



MANAGEMENT V INFORMAČNÍ SPOLEČNOSTI



Business Intelligence

**Jak využít
bohatství
ve vašich
datech**

Ota Novotný, Jan Pour, David Slánský



Česká společnost
pro systémovou integraci

Upozornění pro čtenáře a uživatele této knihy

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

Používání elektronické verze knihy je umožněno jen osobě, která ji legálně nabyla a jen pro její osobní a vnitřní potřeby v rozsahu stanoveném autorským zákonem. Elektronická kniha je datový soubor, který lze užívat pouze v takové formě, v jaké jej lze stáhnout s portálu. Jakékoliv neoprávněné užití elektronické knihy nebo její části, spočívající např. v kopírování, úpravách, prodeji, pronajímání, půjčování, sdělování veřejnosti nebo jakémkoliv druhu obchodování nebo neobchodního šíření je zakázáno! Zejména je zakázána jakákoliv konverze datového souboru nebo extrakce části nebo celého textu, umísťování textu na servery, ze kterých je možno tento soubor dále stahovat, přitom není rozhodující, kdo takovéto sdílení umožnil. Je zakázáno sdělování údajů o uživatelském účtu jiným osobám, zasahování do technických prostředků, které chrání elektronickou knihu, případně omezují rozsah jejího užití. Uživatel také není oprávněn jakkoliv testovat, zkoušet či obcházet technické zabezpečení elektronické knihy.



Copyright © Grada Publishing, a.s.



Copyright © Grada Publishing, a.s.

Edice Management v informační společnosti

Ediční rada:

doc. Ing. Josef Basl, CSc. – ZČU v Plzni, VŠE v Praze – předseda

PhDr. Jiří Adamík – Grada Publishing, a.s. – místopředseda

prof. Ing. Jan Ehleman, CSc. – Technická univerzita Liberec

doc. Ing. Karol Matiaško, CSc. – Žilinská univerzita v Žiline

doc. RNDr. Jaroslava Mikulecká, CSc. – Univerzita Hradec Králové

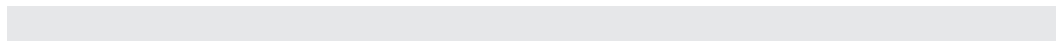
prof. RNDr. Jaroslav Pokorný, CSc. – MFF UK v Praze

doc. Ing. Jan Pour, CSc. – VŠE v Praze

doc. Ing. Karel Richta, CSc. – FEL ČVUT v Praze

doc. Ing. Milena Tvrdíková, CSc. – VŠB-TU Ostrava

prof. Ing. Ivan Vrana, DrSc. – Česká zemědělská univerzita v Praze



Ing. Ota Novotný, Ph.D.

doc. Ing. Jan Pour, CSc.

Ing. David Slánský

Business Intelligence

© Grada Publishing, a.s., 2005

Cover Design © Grada Publishing, a.s., 2005

Vydala Grada Publishing, a.s., U Průhonu 22, Praha 7

tel.: +420 220 386 401, fax: +420 220 386 400, www.grada.cz

jako svou 2113. publikaci

Odpovědný redaktor Mgr. Petr Mušálek

Sazba Milan Vokál

Počet stran 256

První vydání, Praha 2005

Vytiskla tiskárna PBtisk

Prokopská 8, Příbram VI

Nakladatelství děkuje společnosti ADASTRA, s.r.o., za podporu při vydání této knihy.

ISBN 80-247-1094-3 (tištěná verze)

ISBN 978-80-247-6685-0 (elektronická verze ve formátu PDF)

© Grada Publishing, a.s. 2011

Obsah

O autorech	11
Úvod	13
Část I: Principy Business Intelligence	15
1. Business Intelligence – vývoj a základní principy	17
1.1 Vývoj Business Intelligence	17
1.2 Business Intelligence v řízení firmy	17
1.3 Základní principy řešení Business Intelligence	20
1.3.1 Základní principy multidimenzionálních databází	20
1.3.2 Porovnání analytických a OLTP systémů	25
1.4 Hlavní komponenty, typy aplikací a vrstvy v Business Intelligence	26
1.4.1 Produkční (zdrojové) systémy	28
1.4.2 Extraction, Transformation and Loading – ETL	29
1.4.3 Enterprise Application Integration – EAI	29
1.4.4 Dočasné úložiště dat – DSA (Data Staging Areas)	30
1.4.5 Operativní úložiště dat – ODS (Operational Data Store)	30
1.4.6 Datový sklad – DWH (Data Warehouse)	32
1.4.7 Datové tržiště – DMA (Data Mart)	33
1.4.8 OLAP databáze	33
1.4.9 Reporting	34
1.4.10 Manažerské aplikace – EIS (Executive Information Systems)	34
1.4.11 Dolování dat (Data Mining)	35
1.4.12 Oborová znalost know-how	36
1.4.13 Technická znalost	36
1.4.14 Nástroje pro zajištění datové kvality	36
1.4.15 Nástroje pro správu metadat	36
1.4.16 Další komponenty související s BI	37
1.5 Rozdíly v pojetí datových skladů a tržišť	38
1.6 Integrace komponent Business Intelligence	38
1.7 Business Intelligence – příklady základních principů	39
2. Řízení projektů a aplikací Business Intelligence	45
2.1 Přístupy k řešení Business Intelligence	45
2.1.1 Postupné budování datových tržišť	45
2.1.2 Jednorázové vybudování celkového řešení	48
2.1.3 Přírůstkový přístup	49
2.2 Principy a úlohy řízení Business Intelligence	51
2.3 Business Intelligence v kontextu řízení informačního systému	54



2.3.1	Strategie rozvoje BI	55
2.3.2	Plánování a koordinace projektů BI	55
2.3.3	Řízení služeb BI	57
2.3.4	Řízení zdrojů BI	58
2.3.5	Řízení projektů BI	59
2.3.6	Řízení provozu aplikací BI	59
2.4	Řízení a postupy projektů Business Intelligence	60
2.4.1	Úvodní studie BI	61
2.4.2	Specifikace přírůstku	64
2.4.3	Analýza BI řešení	66
2.4.4	Modelování a návrh datového skladu a BI řešení	68
2.4.5	Návrh technologické platformy přírůstku	71
2.4.6	Návrh transformací – ETL	73
2.4.7	Implementace BI řešení	75
2.5	Podstatné parametry pro řízení Business Intelligence	78
2.5.1	Strategické a plánovací parametry	80
2.5.2	Parametry komplexních řešení, poskytovaných služeb a jejich dodavatelů	83
2.5.3	Parametry zdrojů a komponent řešení BI	86
2.5.4	Parametry řízení projektů, přesněji řečeno přírůstků BI	90
2.5.5	Provozní parametry BI	91
Část II: Řešení a realizace úloh Business Intelligence		95
3.	Příprava, plánování, analýzy	97
3.1	Strategie rozvoje Business Intelligence	97
3.1.1	Analýzy uživatelských (business) požadavků	97
3.1.2	Formulace strategických záměrů v BI	101
3.1.3	Analýza připravenosti firmy na BI a analýza stavu IS/ICT	101
3.1.4	Návrh architektur BI	102
3.2	Plánování projektu Business Intelligence	103
3.2.1	Formulace plánu vývoje BI	103
3.2.2	Výběr řešení a jeho dodavatele	104
4.	Návrh řešení Business Intelligence	107
4.1	Detailní specifikace uživatelských požadavků	107
4.2	Modelování Business Intelligence řešení	108
4.2.1	Dimenzionální modelování	108
4.2.2	Modelování datového skladu (logický model)	113
4.2.3	Změny v dimenzích – SCD (Slowly Changing Dimensions)	121
4.3	Návrh technologické platformy	126
4.3.1	Návrh technologické architektury	127
4.3.2	Fyzický model datového skladu	128
4.3.3	Řešení velikosti databáze datového skladu a nárůstu dat	130

