

13

Vysoká škola báňská – Technická univerzita Ostrava  
Fakulta strojní, Katedra automatizační techniky a řízení

# Informační systémy



2006/2007

Ivan Kedroň

1

---

---

---

---

---

---

---

---

## Obsah

- Analytické nástroje SQL serveru.
- OLAP analýza údajů v databázi.
- Data Mining.

- Doporučená literatura:  
Lacko, L. Analytické možnosti produktu  
Microsoft SQL Server 2000. Praha :  
Microsoft, s. r. o. 74 s.



Informační systémy 2006/2007

2

---

---

---

---

---

---

---

---

## Analytické možnosti MS SQL Server

- MS SQL Server
  - Relační databázový systém
- MS Data transformation Services (ETL)
  - Nástroj pro extrakci, transformaci a loading
- MS OLAP Server
  - Systém pro On-line Analytical Processing
- MS Data Mining services
  - Nástroj pro Data Mining – tedy dolování dat



Informační systémy 2006/2007

3

---

---

---

---

---

---

---

---

## Nástroje pro ETL

- Proč transformovat data
  - Data pochází z několika různých nehomogenních zdrojů
  - Data jsou v lepším případě pouze zpracovány do sestav
  - Do skladu (Data Warehouse) je třeba uložit data jednoho typu, aby bylo možno je zpracovávat pro analýzu



---

---

---

---

---

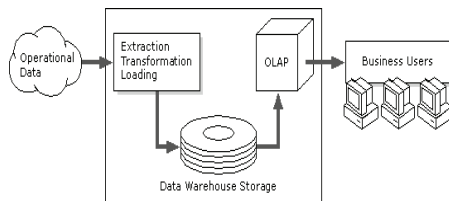
---

---

---

## Nástroje pro ETL

- Schéma datového skladu



---

---

---

---

---

---

---

---

## Nástroje pro ETL

- Etapy procesu ETL
  - Extrakce
    - Výběr dat prostřednictvím určitých metod
  - Transformace
    - Ověření, čištění, integrování a časové označení dat
  - Loading
    - Transport (uložení) dat do datového skladu



---

---

---

---

---

---

---

---

## Nástroje pro ETL

- V MS SQL Server zajišťuje ETL Data Transformation Services pomocí služeb importu dat, který umí pracovat s daty z velkého množství různých zdrojů
- Import probíhá na základě průvodce a definic v něm zadaných nebo za pomoci workflow diagramů procesu ETL



---

---

---

---

---

---

---

---

## Příklad použití DTS

Jako příklad využijeme databázi FoodMart, která je součástí instalace SQL Server. Tato databáze je vytvořena v MS Access a proto je nutno data převést pomocí ETL do formy vhodnější pro zpracování.

Postup:

- Vytvoříme databázi do které budeme importovat data pomocí např. pomocí Enterprise Manageru
- Zadáme import dat, čímž spustíme DTS Import Wizard



