



MANAGEMENT V INFORMAČNÍ SPOLEČNOSTI



# **Self Service Business Intelligence**

**Jak si vytvořit vlastní analytické,  
plánovací a reportingové aplikace**

Jan Pour, Miloš Maryška,  
Iva Stanovská, Zuzana Šedivá



Česká společnost  
pro systémovou  
integraci



**Máte zájem** dozvědět se více o našich produktech a službách?

Případně byste se rádi stali **členy našeho teamu?**

**Kontaktujte nás**, rádi se Vám budeme věnovat.



### Business Intelligence

Jsme odborníci na tvorbu efektivní a hodnotné business intelligence.



### Data warehousing

Návrh data warehouse, jeho tvorba a provoz jsou pro nás rutina.



### Enterprise aplikace

Dokážeme spojit nekompatibilní světy do jediného kompatibilního řešení.



### Plánování

Máme kompletní řešení pro Vaše plánování a jsme připraveni Vám poskytnout naše know-how.

**TARGETTY**<sup>®</sup>  
planning & reporting



## PLÁNOVÁNÍ JAKO **KREATIVNÍ** ČINNOST

**Zjednodušte si plánování**, mějte detailní informaci o plnění, využijte potenciál týmu.

*Našemu kamarádovi Honzovi*

**Upozornění pro čtenáře a uživatele této knihy**

Všechna práva vyhrazena. Žádná část této tištěné či elektronické knihy nesmí být reprodukována a šířena v papírové, elektronické či jiné podobě bez předchozího písemného souhlasu nakladatele. Neoprávněné užití této knihy bude **trestně stíháno**.

Doc. Ing. Jan Pour, CSc.  
Doc. Ing. Miloš Maryška, PhD.  
Ing. Iva Stanovská  
Ing. Zuzana Šedivá, Ph.D.

## **Self Service Business Intelligence**

*Jak si vytvořit vlastní analytické, plánovací a reportingové aplikace*

Vydala Grada Publishing, a.s.  
U Průhonu 22, 170 00 Praha 7  
tel.: +420 234 264 401, fax: +420 234 264 400  
[www.grada.cz](http://www.grada.cz)  
jako svou 6828. publikaci

### **Odborní recenzenti:**

Ing. Michal Valenta, Ph.D.  
Ing. Vladislav Štefaňák

Odpovědný redaktor Petr Somogyi  
Grafická úprava a sazba Milan Vokál  
Návrh a zpracování obálky Michal Němec  
Počet stran 352  
První vydání, Praha 2018  
Vytiskla Tiskárna v Ráji, s.r.o., Pardubice

© Grada Publishing, a.s., 2018

ISBN 978-80-271-0823-7 (ePub)  
ISBN 978-80-271-0816-9 (pdf)  
ISBN 978-80-271-0616-5 (print)

# Obsah

---

|                                                                                              |           |
|----------------------------------------------------------------------------------------------|-----------|
| O autorech .....                                                                             | 12        |
| Řekli o knize .....                                                                          | 14        |
| Úvodem .....                                                                                 | 16        |
| <b>Oddíl A: Business Intelligence (BI) a Self Service Business Intelligence (SSBI) .....</b> | <b>19</b> |
| <b>1. Podstata a principy analytických a plánovacích úloh .....</b>                          | <b>20</b> |
| 1.1 Datové modelování .....                                                                  | 21        |
| 1.2 Dimenzionální modelování .....                                                           | 25        |
| 1.2.1 Dimenzionální modelování – hrubý dimenzionální model .....                             | 26        |
| 1.2.2 Návrh datového modelu pro SSBI .....                                                   | 30        |
| 1.3 Tabulky faktů .....                                                                      | 30        |
| 1.3.1 Granularita v tabulce faktů .....                                                      | 31        |
| 1.3.2 Typy tabulek faktů .....                                                               | 31        |
| 1.3.3 Agregace dat .....                                                                     | 32        |
| 1.3.4 Měrné jednotky .....                                                                   | 33        |
| 1.3.5 Rozsah tabulky faktů .....                                                             | 33        |
| 1.3.6 Zdroje a kalkulace ukazatelů .....                                                     | 33        |
| 1.3.7 Tabulky faktů bez ukazatelů .....                                                      | 33        |
| 1.4 Tabulky dimenzí .....                                                                    | 34        |
| 1.4.1 Principy návrhu dimenzionálních tabulek .....                                          | 34        |
| 1.4.2 Schéma STAR .....                                                                      | 36        |
| 1.4.3 Schéma SNOWFLAKE .....                                                                 | 38        |
| 1.4.4 Referenční dimenze .....                                                               | 40        |
| 1.4.5 Degenerované dimenze .....                                                             | 41        |
| 1.4.6 Dimenze parent-child .....                                                             | 41        |
| 1.4.7 Kauzální dimenze .....                                                                 | 42        |
| 1.4.8 Klíče a umělé klíče .....                                                              | 42        |
| 1.4.9 Chyby a NULL hodnoty v klíčích .....                                                   | 42        |
| 1.4.10 Alternativní struktury dimenzí .....                                                  | 43        |
| 1.4.11 Dimenze času .....                                                                    | 43        |
| 1.4.12 Sběrná dimenze (Junk Dimension) .....                                                 | 44        |
| 1.4.13 Změny v dimenzích: SCD (Slowly Changing Dimensions) .....                             | 44        |
| 1.4.14 Analýzy faktů s respektováním změn v dimenzích .....                                  | 48        |
| 1.5 Analytické úlohy v řízení firmy .....                                                    | 52        |
| 1.5.1 Podstata a cíle analytických úloh .....                                                | 52        |
| 1.5.2 Funkcionalita analytických úloh .....                                                  | 53        |
| 1.5.3 Analytické úlohy v kontextu řízení firmy .....                                         | 53        |

|           |                                                          |           |
|-----------|----------------------------------------------------------|-----------|
| 1.6       | Plánovací úlohy v řízení firmy                           | 54        |
| 1.6.1     | Podstata plánovacích úloh                                | 54        |
| 1.6.2     | Principy plánovacích úloh v BI                           | 54        |
| 1.6.3     | Principy změn v analytické databázi                      | 55        |
| 1.6.4     | Plánovací úlohy v kontextu řízení firmy                  | 56        |
| <b>2.</b> | <b>Prezentace a vizualizace dat</b>                      | <b>57</b> |
| 2.1       | Tabulky                                                  | 59        |
| 2.2       | Typy grafů a jejich účel                                 | 61        |
| 2.2.1     | Sloupcový graf (Column Graph)                            | 61        |
| 2.2.2     | Výsečový graf (Pie Chart)                                | 67        |
| 2.2.3     | Spojnicový graf (Line Chart)                             | 70        |
| 2.2.4     | Paretův diagram                                          | 71        |
| 2.2.5     | Plošný graf (Area Chart)                                 | 72        |
| 2.2.6     | Paprskový graf (Radar Chart)                             | 72        |
| 2.2.7     | Bodový graf (Scatter Plot)                               | 73        |
| 2.2.8     | Teplotní mapa (Heat Map)                                 | 74        |
| 2.2.9     | Bublinový graf (Bullet Graph)                            | 74        |
| 2.2.10    | Stromová mapa (Tree Map)                                 | 76        |
| 2.2.11    | Krabicový graf (Box Plot)                                | 76        |
| 2.3       | Faktory ovlivňující srozumitelnost grafu                 | 79        |
| 2.4       | Dashboards, scorecardy                                   | 84        |
| 2.5       | Zásady správného grafického návrhu dashboardu            | 88        |
| 2.5.1     | Data v dashboardu zobrazit na jedné obrazovce            | 88        |
| 2.5.2     | Jednoduchý design                                        | 88        |
| 2.5.3     | Správný výběr grafu                                      | 91        |
| 2.5.4     | Ukázky chybných a správných grafických návrhů dashboardů | 94        |
| <b>3.</b> | <b>Business intelligence</b>                             | <b>98</b> |
| 3.1       | Podstata Business Intelligence                           | 98        |
| 3.2       | Komponenty řešení Business Intelligence                  | 100       |
| 3.2.1     | Dočasné úložiště dat                                     | 100       |
| 3.2.2     | Operační datový sklad                                    | 101       |
| 3.2.3     | Datový sklad                                             | 101       |
| 3.2.4     | Datové tržiště                                           | 102       |
| 3.2.5     | Extract Transform Load (ETL/ELT)                         | 104       |
| 3.2.6     | OLAP databáze                                            | 104       |
| 3.2.7     | Analytická aplikace                                      | 106       |
| 3.2.8     | Reporting                                                | 107       |
| 3.2.9     | Dolování dat                                             | 107       |
| 3.2.10    | Information Management                                   | 107       |
| 3.3       | Přístupy k řešení projektů Business Intelligence         | 110       |
| 3.3.1     | Postupné vytváření datových tržišť                       | 110       |
| 3.3.2     | Jednorázové vytvoření celého BI řešení                   | 111       |
| 3.3.3     | Přirůstkový přístup                                      | 112       |
| 3.4       | Postupy řešení projektů Business Intelligence            | 112       |
| 3.5       | Zpracování Úvodní studie projektu BI                     | 114       |

|           |                                                                             |            |
|-----------|-----------------------------------------------------------------------------|------------|
| 3.6       | Specifikace přírůstku řešení                                                | 115        |
| 3.7       | Analýza stavu a požadavků na BI                                             | 116        |
| 3.8       | Modelování a návrh řešení BI                                                | 116        |
| 3.9       | Návrh technologické platformy přírůstku BI řešení                           | 118        |
| 3.10      | Návrh transformací dat: ETL/ELT                                             | 119        |
| 3.11      | Implementace BI řešení                                                      | 119        |
| 3.12      | Zavedení BI do provozu                                                      | 120        |
| <b>4.</b> | <b>Principy a postup řešení Self Service BI</b>                             | <b>121</b> |
| 4.1       | Principy řešení a užití SSBI                                                | 121        |
| 4.1.1     | SSBI a posuny v organizaci v IT                                             | 121        |
| 4.1.2     | Hlavní doporučení při zavádění SSBI                                         | 122        |
| 4.1.3     | Postup řešení aplikací SSBI                                                 | 123        |
| 4.2       | Zpracování Úvodní studie pro SSBI                                           | 123        |
| 4.2.1     | Definování cílů a efektů Self Service BI                                    | 123        |
| 4.2.2     | Vytvoření katalogu uživatelů                                                | 124        |
| 4.2.3     | Specifikace požadavků na SSBI                                               | 124        |
| 4.2.4     | Funkční specifikace SSBI řešení                                             | 125        |
| 4.2.5     | Specifikace zdrojů dat                                                      | 125        |
| 4.2.6     | Výběr SSBI nástroje a architektury řešení                                   | 126        |
| 4.2.7     | Určení harmonogramu a ekonomických charakteristik projektu                  | 126        |
| 4.3       | Analýza a návrh SSBI aplikací                                               | 127        |
| 4.3.1     | Vyhodnocení aktuálních uživatelských požadavků a dostupných datových zdrojů | 128        |
| 4.3.2     | Dimenzionální modelování                                                    | 128        |
| 4.3.3     | Návrh výstupů: tabulky, dashboardy, reporty                                 | 128        |
| 4.4       | Implementace SSBI aplikací                                                  | 129        |
| 4.5       | Zavedení do provozu SSBI aplikací                                           | 130        |

