

## **Na co nezapomenout v projektu budování datového skladu**

Iva Valečková, doplnil Marcel Špeta

Základní pravdy o implementaci řešení Business Intelligence jsou v dnešní době většinou již ve všeobecném podvědomí IT pracovníků i vedení společností. Až při samotných realizacích projektů může docházet k problémům, které ovlivňují jejich celkový úspěch a vnímání (skluz harmonogramu, nenaplněná očekávání zadavatele, nepochopení zadání, ...). Ukažme si u jednotlivých kroků realizace projektu datového skladu riziková místa a zkusme tak předcházet v maximální možné míře těmto nepříjemným situacím.

Většina firemních systémů obsahuje velké množství dat, ze kterých se relevantní informace získávají jen obtížně. Cestou ke zhodnocení a efektivnímu využití těchto dat je nasazení nástrojů Business Intelligence (dále BI). BI je cesta ke konkurenční výhodě.

### **Definice Business Intelligence**

Pod označením Business Intelligence rozumíme především výkonné analytické nástroje, které umožňují využít firemní data nejen k analýze již proběhlých jevů, ale také k predikcím budoucího vývoje. Nedílnou součástí BI je i sjednocená datová základna, odkud lze data čerpat, tedy datový sklad.

Před samotným zahájením projektu implementace BI je nutné sjednotit „slovníček pojmů“ zadavatele a dodavatele. Dále je vhodné nastavit u zadavatele správná očekávání vzhledem k přínosům BI. Někdy bývá totiž nasazení BI vnímáno jako spásný proces, který sám o sobě připraví potřebná data v provozních systémech či „narovná“ podnikové procesy.

Základním předpokladem pro správnou funkci a využití datového skladu jsou korektní a pravdivá data již na úrovni provozních systémů (například chyby v evidenci zákazníků se pouhým načtením do datového skladu neopraví). Datový sklad většinou slouží pro integraci a následné využití dat z provozních systémů. Poměrně často datový sklad také poskytuje sdílená data provozním systémům (např. platnou verzi číselníků).

### **Co je datový sklad**

Centralizovaná databáze, kde jsou data uložena ve speciálních datových strukturách, včetně historie. Součástí datového skladu jsou i ETL procesy, které zajišťují plnění datového skladu, případně zajištění čistících mechanismů. Data jsou v datovém skladu uložena vyčištěná a konsolidovaná.

Z globálního pohledu bývá někdy datový sklad vnímán také jako cesta k systémové integraci.

Datový sklad funguje jako základ systémové integrace jen do určité míry. Z manažerského pohledu se opravdu jedná o informační integraci, z pohledu provozu se o systémové integraci hovořit příliš nemůže, datový sklad neřeší a ani nemá řešit vazby mezi provozními systémy navzájem (výjimkou mohou být např. již zmiňované číselníky).

### **Datový sklad ano či ne?**

Jsou-li o potřebnosti a smysluplnosti implementace datového skladu (či celého BI řešení) nějaké pochybnosti, obzvláště u managementu, nebo pokud není jasno v tom, co se dá v průběhu implementace zvládnout vlastními silami a na co přizvat specialisty, pak je vhodné vypracovat Studii připravenosti na datový sklad.

Cílem studie je posouzení připravenosti společnosti na projekt datového skladu. Jedná se o identifikaci silných a slabých stránek společnosti ve vztahu k projektu datového skladu, včetně potencionálních rizik a dále o ohodnocení připravenosti předpokládaných oblastí k realizaci, včetně komentářů a doporučení z hodnocení vyplývajících.

### **Analýza**

Základním krokem každého projektu je analýza prostředí, jejímž hlavním cílem je definovat oblasti použití a funkce systému v dané společnosti. Výstupem této analýzy je návrh řešení, včetně doporučení, jak provádět implementaci. Součástí návrhu řešení je také doporučení vhodné technologie pro implementaci datového skladu, která zohledňuje stávající HW a SW prostředí společnosti, požadavky uživatelů a kapacitní náročnost zpracovávaných úloh.

Nezbytným předpokladem úspěšné implementace datového skladu (obecněji BI systému) je podpora ze strany vedení. Podobně jako implementace jakéhokoli jiného systému představuje zátěž navíc k běžným pracovním povinnostem pro klíčové uživatele. Možné kapacity nasazení klíčových pracovníků na realizaci projektu jsou v průběhu meetingů zjišťovány a zapracovány do analytického dokumentu. Nejintenzivnější spolupráce s klíčovými uživateli je očekávána v kroku analýzy během meetingů a při akceptaci výstupních dokumentů a také při testování a ověřování správnosti implementovaného řešení.

