

On line analytical processing (OLAP) databáze v praxi

Lukáš Matějovský

Lukas.Matejovsky@CleverDecision.com

Jan Zajíc

Jan.Zajic@CleverDecision.com

Obsah



- Představení přednášejících
- Základy OLAP
- Příklady využití v praxi
- Komponenty databáze SQL Server Analysis Services
- Základy jazyka MDX



Company profile

- IT vendor specialized in Business Intelligence / DWH / Planning & Reporting
- Founded 2006, 15+ employees, HQ in Prague
- Microsoft Gold Certified Partner for Business Intelligence solutions
- Microsoft Awards Winner for the best Business Intelligence solution
- 50+ realized BI/DWH projects
- 30+ customers in the EMEA region



Business Intelligence
Business Intelligence Platform
Data Management Solutions
ISV/Software Solutions



Jan Zajíc



- Zakladatel a jednatel Clever Decision
- 20+ let v IT
- 15+ let ve vývoji BI/DWH řešení a řízení projektů

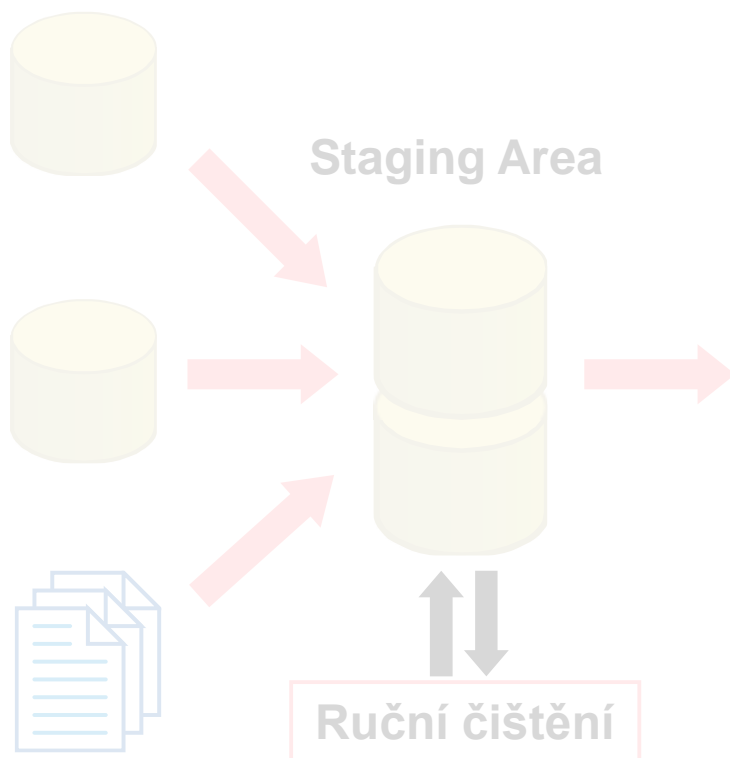
Lukáš Matějovský



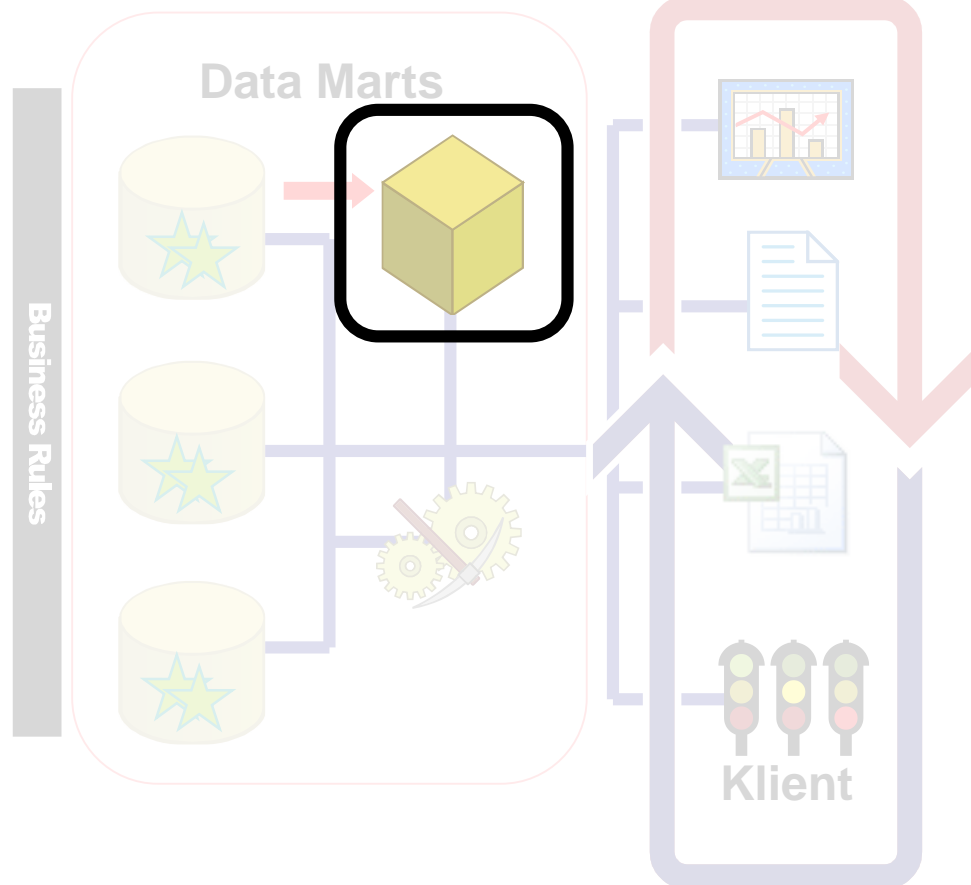
- Senior consultant
- 20+ let ve vývoji Business Intelligence a DWH řešení
- Specializace na SQL Server a datové modelování

Kde jsme?

Datové zdroje



Datový sklad



Základy OLAP



- Úvod do OLAP
- Scénáře
- SQL Server 2014 Analysis Services
- Unified Dimensional Model (UDM)
- Výhody UDM

