

Dátové sklady a OLAP

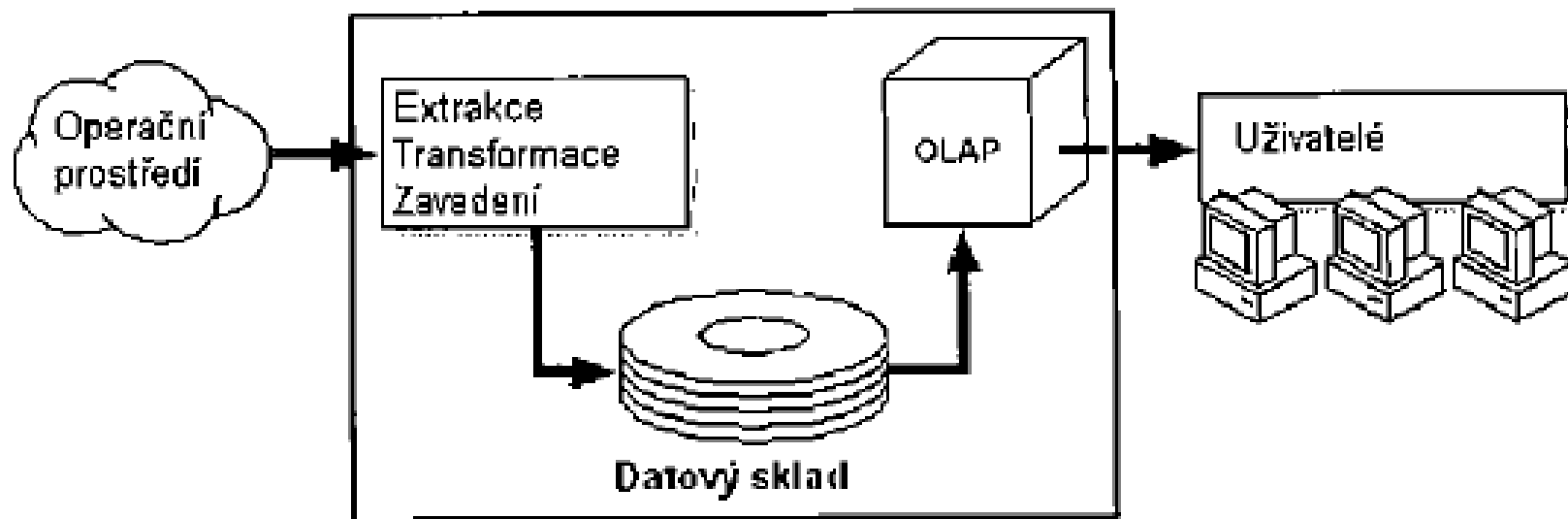
OBSAH

- Dátové sklady
- Dátové trhy
- Metódy budovania dátového skladu
- Mechanizmus naplňania dátového skladu
- OLAP (definícia, typy)
- Operácie OLAP
- Prezentácia údajov v dátovom sklade
- Nástroje pre prezentáciu údajov dátového skladu

Dátové sklady

- B.Inmon: Dátový sklad je úložisko
 - subjektovo orientovaných (orientované na predmet záujmu, nie na aplikáciu, v ktorej boli vytvorené),
 - integrovaných(jednotná terminológia),
 - časovo premenlivých (viac časových úsekov)
 - a nemenných(dáta sa nemenia a neodstraňujú) historických dát použitých pre podporu rozhodovania
 - Atomické a sumárne dáta

Dátové sklady



Dátové trhy

- Podmnožina dátového skladu, určená pre menšie organizačné jednotky firmy (účtovníctvo, marketing)
- Vytvárajú sa pre jednoduchšiu orientáciu používateľov, aj kvôli bezpečnosti citlivých údajov

Budovanie dátového skladu

- **Metóda „veľkého tresku“**

- Vytvorenie dátového skladu implementáciou jedného projektu
- Fázy:
 - Analýza požiadaviek zo strany podniku
 - Vytvorenie dátového skladu
 - Vytvorenie prístupu buď priamo, alebo cez dátové trhy
- Nie je to veľmi dobrý prístup, pretože proces vytvárania dátového skladu je dynamický, môžu sa zmeniť požiadavky zákazníka, ako aj využiteľné technológie
- Je potrebné dlhé čakanie na návratnosť investície

Budovanie dátového skladu II.

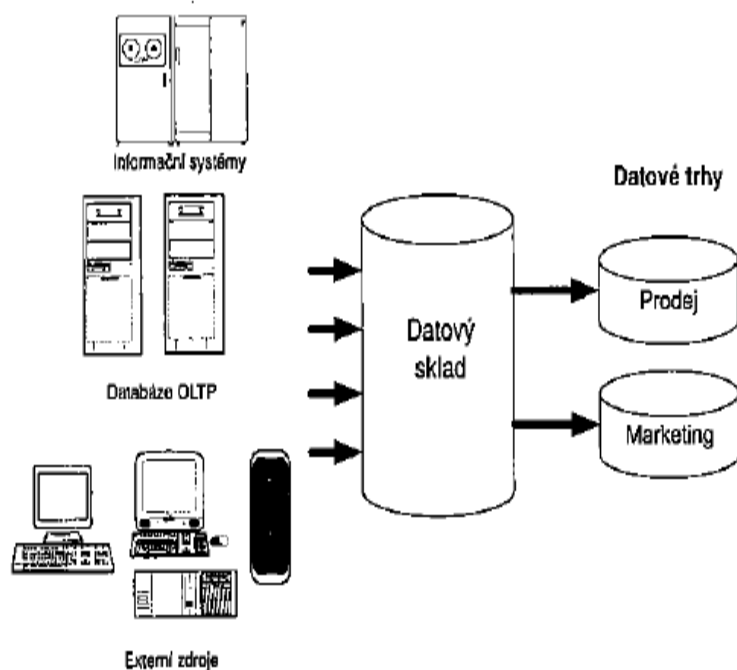
- **Prírastková metóda**

- Vytvorenie dátového skladu po etapách
- Vytvorenie niekoľkých dátových trhov, postupné dopĺňanie projektu
- Iteratívny proces
- Výhody:
 - Kontinuita budovaného projektu
 - Implementácia škálovateľnej, rozšíriteľnej architektúry
 - Rýchlejšia návratnosť investícií
- Fázy budovania:
 - stratégia
 - Definícia
 - analýza
 - návrh
 - Zostavenie
 - produkcia

