

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačních technologií



Bakalářská práce

Platforma Pentaho BI v Business Intelligence

Tomáš Oláh

© 2018 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Provozně ekonomická fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Tomáš Oláh

Informatika

Název práce

Platforma Pentaho BI v Business Intelligence

Název anglicky

Platform Pentaho BI in Business Intelligence

Cíle práce

Téma práce je zaměřeno na open source software Pentaho BI, který poskytuje kompletní sadu nástrojů využívaných během procesu zavádění Business Intelligence do podnikové sféry. Hlavním cílem práce je ukázat jednotlivé nástroje této open source Business Intelligence platformy, které jsou využívány v řadě firem zabývajících se touto problematikou.

Metodika

Metodika řešené problematiky je založena na studiu a analýzy odborné literatury a publikací doprovázené veřejně dostupnými informačními zdroji. Řešení práce je realizováno znalostí potřebných nástrojů a orientace v problematice Business Intelligence. Vlastní řešení bude formulováno v závěru bakalářské práce na základě teoretických a praktických poznatků.

Doporučený rozsah práce

30 – 40 stran

Klíčová slova

Business Intelligence (BI), Data Warehouse (DWH), Pentaho BI, Enterprise Architect (EA), OLAP, Multidimenzionální modelování, Oracle SQL Developer

Doporučené zdroje informací

LABERGE, Robert. Datové sklady: agilní metody a business intelligence. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1

Lacko, Ľuboslav: Business Intelligence v SQL serveru 2008, Praha: Computer Press, 2008. ISBN 9788025128879

MURRAY, Dan. Tableau your data!: fast and easy visual analysis with tableau software. 2nd edition. Indianapolis, IN: John Wiley and Sons, 2015. ISBN 1119001196. 5. HOTKA, Dan. Oracle SQL developer handbook. New York: McGraw-Hill, c2007. ISBN 0071484744

POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business intelligence. Praha: Grada, 2004. EAN 24766850

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

Čestné prohlášení

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

V Praze dne 11.3.2018

Poděkování

Rád bych touto cestou poděkoval panu Ing. Janu Tyrychtrovi, Ph.D. za odborné vedení a za cenné rady při konzultacích během tvorby této bakalářské práce. Dále bych rád poděkoval Ing. Lukášovi Cíglerovi za jeho odborné rady v oblasti Business Intelligence.

Platforma Pentaho BI v Business Intelligence

Abstrakt

Téma bakalářská práce je zaměřeno na open source software Pentaho, který poskytuje kompletní plnohodnotnou sadu softwarových nástrojů využívaných během procesu zavádění Business Intelligence do podnikové sféry. První dvě kapitoly literární rešerše poskytují kritický a ucelený přehled problematiky manažerského rozhodování a oblasti Business Intelligence, včetně její architektury. Závěrečná kapitola literární rešerše se zabývá platformou Pentaho, prostřednictvím které je realizována praktická část práce. Tato kapitola obsahuje celistvý pohled na platformu Pentaho a její dílčí nástroje, jenž jsou k dispozici v open source verzi s názvem Pentaho Community Edition. V praktické části byly autorem práce implementovány nástroje platformy Pentaho, které dohromady představují Business Intelligence řešení, aplikované na volně dostupné datové sadě fiktivní společnosti Northwind Traders, nad kterou je pomocí těchto nástrojů provedena datová analýza. Výstupem praktické části jsou reportingové a ad-hoc self-service sestavy, které reflektují obchodní činnost společnosti Northwind Traders a mohou být užitečné při manažerském rozhodování.

Klíčová slova: Business Intelligence (BI), Data Warehouse (DWH), Pentaho BI, Enterprise Architect (EA), OLAP, Multidimenzionální modelování, Oracle SQL Developer

Pentaho BI platform in Business Intelligence

Abstract

The bachelor thesis is focused on Pentaho open source software, which provides a complete set of software tools used during the Business Intelligence process. Literary research provides a critical overview of the areas of managerial decision-making and Business Intelligence, including its architecture. The final chapter of literary research deals with the Pentaho platform. This chapter provides a complete view of the Pentaho platform along with its sub-tools in the Community Edition version. In the practical part, the Pentaho platform was implemented on the freely available dataset of the fictitious company Northwind Traders. Above this dataset, data analysis was performed using the above mentioned tools. The output of the practical part is reporting and ad hoc self-service assemblies that represent the business of Northwind Traders and can be used for managerial decision-making.

Keywords: Business Intelligence (BI), Data Warehouse (DWH), Pentaho BI, Enterprise Architect (EA), OLAP, Multidimensional modeling, Oracle SQL Developer

Obsah

1 Úvod.....	11
2 Cíl práce a metodika	13
2.1 Cíl práce	13
2.2 Metodika	13
3 Teoretická východiska	15
3.1 Manažerské rozhodování	15
3.1.1 Úvod do manažerského rozhodování.....	15
3.1.2 Různé přístupy k rozhodování	16
3.2 Business Intelligence.....	16
3.2.1 Úvod do Business Intelligence	16
3.2.2 Definice Business Intelligence.....	17
3.2.3 Data, informace, znalost a moudrost	18
3.2.3.1 Životní cyklus zpracování informací.....	19
3.2.4 Klíčové složky Business Intelligence	20
3.2.5 Vliv Business Intelligence na podnik	21
3.2.6 Dopad Business Intelligence na rozhodovací procesy.....	23
3.3 Architektura Business Intelligence	24
3.3.1 Zdrojové systémy.....	26
3.3.2 Extrakce, transformace a uložení dat.....	26
3.3.2.1 Extrakce dat	27
3.3.2.2 Transformace dat	27
3.3.2.3 Uložení dat.....	27
3.3.3 Datový sklad	28
3.3.3.1 Dimenzionální modelování	29
3.3.3.2 Datové tržiště.....	30
3.3.4 Střední analytická vrstva.....	31
3.3.4.1 OLAP kostka	31
3.3.5 Prezentační vrstva (front-end aplikace)	33
3.3.5.1 Reporty	33
3.3.5.2 Dashboardy.....	34
3.3.5.3 Ad-hoc sestavy	34
3.4 Platforma Pentaho	34
3.4.1 Historie platformy Pentaho.....	35
3.4.2 Verze platformy Pentaho	35

3.4.3	Postavení platformy Pentaho na trhu	35
3.4.4	Silné stránky platformy Pentaho	37
3.4.5	Výzvy platformy Pentaho	37
3.4.6	Nástroje platformy Pentaho	38
3.4.6.1	Business Analytics Platform.....	39
3.4.6.2	Data Integration	39
3.4.6.3	Schema Workbench.....	40
3.4.6.4	Report Designer	41
3.4.6.5	Saiku Analytics.....	42
3.4.6.6	Dashboard Maker	42
4	Vlastní práce	43
4.1	Úvod do praktické části.....	43
4.2	Databáze Northwind.....	43
4.2.1	Vlastnosti databáze Northwind.....	43
4.3	Návrh řešení pro společnost Northwind Traders	44
4.3.1	Současný stav	44
4.3.2	Návrh metrik	44
4.3.3	Architektura řešení.....	45
4.4	Budování datového skladu	46
4.4.1	Import datové sady do MS SQL Serveru.....	47
4.4.2	Návrh datového skladu	47
4.4.2.1	Výběr dimenzionálních tabulek.....	47
4.4.2.2	Výběr faktové tabulky	48
4.4.2.3	Vytvoření datového skladu v MS SQL Serveru	49
4.5	Extrakce, transformace a uložení dat do datového skladu	51
4.5.1	Vytvoření ETL transformace v Pentaho Data Integration	51
4.5.2	Automatizace ETL transformací.....	53
4.6	OLAP kostka.....	54
4.6.1	Tvorba OLAP kostky.....	54
4.6.2	Operace nad OLAP kostkou	55
4.7	Reporting.....	56
5	Výsledky a diskuse	59
6	Závěr.....	61
7	Seznam použitých zdrojů	63
8	Přílohy	66

Seznam obrázků

Obrázek 1: Životní cyklus zpracování informací (Eckerson 2003-vlastní zpracování)	19
Obrázek 2: BI prostředí (Eckerson 2003-vlastní zpracování)	21
Obrázek 3: BI architektura (Chaudhuri 2011-vlastní zpracování).....	25
Obrázek 4: Star schéma (Novotný Ota, Pour Jan, Slánský David 2004).....	30
Obrázek 5: Schéma BI (Novotný Ota, Pour Jan, Slánský David 2004)	31
Obrázek 6: OLAP kostka (Novotný Ota, Pour Jan, Slánský David 2004)	32
Obrázek 8: Magický kvadrant (Gartner 2017).....	36
Obrázek 9: Relační model databáze Northwind (Microsoft 2010).....	44
Obrázek 10: Architektura řešení (vlastní zpracování)	46
Obrázek 11: Model datového skladu (vlastní zpracování)	50
Obrázek 12: ETL transformace pro F_Orders (vlastní zpracování).....	51
Obrázek 13: ETL transformace pro D_Customers (vlastní zpracování)	52
Obrázek 14: ETL transformace pro D_Employees (vlastní zpracování).....	52
Obrázek 15: ETL transformace pro D_Products (vlastní zpracování)	52
Obrázek 16: Job ETL transformace (vlastní zpracování)	53
Obrázek 17: Skript BAT souboru (vlastní zpracování)	54
Obrázek 18: Ad-hoc sestava pro objemy prodejů (vlastní zpracování).....	56

Seznam použitých zkratk

BI – Business Intelligence
CRM – Customer Relationship Management
DB - Databáze
DDL – Data Definition Language
DM – Datové tržiště
DWH – datový sklad
ERP – Enterprise Resource Planning
ETL – Extrakce, transformace, uložení
IoT – Internet věcí
IT – Informační technologie
KPI – klíčové ukazatele výkonnosti
MS – Společnost Microsoft
OLAP – Online Analytical Processing
CE – Community Edition
PDM – Pentaho Dashboard Maker
PRD – Pentaho Report Designer
SQL – Strukturovaný dotazovací jazyk
WBS – Pentaho Workbench Schema

1 Úvod

Tato práce se zabývá plnohodnotnou Business Intelligence platformou s názvem Pentaho, která obsahuje dílčí nástroje, pomocí kterých je možné pokrýt implementaci Business Intelligence řešení do organizace v celém jejím kontextu.

Pojem Business Intelligence je znám již řadu let. Nicméně v posledních několika letech se Business Intelligence stává nedílnou součástí podnikových informačních systémů, protože organizacím umožňuje rychlým a efektivním způsobem sbírat, analyzovat a následně prezentovat informace, které jsou pro organizace klíčové při jejich rozhodování na všech úrovních, respektive na strategické, taktické a operativní úrovni. Business Intelligence je zastřešující oblast spadající do světa informačních technologií, která se skládá z dovedností, znalostí a technologií umožňující organizacím holistický pohled do vnitra organizace, díky pokročilým datovým analytickým technikám. Základní podstatou Business Intelligence je přeměnit data organizace na informace, mezi informacemi nalézt skryté vazby, ty následně přeměnit na znalost a tu poté využít k pozdějšímu moudrému rozhodování, které má organizaci pomáhat k dosahování jejich cílů při jejich obchodní činnosti. (Aliev, 2014)

V současné době se na trhu Business Intelligence vyskytuje již mnoho nástrojů od různých dodavatelů, které si v mnohém značně konkurují, nicméně se dá říci, že každý nástroj je svým způsobem specifický, a proto je výběr konkrétních nástrojů potřeba vždy do jisté míry důkladně zvážit, tak aby reflektoval byznys potřeby organizace. (Yoon, 2014)

Platforma Pentaho, která je hlavním předmětem této práce, představuje sadu softwarových nástrojů vhodných pro aplikaci Business Intelligence do prostředí organizace. Platformu Pentaho je možné rozdělit na komerční a open source verzi. Tato práce se zabývá výhradně open source verzí, která je volně ke stažení pro širokou veřejnost. Zásadní rozdíl mezi těmito verzemi je ten, že komerční verze obsahuje nástroje, které disponují pokročilejšími technikami, sofistikovanějšími možnostmi datové analýzy a větším množstvím nástrojů, jenž je možné využít například v oblastech Big Data, Data Science či Internet of Things. Pro využívání komerční verze je však nutné platit za licence. Na druhé straně je open source verze jednou ze dvou plnohodnotných Business

Intelligence platformem, které jsou v současné době dostupné na trhu. Open source verze obsahuje nástroje, které je možné využívat pro datovou extrakci, transformaci a integraci, tvorbu analytických technologií, Business Intelligence Self-service (ad-hoc analytické sestavy), dashboarding a reporting. Centrem všech nástrojů platformy je BI Server, pomocí kterého je možné výsledky datové analýzy prezentovat koncovým uživatelům, kterými může být vedení či management organizace.

Praktická část, která se zabývá implementací platformy Pentaho nad daty fiktivní společnosti Northwind Traders, se zaměřuje na aplikaci jednotlivých nástrojů open source verze od extrakce a transformace dat až po prezentaci výsledků získaných prostřednictvím datové analýzy nad daty společnosti Northwind Traders.

2 Cíl práce a metodika

2.1 Cíl práce

Téma práce je zaměřeno na open source software Pentaho BI, který poskytuje kompletní sadu nástrojů využívaných během procesu zavádění Business Intelligence do podnikové sféry.

Hlavním cílem práce a zároveň cílem praktické části je ukázat jednotlivé nástroje této open source BI platformy, které jsou využívány v řadě firem zabývajících se touto problematikou, prostřednictvím jejich aplikace na volně dostupné datové sadě relační databáze fiktivní společnosti Northwind. Hlavní cíl bude realizován prostřednictvím použití nástrojů platformy Pentaho za účelem vytvoření uceleného BI řešení pro podporu manažerského reportingu společnosti Northwind.

Cílem teoretické části je představit srozumitelným způsobem problematiku manažerského rozhodování, oblastí Business Intelligence a její architektury, a v neposlední řadě představit BI platformu Pentaho, díky které je realizována praktická část práce.

Dílčí cíle práce přímo reflektují hlavní cíl a jsou následující: identifikovat klíčové metriky nad datovým modelem společnosti Northwind Traders, následně navrhnout datový sklad, poté prostřednictvím nástroje platformy Pentaho provést extrakci a transformaci dat ze zdrojové databáze do autorem navrženého datového skladu a posléze nad tímto datovým skladem vytvořit prostřednictvím nástrojů platformy Pentaho analytické výstupy v podobě reportů a ad-hoc analytických sestav, které mohou posloužit managementu výše zmíněné společnosti.

2.2 Metodika

Metodika řešené problematiky je založena na studiu a analýze odborné literatury a publikací doprovázené veřejně dostupnými informačními zdroji.

Literatura, ze které bylo při tvorbě bakalářské práce čerpáno, obsahuje knihy od Ralpa Kimballa (2004) a Billa Inmona (2002), kteří patří k průkopníkům v oblastech

datových skladů a Business Intelligence, a i v současné době patří k nejuznávanějším odborníkům těchto problematik.

Řešení vlastní práce, ve které je realizována implementace BI řešení, je dosaženo prostřednictvím znalostí dílčích nástrojů open source platformy Pentaho a znalostí databázového systému MS SQL Server od společnosti Microsoft. Data, nad kterými je provedena implementace platformy Pentaho v praktické části, jsou data fiktivní společnosti Northwind Traders. Tato data představují volně dostupnou datovou sadu, kterou poskytuje společnost Microsoft pro testovací účely při práci s databázovým systémem MS SQL Server. Součástí datové sady je i datový model této společnosti. Vlastní práce je založena na znalosti metodik a technologií, které jsou nezbytné pro úspěšné nasazení Business Intelligence řešení. Mezi tyto znalosti, metodiky a technologie patří znalost problematiky relačních databází, návrhu datových skladů, transformačních pravidel při aplikování ETL procesů a zavádění oblasti Business Intelligence do podnikového prostředí. (Inmon, 2002), (Kimball, 2013)

Datový sklad obsažený v praktické části je navržen na základě hvězdného schématu. Hvězdné schéma představuje základní typ datového modelu, pomocí kterého jsou běžně navrhovány datové struktury v datovém skladě. (Kimball, 2013)

Aby bylo možné implementovat BI řešení prostřednictvím platformy Pentaho, ve společnosti Northwind Traders, bylo potřeba identifikovat klíčové metriky, na kterých stojí úspěšnost celého řešení. Vybrané klíčové metriky jsou identifikovány na základě analýzy relačního datového modelu společnosti Northwind Traders. Tyto metriky byly stanoveny jako důležité ekonomické ukazatele, které mohou pomáhat společnosti při sledování efektivity jejich obchodní a prodejní činnosti. (Turban, 2010)

Vlastní řešení práce bude formulováno v závěru bakalářské práce na základě teoretických a praktických poznatků.

3 Teoretická východiska

3.1 Manažerské rozhodování

Tato kapitola představuje problematiku manažerského rozhodování, protože hlavním konzumentem výstupních dat z Business Intelligence systémů bývá právě management organizace, kterému tato data poskytují podklady pro rozhodování.

