

Data Quality: Vojsko dvoch svetov má svojho víťaza

Tí, ktorí kvalitu svojich dát zveria nástrojom Business Discovery, porastú, u ostatných to bude beh na dlhé trate. A možno vôbec neuvidia cieľ.

Problémy s kvalitou dát rieši asi každý. Niektorí prežívajú v sladkej nevedomosti, občas narazia na nejakú „nezrovnalosť“ a pritom netuší, že mu v databázach tika časovaná bomba. Potom sú tu organizácie, ktoré zápasia s problémom, ako spojiť dáta uložené v rôznych zdrojoch a moduloch. Iní sa zase intenzívne venujú čisteniu dát v snahe dosiahnuť ich kvalitu ešte pred nasadením riešenia BI. Investovali obrovské množstvo energie, peňazí a času do budovania dátových skladov či rozvíjania technológií SQL. Teraz sa nestíhajú čudovať, že míňajú cieľ. Tieto tradičné nástroje totiž nedokážu dáta spoľahlivo a definitívne vyčistiť.

Podobné projekty sú síce ambiciózne, ale spravidla sa k úspešnému koncu, t. j. kvalitným a spoľahlivým dátam, nedopracujú. Treba za tým vidieť skôr snahu veľkých firiem zabezpečiť si dlhodobé a drahé zákazky.

Na dátové sklady radšej zabudnite

Business Discovery sa ukazuje ako skutočne efektívny spôsob, ako spoľahlivo odhaliť a v priebehu dní či týždňov vyriešiť problémy s kvalitou dát. Tie totiž vďaka logickému, tzv. asociatívnemu spojeniu súvislostí medzi dátami odhalia ich skutočnú kondíciu aj v oblastiach, ktoré by sme tradičnými technológiami SQL neodhalili. Tú pomôžu zlepšiť a za sebou nechajú spoľahlivé dáta, o ktoré sa dá oprieť. A ide to aj bez budovania drahých dátových skladov a rokov čistenia firemných dát.

Tieto vskutku inovatívne riešenia umožnia jednoduchým klikaním filtrovať a čistiť väzby medzi dátami a odhaliť tie chybné. Pri zavádzaní systému sa tak môžete zamerať na oblasti, ktoré „horia“. Navyše to pracovníkov motivuje k svedomitému vyplňaniu informácií, čo predtým vedome či nevedome ignorovali.

Upratovanie vopred sa preceňuje

Pri čistení dát je dôležitá nielen AKO, ale aj KEDY. Mnoho tradičných IT dodávateľov tvrdí, že prv ako nasadia riešenie business intelligence (BI), je nevyhnutné mať „upratané“ dáta. Tento prístup je preceňovaný a vonkoncom nie je nevyhnutný. Pri uplatnení filozofie business discovery sa v praxi osvedčuje pravý opak: Nasadiť BI čím skôr, „natiahnuť“ dáta, priebežne ich čistiť a potom reálne vyhodnocovať. Tým ušetríte drahocenný čas a úsilie.

S týmto prístupom máte dáta v priebehu krátkeho času upratané a zároveň môžete profitovať z kvalitného riešenia BI. Naproti tomu pri tradičnej ceste, vás to bude stáť desiatky tisíc

eur, roky „nervov“ a na výsledok sa ťažko dá spoliehať, keďže sa vám stále budú niekde objavovať nové problémy.

Firmy netušia o nekvalitných dátach

V praxi sa často stretávame s klientmi, ktorí až po napojení svojich dát na riešenie business discovery zisťujú, aký majú vo svojich databázach neporiadok. Často v nemom úžase rátajú chýbajúce, neúplné, duplicitné, nesprávne údaje či dáta, ktoré nemajú vytvorené väzby. Môže byť za tým pohodlnosť alebo nevedomosť zamestnancov, chybu však môže spôsobiť aj systém, v ktorom sú dáta uložené.

K týmto odchýlkam často dochádza pri manuálnom zadávaní dát. Na jednom z projektov, pri ktorom sme riešili reporting v rámci holdingu, sa v reporte zrazu objavili nečakane vysoké výnosy zahraničnej pobočky. Analýza odhalila, že účtovníčka nesprávne nahadzovala kurz a vždy pred poradou ho ručne upravovala. Takáto „pomoc“ nie je nič výnimočné a môže byť cestou do pekla.

Excel náš každodenný

Najväčší zdroj omylov je manuálne zostavovanie reportov v Exceli. Ten sa využíva hojne aj nad zavedenými (najmä tradičnými) nástrojmi BI. Jeho odstránenie na ceste ku kvalite dát je jeden z prvých krokov.

