



# KIV/SI

Přednáška č. 9

Jan Valdman, Ph.D.  
[jvaldman@dns.cz](mailto:jvaldman@dns.cz)

23.4.2013

# Business Intelligence (BI)



# The Top Challenges of Midsize Companies

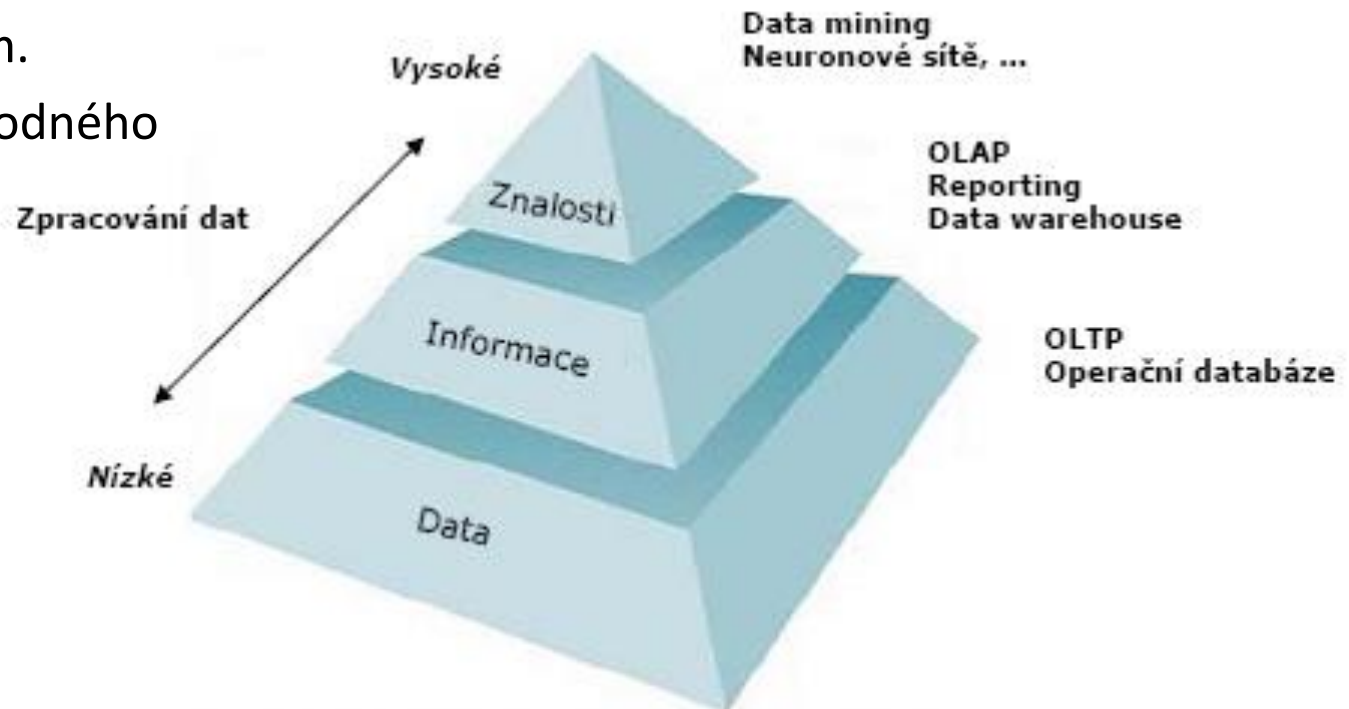
- Improve efficiency, reduce costs
- Strengthen customer relationships, acquisition, sell more
- Predict and respond to trends, optimizing business models
- The highest-priority technology solution chosen by 75 percent of respondents...

**“Turn mountains of data into meaningful insights”**



**Business Intelligence** je soubor nástrojů, technologií a metod, které umožňují nalézt v datech informace (znalosti). Cílem je získávat podklady pro kvalitnější rozhodování (decision support).

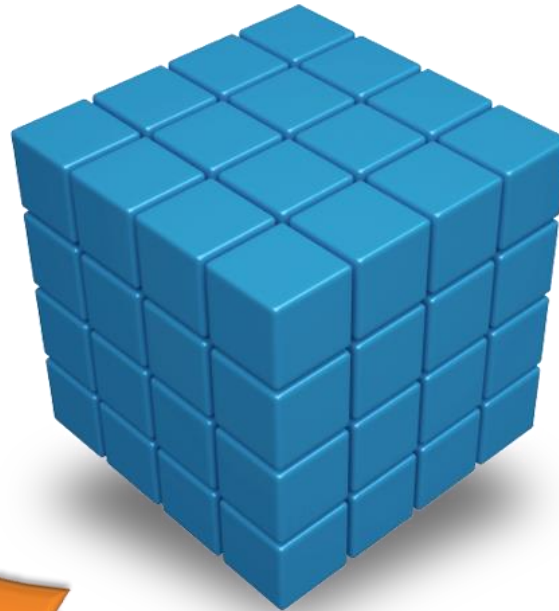
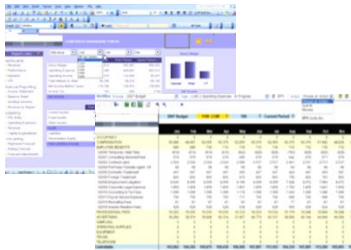
1. Mít data.
2. Vědět, že mám data.
3. Vědět, kde mám data.
4. Mít přístup k datům.
5. Mít data z důvěryhodného zdroje!