---

---

---

---

---

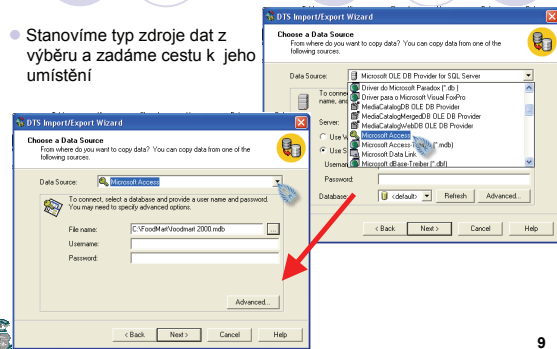
---

---

---

## Příklad použití DTS

- Stanovíme typ zdroje dat z výběru a zadáme cestu k jeho umístění



---

---

---

---

---

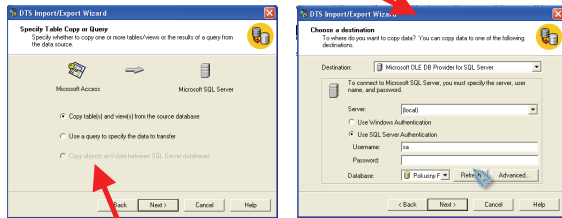
---

---

---

## Příklad použití DTS

- Určíme cílovou databázi a přístup k ní



- Následně zvolíme způsob tvorby tabulek



Informační systémy 2006/2007

10

---

---

---

---

---

---

---

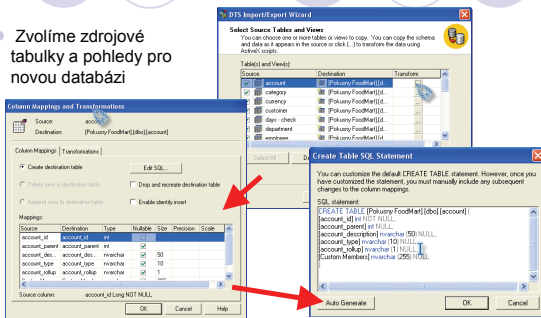
---

---

---

## Příklad použití DTS

- Zvolíme zdrojové tabulky a pohledy pro novou databázi



Informační systémy 2006/2007

11

---

---

---

---

---

---

---

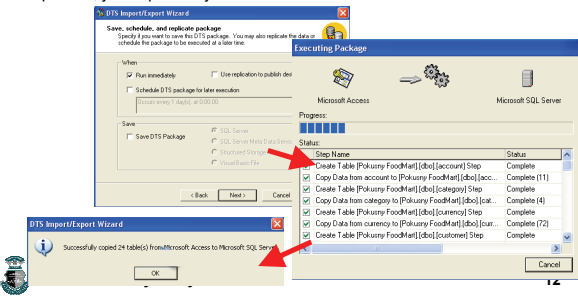
---

---

---

## Příklad použití DTS

- Nakonec ještě nastavíme čas provedení a spustíme samotný převod, jehož průběh je zobrazen a ukončení ohlášeno



Informační systémy 2006/2007

12

---

---

---

---

---

---

---

---

---

---

## On-line Analytical Processing

- Pro analýzu velkého množství údajů
- Výsledkem analýzy jsou souhrny a reporty sloužící v procesu řízení
- Je třeba velkého množství výpočtů a agregací v 'téměř reálném čase'
- Jedná se o definovanou řadu principů, které poskytují dimenzionální rámec pro podporu rozhodování



---

---

---

---

---

---

---

---

## On-line Analytical Processing

- Dvanáct pravidel OLAP
  1. Multidimenzionální konceptuální pohled
  2. Transparentnost
  3. Dostupnost
  4. Konzistentní vykazování
  5. Architektura klient-server
  6. Generická dimenzionalita
  7. Dynamické ošetření řídkých matic
  8. Podpora více uživatelů
  9. Neomezené křížové dimenzionální operace
  10. Intuitivní manipulace s údaji
  11. Flexibilní vykazování
  12. Neomezené dimenze a úrovně agregace



---

---

---

---

---

---

---

---

## On-line Analytical Processing

- Multidimenzionální databázový model je tzv. OLAP kostka (tři dimenze jsou pouze ilustrativní, obvykle je jich mnohem více)



---

---

---

---

---

---

---

---

## On-line Analytical Processing

- Tabulka Faktů
  - Fakty jsou numerické měrné jednotky obchodování
  - Tabulka je největší a má hvězdicové schéma
- Tabulky dimenzí
  - Dimenze jsou textové popisy obchodování
  - Obsahují logicky nebo organizačně hierarchicky uspořádané údaje
  - Obvykle mají stromovou strukturu



---

---

---

---

---

---

---

---

## On-line Analytical Processing

- Postup tvorby kostky v MS SQL Server
  - Připojení ke zdroji dat
  - Volba tabulky faktů
  - Vytvoření dimenzí
  - Konečné vytvoření kostky
  - Výpočet kostky (volba úložiště)
  - Zpracování výsledků  
(výsledky je možno zobrazit např v MS Excel jako kontingenční tabulku)



