

Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra informačních technologií

Implementace Business Intelligence
řešení v podnikovém prostředí

Diplomová práce

Autor: Bc. Martin Urban

Studijní obor: Aplikovaná informatika

Vedoucí práce: Ing. Karel Mls, Ph.D.

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne

Jméno a příjmení

Poděkování

Chtěl bych poděkovat především vedoucímu diplomové práce Ing. Karlu Mlsovi, Ph.D. za cenné rady, připomínky a věnovaný čas při pomoci s vypracováním této diplomové práce. Zároveň bych rád poděkoval Ing. Martinovi Pozdílkovi, Ph.D. ze společnosti MF Servis s.r.o. za umožnění spolupráce na Business Intelligence projektu ZPL2 a odborné rady při tvorbě diplomové práce. V neposlední řadě patří poděkování Ing. Pavlovi Suchému ze společnosti Precheza a.s. za povolení prezentovat zpracování projektu ZPL2 v diplomové práci.

Anotace

Diplomová práce „Implementace Business Intelligence řešení v podnikovém prostředí“ popisuje implementaci Business Intelligence projektu ve společnosti Precheza a.s. Teoretická část diplomové práce charakterizuje základní pojmy Business Intelligence. Představuje moderní přístup tvorby Business Intelligence v cloudu a srovnání se stávající koncepcí. Dále popisuje oblasti tvorby ETL procesů, vytváření datových skladů a OLAP kostek. V praktické části se práce věnuje tvorbě Business Intelligence řešení pro společnost Precheza a.s. Součástí projektu je analýza současného stavu implementace a požadavky na nové řešení. Zahrnuje popis zdrojových dat, která jsou uložena v podnikovém informačním systému. Dále tvorbu ETL procesů, která je vytvořena v produktu SQL Server Integration Services a výstupní data jsou ukládána do datového skladu. Tvorba OLAP kostky je zpracována v produktu SQL Server Analysis Services a je interpretována kontingenční tabulkou v programu Microsoft Excel.

Klíčová slova

Business Intelligence, ETL, datový sklad, OLAP, SQL Server Data Tools, Precheza a.s.

Annotation

Title: Implementation of Business Intelligence solution in the business environment

Diploma thesis „ Implementation of Business Intelligence solution in the business environment” describes the implementation of the Business Intelligence project in Precheza a.s. The theoretical part describes the basic concepts of Business Intelligence. It represents a modern approach to the development of Business Intelligence in the cloud computing and compared with current concepts. It also describes the area of creation ETL processes, data warehousing and OLAP cubes. The practical part is dedicated to creating Business Intelligence solution for the Precheza a.s. company. The project analyses the current state of implementation and demands for new solutions. Includes description of the source data which are stored in the enterprise information system and also the creating of ETL processes, which is created in the SQL Server Integration Services and output data are stored in the data warehouse. Creating OLAP cube is processed in the SQL Server Analysis Services and is interpreted pivot table in Microsoft Excel.

Keywords

Business Intelligence, ETL, data warehouse, OLAP, SQL Server Data Tools, Precheza a.s.

OBSAH

1	Úvod a cíl práce	8
1.1	Úvod	8
1.2	Cíl práce	9
2	Business intelligence	10
2.1	Pojem Business Intelligence	10
2.2	Historie a vývoj Business Intelligence	11
2.3	Architektura Business Intelligence	14
2.3.1	Data ze zdrojových systémů	15
2.3.2	Vrstva extrakce, transformace a nahrávání dat	15
2.3.3	Vrstva pro uložení dat	16
2.3.4	Analytická vrstva zpracování dat	16
2.3.5	Vrstva pro prezentaci koncovému uživateli	18
3	Nejnovější trendy Business intelligence	20
3.1	Cloud computing	20
3.2	Cloud Business Intelligence	23
3.2.1	Tableau	25
3.2.2	Power BI	26
3.2.3	GoodData	27
4	ETL	28
4.1	Extrakce	29
4.2	Transformace a čištění dat	30
4.3	Nahrání dat	32
5	Datové sklady	35
5.1	Pojem datový sklad podle W. H. Inmona	35
5.2	Pojem datový sklad podle R. Kimballa	38
5.3	Dimenzionální model uložení dat	41

5.3.1	Tabulka faktů.....	42
5.3.2	Dimenze.....	43
6	Technologické platformy BI.....	46
6.1	Microsoft.....	46
6.2	IBM.....	48
6.3	SAP.....	49
7	Implementace BI v podniku Precheza a.s.	50
7.1	O společnosti Precheza a.s.....	50
7.2	Analýza a požadavky na BI řešení.....	51
7.3	Návrh BI řešení.....	54
7.3.1	Zdrojová data.....	55
7.3.2	Extraktor zdrojových dat.....	60
7.3.3	Datový sklad.....	62
7.3.4	Implementace v SQL Server Integration Services.....	66
7.3.5	Implementace v SQL Server Analysis Services.....	72
7.3.6	Výstup v Microsoft Excel.....	75
8	Závěr.....	78
	LITERATURA.....	79
	SEZNAM OBRÁZKŮ.....	82
	SEZNAM TABULEK.....	84

1 ÚVOD A CÍL PRÁCE

1.1 Úvod

Většina podniků v dnešní době disponuje jedním nebo častěji více informačními systémy, například CRM (Customer Relationship Management), ERP (Enterprise Resource Planning), SCM (Supply Chain Management) a mnoho dalších. Jednotlivé informační systémy v současné době generují velké množství dat, které je možné analyzovat. Získáme takto cenné informace, případně znalosti.

Pro proces transformace dat z informačních systémů na informace a znalosti je možné využít technologie Business Intelligence (BI). Překlad technologie Business Intelligence do českého jazyka doposud nebyl konkrétně definován, volným překladem by se dala nazvat jako obchodní inteligence. Technologie Business Intelligence je jedním z nejrychleji rostoucích odvětví informačních technologií. Umožňuje manažerům podniků zkvalitnit rozhodovacích činností na základě podložených znalostí. Uplatňuje se ve velkých organizacích, které jsou schopné investovat nemalé finanční prostředky do nového softwarového a hardwarového vybavení. Moderní trendy cloudové technologie snižují počáteční náklady na implementaci a je možné nasazení do středních a malých podniků. Získají tak konkurenční výhodu v efektivním řízení i možnost sledování a předvídání trendů před podniky, jež Business Intelligence nevyužívají.