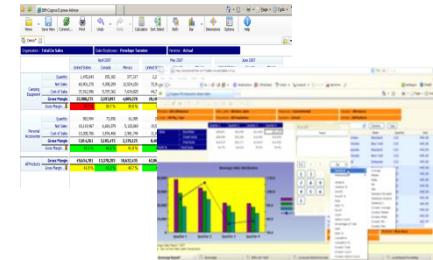
# Dashboards/Reports



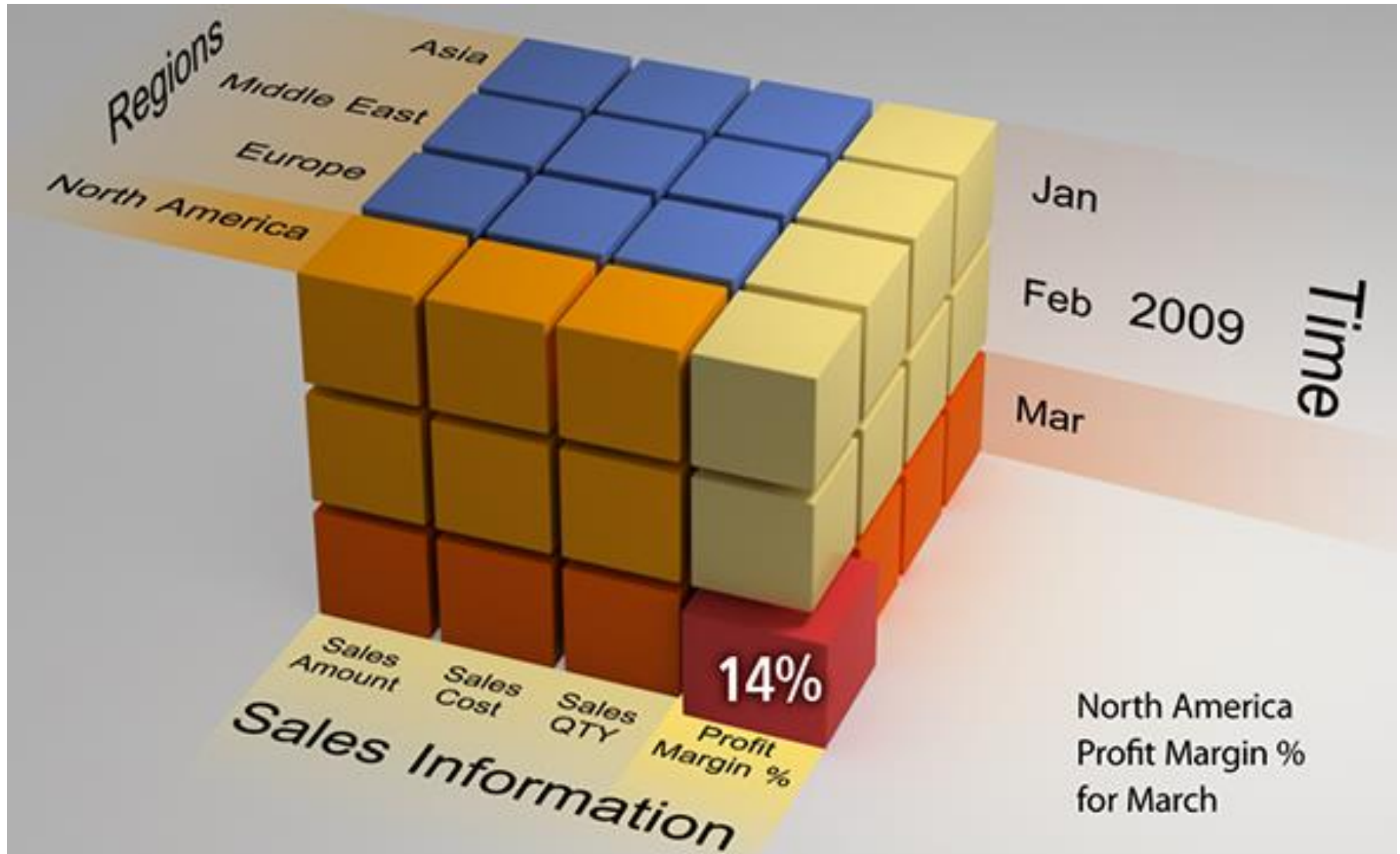
**Planning**

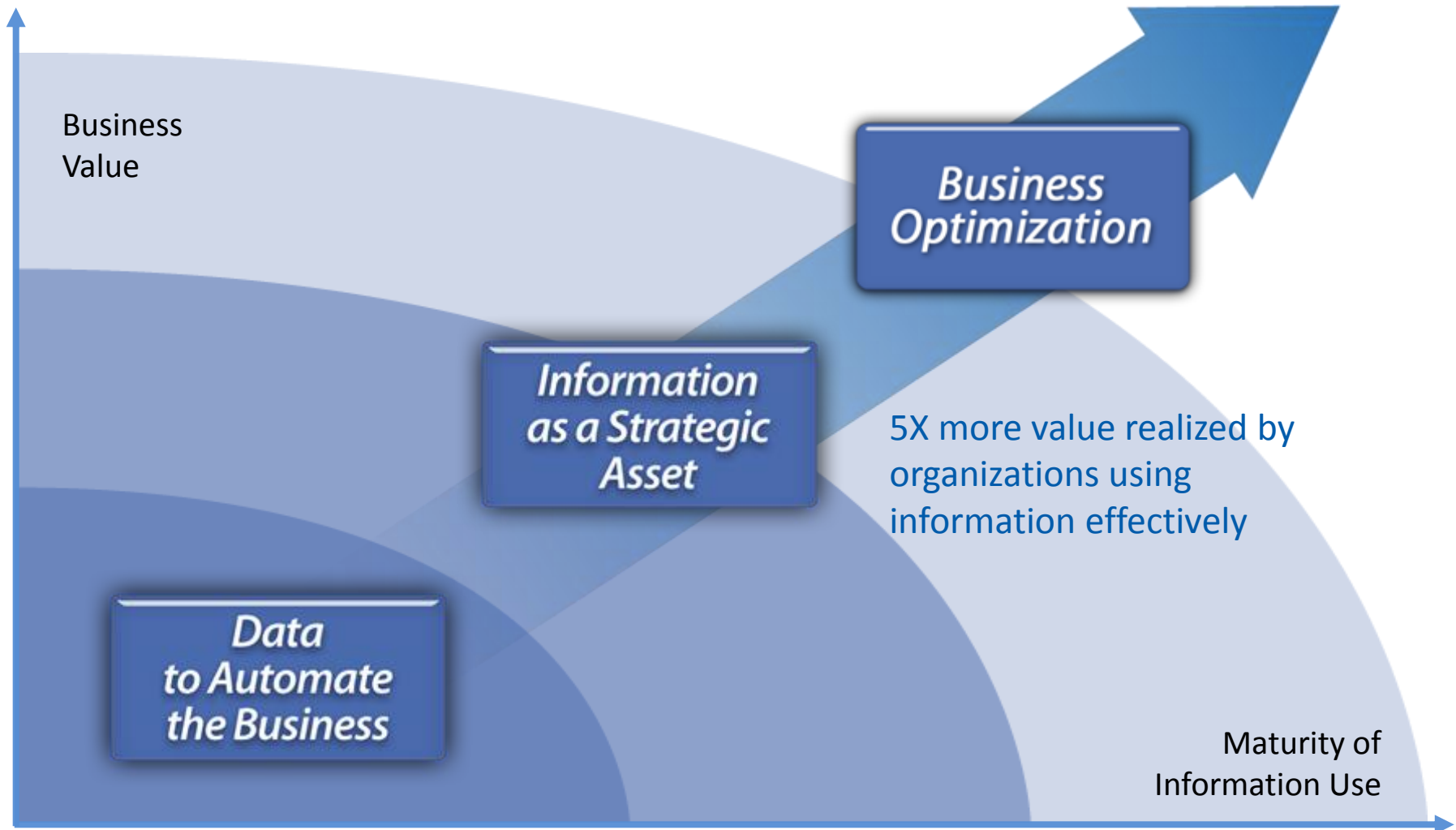


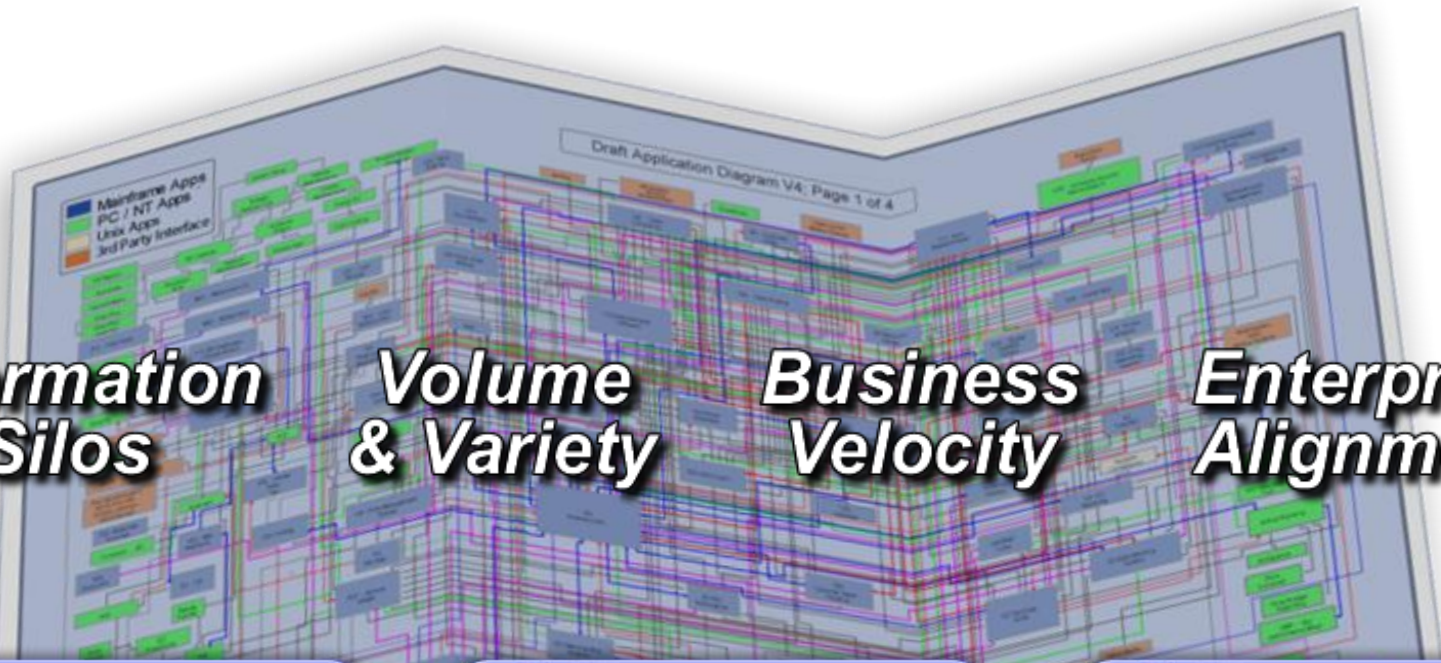
**Analysis**



# Datová analytická multidimenzionální kostka







**Information Silos**

**Volume & Variety**

**Business Velocity**

**Enterprise Alignment**

52% of users don't have confidence in their information<sup>1</sup>

59% of managers miss information they should have used<sup>2</sup>

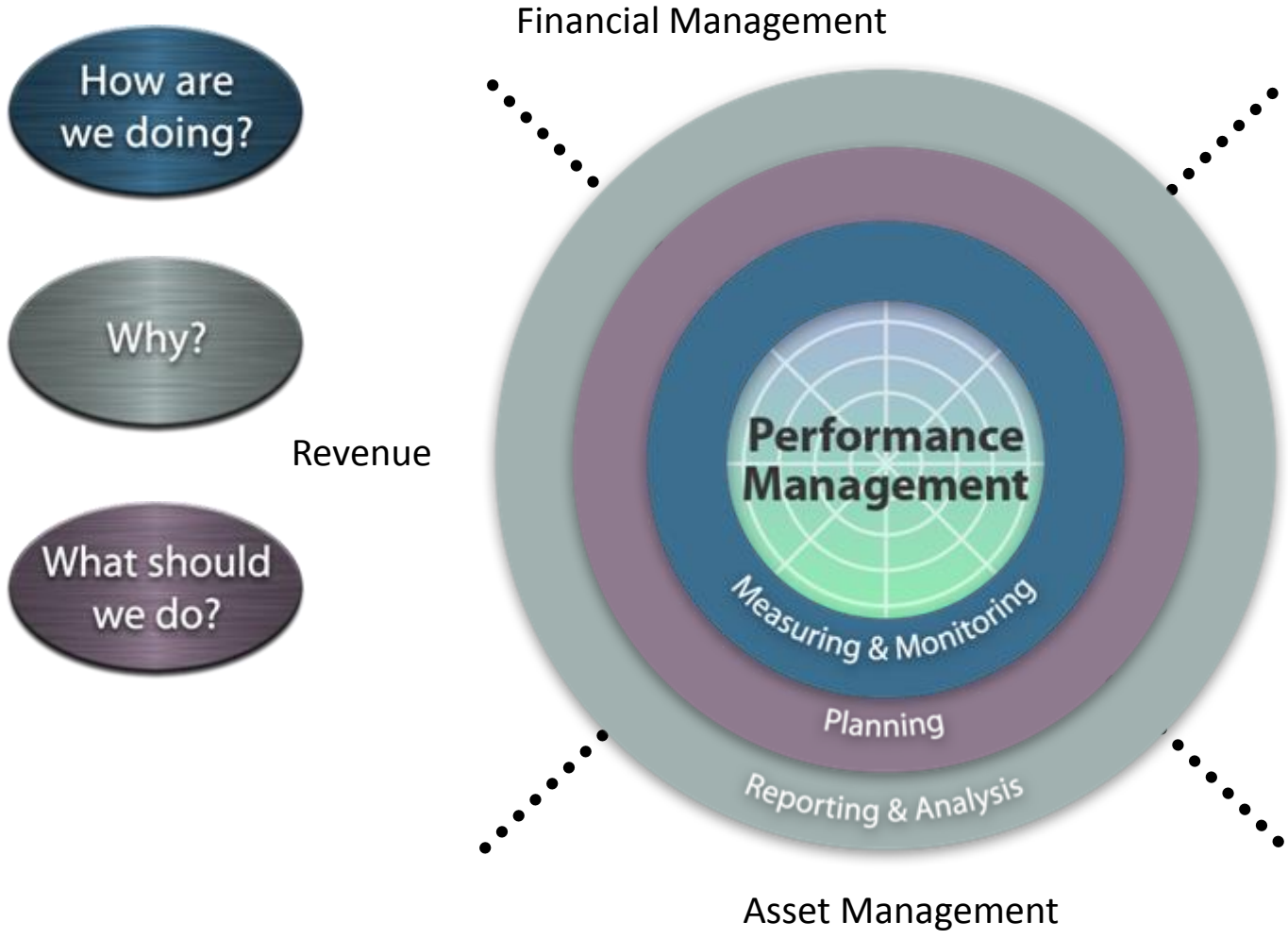
42% of managers use wrong information at least once a week<sup>2</sup>

<sup>1</sup>AIIM 2008 Survey

<sup>2</sup>Accenture 2007 Managers Survey



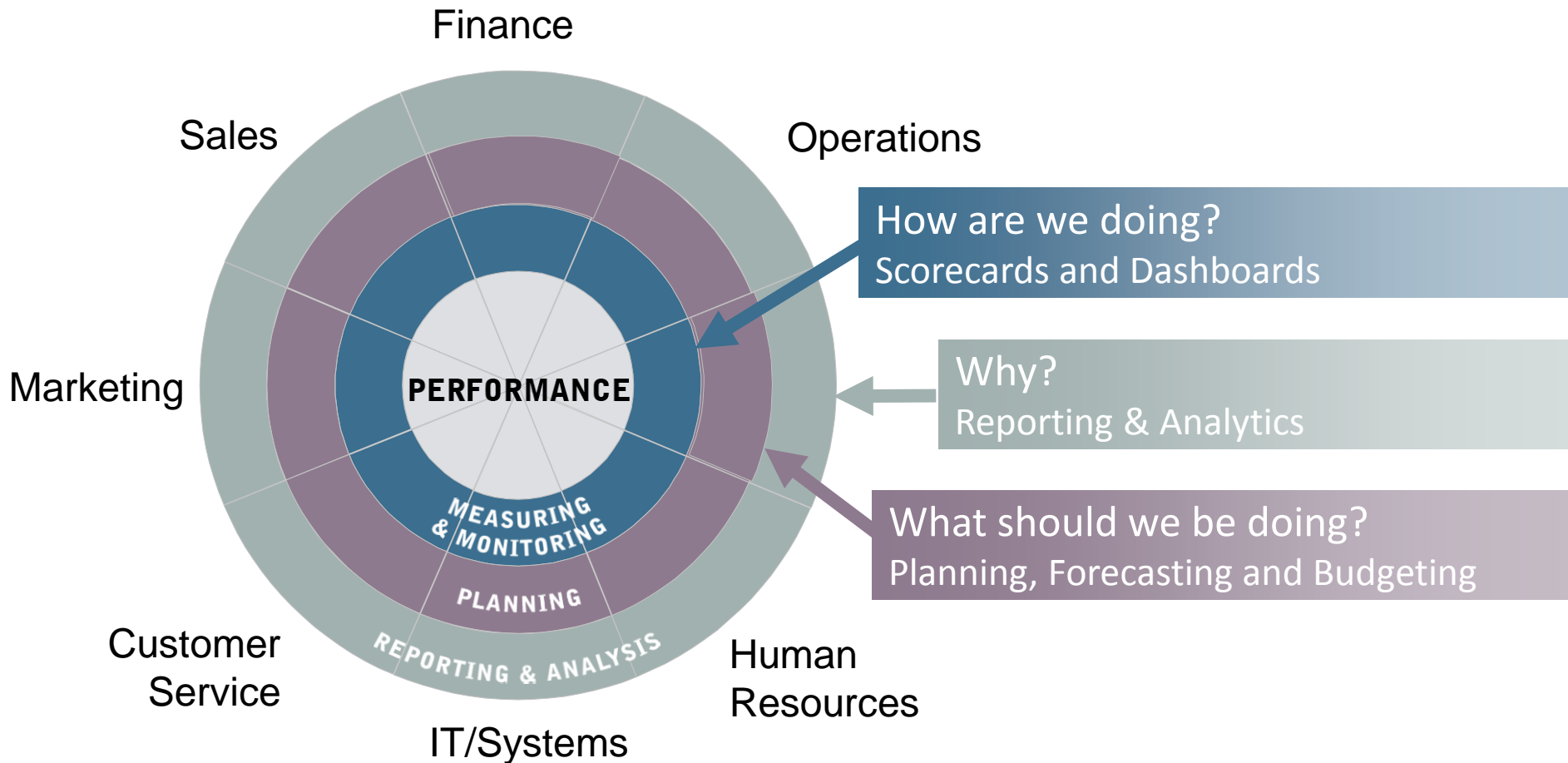
Resolve Issues and Plan for the Future



Act	94.5	102.8	110.6	75
Act	75.5	80.3	86.8	43.9
Admission	44,425	44,005	45,000	1.2
Jan 2	1.2	29.1	13.8	05.6
Populas	74.3	13.9	2.9	2.9
Contributions	13.9	3.4	2.7	2.6
Support (LC)	2.1	2.7	2.7	2.9
Support	2.9	2.8	2.8	0.4
Development	2.1	2.9	2.9	0.4
Development	7.4	12.9	2.8	2
Act Mgt	17.9	2.8	2.8	2
Act Mgt	2.8	2.8	2.8	2
Finance & Admin (Fin)	90.2	85.8	85.8	1
Corporate Office (CO)	8.3	8.3	104.7	1
Total Corporate Expenses	142.5	154.7	154.7	1
Total Other	25.1	14.25	14.25	1
Expenses	17,430	17.3	17.3	1