3.1.1 Úvod do manažerského rozhodování

Podnikům, které chtějí uspět na dnešním rozsáhlém konkurenčním trhu ve své podnikové doméně, by mělo být jasné, co se skrývá v jejich datech. Je to dáno tím, že data jsou jedním z jejich nejdůležitějších aktiv. Díky úspěšné identifikaci skrytých vazeb mezi jednotlivými daty je možné dospět k holistickému pohledu do vnitra organizace, což pomůže podniku získat velmi cennou znalost svého prostředí. V dnešním velmi náročném konkurenčním prostředí určuje vítěze rychlost a přesnost zpracování informací. Následné využití těchto informací při moudrém rozhodování je základním stavebním kamenem při dosahování podnikových cílů, které jsou založené na potřebné znalosti dané obchodní problematiky. (Maister, 2001)

Manažeři se odjakživa považují za vedoucí pracovníky, kteří mají rozhodovací pravomoc ve své kompetenci, a o rozhodování mluví jako o umění. Podle nich je toto umění založené na tvořivosti, úsudku a intuici, které doprovází letité zkušenosti. Je prokázáno, že manažeři, kteří se soustředí na systematický přístup, který podloží metodickým, promyšleným a analytickým rozhodovacím procesem a nebudou brát v potaz a spoléhat na interpersonální komunikační dovednosti, budou více úspěšní při svém rozhodování. (Turban, 2010)

Skutečnost, že rozhodování hraje v organizacích zásadní roli, je na první pohled zřejmá, díky svým jasným dopadům. Tyto dopady mohou přinést příznivé či katastrofální důsledky. Právě proto musí být proces rozhodování podpořen fakty, nejčastěji na základě naměřených metrik či klíčových ukazatelů výkonnosti KPI (Key Performance Indicators), protože ty jsou důležitými ukazateli pro činnost podniku, a mnohokrát i faktorem identifikujícím správné rozhodnutí. V případě, že zvolené rozhodnutí bylo

chybné, musí následovat další navazující rozhodovací proces, který doprovází série dalších rozhodnutí k minimalizaci škod a identifikaci nového a zároveň správného rozhodnutí. Jestliže je analýza rozhodování zaměřená na synoptické racionalitě, musí se manažéři rozhodovat takovým způsobem, při kterém zhodnotí množinu všech možných a vhodných alternativ, a ty následně analyzovat z hlediska finančních nákladů a socioekonomických přínosů (Cost Benefit Analysis). Na závěr je potřeba zvolit vhodnou alternativu, která poskytuje nejvyšší možnou úroveň přínosu. (Aliev, 2014)

Výše uvedené rozhodovací procesy představují volbu mezi dvěma nebo více existujícími alternativami. V případě, že se vyskytnou nám již známé situace, které už někdy v minulosti nastaly, stačí tato rozhodnutí podložit intuitivním úsudkem. Výsledek výběru však vždy doprovází nejisté faktory, které jsou známi až v budoucnosti. Tato myšlenka utvrzuje fakt, že rozhodnutí musí být vždy proaktivní, rigorózní a vždy založené na faktorech, které jsou hodnoceny s určitou mírou nejistoty a rizika. (Aliev, 2014)

3.1.2 Různé přístupy k rozhodování

Ne všechny organizace přistupují k procesu rozhodování stejným způsobem. Rozdílné typy k přístupu procesu rozhodování je možné rozdělit např. podle dlouhodobého a krátkodobého charakteru. (Richards, 2017)

Příklady rozdílných přístupů k rozhodování můžeme najít například v soukromých a státních organizacích. Pokud jde o rozhodování v soukromých podnicích, mají manažéři a vedoucí pracovníci tendenci přijímat rozhodnutí krátkodobé povahy s obvykle omezeným počtem zainteresovaných osob. Státní organizace mají naopak tendenci přijímat rozhodnutí, nesoucí co nejmenší míru rizika s dlouhodobým charakterem, které zahrnují a vyžadují konzultace a diskuze velkého počtu zainteresovaných osob, např. úředníků, zájmových skupin a občanů před finálním rozhodnutím. (Richard, 2017)

3.2 Business Intelligence

3.2.1 Úvod do Business Intelligence

Jak již bylo zmíněno v úvodu, podnikové prostředí je stále čím dál více složitější a neustále měnící se oblast. Organizace musí pružně reagovat na změny pod častým tlakem

tím, že jsou svým způsobem inovativní. Znamená to jejich rychlou adaptaci a následné rozhodování na strategické, taktické a operativní manažerské úrovni. Aby mohla organizace přijímat správná rozhodnutí, musí mít dostatečný přístup k významnému množství relevantních informací a ty následně zpracovat a poté interpretovat. (Turban, 2010) Významné investice do informačních technologií (IT) jsou proto pro organizace a podniky stále důležitější, proto aby získaly konkurenční výhodu díky své efektivnější činnosti. (Rezaie, 2011)

Podnikatelské prostředí doprovází časté organizační reakce na změny vlivem neustále se měnících faktorů (obchodní trhy, náročné požadavky klientů, nové technologie a konkurenční společnosti). Toto prostředí vyvíjí tlak na organizace, a tím nutí organizace neustále vyhledávat příležitosti ke zlepšení a reagovat na změny proaktivním, reaktivním a adaptivním způsobem. Prostřednictvím využívání počítačových systémů, které jsou známi poloviny 90. let a řadíme je pod zastřešující termín Business Intelligence (BI), jsou organizace schopné efektivně a účinně reagovat na tyto události. Tato skutečnost vede k tomu, že se tradiční přístupy k rozhodování stávají zastaralými a neefektivními, a místo toho organizace využívají analytické IT systémy jako je například BI. (Turban, 2010)

3.2.2 Definice Business Intelligence

Neexistuje jednotná a přijatelná definice BI. Výsledkem většiny definic je většinou fakt, že každý autor zpravidla definuje termín BI svým způsobem do jisté míry individuálně. Nicméně je obecné porozumění pojmu BI napříč všemi různými definicemi od různých autorů stále poměrně podobná.

Zde jsou tři příklady konkrétních definic termínu BI:

"Business Intelligence je širokou kategorií technologií, aplikací a procesů za účelem shromažďovat, ukládat, přistupovat a analyzovat data takovým způsobem, aby se uživatelé mohli lépe rozhodovat." (Wixom, 2010)

"Business Intelligence jsou inovativní nástroje pro analýzu dat. BI umožňuje interaktivní přístup a manipulaci s daty tak, aby uživatelé získali cenné informace, a aby

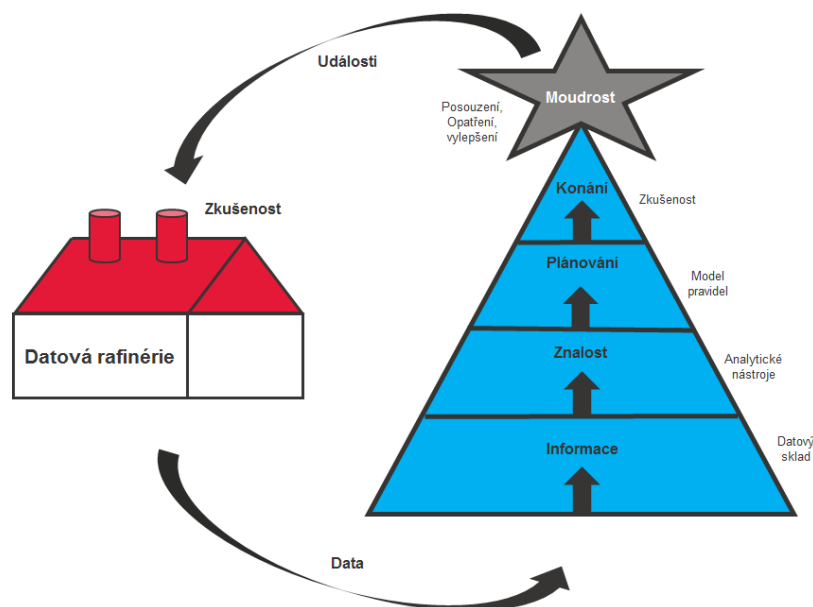
na jejich základě podporovali proces rozhodování v rámci celé organizace při podnikatelských aktivitách. " (Yoon, 2014)

"Business Intelligence jsou organizované a systematické procesy, které se používají k získávání, analýze a interpretaci informací důležitých pro podnikatelskou činnost. Společnosti díky tomu využívají tyto informace a znalosti za účelem podpory taktického a strategického rozhodování." (Hannula, 2003)

Výše uvedené definice BI mohou být na první pohled poměrně obtížné na pochopení a zařazení do celkového kontextu. Tyto definice nicméně stále popisují a vysvětlují podstatu toho, co je BI. BI jednoduše umožňuje uživatelům, vedoucím či řídicím pracovníkům a managementu získávat pomocí datové analýzy cenné informace, z nichž uživatel získá fakta, která je možné použít pro podporu rozhodovacího procesu. (Turban, 2010)

3.2.3 Data, informace, znalost a moudrost

Za účelem zjednodušení vysvětlení principu koncepce BI je možné BI přirovnat k ropné rafinérii, která je znázorněna na obr. 2. Obrázek znázorňuje proces zpracování ropy. Ropná rafinérie je navržena tak, aby převzala surovinu, respektive ropu a zpracovala ji do výsledných produktů jako je benzín, nafta, tryskové palivo či do jiné formy, například ve formě maziva. Toto je ekvivalent BI procesu, a lze tak BI přirovnat k "Datové rafinérii", která přebírá určitý druh suroviny, konkrétně data a transformuje je na jednotlivé produkty s různým informačním charakterem. (Eckerson, 2003)



Obrázek 1: Životní cyklus zpracování informací (Eckerson 2003 – vlastní zpracování)

3.2.3.1 Životní cyklus zpracování informací

Tato kapitola popisuje jednotlivé fáze životního cyklu zpracování informací. Níže je uveden životní cyklus zpracování informací podle Eckersona (2013):

1) Transformace dat na informace

Cílem této fáze je extrahování dat ze zdrojových systémů na jedno centrální místo, kde bude možné tato data sloučit a strukturovat, a pomocí dalších fází z nich získat cenné informace. (Eckerson, 2013)

2) Transformace informací na znalost

Po předchozí fázi, když jsou informace přístupné, je nutné data analyzovat v datovém skladu pomocí analytických nástrojů, které identifikují skryté vazby mezi daty, jejich trendy, vzory chování či anomálie, které se přemění z pouhých informací na konkrétní znalost. (Eckerson, 2013)

3) Využití znalostí při plánování

Na základě získaných znalostí z dat mohou být stanoveny konkrétní politiky organizace, které se pohybují od jednoduchých pravidel, až po komplexní a složitá pravidla vyžadující jistou míru sofistikovanosti. Mezi jednoduchá pravidla můžeme zařadit například situace, když podnik musí objednat a naskladnit nové produkty, když se zásoby sníží. Tato pravidla mají předvídatelnou povahu. Jsou to prognózy založené na historických trendech a předpokladech. Mezi komplexní pravidla

můžeme považovat např. dynamické nastavení cen v reakci na momentální poptávce či nabídky na trhu. Tato pravidla je poté možné implementovat a vytvořit na jejich základech konkrétní scénáře. Podniky tak mohou definovat, které produkty nabízejí zákazníkům na základě analýzy segmentů zákazníků a tím jim poskytnout služby či výrobky, které zrovna poptávají. (Eckerson 2013)

4) Jednání a konání na základě plánu

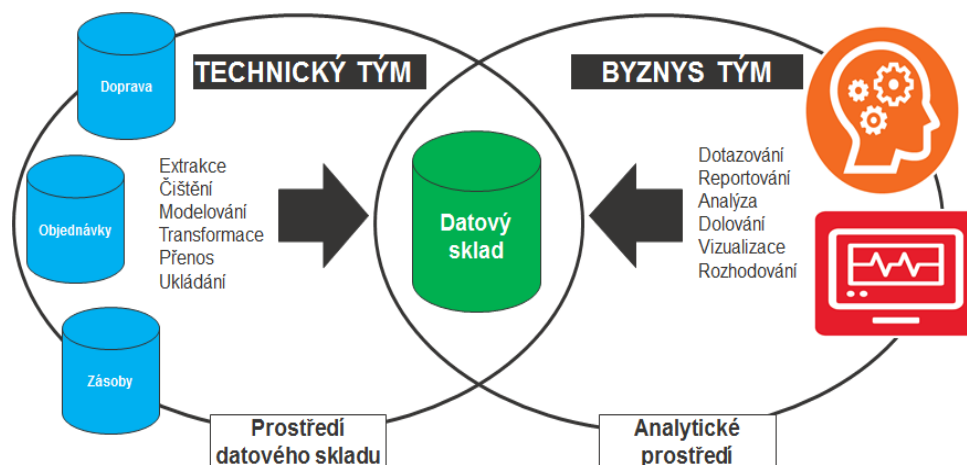
V této fázi životního cyklu zpracování informací podniky vytvářejí ucelené strategické, taktické a operativní plány na základě poznatků založených na pravidlech a znalostech. (Eckerson, 2013)

5) Zpětná vazba

Tato část nám dává velmi cenné údaje o odezvách z akcí, které byly předmětem našeho rozhodnutí. Mohou jimi být například informace o prodeji, které jsou uloženy v datovém skladu podniku. Tyto údaje jsou poté přezkoumány a analyzovány a později se zcela jistě mohou přeměnit na poznatky. (Eckerson, 2013)

3.2.4 Klíčové složky Business Intelligence

BI se skládá ze dvou klíčových složek. Mezi tyto složky patří datový sklad (datový sklad bude vysvětlen ve 3. části této kapitoly) a analytického prostředí. Díky těmto dvěma složkám dochází k interakci dvou různých typů uživatelů, resp. technického a byznys týmu. Technický tým svůj čas tráví v prostředí datového skladu, ve kterém získává, transformuje, ukládá a následně načítá provozní data z jednoho či více operačních systémů do prostředí datového skladu. Tato koncepce je zobrazena na obrázku č. 3. Vzhledem k tomu, že data nejsou vždy čistá nebo konzistentní mezi jednotlivými provozními systémy, bývá jen zřídka snadná integrace. Proto tato velmi náročná práce vyžaduje značné úsilí a odbornou technickou znalost. Byznys tým oproti technickému týmu komunikuje se systémem BI prostřednictvím analytického prostředí, kde tento tým získává data pomocí konkrétních dotazů nad datovým skladem a využívá vizualizační nástroje. (Eckerson, 2013)



Obrázek 2: BI prostředí (Eckerson 2003 – vlastní zpracování)

Turban (2010) založil své chápání principu BI na základě výše zmíněného přirovnání k ropné rafinérii od Eckersona (2003) ukázané na obr. 4. Nicméně argumentuje odlišnou kategorizací BI, která je rozdělena do čtyř složek:

- Datový sklad – přijímá, uchovává a centralizuje data z různých datových zdrojů
- Podniková analýza – soubor nástrojů, které analyzují data v datovém skladu
- Řízení výkonnosti podniku – sledování a analýza výkonnosti na základě KPI ukazatelů
- Uživatelské rozhraní – řídicí panel pro manipulaci dat nad datovým skladem

Podle Eckersona (2004) přebírají BI systémy data a přeměňují je na informace. Pomocí reportingových a analytických nástrojů, uživatelé přeměňují data na informace a později na znalosti. Na základě trendů a skrytých vazeb, které objevují v datech, mohou uživatelé vytvářet pravidla pro jejich operativní, taktické a strategické plány. Tyto plány jsou prováděny na základě určitých scénářů a pokaždé, když organizace prochází rozhodovacím procesem, tak společnost získává novou znalost o tom, jak funguje její podnikání, a na základě jakých rozhodnutí a akcí dosahuje požadovaných cílů. (Eckerson, 2013)

3.2.5 Vliv Business Intelligence na podnik

Výzkum vlivu systémů BI lze rozdělit do dvou různých skupin. Mezi první skupinu patří optimističtější skupina, která tvrdí, že BI hraje v organizacích zásadní roli tím, že

poskytuje užitečné poznatky, podporuje rozhodování a zlepšuje výkonnost podniku. Druhá, kritičtější skupina, která tvrdí, že i když existuje mnoho výhod BI systémů, je obtížné nebo téměř nemožné vypočítat nebo naměřit jejich pozitivní dopad na podnik, protože většina výhod je nehmotného charakteru. Mezi nehmotné výhody patří např. rychlejší a přesnější reporting, lepší rozhodování a komfortnější zákaznický servis. To je důvod, proč někteří manažeři nemohou jednoduše vysvětlit investice do BI vedení společností. (Turban, 2013), (Ramakrishnan, 2012)

BI usnadňuje organizacím ukládat, získávat a analyzovat velké množství informací, což vede k situaci, že se informace mohou přeměnit na reálnou přidanou hodnotu pro organizaci. Tato hodnota může být ve formě učení se z minulých úspěchů či selhání. Úspěšná analýza informací vede k identifikaci příležitostí, zlepšení ziskovosti podniku a dosažení lepší produktivity zaměstnanců s cílem získat významné postavení na trhu. (Ramakrishnan, 2012)

BI systémy nepřímo pozitivně ovlivňují finanční výkonnost tím, že zlepšují interní procesy a výkony zaměstnanců. (Elbashir, 2008), (Hou, 2010) Na druhou stranu existují také výzkumy, které zdůrazňují neutrální nebo dokonce negativní vztah mezi investicemi do IT a zlepšováním organizační výkonnosti. Ačkoli jsou tato tvrzení poměrně zastaralá a jejich význam může být zpochybněn, tak nám přináší řadu faktů. Jedním faktem je skutečnost, že všechny investice do IT nejsou totožné a je nezbytné měřit jejich efektivnost v celkovém kontextu organizace podle specifického účelu, podle kterého byly tyto IT systémy vybudovány. Tradiční míra efektivnosti investic do IT, jako je kupříkladu finanční výkonnost, je příliš široká a měla by být separována na menší sledovatelné oblasti, podle kterých je provedeno jejich následné měření a plnění (např. strategické, informační a produkční dopady). (Weill, 1992) Dalším zajímavým poznatkem je, že rozhodujícím faktorem při měření výhodnosti, jako je např. ziskovost organizace při měření vlivu nebo hodnoty na systémy BI, nemusí být vždy nutně finančního charakteru. Je to dáno tím, že tato měření obvykle nejsou v souladu s úmyslem organizace používat tuto technologii, ani neukazují přímý vliv těchto systémů na organizaci. (Elbashir, 2008)

Většina předchozího výzkumu týkajícího se dopadu BI na výkonnost organizace se soustředila na jeho finanční měření zvýšením výnosů nebo provozním výkonem měřením

zvýšením efektivity v hodnotovém řetězci. Někteří autoři však prezentují poznatky o dopadech BI konkrétněji a specifikují jeho dopady na rozhodovací procesy, které byly popsány v předchozí kapitole.”

