

NÁSTROJE BUSINESS INTELLIGENCE

Milena Tvrdíková

VŠB – Technická univerzita Ostrava, Ekonomická fakulta, Katedra informatiky v ekonomice, Sokolská 33, 701021 Ostrava1, ČR, milena.tvrdikova@vsb.cz

Abstrakt

Nástroje business intelligence se ve stále větší míře uplatňují v podnicích a institucích po celém světě. Jsou speciálně zaměřeny na podporu potřeb řídicích pracovníků. Je to část celkového IS firmy, která pracuje s vybranými nebo upravenými daty, a která se těmito úpravami stává nositelem komplexních informací, charakterizujících příslušné procesy ve firmě. Primárně slouží k identifikaci a lokalizaci určitých jevů ve firmě, v dalším kroku pak k jejich podrobné analýze. Za velmi významné považují rovněž působení nástrojů business intelligence jako integračního faktoru IS firmy. V příspěvku jsou probrány základní principy fungování těchto aplikací.

1. Úvod

Jedním z posunů, ke kterým v podnikové informatice došlo během několika posledních let, byl přesun priorit ke strategickému řízení. Pro takovýto přístup je podstatné mít dostatek správných informací. Toto nám nabízí systém nástrojů, projektových řešení a organizačních opatření známých pod pojmem business intelligence (BI), umožňujících řízení organizací podle znalostí. Rozsáhlá oblast business intelligence se skládá z řady samostatných komponent, má vlastní architekturu i metodiky a s provozními systémy je provázána řadou vazeb. Kromě tradičních aplikací, rozvíjí koncept MIS také EIS (Executive Information Systems) a nástroje dolování dat. Infrastruktura celé oblasti bývá založena na konceptu datového skladu.

Existence datového skladu je předpokladem pro fungování nástrojů BI. Jedná se o ucelenou databázi optimalizovanou pro dotazování a analýzu dat, společně s nástroji, které dotazy, analýzy a kvalitní prezentaci výstupů umožňují. V datovém skladu jsou data integrována a ukládána, ať už se jedná o data z interních nebo externích zdrojů. Konečným cílem je poskytnout čitelné, organizované, analyzovatelné a v reálném čase dostupné informace z maxima podnikových databází i externích zdrojů, které jsou ve velkém rozsahu využitelné při řízení firmy či instituce.

2. EIS – Executive Information System

Prostupování integrovaných informačních systémů (ERP - např. SAP R/3 nebo BAAN) do všech procesů v organizaci spolu s nástupem komunikačních technologií vede k dramatickému růstu objemu dat v těchto systémech. Proto potřeby efektivního řízení vyžadují sledování souhrnných informací, jejichž objem není tak veliký, ale jsou to data se značnou mírou agregace. Tento účel splňují EIS. Jde o efektivní propojení datového potenciálu provozního transakčního systému s řídicími mechanismy organizace a analytickými potřebami jednotlivých složek řízení. Jsou navrhovány tak, aby jednak

umožňovaly přístup k externím datům, jednak byly napojeny na informační systém organizace.

EIS dokáží transportovat značné objemy primárních dat, generovaných základními procesy dané firmy do logických struktur, které reprezentují řídicí a rozhodovací procesy v konkrétní firmě. Jsou charakteristické jednoduchým ovládním (často intuitivním) a obsahují účinné prostředky pro přehlednou prezentaci dat.

2.1 Charakteristika EIS

EIS využívají tzv. OLAP (On Line Analytical Processing) nástrojů, „zajišťujících vysoce efektivní mechanismus vícekritériální analýzy. Pro jejich hodnocení bylo definováno zhruba 18 pravidel, z nichž prvních 12 definoval E.F.Codd.

Základní vlastnosti:

- Multidimenzionální koncept a manipulace s daty:
realizuje uložení dat v kombinaci definovaných dimenzí, umožňuje různé pohledy na data podle dimenzí a dokáže je dynamicky měnit.
- Intuitivní manipulace s daty z pohledu uživatele:
grafické rozhraní, zooming.
- Možnost získávat data z heterogenních datových zdrojů:
OLAP nástroje musí mapovat stav uložení dat, přístup k nim a zajistit příslušné konverze do vlastní datové báze.
- Nabídka vlastních databází pro OLAP i přímý přístup do externích dat.
- Podpora moderních analytických přístupů.
- Klient/server architektura.
- Transparentnost a otevřenost:
schopnost integrace s jinými nástroji bez vlastního ohrožení, uživatel může využívat data z databáze nástroji, na které je zvyklý. Transparentnost vnitřní organizace dat.
- Podpora víceuživatelského provozu:
paralelní přístupy, zajištění integrity a bezpečnosti provozu.

