

BUDOVÁNÍ DATOVÝCH SKLADŮ JAKO PROCES INSPIRACE A UČENÍ

Vladimír Kyjonka

Sybase ČR, Kyjonka@sybase.com

Abstrakt

Informace, obsažené zjevně i potenciálně v datech datových skladů, představují bohatství, jež je velmi žádoucí, ale mnohdy nesnadno dostupné. Tento článek pokouší ukázat některé z možností, jak je vytěžit co nejúčinněji bez nebezpečí devastace informačního terénu necitlivým nasazením "těžké techniky".

Keywords: Datový sklad, databáze, životní cyklus projektu, proces inspirace a učení, míra rizika, otevřené systémy, jedna verze pravdy, princip priorit, doba implementace

1. Datový sklad jako prostředek poznání

Datawarehousing je často tradičně chápán jako proces a řešení, jež svými nároky přesahuje mnohonásobně představy o běžném informačním systému. Předpokládá se, že jde o mnohaterabytové databáze, provozované na mainframe počítačích nebo masivně paralelních systémech, budované v dlouhých časových horizontech (měsíce a roky), vyžadující značné nasazení řešitelských a provozních kapacit a představující rovněž řešení velmi drahé. Přitom jsou většinou koncipovány jako systémy spíše rigidní než pružné, používají všechnu nasazenou výpočetní kapacitu pro uspokojení relativně malého počtu specializovaných uživatelů, informace jsou dosažitelné většinou v pevně definovaných strukturách a omezeným spektrem nástrojů.

Současný vývoj ve všech oblastech IT je však dnes charakterizován snahou dostat informace co nejlíže jejich "skutečným" uživatelům, ve formě, na kterou jsou zvyklí, která je pro ně stravitelná a ve struktuře, jež má právě pro ně největší vypovídací hodnotu. V tomto světle se jeví jako vysoce legitimní požadavek, aby výsledné řešení bylo pružné, výkonné a relativně nenáročné.

Stav který je jistým způsobem příznačný při zavádění nových postupů a systémů zejména v oblasti datových skladů, je často charakterizován větami jako:

- "Všichni to potřebují"
- "Nikdo neví, jak to má vypadat"
- "Každý si pod tím představuje něco jiného" - atd.

V podstatě se nelze této situaci příliš divit; problematika datových skladů je sice mnohde diskutována a rozebírána, na druhé straně však představuje ve využívání IT zcela nový fenomén, se kterým má praktické zkušenosti málokdo. Navíc (a to je odlišnost proti zavádění "klasických" operativních systémů), přináší pro vlastní činnost jejich uživatelů řadu nových možností, které mohou výrazně ovlivnit jejich způsob práce.

Teprve po určité době, kdy jsou nové aplikace a nástroje prakticky využívány, se jejich uživatelé blíže seznámí s obsahem spravovaných dat a jejich informativními možnostmi, a tak si osvojí nové metody a postupy, jak jich využívat. Tato zkušenost je velice cenná a v druhém (a dalších) kolech pak vede k mnohem kvalifikovanější a konkrétní specifikaci požadavků.

Tento proces neustálého **učení a inspirace** je pro moderní pojetí charakteristický a umožňuje shromážděné informace využívat intenzivněji a flexibilněji, neboť jeho budování je založeno a spoluvytvářeno na základě přímého zapojení a zkušenosti jeho účastníků.

Je žádoucí, aby architektura a technologie datového skladu byly zvoleny tak, aby tyto možnosti podporovaly, tedy aby už v raných fázích jeho budování bylo možno uživatelům poskytnout vzhled do jeho obsahu, který jim vytvoří prostor pro zorientování v množství v něm obsažených informací a současně poskytne kvalifikovanější podklady pro analýzu.

2. Životní cyklus projektu Datového skladu a jeho paradoxy

Pro datové sklady je charakteristický životní cyklus, jenž je odlišný (ve většině případů opačný), ve srovnání s provozními OLTP systémy.

Budování OLTP systémů vychází z obecných strategických záměrů a formulování požadavků na budoucí systém. V dalších krocích se pak provádí analýza datových struktur a funkcí systému, jež vyústí v design a následnou implementaci a uvedení do provozu. Teprve potom se příslušné struktury začnou naplňovat daty.