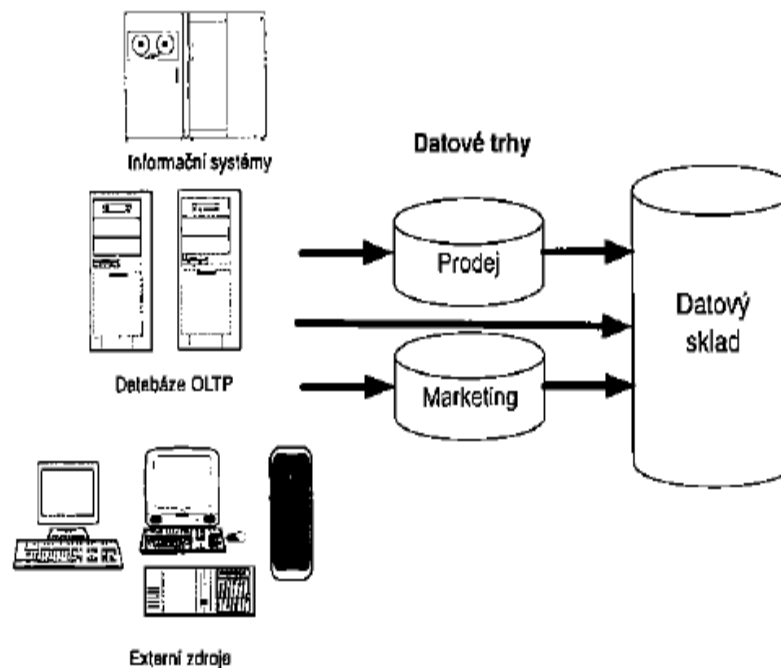
Budovanie dátového skladu III.

- Metódy budovania prírastkovou metódou:

Zhora nadol



Zdola nahor



Mechanizmus naplňania dátového skladu

- Extrakcia, transformácia a zavedenie údajov z produkčných systémov do DW. Etapy:
 1. Definícia toho, kde sa dáta nachádzajú a popísať štruktúry, v ktorých sú uložené
 2. Definovať mapovanie cieľových dát na zdrojové
 3. Popísať spôsob dekódovania a transformácie dát
 4. Určiť procesy, ktoré budú zabezpečovať naplňanie DW
 5. Určiť akým spôsobom budú dáta v DW obnovované

Mechanizmus naplňania dátového skladu

- Spôsoby naplňania:

1. Prvotné naplnenie

2. Naplnenie v prevádzke

1. Vytváranie snímok

- Vytvárajú sa snímky, po ukončení obdobia sa vytvorí konečná snímka, redundantné snímky sa odstránia bez straty informačnej hodnoty.
- Väčšie nároky na kapacitu úložiska, vyššie zaťaženie prenosových liniek
- Transparentnosť implementácie

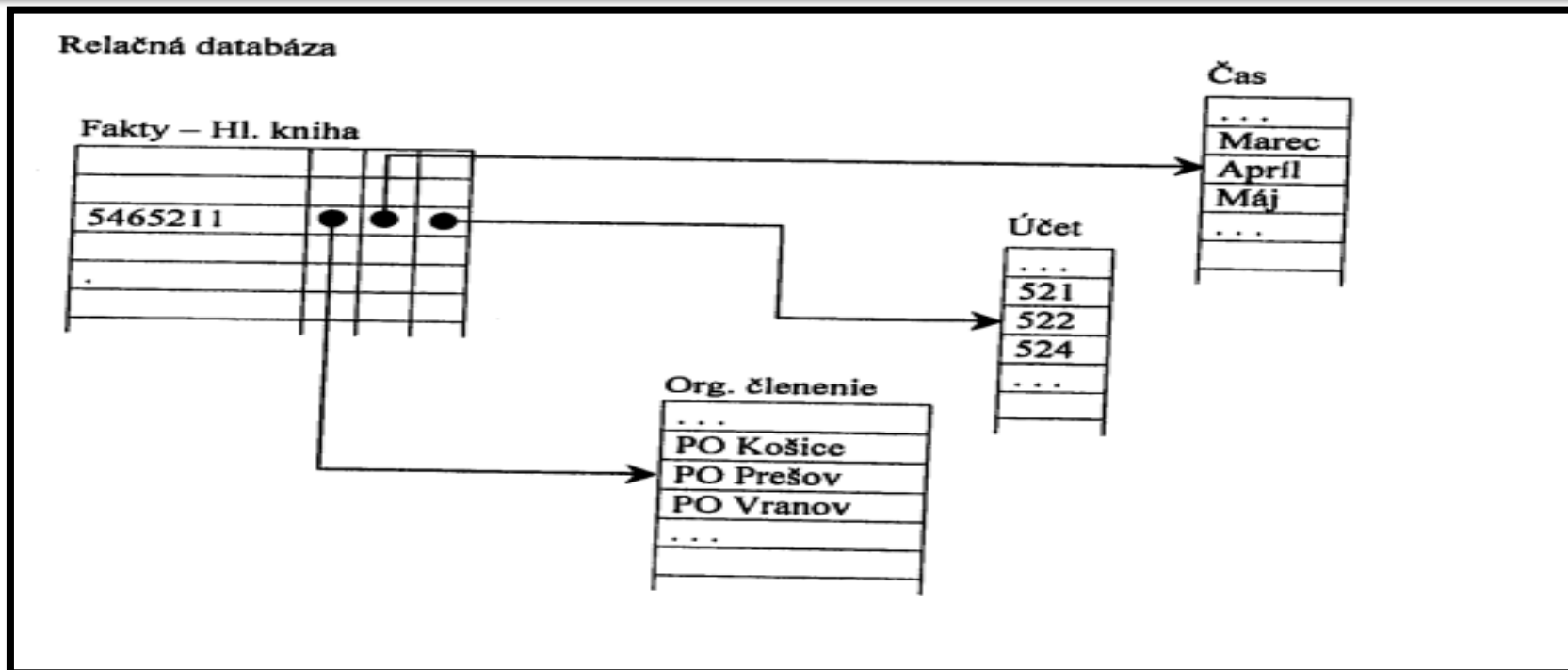
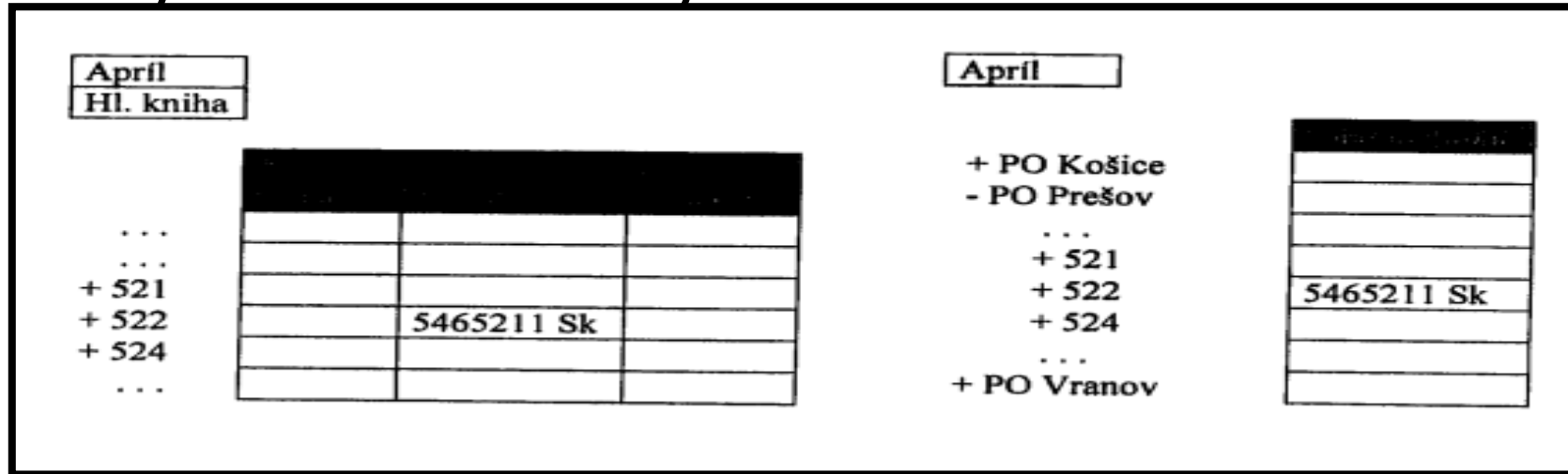
2. Prírastkový spôsob

