

ITA-102 Hurtownie danych

Marcin Gorawski, Sławomir Bańkowski

Moduł 5

Wersja 1.0

Usługi analiz danych: funkcje zaawansowane, skalowalność i optymalizacja

Spis treści

Usługi analiz danych: funkcje zaawansowane, skalowalność i optymalizacja	1
Informacje o module	2
Przygotowanie teoretyczne.....	3
Przykładowy problem.....	3
Podstawy teoretyczne.....	3
Przykładowe rozwiązanie	4
Porady praktyczne	4
Uwagi dla studenta	4
Dodatkowe źródła informacji.....	4
Laboratorium podstawowe.....	5
Różne rodzaje wymiarów (czas realizacji 15min)	5
Definiowanie translacji (czas realizacji 15min)	7
Partycje w kostkach danych (czas realizacji 30min)	9
Laboratorium rozszerzone	17
Zadanie 1 (czas realizacji 45min)	17
Zadanie 5 (czas realizacji 45 min)	17

Informacje o module

Opis modułu

W module znajdziesz opis funkcji zaawansowanych budowania kostek danych: definiowania translacji, dodawania wymiarów o innym typie niż **Regular**, dodawania **KPI**, a także problemu skalowalności kostek danych i stosowania partycji danych.

Cel modułu

Celem modułu jest dokładne poznanie zaawansowanych funkcji budowania kostek danych, co pozwoli na budowanie profesjonalnych struktur pełni wykorzystujących możliwości **Analysis Services**.

Uzyskane kompetencje

Po zrealizowaniu modułu będziesz:

- znał różne rodzaje wymiarów dla kostki danych
- potrafił dodać translację dla dowolnego języka i skorzystać z tej translacji
- wiedział jakie są możliwe typy wymiarów powiązanych z kostkami danych
- rozumiał pojęcie skalowalności kostki danych
- umiał utworzyć nową partycję wraz z agregatami zgodnie z hierarchiami

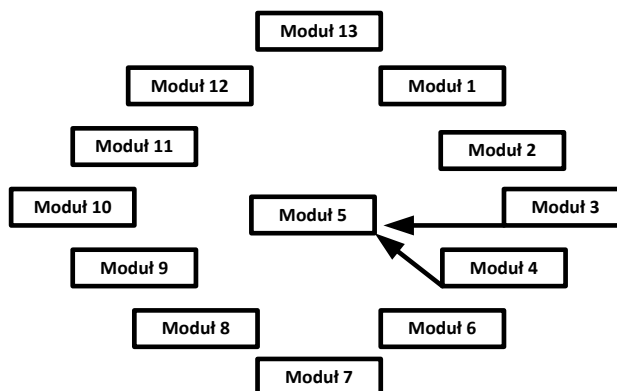
Wymagania wstępne

Przed przystąpieniem do pracy z tym modulem powinienś:

- rozumieć schemat relacyjny i budowę modelu wielowymiarowego
- potrafić utworzyć kostkę danych, a także poziomy, wymiary i hierarchie dla kostki
- potrafić dodać tabele do widoku źródła danych, dodać nowe wymiary do kostki i zdefiniować perspektywy

Mapa zależności modułu

Zgodnie z mapą zależności przedstawioną na Rys. 1, przed przystąpieniem do realizacji tego modułu należy zapoznać się z materiałem zawartym w module 3, 4.



Rys. 1 Mapa zależności modułu

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 skierowanie 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

Działająca poprawnie hurtownia danych, która pozwala ze zgromadzonych w niej danych wydobywać wiedzę, stanowi tylko część sukcesu jej administratora. Oprócz poprawnego działania, hurtownia danych powinna działać wydajnie, niezależnie od faktu jak duża ilość danych (agregatów) jest w niej zgromadzona. W tym miejscu pojawiają się dwa zagadnienia, które zostaną poruszone podczas zajęć: skalowalność oraz optymalizacja.

Pod pojęciem skalowalności rozumiemy tutaj „odporność” na ciągle rosnącą ilość danych. Podstawowym środkiem zapewniającym skalowalność jest partycjonowanie. Dzięki partycjonowaniu możemy uzyskać np.:

- zwiększenie wydajności przetwarzania zapytań MDX,
- połączenie w jedną logiczną kostkę danych rozmieszczonych w kilku niezależnych źródłach danych,
- zmniejszenie skutków ewentualnej awarii nośnika danych (utrata tylko części danych).

Minusem partycjonowania jest większa podatność systemu na awarie, gdyż im więcej jest elementów składowych systemu, tym większe jest prawdopodobieństwo, że w danej chwili przynajmniej jeden z nich ulegnie uszkodzeniu. Jednakże tego typu rozważania nie są przedmiotem zajęć.

Optymalizacja to proces takiego konfigurowania serwera, który pozwala uzyskać maksymalną wydajność, co w przypadku **Analysis Services** przekłada się po prostu na czas realizacji zapytań MDX. Aby osiągnąć cel możemy odpowiednio zaplanować proces budowania (odświeżania) kostki danych korzystając z tzw. buforowania proaktywnego. Oprócz tego na wydajność wpływa liczba zaplanowanych agregacji oraz utworzone hierarchie agregatów.

Proces strojenia może być dokonywany przez administratora, może być również zautomatyzowany i wykonywany z poziomu aplikacji odwołującej się do AS za pośrednictwem interfejsu AMO (**Analysis Management Objects**). Dostępny zestaw obiektów pozwala tworzyć partycje, planować agregacje, itd. z poziomu programowego.

Niniejsze opracowanie przedstawia zagadnienia związane ze skalowalnością **SQL Server 2008 Analysis Services**. W dalszej części omówione zostaną następujące cechy:

- partycje,
- przetwarzanie i strojenie nośnika danych,
- agregacje,
- właściwości składowe,
- agregacje w czasie projektowania AMO.

