

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



**Diplomová práce**

**Business Intelligence**

**Pavel Šefrámek**

© 2015 ČZU v Praze

# ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Katedra informačních technologií

Provozně ekonomická fakulta

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Pavel Šefrámek

Podnikání a administrativa

Název práce

**Business Intelligence**

Název anglicky

**Business Intelligence**

---

### Cíle práce

Hlavním cílem práce je analyzovat současné možnosti využití technologií z oblasti souhrnně nazývané jako Business intelligence (BI) v sektoru zemědělství (zejména jednotlivci a malé a střední podniky), především pak najít takové výstupy aplikace postupu BI (především grafické ukazatele), které přímo pomohou v rozhodovacím a řídicím procesu a v důsledku pak povedou k dosažení lepších ekonomických výsledků podniku.

### Metodika

Teoretická část diplomové práce je založena na studiu a analýze odborných informačních zdrojů věnujících se oblasti BI. Na základě získaných poznatků jsou potom v druhé, praktické části, syntetizována dílčí nestrukturovaná data do přehledných ukazatelů, které přináší jasnou, rychlou a přehlednou orientaci ve zkoumaném problému.

## Doporučený rozsah práce

50 70 stran

## Klíčová slova

Business intelligence, BI, informace, analýza, graf, podnikání, rozhodování, řízení, ukazatel

---

## Doporučené zdroje informací

- DOHNAL, Jan POUR, Jan. Řízení podnikové informatiky a podpora byznysu. Systémová integrace : časopis České společnosti pro systémovou integraci. 2013. v. 20, no. 2, p. 7–17. ISSN 1210-9479
- NOVOTNÝ, Ota POUR Jan SLÁNSKÝ David. Business Intelligence. Jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada Publishing, 2005. 254 p. Management v informační společnosti. ISBN 80-247-1094-3.
- POUR, Jan MARYŠKA Miloš, NOVOTNÝ Ota. Business intelligence v podnikové praxi. Praha: Professional Publishing, 2012. ISBN 978-80-7431-065-2.
- SALLAM, Rita L., et al. Magic quadrant for business intelligence platforms. Gartner Group, Stamford, CT, 2011.
- TRUJILLO, Juan MATÉ, Alejandro. Business intelligence 2.0: a general overview. In: Business Intelligence. Springer Berlin Heidelberg, 2012. p. 98-116.

---

## Předběžný termín obhajoby

2015/06 (červen)

## Vedoucí práce

Ing. Jan Tyrychtr, Ph.D.

Elektronicky schváleno dne 31. 10. 2014

**Ing. Jiří Vaněk, Ph.D.**

Vedoucí katedry

Elektronicky schváleno dne 11. 11. 2014

**Ing. Martin Pelikán, Ph.D.**

Děkan

V Praze dne 26. 03. 2015

### Čestné prohlášení

Prohlašuji, že svou diplomovou práci "Business Intelligence" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 26. 3. 2015

---

## Poděkování

Rád bych touto cestou poděkoval panu Ing. Janu Tyrychtrovi, Ph.D. za obětavou pomoc a cenné rady při vypracování této diplomové práce.

# Business Intelligence

---

## Business Intelligence

### Souhrn

Práce je zaměřena na oblast zpracování velkého množství dat pocházejících z primárních produkčních systémů podniku pomocí analytických konceptů, metod a postupů prováděných za pomoci výpočetní techniky, které se souhrnně nazývají jako Business Intelligence. V práci je rešeršně vymezena oblast Business Intelligence, její historický vývoj, související teoretické pojmy a dále vysvětlení celkového postupu aplikace Business Intelligence na zdrojová data podniku. Na základě teoretické práce jsou v praktické části aplikovány koncepty Business Intelligence a navrženo analytické řešení pro internetový obchod se zemědělskými komoditami. Řešení je realizováno prostřednictvím analytického nástroje PowerPivot, který je součástí tabulkového kalkulátoru Microsoft Excel.

### Summary

This work is about processing big data through computer that comes from various production systems which can be found in today's company infrastructure. These concepts and methods are called Business Intelligence. Business Intelligence is discussed thoroughly from its meaning through historical development of term itself then defining related important concepts and finally clarifying whole process of Business Intelligence. Acquired knowledge is then used in case study that shows use of the Business Intelligence technology in on line store selling agricultural commodities. Demonstrated solution relies on a self-service Business Intelligence tool called PowerPivot which is an optional add-in for Microsoft Excel spreadsheet application.

**Klíčová slova:** Business Intelligence, BI, samoobslužné BI, ETL, PowerPivot, analytika

**Keywords:** Business Intelligence, BI, Self Service BI, ETL, PowerPivot, analytics

## Obsah

Obsah.....	2
1. Úvod.....	3
2. Cíl práce a metodika.....	6
3. Business Intelligence.....	8
3.1 Proces BI.....	13
3.1.1 Tvorba zdrojových dat v produkčních systémech.....	14
3.1.2 Převod informací z produkčních systémů do datového skladu.....	21
3.1.3 Tvorba datových tržišť dle požadovaných oblastí.....	25
3.1.4 Prezentace analytických výstupů.....	26
3.1.4.1 Sestavy.....	27
3.1.4.2 Dotazy.....	27
3.1.4.3 OLAP.....	28
3.1.4.4 Ovládací panely a přehledy výsledků.....	30
3.2 Nástroje BI.....	32
3.3 Doporučení pro implementace BI.....	34
3.4 Nejnovější trendy v oblasti BI.....	36
4. Praktická část.....	38
4.1 Aplikace PowerPivot.....	38
4.1.1 Hardwarové nároky aplikace PowerPivot.....	39
4.1.2 Instalace aplikace PowerPivot.....	40
4.2 Představení internetového obchodu.....	41
4.2.1 Obchodní platforma.....	42
4.2.2 Produkční zdrojový systém.....	42
4.2.2.1 Uložení dat v produkční databázi.....	43
4.3 Transformace produkčních dat do analytické databáze (ETL).....	46
4.4 Analytická práce s importovanými daty.....	50
4.4.1 Tvorba kontingenční tabulky.....	54
5. Závěr.....	58
6. Seznam použitých zdrojů.....	60

# 1. Úvod

Každý podnikatel, který chce dlouhodobě působit na trhu, si musí zajistit dlouhodobou a poměrně stabilní prosperitu a budoucnost, a to prostřednictvím jedné ze základních funkcí podniku, kterou je tvorba zisku. K tvorbě zisku rozhodující měrou přispívá nalezení a postupná aplikace především takových postupů a procesů, které podniku dlouhodobě poskytnou jistou konkurenční výhodu.

Vlastnictví konkurenční výhody je proto zásadním předpokladem pro udržení a pokračování životního cyklu podniku. Úspěšnost získání a hlavně schopnost udržení konkurenční výhody v delším časovém horizontu (což je pro podnikatele klíčové) ovlivňují především výchozí (zdrojové) předpoklady, které svým způsobem pomáhají utvářet celkové chování podniku. Tyto výchozí předpoklady jsou souhrnně nazývány jako podnikové zdroje.

Jedna z definic podnikových zdrojů je vymezuje (Tichá et al., 2002) „*jako zásobu použitelných faktorů, které jsou vlastněny nebo kontrolovány podnikem. Zdroje jsou transformovány do konečných výrobků nebo služeb za použití široké škály aktiv a mechanismů jako jsou technologie, manažerské informační systémy (MIS), pobídkové systémy, vztahy mezi managementem podniku a jeho zaměstnanci a dalšími. Zdroje se tedy skládají z materiálních (např. budovy, stroje a zařízení) a finančních aktiv, lidských zdrojů a know-how, které může být předmětem obchodu (např. patenty, licence) a dalších nehmotných zdrojů:*

- *zkušenost zaměstnanců,*
- *organizační učení,*
- *značka,*
- *reputace.*

Tato definice říká, že jediné dva zdroje, které podnikatelská jednotka v hospodářské soutěži má, jsou hmotné a nehmotné zdroje. Pro úspěšné působení podniku ve stále se



zostřujícím konkurenčním prostředí vysoce globalizovaného trhu 21. století vyvstává logická otázka, jakými zbraněmi s konkurenční silou efektivně a spolehlivě bojovat.

Vlivem všudypřítomného technického pokroku došlo ke stlačení cen běžně dostupných výrobních faktorů (nepřihlíží se ke specializovaným, často na zakázku vyrobeným výrobním faktorům), kdy hlavní otázkou jejich pořízení většinou bývá dostupnost kapitálů. Ten díky nízkým cenám faktorů podnik vlastní, případně si je může bez značných obtíží opatřit např. na kapitálových trzích. Z tohoto důvodu je potřeba najít oblasti, ve kterých se podnikatel může odlišit takovým způsobem, aby získal konkurenční výhodu.

Z tohoto důvodu se nabízí jako klíčová oblast nehmotné podnikové zdroje. Protože je významná část nehmotných zdrojů generována vnitropodnikovými procesy a vyvíjí se prostřednictvím vnitřního procesu učení se (Tichá et al., 2002), spočívá jejich výhoda v tom, že výsledky jsou z větší části přímo závislé a ovlivnitelné samotným podnikatelem.

Jednou z konkurenčních výhod je kvalitní práce v oblasti lidských zdrojů. Pokud podnik svými schopnostmi dovede pro sebe získat lepší lidské zdroje (ve smyslu produktivity práce/přidané hodnoty zaměstnance) než druhý, přinese mu to jistě požadovanou konkurenční výhodu. Vyšší úroveň produktu při neměnných nákladech totiž obvykle znamená vyšší zisk, který je z dlouhodobého hlediska pro trvání podniku klíčový. Největší nevýhodou lidských zdrojů je jejich omezená dostupnost, kdy existují nepřekročitelné stropy růstu.

Druhou oblastí nehmotných podnikových zdrojů, ze které může podnikatel čerpat, a která je předmětem této práce, jsou podniková data. Právě výtěžnost dat, která podnik vlastní, nebo která jsou mu dostupná, ale zatím neviděl potřebu tato data sbírat, protože si neuvědomuje souvislosti a pozitivní důsledky, které mu mohou tato data v konečném důsledku přinést, je klíčovým ukazatelem podnikatelského úspěchu. Spousta podnikatelů sice potřebná data sbírá, ale jelikož nemá zpracován promyšlený systém jejich využití, nedokáže v datech nalézat potřebné souvislosti.

Proto je kladena vysoká důležitost na schopnost data správným způsobem vytěžovat, neboli porozumět jim a zpracovat je na hodnotné informace. Právě onou pomyslnou čarou

dělicí nestrukturovaná (nezpracovaná) data od hodnotných informací je účelný proces zpracování.

Cestou k zvyšování výkonnosti a také ziskovosti pomocí efektivní a přínosné analýzy podnikových dat může být např. použití konceptů, metod a postupů, které se souhrnně označují anglickým souslovím Business Intelligence (dále v textu nazýváno i jako „BI“). Tyto postupy poskytované informačními technologiemi mohou při správné spolupráci managementu a oddělení výpočetní techniky podniku dlouhodobě přinášet požadované ekonomické i mimoekonomické efekty, úspěšnost na trhu, konkurenceschopnost a konkurenční výhody (Dohnal et al., 2013).

Požadovaným výstupem aplikace poté zjednodušeně řečeno bude přetvoření na první pohled nic neříkajících nesouvislých dat na klíčové informace typu:

- Kdo je můj největší odběratel z hlediska počtu objednávek, podílu na celkovém obratu, či zisku a jak se se jeho pozice vyvíjí s časem?
- Který produkt z mého portfolia zákazníci nejvíce reklamují a který dodavatel/výrobce je za jeho dodání zodpovědný?
- Jaké jsou průměrné útraty zákazníků ve všech mých pobočkách v průběhu roku, a mám k tomuto adekvátně uzpůsobeno skladové hospodářství těchto poboček?
- Jak se vyvíjí plánovaný prodej, jak si vedou pracovníci zodpovědní za jeho vykonávání, kde se nacházejí úzká místa projektu a jak můžeme minimalizovat případné chyby?

Právě na tyto a další jim podobné otázky, o kterých současný odpovědný pracovník možná ani neví a jejichž znalost výrazně usnadňuje a zkvalitňuje celkový proces rozhodování, se snaží BI najít odpovědi.

## 2. Cíl práce a metodika

Diplomová práce se zaměřuje na rozvíjející se trend analytické práce s podnikovými daty prostřednictvím výpočetní techniky, souhrnně nazývaný jako Business Intelligence. Hlavním cílem diplomové práce je popsat teoretická východiska oblasti Business Intelligence a následně navrhnout uživatelské analytické výstupy. Dílčí cíle práce jsou:

- vytvořit kritickou literární rešerši Business Intelligence, produkčních zdrojových systémů, transformace ETL, datových tržišť a skladů, uživatelských analytických výstupů,
- identifikovat nástroje Business Intelligence,
- navrhnout řešení Business Intelligence pro internetový obchod se zemědělskými komoditami.

Svým uspořádáním je práce v podstatě rozdělena do dvou tématických oddílů – teoretického a praktického. První teoretická část se zabývá problematikou celého postupu Business Intelligence na základě literární rešerše. Teoretická část začíná historií odvětví, následuje definice samotného Business Intelligence, poté přechází k detailnímu popisu procesu jako takového. V rámci jednotlivých podkapitol jsou také objasňovány různorodé teoretické pojmy, které jsou nezbytné k následnému pochopení praktické části. V rámci teoretické části je využita metoda studia specifických literárních pramenů, především odborných publikací několika zahraničních, zejména však nejvýznamnějších českých odborníků v této oblasti a v neposlední řadě také rozličných elektronických zdrojů.

Teoretická část slouží jako nutné znalostní východisko pro aplikaci získaných poznatků v části praktické, která zároveň při své aplikaci přináší přidanou hodnotu ve formě zkvalitnění podnikového rozhodování a řízení systémem Business Intelligence. Praktická část diplomové práce znázorňuje návrh řešení pro internetový obchod se zemědělskými komoditami, konkrétně krmivem pro hospodářská zvířata. Nejprve je využito metody modelace zdrojové databáze pomocí diagramu E-R, následně metody transformace dat (ETL) z produkční databáze do analytické databáze a v konečné fázi

statistické metody zpracování velkého množství dat dotazovacím jazykem DAX, přičemž výsledky agregace jsou prezentovány kontingenční tabulkou.

### 3. Business Intelligence

Úplně poprvé použil slovní spojení Business Intelligence (již v roce 1865) Richard Millar Devens ve své satirické encyklopedii *Cyclopædia of Commercial and Business Anecdotes*. Popsal BI jako schopnost vytěžit ve svůj prospěch informace získané dříve, než se o nich dozví konkurence (v češtině bychom mohli volně přeložit tuto definici jako mazanost). První definici, která dávala BI do souvislosti s využíváním výchozích dat k vyvozování hodnotných závěrů, uvedl v roce 1958 německý výzkumník významné americké softwarové společnosti International Business Machines Corporation (obecně známou spíše pod svou zkratkou IBM) Hans Peter Luhn. Ve svém článku nazvaném *Systém Business Intelligence* (angl. *A Business Intelligence System*) jako první na světě definoval pojem Business Intelligence jako schopnost vnímat vzájemné souvislosti mezi danými fakty takovým způsobem, který zformuje kroky vedoucí k vytouženému cíli (Luhn, 1958).

Od této poměrně vágní, ale podstatu poměrně přesně vystihující definice uběhlo mnoho let až k současnosti. Díky prudkému rozvoji výpočetní techniky a všeobecné dostupnosti počítačové sítě Internet se upnul odpovídající zájem k oblasti BI a zejména k možnosti její praktické aplikace v každodenním podnikovém řízení. Zároveň rozšíření těchto technologií vede k takové dostupnosti BI, kterou si může dovolit v podstatě opravdu každý.

Laberge (2012) ostatně na vrub technické vyspělosti ve svém díle uvádí, že pokud chtěl podnikový uživatel koncem 70. let zodpovědět dotaz, který vyžadoval analýzu podnikových dat, musel nejprve sepsat podrobnou žádost a následně ji odeslat vnitropodnikovou poštou oddělení výpočetní techniky.

Tento jednoduchý dotaz následně vyžadoval spoustu telefonických spojení, během nichž si obě strany postupně vyjasňovaly jednotlivé kroky a také základní datový model, pomocí kterého budou požadovaná data sdílet. Následovalo zadání práce vývojovému oddělení, které si napsalo vlastní program řešící odpověď na požadovanou otázku (tvorba sestavy dle zadaných požadavků).

V závislosti na složitosti a obsáhlosti původního dotazu trvalo zpracování několik dní a někdy i týdny. Pokaždé se navíc objevovaly komunikační potíže, které mohly celý proces znehodnocovat. Dalším problémem navíc bylo odlišné pochopení požadavku. Zatímco např. pracovník ekonomického oddělení s ekonomickým vzděláním chápal položenou otázku v úplně jiném smyslu, pracovník oddělení výpočetní techniky (bez odpovídajícího ekonomického vzdělání) mohl stejnou otázku pochopit zcela opačným způsobem.

