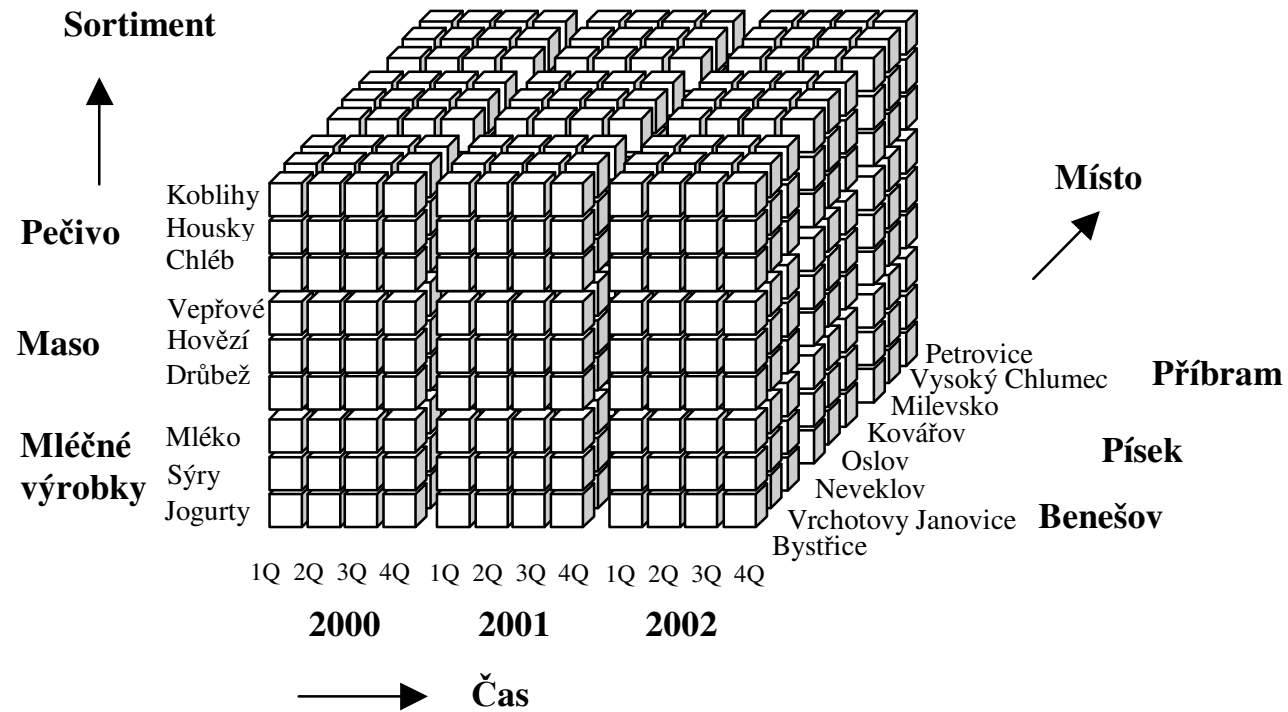


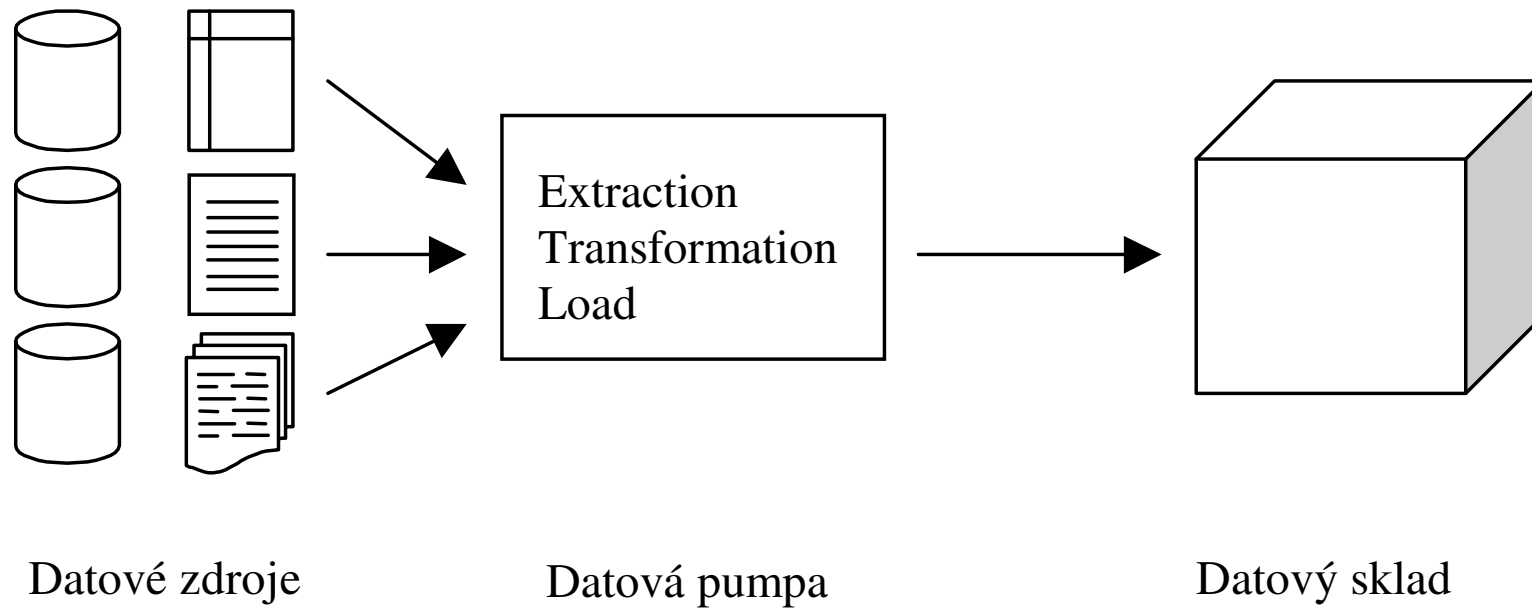
Datové sklady

Zdeněk Kouba

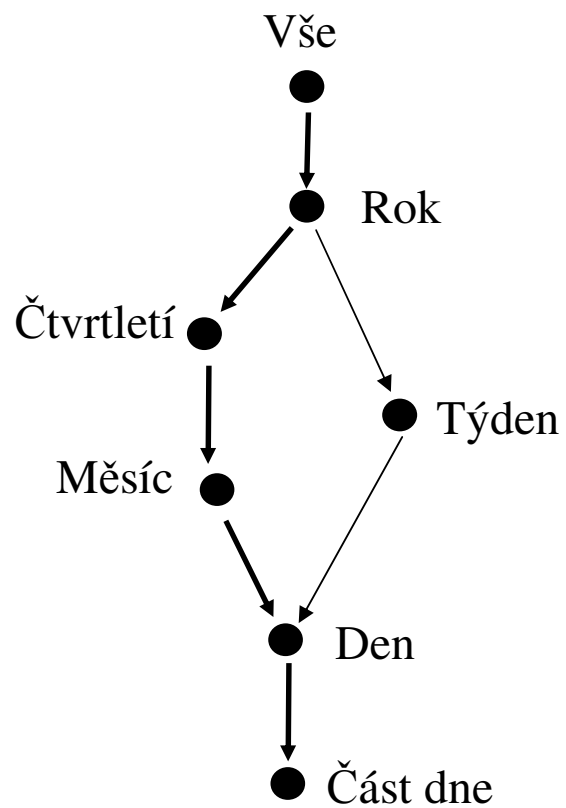
