

OLAP Databáze

V této kapitole se dostáváme ke konečnému stádiu zpracování dat, které má v praxi pravý význam pro firemní plánování a stanovování nových strategií. Dosud jsem se zabýval hlavně problematikou, která se zabývá integrací dat do celopodnikového řešení v podobě budování datových skladů. Tato část se týkala především tzv. „transakčního“ zpracování dat OLTP (On-Line Transactional Processing), které je primárně uzpůsobeno pro relační databázovou základnu, nad kterou běží klíčové aplikace a systémy podporující každodenní chod firmy. Od této chvíle se však situace mění a nadále se budu zabývat technologií OLAP (On-Line Analytical Processing), na jejíž podstatě jsou z velké části založeny aplikace Business Intelligence.

Definice OLAP databáze

Na rozdíl od transakčních databází jsou OLAP databáze určeny pro analýzu dat, která jsou za tímto účelem uložena v multidimenzionální podobě. Tato struktura zahrnuje již předzpracované agregace dat.

Data v datovém skladu jsou sice čistá, konsolidovaná, ale často velmi objemná. Abychom byli schopni tato data efektivně analyzovat, často se ukládají do speciálních datových struktur typu OLAP. Technologie OLAP nám dává možnost prohlížet sumarizované údaje o nákupech, prodeích, apod. v rozpadu podle libovolné dimenze či kombinace více dimenzí (např. časová nebo produktová dimenze) a na libovolném stupni agregace (např. stát – region – okres – město – pobočka) a to v reálném čase bez nutnosti čekat na zdlouhavé procházení celé historie datového skladu. Tímto dostáváme do rukou výkonný analytický a reportovací nástroj oblíbený zejména u manažerů.

Uložení dat v OLAP databázích

Existuje několik způsobů, jakými lze data v OLAP databázích ukládat. Liší se hlavně velikostí výsledné databáze a způsobem, jakým mohou být tato data využívána. Jedná se o následující varianty:

- MOLAP** (Multidimensional OLAP) – asi vůbec nejrozšířenější způsob uložení dat, který nabízí nejvyšší dotazovací výkon ze všech zde uvedených metod. Data jsou uložena v optimalizované multidimenzionální databázi, kde se nachází všechny potřebné agregace.
- ROLAP** (Relational OLAP) – poskytuje uživatelům multidimenzionální zobrazení dat, která však zůstávají uložena v původní relační databázi, což poskytuje vyšší úroveň škálovatelnosti a rychlejší dobu odezvy.
- HOLAP** (Hybrid OLAP) – hybridní uložení dat, které je kombinací předchozích dvou variant a snaží se maximalizovat jejich výhody. Data jsou ponechána v původních relačních tabulkách a agregace jsou uloženy v multidimenzionální podobě. Toto řešení tedy poskytuje propojení mezi velkými objemy dat v relačních tabulkách a zároveň nabízí výhodu rychlejšího zpracování multidimenzionálních agregací.
- DOLAP** (Desktop OLAP) – nejmladší technologie, která umožňuje uživateli stáhnout si požadovanou podmnožinu kostky z OLAP databáze na lokální disk a provádět nad ní analytické operace

Struktura OLAP databáze

Tabulky vytvořené v OLAP databázích jsou oproti transakčním databázím uloženy v denormalizované podobě, nesplňují tedy podmínky třetí normální formy. Tyto tabulky lze rozdělit na dva základní druhy:

Tabulky faktů - jedná se o nejobjemnější tabulky v databázi. Pod pojmem fakta (measures) si můžeme představit číselná vyjádření měrných jednotek obchodování. Jako příklad lze uvést počet prodaných kusů daného zboží, případně zisk z tohoto prodeje. Kromě měrných jednotek obsahují tyto tabulky ještě cizí klíče tabulek dimenzí, pomocí kterých jsou k nim dimenze napojené.