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

- Dla drobnych zmian w schemacie wymiarów nie trzeba przeprowadzać przetwarzania dla całego projektu lub kostki (**Process**), wystarczy tylko opublikować kostkę (**Deploy**).
- Przy translacji zastanów się, jakie znaczenie ma definiowanie każdego elementu osobno: wymiarów, agregatów, poziomów i hierarchii.

Uwagi dla studenta

Jesteś przygotowany do realizacji laboratorium jeśli:

- wiesz co to są translacje
- rozumiesz znaczenie partycji danych
- potrafisz wymienić zalety i wady agregacji

Przed wykonywaniem każdego punktu dokładnie przeczytaj cały punkt, to spowoduje znaczne zmniejszenie liczby pomyłek podczas laboratorium i pozwoli na dokładniejsze poznanie materiału. Jeżeli są błędy niewiadomego pochodzenia, spróbuj cofnąć ostatnie akcje i powtórzyć jeszcze raz. Definiowanie partycji i agregatów jest bardzo czułe na błędy, czasem najmniejsze pomyłki powodują bardzo duże błędy w całym projekcie.

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
2. Erik Thomsen: OLAP Solutions: Building Multidimensional Information Systems
3. Ralph Kimball, Margy Ross: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling


Laboratorium podstawowe




W ramach laboratorium podstawowego przedstawiamy definiowanie translacji na projekcie **AdventureWorksDW**. Na tym samym projekcie prezentujemy utworzenie partycji, a później agregacji. Kolejnym krokiem są agregacje w hierarchiach naturalnych.

Różne rodzaje wymiarów (czas realizacji 15min)

Do tej pory rozpatrywaliśmy kostkę, w której agregaty były w centralnej tabeli faktów, natomiast każdy wymiar był połączony z tabelą faktów za pomocą klucza obcego. Wymiar składał się z kilku tabel, każda kolejna tabela była dołączona za pomocą klucza obcego do jednej z tabeli wymiaru.

Czasami zależy nam na wymiarze, który nie jest bezpośrednio połączony z tabelą faktów, poziomy mogą znajdować się także w samej tabeli faktów. Nie należy mylić jednak poziomu z agregatem – poziom posiadający tyle samo elementów co liczba wierszy w tabeli faktów rzadko ma sens logiczny. Możemy sobie wyobrazić tabele z pomiarami pogody dla pewnego regionu. W tabeli faktów posiadamy siłę wiatru, temperaturę i opady deszczu lub śniegu. Tabelami wymiarów mogą być lokalizacja zmierzonych warunków atmosferycznych, data pomiaru, typ urządzenia mierzącego. Jeżeli chcielibyśmy uzyskać średnią temperaturę dla określonych przedziałów opadów – kostka zbudowana w pierwotny wzór nie umożliwi nam tego. Z całą pewnością możemy dodać kolumnę kalkulowaną, która będzie nam dyskretyzowała wartości opadów na kilka przedziałów: bardzo niskie, niskie, średnie, wysokie, bardzo wysokie. Z tej kolumny znajdującej się w tabeli faktów powinniśmy utworzyć osobny wymiar. Cały czas wiemy, że wymiar ten nie będzie miał zbyt dużej liczby różnych elementów.

Zadanie	Tok postępowania
1. Otwarcie projektu	<ul style="list-style-type: none"> Otwórz Business Intelligence Development Studio (BIDS), następnie z menu wybierz File -> Open -> Project/Solution. Załaduj z dysku projekt Adventure Works DW z poprzedniego modułu. <p> Jeżeli nie masz projektu z poprzedniego modułu, to możesz wykorzystać gotowy projekt dołączony razem z dokumentacją do aktualnego modułu.</p>
2. Definiowanie wymiaru połączanego przez tabelę pośrednią	<ul style="list-style-type: none"> Zakładamy, że zależy nam na wymiarze kategorii i podkategorii produktu nie wchodząc w szczegóły samego produktu. W Solution Explorer kliknij dwukrotnie na kostce Adventure Works DW.dsv aby otworzyć widok źródła danych. Sprawdź, czy na diagramie znajdują się tabele ProductCategory i ProductSubcategory. Jeżeli tych tabel nie ma, dodaj je i zmień nazwy odpowiednio na: Product Category i Product Subcategory. W Solution Explorer kliknij prawym przyciskiem myszy na Dimensions i wybierz New Dimension. Zaznacz Use an existing table i kliknij Next. Jako Data source view wybierz Adventure Works DW i kliknij Next. Jako Main table wybierz Product Subcategory i kliknij Next. W Related tables zaznacz Product Category (powinno być już zaznaczone) i kliknij Next. W available attributes zaznacz English Product Category Name i English Product Subcategory Name i zmień nazwy tych atrybutów odpowiednio na Product Category Name i Product Subcategory Name. Kliknij Next. Przejrzyj utworzony wymiar i kliknij Finish. W widoku edycji kostki wybierz zakładkę Dimension Usage.