Expense

Plan is updated to adjust for rising fuel costs and provides understanding of impact to other expenses and profit



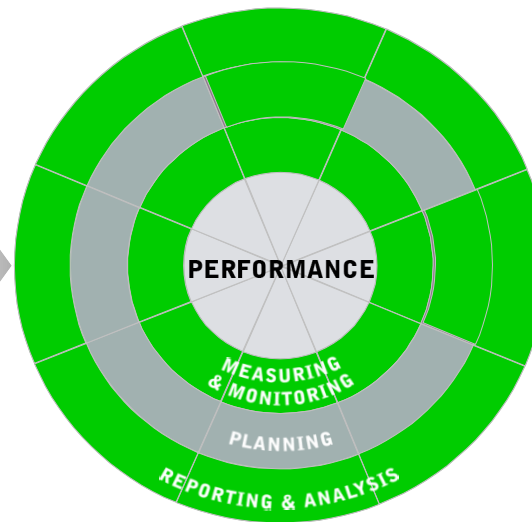
# The Performance Management Journey

## Customer Phases

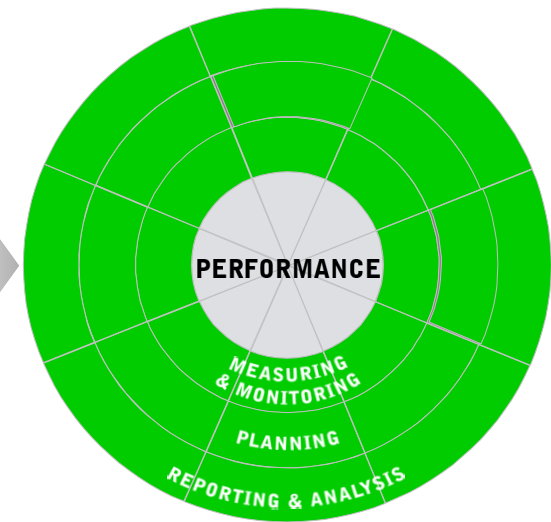
Departmental BI  
Applications



Enterprise BI Capabilities  
Independent Planning

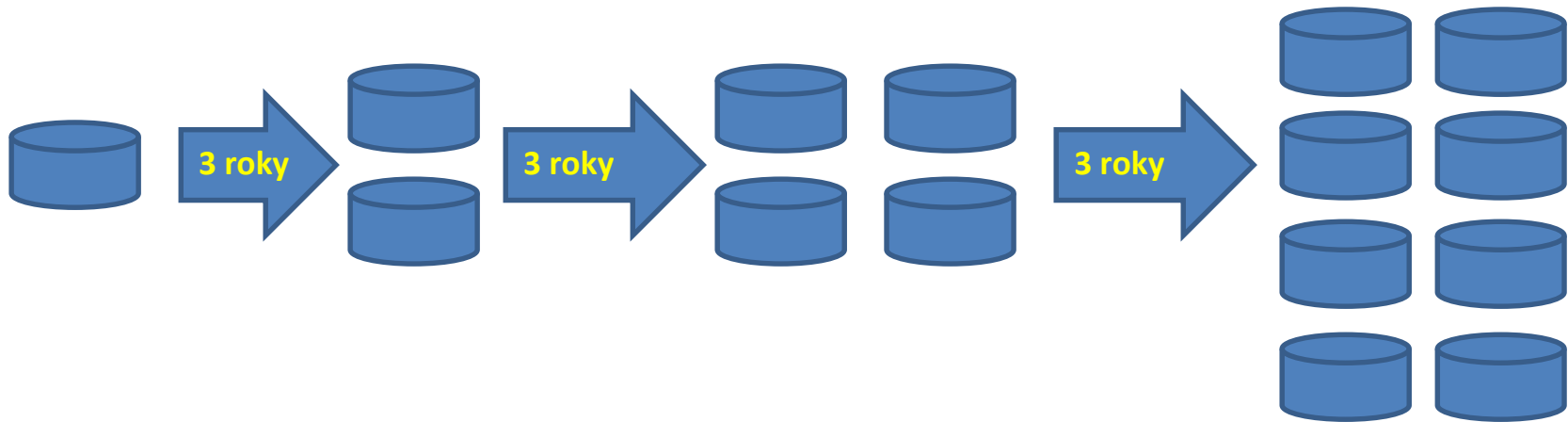


Coordinated Decision  
Making



Common Platform for ALL  
Performance Management needs,  
not a “Suite”

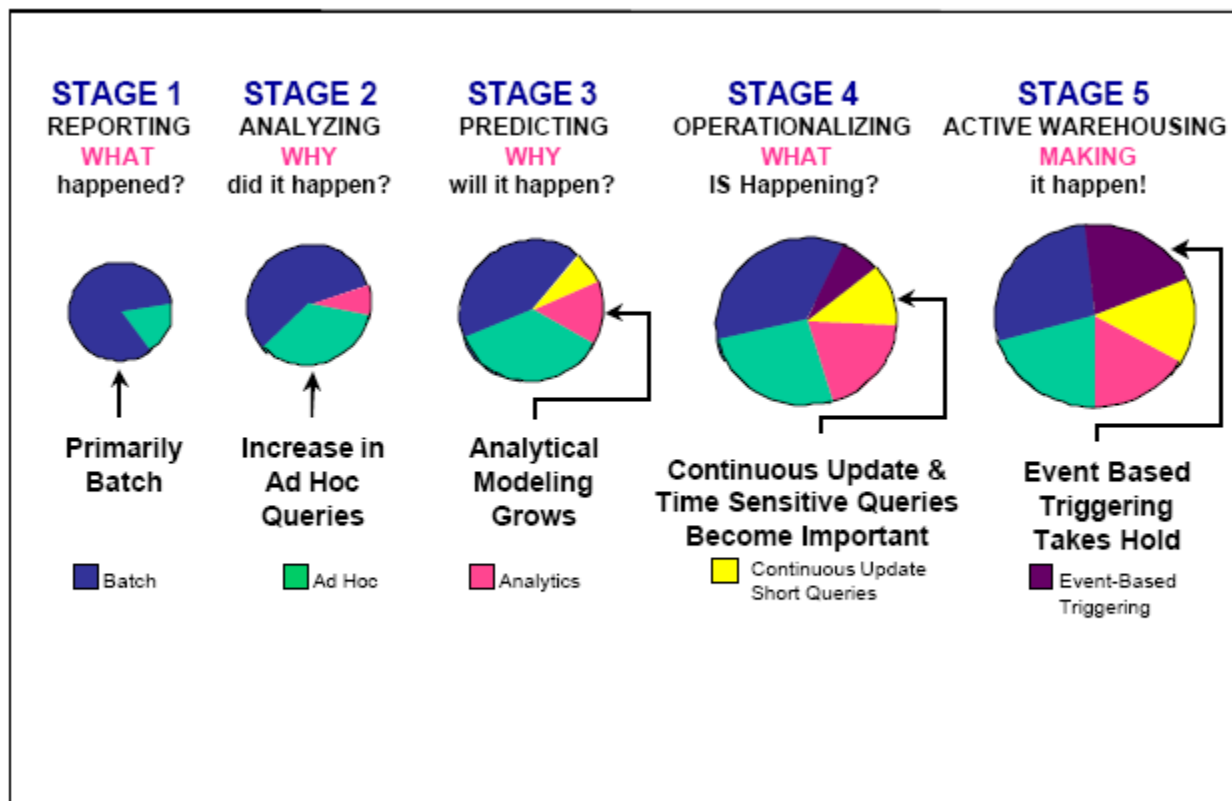
Leading SOA capabilities make BI  
available as a service for easier  
integration



Uvádí se, že každé přibližně **3 roky** se objem dat uložený v databázích **zdvojnásobí**. Tento geometrický nárůst nutně znamená nutnost zkvalitnění procesu zpracování dat.

Podle informací z IBM Infobahn 2009 BI naruby v **téměř 50%** případů uživatelé (manažeři, analytici) pracující s daty:

1. Pracují s nesprávnými nebo neúplnými daty.
2. Nedůvěřují svým datům (a podle 1. zdá se oprávněně).
3. Alespoň 1x týdně činí vědomě (!) rozhodnutí, která z takových dat vycházejí!



Prvotní požadavky na IS se týkaly **hlavně evidence**. Data byla pořizována pokud možno v reálném čase a požadavky na výstupy byly (a jsou) obvykle v podobě:

- Přehledu detailních záznamů (např. přehled prodeje za období).
- Tisku dokladu (faktura, objednávka, dodací list, výpis telefonních hovorů, bankovní výpis, doručenko zásilky apod ...).
- Tisk standardních výkazů (účetní výkazy, přiznání k DPH, atd ...)

Jednotlivé systémy **nebyly (nejsou) vzájemně propojené**. Mnoho informací (např. o zákaznících, produktech, apod ...) bylo vedeno duplicitně.

Byly použity **různé „technologie“**. Kromě IS podporovaných dodavatelskou firmou byly některé systémy vyvíjeny v rámci vlastních zdrojů (obvykle na technologiích FoxPro, Paradox, MS Access apod...) popř. bylo využíváno výstupů do Excelu s následnou úpravou dat.

Pro získání souhrnných informací z více datových zdrojů bylo nutné data transformovat do 1 prostředí (obvykle MS Excel), nejčastěji pomocí různých „automatizovaných“ postupů (maker). Tento proces mohl mít jistou časovou prodlevu a díky možnostem ručního vstupu do dat **nemusí být dostatečně spolehlivý**.

Prvotní požadavky na standardní IS (ERP)

- Chceme vést účetnictví
- Mzdy a personalistiku
- Sklady,
- Odbyt, fakturaci
- Objednávky, ...