Díky dnešní dostupnosti výpočetní techniky, její miniaturizaci a personifikaci si může v dnešní době odpovědný pracovník zajišťovat odpovědi na potřebné dotazy zcela sám, pouze za využití specializovaného softwaru a např. osobního počítače (či v dnešní době tolik populárního tabletu nebo chytrého mobilního telefonu). Díky této dostupnosti může sbírat cenné informace poskytnuté procesem BI opravdu každý.

Právě z těchto důvodů získává obor BI jako takový v současnosti na atraktivitě. Jedná o zcela legitimní a samostatný obor, kterému se věnují celé divize mezinárodních firem dobře známých ve svých oborech (jako např. Oracle, SAP, IBM, apod.) i výzkumníci v akademické sféře, kteří se problematice BI naplno věnují a přinášejí potřebné teoretické základy, které slouží jako stavební kameny k praktickému využití.

I přesto je obor BI relativně nový. Proto asi nikoho nepřekvapí, že nebyla ještě ustálena jedna definitivní definice. Existuje zde tedy jev, který je známý prakticky ze všech oblastí vědeckého bádání, kdy jeden konkrétní pojem vysvětluje vícero definic různých autorů. Přičemž definice procházejí nezbytným časovým testem, kdy některé zastarávají, jiné poslouží jako základ pro definice nově vznikající a jiné naopak pevně odolávají snahám o jejich zpochybnění a používají se do současnosti.

Jednu z prvních novodobých definic v českém jazyce přinesl tým kolem Novotného (2005), která uvádí následující: *„Business Intelligence je sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniku a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data. Aplikace BI pokrývají analytické a plánovací funkce většiny oblastí podnikového řízení, tj. prodeje, nákupu, marketingu, finančního řízení, controllingu, majetku, řízení lidských zdrojů, výroby, IS/ICT*

(pozn. autora: anglické zkratky IS – informační systém a ICT – informační a komunikační technologie) *apod.*“.

V této definici je kladen primární důraz na multidimenzionální pohled na podniková data v celé své šíři, která zajistí účinnou a účelnou podporu většiny oblastí podnikového řízení. V další publikaci na téma BI (Pour et al., 2012) je tato definice rozšířena o vhodnost využití BI na všech úrovních řídicího procesu. Pour zde říká, že procesy BI podporují analytické, plánovací a rozhodovací činnosti organizací a hlavně akcentuje důležitou odchylku *„na všech úrovních a ve všech oblastech podnikového řízení“*.

Třetí definice, tentokrát od zahraničního autora Labergeho (2012) klade důraz hlavně na takticko-strategický význam podnikových informací pro plánování a nabízí možnost využití výstupů BI hlavně pro účely podnikové analýzy, když doslova říká, že: *„Business Intelligence je zastřešující termín, který se vztahuje ke znalostem, procesům, technologiím, aplikacím a postupům, které usnadňují podnikové rozhodování. Technologie Business Intelligence pracuje s použitými (historickými) daty v požadovaném kontextu a pomáhá přijímat podniková rozhodnutí pro budoucnost. Business Intelligence se soustřeďuje hlavně na interní informace o provozních aspektech, které souvisejí s taktickým a strategickým plánováním. Informace jsou obvykle nějakým způsobem uspořádány a přímo nebo nepřímo se odvozují z aktuálních podnikových procesů. Výstup základních dat se používá hlavně v interní analýze, ale lze jej také zkombinovat s externí analýzou, jako je např. kompetitivní analýza SWOT a PEST.“*

Prakticky nejkompaktnější (což pravděpodobně vychází z toho, že jeho kniha je zaměřena více populárně-naučným směrem než 3 předcházející) definici potom přináší Lacko (2011). Ten především akcentuje význam procesu BI jako procesu definujícím podstatné metriky a dále jako procesu, který dokáže z velkého množství obecného vyextrahovat konkrétní poznatky: *„Business Intelligence je množina konceptů a metodik, které zlepšují rozhodovací proces za použití metrik, nebo systémů založených na metrikách. Účelem procesu je konvertovat velké objemy dat na poznatky, které [...] můžeme efektivně použít například v procesu rozhodování a mohou tvořit velmi významnou konkurenční výhodu.“*

Ze všech čtyř výše uvedených definic od různých autorů vyplývá, že ač se liší v méně podstatných dílčích kritériích, jádro zůstává v podstatě stejné. Dalo by se shrnout slovy: BI je sada konceptů, metod a postupů, které se postupně aplikují na velké množství primárních dat pocházejících z různých zdrojů, s cílem najít syntézu, která poskytne potřebné údaje účelně podporující manažerské řízení.

Dále všechny uvedené definice kladou hlavní důraz na konečný efekt celého procesu BI, tedy samotnou syntézu. Z tohoto přirozeně vyplývá, že nejdůležitějším přínosem BI není proces samotný, ale jde především o výsledek.

Dalším nezbytným kritériem celého procesu je celkem logicky čas zpracování konečného výstupu. Odpovědný pracovník zpravidla potřebuje na případnou příležitost či hrozbu reagovat pružně, kvalifikovaně a pokud možno s co nejnižšími náklady. Zde se přímo nabízí otázka, jakou hodnotu má pro dnešní rozhodování analýza, která vyhodnocuje situaci, která již dávno neplatí, protože konkurenční podnik dokázal na příležitost/hrozbu zareagovat rychleji?

Odpověď bude pravděpodobně ve smyslu, že takováto analýza nemá v podstatě žádnou hodnotu, je bezcenná. Z tohoto důvodu je nutno brát v úvahu i dimenzi časovou, kdy ve většině případů požadujeme konečné výstupy v co nejkratším možném časovém horizontu, samozřejmě nejlépe stisknutím jednoho tlačítka - okamžitě.

Legitimní otázkou, kterou se každý podnik uvažující o možnostech zavedení BI v rámci svých procesů bude zabývat, je jaké konkrétní přínosy poplynou aplikujícímu subjektu. Mezi největší výhody datové analýzy přístupem BI zpravidla patří:

1. Efektivní a rychlý proces: při správném výchozím nastavení a následném provedení celého procesu (zejm. správně zpracovaná a vhodně zvolená zdrojová data) poskytuje velmi efektivně (bez zbytečných mezikroků) a poměrně rychle chtěné informace. Poskytuje téměř okamžité zobrazení dat z různých pohledů (pohledem se rozumí např. časové rozmezí poskytované informace, např. počet prodaných kusů skotu za období prvního čtvrtletí roku 2014).



2. Strategický význam: pro výkonné, odpovědné a především kvalitní rozhodování musí mít pověřená osoba k dispozici co nejpřesnější a nejkvalitnější výchozí informace, na základě kterých bude provádět odpovědná a kvalifikovaná rozhodnutí.
3. Aktuálnost a přehlednost: za každé situace poskytuje jasné, přehledné, aktuální, reliabilní a validní informace.
4. Rozšíření stávajících obzorů: proces BI umožňuje spolehlivě a validně odhalovat na první pohled neviděné souvislosti a vyhledávat možná směřování do budoucna.

Z ankety zkoumající nasazení technologií BI v rámci podniků (ICT Revue, 2014), která byla pořádána mezi vrcholovým managementem podniků poskytujících řešení BI v České republice, v jedné z odpovědí vyplynuly jako primární důvody pro investici do BI především nízké pokrytí klíčových podnikových činností prvky analýzy a vykazování. To následně vede ke zvýšeným režijním nákladům spojeným s vytvářením těchto potřebných analýz ad-hoc, nedostupností jejich výstupů v potřebném časovém horizontu a v neposlední řadě také matoucí nejednotností stávajících výstupů.

Jako druhý důvod byla uvedena možnost rozšíření možností současných podnikových systémů o nové funkce jako např. vytěžování nestrukturovaných dat, plánování obchodní výkonnosti, segmentace trhu, apod.

### 3.1 Proces BI

Jelikož celkový proces BI plně závisí na zdrojových datech pořízených v jiných systémech, na základě kterých pracuje, je vcelku jasné, že se nejedná o proces statický. Jde o proces plně dynamický, který musí být neustále prováděn, vyhodnocován a následně upravován na základě získaných zkušeností.

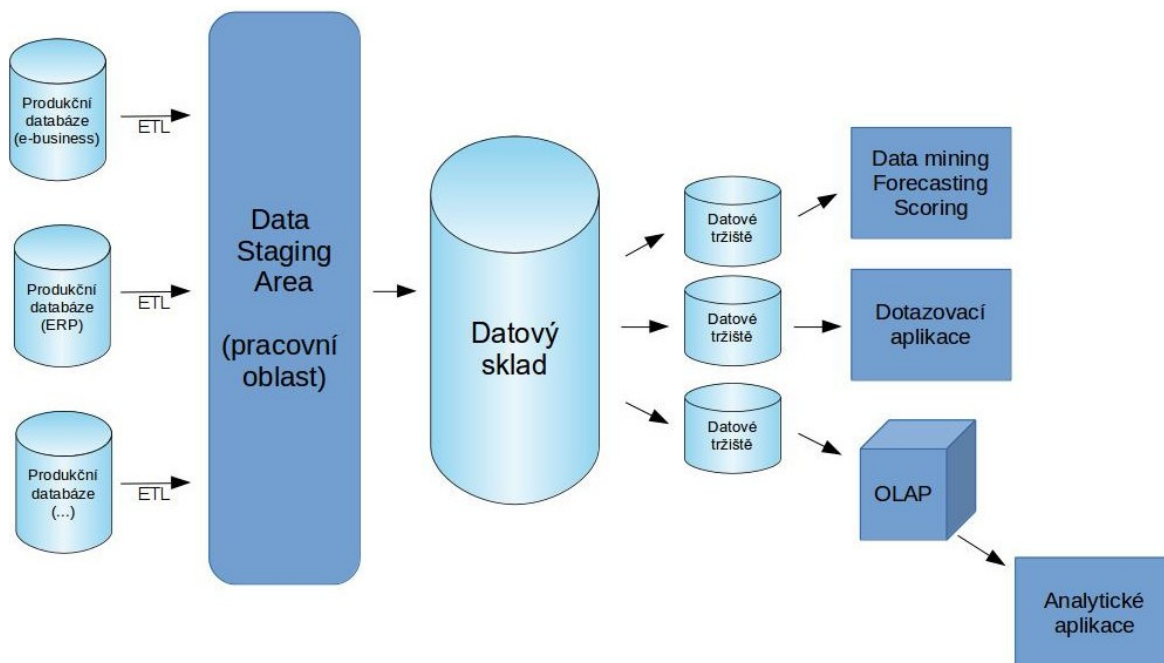
Protože se očekává, že na základě výstupu tohoto procesu budou prováděny kroky, které mají význačný dopad do všech oblastí působení podniku, je více než nutné upozornit na důležitost správnosti provedení celého procesu. Očekává se především bezvadná koordinace zainteresovaných pracovníků, zejména potom řídicích pracovníků (kteří vědí, co přesně potřebují) s oddělením výpočetní techniky, kteří poskytnou nezbytnou technickou – zejména zajištění základního hardwarového a softwarového vybavení – a metodickou podporu (samotná práce se specializovanými programy).

A to nikoli pouze na povrchní úrovni. Pro maximalizaci výnosů celého procesu se očekává téměř dokonalé splynutí informačních technologií s řídicím procesem, kdy hodnota informací je tak velká, že se v posledních několika letech informatika stává sama o sobě hodnototvornou aktivitou, nosným pilířem celé organizace a pevnou součástí celé organizace (Dohnal et al., 2013).

Existují obecně dva přístupy k procesu BI (Novotný et al., 2005), jeden úzce zaměřený, který chápe BI pouze jako jeden z nástrojů prezentace dat nad datovými sklady/tržišti, a druhý širěji zaměřený, který proces popisuje v celkovém kontextu. Tato práce přitom zkoumá BI právě v tomto širším kontextu, kde je proces rozdělen do těchto fází (Novotný et al., 2005):

1. tvorba zdrojových dat v produkčních systémech,
2. transformace produkčních dat do datových skladů,
3. databázové komponenty pro analytické nástroje (datové sklady/tržiště),
4. analytické komponenty nad datovými sklady a prezentace uživatelských výstupů.

Novotný však zároveň upozorňuje, že toto popisované uspořádání celého procesu se může výrazně měnit, zejména dle konkrétní situace a potřeb daného zákazníka či podniku. I přesto lze považovat naznačený postup procesu za nepsaný standard.



**Obrázek č. 1: Schéma procesu BI (podle Poura et al., 2012)**

### 3.1.1 Tvorba zdrojových dat v produkčních systémech

Prvotním článkem celého procesu zpracování dat a její následné transformace na užitečné a hodnotné výstupy jsou zdrojová data. Tato zdrojová data jsou pod kontrolou a ve vlastnictví podniku a jsou typicky pořizována z různých zdrojů a následně zaznamenávána (kódována) do produkčních systémů, resp. databází, které podnik používá.

Produkční (zdrojové) systémy jsou takové systémy podniku, které slouží samotnému procesu BI, získávají se z nich podkladová data pro celý proces, ale zároveň nepatří do skupiny aplikací BI. Poznávacím znamením těchto systémů je jejich celková architektura, která umožňuje vytváření, čtení, úpravu a mazání dat v reálném čase. Souhrn těchto 4 základních operací se někdy označuje anglickým akronymem CRUD tvořeným slovy Create, Read, Update, Delete. Na rozdíl od systému BI tyto systémy primárně nepodporují analytické operace (Novotný et al., 2005), které jsou alfou a omegou BI.

Mezi produkční systémy dlouhou dobu patřily pouze tradiční provozní aplikace, bez kterých se dnes již neobejde prakticky žádná firma. Protože však v poslední době narůstá význam internetových technologií, nabízí se rozdělení zdrojových systémů podniku do těchto dvou kategorií:

1. Tradiční provozní podnikové aplikace. V dnešní době již zcela standardně používané informační systémy typu ERP, CRM, účetní programy, skladové hospodářství, mzdové a personální programy, tabulkové procesory typu Microsoft Excel, apod.
2. Nové zdroje hodnotných informací související s rozvojem internetových technologií (především informace pocházející z interakcí prostřednictvím počítačové sítě Internet – e-mail, internetové prezentace (webové stránky), sociální sítě (Facebook, Twitter), internetové obchody a data pocházející z nákupů v těchto obchodech apod.

Jak vyplývá z výše uvedeného, možných zdrojů dat v podniku je opravdu hodně a v budoucnosti se bude pravděpodobně (pokud se nestane něco nepředvídatelného) jejich objem neustále zvětšovat, a to o přírůstky exponenciální rychlostí.

Pro představu, americká společnost International Business Machines (známější spíše pod zkratkou IBM) publikovala v roce 2006 studii, kterou příznačně nazvala „The toxic terabyte“ (česky doslova „Toxický terabajt“) zaměřenou na informační explozi, která provází vývoj naší současné informační společnosti. Zde uvedla příklad dnes jednoho z nejběžnějších komunikačních prostředků v obchodní i soukromé sféře, e-mailovou komunikaci. Ta zdvojnásobuje svůj objem každý rok o 20 – 30 % (Coles et al., 2006). Jelikož dnes umí systémy řízení vztahů se zákazníky (CRM) analyzovat obsah e-mailové komunikace s obchodními partnery, je nasnadě, že organizace musejí nějakým způsobem uchovávat i tato data. Zvětšování objemu je tedy otázkou, která je zcela legitimně na pořadu dne.

S ohledem na výše uvedené proto velmi záleží na pečlivém zvážení a výběru těch souborů dat, které budou v procesu BI dále využity. Nejdůležitějším činitelem určujícím a spoluvytvářejícím tuto fázi je zejména datová kvalita, případně spolehlivost či integrita.

Tyto termíny vyznačují pomyslnou hranici mezi bezcennými daty a hodnotnými informacemi, na základě kterých je možno se odpovědně rozhodovat.

V případě, že se podnikatelský subjekt snaží o datovou kvalitu, měl by se především zaměřit na následující charakteristiky, kterými lze obecně definovat kvalitní data jako data, která je splňují (Novotný et al., 2005):

- úplnost – data musí být úplná a použitelná,
- soulad – data musejí být v souladu s požadovaným formátem (dodržovat stanovený formát),
- konzistence – nesmějí existovat konfliktní hodnoty,
- přesnost – data musejí být přesná a aktuální,
- unikátnost – data nesmějí být duplicitní,
- integrita – data, která k sobě mají definovány určité důležité vztahy, je musí vzájemně dodržovat.

Důležitost dodržování datové kvality zdůrazňuje Laberge (2012) na příkladu jedné americké pojišťovny, kde její pracovníci vkládali do podnikového interního informačního systému informace o pojistných smlouvách vztahujícím se k pojištěným osobním automobilům svých klientů. Protože dodavatelé softwarového vybavení nezajistili standardy datové kvality již při vývoji aplikace, dovolili do jednoho formuláře v aplikaci vkládat položku „značka vozidla“ vlastními slovy.

Tato implementační chyba pak následně dovolila uvádět automobil Ford jako např. FORD, ford, FROD, Ford, Fort. Pokud by formuláře následně analyzoval člověk, dokázal by asi všech pět slov správně rozpoznat jako značku automobilu Ford. Pokud však tyto záznamy porovnává počítač (který obvykle porovnává rovnost absolutní shodností znaků v poměru 1:1), zjišťuje 5 různých značek automobilů.

