

Vysoká škola ekonomická v Praze
Fakulta managementu v Jindřichově Hradci

Diplomová práce

Kateřina Krsová
2015



Vysoká škola ekonomická v Praze

Fakulta managementu v Jindřichově Hradci

Katedra exaktních metod

Software pro vizualizaci dat na platformě Microsoft v porovnání s vybraným open source produktem

Autor diplomové práce: Bc. Kateřina Krsová

Vedoucí diplomové práce: Ing. Jiří Přibil, Ph.D.

Rok obhajoby: 2015

Čestné prohlášení

Prohlašuji, že diplomovou práci na téma

*„Software pro vizualizaci dat na platformě Microsoft v porovnání s vybraným open
source produktem“*

jsem vypracovala samostatně a veškerou použitou literaturu a další prameny

jsem řádně označila a uvedla v přiloženém seznamu.

V Jindřichově Hradci dne 19. července 2015

podpis

Zadání práce

Název diplomové práce:

Software pro vizualizaci dat na platformě Microsoft v porovnání s vybraným open source produktem

Abstrakt:

Diplomová práce se zabývá porovnáním Business Intelligence software pro vizualizaci dat na platformě Microsoft s vybraným open source produktem. Cílem práce je najít vizualizační software vhodný pro malou společnost, díky kterému by si firma mohla vizualizace vytvářet sama. Po úvodu do problematiky Business Intelligence jsou popsány vizualizační prostředky platformy Microsoft, konkrétně Excel, Power View, Power Map a Reporting. Následně je vybrán open source software, konkurenceschopný s vizualizačními produkty platformy Microsoft, konkrétně CDE Dashboard z platformy Pentaho. Prostřednictvím navržených scénářů a hodnocených kritérií jsou porovnány vybrané produkty. Závěr práce posuzuje vhodnost jednotlivých software pro vizualizaci dat.

Klíčová slova:

Microsoft, Pentaho, business intelligence, CDE Dashboard, vizualizace dat

Obsah

Úvod	13
1 Business intelligence	17
1.1 Zdrojové systémy	18
1.1.1 ERP (Enterprise Resource Planning)	18
1.1.2 CRM (Customer Relationship Management)	19
1.2 Komponenty datové transformace	19
1.2.1 ETL (Extract Transform Load)	19
Extract	19
Transform	19
Load	20
1.2.2 EAI (Enterprise Application Integration)	20
1.3 Databázové komponenty	21
1.3.1 DSA (Data Staging Areas), ODS (Operational Data Store)	21
1.3.2 Data Warehouse (DWH)	21
1.3.3 Data Mart (DMA)	23
1.4 Analytické komponenty	23
1.4.1 OLAP (On Line Analytical Processing)	23
1.4.2 Reporting	24
1.4.3 Dashboard	25
1.4.4 Executive Information Systems (EIS)	27
1.4.5 Data mining	27
1.5 Obecné technologie pro práci s daty	28
1.5.1 Nástroje pro zajištění datové kvality	28
1.5.2 Nástroje pro správu metadat	28
2 Vizualizace dat na platformě Microsoft	28
2.1 MS Excel	29
2.2 Power View	29
2.3 Power Map	31
2.4 Reporting	32
3 Open Source BI	33
3.1 Nabídka open source produktů	33
3.2 Výběr open source produktu - Pentaho	34
4 Metodická část	35
4.1 Hodnocená kritéria	36
4.2 Hodnocené scénáře	37
4.3 Metoda hodnocení	38
4.4 Předpoklady a omezení práce	40

4.4.1 Softwarové a hardwarové omezení.....	40
4.4.2 Omezení výzkumného vzorku.....	41
5 Hodnocení scénářů.....	42
5.1 Scénář 1 – řešení.....	42
5.1.1 Excel.....	42
5.1.2 Power View.....	43
5.1.3 Power Map.....	44
5.1.4 Reporty.....	46
5.1.5 CDE Dashboard.....	47
5.1.6 Řešení scénáře – hodnocení.....	48
5.2 Scénář 2 – řešení.....	50
5.2.1 Excel.....	51
5.2.2 Power View.....	52
5.2.3 Report	53
5.2.4 CDE Dashboard.....	53
5.2.5 Řešení scénáře 2 - hodnocení.....	55
5.3 Scénář 3 – řešení.....	56
5.3.1 Excel.....	57
5.3.2 Power View.....	57
5.3.3 Report	58
5.3.4 CDE Dashboard.....	59
5.3.5 Řešení scénáře 3 - hodnocení.....	60
5.4 Hodnocení vizualizací.....	62
Závěr	66
Seznam literatury.....	69
Přílohy	73

Seznam obrázků

Obrázek 1: Obecná koncepce architektury BI	18
Obrázek 2: ERP systém	19
Obrázek 3: Schéma ETL	20
Obrázek 4: EAI	20
Obrázek 5: Datový sklad - Kimball	22
Obrázek 6: Datový sklad - Inmon	22
Obrázek 7: OLAP	23
Obrázek 8: Ukázka reportu	25
Obrázek 9: Ukázka dashboardu	26
Obrázek 10: MS Excel - ukázka	29
Obrázek 11: Power View - ukázka	30
Obrázek 12: Power Map - ukázka	31
Obrázek 13: Report - ukázka	32
Obrázek 14: Saiku - ukázka	35
Obrázek 15: CDE Dashboard - ukázka	35
Obrázek 16: Excel - vizualizace scénář 1	43
Obrázek 17: Power View - vizualizace scénář 1	44
Obrázek 18: Power Map - vizualizace scénář 1	45
Obrázek 19: Report - vizualizace scénář 1	47
Obrázek 20: CDE Dashboard - vizualizace scénář 1	48
Obrázek 21: Excel - vizualizace scénář 2	51
Obrázek 22: Power View - vizualizace scénář 2	52
Obrázek 23: Report - vizualizace scénář 2	53
Obrázek 24: CDE Dashboard - vizualizace scénáře 2	54
Obrázek 25: Excel - vizualizace scénář 3	57
Obrázek 26: Power View - vizualizace scénář 3	58
Obrázek 27: Report - scénář 3	59
Obrázek 28: CDE Dashboard - vizualizace scénář 3	60

Seznam tabulek

Tabulka 1: Subjektivní váhy kritérií (body)	39
Tabulka 2: Hodnocení scénáře 1 (v bodech)	50
Tabulka 3: Upravené hodnocení scénáře 1 (v bodech)	50
Tabulka 4: Hodnocení scénáře 2 (v bodech)	56
Tabulka 5: Upravené hodnocení scénáře 2 (v bodech)	56
Tabulka 6: Hodnocení scénáře 3 (v bodech)	61
Tabulka 7: Upravené hodnocení scénáře 3 (v bodech)	62
Tabulka 8: Celkové hodnocení přístupů	63

Seznam rovnic

Rovnice 1: Ukazatel Cena	37
Rovnice 2: Ukazatel Množství	37
Rovnice 3: Ukazatel Obrat	38
Rovnice 4: Ukazatel Celkové náklady	38
Rovnice 5: Ukazatel Zisk	38
Rovnice 6: Váha hodnocení	39
Rovnice 7: Hodnocení scénáře	39
Rovnice 8: Celkové hodnocení scénářů	40

Úvod

Práce spadá do oblasti business intelligence, dále jen BI. Pod pojmem BI se skrývá nemálo věcí, které by měly firmám pomoci lépe se rozhodovat. BI je často brán jako zastřešující pojem, který zahrnuje různé aktivity a technologie, které tak či onak využijí údajů a informací pro zvýšení výkonnosti podniku a podpory rozhodování na všech organizačních úrovních (strategické, taktické a provozní úrovni) (Gudfinnsson et al. 2015).

S nárůstem objemu, rychlosti a odlišnosti podnikových dat, se stávají tradiční způsoby analýzy, sledování a pochopení informací neúčinné. S rostoucím využitím Business intelligence řešení napříč podniky, pro podporu rychlého rozhodování, hledají organizace řešení v tomto odvětví. Organizace potřebují jednotný pohled na klíčové obchodní informace a jejich dopad na koncového uživatele, proto je vizualizace podnikových metrik a procesů nutností.

„Na začátku 21. století se na lidstvo valí datové tsunami. Výzkumníci z analytické společnosti IDC spočítali, že celosvětový objem dat se každé dva roky zdvojnásobí. V roce 2011 odpovídal prostoru kolem dvou set miliard DVD (Černý 2012).“

Ve svém důsledku tedy BI umožňuje transformovat velké množství dat z různých zdrojů do podoby, kterou je možné lépe prezentovat a vizualizovat. Proč je tak důležité vizualizovat statistická i jiná data? W. S. Jevons v roce 1849 komentoval své diagramy, kde sledoval změny cen různých produktů v závislosti na některých komerčních jevech (například na objevení australského zlata): *„Jejich smyslem není ani odkaz ke konkrétním číslům, která lze lépe zjistit z odpovídajících tabulek, jako předvést očím obecné výsledky vyplývající z velkého množství čísel, jež nemohou být zachyceny jinak než graficky. Mé diagramy ukazují i ty nejmenší detaily tabulek, ale předčí i výpočty středních hodnot, protože oko či mysl samy zaznamenají obecný trend číselných souborů (Saxl a Ilucová 2004).“* Vizualizace a to zejména grafické umožňují uživateli uchopit velké objemy dat a převést je do podoby, která je srozumitelná.

Vzhledem k rozkvětu využití business intelligence a progresivních technologií pro analýzu, se role vizualizačního softwaru a nástrojů také rozšiřuje a rozvíjí. Důležitou roli zde hraje faktor dostupnosti, který zahrnuje podporu mobilních a vzdálených uživatelů. Přístup může probíhat na bázi prohlížeče, aby bylo možné kontrolovat vše od cloudových platforem pro správu databází a datových skladů, až k analytické platformě. Mobilita umožňuje člověku, který má na starosti rozhodování, samoobslužný přístup k důležitým obchodním informacím. Firmy chtějí být inovativní a nevšedně využívat obchodní informace z oblasti business intelligence v reálném čase, kdykoliv

a kdekoliv. Vizualizace BI fascinuje diváka, díky ní je možné dále zkoumat údaje a informace, aby bylo možné se rychle rozhodovat (Sen 2014).

Už jen pestrost možností jakým způsobem data a informace vizualizovat a prezentovat dává uživateli silný podklad pro vytvoření poutavé vizualizace, která bude přesně vystihovat požadavky pro rychlejší rozhodování. Různorodost vizualizačních možností lze prezentovat kupříkladu na periodické tabulce vizualizačních metod. Periodická tabulka obsahuje příklady možných vizualizací. Autoři Lengler a Eppler ji postavili na dvou dimenzích: období a skupině. Rozměr složitosti vizualizace je nazván „období“ a aplikační prostor „skupiny“. Po klasifikaci jsou tyto dvě dimenze uspořádány podobným způsobem. Jednotlivé sloupce uvádí podobné metody pro podobné účely, pouze jsou víc a víc složitější. Oblast aplikačního prostoru „skupiny“ se dále dělí do následujících kategorií, dle barevného pozadí¹:

- **Data Visualization** - Vizualizace dat obsahuje standardní kvantitativní formáty, například koláčové grafy, sloupcové grafy, spojnicové grafy. Jedná se o vizuální reprezentaci kvantitativních údajů schematicky (s osami nebo bez). Tyto grafy jsou univerzální, používají se především pro získání přehledu údajů.
- **Information Visualization** - Vizualizace informací, jako například sémantické sítě nebo stromové mapy. Použití interaktivní vizuální reprezentace dat k zesílení poznání. Data jsou transformována do obrazu.
- **Concept Visualization** - Konceptová vizualizace, jako myšlenkové mapy nebo Ganttův diagram. Tyto metody jsou vypracovávány, jako kvalitativní koncepty, nápady, plány a analýzy přes pomocná pravidla mapování. Pojetí vizualizace znalostí jsou obvykle prezentována v 2-D grafických vyjádřeních. Většinou se využívá sloupcových, či koláčových grafů.
- **Metaphor Visualization** - Metaforická vizualizace, jako je mapa metra nebo šablona příběhu. Účinnou a jednoduchou šablonou jsou zprostředkovány komplexní postřehy. Vizuální metafory plní dvojí funkci. Za prvé umisťují informace graficky, aby je organizovaly a strukturovaly. Za druhé zprostředkovávají přehled o zastoupení informace prostřednictvím základních znaků metafor.
- **Strategy Visualization** - Strategické vizualizace, jako strategie obrazu nebo technologického plánu, je definována jako systematické použití doplňkových vizuálních reprezentací, zlepšujících analýzu, vývoj, formulaci, komunikaci a implementaci strategie v organizacích.

.....
¹ Periodická tabulka vizualizačních metod tvoří přílohu 2 této práce. (Lengler a Eppler 2007)

- **Compound Visualization** - Kombinované vizualizace se skládají z několika výše uvedených formátů. Mohou to být složité znalostní mapy, které obsahují schématické a metaforické prvky nebo jiná schémata. Pod označení kombinované vizualizace jsou zahrnuty vizualizace, které nespádají do žádného z předchozích rozdělení a jsou tedy jejich kombinací (Lengler a Eppler 2007).

Ve své práci jsem se rozhodla zaměřit na porovnání software vhodných pro prezentaci dat, které nabízí platforma Microsoft s vybraným open source² BI produktem, který se nachází na trhu. Jelikož trh nabízí velké množství open source produktů a není v mých silách v rámci diplomové práce obsáhnout všechny, porovnávám proto pouze jeden vybraný. Dále v práci zohledňuji, na jaká data se daný prezentační software hodí. Mým cílem je zjistit, zda je v současné době na trhu nabízen open source produkt, který by byl do určité míry konkurenceschopným s BI řešením na platformě Microsoft. Rozšířím obzory v oblasti vizualizace dat nejen sobě, ale i čtenářům této práce. Mým cílem je vytvořit doporučení pro společnost, k identifikování konkrétního vhodného vizualizačního produktu pro určitá data. Zaměřím se na menší společnosti, které nemají takové finanční schopnosti, aby si mohli dovolit najmout firmu na zpracování dat v oblasti BI. Dále budu zkoumat, zda existuje produkt (spadající do mé oblasti zkoumání), který by byl natolik uživatelsky přívětivý, že by si dané vizualizace mohla vytvářet sama firma, například po nějakém kratším školení. Informace získané psaním této práce, v budoucnu využiji při výkonu svého povolání.

Z mé diplomové práce vyplynou doporučení pro použití vhodného open source produktu na prezentaci vybraných dat v porovnání s tím, co nabízí platforma Microsoft. Dále by mé zkoumání mělo vést ke zjištění/nalezení open source software, případně software z platformy Microsoft, který mohou využívat firmy samy bez toho, aby jim BI vizualizace musel vytvářet specialista. Zaměstnancům by například stačilo absolvovat pouze nějaká školení.

Podmínky, které má splňovat vybraný open source produkt a vizualizační software od společnosti Microsoft:

- Open source produkt, který poskytuje prezentační software a je konkurenceschopný prezentačním softwarům platformy Microsoft.
- Open source software a software platformy Microsoft, na kterém si zvládne vytvořit vizualizaci sama firma.

.....

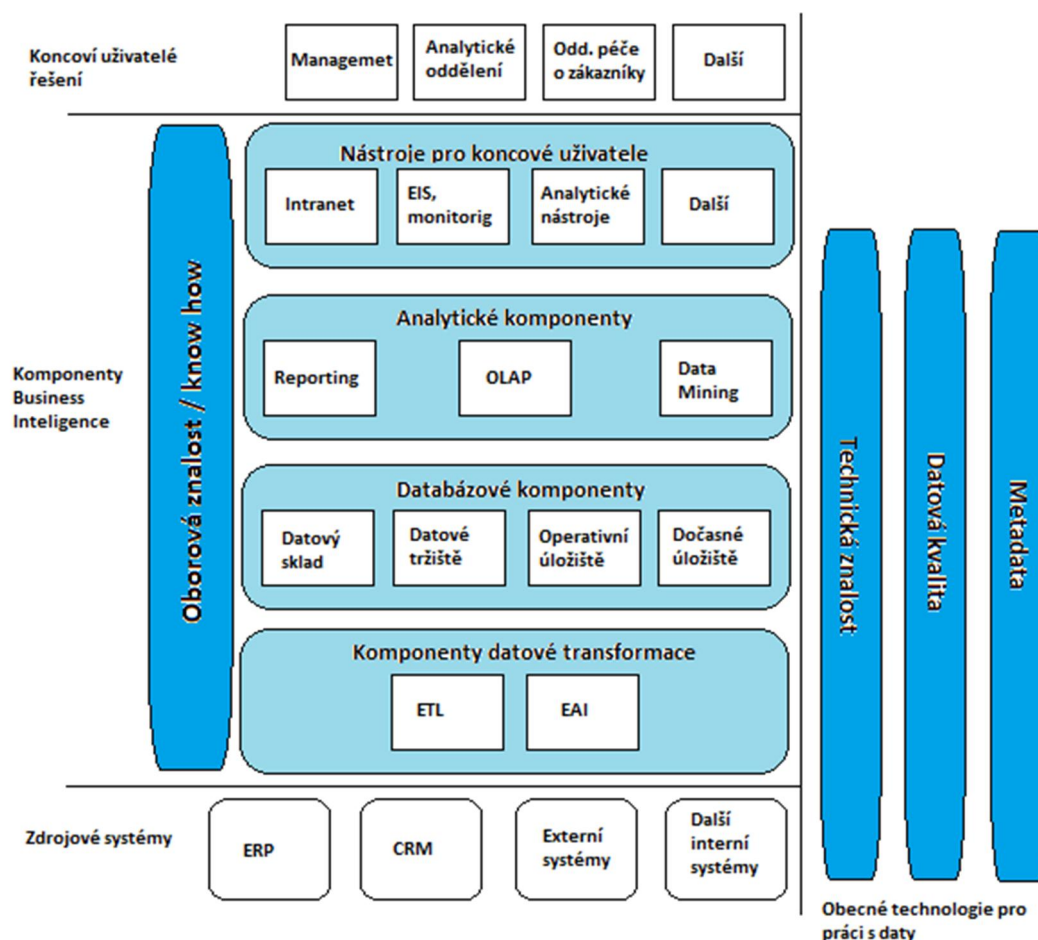
² Open source se do českého jazyka překládá jako otevřený software. Osobně preferuji výraz open source, dle mého názoru to více odpovídá skutečnému pojetí. Výraz bych přeložila spíše jako software s otevřenými (volně dostupnými) zdroji.

1 Business intelligence

BI je pojem užívaný již dlouho, přestože jeho význam nemá ustálenou přesnou definici. Business intelligence systém byl představen H. P. Luhnem v roce 1958, v časopise IBM, ve stejnojmenném článku. Luhnem definoval BI jako schopnost vnímat vzájemné vztahy uvedených skutečností tak, aby vedly činy k požadovanému cíli. Luhn psal o statistické analýze textu konkrétně o tom, co dnes nazýváme „meta-data“. Zaměřoval se výhradně na dokumenty jako zdroj informací. V šedesátých letech ještě nikdo nic nevěděl o počítačových sítích a webových prohlížečích. Dnešní pojetí BI bylo tedy nemyslitelné. Číselně zaměřené BI sahá až do 70. let minulého století, zatímco textově a znalostně zaměřené BI je mnohem mladšího ročníku. Avšak pojem, který vznikl v roce 1958 a zaměřuje se na získávání informací, řízení znalostí, komunikaci a spolupráci, funguje ještě dnes (Grimes 2008).

Jednu z prvních definic, která se podobá dnešnímu pojetí BI, pronesl Howard Dresner v roce 1989. Analytik společnosti Garner Group navrhl business intelligence jako zastřešující termín, který popisuje pojmy a metody s cílem zlepšit podnikatelské rozhodování o použití systémů podpory na základě zjištěných skutečností (Power 2007). Další definice pochází od Dr. Bradforta. Ten BI označuje jako interaktivní způsob zkoumání a analýzy struktur, specifické oblasti informací, které mají rozpoznat trendy nebo vzory, a tím odvozené postřehy a vyvození závěrů (Bradford 2010). Společnými znaky těchto definic je zpracování dostupných dat a informací tak, aby pomohly k lepšímu rozhodování. Pod pojmem zpracování dat je myšleno protažení dat BI systémem skládajícím se z několika komponent, viz obrázek 1. Přičemž není nutné, aby BI obsahovalo všechny komponenty.

Obrázek 1: Obecná koncepce architektury BI



Zdroj: (Novotný et al. 2005)

Poznámka: Vlastní zpracování

1.1 Zdrojové systémy

Zdrojový systém poskytuje základní zdrojová data pro další zpracování do BI úrovně. Může se jednat o jakýkoliv systém poskytující potřebná data, která se dále mohou zpracovávat na informace. Zdrojovým systémem může být ERP, CRM, externí a další interní systémy.

1.1.1 ERP (Enterprise Resource Planning)

ERP je systém, který slouží k plánování zdrojů souvisejících s činnostmi podniku. Udržuje přehled o různých odděleních organizace v centrální databázi. Na obrázku 2 jsou zachyceny některé oblasti, kterých se ERP týká. Tento nástroj udržuje informace o zaměstnancích, projektech, klientech (CRM), platbách, logistice, podpoře, mzdách, lidských zdrojích, atd.

Obrázek 2: ERP systém



Zdroj: (Admin 2012)

1.1.2 CRM (Customer Relationship Management)

CRM je orientované na zákazníka, pomáhá řídit marketingové a obchodní oddělení. CRM spravuje informace o stávajících i potenciálních zákaznících a napomáhá k budování lepších vztahů. Cílem je zachytit co nejvíce informací o zákaznících, aby bylo možné poskytovat lepší nabídky a porozumět jejich potřebám (Admin 2012).

1.2 Komponenty datové transformace

Prostřednictvím transformační vrstvy se data ze zdrojových systémů transformují do datových skladů. Tato vrstva obsahuje nástroje ETL a EAI.

1.2.1 ETL (Extract Transform Load)

Patří mezi nejvýznamnější komponenty celého BI, někdy bývá označováno pojmem datová pumpa. Na rozdíl od předchozích komponent, která data ukládají a spravují, ETL data přemísťuje.