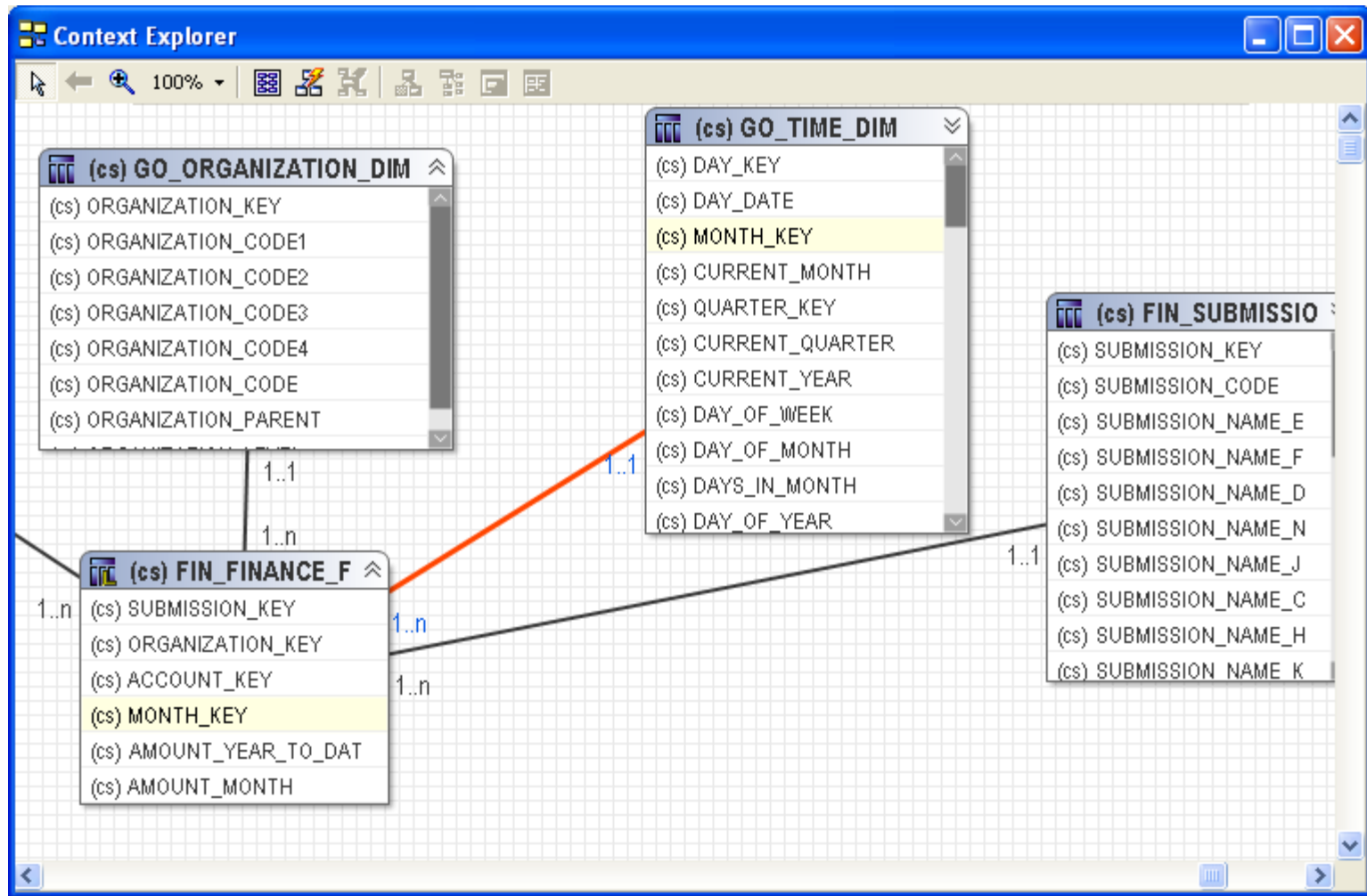
Prvotní požadavky na IS vytvořený na zakázku

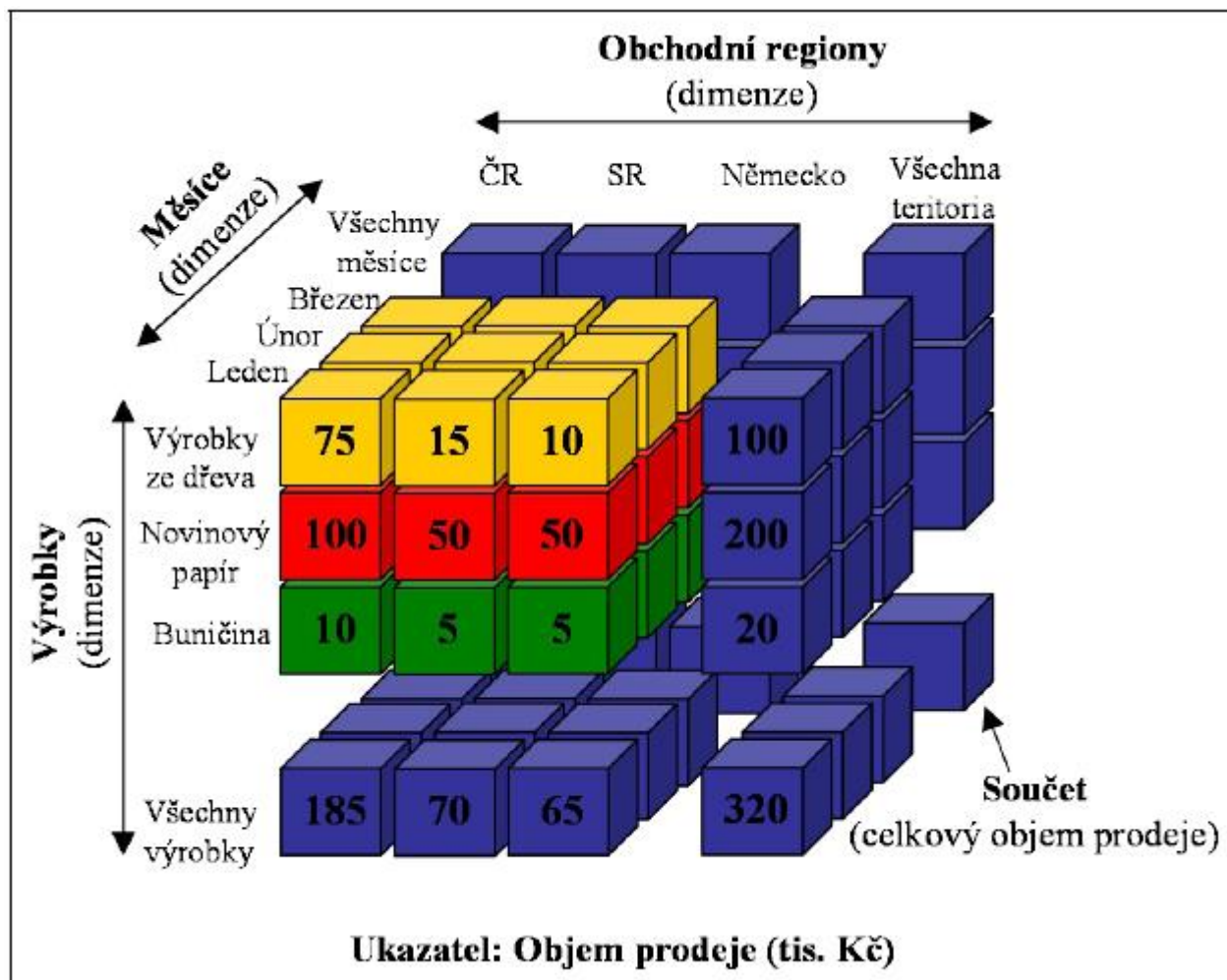
- Evidence dopravy, sledování zásilek online (doprava, logistika)
- Evidence dokladů (např. e-mailů), workflow
- Plánování (výroby, účetnictví, lidských zdrojů, atd ...)
- Sledování hlasových a datových služeb (telekomunikace)
- Výběr peněz z bankomatů, použití platebních karet (banky)

- **ERP** (Enterprise Resource Planning) – kompletní podnikový informační systém (IS)
- **CRM** (Customer Relationship Management) – obvykle součást ERP
- **Relační databáze** (db) – obsahuje tabulky a vazby mezi tabulkami. Používá jazyk SQL.
- **OLTP** (On-Line Transactional Processing) – technologie zpracování dat, typická pro ERP.
- **Multidimenzionální db (mdb)**– obsahuje data uložená do mdb struktur (kostek). Používá obvykle jazyk MDX
- **OLAP** (On-Line Analytical Processing) – technologie zpracování dat typická pro MDB. Existuje relační (ROLAP), multidimenzionální (MOLAP) a hybridní (HOLAP).
- **(De)Normalizace** – způsob optimálního uložení dat v db. Obvykle 3. normální forma.
- **Dimension** (dimenze) – rozměr, podle kterého se sledují data (např. Zákazník, Období)
- **Hierarchy** (hierarchie) – popisuje vztah nadřizenosti mezi prvky dimenze. Dimenze může mít víc hierarchií.
- **Level** (úroveň hierarchie) – hierarchie obvykle obsahuje více úrovní (např. Rok – Měsíc – Den, Rok - Týden).
- **Member** – hodnota na dané úrovni (např. Leden, Únor, ...)
- **Fact** (hodnota) – číselné vyjádření sledovaných hodnot v rámci dimenzí.
- **KPI** (Key Performance Indicator) – klíčový ukazatel.
- **Drill Up / Down** – průchod úrovněmi dané hierarchie, up = slučování, down = zjemnění.
- **Drill Through** – průchod na jinou úroveň (obvykle detailnější), realizován často samostatným reportem.
- **Data Mining** (dolování dat) – získávání skrytých informací a souvislostí z dat.
- **Data Warehouse** (datový sklad) – centrální úložiště dat (další def. dále)
- **Data Mart** – část datového skladu sloužící konkrétnímu účelu nebo určená konkrétní organizační jednotce.
- **Operational Data Store** (ODS) – operační databáze sloužící k okamžitému rozhodování.
- **ETL (ELT)** – proces plnění datového skladu (Extract, Transform, Load).
- **Active Warehouse** – datový sklad pracující (téměř) v reálném čase.



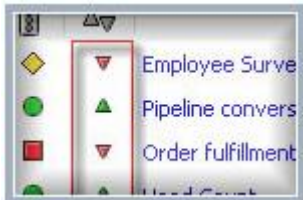
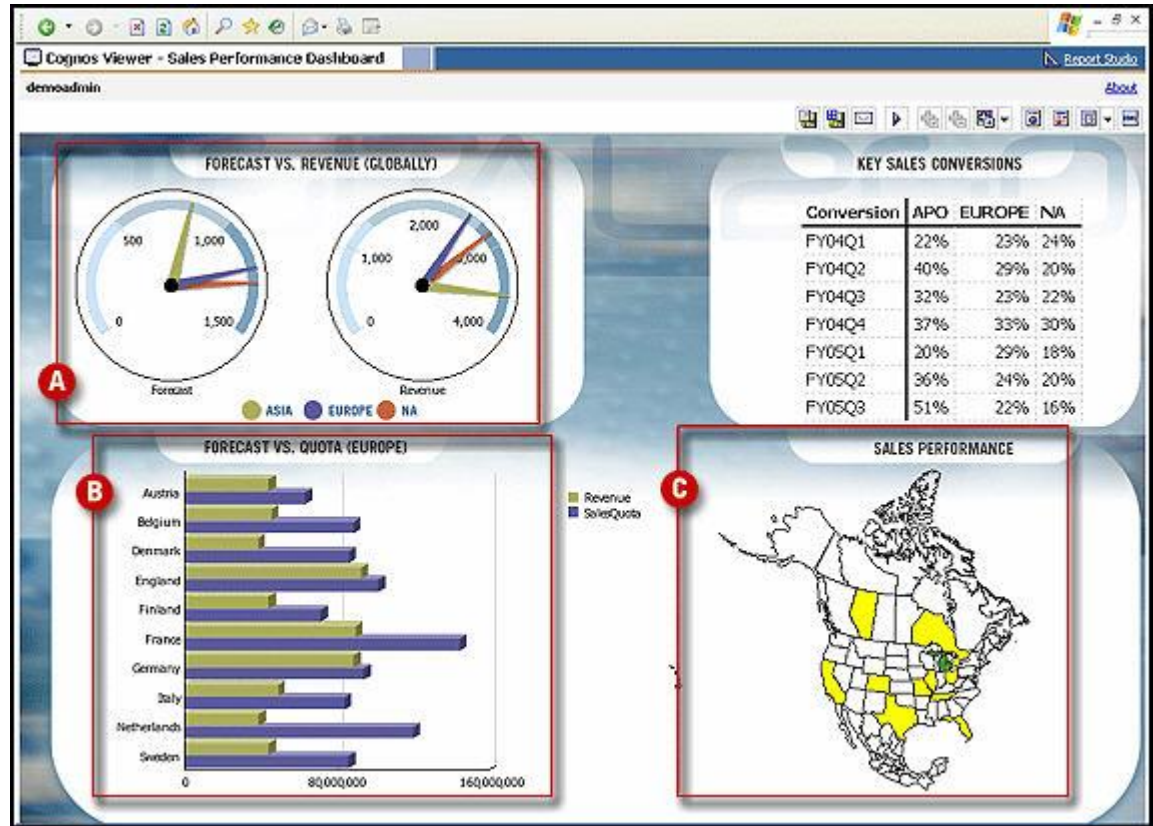
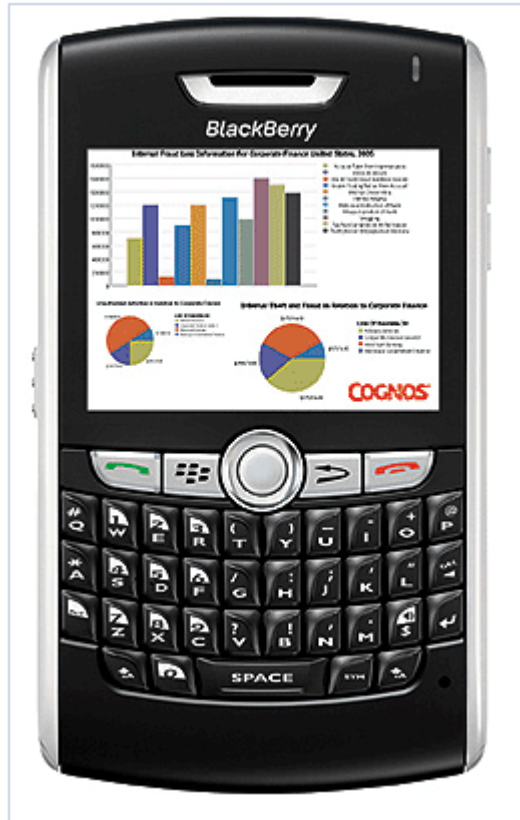
# Relační databáze vs. Analytický slovník