4.4	Řešení transformací a kvality dat	132
4.4.1	Analýzy a řízení kvality dat	132
4.4.2	Návrh transformací dat – datových pump	142
4.5	Návrh klientských aplikací	144
5.	Provoz aplikací Business Intelligence	147
5.1	Řízení provozu Business Intelligence	147
5.1.1	Provoz aplikací BI ve srovnání s provozními systémy	147
5.1.2	Systém řízení provozu BI	148
5.1.3	Organizace provozu BI	150
5.1.4	Role v BI	152
5.2	Úlohy v provozu Business Intelligence	154
5.2.1	Řízení technické infrastruktury	154
5.2.2	Ladění výkonu databáze	154
5.2.3	Udržování dat a metadat	154
5.2.4	Řízení růstu objemu datového skladu	154
5.3	Ekonomika Business Intelligence	155
5.3.1	Náklady na BI	155
5.3.2	Efekty řešení BI	156
6.	Implementace úloh Business Intelligence	159
6.1	Základní východiska implementace úloh Business Intelligence	159
6.2	Příprava produkčních systémů a dalších zdrojů dat pro Business Intelligence	160
6.3	Implementace DSA	160
6.4	Implementace ETL mezi produkčním systémem a DSA	160
6.5	Implementace aplikací datových skladů	161
6.6	Implementace datových pump	164
6.6.1	Zdroje dat	164
6.6.2	Transformační úlohy	165
6.6.3	Řízení průběhu transformace	167
6.6.4	Základní doporučení pro implementaci datových pump	169
6.7	Naplnění datového skladu daty	169
6.8	Implementace OLAP databází	171
6.8.1	Vytvoření OLAP databáze a definice umístění (způsobu připojení) datových zdrojů	171
6.8.2	Vytvoření dimenzí	172
6.8.3	Vytvoření definic OLAP kostek	180
6.8.4	Nastavení přístupových práv	185
6.9	Implementace klientských aplikací	185
6.9.1	Připojení k OLAP databázi	186
6.9.2	Vytvoření kontingenční tabulky	187
6.9.3	Vytvoření kontingenčních grafů	187
6.9.4	Vytvoření dalších speciálních výstupů	190
6.9.5	Pracovní panely uživatelů	190

Část III: Aplikace a užití Business Intelligence	193
7. Aplikační oblasti Business Intelligence	195
7.1 Finance	195
7.2 Marketing	196
7.3 Výroba	197
7.4 Logistika	197
7.5 Řízení vztahů s dodavateli	198
7.6 Lidské zdroje	198
7.7 Informatika	199
7.8 Corporate Performance Management (CPM)	199
7.9 Web Analytics	200
7.10 Customer Intelligence	201
8. Dolování dat	205
8.1 Úlohy dolování dat	205
8.2 Techniky dolování dat	206
8.3 Proces dolování dat	207
8.3.1 Definice problému	207
8.3.2 Výběr dat	207
8.3.3 Příprava dat	207
8.3.4 Data Mining	208
8.3.5 Zprovoznění modelu (Deployment)	208
8.3.6 Obchodní akce	208
8.4 Technologie dolování dat	208
8.5 Aplikace dolování dat	209
8.5.1 Segmentace	209
8.5.2 Automatizovaný skóring zákazníků	213
9. Trendy na trhu Business Intelligence	217
9.1 Trh Business Intelligence	217
9.1.1 Růst	220
9.1.2 Konvergence	220
9.1.3 Posun k dodavatelskému způsobu vývoje řešení BI	221
9.1.4 Vývoj standardů	221
9.1.5 Zaměření na data o zákaznících – Customer Intelligence	221
9.2 Zákazníci a řešení Business Intelligence	221
9.3 Architektury Business Intelligence	224
9.3.1 Vývoj architektur řešení BI	224
9.3.2 Vývoj komponent BI	225
10. Faktory úspěšnosti řešení Business Intelligence	231
10.1 Faktory úspěšnosti ve strategii a plánování Business Intelligence	232
10.1.1 Úroveň řízení firmy, firemní kultura	233
10.1.2 Existence potřeby BI aplikací	233

10.1.3	Kvalita vize a strategie BI	233
10.1.4	Identifikace efektů řešení	234
10.1.5	Definování realizovatelných cílů BI	234
10.1.6	Stav zdrojových informačních systémů a produkčních dat	234
10.1.7	Existence silného sponzora	235
10.1.8	Motivace uživatelů na kooperaci v BI projektech	235
10.1.9	Kvalita a akceptování uživatelských požadavků	235
10.1.10	Volba přístupu k řešení BI	236
10.1.11	Jasná definice rozsahu projektů	236
10.2	Faktory kvality řešení a služeb Business Intelligence	237
10.2.1	Rozsah outsourcingu a kvalita dodavatele	237
10.2.2	Řízení rizik a kvality služeb BI	237
10.2.3	Kvalita smluv SLA	238
10.2.4	Kvalita celkové architektury BI	238
10.3	Faktory úspěšnosti disponibilních zdrojů	239
10.3.1	Objem a doba pro získání financí	239
10.3.2	Alokace dostatečných pracovních kapacit	239
10.3.3	Kvalita technologické platformy BI	239
10.4	Faktory úspěšnosti řízení a řešení projektů Business Intelligence	240
10.4.1	Úroveň řízení projektu	241
10.4.2	Doba trvání projektu	241
10.4.3	Úroveň sdílení informací	241
10.4.4	Kvalita modelování datových skladů a tržišť	241
10.5	Faktory úspěšnosti spojené s provozem aplikací Business Intelligence	242
10.5.1	Celková úroveň provozu BI aplikací	242
10.5.2	Výkonnost BI řešení	242
10.5.3	Způsob aktualizace datových skladů, řízení nárůstu dat	243
10.5.4	Náročnost řešení na administraci	244
10.5.5	Podpora uživatelů	244
10.5.6	Kvalita změnových řízení	244
10.5.7	Úroveň řízení kvality dat	245
Zdroje	247
Literatura	247
Webové adresy související s tematikou	249
Terminologický slovník	250
Rejstřík	251

Color profile: Generic CMYK printer profile
Composite Default screen

O autorech

Ing. Ota Novotný, Ph.D.