3.2.6 Dopad Business Intelligence na rozhodovací procesy

Management BI, by měl zahrnovat účel a strategii BI, jeho implementaci a následnou podporu. Dále by měl zohledňovat kvalitu dat, kvalitu informací a rozsah působnosti BI (počet obchodních funkcí nebo procesů podporovaných BI nástroji). Tyto účinky totiž společně přinášejí pozitivní vliv na kvalitu rozhodovacího procesu organizace. BI management má zejména významný vliv na rozhodovací procesy prostřednictvím shromažďování vysoce kvalitních dat a informací. To má smysl, jelikož pokud podniky spravují svůj BI systém pomocí jasné a transparentní strategie, během které zazněli otázky proč, jak a kde bude BI systém implementován, následně udržován, a kde se budou analyzovat a transformovat data za účelem získávání vysoce kvalitních informací, jenž musí být relevantní a důvěryhodná, mohou docílit větší efektivity. (Wieder, 2015)

BI systémy zkracují čas a zvyšují účinnost rozhodovacího procesu tím, že umožňují získávat informace a znalosti z obrovských objemů zdrojových systémů obsahující nestrukturovaná data. Vysoce kvalitní data a informace mají pozitivní vliv na kvalitu rozhodovacího procesu. Pokud má organizace přístup k velkým objemům dat prostřednictvím BI systému, mají větší šanci, že bude lépe rozhodovat a bude mít k dispozici relevantnější informace, které povedou k lepšímu rozhodování společně s vyšší úsporou času. (Wieder, 2015)

BI má pozitivní dopad na rozhodování a je vhodnou podporou pro podnikový informační systém pro podnikové plánování zdrojů (ERP). BI systém integrovaný do celopodnikového systému zvyšuje rozhodovací výkonnost organizace. Na rozdíl od organizací, které používají pouze ERP systém. To znamená, že organizace zvyšuje efektivnost a přesnost svých rozhodovacích procesů tím, že implementuje a integruje BI systém do jejich současného podnikového systému, který může dále zahrnovat např. systém pro řízení vztahů se zákazníky (CRM), účetní systémy apod. Při implementaci BI systému za účelem zvyšování rozhodovací výkonnosti bylo zaznamenáno několik pozitivních přínosů. Mezi ně patří například - snížená doba potřebná pro rozhodnutí,

zvýšení spolehlivosti těchto rozhodnutí, rychlejší identifikace problémů, poskytnutí více možných alternativ při řešení problému a využití dalších informačních zdrojů při rozhodování. (Hou, 2010)

Výhody implementace BI systému reflektují také pozitivní dopady na rozhodování při provádění případových studií (Case Study). Tato skutečnost byla dokázána na univerzitě v Uppsale (USA), při které bylo zjištěno, že rozhodnutí zaberou méně času, mají lepší kvalitu a prezentují se mnohem snáze a srozumitelně. (Aziz, 2013)

3.3 Architektura Business Intelligence

Existuje mnoho architektur při řešení návrhu implementace BI. V první řadě je potřeba navrhnout architekturu BI systému a vymezit její rozsah, složitost a komplexnost takovým způsobem, který bude nejvíce zapadat do potřeb konkrétní organizace. (Slánský, 2004)

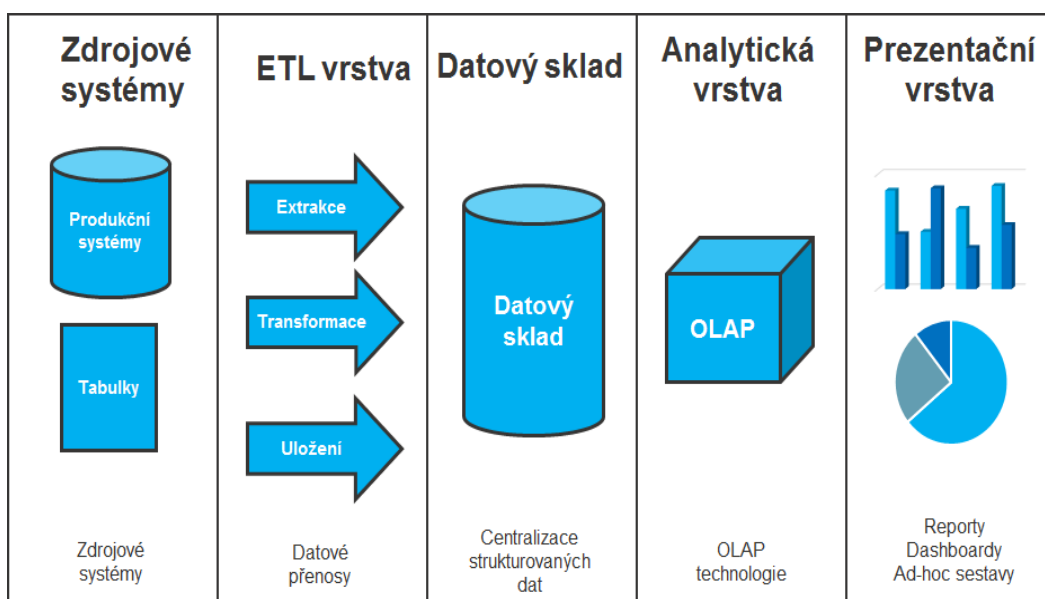
Níže je uvedena struktura BI architektury od Chaudhuriho (2011) z knihy „*An overview of Business Intelligence Technology*“. (Chaudhuri, 2011)

Model BI architektury se skládá z pěti složek (Chaudhuri, 2011):

- 1) Zdrojové systémy: tyto systémy tvoří veškeré zdroje dat podniku prostřednictvím databázových systémů (např. ERP systémy, účetní systémy, transakční či provozní systémy, CRM systémy atd.) Tyto systémy poskytují surová a nestrukturovaná data.
- 2) Integrační vrstva a ETL nástroje: vrstva sloužící pro extrakci, transformaci a ukládání dat ze zdrojových systémů do datového skladu.
- 3) Datový sklad (DWH): DWH je základní komponenta BI řešení, a data jsou zde archivována prostřednictvím multidimenzionálních databází.
- 4) Střední analytická vrstva: zde jsou data analyzována pomocí sofistikovaných analytických metod, kterými BI nástroje disponují.
- 5) Prezentační vrstva (front-end aplikace): tato vrstva slouží pro prezentaci výsledných analýz získaných z dat uložených v datovém skladě, jež jsou určeny pro koncové uživatele.

Celý výše uvedený proces probíhá tak, že data jsou z první částí (zdrojových systémů) extrahována z různých datových zdrojů. V druhé části dochází k jejich následnému čištění, transformaci do požadované podoby a integraci na společné místo. Data jsou poté v požadované struktuře a formátu uložena do datového skladu, který je třetí částí BI modelu. V datovém skladu se data archivují v neměnné formě. Následující střední analytická vrstva načítá data z datového skladu a poté je analyzuje. Ve střední analytické vrstvě se nachází např. technologie OLAP (Online Analytical Processing), která umožňuje nahlížet na jednotlivá data pomocí n-rozměrných dimenzí postavených nad multidimenzionálním modelem. Poslední a neméně důležitou součástí architektury BI je prezentační vrstva, která využívá front-end aplikace, které umožňují koncovým uživatelům nahlížet na výsledky analýzy dat pomocí statických reportů nebo dynamických dashboardů. Celý tento proces umožňuje transformovat data na informace a znalost, a pomocí vizualizačních nástrojů je poté prezentovat ve snadno srozumitelné formě koncovým uživatelům bez nutné znalosti problematiky BI. (Chaudhuri, 2011)

Níže na obrázku je uveden model výše popsané BI architektury.



Obrázek 3: BI architektura (Chaudhuri 2011 – vlastní zpracování)

Každá vrstva v modelu bude blíže zkoumána v následujících podkapitolách.

3.3.1 Zdrojové systémy

BI systém může získávat data z mnoha různých zdrojů. (Chaudhuri, 2011) U tradičních BI aplikací jsou pro datovou analýzu často hlavním zdrojem strukturovaná data. Nicméně tomu tak vždy není, protože značná část dat získaných ze zdrojových systémů je nestrukturovaná, nebo jen částečně strukturovaná (semistrukturovaná). Podnikové ERP systémy pro plánování podnikových zdrojů jsou tomu dobrým příkladem. V ERP systémech nejsou data strukturována, z důvodu jejich normalizace. Jejich následná analýza je tedy velmi složitá a neefektivní. Proto je důležité zvážit následné využití strukturovaných i semistrukturovaných dat při jejich extrakci do BI systémů, protože oba typy dat mohou poskytovat důležité informace, které jsou potřebné pro přesnou datovou analýzu byznysu. (Baars, 2008)

Data lze rozdělit na externí a interní data. Definice externích a interních dat se liší v závislosti na několika aspektech. Interní data jsou deriváty z datových zdrojů v rámci organizace. Tato data jsou generována v rámci organizace a jsou uložena v provozních databázích. Naopak, externí data jsou generována mimo organizaci, například ve formě zpráv, statistik či reportů. (Inmon, 2002)

Protože jsou data získávána z různých zdrojů, tak se kvalita dat může v mnoha případech lišit. Je to dáno tím, že data mohou být zobrazena různými způsoby, které mohou působit případné komplikace. Další komplikací je například situace, kdy jsou stejné sady dat reprezentovány v různé formě ze stejných datových zdrojů. Díky tomu mohou nastat problémy s integrací dat v druhé části BI modelu, respektive v její transformační části, která je uvedena v následující podkapitole. (Chaudhuri, 2011)

3.3.2 Extrakce, transformace a uložení dat

Účelem extrakce, transformace a uložení (ETL) dat je integrace, čištění a standardizace dat pro jejich následné ukládání do datového skladu. ETL nástroje jsou základem pro fungování úspěšného datového skladu. Hlavním cílem extrakční fáze je zajistit, aby byla data získávána za použití co nejmenších zdrojů s co nejnižším zatížením provozního systému. (Chaudhuri, 2011)

3.3.2.1 Extrakce dat

První dílčí proces ETL se zaměřuje na získávání dat z různých datových zdrojů. Jak je uvedeno v předchozí části, každý zdroj dat může mít svůj vlastní formát nebo strukturu. Získávání dat z různých zdrojů často vede k obrovskému operačnímu výkonu, protože je většinou extrahováno obrovské množství dat. Aby se zamezilo případnému výpadku provozních systémů, dochází k pravidelným extrakcím v době, kdy jsou systémy zatíženy co nejméně (např. v noci), kdy nejsou systémy tolik vyžívány. (Stephandis, 2009)

3.3.2.2 Transformace dat

Druhým krokem v procesu je transformace získaných dat. Proces ETL často vyžaduje mnohostrannou transformaci, kterou lze rozdělit na tři dílčí kategorie; problémy na úrovni schématu, problémy na úrovni záznamu a problémy na úrovni hodnoty. (Stephandis, 2009) Mezi problémy týkající se úrovně schématu patří například situace, kdy mají stejné objekty rozdílné názvy. Druhá úroveň problémů se týká záznamů. Mezi tyto problémy patří např. rozporuplné nebo duplicitní (redundantní) záznamy, stejně jako problémy s datovou konzistencí. Například, míra úrovně detailu neboli granularita dat může být jiná mezi stejnými zdroji. Jeden zdroj může například agregovat data za konkrétní měsíc a je rozporu s jiným datovým zdrojem, který agregaci neobsahuje. Řešení těchto problémů s různými úrovněmi agregací je nezbytné, proto aby byly údaje shodné a použitelné pro další potřebnou analýzu. Poslední kategorií problémů jsou problémy s hodnotou. Jedním z těchto problémů může být např. následující problém. Stát Kalifornie může být vyjádřen pomocí zkratky CA, a ačkoli jsou zkratky jiné, vyjadřují stejnou informaci. Druhým příkladem jsou různé interpretační hodnoty. Názorným příkladem je americký (MM-DD-RRRR) a evropský časový formát (DD-MM-RRRR). (Kimball, 2004)

3.3.2.3 Uložení dat

Třetím a posledním krokem v procesu ETL je načítání dat do cílového zdroje před samotným nahráním dat do datového skladu. Vytváření ETL procesů je technicky i časově náročné vzhledem ke složitosti zdrojových systémů. Navrhování řešení ETL patří ke zdaleka nejnáročnější činnosti v oblasti implementace a údržby datového skladu. (Kimball, 2004)

3.3.3 Datový sklad

Datový sklad, v anglickém jazyce častěji nazýván jako Data Warehouse (DWH), se používá při tradičních BI řešeních, a patří tak mezi hlavní BI komponentu. Jedná se o místo, kde se centralizují a ukládají podnikové informace extrahované z provozních zdrojových systémů organizace. DWH umožňuje sbírat data jak z externích mimo podnikových zdrojů, tak i interních podnikových zdrojů, jak již bylo zmíněno v předchozí části této kapitoly. Primárním cílem datového skladu je podporovat podnikové rozhodovací procesy pomocí strukturovaných dat, které je možné použít pro datovou analýzu díky sofistikovaným a ověřeným analytickým metodám, které oblast BI umí řešit. (Kimball, 2013)

Existuje mnoho definic DWH. Jedna z neznámějších, generelně nepoužívanějších a širokou veřejností nejakceptovanějších je ta, kterou zavedl Bill Inmon. Inmon je jedním z průkopníků v oblasti DWH a je často označován jako "Otec datových skladů". Tento DWH specialista definoval termín DWH následovně:

„Data warehouse is integrated, subject oriented, nonvolatile and time variant collection of data in support of management's decisions.“ (Inmon, 2002)

Volný překlad definice od Inmona: Datový sklad je subjektivě orientovaný, integrovaný, neměnný a časově orientovaný sběr dat určený za účelem podpory manažerského rozhodování.

V definici DW od Inmona (2002) je několik klíčových slov, která je potřeba vysvětlit a vyjasnit si jejich význam.

- „Subjektivě orientovaný“ znamená, že DWH umožňuje analýzu určitých témat v oboru, např. marketingu.
- Druhé klíčové slovo je „integrovaný“. Toto slovní spojení charakterizuje vlastnost, pomocí které může DWH integrovat nebo načítat data z více datových zdrojů, jak je popsáno v předchozí kapitole. Tento aspekt je umožňován pomocí ETL technologie.

- Třetí klíčové slovo „Časově orientovaný“ nám říká, že v DWH ukládá historická data obsahující určitou časovou stopu. Jedná se o jeden z nejzásadnějších rozdílů mezi DWH a produkčními systémy.
- Posledním klíčovým slovem v definici je „časově neměnný“, což znamená, že jakmile se data zanesou do DWH, neměla by být již nikdy modifikována. (Inmon, 2002)

Ralph Kimball (2004), další průkopník v oblasti DW, poskytuje stručnější definici DWH a zdůrazňuje jeho funkčnost.

„A data warehouse is a copy of transaction data specifically structured for query and analysis.“ (Kimball, 2004)

Volný překlad definice: Datový sklad je kopií dat transakcí speciálně strukturovaných pro dotaz a analýza.