Teoretická část diplomové práce bude zaměřena na objasnění základních pojmů v Business Intelligence. Seznámení s nejnovějšími trendy v oblasti Business Intelligence, která jsou provozována v cloudu jako SaaS (Software as a Service). Porovnání s klasickou non-cloud koncepcí. Znázornění kladných a záporných aspektů pro konkrétní použití v podnikovém prostředí. Nejobsáhlejší oblastí teoretické části je tvorba ETL (Extract-Transform-Load), ve kterém probíhá čištění dat z dostupných datových zdrojů. Proces ETL patří mezi časově nejnáročnější a nejdůležitější část tvorby Business Intelligence, protože zajišťuje korektnost datových zdrojů. Očištěná data se poté nahrají do datového skladu, ze kterého jsou čerpány pro tvorbu následných analýz, reportů, predikcí a rozhodovacích podkladů. V závěru teoretické části proběhne seznámení s technologickou platformou Microsoft SQL Server 2014, ve které bude provedena implementace Business Intelligence řešení pro společnost Precheza a.s. Srovnáním s konkurenčními technologiemi bude zhodnocena vhodnost výběru technologie od společnosti Microsoft.

Praktická část diplomové práce bude navazovat implementací kompletního řešení Business Intelligence v technologické platformě Microsoft SQL Server 2014. Implementace zahrnuje vyčištění zdrojových dat ETL procesem, vybudování datového skladu, realizace analýz a reportů.

1.2 Cíl práce

Cílem teoretické části práce je seznámení čtenáře s obecnými principy technologie Business Intelligence a ukázat moderní trendy, kterými se Business Intelligence nyní vyvíjí. Tyto trendy budou následně srovnány s klasickými koncepty a na základě porovnání technologických platforem vybraných společností je zhodnocena vybraná platforma pro následnou implementaci řešení ve společnosti Precheza a.s. Výstupem bude vytvoření konzistentních a pročištěných dat, která budou uložena do navrženého datového skladu. Z dat umístěných v datovém skladu budou realizovány analýzy a reporty následně aplikované do podnikové praxe.

2 BUSINESS INTELLIGENCE

Kapitola vysvětluje pojem Business Intelligence a objasňuje výhody aplikace na podniková data získaná v transakčních systémech. Historické vnímání technologie je spojeno se vznikem datových skladů, o které se zasloužili William H. Inmon a Ralph Kimball. Poslední část se zabývá obecnou architekturou Business Intelligence a jsou popisovány konkrétní části.

2.1 Pojem Business Intelligence

V současnosti informační systémy generují velké množství dat. Může se jednat o data pořízená z technologických procesů, firemní administrativy nebo informace o zákazníkovi. Data jsou pořizována z informačních systémů založených na využívání databázových systémů. Tyto informační systémy převážně pracují transakčním způsobem, a proto nejsou příliš vhodné k získávání informací. Samotná data uložená v transakčních systémech zatím nejsou žádnou konkurenční výhodou. Je potřeba s daty pracovat a získat z nich informace. Pojem informace lze definovat jako data, která mají pro uživatele určitý význam a s jejichž pomocí snáze porozumí datům a chodu celého podniku. Hlavním rozdílem Business Intelligence od podnikových informačních systémů je to, že pracuje analytickým způsobem. Umožňuje rychlý přístup ke konkrétním informacím, které jsou kvalitní, relevantní a přesné. Tímto Business Intelligence umožňuje manažerům, získat všechny potřebné informace ve správný čas, na správném místě a v potřebné podobě. Na základě těchto informací je manažer schopen lépe, rychleji a přesněji rozhodnout v oblasti plánování budoucího vývoje trhu či podniku. Pro oblast řízení podniku je důležité získání informací pro podporu rozhodování a právě tyto vlastnosti si Business Intelligence klade za cíl. Pokud manažer disponuje těmito informacemi, lze již hovořit o konkurenční výhodě.

Následuje citace termínu Business Intelligence od předních českých odborníků v této oblasti:

„Business Intelligence (BI) představuje komplex přístupů a aplikací IS/ICT, které téměř výlučně podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principu multidimenzionality, kterým zde rozumíme možnost pohlížet na realitu z několika možných úhlů.“[22]

2.2 Historie a vývoj Business Intelligence

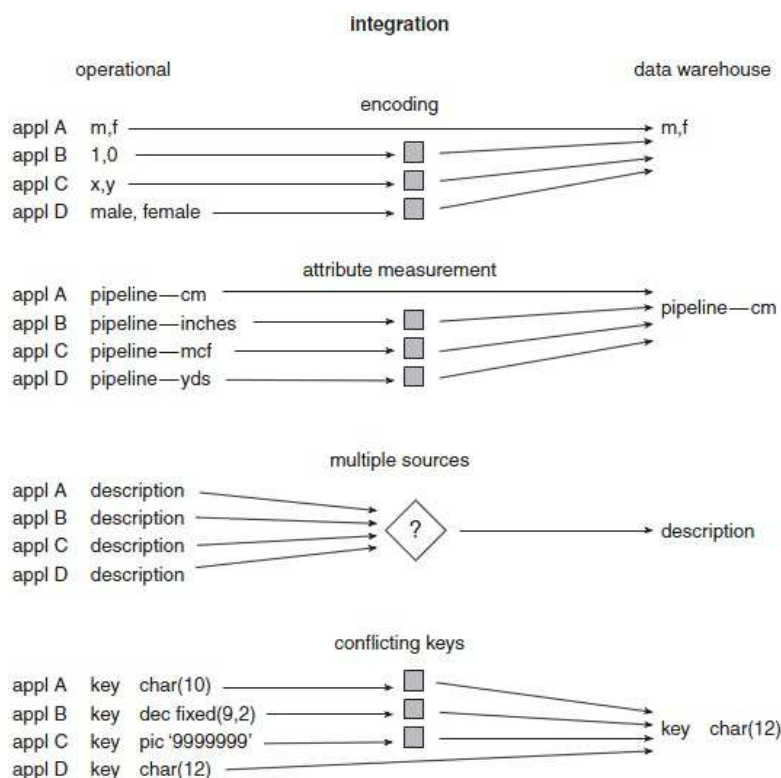
Samotný termín Business Intelligence poprvé v roce 1989 vyslovil analytik společnosti Gartner Group Howard J. Dresner. Tento odborník popisuje Business Intelligence jako sadu konceptů a metod pro zlepšení tvorby manažerských rozhodnutí pomocí podpůrných systémů na základě zjištěných skutečností.

