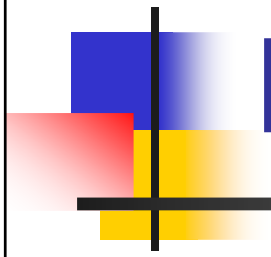
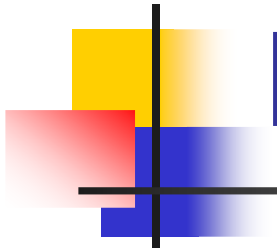


Business Intelligence

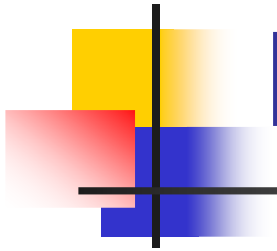
Skorkovský

KAMI, ESF MU



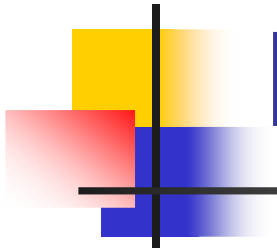
Principy BI

- zpracování velkých objemů dat tak, aby výsledek této akce manažerům pomohl k rozhodování při řízení procesů
- výsledkem zpracování musí být relevantní informace, kterou dostanou manažeři ve správném čase
- základní zdroj dat, která se často ukládají do datových skladů jsou ERP systémy (relační DB)
- získání informací jako výsledek strukturovaných dotazů musí probíhat rychle (krátká odezva)
- používá se pro řízení na strategické, taktické u operační úrovni



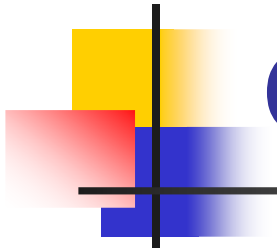
Principy BI

- **Definice 1** : BI je sběr a analýza dat, jejímž cílem je lepší porozumění a reakce na změny, kterým organizace neustále čelí
- **Definice 2** : BI je znalost podniku získaná za použití HW a SW technologií, která umožní přeměnit data organizace v informaci
- **Definice 3** : sada procesů, aplikací a technologií, jejíchž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Tyto procesy podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data



Nástroje BI

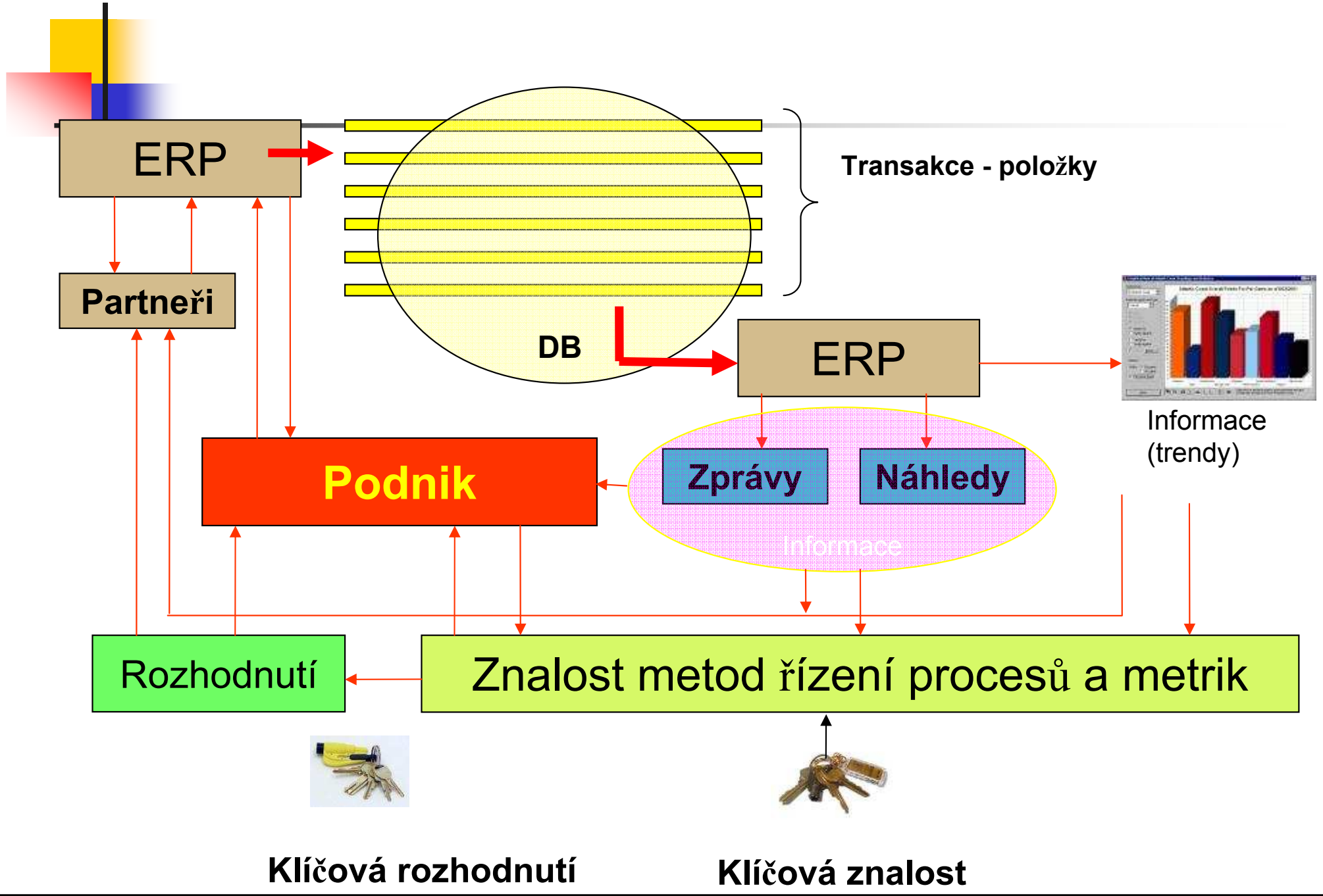
- ERP systémy
- Dočasná úložiště (DSA: Data Staging Area)
- Operativní úložiště (ODS : Operational Data Store)
- Transformační nástroje (ETL : Extraction Transformation Loading)
- Integrační nástroje (EAI : Enterprise Application Integration)
- Datové sklady
- Datová tržiště
- OLAP
- Reportingové nástroje
- EIS (Executive Information Systém)
- Data Mining



