Dobývání znalostí

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Dobývání znalostí

– Úvod do problematiky –

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Dobývání znalostí - úvod

Dobývání znalostí z databází (KDD):

~ Netriviální získávání implicitních, dříve neznámých a potenciálně užitečných informací z dat

- Začátky v 90. letech 20. století:
- Knowledge discovery in databases (KDD)
- Data mining (DM)

Dobývání znalostí - úvod (2)

Začátky, motivace a základy:

- **Umělá inteligence**
  - metody strojového učení
- **Databázové technologie**
  - uchovávání dat, vyhledávání informací
- **Statistika**
  - modelování a analýza závislostí v datech

\+ **potřeba používat (zpracované) údaje pro podporu (strategického) rozhodování ve firmě**

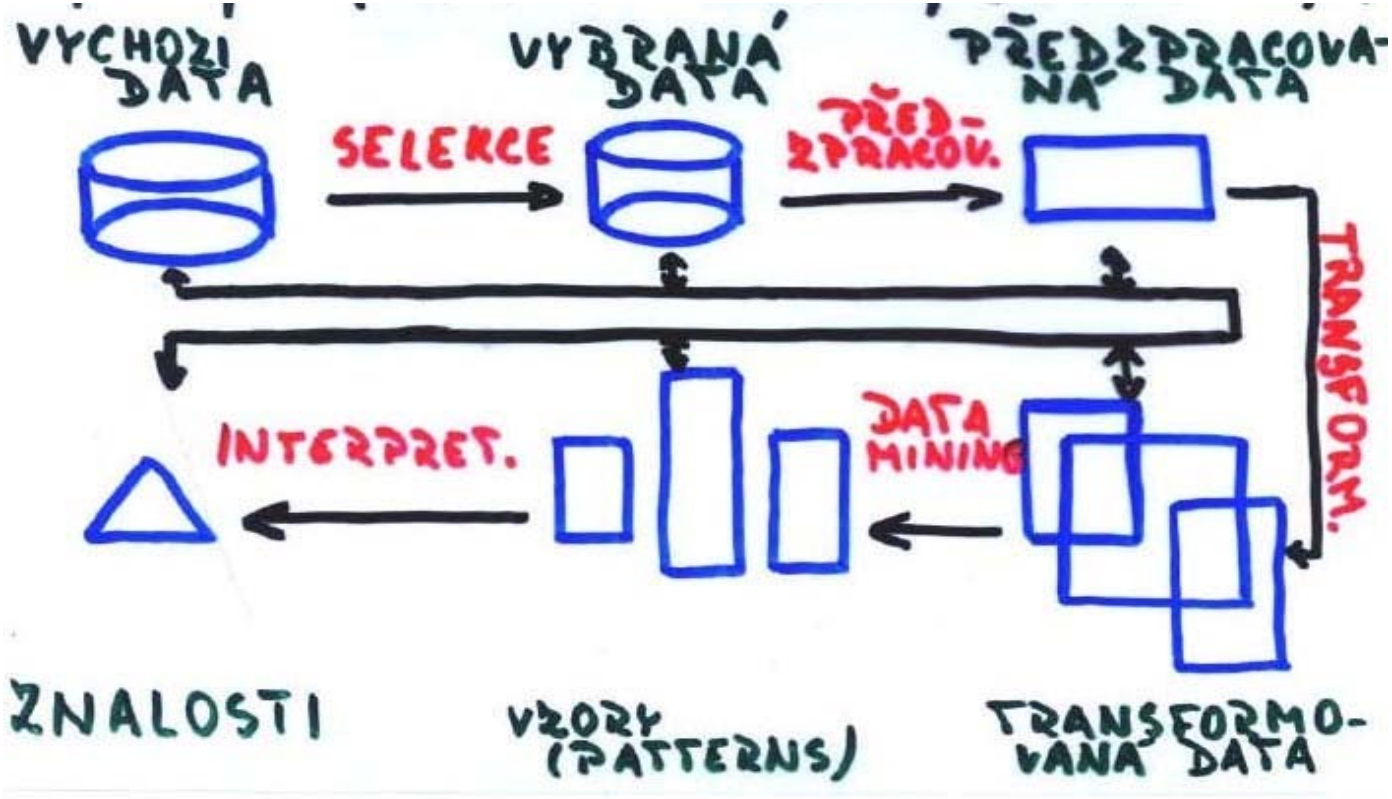
Dobývání znalostí: úvod (3)

~ interaktivní a iterativní proces:

- Příprava dat:
  - Z dat uložených ve složité struktuře (např. datový sklad) se vytváří (jedna) tabulka s relevantními údaji o zkoumaných objektech (klienti banky, zákazníci, …)
  - Selekce
  - Předzpracování
  - Transformace
- Vlastní „dobývání znalostí“ ~ data mining
- Interpretace
  - Nalezené znalosti se hodnotí z pohledu koncového uživatele

Dobývání znalostí - úvod (4)

~ **interaktivní a iterativní proces:**



Manažerský pohled na proces dobývání znalostí z databází

Reálný problém:

→ Impuls pro zahájení procesu dobývání znalostí

Cíl procesu dobývání znalostí:

- Získat co nejvíce relevantních informací vhodných k řešení daného problému

Příklad:

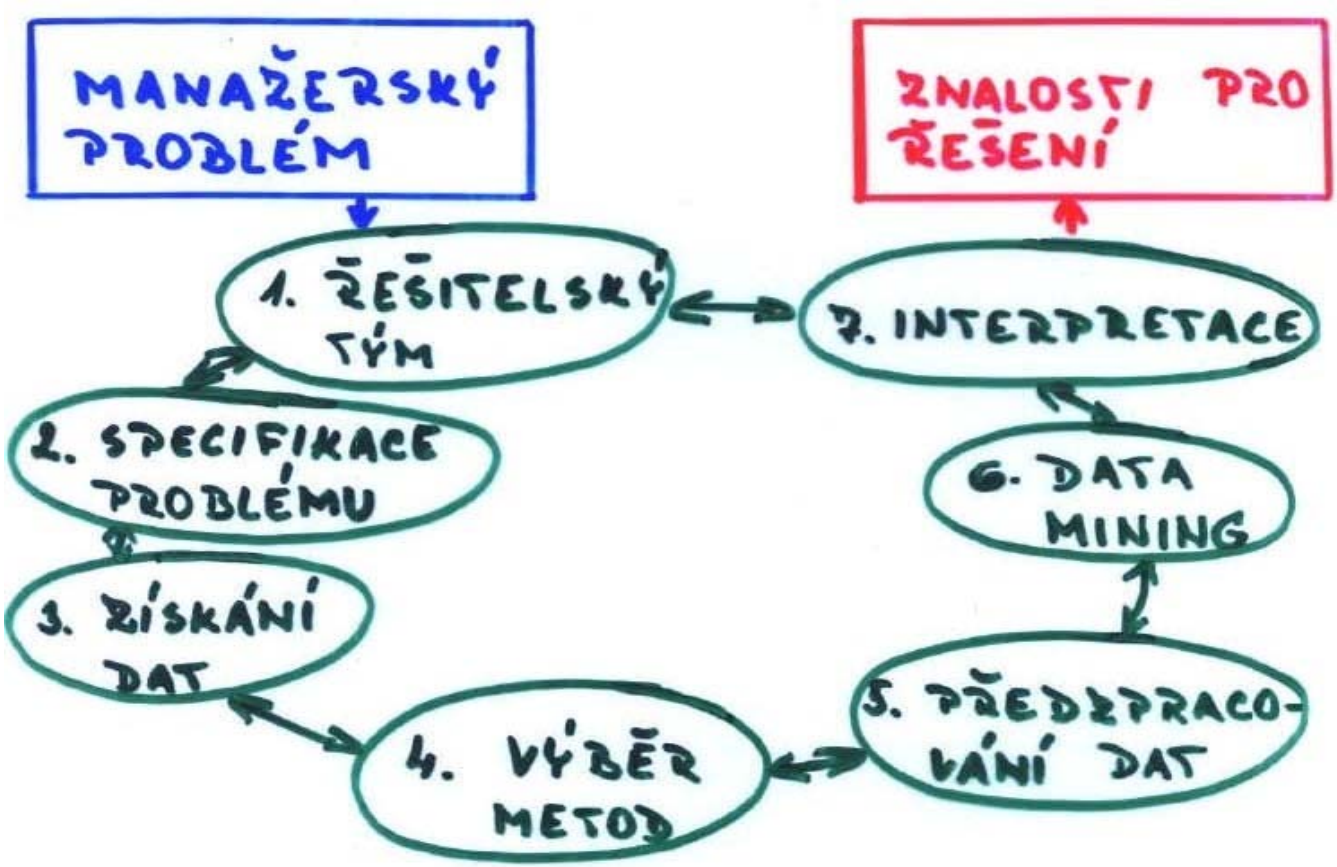
- Nalezení skupin zákazníků obchodního domu nebo skupin klientů banky, kterým lze nabídnout speciální služby
- Nalezené skupiny se interpretují jako segmenty trhu v dané oblasti

