

Obsah

Úvod	11
-------------------	-----------

Kapitola 1

Business Intelligence, datové sklady	13
---	-----------

Business Intelligence jako proces přerodu dat na informace	14
Výhody aplikování BI	16
Fáze budování BI	16
Přechod od transakčních databází k analytickým	16
Kvalita dat pro analýzy	18
Nevhodnost návrhové struktury transakčních databází pro analýzy	18
Zpracování dat z operačního prostředí	20

Kapitola 2

Architektura, modelování a implementace Business Intelligence procesů v SQL Serveru 2008	21
---	-----------

Architektura Business Intelligence na platformě SQL Serveru 2008	22
Modelování procesů Business Intelligence	22
Časová postupnost sekvenčního BI	26
Business Intelligence v reálném čase	27
Nástroje pro práci s analytickými službami	29
Implementace a přístup k UDM přes Business Intelligence Development Studio	30
Pohledy na datové zdroje – Data Source Views	32
Administrace databází a dotazování přes SQL Server Management Studio	33

Kapitola 3

Datový sklad	37
---------------------------	-----------

Definice datového skladu	38
Datový sklad jako jediný zdroj datové pravdy	40
Rozdíly mezi produkčními databázemi a datovým skladem	40
Projekt datového skladu nemůže být neúspěšný	41
Návrh a koncepce projektu datového skladu	41
Datové trhy	43
Metody budování datového skladu	44

Fáze přírůstkové metody	46
Novinky SQL Serveru 2008 využitelné pro datové sklady	48
Kompresce dat	48
Řádková komprese.....	49
Stránková komprese	49
Odhad kompresního poměru pro jednotlivé metody komprimace	50
Zachytávání změn v databázových tabulkách.....	55
Zachytávání změn pomocí mechanismu Change Tracking.....	55
Zachytávání změn pomocí mechanismu Change Data Capture	60
MERGE.....	67

Kapitola 4

Integrační služby 71

Import a export	72
Extrakce, transformace a přenos	75
Oblast vynášení dat	76
Extrakce.....	77
Čištění dat.....	77
Transformace.....	78
Zavedení dat do datového skladu (Loading)	80
Chyby a problémy procesu ETL	81
Testování etapy ETL.....	81
Architektura integračních služeb SQL Serveru 2008	81
Seznámení se s návrhovým prostředím pro integrační služby.....	83
Záložka ControlFlow.....	84
Kontejnery integračních služeb	86
Úlohy pro přípravu dat	89
Workflow úlohy	90
Úlohy pro SQL Server	91
WMI úlohy	91
Skriptové úlohy	92
Úlohy na zajištění zpětné kompatibility	92
Úlohy využívající analytické služby.....	92
Úlohy na správu a údržbu	93
Diagram jednoduchého integračního balíčku.....	93
Příklad: Vytvoření transformačního balíčku pomocí průvodce	95
Námět příkladu.....	96
Možnosti práce s integračním balíčkem v návrhovém prostředí.....	101
Data Flow Task	106

Data Flow Sources	107
Data Flow Transformations	110
Data Flow Destinations	122
Data Flow Path Editor	122
Spuštění balíčku integračních služeb.....	123
Příklad: Import dat z textového souboru.....	125
Alternativní řešení bez využití integračních služeb	134
Příklad pokročilejší konverze textových dat	136
Příklad: Výběr dat ze dvou datových zdrojů, které obsahují duplicitní data	144
Ladění projektů integračních služeb	159
Nasazení integračních projektů.....	161

Kapitola 5

Analytické služby..... 165

Zopakování základů teorie relačních databází	166
Databázová tabulka.....	166
Relační vztahy	168
Vztahy mezi entitami.....	169
Normalizace databází.....	169
Multidimenzionální databáze	172
Multidimenzionální databázový model	172
Porovnání relačního a multidimenzionálního modelu.....	174
Teoretický úvod do problematiky analytických databází	175
Fakta a dimenze	176
Schémata tabulek dimenzí	178
Úložiště multidimenzionálních dat MOLAP, ROLAP, HOLAP	179
Multidimenzionální OLAP (MOLAP).....	180
Relační databázový OLAP (ROLAP)	181
Hybridní OLAP (HOLAP).....	181
Úvodní příklad pro OLAP analýzu	181
Vytvoření relačních struktur, které budou základem pro vytvoření OLAP kostky... ..	188
Vytvoření OLAP kostky rodinného rozpočtu pomocí průvodce	193
Definování datových zdrojů	194
Definování pohledů na datové zdroje	195
Návrh dimenzí	199
Návrh kostky.....	204
Sestavení projektu Analytických služeb.....	207
Prohlížení OLAP kostky.....	214