Data mohou být uložena v DWH různými způsoby, jedním z nejčastějších a běžně používaných způsobů je modelování dat v DWH díky multidimenzionálnímu modelování. (Inmon, 2002)

3.3.3.1 Dimenzionální modelování

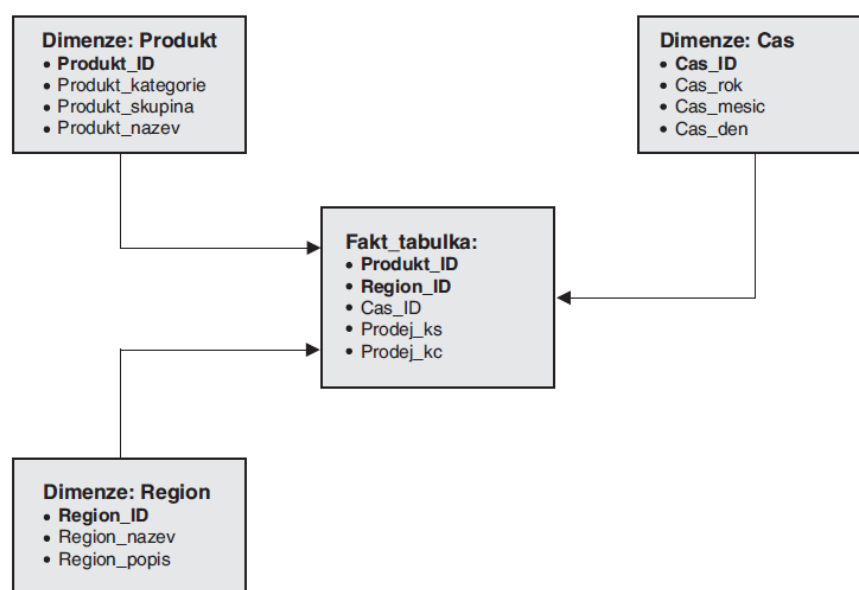
Prostřednictvím dimenzionálního modelování je možné ukládat data do datového skladu v podobě multidimenzionálních modelů. Na rozdíl od klasických ERD databázových modelů, kde jsou data uchovávána v normalizovaném tvaru, jsou data v multidimenzionálním modelu, resp. multidimenzionální databázi denormalizována. Denormalizace je potřebná kvůli žádoucí mnohonásobné redundanci dat, která umožňuje jednodušší a rychlejší analytické dotazování nad datovým skladem. (Kimball, 2013)

3.3.3.1.1 Schémata Star a Snowflake

Dimenzionální model může být reprezentován pomocí schémat Star a Snowflake. Nejčastěji se v praxi využívá realizace prostřednictvím Star schéma díky jeho snadné realizaci a výše zmíněné denormalizaci. Zásadní rozdíl je v tom, že Snowflake schéma obsahuje minimálně jednu normalizovanou tabulku, oproti schématu Star, které obsahuje

pouze denormalizované tabulky. Nicméně lze použít obě schémata, protože jsou obě schémata v principu stejná a rozdělují model na faktové a dimenzionální tabulky, a záleží pouze na specifických obchodních potřebách organizace. Faktové tabulky obsahují sledované metriky, které jsou předmětem analýzy. Mezi sledované metriky můžeme zařadit např. množství produktů nebo celkové tržby za prodané zboží. Je důležité si uvědomit, že metriky jsou pouze číselné ukazatele, které je potřeba specifikovat. K tomu nám slouží dimenze. Dimenzionální tabulky mají popisující charakter a obsahují atributy, které specifikují metriky ve faktových tabulkách podle různých perspektiv a oblastí. Příkladem dimenzionální tabulky může být například časová dimenze s atributy den, měsíc a rok. Dalším příkladem dimenzionální tabulky je dimenze Zákazník obsahující atributy jméno, příjmení, adresa a město. (Kimball, 2013)

Následující obrázek je příkladem hvězdné schématu (Star schéma).

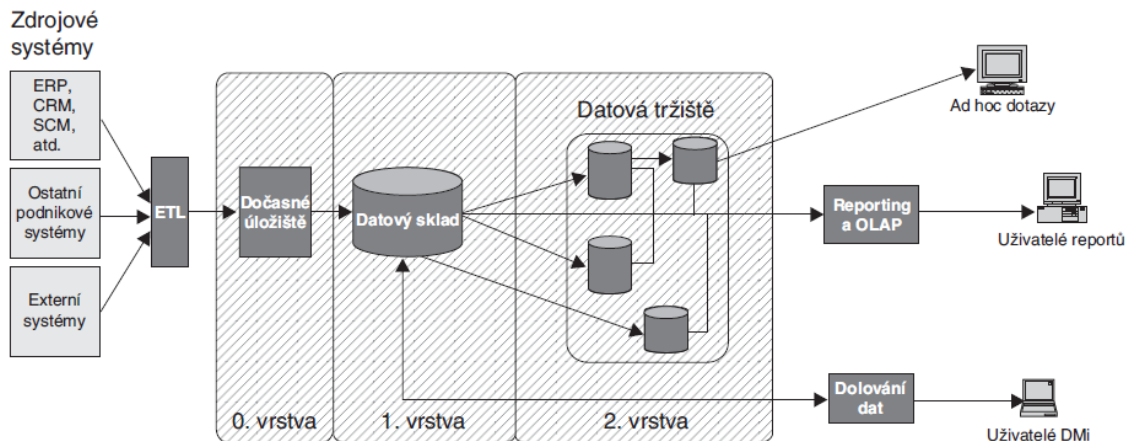


Obrázek 4: Star schéma (Novotný Ota, Pour Jan, Slánský David 2004)

3.3.3.2 Datové tržiště

Datový sklad se může dále členit na tzv. datová tržiště (DM - Data Mart). Princip DM je obdobný jako u DWH. DM můžeme považovat za podmnožinu datového skladu, která je určena pro specifickou podnikovou oblast, jako je například konkrétní oddělení, divize, pobočka nebo závod. DM jsou vytvářeny z důvodu zkrácení doby návratnosti investic, snížení nákladů a redukce rizika při jejich implementaci. (Slánský, 2004)

Příklad architektury BI řešení společně s datovým tržištěm demonstruje obrázek níže.



Obrázek 5: Schéma BI (Novotný Ota, Pour Jan, Slánský David 2004)

3.3.4 Střední analytická vrstva

Střední analytickou vrstvu lze považovat za doplněk datového skladu. Vrstva se také může odborně nazývat Mid-tier. Mezi nejvýznamnější techniku patří Online Analytical Processing (OLAP) – OLAP kostka. Tato technika poskytuje BI pracovníkům specializované funkce a umožňuje jim tak vytvářet řešení pro jejich konkrétní situace a potřeby. (Chaudhuri, 2011)

3.3.4.1 OLAP kostka

OLAP kostka je nástroj pro ukládání multidimenzionálních dat. Hlavním účelem těchto kostek je podpora reportingu. Výkonnost dotazů je díky OLAP kostkám rychlejší a umožňuje tak zpracovávat a následně analyzovat obrovské množství dat, které není možné analyzovat prostřednictvím klasických databází ve formě relačních tabulek. Tvorba OLAP kostek je tedy pro podnikové BI zásadní a určitě by tak jejich využití mělo být bráno v potaz v podnikových zájmech ohledně využití BI. OLAP kostky zlepšují výkon analytických dotazů pomocí předběžných výpočtů přes předem zvolené dimenze a na základě nich jsou poté sledované metriky neboli sledované klíčové ukazatele analyzovány. V praxi to funguje následovně. Když uživatel zadá dotaz, rychle okamžitě načte

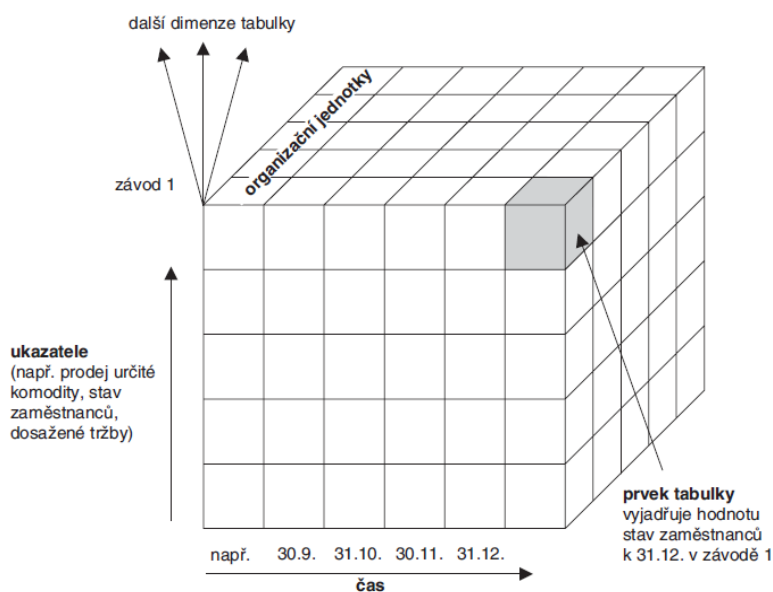
informace z datové struktury a ihned je uživateli prezentuje v požadované formě. (Chaudhuri, 2011)

OLAP kostky jsou postaveny na vícerozměrných neboli multidimenzionálních databázích a pracují se dvěma různými typy dat. Tato data rozdělujeme na metriky a dimenze. (Chaudhuri, 2011)

Metriky jsou množinou hodnot, nejčastěji v číselné formě a jsou uloženy ve faktových tabulkách v datovém skladu. Tyto hodnoty jsou předmětem analýzy a jsou klíčovými ukazateli sledovaného jevu. Běžnými příklady metrik mohou být například metriky prodeje, výdaje nebo objem finančních transakcí. (Chaudhuri, 2011)

Dimenze jsou reprezentovány pomocí dimenzionálních tabulek a slouží k popisu a kategorizaci metrik. Příkladem dimenze může být například dimenze času a trhu. Každá dimenzionální tabulka se skládá z jednotlivých dílčích atributů. Obecně se dá říci, že dimenze mají popisný charakter. Výše zmíněná dimenze času se může rozdělit například na atributy s určitou mírou granularity, jako je rok, kvartál, měsíc či den. Tyto atributy tedy popisují stejnou informaci avšak s rozdílnou úrovní detailu. (Chaudhuri, 2011)

Následující obrázek popisuje příklad jedné možné OLAP kostky.



Obrázek 6: OLAP kostka (Novotný Ota, Pour Jan, Slánský David 2004)

3.3.4.1.1 Operace nad OLAP kostkou

V důsledku struktury a organizace dat uvnitř kostky, můžeme nad kostkou provádět sofistikované dotazy, které jsou následně poskytovány uživateli k zobrazení dat z různých hledisek a perspektiv. (Han, 2012) Nad kostkou můžeme provádět následující operace: drill-down, roll-up, pivoting, slicing a dicing.

- Drill-down: jedná se o navigaci OLAP kostky v hierarchii dimenze směrem k většímu detailu. (Han, 2012)
- Roll-up: umožňuje ve zvolených instancích nastavit vyšší agregační úroveň. (Han, 2012)
- Pivoting: umožňuje „rotovat“ datovou kostku a tím měnit úhel pohledu na data v DWH. (Han, 2012)
- Slicing: operace provádí řezy nad datovou kostkou a umožňuje nalézat pohledy z různých perspektiv. (Han, 2012)
- Dicing: tato operace je inverzí k předchozí operaci. (Han, 2012)

3.3.5 Prezentační vrstva (front-end aplikace)

Mezi hlavní výstupy této vrstvy patří reporty, dashboardy a ad-hoc sestavy, ve kterých mohou uživatelé nalézat cenné informace podložené fakty. Tato vrstva je velmi důležitou částí celého řešení. Poskytuje cenné informace pro strategické, taktické a operativní řízení. Vše ale záleží na úspěšné implementaci předchozích vrstev BI architektury. Bez prezentační vrstvy nezáleží na tom, jak dokonale byl naimplementován datový sklad. Je to dáno tím, že bez této vrstvy. Při implementaci BI řešení máme na úrovni prezentační vrstvy k dispozici hned několik možných variant. Mezi ně patří nejčastěji reporty, dashboardy a ad-hoc sestavy. (Chaudhuri, 2011)

3.3.5.1 Reporty

Reporty jsou využívány v reportingu organizace. Reporting dělíme na následující dva typy – standartní reporting a ad-hoc reporting.

Standartní reporting – reporty jsou generovány v pravidelných časových intervalech na základě nastavených politik. (Chaudhuri, 2011) Ad-hoc reporting – tyto reporty jsou

generovány manuálně na základě jednorázových požadavků, které vytváří uživatele explicitně. (Chaudhuri, 2011)

3.3.5.2 Dashboardy

Další komponentou prezentační vrstvy jsou dashboardy, které představují množinu reportů, které jsou předkládány uživateli takovým způsobem, aby mu poskytoval holistický pohled na klíčové skutečnosti organizace. Dashboard poskytuje přehled o množině klíčových ukazatelů, které mohou být různým způsobem vizualizovány pro jednodušší interpretaci. Podobně jako počítačové palubní desky vozidel zobrazují základní informace potřebné pro sledování aktuálního stavu vozidla. (Chaudhuri et al., 1997) Účelem dashboardu je uspořádat informace takovým způsobem, aby byly poskytovaly přehledné a srozumitelné informace. (Howson, 2014)

3.3.5.3 Ad-hoc sestavy

Ad-hoc sestavy jsou další komponentou pro prezentaci dat. Sestavy umožňují uživatelům BI systému vytvářet jednoduchým způsobem dynamické sestavy, související s obchodními dotazy a reportingovými nástroji. Termín ad-hoc je poněkud zavádějící, neboť jeho význam v českém překladu znamená „spontánní“. To znamená, že jsou tyto sestavy tvořeny pouze na základě jednorázového požadavku, jak už nám napovídá samotný význam slova ad-hoc. Hlavní charakteristikou těchto nástrojů je ten, že sestavy mohou jednoduše vytvářet byznys uživatele a odpadá tak potřeba IT oddělení a jejich zaměstnanců. Ad-hoc sestavy jsou klíčovým faktorem při poskytování BI služeb díky snadnému přístupu k informacím. Je to dáno tím, protože uživatelé prozkoumají data pomocí dotazů spontánně a mohou tak nalézt určité konsekvence v datech, které následně mohou vést k tvorbě standardizovaných reportů a dashboardů. (Howson, 2014)

3.4 Platforma Pentaho

V následující kapitole bude probrána platforma Pentaho BI, pomocí které je realizována praktická část této bakalářské práce.

3.4.1 Historie platformy Pentaho

Platforma Pentaho BI Suite byla vyvinuta americkou společností Pentaho Corporation (se sídlem v Orlandu) v roce 2001. Její název je odvozen od latinského slova „penta“, protože se na jejím vzniku podílelo celkem 5 specialistů, kteří jsou současně také zakladateli této společnosti. Platforma Pentaho je zároveň první BI platformou vůbec, jenž byla vyvinuta a následně poskytnuta široké veřejnosti v podobě open source BI nástrojů s otevřeným zdrojovým kódem. Pentaho je možné použít dvěma možnými způsoby – v podobě otevřeného zdrojového kódu nebo ve formě kompletně vytvořených nástrojů s plnohodnotným grafickým uživatelským rozhraním. (Darrow, 2017)

3.4.2 Verze platformy Pentaho

V roce 2004 Pentaho Corporation rozdělilo svůj primární produkt na dvě verze a obohatila jej kvůli zvýšené poptávce o komerční edici s placenou licencí. Od té doby je tedy Pentaho BI Suite rozdělen na plnohodnotnou open source verzi nazývanou Pentaho Community Edition (Pentaho CE), což je verze s otevřeným zdrojovým kódem, která je popsána v předchozím odstavci, a na verzi komerční s názvem Pentaho Enterprise Edition (Pentaho EE) u které uživatelé a zákazníci platí za její licence. (Darrow, 2017)

Tato práce si neklade za cíl verzi Pentaho EE dále popisovat, protože není její předmětem. Dále se tato práce zabývá výhradně verzí Pentaho CE, díky jejímu využití v praktické části, protože neobsahuje placené licence a je k dispozici zdarma na oficiálních stránkách Pentaho Corporation, kde je možné si ji stáhnout pro osobní účely.

3.4.3 Postavení platformy Pentaho na trhu

Současné (2017) postavení Pentaho na BI trhu je znázorněno na obrázku č. 10, který vizualizuje všeobecně uznávaný Magic Quadrant (magický kvadrant), který každoročně publikuje společnost Gartner. (Sallam, 2017)



Obrázek 7: Magický kvadrant (Gartner 2017)

V roce 2017 se Pentaho umístilo v kvadrantu Niche Players, kde se nachází BI řešení velkých společností, například od společnosti Oracle (Cognos BI). Za toto umístění mohou hlavně nedostatky Pentaha v oblastech Cloud BI a strojového učení (ML – Machine learning), které v současné době dynamicky mění trh BI díky oblasti Industry 4.0. a s ním spojených chytrých technologií (Smart City, Smart Housing). Hodnocení Pentaho a jeho pozici ovlivnili také slabší uživatelská přívětivost jeho zákazníků a analytiků, kteří s BI nástroji pracují. (Sallam, 2017)

Silnou stránkou Pentaho je jeho silná produktová strategie a schopnost porozumět trhu a požadavkům zákazníků. Tyto schopnosti ovlivnily pozici Pentaha v kvadrantu znázorňující vizi. Marketingová divize a strategie společnosti Pentaho se zaměřuje na rozsáhlou analýzu Big Data a v dnešní době na poměrně rozsáhlý fenomén s názvem internet věcí (IoT - Internet of Things). (Sallam, 2017)

V kvadrantu znázorňující schopnost (ability) se odráží relativní umístění Pentaha ve srovnání s ostatními dodavateli v magickém kvadrantu. Dále získal slabší skóre pro kvadrant znázorňující zkušenosti z prodeje. Na závěr lze nicméně konstatovat, že současná schopnost všech dodavatelů BI potažmo jejich schopnost reagovat na trh, jejich výsledky a životaschopnost je u většiny mírně podprůměrná. (Sallam, 2017)

3.4.4 Silné stránky platformy Pentaho

1) Rozsah složitosti a měřítka:

Možnosti platformy Pentaho jsou široké, od integrace dat díky nástroji Pentaho Data Integration, až po pokročilou analýzu dat s možností integrace s nástroji pro Big data. Pentaho zastává přední příčky v horní třetině prodejců ve výše zmíněném magickém kvadrantu pro jeho možnosti při komplexní analýze. To je v souladu se zaměřením společnosti na rozsáhlé nasazení dat na oblast IoT, která potřebuje specializované a pokročilé nástroje při implementaci. (Gartner, 2017), (Underwood, 2017)

2) Dosah dat přes tradiční zdroje:

Pentaho je schopno integrovat a analyzovat tradiční databázová úložiště, ad hoc soubory, NoSQL database, ale dále také nestrukturovaná data (například zdroje sociálních médií, data protokolů a datové toky strojů ze zdrojů IoT). Dva nejdůležitější důvody, proč zákazníci uvádějí, že preferují Pentaho je jeho unikátní přístup k datům a jeho integrace společně s jeho nulovými náklady na licence při výběru verze SE. V případě Big Dat nabízí Pentaho nativní integraci s technologiemi jako Hadoop, Spark, Cassandra a MongoDB, což zákazníci považují za plus. (Gartner, 2017), (Underwood, 2017)