Manažerský pohled na proces dobývání znalostí z databází (1)

Řešení problému:

1. **Vytvořit řešitelský tým**
2. **Specifikace problému**
3. **Získat všechna dostupná data**
4. **Výběr metody**
5. **Předzpracování dat**
6. **Data mining**
7. **Interpretace**

Manažerský pohled na proces dobývání znalostí z databází (2)



Manažerský pohled na proces dobývání znalostí z databází (3)

Řešení problému:

1. **Vytvořit řešitelský tým**
   - experti na řešenou problematiku, na data, na metody KDD
2. **Specifikace problému**
   - v kontextu dobývání znalostí
3. **Získat všechna dostupná data**
   - může vést i k přeformulování problému
   - kvalita datové základny (např. data archivovaná v různých systémech, …)
   - **externí data** popisující prostředí, v němž se analyzované děje odehrávají (např. kalendářní období, reklama, politické události, počasí, …)

Manažerský pohled na proces dobývání znalostí z databází (4)

Řešení problému (pokračování):

4. **Výběr metody pro analýzu dat**
   - často je třeba kombinovat více různých metod:
     - klasifikační metody, metody explorační analýzy dat, metody pro získávání asociačních pravidel, rozhodovací stromy, genetické algoritmy, Bayesovské sítě, neuronové sítě, …
     - metody vizualizace

5. **Předzpracování dat**
   - získaná data se převedou do tvaru požadovaného pro aplikaci zvolených metod
     - např. odstranění odlehlých hodnot, doplnění chybějících hodnot, …
     - výpočetní operace mohou být i značně náročné

Manažerský pohled na proces dobývání znalostí z databází (5)

Řešení problému (pokračování):

6. **Data mining**
   - aplikace zvolených analytických metod pro vyhledávání zajímavých vztahů v datech
   - jednotlivé metody mohou být aplikovány i vícekrát
   - hodnoty vstupních parametrů jednotlivých běhů závisí na výsledcích předchozích běhů
   - jednotlivé typy metod se kombinují na základě dílčích výsledků

Manažerský pohled na proces dobývání znalostí z databází (6)

Řešení problému (pokračování):

7. Interpretace

- **(nezbytné) zpracování obvykle velkého množství výsledků jednotlivých metod**
  - některé výsledky jsou pro uživatele **nezajímavé** anebo **samozřejmé**
  - některé výsledky lze použít přímo, některé je třeba vyjádřit pro uživatele **srozumitelněji**
- výsledky je vhodné uspořádat do **analytické zprávy**
- výstupem může být i provedení vhodné akce
  - např. spuštění monitorovacího programu

Úlohy pro dobývání znalostí

Tři typy úloh:

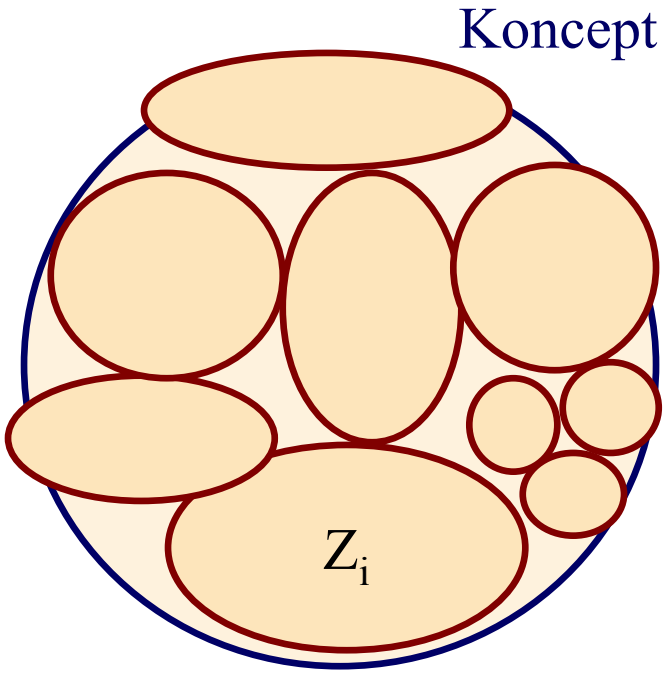
- Klasifikace, resp. predikce
- Deskripce (~ charakteristika, popis)
- Hledání „nugetů“

Úlohy pro dobývání znalostí (2)

Klasifikace (resp. predikce)

- Cílem je nalézt znalosti použitelné pro klasifikaci nových vzorů (případů)
- Získané znalosti by měly co nejlépe odpovídat danému konceptu
- Dáváme přednost přesnosti pokrytí na úkor jednoduchosti
- **Výsledkem je větší množství méně srozumitelných dílčích znalostí**

Úlohy pro dobývání znalostí (3)

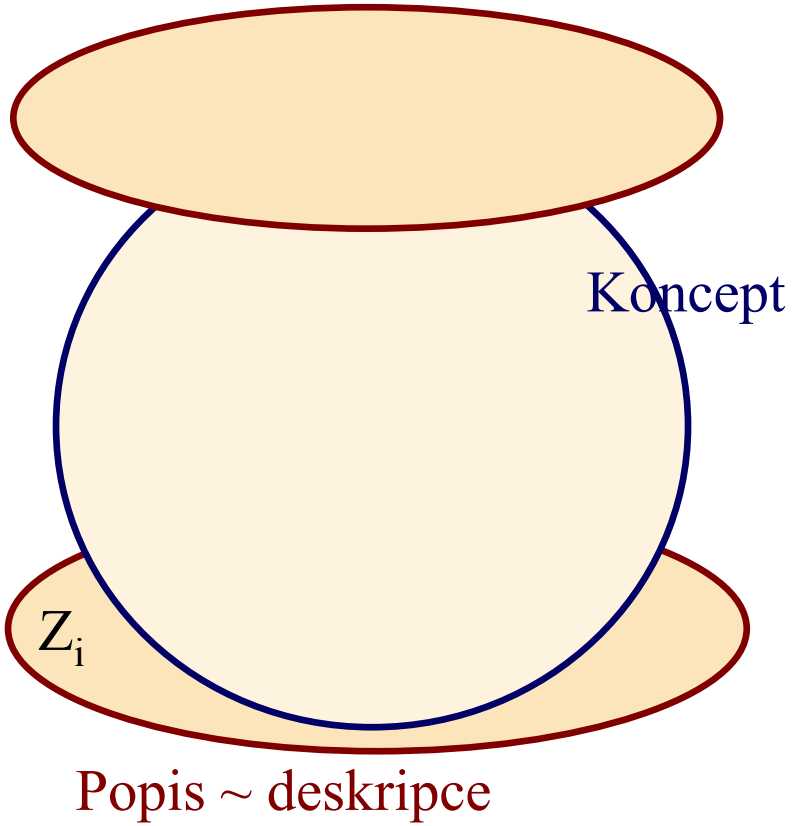


Klasifikace, resp. predikce

Predikce

- ze starších hodnot nějaké veličiny se pokoušíme odhadnout její vývoj v budoucnu
  - např. předpověď počasí, pohyb cen akcií, …

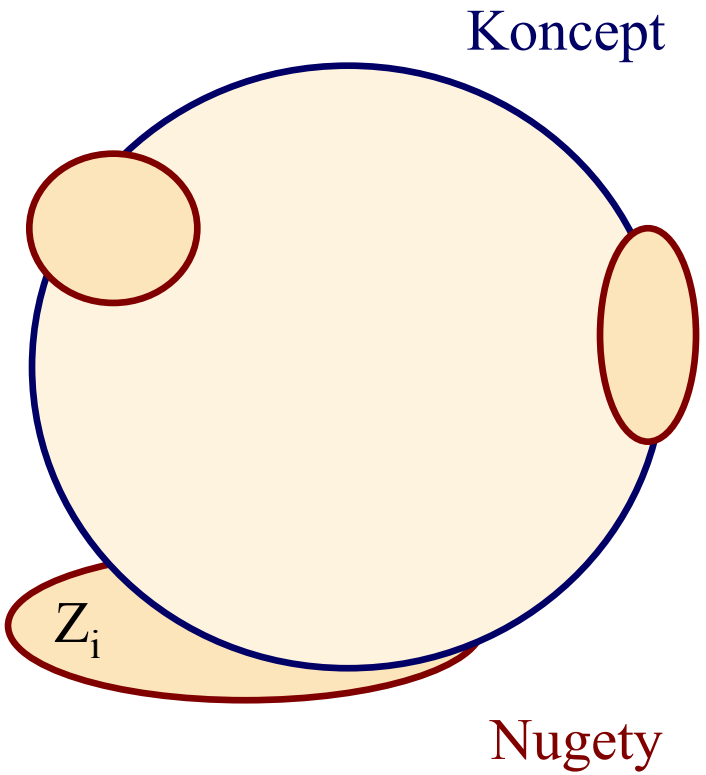
Úlohy pro dobývání znalostí (4)



