

Dátové sklady Prehľad

Jana Dvořáková

1.10.2010

Pokročilé databázové technológie, FIIT STU



IT řešení, které přináší ovoce

Obsah

Prevádzkové systémy

Analytické dotazy

Čo je dátový sklad

Architektúra

Front end, data mining

Zhrnutie a diskusia



Prevádzkové systémy podniku - príklady

Billing – systém pre účtovníctvo

- Príjmy, výdaje
- Účty, hlavná účtovná kniha
- Aktíva, pasíva
- Finančné výkazy, ...

HR – personálny systém

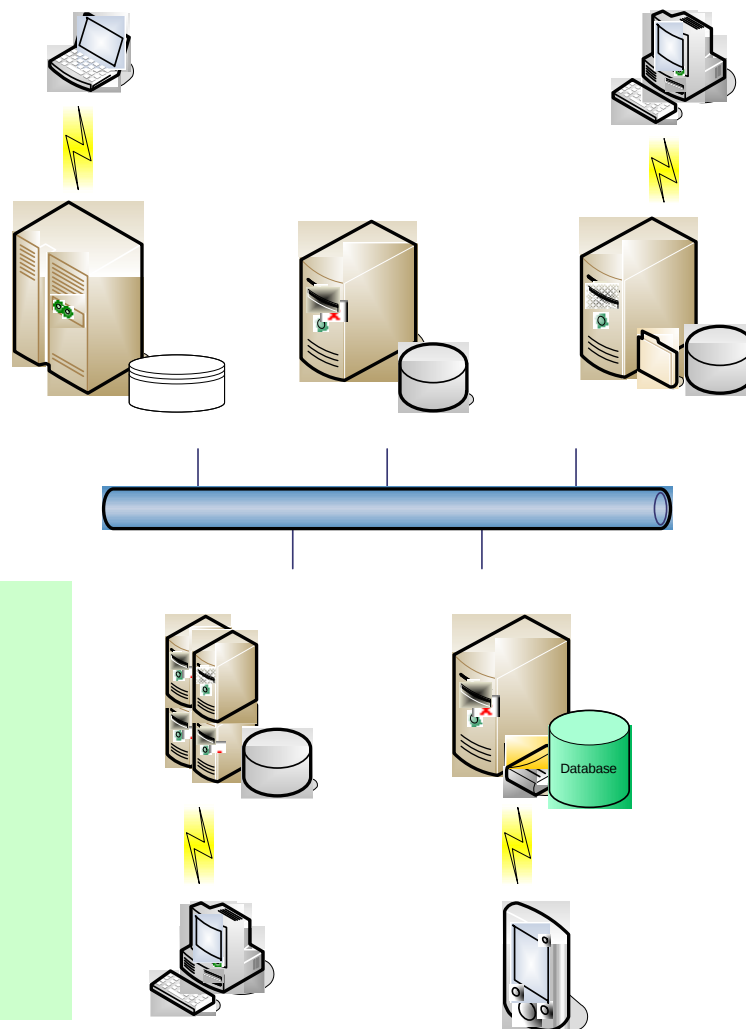
- Zamestnanci, pozície, mzdy,
- Dochádzka, utilizácia, dovolenky
- Školenia, benefity, kariérne postupy, ...

CMS – Content Management System

- Intranet - vnútro podnikový portál, informácie pre zamestnancov
- Internet - webový portál informácie pre klientov, ...

CRM – Customer Relationship Management

- Segmenty zákazníkov, portfólia
- Produktový katalóg
- Podpora predaja
- História kontaktov
- Kampane, ...

