

**OTEVŘENÉ KNIHY**

# **IT A ANATOMIE FIRMY**

**(Podniková analytika)**

Doc. Ing. Ota Novotný, Ph.D.

Prof. Ing. Alena Buchalcevodá, Ph.D.

Ing. Tomáš Bruckner, Ph.D.

Ing. Iva Stanovská,

Ing. Zuzana Šedivá, Ph.D.

Doc. Ing. Jan Pour, CSc.



**Ota Novotný, Alena Buchalcevoá  
Tomáš Bruckner, Iva Stanovská  
Zuzana Šedivá, Jan Pour**

# **IT A ANATOMIE FIRMY**

**(Podniková analytika)**

**PROFESSIONAL PUBLISHING**

**Doc. Ing. Ota Novotný, Ph.D.**  
**Prof. Ing. Alena Buchalceková, Ph.D.**  
**Ing. Tomáš Bruckner, Ph.D.**  
**Ing. Iva Stanovská,**  
**Ing. Zuzana Šedivá, Ph.D.**  
**Doc. Ing. Jan Pour, CSc.**

## **IT A ANATOMIE FIRMY** **(Podniková analytika)**

Kniha byla doporučena k vydání vědeckou radou nakladatelství.

© Autoři  
Edition © Professional Publishing s.r.o.  
Obálka: Jan Mottl  
První vydání, 2023

**ISBN 978-80-88260-70-7**

# Obsah

Úvod.....	13
<b>A) Základní principy podnikové analytiky.....</b>	<b>16</b>
<b>1. Obsah podnikové analytiky.....</b>	<b>17</b>
1.1 Vymezení a kategorizace podnikové analytiky.....	17
1.2 Úlohy, procesy.....	18
1.3 Metriky, dimenze.....	20
1.4 Data a datové zdroje.....	21
1.4.1 Řízení a správa metadat.....	21
1.4.2 Správa metadat.....	21
1.4.3 Zajištění kvality a čištění dat.....	22
1.4.4 Master Data Management.....	23
1.4.5 Data Governance.....	24
1.5 Závěry k obsahu podnikové analytiky.....	25
<b>2. Role v podnikové analytice.....</b>	<b>26</b>
2.1 Uživatelské role.....	27
2.1.1 Vlastník.....	27
2.1.2 Manažer.....	28
2.1.3 Klíčový uživatel, Power User.....	28
2.1.4 Uživatel služeb a aplikací podnikové analytiky.....	29
2.2 Role IT.....	29
2.2.1 Informační manažer (CIO).....	29
2.2.2 Sponzor projektu podnikové analytiky.....	30
2.2.3 Manažer projektu.....	30
2.2.4 Architekt podnikové analytiky.....	31
2.2.5 Byznys analytik / Konzultant.....	32
2.2.6 Datový analytik.....	32
2.2.7 Vývojář softwaru pro podnikovou analytiku.....	33
2.2.8 Dodavatel.....	33
2.3 Závěry k rolím v podnikové analytice.....	34
<b>3. Faktory ovlivňující podnikovou analytiku.....</b>	<b>35</b>
3.1 Byznys prostředí.....	37
3.1.1 Velikost firmy.....	37
3.1.2 Původ a vlastnictví firmy.....	38

3.1.3	Konkurenční prostředí.....	39
3.1.4	Odvětví působnosti firmy .....	39
3.2	Řízení a organizace firmy .....	39
3.2.1	Firemní kultura.....	40
3.2.2	Organizace firmy .....	40
3.2.3	Dislokace firmy.....	41
3.2.4	Business model.....	41
3.2.5	Agilní organizace .....	42
3.2.6	Reengineering podnikových procesů.....	42
3.2.7	Vztah firmy ke svým obchodním a dalším partnerům.....	43
3.2.8	Personální zdroje, úroveň jejich znalostí .....	43
3.2.9	Uplatnění konceptu řízení podnikové výkonnosti.....	44
3.3	Řízení podnikové analytiky .....	45
3.3.1	Řízení IT ve firmě.....	46
3.3.2	Existence silného sponzora .....	46
3.3.3	Strategie řešení podnikové analytiky .....	47
3.3.4	Identifikace potenciálních efektů podnikové analytiky .....	47
3.3.5	Kompetenční centra.....	48
3.3.6	Využití cloud computingu .....	48
3.4	Kvalita informačního systému.....	49
3.4.1	Úroveň dokumentace a optimalizace podnikových procesů.....	49
3.4.2	Kvalita datových zdrojů .....	49
3.4.3	Kvalita podnikových aplikací .....	50
3.5	Závěry k faktorům ovlivňujícím podnikovou analytiku.....	50

#### **4. Vybrané metody řešení podnikové analytiky.....52**

4.1	Dimenzionální modelování .....	53
4.1.1	Principy návrhu tabulek faktů.....	54
4.1.2	Granularita v tabulce faktů .....	55
4.1.3	Typy tabulek faktů .....	56
4.1.4	Měrné jednotky, rozsah, zdroje a kalkulace ukazatelů,.....	56
4.1.5	Tabulky dimenzí – principy návrhu.....	57
4.1.6	Dimenze - STAR .....	59
4.1.7	Dimenze - SNOWFLAKE .....	60
4.1.8	Referenční dimenze a vazby M : N.....	61
4.1.9	Degenerované dimenze.....	62
4.1.10	Dimenze parent-child .....	62
4.1.11	Klíče, umělé klíče.....	63
4.1.12	Chyby a NULL hodnoty v klíčích.....	64
4.1.13	Alternativní struktury dimenzí.....	64

4.1.14	Dimenze času .....	64
4.1.15	Dimenze různých rolí .....	65
4.1.16	Sběrná dimenze.....	65
4.1.17	Změny v dimenzích - SCD (Slowly Changing Dimensions).....	66
4.2	Datové modelování.....	70
4.2.1	Účel datového modelování .....	71
4.2.2	Relační databáze .....	71
4.2.3	Řešení datových modelů .....	73
4.2.4	Podstatné charakteristiky datových modelů.....	73
4.2.5	Normalizace databáze .....	74
4.2.6	Návrh datového modelu pro BI a SSBI .....	74
4.3	Závěry k metodám řešení podnikové analytiky.....	75
<b>B) Řešení podnikové analytiky.....</b>		<b>76</b>
<b>5. Business Intelligence, BI .....</b>		<b>77</b>
5.1	Základní principy Business Intelligence.....	77
5.1.1	Výběr a organizace dat.....	77
5.1.2	Multidimenzionalita uložení a práce s daty .....	79
5.1.3	Nároky na kvalitu dat .....	83
5.2	Efekty BI .....	84
5.3	Omezení, problémy, předpoklady BI.....	85
5.4	Závěry k řešením Business Intelligence .....	85
<b>6. Self-Service Business Intelligence, SSBI.....</b>		<b>86</b>
6.1	Podstatné charakteristiky SSBI .....	86
6.2	Efekty SSBI.....	87
6.3	Omezení, problémy, předpoklady SSBI.....	88
6.4	Závěry k řešení Self-Service Business Intelligence.....	88
<b>7. Mobilní BI .....</b>		<b>90</b>
7.1	Podstatné charakteristiky mobilního BI.....	90
7.2	Efekty mobilního BI.....	91
7.3	Omezení, problémy, předpoklady mobilního BI .....	91
7.4	Závěry k řešení mobilního Business Intelligence.....	91
<b>8. Competitive Intelligence, CI.....</b>		<b>92</b>
8.1	Podstatné charakteristiky CI .....	92

8.2	Efekty CI .....	92
8.3	Omezení, problémy, předpoklady CI .....	93
8.4	Závěry k řešení Competitive Intelligence.....	93

## **9. Postup řešení podnikové analytiky na příkladu Business Intelligence 94**

9.1	Přehled a obsah úloh řešení projektu Business Intelligence .....	94
9.1.1	Zpracování Úvodní studie pro BI .....	95
9.1.2	Specifikace přírůstku řešení.....	99
9.1.3	Analýza stavu a požadavků na BI.....	100
9.1.4	Modelování a návrh řešení .....	103
9.1.5	Návrh technologické platformy přírůstku.....	106
9.1.6	Návrh transformací dat – ETL.....	108
9.1.7	Implementace řešení .....	110
9.1.8	Zavedení BI do provozu, migrace .....	111
9.2	KPI řešení projektu Business Intelligence .....	113
9.3	Data, dokumenty.....	114
9.4	Scénáře, analytické otázky k řešení projektu Business Intelligence.....	116
9.4.1	Implementují se aplikace Business Intelligence.....	116
9.4.2	Řeší se nasazení datového skladu .....	116
9.4.3	Řeší se nasazení datového tržiště.....	116
9.4.4	Řeší se uplatnění ETL / ELT .....	117
9.4.5	Řeší se uplatnění OLAP databází.....	117
9.4.6	Řeší se uplatnění analytických aplikací .....	117
9.4.7	Řeší se uplatnění reportingu.....	117
9.4.8	Řeší se uplatnění dočasného úložiště dat, DSA.....	117
9.4.9	Řeší se uplatnění nástrojů dolování dat.....	118
9.5	Závěry k postupu řešení projektu Business Intelligence.....	118

## **C) Komponenty podnikové analytiky ..... 119**

## **10. Datové báze v podnikové analytice ..... 121**

10.1	Datový sklad, DWH.....	121
10.1.1	Efekty a přínosy datového skladu .....	122
10.1.2	Otázky a problémy datových skladů .....	122
10.1.3	Požadavky na datový sklad .....	123
10.1.4	Úrovně řešení datového skladu .....	123
10.2	Datové tržiště, DMA.....	124
10.2.1	Efekty a přínosy datového tržiště.....	124
10.2.2	Otázky, problémy a omezení spojené s datovým tržištěm.....	125



10.3 Dočasné úložiště dat, DSA .....	125
10.3.1 Efekty a přínosy DSA.....	125
10.3.2 Problémy a omezení spojené s DSA.....	125
10.4 Operační datový sklad, ODS .....	125
10.4.1 Efekty a přínosy operačního datového skladu .....	126
10.4.2 Problémy a omezení spojené s operačním datovým skladem.....	126
10.5 Data Lake .....	126
10.5.1 Efekty a přínosy Data Lake.....	128
10.5.2 Problémy a omezení spojené s Data Lake .....	128
10.6 Závěry k datovým zdrojům v podnikové analytice .....	128
<b>11. Transformace dat.....</b>	<b>129</b>
11.1 Extract Transform Load (ETL/ELT) .....	129
11.1.1 Efekty a přínosy ETL / ELT .....	130
11.1.2 Problémy a omezení spojené s ETL / ELT.....	130
11.2 Zachycení změn ve zdrojových systémech, CDC .....	130
11.2.1 Efekty a přínosy zachycení změn ve zdrojových systémech.....	131
11.2.2 Problémy zachycení změn ve zdrojových systémech.....	131
11.3 Historizace dat.....	132
11.3.1 Efekty a přínosy historizace .....	132
11.3.2 Otázky a problémy historizace.....	133
11.4 Závěry k transformacím dat v podnikové analytice.....	133
<b>12. Datová analytika .....</b>	<b>135</b>
12.1 OLAP databáze .....	135
12.1.1 Efekty a přínosy OLAP.....	135
12.1.2 Problémy a omezení spojené s OLAP databázemi .....	135
12.2 In-Memory Analytics .....	136
12.2.1 Efekty a přínosy In-Memory Analytics.....	136
12.2.2 Problémy a omezení spojené s In-Memory Analytics .....	137
12.3 Sandbox.....	137
12.3.1 Efekty a přínosy Sandbox.....	137
12.3.2 Problémy a omezení spojené se Sandbox .....	137
12.4 Závěry k analytickým nástrojům .....	137
<b>D) Reporting a dashboardy .....</b>	<b>138</b>
<b>13. Reporting.....</b>	<b>139</b>

13.1	Uživatelé reportingu.....	139
13.2	Kategorizace reportů .....	139
13.2.1	Interní reporting.....	139
13.2.2	Externí reporting .....	140
13.2.3	Další kategorie reportů.....	140
13.3	Efekty a přínosy reportingu.....	141
13.4	Problémy a omezení spojené s reportingem .....	141
13.5	Závěry k principům a možnostem reportingu.....	142
<b>14.</b>	<b>Dashboardy .....</b>	<b>143</b>
14.1	Princip dashboardů v podniku jako součást podnikové analytiky .....	143
14.2	Efekty a přínosy dashboardu .....	144
14.3	Problémy a omezení dashboardů.....	145
14.4	Závěry k realizaci dashboardů.....	145
<b>E)</b>	<b>Prediktivní analytika .....</b>	<b>147</b>
<b>15.</b>	<b>Potřeba a obsah prediktivní analytiky .....</b>	<b>149</b>
15.1	Potřeba prediktivní analytiky: prognózování a plánování .....	149
15.1.1	Prognózování v byznysu.....	149
15.1.2	Podstata a principy plánování.....	150
15.1.3	Podstata plánovacích úloh.....	151
15.1.4	Principy plánovacích úloh .....	151
15.1.5	Plánovací úlohy v kontextu řízení firmy .....	152
15.2	Podstatné principy prediktivní analytiky.....	152
15.3	Řešení prediktivní analytiky .....	154
15.4	Efekty uplatnění prediktivní analytiky.....	154
15.5	Omezení, problémy, předpoklady řešení prediktivní analytiky.....	155
15.6	Využití prediktivní analytiky.....	156
15.7	Závěry k potřebě a obsahu prediktivní analytiky.....	157
<b>16.</b>	<b>Prediktivní analytika v kontextu podnikové analytiky .....</b>	<b>158</b>
16.1	Business Intelligence, BI.....	158
16.1.1	Principy Business Intelligence .....	159
16.1.2	Prediktivní analytika a Business Intelligence .....	159
16.2	Competitive Intelligence, CI .....	159
16.2.1	Principy Competitive Intelligence .....	159

16.2.2	Prediktivní analytika a Competitive Intelligence.....	160
16.3	Data Science .....	160
16.3.1	Principy Data Science.....	160
16.3.2	Prediktivní analytika a Data Science.....	161
16.4	Data mining, DMI .....	161
16.4.1	Principy data miningu.....	161
16.4.2	Prediktivní analytika a data mining .....	163
16.5	Text mining.....	163
16.5.1	Principy text miningu.....	163
16.5.2	Prediktivní analytika a text mining.....	164
16.6	Umělá inteligence (Artificial Intelligence).....	165
16.6.1	Principy umělé inteligence .....	165
16.6.2	Prediktivní analytika a umělá inteligence .....	165
16.7	Machine Learning, ML .....	165
16.7.1	Principy machine learning.....	165
16.7.2	Prediktivní analytika a machine learning.....	166
16.8	Big Data Analytics.....	167
16.8.1	Principy Big Data Analytics .....	167
16.8.2	Prediktivní analytika a Big Data Analytics.....	168
16.9	Závěry k prediktivní analytice v kontextu podnikové analytiky.....	168
<b>17.</b>	<b>Role v prediktivní analytice .....</b>	<b>169</b>
17.1	Doménový expert.....	169
17.2	Databázový expert.....	169
17.3	Expert prediktivního modelování .....	170
17.4	Závěry k rolím v prediktivní analytice.....	170
<b>18.</b>	<b>Postupy řešení a organizace prediktivní analytiky .....</b>	<b>171</b>
18.1	Pochopení byznysu („Business Understanding“)... ..	172
18.1.1	Definování vstupních požadavků na prediktivní modely .....	172
18.1.2	Byznys cíle.....	172
18.1.3	Definování dat pro prediktivní modelování.....	173
18.1.4	Definování jednotek analýzy .....	173
18.1.5	Závěry k „Business Understanding“ .....	174
18.2	Porozumění datům („Data Understanding“)... ..	174
18.2.1	Analýza dostupných zdrojů dat.....	175
18.2.2	Analýza datové kvality .....	175
18.2.3	Závěry k „Data Understanding“ .....	175

18.3 Příprava dat („Data Preparation“)	176
18.3.1 Organizace a shromáždění dat	176
18.3.2 Čištění dat	176
18.3.3 Vytvoření odvozených proměnných	177
18.3.4 Závěry k „Data Preparation“	178
18.4 Modelování („Modeling“)	178
18.4.1 Typy prediktivních modelů	179
18.4.2 Hledání vztahů – data mining, cluster analýza	179
18.4.3 Tvorba prediktivního modelu	179
18.4.4 Soubor modelů (ensemble)	180
18.4.5 Závěry k „Modeling“	180
18.5 Vyhodnocení výsledků („Assessing Models“)	181
18.5.1 Validace a ladění prediktivních modelů	181
18.5.2 Vyhodnocení prediktivního modelu	182
18.5.3 Závěry k „Assessing Models“	183
18.6 Využití výsledků („Model Deployment“)	183
18.6.1 Prezentace výsledků	184
18.6.2 Konsolidace informací	184
18.6.3 Závěry k „Model Deployment“	185
<b>19. Závěr</b>	<b>186</b>
Přehled literatury	187
Terminologický slovník	193

# Úvod

Podniková analytika je v současnosti jednou **z nejperspektivnějších oblastí** řízení firem a uplatňování IT v jejich prostředí. Je to dáno díky jejím možnostem efektivní podpory řídicích, tedy analytických, plánovacích i rozhodovacích aktivit podnikových manažerů a specialistů. Tím aplikace podnikové analytiky významnou měrou **přispívají k celkovému zvyšování kvality podnikového řízení** a současně se tak stávají i podstatným faktorem, ovlivňujícím jejich konkurenceschopnost a jejich konkurenční výhody. To dokumentují i výsledky celosvětových průzkumů, prováděné renomovanými analytickými firmami.

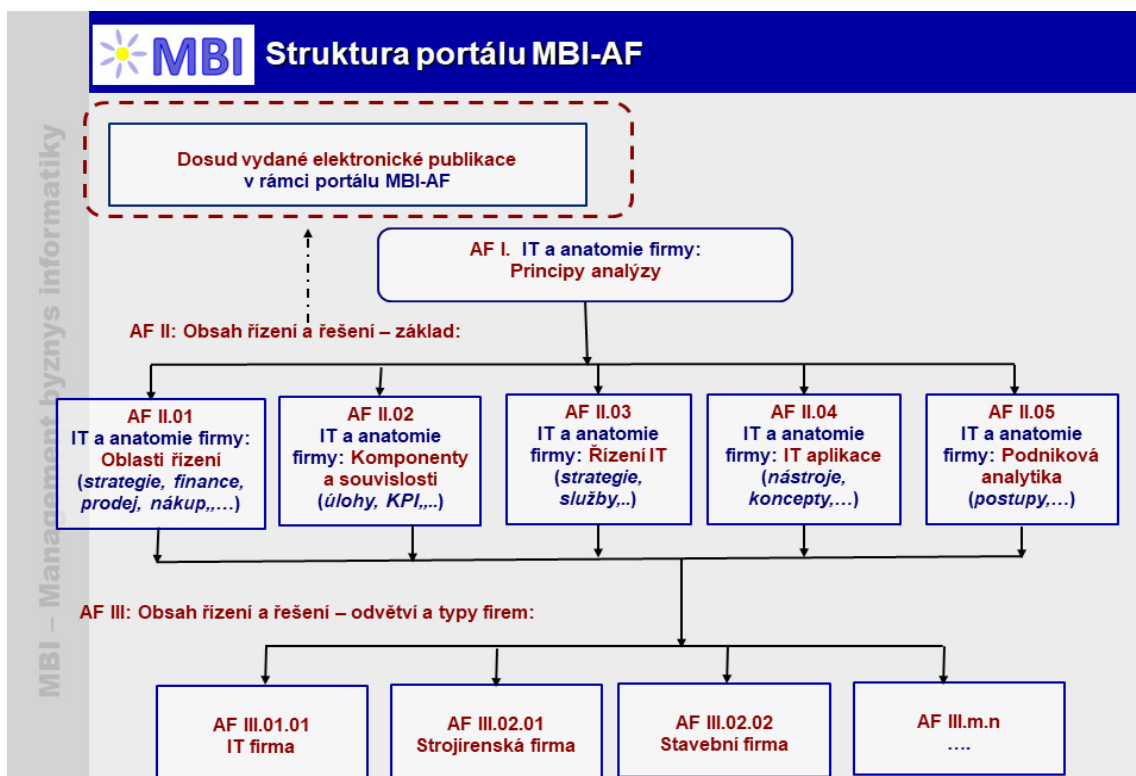
Rovněž v ČR se tato oblast informatiky **rozvíjí již od počátku 90. let** a v současné době představuje zcela běžnou a integrální součást informatiky firem a jejich úroveň, kvalita a rozsah se permanentně zvyšují. Pozornost, věnovaná podnikové analytice a **potřebná kvalita znalostí** u dodavatelů i uživatelů, je proto zcela opodstatněná.

Řešení podnikové analytiky představuje v současnosti obrovské **množství nejrůznějších organizačních, analytických, implementačních a provozních úloh**. Využívá se při něm rozsáhlé spektrum softwarových nástrojů, databázovými počínaje a nejrůznějšími speciálními prostředky, jako např. pro transformace, kontroly a čištění dat konče. Jejich standardním popisům, je věnováno poměrně hodně literatury. Tento text si proto klade za úkol **podívat se na vybavenost a prostředky podnikové analytiky v širším kontextu** a zejména pohledem manažera a analytika. To znamená vedle běžných charakteristik hodnotit, jaké přinášejí firmě **potenciální efekty** a na druhé straně **s jakými omezeními** nebo problémy při jejich realizaci a užití je třeba počítat. **Cílem tohoto textu** je tak poskytnout pokud možno **komplexní pohled** na podnikovou analytiku, a to jak z hlediska jejího **obsahu**, tak použitých **technologií, metod i postupů**.

Text publikace je **součástí portálu** na adrese <https://mbi-af.cz>, jehož základní strukturu dokumentuje Obrázek 0-1:

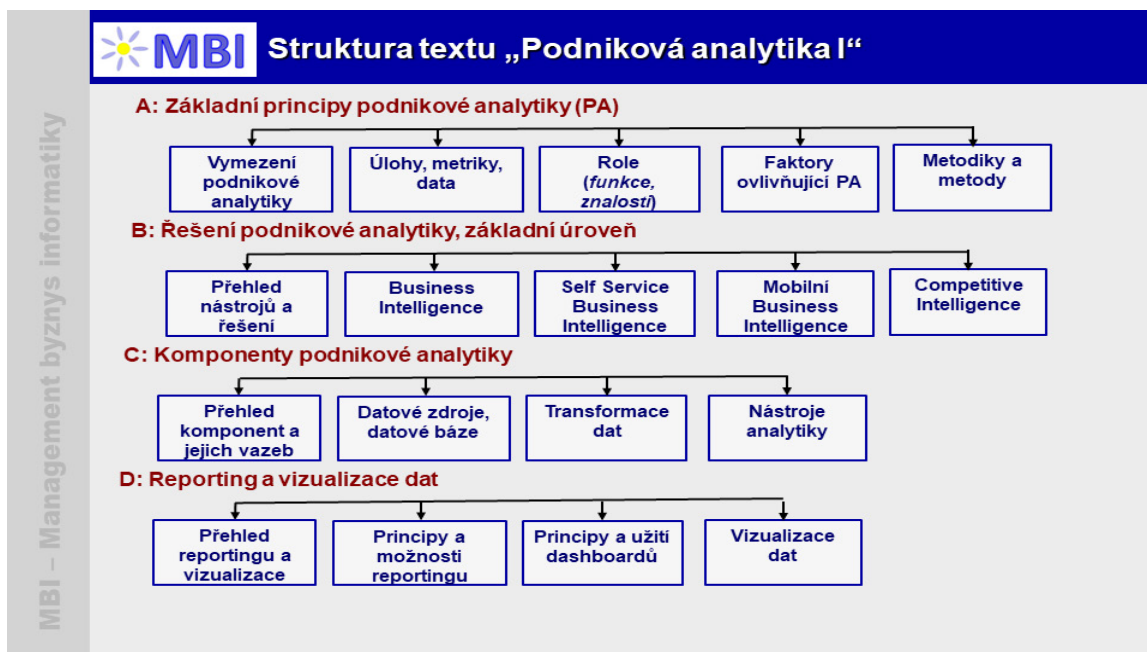
Portál je založen **na sadě dokumentů dvou typů**:

- dokumenty mající charakter **publikací** (na obrázku nahoře) a jsou tak určené pro studijní účely, jako literatura ke kurzům apod.,
- **pracovní dokumenty**, které mají i podle názvu čistě pracovní charakter, jsou průběžně doplňovány a aktualizovány a slouží primárně jako podklady pro řešení analýz a návrhů v rámci projektů, nebo jako podklady pro řešení dílčích úloh v rámci předmětů na VŠ. S ohledem na jejich určení jsou také **podstatně detailněji** účelově uspořádané, než je tomu u publikací.



Obrázek 0-1: Obsah portálu MBI-AF

Strukturu publikace demonstruje Obrázek 0-2:



Obrázek 0-2: Struktura textu „IT a anatomie firmy: Podniková analytika I“

Z pohledu řešení a odpovídajících nástrojů a přístupů je možné publikaci **pracovně členit** do následujících **oblastí**, jak ukazuje předchozí schéma:

- Charakteristika principů a **obsahové náplně** řešení podnikové analytiky, zejména úloh, metrik, datových zdrojů, rolí, faktorů a metod (oddíl A).
- Podniková analytika **na základní úrovni** zahrnuje řešení a nástroje typu business intelligence self service business intelligence, mobile business intelligence a competitive intelligence. Kromě stručného vymezení zahrnuje i specifikaci významných potenciálních efektů a na druhé straně i případné problémy a omezení z pohledu řešení i užití (oddíl B).
- V rámci podnikové analytiky se využívá velmi široká **škála nástrojů, konceptů a technologií**. Obdobně, jako v předchozím případě, text zahrnuje stručné vymezení funkcí a současně i podstatné poskytované efekty i významná omezení jejich implementace a provozu (oddíl C).
- Podstatnou součástí podnikové analytiky je **firemní reporting** a s ním spojená **vizualizace dat**, tj. obsahové, formální i organizační přístupy při definování, vytváření a poskytování reportů manažerům a specialistům firmy pro jejich řídicí aktivity (oddíl D).
- Závěrečnou část publikace představují principy a **možnosti prediktivní analytiky** jakožto jedné z nejintenzivněji využívaných disciplín pokročilé analytiky v reálné praxi (oddíl E).

**Poznámka:** Obsahem jde o texty vycházející většinou ze stránek obdobného portálu MBI, který je na adrese <https://mbi.vse.cz>, ale ten je **momentálně z bezpečnostních důvodů přístupný** uživatelům, využívajícím **pouze doménu vse.cz**. V obrázcích a dokumentech se proto můžeme setkat s logem portálu MBI a s kódy objektů, které se na MBI používají.

**Ještě terminologická poznámka:** V textu jsou použity termíny „**podnik**“ i „**firma**“ ve stejném nebo obdobném smyslu. Termín „**firma**“ pokládáme za základní, ale v mnohém kontextu je využití termínu „**podnik**“, „**podnikový**“ apod. přirozenější. Využíváme je tak podle obvyklých použití v praxi.

## A) Základní principy podnikové analytiky

**Vstupní oddíl** do podnikové analytiky pokrývá zejména její **obsahovou stránku**, tedy co je nebo by mělo být prvotním zájmem analytika při řešení projektů. Patří sem tak všechny **aspekty tvořící logiku podnikové analytiky a současně základnu** pro využití nejrůznějších **technologií a nástrojů**, které jsou předmětem i dalších navazujících oddílů.

Oddíl zahrnuje kapitoly, jejichž náplň stručně objasní **následující přehled**:

- Východiskem je celkové **vymezení a kategorizace** podnikové analytiky, tj specifikace hlavních témat řešení.
- Řešení podnikové analytiky se vztahuje k jednotlivým **úlohám nebo procesům řízení firmy**, přičemž evidenční a transakční úlohy představují z pohledu analytiky pouze přípravu dat a na tomto místě se jimi přímo nezabýváme. **Kapitola 1.2** představuje přehled těchto úloh, jejichž funkcionalita byla náplní již předcházející publikace „IT a anatomie firmy (Oblasti a komponenty řízení)“.
- Další klíčovou komponentou řešení podnikové analytiky je celý **systém metrik** a příslušných **analytických dimenzí**. Rozsah jejich dokumentace je mimořádně velký, a proto na tomto místě (**kapitola 1.3**) se pouze rekapitulují jejich základní **principy**.
- V souvislosti s metrikami i dimenzemi je podstatnou otázkou, kde se pro ně získají data, tedy **datové zdroje (kapitola 1.4)**.
- Pro řešení projektů musí být definovány příslušné **role pracovníků**, a to jak na řešitelské, tak uživatelské straně. **Kapitola 2** obsahuje vymezení vybraných rolí, specifických pro potřeby podnikové analytiky s určením jejich **funkční náplně**.
- Obsahem **kapitoly 3** je charakteristika **faktorů**, majících obvykle **podstatný vliv** na průběh i úspěšnost řešení analytických projektů. Představují zde poměrně široké spektrum těchto faktorů, rozdělených do skupin podle jejich povahy.
- Poslední kapitola oddílu, **kapitola 4**, se zaměřuje na poněkud detailnější prezentaci **dvou vybraných metod**, které se k řešení podnikové analytiky obvykle vztahují, a to **multidimenzionálního a datového modelování**. Metodik a metod je ve spojení s podnikovou analytikou celá řada, prezentovány jsou na různých analytických i firemních portálech.



# 1. Obsah podnikové analytiky



Obsah podnikové analytiky, tedy ekonomickou nebo obchodní náplň řízení firmy, lze **vyjádřit úlohami**, které jsou předmětem řešení, **metrikami, adekvátními dimenzemi a daty** a datovými zdroji pro podnikovou analytiku.

**Účelem** kapitoly je:

- **vymezit** podnikovou analytiku, **kategorizovat součásti** jejího obsahu **a úrovně** řešení,
- **určit charakteristiky vybraných typů úloh** ve vztahu k podnikové analytice,
- **určit** podstatné **charakteristiky metrik**, které jsou předmětem řešení úloh podnikové analytiky a souvisejících **analytických dimenzí**,
- **určit** podstatné **charakteristiky datových zdrojů** pro podnikovou analytiku a jejich využití.

## 1.1 Vymezení a kategorizace podnikové analytiky

Podniková analytika v současnosti tvoří integrální součást podnikového řízení a s tím i součást IT i anatomy firmy. Zatímco termínem **analýza** („*analysis*“) se obvykle označují principy a aktivity spojené s řešením informačních systémů firem, pak termín **analytika** („*analytics*“) se využívá pro vyhodnocování a prezentování firemních aktivit a jejich výsledků v podobě celé škály jejich ukazatelů a souvisejících analytických dimenzí. Zahrnují převážně analytické nebo plánovací činnosti. **Podniková analytika** pak představuje svým způsobem zastřešující pojem, případně i synonymum pro různé specifické, či dílčí disciplíny a pojetí jako je business analytika, datová analytika a další a s nimi spojené dnes již obvyklé nástroje a technologie, jako je business intelligence, competitive intelligence a další.

Pokud bychom měli rozdělit, resp. **kategorizovat** podnikovou analytiku z pohledu jejích aplikací a využívaných metod, pak je pracovním způsobem můžeme **členit na dvě skupiny**:

- **základní**, která má v praxi zřejmě největší rozšíření s aplikacemi a nástroji, majícími již většinou delší historii; sem patří zejména business intelligence (BI), self service business intelligence (SSBI), competitive intelligence (CI) a technologická varianta předchozích možností, tedy mobilní business intelligence apod. další),
- **pokročilá**, která zejména v poslední době nabírá rychle na významu a praktickém rozšíření a která je většinou založena na velmi sofistikovaných metodách, modelech i algoritmech, vycházejících z matematiky a statistiky. Sem zřejmě patří zejména data mining, process mining, text mining, prediktivní analytika, data science, umělá inteligence, strojové učení a další. Je nutné současně přiznat, že tyto disciplíny se často vzájemně překrývají nebo doplňují a je obtížné mezi nimi hledat ostré hranice.

Z pohledu řešení a uplatnění podnikové analytiky je dobré upozornit na její následující **výrazné aspekty**:

- Aplikace a řešení podnikové analytiky jsou převážně **orientovány na potřeby pracovníků vyšších úrovní řízení** firmy, tedy manažery, podnikové specialisty, vlastníky, členy správních rad apod. Tomu musí **odpovídat i způsob a forma poskytovaných informací**. Na druhé straně s rozvojem nových technologií se uplatnění analytiky postupně dostává i na nižší úrovně firemního řízení.
- **Na rozdíl od transakčních úloh a aplikací** je při řešení úloh analytiky nutné **respektovat tyto momenty**:
  - I když i v analytické oblasti jsou v nabídce trhu určité standardní (typové) aplikace, přesto převažující smysl poskytovaných analytických aplikací je **v jejich individualizaci** podle potřeb konkrétních manažerů, specialistů a dalších.
  - Právě individualizace řešení je i základem jejich obvykle skloňovaného efektu, tedy hlavního **přínosu pro konkurenceschopnost firmy** a získávání konkurenčních výhod. Standardní, typové aplikace příliš konkurenčních výhod nepřinášejí, naopak čím „chytřejší“ budou individuální aplikace podnikové analytiky (analytické i plánovací), tím vyšší je potenciál budoucích výhod.
  - Podstatný rozdíl transakčních aplikací a aplikací podnikové analytiky spočívá v tom, že zatímco **transakční aplikace obvykle datové zdroje** vytvářejí (např. na základě vstupních dokladů), pak **podniková analytika tyto zdroje využívá** pro realizaci analýz, plánů, prognóz. To současně vytváří na tyto zdroje i specifické nároky, např. v oblasti jejich dostupnosti, kvality dat apod.
- **Při praktickém užití aplikací podnikové analytiky** manažeři a specialisté firmy vyhodnocují podnikové ukazatele podle různých dimenzí, v delším časovém vývoji, s predikcemi na delší časové horizonty, s identifikací problémů, kterou jsou jinak obtížně identifikovatelné. To vše znamená, že uplatnění podnikové analytiky vede přirozenou cestou ke **zvyšování manažerské kvalifikace**.

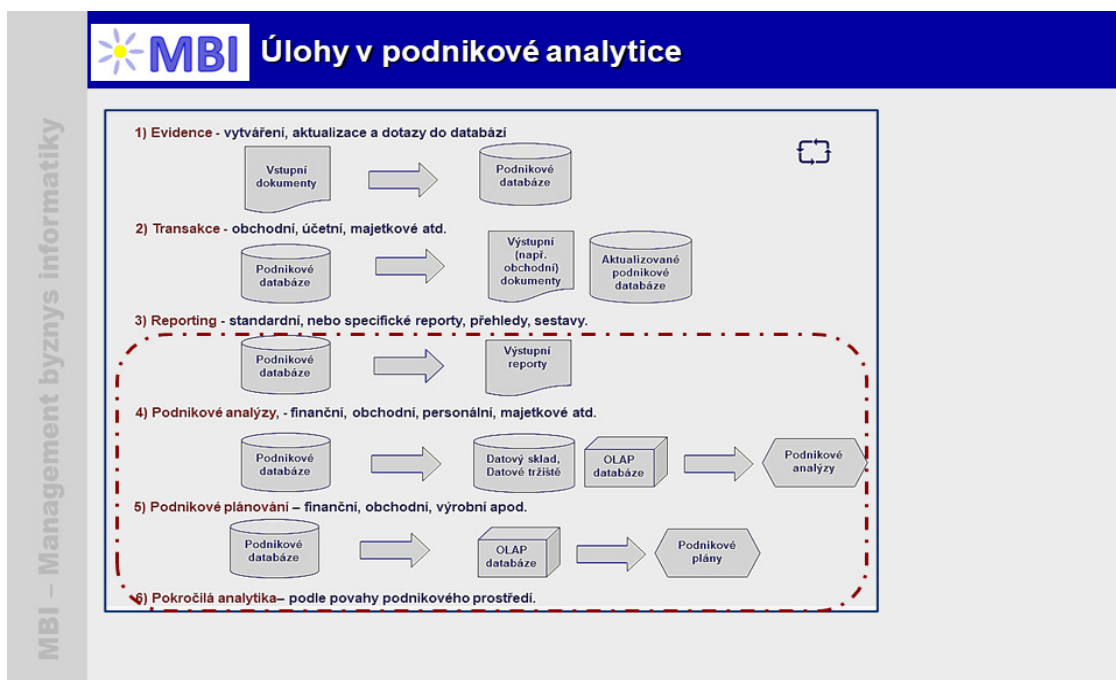
## 1.2 Úlohy, procesy

Úloha je v anatomii firmy **základní komponentou**, a tedy i v podnikové analytice. Úlohy mají zachytit a dokumentovat řízení firmy, tedy co se realizuje, řídí a řeší. **Termín „úloha“** se jeví jako **obecnější**, a to proto, že její součástí **může nebo nemusí být „proces“**, nebo naopak úloha může být součástí nějakého komplexního procesu. „Úlohy“ mohou v podnikové analytice primárně zahrnovat přehled tzv. **„klíčových aktivit“**, tj. činností, které se nejvýraznějším způsobem podílejí na finálním výsledku úlohy. Z pohledu analytických nebo plánovacích úloh to mohou být aktivity spojené s analýzou nebo plánováním vybraných skupin metrik, se sledováním časového vývoje jejich hodnot, porovnáním plánovaných a skutečných hodnot, plánováním hodnot podle určitých časových horizontů apod.

Počet úloh je v řízení firmy obvykle značný, a proto je dobré je nějakým způsobem **členit, resp. kategorizovat**. V našem případě využíváme hlediska dvě:

- úlohy rozdělené **podle oblastí řízení**, např. finance, prodej atd.,
- úlohy rozdělené podle **standardních operací s daty**, tj. evidence, transakce, reporting, analýzy, plánování, pokročilá analytika, případně specifické úlohy.

**Kategorizace úloh z obou pohledů** je využita v publikaci „**IT a anatomie firmy (Oblasti a komponenty řízení)**“. Na tomto místě pouze pro rekapitulaci uvádíme přehled typů úloh podle operací s daty se zvýrazněním těch, které se přímo váží k podnikové analytice (Obrázek 1-1).



**Obrázek 1-1: Typy úloh v podnikové analytice**

Typy úloh v rámci této pracovní klasifikace zahrnují:

- **Evidenční** úlohy zahrnují především pořizování dat a jejich kontroly.
- **Transakční** úlohy zajišťují provádění obchodních, finančních a dalších transakcí, tj. např. realizace obchodního případu „Prodej“, „Nákup“ apod.
- **Reportingové** úlohy obsahují návrh, zpracování a distribuci reportů.
- **Analytické** úlohy představují analýzy finančních, obchodních a dalších aktivit firmy.
- **Plánovací** úlohy zajišťují plánování na bázi podstatných plánovacích ukazatelů a jejich dimenzí.
- **Úlohy pokročilé analytiky** znamenají realizaci pokročilých analytických funkcí, jako je např. prediktivní analytika atd.

**Evidenční a transakční úlohy** představují z pohledu podnikové analytiky zejména vytváření a aktualizace datových zdrojů jako vstupů do jejích úloh. Další typy úloh, tj.

**reportingové, analytické, plánovací analytiky** se k podnikové analytice již bezprostředně váží. Jejich obsah byl rovněž vymezen v publikaci „**IT a anatomie firmy (Oblasti a komponenty řízení)**“, proto se zde na ni pouze odvoláváme. Úlohy pokročilé analytiky budou obsahem jedné z následujících publikací.

### 1.3 Metriky, dimenze

Předmětem řešení v úlohách podnikové analytiky jsou zejména metriky reprezentované **ukazateli, odpovídajícími dimenzemi a datovými zdroji**. Každá metrika je **dokumentována** standardně v těchto bodech:

- obsahové vymezení metriky, resp. ukazatele, její využití i s případným kalkulačním předpisem,
- související metriky, odvozené od základního vymezení, kde jde o pracovní vytváření jejich skupin s ohledem na jejich velký počet, a tedy i zajištění větší přehlednosti celého textu,
- přehled analytických dimenzí, které se v souvislosti s metrikou obvykle využívají,
- datové zdroje pro metriku, obvykle databáze nebo jejich části či speciální dokumenty a výkazy.

**Z hlediska obsahu metrik** je účelné v jednotlivých oblastech řízení sledovat:

- metriky **finančního, resp. ekonomického** charakteru (objem nákladů, výnosů, komplexní finanční metriky jako cash flow, úroveň zadluženosti firmy apod., objem majetku, objem odpisů, počet a objem reklamací apod.),
- metriky **výkonového nebo procesního** charakteru (počet zpracovávaných daňových dokladů, počet prodejních nebo nákupních transakcí, počet vyřízených reklamací, počet realizovaných školení apod.),
- metriky **organizačního charakteru** (vytížení kapacit, počet pracovníků, útvarů, počet zákazníků, dodavatelů apod.),
- metriky **kvalitativního charakteru**, jako jsou sledovaná úroveň kvality produktů a služeb, dodací lhůty, kvality dodávaných materiálů atd.

**Z hlediska dopadů** metrik do systému řízení existují jejich 3 typy – **KPI, KRI a KGI**.

Pro správné určení **KPI ukazatele** je vhodné vypsát **seznam kritických faktorů úspěchů** podniku, který úzce souvisí s KPI. **KPI (Key Performance Indicator)**, klíčový ukazatel výkonnosti splňuje zpravidla tyto body:

- je měřitelný velice často, zpravidla denně,
- musí mu rozumět všichni zaměstnanci,
- má významný vliv na organizaci,
- není finanční povahy,
- má zásadní vliv na výkonnost podniku,
- ukazuje, jak dobře fungují klíčové procesy podniku,
- v podniku existuje maximálně 10 KPI ukazatelů.

U každého **KPI** ukazatele je především podstatný **potřebný zdroj dat, granularita, jeho výpočet a odchylka**:

- **Zdroje dat** identifikují, kde se nachází potřebné **informace pro daný typ KPI** ukazatele. Samotný způsob výpočtu KPI ukazatele se může počítat z více datových zdrojů, proto je důležité hodnotit analytické nástroje podle nezávislosti a existence velkého množství konektorů na různé typy datových zdrojů.
- **Granularita**, která **určuje detail výpočtu pro KPI**. Jednoduše si to lze představit u zjišťování KPI za celý stát, nebo i města, obce nebo ve vazbě na čas, kdy KPI může mít souhrnnou hodnotu za celý rok, měsíc, týden či den.
- **Výpočet KPI** je většinou vyjádřen **v procentech**, nebo jako součet či průměr, mluvíme-li o **jednom datovém zdroji**. Je-li zdrojem pro výpočet KPI potřeba **více datových zdrojů**, je nutné počítat **s podmíněnými výpočty** a v některých případech i s různou kvalitou dat.
- **Výpočet změny oproti určitému časovému období**, například porovnání tržeb ve stejné pobočce oproti minulým obdobím, porovnání zisků za různá období, plnění plánu výroby oproti minulému čtvrtletí atd.

**KRI (Key Result Indicator)** je **výstup aktivit** podniku. Sděluje managementu, jak je podnik úspěšný, ale už nelze zjistit směr, jakým způsobem dále zlepšit výsledek.

**KGI (Key Goal Indicator)** ukazuje, jakým způsobem se firmě daří dosahovat výsledků a cílů a je provázán se strategií podniku.

## 1.4 Data a datové zdroje

Pro analýzy a užití metrik jsou podstatné charakteristiky odpovídajících datových zdrojů. Každý datový zdroj je **dokumentován** jejich obsahem, účelem a alespoň vybranými nebo hlavními atributy. Ve vztahu k řešení podnikové analytiky je ale s daty a datovými zdroji spojena řada témat a metod, z nichž se zde ve stručnosti zaměříme pouze na následující:

- řízení a správa metadat,
- zajištění kvality a čištění dat,
- MDM, Master Data Management,
- Data Governance.

### 1.4.1 Řízení a správa metadat

### 1.4.2 Správa metadat

Metadata jsou ve své podstatě **strukturovaná data o datech**. Metadata představují údaje **nejen o samotných datech**, ale také o technických prostředcích, softwaru, nebo sítích, kde se data nacházejí. Specifikují jejich **kontext, obsah, předpokládanou interpretaci a dostupnost dat**. Hlavní **účelem** metadat je **poskytování informací k analýze**, návrhu, vývoji, implementaci a užití jednotlivých aplikací i celé podnikové analytiky.

V souvislosti s řešeními projektů je **obsah metadat** např. následující:

- celkový **popis zdrojových systémů**,
- u data staging area to jsou popisy dat **ve slovníku datového skladu**,
- u datového skladu **popisy transformačních pravidel** pro každou tabulku a každý datový element a popisy business názvů a transformačních pravidel pro každou tabulku a každý datový element,
- **pro reporting** vysvětlení každého pole na reportu.

**Podstata metadat** je tedy zřejmá, jejich uplatnění jako faktor v řešeních podnikové analytiky je dáno několika důvody:

- řešení podnikové analytiky **se vztahují převážně na celý podnik**, jsou proto velmi komplexní, rozsáhlá a komplikovaná. Uspořádané, jasně strukturované a dostupné informace o tom, co tato řešení obsahují, jaké datové struktury, v jakých vazbách apod., jsou **při této složitosti nutnou podmínkou realizace** projektů;
- s rozsahem řešení analytiky **roste i rozsah jejich metadat**. Pro efektivní zajištění projektů i provozu aplikací se využívají **celé systémy pro správu metadat**, tedy databáze metadat s příslušnou aplikační nadstavbou pro práci s nimi;
- jako jeden z efektů podnikové analytiky, vedle své analytické a plánovací funkcionality, se běžně zdůrazňuje i jejich **úloha ve zvyšování pořádku** (např. čistoty dat) v celém informačním systému. K uskutečnění této úlohy je nutné **disponovat dokonalým přehledem a evidencí** o stávajících datových a dalších zdrojích podnikové informatiky. Takovou evidenci nabízejí metadata.

Proto se otázkám metadat věnují ve velkých společnostech určení specialisté a na jejich využití je postavena naprostá většina projektů.

### 1.4.3 Zajištění kvality a čištění dat

Řešení kvality dat představuje celý **komplex činností, procesů a metod**, které se v rámci systému **řízení kvality** dat realizují. Systém řízení kvality dat (**DQMS, Data Management Quality System**) je obvykle vymezen jako „celkový úhrn aktivit organizace, sloužících k zajištění kvality dat“ a **zahrnuje** tyto hlavní funkce:

- opravu chyb a jiných nedostatků v datech hlášených organizaci zákazníkem a jinými subjekty,
- periodické čištění datových zdrojů a databází,
- čištění aktuálně používaných dat,
- prevenci chyb.

Kvalitní nastavení řízení kvality dat přináší **nejrůznější ekonomické efekty**, např. v úsporách ze ztrát zákazníků vzhledem k nekvalitně poskytovaným informacím, ve **zvýšení důvěryhodnosti firmy** vzhledem k zákazníkům. Podstatným momentem ve využití řízení datové kvality je **snížení časových ztrát při dohledávání chyb** v rámci podnikového reportingu.

Na druhé straně realizace kontrolních a opravných operací na datech představují často **značnou časovou náročnost (až 80 % celkového času řešení) i zvýšené náklady**. S tím je nutné počítat, ale pro úspěch podnikové analytiky jsou právě tyto aktivity klíčovým předpokladem.

#### 1.4.4 Master Data Management

**Master data** jsou vysoce ceněná a klíčová data, používaná **k podpoře kritických podnikových procesů** napříč celým podnikem. Tato data jsou důležitá v každé obchodní transakci, aplikaci, reportu a rozhodovacím procesu.

**Master Data Management** představuje **sadu procesů, přístupů, metodologií, nástrojů, technologií** a řešení, které pomáhají organizacím jednoznačně identifikovat a integrovat celopodnikově sdílená klíčová data. Jako příklad mohou sloužit data zákaznická, data o produktech, službách či lokalitách a další. Obdobné řešení a přístup představuje **Reference Data Management (RDM)** – vztahující se k určitým organizačním jednotkám.

MDM zajišťuje i **správu významu – sémantiky** – klíčových dat a také správu vztahů mezi daty klíčovými a ostatními.

V rámci MDM se rozlišuje **provozní a analytické MDM**:

- **provozní MDM** se soustřeďuje na zajištění „jednotného pohledu“ na master data v hlavních transakčních systémech, tj. na místě, kde jsou master data poprvé vytvářena,
- **analytické MDM** se orientuje na zajištění „jednotného pohledu“ na master data v datovém skladu.

**Podstatnými přínosy** MDM jsou např.:

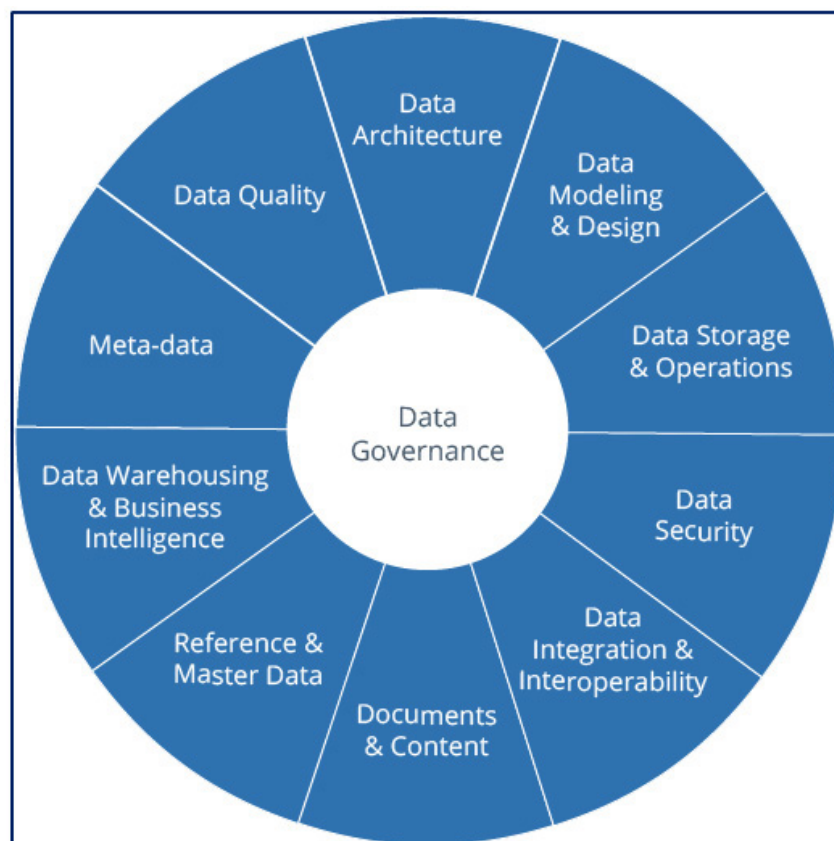
- výrazné **zvýšení kvality klíčových dat, jejich konzistence**, dostupnosti a důvěryhodnosti. Konzistence dat podporovaná MDM snižuje nedůvěru k datům a zefektivňuje rozhodovací procesy,
- **poskytování komplexních informací o zákazníkovi**, tj. úložiště master dat pro zákaznická data poskytuje jediný zdroj pro konsolidaci všech zákaznických aktivit v konzistentní formě a využívá se v transakčních i analytických aplikacích,
- **zkvalitnění zákaznických služeb** – díky dostupnosti informací a schopnosti reagovat na poptávky, objednávky, dotazy atd.,
- **snížení nákladů na podnikovou analytiku** – snižuje se objem replikací stejných dat, a tedy činností a nákladů s nimi spojených (licenční náklady, náklady na infrastrukturu apod.).

**Předpokladem úspěchu** jsou však efektivní **system motivace** mezi pracovníky uživatelské sféry i IT, jasně deklarovaná **pravidla** pro hodnocení kvality dat a zejména **efektivní začlenění MDM** do řízení datových zdrojů v rámci řízení IT, a to se všemi podstatnými souvislostmi, včetně ekonomických.

### 1.4.5 Data Governance

Společnost **Profisee** uvádí k Data governance následující **definici**: „Data governance je sadou principů a pokynů, zaručujících vysokou kvalitu dat po celou dobu jejich životního cyklu.“. Data governance je tedy **programem zaručujícím patřičnou kvalitu dat** v celé organizaci. Aby došlo ke skutečnému posunu při práci s daty a jejich **kvalitou či bezpečností**, je důležité, aby jednotliví zaměstnanci dodržovali jasně nastavená pravidla při jakékoliv práci s daty (Profisee, 2020).

Kromě kvality dat postihuje data governance ještě **další oblasti, spojené s daty**. Základní oblastí je **bezpečnost**. Další oblasti jsou zaměřeny na **architekturu, integraci** nebo například na **datové sklady**.



**Obrázek 1-2: Oblasti postihované data governance (Zdroj: JANOSCHEK, 2019)**

Na obrázku (Obrázek 1.2) je uvedeno, kolik oblastí data governance pokrývá. Několik alespoň **základních bodů spojených a data governance** je v následujícím přehledu:

- určení **jednoznačných pravidel** týkající se metadat, bezpečnosti dat, ale také požadované kvality a přístupu k datům,
- určení **majitelů a osob zodpovědných za data**. Data jsou využívána na různých úrovních a každé oddělení, užívající nebo produkující data, by se mělo podílet na jejich správnosti a mělo by nést určitou zodpovědnost za tato data. Nemělo by se tedy stát, že majitelem dat bude pouze jedno oddělení,



- **specifikace kontroly dat**, určení časového úseku, kdy se budou data testovat, zda data odpovídají požadované kvalitě, případně jak budou probíhat změny v datech a zápisy o těchto změnách,
- **realizace datových auditů** a zjišťování, na jaké úrovni se data nacházejí a jak je dodržována nastavená politika.

## 1.5 Závěry k obsahu podnikové analytiky



- Podniková analytika představuje **komplex zejména analytických a plánovacích úloh** v řízení firmy a lze ji pracovně **rozlišit na základní** (business intelligence a další) **a pokročilou** (data mining, prediktivní analytiku apod.).
- **Řešení úloh a aplikací** podnikové analytiky **se podstatně liší od transakčních aplikací** způsobem **práce s daty, použitými metodami, účastí uživatelské sféry** na řešení.
- Kvalitně připravené aplikace podnikové analytiky přispívají nejen ke **zvyšování kvality řízení a konkurenceschopnosti firmy**, ale svojí povahou často znamenají i **přínos v pozitivních posunech manažerské kvalifikace**.
- **Znalost obsahu řízení** firmy je pro řešení úloh podnikové analytiky **klíčová**, a to ve všech jeho uvedených součástech (úlohy, metriky, dimenze, data).
- **Základem řešení** úloh podnikové analytiky je **návrh a realizace analytických a plánovacích funkcí nad systémem metrik** firmy, které vyjadřují jednotlivé sledované ukazatele a jim odpovídající analytické dimenze.
- **Nejvýznamnější metriky** pro řízení firmy se charakterizují jako **KPI (Key Performance Indicator)** a jsou hlavní náplní manažerských reportů a dashboardů.
- S metrikami souvisí i **komplex otázek a problémů datových zdrojů**, z nichž jsou tyto metriky čerpány.
- Klíčovou otázkou datových zdrojů je **zajištění jejich kvality**, pro které je k dispozici celá řada metod a nástrojů.
- **Obsah řízení a podnikové analytiky** je samozřejmě **velmi variantní** podle odvětví, typu a zaměření firem a jeho zvládnutí je pro analytiku hodně náročné. Přesto je účelné disponovat potřebnou orientací v tomto směru a tu konkretizovat podle podmínek dané firmy.

## 2. Role v podnikové analytice



**Účelem** kapitoly je:

- vymezit **funkční náplň** pracovníků v oblasti podnikové analytiky, zejména manažerů, analytiků, vývojářů,
- nabídnout čtenáři alespoň **základ rolí v podnikové analytice pro jejich další konkretizaci** podle podmínek vlastní praxe.

Role určují, jaké **funkce a kompetence** bude mít pracovník v řešení problémů v určité specifické oblasti podnikové analytiky. **Smyslem rolí** je zde přesněji vyjádřit, jak jsou, nebo mají být, úlohy v řízení a řešení podnikové analytiky personálně zajištěny, resp. jaká je personální náročnost na realizaci jednotlivých úloh. Vedle toho je dobře dokumentovaný systém rolí solidním **základem pro přípravu a realizaci kvalifikačních programů a školení** ve vazbě na IT projekty a v širším kontextu i ve vazbě na řízení celé firmy.

Do kapitoly jsou vzhledem k řešení podnikové analytiky zařazeny tyto **role**:

- **Uživatelské role:**
  - **Vlastník**, resp. člen správní rady nebo dozorčí rady, většinou rozhoduje o investicích do IT včetně podnikové analytiky.
  - **Manažer** představuje v našem případě společnou roli pro různé typy manažerů podle jejich působnosti.
  - **Klíčový uživatel** podnikové analytiky („Power User“) definuje požadavky na obsah a funkcionalitu úloh podnikové analytiky a současně s tím se podílí na analytické přípravě řešených úloh v rámci projektů.
  - **Uživatel služeb a aplikací** podnikové analytiky využívá disponibilní analytické a plánovací aplikace, kvalifikovaně interpretuje dodávané výsledky v rámci řízení dané oblasti.
- **IT role:**
  - **Informační manažer** zodpovídá za začlenění aplikací podnikové analytiky do celého aplikačního portfolia a za jejich efektivní provoz.
  - **Sponzor projektu**, pracovník, odpovídající za celý projekt a jeho výsledky. Jde většinou o osobnost, která je schopna řešení a uplatnění analytiky ve firmě prosadit.
  - **Manažer projektu** řídí projekt a projektový tým v celém průběhu jeho řešení, zajišťuje komunikaci a kooperaci jak s vedením firmy, tak s externími partnery a dodavateli.
  - **Architekt podnikové analytiky** řeší návrhy IT infrastruktury a technologickou realizaci aplikací, vytváří potřebnou aplikační a technologickou architekturu podnikové analytiky.
  - **Byznys analytik** zajišťuje řešení všech aspektů obsahové náplně úloh v rámci projektu, komunikuje a kooperuje s vedením projektu a ostatními členy analytických týmů.

- **Datový analytik** zajišťuje mapování dat ve zdrojových databázích a jejich transformace do analytických databází, podílí se na řešení potřebné kvality dat a jejich konsolidaci.
- **Vývojář softwaru** pro podnikovou analytiku vyhodnocuje požadavky na analytické a plánovací aplikace, na jejich základě je navrhuje a zajišťuje jejich technologickou realizaci a dokumentaci.
- **Dodavatel** v případě řešení externími kapacitami zajišťuje veškeré činnosti řešení projektu a následnou podporu podle uzavřené smlouvy.

Vedle již zmíněného účelu rolí vzhledem k řízení firmy a k řešení podnikové analytiky je dobré uvést i **podstatný předpoklad** jejich úspěšného užití, a tím je jejich **průběžná aktualizace**. Celková funkcionalita, architektura i použité technologie podnikové analytiky **se s vývojem relativně často mění**. Proto je účelné upravovat i náplň jednotlivých rolí, jejich kompetence a nároky na kvalifikační přípravu. Úspěšnost aplikací podnikové analytiky je podstatně více, než je tomu u jiných typů aplikací, **závislá na kvalitě, motivaci a znalostech lidí**, kteří je realizují a následně využívají. Proto je jejich příprava jedním z klíčových faktorů úspěchu a systém dobře dokumentovaných rolí může v tomto směru dobře napomáhat. V dalších podkapitolách jsou uvedeny **podstatné funkční charakteristiky vybraných rolí** podnikové analytiky:

## 2.1 Uživatelské role

Rolí, které se podílejí na přípravě a užití podnikové analytiky, je celá řada a s ohledem na již výše zmíněnou individualizaci řešení je jejich vymezení nezbytné věnovat intenzivní pozornost. Bylo by samozřejmě možné podrobněji rozlišovat i role na uživatelské straně, ale s ohledem na rozsah textu se omezujeme na **charakteristiky vlastníka, manažera, uživatele a klíčového uživatele**, které je už možné promítat a upravovat podle jednotlivých konkrétních uživatelských rolí.

### 2.1.1 Vlastník

Vlastník firmy, resp. člen některého z řídicích orgánů, jehož **vztah k řešení podnikové analytiky se výrazně liší** podle toho, zda jde o velkou nebo malou firmu, zda jde o českou firmu nebo součást některé nadnárodní korporace; zda kultura firmy, kterou vlastní a ovlivňuje, je nakloněná k řešení a užívání podnikové analytiky apod. Obvykle zajišťuje tyto **funkce**:

- posuzování, případně **formulace celkové koncepce rozvoje** firmy včetně uplatňování možností podnikové analytiky,
- monitorování a **hodnocení výkonnosti podniku a jeho managementu** i s využitím aplikací podnikové analytiky,
- posuzování a **schvalování investic** do podnikové analytiky a posuzování efektivnosti těchto investic,
- výběr nebo **schvalování řídicích pracovníků** ve vazbě na aktivity, spojené s rozvojem podnikové analytiky,

- využívání specifických **aplikací pro prognózování a plánování** rozvoje firmy.

### 2.1.2 Manažer

Rolí manažera jsou zde chápány **zobecněné charakteristiky pro různé typy a úrovně manažerů**, jako např. generálního manažera (CEO), finančního manažera (CFO), marketingového manažera (CMO) a dalších. **Jejich funkce** vzhledem k podnikové analytice jsou obvykle tyto:

- určování **cílů, strategie a politiky v rozvoji podnikové analytiky** vzhledem k aktuálním i budoucím potřebám firmy,
- podíl na **plánování, řízení a koordinaci jednotlivých aktivit**, spojených s podnikovou analytikou,
- monitorování a **hodnocení výkonu pracovníků** podle jednotlivých oblastí řízení v souvislosti s účastí na projektech podnikové analytiky,
- **schvalování rozpočtů, kontrolování výdajů** a zajišťování účelného využívání finančních zdrojů, určených pro podnikovou analytiku,
- **řízení spolupráce mezi útvary** firmy i s externími partnery na rozvoji a užití podnikové analytiky,
- **návrh případných organizačních změn** ve firmě v souvislosti s podnikovou analytikou, např. vznik kompetenčních center apod.,
- řešení **transformace byznys modelů a byznys procesů** v souvislosti s výsledky aplikací podnikové analytiky.

### 2.1.3 Klíčový uživatel, Power User

Klíčový uživatel určuje **základní nároky a požadavky na realizaci** aplikací podnikové analytiky ve firmě. Podílí se na analytických činnostech v oblasti úloh podnikové analytiky. Podílí se podle jednotlivých fází a úloh řešení projektů na těchto **činnostech**:

- zajišťuje **hodnocení funkcionality zdrojových aplikací** a IT služeb vzhledem k vývoji analytických a plánovacích potřeb firmy, požadavků partnerů i vývoji v legislativě,
- **konzultuje požadavky** na řešení podnikové analytiky v rámci své oblasti řízení, a to z pohledu potřeb firmy i aktuální legislativy,
- definuje **specifikace sledovaných ukazatelů**, struktury a další charakteristiky dimenzí a navrhuje podstatné **vazby ukazatelů a dimenzí**,
- definuje nároky na **obsah a strukturu hlavních reportů**, na jednotlivé sledované ukazatele a jim odpovídající dimenze, na hlavní analytické a plánovací funkce,
- určuje **charakter složitějších analytických operací**, data-miningu, analýzy nestrukturovaných dat a dalších pokročilých technik datové analýzy,
- konzultuje **požadavky na funkcionality** analytických a plánovacích aplikací,
- **analyzuje definované požadavky uživatelů**, resp. celé uživatelské sféry vzhledem k podnikovým metodikám, směrnícím a standardům,

- poskytuje informace a zkušenosti o **stavu a kvalitě relevantních zdrojových databází**,
- formuluje **zadání nových analytických a plánovacích aplikací** a reportů, specifikuje nároky na jejich kvalitu, tj. funkcionalitu, dostupnost, uživatelské rozhraní apod.,
- konzultuje analytické otázky **transformací dat**,
- podílí se na řešení **problémů spojených s kvalitou dat**, definuje potřebné kontroly dat a podílí se na úpravách dat pro podnikovou analytiku,
- spolupracuje na přípravě a **vytvoření databází** datového skladu a datových tržišť z existujících datových zdrojů,
- podílí se na **kvalifikační přípravě** všech ostatních uživatelů.

#### 2.1.4 Uživatel služeb a aplikací podnikové analytiky

Uživatelem služeb podnikové analytiky se rozumí **nejen manažeři**, jimž byla věnována předchozí část, ale i další **pracovníci převážně na střední úrovni řízení** nebo **specialisté firmy**, např. finanční analytici, obchodní a marketingoví specialisté a další. V tomto kontextu se jimi rozumí pracovníci, kteří na rozdíl od klíčových uživatelů jsou pasivními uživateli podnikové analytiky. Realizují tyto základní **funkce**:

- kvalifikovaně **využívají realizované analytické a plánovací aplikace** a výstupy reportingu,
- **konzultují specifické problémy a otázky** analytických a plánovacích aplikací v rámci své působnosti,
- formulují dílčí **nové požadavky** na aplikace a technologie podnikové analytiky vzhledem k potřebám podniku i ke změnám v legislativě,
- **vyhodnocují vlastní způsob využití** provozovaných aplikací, jejich dostupnost, úroveň komunikace, kvalitu vizualizace a další.