V případě, když by chtěli odpovědní pracovníci např. nechat zpracovat analýzu na téma průměrná výše placeného pojistného u majitelů osobního vozu značky Ford, není třeba uvádět následky rozhodnutí učiněných na základě této zcela špatné analýzy. Přestože by na první pohled poskytovala zcela správnou odpověď na položenou otázku, fakticky by vůbec neodpovídala realitě.

Tyto odchylky bohužel neumí současné systémy řešit a proto je jejich odhalení velice obtížné až zcela nemožné. Protože chybnými rozhodnutími učiněnými na základě těchto *falešných* výstupů analýzy mohou být způsobeny obrovské materiální škody, je nutno se velmi dobře zaměřit na zajištění datové kvality již na úrovni zdrojových systémů.

Tato práce by měla být systematicky plánována a kontrolována příslušnými odpovědnými pracovníky. Jednou chybně získaná data se totiž bez přímého vztahu k době vzniku poměrně nesnadno opravují (přičemž toto zvlášť vynikne u činností, které nemají rutinní charakter, na jejichž vykonávání působí vícero vnějších unikátních vlivů).

Mezi nejčastější oblasti chybných dat mj. patří např. tyto (Pour et al., 2012):

- chyby při manuálních vstupech dat (způsobené samotnými pracovníky zodpovědnými za kódování jako např. překlepy ve slovech, špatně opsané číslo, vynechané znaky, dále např. také špatné pochopení účelu zadávané položky, apod.),
- různé způsoby uložení dat v různých zdrojových systémech (např. jeden systém může ukládat pohlaví jako M/Ž, druhý jako muž/žena, případně třetí analogicky v cizím jazyce M/F, přičemž všechny tři reprezentace znamenají pro svého původce totéž, zatímco pro 4. integrující program znamenají 3 odlišené hodnoty),
- neoprávněné zkoušení a testování aplikací uživateli na ostrých datech (do dat, která by měla splňovat vysokou produkční kvalitu jsou vkládána testovací, často nesmyslná a bezkontextová data),
- chyby vznikající v důsledku přímého přístupu do databáze externími subjekty, kteří neznají všechny zákonitosti a souvislosti podnikové aplikace a mohou tak zanášet do databáze neodpovídající databáze,

- chyby vzniklé špatným prvotním návrhem databáze (ignorace vazeb mezi jednotlivými záznamy, nedostatečná kontrola referenční integrity jednotlivých tabulek, nevytvoření klíčů nad významnými sloupci apod.) a další jiné.

Výše popsána byla kvalita dat především po jejich formální stránce, kdy byl hlavní důraz kladen na to, zda se dvě na pohled stejná slova shodují identicky písmeno po písmenu, zda je celé číslo vždy zadáno jako celé číslo číslicí a ne např. jako desetinné (byť s nulovým desetinným místem – některé programy totiž rozlišuje mezi typy čísel, kde 1 není zcela jistě identicky 1,00) případně zapsané slovy jako „jedna“. Druhou, neméně důležitou kvalitativní charakteristikou informace je její obsahová kvalita.

Když význačný italský ekonom a filosof Vilfredo Pareto zkoumal rovnoměrnost rozdělení národního bohatství (důchodu) mezi obyvatele územní jednotky, přišel na to, že bohatství je rozděleno nerovnoměrně, zhruba v poměru 80:20. Tento princip následně přešel i do jiných oblastí (i relativně vzdálených ekonomii a rozdělování bohatství), až se začal používat jako obecný poměr rozdělení *kvalitního* od *nekvalitního*. V ekonomii se tak např. používá parafráze, že pouze 20% segment sortimentu se podílí na tržbách společnosti z celých 80 %. Podobně se k poměru kvalitního od nekvalitního vyslovil známý americký spisovatel Theodore Sturgeon, který hranici posunul ještě dál, když řekl, že 90 % z celku má nevalnou hodnotu (doslovně: „ninety percent of everything is crap“). Tato myšlenka se poté vžila ve více oblastech pod pojmenováním Sturgeonův zákon.

Uvedená pravidla vztažená na oblast informací, mohou být interpretována tak, že pouze 10-20 % z podniku dostupných zdrojových informací bude mít požadované kvalitativní charakteristiky a mohou tak být bezpečně využity v následném analytickém procesu.

Nabízí se celkem logická námitka, že se jedná o pouhé vyslovení nahodilé filosofické myšlenky, která nemá v exaktním světě informačních technologií své místo. Na legitimitě těchto názorů ovšem přidává zcela exaktní studie renomované celosvětově uznávané americké analytické společnosti Gartner, která se již od svého založení v roce 1979 zabývá výzkumem a hlavně předvídáním trendů v oblasti informačních technologií.

Tato výzkumná organizace také mimo jiné každoročně pořádá ve Spojených státech amerických odbornou konferenci nazvanou *Gartner Business Intelligence & Analytics Summit* věnovanou oblasti Business Intelligence a analýzy dat. Na konferenci pořádané v roce 2012 její pracovníci uvedli výsledky vlastního výzkumu, který zkoumal datovou kvalitu (co do obsahu dat, nikoli formy) podnikových zdrojů. Publikovaný výsledek říká, že nestrukturovaná data (tedy spíše nahodilé shluky bez kontextu) zabírají ve zdrojových systémech celých 60-80 % celkové kapacity (Quagini, 2012), jinak řečeno, kvalitní data tvoří pouze 20-40 %.

Z hlediska celkového objemu informací je tedy nutno zaměřit se nejvíce na nejhodnotnější podíl celku, tedy oněch popisovaných zhruba 20 %. Pro usnadnění kvalifikace obsahové kvality dat je nutno identifikovat především takové informace, které mají největší význam při budování konkurenční výhody podniku. Protože informace patří do podnikových zdrojů, vztahují se na ně obecná kritéria hodnotných podnikových zdrojů. Ta říkají, že aby měly zdroje význam při budování konkurenční výhody podniku, je potřeba, aby splňovaly zejména následující podmínky (Tichá et al., 2002):

1. musejí být hodnotné, (musí umožnit využít svůj potenciál),
2. musejí být vzácné (tedy ne lehce dostupné nebo napodobitelné pro konkurenční podniky),
3. nesmí k nim existovat snadno dostupné substituty.

Po této analýze by mělo být rozhodnuto, která data budou pokračovat dále v rámci procesu BI. Protože budou následně použita v procesu zpracování, je důležité, aby k nim byl zajištěn odpovídající přístup. Tento úkol nepatří mezi úplně triviální, protože prakticky každý systém má většinou svůj vlastní systém uložení dat (co do struktury a integrity).

Obecně existují dvě metody přístupu k datové sadě určitého systému. V horším případě je nutno přistupovat přímo do databáze daného programu, což s sebou nese především riziko nemožnosti k datům v této databázi vůbec přistoupit (z důvodu ochrany



takového jednání tvůrci aplikace), případně možnost poškození produkční databáze nesprávnou manipulací, což může vést až k nenávratné ztrátě důležitých dat.

Při přímém přístupu do databáze produkčního systému je možno v rámci lepší orientace v konkrétním datovém modelu využít specializované diagramy. Tyto diagramy poskytují celkový pohled na uspořádání datových prvků v rámci jednoho systému (databáze). Jejich cílem je přehledně vizualizovat celkový pohled na databázi, kterou vnímají jako soubor entit a jejich vzájemných vztahů. Jeden takový diagram se potom nazývá model E-R (podle anglické zkratky Entity-relationship model), tvorba těchto diagramů potom modelováním.

Při modelaci diagramu se ve své podstatě využívají 3 nosné prvky (Chen, 1976):

1. datová entita (např. tabulka),
2. atributy entity (např. sloupce tabulky),
3. vzájemné vztahy mezi entitami (např. cizí klíče).

Výsledkem modelace může být např. schematické znázornění jednotlivých tabulek, jejich sloupců včetně datových typů a v neposlední řadě vztahů mezi jednotlivými tabulkami. Výhodou těchto diagramů je především rychlý vizuální přehled v často složité síti vztahů a jejich snadná pochopitelnost, kdy uživatel tohoto výstupu nemusí mít technické znalosti databáze, jejíž strukturu diagram znázorňuje.

V lepším případě, pokud dodavatel software poskytuje tzv. API (Application Programming Interface), však nemusí být přímý kontakt s databází vůbec navazován. API je aplikační rozhraní, pomocí kterého se může dotyčná osoba pohodlně dotazovat samotné aplikace na důležitá data bez nutnosti přímého přístupu do databáze. V případě tradičních podnikových aplikací toto nebývá úplně standardem (v dřívějších dobách, kdy neexistovala poptávka po interoperabilitě jednotlivých systémů, nebyl důvod pro uživatele k těmto datům přistupovat), zatímco u dnešních moderních internetových aplikací bývá slušným standardem, kdy se se zpřístupněním API počítá již v začátku vývoje služby.

V situaci, kdy má podnik dokončeno základní zhodnocení zdrojových systémů, zajištěnou potřebnou kvalitu dat, vybrán základní datový vzorek (kvalitní data po stránce obsahové) a zajištěn přístup k těmto datům, může začít tato data transformovat do analytické podoby, která bude vhodná pro systémy BI, které dokáží nad těmito daty tvořit požadované výstupy. Následuje tedy fáze převodu (transformace) zdrojových dat do datových skladů.

### 3.1.2 Převod informací z produkčních systémů do datového skladu

Tato fáze spočívá především v precizním nastavení pravidel převodu zdrojových dat na data analytická, využitelná v procesu BI. Integračním prvkem, který je schopný tato analytická data uchovávat je datový sklad. Probíhá zde převod produkčních dat do datového skladu. Jedná se o rozsáhlý soubor úloh a operací, které se velkou měrou podílejí na pracnosti a časové náročnosti projektů BI, obvykle cca v rozsahu 60 – 80 % celkové pracnosti (Pour et al., 2012).

Datový sklad, často také označován anglickou zkratkou DWH (Data Warehouse), se definuje jako *integrovaný, konsolidovaný, subjektivě orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu* (Pour et al., 2012). Pour dále vysvětluje jednotlivé charakteristiky datového skladu takto:

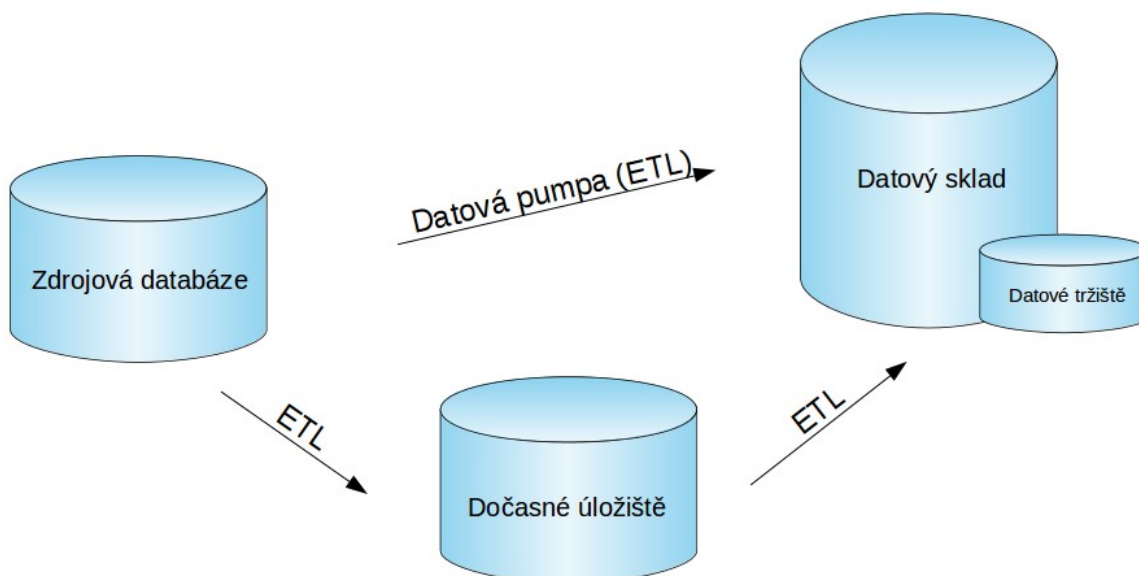
- subjektivě orientovaný – data jsou primárně uložena na základě svého typu, ne podle místa původu (produkční aplikace, kde data vznikla),
- konsolidovaný – i přesto, že zdrojová data pochází z více autonomních zdrojů, v datovém skladu musejí být normalizována do jedné výsledné podoby,
- integrovaný – data jsou integrována v rámci celého podniku nehledě na organizační jednotky,
- stálý – na rozdíl od zdrojových databází, které podporují základní procesy manipulace CRUD, datové sklady musejí zajišťovat konstantní pohled na data. Z tohoto důvodu bývají data přístupná pouze pro čtení a jakékoli manipulace ovlivňující jejich podobu jsou striktně zakázány,

- časově rozlišený – zatímco zdrojové programy většinou obsahují informace vztažené platné v určitý konkrétní čas, datové sklady podporují i jejich historizaci (tedy zobrazení nejenom aktuální hodnoty, ale také vývoje této hodnoty za určitou časovou jednotku).

Procesu transformace dat do datového skladu předchází vytvoření tzv. logické datové mapy, jejímž účelem je detailně zmapovat a hlavně definovat vztahy mezi daty ve zdrojových a cílových databázích. Tato datová mapa je vyjádřena v různých variantách detailnosti a různou formou (nejčastěji tabulkou), přičemž většinou obsahuje tyto konkrétní charakteristiky (Pour et al., 2012):

- název cílové databáze, dále názvy jednotlivých tabulek a jejich typy v této databázi. To platí pro analytická datová úložiště, tj. datové sklady, tržiště, dočasná úložiště apod,
- názvy sloupců v cílových tabulkách,
- názvy zdrojových databází (s produkčními daty), dále názvy a typy tabulek, jejich sloupce,
- popis transformace dat ze zdrojové databáze do cílové.

Tok dat ze zdrojové databáze může být přímý bez prostředníka, ale to je spíše výjimka. V praxi se daleko více používá druhá možnost, tedy využití prostředníka. Jako prostředník je zde využívána pracovní databáze, alternativně dočasné úložiště dat, případně Data Staging Area (Pour et al., 2012).



**Obrázek č. 2: Tok dat ze zdrojové databáze do analytického datového skladu (podle Poura et al., 2012)**

Samotnou transformaci poté zajišťuje specializovaný software označovaný jako datová pumpa, případně anglickou zkratkou ETL. Zkratka ETL zjednodušeně popisuje proces, kterým musejí data projít, aby mohla být uložena v datovém skladu. Každé jednotlivé písmeno zkratky znamená (Pour et al., 2012):

- E (Extract) – výběr dat ze zdrojových databází,
- T (Transform) – transformace zdrojových dat do jiných datových struktur (tabulek a struktur záznamů),
- L (Load) – samotné uložení dat do analytických databází.

Datová pumpa ve své podstatě bývá specializovaný program, který provádí výše popsané operace sekvenčně jednu po druhé. Jako příklad datové pumpy může být uveden program, který na základě uživatelského pokynu získá z produkčního fakturačního systému záznamy všech faktur (vydaných i přijatých) včetně identifikací a adres obchodních partnerů na nich uvedených a tato data jednotně zapíše do datového skladu.

Datové pumpy většinou pracují v dávkovém režimu, kdy zpracovávají velké množství souvisejících datových záznamů, většinou v určitých časových intervalech.

Nejběžněji se v dnešních systémech využívají intervaly denní, týdenní a měsíční (Novotný et al., 2005).

Po softwarové stránce představují datové pumpy veskrze dvě skupiny programů. První tvoří profesionální řešení dodávaná výrobcem softwaru, kteří se specializují na oblast BI, kde můžeme zmínit např. produkty společností Oracle (Warehouse Builder), či Microsoft (SQL Server Integration Services). Tyto programy jsou vybaveny již od výroby sadou standardních zprostředkovatelů propojení tzv. konektorů, které umí automaticky rozpoznat podporovaný produkční systém, připojit se k jeho zdrojové databázi a data následně transformovat do datového skladu. Výhody pro koncového uživatele jsou zřejmé: rychlost, jednoduchost, spolehlivost, stabilita a robustnost takového řešení.

Protože dodavatelé těchto konektorů mají většinou velké zákaznické báze, je velice pravděpodobné, že konektory jsou dobře otestované na různorodých konfiguracích a obvyklé chyby jsou díky tomu opraveny. Konektory navíc zcela jistě prošly mnohými zátěžovými testy a jejich stabilita v praxi, která je v případě automatizovaných dávkových opakovaných úloh ideálně bez lidského zásahu velice žádaná.