# Reporty, Dashboardy

Here is the Blackberry 8800 Scorecard:



# OLAP Operace

Expected volume	2004	2005	2006	2007	Summary
Camping Equipment	4,740,560	5,542,830	6,844,605	4,835,650	<b>21,963,645</b>
Golf Equipment	864,525	1,019,785	1,209,190	885,950	<b>3,979,450</b>
Mountaineering Equipment		2,249,350	3,184,395	3,090,440	<b>8,524,185</b>
Outdoor Protection	4,740,445	3,540,335	1,367,080	594,935	<b>10,242,795</b>
Personal Accessories	35,001,590	43,126,085	50,967,235	38,386,205	<b>167,481,115</b>
<b>Summary</b>	<b>45,347,120</b>	<b>55,478,385</b>	<b>63,572,505</b>	<b>47,793,180</b>	<b>212,191,190</b>

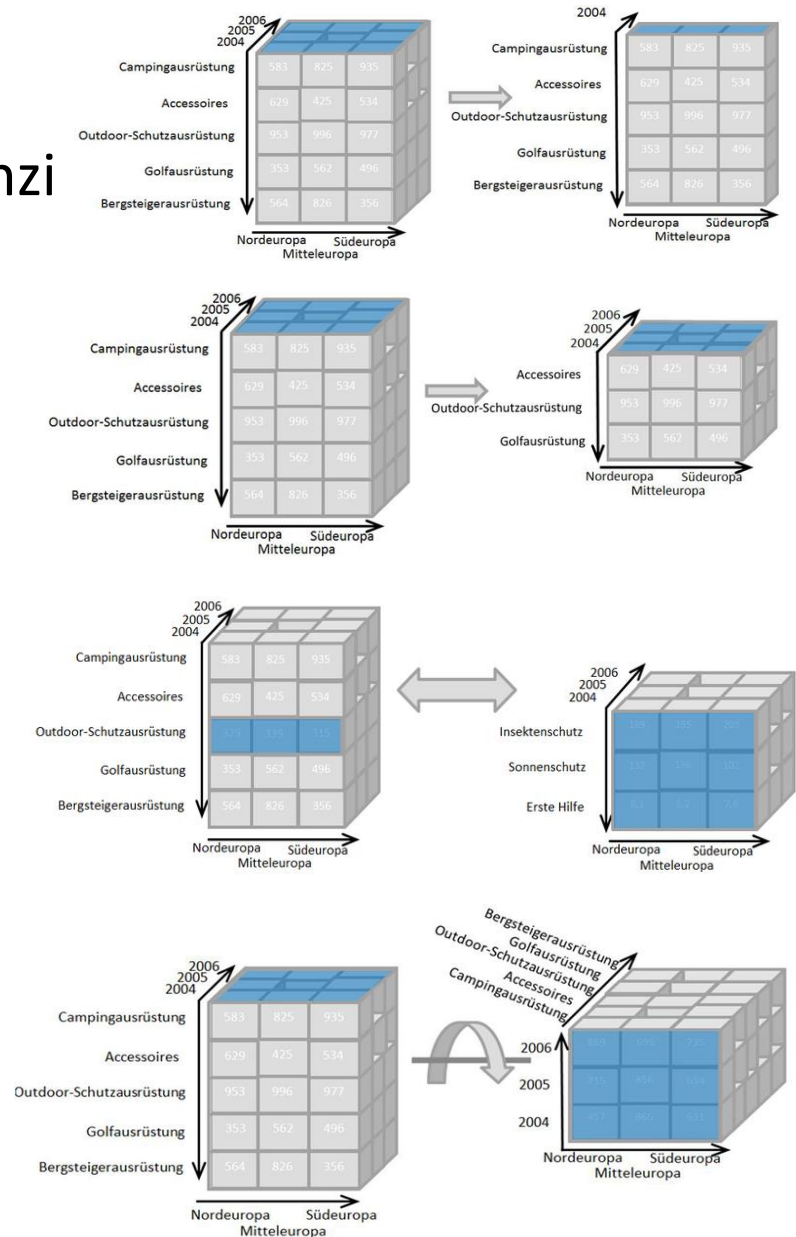
Expected volume	Q1	Q2	Q3	Summary
Camping Equipment	2,080,710	2,050,010	704,930	<b>4,835,650</b>
Golf Equipment	357,270	388,230	140,450	<b>885,950</b>
Mountaineering Equipment	1,302,520	1,330,055	457,865	<b>3,090,440</b>
Outdoor Protection	245,330	257,970	91,635	<b>594,935</b>
Personal Accessories	15,668,620	17,119,740	5,597,845	<b>38,386,205</b>
<b>Summary</b>	<b>19,654,450</b>	<b>21,146,005</b>	<b>6,992,725</b>	<b>47,793,180</b>

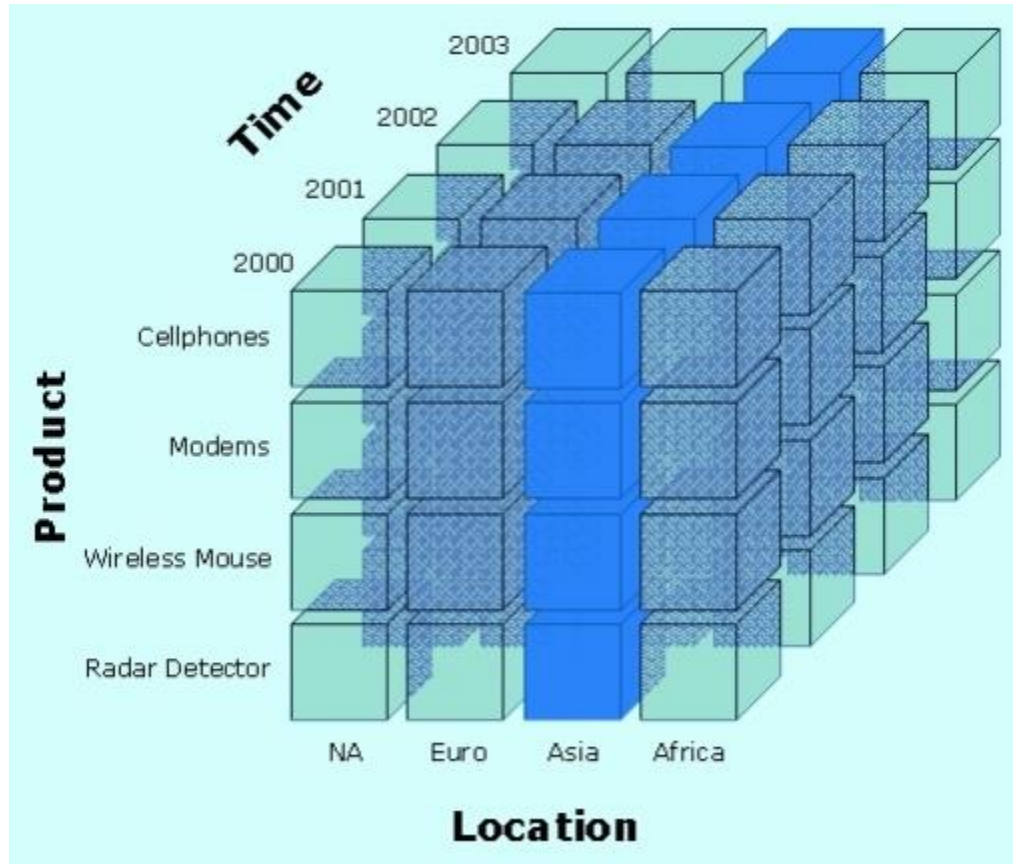


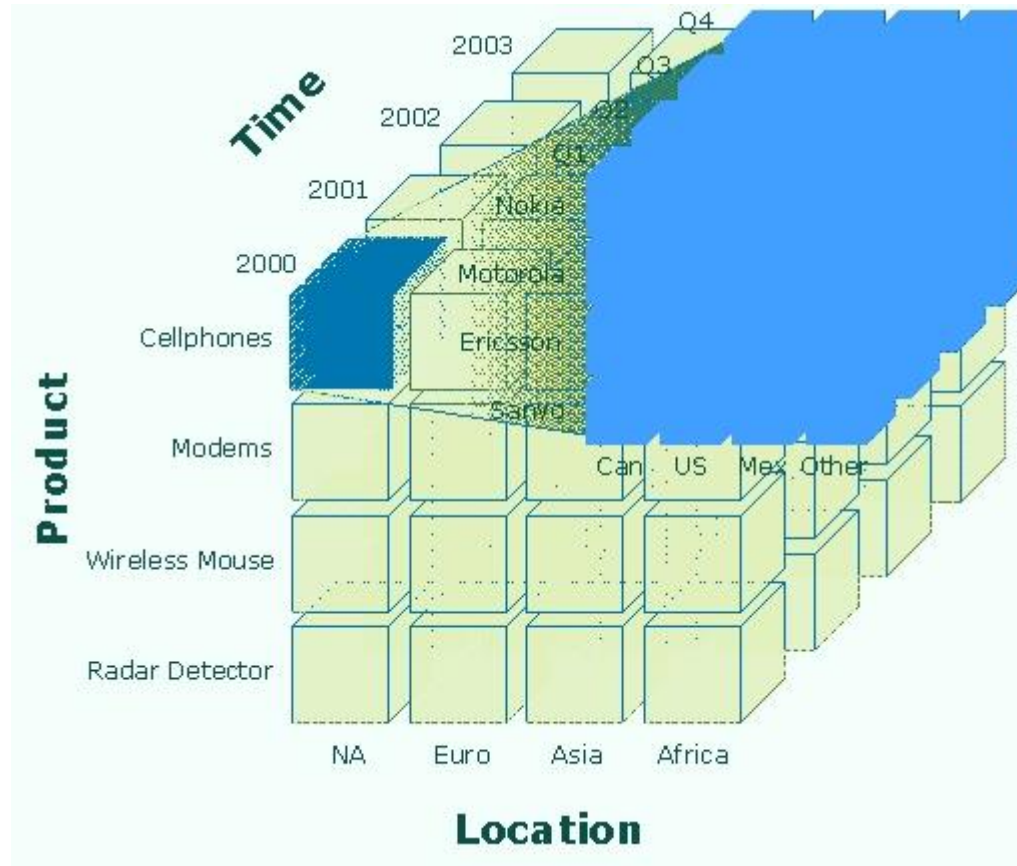
Expected volume	2004	2005	2006	2007	Summary
Cooking Gear	2,260,690	2,665,870	3,162,070	2,218,260	<b>10,306,890</b>
Lanterns	916,450	951,480	1,290,325	897,740	<b>4,055,995</b>
Packs	458,705	564,735	726,490	558,830	<b>2,308,760</b>
Sleeping Bags	539,490	643,515	809,725	577,450	<b>2,570,180</b>
Tents	565,225	717,230	855,995	583,370	<b>2,721,820</b>
<b>Summary</b>	<b>4,740,560</b>	<b>5,542,830</b>	<b>6,844,605</b>	<b>4,835,650</b>	<b>21,963,645</b>