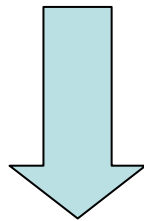
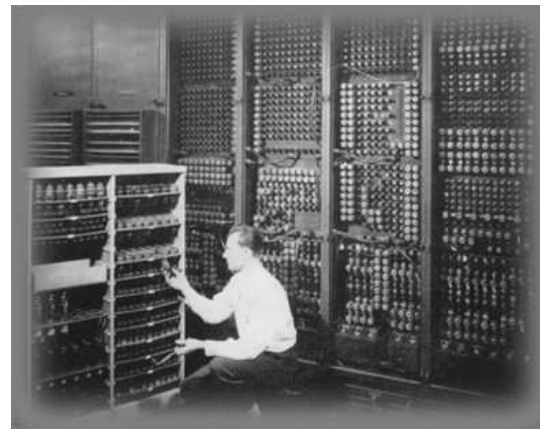


Prevádzkové systémy podniku - vlastnosti

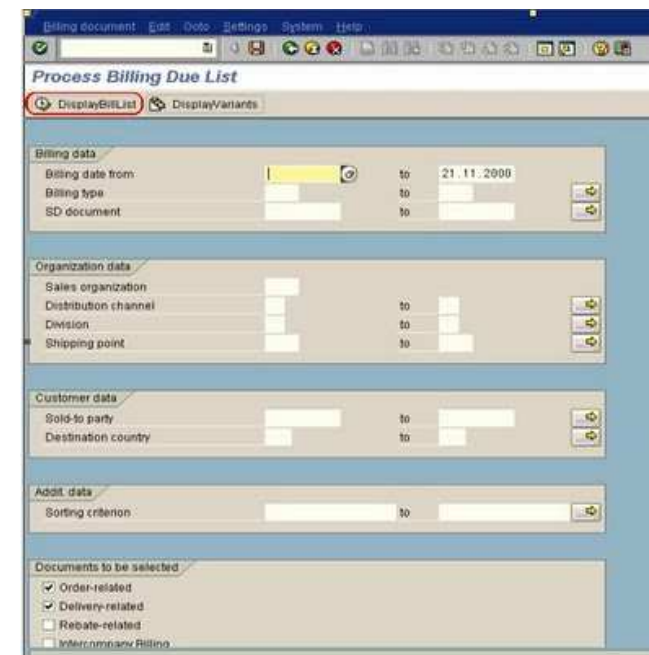
- Prevádzkové (tranzakčné) systémy
 - Operatívne dáta
- Podpora podnikania, prevádzka činností podniku
- Vznikajú v nich údaje
 - Zadávanie dát užívateľmi (napr. založenie účtu na pobočke banky)
 - Automatické plnenie dátami (napr. transakcie výberu hotovosti v bankomatoch)
- Interakcia (integrácia)
 - Dátové rozhrania - výmena dát
 - Aplikačné rozhrania – prepojenie functionality

Prevádzkové systémy podniku – HW a SW

- Rôznorodý HW, SW
 - Historické mainframy, aplikácie v Cobole, konzolové UI, zložitý prístup k dátam



- Moderné aplikácie s grafickým UI a plnohodnotným transakčným DBMS



Obsah

Prevádzkové systémy

Analytické dotazy

Čo je dátový sklad

Architektúra

Front end, data mining

Zhrnutie a diskusia



Analytické dotazy - príklady

Ktoré pridružené produkty
sa najviac predávajú na
východnom Slovensku
spolu s bežným účtom?

BANKA



O koľko percent sa zvýšilo
poistné plnenie za škodu
spôsobenú povodňami za
posledný rok?

POISŤOVŇA



Analytické dotazy - vlastnosti

- Vyžadujú vstupné dáta z viacerých prevádzkových systémov
- Komplexná selekcia a transformácia dát
 - Typicky nie je realizovaná jedným skriptom, ale je rozdelená na predprípravu dát a spracovanie dát
- Orientácia na subjekt
 - Produkt, zákazník, región, ...
- Odpovede na analytické dotazy slúžia na **podporu rozhodovania** pre podnikový manažment

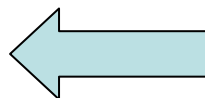


Analytické dotazy – zdrojové systémy



NEW
FRONTIER
SLOVAKIA

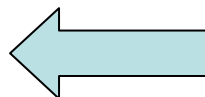
Ktoré pridružené produkty sa najviac predávajú na východnom Slovensku spolu s bežným účtom?



- Databáza aplikácie pre predaj
- Databáza zákazníkov (adresy zákazníkov)
- Centrálny číselník regiónov
- Produktový katalóg

BANKA

O koľko percent sa zvýšilo poistné plnenie za škodu spôsobenú povodňami za posledný rok?