## 2.2 Role IT

Skupina rolí IT ve vazbě na podnikovou analytiku zahrnuje hlavní **manažerské role, role analytiků a vývoje**.

### 2.2.1 Informační manažer (CIO)

Informační manažer se podílí na plánování úloh a aplikací podnikové analytiky, zajišťuje **průběžnou kontrolu** nad průběhem projektů podnikové analytiky a podílí se na **řešení zásadních problémů** v jejich rámci. Monitoruje **využití jednotlivých analytických a plánovacích aplikací** a navrhuje případné provozní nebo koncepční změny. Uskutěčňuje tyto hlavní **činnosti**:

- **formulace a realizace strategie** v oblasti podnikové analytiky, zejména definování jejích cílů v souladu s celopodnikovými cíli, určení způsobu a termínů splnění navržených cílů, metrik splnění cílů, zodpovědnosti za splnění cílů,
- určení **strategických úloh řízení firmy** ve vazbě na aplikace podnikové analytiky,

- **zajištění souladu** řešení podnikové analytiky s podnikovou architekturou, s IT architekturou podniku, vývojovými trendy na trhu i s aktuálními potřebami firmy,
- uplatňování **strategie sourcingu** při přípravě i realizaci projektů podnikové analytiky,
- **vyhodnocování nákladů a přínosů** podnikové analytiky, jejich možná kvantifikace a prezentace managementu firmy,
- formulace **pravidel řízení rozvoje a provozu** podnikové analytiky, změny organizačních struktur v relaci na rozvoj podnikové analytiky,
- **řešení vztahů k externím dodavatelům** produktů a služeb podnikové analytiky, schvalování smluv s dodavateli, koordinace smluv s interními i externími zákazníky a dodavateli IT,
- **řešení vazeb útvaru informatiky na ostatní útvary** firmy v relaci k podnikové analytice,
- **plánování a vyhodnocování investic** do podnikové analytiky,
- dohled nad řešením **operativních úkolů**, souvisejících s řízením rozvoje a provozu podnikové analytiky.

### 2.2.2 Sponzor projektu podnikové analytiky

Sponzor projektu **zodpovídá za celý projekt** a dosažení jeho očekávaných **efektů**. **Je jmenován vlastníky** společnosti nebo jejím nejvyšším vedením (CEO). Obvykle představuje **osobnost**, která nejen že je schopná zajistit finanční prostředky, ale i **přesvědčit a získat pracovníky firmy** na úrovni manažerů a podnikových specialistů pro řešení, a především i následné využití projektu. Je tak schopen projekt prosadit a postarat se o jeho pozitivní přijetí. Je obvykle **členem vrcholového vedení**, se silnou pozicí ve společnosti. Zajišťuje zejména tyto **činnosti**:

- zajišťuje **komunikaci a kooperaci s vedením firmy** v kontextu daného projektu,
- s vedením podniku spoluurčuje celkovou **strategii v oblasti podnikové analytiky**, které typy aplikací vzhledem k podnikovým potřebám budou prioritní,
- je obvykle **představitelem řídicí komise projektu**, má hlavní rozhodovací pravomoc,
- **jmenuje** členy řídicí komise, resp. řídicího týmu, včetně manažera projektu,
- zodpovídá za to, že **projekt je přijatelný z pohledu relace cena / výkon**, za naplnění cílů projektu a očekávaných ekonomických i mimoekonomických efektů,
- získává potřebné **finanční a lidské zdroje** a případně se účastní na získání potřebných technických prostředků,
- při řešení koncepčních a organizačních úkolů projektu **kooperuje převážně** s formačním manažerem, manažerem projektu podnikové analytiky a dodavateli.

### 2.2.3 Manažer projektu

Manažer projektu podnikové analytiky **je jmenován** sponzorem projektu, resp. vedením firmy. Jeho **hlavním úkolem** je řídit projekt a projektový tým, tak aby dosáhl **potřebné**

**kvality a očekávaných efektů** z pohledu vlastníků, vedení firmy, pracovníků firmy, případně i externích partnerů. Role manažera projektu v podnikové analytice představuje zejména tyto **činnosti**:

- kooperuje a **komunikuje se sponzorem projektu** na řešení klíčových koncepčních a organizačních otázek projektu a jeho prostřednictvím komunikuje s vedením firmy,
- zajišťuje **komunikaci i s externími partnery**, resp. dodavateli, podílejícími se na řešení projektu,
- je obvykle **členem řídicí komise projektu**,
- zodpovídá sponzorovi za **naplnění cílů projektu** a očekávaných ekonomických i mimoekonomických efektů,
- v kooperaci **s architektem** podnikové analytiky určuje základní **přístup k řešení** projektu na základě posouzení efektů a omezení možných **variant**, případně jejich modifikací a kombinací,
- s informačním manažerem a sponzorem projektu se podílí **na rozvoji celkové koncepce podnikové analytiky**, tj. na určení dalších aplikací včetně pokročilé analytiky a na specifikaci jejich priorit,
- připravuje **zadání projektu**, specifikaci cílů projektu, metriky projektu, určuje strategii a výběr metod řízení a postupu projektu.

#### 2.2.4 Architekt podnikové analytiky

Architekt podnikové analytiky je specialistou v oblasti celkového návrhu aplikací a návrhu IT infrastruktury pro podnikovou analytiku. Analyzuje a navrhuje vhodnou aplikační a technologickou architekturu celé podnikové analytiky firmy. Připravuje podklady pro konfigurování hardware a software. Monitoruje a řeší problémy s optimalizací výkonu aplikací. Realizuje např. **tyto činnosti**:

- řeší **datovou a technologickou architekturu** řešení s respektováním zvoleného přístupu,
- určuje **umístění datového skladu na servery**, umístění jednotlivých datových tržišť, technické řešení OLAP databází,
- zajišťuje výběr a realizaci **produktů pro klientské aplikace**, využití portálů pro reportování z aplikací,
- analyzuje současný **stav infrastruktury** ve vztahu k dosavadnímu řešení podnikové analytiky,
- vyhodnocuje **vytížení kapacit zdrojů**,
- analyzuje současný **stav infrastruktury vzhledem k aktuálním uživatelským požadavkům**,
- analyzuje **stav datových zdrojů** a jejich technickou dostupnost,
- zajišťuje analýzy, hodnocení a monitorování **technologické infrastruktury** s cílem zajistit, aby byla optimálně konfigurována a škálovatelná,

- zpracovává návrh **technologické infrastruktury s respektováním požadavků** na její výkon včetně špičkových zatížení, bezpečnosti provozu, flexibility, možnosti škálování i pracovní náročnosti obsluhy,
- podílí se na **výběru vhodných technologických prostředků** pro projekty,
- řeší **technologické problémy** spojené s návrhy a předpokládaným provozem transformačních procedur.

### 2.2.5 Byznys analytik / Konzultant

Byznys analytik řeší obsahovou a logickou stránku jednotlivých úloh a přípravu analytických a plánovacích aplikací v rámci projektů podnikové analytiky. Podle povahy projektu zajišťuje podle jednotlivých úloh a fází řešení projektů např. tyto **činnosti**:

- **konzultuje s uživateli** jejich problémy a požadavky na analytické a plánovací aplikace,
- posuzuje **kvalitu zdrojových databází**,
- specifikuje rámcový **obsah a strukturu hlavních reportů**, ukazatelů a jim odpovídajících dimenzí,
- definuje základní **funkcionalitu** analytických a plánovacích aplikací,
- řeší **celkovou koncepci transformací dat** (ETL / ELT apod.),
- zajišťuje **analýzu současného stavu řešení** a disponibilních zdrojových databází a aplikací,
- kooperuje na **analýzách požadavků** uživatelů a jejich konsolidaci,
- v kooperaci s dodavatelem zajišťuje **analytické a projektové činnosti** při návrhu **dočasného úložiště dat**,
- navrhuje a projednává s uživateli **hrubý dimenzionální model** řešení,
- navrhuje **datový model datového skladu i datových tržišť** v rámci daného přírůstku,
- navrhuje strukturu a způsob **využití OLAP databází**,
- řeší **analytické otázky transformací dat**, tj. vlastní transformace dat (např. formáty dat), granularitu transformovaných dat,
- **navrhuje kontrolní a opravné procedury** v souvislosti s čištěním, resp. zajištěním kvality dat,
- navrhuje způsob **aktualizací při změnách dimenzí**,
- kooperuje při **přípravě datové základny** pro zahájení provozu aplikací,
- podílí se na **řešení analytických problémů nebo chyb**, vyplývajících s přípravou provozu.

### 2.2.6 Datový analytik

Datový analytik zajišťuje zejména správné **mapování ze zdrojových do cílového systému**, konzistenci dat apod. Zajišťuje podle jednotlivých úloh řešení projektů tyto **činnosti**:

- zajišťuje **analýzu datových struktur zdrojových systémů**, jejich kvalitu a dostupnost,
- spolupracuje s architektem podnikové analytiky na návrhu **datové architektury**,
- spolupracuje na **vyhodnocení stavu a kvality dat**, na formulaci principů řízení jejich kvality,



- konzultuje a posuzuje možné **varianty přístupů k řešení** projektů z pohledu datové architektury,
- definuje **datové struktury cílového systému**, tj. datového skladu, datových tržišť a dalších,
- spolupracuje na **specifikaci nástrojů pro transformace dat**,
- spolupracuje s byznys analytiky na **určení oblastí**, které budou v rámci datových transformací **prioritní**,
- definuje **tabulky, nutné pro plnění cílové databáze**, např. datového skladu a tržišť, které se nenacházejí ve zdrojových systémech,
- definuje **detailní pravidla procedury transformací dat**, popis transformací polí mezi zdrojovými a cílovým systémem (datové typy, délka polí, plnění konstantami),
- spolupracuje na **implementaci databázových schémat** a transformačních procedur,
- spolupracuje na **vytváření prvotních databází** datového skladu a datových tržišť.

### 2.2.7 Vývojář softwaru pro podnikovou analytiku

Vývojář software pro podnikovou analytiku hodnotí požadavky na stávající nebo nové softwarové aplikace a navrhuje, vyvíjí, dokumentuje a provádí údržbu aplikací. Realizuje **následující činnosti**:

- zajišťuje analýzu a **hodnocení požadavků na softwarové aplikace** a související základní software,
- podílí se na **návruhu vývojové a provozní platformy** aplikačního softwaru,
- konzultuje s architektem datové analytiky **specifikace potřebných technických parametrů infrastruktury**,
- realizuje návrh a **vývoj zejména transformačních procedur**, analytických, plánovacích aplikací a reportů,
- kooperuje při **řízení testování aplikací** a validačních postupů,
- zajišťuje **zpracování dokumentace** transformačních procedur a aplikací,
- kooperuje při posouzení, vývoji, zdokonalování a dokumentaci **postupů údržby** operačních systémů, komunikačních prostředí a aplikačního softwaru.

### 2.2.8 Dodavatel

Externí dodavatel představuje **větší či menší firmy** poskytující jednak produkty, ale zejména **komplexní či specializované služby** pro řešení projektů nebo zajištění provozu podnikové analytiky. Rozsah a způsob dodávek **se odvíjí od sourcingové strategie**, přijaté na úrovni vedení firmy. Dodavatel **zajišťuje tyto funkce**:

- spolupráce na celkové strategii a **přístupu k řešení** podnikové analytiky,
- **formulace jednotlivých typů architektur** relevantních vzhledem k projektům,
- **analýza a návrh řešení**, tj. funkcionality, datového zajištění, technologických zdrojů, jejich testování a dokumentace,

- **implementace a zavedení do provozu** realizovaných aplikací a IT služeb, zajištění instalačních, integračních, školicích a dalších služeb,
- **zajištění kompletního postimplementačního servisu** a dalšího rozvoje řešení.

## 2.3 Závěry k rolím v podnikové analytice



- Řešení podnikové analytiky je charakteristické **vysokými nároky na kooperaci analytiků a uživatelů**. S rostoucí komplexností a složitostí analytických systémů tyto nároky dále rostou. Proto **vymezení struktury a náplně jednotlivých rolí** je z tohoto pohledu velmi významné.
- Výše uvedené role tvoří pouze **podmnožinu**, které se úzce váže na jednotlivé součásti řešení podnikové analytiky. Je nezbytné pak **brát v úvahu i další role, zejména manažerské a role firemních specialistů**.
- Náplň jednotlivých rolí je kromě zmíněné kooperace účelná i pro systematické plánování a **řízení kvalifikačních programů** pro podnikovou analytiku a nejen pro ni, de facto pro vytváření celého komplexu školicích programů a aktivit.
- Náplň jednotlivých rolí je nezbytné **modifikovat podle potřeb** a možností firmy a zejména struktury jejích personálních zdrojů.
- Ekonomické i technologické prostředí se postupně a často i rychle mění. Proto je nezbytné **nastavit i systém průběžné aktualizace** definovaných rolí.
- Výše uvedené vymezení rolí se s ohledem na rozsah textu omezuje na jejich funkční náplň, resp. klíčové činnosti. Pro praktické využití je **účelné je doplnit i požadavky na kvalifikaci, resp. znalosti**, a to podle konkrétních potřeb firmy.
- Uplatnění rolí, jejich náplně a požadavků na znalosti je velmi dobré rovněž při **specifikaci náborových akcí** a následně i v personálním řízení při určování obsahu jednotlivých pracovních pozic.

### 3. Faktory ovlivňující podnikovou analytiku



Každý projekt i aplikace podnikové analytiky jsou ovlivněny celou škálou organizačních, technických a dalších faktorů, které je nutné při jeho řešení, nasazení a užití jejich aplikací respektovat.

**Účelem** kapitoly je:

- určit ty **faktory řízení firmy i byznys prostředí**, které významně ovlivňují řešení a užití úloh podnikové analytiky,
- vymezit **jejich podstatné charakteristiky a vlivy** na podnikovou analytiku,
- vytvořit tak předpoklady, aby na základě **vyhodnocení podmínek, resp. faktorů** bylo možné posoudit, jaké řešení podnikové analytiky bude účelné, případně, zda má taková implementace vůbec smysl.

**Faktorem** se rozumí takové uplatnění osvědčených přístupů, postupů a zkušeností, které povedou ke splnění stanovených cílů a požadavků kladených na projekt a k dosažení očekávaných efektů. Faktory jsou **založeny na různých zdrojích**. Jednak vycházejí **z nejlepších praktických zkušeností**, získaných při realizaci obdobných řešení a projektů, jednak z poznatků a znalostí, získaných na bázi výzkumných aktivit, průzkumů, hodnocení vývojových trendů, nebo teoretických prací, které jsou předmětem činnosti analytických společností i akademické sféry. Jinak řečeno, faktory v sobě zahrnují ve velmi komprimované formě všechny **osvědčené poznatky a zkušenosti z dané oblasti**, v našem případě z implementací a provozu aplikací podnikové analytiky.

Faktory podnikové analytiky jsou svým rozsahem, strukturou i obsahem **velmi rozmanité** a je proto účelné je členit a klasifikovat. Při každé klasifikaci v informatice je však zřejmé, že hranice mezi jednotlivými skupinami faktorů nemohou být vždy ostré, a navíc každá taková **klasifikace má subjektivní charakter** a představuje pouze určitý pohled na danou věc. Proto považujeme i další klasifikaci jako za pouze pracovní.

Kapitola **obsahuje tyto faktory a jejich skupiny**:

- **Byznys prostředí** vyjadřuje **vnější podmínky** a vlivy pro uplatnění podnikové analytiky v daném podniku. Sem např. patří:
  - **velikost firmy**, daná zejména počtem zaměstnanců a velikostí ročního obrátu,
  - **původ a vlastnictví firmy**, zda je firma v českém vlastnictví, nebo je pobočkou nadnárodní společnosti, či je výlučně zahraniční firmou,
  - **konkurenční prostředí**, představující zejména sílu konkurenčních tlaků, a tedy i tlaku na uplatnění podnikové analytiky,
  - **odvětví působnosti firmy**, z něhož vyplývají zejména požadavky na obsahovou stránku řešených analytických a dalších úloh.
- **Řízení a organizace firmy** představují **vnitřní podmínky řízení**, ovlivňující obsah, rozsah i užití úloh podnikové analytiky. Patří sem:

- **firemní kultura**, tj. systém hodnot, které podnik vyznává, zaběhnutá schémata jednání a rozhodování atd., má vliv na styl řízení a na úroveň detailu, na jakém se otázky podnikové analytiky řeší a další aspekty,
  - **organizace firmy** je prostředí pro racionální kooperaci pracovníků a pracovních týmů, je výrazem efektivní dělby práce,
  - **dislokace firmy** vyjadřuje regionální rozmístění centrály firmy a jejích jednotlivých obchodních poboček, detašovaných skladů, výrobních provozů apod.,
  - **existence business modelu** je forma pro kvalitní pochopení základního fungování firmy a pro řešení zejména obsahu podnikové analytiky má klíčový význam,
  - **agilní organizace** používá „*agile*“ jako základ svého fungování, kde jednotlivé týmy jsou organizovány do vyšších celků, které se operativně řídí a orientují se na průběžné poskytování výsledků řešení,
  - **regineering podnikových procesů** zaměřený zejména na snižování časové a finanční náročnosti procesů firmy při zachování jejich vysoké kvality,
  - **vztah podniku ke svým obchodním a dalším partnerům** je založen v tomto kontextu na poskytování výstupů, resp. reportů podnikové analytiky v určitých oblastech řízení (stav zakázek apod.) pro obchodní partnery firmy,
  - **personální zdroje** jsou rovněž klíčovým faktorem, ovlivňujícím provoz a rozvoj podnikové analytiky, zejména profesní, kvalifikační i věková struktura pracovníků a úroveň jejich znalostí (viz kapitola věnovaná rolím).
- **Řízení podnikové analytiky** zahrnuje podmínky řízení, spojené s rozvojem a provozem IT ve firmě. Patří sem:
    - **řízení IT ve firmě**, které musí respektovat specifické nároky IT aplikací, v případě podnikové analytiky je to např. vysoká míra účasti uživatelů na řešení aplikací, specifické nároky na jejich provoz apod.,
    - **existence silného sponzora**, tedy osobnost se značnou mírou vlivu a s nezbytnými rozhodovacími pravomocemi, která je také schopna vidět firmu a její aktivity ve všech podstatných souvislostech,
    - **strategie řešení podnikové analytiky**, která je velmi podstatným faktorem řešení s ohledem na značný rozsah těchto projektů, na jejich finanční a časovou náročnost a rychlý rozvoj technologií,
    - **identifikace potenciálních efektů podnikové analytiky**. Je podstatná s ohledem na to, že analytické přípravě a využití těchto aplikací musí věnovat svůj čas i manažeři a specialisté firmy, neboť musí vědět, proč ho do podnikové analytiky mají investovat,
    - **kompetenční centra**, která organizačně a systematicky sdružují pracovníky uživatelských a IT útvarů a vytvářejí tak kvalifikovanou základnu pro kvalitní rozvoj podnikové analytiky ve firmě,

- **využití cloud computingu**, založeného na sdílení hardwarových a softwarových prostředků v infrastruktuře internetu, umožňující rozvoj podnikové analytiky při nižších nákladech a často i s nižšími časovými nároky,
  - **metody řešení nárůstu dat**, kdy při enormních nárůstech objemu dat je nutné hledat a aplikovat efektivní metody řešení tohoto problému.
- **Kvalita informačního systému** představuje hodnocení kvality jeho zdrojů, zejména úrovně řešení kvality dat a kvality podnikových aplikací. Patří sem:
    - **úroveň dokumentace a optimalizace podnikových procesů** vytvářející nezbytné prostředí pro racionální využívání zdrojů pro podnikovou analytiku a současně její efektivní využívání,
    - **kvalita datových zdrojů**, včetně dostupnosti dokumentace datových zdrojů – tento faktor je zde zahrnut (i když byl již zmíněn) jako zdůraznění jeho zásadního významu pro úspěšnost podnikové analytiky,
    - **kvalita podnikových aplikací**, zejména transakčních, je významná právě jako předpoklad kvality datových zdrojů i s ohledem na předpokládaný rozvoj zdrojových aplikací (jejich náhrady, změny, upgrade atd.).

Analýza, návrh, provoz i užití jakékoli aplikace podnikové analytiky, jakéhokoli projektu, jsou vždy ovlivňovány celou řadou **faktorů, které určují, nebo spoluurčují jejich výslednou kvalitu a konečný úspěch**. Přitom platí, že tyto faktory se mohou u jednotlivých typů projektů a aplikací výrazně lišit. Je proto vždy účelné tyto faktory identifikovat a analyzovat vzhledem k typu řešené aplikace, resp. projektu. Jejich poznání se pak nutně promítá do projektových a provozních postupů, tj. do využití jejich pozitivního potenciálu a do aktivit, směřujících k minimalizaci jejich negativních dopadů.

**Detailněji** jsou jednotlivé uvedené faktory analyzovány **v dalších podkapitolách**.

### 3.1 Byznys prostředí

Skupina faktorů, označených jako „byznys prostředí“, vyjadřuje převážně **vnější podmínky** pro uplatnění podnikové analytiky v daném podniku. Do této skupiny spadají především tyto faktory:

- velikost firmy,
- původ a vlastnictví firmy,
- konkurenční prostředí,
- odvětví působnosti firmy.

#### 3.1.1 Velikost firmy

Velikost firmy výrazně **ovlivňuje způsob a potřeby řešení podnikové analytiky**, a na konec i očekávané a následně reálné výsledky. Je obvykle dána počtem zaměstnanců a výší ročního obrátu, obvykle se dělí na 3 kategorie:

- **Malé** firmy: s počtem 1–100 zaměstnanců, s ročním obrátem, který nepřesahuje 30 milionů Kč.

- **Střední** firmy: s počtem od 101 do 500 zaměstnanců a s ročním obratem od 31 do 100 milionů Kč.
- **Velké** firmy: s počtem nad 500 zaměstnanců a s obratem vyšším než 100 milionů Kč.

Zatímco v minulosti byla řešení podnikové analytiky záležitostí převážně velkých, případně středních firem, v současnosti se uplatňuje prakticky na všech úrovních jejich velikosti. Patří sem i následující charakteristiky:

- Pro **velké a střední** firmy je typické **využití komplexních řešení** na bázi rozsáhlých datových skladů, datových tržišť a dalších technologií. Zejména se zesilují potřeby na **standardní reporting**, na integraci a konsolidaci datových zdrojů a samozřejmě na podstatně rozsáhlejší a komplexnější analytické aplikace.
- **Malé**, resp. menší firmy se zaměřují spíše na menší řešení na bázi self service business intelligence (**SSBI**).
- **Velké a střední** firmy disponují obvykle **potřebnými finančními a personálním zdroji**, nutnými pro trvalý rozvoj analytických řešení.
- Na druhé straně **u větších firem** jsou často podstatně **složitější zdrojové systémy**, a tedy i podstatně náročnější transformace dat do analytických databází, stejně tak je podstatně náročnější řízení a zajištění potřebné kvality dat.
- U menších firem jsou **obvykle vlastníci současně manažery**, a tedy jsou často do rozvoje podnikové analytiky přímo zapojeni, u velkých firem je tento problém s vysokým počtem vlastníků složitější.

Všechny uvedené dílčí faktory je nezbytné pečlivě zvažovat již při plánování a zadávání těchto projektů.

### 3.1.2 Původ a vlastnictví firmy

Vlastnictví firmy je jeden z faktorů, který **vyjadřuje formy a složitost vlastnických vztahů** a v tomto případě zejména také to, zda je firma v českém vlastnictví, nebo je pobočkou nějaké nadnárodní společnosti, či je výlučně zahraniční firmou, působící v českém prostředí a na českých trzích. S tím jsou spojeny tyto dopady:

- Podniková analytika a její nástroje jsou vedle managementu **směřovány na potřeby vlastníků**, pokud mají zájem se v dané oblasti angažovat. Podstatné je to, že **schvalují investice do jejího rozvoje**. Jejich souhlas je pak často výrazně ovlivněn tím, do jaké míry jsou aktivními uživateli produktů podnikové analytiky a nejlépe i jejich spolutvůrci, resp. konzultanty.
- Specifické problémy jsou v situaci, kdy firma představuje **pobočku zahraniční nadnárodní firmy**. V tomto případě nejde pouze o schvalování investic, jde i o souhlas s navrhovaným řešením, kdy **mateřská společnost často vyžaduje dodržování celofiremních standardů** – funkčních i technologických. To je na jedné straně pochopitelné, na druhé straně to omezuje invenci a iniciativu tuzemských pracovníků v rozvoji podnikové analytiky.

### 3.1.3 Konkurenční prostředí

Potřeby řešení analytických projektů jsou výrazně ovlivněny i **tržním, zejména konkurenčním prostředím**, kde podnik působí. V segmentech trhu s relativně nízkým konkurenčním tlakem, např. v hutnictví nebo těžkém strojírenství a případně ve veřejné správě, je potřeba projektů podnikové analytiky, a především zájem uživatelů na využívání těchto aplikací omezen a tím i výsledky a efekty z nich jsou často nepříliš výrazné. Stejně tak je omezen i tlak konkurence na efektivitu a kvalitu obchodních aktivit podniku a tím na existenci a úroveň analytických a plánovacích aplikací.

Konkurenční prostředí firem je popsáno v řadě publikací a modelů, jde o **faktor, který v řešení rozvoje podnikové analytiky často hraje klíčovou roli**:

- Konkurence, zákazníci, dodavatelé a další partneři, jejich hodnocení, jejich očekávaný vývoj, jejich **nároky a očekávání se stávají velmi významnou součástí podnikové analytiky** a obvykle tvoří jádro aplikací, zaměřených na strategické řízení, marketing nebo prodej.
- Síla konkurence je faktor, který, zejména u velkých firem, **posiluje potřeby investic zejména do pokročilé analytiky**, jako např. do prediktivní analytiky.
- Faktor konkurenčního prostředí znamená **výraznější potřebu a uplatnění externích datových zdrojů** a s tím spojené nezbytné aktivity, jako např. jejich kvalifikované hodnocení kvality, dostupnosti, finanční náročnosti.

### 3.1.4 Odvětví působnosti firmy

**Společné charakteristiky** ve vztahu k odvětví, kde firma působí, jsou následující:

- Odvětví působnosti firmy je **významné zejména pro obsahové zaměření analytických a plánovacích úloh** a náročnost a složitost jejich funkcionality.
- Odvětví ekonomiky **se liší i dosavadním historickým vývojem v uplatňování nástrojů podnikové analytiky** a tím i rozsahem zkušeností, které týmy těchto firem v průběhu vývoje získaly. Odvětví s nejdelší historií v této oblasti jsou evidentně bankovníctví a finanční služby vůbec, nebo telekomunikace.
- Firmy v různých odvětvích ekonomiky svou složitostí řízení, tlakem na efektivitu, rozsahem různorodých datových zdrojů, již existující IT infrastrukturou, **vytvářejí jak různé předpoklady pro rozvoj a uplatňování analytiky, tak i vyšší či nižší objektivní náročnost na funkcionalitu** a technologickou úroveň aplikací.

## 3.2 Řízení a organizace firmy

Skupina faktorů, označených jako „**řízení a organizace firmy**“, vyjadřuje převážně **vnitřní podmínky řízení** pro uplatnění podnikové analytiky v daném podniku. Do této skupiny spadají především tyto faktory:

- firemní kultura,
- organizace firmy,
- dislokace firmy,
- existence business modelu,

- agilní organizace,
- reengineering podnikových procesů,
- vztah podniku ke svým obchodním a dalším partnerům,
- personální zdroje.

### 3.2.1 Firemní kultura

Zřejmě nejpodstatnějším faktorem je **potřeba** a samotný **zájem vedení společnosti** o aplikace tohoto typu. Ty jsou určeny primárně **pro manažersky nebo analyticky orientované uživatele**, i když s ohledem na jejich stále vyšší dostupnost se předpokládá jejich užití i na středních a nižších úrovních řízení. Z toho vyplývá, že kvalita jejich **řešení a zejména užití není primárně dána předpisy, metodikami, resp. disciplínou pracovníků, ale zájmem, motivací a invencí** na uživatelské i dodavatelské straně.

Kultura firmy, tj. **systém hodnot, které podnik vyznává**, zaběhnutá schémata jednání a rozhodování atd., má vliv na styl řízení a na úroveň detailu, na jakém se otázky v podniku řeší – např. jak detailně jsou předepsány činnosti jednotlivých podnikových procesů:

- Úroveň kultury firmy má také **vliv na iniciativu a kreativitu pracovníků**, která je pro účast na řešení projektů podnikové analytiky často rozhodující.
- Vysoká podniková kultura umožňuje obvykle **méně problémů s rychlejším zaváděním nových aplikací** a technologií, včetně analytických.
- Firemní kultura se **promítá do efektivnosti řízení projektů** i do efektivnosti řízení celé podnikové analytiky.
- Kultura firmy určuje i priority **v orientaci informatiky na určité typy aplikací**, zejména v případě podnikové analytiky.
- Z pohledu kultury firmy má **klíčový vliv** na pozici, rozvoj a užití podnikové analytiky **přístup managementu**.

### 3.2.2 Organizace firmy

**Organizace** firmy je prostředí pro racionální kooperaci pracovníků a pracovních týmů, výrazem efektivní dělby práce. **Organizační struktura** řeší problém **příjemného rozpětí řízení**, tj. počtu pracovníků, který je schopen daný řídicí pracovník efektivně řídit. Na základě toho vznikají organizační úrovně:

- Jasně definovaná a dokumentovaná organizace firmy **přispívá k efektivnosti řízení** a k uplatnění podnikové analytiky.
- Kvalitní organizace přispívá ke **zvyšování výkonnosti celé firmy**, včetně využití metody Corporate Performance Management, CPM.
- Řešení organizace **navazuje na firemní strategii**, resp. je její součástí. K efektivnímu promítnutí strategie do firemní organizace **se využívá metoda Balance Scorecard, BSC**.
- Jasně definovaná podniková organizace **přispívá k rychlé identifikaci problémů**, jejich zdrojů a příčin, a nakonec k jejich řešení i na bázi podnikové analytiky.
- Organizace firmy má být natolik **flexibilní, aby byla schopna rychle reagovat na vývoj** podnikatelského prostředí a vztahů k externím partnerům.



- Organizace firmy **má efektivně respektovat i možnosti podnikové analytiky.**
- Organizace firmy musí, i **s předstihem, reagovat na předpokládané potřeby** v transformacích byznysu a byznys modelu dané firmy.

### 3.2.3 Dislokace firmy

Dislokace firmy vyjadřuje **regionální rozmístění centrály firmy a jejích jednotlivých obchodních nebo výrobních poboček, detašovaných skladů apod.:**

- Vysoká **decentralizace umožňuje často lepší využití dostupných kapacit** (personálních, technických nebo materiálových) pro řešení projektů analytiky.
- S vysokou nebo novou dislokací **vznikají nároky na změny v řízení firmy** a obvykle i na transformaci celého byznysu a současně i nároky na změny analytických a plánovacích aplikací.
- U vysoké dislokace **do zahraničních regionů dochází i k řešení kulturních rozdílů**, pracovních návyků, disciplíny s dopady na řešení a využití analytických úloh.
- U dislokací **do zahraničí** je nutné řešit i tuzemské **legislativní omezení a rozdíly.**

### 3.2.4 Business model

Business model je **forma a přístup pro kvalitní pochopení základního fungování firmy.**

- Podporuje **uvědomění si souvislostí jednotlivých částí a aspektů firmy** a souvislostí v analytických i plánovacích úlohách.
- Umožňuje **kvalifikovanou aplikaci tohoto přístupu při řešení strategických úloh** v řízení firmy.
- Vyžaduje **motivaci a zájem ze strany vedení firmy** pro uplatnění business modelu v analytice a plánování.
- Zahrnuje rovněž postupné **naplnění tzv. „Lean Canvas“** (viz další obrázek):

Problem	Solution	Unique Value Proposition	Unfair Advantage	Customer Segments
1. Nejdříve identifikujte problém, který chcete řešit.	4. Popište základní prvky Vaší firmy.	3. Pojmenujte klíčové hodnoty Vašeho produktu, kvůli kterým bude mít zákazník zájem koupit.	5. Identifikujte výhodu, kterou konkurence nemá/nezíská.	2. Určete typické zákazníky, kteří tento problém mají a budou mít zájem ho řešit.
	Key Metrics		Channels	
	8. Jak budete měřit úspěch v jednotlivých fázích podnikání?		9. Jaké kanály zvolíte k obsluze zákazníků?	
Cost Structure		Revenue Streams		
7. Specifikujte strukturu nákladů potřebných pro rozjezd a fungování firmy.		6. Stanovte, z čeho budou plynout příjmy.		

Obrázek 3-1: Lean canvas

### 3.2.5 Agilní organizace

**Agilní organizace** je podnik nebo jeho část, která **používá agile jako základ svého fungování**. Jednotlivé týmy (využívající např. SCRUM) jsou organizovány do vyšších celků. Existuje několik přístupů/frameworků (např. Spotify, SAFe, LeSS), které podporují orchestraci agilní organizace.

- Některé společnosti **zvolily agilní transformaci v plném rozsahu**, tedy v řešení projektů podnikové analytiky.
- Agilní transformace velkého rozsahu **musí být vedena vrcholovými manažery**. Efektivní vedení však není možné bez osvojení agilních metod a zapojení do transformace.
- Efekty agilní organizace dokumentuje následující obrázek:



Obrázek 3-2: Efekty agilní organizace (Zdroj: [scaledagileframework.com/case-studies](https://scaledagileframework.com/case-studies))

### 3.2.6 Reengineering podnikových procesů

**Hlavní přístupy představují** úlohy řízení podnikových procesů (BPM, Business Process Management), modelování podnikových procesů a reengineeringu podnikových procesů (BPR, Business Process Reengineering), spojené s možnostmi snižování jejich časové a finanční náročnosti. S tím jsou spojené následující aspekty:

- Vytvoření podkladů pro **realizaci organizačních změn**, kdy organizační změny budou založeny na optimalizovaných procesech, tedy optimalizovaném fungování firmy.
- **Změny organizační příslušnosti a kvalifikace pracovníků** se budou provádět v kontextu zlepšení průběhů všech navržených procesů.
- Procesní reengineering je především záležitostí zvyšování kvality řízení firmy. Proto by **hlavní zájem na jeho řešení a úspěchu měli mít podnikoví manažeři**, zejména nejvyšší podnikový management.
- Projekty procesního reengineeringu vyžadují **aktivní účast většiny pracovníků** firmy, musí se však zajistit jejich kvalitní příprava v oblasti procesního modelování, což může narážet na časové nebo finanční bariéry.

- Pro projekty BPR je nutné volit i **adekvátní nástroje pro specialisty**, včetně nástrojů pro podporu analytiky, tedy specializované modelovací nástroje, následně je nutné zajistit dostupnost výsledků projektu široké uživatelské sféře.

### 3.2.7 Vztah firmy ke svým obchodním a dalším partnerům

Výstupy, resp. reporty z datových skladů a tržišť jsou umísťovány na firemní portály, a to nejen pro využití interními pracovníky podniku, ale v určitých oblastech řízení (stav zakázek apod.) **pro obchodní partnery** firmy. Nabídka této funkcionality s pochopením a respektováním potřeb i externích subjektů tvoří jeden z podstatných efektů podnikové analytiky a ovlivňuje tak i její výsledný úspěch.

### 3.2.8 Personální zdroje, úroveň jejich znalostí

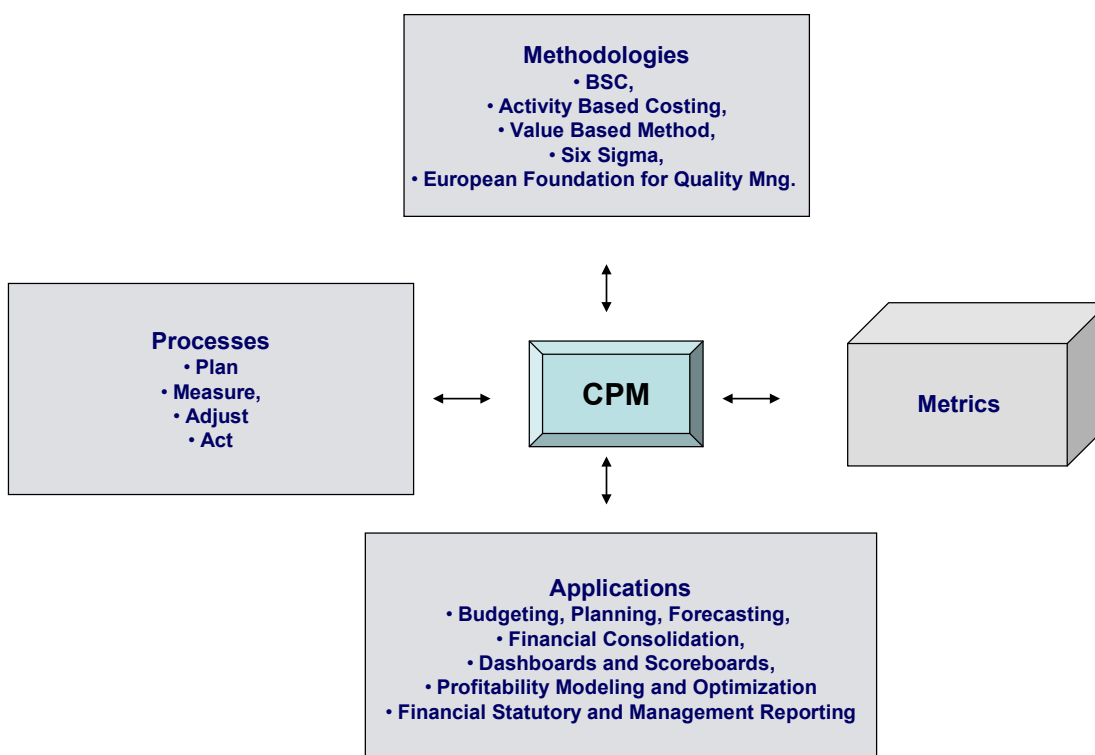
**Ve fázi analýzy** jsou projekty podnikové analytiky charakteristické tím, že **předpokládají velmi intenzivní a kvalifikovanou kooperaci** dodavatelů a řešitelů. To je dáno tím, že vesměs jde o aplikace značně specializované, šité na potřeby konkrétních uživatelů (tedy ne standardní) a jde o aplikace, jejichž **finální přínos** pro firmu bude více než jinde závislý na tom, **jak budou tyto aplikace „chytře“**, v jakém rozsahu v nich budou uplatněny znalosti a zkušenosti jejich dodavatelů i budoucích uživatelů. Čím kvalifikovanější a motivovanější budou tito pracovníci na obou stranách, tím lze očekávat kvalitnější výsledky. To platí u analytických aplikací v podstatně větší míře než u jiných typů aplikací.

Klíčovým faktorem, ovlivňujícím provoz a rozvoj podnikové analytiky, je tak **profesní, kvalifikační i věková struktura pracovníků** a úroveň jejich znalostí, a to jak znalostí pracovníků IT útvaru, tak znalostí jejich uživatelů (uživatelé mohou být i lidé mimo podnik). S tím jsou spojené potenciální efekty a případné problémy:

- **Kvalifikační struktura pracovníků** vyšší úrovně je předpokladem a zdrojem pro rozvoj podnikové analytiky a její efektivní využívání.
- **Znalosti pracovníků**, orientované jak na byznys a manažerské metody, tak na analytické metody v IT, jsou dobrým základem pro kvalitní řešení projektů v podnikové analytice.
- Je účelné orientovat **kvalifikační rozvoj pracovníků IT** na ty oblasti, které nelze efektivně získat v rámci outsourcingu.
- Problém často spočívá v nedostatečné, nebo **špatně orientované motivaci uživatelů a někdy i informatiků** na dalším rozvoji podnikové analytiky.
- Není vytvářen **časový prostor** pro práci klíčových uživatelů na projektech podnikové analytiky.
- **Top management** firmy často **nemá dostatečný nadhled** nad rozvojem a perspektivními možnostmi podnikové analytiky.
- **Kvalifikace CIO** je především technologická, nikoli ekonomická a manažerská.

### 3.2.9 Uplatnění konceptu řízení podnikové výkonnosti

**Řízení podnikové výkonnosti, Corporate Performance Management, CPM**, je kombinace managementu, metodik a metrik, podporovaná aplikacemi, nástroji a infrastrukturou, která umožňuje uživatelům definovat, monitorovat a optimalizovat výsledky a výstupy tak, aby bylo dosaženo cílů osobních či cílů organizační jednotky v souladu se strategickými cíli podniku. Corporate Performance Management (CPM) je hlavním představitelem systémů řízení výkonnosti. Základní vymezení CPM vytvořila společnost Gartner a je v souladu s obecnou definicí systémů řízení výkonnosti. CPM tvoří čtyři základní segmenty, resp. komponenty řešení, které jsou vzájemně provázané (viz Obrázek 3-3).



**Obrázek 3-3: Vymezení CPM (Zdroj: Chandler, 2008)**

Základní **segmenty CPM** tvoří:

- Komplex **manažerských metod**, které tvoří metodologický logický základ podnikového řízení a jejichž principy se respektují v ostatních segmentech CPM. Do těchto metod obvykle patří Balanced Scorecard (BSC), Activity Based Costing (ABC), Value Based Management (VBM), Six Sigma a další.
- **Podnikové procesy**, tj. plánovací, analytické, monitorovací, vytvářející ve svém komplexu procedurální logiku podnikového řízení a navazující na uvedené manažerské metody.
- **Metriky** pro podnikové řízení postavené **na principech podnikové analytiky**, tj. představované klíčovými ukazateli ve vztahu k odpovídajícím dimenzím (zákaznické, komoditní atd.).

- Metody, procesy a metriky jsou základem pro **plánovací a analytické aplikace** postavené na technologiích a přístupech business intelligence.
- CPM tak představuje **komplex metod, procesů, dat a aplikací**, který je nezbytné pro konkrétní podnikové potřeby vždy přiměřeně přizpůsobit. Na druhé straně tento koncept je aplikovatelný jak pro firmy různé odvětvové orientace, tak i různé velikosti.

CPM je založeno na principech a technologiích podnikové analytiky, zejména business intelligence, ale není jeho synonymem. CPM je **rozšiřuje o koncept „řízení“**, který zahrnuje procesy jako plánování, prognózování a základní východiska podnikové strategie a úzce se váže na metody a metodiky pro řízení výkonnosti (např. BSC, ABC a další). Uplatnění aplikací podnikové analytiky podporují uvedené metodiky a je proto jádrem současného konceptu CPM).

K **hlavní funkcionalitě** úloh a nástrojů CPM tak patří komplexní podnikové analýzy na všech úrovních řízení, plánování, rozpočtování a prognózování na bázi klíčových firemních metrik. Lze pracovně vymezit **tři hlavní pojetí řízení výkonnosti**:

- **analytické** pojetí řízení výkonnosti – zaměřené na vymezení a provázanost podstatných komponent řízení podniku,
- **ekonomické** pojetí řízení výkonnosti – obvykle chápáné jako součást manažerského účetnictví,
- **manažerské** pojetí řízení výkonnosti – definující zejména manažerské postupy pro nastavení a hodnocení výkonnosti dílčích útvarů a pracovníků podniku.

Řízení podnikové výkonnosti v analytickém pojetí směřuje k **určení a vymezení hlavních podnikových procesů**, resp. úloh, **metrik, metod a aplikací**, tj. komponent, které významně **ovlivňují úspěšnost podniku**. Definování vazeb mezi uvedenými komponentami **podporuje systematický charakter** řízení. Vymezení postupů a **zodpovědností pracovníků za plnění cílů** ve výkonnosti podniku (manažerské pojetí) zvyšuje úspěšnosti jejich dosažení,

**Ekonomické pojetí** vede k výběru nejdůležitějších ekonomických, převážně finančních ukazatelů a soustřeďuje na ně hlavní pozornost. Hlavním efektem je **provázanost všech tří uvedených pojetí** a vytvoření předpokladů pro kvalitní a systematické řízení podnikové výkonnosti. Principy CPM **posilují i vazby a komunikaci v řízení** na strategické, taktické i operativní úrovni. To znamená i stejné pochopení strategických cílů firmy u pracovníků na všech úrovních řízení a jejich podporování.

### 3.3 Řízení podnikové analytiky

Skupina faktorů, označených jako „**řízení podnikové analytiky**“, vyjadřuje převážně **podmínky řízení**, vyvolané charakterem a nároky IT ve firmě. Do této skupiny **spadají především tyto faktory**:

- řízení IT ve firmě,
- existence silného sponzora,

- strategie řešení podnikové analytiky,
- identifikace potenciálních efektů podnikové analytiky,
- kompetenční centra,
- využití cloud computingu.

### 3.3.1 Řízení IT ve firmě

Aplikace podnikové analytiky jsou dnes již obvyklou, přesto specifickou součástí IT ve firmě. Řízení IT **musí proto respektovat i některé specifické nároky** těchto aplikací, zejména:

- pro analytické aplikace jsou charakteristické **velmi těsné vazby na ostatní aplikace** podnikové informatiky. Jejich úspěch závisí proto i **na kvalitě podnikové a aplikační architektury**, úrovni zajištění integrace podnikové informatiky, na kvalitním plánování a zadávání nových projektů s respektováním možností využití podnikové analytiky (např. využití její analytické funkcionality a nezatěžováním tím transakční aplikace apod.),
- podniková analytika je postavena na **využití datových zdrojů**, vznikajících převážně **v transakčních aplikacích**. Je nutné řešit celý **komplex otázek zajištění kvality dat**, což není záležitostí pouze projektů analytiky, ale procesů řízení celé podnikové informatiky, resp. IT. S jejich nízkou kvalitou klesá kvalita nebo úplně zaniká řešení podnikové analytiky,
- pro kvalitní a systematické řízení rozvoje podnikové analytiky je předpokladem i **systematické a kvalitní řízení datových zdrojů**, v současné době obvykle založené **na principech Data Governance**,
- s tím souvisí **řízení změn** v IT aplikacích. Pokud není zajištěno **efektivní a včasné předávání informací** o změnách v primárních aplikacích správcům analytických aplikací, v řešení dochází **k chybám ve vstupních datových strukturách** v ETL, resp. transformačních procedurách, načítání chybných dat a následně k chybám ve výstupních reportech a a dashboardech,
- úspěch podnikové analytiky ovlivňuje i **úroveň řízení provozu** celé informatiky, zejména správa databázových serverů, plánování a kontrola průběhu ETL procesů, zařazení problematiky analytiky do služeb help-desku atd.

### 3.3.2 Existence silného sponzora

Pravidlo **silného sponzora** je v oblasti podnikové analytiky již všeobecně známé. S ohledem na konečný úspěch by tyto projekty (BI a další) měly být uvnitř firmy vždy podporovány **osobností se značnou mírou vlivu a s nezbytnými rozhodovacími pravomocemi** (viz kapitola 2.2.2). Jde o osobnost, která navíc vedle svého vlivu a prezentovaného zájmu o řešení, je schopna vidět podnik a jeho aktivity ve všech podstatných souvislostech, ve vztahu k podnikovému okolí, je schopna formulovat a rozhodovat o klíčových prioritách řešení a samozřejmě je schopna řešit finanční zajištění projektu a dalšího provozu.

### 3.3.3 Strategie řešení podnikové analytiky

Strategické otázky řešení podnikové analytiky se promítají na nejvyšší úrovni **do informační strategie** firmy, vztahující se k IT firmy jako celku a na nižší úrovni do strategie analytických projektů. Při převládajícím **značném rozsahu** těchto projektů, jejich **finanční a časové náročnosti**, mimořádně rychlém **rozvoji technologií**, na nichž jsou založeny, je určení správné strategie odpovídající potřebám a možnostem firmy **velmi podstatným faktorem** řešení. Stanovení strategie by tak mělo zohlednit i vyhodnocení všech dílčích faktorů, a to zejména:

- určení **rozsahu** projektu a jeho **priorit** ve vztahu k podnikovým procesům, resp. oblastem řízení firmy,
- určení **přístupu** k řešení projektu,
- zaměření řešení na určitý **typ produktu a produkt**, kde se nabízejí následující hlavní varianty:
  - využití **specializovaných analytických aplikací**, vytvářených přesně podle zadání uživatele s využitím nástrojů pro jejich tvorbu, a to **integrovaných do databázových systémů**, jako např. v MS SQL Serveru, nebo **specializovaných nástrojů**, jako např. produkty Informatica apod.,
  - využití **standardních analytických aplikací**, např. pro finanční analýzy, marketing apod.,
  - využití analytických funkcí integrovaných do jiných typů aplikací, např. do aplikací ERP, CRM, APS/SCM apod.,
- určení rozsahu **outsourcingu** v implementaci a provozu analytických řešení, určení podílu vlastních řešitelských kapacit a **způsobu výběru** dodavatelské společnosti,
- možnost využití **specifických modelů** zajištění implementačních a provozních kapacit, např. využití cloud computingu a s ním spojených služeb.

### 3.3.4 Identifikace potenciálních efektů podnikové analytiky

K tomu, aby **potřeba aplikací** podnikové analytiky byla kvalifikovaně posuzována, je nezbytné **formulovat jejich potenciální efekty** s ohledem na danou situaci firmy. Určování a **posuzování efektů** analytiky je v porovnání s ostatními typy aplikací poněkud **specifické**. V každém případě je vymezení očekávaných efektů a sledování jejich naplnění podstatné s ohledem na to, že analytické přípravě a využití těchto aplikací musí **věnovat čas na uživatelské straně manažerů** a podnikoví specialisté, jejichž časové možnosti jsou vesměs omezené. Musí proto, pokud možno, přesně vědět, co jim takto vynaložená časová i finanční investice přinese.

Na druhé straně však efekty podnikové analytiky **nemusí** být vždy zcela **přesně kvantifikovatelné**, resp. ve finančním vyjádření. Často se v těchto případech jedná o **kvalitativní efekty** znamenající dosažení vyšší konkurenceschopnosti firmy, získání lepší pozice na trhu, poskytování kvalitnějších informačních služeb obchodním partnerům atd. Je dobré si v tomto kontextu i položit **otázku: „Jaký bude mít dopad na firmu situace, kdy nebude investovat do podnikové analytiky, zatímco konkurence ano?“**.

Navíc má dosažení těchto efektů **delší časový horizont**, způsobený potřebnou dobou na vytvoření a naplnění datového skladu, vytvoření časových řad sledovaných ukazatelů, osvojení si náročnějších analytických aplikací uživateli apod.

### 3.3.5 Kompetenční centra

Úspěch řešení podnikové analytiky je silně **závislý na efektivní kooperaci dodavatelů a uživatelů** především v analytické fázi řešení. Průzkumy v české i zahraniční praxi však mnohokrát ukázaly, že právě **nedostatečná kooperace** a komunikace mezi byznysem a IT specialisty je překážkou kvalitnějších výsledků.

Jednou z cest je **vytvoření kompetenčních center** pro podnikovou analytiku, které byly i v praxi mnohokrát ověřeny. Jejich **podstatou** je to, že formálně organizačně a systematicky **sdržují pracovníky uživatelských a IT útvarů**, kde společně řeší klíčové problémy a úlohy projektů podnikové analytiky. Jsou postaveny na jasně definované organizaci, pracovních procedurách, dokumentačních a dalších standardech a znamenají tak obvykle významný posun v racionalizaci řešení i provozu aplikací.

### 3.3.6 Využití cloud computingu

**Koncept cloud computingu** je charakterizován sdílením hardwarových a softwarových prostředků v infrastruktuře internetu. Služby cloud computingu jsou významné tím, že uživatelé platí pouze za užívání těchto kapacit a vstupní náklady jsou tak minimální.

Cloud podnikové analytiky, resp. business intelligence, nabízí možnost **soustředit se více na předmět podnikání** a analýzy z byznysového hlediska, než na správu IT infrastruktury. **Cloud business intelligence** lze vymezit jako **model, ve kterém je alespoň některá komponenta BI řešení umístěna v cloudovém prostředí** (tj. mimo infrastrukturu firmy). Komponenty, které se obvykle umísťují do cloudového prostředí, jsou ETL/ELT pumpy, datový sklad a reportingové, analytické a další vizualizační nástroje (Borovec, 2020).

Rovněž lze na Cloud Business Intelligence pohlížet **z hlediska modelů dodávky služeb**. V daném případě poskytování cloudových služeb v modelu software jako služba (**SaaS**) znamená nabízení **předpřipravených BI software**, přičemž z firemního prostředí zákazníka je do cloudového řešení **potřeba migrovat pouze data**. Při dodávání cloudového BI formou platformy jako služby (**PaaS**) se jedná o BI systém, jehož **funkcionalitu si může zákazník upravit podle potřeby**, tak, aby ukazatele z BI analýz mohly přesně měřit výkonnost byznysu (tj. sledovat mimo jiné i naplnění takzvaných klíčových ukazatelů výkonnosti). V daném případě pak jde dodavateli o **nabídnutí co nejširší funkcionality na svých platformách zákazníkům**. Poslední možností je **poskytnutí pouze infrastruktury** jako služby (tj. hardware a software s využitím virtualizace, přičemž veškeré dané infrastrukturní součásti jsou spravované dodavatelem), na kterém **si zákazník samostatně buduje své BI řešení**.

Veškerá cloudová řešení v podnikové analytice, a tedy i v BI, by měla být podložena smlouvou o dodávce služeb (**SLA**), uzavřenou mezi dodavatelem a odběratelem. Součástí by měla být **definice způsobu platby za služby, pojištění** při výpadku služeb



včetně **vymezené hierarchie priorit a kritičnosti** pro zákazníka, tj. které systémy musí být provozuschopné a jaká je jejich kritičnost pro byznys včetně definic, co přesně pro zákaznickou firmu znamená provozuschopnost systémů a co má učinit dodavatel, když nebudou provozuschopné a jaké budou **sankce**, pokud tak neučiní a jak se budou v čase sankce zvyšovat (Borovec, 2020).

### 3.4 Kvalita informačního systému

Skupina faktorů, spojených s informačním systémem podniku, představuje zejména **hodnocení kvality jeho zdrojů**, resp. z čeho může řešení podnikové analytiky vycházet, tj. zejména úroveň řešení podnikových procesů, kvality dat a kvality podnikových aplikací. **Spadají sem tyto faktory:**

- úroveň dokumentace a optimalizace podnikových **procesů**,
- kvalita **datových zdrojů**, včetně dostupnosti dokumentace datových zdrojů,
- kvalita **podnikových aplikací**.

#### 3.4.1 Úroveň dokumentace a optimalizace podnikových procesů

Procesní modely a jejich dokumentace mají obvykle primární užití v transakčních systémech. V podnikové analytice je jejich význam **dán těmito aspekty:**

- díky procesním modelům a reengineeringu podnikových procesů jsou **přesněji definovány problémy a požadavky** i na analytické aplikace, resp. tyto aplikace pak mohou přesněji odpovídat i potřebám podnikových procesů,
- celkové procesní modely podniků jsou **základem pro přesnější a objektivnější specifikaci priorit** v řešení projektů podnikové analytiky, resp. přesnější definici obsahu a pořadí jednotlivých případných přírůstků při přírůstkovém přístupu k řešení,
- současná řešení podnikové analytiky představují nejen jednotlivé analytické a plánovací aplikace, ale zahrnují i **definování analytických a plánovacích procesů** včetně jejich realizací v aplikacích workflow, integrovaných do systémů podnikové analytiky. Procesní podnikové modely pro ně vytvářejí potřebný základ,
- procesní modely jsou rovněž důležitým předpokladem pro **uplatňování konceptu řízení podnikové výkonnosti – Corporate Performance Management**, v němž podnikové procesy jsou jednou ze čtyř součástí.

#### 3.4.2 Kvalita datových zdrojů

Význam kvality datových zdrojů pro řešení podnikové analytiky byl již několikrát zdůrazněn. Do této kapitoly je tento faktor zahrnut pouze jako konstatování jeho klíčového významu pro úspěšnost podnikové analytiky. **Dílní faktory**, ovlivňující datovou kvalitu, lze vymezit v následujících **třech skupinách:**

- **technické prostředí**, zahrnující celopodnikový slovník dat, datový katalog, centralizaci aplikací a jejich datových zdrojů, např. jednotná identifikace zákazníků, kontroly definovaných business pravidel,

- úroveň použité **metodiky**, tj. podnikové metodiky a směrnice, kvalita číselníků a kódových tabulek, systém řízení změn,
- **přístupy k řešení** informačního systému, tj. způsob přípravy dat, přípravy uživatelů, systém motivačních kritérií.

S tím souvisí ještě další podstatné aspekty, a to **dostupnost dokumentace produkčních datových zdrojů**, případně možnost poskytování potřebných dat poskytovateli nebo provozovateli. V některých případech **není zcela jednoduché tyto dokumentace, resp. data získat** s ohledem na autorská práva nebo smlouvy mezi zákazníkem a poskytovateli těchto primárních systémů. V každém případě je dobré si tyto podmínky a možnosti ještě před zahájením projektu ověřit a podle možností je začít řešit.

### 3.4.3 Kvalita podnikových aplikací

Otázka kvality podnikových aplikací, zejména transakčních, je v souvislosti s projekty podnikové analytiky **posuzována v několika úhlech pohledu**:

- do jaké míry jsou tyto aplikace **schopné poskytovat úplná, konzistentní a přesná data**, tj. jak je navržena jejich **datová základna**, jaký **systém kontrol** zahrnuje jejich **funkcionalita**, jak odpovídají potřebám podniku z pohledu poskytovaných funkcí i vytvářených a zpracovávaných dat,
- zda zahrnují **vlastní analytickou a plánovací funkcionalitu**, do jaké míry je využívána, zda je účelné ji nahradit funkcionalitou aplikací podnikové analytiky,
- zda v sobě integrují technologie a aplikace podnikové analytiky, jako je tomu např. u systémů ERP, u CRM s analytickými komponentami, tedy *customer intelligence*, nebo u aplikací APS / SCM pro analýzy a plánování provozu a rozvoje dodavatelských řetězců.

## 3.5 Závěry k faktorům, ovlivňujícím podnikovou analytiku



- „**Faktor**“ představuje významné podmínky firemního prostředí, ovlivňující způsob řešení rozvoje firmy, jednotlivých projektů, a nakonec i jejich i úspěšnost.
- Jedním z **klíčových předpokladů** pro úspěšné řešení a využití analytických aplikací je tak **existence jejich potřeby** z pohledu cílové skupiny uživatelů, tj. manažerů, podnikových analytiků a specialistů. Tato potřeba je buď **dána čistě odborným zájmem jednotlivců, nebo zájmem vedení společnosti** na jejím celkovém úspěchu. Pokud vedení firmy nepovažuje analytické aplikace za účelné, nebo je systém řízení firmy založen spíše na citu a zkušenostech vedoucích pracovníků, pak je lepší projekty podnikové analytiky nezahajovat, nebo je přesunout na pozdější období podle vývoje situace.

- Jednotlivé faktory, zejména **velikost a původ vlastnictví**, obvykle výrazně **určují přístupy vlastníků a managementu k projektům podnikové analytiky**. Ty jsou obvykle pro konečnou úspěšnost projektů zásadní. Pokud není podnikové prostředí ze všech výše uvedených pohledů připravené **analytické aplikace přijmout** a reálně je využívat, pak je jejich řešení problematické a investice do nich nepřinášejí očekávané nebo vůbec žádné efekty. Prakticky všechny uvedené **faktory je proto nezbytné důsledně analyzovat** a posoudit především ve fázi plánování projektu a rozhodnutí o jeho zahájení, případně o rozsahu jeho řešení.
- Je rovněž účelné velmi dobře posoudit dopady **rozdílů, vyplývajících z odvětvové orientace** firmy na obsah řešení podnikové analytiky; právě pochopení a uplatnění rozdílů v obsahové náplni přináší často **konkurenční výhody**, a tedy zásadní efekty, vyplývající z těchto projektů.
- **Klíčovým faktorem** je zde „**Firemní kultura**“, určující prostředí, zájem a motivaci pracovníků firmy na jejím rozvoji, jejich přístup a iniciativu při řešení podnikové analytiky, postavení a zájem managementu na řešení takových projektů a další.
- Je účelné ještě před zahájením projektu **dobře pochopit a vyhodnotit jednotlivé faktory** a na základě takového vyhodnocení modifikovat celkovou strategii a přístup k řešení projektu.
- Na základě vyhodnocení identifikovaných faktorů je na začátku i účelné **posoudit, zda** řešení některých součástí podnikové analytiky **má vůbec smysl**.

## 4. Vybrané metody řešení podnikové analytiky



**Účelem** kapitoly je:

- poskytnout **smysl a vymezení vybraných metod**, uplatňovaných v řešení podnikové analytiky,
- vymezit alespoň hlavní charakteristiky především metod **dimenzionálního a datového modelování**, využívaných při řešení podnikové analytiky.

Úvodem je třeba zdůraznit, že **text této kapitoly navazuje na odpovídající část publikace** „MARYŠKA, M., POUR, J., STANOVSKÁ, I., ŠEDIVÁ, Z.: Self Service Business Intelligence, Praha. Grada, 2018. ISBN 978-80-271-0616-5“. S ohledem na již dřívější rok vydání této publikace a současně na skutečnost, že právě tyto metody představují určité **jádro řešení** podnikové analytiky, pokládáme za účelné uvedenou část i do této publikace zařadit.

V oblasti podnikové analytiky existuje **celá řada metodik obecného i čistě firemního charakteru**. Jako příklady je možné uvést metodiky společností Adastra, IBM, Oracle, Teradata a dalších. Tyto metodiky jsou velmi detailní a obvykle vázané na produkty a služby, poskytované danými společnostmi. Nemělo by zřejmě smysl je zde opakovat, je k nim dostatek zdrojů. Omezíme se pouze na zobecněný a stručný doporučený postup řešení projektu business intelligence v kapitole 9.

Bez ohledu na jejich povahu je nutné zmínit jejich existenci a zejména **racionální způsob využití** jako jednoho z faktorů kvality výsledného řešení. Důraz na uplatnění metodiky je **dán následujícími aspekty**:

- na analýze i provozu analytických aplikací se většinou podílí **řada pracovníků různých profesí** a je proto účelné, aby v jednotlivých částech řešení uplatňovali, pokud možno **společné, standardní postupy**,
- projekty podnikové analytiky jsou často **velmi rozsáhlé**, pokrývají celé nebo velkou část řízení firmy, probíhají **v delším časovém období**, a proto je nezbytné uplatňovat procesní, dokumentační a komunikační **standardy definované metodikou** a omezit tak běžné chyby v kooperaci nebo nedorozumění mezi řešiteli a uživateli,
- **uplatnění** standardních metodik by však mělo být **v praxi efektivní**, tj. mělo by podporovat a racionalizovat standardní operace a postupy, ale na druhé straně neomezovat potřebnou iniciativu a invenci jednotlivých řešitelů (v tom jsou tyto projekty rovněž specifické).

**Další podkapitoly** se váží pouze ke stručnému vymezení **metody dimenzionálního a metody datového modelování** v prostředí podnikové analytiky.

## 4.1 Dimenzionální modelování

Dimenzionální modely, zejména **hrubý dimenzionální model** by měly vytvořit potřebný základ pro návrh a implementaci 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**, tj. úroveň 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**:

1. **dimenze**, s jejichž pomocí se budou ukazatelé analyzovat,
2. **ukazatele** (fakta, metriky), které budou v aplikaci sledovány a analyzovány,
3. **vazby ukazatelů a dimenzí**, tj. který ukazatel se bude analyzovat podle kterých dimenzí.

V následujícím textu jsou uvedeny **charakteristiky dimenzí i ukazatelů**, které 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 pro 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**:

- jednoznačná **identifikace** ukazatele – např. *Tržby\_Objem* – pro účely definování vazeb ukazatele na jednotlivé dimenze a dalšího navrhování tabulek faktů a datových modelů,
- plný název ukazatele, např. *Tržby z prodeje v Kč*,
- obsahové **vymezení a účel** ukazatele – je dobré vymezit ukazatele co nejpřesněji a nejkonkrétněji tak, aby byl vytvořen podklad pro shodu mezi uživateli při 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, např. numerické, textové,
- **měrná jednotka** – např. Kč, % apod.,
- možnost **agregace** ukazatele, tj. aditivní (A), neaditivní (N), semiaditivní (S),
- **KPI**, tj. 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:**

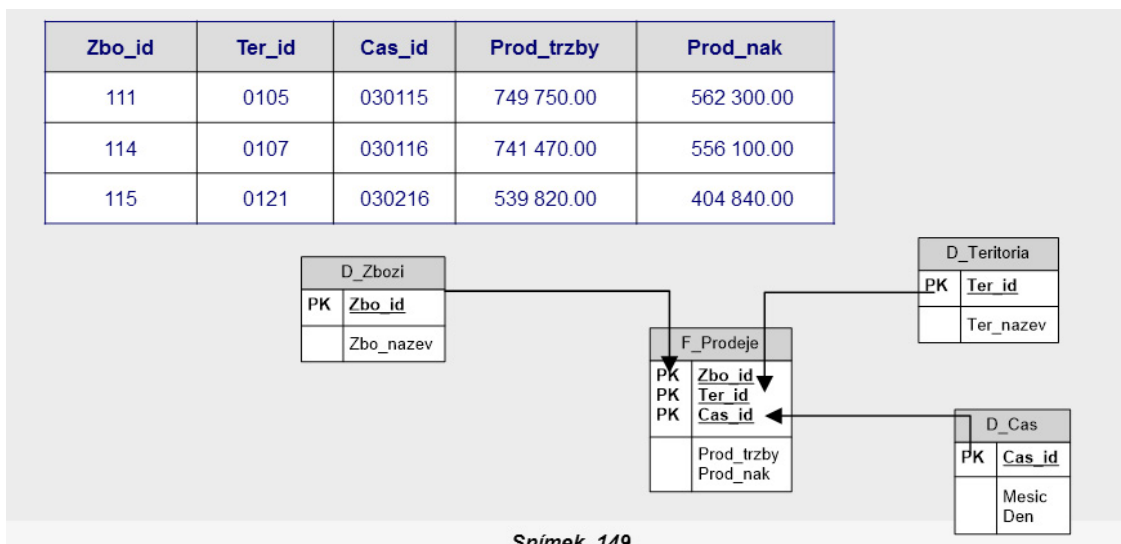
- **identifikace** dimenze – podle stanovených dohodnutých standardů, např. *DI\_Zbozi*,
- plný **název** dimenze, např. *Zboží prodávané firmou*,
- **obsah**, tj. detailnější textové vyjádření obsahu dimenze tak, aby umožňovalo její sjednocené chápání mezi různými uživateli a analytiky,
- **typ** dimenze, tedy zda jde o dimenzi časovou, STAR, SNOWFLAKE, degenerovanou (viz dále),
- **zdroj dat** pro dimenzi, resp. její prvky, např. databáze *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, např.:  
Kategorie zboží – Skupiny zboží – Jednotlivá zboží, zbožové položky,
- **atributy** dimenze, tj. struktura záznamu dimenzionálních tabulek,
- doplňující **poznámky** k vytvoření a využití dimenze, např. požadavky legislativy apod.