	<ul style="list-style-type: none"> • Kliknij na puste pole prawym przyciskiem myszy i wybierz Add Cube Dimension. • Z listy dostępnych wymiarów w oknie Select Dimension wybierz Product Subcategory i kliknij OK. • Kliknij na pole dla wymiaru Product Subcategory i grupy agregatów Internet Sales.  <p><i>Rys. 2 Łączenie wymiaru z grupa agregatów</i></p> <ul style="list-style-type: none"> • W oknie definicji relacji zmień Select relationship type z No relationship na Referenced. • W Intermediate dimension wybierz Product. • W Reference dimension attribute wybierz Product Subcategory Key. • W Intermediate dimension attribute wybierz Product Subcategory Key. • Kliknij OK.  <p><i>Rys. 3 Zadeklarowany wymiar z połączeniem poprzez tabelę pośrednią</i></p>
<p>3. Tworzenie wymiaru z tabeli faktów</p>	<p> W poniższym punkcie zależy nam na wymiarze określającym cenę jednostki. Taki wymiar składałby się z kilku progów przykładowo: 'do 1000', 'od 1000 do 2000', 'od 2000 do 3000', 'powyżej 3000'. W ramach takiego wymiaru można zaprojektować raport: suma zarobionych pieniędzy w zależności od ceny jednostkowej produktu.</p> <ul style="list-style-type: none"> • W widoku edycji źródła danych kliknij prawym przyciskiem myszy na tabeli InternetSales i wybierz New Named Calculation. • Jako Calculation name wpisz Unit Price Discretized. • Jako Description wpisz: Cena jednostki. • Jako Expression wpisz: <pre> case when UnitPrice<1000 then 'do 1000' when UnitPrice<2000 then 'od 1000 do 2000' when UnitPrice<3000 then 'od 2000 do 3000' else 'powyżej 3000' end </pre> • W Solution Explorer kliknij prawym przyciskiem myszy na Dimensions i wybierz New Dimension. • Zaznacz Use as existing table i kliknij Next. • Jako Data source view wybierz Adventure Works DW. • Jako Main table wybierz InternetSales. • Jako Key columns wybierz: SalesOrderNumber i SalesOrderLineNumber. • Jako Name column wybierz Unit Price Discretized. • W Select related tables odznacz wszystkie tabele i kliknij Next. • Odznacz wszystkie kolumny oprócz SalesNumber i kliknij Next. • Wpisz nazwę wymiaru: Unit Price Level i kliknij Finish.

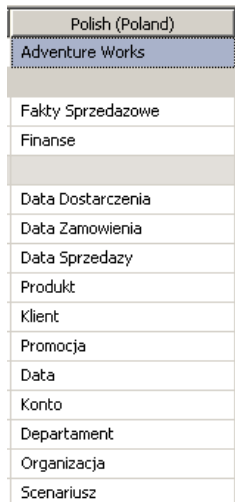


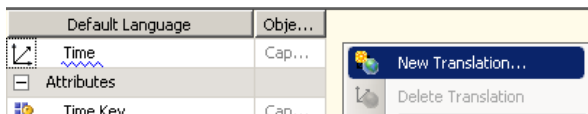
	<ul style="list-style-type: none"> W oknie edycji dla wymiaru wybierz zakładkę Dimension structure. W panelu Data Source View dodaj kliknij prawym przyciskiem myszy na kolumnie Unit Price Discretized i wybierz New Attribute from Column. <div data-bbox="783 331 1177 421" data-label="Image"> </div> <p style="text-align: center;">Rys. 4 Tworzenie nowej kolumny dla wymiaru</p> <ul style="list-style-type: none"> Dla poziomu Sales Order Number ustaw właściwość AttributeHierarchyVisible na False. W oknie projektowania kostki wybierz zakładkę Dimension Usage. Kliknij w puste pole i wybierz Add Cube Dimension. Z listy wymiarów wybierz Unit Price Level i kliknij OK. <div data-bbox="517 689 564 741" data-label="Image"> </div> <p>Można zauważyć, że został dodany nowy wymiar z charakterystyczną ikoną grupy agregatów (taka ikona jest też koło grup agregatów: Internet Sales i Finance).</p> <div data-bbox="552 819 1406 875" data-label="Image"> </div> <p style="text-align: center;">Rys. 5 Nowy wymiar o typie Fact</p>
--	--

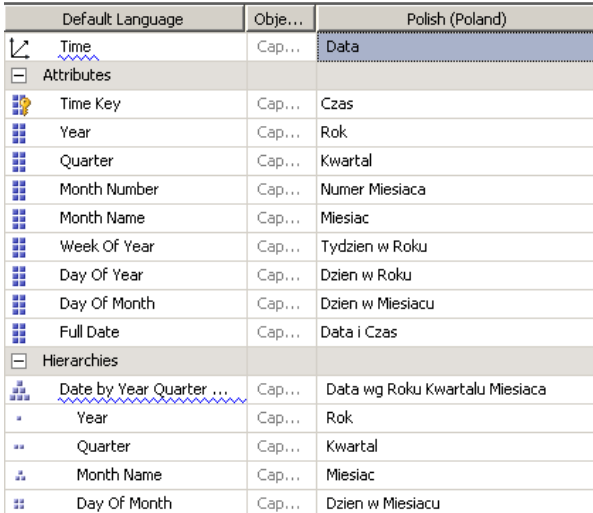
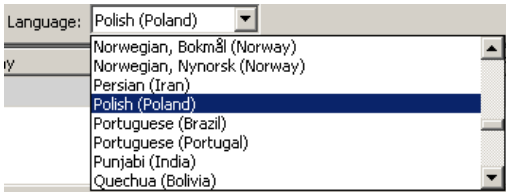
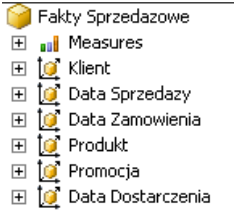
Definiowanie translacji (czas realizacji 15min)

Kostka danych, która została stworzona w poprzednich częściach używa danych i funkcji, co do których zostało przyjęte niejawne założenie, że osoba używająca projektu Analysis Services zna język angielski. Nie zawsze jednak założenie takie jest słuszne i często zdarza się, że konieczne jest przyjęcie założenia, że odbiorcy posługują się tylko językiem ojczystym. Analysis Services umożliwiają całkowite dostosowanie tworzonych kostek danych do wymagań klienta pod względem językowym.

