



Strategic Sponsor



Gold Sponsors



Silver Sponsors



Technical Partners



Academic Partners



Wyższa Szkoła Zarządzania
i Bankowości w Krakowie

Wyższe Szkoły Bankowe

Media Partners



Made in Wro
Do IT here!





Procesowanie i partycjonowanie Analysis Services od podszewki (300)

Adrian Chodkowski

Adrian.Chodkowski@outlook.com



Adrian Chodkowski

- Konsultant Business Intelligence w Jcommerce S.A
- Certyfikowany specjalista baz danych i Business Intelligence z zakresu technologii Microsoft i IBM
- SQL SERVER: 2005/2008(R2)/2012/2014
- Sharepoint 2010,2013 ☹️
- Excel 😊
- Presales, Analiza, Architektura, Development, Trening



O czym nie będziemy mówić?

- Hardware
- ROLAP
- HOLAP
- RT-OLAP
- TABULAR

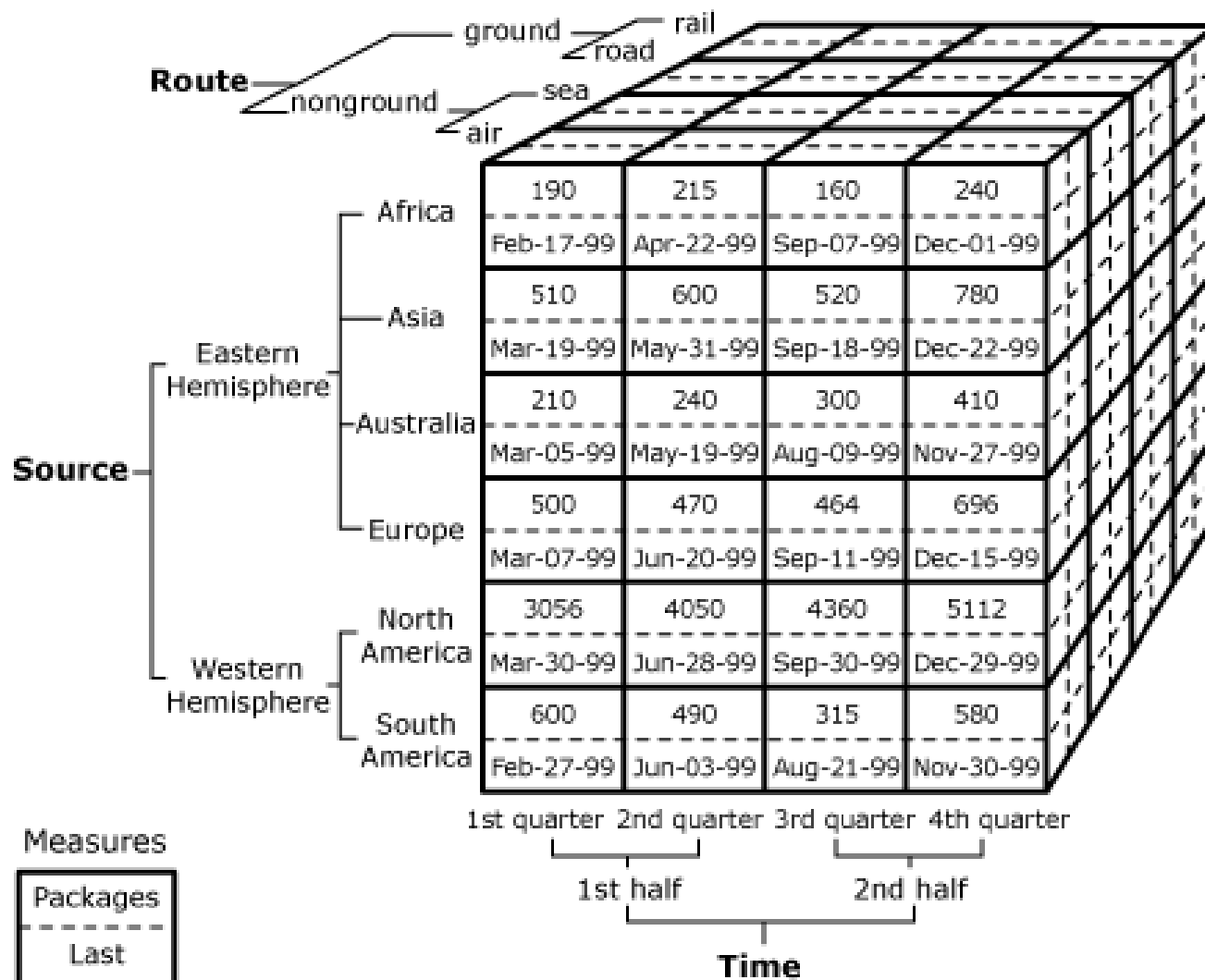


Agenda

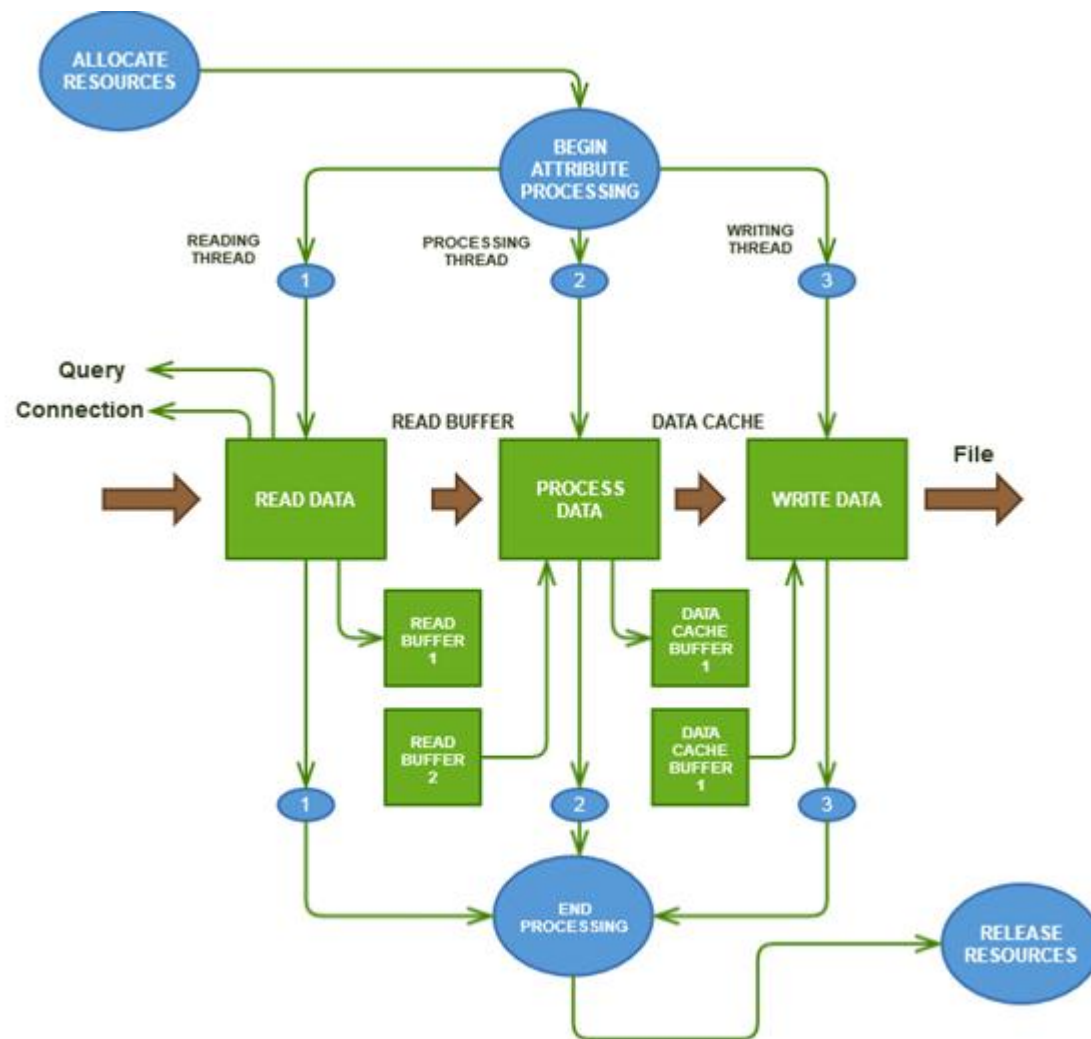
- Pojęcie kostki wielowymiarowej
- Architektura procesowania w SSAS
- Procesowanie wymiarów
- Procesowanie partycji
- Definiowanie partycji
- Lazy Processing



