

Kapitola 3

Argumenty pro implementaci

Existuje mnoho důvodů, proč vybudovat nebo naopak nevybudovat datový sklad nebo řešení business intelligence. Významné podniky mají rozhodně zájem vytvořit jedinou centrální terminologii, a tedy i datové prostředí s jednoznačnou interpretací dat. Vzhledem k jejich velikosti je totiž pro tyto organizace poměrně náročné zajistit, aby si všichni uživatelé rozuměli. Prakticky každá organizace nashromáždila masivní objemy dat, ve kterých je nutné analyzovat trendy, souvislosti, předpovědi, standardní hodnoty a další parametry, aby podnik mohl vyvozovat závěry a získávat konkurenční výhody.

V několika případech jsem spolupracoval s organizacemi, které si již na začátku svého podnikání uvědomily hodnotu a důsledky vhodné správy a analýzy dat. Objednaly si proto vybudování datového skladu a systému business intelligence v době, kdy se teprve začínaly rozvíjet. Ještě předtím, než nabídly své produkty a služby zákazníkům, se tyto začínající firmy rozhodly získat výhodné postavení na trhu díky tomu, že příslušné systémy dokáží analyzovat využití služeb, využití sítě a získávání či ztráty zákazníků s ohledem na podnikovou strategii. V jedné telefonní společnosti jsme například navrhli a vytvořili datový sklad využívající podnikové síťové a fakturační systémy tak, aby zaznamenával podrobnosti o voláních do konkrétních datových trhů business intelligence. Tímto způsobem bylo možné analyzovat trendy získávání zákazníků a využití služeb souběžně s marketingovými programy a kanály. V zásadě lze říci, že technologické společnosti uvažují z technologického hlediska strategicky již od samého začátku. Ze strategického obchodního hlediska to dává dokonalý smysl.

Bez ohledu na svou velikost mají všechny organizace omezené prostředky a časové plány. Každá organizace se také značně zajímá o to, jakou návratnost investic (ROI – return of investment) může datový sklad a systém business intelligence nabídnout. To může být ovšem poměrně těžké spočítat. Kvalitativní výhody datového skladu jsou jasné: zvýší spolehlivost budoucích operací, protože data budou mít vysokou úroveň integrity, zpřesní rozhodování, zpevní datovou základnu a tím poskytne firmě konkurenční výhodu atd. Obtížné je odpovědět na otázku ohledně vyčíslení návratnosti investic do datového skladu.

Návratnost investic závisí na zvolené strategii budování datového skladu. Jedná se o projekt zaměřený výhradně na úsporu nákladů na IT, jako například při migraci ze stávající platformy na jinou? Nebo jde čistě o obchodní možnosti business intelligence, například při snaze o omezení ztrát zákazníků? Ať už jsou důvody jakékoli, je potřeba určit kvantifikovaný cíl – například takto:

- Snížení nákladů na údržbu systémů mainframe o jeden milion euro ročně
- Snížení nákladů na software o 250 tisíc dolarů ročně v průběhu pěti let
- Rozšíření zákaznické základny o 5 procent během jednoho roku
- Omezení ztrát zákazníků o 3 procenta ročně opakovaně po čtyři následující roky

Potenciál výnosů je možné vypočítat třeba jako násobek průměrných nebo odhadovaných příjmů na uživatele (ARPU – average revenue per user) a počtu udržovaných nebo nově získaných zákazníků. Zvýšení výnosů díky informacím získaným z prostředí datového skladu lze také odhadnout předpovídáním prodejů na základě standardních hodnot nebo trendů. Stejně jako u jiných podnikových aktiv je v tomto případě klíčové určit potenciální nárůst příjmů nebo snížení nákladů díky datovému skladu. Jestliže jsou zisky nebo úspory vyšší než náklady na vývoj a investiční výdaje, je projekt datového skladu z hlediska návratnosti investic životaschopný.

Je však důležité připomenout, že projekt datového skladu je často obtížné ocenit z finančního hlediska. Může být značně těžké – ne-li nemožné – určit finanční úspory, které vyplynou ze standardizace podnikové datové terminologie nebo z centralizace dat z více oddělených zdrojů do jediného centrálního prostředí. Z jednoho hlediska je potřeba porozumět důvodům budování datového skladu a řešení business intelligence. Díky jasné představě o typu vytvářeného prostředí může být jednodušší identifikovat a vypočítat náklady a výnosy. Když porozumíte důvodům pro budování systému datového skladu, získáte alespoň představu o tom, jak k projektu přistupovat.

Při vývoji systémů datových skladů se často uplatňují následující scénáře:

- Migrace ze stávajících platform
- Centralizace odlišných datových skladů
- Konsolidace různých datových trhů
- Nové iniciativy
- Scénář IT „s důrazem na budování“
- Scénář „datového zmatku“ (data floundation)

V následujících částech kapitoly popíšeme běžné scénáře budování nebo odmítnutí datového skladu a řešení business intelligence spolu s přehledem příslušných aspektů.

Migrace platformy

Budování datového skladu s cílem migrace může vycházet ze strategie změny platformy. Obvykle k tomu dochází, když datový sklad vychází ze starších systémů, zpravidla počítačových systémů typu mainframe, které v podniku fungovaly mnoho let. Aktuální strategie spočívá v přesunu na servery střední vrstvy. Hlavním důvodem obvykle bývá omezení nákladů nebo likvidace počítačů mainframe. Měsíční náklady na údržbu počítačů mainframe často převyšují

jí součet nákladů na přechod k nové platformě a nákup kompletního nového softwaru. Jedná se o dokonalou příležitost k nákupu hotového datového modelu datového skladu, databáze, disku atd.

Zajištění nepřerušného provozu

Vzhledem k tomu, že systém datového skladu a všechny jeho funkce již jsou k dispozici, podnik očekává, že nedojde k přerušení provozu. Organizace tedy již nějakou dobu používá vvinuté programy pro vykazování, databáze, rutiny pro načítání a další komponenty staršího systému. Ještě důležitější je fakt, že podnik se přizpůsobil aktuálním sestavám. Data mohou být do jisté míry pochybná. Avšak vzhledem k tomu, že starší systém funguje a poskytuje podniku pravidelné sestavy, znamená to, že podnik vychází z určité základní úrovně. Tato pravidelnost a znalost standardních hodnot mají velký význam. Pokud by došlo k výrazné odchylce od aktuální úrovně, mohlo by to podnik zásadně poškodit. Proto je potřeba udržet dojem nepřerušného provozu.

