Title</b> -&gt; <b>Title</b></li> <li>– <b>FirstName</b> -&gt; <b>First Name</b></li> <li>– <b>LastName</b> -&gt; <b>Last Name</b></li> <li>– <b>BirthDate</b> -&gt; <b>Birth Date</b></li> <li>– <b>MaritalStatus</b> -&gt; <b>Marital Status</b></li> <li>– <b>Gender</b> -&gt; <b>Gender</b></li> <li>– <b>Email Address</b> -&gt; <b>Email</b></li> <li>– <b>YearlyIncome</b> -&gt; <b>Yearly Income</b></li> <li>– <b>TotalChildren</b> -&gt; <b>Total Children</b></li> <li>– <b>NumberChildrenAtHome</b> -&gt; <b>Number of Children At Home</b></li> <li>– <b>EnglishEducation</b> -&gt; <b>Education</b></li> <li>– <b>EnglishOccupation</b> -&gt; <b>Occupation</b></li> <li>– <b>HouseOwnerFlag</b> -&gt; <b>House Owner Flag</b></li> <li>– <b>NumberCarsOwned</b> -&gt; <b>Number of Cars Owned</b></li> <li>– <b>AddressLine1</b> -&gt; <b>Address</b></li> <li>– <b>Phone</b> -&gt; <b>Phone</b></li> <li>– <b>ComuteDistance</b> -&gt; <b>Comute Distance</b></li> <li>– <b>EnglishCountryRegionName</b> -&gt; <b>Country Region Name</b></li> </ul> </li> <li>• Dla odpowiednich poziomów zmień właściwość <b>Type</b>: <ul style="list-style-type: none"> <li>– <b>Country Region Name</b> -&gt; <b>Geography \ Region</b></li> <li>– <b>Country Region Code</b> -&gt; <b>Geography \ Region</b></li> <li>– <b>State Province Code</b> -&gt; <b>Geography \ StateOrProvince</b></li> <li>– <b>State Province Name</b> -&gt; <b>Geography \ StateOrProvince</b></li> <li>– <b>City</b> -&gt; <b>Geography \ City</b></li> <li>– <b>Address</b> -&gt; <b>Other \ Location \ Address</b></li> </ul> </li> </ul> <p> Podobnie jak zmiany dla wymiaru <b>Date</b>, typy dla poziomów wymiaru <b>Customer</b> pozwolą na lepsze zrozumienie zależności pomiędzy poszczególnymi poziomami i bardziej skuteczne projektowanie agregacji.</p> <ul style="list-style-type: none"> <li>• Ustaw właściwość <b>AttributeHierarchyDisplayFolder</b> na wartość <b>Location</b> dla poziomów: <ul style="list-style-type: none"> <li>– <b>City</b></li> <li>– <b>Country Region Name</b></li> <li>– <b>Postal Code</b></li> <li>– <b>State Province Name</b></li> </ul> </li> </ul> <p> Niektóre wymiary posiadają wiele poziomów. Ustawienie folderu dla każdego z nich pozwala uporządkować wymiar, aby łatwiej było odnaleźć potrzebne dane.</p> <ul style="list-style-type: none"> <li>• Ustaw właściwość <b>AttributeHierarchyDisplayFolder</b> na wartość <b>Demographic</b> dla poziomów: <ul style="list-style-type: none"> <li>– <b>Commute Distance</b></li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>– <b>Education</b></li> <li>– <b>Gender</b></li> <li>– <b>House Owner Flag</b></li> <li>– <b>Marital Status</b></li> <li>– <b>Number Cars Owned</b></li> <li>– <b>Number Children At Home</b></li> <li>– <b>Occupation</b></li> <li>– <b>Total Children</b></li> <li>– <b>Yearly Income</b></li> </ul>
14. Dodawanie kolumn kalkulowanych	<p> Jak można zauważyć, wymiar <b>Customer</b> posiada osobno imię i nazwisko klienta. Zależy nam jednak na informacji pełnej, która zawiera imię i nazwisko w jednej kolumnie. Można to uzyskać dodając kolumnę kalkulowaną.</p> <ul style="list-style-type: none"> <li>• Otwórz widok edycji widoku źródła danych klikając dwukrotnie <b>Adventure Works DW.dsv</b>.</li> <li>• Kliknij prawym przyciskiem myszy tabelę <b>Customer</b> (jej nazwę) i wybierz <b>New Named Calculation</b>.</li> </ul>  <p>Rys. 17 Dodanie nowej kolumny kalkulowanej</p> <ul style="list-style-type: none"> <li>• Wpisz następujące wartości: <ul style="list-style-type: none"> <li>– <b>Column name:</b> <b>FullName</b></li> <li>– <b>Description:</b> <b>Full name of customer</b></li> <li>– <b>Expression:</b> <pre>case when MiddleName is null then     FirstName + ' ' + LastName else     FirstName + ' ' + MiddleName + ' ' + LastName End</pre> </li> </ul> </li> <li>• Kliknij <b>OK</b>.</li> </ul> <p> Wpisane wyrażenie jest to kod SQL który zostanie dopisany do zapytania SQL do frazy <b>select</b>, co pozwoli na obliczenie pełnej nazwy klienta dla każdego wiersza.</p> <ul style="list-style-type: none"> <li>• Kliknij prawym przyciskiem myszy tabelę <b>Customer</b> i wybierz <b>Explore Data</b>. Sprawdź, czy w dodanej na końcu w kolumnie <b>FullName</b> znajdują się poprawne wartości.</li> <li>• Dodaj poziom <b>Full Name</b> w wymiarze <b>Customer</b>.</li> </ul> <p>Po dodaniu kolumny kalkulowanej lub w przypadku zmiany schematu źródłowego zmiany te są od razu widoczne w widoku źródła danych, dlatego jesteśmy w stanie bardzo szybko zmodyfikować kostkę.</p>