V případě datových skladů je situace zcela odlišná. To, co je dáno hned od začátku, jsou data. Existují v systému ještě před započítáním prvních úvah o datových skladech ve strukturách, daných jejich využitím v OLTP systémech. Teprve po jejich zdokumentování, popisu a faktickém poznání je možno kvalifikovaně formulovat požadavky na jejich využití. Požadavky na funkce systému jsou "paradoxně" formulovány na konci cyklu.

V tomto kontextu je velmi důležitý pohled na datový sklad skutečně jako na záležitost podchycení, ukládání a zpřístupnění dat. Neboť vlastní data v systému existují a budou existovat po celou dobu jeho života, nezávisle na tom jakým způsobem, kdy a v jakém rozsahu budou prezentována.

Teprve vytěžením určitého objemu informací z datového skladu a zorientováním se ve věcné podstatě jeho informačního bohatství nabývají jeho uživatelé větší schopnosti

formulovat cíleněji relevantní dotazy, jež vyústí v požadavky na další funkce a formalizaci struktur datových skladů. Na základě jejich realizace pak obdrží informace v nové kvalitě a hloubce. Tento proces inspirace a formulování nových požadavků může být v podstatě nekonečný, s každým novým cyklem však nabývá podoba datového skladu reálnějších obrysů a cílenější podoby.

2.1 Rizika

Pro současné projekty IT jsou mj. charakteristické dvě skutečnosti, působící ve svém důsledku proti sobě: Vzhledem k tomu, že projekty IT zasahují stále výraznějším způsobem do života organizací, nabývají rizika spojená s jejich nesprávnou funkcí podstatně výraznějšího charakteru pro život podniku. Proto jsou zcela na místě zvýšené nároky na důkladnou a přesnou analýzu požadavků, architektury i samotný návrh systému.

Na druhé straně se výrazně zkrátila doba aktuálnosti a životnosti všech typů informací; zejména v oblasti IT se může rychlost a včasnost realizace systému stát rozhodujícím faktorem úspěchu. "Klasický" přístup vytváření projektů s mnohaletým horizontem realizace je v tomto směru elementem, zvyšujícím riziko. Požadavek na maximální zkrácení projekčního cyklu je tedy zcela legitimní.

Pro datové sklady je tento aspekt zvýrazněn převráceným charakterem životního cyklu projektu; pravděpodobnost "minutí cíle" při nesprávném nasměrování je zde výrazně vyšší. (Podle údajů společnosti Price Waterhouse je z celkového počtu projektů datových skladů úspěšně dokončeno 26 % ! Při průměrné ceně projektu 3 mil USD, kterou uvádí též pramen, je na první pohled zřetelně vidět, o jak velké riziko se jedná.) Čelit tomuto nebezpečí je možno pouze dostatečně flexibilním přístupem, jenž při důkladné analýze umožní zapojit korektivní mechanismy a využít již v jednotlivých etapách výstavby data warehouse procesu učení, jenž se uplatňuje při využívání (tj. fakticky provozování) systému.

3. Jak je to s jednou verzí pravdy

Jako jeden ze stěžejních argumentů pro zavádění datových skladů je často uváděn požadavek na dosažení "jedné verze pravdy". V systémech, jež historicky procházely různými vývojovými formami, jež navíc bývá jí složeny z více nebo méně nesourodých částí, určených především k pokrytí specifických podnikových agend, jsou často tytéž informace uchovávány v různé formě, granularitě, ale hlavně i různém logickém kontextu. Veškeré analýzy pak mohou být znehodnoceny právě odlišným chápáním významu a obsahu jak vstupních, ale i výsledných informací. Především pro rozhodování globálního a strategického charakteru je třeba vytvořit jednotnou základnu, která bude představovat pojmové i obsahové východisko pro interpretaci všech použitých kategorií v rámci celého podniku - "jedna verze pravdy". Tento požadavek bývá často používán jako argument pro "orthodoxní" řešení, jejichž tvůrci zpravidla říkají: Realizaci projektu musí předcházet velmi důkladná analýza, na základě které bude ona "Jedna verze pravdy" stanovena. Až ji budeme mít, pak teprve vybudujeme "Správně fungující systém".