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

#### 4.1.1 Principy návrhu tabulek faktů

**Základní principy návrhu tabulek faktů** dokumentuje Obrázek 4-1. **Sloupce** tabulky faktů jsou vesměs buď **klíčové atributy, resp. primární a cizí klíče** (*Zbo\_id, Ter\_id, Cas\_id*), a **hodnoty** ukazatelů (*Prod\_trzby, Prod\_nak*). **Klíčové atributy, cizí klíče, reprezentují jednotlivé dimenze** a jejich hodnoty jsou prvky těchto dimenzí, např. 111 je prvkem dimenze *Zboží*.

**Řádky v tabulce** představují ve své podstatě **jednotlivá měření** (v obchodu, výrobě apod.) a většinou jsou přiřazovány **co nejdetailněji**, tj. pouze na úrovni listů ve strukturách použitých dimenzí. Klíčové položky (listy ve struktuře dimenzí) jsou, s ohledem na nároky na paměťový prostor, vesměs celočíselného typu.



**Obrázek 4-1: Principy návrhu tabulek faktů**

**Příklad** dokumentující vazby tabulky faktů na tabulky dimenzí dokumentuje Obrázek 4-1, kardinalita vazeb tabulek dimenzí na tabulku faktů je vždy 1:N.

#### 4.1.2 Granularita v tabulce faktů

Granularita určuje **úroveň podrobnosti údajů, resp. faktů, uložených v tabulce faktů**. Granularita údajů v tabulce faktů je přímo **závislá na počtu a úrovni podrobnosti dimenzí**, odpovídajících příslušné tabulce faktů. Pro řešení úrovně granularity existují některé **obecná doporučení**:

- pokud to technické kapacity dovolují, měla by být data uložena **s nejvyšší možnou granularitou**,
- data vstupující do datového skladu 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 (tak zvané k její hlavičce), např. náklady na dopravu, a některé k jednotlivým objednaným zbožovým položkám (např. 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 zbožové 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 taková alokace je možná a má smysl a současně, kdo z uživatelské (byznys) sféry se na formulaci pravidel pro alokaci bude podílet. Pokud se alokace ukáže jako neúčelná, pak souhrnné hodnoty (např. dopravné) musí vytvořit samostatnou agregovanou faktovou tabulku. **Nedoporučuje se kombinovat fakta s různou granularitou do jedné faktové tabulky.**

### 4.1.3 Typy tabulek faktů

V datových skladech existují **tři hlavní typy tabulek faktů** vzhledem k jejich granularitě dat, viz další obrázek.

Charakteristika	Transakční	Periodická snímkováná	Akumulovaná snímkováná
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 (Insert)	Přidávání záznamů (Insert)	Přidávání záznamů (Insert) a aktualizace (Update)
Aktualizace záznamů tabulky faktů	Nerealizuje se	Nerealizuje se	Realizuje se vždy při změně
Časová dimenze	Datum transakce	Datum konce časového intervalu	Více datumů pro standardní provádění změn
Fakta (ukazatele)	Obsah transakční aktivity	Obsah odpovídající definovanému časovému intervalu	Obsah odpovídající celému životnímu cyklu dat

Obrázek 4-2: Typy tabulek faktů

**Transakční tabulky faktů** jsou založeny na tom, že detailní informace, vstupující do datového skladu, 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. Transakční tabulky faktů patří v praxi k těm velmi často využívaným.

**Periodické snímkové tabulky faktů** jsou **v praxi nepoužívanější**. Data vstupují do datového skladu **v pravidelných, předem definovaných časových úsecích** (snímcích, např. dnech) a vyjadřují souhrnné hodnoty ukazatelů za celý časový snímek (např. celkový objem transakcí za daný časový úsek). Tento typ tabulek je nejvíce užívaný i pro odhadování, resp. predikci trendů vybraných ukazatelů.

**Akumulované (též někdy stavové) snímkové 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 než předchozí dva typy.

### 4.1.4 Měrné jednotky, rozsah, zdroje a kalkulace ukazatelů,

Tabulky faktů obsahují ukazatele, které **potřebují různí uživatelé sledovat v různých měrných jednotkách**, např. počty vyrobených produktů v kusech, v tisících, krabicích, v paletách apod. Nabízejí se dvě možnosti, buď umístit jednotky a přepočítací koeficienty např. 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 **umístit je do záznamů 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 (např. každý prodej, každý telefonní hovor apod.). Na druhé straně je proto **snaha omezit jejich rozměr co do počtu sloupců a rozsahu jednotlivých sloupců**. Dalším způsobem řešení je určení granularity dat podle období, např. pro posledních aktuálních 60 dnů se využije denní granularita tabulky faktů, pro starší období pak granularita nižší.

Tabulka faktů obsahuje **základní, elementární hodnoty ukazatelů vstupující ze zdrojových databází, i hodnoty kalkulované**, tedy v tomto případě kalkulace v rámci jednoho záznamu (např.  $\text{Prod\_zisk} = \text{Prod\_trzby} - \text{Prod\_nak}$ ). Kalkulace se mohou provádět na úrovni ETL, datového skladu, resp. tržišť, nebo na úrovni analytických aplikací. Obvykle je **užitečné u aditivních faktů ukládat kalkulované hodnoty přímo do datového skladu, resp. tržišť**, 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.

Dimenzionální modely obsahují i tzv. **tabulky faktů bez ukazatelů**, faktů (**factless fact table**), které nemají žádné ukazatele a využívají se např. pro **zjišťování počtu určitých událostí**. To znamená, že každý výskyt záznamu ve fakt tabulce s daným klíčem indikuje vznik události, např. 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.

#### 4.1.5 Tabulky dimenzí – principy návrhu

Princip návrhu tabulek dimenzí dokumentuje Obrázek 4-3.

Produkt_Id	Skupina_Id	Produkt_Skupina	Kateg_Id	Produkt_Kateg	Nazev
111	11	notebook	1	počítač	Acer Travelmate 292
114	12	PDA	1	počítač	Pocket LOOX 410
116	21	monitor LCD	2	monitor	LG L1730S

Obrázek 4-3: Principy návrhu tabulek dimenzí

Tabulky dimenzí jsou de facto **podnikové číselníky**, např. dimenze zboží, zákazníků, dodavatelů atd., a to se všemi možnostmi a problémy, které jsou s nimi spojeny. Dimenzionální tabulka obsahuje **vedle klíčových (obvykle numerických) atributů další řadu většinou textových atributů**, popisujících podstatné charakteristiky jednotlivých produktů, zákazníků apod. Počet atributů je dán nároky na zpracování nejrůznějších podnikových reportů, vycházejících z dat v datovém skladu. 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 řádek, zatímco u tabulek dimenzí je to právě naopak.

Kvalita návrhu celého datového skladu je tak často dána zejména **kvalitou návrhu dimenzí**.

Jedna řádka tabulky je vymezena pouze pro **jeden prvek dimenze**. Každý řádek dimenzionální tabulky musí být identifikován svým primárním klíčem (v případě obrázku – *Produkt\_Id*), který pak také reprezentuje **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, 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**.

Do atributů dimenzionální tabulky se řadí i některé numericky vyjádřené charakteristiky, např. velikost zboží, ale v kontextu tabulky mají čistě popisný charakter. Při návrhu datového skladu je často problémem, **zda příslušný numerický atribut má být součástí určité tabulky faktů, nebo dimenzí**. Řešení spočívá v zodpovězení otázky, zda takový atribut nabývá velkého množství hodnot, a tedy se i často mění, zda se podílí na běžných kalkulacích, v tom případě patří do tabulky faktů. Pokud jde o atribut, který nabývá diskrétních hodnot, relativně konstantních, 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, 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 to, **zda využívat více jednodušších dimenzí, nebo méně dimenzí, ale složitějších a komplexnějších**. Příkladem mohou být jedna dimenze geografická, resp. teritorií, a vedle toho jedna dimenze jednotlivých prodejů. Oproti tomu stojí varianta jedné společné dimenze prodejů, kde na vyšší úrovni budou teritoria, tedy státy, regiony apod. a na nejnižší úrovni jednotlivé prodejny podle jejich umístění v regionech.

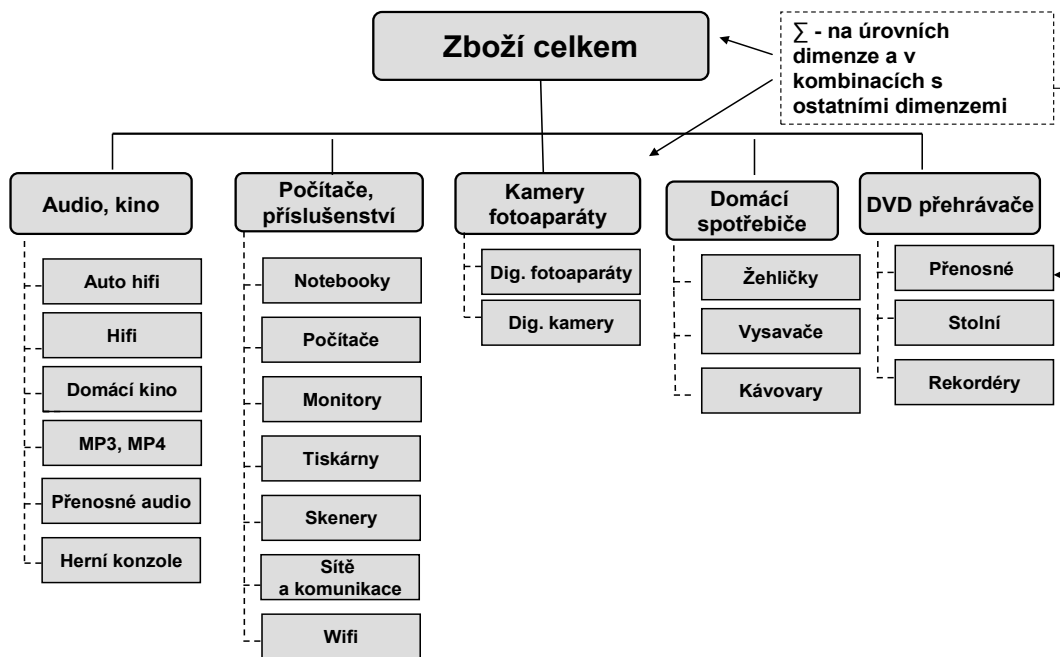
**Varianta více jednodušších dimenzí má tyto výhody:**

- je jednodušší pro pochopení uživateli,
- je jednodušší jejich správa a provádění úprav,
- nabízí více možných kombinací mezi různými dimenzemi v analytických aplikacích.

Oproti tomu **varianta méně dimenzí, ale komplexnějších, je výhodná v tom, že:**

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

Podstatnou charakteristikou dimenzionálních tabulek je také především to, že **data v dimenzích jsou hierarchicky strukturovaná** tak, aby bylo možné na základě těchto struktur získávat agregované hodnoty ukazatelů v připojených tabulkách faktů. Ve své podstatě struktury dimenzí vyjadřují i struktury řízení v podniku. To dokumentuje příklad na obrázku (Obrázek 4-4).



Obrázek 4-4: Struktura dimenze Zboží

Pro vyjádření hierarchie existují **dvě možnosti její realizace – STAR a SNOWFLAKE**.

#### 4.1.6 Dimenze – STAR

U STAR schématu jsou v základní dimenzionální tabulce **zahrnuty i všechny další sloupce pro nadřazené (parent) úrovně v hierarchii** (kategorie zboží – *Zbo\_Kategorie* a skupina zboží – *Zbo\_Skupina*, resp. identifikátory prvků nadřazených úrovní). To představuje poměrně **vysokou redundanci dat**, kdy se hodnoty atributů nadřazených úrovní hierarchie vícenásobně opakují (Obrázek 4-5).

Zbo_id	Zbo_Kategorie	Zbo_Kat_Nazev	Zbo_Skupina	Zbo_Sku_Nazev	Zbo_Nazev
501	1	Audio, kino	101	Auto hifi	Autoradio Logik
502	1	Audio, kino	101	Auto hifi	Autoradio LG LAC3800
503	1	Audio, kino	101	Auto hifi	Autoradio Pioneer
504	1	Audio, kino	101	Auto hifi	Autoradio Logik
505	1	Audio, kino	102	Hifi	Mikro systém Hitachi AXM717
506	1	Audio, kino	102	Hifi	Mikro systém Panasonic SCPM45
507	1	Audio, kino	102	Hifi	Mikro systém Sony CMTEH25
508	2	Počítače	201	Notebooky	Acer Aspire One A150
509	2	Počítače	201	Notebooky	HP Compaq 2133
510	2	Počítače	201	Notebooky	Asus X51L
511	2	Počítače	203	Monitory	LG W1934S BN
512	2	Počítače	203	Monitory	ASUS VW193B
513	4	Domácí spotřebiče	401	Žehličky	Philips GC2528
514	4	Domácí spotřebiče	401	Žehličky	Bosch 250
515	4	Domácí spotřebiče	402	Vysavače	Zelmer 3500
516	4	Domácí spotřebiče	402	Vysavače	Zelmer 5000
517	4	Domácí spotřebiče	403	Kávovary	Espresso Delonghi ESAM2600
518	4	Domácí spotřebiče	403	Kávovary	Espresso Delonghi E2100

Obrázek 4-5: Dimenze STAR

Pro STAR schéma jsou z hlediska provozu a uživatelských aplikací **podstatné tyto charakteristiky**:

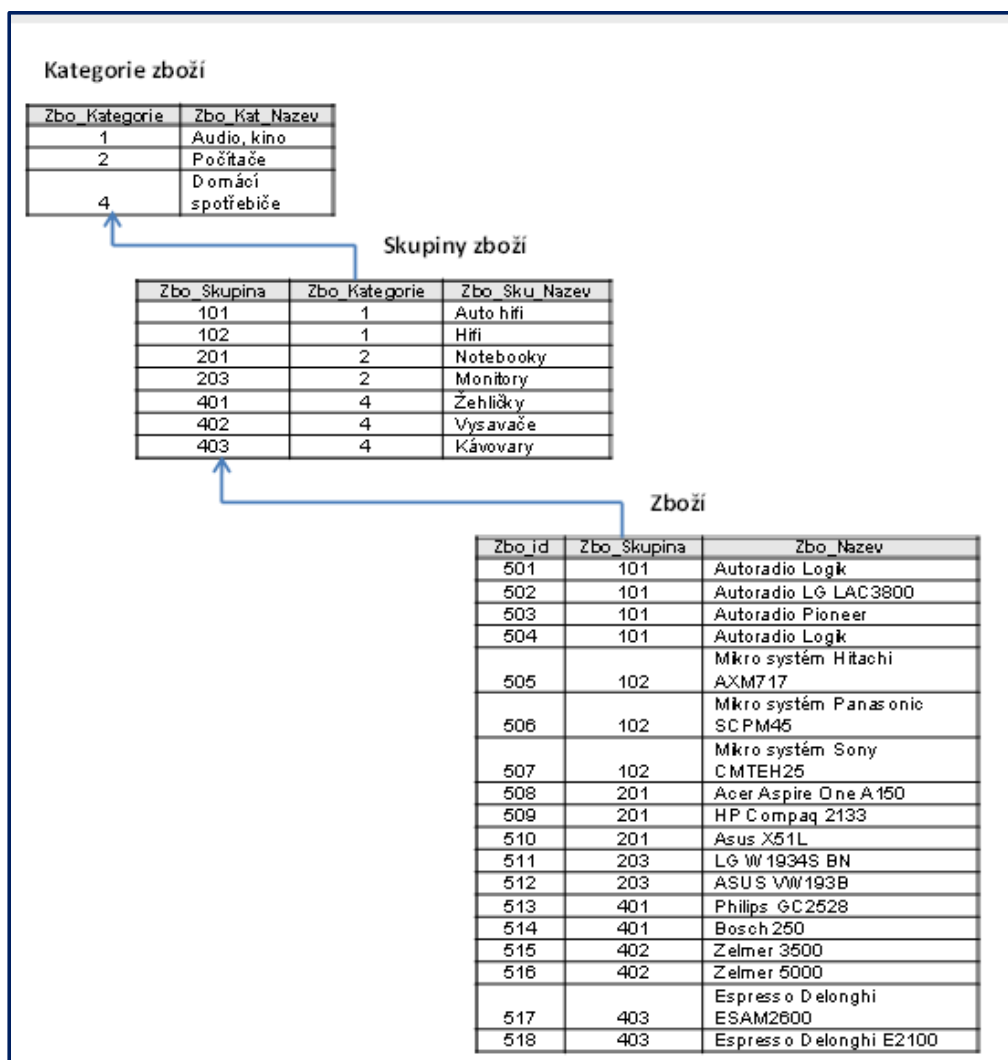
- je evidentně **rychlejší v době odezvy** pro poskytování výstupů, neboť odpadájí operace spojování (join) mezi tabulkami jednotlivých úrovní a stačí zpravidla

jedno spojení mezi tabulkou faktů a dimenzí, proto se využívá u těch komponent analytiky, bezprostředně vázaných k vlastnímu využití (datová tržiště, datové sety v Power BI apod., viz dále),

- umožňuje **jednodušší prohlížení (browsing) dimenzí**, a zadávání filtrů pro všechny hierarchické úrovně dimenze,
- je však **neefektivní při častých změnách** v hierarchiích prvků dimenze, neboť jedna a tatáž změna se musí promítnout do mnoha řádek tabulky.

#### 4.1.7 Dimenze - SNOWFLAKE

Hierarchie ve SNOWFLAKE schématu je **založena na řetězci provázaných tabulek** vždy s kardinalitou **1:N** pro dvě související úrovně hierarchie v dimenzi (v našem případě *Zbo\_Zbozi* – *Zbo\_Skupina* a *Zbo\_Skupina* – *Zbo\_Kategorie*). To znamená, že **došlo k normalizaci dat** v tabulkách, redundance dat je minimální, viz Obrázek 4-6.



**Obrázek 4-6: Dimenze SNOWFLAKE**

Pro SNOWFLAKE schéma jsou **podstatné tyto charakteristiky**:

- díky normalizaci dat je toto řešení **výhodné při častých změnách** v dimenzích a v hierarchické struktuře jejich prvků,

- vede **k úspoře místa** v databázi datového skladu, což je ale v důsledku nízkého objemu dat v dimenzionálních tabulkách v relaci k objemu dat v tabulkách faktů často minimální a z hlediska celkového řešení skladu nevýznamný faktor,
- umožňuje využívat **prostředky pro vynucení referenční integrity** mezi jednotlivými úrovněmi tabulek v hierarchii dimenze,
- poskytuje výhody pro **efektivní tvorbu agregačních tabulek**,
- je **méně přehledné** než schéma STAR,
- realizace **spojení tabulek (joinů) je složitá a komplexní**, a v současných databázových systémech i časově náročná.

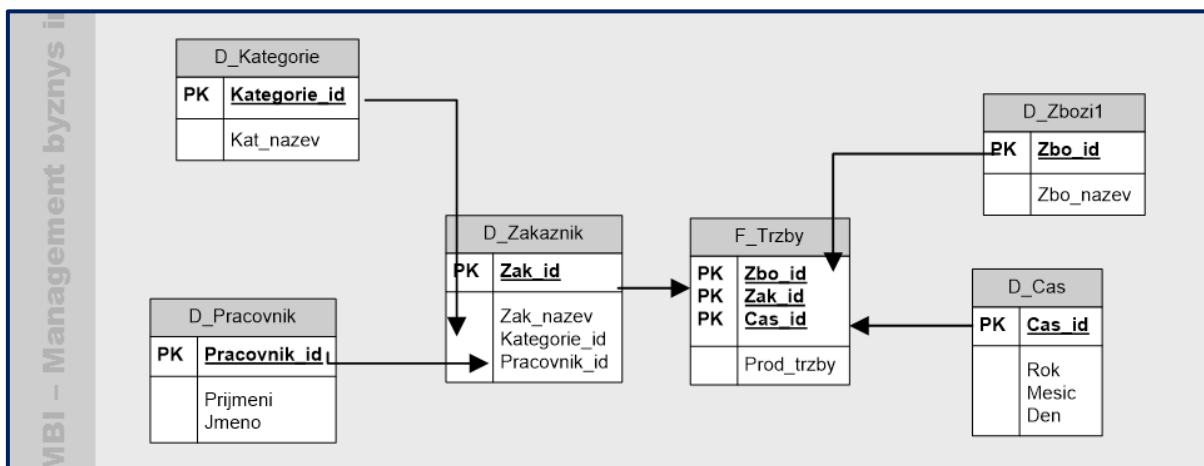
V rámci jednoho datového skladu a jednoho schématu mohou být **současně definovány některé dimenze ve schématu STAR, další tabulky dimenzí ve schématu SNOWFLAKE** a záleží pouze na efektivnosti řešení jednotlivých dimenzí. V dalším textu se budeme zabývat různými variantami řešení dimenzí.

#### 4.1.8 Referenční dimenze a vazby M : N

V terminologii dimenzionálního modelování se pro **dimenzi, která se na tabulku faktů odvolává prostřednictvím jiné dimenze**, používá termín **referenční dimenze**, nebo referenční dimenzionální tabulka. V našem případě je referenční dimenzí *D\_Zakaznik*, viz Obrázek 4-7.

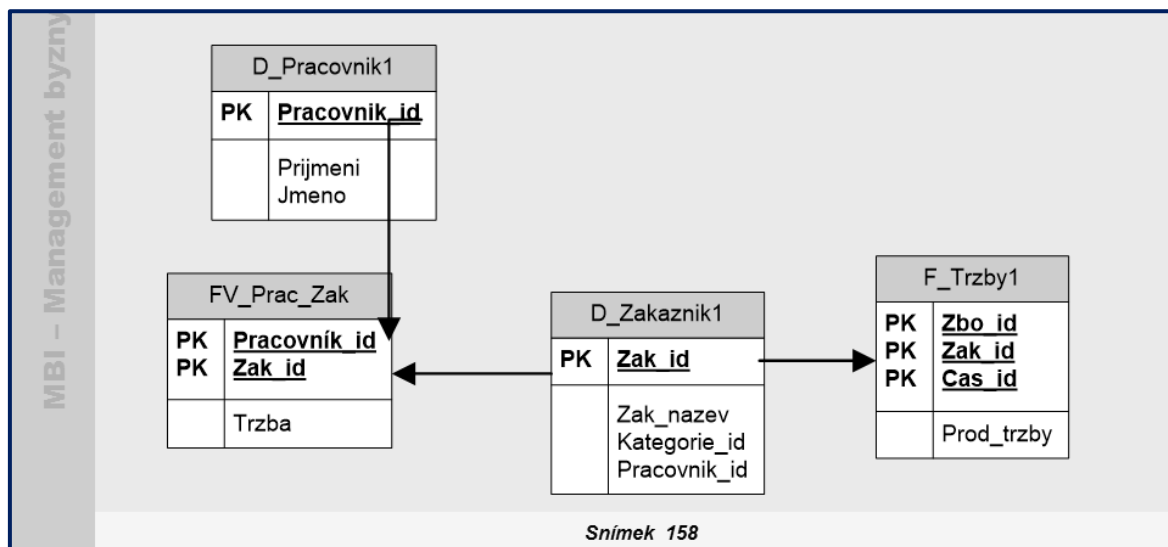
Tyto vztahy nemusí být vždy tak jednoduché, vyjadřující hierarchii určité dimenze, ale mohou **řešit relevantní vazby i na jiné objekty** a mohou být i **více rozvětvené**. Příklad na obrázku ukazuje, kdy na tabulku faktů sledující *Tržby od zákazníků* je vázána dimenze *Zákazník* a na ni další dvě dimenze *Kategorie zákazníků* a *Pracovníci* (mající daného zákazníka na starosti).

Z obrázku vyplývá, že jsme schopni sledovat a agregovat tržby podle kategorií zákazníků a v rámci toho podle jednotlivých zákazníků, nebo pracovníků zodpovědných za zákazníky a v rámci toho podle jednotlivých zákazníků. Toto řešení je postaveno na předpokladu kardinality vazeb mezi nadřazenou a podřízenou dimenzí **1:N (One-to-Many)**. To je zcela obvyklé např. u vztahu *Kategorie zákazníka – Zákazník*.



Obrázek 4-7: Referenční dimenze

U vztahu *Pracovník – Zákazník* to ale tak nemusí být. Se zákazníkem může pracovat více pracovníků a pak se kardinalita vztahu změní na **M : N (Many-to-Many)**. Řešením je využití vazební (pracovní, *bridge*) tabulky faktů, řešící dílčí tržby zákazníků s vazbou na jednotlivé pracovníky (pokud je samozřejmě věcně možné tržby na jednotlivé pracovníky alokovat), viz Obrázek 4-8.



Obrázek 4-8: Vazby dimenzí M : N

#### 4.1.9 Degenerované dimenze

V reálných řešeních existují i tzv. **degenerované dimenze**, tzn., že existuje **dimenze pouze na základě příslušného atributu v tabulce faktů a nemusí pro ni existovat dimenzionální tabulka**. Příklady jsou dimenze *Objednávky*, *Faktury*, jejichž prvky jsou určeny pouze číslem objednávky nebo faktury bez slovního vyjádření. To znamená, že číslo objednávky, faktury apod. je součástí struktury tabulky faktů bez vazby na dimenzionální tabulku. **Analýzy lze tak realizovat i podle těchto identifikací dokumentů**, a tedy i dopovídajících transakcí. Navíc je možné, v případě potřeby, následně získávat na jejich základě i další potřebné informace z OLTP systémů, tj. k jednotlivým objednávkám atd. Prvky degenerovaných dimenzí se mohou účelně využívat v kombinaci se standardními dimenzemi. Např. číslo nákupu (nákupní transakce) v maloobchodě představuje degenerovanou dimenzi, která se často váže na identifikaci pokladny (*POS*, *Point of Sale*), která je standardní dimenzí.

Avšak v případě, že **atribut a současně primární klíč takové dimenze je příliš rozsáhlý** (např. u čísla faktury), nebo je třeba s dokumentem (fakturou, objednávkou apod.) vést další atributy, pak se původní klíč nahrazuje umělým klíčem a příslušnou dimenzionální tabulkou a charakter degenerované dimenze se logicky ztrácí.

#### 4.1.10 Dimenze parent-child

Existují a využívají se i tzv. **parent-child** dimenze, což je např. **dimenze zaměstnanců**, kde:

- na úrovni listu je vždy pouze jeden *zaměstnanec*,

- na vyšší úrovni je *manažer*, který je většinou nadřizený pro několik zaměstnanců, ale ten se odkazuje zpět na jednoho konkrétního zaměstnance – manažera.

#### 4.1.11 Klíče, umělé klíče

Pro identifikaci jednotlivých prvků i vyšších úrovní v dimenzi se doporučuje využívat především **umělých**, systémem **automaticky generovaných klíčů** (*surrogate key*) oproti **operačním**, ve zdrojových **transakčních systémech** využívaným klíčům. Tyto umělé klíče tedy slouží jako primární klíče v tabulkách dimenzí a jako cizí klíče v tabulkách faktů, a tak současně k řešení vazeb (*join*) mezi tabulkami faktů a tabulkami dimenzí. V tabulkách faktů tyto cizí klíče vytvářejí složený primární klíč tabulky faktů.

**Důvody pro umělé primární klíče** v dimenzích jsou:

- umělé klíče **odstiňují datové sklady a tržiště od změn klíčů** ve zdrojových databázích, jsou na těchto změnách nezávislé,
- umožňují **efektivnější a kvalitnější konsolidaci dat** v situacích, kdy operační klíče více transakčních systémů se překrývají, nebo jsou vzájemně nekonzistentní,
- jejich řešení je **jednodušší, menší rozsahem** (obvykle typu integer, 4 byty) a přispívají tak i k vyššímu výkonu datového skladu,
- i při malém rozsahu (4 byty) nabízejí **identifikaci pro cca 2 miliardy záznamů**,
- **úspory v rozsahu** jsou zejména patrné u složených klíčů ve faktových tabulkách (s miliardami záznamů), kde celý rozsah klíče je dán  $n \times 4$  byty, kde  $n$  je počet cizích klíčů (tj. dimenzí), oproti podstatně rozsáhlejšímu operačním klíčům transakčních systémů,
- mohou **vyjádřit i specifické stavy prvků** dimenze (např. „není známo“ apod.),
- pro časové dimenze a při dodržení standardního pořadí, očíslování časových úseků, nabízejí pak **lepší uspořádání dat ve fakt tabulce** a lepší možnosti pro rozdělení faktové tabulky na části, resp. partitioning,
- protože datový sklad nebo tržiště obsahují historická data, **u operačních klíčů** se může stát, že budou po určitém období v transakčních systémech **restartovány**, a to může evidentně způsobit problémy,
- jsou výhodné **pro řešení změn v dimenzích, SCD (Slowly Changing Dimension)**.

Umělé klíče **se obvykle vytvářejí v pracovní oblasti DSA** a současně se zde řeší i vazby dimenzionálních a faktových tabulek. V případě, že složené klíče ve faktových tabulkách nezajistí jednoznačnost záznamů (např. při řešení obchodních operací v maloobchodu), pak se tato **jednoznačnost zajišťuje umělým klíčem** faktové tabulky, nebo kombinací složeného klíče a časového razítka, doplněním čísla obchodní transakce, faktury apod., tedy degenerovanou dimenzí.

**Přirozené klíče** (nikoli umělé) se mohou využít pouze v případech klíče času pro datum (např. 20120910, vyjadřující strukturu data ve tvaru *rrrrmmd*), pro klíče identifikující intervaly hodnot umožňující lépe vyjadřovat jednotlivé intervaly a pro klíče sběrných (*junk*) dimenzí.

#### 4.1.12 Chyby a NULL hodnoty v klíčích

V některých případech se lze setkat s výskytem NULL hodnot v cizích klíčích. Tuto možnost nabízí následně **řešení OLAP kostek, které se tak mohou s NULL hodnotami v cizích klíčích vyrovnat**. Na druhé straně se velmi silně doporučuje se NULL hodnotám v cizích klíčích vyhnout. Je v tomto případě vždy efektivnější příslušnou hodnotu do dimenze doplnit a zajistit tak korektní vazbu mezi dimenzionální a faktovou tabulkou. Souhrnně je tak nutné řešit veškeré možné chyby v konzistenci dat a vazbách mezi primárními a cizími klíči.

#### 4.1.13 Alternativní struktury dimenzí

Pohyb v datech tabulky faktů podle hierarchie prvků v příslušné dimenzi znamená pohybovat se po různé úrovni detailu dat. Jde tedy o operaci drill down / drill up. V rámci jedné dimenze se mohou tvořit tzv. **alternativní struktury**, tj. např. v časové dimenzi (*rok – měsíc – den, rok – kvartál – den* apod.), pro různé organizační struktury v rámci jednoho podniku apod. V tomto případě se uživateli nabízí více možností, jak se pohybovat po datech v různých strukturách hierarchie. **Realizace řešení spočívá v tom, že:**

- nejnižší úrovně hierarchie – listy – obsahují sloupce pro identifikátory, resp. klíče pro více nadřazených struktur,
- agregační operace se pak realizují rozdílně pro různé hierarchické struktury dimenze.

#### 4.1.14 Dimenze času

Dimenze času je v řešeních BI jejich **standardní součástí**. Prakticky vždy se vývoj ukazatelů sleduje v čase, a proto musí být tato dimenze prakticky vždy definována. Obvykle se **generuje z podnikového kalendáře**, nebo se vytváří manuálně, a to tak, aby pokryla existující data v potřebném rozsahu zpět a s potřebným počtem let dopředu (např. 10 let).

Příkladem je vyjádření časové dimenze *20120610*, která jednoznačně vyjadřuje strukturu *rok – měsíc – den*).

Z praktických důvodů **se časová dimenze obvykle dělí na dvě dimenze:**

- dimenzi data (*date dimension*) – rok, kvartál, měsíc, den apod.,
- dimenzi času dne (*time-of-day dimension*) – hodina, minuta, příp. sekunda.

**Není účelné tyto dvě dimenze spojovat** dohromady, neboť dimenze času dne má naprosto standardní strukturu a pro každý den by se tedy její řádky stále opakovaly. To znamená, že každý den v této dimenzi by při rozpadu jen na minuty představoval 1440 řádků, pro jeden rok by to pak bylo 525 600 řádků navíc.

V dimenzi data se jako atributy, kromě celočíselného primárního klíče, **definují jednak běžné součásti data podle stanovené struktury**, např. číslo roku, měsíce, dne, ale využívají se i **další atributy**, jinak označující stanovené časové jednotky, a to **pro reporty a různé výpočty**, např.:



- název měsíce, dne (např. březen, středa apod.),
- pořadí dne v týdnu (1 – 7), v měsíci, v kvartálu, v roce pro účely porovnávání stejných dnů,
- číslo kvartálu,
- datum fiskálního roku (kvartál, měsíc, den),
- identifikátor všedního dne, soboty, neděle, svátků,
- časová pracovní kapacita dne,
- určení sezónního dne, např. prodejní sezóny,
- případně další.

#### 4.1.15 Dimenze různých rolí

Různé role jedné dimenze (*Role-Playing Dimension*) se využívají v těch situacích, kdy **potřebujeme jednu a tutéž dimenzi využívat ve vztahu k tabulce faktů s různým kontextem**, resp. různým obsahovým zaměřením. Příkladem může být uplatnění dimenze času (viz kapitola 4.3.11) ve vztahu k tabulce faktů skladových zásob. V tomto případě můžeme v této tabulce disponovat různými časy nebo daty, např. datem objednání na sklad, datem přijetí na sklad, datem výdeje ze skladu apod. (viz historizace dat). Pracujeme zde **s několika atributy data a v různém kontextu**. Jedním z možných řešení je vytvořit fyzicky různé časové dimenze, což je ale neefektivní a náchylné k chybám, vzhledem k zachování konzistence jednotlivých tabulek. Druhou možností je **na základě jedné fyzické dimenzionální tabulky vytvořit různé databázové pohledy** s respektováním různých názvů položek (pro uplatnění v různých reportech) a ty pak připojovat k faktové tabulce podle aktuálního kontextu. Např. pro výše zmíněnou tabulku zásob může vytvoření pohledu nad časovou tabulkou *DATUM* vypadat takto:

```
CREATE VIEW DATUM_PRIJMU (DATUM_PRIJMU_KLIC, DATUM_PRIJMU_ROK,
DATUM_PRIJMU_MESIC, DATUM_PRIJMU_DEN)
```

```
AS SELECT DATUM_KLIC, DATUM_ROK, DATUM_MESIC, DATUM_DEN.
```

K tabulce faktů bude pak jako dimenze připojen tento pohled a do hlavičky příslušných reportů se vloží titulky odpovídající datu příjmu zboží na sklad. Obdobně lze definovat pohledy i pro další data, spojená s průběhem zboží ve skladu. Takto využitá dimenze v různých rolích se označuje jako **Role-Playing Dimension**.

#### 4.1.16 Sběrná dimenze

Ve zdrojových systémech se často pro rozlišení ukazatelů využívají **nejrůznější kódy a příznaky**, obsahující vždy pouze několik znaků. Příkladem může být způsob platby (hotově, kartou, šekem), plánovaná nebo skutečná hodnota ukazatele apod. Jednou z možností je **pro každý z těchto příznaků vytvořit samostatnou dimenzi** (pak narůstá jejich počet), ponechat je ve faktové tabulce (narůstá rozsah záznamů), **nebo z nich vytvořit společnou, sběrnou dimenzi (junk dimension)**, která bude v jednotlivých záznamech obsahovat kombinace příznaků, kde každý záznam s takovou kombinací bude mít svůj umělý klíč. Příklad dokumentuje Obrázek 4-9.

MBI – Management by	Klíč	Plán_Skut	Způsob platby
	1	Skutečná	Hotově
	2	Skutečná	Kartou
	3	Skutečná	Šekem
	4	Plánovaná	Hotově
	5	Plánovaná	Kartou
	6	Plánovaná	Šekem

**Obrázek 4-9: Sběrná dimenze**

**Na celou kombinaci** se pak záznamy ve faktové tabulce odkazují **příslušným klíčem**. To znamená, že v našem případě nahradí ve faktové tabulce dva cizí klíče pouze jeden, v případě pěti různých příznaků nahradí pět různých cizích klíčů rovněž jeden apod.

Při větším počtu různých kódů a příznaků a jejich hodnot samozřejmě narůstá počet kombinací a záznamů ve sběrné dimenzi. Při jejím vytváření je proto účelné **do ní zahrzovat pouze smysluplné kombinace** a případně je pak doplňovat podle skutečných výskytů. Již vytvoření takové dimenze je užitečnou analytickou úlohou, která určuje, co s čím (jaké kódy) má smysl kombinovat, a navíc výsledné řešení datového skladu zjednodušuje. Tím se tak zvyšuje i orientace uživatele v uspořádání datového skladu a jeho využívání.

Dalším účelným způsobem využití sběrné dimenze je **doplňování poznámek a komentářů k jednotlivým ukazatelům ve faktové tabulce**. Kombinace těchto komentářů mohou být rovněž založeny na principu sběrné dimenze s tím, že jednou z hodnot musí být vždy *Bez komentáře*.

Na druhé straně je se sběrnými dimenzemi spojeno i určité **riziko, kdy při narůstajícím počtu výskytů jednotlivých znaků** a tím i jejich kombinací lze dojít k závěru, že sběrnou dimenzi bude třeba **rozdělit do více základních**. Pak to obvykle znamená i přepracování struktury datového skladu a poměrně vysokou pracnost takového řešení.

#### 4.1.17 Změny v dimenzích – SCD (Slowly Changing Dimensions)

Jedním ze závažných problémů, které se musí v BI, resp. SSBI řešit, jsou **změny v dimenzionálních tabulkách**. Tyto změny se musí **rozlišovat podle významu**, zejména z pohledu analýz podle dimenze času. Některé změny v dimenzích mají **pro hodnocení vývoje a srovnatelnost sledovaných ukazatelů zásadní význam**, některé minimální nebo žádný. Principy řešení změn v dimenzích se označují termínem **SCD – Slowly Changing Dimensions**.



**SCD – Slowly Changing Dimensions** – představuje **změny ve struktuře a prvcích dimenzí, tedy číselníků v čase**, přičemž při řešení úloh BI a SSBI existuje reálná potřeba zachovat konzistenci dat z časového hlediska. V tomto kontextu existují následující **možné změny v dimenzích**, tj. jejich prvků, příp. vyšších úrovní v rámci hierarchie dimenze:

- **přidání** nových prvků (např. nová zboží, zákazníci apod.),
- **zrušení** prvků (např. zrušení zboží),
- **změny hodnot atributů** (názvy produktů, jména pracovníků, adresy zákazníků nebo dodavatelů apod.),
- **změny ve struktuře** – zatřídění (zboží, pracovníků) do jiných skupin v rámci hierarchické struktury prvků dimenze.

Při řešení problémů spojených se změnami obsahu dimenzí je třeba **počítat se složitějším návrhem analytické databáze**, s většími nároky na paměťový prostor dimenzí i s prodloužením běhu datových transformací.

**Historii v datech** lze uchovávat dvojím způsobem, a to **v rámci tabulky faktů**, nebo tzv. **historickými tabulkami dimenzí**. Řešení historie prostřednictvím tabulky faktů je založeno na principu, kdy se měnící se atribut přesune z tabulky dimenze (číselníku) do tabulky faktů. To je výhodné v těch případech, když je takový atribut součástí dalších kalkulovaných ukazatelů, jako je např. prodejní plocha, cena apod.

V dalším textu se budeme zabývat řešením změn přímo v dimenzionálních tabulkách. To je založeno **na různých přístupech k řešení**, označovaných v dimenzionálním modelování jako **typ (Type)** a **číslo typu**, např. *Typ 1*. V každém případě **řešení změn** musí být proaktivní, a tedy již **na úrovni hrubého dimenzionálního modelu** musí být pro každý atribut dimenzionální tabulky uveden typ řešení jeho změn.

**Typ 0** určuje, že změny jsou **ignorovány** a do datového skladu se nepromítají. Skutečné promítání změn dimenzí se rozlišuje několika dalšími typy řešení.

#### 4.1.17.1 Typ 1 – Přepis hodnoty atributu

Řešení *Typ 1* je nejjednodušší, neboť **pouze nahrazuje původní hodnotu atributu hodnotou novou**, změněnou ve stejné řádce dimenzionální tabulky. Předpokládejme, že máme dimenzionální tabulku *Zboží*, kde každý záznam má mimo jiné atributy *Primární klíč* (umělý klíč), *Číslo zboží* (operační klíč z transakčního systému), *Zodpovědný pracovník* (pracovník, který kompletně zodpovídá za prodej daného produktu), viz Tabulka 4-1.

**Tabulka 4-1: Původní záznamy dimenzionální tabulky zboží**

Primární klíč	Název zboží	Číslo zboží	Číslo skupiny zboží	Název skupiny	Zodpovědný pracovník
111	Acer Travelmate 292	NT-0015	01	Notebook	J. Malá
112	IBM ThinkPad R40	NT-0023	01	Notebook	P. Novák

Pokud dojde např. ke změně zodpovědného pracovníka a místo J. Malé bude produkt NT-0015 zajišťovat F. Klas, pak se tabulka změní pouze přepisem původního jména novým, viz Tabulka 4-2.

**Tabulka 4-2: Dimenzionální tabulky zboží s aktualizovaným jménem zodpovědného pracovníka**

Primární klíč	Název zboží	Číslo zboží	Číslo skupiny zboží	Název skupiny	Zodpovědný pracovník
111	Acer Travelmate 292	NT-0015	01	Notebook	F. Klas
112	IBM ThinkPad R40	NT-0023	01	Notebook	P. Novák

Pro změny v dimenzi **typu 1 je charakteristické:**

- **nebyl změněn žádný klíč** (primární, ani operační), a to ani v dimenzionální, ani ve faktové tabulce,
- jde o **nejjednodušší** způsob řešení změn dimenzí,
- **faktová tabulka se nijak nemění**,
- na základě uplatnění typu jedna se však **ztrácí historie sledovaných ukazatelů** podle změn v atributech (nevíme, za jaké prodeje byla ještě zodpovědná J. Malá),
- typ 1 je **vhodný** především tam, kdy **jde o opravu chybných hodnot atributů**, nebo původní hodnoty nemají podle požadavků uživatelů na výstupy nebo aplikace žádný význam,
- všechny **předchozí výstupy**, přehledy, agregace **se musí znovu vytvořit** s novou hodnotou atributu (tam, kde je daný atribut ve výstupech uplatněn).

#### 4.1.17.2 Typ 2 – Přidání nové řádky do dimenzionální tabulky

Typ 2 je **nejčastěji používaný způsob** řešení změn dimenzí. Znamená **přidání nového řádku do dimenzionální tabulky**, který obsahuje změněnou hodnotu atributu, případně atributů. Tím se zachovají historické údaje, ale narůstá objem dimenze. V tomto případě **záznam dimenze obvykle obsahuje:**

- **nový primární klíč** pro nový, resp. doplněný záznam,
- **operační klíč** (původní z transakčního systému), který se se změnami nemění, tj. v novém záznamu obsahuje stejnou hodnotu,
- **datum počátku** platnosti záznamu (DATUM\_OD),
- **datum konce** platnosti záznamu (DATUM\_DO) – u aktuálních záznamů se vyplňuje dostatečně vysokou konstantou (např. 1.1.2200).

Způsob řešení předchozí změny u typu 1 dokumentuje pro typ 2 Tabulka 4-3.

**Tabulka 4-3: Změna dimenzionální tabulky zboží pro typ 2**

Primární klíč	Název zboží	Číslo zboží	Číslo skupiny zboží	Název skupiny	Zodpovědný pracovník	Datum od	Datum do
111	Acer Travelmate 292	NT-0015	01	Notebook	J. Malá	1.1.2011	5.5.2011
321	Acer Travelmate 292	NT-0015	01	Notebook	F. Klas	6.5.2011	1.1.2200
112	IBM ThinkPad R40	NT-0023	01	Notebook	P. Novák	1.1.2011	1.1.2200

Pro změny v dimenzi **typu 2** je, kromě již uvedeného způsobu realizace změny, **charakteristické**:

- **data ve faktové tabulce nebudou** v důsledku změny dimenzionální tabulky **změněna**, tj. záznamy ve faktové tabulce se budou po datu změny (5.5.2011) odkazovat na záznam v tabulce dimenze zboží s již novým primárním klíčem, tedy nikoli 111, ale na záznam 321,
- **výběry dat** podle **zodpovědného pracovníka budou probíhat správně** podle různých hodnot daného atributu, tedy dostaneme např. správné objemy prodeje pro J. Malou i pro F. Klase,
- tím, že zůstává stejný operační klíč (NT-0015), dostáváme **správné hodnoty i pro celkové objemy prodeje** tohoto produktu bez ohledu na změnu odpovědného pracovníka,
- **hodnoty atributů Datum\_od, Datum\_do** slouží **pro pokročilé analýzy** a přesné sledování historie prodeje podle změn atributů, pro základní přehledy (viz předchozí body) se ale nemusí využívat,
- výhodou typu 2 je zejména **možnost sledování vývoje ukazatelů podle změn atributů** a současně možnost sledovat tolik změn a u tolika atributů v dimenzionální tabulce, kolik je třeba,
- rovněž platí, že **nadřízené (master) tabulky** (ve SNOWFLAKE schématu), **využívající změny typu 2, musí mít i podřízené tabulky (detail) se změnami typu 2**, naopak to ale není nutné,
- vzhledem k tomu, že je přesně dáno, kdy došlo ke změně, **není třeba zpětně vytvářet původní výstupy a agregace**.

#### 4.1.17.3 Typ 3 – Přidání nového sloupce do dimenzionální tabulky

Typ 3 představuje **přidání nového atributu do dimenze** (nového sloupce) a **přesunutí staré hodnoty do tohoto nového atributu**. Do původního (aktuálního) atributu je pak možné vložit novou hodnotu. Příklad řešení změny typu 3 ukazuje Tabulka 4-4.

**Tabulka 4-4: Změna dimenzionální tabulky pro typ 3**

Primární klíč	Název zboží	Číslo zboží	Číslo skupiny zboží	Název skupiny	Zodpovědný pracovník	Původní zodpovědný pracovník
111	Acer Travelmate 292	NT-0015	01	Notebook	F. Klas	J. Malá
112	IBM ThinkPad R40	NT-0023	01	Notebook	P. Novák	P. Novák

Pro změny v dimenzi **typu 3** je charakteristické:

- lze efektivně **kombinovat výběry** podle původní i nové hodnoty atributu, např. pro srovnání výkonnosti obou pracovníků apod.,
- na rozdíl od typu 2 **může být předchozí i současná hodnota atributu využívána jako platná** současně,
- typ 3 se však v **praxi často nepoužívá**,
- při větším očekávaném počtu změn se může **využít i větší počet nových atributů**, ale to je možné pouze v případech, kdy počet změn lze u dimenze přesněji odhadovat.

#### 4.1.17.4 Typ 4 – Vytvoření nové historické dimenze

Typ 4 je založen na vytvoření nové **historické dimenze, do níž se přesouvají historické údaje**. Tento přístup se používá obvykle tehdy, kdy v dimenzích dochází k rychlým, respektive častým změnám. Jestliže dojde ke změně, je daná **existující dimenze nejdříve zkopírována do historické dimenze** a následně jsou **v původní tabulce dimenze provedeny potřebné změny**.

Při dimenzionálním modelování ve vztahu k SCD je třeba brát v úvahu, že **různé atributy**, které jsou součástí záznamu dimenzionální tabulky, **budou mít různý typ SCD**. Např. datum narození zákazníka se nebude měnit (kromě chyby v datech), bude mít tedy typ 1, zatímco adresa zákazníka se měnit bude a bude účelné ji zpětně sledovat, tedy typ 2 nebo vyšší. V každém případě je nutné typy SCD u jednotlivých atributů definovat a realizovat v transformačních procedurách (ETL), což samozřejmě zvyšuje jejich složitost.

**Nejčastěji užívanými typy SCD jsou 1 a 2.** Při návrhu analytické databáze se v této souvislosti můžeme rozhodovat **mezi dvěma variantami**:

- **jednu dimenzi lze rozdělit na dvě dimenzionální tabulky**, přičemž v jedné budou atributy typu 1 a v druhé typu 2. Toto řešení je relativně jednoduché a často populární, ale nese s sebou dvě rizika – zvyšuje se počet dimenzí, v dotazech je třeba řešit navíc vazby mezi těmito dvěma dimenzemi, což je tvoří složitějšími a časově náročnějšími;
- **zachovat jednu společnou dimenzi s atributy typu 1 i 2**, což řeší problémy předchozí varianty, ale naopak vede ke složitějším transformacím dat.

Provádění změn v dimenzích **při přípravě dat** je poměrně **časově náročné**. Používá se proto kontrolních součtů v každém databázovém záznamu (jako provozní položka). Teprve v případě nesouhlasu kontrolních součtů se kontrolují jednotlivé položky záznamů a zjišťují se změny, které se pak u jednotlivých atributů řeší podle stanoveného typu změny.

## 4.2 Datové modelování

I když je datové modelování svébytnou disciplínou, v souvislosti s analytikou je dobré rekapitulovat některé základní principy.

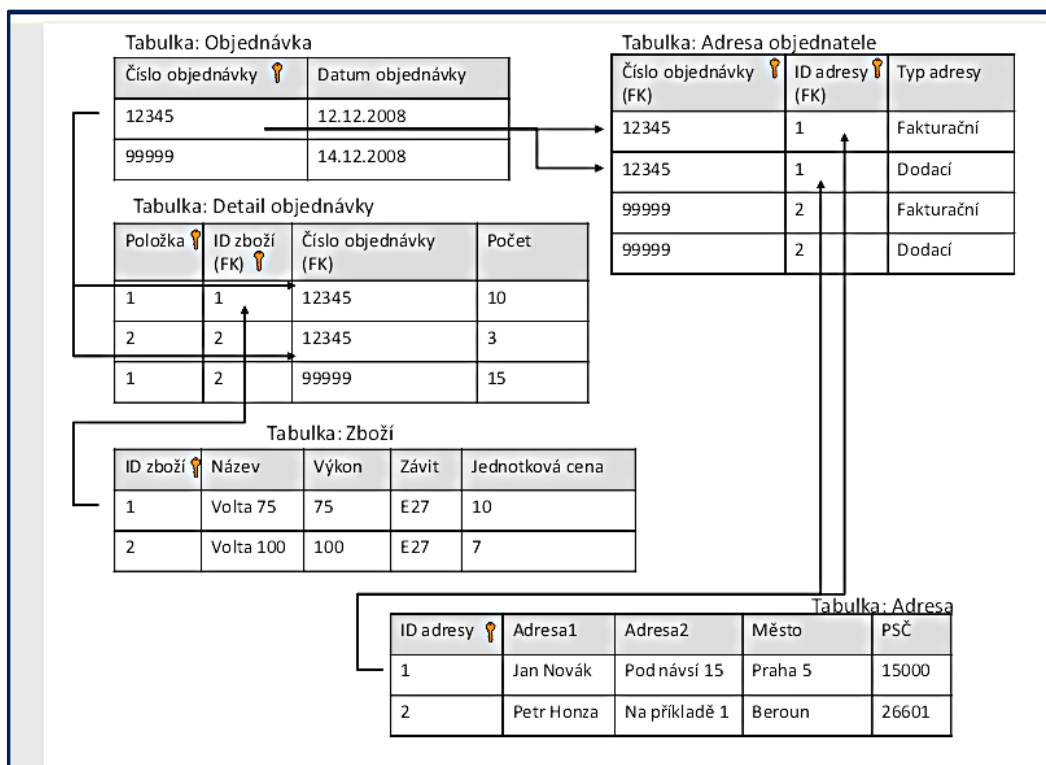
### 4.2.1 Účel datového modelování

Datové modelování je **metoda**, která slouží **k navrhování struktury dat v databázích** a k jejich dokumentaci. **Účelem** datového modelování je přispívat **ke kvalitnímu návrhu databází, jak transakčních, zdrojových, tak analytických**. To znamená, že:

- databáze jsou **konstruovány efektivně s minimalizací redundancí** dat (tj. 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 databáze bylo možné relativně bezpečně rozšiřovat, udržovat a modifikovat. Úpravy databáze nesmějí znamenat nepřiměřená rizika pro kvalitu dat,
- v relační databázi je **řešena integrita na úrovni položek (atributů), tabulek i vazeb mezi tabulkami**, zajišťující přesnost a konzistenci 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.

### 4.2.2 Relační databáze

Datové modelování je spojené obvykle s **relačními databázemi**. Je účelné v tomto kontextu **zrekapitulovat** několik základních termínů.



Obrázek 4-10: Tabulky a jejich vazby v relační databázi

#### 4.2.2.1 Relace

**Relace** – **dvourozměrná datová struktura** představovaná **tabulkou**, kde jeden řádek tabulky odpovídá jednomu záznamu (např. jednomu záznamu zboží) a sloupec odpovídá položce (např. *ID zboží*, *Název*, *Jednotková cena* atd.).

V databázovém pojetí se **sloupec tabulky** označuje jako **atribut**, jehož hodnoty spadají do tzv. **domény**.

#### 4.2.2.2 Integrita

**Integrita** 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). S termínem primární klíč souvisí i **umělý primární klíč (surrogate key)**, nemající nějaký specifický podnikový význam (např. *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 buďto hodnoty *Fakturační* nebo hodnoty *Dodací*.

Posledním typem integritního omezení je tzv. **referenční integrita**, která je **reprezentována cizím klíčem (foreign key)**. Cizí klíč zajišťuje provázanost tabulek, např. mezi tabulkou *Zboží* a *Detail objednávky* je cizím klíčem *ID zboží*, umístěný v tabulce *Detail objednávky*. Na obrázku (Obrázek 4-10) 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 s primárním klíčem. Tak je např. 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 možno odstranit záznam**, jehož hodnota je uvedena jako cizí klíč v tabulce jiné. To např. 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*.

#### 4.2.3 Řešení datových modelů

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 následují obvykle od konceptuálního návrhu, přes logický k fyzickému:



- **Konceptuální návrh** databáze – prvotní návrh databáze, její **struktury, vazeb a všech potřebných charakteristik**:
  - je zaměřen 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**), přiřazenou **identifikaci, vazby** mezi entitami a jejich vlastnosti,
  - v entitách **se nemodelují cizí klíče**, vazby mezi entitami jsou vyjádřeny pouze graficky.
- Na základě konceptuálního návrhu databáze se řeší logický návrh databáze:
  - **cílem** je *detailně vymezit strukturu* databáze založené již na určitém logickém schématu dat, tj. v našem případě relačním,
  - především definovat **struktury databázových tabulek**, jejich atributy, primární klíče a relace mezi tabulkami prostřednictvím cizích klíčů,
- Vytvořením, případně generováním **fyzického návrhu** databáze na základě logického návrhu databáze se definuje fyzický datový model:
  - specifikuje již potřebné **implementační charakteristiky** ve vztahu k příslušnému **databázovému prostředí**, v němž bude databáze realizována,
  - např. **formáty jednotlivých datových položek**, jejich rozsah, resp. délka, definování cizích klíčů a vztahů mezi tabulkami a další charakteristiky.

#### 4.2.4 Podstatné charakteristiky datových modelů

**Entita** představuje z pohledu konceptuálního datového modelu dále **nedělitelnou jednotku reálného světa**, o které se budou v databázi vést příslušná data:

- v databázi entitě obvykle odpovídá **jeden řádek** v databázové tabulce,
- **typ entity** pak představuje zobecnění pro všechny entity se stejnými vlastnostmi (např. typem entity je *Zákazník* a entitou 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,
- v rámci logického návrhu databáze je základem pro **definování struktury** databázové tabulky.
- **Poznámka:** Entita a Typ entity se oba běžně používají pro termín „typ entity“, Výskyt pak pro jednu konkrétní entitu, tj. data o konkrétním sledovaném objektu.

**Vazby** se definují mezi dvěma entitami použitím symbolu vazby s doplněním názvu vazby a dalších jejich charakteristik, včetně kardinality.

**Kardinalita vazby** – násobnost znamená vyjádření **možnosti výskytu počtu entit** na obou stranách vazby:

- existují základní vztahy **1 : 1**, **1 : M**, **M : N**,
- 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,
- vazba **M : N** (**Many to Many**) bude znamenat, že jeden obchodník vyřizuje více objednávek, ale na druhé straně na jedné objednávce se může podílet více obchodníků.

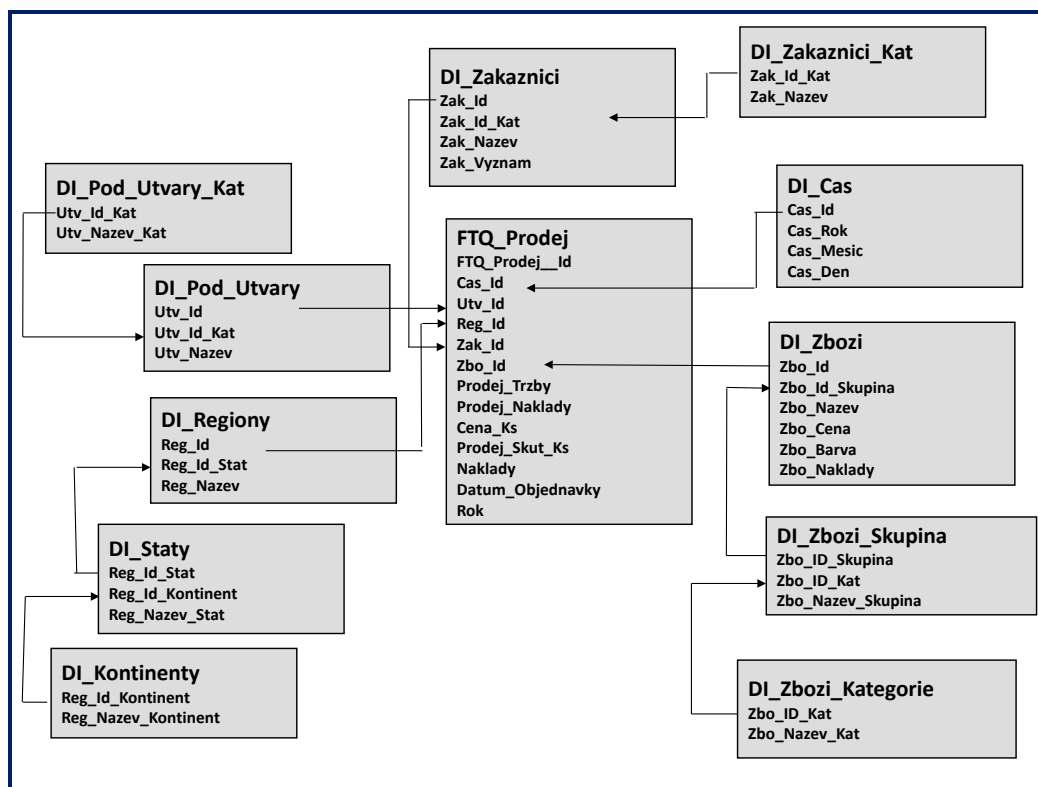
**Povinnost vazby** vyjadřuje **nutnost existence** dané vazby ke každému výskytu entity, například obchodník nemusí být odpovědný za žádnou objednávku, ale objednávka musí mít obchodníka, který je za ni odpovědný.

#### 4.2.5 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.

#### 4.2.6 Návrh datového modelu pro BI a SSBI

Další aktivitou dimenzionálního modelování je analýza a návrh dimenzionálních modelů na bázi principů datového modelování. V návaznosti na hrubý dimenzionální model se



Obrázek 4-11: Příklad: datový model pro BI a SSBI aplikace

**upřesňuje obsah a vazby databázových tabulek dimenzí a faktů s určením běžných charakteristik** (formátu, rozsahu) jejich jednotlivých atributů.

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 4-11.

### 4.3 Závěry k metodám řešení podnikové analytiky



- Před zahájením řešení je účelné rozhodnout o použitých **analytických metodách, metodikách a pracovních postupech**.
  - Tento text obsahuje hlavní principy 2 vybraných metod, a to **dimenzionálního modelování a datového modelování**.
  - S metodami by se měli seznámit a ovládat je nejen informatici (analytici, vývojáři), ale i uživatelé, především **klíčoví uživatelé**. To je významné především u aplikací SSBI, kde se vyšší aktivita uživatelů při řešení předpokládá.
- V souvislosti s přípravou analýzy v rámci řešení jednotlivých projektů je dobré si připravit **sady analytických otázek** a ty pak by měly tvořit **náplň pracovních schůzek** analytiků s jednotlivými manažery a specialisty firmy. Uvedené a další otázky je nutné **konkretizovat podle jednotlivých oblastí řízení** a typu firmy.
- Pro naprostou většinu úloh podnikové analytiky je nutné specifikovat odpovídající **sady metrik**, resp. podnikových ukazatelů a analyzovat jejich hlavní charakteristiky, zejména **analytické dimenze a datové zdroje** pro ukazatele i dimenze.
- Po nezbytném vyjasnění obsahu úloh podnikové analytiky je třeba řešit **technologickou realizaci**, výběr příslušných nástrojů a aplikací a jejich implementace.
  - **Charakteristiky a hodnocení** potenciálních efektů a problémů nástrojů **základní analytiky** jsou náplní **oddílu B**.
  - Řešení aplikací podnikové analytiky je obvykle složeno z celé škály **různých technologických komponent**.
  - Součástí realizace aplikací jsou specifické nároky na **reporting a vizualizace dat**.

## B) Řešení podnikové analytiky

Podniková analytika je v tomto dokumentu pracovně **rozdělena na „základní“ a „pokročilou“**. Tato publikace se vztahuje k řešením, nástrojům a aplikacím, které mají **charakter základní podnikové analytiky**, tj. obvykle tvoří její jádro a je dnes již prakticky běžnou součástí informačních systémů. Z pohledu pokročilé analytiky se text věnuje převážně **prediktivní analytice** v oddílu E.

Do této skupiny **patří**:

- **Business Intelligence (BI)**, tj. řešení, založená na celém spektru **velmi výkonných komponent** (zejména databázových systémů, transformačních nástrojů, analytických a dalších nástrojů), které jsou vzájemně propojeny a tvoří často velmi mohutný **komplex** datových skladů, datových tržišť a dalších komponent. Jsou určeny obvykle pro **rozsáhlé týmy pracovníků** různých útvarů a často i dislokovaných obchodních jednotek.
- **Self Service Business Intelligence (SSBI)** jsou řešení, navazující na původní Business Intelligence, ale založené na **technologiích nižších kapacit**. Jsou určeny pro relativně menší týmy, resp. menší podniky, případně jako doplňky komplexních BI systémů. Pro řešení i užití je charakteristický jejich **operativní charakter**, kratší doba řešení a podstatně nižší náklady, než je tomu u systémů Business Intelligence.
- **Mobilní Business Intelligence** se vztahuje jak ke komplexním BI systémům, tak aplikacím SSBI a jak název napovídá, tak jako koncová (a prezentační) zařízení se užívají tablety nebo smartphony.
- **Competitive Intelligence** se zabývá sběrem, zpracováním a ochranou informací **s cílem získat konkurenční výhodu pro firmu**. Je postaveno primárně na externích informačních zdrojích, často nestrukturovaných, s cílem provádět na nich analýzy převážně prostředí trhu.
- **Poslední kapitolou** tohoto oddílu je možný nebo doporučený **postup řešení projektu Business Intelligence**, pro ostatní typy projektů jsou principy obdobné, ale s ohledem na rozsah je neuvádíme.