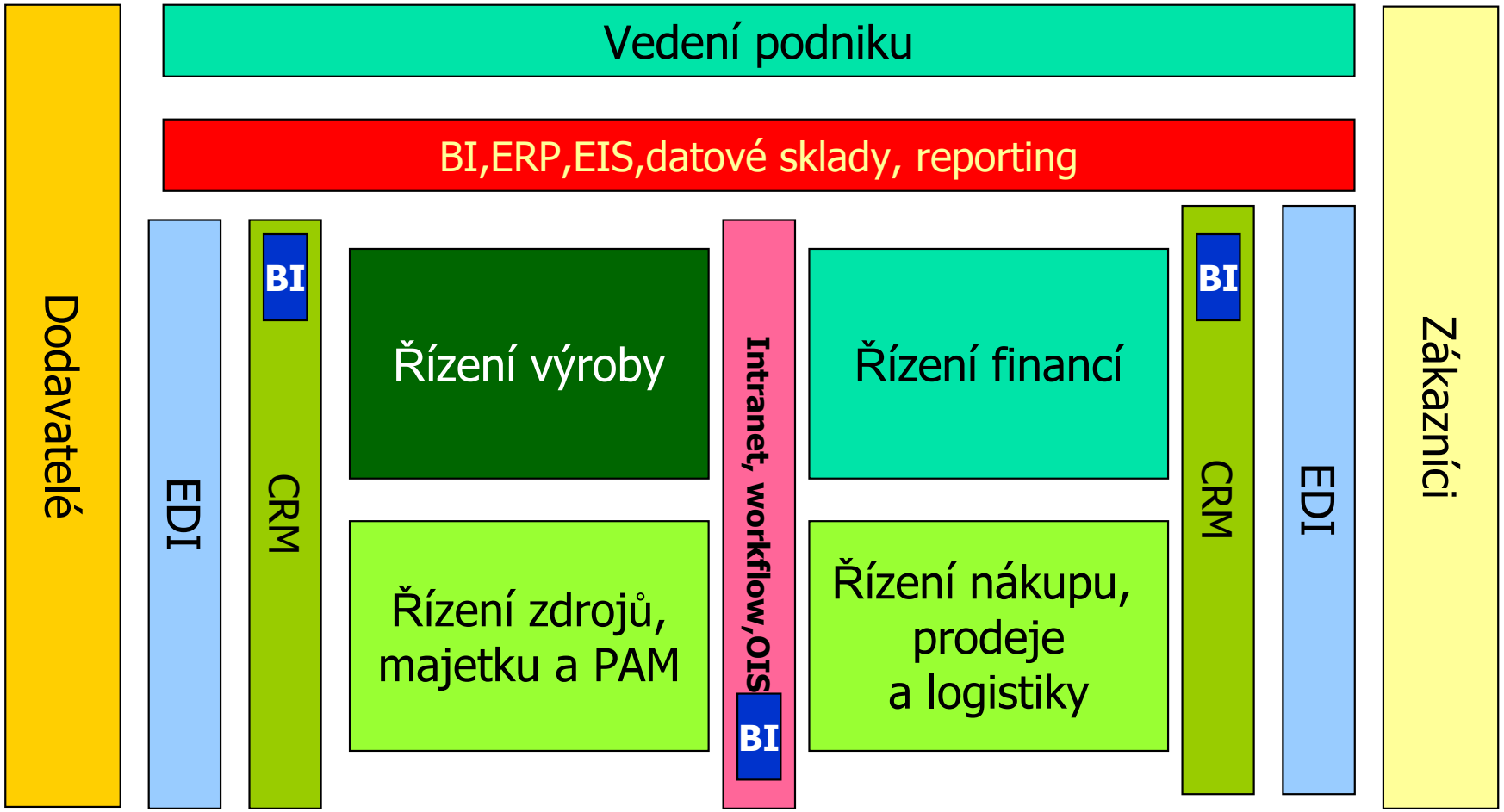
Omezení ERP jako poskytovatele dat

- Neumožňují rychle a pružně měnit kriteria výběru
- Okamžitý přístup uživatelů k velkým objemům agregovaných dat
- ERP jsou primárně určeny k pořizovaní dat a jejich aktualizaci
- V každém podniku se objem dat za každých pět let zdvojnásobí, což ovšem také znamená, že systém je zahlcen redundantními daty
- Vícedimenzionální pohled na data v ERP je problematický. DB ERP není pro tento pohled stavěná. Databáze, které vzniknou přeměnou primárních dat z ERP a jsou využívány např. OLAP technologií jsou pro *drilling* a *slice* operace optimalizovány

