

Datové sklady a integrace podnikových aplikací

Jaromír Habáň

Centrum pro výzkum informačních systémů – odborná sekce ČSSI

jaromir.haban@cvis.cz

Abstrakt

Integrace podnikových aplikací patří ke stěžejním součástem systémové integrace. Možnosti integrace podnikových aplikací se vyvíjejí řadu let společně s rozvojem nových technologií a změnami požadavků zákaznických organizací. Z technologického hlediska zasáhl do tohoto procesu nevýrazněji nástup osobních počítačů, zavádění architektury klient/server a využití internetu v podnikové sféře. Požadavky zákazníků se měnily společně s rostoucí konkurencí, globalizací trhu a rozvojem podnikání v rámci dodavatelských řetězců. Pro řešení integrace je dnes k dispozici řada přístupů a softwarových nástrojů. Jak při integračním procesu mohou pomoci datové sklady?

1. Datové sklady

1.1 Jak definujeme datový sklad?

Datový sklad (data warehouse) můžeme charakterizovat jako jednotné konsolidované úložiště podnikových dat, které je optimalizované pro poskytování rychlých odpovědí na obchodní dotazy. Datový sklad přináší také „pořádek v podnikových datech“ s možností jejich analytického zpracování. Nemůže však být chápán pouze jako nástroj analýzy, který poskytuje uživatelům souhrnný pohled na data vyprodukovaná jejich informačními systémy. Ve skutečnosti jde o komplexní nikdy nekončící proces, v jehož průběhu je třeba příslušná data transformovat z operativních zdrojů, očistit, uložit do odpovídajících struktur a zajistit jejich doručení uživatelům ve struktuře, formě a čase, který by byl pro ně užitečný. Jeho stěžejním úkolem je podpora plánování a řízení firmy, a to nejen na strategické úrovni. Datový sklad získává svou obsahovou náplň především z tzv. OLTP systémů, tedy z ERP, CRM či APS řešení. Konkrétními interními datovými zdroji využitelnými pro analýzy v datových skladech jsou databáze zákazníků nebo databáze historií, vytvářené zásobami, fakturováním, dodávkami a obchodními systémy. Využívá také externí data, a to v podobě katalogů, telefonních a jiných seznamů, obchodních adresářů apod.

1.2 ETL nástroje

O plnění dat do databáze se starají datové pumpy, nebo-li ETL (Extraction, Transaction, Loading) nástroje. Tyto nástroje získávají data ze vzájemně nekompatibilních zdrojů, následně je transformují do nových odpovídajících struktur a ukládají do databáze datového skladu. Data jsou tak připravena k pozdějším analýzám.

Datové pumpy pracují v dávkovém režimu, přenášejí tedy data v určitých časových intervalech (dny, týdny měsíce). ETL nástroje mají také za úkol „čištění“ dat. Kvalita ETL nástrojů tedy přímo ovlivňuje kvalitu dat uložených v datovém skladu a tím pádem také kvalitu informací, které můžeme z něj získat.

1.3 Jiný pohled na data

Při detailnějším pohledu na OLTP systémy narazíme na další podstatnou odlišnost datových skladů od běžné podnikové databáze. Zatímco databázová platforma OLTP je optimalizována pro zpracování velkého množství malých transakcí a uchovává pouze aktuální data, OLAP nástroje analyzují historická data obsažená v datovém skladu. Z nich jsou pak schopny vytvářet obsáhlé statistické sestavy prostřednictvím kladení složitých dotazů. Databáze využívané OLAP nástroji tvoří stabilní strukturu, která je na rozdíl od klasické podnikové databáze neustále rozšiřována, a to bez jakékoliv redukce jejího obsahu. V databázi datového skladu jsou uložena především historická data. Analýzy dat pak na základě historického srovnávání nabývají smysl. Díky systematickému přístupu k budování datového skladu lze také nalézat vztahy mezi daty, které byly získány z různých zdrojů a které by jinak uživatel pravděpodobně nikdy neodhalil.

Datový sklad obsahuje data na různých úrovních podrobnosti. Ta bezprostředně ovlivňuje podrobnost získaných informací. S narůstajícím množstvím dat se zvyšují požadavky na stabilitu výkonu, administraci a škálovatelnost databáze.

Problémem datových skladů bývá aktuálnost získávaných informací, a také jejich bezprostřední dostupnost. Pro EAI platformy je naopak zabezpečení dostupnosti dat v reálném čase kritickým úkolem. Proto také není možné přímo využít datových skladů jako integračního prvku EAI platformy.

