

Datové sklady

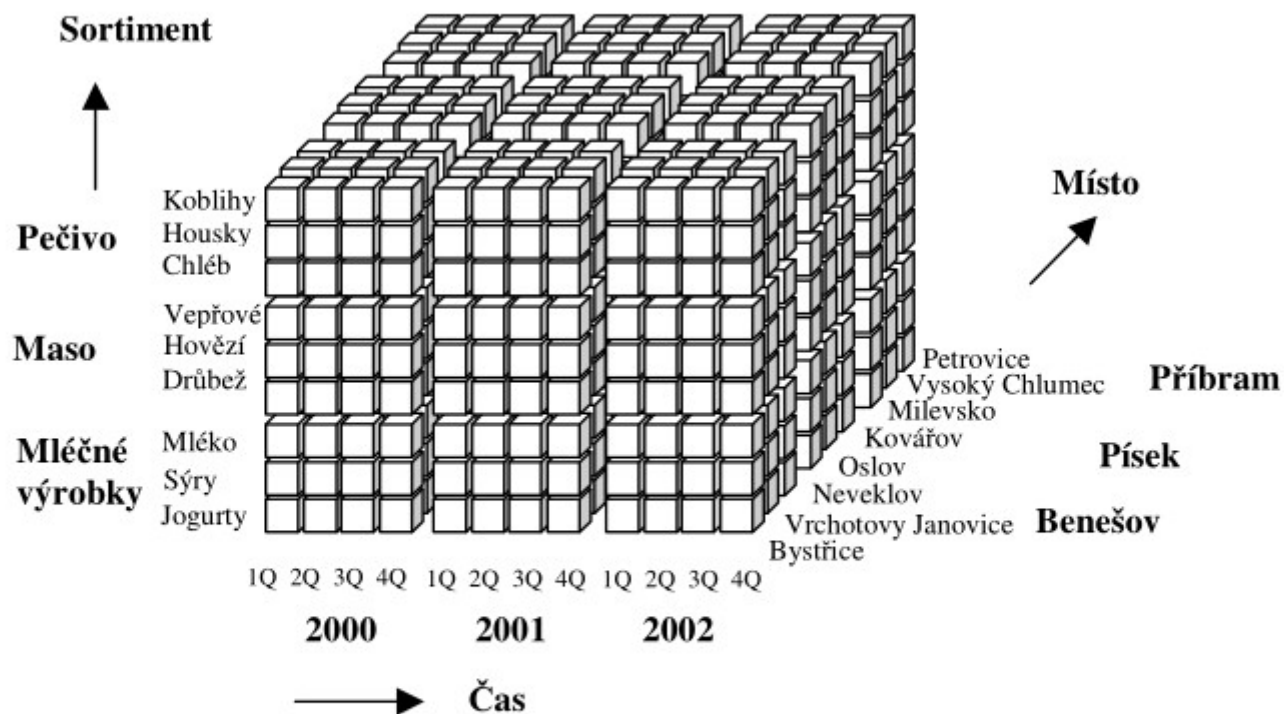
Operace s daty

- **OLTP** (*OnLine Transaction Processing*)
 - Hlavní poslání: bezchybné zpracování transakcí (zápisů)
 - ACID, 3NF, atp.
 - Vedlejší poslání: čtení dat z databáze (SELECT)
 - „provozní systém“
- **OLAP** (*OnLine Analytical Processing*)
 - Hlavní poslání: rychlé zpracování reportů
 - Optimalizace i za cenu duplicity dat (běžně nemusí být v 3NF)
 - „manažerský systém“