Deskripce (~ popis)

- Cílem je nalézt dominantní strukturu nebo vazby, které jsou obsažené v daných datech
- Požadujeme srozumitelné znalosti pokrývající daný koncept
- **Výsledkem je menší množství méně přesných znalostí**

Úlohy pro dobývání znalostí (5)



Hledání nugetů

- Hledáme **zajímavé (nové, překvapivé)** znalosti, které **nemusí** plně pokrývat daný koncept

Úlohy pro dobývání znalostí a jejich aplikace

- Segmentace a klasifikace klientů banky
  - např. rozpoznávání problémových anebo vysoce bonitních klientů
- Predikce vývoje kurzu akcií
- Predikce spotřeby elektrické energie
- Analýza příčin poruch v telekomunikačních sítích
- Analýza důvodů změny poskytovatele služeb
  - Internet, mobilní telefony, …

Úlohy pro dobývání znalostí a jejich aplikace (2)

- Segmentace a klasifikace klientů pojišťovny
- Určení příčin poruch automobilů
- Rozbor databáze pacientů v nemocnici
- Analýza nákupního košíku
  - MBA ~ Market Basket Analysis
  - Walmart (u nás Delvita, Meinl, …)
  - Řetězce supermarketů

Úlohy pro dobývání znalostí a jejich aplikace (3)

- Analýza nákupního košíku (pokračování)
  - Data tvoří např. charakteristiky zákazníků a údaje o jednotlivých nákupech
  - Data předzpracovaná do relační tabulky

→ lze hledat souvislosti mezi jednotlivými typy zboží

- Existují skupiny produktů, které si zákazníci kupují současně?
- Čím se vyznačují jednotlivé skupiny zákazníků?
  - nízký příjem, …

Metodiky pro dobývání znalostí

Cíl: poskytnout uživatelům jednotný rámec pro řešení různých úloh z oblasti dobývání znalostí

- Metodiky vyvinuté producenty programových systémů (5A, SEMMA)
- Metodiky vyvinuté ve spolupráci výzkumných a komerčních institucí jako „softwarově nezávislé“ (CRISP-DM)

→ sdílení a přenos zkušeností z úspěšných projektů

Metodika 5A

- **ASSESS** – posouzení potřeb projektu
  - Stanovení kontextu – cílů, strategií a procesů
- **ACCESS** – shromáždění potřebných dat a jejich příprava
- **ANALYZE** – provedení analýz
  - Data se přeměňují na informace a znalosti

  → použít vícero metod a porovnat jejich výsledky a efektivitu

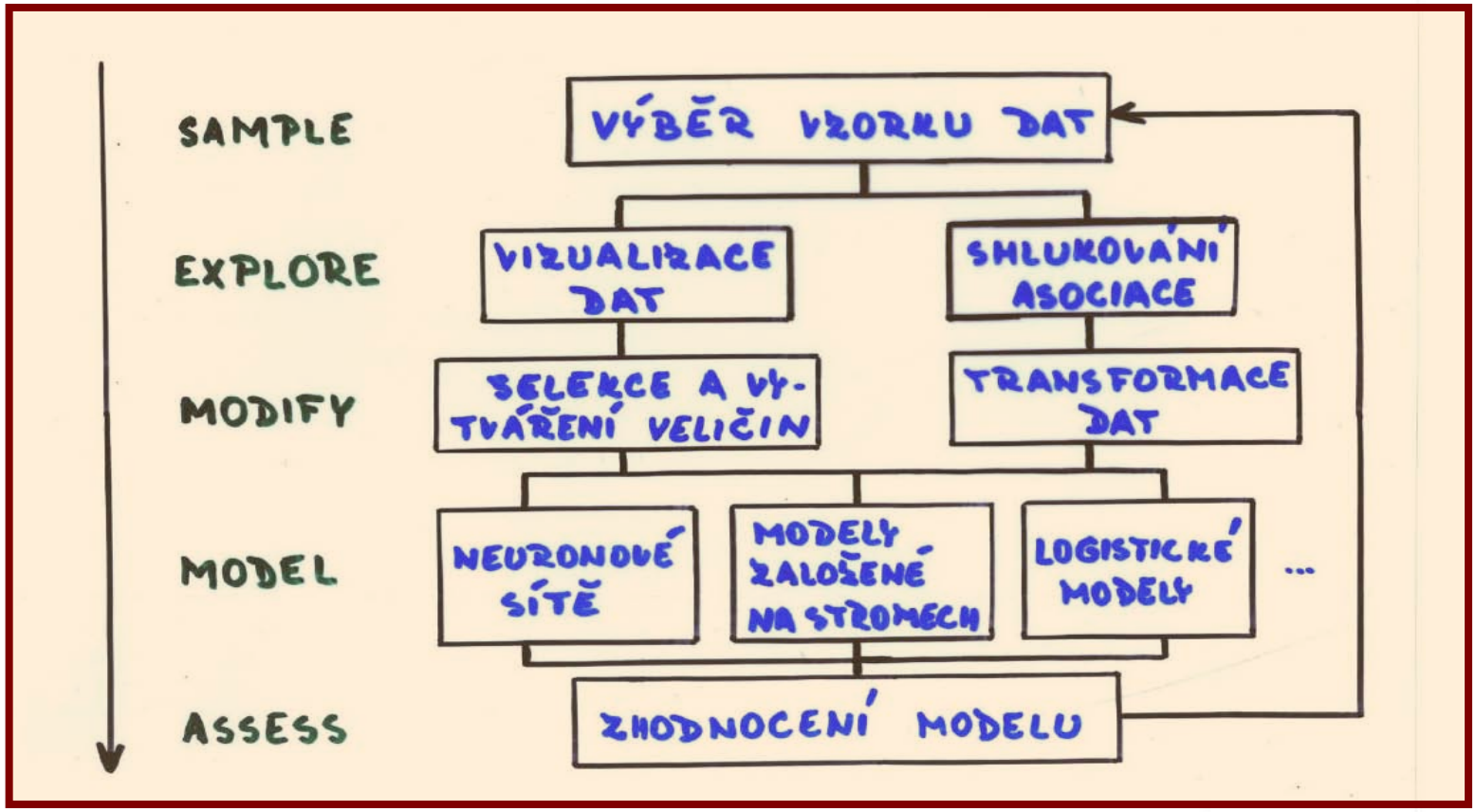
Metodika 5A (pokračování)

- **ACT** – přeměna znalostí na akční znalosti
  - Doporučení, dodatečné otázky a následná rozhodnutí
    → nalezené výsledky by měly být prezentovány jasně a srozumitelně
- **AUTOMATE** – převedení výsledků analýzy do praxe
  - Může zahrnovat např. i vytvoření praktického rozhraní pro snadné použití
  - Umožnit aktualizaci modelů podle nových výsledků

Metodika SEMMA (Enterprise Miner)

- **SAMPLE** – výběr vhodných objektů
- **EXPLORE** – vizuální explorace a redukce dat
- **MODIFY** – seskupování objektů a hodnot atributů, datové transformace
- **MODEL** – analýza dat
  - Neuronové sítě, rozhodovací stromy, statistické techniky, asociace a shlukování
- **ASSESS** – porovnání modelů a interpretace
  - Srozumitelnost pro uživatele

Metodika SEMMA
(Enterprise Miner)



Metodika CRISP-DM

~ CRoss-Industry Standard Process for Data Mining

- Vznik v rámci evropského výzkumného projektu

Cíl:

- Navrhnout univerzální postup použitelný v nejrůznějších komerčních aplikacích

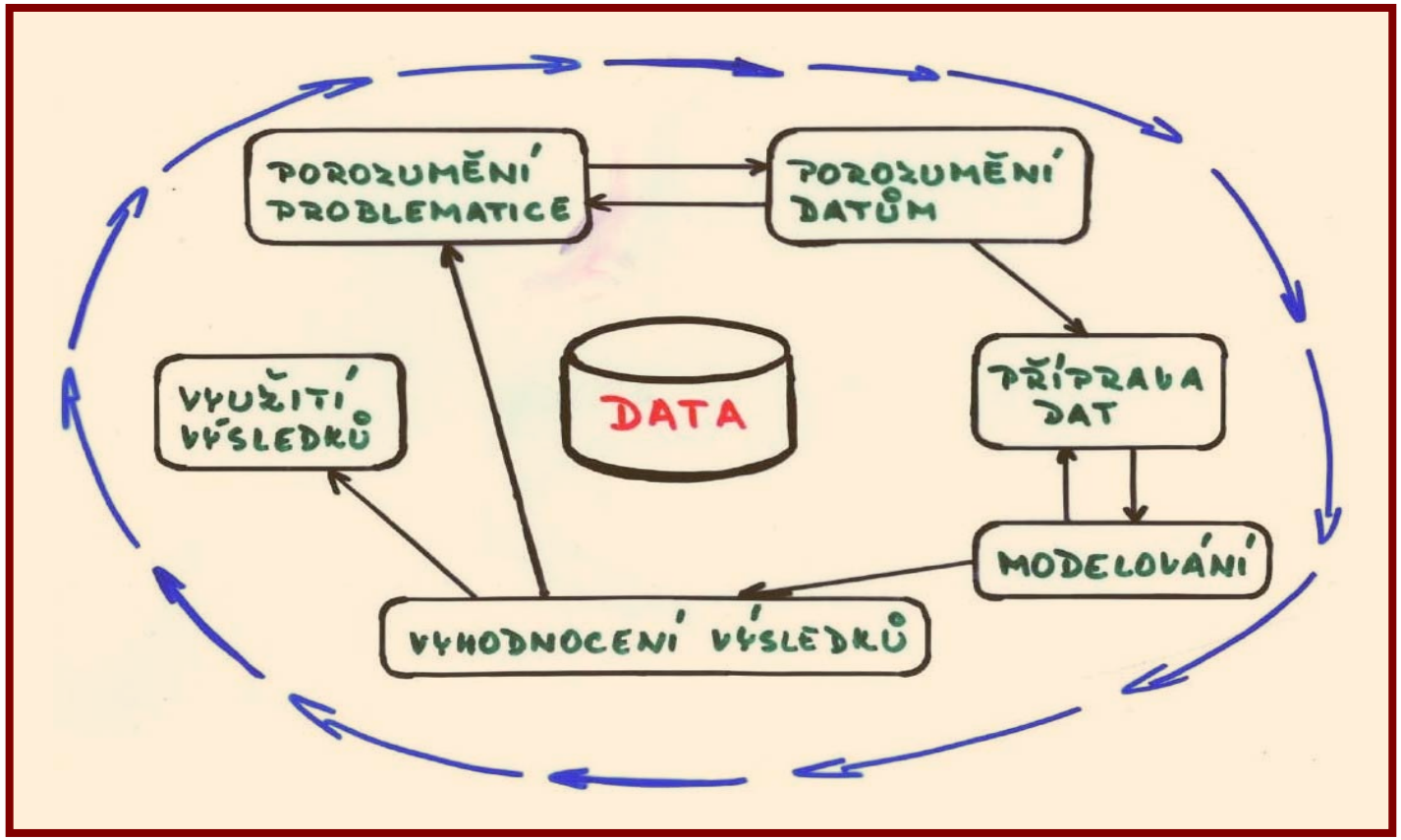
  → Standardní model procesu dobývání znalostí (z databází)

  \+ „průvodce“ možnými problémy a jejich řešením v reálných aplikacích

Metodika CRISP-DM (2)

- Proces dobývání znalostí má 6 fází
  - X pořadí fází není přesně určeno
- Výsledky získané v jedné fázi ovlivňují volbu dalších kroků
- Některé kroky a fáze je třeba provádět opakovaně

Metodika CRISP-DM (3)



Metodika CRISP-DM (4)
(NCR, Daimler-Chrysler, ISL, OHRA)

Porozumění problematice
(~ Business understanding)

- Pochopení cílů úlohy a požadavků na řešení (formulovaných z pohledu manažera)
- Manažerskou formulaci je nutné převést na zadání úlohy pro dobývání znalostí z databází
- „Revize“ zdrojů (datových, výpočetních i lidských)
  - Hodnotí se možná rizika, náklady a přínos
- Stanoví se předběžný plán prací

Metodika CRISP-DM (5)

Porozumění datům

(~ Data understanding)

- Prvotní sběr dat
- Získání základní představy o datech
  - Posouzení kvality dat, vytipování zajímavých podmnožin záznamů v databázi, …
- Výpočet deskriptivních charakteristik dat
  - Četnost atributů, průměrné hodnoty, …
- Výhodou jsou vizualizační techniky

Metodika CRISP-DM (6)

Příprava dat

(~ Data preparation)

- Vytvoření datového souboru, který bude zpracováván jednotlivými analytickými metodami
- Data by měla obsahovat údaje podstatné pro danou úlohu a měla by být ve tvaru vyžadovaném algoritmy pro analýzu
- Příprava dat zahrnuje:
  - Selekci dat, čištění dat, transformaci dat, vytváření dat, integrování dat, formátování dat, …
- Jednotlivé úkony se obvykle provádějí opakovaně a v nejrůznějším pořadí

Metodika CRISP-DM (7)

Modelování

(~ Modeling)

- Použití analytických metod pro dobývání znalostí
  - Z možných metod vybrat ty nejvhodnější a adekvátně nastavit jejich parametry
- Iterativní činnost
  - Opakovaná aplikace algoritmů s různými parametry
- Může vést k potřebě modifikovat data
- Ověření nalezených znalostí

Metodika CRISP-DM (8)

Vyhodnocení výsledků

(~ Evaluation)

- Z pohledu manažerů
  - Byly splněny cíle formulované při zadání úlohy?
- Rozhodnutí o způsobu využití výsledků

Metodika CRISP-DM (9)

Využití výsledků

(~ Deployment)

- Upravit získané znalosti do podoby použitelné pro zákazníka (manažera, zadavatele)
  - Zákazník musí pochopit, co je třeba učinit pro efektivní využití dosažených výsledků!
  - Implementace klasifikačního algoritmu v user-friendly podobě
  - Příprava uživatelského manuálu
  - Instalace programu na pobočkách banky a zaškolení uživatelů
  - Změna metodiky poskytování úvěrů a příslušná změna vnitřních předpisů banky
  - …..

Databáze

Relační databáze:

- Datový soubor je rozdělen do řady relací (tabulek)
  - Množina relací
    - Relace je reprezentovaná dvourozměrnou tabulkou (řádky odpovídají záznamům, sloupce odpovídají atributům)
    - Jednotlivé záznamy jsou jednoznačně identifikovány pomocí primárního klíče

Databáze (2)

Relační databáze (pokračování):

- Operace pro manipulaci s tabulkami
  - Selekce: slouží k výběru záznamů (~ řádků) tabulky
  - Projekce: slouží k výběru atributů (~ sloupců) tabulky
  - Spojení: slouží k propojování tabulek – spojují se řádky se stejnou hodnotou atributu, obvykle klíče
- Kladení dotazů
  - QBE (~ Query By Example)
  - SQL (~ Structured Query Language)

Databáze (3)