3) Hlavní funkce jsou zaměřené na data:

Z hlediska funkčních schopností Pentaho v oblastech – správy, zabezpečení a architektury, ETL, vestavěné pokročilé analýzy, vestavěný analytický nástrojů a možnostem připojení ke zdrojům dat a správě metadat. (Gartner, 2017)

3.4.5 Výzvy platformy Pentaho

1) Zákaznická zkušenost:

Pentaho čelí neustálému problému v oblasti zkušenosti zákazníků. Záporné ohlasy zákazníků se podepsali na umístění Pentaha ve spodní oblasti magického kvadrantu. Společnost Hitachi s produktem Pentaho se v současné době zaměřuje na zlepšování zákaznické péče a na rozvoj jeho atraktivity. Společnost Pentaho

proto nyní vyčleňuje některé finanční zdroje, kterými nyní disponuje (jako součást Hitachi Data Systems) k řešení těchto otázek. (Gartner, 2017), (Underwood, 2017)

2) Povědomí na přeplněném trhu BI:

Díky tomu, že je světový trh s BI produkty poměrně rozsáhlý, je získání většího povědomí o Pentaho a potažmo jeho nových klientů velmi náročné. (Gartner, 2017)

3) Snadnost použití:

Zákazníci uvádějí, že je ovládání Pentaha poměrně náročné. Platforma Pentaho se proto umístila v dolní části magického kvadrantu, který vizualizuje snadnou administraci nástrojů, jeho vývoj a současně také vizuální atraktivitu ve srovnání s ostatními dodavateli. Zákazníci nejvíce poukazují na obtíže při implementaci platformy a její následné využívání. Dále uvádějí, že při nasazení Pentaha na komplexní datové integrace jim způsobuje značné problémy. Zákazníci Pentaho uvádějí, že nejčastěji hlásí problémy ohledně značného omezení a širšího nasazení při integraci. Dále uvedli, že se setkávají se špatným výkonem, který je způsoben díky větším objemům dat. (Gartner, 2017), (Underwood, 2017)

4) Funkční mezery:

Zatímco je v některých oblastech Pentaho silné, v některých základních schopnostech poměrně zaostává. Nižší hodnocení získává v oblasti Cloud Computing BI a v kategoriích zaměřených na uživatelskou přívětivost, do které patří například podpora sdílení výsledných analýz, které je obtížné sdílet napříč podnikem. Dále chybí schopnost automatického generování statistik a nízký počet jazykových mutací. (Gartner, 2017), (Underwood, 2017)

3.4.6 Nástroje platformy Pentaho

Jak již bylo zmíněno, Pentaho poskytuje širokou škálu nástrojů od integračních a transformačních ETL nástrojů, přes OLAP technologie, reporting, dashboardingu, až po dolování velkých dat. O produkty Pentaho CE se stará velké množství IT specialistů, kteří pomáhají s jejím neustálým zlepšováním a technickou podporou. Pentaho má proto neustálý servisní dohled, což je pro jeho zákazníky určitě výhodné. Sada Pentaho nástrojů

se skládá z několika nástrojů, které lze vzájemně propojit a tvoří tak kompletní plnohodnotný BI systém. (Bouman, 2009)

3.4.6.1 Business Analytics Platform

Hlavním jádrem Pentaha je nástroj Business Analytics Platform (BAP), který pokrývá nástroje celé platformy a umožňuje jejich funkčnost a poskytuje jim serverové prostředí. Dále pomáhá s vykreslováním výsledných reportů, analytických výstupů a dashboardů. Je poskytován podobě plnohodnotného webového uživatelského rozhraní, které je možné do určité míry modifikovat. (Bouman, 2009)

Pentaho BA platforma umožňuje:

- řídit uživatelské účty, role a práva uživatelů a práva k datovým objektům;
- řídit zdroje dat (databázové relační systémy, OLAP technologie apod.);
- nastavovat servery pro automatizaci reportingu;
- designovat a řídit automatizaci BI procesů;
- řídit BI katalogy reportů, analýz a dashboardů;
- modifikovat jazykové mutace. (Bouman, 2009)

Součástí BAP je také centralizované úložiště, které umožňuje jednotlivým nástrojům platformy Pentaho sdílet mezi sebou datové struktury z jednotného úložiště. Centralizované úložiště běží na systému HyperSQL, ale lze jej nahradit systémem Oracle, MySQL či PostgreSQL. (Bouman, 2009)

3.4.6.2 Data Integration

Nástroj, který je k dispozici pro ETL techniky se nazývá Pentaho Data Integration (PDI), také jinak nazývaný jako Kettle či Spoon. PDI má plnohodnotné grafické uživatelské rozhraní. Uživatel zde může dát dohromady všechny druhy ETL, respektive od jednoduchých až po komplexní řešení pro potřebné transformace dat. PDI se skládá z Data Integration Serveru (DI Server) a grafického editoru Spoon. Pomocí serveru v něm mohou být realizovány všechny potřebné extrakce a transformace na úrovni jednotlivých datových struktur. (Bouman, 2009)

PDI je možné napojit na mnoho databázových systémů díky jeho kompatibilitě, která je zajištěna jeho vývojem v jazyce Java. PDI je ho možné využívat na libovolném operačním systému. Dále je možné PDI napojit na příkazový řádek systému pro nastavení automatizace datových přenosů například prostřednictvím BAT souborů. (Bouman, 2009)

Uživatelské rozhraní v PDI je do jisté míry intuitivní a tím dostupné i pro méně zdatné uživatele, kteří nemají velké zkušenosti s ETL nástroji. Nicméně je nutné mít při práci s PDI již předchozí zkušenosti s jinými nástroji. (Bouman, 2009)

V PDI se skládají ELT procesy z těchto stavebních bloků:

Step: Step představuje jeden krok ETL procesu. Existuje mnoho různých nástrojů pro stavy k dispozici v PDI a mnohým uživatelům při jejich práci stačí. Pokud nestačí, může si nové přidat prostřednictvím skriptů, které je možné v PDI využít buď pomocí databázového jazyka SQL nebo skriptovacího jazyka JavaScript. (Bouman, 2009)

Transformace: Transformace je množina jednotlivých stepů. Pro úspěšnou transformaci je potřeba všechny stavy vzájemně propojit a současně nastavit jejich správné vstupní a výstupní parametry. V případě chybných parametrů je uživateli k dispozici chybové protokolování, aby mu pomohla zjistit, co se stalo, pokud se transformace nezdaří. Zde je důležité zmínit, že transformace je procesního charakteru – má start (vstupní soubor, tabulka atd.) a příslušný konec (výstup do souboru). Během transformace se vytváří streamy, které je možné využít pro napojení na další transformace. (Bouman, 2009)

Job: Úloha se skládá z jedné nebo více transformací, které jsou spuštěny v pořadí podle toho, jak je úloha nastavena. Úlohy je možné nastavit plynule z jedné transformace na druhou, nebo jednoduše nastavit jejich sekvenci paralelně. Úlohy lze použít k naplánování úkolů, jako je spuštění všech transformací pro aktualizaci dat datovém skladu na základě předem nastavených scénářů. (Bouman, 2009)

3.4.6.3 Schema Workbench

Nástroj Schema Workbench (SWB) je v platformě pro tvorbu OLAP technologií, resp. pro tvorbu OLAP kostek. OLAP schémata jsou v Pentaho realizována

prostřednictvím XML skriptů. OLAP kostky je možné v WBS vytvářet buď pomocí skriptování v XML jazyce, nebo pomocí plnohodnotného grafického uživatelského rozhraní. Tvorba je ve WBS poměrně intuitivní. Nicméně může méně zdatným uživatelům způsobovat problémy při jeho napojení na BI server, protože neposkytuje protokol chyb. (Bouman, 2009)

Ve WBS uživatel musí nadefinovat datové spojení nad databází datového skladu. Následně v něm nastaví potřebnou strukturu v podobě dimenzionálních a faktových tabulek se sledovanými metrikami. U každé tabulky je potřeba nastavit potřebnou hierarchii, ve které se mají atributy zobrazovat. Celé schéma je poté možné nainportovat do prostředí BI serveru, kde je možné využívat služeb Pentaho Saiku v podobě Self-Service BI. (Bouman, 2009)

3.4.6.4 Report Designer

Pentaho report Designer (PRD) je grafickým rozhráním pro vytváření, úpravu a publikování sestav platformy Pentaho BI. Jednou z hlavních výhod při používání PRD oproti ostatním konkurenčním reportingovým nástrojům je schopnost využívat modely metadat jako zdroje dat. Reporty lze také publikovat přímo na Pentaho BI Server. Tvůrce reportu tak může výsledný report publikovat přímo v uživatelském portálu Pentaha BAP. (Bouman, 2009)

PRD nemá plnohodnotné návrhové rozhraní, takže většinou práce na reportech stráví tvůrce pouze s návrhovou obrazovkou, která zobrazuje strukturu sestavy, a nikoli obsah a konečné uspořádání a vizualizaci. Reporty se vytváří nad datovým skladem buď pomocí SQL dotazů, nebo prostřednictvím MDX dotazů, které jsou pro Pentaho charakteristické. Dále umožňuje vytvářet libovolné statické reporty v podobě agregovatelných tabulek či seznamů, které může doprovázet řada rozmanitých grafů. Reporty lze také generovat v různých dostupných výstupních formátech: PDF, HTML, XLS, RTF a CSV. (Bouman, 2009)

3.4.6.5 Saiku Analytics

Saiku je analytický ad-hoc nástroj pro tvorbu datových souprav. Jeho uživatelské prostředí je intuitivní a ad-hoc dotazy v něm dokáže vytvořit i méně zdatný uživatel i bez potřebné znalosti MDS či SQL jazyka. (Barber, 2017)

Saiku nabízí uživatelsky přívětivé webové řešení pro analýzu dat, které uživatelům umožňuje rychle a snadno analyzovat firemní data a vytvářet a sdílet zprávy. Řešení se připojuje k řadě serverů OLAP, včetně služeb Mondrian, Microsoft Analysis Services, SAP BW a Oracle Hyperion, a lze je rychle a nákladově efektivně nasadit a umožnit tak uživatelům prohlížet data v reálném čase. (Barber, 2017)

V současné době je nejpoužívanějším analytickým nástrojem na platformě Pentaho CE nástroj Saiku Analytics (dále pouze Saiku). Saiku umožňuje tvorbu dynamicky vizualizovaných grafů na základě výsledků analýzy. Mezi funkcemi Saiku chybí možnost tvorby dopočítaných ukazatelů definovaných na uživatelské úrovni a možnost tvořit součtové kumulativní řádky. Tuto možnost je nicméně možné řešit pomocí MDX dotazů, které ovšem vyžadují potřebnou znalost uživatele. (Bouman, 2009)

3.4.6.6 Dashboard Maker

Pentaho Dashboard Maker (PDM) umožňuje vytvářet statické i dynamické dashboards, poskytující okamžité nahlédnutí do celopodnikového výkonu tím, že uživatelé vizualizují klíčové metriky, reporty a analytické sestavy v soudržném a vizuálně bohatém rozhraní. Vytvoření dashboardu v PDM je jednoduché a intuitivní. (Bouman, 2009)

PDM obsahuje:

- Grafy za účelem jednoduché vizualizace dat s možností tvorby dynamických přechodů;
- Datové tabulky;
- URL adresy s konkrétními webovými stránkami namapovanými na Pentaho Server, které chcete uživatel na dashboardu zobrazit. (Bouman, 2009)

Dashboard Designer má také ovládací prvky pro dynamické filtry, které umožňují koncovým uživatelům měnit podrobnosti panelu výběrem různých hodnot z rozevíracího seznamu a tím kontrolovat obsah dashboardu. (Bouman, 2009)

4 Vlastní práce

4.1 Úvod do praktické části

Cílem praktické části bakalářské práce je demonstrovat ukázkou aplikace jednotlivých nástrojů platformy Pentaho CE na veřejně dostupné datové sadě Northwind, která obsahuje data fiktivní společnosti Northwind Traders. Cíl práce je realizován prostřednictvím identifikace klíčových metrik, konstrukce datového skladu nad databází Northwind, extrakcí a transformací dat do nově vytvořeného datového skladu, tvorbou OLAP kostky, analýzou dat nad datovým skladem a následnou prezentací analyzovaných dat. Výstupem praktické části jsou reporty a ad-hoc Self-service sestavy, které vzniknou na základě analýzy dat obsažených v databázi Northwind. Výstupy budou srozumitelné tak, aby mohli vedení společnosti a jeho managementu dávat cenné informace.

4.2 Databáze Northwind

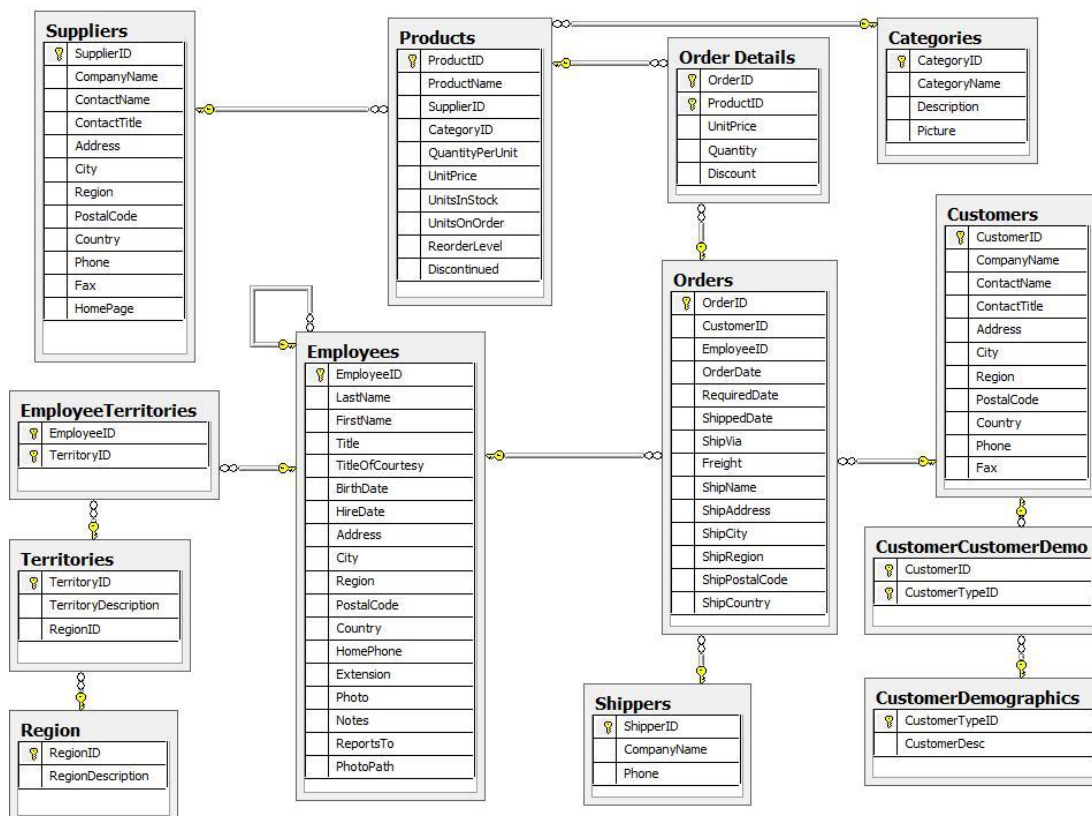
Pro realizaci praktické části byla vybrána volně distribuovaná datová sada v podobě databáze Northwind. Tato databáze je volně ke stažení na stránkách společnosti Microsoft. Společnost Microsoft je zároveň tvůrcem této datové sady.

4.2.1 Vlastnosti databáze Northwind

Databáze Northwind představuje podniková obchodní data fiktivní společnosti Northwind Traders. Předmětem podnikání této společnosti je dovoz a vývoz potravinářských produktů po celém světě. Atributy databáze jsou záznamy ze všech jejich obchodních transakcí, které reflektují prodejní a nákupní transakce mezi dodavateli a zákazníky společnosti Northwind Traders.

V datové sadě je celkem 13 tabulek. Níže je uvedený relační databázový model společnosti Northwind Traders. Model zobrazuje jednotlivé entity a jejich vzájemné vazby,

a tím umožňuje získat lepší přehled o struktuře uložených dat. Databázový relační model vypadá následovně:



Obrázek 8: Relační model databáze Northwind (Microsoft 2010)

4.3 Návrh řešení pro společnost Northwind Traders

4.3.1 Současný stav

Současný systém společnosti Northwind Traders neumožňuje sledovat a analyzovat klíčové metriky, které by vedení společnosti pomohli při jejím obchodním rozhodování.

4.3.2 Návrh metrik

Na základě této skutečnosti se pro společnost Northwind Traders nabízí možnost naimplementovat BI řešení, které bude obsahovat data z jejich relačního databázového modelu. Protože společnost postrádá manažerský reporting, který by ji pomohl sledovat její efektivitu, jeví se jako možné metriky pro pozdější klíčové ukazatele výkonnosti (KPI) následující metriky:

- Objemy tržeb v dolarech (v jednotkách \$);
- Množství objednávek v kusech (v jednotkách ks).

Podle objemu tržeb, které mohou být v datovém skladu agregovány podle určité míry granularity, může společnost sledovat například:

- měsíční a roční tržby;
- zisk za jednotlivé objednávky
- tržby podle regionů;
- produkty, které společnosti vytváří nejvyšší zisk.