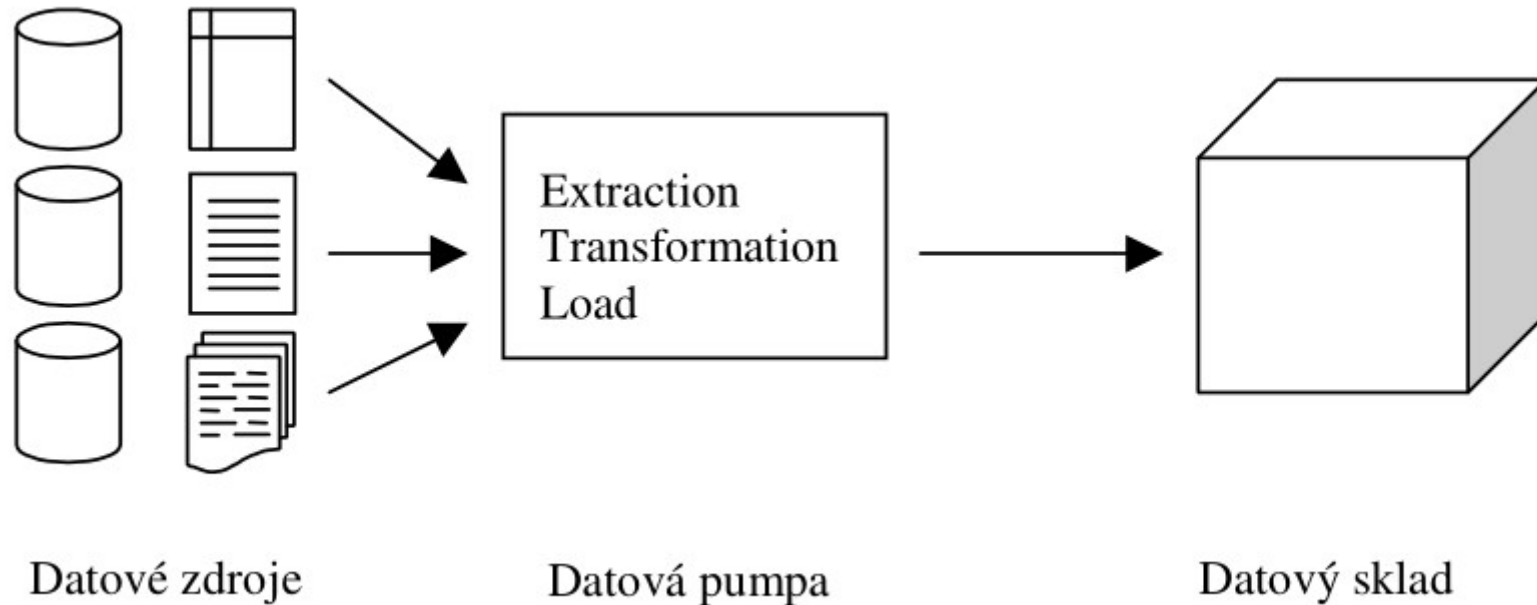
OLAP – Datová krychle

- **Fakta** (faktová tabulka)
 - Obsahuje hodnoty
(typicky vzniklé agregací zdrojových dat)
- **Dimenze** (dimenzionální tabulky)
 - Reprezentují dimenze, tj. popisy objektů, ke kterým se konkrétní fakt váže