Ve skutečnosti ovšem v praxi neexistují žádná více neurčitá zadání, než jsou zadání pro řešení datových skladů. Ona "Pravda" je totiž cílem a smyslem celého řešení a datawarehousing má být nástrojem pro její nalezení. Je bláhové se domnívat, že uživatelé (zadavatelé) sami ji naformulují na začátku projektu, a ještě bláhovější tvrdit, že ji stanoví externí analytici dodavatelské firmy. Její vyvíjející se poznání, hledání a nalezení je plně v rukou těch, kterým příslušné informace patří a mají sloužit. Vzhledem k tomu, že je tato kategorie do značné míry rozhodující pro strukturu a obsah řešení datového skladu, je třeba myslet velmi vážně na to, aby návrh systému, jeho projekt a realizace počítal s možností aktivního přístupu k jeho datům ze strany koncových uživatelů už v jeho průběhu a umožnil nové poznatky a požadavky flexibilně začleňovat tak, jak přicházejí.

4. Proces inspirace a učení

Při řešení datových skladů se často setkáváme s tendencí zužovat celou problematiku především na otázku prezentace dat. Většinou se pak jedná o diskusi nasazení specializovaných nástrojů pro OLAP, ROLAP, data mining, automaticky se předpokládá multidimensionální analýza atd. Omezení se pouze na tuto jevovou stránku problematiky vede ovšem k některým důsledkům, jež mohou snížit užitnou hodnotu celého řešení a způsobit neefektivní vynaložení prostředků. Jde zejména o

- **Nutnost přizpůsobit hned od začátku datovou základnu použitým speciálním prostředkům.** Tyto nástroje většinou vyžadují speciální schémata, design databází, někdy je jejich použití dokonce limitováno na jediný databázový systém. Důsledkem je pak na jedné straně omezená flexibilita a větší či menší uzavřenost systému, na druhé straně je praktickým důsledkem zpravidla nepředpokládaný nárůst dat, způsobený enormními nároky na denormalizaci, označovaný jako datová exploze.
- **Orientace na specializované uživatele.** Uvedené prostředky jsou zpravidla vybaveny náročným aparátem, umožňujícím provádět komplikovanou a sofistikovanou analýzu dat s použitím matematických a statistických metod. Tento přístup je velmi užitečný a pro odhalování "skrytých" dějů, analýzu trendů a důkladné vyhodnocení "informačního bohatství" téměř nezastupitelný. V praxi však uživatelé tohoto typu, kteří jsou jednak vybaveni znalostmi, jež jim umožní tyto nástroje využívat, a pro něž jsou takovéto analýzy hlavní pracovní náplní, představují jenom zlomek z těch, kteří by mohli a měli informací z datových skladů využívat.

- **Náročná příprava, údržba, provoz.** Speciální prostředky, schémata a design, zejména při vysoké úrovni denormalizace, podstatným způsobem zvyšují složitost systému. Jestliže jsou tyto postupy uplatněny od počátku na systém jako celek jako univerzální šablona datového skladu, vyžaduje pak jejich dokumentace, údržba, správa a aktualizace enormního úsilí. Připočteme-li k tomu faktor exploze dat, jsou výsledkem téměř nepředpověditelné provozní nároky a velmi nesnadná realizace případných změn. Logickým důsledkem a průvodním znakem tohoto přístupu je relativně dlouhá doba implementace.

V současné době je vývoj IT charakterizován dostupností prostředků IT pro mnohem větší okruh uživatelů, než tomu bylo ještě v nedávné minulosti. Zejména v organizacích, ve kterých jsou počítačové systémy využívány jako běžné pracovní nástroje, se dá hovořit o tom, že uživatelé mohou být prakticky všichni zaměstnanci. Práce s prostředky výpočetní techniky pro většinu z nich však neznamena hlavní předmět jejich pracovní náplně, ale skutečně jen pouhý prostředek pro realizaci jejich úkolů.

V případě datového skladu jde mimo jiné o to, jak z něj učinit nástroj, který může všem potenciálním uživatelům pomoci řešit jejich věcné odborné problémy, aniž by neúměrné nároky na používání speciálních metod vyžadovaly více odborných znalostí přímo nesouvisejících s hlavním oborem činnosti.