Data cube



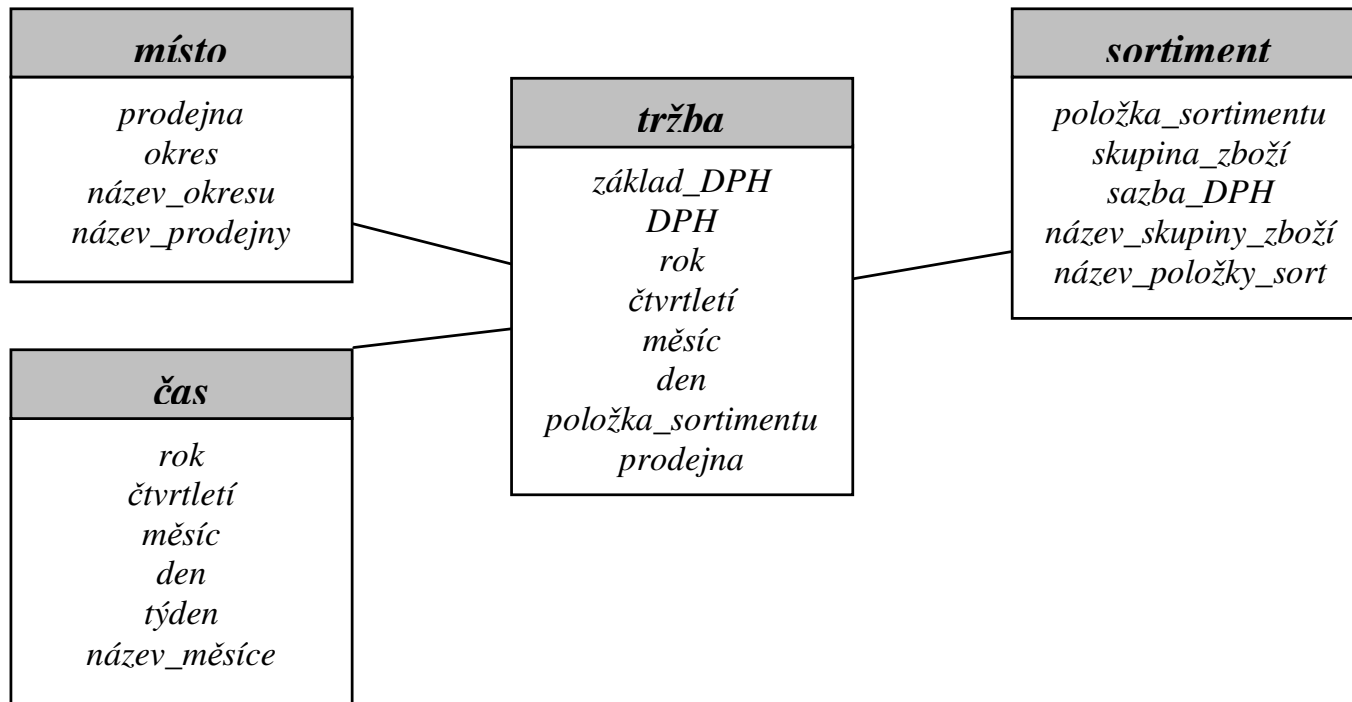
Proces ETL



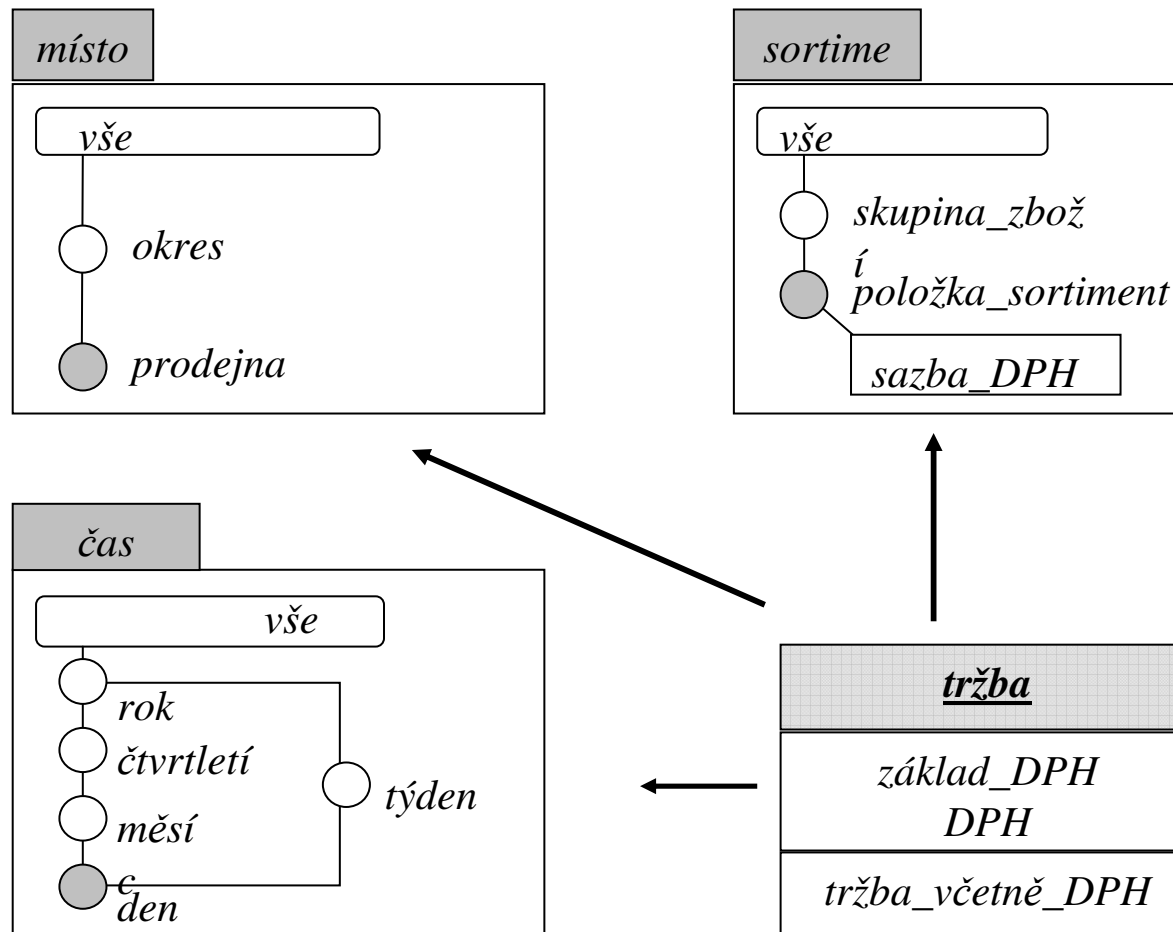
Agregační hierarchie dimenze “čas”