Extract

Získávání zdrojových dat. Zdrojová data mohou pocházet z různých zdrojových souborů, jak zobrazuje obrázek 3. Zdrojem může být například XLS, CSV soubor nebo soubor z databáze (z SQL Serveru nebo ORACLU).

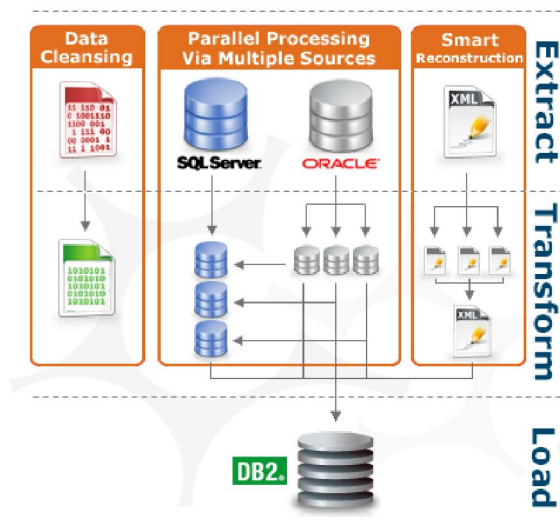
Transform

Transformace neboli přeměna dat. Obrázek 3 ukazuje některé případy transformací, například čištění dat, kombinování s jinými daty a jiné.

Load

Vytvoření databáze ze získaných dat, případně datového skladu (DWH).

Obrázek 3: Schéma ETL

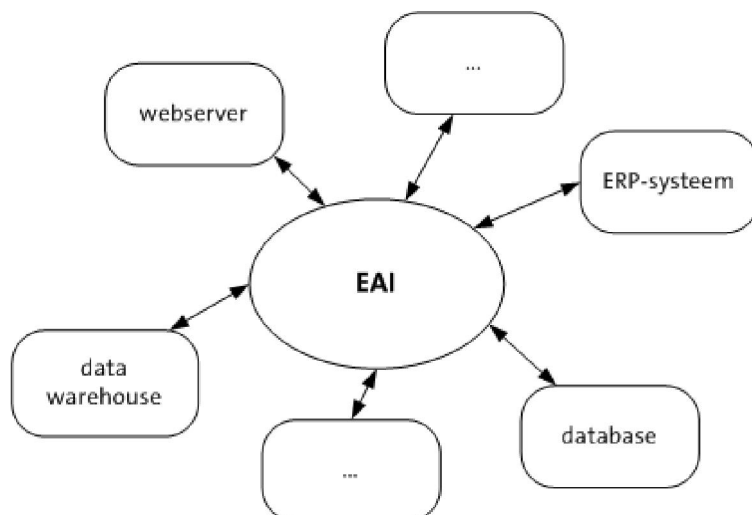


Zdroj: (ANON. 2015b)

1.2.2 EAI (Enterprise Application Integration)

Dříve byly navrhovány softwarové programy, pro každou oblast společnosti zvlášť, bez interakce. Nyní společnosti investují do EAI, do software, který jim umožní přenášet a sdílet informace napříč společnostmi. Tyto software platformy fungují v reálném čase.

Obrázek 4: EAI



Zdroj: (nl.wikipedia 2004)

1.3 Databázové komponenty

Databázové komponenty představují prvky určené pro práci s daty uloženými v databázi. Tuto část systému reprezentují datové sklady, datového tržiště, operativní úložiště a dočasné úložiště.

1.3.1 DSA (Data Staging Areas), ODS (Operational Data Store)

DSA představuje dočasné úložiště dat znázorňující meziprostor mezi zdroji dat a datovým skladem. Slouží ke shromažďování detailních dat, která neobsahují historii (jsou to pouze aktuální data). Nejdůležitějším úkolem dočasného úložiště dat, je urychlit proces extrakce dat. Nemusí být součástí BI řešení (Novotný et al. 2005).

Operativní úložiště dat (ODS) slouží k uchování aktuálních dat. Data mohou mít nízkou dobu odezvy. Nemusí být součástí BI řešení. Uplatňuje se stejně jako DSA u rozsáhlejších systémů (Novotný et al. 2005).

Rozdíl mezi DSA a ODS:

- DSA - dočasná úložiště dat, skladují se zde data, než dojde k jejich zpracování v datovém skladu (po zpracování jsou smazána).
- ODS – databáze, která obsahuje data pro analýzy či dotazy a je přístupná uživateli. Příkladem může být referenční databáze obsahující informace o produktech (Novotný et al. 2005).

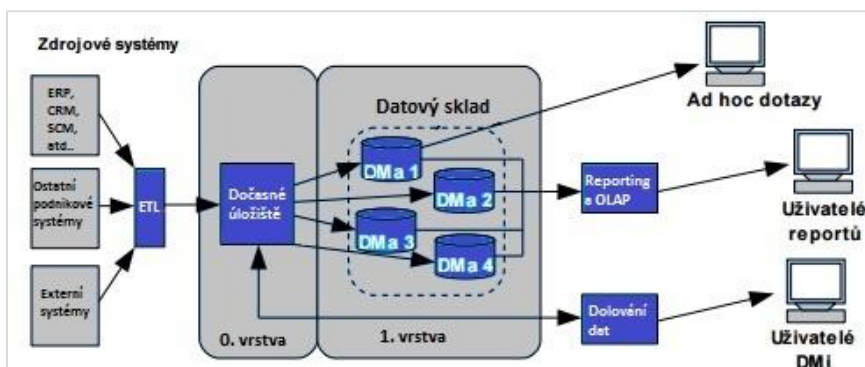
1.3.2 Data Warehouse (DWH)

Datový sklad je centrální úložiště dat, vytvořené pomocí integrace dat z jednoho nebo více různých zdrojů. Datové sklady obsahují současné i historické údaje a jsou používány pro vytváření sestav a pro sledování trendů, které mohou být zprávami pro osoby ve vedoucích funkcích, jako je například roční a čtvrtletní srovnání. Datové sklady mohou být projektovány několika způsoby a to v závislosti na nákladech na vývoj, potřebách, případně struktuře společnosti. Tato řešení jsou dělena podle dvou nejznámějších autorů:

- **R. Kimball** – samostatná datová tržiště,
- **W. H. Inmon** – konsolidovaný datový sklad.

R. Kimball znázorňuje datový sklad prostřednictvím spojení několika dílčích skladů (datových tržišť), takto vzniklá databáze je určena pouze pro jednu skupinu uživatelů. Výhodou je možnost rychlého řešení konkrétních požadavků (Novotný et al. 2005).

Obrázek 5: Datový sklad - Kimball



Zdroj: (Novotný et al. 2005)

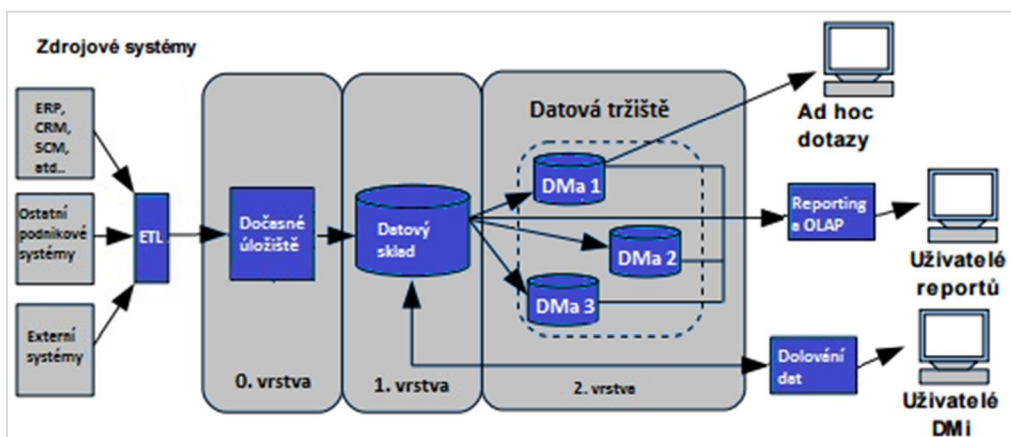
Poznámka: Vlastní zpracování

W. H. Inmon uvažuje jednotný datový sklad se závislými datovými tržišti, ke kterým mají přístup všechny skupiny. Používá třívrstvou architekturu, může být použito i dočasné datové úložiště. Tento přístup vyžaduje na počátku provedení důkladné analýzy potřeb jednotlivých oddělení (Novotný et al. 2005).

Definice datového skladu od Billa Inmona:

„Datový sklad je podnikově strukturovaný depozitář subjektivě orientovaných, integrovaných, časově proměnných, historických dat použitých pro získávání informací a podporu rozhodování. V datovém skladu jsou uložena atomická a sumární data (Lacko 2009).“

Obrázek 6: Datový sklad – Inmon



Zdroj: (Novotný et al. 2005)

Poznámka: Vlastní zpracování

1.3.3 Data Mart (DMA)

Datová tržiště jsou zmenšenou verzí datových skladů, která se zaměřují na potřeby konkrétních skupin uživatelů. Mohou se zaměřit na specifický cíl (Piccoli 2012).

1.4 Analytické komponenty

Mezi analytické komponenty patří například OLAP, Reporting, Dashboardy a doložení dat. Některé analytické komponenty zároveň slouží konečným uživatelům, například v rámci vizualizací.

1.4.1 OLAP (On Line Analytical Processing)

OLAP databáze slouží k analytickému zpracování dat v reálném čase. OLAP obsahuje data z datových skladů, případně datových tržišť a vytváří tak multidimenzionální kostku. Cílem je připravit data takovým způsobem, aby pomohla koncovým uživatelům při rozhodování. Takzvaná kostka se stává multidimenzionálním zobecněním dvou nebo trojrozměrné tabulky. Umožňuje například shrnout finanční údaje za vybrané město, prodejnu, prodejce, produkt, měsíc. Může porovnat, jaké množství určitého produktu prodal vybraný prodejce. Případně porovnávat plánované množství se skutečností.

Obrázek 7: OLAP



Zdroj: (SQIAR Ltd 2014)

Existuje několik klonů technologie OLAP, například:

- **MOLAP** (Multidimensional Online Analytical Processing – Multidimenzionální OLAP) – pro ukládání dat je používán multidimenzionální formát. Data uložená v MOLAP nejsou dostupná v reálném čase, jedná se o pouhé „přestěhování“ z datového skladu do nového uložení.

- **ROLAP** (Relational Online Analytical Processing – Relační OLAP) – k uložení dat je určena relační databáze (ANON. 2015a). U ROLAP dochází ke zpřístupnění dat v reálném čase, tak jak přibývají v datovém skladu.
- **HOLAP** (Hybrid Online Analytical Processing - Hybridní OLAP) – spojení MOLAP a ROLAP - Část dat je uložena jednou metodou a část druhou. (ANON. 2014b)
- **DOLAP** (Desktop Online Analytical Processing – Desktopová OLAP) – vznikla až na konci devadesátých let, umožňuje připojení k centrálnímu OLAP a následné stažení potřebné podmnožiny na lokální PC. Uživatel poté nemusí využívat připojení k serveru. Vhodné například pro mobilní aplikace (Novotný et al. 2005).

1.4.2 Reporting

Reporting je druh BI výstupu, který vzniká ze spolupráce mezi IT pracovníky a podniky. Report se zaměřuje na jednu oblast pozorování do většího detailu, s možností různých „prokliků“ hlouběji. Nástroje reportingu slouží k přípravě různých sestav, mezi nejznámější druhy reportů patří:

- **Standardní reporty** – jsou tvořeny předpřipravenými dotazy, které se v určitých časových periodách spouští.
- **Ad-hoc reporty** – generují se dle potřeb zákazníka. Jsou tvořeny specifickými dotazy, které explicitně vytváří uživatel.

Další možné členění reportů:

- **Statické** – reporty podobné tradičním papírovým reportům. Dá se v nich listovat a číst.
- **Interaktivní** – reporty, které lze přizpůsobovat pomocí různých ovládacích prvků.

Dělení dle oblasti a nasazení:

- **Enterprise**

Do oblasti Enterprise patří data v podnikové informatice.

Spadají sem reporty, týkající se jednotlivých oblastí podniku. Například pro obchodní nebo finanční oddělení, lidské zdroje a podobně. Data jsou obsažena v datových skladech, případně v podnikových databázích. Výhodou těchto reportů je, že data jsou již přetransformována v ETL etapě a přeplněna do DWH (datových skladů). Tyto reporty bývají nasazovány na úrovni podnikových sítí. Uživatelé k nim mohou přistupovat prostřednictvím intranetu, v rámci podniku.

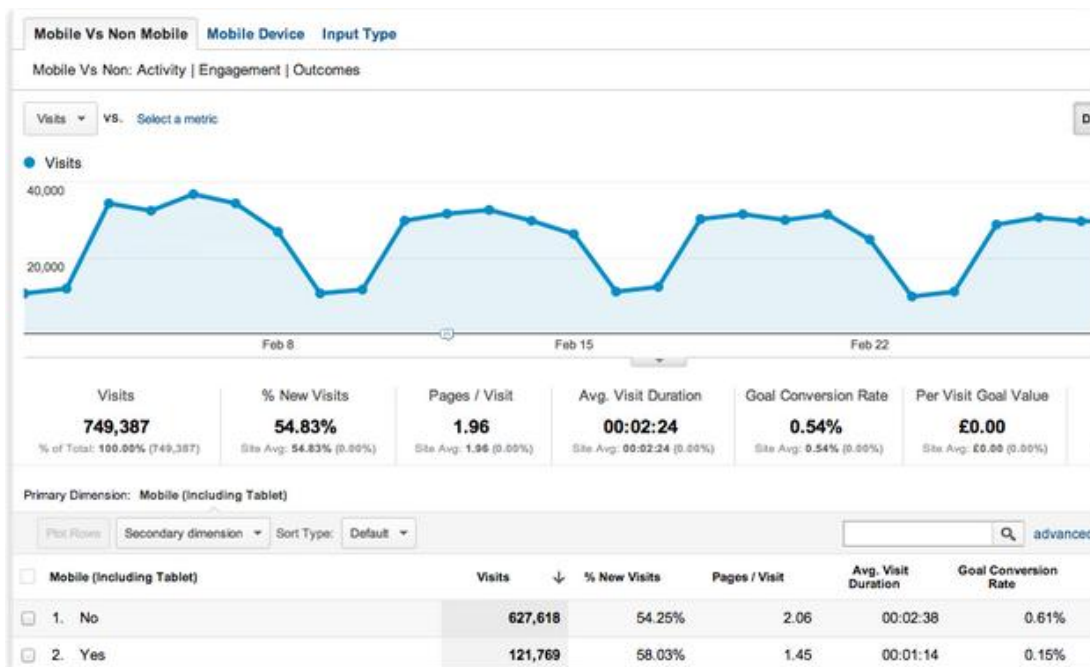
- **Embedded**

Embedded reporty jsou integrovány do aplikací.

- **B2B**

Reporty B2B jsou vytvářeny společnostmi pro jinou firmu. Data jsou zde poskytována externím subjektům. Při vytváření těchto reportů je mnohdy žádoucí skrýt informace, které pro budoucí rozhodnutí nebudou potřeba (Lacko 2009).

Obrázek 8: Ukázka reportu



Zdroj: (Charlton 2013)

1.4.3 Dashboard

Příkladem další vizualizace výstupů je Dashboard. Dashboard slouží primárně jako manažerská deska k rychlé orientaci, zda vše ve společnosti funguje, jak má. Ve většině případů jsou vizualizace používány jako přehled obsahující informace z více oblastí, není to však podmínkou. Tento druh vizualizace, stejně jako reporty, umožňuje zobrazování větších detailů. Dashboard zobrazuje sestavy a přehledy, které jsou předdefinované a obsahují aktuální informace o společnosti. Prostřednictvím dashboardů je možné pozorovat průběh výroby, prodeje a informace o zákaznících (sledování spokojenosti).

Dashboards tvoří jednotlivé sestavy, které jsou uspořádány dohromady například v SharePointu³. Dashboards je možné vytvářet například pomocí PerformancePoint

³ SharePoint je služba, která pomáhá vytvořit webové stránky, často je používána jako bezpečné místo k uložení, organizování a sdílení přístupu k informacím prostřednictvím webového prohlížeče (Microsoft 2015f)

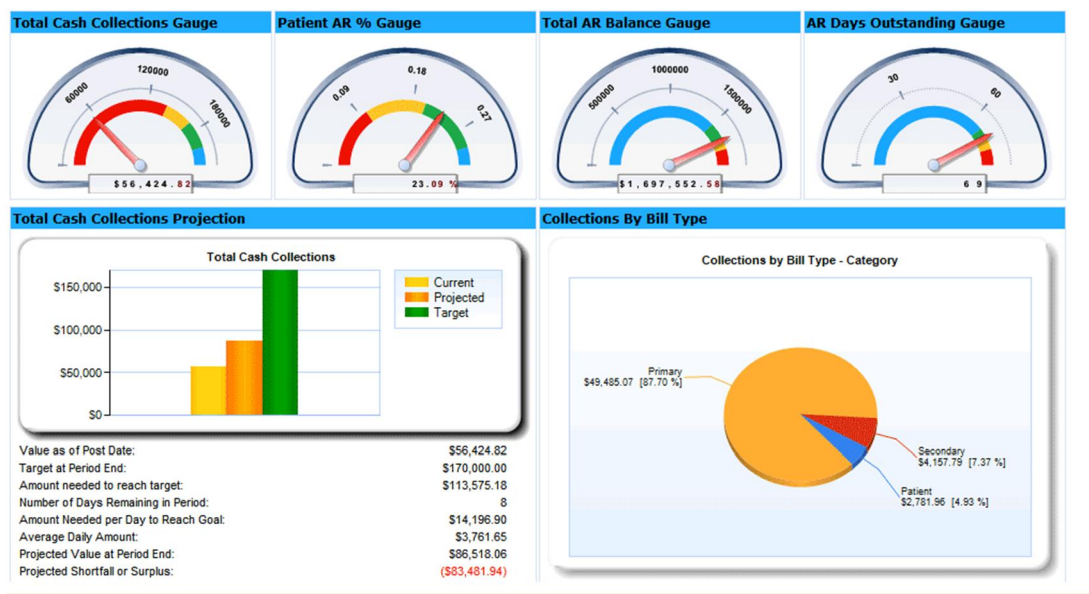
Dashboard Designer, případně pomocí SharePoint Serveru. Výběr nástroje, který použijeme pro tvorbu dashboardu závisí na několika faktorech: pro jak velkou skupinu je vizualizace vytvářena, co má být zobrazováno a další (Microsoft 2015b).

Pokud chceme vytvořit dashboard, který obsahuje jednu nebo více stránek zobrazující reporty opakovaně použitelných zpráv a informací. Tyto zprávy mají být použity v rámci velké skupiny nebo celé organizace. Dashboard může obsahovat robustní ukazatele, zprávy a stejně tak zprávy zobrazované reporty, které jsou umístěné na jiných serverech. V tomto případě je vhodné využít nástroj PerformancePoint Dashboarded Designer (Microsoft 2015b).

Pokud nám stačí stránka SharePointu, která obsahuje jednu nebo více zpráv a tyto zprávy mají být použity týmem nebo malou skupinou, pak je vhodné vytvořit stránku webu SharePoint. Web SharePointu je také vhodný pro jednoduchý ukazatel nebo zprávu, která informuje o výkonnosti a má být zobrazena na jedné stránce (Microsoft 2015b).

Výše zmíněné informace se týkají dashboardů platformy Microsoft. Dashboardy jsou vyvinuty i na jiných platformách, například open source platforma Pentaho obsahuje CDE⁴ Dashboardy. Tyto dashboardy lze také umístit na web a vytvářet vizualizace je možné přímo v Community Dashboard Editoru.

Obrázek 9: Ukázka dashboardu



Zdroj: (ANON. 2015c)

⁴ Community Dashboard Editor

1.4.4 Executive Information Systems (EIS)

Analytické aplikace, které si kladou za cíl podporovat manažerské procesy (např. analýzy, plánování, rozhodování). Tento analytický a prezentační nástroj se objevil již v osmdesátých letech.

1.4.5 Data mining

„Data mining je proces analýzy dat z různých perspektiv a jejich přeměna na užitečné informace. Z matematického a statistického hlediska jde o hledání korelací tedy vzájemných vztahů nebo vzorů v datech. Data mining je proces, jehož cílem je těžba informací v databázích. Využívá statistické metody a další metody hraničící s oblastí umělé inteligence (Lacko 2009).“

Dolování dat může obsahovat různé techniky. Mezi nejznámější techniky patří:

- **Market Basket Analysis** (Analýza nákupního košíku) – Forma clusteringu, sloužící k vyhledávání shluků (skupin a prvků), které se vyskytují pospolu.
- **Memory-Based Reasoning** (Dedukce) – Předpovídá neznámé skutečnosti za pomoci skutečností známých. Pozoruje blízké okolí známých údajů a sestavuje jejich hodnoty tak, aby bylo možné provést odhady hodnot, které je potřeba předpovědět.
- **Cluster Detection** (Detekce shluků) - Objevování skupin a struktur v datech, která si jsou „podobná“, bez použití známých struktur v datech.
- **Link Analysis** (Analýza závislostí) – Odlišuje se od výše zmíněných technik především tím, že se zaměřuje na vztahy mezi jednotlivými složkami, nikoli na jejich vlastnosti.
- **Decision Trees and Rule Induction** (Rozhodovací stromy a indukce) – Tyto modely jsou tvořeny výstupy statistických a nestatistických technik. Řadí se sem například: regresní a klasifikační stromy, chi-kvadrát, automatická indukce a další.
- **Artificial Neural Networks** (Neuronové sítě) - Neuronová síť patří mezi výpočetní modely používané v umělé inteligenci. Vzorem pro neuronové sítě byl lidský mozek. Principem této techniky je nastavení parametrů jednotlivým neuronům, v průběhu procesu učení se. Učení probíhá na tréninkovém vzorku dat. Výsledkem by měla být taková konfigurace, která bude co nejlépe vyhovovat následné predikci.
- **Genetic Algorithms** (Genetické algoritmy) – Genetické algoritmy, na rozdíl od všech výše zmíněných technik, neslouží k predikci hodnot, ale k vývoji. Jejich prostřednictvím je možné vyhledávat optimální množiny parametrů, které mohou sloužit k predikci zprostředkované jinou technikou (Novotný et al. 2005).