## **Oddíl B: Nástroje pro tvorbu řešení Self Service Business Intelligence** .....133

|           |                                                               |            |
|-----------|---------------------------------------------------------------|------------|
| <b>5.</b> | <b>Power Pivot</b>                                            | <b>134</b> |
| 5.1       | Hlavní principy Power Pivot                                   | 134        |
| 5.2       | Zpracování úloh v Power Pivot                                 | 135        |
| 5.3       | Aktivace Power Pivot                                          | 136        |
| 5.4       | Postup vytvoření databáze a aplikace v Power Pivot            | 136        |
| 5.4.1     | Otevření Správy databáze                                      | 137        |
| 5.4.2     | Import dat z datových zdrojů                                  | 138        |
| 5.4.3     | Definování relací mezi tabulkami                              | 140        |
| 5.4.4     | Vytvoření kontingenční tabulky/grafu nad databází Power Pivot | 142        |
| 5.4.5     | Vytvoření průřezů v kontingenční tabulce                      | 144        |
| 5.5       | Zdroje dat pro import do databází Power Pivot                 | 145        |
| 5.5.1     | Excel                                                         | 146        |
| 5.5.2     | Textové soubory                                               | 146        |
| 5.5.3     | Databáze                                                      | 148        |
| 5.5.4     | SQL Server Analysis Services                                  | 149        |
| 5.5.5     | Clipboard (Copy/Paste)                                        | 150        |

|           |                                                              |            |
|-----------|--------------------------------------------------------------|------------|
| 5.5.6     | Aktualizace dat                                              | 150        |
| 5.6       | Další funkce Power Pivot                                     | 151        |
| 5.6.1     | Výpočty ukazatelů                                            | 151        |
| 5.6.2     | Vytváření klíčových indikátorů výkonnosti                    | 151        |
| 5.6.3     | Zjišťování zdrojových dat                                    | 151        |
| 5.7       | Příprava dat pro Power Pivot                                 | 152        |
| <b>6.</b> | <b>Power BI</b>                                              | <b>155</b> |
| 6.1       | Principy Power BI                                            | 155        |
| 6.1.1     | Stručné představení postupu práce v Power BI                 | 157        |
| 6.1.2     | Vizualizace                                                  | 157        |
| 6.1.3     | Datové sady                                                  | 158        |
| 6.1.4     | Sestavy (reports)                                            | 158        |
| 6.1.5     | Řídicí panely (dashboardy)                                   | 159        |
| 6.1.6     | Dlaždice (tiles)                                             | 159        |
| 6.2       | Instalace Power BI                                           | 159        |
| 6.2.1     | Instalace Power BI Desktop                                   | 160        |
| 6.2.2     | Instalace Power BI mobile apps                               | 162        |
| 6.3       | Postup vytvoření datové sady a aplikace v Power BI desktop   | 162        |
| 6.3.1     | Zdroje dat pro vytvoření datové sady Power BI desktop        | 162        |
| 6.3.2     | Import dat z datových zdrojů a jejich transformace           | 164        |
| 6.3.3     | Definování relací mezi tabulkami                             | 171        |
| 6.4       | Vytvoření uživatelského rozhraní – sestavy a vizualizace dat | 177        |
| 6.4.1     | Sestavy v Power BI Desktop                                   | 177        |
| 6.4.2     | Tvorba vizuálů v sestavách                                   | 178        |
| 6.4.3     | Filtrování, průřezy, zvýrazňování a drilování                | 181        |
| 6.4.4     | Analytické funkce v grafech Power BI                         | 190        |
| 6.5       | Power BI Service                                             | 191        |
| 6.5.1     | Publikování připravených reportů do služby Power BI          | 191        |
| 6.5.2     | Vybrané funkce Power BI Service                              | 193        |
| 6.5.3     | Vytvoření a sdílení dashboardů                               | 196        |
| 6.5.4     | Cortana intelligence v Power BI Service                      | 200        |
| 6.5.5     | Použití datové brány pro obnovu dat ve službě Power BI       | 200        |
| 6.6       | Mobilní aplikace Power BI Mobile                             | 202        |
| 6.7       | Režim přímého dotazování DirectQuery                         | 204        |
| <b>7.</b> | <b>Jazyk DAX</b>                                             | <b>207</b> |
| 7.1       | Principy DAX                                                 | 207        |
| 7.1.1     | Výpočetní předpis v DAX                                      | 207        |
| 7.1.2     | Datové typy a operátory                                      | 208        |
| 7.2       | Kontext vyhodnocování výpočetních předpisů                   | 208        |
| 7.3       | Kalkulované sloupce a řádkový kontext                        | 210        |
| 7.3.1     | Vytvoření kalkulovaného sloupce                              | 210        |
| 7.4       | Míry a filtr kontext                                         | 212        |
| 7.5       | Funkce IF ()                                                 | 215        |
| 7.6       | Práce s více tabulkami                                       | 215        |
| 7.6.1     | Řádkový kontext s vazbami tabulek                            | 215        |

|           |                                                     |            |
|-----------|-----------------------------------------------------|------------|
| 7.6.2     | Filtr kontext s vazbami tabulek                     | 216        |
| 7.7       | Funkce CALCULATE ()                                 | 217        |
| 7.7.1     | Podstata funkce CALCULATE()                         | 217        |
| 7.7.2     | Parametr ALLSELECTED                                | 218        |
| 7.8       | Využití iterací a iterátorů                         | 219        |
| 7.8.1     | CALCULATE () a využití FILTER ()                    | 219        |
| 7.9       | Využití hierarchií prvků dimenzí                    | 220        |
| 7.9.1     | Výpočty v hierarchiích prvků                        | 220        |
| 7.10      | Key Performance Indicators (KPI)                    | 221        |
| 7.11      | Využití dimenze času (Time Intelligence)            | 223        |
| 7.11.1    | Základní principy Time Intelligence                 | 223        |
| 7.11.2    | Ukazatele pro pracovní dny                          | 224        |
| 7.11.3    | Funkce YTD, QTD, MTD                                | 224        |
| 7.11.4    | Sledování hodnot za minulé roky                     | 226        |
| 7.11.5    | Další funkce                                        | 226        |
| 7.12      | Další analytické funkce založené na DAX             | 227        |
| 7.12.1    | Seskupování dat (banding)                           | 227        |
| 7.12.2    | Vytvoření pořadí (ranking)                          | 227        |
| 7.12.3    | Růst počtu zákazníků                                | 228        |
| 7.13      | Příklady použití jazyka DAX v Power BI Desktop      | 228        |
| 7.13.1    | Definice kalkulovaného sloupce v Power BI           | 230        |
| 7.13.2    | Definice měř v Power BI Desktop                     | 233        |
| 7.13.3    | Podpora pro výpočet měř v Power BI (Quick Measures) | 235        |
| <b>8.</b> | <b>Qlik Sense</b>                                   | <b>239</b> |
| 8.1       | Principy Qlik Sense                                 | 239        |
| 8.2       | Instalace Qlik Sense                                | 241        |
| 8.3       | Zdroje dat pro import do Qlik Sense                 | 241        |
| 8.4       | Postup vytvoření databáze a aplikace v Qlik Sense   | 242        |
| 8.4.1     | Import dat z datových zdrojů                        | 242        |
| 8.5       | Definování relací mezi tabulkami                    | 247        |
| 8.6       | Vytvoření uživatelského rozhraní (vizualizace)      | 248        |
| 8.7       | Vybrané funkce Qlik Sense                           | 251        |
| 8.7.1     | Grafy                                               | 251        |
| 8.7.2     | Master Items                                        | 253        |
| 8.7.3     | Pole                                                | 254        |
| 8.7.4     | Proměnné (variables)                                | 254        |
| 8.7.5     | Hledání v datech                                    | 255        |
| 8.8       | Komplexní příklad tvorby uživatelského rozhraní     | 256        |
| <b>9.</b> | <b>QlikView</b>                                     | <b>270</b> |
| 9.1       | Principy QlikView                                   | 270        |
| 9.2       | Instalace QlikView                                  | 271        |
| 9.3       | Zdroje dat pro import do QlikView                   | 272        |
| 9.4       | Postup vytvoření databáze a aplikace v QlikView     | 272        |
| 9.5       | Import dat z datových zdrojů                        | 273        |
| 9.6       | Úprava navržených objektů a relací                  | 276        |

|                                  |                                                      |            |
|----------------------------------|------------------------------------------------------|------------|
| 9.7                              | Vytvoření uživatelského rozhraní (vizualizace)       | 277        |
| 9.8                              | Vybrané funkce QlikView                              | 281        |
| <b>10.</b>                       | <b>Tableau</b>                                       | <b>282</b> |
| 10.1                             | Principy Tableau                                     | 282        |
| 10.2                             | Instalace Tableau                                    | 284        |
| 10.3                             | Vytvoření databáze a aplikace v Tableau              | 285        |
| 10.4                             | Import dat z datových zdrojů                         | 285        |
| 10.5                             | Definování relací mezi tabulkami                     | 287        |
| 10.6                             | Vytvoření uživatelského rozhraní (vizualizace)       | 290        |
| 10.7                             | Další funkce Tableau                                 | 292        |
| 10.7.1                           | Nastavení filtru                                     | 292        |
| 10.7.2                           | Využití geografických dat                            | 293        |
| 10.7.3                           | Hierarchie dimenzí                                   | 295        |
| 10.7.4                           | Export vizualizace do datového souboru               | 296        |
| 10.7.5                           | Vytvoření kontingenční tabulky a její úpravy         | 298        |
| 10.7.6                           | Počítané pole (Calculated Field)                     | 301        |
| 10.7.7                           | Proměnné (Parameters)                                | 304        |
| 10.7.8                           | Určení formátu číselné hodnoty                       | 305        |
| 10.8                             | Komplexní příklad uživatelského rozhraní             | 306        |
| 10.8.1                           | Základní přehled ukazatelů                           | 306        |
| 10.8.2                           | Určení úrovně agregace pro vybrané ukazatele         | 308        |
| 10.8.3                           | Sledování podílových ukazatelů                       | 313        |
| 10.8.4                           | Porovnání plánovaných a skutečně dosahovaných hodnot | 313        |
| 10.8.5                           | Vytvoření dashboardu                                 | 318        |
| <b>11.</b>                       | <b>Targetty</b>                                      | <b>321</b> |
| 11.1                             | Principy Targetty, základní funkcionalita            | 322        |
| 11.1.1                           | Sestavování podnikových plánů                        | 323        |
| 11.1.2                           | Finanční a další funkce spojené s tvorbou plánů      | 324        |
| 11.1.3                           | Tvorba dashboardů                                    | 325        |
| 11.1.4                           | Podpora workflow v plánování                         | 325        |
| 11.2                             | Postup využití nástroje Targetty                     | 326        |
| 11.2.1                           | Postup práce v Targetty při přípravě plánu           | 327        |
| 11.2.2                           | Výpočty plánovacích hodnot                           | 328        |
| 11.2.3                           | Poznámky k plánu                                     | 329        |
| 11.2.4                           | Teamwork                                             | 330        |
| 11.2.5                           | Reporty                                              | 330        |
| 11.2.6                           | Možnosti a nápověda                                  | 330        |
| 11.3                             | Oblasti využití Targetty                             | 331        |
| <b>Závěr</b>                     |                                                      | <b>333</b> |
| <b>Příloha: Souhrnný příklad</b> |                                                      | <b>335</b> |
| Dimenze                          |                                                      | 336        |
| Dimenze Čas                      |                                                      | 336        |
| Dimenze Útvary                   |                                                      | 336        |

|                            |            |
|----------------------------|------------|
| Dimenze Regiony .....      | 337        |
| Dimenze Zákazníci.....     | 337        |
| Dimenze Zboží.....         | 338        |
| Tabulka faktů .....        | 338        |
| <b>Literatura .....</b>    | <b>339</b> |
| <b>Slovník pojmů .....</b> | <b>343</b> |
| <b>Rejstřík .....</b>      | <b>349</b> |

# O autorech

---

## **Doc. Ing. Jan Pour, CSc.**

Jan Pour vystudoval fakultu národohospodářskou na VŠE v Praze. Je docentem katedry informačních technologií Vysoké školy ekonomické v Praze. Profesně se orientuje na otázky řízení podnikové informatiky, koncepcí informačních systémů podniků a úlohy Business Intelligence. Je autorem nebo spoluautorem čtrnácti knižních publikací a řady vysokoškolských skript, článků v časopisech a příspěvků na konferencích. Podílel se na řešení cca sto dvaceti projektů pro nejrůznější podniky a organizace v energetice, průmyslu, obchodu i ve státní správě. V současné době je členem Rady České společnosti pro systémovou integraci a redaktorem časopisu Systémová integrace.