Migrace z jedné platformy na jinou přináší mnoho technických změn, které mohou vést i ke změnám vykazovaných informací. Často se stává, že při změně nebo kompletním přepsání programů při migraci se zjistí, že neposkytují stejné výsledky jako jejich předchůdci. Příčinou rozdílů může být nedokumentovaná oprava chyby před několika lety, která nebyla správně přenesena, nebo nesprávná interpretace logiky zdrojového programu. V každém případě se to však projeví odlišným výstupem. Ještě horší je fakt, že výsledné rozdíly se nemusí projevit na úrovni počátečního programu, ale v závislém programu v další fázi zpracování. Jinými slovy: jeden program může změnit data, ale změna se ukáže až na výstupu jiného programu.

V důsledku toho finální sestavy nového systému nemusí obsahovat stejná čísla jako sestavy generované starším systémem. Pokud je k dispozici záznam pro audit, lze zpracování položek sestavy vysledovat. V opačném případě však může vrcholné vedení zažít trapné chvíle, protože nelze zjistit, jak byly hodnoty agregovány nebo odvozeny.

Zpětný překlad

Další významnou fází scénáře se změnou platformy představuje zpětný překlad aktuálního systému datového skladu, aby bylo možné zjistit, jak byl vytvořen. Pokaždé se jedná o náročný úkol. Podle mého názoru vyžaduje zpětný překlad vždy více času než napsání nových programů od začátku.

Nejhorší situace nastává u mnoha starších systémů, jejichž programy byly nedostatečně dokumentovány nebo nemají žádnou dokumentaci a ve většině případů se v průběhu let nezaznamenávaly aktualizace ani opravy. Systém jednoho zákazníka načítal moduly programu COBOL, ke kterým nebyl dostupný zdrojový kód. Jinými slovy program fungoval, ale samotný jeho kód neměl žádnou dokumentaci. Nikdo přesně nevěděl, co přesně program dělá. Program se spouštěl jen jednou ročně při uzávěrce, ale chybějící dokumentace znamenala pro podnik značné riziko. Pokud by program z nějakého důvodu (např. nesprávných datových hodnot či datového formátu) přestal fungovat, znamenalo by to pro podnikové vykazovací prostředí katastrofu.

Zpětný překlad je v některých případech nutností. Tyto projekty musí vyhradit čas na analýzu toho, co procesy skutečně dělají a jak přitom postupují. Poté je nutné znovu vytvořit obdobné procesy v novém jazyce, případně na nové platformě. To vyžaduje čas a úsilí. Téměř všechny

projekty, se kterými jsem se setkal, podcenily náročnost tohoto úkolu, a došlo proto k překročení rozpočtu. Projekty migrace ze starších systémů na novou platformu proto trvají značně déle, než bylo plánováno, vyžadují oproti prvním odhadům více pracovníků a stojí více, než se čekalo. Možnost outsourcingu těchto projektů se v některých případech poměrně osvědčila a v jiných naprosto selhala.

Z výše uvedeného vyplývá, že projekty zpětného překladu mohou být nákladné a občas je nelze úspěšně dokončit. Pokud je však starší systém dokumentován a z programů lze odvodit stávající softwarovou logiku, je projekt zpětného překladu reálný, ale v každém případě si vyžádá čas a odpovídající prostředky.

Kvalita dat

Vyšší náklady nejsou ve většině případů způsobeny samotnou migrací, ale problémy s kvalitou dat, které se objevují vždy. Často se zdá, že kvalita dat je v datovém skladu kořenem všeho zla. V kapitole 2 jsme rozebírali, co to je kvalita dat. Kolik úsilí je však potřeba věnovat opravám dat, aby bylo možné zvládnout plánované úkoly? Zvýšení kvality dat obvykle vyžaduje určitý čas, protože je nutné vytvořit profil a najít vlastníka a management musí zvolit řešení: obejít, opravit u zdroje, opravit ve fázi ETL atd. Optimální postupy doporučují, aby se potíže s kvalitou dat opravily tam, odkud data pocházejí. Opravy dat ve zdrojovém systému však nemusí být praktické a občas jim může bránit správce zdrojového systému. Výkonné vedení proto musí najít a schválit pracovní postup a správce datového skladu musí poté vytvořit příslušný návrh a realizovat jej.

U většiny migrací starších systémů na jinou platformu jsou potíže s kvalitou dat odpovědné za velkou část celkového úsilí, zejména v případech, kdy je starší systém datového skladu poměrně letitý. V dřívějších dobách nepředstavovala kvalita dat takový problém, protože data byla častěji určena k jedinému využití a v mnoha případech pocházela z jediného zdroje. V současnosti se všeobecně uznává, že je důležité sdílet data v rámci celého podniku. Data proto mají různé zdroje a objevují se zásadní potíže, například s terminologií a datovými hodnotami. Komplikace s kvalitou dat ze starších systémů způsobuje opakované použití datových polí. Staré programy jazyka COBOL používaly 77 změn definice úrovně. To znamená, že datové pole mohlo v jedné instanci označovat hodnotu X a v jiné instanci se vztahovat k hodnotě Y. V těchto případech prakticky docházelo ke směšování dat z různých zdrojů, která bylo velmi obtížné (i když ne nemožné) oddělit.

Jeden zákazník měl starý zdroj, který poskytoval data starému datovému skladu. Určitý program pro načítání hledal ve zdrojovém souboru hodnotu „X“ na pozici 286. V případě výskytu hodnoty „X“ provedl určitou logiku. Nikdo nevěděl, co to „X“ znamená a kde se tam vzalo. Když se tento problém začal na poradě vedení řešit, ukázalo se, že vedoucí zdrojového systému tuto logiku před deseti lety sama naprogramovala. Bohužel si však nedokázala vzpomenout, o co šlo a proč se to do programu dostalo. Pamatovala si pouze, že příslušný kód poskytoval přesnější výsledky. Problém s kvalitou dat se přenášel do dalších fází zpracování, a mohl se proto později znovu projevit. Co by se však stalo, pokud by se tento prvek ve zdrojovém souboru přestal vyskytovat? Ovlivnilo by to stávající sestavy? Všechny potíže s kvalitou dat je potřeba dokumentovat a prozkoumat. Neodkládejte řešení těchto problémů na pozdější dobu.