Należy pamiętać, że oprócz języka dostosowania wymagają także formaty wyświetlania danych takie jak: czas (sposób zapisu daty), waluta (inna jednostka monetarna), separator części ułamkowej liczb zmiennoprzecinkowych (kropka vs. przecinek).


Zadanie	Tok postępowania
<p>4. Definiowanie nowej translacji</p>	<ul style="list-style-type: none"> W Solution Explorer kliknij dwukrotnie na kostce Adventure Works DW aby otworzyć widok edycji kostki danych. Wybierz zakładkę Translations. <div data-bbox="517 1529 564 1581" data-label="Image"> </div> <p>Aktualnie nie ma zdefiniowanych translacji, są tylko oznaczenia dla domyślnego języka. Oznaczenia te możemy zmieniać we właściwościach odpowiednich elementów.</p> <ul style="list-style-type: none"> Kliknij prawym przyciskiem myszy w puste pole i wybierz New Translation. <div data-bbox="783 1753 1177 1877" data-label="Image"> </div> <p style="text-align: center;">Rys. 6 Definiowanie nowej translacji</p> <ul style="list-style-type: none"> Wybierz język polski (Polish) i kliknij OK. <div data-bbox="517 1989 564 2040" data-label="Image"> </div> <p>Nowa translacja została utworzona, należy teraz zmodyfikować nazwy</p>





	<p>elementów dla translacji: grupy agregatów, agregaty, poziomy, perspektywy i agregaty kalkulowane.</p>
<p>5. Definiowanie nazw dla nowej translacji</p>	<ul style="list-style-type: none"> Wypełnij nazwy dla nowej translacji.  <p><i>Rys. 7 Przykładowe nazwy dla translacji</i></p> <p> Przykładowe uzupełnienie nazw dla kilku elementów znajduje się na rysunku powyżej.</p>
<p>6. Wypełnianie translacji dla wymiaru</p>	<ul style="list-style-type: none"> W Solution Explorer kliknij dwukrotnie na wymiarze Time aby otworzyć okno edycji wymiaru. Przełącz się na zakładkę Translations.  Dla kostki i każdego wymiaru należy definiować translacje osobno. Jeżeli nie będzie zdefiniowana translacja wybrana przez użytkownika, nazwy będą pochodzić z translacji domyślnej. Kliknij prawym przyciskiem myszy na puste pole i wybierz New Translations.  <p><i>Rys. 8 Nowa translacja dla wymiaru</i></p> <ul style="list-style-type: none"> Wybierz język polski i kliknij OK. Wypełnij nazwy elementów dla nowo utworzonej translacji:




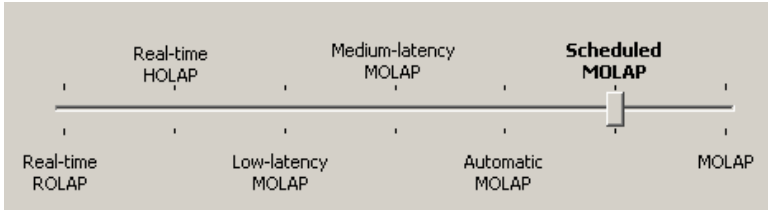


	 <p><i>Rys. 9 Tłumaczenia nazw elementów wymiaru Daty</i></p>
7. Translacja dla wymiaru Product	<ul style="list-style-type: none"> Zdefiniuj translacje dla wymiaru Product dla języka polskiego.
8. Weryfikacja translacji	<ul style="list-style-type: none"> Przeprowadź Process dla kostki danych. Otwórz widok edycji kostki, przełącz się na zakładkę Browser. Przełącz się na język polski.  <p><i>Rys. 10 Zmiana języka na polski</i></p> <ul style="list-style-type: none"> Przełącz się na perspektywę Fakty Sprzedazowe. Sprawdź, jak nazywają się teraz wymiary i poziomy w wymiarze Data.  <p><i>Rys. 11 Polskie nazwy kostki i wymiarów</i></p>



Partycje w kostkach danych (czas realizacji 30min)

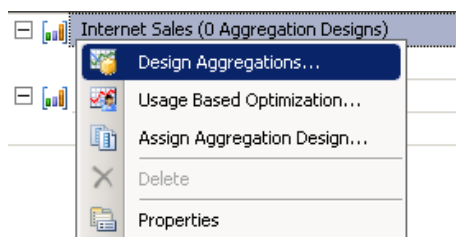
Celem prezentacji jest przedstawienie pojęć partycjonowania i projektowania agregacji. Pokazane zostanie jak utworzyć dwie partycje w oparciu o jedną tabelę faktów oraz jak określić parametry przetwarzania i nośnika dla każdej partycji.

Zadanie	Tok postępowania
9. Sprawdzanie dostępnych partycji dla każdej grupy agregacji	<ul style="list-style-type: none"> W Solution Explorer kliknij dwukrotnie na Adventure Works DW.cube. Wybierz zakładkę Partitions. <p> Zwykle pierwsza partycja jest domyślnie rozwinięta, a pozostałe są zwinięte, co ułatwia przeglądanie wszystkich partycji. W naszym</p>