OLAP kostka v jazyku SQL – klauzule CUBE.....	216
Příklad využití klauzule CUBE.....	216
Vytvoření OLAP kostky z datového skladu.....	227
Databáze AdventureWorksDW2008 jako zdroj dat	228
Vytvoření projektu	229
Návrh kostky.....	230
Definování atributů časové dimenze	235
Definování atributů časové dimenze	236
Definování atributů a hierarchie zákaznické a geografické dimenze	239
Práce s OLAP kostkou v prostředí BI Dev Studio.....	246
Záložka Cube Structure	246
Záložka Dimension Usage	246
Záložka Calculations.....	247
Záložka KPIs.....	250
Záložka Actions	252
Záložka Perspectives	253
Záložka Partitions	254
Záložka Aggregations.....	254
Záložka Translations	255
Záložka Browser	256

Kapitola 6

Dolování dat – data mining263

Oblasti použití data miningu	265
Co data mining neumožňuje.....	266
Teoretický úvod – stručné statistické minimum	267
Rozdělení pravděpodobnosti a testování hypotéz.....	267
Statistické metody využívané data miningovými modely.....	269
Model procesu data miningu.....	272
Procesní schéma data miningu	272
Algoritmy pro data mining	274
Rozhodovací stromy	274
Shlukování.....	276
Sekvenční shlukování.....	276
Asociační pravidla	276
Časové řady.....	277
Neuronové sítě	277
Naive Bayes.....	280
Typické okruhy úloh a výběr algoritmů pro jejich řešení	280

Klasifikace	280
Regrese	281
Segmentace	282
Přiřazování	282
Analýzy textu	282
Úvodní příklad pro data mining	283
Příprava dat	284
Definování datových zdrojů a pohledů	286
Návrh data miningového modelu	288
Prohlížení modelu	293
Predikce nových případů	294
Příklad z praxe – rozlišení jedlých a jedovatých hub	296
Příprava vstupních dat	297
Projekt data miningového modelu	301
Prohlížení rozhodovacího stromu	303
Predikce na základě výsledků analýzy	307
Příklad z byznysu – predikce obchodních případů	311
Příprava vstupních dat	311
Vytvoření data miningových modelů	311
Umístění modelů na server, jejich procesování a prohlížení	315
Porovnání přesnosti jednotlivých modelů	317
Predikce	317

Kapitola 7

Reportovací služby 323

Kategorizace reportů	324
Enterprise Reporting	324
Embedded Reporting	325
B2B Reporting	325
Architektura a filozofie reportovacích služeb	325
Porovnání architektury reportovacích služeb s předcházející verzí	326
Životní cyklus reportu	329
Návrh reportu	330
Správa reportu	330
Doručení reportu	330
Úvodní příklad – návrh výstupní sestavy dat z databáze	331
Výběr zdroje dat	332
Návrh dotazu pro výběr dat	333
Návrh vzhledu reportu pomocí průvodce	339

Deployment reportu	343
Letmý pohled do záznamů na zdrojový kód reportu.....	343
Prohlížení reportu v návrhovém prostředí.....	345
Pokračování návrhu v integrovaném návrhovém prostředí.....	346
Ovládací prvky zobrazení reportu.....	347
Možnosti exportu výstupní sestavy.....	349
Umístění reportu pod správu reportovacích služeb	351
Pohled „za oponu“ reportovacích služeb	353
Nastavení přístupových práv pro spuštění aplikace Report Server na nových operačních systémech Windows Server 2008, Windows Vista a Windows 7	354
Nástroj pro konfiguraci reportovacích služeb	354
Správa reportu.....	356
Příklad návrhu reportu bez využití průvodce	362
Definování zdroje dat a dotazu pro výběr	363
Grafický návrh reportu	364
Formátování údajů.....	369
Seskupování údajů v reportu.....	369
Řádky se sumárními údaji	370
Interaktivní parametrický report.....	371
Výběr nabízených hodnot parametru z datasetu.....	376
Vícenásobné hodnoty parametrů	380
Odkaz na detailní report.....	381
Využití prvku Matrix pro zobrazení sumárních údajů	383
Grafická prezentace údajů v reportu.....	388
Grafické vyjádření klíčových ukazatelů výkonnosti	393
Vyjádření KPI změnou barvy pozadí buněk tabulky	393
Poměrové ukazatele typu „gauges“	395
Struktura jazyka RDL pro návrh reportů	397
Agregační funkce	399
Doručení reportu	401
Data – Driven Subscription.....	402
Report Builder	403
Vytvoření modelu reportu	404
Model reportu pod správou databázového serveru	408
Návrh reportu pomocí Report Builderu	409
Report Builder 2.0.....	412

