

Metody przetwarzania dużych ilości danych

Sprawozdanie z laboratorium

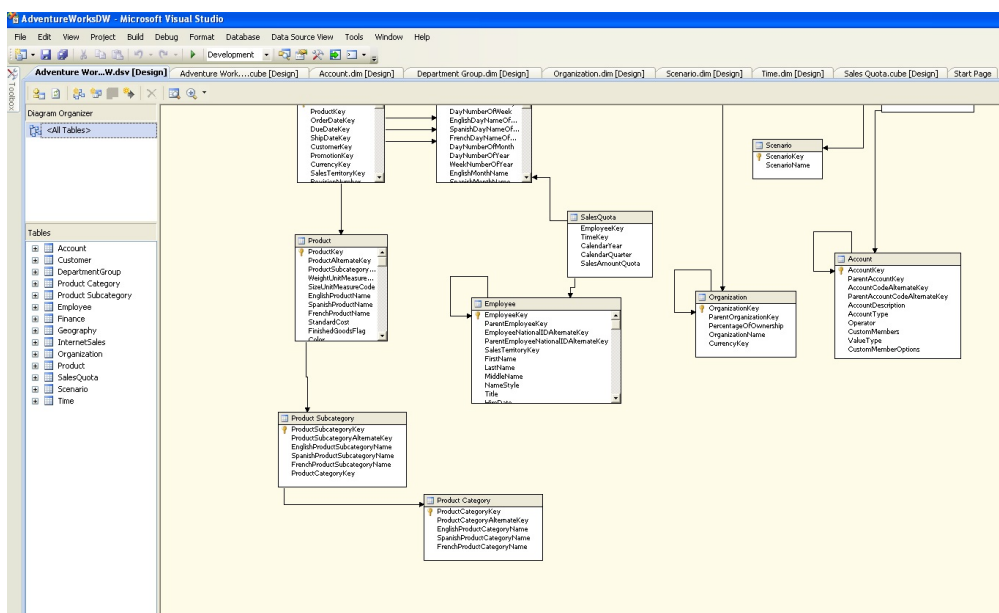
Data	Tytuł zajęć	Uczestnicy
01.12.2020 18:55	Zaawansowana analiza danych	Adam Hyjek (234987) Bartosz Rodziewicz (226105)

Streszczenie laboratorium

Instrukcja laboratoryjna nr 5 przedstawia metodę definiowania nowego wymiaru z wykorzystaniem tabeli pośredniej oraz metodę definiowania wymiaru wykorzystując tabelę faktów. Przedstawiony tam jest również proces tworzenia tłumaczenia (translacji) na inny język opisów kostki danych oraz wymiarów. Opisana również jest metoda definiowania własnych partycji w kostkach danych oraz tworzenia własnych agregacji danych w tych partycjach. Wszystkie te zadania (podstawowe) są przedstawione poniżej w tym sprawozdaniu.

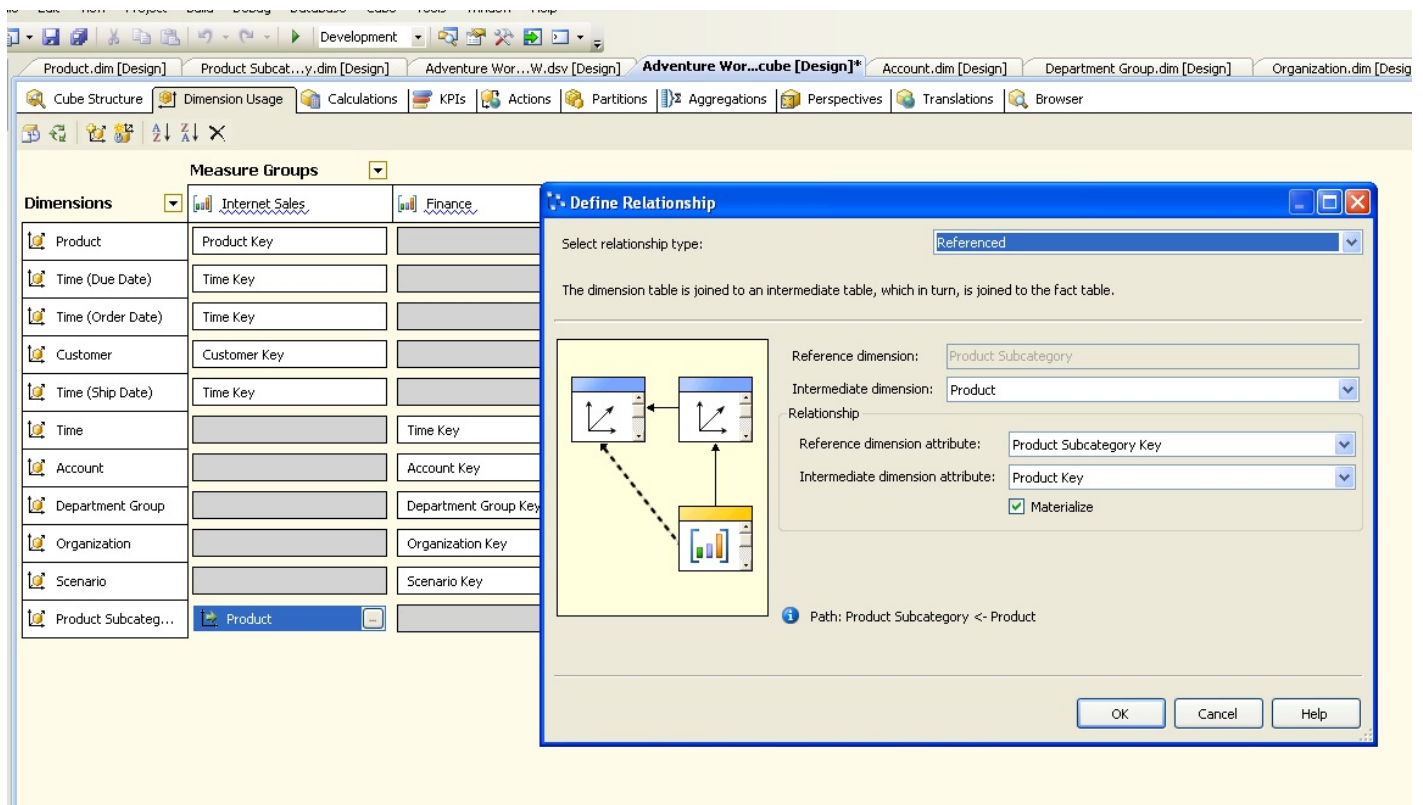
Różne rodzaje wymiarów

Definiowanie wymiaru połączonego przez tabelę pośrednią



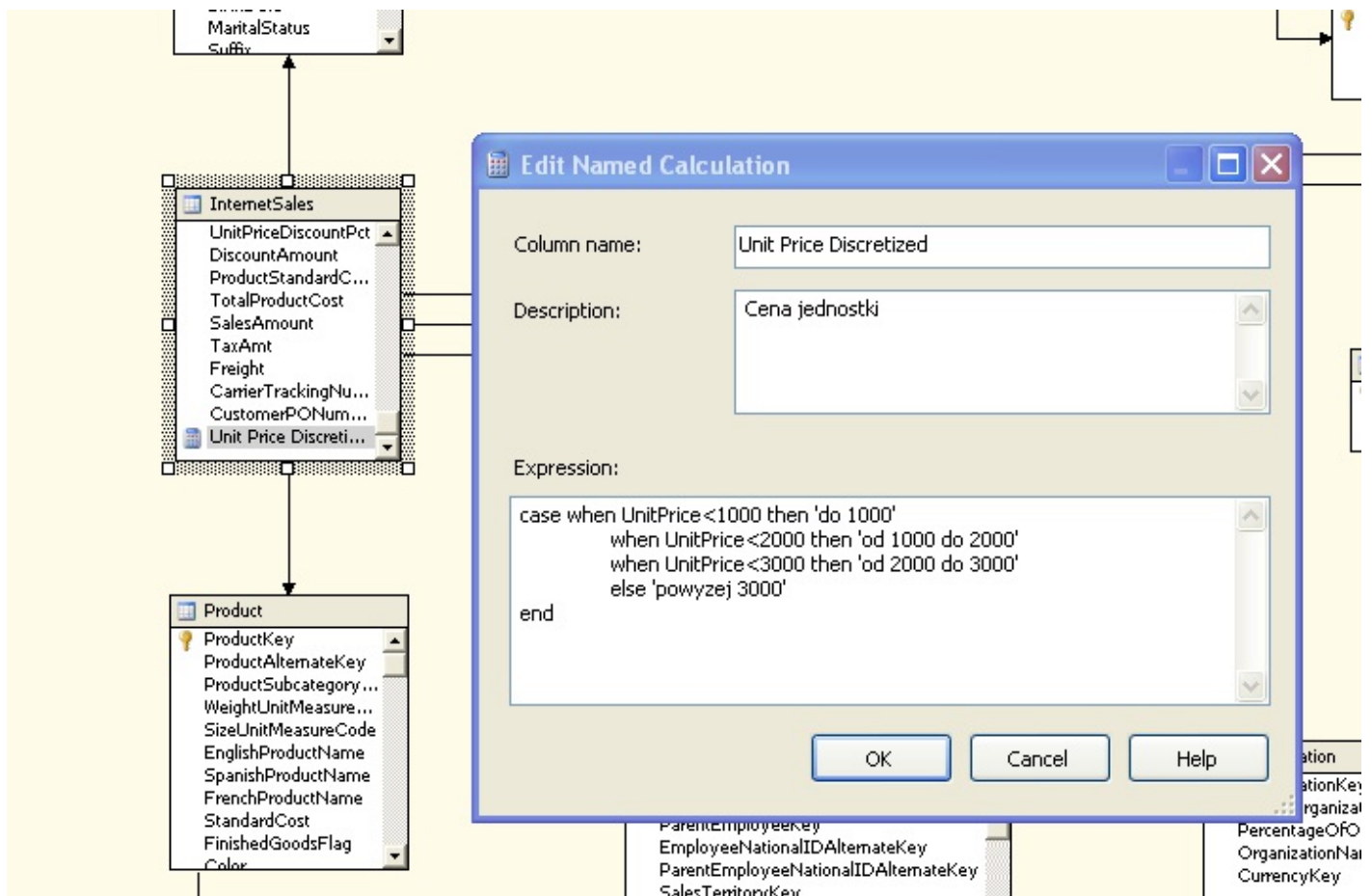
Dodanie brakujących tabeli Product Category i Product Subcategory.