Zjednodušené schéma využívání ERP

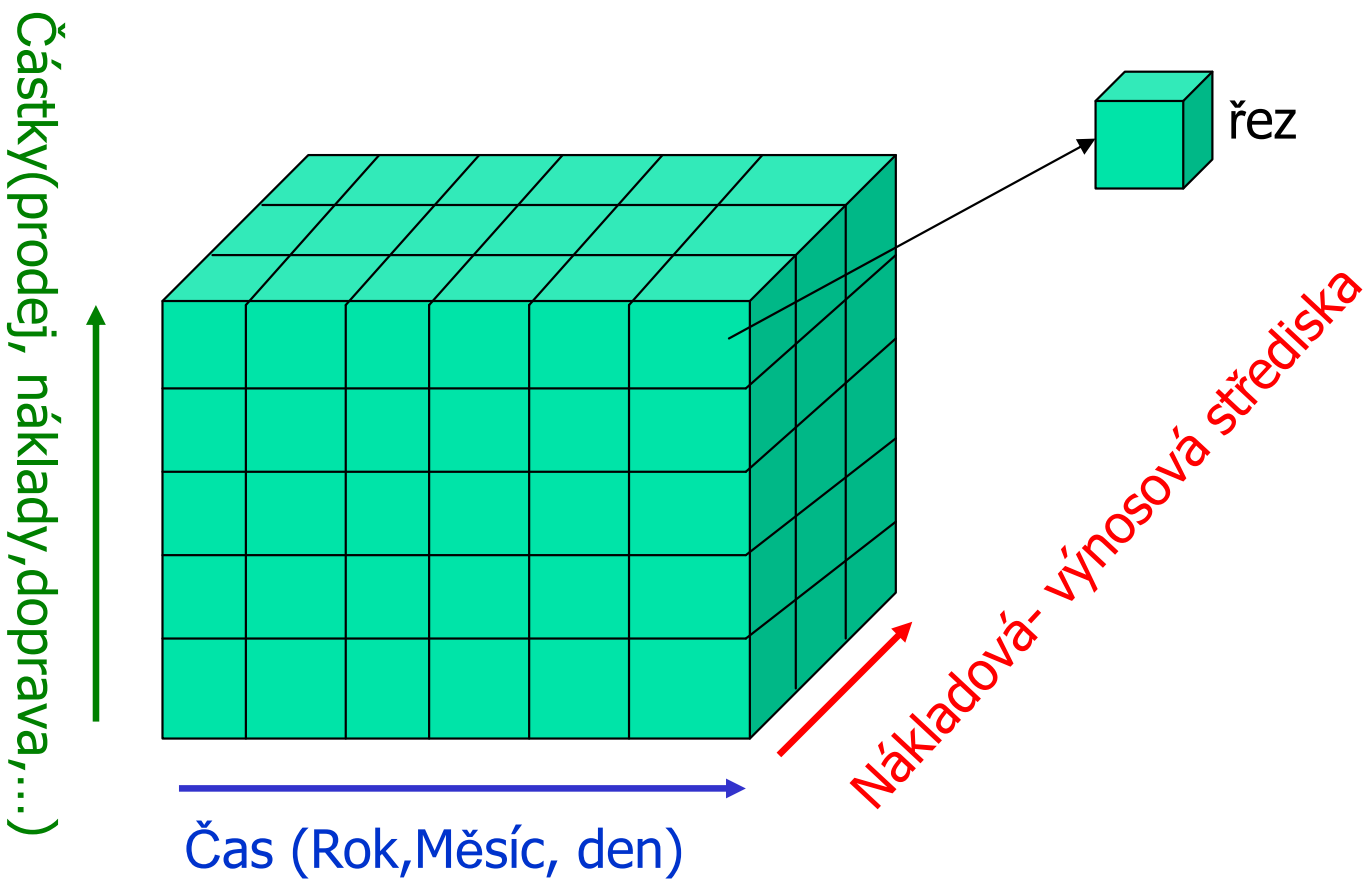


Organizačně-technologické schéma podniku

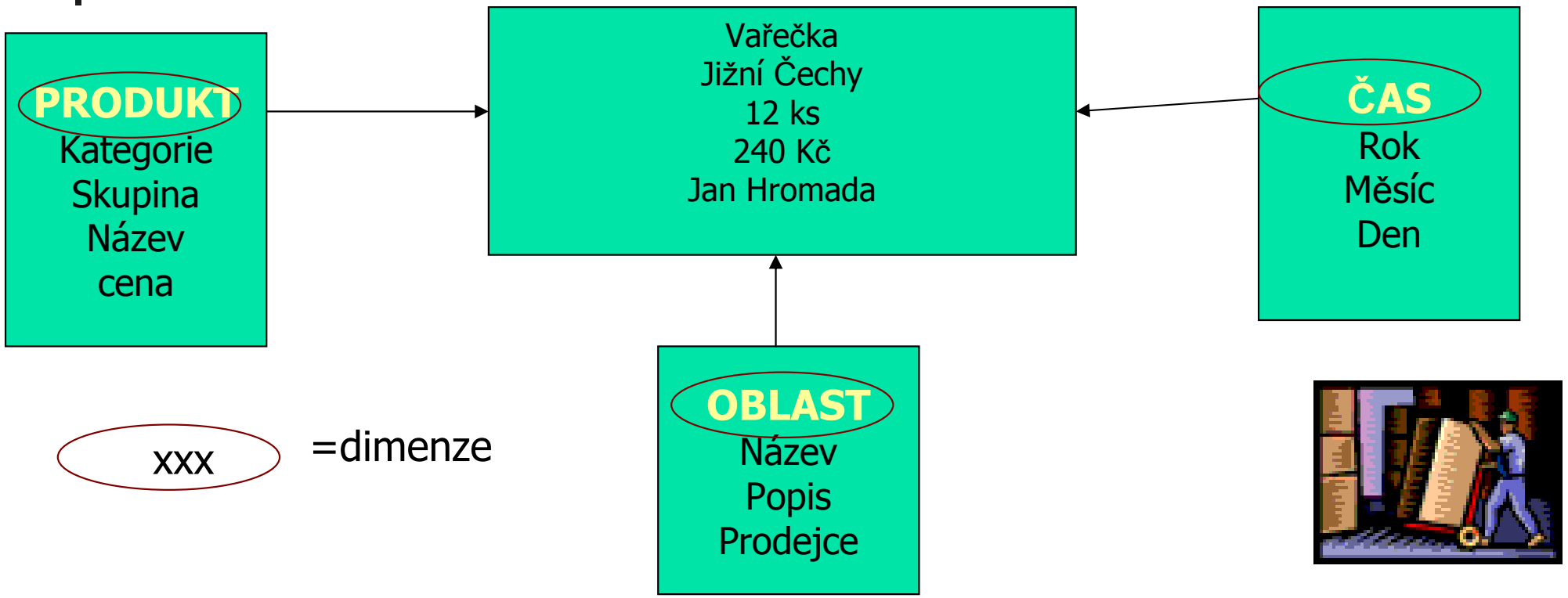


OLAP kostka

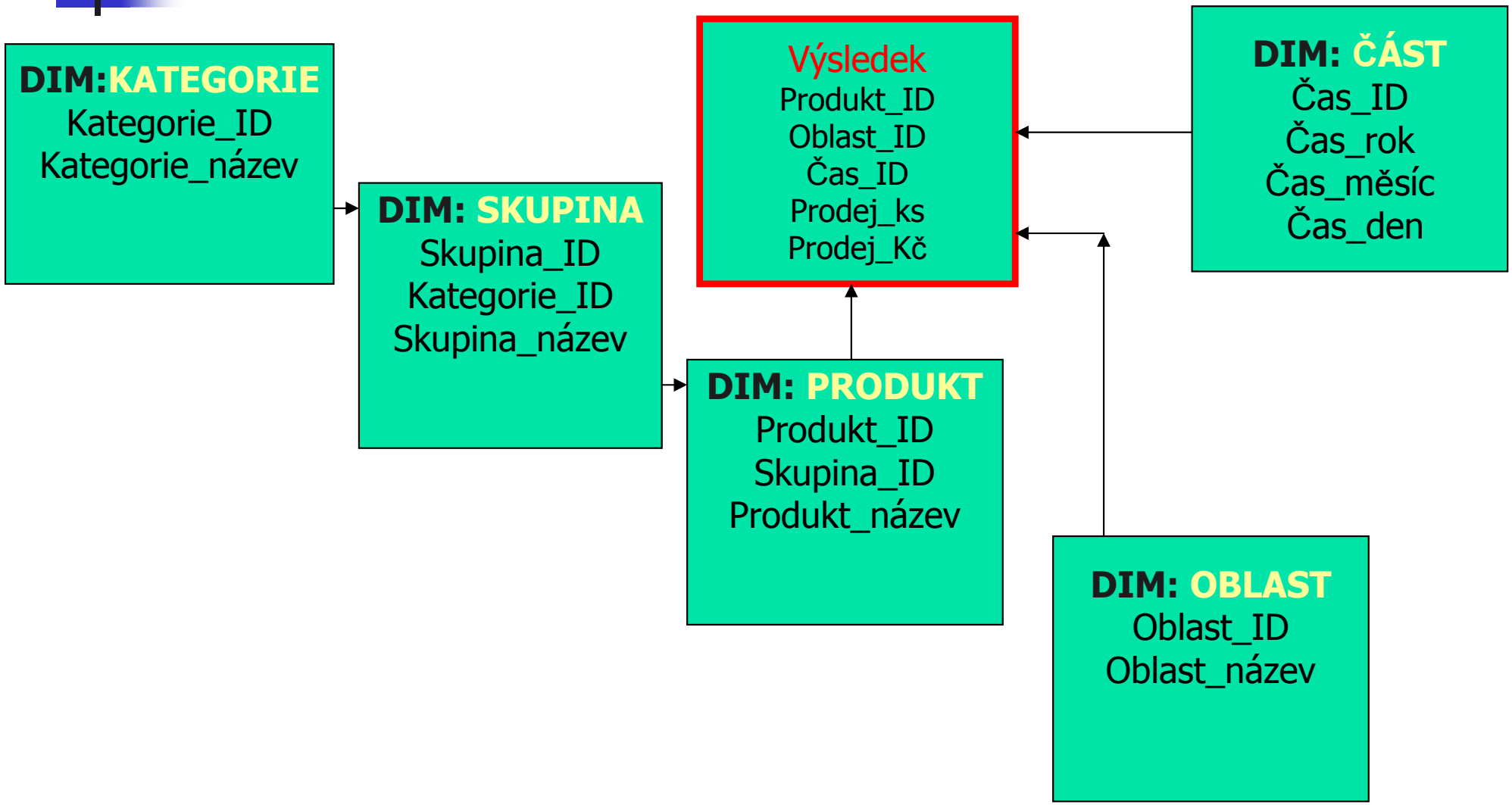
- http://www.databaseanswers.org/designing_olap_cubes.htm



