



MASARYKOVA UNIVERZITA

Základy business intelligence

Jaroslav Šmarda

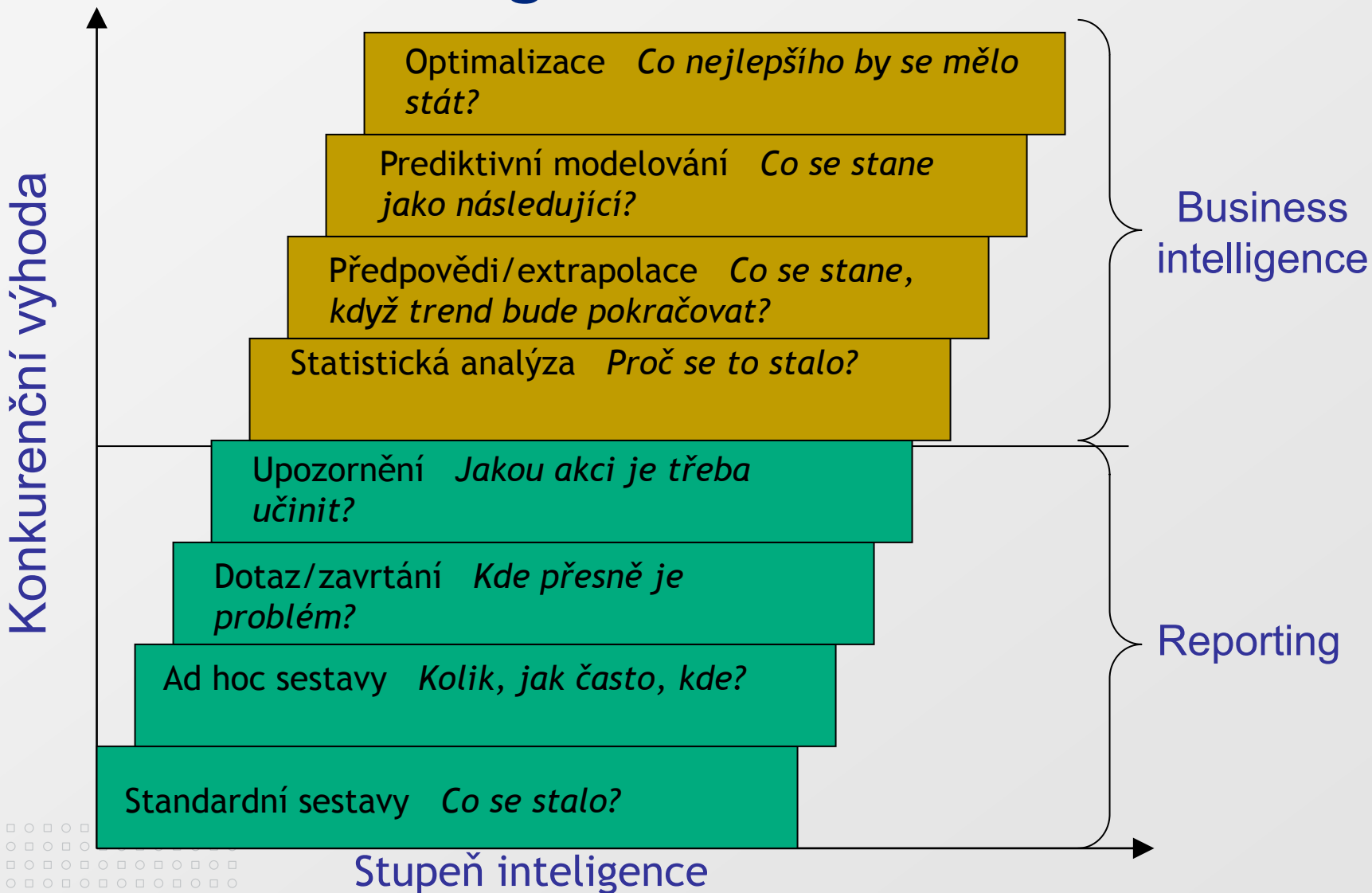
Základy business intelligence

- Business intelligence
- Datový sklad
- On-line Analytical Processing (OLAP)
- Kontingenční tabulky v MS Excelu jako příklad OLAP
- Dolování dat

Business intelligence a MIS

- ❏ Business intelligence:
 - ❏ množina technologií a procesů, které využívají data k porozumění a analýze podnikatelské výkonnosti
- ❏ Analytika:
 - ❏ podmnožina BI
 - ❏ extenzivní využívání dat, statistické a kvantitativní analýzy, vysvětlovacích a prediktivních modelů k rozhodování a následnému jednání
 - ❏ vstup pro rozhodování:
 - ❏ lidí
 - ❏ plně automatické