Expected volume	February	January	March	Summary
Camping Equipment	754,950	577,090	748,670	<b>2,080,710</b>
Golf Equipment	124,290	101,950	131,030	<b>357,270</b>
Mountaineering Equipment	465,620	360,850	476,050	<b>1,302,520</b>
Outdoor Protection	83,485	69,045	92,800	<b>245,330</b>
Personal Accessories	5,275,500	4,322,745	6,070,375	<b>15,668,620</b>
<b>Summary</b>	<b>6,703,845</b>	<b>5,431,680</b>	<b>7,518,925</b>	<b>19,654,450</b>

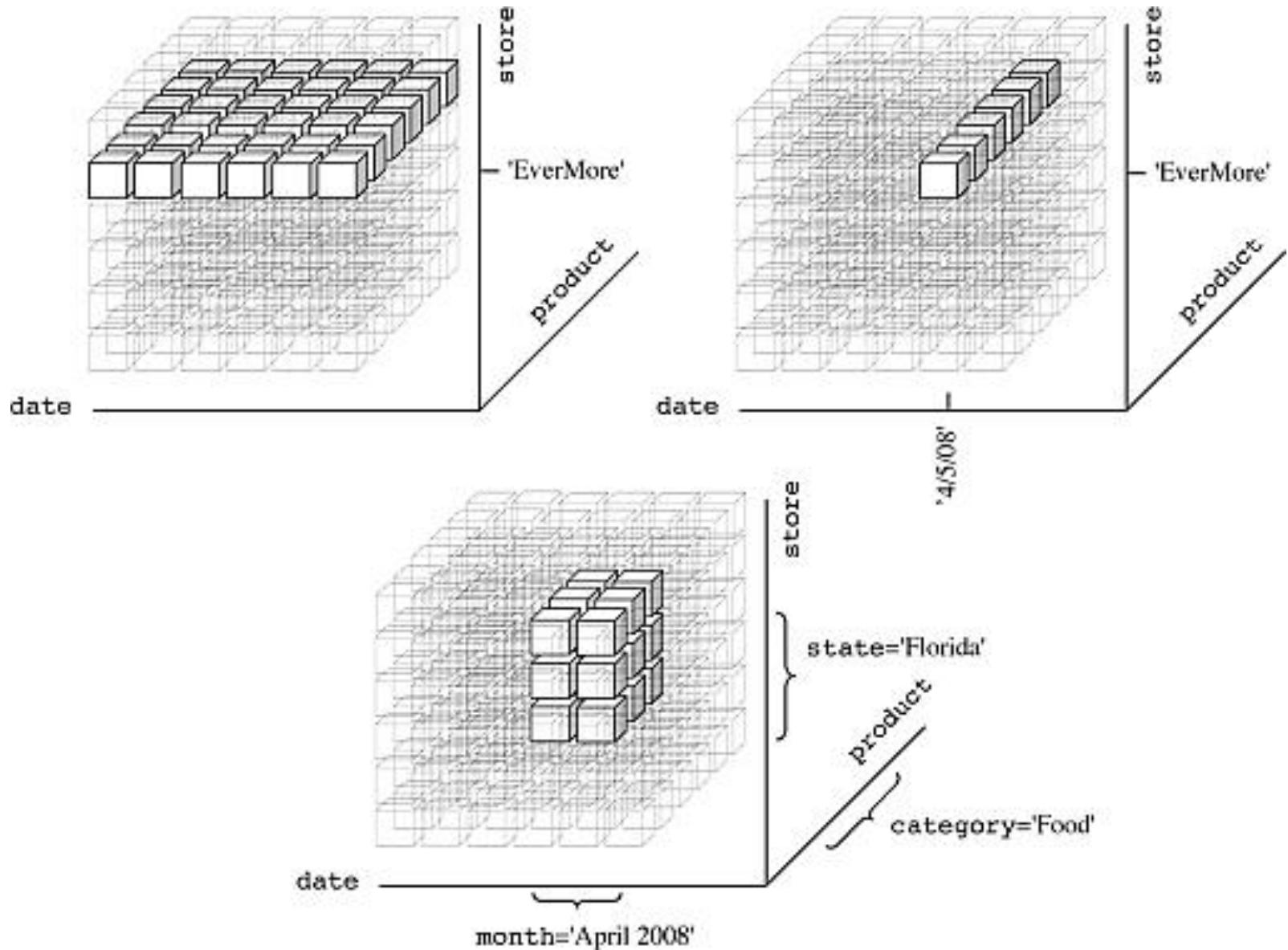
- Slice and dice
  - Omezení počtu kategorií v dimenzi
- Drill-down
  - Detail v rámci dimenze
- Drill-up
  - Sumarizace v rámci dimenze
- Roll-up
  - sumarizace nad všemi daty
- Pivot
  - Otočení os



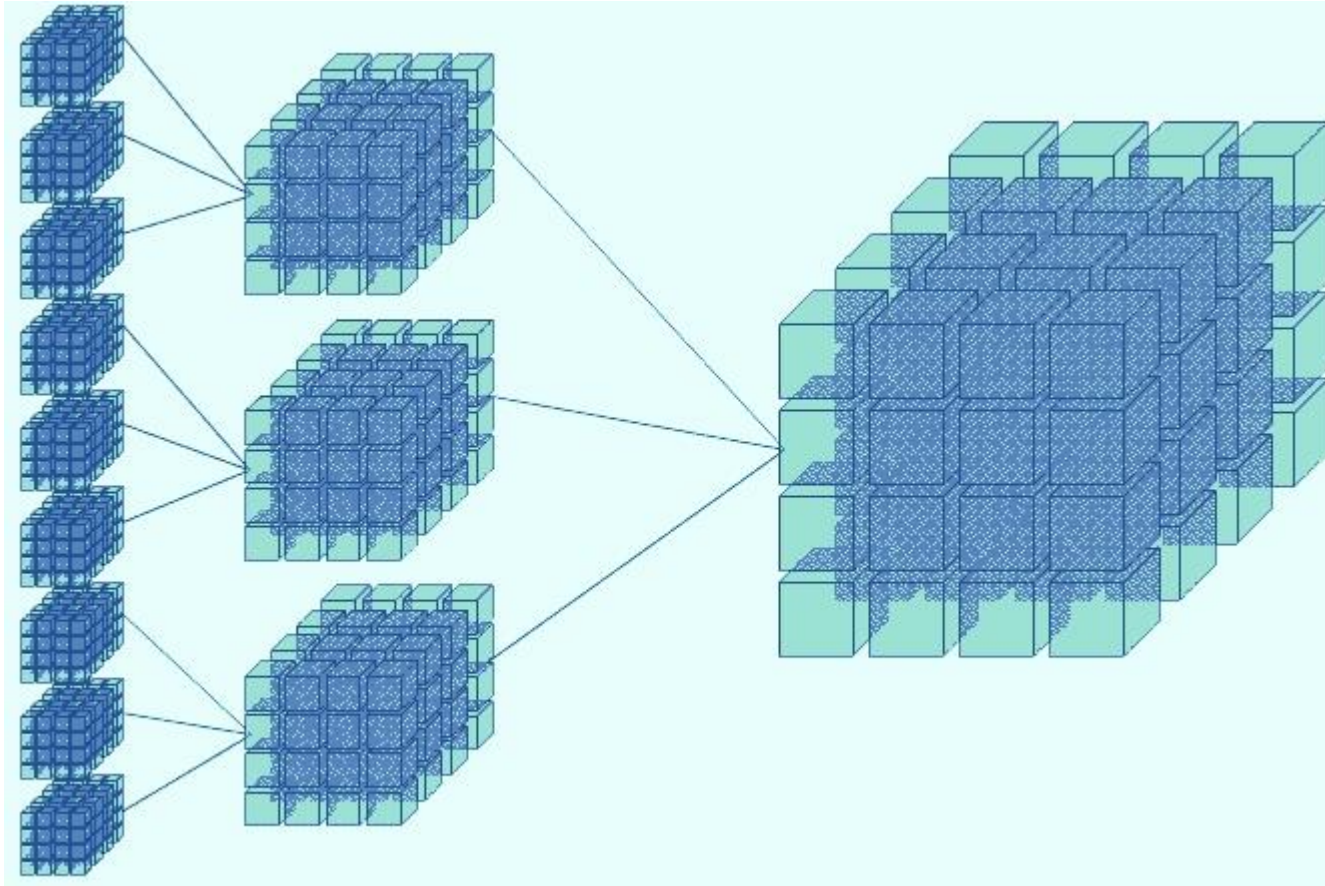


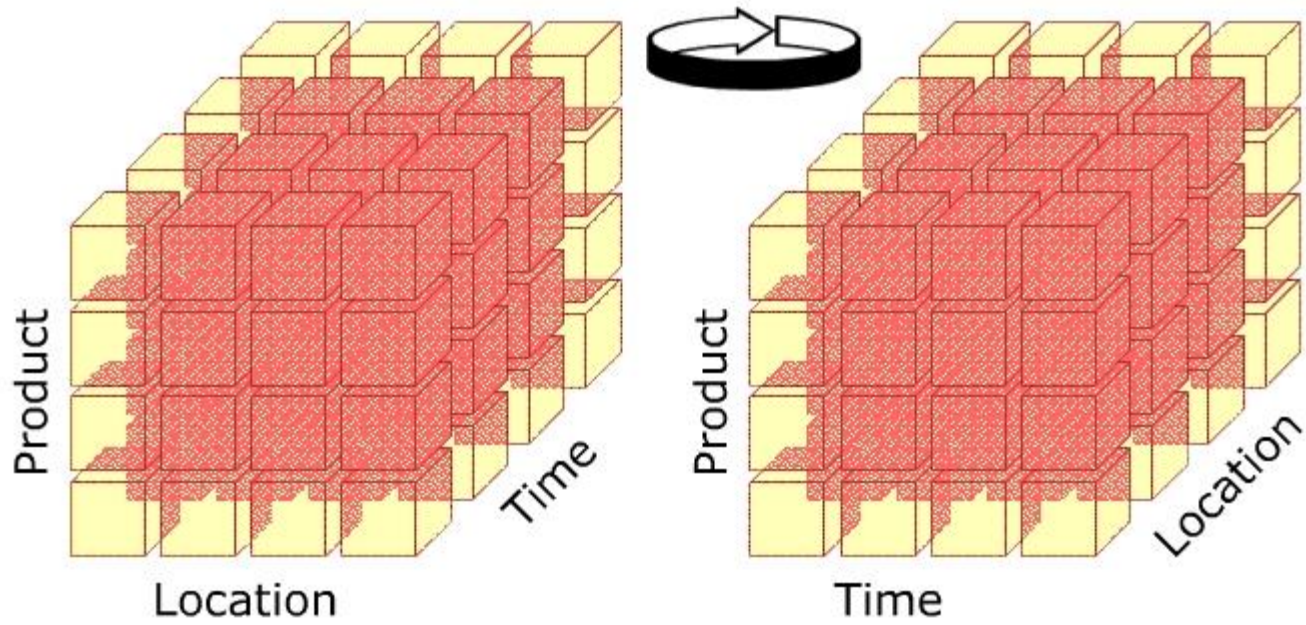


# Drill down



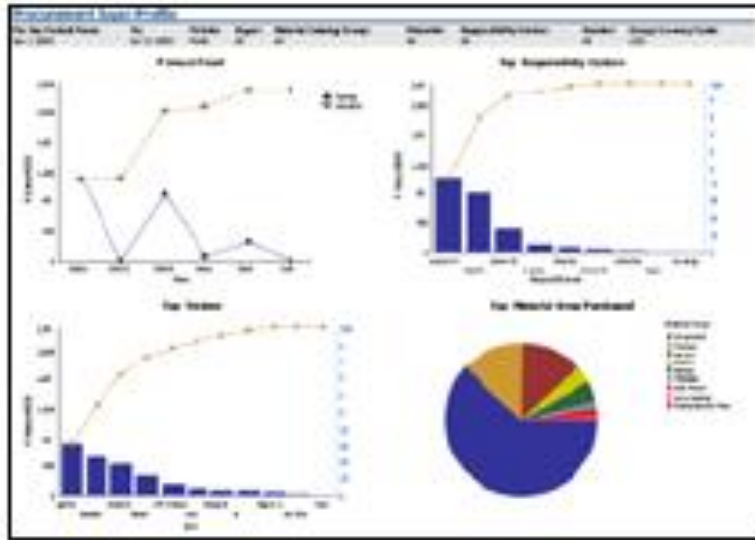






# Reporting, Dashboardy

Figure 5: Sample tactical dashboards using IBM Cognos 8 BI



## Scorecard

Status	Office	Office Manager	Actual	Plan	Difference	% of Plan
●	Miami	Marcus San Antonio	\$210,060	\$220,000	-\$9,940	95.48
●	Dallas	Jane Johannsen	\$269,522	\$354,000	-\$84,478	76.14
●	Atlanta	Helena Krack	\$133,773	\$177,000	-\$43,227	75.58
●	Houston	Robert Roland	\$211,939	\$429,000	-\$217,061	49.4