Relační dimenzionální model: STAR



Relační dimenzionální model: SNOWFLAKE



Datový sklad

Transakce

Podniková DB

Zákazníci

Zakázky

Dodavatelé

Atd.…

Atd.…

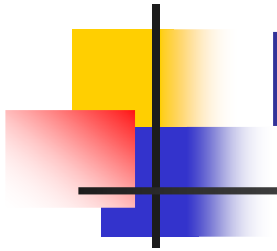
Kopie, ,
organizace dat
Sumarizace dat

Datový sklad

Dolování dat

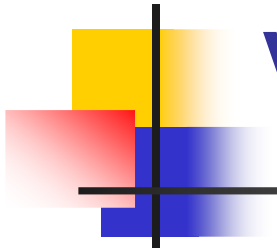
Datoví horníci :

- “Profíci” – vědí co chtějí
- “Výzkumníci” – nepředvídané výsledky



Definice

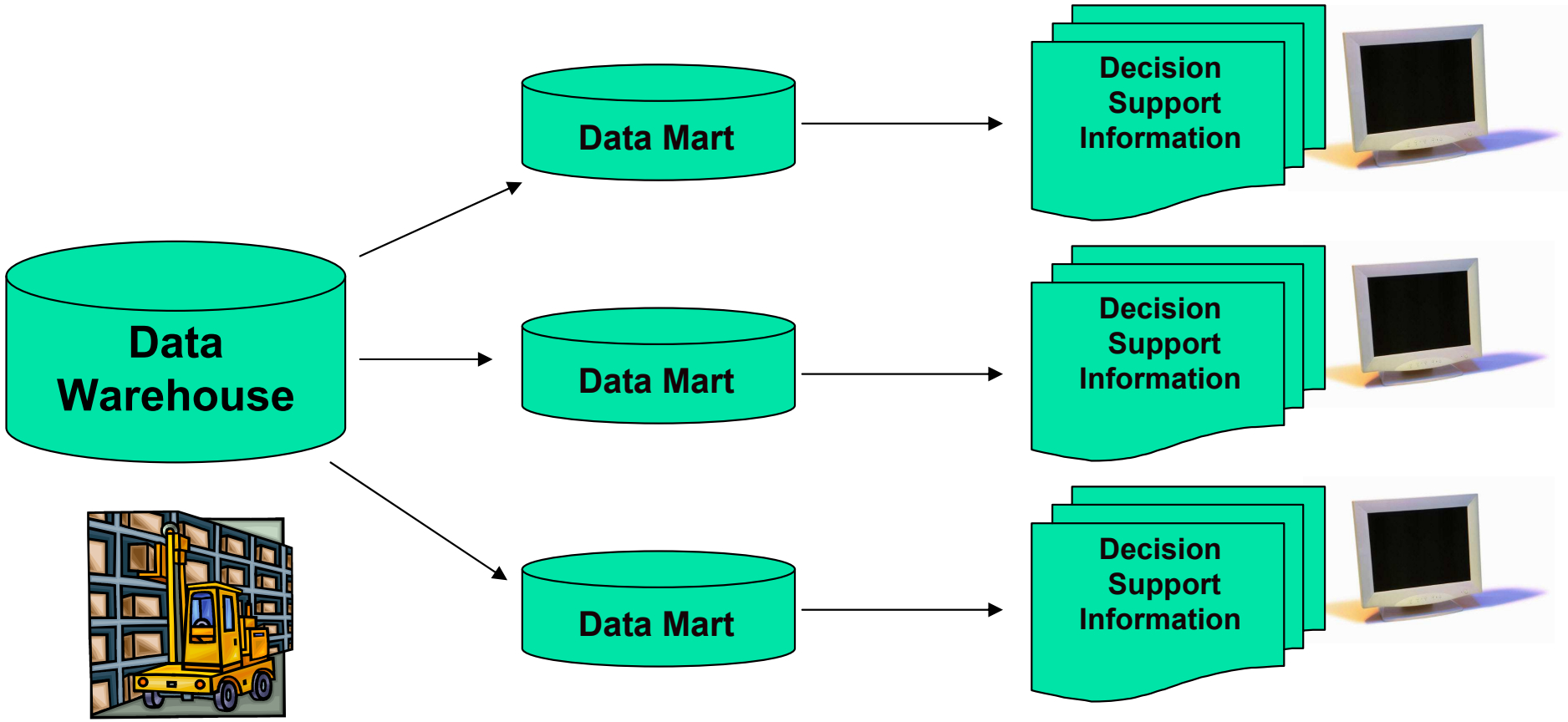
- Datový sklad: základní komponenta BI
- Datové tržiště : subjektově orientované analytické DB- součást datového skladu
- Operativní datová úložiště : podpůrné analytické DB
- Dočasná úložiště dat : úložiště dat před jejich zpracování do databázových komponent řešení BI

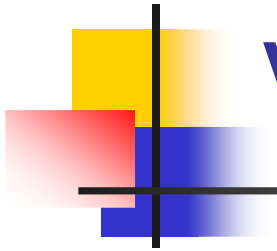


Vrstvy pro analýzu dat

- Reporting : ad hoc dotazovací proces do DB komponent BI
- OLAP : pokročilé a dynamické analytické úlohy
- Data Mining (dolování dat) : sofistikovaná analýza většího množství dat
- Algoritmy pro dolování dat :
  - rozhodovací stromy
  - Neuronové sítě
  - Clustering a klasifikace

Datový sklad->datové tržiště (anglická verze)