Speciální vlastnosti:

- Efektivní zpracování řídkých matic:
chybějící hodnoty musí být ignorovány OLAP analyzátoři bez ohledu na jejich zdroj.
- Zpracování nenormalizovaných dat:
integrace OLAP stroje a nenormalizovaných zdrojových dat.
- Odlišení chybějících hodnot od nulových.

Přehledové vlastnosti (reporting):

- Flexibilní poskytování výstupů:
snadné úpravy výstupů, zobrazení a kombinace výstupů podle okamžitých potřeb uživatele.
- Konzistentní výkon na výstupech:
výkon by neměl být ovlivněn počtem definovaných dimenzí.

Řízení dimenzí:

- Generická dimenzionalita:
dodatečně dodávané funkce musí být zajištěny pro jakoukoliv dimenzi.

- Neomezený počet dimenzí a agregačních úrovní:
počet dimenzí pro model by neměl být limitován.
- Neomezené operace napříč dimenzemi:
operace s daty mezi jednotlivými dimenzemi nemohou být omezeny počtem dimenzí.

Dodržení OLAP standardů, spolu s posouzením jednoduchosti nasazení a ovládní systému a schopnosti dodavatele zhotovit v krátké době prototyp budoucí aplikace nad vzorkem vašich skutečných dat, jsou jednoznačným kritériem pro výběr konkrétního EIS.

2.2 Multidimenzionalita EIS

Typickou vlastností EIS je multidimenzionalita, která umožňuje rychle a jednoduše vytvářet nové pohledy na data, jejich řazení do nových souvislostí, vyhledávání zákonitostí (trendových charakteristik), indikaci odchylek klíčových ukazatelů od plánovaných hodnot, práci s historií a anticipací budoucího vývoje. Aplikace EIS jsou založeny především na využití tzv. multidimenzionálního uložení dat. Základem je multidimenzionální tabulka umožňující rychle a pružně měnit jednotlivé dimenze, tedy měnit pohledy uživatele na modelovanou ekonomickou realitu.

Pohled na ekonomiku jakékoliv organizace je multidimenzionální. Pro ekonomické aplikace, jež jsou rozvinutelné v čase, je počet tří dimenzí, užívaný v tabulkových kalkulátorech, nedostatečný. U malých organizací manažer neztrácí globální přehled a dokáže sledovat i podrobnosti. U větších organizací musí být informace pro manažera pečlivě připravovány. Multidimenzionální forma informací je pak velmi praktická, možná i nutná.

Úlohy tohoto charakteru jsou typickými úlohami v podnicích, ve státní správě, v bankovníctví, kontrolních úřadech apod. Výhoda EIS při řešení těchto úloh spočívá ve schopnosti velmi operativně pracovat v nadprostorové matici s výrazně větším počtem buněk.

Několika dimenzionální uložení dat je většinou realizováno na bázi metadatové nadstavby nad relačními tabulkami. Metadata přiřazují řádky a sloupce relačních databází jednotlivým dimenzím a buňkám v n-dimenzionální tabulce. V metadatach jsou také obsažena pravidla agregace dat na jednotlivých úrovních definovaných dimenzí.

Toto je princip OLAP technologie ukládání dat neboli on-line analytického zpracování dat.

Odezva na změnu definice pohledu je pak v nástrojích využívajících OLAP technologie téměř okamžitá. Problém expanze multidimenzionálních dat se projevuje až v určitých situacích, jako jsou přílišná hierarchická složitost dimenzí, zvětšující se počet dimenzí a s tím související řídkost dat apod. Otázka administrace multidimenzionálních dat rovněž úzce souvisí s typem ukládání a dalším zpracováním těchto dat.