Opravdové základy a počátky se objevovaly již v 60. letech 20. století. Dříve byly nazývány DSS (Decision Support Systems) a lze je přeložit jako systémy pro podporu rozhodování. Z názvu vyplývá, že tyto informační systémy měly sloužit manažerům k pomoci při jejich rozhodování. Existovaly různé typy DSS podporující rozhodování v různých oblastech řízení podniku – strategické, taktické nebo operativní. Pro strategické plánování se uvažovalo v řádu několika let oproti taktickému, který byl zaměřen na plánování v horizontu dnů až týdnů. Korektně navržený datový sklad by měl být schopný spolupracovat se všemi DSS, které jsou ve společnosti nasazeny. [25]

S Business Intelligence je úzce spojeno budování datových úložišť neboli datových skladů (Data Warehouse). Počátkem 80. let 20. století zakladatel William H. Inmon jako první definoval termín datového skladu a jeho architekturu. Datový sklad je subjektivě orientovaný, integrovaný, stálý a má časově rozlišitelné uspořádání dat, která jsou k dispozici potřebám manažerům společnosti. Uvedené termíny lze popsat následujícím způsobem: [14]

- Subjektivě orientovaný – datový sklad je použit k analýze pro určitou oblast dat, která jsou rozdělena podle jejich typu. Například data o prodejkách společnosti mohou být obsažena ve více informačních systémech. Z těchto systémů jsou v datovém skladu uložena všechna data stejného typu pouze na jednom místě.

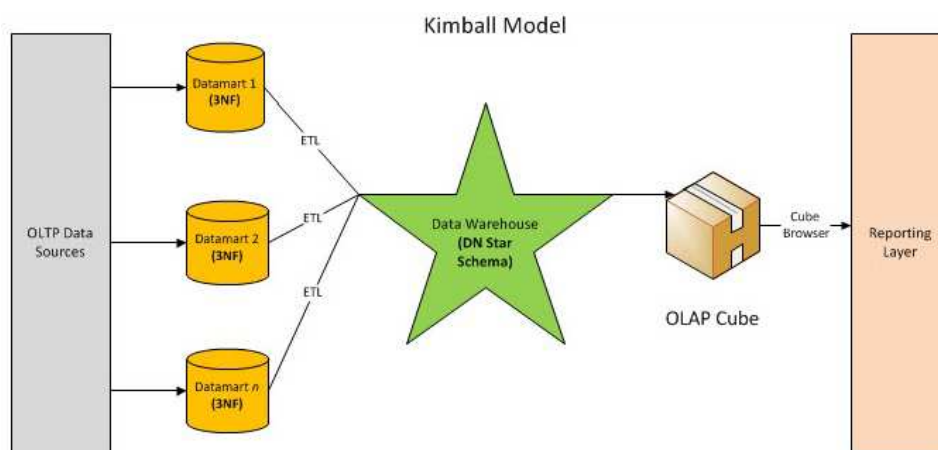
- Integrovaný – datový sklad integruje a ukládá data z různých datových zdrojů. Proto data musí být nejprve transformována do jednotné podoby a následně uložena v datovém skladu. Například první a druhý zdroj mohou mít rozdílné způsoby identifikace zaměstnance, ale v datovém skladu bude pouze jeden způsob jak daného zaměstnance identifikovat. Obrázek 1 znázorňuje transformace z různých provozních dat na jednotnou podobu.



Obr. 1: Zobrazení integrovanosti datového skladu dle Inmona; Zdroj: [14]

- Stálý – v datovém skladu nelze ručně vkládat data ani jakkoliv data modifikovat. Data jsou načítána pouze ze zdrojových či externích databází. Data takto uložená jsou k dispozici po celou dobu životnosti datového skladu.
- Časově rozlišitelný - Pro potřeby analýzy jsou nezbytná historická data, která jsou uložena v datovém skladu. K uloženým datům musí být přidán časový údaj, který umožňuje zadat časový horizont požadované analýzy. Například databáze transakčního systému má uloženou pouze poslední adresu zaměstnance, v datovém skladu jsou uloženy všechny adresy spojené s konkrétním zaměstnancem.

Druhým významným výzkumníkem, který se zasloužil o vývoj v této oblasti, je Ralph Kimball. Jeho pojetí datového skladu je definováno tvorbou datových tržišť, které jako celek vytvoří kompaktní datový sklad. Datová tržiště jsou zpravidla vyhrazena pouze pro určité oblasti zájmu společnosti (pobočka, oddělení apod.) Obrázek 2 je vizualizací modelu R. Kimballa. O pohledu pana Kimballa na datový sklad vypovídá jeho následující definice: „*Datový sklad je spojení všech datových tržišť v rámci společnosti. Informace jsou vždy uloženy v dimenzionálním modelu.*“ [17]



Obr. 2: Schéma modelu Ralpa Kimballa; Zdroj: [3]

Datové tržiště využívá pouze omezený počet uživatelů a lze na něj nahlížet jako na „malý datový sklad“. Výhodou datových tržišť je kratší doba návratnosti investic oproti tvorbě datového skladu. Další výhodou jsou nižší náklady a menší rizika při jejich zavádění.

