

Enterprise Reporting

Potreba aktuálnych a správnych informácií pre chod podniku býva primárny dôvod na realizáciu projektov budovania dátových skladov. Výsledky všetkých predchádzajúcich fáz takéhoto projektu - extrakcia z primárnych systémov, transformácia, nahrávanie dát (ETL - Extract, Transform, Load) alebo budovanie analytických datamartov - sú v konečnej fáze prezentované koncovým používateľom či už ako prostriedky na podporu rozhodovania, alebo ďalších analytických činností, ako je napríklad Data Mining. Keďže reportingová vrstva je jediné rozhranie, ktoré používa väčšina používateľov riešení Business Intelligence, treba jej venovať náležitú pozornosť.

Základné otázky

Každý projekt budovania dátového skladu sa začína analýzou požiadaviek biznis používateľov. Ich pripomienky a požadované funkcie sú zozbierané, zosumarizované a premietnuté do procesov budúceho úložiska dát. Pokiaľ sa nám podarí presne definovať potreby používateľov (ak máme šťastie, vedia ich definovať sami, ale to nebýva vôbec pravidlom), treba ich zladíť s prevádzkou plánovaného riešenia. Pokiaľ budujeme dátový sklad pre niekoľko stoviek cieľových konzumentov reportov, táto úloha sa môže stať veľmi komplexnou. V rámci dizajnu reportingového riešenia si preto musíme odpovedať na niekoľko základných otázok.

Komu reportovať?

Prípravné fázy projektu dátového skladu identifikujú budúcich používateľov systému. Je potrebné predpokladať, že potenciálnym používateľom reportingového systému nad dátovým skladom je každý zamestnanec podniku. Systém preto musí byť schopný integrovať všetkých používateľov s ohľadom na ich informačné požiadavky a predpokladaný spôsob práce so systémom. Ideálne riešenie je, ak systém umožňuje integráciu so systémom na správu používateľov (LDAP, Tivoli a iné). Veľké spoločnosti tvoria divízie, odbory alebo oddelenia, pričom reportingové požiadavky pre používateľov v rámci jednotlivých súčastí firmy bývajú podobné. Najčastejším scenárom preto býva implementácia organizačnej schémy do používateľských skupín reportingového systému. Je logické, že marketingové reporty bude využívať oddelenie predaja a account manažéri a informácie o výkonnosti pobočiek ich riadiaci pracovníci, prípadne zamestnanci. Nevyhnutnou požiadavkou je možnosť zaradiť používateľa do viacerých skupín, teda vedúci oddelenia bude mať prístup k reportom o výkonnosti svojich podriadených, ale aj k informáciám určeným všetkým riaditeľom odborov. Tento príklad ukazuje dva základné prístupy k rozdeleniu používateľov a realizáciu ich reportingových požiadaviek:

Horizontálny - rozdelenie používateľov podľa línií fungovania podniku (financie, predaj, marketing, ľudské zdroje...). Pri tomto pohľade identifikujeme spoločné oblasti dát, ktoré používateľia potrebujú.

Vertikálny - hierarchická úroveň zaradenia konzumenta reportov. Rozlišujeme reporty pre C-level manažment, stredný manažment, operatívnu, externých konzumentov a iné. V tomto prípade analyzujeme hlavne úroveň detailu/agregovanosti dát, ktorú reporty poskytujú.

Čo reportovať?

Ako sme už spomenuli, biznis používatelia v priebehu úvodnej analýzy projektu často nie sú schopní presne definovať svoje budúce dátové požiadavky. Treba preto predpokladať, že v ďalších fázach (hlavne pri dizajne reportov) bude veľa pôvodných požiadaviek prehodnotených a zmenených. Pokiaľ je použitý dátový model navrhnutý efektívne a flexibilne a zároveň máme prístup k potrebným dátam z primárnych systémov, takéto úpravy by nemali predstavovať závažný technický problém. Rozdelenie reportingových požiadaviek na horizontálne a vertikálne nám pomôže odpovedať aj na otázku, čo reportovať.