1.5 Obecné technologie pro práci s daty

Obecné technologie pro práci s daty zahrnují základní pojmy týkající se dat a celého procesu tvorby BI řešení. Patří sem nástroje pro zajištění datové kvality, nástroje pro správu metadat a technické znalosti.

1.5.1 Nástroje pro zajištění datové kvality

Datová kvalita je jedna z velmi důležitých vlastností, která určuje spolehlivost daných dat. Nástroje, které mají za cíl zpracovat data a zajistit jejich:

- **Úplnost** - identifikace nepoužitelných nebo chybějících dat a jejich následné ošetření,
- **soulad** – identifikace dat, která nemají standardní formát a jejich následné ošetření,
- **konzistence** – identifikace dat, které představují rozporné informace a následné ošetření,
- **přesnost** – identifikace dat, která jsou nepřesná, případně zastaralá a následné ošetření,
- **unikátnost** – identifikace duplicitních dat a následné ošetření,
- **integrita** – identifikace dat postrádajících vztahy vůči zbylým datům a následné ošetření (Novotný et al. 2005).

1.5.2 Nástroje pro správu metadat

Metadata jsou data, která popisují jiná data (vlastnosti nebo jiný obsah). Pro popis obsahu a principů fungování komponent IS/ICT⁵ řešení je nutná existence metadat.

2 Vizualizace dat na platformě Microsoft

Platforma Microsoft nabízí možnosti vizualizace jak v oblasti BI jako takové, tak i v oblasti Self BI. Tento druh vizualizace si je schopen vytvořit i člověk, který nemá s programováním zkušenosti.

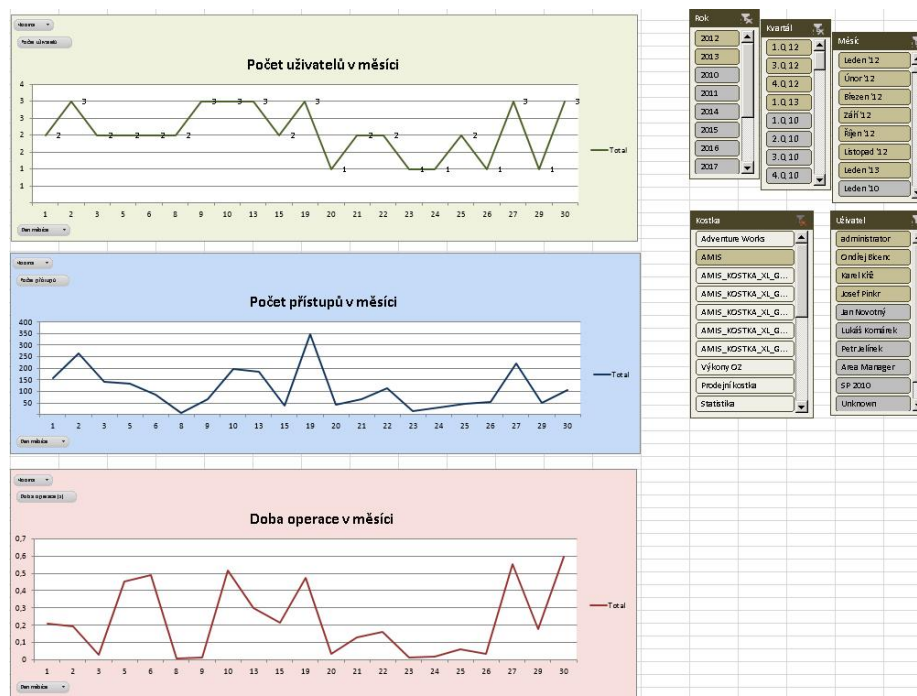
Microsoft nabízí Reporting services, Dashboardy, z oblasti Self services jsou to doplňky MS Excelu s názvy Power View a Power Map. Zvláštním případem je MS Excel sám.

.....
⁵ IS – informační systém, ICT – informační a komunikační technologie

2.1 MS Excel

MS Excel patří mezi nejznámější nástroje pro analýzu (samoobslužnou) a vizualizaci dat. Prostřednictvím Excelu lze pracovat s OLAP kostkami SQL Serveru (respektive daty uloženými v nich) a mnohými dalšími zdroji, například ze služby Analysis Services⁶, z Microsoft Power Query⁷ a dalšími. Excel umožňuje jednoduchý návrh grafických prvků a jejich změny. Pokud se rozhodneme například pro spojnicový graf a poté zjistíme, že by sloupcový byl lepší, můžeme jej bez větších problémů změnit. V tomto nástroji postačí několik kliknutí a máme jiný graf. Další výhodou, kterou poskytuje vizualizace v Excelu, je možnost filtrování v rámci jednotlivých listů (případně na úrovni sešitů). Obrázek 10 zobrazuje vizualizaci v Excelu.

Obrázek 10: MS Excel - ukázka



2.2 Power View

Power View je bezplatným doplňkem MS Excel, který slouží k vizualizaci a prezentaci dat. Konkrétně je funkcí MS Excel 2013 a Microsoft SharePoint Server 2010, 2013 a je také obsažen v Office 365 BI. Zdrojem dat pro Power View může být MS Excel, který

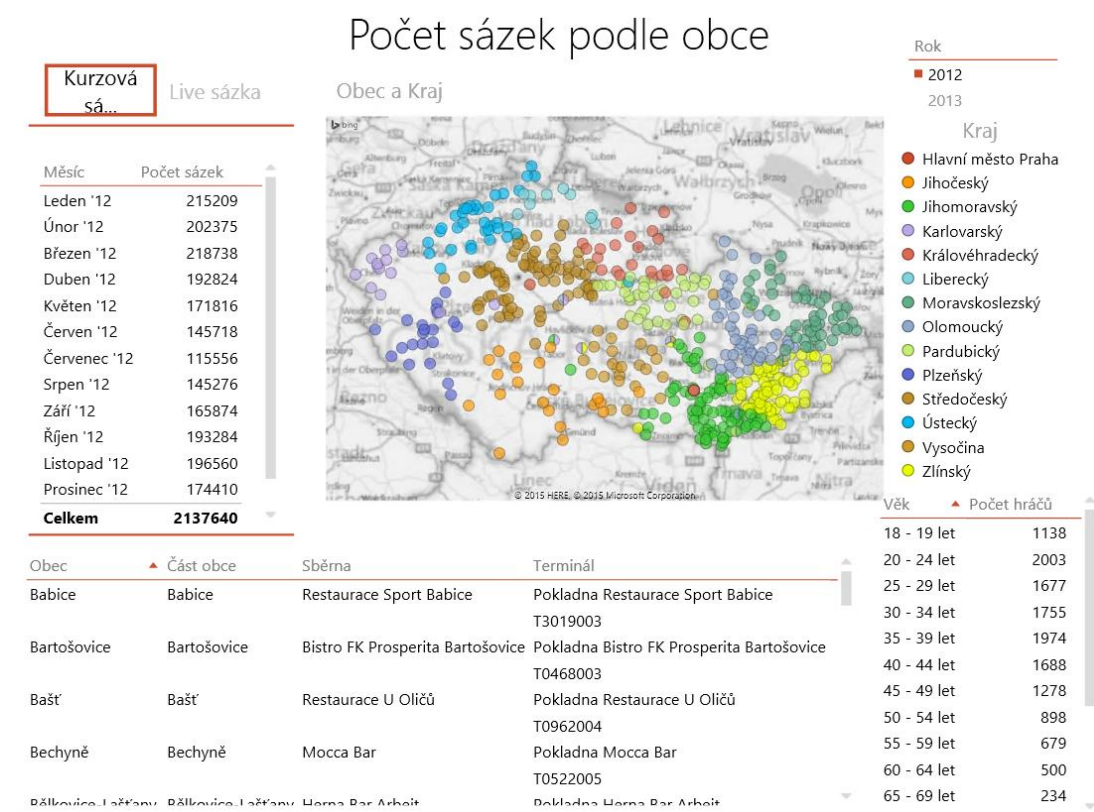
⁶ Analysis services je online analytický data engine používaný pro podporu rozhodování a BI řešení, poskytující analytické informace pro reporty a klientské aplikace, jako je například MS Excel, Reporting services a BI nástroje třetích stran.

⁷ Microsoft Power Query je doplněk aplikace MS Excel, prostřednictvím kterého lze načítat data z externích zdrojů. (Podporován Microsoft Office 2010 Profesionál Plus s programem SW Assurance, Microsoft Office 2013.)

obsahuje, jak zdrojová data, tak i list se samotným Power View (Microsoft 2015d). Dále může být Power View obsaženo v SharePointu⁸, zdrojem pro něj pak může být Power Pivot⁹, tabulkový model na serveru Analysis Services nebo multidimenzionální model (OLAP kostka) (Microsoft 2015e).

Nabízí vizualizace formou tabulek, sloupcových, spojnicových, výsečových a dalších sad grafů. Power View obsahuje i 2D grafy map, do kterých je možné zaznamenávat například údaje o množství. Umožňuje vytvořit bublinové grafy, které se mohou „pohybovat“ v čase. Dále poskytuje „dlaždice“, prostřednictvím kterých lze filtrovat celé listy (všechny grafy obsažené na jednotlivých listech). Při použití filtrů má uživatel možnost skrýt vybrané. V neposlední řadě lze filtrovat všechny objekty (grafy a tabulky) v rámci jednotlivých listů a to prostřednictvím kliknutí do vybraného objektu. Například pokud klikneme na sloupec označující rok 2011, upraví se podle toho všechny grafy a tabulky (Microsoft 2015d). Obrázek 11 prezentuje příklad vizualizace v Power Map.

Obrázek 11: Power View - ukázka



⁸ Více informací o specifikacích SharePointu viz zdroj (Microsoft 2015e)

⁹ Doplněk MS Excel, který pomáhá zpracovávat velké objemy dat z různých zdrojů, rychle je analyzovat a snadno sdílet poznatky (Microsoft 2015c)

2.3 Power Map

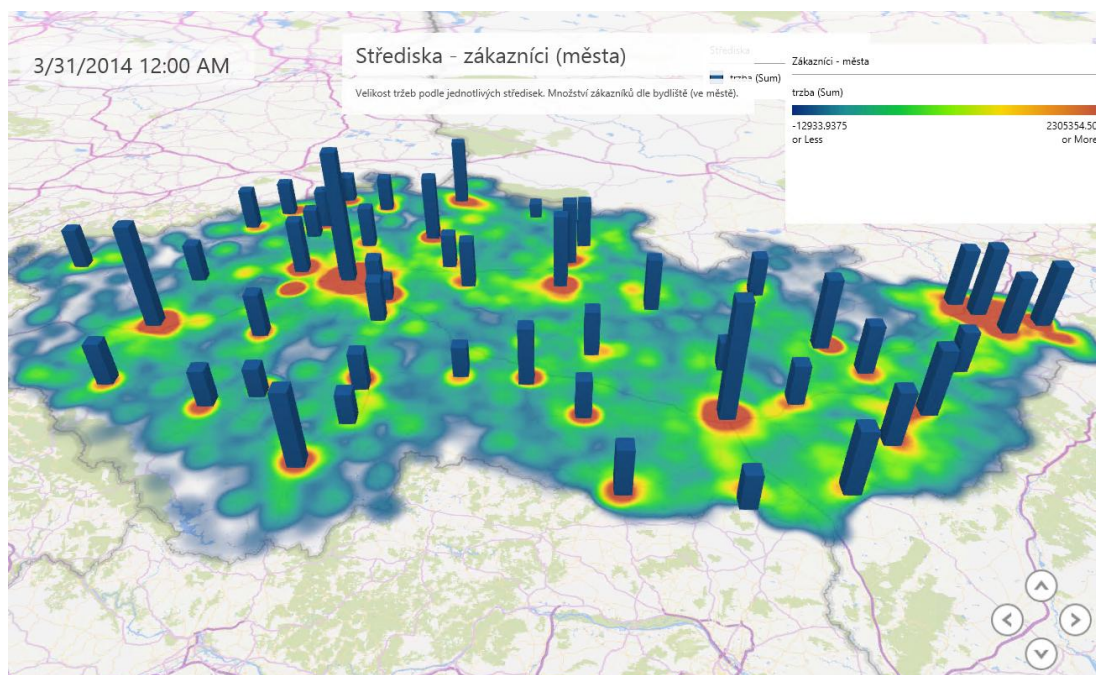
Power Map je bezplatným doplňkem MS Excel, který slouží k vizualizaci a prezentaci dat formou 3D map. Dále ho můžeme najít v Office 365 BI. Tento druh vizualizace je vhodný pro údaje zakreslené do mapy (geografická data). Konkrétně pro data sesbíraná v průběhu času, jelikož je zde možnost zobrazovat data v průběhu času (Microsoft 2015a).

Zdrojem dat pro Power Map může být MS Excel obsahující zdrojová data, případně Power Pivot. Power Map umí zpracovávat zeměpisné souřadnice a další geografické informace:

- ulice,
- města,
- kraje,
- státy/provincie,
- PSČ/poštovní kódy,
- země/regiony.

Celé vizualizace je pak možné exportovat do video formátu. Největší výhodou vizualizace Power Map je možnost animovaného zobrazení v čase. Pokud jsou k dispozici například kumulované hodnoty týkající se množství, zisku nebo nákladů, tak je možné pozorovat vývoj těchto hodnot jako animaci. Příklad vizualizace v Power Map ukazuje obrázek 12.

Obrázek 12: Power Map - ukázka



2.4 Reporting

Reporting umožňuje vytváření datových sestav s různými grafickými prvky (grafy, mapami a podobně). Reporty na platformě Microsoft mohou být vytvářeny prostřednictvím Report Designeru ve Visual Studiu¹⁰, případně Report Builderu. Zdrojem může být SQL databáze, OLAP kostka. Reporty je možné „publikovat“ prostřednictvím webového prostředí, tím je zprostředkován přístup odkudkoli.

Reporty nabízí vizualizace formou grafů (sloupcové, koláčové, bodové...), tabulek, různých „budíků“, obrázků, po stažení mapového podkladu i map a mnoho jiného. Do pohledu lze připojit i filtry (parametry). Umožňuje „proklik“ hlouběji v hierarchii. Například z kategorie zboží je možné se dostat na subkategorii.

V rámci této práce s reporty je k prezentacím využit Report Designer. Při vytváření reportů je potřeba vytvoření datasetů, což jsou datové zdroje obsahující v tomto případě SQL dotazy, jejichž prostřednictvím jsou získávána požadovaná data.). Oproti ostatním vizualizacím v tomto spatřuji nevýhodu reportů.

Další možností vizualizace jsou Dashboardy (viz. Kapitola 1.4.3).

Obrázek 13: Report - ukázka

Stav sázení po kanálech											
Datum od - do: 1.2.2014 - 25.2.2014											
Kanál: Administration,BT,CallCentrum,Internet,IVR,IVT,Mobile Application,Mobile Web,SMS,TNT											
	Administration	BT	CallCentrum	Internet	IVR	IVT	Mobile Application	Mobile Web	SMS	TNT	Celkem
Vsazeno reg. - počet		288	239	38403	350		596	4747		12953	57576
Vsazeno nereg. - počet		2780								13011	15791
Vsazeno celkem - počet		3068	239	38403	350		596	4747		25964	73367
Vsazeno reg.		52 609	21 831	3 565 990	25 151		84 613	626 098		1 127 385	5 503 677
Vsazeno nereg.		296 989								1 115 289	1 412 278
Vsazeno celkem		349 598	21 831	3 565 990	25 151		84 613	626 098		2 242 674	6 915 955
Rozhodnuto reg. - počet		30	21	7 997	56		182	1 541		2 250	12 077
Rozhodnuto nereg. - počet		283								1 430	1 713
Rozhodnuto celkem - počet		313	21	7 997	56		182	1 541		3 680	13 790
Rozhodnuto reg.		33 414	14 146	2 596 461	10 606		53 494	559 905		772 949	4 040 975
Rozhodnuto nereg.		167 546								799 265	966 811
Rozhodnuto celkem		200 960	14 146	2 596 461	10 606		53 494	559 905		1 572 214	5 007 786
Rozdíl reg.		19 195	7 685	969 529	14 545		31 119	66 193		354 436	1 462 702
Rozdíl nereg.		129 443								316 024	445 467
Rozdíl celkem		148 638	7 685	969 529	14 545		31 119	66 193		670 460	1 908 169
Vyplaceno reg. - počet		30	21	7997	56		182	1541		2250	12077
Vyplaceno nereg. - počet		318								1465	1783
Vyplaceno celkem - počet		348	21	7997	56		182	1541		3715	13860
Vyplaceno reg.		33 414	14 146	2 596 461	10 606		53 494	559 905		772 949	4 040 975
Vyplaceno nereg.		222 151								949 528	1 171 679
Vyplaceno celkem		255 565	14 146	2 596 461	10 606		53 494	559 905		1 722 477	5 212 654
Výtěžnost reg.		36,49 %	35,20 %	27,19 %	57,83 %		36,78 %	10,57 %		31,44 %	26,58 %
Výtěžnost nereg.		43,59 %								28,34 %	31,54 %
Výtěžnost celkem		42,52 %	35,20 %	27,19 %	57,83 %		36,78 %	10,57 %		29,90 %	27,59 %

¹⁰ Visual Studio je software umožňující programování například v jazycích C#, C++, .NET, dále je možné využít jeho nástroje z oblasti BI (analytické, integrační a reportovací služby) a mnohé další. V tomto případě bude použito reportovacích služeb.

3 Open Source BI

Open source se podobá tradičnímu BI, ale na vývoj jsou používány volně dostupné zdrojové kódy. Open source BI pokrývá stejně jako komerční verze celou oblast BI, od datových skladů, přes získávání zdrojových dat, až po vizualizaci (Wise 2012).

Jednotlivé open source produkty nemusí pokrývat celá řešení v oblasti BI, zaměřují se převážně pouze na některou jeho část. Open source software je poskytován zdarma (v mnohých případech však pouze některé jeho části).

3.1 Nabídka open source produktů

O oblastech open source produktů bylo napsáno několik článků. Z oblasti produktů open source BI, které jsou zdarma, byl vytvořen například seznam 16-ti produktů. V seznamu jsou obsaženy produkty obsahující kompletní řešení v oblasti BI i produkty, které jsou tvořeny jen některou z jeho částí. Seznam obsahuje:

- Pentaho Community edice – obsahuje kompletní BI řešení (datový sklad je nutné řešit samostatně),
- OpenText Actuate Information Hub, Free edice (BIRT) - vizualizace,
- Report Server – webová aplikace pro vizualizaci,
- JasperReports BI – pro vizualizace a možnost použití add-on funkce k analýze dat na bázi OLAP,
- Jedox Base – poskytuje OLAP server, Jedox Excel Add-in (vizualizace),
- SpagoBI – analytický model (vizualizace), model chování (práva a role koncových uživatelů),
- ART – Reporting Tool - vizualizace,
- Pentaho Reporting – vizualizace,
- JMagallanes – uživatelská aplikace po OLAP (vizualizace na základě statických posudků a měnící se kontingenční tabulky),
- OpenReports - vizualizace,
- Seal Report – vizualizace pro rozhraní Microsoft .NET Framework (psán v C#),
- Openl – webová aplikace pro vizualizace,
- NextReports – vizualizace na libovolném operačním systému, kde je spuštěna java,
- Rapidminer – data mining a vizualizace dat,
- Mondrian – OLAP umožňující dotazování pomocí jazyka MDX,
- KNIME – obsahuje analytickou část, reporting a integrační platformu (ANON. 2014a).

3.2 Výběr open source produktu - Pentaho

Žebříček top 3 open source produktů obsahuje reportovací nástroje BIRT, JasperReport a Pentaho. Pentaho obsahuje na rozdíl od BIRTU a JasperReportu kompletní BI řešení (kromě datového skladu, který je nutné řešit samostatně). Pentaho Reporting má bohatou sadu funkcí připravených pro použití v podnikových organizacích. Pentaho funguje na bázi webového rozhraní a je vyvíjen ve dvou edicích – Enterprise a Community. Enterprise má rozšířené možnosti oproti Community verzi, ale je zpoplatněna. K tomuto produktu existuje rozsáhlá dokumentace, která napomáhá k lepšímu využívání celého produktu. Pentaho ke svému fungování používá platformu Java, konkrétně Java Enterprise Edition a může být použito na operačních systémech Windows, Linux a Mac (všude, kde je možné spustit Javu) (Muilwijk 2014). Podrobné informace jsou obsaženy na oficiálních stránkách Pentaho.¹¹

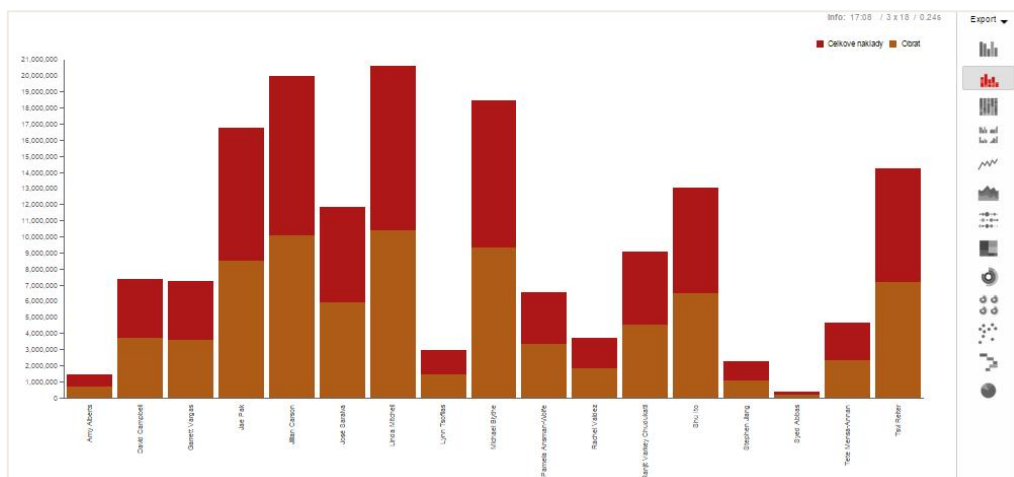
Pentaho jsem vybrala, kvůli nabídce nejucelenějšího BI produktu zdarma, univerzálního použití bez závislosti na druhu operačního systému a dostupnosti na webovém prostředí.