- Evidencia poistných zmluv
- Databáza aplikácie pre spracovanie poistných udalostí
- Produktový katalóg neživotného poistenia

POIŠŤOVŇA

Analytické dotazy – podpora rozhodovania

Otázka?

Odpoveď

Rozhodnutie → zisk (BC)

Ktoré pridružené produkty sa najviac predávajú na východnom Slovensku spolu s bežným účtom?

Internet Banking a Visa Electron

Predajcovia na pobočkách vých. Slovenska budú automaticky ponúkať k bežnému účtu aj Internet Banking a Visa Electron → zvýšenie predaných IB a Visa Electron na vých. Slovensku

BANKA

O koľko percent sa zvýšilo poistné plnenie za škodu spôsobenú povodňami za posledný rok?

O 50%

Zvýši sa poistné pre poistenie proti povodňam → udržanie rovnakých ziskov aj pri zvýšení pravdepodobnosti poistnej udalosti

POIŠŤOVŇA

Realizácia analytických dotazov

1. Manuálne

- Skupina skriptov pre prípravu a spracovanie dát z prevádzkových systémov
- Neprehľadné, ťažko udržiavateľné
(najmä vo veľkých podnikoch pri veľkom počte prevádzkových systémov)

2. Pomocou dátového skladu

- Samostatná databáza obsahujúca dáta predpripravené pre analytické dotazy

Obsah

Prevádzkové systémy

Analytické dotazy

Čo je dátový sklad

Architektúra

Front end, data mining

Zhrnutie a diskusia



Čo je dátový sklad?

Formálne:

- Subjektovo orientovaná
- Integrovaná, jednotná
- V čase organizovaná
- Trvalo uložená

množina dát umožňujúca
podporu rozhodovania.

(Bill Inmon, 1990)



Čo to znamená...

- **Subjektovo orientovaná**
 - Dáta poskytujú informáciu o konkrétnom subjekte (produkt, zákazník, ..), nie o operáciach/prevádzke podniku
- **Integrovaná, jednotná**
 - Dáta sú zoskupené z rôznych zdrojových systémov do jednotného celku
- **V čase organizovaná (historizovaná)**
 - Všetky dáta v dátovom sklade sú identifikované časovou značkou
- **Trvale uložená**
 - Dáta sú stabilné, read-only. Dáta pribúdajú, ale nie sú prepisované ani odmazávané.

Dátový sklad - vlastnosti

Príprava dát

- **Spája dáta** z viacerých zdrojových systémov, **transformuje dáta** do podoby potrebnej pre analytické dotazy
 - Automatický batchový proces
 - Pravidelné opakovanie rovnakých operácií

Prínosy

- Analytici dostanú pred-pripravené dáta a môžu ich analyzovať
- Chybovosť prípravy dát eliminovaná

OLTP vs. OLAP technológia

- OLAP – Online Analytical processing, súbor technológií a nástrojov pre podporu analytického dotazovania
- Viac v časti „Techniky“