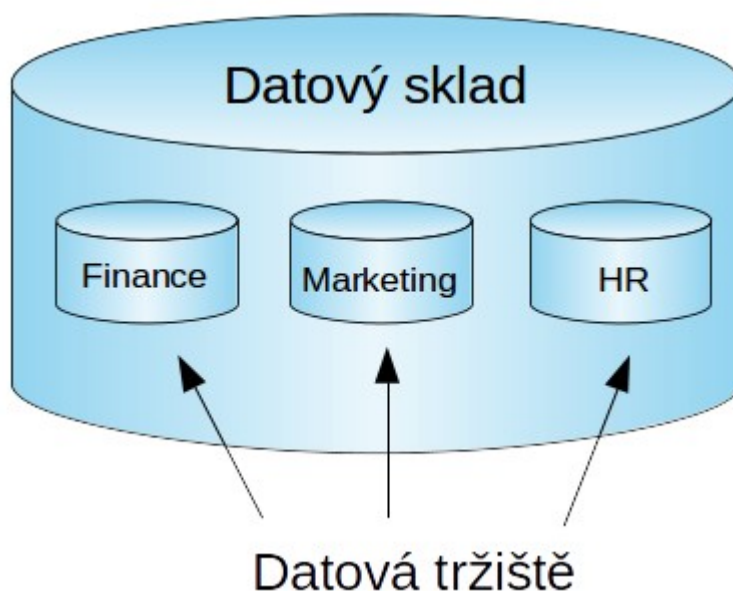
Největší nevýhoda tohoto řešení spočívá v nutnosti vlastnit konektorem podporovaný produkční systém. Pokud zde podpora chybí, je nutné využít druhý přístup k pořízení nástroje ETL a tím bývá nechat si datovou pumpu vytvořit na zakázku.

Výhodou tohoto přístupu je především těsná a velice přesná integrace zdrojového systému s datovým skladem, nevýhoda často spočívá ve vyšší ceně zakázkové výroby a především neuniverzálnosti tohoto řešení.

Během celé fáze musí být na paměti pečlivost provedení celého procesu. Zatímco vstupující datová masa a její kvalita je pevně daná výstupem první fáze, ve fázi transformace mohou být data dále pozměněny (i směrem k odlišné interpretaci). Z tohoto důvodu zde musí být především zajištěna pokud možno dokonalá transformace s co nejvyšší kvalitou výsledných dat v datovém skladu (i za využití tzv. čištění zdrojových dat).

### 3.1.3 Tvorba datových tržišť dle požadovaných oblastí

Nyní, když má podnik zdrojová data bezpečně převedena a uložena v datovém skladu, začíná práce s analytickými daty, které představují úplnou abstrakci nad daty zdrojovými. Protože se předpokládá, že jednotlivé organizační skupiny (např. pobočky, oddělení, divize) budou mít v určitých oblastech specifické požadavky, jsou nad datovými sklady vytvářena tzv. datová tržiště, DMA (Data Mart). Datová tržiště jsou ve své podstatě zmenšené datové sklady určené pro omezený (specifický) okruh uživatelů, fakticky tvoří podmnožinu datového skladu.



**Obrázek č. 3: Vztah mezi datovým skladem a datovými tržišti**

Jednotlivá datová tržiště jsou v současnosti nejčastěji budována na principu tzv. sběrníkové architektury, kdy je hlavní důraz kladen na integrované budování datových tržišť nad jednou společnou dimenzí. Oborová tržiště potom obsahují specifické informace a sdílená dimenze slouží k uchování společných informací. Postup budování tržišť na sběrníkové architektuře popisuje Novotný (2005) takto:

1. první datové tržiště je vybudováno na základě požadavků organizační skupiny, pro kterou je tvořeno. Např. tržiště pro obchodní oddělení bude obsahovat údaje o vydaných fakturách, o učiněných objednávkách, o expedovaných artiklech apod.

Při analýze jsou vybrány ty oblasti, které mohou být sdílené napříč ostatními tržišti. Zde se nabízí kontaktní informace o obchodních partnerech, dále skladové karty, které jsou potřeba napříč agendami, či případně obecné číselníky typu sazba DPH, měna apod,

2. další datová tržiště jsou budována na základě zkušeností s prvním tržištěm, kdy existuje snaha maximálně využít sdílenou dimenzi a omezit tak potenciální nechtěnou duplicitu dat na minimum. V případě, kdy existuje nemožnost nalezení rozumného kompromisu mezi sdílenými a privátními daty, je taková duplicita na místě.

Nad těmito nezávislými datovými tržišti může být dále vytvořena souhrnná datová vrstva (buď virtuální, nebo fyzická), která tak usnadňuje celopodnikový komplexní reporting.

### **3.1.4 Prezentace analytických výstupů**

Po transformaci zdrojových dat pomocí nástrojů ETL do datového skladu a následném vytvoření oborových datových tržišť nyní přichází fáze, kvůli které vlastně celý proces BI probíhá. Fáze, která by měla podniku přinést právě onu přidanou hodnotu celého procesu.

V této chvíli budou využita uložená analytická data pomocí specializovaných nástrojů, které umožní uživatelskou analýzu a následně prezentaci výsledků, která může mít např. následující formy (Laberge, 2012):

- sestavy,
- dotazy,
- OLAP
- ovládací panely,
- přehledy výsledků.

### 3.1.4.1 Sestavy

Sestava (někdy také označována jako tisková sestava) je zřejmě nejklasičtějším a nejběžnějším obecně známým typem datového výstupu. *Sestavy jsou statické, obvykle předem plánované a spouštěné rutiny, které vytvářejí konkrétní přehledy. Sestavy pak poskytují požadované informace* (Laberge, 2012). Sestavy mají nejtýpičtější formu klasické tabulky s hlavičkou a případně zápatím, které většinou obsahuje agregační funkce.

**Tabulka č. 1: Ukázka jednoduché sestavy**

Datum	Položka	Počet prodaných ks	Částka [Kč]	Celkem [Kč]
1. 5. 2014	Svetr bílý	5	100	500
12. 5. 2014	Svetr modrý	10	120	1 200
14. 5. 2014	Svetr zelený	10	100	1 000
<b>Celkem</b>		<b>25</b>	<b>220</b>	<b>2 700</b>

Zdroj: vlastní tvorba

### 3.1.4.2 Dotazy

Dotazy vyžadují pokročilou znalost tzv. dotazovacího jazyka. Dotazovací jazyk je speciální způsob komunikace s určitou databází, která obsahuje jednotlivé tabulky. Pomocí dotazovacího jazyka probíhá komunikace se správcem databáze (většinou specializovaný databázový software) ve formě otázek, na které chce uživatel znát odpovědi. Výhodou těchto jazyků je, že na rozdíl od klasických programovacích jazyků určených k tvorbě počítačových programů, jsou méně složité, více srozumitelné a podobné standardní lidské řeči (forma zadávání dotazu je podobná, jako kdyby chtěl uživatel počítači přikázat určitý úkon v anglickém jazyce).

Naopak nevýhoda je zřejmá: zatímco výše uvedenou tabulku dovede vytvořit každý průměrně technicky zblhlý uživatel, pokládat dotazy ve specializovaném jazyce již vyžaduje určité úsilí a uživatelskou znalost.

Nejtýpičtějším příkladem dotazovacího jazyka je jazyk SQL (Structured Query Language), který je od roku 1987 standardizován organizací ISO. Mimo tento průmyslový standard existují také další dotazovací jazyky, které většinou fungují pouze v určitém



programu. Je zde typické, že program i dotazovací jazyk vytvořila stejná firma. Příkladem takového jazyka je dotazovací jazyk DAX (Data Analysis Expressions), vyvinutý firmou Microsoft.

Ukázkou použití dotazovacího jazyka SQL může být dotaz na počet prodaných kusů v měsíci listopadu roku 2014:

```
SELECT COUNT(pocet) FROM prodej WHERE MONTH(datum) = 11 AND YEAR(datum) = 2014
```

Tento dotaz se dá volně parafrázovat do českého jazyka přibližně tímto příkazem: *Zobraz součet prodaných kusů z tabulky prodejů, které se uskutečnily v 11. měsíci roku 2014.* Jak je vidět, jazyk je velmi podobný lidské mluvené řeči a díky této vlastnosti i snadněji naučitelný.

### 3.1.4.3 OLAP

Technologie OLAP (Online Analytical Processing) je o hodně chytřejším příbuzným klasické sestavy. Na rozdíl od statické sestavy, jejíž struktura je od okamžiku vytvoření prakticky neměnná, je tento typ výstupu plně dynamický a v reálném čase (online) přizpůsobitelný uživatelskému zobrazení. Každý výstup je jedinečnou kombinací předem zvolených parametrů, které mohou být v průběhu libovolně upravovány.

Tvorba výstupu je intuitivní jednoduchý proces nevyžadující pokročilé znalosti počítače a může tedy být plně v režii odpovědného pracovníka, který potřebuje výsledné informace. Díky tomu odpadá nepříjemná možnost vzniku komunikačního šumu, který by při zapojení více článků logicky vznikl.

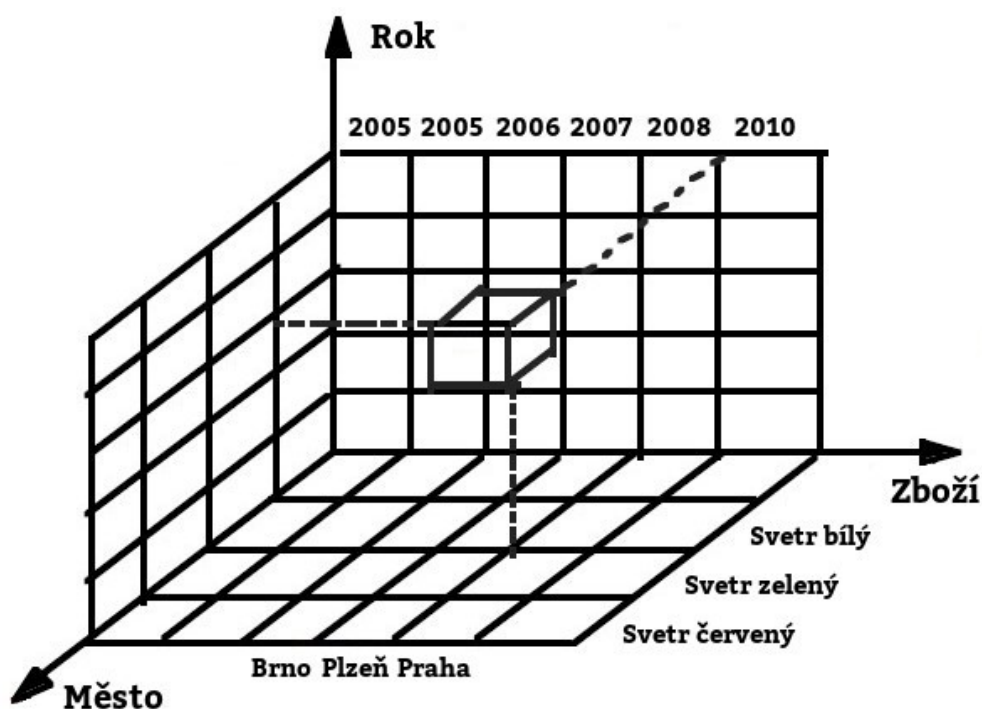
Proces sestavení požadovaných výstupů technologií OLAP může být znázorněn na příkladu rozbalovacího stromu, kde uživatel postupně rozbaluje nebo sbaluje jeho jednotlivé dimenze, podle toho, jak podrobná data zrovna potřebuje.

Na vrcholu stromu, kde se nachází počet prodaných kusů za rok 2014, postupně přechází na další úroveň – počet prodaných ks v měsíci lednu, dále počet prodaných kusů ve Středočeském kraji, následně počet prodaných ks v prodejně v Benešově a nakonec

počet prodaných ks pracovníkem Janem Novákem. Jak je vidět, lze takto velmi jednoduchým a flexibilním způsobem získávat potřebné informace bez nutnosti jakékoli komunikace s další osobou, případně nutnosti pořizovat další data.

V případě použití klasické sestavy by celý proces byl o poznání zdlouhavější a složitější. Celému procesu by předcházela fáze prvotního nastavení a následného vytvoření. To by se muselo pro každou unikátní konfiguraci opakovat. V tomto případě by šlo o 5 různých sestav, které technologií OLAP nahradila jedinou. V případě požadavku na přidání dalšího sloupce do předpočítané sestavy by to znamenalo další přenastavení všech 5 sestav. V modelu OLAP by se pravděpodobně jednalo o jednu malou úpravu.

Aby mohl podnikový uživatel takto jednoduše vytvářet uživatelské výstupy technologií OLAP je potřeba, aby byla správně připravena podkladová data. Nejčastější datovou strukturou uspořádání dat pro model OLAP je tzv. datová kostka (Laberger, 2012).



Obrázek č. 4: Datová kostka OLAP (podle Shoshani, 1997)

Datová kostka drží v každé ze svých dimenzí hodnoty pro danou charakteristickou proměnnou, a umožňuje tím snadno provádět jednotlivé řezy dimenzemi. Z obrázku vyplývá, že díky dimensionalitě kostky lze získat odpověď na otázku, kolik se ve kterém

roce, v tom kterém měště prodalo zelených a červených svetrů. Nebo taky kolik se prodalo všech svetrů za dva roky bez ohledu na geografickou oblast (za všechny oblasti).

Díky okamžitému online zpracování a chytrému uložení dat v kostce je možno zodpovídat tyto a podobné dotazy prakticky okamžitě. Není potřeba vytvářet a předpočítávat žádné další nebo nové sestavy. Záleží jen na uživateli, jak podrobnou kostku si sestaví. Někomu může stačit časový horizont čtvrtletí, někdo potřebuje měsíce, jiný naopak dny či hodiny. Taková vlastnost se potom nazývá granularita dat – tedy jak moc podrobně lze danou charakteristiku zobrazit.

Žádanou vlastností při práci s datovým modelem typu OLAP je rychlost odezvy a zpracování. Při postupném prolézání stromu je potřeba znát odpovědi na podrobnější otázky v co nejkratším čase. V běžném provozu jsou schopny specializované programy umožňující technologii OLAP odpovídat na dotazy v řádech vteřin, maximálně minut. V případě komparace s klasickými sestavami je nutno podotknout, že v tomto případě nemohou konkurovat.

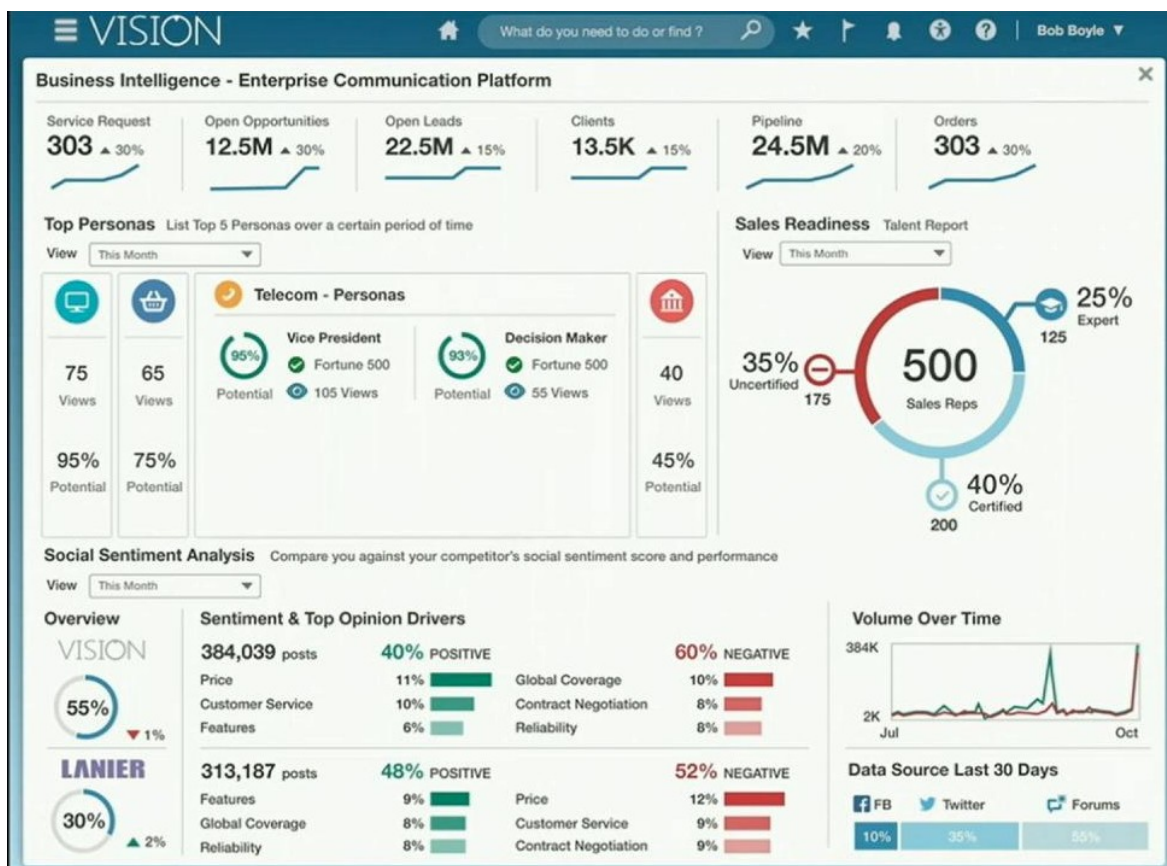
#### **3.1.4.4 Ovládací panely a přehledy výsledků**

Zatímco všechny tři výše uvedené způsoby prezentace dat jsou v podstatě standardní a dlouholetou výbavou systému BI na všech úrovních složitosti, představují ovládací panely (dashboard) a přehledy výsledků (scorecard) nejnovější trendy v prezentaci datových výstupů. Zatímco sestavy, dotazy a OLAP poskytovaly výstupy v lety prověřené a všem bezpečně známé podobě – klasické tabulce – hodí se tyto panely a přehledy především k prezentaci tzv. KPI (Key Performance Indicator), klíčových ukazatelů výkonnosti. Klíčové ukazatele výkonnosti (někdy také jako klíčové metriky výkonnosti) jsou výsledky měření pro podnik nejdůležitějších výkonnostních veličin, jejichž hodnoty jsou kritické pro současný i budoucí úspěch organizace (Parmenter, 2007).