Vystudoval Vysokou školu ekonomickou v Praze, kde získal i titul Ph.D. v oboru informatika. Od roku 1996 zde na katedře informačních technologií přednáší problematiku business intelligence, ERP produktů, měření výkonnosti informačních systémů a jejich auditu. V současné době se zaměřuje na aplikaci principů business intelligence v řízení podnikové informatiky a při strategickém řízení podniku. Dále zastupuje Českou republiku v mezinárodním výboru ISO/IEC JTC1 SC7 – Softwarové inženýrství, kde připravuje a připomínkuje mezinárodní normy ISO/IEC z oblasti měření jakosti informačních systémů a softwarových produktů. Ota Novotný se účastní řady projektů spojených s návrhem a řízením implementace informační podpory firemních procesů a následného auditu. Je spoluautorem dvou knih, autorem či spoluautorem několika skript a řady článků či příspěvků na konferencích.



Doc. Ing. Jan Pour, CSc.

Vystudoval Vysokou školu ekonomickou v Praze v roce 1970. Je pracovníkem katedry informačních technologií VŠE v Praze. Orientuje se na otázky řízení podnikové informatiky, systémovou integraci a řešení úloh business intelligence a datových skladů. Je autorem nebo spoluautorem čtyř knižních publikací, 25 vysokoškolských skript a řady článků a příspěvků na konferencích. V současné době je členem Rady České společnosti pro systémovou integraci a šéfredaktorem časopisu Systémová integrace. Je členem programových výborů mezinárodních konferencí, působí rovněž jako analytik v konzultační a analytické společnosti ITG se sídlem v Praze. Jan Pour se podílí na řešení projektů v oblasti řízení podnikové informatiky a v oblasti návrhů informačních systémů v průmyslu, obchodu, energetice i v organizacích státní správy.



Ing. David Slánský

Vystudoval katedru informačních technologií Vysoké školy ekonomické v Praze, působil rovněž v jaderné laboratoři CERN ve Švýcarsku. V současné době pracuje ve společnosti Adastra, kde zastává pozici Business Analyst Manager. David Slánský se intenzivně věnuje oblasti business intelligence a následnému využití v zákaznický orientovaných procesech v podnicích, zejména v telekomunikačních společnostech a finančních institucích. Kromě business intelligence se specializuje na analýzu, design a implementaci IS/ICT architektur a na související problematiku měření přínosů a nákladů IS/ICT řešení.



Color profile: Generic CMYK printer profile
Composite Default screen

Úvod

Business Intelligence (BI) je termín, označující celý komplex činností, úloh a technologií, které dnes stále častěji tvoří běžnou součást řízení podniků a jejich informačních systémů. Business Intelligence není právě ryzí český termín, ale v odborné praxi se již tak vžil a současně je natolik obtížné pro něj najít nějaký přijatelný překlad (podniková inteligence, obchodní inteligence, inteligentní obchod, ...), že jsme u něj raději zůstali, za což se všem čtenářům omlouváme.

Na úvod této publikace si musíme zodpovědět tři obvyklé otázky:

- Co je jejím účelem?
- Komu je určena?
- Co je jejím obsahem a na co je orientována?

Účel: Text je koncipován jako základ pro práci a další studium v oblasti Business Intelligence (BI). Jeho účelem je podat v poměrně komprimované formě celkový přehled o principech, na nichž jsou úlohy a aplikace BI založeny, o přístupech k jejich řešení a rovněž o možnostech jejich využití v řízení podniku nebo instituce. Záměrem autorů proto není podat detailní výklad jednotlivých produktů nebo technologií, které se v dané oblasti používají. Na druhé straně je snahou dokumentovat prezentované principy řešení na konkrétních příkladech. Pro tento účel jsou zde převážně využívány prostředky spojené s databázovým prostředím Microsoft SQL Server 2000, a v jeho rámci integrovanými produkty BI, zejména Analysis Services a Data Transformation Services (DTS).

Vzhledem k tomu, že jde o oblast informatiky, která se dynamicky rozvíjí, jsou v této textové formě obsaženy především obecně platné charakteristiky a metody spojené s řešením BI, zatímco aktuální informace o produktech, novinky na trhu BI a další doplňující informace najde čtenář na [www stránkách](#) různých informačních serverů, např. České společnosti pro systémovou integraci, Datawarehouse klubu, Centra pro BI a dalších uvedených v závěru knihy.

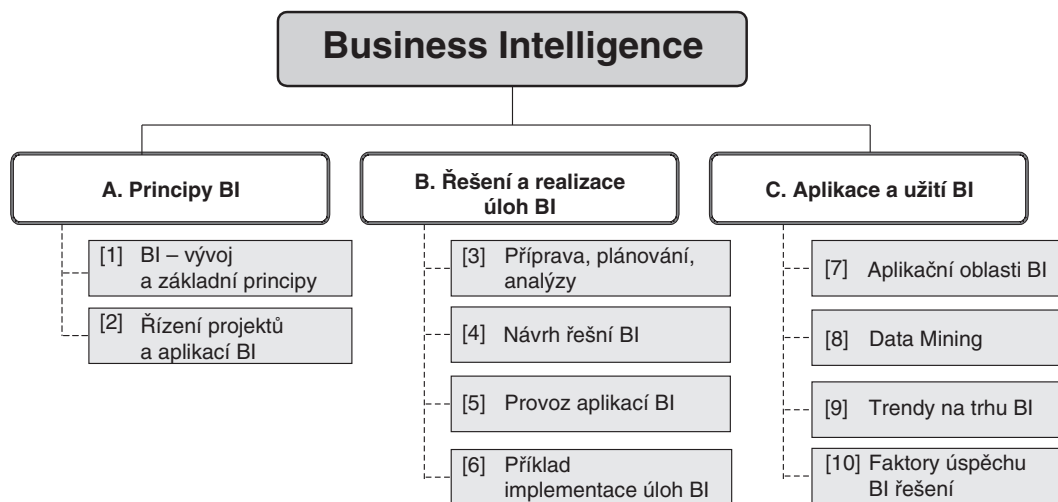
Určení: Publikace je primárně určena **pracovníkům v oblasti informatiky**, kteří až dosud pracovali v oblasti transakčních aplikací a chtějí rozšířit svoji specializaci i pro oblast analytických a rozhodovacích úloh v rámci řešení podnikových informačních systémů, dále pracovníkům působícím v **uživatelské sféře informatiky**, kteří se budou podílet na řešení uvedeného typu aplikací jako konzultanti nebo zadavatelé těchto úloh, a v neposlední řadě i **studentům** oborů informatiky jako studijní literatura pro kursy BI, resp. aplikačně orientované kursy informačních systémů.

Obsah: Obsah celé publikace se dělí do tří částí, které na sebe velmi úzce navazují, a to:

- **Část I – Principy Business Intelligence**, podává celkový přehled o obsahové i technologické podstatě Business Intelligence, jeho hlavních komponentách a úlohách spojených s jeho řešením a řízením.
- **Část II – Řešení a realizace úloh Business Intelligence**, se detailněji zabývá jednotlivými otázkami a úlohami BI, počínaje určením jejího strategického zaměření, plánováním projektů, přes analýzu, modelování, návrh a implementaci až po zajištění organizace, řízení provozu BI aplikací a hodnocení jejich efektů vzhledem k předpokládaným nebo vynaloženým nákladům. Každá z dílčích úloh v této části zahrnuje vymezení jejich účelu (proč se realizují), jejich obsahové podstaty (co je jejich náplní) a faktorů jejich úspěšné realizace (co je při jejich řešení dobré respektovat a které dosavadní zkušenosti zohlednit).