Úvod do OLAP

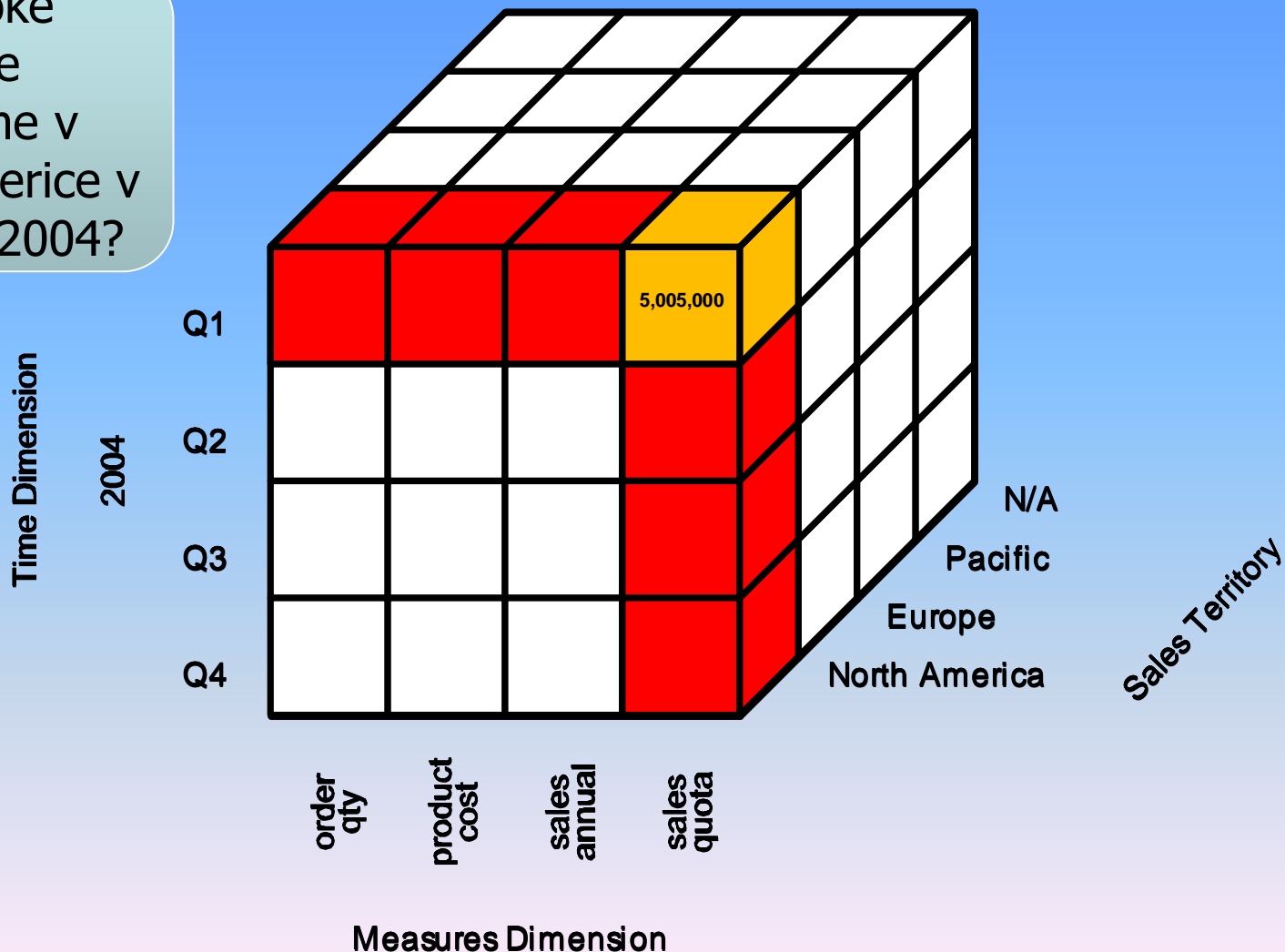


- Obsahuje struktury které umožňují rychlý ad-hoc přístup k informacím
- Jeho součástí je také výpočetní nástroj pro rychlé a flexibilní transformace základních dat
- Podporuje vyhledávání trendů a statistik, které nejsou přímo viditelné na základě dotazů z datového skladu

OLAP = **O**n**L**ine **A**nalytical **P**rocessing

Dotazy nad kostkou

Jak vysoké
prodeje
očekáváme v
Severní Americe v
1. kvartále 2004?



Scénáře



- Analýzy prodeje
- Rozpočtování, odhady budoucího vývoje, reportování
- Finanční reportování
- Analýza Webových statistik
- Vyhodnocování průzkumů
- Analýza ETL procesů

Může být použito pro jakýkoliv scénář, který používá agregovaná data a odpovědi na dotaz mají být rychlé pro různé úrovně detailu

Ukázka využití



Jeden ze scénářů využití -> Excel jako OLAP klient

SQL Server 2014 Analysis Services



- OLAP



- Uspořádává a agreguje data z datových zdrojů
- Provádí výpočty, které je obtížné provést v rámci dotazu nad relační databází
- Podporuje rozšířené funkcionality BI, např. Klíčové ukazatele výkonu (KPI)