Paralelní prostředí

Téměř všechny podniky, které realizují projekt s migrací platformy, se z nějakého důvodu domnívají, že budou starý systém mainframe a nově vybudovaný systém provozovat souběžně jen krátkou dobu, například měsíc či dva, a poté starší systém odpojí. Z vlastní zkušenosti mohou říci, že se to nikdy nepodařilo. Oba systémy obvykle paralelně fungují alespoň tři až šest měsíců, což závisí na granularitě vykazování: denní, měsíční nebo kvartální.

To lze vysvětlit následovně: Podnik již nějakou dobu používá starší datový sklad a při svém rozhodování vychází z jeho výstupu. Pokud se objeví potíže s kvalitou dat, může to vést k tomu, že se změní klíčové indikátory výkonu (tj. měřítko či metriky). To znamená, že podnik již nepracuje se stejnými fakty, což má dopad na rozhodovací proces a může se to projevit změnou podnikových aktivit nebo zaměření podniku. V těchto případech se vedení může rozhodnout, že přechod pozastaví, aby bylo možné zjistit, proč se objevují tyto rozdíly ve výstupech staršího systému a nově vybudovaného systému.

Je rozumné provozovat současně starý i nový systém po celou plánovanou dobu. V každém případě je vhodné mít k dispozici srovnání za více než jeden či dva měsíce souběžného provozu. V případě měsíční granularity vykazování je například praktické, aby obě prostředí fungovala alespoň tři až šest měsíců. Podle mých zkušeností mohou organizace, které používají staré i nové systémy během tohoto období, ukončit provoz starého systému s mnohem větší jistotou. Důvod spočívá v tom, že první měsíc obvykle ukáže nepředpokládané neshody. Pokud vše běží dobře a problémy jsou opraveny, mohou se opravené výsledky ověřit druhý měsíc. Třetí měsíc slouží k dvojnásobnému potvrzení, že vše funguje správně, a poskytne managementu vyšší jistotu, kterou k migraci na nový systém potřebuje. V tomto případě by stačily tři měsíce, ale v případě výskytu dalších zásadních nesouladů může být potřeba období prodloužit.

Před několika lety přecházela velká banka ze staršího datového skladu založeného na počítači mainframe na novou platformu. Přes všechno úsilí o vyřešení potíží s kvalitou dat a investice do zpětného překladu se však dostala do situace, kde nový a starý systém neposkytovaly stejné výsledky a banka nedokázala najít důvod těchto rozdílů. Pokud by situaci pasivně přijala a poskytla nové finanční výsledky zákazníkům, mezi které patřilo i několik velkých investičních fondů, určitě by dramaticky poklesla hodnota jejich akcií, protože finanční výsledky by se zásadně lišily od předchozího fiskálního období. V tomto případě se management rozhodl implementaci datového skladu zpomalit, aby bylo možné rozdíly ve výsledcích sladit. Trvalo to přitom pouze o několik měsíců déle.

Přidaná hodnota

Když v případě přístupu, který je založen na pouhé změně platformy, dojde k překročení nákladů, obvykle je připravena nabídka projektu pro koncové uživatele. Když současně probíhá migrace a nový systém poskytuje přidanou hodnotu pro konkrétní účel nebo zákazníky (tj. koncové uživatele), může rozpočet na nový projekt pomoci při dofinancování projektu migrace na novou platformu.

Tento ekonomicky výhodný přístup může také celému úsilí o budování datového skladu dodat novou energii, protože získá podporu více podnikových uživatelů. Řízení odlišných programů sice představuje dodatečnou komplikaci, ale projekt lze rozšířit o program interního marketingu. Úsilí o změnu platformy obvykle zahrnuje nové funkce business intelligence, ale projekt BI musí obvykle počkat, než jsou dokončeny fáze zajištění kvality dat a zpětného překladu.

Centralizace datového skladu

Myšlenka centralizace datového skladu vychází z toho, že je vhodné shromáždit veškeré informace o datech, procesech, architektuře a použití z několika stávajících datových skladů a sloučit je do jediného centrálního prostředí neboli systému. Tento projekt může mít logickou nebo fyzickou povahu.

Migrace několika datových skladů do centralizovaného umístění může zahrnovat změnu platformy, protože jeden nebo více stávajících datových skladů se mohou nacházet na samostatných serverech s odlišnými nebo podobnými operačními systémy. Celý projekt sleduje dva cíle. Sjednocení na jedné platformě ve fyzickém scénáři sníží náklady na hardware a podporu a v logickém i fyzickém scénáři umožní, aby se rozvinula centralizovaná strategie zpracování dat.

Tyto typy centralizačních scénářů se obvykle realizují v situacích, kdy společnost zakoupí jinou firmu, nebo když jedna organizace vytvořila dva nebo více datových skladů a nyní je potřeba sloučit.

Fúze podniků

Ve scénáři fúze dvou podniků je nutné začít odbornou analýzou. To znamená, že před zahájením jakýchkoli změn je potřeba podrobně prozkoumat všechny datové sklady. Je každý datový sklad na úrovni oddělení nebo celého podniku? Má podobné nebo zcela odlišné zdroje? Měl by se plán řídit nadřazenou architekturou a strategií, sledovat nově získané řešení, nebo by měl vzniknout jako hybrid obou přístupů? Fyzická analýza samozřejmě zahrnuje i celý podnik včetně analýzy odborné úrovně pracovníků a hodnocení technické architektury.

Slučování v rámci organizace

Pokud dochází k internímu slučování systémů, je pravděpodobné, že nastal jeden ze dvou případů. Buď se zjistilo, že náklady na provoz dvou nebo více kompletních systémů datového skladu jsou nesmyslně vysoké, nebo podnik dospěl k tomu, že z organizačního hlediska je nejvhodnější mít veškerá data v jednom centrálním prostředí, což je primárním účelem datového skladu. V mnoha případech jsou jednotlivá oddělení politicky motivována k tomu, aby si vytvářela vlastní řešení business intelligence. Platí to zejména v sektoru finančnictví a pojišťovnictví. Vedoucím pracovníkům v těchto organizacích značně záleží na jejich každoročních odměnách, které mohou být poměrně vysoké. Snaží se proto dosáhnout dobrých výsledků bez ohledu na výsledky jiných oddělení. Projekty datových skladů na úrovni jednotlivých oddělení pak získávají kompletní podobu.