	<p>przypadku mamy dostępne dwie partycje zdefiniowane domyślnie dla dwóch tablic faktów dostępnych w kostce: Internet Sales, Finance. Domyślnie każda grupa pomiarów ma przypisaną jedną partycję.</p> <ul style="list-style-type: none"> Rozwiń wszystkie partycje. <p> Należy zauważyć, że grupa pomiarów Internet Sales posiada obecnie jedną partycję.</p>
10. Utworzenie nowych partycji	<p> Gdy zachodzi potrzeba utworzenia wielu partycji, konieczne jest zróżnicowanie danych zawartych w każdej z nich. Jedną z możliwości jest posiadanie osobnych tabel faktów. Jest to jednak czasem kłopotliwe. Inna możliwość to określenie podzbioru danych dla każdej partycji przy użyciu klauzuli WHERE. Próba modyfikacji lub utworzenia nowej partycji wykorzystującej tablicę faktów będącą już w użyciu a nieposiadającej klauzuli WHERE spowoduje błąd. Aby utworzyć nową partycję dla istniejącej tablicy faktów konieczne jest ręczne usunięcie istniejącej partycji a następnie utworzenie nowej.</p> <ul style="list-style-type: none"> Kliknij prawym przyciskiem myszy na numerze przy partycji Internet Sales i wybierz Delete. Kliknij na łącze New Partition. Jako Measure group wybierz Internet Sales, w Available tables zaznacz pole przy InternetSales i kliknij Next. Zaznacz pole Specify a query to restrict rows. Przeanalizuj zapytanie które pokazało się w oknie edycji Query. <p> Zostanie automatycznie utworzone zapytanie zawierające pełną listę kolumn tabeli wybranej z Data Source View. W przypadku, gdy jest to pojedyncza tabela można zamienić listę kolumn na "select * from" co automatycznie uwzględni nowe kolumny, jeżeli takie zostaną dodane później do tabeli.</p> <ul style="list-style-type: none"> Na końcu zapytania (po WHERE) dodaj: orderdatekey >= 915 <p> Rekord o kluczu 915 odpowiada dacie 1 Styczeń 2004. Bieżący rok w danych projektu AdventureWorksDW to 2004, więc uwzględnione będą tylko dane z roku 2004. Po dodaniu klauzuli WHERE partycja będzie zawierać tylko podzbiór całej tabeli określony przez wprowadzone wyrażenie warunkowe.</p> <ul style="list-style-type: none"> Kliknij Next. Wybierz Current Server instance i Default Server location i kliknij Next. Zmień nazwę na Fact Internet Sales 2004, zaznacz Design aggregation later. Kliknij Finish.
11. Zadanie utworzenia partycji	<ul style="list-style-type: none"> Zaprojektuj partycje dla pozostałych lat: <ul style="list-style-type: none"> Fact Internet Sales 2003: orderdatekey >= 550 and orderdatekey <= 914 Fact Internet Sales 2002: orderdatekey >= 185 and orderdatekey >= 549 Fact Internet Sales 2001:


	orderdatekey <= 184
12. Zmiana agregacji dla partycji	<p> Może się zdarzyć, że dla pewnej partycji dane zostaną dodane. Co nastąpi w momencie dodania danych? Czy uaktualnienie pojawi się także w kostce? MOLAP czyli Multidimensional OLAP powoduje, że kostka danych korzysta z agregacji obliczanych przy procesie uruchamianym ręcznie co pewien czas. Oznacza to, że najnowsze dane nie znajdą się w kostce dopóki nie wykonamy ponownego procesu. Można to zmienić poprzez zmianę trybu synchronizacji dla partycji.</p> <ul style="list-style-type: none"> W widoku projektu kostki na zakładce Partitions kliknij na partycję Fact Internet Sales 2004 i kliknij na Storage Settings. <p>Sprawdź jakie opisy są przy przeciągnięciu suwaka z pozycji MOLAP po ROLAP.</p> <p> • Kliknij Options.</p> <p> W tym miejscu można ustawić parametry, które są już przygotowane w standardowych wariantach predefiniowanych. MOLAP storage jest specyficzną strukturą przeznaczoną do składowania danych wielowymiarowych. Domyślnie nie jest włączony tryb aktywnego buforowania (proactive caching). Oznacza to, że jedynym sposobem na odświeżenie danych jest ręczne uruchomienie przetwarzania partycji lub poprzez zdefiniowanie zadania za pomocą pakietu Integration Services.</p> <ul style="list-style-type: none"> Z MOLAP przestaw suwak na Scheduled MOLAP.  <p><i>Rys. 12 Opcje przechowywania danych: od ROLAP przez HOLAP do MOLAP</i></p> <ul style="list-style-type: none"> Sprawdź jakie konsekwencje pociąga za sobą poprzednia zmiana – przeczytaj opis i zastanów się co w rzeczywistości się zmieni. Wybierz opcję Custom settings, a następnie naciskamy przycisk Options. <p> Aktualnie jako Storage Settings jest zaznaczony MOLAP, ale można już zaznaczyć lub odznaczyć opcję Enable proactive caching. Poniżej znajduje się zaznaczona opcja Update the cache periodically. Bufor będzie automatycznie uaktualniany co pewien okres czasu. Domyślnie będzie to co 1 dzień, ale czy wiadomo, o której godzinie?</p> <ul style="list-style-type: none"> Zmieniamy okres przebudowy (Rebuild Interval): 10 minutes. <p> Jeżeli chcemy uzyskać przetwarzanie OLAP niemal w czasie rzeczywistym, ustawiamy znacznie częstsze aktualizacje. Oczywiście częstsze aktualizacje to również obciążenie dla systemu. Stosunkowo małe partycje mogą być przetwarzane bardzo często, jednak, jeżeli rozmiar partycji jest dość duży, lepiej żeby nie było potrzeby aktualizowania jej zbyt często.</p> <ul style="list-style-type: none"> Kliknij przycisk Cancel, a następnie ustaw suwak na pozycję Medium-