4.1 Typy uživatelů datového skladu

Z pohledu možného využívání služeb datového skladu můžeme uživatele rozdělit do několika skupin, které se vyznačují různým zaměřením, postavením v hierarchii rozhodování, zkušenostmi s IT, a které vyžadují odlišný přístup při návrhu funkcí datového skladu a použitých prostředků.

- **Exekutiva**
Hlavním předmětem činnosti těchto uživatelů je rozhodování a řízení. Práci s počítačem mohou zpravidla věnovat čas v rozsahu několika hodin za měsíc. Nedá se u nich předpokládat možnost detailního zaškolení v používání specializovaných nástrojů, znalostí o datových strukturách a detailním obsahu dat. Pro svou práci potřebují získávat hotové odpovědi na své otázky. Ty mohou být realizovány určitými standardizovanými a pravidelnými výstupy, často však mohou být jejich požadavky velmi proměnlivé a nestandardní. Většinou vědí, jaký typ odpovědi potřebují, nedisponují však potřebným časem a dovednostmi, aby si příslušné informace mohli sami zpřístupnit. Pracují většinou s menší či větší podporou specialistů, ať problémově orientovaných, nebo z oblasti IT. Kritickým kriteriem u této skupiny bývá čas, a to i u požadavků velmi nestandardních, protože příslušná informace je zpravidla pro ně zajímavá pouze v době, kdy mají provést rozhodnutí a později je její hodnota podstatně menší.
- **Koncoví uživatelé ("běžní uživatelé IT")**
Pro tyto uživatele představuje jejich hlavní pracovní náplň jejich odborná problematika (např. marketing, obchodování atd.). Výpočetní techniku využívají jako standardního pracovního nástroje a u počítače tráví denně hodiny. Zpravidla jsou velmi zblhlí ve využívání svých aplikačních programů, často jsou zkušení v práci s různými kancelářskými softwarovými produkty (spreadsheets). Nejsou školenými odborníky na výpočetní techniku a pro realizaci dotazů, které jim neposkytují jejich standardní nástroje, rovněž vyžadují podporu IT specialistů.

Vzhledem k tomu, že jsou většinou spjati s operativními agendami každodenního provozu, jsou pro ně zajímavé informace, jež jim umožňují na základě informací globálního charakteru "vystopovat" příčiny příslušných dějů pohledem do detailních dat. Pro přístup k detailu je pro ně rovněž důležitá doba odezvy, jež jim umožní interaktivní reakci, a aktuálnost dat.

- **Analytici**

Jsou zde míněni specialisté, jejichž hlavní činností je business analýza (např. pracovníci zaměřeni na vyhodnocování rizik, vyhodnocování dopadu strategických rozhodnutí, analýzu poptávky po určitých typech služeb atd.). Jsou to lidé, kteří po obsahové stránce nejlépe rozumí datům a dokážou je interpretovat v patřičných souvislostech. Zpravidla jsou vybaveni aparátem využívajícím metod statistiky a pravděpodobnosti, modelování, operačního výzkumu apod. Neobejdou se bez specializovaných softwarových nástrojů a u počítačů tráví hodiny denně.

Představují skupinu uživatelů, pro něž jsou určeny sofistikované nástroje pro statistickou analýzu, data mining apod.

- **Informatici**

Specialisté z oboru IT nejsou uživateli datových skladů v pravém slova smyslu. Získané informace většinou neslouží přímo jim, ale jsou poskytovány jiným subjektům. Jejich úkol spočívá jednak ve vytváření a provozování infrastruktury, jež umožní potřebné informace zpřístupnit ostatním uživatelům, a jednak jim poskytnout podporu pro realizaci konkrétních požadavků. Jsou mezi nimi specialisté na programové vybavení včetně databází, ale většinou nejsou s vlastními daty seznámeni po obsahové stránce tak dokonale, jako koncoví uživatelé nebo specializovaní analytici. Je proto pro ně důležité, aby vývoj, správa, struktury dat, ale i formální stránka celého procesu byla dostatečně přehledná, strukturovaná a srozumitelná. Vzhledem k tomu, že oni jsou těmi, kdo v praxi musí realizovat nejrozsáhlejší požadavky koncových uživatelů, je pro ně určující flexibilita daného systému, jež jim to umožní. Na druhé straně jsou to oni, kdo v průběhu návrhu, budování, ale i provozu datového skladu podstatným způsobem ovlivňují jak jeho obsah, tak možné způsoby komunikace ostatních - problémově orientovaných - uživatelů.