U každého uvedeného typu řešení nebo aplikace je v jednotlivých kapitolách obsaženo jejich **funkční vymezení** a hlavní převážně **technologické charakteristiky** a následně i jejich **hlavní efekty** pro firmu na druhé straně **omezení** nebo potenciální problémy. Závěr kapitol tvoří podstatná **doporučení** pro jejich řešení a užití.

## 5. Business Intelligence, BI



**Účelem** kapitoly je:

- poskytnout **podklady pro hodnocení IT aplikací**, nástrojů a technologií, zaměřených na Business Intelligence a určit doporučení pro přístupy k jejich analýze, návrhu a implementaci,
- definovat **podstatné charakteristiky** těchto aplikací a nástrojů,
- vymezit významné **efekty a případná omezení** aplikací a nástrojů Business Intelligence jako podstatné vstupy pro jejich výběr a užití.

### 5.1 Základní principy Business Intelligence

Pokud zatím odhlédneme od realizačních charakteristik a jednotlivých produktů, pak **základní principy** řešení Business Intelligence můžeme shrnout do následujících bodů:

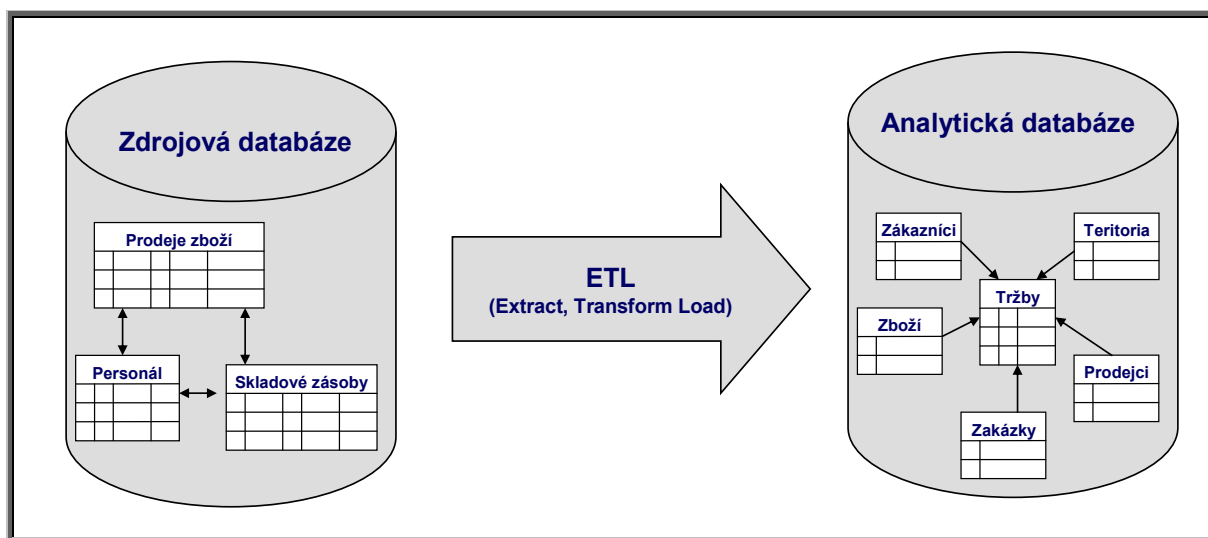
- zatímco transakční systémy jsou primárně určeny pro pořizování a aktualizace dat a tomu odpovídá i organizace dat v databázi, pak řešení Business Intelligence jsou určena pro analytické aplikace a tomu musí odpovídat i **výběr dat** ze zdrojových databází a jejich **organizace** v analytických databázích,
- zatímco transakční systémy udržují data na maximální úrovni detailu (většinou na úrovni jedné transakce se všemi jejími detailními atributy), BI řešení ukládají pouze data relevantní pro analýzy, tedy na potřebné **úrovni detailu (granularity)**, tedy **detailní i agregovaná** podle požadovaných hledisek řízení firmy,
- analytické aplikace pracují primárně s daty podnikových ukazatelů a ty vyhodnocují podle nejrůznějších hledisek, dimenzí a jejich kombinací. To znamená, že BI řešení jsou vesměs založena na **multidimenzionalitě** uložení a zpracování dat,
- zatímco transakční systémy obvykle udržují databáze většinou s aktuálními daty, BI řešení jsou založena na využití **časové dimenze**, to znamená, že ukládají data do analytických databází většinou postupně v jednotlivých časových snímcích,
- uvedené předchozí principy (agregace, multidimenzionalita, časová dimenze) vedou ke specifickým, podstatně vyšším nárokům Business Intelligence na **kvalitu dat**.

Business Intelligence je postavena na celé řadě dalších dílčích pravidel a principů a současně i jejich **různých modifikací podle konkrétního technologického prostředí**. Výše uvedená pravidla však pokládejme pouze za výchozí.

#### 5.1.1 Výběr a organizace dat

Aplikace Business Intelligence (kromě plánovacích aplikací) nevytvářejí nebo nepožizují nová data, ale využívají data, vytvořená transakčními aplikacemi (ERP, CRM atd.). Databáze těchto transakčních aplikací se proto z pohledu BI označují

jako **zdrojové**. Podstatnou vlastností těchto databází je organizace jejich dat podporující přístupy k detailním datům, ukládání a aktualizace dat, tj. že jsou pro tyto operace optimalizované (např. na základě normalizace datových struktur). Oproti tomu analytické BI aplikace jsou optimalizované na efektivní poskytování analytických informací, tj. data zde musí být organizována ve shodě s potřebami analytických úloh, tj. musí obsahovat **hodnoty ukazatelů ve vazbě na analytická hlediska, tedy dimenze**. Z toho vyplývá, že mezi zdrojovými databázemi a analytickými databázemi musí proběhnout transformace dat. Detailněji tento princip dokumentuje Obrázek 5-1.



**Obrázek 5-1: Transformace dat**

**ETL**, resp. **ELT** (Extract, Transform, Load – viz dále), nebo také **datová pumpa** je program, resp. soustava programů, který zajišťuje výběr dat (*Extract*) ze zdrojových databází, jejich transformace (*Transform*) do jiných datových struktur (jiných tabulek a struktur záznamů) a fyzické uložení dat (*Load*) do analytických databází. Z toho dále vyplývají následující charakteristiky ETL:

- ze zdrojových databází musí být vybrána pouze taková **data**, která jsou určena **pro analytické, plánovací a rozhodovací aktivity** podniku, tedy nikoli všechna. To je jedním z prvních úkolů analytiků Business Intelligence takový kvalifikovaný výběr dat určit, tj. rozhodnout, co se bude z datových zdrojů vybírat;
- data jsou transformována do nových datových struktur analytických databází, které musí být předem navrženy tak, aby nejlépe odpovídaly potřebám řízení podniku. To souvisí s možnostmi řešení **multidimenzionality a granularity dat**, k nimž se vrátíme v dalších částech,
- data do Business Intelligence vstupují z různých zdrojových databází (ERP, e-Business, CRM atd.), přičemž v těchto různých zdrojích mohou být jedna a táž data uložena vícekrát (a k tomu ještě různě), např. různé databáze prodej-

ců, zákazníků apod. Ale do analytických databází musí vstoupit pouze jednou. V této souvislosti se o BI často mluví jako o „jedné verzi pravdy“. V transformační vrstvě musí dojít ke **konsolidaci dat**, tj. určení vstupujících dat s vyloučením duplicit či multiplicit. I toto je důležitý úkol pro analytiku a následně provozovatele aplikací Business Intelligence,

- se zajištěním konsolidace dat úzce souvisí i dosažení potřebné **kvality dat, tj. vyloučení chyb, nepřesností atd.**

Právě tato část projektů BI, tedy řešení transformační vrstvy, je pracovně, časově i finančně nejnáročnější a obvykle představuje cca 60 – 80 % vynaložených pracovních kapacit. Pro úspěšné BI řešení však znamená zcela nezbytný předpoklad.

### 5.1.2 Multidimenzionalita uložení a práce s daty

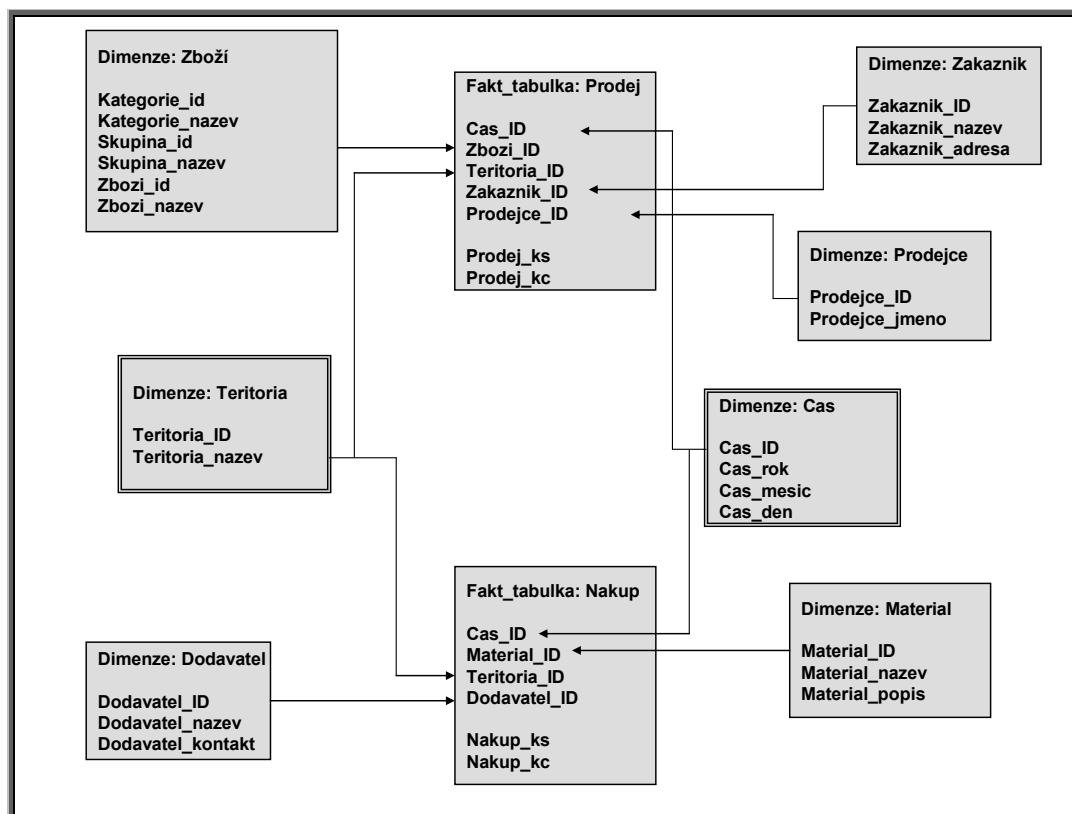
Požadavek pohledů uživatele na sledované ukazatele z více hledisek (dimenzí) a jejich kombinací je základem pro užití principu **multidimenzionality** uložení a práce s daty. To s sebou současně přináší i požadavek na **specifickou organizaci dat** v databázi. Způsob realizace multidimenzionality v datech poskytuje klasicky základní možnosti a podle nich rozdělíme i další objasnění tohoto principu, a to na:

- Multidimenzionalitu, vyjádřenou v relačních databázích,
- multidimenzionalitu dat, realizovanou pomocí tzv. OLAP technologie (On Line Analytical Processing), tj. technologie, vyvinuté speciálně pro tyto účely,
- další možnosti, jako je např. tabulární forma organizace dat.

#### **Multidimenzionalita dat v prostředí relační databáze**

Datové modely produkčních systémů jsou komplexní, obsahují mnoho tabulek a jejich vazeb. Takto organizovaná data jsou z hlediska jejich vytváření a aktualizace velmi efektivní, ale pro běžného uživatele se stávají dosti nepřehledná. Pro výše uvedený nedostatek se objevily snahy o zjednodušení takového uložení dat a jeho přizpůsobení pro tvorbu BI řešení. Vznikl tak relační „**dimenzionální model**“, kterému se také běžně říká „**Schéma hvězdy**“ (*STAR scheme*), resp. „**Schéma sněhové vločky**“ (*SNOWFLAKE scheme*) – viz podkapitoly 4.1.6 a 4.1.7.

Je zřejmé, že databáze datového skladu či datových tržišť (viz dále) nemůže být řešena jedním schématem typu STAR či SNOWFLAKE, ale těchto schémat je v jejich databázích obvykle obrovské množství. Základní princip je v tom, že se sledované ukazatele dělí do tabulek faktů podle logické příbuznosti a tedy i podle stejných přiřazených dimenzí a stejných složených klíčů. Takto se vytvářejí jednotlivá schémata, která se pak provazují pomocí společných, resp. **sdílených dimenzí**. Obrázek 5-2 tento princip dokumentuje na příkladu, kde vidíme, že těmito sdílenými dimenzemi mezi tabulkami faktů *Prodej* a *Nakup* jsou dimenze *Teritoria* a *Cas*.



Obrázek 5-2: Propojení jednotlivých databázových schémat pomocí sdílených dimenzí

### **Multidimenzionalita dat v prostředí OLAP technologie**

**Multidimenzionální databáze** jsou optimalizované pro uložení a interaktivní využívání multidimenzionálních dat. Výhodou multidimenzionality, resp. nasazení OLAP (On Line Analytical Processing) technologií je rychlost zpracování a efektivní analýzy multidimenzionálních dat (drilling, slice and dice apod.). OLAP technologie tak nabízejí efektivní přístup k datům v jednoduché struktuře, vhodné pro analytické činnosti podnikových manažerů a podporu jejich rozhodování. Se zavedením pojmu BI a současně s rozvojem nástrojů a technologií pro širokou podporu analytických činností v organizaci se však výraz OLAP poněkud zúžil.

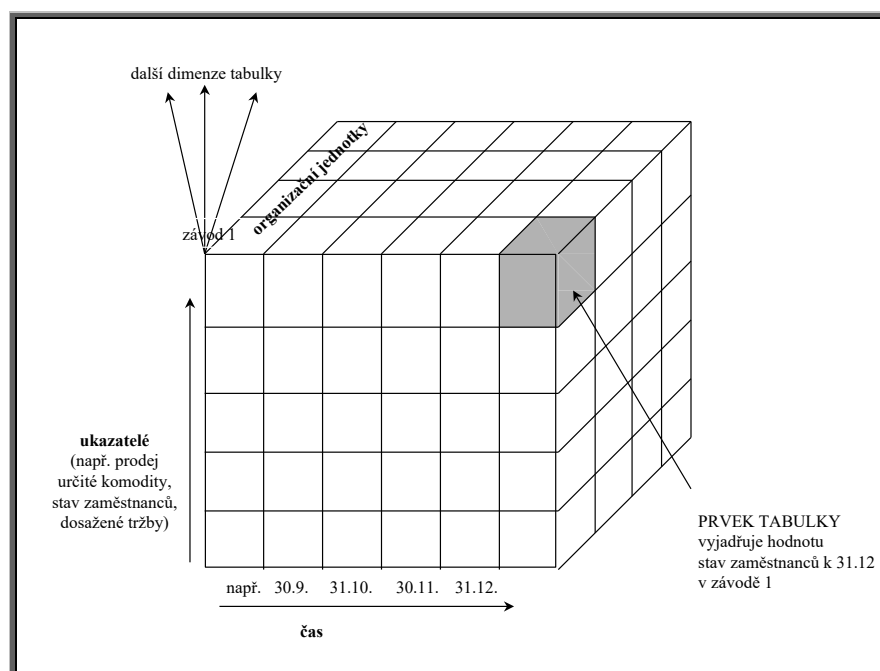


Užší význam definuje **OLAP** čistě technologicky, tedy jako „informační technologii, založenou především na koncepci multidimenzionálních databází. Jejím hlavním principem je několikadimenzionální tabulka, umožňující rychle a pružně měnit jednotlivé dimenze a měnit tak pohledy uživatele na modelovanou ekonomickou realitu. Tento text bude nadále pracovat s užším – technologickým významem výrazu OLAP.

Základním **principem** technologie OLAP je **několikadimenzionální tabulka**, umožňující velmi rychle a pružně měnit **jednotlivé dimenze** a nabízet tak uživateli různé pohledy na ekonomickou realitu. Nabízí možnosti drill-down, drill-up nebo slice and dice



pro výběr („vykrojení“) dat z OLAP kostky, jak dokumentuje Obrázek 5-3. OLAP kostky tak zahrnují předzpracované agregace dat podle výše uvedených hierarchických struktur dimenzí a jejich kombinací. Jde tak v podstatě o princip „n-dimenzionální Rubikovy kostky“, naplněné nejdůležitějšími podnikovými daty.



**Obrázek 5-3: Princip multidimenzionální databáze na bázi OLAP**

Z obrázku vyplývá, že standardními dvěma dimenzemi jsou tu ukazatelé (ekonomické proměnné) a čas. Ostatní dimenze se pro jednotlivé modely definují podle potřeby, např. organizační jednotka, zboží, zákazník, dodavatel, teritorium, konkurent apod. Obsah dimenzí je tvořen **prvky dimenzí** (viz předchozí text). Promítnutí všech dimenzí do jednoho bodu tvoří **prvek OLAP kostky**. Na základě technologie OLAP jsou vytvářeny **OLAP databáze**, které představují jednu nebo několik souvisejících OLAP kostek.

Technologie OLAP se prakticky realizuje **v řadě variant**, z nichž uvedeme alespoň základní:

- MOLAP (*Multidimensional OLAP*) je charakteristická speciálním uložením dat v multidimenzionálních – binárních OLAP kostkách,
- ROLAP (*Relational OLAP*) řeší multidimenzionalitu s využitím technologie relačních databází,
- HOLAP (*Hybrid OLAP*) je kombinací předchozích přístupů, kdy detailní data jsou uložena v relační databázi a agregované hodnoty jsou uloženy v binárních OLAP kostkách.

Multidimenzionální databáze v prostředí OLAP jsou **optimalizované pro uložení a interaktivní využívání multidimenzionálních dat** a jejich výhodou je zejména rychlost zpracování a krátká doba odezvy.

## **Tabulární model**

Tabulární model lze definovat jako „sadu metadat, zahrnující tabulky, vztahy, metriky, KPI, skupiny kalkulací (calculation groups), hierarchie, překlady, bezpečnostní role a mnoho dalších prvků, které tvoří tzv. sémantický model, zajišťující systém navigace v klientských nástrojích, jako je realizace reportů v Power BI nebo v Excelu.

Model odpovídá datové sadě ve službě Power BI anebo databázi ve službě Analysis Services (Azure Analysis Services anebo SQL Server Analysis Service). Datové sady a databáze jsou modely naplněné daty. Definice modelu bez jakýchkoli dat je „tabulární model“ (Russo 2022).

Datové sady anebo databáze, založené na tabulárních modelech, jsou speciální **analytické databáze**, které pro ukládání dat mohou používat dva druhy řešení:

- **In-memory**, což znamená, že vykonávají dotazy na data v paměti aplikace nad celou databází,
- **DirectQuery**, což znamená, že data nejsou kopírována do aplikace, přes kterou k nim přistupujeme (například Power BI Desktop, Microsoft SQL Server Analysis Services, Azure Analysis Services), ale dotazy jsou předávány přímo do zdrojových databází, kde probíhá zpracování a do aplikace je předáván až výsledek dotazu. V tomto případě je tedy tabulární model založen v příslušném nástroji, ale data nejsou jeho součástí.

Tabulární modely **kombinují technologii MOLAP a relačních databází** (Rabeler et al. 2018).

Zatímco **in-memory** modely jsou **výchozí**, **DirectQuery režim se používá pro** příliš velké objemy dat na to, aby se vešly do paměti, nebo pokud volatilita dat vylučuje rozumnou strategii zpracování. **DirectQuery** dosahuje shody s in-memory modely prostřednictvím **podpory široké škály datových zdrojů**, schopností zpracovávat kalkulované tabulky a sloupce v DirectQuery modelu, zabezpečení na úrovni řádků prostřednictvím výrazů DAX, které se dostanou do databáze typu back-end, a optimalizace dotazů, jejichž výsledkem je rychlejší průchodnost (propustnost) (Rabeler et al. 2018).

Tabulární modely jsou definovány strukturami a protokoly dokumentovanými Microsoftem (Russo 2022):

- MS-SSAS-T: SQL Server Analysis Services Tabular Protocol.
- Tabular Object Model (TOM).
- Tabular Model Scripting Language (TMSL).
- XML for Analysis (XMLA) Reference.

Sama definice tabulárního modelu má díky TOM strukturu JSON souboru, známá jako model.bim. Díky tomu lze model vytvářet prostřednictvím:

- Power BI.
- Microsoft Visual Studio (Analysis Services Project).
- Tabular Editoru.

Tabulární modely mohou být **nasazeny do služby Azure Analysis Services**, nebo na instance služby **SQL Server Analysis Services (SSAS)**, nakonfigurované na režim Tabular server, nebo do služby Power BI Premium. Nasazené tabulární modely je možné spravovat přes SQL Server Management Studio (Rabeler et al. 2018).

Nástroje tabulárního modelu původně podporovaly mezi tabulkami **pouze vazby 1 : 1 nebo 1 : N s jednosměrnou i obousměrnou filtrací**. Od roku 2020 jsou implementovány i vazby M : N, i když je doporučeno tento typ vazeb používat pouze ve specifických scénářích užití, stejně jako obousměrné vazby 1 : N.

### 5.1.3 Nároky na kvalitu dat

Jedním z častých problémů současné podnikové informatiky je nízká kvalita dat. To se nejvíce promítá právě do řešení aplikací Business Intelligence, které jsou na kvalitu dat zvláště citlivé, neboť každá dílčí chyba se může v souhrnných reportech nebo analytických aplikacích mnohonásobně projevit nebo zvětšit. Kvalita dat se posuzuje podle čtyř základních hledisek:

- **dostupnost (availability)** – vyjadřuje možnosti, jak může uživatel přistupovat k informacím v okamžiku aktuální potřeby. Dělí se na:
  - dostupnost v čase,
  - dostupnost v místě, lokalitě, kde mají být informace využity,
  - dostupnost v požadované struktuře,
  - dostupnost v požadovaném formátu,
- **přesnost (accuracy)** – sleduje, zda všechna data jsou ve správném kontextu (např. zda odpovídá dané PSČ zadané adrese),
- **úplnost (completeness)** – určuje, do jaké míry jsou k dispozici všechny informace pro daný kontext, např. existuje správné PSČ, ale celá adresa je neúplná. Dalším příkladem je situace, kdy nemusí platit, že jsou všechny požadované transakce faktů neimportovány, tj. jejich celkový souhrn má pravdivý smysl jenom tehdy, když ho tvoří všechny detailní transakce,
- **konzistence (consistency)** – specifikuje možné problémy v porušení standardů nebo vazeb mezi daty. Např. databáze zákazníků v různých lokalitách mají pro stejného zákazníka různá identifikační čísla.

Významným aspektem kvality dat je určení a zjišťování možných zdrojů chyb, nepřesností, neúplností a porušených konzistencí. Při analýze zdrojových datových bází je proto nutné identifikovat nejen syntaxi dat (tabulky, atributy, klíče), ale i jejich sémantiku (význam jednotlivých atributů, kombinace různých příznaků apod.).

Největším zdrojem chyb a různých poruch v datech je obvykle provoz transakčních aplikací, resp. jejich využití koncovými uživateli. V souvislosti s tím existují následující případy, způsobující **obvyklé chyby v datech**:

- chyby, které vznikají při manuálních vstupech dat, např. prohození číslic, pravopisné chyby, špatně zadané kódy, hodnoty zapsané do nesprávného pole,

- data pocházející z různých zdrojů (různých zdrojových aplikací), s tím jsou spojeny problémy sjednocení číselníků, dodržení formálních standardů v datech apod.,
- neoprávněné zkoušení a testování aplikací uživateli na „ostrých“ datech,
- rozdíly v interpretaci dat, kdy různí uživatelé chápou sémantiku datové položky různě, což je dáno většinou nepřesným vymezením datových objektů, nebo nedostatky v dokumentaci,
- zásahy do zdrojových aplikací, kdy např. administrátor přidává do struktury záznamu novou datovou položku a zapomene to oznámit správcům ostatních aplikací, včetně BI,
- chyby přicházející z existujících externích datových zdrojů v důsledku špatně nastavených vazeb k externím datům,
- chyby vznikající v důsledku přímých vstupů do databází externími partnery (zákazníky, dodavateli) v rámci sdílených nebo otevřených aplikací pro vnější subjekty,
- chyby v návrhu databází, zejména nedodržovanou doménovou nebo referenční integritou,
- chyby v nedodržení relačních zásad – neexistují primární klíče, není vynucována povinnost atributů apod.,
- problémy s konverzí datových typů (zejména typu datum).

Přesné **specifikace a analýzy zdrojů chyb** jsou tak velmi důležitým předpokladem úspěšnosti řešení BI aplikací

## 5.2 Efekty BI

- Umožňují **lépe pochopit a analyzovat** podstatu vlastní obchodní a **manažerské činnosti**, proniknout do jejich hlubších a složitějších souvislostí, umožňují posun k multidimenzionálnímu řízení v reálném čase.
- Uplatnění časové dimenze nabízí **sledování vývojových trendů** z nejrůznějších pohledů.
- Analytická pravidla podle stanovených limitních hodnot jednotlivých ukazatelů a jejich dimenzí **umožňují upozorňovat na kritické nebo mimořádné stavy**.
- Možnosti rychle se pohybovat **na různé úrovni detailu informací** (agregačních úrovních) odpovídajících právě řešenému problému v řízení.
- Podpora **řešení skrytých problémů** na základě identifikace složitých závislostí mezi daty, odhalováním podobností mezi např. zákazníky, obchodními případy apod.
- **Kvalifikační efekty** – BI přirozenou cestou posiluje schopnosti manažerů a specialistů při řešení svých úloh „multidimenzionálně uvažovat“.
- Aplikace BI pro práci s ukazateli ve vazbě na podnikové procesy přinášejí i do manažerské sféry své specifické efekty a **racionalizují jak řízení jednotlivých dílčích úloh a procesů, tak celého podniku**.
- **Integrační efekty** podporou integrace dezintegrovaných informačních zdrojů (např. z věcného či geografického hlediska).

### 5.3 Omezení, problémy, předpoklady BI

- Úspěšnost BI řešení **silně závisí na zájmu uživatelů** a zejména vedení podniku BI implementovat.
- BI aplikace jsou velmi silně **závislé na kvalitě dat transakčních systémů**, z kterých data čerpají.
- BI řešení jsou velmi **náročná na konsolidaci** vstupních dat z různých datových zdrojů.
- Efekty BI jsou **obtížně měřitelné** a mají spíše měkký charakter.
- Pro úspěšnost BI je podstatné **zvolit správně přístup** k řešení a navrhnout správně architekturu.

### 5.4 Závěry k řešením Business Intelligence



- Nástroje **Business Intelligence**, BI představují již dlouhodobě **základ řešení** analytických a plánovacích aplikací a souvisejícího reportingu.
- Řešení aplikací Business Intelligence je charakteristické relativně vysokou individualizací, tj. aplikace jsou obvykle vytvářeny **podle konkrétních specifických potřeb jednotlivých manažerů** a specialistů firmy.
- Na trhu je mimořádně **silná nabídka produktů BI, včetně souvisejících služeb**. To vytváří i podstatně vyšší nároky na kvalifikovaný výběr těchto produktů a služeb na základě vyhodnocení reálných potřeb a možností firmy.
- Systémy BI jsou založené **na architekturách, zahrnujících celou škálu komponent** různých typů a určení (datový sklad, datová tržiště, ETL a další). Návrh takové architektury musí vycházet z rozsáhlé **analýzy stavu a potřeb řízení** jak celé firmy, tak řízení a architektury IT.
- Přehled a **základní charakteristiky komponent** obsahuje následující text, kde pro každou komponentu jsou uvedeny obvyklé **efekty a případné problémy**. Ty je třeba při návrhu architektury BI kvalitně konkretizovat a analyzovat pro další návrhy.
- Systémy IT ve firmě, jak zdrojové, tak BI, procházejí trvale **vývojem a řadou změn**. To při provázanosti BI na zdrojové databáze a aplikace vytváří vysokou **potřebu kvalitního řízení, právě na rozhraní zdrojů a BI**.
- BI produkty poskytují, kromě vlastního uložení dat a analytiky **množství nástrojů** pro zajištění kvality dat, správu metadat i pro pokročilou analytiku.
- BI systémy prošly vývojem směřujícím **ke konvergenci nástrojů**, tj. současné databázové systémy zahrnují celou funkcionalitu pro BI. Uživatelům tak odpadají časté starosti se zajištěním integrace mezi různými nástroji. I tak na druhé straně existuje nabídka specializovaných nástrojů, např. pro ETL.
- Užití BI aplikací vytváří u manažerů dobré předpoklady pro **zvysování manažerské kvalifikace**, nutí je přirozenou cestou uvažovat o problémech „**multi-dimenzionálně**“.
- Pro řešení úloh BI je účelné definovat **sadu analytických otázek**, které by měly být předmětem řešení a diskusí mezi manažery a analytiky.

## 6. Self-Service Business Intelligence, SSBI



**Účelem** kapitoly je:

- vymezit hlavní **charakteristiky aplikací a produktů SSBI**, jejich pozitiva a problémy,
- vytvořit základní podklad pro **posouzení jejich možností** při jejich návrzích samotnými uživateli.

Samoobslužné BI aplikace a řešení (Self-Service BI) představují **jeden z nejnvýraznějších trendů** v oblasti Business Intelligence.

### 6.1 Podstatné charakteristiky SSBI

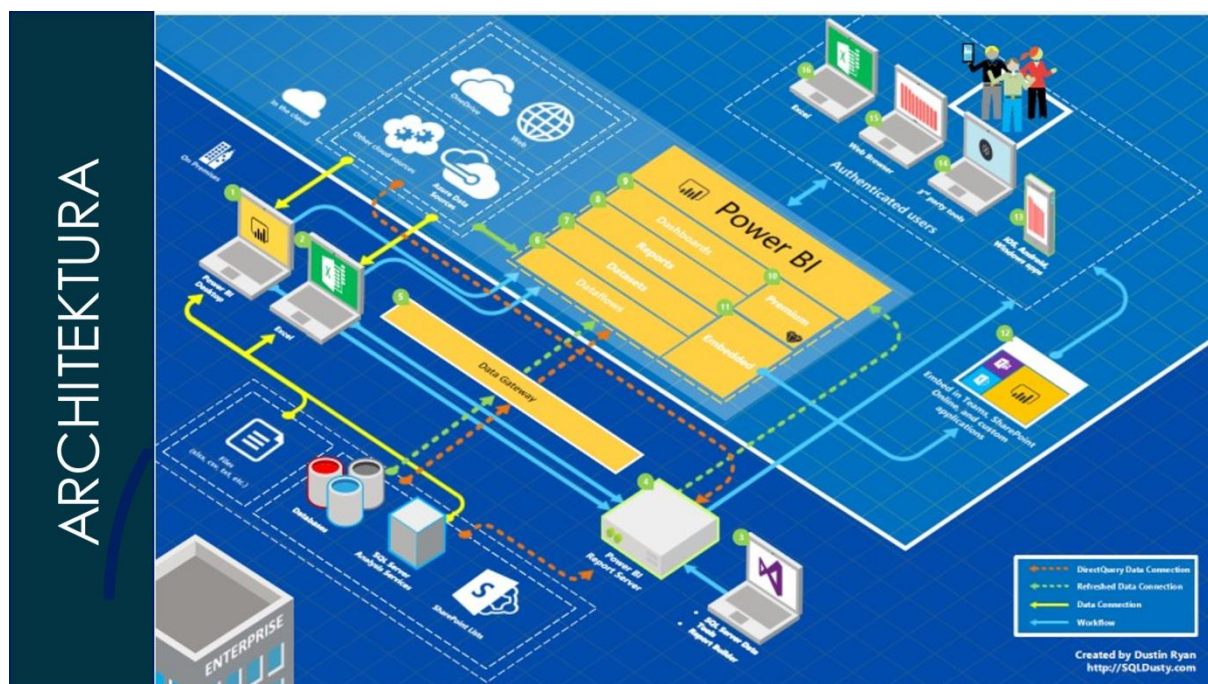
**Smyslem SSBI je** na základě nových technologií poskytnout uživatelům prostředí pro realizaci svých analytických úloh bez nutnosti využívání komplexních a obvykle velmi složitých systémů BI.

Samoobslužné BI umožňují např. realizovat **multidimenzionální uložení a zpracování dat**, nabízejí efektivní a **jednoduché přístupy k datům**, poskytují prostředky, např. **jazyk DAX**, pro výpočty a další operace v dimenzionálním prostředí apod. Kromě toho jsou tyto aplikace i **vhodným prostředkem pro pochopení podstaty a způsobu využití větších Business Intelligence systémů**.

Self-Service BI **se snaží vyřešit rozpor mezi dvěma protichůdnými podnikovými silami** – potřebou **flexibility a svobody** při analýze a zkoumání dat koncovými „spotřebiteli“ BI výstupů, které ovšem bohužel nelze docílit s pomocí „dodavatele“ ve formě podnikového IT oddělení. Proti pak stojí **potřeba IT oddělení mít neustálou kontrolu nad daty** vytvářením a distribucí informací uvnitř firmy.

SSBI rozšiřuje klasické tradiční BI prostředí o **možnosti provádění vlastních analýz** nad zpřístupněnými daty a jejich reporting bez nutnosti potřeby zásahu IT oddělení. Dává ne-IT pracovníkům **více možností, větší flexibilitu a větší samostatnost**, kdy je možné získat odpověď na danou otázku ve výrazně kratší době. Přístup SSBI tedy obchází **problém s neustále se měnícími a novými požadavky** a potřebami uživatelů, kdy každý může získat potřebnou informaci kdykoliv si zamane. Podnikové IT se tak stává do zcela nové role.

Typickým **představitelem** produktů SSBI je **Power BI** společnosti Microsoft, který v současné době patří i nejrozšířenějším. Schéma jeho architektury představuje Obrázek 6-1:



Obrázek 6-1: Architektura Power BI (Zdroj: Seyfor)

## 6.2 Efekty SSBI

- Podstatně se zkracuje **doba potřebná na implementaci** analytických aplikací.
- Analýza dat **samotnými byznys uživateli** a dodání požadovaných výstupů/informací je rychlejší. Je možné využít i jiná data než z podnikového datového skladu.
- Self-Service BI aplikace nabízejí obdobnou **flexibilitu a výkonnost**, jako je tomu většinou u aplikací, založených na OLAP databázích.
- Základní **příprava aplikací** včetně transformací zdrojových dat je výrazně zjednodušená, takže je dostupná i ne IT pracovníkům, samozřejmě po nezbytném zaškolení.
- Součástí technologií jsou i **programovací prostředky**, jako např. jazyk DAX (Data Analysis Expressions) pro Microsoft Power BI.
- **Finanční, zdrojová i provozní náročnost** Self-Service BI je oproti standardním projektům a provozovaným aplikacím výrazně nižší.
- Self-Service poskytuje i velmi dobrou podporu pro tvorbu **analytických i vizuálně náročnějších aplikací**, jako např. dashboardů, klikovacích map apod.
- **Uživatelé dostávají do rukou data a přístup k nim.** Data si mohou kdykoliv podle svých aktuálních potřeb zanalyzovat a vyhodnotit za pomoci poskytnutých nástrojů. Nemusí tak dlouho čekat na výstupy, které jim dodá pomalé IT na základě jejich požadavku.
- Namísto toho, aby **IT oddělení** bylo zahlceno stále novými a měnícími se požadavky na reporty a analýzy, předává „moc“ práce s daty do rukou samotných uživatelů. Zároveň **se může soustředit na řízení a aktivity, které přinášejí společnosti větší hodnotu.**

- Změna vede **k celkové větší spokojenosti uživatelů s IT oddělením** v organizaci, které se v jejich očích přetváří z podoby úzkého hrdla v rychlosti dodávání BI výstupů, do role flexibilního pomocníka, který lépe dokáže uspokojovat uživatelské byznys požadavky.

### 6.3 Omezení, problémy, předpoklady SSBI

- Self-Service BI řešení se mohou realizovat **pouze pro úlohy určité kategorie**, tedy relativně jednodušší a izolovanější úlohy.
- Self-Service BI má **omezené možnosti čištění a konsolidace dat**, které u standardních BI aplikací představují jednu z pracovně nejnáročnějších, ale i jejich finálně nejefektivnějších částí.
- Problémem je dosažení **integrace dat** v rámci podniku, které nabízejí celopodnikové datové sklady.
- S předchozí poznámkou souvisejí i podstatně omezené **možnosti celopodnikového reportingu**, např. centrálního reportingu v rámci nadnárodních společností apod.
- I když je příprava analytických aplikací na bázi Self-Service BI uživatelsky výrazně dostupnější, přesto je nutné, aby pro kvalitnější aplikace **uživatelé ovládali alespoň základní principy analytických metod**, tj. především dimenzionálního a datového modelování.
- Je zde zvýšené riziko **špatných rozhodnutí** na základě „špatných dat“ od byznys uživatelů.
- Mění se **pozice IT** z producenta výstupů do pozice tzv. „supervizora“. Je na něj tedy kladena tíha řízení a kontroly celého prostředí a produkováných výstupů.
- Pouhou instalací SSBI nástrojů se požadované přínosy nezískají. Je potřeba, **aby uživatelé pochopili přínos** a možnosti daných nástrojů a začali je využívat.

### 6.4 Závěry k řešení Self-Service Business Intelligence



- Vznik Self-Service Business Intelligence, SSBI, znamenal významný mezník v rozvoji podnikové analytiky. Vytvořil **předpoklady pro samostatnou práci uživatelů** v návrhu a využití analytických a plánovacích aplikací.
- Návrhy SSBI **na straně uživatelů** vytvářejí nároky na **zvyšování znalostí**, nejen pokud jde o provozované produkty, ale i **metody**, uplatňované při analýze a návrzích, zejména multidimenzionálního modelování a datového modelování.
- Nasazení aplikací SSBI se **liši obvykle podle velikosti a potřeb firm**. U menších firem může pokrývat jejich potřeby v celém rozsahu a nahradit komplexní BI systémy. U větších firem tomu tak často není a aplikace SSBI se stávají doplňkem komplexních BI aplikací, poskytující uživatelům větší operativnost při jejich práci.



- Na trhu je již adekvátní **nabídka produktů SSBI i souvisejících služeb**. To vytváří i odpovídající nároky na kvalifikovaný výběr těchto produktů a služeb.
- Současné kvalitní SSBI produkty poskytují **možnosti vstupů z nejrůznějších zdrojů**, tedy databází, internetových zdrojů, typů souborů atd.
- Systémy IT ve firmě, jak zdrojové, tak BI, procházejí trvale **vývojem a řadou změn**. To při provázanosti SSBI na zdrojové databáze a aplikace vytváří vysokou **potřebu kvalitního řízení, právě na rozhraní zdrojů a SSBI**.
- I pro řešení úloh SSBI je účelné definovat **sadu analytických otázek**, které by měly být předmětem řešení a diskusí mezi manažery, analytiky a uživateli.

## 7. Mobilní BI



**Účelem** kapitoly je:

- vymezit základní **charakteristiku mobilních BI aplikací**, jejich pozitiva a problémy,
- vytvořit podklad pro **analýzu jejich možností** při analytické podpoře úloh zejména transakčního charakteru, např. řízení prodeje, nákupu, skladů, majetku.

Mobilní Business Intelligence slouží k zobrazování a **práci s BI nástroji na mobilních telefonech** (smartphones) a tabletech.

### 7.1 Podstatné charakteristiky mobilního BI

Typickým uživatelským scénářem mobilního BI je **zobrazení dashboardů, reportů** a možnost s nimi pracovat a využívat základní analytické funkce. Vede to **k efektivnímu využití informace a možnosti flexibilního rozhodování**, založeného na relevantních informacích. **Cílem mobilního BI** není nabízet kompletní funkcionalitu jako desktopové verze, ale využít dostupné aktuální informace k okamžitému rozhodování (Ševčík, 2020).

„*Mobilní BI je trendem, kdy byznys uživatelé přistupují ke svým datům a dashboardům pomocí mobilních telefonů a tabletů*“ [Logi Analytics, 2020]. Představuje **jednoduchý přístup** k firemním datům, reportům, klíčovým ukazatelům výkonu či dashboardům pomocí mobilních zařízení. Zároveň některé aplikace umožňují **sdílení získaných dat** s kolegy, okomentování dashboardů pro zbytek týmu či dotazování dat v přirozeném jazyce [MICROSOFT, 2020].

K těmto datům lze **přistupovat** v rámci mobilního BI řešení pomocí **aplikace stažené přímo v konkrétním zařízení**. Tyto aplikace by měly být navrženy speciálně pro tato zařízení a práce s nimi by měla být pro uživatele jednoduchá a poměrně intuitivní. Správně vytvořené aplikace mobilního BI si přizpůsobí zobrazení dashboardu tak, **aby byl efektivně využit prostor**, ale zároveň bylo vše pro uživatele čitelné a srozumitelné.

Druhou možností je **přístup přes internetový prohlížeč** v mobilním zařízení, pomocí kterého se lze dostat na webovou aplikaci, kterou je třeba zpřístupnit. To umožňuje používat **stejné funkce jak na počítači, tak i na mobilu** [Klipfolio, 2020]. Problém tohoto řešení může být v tom, že pokud se přistupuje k datům pomocí **webového prohlížeče**, může se stát, že **grafy nebudou vypadat správně**, popisky nebudou sedět, budou nečitelné, nebo se jen bude s webovou aplikací špatně pracovat, protože byla primárně navrhována pro počítačové využití. Webová aplikace musí tak být **kompatibilní i s mobilním prohlížečem**, a ne pouze s počítačovým prohlížečem [Logi Analytics, 2020].

Velké úskalí mobilního BI se podle společnosti BARC nachází v nutnosti zabezpečení mobilního zařízení i samotné aplikace. **Bezpečnostní opatření** by podle mnohých ex-

pertů a společností měla být na vysoké úrovni kvůli citlivosti dat, se kterými se v aplikaci pracuje a ke kterým by při ztrátě mobilního zařízení mohl získat kdokoli přístup [BARC, 2020].

## 7.2 Efekty mobilního BI

- Existují možnosti **distribuce obsahu** do další početné množiny zařízení.
- **Dostupnost** informací umožňuje využití hodnoty informace okamžitě po jejím získání. Jinak hodnota informace po jejím získání rychle klesá, čím dřív je využita, tím větší efekt přinese.
- Nabízí se vysoká **jednoduchost ovládní** a popisnost poskytnutých dat.
- Umožňuje využití **funkcí**, které **smartphone nebo tablet** nabízí – GPS senzor, kamera, WiFi.
- Mobilní BI je v plném souladu s cloud paradigmatem – **pracovat odkudkoli**, kde je dostupný internet.

## 7.3 Omezení, problémy, předpoklady mobilního BI

- Návrh uživatelsky **přívětivého rozhraní** znamená nutnou optimalizaci pro smartphones a tablety. Uživatelské rozhraní a design aplikace se zcela liší od desktopové verze – jsou použity jiné technologie interakce (dotyková obrazovka) a jiné uživatelské scénáře.
- Zvyšují se nároky na **bezpečnost** – aplikace pracuje s citlivými podnikovými daty. Organizace musí řešit různé úrovně zabezpečení a kontrol přístupu.
- Je obvykle nutné vyvíjet aplikaci **pro každou platformu zvlášť** – Android, iOS, Windows Phone.
- **Internetová služba**, založená na standardu HTML5, řeší předchozí bod a nabízí multiplatformní řešení, avšak HTML5 hrozí rychlé zastarání z důvodu nízké flexibility.

## 7.4 Závěry k řešení mobilního Business Intelligence



- Nasazení aplikací mobilních BI se **liší obvykle podle charakteru firmy**. Na významu nabývají tam, kde jde o vysokou dislokaci jednotek a aktivit uživatelů, např. v řízení prodeje, nákupu, detašovaných skladů, v řízení investic apod.
- Na trhu je již adekvátní **nabídka produktů mobilních BI**. To vytváří i odpovídající nároky na kvalifikovaný výběr těchto produktů a služeb.
- I pro řešení úloh mobilních BI je účelné definovat **sadu analytických otázek**, které by měly být předmětem řešení a diskusí mezi manažery, analytiky a uživateli.

## 8. Competitive Intelligence, CI



**Účelem** kapitoly je:

- vymezit základní **charakteristiku aplikací Competitive Intelligence**, jejich pozitiva a problémy,
- vytvořit podklad pro **analýzu jejich možností** při řešení úloh strategického řízení firmy, marketingu, nebo řízení prodeje.

Competitive Intelligence (CI) je **systematická a etická disciplína**, zabývající se sběrem, zpracováním a ochranou informací **s cílem získat konkurenční výhodu** a zlepšit tak celkové postavení podniku na trhu. K tomu využívá jak **primárních informačních zdrojů** – lidí, tak **sekundárních artefaktů** – dokumentů, audiovizuálních záznamů, tiskových zpráv, dat ze sociálních sítí atp.

### 8.1 Podstatné charakteristiky CI

CI pracuje výhradně **s legálními informačními zdroji**, nejedná se o průmyslovou špionáž. Disciplína má primární **vazbu na strategické řízení** podniku a podnikové IT, ale **podporuje analytickou, plánovací a rozhodovací činnost organizace na všech úrovních** a ve všech oblastech podnikového řízení, tj. prodeje, nákupu, marketingu, finančního řízení, controllingu, majetku, řízení lidských zdrojů, výroby a dalších.

**Konkurenční zpravodajství (CI, Competitive Intelligence)** je, na rozdíl od průmyslové špionáže, **etická a legální činnost**, využívající veřejně dostupných zdrojů s cílem zvýšit konkurenceschopnost společnosti za pomoci podpory rozhodování, analýzy trhu, identifikace rizik a příležitostí, ať už současných, nebo budoucích, a to v nekončícím systematickém procesu.

### 8.2 Efekty CI

- **Tradiční forma „aktivního CI“** se primárně zaměřuje na **externí informační zdroje a tržní okolí** podniku (odběratelé, dodavatelé, partneři, konkurence, legislativní rámec, profesní a zájmové komunity, zpravodajské portály, sociální sítě apod.) s tím, že **racionálně umožňuje**:
  - zmapovat tržní prostředí a aktivity konkurence,
  - zhodnotit pozici firmy vůči konkurentům,
  - odhalit případné vnější hrozby,
  - identifikovat možné příležitosti pro další růst a rozvoj.
- **Doménou tzv. „pasivního CI“ (counter-CI)** je pak **ochrana vlastních interních zdrojů** před CI aktivitami konkurence a dále snaha zmapovat a vytěžit tyto interní zdroje. Nabízí tyto možnosti:
  - omezit efekt CI aktivit konkurence,
  - upevnit / zlepšit stávající pozici firmy na trhu,

- identifikovat klíčové nositele znalosti z řad zaměstnanců,
- odhalit a využít možný skrytý potenciál uvnitř vlastní firmy.
- Podporuje **strategické plánování a řízení** firmy díky aktuálním informacím o stavu trhu a aktivitách konkurence.
- **Zkracuje reakční doby** v případě nenadálé události či hrozby díky její včasné identifikaci.
- **Proaktivně vyhledává příležitosti pro inovace** a růst a tím i zlepšování pozice firmy na trhu.

### 8.3 Omezení, problémy, předpoklady CI

- **Efekty CI jsou obtížně měřitelné** a mají spíše měkký / podpůrný charakter.
- Aktivity CI **nelze plně automatizovat**, je nutné je neustále přizpůsobovat okolnostem a aktuálním podmínkám.
- **Úspěšnost CI** řešení silně závisí na použitých informačních zdrojích – dostupnost kvalitních a spolehlivých informačních zdrojů je zcela klíčová pro efektivní CI.
- CI přináší **největší efekt v dlouhodobém horizontu** a vyžaduje trvalou investici do zdrojů (lidských i finančních).
- Pro úspěšnost CI je podstatné zvolit **vhodný přístup k celému řešení a navrhnout správně architekturu** s ohledem na další části firmy a dostupné informační zdroje – a to jak koncepčně, tak i z hlediska samotné použité technologie.

### 8.4 Závěry k řešení Competitive Intelligence



- V rámci Competitive Intelligence se stále uplatňují především **individuální řešení na míru** marketingovým a obchodním specialistům a manažerům.
- Analýza a návrh aplikací CI musí vycházet z detailního pochopení **aktuálních i budoucích obchodních aktivit a potřeb firmy**.
- S ohledem na individuální charakter CI je prakticky **nezbytná účast manažerů a specialistů firmy** na jejich řešení.
- Analytická příprava aplikací je zde vesměs založena na **specifikaci externích zdrojů a jejich charakteru a kvality**.
- Pro řešení úloh CI je rovněž účelné definovat **sadu analytických otázek**, které by měly být předmětem řešení a diskusí mezi manažery a analytiky.

## 9. Postup řešení podnikové analytiky na příkladu Business Intelligence



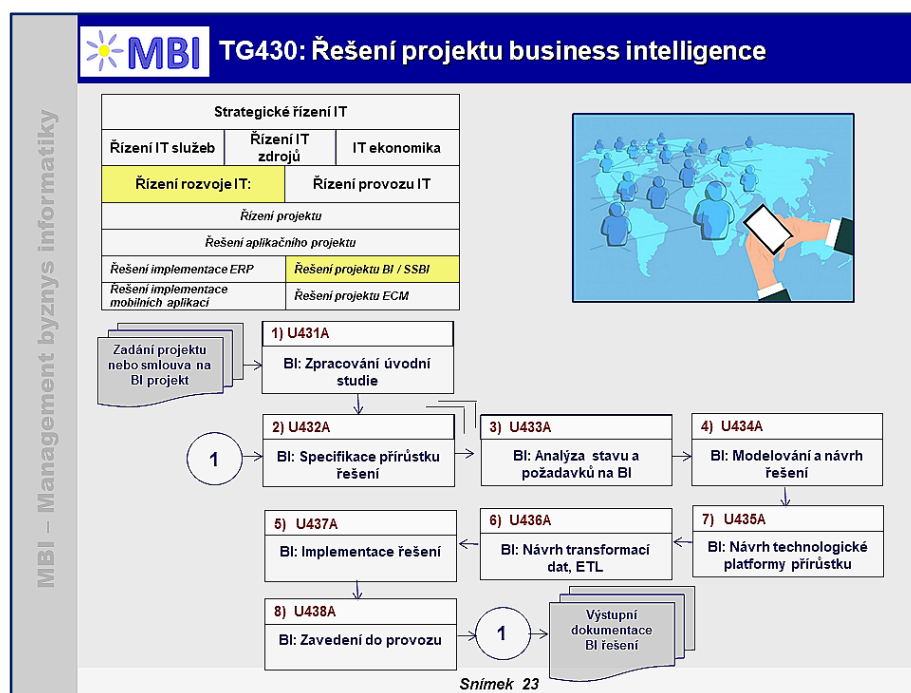
Předpokládejme, že projekt je řešen **převážně dodavatelským způsobem**, ale **s vysokou účastí zákazníka a uživatelské sféry**. Úlohy jsou postaveny na specificky zaměřených metodikách a některých metodách, odpovídajících BI technologiím a BI funkcionalitě. **Účelem** je:

- **zvýšit kvalitu analytických a plánovacích činností** v rámci podnikového řízení,
- řešit **aplikace respektující principy řešení business intelligence** – zejména s orientací na analytické a plánovací úlohy, multidimenzionalitu uložení dat atd.,
- poskytnout uživatelům **efektivní prostředí analytických a plánovacích úloh** a zvýšit jejich dostupnost a flexibilitu pro uživatele,
- **snížit závislost uživatelské sféry na kapacitách IT útvarů** a současně redukovat zatížení pracovníků IT úlohami, které mohou samostatně realizovat uživatelé.

V případě dalšího zde prezentovaného postupu se vychází z tzv. **přírůstkového přístupu k řešení**.

### 9.1 Přehled a obsah úloh řešení projektu Business Intelligence

Celkový přehled úloh řešení projektu Business Intelligence dokumentuje Obrázek 9-1.



Obrázek 9-1: Řešení projektu Business Intelligence, přehled úloh

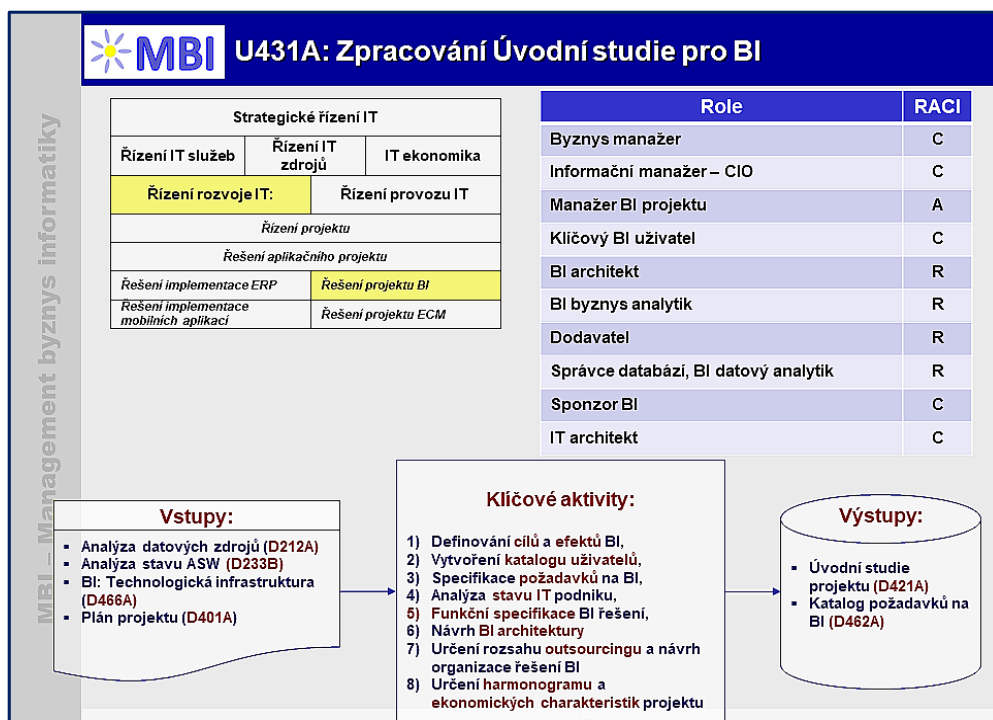
Do řešení BI spadají **tyto úlohy**:

- **Zpracování Úvodní studie**, která zejména definuje priority řešení BI pro jednotlivé oblasti řízení podniku (finance, prodej atd.) a jim odpovídající přírůstky řešení a obvykle i jejich pořadí v řešení projektu.
- **Specifikace přírůstku** řešení, tj. oblasti, kterou pokrývá, jeho obsahu, resp. funkcionality, datových zdrojů.
- **Analýza stavu**, zejména zdrojových databází a aktuálních uživatelských požadavků na BI řešení.
- **Modelování a návrh** řešení, tj. zejména zpracování hrubého dimenzionálního modelu a následně datového modelu datového skladu a tržiště.
- **Návrh technologické platformy**, pokud se bude měnit nebo upravovat oproti již realizované platformě.
- **Návrh transformací dat – ETL / ELT** a s tím spojené kontroly dat a nároky na jejich úpravy, resp. čištění a konsolidaci.
- **Implementace řešení**, tj. realizace datových skladů, reportingu a analytických a plánovacích aplikací.
- **Zavedení do provozu**, migrace dat a příprava uživatelů.

Další paragrafy obsahují **přehled úloh a jejich stručný obsah**.

### 9.1.1 Zpracování Úvodní studie pro BI

**Cílem úvodní studie** BI je zajistit **komplexní zmapování prostředí**, do něhož má být BI řešení zasazeno, určit **přístup k řešení BI, určit priority oblastí** řešení, resp. podnikových procesů, které má BI pokrývat, vytvořit celkovou koncepci BI řešení, viz Obrázek 9-2.



**Obrázek 9-2: Zpracování Úvodní studie pro BI**

### 9.1.1.1 Obsah úvodní studie

Úloha **zahrnuje definování cílů a efektů BI**, vytvoření katalogu uživatelů, specifikace požadavků na BI, analýzu stavu podnikové informatiky, funkční specifikaci BI řešení, návrh BI architektury, návrh organizace řešení BI a určení harmonogramu a ekonomických charakteristik projektu.

V dalších kapitolách **se předpokládá přírůstkový přístup**. Od návrhu architektury se pak odvíjejí návrhy jednotlivých dílčích přírůstků BI řešení, jejich vzájemných vazeb, případné dopady do ostatních aplikačních software, např. ERP.

#### **Možnosti řešení BI projektů**

Způsob řešení BI projektů ovlivňuje i **případné využití referenčních modelů nebo typových BI aplikací**. V tomto případě existují tyto **možnosti**:

- využití **předem připravených, referenčních dimenzionálních modelů** pro určité typy podniků nebo odvětví ekonomiky (bez softwarové realizace),
- využití **předem připravených typových aplikací BI** rovněž pro určité typy podniků, přičemž takové typové aplikace většinou nabízejí i systémy ERP,
- **vývoj aplikací zcela „na míru“** vlastními nebo externími kapacitami s využitím vývojových nástrojů, specializovaných pro tyto aplikace.

Všechny uvedené možnosti více či méně **ovlivňují postup řešení celého projektu**, u některých fází nebo činností evidentně redukuje jejich rozsah, nebo naopak posilují jejich význam. Úvodní studie tak postihuje celý rozsah BI projektu, zatímco **další fáze** se již vztahují pouze **k jednotlivým přírůstkům**.

### 9.1.1.2 Klíčové aktivity

#### **Definování cílů a efektů BI**

Definování cílů a efektů BI musí **vycházet ze strategických cílů firmy** a od nich odvozených cílů informatiky, tedy cílů rozvoje celého informačního systému firmy. **Efekty BI** nemusí být vždy zcela přesně kvantifikovatelné, resp. ve finančním vyjádření. Často se v těchto případech jedná o **kvalitativní efekty**, znamenající dosažení vyšší konkurenceschopnosti firmy, získání lepší pozice na trhu, poskytování kvalitnějších informačních služeb obchodním partnerům atd.

Navíc má dosažení těchto efektů **delší časový horizont** způsobený potřebnou dobou na vytvoření a naplnění datového skladu, vytvoření časových řad sledovaných ukazatelů, osvojení si náročnějších analytických aplikací uživateli apod.

#### **Vytvoření katalogu uživatelů**

Vytvoření katalogu uživatelů, tj. manažerů, finančních, výrobních, marketingových analytiků a dalších specialistů, většinou s nestandardními informačními požadavky. Úvodní studie proto zahrnuje vytvoření **katalogu uživatelů, jejich profesní orientace a jejich počtů**, jako vstup pro následující zjišťování a analýzy byznys požadavků.



## **Specifikace požadavků na BI**

Specifikace požadavků na BI definuje potřeby firemních procesů a požadavků uživatelů na analytické a plánovací funkce informatiky, tedy BI. Výsledkem je **mapa vybraných procesů a činností řízení firmy a jim přiřazené nároky** na funkcionalitu BI.

Podstatnou charakteristikou řešení v této fázi je i **určení priorit procesů** vzhledem k řešení BI, a tedy **prvotní určení priorit** jednotlivých přírůstků řešení. Na této úrovni jde o rámcově definované požadavky, jejichž zpřesnění bude náplní dalších analytických fází.

## **Analýza stavu IT podniku**

Analýza stavu IT firmy zahrnuje hodnocení jejich aktuálního stavu, tedy **SW a HW zdrojů, všech dostupných interních i externích datových bází**, jejich validity vzhledem k požadavkům BI, technické, ekonomické a organizační dosažitelnosti. **Otázka kvality aplikací ve firmě, zejména transakčních**, je v souvislosti s BI posuzována v několika úhlech pohledu:

- do jaké míry jsou tyto **aplikace schopné poskytovat úplná, konzistentní a přesná data**, tj. jak je navržena jejich datová základna, jaký systém kontrol zahrnuje jejich funkcionalita, **jak odpovídají potřebám podniku** z pohledu poskytovaných funkcí i vytvářených a zpracovávaných dat,
- zda **zahrnují vlastní analytickou a plánovací funkcionalitu**, do jaké míry je využívána, zda je účelné ji nahradit funkcionalitou BI aplikací,
- zda v sobě **integrují BI technologie a BI aplikace**, jako je tomu např. u systémů ERP, u CRM, nebo u aplikací APS / SCM pro analýzy a plánování provozu a rozvoje dodavatelských řetězců.

## **Funkční specifikace BI řešení**

Funkční specifikace **definuje předmět, rozsah řešení a jeho funkční specifikaci**, a to pouze na hrubé úrovni. K funkcím nebo skupinám funkcí jsou **přiřazeny jejich priority**. Výsledkem je tak struktura požadovaných funkcí jejich priorit a podklad pro návrh BI architektury.

## **Návrh BI architektury**

Návrh BI architektury představuje **specifikaci hlavních BI aplikací, datového skladu, datových tržišť, případně data lake a dalších komponent** a všech podstatných vazeb mezi nimi. Při volbě architektury celého BI je třeba vybrat adekvátní kombinaci aplikačních a technologických komponent, jež by měly tvořit celkové řešení BI. Ta ovlivňuje základní **parametry úspěšného řešení, finanční stránku, vnímání řešení managementem organizace a koncovými uživateli i časovou náročnost** jednotlivých implementačních kroků. Při návrhu architektury se musí **respektovat celá řada klíčových požadavků**, zejména:

- **otevřenost** – architektura musí podporovat připojení nových řešení BI, ale i nových systémů či zapojení nových dodavatelů. Současně musí být schopna pojmout organizační i procesní změny,

- **škálovatelnost** – architektura musí umožnit libovolně rozšiřovat řešení jak po věcné, tak technologické stránce,
- **schopnost integrace na ostatní produkty a projekty**, výstupy z řešení BI musí být schopné integrovat do jiných řešení organizace (data musí být např. využitelná v aplikacích pro podporu kontaktních center, marketingu, prodeje apod.),
- **jednoduchost (transparentnost)** ve zvládnutí poměru komplexnost a šíře funkcí versus jednoduchost řízení a manipulace s aplikacemi,
- **výkonnost, požadovaná funkcionalita** a další.

### **Určení rozsahu outsourcingu a návrh organizace řešení BI**

Určení rozsahu outsourcingu se vztahuje k implementaci a provozu BI řešení, zahrnuje **určení podílu vlastních řešitelských kapacit a dodavatelské společnosti a návrh způsobu kooperace**. S tím se váže i návrh využití specifických modelů zajištění implementačních a provozních kapacit, zejména **využití cloud computingu, např. modelu SaaS**.

Návrh organizace řešení BI zahrnuje **specifikace jednotlivých rolí podílejících se na projektování a provozu BI**, organizačních struktur s jasně vymezenými zodpovědnostmi a kompetencemi. Návrh organizace obsahuje i vymezení projektových a provozních standardů a principy kooperace dodavatele, interních analytiků a uživatelů.

### **Určení harmonogramu a ekonomických charakteristik projektu**

Určení harmonogramu a ekonomických charakteristik projektu je **postaveno na rámcové specifikaci jednotlivých přírůstků a termínů** jejich řešení. Zahrnuje rámcové určení výše nákladů na projekt, resp. jednotlivé přírůstky a odhadované finanční přínosy.

#### **9.1.1.3 Podmínky úspěšnosti**

- Úvodní studie vyjadřuje **celkovou koncepci** řešení BI ve společnosti a má proto **na jeho kvalitu zásadní vliv**, zejména je to promyšlená a otevřená architektura, kterou lze postupně a systematicky naplňovat jednotlivými přírůstky.
- Předpokladem úspěšné Úvodní studie i celého projektu BI je **zájem vedení společnosti** o aplikace tohoto typu. Ty jsou určeny primárně pro manažersky nebo analyticky orientované uživatele, i když se s ohledem na jejich stále vyšší dostupnost předpokládá jejich užití i na středních a nižších úrovních řízení. Z toho vyplývá, že **kvalita jejich řešení** a zejména užití není primárně dána předpisy, metodikami, resp. disciplínou pracovníků, ale **zájmem, motivací a invencí na uživatelské i dodavatelské straně**.
- Jedním z klíčových předpokladů pro úspěšné řešení a využití BI aplikací je **existence jejich potřeby z pohledu cílové skupiny uživatelů**, tj. manažerů, podnikových analytiků a specialistů. Tato potřeba je buď dána čistě odborným zájmem jednotlivců, nebo zájmem vedení společnosti na jejím celkovém úspěchu. Pokud vedení podniku nepovažuje BI aplikace za účelné, nebo je systém řízení firmy založen spíše na citu a zkušenostech vedoucích pracovníků, pak je lepší BI projekty nezahajovat, nebo je přesunout na pozdější období podle vývoje situace.