Tato problematika je podrobně zpracována v kapitole 5 - Datové sklady

S vývojem datových skladů se do popředí dostávají metody multidimenzionálního uložení dat. Multidimenzionální databáze jsou koncipovány pro rychlé vyhodnocování komplexních dotazů na analytické bázi a umožňují ukládání agregovaných dat. Do databáze se k agregovaným datům uloží také jejich základní data v nejvyšším stupni detailu. [22]

Data v multidimenzionálním úložišti dat rozdělujeme na dva typy. Prvním typem jsou fakta, která popisují skutečnosti ve společnosti. Převážně se jedná o numerické hodnoty (výše prodeje, cena objednávky apod.). Druhým typem dat jsou dimenze. Obsahují popisné informace ke skutečnostem ve společnosti. V dimenzích se objevují

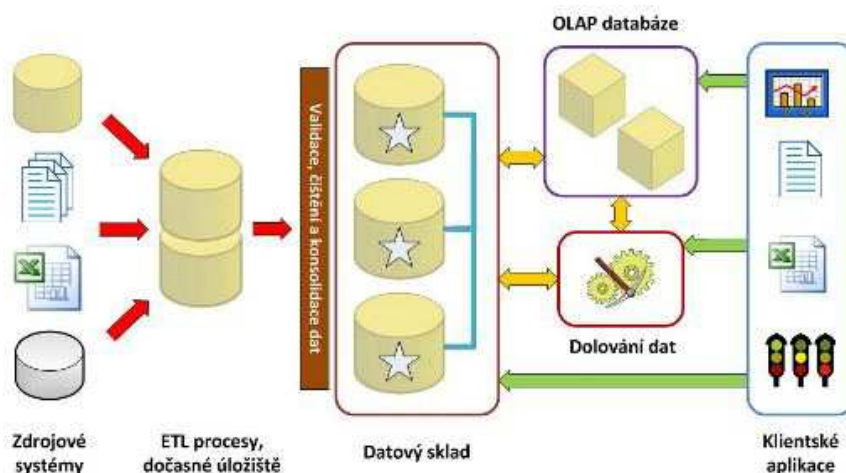
atributy, které v úložišti nejsou často měněné. Například uvažujeme-li dimenzi produkt, která může obsahovat atributy: jméno produktu, kategorie, barva apod. [17]

Mezi výhody multidimenzionálního uložení dat jsou považovány: rychlý přístup k obsáhlému datovému zdroji, který může být na nízkém stupni detailu, možnost vytváření komplexních datových analýz a podpora tvorby trendů a modelů. Nevýhoda může být spatřována ve vyšších nárocích na diskové úložiště. Nahrávání dat do databáze může být časově a systémově velmi náročné.

Používání dimenzionální techniky modelování umožňuje prezentovat relevantní informace uživatelům ve velmi krátkém čase a tím nejjednodušším způsobem. Toto jsou i základní kritéria pro tvorbu Business Intelligence řešení.

2.3 Architektura Business Intelligence

Koncepce architektury Business Intelligence a uspořádání aplikovaných komponent se převážně odvíjí od specifických požadavků zákazníků. Lze implementovat od nejjednoduššího řešení až po náročné a komplexní. Na složitosti řešení závisí alokace technologických, finančních a pracovních prostředků. [22] Přesto na základě zkušeností z vývoje v této oblasti je možné znázornit obecnou koncepci architektury Business Intelligence, který ilustruje obrázek 2.



Obr. 3: Obecný koncept architektury Business Intelligence; Zdroj: [24]

Konkrétně se rozdělují do následujících pěti základních vrstev:

- Data ze zdrojových systémů
- Vrstva extrakce, transformace a nahrávání dat
- Vrstva pro uložení dat
- Analytická vrstva zpracování dat
- Vrstva pro prezentaci koncovému uživateli

Následně jsou uvedeny bližší informace k jednotlivým vrstvám architektury.

2.3.1 Data ze zdrojových systémů

Podnikové informační systémy generují data, která jsou ukládána v transakčních databázích. Nejsou navrhnuté pro analytické zpracování dat, ale jsou optimalizované pro ukládání velkého množství transakcí. Do transakčních systémů můžeme například zařadit CRM, ERP, SCM systémy a mnoho dalších pro podporu existence konkrétních oddělení. [22]

Podnikové systémy lze považovat za hlavní zdroj dat. Čerpání dat je možné z i externích souborů jako například tabulky vytvořené v MS Excel, nebo z veřejných databází kurzového vývoje měny či meteorologických dat.

2.3.2 Vrstva extrakce, transformace a nahrávání dat

Zkráceně vrstvu označujeme jako ETL (Extract-Transform-Load), lze pro ni také používat český ekvivalent datová pumpa. Hlavním úkolem vrstvy je načtení dat ze zdrojových systémů, vyčištění a transformace a finální nahrání do datového skladu. Právě tato vrstva projektu Business Intelligence je nejdůležitější a časově i finančně nejnáročnější. Data musí projít denormalizací a deduplikací z heterogenních podnikových systémů, konverzí datových typů a jednotek, vytvořením agregovaných dat a multidimenzionálních struktur. Pokud je tato fáze projektu zanedbaná, manažer může přijímat zásadní rozhodnutí na základě chybných informací. V případě zjištění jakékoliv chyby, uživatelé ztrácejí důvěru k vytvořenému projektu a odmítají ho používat. [24]

Tato problematika je podrobně zpracována v kapitole 4 – ETL

2.3.3 Vrstva pro uložení dat

Další vrstva obsahuje databázové komponenty pro ukládání vyčištěných dat v prostředí Business Intelligence. Komponenty se velmi často liší podle požadavků zákazníka. Avšak nejčastěji je možné setkat se s datovými sklady Billa Inmona nebo s datovými tržišti Ralpa Kimballa. Kapitola 2.2 obsahuje popis principů obou přístupů. Dále je možné zaznamenat komponenty dočasného úložiště dat (DSA) a operativního úložiště dat (ODS), které jsou blíže představeny v kapitole 5 – Datové sklady.