---

---

---

---

---

---

---

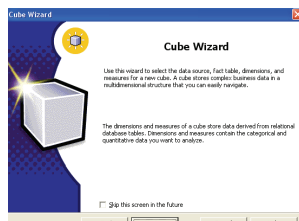
---

## Příklad OLAP analýzy

Jako příklad využijeme opět databázi FoodMart, tentokrát je již převedenou do MS SQL Server 2000 v předchozím příkladě. Budeme pracovat s Analysis Managerem – nástrojem analytických služeb

Postup:

- Založíme novou databázi tentokrát v Analysis Manageru
- V záložce Cubes dáme vytvořit novou kostku pomocí průvodce



---

---

---

---

---

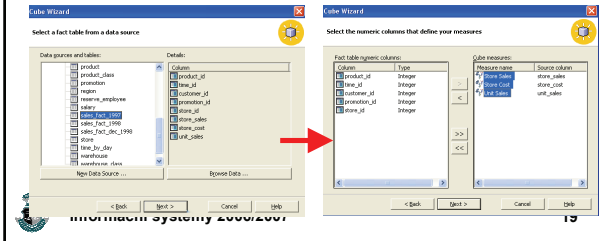
---

---

---

## Příklad OLAP analýzy

- Zvolíme tabulku faktů, tedy data se kterými chceme pracovat - v našem případě to může být např. tabulka sales\_fact\_1997
- Z ní pak vybereme sloupce které budou měřnými jednotkami pro analýzu (Store\_Sales, Store\_Cost, Unit Sales)




---

---

---

---

---

---

---

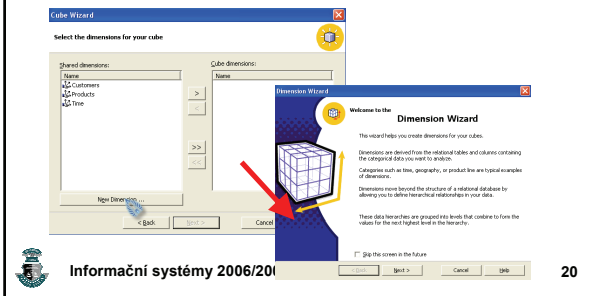
---

---

---

## Příklad OLAP analýzy

- Dále se dostaneme k vytvoření jednotlivých dimenzí, k čemuž slouží další průvodce




---

---

---

---

---

---

---

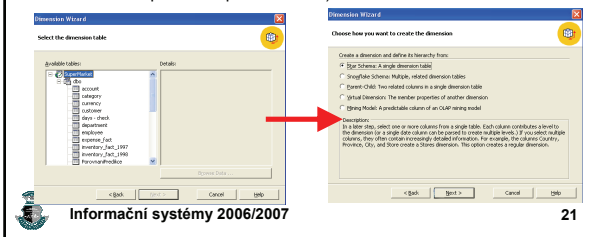
---

---

---

## Příklad OLAP analýzy

- V tomto průvodci zvolíme schéma z nabídky (Star Schema)
- Zvolíme tabulku, případně tabulky ze kterých se budou dimenze tvořit (pro náš případ budou tři dimenze – jedna z tabulky customer, druhá z time\_by\_day a třetí ze spojení tabulek product a product class)




---

---

---

---

---

---

---

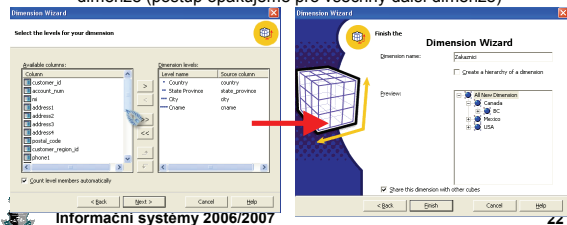
---

---

---

## Příklad OLAP analýzy

- Následně z tabulky vybereme jednotlivé sloupce tak, aby tvořily strukturu dimenzí (na obr. Je vidět zvolené sloupce pro dimenzi Customers)
- Zadáme název, podíváme se na náhled a dokončíme tvorbu dimenze (postup opakujeme pro všechny další dimenze)