4.2 Hlavní akcenty při budování datových skladů

Pro splnění požadavků, umožňujících přístup, diskutovaný v předcházející části, je třeba zdůraznit následující základní směry:

4.2.1 Orientace na data

Vlastní data jsou tím nejcennějším, čím datový sklad disponuje. V okamžiku jeho budování již existovala a smyslem jejich využití novým způsobem je širší využití jejich informačního potenciálu. Způsob práce s nimi, ukládání, zpřístupňování považujeme za jednu z nejdůležitějších podmínek úspěšného budování a provozování datových skladů.

Nerespektování přirozené povahy dat v jejich plné šíři a různorodosti vede zpravidla k tomu, že jejich informační hodnota se zúží - z důvodu přizpůsobení jejich struktur konkrétním prezentačním nástrojům, nutnosti výrazně agregovat i data, jež je žádoucí mít k dispozici v detailní podobě, případně volbou agregačních kritérií, která se časem ukáží jako neodpovídající - nebo jednosměrným a uzavřeným způsobem prezentace, neumožňujícím

využívání jiných specializovaných nástrojů a případné rozšíření o nové možnosti, jež přinese další vývoj.

V tomto smyslu je zásadní pozornost třeba věnovat prostředkům pro ukládání a správu dat, jejichž nosným prvkem jsou specializované databázové systémy.

4.2.2 Orientace na uživatele

Druhým podstatným elementem, jenž hraje klíčovou roli v každém projektu, je uživatel. Představujeme si jej jako konkrétního člověka, majícího na jedné straně konkrétní požadavky a představy, na druhé straně disponuje vždy omezenými možnostmi. Cílem je pokrýt odpovídajícím způsobem potřeby všech tříd uživatelů tak, aby to bylo užitečné pro jejich vlastní práci. V praxi to znamená snahu využít všech nástrojů, jež se pro ně nabízejí, nebo které už využívají. I zde je smyslem respektování přirozené povahy všech uživatelů systému.

4.2.3 Proces inspirace a učení

Tento princip představuje jednu z hlavních myšlenek procesu budování a využívání datového skladu. Nejpřesnějšími a nejrelevantnějšími podklady pro formulování požadavků na funkce datového skladu jsou ty, jež jsou získány při jeho využívání.

Respektovat tento princip znamená využívat takových postupů a technologií, jež umožní využít na patřičné úrovni aktivní zkušenosti uživatelů ve všech etapách vývoje i provozu datového skladu.

Tento na první pohled paradoxní požadavek je možno splnit pouze při využití flexibilní technologie, nasazení prostředků různé třídy a určení a pružnou škálovatelnou architekturou.

5. Základní rysy metodologie SAFE/DW

5.1 Princip budování datových skladů

Příkladem respektování hlavních principů budování datových skladů v procesu inspirace a učení, popsaném výše, je etapovitá výstavba s využitím principu stavebních kamenů, realizovaných metodologií SAFE /DW. Základní myšlenkou tohoto přístupu je využití výstupů dílčích kroků jednotlivých etap jako analytického podkladu pro další etapy.

Každá etapa řešení představuje samostatný projekt (subprojekt), obsahující všechny základní kroky: business analýzu, návrh architektury, implementaci. Poté, co je dokončena, mohou koncoví uživatelé okamžitě začít datový sklad (ve smyslu skutečného skladiště informací, nikoliv např. OLAP aplikace) využívat. V praxi se ukázalo, že teprve první kroky uživatelů v novém terénu jim otevrou oči natolik, aby si uvědomili, co od tohoto systému mohou očekávat za informace. Teprve pak začne skutečná analýza jejich požadavků.

"Dej mi, co jsem ti řekl, že chci, a pak ti řeknu, co skutečně chci" - tak tento stav charakterizuje Bill Inmon ve své knize Building the Data Warehouse.