2.3.4 Analytická vrstva zpracování dat

Pro analýzu dat se používají databáze s multidimenzionálními strukturami. Ve srovnání s relačními databázemi jsou převážně používány nenormalizované tabulky, které dělíme na tabulky faktů a dimenzí (viz kapitola 2.2). Analytické databáze se také často označují pod zkratkou OLAP (Online Analytical Processing). Typickým příkladem může být trojdimenzionální kostka s dimenzemi času, produktu a regionu. Specifikací parametrů v jednotlivých dimenzích získáváme výsledná data v dimenzionálních průnicích.[19]

Termín OLAP definoval Dr. E. F Codd, který v roce 1993 stanovil 12 pravidel, které by měl splňovat ideální OLAP systém: [5]

1. **Multidimenzionální konceptuální pohled** – pro analýzu shromážděných dat by měl OLAP uživateli poskytovat multidimenzionální model, který bude odpovídat podnikatelským požadavkům.
2. **Transparentnost** – technologie OLAP systémů, databáze či výpočty by měly být pro uživatele transparentní, aby uživatel mohl naplno využít svůj potenciál při použití front-end nástrojů.
3. **Dostupnost** - OLAP by měl přistupovat pouze k datům, které jsou nezbytné pro analýzu. Navíc by OLAP měl být schopný přistoupit ke všem datům nezávisle na tom, z jakého heterogenního informačního systému data pocházejí.
4. **Konzistentní vykazování** – pokud roste objem uložených záznamů v OLAP databázi, měla by délka přístupu zůstat konstantní nebo se jen mírně zvýšit.

5. **Klient-server architektura** – umožňuje OLAP systému mít vyšší výkon, flexibilitu a nižší náklady.
6. **Generická dimenzionalita** – každá dimenze dat musí být ekvivalentní svojí strukturou a operačními schopnostmi
7. **Dynamické ošetření řídkých matic** – systém OLAP musí být schopný dynamicky přizpůsobovat fyzické schéma konkrétnímu analytickému modelu podle hustoty analyzovaných dat.
8. **Víceuživatelský** – systém OLAP umožňuje práci více uživatelů současně na konkrétním modelu
9. **Neomezené křížové dimenzionální operace** – OLAP musí být schopný rozeznávat hierarchie dimenzí a automaticky provést kumulované kalkulace v různých kombinacích dimenzí.
10. **Intuitivní manipulace s daty** – uživatelské prostředí musí být intuitivní a přívětivé. Umožňuje základní analytické operace (drill down, drill up, slicing, pivoting a dicing)
11. **Flexibilní vykazování** – musí existovat způsob, jak uspořádat řádky a sloupce dat způsobem, který uživateli umožní intuitivní analýzu. Nebo může upořádat data podle logických struktur, které se nacházejí ve společnosti.
12. **Neomezené dimenze a úrovně agregace** – analytický model smí obsahovat více dimenzí a v každá z nich může mít vícenásobnou hierarchii. Implementace OLAP systému by neměla omezovat počty dimenzí.

Technologie OLAP je možné rozdělit podle typu implementace na následující varianty: (MOLAP, ROLAP a HOLAP)

- MOLAP – analýzu vykonávanou nad datovým skladem, ve kterém jsou data ukládána v multidimenzionální struktuře, nazýváme multidimenzionální OLAP analýzou (MOLAP). Zpracování MOLAP uloží analytická data do vlastních databázových struktur. Výstupy z analýzy jsou rychlé, protože data nejsou uložena v relační struktuře. Používají se multidimenzionální struktury, které jsou vhodnější k tvorbě analýz. Nevýhodou jsou požadavky na úložnou kapacitu, které s rostoucím počtem dimenzí mohou narůstat. [19]

- ROLAP – analýzu prováděnou nad daty, která jsou uložena v datových skladech s relační strukturou, nazýváme relační OLAP (ROLAP). Zpracování analýzy je náročné z důvodu nutnosti použití SQL příkazů pro načtení dat z relační struktury. Výhody tohoto modelu je možné spatřovat v práci s velkými objemy dat a nevznikají problémy s redundancí. Avšak k nevýhodám patří větší časové nároky na analýzu a vyšší výkon databáze. [23]
- HOLAP - jedná se o kombinaci MOLAP a ROLAP úložišť. Jsou aplikovány výhody jednotlivých řešení a potlačeny nevýhody. Data jsou ukládána do relačních databází a do multidimenzionálních struktur jsou ukládány předpočítané agregace. Tento typ implementace minimalizuje požadavky na ukládání dat a zároveň se vyznačuje rychlým zpracováním dat. [37]

Do analytické vrstvy zpracování dat také řadíme dolování dat, které je často nazýváno anglickým spojením data mining. V současnosti je to jedna z možností, jak se vyznat v obrovském množství ukládaných dat. Charakterizuje ho objevování strategických informací pocházejících z velkého datového zdroje pomocí matematických a statistických technik. Mezi ně je možné zahrnout rozhodovací stromy, genetické algoritmy, neuronové sítě a další. Z hlediska matematicko-statistického lze hovořit o testování hypotéz, hledání korelací a vzorů. [19]

Procesem extrakce neznámých informací jsou objeveny nové skutečnosti, které umožní manažerům zaměřit se na důležité faktory podnikání. Používá se například pro určení cílové skupiny pro konkrétní produkt, predikci prodeje konkrétního zboží a třeba i v medicíně k vyhledávání skrytých vztahů mezi jednotlivými onemocněními.

2.3.5 Vrstva pro prezentaci koncovému uživateli

Poslední vrstvou v obecném konceptu architektury pro Business Intelligence je prezentace koncovému uživateli. Jedná se o sumarizaci požadavků na analytické operace a následující prezentaci zjištěných výsledků. Pro formu prezentace je možné použít různorodé přístupy, například se může jednat o aplikace založené na webových technologiích, kancelářské aplikace a systémy EIS (Executive Information Systems). [22]

Právě poslední zmíněný systém EIS si klade za cíl podporovat v rozhodování střední a vyšší úroveň managementu. Pro podporu nižšího managementu jsou určeny nástroje reportingu. Rozdíl mezi těmito dvěma nástroji spočívá v tom, z jaké vrstvy

čerpají data pro vytvoření analýzy. Nástroje reportingu přistupují do datových skladů nebo přímo do transakčních databází, zatímco systém EIS si vytváří vlastní multidimenzionální sémantickou vrstvu. Systémy EIS jsou definovatelné následujícími hlavními body: [22]

- poskytují informace o chodu a plnění cílů celého podniku
- pracují s velkým množstvím datových zdrojů (interní i externí)
- možnost přistupovat ke konkrétním datům a taktéž z nich vytvářet agregace
- umožňují on-line základní analytické operace (drill down, drill up, slice, a dice)
- grafické a uživatelské prostředí je velmi intuitivní a lehce ovladatelné