## Budowanie raportów za pomocą tabel przestawnych (czas realizacji 10 min)

Po zmianach w wymiarach, hierarchiach i poziomach można przejść do budowania prostych raportów za pomocą wbudowanych tabel przestawnych. Jest to bardzo prosty sposób na eksplorację danych i sprawdzenie, jak prezentuje się zbudowana kostka danych.

Zadanie	Tok postępowania
15. Deploy kostki danych	<ul style="list-style-type: none"> <li>Kliknij prawym przyciskiem myszy projekt <b>Adventure Works DW</b> i wybierz <b>Process</b>.</li> <li>Jeżeli pokaże się okienko o ponownym budowaniu projektu, kliknij <b>Yes</b>.</li> <li>Kliknij <b>Run</b>, aby rozpocząć przetwarzanie kostki danych.</li> </ul> <p> W razie ewentualnych błędów należy dobrze sprawdzić wymiary, poziomy i hierarchie. Po przetworzeniu procesu można sprawdzić, jaka liczba wierszy została przetworzona i ile czasu trwało samo przetwarzanie.</p>  <p style="text-align: center;"><i>Rys. 18 Log z publikowania kostki danych</i></p> <ul style="list-style-type: none"> <li>Kliknij <b>Close</b> i jeszcze raz <b>Close</b>.</li> <li>W oknie <b>Adventure Works DW.cube</b> przejdź na zakładkę <b>Browser</b>.</li> </ul> <p> Jakie wymiary i agregaty zawiera kostka <b>Adventure Works DW</b>?</p>
16. Definiowanie prostego raportu	<ul style="list-style-type: none"> <li>Przeciągnij agregaty: <b>Internet Sales Count</b> i <b>Unit Price</b> do pola szczegółów (<b>Drop Totals Or Detail Fields Here</b>).</li> <li>Przeciągnij poziom <b>Color</b> z wymiaru <b>Product</b> do pola wierszy (<b>Drop Row Fields Here</b>).</li> </ul>  <p style="text-align: center;"><i>Rys. 19 Raport przedstawiający cenę jednostek i liczbę sprzedaży w zależności od koloru produktu</i></p> <ul style="list-style-type: none"> <li>Na raporcie kliknij <b>Internet Sales Count</b> i naciśnij <b>Delete</b>, kliknij także <b>Color</b> i naciśnij <b>Delete</b>.</li> <li>Przeciągnij do raportu agregat <b>Tax Amt</b> i poziom <b>Year</b> z wymiaru <b>Ship Date</b>.</li> </ul>