QBE – uživatel vyplní (vybere) ve formuláři, co ho zajímá

→ zadá „masku“, které by měly odpovídat nalezené záznamy

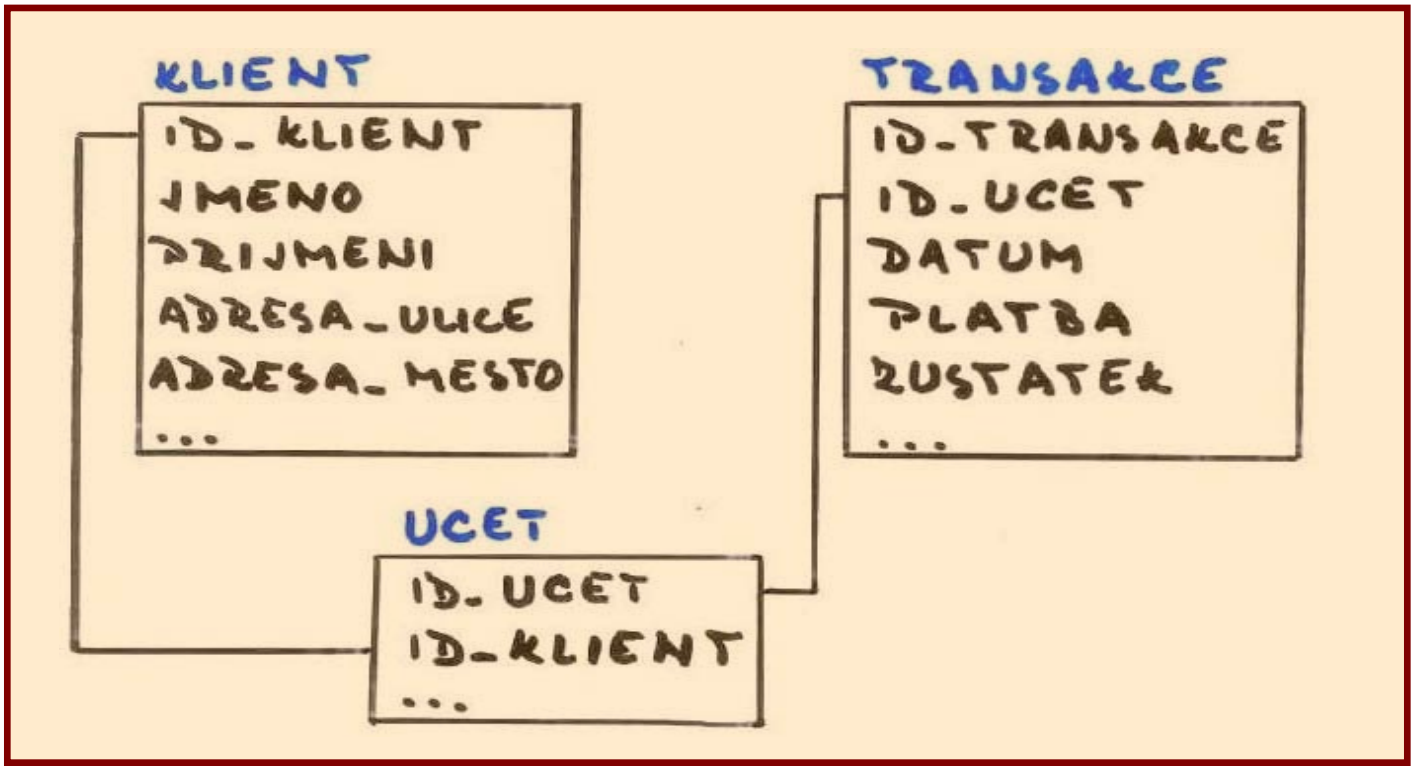
- Relativě jednoduchý, intuitivní způsob kladení dotazů
- Vhodnější pro méně zkušené uživatele

SQL – jednoduchý programovací jazyk pro definování dat a manipulaci s nimi

- Určeno pro zkušenější uživatele

Databáze (4)

Příklad relační databáze



Databáze (5)

Příklad dotazu v jazyce SQL

```
SELECT    klient.jmeno, klient.prijmeni,
          klient.adresa_ulice,
          klient.adresa_mesto, ucet.cislo_uctu,
          transakce.zustatek
FROM      klient, ucet, transakce
WHERE     klient.id_klient = ucet.id_klient;
AND       transakce.id_ucet = ucet.id_ucet;
AND       transakce.zustatek < 100;
GROUP BY klient.adresa_mesto
```

Databáze (6)

EIS ~ Executive Information Systems:

- První pokus přiblížit dotazování do databáze manažerům
- Snadné ovládání
- Analýzu provádí analytik sám
  - Uživatel vybere v menu dotaz
  - Následně je dotaz převeden do jazyka SQL
  - Poté je dotaz proveden standardním způsobem
- Nevýhody:
  - omezený počet předem připravených dotazů
  - Malá flexibilita

Databáze (7)

OLAP ~ On-Line Analytical Processing:

(E. F. Codd – 80. léta 20. století)

- Rychlost a flexibilita
- Intuitivní ovládání
- Možnost vizualizace
- Grafické rozhraní
  - Nahlížení na data v numerické podobě i v podobě grafů

Databáze (8)

OLAP (pokračování):

- Multidimenzionální koncept ukládání i manipulace s daty
- Intuitivní manipulace s daty
- Práce s daty z heterogenních datových zdrojů
  - Provádí se konverze dat
- Použití analytických metod
  - Statistické přehledy
  - What-if analýzy

Databáze (9)

OLAP (pokračování):

- client/server architektura
- Podpora multiuživatelského pohledu
- Ukládání výsledků OLAP mimo zdrojová data
- Dynamická manipulace s řídkými maticemi
- Zpracování chybějících hodnot
- Neomezený počet dimenzí a agregačních úrovní

Databáze (10)

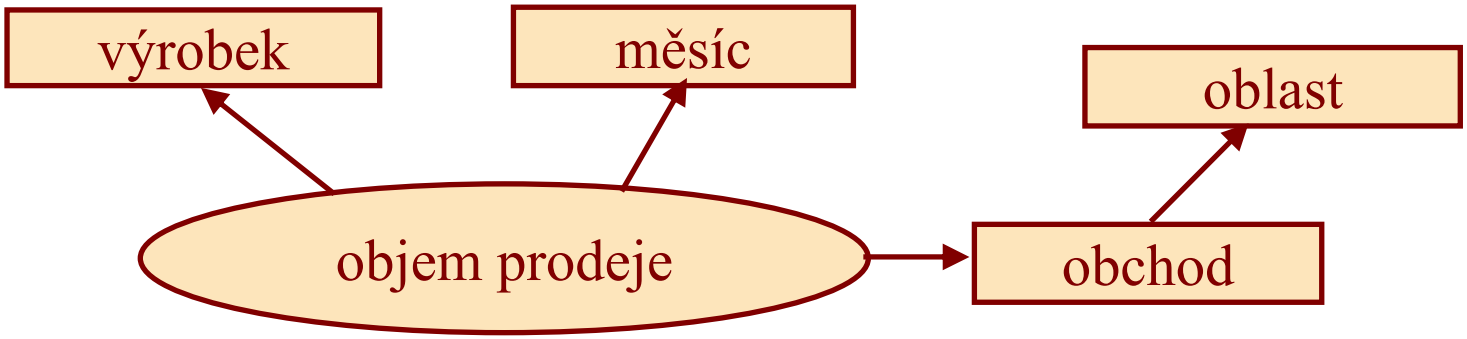
OLAP (pokračování):

Základ OLAP ~ pohled na data jako na mnoharozměrnou tabulku

→ datová krychle (~ data cube)

Příklad jednoduché databáze:

Údaje o prodeji různých výrobků v jednotlivých měsících v různých obchodech



Databáze (11)

OLAP (pokračování):

Převod databáze na datovou krychli

- Sledované atributy tvoří dimenze krychle
- Záznamům v databázi odpovídají buňky krychle

→ různé pohledy na data

X plýtvá se místem

- řada buněk je prázdná

Databáze (12)

OLAP (pokračování):

Příklad – záznamy v databázi PRODEJ:

Měsíc	výrobek	obchod	množství
Leden	káva	Praha	23
Leden	čaj	Plzeň	18
Leden	káva	Plzeň	27
Leden	čaj	Písek	4
Únor	mléko	Praha	40
Únor	káva	Praha	27
Únor	mléko	Plzeň	12

Databáze (13)

OLAP (pokračování):

Příklad – záznamy v databázi PRODEJ:

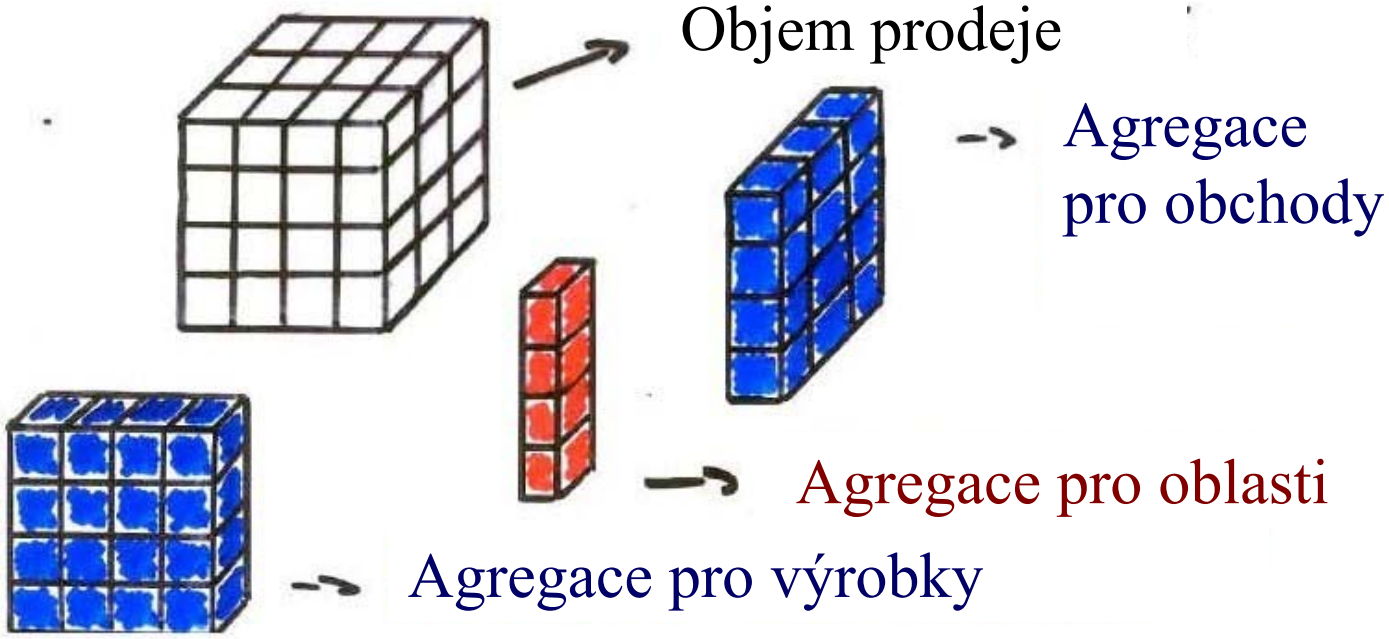
⇒ Řídká matice:

	Praha			Plzeň			Písek		
	K	Č	M	K	Č	M	K	Č	M
Leden	23			27	18			4	
Únor	27		40			12			

Databáze (14)

OLAP (pokračování):

Datová krychle



Databáze (15)

OLAP (pokračování):

Datová krychle (logický model)

- Obsahuje data z operačních databází
- \+ dílčí souhrny ~ **agregace**

  => rychlá odezva na „nové“ dotazy uživatele

  => flexibilita systému

Práce s krychlí:

- Natáčení (~ pivot)
- Provádění řezů (~ slice)
- Výběr určitých částí (~ dice)
- Zobrazování různých agregovaných hodnot

Databáze (16)

OLAP (pokračování):

Hodnoty atributů lze sdružovat do hierarchií:

- Úrovní v hierarchii může být více
  - Př.: obchod → oblast → kraj → země
- Operace:
  - Roll-up
    - přechod na hierarchicky vyšší – obecnější – úroveň
    - Zobrazované údaje mají podobu **souhrnů**
  - Drill-down
    - Podrobnější pohled
    - Různé úrovně - **granularita**

Databáze (17)

OLAP (pokračování):

Implementace (datové krychle):

- Velmi řídká data
- Nerovnoměrně rozmístěná data

Hyperkrychle (hypercube)

- Jedna velká krychle
- nástroje pro práci s řídkými daty
- Jednoduchá struktura a srozumitelnost pro uživatele

Multikrychle (multicube)

- Větší počet navzájem propojených menších krychlí (obsahujících jen několik dimenzí)
- Efektivní uložení dat

Databáze (18)

OLAP (pokračování):

Implementace (datové krychle):

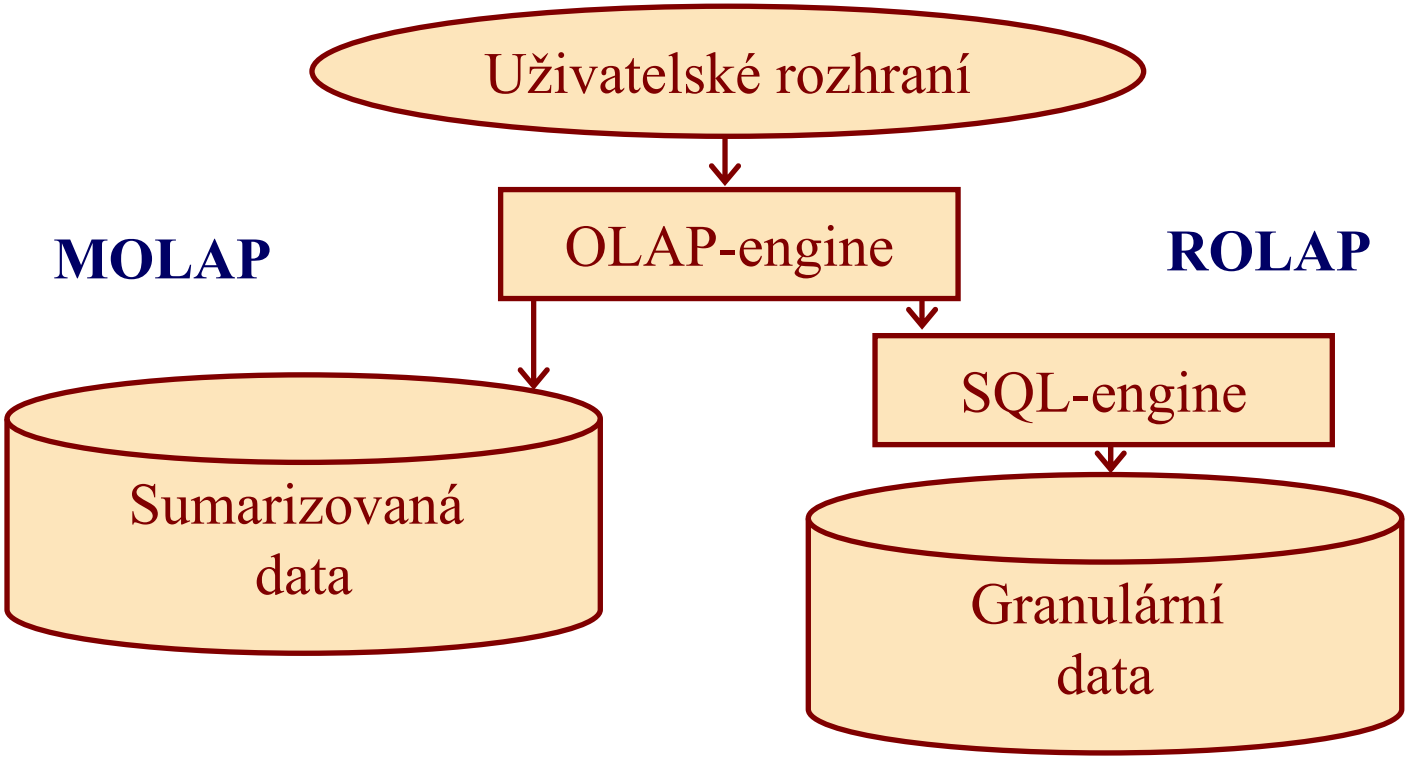
- rychlý přístup k datům klade vysoké nároky na datový server (a jeho cenu)

→ Namísto OLAP (~ MOLAP ~ Multidimenzionální OLAP) použít ROLAP ~ Relační OLAP (založený na klasické relační databázi)

  - Dotazy OLAP se převádějí do klasických dotazů SQL

Databáze (19)

Porovnání MOLAP x ROLAP:



Databáze (20)

MOLAP

~ „klasický“ OLAP

\+ vhodné pro středně velké, statiské aplikace

- např. analýza historických dat o prodeji nějakého výrobku

\- Nehodí se pro dynamické aplikace s průběžnou aktualizací dat (výpočty souhrnů!)

ROLAP

~ relační OLAP

\+ vhodné pro rozsáhlé aplikace využívající transakční data

\+ zpracování rozsáhlých dat za použití existujících databázových technologií

\- nepoužívá se příliš pro obchodní aplikace

Databáze (21)

Implementace ROLAP:

- Schéma hvězdy (star schema)
- Schéma sněhové vločky (snowflake schema)

Star schema – hvězda:

- Vychází z jedné centrální **tabulky faktů**, která obsahuje složený **primární klíč** (jeden segment klíče pro každou dimenzi) a **detailní data** (např. objem prodeje daného výrobku v daném obchodu za dané období)
  - Může obsahovat i **agregovaná data**

Databáze (22)

Star schema – hvězda (pokračování):

- Pro každou dimenzi existuje jedna tabulka s údaji na různé úrovni příslušné hierarchie → **tabulka dimenzí**
- **Úroveň v hierarchii (level)** se zaznamenává jako další indikátor do tabulky dimenzí
  → nutná při dotazování do tabulky, která obsahuje současně data detailní i agregovaná

Klady: srozumitelnost, snadné definování hierarchií, jednoduchá metadata, rychlý přístup k datům

Nevýhody: problémy s velkými tabulkami dimenzí, předpokládá statická data neaktualizovaná on-line

Databáze (23)

Snowflake schema – sněhová vločka:

- **Normalizované tabulky dimenzí**
  - Každá tabulka nějaké dimenze ukazuje na příslušnou **agregovanou tabulku faktů**
- **Tabulky dimenzí** obsahují jediný primární klíč pro danou úroveň dimenze spolu s **odkazem na nejbližšího rodiče v hierarchii dimenzí**
- Odpadá nutnost používat indikátor úrovně v hierarchii (v každé tabulce údaje jen z jedné úrovně)

Klady: dotazy na agregované hodnoty

Nevýhody: údržba, velký počet tabulek

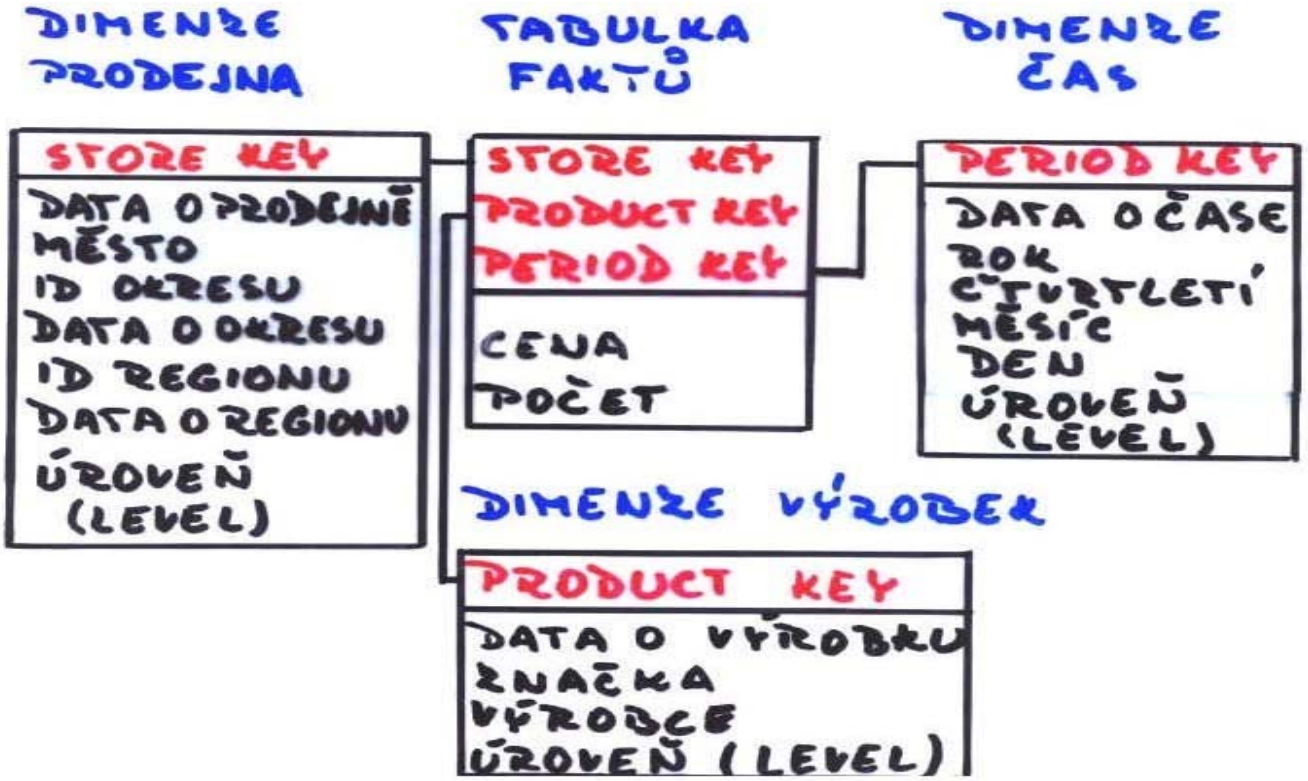
Databáze (24)

Příklad:

- Databáze má 3 dimenze: prodejna, výrobek, čas
- Dimenze prodejen je tvořena hierarchií:
  - obchod → okres → region
- Dimenze výrobku je tvořena hierarchií:
  - výrobek → značka → výrobce
- Dimenze času je tvořena hierarchií:
  - datum → měsíc → čtvrtletí → rok

Databáze (25)

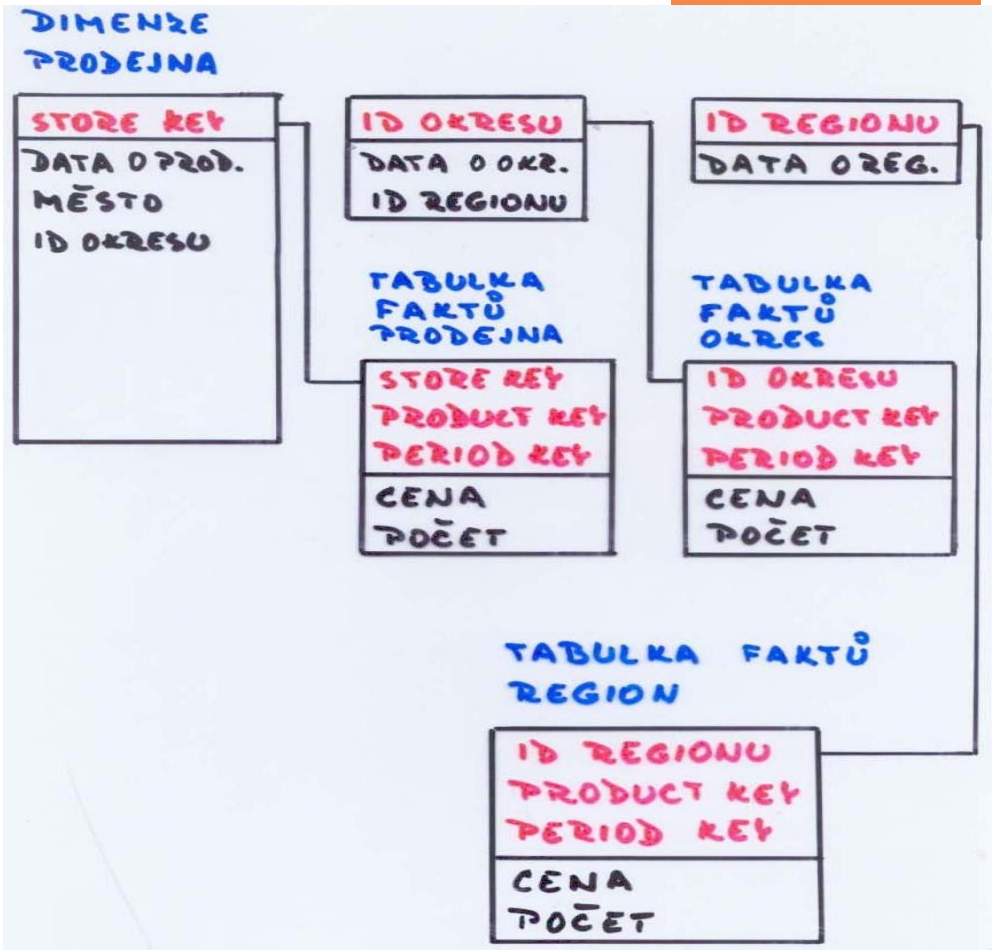
Příklad (pokračování): **hvězda**



Databáze (26)

Příklad (pokračování):

sněhová vločka



Databáze (27)

Datové sklady a datová tržiště:

- **OLAP** ~ nástroj pro analýzu a vizualizaci dat o firmě
- **Datový sklad** ~ místo, kde jsou analyzovaná data uložena
  - **Orientován na subjekty, kterými se daná firma zabývá**
    - Např.: zákazník, dodavatel, produkt, aktivita