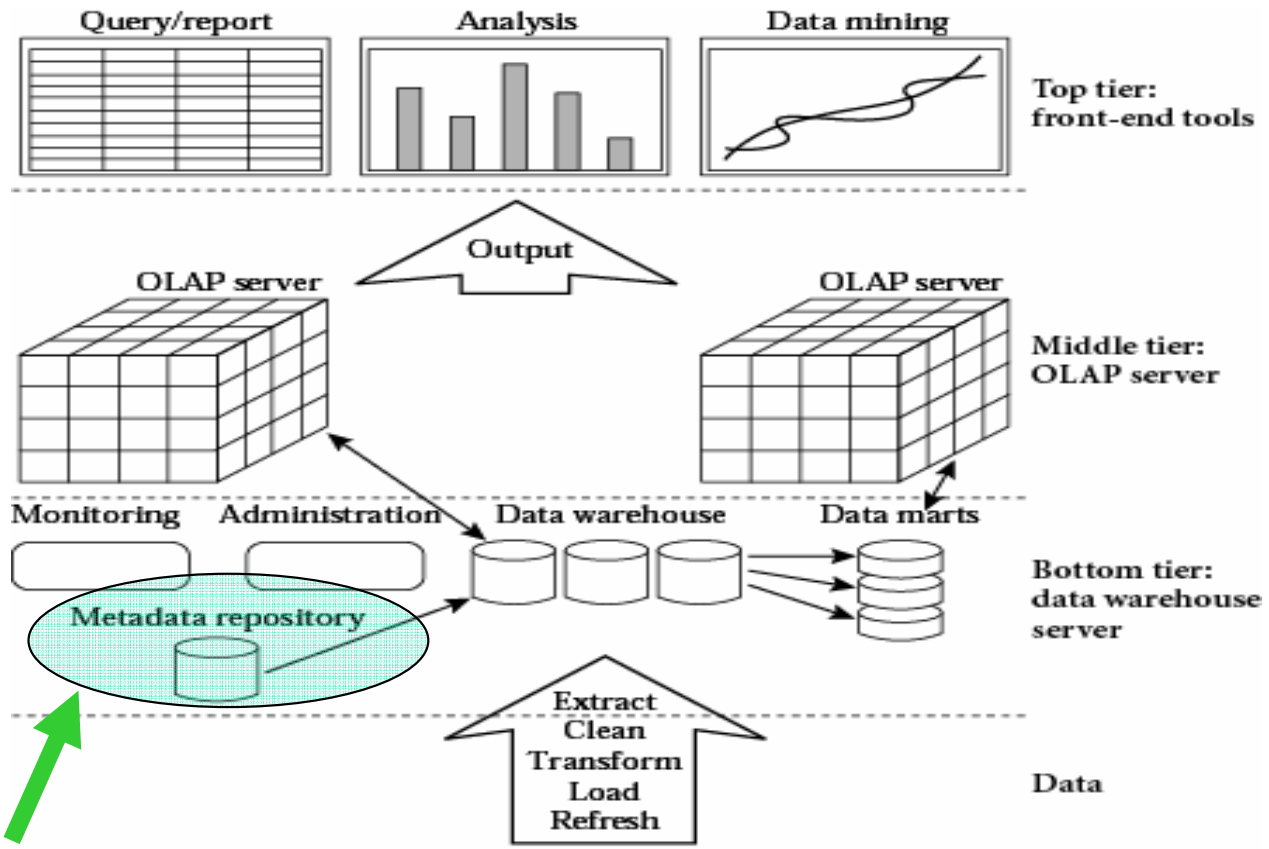


Vysvětlení pojmu METADATA

Metadata jsou data o datech, kde pomocí předem definovaných dat s jasně danou a popsanou strukturou uchováváme informace o jiných datech.

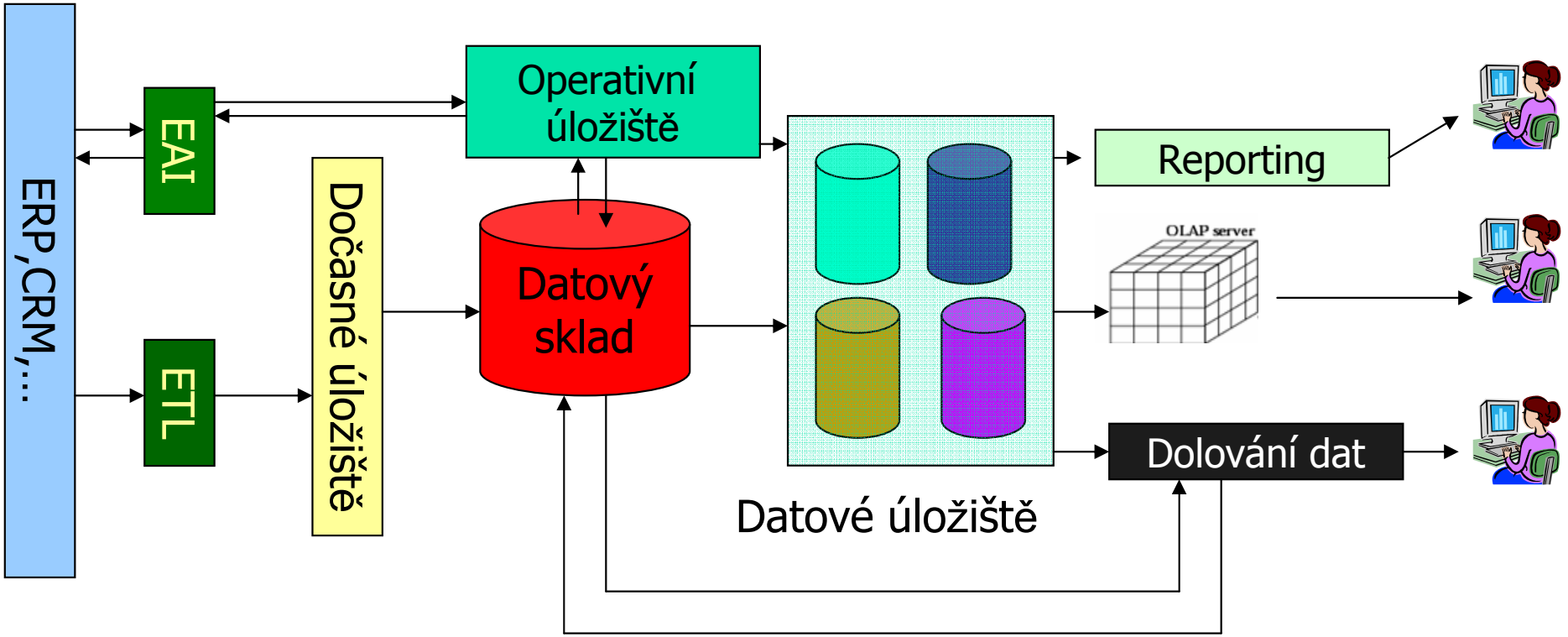
Typickým příkladem metadat jsou katalogizační záznamy v knihovnách, což byla jejich původní funkce.

Architektura OLAP (anglická verze)