- Ship Date
  - Ship Date.Day Of Month
  - Ship Date.Day Of Year
  - Ship Date.Full Date
  - Ship Date.Month Name
  - Ship Date.Month Number
  - Ship Date.Quarter
  - Ship Date.Week Of Year
  - Ship Date.Year
  - Ship Date.Date by Year Quarter Month Day

Year	Unit Price	Tax Amt
2001	3105587.32920001	248446.990999999
2002	6576978.98109993	526158.339500008
2003	9517548.5304037	761403.909200116
2004	10158562.3800044	812684.990400158
Grand Total	29358677.2207081	2348694.23010028

Rys. 20 Poziomy i hierarchia dla wymiaru Ship Date, raport przedstawiający cenę jednostek i podatek w zależności od roku

Można zauważyć, że nie da się rozwinąć lat, ponieważ poziom Year został wybrany bezpośrednio z wymiaru, a nie z hierarchii.

- Usuń z raportu Year, a następnie przeciągnij Year z wymiaru Ship Date z hierarchii Date by Year Quarter Month Day.

- Ship Date
  - Ship Date.Day Of Month
  - Ship Date.Day Of Year
  - Ship Date.Full Date
  - Ship Date.Month Name
  - Ship Date.Month Number
  - Ship Date.Quarter
  - Ship Date.Week Of Year
  - Ship Date.Year
  - Ship Date.Date by Year Quarter Month Day
- Members
  - Year
  - Quarter
  - Month Name
  - Day Of Month

Year	Quarter	Month Name	Unit Price	Tax Amt
2001	3	July	349490.1938	27959.2159
		August	518096.417600001	41447.7142
		September	481550.124800001	38524.0106
	4	Total	1349136.7362	107930.9407
Total		1756450.593	140516.0503	
Total		3105587.32920001	248446.990999999	
2002			6576978.98109993	526158.339500008
2003			9517548.5304037	761403.909200116
2004			10158562.3800044	812684.990400158
Grand Total			29358677.2207081	2348694.23010028

Rys. 21 Poziomy i hierarchia dla wymiaru Ship Date, raport przedstawiający cenę jednostek i podatek w zależności od roku rozwijanego do kwartału i miesiąca

17. Tworzenie zaawansowanego raportu

- Do pól szczegółów przenieś agregaty Order Quantity i Sales Amount.
- Do pól kolumn przenieś poziom Education z wymiaru Customer.
- Do pól wierszy przenieś poziom Product Line z wymiaru Product.
- Do pól filtrów przenieś poziom Year z wymiaru Ship Date. Rozwiń lata i zaznacz tylko wartości 2002 i 2003.

Ship Date.Date by Year Quarter Month Day (Multiple Items)					
		Gender		Grand Total	
		F		M	
Product Line	Order Quantity	Sales Amount	Order Quantity	Sales Amount	Order Quantity
M	3694	2888929.463	3708	2667010.4214	7402
R	3894	4505304.64120002	3935	4478758.32590002	7829
S	4474	115303.34	4616	118016.73	9090
T	838	677058.920000001	819	644145.67	1657
Grand Total	12900	8186596.36420001	13078	7907931.14730001	25978

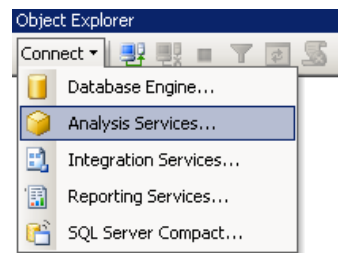
Rys. 22 Raport przedstawiający liczbę jednostek zamówionych i sumę sprzedaży w zależności od linii produktu i płci

## Przeglądanie opublikowanej kostki za pomocą Microsoft SQL Server Management Studio (czas realizacji 10 min)

Opublikowana kostka znajduje się w Analysis Services. Można ją przeglądać nie tylko za pomocą projektu, ale także za pomocą konsoli zarządzającej bazami SQL Server. Przeglądanie kostki i tworzenie raportów wielowymiarowych możliwe jest także za pomocą Reporting Services i dowolnego narzędzia, które jest klientem zgodnym ze specyfikacją XML/A.