Business intelligence



Nástroje BI

- Používané analytické technologie
 - tabulkové procesory (MS Excel)
 - OLAP nástroje (On Line Analytical Processing)
 - statistické a kvantitativní algoritmy
 - nástroje pro dolování dat
 - nástroje pro textové dolování (data crawler)
 - simulační nástroje

Ukládání dat

- OLTP (On Line Transaction Processing)
 - velké množství transakčních dat
 - detailní data z každodenních transakcí v podniku
- OLAP (On Line Analytical Processing)
 - analýzy obvykle historických dat pro hledání trendů potřebných pro strategická rozhodnutí managementu
 - periodické ukládání dat (detailních nebo agregovaných)

Datový sklad (Data Warehouse)

- ▣ předměťově orientované, integrované, časově proměnné kolekce dat určené pro podporu rozhodování
- ▣ předměťově orientované – prodeje, zákazníci, dodavatelé, produkty
- ▣ určený pro OLAP

- ▣ *Pozn.: Na rozdíl od OLAP je OLTP procesně orientované – zpracování mezd, účetnictví, výroba*

Datový sklad

- data integrována z různých OLTP datových zdrojů
- není aktualizován v reálném čase (okamžitě)
- aktualizován periodicky (např. denně, týdně, měsíčně)
- potenciálně konkurenční výhoda
- zvyšuje produktivitu managementu při rozhodování