	<p>latency MOLAP. Kliknij przycisk Options.</p> <p> W tym wariantie dostępne jest nieco inne podejście do aktualizacji. Wybierając opcję Medium-latency MOLAP lub Low-latency MOLAP decyzję o tym, kiedy zawartość bufora staje się nieaktualna podejmujemy na podstawie sposobu zmian danych. Przyjrzyjmy się parametrowi Silence interval. Założmy, że szczytowa liczba transakcji przypada na pierwsze 5 minut każdej godziny. Ponieważ w tym czasie występuje ogromna liczba modyfikacji tabeli faktów, chcemy poczekać aż będzie nieco spokojniej i dopiero wtedy wykonamy agregację. W ten sposób zabezpieczamy się przed aktualizacją po wstawieniu każdego nowego rekordu. Parametr Silence interval określa jak długa cisza jest konieczna aby nastąpiło przetwarzanie danych przez mechanizm Proactive Cache. Jeżeli jednak czas szczytu trwa zbyt długo to ustawiając parametr Silence Override Interval możemy mimo wszystko wymusić przetwarzanie, aby udostępnić najnowsze dane. Takie rozwiązanie najlepiej sprawdza się w przypadku poklastrowanych tabel faktów. Nawet po zmianie parametrów, nie jest możliwe kliknięcie OK – przycisk jest nieaktywny, jest to spowodowane typem partycji, która jest ograniczona poprzez frazę WHERE. Należy zdefiniować zapytanie, które będzie uzupełniało dane z tabeli źródłowej.</p> <ul style="list-style-type: none">• Kliknij na zakładkę Notifications.• Zaznacz Scheduled polling, a następnie kliknij obok <Add Query>.• W dolnym panelu wpisz zapytanie: <pre>SELECT * FROM [dbo].[FactInternetSales] WHERE OrderDateKey >= 915</pre>• Kliknij na tabele powyżej. Sprawdź jak zmieniło się zapytanie w panelu oraz jak wygląda wiersz dla OrderDateKey. <table border="1"><tr><td>OrderDateKey</td><td></td><td>FactIntern...</td><td><input checked="" type="checkbox"/></td><td></td><td></td><td>>= 915</td></tr></table> <p><i>Rys. 13 Dane na temat kolumny OrderDateKey</i></p> <ul style="list-style-type: none">• Kliknij Cancel. <p> Przyjrzelismy się różnym możliwościom realizacji aktualizacji danych w kostce. Można zdecydować się na ręcznie sterowaną aktualizację, jak również można zastosować aktualizacje automatyczne: planowane lub dobierane dynamicznie.</p> <ul style="list-style-type: none">• Ustaw suwak na pozycję MOLAP i kliknij OK.	OrderDateKey		FactIntern...	<input checked="" type="checkbox"/>			>= 915
OrderDateKey		FactIntern...	<input checked="" type="checkbox"/>			>= 915		
13. Konfiguracja opcji agregacji	<ul style="list-style-type: none">• Wybierz zakładkę Aggregations.• Kliknij prawym przyciskiem myszy na Internet Sales i wybierz Design Aggregations.							




Rys. 14 Tworzenie nowych agregacji dla kostki










- Zaznacz partycję dla roku **2004** i kliknij **Next**.
- Dla wymiaru **Order Date** zmień opcję **Default** na **Full** dla poziomów: **Year, Quarter, Month Name**.

 Przy zaznaczaniu pierwszego poziomu na **Full** może pokazać się okno z ostrzeżeniem, że przy poziomach z dużą licznoscią wierszy agregacje mogą zająć bardzo dużo czasu i zasobów, dlatego nie jest zalecane włączanie takich poziomów do agregacji.


- Kliknij **Next**.

 Aktualnie zliczone wiersze są tylko dla grupy agregatów **Internet Sales**. Można sprawdzić, ile jest wierszy w tabelach. Zakładamy, że są to tylko dane testowe pomocne przy budowie kostki i nie odzwierciedla to danych produkcyjnych, których może być kilka lub nawet kilkadziesiąt razy więcej.

- Kliknij **Count** aby zliczyć wiersze dla wszystkich poziomach dla których będzie utworzona agregacja.
- Sprawdź jakie są licznosci poszczególnych poziomów dla wszystkich wymiarów, w szczególności jakie są licznosci wymiarów zaznaczonych w poprzednim kroku.




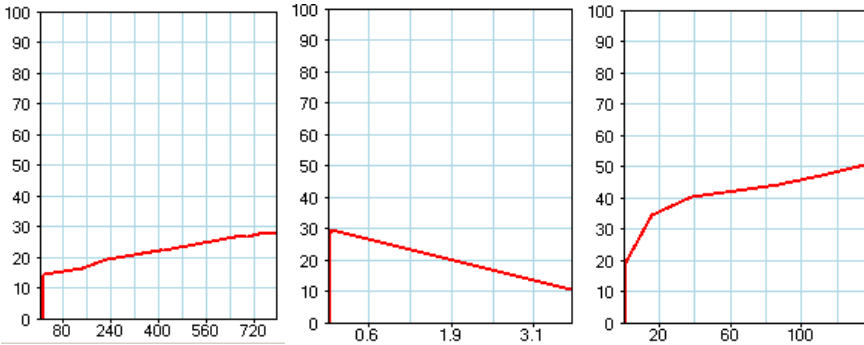
Cube Objects	Estimated Count	Partition Count
 Internet Sales	60,398	32,265
 Ship Date		
 Time Key	1,158	0
 Year	4	
 Order Date		
 Time Key	1,158	0
 Year	4	0
 Quarter	4	0
 Month Name	12	0

Rys. 15 Określanie przybliżonej liczby elementów dla poziomów





 Liczba szacunkowa (**Estimated Count**) jest domyślną dla kostki. Jeżeli podana zostanie liczba wierszy dla partycji, to ona będzie brana pod uwagę. Jeżeli zarówno **Estimated Count** jak i **Partition Count** są równe zero, nie można zdefiniować agregacji. Bieżąca liczba wierszy może się różnić od liczby wierszy na etapie produkcyjnym. Nie trzeba być zbyt dokładnym, jednak nie należy podawać zbyt niskiej wartości. Najgorszym przypadkiem jest pozostawienie wartości testowej i użycie jej w środowisku produkcyjnym bez żadnych zmian. Może to spowodować znaczne obniżenie wydajności.