- **Pravidlo silného sponzora;** s ohledem na konečný úspěch by projekty BI měly být uvnitř firmy vždy podporovány osobnostmi se značnou mírou vlivu a s nezbytnými rozhodovacími pravomocemi.
- Úvodní studie **má jasně definovat priority podnikových procesů**, které bude BI řešení pokrývat a od nichž se odvíjejí i priority jednotlivých přírůstků.
- Hned na počátku řešení je účelné **jasně specifikovat očekávané efekty** řešení pro podnik a jednotlivé uživatele. Určování a posuzování BI efektů je v porovnání s ostatními typy aplikací poněkud specifické. V každém případě je vymezení očekávaných efektů a sledování jejich naplnění podstatné s ohledem na to, že **analytické přípravě a využití těchto aplikací musí věnovat čas na uživatelské straně manažerů** a podnikoví specialisté, jejichž časové možnosti jsou vesměs omezené. Musí proto, pokud možno, přesně vědět, co jim takto vynaložená časová i finanční investice přinese.
- Již v rámci Úvodní studie je třeba jasně **definovat zasazení řízení BI rozvoje a provozu do celého řízení informatiky**. Zejména je nezbytné nastavit **systém řízení změn** v produkčních systémech a databázích s okamžitou informací pro správu BI.
- **Využití konceptu řízení výkonnosti** a zejména řízení podnikové výkonnosti představuje díky své komplexnosti a provázanosti jednotlivých komponent a BI aplikací faktor úspěchu nejen BI samotného, ale i celého řízení firmy.

### 9.1.2 Specifikace přírůstku řešení

**Cílem** definování přírůstku BI je vytvořit **konkrétní zadání** pro analýzu a implementaci dalšího přírůstku, tedy rozšíření stávajících řešení a aplikací BI.

#### 9.1.2.1 Obsah definování přírůstku BI

Přírůstkový přístup je v současné praxi využíván často. Obsahuje vyhodnocení aktuálního stavu BI, **definování přírůstku BI řešení a určení harmonogramu** a ekonomiky navrženého přírůstku. Po celkové verifikaci návrhu zadání přírůstku a odsouhlasení kompetentními pracovníky je **zadávací dokumentace zkompletována** a slouží jako vstup do dalších fází řešení.

#### 9.1.2.2 Klíčové aktivity

- **Vyhodnocení aktuálního stavu** řešení BI,
- Specifikace **nových nebo změněných uživatelských požadavků** oproti Úvodní studii nebo předchozím přírůstkům řešení a vyhodnocení aktuálního stavu provozu datového skladu, tržišť, ETL atd. podle provozní dokumentace **s určením klíčových problémů a omezení**, nároků na aplikační nebo technologický rozvoj,
- **Definování přírůstku**, vytvoření zadávací dokumentace přírůstku BI, tj. identifikační a organizační údaje, jako např. cílová skupina uživatelů, sponzor, cíl přírůstku, očekávané efekty, základní funkcionalita.

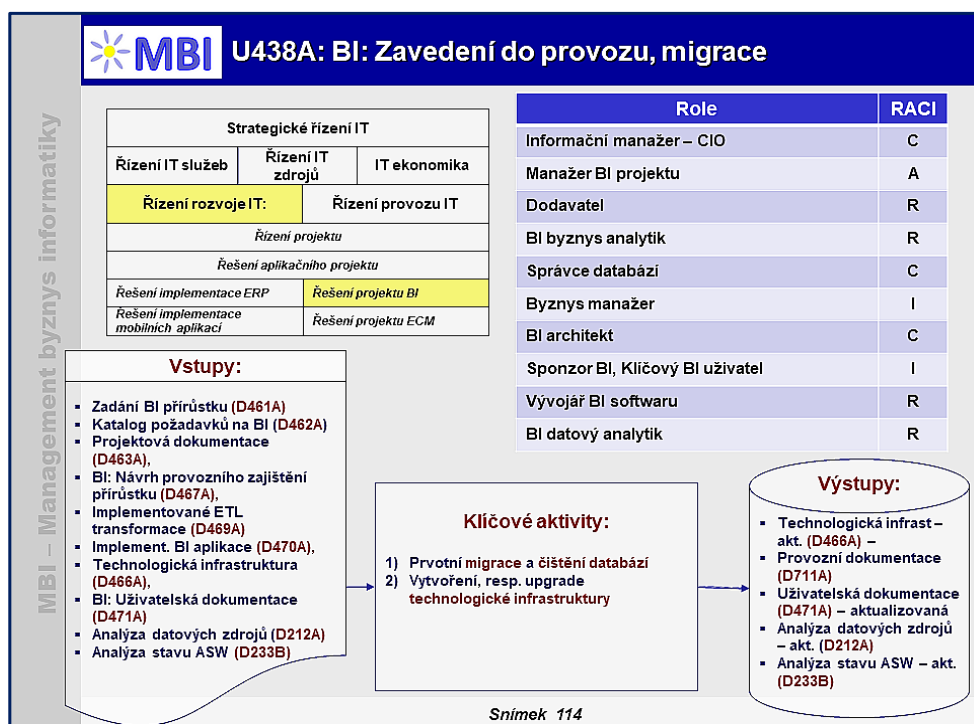
- Určení **harmonogramu a ekonomiky** přírůstku, stanovení především jeho nákladové stránky a celého plánovaného harmonogramu řešení podle jednotlivých etap a termínů.

### 9.1.2.3 Podmínky úspěšnosti

- Pro reálné definování přírůstku BI je předpokladem **velmi dobrá a kompletní znalost aktuálních zdrojů**, tj. zdrojových databází a aplikací, zejména jejich kvality a dostupnosti.
- Priority a **harmonogram přírůstků** je nezbytné **odsouhlasit s vedením podniku**, případně s vlastníky.
- Předpokladem správné definice přírůstku je **kvalitní dokumentace současného stavu** datového skladu, tržišť a ETL procedur.
- Při dodavatelském způsobu řešení je otázkou, zda **ponechat řešení dalšího přírůstku na původním dodavateli**, nebo vybírat nového. Pokud je s původním dodavatelem alespoň přiměřená spokojenost, pak je lepší zůstat u původního.
- Při definování přírůstku je účelné **posoudit možný vývoj nebo změny v produkčních systémech**, případně i v dalších aplikacích s vazbami na BI (portály apod.).

### 9.1.3 Analýza stavu a požadavků na BI

**Cílem** analýzy požadavků a stavu BI je na jedné straně detailně posoudit **aktuální uživatelské požadavky** a vyhodnotit jejich **realizovatelnost**, případně, zda nejsou nebo nebyly již dříve řešeny. Kromě toho je smyslem **zhodnotit existující datové, technologické a další zdroje**, zejména jejich kvalitu a dostupnost, ekonomickou náročnost, vazby na ostatní existující zdroje mimo BI (Obrázek 9-3).



Obrázek 9-3: Analýza stavu a požadavků na BI

### 9.1.3.1 Obsah analýzy požadavků a stavu

Úloha **vychází z interview** a verifikace uživatelských požadavků, **obsahuje**:

- detailní analýzu a specifikaci **byznys problémů a požadavků uživatelů**,
- analýzu **podnikových procesů a jejich priorit** vzhledem k řešení BI aplikací,
- analýzu **zdrojových aplikací**, dostupnosti a kvality produkčních dat,
- specifikaci **případných změn v produkčních databázích**.

### 9.1.3.2 Klíčové aktivity

#### **Interview a verifikace uživatelských požadavků**

Požadavky je nezbytné verifikovat v rámci **předem rozvržených interview**. Důvodem je to, že v rámci fáze „Specifikace přírůstku“ nemusely být zachyceny všechny uživatelské požadavky, nebo byly definovány na hrubé úrovni a je třeba je nyní zpřesnit.

#### **Analýza podnikových procesů vzhledem k BI**

Procesní modely a jejich dokumentace mají obvykle primární užití v transakčních systémech. V Business Intelligence je jejich **význam dán těmito aspekty**:

- díky procesním modelům a reengineeringu podnikových procesů jsou přesněji **definovány požadavky i na analytické aplikace**, resp. tyto aplikace pak mohou přesněji odpovídat i potřebám podnikových procesů,
- celkové procesní modely firmy jsou základem **pro přesnější a objektivnější specifikaci priorit v řešení BI projektů**, resp. přesnější definici obsahu a pořadí řešení BI,
- současná BI řešení představují nejen jednotlivé analytické a plánovací aplikace, ale zahrnují i **definování analytických a plánovacích procesů včetně jejich realizací v aplikacích workflow** integrovaných do BI systémů, procesní modely pro ně vytvářejí potřebný základ,
- procesní modely jsou rovněž důležitým **předpokladem pro uplatňování konceptu řízení podnikové výkonnosti – Corporate Performance Management**, v němž podnikové procesy jsou jednou ze čtyř součástí.

#### **Detailní analýza a specifikace byznys problémů a požadavků**

Detailní analýza a specifikace byznys problémů a požadavků se realizuje s cílem **vyloučit duplicitní / již existující a provozované funkcionality** a současně posoudit **realizovatelnost požadavků vzhledem ke stávajícím zdrojům**, tj. finančním, personálním, datovým a technologickým. Požadavky jsou dále doplněny o další charakteristiky, jako např. zda požadavek bude řešen rozšířením některé stávající aplikace nebo aplikací novou, zda vyžaduje nové produkční datové zdroje nebo ho lze realizovat na existujících databázích datových skladů a tržišť atd.

### **Analýza dostupnosti a kvality produkčních dat**

Analýza dostupnosti a kvality produkčních dat vychází ze specifikace vazeb BI řešení na ostatní aplikace v IT architektuře tak, aby bylo plně zřejmé, **kteřá aplikace tyto zdroje produkuje (ERP, CIS, CRM)** a jaké budou **předpokládané nároky na transformace** dat z těchto aplikací. V dalším kroku se realizuje detailní **analýza kvality**, příp. chybovosti, integrity, dostupnosti potřebných datových zdrojů.

**Příprava a zajištění požadované kvality dat** na straně vstupů z úloh typu ERP apod. většinou představuje **největší část** projektu. Znamená to zajistit kontrolu dat, dostupnost požadovaných dat, jejich předzpracování, opravy, doplnění apod. Výsledkem je – na konceptuální úrovni (detailní specifikace jsou prováděny později) – **přehled nezbytných transformací a operací s produkčními datovými zdroji**, jejich obsah, předpokládaný rozsah dat a příp. další organizační a technické nároky. Do této oblasti spadá i **analýza potřeb a dostupnosti externích dat**, tj. jejich určení, technické zajištění, smluvní a další organizační opatření.

**Díličí faktory ovlivňující datovou kvalitu** lze vymezit v následujících třech skupinách:

- **technické prostředí**, zahrnující celopodnikový slovník dat, centralizaci aplikací a jejich datových zdrojů, např. jednotnou identifikaci zákazníků, kontroly definovaných byznys pravidel,
- **úroveň použité metodiky**, tj. firemní metodiky a směrnice, kvalita číselníků a kódových tabulek, systém řízení změn,
- **přístupy k řešení informačního systému**, tj. způsob přípravy dat, přípravy uživatelů, systém motivačních kritérií.

### **Specifikace změn v produkčních databázích**

Specifikace změn v produkčních databázích představuje možný zpětný **dopad BI do produkčních datových zdrojů**. Nároky na analytické a plánovací aplikace mohou někdy znamenat požadavky na změny v produkčních datech, tj. např. doplnění nových datových struktur.

#### **9.1.3.3 Podmínky úspěšnosti**

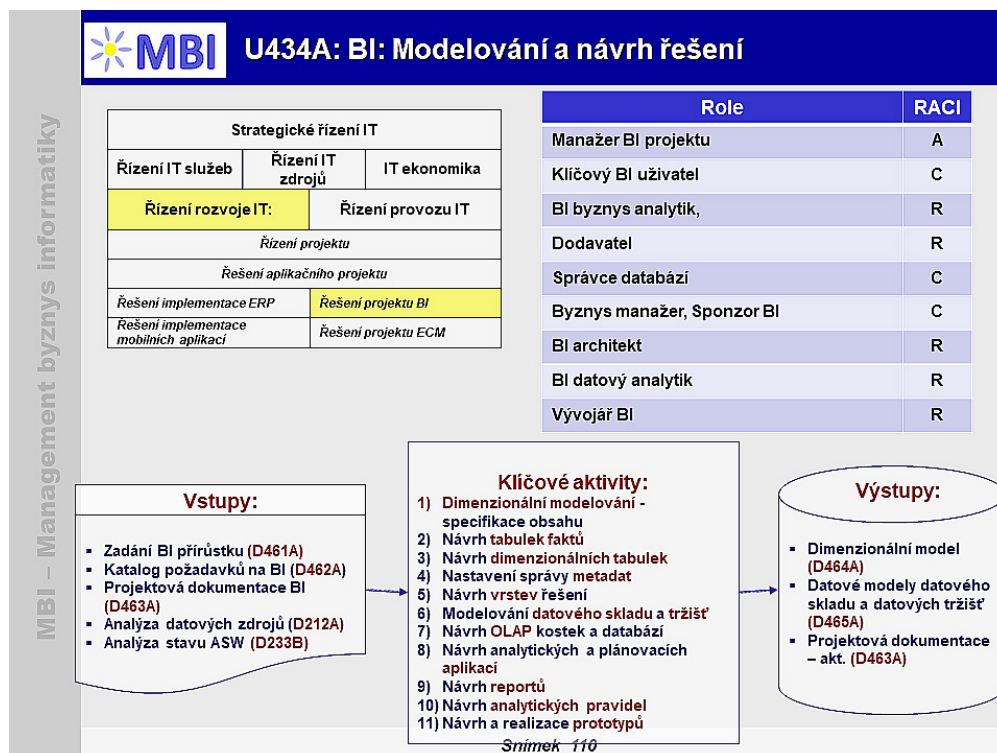
- Pro BI aplikace jsou charakteristické velmi těsné vazby na ostatní aplikace podnikové informatiky. Úspěch BI **závisí proto i na kvalitě podnikové a aplikační architektury**, úrovni zajištění systémové integrace podnikové informatiky, kvalitním plánování a zadávání nových projektů s respektováním možností využití BI (např. využití její analytické funkcionality a nezatěžováním tím transakční aplikace apod.).
- BI je postaveno na využití datových zdrojů, vznikajících převážně v transakčních aplikacích. Je nutné řešit celý **komplex otázek zajištění kvality dat**, což není záležitostí pouze BI projektů. S jejich nízkou kvalitou klesá kvalita nebo úplně zaniká BI řešení.
- Úspěch a míra využití BI aplikací nejsou často dány předpisy a disciplínou v podniku, ale **zájmem, motivací a invencí jeho pracovníků**. S tím se váže i **kva-**

**lifikace pracovníků**, jejich schopnost pracovat s novými datovými strukturami, s aplikacemi, vybavenými vysokou flexibilitou a širokou funkcionalitou.

- Úspěch BI řešení je silně závislý **na efektivní kooperaci dodavatelů a uživatelů** především v analytické fázi řešení. Průzkumy v české i zahraniční praxi však mnohokrát ukázaly, že právě nedostatečná kooperace a komunikace mezi byznysem a IT specialisty je zde překážkou kvalitnějších výsledků. Jednou z cest je **vytvoření kompetenčních center** pro BI, které byly v praxi mnohokrát ověřeny.
- **Řízení změn v podnikové informatice** – pokud není zajištěno efektivní a včasné předávání informací o změnách v primárních aplikacích správcům BI řešení, dochází k chybám ve vstupních datových strukturách v ETL transformacích, načítání chybných dat a následně k chybám ve výstupních reportech a analytických aplikacích.

### 9.1.4 Modelování a návrh řešení

**Cílem** modelování a návrhu řešení BI je **analyzovat a navrhnout obsah řešení** s využitím dimenzionální analýzy, na základě výsledků navrhnout datové modely datového skladu, resp. datových tržišť a dalších databázových komponent řešení (DSA, ODS a OLAP databázi) a definovat funkcionalitu analytických a plánovacích aplikací (Obrázek 9-4).



**Obrázek 9-4: Modelování a návrh řešení**

#### 9.1.4.1 Obsah

Úloha Modelování a návrh **BI řešení obsahuje:**

- dimenzionální modelování,
- návrh jednotlivých vrstev řešení,

- řešení datových modelů datového skladu a tržišť,
- návrh OLAP kostek a OLAP databází,
- nastavení systému metadat,
- návrh analytických a plánovacích aplikací, návrh dashboardů,
- návrh struktury a obsahu jednotlivých reportů,
- návrh analytických pravidel,
- návrh pilotních řešení a realizace prototypů.

Jádrem řešení této fáze je **dimenzionální modelování** a na jejím základě vytvoření odpovídajících databázových schémat na různých úrovních databázových komponent řešení BI. Součástí řešení jsou obvykle návrhy aplikací dotažené do prototypů, jejichž posouzení a schválení je pak racionálním základem pro následující návrhy technické infrastruktury a implementační práce.

#### 9.1.4.2 Klíčové aktivity

##### **Dimenzionální modelování – hrubý dimenzionální model**

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) a jejich dílčích charakteristik,
- **specifikace vazeb** mezi ukazateli a odpovídajícími dimenzemi.

Detailně podle metody dimenzionálního modelování, viz podkapitola 4.1

##### **Nastavení správy metadat**

Metadata jsou ve své podstatě strukturovaná data o datech. Metadata představují údaje nejen o samotných datech, ale také o technických prostředcích, softwaru, nebo sítích, kde se data nacházejí. Specifikují jejich kontext, obsah, předpokládanou interpretaci a dostupnost. **Hlavním účelem metadat** je poskytování **informací k analýze**, návrhu, vývoji, implementaci a užití jednotlivých aplikací i celého IT firmy. V souvislosti s BI řešeními je **obsah metadat např. následující**:

- celkový popis zdrojových systémů,
- popisy dat ve slovníku datového skladu,
- u datového skladu popisy transformačních pravidel pro každou tabulku a každý datový element a popisy byznys názvů a transformačních pravidel pro každou tabulku a každý datový element,
- pro reporting vysvětlení každého pole na reportu.

Podstata metadat je tedy zřejmá, jejich **uplatnění jako faktor úspěšnosti** v BI řešeních je dáno **několika důvody**:



- BI řešení se vztahují převážně na celou firmu, jsou proto velmi **komplexní, rozsáhlá a komplikovaná**. Uspořádané, jasně strukturované a dostupné informace o tom, co tato řešení obsahují, jaké datové struktury, v jakých vazbách apod., jsou **při této složitosti nutnou podmínkou realizace** těchto projektů,
- s rozsahem BI řešení **roste i rozsah jejich metadat**. Pro efektivní zajištění projektů i provozu BI se využívají celé systémy pro správu metadat, tedy databáze metadat s příslušnou aplikační nadstavbou pro práci s nimi,
- jako jeden z efektů BI, vedle své analytické a plánovací funkcionality, se běžně zdůrazňuje i jejich **úloha ve zvyšování pořádku** (např. čistoty dat) v celém informačním systému. K uskutečnění této úlohy je nutné disponovat dokonalým přehledem a evidencí o stávajících datových a dalších zdrojích IT firmy; takovou evidenci nabízejí metadata.

### **Návrh vrstev řešení**

Návrh vrstev řešení znamená již konkrétní návrh jednotlivých vrstev řešení, tj. jejich **obsahové náplně, vzájemných vazeb a příp. návrhy změn do existujících vrstev**. Definuje se tak základní obsah dat na úrovni dočasného úložiště dat – Data Staging Area, operativního datového skladu (ODS), datového skladu (DWH), datových tržišť (DMA).

### **Modelování datového skladu a tržišť**

Modelování datového skladu a tržišť zahrnuje **návrh relační databáze datového skladu a/nebo datových tržišť (tvorba logického datového modelu)**. Je postaven na běžných principech datového modelování, avšak s respektováním databázových schémat typických pro datové sklady, tj. STAR a SNOWFLAKE. Do této úlohy se obvykle řadí i odhad velikosti databází datového skladu a datových tržišť a odhady jejich nárůstů objemu (detailněji podkapitola 4.2).

BI projekty, zejména v oblasti datových skladů a tržišť jsou specifické rovněž **nutností řešení nárůstu dat v těchto databázích**. Jestliže jsme dříve hovořili o gigabytech dat jako větších datových skladech, dnes to jsou terabyty nebo desítky a stovky terabytů dat. Je tedy nutné hledat a aplikovat efektivní **metody řešení tohoto problému**, a to i za situace stále se enormně zvyšujících kapacitních parametrů paměťových zařízení. K těmto metodám patří např. metoda komprese dat, metody segmentování dat a jejich uložení na média podle předpokládaných nároků na dostupnost apod.

### **Návrh OLAP kostek a databází**

Návrh OLAP kostek a databází představuje určení všech jejich podstatných charakteristik. Určují se zejména **zdrojové tabulky datového skladu, tržiště, příp. produkčních databází, vstupujících do OLAP** kostek, jejich obsahová struktura, tj. ukazatele, dimenze, jejich vnitřní struktura, sdílení dimenzí, vypočítané ukazatele a další dílčí charakteristiky.

### **Návrh analytických a plánovacích aplikací**

Návrh analytických a plánovacích aplikací zahrnuje zejména **funkcionalitu aplikací** nad datovým skladem a/nebo OLAP kostkami a strukturu jejich komunikace s uživatelem.

### **Návrh reportů**

Návrh reportů zahrnuje **určení obsahu požadovaných reportů, včetně jejich formátu**, standardních struktur přehledových tabulek, záhlaví, komentářů, příp. i pro výstupní textové dokumenty na základě firemních standardů pro reporting.

### **Návrh analytických pravidel**

Návrh analytických pravidel je **určení limitních hodnot ukazatelů nebo jejich intervalů** ve vazbě na příslušné dimenze, které mohou být pro uživatele nějakým způsobem zajímavé. Představují např. „problém“ (cash flow je pod určitou hranicí), „varování“, „příležitost“ apod. Je zřejmé, že nastavení analytických pravidel je úzce **svázáno s vlastní ekonomickou nebo obchodní interpretací** jednotlivých ukazatelů, vycházející z praxe firmy, dalších získaných zkušeností apod.

### **Návrh a realizace prototypů**

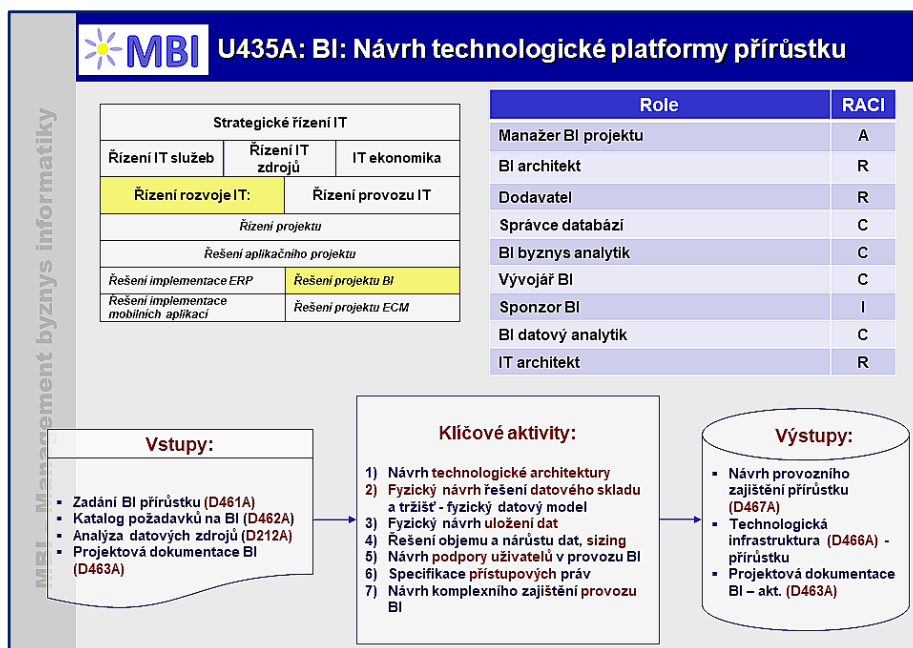
*Návrh a realizace prototypů* zahrnuje zejména **zpracování prototypového řešení na vzorku dat uživatele**, případně vygenerovaných datech, zpracování připomínek k návrhu, zpracování protokolu k oponentuře prototypu s návrhem řešení připomínek a promítnutí úprav do výsledného řešení.

#### **9.1.4.3 Podmínky úspěšnosti**

- **Úroveň zpracování dimenzionálních modelů**, v potřebné úrovni detailu je rovněž důležitým faktorem úspěšnosti řešení. S ohledem na potřebu kooperace řešitelů s uživateli by znalost principů dimenzionálního modelování u uživatelů měla patřit k podmínkám racionální kooperace.
- Předpokladem je **dostupnost dokumentace produkčních datových zdrojů**, případně možnost poskytování potřebných dat poskytovateli nebo provozovateli. V některých případech není zcela jednoduché tyto dokumentace, resp. data získat s ohledem na autorská práva nebo smlouvy mezi zákazníkem a poskytovateli těchto primárních systémů. V každém případě je dobré si tyto podmínky a možnosti ověřit a podle možností je začít řešit.

#### **9.1.5 Návrh technologické platformy přírůstku**

**Cílem** úlohy je specifikovat komplexní nároky na odpovídající **softwarové a technické zajištění** a současně i některá **provozní opatření**, vzhledem k potřebám daného přírůstku (Obrázek 9-5).



Obrázek 9-5: Návrh technologické platformy přírůstku

#### 9.1.5.1 Obsah návrhu technologické platformy

Úloha Návrh technologické platformy **obsahuje**:

- **návrh technologické architektury**, fyzický návrh řešení datového skladu a tržišť, fyzický návrh uložení dat,
- **řešení objemu** a nárůstu dat, sizing,
- **návrh podpory uživatelů** v provozu BI, specifikace přístupových práv, návrh komplexního zajištění provozu BI aplikací.

Úlohy BI musí navazovat a řídit se běžnými principy celého IT firmy, ale na druhou stranu jejich provoz musí respektovat zvláštnosti BI.

#### 9.1.5.2 Klíčové aktivity

- **Návrh technologické architektury** představuje vymezení všech podstatných technologických **komponent řešení BI, resp. jejich upgrade**. To znamená, že se určuje struktura a parametry technických prostředků, zejména serverů, příp. i úpravy nebo upgrade komunikační infrastruktury. Určuje se i struktura, parametry nebo upgrade operačních systémů, databázových systémů, speciálních prostředků pro správu dat, řízení kvality dat apod.
- **Fyzický návrh řešení** datového skladu a tržišť, tj. fyzický datový model, návrh jednotlivých databázových tabulek, řešení jejich vazeb, rozdělení do datového skladu a datových tržišť.
- **Fyzický návrh uložení dat** zahrnuje dosažení takových technologických a provozních parametrů datového skladu nebo tržiště, které by splňovaly běžné nároky na interaktivní komunikaci uživatele s BI řešením.
- **Řešení objemu a nárůstu dat, sizing**, řeší rozsah dat, který narůstá téměř geometrickou řadou a vzhledem k potřebným diskovým kapacitám je třeba realizo-

vat s perspektivou nárůstu dat pro dané časové období několika let. Znamená to však řešit i optimalizaci a efektivní organizaci takto narůstajících dat.

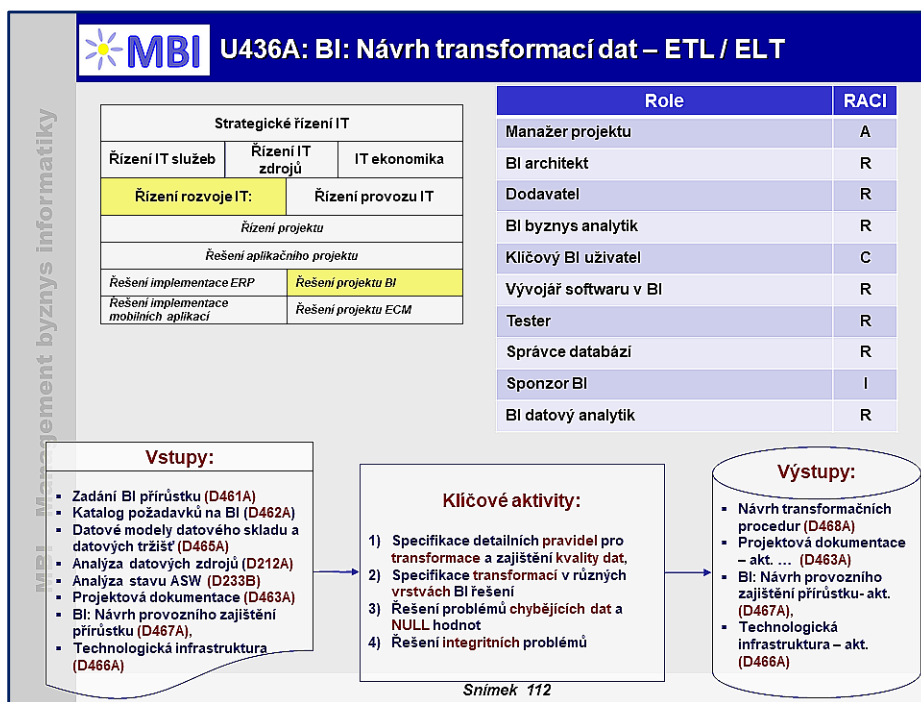
- **Návrh podpory uživatelů v provozu BI** zahrnuje školení uživatelů, vzhledem k plánovaným aplikacím a technologiím a návrh podpůrných služeb, ať speciálních nebo těch, které jsou součástí služeb informačního systému jako celku (help desk, service desk apod.).
- **Specifikace přístupových práv** je specifická v tom, že i systém přístupových práv respektuje uspořádání dat, tedy přístupová práva se neurčují pouze k jednotlivým reportům, analýzám, případně OLAP kostkám, ale i k úrovním hodnot podle struktury dimenzí apod.
- **Návrh komplexního zajištění provozu BI** definuje nároky na zajištění provozu BI aplikací, zajištění bezpečnosti, nároky na provoz ETL procedur apod.

### 9.1.5.3 Podmínky úspěšnosti

Uplatnění SOA, Service Oriented Architecture, umožňuje **vytváření analytických aplikací a jejich řešení v úzké vazbě na podnikové procesy a jejich klíčové funkce**. BI na bázi SOA tak představují procesně orientované analytické aplikace. To znamená, že se tím i fyzicky naplňuje výše uvedený koncept řízení podnikové výkonnosti, uplatnění workflow mechanismů a ve svém důsledku se posiluje úroveň integrace a současně flexibility celého informačního systému firmy.

### 9.1.6 Návrh transformací dat – ETL

**Cílem** úlohy je detailně definovat **transformační pravidla mezi produkčními daty a analytickými daty** uloženými v BI databázích, zajistit odpovídající kvalitu dat a navrhnout komplex čisticích a transformačních procedur (Obrázek 9-6).



Obrázek 9-6: Návrh transformací dat

### 9.1.6.1 Obsah návrhu transformací dat

Úloha Návrh transformací dat **obsahuje**:

- specifikace detailních pravidel pro transformace a zajištění kvality dat,
- specifikace transformací v různých vrstvách BI řešení,
- řešení problémů chybějících dat a NULL hodnot,
- řešení integritních problémů v databázích.

### 9.1.6.2 Klíčové aktivity

#### **Specifikace detailních pravidel pro transformace a zajištění kvality dat**

Specifikace detailních pravidel pro transformace a zajištění kvality dat je detailní analýza existujících datových zdrojů a **definování pravidel pro dosažení požadované kvality dat**. Zahrnuje i stanovení parametrů pro kontroly a transformace dat po stránce obsahové (datových struktur), technické i organizační (stanovení periodicity aktualizace modelů, zodpovědnosti za vstupní data). Klíčovou náplní je tu však skutečné zajištění kvality dat **návrhem systému kontrol a opravných operací**, neboť, jak jsme již zdůraznili, čistota dat je jedním z kritických faktorů úspěšnosti jakékoli BI aplikace.

#### **Specifikace transformací v různých vrstvách BI řešení**

Na základě pravidel definovaných koncovými uživateli (byznys pravidel) a navržené technologické architektury se určují **požadované transformace mezi jednotlivými komponentami**, resp. vrstvami BI řešení, případně se upravují stávající ETL procedury podle nároků na transformace řešeného přírůstku. Definují se zde **obsah a organizační stránky transformace** mezi produkčními databázemi, DSA, operativním datovým skladem, datovým skladem a datovými tržišti.

#### **Řešení problémů chybějících dat a NULL hodnot**

Neúplné datové struktury a častý výskyt NULL hodnot je zcela běžným problémem produkčních databází. Důvodů pro to je několik, od špatné disciplíny uživatelů transakčních aplikací, přes neúmyslné omyly a chyby až po objektivní neznalost nebo nemožnost získání některých informací (např. identifikační nebo kontaktní údaje zákazníka apod.). Zatímco transakční aplikace se s tímto problémem obvykle vyrovnají tím, že chybějící data prostě nezobrazí, u analytických aplikací je to podstatně složitější v tom, že **chybějícími daty může být narušena vypovídací schopnost a kvalita** výsledných analýz a reportů.

#### **Řešení integritních problémů**

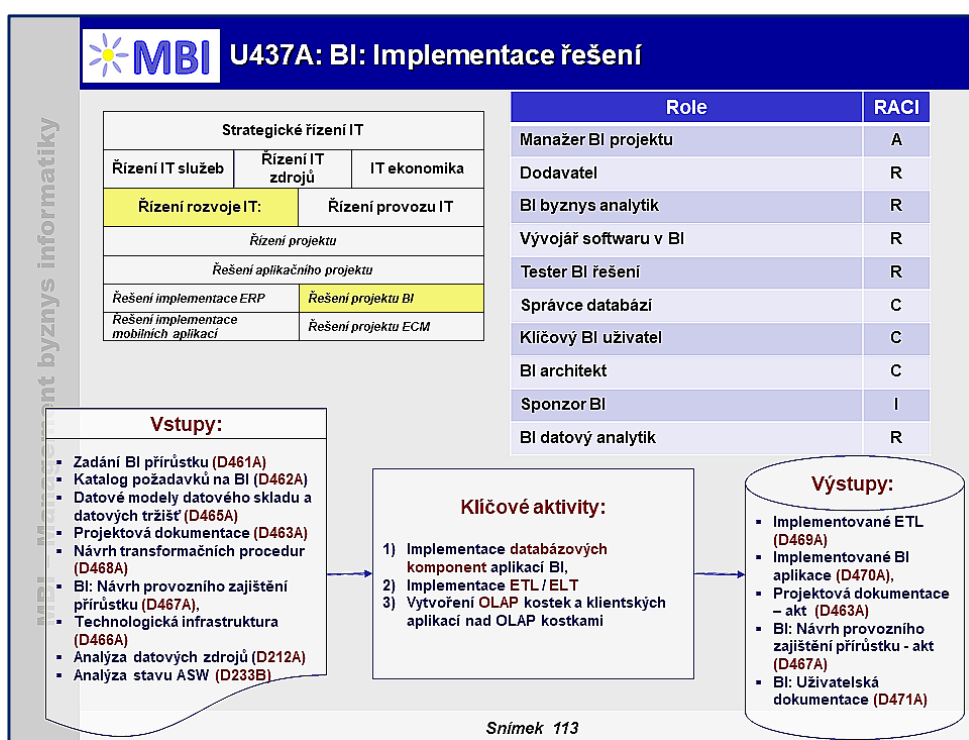
Řešení integritních problémů je obdobné jako u běžných databázových aplikací a spočívá v **porušení nebo neexistenci předpokládaných nebo definovaných vazeb** mezi záznamy jednotlivých databázových tabulek.

### 9.1.6.3 Podmínky úspěšnosti

- Předpokladem je **dostupná a kvalitní dokumentace zdrojových databází**.
- Je nezbytný kvalitní systém, resp. **pravidla pro zajištění kvality**, čištění dat a jejich konsolidaci.
- Často rozhoduje správné **využití vhodných SW prostředků** pro ETL transformace, buď integrovaných do databázových systémů pro BI, nebo specializovaných pro ETL (např. Informatica apod.).
- V harmonogramu řešení je nutné **vyhradit ETL potřebnou časovou a pracovní kapacitu**, většinou ETL představují cca 60 % celého řešení BI

### 9.1.7 Implementace řešení

**Cílem** úlohy je zajistit **vytvoření a otestování požadovaných aplikací a nástrojů** BI v rámci celého přírůstku, vytvořit a kompletovat dokumentaci BI řešení (Obrázek 9-7).



Obrázek 9-7: Implementace řešení

#### 9.1.7.1 Obsah implementace řešení

Úloha Implementace **obsahuje** implementaci databázových komponent, dotazů a aplikací BI, ETL procedur, OLAP kostek a klientských aplikací nad OLAP kostkami.

**Nejčastějším nástrojem** pro tvorbu BI aplikací nad OLAP kostkami jsou **kontingenční tabulky** a grafy Excel. Vedle dalších standardních programových prostředků (SQL, VBA) se pro implementaci BI aplikací používají i některé speciální nástroje, které jsou vlastní pouze BI, např. prostředky pro vytváření OLAP kostek, nástroje data miningu apod.

### 9.1.7.2 Klíčové aktivity

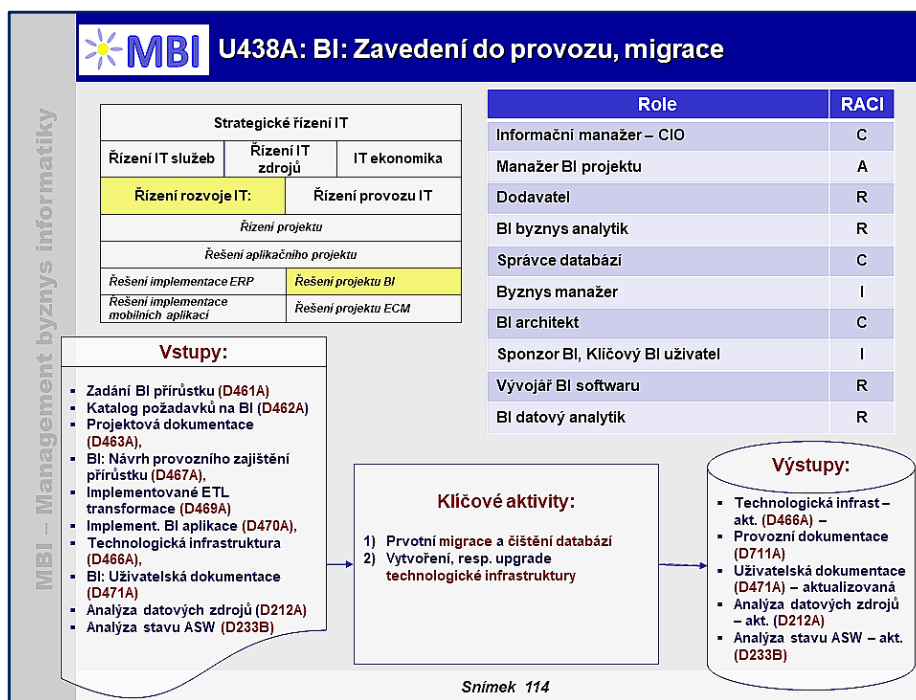
- **Implementace databázových komponent** aplikací BI mají obdobný charakter jako u ostatních typů databází s tím, že charakter prováděných SQL dotazů musí ve svém obsahu respektovat charakter databázových schémat STAR a SNOWFLAKE (pokud jsou použity) a respektovat tak současně analytické nebo plánovací zaměření řešených aplikací.
- **Implementace ETL** je programová realizace datových pump, otestování provozu a zejména jejich časových nároků, nastavení jejich provozních parametrů.
- **Vytvoření OLAP kostek a klientských aplikací** nad OLAP kostkami jsou aplikace v běžných klientských nástrojích, převážně Excelu, nebo řešení těchto aplikací ve specializovaných nástrojích pro BI aplikace. Další náročnější variantou je využití speciálních dotazovacích jazyků, např. v jazyku MDX (Multidimensional Expressions) nebo DAX (Data Analysis Expressions).

### 9.1.7.3 Podmínky úspěšnosti

- **Výstupy, resp. reporty** z datových skladů a tržišť jsou **umístovány na podnikové portály**, a to nejen pro využití interními pracovníky firmy, ale v určitých oblastech řízení (stav zakázek apod.) pro obchodní partnery podniku. Nabídka této funkcionality s pochopením a respektováním potřeb i externích subjektů tvoří jeden z podstatných efektů BI a ovlivňuje tak i jeho výsledný úspěch.
- **Využití adekvátních analytických nástrojů** je třeba realizovat vzhledem k nárokům uživatelů i ekonomické výhodnosti jejich pořízení a provozu.
- **Implementace aplikací BI by měly zahrnovat nejrůznější oblasti řízení**, včetně aplikací pro řízení podnikové informatiky.
- Pro efektivní využívání analytických aplikací je třeba **zajistit kvalitní školení uživatelů** s důrazem na principy a efekty využití multidimenzionálních aplikací a jejich flexibilitu.
- Vedle vytváření specializovaných aplikací, resp. vytvářených přímo na míru, lze využívat **aplikace, které jsou integrovány do jiných produktů**, zejména ERP; tyto varianty je nutné dobře posoudit.

### 9.1.8 Zavedení BI do provozu, migrace

**Cílem** úlohy je zajistit **vytvoření provozního prostředí pro ETL, datové sklady, datová tržiště, OLAP** databáze, aplikace a další součásti BI řešení. Připravit uživatele BI na práci s aplikacemi BI a jejich efektivní využití (Obrázek 9-8).



**Obrázek 9-8: Zavedení BI do provozu, migrace**

### 9.1.8.1 Obsah

Úloha **obsahuje zejména prvotní migrace** a čištění databází a vytvoření, resp. upgrade technologické infrastruktury, plošné školení uživatelů.

### 9.1.8.2 Klíčové aktivity

- **Prvotní migrace a čištění databází** se vzhledem k možným odlišnostem od standardního běhu a z toho vyplývajících problémů typicky provádí ještě ve fázi implementace pod dohledem vývojového týmu.
- **Vytvoření, resp. upgrade technologické infrastruktury** vyplývá z návrhů úlohy „Návrh technologické platformy“ a je de facto jejich fyzickou realizací.

### 9.1.8.3 Podmínky úspěšnosti

- Úspěch BI migrace ovlivňuje i úroveň řízení provozu celé informatiky, **zejména správu databázových serverů, plánování a kontrolu průběhu ETL procesů**, zařazení problematiky BI do služeb service-desku.
- Zajištění potřebné **kvality dat**, vstupujících do datových skladů a tržišť a následně analytických aplikací, je klíčovou podmínkou pro vůbec smysluplnou existenci těchto součástí informačního systému. Na druhé straně je **tato podmínka v konkrétních podnicích často nesplněna** nebo je splněna pouze rámcově. To bylo důvodem pro vznik různých **metod a přístupů k řízení kvality dat (jako je MDM, Data Governance apod.)**.

Na závěr této část **uvádíme přehled rolí a jejich podíl** na jednotlivých úlohách postupu řešení projektu.



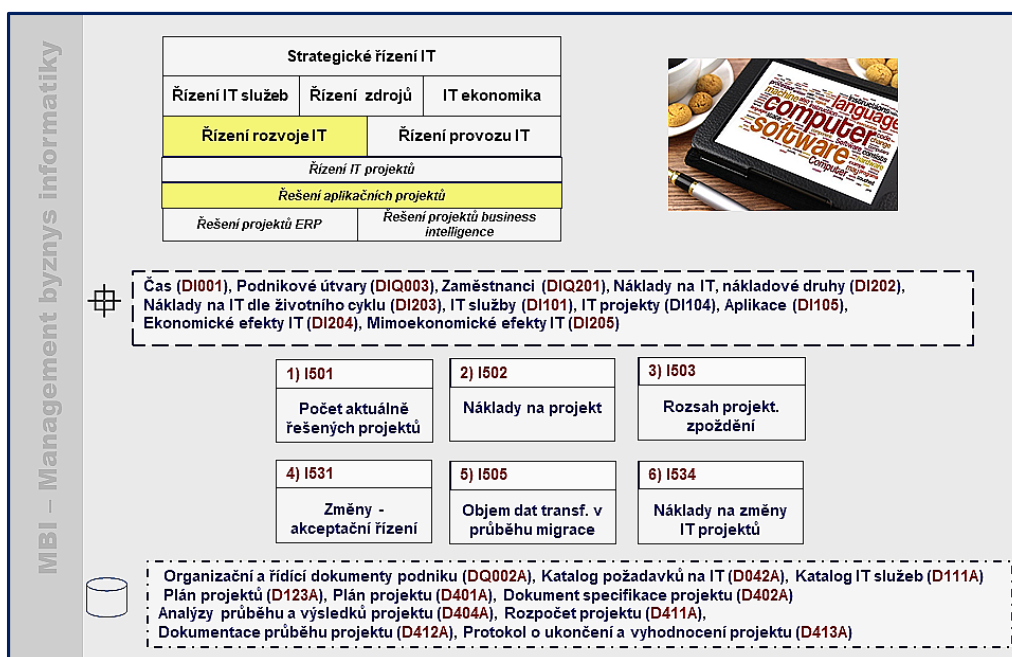
Jen existující vazby? ☑		UST	Přír.s.	Anal.	Model.	IT	ETL	Impl.	Prov.
Typ: RACI kód		U431A	U432A	U433A	U434A	U435A	U436A	U437A	U438A
Role:									
Informační manažer (CIO)	R101	C	C						C
Dodavatel	R109	R	R	R	R	R	R	R	R
IT architekt	R401	C	C			R			
Správce databázi	R502	R	C	C	C	C	R	C	C
Byznys manažer	RQ009	C	C	C	C				I
Sponzor BI projektu	RQ101	C	C	C	C	I	I	I	I
Manažer BI projektu	RQ102	A	A	A	A	A	A	A	A
BI byznys analytik / BI Konzultant	RQ111	R	R	R	R	C	R	R	R
BI datový analytik	RQ112	R	C	R	R	C	R	R	R
BI architekt	RQ121	R	R	R	R	R	R	C	R
Vývojář softwaru v BI	RQ122				R	C	R	R	R
Klíčový BI uživatel, Power User	RQ132	C	C	C	C		C	C	C

Obrázek 9-9: Role v řešení projektu Business Intelligence

# 10

## 9.2 KPI řešení projektu Business Intelligence

Další přehled metrik představuje pouze **vybrané**, které lze považovat za **KPI**. Přehled KPI řešení projektu Business Intelligence představuje Obrázek 9-10.



Obrázek 9-10: Přehled KPI řešení projektu Business Intelligence

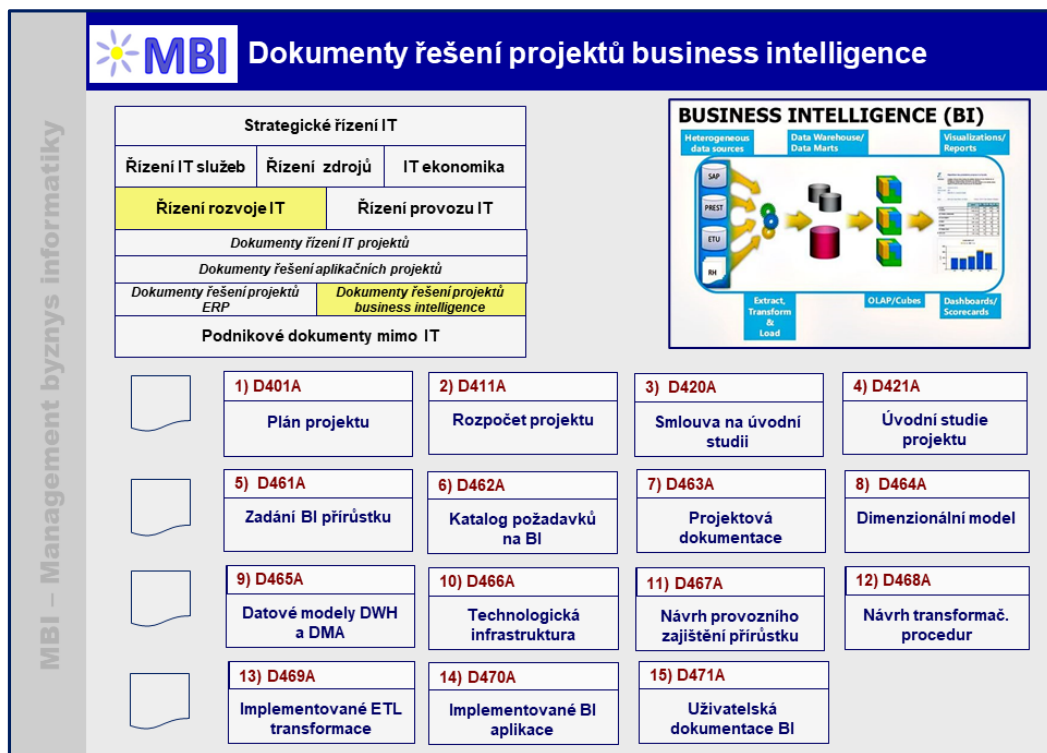
V dané oblasti řízení se využívají zejména tyto základní a související **metriky**:

- **Počet aktuálně řešených BI projektů:**
  - **Aktuální doba trvání projektů** ve dnech ve sledovaném období podle projektů a dodavatelů.
  - **Počet dokončených a předaných release projektů** podle projektů a dodavatelů.
  - **Počet nedodržených milníků** projektů v projektovém plánu podle projektů a dodavatelů.
  - **Počet identifikovaných rizik** v projektech podle projektů a dodavatelů.
- **Náklady na BI projekt:**
  - **Podíl počtu projektů dokončených** v rámci stanoveného rozpočtu v % podle typů projektů a dodavatelů.
  - **Podíl objemu skutečně čerpaných nákladů** na projekt vzhledem k rozpočtovaným v % podle projektů a dodavatelů.
- **Čas projektu:**
  - **Rozsah projektových zpoždění** ukazuje rozsah projektových zpoždění v důsledku zdržení investičních rozhodnutí nebo v důsledku nedostatku fondů.
  - **Podíl objemu nákladů zpožděných projektů** na jejich celkovém objemu nákladů na projekty.
- **Počty změn, vyžádaných v důsledku akceptačního řízení.**
- **Objem dat, transformovaných v průběhu migrace:**
  - **Podíl úspěšně migrovaných dat** do nového systému na celkovém objemu migrovaných dat v %.
- **Náklady na změny IT projektů:**
  - **Poměr požadavků zvyšující náklady** proti požadavkům snižující náklady v % podle projektů a dodavatelů,
  - **Podíl objemu nákladů na změnu** podle typu změny projektu na celkových nákladech na změnu v % podle projektů a dodavatelů.



### 9.3 Data, dokumenty

**Přiřazení dat a dokumentů k jednotlivým úlohám** dokumentují schémata v kapitole 9.1. **Souhrnný přehled** dat, dokumentů v řešení projektu Business Intelligence, dokumentuje Obrázek 9-11.



**Obrázek 9-11: Vstupy a výstupy úloh řešení projektu Business Intelligence**

V dalším přehledu uvádíme data a dokumenty, a to v pořadí podle schématu:

- Plán projektu.
- Rozpočet projektu.
- Smlouva na úvodní studii.
- Úvodní studie projektu.
- Zadání BI přírůstku.
- Katalog požadavků na Business Intelligence.
- Projektová dokumentace BI.
- Dimenzionální model.
- Datové modely datového skladu a datových tržišť.
- Technologická infrastruktura BI.
- Návrh provozního zajištění přírůstku.
- Návrh transformačních procedur pro BI.
- Implementované ETL transformace.
- Implementované BI aplikace.
- Uživatelská dokumentace BI aplikací.



## 9.4 Scénáře, analytické otázky k řešení projektu Business Intelligence

Další scénáře obsahují **náměty analytických otázek** pro řešení projektů a pro přípravu diskusí s manažery a pracovníky firmy. Podle konkrétní situace je třeba vybrat relevantní otázky.

### 9.4.1 Implementují se aplikace Business Intelligence

- Jak zajistit **efektivní a kvalitní průběh** implementace?
- Jak definovat **jednotlivé přírůstky BI** řešení a jejich priority?
- Jakou zvolit **architekturu BI řešení**?
- Jak připravit jednotlivé aplikace a přírůstky **do provozu**?
- Jak zpracovat veškerou **potřebnou dokumentaci**?

### 9.4.2 Řeší se nasazení datového skladu

- Představuje datový sklad **přínosný bod ve vývoji firmy** z pohledu strategie společnosti? Bude tento přínos dostatečným faktorem k další investici do budoucího DWH řešení?
- Jakým způsobem bude sklad **vyvíjen** (využijí se interní zdroje, dodavatelské, popř. aplikační balík)?
- Bude nové DWH řešení představovat **rozvoj stávajícího DWH řešení**, nebo se bude jednat o implementaci nového DWH řešení?
- Jaká **platforma** bude pro implementaci daného DWH řešení využita?
- Jaké **nástroje** budou při implementaci DWH řešení využity?
- Bude vhodné využít způsob **ETL, nebo ELT** řešení?
- Jak se bude rozvíjet **obsah DWH**?
- Jak rychle budou **data zprostředkována** prostřednictvím DWH?

### 9.4.3 Řeší se nasazení datového tržiště

- Představuje datové tržiště (dále DM) **přínosný bod** ve vývoji firmy z pohledu strategie společnosti? Bude tento přínos **dostatečným faktorem** k rozhodnutí o investici do vývoje DMA?
- Jakým způsobem bude DMA **vyvíjeno**? Budou při vývoji využity **dodavatelské zdroje, interní zdroje**, popř. aplikační balík?
- Bude nový DMA **rozvíjet nynější DM řešení**, nebo se bude jednat o implementaci nového DMA?
- Jaká **platforma** bude při implementaci DMA využita?
- Jaké **nástroje** budou při implementaci DMA využity?
- Jakým způsobem budou **data do DMA** nahrávána?
- Jaký **typ procedury** bude při nahrávání dat využíván?

#### 9.4.4 Řeší se uplatnění ETL / ELT

- Je pro danou situaci vhodnější využití **ELT, nebo ETL** procedury?
- Jakou **platformu** lze využít při řešení daného projektu?

#### 9.4.5 Řeší se uplatnění OLAP databází

- Je třeba v případě nového řešení, založeného na rozvoji předešlého BI projektu, **OLAP kostku modifikovat**, nebo vytvořit zcela novou?
- Jaké technologie budou při řešení využívány?
- Má nějaká **platforma** větší předpoklad pro využití v konkrétním projektu?
- Disponuje **řešitel dostatečnou znalostí byznys logiky** daného problému?

#### 9.4.6 Řeší se uplatnění analytických aplikací

- Představují analytické aplikace vhodnou **podporu managementu** společnosti?
- Poskytuje dostatečné **ekonomické analýzy** prostředí firmy?
- Je na jejich základě možné popisovat a vyhodnocovat **podnikové hrozby**?
- Zohledňují **legislativní a jiná omezení**?
- Představují pro danou společnost **konkurenční výhodu**?
- Disponují dostatečným **zabezpečením** citlivých dat?

#### 9.4.7 Řeší se uplatnění reportingu

- Zachycuje reporting **vhodné metriky a atributy**?
- Jaké jsou cíle a očekávané **přínosy** reportingu?
- Budou moci daný **reportingový nástroj** obsluhovat i nezainteresovaní uživatelé bez předešlé znalosti reportingových nástrojů?
- Zohledňuje konkrétní reportingový nástroj **legislativní a jiná omezení**?
- Představuje daný reportingový nástroj **konkurenční výhodu**?
- Disponuje reportingový nástroj dostatečným **zabezpečením** citlivých dat?
- Jsou **přínosy** reportingu prokazatelné?
- Které **oblasti** má reporting pokrývat?
- Jakého charakteru bude konečně řešení (**centralizované / decentralizované**)?
- V jakém potřebném/přípustném **časovém horizontu** má být řešení implementováno?
- Jak velký **rozpočet** má být alokován na implementaci a údržbu systému?
- Bude k implementaci docházet formou **cloud technologie**, bude outsourcován, nebo bude probíhat formou využití interních zdrojů?

#### 9.4.8 Řeší se uplatnění dočasného úložiště dat, DSA

- Představuje DSA **efektivní uvolnění výkonů** primárních systémů?
- Přináší DSA danému projektu **pozitivní efekt** z finanční a implementační perspektivy?

- Disponuje společnost zavádějící DSA **dostatkem zdrojů a prostředků** pro realizaci DSA?

#### 9.4.9 Řeší se uplatnění nástrojů dolování dat

- Bude třeba při DMI **využít DWH**?
- Jak lze poznatky z DMI řešení **využít v praxi** řízení firmy?
- Budou mít **pro zákazníka výstupy** DMI jasné přínosy?
- Lze z výsledků DMI **vypozorovat daný trend**, popřípadě jakým směrem se tento trend vyvíjí?
- Bude zvolený **DMI nástroj** obsahovat intuitivní uživatelské rozhraní?
- Jakou **platformu a nástroje při DMI** řešení zvolit?

### 9.5 Závěry k postupu řešení projektu Business Intelligence

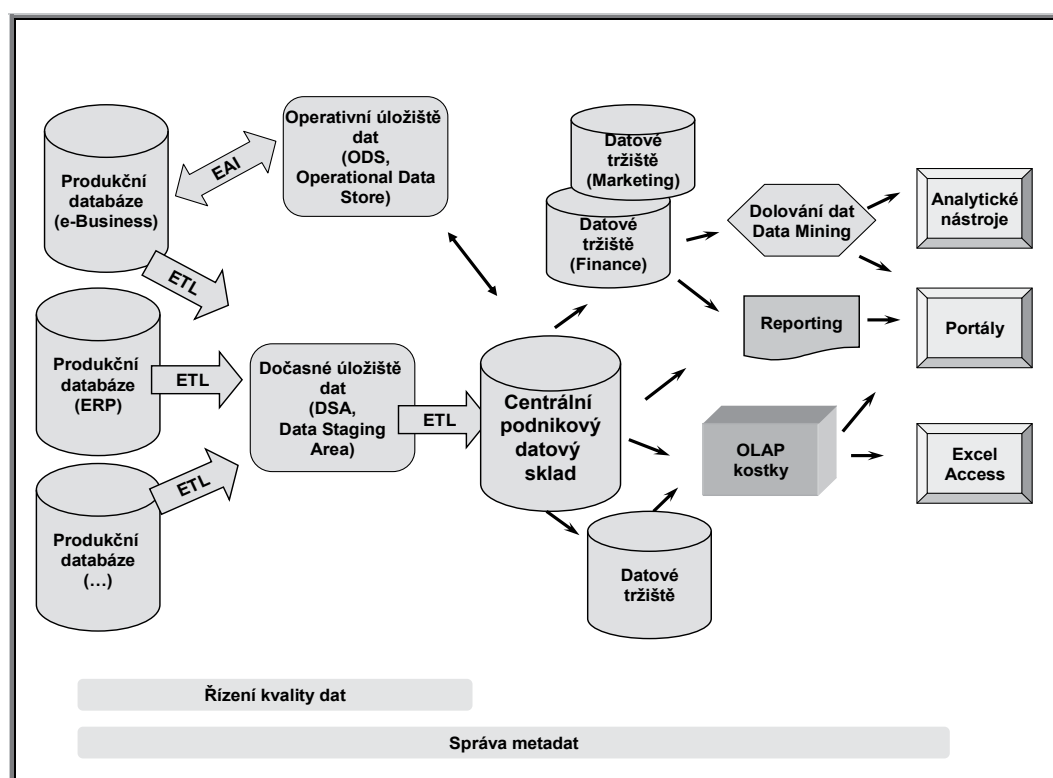


- Pro implementaci BI řešení existuje celá **řada implementačních metodik**, obvykle je třeba využívat tu, kterou využívá dodavatel řešení.
- V rámci BI se jako klíčová využívá **metoda dimenzionálního modelování** – s tou by se měli dobře seznámit i uživatelé, podílející se na BI projektech.
- Řešení BI aplikací **ovlivňuje celá řada faktorů**, např. tzv. Self service BI, Big Data Analytics apod.
- Členové pracovních týmů **z uživatelské sféry** by měli být připraveni využívat standardní **analytické metody**.
- U každého aplikačního projektu musí být **definováno akceptační řízení** a na základě jeho výsledků je třeba definovat postupy, jak aplikaci zavést do provozu.
- **Zajištění akceptačního protokolu** je klíčová věc, bez které nelze projekt úspěšně uzavřít.
- Je dobré vytvořit **prototypové řešení**, resp. Proof of Concept, kde si uživatel může lépe představit finální produkt.
- Pro zpracování úvodní studie je dobré vyčlenit **dostatečný časový prostor**, protože její kvalita často ovlivňuje výslednou kvalitu celého projektu.
- Úvodní studie se často stává podkladem pro přípravu kontraktu na celý projekt a pak je třeba zajistit **provázanost Úvodní studie a kontraktu**.
- Pro efektivní kooperaci implementačních týmů s klíčovými uživateli je nezbytné těmto **uživatelům vytvořit, pokud je to možné, dostatečný časový prostor v rámci jejich pracovní náplně** a odpovídající motivaci na řešení projektu.
- Je účelné dobře vyhodnotit **kvalitu původních databází** a podle toho upravit i plán a harmonogram migrace.
- V rámci strategie a následně plánu migrace je třeba **definovat tzv. roll out**, tj. instalaci aplikace na jednotlivé dislokované jednotky podniku (závo-  
dy, pobočky apod.).
- Pro datovou migraci je třeba disponovat **dokumentací dat původního systému**.

## C) Komponenty podnikové analytiky

Oddíl zahrnuje podstatné **charakteristiky jednotlivých dílčích komponent a konceptů**, zejména BI a SSBI řešení, rozdělených do několika skupin. Kromě stručného **funkčního a technologického vymezení** zahrnují jednotlivé kapitoly, resp. podkapitoly věnované jednotlivým komponentám, zejména hlavní poskytované **efekty** na jedné straně a případná **omezení** nebo **potenciální problémy** na straně druhé.

Celkové schéma uspořádání klíčových komponent v rámci komplexního řešení podnikové analytiky dokumentuje další obrázek:



Obrázek C-1: Obecné uspořádání hlavních komponent v systému BI

Konkrétní **uspořádání jednotlivých součástí** v řešení podnikové analytiky se může výrazně měnit **podle situace a potřeb dané firmy**. To znamená v rozsahu od těch nejjednodušších řešení až po řešení nejkomplexnější a také technologicky, finančně i pracovníčně nejnáročnější.

V návaznosti na předchozí schéma jsou **v dalším přehledu** v pracovním členění uvedeny **komponenty a koncepty pouze s nejstručnější charakteristikou**, detailnější popis je v následujících kapitolách nebo podkapitolách:

- **Datové báze podnikové analytiky:**
  - **Datový sklad, DWH**, je integrovaný, konsolidovaný, subjektivě orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu (INMON, 2002).

- **Datové tržiště, DMA**, je specifický datový sklad, určený pro omezený okruh uživatelů (oddělení, divize, pobočka, závod).
- **Dočasné úložiště dat, DSA**, představuje dočasné uložení extrahovaných dat z produkčních databází s cílem zajistit jejich přípravu a potřebnou kvalitu před vstupem do datového skladu.
- **Operační datový sklad, ODS**, je místo datové integrace aktuálních dat z primárních systémů, podporující relativně jednoduché dotazy nad menším množstvím aktuálních dat.
- **Data Lake** je úložištěm veškerých podnikových dat (strukturovaných i nestrukturovaných) i těch v takzvané surové podobě.
- **Transformace dat:**
  - **Extract Transform Load (ETL / ELT)**, resp. datová pumpa, data ze zdrojových systémů vybírá (*Extract*), upravuje do požadované formy a uspořádání (*Transform*) a nahrává je do datových schémat datového skladu nebo tržiště (*Load*).
  - **Zachycení změn ve zdrojových systémech, Change Data Capture (CDC)**, reprezentuje schopnost zachytávat změny v datech ve zdrojových systémech a získávat tyto změny efektivně, a pokud je to možné, dosáhnout co nejmenšího zatížení zdrojových systémů.
  - **Historizace dat** představuje možnost určit, kdy, která data byla či jsou platná, resp., že všechna data v datovém skladu jsou určitým způsobem identifikována prvky dimenze času.
- **Analytika:**
  - **OLAP databáze** představují jednu nebo několik souvisejících a vzájemně propojených OLAP kostek, které již zahrnují předzpracované agregace dat podle definovaných hierarchických struktur dimenzí a jejich kombinací.
  - **In-Memory Analytics** je analytická funkcionalita realizovaná v operační paměti, představuje součást produktů, umožňující realizovat multidimenzionální analýzy, a to i s detailními daty, uloženými v operační paměti.
  - **Sandbox** představuje možnosti pro uživatelské zkoušení, prototypování a případně testování nových nástrojů a analytických úloh.



## 10. Datové báze v podnikové analytice



Účelem kapitoly je:

- vymezit základní **charakteristiky databází**, majících již analytický charakter, tj. datových skladů, datových tržišť a dalších, jejich pozitiva a problémy,
- vytvořit podklad pro **analýzu jejich možností a nároků**, zejména při řešení architektury BI systémů a výběru nejvhodnějších databázových produktů.
- Do této kapitoly **nejsou zařazeny zdrojové databáze**, jejichž otázky jsou řešeny např. s řízením kvality dat.