Databáze OLTP vs. Datový sklad

Databáze OLTP

Obsahuje aktuální data

Ukládá jen detailní údaje

Data jsou dynamická

Dotazy jsou rychlé a přistupují k několika řádkům tabulek

Řízeny transakcemi, podporují každodenní operace

Procesně orientovány

Slouží velkému množství paralelně pracujících uživatelů

Datový sklad

Obsahuje historická data

Ukládá detailní data společně s agregovanými

Data jsou statická kromě periodického dohrávání dalších období

Dotazy trvají dlouho a přistupují k mnoha řádkům tabulek

Řízeny analýzami, podporují rozhodování

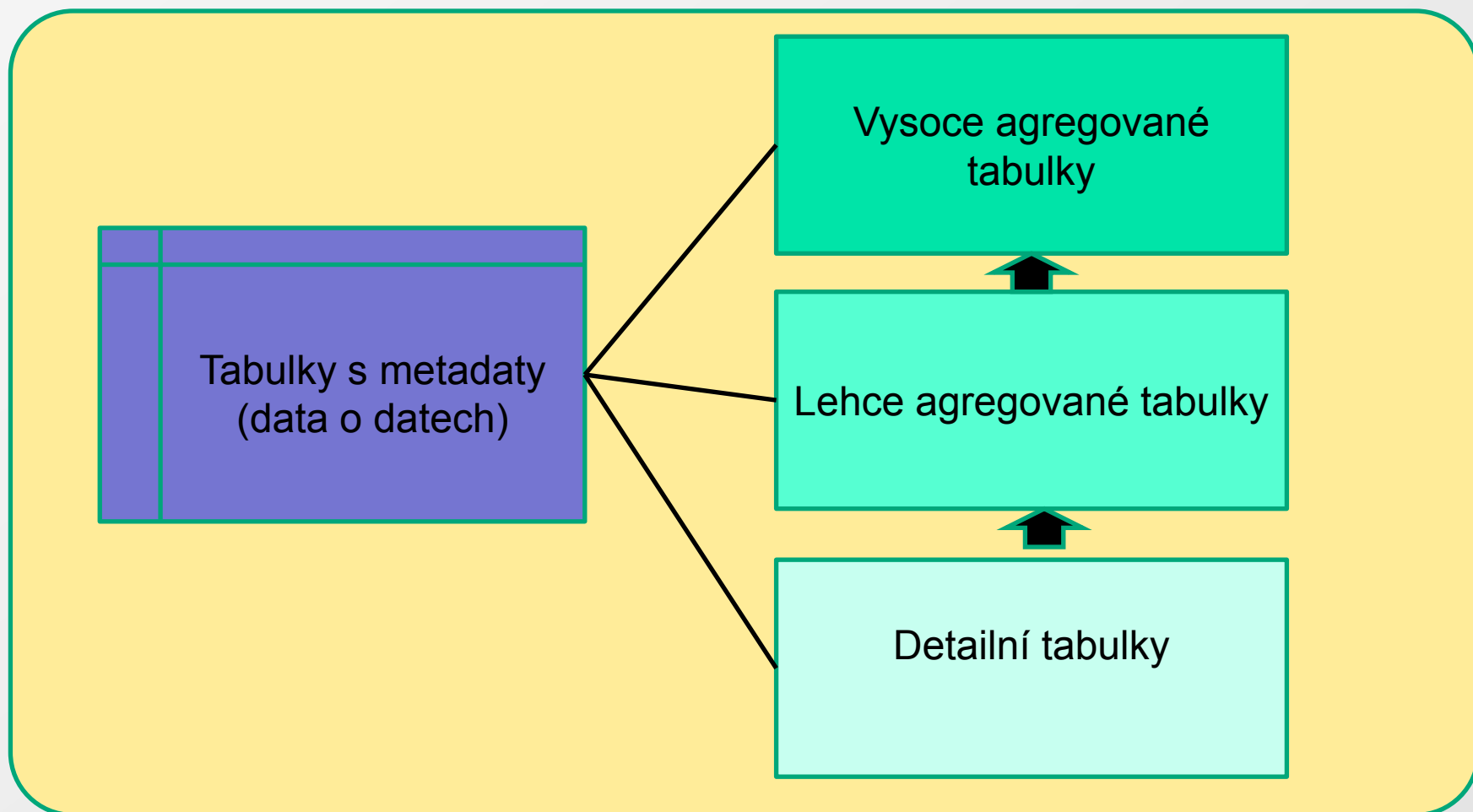
Předmětově orientované

Slouží relativně malému počtu manažerů, kteří rozhodují

Architektura datových skladů

- Agregované tabulky
- Hvězdicová schémata

Datový sklad – agregované tabulky

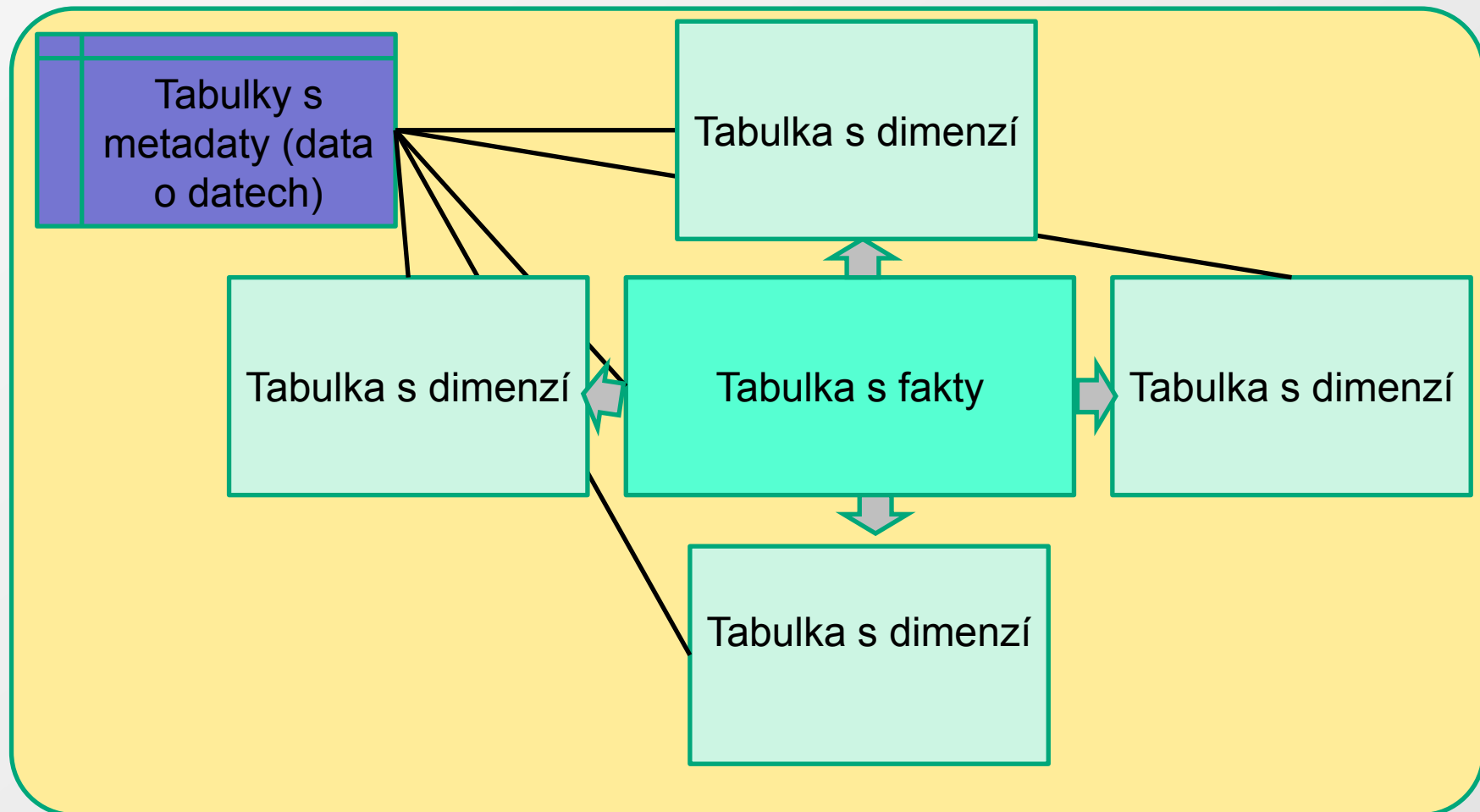


Pozn.: Základem úspěchu je navrhnout správné agregované tabulky

Datový sklad – hvězdicová schémata

- Jedna tabulka s detailními daty – tabulka s fakty (fact table)
 - Obklopená řadou tabulek s dimenzemi (dimension tables)
 - dimenze nebo také kategorie