Zadanie	Tok postępowania
18. Przeglądanie kostki danych za	<ul style="list-style-type: none"> <li>Otwórz SQL Server Management Studio, a następnie w oknie <b>Object Explorer</b> wybierz <b>Analysis Services</b>.</li> </ul>

pomocą konsoli  
zarządzającej



*Rys. 23 Połączenie do serwera Analysis Services*

- Wpisz nazwę serwera (**localhost**), a następnie kliknij **Connect**.
- Rozwiń **Databases** -> **Cubes**, a następnie kliknij prawym przyciskiem myszy kostkę i z kontekstowego menu wybierz **Browse**.
- Utwórz kilka przykładowych raportów.

## Laboratorium rozszerzone

### Zadanie 1 (czas realizacji 10 min)

W wymiarze **Time** utwórz hierarchię **Year – Month Number – Day of Month**.

### Zadanie 2 (czas realizacji 10 min)

Kwartał jest wyświetlany jako numer kwartału: 1, 2, 3, 4. W raportach powinien być wyświetlany jako **Q1, Q2, Q3 i Q4**.

Używając nowej kolumny, dodaj poziom **Quarter Name** i utwórz hierarchię: **Year – Quarter Name – Month Name**.

### Zadanie 3 (czas realizacji 15 min)

**Product Line** zawiera oznaczenia: **M, R, S i T**. Nie jest to przejrzyste z punktu widzenia użytkownika biznesowego. Należy zmienić wyświetlanie tych wartości na raportach na nazwy znaczące.

Używając kolumny kalkulowanej i polecenia

```
case
  when <warunek1> then <rezultat1>
  when <warunek2> then <rezultat2>
  when <warunek3> then <rezultat3>
  else <rezultat4>
end
```

zmień poziom **Product Line** na następujące wartości:

- **M – Mountain**
- **R – Road**
- **S – Accessory**
- **T – Touring**
- w pozostałych przypadkach – **Components**

utwórz nową kolumnę kalkulowaną **ProductFullLine**, a następnie dodaj poziom **Product Full Line** oparty na nowo utworzonej kolumnie.

### Zadanie 4 (czas realizacji 15 min)

Wymiar klienta (**Customer**) nie posiada żadnych hierarchii. W celu sprawnego projektowania raportów konieczne jest zdefiniowanie następujących hierarchii:

- **Customer by Location: Country Region Name -> State Province Name -> City**
- **Customer By Education and Occupation: Education -> Occupation**

Zdefiniuj wspomniane hierarchie i utwórz dwa raporty: wszystkie agregaty w zależności od danej hierarchii.

### Zadanie 5 (czas realizacji 40 min)

Firma **Czysta Energia** jest jednym z producentów i dostawców mediów do mieszkań. Właśnie wdrożone zostały odczyty elektroniczne z liczników gazu, wody i energii elektrycznej. Odczyty, w odróżnieniu od tradycyjnych przechowują stan różnicowy zużycia medium (oznacza to, że wartości wyznaczają zużycie od ostatnio przesłanego pomiaru, nie od początku założenia licznika). Dane są przechowywane w bazie danych, co umożliwia wystawianie rachunków dla odbiorców indywidualnych.

Problemem są raporty, które należy generować co pewien czas – zajmuje to znaczące zasoby sprzętowe. Pojedynczy pomiar odczytywany jest co pół godziny, dlatego liczba pomiarów

przekracza kilkaset milionów miesięcznie. Powoduje to znaczne wykorzystanie bazy przy przekrojowych raportach analitycznych.

Rozwiązaniem jest zbudowanie kostki danych z kilkoma wymiarami, dzięki czemu możliwe będzie bardzo szybkie tworzenie raportów za pomocą tabel przestawnych, a także późniejsza ich prezentacja na stronie WWW (za pomocą **Reporting Services**).

Skrypt tworzący bazę danych wraz z przykładowymi danymi znajduje się w pliku **DigitalMeters.sql**.

Zapoznaj się ze schematem relacyjnym, a następnie utwórz kostkę danych **Measures** z następującymi wymiarami: **Meter, Date, Vendor, Temperature**.

Za pomocą tabel przestawnych zaprojektuj następujące raporty:

- Suma zużycia mediów w zależności od roku i typu licznika.
- Suma zużycia mediów w zależności od dostawcy, tylko dla energii.
- Suma zużycia mediów w zależności od siły wiatru, tylko dla gazu.