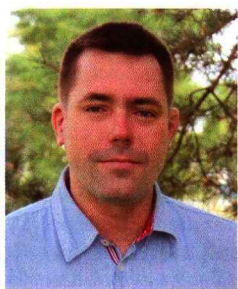


Analýza a vizualizace dat – rychlá cesta k informacím

Rozhovor poskytl Jiří Bohuslav, managing consultant a team leader DWH ze společnosti Sophia Solutions. V oblasti datových skladů a reportingu se pohybuje již dlouhých 15 let, působil na mnoha projektech v rolích konzultanta, analytika, architekta řešení, pracoval například pro společnosti Český Aeroholding, PPF banka, Lukoil, ČSOB a řadu dalších.



1. BI a analytika procházejí v posledních letech bouřlivým rozvojem. Co by se dalo označit jako největší změna v uplynulém roce?

Máte pravdu, že oblast BI je skutečně velmi rychle se rozvíjející oblastí. Za největší změnu v posledních letech považuji jednoznačně orientaci na cloudová řešení,

kteří se podle mého názoru stanou novým standardem v horizontu několika málo let.

2. Vizualizace dat představuje ideální řešení, jak představit výsledky analýzy i lidem, kteří se v analytice nepohybují, tedy manažeri či běžní uživatelé. Jaký vývoj se dá v této oblasti očekávat, třeba v souvislosti s mobilními či cloudovými řešeními?

Vizuální interpretace dat je něco, na co ještě nejsou zákazníci úplně zvyklí. Považují takové systémy za „třešinku na dortu“ jejich BI řešení, mají pocit, že se bez takovéto komponenty obejdu, a větší důraz kladou na nezbytný reporting v podobě obsáhlých sestav se spoustou položek, čísel a parametrů. My však tvrdíme, že uživatelé velmi často s vizuální analýzou vystačí, ne každý uživatel potřebuje znát detailní číslo v reportu. Pro úspěšné rozhodnutí často postačí obecná znalost trendů, jednoduše

zobrazitelné korelace mezi daty, zobrazení časových řad, aplikace správných typů grafů. Tato vizuální analýza se teprve stává odrazovým můstkem pro skutečnou detailní analýzu, kde se již bez kvalitního analytického nástroje neobejdete. Cloudová řešení takovému přístupu nahrávají, protože vizuální analýza může být přístupná odkudkoli/komukoli/kdykoli na jakémkoli zařízení (PC, tablet, mobil), přičemž již

dnes existují plnohodnotné nástroje, které práci s analytickým systémem umožní.

3. Řada organizací v současnosti sbírá všemožná data – i když pro ně v současnosti nemají velký užitek, jsou přesvědčené, že tomu tak v budoucnu díky využití analytiky bude. Existují už dnes nějaké nástroje, které by dokázaly taková data předtřídit, a zmenšit tak například velikost potřebných úložišť?

Já si myslím, že spousta firem dnes nezpracovává ani ta data, která potřebují – a to jim generuje zbytečné ztráty. V případě těch pokrokovějších organizací, které tak již činí, řešením je zamyslet se nad architekturou a celkovou infrastrukturou systémů, která data sbírají a ukládají. Ukazuje se, že ideální způsob je kombinace standardního datového skladu běžícího na výkonných databázích, který integruje pouze vybraná data z různých zdrojových systémů s relativně velkou přesností, a řešení postaveného na Hadoopu, které umožňuje sbírat velká data v surové podobě, jejichž uložení je levné (řádově levnější než v případě datových skladů). Data uložená v Hadoopu svojí povahou neprocházejí velkým čištěním ani transformacemi, nebdá se tolik na absolutní datovou kvalitu a čistotu jako v případě datových skladů, nicméně dokážou odpovědět na základní otázky, přičemž již dnes existují výkonné prostředky a nástroje pro jejich analýzu.

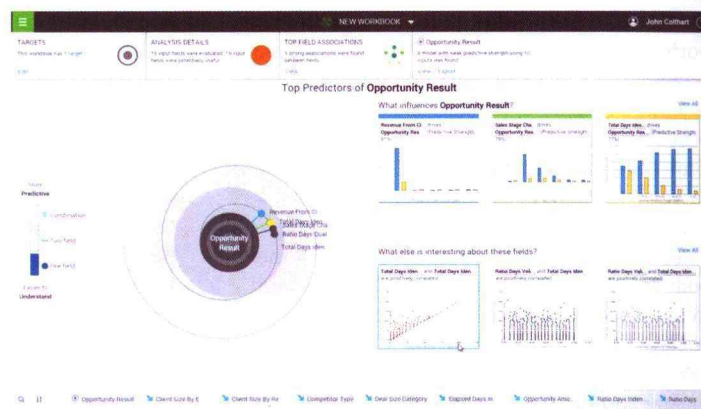
4. Potřeba analyzovat data se může v průběhu času měnit – od velkého objemu až jen po nějaký udržovací režim. Co byste poradil firmám, jejichž požadavky na analytická řešení se průběžně mění a nechtějí jít cestou alokovaní prostředků pro tu nejvyšší zátěž?

Pro takovéto firmy se právě hodí cloudová řešení. Velmi snadno lze v čase dedikovat různý datový prostor, výkon serveru posílit jen tehdy, když je skutečně zapotřebí, a toto vše řídit podle aktuálních potřeb společnosti. Zároveň je pro takové zákazníky atraktivní, že veškeré náklady jdou z provozních nákladů (OPEX), což opět umožňuje velkou pružnost a ochrana investice je tímto dobře zajištěna.

5. Jak vidíte bezpečnost dat v cloudu, nepanují obavy ohledně privátnosti dat?

Na obavy zákazníků ohledně bezpečnosti skutečně často narážíme. Paradoxně se ale ukazuje, že zabezpečení dat v cloudu je lepší než u lokálně uložených dat. Snažíme se posunout vnímání zákazníků tímto směrem a vidíme, že se to postupně daří. Je běžné, že již dnes řada firem má svá citlivá data uložena v hostingovém centru. A není pochyb o tom, že firmy časem do cloudu přesunou i svoje analytické aplikace, stejně jako tomu už je u e-mailů, sdílených úložišť a podobně.

6. V souvislosti s BI se občas uvádí, že tuto oblast může vpřed posunout technologie jako strojové učení či datamining. Jak vidíte jejich potenciál právě ve spojení s analytikou a případně jak velká cesta se musí pro jejich úspěšné zapojení ještě udělat?



Oblast dataminingu a strojového učení se poslední dobou dostává do popředí a stává se nedílnou součástí BI systémů. Zde se bavíme o rozhodovacích stromech, segmentaci zákazníků, různých prediktivních analýzách. Datamining je ve skutečnosti složitá disciplína, a to, co se změnilo za poslední roky, je, že dataminingové komponenty se stávají integrovanou součástí pokročilých analytických nástrojů. Jinými slovy, datamining nemusí být výsadou jen studovaných analytiků se znalostí matematiky a statistiky, nýbrž v omezeném, ale velmi často postačujícím rozsahu tyto metody dokáže použít běžný poučený uživatel. Toto je směr, ve který věřím, a který bude mít čím dále větší potenciál.

POMÁHÁME DELAT SPRÁVNÁ ROZHODNUTÍ
Sophia Solutions | sophia