Tworzenie nowego wymiaru Product Subcategory.

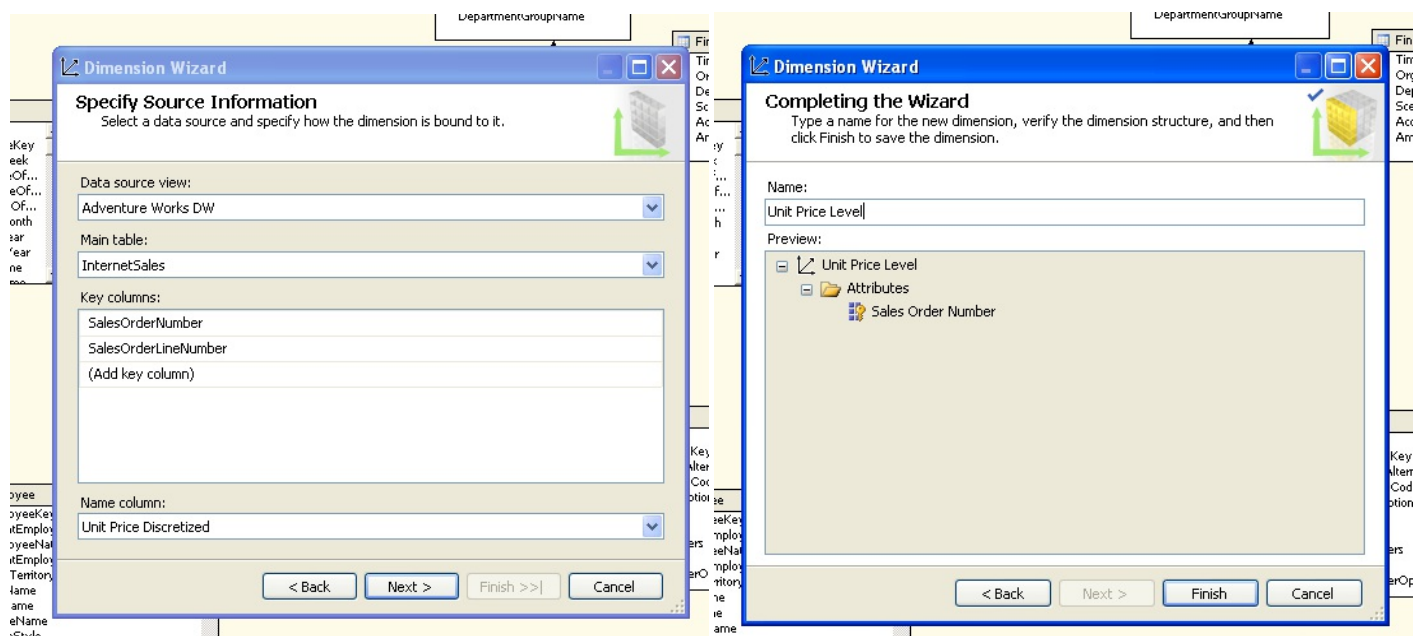


Dodanie nowego wymiaru do measure grupy i ustawienie jego parametrów.

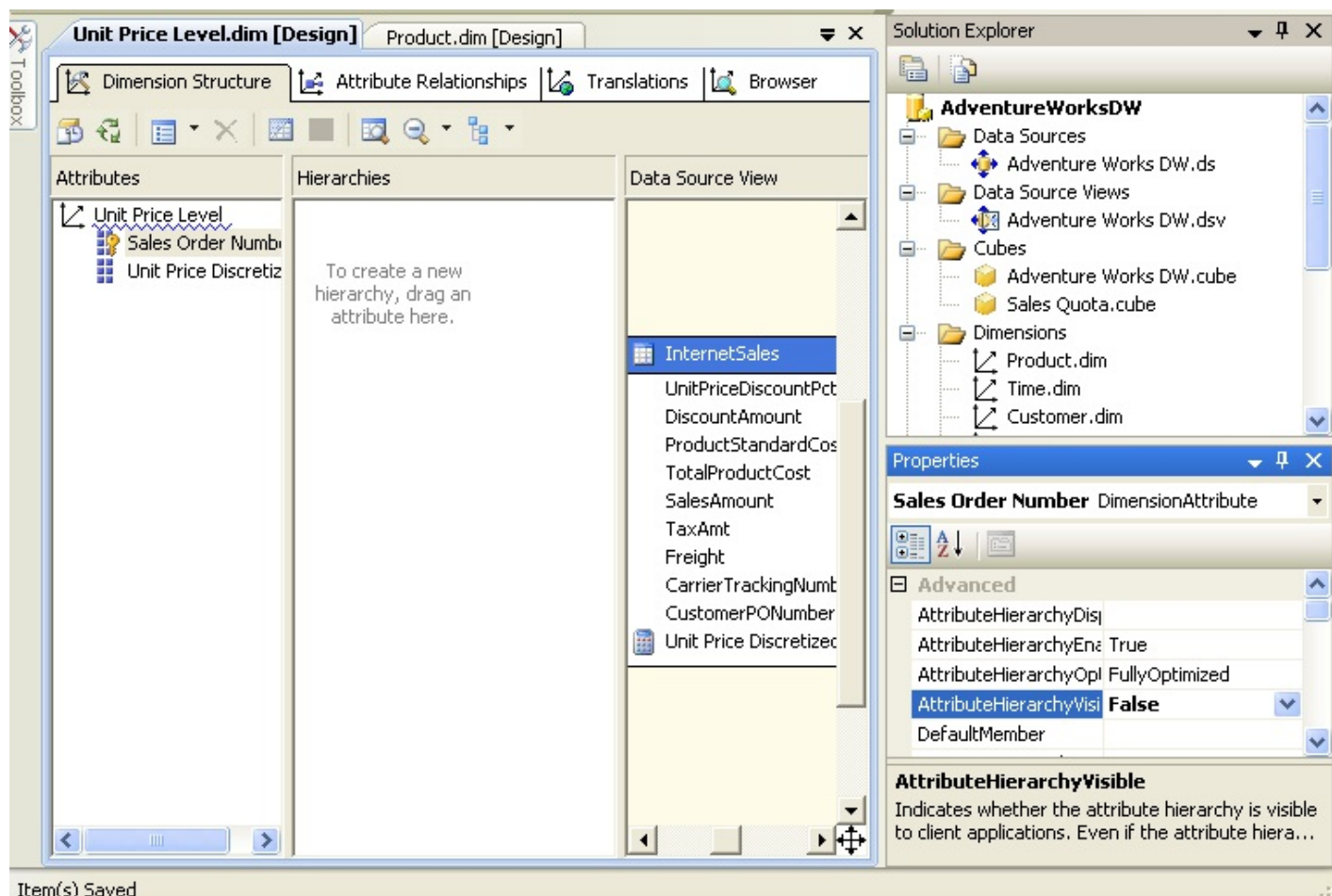
Tworzenie wymiaru z tabeli faktów



Dodanie nowego wyliczania.