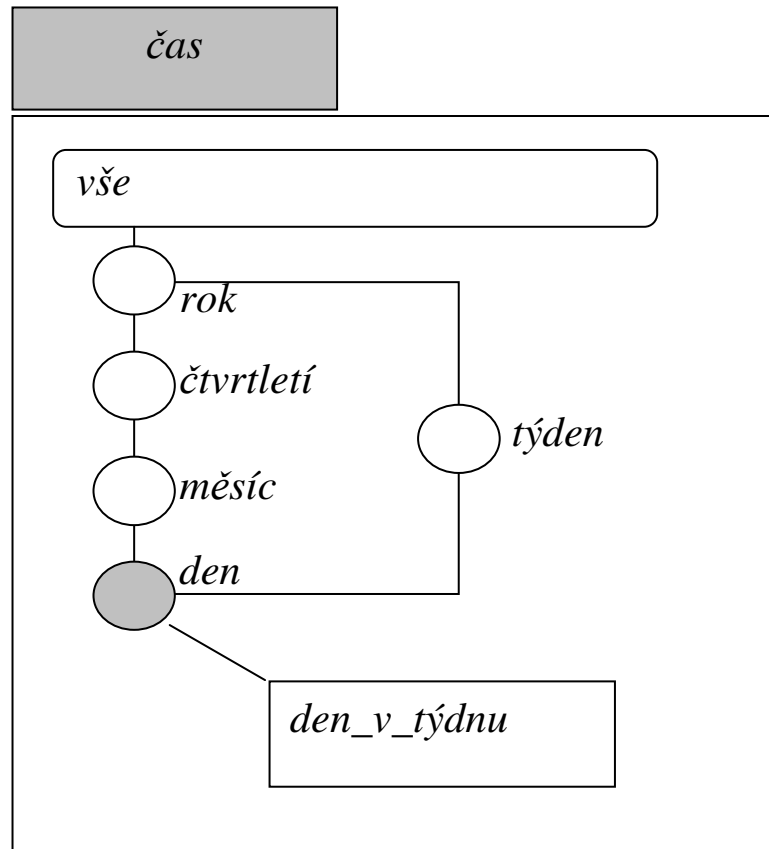
Star schema



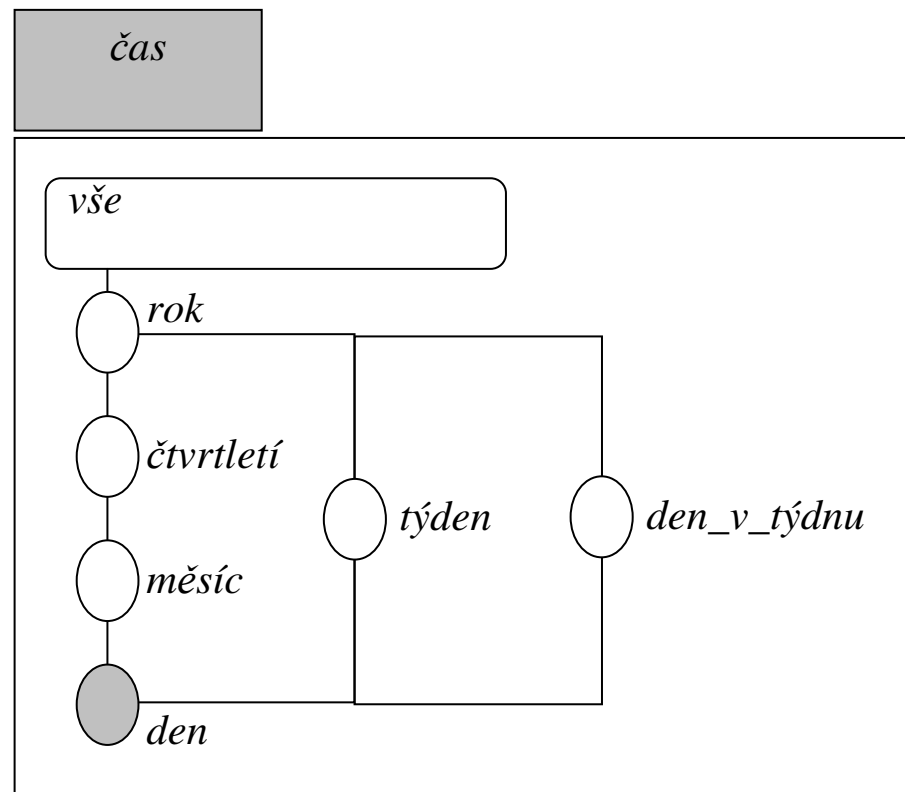
Konceptuální model motivačního příkladu