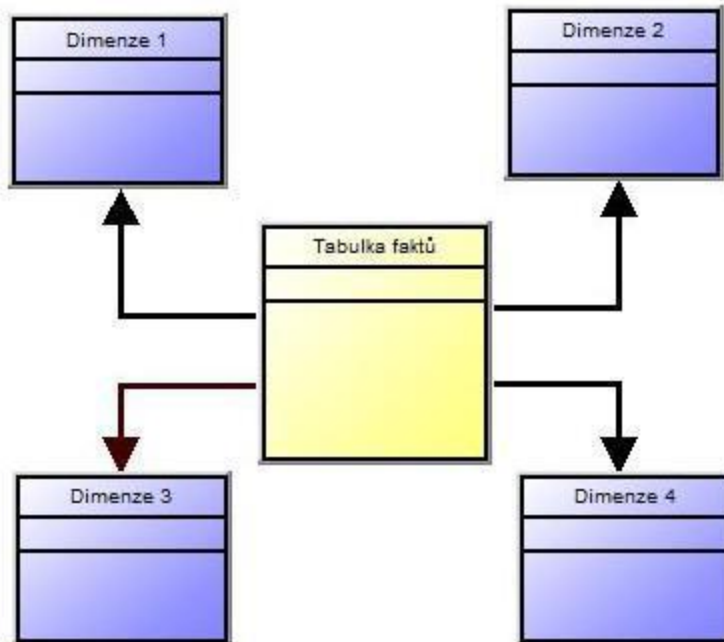
Tabulky dimenzí – obsahují logicky nebo organizačně uspořádané údaje, které popisují různé aspekty obchodování. Prakticky jde o jakousi obdobu podnikových číselníků, protože jejich účel i obsah jsou velmi podobné.

„Tabulky dimenzí vysvětlují všechna „proč“ a „jak“, pokud jde o obchodování a transakce prvků. Zatímco dimenze obecně obsahují relativně stabilní data, dimenze zákazníků se aktualizují častěji. Nejčastěji se používají časové, produktové a geografické dimenze.“

Tabulky faktů a dimenzí mohou tvořit různá topologická uspořádání, která se liší především rozdílností implementace hierarchií v jednotlivých dimenzích. Více viz. následující podkapitoly.

STAR schéma

Jedná se o tzv. „hvězdicové schéma“, které se skládá z centrální tabulky faktů a okolních dimenzí. Každá dimenze je vždy tvořena právě jednou tabulkou dimenzí a mezi těmito tabulkami nenajdeme žádné vzájemné relační propojení. Dimenze v tomto schématu nejsou normalizovány, což často logicky vede k redundantnímu uložení dat a větší zátěži dimenzí samotných. Tato denormalizace je zapříčiněna především výskytem hierarchického uspořádání dat v jednotlivých dimenzích.