### 10.1 Datový sklad, DWH

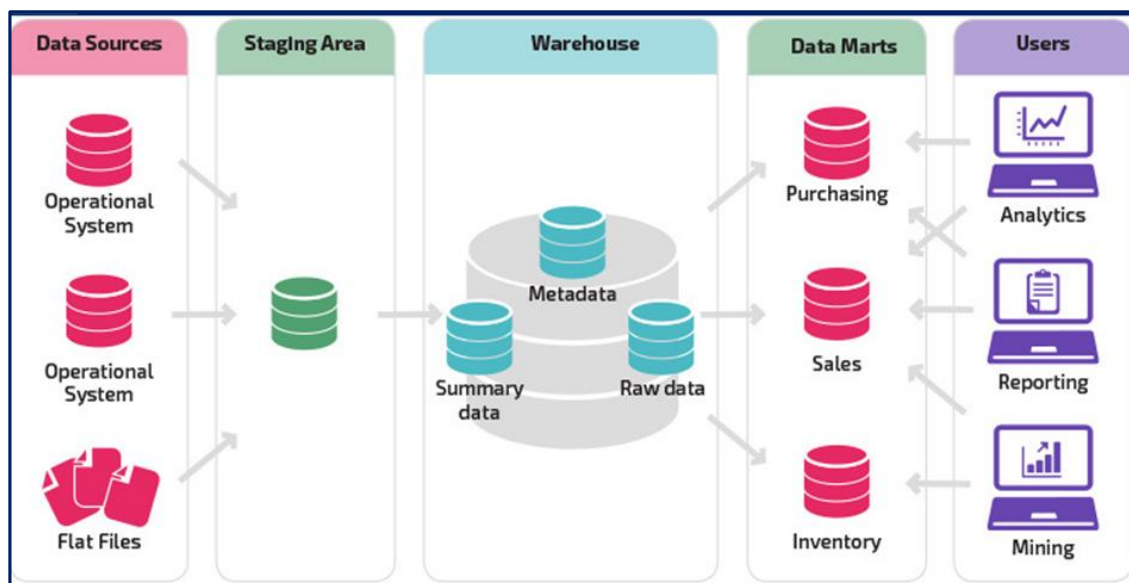
Technologie datových skladů (**Data Warehouse, DWH**) představuje v současné době již běžnou součást firemních IT. Podle (INMON, B.: Building the Data Warehouse. Indianapolis, John Wiley and Sons 2002) je **datový sklad vymezen** takto: „Datový sklad je integrovaný, konsolidovaný, subjektivě orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu“. Tyto pojmy **lze interpretovat** takto:

- **subjektivě orientovaný** – data jsou rozdělována podle jejich typu, ne podle aplikací, ve kterých vznikla,
- **konsolidovaný** – data jsou konsolidována z různých zdrojů, struktur a forem do jedné výsledné formy (do jedné verze pravdy),
- **integrovaný** – data jsou ukládána v rámci celé firmy, a ne pouze v rámci jednotlivých útvarů,
- **stálý** – datové sklady jsou koncipovány převážně jako pouze pro čtení (read only), až na výjimky se zde žádná nová data nevytvářejí ani neaktualizují,
- **časově rozlišený** – do datového skladu je uložena i historie dat, tedy obsahují dimenzi času.

V současné době datové sklady obvykle **obsahují normalizovaná data**, zatímco denormalizace, např. na bázi STAR schémat, se realizuje až na úrovni datových tržišť.

Datové sklady (obdobně datová tržiště) jsou **realizované v prostředí relačních databázových systémů**, např. MS SQL Server, Oracle, DB/2, Teradata a další (Snítíl, 2018).

Příklad uspořádání komponent datového skladu představuje Obrázek 10-1:



Obrázek 10-1: Data warehouse (Zdroj: panoply.io)

### 10.1.1 Efekty a přínosy datového skladu

- Disponuje **konsolidovanými, konzistentními, subjektivě orientovanými a historizovanými daty**.
- DWH poskytuje **optimální informace** podporující manažerské rozhodování na strategické, taktické a operativní úrovni řízení.
- Umožňuje získat **strukturovaný přehled** z velkého množství dat, uchovávaných v různorodých databázích.
- Nabízí rychlý a **komplexní přístup** k velkému množství dat.
- Disponuje schopnostmi pro **modelování a prognózy** podle potřeb řízení firmy.
- Funguje na principu „**read only**“.
- Data jsou do DWH **nahrávána** obvykle **v definovaném časovém úseku**, méně často průběžně (on line).
- Disponuje schopností aktivního **využití externích datových zdrojů** třetích stran (např. dlouhodobé zachycení vývoje produktu).
- Disponuje **operativními dotazy**, analýzami časových řad, multidimenzionální analýzou a finančními analýzami.
- Disponuje informacemi, uloženými **v metadatech**.
- DWH představuje datový **zdroj pro analytické aplikace a reportingové nástroje**.

### 10.1.2 Otázky a problémy datových skladů

- Vysoká **technologická a finanční náročnost** plyne z možných kombinací zdrojových systémů a zákaznických požadavků, nutných pro realizaci DWH.
- Při zavádění nového podnikového informačního systému je ve valné většině případů nutné **realizovat nový DWH**.
- Oproti klasickým relačním databázím je **paměťově náročnější** a dochází k delší době odezvy.

- Řešení představují komplikované **načítací procedury**.
- Obecná **neznalost funkcionality** a možností datových skladů je často na straně podnikových manažerů.

### 10.1.3 Požadavky na datový sklad

**Požadavky** na datový sklad lze formulovat následovně.

- Informace z DWH **musí být jednoduše dostupné** – DWH musí poskytovat informace jednoduše, obsah DWH musí být **pochopitelný** pro uživatele, musí nabízet informace v nejrůznějších kombinacích (slicing and dicing), zajišťovat co nejkratší dobu odezvy.
- Informace z DWH musí být **prezentovány konzistentně** – musí poskytovat věrohodné informace, tj. data musí být shromážděna z různých zdrojů, pečlivě **kontrolována a čištěna** a poskytována až tehdy, pokud jsou v pořádku.
- Data musí být postavena na **jasném vymezení obsahu** a jasném odlišení jejich identifikátory, resp. názvy.
- DWH musí být **flexibilní vzhledem ke změnám** – musí být adaptabilní ke změnám v uživatelských požadavcích, podnikovému prostředí, datovým zdrojům, technologiím.
- DWH musí **zajišťovat bezpečnost dat** – DWH obsahuje kompletní a často citlivé informace, musí být zajištěno řízení přístupů k jednotlivým částem DWH.
- iDWH musí tvořit **základ pro zkvalitňování řídicích a rozhodovacích procesů** – pro DWH musí být deklarovány efekty, které přinese (i když nemusí být vždy exaktně vyjádřeny a ve finančních ukazatelích).
- Podniková **komunita** musí DWH **akceptovat adekvátním způsobem** – využití DWH (a BI) není, na rozdíl od transakčních systémů, nezbytné a je závislé na ochotě a invenci uživatelů. Proto je také nezbytné **naplnit všechny předchozí požadavky**.

### 10.1.4 Úrovně řešení datového skladu

DWH má architekturu v následujících hlavních úrovních. Dále využívá databáze pro potřeby logování a správy metadat. Celkově se tak skládá z několika databází (Vaněk, 2020):

- **DWH\_Stage (L0): Vstupní databáze**
  - Obsahuje objekty se strukturou 1:1 oproti objektům ze zdrojových systémů. Je jedno, z jakého zdroje objekt či data v objektu přicházejí, všechny proudí do této databáze. Uživatel může jasně rozpoznat, z jakého zdroje objekty jsou pomocí dvou ukazatelů. První je atribut v samotné tabulce, který má defaultní hodnotu jako název zdrojového systému. Druhý je schéma, ve kterém je objekt uložen. Počet schémat je tedy roven počtu zdrojových systémů.
  - Tato vrstva se neudržuje. Před každým iterativním nahráváním dat se všechny tabulky přemažou.

- **DWH (L1): Přestupní vrstva mezi vstupní a výstupní**
  - Obsahuje více schémat, ovšem nejdůležitější je PreComputed a dbo. PreComputed schéma obsahuje stejné struktury jako L0, ovšem s tím rozdílem, že po každém iterativním nahrání dat se velikost dat pouze rozšiřuje a nemaže. Tedy slouží jako persistentní (stálá) vstupní vrstva. Schéma dbo již obsahuje data transformovaná do podoby podle potřeb budoucích dimenzních a faktových tabulek na úrovni L2. Objekty a atributy se zde řídí podle jmenné konvence, získávají optimální datové typy a ukládají pouze ty atributy, které se nyní využívají. Pokud některý atribut bude v budoucnu potřebný pro reporting, dá se celá jeho historie transformovat ze schématu PreComputed.
  - Všechny cizí klíče se ve schématu dbo transformují na Surrogate Keys (úspora prostoru, transformace složených klíčů na jednoduché).
- **DWH\_Mart\_Service (L2): Výstupní vrstva**
  - Obsahuje již konečnou podobu dimenzí a faktů tak, aby odpovídaly potřebám reportingu.
- **DWH\_Meta:**
  - Ukládá data o mapování objektů pro potřeby ETL a všechny procedury pro tvorbu procedur z těchto metadat.
- **DWH\_Log:**
  - Zde se ukládají informace o průběhu procedur. Je zdrojem pro monitoring a monitorovací reporty.

## 10.2 Datové tržiště, DMA

**Princip datových tržišť (Data Mart, DMA) je obdobný**, jako v případě datových skladů. **Rozdíl** je v tom, že datová tržiště jsou určena pro **omezený okruh uživatelů** (oddělení, divize, pobočka, závod). Podstatou jsou tak **decentralizované „datové sklady“**, které se pak mohou postupně integrovat do celopodnikového řešení.

Datové tržiště je tak **problémově orientovaný datový sklad, určený pro pokrytí konkrétní problematiky** vymezeného okruhu uživatelů a umožňující flexibilní „ad-hoc“ analýzy dat. **Výsledkem** vytváření datových tržišť je **zkrácení doby návratnosti investic**, snížení nákladů a podstatné zmenšení rizika při jejich zavádění.

### 10.2.1 Efekty a přínosy datového tržiště

- DMA poskytuje **informace, podporující rozhodování** manažerů na strategické, taktické a operativní úrovni řízení.
- Nabízí uživatelům **flexibilní přístup** k datům.
- V porovnání s DWH **lze DMA vyvinout levněji** a rychleji.
- V porovnání s DWH poskytuje **kratší dobu odezvy** při dotazování.
- Disponuje konsolidovanými, konzistentními, agregovanými, subjektivě orientovanými **daty** a historizovanými daty.

- Nabízí schopnost získat **strukturovaný přehled** z velkého množství dat, uchovávaných v různých databázích.
- Umožňuje rychlý a **komplexní přístup** k rozsáhlému množství dat.
- Přístup je zajištěn pouze **oprávněným uživatelům** právě v takovém rozsahu, který jim přísluší v návaznosti na konkrétní právní politiku příslušné organizace.

### 10.2.2 Otázky, problémy a omezení spojené s datovým tržištěm

- Při zavádění nového IT firmy je ve většině případů nutné **realizovat nové DMA**.
- Při zásahu do DMA je v případě změn dimenzí nutné provést **opětvornou agregaci dat**.
- Oproti klasickým relačním databázím je DMA **paměťově náročnější** a dochází ve většině případů k delší době odezvy.
- Data, uložená v DMA, **nelze modifikovat**.
- Uživateli je zpřístupněn **pouze omezený soubor dat**, což může u zaměstnance vést ke špatnému pochopení a k dezinformaci kontextu podnikových dat.

## 10.3 Dočasné úložiště dat, DSA

**Úkolem** dočasného úložiště dat (**Data Staging Area, DSA**) je dočasné uložení extrahovaných dat z produkčních databází s cílem **zajistit jejich přípravu a potřebnou kvalitu** před vstupem do datového skladu. Pro data v dočasném úložišti je podstatné, že jsou to **data detailní, neagregovaná, často nekonzistentní, bez časové dimenze**. DSA obsahuje **pouze aktuální data**, tj. po jejich zpracování v DSA a přenosu do datového skladu nebo tržiště se z DSA odstraní.

### 10.3.1 Efekty a přínosy DSA

- Představuje **uvolnění výkonu** primárních systémů.
- Urychluje proces **extrakce** dat.
- Nejedná se o **povinnou** komponentu.
- Snižuje celkové **náklady** na provoz DWH.

### 10.3.2 Problémy a omezení spojené s DSA

- Data **nejsou agregována** ani historizována.
- Data jsou **nahrazena** po jejich zpracování další dávkou dat.
- Zvyšuje **náklady** na realizaci DWH.

## 10.4 Operační datový sklad, ODS

Operační datový sklad (**Operational Data Store, ODS**) je komponenta datové vrstvy, která je charakteristická následujícím **využitím**:

- Jedná se o **jednotné místo datové integrace** aktuálních dat z primárních systémů.
- Podporuje **relativně jednoduché dotazy** nad menším množstvím aktuálních dat.

#### 10.4.1 Efekty a přínosy operačního datového skladu

- Představuje jednotné místo **datové integrace** aktuálních dat.
- Nabízí možnost **rychlého dotazování**.
- Představuje zdroj pro sledování konsolidovaných, agregovaných dat **s minimální dobou odezvy** po zpracování.
- Podporuje **dotazování** nad omezeným množstvím aktuálních dat.
- Obsahuje **aktuální záznamy** vybraného množství dat.
- Podporuje **interaktivní komunikaci** se zákazníkem.
- Umožňuje schopnost napojení na **integrační platformy**.
- Je zdrojem pro **reportingové nástroje**.
- Nejedná se o **povinnou** komponentu.
- Obsahuje **konzistentní** a v některých případech agregovaná data.

#### 10.4.2 Problémy a omezení, spojené s operačním datovým skladem

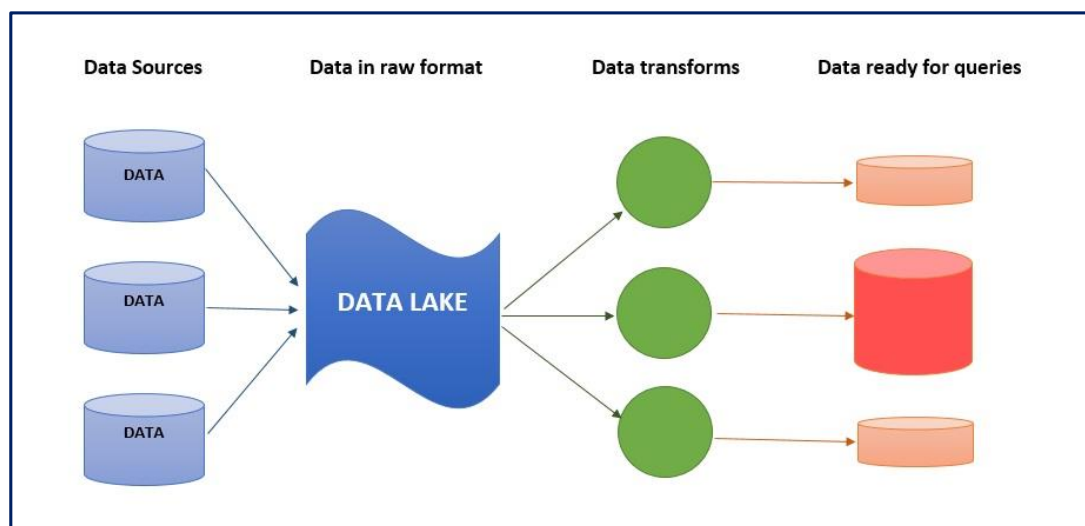
- Zvyšuje **náklady** na realizaci DWH.
- Neposkytuje pohled na data **v časovém srovnání**.
- Je často náročné na **integraci** se zdrojovými systémy.

### 10.5 Data Lake

Vzhledem k **množství nestrukturovaných dat** a jejich nárůstu v poslední době bylo potřeba umět s těmito daty efektivně pracovat a nakládat. Právě z tohoto důvodu **vznikl Data Lake** (Ševčík, 2020). Toto řešení zatím není úplně dokonalé a očekává se zlepšení do budoucnosti, aby nedocházelo k takzvaným **data swamps**, ale také, aby práce s daty, v uloženými v Data Lake, byla jednodušší a efektivnější.

Data Lake je **úložištěm veškerých podnikových dat** i těch **v tak zvané surové podobě**. Jedná se tedy o data, která nebyla zatím nijak zpracována a jejich další účel nemusí být zatím znám nebo určen. Jedna společnost může mít v tomtéž repositáři **uložena strukturovaná data**, například ze svých relačních databází, **semistrukturovaná data**, jako jsou například soubory, vytvořené pomocí Microsoft Excel, ale **i nestrukturovaná data**, u kterých si jako příklad můžeme uvést obrázky nebo videa [Talend, 2020]. Data Lake se tím stává **centrem datové integrace** podniku, které ale zároveň poskytuje schopnost dotazování a získání tím námi požadovaných informací nad velkým množstvím dat v reálném čase.

Obrázek 10-2 dokumentuje, že při ukládání dat do Data Lake je **využíván převážně proces ELT** (Extract, Load, Transform), díky kterému by mělo být umožněno nejprve nahrání a **uložení dat s následnou možností provedení více nezávislých transformací** nad stejnými daty. Data, která jsou ukládána, nemají nijak určený specifický formát pro jejich uložení a mohou vstupovat v reálném čase, stejně tak jako v dávkových vstupech [DARKWAH, 2020].



**Obrázek 10-2: Způsob práce s daty v data lake (Zdroj: Sameh, 2020)**

Je nutné zajistit, že Data Lake bude vhodně spravován a surová data budou mít požadovanou kvalitu. **Nároky na správu dat** jsou poměrně vysoké, protože se v Data Lake pracuje s obrovským objemem dat, která mají nejrůznější struktury. Pokud by došlo k tomu, že by data nebyla správně spravována nebo nebyla spravována vůbec a Data Lake by tím **přestal poskytovat požadovanou a zamýšlenou hodnotu** informací, respektive by nebylo možné jej nadále efektivně využívat, tak se jedná o Data Swamp [KNIGHT, 2020].

**Nezákladnější rozdíly** mezi Data Warehouse a Data Lake jsou následující:

- **Data Lake** obsahuje **surová data**, u kterých často **není určen účel**, za kterým jsou ukládána. Může tak dojít k tomu, že budou uložena data, která ani nikdy nebudou reálně použita. **Data Warehouse** je určen pro **strukturovaná data**, která již byla anebo jsou používána a zpracována [Talend, 2020].
- Rozdílem je také **míra nahrání a uložení dat**. Jsou využívány zejména **dva procesy**. Pro Data Lake je to již zmiňovaný proces **ELT** a pro Data Warehouse proces **ETL** (Extract, Transform, Load) [DARKWAH, 2020]. Základním rozdílem je zde to, že při ETL procesu dochází k transformaci dat do určitého jasně daného formátu ještě před nahráním do Data Warehouse. U ELT dochází po extrakci dat k jejich nahrání do Data Lake a až poté k případné transformaci. Zároveň proces **ELT umožňuje provádět více různých transformací**, využívajících stejná data [PEARLMAN, 2020].
- Jelikož **Data Lake** nemá prakticky jasnou strukturu, je **mnohem snazší jej měnit** a jakékoliv změny dat mohou být provedeny velmi rychle a nenáročně. U **Data Warehouse** je tomu právě naopak. Každá změna, provedená na datech, je velmi náročná a nákladná.
- Poslední základní rozdíl je **spektrum uživatelů**, schopné využít a porozumět datům, uloženým v jednom či druhém úložišti. Data, uložená v **Data Lake**, jsou určena především **pro informační specialisty** z důvodu náročnosti porozumění nezpracovaných surových dat. Na druhou stranu datům, uloženým v Data Warehouse, by měl být schopen porozumět uživatel. Jedná se například o grafy nebo tabulky [Talend, 2020].

### 10.5.1 Efekty a přínosy Data Lake

- Data Lake poskytuje podstatně **větší flexibilitu** při zpracování analytických úloh.
- Představuje jednotné místo **datové integrace** z původních oddělených komponent dočasného i operativního úložiště dat.
- Nabízí možnost **rychlého dotazování** prakticky v reálném čase s využitím kvalitních uživatelsky orientovaných nástrojů.
- Zahrnuje řízení a **zpracování velkých objemů** z různých typů a zdrojů dat.
- Představuje zdroj pro sledování konsolidovaných, agregovaných dat **s minimální dobou odezvy** po zpracování.
- Zajišťuje požadovanou **úroveň kvality** komplexu dat pro podnikovou analytiku, s odpovídajícími nástroji a funkcemi pro čištění a konsolidaci dat.
- Umožňuje ukládání, zpracování a využívání **i nestructurovaných dat**,
- Podporuje jak **dávkové vstupy** dat, jak **vstupy v reálném čase**.
- Podporuje **interaktivní komunikaci** se zákazníkem.
- Umožňuje schopnost napojení na **integrační platformy**.
- Je zdrojem pro **reportingové nástroje**.
- Obsahuje **konzistentní** a v některých případech agregovaná data.
- Zahrnuje i analytické nástroje a nástroje **pro Data Governance**.
- V jistých případech může Data Lake **nahrazovat i datový sklad** a integrovat tak všechny tři původní komponenty a vytvořit tak Logický datový sklad.

### 10.5.2 Problémy a omezení spojené s Data Lake

- Představuje relativně vysoké **nároky na Data Governance** s ohledem na rizika spojená s velkými objemy dat a jejich složitými strukturami.
- Při nezvládnutém řízení Data Lake může **docházet k problémům s jeho efektivním využitím**.
- Musí být jasně nastavená pravidla pro **řízení změn v produkčních zdrojích**, včetně zodpovědností za poskytování informací o plánovaných a realizovaných změnách.
- Znamená vysoké nároky na kvalitní **provázání s celým systémem řízení dat** v podniku.

## 10.6 Závěry k datovým zdrojům v podnikové analytice



- Analytické databázové systémy se **liší svojí mohutností a pozicí** v architekturách BI řešení.
- Jednou z klíčových otázek je, zda bude systém postaven na **jednom centralizovaném datovém skladu** nebo na **datových tržištích**, resp. **kombinaci** datového skladu a datových tržišť.
- Další aktuální variantou může být **vytvoření „Data Lake“**, integrující všechny analytické databáze nebo jejich části.
- Analytické databáze mají řadu svých specifik a je tedy účelné disponovat **specialisty** pro tato řešení.
- S ohledem na různé účely je nutné dobře rozlišit řešení **normalizace dat** na úrovni datového skladu i datových tržišť.



# 11. Transformace dat



**Účelem** kapitoly je:

- vymezit základní **charakteristiky těch součástí BI systémů**, které se podílejí na transformacích dat mezi zdrojovými a analytickými databázemi, resp. mezi analytickými zdroji navzájem,
- vytvořit podklad pro **analýzu možností a nároků** jednotlivých typů produktů pro datové transformace, zejména při řešení architektury BI systémů a výběru nejvhodnějších technologií a řešení,
- poskytnout základní informace o funkcích, souvisejících s procesem transformace, tj. o možnostech **efektivního zachycování změn ve zdrojových systémech** a o řešení **historizace dat**.

## 11.1 Extract Transform Load (ETL/ELT)

**ETL (Extract, Transform, Load)** je jednou z nejvýznamnějších komponent celého komplexu Business Intelligence. Běžným označením pro prostředky ETL je rovněž **datová pumpa**. Jejím **úkolem** je data ze zdrojových systémů **vybrat (Extract)**, **upravit** data do požadované formy a uspořádání (**Transform**) a **nahrát** je do specifických datových struktur, resp. datových schémat datového skladu nebo tržiště (**Load**). Tyto nástroje lze tedy použít pro **přenos dat mezi dvěma (či více) libovolnými databázemi nebo datovými soubory** (textovými, tabulkovými, XML soubory atd.).

ETL nástroje pracují obvykle **v dávkovém režimu**, data jsou tedy přenášena najednou v určitých časových intervalech, např. denních, týdenních apod. Právě transformace dat, spojené s ETL, jsou **pracovně, časově i finančně nejnáročnější** a obvykle představují cca 60 % vynaložených pracovních kapacit. Pro úspěšné BI řešení však znamenají zcela nezbytný předpoklad.

Pro ETL jsou **podstatné** následující **charakteristiky**:

- ze zdrojových databází jsou obvykle vybrána pouze taková **data**, která jsou určena pro **analytické, plánovací a rozhodovací aktivity** podniku, což je jedním z prvních úkolů analytiků Business Intelligence,
- data jsou transformována do nových datových struktur **analytických databází**, které musí být předem navrženy tak, aby nejlépe **odpovídaly potřebám řízení podniku**. To souvisí s možnostmi využití multidimenzionality a granularity dat,
- data datových skladů nebo tržišť **vstupují z různých zdrojových databází** (ERP, e-Business, CRM atd.), přičemž v těchto různých zdrojích mohou být jedna a **tatáž data uložena vícekrát** (a k tomu ještě různě), např. různé databáze prodejců, zákazníků apod. Do analytických databází však musí vesměs vstoupit pouze jednou. V transformační vrstvě (s využitím ETL) **musí dojít ke konsolidaci dat**, tj. určení vstupujících dat s vyloučením duplicit či multiplícit,
- se zajištěním konsolidace dat úzce souvisí i **dosažení potřebné kvality dat**, tj. vyloučení chyb, nepřesností atd.

### 11.1.1 Efekty a přínosy ETL / ELT

- Díky ETL lze z množiny nestructurovaných dat **vyjmout** právě ta **data**, která mají být **relevantní** vzhledem k zákaznickým požadavkům na podporu řízení, resp. k řešení datového skladu (DWH).
- Velké množství ETL nástrojů umožňuje **flexibilní přístup** a schopnost identifikovat právě nástroj, který nejvíce odpovídá řešení konkrétní úlohy.
- Pomocí ELT metody lze rozdělit DWH procesy na **diverzní části**, které následně dělí projekt na subprojekty a nabízí tak transparentní pohled na celkový proces.
- **Čištění dat** a odstranění duplicitních dat zajišťuje řešení chyb, způsobených lidským faktorem atd.
- ETL procesy v běžném případě probíhají **ve stejné časové frekvenci** a ve stanoveném čase (většinou v noci). Tento způsob nasazení je zcela vyhovující pro většinu operativních analýz.
- Podporuje **speciální typy analýz**, vyžadujících okamžité vyhodnocení, např. fraud systémy.
- **Centralizovaná správa** transformací bývá uložena v repozitáři.
- Nabízí se možnost vytváření a **správy metadat**.
- Poskytuje řízené **plánování a spouštění transformací**.
- Správně navržený ETL systém **zaznamenává všechny chyby** do chybového žurnálu.

### 11.1.2 Problémy a omezení, spojené s ETL / ELT

- Modifikace původní ETL procedury představuje **nárůst nákladů a času** na její řešení.
- V případě změny ETL technologie je nutná **znalost aktuálního i původního skriptovacího jazyka**.
- Pro řešitele je nezbytně nutná **znalost transformačních technik**, metodik a funkcionalit.
- Při **nedodržení pravidel**, uplatňovaných při ETL (tj. funkcionalita a metodika), je princip ETL prakticky nevyužitelný a stává se při opětovném využití nepřehledným (nečitelná dokumentace způsobuje nepřehledný systém ETL procedur).
- Existuje vysoká **míra závislosti na analýze** primárního systému.
- Je třeba řešit **technologické bariéry** (např. systém není schopen transformovat zdrojový datový formát).

## 1.2 Zachycení změn ve zdrojových systémech, CDC

Účelem fungování **Change Data Capture (CDC)** je **schopnost zachytávat změny v datech ve zdrojových systémech**. Snahou použití CDC je získávat tyto změny efektivně, a pokud je to možné, dosáhnout co nejmenšího zatížení těchto zdrojových systémů. Získání zdrojových dat by tedy nemělo mít **žádný anebo pouze zanedbatelný**

**dopad do běžného používání zdrojových systémů**, a to zejména tak, aby nebyla narušena jejich primární funkce (Snítil, 2018).

Zachycení změn ve zdrojových systémech v architektuře datového skladu se obvykle nachází **mezi zdrojovými systémy** (zejména jejich databázemi) **a dočasným úložištěm** dat.

CDC **nabízí alternativu k použití ETL**, které vždy **zatíží zdrojovou databázi**, a to z důvodu, že se do zdrojové databáze připojuje **přes konektor**, z kterého čte data (nehledě na to, zda se jedná o nativní konektor konkrétní databáze nebo o generický konektor, jakým je například ODBC či JDBC). Naproti tomu **CDC** se dokáže vyhnout tomuto zatížení, protože **dokáže změny provedené v datech vyčítat z databázových logů**, obvykle se tedy nevytváří žádné databázové připojení, databáze nemusí tvořit exekuční plán a provádět načítání dat z disků.

Jedním z nástrojů, které je možné použít pro implementaci tohoto rozšiřujícího konceptu, je například *IBM InfoSphere Change Data Capture*. Ve většině případů tedy **není potřeba nahrávat data z těchto databází a z jejich diskových úložišť**, ale pomocí znalosti databázových logů lze odvodit, jaká data byla jak změněna a na základě toho vyvodit potřebný stav v cílovém úložišti.

### 11.2.1 Efekty a přínosy zachycení změn ve zdrojových systémech

- Uplatněním konceptu se lze **vyhnout** výraznému **zatížení zdrojových systémů** a zároveň je možné potřebná data zpřístupnit pro další zpracování.
- Efektem je relativně **rychlé zachycení změn** ve zdrojových systémech, takže tyto změny mohou být velmi rychle zapisovány do cílového systému a zpřístupněny tak k dalšímu zpracování.
- Nasazení specializovaného CDC nástroje v konkrétním prostředí je velkou výhodou zejména v případě, kdy zvolený **nástroj umí změny získávat ze všech zdrojových systémů**, které jsou provozované v daném prostředí.
- Nasazení specializovaného nástroje vede ke **zjednodušení používaného řešení**, kdy místo různých způsobů získávání dat ze zdrojových systémů by existoval jednotný způsob získávání zdrojových dat. Správa takového řešení a jeho provoz by tak mohly být efektivnější.

### 11.2.2 Problémy zachycení změn ve zdrojových systémech

- Uvedený způsob získání datových změn ze zdrojových systémů **na první pohled nepřináší výrazné nevýhody**.
- Určitým rizikem a omezením jsou případné **změny ve strukturách tabulek na zdrojovém systému**. V takovém případě může dojít k zastavení přenášení změn ze zdrojového systému do cílového a je nutné provést znovu mapování dat.

### 11.3 Historizace dat

Účelem historizace je zajistit **potřebu udržovat historickou kolekci dat v cílovém systému** (v datovém skladu) a podle toho i data vhodně transformovat (Snítíl, 2018).

**Koncept Historizace** se zabývá splněním části **definice** datového skladu Williama Inmona v tom smyslu, že *datová kolekce je historická*. To znamená, že je možné **určit, kdy, která data byla či jsou platná**. Z jiného úhlu pohledu lze také říci, že **všechna data** v datovém skladu jsou určitým způsobem **identifikována prvky dimenze času**.

Jelikož se jedná o **základní funkčnost** datového skladu (ať v užším či širším pojetí), **existuje mnoho řešení**, jak splnit tuto definiční podmínku. V tomto kontextu jsou podstatné **možnosti historizace po použití konceptu CDC**, tj. jaký dopad bude mít pro datový sklad situace, kdy přicházejí ze zdrojových systémů **pouze provedené změny**. Tím tento koncept logicky navazuje na rozšiřující koncept CDC.

V takovém případě **lze výstup z CDC považovat za první vrstvu datového skladu**, kterou je možné označit jako první nultou vrstvu **L0 (z anglického slova layer)**. Z této vrstvy je potřeba **data transformovat do další vrstvy L1**, která již obsahuje historickou kolekci dat. Pokud zanedbáme přenosové a technologické zpoždění mezi zdrojovým systémem a vrstvou **L0**, tato vrstva obsahuje **identický stav dat jako je na zdrojovém systému**. Naopak **vrstva L1** obsahuje **poslední známý stav zdrojových systémů**, provedený v rámci **poslední transformace**. Tím, že nemáme k dispozici informaci, která data byla změněna v rámci vrstvy L0, musí se během této transformace L0 do L1 provést **porovnání dat L0 vůči poslednímu dostupnému stavu dat ve vrstvě L1**.

Toto **porovnání, nutné pro identifikaci změn** ve zdrojovém systému vůči historické kolekci ve vrstvě L1, je výkonově poměrně náročné a je spojené s dalšími možnými nevýhodami, které jsou podobné jako v případě použití celkového exportu dat ze zdrojových systémů.

**Další možností** je ta, kdy každý **přenesený záznam obsahuje časovou značku** určující čas změny na zdrojovém systému, ke které se daný záznam váže. Oproti předchozím dvěma mapováním přenáší do cílového systému **pouze změny ze zdrojového systému**. Při použití tohoto mapování **nelze jednoduše zjistit stav zdrojového systému**, ale tento stav se musí ze získaných změn dopočítat.

#### 11.3.1 Efekty a přínosy historizace

- Celý koncept je použitelný a **umožňuje zpracování dat z výstupů CDC**, kde jsou obsaženy pouze změny provedené na zdrojovém systému do další vrstvy, která reprezentuje celkový obraz dat na zdrojovém systému včetně historie.
- Je možné dosáhnout značné **univerzálnosti celého konceptu**, kdy nehledě na zdrojový systém je možné vytvářet relativně jednoduše historické kolekce dat. Tyto historické kolekce dat poté mohou sloužit jako vstup pro následné transformační úlohy, které jsou obvykle pracnější a časově náročnější na vytvoření. Zvláště v případech, kdy tyto transformační úlohy jsou teprve vyvíjené, lze jim po jejich vytvoření a otestování nabídnout historická data a doplnit tak i zpětně nové datové struktury v dalších konsolidovaných vrstvách řešení datového skladu.

- **Možnost zpracování dat k určeným časům**, tj. pokud je pro specifické potřeby nezbytné vycházet z dat platných k určeným časovým bodům například z důvodů porovnatelnosti reportů, nebo z potřeb sledovat vývojové trendy, toto řešení takové zpracování umožňuje, a to i zpětně, kdy s využitím auditních dat lze dopočítat stav zdrojového systému k danému okamžiku.
- Data, zpracovaná pomocí tohoto konceptu, lze také **použít pro důkladné a/ nebo velmi složité analýzy**. Tato data totiž obsahují historii dat zdrojových systémů a zároveň jsou ve stejné podobě jako data zdrojových systémů. Je zde tedy minimální riziko možné ztráty určité informace, které existuje v obvyklých transformačních a agregačních datových úlohách. Také je zde minimalizované riziko možných chyb, které mohou být omylem způsobené transformačními úlohami.

### 11.3.2 Otázky a problémy historizace

Při použití tohoto rozšiřujícího konceptu je **potřeba zvážit četnost zpracování** vzhledem k výpočetní náročnosti. I když v tomto případě nedochází k porovnání dvou plných snímků, tedy k celkovému stavu dat ve zdrojových systémech a poslednímu dostupnému stavu dat v historické kolekci, ale jsou v tomto případě zpracovávány pouze provedené změny na zdrojovém systému, určitá **výpočetní náročnost na udržování historické kolekce** zde existuje, zvláště pak v porovnání s řešeními, které historickou kolekci dat zdrojových systémů neudržují. V každém případě lze doporučit vhodně využívat optimalizace, které zvolená technologie pro zpracování dat nabízí.

Určitým rizikem, které je spojeno s tímto rozšiřujícím konceptem, je zejména v dlouhodobém horizontu **nárůst objemu uchovávaných dat**. Před zavedením tohoto konceptu je vhodné provést analýzy za účelem odhadnutí datového objemu historické kolekce a tempa jejího růstu. Řešením tohoto problému může být **přesouvání starších dat z historické kolekce do jiného systému**, který nabízí relativně levné uložení těchto dat. V případě potřeby je možné i historickou kolekci dat odmazávat a udržovat v ní například pouze poslední rok, aby bylo dosaženo optimálního poměru mezi cenou datového úložiště a přidanou hodnotou v ní obsažených dat.

## 11.4 Závěry k transformacím dat v podnikové analytice



- Transformace dat patří k pracovně i časově nejnáročnějším aktivitám v rámci řešení podnikové analytiky. Je tomu tak nutné věnovat **v harmonogramech dostatečný časový prostor**.
- V průběhu transformací dat se řeší i nezbytné **kontroly jejich kvality** a odpovídající **čištění** dat, neboť datová analytika je na kvalitu dat značně citlivá, např. na úrovni dočasného úložiště dat.
- Jednou z otázek je, **jakými prostředky** bude efektivní transformace realizovat, tj. integrovanými prostředky v rámci databází (např. MS Integration Services), specializovanými prostředky (např. Informatica), případně vytvořením vlastních nástrojů, např. na bázi SQL skriptu.

- V současnosti se nabízejí i prostředky pro **automatické zachycování změn** ve zdrojových systémech, označených jako **CDC, Change Data Capture**.
- V souvislosti s transformacemi dat je nutné posoudit i nároky na řešení **historizace dat**, tj. zachycení jejich hodnot v průběhu časového vývoje.

## 12. Datová analytika



Účelem kapitoly je:

- vymezit základní **charakteristiky součástí BI systémů**, které se vztahují přímo k realizaci analytických funkcí, kde obvyklou základní komponentou jsou OLAP databáze,
- současně i poskytnout základní informace o technologiích **in-memory analytics** jako o jedněch z výrazných trendů podnikové analytiky,
- vytvořit podklad pro **analýzu možností a nároků** jednotlivých typů technologií a produktů pro analytické funkce,
- poskytnout základní informace o funkcích a možnostech **uživatelského testování analytiky, tzv. sandbox**.

### 12.1 OLAP databáze

**OLAP databáze** představují jednu nebo několik souvisejících a **vzájemně propojených OLAP kostek**. Ty většinou, na rozdíl od datových skladů, již **zahrnují předzpracované agregace dat** podle definovaných hierarchických struktur dimenzí a jejich kombinací (viz kapitola 5.1.2).

#### 12.1.1 Efekty a přínosy OLAP

- Nabízí využití **multidimenzionálních analýz**.
- Výhodou je transparentní pohled na **aktuální data**, sloužící jako podklad při rozhodování.
- Umožňuje **podporu reportingu**.
- Představuje vstup pro **realizaci metod** (např. BSC, dynamický BSC atd.).
- Podporuje automatizaci **plánovacích procesů**.
- Umožňuje **uvolnění lidských zdrojů**, kontrolu vývoje reality a porovnání s plánem.
- Disponuje funkcemi **drill down, drill up, drill through**.
- Umožňuje předdefinovat **jednotnou obchodní logiku**, jako například způsob výpočtu ceny produktu apod.
- Disponuje **agregovanými ukazateli**.
- Nabízí **předdefinovanou funkcionalitu** (např. identifikaci KPI).
- **MOLAP**: Je navržený pro rychlé přijímání dat, podporuje slice & dice operace, obsahuje předpřipravené kalkulace a dále je specifický rychlou odezvou při dotazování.
- **ROLAP**: Je specifický podporou rozsáhlého množství dat, kdy rozsah závisí na velikosti zdrojové databáze, umožňuje využití funkcionality zdrojové databáze.
- **HOLAP**: Zahrnuje kumulaci a možné kombinace výše popsanych výhod.

#### 12.1.2 Problémy a omezení, spojené s OLAP databázemi

- **MOLAP**: Je charakteristický prací s omezeným rozsahem dat, kdy dochází ke generování kalkulací již při samotné tvorbě OLAP kostky.

- **ROLAP:** Omezení ROLAP vychází z SQL funkcionality zdrojové databáze, oproti MOLAP disponuje horší dobou odezvy při SQL dotazování.
- **HOLAP:** představuje složitější řešení při kumulaci a kombinaci výše popsaných nevýhod.

## 12.2 In-Memory Analytics

Analytická funkcionalita v paměti, resp. **analytické aplikace realizované v operační paměti** (*In Memory Analytics*), je nabízena jako technologie, resp. součást nových produktů, umožňující realizovat multidimenzionální analýzy, a to i s detailními daty, uloženými v operační paměti. Je to pojetí konceptu BI, které **pro zvýšení výkonu** ve své architektuře používá **komponentu in-memory databázi**.

Existuje několik přístupů k in-memory koncepci. Prvním přístupem je **kompletní in-memory řešení**, nebo **integrace in-memory databáze** do stávajícího řešení BI.

### 12.2.1 Efekty a přínosy In-Memory Analytics

- Umožňuje díky výkonu in-memory databáze analyzovat data doslova **v reálném čase**, s dobou odezvy pouhých několika sekund. Hodí se pro zpracování velkých objemů dat a pro složité výpočty nad těmito daty.
- Pomáhá **odstínit zátěž vedlejších i zdrojových systémů**. Různé systémy disponují různou dobou odezvy, která se navíc mění se stupněm zátěže těchto systémů. Přidáním komponenty in-memory databáze do řešení BI se tak značně odlehčí zátěž vedlejších systémů, včetně systémů zdrojových.
- Představuje nastartování procesu **zlepšení datové kvality**. Dostupnost informací je základním předpokladem datové kvality. Díky zvýšení výkonnosti analýzy dat uživatel získá informace v požadovaném čase (nebo alespoň výrazně aktuálnější informace).
- Nabízí **jednodušší provoz a údržbu** systému. Jelikož analytické aplikace pracují nad obrovským množstvím dat, vyžadují vysoký výpočetní výkon. S kompletním in-memory řešením není již potřeba složitě budovat architekturu BI po výše uvedených vrstvách (nákup HW, SW, implementace, integrace aplikací a dat, ladění systému). In-memory řešení je možné získat od jediného dodavatele v mnohem kratším čase a tím snížit náklady na provoz a údržbu systému.
- **Standardizované BI aplikace** – s příchodem koncepčně „jednoduchého“ in-memory řešení BI není třeba vyvíjet aplikace šité zákazníkovi na míru, nebo vyvíjet kompletní řešení na „zelené louce“. Již dnes několik dodavatelů nabízí hotové řešení BI pro standardizované systémy typu ERP, CRM a další. Ty obsahují předdefinované komponenty pro jednotlivé vrstvy (ETL, vlastní datový model, sady reportů a dashboardů).
- **Masové rozšíření BI** je také jednodušší díky nasazení, nárůstu výkonu i novému typu BI aplikací; postupně dojde k rozšíření řešení BI mezi širší okruh uživatelů. Vzhledem k flexibilitě škálování řešení si ho budou moci v budoucnu dovolit i malé a střední podniky.



## 12.2.2 Problémy a omezení, spojené s In-Memory Analytics

- **Cena nejvýkonnějších řešení** zůstává díky vysoké ceně operačních pamětí prozatím velmi vysoká. Díky tomu si mohou in-memory BI pro zpracování opravdu velkých objemů dat dovolit jen velké společnosti. Na druhé straně se však in-memory používá i pro self service BI, a tedy i pro menší řešení.
- Pro úspěšnou implementaci řešení je potřeba zvolit **vhodný přístup pro danou situaci**, ve které se podnik nachází. Na základě potřeb podniku je nezbytné vybrat komponenty řešení tak, aby plně vyhovovaly jeho potřebám.
- Využití in-memory databází **bez použití energeticky nezávislých operačních pamětí** typu NVRAM s sebou nese zásadní nedostatek takového řešení. Tím je potřeba nahrání dat do in-memory databáze ze zdrojových systémů nejen při plánované aktualizaci, ale i při výpadku systému.

## 12.3 Sandbox

Poskytuje prostor pro **uživatelské zkoušení, prototypování** a případně testování nových nástrojů a analytických úloh (Slánský, 2018).

### 12.3.1 Efekty a přínosy Sandbox

- Nabízí možnost **rychlého zkoušení a dotazování** s využitím nových uživatelsky orientovaných nástrojů.
- **Nejedná se o povinnou** komponentu.
- S větším využitím podporuje efektivně **zvyšování kvalifikace a znalostních předpokladů** uživatelů pro další běžný provoz v podnikové analytice.

### 1.3.2 Problémy a omezení, spojené se Sandbox

- Sandbox **nemůže být využíván v běžném** provozním režimu.
- **Nepodporuje obvyklou dokumentaci** ani běžně předpokládanou datovou kvalitu dat.
- **Nepodporuje backupy** a obnovu dat po výpadcích.
- **Nenabízí obvyklou úroveň zajištění bezpečnosti** zpracování a využívání dat.
- Je otázkou, **jaká data a s jakými zodpovědnostmi** do Sandboxu ukládat.

## 12.4 Závěry k analytickým nástrojům



- Uvedené prostředky a přístupy představují ty, které bezprostředně **realizují analytické operace** s daty.
- Jádrem prostředků analytických operací jsou **OLAP databáze** a jejich různé formy (viz výše).
- Významným příspěvkem ke zvyšování výkonu analytických operací je koncept **in-memory analytics**.
- Velmi častým problémem je zajištění **přípravy uživatelů** s možností odzkoušení si některých analytických funkcí, k tomu se využívá prostředí, označované jako „**sandbox**“.

## D) Reporting a dashboardy

Oddíl se zabývá dvěma součástmi na prezentační úrovni úloh podnikové analytiky:

- **Část věnovaná reportingu** zahrnuje **členění uživatelů** reportů a jejich potřeby, **kategorizaci** reportů podle vybraných hledisek podnikového řízení a vymezení **potenciálních efektů**, které kvalitní reporting přináší firmě, na druhé straně případných **problémů**, které jsou s řešením reportů a jejich organizací spojeny.
- **Dashboardy** lze charakterizovat jako **grafické vyjádření sledovaných metrik**, významných pro řízení firmy. Jsou prezentovány formou grafů, tabulek, budíků a jiných grafických ukazatelů a využívají se např. při řízení podnikové výkonnosti (CPM). Kapitola obsahuje **podstatné charakteristiky** dashboardů, jejich **klasifikaci**, **možnosti zobrazení** a rovněž hlavní **efekty** pro firmu a specifikace případných **problémů** s jejich řešením.

## 13. Reporting



**Účelem** kapitoly je:

- nastítnit **klasifikaci** různých možností a forem reportingu,
- analyzovat funkce a možnosti reportingu z pohledu jejich **potenciálních efektů** a na druhé straně **omezení** jako podklad pro efektivní návrhy reportingových řešení.

Reporting slouží jako **systematický zdroj informací a ukazatelů**, vypovídající o činnosti společnosti pro své uživatele, kterých v podniku mohou být desítky, stovky nebo i tisíce. Za tímto zdrojem informací pro uživatele stojí většinou **dotazy na data** do databází firmy.

Každá skupina uživatelů nebo dokonce každý jednotlivý uživatel mají **různé požadavky** týkající se obsahu, formy i času distribuce výkazů, které ke své úloze ve firmě potřebují. Důležitou součástí je zajištění takové **distribuce**, která pracovníkům ve firmě zajistí **přístup pouze k jim relevantním údajům** a bude chránit důvěrná data před nežádoucím šířením, dále podle (Fibírová, J. a další, 2015).

### 13.1 Uživatelé reportingu

Uživatele lze rozdělit na **dvě základní skupiny**:

- **Interní uživatelé** zahrnují manažerské a další interní pozice ve firmě.
- **Externí uživatelé** přistupují pouze k externím informacím – tedy k informacím, které podnik zveřejňuje. Zpravidla se jedná o externí výkazy. Patří sem např. spolupracující firmy – dodavatelské a odběratelské společnosti, banky, státní orgány, například finanční úřady, správy sociálního zabezpečení, krajské úřady, zastupitelské orgány obcí a měst, vlastníci podniku a další.

### 13.2 Kategorizace reportů

Tato část se zabývá kategorizací reportů, která může pomoci poznat jejich účel a obsah podrobněji.

#### 13.2.1 Interní reporting

Reporting se z časového hlediska dělí na standardní a mimořádný (Fibírová, Šoljaková 2010):

**Standardní reporting** se vyznačuje zprávami dodávanými **v pravidelných časových úsecích** – obvykle jednou za měsíc, za čtvrt roku nebo za celý rok. **Struktura** zprávy z hlediska obsahu informací, výpočtů nebo analýz a podobně je **předem stanovená**. Výjimkou nejsou reporty týdenní nebo i denní. Pracovníci reportingu by však měli u každého periodického reportu **zvážit nákladovou náročnost** generování takového reportu.

**Mimořádný reporting** představuje zprávy **generované na požadavek** nebo reporty, které se běžně negenerují – například analýza rizika, analýza sortimentních skupin a podobně.

Každý podnik může mít jiný **system interních reportů**, lze však často vysledovat **dvě elementární kategorie reportů – základní souhrnné** (či přehledové) výkazy a **dílčí** zprávy (Fibírová a Šoljaková 2010).

- **Souhrnný reporting** podává přehled o činnostech podniku **za určité období**; do jeho obsahu tradičně patří **základní finanční ukazatelé** včetně srovnání těchto hodnot s hodnotami plánovanými nebo s hodnotami minulé časové periody. Fibírová a Šoljaková (2010) zde zdůrazňují důležitost doplňujícího komentáře, jež se zabývá rozdíly mezi plánem a skutečností a analýzou, která hledá odpovědi na to, proč odchylky nastaly.
- **Dílčí reporting** se věnuje **konkrétnějším informacím**, struktura více odpovídá daným požadavkům pracovníků podniku. Jejich potřeby se navíc liší od odboru k odboru, ve kterém svou práci vykonávají (obchod, výroba, marketing a podobně).

### 13.2.2 Externí reporting

**Povinnost ze zákona** předkládat výsledky hospodaření formou **auditovaných finančních výkazů** představuje velkou část externího reportingu. Jde o standardní **periodické výkazy o hospodaření**, které vycházejí jednou ročně, tedy o rozvahy, výsledovky a výkazy peněžních toků. Důležitým prvkem externích výkazů je úplnost, srovnatelnost a spolehlivost údajů.

### 13.2.3 Další kategorie reportů

Samotné reporty se dělí na **standardní a ad hoc**. Podobné dělení již bylo zmíněno, zde se však neklade důraz na časovou dimenzi reportů, nýbrž na dimenzi uživatelskou; přesněji řečeno na míru, se kterou může uživatel ovlivňovat datový obsah reportu.

**Ad hoc zprávy** si může vytvořit sám uživatel vytvořením konkrétního jednorázového dotazu nad databázemi.

**Standardní reporty** lze ještě dále členit na statické a dynamické. Uživatel si **statický report** již nemůže upravovat, kdežto s daty **v dynamickém reportu** může manipulovat podle svých potřeb. Je však vždy omezen nástroji, které mu byly ve zprávě vyhrazeny. Nejčastěji se používají filtry. Uživatel si tak může například vymezit jiné časové období, než je v reportu defaultně nastaveno, může si nechat vypsat data se společnými parametry.

**Podle obsahové stránky** dat, jež předurčuje i uživatele zprávy, **se rozlišují kategorie datové domény**. Mezi ně patří *finance, prodej, marketing, zákaznický servis, nákup, sklady, lidské zdroje, výroba, informační technologie a další*.

### 13.3 Efekty a přínosy reportingu

- Umožňují **monitoring procesů** ve firmě, s čímž souvisí i vyhodnocování takových dat, na jejichž základě lze **identifikovat problém a jeho příčinu**. Díky těmto informacím se vylepšují a **optimalizují podnikové procesy**. Zkvalitnění podnikových procesů vede ke **zlepšení rozhodování**, optimalizace výkonnosti a určení správného budoucího směru fungování společnosti.
- Kvalitní reporting přináší také **odhalení příčin problémů** pomocí včasných a správných informací s možností prohlížení z různých perspektiv a úrovně detailu.
- Reporting také znamená **zvýšení transparentnosti** ve firmě. Není v rozporu s nežádoucím šířením citlivých informací. Jde o poskytování **úplných relevantních informací** všem, kteří by přístup k nim měli mít, k nimž může přistupovat veřejnost nebo zájmové skupiny (*stakeholders*). To přináší i **zlepšení prestiže firmy**.
- **Jednotný pohled** na firemní realitu na základě konsolidovaných reportů vede ke **zkvalitnění komunikace** mezi zaměstnanci a manažery na všech úrovních řízení.
- **Podpora efektivity** – zaměstnanci mají více času na hlubší analýzy a akce, nemusí ručně konsolidovat informace do pravidelných reportů, pracovníci neřeší tvorbu reportů, ale jejich obsah.
- Na druhé straně možnost pro zaměstnance vytvářet **vlastní reporty bez podpory IT útvarů** radikálně snižuje dobu dodání informací k příjemci.
- Nabízejí možnosti **předdefinované logiky**, díky které nedochází k situacím, kdy výstup dvou reportů tvoří odlišná čísla na základě stejné vypočítávané veličiny.
- **Integrate** reportingu s kancelářskými aplikacemi umožňuje zefektivnění tvorby reportů nejrůznějšího druhu, včetně **exportu reportů** do standardních aplikací pro další zpracování a sdílení. S tím souvisí i **navázání dalších aplikací** na report (např. kliknutím na jméno dodavatele se otevře mapa s jeho sídlem atd.).

### 13.4 Problémy a omezení, spojené s reportingem

- Vysokou pozornost je třeba věnovat kvalitní **strukturalizaci reportů** tak, aby uživatel se např. mohl efektivně pohybovat po různých úrovních detailu obsažených dat a mohl se v nich rychle orientovat.
- Velmi častým problémem je, že reporty se vytvářejí překotně a **ve svém obsahu se různě překrývají**, pracovníci firmy jsou zahlceni někdy i zbytečnými reporty. Příprava každého reportu by se neměla obejít bez **nezbytné analýzy** v uvedeném smyslu.
- V řadě případů nemají pracovníci zejména ve větších firmách souhrnný, centralizovaný přehled o existujících reportech, což vede k **vytváření nových reportů čistě duplicitních**, nebo s částečnou duplicitou.
- Výstup reportingu je vysoce **závislý na kvalitě zdrojových dat**.

- Některé reporty, např. u poboček zahraničních společností, musí **dodržovat mezinárodní standardy**, např. **IFRS** (*International Financial Reporting Standards*) nebo **US GAAP** (*United States General Accepted Accounting Principles*). Díky standardům lze zajistit nejen celistvé vymezení obsahu reportů, ale i takovou strukturu, která bude srozumitelná pro všechny.
- **Finanční náročnost** sofistikovanějších reportingových nástrojů vede někdy k jejich odmítnutí bez vyhodnocení jejich potenciálních efektů.
- Pro efektivní využití reportů je podstatná **kvalitní vizualizace dat**, která však musí dodržovat základní principy, např. při výběru grafů, schémat apod.

### 13.5 Závěry k principům a možnostem reportingu



- Účelem reportingu je data vybírat a **poskytovat vzhledem k uživatelským problémům** a požadavkům.
- Reporty musí nabízet i vhodné **formy prezentace**, založené na celé škále možností **vizualizace** dat.
- V některých případech se věnuje **větší pozornost vizualizaci** před kvalitním obsahem, což může být problém.
- Reporty se různě člení s tím, že základní rozdělení je na **interní a externí**.
- S reportingem souvisí i jejich nezbytná **ochrana** vzhledem k oprávněným jednotlivých uživatelů.
- Častým problémem je i **neúměrné množství** různě se překrývajících, duplicitních reportů, což zatěžuje zejména orientaci uživatelů v jejich využití.

## 14. Dashboardy



**Účelem** kapitoly je:

- nastítnit **klasifikaci a kategorizaci** dashboardů v praxi,
- analyzovat funkce a možnosti dashboardů z pohledu jejich **potenciálních efektů** a na druhé straně **omezení**,
- popsat **doporučované postupy** při řešení dashboardů.

Dashboard je **grafickým vyjádřením sledovaných metrik**, které jsou důležité z pohledu byznysu organizace. Metriky jsou prezentovány formou grafů, tabulek, budíků a jiných grafických ukazatelů. Dashboardy jsou **využívány** např. při **řízení podnikové výkonnosti (CPM)**. Dashboardy by měly respektovat **principy** názornosti a viditelnosti všech sledovaných metrik, umožnit posouzení plánu vs. skutečnosti, poskytnout přehled o zlepšujících se a zhoršujících se metrikách a sledovat výkonnost zaměstnanců.

### 14.1 Princip dashboardů v podniku jako součást podnikové analytiky

Ve spojitosti s dashboardy vyvstává **několik otázek**. Co má správný dashboard obsahovat, aby byl cenný pro byznys uživatele? Jaké dopady má data science na oblast dashboardů, speciálně na prezentaci informací, tak, aby je uživatel vyhodnotil co nejdříve? Co je základem dashboardu? Pro jakou skupinu uživatelů se dashboardy vytvářejí?

Dashboard podle principu **SMART** musí obsahovat **základní vlastnosti**, které jsou nezbytné pro úspěch:

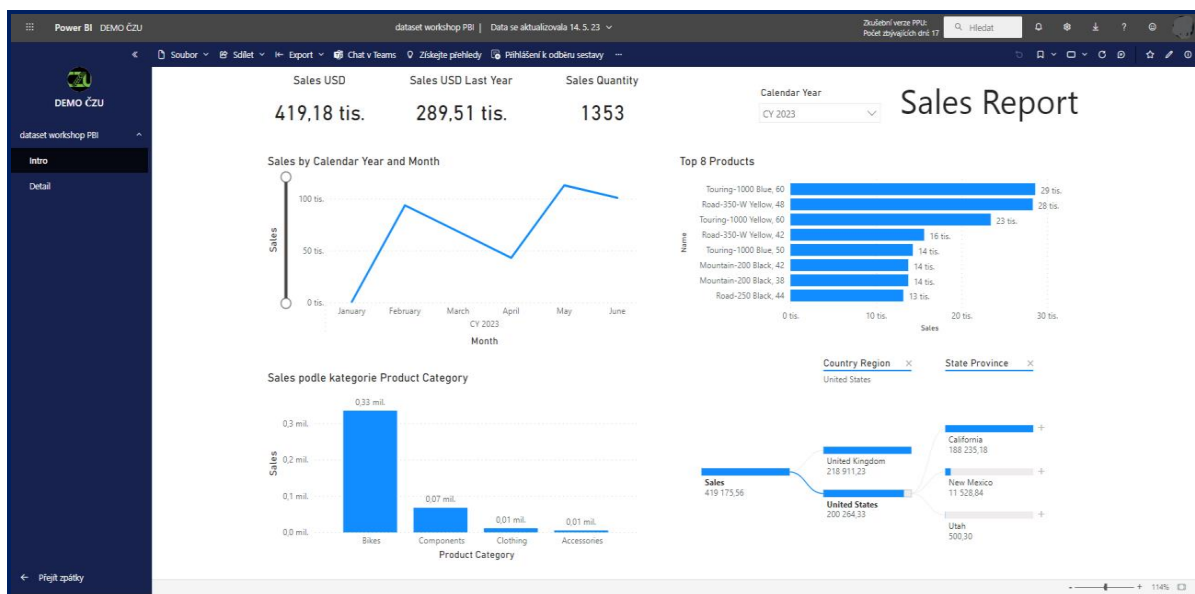
- **Synergy** – měl by být synergicky navržen tak, aby informoval byznys uživatele o rozdílech a událostech na jedné obrazovce.
- **Monitoring** – musí umět monitorovat KPIs a být schopen zobrazit kritické KPI ukazatele pro efektivní rozhodování v daný časový okamžik.
- **Accuracy** – prezentované informace na dashboardu musí být přesné, aby jim uživatel věřil. Data, vstupující do dashboardu, musí tedy být testována a správně validována.
- **Responsive** – citlivý, tj. musí umět rozpoznat prahové hodnoty tak, aby byznys uživateli oznámil prostřednictvím semaforů, KPI měřítek či e-mailových oznámení změnu, která má vliv na rozhodnutí.
- **Timely** – včasný, musí umět zobrazit nejaktuálnější informace vždy včas z hlediska efektivního rozhodnutí. Informace by měly být byznys uživateli podávány v reálném čase.

Musí dále naplňovat ještě **další požadavky**, a to:

- **Interaktivní** – nabízí využití funkce Drill-Down pro zjištění větších detailů.
- **Větší historie dat** – musí umožnit uživatelům zjistit historické a aktuální trendy daného KPI.

- **Customizace** – umožňuje změnit si dashboard podle země, určitých privilegií, pracovní pozice aj.
- **Analytický** – uzpůsobuje uživatelům analýzu dat v dashboardu pomocí dalších funkcí.
- **Společný pro více uživatelů** – umožňuje uživatelům sdílení dashboardu napříč organizací.
- **Sledovatelnost** – měla by umožnit každému uživateli zvolit si, jaké metriky chce sledovat v závislosti na pracovní pozici, regionu, zemi, produkci a na dalších atributech.

Příklad dashboardu v prostředí Power BI dokumentuje následující obrázek 14.1:



Obrázek 14-1. Příklad dashboardu v Power BI (Zdroj: Seyfor)

## 14.2 Efekty a přínosy dashboardu

- Poskytuje aktuální informace, které **umožňují uživatelům zlepšit rozhodování, optimalizaci procesů a efektivitu práce**. Umožňují tedy všem uživatelům (od manažerů až po běžné zaměstnance) **monitorovat klíčové aktivity** a procesy, potřebné k dosažení cílů firmy.
- Je založen na **definování podnikové strategie** pomocí **cílů, opatření a iniciativ**, které jsou určeny jak skupinám ve firmě, tak i některým jednotlivcům. Uživatelé mají přehled, co přesně musí udělat ve svých oblastech k dosažení firemních cílů.
- Existence dashboardů má pozitivní vliv **na ladění a provedení firemní strategie** s mnohem menšími výkyvy než bez nich.
- Zvýšení **predikce pomocí KPI ukazatelů**, pomocí nichž manažeři mohou lépe predikovat prognózu na základě minulé činnosti. Tento proces pomáhá společnosti vyhnout se nečekaným výpadkům.



- Zlepšuje **koordinaci a komunikaci** mezi manažery jednotlivých úseků či oddělení, aby více spolupracovali mezi sebou.
- Zvyšuje **motivaci zaměstnanců** pomocí vyplácení odměn za vyšší efektivitu práce, podloženou jednoznačně lepšími KPI. Poskytuje **sledování výkonnosti**, hodnocení a motivaci zaměstnanců.
- Nabízí **jednoznačný pohled na podnikání**, jelikož se udržuje pouze jedna společná verze, tudíž nedochází k nedorozumění mezi IT manažery a analytiky v organizaci, jaká verze dat je správná.
- **Snižuje náklady a redundance** pomocí konsolidace a standardizace informací.
- Umožňuje pracovníkům **měnit si dashboardy podle svých požadavků** bez podpory IT oddělení a analyzovat informace z různých úhlů pohledů.
- Obsahuje vysoce názorné, **grafické zobrazení** sledovaných metrik.
- Poskytuje okamžité posouzení sledovaných metrik z pohledu **plán / skutečnost**.

### 14.3 Problémy a omezení dashboardů

- Řešení dashboardů vyžaduje **specifické analytické přístupy** s důrazem na adekvátní výběr relevantních ukazatelů (KPI) pro jednotlivé role v řízení firmy.
- Často se objevuje **přeceňování analýz** a výstupů z dashboardů, zejména jejich grafické stránky.
- Pro grafické zobrazení metrik je potřebný relevantní **soubor dat za určitý časový úsek**.
- Problémy **související s kvalitou** a uložením dat znamenají, že při špatné kvalitě dat prezentují dashboardy špatné výsledky a poskytují nekvalitní podklady pro rozhodování a řízení firmy.
- **Nevhodně zvolené nástroje** pro prezentaci informací nepodporují kvalitní funkce pro analýzu a práci s daty.
- **Aktualizace dashboardu** neprobíhá zcela automaticky z důvodu velkého množství informací, byznys uživatelé stráví velké množství času, většinou několik hodin v týdnu aktualizací stavu namísto jeho včasného vyhodnocení, což není udržitelné v dlouhodobém horizontu.
- Objevuje se **velká izolovanost a granularita dashboardu** oproti zbytku řešení. KPI ukazatele poskytují příliš úzký pohled na byznys vůči celé firmě.

### 14.4 Závěry k realizaci dashboardů



- Dashboardy jsou obvykle **grafickým vyjádřením ukazatelů podle klíčových dimenzí**, a to formou grafů, tabulek, budíků a podobně.
- Řešení dashboardů má **specifické nároky na jejich analýzu** a řešení, neboť musí přesně vystihnout hlavní souhrnné informace o dané oblasti řízení podle potřeb příslušného manažera, kterému je určen.

- Základní **členění dashboardů** je na typy:
  - strategické,
  - taktické,
  - provozní.
- Každý z uvedených typů dashboardů má **odlišné určení, funkcionality i způsob využití** v praxi.
- Dashboard musí současně nabízet **cesty k detailním informacím**, z nichž jsou vytvořeny souhrnné.
- Charakter dashboardů se rozlišuje **podle úrovní řízení** a manažerů, kteří na nich pracují.