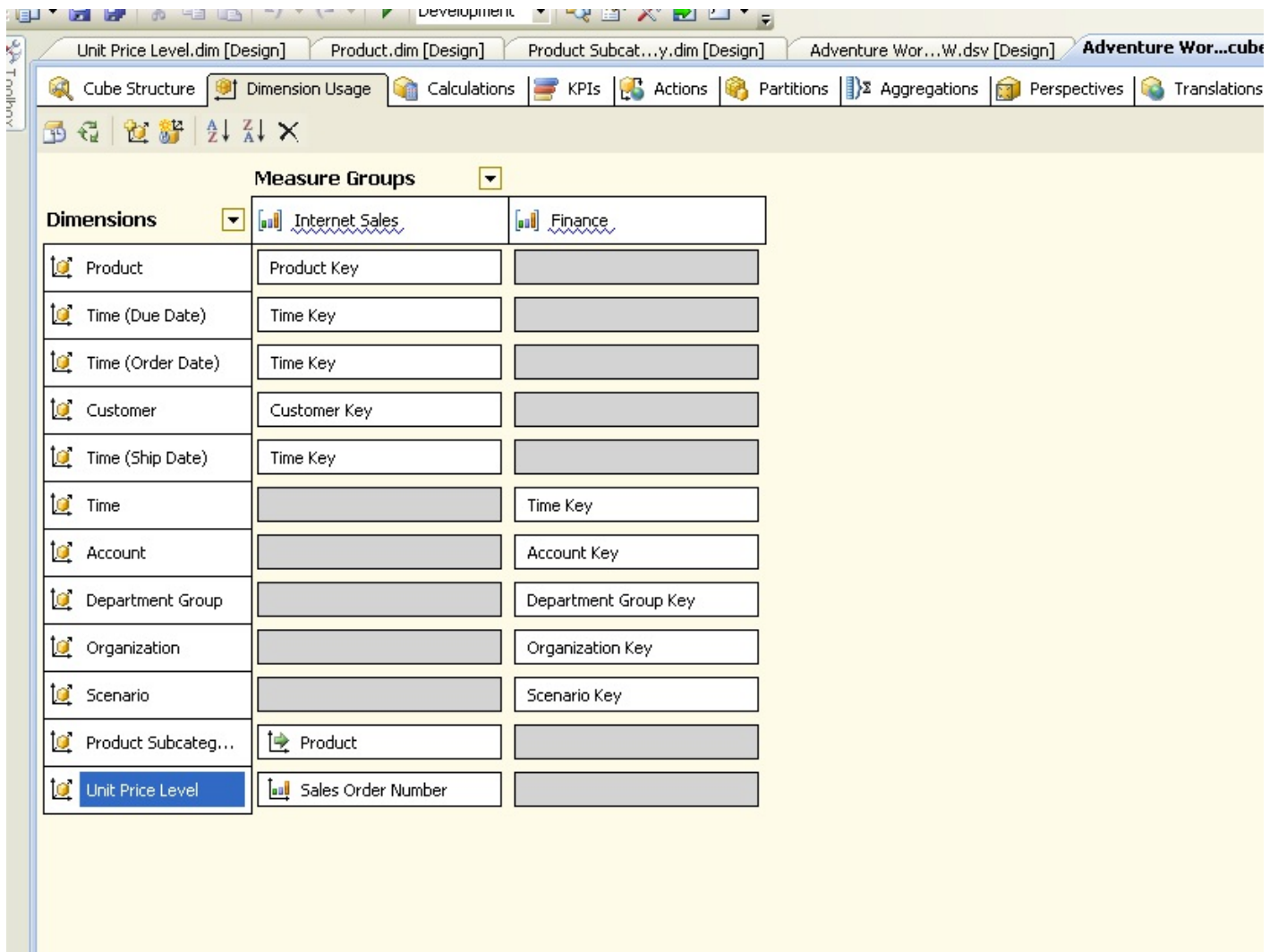


Tworzenie nowego wymiaru.



Item(s) Saved

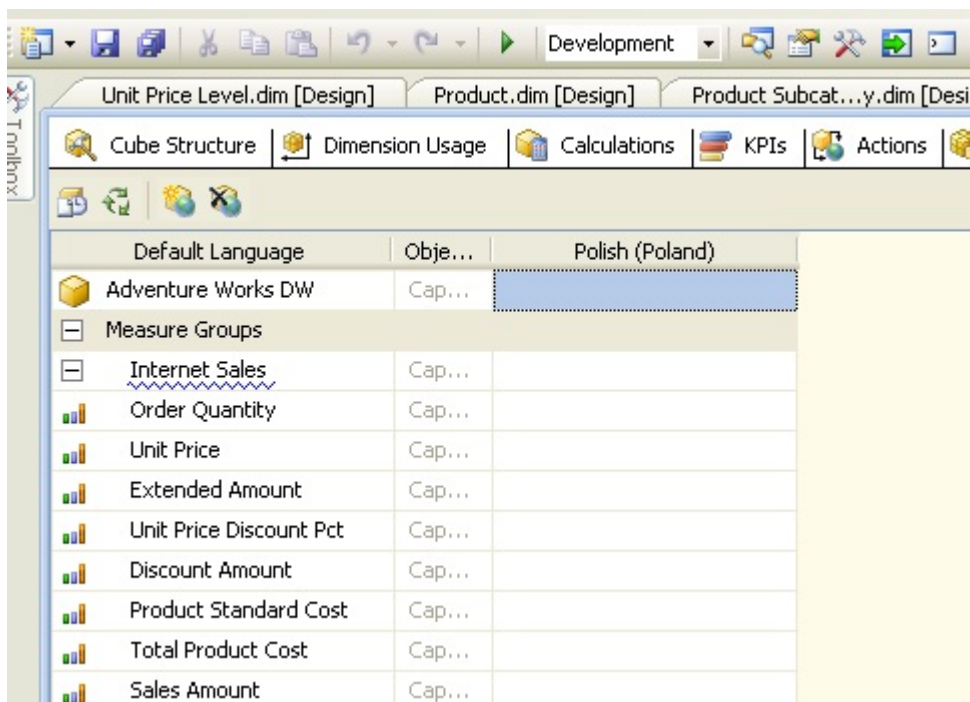
Dodanie atrybutu z kolumny oraz zmiana parametru.



Dodanie wymiaru do measure grupy.

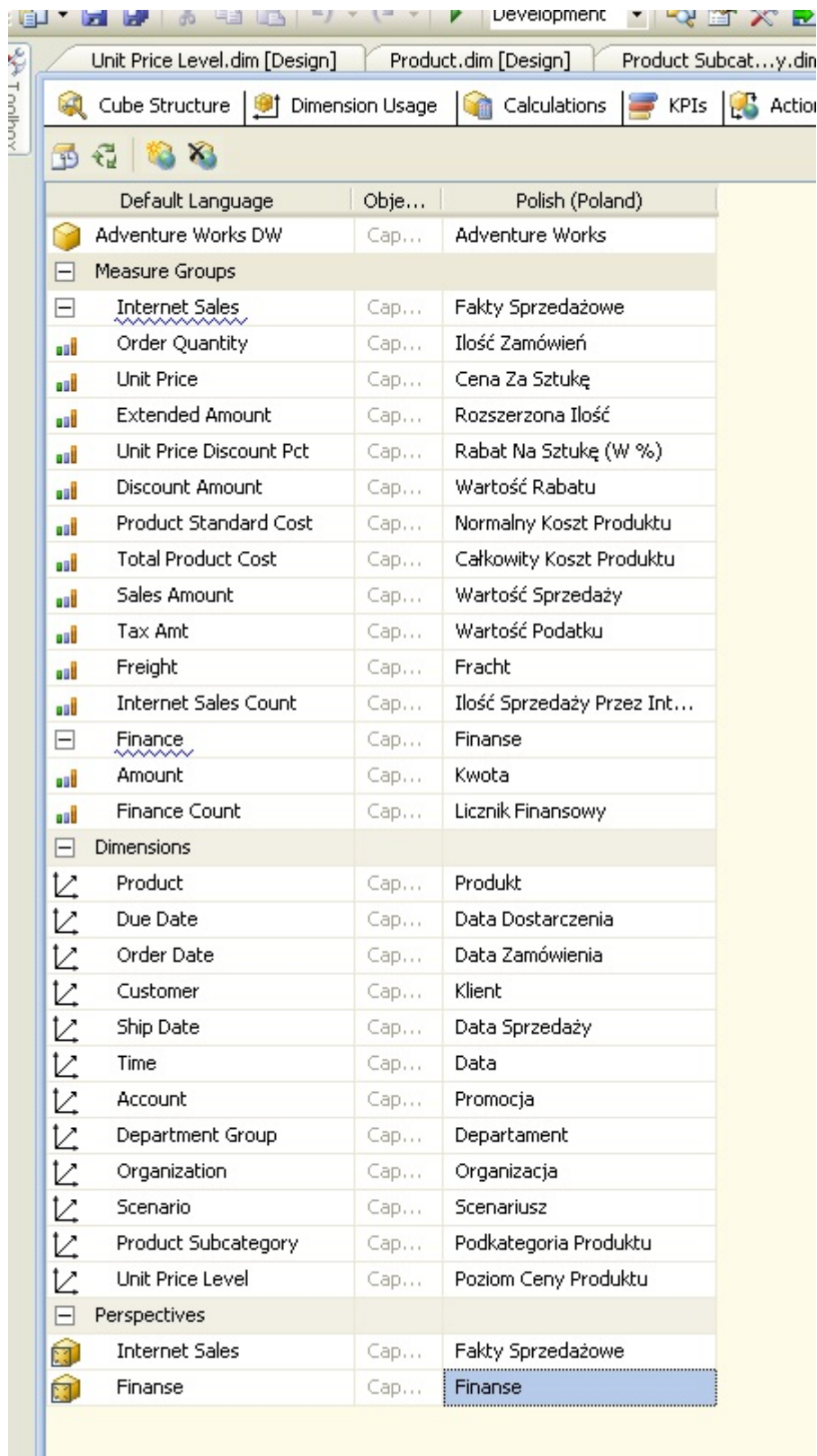
Definiowanie translacji

Definiowanie nowej translacji



Dodanie nowej translacji na j. polski.