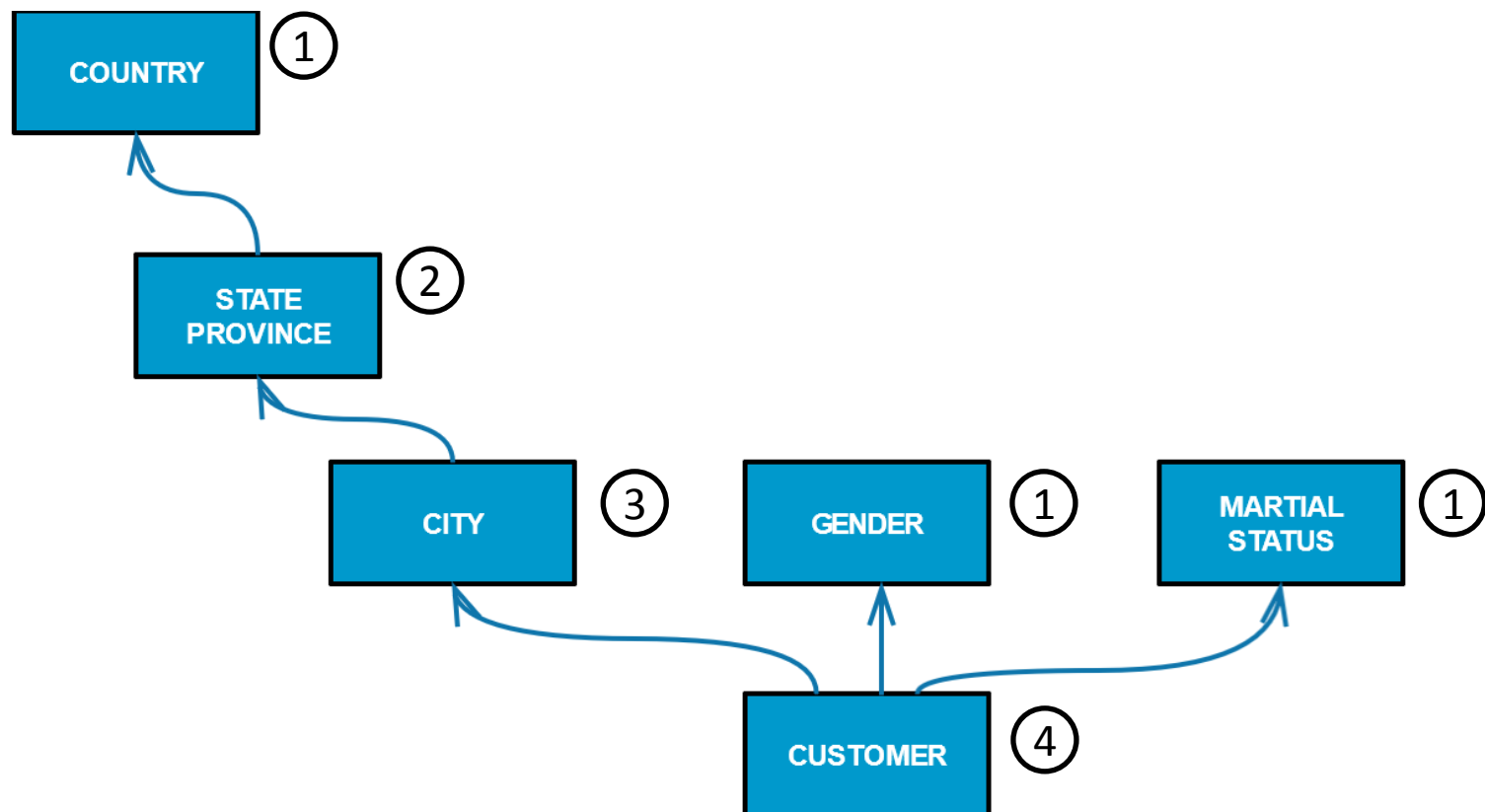
Czym jest kostka?



Architektura pobierania danych w SSAS



Procesowanie wymiarów: plan procesowania



Procesowanie - pierwsze usprawnienia

Spójrz na źródło danych!