## E) Prediktivní analytika

Otázky prediktivní analytiky představují **velmi široký rozsah informací, přístupů, analytických úvah a nástrojů**, které mají jak obecnou, společnou povahu, tak zcela konkrétní vazbu na odvětví, prostředí a potřeby jednotlivých typů firem. Oddíl „Prediktivní analytika“ vychází dále **z celé řady zdrojů**, které jsou souhrnně uvedeny v závěru publikace. Za **hlavní** pro účely tohoto oddílu považujeme tyto:

- ABBOTT, D.: ***Applied Predictive Analytics. Principles and Techniques for the Professional Data Analyst***. John Wiley & Sons, Indianapolis, 2014. ISBN: 978-1-118-72796-6. K této publikaci s ohledem na její význam doplníme **několik podstatných poznámek autora**:
  - *Kniha popisuje prediktivní analytiku spíše pohledem praktika než teoretika.*
  - *Dobry praktik nemusí rozumět matematickým principům algoritmů, aby byl schopen je úspěšně aplikovat.*
  - *Na druhé straně dobrý praktik má rozumět dopadům změn parametrů pro modely, vlivům předpokladů na algoritmy predikcí a omezení algoritmů – zejména v prostředí velmi náročných projektů.*
  - *V praxi je často mnoho způsobů, jak řešit úspěšně problémy a výběr vždy závisí na mnoha faktorech, které je ovlivňují. Je proto třeba je jasně identifikovat a vyhodnocovat.*
  - *Úspěšné řešení prediktivních modelů závisí více na dobrém pochopení dat (a jejich obsahu), než na využití sofistikovaných algoritmů.*
- WILSON, J., E.: ***Predictive Analytics for Business Forecasting and Planning***. Graceway Publishing Company, 2021. ISBN 978-0-9839413-8-5. V rámci publikace jsou z pohledu významu a pojetí prediktivní analytiky **vyjádření několika špičkových expertů**, např.:

*„Predictive analytics is the most important business topic of our time extending an organization’s ability to optimize efficiency while improving speed to market...”*

Joe Eschenbrenner, ACPF  
Director, Demand & Supply Planning  
**PUMA Group**

*“Most works on predictive analytics discuss esoteric algorithms and techniques, all too often focusing on buzzwords instead of how practitioners can benefit from their use...”*

Dr. Larry Lapide  
Research Affiliate  
**MIT Center for Transport & Logistics**

*“Cause-and-effect techniques have transformed business forecasting and analytics...”*

Dr. Barry Keating

Professor of Business Economics & predictive Analytics

**University of Notre Dame**

*“Integrating predictive data analytics and scalable new techniques is the future of demand planning...”*

Tim Hotze

SVP, Global SCM & Logistics,

Network Planning & Global Intelligence

**Target**

## 15. Potřeba a obsah prediktivní analytiky



**Obsah řízení a prediktivní analytiky** je **velmi variantní** podle odvětví, typu a zaměření firem a jeho zvládnutí je pro analytiky hodně náročné. Z pohledu řízení firmy je ale samotným základem pro prediktivní analytiku plánování, resp. plánování poptávky („*demand planning*“) a prognózování („*forecasting*“), a to s uplatněním pokročilých principů, např. umělé inteligence a „*machine learning*“.

**Účelem** kapitoly je:

- definovat **potřeby byznysu** vzhledem k prediktivní analytice, tedy potřeby a požadavky plánování a prognózování ve firmě,
- **vymezit** prediktivní analytiku, **kategorizovat součásti** jejího obsahu, a to ve vztahu k potřebám prognózování a plánování,
- specifikovat **klíčové principy**, na nichž je prediktivní analytika založena,
- zhodnotit podstatné **efekty a omezení** prediktivní analytiky,
- určit **oblasti využití** prediktivní analytiky.

### 15.1 Potřeba prediktivní analytiky: prognózování a plánování

**Prognózování a plánování** v řízení firmy jsou aktivity, které **spolu souvisejí**, i když reálně jsou v některých případech realizovány nezávisle na sobě. Platí ale, že čím kvalitnější a přesnější jsou prognózy, tím kvalitnější jsou plány a plánování.

**Rozvoj** plánování a prognózování, jejich funkcionality ve vazbě na prediktivní analytiku, je možné považovat za jednu z **klíčových oblastí rozvoje byznysu** a investic ve firmách bez ohledu na to, v kterém odvětví působí, jaká je jejich velikost, složitost apod. Současné firmy disponují obrovskými objemy dat a smyslem konceptu, metod a nástrojů prediktivní analytiky je z nich vytěžit právě takové informace, které povedou k **posilování kvality plánování a prognózování** a tím i jejich konkurenceschopnosti a získávání nových konkurenčních výhod.

#### 15.1.1 Prognózování v byznysu

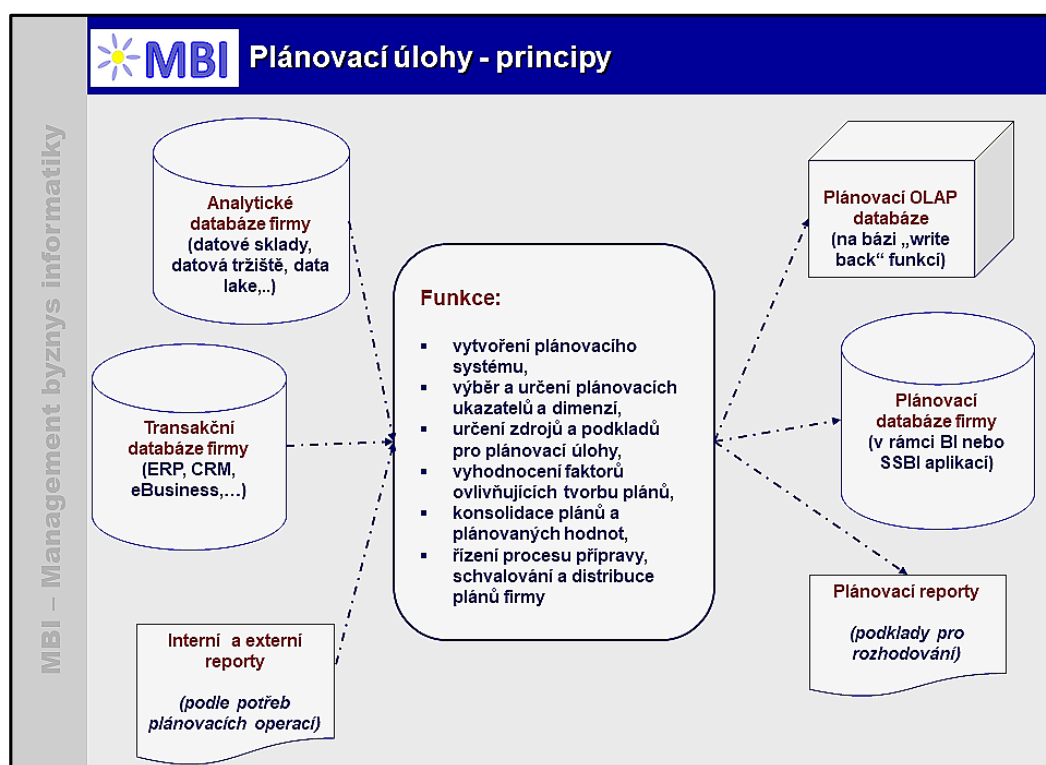
**Prognózování v byznysu** („*business forecasting*“) lze chápat jako proces, který využívá data, analytiku a vyhodnocování zkušeností pro vytváření predikcí podle potřeb byznysu. Prognózování **lze rozdělit**, obdobně jako jiné řídicí aktivity, na (Wilson, 2021):

- **strategické**, na dlouhé období, ale s rizikem větších chyb,
- **taktické**, se střední dobou odhadů, obvykle s vysokou přesností,
- **operativní**, real time, umožňující bezprostředně reagovat na změny tržního prostředí, změny poptávky, změny v disponibilních zdrojích apod.

Prognózování je většinou založené na **sledování trendů, sezónních vlivů, nejrůznějších faktorů**, na jejichž základě se formulují odhady, prognózy budoucího vývoje, např. obchodu, ekonomiky, technologií ve firmě.

### 15.1.2 Podstata a principy plánování

**Plánování v byznysu** se přímo váže na prognózování a v některých případech staví na výstupech prognózování, resp. jednotlivých prognóz. Je to de facto projekce budoucnosti, která kombinuje znalost minulosti s vyhodnocením budoucích potřeb produktů a služeb. Podstatu a principy plánovacích úloh dokumentuje Obrázek 15-1.



**Obrázek 15-1: Principy plánování a plánovacích úloh**

Hlavním smyslem plánování je **snížení nejistoty budoucího vývoje**, tj. plány a rozpočty poskytují kontrolní mechanismus, zda jsou naplánované cíle plněny s určitými povolenými odchylkami (Žůrková 2007, s. 9). Klíčovým aspektem systému plánů a rozpočtů je **časový předstih**, s nímž jsou možná rizika a úzká místa plnění cílů a řízení firmy identifikována.

**Úrovně plánování** zahrnují (v návaznosti na úrovně prognóz) tři základní úrovně plánů:

- strategické, cca na 10 let,
- taktické (manažerské), 1 – 3 roky,
- operativní, denní, týdenní, měsíční či kvartální.

**Klouzavé plánování** zahrnuje na ně navázané klouzavé rozpočty. Funguje na principu neustálé aktualizace plánů a rozpočtů, a to na základě skutečně sledovaného vývoje.

Umožňuje průběžně porovnávat plány a rozpočty se skutečností a vyhodnocovat jejich dosavadní přesnost a příslušně je upravovat pro další období.

### 15.1.3 Podstata plánovacích úloh

Pro řešení plánovacích úloh (plánů tržeb, nákladů, investic atd.) je **třeba nejen data zpřístupňovat podle nejrůznějších definovaných dimenzí** a analyzovat je podle nich, ale i nová data, tedy **plány tvořit, rovněž s respektováním těchto dimenzí** a jejich hierarchických struktur. Plánovací úlohy sledují **několik základních cílů**, zejména:

- vytvořit a **využít plánovací systém**, respektující ve firmě uplatňované plánovací a rozvrhové metody,
- **zajistit konsolidace vytvářených plánů**, vznikajících na různých organizačních jednotkách, tj. závodech, divizích, odděleních, nebo naopak rozpouštění centrálně stanovených plánů na tyto jednotky,
- **zajistit konsolidace hodnot z různých druhů plánů**, např. plánu investičního, výrobního nákupního, prodejního, personálního apod. do výsledného, obvykle finančního plánu,
- **zajistit konsolidace plánů z pohledu různých měn** a přepočítat na výslednou požadovanou měnu,
- automatizovat **řízení pracovního toku (workflow)** při přípravě plánu, resp. plánů, na kterém se podílejí různí manažeři, plánovači a další pracovníci podniku,
- efektivně **zpřístupňovat sestavené plány** zainteresovaným pracovníkům podniku,
- zajistit **potřebnou bezpečnost a nastavení přístupových práv** pro zpracování plánů i pro jejich prezentaci ve firmě, případně mimo ni, kde jde o možnosti jejich čtení, zápisu a schvalování.

### 15.1.4 Principy plánovacích úloh

Data do analytických databází se obvykle **nahrávají z primárních zdrojů**. Jinou možností je však tzv. **zpětný zápis**, což je funkce, kdy data do analytické databáze může zadávat přímo uživatel, nebo určitá, většinou plánovací aplikace. To znamená, že uživatel, resp. uživatelská aplikace zde může měnit hodnoty ukazatelů a dimenzí. Takové možnosti vytváření plánů a realizace změn, prováděných uživatelem zpětným zápisem, jsou základem řešení efektivních a výkonných plánovacích aplikací, které přinášejí **několik podstatných efektů**:

- **využití již existujících hodnot** dříve vytvořených plánů nebo hodnot o skutečnosti z minulých období, které se tak stávají základem pro tvorbu nových plánů,
- **využití časových řad**, vyhodnocování podstatných faktorů a proměnných a na jejich základě vytvoření jejich extrapolací,
- **pružné zpracování změn plánů**, kdy je třeba promítnout změny v celé plánovací struktuře, např. pokud se změní plán v jednom oddělení, je nutné měnit i konsolidované plány na vyšších úrovních podnikového řízení, a naopak kdy je nutné promítnat změny z vyšších organizačních jednotek na nižší,

- **promítání změn zpět do zdrojových databází** transakčních aplikací lze efektivně využívat k distribuci aktuálních plánů do aplikací nebo databází podřízených útvarů, divizí, jednotlivých poboček apod.,
- při využití efektivních **rozvrhových algoritmů, aplikovaných na struktury jednotlivých dimenzí** a jejich kombinací (např. procentuálních nebo obdobných rozvrhových schémat), lze podnikové plány sestavovat podstatně efektivněji a zajistit jejich konzistenci,
- je možné definovat a využívat tzv. **podniková pravidla (business rules)** nebo omezení (*business constraints*), tj. pravidla pro kontroly plánů, např. objem investic nepřesáhne 20 % plánovaných nákladů celého podniku apod.,
- podnikové plány jsou **centrálně uloženy v analytické databázi**, což znamená jejich rychlé a konzistentní vytváření s rozpadem či agregací plánovacích ukazatelů podle definovaných dimenzí a možnost rychlého porovnání skutečnosti s plánem, pokud je použita stejná struktura ukazatelů a dimenzí.

### 15.1.5 Plánovací úlohy v kontextu řízení firmy

Plánovací úlohy jsou rovněž **součástí naprosté většiny oblastí řízení** firmy i firmy jako celku. Na základě definovaných principů lze určit jejich **uplatnění v různých oblastech řízení firmy**, což je pak konkretizováno podle oblastí řízení. **Pracovně** lze přijmout následující **klasifikaci**:

- **Sestavení plánu** za vybranou oblast **podle vybraných ukazatelů** s využitím relevantních dimenzí, např. plán tržeb v daném období, plán objemu nákupů vybraných materiálů, plán počtu servisních zásahů na dodaném zboží apod.
- **Zpracování výhledu podle vybraných ukazatelů a s využitím relevantních dimenzí**, např. na 12 měsíců klouzavě v průběhu roku.
- **Sestavení rozpočtu**, tedy finančních hodnot ukazatelů, které silně závisí na prostředí konkrétní organizace, charakteristice trhu, druhu nabízeného produktu a dalších.

V každém případě jsou dílčí charakteristiky prognózování a plánování a plánovacích úloh podstatným předpokladem **pochopení potřeb prediktivní analytiky**, jejího řešení a uplatnění v manažerské praxi.

## 15.2 Podstatné principy prediktivní analytiky

**Výchozím principem a současně předpokladem** řešení a užití prediktivní analytiky je to, že jejím hlavním účelem je poskytovat **kvalitní a sofistikovanou podporu prognózování a plánování** v řízení firmy (viz výše) a tím přispívat k posilování její pozice na trhu a celkové ekonomické a obchodní úspěšnosti.

Prediktivní analytika je **termín relativně** nový a staví na jiných analyticky orientovaných disciplínách jako data mining, machine learning, statistika nebo rozpoznávání vzorů. Prediktivní analytika (Praus, 2013) je **typ analýzy, využívající data a prediktivní modely pro předpověď jevů na mikroekonomické úrovni**. Prediktivní analytika využívá



technologie, které se učí ze zkušeností (dat), aby předvídala budoucí chování jedinců k lepšímu rozhodování (Siegel, 2013).

**Hlavním předpokladem**, se kterým pracuje, je to, že **v lidském chování lze nalézt podobnosti**, vztahy a vzorce. Z dostupných dat je prediktivní analytika schopná, **pomocí technik dolování dat**, identifikovat důležité podobnosti a vztahy a využít je k předpovědi nejrůznějších jevů, událostí a aktivit.

**Hlavní principy**, na nichž je založeno řešení a využití prediktivní analytiky v praxi, je možné shrnout **do následujících bodů** (Abbott, 2014, Siegel, 2013, Wilson, 2021):

- Řešení prediktivní analytiky často závisí na potřebě **analyzovat hodnoty, které nejsou přesně známy**, ale jsou mimořádně užitečné. V některých případech jsou takové hodnoty známé historicky, ale není známo, kdy má docházet k rozhodnutí, resp. predikovanému rozhodnutí.
- **Využití prediktivní analytiky** se váže k různým sférám zájmu, např. k odhadům prodeje zboží, vývoje cen, podílu na trhu, měnových kursů, marketingových aktivit, promo akcí, retence zákazníků, optimálního portfolia vyráběných produktů, ale i makroekonomických nebo demografických trendů, počasí a dalších.
- Jde o proces **využití výpočetních metod** k nalézání a reportování významných vzorů, resp. schémat v datech. Prediktivní analytika vyhodnocuje historická data a je **realizována ve vztahu k disciplínám a metodám** jako je např. Business Intelligence, Competitive Intelligence, data science, data mining, text mining, umělá inteligence, machine learning, big data analytics, statistika, případně další. Základní charakteristiky těchto disciplín a jejich vztah k prediktivní analytice je obsahem **kapitoly 16**.
- Zvyšování pozornosti prediktivní analytice je založeno na tom, že **firmy disponují stále většími objemy dat** a přirozená cesta je využívat je ke zlepšování odhadů, prognóz, rozhodovacích činností a zvyšování celkové efektivity, a to v podstatně kratším čase, než tomu bylo dosud.
- Prediktivní analytika **zahrnuje nejistotu**: prognózy nemohou být nikdy přesné, výsledky na 100 % neexistují. Do odpovědí musí být vždy zahrnuty pravděpodobnost a ohodnocení chyb.
- Na druhé straně, zatímco tradiční metody prognózování a plánování vyžadovaly relativně přesně otázky a požadavky, metody prediktivní analytiky **hledají odpovědi i na špatně formulované nebo i žádné otázky**.
- **Řešení** prediktivní analytiky **zahrnuje** jak „**vědu**“, resp. vědecké disciplíny, tak „**umění**“, reprezentované zejména praktickými zkušenostmi řešitelů projektů. Přístup jeden bez druhého, ale nevede obvykle k úspěchu.
- Na prediktivní analytice se podílejí zejména **tyto role**: doménový expert, data a databázový expert a expert prediktivního modelování.
- Podmínkou efektivního řešení a využití prediktivní analytiky je schopnost **správně formulovat analytické otázky** ve vztahu k jejímu řešení.
- Prediktivní analytika je **méně přesná, pokud jde do většího detailu**: vysoká granularita v attributech, vlastnostech produktů apod. vždy povede k větším ne-

přesností, než když bude úroveň detailu nízká. I v dimenzi času, pokud se prognózy pohybují od roků, měsíců, např. ke dnům, klesá jejich přesnost.

- Prediktivní analytika je tím **přesnější, čím je blíže k události, k níž se vztahuje**: pokud se bude prognóza realizovat s relativně větším časovým zpožděním ve vztahu k nastalé situaci, pak se přesnost snižuje. Cílem např. je dostat se co nejbliže k aktuální zákaznické poptávce.
- Prediktivní analytika **se zlepšuje s objemem informací a znalostí** o daném jevu: cílem je získat co největší a nejkompaktnější obrázek o situaci, která má být předmětem analytiky.
- Nejvýznamnější součástí prognózování a plánování, a tedy i prediktivní analytiky jsou **předpoklady**, jsou důležitější než vlastní čísla. Předpoklady jsou velmi četné a významně přispívají k pochopení komplexity a nejistoty reality.
- Dalším podstatným principem jsou **scénáře v prediktivní analytice**. Představují potenciální okolnosti a kombinace předpokladů, které mohou mít významný dopad na firmu. Při větším počtu definovaných předpokladů je velmi pravděpodobné, že prediktivní analytika povede i k většímu počtu výsledků. Je pak nezbytné vyhodnotit faktory, které budou ovlivňovat, které z výsledků budou mít v dané situaci vyšší **pravděpodobnost** uplatnění a které méně.

### 15.3 Řešení prediktivní analytiky

Celkový **detailnější postup** řešení úloh prediktivní analytiky podle metodiky CRISP-DM je **náplní celé kapitoly 18**. Na tomto místě shrneme pouze základní aspekty řešení s odkazy na další části textu.

**Vstupem prediktivní analytiky** jsou data se **znalostí cílové proměnné a její hodnoty (targetu)**, která má být v budoucnu predikována. Data se v rámci prediktivní analytiky uspořádají, pročistí a vytvoří se nový datový zdroj. Na něj jsou poté **aplikovány funkce a metody**. **Výstupem** je nejúspěšnější prediktivní model, schopný s určitou pravděpodobností hodnotu cílové proměnné předpovídat.

Řešení, výsledky i využití úloh prediktivní analytiky jsou významně ovlivňovány celou **soustavou faktorů**, které je pro kvalitu řešení nutné dobře pochopit a vyhodnotit.

### 15.4 Efekty uplatnění prediktivní analytiky

Efekty uplatnění prediktivní analytiky se vztahují ke zkvalitnění řízení a rozhodovacích aktivit firmy, a to primárně v následujících bodech:

- Schopnost úspěšné predikce je **využita ke zlepšení rozhodnutí**, které je tak více postavené na faktech (vztazích, trendech) nalezených v datech než na intuici. Postupně se realizuje **promítání aktuálních událostí do prediktivních modelů** v reálném čase.
- Předpovědi, poskytnuté prediktivní analytikou, se týkají převážně **mikroekonomických efektů**, vyhodnocuje se chování jednoho člověka a ne masy lidí. Například se předvídá, kdo si s jakou pravděpodobností koupí nějaký produkt (Siegel, 2013).

- **Prediktivní analytika a plánovací scénáře** umožňují firmě řešit alternativní situace rychleji a efektivněji. Prediktivní analytika, která je zaměřena na **identifikaci parametrů a faktorů** ovlivňujících prognózy, přispívá k určování aktivit, které mají být realizovány pro dosažení požadovaných budoucích výsledků.

## 15.5 Omezení, problémy, předpoklady řešení prediktivní analytiky

Problémy a předpoklady řešení prediktivní analytiky se promítají do jednotlivých fází řešení (Abbott, 2014), které jsou pak konkretizovány v **kapitole 18**.

- **Problémy v řízení:**
  - Řešení a využití prediktivní analytiky silně závisí na vůli a pochopení managementu, do jaké míry je ochotné poskytnout pro řešení zdroje.
  - Obdobně určité zdroje vyžaduje i nasazení prediktivní analytiky do provozu, a proto mnohé modely a aplikace nejsou vůbec využity
- **Problémy v datech:**
  - Obvyklým problémem je zajištění potřebné kvality dat, redundance, duplicity, chyby, absence unifikovaných dat.
  - Specifickým problémem je potřeba transformovat data ze zdrojových systémů (např. ERP) do pouze dvourozměrných tabulek pro prediktivní analytiku. To může být při složitých datových strukturách problém.
  - Problém je často spojen i s nedostatkem dat pro vyvinutí úspěšných prediktivních modelů, i s potřebnou aktualizací dat v konzistenci se zdrojovými systémy.
- **Problémy s modely:**
  - Hlavní a častý problém modelů je přeučení („*overfitting*“), modely jsou příliš komplikované.
  - To znamená, že model špatně funguje na nových datech a interpretace výsledků je nespolehlivá. Často se přeučení v průběhu řešení ani neprojeví, ale až při předání do provozu se ukáže jeho slabší výkon.
  - Dalším problémem je špatný odhad náročnosti řešení modelu vzhledem k dostupnosti dat i komplikovanosti funkcí a tím i problém s dodržením časových termínů pro předání do používání. Je lepší proto začít s řešením relativně jednodušším a následně přidávat další funkce.
- **Problémy s distribucí modelů:**
  - Problémy mohou být spojeny s integrací modelů se zdrojovými systémy a jejich dostupností, případně s jejich upgrady.
  - Realizace prediktivní analytiky je časově náročná s nejistým výsledkem.

## 15.6 Využití prediktivní analytiky

V současné době se využití prediktivní analytiky **posunulo od velkých společností i k menším** nebo středně velkým firmám. Současně se analytici často posouvají od Business Intelligence k prediktivní analytice, jak stoupá tlak na co nejvyšší využití disponibilních dat. Existuje **velké množství aplikací** prediktivní analytiky napříč mnoha obory lidské činnosti. Od bankovníctví, pojišťovnictví, marketingu po medicínu, bezpečnost a další, jako např. (SIEGEL, 2013), nebo:

- americká města Chicago, Los Angeles, Memphis (TN), Richmond (VA), Santa Cruz (CA) a Vineland (NJ) směřují policejní hlídky do oblastí, kde je predikováno nejvyšší riziko kriminality,
- FedEx predikuje s 65% až 90% úspěšností zákazníky, kteří přejdou ke konkurenci,
- odhadem 40 % obchodů na londýnské burze je prováděno algoritmy.

Prediktivní analytiku **využívají především podniky**, jako např.:

- **výrobní a obchodní firmy:**
  - odhady objemu financí a finančních zdrojů,
  - predikce objemu tržeb podle zákazníků,
  - prediktivní údržba,
- **banky:**
  - hodnocení klienta a kredibility,
  - identifikace a hodnocení rizikovosti půjček,
  - vyhodnocování trendů,
- **pojišťovny:**
  - při odhalování podvodných pojistných událostí,
  - při upisování,
  - při cenění rizik a pojistného,
  - pro zlepšení efektivity marketingových kampaní,
- **telekomunikační společnosti:**
  - vyhodnocování trendů v poptávce na trhu,
  - lepší cílení marketingových kampaní,
- **burzy:**
  - predikce vývoje akcií a komodit,
  - obecně se najde využití v marketingu, prodeji, finančních službách, medicíně.

Dá se však předpokládat, že se úlohy prediktivní analytiky využívají nebo budou využívat **v různých oblastech řízení** napříč odvětvími ekonomiky.

## 15.7 Závěry k potřebě a obsahu prediktivní analytiky



- **Potřeba** prediktivní analytiky se váže převážně k **řešení prognóz** v rámci firemního řízení a následně k **přípravě plánů a rozpočtů**. Musí respektovat principy a specifické nároky na funkcionalitu a data těchto úloh.
- Řešení a využití prediktivní analytiky je **založeno na řadě principů** účelných pro pochopení jejích možností, zejména „řešení prediktivní analytiky zahrnuje jak „vědu“, resp. vědecké disciplíny, tak „umění“, reprezentované zejména praktickými zkušenostmi řešitelů projektů“. Tomu musí odpovídat i role expertů, podílejících se na těchto řešeních apod.
- Obdobně je při řešení prediktivní analytiky nezbytné kvalifikovaně **vyhodnotit efekty**, které má a může přinést, a **omezení**, s nimiž je třeba počítat. To je podstatným předpokladem pro úspěšnou realizaci projektů.
- Využití prediktivní analytiky pokrývá již nejen **velké společnosti, ale i střední a menší**, přes nejrůznější odvětví a typy firem a napříč oblastmi jejich řízení.
- Prediktivní analytika je součástí celého **komplexu disciplín**, tvořících **podnikovou analytiku**. Ty jsou náplní další kapitoly.

## 16. Prediktivní analytika v kontextu podnikové analytiky



Jak již bylo výše uvedeno, je prediktivní analytika podstatnou **součástí podnikové analytiky** a podnikového řízení. Podniková analytika představuje v současné době celou **škálu disciplín, metod, přístupů a nástrojů**, které se různě **doplňují**, ale i **překrývají** a zejména se velmi rychle významně **rozvíjejí**. To je důvodem i pro to, že najít nějaké sjednocující schéma s potřebnými vazbami je dnes velmi obtížné.

**Účelem** kapitoly je specifikovat **alespoň rámcově podstatu těchto disciplín, metod a nástrojů** tak, aby bylo zřejmé, kde a jak prediktivní analytika v komplexu těchto ostatních disciplín, metod a nástrojů v praxi řešit a v řízení využívat. V této kapitole jsou uvedeny pouze stručné shrnutí a vazby na prediktivní analytika.

Pro vymezení kontextu prediktivní analytiky **v rámci podnikové analytiky** byly do dalších podkapitol zařazeny tyto její součásti:

- **Business Intelligence (BI)**, tj. řešení, založená na celém spektru velmi výkonných komponent (zejména databázových systémů, transformačních nástrojů, analytických a dalších nástrojů).
- **Competitive Intelligence (CI)** se zabývá sběrem, zpracováním a ochranou informací s cílem získat konkurenční výhodu pro firmu.
- **Data Science** zahrnuje porozumění business logiky dat, přípravu dat, modelování / optimalizaci / simulaci, vyhodnocení a nasazení analytického modelu.
- **Data Mining**, dolování dat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází.
- **Text Mining**, resp. textová analytika představuje analýzu textových zdrojů a získávání z nich nových informací, kde zdroje mohou být velmi různorodé od knižních nebo novinových publikací přes blogy a další.
- **Machine Learning** obsahuje vymezení podstaty a kroků řešení úloh ML, vymezení možností produkcionalizace a governance řešení.
- **Big Data Analytics** pokrývá velmi složité analytické operace na rozsáhlých zdrojích dat, např. ve výrobních podnicích s využitím technologií IoT.

Další podkapitoly se věnují stručné charakteristice uvedených disciplín, metod a nástrojů s návaznostmi na prediktivní analytika.

### 16.1 Business Intelligence, BI

Business Intelligence zejména sleduje podstatné charakteristiky dat, KPI, vytváří reporty a KPI reporty, určované uživateli („*user driven*“) (Abbott, 2014).

### 16.1.1 Principy Business Intelligence

Pokud odhlédneme od realizačních charakteristik a jednotlivých produktů, pak **základní principy** řešení Business Intelligence jsou obsahem kapitoly 5.

### 16.1.2 Prediktivní analytika a Business Intelligence

Zatímco aplikace Business Intelligence se orientují na funkce a proměnné, které jsou určované uživateli „*user driven*“, **prediktivní analytika** specifikuje funkce a hledané vzory, určované daty „*data driven*“. Navíc **prediktivní modely** zajišťují více než jen identifikace atributů, které jsou spojené s cílovou proměnnou, **identifikují** také **nejlepší cestu, jak predikovat cíl**. Neidentifikují jen, které proměnné jsou predikované, ale také jak dobře predikují cíl. Rozdíl mezi oběma disciplínami dobře dokumentují i rozdíly **v otázkách zákaznické analytiky**, které obě disciplíny řeší (Abbott, 2014):

- **Business Intelligence:**
  - Jaký je poměr návratnosti emailů, odpovědí apod.?
  - Jaké regiony mají jakou návratnost odpovědí v průzkumech?
  - Jaké jsou poměry návštěvnosti webu podle produktů?
  - Kolik bylo opakujících se nakupujících podle regionů, produktů apod.?
  - Kolik je účastníků loajality programů podle regionů, skupin zákazníků apod.?
- **Prediktivní analytika:**
  - Jaká je pravděpodobnost návratnosti emailů, odpovědí apod.?
  - Které produkty se budou pravděpodobně nejvíce prodávat?
  - Kolik emailů je pravděpodobně třeba poslat zákazníkům pro zvýšení objemu prodeje?
  - Jaká bude očekávaná návštěvnost webu v příštím týdnu?
  - Jaká je pravděpodobnost vyprodání produktu v příštím měsíci?

Zatímco funkcionality Business Intelligence konstatuje a analyzuje stav „co se stalo a proč se to stalo“, funkcionality **prediktivní analytiky** zahrnuje odhady budoucích stavů na základě pravděpodobnosti.

## 16.2 Competitive Intelligence, CI

Competitive Intelligence využívá **externí informační zdroje** za účelem **vyhodnocování tržního prostředí** a trendů, které zde probíhají, **konkurence**, konkurenčních subjektů a jejich významných charakteristik.

### 16.2.1 Principy Competitive Intelligence

Řešení a využití competitive intelligence ve firemní praxi je obvykle spojeno s principy, obsaženými v kapitole 8.

## 16.2.2 Prediktivní analytika a Competitive Intelligence

Je zřejmé, že Competitive Intelligence se, obdobně jako Business Intelligence, zaměřuje především na získávání a **vyhodnocování informací o aktuálním stavu**, v tomto případě převážně **externích informací** o externím tržním prostředí.

Na druhé straně díky aktuálním informacím o stavu trhu a aktivitách konkurence představuje podporu **prognózování a plánování** v řízení firmy. To znamená, že v tomto místě se obě disciplíny propojují a informace v rámci Competitive Intelligence mohou představovat i významný datový zdroj pro řešení prediktivních modelů.

## 16.3 Data Science

Data Science je přístup k řešení IT, který **z velkého množství komplexních dat**, které navíc i velmi rychle narůstá, **získává informace a znalosti** (F. Provost a T. Fawcett, 2013). Z pohledu tohoto textu představuje Data Science také **společnou nadstavbu** nad řadou dílčích a rozdílných disciplín, metod a nástrojů.

### 16.3.1 Principy Data Science

Data Science, datová věda představuje **široký koncept pro rozvoj a užití dalších dílčích disciplín**, spadajících do podnikové analytiky. V tomto smyslu je postavena na těchto principech:

- V praxi je možné definovat **sadu základních konceptů**, které jsou **podstatou pro formování základních disciplín** pro získávání informací a znalostí z dostupných dat. Tyto koncepty pokrývají celý proces od pochopení byznys problémů, přes využití technik Data Science až po poskytování výsledků manažerům a analytikům s cílem zkvalitňovat jejich rozhodování,
- V první řadě jsou to **koncepty, ukazující**:
  - jak Data Science začlenit do firmy a organizace v rámci konkurenčního prostředí,
  - jak efektivně sestavovat a připravovat týmy pro Data Science,
  - jak formovat způsob analytického myšlení, zaměřený na získávání konkurenčních výhod,
  - jak efektivně postupovat při řešení projektů, orientovaných na Data Science.
- **Koncepty, zaměřené na způsob analytického myšlení**, umožňují:
  - identifikovat vhodná data a metody pro řešení úloh Data Science,
  - aplikovat různé procesy data miningu a dalších přístupů a efektivně přispívat k řešení úloh.
- Uplatnění **konceptů pro získávání znalostí z dat** je další součástí řešení úloh Data Science.
- Zásadním a výchozím **předpokladem** pro efektivní využívání uvedených **konceptů** Data Science je jejich zasazení do kontextu řízení celé firmy, tj. do soustavy typů aplikací a odpovídajících nástrojů, které se vzájemně doplňují, ale i pře-



krývají, tj. respektování takových souvislostí a jejich efektivní využití vzhledem k potřebám dané firmy.

### 16.3.2 Prediktivní analytika a Data Science

Je zřejmé, že objasňování principů Data Science na základě uvedených konceptů výrazně **usnadní komunikaci** mezi specialisty v Data Science a současně i v prediktivní analytice, tj. mezi doménovými experty, datovými experty a experty prediktivního modelování. To znamená, že se vytvářejí předpoklady pro vzájemné pochopení a **podstatně přesnější formulaci problémů** firmy, řešitelných na bázi úloh Data Science a prediktivní analytiky.

Využití prediktivní analytiky na bázi konceptů Data Science **v rámci jednotlivých oblastí řízení** firmy znamená předpoklad jejího efektivního řešení v přímých vazbách na potřeby byznysu.

Příkladem takového konceptu je **určování podobnosti** dvou entit, k nimž se vztahují určitá data. Takový koncept **vytváří základnu pro řešení nejrůznějších úloh**, např. pro hledání podobnosti zákazníka s jinými zákazníky. To pak vytváří podklady pro prediktivní úlohy zaměřené na řešení cíleného marketingu, hodnocení zákazníka z hlediska obchodních možností (tržeb), pravděpodobnosti, že právě on využije prezentované nabídky firmou.

Tyto koncepty jsou také např. základnou pro **úlohy clusteringu**, sdružující objekty (např. zákazníky, zboží apod.) do skupin na bázi jejich sdílených vlastností. S tím souvisí i **využívání vzorů („pattern“)** pro analýzy vlastností zkoumaných objektů, a to v různém kontextu, což je rovněž součástí prediktivní analytiky.

## 16.4 Data mining, DMI

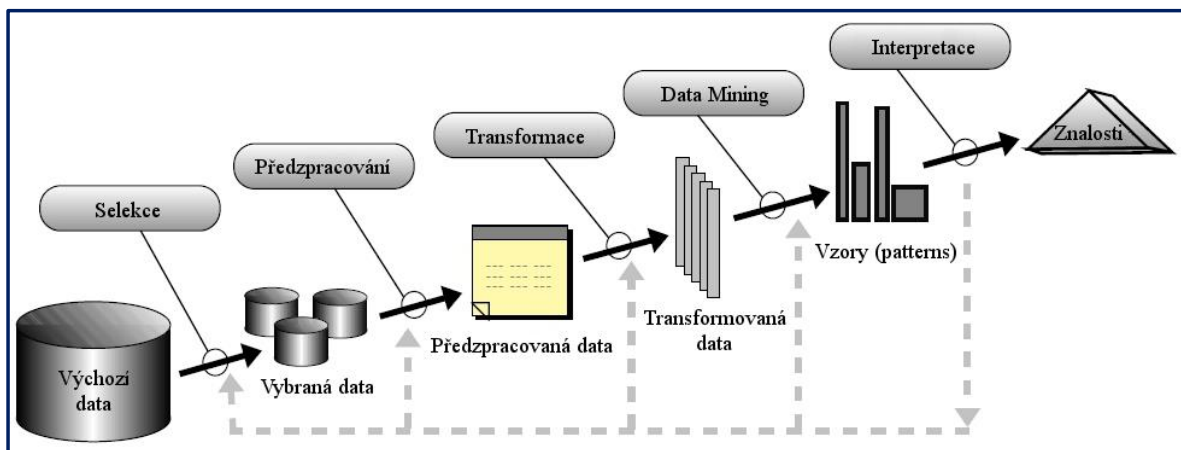
**Data mining, DMI** (dolování dat) je proces **extrakce relevantních, předem neznámých nebo nedefinovaných informací** z velmi rozsáhlých databází.

### 16.4.1 Principy data miningu

Data mining v analytice a pokročilé analytice představuje „klasickou“ disciplínu s dlouhou historií, jejíž principy a přístupy využívají i další analytické disciplíny:

- důležitou vlastností dolování dat je, že se jedná o **analýzy odvozované z obsahu dat**, nikoli předem specifikované uživatelem nebo implementátorem a že se jedná především o odvozování prediktivních informací,
- data mining je součástí **procesu „dobývání znalostí“**, který představuje **5 fází**, které jsou prováděny opakovaně. Každá fáze tak může být (a většinou je) prováděna opakovaně, s cílem nalezení nejlepšího možného výsledku dané fáze a jeho předání jako vstupu fáze následující,
- **fáze dobývání znalostí** jsou (Berka, 2003), viz Obrázek 16-1:
  - **selekce** – vytvoření či shromáždění cílových dat, tj. těch dat, která poslouží pro dobývání znalostí,

- **předzpracování** – „čištění“ špatných dat, zpracování chybějících hodnot, příprava odvozených atributů apod. Předzpracování dat je pokládáno za nejdůležitější část celého procesu dobývání znalostí,
- **transformace** – převedení dat do formátů vhodných pro různé algoritmy data miningu či různé softwarové nástroje. Tato fáze zároveň může obsahovat různé agregace a výběry vzorků dat,
- **data mining** – aplikace analytických metod. Výstupem této fáze jsou nalezené vzory a modely,
- **interpretace** – fáze s cílem vyhodnocení nalezených vzorů a modelů. Tato fáze vyžaduje znalost odborníka v řešené oblasti, neboť identifikuje výsledky obsahující znalosti nové, předem známé, samozřejmé, nezajímavé apod. Výstupem se pak stávají analytické zprávy, vizualizace či rozhodnutí provést vhodné akce.



**Obrázek 16-1: Přehled fází, tvořících proces Dobývání znalostí z databází (Fayyad, a další, 1996 str. 41)**

Data mining využívá celou **škálu metod** jako rozhodovací stromy, rozhodovací pravidla, asociační pravidla, neuronové sítě a další. **Aplikace** data miningu lze rozdělit podle mnoha kritérií a do různých skupin (Máša, 2008). Podle potřeb řízení a uplatnění ve firmách zahrnují:

- **segmentace** – rozdělení objektů do skupin, které mají podobné charakteristiky. Nejčastěji se jedná o segmentaci zákazníků, a to podle různých charakteristik,
- **predikce odchodu zákazníků**, kde účelem je předpovědět klienty, kteří hodlají odejít ke konkurenci, přestat využívat určitý produkt apod.,
- **cílený marketing**, tj. marketingové akce, specializující se pouze na určitý okruh klientů, takových, u kterých je oproti ostatním klientům vyšší pravděpodobnost, že zareagují na konkrétní kampaň,
- **credit scoring** je ohodnocením klienta podle úvěrového rizika. Tedy rizika, zda budou splácet např. svůj úvěr bance,

- **fraud detection**, aplikace, kdy jsou hledány podvody (ať již v rámci pojistných událostí, bankovních transakcí či v systémech telekomunikačních operátorů),
- **analýza nákupního košíku** zjišťuje souvislosti mezi produkty, které kupují zákazníci společně (např. párky společně s hořčicí). Díky tomu lze zákazníkovi vhodně nabízet kombinace produktů (cross selling), pracovat s rozmístěním produktů v rámci prodejní plochy prodejny či tvořit stránky s produktem v rámci eShopu.

### 16.4.2 Prediktivní analytika a data mining

**Provázání** prediktivní analytiky a data miningu je **velmi silné**. Využívá stejných nebo obdobných metod a obvykle i datových zdrojů.

Rozdíl (a současně shoda) je v tom, že **cíle data miningu** jsou predikce *i* deskripce, kde z pohledu data miningu je:

- **„predikce** funkcí, která využívá některých proměnných k předpovědím neznámých či budoucích hodnot jiných proměnných, které nás zajímají“,
- **deskripce** „se zaměřuje na hledání lidem srozumitelných vzorů, popisujících data“.

V každém případě se řešení úloh prediktivní analytiky silně propojuje s postupy a metodami řešení úloh data miningu.

## 16.5 Text mining

**Text Mining, resp. textová analytika** představuje analýzu textových zdrojů a získávání nových informací.

### 16.5.1 Principy text miningu

Text mining se stává stále významnější součástí podnikové analytiky a je založen na těchto principech:

- **zdroje** dat mohou být **velmi různorodé** – od knižních nebo novinových publikací, přes blogy, pracovní poznámky zaměstnanců, např. o kontaktech se zákazníky, o situaci na trhu apod.,
- nestrukturovaná data, která jsou předmětem text miningu, zahrnují:
  - standardní texty (Word, PDF apod.), prezentace (Powerpoint),
  - texty vyjádřené v HTML, XML apod.,
  - emaily,
  - obrázky, videa,
  - audio soubory,
- funkce text miningu vyžadují **pochopení** daného jazyka, struktur vět, gramatiky, slangu, souvislostí v textech.

**Využití text miningu** se váže např. k následujícím oblastem:

- význam text miningu se uplatňuje **s využitím sociálních sítí** zejména pro oblast **marketingu** a marketingové analýzy informací, které se např. na sociálních sítích k firmě váží,
- text mining nachází silné uplatnění i ve **vyhodnocování nejrůznějších průzkumů**, kdy dotazníky a otázky mají otevřený charakter a respondent na ně má odpovídat volným textem, obdobně při analýzách požadavků a dotazů na help desku,
- text mining se velmi silně **uplatňuje v aplikacích Competitive Intelligence**,
- do určité míry s text mining souvisí i oblast analýzy obsahu, **Content Analytics**, která rozšiřuje zdroje o fotografie, multimédia, hlas, případně další,
- významnou aplikační oblast text miningu představuje **zdravotnická analytika**, kde většina informací (např. diagnóz apod.) má textový charakter.

### 16.5.2 Prediktivní analytika a text mining

Text mining **není** zatím přímo **ve středu zájmu** firem ve spojení s prediktivní analytikou. Na druhé straně jeho **význam v tomto kontextu rychle stoupá** tak, jak se ukazuje bohatstvím informací, obsažených v textových zdrojích a využitelných v prediktivních úlohách. K tomu je třeba uvést:

- **řešení** úloh text miningu je relativně **hodně náročné**, vyžaduje si od analytiků prediktivní analytiky značné penzum dodatečných znalostí,
- úspěšnost text miningu v souvislosti s prediktivní analytikou dokumentuje i **příklad IBM počítače Watson**, který kombinuje techniky text miningu, prediktivní analytiky a umělé inteligence,
- **analýza textů**, která je často velmi komplexní, obvykle **přináší** do úloh prediktivní analytiky významná zlepšení a **vyšší přesnost**,
- zatímco prediktivní analytika staví převážně na čtyřech hlavních druzích analýzy, text mining využívá sedm a více druhů analýz, z nichž následujících pět je využíváno i v úlohách prediktivní analytiky, a to (Abbott, 2014):
  - „*information retrieval*“ – vyhodnocuje dokumenty podle přehledu klíčových slov,
  - „*document clustering*“ – sdružuje dokumenty na základě klíčových slov a frází z každého dokumentu,
  - „*document classification*“ – označuje dokumenty podle klíčových slov a frází z každého dokumentu,
  - „*sentiment analysis*“ – je speciální případ klasifikace dokumentů, predikuje pozitivní, negativní nebo neutrální pocity, dojmy z různých zákaznických poznámek, hodnocení nebo informací na sociálních sítích,
  - „*information extraction*“ – vybírá a sumarizuje klíčová slova a fráze, které odpovídají dokumentům.

## 16.6 Umělá inteligence (Artificial Intelligence)

### 16.6.1 Principy umělé inteligence

Umělá inteligence je schopnost stroje nebo systému zvýšit nebo rozšířit jakýkoli proces nebo výstup, které by normálně vyžadovaly lidské znalosti a inteligenci. Umělá inteligence se chápe **jako věda**, studující inteligentní agenty, což je jakákoliv entita, která vnímá prostředí a podniká kroky, které maximalizují svou šanci na úspěšné dosažení stanovených cílů. Často však bývá ztotožňována s termínem „*strojové učení*“, resp. „*Machine Learning*“.

### 16.6.2 Prediktivní analytika a umělá inteligence

Umělá inteligence přinesla do prediktivní analytiky obrovský posun, protože před jejím uplatněním bylo možné pracovat pouze s určitým omezeným počtem proměnných, zatímco s jejím vstupem do prediktivní analytiky se tyto počty enormně zvýšily.

Technologie, spojené s umělou inteligencí, představují významný pokrok v prognózování byznysu a obdobně v celém komplexu plánovacích úloh.

## 16.7 Machine Learning, ML

Machine Learning (Wilson, 2021) je **podmnožina umělé inteligence**. Je to soubor různých technik, metod, modelování a programování, které dovolují systému se automaticky učit.

### 16.7.1 Principy Machine learning

Machine learning je postaven na následujících principech:

- je to **algoritmus nebo technika**, která dovoluje systému být **trénován a učit se vzory** ze vstupů a následně je upravovat podle zkušeností, aniž by musely být explicitně programovány.
- **Rozlišují se:**
  - **řízené („supervised“) učící se algoritmy** – pokoušejí se modelovat vztahy a závislosti mezi cílovou predikcí, prognózou a vstupními vlastnostmi tak, že mohou predikovat výstupní hodnoty pro nová data založená na vztazích, které byly „naučeny“ na předchozích datových sadách. Spoléhají se zde na experty, kteří fungují jako učitelé k naplnění trénovacích dat, která pomáhají modelu získávat správné odpovědi,
  - **neřízené („unsupervised“) učící se algoritmy** – učí se z prostých, jasných příkladů bez připojených odpovědí, kde algoritmu je ponechána možnost určit, které datové vzory využije.
- **Využití** machine learning v jednotlivých **typech analytických úloh** je následující:
  - **deskriptivní analytika** – využívá obdobných metod jako data mining. Využívá metody clusteringu a asociační pravidla. Příkladem klastrů je vytváře-

ní zákaznických skupin podle nákupních zvyklostí. Příkladem asociačního pravidla může být „kdo nakupuje zboží A, obvykle nakupuje i zboží B“,

- **diagnostická analytika** – využívá machine learning pro identifikace různých anomálií např. u detekcí podvodů, nebo hledají vzory dat pro klasifikace událostí, vlastností produktů, nebo pro řešení úloh regrese,
  - **preskriptivní a kognitivní analytika** – zahrnuje i „deep learning“, což je podmnožina machine learning, založená na algoritmech, které se pokoušejí emulovat lidské myšlení a chování při řešení komplexních problémů. Deep learning nastavuje základní parametry dat a počítač má s využitím mnoha vrstev zpracování vytvářet potřebné vzory.
- **Řešení machine learning** typicky obsahuje **tyto kroky** (Žydyk, 2020):
- **získání dat** – pro výběr a naučení modelu budou potřeba určitá podkladová data,
  - **explorace a validace** – prvotní prozkoumání a tzv. profiling dat – zjišťuje se, jak data vypadají jako celek, co v nich je, co by mohlo chybět, jaká je struktura atp.,
  - **čištění (transformace, předzpracování) dat** je nejnáročnější operace procesu – data je třeba řádně připravit (vyřešit chybějící či odlehlé hodnoty, provést kódování textových proměnných, vytvořit nové prediktory atp.),
  - **rozdělení dat** – když jsou data připravena, rozdělí se na trénovací a testovací množinu pro učení a ověření kvality modelu,
  - **výběr a trénování modelu** – podle povahy úlohy je třeba vybrat druh modelu (regrese, klasifikace...) a provést trénování (učení),
  - **vyhodnocení výkonu modelů** – pro dosažení co nejlepšího výsledku se zpravidla testují různé druhy modelů s různými hyperparametry,
  - **uložení nejlepšího modelu** – nejlepší model je uložen a zaverzován pro další použití,
  - **převedení do produkce** – model je třeba zpřístupnit pro použití v konzumujících aplikacích,
  - **monitoring (správa) modelu v produkci** – model v produkčním prostředí je třeba dále monitorovat (chyby, rychlost, zátěž...).

### 16.7.2 Prediktivní analytika a machine learning

Pro prediktivní analytiku nabízí machine learning nové možnosti a pokročilé metody při jejím využití, dosahuje se lepších prognóz. Představuje dnes **součást všech typů analytických úloh**, od deskriptivní po kognitivní včetně prediktivní analytiky. ML je **využíváno v mnoha aplikacích**, jako např. Amazon, YouTube, Netflix v souvislosti s dalším doporučováním obsahu podle předchozích preferencí a nákupů zákazníka. Např. Amazon podle předchozích nákupů doporučuje další tituly knih nebo publikací, souvisejících s předchozími nakupovanými tématy. Business Intelligence aplikace využívají machine learning algoritmy k identifikování dalších použitelných dat.

**Prediktivní analytika** – využívá machine learning a deskriptivní proměnné pro prediktivní modelování jako neuronové sítě, extrapolace, regresní modely, případně pro spojování několika modelů do jednoho celku. Např. u neuronových sítí lze využít data z předchozích marketingových kampaní (jako věk, příjem atd.) a určovat příležitosti a rizika, spojená s novými kampaněmi.

## 16.8 Big Data Analytics

**Příklady použití Big Data Analytics** lze objevit téměř v každém sektoru podnikání, kde se pracuje s rozsáhlými objemy dat.

### 16.8.1 Principy Big Data Analytics

Jednou z hlavních předností Big Data Analytics je fakt, že umožňují hledat odpovědi na otázky, na které se organizace v minulosti ani nenapadlo ptát (Halama, 2021).

Výsledkem mohou být nové a lepší produkty, nebo optimalizace stávajících služeb. Analytik Jeff Kelly ve svém Big Data manifestu (Kelly, 2017) uvedl několik příkladů použití Big Data nástrojů a technologií v současnosti v praxi již zavedených a osvědčených:

- **Systém doporučení:** Řada elektronických obchodů využívá Big Data nástroje pro **doporučení produktu nebo služby** spotřebiteli na základě analýzy jeho chování, historie a profilu v reálném čase. Amazon tímto způsobem nabízí související produkty a sociální síť LinkedIn využívá Big Data nástroje pro doporučení možných přátel a známých.
- **Názorová analýza:** Big Data nástroje umožňují pokročilé **analýzy textu i nestrukturovaných dat např. ze sociálních médií** jako jsou Twitter a Facebook. Jednou z pokročilých funkcí je názorová analýza, která v daném textu hledá a analyzuje slova vyjadřující názor (pozitivní a negativní indikátory), týkající se zkoumaného produktu nebo služby.
- **Detekce podvodů:** V oblasti bankovníctví se Big Data nástroje používají k **detekci podvodů a krádeží**. Analýzou chování uživatele a historií jeho transakcí lze odhalit transakce, u kterých je vysoká pravděpodobnost podvodu, například kompromitace účtu nebo krádež kreditní karty.
- **Analýza sociálních vztahů:** Analýza sociálních vztahů a jejich **vizualizace pomocí sociálních grafů a sociogramů** pomáhá určit vztahy **uvnitř sociální skupiny** a vliv jednotlivců na ostatní členy skupiny. Tyto informace o spotřebitelích jsou pro podniky cenné z toho důvodu, že určují nejdůležitější zákazníky jako ty, kteří nejvíce ovlivňují ostatní ke koupi, a ne ty, kteří sami kupují nejvíce zboží. Pro dosažení úspěchu produktu je tedy nutné přesvědčit právě spotřebitele s velkým vlivem na ostatní.
- **Analýza zákaznické zkušenosti:** Big Data technologie se také používají k integraci a **analýze řady různých datových zdrojů o interakci se zákazníkem** jako jsou call centra, web a sociální média. Kombinace těchto zdrojů dává ucelený a komplexní pohled na zákaznickou zkušenost a umožňuje lepší porozumění a optimalizaci zákaznické komunikace.

- **Monitoring logů:** Big Data nástroje a technologie jsou vhodné pro skladování a **analýzu velkých objemů automaticky generovaných strojových dat** při provozu počítačové sítě jako jsou síťové a serverové logy a data, generovaná dalším hardwarem. Analýza a monitoring těchto dat pomáhají diagnostovat problémy v síti a eliminovat slabé stránky sítě.

### 16.8.2 Prediktivní analytika a Big Data Analytics

- Z předchozího přehledu je patrné, že prediktivní analytika může představovat velmi podstatnou součást Big Data Analytics tam, kde **se jedná o prognostické a zejména plánovací úlohy**, které je nezbytné realizovat na obrovských objemech dat.
- **Retenční analýza: Analýza zákaznické věrnosti**, respektive pravděpodobnosti, že **zákazník přejde ke konkurenci**, je další z oblastí, kde mnoho podniků využívá Big Data technologie. Pravidelným rozbořením zákaznického chování, využívání služeb, výdajů a dalších vzorců chování lze proaktivně řešit zákaznickou retenci.
- **Analýza marketingových kampaní:** V současnosti existuje řada nástrojů pro **monitoring a kvantifikaci efektivnosti marketingových kampaní**. Big Data kromě standardních funkcí umožňuje tyto analýzy provádět v reálném čase a na rozsáhlejších objemech dat, umožňujících vyšší granularitu dat a na základě toho plánovat další marketingové akce, např. v telekomunikacích, retailu apod.
- **Věda a výzkum:** Big Data technologie se často používají k analýze obrovských objemů dat **v oblasti zdravotnictví a farmacie**, zejména pro prediktivní modelování při vývoji léků a analýze dat z klinických testů.

## 16.9 Závěry k prediktivní analytice v kontextu podnikové analytiky



- Existuje celá **řada disciplín**, které se orientují na analytiku a analytické úlohy. Problém je v tom, že tyto disciplíny se nejen **doplňují**, ale i **vzájemně překrývají**, poskytují často obdobnou funkcionalitu, využívají stejné nebo obdobné metody. Specifikovat přesně jejich vzájemné vazby je proto hodně obtížné.
- Pro jejich využití v praxi je proto důležité **definovat jejich možnosti** vzhledem k definovaným a vyhodnoceným potřebám byznysu.
- Z pohledu využití prediktivní analytiky je dobré vedle specifikace základních charakteristik těchto disciplín **vyhodnotit** alespoň rámcově jejich **vztahy a rozdíly** k jejím podstatným charakteristikám. V případě této kapitoly byla tato hodnocení uvedena k následujícím disciplínám:
  - Business Intelligence.
  - Competitive Intelligence.
  - Data Science.
  - Data mining.
  - Text mining.
  - Machine learning.
  - Big Data Analytics.



## 17. Role v prediktivní analytice



Role v prediktivní analytice a při řešení projektů v této oblasti zahrnují **obdobnou škálu rolí**, jako v jiných typech projektů.

Na druhé straně se zde zdůrazňují role na **principu stoličky o třech nohách** („*The Three-Legged Stool*“), a to:

- doménový expert,
- data a databázový expert,
- expert prediktivního modelování.

**Účelem** kapitoly je vymežit hlavní **funkční náplň** uvedených rolí.

Další podkapitoly obsahují hlavní úkoly a funkce uvedených expertů.

### 17.1 Doménový expert

Doménový expert **rozumí byznys problematice**, a tedy potřebám byznysu, jednotlivých oblastí řízení vzhledem k možnostem prediktivní analytiky:

- specifikuje charakter a **nároky na prognózování a plánování** v rámci firmy i k jednotlivým oblastem jejích aktivit,
- formuluje **zadání nových plánovacích aplikací** a reportů, specifikuje nároky na jejich kvalitu, tj. funkcionalitu, dostupnost, uživatelské rozhraní apod.,
- **konzultuje požadavky** na řešení prediktivní analytiky v rámci své oblasti řízení, a to z pohledu hlavních potřeb firmy,
- zajišťuje **hodnocení funkcionality** stávajících řešení a aplikací prediktivní analytiky vzhledem k vývoji potřeb firmy, případně k vývoji požadavků partnerů,
- určuje **požadavky na funkce operací** prediktivní analytiky,
- **analyzuje požadavky** uživatelů na prediktivní funkce, řeší jejich konsolidaci,
- poskytuje informace a zkušenosti o **stavu a kvalitě relevantních zdrojových databází**,
- definuje specifikace **cílových proměnných**, resp. ukazatelů,
- konzultuje specifické otázky **transformací dat**,
- podílí se na řešení **problémů, spojených s kvalitou dat**,
- podílí se na **kvalifikační přípravě** všech ostatních uživatelů.

### 17.2 Databázový expert

Databázový expert se orientuje v datových zdrojích a datových strukturách. Zajišťuje podle jednotlivých úloh řešení prediktivní analytiky tyto **činnosti**:

- zajišťuje **analýzu datových struktur zdrojových systémů**, jejich kvalitu a dostupnost,
- spolupracuje na **vyhodnocení stavu a kvality dat** na formulaci principů řízení jejich kvality,

- konzultuje a posuzuje možné **varianty přístupů k řešení** prediktivní analytiky z pohledu datové architektury,
- spolupracuje na **specifikaci nástrojů pro transformace dat**,
- spolupracuje s doménovým expertem na **určení oblastí**, jež budou z pohledu datových zdrojů **prioritní**,
- definuje **detailní pravidla a procedury transformací dat**, popis transformací polí mezi zdrojovými a cílovým systémem (datové typy, délka polí, plnění konstantami).

### 17.3 Expert prediktivního modelování

**Expert prediktivního modelování** je specializovaný na vytváření modelů a využívání jednotlivých metod a algoritmů. Zajišťuje podle jednotlivých oblastí řešení tyto **činnosti**:

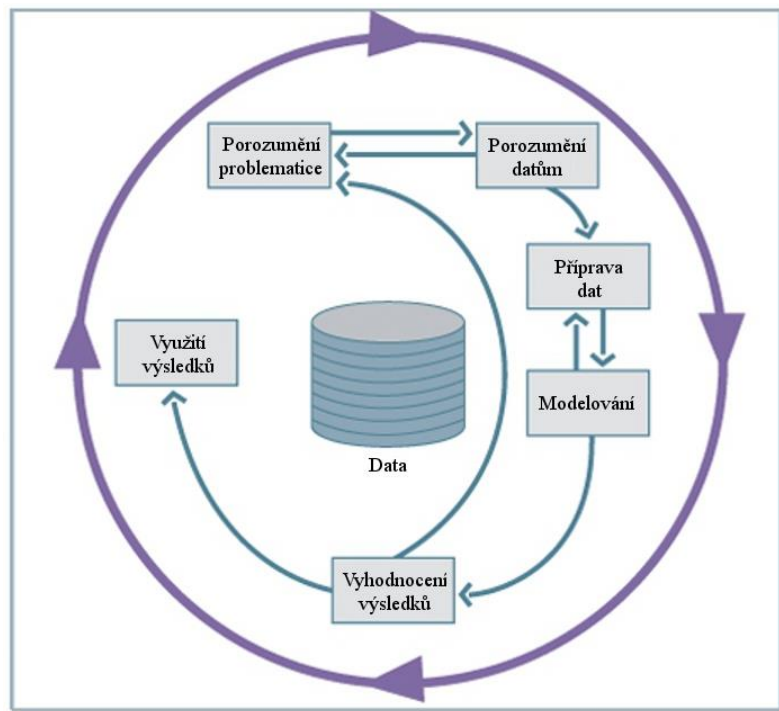
- **konzultuje s doménovými experty** jejich problémy a požadavky na aplikace pro podporu prognózování a plánování ve firmě,
- kooperuje na **analýzách požadavků** na prediktivní modely a na jejich konsolidaci,
- posuzuje **kvalitu zdrojových databází**,
- specifikuje rámcový **obsah a strukturu výstupů**, cílových proměnných, jejich hodnot a vizualizace,
- definuje základní **funkcionalitu** prediktivních modelů,
- zajišťuje **analýzu současného stavu** disponibilních zdrojových databází a aplikací,
- řeší **analytické otázky transformací dat**, tj. vlastní transformace dat,
- **navrhuje kontrolní a opravné procedury** v souvislosti s čištěním, resp. zajištěním kvality dat,
- podílí se na **řešení problémů nebo chyb**, vyplývajících z přípravy provozu prediktivní analytiky.

### 17.4 Závěry k rolím v prediktivní analytice



- Řešení prediktivní analytiky je charakteristické **vysokými nároky na kooperaci doménových expertů, databázových expertů i expertů prediktivního modelování**. S rostoucí komplexností a složitostí prediktivních systémů tyto nároky dále rostou. Proto **vymezení struktury a náplně jednotlivých rolí** jsou z tohoto pohledu velmi významné.
- Výše uvedené role tvoří pouze **podmnožinu**, které se úzce váže na jednotlivé součásti řešení prediktivní analytiky. Je nezbytné pak **brát v úvahu i další role, zejména manažerské**.
- Náplň jednotlivých rolí je kromě zmíněné kooperace účelná i pro systematické plánování a **řízení kvalifikačních programů** pro prediktivní analytiku.
- Náplň jednotlivých rolí je nezbytné **modifikovat podle potřeb** a možností firmy a zejména struktury jejich personálních zdrojů.

## 18. Postupy řešení a organizace prediktivní analytiky



**Obrázek 18-1: Fáze procesního modelu CRISP-DM** (CRISP-DM, n.d.)  
(Berka, 2003, stránky 24-28)

**Nejčastěji užívanou metodikou již od roku 1990 je CRISP-DM.** Primárně se váže k projektům data miningu. Je to dáno tím, že prediktivní analytika a data mining mají k sobě velmi blízko, používají i obdobné algoritmy. Prediktivní analytika tvoří jistou nadstavbu nad data miningem.

**CRISP-DM** značí „**Cross-Industry Standard Process for Data Mining**“ (CRISP-DM, n.d.). **Cílem** je vytvoření standardního **procesního modelu, bez vazby na konkrétní prostředí**, zdarma a veřejně dostupného, se zacílením na praktickou aplikaci prediktivní analytiky (povětšinou komerční). Obsahuje nejen jednotlivé kroky a postupy, ale i nezbytnou dokumentaci, která má být v projektu k dispozici. Metodika je užitečná pro analytiky i proto, že jim poskytuje argumentaci pro manažery, co v rámci projektu dělají a proč to dělají.

Návraty na předchozí fáze jsou dány např. nedostatečnými daty a jejich kvalitou pro vytváření modelů, nebo se při vyhodnocení modelů dojde k tomu, že výsledky nejsou adekvátní předpokladům apod.

CRISP-DM zahrnuje přímo kroky, věnující se **formulaci zadání společně s pochopením problematiky a finálního využití výsledku.**

## 18.1 Pochopení byznysu („Business Understanding“)



Nejdůležitější částí každého projektu prediktivní analytiky je úvodní část, **definující celý projekt**, kde hlavním předmětem zájmu v této fázi jsou vymezení a pochopení byznysu dané firmy a analýza potřeb v oblasti prediktivní analytiky.

**Účelem** fáze je:

- specifikovat hlavní **požadavky** na prediktivní modely,
- definovat **potřeby a cíle byznysu** v souvislostech plánování a prognózování tak, aby bylo zřejmé, jaké jsou důvody pro řešení prediktivní analytiky,
- definovat **základní charakteristiky dat** vzhledem k potřebám analytiky, zejména cílové proměnné, jednotky analýzy a další,
- formulovat **zadání**, vhodné pro úlohy prediktivní analytiky.

Na úvod jde o velmi náročnou fázi vyžadující **znalosti dané domény byznysu, znalosti dat** a databází a **znalosti prediktivních modelů**. Z tohoto důvodu musí být většinou **řešena týmově**. Klíčová role je zde přisuzována praktikům.

### 18.1.1 Definování vstupních požadavků na prediktivní modely

V této úvodní fázi metodiky CRISP-DM pro prediktivní analytiku se analyzují všechny **problémy a požadavky na řešení** podle konkrétních podmínek firmy. **Cílem** je přesně určit a poznat uživatele a účel prediktivního modelu. Musí být známo, **k čemu model bude sloužit**, na co bude aplikován a jak svojí predikcí pomůže efektivně plánovat byznys. Určí se odpovědi na **dvě základní otázky** (SIEGEL, 2013):

- Co má být predikováno?
- Jak bude s predikcí naloženo?

Definují se **metriky vyhodnocení modelu, postup řešení** a implementace prediktivního modelu.

### 18.1.2 Byznys cíle

Fáze rovněž spočívá v **identifikaci cílů** úlohy, **porozumění zadání** formulovaného manažery (vycházející z byznys problémů a cílů). Na identifikaci cílů se podílejí všichni tři výše uvedení experti. Prediktivní modelování pokrývá široký rozsah byznys cílů. **Existuje šest klíčových problémů**, které je třeba ve fázi „Pochopení byznysu“ řešit:

- Jaké jsou hlavní („core“) byznys cíle, na které se mají zaměřit prediktivní modely?
- Jak mohou být byznys cíle kvantifikovány?
- Jaká data jsou k dispozici pro kvantifikaci byznys cílů?
- Jaké metody modelování mohou být využity pro popis nebo predikci byznys cílů?
- Jak lze kvantifikovat správnost modelu vzhledem k byznys cílům, jak výsledky modelu odpovídají potřebám byznysu?
- Jak mohou být prediktivní modely distribuovány a dodávány uživatelům?