Platforma Pentaho nabízí vizualizace v rámci Community edice v Saiku a v CDE Dashboardech. Saiku¹² poskytuje uživateli možnost připojení se do OLAP kostky, případně do databáze, odkud si nahraje data. Získaná data může uživatel prostřednictvím Saiku analyzovat formou tabulek a přednastavených grafů. Pro prostředí Saiku je uživatelsky přívětivé. Nevýhodou Saiku je, že umožňuje vytvoření pouze jednoho objektu. V jednom Saiku můžeme vizualizovat například jen jeden graf. Obrázek 14 ukazuje možné zobrazení dat v Saiku. CDE Dashboard je editor umožňující vytváření dashboardů. Má do jisté míry uživatelsky přívětivé prostředí, ale bohužel ne všechny jeho komponenty jsou vždy funkční. Nabízí například možnost převést vizualizaci vytvořenou v Saiku do prostředí CDE, ale tato možnost nefunguje. Obrázek 15 představuje možnou vizualizaci v CDE dashboardu. Praktická vizualizace v rámci produktu Pentaho bude zpracovávána v CDE Dashboard editoru.

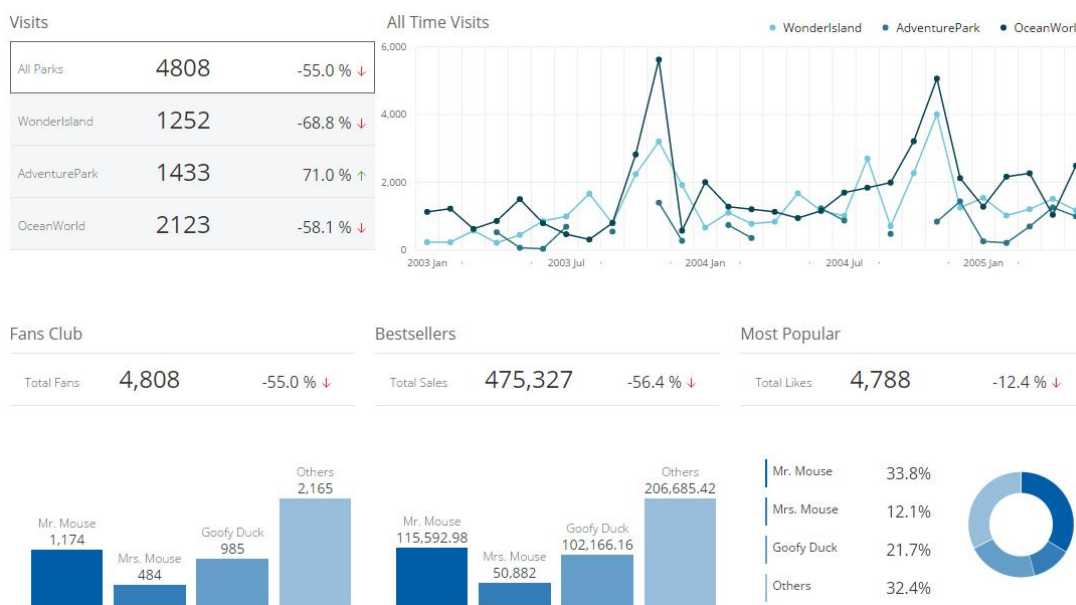
.....
¹¹ <http://community.pentaho.com/>

¹² Saiku existuje i jako samostatný open source software, bez propojení s Pentaho. V této práci je Saiku bráno jako doplněk v rámci Pentaho platformy.

Obrázek 14: Saiku – ukázka



Obrázek 15: CDE Dashboard - ukázka



Zdroj: (Pentaho 2015)

4 Metodická část

Metodická část obsahuje několik podkapitol, které se zabývají postupem práce při komparaci software vhodných pro vizualizaci dat na platformě Microsoft se software pro vizualizaci dat na vybraném open source produktu. Ve druhé kapitole této práce jsou vyjmenovány možnosti vizualizačních software, které poskytuje platforma Microsoft, viz kapitola 2. V kapitole 3.2 je vybrán open source vizualizační produkt, konkurenceschopný produktům platformy Microsoft. V rámci hodnocených kritérií jsou nejprve definovány pojmy, použitelnost a uživatelská zkušenost, které jsou důležité pro výběr produktů vhodných k vizualizaci dat. Následuje stanovení

hodnocených kritérií, která napomáhají spolu s hodnocenými scénáři k výběru BI vizualizačních softwarů. Hodnocená kritéria jsou vybírána s přihlédnutím k požadavkům na vhodné vizualizační prostředky z oblasti BI. Hodnocené scénáře znázorňují konkrétní příklady možných požadavků uživatelů, které jsou vizualizovány tak, aby napomohly k lepšímu rozhodování.

4.1 Hodnocená kritéria

Před uvedením základních hodnocených kritérií, která napomáhají při určování vhodnosti vybraného software, je důležité definovat použitelnost a uživatelskou zkušenost.

Základní rozdíl mezi použitelností a uživatelskou zkušeností:

- **Použitelnost** vyjadřuje do jaké míry systém, služba nebo produkt může být použit určenými uživateli za účelem dosažení konkrétních cílů s účinností, efektivností a spokojeností v zadaném kontextu použití.
- **Uživatelská zkušenost** zahrnuje subjektivní vnímání uživatele a reakce, které vyplývají z použití a/nebo očekávání používání výrobku, systému nebo služby (Jooste et al. 2013).

Kritéria hodnocení použitelnosti BI aplikací:

1. **Přehledné a dobře strukturované informace** - Informace by měly být na pracovní ploše zobrazeny přehledně a strukturovaně, aby se v nich uživatel snadno zorientoval. Pokyny a volby navigace by měly být vhodně umístěny.
2. **Vizualizace, jako podpora rozhodování** - Vhodně zvolené zobrazení dat s ohledem k vizualizovaným datům (tabulky, grafy, mapy...).
3. **Volby navigace** – Viditelně zobrazené ovládací prvky by měly uživatele vést k intuitivnímu ovládání. Uživatel by měl rozumět ovládání aplikace.
4. **Pocit kontroly nad aplikací** - Uživatel by měl cítit kontrolu nad aplikací. Měl by umět využívat vizualizované informace k podpoře rozhodování. Umět předkládané informace správně interpretovat a interpretace vytvářet, dle vlastního uvážení.
5. **Podpora sdílení** – Aplikace by měla být přizpůsobitelná individuálnímu použití nebo použití v rámci spolupracovníků v týmu zabezpečeným způsobem (Jooste et al. 2013).

4.2 Hodnocené scénáře

Hodnocené scénáře představují modelové příklady požadavků uživatelů, které mají být vizualizovány v rámci BI, aby napomohly k lepšímu rozhodování. Jsou zohledněny různé aspekty oblastí možných vizualizací. Například zobrazení v rámci mapových podkladů. Hodnocené scénáře jsou zaměřeny na oblast prodeje/nákupu produktů. Zaměření na oblast prodeje je z důvodu volně dostupné databáze, použité v této práci (bližší specifikace v kapitole 4.4.2).

První scénář se zaměřuje na oblast nákupu zboží zákazníky. Je důležité mít přehled o tom, které zboží, zákazníci nakupují. Stejně důležité je mít přehled o tom, v jakých oblastech zákazníci nakupují. Vizualizace by měla ukázat, ve kterých lokalitách v rámci jednotlivých států se nakupují jaké subkategorie produktů a v jakém množství. Pro přehledné zobrazení jsou vytvořeny ukazatele Cena a Množství, jež přehledně znázorňují, za kolik dolarů zákazníci dané produkty (subkategorie produktů) nakoupili a jaké množství produktů pořídili. Ukazatel Cena tvoří suma součinů jednotkových cen a prodaného množství, viz rovnice 1.

Rovnice 1: Ukazatel Cena

$$Cena = \sum (jednotková\ cena \times množství)$$

Ukazatel Množství je tvořen součtem počtu kusů vybraných produktů, spadajících do dané subkategorie, viz rovnice 2.

Rovnice 2: Ukazatel Množství

$$Množství = \sum (množství)$$

Celou vizualizaci by mělo být možné filtrovat v rámci produktových kategorií, jednotlivých let a zemí, ve kterých zákazníci nakupují.

Druhý scénář se zaměřuje na zaměstnance. Pro každou společnost je důležité sledovat produktivitu zaměstnanců. V tomto případě se vizualizace orientuje na množství peněz, které přinesli zaměstnanci společnosti prostřednictvím prodeje produktů (bez zohlednění nákladů). Ukazatel Obrat představuje sumu součinu množství a jednotkové ceny zboží, viz rovnice 3. Dalším ukazatelem, použitým v této vizualizaci, je ukazatel Množství, viz rovnice 2.

Rovnice 3: Ukazatel Obrat¹³

$$Obrat = \sum (jednotková\ cena \times množství)$$

Vizualizace by měla ukázat zaměstnance prodávající větší množství produktů a také ty, co jsou v prodeji horší. Celou vizualizace by mělo být možné filtrovat v rámci let, produktových kategorií a může být provedeno i srovnání v rámci pohlaví¹⁴ (zda jsou lepšími prodejci muži nebo ženy). Pokud to prostřednictvím daného vizualizačního software bude možné, budou pracovníci barevně odlišeni v závislosti na velikosti získaného obratu (případně průměrného obratu).

Třetí scénář se zaměřuje na produkty. Při vizualizaci bude použita produktová hierarchie (Produktová kategorie – Produktová subkategorie – Produkt). Produkty by měli být zobrazovány v souvislosti s ukazateli Obrat (viz rovnice 3), Celkové náklady, Množství (viz rovnice 2), a Zisk. Ukazatel Celkové náklady tvoří suma nákladů na konkrétní produkty, viz rovnice 4.

Rovnice 4: Ukazatel Celkové náklady

$$Celkové\ náklady = \sum (náklady)$$

Ukazatel Zisk tvoří suma rozdílu obrátů a celkových nákladů (viz rovnice 5).

Rovnice 5: Ukazatel Zisk

$$Zisk = \sum (Obrat - Celkové\ náklady)$$

Prostřednictvím těchto ukazatelů by mělo být znázorněno, které produkty se prodávaly a jaký zisk¹⁵ jejich prodej přinesl společnosti. Vizualizaci lze filtrovat podle jednotlivých let a zemí, ve kterých jsou produkty prodávány. Dále by zde měl být znázorněn zisk z produktů v průběhu roku (v rámci jednotlivých měsíců). Celá vizualizace má poukázat na produkty, které se prodávají nejvíce a v jaké části roku.

4.3 Metoda hodnocení

Metoda hodnocení vhodných vizualizačních prostředků, které jsou reprezentovány hodnocenými kritérii, je postavena na bodovém hodnocení. Je zde použita škála od 1 do 10 bodů, kdy nejméně je 1 bod a nejvyšší bodová hodnota představuje 10.

¹³ Pro ukazatel Obrat je použit stejný výpočet, jako pro ukazatel Cena. Rozdíl je pouze v názvu, z důvodu vhodnosti slov. Zákazníci nakupují za určitou cenu, nikoli za obrat

¹⁴ Filtr pohlaví bude uveden pouze v některých vizualizacích, nejedná se pouze o zajímavost, nikoli o kritérium hodnocení.

¹⁵ Zisk je v této práci stanoven rovnicí 5.

Vzhledem k 5-ti kritériím hodnocení, která jsou obsažena v kapitole 4.1, musí být stanoveny váhy jednotlivých kritérií. Prostřednictvím váhy kritérií neboli koeficientů významnosti jsou číselně vyjádřeny důležitosti jednotlivých kritérií. Čím je váha vyšší, tím vyšší je významnost daného kritéria. Aby bylo dosaženo srovnatelnosti vah kritérií, musí být jejich váhy stanoveny prostřednictvím vhodné metody. V tomto případě je použita metoda párového srovnávání.

V metodě párového srovnávání se pro každou dvojici kritérií stanoví preference, to se opakuje pro všechny dvojice. Porovnává se vždy kritérium v řádku s kritériem ve sloupci, pokud je preferováno kritérium v řádku, zapíšeme do pole 1, pokud preferujeme kritérium ve sloupci, zapíšeme do daného pole 0. Pro každé kritérium v řádku se součtem 1 stanoví celkové váhy kritérií (Fotr et al. 2010). Stanovené subjektivní váhy kritérií uvádí tabulka 1.

Tabulka 1: Subjektivní váhy kritérií (body)

Hodnocené kritérium	K ₁	K ₂	K ₃	K ₄	K ₅	Celkové váhy kritérií
K ₁	X	0	1	1	1	3
K ₂	1	X	1	1	1	4
K ₃	0	0	X	1	0	1
K ₄	0	0	0	X	1	1
K ₅	0	0	1	0	X	1

Váha hodnocení jednotlivých kritérií se vypočítá jako podíl celkové váhy kritérií a sumy celkové váhy kritérií, viz rovnice 6. Hodnocení scénáře je pak určeno vynásobením dosaženého hodnocení a váhy hodnocení kritéria, viz rovnice 7.

Rovnice 6: Váha hodnocení

$$\text{Váha hodnocení kritéria} = \frac{\text{celkové váhy kritérií}}{\sum(\text{celkové váhy kritérií})}$$

Zdroj: (Fotr et al. 2010)

Rovnice 7: Hodnocení scénáře

$$\text{Hodnocení scénáře} = \text{dosažené hodnocení} \times \text{váha hodnocení kritéria}$$

Hodnocení nejvhodnějšího softwaru pro vizualizace dat (software vhodný pro vizualizaci v rámci malé společnosti) je provedeno za pomoci několika scénářů

uvedených v kapitole 3.2 a subjektivních vah kritérií. Aby bylo dosaženo objektivních výsledků, probíhá hodnocení kritérií za použití několika scénářů (opakovaně). Některé scénáře povedou k doporučení vizualizačního software pro určitý druh vizualizace například pro vizualizace v mapách.

Hodnocená kritéria spolu se scénáři vedou k výběru nejvhodnějšího vizualizačního software, při zohlednění různých vizualizací. Pro zjištění nejvhodnějšího vizualizačního produktu je použit výpočet průměru, konkrétně tvořen sumou hodnocení scénářů podělenou počtem scénářů, viz rovnice 8. Prostřednictvím tohoto výpočtu bude zvolen nejvhodnější software.

Rovnice 8: Celkové hodnocení scénářů

$$\text{Celkové hodnocení scénářů} = \frac{\sum(\text{hodnocení scénáře})}{\text{počet scénářů}}$$

4.4 Předpoklady a omezení práce

Předpokladem celé práce je vizualizace na platformě Microsoft a open source produktech, které jsou v oblasti vizualizací konkurenceschopné platformě Microsoft. Vzhledem k rozsahu diplomové práce nelze brát v úvahu všechny společnosti pracující v oblasti BI, ani obsáhnout všechny open source produkty nabízející produkty určené k vizualizacím v oblasti BI. Vzhledem k záměru budoucího využití získaných zkušeností při výkonu povolání jsem zvolila platformu Microsoft. Z platformy Microsoft jsou použity nejnovější vizualizační prostředky, konkrétně - MS Excel 2013 (ve verzi Professional Plus), Power View, Power Map z téhož roku a Reporty budou vytvářeny v SQL Server Data Tools for Visual Studio 2013. Z oblasti open source je vybrán vizualizační produkt v kapitole 3.2, konkrétně se jedná o produkt CDE Dashboard z platformy Pentaho, který je ve výše zmíněné kapitole popsán.

4.4.1 Softwarové a hardwarové omezení

Jedním z mých cílů v rámci této práce je najít vhodný produkt z oblasti vizualizace BI, který by mohla využívat menší společnost (například podnikatel), proto bude z hlediska hardware využít osobní počítač. Bližší specifikace počítače – operační systém Windows 8.1 (64 bit), procesor Intel Core i5 se dvěma jádry, rychlostí 1,7 GHZ a konfigurace 8 GB operační paměti (RAM).

Omezení z oblasti software – z platformy Microsoft je využit MS Excel 2013 z kancelářského balíku Microsoft Office Professional Plus, Power View a Power Map jsou získány, jako doplňky pro tuto verzi MS Excel. Dále je využit SQL Server Data Tools for Visual Studio 2013 a Microsoft SQL Server 2014 Enterprise edice.

Z oblasti open source je vybrán a blíže specifikován vizualizační produkt v kapitole 3.2. Konkrétně se jedná o produkt CDE Dashboard z platformy Pentaho, na němž bude vizualizace prováděna v prostředí webového prohlížeče Google Chrome¹⁶. Dále je potřeba mít nainstalovanou Java Enterprise Edition a soubory potřebné ke spuštění Pentaho na webovém prostředí.

Primárním zdrojem pro všechny vizualizační produkty je databáze databázi blíže specifikovaná v kapitole 4.4.2, která běží na databázovém enginu SQL serveru 2014, prostřednictvím kterého k ní budu přistupovat. Software Power View a Power Map čerpají data z doplňku Microsoft Excel Power Pivot. Proto je nutné nahrát data z SQL databáze nejprve do Power Pivotu a následně vytvářet požadované vizualizace. Ostatní produkty (MS Excel, Reporty, EDE Dashboard) mají možnost čerpat přímo z SQL databáze.

4.4.2 Omezení výzkumného vzorku

Dalším omezením této práce jsou zdrojová data, která budou použita při testování vizualizačních BI produktů. V soukromé sféře je v dnešní době téměř nemožné získat potřebná data, která mohou být zveřejněna (mnohdy je problém i jejich zpracování). Společnosti se „bojí“ data poskytovat. Kvůli této neochotě a zároveň záměru otestovat vizualizace různých dat, od dat vhodných do map až po množství prodaných výrobků, jsem se rozhodla využít volně dostupnou databázi vhodnou pro oblast BI od společnosti Microsoft s názvem AdventureWorks. Databáze je poskytována s SQL serverem a následně využívána například při různých školeních od společnosti Microsoft, případně jiných školících společností (Microsoft 2014b). Konkrétně jsem zvolila databázi AdventureWorksDW2014(Microsoft 2014a). Tato databáze obsahuje do určité míry vyčištěná data. Data, která prošla zpracováním v SQL serveru a jsou připravena pro další zpracování v rámci OLAP kostky nebo vizualizace. Jak zmiňuje kapitola 4.2, záměrem této práce je najít vhodný vizualizační produkt a proto do zdrojové databáze zasahuji v minimálním rozsahu.

AdventureWorksDW2014 je databáze simulující nadnárodní společnost vyrábějící a prodávající své produkty na mezinárodních trzích. Konkrétně se zabývá výrobou a prodejem kol, jejich příslušenstvím, komponentami a oblečením na kolo. V rámci většího rozsahu této databáze jsem zvolila pro zjednodušení práce s databází pouze

.....

¹⁶ Lze využít, jakýkoli webový prohlížeč, volba Google Chrome je založena pouze na osobních preferencích.

její část, konkrétně tabulky související s prodejem.¹⁷ Konkrétně jsou z důvodu splnění scénářů použity tabulky s informacemi o zákaznících, zaměstnancích, produktech, čase a prodejích/nákupch.

5 Hodnocení scénářů

Kapitola hodnocení scénářů zahrnuje praktické aplikace vizualizací vytvořených na základě hodnocených scénářů, které jsou blíže specifikovány v kapitole 4.2. Kapitoly 2 a 3 obsahují blíže specifikované produkty, jejichž prostřednictvím jsou vytvářeny vizualizace. Vizualizační software, který představuje možnosti zpřehlednění dat pro zákazníka, je reprezentován z platformy Microsoft produkty Excel, Power Map, Power View a Reporty. Platformu open source zastupuje Pentaho, konkrétně produkt DCE Dashboard. Hodnocení vizualizací a produktů jako takových probíhá prostřednictvím bodově ohodnocených kritérií, která blíže specifikuje kapitola 4.3. Každá podkapitola obsahuje popis řešení dané vizualizace doplněným obrázkem konkrétní prezentace.