- Prenáša do DW iba zmeny v transakčnom systéme
- Minimalizuje zaťaženie prenosových liniek
- Neustále sledovanie zmien a stavu DW môže priniesť skryté problémy

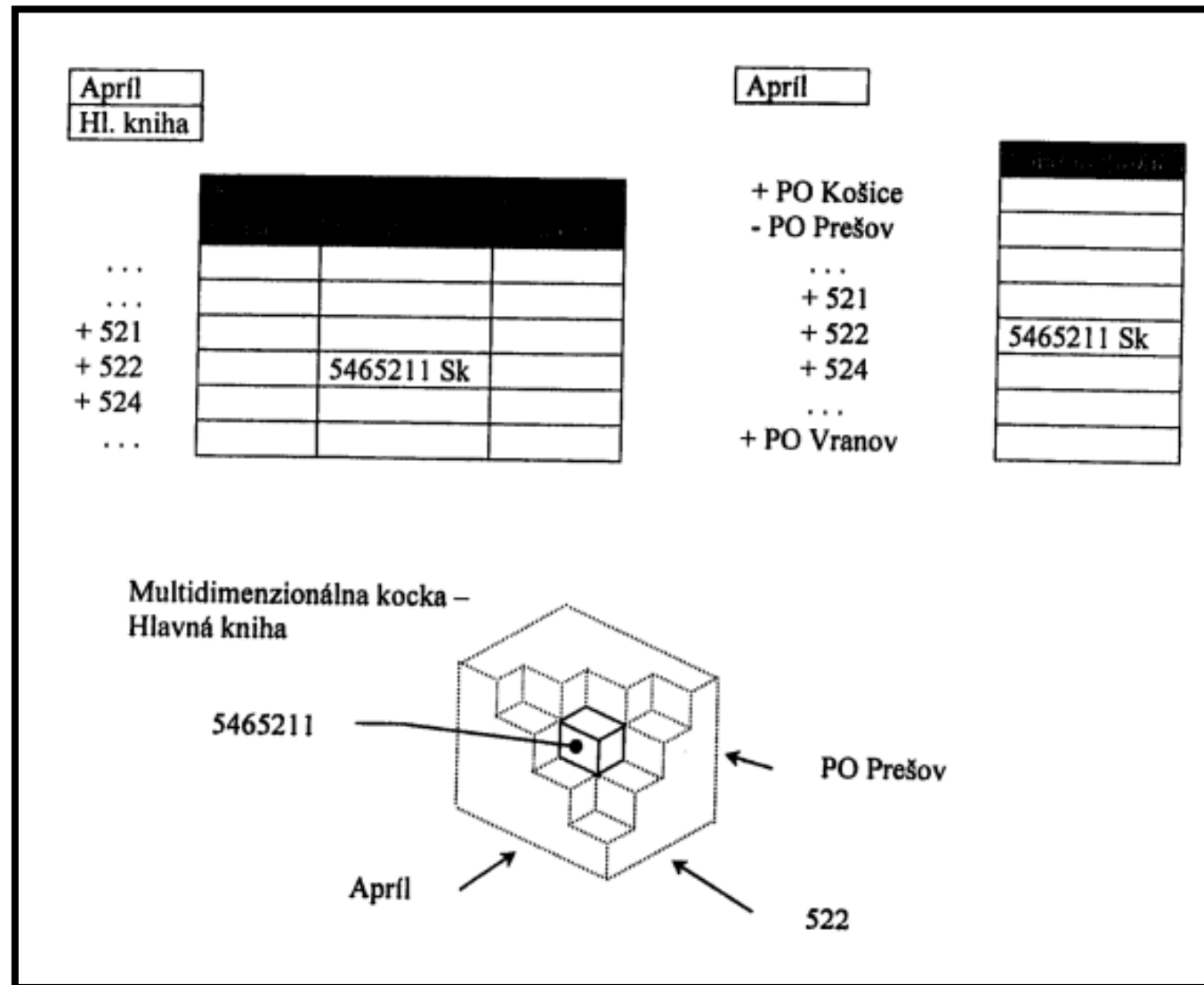
Reprezentácia dát v dátovom sklade

- Dáta môžu byť reprezentované
 1. Relačným dátovým modelom
 - Dáta uložené v tabuľkách,
 - Každý riadok obsahuje dáta ,stĺpce predstavujú údaje o atribútoch
 2. Multidimenzionálnym databázovým modelom
 - Dátový model je možné zobrazit' ako n-rozmernú kocku

Relačný databázový model



Multidimenzionálny databázový model



Relačný vs. Multidimenzionálny model

Relačný model

+ potenciál odborníkov vo firmách, rutinná práca s modelom

+ využitie v transakčných databázach, aj dátových skladoch

Multidimenzionálny model

+ rýchly a komplexný prístup k veľkému množstvu údajov

+ prístup k multidimenzionálnym aj relačným dátovým štruktúram

+ možnosť komplexných analýz

+ schopnosti pre modelovanie a prognózy

Relačný model

- Potenciálne kapacitné obmedzenia

- Obmedzenia dostupnosti dát v rozumnou čase

Multidimenzionálny model

- Problémy pri zmenách dimenzií, menšia flexibilita

- Vyššie kapacitné nároky

OLAP

- On-Line Analytical Processing
- Voľne definované princípy tvoriace rámec pre podporu rozhodovania