Definiowanie nazw dla nowej translacji



Default Language	Obje...	Polish (Poland)
Adventure Works DW	Cap...	Adventure Works
Measure Groups		
Internet Sales	Cap...	Fakty Sprzedażowe
Order Quantity	Cap...	Ilość Zamówień
Unit Price	Cap...	Cena Za Sztukę
Extended Amount	Cap...	Rozszerzona Ilość
Unit Price Discount Pct	Cap...	Rabat Na Sztukę (W %)
Discount Amount	Cap...	Wartość Rabatu
Product Standard Cost	Cap...	Normalny Koszt Produktu
Total Product Cost	Cap...	Całkowity Koszt Produktu
Sales Amount	Cap...	Wartość Sprzedaży
Tax Amt	Cap...	Wartość Podatku
Freight	Cap...	Fracht
Internet Sales Count	Cap...	Ilość Sprzedaży Przez Int...
Finance	Cap...	Finanse
Amount	Cap...	Kwota
Finance Count	Cap...	Licznik Finansowy
Dimensions		
Product	Cap...	Produkt
Due Date	Cap...	Data Dostarczenia
Order Date	Cap...	Data Zamówienia
Customer	Cap...	Klient
Ship Date	Cap...	Data Sprzedaży
Time	Cap...	Data
Account	Cap...	Promocja
Department Group	Cap...	Departament
Organization	Cap...	Organizacja
Scenario	Cap...	Scenariusz
Product Subcategory	Cap...	Podkategoria Produktu
Unit Price Level	Cap...	Poziom Ceny Produktu
Perspectives		
Internet Sales	Cap...	Fakty Sprzedażowe
Finance	Cap...	Finanse

Uzupełnienie tłumaczeń kostki dla tłumaczenia na j. polski (mamy nadzieję, że nie będzie oceniana jakość tłumaczenia).

Wypełnianie translacji dla wymiaru

Default Language	Object Type	Polish (Poland)
Time	Caption	Data
[All Dates]	AttributeAllMemberName	[Wszystkie Daty]
[Unknown Date]	UnknownMemberName	[Nieznana Data]
Attributes		
Year	Caption	Rok
Quarter	Caption	Kwartał
Month Number	Caption	Numer Miesiąca
Month Name	Caption	Miesiąc
Week Of Year	Caption	Tydzień Roku
Day Of Year	Caption	Dzień Roku
Day Of Month	Caption	Dzień Miesiąca
Full Date	Caption	Pełna Data
Hierarchies		
Date by Year Quarter Month	Caption	Data Podzielona Na Rok K...
Year	Caption	Rok
Quarter	Caption	Kwartał
Month Name	Caption	Miesiąc
Day Of Month	Caption	Dzień Miesiąca
Full Date	Caption	Pełna Data

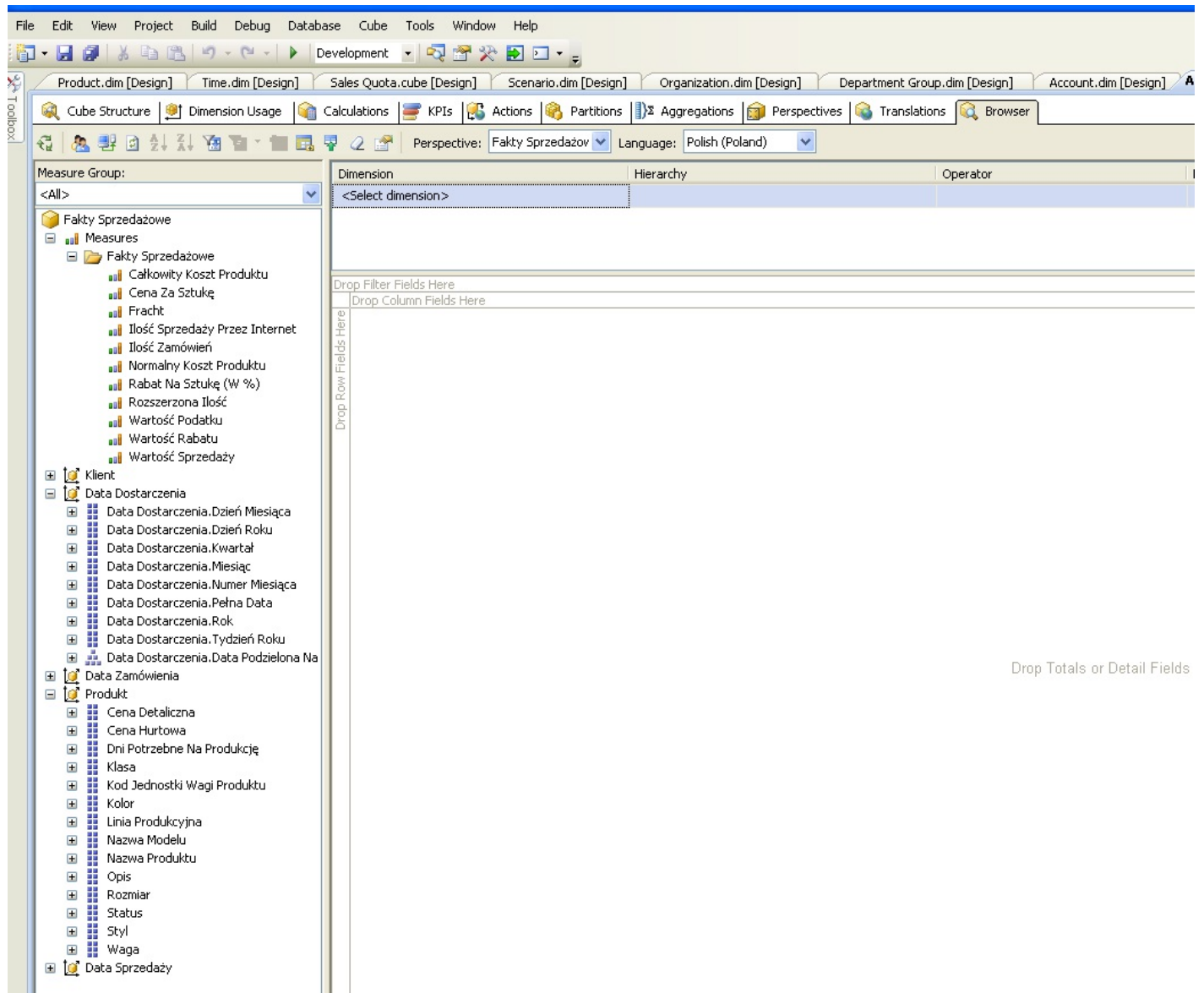
Uzupełnienie tłumaczeń wymiaru Time dla tłumaczenia na j. polski (mamy nadzieję, że nie będzie oceniana jakość tłumaczenia).

Translacja dla wymiaru Product

Default Language	Object Type	Polish (Poland)
Product	Caption	Produkt
[All Products]	AttributeAllMemberName	[Wszystkie Produkty]
[Other Products]	UnknownMemberName	[Inne Produkty]
Attributes		
Product Name	Caption	Nazwa Produktu
Weight Unit Measure Code	Caption	Kod Jednostki Wagi Produktu
Color	Caption	Kolor
Size	Caption	Rozmiar
Weight	Caption	Waga
Product Line	Caption	Linia Produkcyjna
Dealer Price	Caption	Cena Detaliczna
List Price	Caption	Cena Hurtowa
Days To Manufacture	Caption	Dni Potrzebne Na Produkcję
Class	Caption	Klasa
Style	Caption	Styl
Model Name	Caption	Nazwa Modelu
Description	Caption	Opis
Status	Caption	Status

Uzupełnienie tłumaczeń wymiaru Produkt dla tłumaczenia na j. polski.

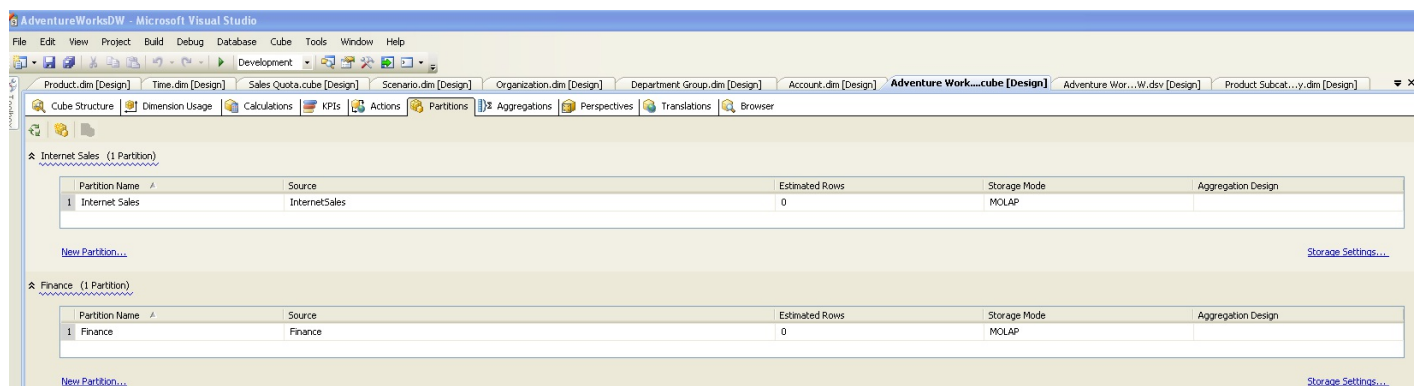
Weryfikacja translacji



Polskie nazwy widoczne w zakładce Browser.