Sestavení klíčových ukazatelů výkonnosti je v režii každého podniku, protože podnikatel sám by měl nejlépe vědět, které veličiny nejvíce ovlivňují výkonnost jeho podniku. I zde platí ono známé pravidlo, že méně je někdy více a proto se doporučuje držet celkový počet ukazatelů v rámci jednoho podniku v počtu mezi 10-20 ukazateli (Parmenter, 2007). V opačném případě může nastat záporný efekt z jinak očekávaného

přínosu zavedení KPI, který zastihuje jeho největší výhodu – rychlost, jasnost a přehlednost. Toto navíc popírá základní pravidlo KPI, které říká, že ke sledování by se měly vybrat pouze ty neklíčovější oblasti.

Prezentace KPI prostřednictvím panelu ve specializovaném softwaru Oracle Business Intelligence Suite s cca 15 klíčovými ukazateli výkonu může vypadat například takto:



Obrázek č. 5: Snímek aplikace Oracle Business Intelligence Suite (Soat, 2014)

Zde jasně vyplývá největší výhoda tohoto způsobu prezentace. Rychlá účinná a přehledná prezentace velkého množství klíčových informací. Zatímco klasická sestava (ať už ve statické formě, či ve formě dynamické pomocí technologie OLAP), poskytuje veskrze úzce zaměřený pohled na jedinou zkoumanou problematiku v jediný okamžik, poskytuje ovládací panel velmi široký pohled (co do počtu zobrazovaných dat) speciálně zacílený na nejdůležitější oblasti.

Žádanou vlastností u ovládacích panelů je jejich rychlé a pohodlné sdílení napříč celými týmy zainteresovaných osob. Ovládací panely mohou být umístovány do vnitropodnikové sítě, kde jsou ihned k dispozici autorizovaným uživatelům. Hodí se také výborně k motivaci jednotlivých pracovníků, kdy jsou tematicky zaměřené panely (např. a aktuální prodej zboží za aktuální časovou jednotku) promítány na velkou obrazovku. Každý člen týmu poté prakticky okamžitě sleduje, jak se jeho týmu vede, a týmy tak mezi sebou mohou přirozeně soupeřit a motivovat se k lepším výsledkům.

### 3.2 Nástroje BI

Pokud podnikatel chce aplikovat procesy BI zpravidla se neobejde bez pomoci výpočetní techniky (BI je především o zpracování velkého množství dat), především pak softwarového vybavení, nástrojů, které umí zpracovávat data a co je hlavní, pomáhat z nich vytvářet analýzy a doporučení.

Softwarové (programové) vybavení potřebné k aplikaci BI dodávají především specializované firmy, které v rámci své podnikatelské činnosti tyto nástroje vyvíjejí vlastními vývojovými týmy, či pouze dodávají hotové produkty vytvořené jinými softwarovými společnostmi. Nejběžnějším kritériem dělení těchto dodavatelů je dle velikosti firmy, která přibližně koreluje s velikostí (co do počtu funkcí a robustnosti) dodávaného produktu.

Základní dělení rozlišuje 4 velikostní skupiny (Pour et al., 2012):

1. dodavatelé komplexních řešení s největším podílem na trhu (angl. Megavendors). Nabízejí úplné portfolio potřebné k aplikaci BI od databázových systémů určených k tvorbě datových skladů a tržišť, transformačních nástrojů ETL až po analytické databáze OLAP. Největší z firem patřících do této skupiny jsou:
  - Oracle (včetně společnosti Hyperion) s databázovým systémem Oracle,
  - Microsoft s databázovým systémem MS SQL Server,
  - IBM (včetně Cognos) s databázovým systémem DB2,

- SAP (včetně Business Objects),
2. dodavatelé specializovaných řešení a nástrojů (angl. Pure players), mezi které se řadí společnosti:
- Actuate,
  - Information Builders,
  - Microstrategy,
  - SAS Institute,
3. menší dodavatelé nabízející vysoce specializované produkty (angl. Up-and-Comers), kam patří:
- Arcplan,
  - Board,
  - Panorama,
  - QlikTech,
  - Spotfire (Tibco),
  - Targetik,
4. komunitně (obvykle na bázi open-source) vyvíjené produkty BI:
- Pentaho,
  - JasperSoft.

Do tohoto seznamu může patřit i pátá skupina, která v současnosti získává na významu, a tou je software umožňující tzv. samoobslužné (Self Service) BI, na které bude práce dále zaměřena.

Zde patří např. produkty:

- PowerPivot společnosti Microsoft,
- Tableau Self-Service BI společnosti Tableau Software.

Asi málokoho překvapí fakt, že s velikostí výrobce a šíře jeho sortimentů bude v korelaci i cena výsledného řešení. Na samotném vrcholu žebříčku budou společnosti typu Oracle, kde se cena balíku Hyperion v době psaní tohoto textu pohybuje okolo cca 370 tisíc Kč (Oracle, 2014), přičemž na samotném dnu budou komunitně vyvíjené produkty typu open-source, kde se nabízejí základní verze bez technické podpory zcela zdarma.

Praktická část této práce se věnuje procesu BI prováděnému ve formě samoobslužného BI. Proto přichází v úvahu mnohem levnější varianta (na opačném konci cenového spektra), konkrétně velice známý software – kancelářský balík Office, který dodává firma Microsoft. Pořizovací cena licence verze Microsoft Office 2013 pro podnikatele se v době psaní tohoto textu pohybuje kolem 6 000 Kč vč. DPH (Zboží.cz, 2015).

Právě tento kancelářský balík, konkrétně jeho součást Microsoft Excel, je nezbytný pro běh aplikace umožňující samoobslužné BI. Microsoft pro něj dodává softwarový doplněk jménem PowerPivot, který přidává do aplikace Excel možnost využívat metody BI. Doplněk je k dispozici zcela bez poplatků pro Microsoft Excel verze 2010 a novější.

### **3.3 Doporučení pro implementace BI**

Nakonec se nabízí otázka, zda podnik dosáhne na proklamované výhody aplikace BI. Pokud se bude na projekt nahlížet z hlediska základních ekonomických charakteristik, je zvláště v případě celkové nákladovosti projektu poměrně snadné spočítat jeho celkové dopady. Pokud se jedná o výnosy, zde už bude situace těžší, protože v důsledku nových

analýz, a tím i nového pohledu na staré problémy, může vyniknout mnoho vedlejších pozitivních efektů, které nejsou prvoplánově změřitelné.

Z tohoto důvodu je v případě implementace klíčovým faktorem zaměřit se na dvě oblasti, a to především:

1. celkovou nákladovost projektu a hlavně důsledné sledování současného průběhu versus projektovaný rozpočet (velmi dobře sledovatelná veličina),
2. podstatné faktory úspěšnosti projektu IT, zvláště poté projektů typu Business Intelligence.

Jelikož v prvním případě přicházejí do úvahy obecné metody hlídání nákladovosti projektů, které jsou dobře popsány i v jiných oblastech mimo informační technologie, nebude o nich v této práci dále řečeno. Rozebrán bude především bod č. 2, faktory specifické pro projekty BI.

Pour (2012) definuje faktory úspěchu jako *takové uplatnění osvědčených přístupů, postupů a zkušeností, které povedou ke splnění stanovených cílů a požadavků kladených na projekt a k dosažení očekávaných efektů*. Jako zdroje těchto přístupů a postupů vedoucích k úspěchu především vidí vlastní zkušenosti s předchozími projekty, případně vlastní výzkumnou činnost (pokud dosud jednotka potřebné zkušenosti nemá). V otázce nejdůležitějších kategorií faktorů úspěšnosti potom přichází s následujícím rozdělením:

1. řízení podniku a podnikové informatiky (zde je vyzdvihována především vzájemná koordinace odpovědných oddělení, která je nezbytná pro hladký průběh projektu),
2. řízení Business Intelligence – organizace musí vlastnit potřebné organizační a personální zdroje, které dokáží zajistit implementaci technologie BI,
3. pracovní metody a nástroje – zejména znalost metodických postupů a dostupných nástrojů, které slouží jako významný pomocník celého procesu,
4. informační systém podniku – celková koncepce produkčních systémů podniku, jejich kvalita, dostupnost jejich dat, popsané a dodržované podnikové procesy,



5. informační a komunikační technologie – ovlivňují celkovou úroveň spolupráce při realizaci projektu.

### 3.4 Nejnovější trendy v oblasti BI

Protože je obor BI relativně nový, v současnosti velice žádaný a jsou na něj takřkajíc upřeny zraky všech významných hráčů v sektoru informačních technologií, vyskytují se zde významné investice, které přispívají prudkému rozvoji celé oblasti. V současnosti se proto setkáváme se třemi zajímavými trendy, které formují aktuální podobu celého odvětví. Pour (2012) uvádí především trendy spočívající v technologiích Cloud computing a SaaS. Kromě toho existuje ještě jeden významný proud, který se nazývá Self Service BI.

**Cloud computing** představuje odklon od tradiční distribuce softwaru, kde uživatel typicky zakoupí licenci k používanému softwaru, který si nainstaluje do svého počítače na místní fyzické datové úložiště. Data produkovaná tímto softwarem poté zůstávají prostřednictvím tohoto datového úložiště pevně svázána s jedním konkrétním počítačem, ke kterému má většinou přístup pouze jediný pracovník. V případě, kdy tedy nemůže fyzicky přistoupit k pracovní stanici, která obsahuje datové úložiště s daty programu, dočasně ztrácí veškerý přístup k datům uloženým na tomto médiu. Cloud computing se proto snaží od fyzického svázání dat s jednou stanicí či uživatelem oprostit a nabízí formou placené služby přístup k výpočetnímu výkonu (včetně datového úložiště) prakticky kdekoli a kdykoli. Hlavní princip spočívá v tom, že funkci datového úložiště (a někdy i funkci výpočetní) na sebe přebírá poskytovatel Cloud computingu (většinou datové centrum), který garantuje jeho trvalou dostupnost omezenou pouze dostupností sítě Internet.

**SaaS neboli Software as a Service** představuje princip poskytování softwaru na bázi Cloud computingu. Koncový zákazník se zde nestará o instalaci softwaru ani jeho aktualizace. Místo toho si od dodavatele pronajímá (obvykle za stálý poplatek za určité období, většinou měsíční) kompletní přístup k požadovanému softwaru. Veškeré starosti s potřebným hardwarem, instalací softwaru a jeho aktualizací, připojením do sítě Internetu obstarává dodavatel. Poskytování těchto aplikací má nejčastěji formu tenkého klienta, kdy celou aplikační logiku zpracovává dodavatelsky zajištěný stroj a koncový zákazník obsluhuje pouze grafické rozhraní, nejčastěji ve formě webové aplikace (vystavěné na standardu HTML).

Zatímco u předchozích dvou trendů byl význačný především způsob poskytování stávajícího softwaru, **Self Business Intelligence** (česky doslova jako „samoobslužné BI“) přináší poněkud větší revoluci, protože podstatně mění průběh celého procesu. Pour (2014) vidí smysl metod samoobslužné BI především v tom, že *„na základě nových technologií poskytují uživatelům prostředí pro realizaci svých analytických úloh bez nutnosti používání komplexních a obvykle velmi složitých systémů BI“*. Je zde vidět jasné zaměření na zjednodušení klasického procesu BI (který jsme v předchozích kapitolách popsali) s tím, že celý proces by po určitém zjednodušení měl zvládnout jeden uživatel („samoobslužné“ BI) nejlépe s jediným nástrojem.

Takový nástroj poté obvykle obsahuje v jednom balení jednoduchý importér zdrojových dat, díky kterému je možno tato data převést do multidimenzionálního úložiště, s kterým umí nástroj pracovat, dále pak jednoduché nástroje na transformaci dat (ETL) a v neposlední řadě podporu pro rozličné operace v dimenzionálním prostředí (zejména tvorbu pohledů typu kontingenční tabulky, či ukazatelů typu KPI).

## 4. Praktická část

Smyslem této praktické části diplomové práce je uplatnění principů BI v podnikové praxi. Na příkladu samoobslužného BI prezentuje využití zdrojových dat od jejich zjištění z produkčního systému, přes transformaci do analytické databáze až po sestavení dynamického výstupu. Výstupem je aplikace dynamické kontingenční tabulky včetně využití klíčového ukazatele výkonnosti. Ta je vytvořena za použití samoobslužného softwarového nástroje.

Důvody volby samoobslužné BI jsou především následující:

- administrativní jednoduchosti: ovládat samoobslužné BI by mělo být tak snadné, že tento proces zvládne jediná osoba bez externí pomoci,
- nákladovosti: pro celý proces vystačí pouze jediný software a nebudou tak vynaloženy další investice do většího množství produktů od vícero dodavatelů.

Jako základní a jediná aplikace k tvorbě samoobslužného BI je potom zvolena aplikace společnosti Microsoft jménem PowerPivot. Následuje popis samotné aplikace, představení nároků na výpočetní techniku a také instalace produktu.

### 4.1 Aplikace PowerPivot

Aplikace PowerPivot je softwarový doplněk kancelářského balíku Microsoft Office přímo vyvíjený společností Microsoft. Slouží pro účely doplnění tabulkového procesoru Microsoft Excel o schopnosti BI. Doplněk funguje s verzemi Microsoft Office 2010 a novější (v době psaní textu konkrétně ještě verze 2013). Doplněk je osobě vlastníci licenci k balíku nabízen zcela zdarma.

V rámci návrhu řešení je použita softwarová konfigurace spočívající z doplňku PowerPivot spolu s kancelářským balíkem Office verze 2010, to vše spuštěné pod operačním systémem Microsoft Windows XP (s nainstalovanou aktualizací Service Pack 3, která je nezbytná ke zdárné instalaci aplikace PowerPivot). Následují hardwarové a

softwarové nároky nezbytné pro úspěšné provozování a poté i samotnou instalaci doplňku do kancelářského balíku Office.

#### 4.1.1 Hardwarové nároky aplikace PowerPivot

Doporučení k instalaci vydává přímo výrobce softwaru, společnost Microsoft, která na stránkách věnovaných produktu uvádí mj. tyto hodnoty:

**Tabulka č. 2: Hardwarové nároky produktu PowerPivot**

Objekt	Popis
Procesor	32 nebo 64bitový procesor o taktu 500 MHz
Operační paměť (RAM)	Samotný doplněk vyžaduje při nulovém využití nejméně 25 MB operační paměti. Pro vytvoření prázdného sešitu potřebuje dalších 33 MB operační paměti  Velikost operační paměti je důležitá pro práci s datovými soubory. V 32bitovém operačním systému PowerPivot podporuje otevření maximálně 2GB datového souboru, v 64bitovém pak maximálně 4 GB
Diskový prostor	Samotný doplněk vyžaduje nejméně 100 MB volného místa na disku. Pro práci s daty je nutno zajistit odpovídající kapacitu vzhledem k aktuální velikosti dat

Zdroj: Hart, 2015

Jak je vidno z výše uvedené tabulky, hodnoty jsou velice příznivé a neměly by způsobovat dodatečné náklady na zakoupení nového, či aktualizaci stávajícího výpočetního vybavení podniku. Pro jednotky zpracovávající větší množství dat najednou (konkrétně více než 2 GB dat), se ovšem doporučuje aktualizace počítače na 64bitovou architekturu. Zde především závisí na souhře hardwaru i softwaru. Je nutno vlastnit nejen 64bitový procesor, ale také 64bitový operační systém (Microsoft nabízí 64bitové operační systémy počínaje verzí Windows XP).

Výrobce dále uvádí softwarové vybavení počítače nezbytné k provozu a instalaci doplňku PowerPivot:

**Tabulka č. 3: Softwarové nároky produktu PowerPivot**

<b>Objekt</b>	<b>Popis</b>
Operační systém (OS)	Microsoft Windows XP (pouze 32bitová verze), Windows Vista SP2, Windows 7, Windows 8 a Windows Server 2003/2008 v určitých konfiguracích
Nezbytné součásti OS Windows	Microsoft .NET Framework 3.5 SP1
Požadavky na Microsoft Excel	Minimální verze 2010 (32 nebo 64bitová)
Požadavky na Microsoft Office	Sdílené komponenty Microsoft Office, podpora programování .NET pro Excel

Zdroj: Hart, 2015

Z této tabulky nevyplývají žádné speciální požadavky, kterým by standardní uživatel neměl bezpečně vyhovět. Jediné produkty poskytované za úplatu zde jsou operační systém a kancelářský balík. Ty by však vzhledem k faktu, že prakticky patří ke standardnímu vybavení podnikového počítače, neměly vyvolat žádné dodatečné náklady.