DWH – súčasť modernej podnikovej IT architektúry

Dátové sklady sú súčasťou širšieho pojmu **Business Intelligence**

Obsah

Prevádzkové systémy

Analytické dotazy

Čo je dátový sklad

Architektúra

Front end, data mining

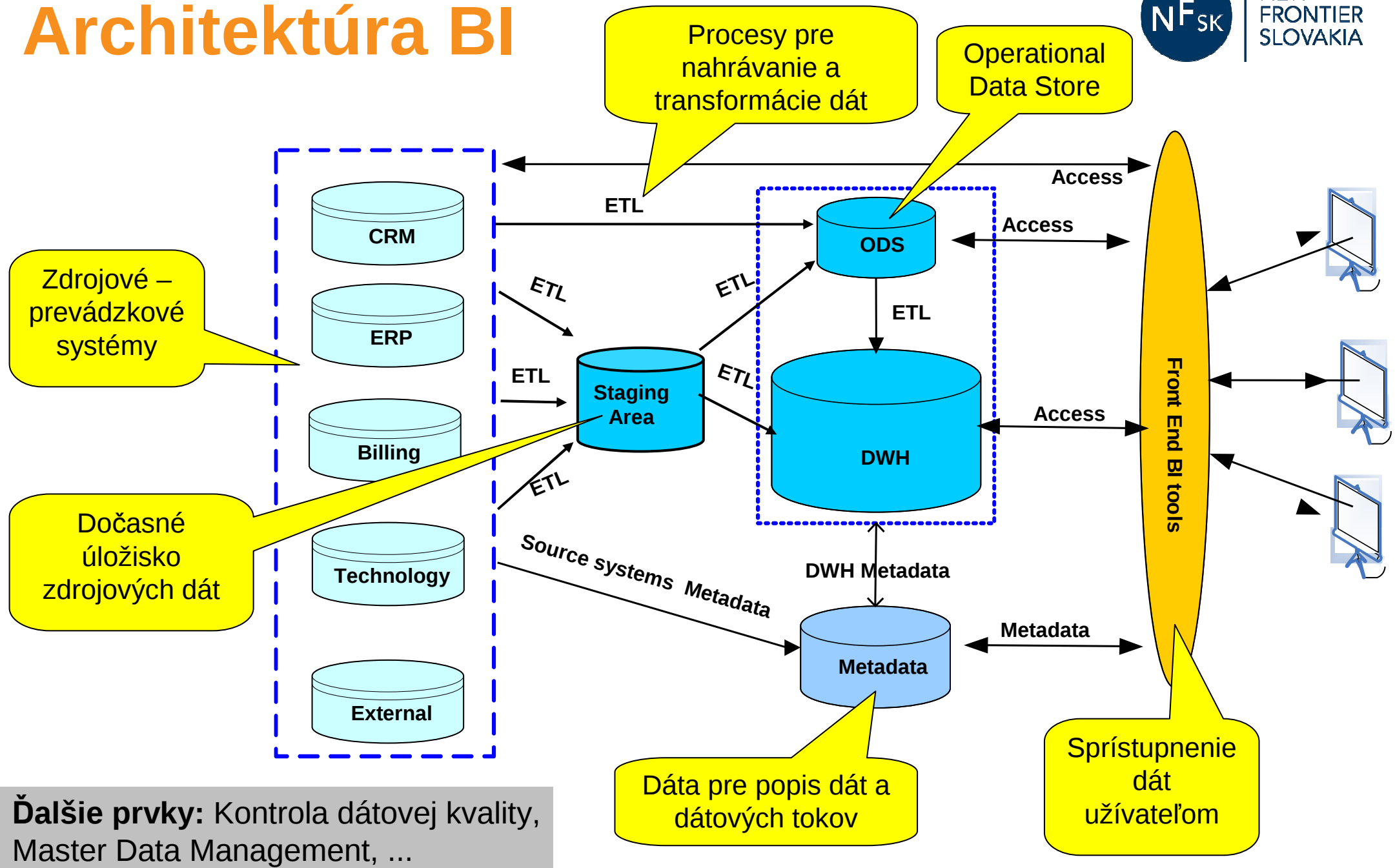
Zhrnutie a diskusia



Architektúra BI



NEW
FRONTIER
SLOVAKIA



DWH – prístupy k architektúre

Datamart

= podmnožina dát z DWH, orientovaná na určité oddelenie, alebo typ úloh

CIF – Corporate Information Factory: 3-vrstvová architektúra (Bill Inmon)

1. Dočasné úložisko (staging area)
2. Centrálné úložisko dát (normalizované)
3. Datamarty – nezávislé, vytvorené nad centrálnym úložiskom, každé oddelenie podniku má vlastné, prispôsobené datamarty

DWHBus – Data Warehouse Bus: 2- vrstvová architektúra (Ralph Kimball)

1. Dočasné úložisko (staging area)
2. Datamarty - konsolidované, zdieľané dimenzie

Obsah

Prevádzkové systémy

Analytické dotazy

Čo je dátový sklad

Architektúra

Front end, data mining

Zhrnutie a diskusia



Front End

= sprístupnenie informácií
používateľom

Formy prístupu

- Reporting, dashboard
- OLAP
- Plánovanie, forecasting, what-if analýzy
- Ad-hoc query