Centrální návrh a místní použití

V jiných scénářích s interním slučováním vystupují globální organizace, které mají pobočky po celém světě. Tyto organizace mají strategickou vizi, která požaduje konsolidovat podniková aktiva do centrálně integrované architektury a datového prostředí. Tento požadavek nemusí nutně vést k centrálnímu databázovému úložišti, ale někdy je výsledkem federativní nebo satelitní řešení datového skladu s centralizovaným návrhem a centralizovanou koordinací. Jednotlivé pobočky mohou spravovat své vlastní datové sklady, ale návrh a architekturu určuje centrum.

Tím je zajištěna podniková datová terminologie, distribuce strategií vykazování, a tedy i standardizace finančního účetnictví a znalostí.

V obou scénářích je nutné před zahájením vlastní konsolidace shromáždit příslušné informace. Je nutné získat znalosti o každém prostředí, které se podílí na slučování. Tímto způsobem je možné určit základní datové komponenty, procesy, architektury, způsoby použití atd. Hlavním účelem těchto postupů je z podnikového hlediska centralizace veškerých podnikových dat ve společné oblasti, do které získají přístup všechny zainteresované strany. To znamená, že všichni budou diskutovat o jablkách a nebudou si plést jablka s hruškami. Z datového hlediska oddělení IT je hlavním cílem vytvořit společnou terminologii a vytvořit vhodnou strukturu dat, která zajistí celopodnikovou standardizaci.

Tento přístup obvykle začíná průzkumem oddělení IT nezávisle na tom, zda podnik tento projekt podporuje. V praxi jde o fungování celého podniku, ale první krok, který popisuje dostupná interní aktiva, vychází z iniciativy oddělení IT. O následném postupu již rozhoduje podnikové vedení v závislosti na podnikovém využití a hodnotě. Obvykle nejde o přístup „s důrazem na budování“ (který si vysvětlíme dále v této kapitole), protože celý projekt se stává součástí celopodnikové datové strategie a ke svému úspěchu vyžaduje podnikové sponzory a zástupce.

Konsolidace datových trhů

V tomto scénáři má několik oddělení svá vlastní prostředí datových trhů. Zdroje mohou a nemusí být stejné a vykazování se obvykle liší, ale může existovat určitá konformita na úrovni základních dat.

Datové sklady lze spravovat čistě na úrovni podnikových oddělení nebo je může částečně spravovat oddělení IT. V každém případě bylo zjištěno, že tato nezávislá prostředí představují nevyhodu a nadále nejsou v nejlepším zájmu ani jednotlivých oddělení, ani celé organizace.

Podniková oddělení mohou potřebovat vlastní týmy IT a dospívají k závěru, že získávání dat z mnoha zdrojových systémů je poněkud náročné a komplikované. V této fázi v podniku převládne názor, že by bylo výhodnější, kdyby mohlo celkové fyzické prostředí spravovat a zajišťovat jeho činnost specializované oddělení IT. Oddělení IT zaměstnává pracovníky, kteří umí extrahovat data ze zdrojových (neboli provozních či funkčních) systémů a lépe dokáží navrhovat a budovat robustnější prostředí než malé týmy IT v odděleních koncových uživatelů. V podnikových odděleních obvykle pracují podnikoví analytici, tvůrci sestav a občas i méně zkušení správci databází.

Tento scénář se z hlediska vývoje datového skladu liší od lehkého po intenzivní nasazení. Myšlenka odpovídá strategii centralizace datového skladu, ale v tomto scénáři nemuselo být prostředí datového skladu v praxi dosud vytvořeno. Malé podnikové týmy IT mohou v současnosti zajišťovat analytické vykazování, které však neodpovídá robustní strategické perspektivě datového skladu. Mohou používat datové trhy na úrovni oddělení, které jsou svým rozsahem a funkcemi přizpůsobeny pouze účelům jejich oddělení. V tomto scénáři se k vykazování obvykle používají aplikace Microsoft Access a Excel. Návrh dat může mít různou podobu: od amatérských dimenzí a tabulek faktů, které nijak nezávisí na žádném potvrzení, po hybridní návrh nebo modely hvězdice, sněhové vločky a třetí normální formy spolu s tabulkami, které jednoduše fungují. Na druhé straně mohou mít návrhy správné potvrzené dimenze a elegantní hvězdicová schémata, která však nejsou potvrzena v datových skladech jiných oddělení.

Tento scénář se obvykle uplatňuje v menších odděleních, která intenzivně využívají aplikace MS Access a Excel. Zpravidla zaměstnávají jednoho až dva techniky, kteří přebírají data z několika zdrojových systémů a převádějí je do formátu, který je podle jejich názoru pro podnikové uživatele praktický. Tento přístup vychází z podnikového využití a umožňuje vyvinout a používat mnoho sestav. Problém spočívá v tom, že v nejhorším případě může existovat mnoho redundantních sestav, které by dobrý nástroj BI dokázal sloučit do jediné přehledné krychle. Rozbor naleznete v části „Uživatelská prezentace business intelligence“ kapitoly 1.

Tyto typy systémů datových trhů v odděleních obvykle pocházejí z prvních projektů datových skladů. Vznikly kvůli ukládání historických transakcí, protože zdrojové systémy obvykle ukládají transakce jen na krátké aktuální období. Zpravidla se používá určitý typ obecné agregace, která připraví data pro libovolné uživatele oddělení, aby je mohli zpracovávat podle potřeby. To však znamená, že pokaždé nejsou na 100 procent známe skutečné požadavky podnikového použití. Když počet těchto datových trhů překročí tři až čtyři, komplikuje se správa a objevují se požadavky, aby oddělení IT pomohlo při centralizaci a importu dat. Pochopitelně se na mnoha místech objevují potíže s kvalitou dat, protože koncoví uživatelé si jednoduše zvykli na to, že data bývají nesprávná nebo neúplná, například kvůli úpravám délky období.