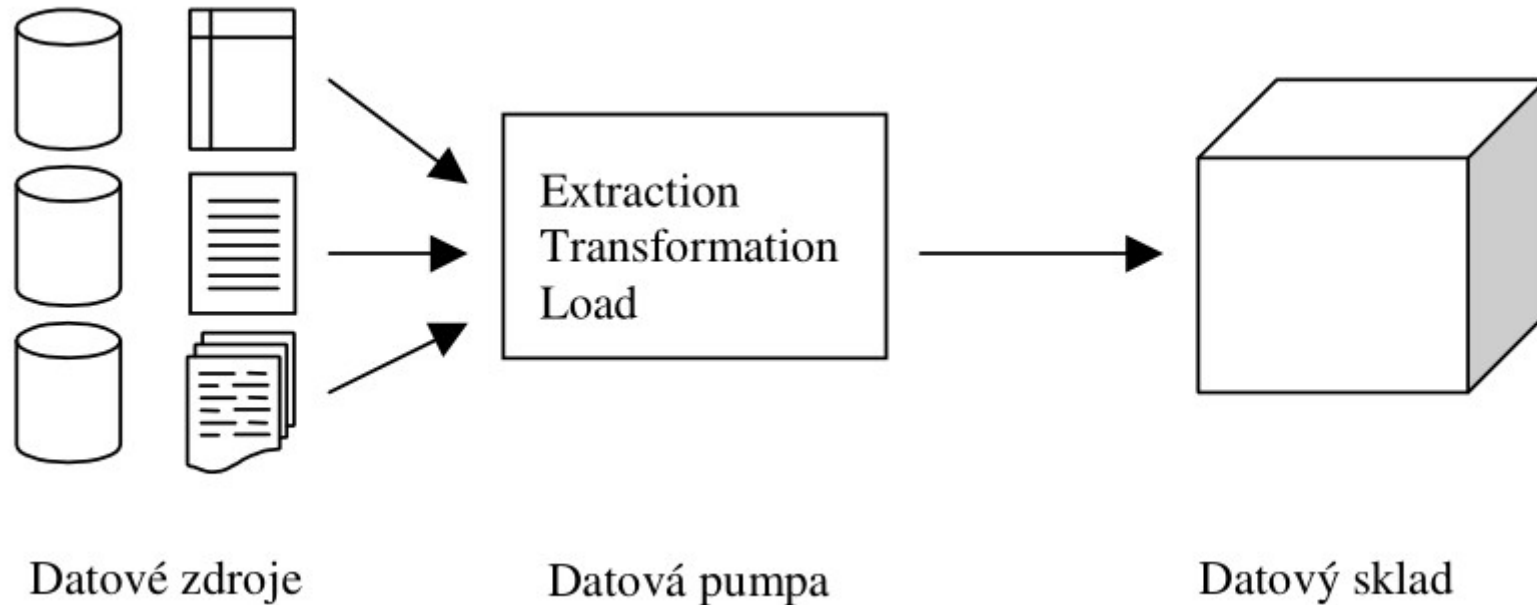
OLAP – Datová krychle



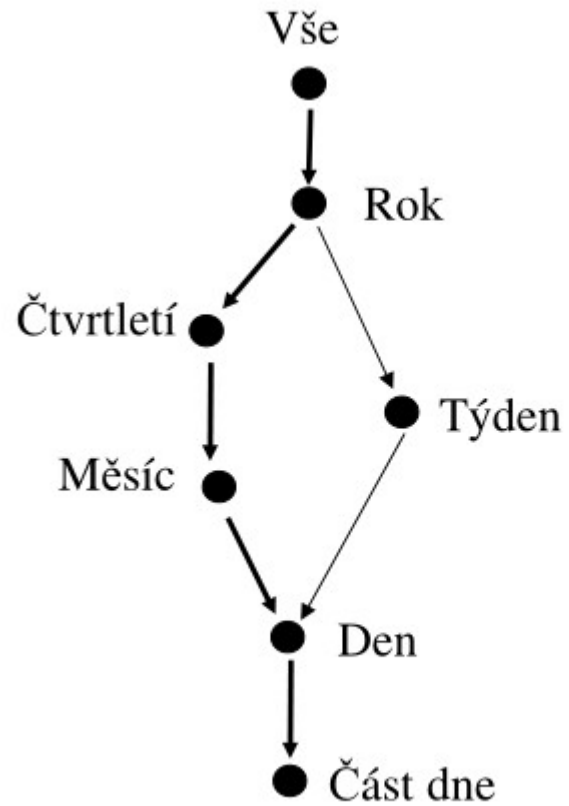
OLAP – Datová pumpa



OLAP – Datová pumpa



OLAP – Příklad dimenze



OLAP – operace

- **Kostkování**

- Omezení dimenzionálního atributu na množinu hodnot
- Speciální případ: **krájení**
 - Omezení na jednu hodnotu

- **„Roll up“ a „Drill down“**

- Pohyb na vyšší/nížší stupeň
 - např. rok => kvartál

- **Pivotování**

- Změna pořadí dimenzionálních atributů
 - Hledání závislostí a trendů

| Zboží | Rok | Zisk |
|----------|------|------|
| Drogerie | 2015 | 40 |
| | 2016 | 25 |
| Pečivo | 2016 | 10 |