FE – reporting, dashboard

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window WhiteLight Help

Arial 10 B I U

A14 = SK

OLAP Report1

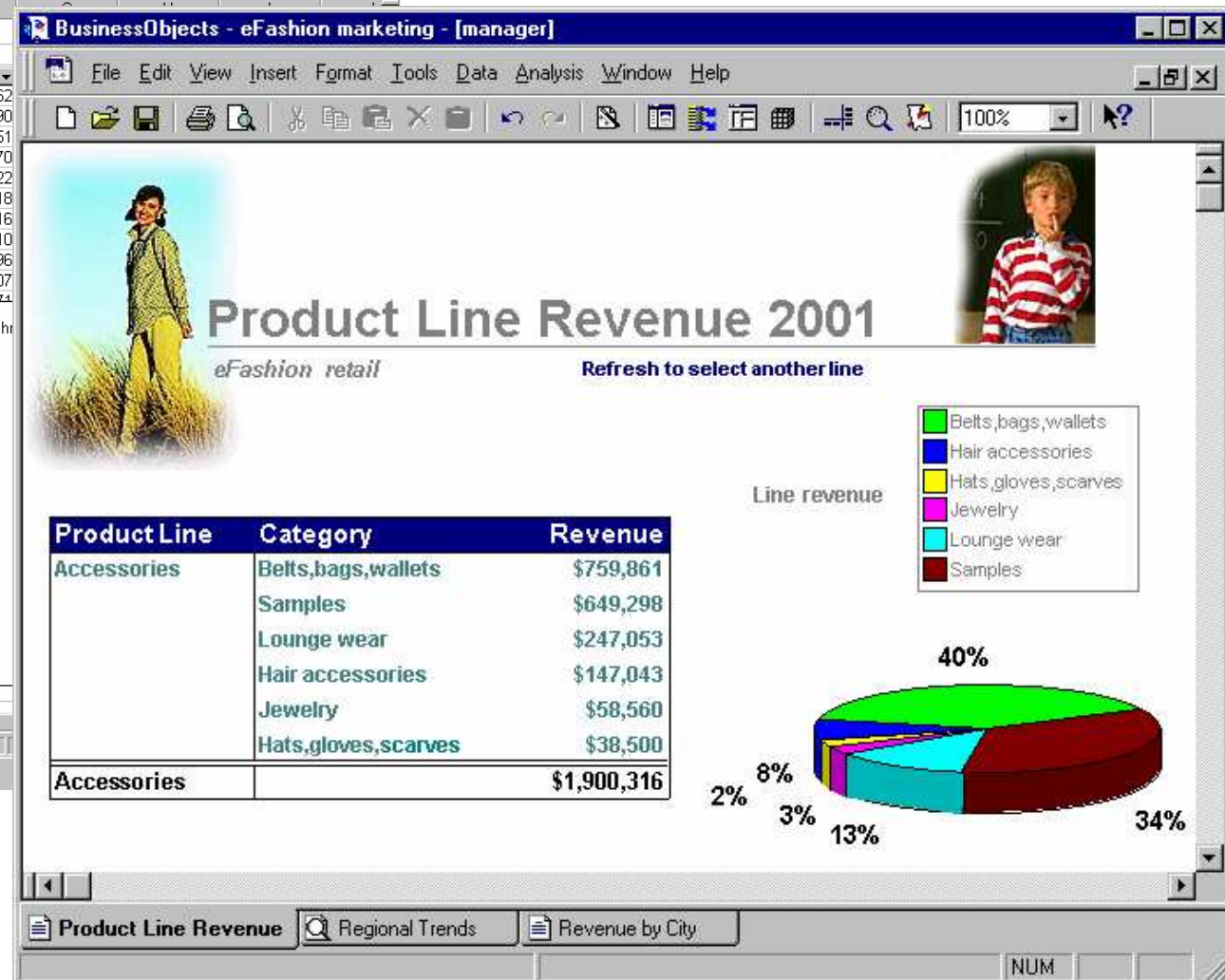
	A	B	C	D	E	F
1	OLAP Report1					
2						
3		Minula obdobi	1991	Q11991	1992	1993
4	Partneri podle regionu - vsichni	2,07218E+11	16579986428	3771957208	16778919980	16643291162
5	CZ	1,86669E+11	14941686282	3407757981	15105508273	15001772990
6	JC	12545973452	1028355373	245810090,7	1021914155	1025558951
7	JM	28984243525	2326577271	510314392,7	2344974718	2392605170
8	Pra	48441328605	3841039751	863532636,6	3955653846	3863906122
9	SC	15378665205	1235800160	280263242,9	1255454260	1240647418
10	SM	33897107978	2653175398	624287697,7	2669725848	2700855916
11	StC	17262986295	1370389948	326553820,5	1395682726	1368242710
12	VC	17213870485	1431624102	328181532,6	1401407611	1348308696
13	ZC	12945076290	1054724280	228814567,6	1060695108	1061648007
14	SK	20549430258	1539200146	354100206,0	1572441707	1541518174