Z horizontálneho hľadiska môžeme reporty deliť podľa dátových oblastí, ktoré pokrývajú. Dá sa predpokladať, že architektúra dátového skladu alebo nadväzujúcich datamartov je navrhnutá predmetne, takže jednotlivé oblasti sú logicky rozlíšiteľné, a teda identifikácia potrebných podkladových dát je jednoduchá.

Pri vertikálnom pohľade sledujeme úroveň agregácie reportovaných dát. Vysokopostavení manažéri vyžadujú agregované dáta na korporátnej úrovni, porovnania s predošlými obdobiami, vývojové trendy. Naproti tomu, čím nižšie klesáme v rámci organizačnej hierarchie, tým viac klesá agregovanosť dát, povedzme na úroveň objemu predaja za predošlý deň, počtu nových klientov alebo telefónnych liniek odpojených z dôvodu neplatenia.

Ako reportovať?

Táto otázka je najkomplexnejšia a dala by sa rozdeliť na viaceré samostatných podotázok. V rámci rozsahu článku sa ich pokúsme identifikovať a zodpovedať v stručnejšej forme. Predovšetkým treba identifikovať jednotlivé druhy reportov z hľadiska ich funkčnosti:

Statické reporty - reporty, ktoré sú dizajnované ako statické nosiče informácií, bez ďalších špeciálnych interaktívnych funkcií. V našich končinách sa na označenie takéhoto typu dokumentov často používa aj slovo zostavy. Sú dennými podkladmi pre operatívnu prácu a rovnako dlhá je aj ich životnosť.

Špeciálnou skupinou v rámci statického reportingu sú takzvané dashboardy, ktoré umožňujú intuitívny pohľad na dáta prostredníctvom ich grafických reprezentácií, ako sú tachometre, semafore, ukazovatele teploty a podobne. Tento typ využíva hlavne manažment na sledovanie KPI (Key Performance Indicators - kľúčové ukazovatele výkonnosti podniku, ako sú výnosy, počty klientov, obrat, profitabilita atď.)

Dynamické reporty

- dynamiku v tomto zmysle chápeme ako možnosť interakcie používateľa s preddefinovaným reportom prostredníctvom filtrov, výberu dátových položiek, kalkulácie nových na základe existujúcich dát atď. Za jeden z druhov dynamického reportingu môžeme považovať OLAP reporting, ktorý umožňuje vnorenie do väčších úrovni detailu (drill-down) alebo postup smerom hore (drill-up), prípadne hlbšiu analýzu a vyhľadávanie dát (slice and dice).

Ad hoc reporty

- v tomto prípade ponechávame používateľovi voľnú ruku v tom, aké dáta si vyberie do svojho reportu. Predpokladáme pritom, že používateľ neovláda niektorý z jazykov na prístup k dátam (SQL, MDX), musíme mu preto umožniť, aby vedel jednotlivé dátové položky a podmienky definovať prostredníctvom vizuálneho rozhrania metódou point-and-click.

Tri úrovne interakcie s reportmi indikujú aj množstvo a charakteristiku používateľov, ktorí s nimi pracujú. Najčastejšie sú generované statické reporty, počet používateľov dynamických a ad hoc reportov klesá spolu so znalosťou dát a analytickou povahou pracovného zaradenia používateľov. Zároveň stúpa potenciálna náročnosť spracovania dát pre tieto reporty.

Dôležitou otázkou je forma, akou používatelia k reportom pristupujú. Vo veľkých korporáciách je takmer nemysliteľné, aby mal každý jednotlivý používateľ nainštalovaný samostatný program na prehliadanie reportov a prácu s nimi. Tu na javisko vstupujú hlavné komunikačné platformy súčasnosti, teda internet a e-mail. Prostredníctvom nich dokážeme distribuovať reporty množstvu používateľov bez potreby náročných inštalácií a zaškoľovania. Portálové systémy umožňujú integráciu korporátnych reportov na jednom mieste, pričom používatelia by mali mať možnosť nastaviť si personalizované prostredie s vlastnými reportmi, oddelenia môžu mať vlastné stránky pokrývajúce nimi využívané dáta atď. Portálové riešenia svoju opodstatnenosť v podnikovom prostredí potvrdili natoľko, že softvérové spoločnosti pracujú na spoločných štandardoch, ktoré umožnia interoperabilitu portálov a dosiahnu tým synergický efekt webového prístupu k podnikovým dátam (JSR 168: Portlet Specification).