1.4 Výkonnost datového skladu

Jako každý jiný informační systém i datový sklad je charakterizován výkonností. Čím podrobnější data sklad obhospodařuje, tím více se snižuje jeho výkonnost. Data je tedy nutno agregovat do formy definovaných číselných ukazatelů. Takovýmto ukazatelem může být např. věkový průměr osob ve zkoumaném souboru. Výpočty těchto ukazatelů se provádějí během úvodních agregací atomizovaných dat, přičemž jejich zdrojem jsou relační databáze vlastního datového skladu. Takto agregovaná data je pak možno využít k zjednodušenému „typizovanému“ dotazování, které nebude snižovat celkový výkon zbytečným opakováním složitých dotazovacích operací.

Celkový výkon datového skladu významně ovlivňuje také proces agregace dat pro všechny odvozené sklady, tzv. datová tržiště (Data Marts). Tento proces totiž probíhá opakovaně po každé aktualizaci relační databáze hlavního datového skladu. Datová tržiště slouží ke specializovaným analýzám obohacujícím data o hodnotu (např. ziskovost zákazníků), která má význam při rozhodování o dílčích strategiích firemního byznysu (zaměření na klíčové skupiny zákazníků). Samotná datová tržiště tvoří autonomní jednotky obsahující odlišná data a využívající jiné druhy algoritmů pro analýzy. Data je tedy třeba agregovat pro každé tržiště zvlášť. S růstem počtu jednotlivých tržišť pak samozřejmě klesá i celkový výkon (roste časová náročnost na zpracování) datového skladu.

Nesmíme ovšem zapomínat na údržbu těchto rozdílných nástrojů, které v podstatě sdílejí stejná data (přidání atributu v OLTP systému se projeví v datové pumpě, databázi, OLAP nástrojích i v datových tržištích). Pokud mají jiné aplikace využívat služeb datového skladu, musí být přesně popsáno, co přesně datový sklad obsahuje. Každý z těchto nástrojů obsahuje vlastní způsob popisu dat (databáze - systémový katalog, ETL nástroje - popisy vstupních a výstupních polí a popisy transformací, OLAP nástroje - popisy dimenzí a faktů), tím jsou metadata.

2. Integrace podnikových aplikací

Integraci podnikových aplikací (EAI) můžeme charakterizovat jako propojení původně nezávislých dílčích řešení či informačních systémů, které jsou vzájemně nekompatibilní a jejichž správa a údržba probíhá nezávisle. EAI jako jednotná aplikační platforma pak logicky musí obsahovat soubor nástrojů a technologií, které umožní kompaktní správu a efektivní spolupráci doposud nezávislých aplikací.

Může nám EAI nabídnout i další služby? Nejde přece jenom o náklady na implementaci, provoz a údržbu, trápí nás často celková efektivita zavádění nových aplikací a procesů. Nedokážeme posoudit, jak jsou tyto procesy monitorovány a kolik na ně vynakládáme lidských a systémových zdrojů. Chybí nám konkrétní informace o tom, jak dlouho např. trvá vyřízení obchodního případu. A pokud bychom takovouto informaci získali, dokážeme rychle dané procesy změnit podle aktuálních požadavků?

2.1 Point-to point

Začněme ale nejprve u způsobů integrace podnikových aplikací. Existují v podstatě dva. Ten první, se nazývá „point-to-point“ – každý s každým. Spočívá ve vytváření účelových propojovacích řešení zabezpečujících synchronní komunikaci mezi aplikacemi. Tento přístup je v praxi doposud hojně využíván, avšak není příliš efektivní. Generuje nadbytečné náklady na vývoj každého jednotlivého komunikačního řešení a předpokládá značnou pracnost s rozdílnými komunikačními protokoly či zdrojovými kódy nestandardních aplikací. Dále také počítá s připraveností aplikací na synchronní komunikaci. Pokud však aplikace dostupná není, transakce nemůže být provedena.

2.2 Jak funguje EAI řešení?

EAI platforma umožňuje uskutečnit propojení na principu middlewaru. Slouží jako sdílené místo, na něž jsou zasílána data z jednotlivých aplikací a kde jsou také transformována a předána ke zpracování ve srozumitelném formátu. EAI řešení zabezpečuje přenos dat asynchronně. Data jsou totiž umístěna „ve frontě“ middlewaru. Příslušná transakce se pak může uskutečnit bez nároku na okamžité propojení zdrojové a cílové aplikace.

Samotná integrace prostřednictvím EAI platformy pak může probíhat na několika úrovních – datové, uživatelského a aplikačního rozhraní a obchodní logiky. Využití integrace na konkrétní úrovni by pak mělo být odrazem potřeb zákaznické organizace a vlastností integrovaných aplikací.