Informační systémy 2006/2007

---

---

---

---

---

---

---

---

---

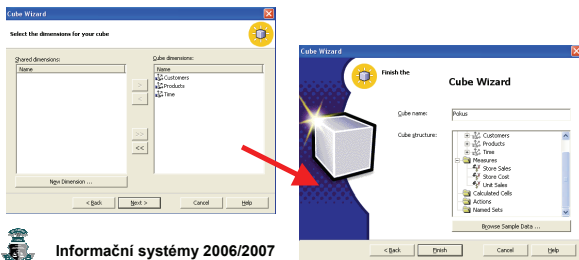
---

---

---

## Příklad OLAP analýzy

- Jakmile máme vytvořeny všechny dimenze, vybereme je v průvodci tvůrce kostky a dokončíme tvorbu kostky zadáním názvu



Informační systémy 2006/2007

---

---

---

---

---

---

---

---

---

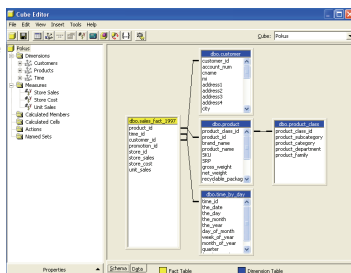
---

---

---

## Příklad OLAP analýzy

- Tím se dostaneme do Cube Editoru, kde si můžeme prohlédnout schéma, provést kontrolu a případné úpravy
- Pokud je vše v pořádku, pak zvolíme Process Cube čímž dojde k výpočtu kostky



Informační systémy 2006/2007

24

---

---

---

---

---

---

---

---

---

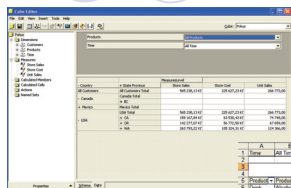
---

---

---



## Příklad OLAP analýzy



Po ukončení výpočtu si můžeme v záložce data prohlédnout výsledek analýzy s možností libovolného vnoření