- Oprzyj tabele w twoim DSV na widokach
- Stworzone widoki zoptymalizuj za pomocą indeksów i hintów (nolock)
- Unikaj złączeń widokach źródłowych
- Zmień domyślne ustawienia:
 - **PacketSize** domyślnie to tylko 4096 bajtów zmień na 32767 bajtów
 - Sterownik **SQL Native Client** jest zazwyczaj znacznie wolniejszy niż **Microsoft OLE DB Provider for SQL Server**
 - **Maximum number of connections** ustaw na tyle ile masz równoległych procesowań
- Wyeliminuj **konwersje typów** i inne operacje jak np. Trimming



Process Full

- **ProcessFull** – powoduje, że wszystkie połączone z wymiarem partycje muszą być partycjonowane, należy stosować w wyjątkowych przypadkach
- Czasochłonne ale akceptowalne gdy do dyspozycji jest duży interwał czasowy na procesowanie

~~Process Full~~ --> Process Data + Process Index



Process Add

- Najszybsza opcja procesowania
- Dodaje tylko nowe wiersze
- **DuplicateKey= IgnoreError**
- Możliwa implementacja Update i Delete ale wymaga specjalnego zaprojektowania hurtowni danych
- Jak to zrobić?

Quantity ▼	Status ▼
100	
20	Deleted
50	
-20	Deleting row



DEMO

Process Add



Process Update

Inteligentnie sprawdza zmiany w wymiarze, bardzo wolny proces (wolniejszy niż ProcessFull).

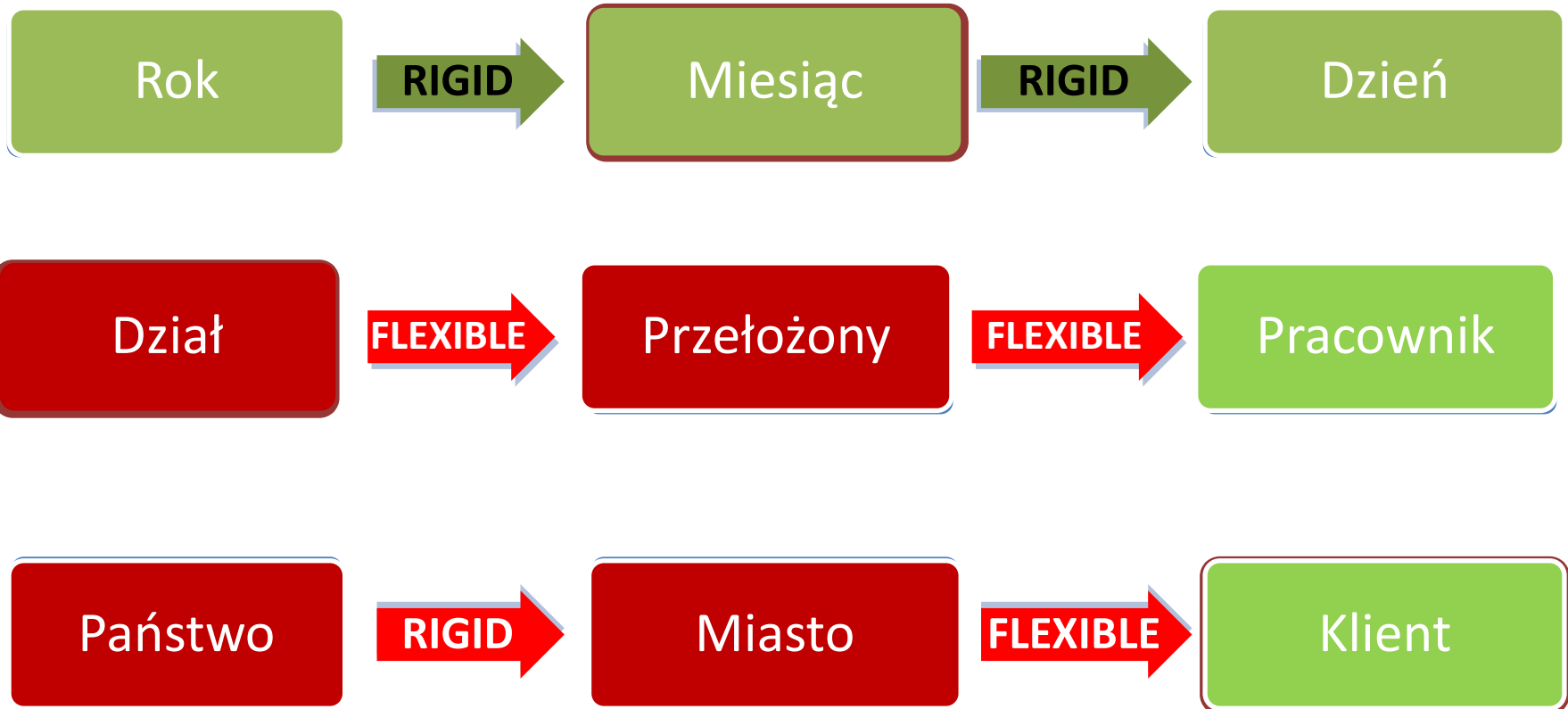
Relacje pomiędzy atrybutami przyspieszają procesowanie ale również są istotne z punktu widzenia agregacji.