Ne nepodstatným aspektem tohoto přístupu je i finanční efekt. Vzhledem k faktu, že v praxi je úspěšně dovedena do konce jen menšina zahájených projektů datových skladů, je riziko promarnění vynaložených nákladů tím významnější, čím větší objem prostředků byl investován. Tedy, jinak řečeno, čím větší (rozuměj časově delší) projekt, tím větší ztráty způsobí jeho pozdější přizpůsobení a rozšíření potřebná pro zohlednění nových požadavků, vzniklých hlubším poznáním informačního potenciálu, případně změnami prostředí, jež nastaly v jeho průběhu. Návrat "ke kořenům" je v tomto případě mnohem bolestivější a

nákladnější, nehledě na to, že z hlediska poznání **skutečných požadavků** se často ocitáme znovu na prvotní startovní čáře, ovšem v podmínkách, kdy okolí (management) bude podstatně méně nakloněno dalším nákladným experimentům.

V případě inkrementálního postupu s možností zahájení uživatelského interaktivního dotazování co nejdříve, je aktivita koncových uživatelů velmi cennou součástí procesu analýzy systému (jakýsi "průzkum bojem"). Rozsah jednotlivých kroků je dán rozsahem dané etapy a především **prioritou**, jež je určující pro její realizaci.

		Implementace
		Architektura
	Implementace	
	Architektura	
Implementace		
Architektura		
Business analýza		

Obr.: Škálovatelný projekt DW

5.2 Princip priorit

Pro realizaci jednotlivých kroků řešení, ať pro centralizovaný datový sklad, nebo, obzvláště pro specializované Data Marty, je stanovení základních priorit stěžejní podmínkou jejich realizace. Tyto priority pak určují vlastní charakter konkrétního řešení, použitou technologii, vynaložené náklady, ale i jeho postavení v celkové struktuře systému.

6. Požadavky na softwarové řešení

Jestliže se podíváme na proces vývoje a návrhu datových skladů "technologickými očima", je zřejmé, že popsané postupy nemohou být realizovány, nejsou-li podepřeny odpovídající technologií. Vlastnosti softwarového řešení jako celku i jeho jednotlivých komponent, mají jinou důležitost a priority, než tomu může být u "klasických" OLTP systémů. Patří mezi ně zejména:

6.1 Flexibilita

Tato vlastnost je zde záměrně jmenována na prvním místě. Při obvyklé mnohotvarosti prostředí, požadavků a představ budoucího řešení musí být výsledný systém, ale i proces jeho vytváření, schopen reagovat nejen na různé požadavky na jeho funkčnost; musí rovněž umožňovat pružné změny struktur a logiky dílčích kroků. Tuto flexibilitu je možno blíže charakterizovat jako:

- **Schopnost využívat různých primárních dat** (tento požadavek je ovšem samozřejmý pro většinu koncepcí datových skladů)
- **Schopnost pracovat s různými formálními datovými strukturami** - Některé požadavky mohou vyžadovat data, jež nejsou upravena (např. denormalizována, optimalizována) do podoby, se kterou pracují „klasické“ warehouseové systémy. To může být dáno buď tím, že při řešení příslušných požadavků je např., z hlediska priorit na prvním místě rychlost uvedení dané aplikace do provozu (prostor pro náročnou analýzu a ladění datových struktur je omezen). Jiným důvodem může být skutečnost, že optimální strukturu nelze předem určit; zejména při zavádění nových funkcí a pohledů nemusí být k dispozici konkrétní požadavky, ze kterých by bylo možno předem stanovit, jaká bude finální optimalizovaná struktura systému, a bude nutno ji měnit postupně v závislosti na stupni poznání informačního potenciálu datového skladu a zkušeností s jeho využíváním.
- **Schopnost pracovat s různými koncovými (prezentačními) nástroji** - v podniku může být používána řada softwarových nástrojů - od tabulkových kalkulátorů MS Excell, přes prostředky jako MS Access až po řadu aplikací pracujících např. v režimu klient - server. Současně už mohou být testovány a (v pilotních projektech využívány) i specializované nástroje pro Data Warehouse. Velký zásah do struktury koncových nástrojů bude pravděpodobně znamenat masovější uplatnění přístupu přes intranet. Pro specifické funkce statistických analýz, dataminingu apod. mohou být zvoleny v různých etapách specializované nástroje, které budou vyžadovat více nebo méně přizpůsobené podmínky.
- **Schopnost reagovat pružně na nové typy požadavků** - pro některé skupiny činností v některých odděleních a útvech je důležitá možnost "okamžité analýzy na míru" při vzniku nových skutečností (např. vyhodnocování nových obecných trendů v marketingových agendách, odhad dopadu nových typů rizik atd.). Lze očekávat, že řada důležitých požadavků bude formulována teprve v procesu využívání datových skladů, tak, jak budou jeho účastníci (uživatelé, ale i správci a administrátoři) postupně odkrývat další obsahové a technické možnosti.