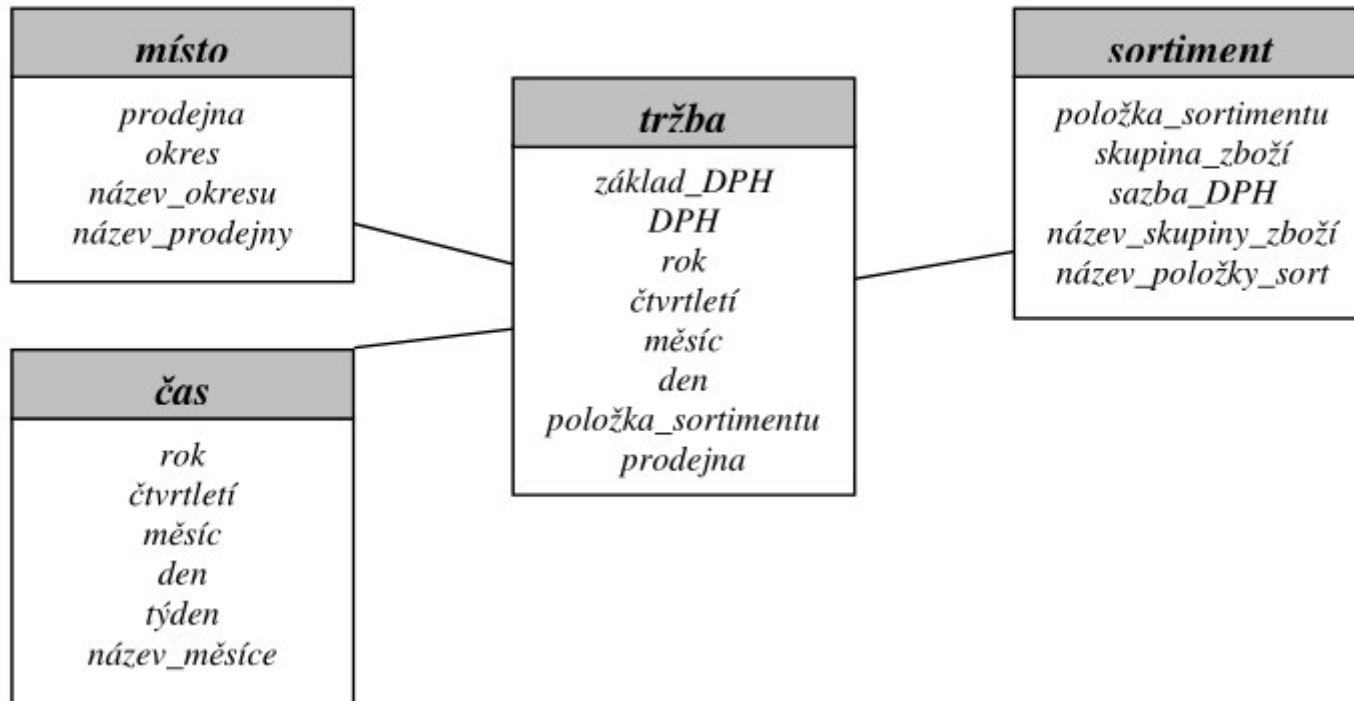
| Rok | Zboží | Zisk |
|------|----------|------|
| 2016 | Pečivo | 10 |
| | Drogerie | 25 |
| 2015 | Drogerie | 40 |

OLAP – star schema

- Typicky:
 - 1 faktová tabulka
 - Pro každou dimenzi jedna dimenzionální tabulka
 - Výhody: Malý počet tabulek, rychlé vyhledávání
 - Nevýhody: Redundantně uložená dimenzionální data (2NF)

| ID | Rok | Kvartál | Měsíc | Den | DenTyd |
|----|--------|---------|--------|--------|--------|
| #1 | 2016 | 2 | 5 | 3 | Út |
| #2 | 2016 | 2 | 5 | 4 | St |
| #3 | 2016 | 2 | (null) | (null) | (null) |
| #4 | (null) | (null) | (null) | (null) | St |