Obrázek č. 5: STAR schéma, vlastní zpracování

Toto schéma má následující výhody a nevýhody

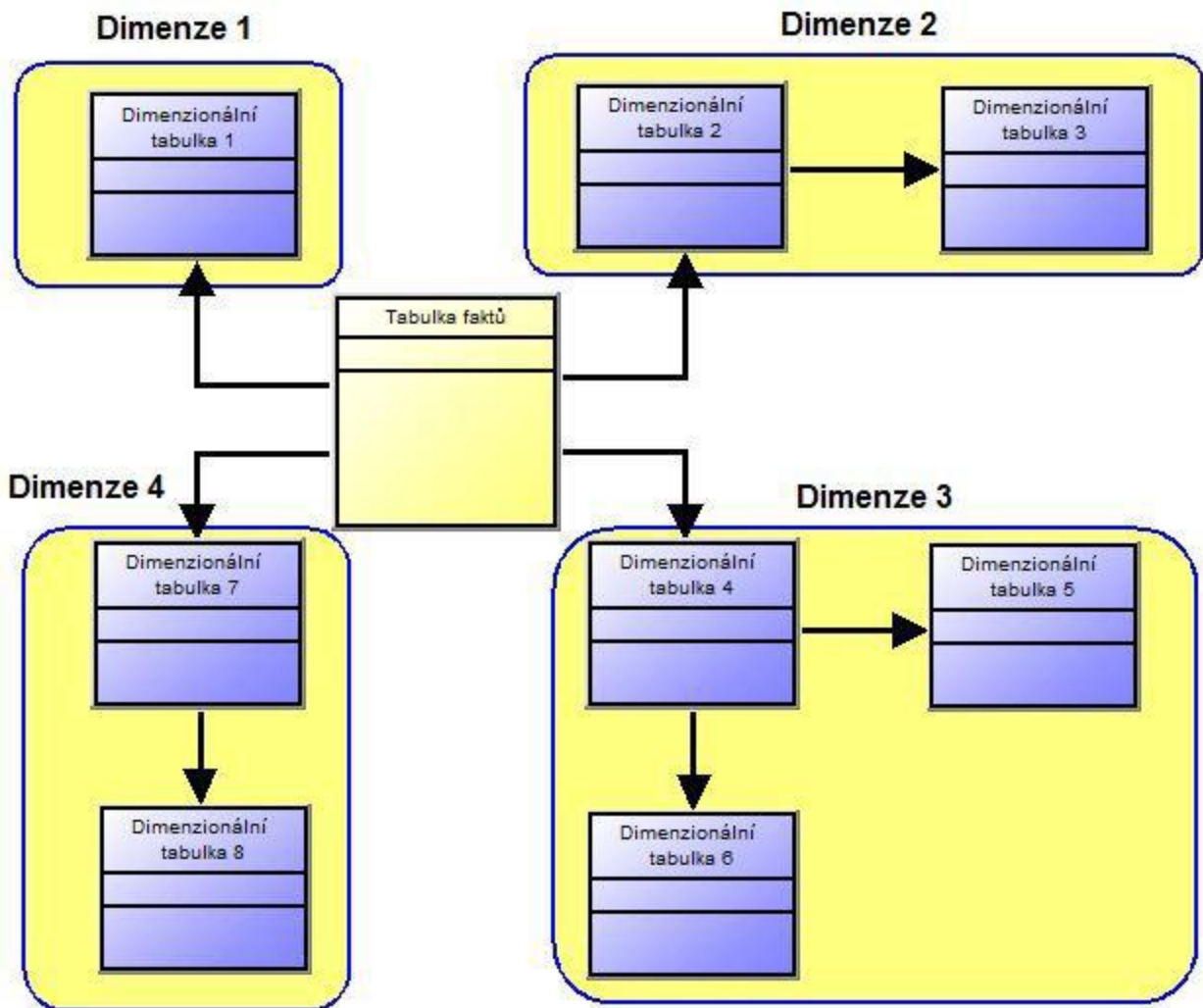
□□ **Výhody** – vysoký dotazovací výkon

□ **Nevýhody** – v důsledku nenormalizovaných dimenzí je vytvoření takového modelu relativně pomalé

SNOWFLAKE schéma

Toto schéma také označováno jako tzv. „sněhová vločka“. Na rozdíl od hvězdy má o něco složitější strukturu, protože každá dimenze je obvykle tvořena více vzájemně propojenými dimenzionálními tabulkami, které jsou normalizovány za účelem snížení redundance dat v jednotlivých dimenzích. Výsledný nárůst velikosti databáze není tedy tak rapidní jako v případě hvězdicového schématu.

Schéma sněhové vločky lze znázornit následujícím příkladem:



Obrázek č. 6: SNOWFLAKE schéma, vlastní zpracování

I toto schéma má své následující výhody a nevýhody

□□ **Výhody** – rychlé zavedení údajů do normalizovaných tabulek

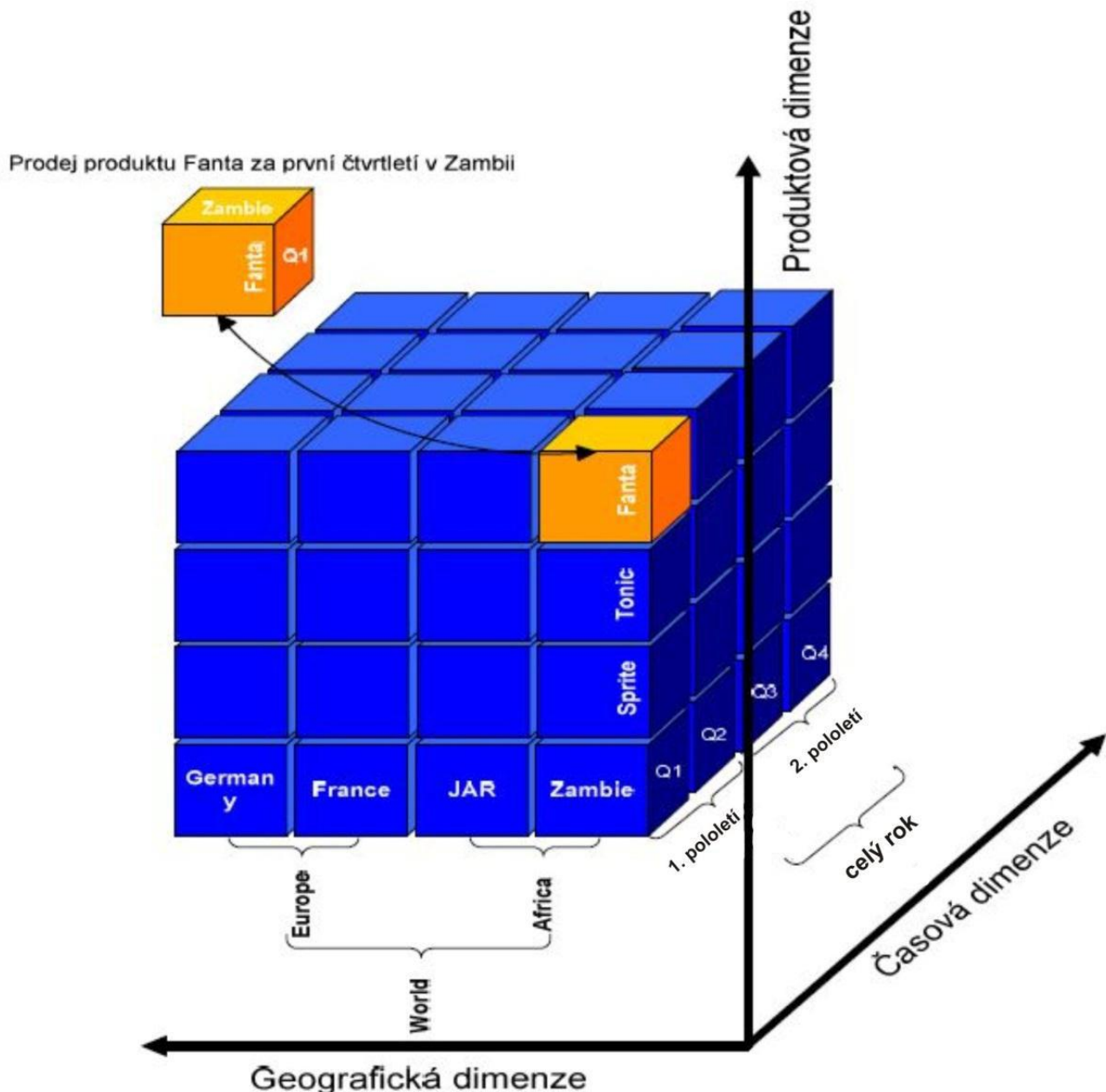
□ **Nevýhody** – v důsledku nenormalizovaných dimenzí je vytvoření takového modelu relativně pomalé

OLAP datová kostka

Pojem OLAP kostka (OLAP cube) vychází přímo z koncepce multidimenzionální databáze. Prakticky se dá říct, že jedna taková databáze může být základem pro více datových kostek (kostka je tedy podmnožinou OLAP databáze) a záleží pouze na pohledu, ze kterého na data v ní obsažená nahlížíme. Toto pojmenování vychází ze skutečnosti, že se jedná o vícerozměrnou tabulku, jejíž struktura je tvořena daty, která pocházejí z jedné nebo více tabulek faktů a informacemi prezentovanými formou dimenzí.

Podle L. Lacka [9, s. 390] je definice OLAP kostky následující: „*Vícerozměrná kostka jako soubor skupin měřítek je prostředkem pro kombinace faktů s různou granularitou. Z hlediska interpretace se jedná o jakousi virtuální strukturu či kontejner pro skupinu měřítek, dimenzí, politik, nebo jinak řečeno, je to spíše logická struktura nežli fyzické údaje.*”

Díky multidimenzionální struktuře datové kostky máme možnost pohlížet na realitu z několika možných úhlů pohledu. Pro lepší názornost nám poslouží následující obrázek:



Obrázek č. 7: Princip datové kostky, převzato z [11]

Datová kostka, znázorněná na obrázku č. 7, je pro snadné pochopení této problematiky tvořena třemi dimenzemi. V praxi však žádné omezení pro počet dimenzí v kostce neexistuje a není tedy neobvyklé, aby byla tvořena třeba i několika desítkami dimenzí. Grafické znázornění takovéto struktury by bylo velmi složité a nepřehledné.