2.3 Využívané technologické principy

OLAP technologie jsou rozdělovány do tří skupin – podle typu ukládání a způsobu zpracování dat:

- ROLAP (Relational On-Line Analytical Processing)
- MOLAP (Multidimensional On-Line Analytical Processing)
- HOLAP (Hybrid On-Line Analytical Processing)

ROLAP		MOLAP		HOLAP
výhody	nevýhody	výhody	nevýhody	
<ul style="list-style-type: none"> • Přímý přístup k datům až na konkrétní doklad • Bez časové prodlevy od změn v primárním systému 	<ul style="list-style-type: none"> • Dlouhá doba odezvy • Musí se definovat metadata popisující relační databázi 	<ul style="list-style-type: none"> • Jednodušší nastavení komunikace • Rychlejší zpracování požadavku • Bez omezení použitého SQL jazyka • Nižší zátěž využívaného hardware 	<ul style="list-style-type: none"> • Lze získat pouze agregovaná data • Realizace databáze s určitým zpožděním 	Kombinace obou předešlých technologií

ROLAP reprezentuje přímý přístup k datům relačního primárního systému, což znamená, že data prezentovaná v zobrazovacím nástroji jsou získávána přímo z původních datových zdrojů, např. z tabulek databáze Oracle (přístup do těchto tabulek je obvykle realizován prostřednictvím „ODBC“ ovladačů v okamžiku potřeby). Pro uložení dat se tedy používají standardní relační databáze a data z nich jsou vybírána pomocí SQL dotazů.

Produkty kategorie **MOLAP** využívají specifickou multidimenzionální databázi. Informace jsou v této speciální databázi, navržené jako množina multidimenzionálních matic a jsou aktualizovány a doplňovány v určeném pravidelném intervalu.

HOLAP představuje přístup kombinující obě předešlé technologie. V tomto případě jsou data čerpána přímo z primárních zdrojů, přičemž jejich část je ukládána do multidimenzionálních matic (model MOLAP). V zásadě se do multidimenzionální databáze ukládají agregované údaje, jejichž získávání bývá nejvíce časově náročné. Rozlišení správného stupně agregace, který bude představovat hranici mezi podrobnějšími daty ukládanými do relační datové struktury a agregovanými informacemi zaváděnými do multidimenzionálních databází, je významným úkolem pro implementátora EIS.

Lze říci, že nástroje dostupné na trhu v ČR podporují HOLAP přístup, přičemž stanovení hranice mezi ukládáním dat do OLAP struktur a přímým přístupem do relačních zdrojů je možno individuálně stanovit v rozmezí 0-100% (tedy od případu, kdy se v multidimenzionálních databázích neukládá nic, až po situaci, kdy zobrazovací nástroj čerpá data pouze z OLAP zdrojů).