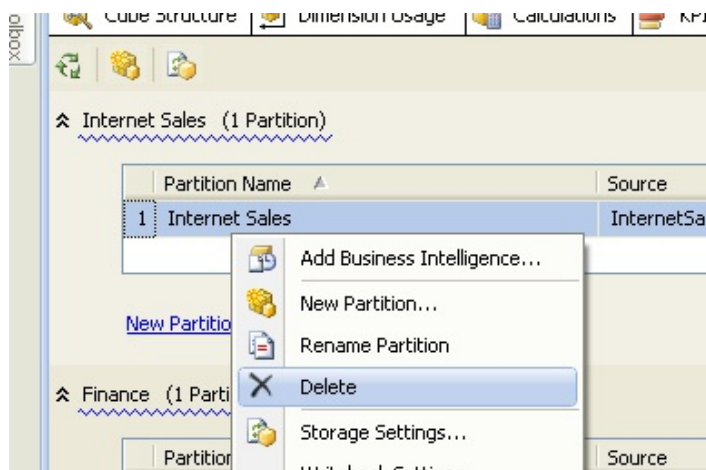
Partycje w kostkach danych

Sprawdzanie dostępnych partycji dla każdej grupy agregacji

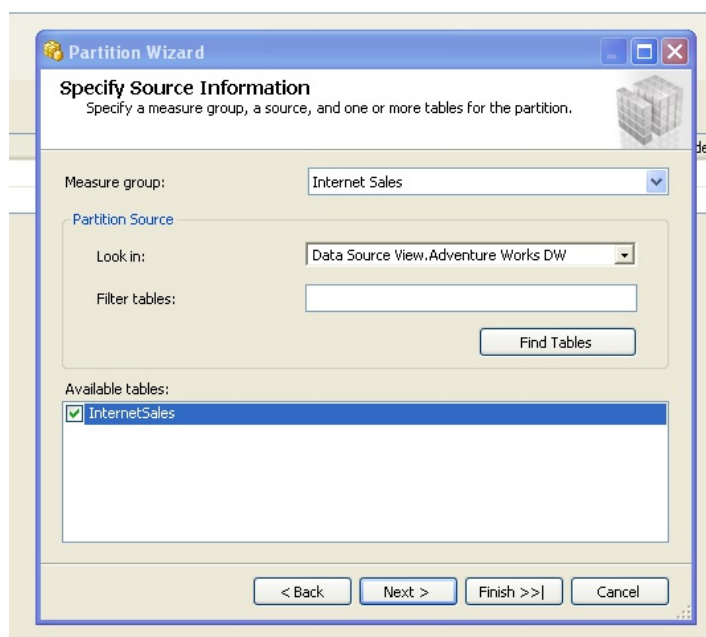


Domyślne partycje przypisane do grup pomiarów. Grupa Internet Sales posiada tylko jedną partycję.

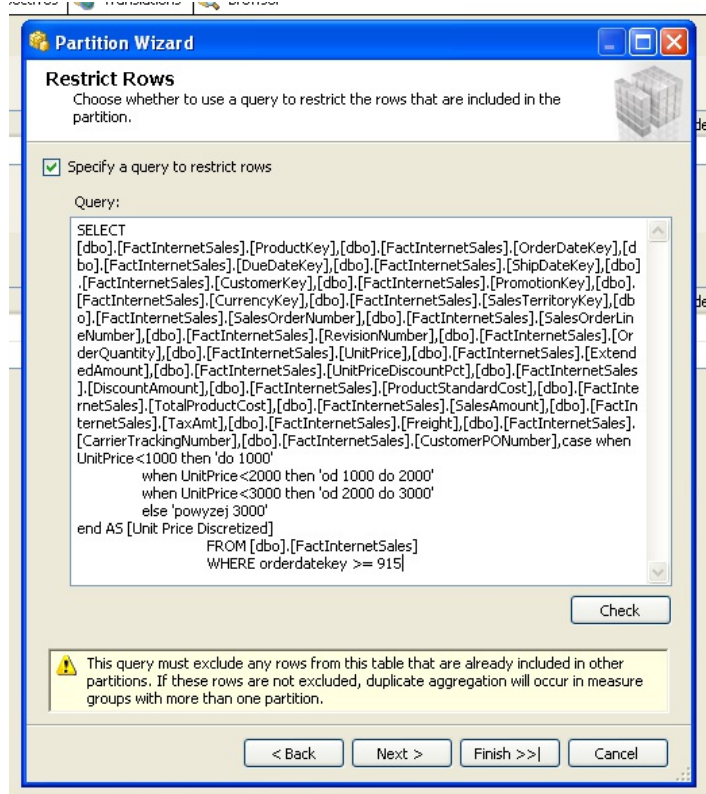
Utworzenie nowych partycji



Do utworzenia ręcznych partycji konieczne jest usunięcie domyślnej.



Tworzenie nowej partycji - dane źródła.



Tworzenie nowej partycji - zapytanie ograniczające.

Partition Wizard

Completing the Wizard

Provide a name for the partition, set aggregation options, and then click Finish to save the partition.

Name:

Aggregation options:

☐ Design aggregations for the partition now

☒ Design aggregations later

☐ Copy the aggregation design from an existing partition

Copy from:

☐ Deploy and Process Now

< Back Next > Finish Cancel

Tworzenie nowej partycji - ustalenie poprawnej nazwy partycji.

Zadanie utworzenia partycji

Partition Wizard

Restrict Rows

Choose whether to use a query to restrict the rows that are included in the partition.

☒ Specify a query to restrict rows

Query:

```

} [DiscountAmount],[dbo].[FactInternetSales].[ProductStandardCost],[dbo].[FactInternetSales].[TotalProductCost],[dbo].[FactInternetSales].[SalesAmount],[dbo].[FactInternetSales].[TaxAmt],[dbo].[FactInternetSales].[Freight],[dbo].[FactInternetSales].[CarrierTrackingNumber],[dbo].[FactInternetSales].[CustomerPONumber],case when
UnitPrice<1000 then 'do 1000'
when UnitPrice<2000 then 'od 1000 do 2000'
when UnitPrice<3000 then 'od 2000 do 3000'
else 'powyzsze 3000'
end AS [Unit Price Discretized]
FROM [dbo].[FactInternetSales]
WHERE orderdatekey >= 550 and orderdatekey <= 914

```

Check

This query must exclude any rows from this table that are already included in other partitions. If these rows are not excluded, duplicate aggregation will occur in measure groups with more than one partition.

< Back Next > Finish >> Cancel

Tworzenie 3 kolejnych partycji.

Cube Structure Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

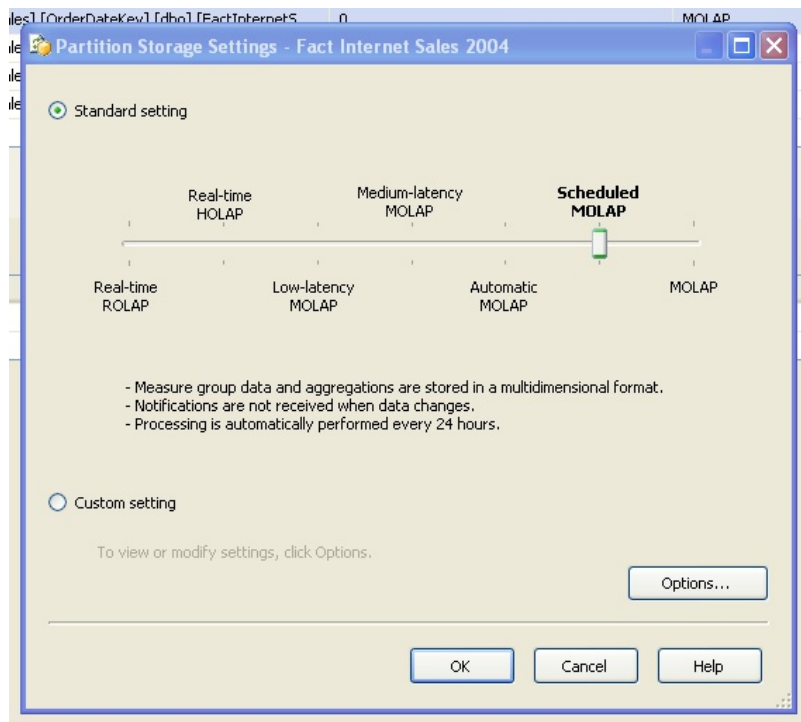
Internet Sales (4 Partitions)