Relacje pomiędzy wymiarami problemy:

- **RIGID** – błąd w przypadku wykrycia zmiany w źródle
- **FLEXIBLE** – w przypadku zmiany indeksy wymagają przeprocesowania



Process Update – wpływ na agregacje



DEMO

Process Update



Zwiększanie wydajności procesowania wymiarów

• Przechowywanie wymiarów niepotrzebnych atrybutów

- Doc
- Uży
- Atti
- Atti
- Atti
- Usu

SELECT

DISTINCT

```
[dbo_DimCustomer].[dbo_DimCustomerCustomerKey0_0] AS [dbo_DimCustomerCustomerKey0_0]
,[dbo_DimCustomer].[dbo_DimCustomerFullName0_1] AS [dbo_DimCustomerFullName0_1]
,[dbo_DimCustomer].[dbo_DimCustomerDateFirstPurchase0_2] AS [dbo_DimCustomerDateFirstPurchase0_2]
,[dbo_DimCustomer].[dbo_DimCustomerPhone0_3] AS [dbo_DimCustomerPhone0_3]
,[dbo_DimCustomer].[dbo_DimCustomerNumberCarsOwned0_4] AS [dbo_DimCustomerNumberCarsOwned0_4]
,[dbo_DimCustomer].[dbo_DimCustomerEnglishOccupation0_5] AS [dbo_DimCustomerEnglishOccupation0_5]
,[dbo_DimCustomer].[dbo_DimCustomerEnglishEducation0_6] AS [dbo_DimCustomerEnglishEducation0_6]
,[dbo_DimCustomer].[dbo_DimCustomerNumberChildrenAtHome0_7] AS [dbo_DimCustomerNumberChildrenAtHome0_7]
,[dbo_DimCustomer].[dbo_DimCustomerTotalChildren0_8] AS [dbo_DimCustomerTotalChildren0_8]
,[dbo_DimCustomer].[dbo_DimCustomerYearlyIncome0_9] AS [dbo_DimCustomerYearlyIncome0_9]
,[dbo_DimCustomer].[dbo_DimCustomerEmailAddress0_10] AS [dbo_DimCustomerEmailAddress0_10]
,[dbo_DimCustomer].[dbo_DimCustomerBirthDate0_11] AS [dbo_DimCustomerBirthDate0_11]
,[dbo_DimCustomer].[dbo_DimCustomerMaritalStatus0_12] AS [dbo_DimCustomerMaritalStatus0_12]
,[dbo_DimCustomer].[dbo_DimCustomerGender0_13] AS [dbo_DimCustomerGender0_13]
,[dbo_DimCustomer].[dbo_DimCustomerHouseOwnerFlag0_14] AS [dbo_DimCustomerHouseOwnerFlag0_14]
,[dbo_DimCustomer].[dbo_DimCustomerCommuteDistance0_15] AS [dbo_DimCustomerCommuteDistance0_15]
,[dbo_DimCustomer].[dbo_DimCustomerAddressLine10_16] AS [dbo_DimCustomerAddressLine10_16]
,[dbo_DimGeography].[PostalCode] AS [dbo_DimGeographyPostalCode0_0]
```

ized

```
SELECT
DISTINCT
[dbo_DimGeography].[PostalCode] AS [dbo_DimGeographyPostalCode0_0]
,[dbo_DimGeography].[City] AS [dbo_DimGeographyCity0_1]
,[dbo_DimGeography].[StateProvinceCode] AS [dbo_DimGeographyStateProvinceCode0_2]
FROM [dbo].[DimGeography] AS [dbo_DimGeography]
```