V jiných scénářích si více oddělení vytváří vlastní prostředí datových trhů. Některé mohou být na výše popsané úrovni a jiné mohou být pokročilejší. V každém případě by zde vývojáři měli porozumět tomu, co podnik z hlediska vykazování aktuálně používá, a poté extrahovat datové komponenty. To znamená porozumět podnikové terminologii neboli slovníku a provést konverzi na základní datové komponenty. Těmto projektům může dokonale vyhovovat model s potvrzenými dimenzemi spolu s architekturou sběrnice nebo normalizovaným logickým návrhem, který lze integrovat do podnikového modelu.

Výhodou těchto prostředí je, že podnik již může mít k dispozici poměrně zkušené podnikové a zdrojové analytiky. Vzhledem k tomu, že se o sebe museli sami postarat tak dlouho, obvykle již hodně vědí o podnikovém použití a zdrojových systémech. Hlavní podnikový analytik se v těchto scénářích obvykle stává expertem na problematiku oddělení. Je však potřeba dbát opatrnosti. Tito analytici mohou znát data a činnost podniku, ale nemusí nutně vědět, co s výsledky podniku druhý nebo třetí pracovník v řadě.

Nevýhoda těchto scénářů spočívá v tom, že podnik si obvykle chce udržet jistou míru kontroly nad vývojem svého prostředí. To platí zvláště tehdy, jestliže podnik pracuje s pokročilými nástroji BI. Panuje přesvědčení, že podnikoví analytici dokáží navrhnout datové požadavky sami a oddělení IT by mělo pouze zajišťovat podnikový servis. Je důležité, aby se oddělení IT v této fázi zapojilo do příslušných podnikových aktivit a usnadnilo správu podnikových dat. Díky tomu lze určit skutečné podnikové požadavky a rozsah použití. Projekt business intelligence zde nespočívá pouze v přenesení odpovědnosti za vkládání dat na oddělení IT, ale vyžaduje spolupráci při budování skutečného prostředí BI od používání zpět k návrhu, poté ke zdrojům a nakonec k celkové struktuře. Datový architekt by měl mít s těmito činnostmi bohaté praktické zkušenosti. Zároveň je důležité, aby oddělení IT nepřebíralo iniciativu a nepokoušelo se obejít analytiky na straně podnikových uživatelů a obracet se přímo na jejich uživatele. To by způsobovalo konflikty a odpor dotčených vedoucích pracovníků.

Na začátku projektu se zaměřte na vlastní řešení, nikoli na vlastníky jednotlivých prvků či na žabomyší spory. Snažte se dospět ke konkrétnímu a použitelnému řešení, které zvýší hodnotu podniku v co nejkratší době a bude přitom volit maximálně strukturovaný postup s mini-

málními potížemi. Jakmile se řešení objeví, podnikoví uživatelé si uvědomí výhody nového systému, což podpoří užší partnerství mezi techniky z podnikových oddělení a zaměstnanci oddělení IT, které tak bude mít otevřené dveře k prosazování svých představ.

Nová iniciativa

Přístup „nové iniciativy“ je založen na partnerství mezi podnikem a oddělením IT. Podnik chce obvykle vytvořit řešení business intelligence pro vykazování a případně také pokročilé řešení, které obsahuje ovládací panely a přehledy výsledků. Podnik již může mít k dispozici vykazovací prostředí, ale v zásadě se to považuje za projekt „na zelené louce“. Scénář vychází z toho, že podnik má strategii včasného a přesného vykazování, která umožní analyzovat základní parametry podniku, omezit náklady a zvýšit výnosy.

Odborníci na dodavatelský řetězec to označují jako *výroba na objednávku*. Projekt spočívá v budování datového skladu, který umožní rychle realizovat scénář business intelligence s důrazem na přidanou hodnotu pro podnik. Zásadní výhoda tohoto scénáře leží v tom, že podnik je ochoten podpořit nový vývojový projekt a vedoucí pracovníci se obvykle o projekt intenzivně zajímají.

Začátek nové iniciativy datového skladu je často doprovázen vysokými očekáváními podnikových uživatelů i oddělení IT. Zainteresané osoby se tak zahájení projektu nemohou dočkat. Zpočátku se tolerují i větší zpoždění, protože všichni chápou, že takové prostředí v organizaci dosud nikdo nevybudoval. Projekt může postupovat přiměřeně pomalu, protože zaměstnanci se stále seznamují se svými úkoly a s rozdíly mezi fungováním datového skladu a aktuálních transakčních systémů. Do počátečních fází těchto projektů je vždy vhodné zapojit zkušeného externího konzultanta.

Oddělení IT má za úkol implementovat projekt a vybudovat celé prostředí s plnou podporou podniku. To znamená, že oddělení IT může tento projekt integrovat do podnikové datové strategie a využít optimálních postupů. Nesmí však ztratit ze zřetele hlavní cíl, tj. zvýšení hodnoty pro podnik. Kromě toho je nutné pečlivě předcházet průběžnému zvětšování rozsahu projektu. V této fázi je projekt veřejný a vzbuzuje vysoká očekávání. Proto se často rozšiřuje o nové požadavky – pouze jedna datová oblast pro jednu nebo dvě další sestavy, které však mají značný dopad. Pečlivě plánujte počáteční fáze a hlídejte doplňování dalších požadavků. Jestliže se organizace o vytvoření datového skladu zatím nepokusila, mějte na paměti, že první fáze projektu datového skladu vyžaduje vybudování pevných základů. Můžete přidávat hodnotu pro počáteční podnikové použití, ale uvědomte si, že značné úsilí je nutné zaměřit na stavbu základů. Pokud se datový sklad již dříve budoval a tento projekt tehdy nejspíš selhal (proto začíná znovu od začátku), uvědomte si, proč poprvé nebyl úspěšný, a opět omezte nárůst jeho rozsahu.