## **Doc. Ing. Miloš Maryška, PhD.**

Miloš Maryška studoval na Katedře informačních technologií Vysoké školy ekonomické, kde obhájil disertační práci na téma „Referenční model řízení ekonomiky podnikové informatiky“. Od roku 2015 působí na Katedře informačních technologií jako docent. V rámci vědecké a pedagogické práce se zaměřuje na oblast řízení ekonomiky podnikové informatiky, Business Intelligence, Corporate Performance Management, finanční řízení podnikové informatiky a ERP systémy. Je spoluautorem sedmi monografií a více než sta odborných a vědeckých článků v časopisech a ve sbornících z mezinárodních konferencí. Ve společnosti VIG Re působí jako projektový manažer a team leader oddělení Business Processes and IT Solutions.



## **Ing. Iva Stanovská**

Iva Stanovská je absolventkou fakulty řízení Vysoké školy ekonomické v Praze. V současné době působí jako ředitelka pro rozvoj a konzultační služby ve společnosti Solitea Business Solutions s.r.o., kde je odpovědná za řízení týmu Business Intelligence, předprodejní aktivity a podporu obchodu, firemní procesy a interní informační systém. Současně na částečný úvazek působí jako odborná asistentka na katedře informačních technologií Fakulty informatiky a statistiky VŠE v Praze, kde se zaměřuje na metodiky a postupy specifikace informačních systémů. Je spoluautorkou několika učebních textů, článků v časopisech a příspěvků na konferencích.



### **Ing. Zuzana Šedivá, Ph.D.**

Zuzana Šedivá vystudovala obor Automatizované systémy řízení na Fakultě informatiky a statistiky na Vysoké škole ekonomické v Praze, kde od absolvování působí na Katedře informačních technologií. V pedagogické práci i ostatních odborných aktivitách se zaměřuje především na oblast aplikací grafických dat, vizualizací a analytické úlohy v BI. Současně svou odbornost rozvíjí v oblasti podpory informačních technologií ve vzdělávání. Kromě krátkodobých praktických projektů, kterých se zúčastnila, spolupracovala dlouhodobě s organizací OBSE v oblasti správy a řízení elektronických dokumentů.



# Řekli o knize

---

Kniha je praktickým a kompletním průvodcem po světě tzv. Self Service Business Intelligence (SSBI) nástrojů. Praktickým proto, že se zaměřuje na reálné případy použití této technologie v kontextu řízení a plánování v menších a středních podnicích. Kompletním průvodcem je ve více směrech:

1. Kromě popisu vybraných nástrojů SSBI a způsobu jejich použití poskytuje také příslušnou teorii, zejména úvod do datového modelování, architekturu a postup návrhu datových skladů a metodiku využití nástrojů SSBI pro řízení a plánování.
2. Jednotlivé nástroje SSBI popisuje od jejich stažení přes instalaci, konfiguraci až pro reálné použití. Kniha nabízí sedm různých rozšířených a v praxi používaných nástrojů SSBI. Většina z nich je dostupná volně nebo v tzv. community edition. Čtenář si tak může udělat přesný obraz o nástrojích a způsobu práce s nimi a vybrat si ten, který mu nejvíce vyhovuje.

Kniha je napsaná velmi čtivě a srozumitelně, přináší i množství ilustrativních příkladů. Teoreticky laděný první oddíl je terminologicky konzistentní a dostatečně přesný pro pochopení problematiky, aniž by zabíhal do složitých detailů, které nejsou pro uživatele SSBI nástrojů potřebné.

Jsem přesvědčen, že knihu ocení široká skupina čtenářů z řad manažerů středních a menších firem, analytiků, zaměstnanců IT oddělení, kteří poskytují v podniku informační servis, stejně jako studenti ekonomických a technických oborů se specializací na informační systémy a management.

Další přidanou hodnotou knihy je její provázanost s portálem MBI (Management Byznys Informatiky). Jedná se o živý informační zdroj, do něhož přispívá řada specialistů na jednotlivé oblasti řízení a provozu IT a který poskytuje stále aktualizované a doplňované informace o operativním, taktickém a strategickém řízení IT v podniku. Právě tam tematika SSBI patří. Čtenář knihy tedy navíc dostává k dispozici další cenný informační zdroj.

*Ing. Michal Valenta, Ph.D.*

*České vysoké učení technické v Praze*

*Fakulta informačních technologií*

*Katedra softwarového inženýrství*

Podle studie společnosti Gartner (2017) je dnes penetrace BI nástrojů ve firmách na úrovni 30 %, tedy v průměrné organizaci pouze 30 % uživatelů aktivně používá některý z nástrojů BI. Pokud se navíc podíváme do historie před pěti lety, zjistíme, že před zavedením metod Self Service Business Intelligence (SSBI) byl tento ukazatel na úrovni 15–17 %. Lze tedy říct, že to jsou právě nástroje SSBI, které významně přispívají k většímu rozšíření nástrojů BI v organizacích.

Publikace přináší komplexní přehled oblasti jak v metodické rovině (oddíl A), tak na příkladu konkrétních nástrojů (oddíl B) a velice dobře funguje jako komplexní úvod do oblasti SSBI jak pro pracovníky a manažery, tak pro studenty oborů IT.

Oddíl A předkládá úvod do problematiky logickým přehledem základů nástrojů Business Intelligence od analytických databází přes budování datových skladů a vizualizační nástroje až po nástroje SSBI v jejich dnešní podobě.

Oddíl B čtenáře provede ukázkami sedmi vybraných nástrojů z kategorie SSBI, včetně reálných ukázek z uživatelského rozhraní, pracovních postupů při budování analytického modelu i vizualizačního rozhraní.

Za velmi přínosné považují především propojení metodiky budování SSBI řešení v organizaci s metodikou MBI, která umožňuje čtenáři jednoduchou formou proniknout do logiky business procesů v organizaci. Lze konstatovat, že u nástrojů typu SSBI je jejich vhodné a pevné ukotvení do metodického rámce a procesů organizace kritickým faktorem jejich finálního úspěchu či neúspěchu.

V publikaci je několikrát akcentováno, že řešení typu SSBI nejsou z pohledu uživatele zcela nenáročné na technologické znalosti, naopak: vyžadují od uživatele určitý mentální posun z role příjemce reportů do role reálného správce dat a jejich business logiky. Právě zakotvení nástrojů SSBI v pevných pravidlech, jednotné metodice užití dat a také jednotné pravdě v rámci organizace představuje základní kámen SSBI řešení. Toto téma vnímám jako jeden z nejvýraznějších přínosů publikace, která může přispět ke skutečně efektivnímu použití SSBI nástrojů v organizaci. Pro čtenáře jsou užiteční a velmi příjemné i reálné postupy a doporučení v oblasti vizualizací v oddílu A, které jsou následně doplněné o postupy přímo v konkrétních nástrojích (oddíl B).

Z pohledu možného dalšího rozvoje publikace, resp. jejího případného pokračování bych doporučil následující oblasti:

- Zahnutí návrhů na organizační kroky a změny procesů, které povedou k „udržitelnému rozvoji“ SSBI nástrojů ve firmě. Tedy kroky, které následují po vlastním nasazení SSBI nástrojů a mají za cíl zajistit jejich efektivní a dlouhodobé používání. Konkrétně může jít například o rozvoj nástrojů samostatně v jednotlivých útvech, avšak pod jednotným metodickým vedením, sdílení definic KPI, zajištění odpovědnosti za kvalitu dat (data stewardship) atp.
- Rozšíření oblasti plánovacích a prediktivních SSBI nástrojů o přehled dalších nástrojů. Oblast plánovacích SSBI nástrojů je specifický trh, kde se setkávají standardní poskytovatelé SSBI s řešeními specializovanými čistě na oblast plánování či prediktivní logiky. Podle studie společnosti Gartner (2017) je tento segment nejrychleji rostoucí oblastí SSBI a do budoucna může zejména prediktivní SSBI výrazně posunout možnosti nasazení BI nástrojů v organizacích.

Publikaci vřele doporučuji všem čtenářům pro její výraznou přidanou hodnotu a pro skloubení metodiky a praxe SSBI do přehledného a logického celku. Současně bych chtěl kolektivu autorů poděkovat, protože podobná publikace na českém knižním trhu zatím chyběla.

*Ing. Vladislav Štefaňák  
Chief Executive Officer  
MIBCON, a.s.*

# Úvodem

---

**Business Intelligence (BI)** představuje jednu z klíčových aplikací IT, která nejvýrazněji ovlivňuje celkovou kvalitu a výkonnost řízení podniků a jejich obchodních a dalších aktivit. Například průzkumy a analýzy společnosti Gartner ukazují, že investice do BI představují celosvětově a dlouhodobě nejvyšší investiční prioritu podniků do IT. Je to dáno zejména rozsáhlými možnostmi BI při podpoře analytických, plánovacích i rozhodovacích aktivit a jejího vlivu na celkovou úspěšnost a konkurenceschopnost firmy.

Výrazným vývojovým proudem v rámci Business Intelligence je uplatňování postupů, nástrojů a aplikací charakterizovaných termínem **Self Service Business Intelligence (SSBI)**, resp. **samoobslužná Business Intelligence**. V tomto textu používáme první z uvedených termínů. Zatímco původní standardní komplexní systémy se vyznačují celopodnikovým záběrem, obrovskými objemy dat, vysokou složitostí řešení a vysokými náklady, řešení Self Service Business Intelligence jsou zaměřená zejména na individuální potřeby uživatelů nebo jejich menších skupin. Tato řešení jsou založena – oproti standardním BI systémům – také na odlišné škále produktů a nástrojů (např. Power BI, PowerPivot, Tableau, Qlik Sense atd.). Základní principy řešení, analýza a návrh SSBI aplikací jsou však v mnoha směrech obdobné jako u komplexní BI. V tomto kontextu se také často zdůrazňuje, že aplikace SSBI je třeba vesměs chápat jako efektivní doplněk komplexních BI systémů, nikoli jako jejich náhradu.

Tato publikace se primárně orientuje na řešení a nástroje kategorie SSBI, a to jak na jejich analytickou část, tak na implementační. S ohledem na silné vazby na komplexní BI systémy postihuje také jejich základní principy, komponenty a postupy řešení. Kniha je určena IT analytikům i uživatelům (podnikovým specialistům, manažerům), kteří mají zájem samostatně připravovat vlastní analytické a plánovací aplikace, případně se na jejich přípravě a vývoji aktivně podílet. Je rovněž určena studentům, a to jak těm z IT oborů, tak i z ostatních ekonomicky nebo manažersky zaměřených oborů.

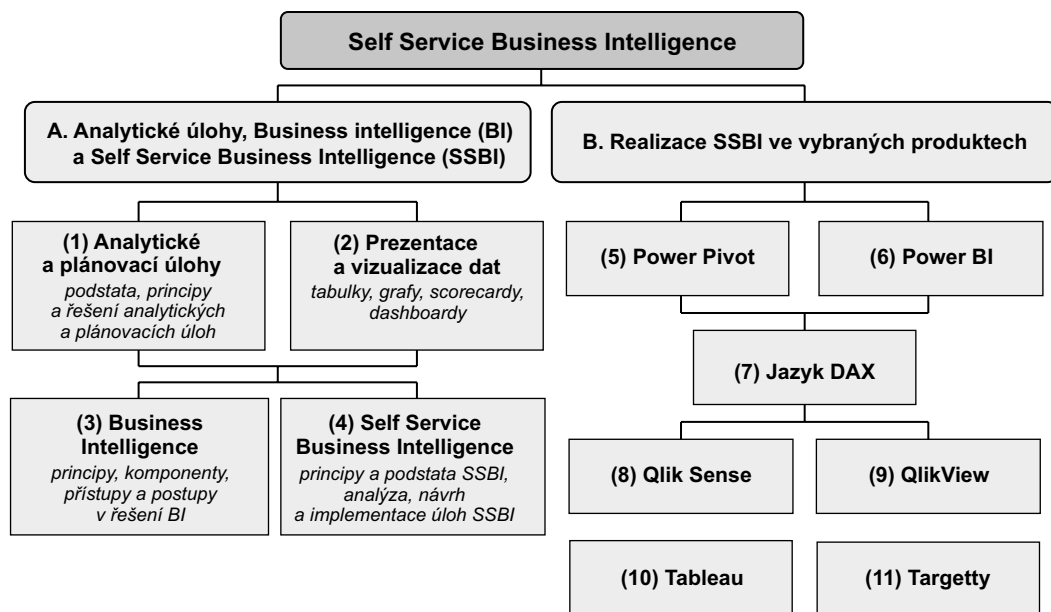
Cílem knihy je poskytnout komplexní pohled na přípravu a realizaci zmíněných typů aplikací, resp. úloh spadajících do oblasti označované jako „business analytics“. Strukturu celého textu dokumentuje obrázek na následující straně.