Podle druhé metriky zachycující množství objednávek v kusech může společnost sledovat například:

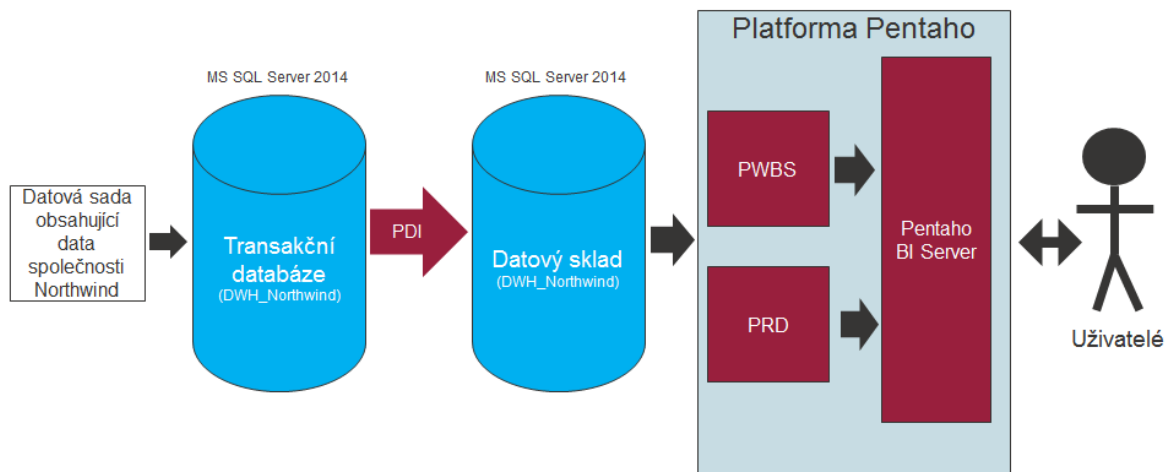
- měsíční a roční množství objednávek podle regionu;
- kteří zákazníci si kupují nejvíce produktů a tím je zařadit např. do sekce VIP zákazníků;
- zaměstnance, kteří prodají nejvíce zboží.

Platforma Pentaho Community Edition se jevila jako vhodná platforma, protože je volně dostupná a obsahuje nástroje, které dohromady umožňují vytvořit ucelené BI řešení. Verze Pentaho Community Edition navíc představuje nulové náklady na licence. Jediným nákladem jsou počáteční náklady na implementaci.

4.3.3 Architektura řešení

Před samotným začátkem budování datového skladu a jeho realizace pomocí nástrojů platformy Pentaho bylo nutné specifikovat a navrhnout architekturu celého řešení.

Schéma níže ukazuje architekturu celého řešení, která je vizualizována prostřednictvím nástroje MS PowerPoint 2016.



Zkratky

- **PDI** – Pentaho Data Integration (extrakční, transformační a integrační nástroj)
- **PWBS** – Pentaho Workbench Schema (nástroj pro tvorbu analytických OLAP kostek)
- **PRD** – Pentaho Report Designer (nástroj pro tvorbu reportingových sestav)
- **Pentaho BI Server** – webové rozhraní umožňující uživatelům vytvářet ad-hoc sestavy nad OLAP kostkou a nahlížet na vygenerované reportingové sestavy z PRD

Obrázek 9: Architektura řešení (vlastní zpracování)

První objekt schématu představuje datovou sadu společnosti Northwind, která bude nejdříve naimportována do transakčního databázového systému MS SQL Server 2014. Poté, když budou k dispozici data v transakční databázi, bude následovat tvorba fyzického datového skladu, do kterého budou později extrahována a transformována data pomocí integračního nástroje **Pentaho Data Integration**. Jakmile bude datový sklad naplněn daty v požadovaném formátu, bude vytvořena OLAP kostka nad datovým skladem prostřednictvím nástroje **Schema Workbench**, která poté bude naimportována do webového prostředí nástroje **Pentaho BI Server**, ve kterém bude mít management společnosti možnost vytvářet si své vlastní ad-hoc sestavy díky nástroji **Pentaho Saiku Analytics**. Posledním nástrojem celého řešení je **Pentaho Report Designer**, pomocí něhož budou vytvořeny reportingové sestavy.

4.4 Budování datového skladu

Tato kapitola řeší návrh datového skladu společnosti Northwind Traders a je rozdělena do následujících podkapitol:

1. Import datové sady Northwind do MS SQL Serveru 2014
2. Návrh datového skladu (databázový model)

3. Vytvoření datového skladu (fyzické vytvoření v databázovém systému)

4.4.1 Import datové sady do MS SQL Serveru

Aby bylo možné nad data společnosti vytvořit datový sklad, musel být nejdříve naimportován datový balíček Northwind do databázového systému. Pro účely praktické části byl vybrán databázový systém on společnosti Microsoft s názvem MS SQL Server 2014. MS SQL Server je v celé bakalářské práci využíván ve verzi express, kterou je možné stáhnout na oficiálních stránkách společnosti Microsoft zcela zdarma.

4.4.2 Návrh datového skladu

Poté, když byla data, která představují transakční databázi společnosti, naimportována do databázového systému MS SQL Serveru, bylo potřeba navrhnout datový sklad, který by společnosti umožňovat sledovat výše zmíněné metriky, jenž byly stanoveny v předchozí kapitole.

Aby bylo možné navrhnout datový sklad, bylo potřeba provést tyto následující kroky:

1. Vybrat vhodné dimenzionální tabulky
2. Vybrat vhodné faktové tabulky
3. Vytvořit fyzický datový sklad v databázovém systému

Všechny výše zmíněné prerekvizity, které jsou nezbytné pro vybudování datového skladu, musely být provedeny v souladu s byznys potřebou společnosti Northwind Traders tak, aby tato služba vytvářela přidanou hodnotu jejím koncovým zákazníkům, resp. vedení a managementu společnosti.

4.4.2.1 Výběr dimenzionálních tabulek

Aby společnost Northwind Traders mohla sledovat výše zmíněné metriky výkonnosti, musel být vytvořen datový sklad obsahující data z relačního modelu jejich IS, který splňuje všechny potřebné náležitosti datového skladu, které byly zmíněné v teoretické části práce.

Na základě výše zmíněných požadavků, se jeví jako vhodné dimenze následující dimenzionální tabulky:

Dimenze zákazníků s názvem D_Customers

Jako konkrétní atributy této dimenze byly stanoveny tyto atributy - CustomerID, CompanyName, ContactName, ContactTitle, Address, City, Region, PostalCode, Country, Phone, Fax.

Dimenze zaměstnanců s názvem D_Employees

Jako konkrétní atributy této dimenze byly stanoveny tyto atributy - EmployeeID, FullName (tento atribut představuje sloučení jména a příjmení zaměstnance do jednoho atributu).

Dimenze produktů s názvem D_Products

Jako konkrétní atributy této dimenze byly stanoveny následující atributy - ProductID, ProductName.

Dále bylo potřebné vytvořit degenerované dimenze, které budou součástí faktové tabulky. Tyto degenerované dimenze jsou blíže popsány v následující podkapitole při výběru faktové tabulky datového skladu.

4.4.2.2 Výběr faktové tabulky

Pro model datového skladu společnosti Northwind Traders postačila tvorba pouze jedné faktové tabulky, která nicméně plně splňovala požadavky, které byly stanoveny v předchozí kapitole, jenž se zabývala návrhem metrik. Řešení obsahuje pouze jednu faktovou tabulku, protože předmětem výsledných analytických výstupů budou výhradně informace týkající se prodeje. Pro faktovou tabulku, reflektující stejnojmennou konvenci výše zmíněných dimenzionálních tabulek, byl zvolen název F-Orders.

Faktová tabulka obsahuje primárně číselné atributy o objednávkách. Z tohoto byly stanoveny následující atributy - OrderID, ProductID, UnitPrice, Quantity, EmployeeID, CustomerID, OrderDate, ShipCountry.

Poslední dva zmíněné atributy (OrderDate a ShipCountry) představují degenerované dimenze, protože obsahují pouze jeden popisný atribut (více v tomto případě není žádoucí) a jsou uloženy ve faktové tabulce.

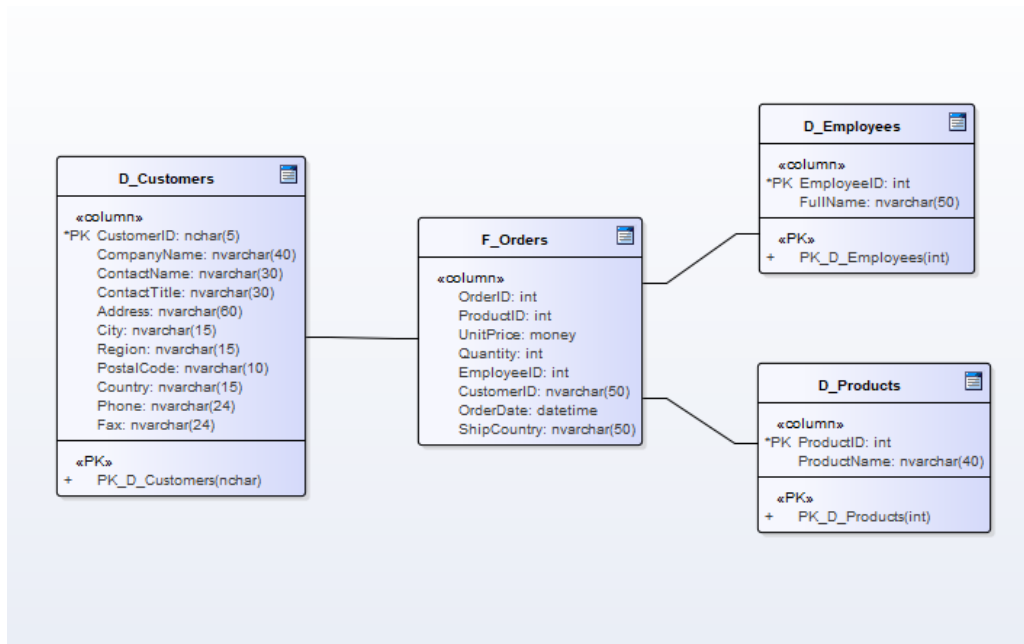
4.4.2.3 Vytvoření datového skladu v MS SQL Serveru

Poté, když byla specifikována struktura datového skladu, bylo nutné vytvořit výše zmíněné tabulky (faktová tabulka společně s dimenzionálními tabulkami), aby je bylo možné následně naplnit daty pomocí integračního ETL nástroje Pentaho Data Integration.

Pro vytvoření fyzického datového skladu bylo nutné nejdříve vytvořit DDL (Data Definition Language) skripty, pomocí kterých bylo možné postupně vytvořit všechny potřebné tabulky datového skladu. Jednotlivé DDL skripty jsou uvedeny v příloze bakalářské práce.

- Pro dimenzionální tabulku zaměstnanců (D_Employess) – příloha č. 1.
- Pro dimenzionální tabulku produktů (D_Products) – příloha č. 2.
- Pro dimenzionální tabulku zákazníků (D_Customers) – příloha č. 3.
- Pro faktovou tabulku prodejů (F_Orders) – příloha č. 4.

Jakmile byly v specifikovány výše popsané tabulky, bylo nutné je zanást do databázového modelu, který vizualizuje přehledným způsobem strukturu datového skladu. Databázový diagram datového skladu s názvem Northwind_DWH vypadá následovně:



Obrázek 10: Model datového skladu (vlastní zpracování)

Faktová tabulka obsahuje metriky Quantity (množství) a UnitPrice (cena za jednotku zboží). Dále obsahuje atributy, které představují degenerované dimenze. Těmito atributy jsou ShipCountry (země kam je zboží doručeno) a OrderDate (datum objednávky). Dále faktová tabulka obsahuje množinu cizích klíčů (FK), pomocí nichž bylo realizováno propojení s dimenzionální tabulky.

Dimenzionální tabulky obsahují primární klíče, které jsou napojeny na cizí klíče faktové tabulky. Zároveň je nutné dodat, že kardinalita mezi dimenzionální a faktovou tabulkou má vždy kardinalitu typu 1:N (právě jednomu záznamu dimenzionální tabulky odpovídá více záznamů z faktové tabulky). Dimenzionální tabulky obsahují také popisné atributy, které specifikují vždy danou dimenzi a tím pomáhají sledované metriky blíže specifikovat a nahlížet na ně z různých perspektiv.

Výše zmíněný datový model datového skladu Northwind_DWH byl vytvořen v programu Enterprise Architect, který představuje komplexní nástroj pro systémovou analýzu. Modelování v Enterprise Architect má své výhody, protože uživateli může tento program vygenerovat automaticky DDL skripty, čímž odpadá pracnost s manuální tvorbou těchto skriptů. Další výhodou automatického generování DDL skriptů je minimalizace

možných chyb. V této práci byly nicméně DDL skripty, které reprezentující tabulky v datovém skladě, vytvořené manuálně v prostředí databázového systému MS SQL Server.

4.5 Extrakce, transformace a uložení dat do datového skladu

Poté, když byl vytvořen datový sklad společnosti Northwind Traders, bylo nutné tento sklad naplnit daty z produkčního systému, jenž byla na začátku této kapitoly naimportována do prostředí databázového systému MS SQL Serveru 2014.

4.5.1 Vytvoření ETL transformace v Pentaho Data Integration

Pro extrakci a transformaci dat do datového skladu byl použit integrační nástroj Pentaho Data Integration (PDI) z platformy Pentaho, ve kterém byly vytvořeny ETL transformace.

Aby bylo možné provést datovou pumpu ze zdrojové databáze do datového skladu, musel být v prostředí nástroje PDI provedeny následující kroky.

1. Vytvoření ETL transformací;
2. Vytvoření databázového připojení na vstupní i výstupní databázi;
3. Namapování zdrojových atributů na atributy uložené v nově vytvořeném datovém skladě.

První krok byl realizován tak, že byly v prostředí PDI sestaveny všechny potřebné ETL transformace pro všechny nově vzniklé čtyři tabulky v datovém skladě. Tyto transformace vypadaly následovně:

Pro faktovou tabulku objednávek F_Orders:



Obrázek 11: ETL transformace pro F_Orders (vlastní zpracování)

Pro dimenzionální tabulku zákazníků D_Customers:



Obrázek 12: ETL transformace pro D_Customers (vlastní zpracování)

Pro dimenzionální tabulku zaměstnanců D_Employees:



Obrázek 13: ETL transformace pro D_Employees (vlastní zpracování)

Pro dimenzionální tabulku produktů D_Products:



Obrázek 14: ETL transformace pro D_Products (vlastní zpracování)

Transformace jsou triviální a jejich kardinalita je typu 1:1 (pro jeden datový vstup je přiřazen právě jeden datový výstup), protože data není potřeba nijak speciálně transformovat do výsledného formátu, ve kterém budou data uložena v datovém skladě. Smyslem ETL transformací bylo data uložit do potřebné datové struktury, respektive do schématu Star, které bylo navrhuto, namodelováno a později fyzicky vytvořeno v databázovém prostředí MS SQL Server 2014.

Poté, když byly vytvořeny jednotlivé ETL transformace pro datové vstupy ze zdrojové tabulky databáze Northwind do datového skladu, bylo nutné nastavit databázové připojení nástroje Pentaho Data Integration k MS SQL Serveru. Pro step ETL transformace pro datový výstup, bylo zároveň nutné nastavit tzv. datový mapping. Datový mapping umožňuje tvůrci ETL transformace přiřadit jednotlivé atributy zdrojové tabulky k atributům cílové tabulky. V tomto případě byly v datovém mappingu namapovány atributy, které byly zvoleny ve stepu pro datový vstup transformace. Aby bylo možné celou

transformaci provést, museli být oba předchozí stepy propojeny, aby mohl být proveden datový tok ze zdrojové databáze dbo.Northwind do cílového datového skladu s názvem dbo.Northwind_DWH.

Následně, když byly všechny patřičné parametry nastaveny, bylo potřeba tyto transformace spustit. V nástroji PDI je k dispozici informativní protokolování chyb, které umožňuje identifikovat případné chyby při extrakci, transformaci a ukládání dat.

Když byl datový sklad Northwind_DWH pomocí nástroje Data Integration naplněn daty, bylo potřeba tyto transformace patřičně automatizovat, aby měla společnost Northwind Traders v datovém skladě k dispozici neustále aktuální data, a nemusela tyto transformace v prostředí nástroje Data Integration opakovaně manuálně spouštět. Z tohoto důvodu musel být chod transformací patřičným způsobem kontinuálně zajištěn prostřednictvím automatizace ETL procesů.

4.5.2 Automatizace ETL transformací

Problém s automatizací ETL transformací byl vyřešen prostřednictvím dávkového souboru systému Windows. Aby bylo možné automatizovat proces, který automatizaci ETL transformací provádí, musely být v nástroji Data Integration všechny realizované ETL transformace naimportovány do tzv. Jobu. Job se skládá ze zanořených výše realizovaných ETL transformací, a strukturuje tak tyto transformace do jediné sekvence, kterou je možné prostřednictvím doplňkového nástroje Kitchen v Pentaho Data Integration a Plánovače úloh ve Windows automaticky spouštět podle předem nastavených pravidel přes příkazový řádek operačního systému Windows.

Níže je uveden sekvenční Job, které sjednocuje všechny výše zmíněné ETL transformace pro tabulky uložené v datovém skladě DWH_Northwind.