Neraz si pracovníci vôbec neuvedomujú svoje chyby, ktoré sa spravidla nazbierajú nezávisle od seba. Môžu tak narásť do hrozivých rozmerov. Len na ilustráciu, zmena forecastu v nemeňovanej banke spočívala v prepísaní sumy

v políčkach Excelu, často bez poznámky a neriadeneým spôsobom. Výsledkom takto upravovaných reportov sú omyly v hodnotách stoviek tisíc eur. Manuálne úpravy sú často nevyhnutné, ale mali by sa diať riadeným, nie živelným spôsobom.

Spájanie viacerých dátových zdrojov už nie je problém

Často sa stretávame s nekonzistenciou dát naprieč jednotlivými zdrojmi. Mnoho firiem využíva viacero modulov v rámci svojho IT – účtovníctvo, CRM, evidenciu výkonov, logistiku a iné. Okrem toho je tu všadeprítomný Excel. Dáta navzájom nedokážu prepojiť, keďže ich zdroje boli vytvárané individuálne a neexistuje centrálny číselník, ktorý by ich spájal. Aby sa na dáta dalo pozeráť jednotne, treba ich spojiť na jednom mieste. Tento prístup sa dá úspešne využiť aj pri migrácii dát medzi systémami.

Vo svete heterogénnych informácií je to skutočná výzva, ale len ak sa to robí tradičnou cestou. Vtedy je potrebné dátové zdroje rozložiť a následne postaviť na jednotnú úroveň. Problémom však je neustále sa meniaci štruktúra. Tá spôsobuje, že po zmenách treba dáta znovu spájať a čistiť vzťahy medzi nimi. Keďže život firiem je dynamický, nie statický, trochu to pripomína bájneho Sisyfa a jeho balvan, ktorý ho zakaždým prevaluje.

S využitím prístupu business discovery dokážete spojiť všetky súvisiace dáta z rôznych dátových zdrojov a v podstate ich tak integrovať bez toho, aby ste ich museli zladovať na úrovni databázy. Takéto riešenie možno uplatniť v priebehu dní až týždňov, nehovoriac o finančných benefitoch.

Anička nie je vanička

Iný problém, s ktorým sa často stretávame na úrovni reportov a analýz, je nejednotnosť v pojmoch. Stáva sa to, keď o tom istom probléme hovoria ľudia z obchodu, financií, prevádzky a každý si pod maržou, výkonomi či tržbami predstavuje niečo iné. Dohľad nad kvalitou v tomto prípade spočíva v zladovaní a validácii reportov a v nich obsiahnutých ukazovateľov (KPIs). Jedna vec je vyčistiť dáta z technickej



stránky a druhá je hovoriť rovnakým manažérskym jazykom.

Prístup business discovery zahŕňa aj automatický data governance nad všetkým, čo vstúpi do systému, až do momentu, keď odtiaľ vystúpi. Vidíme teda, odkiaľ dáta prišli, ako boli transformované, prepočítavané a ako sa používajú v reportoch. Takto vieme veľmi jednoducho na pár kliknutí zistiť, či sa napríklad pojem marža v rôznych systémoch nevníma ináč, či má rovnaký vzorec, či sa nepoužíva pod rôznymi názvami.

Big Brother v metadátach

Nad kvalitou reportov bdie „Big Brother“ v podobe nástrojov governance dashboard a monitoringu. Tie dokážu na základe nastavených vzťahov alebo predikčných analýz automaticky odhaliť odchýlku od normálu a upozorniť na ňu. Kliknutím na problematický záznam veľmi rýchlo zistíte, kde problém vznikol, a môžete sa k nemu prekliknúť až na úroveň dátového zdroja. Čo je na týchto automatických upozorneniach dôležité, je dynamický pohľad na kvalitu dát, ktorú tak možno udržiavať aj priebežne.

Pri tvorbe reportov je kritické zdefinovať si manažment metadát. Pri tradičnej architektúre to nie je nič jednoduché, spravidla sa o to stará špecialista, ktorý si vypýta slušný honorár. S business discovery ste však v pozícii, keď je definovanie metadát automatické, samozrejme, treba ich správne interpretovať. To je však používateľsky veľmi pohodlné a zrozumiteľné.

Inovácie majú odľahčiť biznis

Prístup business discovery dokázal zdanlivo náročnú tému, ako je dátová kvalita veľkých objemov heterogénnych dát, priblížiť bežnému používateľovi. Kvalitné dáta sú tak na dosah aj pre „bežného človeka“, ktorý s vhodným nástrojom dokáže samostatne analyzovať a kontrolovať ich kvalitu. Takýto systém riadenia kvality dát odľahčuje organizácie, ktoré tak nie sú odkázané na drahých externých špecialistov, resp. s ich pomocou dokážu dosiahnuť výrazne viac.



MARTIN KOSTIĆ
riaditeľ spoločnosti EMARK,
kostić@emark.sk