## Struktura textu knihy

Příprava a realizace aplikací BI i SSBI (stejně jako i naprosté většiny ostatních) vyžaduje většinou kombinaci tří druhů znalostí:

- věcných: znalosti obsahu podnikání, podnikových procesů a funkcí, systémů ukazatelů a jejich analytických hledisek,
- metodických: znalosti postupů řešení a možností metod, které se v dané oblasti používají (dimenzionální modelování, datové modelování a další),
- technologických: znalosti funkcionality a způsobů využití softwarových nástrojů pro implementaci aplikací (např. Power BI, Tableau atd.).

Na tyto požadavky se snaží reagovat i tato kniha. Problém ale spočívá v tom, že rozsah k tomuto účelu důležitých informací je značný, přičemž uživatel je potřebuje v různých okamžicích řešení, v různém rozsahu a na různé úrovni detailu. K tomu se ještě váží změny přicházející především v technologické rovině řešení, tedy v dostupných nástrojích. **Z těchto důvodů jsme se rozhodli provázat text knihy s informacemi na portálu MBI (Management Byznys Informatiky), který najdete na adrese <http://mbi.vse.cz>.**



Poslední informací vážící se k obsahu a organizaci celé publikace je vymezení použitých **formálních standardů**, zejména piktogramů, které uvozují standardní části jednotlivých kapitol:



Účel kapitoly nebo části textu.



Vymezení nebo definice nejdůležitějších pojmů.



Příklady.



Diskuse, doporučení, praktiky.



Shrnutí kapitoly nebo vybrané části textu.



# Oddíl A: Business Intelligence (BI) a Self Service Business Intelligence (SSBI)

---

První oddíl je zaměřen na společné charakteristiky analytických a plánovacích úloh v podnikovém řízení a na jejich řešení. V návaznosti na to poskytuje obecně pojaté informace o aplikacích a nástrojích jak pro úlohy Business Intelligence, tak Self Service Business Intelligence. Oddíl má odpovědět na tyto otázky:

1. Jaké jsou specifické charakteristiky analytických a plánovacích úloh v podniku na rozdíl od ostatních, zejména transakčních?
2. Jaké jsou používané metody při řešení aplikací pro analytické a plánovací úlohy?
3. Co je podstatou datového modelování a jak se využívá při řešení analytických a plánovacích aplikací?
4. Co je podstatou dimenzionálního modelování, jaké jsou hlavní charakteristiky tabulek faktů a dimenzionálních tabulek a jejich vzájemných vazeb?
5. Jaké jsou současné možnosti vizualizace a prezentace dat v analytických a plánovacích aplikacích, tedy možnosti využití různých typů grafů, map, dashboardů apod.?
6. Jaké jsou principy aplikací a technologií komplexních Business Intelligence systémů, jaké jsou možné přístupy a postupy řešení v této oblasti?
7. Jaké jsou podstatné charakteristiky aplikací a nástrojů Self Service Business Intelligence, jaké jsou doporučené postupy řešení v této oblasti?

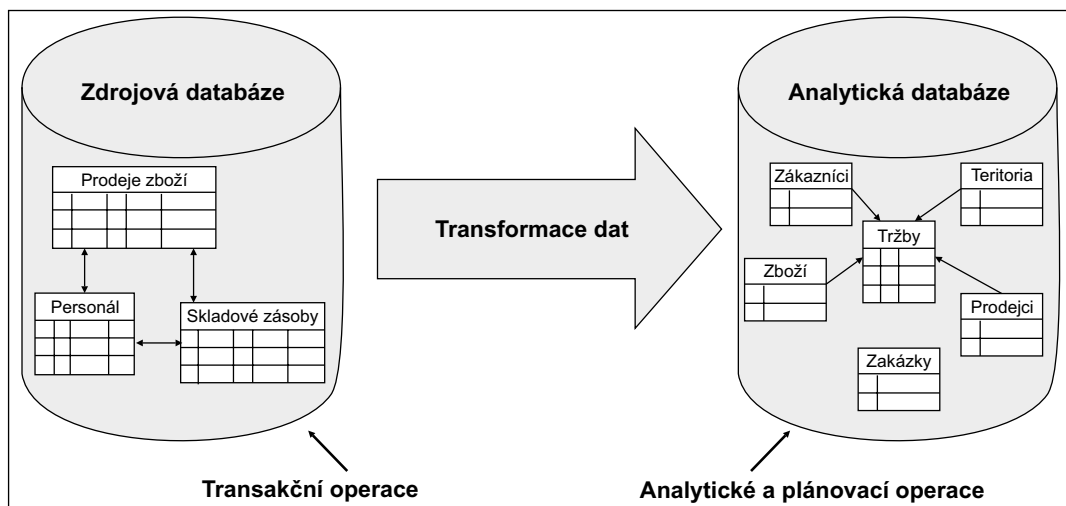
# 1. Podstata a principy analytických a plánovacích úloh



Analytické a plánovací úlohy, označované též jako *Business Analytics*, představují disciplínu a společný termín pro podnikové analýzy, plánování, dolování dat, prediktivní analýzy a další úlohy a možnosti podporující a zvyšující kvalitu podnikového řízení a rozhodování. Standardně se tyto funkce realizují na základě ukazatelů (resp. faktů, metrik) a podle zvolených hledisek, resp. dimenzí. Business Intelligence již po několik desítek let, Self Service Business Intelligence nyní dává těmto principům systematickosti, provázanosti a odpovídající technologickou základnu. Základem jejich řešení je tzv. **dimenzionální modelování** orientované na určení požadovaných ukazatelů a jim odpovídajících dimenzí a jejich podstatných charakteristik. **Předpokladem** pochopení a řešení dimenzionálních modelů je **datové** modelování, a proto mu jako východisku pro další kapitoly věnujeme úvodní část tohoto textu. Cílem kapitoly je vymezit principy datových a dimenzionálních modelů a určit funkce zejména analytických a plánovacích úloh v podnikovém řízení a s tím i požadavky na řešení aplikací BI i SSBI. Jedním z velmi podstatných požadavků na tyto aplikace je vysoká úroveň prezentace dat, využití nejrůznějších typů grafů a schémat, které by měly přispívat k co nejrychlejší orientaci uživatele v analytických tabulkách, reportech apod. Proto je poměrně rozsáhlá část textu věnovaná možnostem vizualizace a prezentace dat jak v aplikacích BI, tak SSBI.

Doplňující informace jsou dostupné na webu MBI. Kromě toho příloha obsahuje komplexní příklad využitý zejména při prezentaci jednotlivých produktů v kapitolách 5–11. V rámci této kapitoly využíváme také několika zjednodušených dílčích příkladů pro objasnění podstaty analytických a plánovacích operací.

**Analytické aplikace** (BI i SSBI) nevytvářejí nebo většinou nepořizují nová data, ale využívají data vytvořená transakčními aplikacemi (ERP, CRM atd.). Databáze **transakčních aplikací** se proto označují jako **zdrojové**. Podstatnou vlastností těchto databází je taková organizace jejich dat, která podporuje přístupy k detailním datům, ukládání a aktualizace dat. Databáze jsou pro tyto operace optimalizované. Oproti tomu analytické BI a SSBI aplikace jsou optimalizované na efektivní poskytování analytických informací, tedy data zde musí být organizována ve shodě s potřebami analytických úloh a musí obsahovat **hodnoty ukazatelů ve vazbě na analytická hlediska, tedy dimenze**. Z toho také vyplývá, že mezi zdrojovými databázemi a analytickými databázemi probíhá transformace dat (viz obrázek 1.1). Analytické databáze mají v prostředí Business Intelligence formu datových skaldů, datových tržišť, případně OLAP databází (viz kapitolu 3), v prostředí Self Service BI disponují jednotlivé produkty vlastními technologickými nástroji pro uložení a správu analytických dat (viz kapitoly 5–11).



**Obrázek 1.1** Transformace dat ze zdrojových do analytických databází

Kvalitní a efektivní návrh analytických databází je předmětem **datového a dimenzionálního modelování**. Jejich hlavním úkolem je vytvořit základní logiku uložení a uspořádání dat tak, aby vyhovovala požadavkům na analytické a plánovací aplikace, a vytvořit flexibilní datový model, který bude plně podporovat rozsah podnikových analýz a plánování, a to jak aktuálně požadovaných, tak očekávaných v budoucím období.

Je dobré zdůraznit, že při řešení Business Intelligence i Self Service Business Intelligence je předpokladem znalost datového i dimenzionálního modelování, a to nejen u IT analytiků a vývojářů, kde se běžně předpokládá, ale i u uživatelů. Mají-li se uživatelé na těchto řešeních podílet (a v případě SSBI je pokud možno i samostatně implementovat), pak jsou tyto základní znalosti nezbytné.

## 1.1 Datové modelování



Datové modelování je metoda, která podle názvu slouží k navrhování struktury dat v databázích a k jejich dokumentaci. **Účelem datového modelování** je přispívat ke kvalitnímu návrhu databází (jak transakčních, zdrojových, tak v daném kontextu zejména analytických), což znamená, že:

Databáze jsou konstruovány efektivně s minimalizací redundancí dat (tedy jejich duplicit, či multiplicit), se zajištěním požadovaného výkonu a doby odezvy při nejčtetnějších dotazech a dalších operacích.

- Databáze se navrhují tak, aby poskytovaly data uživatelům, která odpovídají jejich aktuálním požadavkům, ale i požadavkům, které lze v budoucnu očekávat. To znamená tak, aby bylo databáze možné relativně bezpečně rozšiřovat, udržovat a modifikovat. Úpravy databáze nesmí znamenat nepřiměřená rizika pro kvalitu dat.
- V relační databázi je řešena integrita na úrovni položek (atributů), tabulek i vazeb mezi tabulkami zajišťující přesnost a konsistenci uložených dat po celou dobu provozu,
- Aplikace postavené na databázi lze pak řešit efektivně a návrh databáze neznámá významné omezení pro jejich další údržbu a rozvoj.