Operace spojené s OLAP kostkou

Protože se v případě datové kostky jedná o vícerozměrnou strukturu, musí také existovat způsob, jakým na požadovaná data nahlížet a analyzovat je. Pro tuto potřebu existují speciálně navržené operace, pomocí kterých lze s kostkou pracovat. Jedná se o tyto následující operace:

- Drill-down** – umožňuje sestoupit na nižší úroveň agregace a dostat se tak k více detailním informacím.
- Roll-up** – jde o opak operace drill-down, který umožňuje přechod do vyšších agregačních úrovní z úrovně aktuální.
- Pivoting** - umožňuje „otáčet“ datovou krychli, tj. měnit úhel pohledu na data. Díky této operaci lze pohlížet na data z pohledu různých dimenzí.
- Slicing** – umožňuje provádět tzv. „řezy“ kostkou, tj. nalézt pohled, v němž je jedna dimenze fixována v určité agregační úrovni. Jinými slovy, tato dimenze aplikuje filtr na instance příslušné agregační úrovně dané dimenze.
 - Dicing** – jedná se o obdobu slicing, jenž umožňuje nastavit filtr pro více dimenzí.

Reporting

Reporting se dá popsat jako činnost, která slouží k získávání a interpretaci informací, jež jsou vyžadovány firmou pro podporu rozhodování. Nosným základem reportingu obvykle není nic jiného, než dotazování se do databáze pomocí jejího standardního rozhraní. Přestože je často reporting prováděn i nad transakčními databázemi, rozhodl jsem se jej zařadit do této kapitoly, protože se přeci jen jedná o analytický typ činnosti a má tak s aktuálně probíranou tematikou mnoho společného.

Ať už chceme znát stav a vývoj kritických aspektů ve firmě nebo efektivitu podřízených a výroby produktů, nabízí se nám rozdílné situace z pohledu množství potřebných dat, rychlosti rozhodovacího procesu a potřeby plánování.

Co je však v těchto i ostatních situacích společné je to, že se potřebujeme rychle a správně rozhodnout, aniž bychom znali celý proces Business Intelligence. Důležitý je pro nás až validní výsledek. Reportingem tak lze nazvat právě tento poslední stupeň celého procesu. Prostřednictvím reportingu získá oprávněná osoba informace ve formě, která je mu graficky příjemná, funkcionálně dostatečná a se snadnou obsluhovatelností tak, aby nebyl uživatel zdržován a zahlcován úkony, které nejsou bezpodmínečně nutné

Existuje několik typů reportingu, které se liší způsobem vytvoření, využitím a svým přínosem

Statický reporting

Je vhodný zejména pro vizualizaci informací standardní struktury a vzhledu s takřka neměnnými vstupními parametry. Hodí se dobře pro finanční výkaznictví, přehledy o prodejích produktů nebo automaticky pravidelně zasílané reporty e-mailem. Výhodou tohoto typu je, že spotřebitel informací získá informace „jedním klikem“.

Dynamický reporting

Jeho charakter je podobný statickému reportingu, s tím rozdílem, že uživatel může ovlivňovat obsah a formu reportu zadáváním vstupních parametrů. Tento typ je vhodný pro přehledy vztahující se k předem neznámým časovým obdobím, kategoriím produktů a zákazníků, v případě potřeby lze částečně ovlivnit i design a formu samotného reportu. Výhodou dynamického reportu je přizpůsobení reportu potřebám konkrétního uživatele.

Ad hoc reporting

Jestliže si uživatel nevystačí s jedním z výše zmíněných typů reportů, má možnost si vytvořit report dle vlastních požadavků. Ad hoc reporty se hodí v situacích, kdy je těžko dopředu určit, jaký obsah a formu má daný report splňovat, případně tyto informace ještě vůbec nejsou známy. Výhodou ad hoc reportů je jejich nezávislost na vývojářích reportovacích systémů (uživatel si může report vytvořit sám) a možnost vytvořit si report až ve chvíli, kdy jeho potřeba nastane a budou známy všechny informace potřebné k definici reportu.