Partition Name	Source	Estimated Rows	Storage Mode	Aggre
1 Fact Internet Sales 2004	SELECT [dbo].[FactInternetSales].[ProductKey],[dbo].[FactInternetSales].[OrderDateKey],[dbo].[FactInternetSales].[SalesAmount],[dbo].[FactInternetSales].[TaxAmt],[dbo].[FactInternetSales].[Freight],[dbo].[FactInternetSales].[CarrierTrackingNumber],[dbo].[FactInternetSales].[CustomerPONumber],case when UnitPrice<1000 then 'do 1000' when UnitPrice<2000 then 'od 1000 do 2000' when UnitPrice<3000 then 'od 2000 do 3000' else 'powyzsze 3000' end AS [Unit Price Discretized] FROM [dbo].[FactInternetSales] WHERE orderdatekey >= 550 and orderdatekey <= 914	0	MOLAP	
2 Fact Internet Sales 2003	SELECT [dbo].[FactInternetSales].[ProductKey],[dbo].[FactInternetSales].[OrderDateKey],[dbo].[FactInternetSales].[SalesAmount],[dbo].[FactInternetSales].[TaxAmt],[dbo].[FactInternetSales].[Freight],[dbo].[FactInternetSales].[CarrierTrackingNumber],[dbo].[FactInternetSales].[CustomerPONumber],case when UnitPrice<1000 then 'do 1000' when UnitPrice<2000 then 'od 1000 do 2000' when UnitPrice<3000 then 'od 2000 do 3000' else 'powyzsze 3000' end AS [Unit Price Discretized] FROM [dbo].[FactInternetSales] WHERE orderdatekey >= 185 and orderdatekey <= 549	0	MOLAP	
3 Fact Internet Sales 2002	SELECT [dbo].[FactInternetSales].[ProductKey],[dbo].[FactInternetSales].[OrderDateKey],[dbo].[FactInternetSales].[SalesAmount],[dbo].[FactInternetSales].[TaxAmt],[dbo].[FactInternetSales].[Freight],[dbo].[FactInternetSales].[CarrierTrackingNumber],[dbo].[FactInternetSales].[CustomerPONumber],case when UnitPrice<1000 then 'do 1000' when UnitPrice<2000 then 'od 1000 do 2000' when UnitPrice<3000 then 'od 2000 do 3000' else 'powyzsze 3000' end AS [Unit Price Discretized] FROM [dbo].[FactInternetSales] WHERE orderdatekey <= 184	0	MOLAP	
4 Fact Internet Sales 2001	SELECT [dbo].[FactInternetSales].[ProductKey],[dbo].[FactInternetSales].[OrderDateKey],[dbo].[FactInternetSales].[SalesAmount],[dbo].[FactInternetSales].[TaxAmt],[dbo].[FactInternetSales].[Freight],[dbo].[FactInternetSales].[CarrierTrackingNumber],[dbo].[FactInternetSales].[CustomerPONumber],case when UnitPrice<1000 then 'do 1000' when UnitPrice<2000 then 'od 1000 do 2000' when UnitPrice<3000 then 'od 2000 do 3000' else 'powyzsze 3000' end AS [Unit Price Discretized] FROM [dbo].[FactInternetSales] WHERE orderdatekey <= 184	0	MOLAP	

New Partition...

Utworzone 4 partycje.

Zmiana agregacji dla partycji



Zmiana trybu agregacji danych.

W programie dostępne są następujące presety:

- MOLAP
 - dane measure grup oraz agregacji przechowywane są w formacie wielowymiarowym
 - powiadomienia są pokazywane, gdy dane ulegną zmianie
 - wczytywanie nowych danych jest wykonywane ręcznie
- Scheduled MOLAP
 - różni się tym od MOLAP, że wczytywanie nowych danych jest wykonywane automatycznie co 24h
- Automatic MOLAP
 - różni się tym od MOLAP, że powiadomienia o zmianie danych wywołają automatyczne wczytywanie danych, co powoduje, że nowe dane są wczytywane bez opóźnień
- Medium-latency MOLAP
 - różni się tym od Automatic MOLAP, że dane nie są wczytywane bez opóźnienia, a z opóźnieniem maksymalnie kilku godzinnym.
- Low-latency MOLAP
 - różni się tym od Automatic MOLAP, że dane nie są wczytywane bez opóźnienia, a z opóźnieniem maksymalnie 30 minutowym.
- Real-time HOLAP
 - różni się tym od MOLAP, że dane measure grup przechowywane są w formacie relacyjnym, natomiast agregacje w formacie wielowymiarowym; dane zawsze oddają stan rzeczywisty
- Real-time ROLAP
 - różni się tym od Real-time HOLAP, że i measure grupy, jak i agregacje są przechowywane w formacie relacyjnym

Zmiana z MOLAP na Scheduled MOLAP powoduje, że hurtownia danych będzie automatycznie co 24h aktualizowała swoje dane.

Storage Options

Storage mode: MOLAP

☒ Enable proactive caching

General | Notifications

Cache Settings

☐ Update the cache when data changes

Silence interval: (Not enabled)

Silence override interval: (Not enabled)

☐ Drop outdated cache

Latency: (Not enabled)

☒ Update the cache periodically

Rebuild interval: Minutes

Options

☐ Bring online immediately ☐ Apply settings to dimensions

☐ Enable ROLAP aggregations

OK Cancel Help

Szczegółowe ustawienia Scheduled MOLAP.

Storage Options

Storage mode: MOLAP

☒ Enable proactive caching

General | Notifications

Cache Settings

☒ Update the cache when data changes

Silence interval: Seconds

Silence override interval: Minutes

☒ Drop outdated cache

Latency: Hours


☐ Update the cache periodically

Rebuild interval: (Not enabled)

Options

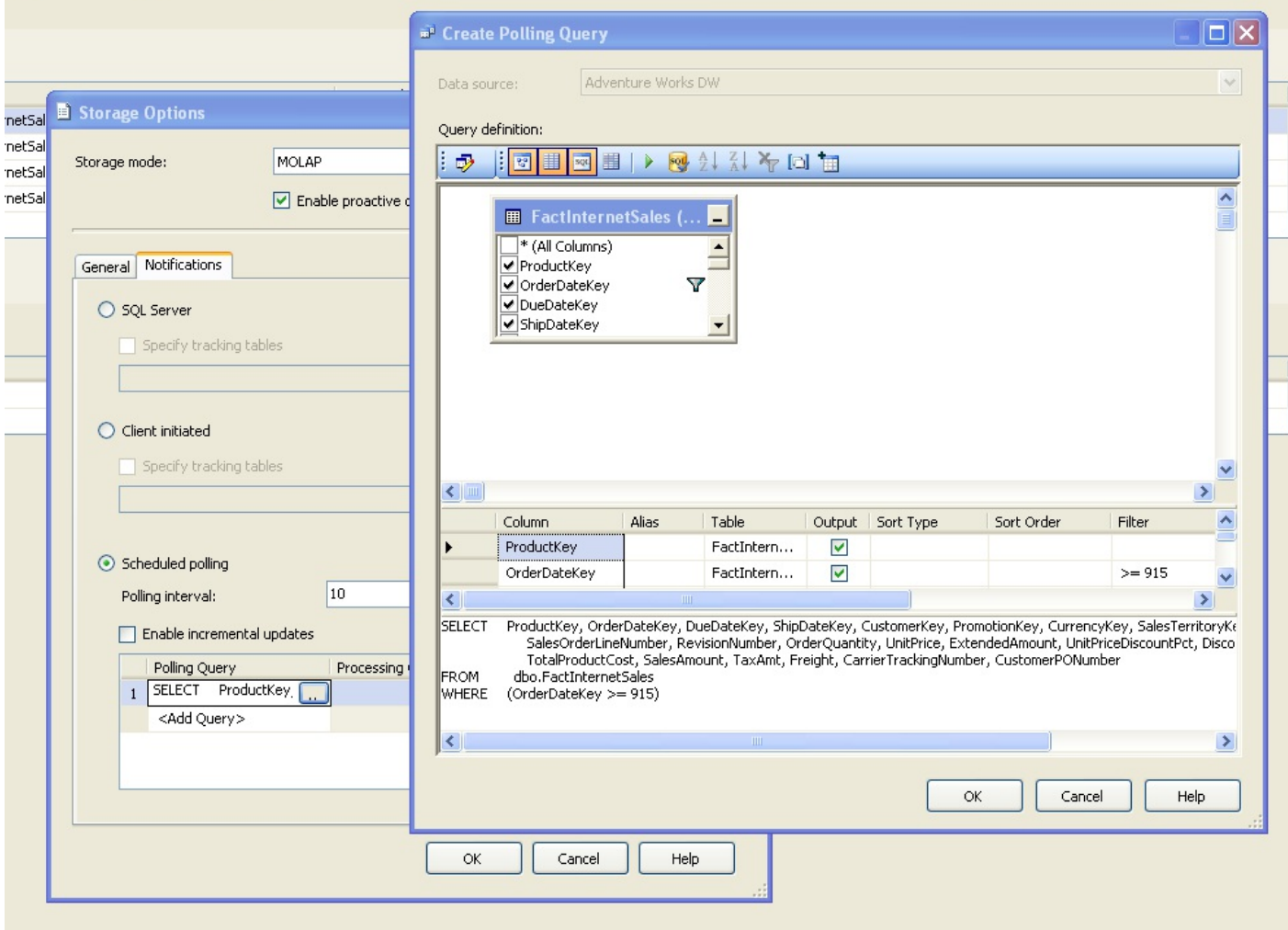
☒ Bring online immediately ☐ Apply settings to dimensions

☐ Enable ROLAP aggregations

 This partition is bound to a query, and the current notification method is SQL server notification. In order for change notifications to be received, you must either specify one or more tracking tables or set the notification method to Scheduled Polling.