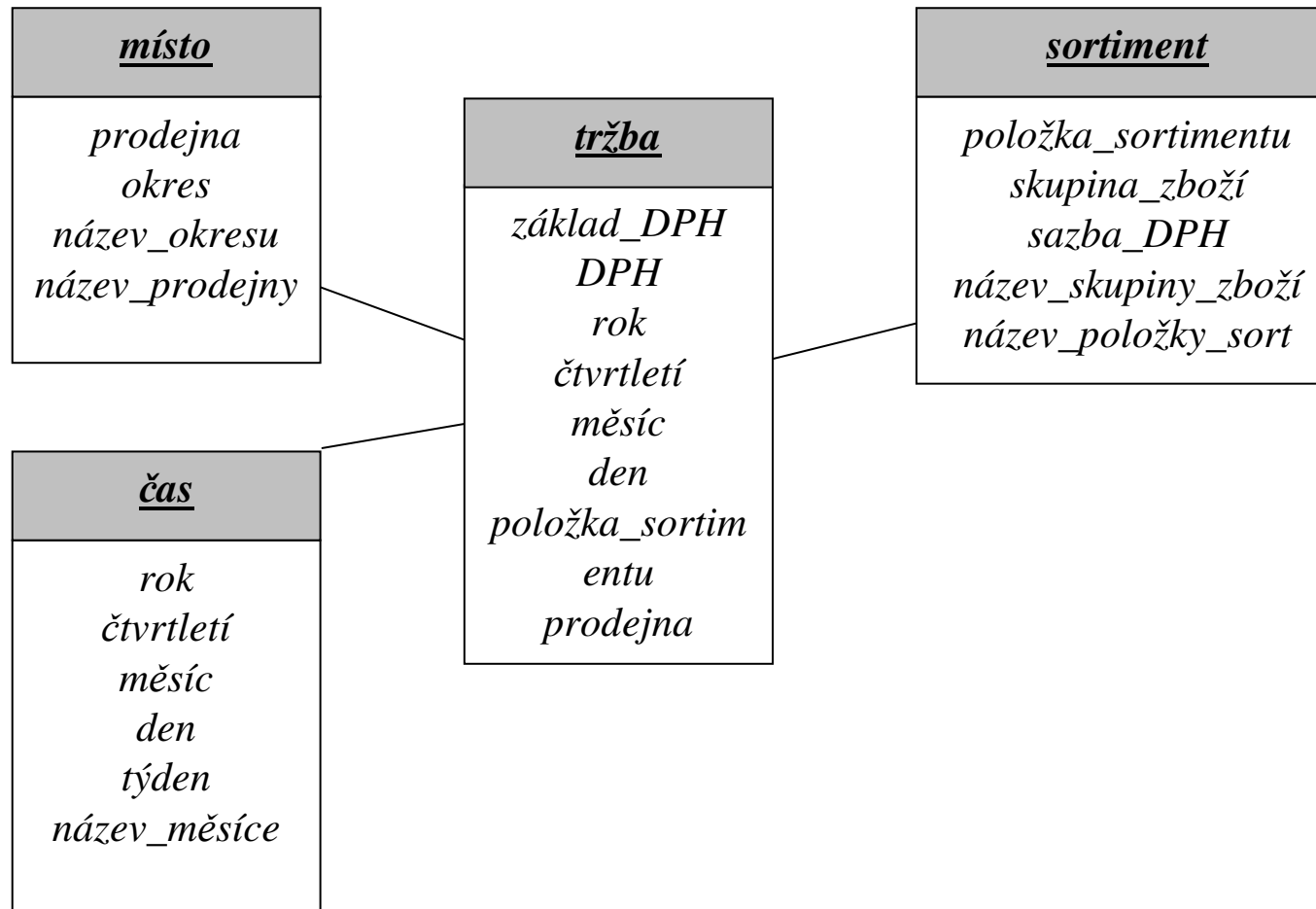
Den v týdnu jako dimenzionální atribut



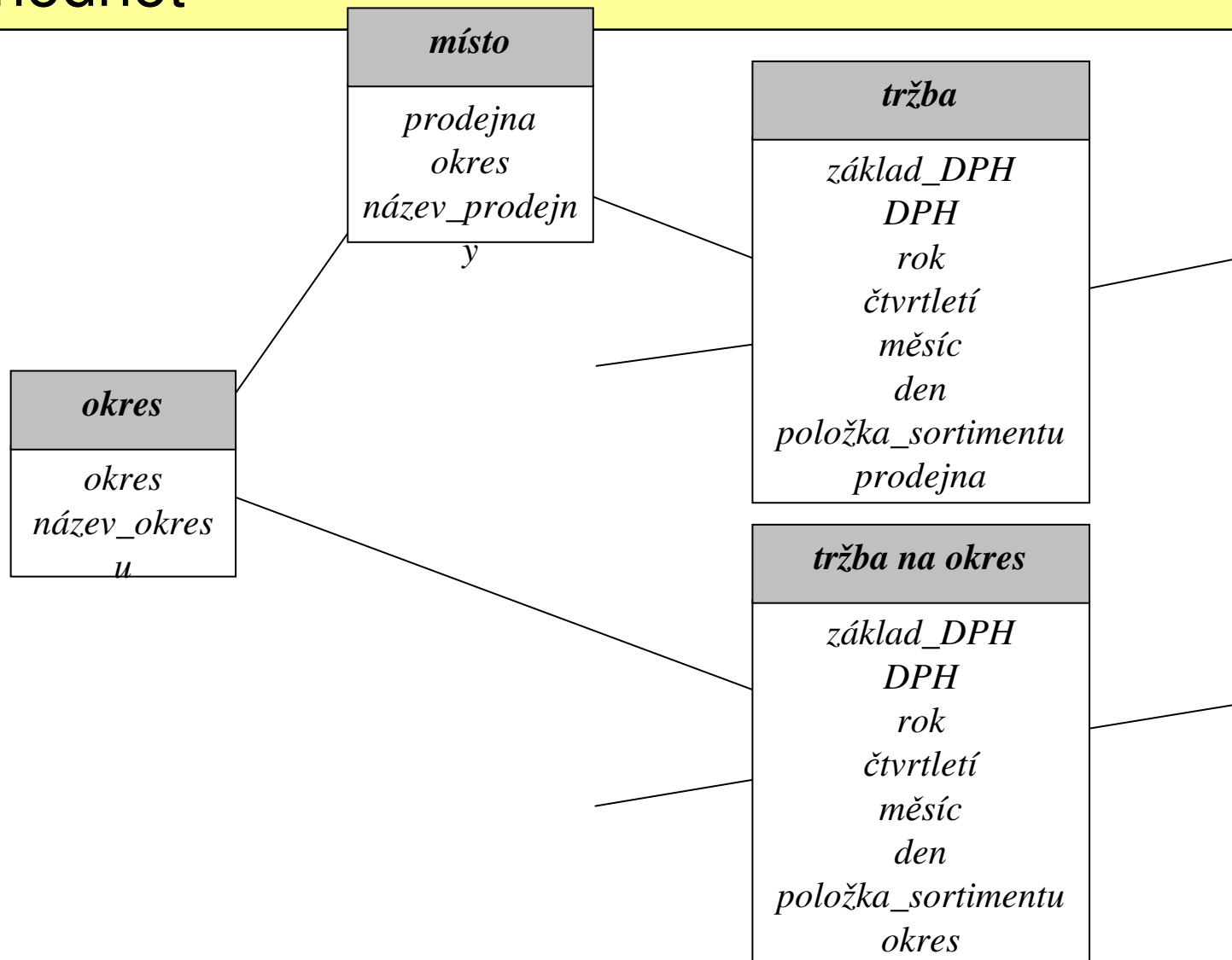