- **Část III – Aplikace Business Intelligence**, nabízí základní představu o možnostech užití technologií a aplikací BI při řízení firmy nebo instituce, a to od dnes již vcelku obvyklých řešení v oblasti finančních, obchodních nebo marketingových analýz až po stále atraktivnější typy aplikací, jako je tzv. *Customer Intelligence*, aplikace BI ve strategickém řízení firmy apod. Část III má tak odpovědět na otázky, kde lze BI ve firmě využít, jaké jsou zde jeho možnosti a případná omezení.

Celkovou strukturu textu dokumentuje úvodní obrázek.



Struktura textu

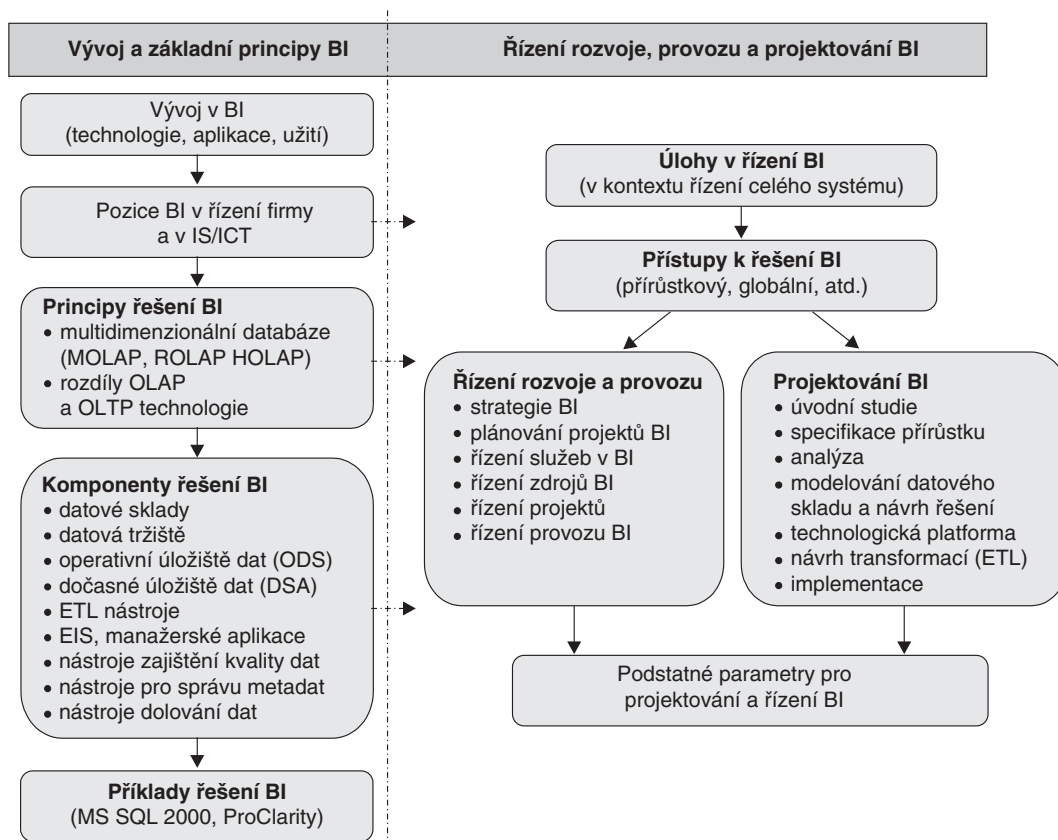
S obsahem souvisí i forma textu a proto na tomto místě upozorníme na použité piktogramy uvozující standardní části jednotlivých kapitol:

	Orientace a účel kapitoly nebo části textu.
	Vymezení nebo definice pojmu.
	Příklady nebo dílčí kroky při realizaci jednotlivých funkcí.
	Doporučení pro řešení úlohy, problému nebo operace.
	Shrnutí kapitoly nebo vybrané části textu.

Na závěr by autoři chtěli touto cestou poděkovat spolupracovníkům i absolventům Katedry informačních technologií Vysoké školy ekonomické v Praze, kteří se podíleli na přípravě podkladů. Jejich práce jsou uvedeny v celkovém přehledu literatury. Zvláštní dík patří zaměstnancům společnosti Adastra (zejména Ing. Ondřeji Šňupárkovi, Ing. Jaroslavu Vlčkoví, Ing. Tomáši Kočkovi, Ph.D., a Ing. Martinu Šálymu), kteří poskytli pro tuto publikaci cenné materiály, připomínky a náměty.

ČÁST I: PRINCIPY BUSINESS INTELLIGENCE

Cílem celé první části je objasnit podstatu Business Intelligence a způsob řízení a řešení jeho provozu a rozvoje. Celkový pohled na témata prezentovaná v této části poskytuje následující obrázek.



Struktura témat prezentovaných v části I

Color profile: Generic CMYK printer profile
Composite Default screen

1. Business Intelligence – vývoj a základní principy



Business Intelligence (BI) představuje **komplex přístupů a aplikací IS/ICT**, které téměř výlučně podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principu multidimenzionality, kterým zde rozumíme možnost pohlížení na realitu z několika možných úhlů. **Cílem** této kapitoly je vysvětlit dosavadní vývoj a základní principy těchto technologií a aplikací.

1.1 Vývoj Business Intelligence

Řešení směřující k podpoře manažerských a analytických úloh v podnikovém řízení se začala objevovat již na konci sedmdesátých let minulého století v souvislosti s rozvojem on-line zpracování dat. První pokusy a aplikace jsou spojeny s americkou firmou Lockheed. V polovině osmdesátých let byly publikovány první významné práce k tomuto typu aplikací (prof. Rockart: „CEO Goes On-line“ a některé další). V druhé polovině osmdesátých let přišly na trh v USA první firmy s komerčními produkty, založenými na multidimenzionálním uložení a zpracování dat, označovanými jako EIS (Executive Information System), a to firmy Comshare a Pilot. Trh s EIS produkty se pak velmi rychle rozvíjel a na začátku devadesátých let (od roku 1993) se tyto produkty začaly prosazovat i na českém IS/ICT trhu.

Koncem osmdesátých a začátkem devadesátých let se v USA začal velmi silně prosazovat i další trend v multidimenzionálních technologiích, a to datové sklady (Data Warehouse) a datová tržiště (Data Marts). Za rozvojem těchto technologií stáli především Ralph Kimball a Bill Inmon. Větší uplatnění datových skladů a tržišť je na českém trhu patrné spíše až v druhé polovině devadesátých let. V souvislosti s datovými sklady a narůstajícím objemem dat v tomto prostředí se v průběhu devadesátých let začaly prosazovat i technologie a nástroje tzv. dolování dat (Data Mining)¹ založené na vysoce sofistikovaných analýzách dat s pomocí nejrůznějších matematických a statistických metod.

K jednotlivým technologiím, nástrojům a aplikacím zařazovaným do Business Intelligence se budeme postupně vracet v dalších částech tohoto textu.