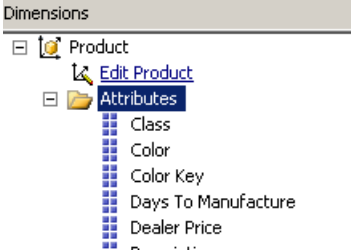


- Ustaw **Estimated Count** dla tabeli faktów: 100,000,000.

 Podanie liczby elementów składowych każdego wymiaru jest konieczne. W kolumnie **Estimated Count** wpisać należy ile rekordów

	<p>przynależy do wybranego atrybutu. Liczbę tę można nadpisać mniejszą lub większą wartością w kolumnie Partition Count.</p> <ul style="list-style-type: none"> Ustaw Estimated Count dla określonych poziomów: <ul style="list-style-type: none"> Product Key: 1000 Customer Key: 20000 Promotion Key: 10000 Promotion Category: 100 Ustaw Partition Count dla tych poziomów na jedną czwartą ich wartości Estimated Count. <p> Zauważmy, iż jedynymi wierszami akceptującymi parametr Partition Count są atrybuty reprezentujące poziom liści. W tej kostce mamy cztery wymiary, z których każdy posiada jeden atrybut, który może być poddany agregacji.</p> <ul style="list-style-type: none"> Kliknij przycisk Next.
14. Strojenie agregacji	<p> Opcje agregacji są kolejną drogą kontrolowania ile agregatów będzie obliczanych.</p> <ul style="list-style-type: none"> Ustaw Performance Gain Reaches na wartość 50% i naciśnij przycisk Start. <p> Wybierając tę opcję nie utworzymy wszystkich możliwych agregacji. Utworzone zostanie tylko tyle, aby zysk wydajności w porównaniu do zapytań wykonywanych bez danych zagregowanych wyniósł 50%. Możemy następnie zobaczyć, ile agregacji zostało zaproponowanych dla bieżącej kombinacji wymiarów w kostce danych.</p> <p>Patrząc na wykres po prawej stronie możemy zaobserwować związek pomiędzy zyskiem wydajności, a rozmiarem miejsca wymaganego na dysku na przechowywanie wyników agregacji. Czasem znacznie większe użycie dysku pociąga za sobą relatywnie niewielki wzrost wydajności. Innym razem niewiele więcej nośnika pozwala na znaczny wzrost wydajności, by za chwilę znów wykres spłaszczył się. Gdy dojdzie już do spłaszczenia wykresu wydajności, kolejne agregacje kompletnie nic nie dają, zajmują tylko niepotrzebnie miejsce na dysku. W przypadku większych i bardziej wyszukanych kostek danych wykres może mieć znacznie bardziej złożony kształt (krzywa) lecz, gdy już dojdzie do jego spłaszczenia, zwiększanie liczby agregacji nie będzie przynosić korzyści.</p> <ul style="list-style-type: none"> Przeanalizuj, ile agregacji zostało wykonanych. <div data-bbox="544 1668 1412 2011">  </div>

Rys. 16 Krzywe określające zysk z agregacji

	<ul style="list-style-type: none"> • Zmień Performance Gain Reaches na wartość 30%, kliknij Reset a następnie Start. <p> Jak widać liczba wierszy każdego z atrybutów wymiarów ma wpływ na zaplanowane agregacje.</p> <ul style="list-style-type: none"> • Przetestuj parametr Gain Reaches dla wartości 55% i 60%. • Naciśnij przycisk Back i zmień parametr Estimated Count poziomu Product Key na 100000. Zmień także Estimated Count dla poziomu Customer Key na 20000. Podobnie jak poprzednio ustaw Partition Count na ¼ wartości Estimated Count. <p> Gdy jeden lub więcej wymiarów posiada większą liczbę rekordów zaobserwujemy, że liczba proponowanych agregacji będzie inna. Czasem może być to mniej agregacji niż poprzednio, gdyż mogłoby się okazać, że rozmiar zbioru agregacji byłby większy o 30% niż sama tabela faktów.</p> <ul style="list-style-type: none"> • Naciśnij przycisk Next i wybierz opcję Save the aggregations but do not process them aby zapisać agregacje bez ich przetwarzania. • Kliknij Finish.
<p>15. Przegląd agregacji utworzonych w zbiorze AdventureWorks. partition</p>	<ul style="list-style-type: none"> • W oknie Solution Explorera naciśnij przycisk Show All Files. <div data-bbox="794 954 1161 1066" data-label="Image"> </div> <p><i>Rys. 17 Przycisk pokazania wszystkich plików w ramach projektu</i></p> <ul style="list-style-type: none"> • Kliknij prawym przyciskiem myszy na Adventure Works DW.partitions i wybierz View Code. • Przewiń otwarte okno w poszukiwaniu znaczników <Aggregation>. <p> Odszukać podany ciąg znaków można za pomocą skrótu klawiszowego Ctrl-F. Przeglądając kod pliku partycji można dostrzec informacje o grupach pomiarów oraz zaplanowanych agregacjach. Można sprawdzić, czy liczba agregacji zgadza się z liczbą znaczników <Aggregation>. Przyjrzyjmy się agregacji Aggregation 0. Jeżeli nie zawiera ona ID atrybutu dla wskazanego wymiaru, to znaczy, że wykorzystuje wszystkie poziomy wskazanego wymiaru. Jeżeli agregacja ma określony atrybut, będzie to agregacja na poziomie liścia. Podsumowując, dla każdej agregacji będzie utworzona tabela przechowująca agregaty konkretnych atrybutów w kombinacji z wszystkimi poziomami wskazanymi wymiarów, dla których nie określono atrybutu. Analizując definicje agregacji jesteśmy w stanie stwierdzić, które wymiary będą agregowane na poziomie atrybutów, a które będą agregowane całościowo.</p>
<p>16. Zmiana agregacji dla poziomów</p>	<p> Jeżeli podczas agregacji chcemy brać pod uwagę szczególne atrybuty, możemy zmienić im typ agregacji na pełny, co spowoduje tworzenie agregacji w oparciu o ten agregat.</p> <ul style="list-style-type: none"> • W widoku edycji kostki Adventure Works DW otwórz zakładkę Cube Structure. • W panelu Dimensions rozwiń atrybuty dla wymiaru Product.