Byznys cíle vycházejí primárně z kvalifikovaného pochopení obsahu řízení firmy a nároků na analytiku podle jednotlivých oblastí řízení.

### 18.1.3 Definování dat pro prediktivní modelování

V úvodní fázi se rovněž specifikují charakteristiky dat pro další úlohy. Platí zde tyto **základní principy** (Abbott, 2014):

- Data pro prediktivní modelování musí být **dvoudimenzionální – řádky, sloupce**.
- Každá **řádka** reprezentuje **jednotku analýzy** („unit of analysis“). Např. pro zákaznickou analytiku je to **zákazník**. Pro analytiku call center to může být jeden **hovor**. Pro analytiku podvodů jedna transakce. Jednotky analýzy jsou úzce vázány na byznys obsah prediktivní analytiky.
- Sloupce v tabulkách se často označují jako **atributy, deskriptory, proměnné, pole, vlastnosti**.
- **Proměnné** se chápou jako **ukazatelé (measures)**, které se vztahují k záznamu nebo ho popisují. Prediktivní modelování považuje sloupec za separátní ukazatel, pořadí a struktura musí být standardní.
- Data jsou obvykle získávána **ze souborů, různých formátů**, např.:
  - „Flat files“ (csv soubory), položky mohou být odděleny čárkou, specifickým znakem, tabulátorem.
  - „Flat files s pevnou délkou záznamu“ s pevným počtem znaků na položku.
  - Ostatní kustomizované „Flat files“.
  - Binární soubory na bázi specifických formátů, odpovídajících softwarovým balíkům (SPSS, SAS, Matlab a další).
- Data mohou být zpřístupněna **přímo z databází** (tabulky, views) na bázi nativních driverů nebo ODBC. Některé software umožňují komplexní dotazy pro získávání data pro predikce, což umožňuje:
  - data nemusí být speciálně ukládána na disky,
  - data se mohou udržovat přímo v databázi, nemusí být udržována ve flat souborech,
  - analytici mají větší kontrolu a flexibilitu ve využití dat.
- Musí být zajištěno, že tabulky a pohledy **musí zůstat v průběhu predikcí stejné**.

Důležitá součást fáze je definování **jedné nebo více cílových proměnných**. Cílová proměnná je taková, jejíž **hodnota je predikována nebo hodnocena** v prediktivním modelu a může být **numerická nebo kategoriální** (případně binární).

### 18.1.4 Definování jednotek analýzy

Prediktivní modelování předpokládá, že každý **záznam** (řádka) je **nezávislé pozorování**. To znamená, že algoritmy neuvažují přímé vazby mezi záznamy. **Není ale vždy jasné**, které jednotky analýzy použít. Obvyklým příkladem řešení jednotky analýzy je **osoba: zákazník**.

Pokud ale např. zákazník – host **navštíví např. hotel vícekrát**, pak každá návštěva bude jednotka analýzy. Záleží na požadovaném detailu analýzy. V některých případech se vytvářejí **variantní modely** podle jednotek analýzy (zákazník, návštěva, sumarizace návštěv), podle potřeb rozhodování a byznysu.

### 18.1.5 Závěry k „Business Understanding“



- Základem je **identifikace cílů** úlohy, **porozumění zadání** formulovaného manažery (vycházející z business cílů).
- Identifikace cílů se řeší **týmově**, podílejí se na ni všichni **tři experti**, tedy doménový expert, databázový expert i expert pro tvorbu modelů.
- **Existuje šest klíčových problémů** „Pochopení byznysu“. Jedním z hlavních je „**Jaké jsou hlavní („core“) byznys cíle, na které se mají zaměřit prediktivní modely?**“
- Na úvod se analyzují všechny **problémy a požadavky na řešení** podle konkrétních podmínek firmy.
- Data pro prediktivní modelování musí být **dvoudimenzionální – řádky, sloupce**.
- Každá **řádka** reprezentuje **jednotku analýzy** („unit of analysis“).
- Sloupce v tabulkách se často označují jako **atributy, deskriptory, proměnné, pole, vlastnosti**.
- **Proměnné** se chápou jako **ukazatelé (measures)**, které se vztahují k záznamu nebo ho popisují. Podstatnou součástí je definování **jedné nebo více cílových proměnných**.
- Cílová proměnná je taková, jejíž **hodnota je predikována nebo hodnocena** v prediktivním modelu.
- Prediktivní modelování předpokládá, že každý **záznam** (řádka) je **nezávislé pozorování**. To znamená, že algoritmy neuvažují přímé vazby mezi záznamy.

## 18.2 Porozumění datům („Data Understanding“)



Fáze se zabývá **získáním dat** a jejich následnou **analýzou**.

**Účelem** fáze je:

- zajistit identifikaci a **vyhodnocení datových zdrojů**, použitelných pro úlohy prediktivní analytiky,
- analyzovat jak **interní**, tak **externí** datové zdroje,
- připravit a realizovat potřebné aktivity v souvislosti se zajištěním **potřebné kvality dat**.

V návaznosti na pochopení byznysu a jeho potřeb z předchozí fáze se řeší **problematika „pochopení dat“** a datových zdrojů, na kterých je řízení firmy založeno, a současně i **identifikace hlavních problémů v datech**, které je třeba v souvislosti s přípravou prediktivních modelů řešit.



- Pro porozumění datům a řešení jejich kvality se ve větší míře využívají i **vizualizační nástroje**.
- Při získávání a sběru nových dat je účelné **identifikovat** ta, která lze efektivněji získat z **existujících databází** a dostupných datových zdrojů.
- Při sběru nových dat je nejprve třeba je **vyhodnotit z pohledu jejich důvěryhodnosti** a využitelnosti v řešení prediktivních modelů.
- Při čištění dat je účelné se zaměřovat i na **anomálie v datech**.
- Testování a čištění dat je dobré v průběhu řešení realizovat **co nejdříve**.

### 18.3 Příprava dat („Data Preparation“)



Úkolem je **přípravit finální datový soubor**, který bude zpracováván konkrétními algoritmy. **Účelem** fáze je:

- realizovat přípravu finálního datového souboru,
- zajistit **kontroly, čištění dat**,
- realizovat **doplňující výpočty**.

Příprava dat je fáze časově i pracovně mimořádně náročná a podle různých studií představuje až **80 % celkového času řešení** prediktivní analytiky. Znamená **shromáždění dat, jejich kontroly, čištění a vytváření doplňujících dat** na základě odvozených proměnných. Navíc obvykle platí, čím více dat se pro prediktivní modely zajišťuje, tím větší riziko chyb a nekvality je s nimi spojeno, a je tedy i větší náročnost na jejich čištění.

#### 18.3.1 Organizace a shromáždění dat

Jedná se většinou o **nejpracnější část** celého procesu prediktivní analytiky. Tato fáze zahrnuje už vlastní operace s daty nad datovým zdrojem /zdroji. **Hlavní činností** je především **shlukování, třídění a řazení dat k sobě podle byznys významu a potřeb predikce**. Tato část vyžaduje výbornou technickou a byznys znalost dat. Je nutné, aby data byla obsáhlá a úplná, protože i zdánlivě nerelevantní proměnné a vztahy mezi nimi se mohou ukázat jako důležité.

**Nerelevantní proměnné** (například s nulovými hodnotami, nebo náhodně generované hodnoty a identifikátory) se poté **vyřadí ve fázi tvorby modelu**. Data je nutné mít co nejúplnější a bezchybná, protože **kvalita datového základu** přímo ovlivňuje úspěšnost predikce prediktivních modelů.

#### 18.3.2 Čištění dat

Čištění dat v sobě zahrnuje **procházení záznamů a hledání chybných, neaktuálních nebo nekompletních dat**. Takové záznamy mohou mít dopad na přesnost prediktivní-



ho modelu a ideálně by měly být opraveny nebo doplněny, aby byl konečný model co nejpřesnější (NYCE, 2007). Předpokladem pro efektivní řešení kvality dat je **jasná specifikace zdrojů dat**, odkud data pocházejí a s jakými riziky je třeba vzhledem k existujícím zdrojům počítat (např. externí zdroje, zcela nové zdroje dat, struktura a kvalifikace pracovníků pořizujících data apod.). Na specifikaci a analýzu zdrojů dat pak váže **profilování dat**, tj. jejich struktur, vlastníků a zodpovědností, potenciálních problémů apod.

**Problémy nebo chyby**, které je třeba řešit, jsou obvykle tyto (Wilson, 2021):

- **chybějící data** – naskýtají se tyto možnosti řešení:
  - vynechání celého řádku nebo sloupce,
  - doplnění předpokládaných nebo fixních hodnot,
  - doplnění průměrné hodnoty, mediánu nebo modu, podle konkrétní situace,
- **data mimo obvyklé hodnoty („outliers“)** – jsou to hodnoty, které jsou extrémně velké, malé, nebo leží mimo obvyklé tolerance. Problém je však v tom, že tyto hodnoty nemusí být nakonec „outliers“ a záleží na jejich kvalifikovaném prověření. Řešením mohou být:
  - eliminace takových hodnot,
  - využití průměrné hodnoty, mediánu či modu, nebo nahrazení obvyklou hodnotou,
  - řešení pomocí agregace a úpravy stávajících hodnot,
- **duplicitní data** představují častý problém a jejich řešení je následující:
  - vynechání celého řádku nebo sloupce,
  - využití klíčů pro identifikaci duplicit a jejich následné řešení,
  - využití vizualizace k určení míst duplikátů a určení jejich řešení,
- **problémy formátování dat** mohou být nejvíce časově náročné na jejich řešení, možnosti jsou např. následující:
  - rozdělení sloupce na více sloupců podle různých formátů dat,
  - zjištění, kde jsou numerická a kategoriální data a podle toho jejich rozdělení,
  - přesná identifikace formátovacích problémů (např. s využitím klíčových hodnot) a jejich následné řešení.

### 18.3.3 Vytvoření odvozených proměnných

V této fázi prediktivní analytiky se dále vytvářejí **odvozené vypočítané hodnoty / ukazatele** na základě analýz, zkušeností a známých faktů z daného oboru. Tyto hodnoty mohou mít v prediktivních modelech potencionálně velice **vysoké váhy**, proto je potřeba věnovat jim zvýšenou pozornost a konzultovat je s ostatními pracovníky.

Finální datový soubor je připravován s přihlédnutím **ke konkrétnímu zvolenému algoritmu** (schopnost algoritmu zacházet s chybějícími hodnotami, vyžadované datové typy na vstupu, rozsah datového souboru).

Poměrně často dochází **k návratu do této fáze** z kroku modelování, ať již z důvodů odhalených nepokrytých chyb v datech, či v případě nutných změn, souvisejících s konkrétním algoritmem.

### 18.3.4 Závěry k „Data Preparation“



- Příprava dat představuje většinou **nejpracnější část** celého řešení prediktivní analytiky.
- Zahrnuje **shlukování, třídění a řazení dat k sobě podle byznys významu a potřeb predikce**.
- Čištění dat v sobě zahrnuje **procházení záznamů a hledání chybných, neaktuálních nebo nekompletních dat**.
- Obvykle se musí **řešit tyto problémy**:
  - chybějící data,
  - data mimo obvyklé hodnoty („outliers“),
  - duplicitní data,
  - problémy formátování dat.
- Součástí této fáze je i to, že se dále vytvářejí **odvozené vypočítané hodnoty / ukazatele** na základě analýz, zkušeností a známých faktů z daného oboru.

## 18.4 Modelování („Modeling“)



V rámci fáze jsou aplikovány **analytické metody** (algoritmy pro úlohy prediktivní analytiky), dochází k hledání optimálních **nastavení parametrů** pro jednotlivé algoritmy.

**Účelem** fáze je:

- specifikovat **vhodnou metodu** pro modelování úloh prediktivní analytiky,
- vytvářet **prediktivní model**,
- určovat a vytvářet **větší celky** pro řešení modelů.

„*Prediktivní model je kombinace metod, kde primárním cílem je predikovat pravděpodobnost kategoriálních nebo kontinuálních hodnot s využitím hodnot prediktorů*“ (Wilson, 2021).

Vzhledem k faktu, že **neexistuje optimální metoda** pro všechny úlohy, doporučuje se hledat vhodnou metodu a vhodné nastavení parametrů, pro definitivní výběr nejlepší metody porovnat výsledky různých nastavení a různých metod.

Pokud jsou data připravena a zkontrolována, **je součástí této fáze** výběr přístupů a metod, specifikace odhadů, specifikace předpokladů, generování predikcí i ověřování získaných výsledků.

### 18.4.1 Typy prediktivních modelů

Moderní prediktivní modely a jejich algoritmy jsou **postavené na principech strojového učení** (machine learning) (SIEGEL, 2013). Modely se různými způsoby **učí z historických dat**, v nichž **hledají významné vztahy a proměnné** vztahující se k cílové proměnné (proměnným), která je v historických datech známa a kterou je v konkrétním případě žádoucí predikovat.

Nejdůležitější vlastností prediktivních modelů je **generalizace** – schopnost naučit se v dostupných datech **jen to, co je důležité** a správně **vyhodnotit náhodné jevy a šum** v datech jako nevýznamné. Problémy přeučení a nedoučení, zapříčínující špatnou generalizaci, se týkají všech typů modelů. Řeší je různé techniky, od prořezávání větví rozhodovacího stromu, až po skladbu několika i různých modelů do jednoho velkého modelu (ELDER IV, 2003). V současnosti existuje množství prediktivních modelů.

### 18.4.2 Hledání vztahů – data mining, cluster analýza

Podstatou je **rozbor množiny dat, vybraných pro prediktivní analytiku**. Je to analýza dat, sloužící k **identifikování skrytých vazeb, vzorů a vztahů**. **Data mining** je **důležitá část prediktivní analytiky**, protože data a vztahy, které identifikuje jako relevantní, mohou být použity při vývoji prediktivního modelu. Data mining v procesu prediktivní analytiky představuje **získávání znalostí o vztazích** a výsledný **prediktivní model je aplikací těchto znalostí**. Hlavní předností data miningu je to, že zaznamenává všechny vztahy (nebo korelace), které jsou v datech přítomny, bez ohledu na to, co je zapříčinilo (NYCE, 2007).

**Cluster analýza** je v prediktivní analytice využita pro **hledání podobností v datech**. Pomocí algoritmů a metod seskupuje **objekty podobných vlastností do skupin**. Tato analýza může být využita k odhalení struktur v datech, ale neposkytuje interpretaci nebo vysvětlení, proč tyto struktury existují (STATSOFT, 2013).

### 18.4.3 Tvorba prediktivního modelu

„*The machine actually learns more about your next likely action by studying others than by studying you.*“ (SIEGEL, 2013).

Prediktivní modely jsou jedním **z nejdůležitějších základů prediktivní analytiky**, která na nich přímo staví. Modely z dostupných dat **analyzují historická chování** k posouzení **pravděpodobnosti výskytu predikovaného jevu**, např. že zákazník s určitými vlastnostmi zakoupí nějaký produkt, využije nějakou službu, onemocní, pokusí se o podvod, klikne na reklamní banner na webových stránkách atp.

**Nejdůležitější vlastností** prediktivních modelů je **generalizace**. Ta zaručuje, že model, naučený z historických dat (in-sample), **dokáže správně vyhodnotit data nová** (out-of-sample), která do tvorby a určení modelu nevstoupila.

V současnosti se používají komplexní prediktivní modely, využívající **principy strojového učení** (SIEGEL, 2013). Modeláři je tvoří jednoduše a rychle v moderních nástrojích, určených pro dolování dat.

Vytvořené modely se v programech **samy validují, optimalizují a vyhodnocují**. Existuje celá řada typů modelů, které se snaží o co nejpřesnější předpověď daného jevu a jsou vhodné na různé situace.

**Vývoj a výběr prediktivního modelu** znamená, že modely mohou mít různou velikost a tvar v závislosti na jejich složitosti a využití, pro které jsou navrženy (NYCE, 2007). Pro co nejvyšší přesnost může být použito **více modelů**, které jsou následně **porovnávané a kombinovány**.

**Vývoj a výběr** modelu se skládá z **následujících kroků**:

- Definují se **vstupní proměnné a jejich váhy**. Váhy proměnných jsou určeny automatizovaně pomocí algoritmů a strojového učení, nebo je modelář nastaví sám.
- Vybere se **co nejvhodnější model**, který co nejlépe „sedí“ na historická data (in-sample) a na nová a testovací data (out-of-sample).
- Model se **ladí a vylepšuje, probíhá úprava proměnných, jejich vah a samotného modelu** na základě testování, zpětné validace a nových dat.

#### 18.4.4 Soubor modelů (ensemble)

**Modely a jejich výstupy (predikce)** je možné **skládat do většího celku**. Vznikne tak jeden velký a robustní model, postavený na několika různých prediktivních modelech. Z povahy tohoto modelu se může zdát, že bude náchylnější k přeučení a na out-of-sample datech bude vykazovat horší výsledky než samostatné modely, ze kterých je složený. Je-li složený z více modelů, vstupuje do něj více proměnných, což popírá princip generalizace prediktivních modelů. Paradoxně ale **výsledky praxe ukazují, že tento model dosahuje lepší predikce** než samostatné modely, ze kterých je složený.

#### 18.4.5 Závěry k „Modeling“



- Prediktivní modely a jejich algoritmy jsou **postavené na principech strojového učení** (machine learning).
- Je nezbytné dobře pochopit klíčové **faktory, proměnné a parametry**.
- Modely se různými způsoby **učí z historických dat**, v nichž **hledají významné vztahy a proměnné**, vztahující se k cílové proměnné.
- Základem je zde **rozběr množiny dat, vybraných pro prediktivní analytiku**, tedy analýza dat, sloužící k **identifikování skrytých vazeb, vzorů a vztahů**.
- Modely z dostupných dat **analyzují historická chování** k posouzení **pravděpodobnosti výskytu predikovaného jevu**.
- Používají se komplexní prediktivní modely, využívající **principy strojového učení**.
- Existuje **celá řada typů modelů**, které se snaží o co nejpřesnější předpověď daného jevu a jsou vhodné na různé situace.
- **Vývoj a výběr prediktivního modelu** znamená, že modely mohou mít různou velikost a tvar v závislosti na jejich složitosti a využití.

- **Modely a jejich výstupy (predikce)** je možné **skládat do většího celku**. Vznikne tak jeden velký a robustní model.
- Klíčovou součástí návrhu modelu je to, že se definují **vstupní proměnné a jejich váhy**. Váhy proměnných jsou určeny automatizovaně pomocí algoritmů a strojového učení, nebo je modelář nastaví sám.

## 18.5 Vyhodnocení výsledků („Assessing Models“)



V této fázi dochází **ke kontrole dosažených výsledků** oproti manažerskému zadání. Jsou již k dispozici výsledné znalosti (modely, vzory), dosažené analytickými metodami.

**Účelem** kapitoly je:

- definovat **principy validace a ladění** prediktivních modelů,
- analyzovat **kvalitu a obsah** navržených prediktivních modelů,
- definovat **metriky** pro hodnocení úspěšnosti modelu,
- vymezit problémy **overfitting, tedy přeučení modelu** a na druhé straně **underfitting, resp. nedoučení modelu**,
- specifikovat celkový **přístup k hodnocení** modelu.

V této fázi je třeba zpětně zkontrolovat **celý proces** v souvislosti s manažerským zadáním a případně identifikovat nedostatečně zohledněné obchodní aspekty. Součástí tohoto kroku je i **rozhodnutí o využití získaných** znalostí.

### 18.5.1 Validace a ladění prediktivních modelů

Aby se zajistilo, že je naučený prediktivní model co nejpřesnější, **musí být testován pomocí skupiny dat out-of-sample, testovacích dat**, která nijak nevstoupila do vývoje a učení modelu. Ověřuje se schopnost predikce modelu na této skupině dat a **porovná se, jak moc se odchyluje od výsledků z učících dat**. Pokud je odchylka velká, znamená to, že model není optimálně generalizován. Nabyté znalosti jsou poté aplikovány na model, který se podle nich upraví a následně se opět testuje.

Z důvodu vlastnosti generalizace **nejsou prediktivní modely při predikci úspěšné na 100 %**. V každém prediktivním modelu se objevují **2 základní typy chyb** (NYCE, 2007):

- Typ I. – chybně predikovaný výskyt ve skutečnosti není výskyt.
- Typ II. – chybně predikovaný nevýskyt je ve skutečnosti výskyt.

**Využívané metriky modelu**, pomocí kterých se vyhodnocuje **úspěšnost modelu, jsou Lift, ROC, Missclassification Rate**.

**Lift** je **ukazatel popisující úspěšnost predikce** a použitelnost prediktivního modelu nad bází dat. Definuje, jak moc je model úspěšnější než náhodný výběr (SIEGEL, 2013). Úspěšnost prediktivních modelů se **nad mění se hloubkou báze dat různí**. Pro usnadnění porovnání modelů nad různými hloubkami báze se zavedl pojem decil.

**Jeden decil obsahuje 10 % záznamů skórované báze**, seřazené sestupně podle predikovaného skóre. V prvním decilu obsahuje náhodný výběr 10 % všech pozitivních targetů, ve dvou decilech 20 % a v celé bázi dat jsou odhaleny všechny pozitivní targety. Nejvyššího liftu modely zpravidla dosahují na začátku v prvním decilu, dále má klesající tendenci. Lift náhodného výběru je 1 nad celou bázi dat. **Úspěšný model správně určí v prvních čtyřech decilech 80 % pozitivních targetů** (ČIČO, 2013).

**ROC (Receiver operating characteristic)** je dalším typem zobrazení úspěšnosti predikce modelu. Využívá se pro vyhodnocení úspěšnosti predikce binárního targetu.

**Misclassification Rate (míra chybovosti)** je poměr špatně vyhodnocených případů k celku. Výpočet hodnoty misclassification rate = počet špatně klasifikovaných (pozitivních i negativních targetů) / celkový počet klasifikací (ZHONG, 2007).

**Overfitting (také overlearning), přeučení modelu** znamená, že model **špatně vyhodnocuje náhodný šum v datech**, určuje důležité vztahy na základě náhodných proměnných a postrádá schopnost generalizace. Úspěšnost predikce modelu na nových datech je oproti ideálnímu stavu snížena. Tento nežádoucí stav může být mimo jiné **způsobený následujícími případy**:

- špatné nastavení modelu, například příliš velký (hluboký, rozvětvený) rozhodovací strom, příliš komplexní neuronová síť,
- příliš malý vzorek učicích dat,
- chyby ve vstupních proměnných, které nebyly řádně očištěny.

Přeučení znamená, že **model z dostupných dat předpokládá příliš mnoho**. Pokud bude například rozhodovací strom příliš hluboký a každý jeho list bude mít pouze jeden záznam, pak je strom, se všemi svými pravidly pouze jinou interpretací tréninkové množiny dat (SIEGEL, 2013). Na tréninkových datech bude tento model vykazovat 100% úspěšnost predikce, ale na testovacích, a hlavně nových datech bude model méně úspěšný, než kdyby byl optimálně generalizovaný.

**Underfitting (nedoučení modelu)** znamená, že **učení modelu bylo chybou nastavení modelu** nebo nedostatkem dat **zastaveno příliš brzy** a nebyly odhaleny všechny důležité vztahy. Model je příliš obecný a jednoduchý. Například nedoučený strom se skládá z příliš malého počtu pravidel a má málo listů.

**Pruning (prořezávání)** je jedna z metod **optimalizace modelů**. Účelem je **snížení komplexity** modelu. Například u rozhodovacích stromů představuje pruning „prořezávání“ větví modelu, za účelem snížení počtu větví a listů a tím snížení rizika overfittingu (SIEGEL, 2013), u neuronových sítí zase snížení počtu neuronů a vrstev Hidden Layer (MATIGNON, 2005). V praxi se k automatizaci tohoto procesu využívá validační skupina dat (SIEGEL, 2013).

### 18.5.2 Vyhodnocení prediktivního modelu

Po vytvoření prvních modelů se **výsledky těch nejúspěšnějších testují v praxi na nových datech**. Na základě predikce se **uskuteční rozhodnutí** a nastane nějaká akce. Poté se vyhodnocuje, jak moc predikce odpovídá realitě, **vyhodnocuje se úspěšnost**

**modelu** v praxi. V některých případech může být **objektivní vyhodnocení problematické**, protože provedená akce ovlivní chování jedince a není tak možné zjistit jeho chování, když by akce nenastala.

**Příkladem** může být **marketingová kampaň** telekomunikační společnosti zaměřená na odcházející klienty, kde prediktivní modely vyhodnocují pravděpodobnost odstoupení klienta od smlouvy a jeho přechodu ke konkurenci. Klienti s nejvyšším skóre jsou společnostmi osloveni a je jim nabídnuta nějaká výhoda. V tomto případě se však nedá přesně ověřit, zda klienti, vyhodnocení jako nejnáchylnější k odstoupení, skutečně odstoupit chtěli, tedy zda je model vyhodnotil správně jako nejnáchylnější k odstoupení. Dá se pouze měřit, kolik lidí skutečně odstoupilo a kolik ne. Řešení tohoto problému spočívá ve **využití kontrolních skupin**, kdy se záměrně část klientů s vysokým skóre neosloví a sleduje se u nich jejich chování, neovlivněné prediktivním modelem.

### 18.5.3 Závěry k „Assessing Models“



- Prediktivní model **musí být testován pomocí skupiny dat out-of-sample, testovacích dat**, která nijak nevstoupila do vývoje a učení modelu.
- Hlavní **metriky modelu**, pomocí kterých se vyhodnocuje **úspěšnost modelu, jsou Lift, ROC, Missclassification Rate**.
- **Overfitting**, resp. přeučení modelu znamená, že model **špatně vyhodnocuje náhodný šum v datech**, určuje důležité vztahy na základě náhodných proměnných a postrádá schopnost generalizace.
- **Problémy overfitting**, mohou být způsobeny následujícími případy:
  - špatné nastavení modelu, například příliš velký (hluboký, rozvětvený),
  - příliš malý vzorek učicích dat,
  - chyby ve vstupních proměnných.
- **Underfitting** je dáno tím, že **učení modelu bylo chybou nastavení modelu**, nebo nedostatkem dat **zastaveno příliš brzy** a nebyly odhaleny všechny důležité vztahy.
- Při vyhodnocování modelu se na základě predikce **uskuteční rozhodnutí**. Poté se vyhodnocuje, jak moc predikce odpovídá realitě, **vyhodnocuje se úspěšnost modelu** v praxi.

## 18.6 Využití výsledků („Model Deployment“)



V rámci fáze se specifikuje **další využití** získaných znalostí.

**Účelem** kapitoly je:

- specifikovat a zajistit dodání výsledků modelů konkrétním uživatelům,
- realizovat efektivní a názorné prezentace výsledků pro uživatele,
- specifikovat potřeby konsolidace výsledků modelu pro uživatele.

**Výstup** této fáze může mít podobu, sahající od seznamu pravidel, přes formulace získaných znalostí či analytické zprávy, až po praktická nasazení získaných modelů (např. pro klasifikaci zákazníka) do produkčních systémů podniku.

### 18.6.1 Prezentace výsledků

Prezentace získaných výsledků z modelů směřuje primárně na uživatele, tj. manažery, specialisty v oblasti prognózování nebo plánování v různých oblastech byznysu. Je zřejmé, že prediktivní analytika může přinést firmě efekty, pokud tito pracovníci výsledky dobře pochopí a budou je kvalifikovaně využívat. Proto právě prezentace výsledných informací patří k velmi významným součástem řešení úloh prediktivní analytiky. K tomu **několik poznámek a doporučení** (Wilson, 2021):

- prezentace by měla být orientována především **na byznys výsledky, získané efekty** a řešení stávajících byznys problémů, podstatně méně na technické aspekty řešení,
- základem efektivní prezentace má být **poznání účastníků prezentace** na uživatelské straně, zejména jejich potřeb,
- jádrem prezentace by mělo být **manažerské shrnutí** („*management summary*“),
- i když manažeři a plánovači pracují převážně s čísly, v případě prezentací je vhodnější se zaměřit převážně na „příběh“, tedy **řešení konkrétních problémů**, které výsledky modelů přinášejí,
- **cílem** má být nejen přinášet čísla, ale oslovit a přesvědčit uživatele, aby je využívali,
- velmi podstatnou, i když dnes už běžnou, má být **kvalitní vizualizace** prezentovaných výsledků.

### 18.6.2 Konsolidace informací

Kvalitní informace pro uživatele jsou ty, **kteří potřebuje, ale ne zbytečně víc**, neboť se s nimi bude obtížně vyrovnávat a složitě se v nich bude orientovat. Rovněž zde **několik poznámek**:

- obvyklá kritika se vztahuje k poskytovaným informacím, které dostatečně **rychle nevedou k požadovaným výsledkům**,
- **příliš mnoho času se věnuje způsobu**, resp. postupu analýzy oproti byznys podstatě výsledných informací,
- obvykle se doporučuje **samotnému postupu** řešení věnovat při prezentacích pouze **cca 10 % času** (oproti vlastnímu obsahu),
- naopak je účelné konsolidovat výstupní informace do logických celků, tedy v kontextu řízení celé firmy,
- konsolidované informace v širokém kontextu řešení **přispějí nejen k efektivnímu řízení**, ale svým způsobem budou podporovat zvyšování manažerské a analytické kvalifikace klíčových uživatelů.



### 18.6.3 Závěry k „Model Deployment“



- **Prezentace výsledných informací**, získaných na základě prediktivních modelů, patří ke konečné, ale **velmi významné součásti** celého postupu řešení prediktivní analytiky.
- Důležitým předpokladem je **poznání očekávání a potřeb zúčastněných uživatelů**, tedy manažerů a specialistů firmy.
- Efektivní využití poskytovaných výstupů z prediktivní analytiky závisí i na jejich **konsolidaci v širokém kontextu vzhledem k řízení** firmy.

## 19. Závěr

Text, věnovaný podnikové analytice, navazuje na publikace, vydané v několika posledních letech. Jejím účelem bylo v současné době, kdy podniková analytika je již běžnou součástí řízení firem prakticky napříč odvětvími i velikostními skupinami, poskytnout celkový pohled na hlavní principy a přístupy k řešení v této oblasti.

V textu, obdobně, jako i v jiných publikacích této řady, se zdůrazňuje analytický pohled, to znamená, že se zde neprezentují pouze jednotlivé charakteristiky řešení nebo nejrůznějších komponent podnikové analytiky, ale především se specifikují jejich očekávané efekty a potenciální problémy, které je třeba v souvislosti s jejich implementací a využitím řešit.

Tento text, věnovaný podnikové analytice, představuje základní přehled a pohled na její obsah. V rámci řady textů se dále předpokládá, že na něj budou následně navazovat texty, zaměřené na její konkrétní řešení a využití s promítáním odvětvových specifik, jako je „Výrobní analytika“, „Obchodní analytika“, „IT analytika“ a další.

## Přehled literatury

AALST, WIL van der. *Process Mining Data Science v akci*. Berlín : Springer, 2016. 978-3-662-49850-7.

ABBOTT, D.: *Applied Predictive Analytics. Principles and Techniques for the Professional Data Analyst*. John Wiley & Sons, Indianapolis, 2014. ISBN: 978-1-118-72796-6.

ANGEL, I. O., SMITHSON, S.: *Information Systems Management - Opportunities and Risks*, Macmillan, 1991.

BACAL, R.: *Manager's Guide to Performance Management*. New York, McGraw-Hill 2012. ISBN 978-0-07-177225-9.

BARC. *Business Intelligence, Analytics & Planning Tool Reviews* [online]. 2020. [cit.2020-02-12]. Dostupné z: <https://bi-survey.com/>

BERKA, P. 2003. *Dobývání znalostí z databází*. Praha: Academia, 2003. str. 366. ISBN 80-200-1062-9.

BOROVEC, J.: *Analýza a návrh využití variant řešení Cloud Business Intelligence v podnicích*. DP, VŠE, 2020.

BOTHE, O., KUBERA, O., BEDNÁŘ, D., POTANČOK, M., NOVOTNÝ, O.: *Managing Analytics for Success*, CRC Press, 2022. ISBN 978-1-032-20851-0.

BRUCKNER, T. VOŘÍŠEK, J., BUCHALCEVOVÁ, A. a kolektiv: *Tvorba informačních systémů: Principy, metodiky, architektury*, Grada, 2012, ISBN 978-802477-9027.

BUCHALCEVOVÁ, A.: *Metodiky vývoje a údržby informačních systémů*. Praha, Grada 2004. ISBN 80-247-1075-7.

BUCHALCEVOVÁ, A., PAVLÍČKOVÁ, J., PAVLÍČEK, L.: *Základy softwarového inženýrství, materiály ke cvičením*. Praha, VŠE 2008.

CAO, L.: *Data Science Thinking*. Springer, 2018. ISBN 978-3-319-95091-4.

CIMLER, P., ZADRAŽILOVÁ, D. a kol.: *Retail management*. Praha, Management Press, 2007. ISBN: 978-80-7261-167-6.

CRISP-DM - Home. *CRISP-DM - Home*. [Online] CRISP-DM. [Cited: 11 06, 2010.] <http://www.crisp-dm.org/>.

ČIČO, M. 2013. *Konzultace prediktivní analýzy*. Praha, 2013.

DARKWAH, Johnon. *Rozdíly mezi Data Lake a Data Warehouse*. In: *Gauss Algorithmic* [online]. 2018-06-20 [cit. 2020-03-13]. Dostupné z: <https://www.gaussalgo.com/datova-integrace/rozdily-mezi-data-lake-a-data-warehouse>.

DOHNAL, J., POUR, J.: *IT v řízení podniku*, Praha, Professional publishing 2016. ISBN 978-80-7431-160-4.

DOHNAL, J., PŘÍKLENK, O.: *CIO a podpora byznysu*. Praha, Grada, 2011. ISBN 978-80-247-4050-8.

- DOHNAL, J.: Řízení vztahů se zákazníky – procesy, pracovníci, technologie. Praha, Grada 2002.
- DRESNER, H.: Profiles in Performance. New York, John Wiley and Sons, 2010. ISBN: 978-0-470-40886-5.
- ECKERSON, W., W.: Performance Dashboards. New Jersey, John Wiley & Sons 2006.
- ELDER IV, J. 2003. The generalization paradox of ensembles. Journal of Computational and Graphical Statistics, 2003. 853-864.
- ENGLISH, L. P.: *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing profits*. New York, John Wiley & Sons 2003. ISBN 0-471-25383-9.
- FAUSKE, MAGNE, K. 2006. Example: Neural network. texample.net. [Online] 2006. [cit. 2013-11-04]. Dostupné z: <http://www.texample.net/tikz/examples/neural-network/>.
- FAYYAD, USAMA, M., a další: *Advances in Knowledge Discovery and Data Mining*. Cambridge : AAI Press/MIT Press, 1996. str. 560. ISBN: 0-26-256097-6.
- FIBÍROVÁ, J., ŠOLJAKOVÁ, L. 2010. *Reporting*. 3. vydání. Praha: Grada Publishing, a.s. ISBN 978-80-247-2759-2.
- FIBÍROVÁ, J., ŠOLJAKOVÁ, L., WAGNER, J., PETERA, P.: Manažerské účetnictví. Nástroje a metody. Praha, Wolters Kluwer, 2015. ISBN: 978-80-7478-743-0.
- FRIEBELOVÁ, J.: Rozhodovací stromy. *Osobní stránka - Jana Friebelová*. [Online] 28. 12 2006. [Citace: 10. 11 2010.] [http://www2.ef.jcu.cz/~jfrieb/rmp/data/teorie\\_oa/STROMY.pdf](http://www2.ef.jcu.cz/~jfrieb/rmp/data/teorie_oa/STROMY.pdf).
- GUYER, C., MILENER, G., KUMAR, S., HAMILTON, B., ROTH, J., LAUDENSCHLAGER, D. a RABELER, C.: 2017. *Master data services Overview (MDS)* [online]. Microsoft Docs. [cit. 2019-04-30]. Dostupné z: <https://docs.microsoft.com/en-us/sql/master-data-services/master-data-services-overview-mds?view=sql-server-2017>
- HALAMA, J.: Řízení datové kvality v Hadoop Ecosystem, DP, VŠE, Praha, 2021.
- HAMAN, Martin. 2018. *Data Day 2018* [online]. Dostupné z: [https://www.gopas.sk/getattachment/47219b58-3c18-427e-b304-c940604185a8/Martin\\_Haman\\_all.pdf.aspx;.php;.html](https://www.gopas.sk/getattachment/47219b58-3c18-427e-b304-c940604185a8/Martin_Haman_all.pdf.aspx;.php;.html).
- HAND, D, MANNILA, H a SMYTH, P. 2001. Principles of data mining (adaptive computation and machine learning). Cambridge, MIT Press, 2001. ISBN 978-0262082907.
- HINDLS, R., HRONOVÁ, S., NOVÁK, I.: Analýza dat v manažerském rozhodování. Grada, 1999. ISBN: 80-7169-255-7.
- HOLTSNIDER, B., JAFFE, B. D.: IT Managener's Handbook. Amsterdam, Elsevier 2012. ISBN 978-0-12-415949-5.
- CHALUPNÍK, V. 2012. Biologické algoritmy (5) - Neuronové sítě. Root.cz. [Online] 2012. [Cit. 2013-8-14]. Dostupné z: <http://www.root.cz/clanky/biologicke-algoritmy-5-neuronove-site/#ic=serial-box&icc=text-title>.

CHANDLER, N.: The CPM Scenario. Gartner BI Summit 2008.

CHRAMOSTOVÁ, V., POTANČOK, M., POUR, J.: Byznys analytika pro manažery, Oeconomia, Praha, 2020.

CHUDÁN, D.: Vymezení pojmů pokročilé analytiky, VŠE, 2023.

ISO certifikace – (MANages, a.s., [Online] @2009. [Citace: 26. září 2014.] 2009).

JANOSCHEK, N. Data Governance: Definition, Challenges & Best Practices. In: *BI Survey* [online]. [cit. 2020-03-01]. Dostupné z: <https://bi-survey.com/data-governance>.

JOYFUL CRAFTSMAN. 2018. *Hitchhiker's Guide to the SSAS Tabular* [online]. [cit. 2019-04-14]. Dostupné z: <https://www.sqlsaturday.com/779/Sessions/Details.aspx?sid=85530>.

KELLY, J.: *Big Data Vendor Revenue and Market Forecast 2012 - 2017*. [online] 2012 [cit. 18. září 2013]. Dostupné z: [http://wikibon.org/wiki/v/Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2012-2017](http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017).

KENNEDY, J. a SATRAN, M.: 2018. *Access Control Lists*. Microsoft Docs [online]. [cit. 2019-04-23]. Dostupné z: <https://docs.microsoft.com/en-us/windows/desktop/secauthz/access-control-lists>.

KIMBALL, R., ROSS, M.: *Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. Indianapolis, John Wiley Publishing 2010. ISBN 978-0-470-56310-6.

KLIPFOLIO.COM [online]. [cit. 2020-03-20]. Dostupné z: <https://www.klipfolio.com/resources/articles/what-is-mobile-business-intelligence>.

KNIGHT, M.: Data Lake vs. Data Swamp: Leveraging Enterprise Data. In: *Dataversity* [online]. 2018 [cit. 2020-03-14]. Dostupné z: <https://www.dataversity.net/data-lake-vs-data-swamp-leveraging-enterprise-data/>.

KOTTER, J., P.: *Vedení procesu změny: osm kroků úspěšné transformace – (Management Press 2000) – ISBN8072610155*.

KRÁL, B. a kol.: *Manažerské účetnictví*. Praha, Management Press 2006. ISBN 80-7261-141-0.

KRÁL, B., FIBÍROVÁ, J., MATYÁŠ, O., MENŠÍK, M., STRÁNSKÝ, J., ŠOLJAKOVÁ, L., WAGNER, J., ZRALÝ, M. a HALÍŘ, Z.: 2012. *Manažerské účetnictví*. 3. doplněné a aktualizované vydání. Praha: Management Press. ISBN 978-80-7261-217-8.

KUHN, M., JOHNSON, K.: *Applied Predictive Modeling*, Groton, Saline, USA, Springer, 2016. ISBN 978-1-4614-6848-6.

KULHAVÝ, L.: *Praktické uplatnění technologií Data Mining v pojišťovnictví - (VŠE, Praha 2011)*.

KUNSTOVÁ, R.: *Efektivní správa dokumentů*. Praha, Grada, 2009. ISBN 978-80-247-3257-2.

LABERGER, R.: *The Datawarehouse Mentor*. New York, McGraw Hill, 2011. ISBN: 978-0-07-174532-1.

LABERGER, R.: Datové sklady. Agilní metod y Business Intelligence. Praha, Computer Press, McGraaw Hill, 2012. ISBN 978-80-251-3729-1.

LANEY, D., B.: Infonomics, Bibliomotion, Inc., New York, 2018. ISBN 978-1-138-09038-5.

LOGI ANALYTICS [online]. [cit. 2020-03-19]. Dostupné z: <https://www.logianalytics.com/resources/bi-encyclopedia/mobile-bi/> .

MÁŠA, P. 2008. 4IZ560 – Data mining – praktické aplikace. 2008.

MATIGNON, R. 2005. Neural Network Modeling Using Sas Enterprise Miner. Bloomington : AuthorHouse, 2005. ISBN 978-1-4184-2341-4.

MATINGON, R. 2007. Data Mining Using SAS Enterprise Miner. Bloomington : Wiley, 2007. ISBN 978-0-470-14901-0.

MICROSOFT. *Dokumentace k Power BI* [online]. [cit.2020-02-12]. Dostupné z: <https://docs.microsoft.com/cs-cz/power-bi> .

NOVOTNÝ, O., POUR, J., BASL, J., MARYŠKA, M.: Řízení výkonnosti podnikové informatiky. Professional Publishing, Praha, 2010. ISBN 978-80-7431-040-9.

NYCE, C. 2007. *Predictive Analytics White Paper*. American Institute for CPCU/Insurance Institute of America, autor neznámý, 2007.

PALADINO, B.: Innovative Corporate Performance Management: Five Key Principles to Accelerate Results. Indianapolis, Wiley Publishing, 2011. ISBN: 978-0-470-62773-0.

PARMENTER, D.: Key Performance Indicators (KPI): Developing, Implementing, and Using Winning KPIs.

PEARLMAN, S.: ELT vs. ETL: Defining the Difference. In: *Talend* [online]. 2019-08-09 [cit. 2020-03-14]. Dostupné z: <https://www.talend.com/resources/elt-vs-etl/> .

POUR, J., MARYŠKA, M., STANOVSKÁ, I., ŠEDIVÁ, Z.: Self Service Business Intelligence, Praha. Grada, 2018. ISBN 978-80-271-0616-5.

PRAUS, O.: Prediktivní analýza – postup a tvorba prediktivních modelů. DP, VŠE, 2013.

*Profisee: Data Governance – What, Why, How, Who & 15 Best Practices*. In: *Profisee* [online]. 2020 [cit. 2020-02-25]. Dostupné z: <https://profisee.com/data-governance-what-why-how-who/> .

PROVOST, F., FAWCETT, T.: Data Science for Business. What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media. Sebastopol. 2013. ISBN: 978-1-449-36132-7.

RABELER, C., GUYER, C. a DUNCAN, O.: 2018. *Tabular Models*. Microsoft Docs [online]. [cit. 2019-04-14]. Dostupné z: <https://docs.microsoft.com/en-us/sql/analysis-services/tabular-models/tabular-models-ssas?view=sql-server-2017> .

RABELER, C., GUYER, C. a DUNCAN, O.: 2018. *Perspectives in tabular models* [online]. Microsoft Docs. [cit. 2019-04-29]. Dostupné z: <https://docs.microsoft.com/en-us/sql/analysis-services/tabular-models/perspectives-ssas-tabular?view=sql-server-2017> .

- ROOSEVELT, M. 2005. The use of predictive modeling in the insurance industry. místo neznámé, PINNACLE Actuarial Resources, INC, 2005.
- RUSSO, M. 2022 *Development tools for Tabular models in 2021*. [cit. 2022-05-22]. Dostupné z: <https://www.sqlbi.com/articles/development-tools-for-tabular-models-in-2021/> .
- RUSSO, M , FERRARI, A.: *Header/Detail vs Star Schema models in Tabular and Power BI* . 2021 [cit. 2021-10-11]. Dostupné z: <https://www.sqlbi.com/articles/header-detail-vs-star-schema-models-in-tabular-and-power-bi/>.
- RUSSO, M , FERRARI, A.: *Header/Detail vs Star Schema models in Tabular and Power BI* . 2021 [cit. 2021-10-11]. Dostupné z: <https://www.sqlbi.com/articles/header-detail-vs-star-schema-models-in-tabular-and-power-bi/> .
- REŠL, Š.: Využití prediktivní analytiky pro finanční plánování firmy, BP, VSE, 2020.
- ŘEPA, V.: Podnikové procesy. Praha, Grada 2007.
- SAMEH, S.: Databricks Delta Lake – A Friendly Intro. In: *DEV Community* [online]. 2020-01-09 [cit. 2020-03-13]. Dostupné z: <https://dev.to/samehsharaf/databricks-delta-lake-a-friendly-intro-2lap>
- SEGAL, T.: How Prescriptive Analytics Can Help Businesses. In: Investopedia [online]. 2019-07-05 [cit. 2020-03-16]. Dostupné z: <https://www.investopedia.com/terms/p/prescriptive-analytics.asp> .
- SHARPE, P. K. and SOLLY, R. J.: Dealing with missing values in neural network-based diagnostic systems. *NEURAL COMPUTING & APPLICATIONS*. 1995, pp. 73-77.
- SCHIESSER, R.: IT Systems Management. New York, Prentice Hall 2010. ISBN 978-0-13-702506-0.
- SIEGEL, E: Predictive Analytics. New York, John Wiley & Sons, 2016. ISBN 978-1-119-14567-7.
- SLÁNSKÝ, D.: Data and Analytics for the 21st Century: Architecture and Governance, Professional Publishing, 2018. ISBN 978-80-88260-16-5.
- SMITH, A. 2004. Branch Prediction with Neural Networks: Hidden layers and Recurrent. Connections. Department of Computer Science University of California : San Diego, 2004.
- SNÍTIL, J.: Datový sklad na technologiích IBM a jeho možnosti. DP, VŠE Praha, 2018.
- STATSOFT. 2013. Electronic Statistics Textbook. [Online] 2013. [Cit. 2013-7-15]. Dostupné z: <http://www.statsoft.com/Textbook> .
- ŠEVČÍK, O.: Současné trendy v Business Intelligence, BP, VŠE Praha, 2020.
- ŠOLJAKOVÁ, L. FIBÍROVÁ, J.: Reporting. Praha, Grada, 2010. ISBN 978-80-247-2759-2.
- SYNEK, M. a kol.: Manažerská ekonomika. Praha, Grada 2011. ISBN 978-80-247-3494-1.

SYNEK, M., KISLINGEROVÁ, E. a kol.: Podniková ekonomika. Praha, C H Beck 2015. ISBN 978-80-7400-274-8.

TALEND: Data Lake vs Data Warehouse: Key Differences. In: *Talend* [online]. [cit. 2020-03-13]. Dostupné z: <https://www.talend.com/resources/data-lake-vs-data-warehouse/> .

TEASLEY, B. :2005. Using Predictive Models, Part 1. Clickz.com. [Online] 2005. [Cit. 2013-9-4]. Dostupné z: <http://www.clickz.com/clickz/column/1707212/using-predictive-modelspart> .

TEJADA, Z.: 2018. *Technologie OLAP (Online Analytical Processing)* [online]. Microsoft Docs. [cit. 2019-04-14]. Dostupné z: <https://docs.microsoft.com/cs-cz/azure/architecture/data-guide/relational-data/online-analytical-processing> .

TOMEK, G., VÁVROVÁ, V.: Průmysl 4.0 aneb nikdo sám nevyhraje. Praha, Professional Publishing, 2017. ISBN 978-80-906594-4-5.

UČEŇ, P.: Zvyšování výkonnosti firmy na bázi potenciálu zlepšení. Praha, Grada 2008. ISBN: 978-80-247-2472-0.

VAN DECKER, J, CHANDLER, N.: Top Processes for Corporate Performance Management, Gartner, 2011.

VANĚK, J: Architektury datových skladů a jejich historický vývoj. DP, VŠE, 2020.

VEBER, J. a kol.: Management – základy, prosperita, globalizace. Praha, Management Press 2000.

*VertiPaaS engine in SSAS Tabular Database* [online]. 2017. [cit. 2019-04-30]. Dostupné z: <http://nialbi.blogspot.com/2017/11/vertipaas-engine-in-ssas-tabular-databse.html> .

VOMLELOVÁ, M.: Rozhodovací stromy. *Výuka - Marta Vomlelová*. [Online] 10 20, 2009. [Cited: 11 10, 2010.] <http://kti.mff.cuni.cz/~marta/slistromy.pdf>.

VOŘÍŠEK, J., POUR, J. a kol.: Management podnikové informatiky, Professional Publishing, 2012, ISBN 978-80-7431-102-4.

WILSON, J., E.: Predictive Analytics for Business Forecasting and Planning. Graceway Publishing Company, 2021. ISBN 978-0-9839413-8-5.

ZHONG, M. 2007. An Analysis of Misclassification Rates for Decision Trees. Florida, 2007. PhD Thesis. University of Central Florida Orlando.

ŽŮRKOVÁ, H.: Plánování a kontrola – klíč k úspěchu. 1. vydání. Praha: Grada Publishing, 2007. ISBN 80-247-1844-8.

ŽYDYK, R.: Návrh MLOps řešení pro výrobní společnost. DP, VŠE, 2022.



# Terminologický slovník

Application Programming Interface	<b>Application Programming Interface</b> – API – rozhraní aplikačních programů.
Atribut	<b>Atribut</b> – vlastnost entity nebo sloupec/pole u databázové tabulky či položka v datovém souboru. Datová část objektu, tj. proměnná, která je součástí objektu.
Big Data Analytics	<b>Big Data Analytics</b> : je výrazným fenoménem současné doby, který je dán extrémním nárůstem dat v souvislosti s uplatňováním IoT, sociálními sítěmi a dalšími.
Business Intelligence	<b>Business Intelligence</b> – BI – souhrn technologií, a nástrojů, které umožňují na základě aplikace multidimenzionálního přístupu k měření vytvářet a optimalizovat aplikace na podporu analytických a plánovacích aktivit podniku.
Cloud computing	<b>Cloud computing</b> – metoda přístupu k využití výpočetní techniky, která je založena na poskytování sdílených výpočetních prostředků a jejich využívání formou služby. Existují nejrůznější modely služeb a možnosti jejich poskytování, ale všem typům cloud computingu je společná schopnost poskytovat prostředky na vyžádání, elasticky, samoobslužně a prostřednictvím přístupu z rozsáhlé sítě a také schopnost měřit spotřebované služby v rámci sdíleného fondu prostředků.
Clustering	<b>Clustering</b> je přístup nebo kombinace metod, kde hlavním <b>cílem je seskupovat data</b> , která se k sobě váží do jednotlivých skupin.
Competitive Intelligence	<b>Competitive Intelligence</b> – CI – systematický a etický program pro získávání, analyzování a řízení externích informací o trhu, konkurenci a trendech, které mohou mít vliv na podnikové plány, rozhodování a jeho činnost.
Data Mining	<b>Data Mining</b> – dolování dat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází.
Data Science	<b>Data Science</b> – zahrnuje porozumění business logiky dat, přípravu dat, modelování / optimalizace / simulace, vyhodnocení a nasazení analytického modelu.
Databáze	<b>Databáze</b> – <b>DB</b> – data, které slouží více aplikacím, jsou v nich minimalizovány redundance dat a existuje vhodně centralizovaná správa těchto dat. Cílem databázového systému je uspořádat datové zdroje (datovou základnu) na počítači tak, aby tyto zdroje mohly být využívány.
Databázový server	<b>Databázový server</b> – Database Server, server, orientovaný na zpracování dat v databázích.
Datové tržiště	<b>Datové tržiště</b> – Data Mart – DMA – analytická databáze, určená pouze pro určitou část podniku (útvary, oblast řízení apod.).
Datová pumpa	<b>Datová pumpa</b> – ETL – Extract, Transform, Load – datová pumpa – software pro realizaci transformací dat mezi různými datovými zdroji.
Datový sklad	<b>Datový sklad</b> – Data Warehouse – DWH – integrovaný, subjektivně orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu.
Datový typ	<b>Datový typ</b> – označuje typ dat uložených v proměnné, např. číslo, text apod. Určuje množinu přípustných hodnot a operace s těmito hodnotami.
Diagram tříd	<b>Diagram tříd</b> – Class diagram – jeden z diagramů UML, vyjadřuje třídy a vztahy mezi nimi.
Diagram užití	<b>Diagram užití</b> – Use Case diagram – jeden z diagramů UML, vyjadřuje chování systému a vztah k okolí.

Dimenze	<b>Dimenze</b> – analytické hledisko pro identifikaci a hodnocení sledovaných ukazatelů, je tak součástí de facto každé metriky. Z informatického pohledu se jeví jako struktura dat, případně jako databázová tabulka obsahující záznamy o jednotlivých prvcích dimenze, tj. např. o jednotlivých službách informatiky, dodavatelích IT apod.
Dimenzionální modelování	<b>Dimenzionální modelování</b> – vymezení všech dimenzí, jejich obsahu, včetně vnitřní hierarchie prvků, a dílčích charakteristik jednotlivých dimenzí, určení soustavy sledovaných ukazatelů (faktů) a jejich dílčích charakteristik, specifikace vazeb mezi ukazateli a odpovídajícími dimenzemi.
Dimenzionální tabulka	<b>Dimenzionální tabulka</b> – tabulka, obsahující kontext k datům fakt tabulky, obsahuje obvykle textové popisné údaje.
Dotazovací jazyk	<b>Dotazovací jazyk</b> – umožňuje uživateli formulovat dotazy či získávat informace z databáze formulací požadavku podle definovaných standardů.
DSA	<b>DSA</b> – Data Staging Area – jde o místo v architektuře datového skladu, kde se data ze zdrojových systémů zpracovávají a transformují do podoby dat, definovaných dimenzionálním (cílovým) modelem.
Entita	<b>Entita</b> – označení obecného jasně definovaného prvku.
ERP	<b>ERP</b> – Enterprise Resource Planning – aplikace, mající celopodnikový charakter a pokrývající funkcionalitou většinu, resp. velkou část funkcí podnikového řízení.
ETL	<b>ETL</b> – Extract, Transform, Load – datová pumpa – software pro realizaci transformací dat mezi různými datovými zdroji.
Fakt	<b>Fakt</b> – metrika (zpravidla číselná hodnota), která je v podniku sledována.
Fakt tabulka	<b>Fakt tabulka</b> – tabulka, obsahující obchodní fakta a hodnoty, zpravidla číselné a aditivní.
Faktor	<b>Faktor</b> – souhrnné vyjádření pro organizační, ekonomické, technologické a další podmínky řešení a realizaci úloh řízení podniku a informatiky, např. úroveň managementu, existence nabídky cloud computingu, různých typů aplikací apod.
Granularita	<b>Granularita</b> – úroveň podrobnosti faktů, uložených ve fakt tabulce.
Implementace	<b>Implementace</b> – realizace navrženého systému v programovém prostředí.
In-memory computing	<b>In-memory computing</b> – výkonná databáze, umožňující real-time zpracování enormního množství operací s cílem zrychlení jednotlivých transakcí jak na úrovni procesů, tak celých modulů.
Key Goal Indicator	<b>Key Goal Indicator</b> – KGI – cílově orientovaná metrika, tj. měří výsledek procesu, např. počet úspěšně vyřešených incidentů v procesu řízení incidentů.
Key Performance Indicator	<b>Key Performance Indicator</b> – KPI – metrika výkonnosti jednotlivých složek procesu (činností a zdrojů procesu), tj. měří jednotlivé činnosti a zdroje procesu, např. dobu trvání činnosti, objem zpracovaných dokumentů, % využití doby operátora service desku apod.
Klíč	<b>Klíč</b> – identifikace entity – zvolený atribut entity, rozlišují se primární a sekundární klíče entity.
Kostka	<b>Kostka</b> – označení pro multidimenzionální databázi (nebo její princip). Kostka je realizována buď prostřednictvím multidimenzionální databáze, nebo zvláštní struktury relační databáze.
Neuronové síť	<b>Umělé neuronové síť</b> – vycházejí ze znalostí o biologických neuronových sítích. Umělé neuronové síť <b>se skládají z jednotlivých navzájem propojených neuronů.</b>

Machine Learning	<b>Machine Learning</b> obsahuje vymezení podstaty a kroků řešení úloh ML, vymezení možností produkcionalizace a governance řešení.
Metoda	<b>Metoda</b> – představuje známou a v rámci úloh modelu MBI aplikovatelnou metodu manažerského (např. BSC, RPZ atd.), ekonomického (např. ABC, ROI atd.) nebo infromatického charakteru (datové modelování, dimenzionální modelování atd.).
Metodika	<b>Metodika</b> – doporučený souhrn fází, etap, přístupů, zásad, postupů, pravidel, dokumentů, řízení, metod, technik a nástrojů pro tvůrce informačních systémů, který pokrývá celý životní cyklus informačního systému.
Metrika	<b>Metrika</b> – sledovaná a měřená hodnota ukazatele pro potřeby řízení podniku a informatiky. K ukazatelům se váží dimenze pro jejich identifikaci, analýzy a plánování. Vychází se přitom z principů dimenzionálního modelování.
Multidimenzionální databáze	<b>Multidimenzionální databáze</b> – databáze, kde jsou data uložena na principu vícerozměrové matice. Hodnoty jsou přístupné přímo pro danou kombinaci prvků dimenzí.
OLA	<b>OLA</b> – Operational Level Agreement – je formální mechanismus, který zajišťuje spolupráci interních a externích dodavatelů služeb IT při plnění jejich nejčastěji outsourcingového kontraktu.
OLAP	<b>OLAP</b> – Online Analytical Processing – informační technologie, založená především na koncepci multidimenzionálních databází. Jejím hlavním principem je několikadimenzionální tabulka, umožňující rychle a pružně měnit jednotlivé dimenze a měnit tak pohledy uživatele na modelovanou ekonomickou realitu.
OLAP server	<b>OLAP server</b> – technologie pro uložení a správu dat v multidimenzionálních databázích, pro jejich vytváření a plnění.
Rozhodovací strom	<b>Rozhodovací strom</b> je grafickou reprezentací logického vývoje časově na sebe navazujících alternativních rozhodnutí. (Friebeľová, 2006). Jde o zvláštní případ grafu, kdy rozhodovací strom je tvořen z: (Vomlelová, 2009).
Podnikový proces	<b>Podnikový proces</b> – proces, kterým podnik zajišťuje naplnění podnikových cílů, reaguje na významné události a zajišťuje produkci plánovaných výstupů (produktů, služeb apod.).
Prediktivní analytika	<b>Prediktivní analytika</b> je <i>typ analýzy, využívající data a prediktivní modely pro předpověď jevů na mikroekonomické úrovni</i> . Prediktivní analytika využívá technologie, které se učí ze zkušeností (dat), aby předvíдалa budoucí chování jedinců k lepšímu rozhodování (Siegel, 2013).
Produkční databáze	<b>Produkční databáze</b> – databáze uchovávající data, která vytvářejí a spravují transakční informační systémy; představují hlavní vstup do řešení Business Intelligence.
RACI	<b>RACI</b> – matice RACI přiřazuje a zobrazuje odpovědnosti jednotlivých osob či pracovních míst za danou oblast v organizaci: R - Responsible (vykonává), A - Accountable (zodpovídá), C - Consulted (konzultuje), I - Informed (má být informován).
Referenční model	<b>Referenční model</b> – model, který je, vedle metodické stránky řešení, naplněn předdefinovaným obsahem. Tento obsah, který tvoří například podnikové procesy a jejich prvky, vzniká a postupně se rozvíjí na základě poznatků a zkušeností z dosud realizovaných projektů pro různé zákazníky v praxi.
Referenční proces	<b>Referenční proces</b> – proces či skupina procesů v rámci referenční metodiky.
Relační databáze	<b>Relační databáze</b> – databáze splňující dvě základní podmínky: databáze je chápána jako množina relací, jsou k dispozici minimálně operace selekce, projektování a spojení.

Report	<b>Report</b> – výstup software s výsledky hodnot ukazatelů. Může být vyjádřen tabulkou, grafem atd.
RI	<b>RI</b> – Results Indicators – se vztahují k dějům, co byly ukončeny. Může jít o děje dlouhodobé, např. fáze, dlouhodobý projekt, nebo o děje krátkodobé, např. aktivita nebo proces.
Role	<b>Role</b> – představují typové skupiny pracovníků, charakterizované vykonáváním obdobných činností. Jeden člověk může zastávat více rolí. Role mají definované odpovědnosti, které zahrnují provádění určitých činností.
Řízení výkonnosti	<b>Řízení výkonnosti</b> – kombinace managementu, metodik a metrik, podporovaná aplikacemi, nástroji a infrastrukturou, která umožňuje uživatelům definovat, monitorovat a optimalizovat výsledky a výstupy tak, aby bylo dosaženo cílů osobních či cílů organizační jednotky v souladu se strategickými cíli, stanovenými na různých úrovních řízení podniku (osobní, procesní, skupinové a korporátní cíle podniku nebo podnikatelského ekosystému).
Řízení výkonnosti IT	<b>Řízení výkonnosti IT</b> – IT PM – IT Performance Management – koncept řízení výkonnosti IT, založený na vztazích metod, procesů, metrik a aplikací řízení IT.
Scénář	<b>Scénář</b> , resp. životní situace – v řízení podniku a informatiky vymezuje jeho podstatu a možnosti, jak danou životní situaci řešit pomocí úloh.
Slovník dat	<b>Slovník dat</b> – Data Dictionary – prostředek centrálního popisu datových struktur.
Složený klíč	<b>Složený klíč</b> – klíč, tvořený více atributy.
Snowflake schéma	<b>Snowflake schéma</b> – Schéma sněhová vločka – typ dimenzionálního databázového modelu, tvořeného více normalizovanými a nenormalizovanými tabulkami dimenzí.
Softwarová architektura	<b>Softwarová architektura</b> – určuje, z jakých modulů bude software postaven a jaké vazby budou existovat mezi těmito moduly.
Sourcing	<b>Sourcing</b> – podnikový proces, jehož cílem je rozhodnutí o tom, které služby, procesy a zdroje má podnik zajišťovat sám a které přenechat externím poskytovatelům, výběr nejvhodnějších poskytovatelů externích služeb, sepsání smluv s poskytovateli o obsahu a úrovni poskytovaných služeb a kontrola poskytovaných služeb a řízení vztahů s externími poskytovateli.
Star schéma	<b>Star schéma</b> – schéma hvězdy je typ dimenzionálního databázového modelu, tvořeného pouze nenormalizovanými tabulkami dimenzí.
Text Mining	<b>Text Mining</b> , resp. textová analytika představuje analýzu textových zdrojů a získávání z nich nových informací, kde zdroje mohou být velmi různorodé, od knižních nebo novinových publikací přes blogy a další.
Transakční zpracování	<b>Transakční zpracování</b> – metoda zpracování dat, při které se každý požadavek zpracuje okamžitě po svém příchodu. Uživatel může v průběhu zpracování svůj požadavek upřesňovat.
Ukazatel	<b>Ukazatel</b> – sledovaný údaj (Fact, Measure) v organizaci, například objem prodeje, počet reklamací, obrat atd.
Úloha	<b>Úloha</b> – je základní logickou jednotkou řízení podnikové informatiky. Úloha je dokumentována řadou atributů, tj. cílem, obsahem, vstupy, výstupy, procesy řízení, použitými metodami, metrikami, aplikacemi a faktory úspěšnosti.
Umělá inteligence	<b>Umělá inteligence</b> – schopnost stroje nebo systému zvýšit nebo rozšířit jakýkoli proces nebo výstup, které by normálně vyžadovaly lidské znalosti a inteligenci.

XML	<b>XML</b> – eXtensible Markup Language – značkovací jazyk, jehož aplikací na textové soubory vznikají jednotlivé XML dokumenty. Vlastní specifikace jazyka uvádí způsob zápisu struktury dokumentu, mechanismus vytváření logických struktur v dokumentu, pravidla deklarace elementů a vlastností apod.
XQuery	<b>Xquery</b> – dotazovací nástroje pro získávání dat z XML dokumentů na základě definovaných podmínek a pravidel.
XSD	<b>XSD</b> – XML Schema Definition Language – jazyk pro definování XML schémat. Mechanismus popisu a uložení struktury XML dokumentu. Využívá se u složitých datových struktur (DTD).
XSLT	<b>XSLT</b> – Extensible Stylesheet Language Transformation – prostředek, který transformuje XML dokumenty do dokumentů XML s jinou datovou strukturou (XSD nebo DTD) nebo do jiných datových formátů (HTML, PDF, RTF, databázových formátů atd.).