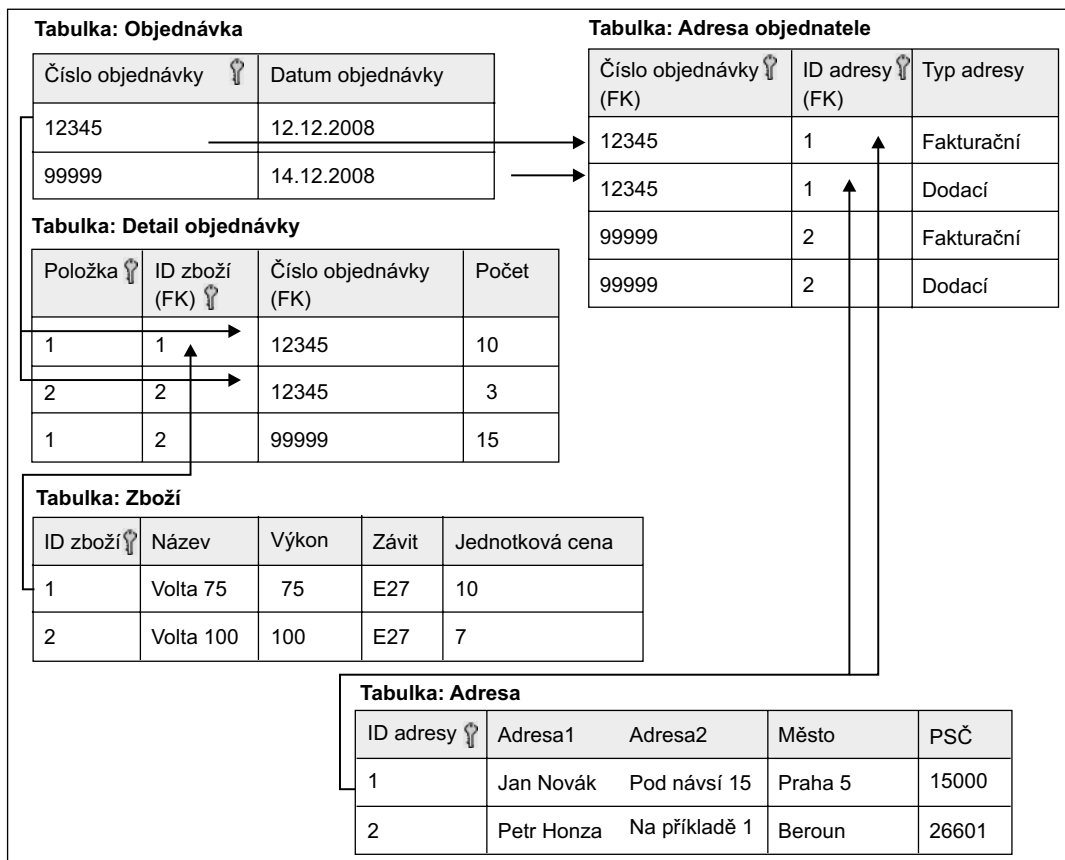
V dalších částech této podkapitoly nemůžeme postihnout celou šíři otázek a postupů datového modelování (k tomu slouží rozsáhlá specializovaná literatura), ale pokusíme se vystihnout pouze nejdůležitější principy a termíny potřebné pro řešení BI i SSBI projektů i jako nezbytný základ pro pochopení dimenzionálního modelování, které je obsahem další podkapitoly.

Datové modelování tak, jak ho budeme dále prezentovat a využívat, je spojené obvykle s **relačními databázemi**. S ohledem na další text je proto účelné zde **rekapitulovat** pouze několik základních termínů.

**Relace** je dvourozměrná datová struktura představovaná tabulkou, kde jeden řádek tabulky odpovídá jednomu záznamu (například jednomu záznamu zboží) a sloupec odpovídá položce (například *ID zboží*, *Název*, *Jednotková cena* atd.). V databázovém pojetí se sloupec tabulky označuje jako atribut, jehož hodnoty spadají do tzv. **domény**. Příklad části takové databáze zachycuje obrázek 1.2. Z obrázku je patrné, že databázi tvoří pět relací, resp. tabulek. Relace (tabulka) *Zboží*, *Adresa* a *Objednávka* se označují jako entitní relace, relace *Adresa objednatele* a *Detail objednávky* popisují vztah mezi entitními relacemi a jsou proto označovány pojmem relace vztahová (*asociativní*).

S databází je spojena také **integrita**, což je pravidlo omezující možné hodnoty atributů, případně možné manipulace se záznamy, které existují ve vazbě k záznamům jiné tabulky. K významným integritním omezením patří existence **primárního klíče** (tzv. **entitní integrita**). Tím se zajišťuje jednoznačná identifikace záznamu v tabulce (žádné dva záznamy nemohou mít stejnou hodnotu primárního klíče). V příkladu (viz obrázek 1.2) jsou položky, které hrají roli primárního klíče, označeny symbolem klíče. Primární klíč může vzniknout i jako kombinace hodnot zvolených atributů, která se pak označuje jako **složený klíč**. Příkladem primárního klíče může být *Číslo objednávky* v tabulce *Objednávka*. S termínem primární klíč souvisí i umělý primární klíč (*surrogate key*), který nemá nějaký specifický podnikový význam (například *ID zboží*), ale vzniká postupným načítáním stanovené hodnoty (většinou 1) při přidávání dalších záznamů do tabulky databáze.

Další typ integritního omezení je **doménová integrita**, která zajišťuje, aby údaj uvedený jako hodnota atributu byl vybrán z množiny definovaných přípustných hodnot. Například pro sloupec (doménu) *Typ adresy* v tabulce *Adresa objednatele* může být takovým doménovým omezením možnost uvedení buď hodnoty *Fakturační*, nebo hodnoty *Dodací*.



**Obrázek 1.2** Tabulky a jejich vazby v relační databázi (Zdroj: Gála a kol.: Podniková informatika, Grada Publishing, 2009)

Posledním typem integritního omezení je tzv. **referenční integrita**, která je reprezentována **cizím klíčem** (*foreign key*). Cizí klíč zajišťuje provázanost tabulek, například mezi tabulkou *Zboží* a *Detail objednávky* je cizím klíčem *ID zboží* umístěný v tabulce *Detail objednávky*. Na obrázku je položka, která je cizím klíčem označena zkratkou FK. Princip referenční integrity zajišťuje:

- Aby nemohla být vložena jako hodnota atributu, který je označen jako cizí klíč, hodnota, jež nemá odpovídající hodnotu atributu v tabulce, jejíž klíč jako cizí uvádíme. Tak je například zajištěn vztah mezi tabulkou *Objednávka* a *Detail objednávky*, kdy platí, že v tabulce *Detail objednávky* se jako hodnota atributu *Číslo objednávky* nesmí objevit jiné číslo než to, které je uvedeno u některého ze záznamů v tabulce *Objednávka*.
- Aby z tabulky nebylo možné odstranit záznam, jehož hodnota je uvedena jako cizí klíč v tabulce jiné. To například znamená, že nelze odstranit záznam z tabulky *Zboží*, jehož ID zboží je rovno 1, protože se tato hodnota vyskytuje jako hodnota cizího klíče v tabulce *Detail objednávky*.

- Uvedené základní principy relačních databází se promítají do řešení datových modelů. Návrh datových modelů se realizuje na několika úrovních, resp. v několika základních krocích. Tyto kroky postupují obvykle od konceptuálního návrhu přes logický až k fyzickému.

1. **Konceptuální návrh databáze:** Prvotní návrh databáze, tedy její struktury, vazeb a všech potřebných charakteristik, obvykle začíná na konceptuální úrovni. Ta je zaměřena na věcnou podstatu dat a nezohledňuje fyzické, resp. technologické charakteristiky databáze, ale pouze její požadovanou strukturu a logiku uspořádání dat, které jsou nezávislé na konkrétním softwaru pro řízení databáze. Účelem konceptuálního návrhu databáze je specifikovat typy datových objektů (typy entit), jejich vlastnosti (atributy), přirozenou identifikaci, vazby mezi entitami a jejich vlastnosti. V konceptuálním modelu se v entitách nemodelují cizí klíče, vazby mezi entitami jsou vyjádřeny pouze graficky.
2. **Logický návrh databáze:** Na základě konceptuálního návrhu databáze se řeší logický návrh databáze, jehož cílem je detailně vymezit strukturu databáze založené již na určitém logickém schématu dat, v našem případě relačním. Tedy především definovat struktury databázových tabulek, jejich atributy, primární klíče a relace mezi tabulkami prostřednictvím cizích klíčů.
3. **Vytvoření, případně generování fyzického návrhu databáze:** Na základě logického návrhu databáze se definuje fyzický datový model, který specifikuje již potřebné implementační charakteristiky ve vztahu k příslušnému databázovému prostředí, v němž bude databáze realizována. Tím se myslí například formáty jednotlivých datových položek, jejich rozsah, resp. délka, definování cizích klíčů a vztahů mezi tabulkami a další charakteristiky.

Datové modely budou jednou ze součástí řešení dimenzionálních modelů v dalších částech kapitoly a následně budou prezentovány v souvislosti s jednotlivými produkty SSBI (viz kapitoly 5–11). V souvislosti s datovým modelováním je ještě třeba vymezit tyto termíny:

- **Entita** představuje z pohledu konceptuálního datového modelu dále nedělitelnou jednotku reálného světa, o které se budou v databázi vést příslušná data. V databázi obvykle entitě odpovídá jeden řádek v databázové tabulce. Typ entity pak představuje zobecnění pro všechny entity se stejnými vlastnostmi (například typem entity je *Zákazník* a entitou je jeden každý zákazník). Typu entity odpovídá v databázi jedna databázová tabulka. Typ entity má vlastní název, identifikátor a její obsah je dále definován skupinou atributů, které jsou následně, v rámci logického návrhu databáze, základem pro definování struktury databázové tabulky. (*Entita* a *Typ entity* se oba běžně používají pro termín „typ entity“, *Výskyt* pak pro jednu konkrétní entitu, tedy data o konkrétním sledovaném objektu.)
- **Vazby** se definují mezi dvěma entitami použitím symbolu vazby s doplněním názvu vazby a dalších jejích charakteristik, včetně kardinality.
- **Kardinalita vazby** (násobnost a povinnost) znamená vyjádření možnosti výskytu počtu entit na obou stranách vazby. Existují základní vztahy 1 : 1, 1 : M, M : N. Tedy pro vazbu 1 : M (*One to Many*) to znamená, že například jeden obchodník odpovídá za více objednávek zboží a jedna objednávka zboží je vyřizována pouze jedním obchodníkem. Pokud bychom danou vazbu definovali vztahem M : N (*Many to Many*), pak to bude znamenat, že jeden obchodník vyřizuje více objednávek, ale na druhé straně se na jedné objednávce může podílet více obchodníků. Povinnost vazby pak vyjadřuje nutnost existence dané vazby ke každému výskytu entity, například obchodník nemusí být odpovědný za žádnou objednávku, ale objednávka musí mít obchodníka, který je za ní odpovědný.

- **Normalizace databáze:** Normalizace databáze se chápe jako postup, kdy je struktura dat v relační databázi přeorganizována tak, aby byly minimalizovány duplicity nebo multiplicity uložení dat v databázi tak, aby ideálně jedna změna v realitě vyžadovala pouze jednu změnu v datech v databázi (viz dále).

S databázemi a datovým modelováním je spojena celá škála dalších termínů a pravidel, o nichž je řeč v různé specializované literatuře, například HERNANDEZ, M. J. (2006), KIMBALL, R., ROSS, M. (2010), PROVOST, F., FAWCETT, T. (2013), ale jako předpoklad pro další text zůstaneme zatím pouze u výše vybraných.

## 1.2 Dimenzionální modelování

Uplatnění dimenzionálního modelování je předpokladem pro naplnění účelu analytických aplikací (BI i SSBI), tedy:

- Prezentovat uživatelům potřebné informace co nejjednodušším způsobem.
- Poskytovat odpovědi na dotazy s minimální dobou odezvy.
- Zajišťovat relevantní informace přesně odpovídající definovaným podnikovým procesům nebo úlohám.

Dimenzionální modely a na nich postavené aplikace jsou pak pro uživatele mnohem pochopitelnější než normalizované modely pro transakční aplikace.