Význam Business Intelligence pro malé a střední podnik

Nyní se dostávám k závěrečné kapitole teoretické části. Doposud jsem se zabýval vysvětlením pojmu Business Intelligence a podrobným popisem jeho jednotlivých komponent a procesů, které díky své provázanosti tvoří jeho vrstvy. Na základě těchto teoretických poznatků bych rád nastínil možnosti jeho využití v podmínkách malých a středních podniků, protože za svou poměrně krátkou historii existence se tento pojem stal výsadou především velkých společností. Není se ani čemu divit, protože právě tyto společnosti většinou disponují velkými objemy dat, pro jejichž analýzu a následné přetvoření ve smysluplné informace hraje BI klíčovou roli. Správné zavedení technologie BI totiž přináší většině velkých podniků i přes nezanedbatelné počáteční náklady do její implementace zvýšení celkového obrátu a ziskovosti. Zda se však zavedení technologie BI vyplatí i pro podniky menšího charakteru je předmětem této kapitoly.

Jako první předpoklad pro implementaci BI by si měl podnik uvědomit, zda dokáže nashromáždit dostatečné množství dat, aby mělo smysl je dále analyzovat a hledat v nich skryté závislosti [3]. Jestliže je objem podnikových dat v tak malém rozmezí, že je možné z nich vyvozovat dostatečně přesné závěry bez nutnosti využití pomocných nástrojů k tomuto účelu navržených, nebude zde patrně důvod, aby se podnik touto otázkou nadále zabýval.

Dalším důležitým aspektem jsou omezené finanční prostředky, které jsou tyto firmy obvykle schopny vynaložit na realizaci takovýchto řešení. To ovšem automaticky neznamená, že by se neměly pokoušet alespoň určitou část BI technologií implementovat. Nakonec není přeci ani nutné, aby se vydávaly směrem komplexního budování datového skladu, jehož plný potenciál by vzhledem k jejich potřebám pravděpodobně zůstal nevyužitý (nemluvě o velké finanční a časové náročnosti takového projektu). Tento problém interpretuje ve své dokumentaci firma SAP následovně:

„Organizace středních velikostí mají omezené zdroje. Řešení BI by mělo nabídnout nízké náklady na vlastnictví prostřednictvím standardní integrace. Dále by mělo poskytnout šablony a nástroje, které zjednodušují a urychlují vytváření vlastních výkazů, dotazů a dashboardů, a eliminují tak potřebu najímat drahé externí konzultanty.

Řešení BI by mělo zahrnovat důležité komponenty, aby nebylo nutné nakupovat další software.“

Výběr vhodného řešení tak závisí na důkladné analýze těchto potřeb, které jsou pro konkrétní podnik nejzásadnější.

Na jejich základě je pak třeba formulovat požadavky pro aplikaci odpovídajícího řešení, které může být v budoucnu rozšířeno o další prvky. Různé komerční varianty, zaměřené pro potřeby malých a středních podniků, dnes nabízí již většina hlavních dodavatelů těchto služeb, jako např. Microsoft, Oracle, IBM, SAP, SAS a další. V současnosti již existují i poskytovatelé několika open source BI řešení (např. JasperSoft, Pentaho), která se možná nemohou svou komplexností zcela rovnat jejich výše zmíněným konkurenčním produktům, pro potřeby menších podniků však mohou být zcela dostačující.

Musím přiznat, že já sám jsem byl ještě poměrně nedávno skeptický vůči významu BI pro menší podniky, aniž bych se nad tím nějak do hloubky zamýšlel. Byl jsem přesvědčen, že přínos BI technologií bude u těchto podniků zanedbatelný v porovnání s investovaným časem a kapitálem. Tato investice je však úměrná rozsahu konkrétního řešení a logicky tak bude pro potřeby menšího podniku nižší nebo téměř nulová. To ostatně potvrzuje již výše zmíněná nabídka komerčních i open-source řešení, která odpovídá požadavkům této cílové skupiny.

Osobně si tedy myslím, že určitý potenciál Business Intelligence pro malé a střední firmy zde je a záleží jen na nich, jakým směrem se vydají.