Na základě těchto požadavků potom přichází zjištění, že v nejběžnějších případech by implementace produktu neměla způsobovat žádnou dodatečnou investici.

#### **4.1.2 Instalace aplikace PowerPivot**

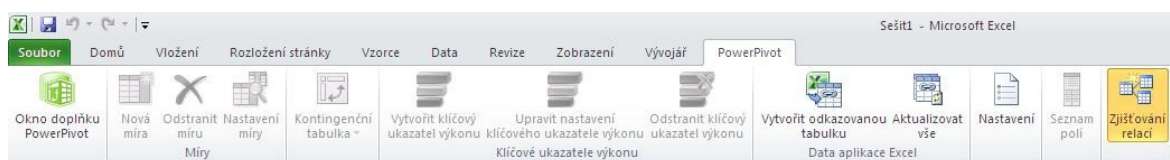
Na začátku implementace BI v podniku musí být instalace nezbytného nástroje. Kromě samotného doplňku je také nainstalována nezbytná součást Visual Studio 2010 Tools for Office Runtime, bez které je doplněk v rámci kancelářského balíku „neviditelný“.

Postup instalace poté spočívá v těchto krocích:

1. stáhnutí a instalace nezbytné součásti doplňku PowerPivot „Visual Studio 2010 Tools for Office Runtime“ z internetové stránky <http://www.microsoft.com/en-us/download/details.aspx?id=44074>,
2. jako druhý krok stáhnutí a instalace samotného doplňku PowerPivot z internetové stránky <http://www.microsoft.com/cs-cz/download/details.aspx?id=29074>

Po splnění těchto dvou jednoduchých kroků stačí spustit aplikaci Microsoft Excel a zkontrolovat, zda instalace proběhla úspěšně, nejjednodušeji pravděpodobně vizuální kontrolou, která by měla splňovat 2 podmínky:

1. kliknout na zelené tlačítko Soubor, dále pak Možnosti a v nově otevřeném okně se přepnout do záložky Doplnky. Zde by se měla v sekci nazvané *Aktivní doplňky aplikací* nacházet položka s názvem *PowerPivot for Excel* a
2. v hlavním pásu karet (tzv. nabídka nástrojů „Ribbon“) se nově objevila záložka s názvem PowerPivot, jejíž obsah je následující:



**Obrázek č. 6: Pás karet aplikace Microsoft Excel s přidaným doplňkem PowerPivot**

Nyní, po úspěšné instalaci a následné kontrole doplňku PowerPivot, následuje ukázka jeho využití v praxi. To představuje návrh řešení BI pro internetový obchod se zemědělskými komoditami, konkrétně s krmivou pro hospodářská zvířata.

## 4.2 Představení internetového obchodu

Návrh řešení je koncipován jako vhodný k využití pro malou a střední firmu, která v současné době vlastní licenci na operační systém Microsoft Windows a kancelářský balík Microsoft Office a provozuje internetový obchod. Internetový obchod většinou bývá zaměřen na pohodlnou a rychlou práci s daty, které se věnují operativnímu řešení určitého problému (nová objednávka, správa katalogu zboží, evidence kontaktů apod.). Většinou však neobsahuje takové analytické metody, které by jeho majiteli přinášeli názorný obraz o celkovém chodu, přispěly ke zlepšení jeho řízení a poskytly potřebná data pro odpovědné plánování.

Tyto nedostatky doplňuje nástroj samoobslužného BI, který potřebné analytické funkce dodává. Právě díky této technologii může prezentovaný internetový obchod objevit

nové, dosud neviděné souvislosti, které mu pomohou ke zvýšení celkového výkonu společnosti.

#### 4.2.1 Obchodní platforma

Dle dříve popsaného postupu BI jsou nejprve získána primární data ze zdrojových systémů, která jsou následně transformována do analytické databáze nástroje PowerPivot. Zdrojový systém, v tomto případě internetový obchod, je vystavěn na technologii souhrnně označované jako LAMP, což je zkratka 4 slov – Linux, Apache, MySQL a PHP, které budou v krátkosti představeny.

Spojovacím prvkem všech 4 softwarových produktů je, že jsou komunitně vyvíjeny na bázi open-source a uvolněny pro komerční i nekomerční použití bez jediného poplatku. Z hlediska nákladů pro podnikatele tedy velmi výhodná volba. **Linux** je operační systém Unixového typu, **Apache** je populární a široce používaný server sloužící k poskytování internetových stránek. **MySQL** je relační databázový server podporující standard dotazovacího jazyka SQL. **PHP** je interpretovaný programovací jazyk, který zde slouží jako hlavní pilíř aplikační logiky.

V otázce využití BI bude nejdůležitější součástí databázový server MySQL, ve kterém jsou v současnosti uložena veškerá důležitá data týkající se provozu obchodu. Databázový server zde slouží v terminologii BI jako zdroj produkčních dat.

#### 4.2.2 Produkční zdrojový systém

Jelikož jsou současné ovládací nástroje internetového obchodu zaměřeny především na funkce, ke kterým jsou primárně produkční zdrojové systémy určeny – tedy k rychlému vytváření, čtení, úpravě a mazání záznamových údajů – vystává logická potřeba tato data analyzovat z hlediska jejich vzájemnosti a zajímavých vztahů.

Jinými slovy, hlavní zájem není kladen na informaci, že zákazník A objednal toho a toho dne objednávkou B zboží C v hodnotě D. Je potřeba znát komplexnější pohled na celou problematiku s možností časové projekce a vzájemných závislostí zkoumaných veličin. Z tohoto důvodu se převádí data v současnosti uložená v produkční databázi do analytické databáze doplňku PowerPivot.

#### 4.2.2.1 Uložení dat v produkční databázi

Systém celého internetového obchodu zaplňuje jednu databázi a určité množství tabulek, které jsou pro jeho hladký a bezvadný chod nezbytné. Tabulky jsou v rámci zachování referenční integrity vzájemně propojeny pomocí cizích klíčů dle principů relační databáze.

V této práci je využit pouze malý zlomek celé databáze, a to konkrétně těchto pět tabulek:

1. adresář dodavatelů a odběratelů,
2. kniha vydaných faktur,
3. výčet položek náležících k jednotlivým fakturám,
4. skladové karty jednotlivého zboží,
5. číselník kategorií zboží.

Tabulka **adresáře** obsahuje adresní záznamy jednotlivých obchodních partnerů, se kterými podnikatelská jednotka spolupracuje. Pro každého partnera si eviduje jeho název a poštovní adresu sestávající z ulice, města a poštovního směrovacího čísla.

**Tabulka č. 4: Ukázka tabulky adresar**

id	typ	nazev	ulice	mesto	psc
1	dodavatel	Družstvo Adam, a.s.	Na dolínách	Praha	14700
2	dodavatel	Kombinát Cyril, s.r.o.	Struhařov	Struhařov	25601
3	odberatel	Hana Novotná	Borůvková	Olomouc	77900
4	odberatel	Karel Krejčí	Portmanka	Litomyšl	57001
5	odberatel	Miroslav Němec	Podskalská	Kolín	33040

Zdroj: vlastní tvorba



Tabulka **faktury** obsahuje knihu vydaných faktur, které korespondují s učiněnými objednávkami zákazníků internetového obchodu. Pro každou vydanou fakturu se eviduje její číslo (zejm. pro daňové účely v případě daňového dokladu), údaj o odběrateli (kupujícím), datum vydání a v neposlední řadě celkovou částku faktury. Tabulka využívá relačního propojení, kdy hodnota odběratele je určena cizím klíčem z tabulky adresáře a je tak zaručena snadná výsledovatelnost a unikátnost obchodního partnera.

**Tabulka č. 5: Ukázka tabulky faktury**

id	cislo	odberatel_id	datum_vytvoreni	castka
1	FV00001	30	2014-01-01	6870.00
2	FV00002	56	2014-01-01	1250.00
3	FV00003	16	2014-01-01	1480.00
4	FV00004	80	2014-01-01	650.00
5	FV00005	117	2014-01-01	8210.00

Zdroj: vlastní tvorba

Tabulka **položky** slouží jako pomocná tabulka pro knihu faktur, kdy se zpravidla předpokládá, že jedna faktura bude obsahovat více položek (jednostranný vztah 1:N). Tabulka využívá dvou cizích klíčů: jeden pro položku faktury, kde se eviduje číslo faktury, ke které patří, druhou pak pro zboží, které je reprezentováno číslem skladové karty.

**Tabulka č. 6: Ukázka tabulky položky**

id	faktura_id	zbozi_id	pocet	castka	soucet
1	1	5	6	200.00	1200.00
2	1	18	3	220.00	660.00
3	1	2	4	210.00	840.00
4	1	19	6	270.00	1620.00
5	1	23	6	250.00	1500.00

Zdroj: vlastní tvorba

Tabulka **zboží** slouží jako seznam skladových karet, artiklů, které internetový obchod nabízí k prodeji. Tabulka obsahuje dva cizí klíče, a to vazbu vzhledem k tabulce adresáře a

kategorie zboží. V rámci těchto vazeb je znázorněn dodavatel té které skladové položky, resp. kategorie zboží.

**Tabulka č. 7: Ukázka tabulky zboží**

id	dodavatel_id	kategorie_id	nazev	cena
1	1	1	Bažant základ	160.00
2	1	1	Bažant plus	210.00
3	1	1	Bažant komplet	310.00
4	1	2	Kachna základ	150.00
5	1	2	Kachna plus	200.00

Zdroj: vlastní tvorba

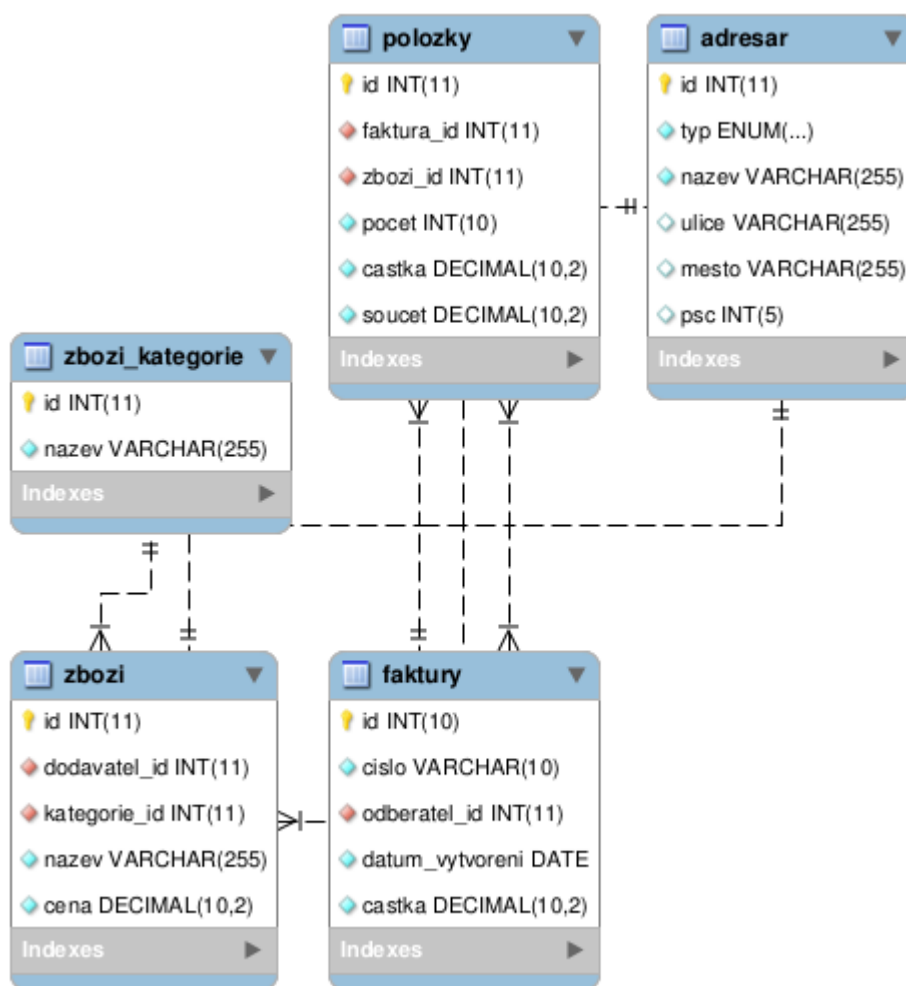
Poslední tabulkou je tabulka **kategorie zboží**, která slouží jako výčtový typ pro tabulku zboží a reprezentuje jednotlivé kategorie prodávaných krmiv. Kategorie v tabulce zboží díky tomu nemusí být uváděny svým názvem, ale používá se zde cizího klíče, který na daný název odkazuje. Toto vydělení kategorie do samostatné tabulky může usnadnit analytickou práci s tabulkou.

**Tabulka č. 8: Ukázka tabulky zboží\_kategorie**

id	nazev
1	Krmiva pro bažanty
2	Krmiva pro kachny
3	Krmiva pro koně
4	Krmiva pro králíky
5	Krmiva pro ovce

Zdroj: vlastní tvorba

Celkové schéma je pak pro větší přehlednost a lepší orientaci znázorněno tímto modelem E-R (Entity–relationship model):



Obrázek č. 7: Logické schéma tabulek v relační databázi

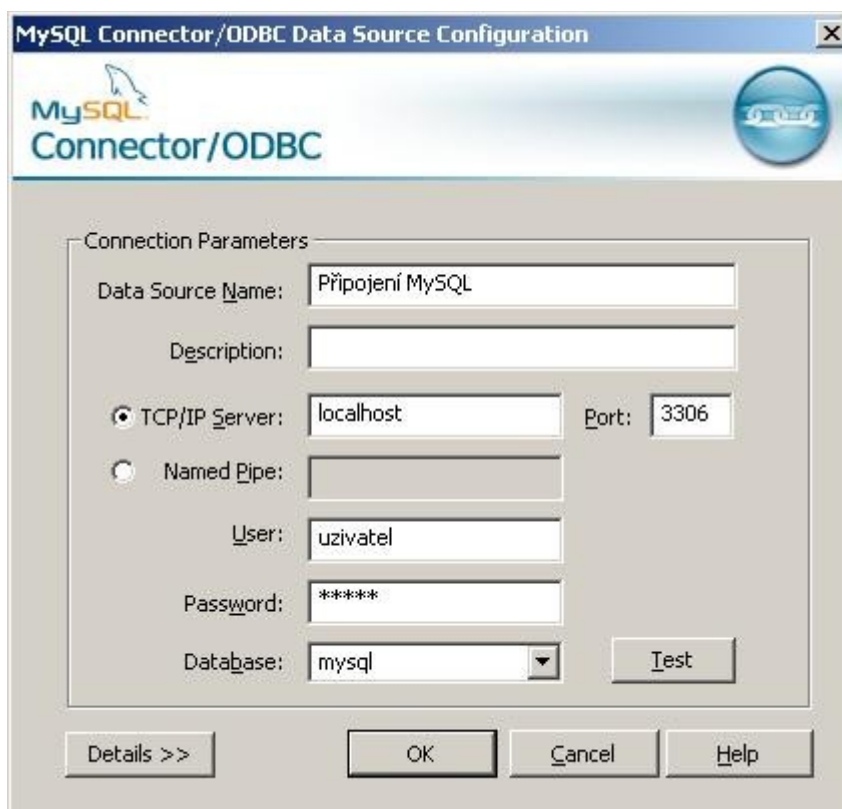
### 4.3 Transformace produkčních dat do analytické databáze (ETL)

Pro možnost práce s daty v analytickém nástroji PowerPivot je potřeba provést import dat ze zdrojové databáze MySQL do doplňku PowerPivot. Ten naštěstí již od výroby podporuje standard ODBC (Open Database Connectivity), díky kterému dojde snadno k propojení databázového serveru (MySQL) s analytickým datovým skladem (PowerPivot).

Prvním krokem je stažení a instalace ovladače databáze MySQL, který splňuje standard ODBC. Z internetových stránek <http://dev.mysql.com/downloads/connector/odbc/> je stažena 32bitová verze ovladače pro operační systém Microsoft Windows a následně provedena instalace.