	<div data-bbox="804 208 1155 456"></div> <div data-bbox="517 468 564 524"></div> <div data-bbox="592 468 1441 647"><p>Istnieje wiele produktów, ale mają relatywnie mało kolorów. Jeżeli liczba elementów dla wymiaru jest mała, koszt wykonania agregacji dla poziomu jest także niski. Należy zadać pytanie, czy określony wymiar jest potrzebny i czy będą wykonywane zapytania z tym poziomem.</p></div> <div data-bbox="525 658 1441 960"><ul style="list-style-type: none">• Na atrybucie Color kliknij prawym przyciskiem myszy i wybierz Properties.• W oknie właściwości zmień właściwość AggregationUsage na Full.• Przełącz się na zakładkę Aggregations.• Na Internet Sales kliknij prawym przyciskiem myszy i wybierz Design Aggregations.• Zaznacz Fact Internet Sales 2004 i kliknij Next.• Sprawdź atrybut Color w wymiarze Product.</div> <div data-bbox="517 972 564 1028"></div> <div data-bbox="592 972 1441 1016"><p>Jak można zauważyć, atrybut Color ma już ustawioną wartość na Full.</p></div> <div data-bbox="525 1028 732 1061"><ul style="list-style-type: none">• Kliknij Cancel.</div>
--	--

Laboratorium rozszerzone

Zadanie 1 (czas realizacji 45min)

Utwórz nową kostkę **Finance** zawierającą wymiary: **Time, Account, Department, Scenario, Organization, Currency** (wymiar poprzez tabelę **DimOrganization**). Dla wymiarów utwórz stosowne hierarchie. Dla kostki utwórz perspektywy zawierające następujące wymiary:

- **Time, Account**
- **Time, Scenario, Currency.**

Utwórz nowe agregaty kalkulowane, które wyznaczają podatki od sumy operacji. Wartości ujemne powinny się wyświetlać w kolorze czerwonym. Wartości dodatnie powyżej 1000 powinny się wyświetlać w kolorze niebieskim.

- **Tax Amount 7**
- **Tax Amount 22**

Utwórz wymiar pozwalający obliczyć, na jaką sumę były operacje w następujących przedziałach (wartości dodatnie i ujemne):

- Poniżej -20000
- Od -20000 do 0 (rozłącznie)
- Od 0 do 5000
- Od 5000 do 20000
- Powyżej 20000

Utwórz translację języka polskiego dla kostki **Finance**.

Utwórz partycje dla kostki **Finance** wg roku i scenariusza (4 lata * 3 scenariusze). Dla partycji z ostatniego roku ustaw opcje przechowywania (**Storage Settings**) na **Medium-latency MOLAP**.

Utwórz agregacje dla kostki **Finance** zgodnie z raportami przedstawionymi poniżej.

Za pomocą tabel przestawnych utwórz następujące raporty:

- Suma wypłat i podatku 7% w zależności od scenariusza, roku i kwartału.
- Suma wypłat i podatku 7% w zależności od departamentu (3 poziomy).
- Suma wypłat w zależności od scenariusza i waluty.
- Suma wypłat w zależności od wielkości wypłat (w odpowiednio utworzonych wcześniej przedziałach).

Zadanie 5 (czas realizacji 45 min)

Firma **DziekanatSoftware** zaprojektowała i wykonała system dziekanatowy **SuperDziekanat** do obsługi dziekanatu na uczelniach. System bazodanowy złożony jest z kilkudziesięciu tabel połączonych ze sobą kluczami obcymi. Problem jest z modulem raportującym dla tego systemu. Każdy raport napisany jest w języku SQL, następnie dodawany do aplikacji i wykonywany na bazie danych. Raporty są przekrojowe, dlatego cała baza jest skanowana, co zajmuje sporo czasu i odbywa się w czasie największego użycia systemu. Przy kilkunastu tysiącach studentów w największych tabelach jest nawet kilka milionów rekordów z kilku ostatnich lat, dlatego użycie bazy danych dla raportów nie wchodzi w grę. Ponadto, użytkownicy biznesowi ze znajomością narzędzia **MS SQL Server Management Studio** chcieliby mieć możliwość tworzenia raportów za pomocą tabel przestawnych.

Zadanie to jest kontynuacją zadania z poprzedniego modułu.

Rozwiązaniem jest utworzenie kostek danych ze schematu relacyjnego, a następnie zdefiniowaniem agregatów, wymiarów, poziomów i hierarchii.

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

Zapoznaj się ze schematem relacyjnym, a następnie utwórz kostki danych (w nawiasie są proponowane wymiary): ocena (student, data studenta, data, grupa, typ oceny), obecność (student, zajecie, data dla zajecia), logowanie (uzytkownik, data), zajecie (grupa, termin).

Za pomocą tabel przestawnych zaprojektuj następujące raporty:

- Średnia ocena w zależności od typu studiów.
- Średnia ocena w zależności od instytutu i roku dziekańskiego.
- Liczba nieobecności w zależności od przedmiotu.
- Procentowa obecność w zależności od prowadzącego.

Utwórz agregacje i partycje wg typowych raportów.

Utwórz perspektywy zawierające czas i jeden wybrany wymiar.