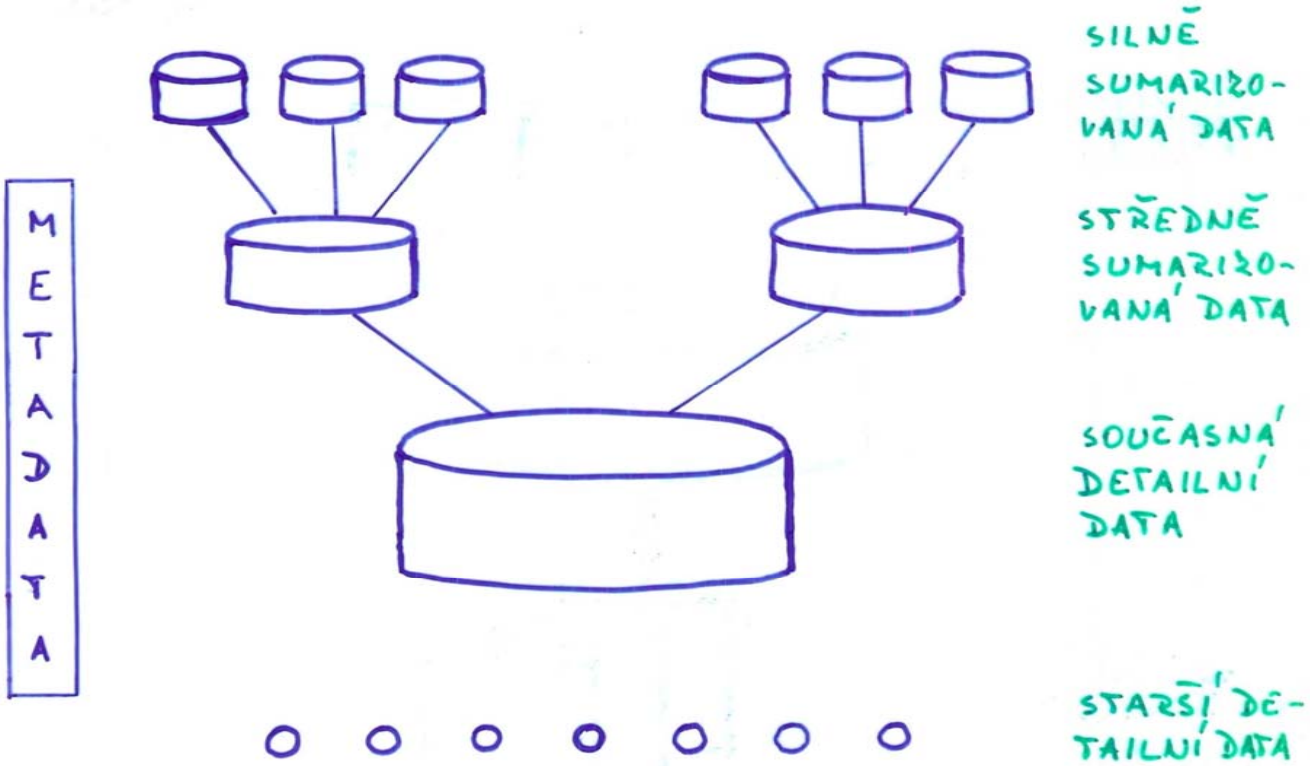
      → neuchovává data, která nejsou vhodná pro podporu rozhodování na manažerské úrovni
  - **Integrace a sjednocení dat**
    - Např.: názvů stejných ukazatelů, měřítek, kódování, …

Databáze (28)

- **Datový sklad** (pokračování)
  - **Časově proměnný**
    - Zafixování dat z produkčních databází
    - Off-line aktualizace v určitých časových intervalech (měsíčně, ročně, …)
    - Analýza odděleně od produkčních databází
      - Nešetrný zásah neovlivní operativní řízení firmy
      - Rychlejší odezva na položený dotaz
      - **X nevýhodou je zastarávání dat**
    - Časové údaje jsou v datovém skladu explicitně přítomny jako jedna z důležitých informací
  - **Stálý** ~ dotazy, které do datového skladu směřují uživatelé, nezpůsobují změnu zde uložených dat

Databáze (29)

Struktura datového skladu:



Databáze (30)

Struktura datového skladu:

- Operační data uložená v daném okamžiku
- Starší operační data
- Souhrny na různých úrovních abstrakce
- Metadata ~ zachycují informace o datech

Vytvoření datového skladu:

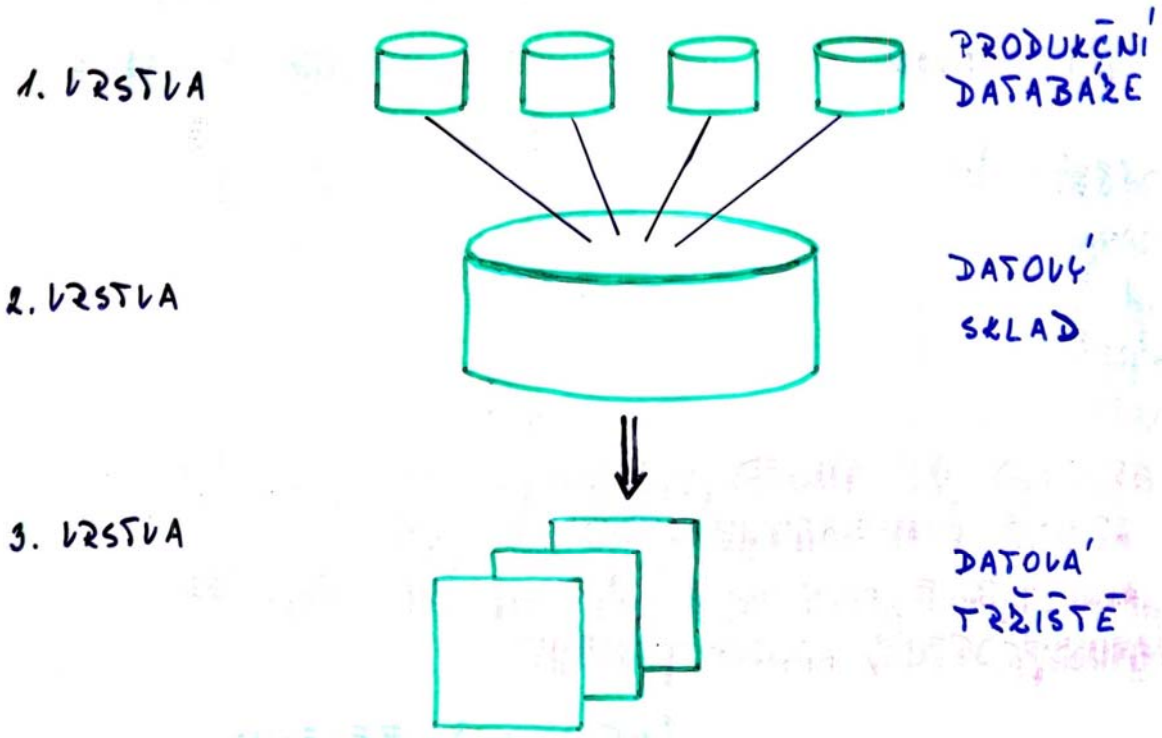
- Načtení dat
- Konverze dat
- Čištění, transformace, …

\+ **datová tržiště (data mart)**

- Přesun dat relevantních pro určitý typ analýz
- Třívrstvá architektura datového skladu

Databáze (31)

Třívrstvá architektura datového skladu



Databáze (32)

Dotazovací jazyky pro dobývání znalostí z databází:

- **MINE RULE**
  - Umožňuje klást dotazy na asociační pravidla:
    - Implikace typu „Jestliže platí předpoklad, platí i závěr“ doplněné o kvantitativní charakteristiky odvozené z počtu záznamů v databázi splňujících předpoklad, resp. závěr pravidla
  - Příkazy: FROM, WHERE, GROUP BY, CLUSTER BY, SELECT, EXTRACTING RULES
  - Příklad: IF produkt_1 & produkt_2 & … & produkt_n THEN produkt (SUPPORT, CONFIDENCE)

Databáze (33)

Dotazovací jazyky pro dobývání znalostí z databází:

- **MINE RULE** (pokračování)
  - SUPPORT (podpora)
    - ~ podíl počtu záznamů, ve kterých současně platí předpoklad i závěr pravidla, a celkového počtu záznamů vybraných na základě podmínky WHERE
  - CONFIDENCE (spolehlivost)
    - ~ podíl počtu záznamů, ve kterých současně platí předpoklad i závěr pravidla, a počtu záznamů, ve kterých platí pouze předpoklad

Databáze (34)

Dotazovací jazyky pro dobývání znalostí z databází:

- **Příklady dotazů**
  - **Dotaz v MINE RULE:**

```
MINE RULE Priklad AS
SELECT DISTINCT 1..n produkt AS BODY, 1..1
       produkt AS HEAD, SUPPORT, CONFIDENCE
FROM Prodej
WHERE BODY.město = HEAD.město
       AND BODY.datum = HEAD.datum
EXTRACTING RULES WITH SUPPORT: 0.1,
                      CONFIDENCE: 0.5
```

Databáze (35)

Dotazovací jazyky pro dobývání znalostí z databází:

- **Příklady dotazů**
  - **Dotaz v MSQL – hledání pravidel:**

    (Odvodit podle věku a pohlaví, jaké má daný zaměstnanec auto.)

```
Emp(Id,Age,Sex,Salary,Position,Car)

GetRules(Emp)
into R
where support > 0.1 and confidence > 0.9

SelectRules(R)
where body has {(Age=*),(Sex=*)}
      and head is {(Car=*)}
```