1.2 Business Intelligence v řízení firmy

V prostředí stále tvrdší konkurence musí podnikoví analytici a manažeři rozhodovat pod časovým tlakem a současně s vysokou zodpovědností. To znamená, že pro tato rozhodnutí musí mít dostatek relevantních a objektivních informací, které jsou dostupné rychle, s minimální technickou náročností na manipulaci, a přitom s možností rychle formulovat nové požadavky na další informace odpovídající aktuální obchodní nebo výrobní situaci.

Zpracování a uložení dat v transakčních systémech, především v **aplikacích ERP**, je založeno vesměs na využití **relačních databázových systémů**. Toto řešení je v mnohém ohledu velmi výhodné. Data jsou zde přehledně uspořádána, a v případě efektivně navržené datové základny umožňují rychlé provádění jednotlivých

¹ V této publikaci jsou oba termíny používány ve stejném významu

transakcí a poskytují odpovídající dobu odezvy na zadané dotazy. Navíc zajišťují integritu dat, bezpečnost přístupu k datům a další potřebné charakteristiky spojené s řízením firmy na taktické nebo operační úrovni. ERP aplikace však mají z hlediska analytických a plánovacích činností podniku některá **omezení**:

- Neumožňují rychle a **pružně měnit kritéria** pro analýzy podnikových dat (např. sledovat data o prodeji – v čase, podle zákazníků, produktů, segmentů trhu, obchodních zástupců, podnikových útvarů, a dále i v nej-různějších kombinacích uvedených kritérií).
- Stejně tak se v obrovských objemech dat současných databází obtížně řeší zajištění okamžitého **přístupu** pracovníků k **agregovaným datům**, a to na nej-různějších úrovních agregace (za podnik, útvar, za všechny zákazníky, skupiny zákazníků, jednotlivé zákazníky atd.).
- ERP a ostatní transakční aplikace jsou **primárně určeny pro pořizování a aktualizace dat**, přičemž některé z nich pracují neustále téměř na 100 % svého možného výkonu; analytické úlohy tyto systémy nadměrně zatěžují a v mnohých případech nejsou ani díky jejich vytížení možné.
- Dalším problémem je narůstající objem dat v podniku, který se v průměru zdvojnásobí každých pět let. Většina firem tak nemá problém s nedostatkem dat, ale naopak firmy jsou jimi **zahlceny**, a to často **redundantními a nekonzistentními daty**, která jsou v rozhodovacích procesech obtížně využitelná.

Pokud o těchto omezeních mluvíme v souvislosti se systémy ERP, pak to neznamená, že by aplikace ERP a další transakční úlohy nebyly schopny zmíněné operace realizovat. Jde však o jejich rychlost a pružnost vzhledem k uživatelským požadavkům. Řešení uvedených problémů se tak postupně stalo doménou speciálních technologií a aplikací Business Intelligence.

Termín Business Intelligence zavedl v **roce 1989 Howard J. Dresner**, analytik společnosti Gartner Group, který jej popsal jako „**sadu konceptů a metod určených pro zkvalitnění rozhodnutí firmy**“. Vyzdvihuje zde význam datové analýzy, reportingu a dotazovacích nástrojů, které provádějí uživatelé množstvím dat a pomáhají mu se syntézou hodnotných a užitečných informací.

Možná právě díky krátké době existence tohoto termínu není pro něj v dnešní době zavedena jednotná definice, podporovaná jakoukoliv organizací, zabývající se standardy (např. ANSI). Například definice ze serveru **searchCRM.com** definuje Business Intelligence jako určitou kategorii aplikací a technologií pro sběr, skladování, analyzování a zpřístupňování dat, jejichž účelem je pomoci podnikovým uživatelům dělat lepší rozhodnutí. BI aplikace podle searchCRM.com zahrnují funkčnost systémů pro podporu rozhodování, dotazování a reportingu, statistických analýz, vytváření prognóz a Data Mining.

Server firmy iOLAP charakterizuje BI jako sběr a analýzu dat, jejímž cílem je lepší porozumění a reakce na změny, kterým organizace čelí². Definici založenou na znalostech a informacích nabízí server DM Review³, který definuje BI jako znalosti o podniku získané za pomoci rozličných hardwarových a softwarových technologií, které umožňují organizaci přeměnit data na informace. Tento popis považuje technologie pouze za prostředek, nikoli za podstatu Business Intelligence.

Pro další účely tohoto textu využijeme definici na serveru České společnosti pro systémovou integraci (www.cssi.cz):

² iOLAP – http://www.iolap.com/bi_getting_started.htm

³ DM Review – <http://www.dmreview.com>



Business Intelligence je sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data.

Aplikace BI pokrývají **analytické a plánovací funkce většiny oblastí podnikového řízení**, tj. prodeje, nákupu, marketingu, finančního řízení, controllingu, majetku, řízení lidských zdrojů, výroby, IS/ICT apod.

Do nástrojů a aplikací Business Intelligence se zahrnují:

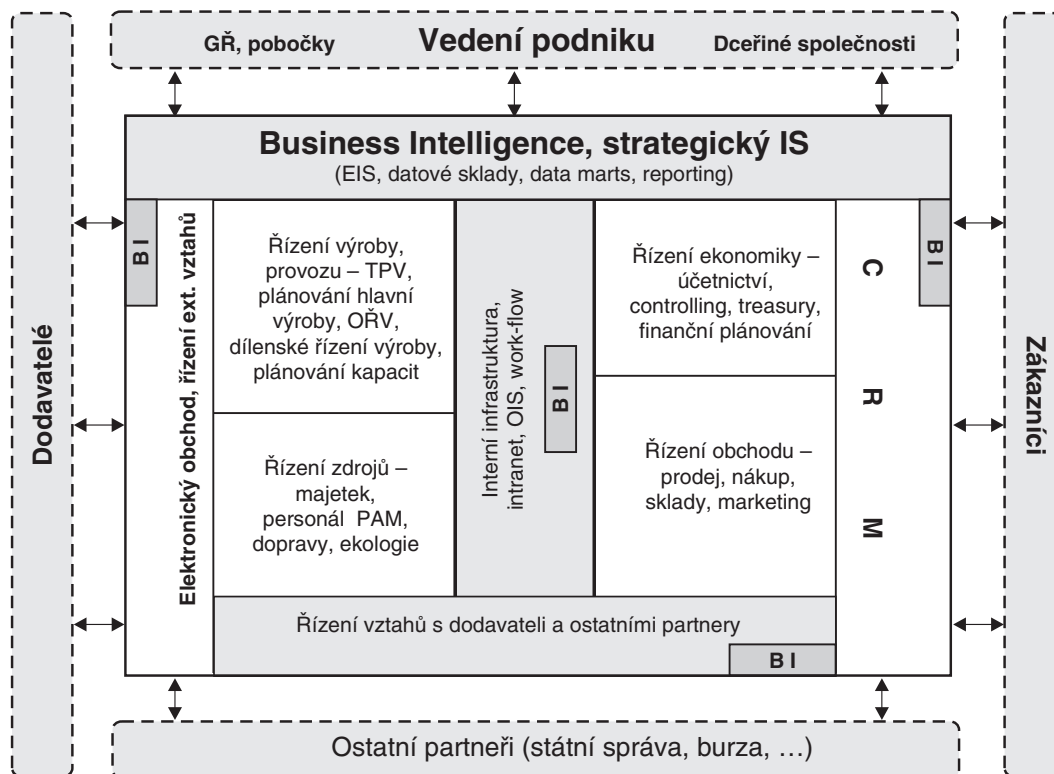
- produkční, zdrojové systémy,
- dočasná úložiště dat (DSA – Data Staging Area),
- operativní úložiště dat (ODS – Operational Data Store),
- transformační nástroje (ETL – Extraction Transformation Loading),
- integrační nástroje (EAI – Enterprise Application Integration),
- datové sklady (DWH – Data Warehouses),
- datová tržiště (DMA – Data Marts),
- OLAP,
- reporting,
- manažerské aplikace (EIS – Executive Information Systems),
- dolování dat (Data Mining),
- nástroje pro zajištění kvality dat,
- nástroje pro správu metadat,
- ostatní.