Partneri podle regionu - vsichni

Sheet1 Sheet2 Sheet3

Ready

Start Microsoft Excel - Book1 Windows NT Task Manager



FE – OLAP kocka

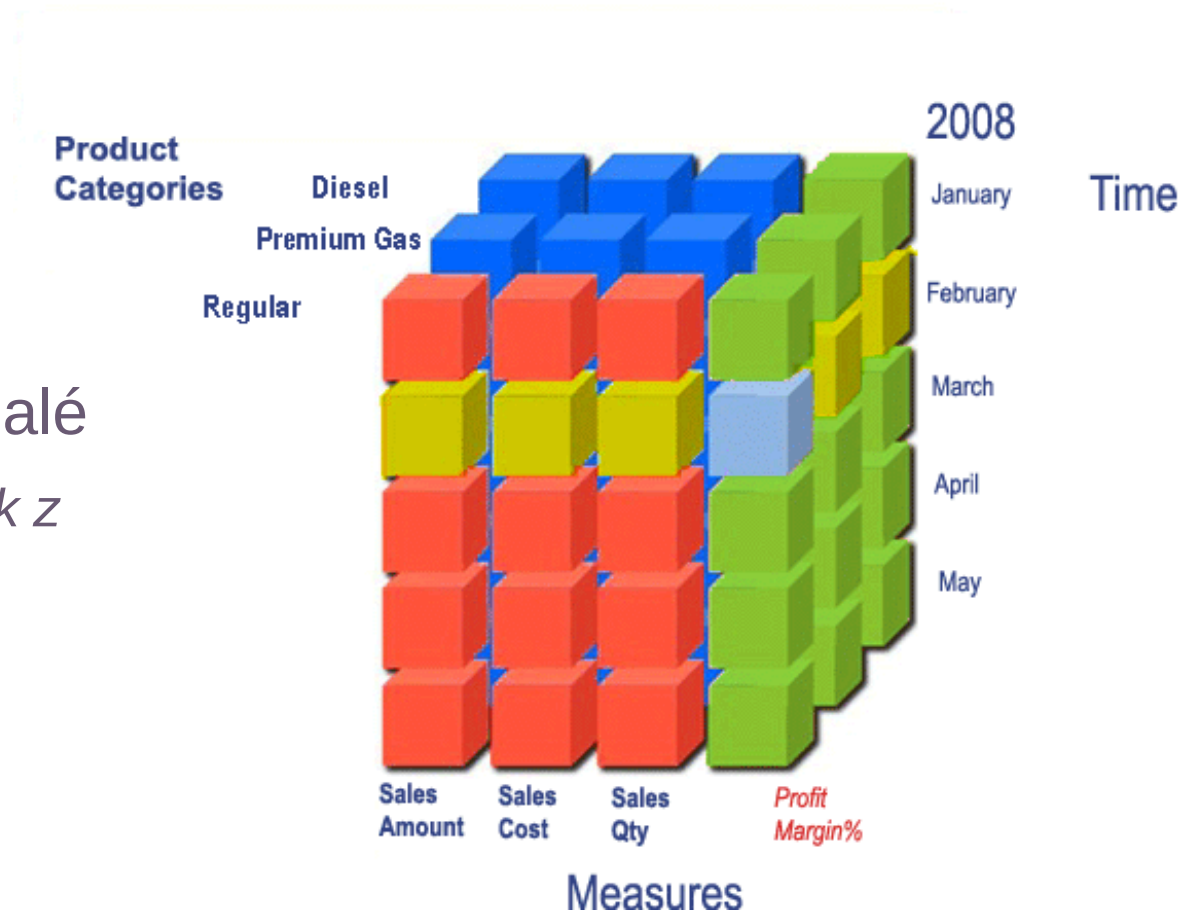
= dátová štruktúra, ktorá umožňuje rýchlu analýzu dát cez viacero dimenzií

- Dotazovanie použitím relačných databáz je pomalé
- Napr. Aký je percentuálny zisk z predaja benzínu „Regular“ vo februári 2008?