Typy OLAP-u

- 1. Relačný OLAP (ROLAP)
 - Využíva dáta z relačného dátového skladu, ktoré poskytuje používateľovi ako multidimenzionálny pohľad
 - Data a metadata sú takisto uložené v relačnej databáze a sú využité k tomu, aby boli pomocou nich generované príkazy SQL potrebné k vyhotoveniu používateľského dopytu
 - Nevzniká redundancia
- 2. Multidimenzionálny OLAP (MOLAP)
 - Ukladá dáta vo vlastných dátových štruktúrach a sumároch
 - Počas tohto procesu vykoná čo najviac medzivýpočtov
 - Výhodou je rýchly prístup k dátam, nevýhodou redundancia údajov
- 3. Hybridný OLAP (HOLAP)
 - Údaje zostávajú v relačných databázach, agregáty sú uložené v multidimenzionálnej štruktúre

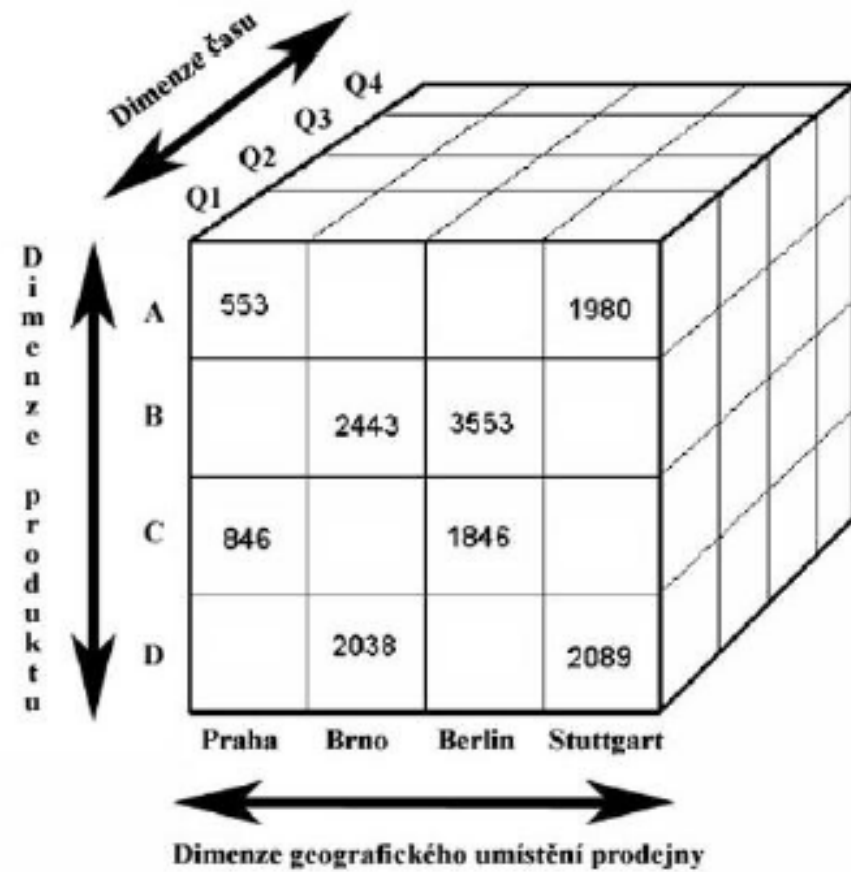
Analýza OLAP – Pojmy II.

Dimenzia

- Dimenzia – obsahuje logicky alebo hierarchicky usporiadané údaje
- Tabuľky dimenzií
 - Menšie než tabuľky faktov
 - Dáta sa v nich často nemenia (výnimkou napr. dáta o zákazníkoch)
 - Často sa používajú časové, produktové a geografické dimenzie

Analýza OLAP – pojmy III.

Dimenzia



Analýza OLAP – Pojmy

Fakt

- Fakt – numerická merná jednotka
- Tabuľka faktov
 - najväčšia tabuľka v databáze, obsahuje veľký objem dát
 - Sledovaný fakt a odkazy (cudzíe kľúče) na tabuľky jednotlivých dimenzií