K uvedenému přehledu nástrojů, které zahrnujeme do BI, je nutné poznamenat, že i na tuto otázku existují různé názory. Vedle toho, který chápe BI jako široký rámec od manažerských aplikací po reporting, existuje i jiný pohled, který chápe BI pouze jako jeden z nástrojů vedle, resp. nad datovými sklady nebo datovými tržišti. V našem případě se budeme nadále držet první z uvedených variant.

Ze všech předcházejících vymezení a definic však vyplývá, že BI je orientován na vlastní využití informací v řízení a rozhodování, a nikoli na základní zpracování dat a realizaci běžných obchodních, finančních a dalších transakcí. To, jak jsou možnosti BI využity, dnes do značné míry ovlivňuje výkonnost a kvalitu řízení firmy, a v souvislosti s tím nakonec i její celkovou úspěšnost a konkurenceschopnost.

Business Intelligence představuje již široký komplex nástrojů a aplikací, jejichž **postavení v architektuře IS/ICT** dokumentuje obrázek 1.1.

Z obrázku vyplývá, že BI je úzce provázána s ostatními aplikacemi IS/ICT, čerpá z nich vstupní data a v poslední době i data často do ostatních aplikací vrací. To vede k dílčímu pracovnímu závěru, že kvalita řešení BI je úzce závislá na kvalitě ostatních (transakčních) aplikací, zejména na kvalitě jejich – tzv. produkčních dat, resp. databází.



Obrázek 1.1 Postavení BI v aplikační architektuře IS/ICT

1.3 Základní principy řešení Business Intelligence

V předchozí části kapitoly jsme vymezili hlavní komponenty BI řešení (manažerské aplikace, datové sklady atd.). I když jsou mezi nimi podstatné rozdíly, přesto celá řada základních principů je společná. Na tomto místě se proto zaměříme na:

- základní principy multidimenzionálních databází,
- hlavní rozdíly mezi analytickými a transakčními systémy.

1.3.1 Základní principy multidimenzionálních databází

Z předchozího textu vyplývá, že informační systémy mohou pracovat se dvěma základními typy informací – **operativními a analytickými**. První typ, **operativní informace**, slouží pro realizaci obchodních a dalších transakcí v podniku. Jsou uloženy většinou v relačních databázích, zobrazují aktuální stav podniku a v průběhu jednoho dne se mohou i několikrát měnit. Příkladem může být např. účetnictví, data v dokumentech obchodních případů apod. Transakční systémy realizují jejich zpracování v reálném čase a označují se jako **OLTP** (On Line Transaction Processing) systémy. Vzhledem k analytickým aplikacím se data OLTP systémů chápou jako **pri-mární, zdrojová nebo produkční**.

Na druhé straně systémy pracující s **analytickými** informacemi využívají primární data vytvořená v OLTP systémech.

Pro své uložení a operace s daty se pro tyto systémy vžil v osmdesátých letech minulého století název **OLAP – On Line Analytical Processing**. Se zavedením pojmu BI (který ve své podstatě kopíruje výše zmíněný význam výrazu OLAP) a současně s rozvojem nástrojů a technologií pro 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.“

Tato kniha bude nadále pracovat s užším – technologickým významem výrazu OLAP. Pro širší význam budeme používat výraz **analytické** nebo **BI** systémy.

Na rozdíl od OLTP systémů jsou analytické systémy charakteristické především těmito **vlastnostmi**:

- informace poskytují na základě vstupů získaných z primárních dat (viz předchozí schéma architektury IS/ICT);
- jejich data jsou uložena multidimenzionálně, resp. v multidimenzionálních databázích;
- obsahují různé úrovně agregace dat, podle hierarchické struktury dimenzí (viz dále);
- zachycují faktor času a umožňují realizovat časová srovnání, časové řady, predikovat možný vývoj sledovaných ukazatelů apod.

Pro snadnější pochopení principů fungování analytických systémů, a jejich rozdílu oproti systémům transakčním (OLTP), budeme v následujících odstavcích vycházet z předpokladu, že analytické řešení bude pokrývat typické požadavky na reporting, tedy nejstarší, a ještě stále nejrozšířenější, funkcionalitu analytických aplikací.

1.3.1.1 Architektura multidimenzionální databáze

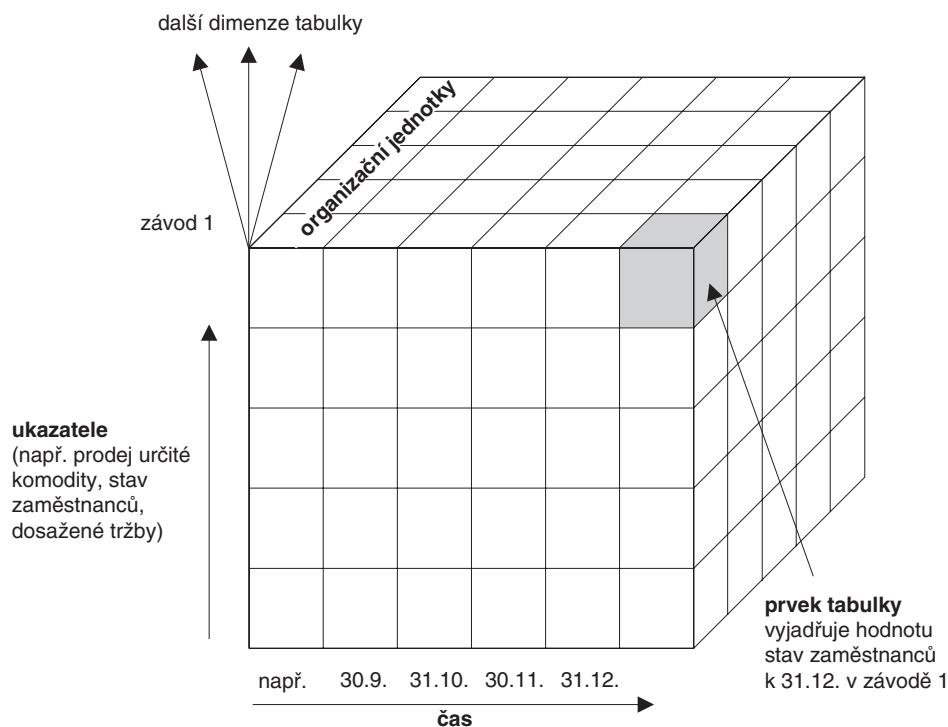
Pro data analytického typu se nehodí, aby byla ukládána v relačních databázích do podoby třetí normální formy (typické pro transakční systémy). Aby analytické systémy mohly poskytovat různé analýzy a přehledy sloužící pro strategické rozhodování, je nutné, abychom se na jejich data mohli dívat z více hledisek současně. Mělo by tedy být možné vytvářet tzv. multidimenzionální pohledy, což je pro data uložená v třetí normální formě velký problém. Nástroje koncového uživatele musí umožňovat analýzu ve smyslu nacházení souvislostí, které nejsou z primárních dat na první pohled zřejmé. Navíc je nutné procházet velká množství dat, vypočítávat agregace (které v databázích modelovaných ve třetí normální formě nejsou automaticky uloženy), rychle měnit pohledy na data, rychle, a co možná automatizovaně, je ukládat do přehledných tabulek a grafů.