Dimenzionální modelování vychází z poznání a zhodnocení potřeb řízení dané organizace. Jeho obsahem je:

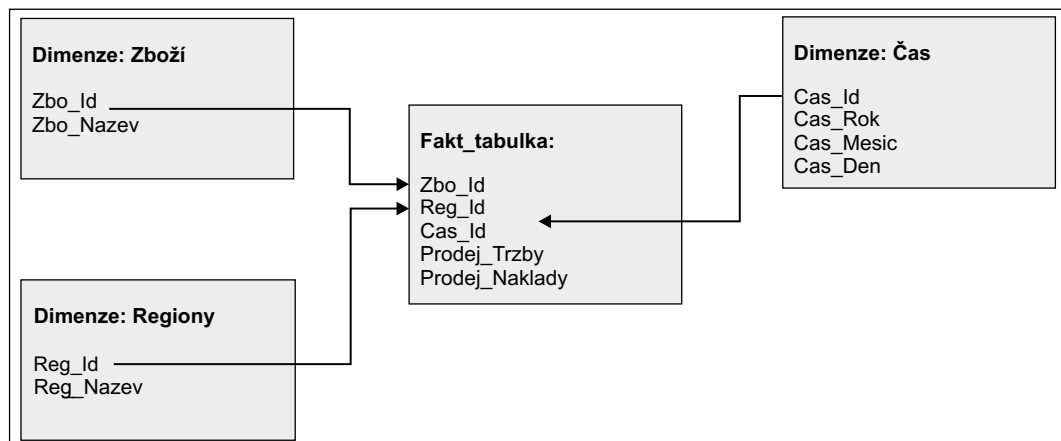
- Vymezení všech dimenzí, jejich obsahu, včetně vnitřní hierarchie prvků, a dílčích charakteristik jednotlivých dimenzí.
- Určení soustavy sledovaných ukazatelů (faktů nebo metrik) a jejich dílčích charakteristik.
- Specifikace vazeb mezi ukazateli a odpovídajícími dimenzemi.

Uložení dat a operace s nimi jsou zde založeny na **multidimenzionalitě**, a to vhodným řešením organizace dat, resp. databázových schémat (viz obrázek 1.3).

V centru schématu je **tabulka faktů**, tedy tabulka sledovaných hodnot ekonomických a dalších ukazatelů (*Prodej\_Trzby*, *Prodej\_Naklady*) identifikovaných klíčem složeným z cizích klíčů dimenzionálních tabulek (*Zbo\_Id*, *Reg\_Id*, *Cas\_Id*), nebo vlastním umělým primárním klíčem. Cizí klíče potom slouží pouze pro vazby do dimenzionálních tabulek. Tabulka faktů se tak váže na dimenzionální tabulky, tabulky faktů se ale neváží na jiné faktové tabulky.

**Dimenzionální tabulky** slouží jako úložiště vlastností, tedy popisných a doplňujících informací o hodnotách ukazatelů uložených v tabulce faktů. Například k číslu zákazníka v tabulce faktů jsou v dimenzionální tabulce obsaženy všechny potřebné informace o zákazníkovi s odpovídajícím číslem. Většinou si dimenzionální tabulku lze představit jako **číselník**. Pro reálné dimenzionální tabulky je ale typické velké množství atributů, pro něž se nejlépe hodí atributy textové a diskrétní.

Uvedené schéma na obrázku 1.3 představuje vybranou část souhrnného příkladu, který je podrobněji popsán v příloze.



**Obrázek 1.3** Multidimenzionální uložení dat

## 1.2.1 Dimenzionální modelování – hrubý dimenzionální model

Výsledné dimenzionální modely by měly vytvořit potřebný základ pro realizaci analytických aplikací, a to s těmito vlastnostmi:

- Jejich výsledná podoba je relativně jednoduchá, založená na určité symetrii a standardech (fakta – dimenze), což umožňuje poměrně dobrou a rychlou orientaci v logice řešení jak informatikům, tak uživatelům.
- Jednoduchost řešení nabízí i lepší možnosti zvyšování provozního výkonu aplikací.
- Jednoduchost a přehlednost je rovněž základem i pro snadnější úpravy a rozšiřování obsahu řešení, tedy doplňování nových dimenzí, atributů apod.

Podstatou **hrubého dimenzionálního modelu** je vymezit obsah řešených BI a SSBI aplikací bez ohledu na jejich technickou realizaci v konkrétní databázi. Součástí řešení je i stanovení granularity, tedy úrovně detailu sledovaných hodnot v analytické databázi. Ta by měla být co nejvyšší, aby bylo možné realizovat co nejpodrobnější analytické operace. Na druhé straně je nutné při dané granularitě odhadnout objem databáze a její očekávaný růst. Znamená to určit formou běžného textu nebo s použitím tabulek, matic, případně specifických schémat následující:

1. Dimenze, s jejichž pomocí se budou ukazatele analyzovat.
2. Ukazatele (fakta, metriky), která budou v aplikaci sledována a analyzována.
3. Vazby ukazatelů a dimenzí, tedy který ukazatel se bude analyzovat podle kterých dimenzí.

V následujícím textu jsou uvedeny charakteristiky dimenzí i ukazatelů, jež má obsahovat hrubý dimenzionální model. I v tomto případě je nutné zdůraznit možnosti **modifikace nebo redukce** jeho obsahu. Řada charakteristik totiž vychází z potřeby sladění představ o výsledných řešeních mezi širokou škálou uživatelů, což je typické zejména pro komplexní systémy Business Intelligence. Na druhé straně aplikace zaměřené pouze na individuální potřeby jednotlivců rozsáhlou dokumentaci dimenzionálních modelů někdy nepotřebují. Je proto

na úvaze konkrétních uživatelů i IT analytiků, v jakém rozsahu a podrobnosti budou hrubé dimenzionální modely tvořit.

V případě návrhu **ukazatelů** se určují tyto charakteristiky (podrobněji viz kapitolu 1.3):

- Jednoznačná **identifikace** ukazatele – například *Prodej\_Tržby* – pro účely definování vazeb ukazatele na jednotlivé dimenze a dalšího navrhování tabulek faktů a datových modelů.
- Plný název ukazatele, například *Tržby z prodeje v KČ*.
- Obsahové **vymezení a účel** ukazatele – je dobré vymezit ukazatel co nejpřesněji a nejkonkrétněji tak, aby byl vytvořen podklad pro shodu mezi uživateli na jeho interpretaci.
- **Zdroje dat** pro ukazatel – z jakých databází nebo souborů lze hodnoty ukazatele získávat, případně je pořizovat manuálním způsobem, expertním odhadem apod.
- **Kalkulace** ukazatele – výpočty pro základní i související ukazatele.
- **Typ**, formát dat, například numerické, textové.
- **Měrná jednotka** – například KČ, % apod.
- Možnost **agregace** ukazatele, tedy aditivní (A), neaditivní (N), semiaditivní (S).
- **KPI**, tedy zda ukazatel představuje klíčový indikátor výkonnosti (*Key Performance Indicator*), nebo klíčový cílový indikátor byznysu (*KGI, Key Goal Indicator*).
- **Vazby** ukazatele na definované dimenze formou jednoduché matice.

V případě návrhu **dimenzí** se určují a dokumentují následující charakteristiky (podrobněji viz kapitolu 1.4):

- **Identifikace** dimenze – podle stanovených dohodnutých standardů, například *DI\_Zbozi*.
- Plný **název** dimenze, například *Zboží prodávané podnikem*.
- **Obsah**, tedy podrobnější textové vyjádření obsahu dimenze tak, aby umožňovalo její sjednocené chápání mezi různými uživateli a analytiky.
- **Typ** dimenze, tedy zda jde o dimenzi časovou, STAR, SNOWFLAKE, degenerovanou (viz kapitoly 1.4.2 a 1.4.3).
- **Zdroj dat** pro dimenzi, resp. její prvky, například databáze *DB\_Zbozi*, tabulka číselníku v Excelu, textový soubor apod.
- Struktura **prvků** dimenze, resp. jejich hierarchie tak, aby bylo zřejmé, v jaké celkové struktuře budou prvky dimenze uspořádány, samozřejmě bez jejich výčtu, ale s příklady:
  - (1) Kategorie zboží (například *Audio a kino, Počítače, Monitory, Spotřební elektronika* atd.).
    - (a) Skupiny zboží (například *Auto hifi, Notebooky, Ledničky* atd.).
    - (i) Jednotlivá zboží, zbožové položky, například (*Autoradio Logik, Acer Aspire One A150* atd.).
- **Atributy** dimenze, tedy struktura záznamu dimenzionální tabulky, viz například tabulku 1.3.
- Doplnující **poznámky** k vytvoření a využití dimenze, vazby na legislativní nároky apod.

Na tomto místě jsme uvedli základní charakteristiky navrhované dimenze, které se podle potřeby v průběhu dimenzionálního modelování dále doplňují a zpřesňují.

Následující dvě tabulky dokumentují specifikaci jednotlivých dimenzí, ukazatelů a jejich vazby na dimenze.

**Tabulka 1.1** Přehled dimenzí a jejich charakteristik

| ID            | Název               | Obsah                                                                            | Typ       | Zdroj           | Struktura(y)                                                                                             | Atributy                                   | Poznámky                                     |
|---------------|---------------------|----------------------------------------------------------------------------------|-----------|-----------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------|----------------------------------------------|
| DI_Cas        | Čas                 | Struktura časových jednotek pro sledování ukazatelů v časovém vývoji             | časová    | manuální        | Rok – Kvartál – Měsíc<br>Rok – Měsíc –<br>– Týden – Den                                                  | Cas_Id<br>Cas_rok,<br>...                  | Datová dimenze sleduje čas vzniku obchodu    |
| DI_Regiony    | Regiony             | Regiony, kde podnik má své obchodní a další aktivity                             | SNOWFLAKE | DB: Obchod      | Kontinent – Stát –<br>– Region                                                                           | Reg_Id<br>Reg_Nazev<br>Reg_...             |                                              |
| DI_Pod_Utvary | Podnikové útvary    | Organizační útvary – divize, oddělení – pro sledování obchodní výkonnosti útvarů | SNOWFLAKE | Organizační řád | Kategorie útvaru (například Obchod, Výroba atd.) – Útvar (například Prodej Audio, Prodej počítače atd.), | Utv_Id<br>Utv_Kod<br>Utv_Nazev<br>Utv_.... | Může zahrnovat i další organizační struktury |
| DI_Zakaznici  | Struktura zákazníků | Struktura a podstatné charakteristiky zákazníků podniku                          | STAR      | Controlling     | Kategorie zákazníka – Zákazník                                                                           | Nak_id_druhu<br>Nak_nazev_druhu<br>...     | Lze doplnit i další členění zákazníků        |
| DI_Zbozi      | Zboží               | Struktura položek zboží a služeb nabízených a prodávaných zákazníkům             | STAR      | DB: Zbozi       | Kategorie zboží – Skupiny zboží –<br>– Zbožové položky                                                   | Zbo_id<br>Zbo_Kategorie_id ...             | Vychází z katalogu zboží                     |
| ...           |                     |                                                                                  |           |                 |                                                                                                          |                                            |                                              |

**Tabulka 1.2** Přehled ukazatelů (faktů)

| ID               | Název           | Obsah                                                             | Zdroj/Vyp      | Typ, formát    | Jednotka  | Agregace | KPI | Dim.: |     |     |     |     |     |
|------------------|-----------------|-------------------------------------------------------------------|----------------|----------------|-----------|----------|-----|-------|-----|-----|-----|-----|-----|
|                  |                 |                                                                   |                |                |           |          |     | Cas   | Reg | Utv | Zak | Zbo | ... |
| Prodej_Trzby     | Objem tržeb     | Objem tržeb za prodané zboží a služby                             | DB: Ucetnictvi | Numeric (10,2) | tis. Kč   | A        | Ne  | X     | X   | X   |     | X   |     |
| Prodej_Naklady   | Objem nákladů   | Objem veškerých nákladů spojených s prodejními aktivitami podniku | DB: Ucetnictvi | Numeric (10,2) | tis. Kč   | A        | Ne  | X     | X   | X   | X   |     |     |
| Prodej_Reklamace | Počet reklamací | Počet reklamací od zákazníků na zboží a služby                    | CRM            | Numeric (6,0)  | reklamace | A        | Ano | X     | X   | X   | X   | X   |     |
| ...              |                 |                                                                   |                |                |           |          |     |       |     |     |     |     |     |

Podstatným aspektem této úrovně řešení je určení náplně dimenzí a jejich prvků, to znamená například jací konkrétní zákazníci budou naplňovat dimenzi *DI\_Zakaznici*, jaká konkrétní zboží dimenzi *DI\_Zbozi* apod., podle struktur definovaných v předchozí tabulce.