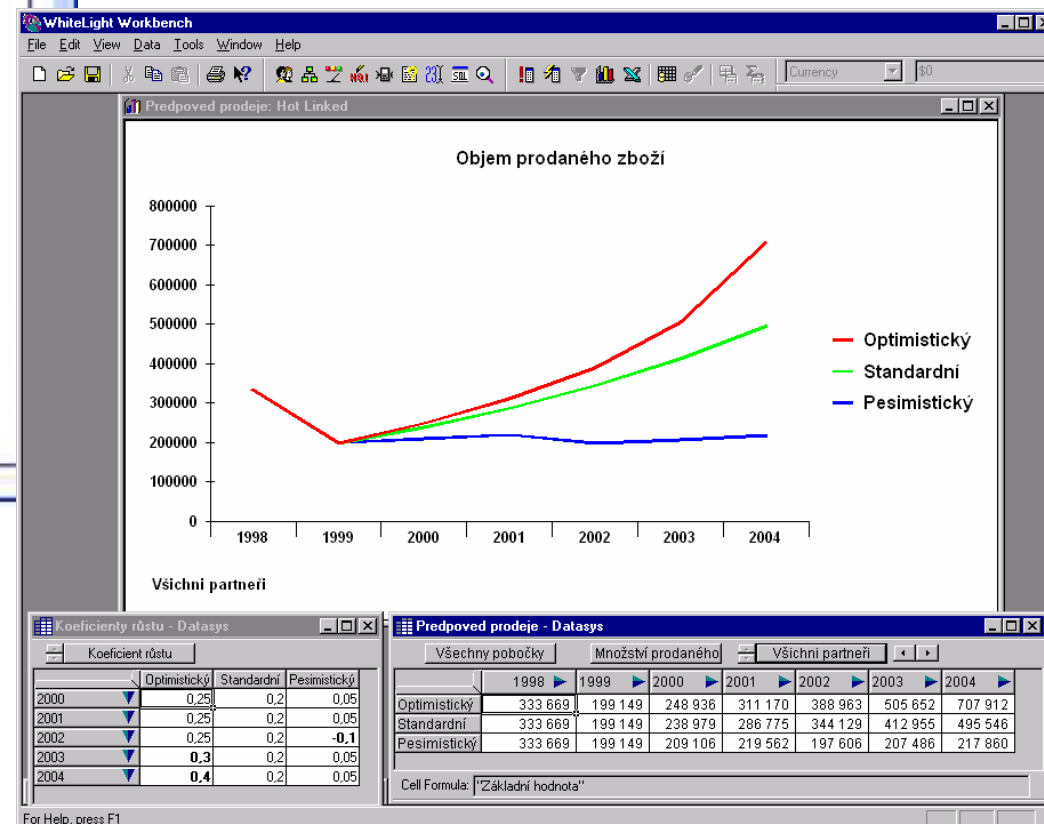
Profit margin – **zelený rez**

„Regular“ – **červený rez**

Február 2008 – **žltý rez**



FE – planning, forecasting



FE – Ad-hoc query

= získavanie informácií podľa potreby

Pre pokročilých používateľov:

- Prístup priamo do databázy DWH
- Vyžaduje znalosti:
 - Dátového modelu DWH
 - SQL

Data mining



NEW
FRONTIER
SLOVAKIA

= špecifická forma ETL

- Získavanie skrytých hodnotných informácií
- Analýza historického správania sa (dát)

⇒ predikcia správania sa v budúcnosti



- Využitie štatistík, rozhodovacích stromov, neurónových sietí, ...
- Napr. identifikácia podvodov v poisťovníctve, risk management v bankovníctve

Diskusia