METADATA
viz definice
na předchozím snímku

Hlavní komponenty BI a jejich vazby



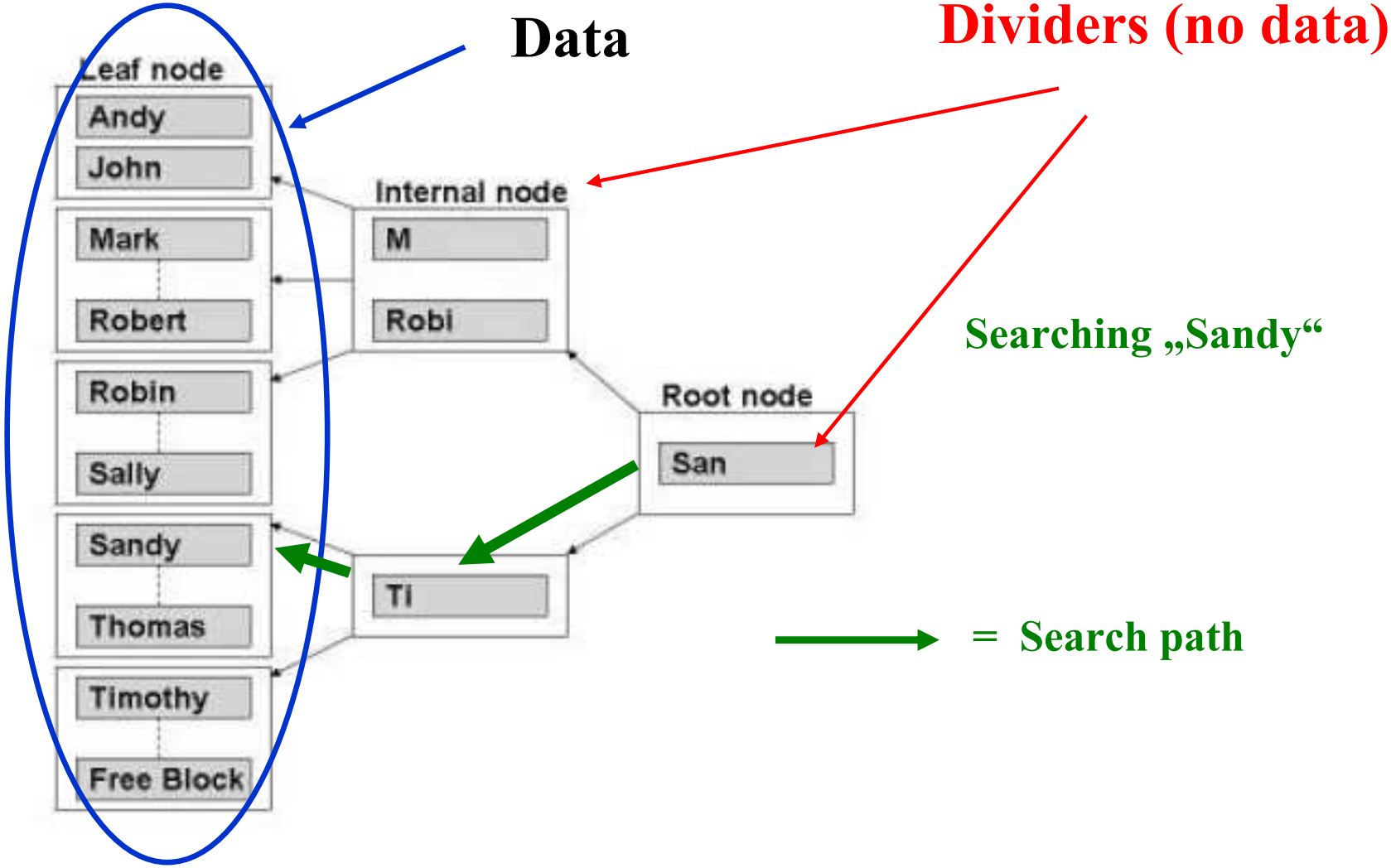
Dolování dat

- Rozhodovací stromy
- Neuronové sítě
- Genetické algoritmy
- Clustering a klasifikace

Dolování dat

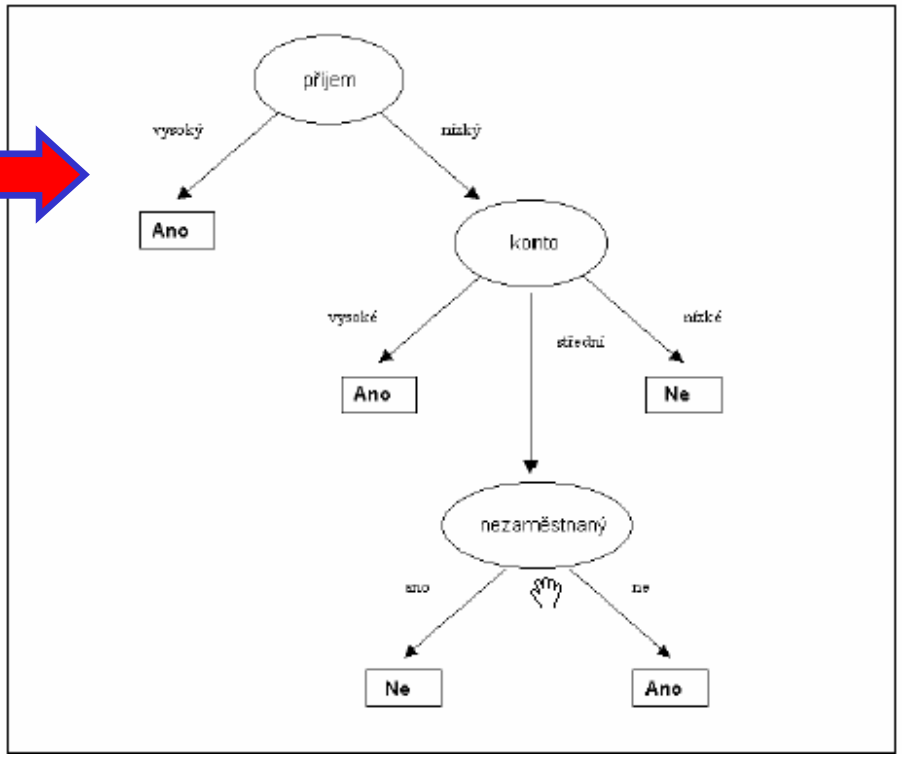
- Rozhodovací stromy (RS) - prediktivní model, který se zobrazuje v podobě stromu, kde každý uzel určuje kritérium pro následní rozvětvení. Strom rozděluje veškerá zdrojová data do segmentů, kde každý list odpovídá určitému segmentu definovanému předešlými uzly.Data v jednom segmentu mají shodné vlastnosti.

B + tree-jeden z příkladů RS



Příklad vytvoření RS

klient	příjem	konto	pohlaví	nezaměstnaný	úvěr
k1	vysoký	vysoké	žena	ne	ano
k2	vysoký	vysoké	muž	ne	ano
k3	nízký	nízké	muž	ne	ne
k4	nízký	vysoké	žena	ano	ano
k5	nízký	vysoké	muž	ano	ano
k6	nízký	nízké	žena	ano	ne
k7	vysoký	nízké	muž	ne	ano
k8	vysoký	nízké	žena	ano	ano
k9	nízký	střední	muž	ano	ne
k10	vysoký	střední	žena	ne	ano
k11	nízký	střední	žena	ano	ne
k12	nízký	střední	muž	ne	ano