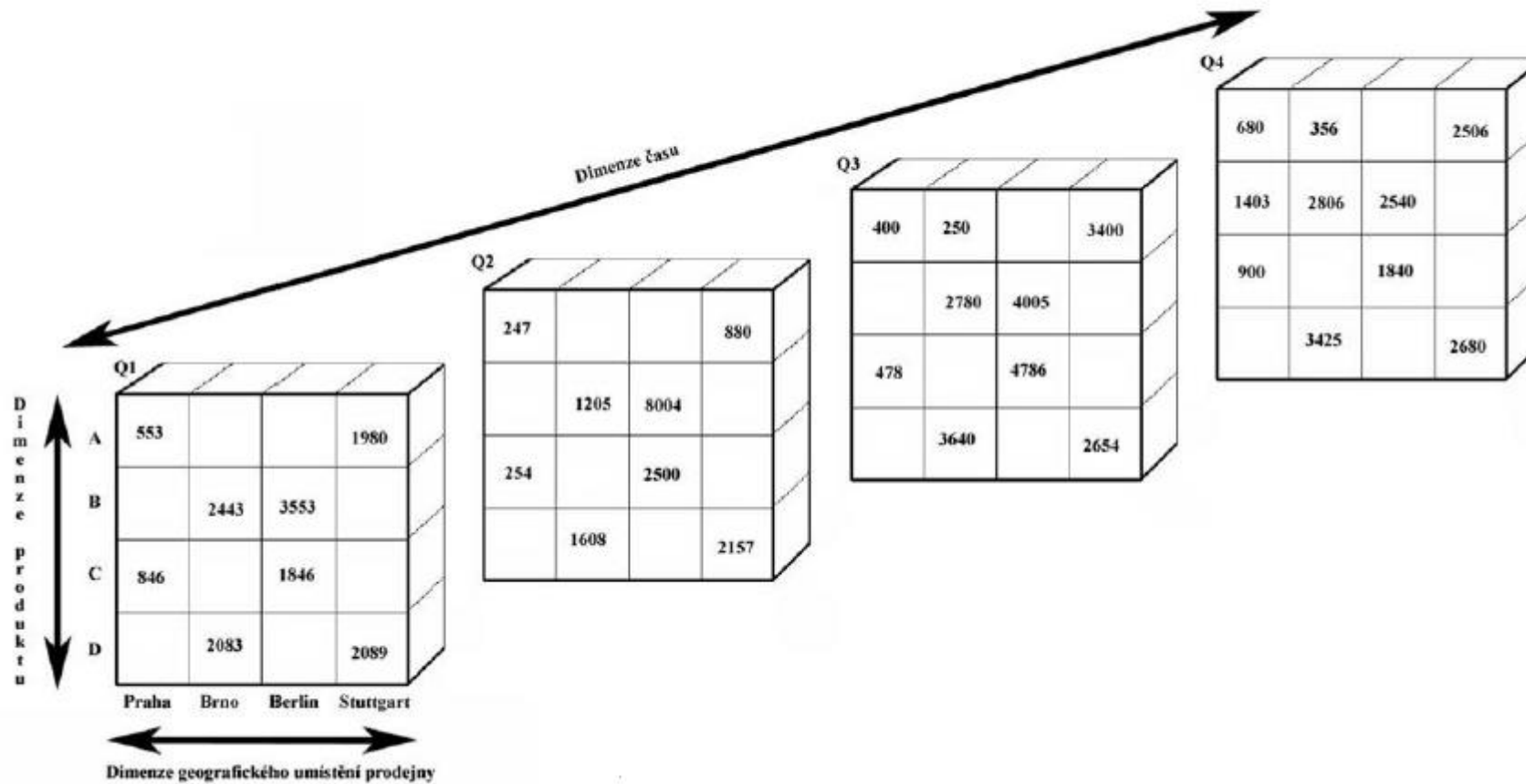
Analýza OLAP – Pojmy

Tabuľka Faktov

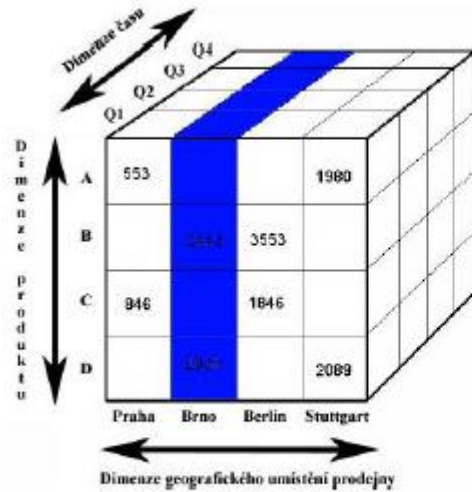
POCET_ID	POCET	FK_PREDAJNA_ID	FK_DATUM_ID	FK_PRODUKT_ID	FK_Miesto_Predajne_ID
1	6.8	84	1	15	2
2	32.8	84	1	16	2
3	154.4	84	1	17	2
4	1.9	84	1	18	2
5	4.6	84	1	22	2
6	66.6	84	1	24	2
7	74.2	84	1	25	2
8	6.3	84	1	26	2
9	8.7	84	1	1	2
10	19.7	84	1	29	2
11	29	84	1	3	2
12	29.8	84	1	30	2
13	2	84	1	4	2
14	2.5	84	1	5	2
15	2.4	84	1	6	2
16	1.4	84	1	7	2

Analýza OLAP – pojmy IV.

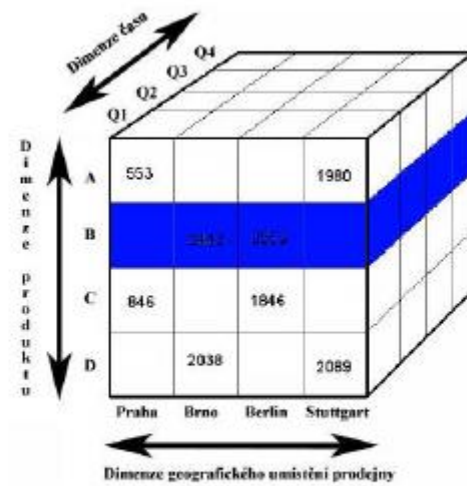
Dimenzia



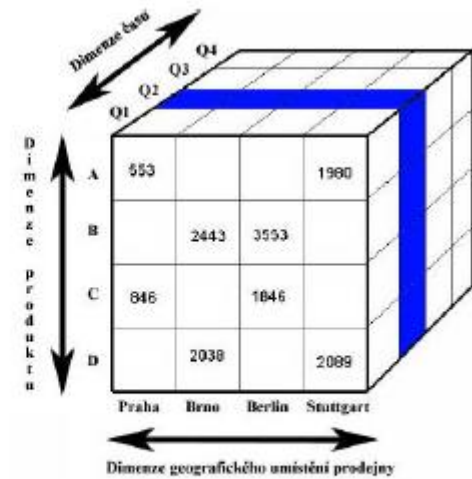
Analýza OLAP – pojmy V. Dimenzia



A - Analýza údajů podle geografických kritérií



B - Analýza údajů podle produktu



C - Analýza údajů podle časových kritérií

Analýza OLAP – pojmy VI.

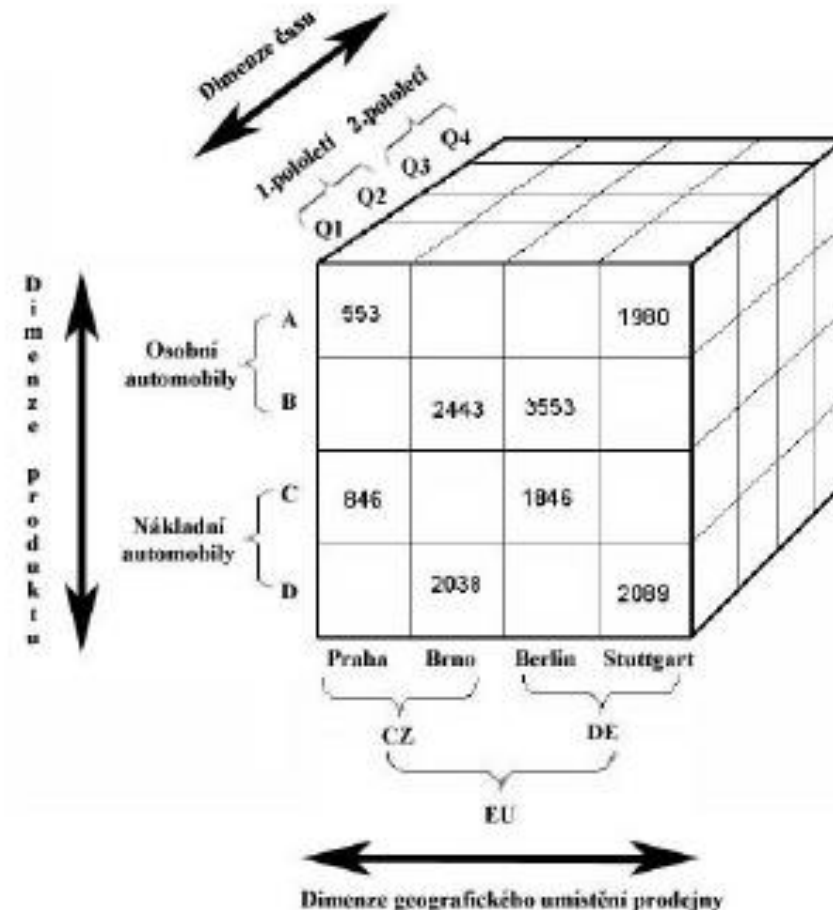
Hierarchie

- Hierarchie – dimenzie majú obvykle stromovú (hierarchickú štruktúru)
- Napr.:

Dimenzia/ úrovne	REGIÓN	PRODUKT	ČAS
NAJVYŠŠIA	Kontinent	Druh	Rok
....	Štát	Kategória	Kvartál
....	Kraj	Subkategória	Mesiac
NAJNIŽŠIA	Mesto	Názov	Deň

Analýza OLAP – pojmy VII. Hierarchie II.