3 NEJNOVĚJŠÍ TRENDY BUSINESS INTELLIGENCE

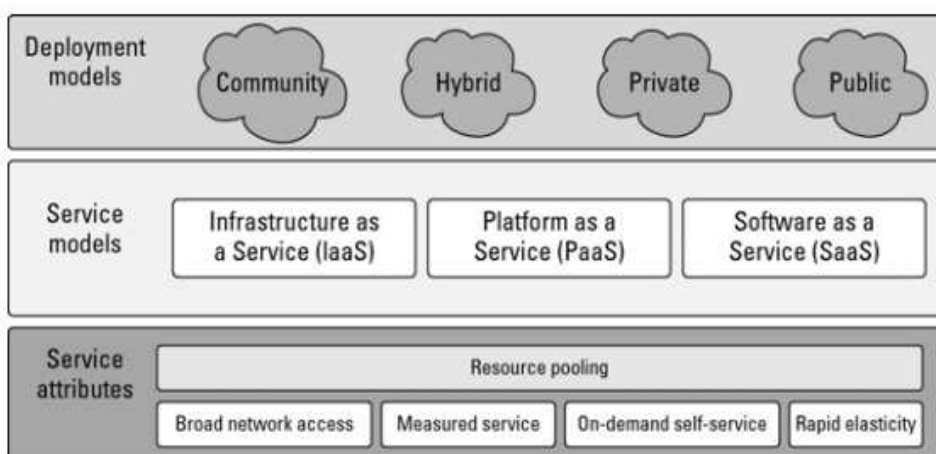
Vývoj v Business Intelligence lze spatřovat především ve spojení s technologií cloud computing. Pro zákazníky je atraktivní z důvodu úspory počátečních a provozních výdajů na fungování Business Intelligence, rychlosti vývoje a jednoduchosti používání pro koncové uživatele, analytiku či manažery.

3.1 Cloud computing

Pro cloud computing jsou zásadní dva hlavní termíny: [34]

- Abstrakce: Odstínění uživatelů a vývojářů od podrobností implementace systému. Používané aplikace, které jsou spuštěny na fyzickém stroji, nejsou nainstalovány na používaném stroji. Data se ukládají do neznámého datového úložiště. Přístup k aplikaci a datům je možný z jakéhokoli fyzického stroje.
- Virtualizace: Umožňuje systému rozdělovat výkon podle aktuální potřeby. Systém a úložiště může být spravováno z centrálního místa a náklady jsou účtovány pouze za využití zdroje.

Cloud computing je v pohledu převážně laické veřejnosti vnímán jako synonymum ke slovu internet. Ano, v praxi je možné se setkat s velkým množstvím rozdílných řešení cloud technologie, avšak předchozí tvrzení nelze vnímat jako správné. Správné definování cloud computing je popsáno podle [20], kteří jej dělí na pět základních charakteristik, model služeb (Services Model) a model publikování (Deployment Model) dle obrázku 4.



Obr. 4: Definice cloud computing podle NIST; Zdroj: [34]

Pět základních charakteristik (Service attributes)

Podle NIST (National Institute of Standards and Technology) musí splňovat následující charakteristiky, aby bylo možné hovořit o označení cloud computing: [20]

1. **Samoobslužný systém (On-demand self-service)** – zákazník si může sám určovat poskytování výpočetních kapacit. Hovoříme zde například o přidělení výkonu serveru a velikosti síťového úložiště, a to bez jakékoliv interakce s lidským faktorem.
2. **Široký síťový přístup (Broad Network access)** – Možnost přístupu přes standardní internetové protokoly. Pro připojení lze použít tlusté nebo tenké klienty (tablet, notebook a mobilní telefon)
3. **Sdílení zdrojů (Resource pooling)** – poskytovaný výpočetní výkon, datové úložiště, operační paměť apod. Vše je sdílené, aby bylo možné výpočetní zdroje použít pro více zákazníků a tím docílit vysoké efektivity. Zákazník často neví v jakém státě, oblasti a datovém centru se jeho data zpracovávají a ukládají.
4. **Rychlá flexibilita (Rapid elasticity)** – schopnost v době potřeby okamžitě uspokojit velkou poptávku po datových zdrojích a naopak v době nižší poptávky okamžitě uvolnit datové zdroje.
5. **Měřitelné služby (Measured service)** – cloud systémy automaticky řídí a optimalizují využívání výpočetních zdrojů. Disponují měřicími schopnostmi na určité úrovni abstrakce, které odpovídají druhu služby (úložiště, šířka pásma nebo počet aktivních uživatelských účtů). Využití zdrojů je monitorováno, kontrolováno nebo reportováno pro zákazníka a taktéž i pro poskytovatele.

Rozdělení podle modelu služeb (Services Model)

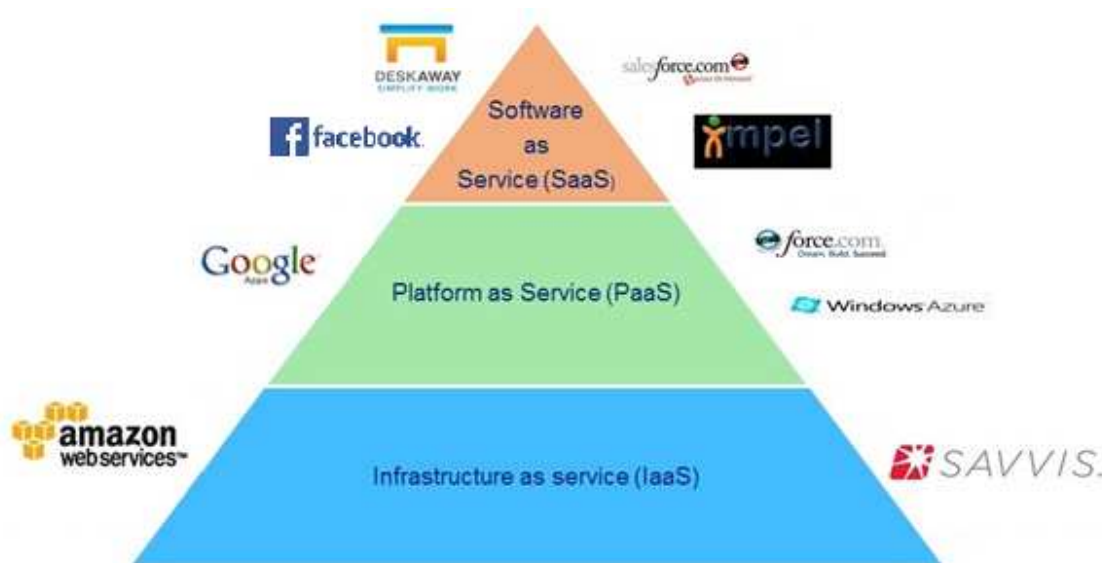
V praxi se zákazníci mohou setkat s rozdělením cloud computing na tři základní skupiny: [15]