5.1 Scénář 1 – řešení

První scénář se zaměřuje na oblast nákupu zboží. Jeho prostřednictvím je řešena přehledná vizualizace obsahující informace o zákaznických nákupech. Jaké subkategorie produktů, v jakém množství a kde zákazníci nakupují. Zdrojové tabulky pro vizualizaci tohoto scénáře tvoří: dbo.DimDate, DimCustomer, DimSalesTerritory, DimProduct, DimProductCategory, DimProductSubcategory, FactInternetSales.¹⁸

5.1.1 Excel

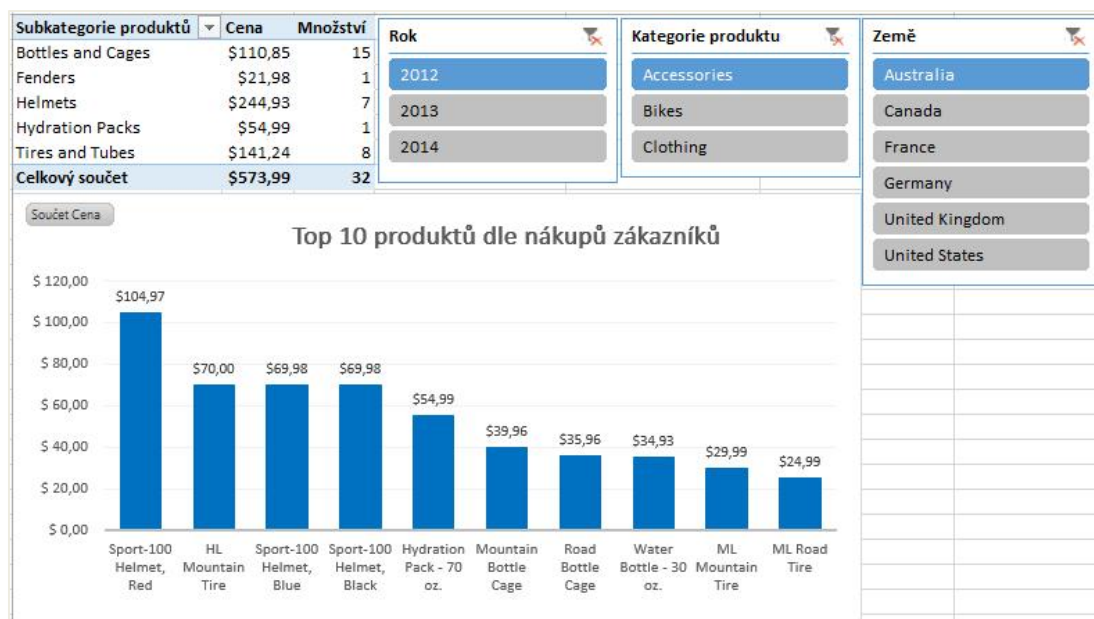
Pro realizaci prvního scénáře jsem jako první zařadila Excel. Tento produkt umožňuje na kartě Data vytvořit připojení k tabulkám ležícím na SQL serveru. Po natažení potřebných tabulek začíná samotná vizualizace. V Excelu jako takovém není možné vizualizovat data v rámci mapových pokladů, proto jsem se rozhodla použít tabulku a graf s připojením filtrů, jak je vidět na obrázku 16. Tabulka v levém horním rohu zobrazuje subkategorie s celkovou cenou, za kterou byly produkty spadající pod danou subkategorii prodány v daném roce, na daném území, další sloupec obsahuje množství prodaných produktů v kusech. Sloupcový graf znázorňuje top 10 produktů, za které zákazníci v daném roce, zemi utratili nejvíce dolarů. Vizualizace je filtrována na rok 2012, země Austrálie a kategorie produktu Accessories (příslušenství).

.....
¹⁷ Jak je zmíněno výše, cílem není vytvoření celého BI řešení, pouze srovnání vizualizační části.

¹⁸ Tabulky nejsou použity ve formátu, který byl převzat ze zdrojové databáze, bylo nutné provést určité úpravy (např. spojení tabulek). Jelikož nižší vrstvy BI nejsou předmětem této práce, nebudou tyto úpravy detailně popisovány.

Dostupnost Excelu pro potenciálního uživatele je dána dostupností balíčku Microsoft Office na serveru¹⁹, kde si chce daný člověk vizualizaci prohlédnout. Dále záleží na tom, zda je pro prohlížečícího dostačující pohled na statickou databázi. Pokud jsou data do databáze průběžně doplňována a uživatel chce vidět aktuální verzi, je nutné mít kromě Excelu s prezentací i přístup do zdrojové databáze. Jestliže je dostačující statická verze Excelu, je možná samostatná distribuce souboru obsahujícího vizualizace. Přístup do aplikace zabezpečeným způsobem, případně zobrazení dat dle uživatelského přístupu není možné.

Obrázek 16: Excel - vizualizace scénář 1



5.1.2 Power View

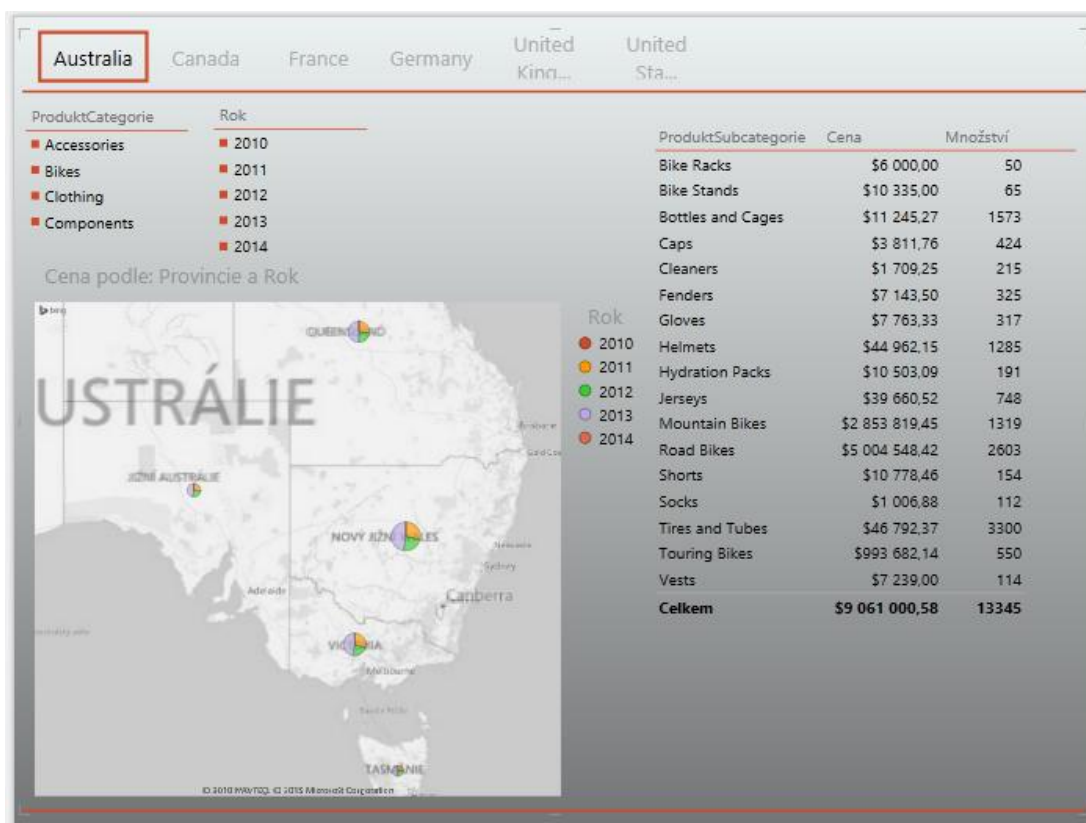
Pro realizaci řešení prvního scénáře prostřednictvím Power View bylo nutné nahrát zdrojová data z SQL serveru do Power Pivotu v Excelu, kde jsem vytvořila vazby mezi tabulkami. Power Pivot jsem vytvořila jeden pro všechny tři scénáře, dochází pouze ke změnám ve výběru zdrojových tabulek. Následnou vizualizaci v Power View jsem složila z tabulky obsahující seznam produktových subkategorií s celkovou cenou v dolarech, za kterou byly prodány produkty spadající do vybraných subkategorií a související množství. Dále obsahuje mapu, ve které je zobrazena cena prodaného zboží v závislosti na lokalitě zákazníků, na provincii kde nakupovali a letech ve kterých nakupovali. Velikost části kolečka na mapě zobrazuje množství peněz, za které

¹⁹ Server je pro účely této diplomové práce chápán jako počítač, notebook, případně jiné zařízení, na kterém je možné prohlížení vizualizací v Excelu a jiné.

nakoupili zákazníci v daném roce. Zobrazanou mapu a tabulku je možné filtrovat prostřednictvím produktových kategorií, v této vizualizaci není využito této možnosti. Dalším filtrem jsou roky, které umožňují časové omezení vizualizace. Jak můžeme vidět na obrázku 17, je možné celou vizualizaci filtrovat prostřednictvím vybrané země, která je na „dlaždici“ umístěné v horní části. Pro vizualizaci jsem vybrala Austrálii, takže všechny zobrazené údaje se týkají tohoto státu.

Dostupnost Power View pro potenciálního uživatele je v tomto případě dána dostupností Excelu s jeho doplňkem pro Power View a konkrétního souboru obsahujícího danou vizualizaci. Pokud pro prohlížečícího není nutné rozšiřovat data, není nutné připojení k databázi. V Excelu obsahující prezentaci v Power View jsou obsažena zdrojová data v rámci Power Pivotu. Pro distribuci je tedy postačující předání potřebného Excelu. Vizualizaci uživatel může prohlížet, filtrovat, upravovat i rozšiřovat. V rámci sdílení není možné nastavit uživatelské přístupy a omezit tak zobrazení dat pro určité uživatele.

Obrázek 17: Power View - vizualizace scénář 1



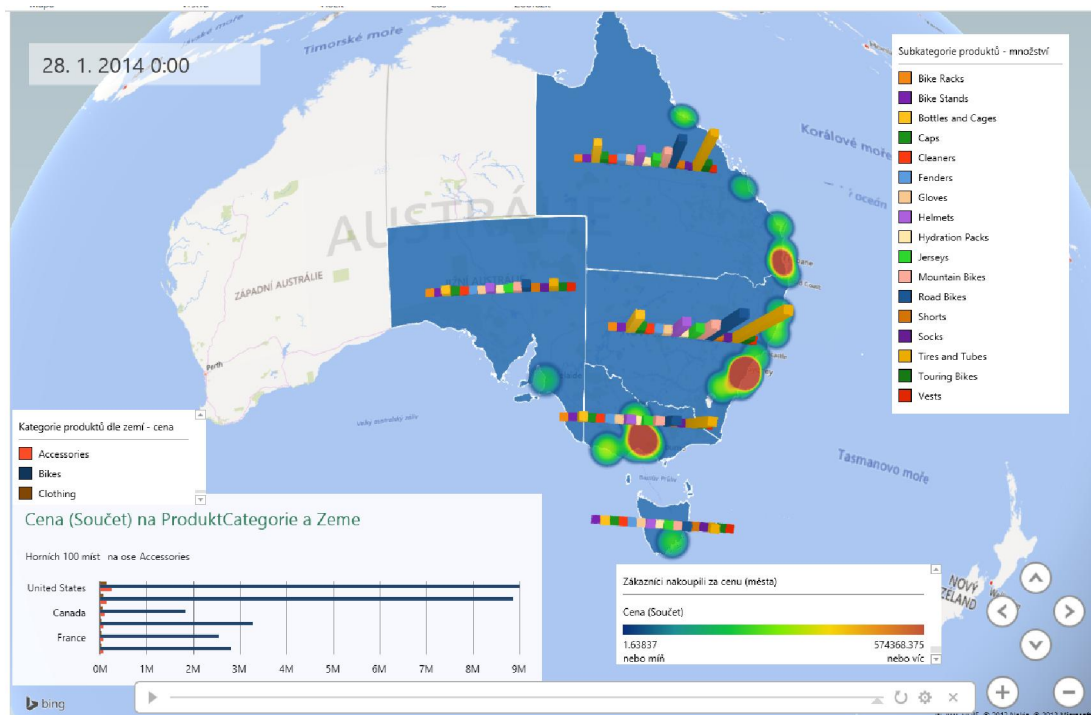
5.1.3 Power Map

Pro realizaci řešení v Power Map bylo nutné, stejně jako u Power View, nejprve data natáhnout do Power Pivotu. Stejný Power Pivot je použit pro Power View a Power

Map. Jak je vidět na obrázku 18, ústředním bodem vizualizace v Power Map je mapa, v tomto případě má tvar globusu. Stejně jako u předchozí vizualizace jsem se zaměřila na Austrálii. Modře vybarvené části představují provincie, ve kterých zákazníci nakupovali, tato část tvoří první vrstvu. V rámci jednotlivých provincií jsou zakresleny sloupcové 3D grafy představující množství jednotlivých subkategorií produktů, které bylo prodáno v rámci vybraných let (druhá vrstva). Do mapy je zakreslena ještě heat mapa, reprezentující cenu v dolarech za kterou zákazníci nakoupili (třetí vrstva). Pro geografické rozložení heat mapy slouží města, ve kterých nakupující bydlí. Do Power Map je možné vložit i 2D graf, v tomto případě reprezentuje cenu v závislosti na zemi a produktové kategorii. Vizualizace v Power Map jsou postaveny na animacích v čase. V tomto případě je vizualizace v posledním bodu daného času, který můžeme vidět v levé horní části obrázek 18. Ve spodní části je časová osa prostřednictvím, které je možné celou vizualizaci spustit. Časová osa v dané vizualizaci může být, ale není to podmínkou. Já jsem časovou osu navázala na druhou a třetí vrstvu, pokud tedy pustíme vizualizaci v čase, začnou 3D grafy růst a heat mapy se zvětšovat a měnit barvu. Jelikož Power Map umožňuje vizualizace pouze prostřednictvím map, tak je její využití vhodné pouze pro tento scénář a nebude již dále vytvářena.

Dostupnost Power Map je dána stejnými předpoklady jako u Power View, pouze s tím rozdílem, že Excel musí obsahovat doplněk pro Power Map.

Obrázek 18: Power Map – vizualizace scénář 1

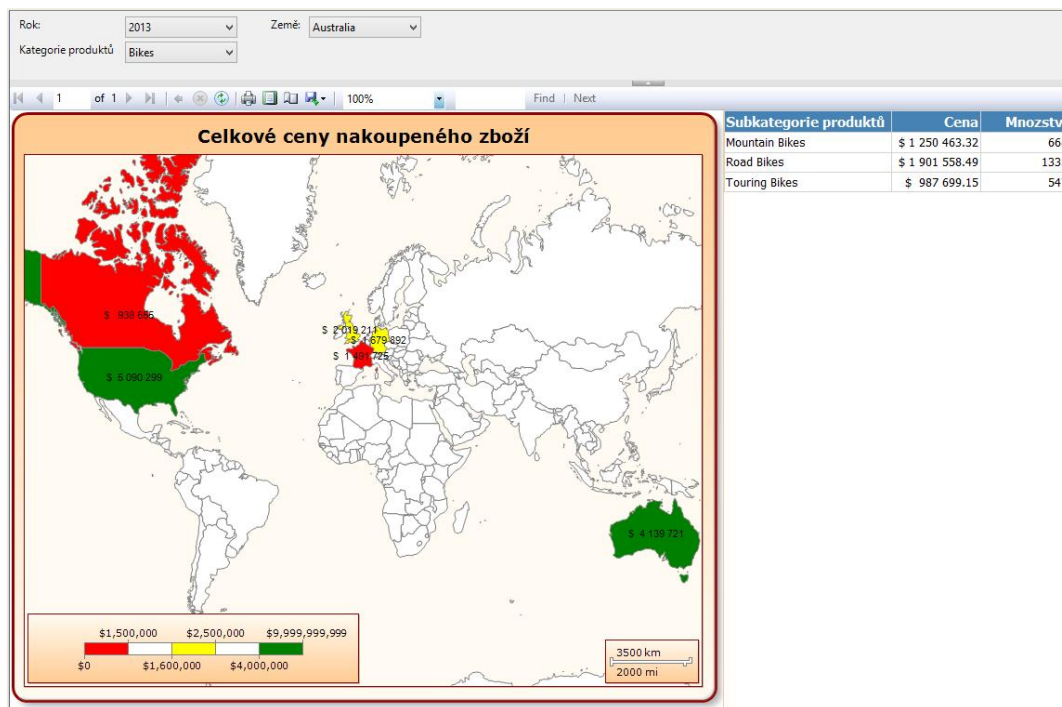


5.1.4 Reporty

Vizualizace reportu, která v této práci reprezentuje poslední vizualizační software z platformy Microsoft, je realizována prostřednictvím Report Designeru v prostředí Visual Studia. Nejprve jsem vytvořila sdílený Data Source a následně pokračovala ve tvorbě reportu. Obrázek 19 reprezentuje konečnou prezentaci řešení. Mapový podklad pochází ze zdroje (SANDVIK 2009) a zobrazuje země, ve kterých zákazníci ve vybraném roce a v dané produktové kategorii nakupovali výrobky. Barevné odlišení zemí je následující: země s nejnižší celkovou cenou prodaných výrobků jsou vybarveny červenou barvou, země s nejvyšší celkovou cenou prodaných výrobků jsou označeny zelenou barvou a země spadající mezi tyto kategorie má barvu žlutou. Každá barevně označená země obsahuje sumu všech cen v amerických dolarech za nakoupené výrobky v dané lokalitě. Na pravé straně od mapy leží tabulka obsahující informace o produktových kategoriích, celkových cenách za nakoupené produkty spadající do daných subkategorií a množství prodaných výrobků v kusech. Na celý report jsem použila filtrování v rámci let a produktových kategorií, na tabulku je použit ještě filtr země, v tomto případě je vybrán rok 2013, produktová kategorie Bikes (kola), a země Austrálie.

Přístupnost reportů pro potenciálního uživatele je dána dostupností webového prohlížeče a samozřejmě přístupem k internetu. Pokud report publikuji na internet, tak potenciálnímu uživateli stačí znát potřebnou adresu webové stránky a mít k ní zřízen přístup a může si prohlížet vizualizace. Pomocí přístupových práv lze řídit data zobrazovaná jednotlivým uživatelům.

Obrázek 19: Report – vizualizace scénář 1



Zdroj podkladu pro mapu: (SANDVIK 2009)

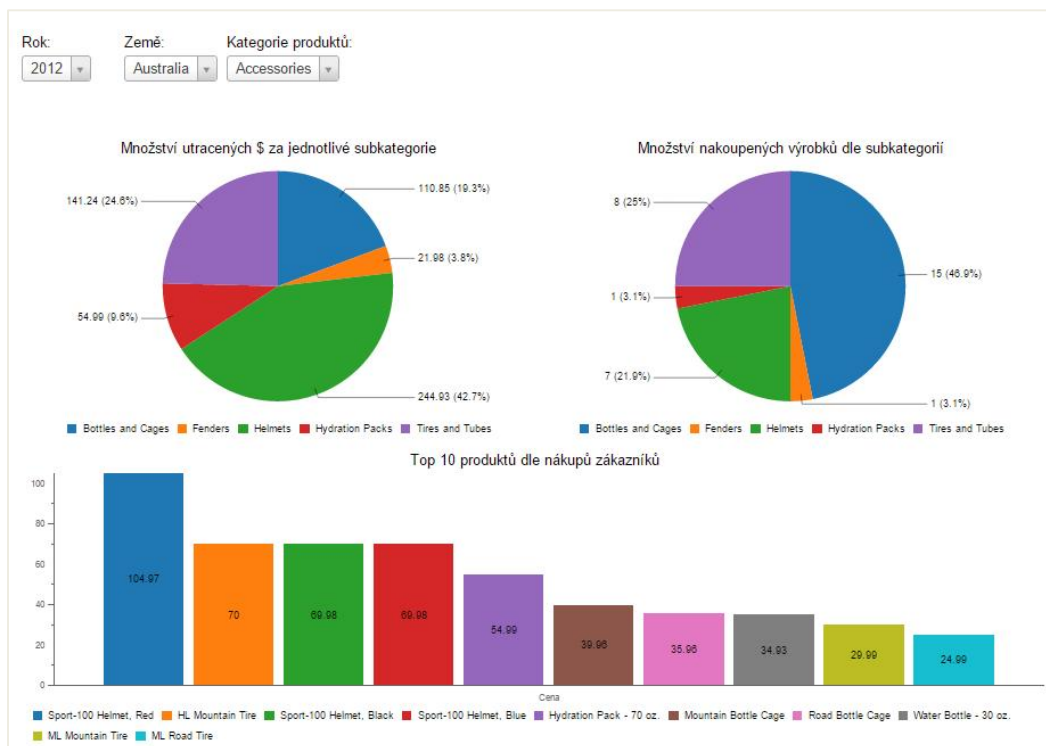
5.1.5 CDE Dashboard

Vizualizace v rámci open source platformy Pentaho je vytvářena prostřednictvím CDE Dashboardů. Aby bylo možné použití zdrojové SQL databáze, je nutné stažení ovladače sqljdbc_4.1²⁰. Java Database Connectivity (JDBC) umožní přístup k databázovému serveru. Následovně jsem vytvořila data source a nad tímto datovým zdrojem jsem vytvořila analytickou vrstvu. Po nadefinování data source je možné přejít k samotné vizualizaci. Pro vizualizace v rámci tohoto open source produktu existují mapy, ale není možné je zprovoznit (nefungují). Vizualizaci jsem vytvořila, jako kombinaci tří grafů a pomocných filtrů, jak ukazuje obrázek 20. První koláčový graf ukazuje, za kolik dolarů nakoupili zákazníci výrobky rozdělné dle subkategorií, kromě částky v dolarech je zaznamenán i procentuální podíl dané subkategorie na celku. Druhý koláčový graf reprezentuje množství nakoupených výrobků rozdělných dle subkategorií. Třetí graf (sloupcový) zobrazuje top 10 produktů, za které zákazníci nejvíce utratili. Celou vizualizaci je možné filtrovat prostřednictvím let, země a kategorie produktů. Mnou vytvořená vizualizace má data filtrována na rok 2012, v rámci Austrálie a Accessories (příslušenství).

²⁰ Lze stáhnout zde: <https://www.microsoft.com/en-us/download/details.aspx?id=11774>

Dostupnost CDE Dashboardů pro potenciálního uživatele je dána jednak dostupností webového prohlížeče a samozřejmě přístupem k internetu. Následně je postup obdobný jako u Reportů.

Obrázek 20: CDE Dashboard - vizualizace scénář 1



5.1.6 Řešení scénáře – hodnocení

Tato podkapitola obsahuje hodnocení řešení v rámci prvního scénáře. Požadavkem bylo zobrazit lokality, ve kterých zákazníci nakupovali v rámci jednotlivých států a jaké subkategorie produktů v jakém množství a za jakou celkovou cenu nakoupili. Výsledné hodnocení v rámci scénáře 1 je zaneseno v tabulce 2. Upravené bodové hodnocení, s přihlédnutím k vahám kritérií zobrazuje tabulka 3.