Datový sklad – hvězdicová schémata



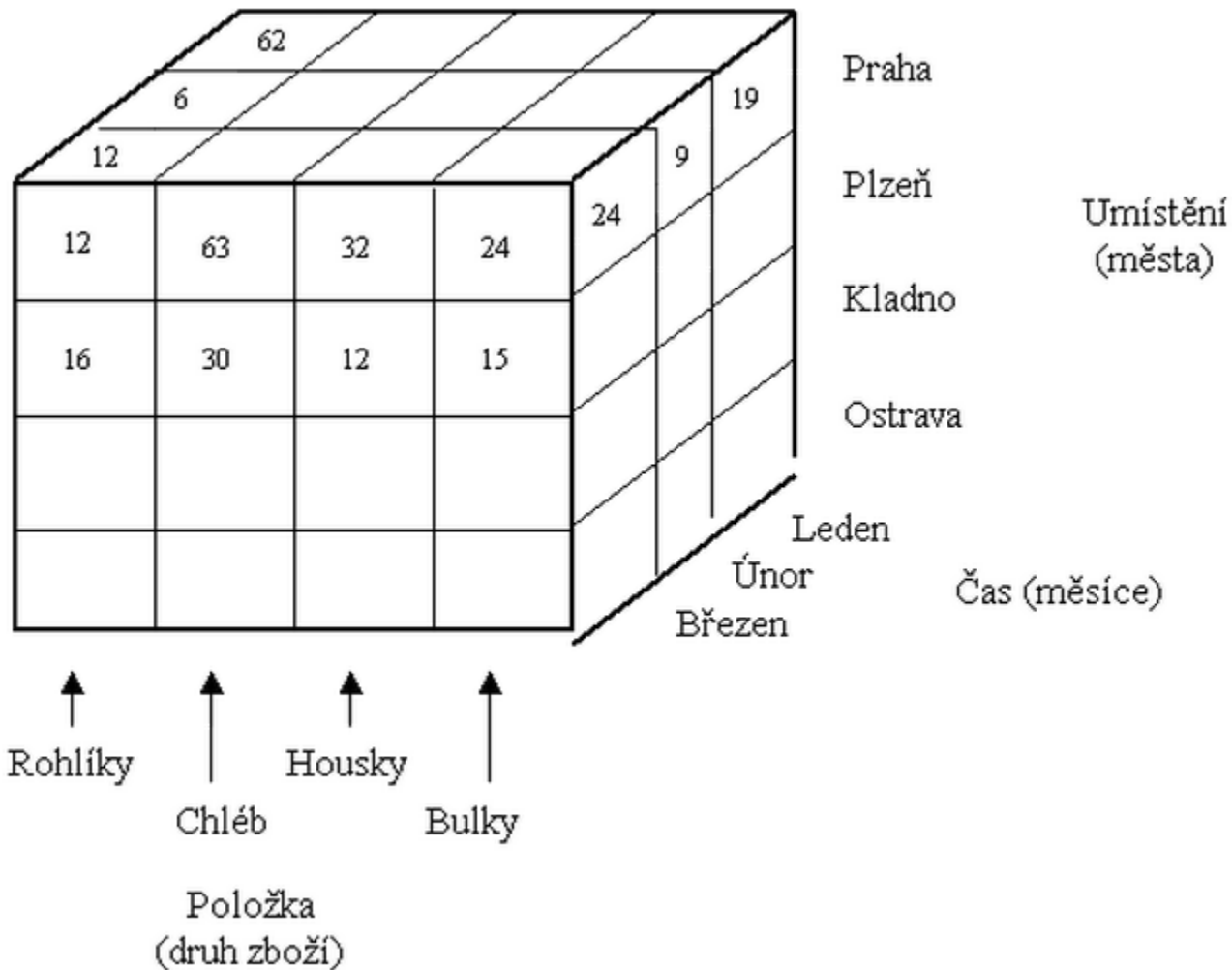
Datový sklad – hvězdicová schémata

- Tabulka s fakty v dimenzionálním modelu obsahuje:
 - metriky a fakta pro podnikatelské procesy
- Pozn.: Základem úspěchu je navrhnout správnou tabulku s fakty se správnými dimenzemi

Multidimenzionální databáze

- Označované jako MOLAP (Multidimensional On Line Analytical Processing)
- Výrobci:
 - Oracle Essbase, MicroStrategy, Cognos PowerPlay

Multidimenzionální databáze jako datová kostka



OLAP – interaktivní analýzy dat

- OLAP (On Line Analytical Processing)
- Například kontingenční tabulky v MS Excelu
- Vstup: Datová tabulka
 - Sloupce obsahují:
 - Kategorie (tvoří dimenze kontingenční tabulky nebo osy grafu)
 - Hodnoty (údaje v kontingenční tabulce nebo grafu)

OLAP – vstupní data

Datum	Prodejce	Značka	Model	Cena základní	Cena výbavy
2.5.2011	Brno	Fabia Combi	Scout	398,9	8,5
2.5.2011	Ždár n. S.	Superb	Exclusive	904,9	25,3
2.5.2011	Jihlava	Superb Combi	Ambition	796,9	55,9
2.5.2011	Praha	Yeti	Active	503,8	58,8
2.5.2011	Třebíč	Praktik	Praktik	262,8	20,7
2.5.2011	Praha	Roomster	Style	349,9	23,6
3.5.2011	Brno	Octavia	Classic	428,8	34,8
3.5.2011	Praha	Octavia Combi	Elegance	489,9	45,9
3.5.2011	Pardubice	Fabia	Ambiente	264,9	15,6
3.5.2011	Plzeň	Superb Combi	Ambition	759,9	55,9
3.5.2011	Praha	Yeti	Classic	497,8	53,9
4.5.2011	Pardubice	Praktik	Praktik	262,8	20,7
5.5.2011	Brno	Roomster	Style	349,9	23,6
5.5.2011	Brno	Octavia	Classic	439,9	38,8
5.5.2011	Brno	Octavia Combi	Elegance	489,9	45,9
5.5.2011	Praha	Superb Combi	Ambition	725,8	43,5
5.5.2011	Pardubice	Yeti	Elegance	538,9	65,8
5.5.2011	Jihlava	Praktik	Praktik	262,8	20,7
5.5.2011	Plzeň	Roomster	Style	349,9	23,6
5.5.2011	Jihlava	Octavia	Classic	439,9	38,8
6.5.2011	Pardubice	Octavia Combi	Elegance	478,8	46,9
6.5.2011	Praha	Fabia	Ambiente	256,9	10,9

Datamart

- Podmnožina datového skladu pro:
 - podporu některé organizační složky nebo podnikatelské funkce (např. obchod)

Dolování dat

- Proces extrahování
 - validních,
 - dříve neznámých,
 - komplexních informací
- z rozsáhlých databází a jejich použití pro důležitá podnikatelská rozhodnutí

Dolování dat

- Strukturovaných dat z relačních databází
- Dolování textových dat
 - Amazon: Ti, co kupují tuto knihu ..., často kupují také tuto knihu ...
 - Google Analytics

Dolování dat ze sociálních sítí

⇒ Příklady:

- ⇒ Twitter: Friends/Followers, retweets
- ⇒ Folksonomie (folksonomy)
 - ⇒ základním aspektem lidské inteligence je potřeba klasifikace a klasifikační hierarchie
 - ⇒ klasifikace pomocí tagů (tags)
 - ⇒ např. Twitter: hashtags (začínají #)
- ⇒ Analýza textů přirozených jazyků