2.2.1 Propojení přes data

Integrace aplikací na datové úrovni využívá úložišť dat a možností jejich přesunu. Přitom v podstatě ignoruje existenci aplikací jako takových. Tento přístup nemusí být na škodu, neboť většina aplikací funguje obdobně – ukládá data do relačních databází. Data z jedné databáze jsou pak za pomoci ETL nástrojů automaticky čištěna, transformována a uložena v požadované struktuře do databáze sloužící jako datové úložiště jiné aplikace. Jde o docela dobré a rychlé řešení s jasně čitelnými přínosy spočívající zejména v úspoře nákladů v porovnání s ostatními způsoby integrace. Hodí se také k budování operačních datových skladů (ODS). Může však narazit na obtíže s udržováním velkého množství databázových tabulek. Pokud od EAI požadujeme vysokou flexibilitu a schopnost práce v reálném čase, pak integrace na datové úrovni není příliš šťastné řešení.

2.2.2 Uživatelské rozhraní a portál

Zatímco „datový přístup“ ignoroval existující aplikace, propojení přes uživatelské rozhraní jde přesně opačným směrem. Snaží se automatizovat práci uživatele a to tak, že odpovědi systému na vznesený požadavek jsou výsledkem simulací uživatelského postupu. Jedná se o způsob využívaný především tehdy, kdy je potřeba integrovat velmi staré aplikace, u nichž není k dispozici zdrojový kód, a kdy nelze využít jiného integračního postupu. V jiných případech by ale toto řešení bylo velmi neefektivní. Je totiž závislé na neměnnosti aplikace, s čímž můžeme u modernějšího softwaru jen velmi těžko počítat.

K integraci přes uživatelské rozhraní lze jistě zařadit také moderní trend nasazování portálových technologií – jednu z klíčových součástí EAI řešení. Portál však neslouží primárně k tomu, aby zabezpečoval výměnu dat mezi aplikacemi. Jeho úkolem je sjednotit uživatelské rozhraní a zabezpečit intuitivní přístup k aplikacím, aniž by byl uživatel zatěžován jejich sledováním v rámci svého podnikového procesu.

2.2.3 Příliš těsné propojení

Při integraci je možné využít také aplikačního programového rozhraní (API), které poskytuje výrobce konkrétní aplikace. Pomocí API lze přistupovat jak k datovým zdrojům, tak k obchodní logice jednotlivých aplikací. Propojení pomocí API se využívá u složitých, nestandardních aplikací. Je však zcela závislé na výrobci daného API.

Integrace na bázi obchodní logiky znamená způsob zasahující do struktury stávajících podnikových aplikací. Cílem propojování na této úrovni je sdílet klíčové postupy sloužící při řízení celé organizace. Operace prováděné různými aplikacemi pak budou vykonávány stejně. Tento přístup také zjednodušuje správu a aktualizaci jednotlivých komponent. Na druhé straně je také nositelem mnoha rizik. Především vytváří velmi těsnou vazbu mezi aplikacemi a vzájemnou závislost. Ve své podstatě je příliš složitý, nákladný a pracný pro vývojáře či servisní podporu. Podobná rizika přináší integrace na úrovni dat, uživatelského i aplikačního rozhraní. Předpokládá totiž sdílení interních charakteristik jednotlivých aplikací.

2.2.4 Orientace na služby

Východiskem z příliš těsné a složité integrace je koncept SOA. Tato architektura je zaměřena na opakované a sdílené poskytování služeb za použití takových standardů, které tuto spolupráci umožní. Služba, je autonomní část softwaru, která slouží k plnění přesně definované úlohy a je dostupná po síti. Konkrétně může jít například o zadání faktury do ERP systému nebo získání údajů o zákazníkovi z CRM systému.

SOA jako ucelený koncept vede společně s webovými službami k vývoji tzv. kompozitních aplikací, které slučují dosud dvě oddělené součásti – vývoj aplikací a jejich integraci. Propojováním heterogenním systémů jsou zároveň definovány nové obchodní procesy.

Tvorba kompozitních aplikací spočívá v převedení stávající funkcionality do podoby webové služby a v koordinaci jejího využívání. To má za důsledek výraznou redukci nákladů na vývoj a provoz systému a rychlé zavádění nových služeb do praxe. Samotná koordinace webových služeb pak představuje asynchronní i synchronní formu komunikace na úrovni, která je pro daný přenos dat nejvhodnější. Pro popis těchto nově definovaných procesů se používá standardní jazyk (BPEL), přijatý předními IT dodavateli jako standard.