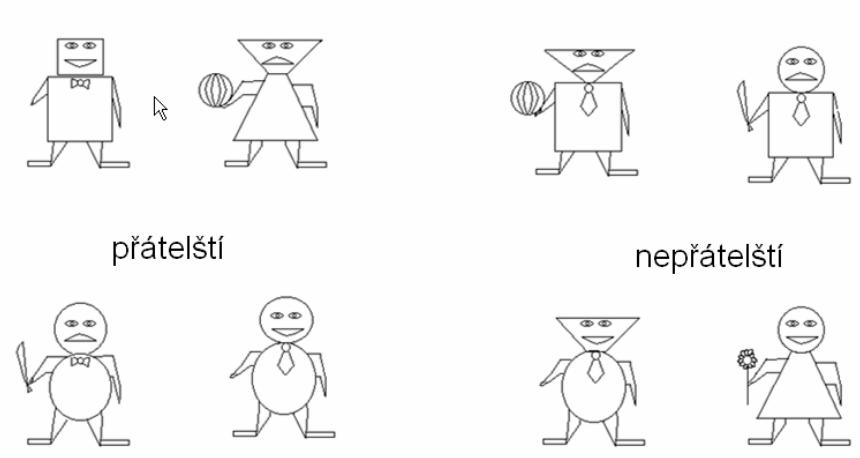
Typy stromů :

a) CART=Classificaion and Regression Trees (kriterium redukce směrodatné odchylky)

b) CHAID =Chi-squared Automatic Interaction Detector

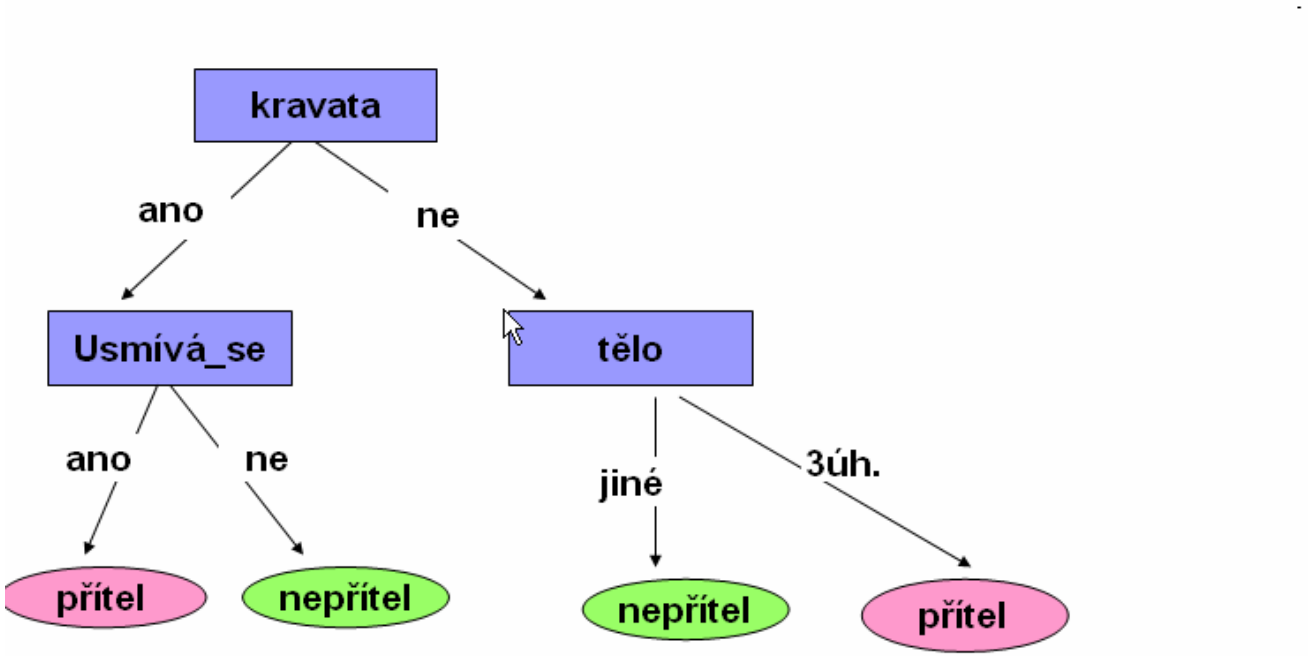
http://lisp.vse.cz/~berka/docs/izi456/SL-IDT.PDF

Rozdělení postaviček podle atributů



Hlava	Úsměv	Ozdoba	Tvar těla	Předmět	Přátelský
Kruh	Ne	Kravata	Čtverec	Šavle	NE
Čtverec	Ano	Motýlek	Čtverec	NIC	ANO
Kruh	Ne	Motýlek	Kruh	Šavle	ANO
Trojúhelník	Ne	Kravata	Čtverec	Balon	NE
Kruh	Ano	NIC	Trojúhelník	Květina	NE
Trojúhelník	Ne	NIC	Trojúhelník	Balon	ANO
Trojúhelník	Ano	Kravata	Kruh	NIC	NE
Kruh	Ano	Kravata	Kruh	NIC	ANO

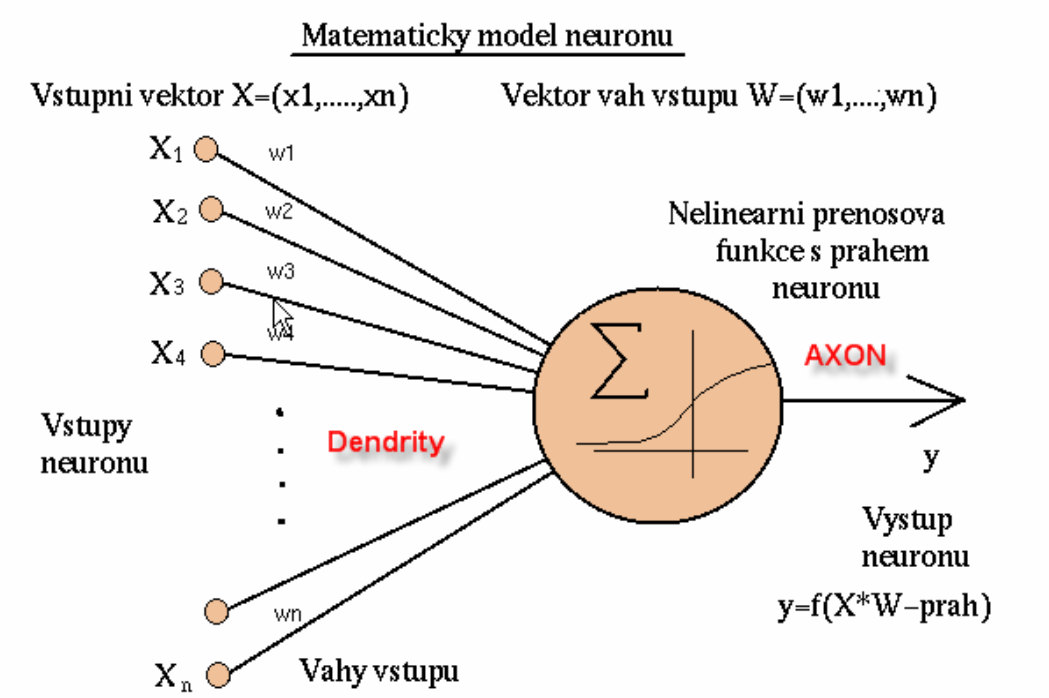
Rozhodovací strom jako logický výraz

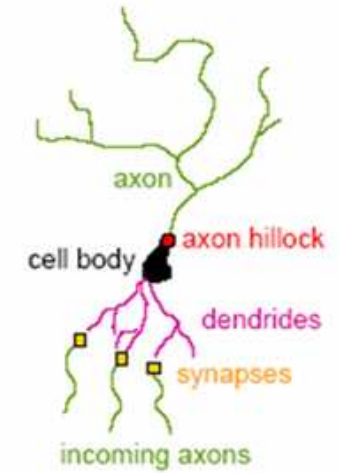


(Kravata=ano & usmívá_se=ano) **V** (Kravata=ne & tělo=3úh.)