- Data mining

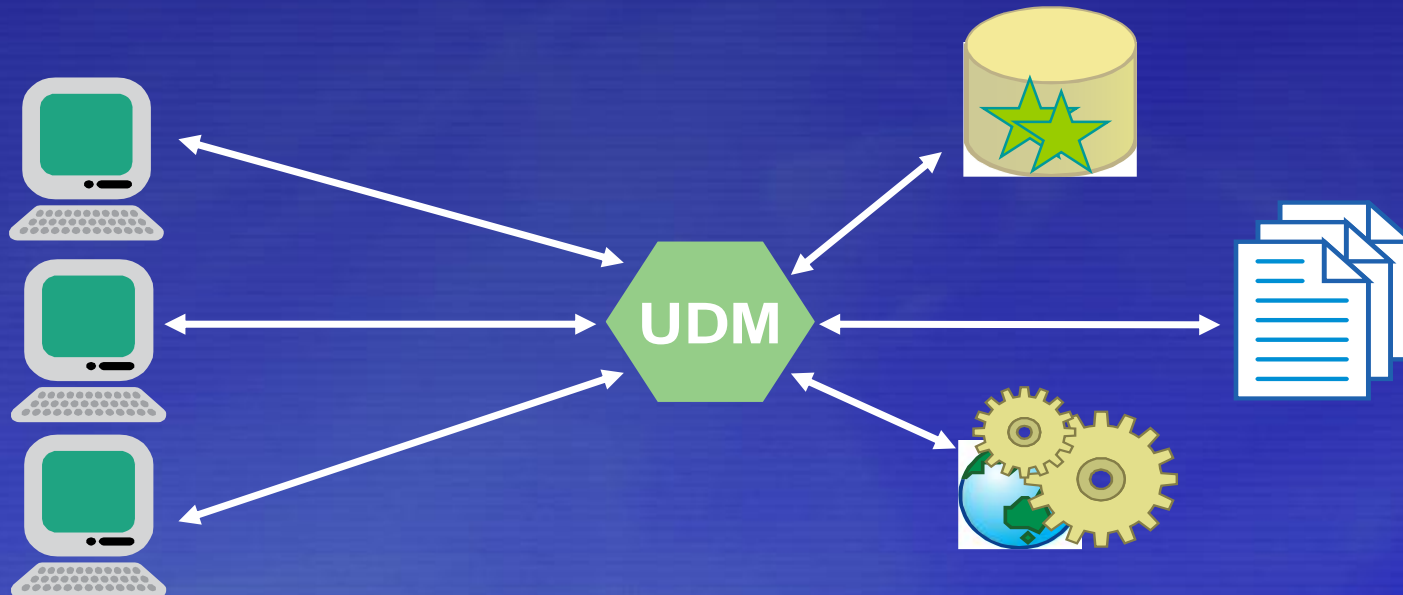


- Vyhledává vzory jak v relačních tak OLAP datech
- O jeho výsledky lze rozšířit OLAP databázi

Unified Dimensional Model (UDM)



- Představuje most mezi koncovým uživatelem a zdrojovými daty
- Konsoliduje obchodní pravidla do jediného modelu
- Je lepší alternativou k tradičním modelům
- V SSAS je tento model nazýván Kostka (Cube)



Výhody UDM



- Umožňuje rozsáhlé rozšíření uživatelského modelu
- Umožňuje velmi výkonné získávání informací, rychlé interaktivní analýzy i nad velkým objemem dat
- Díky zahrnutí obchodních pravidel umožňuje širší analýzy
- Výsledný model abstrahuje od použitých datových zdrojů

Komponenty databáze



- Data Source
- Data Source View
- Cube
- Dimensions
- Measures
- Calculations
- Key Performance Indicators
- Actions
- Perspectives
- Translations

Data Source

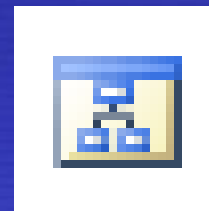
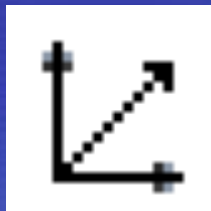


- Obsahuje informace o připojení k datovému zdroji
 - OLE DB Provider
 - Managed .NET Provider
- Je podporován SQL Server i další populární databáze

Data Source View



- Popisuje definici jednotlivých prvků v datovém zdroji
- Pro klientské aplikace je neviditelný
- Výhody:
 - Zaměřuje se pouze na podmnožinu tabulek z datových zdrojů
 - Umožňuje integrování více datových zdrojů