### **Datový sklad není smetiště**

V průběhu analytických prací jsou mapovány uživatelské požadavky na systém a k nim odpovídající a dostupné zdroje dat. Při jednáních s klíčovými uživateli je běžné, že na dotaz, jaká data skutečně potřebují v datovém skladu, se analytikům a architektům datového skladu dostává odpovědi: "Raději tam dejte všechno, aby to tam později nechybělo." Při návrhu datového skladu je nutné dbát na efektivnost návrhu tak, aby data nebyla v datovém skladu ukládána redundantně. Data v datovém skladu by měla být ukládána s nějakým konkrétním záměrem. Data, pro která není v danou chvíli zřejmé využití, mohou být do datového skladu přidána později.

### **Přístupy k realizaci**

Celý projekt vytvoření datového skladu lze řešit dvojím způsobem:

- **„velký třesk“** - nejprve jsou provedeny detailní analýzy všech požadavků uživatelů a k nim patřících zdrojů dat, je vypracován souhrnný návrh řešení a poté probíhá implementace. U velkých projektů může fáze realizace trvat i více než rok.

#### **Výhody**

- od začátku je jasná datová struktura celého datového skladu

#### **Nevýhody**

- dlouhá doba od zadání po předání do provozu - zadavatel může mít pocit, že se dlouho nic neděje a ztratí o dokončení datového skladu zájem  
- možné změny zdrojů dat a požadavků ještě před ukončením implementace a předáním do ověřovacího provozu  
- komplikované zohlednění změny priorit požadavků v průběhu implementace  
- velké riziko neúspěšnosti projektu

- **„inkrementální způsob“** - v průběhu úvodní analýzy jsou určeny jednotlivé etapy a to podle priorit řešení požadavků, každá z etap by neměla trvat déle než 3-4 měsíce. Etapou se v tomto smyslu rozumí sled kroků detailní analýza, návrh řešení, implementace, testování, školení, ověřovací provoz, předání do provozu. Kroku školení je věnován samostatný odstavec níže.

#### **Výhody**

- zadavatel může v krátkém časovém úseku využívat výstupy pro další rozhodování - návratnost investic, možné i samofinancování dalších etap ze vzniklých úspor  
- pořadí etap může být upraveno dle aktuálních potřeb společnosti  
- snazší reakce na změny datových zdrojů v průběhu projektu

#### **Nevýhody**

- možné komplikace při změně datové struktury - zásadní úpravy stávajících a již používaných datových struktur, doplnění historických dat

V současné době se téměř zcela ustoupilo od realizace datového skladu celého najednou a implementace se realizují inkrementálním způsobem.

I u inkrementálního způsobu existuje riziko přílišného nárůstu objemu prací v rámci etapy, když se nepodaří určit priority požadavků a „všichni chtějí všechno v prvním inkrementu“. Pak musí nastoupit řídicí moc vedení projektu a striktně určit úlohy, které budou v etapě řešeny.

## **Další využití dat v datovém skladu**

Data v datovém skladu lze rozdělit na dvě části - vlastní data a metadata (data o datech). Při návrhu využití informací uložených v datovém skladu jsou metadata neprávem opomíjena.

Příkladem využití metadat je importem vytvořená definice rozhraní pro reporting (přinese ušetření nákladů na implementaci), dalším příkladem může být dokumentace přístupových práv (přístupová práva jsou zdokumentována podle skutečnosti, nikoli podle představ) nebo distribuce jednotných metadat do provozních systémů.

## **Životní cyklus datového skladu**

Při implementacích řešení BI bývá jedním z problematických míst nenaplnění očekávání sponzora projektu. Implementaci BI nelze srovnávat s implementací provozního systému, kdy mnohdy stačí systém nainstalovat, vyškolení uživatele a pak jen vkládat data a používat existující procesy a výstupy, tedy dále neinvestovat a to do té doby, dokud je provozní systém dostačující.