- **Software as a Service (SaaS)** – zákazník přistupuje k aplikaci přes uživatelské rozhraní. Aplikace běží na serverech poskytovatele.
- **Platform as a Service (PaaS)** – zákazníkům jsou nabídnuty operační systémy, vývojové a administrační nástroje, s pomocí kterých mohou vytvářet vlastní

aplikace. Součástí tohoto řešení je i využívání serverů a úložného prostoru, které zajišťuje poskytovatel.

- **Infrastructure as a Service (IaaS)** - zákazníkovi poskytuje virtualizační technologie, úložný prostor a síťové prostředky. Spravuje si vlastní operační systémy a aplikace.

Na obrázku 5 je možné sledovat hierarchii jednotlivých skupin. Po stranách jsou uvedeny příklady služeb, které patří do dané vrstvy.



Obr. 5: Hierarchie modelu služeb a konkrétní příklady; Zdroj: [11]

Rozdělení podle modelu nasazení (Deployment Model)

Podle NIST rozdělujeme cloud computing do následujících čtyř skupin: [34]

- **Public cloud** – je dostupný pro používání veřejnosti, popřípadě pro velké průmyslové skupiny. Vlastní ho organizace, která poskytuje cloud služby.
- **Private cloud** – infrastruktura je používána pouze pro účely organizace. O správu cloudu se může starat samotná organizace nebo společnost třetí strany. Z toho vyplývá, že může být umístěná přímo v organizaci nebo ve společnosti poskytovatele. Hlavním přínosem je poskytnutí cloudu pouze skupině nebo uživatelům skupiny, která jej vlastní nebo je její součástí.
- **Community cloud** – podporována je specifická komunita, která je sdílená v rámci několika organizací se stejnými zájmy. Například konkrétní obchody, které spadají pod obchodní řetězec.

- **Hybrid cloud** – jedná se o spojení dvou či více cloudů předchozích skupin do jednoho kompaktního celku.

3.2 Cloud Business Intelligence

Společnost Gartner se zabývá analyzováním a sledováním vývoje konkrétních oblastí v oboru informačních technologií. Podle této společnosti je možné v oblasti Business Intelligence sledovat velký podíl společností, které nabízejí vlastní řešení Business Intelligence založené na cloud technologii.

Spojením dvou zmíněných technologií obecně získáme následující výhody:

- Nižší počáteční investice na pořízení potřebného softwaru a hardwaru
- Kratší doba implementace zadaného projektu
- Platba pouze za využitý a reálně spotřebovaný strojový čas v datovém centru.
- Všudypřítomná dostupnost dat prostřednictvím připojení k internetu s použitím notebooku, tabletu či mobilního telefonu.
- Intuitivní používání a práce s grafy.

Ale také je možné sledovat určité nevýhody, které mohou potenciálního zákazníka odradit: Především jde o nevědomost, v jaké lokalitě a datovém centru se pro zákazníka cenná data nacházejí. Například bankovní instituce jsou z důvodu bezpečnosti velice opatrné s používáním cloud technologií. Avšak moderní datová centra splňující certifikace, mohou být mnohdy bezpečnější než lokální úložiště ve společnosti. Příkladem může být datové centrum Swiss Fort Knox vybudované ve švýcarských Alpách (viz obrázek 6). Ve vzdálenosti 10 kilometrů od datového centra je z důvodu redundance ve skalním masívu vybudované totožné datové centrum. [35] Další nevýhoda může být spatřována při rozhodnutí změnit využívanou platformu. Stávající poskytovatel platformy si může klást nevýhodné požadavky na převedení dat do jiného úložiště, proto je vhodné na takové případy pamatovat ve smluvních podmínkách.



Obr. 6: Datové centrum Swiss Fort Knox; Zdroj: [35]

Gartner pro hodnocení postavení firem v oblasti Business Intelligence využívá vlastní metodu zobrazení. Nazývá ho Magický kvadrant (Magic Quadrant). Aktuální verzi pro rok 2015 ilustruje obrázek 7. Rozděluje se do čtyř čtverců ležících ve dvou osách. Vodorovná osa reprezentuje úplnost vize (completeness of vision) a svislá osa popisuje připravenost k růstu (ability to execute). Následně jsou uvedeny charakteristiky jednotlivých čtverců: [31]

- **Lídři (Leaders)** – tento kvadrant patří společnostem, které jsou ve svém oboru silné, často jsou to velké a vyzrálé podniky. Do jisté míry určují, jakým směrem se daný trh bude vyvíjet.
- **Vyzyvatelé (Challengers)** – v tomto kvadrantu jsou uvedeny společnosti, které mají velký potenciál růstu. Avšak mohou být omezeny technickým prostředím nebo specifickými způsoby užití. Jejich vize může být brzděná nedostatkem koordinace různých produktů v jejich portfoliu nebo špatným marketingem.
- **Vizionáři (Visionaries)** – společnosti nacházející se v tomto kvadrantu mají velkou vizi na dodávání své Business Intelligence platformy. Mají širokou nabídku funkcionalit a jejich platforma je flexibilní. Typicky se jedná o menší společnosti, o kterých mají zákazníci velice malé povědomí.

- **Niche Players** – pro název tohoto kvadrantu v českém jazyce neexistuje ekvivalent. Společnosti v tomto kvadrantu jsou specifické tím, že se svou platformou snaží zaplnit prázdné místo na daném trhu.