- **Roll-up** – vystupovanie na vyššie úrovne danej dimenzie
- **Drill-down** – zostupovanie na nižšie úrovne
- **Slice** – projekcia cez jednu dimenziu
- **Dice** – projekcia cez viac dimenzií



Prezentácia údajov v dátovom sklade

- Možnosti prezentácie údajov v dátovom sklade:

1. SQL Používateľský dopyt

- Na základe SQL dopytu sa vytvorí zostava dát
- Výsledky takýchto dopytov sú spravidla v RD uložené ako pohľady a sprístupnené patričnej skupine používateľov
- Typy dotazov:
 - Základné dopyty
 - Hierarchické dopyty
 - Dopyty s využitím kalkulácií
 - Využitie atribútov pre agregáciu

Príklad SQL dopytu

```
SELECT customer, ROUND(MIN(sales)) minimum, ROUND(MAX(sales)) maximum,  
       ROUND(AVG(sales)) average  
FROM  
  (SELECT customer_ldsc customer, time_ldsc time, sales  
    FROM units_cube_cubeview  
    WHERE time_calendar_yea_prnt = 'QUARTER_67'  
    AND product_ldsc = 'Sentinel Standard'  
    AND customer_market_segme_prnt = 'MARKET_SEGMENT_4'  
    AND channel_level = 'TOTAL_CHANNEL'  
  )  
GROUP BY customer  
ORDER BY customer;
```

Výsledek vnoreného dopytu

CUSTOMER	TIME	SALES
Dept. of Communication	Aug-01	1752.06
Dept. of Communication	Jul-01	5344.2
Dept. of Communication	Sep-01	3507.06
Dept. of Labor	Aug-01	
Dept. of Labor	Sep-01	1753.53
Dept. of Labor	Jul-01	3562.8
Ministry of Finance	Jul-01	1781.4
Ministry of Finance	Aug-01	3504.12
Ministry of Finance	Sep-01	7014.12
Ministry of Intl Trade	Jul-01	5344.2
Ministry of Intl Trade	Sep-01	5260.59
Ministry of Intl Trade	Aug-01	5256.18
Royal Air Force	Jul-01	3562.8
Royal Air Force	Sep-01	3507.06
Royal Air Force	Aug-01	8760.3
UK Environmental Department	Aug-01	
UK Environmental Department	Sep-01	
UK Environmental Department	Jul-01	3562.8
US Dept. of Research	Jul-01	1781.4
US Dept. of Research	Aug-01	1752.06
US Dept. of Research	Sep-01	1753.53
US Marine Services	Sep-01	
US Marine Services	Aug-01	
US Marine Services	Jul-01	

Výsledok vonkajšieho dopytu

CUSTOMER	MINIMUM	MAXIMUM	AVERAGE
Dept. of Communication	1752	5344	3534
Dept. of Labor	1754	3563	2658
Ministry of Finance	1781	7014	4100
Ministry of Intl Trade	5256	5344	5287
Royal Air Force	3507	8760	5277
UK Environmental Department	3563	3563	3563
US Dept. of Research	1752	1781	1762
US Marine Services			

Prezentácia údajov v dátovom sklade

2. Tabuľka

- Umožňuje zobrazenie výsledku používateľského dopytu
- Typy:
 - 1. tabuľka – hodnota sledovaného faktu a požadovaných dimenzií sa nachádza len na jednej osi
 - 2. krížová tabuľka – hodnoty požadovaných dimenzií sú na oboch osiach ,v ich priesečníku je zobrazená hodnota sledovaného faktu pre požadované hodnoty

Prezentácia údajov v dátovom sklade

- Jednorozmerná tabuľka





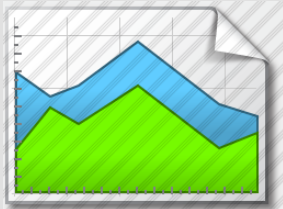

	Cena SUM	Predajna	Skupina	Rok
▶ 1	565545,43	Urob si sam	stavebny material	2004
▶ 2	281555,53	Hornbach	stavebny material	2004
▶ 3	187195,23	Kovomat	stavebny material	2004
▶ 4	174692,28	Tesco	stavebny material	2004
▶ 5	143469,95	Zeleziarstvo	stavebny material	2004
▶ 6	142035,92	Urob si sam	spotrebny material	2004
▶ 7	130360,08	Baumax	stavebny material	2004
▶ 8	108872,05	MIBU zeleziarstvo	stavebny material	2004
▶ 9	100527,4	Tesco	naradie	2004
▶ 10	89841,69	Mega zeleziarstvo	stavebny material	2004
▶ 11	69349,59	Zaleziarstvo merkur	stavebny material	2004
▶ 12	58059,16	Hornbach	spotrebny material	2004
▶ 13	44775,79	Tesco	spotrebny material	2004
▶ 14	43701,47	Kovomat	spotrebny material	2004
▶ 15	38274,57	Lapal	stavebny material	2004
▶ 16	36548,06	Zeleziarstvo	spotrebny material	2004
▶ 17	31724,39	Baumax	spotrebny material	2004
▶ 18	27782,09	MIBU zeleziarstvo	spotrebny material	2004
▶ 19	27571	Baumax	naradie	2004
▶ 20	21140,88	Mega zeleziarstvo	spotrebny material	2004
▶ 21	17126,8	Mega zeleziarstvo	naradie	2004
▶ 22	14168,1	Zaleziarstvo merkur	spotrebny material	2004
▶ 23	12282,4	Proving	naradie	2004
▶ 24	10419,1	Urob si sam	naradie	2004
▶ 25	6605,7	Lapal	spotrebny material	2004
▶ 26	172,4	Proving	spotrebny material	2004

- Krížová tabuľka

	Sales	Sales From Prior	Change in Sales	Percent Change
▶ North America	58,255,803	48,031,032	10,224,771	0
▶ Asia Pacific	46,706,721	24,034,447	22,672,274	1
▶ Europe	12,773,568	12,455,443	318,125	0
▶ North America	7,676,832	4,735,316	2,941,516	1
▶ Asia Pacific	2,401,239	1,866,945	534,295	0
▶ Europe	2,462,351	1,392,112	1,070,239	1

Prezentácia údajov v dátovom sklade

- 3. graf
- Prezentácia údajov v grafickej forme
- Typy grafov:

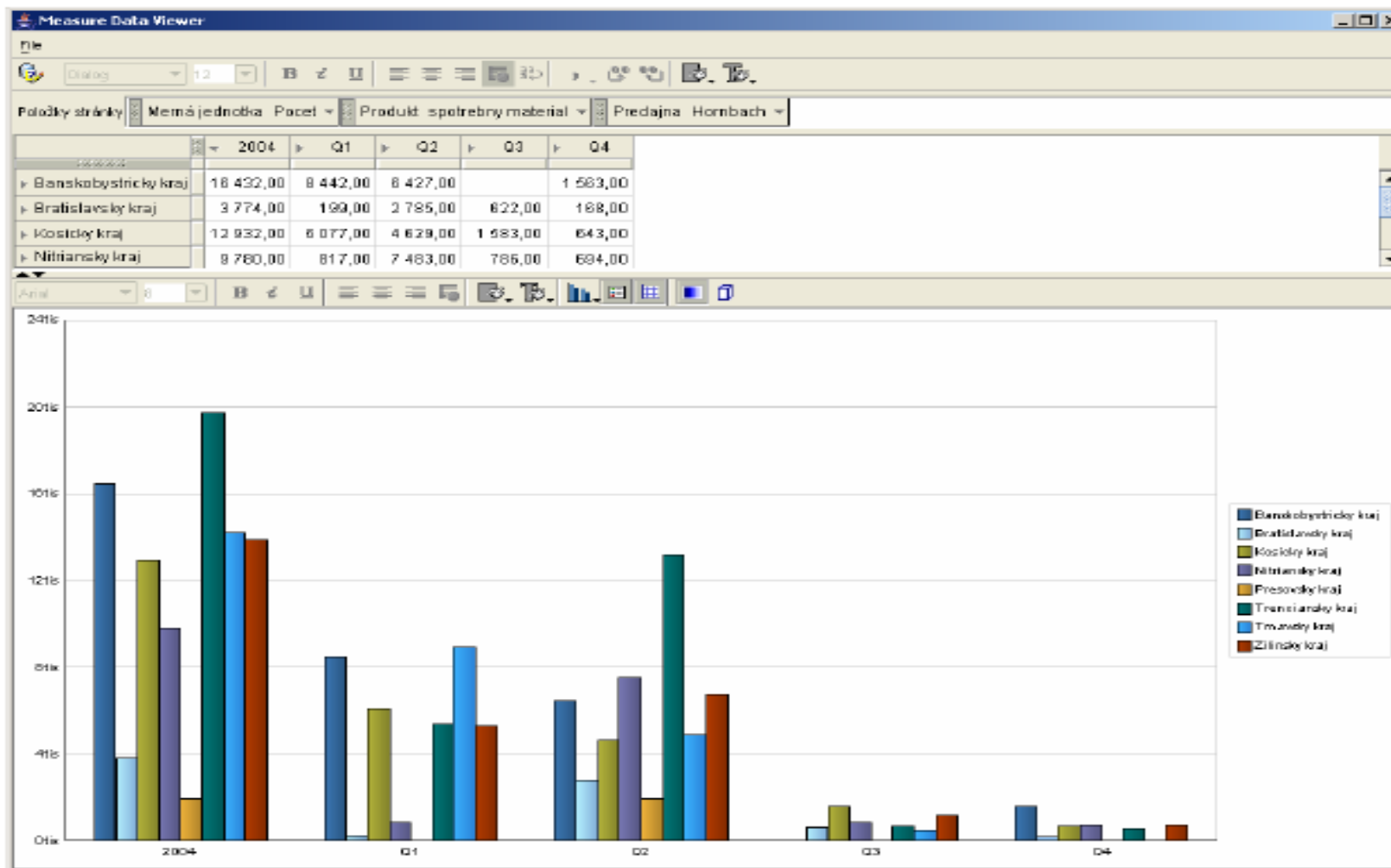
Stĺpcový	Koláčový	Čiarový
		
Pruhový	Plošný	Kombinovaný
		

Nástroje pre prezentáciu údajov dátového skladu

1. Analytic Workspace Manager (AWM)

- Slúži na realizáciu multidimenzionálnej dátovej kocky a následnú realizáciu jednoduchých analýz
- Pracuje na multidimenzionálnej úrovni (MOLAP)
- Oproti ostatným analytickým nástrojom umožňuje len malé spektrum možností pre realizáciu analýz

Ukážka AWM



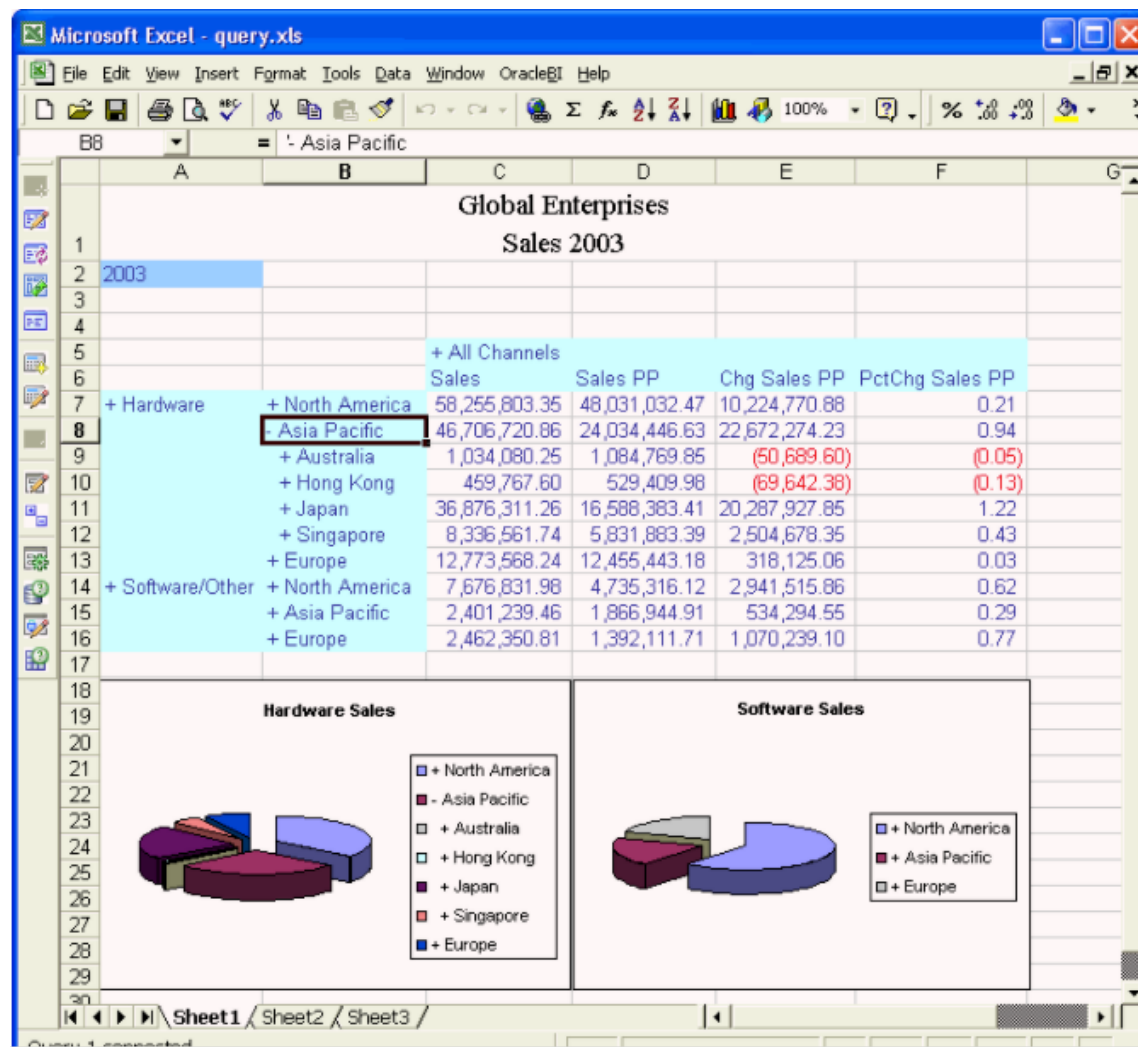
Ročný a kvartálny pohľad spotrebného materiálu v predajni Hornbach.