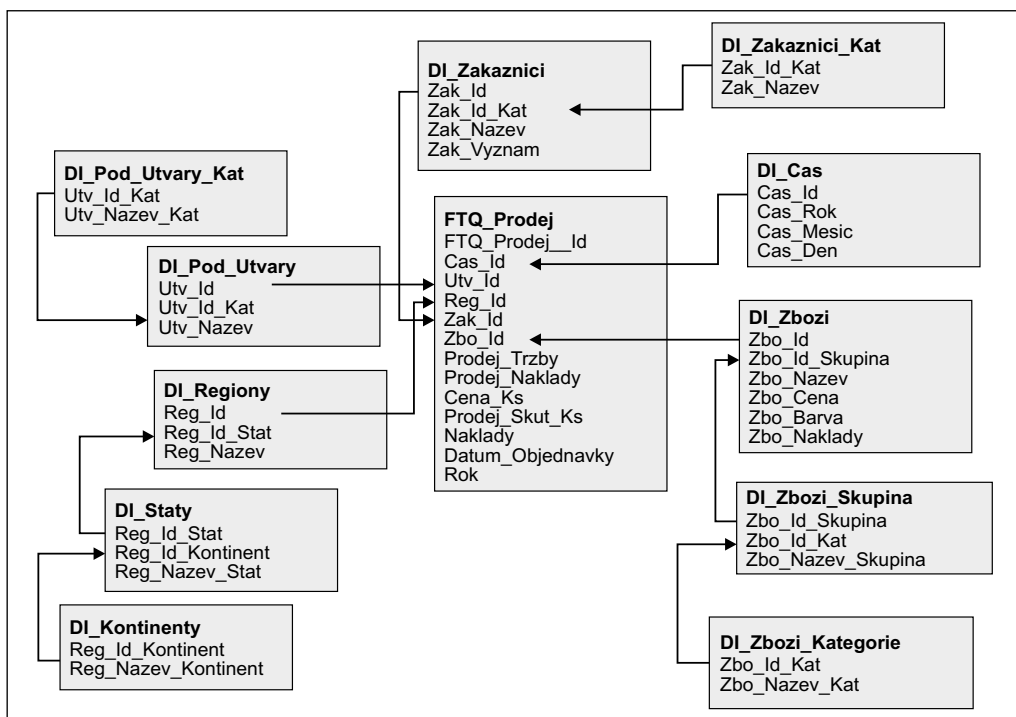
Jedním z problémů je však to, že vazby ukazatele s definovanými dimenzemi označenými v matici v buňkách příslušné řádky X musí mít adekvátní obraz ve zdrojových databázích. To znamená, že například hodnoty ukazatele *Prodej\_Trzby* musí být identifikovány pomocí prvků všech uvedených dimenzí. Pokud tomu tak není,

pak je hodnota ukazatele neúplně identifikována, musí být řešena například v rámci čištění a transformací dat, nebo specifickou funkcí databázového systému.

Speciální místo v modelu má časová dimenze, tedy jaká bude struktura časových intervalů (roky, kvartály, měsíce), zda těchto časových struktur bude současně více, zda se bude k aktuálnímu datu nějakým způsobem měnit (například na dekády, dny), zda se budou nějaké starší časové úseky přesouvat z provozního řešení do archivu (tzv. *aging*) apod.

Další aktivitou dimenzionálního modelování je analýza a návrh dimenzionálních modelů na bázi principů datového modelování. V návaznosti na hrubý dimenzionální model se upřesňuje obsah a vazby databázových tabulek dimenzí a faktů s určením běžných charakteristik (formátu, rozsahu) jejich jednotlivých atributů. Podstata a různé možnosti realizace tabulek faktů jsou dále obsahem podkapitoly 1.3, tabulek dimenzí pak podkapitoly 1.4. V souvislosti s produkty SSBI (kapitoly 5–11) je však dobré upozornit na to, že se často struktura tabulek vygeneruje automaticky při vstupu, resp. transformaci dat ze vstupních zdrojů, například z tabulky Excelu, zdrojových databázových tabulek apod. Minimálně konceptuální datový model je však účelné vždy udělat.

V rámci této fáze je třeba ještě znovu verifikovat, zda data pro navrhované dimenzionální modely jsou ve zdrojových databázích k dispozici a jsou kvalitní. Podstatou této aktivity je transformace výsledků hrubého dimenzionálního modelu do relačních databázových struktur. Příklad datového modelu pro analytickou aplikaci řízení prodeje dokumentuje obrázek 1.4.



**Obrázek 1.4** Příklad – datový model pro BI a SSBI aplikace

## 1.2.2 Návrh datového modelu pro SSBI

V dalších podkapitolách uvedeme podstatné charakteristiky hlavních součástí dimenzionálních modelů, tedy tabulek faktů, tabulek dimenzí a jejich vazeb.

## 1.3 Tabulky faktů

Tabulka faktů uchovává hodnoty sledovaných ukazatelů a jsou pro ni podstatné vlastnosti, které budeme dokumentovat na následujícím příkladu obchodního podniku (viz tabulku 1.3). V souvislosti s ukazateli se používají i termíny fakta nebo metrika.

**Tabulka 1.3** Příklad tabulky faktů

| Zbo_Id | Reg_Id | Cas_Id | Prodej_Trzby | Prodej_Naklady |
|--------|--------|--------|--------------|----------------|
| 111    | 0105   | 030115 | 749 750.00   | 562 300.00     |
| 114    | 0107   | 030116 | 741 470.00   | 556 100.00     |
| 115    | 0121   | 030216 | 539 820.00   | 404 840.00     |

Zjednodušená struktura tabulky faktů může být následující:

- **Zbo\_Id**: Identifikátor prodaného zboží.
- **Reg\_Id**: Identifikátor regionu, místa, kde se zboží prodalo.
- **Cas\_Id**: Identifikátor času, kdy se zboží prodalo.
- **Prodej\_Trzby**: Tržby za prodané zboží (daného zboží, na daném místě a v daném čase).
- **Prodej\_Naklady**: Náklady na prodané zboží (daného zboží, na daném místě a v daném čase).

**Sloupce** tabulky faktů jsou vesměs buď klíčové atributy (**Zbo\_Id**, **Reg\_Id**, **Cas\_Id**), nebo hodnoty ukazatelů (**Prodej\_Trzby**, **Prodej\_Naklady**). Klíčové atributy reprezentují jednotlivé dimenze a jejich hodnoty jsou prvky těchto dimenzí, například 111 je prvkem dimenze Zboží.

**Řádky** v tabulce představují ve své podstatě jednotlivá měření (v obchodě, výrobě apod.) a většinou jsou přiřazována na co nejnižší úrovni detailu, tedy pouze na úrovni listů ve strukturách použitých dimenzí. Na odpovídající tabulky dimenzí tabulka faktů odkazuje prostřednictvím **cizích klíčů**, například **Zbo\_Id** váže tabulku faktů na dimenzionální tabulku **DI\_Zbozi** apod. Všechny cizí klíče reprezentující vazby do odpovídajících tabulek dimenzí tvoří v tomto případě **složený klíč** tabulky faktů, například **Zbo\_Id**, **Reg\_Id**, **Cas\_Id**. Jinou možností je, že každý řádek tabulky faktů má vlastní, většinou umělý klíč, což je mnohdy s ohledem na riziko duplicit klíčů bezpečnější. Klíčové položky jsou vesměs celočíselného typu.

Příklad vazby tabulky faktů na tabulky dimenzí dokumentuje obrázek 1.3. Kardinalita vazeb tabulek dimenzí na tabulku faktů je vždy 1 : N. Tabulka faktů tak vlastně současně vyjadřuje vazby M : N mezi tabulkami dimenzí navzájem.

### 1.3.1 Granularita v tabulce faktů

**Granularita** určuje úroveň podrobnosti údajů uložených v tabulce faktů. Granularita údajů v tabulce faktů je přímo závislá na počtu a úrovni podrobnosti dimenzí připojených k příslušné tabulce faktů. Například máme-li v časové dimenzi definovanou strukturu až na jeden den a v dimenzi *DI\_Zboží* na jeden dílčí produkt, pak každý záznam v tabulce faktů (*grain, zrno*) je na úrovni *objemu prodeje jednoho dílčího zboží za jeden den*. Tím je dána granularita tabulky faktů, obdobné je to i ve vztahu k ostatním dimenzím.

Nízká granularita (tedy nízká úroveň detailu uložených dat) znamená nemožnost pracovat s podrobnými daty, tedy podle dnů, jednotlivých zboží, regionů, prodejců apod. Naopak vysoká granularita (tedy vysoká úroveň detailu dat) možnosti detailních analýz nabízí, ale na druhé straně znamená i podstatně vyšší nároky na diskový prostor analytické databáze a jeho další provozní charakteristiky.



Pro řešení úrovně granularity existují některá obecná doporučení:

- Pokud to technické kapacity dovolují, měla by být data uložena s nejvyšší možnou granularitou.
- Data vstupující do analytické databáze z různých zdrojů je účelné transformovat na stejnou nebo srovnatelnou granularitu.

V případě druhého uvedeného doporučení existují v praxi určité problémy a s nimi spojené analytické úlohy. Typickým příkladem mohou být data získávaná z obchodních objednávek, faktur apod. Některé údaje (fakta) se mohou vázat k objednávce jako celku (takzvaně k její hlavičce, například náklady na dopravu), a některé k jednotlivým objednávaným zbožíovým položkám (například hodnota dodávky příslušného zboží). Jde evidentně o různou granularitu uvedených dat.

Úlohou analytika je převést tyto údaje na stejnou – vyšší – granularitu, tedy v našem případě rozpočítat náklady na dopravu na jednotlivé zbožíové položky. Tato operace se v dimenzionálním modelování označuje jako **alokace** (nikoli pouze nákladů, jak je známá z ekonomických disciplín). Při řešení této úlohy je však nutné nejprve rozhodnout, zda je taková alokace možná, zda má smysl a současně posoudit, kdo z uživatelské (byznys) sféry se na formulaci pravidel pro alokaci bude podílet. Pokud se alokace ukáže jako neúčelná, pak souhrnné hodnoty (například dopravné) musí vytvořit samostatnou agregovanou faktovou tabulku. Nedoporučuje se kombinovat fakta s různou granularitou do jedné faktové tabulky.

### 1.3.2 Typy tabulek faktů

V analytických databázích existují tři hlavní **typy tabulek faktů** vzhledem k jejich granularitě dat.

**Transakční tabulky faktů** jsou založeny na tom, že detailní informace vstupující do analytické databáze jsou vázány na jednotlivé transakce a pohybují se na nejvyšší možné granularitě dat. Z toho vyplývá, že časový úsek nebude stejný, ale bude záviset na době výskytu jednotlivých transakcí. Je potom vhodné rozdělit časovou dimenzi na „den“ a „čas“, tedy čas jednotlivých transakcí v rámci dne. Transakční tabulky faktů patří v praxi k těm často využívaným.

**Periodické snímkané tabulky faktů** jsou v praxi nejpoužívanější. Data vstupují do datového skladu v pravidelných, předem definovaných časových úsecích (snímcích, například dnech) a vyjadřují souhrnné hodnoty ukazatelů za celý časový snímek (například celkový objem transakcí za daný časový úsek). Tento typ

tabulek se nejvíce využívá i pro odhadování, resp. predikci trendů vybraných ukazatelů. Návrh periodických snímkaných tabulek se obvykle úzce váže na návrh transakčních tabulek faktů a sdílí většinu společných dimenzí. Volba časového úseku pak samozřejmě ovlivňuje nejen granularitu, ale i nárůst objemu dat v datovém skladu.

**Akumulované** (též někdy **stavové**) **snímkané tabulky faktů** jsou rovněž závislé na výskytu transakcí, ale jejich hodnoty se v čase postupně aktualizují. Například při postupném objednávání zboží na sklad se tak udržuje přehled o aktuálním stavu a vývoji dané objednávky. Tyto tabulky se v praxi využívají méně často.

Souhrnné porovnání jednotlivých typů tabulek faktů dokumentuje tabulka 1.4.