Kapitola 8

Klientský přístup 419

Kategorizace klientského přístupu	420
Tlustý klient	421
Tenký klient.....	421
Požadavky na Smart klienta	421
Úvod do předpokladů a možností kancelářského balíku MS Office v roli klienta analytických služeb	422
Kontingenční tabulka (Pivot Table) a kontingenční graf	422
Kontingenční tabulka (Pivot Table) v prostředí aplikace Excel	424
Excel 2007 jako klient analytických služeb.....	426
Připojení k analytickému serveru a výběr analytických struktur	427
Zobrazení analytických dat ve formě kontingenční tabulky.....	431
Vnořování do hierarchických struktur.....	431
Filtrování dat	433
Grafická vizualizace dat.....	433
Podmíněné formátování buněk.....	435
Kontingenční graf.....	436
Práce s daty OLAP v režimu offline.....	437

Přílohy 443

Příloha 1: Cvičná databáze Adventure Works 2008.....	444
AdventureWorks 2008	444
Příloha 2: Cvičná databáze Adventure WorksDW 2008	445

Rejstřík 447

Úvod

Současná situace v oblasti využívání údajů pro podporu rozhodování je podmíněná neustále se měnícími požadavky byznysu, přičemž údaje uchovávané jednotlivými subsystémy podnikové informatiky v mnohých firmách nedokážou na tyto požadavky operativně reagovat. Aby byly informace pro podporu rozhodování relevantní, musí být poskytnuté v reálném čase a v požadovaném formátu, a to na úrovni strategické, taktické i operační. Úspěšné řízení firem organizací nezajistí množství údajů, ale rychlost a přesnost, s jakou se manažerům daří z těchto údajů získávat klíčové informace pro řízení podniku. Největší výzvou dneška je porozumění celkovým obchodním datům, speciálně tomu, jaká klíčová data existují, kde jsou uložena, co je jejich zdrojem, jaké jsou jejich závislosti (jestli jsou) a jaký je způsob integrace s ostatními systémy.

Business Intelligence jako proces přeměny údajů na informace a poznatky je jedním z nejdynamičtěji rostoucích odvětví informačních technologií. Nejvíce se využívá v podnikových informačních systémech, ale nachází si cestu i do různých laboratoří, například ve zdravotnictví, a také do oblasti výzkumu a vývoje. Problematiku analytických databází a datových skladů, dolování dat (data mining) zahrnuje nejen technologické aspekty, ale i plánování, návrh, zavedení a řízení konkrétního projektu databáze nebo datového skladu.

BI je nadále oblastí vysoké priority a zvyšujících se výdajů pro organizace, které chtějí zlepšit svůj rozhodovací proces založený na údajích sesbíraných z procesů, aplikací a od lidí. Proto je BI hlavní prioritou pro IT exekutivu a zároveň technologií s jednou z nejvyšších návratností.

Hlavní přínosy řešení Business Intelligence spočívají v přechodu z intuitivního rozhodování na rozhodování kvalifikované, realizované na základě kvalitních, konsolidovaných a operativně dostupných informací dodaných ve správný čas správným osobám. To umožňuje zlepšení obchodních a marketingových aktivit, možnost sledování a předvídání trendů, a tedy v konečném důsledku hlavně zvýšení konkurenceschopnosti firmy.

Tato publikace vychází z koncepce knihy Business Intelligence na platformě SQL Server 2005, ale není to ani její jednoduchý upgrade z verze 2005 na verzi 2008 a už vůbec ne rozdílová publikace. Jak vyplývá z názvu, je kniha zaměřená na verzi SQL Server 2008 a snaží se přinést komplexní pohled na problematiku Business Intelligence na této platformě.

Osnova je koncipovaná tak, že po nenáročném teoretickém úvodu do jednotlivých oblastí je těžiště kladeno na praktické zvládnutí dané problematiky na jednoduchých a srozumitelných příkladech. Knihu mohou používat i začátečníci, ale určitá znalost problematiky databází a databázového jazyka SQL bude velkým přínosem.

Přestože problematika analýz a datových skladů se v podnicích pojí s datovými centry s výkonnými servery a mnoha terabytovými úložišti, všechny příklady v knize jsou koncipovány tak, aby je bylo možné vyzkoušet na běžném PC nebo na notebooku. Pro zajímavost uvádíme, že všechny příklady z této knihy byly realizovány a odzkoušeny na notebooku s procesorem Intel Core 2 Duo L7100 taktovaným jen n 1,20 GHz s 2 GB operační pamětí, což je v době vydání knihy hardwarová konfigurace dimenzovaná hodně pod optimálním poměrem cena/výkon. Přesněji, na tomto notebooku běžel vir-

tuální počítač Virtual PC 2007 využívající 1 GB paměti. Až na tomto virtuálním počítači byla nainstalovaná verze SQL Server 2008 Enterprise Edition.