6.1.1 Snížení rizika

Realizace projektů datových skladů je ve světě i u nás problematikou relativně mladou a většina organizací, provozujících počítačové systémy, má s jejich budováním pouze omezené zkušenosti.

K této skutečnosti pak přistupuje fakt obecně menší možnosti vytváření cílených funkčních celků, než jak je tomu v případě "klasických" operativních systémů (určených pro OLTP - On line transakční zpracování). Riziko, jež vyplývá z použití značného objemu prostředků a kapacit pro aktivity, jež se ve svém důsledku mohou minout účinkem, je proto v tomto případě podstatně zřetelnější.

Z uvedeného vyplývá jako legitimní požadavek na použití metodiky a technologie, jež by umožňovala rizika tohoto typu minimalizovat.

6.1.2 Otevřenost

Otevřená koncepce systému datového skladu bude pravděpodobně představovat jeden ze stěžejních rysů projektu.

Představuje logický požadavek, vyplývající z různorodé datové a aplikační základny, jež je zdrojem všech informací pro Data Warehouse a z předpokládaných různých typů uživatelů a různých způsobů práce s datovým skladem.

Pro proces jeho vytváření však je kromě toho otevřenost důležitá ještě ze dvou hledisek:

- **Postupné modulární budování datového skladu**

Zkušenosti, jež získají tvůrci a uživatelé z prvních etap jeho budování a provozu, budou s největší pravděpodobností východiskem pro návrh a realizaci dalších etap. To bude vyžadovat schopnost začlenění nástrojů a postupů, jež v prvotních fázích nemusely mít vůbec uplatnění.

- **Uplatnění nových technologií**

Během vývoje a života projektu lze očekávat bouřlivý vývoj ve všech oblastech IT; datawarehousing bude zřejmě oborem, jenž během nejbližších let přinese nové poznatky, technologie a přístupy ještě ve větší míře. V tomto smyslu je otevřená koncepce jednotlivých komponent i projektu jako celku nutnou podmínkou, jestliže má uvažovaný systém držet krok s vývojem, případně nemá být zastaralý už při svém uvedení do provozu.

6.1.3 Rychlá implementace

Pro některé části systému nabude kritické povahy podmínka rychlého uvedení do provozu. Efekt z informací, získávaný z datových skladů může být závislý např. na rychlosti mobilizace příslušných zdrojů.

V tomto směru jde především o to, aby systém disponoval prostředky, jež umožní relativně snadno, s nepříliš velkými náklady a s minimálním rizikem vytvářet části datových skladů, jež začnou téměř okamžitě produkovat užitečné výsledky. Tuto možnost je třeba chápat jako součást vytváření koncepčně vyzrálého datového skladu, jež ovšem může být realizována paralelně. Ve svém důsledku pak mohou být jejich výsledky podkladem pro důkladnější a přesnější (správnější) analýzu skutečných rysů "komplexního" řešení.

7. Příklady různých typů přístupů z praxe

Potřeba takových nástrojů a metodiky, jež umožní co možná největší flexibilitu, danou tím, že dílčí, ale i konečný cíl, se budou vyvíjet a měnit (a mnohdy dramaticky) v závislosti na stupni poznání obsahu postupně budovaného datového skladu, může být dokumentována následujícími v praxi realizovanými příklady:

Příklad 1:

Prvním z nich je případ velké americké zásilkové společnosti, kde byl navržen a realizován datový sklad na základě klasického přístupu a odpovídající "tradiční" technologie relační databáze a OLAP nadstavby. Po jeho dokončení a uvedení do provozu byl tvořen strukturou několika multidimenzionálních matic, které, dle původní představy zadavatelů i analytiků, měly "spolehlivě pokrýt veškerou škálu možných dotazů". Během krátké doby začala společnost zaměstnávat 3 analytiku, jejichž jedinou pracovní náplní bylo vytvářet nové a nové datové modely, které by umožnily pokrýt poptávku uživatelů po nových dotazech. Po

několika měsících, během kterých takto vznikla téměř stovka multidimenzionálních modelů, se stala situace nezvladatelnou, projekt byl přerušen a stávající systém byl nahrazen novou technologií, umožňujícími flexibilní přístup. Systém je od té doby spravován jedním databázovým administrátorem. Přitom umožňuje téměř okamžitě realizovat jakékoliv nové a nepredikovatelné požadavky.

Příklad 2:

Velká tuzemská státní finanční organizace byla postavena před problémem v podstatě ze dne na den umožnit získávat odpovědi na velmi nespécifické a předem nepredikovatelné dotazy nad naše poměry velmi rozsáhlými daty. Během několika týdnů byl uveden do provozu datový sklad se 150 GB **čistých dat**, jež byl okamžitě využíván především relativně jednoduchými nástroji pro tvorbu dotazů a klient-server aplikacemi a připojenými tabulkovými procesory. Po zhruba jeden a půl ročním provozování, vytrání uživatelů a hlavně poznání informačního potenciálu datového skladu bylo na základě podstatně lépe kvalifikovaných požadavků přistoupeno k budování analytické nadstavby na bázi osvědčené technologie OLAP.

Příklad 3:

Soukromá česká organizace, operující rovněž ve finančním sektoru se rovněž rozhodla pro budování datového skladu, dokonce používala touž technologii jako v předchozím případě. Celkový projekt byl ale koncepčně, technologicky i dodavatelsky pojat jako "klasický" projekt informačního systému na klíč, založeného především na využití OLAP nástrojů. Po několikaměsíční usilovné analýze a programování stále nebyly k dispozici žádné výsledky, projekt se dostal do skluzu, a hlavně: zadavatel zjistil, že během této doby velmi akutně vyvstala potřeba informací, se kterými se v původním zadání nepočítalo. Dodavatel řešení nebyl schopen (ani ochoten) vzhledem k celkové koncepci projektu nové požadavky realizovat. Datový sklad přestal uživatelům plně sloužit ještě před tím, než byl uveden do provozu. Náprava tohoto stavu se stala předmětem dalšího koncepčního rozhodování o architektuře celého řešení.

8. Závěr

Systémy pro podporu rozhodování, manažerské informační systémy, datové sklady, business intelligence, datamarty ... tyto a ještě další pojmy jsou v současnosti skloňovány mnohými ústy, ačkoliv ještě nedávno byly diskutovány spíše jen v odborné literatuře a byly chápány jako záležitost spíše okrajová. Jejich právo na existenci je dnes už zřejmé. Dokonce jsou u nás i ve světě doloženy mnohé pokusy o jejich zdařilou, ale i méně úspěšnou realizaci.

Zkušenosti, získané při jejich budování orientují kroky jejich dnešních tvůrců poněkud jiným směrem, než tomu bylo v minulosti. Snaha řešit problémy spíše vtípem a elegancí než nasazením hrubé síly, je dána zcela jednoznačně skutečností, že počítačové systémy nejsou už chápány jako nenasytná božstva, jimž je třeba se kořit a přinášet oběti. Ve středu pozornosti není počítač, ani operační systém, databáze nebo jakákoliv jiná technologie sama o sobě. Ústřední postavou v tomto ději je člověk, jenž má právo požadovat, aby mu systém sloužil, nikoliv vládl. Aby se mohl ptát a získávat odpovědi na to, co ho zajímá a co potřebuje vědět.

Má právo chtít, aby s ním systém komunikoval způsobem, který mu vyhovuje, ať je povoláním manager, analytik informatik nebo účetní. A konečně, má právo se zajímat o to, jak dlouho bude trvat, než mu systém začne poskytovat nějaké výsledky a v neposlední řadě - kolik to všechno bude stát.

Jednoduše řečeno, budování datového skladu vyžaduje spíše než masivní nasazování drahého hardwaru, softwaru, lidské síly a umu přístup, který ukazuje cestu, vedoucí k cíli.