Obrázek 15: Job ETL transformace (vlastní zpracování)

Dalším krokem k úspěšné automatizaci bylo nutné vytvořit dávkový soubor BAT, který spouští výše zmíněný sekvenční ETL Job s názvem “main_job.kjb“ v nástroji Pentaho Data Integration. Skript obsažený v BAT souboru byl realizován přes příkazový řádek a jeho syntaxe je následující:

```
C:\WINDOWS\system32>call kitchen.bat /file:C:\Users\olazn\Desktop\BP\main_job.kjb
```

Obrázek 16: Skript BAT souboru (vlastní zpracování)

Plánovač úloh spouští soubor „main_job.kjb“ každý den ve 22 hodin a 50 minut a přenáší data pomocí datové pumpy, která byla nastavena v nástroji Data Integration do datového skladu na denní bázi tak, aby mohli vedoucí pracovníci společnosti Northwind sledovat svou obchodní činnost díky aktuálním informacím uložených v datovém skladě.

Je potřeba zmínit, že výše uvedené řešení, při kterém byla provedena automatizace ETL procesu, je možné řešit pomocí integrace nástrojů Pentaho Data Integration a Pentaho BI serveru. Touto funkcionalitou disponuje pouze komerční verze Pentaho Enterprise Edition.

4.6 OLAP kostka

4.6.1 Tvorba OLAP kostky

Pro tvorbu samoobslužných analytických ad-hoc sestav, bylo potřeba vytvořit OLAP kostku, která umožňuje vedení společnosti vytvářet své vlastní Self-service BI sestavy. Tuto kostku bylo proto potřeba vytvořit v nástroji Schema Workbench (WBS) a následně tuto ji nahrát na BI server, který je v platformě Pentaho k dispozici ve formě webového rozhraní.

V nástroji WBS bylo nutné vytvořit OLAP kostku, která bude umožňovat managementu společnosti Northwind náhled na objemy tržeb podle dimenzí zákazníků, produktů a zaměstnanců pomocí Self-Service ad-hoc nástroje Saiku Analytics. Aby bylo možné toto cíle realizovat, musela být nejdříve v nástroji WBS vytvořena OLAP kostka – viz příloha č. 5. Pomocí nástroje WBS bylo nutné vytvořit OLAP kostku v podobě XML

souboru, kterou bylo nutné následně nahrát na Pentaho BI Server, prostřednictvím kterého bude mít management společnosti možnost vytvářet si své vlastní ad-hoc sestavy libovolně bez potřeby hlubší znalosti BI.

Pro realizaci OLAP kostky v nástroji WBS, muselo být nejdříve provedeno propojení nástroje WBS s databázovým systémem MS SQL Server 2014. Propojení bylo obdobného typu, jako u nástroje PDI. Poté co bylo toto propojení úspěšně provedeno, musela být namapována faktová tabulka a dimenzionální tabulky na jednotlivé atributy datového skladu. V případě sledované metriky Objemy tržeb v dolarech bylo potřeba si metriku nejdříve spočítat nad datovým skladem prostřednictvím dotazovacího jazyka DMX, protože nástroj WBS neumožňuje dopočítávat metriky, jak bylo zmíněné v teoretické část této práce.

Výstupem programu WBS byl XML soubor, který byl následně nainportován a publikován na Pentaho BI Server, do kterého bylo potřeba nainstalovat dílčí plugin pro Saiku Analytics, který umožňuje využívat ad-hoc sestavy.

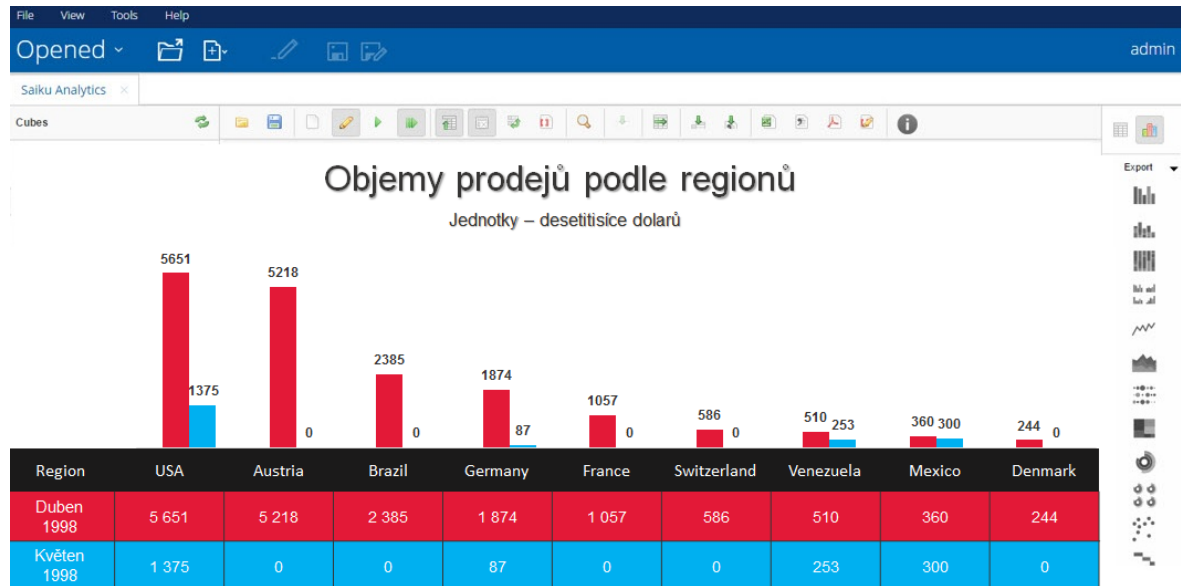
Zároveň je nutné zmínit, že výše uvedené řešení provedené v nástroji WBS bylo realizováno pouze pro metriku sledující objemy tržeb. Pokud by společnost potřebovala přidat i druhou metriku, kterou je množství objednávek, bude potřeba opět vytvořit novou OLAP kostku pro tuto metriku v nástroji WBS.

4.6.2 Operace nad OLAP kostkou

Poté, když byla OLAP kostka nainportována v podobě XML souboru na Pentaho BI Server, bylo možné vytvářet ad-hoc sestavy. Pro úspěšné nahrání OLAP kostky na BI Server, bylo nutné stáhnout do Pentaho BI Serveru plugin, který umožňuje uživatelům využívat nastavení serveru v podobě nástroje Pentaho Saiku Analytics. V nástroji Saiku si uživatelé mohou vytvořit jednoduché ad-hoc sestavy nad danou OLAP kostkou.

Pro potřeby managementu byla do prostředí Pentaho BI Serveru potažmo Saiku Analytics nainportována OLAP kostka, která vypočítává objemy prodeje v dolarech. Nad touto kostkou je možné provádět sestavy, které sledují objemy tržeb podle metrik produktů, zákazníků a zaměstnanců.

Níže je ukázka výsledného ad-hoc dotazu pro objemy tržeb agregovaných podle regionu, který zároveň vizualizuje data za poslední 2 měsíce, resp. duben a květen 1998).



Obrázek 17: Ad-hoc sestava pro objemy prodejů (vlastní zpracování)

4.7 Reporting

Pro tvorbu reportingových sestav slouží v platformě Pentaho nástroj Pentaho Report Designer (RD). Aby bylo možné vytvořit reporty, bylo nejdříve nutné nastavit databázové připojení, které umožnilo nástroji Pentaho Data Integration čerpat data přímo z datového skladu DWH_Northwind.

Podle definovaných metrik, které byly specifikovány v úvodní části praktické části, bylo vytvořeno následujících pět reportů.

Report č. 1 - TOP 25: nejprodávanější zboží (v tis. ks)

Tento report zobrazuje 25 celosvětově nejprodávanějších výrobků společnosti Northwind Traders za poslední aktuální období (květen 1998), které je v databázi dostupné – viz. Příloha č. 6.

Report č. 2 (sub-report) - TOP 5: nejprodávanější zboží podle regionu (v tis. ks) a TOP 10: nejprodávanější zboží (v tis. ks) v regionu USA

Tento report je rozdělen na dva menší reporty. První sub-report zobrazuje 5 nejprodávanějších produktů podle regionu a druhý sub-report zobrazuje 10 nejprodávanějších produktů v USA. Oba sub-reporty obsahují název konkrétního produktu a počet prodejů v kusech – viz. Příloha č. 7.

Report č. 3 - Objemy tržeb (v tis. \$) podle zaměstnanců

Tento report zobrazuje objemy tržeb v tisících dolarů, za poslední čtyři měsíce (únor až květen 1998). Na reportu je možné vidět všechny obchodníky společnosti Northwind Traders, kteří jsou seřazeni abecedně. Tento report může společnosti Northwind Traders sloužit jako ukazatel výkonnosti jednotlivých zaměstnanců – viz. Příloha č. 8.

Report č. 4 - Objemy prodejů (v tis. \$) podle jednotlivých regionů

Tento report vizualizuje peněžní objemy prodejů podle jednotlivých regionů, do kterých společnost Northwind Traders dodává zboží. Report obsahuje tabulkovou i grafickou vizualizaci – viz. Příloha č. 9.

Report č. 5 - Tržby (v tis. \$) a objemy (v tis. ks) za posledních 23 měsíců

Poslední reportingová sestava vizualizuje tržby (v tis. \$) a objemy prodejů (v tis. ks) za posledních 23 měsíců. Aby byla vizualizace pro manažery přehlednější, jsou v reportu uvedeny 2 spojnicové grafy, které vizualizují trendy prodejů v čase. První graf vizualizuje peněžní tržby a druhý graf množství prodaných kusů – viz. Příloha č. 10.

Všechny výše uvedené reportingové sestavy byly nastaveny tak, aby je bylo možné generovat automaticky jednou měsíčně na základě aktuálních dat uložených v datovém skladě. Tato automatizace byla provedena prostřednictvím nástroje PRD, ve spolupráci s nástrojem Pentaho BI Server, který disponuje automatizací reportingových sestav. Nástroj PRD touto funkcí disponuje, na rozdíl od nástroje Pentaho Data Integration, ve kterém byla automatizace ETL transformací řešena prostřednictvím Plánovače úloh a příkazového řádku systému Windows, na základě absence této funkčnosti. Díky tomuto

automatizovanému řešení, umožňující automatické generování reportingových soustav, má společnost Northwind Traders k dispozici vždy aktuální reporty, které jsou generovány na měsíční bázi.

5 Výsledky a diskuse

Open source platforma Pentaho umožňuje implementovat BI řešení v celém kontextu, respektive od extrakce a transformace dat, přes datovou analýzu nad datovým skladem ve formě OLAP kostek, které je možné využívat při analytických ad-hoc sestavách, až po tvorbu reportů a dashboardů, které mohou pomáhat řídicím pracovníkům a managementu společnosti při jejich podnikovém rozhodování.

Výsledkem bakalářské práce jsou reportingové a ad-hoc self-service sestavy, které umožňují společnosti Northwind Traders vytvářet libovolné jednoduché ad-hoc analýzy bez nutnosti znalosti problematiky BI. Na základě analýzy relačního datového modelu společnosti, byly v rámci řešení identifikovány a definovány dvě klíčové metriky, jenž představují objemy prodejů v dolarech a množství objednávek v kusech, které je možné sledovat z různých perspektiv podle dimenzí produktů, zákazníků, zaměstnanců, regionů a časového období.

Všechny uvedené reportingové a ad-hoc self-service sestavy byly plně automatizovány na denní bázi, a díky tomu má vedení a management společnosti k dispozici aktuální obchodní informace. Automatizace byla provedena prostřednictvím systému Windows pomocí dávkového souboru BAT spouštěného formou příkazového řádku. V tomto případě se jedná o úskalí open source verze Pentaho, protože nedisponuje možností plně automatizovaného řešení prostřednictvím platformy Pentaho. Touto možností disponuje pouze komerční verze platformy Pentaho. Další úskalím práce je fakt, že každá další modifikace vyžaduje odbornou práci IT odborníka se znalostí BI, protože open source verze oproti komerční verzi vyžaduje určitou znalost této problematiky.

Implementace uvedená v praktické části představuje pouze investiční náklady spojené výhradně s náklady spojenými s implementací, které obsahují lidskou činnost. Odpadají tedy náklady zahrnující cenu licencí.

Implementované řešení umožňuje společnosti sledovat základní ekonomické ukazatele reflektující jejich obchodní činnost. Řešení je nicméně možné časem rozšířit o další funkčnosti. Pokud by se řešení ukázalo pro společnost jako velmi výhodné, a

společnost by chtěla BI řešení rozšířit o pokročilejší technologie, nastává možnost zvážení investice do komerční verze Pentaho, která disponuje širší škálou funkcí, které by mohla společnost Northwind Traders později chtít využívat.

6 Závěr

Bakalářská práce se zabývala implementací open source platformy s názvem Pentaho, která představuje plnohodnotnou sadu BI softwarových nástrojů. Hlavním cílem této práce a zároveň cílem praktické části bylo ukázat jednotlivé nástroje této open source platformy, které je možné využít při implementaci BI řešení.

Aby bylo možné realizovat implementaci v praktické části, bylo nutné nejdříve vysvětlit podstatu manažerského rozhodování, které je nezbytné pro úspěšné přežití podniků na dnešním vysoce konkurenčním trhu. Poté byla vysvětlena problematika BI. Kapitola zabývající se Business Intelligence poskytla ucelený a srozumitelný výklad od úvodu do této problematiky, přes její definici, procesu transformace dat na znalost a moudrost, klíčových složek, vlivu BI na celý podnik, dopady na rozhodovací procesy, až po faktory, které ovlivňují úspěšné přijetí těchto systémů do prostředí organizace. Další kapitola se zabývala architekturou BI od zdrojových systémů, extrakčních a transformačních nástrojů, přes jádro datového skladu a datových tržišť, dále následovala střední analytická vrstva, jež se skládá z OLAP technologií, až po finální prezentační vrstvu, která slouží koncovým uživatelům a konzumentům analytických výstupů. Ve finální prezentační vrstvě byl vysvětlen princip reportingu, dashboardingu a ad-hoc self-service analytických sestav. Poslední kapitola literární rešerše se zabývala samotnou platformou Pentaho, pomocí které je realizována praktická část práce. V této kapitole byla popsána historie platformy, její komerční a open source verze, technická specifikace, současná situace a současné postavení na světovém BI trhu prostřednictvím analýzy magického kvadrantu od společnosti Gartner. Dále byly rozebrány výhody a výzvy této platformy. V závěru kapitoly byly představeny a následně rozebrány jednotlivé nástroje open source verze, pomocí nichž byla následně realizováno vlastní řešení této bakalářské práce.

Hlavním cílem práce, jak již bylo zmíněno, bylo ukázat jednotlivé nástroje open source platformy Pentaho verze Community Edition. Tento cíl byl realizován prostřednictvím implementace vybraných dílčích nástrojů platformy na volně dostupné datové sadě relační databáze fiktivní společnosti Northwind Traders, za účelem vytvoření uceleného BI řešení pro podporu manažerského reportingu této společnosti. Aby bylo

možné tento hlavní cíl realizovat, bylo nutné nejdříve navrhnout datový sklad, který by nejlépe reflektoval byznys potřeby společnosti. Z tohoto důvodu bylo nutné identifikovat klíčové metriky, prostřednictvím nichž mohlo být celé řešení realizováno. Na základě identifikovaných klíčových ekonomických metrik byla vybrána faktová tabulka následovaná dimenzionálními tabulkami. Poté, když byly výše zmíněné metriky, fakta a dimenze zakomponovány do uceleného řešení, bylo možné navrhnout model datového skladu a ten poté následně fyzicky vytvořit v prostředí databázového prostředí v MS SQL Serveru. Jakmile byl datový sklad vytvořen, byl následně naplněn daty pomocí extrakčního a transformačního nástroje platformy Pentaho s názvem Pentaho Data Integration. Všechny vytvořené ETL transformace byly následně automatizovány, z toho důvodu, aby mělo vedení společnosti k dispozici vždy aktuální a konzistentní informace. ETL transformace byly automatizovány na denní bázi. Poté, když byly v datovém skladě k dispozici aktuální data, bylo možné je následně analyzovat. Nad daty byla následně vytvořena analytická OLAP kostka prostřednictvím nástroje Pentaho WBS, která byla poté publikována v nástroji Pentaho BI Server, ve kterém si mohou řídicí pracovníci společnosti vytvářet ad-hoc self-service analytické sestavy. Na závěr praktické části byla vytvořena sada reportingových sestav pro podporu manažerského rozhodování prostřednictvím nástroje PRD.

7 Seznam použitých zdrojů

YOON, Tom E. *User Acceptance of Business Intelligence (BI) Application: Technology, Individual Difference, Social Influence, and Situational Constraints*. 2014. Waikoloa, HI, USA: IEEE, 2014. ISBN 978-1-4799-2504-9.

DAVID H. MAISTER, CHARLES H. GREEN a ROBERT M. GALFORD. *The trusted advisor*. New York: Touchstone, 2001. ISBN 0743212347.

WEILL, Peter. *The Relationship Between Investment in Information Technology and Firm Performance: A Study of the Valve Manufacturing Sector*. Melbourne, 1992. Studie. University of Melbourne.

HOU, Chung Kuang a K. Nadia PAPAMICHAIL. *The impact of integrating enterprise resource planning systems with business intelligence systems on decision-making performance: An empirical study of the semiconductor industry*. Manchester, 2010. Empirická studie. The University of Manchester.

WIEDER, Bernhard. *The impact of Business Intelligence on the quality of decision making – a mediation model*. Sydney, 2007. Studie. University of Technology Sydney.

AZIZ, Mustafa Nizamul a Ziyad SARSAM. *The impact and power of Business Intelligence (BI) on the Decision making process in Uppsala University: A case study*. Uppsala, 2013. Studie. Uppsala University, Sweden.