● = Excellent (Above 90.0%) ● = On Track (Between 70.0 - 90.0%) ● = Warning (Below 70.0%)

## Náhled -> Trend -> Detaily



## Moje alerty

Order Alerts		
Urgent	<a href="#">Shipping of Order 898982 has been delayed</a>	01/11/05 02:47 PM
High	<a href="#">Order 898985 has shipped</a>	01/11/05 02:47 PM
Account Alerts		
High	<a href="#">Credit approval for AMTEC</a>	01/11/05 02:47 PM
High	<a href="#">Cal Industrial has registered for customer webinar</a>	01/11/05 02:47 PM
Company Alerts		
High	<a href="#">Sales meeting moved to Jan 1st</a>	01/11/05 02:47 PM
High	<a href="#">Quarterly expenses due by end of month</a>	01/11/05 02:47 PM

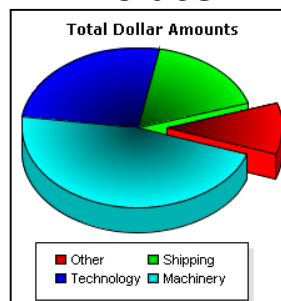
## Zanořené tabulky

Opportunity Name	Rep	Amount	Stage	Committed	Initiated	Forecast Date
<a href="#">T&amp;T Support Contract</a>	Fred Carrier	\$35,500	1	N	12/01/04	03/31/05
<a href="#">AlcaTec server upgrades</a>	Fred Carrier	\$25,500	2	N	09/01/04	03/31/05
<a href="#">AlcaTec.com Next Rev</a>	Fred Carrier	\$25,000	1	N	10/01/04	05/31/05
<a href="#">AlcaTec services contract</a>	Fred Carrier	\$19,000	3	N	10/01/04	03/31/05
<a href="#">T&amp;T New Architecture</a>	Fred Carrier	\$15,000	2	N	10/01/04	03/31/05
<a href="#">T&amp;T Server Upgrades</a>	Fred Carrier	\$13,500	5	Y	09/01/04	12/01/04

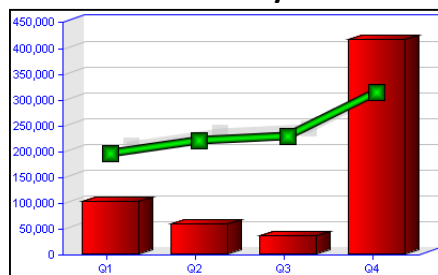
## Indikátory

	2003	2004	Change
YTD Operating Profit	32,202	31,942	-0.81%
YTD ROE	24.94%	23.97%	-3.89%
YTD Sales	1,120,783	1,220,060	8.86%
YTD Gross Margin	258,927	282,078	8.94%
YTD Loss Absorption	14.10%	20.40%	44.68%
Average Company Size	2,700	2,870	6.30%

## Koláče



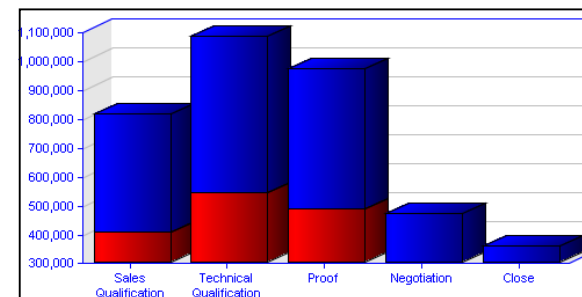
## Trendy



## Filrování dat

Region: Southeast US | Office: Houston | Rep: Fred Carrier

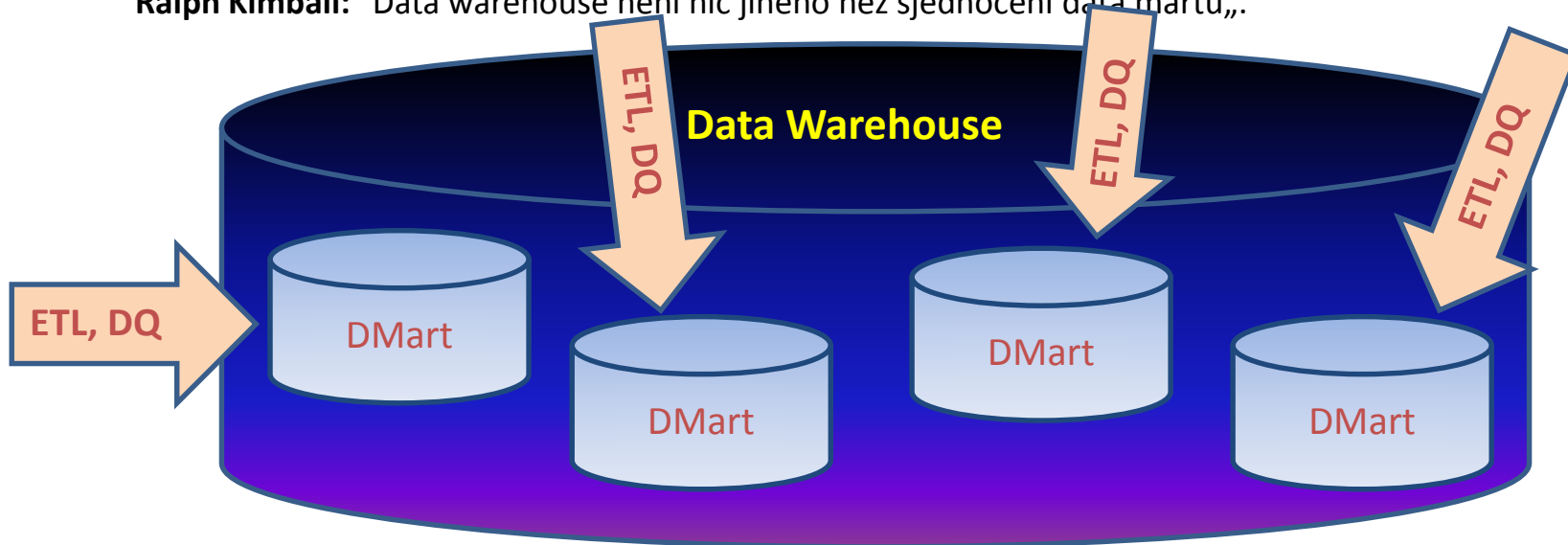
## Graf -> Tabulka



# Dva způsoby budování datového skladu

## 1. Data warehouse jako množina data martů (bus architecture)

Ralph Kimball: "Data warehouse není nic jiného než sjednocení data martů,,.



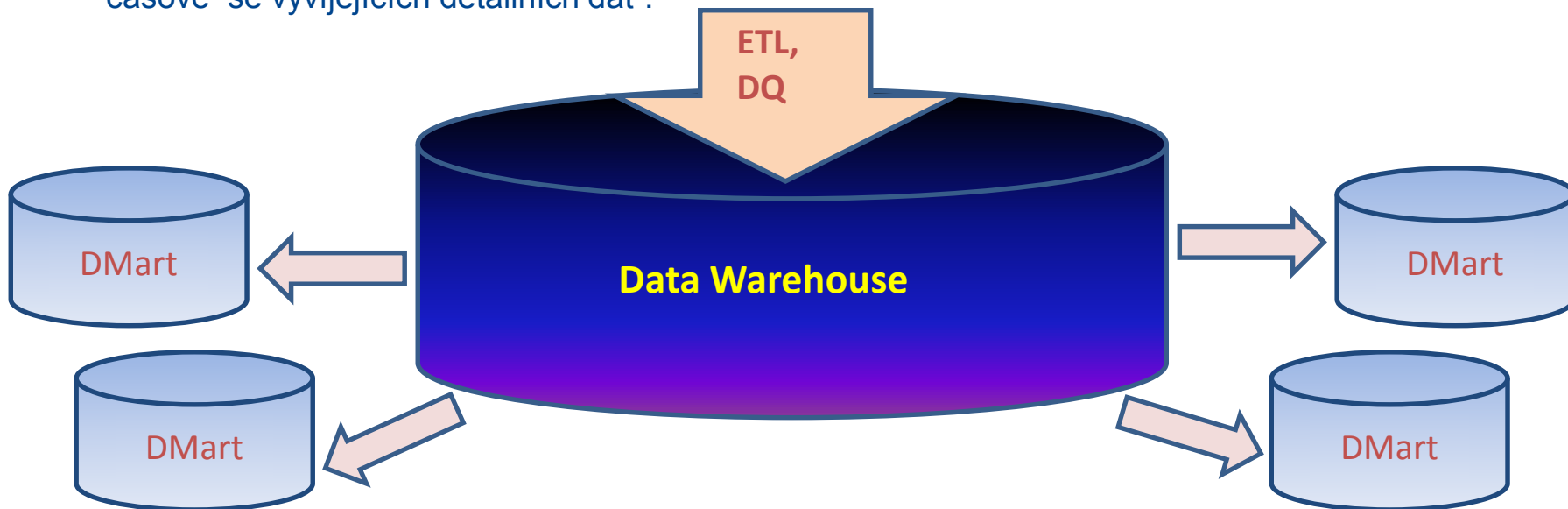
Data mart je možné sjednotit pouze za předpokladu tzv. "všeobecně přijatých" **dimenzí a faktů** (conformed dimension). V opačném případě není možné DM spojovat do 1 celku resp. pokud by se spojovaly, výsledkem budou špatná data!

Plus	Minus
Rychlá implementace data martů.	Redundance dat.
Nízké počáteční náklady.	Každý DM má vlastní historii, ETL, dimenze, řešení datové kvality.
	Hůř monitorovatelné procesy, vyšší HW i SW nároky na údržbu.