Wpływ sortowania na procesowanie

Data File	Sorted	Not Sorted	% Diff
fact.data	195,708,592	344,502,968	43.19%
agg.rigid.data	106,825,677	106,825,677	0.00%
dim1.dim2.fact.map	17,332,729	32,989,946	47.46%
dim1.dim3.fact.map	16,923,276	32,222,813	47.48%
dim1.dim4.fact.map	6,079,396	12,286,978	50.52%
dim5.dim6.fact.map	2,630,888	6,057,334	56.57%
dim1.dim7.fact.map	1,809,725	3,904,004	53.64%
dim8.dim9.fact.map	1,592,886	3,793,452	58.01%
dim1.dim10.fact.map	1,419,255	3,108,248	54.34%
dim8.dim11.fact.map	1,301,221	3,042,638	57.23%
dim1.dim12.fact.map	2,949,432	2,949,432	0.00%
dim1.dim13.fact.map	2,934,836	2,934,836	0.00%
dimA.dimA.fact.map	1,101,552	2,716,289	59.45%
dim8.dimB.fact.map	961,332	2,451,956	60.79%
dim1.dimC.fact.map	1,027,305	2,323,906	55.79%
dim8.dim8.fact.map	1,592,886	2,308,232	30.99%
dimA.dimD.fact.map	851,095	2,170,962	60.80%

- <http://dennyglee.com/2013/09/30/analysis-services-multidimensional-it-is-the-order-of-things/>



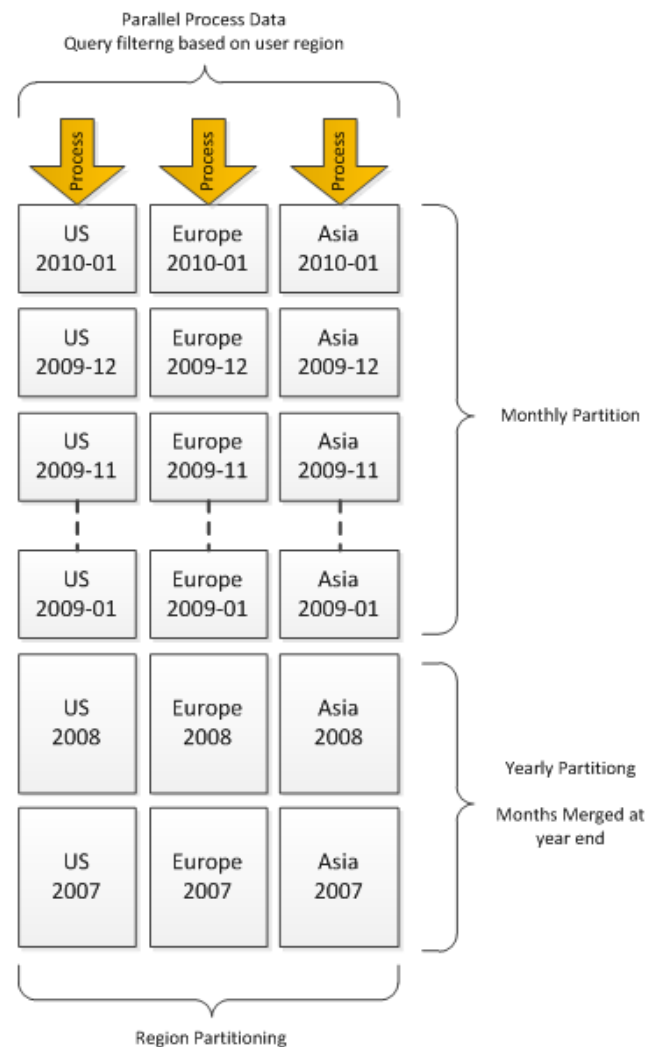
Partycje

- Dzielą dane w grupach miar na osobno przechowywane i zarządzane jednostki
- Pozwalają na to aby SSAS przeszukiwał tylko określoną część danych i szybciej zwracał dane do użytkowników
- Umożliwiają stworzenie planu zasilania danymi
- Znacząco zwiększają wydajność procesowania