Excel splnil požadavek v rámci prvního scénáře pouze částečně, nejsou pro něj k dispozici žádné mapové podklady, tudíž údaje nemohly být zakreslovány na určená místa. Zbylou část včetně požadovaných filtrů splnil. Splnění požadavků scénáře napomáhá k hodnocení kritérií. Přehledné a dobře strukturované informace jsou splněny. Vhodně zvolené zobrazení s ohledem k datům není plně splněno vzhledem k absenci map. Volby navigace z pohledu konečného uživatele²¹ hodnotím kladně a z pohledu uživatele vývojáře taktéž. Pocit kontroly nad aplikací je pro uživatele

²¹ Konečný uživatel – výsledné vizualizace pouze sleduje, nevytváří

se středně pokročilými nebo vyššími znalostmi Excelu jistý, ale vizualizace je možné upravovat v rámci rozsahu Excelu, není zde úplná volnost. Podpora sdílení je popsána v kapitole 5.1.1, sdílení je možné prostřednictvím Excelu obsahujícího vizualizace, ve kterém bez připojení k datovým zdrojům není možné upravovat zobrazení.

V rámci vizualizace prostřednictvím Power View byl splněn první scénář včetně zakreslených údajů o zákaznících do mapy. První scénář je jediný, který je možné plně realizovat v Power Map, ostatní scénáře jsou pro mapový podklad ne úplně vhodné. Power Map splnila zakreslení do mapy, avšak prostřednictvím tohoto prezentačního prostředku není možné filtrovat zobrazovaná data. Pokud by bylo dáno přesné datové omezení v rámci konečné vizualizace, musela bych to provést již o úroveň níž, v datovém skladu. Požadavek na přehledné a dobře strukturované informace Power View splňuje, vizualizace v Power Map nemá tak přesná zobrazení, je zde povolováno značné zkrácení prostřednictvím autora vizualizace. Kritérium vhodné zvolené zobrazení dat je v obou případech splněno. Volby navigace jsou přehledně zobrazeny v obou případech Power zobrazení pro oba dva druhy uživatelů. (Mírně horší hodnocení dávám Power Map, kvůli absenci filtrů a dalších druhů grafů.) Opět kladně hodnotím pocit kontroly nad aplikací, jelikož tyto produkty byly vyvinuty pro samostatné vizualizace, které zvládne i uživatel neznalý programovacích jazyků. Podpora sdílení je uvedena v kapitole 5.1.2, sdílení je možné pomocí soubore MS Excel, který obsahuje jak vizualizaci, tak zdrojová data.

Pro zobrazení v mapovém podkladu v reportu jsem si stáhla podklad, který umožňuje zobrazení dat v rámci jednotlivých států světa. S přihlédnutím ke zbylým zobrazovaným datům a filtrům jsem požadavky z převážné části splnila. (Ačkoli vizualizace v mapě je možná pouze na úrovni států.) Kritérium přehledné a dobře strukturované informace je splněno pro konečného uživatele, pro uživatele, který provádí i vývoj to není úplně pravdivé tvrzení. Vhodně zvolené zobrazení dat do určité míry ano. Kritérium volby navigace při rozdělení uživatelů je pro konečného uživatele viditelnost zobrazených prvků výborná, ale pro vývojáře to není tak úplně pravdivé tvrzení. Pocit kontroly nad aplikací se u Reportů dostaví. Podpora sdílení probíhá prostřednictvím webu, takže výborná.

CDE dashboard stejně jako Excel nesplňuje lokalizační část požadavku. Přestože pro platformu Pentaho existují mapové podklady, tak s nimi nelze pracovat, nefungují. Zbylá část vizualizace včetně filtrování odpovídá nárokům. Kritérium přehledné a dobře strukturované informace je pro konečného uživatele splněno, pro vývojáře už to tak úplně pravda není. Vizualizace jako podpora rozhodování, toto kritérium není plně splněno, vzhledem k nefunkčnosti map. Volby navigace pro koncového uživatele jsou intuitivní, ale pro vývojáře ne tak úplně. Pokud se uživatel proškolí, tak

může mít plný pocit kontroly nad aplikací, jelikož při vývoji má nespočet možností. Podpora sdílení probíhá prostřednictvím webového rozhraní, takže toto kritérium hodnotím kladně.

Tabulka 2: Hodnocení scénáře 1 (v bodech)

Hodnocené kritérium	Excel	Power View	Power Map	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	10	10	9	10	10
2. Vizualizace, jako podpora rozhodování	8	10	9	9	8
3. Volby navigace	10	10	10	9	9
4. Pocit kontroly nad aplikací	9	8	8	10	10
5. Podpora sdílení	8	9	9	10	10
Hodnocení scénáře	45	47	44	48	47

Tabulka 3: Upravené hodnocení scénáře 1 (v bodech)

Hodnocené kritérium	Váha kritéria	Excel	Power View	Power Map	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	0,3	3	3	2,7	3	3
2. Vizualizace, jako podpora rozhodování	0,4	3,2	4	3,6	3,6	3,2
3. Volby navigace	0,1	1	1	1	0,9	0,9
4. Pocit kontroly nad aplikací	0,1	0,9	0,8	0,8	1	1
5. Podpora sdílení	0,1	0,8	0,9	0,9	1	1
Hodnocení scénáře	1	8,9	9,7	9	9,5	9,1

5.2 Scénář 2 – řešení

Druhý scénář se orientuje na zaměstnance. Jeho prostřednictvím je řešena otázka produktivity zaměstnanců. Tento scénář je konkrétně zaměřen na množství peněz, které přinesli zaměstnanci společnosti prostřednictvím prodeje produktů. Zdrojové

tabulky potřebné k vizualizacím: DimDate, DimEmployee, DimProduct, DimProductCategory, DimProductSubcategory, DimResellerGeography, FactResellerSales.²²

5.2.1 Excel

Před samotnou vizualizací třetího scénáře v prostředí Excelu jsem vytvořila nové připojení do databáze, z důvodu přehlednosti načítání datových zdrojů. Obrázek 21 představuje konečný pohled na řešení. V horní části jsou dvě tabulky obsahující informace o zaměstnancích, obratech získaných za prodej produktů a množství prodaných výrobků v kusech. První tabulka reprezentuje celkové hodnoty ukazatelů a druhá průměrné hodnoty. K ukazatelům obrát a průměrný obrát jsem připojila barevné indikátory, ze kterých je na první pohled vidět, jak si jednotliví zaměstnanci mezi sebou stojí v rámci získávání peněz za prodej produktů. Pod tabulky jsem umístila graf představující obraty zaměstnanců v závislosti na jednotlivých letech. Celý pohled je možné filtrovat prostřednictvím roku, pohlaví zaměstnance a kategorie produktů. Současné nastavení filtrů: rok 2012, zbylé filtry nejsou použity.

Obrázek 21: Excel - vizualizace scénář 2



²² Tabulky nejsou použity ve formátu, který byl převzat ze zdrojové databáze, bylo nutné provést určité úpravy (např. spojení tabulek). Jelikož nižší vrstvy BI nejsou předmětem této práce, nebudou tyto úpravy detailně popisovány.

5.2.2 Power View

Při řešení druhého scénáře prostřednictvím Power View jsem převzala Power Pivot s připravenými zdrojovými daty z řešení předchozího scénáře. Jak zobrazuje obrázek 22 vizualizace je tvořena tabulkou, sloupčovým grafem a filtry. Tabulka se skládá ze jmen zaměstnanců, průměrného obrátu jednotlivých zaměstnanců, maximálního obrátu jednotlivých zaměstnanců, celkem dosaženého obrátu a množství ks prodaného zboží. V řádku celkem u sloupce s průměrným obrátem hodnota udává průměrný obrát (pro všechny zaměstnance), dle této hodnoty je možné srovnávat, kdo je v tomto ohledu podprůměrný. Analogické je to u sloupce maximum. Ve sloupci obrát je celková hodnota obrátů a stejné je to i s množstvím. Celou tabulku jsem seřadila prostřednictvím sloupce Obrát, od nejvyšší po nejnižší hodnoty. Graf ve spodní části ukazuje obrát podle zaměstnance a roku. Vizualizaci jsem vyfiltrovala prostřednictvím roku 2011, dále je možné využít filtr produktové kategorie a pro zajímavost byl přidán filtr pohlaví. Při testování těchto dat jsem neobjevila žádné extrémní výkyvy mezi pohlavími.

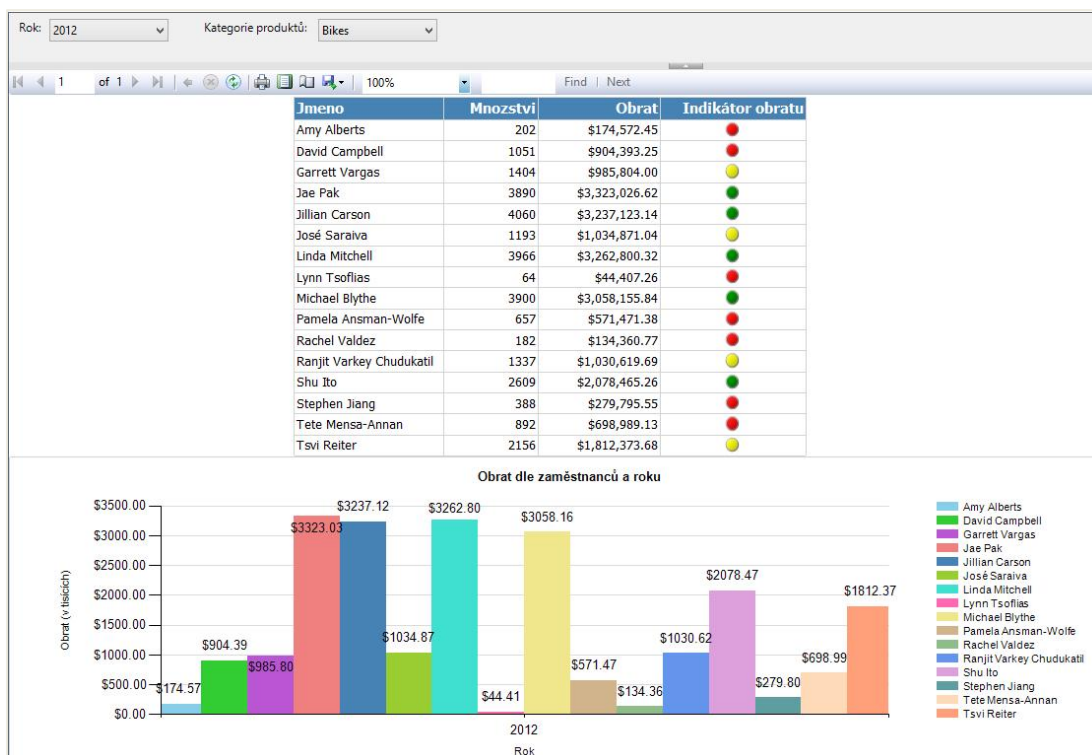
Obrázek 22: Power View – vizualizace scénář 2



5.2.3 Report

Při řešení druhého scénáře jsem pro získání zdrojových dat použila sdílený datový zdroj vytvořený při tvorbě prvního scénáře. Obrázek 23 prezentuje konečné řešení vizualizace. V horní části reportu jsem navrhla tabulku obsahující jména zaměstnanců množství prodaného zboží, obrat za jednotlivé zaměstnance a indikátor obratu. Poslední sloupec tabulky je tvořen barevnými indikátory, které slouží jako vizuální ukazatel, jak si jednotliví zaměstnanci mezi sebou stojí, v rámci získávání peněz za prodej zboží. Červená kolečka značí ty zaměstnance, kteří dosáhli menšího obratu, naopak zelená kolečka značí zaměstnance s vysokými obraty, žlutá kolečka mají zaměstnanci, kteří skončili uprostřed stupnice. Spodní část reportu patří sloupcovému grafu, zobrazujícímu přehled obrátů jednotlivých zaměstnanců ve vybraném roce. Hodnoty grafu jsou udány v tisících dolarů. Celý report lze filtrovat prostřednictvím roku a kategorie produktu, v tomto případě jsem nastavila rok 2012 a kategorii Bikes (kola).

Obrázek 23: Report - vizualizace scénář 2

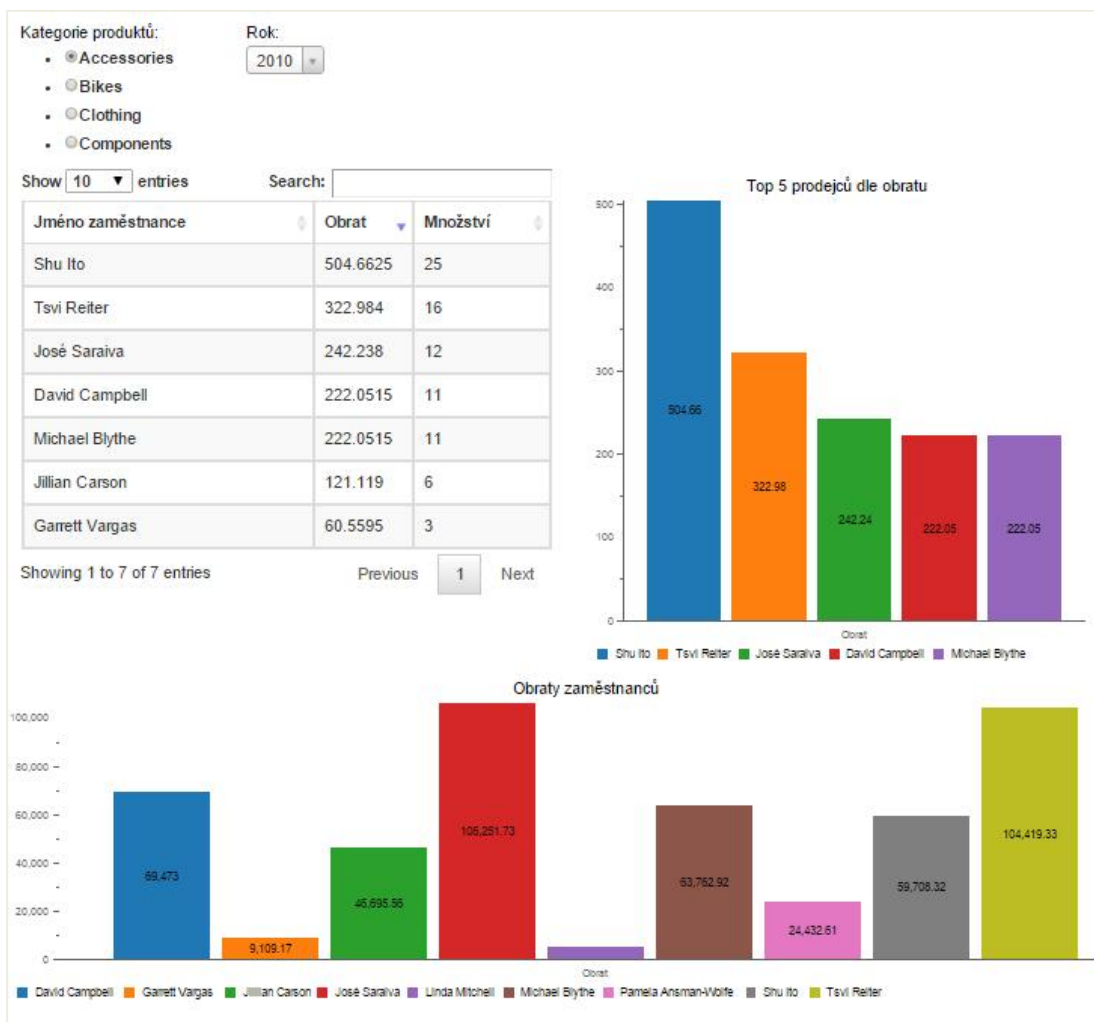


5.2.4 CDE Dashboard

Připojení zdrojových dat je pro řešení druhého scénáře stejné jako u prvního. Pro každou faktovou tabulku a k ní připadající dimenzionální tabulky je potřeba vytvořit nový data source (Analysis). Vizualizaci v CDE dashboardech v rámci řešení druhého

scénáře tvoří tabulka, dva sloupcové grafy a dva filtry, viz obrázek 24. Tabulka umístěna v horní části dashboardu znázorňuje získané obraty a množství prodaných produktů za jednotlivé zaměstnance (řazena podle velikosti obratu). Vpravo od tabulky je graf pěti prodejců, kteří měli ve vybraném roce největší obraty z prodeje produktů, spadajících do vybrané kategorie. Graf ve spodní části udává celkový obrat zaměstnanců za všechny produkty, na tento graf není připojen filtr kategorie produktu, je připojen pouze k filtru rok. Filtr obsahující produktové kategorie je vytvořen vizuálně jinak než filtry v předchozím řešení CDE dashboardu a filtr roky. U obou filtrů je umožněn výběr pouze jedné hodnoty. Vizualizace je filtrována na Accessories (příslušenství) a rok 2011.

Obrázek 24: CDE Dashboard - vizualizace scénáře 2



5.2.5 Řešení scénáře 2 - hodnocení

Požadavkem na řešení scénáře 2 je zobrazení zaměstnanců v souvislosti s obratem, který získali a množstvím prodaných výrobků. Prostřednictvím vizualizace má být na první pohled zřejmé, kteří zaměstnanci jsou v prodeji lepší a kteří horší. Filtrování má být umožněno v rámci roku a produktové kategorie. Výsledné hodnocení v rámci scénáře 2 je obsaženo v tabulce 4. Tabulka 5 zachycuje hodnoty po úpravách váhami kritérií.

Prostřednictvím řešení druhého scénáře v prostředí Excelu jsou požadavky na vizualizaci splněny. Zobrazený pohled obsahuje informace o obrazech i množství prodaných produktů. Vedle získaných obrátů/průměrných obrátů jsou vloženy indikátory související s velikostí obrátu/průměrného obrátu. Kritérium přehledných a dobře strukturovaných informací je splněno, uživatel se zde snadno zorientuje. Zobrazení dat je vhodně zvoleno a ovládací prvky vedou uživatele k intuitivnímu ovládání. Uživatel má do určité míry kontrolu nad aplikací. Podpora sdílení je specifikována stejně, jako při řešení předchozího scénáře.

Řešení druhého scénáře za pomoci prezentace dat v Power View splňuje požadované zadání, ačkoli zde nejsou použity na první pohled viditelné vizualizační prostředky určující velikost obrátů. V tabulce jsem hodnoty seřadila podle velikosti obrátu. Kritérium přehledného zobrazení dat je prostřednictvím této vizualizace splněno, grafy a tabulky jsou vhodně zvoleny. Ovládací prvky vedou k intuitivnímu ovládání. Pocit kontroly nad touto aplikací je pro uživatele značný, vizualizace dat je vytvářena poměrně intuitivně, avšak omezeně. Podpora sdílení dat je popsána v rámci předchozího scénáře.

Vytvořený report s obsahem řešení druhého scénáře splňuje požadavky včetně barevných indikátorů týkajících se získaných obrátů. Kritérium týkající se přehledných informací je v reportech splněno, stejně tak jako vhodně zvolené zobrazení dat s ohledem k vizualizovaným skutečnostem. Zobrazené ovládací prvky vedou k intuitivnímu ovládání a uživatel má kontrolu nad aplikací. Podpora sdílení u reportů probíhá v rámci webu.

CDE dashboard splňuje požadavky stanovené v rámci druhého scénáře, přestože neobsahuje indikátory zobrazující na první pohled velikost získaných obrátů. Kritérium přehlednosti informací a vhodného zobrazení dat jsou splněny. Volby navigace vedou k intuitivnímu ovládání aplikace. Uživatel cítí kontrolu nad aplikací, po ovládnutí tvorby řešení jsou možnosti vizualizace dat rozsáhlé. Podpora sdílení probíhá stejně jako u reportů v rámci webového rozhraní.

Tabulka 4: Hodnocení scénáře 2 (v bodech)

Hodnocené kritérium	Excel	Power View	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	10	10	10	10
2. Vizualizace, jako podpora rozhodování	10	9	10	9
3. Volby navigace	10	10	9	9
4. Pocit kontroly nad aplikací	9	8	10	10
5. Podpora sdílení	8	9	10	10
Hodnocení scénáře	47	46	49	49

Tabulka 5: Upravené hodnocení scénáře 2 (v bodech)

Hodnocené kritérium	Váha kritéria	Excel	Power View	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	0,3	3	3	3	3
2. Vizualizace, jako podpora rozhodování	0,4	4	3,6	4	3,6
3. Volby navigace	0,1	1	1	0,9	0,9
4. Pocit kontroly nad aplikací	0,1	0,9	0,8	1	1
5. Podpora sdílení	0,1	0,8	0,9	1	1
Hodnocení scénáře	1	9,7	9,3	9,9	9,5

5.3 Scénář 3 – řešení

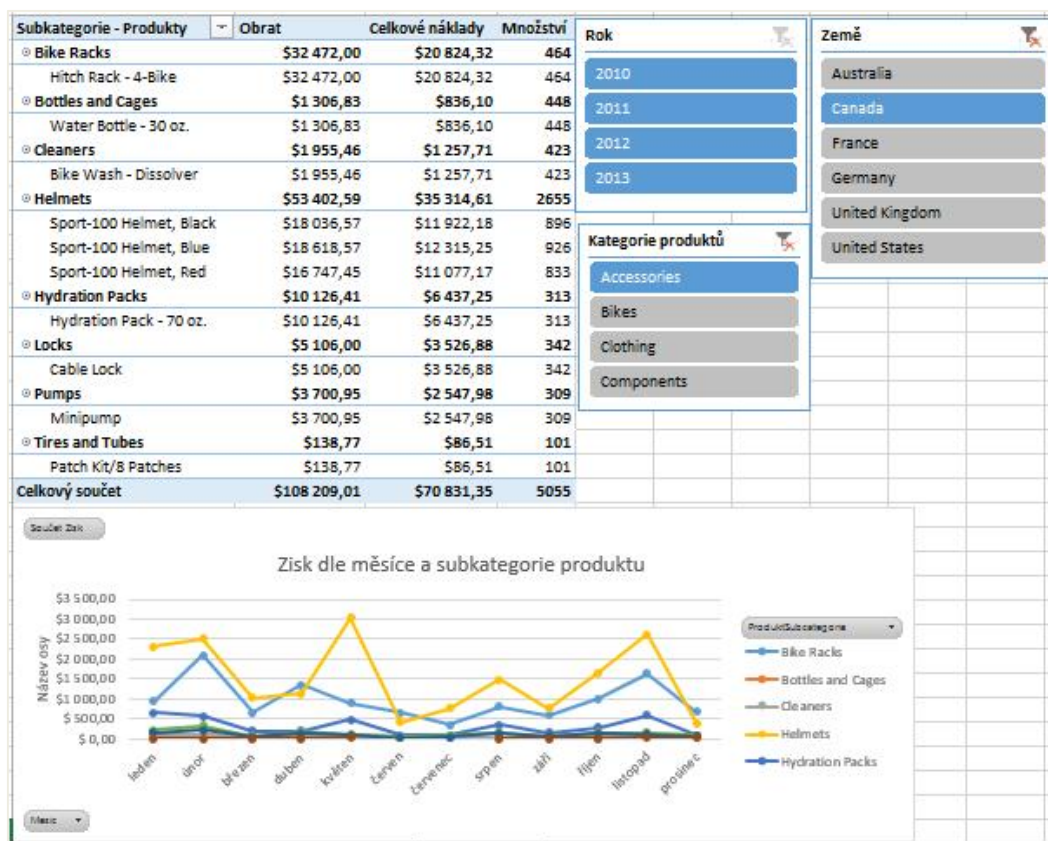
Třetí scénář se zabývá produkty. Prostřednictvím vizualizací tohoto scénáře je ukázáno, které produkty se prodávají a jaký přinášejí zisk, se zaměřením na vývoj v rámci jednotlivých měsíců v roce. Zdrojové tabulky pro zpracování tohoto scénáře: DimDate, DimProduct, DimProductCategory, DimProductSubcategory, DimGeography, DimReseller, FactResellerSales.²³

²³ Tabulky nejsou použity ve formátu, který byl převzat ze zdrojové databáze, bylo nutné provést určité úpravy (např. spojení tabulek). Jelikož nižší vrstvy BI nejsou předmětem této práce, nebudou tyto úpravy detailně popisovány.