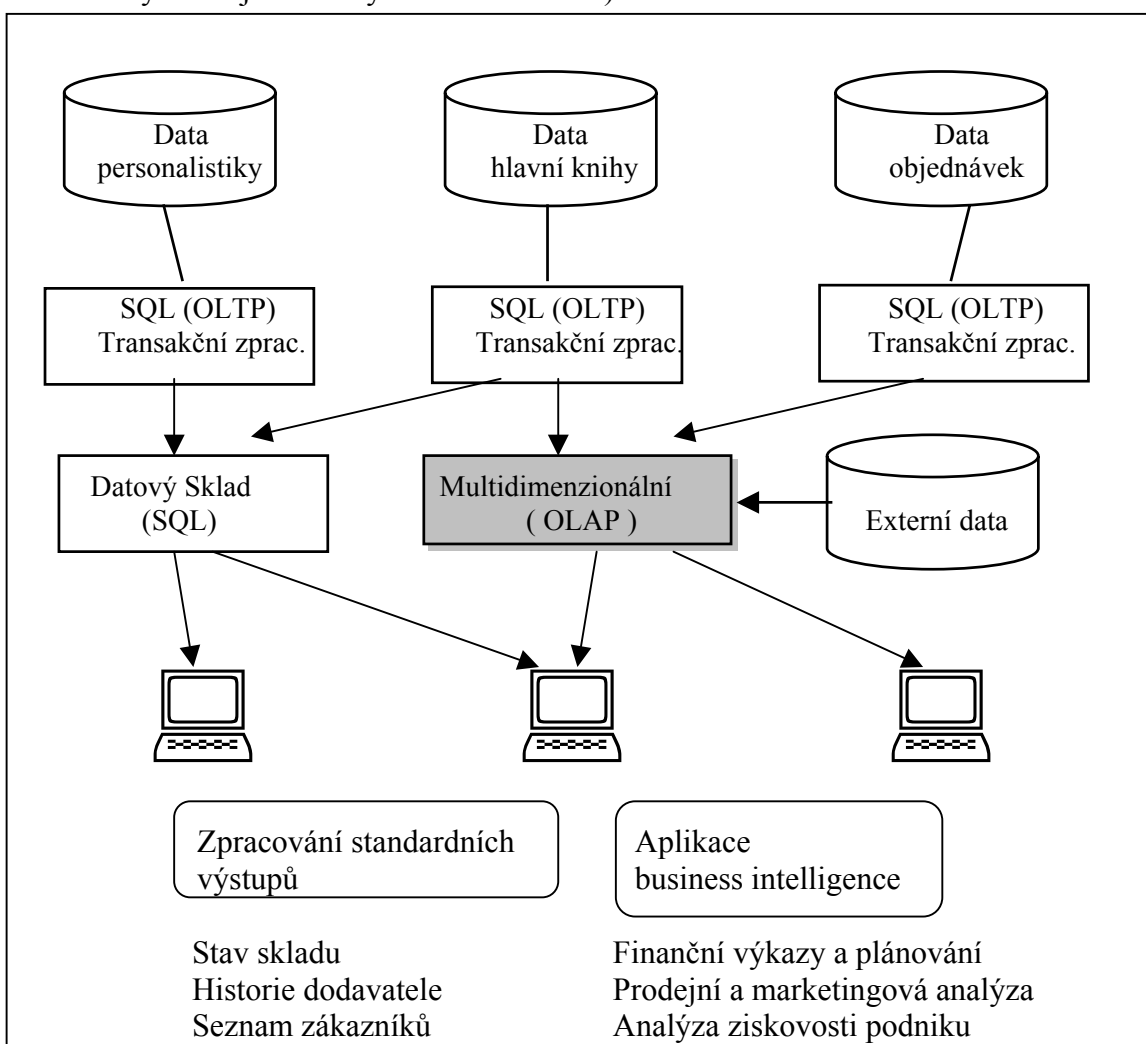
3. Datové sklady

Datový sklad (DS) je dlouhodobým úložištěm, kam data shromážděná klasickými informačními systémy přibývají po jednotlivých dávkách.

V datovém skladu:

- odezvy na dotaz nemusí být okamžité
- připouští se určitá redundance dat

- data se z datového skladu nikdy nelikvidují (mohou se provádět souhrnné agregace některých údajů a zálohy na externí média).



Obr.1. Schéma technologie OLAP – komplexního řešení podpory rozhodovacích procesů, nazývané též business intelligence

Charakteristickou vlastností datového skladu je různorodost zdrojů. Zdroje pro datové sklady jsou uloženy ve zcela odlišných strukturách, formátech, mají odlišnou filozofii záznamu, jsou uloženy na různých médiích apod.

Datové sklady fungují na principu tří kroků (ETL):

- Extrakce – schopnost převzít data z co nejširšího spektra datových zdrojů nejrůznějšího charakteru. (Jedním ze způsobů přístupu může být standardizované rozhraní ODBC.)
- Transformace – postupná řada operací, které extrahovaná data připraví pro vlastní načtení do datového skladu. (Dochází ke kontrole, doplnění nebo změně dat, převodu na stejné formáty a odstranění nekonzistence.)
- Load – Data jsou natažena do vlastního fyzického prostoru datového skladu. (Jsou připravena na vytěžování čili pokládání dotazů.)

V architektuře datových skladů se prosadily dva koncepty – nezávislé data marty (virtuální datové sklady),
- integrované datové sklady.

Nezávislé data marty jsou samostatná datová úložiště pro jednotlivé aplikace nebo útvary. Jejich nevýhodou je možná nekonzistence mezi jednotlivými úložišti a komplikované načítací procesy.

Integrovaný datový sklad je centrální datové úložiště, kde je požadavek konzistence naprosto zásadní (DS musí poskytovat „jedinou verzi pravdy“). Tento přístup v současnosti převládá.