3. Datové sklady v EAI a jejich přínosy

3.1 Uplatnění datových skladů při integraci

Z výše uvedeného tedy vyplývá, že datové sklady nejsou přímo integračními nástroji. Zabezpečují však podporu EAI platforem, nebo jsou přímo jejich součástí. Jednou z možností řešení problému integrace je aplikace dnes již klasických možností propojení různorodých dat z různého databázového prostředí na bázi stále výkonnějších ETL nástrojů, jako součástí řešení datových skladů. Nezbytným předpokladem nadále zůstává velké množství analytické práce, jejíž kvalita je vždy rozhodujícím faktorem úspěšné nasazení datových skladů a v konečném důsledku i úspěchem integrace celého informačního systému. ***Datové sklady se tedy při integraci aplikací uplatní tam, kde jsou nástroje EAI využity pro přenos dat do datových úložišť v reálném čase. EAI tak doplňuje dávkový přenos a umožňuje vznik nové generace datových skladů. Podstatně širší a zajímavější je při integraci aplikací využití tzv. operačního datového skladu (ODS).***

3.2 Jak pomáhá ODS?

Mnoho organizací se potýká s tím, že tatáž informace (např. adresa zákazníka) pochází potenciálně z více zdrojových systémů. Jak ale řešit konflikty v případě, že se data z těchto zdrojů rozcházejí? Abychom získali celostní pohled na důležité objekty (klíčoví zákazníci) a podnikové procesy (servisní služby), je třeba přistupovat k více systémům zároveň, aniž by byla garantována kvalita takto získaných dat. Aplikace se často přímo dotazují na data vlastněná jinými aplikacemi, přičemž využívají znalost databázového schématu jiné aplikace a předpokládají, že je toto schéma neměnné. Tím však pěstují vzájemnou těsnou integraci, která neprospívá flexibilitě celého systému.

Zavedením ODS, implementací řízených mechanismů, jak do datového skladu přispívat a jak z něj data získávat, včetně jeho zapojení do EAI lze problémy tohoto typu velice efektivně řešit. ***Využití ODS eliminuje provázanost přes podniková data a umožňuje tak integraci na úrovni obchodní logiky nebo aplikačního rozhraní.***

3.3 Další přínosy datových skladů

Kromě zmíněných podpůrných integračních efektů přináší datové sklady především zvýšení produktivity v oblasti analýz podnikových dat, což bezprostředně podporuje strategické řízení podniku. Datové sklady se tak stávají stále častěji samozřejmou součástí architektury podnikového informačního systému. Pokud jsou integrovány v rámci EAI platformy, pak snižují celkové náklady na vlastnictví. K dosažení uvedených přínosů je třeba při implementaci věnovat velkou pozornost navrhovanému rozsahu a struktuře celého datového skladu.

Nezanedbatelná je ovšem i podpora taktického a operativního řízení, zejména pak prostřednictvím ODS (Operational Data Store). Ta se uplatňuje především formou automatizace rozhodovacích procesů, jako např. při řízení dodavatelsko-odběratelských vztahů u supermarketových řetězců.

Literatura

1. HABÁŇ, J. Manažerský přehled řešení pro inteligenci firem na českém trhu. BIZ, 2004, roč. 5, č. 12, s. 76-79. ISSN 1214-8431
2. HABÁŇ, J. Pumpy na trhu i pro menší firmy. Také menší mohou být inteligentní. BIZ, 2004, roč. 5, č.12, s. 80-81. ISSN 1214-8431

3. HABÁŇ, J., SODOMKA, P. Efektivní tvorba a provoz datových skladů. In Systems Integration 2003. Prague: Prague University of Economics, 2003, Proceedings, s. 61-68. ISBN 80-245-0522-3
4. INMON, W. Building the data warehouse, John Wiley and Sons, 1999
5. KIMBALL, R. The Data Warehouse Toolkit - Expert Methods for Designing, Developing and Deploying Data Warehouses, John Wiley, 1998
6. LACKO, L.: Datové sklady, analýza OLAP a dolování dat, Computer Press, 2003, ISBN 80-7226-969-0
7. POUR, J. Datové sklady jako nástroj integrace informačních systémů. In Svět informačních systémů 2004. Zlín: UTB ve Zlíně, CVIS – odborná sekce ČSSI, 2004, Sborník přednášek z mezinárodní konference, s. 91-97. ISBN 80-7318-166-5
8. SODOMKA, P., HABÁŇ, J. Téma Systémová integrace: Strategie, projekty, lidé a technologie (základy moderního pojetí SI) EBIZ, 2003, roč. 4, č. 9, s. 48-49. ISSN 1213-063X
9. SODOMKA, P. Jak snížit náklady na IT? BIZ, 2005, roč. 6, č. 4, s. 56-59. ISSN 1214-8431
10. SODOMKA, P. Integrace nové generace. BIZ, 2005, roč. 6, č. 4, s. 60-65. ISSN 1214-8431
11. SODOMKA, P., HABÁŇ, J., KLČOVÁ, H. Výroční výzkumná zpráva CVIS 2003. In TRNKA, F. Výzkum konkurenční schopnosti českých průmyslových výrobců. Výzkumný záměr UTB, FaME ve Zlíně. Zlín: 2004. MŠMT 265 300021.