5.3.1 Excel

Pro řešení třetího scénáře v Excelu jsem připojila stejný datový zdroj jako u předchozího scénáře. Celou vizualizaci tvoří tabulka obsahující informace o subkategoriích produktů a pod ně spadajících produktech. Informace tvoří velikost obrátu (množství dolarů získaných prodejem daných výrobků), celkové náklady na jednotlivé subkategorie a produkty v dolarech a množství produktů v kusech. Spodní část obsahuje graf zobrazující průběh zisku jednotlivých subkategorií produktů v rámci měsíců v roce. Celý pohled je možné filtrovat prostřednictvím Produktových kategorií, země a jednotlivých let. Na obrázku 25 můžeme vidět nastavení filtrů v rámci zemí na Kanadu, kategorie produktů na Accessories (doplňky) a filtr rok není nastaven.

Obrázek 25: Excel - vizualizace scénář 3

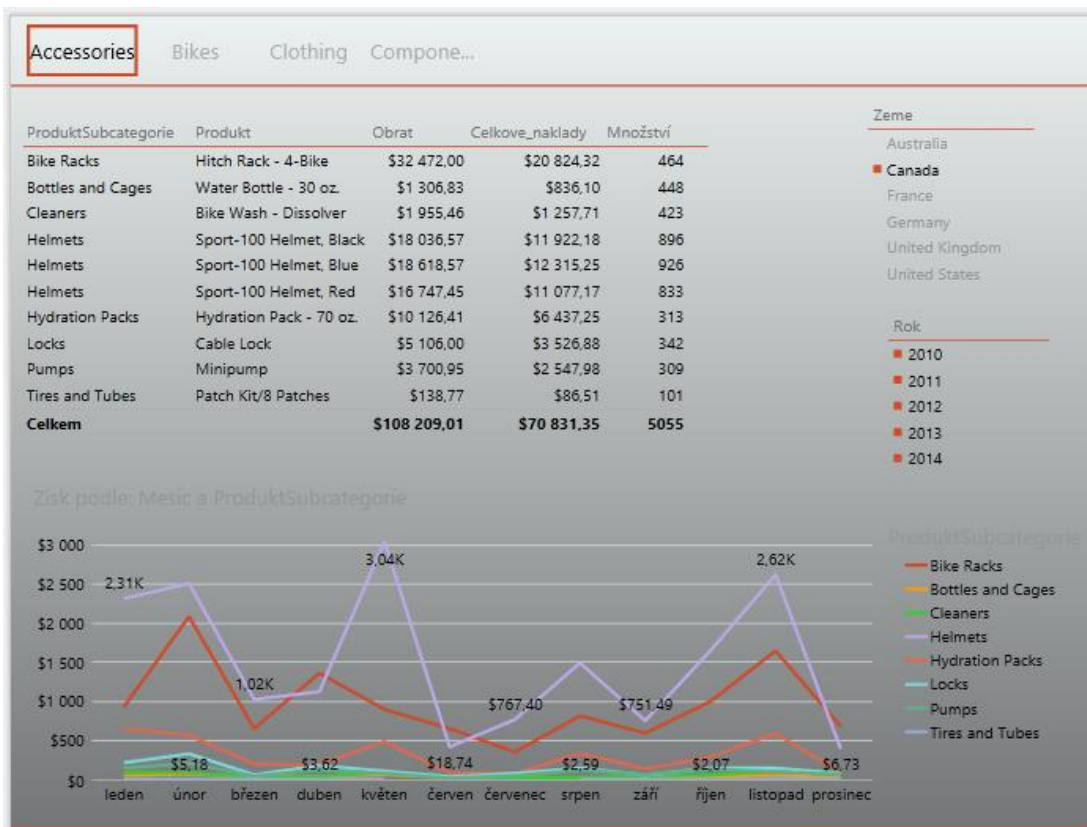


5.3.2 Power View

Při řešení třetího scénáře prostřednictvím Power View jsem použila stejný Power Pivot, jako u předchozích scénářů. Při návrhu řešení jsem použila „dlaždice“ s údaji týkajícími produktových kategorií, jako filtr v rámci celé vizualizace. Na obrázku 26 je vidět, co všechno obsahuje řešení třetího scénáře v rámci této vizualizace. Horní část zaplňuje tabulka obsahující údaje o produktových subkategoriích a samotných

produktech, obratech za jednotlivé produkty, celkové náklady a množství za jednotlivé produkty. Spodní část obsahuje graf, ve kterém je zakreslen zisk, který přináší jednotlivé výrobky v rámci měsíců ve vybraném roce. Celou vizualizaci je možné filtrovat v rámci jednotlivých zemí a let, já jsem nastavila filtr na Kanadu.

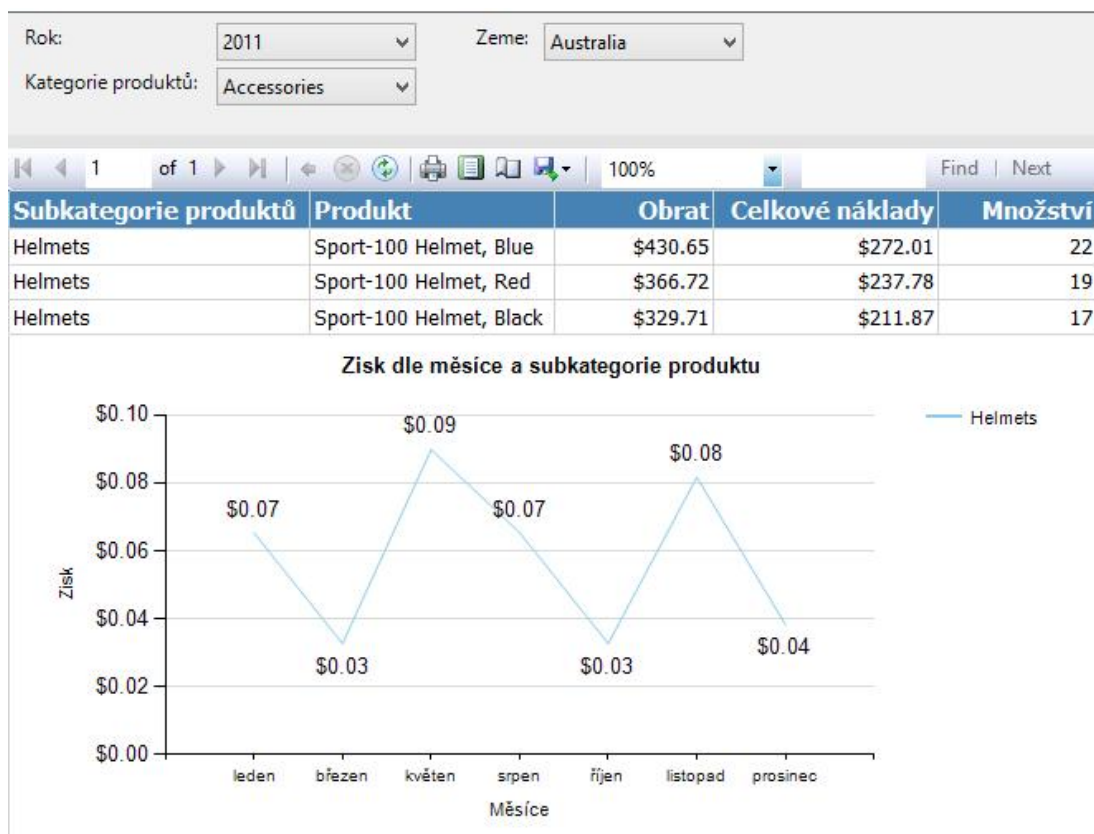
Obrázek 26: Power View – vizualizace scénář 3



5.3.3 Report

Pro vizualizaci třetího scénáře jsem použila sdílený datový zdroj, stejně jako při tvorbě prvních dvou scénářů. Skladbu výsledného reportu prezentuje obrázek 27. V horní části je vytvořena tabulka obsahující údaje o subkategoriích produktů a produktech samotných, o jejich obratech, celkových nákladech a množství. Ve spodní části reportu je obsažen spojnicový graf prezentující zisk jednotlivých subkategorií produktů v rámci měsíců ve vybraném roce, kde zobrazené hodnoty jsou v tisících dolarech. V reportu jsou obsaženy tři filtry, a to rok, země a kategorie produktů. Já jsem zvolila počáteční nastavení hodnot 2011, Accessories (příslušenství) a Austrálie.

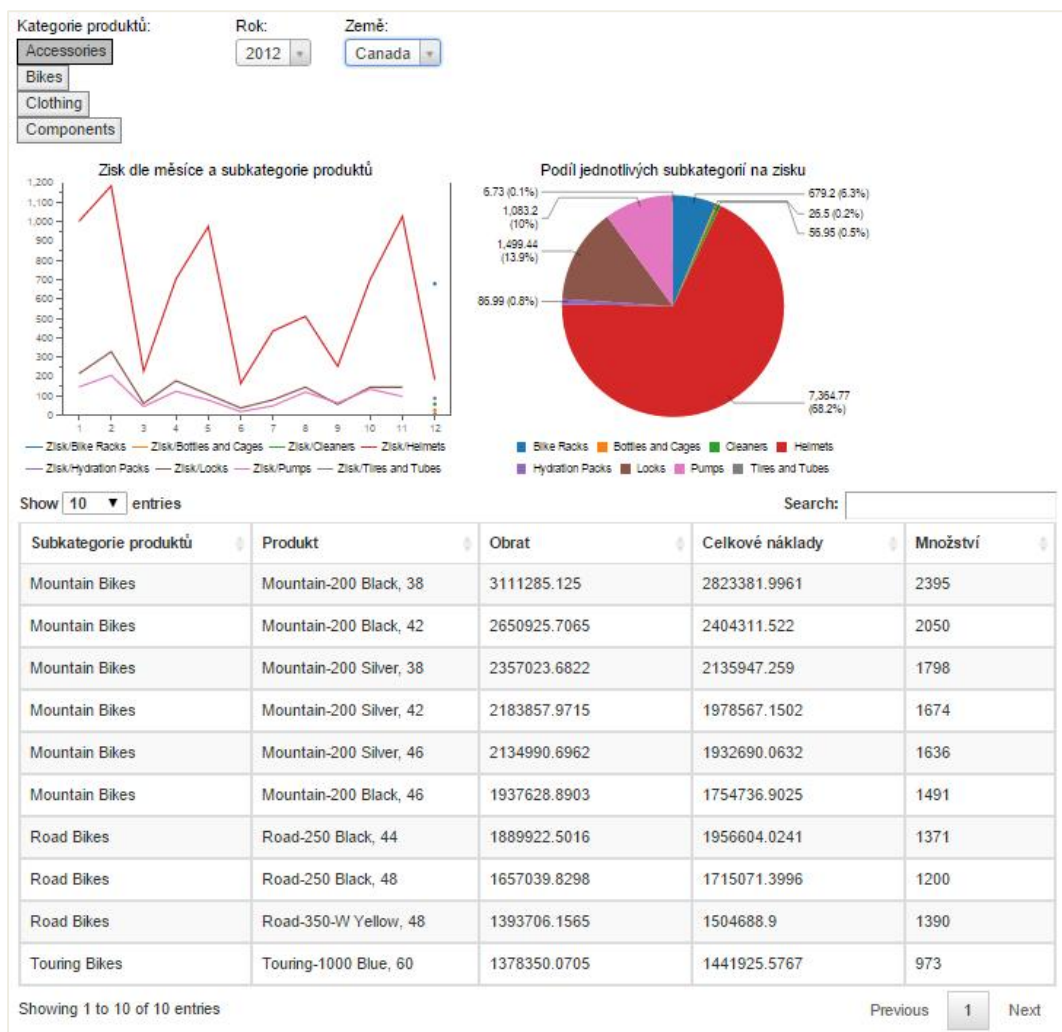
Obrázek 27: Report - scénář 3



5.3.4 CDE Dashboard

Pro vytvoření vizualizace k řešení třetího scénáře jsem v platformě Pentaho použila stejné datové připojení jako u prvních dvou řešení. CDE dashboard tvoří spojnicový graf, který podává informace o velikosti zisku subkategorií produktů v rámci měsíců ve vybraném roce (údaje v grafu jsou v dolarech). Dalším objektem obsaženým v reportu je koláčový graf zobrazující podíl jednotlivých subkategorií produktů na zisku. Údaje v grafu jsou uváděny v dolarech a závorky obsahují procentní podíl. Do spodní části jsem vložila tabulku, nezávislou na filtrech, která ukazuje deset produktů s největšími obraty. Tabulka obsahuje subkategorie a produkty s obraty a celkovými náklady jednotlivých produktů v dolarech a množství v kusech. Grafy v dashboardu je možné filtrovat prostřednictvím kategorií produktů, roku a země. Jak ukazuje obrázek 28, filtr kategorie produktů je rozdílný od všech do této chvíle použitých filtrů, nastavila jsem ho na Accessories (doplňky) a umožňuje výběr více kategorií. Další filtr rok jsem nastavila na hodnotu 2012 a zemi na Kanadu.

Obrázek 28: CDE Dashboard - vizualizace scénář 3



5.3.5 Řešení scénáře 3 - hodnocení

Požadavkem na řešení scénáře 3 je zobrazení produktů a jim nadřazených prvků v souvislosti s ukazateli obrat, celkové náklady, množství a zisk, dále zde má být zobrazen zisk v průběhu měsíců v roce. Tabulka 6 zachycuje výsledné hodnocení scénáře 3. Hodnocení po úpravách váhami kritérií je umístěno v tabulce 7.

Řešení třetího scénáře prostřednictvím Excelu splňuje zadané požadavky na vizualizaci. Vytvořený pohled zachycuje tabulku obsahující produktové subkategorie a pod ně patřící produkty se souvisejícími obraty, celkovými náklady a množstvím, jelikož je produktů mnoho, zvolila jsem kategorie produktů jako filtr. Do vizualizace jsem zakomponovala i graf obsahující vývoj zisku produktových subkategorií v jednotlivých měsících roku. Informace jsou na listu zobrazeny přehledně a strukturovaně. Zobrazení dat je zvoleno vhodně a ovládací prvky vedou uživatele

k intuitivním ovládání. Do určité míry má uživatel kontrolu nad aplikací. Sdílení je možné pouze prostřednictvím Excelu.

V aplikaci Power View vytvořené řešení třetího scénáře splňuje určené požadavky. Obsahuje požadované zobrazení ukazatelů v kombinaci s produkty a jejich subkategoriemi, včetně grafu vývoje zisku v rámci měsíců roku. Kritéria přehlednosti zobrazovaných informací i vhodně zvolených zobrazení jsou splněna. Ovládací prvky vedou uživatele k intuitivnímu ovládání. Uživatel cítí kontrolu nad aplikací, ale z hlediska vývoje vizualizace jsou jeho možnosti omezené. Sdílení je možné prostřednictvím poskytnutí souboru s vizualizací jiným osobám.

Report obsahující řešení třetího scénáře splňuje příslušné požadavky. Obsahuje tabulku s požadovanými informacemi i spojnicový graf zobrazující vývoj zisku za subkategorii produktů. Kritéria přehlednosti, strukturovanosti informací a vhodného zobrazení jsou splněna. Pokud uživatel nehledá krásu, tak je toto zobrazení dostačující. Ovládací prvky jsou viditelně zobrazeny a vedou k intuitivnímu ovládání. Uživatel má velkou kontrolu nad aplikací, za předpokladu získání potřebných znalostí. Podpora sdílení reportů probíhá v rámci webu.

Vizualizace v CDE dashboardu splňuje požadavky obsažené ve specifikaci třetího scénáře. Dashboard zobrazuje stanovené ukazatele obrát, celkové náklady, množství a zisk v souvislosti se subkategoriemi produktu/produkty. Navíc oproti ostatním vizualizacím jsem zde uvedla podíl jednotlivých subkategorií na zisku. Informace jsou rozmístěny přehledně, aby se uživatel snadno zorientoval. Kritéria vhodného zobrazení dat a viditelných ovládacích prvků vedoucích k intuitivnímu ovládání jsou také splněna. Uživatel po získání potřebných znalostí pocítí velkou kontrolu nad aplikací. Podpora sdílení probíhá prostřednictvím webového rozhraní.

Tabulka 6: Hodnocení scénáře 3 (v bodech)

Hodnocené kritérium	Excel	Power View	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	10	10	10	10
2. Vizualizace, jako podpora rozhodování	10	10	10	10
3. Volby navigace	10	10	9	9
4. Pocit kontroly nad aplikací	9	8	10	10
5. Podpora sdílení	8	9	10	10
Hodnocení scénáře	47	47	49	49

Tabulka 7: Upravené hodnocení scénáře 3 (v bodech)

Hodnocené kritérium	Váha kritéria	Excel	Power View	Report	CDE Dashboard
1. Přehledné a dobře strukturované informace	0,3	3	3	3	3
2. Vizualizace, jako podpora rozhodování	0,4	4	4	4	4
3. Volby navigace	0,1	1	1	0,9	0,9
4. Pocit kontroly nad aplikací	0,1	0,9	0,8	1	1
5. Podpora sdílení	0,1	0,8	0,9	1	1
Hodnocení scénáře	1	9,7	9,7	9,9	9,9

5.4 Hodnocení vizualizací

Cílem této části diplomové práce je zhodnocení vizualizačních systémů s přihlédnutím k vhodnosti použití. Hodnocení probíhá na dvou úrovních, nejprve na úrovni jednotlivých kritérií, jež jsou specifikovány v kapitole 4.3. Následně je uskutečněno hodnocení na základě scénářů. Posouzení na základě jednotlivých kritérií povede ke konkretizaci vlastností porovnávaných software a tím dojde ke specifikaci oblasti použití. Například pokud bude produkt využívat malá společnost s jedním vedoucím, tak není nutné vybírat software s přihlédnutím k přístupovým právům. Cílem celkového hodnocení scénářů je zjistit, zda byl vybrán open source produkt, který může konkurovat platformě Microsoft. A dále ukázat, jestli není některý z posuzovaných software lépe použitelný ve všech oblastech vizualizací než ostatní, případně zda některý ze zkoumaných produktů není vhodný pro konkrétní oblast, například pro vizualizaci geografický údajů. Hodnocení kritérií nadefinovaných v kapitole 4.1 je zobrazeno v tabulce 8 spolu s celkovým hodnocením scénářů. Při tvorbě řešení jednotlivých scénářů jsem si vyzkoušela práci s jednotlivými software, a proto do závěrečného zhodnocení zahrnu hlediska použitelnosti a uživatelské zkušenosti, specifikované v kapitole 4.1.