# Dva způsoby budování datového skladu

## 2. Centrální Data Warehouse (hub architecture)

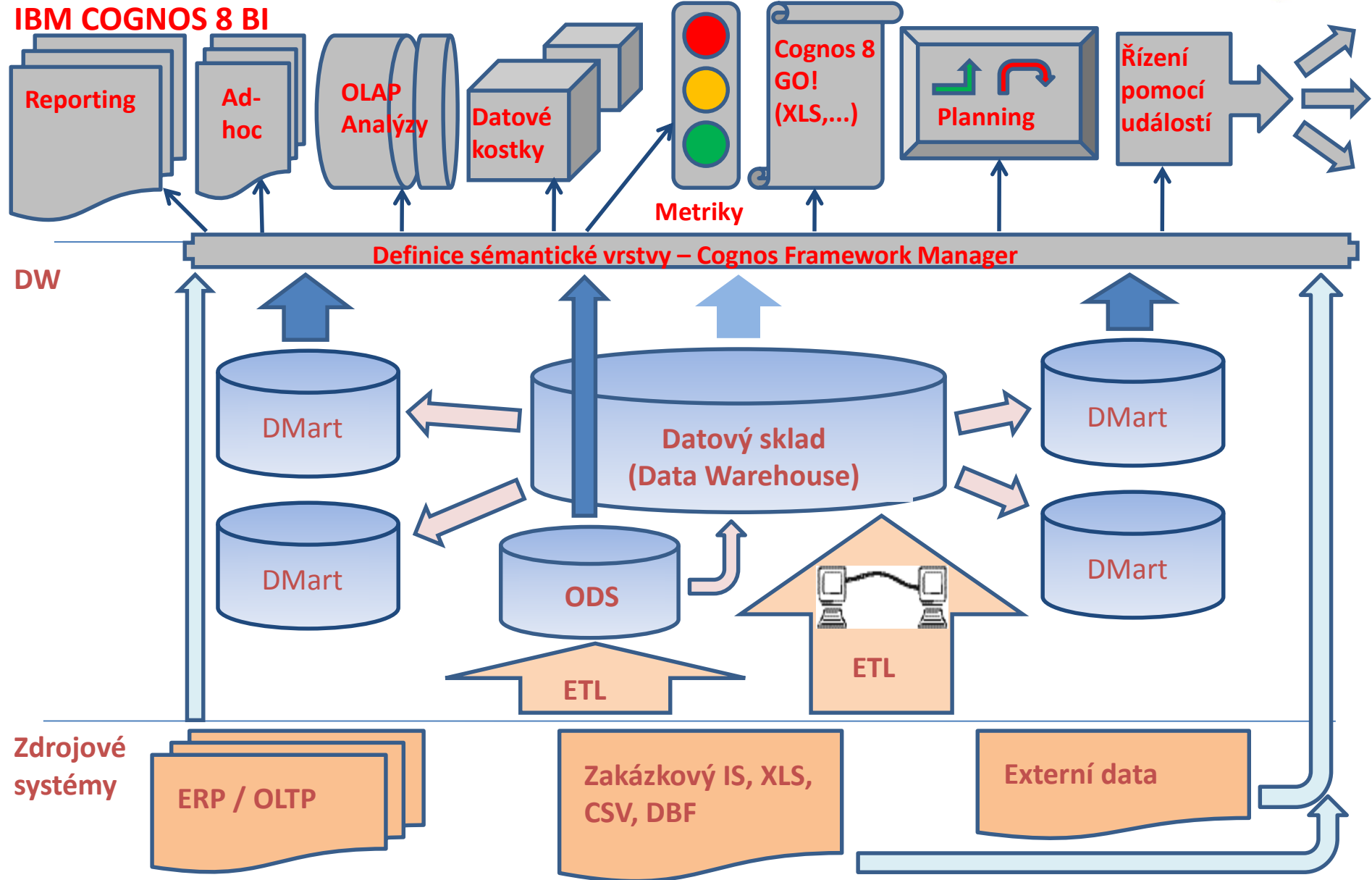
**Bill Inmon:** „(Centrální) datový sklad je soubor integrovaných, předmětově orientovaných, stálých, časově se vyvíjejících detailních dat“.















Plus	Mínus
Centrální datový sklad plněný jednotným ETL postupem, použito 1 řešení datové kvality a vytvořeny společné dimenze.	Vyšší náklady na návrh a implementaci centrálního skladu.
Minimalizována redundantnost dat.	Delší „přípravný“ čas.
Možnost centrálního monitorování datového skladu	

**ODS** je soubor integrovaných, předmětově orientovaných, nestálých aktuálních detailních dat vytvořená pro aktuální potřeby uživatelů.

# Schéma datového skladu a Cognos



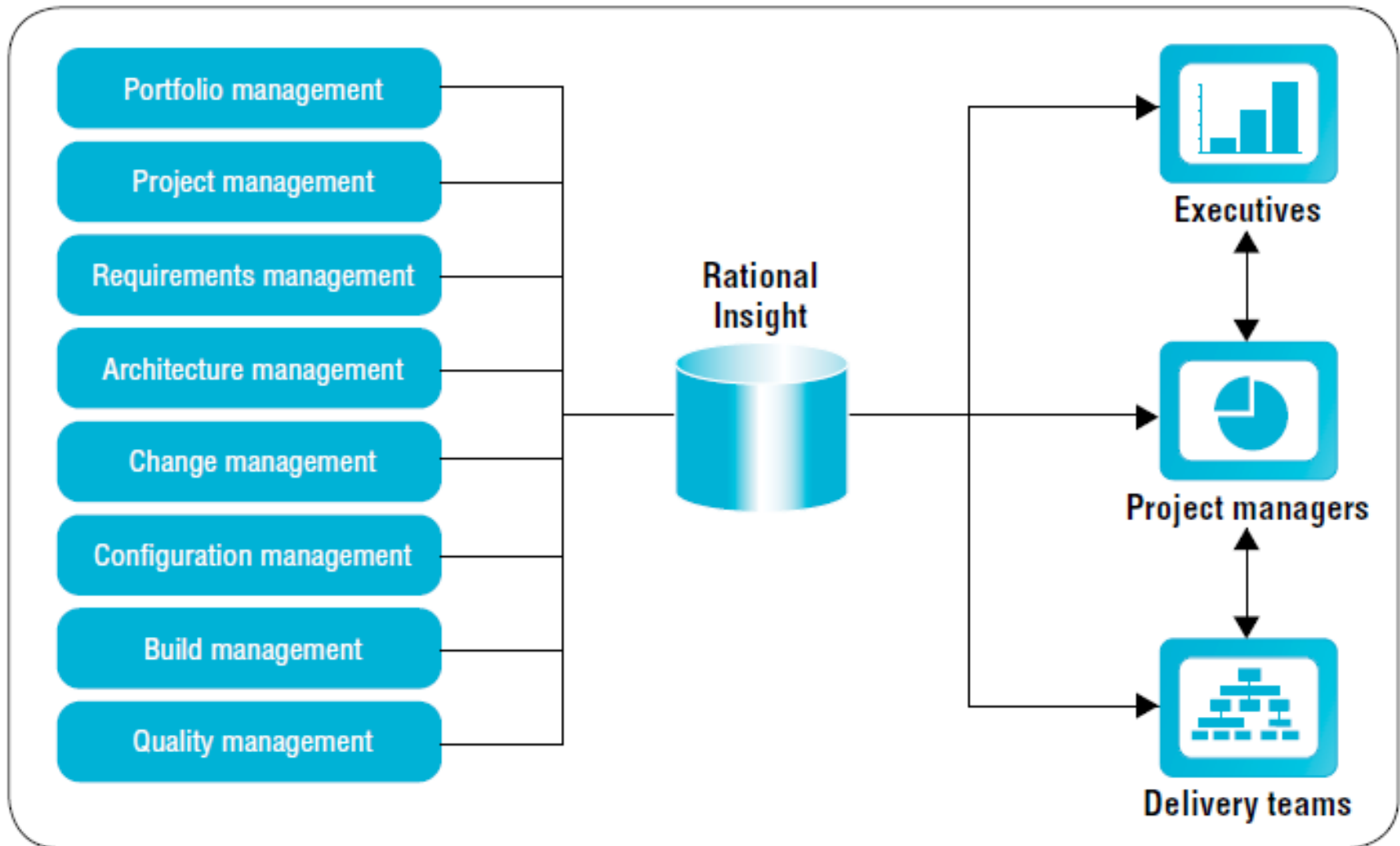
# Příklad Cognos 7 – co vše zahrnuje

Oblast		Cognos	Jiné (IBM) řešení
Budování DW			InfoSphere Warehouse
MDM			InfoSphere MDM
ETL, DQM		Data Manager	InfoSphere DataStage
Správa metadat		<b>Framework Manager</b>	
Reporting		<b>Report Studio</b>	
Ad-Hoc		<b>Query Studio</b>	
OLAP, datové kostky		<b>Analysis Studio, Transformer, PowerPlayer</b>	
Scorecarding		<b>Metric Studio, Metric Designer</b>	
Data Mining		Cognos Scenario	InfoSphere Warehouse Data Mining
Plánování		<b>Planning</b>	
Řízení pomocí událostí		<b>Event Studio</b>	
Napojení na externí nástroje		<b>Cognos GO! Office a GO! Mobile</b>	



# Porovnání transakčních systémů (OLTP) a analytických systémů (OLAP)

Znak	OLTP	OLAP
Charakteristika	Provozní zpracování	Informační zpracování
Orientace	Transakční	Analytická
Uživatel	Běžný uživatel, databázový administrátor	Znalostní pracovník (manažer, analytik)
Funkce	Každodenní operace	Dlouhodobé informační požadavky, podpora rozhodování
Návrh databáze	Entitně-relační základ, aplikačně orientovaný	Hvězda/sněžná vločka, věcná orientace
Data	Současná, zaručeně aktuální	Historická
Sumarizace dat	Základní, vysoká podrobnost dat	Shrnutá, kompaktní
Náhled	Detailní	Shrnutý, multidimenzionální
Jednotky práce	Krátké, jednoduché transakce	Komplexní dotazy
Přístup	Číst, pořizovat a aktualizovat	Pouze číst
Zaměření	Vkládání dat	Získávání informací
Počet dostupných záznamů	Desítky	Miliony
Počet uživatelů	Stovky – tisíce.	Desítky – stovky.
Velikost databáze	100 MB až GB	100 GB až TB
Přednosti	Vysoký výkon, vysoká přístupnost	Vysoká flexibilita, nezávislost koncového uživatele
Míry hodnocení	Propustnost transakcí	Propustnost dotazů a doba odezvy



# Rational Insight

**Filters**

Program:

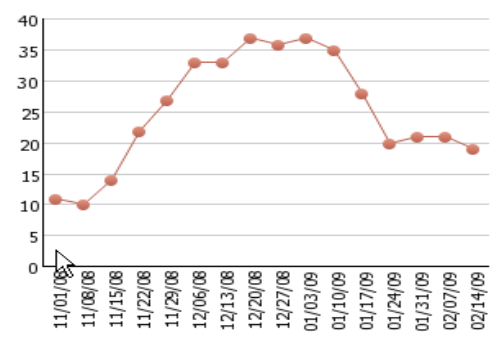
Project:

**Information**

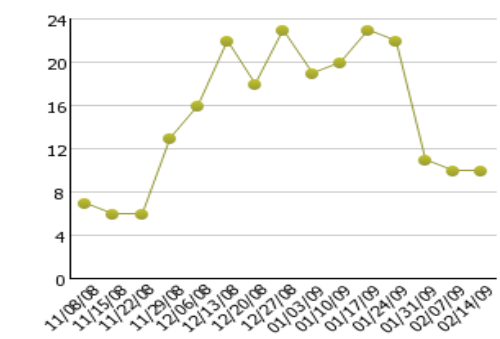
This dashboard is at project level.

Smarter Planet Online Auction

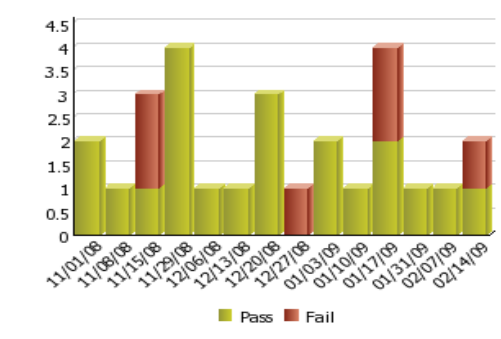
**Outstanding Work**



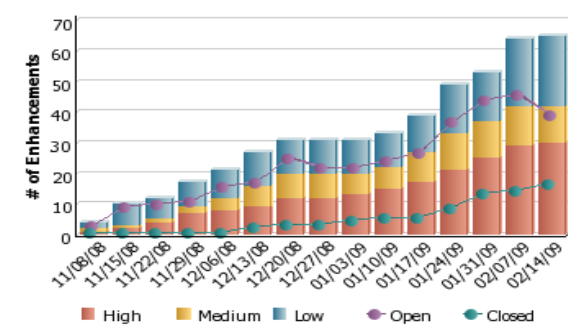
**Work Completed**



**Build Health**



**Enhancement Request Backlog**



**Actual Defect Trends**