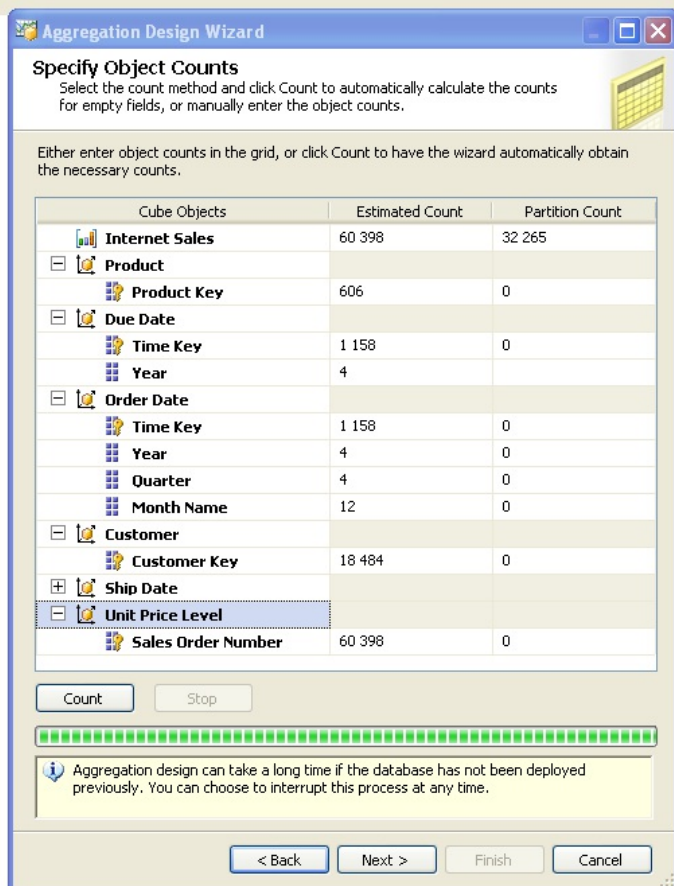
OK Cancel Help

Nieskonfigurowane szczegółowe ustawienia Medium-latency MOLAP.



Konfiguracja zapytania powiadamiającego o zmianie, po zapisaniu zapytania zapytanie rozwinęto *

Konfiguracja opcji agregacji



Specify Object Counts
Select the count method and click Count to automatically calculate the counts for empty fields, or manually enter the object counts.

Either enter object counts in the grid, or click Count to have the wizard automatically obtain the necessary counts.

Cube Objects	Estimated Count	Partition Count
Internet Sales	100 000 000	32 265
Product		
Product Key	1 000	250
Due 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
Customer		
Customer Key	20 000	5 000
Ship Date		
Unit Price Level		
Sales Order Number	60 398	0

Count Stop

< Back Next > Finish Cancel

Strojenie agregacji

Set Aggregation Options
Choose an aggregation option to optimize storage and query performance for your system.

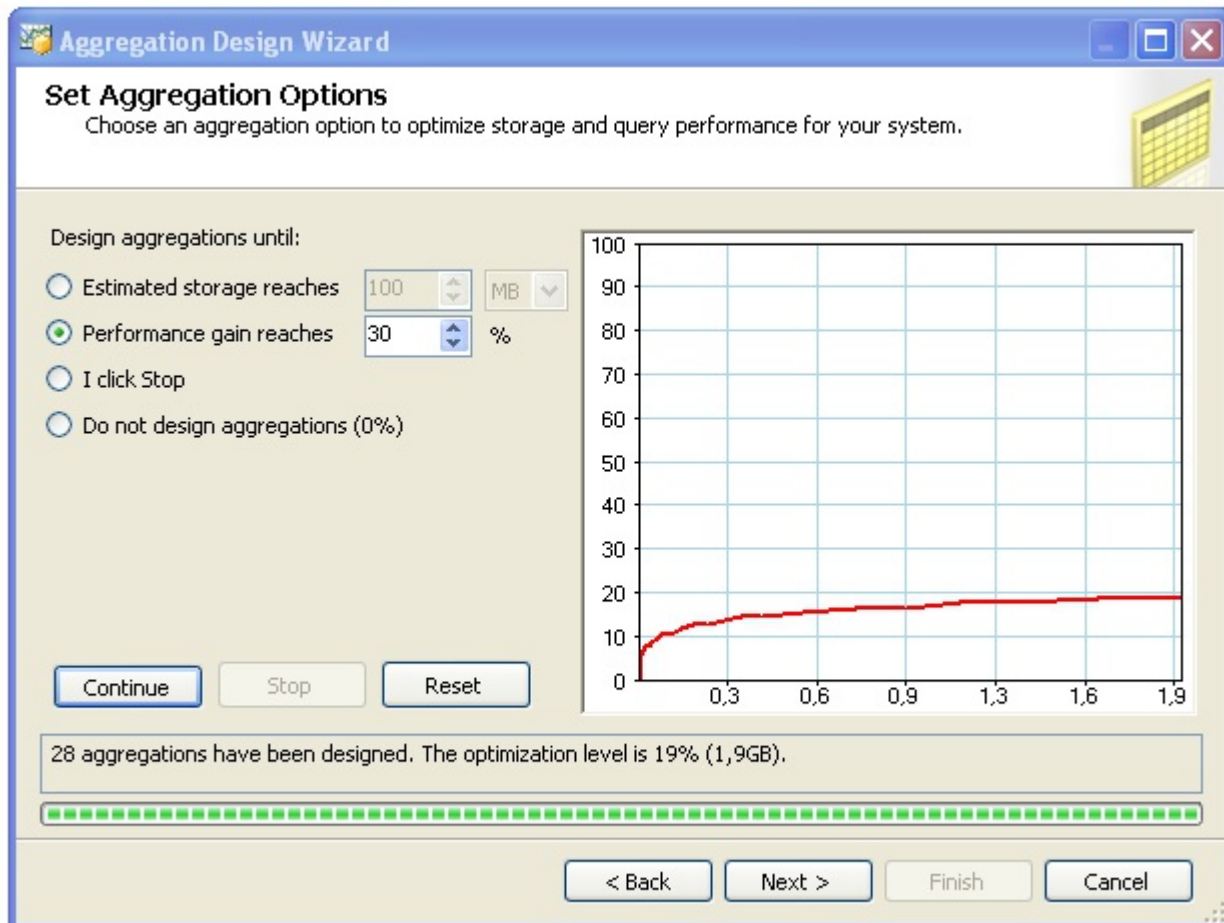
Design aggregations until:

- ☐ Estimated storage reaches 100 MB
- ☒ Performance gain reaches 50 %
- ☐ I click Stop
- ☐ Do not design aggregations (0%)

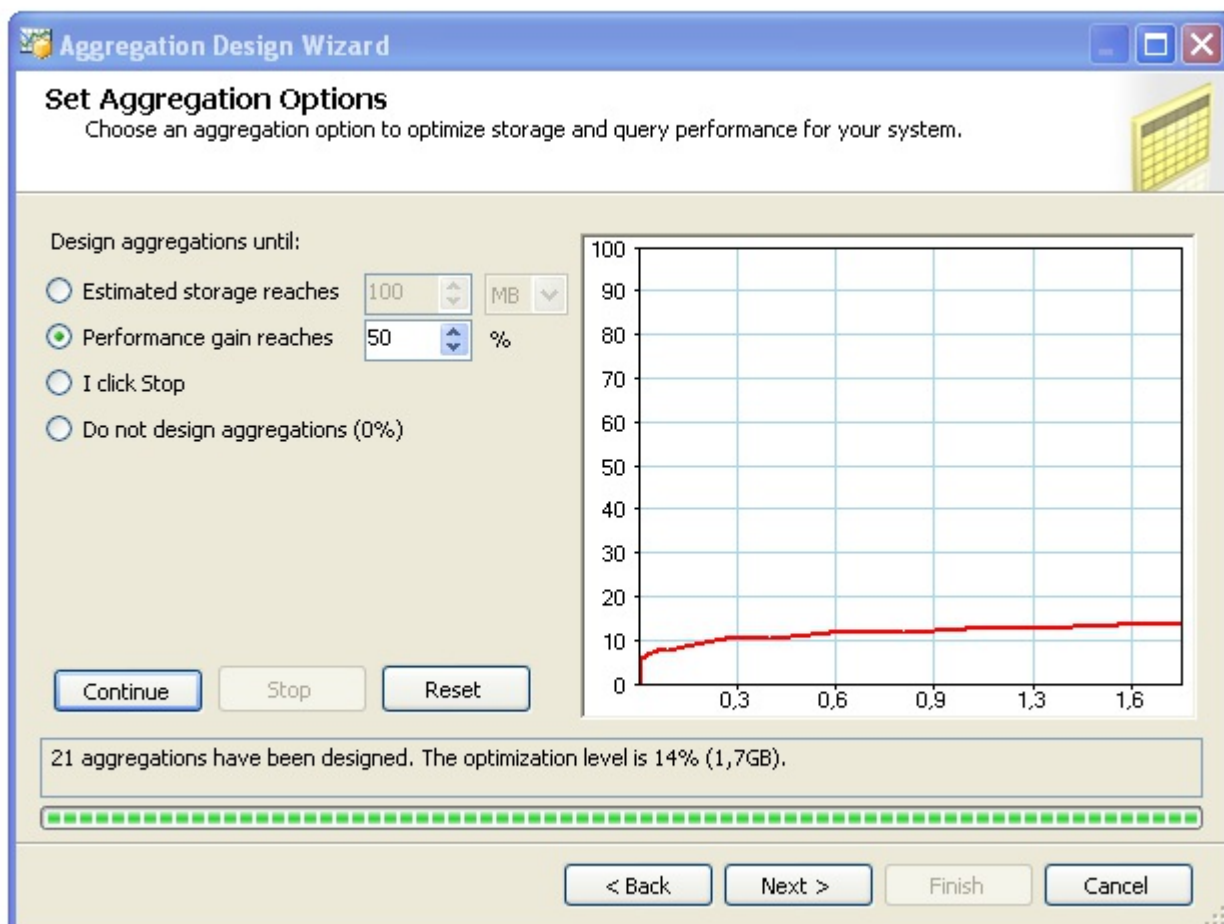
Continue Stop Reset

28 aggregations have been designed. The optimization level is 19% (1,9GB).