Den v týdnu jako agregační úroveň



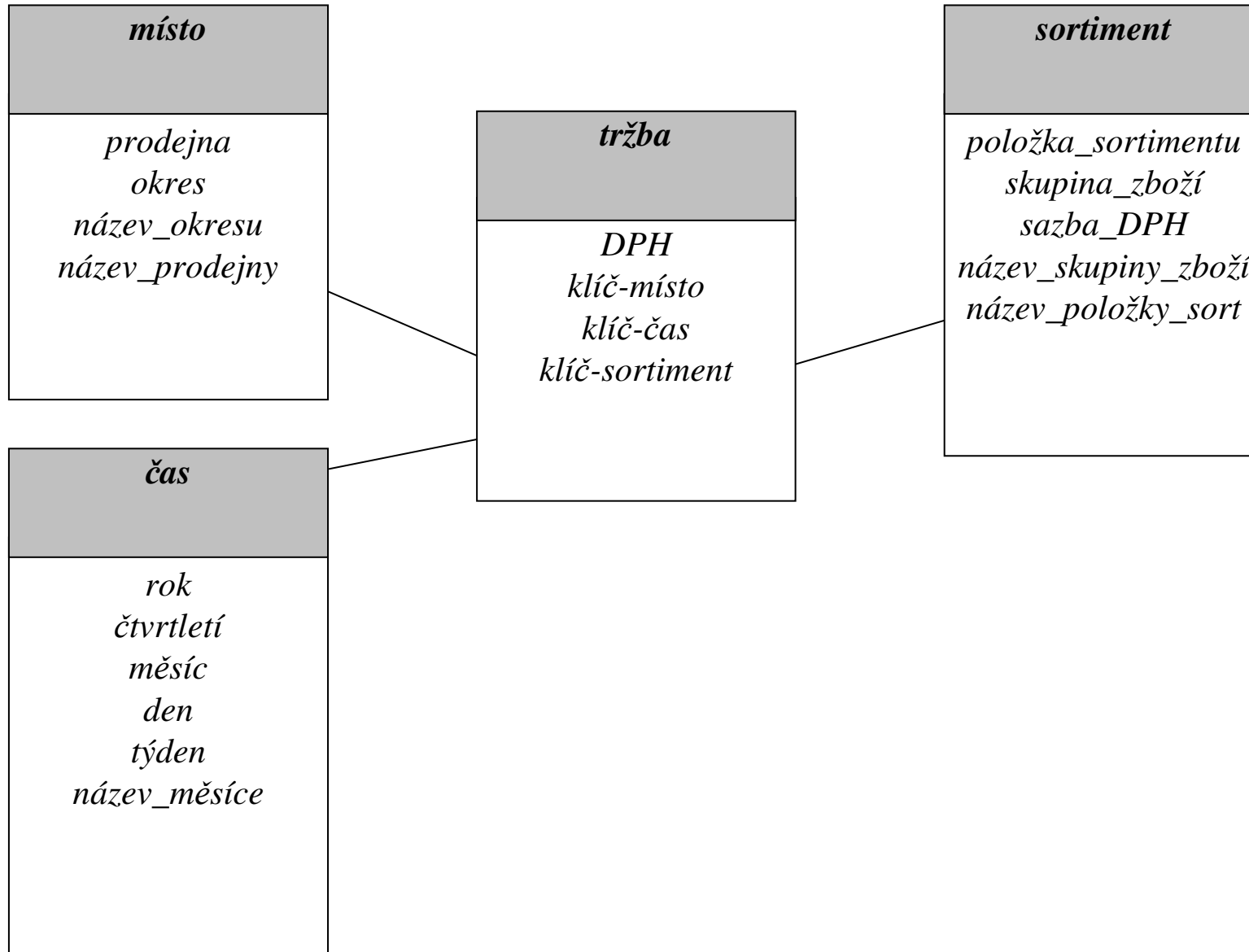
Logický model (*star-schema*) motivačního příkladu