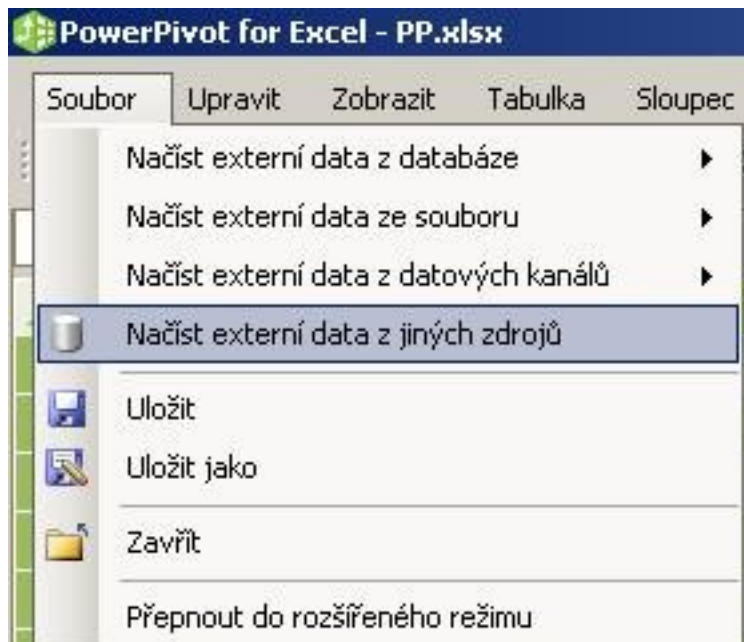
Po úspěšné instalaci dojde prostřednictvím tohoto ovladače k vytvoření připojení k databázi, pomocí kterého se data do doplňku PowerPivot importují. Připojení k databázi se nastavuje nástrojem *Nastavení* → *Ovládací panely* → *Nástroje pro správu* → *Datové zdroje (ODBC)*. Zde v záložce *Uživatelské DSN* kliknutím na tlačítko *Přidat*. V nově otevřeném okně, které vyzývá k výběru ovladače, dále označením položky *MySQL ODBC 5.3 Unicode Driver*. Celý výběr se potvrdí kliknutím na tlačítko *Dokončit*.

Následně se vyplní potřebné údaje, zejména *Data Source Name* (libovolný název připojení), *TCP/IP Server* (adresa, na které poslouchá server MySQL) a v neposlední řadě pole *User* (uživatelské jméno) a *Password* (heslo) a *Database* (jméno naší databáze). Poté může být připojení otestováno kliknutím na tlačítko *Test*. V případě úspěšné operace by nyní mělo být spojení s databází umožněno napříč operačním systémem Microsoft Windows. Toto je nezbytný předpoklad následné práce v doplňku PowerPivot.

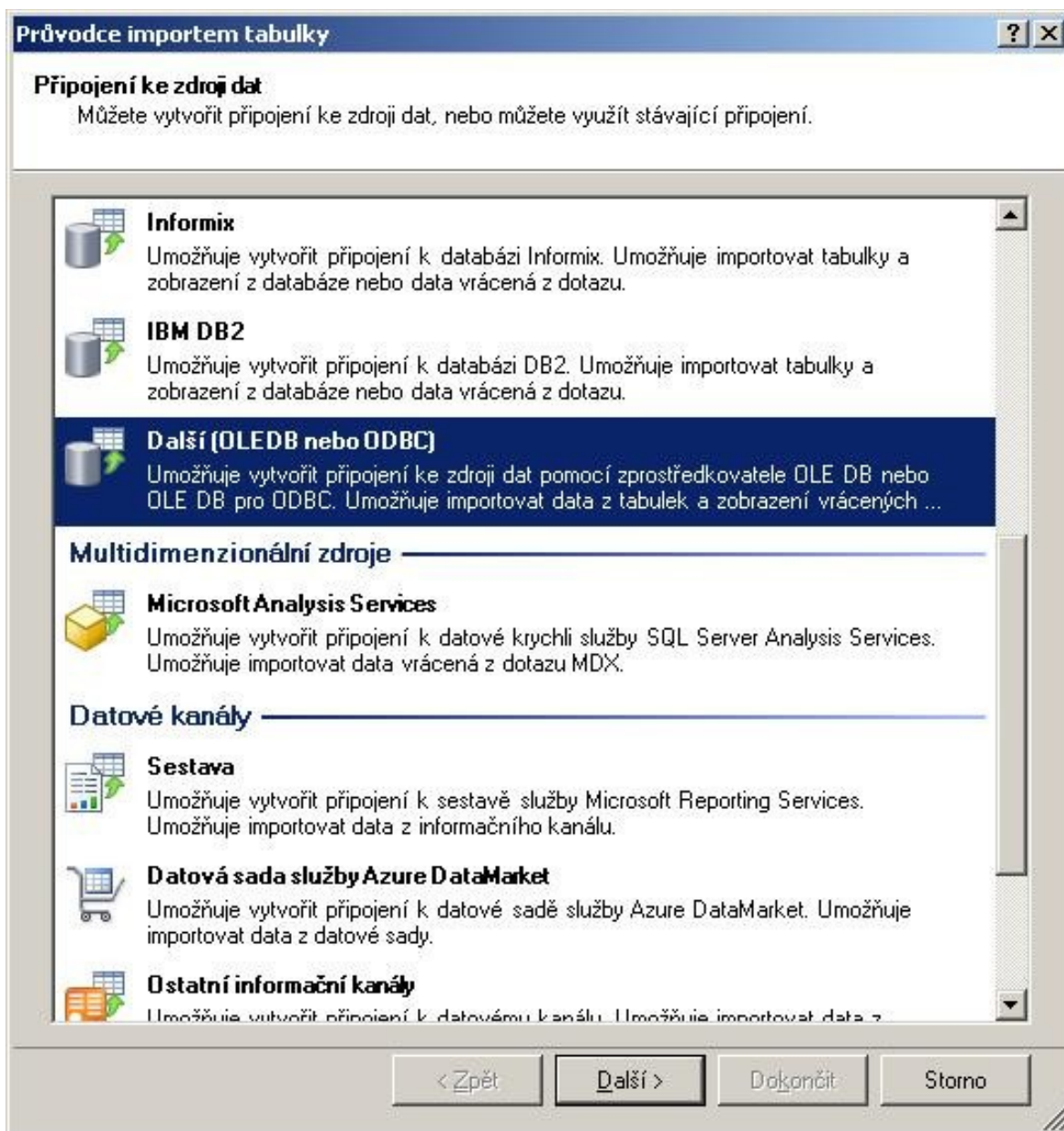


**Obrázek č. 8: Nastavení spojení serveru MySQL**

Následuje přepnutí se zpátky do aplikace Excel, kde kliknutí na tlačítko *Okno doplňku PowerPivot* v záložce *PowerPivot* pásu karet otevře hlavní okno doplňku. Zde následuje výběr položky z hlavní nabídky *Soubor* → *Načíst externí data z jiných zdrojů*.



**Obrázek č. 9: Výběr volby z nabídky doplňku PowerPivot**



V nově otevřeném okně ze seznamu nabízených možností následně vybere položka *Další (OLEDB nebo ODBC)*, která se nachází v kategorii *Relační databáze*.

### Obrázek č. 10: Výběr typu datového zdroje

V dalším kroku se zadává libovolný popisný název pro nově vytvořené spojení a sestavuje se připojovací řetězec kliknutím na tlačítko *Sestavení...*. V nově otevřeném okně se pod bodem č. 1 *Zadejte zdroj dat* zaškrtně možnost *Použít název zdroje dat* a v seznamu pod touto možností se následně vybere připojení dle názvu, které bylo předtím vytvořeno v

rámci nastavení *Datové zdroje (ODBC)*. Krok je potvrzen kliknutím na tlačítko *OK* a poté na tlačítko *Další*.

V následujícím kroku zůstává zaškrtnutá možnost *Vybrat ze seznamu tabulek a zobrazení data, která mají být importována* a pokračuje se na dále kliknutím na tlačítko *Další*.

Pokud je postupováno správně, přichází poslední krok importu dat do doplňku PowerPivot. Tento krok zajišťuje výběr konkrétních tabulek, které mají být do analytické databáze importovány. Výběr je zajištěn zobrazeným výčtem s možností selekce konkrétních tabulek. Dojde tedy k zaškrtnutí požadovaných tabulek a proces je zakončen tlačítkem *Dokončit*.

Každá takto přidaná tabulka se zařadí jako nový list do spodní části okna doplňku PowerPivot. V rámci doplňku se tabulka v podstatě zobrazuje jako klasická ručně vytvořená tabulka v aplikaci Microsoft Excel. Nyní by se v okně doplňku mělo nacházet 5 těchto záložek, každá reprezentující jednu původní tabulku z databáze MySQL.



**Obrázek č. 11: Seznam importovaných tabulek do doplňku PowerPivot**

#### **4.4 Analytická práce s importovanými daty**

Nyní přichází tolik žádaná analytická práce nad víceméně „surovými“ daty, která byla importem do doplňku PowerPivot přidána. Součástí tohoto příkladu je definice a deklarace vlastní proměnné, využití dotazovacího jazyka DAX, definice vlastní míry a klíčového ukazatele výkonu (KPI). Výstupem příkladu pak přehled měsíčních tržeb v roce 2014 s klíčovým ukazatelem výkonu, který vizuálně indikuje, zda tržby v daném měsíci byly pod či nad průměrem tržeb celého roku.

**Tabulka č. 9: Seznam použitých funkcí v tomto příkladu**

Název funkce	Popis funkce	Syntaxe
CALCULATE	Vyhodnotí výraz s hodnotami upravenými filtry	CALCULATE(<výraz>, <filtr1>, <filtr2>...)
ALLSELECTED	Vrátí všechny řádky v tabulce nebo všechny hodnoty ve sloupci, přičemž jsou ignorovány všechny filtry, které mohly být použity uvnitř dotazu, ale je zachováno filtrování určené zvenčí.	ALLSELECTED([<tabulka>   <sloupec>])
DISTINCTCOUNT	Spočítá počet jedinečných hodnot v určeném sloupci	DISTINCTCOUNT(<sloupec>)
SUM	Sečte hodnoty všech řádků v určeném sloupci	SUM(<sloupec>)
MONTH	Vrátí číslo od 1 (leden) do 12 (prosinec) představující měsíc	MONTH(<datum>)
YEAR	Vrátí rok data jako čtyřmístné celé číslo	YEAR(<datum>)

Zdroj: dokumentace k jednotlivým příkazům na stránkách Microsoft Developer Network ([msdn.microsoft.com](http://msdn.microsoft.com))

Vše začíná přidáním dvou nových sloupců do tabulky faktury, které usnadní následný výpočet. Bude se jednat o sloupce `mesic` a `rok`. Přidání sloupců proběhne v datové tabulce faktury, kde kliknutím na záhlaví *Přidat sloupec* dojde v oblasti určené pro zapsání funkce (označené stejně jako v aplikaci Excel symbolem *fx*) k zapsání příkazu v jazyce DAX:

```
=MONTH(faktury[datum_vytvoreni])
```

Výstupem této funkce je, že pro každý řádek v tabulce zobrazí pořadové číslo měsíce v roce. Výraz `faktury[datum_vytvoreni]` znamená, že se odkazuje na sloupec `datum_vytvoreni`, který se nachází v tabulce faktury.



Analogický postup je uplatněn v případě sloupce roku. Do nového sloupce je vepsán vzorec:

```
=YEAR(faktury[datum_vytvoreni])
```

Pod tabulkou s daty se nachází oblast nazývaná *Oblast výpočtu*. Tato oblast se vlastně podobá klasické tabulce s tím, že je přímo určená ke psaní vlastních vzorců (výpočtů). V textu je jí dále využito k zapsání vlastních funkcí pomocí jazyka DAX.

Nejprve je vytvořena proměnná `pocetMesicu`, která reprezentuje počet unikátních hodnot měsíců zahrnutých do datové sady. V případě ročního období by těchto hodnot mělo být 12. Pro výpočet výskytu unikátních hodnot se používá funkce `DISTINCTCOUNT()`.

Zde přichází na řadu vlastnost, která se v terminologii doplňku PowerPivot nazývá kontext. Kontext je v podstatě souhrn datových hodnot, které jsou brány v potaz při určitém dotazu. Doplňek PowerPivot rozlišuje celkem 3 kontexty (Context in DAX Formulas, 2015):

1. kontext řádku,
2. kontext dotazu,
3. kontext filtru.

Ve většině případů program implicitně používá řádkový kontext, který způsobuje, že zapsané funkce jsou vždy vyhodnocovány proti hodnotám, které se nacházejí na stejném řádku jako výsledek funkce. Protože je však v tomto případě potřeba zjistit hodnoty měsíců z celé tabulky, musí být vlastnost kontextu potlačena. K tomu slouží funkce `ALLSELECTED()`, která vrací všechny řádky tabulky bez ohledu na současný kontext.

Funkci `ALLSELECTED()` je aplikována jako filtr prostřednictvím funkce `CALCULATE()`. Do prvního řádku oblasti výpočtu se zapíše vzorec, který explicitně definuje proměnnou `pocetMesicu`:

```
pocetMesicu:=CALCULATE(DISTINCTCOUNT(faktury[mesic]);  
ALLSELECTED())
```

Následuje proměnná `mesicniPrumer`, která zjišťuje průměrnou měsíční tržbu v daném roce. Její výpočet je klasický aritmetický průměr, kdy je součet tržeb za všech 12 měsíců (opět se využívá funkce `ALLSELECTED()`) vydělen celkovým počtem měsíců. Celý vzorec je následně zapsán do druhého řádku oblasti výpočtu:

```
mesicniPrumer:=CALCULATE(SUM(faktury[castka]); ALLSELECTED())  
/ faktury[pocetMesicu]
```

Předposlední proměnná je `soucetMesic`, která provádí sumu všech tržeb učiněných v daném měsíci. Zde je využita klasická, z Excelu dobře známá funkce `SUM()`:

```
soucetMesic:=SUM(faktury[castka])
```

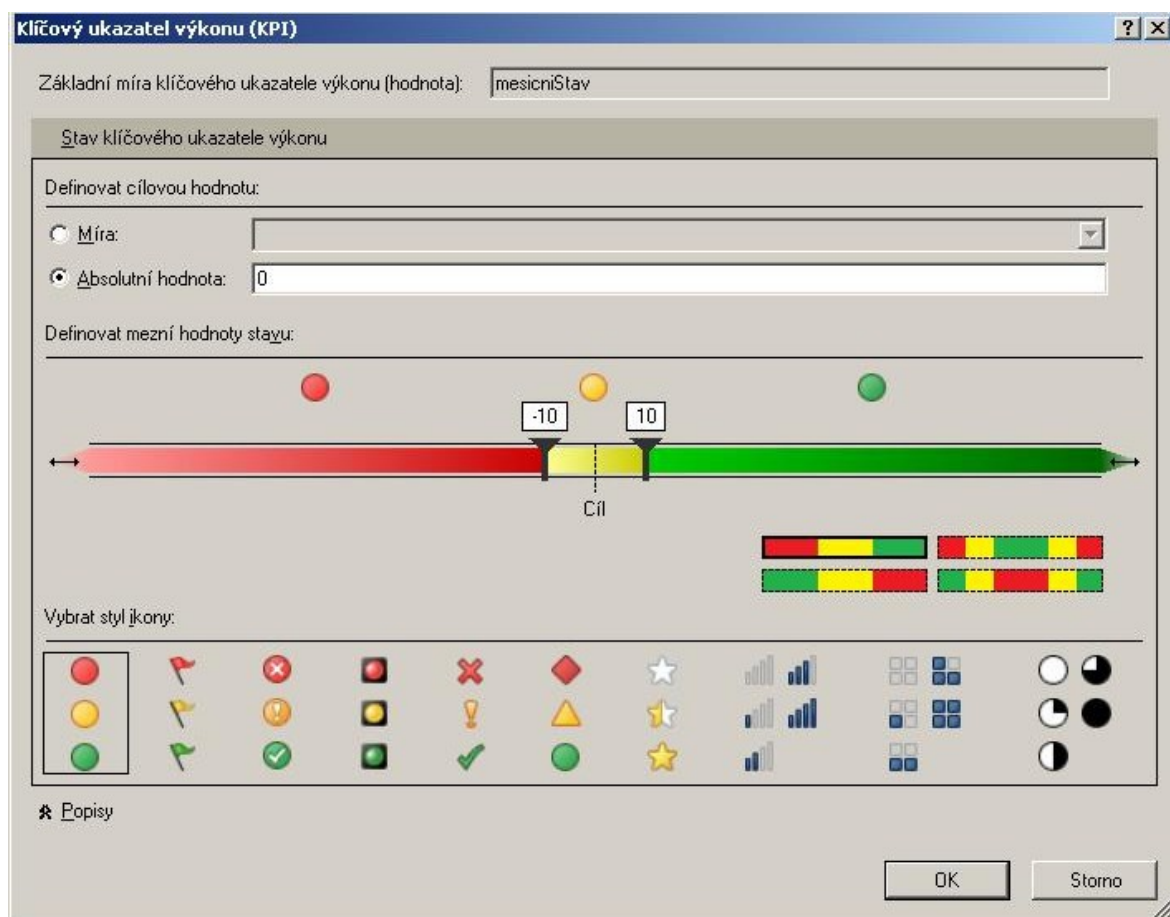
Poslední proměnnou je `mesicniStav`, která slouží jako aktuální indikátor, zda se v daném měsíci prodalo méně, či více, než je měsíční průměr pro daný rok. Deklarace proměnné vypadá takto:

```
mesicniStav:=((faktury[soucetMesic] - faktury[rocniPrumer]) /  
faktury[rocniPrumer]) * 100
```

Vedle použití vlastních proměnných umí doplněk PowerPivot mimo jiné v rámci kontingenční tabulky zobrazit plnění uživatelem definovaného klíčového ukazatele výkonnosti. Tato funkce umožní podnikateli přehledně sledovat, jak si jeho obchod vedl v daném měsíci. Ukazatel lze přidat kliknutím pravým tlačítkem do řádku s proměnnou `mesicniStav` a vybráním volby *Vytvořit klíčový ukazatel výkonu...*

V nově otevřeném okně v sekci *Definovat cílovou hodnotu:* se vybere přepínač *Absolutní hodnota*, a do editačního políčka se zadá číslo 0. Následně proběhne definice mezní hodnoty stavu, aby mohl program barevně zobrazit, zda byl měsíční prodej podprůměrný nebo nadprůměrný.