Tabulka 8: Celkové hodnocení přístupů

Software	Scénář	Kritéria hodnocení					Celkové hodnocení scénářů
		1	2	3	4	5	
Excel	Zákazníci	10	8	10	9	8	8,9
	Zaměstnanci	10	10	10	9	8	9,7
	Produkty	10	10	10	9	8	9,7
Hodnocení kritéria		10	9,33	10	9	8	9,43
Power Map	Zákazníci	9	9	10	8	9	9
Power View	Zákazníci	10	10	10	8	9	9,7
	Zaměstnanci	10	9	10	8	9	9,3
	Produkty	10	10	10	8	9	9,7
Hodnocení kritéria		10	9,67	10	8	9	9,57
Report	Zákazníci	10	9	9	10	10	9,5
	Zaměstnanci	10	10	9	10	10	9,9
	Produkty	10	10	9	10	10	9,9
Hodnocení kritéria		10	9,67	9	10	10	9,77
CDE Dashboard	Zákazníci	10	8	9	10	10	9,1
	Zaměstnanci	10	9	9	10	10	9,5
	Produkty	10	10	9	10	10	9,9
Hodnocení kritéria		10	9	9	10	10	9,5

První kritérium se zabývá přehlednými a dobře strukturovanými informacemi. Informace by měly být na pracovní ploše zobrazeny přehledně a strukturovaně, aby se v nich uživatel snadno zorientoval. První kritérium je splněno ve všech případech kromě jediného. Vizualizace v Power map dosáhla bodového hodnocení 9, důvodem ke snížení je nepřehlednost vizualizace a nemožnost filtrování. Druhým kritériem je vizualizace jako podpora rozhodování. V tomto kritériu je hodnoceno vhodně zvolené zobrazení dat. V Excelu a CDE Dashboardu byly strženy dva body při vytváření prvního scénáře z důvodu neexistence, nemožnosti použití mapových podkladů, bylo tedy nutné použít jiný druh vizualizace. Report má stržený jeden bod za vizualizaci v mapách, mapové podklady fungují, ale neumožňují zanesení údajů ve větším detailu. Jeden bod, za možná zkrácení dat ze strany uživatele, byl stržen i vizualizaci v Power Map. Není možné zjistit údaje o datech v rámci všech zobrazených sloupců

ve sloupcových grafech). V rámci řešení druhého scénáře bylo sníženo bodové hodnocení u vizualizací v Power View a CDE Dashboardů za nevložené vizuální indexy sloužící k posouzení velikostí obrátů na první pohled. Třetí kritérium jsou volby navigace. Ovládací prvky by měly být viditelné a vést uživatele k intuitivnímu ovládání, uživatel by měl rozumět ovládání aplikace. Při hodnocení tohoto kritéria dosáhly nižších hodnot Reporty a CDE Dashboardy z důvodu složitosti vytváření vizualizací pro nezkušené uživatele. Čtvrtým kritériem je pocit kontroly nad aplikací, který by měl uživatel cítit. U tohoto kritéria jsem hodnotila možnosti daných aplikací a jejich případná omezení. Nejlépe dopadly Reporty a CDE Dashboardy, jelikož tyto aplikace dávají určitou volnost uživateli při vytváření prezentací dat. Excel při hodnocení ztratil jeden bod z důvodu omezení aplikací Excel, která toho uživateli nedovolí udělat mnoho. Dva body ztratily aplikace Power View a Power Map z důvodu omezených možností vizualizací. Například nemožnost přejmenovat sloupce a jiné části dat přímo ve vizualizaci (Obrat podle: Rok a Jméno). U těchto doplňků MS Excel je umožněno vizualizovat daná data pouze podle „nabídky objektů“. Posledním pátým kritériem je podpora sdílení. Aplikace by měla být přizpůsobitelná individuálnímu použití nebo použití mezi pracovníky. V rámci hodnocení tohoto kritéria dopadl nejhůře s 8 body Excel. Tento software je možné sdílet prostřednictvím souboru obsahujícího prezentaci dat. Pokud uživatel, kterému je soubor sdílen nemá vytvořeno napojení na potřebná data nebo v horším případě nemá ke zdrojovým datům vůbec přístup, tak si může pouze prohlédnout vytvořenou vizualizaci, bez možnosti filtrovat a jinak upravovat zobrazovaná data. O něco lépe v hodnocení dopadly doplňky MS Excel Power View a Power Map. Jejich sdílení je možné prostřednictvím souboru, který obsahuje vytvořené vizualizace. Zobrazovaná data je možné upravovat a jinak s nimi pracovat, jelikož obsahují zdrojová data v Power Pivotu. Pokud chce uživatel daná data aktualizovat, musí vytvořit připojení do databáze, ze které pocházejí. Nejvíce bodů v oblasti sdílení získaly Reporty a CDE Dashboardy, které umožňují sdílení prostřednictvím webového prohlížeče a umožňují na rozdíl od ostatních spravovat přístupová práva k jednotlivým vizualizacím.

Celkové hodnocení scénářů, sloužící k výběru nejvhodnějšího software, je určeno výpočtem průměrného hodnocení scénářů. Mezi jednotlivými produkty nejsou značné rozdíly v rámci celkového hodnocení, viz tabulka 8. Nejmenšího celkového hodnocení dosáhl CDE Dashboard (9,50) a nejvyššího bodového hodnocení dosáhl Report (9,77). Rozdíl mezi nejnižším a nejvyšším hodnocením je tedy pouhých 0,27. Požadavek konkurenceschopnosti je tedy splněn. Rozdíly v případě jednotlivých scénářů také nejsou nikterak markantní. U prvního scénáře dosáhl nejvyšší hodnocení Power View (9,7) a nejnižšího hodnocení Excel (8,9). Rozdíl mezi těmito hodnoceními je 0,8. U druhého a třetího scénáře je nejvyšší bodový zisk 9,9 u Reportu

a v případě třetího scénáře i u CDE Dashboardu. Nejnižší hodnoty v případě druhého scénáře dosáhl produkt Power View a to 9,3, v případě třetího scénáře získal nejvyšší hodnotu 9,5 CDE Dashboard. Rozdíl mezi „nejlepším“ a „nejhorším“ je tedy 0,6 v případě druhého scénáře, a ve třetím je rozdíl pouze 0,4.

Žádný vizualizační software tedy není všeobecně výhodnější a uplatitelnější před ostatními produkty. Zde přichází na řadu pojmy použitelnost a uživatelská zkušenost, definované v kapitole 4.1. V úvodu této diplomové práce specifikuji cíle práce. Porovnávány produkty mají být vhodné pro použití v malé společnosti a uživatelé by jejich ovládnutí měli zvládnout po kratším školení. V rámci použitelnosti a uživatelské zkušenosti specifikuji dva druhy uživatelů. První uživatel je laik, který nezná žádný programovací (dotazovací) jazyk a nemá s vizualizací dat žádné zkušenosti, pracuje sám nebo s malým množstvím lidí. Druhým uživatelem je osoba, mající základní zkušenosti s programováním, vizualizací a data chce dále sdílet s dalšími spolupracovníky. Pro uživatele, který je popsán v prvním případě bych doporučovala použití produktu Power View. Tento produkt vychází v rámci hodnocených kritérií, viz tabulka 8 velmi dobře. Je uživatelsky přívětivý a v rámci prvních tří kritérií, týkajících se přehledných, dobře strukturovaných informací, vhodně zvoleného zobrazení dat a intuitivnímu ovládnutí, získal nejvíce bodů, ztratil z plného počtu pouze jeden. Pokud bude uživatelem laik, který chce, co nejjednodušším způsobem vizualizovat data a nepotřebuje je sdílet, jsou pro něj poslední dvě kritéria (pocit kontroly nad aplikací a podpora sdílení) druhořadé. Pokud druhý uživatel bude chtít mít plnou kontrolu nad vizualizacemi a chce vytvořené vizualizace sdílet, tak je pro něj vhodný spíše Report nebo CDE Dashboard. Jeho preferencí by byly poslední dvě kritéria a předchozí považuje za nepodstatné.

Závěr

Cílem mé diplomové práce bylo porovnat BI software pro vizualizaci dat na platformě Microsoft s vybraným open source produktem. Dílčím cílem bylo vybrat open source software, nacházející se na trhu, který by byl do určité míry konkurenceschopný s vizualizačními produkty platformy Microsoft. Vybrané software měly odpovídat požadavkům malé společnosti, kde by vizualizaci jako takovou mohli vytvářet sami zaměstnanci, například po kratším školení.

Dříve než jsem mohla začít s výběrem open source produktu, bylo nutné vysvětlit pojem business intelligence a popsat jeho komponenty. Následně jsem blíže specifikovala, jaký software vhodný k vizualizaci dat v oblasti business intelligence nabízí platforma Microsoft. V kapitole 2 je uveden výčet software, použitých při praktických ukázkách, obsahující produkty MS Excel, Power View, Power Map a Reporting, s jejich bližší specifikací. V kapitole 3 jsem popsala, jaké produkty z oblasti open source BI nabízí trh a následně vybrala software Pentaho a tento výběr zdůvodnila. V podkapitole 3.2 je blíže specifikován konkrétní vizualizační produkt (CDE Dashboard) použitý v praktických ukázkách.

Z důvodu objektivního zhodnocení a porovnání vybraných vizualizačních software bylo nutné vytvořit metodiku hodnocení, která je obsažena v kapitole 4. V podkapitole 4.1 jsem vybrala kritéria, jejichž prostřednictvím jsem v praktických ukázkách provedla zhodnocení. Zvolila jsem 5 kritérií a hodnocení prováděla na bodové stupnici od 1 do 10 bodů. Pro zaručení objektivity srovnávání jsem v podkapitole 4.2 vytvořila tři scénáře spadající do oblasti prodeje/nákupu produktů. Poslední podkapitola této části je tvořena předpoklady a omezeními.

V poslední kapitole následovalo vytvoření vizualizací dle stanovených scénářů a jejich zhodnocení. Tabulka 8 obsahuje celkové hodnocení kritérií a scénářů v rámci jednotlivých software.

Celkové hodnocení kritérií napříč všemi scénáři nepoukazuje na všestrannou vhodnost nebo naopak nevhodnost některého ze srovnávaných produktů. Avšak dokazuje to, že jsem vybrala open source produkt, který je konkurenceschopný platformě Microsoft. Při porovnávání v rámci prvních dvou kritérií, přehledných a dobře strukturovaných informací a vhodně zvoleného zobrazení dat, jsem dospěla pouze k drobným odchylkám v hodnocení. Tyto odchylky se týkají nezobrazení některých objektů, například chybějící mapové podklady. V rámci třetího kritéria, volby navigace, jsem body sebrala za složitější vytváření vizualizací. Při hodnocení posledních dvou kritérií, pocitu kontroly nad aplikací a podpory sdílení, jsem ubrala body

produktům, které mají omezené možnosti v rámci vývoje vizualizace a možností sdílení vizuálních prezentací. Přes drobné odchylky však hodnotím všechny software kladně a při takto vytvořeném zadání bych žádný nevyzdvihovala nad ostatní.

Výhodné použití testovaných software v rámci jednotlivých scénářů z bodového hodnocení nevyplývá. Jako software vhodný pro práci s mapovými podklady bych doporučila Power View. Prostřednictvím tohoto produktu se dají do mapy zaznamenat nejen země, ale i kraje, města a jiné geografické celky.

Z celkového hodnocení v rámci scénářů plyne, že srovnávané software poskytují obdobné možnosti pro vizualizaci dat. Žádný produkt není výrazně lepší ani horší, proto jsem přidala kritéria použitelnosti a uživatelské zkušenosti, definované spolu s kritérii v podkapitole 4.1. Jejich prostřednictvím jsem definovala dva vzorové uživatele. Jeden byl definován jako laik, neznalý programovacích (dotazovacích) jazyků, který nemá s vizualizací žádné předchozí zkušenosti a nepotřebuje data sdílet. Druhý uživatel měl základní znalosti z oblasti programování a vizualizací a výsledky své práce chtěl sdílet s kolegy. Pro prvního uživatele jsem doporučovala využít produkt Power View, jelikož v rámci prvních tří kritérií, přehledných a dobře strukturovaných informací, vhodně zvoleného zobrazení dat a volby navigace, vycházel velmi dobře. Kritéria pocitu kontroly nad aplikací a podpory sdílení jsou pro jeho užívání irelevantní, jelikož omezené možnosti aplikace mu nevaří a sdílet data nepotřebuje. Prostředí, sloužící k vytváření a prohlížení vizualizací prostřednictvím Power View, je uživatelsky přívětivé a výsledné vizualizace jsou na pohled hezké. Pro druhého uživatele, který chce mít plnou kontrolu nad aplikací a chce vytvořené vizualizace sdílet, bych doporučovala spíše použít vizualizaci v rámci Reportů, případně CDE Dashboardů. Tyto dva produkty získaly nejvyšší bodové hodnocení v rámci čtvrtého a pátého kritéria, tedy pocitu kontroly nad aplikací a podpoře sdílení. Report a CDE Dashboard poskytují uživateli možnost přizpůsobit si vizualizace svým představám, prostřednictvím programovacích (dotazovacích) jazyků. Dále také umožňují sdílení výsledných vizualizací prostřednictvím webového prostředí a nastavení přístupových práv pro uživatele. Největší výhodou CDE Dashboardů oproti vizualizacím na platformě Microsoft je možnost použití na jakémkoli operačním systému.

Z výsledku této práce usuzuji, že při přesné specifikaci požadavků od zákazníka (budoucího uživatele vizualizací) je možné vybrat software pro vizualizaci dat přesněji. Na základě této skutečnosti bych doporučovala přidání dalších kritérií, například časové náročnost, finanční stránky věci, případně požadavků na objem zpracovávaných dat. Dále považuji za důležité přesně specifikovat zaměření podnikatelské činnosti. Ve své diplomové práci jsem se zaměřila pouze na oblast

prodeje/nákupu. Dle bližší specifikace činnosti by následně byly upraveny i jednotlivé scénáře. Na základě těchto úprav by byl možný přesnější výběr vhodného software pro vizualizaci dat.

Seznam literatury

ADMIN, 2012. *Difference between ERP and CRM / WebERP4* [online]. [vid. 14. duben 2015]. Dostupné z: <http://blog.weberp4.com/difference-between-erp-and-crm/>

ANON., 2014a. 33 Open Source and Free Business Intelligence Solutions. *Predictive Analytics Today* [online] [vid. 30. červen 2015]. Dostupné z: <http://www.predictive-analytics.com/open-source-free-business-intelligence-solutions/>

ANON., 2014b. Computer Companies; Patent Issued for Inconsistency Robustness in Scalable OLAP Cubes. *Computer Weekly News* [online]. 16.10. [vid. 7. červen 2015]. Dostupné z: <http://search.proquest.com.zdroje.vse.cz/docview/1609126799?pq-origsite=summon>

ANON., 2015a. Computer Companies; Patent Issued for Method of Processing and Decomposing a Multidimensional Query against a Relational Data Source. *Journal of Engineering* [online]. 28.1. [vid. 7. červen 2015]. ISSN 1945-8711. Dostupné z: <http://search.proquest.com.zdroje.vse.cz/docview/1647178947?OpenUrlRftd=info:xri/sid:summon>

ANON., 2015b. *Data Integration & ETL / Jitterbit / Integrate data from any source* [online]. [vid. 14. duben 2015]. Dostupné z: <http://www.jitterbit.com/solutions/etl-data-integration/>

ANON., 2015c. *Medical Billing Online Dashboard* [online] [vid. 24. duben 2015]. Dostupné z: <http://www.claimcare.net/medical-billing-companies-claimcare/actionable-reporting/medical-billing-online-dashboard/>

BRADFORD, Mariannne, 2010. *Modern ERP: Select, Implement & Use Today's Advanced Business Systems*. Second Edition. Great Britain: Lulu. ISBN 978-0-557-43407-7.

ČERNÝ, Michal, 2012. *Vizualizace dat: Jak odhalit utajené souvislosti* [online] [vid. 26. duben 2015]. Dostupné z: <http://vtm.e15.cz/node/5048>

FOTR, Jiří, Lenka ŠVECOVÁ a KOLEKTIV AUTORŮ, 2010. *Manažerské rozhodování: postupy, metody a nástroje*. 2 přepracované vydání. Praha: Ekopress. ISBN 978-80-86929-59-0.

GRIMES, Seth, 2008. BI at 50 Turns Back to the Future. *InformationWeek* [online] [vid. 12. duben 2015]. Dostupné z: <http://www.informationweek.com/software/business-intelligence/bi-at-50-turns-back-to-the-future/211900005>

GUDFINNSSON, Kristens, Mattias STRAND a Mikael BERNDTSSON, 2015. Analyzing business intelligence maturity. *Journal of Decision Systems*. roč. 24, č. 1, s. 37–54. ISSN 12460125.

CHARLTON, Graham, 2013. 10 valuable Google Analytics custom reports. *Econsultancy* [online] [vid. 24. duben 2015]. Dostupné z:

http://econsultancy.com/blog/62271-10-valuable-google-analytics-custom-reports-2/?utm_campaign=bloglikes&utm_medium=socialnetwork&utm_source=facebook

JOOSTE, Chrisna, Judy VAN BILJON a Jan MENTZ, 2013. Usability Evaluation Guidelines for Business Intelligence Applications. In: *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference* [online]. New York, NY, USA: ACM, s. 331–340 [vid. 19. červen 2015]. SAICSIT '13. ISBN 978-1-4503-2112-9. Dostupné z: doi:10.1145/2513456.2513478

LACKO, Luboslav, 2009. *Business Intelligence v SQL Serveru 2008 Reportovací, analytické a další datové služby*. Brno: Computer Press, a. s. ISBN 978-80-251-2887-9.

LEGLER, R. a M. EPPLER, 2007. Towards A Periodic Table of Visualization Methods for Management. In: *ASTED Proceedings of the Conference on Graphics and Visualization in Engineering (GVE 2007)* [online]. Dostupné z: http://www.visual-literacy.org/periodic_table/periodic_table.pdf

MICROSOFT, 2014a. Microsoft SQL Server Product Samples: Database - Download. *CodePlex* [online] [vid. 23. červen 2015]. Dostupné z: <https://msftdbprodsamples.codeplex.com/downloads/get/880664?ProjectName=msftdbprodsamples>

MICROSOFT, 2014b. Microsoft SQL Server Product Samples: Database Product Samples. *CodePlex* [online] [vid. 23. červen 2015]. Dostupné z: <https://msftdbprodsamples.codeplex.com/releases?ProjectName=msftdbprodsamples>

MICROSOFT, 2015a. *Get started with Power Map* [online]. Dostupné z: <https://support.office.com/en-us/article/Get-started-with-Power-Map-88a28df6-8258-40aa-b5cc-577873fb0f4a>

MICROSOFT, 2015b. *Getting started with dashboards* [online] [vid. 16. červen 2015]. Dostupné z: <https://support.office.com/en-sg/article/Getting-started-with-dashboards-edc6a32c-85c5-443d-9103-56dcc7bb24a4>

MICROSOFT, 2015c. *Power Pivot: Powerful data analysis and data modeling in Excel* [online]. Dostupné z: <https://support.office.com/en-gb/article/Power-Pivot-Powerful-data-analysis-and-data-modeling-in-Excel-d7b119ed-1b3b-4f23-b634-445ab141b59b?omkt=en-GB&ui=en-US&rs=en-GB&ad=GB>

MICROSOFT, 2015d. *Power View: Explore, visualize, and present your data* [online]. Dostupné z: <https://support.office.com/en-us/article/Power-View-Explore-visualize-and-present-your-data-98268d31-97e2-42aa-a52b-a68cf460472e?ui=cs-CZ&rs=en-US&ad=CZ&omkt=en-US>

MICROSOFT, 2015e. *Power View in SharePoint Server: Create, save, and print reports* [online]. Dostupné z: <https://support.office.com/en-us/article/Power-View-in-SharePoint-Server-Create-save-and-print-reports-f616575e-02bf-46cd-b94a-b744245ef9d1?ui=cs-CZ&rs=en-US&ad=CZ&omkt=en-US>

- MICROSOFT, 2015f. *What is SharePoint?* [online]. Dostupné z: <https://support.office.com/en-us/article/What-is-SharePoint-97b915e6-651b-43b2-827d-fb25777f446f?ui=cs-CZ&rs=en-US&ad=CZ&omkt=en-US>
- MUILWIJK, Robin, 2014. *Top 3 open source business intelligence and reporting tools* [online]. Dostupné z: <http://opensource.com/business/14/6/three-open-source-business-tools>
- NL.WIKIPEDIA, Original uploader was SanderSpek at, 2004. *Enterprise Application Integration*. [online]. 20. leden 2004. [vid. 14. duben 2015]. Dostupné z: <http://commons.wikimedia.org/wiki/File:EAI.png>
- NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ, 2005. *Business Intelligence: Jak využít bohatství ve vašich datech*. 1. vydání. Havlíčkův Brod, Česká republika: Grada Publishing, a. s. ISBN 80-247-1094-3.
- PENTAHO, 2015. CDE Open Demos | Webdetails • Business Analytics. *Webdetails* [online] [vid. 18. červenec 2015]. Dostupné z: <http://www.webdetails.pt/cto-ols/knowledge/cde-open-demos/>
- PICCOLI, Gabriele, 2012. *Information Systems for Managers: Text and Cases*. Second Edition. United States of America: John Wiley & Sons. ISBN 978-1-118-05761-2.
- POWER, D. J., 2007. *A Brief History of Decision Support Systems* [online]. 10. březen 2007. B.m.: DSSResources.COM. [vid. 12. duben 2015]. Dostupné z: <http://dssresources.com/history/dsshhistory.html>
- SANDVIK, Bjørn, 2009. World Borders Dataset. *thematicmapping.org* [online] [vid. 16. červenec 2015]. Dostupné z: http://thematicmapping.org/downloads/world_borders.php
- SAXL, Ivan a Lucie ILUCOVÁ, 2004. Historie grafického zobrazování statistických dat. In: *ROBUST* [online]. Praha: JČMF, s. 363 – 386. ISBN 80-7015-972-3. Dostupné z: http://www.statspol.cz/robust/2004_robust2004.pdf
- SEN, Sandip, 2014. BI Visualization - New Trends for an Effective BI Solution. *Communications Today* [online]. 4. [vid. 26. duben 2015]. Dostupné z: <http://search.proquest.com.zdroje.vse.cz/docview/1517370952/abstract/6C682F50CD1E4225PQ/1?accountid=17203>
- SQIAR LTD, 2014. *OLAP DESIGN « Tableau Software Consultants* [online]. [vid. 15. červenec 2015]. Dostupné z: <http://www.sqiar.com/services/data-warehousing/olap-design/>
- WISE, Lyndsay, 2012. *Using Open Source Platforms for Business Intelligence: Avoid Pitfalls and Maximize ROI*. B.m.: Newnes. ISBN 978-0-12-415811-5.

1. Zdrojové kódy v elektronickém formátu na CD
2. Tabulka vizualizačních metod

[illegible]

Process Visualization	Structure Visualization	Hybrid Visualization
Overview	Detail AND Overview	Detail
Divergent thinking	Convergent thinking	
Ed different idea	Sd depth detail than	Sd depth detail than
Pf portfolio display	Pr performance display	Pr performance display
Sg strategic plan, want	St strategy map	St strategy map
Mz marketing campaign	Mc marketing cost	Mc marketing cost
Z zero/ly marketing cost	Ho hook of quality	Ho hook of quality
Ad advertising display	Fd feedback display	Fd feedback display
De decision display	Ft future map	Ft future map
Bm big market cost	Mq marketing quality	Mq marketing quality
Sttc strategy cost	Sr strategy map	Sr strategy map
Vc value chain map	Po product for focus	Po product for focus
Hy hybrid display	S strategy map	S strategy map
Is information display	Sm strategy map	Sm strategy map
Ta target map	Ld life cycle display	Ld life cycle display
Sd strategy display	Tc technology display	Tc technology display