Neuronové sítě

- Neuronové sítě (NS) - užívané pro tvorbu prediktivních modelů, Jsou založeny na obdobných principech, které napodobují organizaci nebo způsob chování lidského mozku, založeném na systému neuronů.





Synapse je vazba a má dva typy : Excitační (vybuzující) a Inhibiční (tlumící)

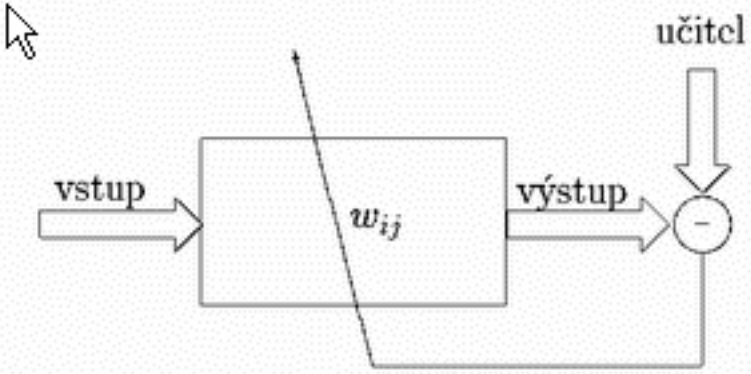
Učení neuronových sítí

Učení neuronových sítí

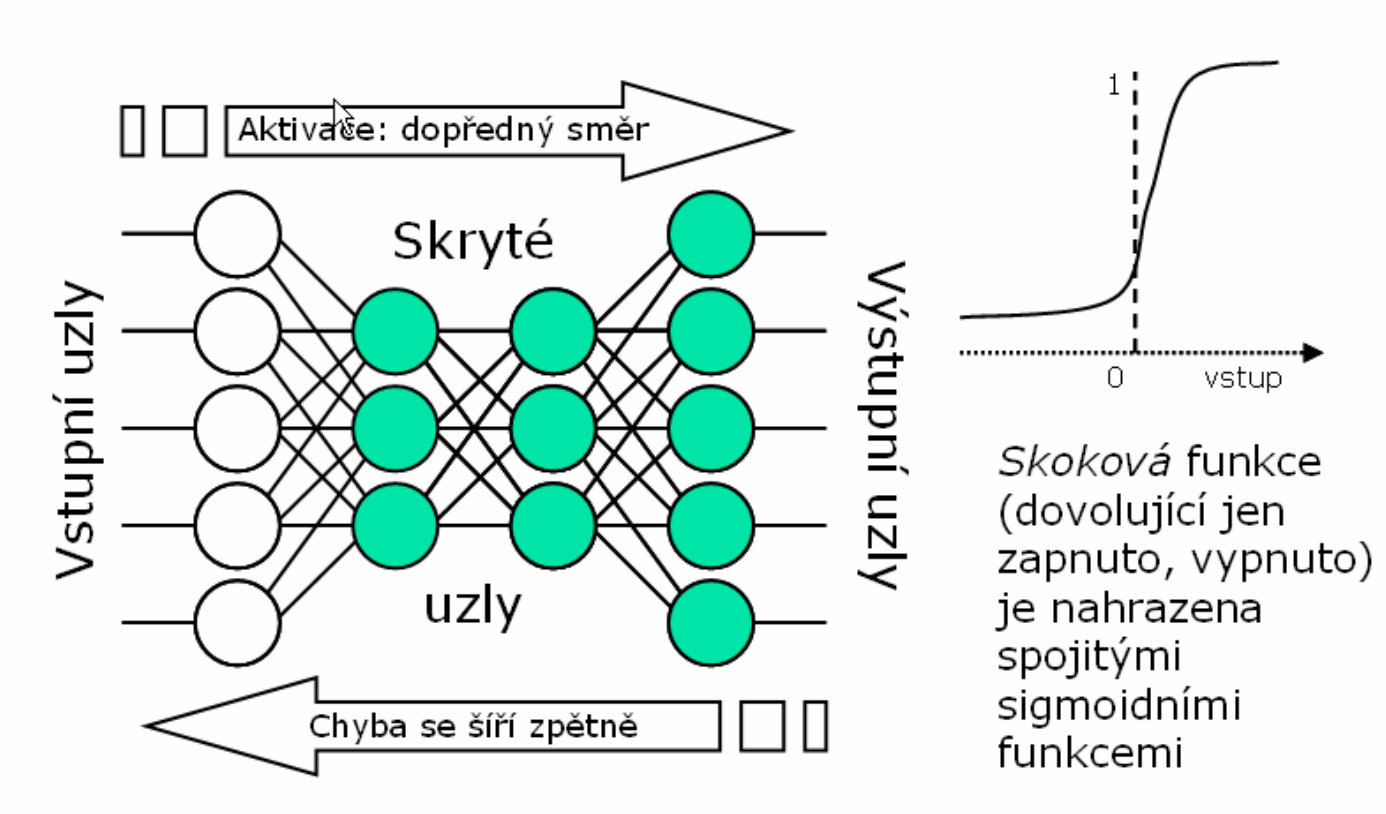
Cílem učení je nastavit váhy spojení $w_{i,j}$ tak, aby síť vytvářela správnou odezvu na vstupní signál.

Základní způsoby učení:

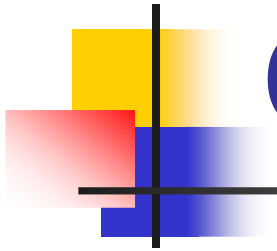
- **učení s učitelem (supervised learning)**
  Neuronová síť se učí srovnáváním aktuálního výstupu s výstupem požadovaným (učitel) a nastavováním vah synapsí tak, aby se snížil rozdíl mezi skutečným a požadovaným výstupem.



Vícevrstvé neuronové sítě



Skoková funkce (dovolující jen zapnuto, vypnuto) je nahrazena spojitými sigmoidními funkcemi



OLAP databáze

- **OLAP** DB představují jednu nebo více souvisejících OLAP kostek
- **OLAP** kostka na rozdíl od datových skladů zahrnuje předzpracované agregace dat podle definovaných hierarchických struktur dimenzí a jejich kombinací
- Technologie **OLAP** má několik variant (uvádím zde pouze dvě z nich):
  - **MOLAP** - Multidimensional OLAP (speciální uložené v multidimenzionálních-binárních kostkách)
  - **ROLAP** – Relational OLAP (uloží data do relační DB)

Datová pumpa

Primární transakční systém (ERP,CRM,..)

Datová pumpa

Datový sklad

Datová pumpa (kritické místo celé aplikace)

Datová pumpa, nebo-li ETL nástroj umožňuje efektivní zpracování velkých objemů z různých zdrojů a jejich uložení do datového skladu. Každý ETL nástroj musí umět:

a) zpracovávat různorodá data obvykle fyzicky umístěná na různých místech,
b) navrhovat transformace pro přenos dat mezi různými datovými formáty

Zpracování = odstranění redundancí, agregace podle dimenzí, zapomínání dat
Zapomínání dat = úmyslné odstranění nepotřebných dat z datového skladu

Datová pumpa (kritické místo celé aplikace)

Datová pumpa = Extraction Transformation and Loading = ETL