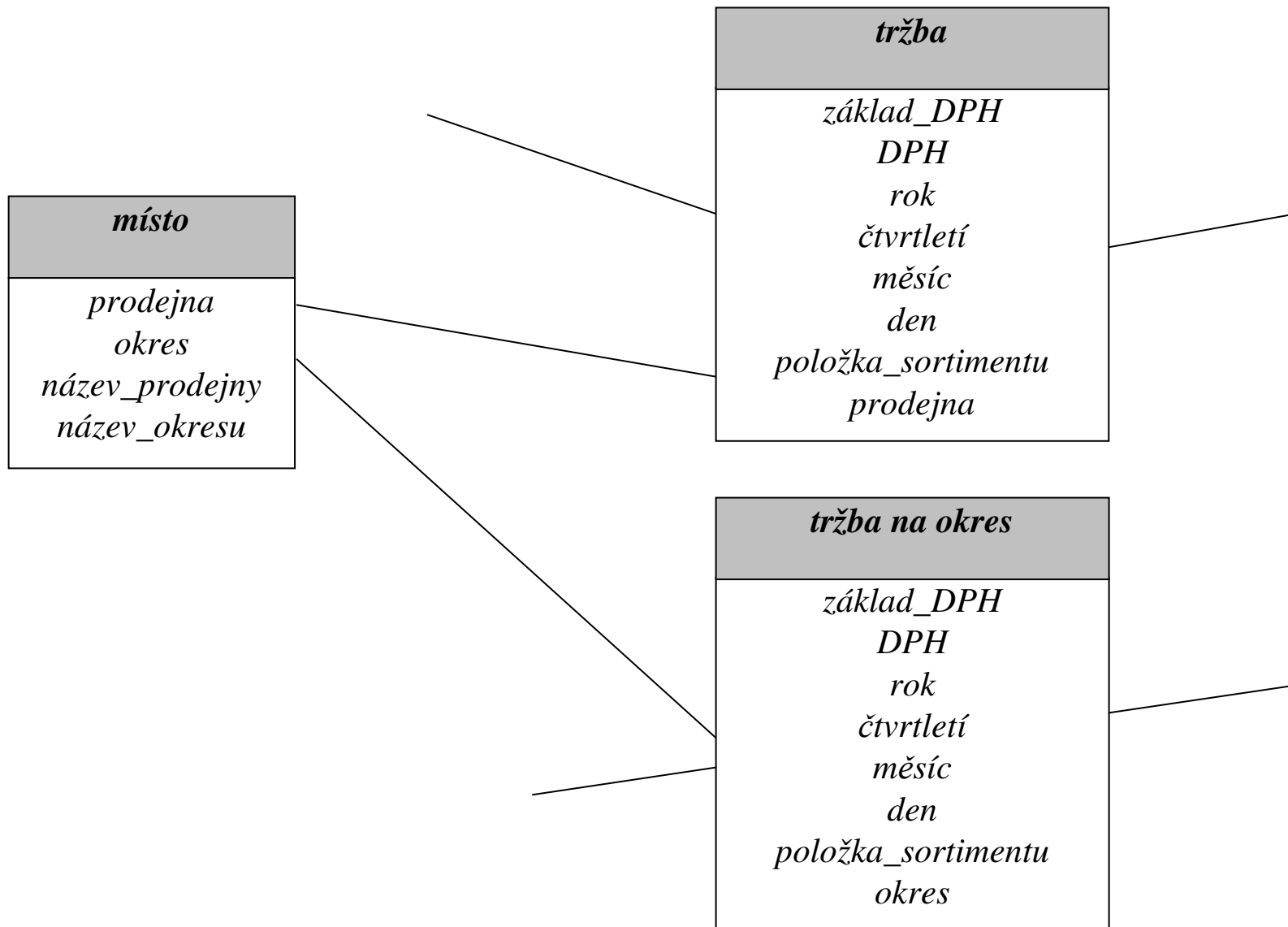
Snowflake schema s tabulkou agregovaných hodnot



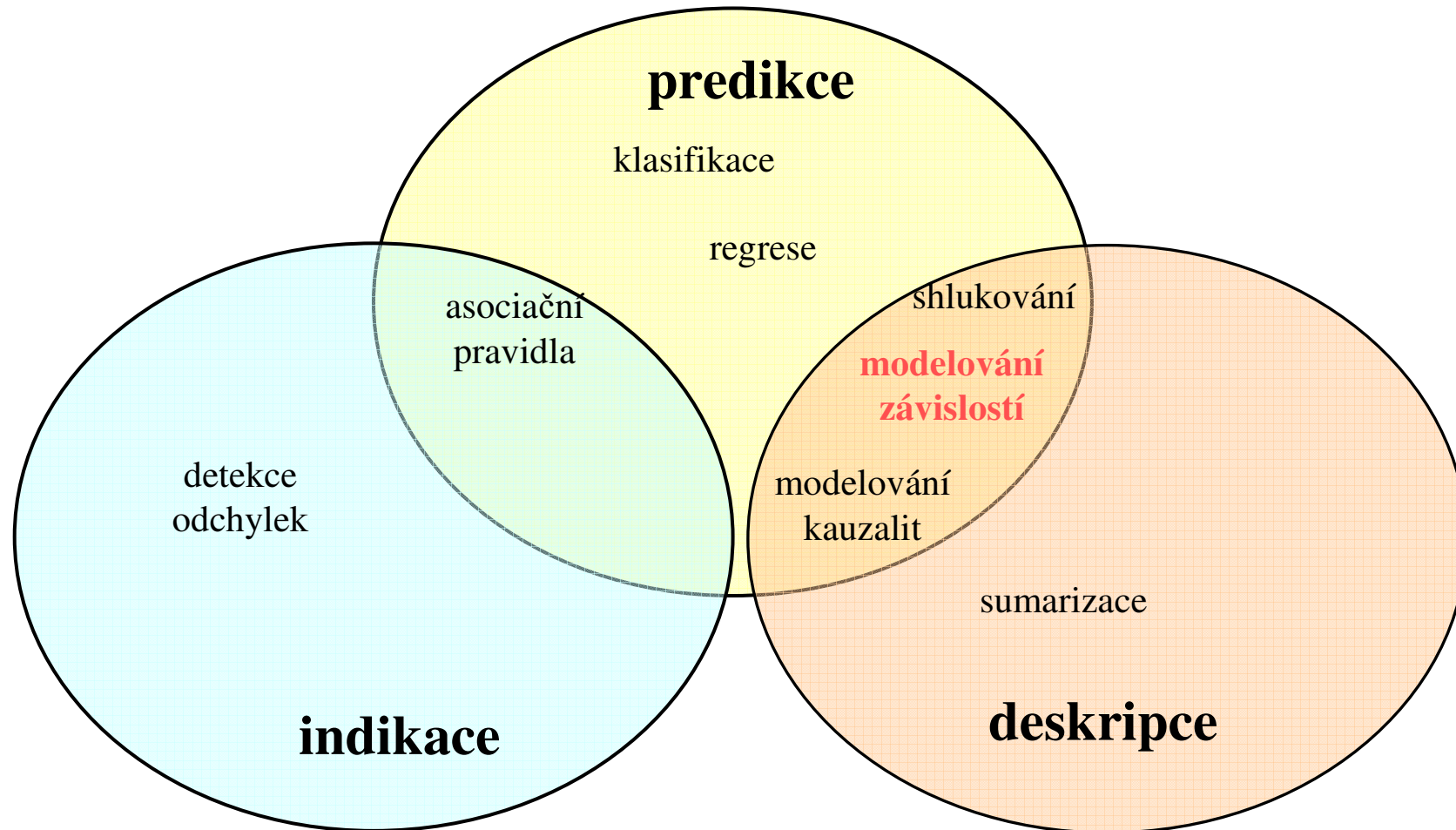
Konsolidované star-schema motivačního příkladu



Schema konsolidovaných faktů motivačního příkladu



Data mining (vytěžování/dolování dat)



Využití nalezených závislosti k predikci

Příklad aplikace datového skladu v predikční úloze:

Predikce spotřeby pitné vody

Faktory, které mohou spotřebu ovlivňovat, nalezeny metodami data mining:

- počasí
- den v týdnu
- roční období
- poloha (zahradkářská kolonie/sídliště)

Využití nalezených závislosti k predikci

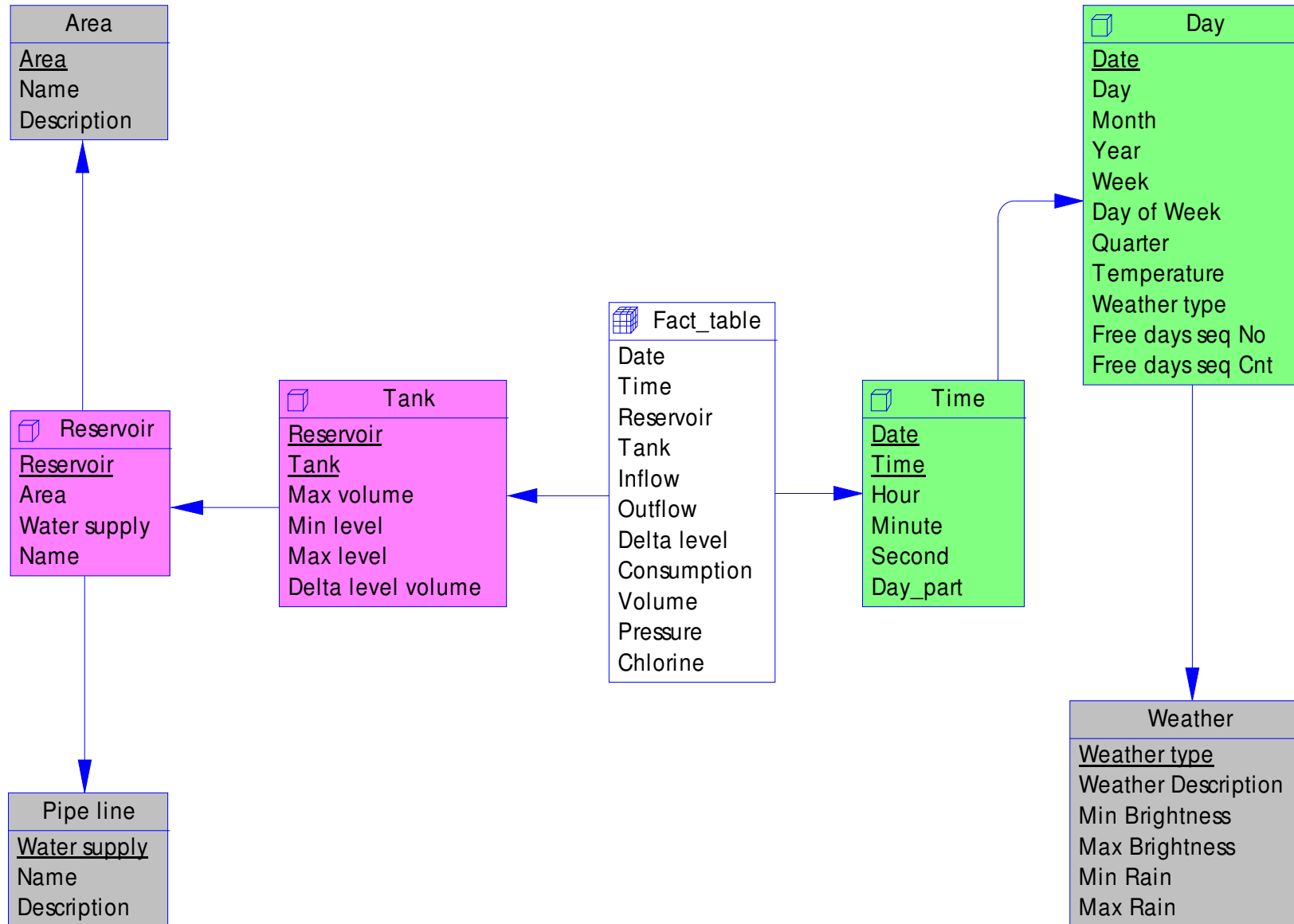
Příklad aplikace datového skladu v predikční úloze:

Predikce spotřeby pitné vody

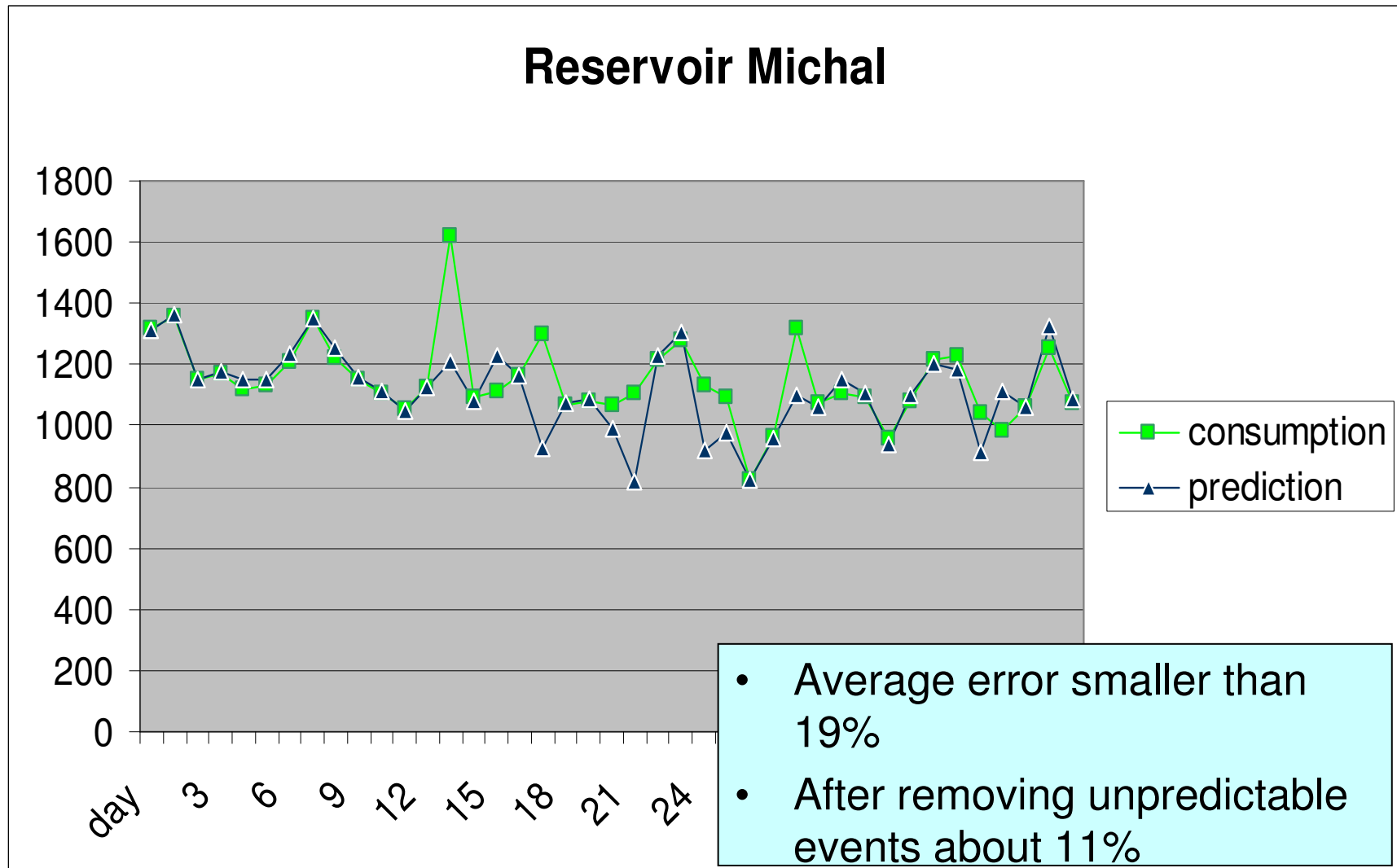
Faktory, které mohou spotřebu ovlivňovat, nalezeny metodami data mining:

- počasí
- den v týdnu
- roční období
- poloha (zahrádkářská kolonie/sídliště)

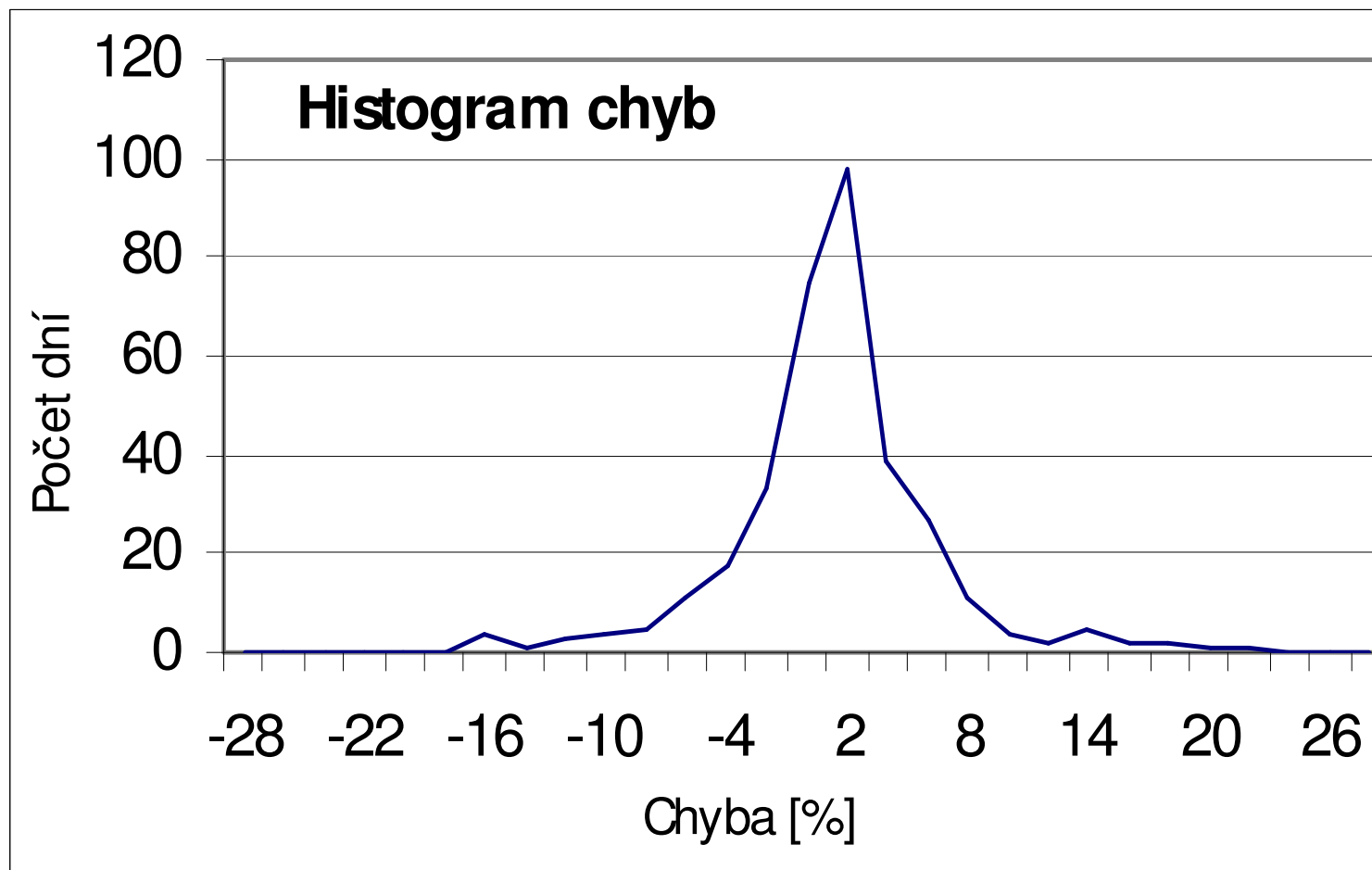
Star schema datového skladu



Výsledky



Histogram chyb



Senzitivita

Senzitivita je definována jako relativní četnost případů patřících do třídy T_1 , které dané klasifikační pravidlo správně zařadilo do třídy T_1 (správně pozitivně klasifikované případy), tedy

$$\text{Sens}(x, \theta) = P(d_{T_1}(x) > \theta | T_1).$$

Specificita

Specifičnost definujeme jako relativní četnost případů patřících do třídy T2, které však byly nesprávně zařazeny do třídy T1 (nesprávně pozitivně klasifikované případy), tedy

$$Spec(x, \theta) = P(d_{T_1}(x) > \theta | T_2)$$

ROC křivka