Strategie partycjonowania

- Partycjonowanie po czasie
 - Partycjonowanie mieszane
- Partycje powinny zazwyczaj mieć od 2 do 20 milionów wierszy
 - Zbyt duża ich ilość jest równie szkodliwa co ich brak



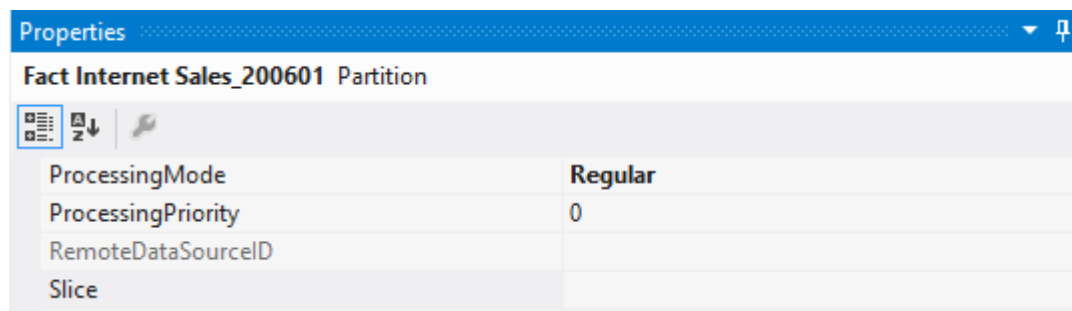
DEMO

Dynamiczne tworzenie partycji z wykorzystaniem SSIS/XMLA



Eliminowanie partycji

- W dobrze przygotowanym projekcie SSAS odpyttywana są tylko partycje zawierające istotne dla danego zapytania dane – pozostałe z nich są eliminowane
- Dobrą praktyką jest ręczne ustawienie właściwości **SLICE** w przypadku trybu MOLAP
- Ustawienie tej właściwości jest obligatoryjne w przypadku ROLAP



The screenshot shows the 'Properties' window for a partition named 'Fact Internet Sales_200601'. The window has a blue header bar with the title 'Properties' and a search icon. Below the header, the partition name is displayed. There are three icons: a list view icon (selected), a zoom in icon, and a zoom out icon. The main area contains a table with properties and their values.

Fact Internet Sales_200601 Partition	
ProcessingMode	Regular
ProcessingPriority	0
RemoteDataSourceID	
Slice	

DEMO

Slice



Zwiększanie wydajności procesowania partycji

- Używaj odpowiednich typów danych:
 - Dla kluczy - INTEGER
 - Dla miar całkowitoliczbowych – INTEGER
 - Dla wartości zmienne-przecinkowych -MONEY, REAL, FLOAT(~~VARDECIMAL~~)
- Procesuj partycje dla tej samej grupy miar równolegle
- Używaj ProcessData +ProcessIndex zamiast ProcessFull
- Dostosuj równoległe procesowanie do sprzętu jaki posiadasz
- Dobrą praktyką jest dostosowanie partycjonowania SSAS do partycji w bazie danych






Lazy Processing

- Dane dostępne są dla użytkowników zaraz po pojawieniu
- Agregacje i indeksy tworzone są w późniejszym czasie w tle
- Specjalny wątek sprawdza systematycznie, które partycje wymagają stworzenia indeksów i agregacji
- Porządek sprawdzania ustala właściwość **Processing Priority**
- Wątek sprawdza zużycie zasobów i tworzy obiekty dla jak największej ilości partycji aby zmaksymalizować zużycie zasobów
- **Brak możliwości predykcji kiedy SSAS skończy tworzenie obiektów**



Lazy Processing

- OLAP\LazyProcessing\Enabled
- OLAP\LazyProcessing\MaxCPUUsage
- OLAP\LazyProcessing\MaxObjectsInParallel
- OLAP\LazyProcessing\SleepIntervalSecs

Fact Internet Sales Partition	
  	
LastSchemaUpdate	2014-12-28 20:13
ProcessingMode	LazyAggregations
ProcessingPriority	0



DEMO

Lazy Processing





Dziękuję za uwagę!

Adrian Chodkowski

Adrian.Chodkowski@outlook.com





Strategic Sponsor



Gold Sponsors



Silver Sponsors



Technical Partners



Academic Partners



Wyższa Szkoła Zarządzania
i Bankowości w Krakowie

Wyższe Szkoły Bankowe

Media Partners



Made in Wro
Do IT here!