Multidimenzionální databáze jsou již optimalizované pro uložení a interaktivní využívání multidimenzionálních dat. Výhodou multidimenzionality, resp. nasazení OLAP technologií (v užším chápání), je rychlost zpracování a efektivní analýzy multidimenzionálních dat (*drilling, slice and dice* apod.).

Základním **principem**, na němž jsou aplikace Business Intelligence založeny, je **několikadimenzionální tabulka** umožňující velmi rychle a pružně měnit **jednotlivé dimenze**, a nabízet tak uživateli různé pohledy na mo-

1

delovanou ekonomickou realitu. Jde tak v podstatě o princip „n-dimenzionální Rubikovy kostky“ naplněné nejdůležitějšími podnikovými daty – viz obrázek 1.2.



Obrázek 1.2 Princip multidimenzionální databáze

Z obrázku vyplývá, že **standardními dvěma dimenzemi jsou tu ukazatele (ekonomické proměnné) a čas**. Ostatní dimenze se pro jednotlivé modely definují podle potřeby, např. organizační jednotka, komodita, zákazník, dodavatel, teritorium, konkurent apod. Obsah dimenzí je tvořen **prvky dimenzí**, tj. konkrétními závody, provozy nebo zákazníky, dodavateli, komoditami apod. Promítnutí všech dimenzí do jednoho bodu tvoří **prvek multidimenzionální databáze**. Každý prvek může pak obsahovat data nebo předpisy (algoritmy) pro jejich transformace.

Je třeba však současně připomenout, že různé produkty různých výrobců používají pro výše uvedené termíny různou terminologii, v tomto případě odkazujeme na příslušnou firemní dokumentaci. Způsoby konkrétní realizace uvedených principů jsou uvedeny v částech věnovaných příkladům vybraných produktů.

Prvky dimenzí jsou většinou uspořádány v **hierarchické struktuře**, tzn. že se rozdělují na skupiny prvků, podskupiny až na jednotlivé prvky. Produkty Business Intelligence pak zajišťují automatické agregace hodnot (např. ekonomických proměnných), a to podle definovaných hierarchických úrovní dimenzí.

22

Business Intelligence

Jednoduchým příkladem takových struktur může být dimenze *organizační struktura*:

Podnik celkem
 Závod 1
 Provoz 11
 Dílna 111
 atd.
 Závod 2
 atd.

Aplikace Business Intelligence pak budou automaticky sledovat např. hodnoty výroby, prodeje zboží, zásob, zisku apod. podle uvedené organizační struktury a budou zajišťovat agregace podle takto definované hierarchické struktury. Tato metoda je využívána pro urychlení odezvy systému na analytické požadavky. Pokud by totiž bylo nutné on-line vypočítávat součty při zobrazování tabulky či grafu ze statisíců či milionů hodnot, odezva systému by mohla být neúnosně velká. Hierarchie uložení agregovaných dat pak uživateli umožňuje se pružně po požadovaných úrovních agregace pohybovat (na úrovni závodů, provozů apod.), aniž by bylo nutné vždy znovu požadované agregace počítat. Tento princip se označuje **drill-down** (pohyb – zpřístupnění dat pro nižší úroveň agregace – detailnějších dat) nebo **drill-up** (v opačném směru).

Multidimenzionalita dat může být implementována na dvou úrovních:

- na úrovni relační databáze, v tzv. „*STAR scheme*“ nebo „*SNOWFLAKE scheme*“ (viz níže),
- na úrovni speciální binární databáze.

1.3.1.2 Implementace na úrovni relační databáze

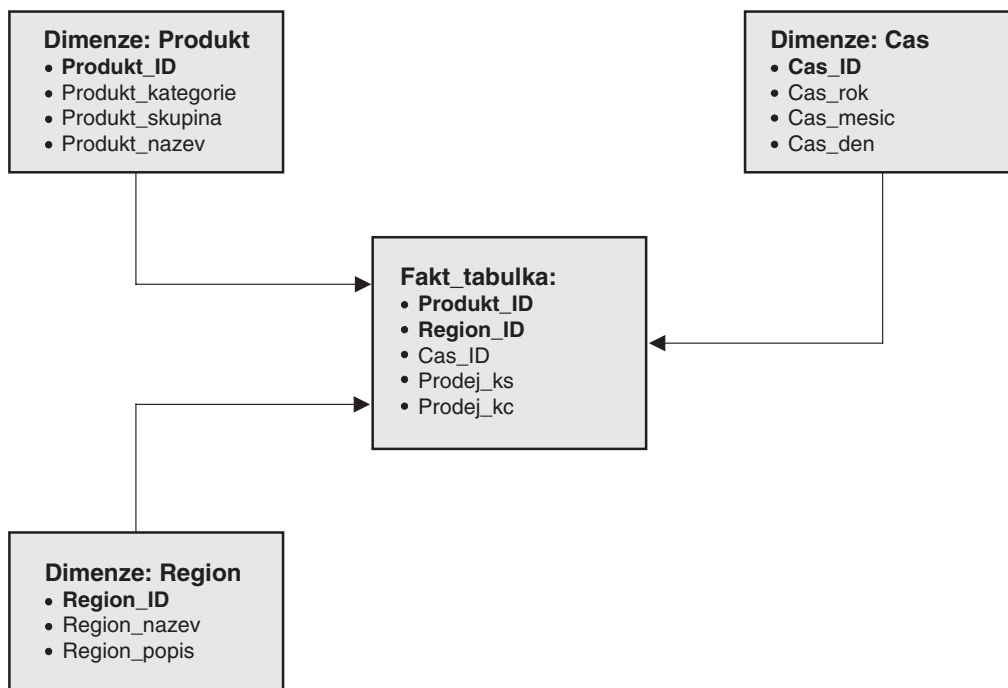
Jak již bylo řečeno, databáze produkčních – transakčních systémů bývají modelovány v tzv. třetí normální formě. Jedná se o design databáze, který nejvíce vyhovuje požadavkům na tyto systémy – tedy možnosti data rychle a jednoduše ukládat a současně optimalizovat velikost databáze.

Datové modely produkčních systémů jsou komplexní, obsahují mnoho tabulek a jejich vazby. Neexistuje jediný způsob, jak provést dotaz do databáze. Těchto způsobů je větší množství a záleží jen na uživateli, kterou vazbu mezi tabulkami zvolí. Různé dotazy tedy mohou dát stejné výsledky. Tím se ovšem pro běžného uživatele stává databáze velmi nepřehledná. Pro dotazy do více tabulek navíc musí vytvářet jakési propojovací můstky, a jak si dále ukážeme, právě tato propojení jsou největší zátěží systému.

