

Operační datový sklad nabízí integraci znalostí

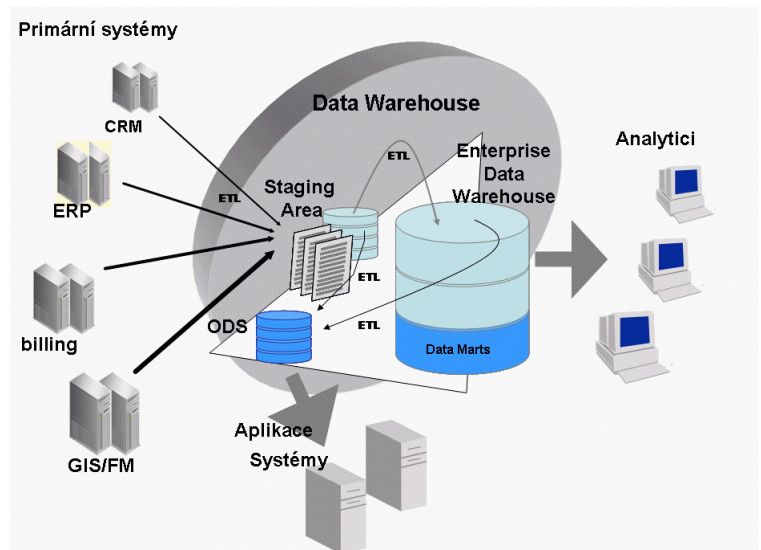
Operační datový sklad (ODS) je jednou z možných součástí obecného datového skladu (DW), součástí se specifickým posláním. V plánech podniků budoucích DW je s umístěním ODS v architektuře systémů málokdy počítáno. Neprávem.

Datový sklad je vnímán zejména analyticky

Role datového skladu (DW, z angl. Data Warehouse) v podniku je tradičně chápána jako analytická. Dostupnost takové analytické funkcionality, která by během analytikova konání nezatěžovala primární podnikové systémy, je jedním ze základních paradigmat Business Intelligence (BI).

Příprava a integrace dat tak, aby dávaly komplexnější obraz analyzovaného subjektu, je ve smyslu BI pouček krokem, který "mění data v informaci". Údaje o zákazníkovi, nastavení jednotlivých služeb, jejich objemy poskytnuté v určitém časovém období, vystavené faktury, uskutečněné či dlužné platby, zákaznickovy preference... To vše, či ještě další ukazatele, umístěné v čase, představují příklady dat roztržitých jednak v systémech specifických pro určitý druh podnikání, dále pak v systémech CRM, ERP a dalších.

Specifickým systémem mohou být například zákaznické systémy bank, poskytovatelů energií a užitit, systémy telekomunikační, logistické, geograficky orientované nástroje pro správu technické infrastruktury...



Vyjmenoval jsem zde několik příkladů, ale není důvod diskvalifikovat žádný další druh systému, v praxi rozhoduje odhodlání nacházet odpovědi na otázky.

Odměnou za investice a úsilí věnované integraci dat ze všech systémů je pak analytická výtěžnost těchto dat. Tu lze následně vyjádřit i finančně, ziskem či úsporou, která vznikla na základě analýzy, a tudíž kvalifikovanějšího rozhodnutí.

Datový sklad zná svoje konvenční členění

Připomeňme členění datového skladu z pohledu tradiční architektury. *Staging Area* (SA) je logická oblast, která je čekárnou pro nezpracovaná data. Fyzicky je realizována například soubory, které jsou doručeny ze systémů výše zmíněných, primárních. Relační databáze je další možná podoba SA, kombinace obojího je téměř pravidlem. Předání dat mezi podnikovým systémem a datovým skladem může mít ovšem i řadu dalších technických variant.

Enterprise Data Warehouse (EDW) je samotné analytické srdce datového skladu, obsahuje zpracovaná data zušlechťená ve smyslu jejich vlastností a předností zmíněných výše. Data v EDW mají nezanedbatelnou historii a jsou po nahrání do DW dále neměnná.