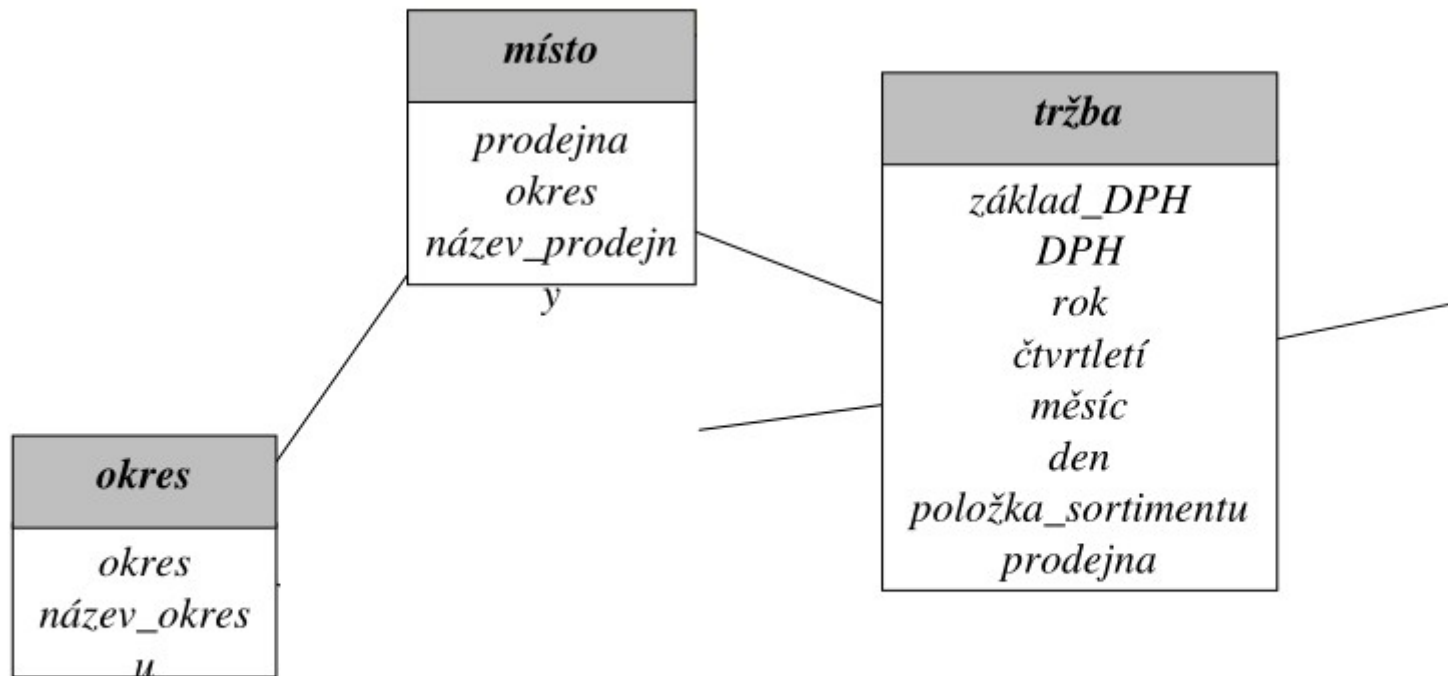
OLAP – star schema



OLAP – snowflake schema

- Typicky:
 - 1 faktová tabulka
 - Dimenzionální tabulky splňují 3NF
 - Výhody: Neredundantní uložená dimenzionální data
 - V principu lze kombinovat s OLTP systémem
 - Nevýhody: Nutnost počítat agregace pro jiné než základní dimenze

OLAP – snowflake schema



OLAP – konsolidované snowflake schema

- Typicky:
 - Pro každou kombinaci nezávislých dimenzionálních tabulek je 1 faktová tabulka
 - Výhody:
 - Neredundantní uložená dimenzionální data
 - Předpočítané agregace pro všechny dimenze
 - Nevýhody: mnoho faktových tabulek, vyžaduje netriviální navigaci

OLAP – konsolidované snowflake