Data můžeme také přenést do

MS Access a

zobrazit pomocí

kontingenční tabulky

Informační systémy 2006/2007

Time	Sub-Product	Product	State	Sum Sales
2000	Sub-Product 1	Product 1	State 1	100,000
2000	Sub-Product 1	Product 1	State 2	150,000
2000	Sub-Product 1	Product 1	State 3	200,000
2000	Sub-Product 1	Product 1	State 4	250,000
2000	Sub-Product 1	Product 1	State 5	300,000
2000	Sub-Product 1	Product 1	State 6	350,000
2000	Sub-Product 1	Product 1	State 7	400,000
2000	Sub-Product 1	Product 1	State 8	450,000
2000	Sub-Product 1	Product 1	State 9	500,000
2000	Sub-Product 1	Product 1	State 10	550,000
2000	Sub-Product 1	Product 1	State 11	600,000
2000	Sub-Product 1	Product 1	State 12	650,000
2000	Sub-Product 1	Product 1	State 13	700,000
2000	Sub-Product 1	Product 1	State 14	750,000
2000	Sub-Product 1	Product 1	State 15	800,000
2000	Sub-Product 1	Product 1	State 16	850,000
2000	Sub-Product 1	Product 1	State 17	900,000
2000	Sub-Product 1	Product 1	State 18	950,000
2000	Sub-Product 1	Product 1	State 19	1,000,000
2000	Sub-Product 1	Product 1	State 20	1,050,000
2000	Sub-Product 1	Product 1	State 21	1,100,000
2000	Sub-Product 1	Product 1	State 22	1,150,000
2000	Sub-Product 1	Product 1	State 23	1,200,000
2000	Sub-Product 1	Product 1	State 24	1,250,000
2000	Sub-Product 1	Product 1	State 25	1,300,000
2000	Sub-Product 1	Product 1	State 26	1,350,000
2000	Sub-Product 1	Product 1	State 27	1,400,000
2000	Sub-Product 1	Product 1	State 28	1,450,000
2000	Sub-Product 1	Product 1	State 29	1,500,000
2000	Sub-Product 1	Product 1	State 30	1,550,000
2000	Sub-Product 1	Product 1	State 31	1,600,000
2000	Sub-Product 1	Product 1	State 32	1,650,000
2000	Sub-Product 1	Product 1	State 33	1,700,000
2000	Sub-Product 1	Product 1	State 34	1,750,000
2000	Sub-Product 1	Product 1	State 35	1,800,000
2000	Sub-Product 1	Product 1	State 36	1,850,000
2000	Sub-Product 1	Product 1	State 37	1,900,000
2000	Sub-Product 1	Product 1	State 38	1,950,000
2000	Sub-Product 1	Product 1	State 39	2,000,000
2000	Sub-Product 1	Product 1	State 40	2,050,000
2000	Sub-Product 1	Product 1	State 41	2,100,000
2000	Sub-Product 1	Product 1	State 42	2,150,000
2000	Sub-Product 1	Product 1	State 43	2,200,000
2000	Sub-Product 1	Product 1	State 44	2,250,000
2000	Sub-Product 1	Product 1	State 45	2,300,000
2000	Sub-Product 1	Product 1	State 46	2,350,000
2000	Sub-Product 1	Product 1	State 47	2,400,000
2000	Sub-Product 1	Product 1	State 48	2,450,000
2000	Sub-Product 1	Product 1	State 49	2,500,000
2000	Sub-Product 1	Product 1	State 50	2,550,000
2000	Sub-Product 1	Product 1	State 51	2,600,000
2000	Sub-Product 1	Product 1	State 52	2,650,000
2000	Sub-Product 1	Product 1	State 53	2,700,000
2000	Sub-Product 1	Product 1	State 54	2,750,000
2000	Sub-Product 1	Product 1	State 55	2,800,000
2000	Sub-Product 1	Product 1	State 56	2,850,000
2000	Sub-Product 1	Product 1	State 57	2,900,000
2000	Sub-Product 1	Product 1	State 58	2,950,000
2000	Sub-Product 1	Product 1	State 59	3,000,000
2000	Sub-Product 1	Product 1	State 60	3,050,000
2000	Sub-Product 1	Product 1	State 61	3,100,000
2000	Sub-Product 1	Product 1	State 62	3,150,000
2000	Sub-Product 1	Product 1	State 63	3,200,000
2000	Sub-Product 1	Product 1	State 64	3,250,000
2000	Sub-Product 1	Product 1	State 65	3,300,000
2000	Sub-Product 1	Product 1	State 66	3,350,000
2000	Sub-Product 1	Product 1	State 67	3,400,000
2000	Sub-Product 1	Product 1	State 68	3,450,000
2000	Sub-Product 1	Product 1	State 69	3,500,000
2000	Sub-Product 1	Product 1	State 70	3,550,000
2000	Sub-Product 1	Product 1	State 71	3,600,000
2000	Sub-Product 1	Product 1	State 72	3,650,000
2000	Sub-Product 1	Product 1	State 73	3,700,000
2000	Sub-Product 1	Product 1	State 74	3,750,000
2000	Sub-Product 1	Product 1	State 75	3,800,000
2000	Sub-Product 1	Product 1	State 76	3,850,000
2000	Sub-Product 1	Product 1	State 77	3,900,000
2000	Sub-Product 1	Product 1	State 78	3,950,000
2000	Sub-Product 1	Product 1	State 79	4,000,000
2000	Sub-Product 1	Product 1	State 80	4,050,000
2000	Sub-Product 1	Product 1	State 81	4,100,000
2000	Sub-Product 1	Product 1	State 82	4,150,000
2000	Sub-Product 1	Product 1	State 83	4,200,000
2000	Sub-Product 1	Product 1	State 84	4,250,000
2000	Sub-Product 1	Product 1	State 85	4,300,000
2000	Sub-Product 1	Product 1	State 86	4,350,000
2000	Sub-Product 1	Product 1	State 87	4,400,000
2000	Sub-Product 1	Product 1	State 88	4,450,000
2000	Sub-Product 1	Product 1	State 89	4,500,000
2000	Sub-Product 1	Product 1	State 90	4,550,000
2000	Sub-Product 1	Product 1	State 91	4,600,000
2000	Sub-Product 1	Product 1	State 92	4,650,000
2000	Sub-Product 1	Product 1	State 93	4,700,000
2000	Sub-Product 1	Product 1	State 94	4,750,000
2000	Sub-Product 1	Product 1	State 95	4,800,000
2000	Sub-Product 1	Product 1	State 96	4,850,000
2000	Sub-Product 1	Product 1	State 97	4,900,000
2000	Sub-Product 1	Product 1	State 98	4,950,000
2000	Sub-Product 1	Product 1	State 99	5,000,000
2000	Sub-Product 1	Product 1	State 100	5,050,000