3. Dolování dat

Datový sklad je pak nástroji business intelligence využíván přes službu nazývanou dolování dat. Dolování dat na základě určitého předpokladu umožňuje vyhledat ve velkém objemu dat souvislosti a vzájemné vztahy, které nebyly známy dopředu. Tím, že umí vyhledávat data o souvislostech, které nebyly dopředu známy, se dolování dat liší od jiných metod počítačem zpracovávaných datových analýz.

Dolování dat by mělo mít vždy za cíl řešení konkrétního obchodního problému nebo nalezení cesty k vylepšení procesu. Cíl musí být předem definován a na jeho základě by se měla připravovat data.

Dolování dat je hledáním skrytých souvislostí, procesem výběru, prohledávání a modelování ve velkých objemech dat. Slouží k odhalení dříve neznámých vztahů mezi daty.

Data musí být samozřejmě očištěna od chyb, úplná a formáty z různých systémů musí být sjednoceny. To zajistí jejich uložení v datovém skladu. Charakter dolování dat vyžaduje, aby data byla průběžně aktualizována.

Spektrum metod, které se využívají při budování modelu dolování dat je velmi rozsáhlé. O žádném modelu nelze říci, že je univerzální a nejlepších výsledků se dosahuje kombinací různých přístupů.

Dnes užívanými metodami dolování dat jsou například:

- odhady hodnot vysvětlované proměnné (lineární regresní analýza, nelineární regresní analýza, neuronové sítě)
- klasifikace (diskriminační analýza, logistická regresní analýza, rozhodovací stromy, neuronové sítě),
- segmentace – shlukování (shluková analýza, genetické algoritmy, neuronové shlukování – Kohonenovy mapy)
- analýza vztahů (asociační algoritmus pro odvozování pravidel typu „if X then Y“)
- predikce v časových řadách (Boxova-Jenkinsonova metoda, neuronové sítě)
- detekce odchylek (vizualizace, statistické postupy).

Některé přístupy jsou založeny na přesně popsaném matematickém modelu a aplikace na konkrétní data se sestává z testování hypotéz a výpočtu neznámých koeficientů. Druhá skupina modelů mění svou strukturu dynamicky, na základě dat, která zpracovávají.

Užívané metody dolování dat jsou bedlivě strážným know-how SW firem.

4. Integrovaný charakter nástrojů business intelligence

Za velmi významné považují působení nástrojů business intelligence jako integračního faktoru IS firmy.

Prispívají k:

- integraci řídicích procesů
 - data jsou dostupná pro všechny úrovně řízení a to v téměř stejném časovém úseku,
- integraci dat
 - tvorba datového skladu a realizace dolování dat
- integraci informačních technologií
 - minimálně prostřednictvím dat propojují i různé ve firmě užívané informační technologie a technologické platformy, stávají se motorem propojení IS firmy do jednotně fungujícího celku,
- celkové integraci systému
 - aplikace BI patří v rámci firmy k aplikacím strategickým. Přenášejí data mezi všemi úrovněmi řízení ve firmě a tak představují významný faktor pro integraci celého IS firmy. Při jejich vývoji a implementaci dochází k uvědomování si jednotlivých částí IS firmy a tím dochází k jejich konečné integraci do synergicky fungujícího celku – integrovaného IS firmy.

Naše století se dá bez nadsázky charakterizovat nebyvalou informační explozí. Největšímu informačnímu nátlaku jsou vystaveni manažeři a řídicí pracovníci firem. Čím je firma rozsáhlejší nebo rozmístěna na více místech, tím roste i tlak na kvalitu, relevantnost, spolehlivost i kvantitu předávaných informací.

Schopný a vzdělaný manager potřebuje pro svou důležitou činnost adekvátní nástroj – kvalitní manažerskou nadstavbu informačního systému. Lze říci, že nástroje business intelligence integrují informační a podnikatelskou strategii a pomáhají podniku přežít v současném turbulentním prostředí.

Literatura:

1. Systémová integrace, roč.4., č.3., 1997
2. Firemní příručka Media M, Speedware Corporation Inc., 1998
3. Geldner, D.: Role a architektury Business Intelligence ve strategickém řízení IS/IT finančních institucí, Systémová integrace, č.1., 2002
4. Tvrdíková, M.: Materiály k výuce předmětu Systémový hardware a software