**Tabulka 1.4** Porovnání typů tabulek faktů (zdroj: Kimball, Ross, 2010)

| Charakteristika                   | Transakční                          | Periodická snímkaná                                | Akumulovaná snímkaná                                                |
|-----------------------------------|-------------------------------------|----------------------------------------------------|---------------------------------------------------------------------|
| Časová perioda                    | Časový okamžik (čas transakce)      | Pravidelné, předem určené intervaly                | Nedefinovaný časový rozsah                                          |
| Granularita                       | 1 záznam = 1 transakce              | 1 záznam = 1 časový interval                       | 1 záznam (postupně aktualizovaný)                                   |
| Plnění (load) tabulky faktů       | Přidávání záznamů ( <i>Insert</i> ) | Přidávání záznamů ( <i>Insert</i> )                | Přidávání záznamů ( <i>Insert</i> ) a aktualizace ( <i>Update</i> ) |
| Aktualizace záznamů tabulky faktů | Nerealizuje se                      | Nerealizuje se                                     | Realizuje se vždy při změně                                         |
| Časová dimenze                    | Datum transakce                     | Datum konce časového intervalu                     | Více kalendářních dat pro standardní provádění změn                 |
| Fakta (ukazatele)                 | Obsah transakční aktivity           | Obsah odpovídající definovanému časovému intervalu | Obsah odpovídající celému životnímu cyklu dat                       |

### 1.3.3 Agregace dat

Značná část **dotazů** do analytické databáze směřuje na agregovaná data, například kolik se prodalo zboží X za první čtvrtletí v daném regionu. Jestliže máme uložena fakta v granularitě odpovídající dnům, pak je pro získání výsledku tohoto dotazu třeba sečíst jednotlivé prodeje podle zadaných kritérií. Z toho plyne významná vlastnost údajů v tabulce faktů, tedy jejich možností agregací. Data v tabulkách faktů se podle toho rozdělují na:

- **Plně aditivní**, u nichž lze prakticky vždy získat smysluplné agregované údaje podle všech využitých dimenzí, například u hodnot *Prodej\_Trzby*, *Nakup\_Objem* apod.
- **Neaditivní**, v tomto případě nemají agregované hodnoty smysl, například *Sazba\_Marže* %, a nahrazují se v agregacích počty jejich hodnot nebo průměrnými hodnotami.
- **Semiaditivní** agregované hodnoty mají smysl pouze podle určitých dimenzí, například u položky *Zasoba\_Ks* mají agregace smysl podle dimenze zboží, ale nikoli v čase. Místo agregovaných hodnot se v tomto případě používají průměrné hodnoty za určitá časová období.

Tuto charakteristiku je tedy nutné brát vždy pozorně v úvahu, aby aplikace nebo dotazy do databází poskytovaly smysluplné výsledky. Nejpoužívanějšími ukazateli v tabulkách faktů jsou tak numerické a plně aditivní ukazatele.

### 1.3.4 Měrné jednotky

Tabulky faktů obsahují ukazatele, které potřebují různí uživatelé sledovat v různých měrných jednotkách, například počty vyrobených produktů v kusech, v tisících, krabicích, v paletách apod. Nabízejí se dvě možnosti: buď umístit jednotky a přepočítací koeficienty například do produktové dimenze, nebo je umístit přímo do jednotlivých záznamů tabulky faktů. S ohledem na riziko chyb a možné změny v koeficientech se doporučuje využívat spíše druhou variantu, tedy umísťovat je do záznamů tabulky faktů.

### 1.3.5 Rozsah tabulky faktů

Tabulky faktů zabírají v datovém skladu obvykle kolem **90 % jeho celkové kapacity** (oproti cca 10 % tabulek dimenzí). Je pro ně charakteristické, že tento rozsah je dán obrovským počtem jejich řádků, záznamů (například pro každý prodej, každý telefonní hovor apod.). Na druhé straně proto existuje snaha **omezit jejich rozměr** co do počtu sloupců a rozsahu jednotlivých sloupců. Dalším způsobem řešení je určení **granularity** dat podle období, například pro posledních aktuálních 60 dnů se využije denní granularita tabulky faktů, pro starší období pak granularita nižší.

### 1.3.6 Zdroje a kalkulace ukazatelů

Tabulka faktů obsahuje základní, elementární hodnoty ukazatelů vstupující **ze zdrojových databází i hodnoty kalkulované**, tedy v tomto případě kalkulace v rámci jednoho záznamu (například  $Prodej\_Zisk = Prodej\_Tržby - Prodej\_Naklady$ ). Kalkulace se mohou provádět na úrovni transformací dat, analytické databáze, nebo na úrovni analytických aplikací. Obvykle je užitečné u aditivních faktů ukládat kalkulované hodnoty přímo do analytické databáze, neboť se tak zajistí dostupnost těchto dat všem uživatelům, bez nutnosti kalkulace opakovat v různých aplikacích. Oproti tomu stojí někdy fakt, že výpočty v analytických nástrojích (Excel atd.) mohou být jednodušší.

### 1.3.7 Tabulky faktů bez ukazatelů

Dimenzionální modely obsahují i tzv. **tabulky faktů bez ukazatelů, faktů** (*factless fact table*), které nemají žádné ukazatele a využívají se například pro zjišťování počtu určitých událostí. To znamená, že každý výskyt záznamu v tabulce faktů s daným složeným klíčem indikuje vznik události: například daná činnost je součástí procesu, zboží bylo zařazeno do marketingové akce apod. U nich lze pak sledovat souhrnné hodnoty pouhou sumarizací počtu záznamů v členění podle klíče.



Při řešení tabulek faktů jde – kromě uvedených principů – také o obsahové vymezení relevantních **ukazatelů (faktů, metrik)** pro příslušné úlohy nebo procesy v podnikovém řízení. Jejich přehled by překračoval zaměření této knihy, a proto odkazujeme na web MBI (<http://mbi.vse.cz>, volba **Metriky**), kde najdete jejich charakteristiky, včetně vazeb na příslušné úlohy a dimenze.

## 1.4 Tabulky dimenzí

Tabulky dimenzí jsou de facto podnikové číselníky, například dimenze zboží, zákazníků, dodavatelů atd., a to se všemi možnostmi a problémy, jež jsou s nimi spojeny. Dimenzionální tabulka obsahuje vedle klíčových (obvykle numerických) atributů další řadu většinou textových atributů popisujících podstatné charakteristiky jednotlivých produktů, zákazníků apod. Jednoduchý příklad struktury dimenzionální tabulky ukazuje tabulka 1.5. Počet a charakter atributů je dán nároky na zpracování nejrůznějších podnikových reportů vycházejících z dat v analytických databázích. Není proto výjimkou, že dimenzionální tabulky obsahují i padesát a více atributů.

**Tabulka 1.5** Příklad struktury dimenzionální tabulky

|             |                           |
|-------------|---------------------------|
| Zbo_Id      | Identifikátor, klíč zboží |
| Zbo_Nazev   | Plný název zboží          |
| Zbo_Nazev_Z | Zkrácený název zboží      |
| Zbo_Cena    | Cena zboží                |
| Zbo_Barva   | Barva zboží               |
| Zbo_Naklady | Standardní náklady        |
| Zbo_MJ      | Měrná jednotka zboží      |
| Zbo_Povrch  | Povrchová úprava zboží    |
| Zbo_Sklad   | Skladovací nároky         |
| Zbo_Balení  | Způsob balení             |
| ...         | ...                       |

Rozdíl mezi dimenzionálními tabulkami a tabulkami faktů je vedle jejich samotného účelu v tom, že tabulky faktů mají relativně malý počet sloupců, ale enormní počet řádků, zatímco u tabulek dimenzí je tomu právě naopak.

Kvalita návrhu celé analytické databáze je tak často určena zejména **kvalitou návrhu dimenzí**.

### 1.4.1 Principy návrhu dimenzionálních tabulek

Z příkladu (tabulka 1.5) vyplývá, že jeden řádek tabulky je vymezen pouze pro **jeden prvek dimenze**. Každý řádek dimenzionální tabulky musí být identifikován svým primárním klíčem (v našem případě identifikátorem zboží: *Zbo\_Id*), který pak reprezentuje i vztah k tabulce faktů (ve vazbě 1 : N), kde je cizím klíčem a zajišťuje podmínku spojení (*join*) mezi tabulkou faktů a tabulkou dimenzí. Musí tedy zachovávat pravidla referenční integrity. Jednotlivé další popisné atributy slouží jako výběrová kritéria v dotazech, nebo jako obsah hlaviček v reportech a pro další operace v uživatelských aplikacích.

Hodnoty dalších **atributů** by měly být převážně textové a diskrétní a měly by spíše obsahovat plné a jasné vyjádření dané charakteristiky, s co nejmenším používáním různých kódů a zkratk. Je nutné si uvědomit, že reporty slouží široké škále uživatelů s nejrůznějším kvalifikačním a informačním vybavením. Na druhé straně v případě potřeby jednoznačného a stručného vyjádření hodnot atributu se nejrůznějším standardním kódům nevyhneme.

Do atributů dimenzionální tabulky se řadí i některé numericky vyjádřené charakteristiky, například velikost zboží, ale v kontextu tabulky mají čistě popisný charakter. Při návrhu analytické databáze je často problémem, **zda příslušný numerický atribut má být součástí určité tabulky faktů nebo tabulky dimenze**. Řešení spočívá v zodpovězení otázky, jestli takový atribut nabývá velkého množství hodnot (tedy se i často mění) a jestli se podílí na běžných kalkulacích – v takovém případě patří do tabulky faktů. Pokud jde o atribut, který nabývá diskrétních, relativně konstantních hodnot a používá se převážně pro výběrové operace v dotazech, pak patří do dimenzionální tabulky. Příkladem může být již zmíněná cena zboží: pokud je relativně stálá a často se nemění, pak je účelné ji zařadit do dimenzionální tabulky, pokud se mění téměř s každým prodejem, patří do jednotlivých řádků prodejů v tabulce faktů.

Další běžnou otázkou při řešení dimenzionálního modelu je, **zda využívat více jednodušších dimenzí, nebo méně dimenzí, ale složitějších a komplexnějších**. Příkladem mohou být jedna dimenze geografická (resp. dimenze regionů) a vedle toho jedna dimenze jednotlivých prodejen. Oproti tomu stojí varianta jedné společné dimenze prodejen, kde na vyšší úrovni budou regiony a na nejnižší úrovni jednotlivé prodejny podle jejich umístění v regionech.

Varianta většího počtu jednodušších dimenzí má tyto výhody:

- jednodušší pro pochopení ze strany uživatelů,
- jednodušší správa a provádění úprav,
- nabízí více možných kombinací mezi různými dimenzemi v analytických aplikacích.

Oproti tomu varianta menšího počtu komplexních dimenzí je výhodná z následujících důvodů:

- zjednodušuje a zpřehledňuje celý dimenzionální model (dimenzí je méně),
- je efektivnější při prohlížení celé struktury a ukazatelů, k nimž se tato dimenze váže.

Obvykle by mělo být k jedné faktové tabulce navázáno maximálně 15 dimenzí. Pokud je jich více, je účelné přistoupit ke slučování více dimenzí do jedné s tím, že v jedné společné dimenzi by měly být prvky se sjednocenými, nebo alespoň souvisejícími atributy. Je tedy nutné volit odpovídající variantu více nebo méně dimenzí podle konkrétních podmínek, složitosti řešení a uživatelských požadavků na reporty a analytické aplikace.

Podstatnou charakteristikou dimenzionálních tabulek je především skutečnost, že **data v dimenzích jsou hierarchicky strukturovaná** tak, aby bylo možné na základě těchto struktur získávat agregované hodnoty ukazatelů v připojených tabulkách faktů, tedy na různé úrovni agregace. Ve své podstatě struktury dimenzí vyjadřují i struktury řízení v podniku. Obvyklé příklady dimenzí tak představuje čas, útvary (resp. organizační struktura podniku), obchodní zástupci, zákazníci, teritoria, zakázky, zboží, služby, dodavatelé, konkurenti a další.

Příklad takové struktury dokumentuje obrázek 1.5.