Je-li celá snaha orientována na řešení na úrovni oddělení, je projekt obvykle zaměřen na datové trhy a zdroje. Datová architektura přitom sleduje podnikové standardy. Jestliže projekt začíná na úrovni oddělení, ale vedení společnosti jej schválilo jako celopodnikovou strategii, projekt se postupně mění na iniciativu komplexního podnikového datového skladu, kdy počáteční fáze umožní podporu požadujícího oddělení. Oba typy projektů jsou v praxi stejné, ale každý z nich klade důraz na poněkud odlišné aspekty. V prvním případě je v centru pozornosti konkrétní podniková hodnota, kdy oddělení IT zajišťuje soulad s terminologií, datovými strukturami a úrovní kvality dat a uplatňuje přitom optimální postupy. Druhý přístup se zaměřuje spíše na

zajištění strategie podnikového datového skladu a business intelligence pro celou organizaci. Přitom v první fázi poskytuje konkrétní podnikovou aplikaci.

Je velmi důležité získat sponzora na úrovni technologického ředitele, který zajistí, že oba tyto scénáře mohou určovat budoucí vývoj a datový sklad zůstane v souladu s jednou centrální verzí dat. Pokud se první scénář změní na projekt konsolidace datových trhů nebo druhý na projekt centralizace datových skladů, došlo k opuštění počáteční podnikové strategie. Datový architekt, vedoucí projektů a výkonné vedení spolu s odpovědnou osobou ve vrcholovém vedení musí spolupracovat na podnikovém řešení.

Potíže nové iniciativy mohou být způsobeny tím, že podnik plně nerozumí fázi návrhu neboli logickému datovému modelování a domnívá se, že k vybudování datového skladu postačuje vytvořit databázové tabulky a naplnit je daty. Z počátku existuje dobrá vůle napravit všechny problémy s kvalitou dat, ale jak projekt postupuje a potíží s kvalitou dat se objevuje stále více, prioritou oprav dat klesá, zejména když se blíží jednotlivé projektové termíny. Obecně převládá názor, že je potřeba budovat rychle, i když zdravý rozum napovídá, že vše může nějakou dobu trvat. Zainteresované osoby by se měly snažit, aby se nevyvinula situace „datového zmatku“ (kterou si vysvětlíme v další části kapitoly).

Nová iniciativa může usilovat o nižší zatížení provozních systémů díky vytvoření speciálně optimalizovaného prostředí, které je určeno k vykazování nebo analýze OLAP. V každém případě by měl návrh umožnit snadné používání a zároveň zajistit, že data ze samostatných zdrojových systémů budou shromážděna v jedné centrální oblasti, která zaručí bezpečnost a kvalitu dat. Speciální návrhy datových struktur také zajistí efektivitu, protože tyto struktury v systémech business intelligence a v transakčně orientovaných systémech se vzájemně liší. Podnikové centrální umístění s potvrzenou terminologií a koordinací správy a použití představuje pro organizaci skutečnou konkurenční výhodu a podnikové aktivum.

Nová iniciativa: dynamické vykazování

Projekty business intelligence nejčastěji vypadají tak, že oddělení IT převezme roli najatého vývojářského týmu a zbytek podniku se na projektu podílí jen málo. Tým datového skladu dostane za úkol vyvinout řešení BI pro konkrétní potřeby vykazování. Projekt zahrnuje pouze vytvoření určitých sestav v dynamickém prostředí, které umožní přechod na nižší a vyšší úroveň podrobností. Projekt je zaměřen na funkce business intelligence, ale přistupuje k řešení výhradně z perspektivy datového skladu, protože oddělení IT dostane pevný úkol implementovat určitý počet sestav. Řešení má podniku poskytnout dynamický pohled na podniková data v předem definovaných sestavách díky vizualizaci OLAP.

Koncepce tohoto scénáře vychází z toho, že podnikoví uživatelé budou moci pracovat s určitými sestavami podle potřeby a dokáží analyzovat data na základě hranic vykazování pomocí funkcí procházení mezi úrovněmi podrobností. Potenciální nevýhoda tohoto scénáře spočívá v tom, že tým datového skladu nepracuje jako aktivní součást podnikového řešení. Tým v tomto scénáři jednoduše dostává úkol zjistit, jak celý projekt realizovat. Tým by se měl zapojit do podnikového rozhodování nejen při hledání způsobu, jak zprovoznit nové prostředí vykazování, ale také při analýze a vytváření správného a vhodného řešení vykazování, které bude splňovat podnikové potřeby. Mnoho projektů datových skladů nebo business intelligence končí neúspěšně kvůli nevhodnému přístupu. Nechovejte se jako číšník, který pasivně přijímá objednávku: poraďte se s kuchařem, které jídlo je pro hosta nejlepší. Když nic jiného, návrhy

sestav by neměly být založeny pouze na konkrétních dostupných sestavách, ale měly by zohlednit základní datové souvislosti.

V tomto scénáři může tým dostat sadu sestav (100 nebo i 120 000 jako v dříve zmíněném případě) a zadání vybudovat datový sklad, který umožní tyto sestavy používat. Který krok je potřeba učinit nejdříve? Obvykle se pokusíte uspořádat sestavy do nějakých skupin, které se označují jako cílové podnikové oblasti. Vždy se však vyskytnou sestavy, které se plně nehodí ani do jedné z oblastí. Sestavy proto dostávají prioritu s ohledem na různé typy interních skupin: jednoduché, průměrné a komplikované. Potom je potřeba jednotlivé oblasti postupně analyzovat. Pohled na priority sestav vychází ze zběžného vyhodnocení, kolik odlišných dat sestavy obsahují. Poté proběhne diskuze s podnikovými uživateli nad jednotlivými sestavami. Jedná se o poměrně nudný a zdoluhavý proces. Nakonec jsou položky umístěné vespod jednoduše označeny stupněm priority. To zní povědomě.

Nakonec začíná skutečná analýza, která může vyžadovat hodně času. Tato první fáze zahrnuje analýzu a hledání základních datových komponent každé sestavy. Samotné sestavy jsou obvykle tvořeny kombinací jiných sestav buď ze stávajících systémů, nebo odvozených z aplikace Excel. Podnikoví uživatelé obvykle doplňují další „přidanou hodnotu“, což je v praxi seznam požadavků na sestavy, které rozšíří množinu již existujících. Upřesněný seznam sestav se stává počátečním cílem projektu. Hlavním smyslem této práce je určit základní datové komponenty a uspořádat je optimálním způsobem, aby umožňovaly efektivní použití. V mnoha případech podnik dodává jednoduché specifikace sestav, které jsou nejasné a vyžadují doladění. Podnikové SME zpravidla není k dispozici, což snahu o analýzu téměř znemožňuje a zpomaluje plánovaný postup. Zajistěte, aby byly požadavky na sestavy zdokumentovány dříve, než začnou běžet přísné termíny pro analýzu a vývoj.