Tento prístup však implikuje, že všetka funkčnosť, ako sme ju už opísali, musí byť dosiahnuteľná prostredníctvom webového rozhrania. Súčasný vývoj na poli internetových technológií (technológie ako napríklad AJAX) umožňuje priblížiť ich správanie a funkčnosť samostatne stojacim aplikáciám a minimalizovať tak ich rozdiely pre koncového používateľa. Koncoví používatelia majú často problém s prechodom z používaných aplikácií na webové rozhranie. Snahou dodávateľov softvéru je preto priblížiť ich webové riešenia funkčnostiam tabuľkových kalkulátorov ako Microsoft Excel. Používateľ preto môže pracovať s tabuľkami, meniť ich na grafy, používať podmienené formátovanie atď. Neoddeliteľnou súčasťou musí byť možnosť exportu dát do zvolených formátov na ďalšie spracovanie.

Pokiaľ postúpime od aktívneho prístupu používateľa k prehliadaniu reportov prostredníctvom špecializovaných intranetových portálov smerom k viac pasívnym prístupom, narazíme na technológie ako RSS (Really Simple Syndication alebo Rich Site Summary). Táto technológia umožňuje prístup k novému obsahu webového servera bez nutnosti jeho prehliadania, čím šetrí čas používateľa.

Distribúcia generovaných reportov e-mailom priamo používateľovi je veľmi efektívny spôsob distribúcie podnikových informácií. Systém automaticky spracuje dáta, vygeneruje potrebné reporty a zašle ich používateľom alebo skupinám používateľov podľa zvolených kritérií. Tento postup je praktický napríklad pri terénnych pracovníkoch, ktorí nemajú priamu možnosť prístupu do firemných sietí. Ale je aj spôsobom optimalizácie riešenia. Pokiaľ má banka

niekoľko desiatok alebo stoviek retailových úverových poradcov, je praktické zasielať im informácie o čerpaní úverov ich klientov mailom, čím šetrí portál od množstva požiadaviek a poradcov od nevyhnutnosti naučiť sa v ňom pohybovať.

Pri e-mailovej distribúcii reportov sa vynára ďalšia otázka, ktorou je formát generovaných reportov. Aj tu sa musíme oprieť o výsledky úvodnej analýzy požiadaviek biznis používateľov. Pokiaľ distribuujeme statické reporty, ideálny je napríklad formát PDF, prípadne RTF. Vzhľadom na požadovanú interaktivitu a zvyky používateľov je veľmi často využívaný aj formát Microsoft Excel.

Kedy reportovať?

Štandardné spracovanie dátového skladu prebieha na dennej báze. Od tohto intervalu sa odvíja aj periodicita generovaných reportov (samostatnou kapitolou sú Operational Data Stores, ktoré integrujú dáta z primárnych systémov v takmer reálnom alebo reálnom čase). Operatívne reporty sú generované denne, ďalšími častými frekvenciami sú týždenné, koncomesačné či výročné reporty. Treba brať do úvahy, že čím častejšia je frekvencia spracovania reportu (predpokladáme reporty za predošlý deň), tým menej podkladových dát ho tvorí a tým rýchlejšie prebehne jeho spracovanie. To, samozrejme, platí aj naopak. Generálnou požiadavkou pri operatívnych reportoch je potreba mať ich k dispozícii hneď na začiatku pracovného času. Operatívne denné reporty a ich podkladové agregácie sú teda spracúvané hneď po skončení spracovania denných dát v dátovom sklade. Menej frekventované reporty, vyžadujú

OLAP – On Line Analytical Processing

Technológia umožňujúca rýchly prístup k agregovaným multidimenzionálnym dátam prostredníctvom takzvaných kociek OLAP. Ich štruktúra pozostáva z dimenzií a faktov. Dimenzie pokrývajú jednotlivé oblasti, z ktorých agregujeme dáta (klient, produkt, čas, organizácia atď.). Jednotlivé dimenzie sú tvorené hierarchiami, úrovňami a členmi. Fakty tvoria jednotlivé dátové položky, ktoré sú agregované v priesečníkoch jednotlivých dimenzií a ich členov (predaj, náklady, počty transakcií atď.). V okamihu prístupu k dátam špecializovaný server pristupuje k agregovaným dátam, čím minimalizuje čas reakcie.

ce väčšie objemy dát, a teda aj čas, sú spracúvané následne.