Barevná škála v intervalu  $(-100;100)$  se k tomuto účelu rozčlení na 3 barevně odlišené skupiny, když se do políčka nad červeným puntíkem vepíše hodnota -10 a do políčka nad zeleným puntíkem 10. Celá operace se potvrdí tlačítkem *OK*. Nastavení ukazatele je znázorněno na následujícím obrázku:



**Obrázek č. 12: Nastavení klíčového ukazatele výkonu (KPI)**

Nyní práce s datovým pohledem doplňku končí a na základě takto připravené databáze je vytvořena analytická tabulka s přehledem tržeb.

#### 4.4.1 Tvorba kontingenční tabulky

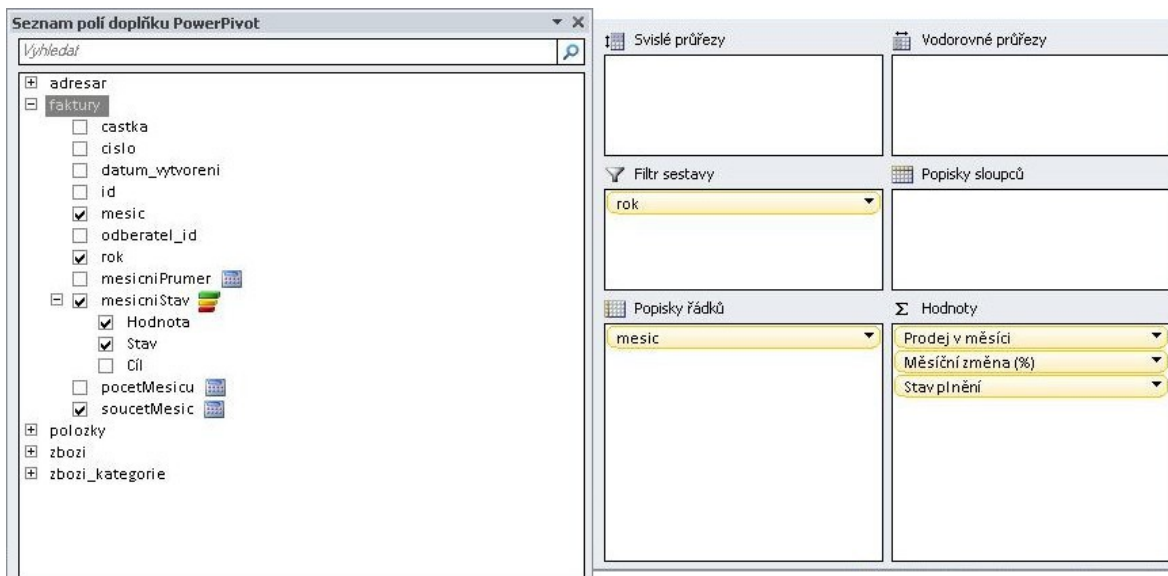
K tvorbě kontingenční tabulky je nutno přepnout se do jiného zobrazení, z doplňku PowerPivot do okna klasického Excelu. Přepnutí je možno provést kliknutím na položku v hlavní nabídce *Zobrazit* → *Kontingenční tabulka*. V potvrzovacím dialogu stačí ponechat výchozí volbu *Nový list*, která pro práci s tabulkou vytvoří nový čistý list.

Nový list je v aplikaci rozdělen na dvě nestejně velké části, přičemž levá část obsahuje prostor alokovaný pro nově vzniklou kontingenční tabulku, ta pravá zase její nastavení a nazývá se Seznam polí doplňku PowerPivot. Právě pravá část je klíčovou oblastí tvorby tabulky a je pro tento účel využita.

Největší plochu zabírá seznam datových tabulek a jejich sloupců, jejichž data mohou být využita k tvorbě výsledné sestavy. Následují sekce *Svislé průřezy*, *Vodorovné průřezy* a *Filtr sestavy*. Všechny tři slouží k detailnímu filtrování kontingenční tabulky, kdy umožňují zadávat různorodé podmínky výběru a zvyšovat tak granularitu zobrazovaných údajů. V tomto případě se využívá sekce *Filtr sestavy*, kdy jsou zobrazeny pouze ty řádky, které splňují danou podmínku roku, zde roku 2014. Do sekce *Filtr sestavy* je přidána proměnná `rok`.

Poslední 3 sekce se jmenují *Popisky sloupců*, *Popisky řádků* a *Hodnoty*. Ty na rozdíl od předchozích 3 jmenovaných neslouží k nastavení granularity dat, ale nastavují výběr zobrazených údajů. V tomto případě se nastavují řádky tabulky, které zobrazují jednotlivé měsíce v roce. V sekci *Popisky řádků* je proto přidána proměnná `mesic`. Poslední sekci, která je v tomto příkladu využita, je sekce *Hodnoty*. Ta definuje datové řádky, jejichž celkový součet budou zobrazeny pro každou hodnotu vycházející z nastavení v sekci *Popisky řádků* (v tomto případě tedy jednotlivé součty za každý kalendářní měsíc). První kalkulovanou proměnnou je `soucetMesic`, která zobrazuje sumu prodejů v daném kalendářním měsíci a roce daném filtrační podmínkou.

Další dva záznamy v sekci *Hodnoty* se týkají předpočítaného klíčového ukazatele výkonnosti `mesicniStav`. Pro jeho zobrazení v kontingenční tabulce se v rozbalovacím stromu tabulky *faktury* vybere vlastnost `Hodnota` a `Stav`, které obě patří pod ukazatel `mesicniStav`. Přidání tohoto ukazatele zajistí rychlou a přehlednou vizuální orientaci, která bude ihned indikovat, jak byl který měsíc v roce úspěšný. Kompletní nastavení kontingenční tabulky v pravé části nového listu pak vypadá takto:



**Obrázek č. 13: Nastavení datových prvků kontingenční tabulky**

Po správném výběru všech podstatných proměnných, které se mají účastnit tvorby kontingenční tabulky vypadá výsledná tabulka jako na následujícím obrázku (*ukázková tabulka je dále upravena – především změnou barevného formátu, formátu čísel a dále přejmenováním záhlaví na srozumitelné české výrazy*).

rok	2014		
Měsíc	Prodej v měsíci	Měsíční změna (%)	Stav plnění
1	377 430,00 Kč	-11,86	●
2	360 420,00 Kč	-15,84	●
3	382 030,00 Kč	-10,79	●
4	337 100,00 Kč	-21,28	●
5	400 830,00 Kč	-6,40	●
6	471 450,00 Kč	10,09	●
7	469 900,00 Kč	9,73	●
8	481 280,00 Kč	12,39	●
9	424 730,00 Kč	-0,82	●
10	483 720,00 Kč	12,96	●
11	458 290,00 Kč	7,02	●
12	491 630,00 Kč	14,80	●
<b>Celkový součet</b>	<b>5 138 810,00 Kč</b>		

**Obrázek č. 14: Výsledná kontingenční tabulka**

Z kontingenční tabulky je patrné, že poskytuje okamžitý vizuální přehled, jak se danému podniku dařilo v průběhu celého roku 2014. Zatímco na začátku roku až do měsíce dubna byly tržby ve vztahu k ročnímu průměru spíše podprůměrné, což značí červený puntík umístěný na daném řádku. Od tohoto měsíce a dále nabraly měsíční tržby trajektorii pozvolného růstu. Tyto hodnoty jsou vyjádřeny žlutou (změna pod 10 %) a zelenou barvou (změna nad 10 %). K největší kulminaci potom došlo v posledním měsíci roku 2014, kde byl zaznamenán roční rekord, který znamenal zhruba 15 % nárůst oproti průměrné roční tržbě.

## 5. Závěr

Cílem této práce bylo seznámení s relativně novým, v současné době rychle rostoucím a populárním odvětvím analytického zpracování dat výpočetní technikou, které se souhrnně označuje jako Business Intelligence. Nejprve byl vysvětlen samotný pojem Business Intelligence, dále byla rozebrána historie vývoje odvětví a nakonec popsán celkový proces jako takový. Nakonec byly zmíněny dostupné softwarové nástroje a aktuální trendy tvořící toto odvětví.

V rámci návrhu řešení BI bylo na datech internetového obchodu s krmivy pro hospodářská zvířata předvedeno, jakým způsobem může zájemce začít s informacemi, které vlastní a sbírá v rámci podnikatelské činnosti, analyticky pracovat. Záměrně byla zvoleno samoobslužné BI, které mělo ukázat, že není potřeba se tohoto nového nástroje bát a že i menší podnikatel, který analytickou práci s daty doteď neprováděl, může prakticky okamžitě začít. To vše se zdarma dostupným doplňkem PowerPivot tabulkového procesoru Excel, který usnadňuje periodické plánování a vyhodnocování vlastní podnikatelské činnosti.

Díky použití zvoleného nástroje byla zajištěna především rychlost, jednoduchost a efektivnost celého procesu a v neposlední řadě samoobslužnost celé operace, kdy k provedení stačila pouze jediná osoba. Další výhody spočívají v nulových investicích na pořízení produktu, minimálních investicích na zaškolení odpovědného personálu (podíl práce probíhá v dobře známém tabulkovém procesoru), nevznikla zde nutnost investovat do tvorby transformačních datových pump na zakázku a mohla být plně využita stávající softwarová i hardwarová architektura podniku.

Z hlediska ekonomického přínosu pro podnikatele se jedná především o snadnou a pohodlnou cestu, jak pracovat s důležitými daty o uskutečněných prodejkách a vyvozovat z nich klíčové závěry pro ekonomickou činnost. Zatímco do této chvíle neměl podnikatel téměř žádný přehled o vývoji svého podnikání (resp. mu tyto přehledy samostatný internetový obchod neumožňoval), tak s využitím samoobslužného BI nyní může začít vyhodnocovat, jak se jeho podniku daří na roční, měsíční, či týdenní bázi. Díky využití barevně odlišeného klíčového ukazatele výkonnosti navíc okamžitě rozpozná trend vývoje

a může tak na danou situaci rychle zareagovat odpovídajícím způsobem. Přehledná sestava historického vývoje může sloužit jako velice kvalitní podklad pro plánování na další časová období.

V práci byla ukázána pouze část možností využití samoobslužného BI, ale základní princip, nejdůležitější komponenty, byl odhalen v celé své šíři. Jakmile totiž jednou dojde k převodu dat z provozních systému do transakční databáze, jsou analytické možnosti nástroje prakticky neomezené. Poté s nimi může podnik pohodlně pracovat a vytvářet nekonečné množství uživatelských sestav, ať už v režimu ad-hoc, případně v režimu souboru stálých sestav, které se budou v rámci daných vnitropodnikových směrnic periodicky vyhodnocovat a poté je na ně reagováno příslušnými kroky. Všechny operace jsou dostupné na pár dalších kliknutí či zapsání nových funkcí v jazyce DAX. To vše zcela bezpečně bez dalšího kontaktu se zdrojovými daty, která zůstávají nezměněna v produkční databázi.

Na druhou stranu nemůže být samoobslužné BI považováno za dokonalého spasitele ve všech situacích. Zvláště v prostředí velkých podniků, které zpracovávají řádově vyšší objemy dat, budou muset tyto jednotky uvažovat o robustnějších produktech v samozřejmě úplně jiných cenových relacích pořízení. Budou existovat také menší procenta případů, ve kterých bude nutno sáhnout k zakázkovým úpravám a tím se celý proces může dále prodražit. Ve velké většině případů malých a středních podnikatelů by však prezentované řešení mělo být dostupné a plně funkční bez jakýchkoli úprav již od výroby.

Jelikož je celé odvětví BI v současnosti poměrně populární, do budoucna můžeme očekávat rostoucí zájem na všech úrovních podnikového prostředí. Kde již technologie BI zavedeny jsou, tam se nepochybně bude pracovat na jejich vyvíjení, zlepšování a postupném prorůstání celou organizační strukturou. Protože je však odvětví relativně nové, dá se očekávat, že ve spoustě firem bude buď úplně neznámým pojmem, případně ve smělych začátcích, kde budou probíhat první implementace metodou pokusu a omylu. Tak či tak, do dalších let přináší odvětví jako takové potenciál velkého a zajímavého růstu.



## 6. Seznam použitých zdrojů

COLES, Paul, Tony COX, Chris MACKEY a Simon RICHARDSON. The toxic terabyte. In: [online]. [cit. 2014-07-15]. Dostupné z: [http://www-935.ibm.com/services/no/cio/leverage/levinfo\\_wp\\_gts\\_thetoxic.pdf](http://www-935.ibm.com/services/no/cio/leverage/levinfo_wp_gts_thetoxic.pdf).

POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business Intelligence v podnikové praxi. 1. vyd. Praha: Professional Publishing, 2012. ISBN 978-80-7431-065-2.

LABERGE, Robert. Datové sklady: Agilní metody a Business Intelligence. 1. vyd. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.

DOHNAL, Jan; POUR, Jan. Řízení podnikové informatiky a podpora byznysu. Systémová integrace, 2013. ISSN 1804-2716.

QUAGINI, Luca. Analytic Excellence: what are the key trends from the Gartner BI Summit 2012. [online]. 2012 [cit. 2014-10-05]. Dostupné z: <http://www.sdgggroup.com/es/blog/analytic-excellence-what-are-key-trends-gartner-bi-summit-2012>.

TICHÁ, Ivana; HRON, Jan. Strategické řízení. Vyd. 1. Praha: Provozně ekonomická fakulta ČZU v Praze ve vydavatelství Credit, 2002. ISBN 978-802-1309-227.

NOVOTNÝ, Ota; POUR, Jan; SLÁNSKÝ, David. Business Intelligence: jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada, 2005. ISBN 80-247-1094-3.

LUHN, Hans Peter. A Business Intelligence system. IBM Journal of Research and Development, 1958. ISSN 0018-8646.

PARMENTER, D. Key performance indicators. Developing, implementing, and using winning KPIs. John Wiley & Sons, Inc, Hoboken, New Jersey, 2007. ISBN 0-470-09588-1.

POUR, Jan. Self-service Business Intelligence. Systémová integrace, 1-2/2014. ISSN 1804-2716.

LACKO, Ľuboslav. 1001 tipů a triků pro SQL. Vyd. 1. Brno: Computer Press, 2011. ISBN 978-80-251-3010-0.

ICT Revue. Možnosti a funkcionality Business Intelligence nástrojů pohledem dodavatelů těchto řešení. [online]. [cit. 2015-03-01]. Dostupné z: <http://archiv.ihned.cz/c1-63230010-anketa-moznosti-a-funkcionality-business-intelligence-nastroju-pohledem-dodavatelu-techto-reseni>.

PRICE, Ed. PowerPivot: Install the PowerPivot Add-In for Excel. [online]. [cit. 2015-01-15]. Dostupné z: <http://social.technet.microsoft.com/wiki/contents/articles/714.powerpivot-install-the-powerpivot-add-in-for-excel.aspx>.

PowerPivot DAX: CALCULATE is a supercharged SUMIF. [online]. [cit. 2015-01-12]. Dostupné z: <http://www.powerpivotpro.com/2009/12/powerpivot-dax-calculate-is-a-supercharged-sumif/>.

Context in DAX Formulas. MICROSOFT. [online]. [cit. 2015-02-15]. Dostupné z: <https://msdn.microsoft.com/en-us/library/gg413423%28v=sql.110%29.aspx>.

Create a Measure and KPI (Tutorial). [online]. [cit. 2015-02-15]. Dostupné z: <https://msdn.microsoft.com/en-us/library/hh272049%28v=sql.110%29.aspx>.

SHOSHANI, Arie. OLAP and statistical databases: Similarities and differences. In: Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems. ACM, 1997.

CHEN, Peter Pin-Shan. The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1976.

Oracle Business Intelligence Applications Global Price List: Software Investment Guide. [online]. 2014 [cit. 2014-11-11]. Dostupné z: <http://www.oracle.com/us/corporate/pricing/price-lists/business-intelligence-price-list-071347.pdf>.

Zboží.cz: Microsoft Office 2013 pro podnikatele CZ. [online]. [cit. 2015-01-10]. Dostupné z: <http://www.zbozi.cz/vyrobek/microsoft-office-2013-pro-podnikatele-cz/>.

SOAT, John. Oracle Advances Big Data, Enterprise Mobility, and Cloud Services. [online]. [cit. 2014-12-11]. Dostupné z: <http://www.oracle.com/us/corporate/oracle-openworld-news/thomas-kurian-keynote/index.html>.

HART, Michele. PowerPivot: Install the PowerPivot Add-In for Excel. [online]. [cit. 2015-01-18]. Dostupné z: <http://social.technet.microsoft.com/wiki/contents/articles/714.powerpivot-install-the-powerpivot-add-in-for-excel.aspx>.