KIMBALL, Ralph a Margy ROSS. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling 3rd Edition*. 3. Indianapolis, IN: John Wiley, 2013. ISBN 1118530802.

KIMBALL, Ralph a Joe CASERTA. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data*. 1. San Francisco, USA: Wiley, 2010. ISBN 978-8126505548.

WIXOM, Barbara a Hugh WATSON. The Bi-Based organization. *International Journal of Business Intelligence Research*. 2010, **2010**(10), 4.

HOWSON, Cindi. *Successful Business Intelligence: Unlock the Value of BI & Big Data*. 2. New York: McGraw-Hill Education, 2014. ISBN 978-0071809184.

ECKERSON, Wayne. Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligence Solutions. In: *REPORT SERIES* [online]. Chatsworth, CA: The Data Warehousing Institute, s. 40 [cit. 2018-03-10]. Dostupné z: http://download.101com.com/tdwi/research_report/2003BIRreport_v7.pdf

BARBER, Tom. *Saiku Documentation*. 3. New York, 2017. Dostupné také z: <https://media.readthedocs.org/pdf/saiku-documentation/latest/saiku-documentation.pdf>

BOUMAN, Roland a Jos van DONGEN. *Pentaho solutions: business intelligence and data warehousing with pentaho and mysql*. 1. New York, USA: Wiley, 2009. ISBN 978-0470484326.

ELBASHIR, Mohamed Z., Philip A. COLLIER a Michael J DAVERN. *Measuring the effects of*

business intelligence systems: The relationship between business process and organizational performance. Melbourne, 2008. Studie. The University of Melbourne.

BAARS, Henning. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. *Information Systems Management*. 2008, **2008**(25), 50.

SALLAM, Rita L. Magic Quadrant for Business Intelligence and Analytics Platforms. *Cdn2.hubspot.net* [online]. New York: Gartner, 2017, 16 February 2017 [cit. 2018-03-10]. Dostupné z: <https://cdn2.hubspot.net/hubfs/2172371/Q1%202017%20Gartner.pdf?t=149626062>

DARROW, Barb. Hitachi Undergoes Major Reorganization to Compete With GE. *Finance.yahoo.com* [online]. San Francisco, USA: yahoo.com, 2017, September 19, 2017 [cit. 2018-03-10]. Dostupné z: <https://finance.yahoo.com/news/hitachi-undergoes-major-reorganization-compete-154547835.html>

Hitachi Introduces Hitachi Vantara: A New Digital Company Committed to Solving the World's Toughest Business and Societal Challenges. In: *Hitachi.com* [online]. Las Vegas, USA: Hitachi, 2017 [cit. 2018-03-10]. Dostupné z: <http://www.hitachi.com/New/cnews/month/2017/09/170920b.pdf>

RAMAKRISHNAN, Thiagarajan. *FACTORS INFLUENCING BI DATA COLLECTION STRATEGIES: AN EMPIRICAL INVESTIGATION*. TEXAS, 2010. Dezertační práce. UNIVERSITY OF NORTH TEXAS.

UNDERWOOD, JEN. EXPLORING 2017 GARTNER BI MAGIC QUADRANT RESULTS. *.jenunderwood.com* [online]. New York: Gartner, 2017 [cit. 2018-03-10]. Dostupné z: <http://www.jenunderwood.com/2017/02/22/2017-gartner-bi-magic-quadrant-results/>

REZAEI, Kamran. *Evaluating the Business Intelligence Systems Performance Criteria Using Group Fuzzy AHP Approach*. 2011. Cambridge, UK: IEEE, 2011. ISBN 978-1-61284-705-4.

ALIEV, R. A. a Oleg H. HUSEYNOV. *Decision theory with imperfect information*. [Hackensack] New Jersey: World Scientific, 2014. ISBN 978-9814611039.

TURBAN, Efraim., Ramesh. SHARDA a Dursun. DELEN. *Decision support and business intelligence systems*. 9th ed. Boston: Prentice Hall, c2011. ISBN 01-361-0729-X.

HAN, Jiawei, Micheline KAMBER a Jian PEI. *Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition*. 1. Waltham, USA: Morgan Kaufmann, 2011. ISBN 0123814790.

Creating Pentaho Dashboards. San Francisco, 2010. Dostupné také z: http://docs.huihoo.com/pentaho/pentaho-business-analytics/4.1/dashboard_guide.pdf

HASAN, NooreDdilla Abu, Suraya MISKON a Norasnita AHMAD. *Business Intelligence readiness factors for higher education institution*. Johor, Malaysia, 2016. Studie. Universiti Teknologi Malaysia.

NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business Intelligence: Jak využít bohatství ve vašich datech*. 1. Praha: Grada, 2005. ISBN 9788024766850.

HANNULA, Mika. *Business Intelligence empirical study on the top 50 Finnish companies*.

Cambridge, 2003. Empirical study. Tampere University of Technology, Finland.

INMON, William H. *Building the data warehouse*. 4th ed. Indianapolis: Wiley Publishing, c2005. ISBN 07-645-9944-5.

RICHARDS, Gregory B. *Big Data and Analytics Applications in Government: Current Practices and Future Opportunities*. 2017. Boca Raton, Fla: CRC Press, 2017. ISBN 978-1498764346.

CHAUDHURI, Surajit, Umeshwar DAYAL a Vivek NARASAYYA. An Overview of Business Intelligence Technology. *Communications of the ACM*. 2011, **2011**(54), 88-98.

Přílohy

Příloha č. 1 – DDL skript D_Employess.....	49
Příloha č. 2 – DDL skript D_Productss.....	49
Příloha č. 3 – DDL skript D_Customers	49
Příloha č. 4 – DDL skript D_Orders.....	49
Příloha č. 5 – OLAP kostka (Schema Workbench).....	54
Příloha č. 6 – Report č. 1.....	55
Příloha č. 7 – Report č. 2.....	56
Příloha č. 8 – Report č. 3.....	56
Příloha č. 9 – Report č. 4.....	56
Příloha č. 10 – Report č. 5.....	56

Příloha č. 1 - DDL skript D_Customers

```
CREATE TABLE [dbo].[D_Customers](
    [CustomerID] [nchar](5) NOT NULL,
    [CompanyName] [nvarchar](40) NOT NULL,
    [ContactName] [nvarchar](30) NULL,
    [ContactTitle] [nvarchar](30) NULL,
    [Address] [nvarchar](60) NULL,
    [City] [nvarchar](15) NULL,
    [Region] [nvarchar](15) NULL,
    [PostalCode] [nvarchar](10) NULL,
    [Country] [nvarchar](15) NULL,
    [Phone] [nvarchar](24) NULL,
    [Fax] [nvarchar](24) NULL
)
```

Příloha č. 2 – DDL skript D_Employees

```
CREATE TABLE [dbo].[D_Employees](
    [EmployeeID] [int] IDENTITY(1,1) NOT NULL,
    [FullName] [nvarchar](50) NOT NULL
)
```

..

Příloha č. 3 – DDL skript D_Products

```
CREATE TABLE [dbo].[D_Products](
  [ProductID] [int] IDENTITY(1,1) NOT NULL,
  [ProductName] [nvarchar](40) NOT NULL
)
```

Příloha č. 4 – DDL skript F_Orders

```
CREATE TABLE [dbo].[F_Orders](
  [OrderID] [int] NOT NULL,
  [ProductID] [int] NOT NULL,
  [UnitPrice] [money] NULL,
  [Quantity] [int] NULL,
  [EmployeeID] [int] NULL,
  [CustomerID] [nvarchar](50) NULL,
  [OrderDate] [datetime] NULL,
  [ShipCountry] [nvarchar](50) NULL
)
```

Příloha č. 5 – OLAP kostka (Schema Workbench)

The screenshot shows the Schema Workbench interface. The left pane displays a tree view of the schema 'northwind_olap1 (Schema1.xml)'. Under the 'D_Customers' dimension, the 'CompanyName' attribute is selected. The right pane shows the configuration for this attribute, with a table listing various properties and their values.

Attribute	Value
name	CompanyName
description	
table	
column	CompanyName
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	CompanyName
type	String
internalType	String
uniqueMembers	<input type="checkbox"/>
levelType	Regular
hideMemberif	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Příloha č. 6 – Report č. 1

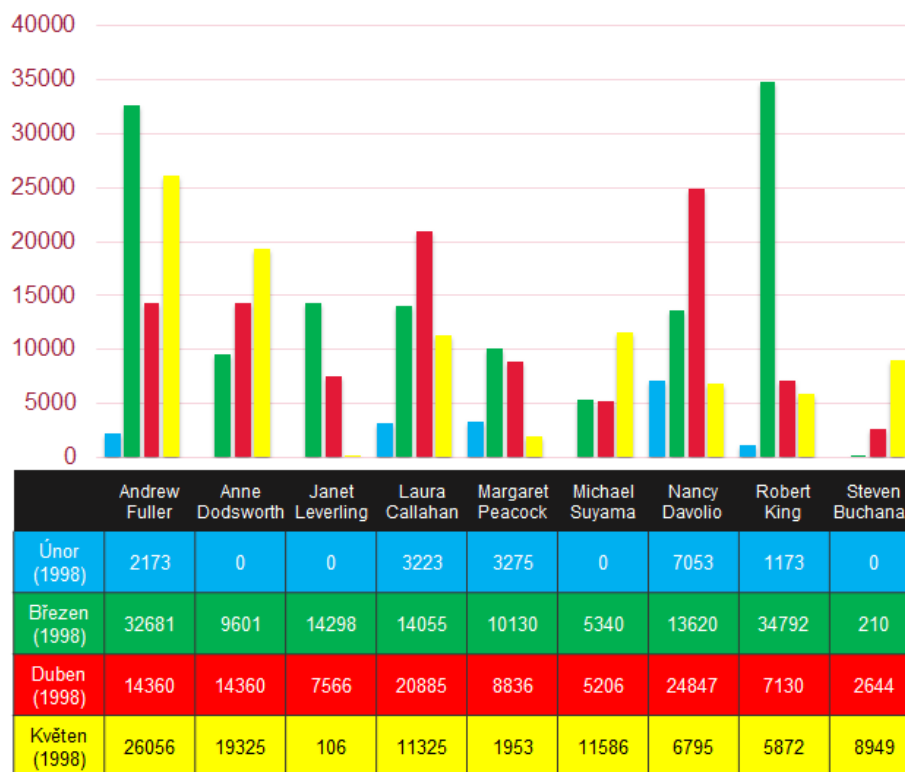
TOP 25: Nejprodávanější zboží			
Pořadí	Název zboží	Množství ks v tis.	Období
1	Wimmers gute Semmelknödel	132	1998-05
2	Alice Mutton	77	1998-05
3	Jack's New England Clam Chowder	64	1998-05
4	Chang	62	1998-05
5	Scottish Longbreads	55	1998-05
6	Teatme Chocolate Biscuits	52	1998-05
7	Pavlova	49	1998-05
8	Chai	40	1998-05
9	Ipoh Coffee	36	1998-05
10	Sasquatch Ale	35	1998-05
11	Spegesild	33	1998-05
12	Original Frankfurter grüne Soße	30	1998-05
13	Perth Pasties	25	1998-05
14	Chrtreuse verte	22	1998-05
15	Valkoinen suklaa	22	1998-05
16	Tofu	21	1998-05
17	Grandma's Boysenberry Spread	21	1998-05
18	Guaraná Fantástica	20	1998-05
19	Gorgonzola Telino	20	1998-05
20	Tourtière	20	1998-05
celkem		836	

Příloha č. 7 – Report č. 2

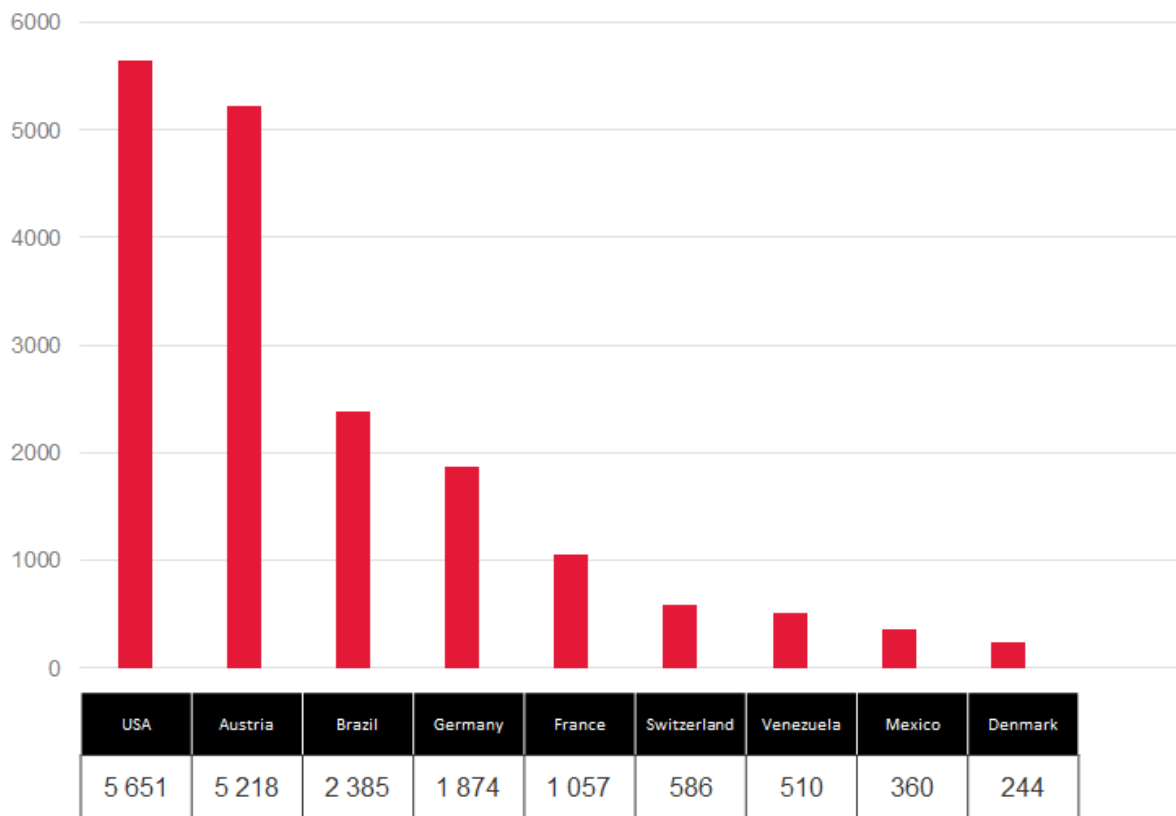
TOP 5: nejprodávanější zboží podle regionu				
Číslo	Název zboží	Množství v		
		tis. ks	Období	Region
1	Wimmers gute Semmelknödel	130	1998-05	Austria
2	Alice Mutton	77	1998-05	USA
3	Scottish Longbreads	55	1998-05	USA
4	Teatime Chocolate Biscuits	42	1998-05	USA
5	Chai	40	1998-05	Germany
Celkem:		344		

TOP 10: nejprodávanější zboží v USA				
Číslo	Název Zboží	Množství v		
		tis. ks	Období	Region
1	Alice Mutton	77	1998-05	USA
2	Scottish Longbreads	55	1998-05	USA
3	Teatime Chocolate Biscuits	42	1998-05	USA
4	Sasquatch Ale	35	1998-05	USA
5	Perth Pasties	25	1998-05	USA
6	Chang	24	1998-05	USA
7	Jack's New England Clam Chowder	15	1998-05	USA
8	Pâté chinois	6	1998-05	USA
9	Pavlova	5	1998-05	USA
10	Rhönbräu Klosterbier	4	1998-05	USA
Celkem:		288		

Objemy tržeb (v tis. dolarech) podle zaměstnanců

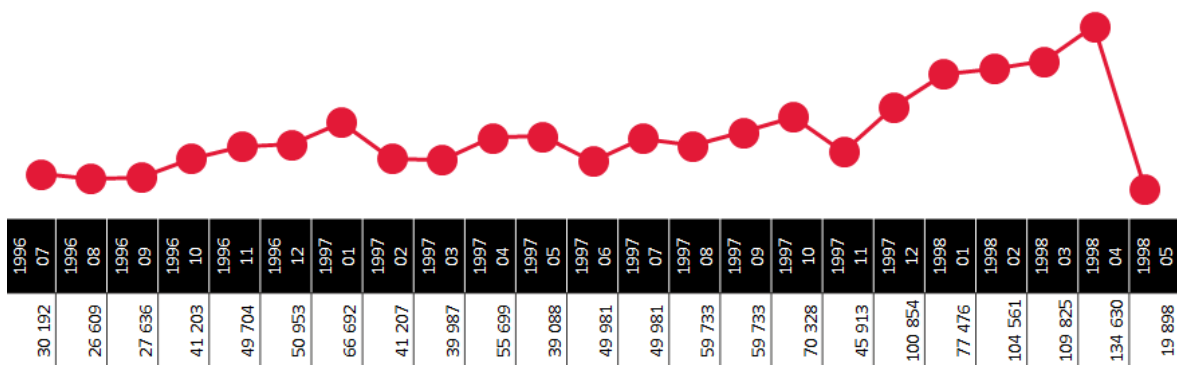


Objemy prodejů (v tis. dolarech) podle jednotlivých regionů – duben 1998



Příloha č. 10 – Report č. 5

Tržby prodeje (v tis. dolarech) za posledních 23 měsíců



Objemy prodeje (v tis. ks) za posledních 23 měsíců