Záver

Reporting je pre koncových používateľov „výkladná skriňa“ projektu Business Intelligence. Preto naň treba kľásť rovnaké kvalitatívne nároky ako na ostatné súčasti dátového skladu. Množstvo a rozmanitosť reportov generovaných v podnikovom prostredí býva značné, o to viac tu preto do popredia vystupuje potreba flexibilných riešení a integrovanej správy dátových skladov ako homogénnych prostredí.



MILOŠ HALEČKA
milos.halecka@svk.sas.com

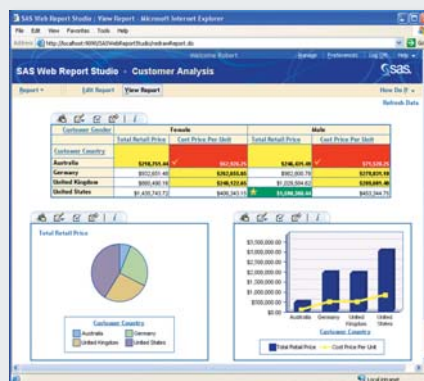
Webové riešenia na reporting

Výhodami reportingovej platformy postavené na webových technológiách sú nulové softvérové požiadavky na strane klienta – ak nerátame štandardný internetový prehliadač – a centralizovaná správa prostredia. SAS Institute v rámci svojich riešení Business Intelligence dodáva tri webové aplikácie na prístup a analýzu dát v dátových skladoch.

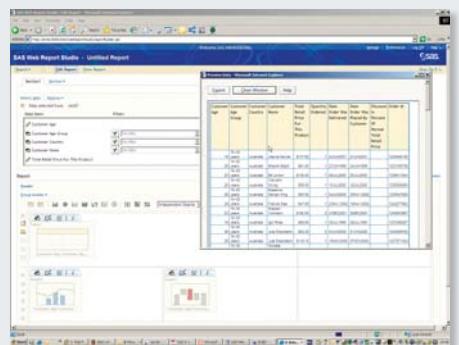
SAS Web Report Studio – aplikácia na tvorbu a distribúciu reportov. Umožňuje výber dát, dizajn tabuliek, grafov, filtrov, zoradovanie dát, podmienené formátovanie, distribúciu a časovanie spracovania reportov, prepájanie reportov, vkladanie log a množstvo ďalších funkcií. Reporty môžu byť vytvorené na relačných alebo multidimenzionálnych dátach, prípadne generované procesmi na dátových serveroch, ktoré poskytujú plnú analytickú silu SAS používateľom webového rozhrania. Služí aj ako centrálny bod na správu reportov a ich pridelovanie skupinám používateľov. Obsahuje sprievodcov na tvorbu reportov. Na jej používanie nie sú potrebné nijaké technické predpoklady a znalosti.

SAS Web OLAP Viewer – aplikácia na prehliadanie a analýzu kociek OLAP. Možnosť definovania vlastných pohľadov, záložiek, pohľadu na detailné dáta (drill-through). Dostupná je ako aplikácia pre Java servlet container a technológiu .NET.

SAS Information Delivery Portal – centrálna aplikácia pre koncových používateľov. Umožňuje publikovanie reportov tvorbu špecializovaných stránok podľa pokrývaných dátových oblastí, prehliadanie zdrojových dát,



vých oblastí, prehliadanie zdrojových dát, kociek OLAP. Používatelia si môžu obsah a vzhľad prispôbiť podľa svojich požiadaviek.



Prostredníctvom portletov poskytuje rozmanitú funkčnosť, ako prístup k rôznym typom obsahu (pdf, xls...), spúšťanie sofistikovaných procesov a analýz v dátovom sklade. Systém riadenia používateľských práv a prístupov umožňuje administrátorom riadiť prístup k jednotlivým objektom v rámci portálu, ako aj preddefinovať obsah podľa používateľských skupín. Portál slúži ako centrálny bod prístupu používateľov k dátam v rámci SAS Enterprise Intelligence Platform.