Podíváme-li se ze stejného úhlu pohledu na řešení BI, najdeme podstatný rozdíl. Ze zkušenosti z řady projektů vyplývá, že tvorba datového skladu a využívání dat uložených v něm je proces nastartovaný při zahájení projektu, který dále pokračuje a mění se spolu s firmou. Datový sklad musí reagovat například na následující změny:

- změny ve zdrojových systémech (akvizice společnosti jinou společností vyvolává potřebu integrace nových dat)
- nové požadavky uživatelů (poptávka po nových reportech obohacených o další data je nekončící proces)
- vzhledem k nárůstu objemu dat dojde ke zpomalení DW (je nutné udělat optimalizaci ETL procesů, nastavení databáze, či zvážit posílení řešení již na úrovni HW)

Je velmi důležité ověřit na začátku projektu porozumění sponzora projektu těmto faktům.

## **Dokumentace**

Nedílnou součástí projektu a přesto občas opomíjenou či zanedbávanou je dokumentace, přestože provází všechny etapy projektu. Vedle standardní projektové dokumentace je vhodné již ve stádiu analýzy stanovit metodiky pro jmenné konvence a měly by být stanoveny i další postupy a principy, například principy vedení projektu, formuláře pro řízení změn a další. Tato dokumentace, přestože je někdy vnímána jako „papírování“ a zátěž pro členy realizačního týmu, se „vyplatí“ při řešení krizových momentů projektu, kdy například může odejít klíčová osoba realizačního týmu ze strany zadavatele a projekt by mohl nabrat zpoždění oproti harmonogramu.

Další skupinou je technická dokumentace řešení - definice datové struktury, ETL procesy, dokumentace k instalacím (přístupové adresy, přístupová hesla a podobně).

Poslední skupina dokumentace je určena pro uživatele - zde je vhodné si určit formu dokumentace (elektronická, tištěná) a hlavně jazyk, v jakém bude dokumentace pro uživatele dostupná. Většina dokumentací k produktům nadnárodních společností je k dispozici pouze v elektronické podobě (formou helpu) a v angličtině, s čímž mohou mít někteří koncoví uživatelé problém a neradi s novým nástrojem pracují. Také tento fakt druhotně ovlivňuje vnímání celé implementace - jedním z cílů by měl být také spokojený a poučený uživatel.

## **Zaškolení uživatelů**

Součástí projektu má být i zaškolení koncových uživatelů. Při implementaci se zaškolují také administrátoři a správci jednotlivých systémů.

Je vhodné ještě před zahájením projektu pro skupinu klíčových uživatelů udělat prezentaci možností a přínosů implementovaného řešení a ukázat práci s koncovými nástroji. Tato investice času se vrátí formou lepší a přesnější definice požadavků klíčových uživatelů.

Obecnou pravdou je, že se investice do zaškolení uživatelů omezuje na nejmenší možnou míru. Na certifikovaná školení jsou posíláni vybraní jedinci, kteří pak mají šířit osvětu o produktu dále do firmy. Na co už se sponzor projektu neohlíží je velmi podstatný fakt, že ne každý (byť je skvělým odborníkem a specialistou) má schopnost dále předat srozumitelnou formou získané vědomosti. Dalším rizikem je přirozený výběr informací každého jedince - co je důležité pro jednoho posluchače nemusí být důležité pro druhého posluchače, každý z účastníků si odnáší ze školení jiný zážitek.

Kvalitní zaškolení uživatelů má velký vliv na vnímání výsledku projektu a na práci s nástroji pro koncové uživatele.

V praxi se osvědčilo použití tzv. superusers tj. vybraní jedinci z každého podnikového oddělení (profesní oblasti), kteří projdou speciálním školením, jak vytvářet ad-hoc reporty a mající vyšší oprávnění v DW. Výhodou je, že daný jedinec výborně rozumí podnikové problematice a je tak podstatně rychleji schopen přetransformovat požadavky oddělení do jim srozumitelného reportu či dashboardu. Navíc tato osoba slouží jako styčný důstojník pro BI specialistu a designera

#### *Inkrementální vývoj DW*

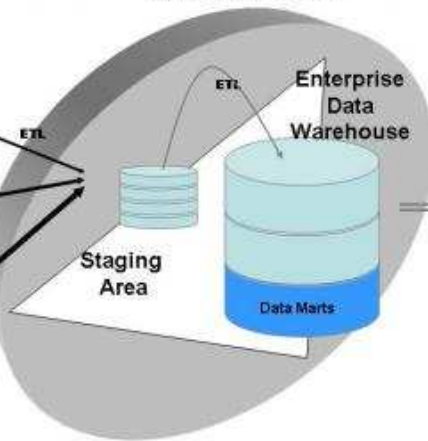


#### *Schéma datového skladu*

**Zdrojové systémy**



**Data Warehouse**



**Využití dat**