Tento scénář také tvoří část mnoha jiných scénářů budování. Migrace platformy, centralizace datového skladu i konsolidace datových trhů obvykle v nějaké formě zahrnují i uvedený scénář.

Důraz na budování

Tato strategie datového skladu je založena výhradně na přístupu IT. V terminologii dodavatelského řetězce se to označuje jako projekt „výroby na sklad“. Přitom se očekává, že po vybudování prostředí se rozvine i zájem o jeho využívání. Tento projekt obvykle iniciují datoví architekti, členové databázové skupiny nebo noví vedoucí oddělení IT, kteří již znají výhody datového skladu a jsou přesvědčeni, že organizace by se měla vydat tímto směrem, ať už je to v souladu s podnikovým strategickým směrem nebo nikoli.

Scénář s důrazem na budování spočívá na přístupu výhradně shora dolů, který je založen na návrhu centrální datové vrstvy, který identifikuje všechna data v rámci organizace. Uvedený přístup samozřejmě zahrnuje stanovení určitých priorit, což zajistí zachycení základních dat organizace. Obvykle se jedná o hlavní transakční data organizace spolu s primárními tematickými oblastmi neboli datovými pilíři. U telekomunikačního operátora se například hlavní datové komponenty ve většině oddělení nazývají záznamy podrobností o voláních (CDR – call detail record). Budou se tedy zaznamenávat všechny podrobnosti o voláních, včetně volajících a volaných čísel, přenašečů podílejících se na každém volání, trvání hovorů, jejich ceně, času volání, jeho typu (hlas, SMS, data atd.) spolu se základními pilíři: data o zákazníkovi, produktu a umístění. V případě maloobchodních prodejců se první fáze projektu s důrazem na budování

zaměří na pokladní transakce: identifikace produktu, částka prodeje, počet prodaných položek, obchod atd.

Přístup s důrazem na budování je výhodný v tom, že oddělení IT má vizi a pracuje na něčem, co celé organizaci pomůže z dlouhodobého hlediska (alespoň se tak domnívá). Nevýhoda tohoto přístupu spočívá v tom, že oddělení IT buduje podle svých představ bez důrazu na konkrétní podnikové použití. Projekt je tedy značně omezený a pravděpodobně se jedná o hračku jednoho z výše postavených datových odborníků nebo některého vedoucího. Oddělení IT obvykle v rámci podnikové komunity usiluje o podporu mezi středním managementem, ale nezískává na svou stranu vrcholové vedení.

V těchto scénářích zpravidla vzniká nestandardní datový model. Jestliže je tento model vytvořen s ohledem na flexibilitu, může být vše v pořádku. Pokud je ale model založen na snaze analyzovat všechna data najednou, celý projekt se dostane do slepé uličky jednoduše proto, že se snaží zvládnout příliš mnoho úkolů. Pravděpodobně jste zaslechli, že se podnik o takový projekt pokusil před několika lety, ale úkol byl příliš náročný a nestačily na něj prostředky, takže musel být opuštěn. Tyto projekty obvykle fungují nejlépe při nákupu předem připraveného datového modelu, který může celému úsilí poskytnout strukturu a organizaci. Pokud z projektových prací vyplyne interní datový model, může se ukázat, že se tento model v druhém, třetím a dalších projektech musí změnit. Podnik se tak od projektu postupně odvrací, protože je nutné návrh neustále přepracovávat. Pořízený hotový model nabízí určitou předem připravenou strukturu, která dovoluje vyzkoušet umístění prvků, a poskytuje dobrou metodologii k organizaci dat do budoucna. Další informace o struktuře datových modelů uvedeme v druhé části této knihy.

Tyto projekty s důrazem na budování mívají špatnou pověst, protože je jejich zastánci v organizaci často propagují téměř nekritickým způsobem. Podnikoví uživatelé brzy ztratí chuť poslouchat, co by měli dělat, a raději se od projektu odvrací. Po získání sponzora však může projekt nabrat směr a vývoj může postupovat vpřed poměrně rychle.

Projekty s důrazem na budování se vyznačují také tím, že bývají nákladné a bez jasného cíle. Některé projekty usilují o vytvoření podnikové terminologie a podnikového logického datového modelu bez aspektů vykazování, bez databázového prostředí a bez snahy o funkce ETL. Jedná se pouze o cvičení v datovém návrhu, která mohou, ale nemusí zahrnovat získávání dat. V tomto případě je projekt na místě. Pamatujte, že datový sklad by měl podniku poskytovat přidanou hodnotu. Pokud projekt zahrnuje vytvoření podnikového datového úložiště, může být zdánlivě užitečný. Jestliže však nemá žádnou vazbu na návratnost investic, nemá žádné podnikové použití a neposkytuje žádnou hodnotu. Na kterých oblastech by měl datový modelář pracovat: na zákaznících, produktech nebo událostech? Jestliže je součástí strategie IT vytvoření takového prostředí, mělo by oddělení IT získat sponzora, rozpočet, cíl, a v důsledku tedy projektem zvýšit podnikovou hodnotu. Jedná-li se o iniciativu některého člena vedení, může projekt souviset s náklady ušlé příležitosti, které jsou v rozporu s aktuálními podnikovými iniciativami.

Datový zmatek

Takový scénář skutečně existuje.

Datový zmatek (data floundration) popisuje stav, kdy se data v datovém skladu nechovají podle očekávání. Jinými slovy se jedná o neobratnou snahu zajistit optimální postupy týkající se dat. Tento přístup není vhodný a bohužel se vyskytuje častěji, než by bylo žádoucí. Určitá organizační složka podniku požádá o to, aby její datové požadavky doplnily stávající iniciativu datového skladu, která již může fungovat nebo se může nacházet v počátečních fázích. V druhém případě vedoucí datového skladu obvykle usiluje o podporu podnikových uživatelů a o zvýšení rozpočtu. Projekt se kvůli tomu obzvláště snaží získat přízeň uživatelů, a proto mu hrozí, že se dostane do situace datového zmatku.