< Back Next > Finish Cancel

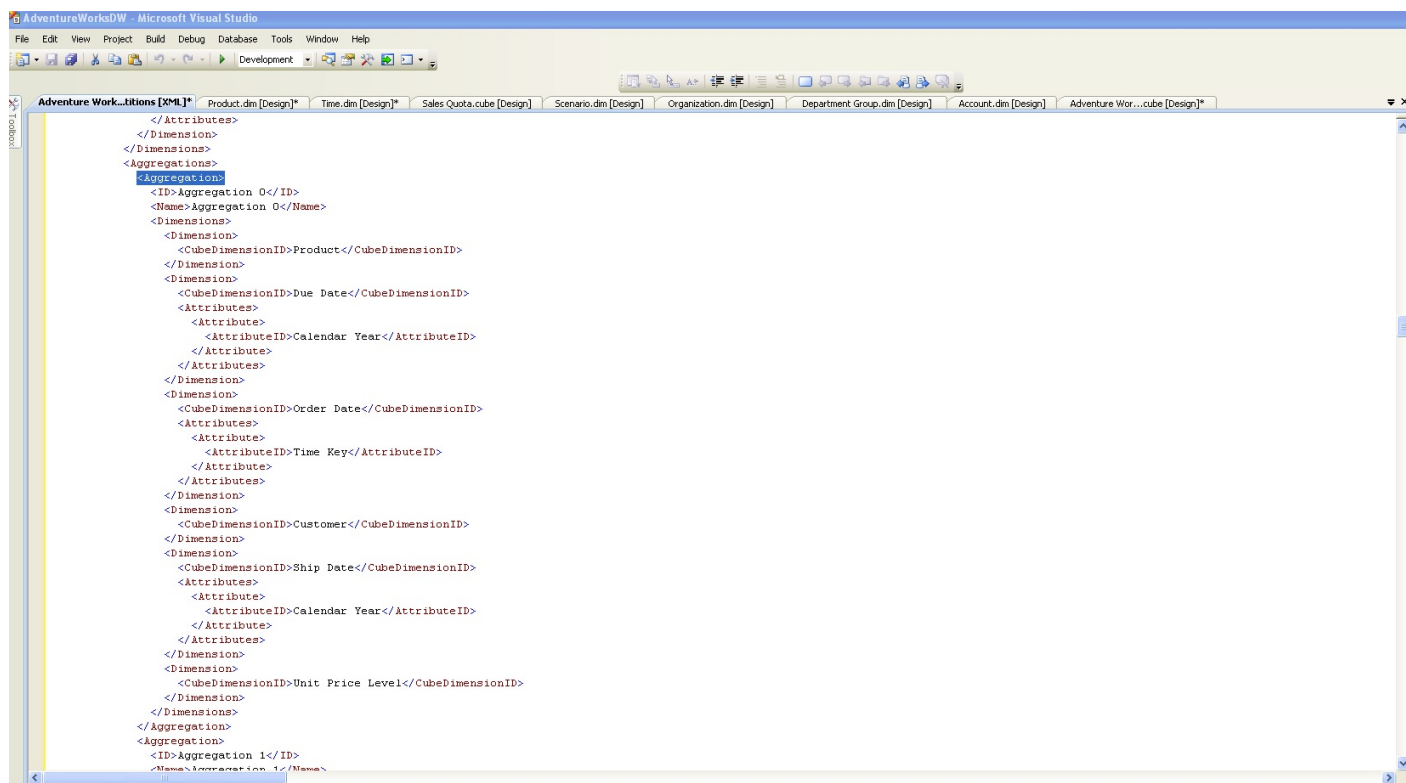


Konfiguracja agregacji - domyślne szacunki strojenia agregacji dla 30% zysku wydajności. Również zostało wykonane 28 agregacji.



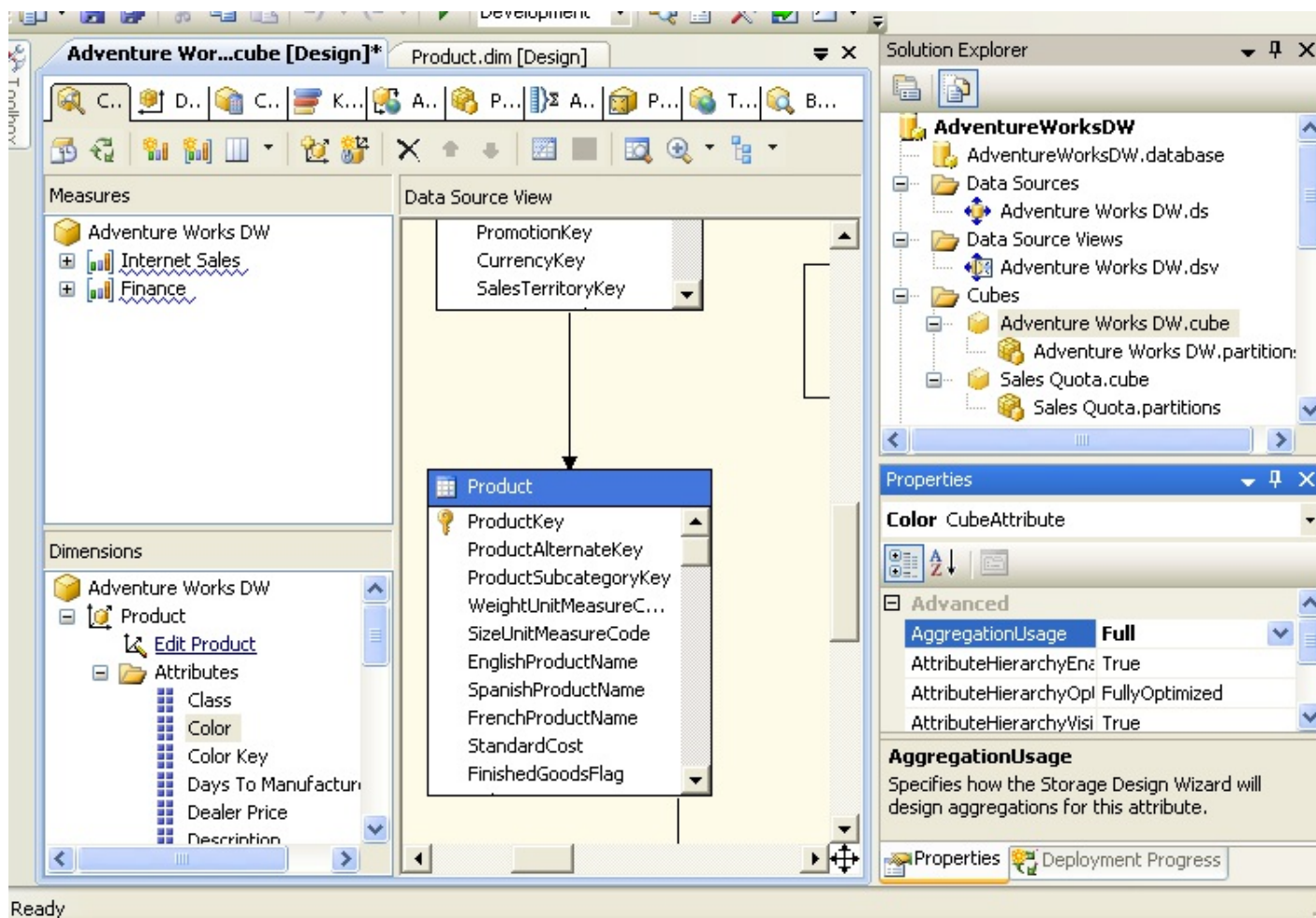
Konfiguracja agregacji - szacunki strojenia agregacji po zwiększeniu ilości Produktów. Zostało wykonane 21 agregacji.

Przegląd agregacji utworzonych w zbiorze `AdventureWorks.partition`



Definicja agregacji w kodzie źródłowym.

Zmiana agregacji dla poziomów



Ready

Zmiana metody agregacji dla atrybutu Color wymiaru Product.

Aggregation Design Wizard

Review Aggregation Usage

Aggregation usage settings can be modified to include or exclude attributes from consideration.

Cube Objects	Default	Full	None	Unrestricted
Product	15	1	0	0
Product ...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Product ...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Weight ...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Color	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Size	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Weight	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Product ...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dealer P...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
List Price	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Days To ...	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Class	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Set All to Default

< Back

Next >

Finish

Cancel

Przy tworzeniu nowej agregacji atrybut Color automatycznie ustawiony jest na full.