Extract-Transform-Load (ETL) je komponenta procesní (ovšem i ta má svoje hmatatelné nástroje) s překryvem nad všemi částmi DW. Extrakce z primárních systémů, přesun dat do datového skladu, nadefinované transformace, čištění a obohacení dat, load do EDW. Toliko alespoň z těch nejdůležitějších funkcí.

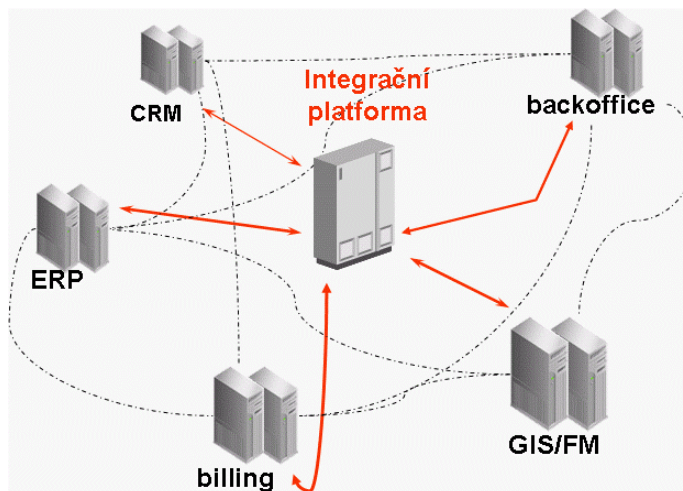
Konečně *ODS* je oblast dat, které mají zanedbatelnou historii a maximální možnou aktuálnost. V procesu budování datového skladu v "bodě nula" není s možným přínosem ODS obvykle vůbec počítáno. Což je škoda. Proto se nad funkcí ODS zamyslíme více.

Malá odbočka - integrační platformy

Hitem zejména posledního období v integraci podnikových aplikací jsou vyhrazené integrační platformy (IP). Z uspořádání podle obrázku plynou zřejmě výhody, integrované systémy jsou propojené centrálně ve hvězdicové architektuře a nikoliv nežádoucími (na budování a údržbu nákladnými) vazbami m:n. Propojení dvou systémů přes centrální bod probíhá buďto komunikací na bázi "odpověď na dotaz", nebo technikou broadcaster-consumer, kdy systém přes IP posílá zprávy konzumentským systémům. Z druhé vlastnosti může těžit i DW, pokud je proces extrakce z podnikového systému do DW realizován zasíláním změn prostřednictvím IP. Kde jsou slabiny této techniky integrace?

ODS znamená integraci znalostí

V módu dotaz-odpověď se totiž IP (samozřejmě zástupně za jiné systémy) chová v důsledku obdobně nežádoucím způsobem jako případný analytik, který zatěžuje svými dotazy i takový podnikový systém, který na takový druh zátěže není dimenzovaný. Přitom nechci být podezírán ze snahy prezentovat jakýkoliv podnikový systém jako choulostivý na dotazy z okolního prostředí. Připusťme jen, že vyhodnocení jednoduše formulovaných dotazů může být netriviální. Pro vysvětlení použijme opět příkladu: Zákazník vznáší dotaz na call centrum poskytovatele a žádá o rozšíření aktivace služeb. Podmínkou aktivace je uhrazení všech závazků, strategie získání informace jsou dvě. Možná (ale ne nezbytná) je u obou strategií zprostředkující funkce IP.