Podnik jednoduše požaduje, aby něco rychle fungovalo, protože již pravděpodobně realizuje vlastní program business intelligence. Neberou se žádné ohledy na datové modelování, podnikovou terminologii ani synchronizaci se stávajícími projekty. Všichni se totiž domnívají, že po několika měsících se vše stejně přestane používat, až bude schválen rozpočet kompletního programu business intelligence. Oddělení IT tak stojí před dilematem, zda má prosazovat kvalitu dat, optimální postupy a zásady datové architektury. Snažte se tomuto scénáři vyhnout, ať už je plánovaný nebo nikoli. Datový sklad by měl vždy dodržovat pravidla správného řízení projektů včetně analýzy dat, návrhu datových modelů, získávání dat, mapování dat, kvality dat a postupů implementace. Provizorní systémy mají tu nemilou vlastnost, že se jich mnoho let nelze zbavit. Není nic horšího než zkušební projekt, který porušuje všechna pravidla a poté jej nelze ukončit. Dá se to přirovnat k nepozvanému hostu na večírku, který se prostě objeví a nemá v úmyslu odejít.

Tyto typy doplňků k datovému skladu obvykle vznikají jako iniciativa podnikových uživatelů, kteří nemají k dispozici potřebné prostředky, a jsou proto ochotni oddělení IT odměnit za rychlou a provizorní práci. To sice posiluje rozpočet datového skladu a umožňuje rychle získat podporu podniku, ale z dlouhodobého hlediska datový zmatek více škodí, než prospívá. Stručně řečeno: do těchto iniciativ se nezapojujte. Dejte podniku na vědomí, že všechny iniciativy datového skladu a business intelligence musí od začátku dodržovat optimální postupy.

Pokud ve vašem oboru podnikání představují datové sklady a business intelligence novinku, nemusí uživatelé chápat, jak je vybudování příslušného prostředí složité, a mohou proto správu dat podceňovat. Z nějakého důvodu se může šířit dojem, že data jsou k dispozici ve zdrojových systémech, které jsou přístupné, a stačí tedy tato data „vytáhnout“ do nové databáze, která bude určena pro funkce business intelligence. Poté mohou vykazovací nástroje snadno získat přístup k databázi a začít vytvářet sestavy. Pochopitelně jsou-li data špatná, budou podle zadání projektu opravena. Tento prostý přístup způsobuje, že mnoho projektů končí neúspěšně. Podnik totiž nedokáže vůbec ocenit související úsilí, které je předpokladem dokončení projektu.

V mnoha případech se projekty datových skladů ocitají v této situaci, když podnikoví uživatelé získají přímý vliv na zadání a mohou do něj „vpašovat“ další datové komponenty. Poté se objevují další a další, což může v nejhorším případě vést k tomu, že se objeví zcela nový zdroj dat, který vyžaduje analýzu, iniciativy kvality dat a další aktivity, jaké souvisejí s přidáním zcela nového zdrojového systému. Další nežádoucí situace může nastat, když se podnik rozhodne projekt ukončit poté, co překročí odhadované personální a finanční požadavky. To se negativně projeví na prostředí datového skladu a problém se dominovým způsobem rozšíří do dalších oddělení, která se v této fázi nechtějí na podnikovém projektu podílet. V jiných podnikových

oblastech nastává scénář typu „počkejme a uvidíme“, což může ovlivnit postup celé iniciativy podnikového datového skladu.

V jiném konkrétním případě někde mezi scénáři „s důrazem na budování“ a „datového zmatku“ se poněkud překračuje jednoduché datové hledisko. V jisté organizaci se domnívali, že když zakoupili servery, databázovou licenci, elegantní datový model od renomovaného dodavatele a nejnovější nástroje BI, postačí jim pouze načíst data a vše bude fungovat. Oddělení IT bylo přímo podřízeno finančnímu řediteli, který se zajímal o to, proč tým datového skladu nemůže jednoduše „stisknout tlačítko“ a generovat sestavy. Očekával také, že data budou dokonalá. Firma přece fungovala, takže kde by se vzaly potíže s kvalitou dat? Domníval se, že pokud by se jakýkoli problém s kvalitou dat přece jen objevil, bylo by jej možné opravit ihned po zjištění nekonzistence v sestavách. Není potřeba dodávat, že projekt narazil na mnoho překážek.

Z toho plyne poučení, že všechny projekty datového skladu by měly dodržovat optimální postupy, které byly mnohokrát vyzkoušeny. Nepoužívejte systém datového skladu jako průchozí systém k pouhému uložení dat. Datový sklad vzniká v první řadě na základě standardní terminologie, která umožní identifikovat všechny datové komponenty. Za druhé platí, že datový sklad má zahrnovat společné struktury k uložení vyčištěných dat.

Důvody, proč datový sklad NEBUDOVAT

Pro vytvoření datového skladu lze najít mnoho důvodů, např.:

- Snížení provozních nákladů
- Aktuální systém nedokáže zvládat režii
 - Kapacita serverů, objem dat nebo použití
- Samostatné systémy, které způsobují nekonzistenci v sestavách
- Potíže s terminologií, které vedou k nekonzistenci dat a sestav
- Vytvoření centrálního důvěryhodného prostředí pro analýzu dat
 - Data, terminologie, datové struktury
- Jednoduchost použití

Existuje také hodně důvodů, proč datový sklad nebudovat nebo jeho projekt odložit. Patří k nim technické, politické i praktické příčiny. V následujících odstavcích uvedeme několik důvodů, proč může být vhodné projekt datového skladu pozastavit.

Nedostatečná kvalita dat

Jestliže jsou historická data tak nekvalitní, že je není možné opravit, budování systému na uložení těchto historických dat je zbytečné. Kvalita dat v kontextu má pro podnik klíčový význam. Pokud tedy podnik data nedokáže použít, není důvod vytvářet systém, který by tato data obsahoval. Jestliže však podnik přijme strategické rozhodnutí zlepšit kvalitu dat v provozních systémech, pak je návrh systému pro budoucí použití vhodným cílem.