Pro výše uvedené nedostatky se objevily snahy o zjednodušení ERD diagramu, přizpůsobení jej pro tvorbu datových skladů, a především přiblížit data koncovému uživateli. Vznikl tak relační „**dimenzionální model**“ kterému se také běžně říká „**Schéma hvězdy**“ (*STAR scheme*), resp. „**Schéma sněhové vločky**“ (*SNOWFLAKE scheme*) – viz obr. 1.3 a obr. 1.4.

V centru schématu je tzv. tabulka faktů, tedy tabulka sledovaných ekonomických a dalších ukazatelů identifikovaných klíčem složeným z klíčů tzv. dimenzionálních tabulek, v nichž jsou uloženy prvky jednotlivých dimenzí. **Dimenzionální tabulky** slouží jako úložiště textových informací o hodnotách uložených v tabulce faktů. Typicky si ji lze představit jako **číselník**. Pro reálné dimenzionální tabulky je typické velké množství atributů, pro něž se nejlépe hodí atributy textové a diskrétní. Přesto občas bývá problematické rozhodnout, které pole bude zařazeno do fakt tabulky, a které do tabulky dimenzionální. Naše rozhodnutí je většinou závislé na tom, je-li sledovaná veličina měřitelná a měnící se v čase – pak patří do fakt tabulky, či zda je diskrétní a vystupuje spíše jako konstanta – pak jde o položku z dimenzionální tabulky.

1



Obrázek 1.3 Datový model Schéma hvězdy (STAR scheme)

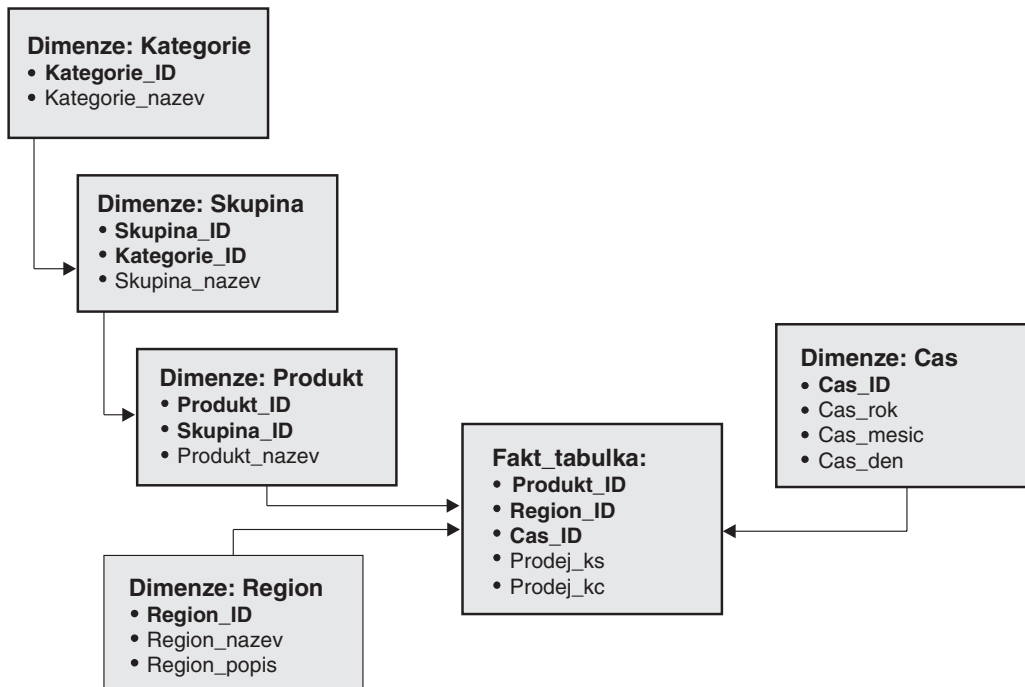
Dotazování do datového skladu téměř vždy probíhá prostřednictvím SQL příkazů. Přestože se zde nechceme věnovat oblasti dotazování pomocí SQL do datového skladu, na tomto místě je pouze demonstrován základní princip dotazování. Typický SQL dotaz, který se využívá ve STAR schématu, může vypadat např. takto:



```
Select Distinct A.Region_ID, A.Cas_ID, A.Prodej_kc
From Fakt_tabulka A
Where A.Region_ID in (Select B.Region_ID
From Region B
Where Region_nazev = „Jižní Čechy“)
```

Příkaz SELECT DISTINCT je nejčastěji používaným příkazem pro dotazování do datového skladu. Tento příkaz se provádí prakticky při každé akci, která zobrazuje nějaká data. Příkaz SELECT je sám o sobě velmi rychlý. Tabulka faktů, pro kterou je nejčastěji používán, může obsahovat mnohdy statisíce až miliony řádků, a přesto lze obdržet odpovídající doby odezvy systému. Nad datovými sklady lze pomocí SQL dotazů samozřejmě vytvářet různé analýzy a reporty, včetně požadovaných agregovaných hodnot.

V některých případech je ovšem i toto řešení z řady důvodů nevhodné. V těchto případech se proto dimenzionální tabulky upravují, resp. normalizují. Vznikne tak **SNOWFLAKE (sněhová vločka)** – viz obr. 1.4.



Obrázek 1.4 Schéma sněhové vločky (SNOWFLAKE scheme)

1.3.2 Porovnání analytických a OLTP systémů

Požadavek pohledů na data z více hledisek (dimenzí) tedy s sebou současně přináší i požadavek na **optimalizované fyzické ukládání dat**, přičemž se většinou jedná o **data historická, agregovaná, průběžně rozšiřovaná** a ukládaná v jednoduché struktuře vhodné pro analýzu a přizpůsobené potřebám managementu.

Jsou tedy zřejmé tyto **základní rozdílové charakteristiky mezi OLTP a analytickými systémy**:

- OLTP systémy jsou primárně určeny pro pořizování dat. Tomu odpovídá jejich celková architektura, a zejména pak databázový model. Ten je charakteristický tabulkami ve třetí normální formě, velkým počtem tabulek a jejich spojení, snahou o nulovou redundanci dat. Současně jsou tabulky indexovány pouze v nejn nutnějších případech.
- Data jsou do OLTP systémů pořizována v reálném čase, v běžném provozu probíhají desítky až statisíce transakcí za minutu. Systémy jsou tak zatěžovány kontinuálně.
- Analytické systémy jsou oproti OLTP určeny primárně pro podporu dotazování. Z toho vyplývá jejich architektura (několikavrstvé databáze) a databázové modely, vyznačující se zmenšeným počtem tzv. denormalizovaných tabulek, vyšší frekvencí indexů, duplicitou uložení dat apod.
- Drtivá většina analytických systémů aktualizuje svá data periodicky (nejběžněji v denních a měsíčních intervalech). Nejnovější trendy sice umožňují i aktualizaci dat v reálném čase, avšak v dnešní době se jedná o případy výjimečné. Zatížení analytických systémů je z tohoto důvodu nárazové – velké zatížení je typické