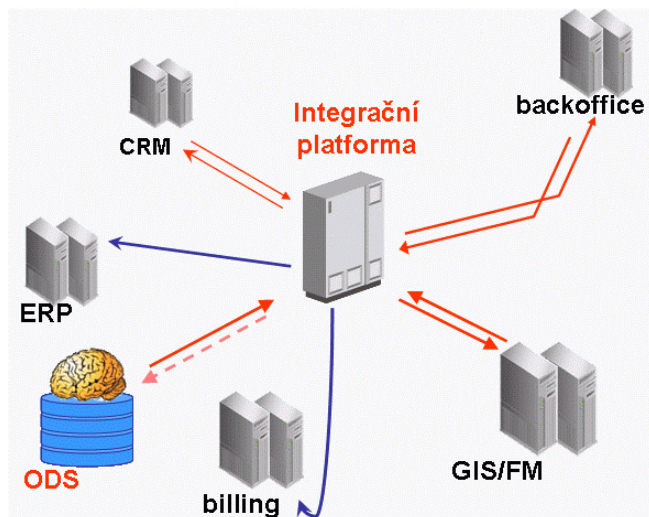


První z cest je dotaz do ERP systému, který ve svém důsledku může znamenat vyhledání všech zákazníkovi již vystavených faktur, párování faktur a plateb, a další akce. (Není to již de facto analytický dotaz?)

Druhou možnou strategií je dotaz do ODS, kde je odpověď na otázku připravena přímo ve formě příznaku: "Tento zákazník nemá žádný dluh".

Informačně zralý podnik s výhodou používá strategii druhou - ODS. Od úzkého okruhu charakterem jednoduchých otázek a odpovědí je potom cesta k sofistikovanějším: Měl zákazník v minulosti potíže s placením? Nakolik je bonitní (credit a risk skóre)? Chová se tak, jako by chtěl odejít? Příklady opět doufám potvrzují potřebu maximální aktuálnosti dat, některá se získávají snáze, jiné (např. scoringy) jsou výstupem BI technik - dolování v datech a modelů, aplikovaných nad EDW částí datového skladu.

Integrační platforma je nástroj, který má nesporné komunikační dovednosti. Schopnost poskytovat odpovědi na předem známý okruh otázek ovšem nemá, tuto způsobilost nese ODS. Kvalitu do systémové integrace pak přináší kombinace obou.



Podnik, který "zapíná" ve svém datovém skladu přihrádku ODS, využívá data doručená do DW více než jen analyticky pro rozhodování managementu. Znalost získaná z dat se prostřednictvím ODS nabízí k využití všemi podnikovými systémy. Ku prospěchu jejich efektivního běhu. ODS může dalším IS komponentám a aplikacím podniku nabídnout k využití výsledky z tak pokročilých analýz, že je ani jiné operativní systémy nabídnout nemohou. V datovém skladu v módu "ODS enabled" více než jinde platí, že informace musí být aktuální a přesná. Nutným předpokladem je proto kázeň podepřená metodikou a důrazem na kvalitu dat.

Závěr

DW je tradičně vnímán jako analytické prostředí, ale druhá z jeho rolí nemusí být svým dosahem pouze druhotná. Operační datový sklad je jako stvořen pro to, aby se v moderním informačním systému doplňoval s integračními nástroji a sloužil jim v roli "cache" znalostí. Slabinu integračních nástrojů, totiž chybějící schopnost poskytnout odpověď, aniž by se dále ptaly, odhalují i producenti ETL technologií. Snaží se posunout svoje nástroje od dávkových k takovým řešením schopným z on-line dat dodat znalost.

Cílem tohoto článku je roli ODS popularizovat, ovšem v mezích serióznosti. Proto ještě položíme a zodpovíme otázku, jaké jsou typické důvody pro implementaci ODS:

1. Ve hře je integrace mnoha systémů, z níž pocházejí vazby s charakterem výrazně m:n; integrací putuje druhově omezený okruh, ale velký objem stejných dat.
2. Podnik je natolik zralý, že usiluje o zvýšení svojí konkurenceschopnosti zapojením business intelligence výstupů do podnikových aplikací, např. scoringy, customer value apod.

3. Alternativně - klíčové systémy podniku nejsou způsobilé pro zátěž plynoucí z přímé integrace, přitom není strategicky (tj. v praxi politicky) únosné tyto systémy obměnit. Tento případ je zajímavý i pro společnost, která s datovým skladem začíná.