 - Vyžaduje pouze read-only přístup ke zdrojovým datům
 - Lze vyvíjet databázi i bez nutnosti připojení k datovým zdrojům
 - Izoluje databázové objekty od změn provedených v datovém zdroji:



- Kombinace dimenzí a ukazatelů v jednotném koncepčním modelu
 - Ukazatelé vycházejí z faktových tabulek
 - Dimenze vycházejí z dimenzních tabulek
- Bohatý datový model je rozšířen o:
 - Kalkulace
 - Klíčové ukazatele výkonu (KPIs)
 - Akce
 - Perspektivy
 - Překlady
 - Partice

Dimenze



- Jednotlivé atributy jsou vytvořeny na základě jedné či více tabulek nebo pohledů z data source view
- Atributy odpovídají sloupcům dimenzní tabulky
- Atributy jsou organizovány do hierarchií
 - Hierarchie Atributu obsahují jednu úroveň All a jednu úroveň typu List
 - Uživatelské hierarchie jsou víceúrovňové kombinace atributů
 - Mohou být zobrazeny ve složkách
- Přirozená hierarchie

Typy a charakteristiky dimenzí



- Time (Server time dimension)
 - Obsah dimenze je založen na zvoleném období nikoliv na obsahu dimenzní tabulky
 - Je užitečná, pokud je kostka založena na OLTP databázi
- Speciální: Account, Currency
 - Řídí způsob agregace účtů za různé časové období
 - Do kostky přidá schopnost provádět přepočty směnných kurzů
- Charakteristiky:
 - Parent-child
 - Write-enabled
 - Linked