25

## Data Mining

- Zjišťování závislostí jednotlivých údajů mezi sebou na základě nashromážděných údajů
- Umožňuje vyhledávat vzory informací v údajích
- Je založený na heuristických algoritmech, neuronových sítích apod.
- Pomáhá sledovat a analyzovat trendy a předvídat události



Informační systémy 2006/2007

26

## Data Mining

- Data Mining v MS SQL Server
  - Volba typu zdroje údajů (relační nebo OLAP)
  - Výběr tabulek obsahujících data pro dolování
  - Výběr typu algoritmu pro analýzu údajů
    - MS clustering – vícerozměrné zhlukové diagramy
    - MS decision trees – nevyvážený rozpadový strom
  - Výběr vstupních a predikovaných sloupců
  - Výsledkem je diagram který ukazuje závislosti
  - Na základě výsledků je možno provádět predikci



Informační systémy 2006/2007

27

## Příklad Data Miningu

Jako příklad využijeme již oblíbenou databázi FoodMart. Tentokrát se však pokusíme odhadnout našeho zákazníka. Zkusíme zjistit co ovlivňuje jeho příjem a jak výrazně.

### Postup:

- V Analysis Manageru zadáme vytvoření nového Data Mining modelu, čímž se nám spustí průvodce




---

---

---

---

---

---

---

---

---

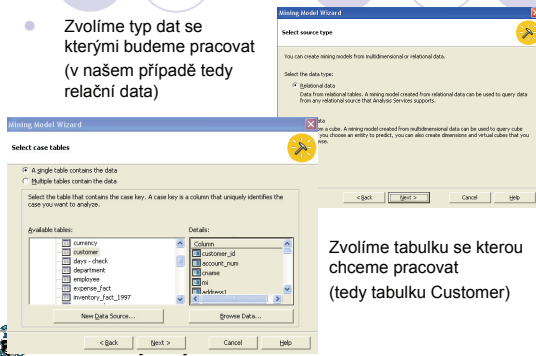
---

---

---

## Příklad Data Miningu

- Zvolíme typ dat se kterými budeme pracovat (v našem případě tedy relační data)



Zvolíme tabulku se kterou chceme pracovat (tedy tabulku Customer)




---

---

---

---

---

---

---

---

---

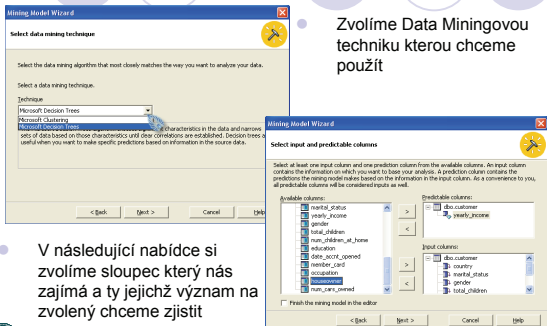
---

---

---

## Příklad Data Miningu

- Zvolíme Data Miningovou techniku kterou chceme použít



- V následující nabídce si zvolíme sloupec který nás zajímá a ty jejichž význam na zvolený chceme zjistit




---

---

---

---

---

---

---

---

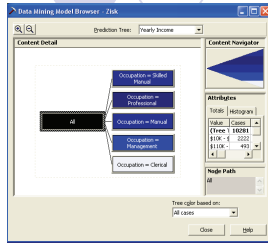
---

---

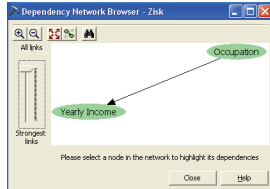
---

---

## Příklad Data Miningu



- Po provedení se nám pomocí barevného schématu zobrazí výsledné vlivy



- A pomocí Dependency Network Browseru se můžeme podívat obdobně na nejvýraznější vlivy



---

---

---

---

---

---

---

---