Nástroje pre prezentáciu údajov dátového skladu

2. OracleBI Spreadsheet Add-In for Excel

- Umožňuje spracovanie reportov z multidimenzionálnej kocky v programe Microsoft Excel
- Pracuje s multidimenzionálnou databázou (MOLAP)
- Umožňuje pre realizáciu reportov využívať okrem funkcionality BI pluginu aj samotné vstavané funkcie MS Excel

Ukážka BI Add-in for Excel

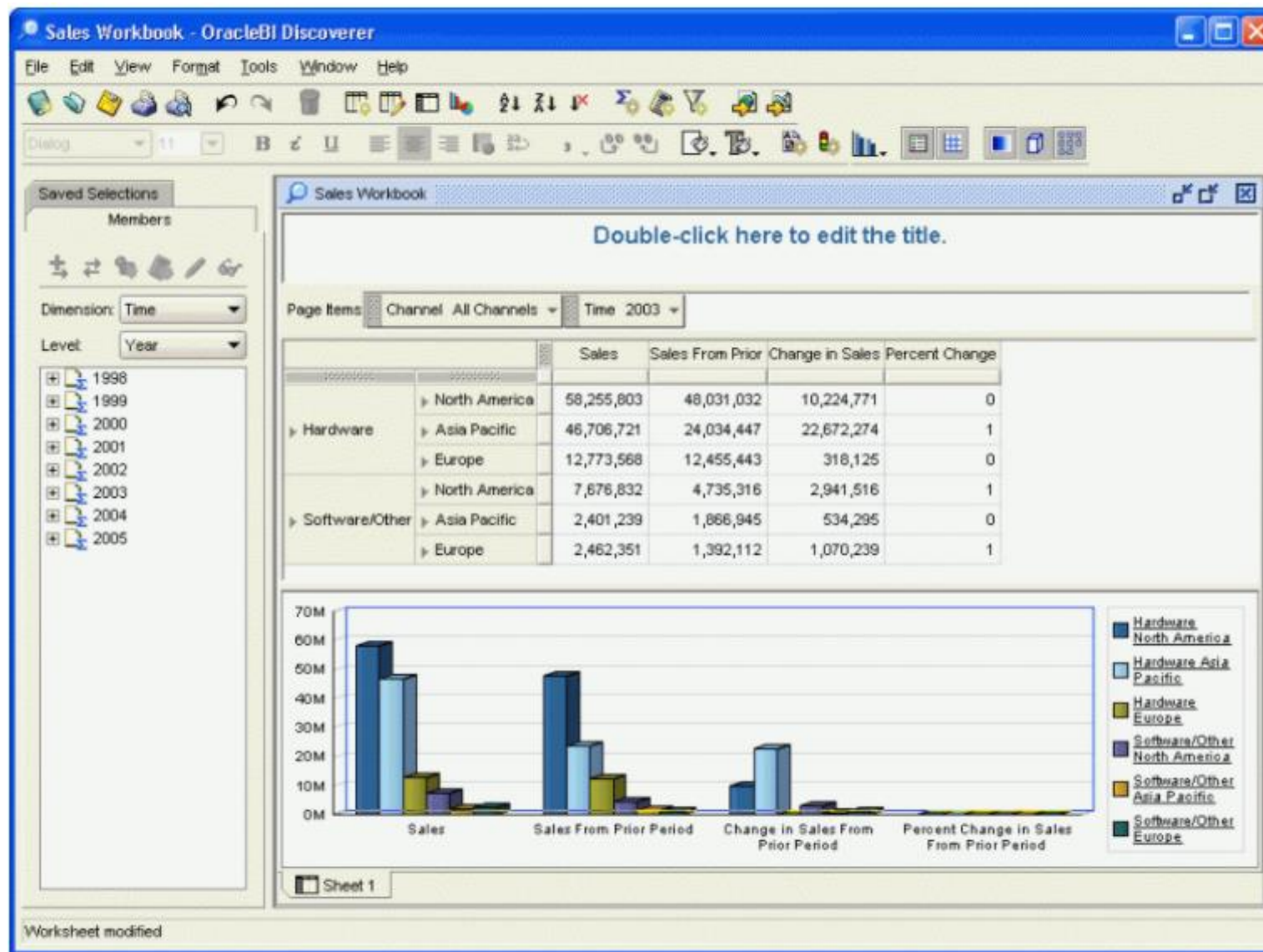


Nástroje pre prezentáciu údajov dátového skladu

3. Oracle Discoverer

- Umožňuje tvorbu analýz a používateľských zostáv v prostredí špeciálne pre to určenom
- Pracuje s dátami na relačnej úrovni (ROLAP)
- Má 2 časti:
 - Administrator – administrátorská časť realizácie používateľského dopytu (definícia hierarchií, vytvorenie väzieb, definovanie privilégií...)
 - Desktop – časť pre realizáciu používateľských zostáv zo schém definovaných v časti Administrator

Ukážka Oracle Discoverer



Ďalšie možnosti spracovania a analýzy dát

- Vytváranie kalkulovaných (vypočítavaných) faktov
- Vytváranie predpovedí
- Zložité agregácie
- Vytváranie reportov
- Vývoj JAVA aplikácií pre OLAP