Vztahy dimenzí



- Definuje vztah mezi dimenzemi a skupinami ukazatelů
- Vztahy mohou být:
 - Regular
 - Reference
 - Fact (Degenerate)
 - Many-to-many
 - Data mining

Measures



- Ukazatelé vycházejí z faktových tabulek
- Ukazatel je číselná hodnota, kterou lze agregovat
 - Sum
 - Count
 - Avg
 - Distinct count
 - Min, Max
- Ukazatelé z jedné faktové tabulky jsou zastřešeny zpravidla jednou skupinou ukazatelů
- Skupina ukazatelů je propojena s jednotlivými dimenzemi (viz. vztahy dimenzí)

Základy MDX



- MDX je dotazovací jazyk na OLAP kostkou
- Pracuje s až 128 osami
 - COLUMNS (0)
 - ROWS (1)
 - WHERE (2)
 - PAGE (3)
- TUPLE - ()
- SET - {}

MDX = **M**ulti**D**imensional **E**xpressions

Základní MDX dotaz



- Příklady MDX dotazů:

```
SELECT  
FROM [CDT]
```

```
SELECT  
NON EMPTY {[Measures].[Work Task Count]} ON COLUMNS,  
NON EMPTY {[Date].[Month].MEMBERS} ON ROWS  
FROM [CDT]
```

```
WITH MEMBER [Measures].[Calc] AS  
    [Measures].[Work Task Count] * 1.1  
SELECT  
NON EMPTY {[Measures].[Calc]} ON COLUMNS,  
NON EMPTY {[Date].[Month].MEMBERS} ON ROWS  
FROM [CDT]
```


Tuple



- Jedná se o referenci na konkrétní adresu v kostce
- Reference se skládá z jednoho ukazatele a jednoho člena z jedné či více dimenzí
- Tuple je v MDX dotazu syntakticky uveden v kulatých závorkách
- Závorky nemusí být uvedeny v případě jednoduchého tuple, který se skládá z jednoho člena jedné dimenze

```
([Date].[Month].&[2014][10], [Measures].[Work Task Count])
```

Set



- Jedná se o sadu žádného, jednoho či více tuplů
- Nejčastěji se používají na jednotlivých osách MDX dotazu
- V MDX je set uveden ve složených závorkách

```
{[Date].[Month].&[2014][10] : [Date].[Month].&[2014][12]}
```

```
{[Date].[Month].&[2014][10] , [Date].[Month].&[2014][12]}
```

- Jedná se o výrazy vyhodnocované až při zpracování dotazu jež nelze uložit ve faktové tabulce
- Typy výpočtu:
 - Kalkulované atributy
 - Pojmenované sady
 - Přiřazení rozsahu (Scope)
- Výpočty jsou definovány v jazyce MDX

MDX = **M**ulti**D**imensional **E**Xpressions

Klíčové ukazatele výkonu (KPIs)



- Kvantifikovatelné ukazatele porovnávající skutečný výkon s definovanými cíly
- Pokud jsou zobrazeny v souhrnu výsledků, ukazují celkové zdraví společnosti

Define KPI



- Value
- Goal
- Status
- Trend

	Actual	Goal and Status	Trend
<input type="checkbox"/> Sales Targets			
<input type="checkbox"/> Sales Target			
Pak, Jae	4172459	4000000 	
Reiter, Tsvi	2222128	2225000 	
Saraiva, José	1837614	1800000 	
Tsoflias, Lynn	701487	545000 	
Valdez-Smythe, Rachel	961128	1100000 	
Vargas, Garrett	1378290	1450000 	

Actions



- Akce je uživatelem vyvolaná operace na základě vybrané kostky či její části
- Akce může být různého typu:
 - Drillthrough
 - Report
 - Zavolání Url
- Akce může být definována nad různými objekty v kostce:
 - Attribute member
 - Cell
 - Cube
 - Dimension member
 - Hierarchy
 - Hierarchy member
 - Level
 - Level member

Perspectives



- Perspektiva je podmnožina kostky
- Kostka je implicitní perspektivou
- Perspektiva je uživateli zobrazována jako další kostka

Q & A



Děkujeme za pozornost :o)

Prostor pro Vaše dotazy...

...a pokud Vás téma zaujalo, kontaktujte nás.

Hledáme šikovné spolupracovníky do teamu (i na part-time).

Kontakt: informace@CleverDecision.com