Obr. 7: Magický kvadrant pro oblast Business Intelligence; Zdroj: [31]

Blíže budou uvedeny tři společnosti, které poskytují řešení Business Intelligence ve spojení s cloud computing technologií. Prvním představitelem bude platforma společnosti Tableau, která je podle Gartnera jedním z lídrů v dané oblasti a má nejlepší předpoklady k růstu společnosti. Jako druhé bude představeno řešení od společnosti Microsoft, konkrétně produkt Power BI. Třetí představenou společností bude česká GoodData, která umožňuje pracovat s reporty pouze za pomoci webového prohlížeče.

3.2.1 Tableau

Společnost Tableau se primárně zaměřuje na vizualizaci dat. Jeho uživatelské prostředí je velice intuitivní a pro používání není potřebná spolupráce s IT specialisty. Pro tvorbu dotazů do databáze vytvořil vlastní vizualizační databázový jazyk nazvaný VizQL (Visual Query Language). Dotaz lze formulovat absolutně vizuálním přístupem bez použití jediného řádku kódu. [12, 36]

Tableau disponuje celou řadou již předpřipravených konektorů do různých datových zdrojů. Například lze čerpat datové zdroje z Microsoft SQL Server, Google Analytics, IBM DB2, Oracle Database, SAP HANA a mnoho dalších. [36]

Produkt, který využívá spojení s technologií cloud computing se nazývá Tableau Online. Cena je určena na 500\$/rok pro jednoho uživatele. Umožňuje využít úložný prostor o kapacitě 100 GB. [36]

3.2.2 Power BI

Společnost Microsoft nabízí Business Intelligence platformu provozovanou v cloud technologii pod názvem Power BI. Jedná se o doplněk velice oblíbené a často vyskytované aplikace Microsoft Excel, který používá velké množství firem. Uživatelem vytvořené tabulky či reporty v prostředí MS Excel lze nahrát do prostředí Power BI. Samozřejmě lze využít i řadu dalších zdrojových dat z Microsoft SQL Server, MySQL, IBM DB2, PostgreSQL, SAP HANA a přibližně dalších padesáti. Na data uložená v reportech se uživatel může dotazovat pomocí funkce Q&A (Otázky a Odpovědi). Uživatel dotazy formuluje pomocí přirozeného anglického jazyka a systém mu graficky prezentuje odpověď. [26]

Microsoft nabízí produkt ve dvou variantách. První varianta je Power BI, která je zcela zdarma a zákazníkovi nabízí datové úložiště o kapacitě 1 GB. Data mohou být aktualizovaná v denním horizontu rychlostí 10 000 řádků za hodinu. Druhá varianta je nazvána Power BI Pro a zákazníkovi bude nabídnuta za 9.99\$/měsíc. Přináší výhody v datovém úložišti o kapacitě 10 GB. Možnost aktualizovat data každou hodinu rychlostí 1 000 000 řádků za hodinu. [26]

3.2.3 GoodData

Společnost GoodData byla založena v roce 2007 Čechem Romanem Staňkem. GoodData svým zákazníkům nabízí vlastní platformu, kterou poskytuje jako SaaS řešení. Pro ukládání dat a provádění výpočtů jsou využívány servery společnosti Amazon. Zákazníkovi pro používání produktu dostačuje webový prohlížeč, proto nemá nadbytečné náklady spojené s koupí speciálního softwaru ani hardwaru. Nákup GoodData platformy zahrnuje také implementaci pro konkrétní společnost. Používání vytvořených reportů a dashboardů je velmi intuitivní a dovoluje menším firmám absenci IT oddělení. [9]

Propojení firemních dat, mezi které může patřit například Google Adwords, Sklik, Facebook, Microsoft Dynamics, SAP a mnoho dalších, obstarává také česká firma s názvem Keboola. Toto spojení dovolí zákazníkovi porozumět datům, která může čerpat z velké škály zdrojů.

Přes veškeré výhody, které platforma GoodData a Keboola nabízí, se cena ročního provozu s implementací se pohybuje okolo 50 000\$. Avšak, v porovnání s klasickým klient-server řešením a provozem IT oddělení, jsou nabízené služby výhodnější ve prospěch GoodData platformy. [38]

4 ETL

Nejdůležitější a časově i finančně nejnáročnější je v Business Intelligence projektu fáze ETL. Zkratka ETL je odvozena ze tří základních částí tohoto procesu – Extrakce, transformace a nahrání. ETL čerpá data z datových zdrojů. Obecně se může jednat o souborové databáze (Access), databáze spravované databázovým serverem (Microsoft SQL Server, Oracle, Informix apod.), data formátu XML a flat souboru exportovaná z informačních systémů, například ze SAP. Získaná data jsou nehomogenní, duplicitní a obsahují velké množství chyb. Proto je nezbytné provést transformaci a čištění dat. Na výstupní straně procesu se nachází datový sklad, který je již plněn upravenými daty a lze ho považovat jako důvěryhodný zdroj. Z datového skladu jsou následně plněna datová tržiště pro konkrétní oblasti podniku. Umístění ETL fáze v procesu zpracování dat a zvýraznění části náročné na lidské zdroje ilustruje obrázek 8.



Obr. 8: ETL schéma zpracování dat; Zdroj: [2]

Hlavní cíle a požadavky, které jsou kladeny na fázi ETL při vývoji Business Intelligence projektu popisují následující body:

- Centralizace dat z mnoha nehomogenních a odlišných datových zdrojů
- Transformace dat do požadovaného formátu
- Očištění zdrojových dat o duplicitní výskyty a chybové záznamy

- Dostatečný výkon pro práci s velkými objemy dat
- Nezávislost na technologické platformě
- Naplnění datového skladu relevantními daty a průběžné plnění nově vznikajícími daty

V následujících podkapitolách jsou detailně popsány tři základní fáze procesu ETL. Jedná se o extrakci dat ze zdrojových informačních systémů, transformaci získaných dat do potřebného formátu a nahrání výsledných dat do datového skladu.

4.1 Extrakce

Počáteční fází celého procesu je extrakce zdrojových dat z podnikových informačních systémů, externích databázových serverů, případně z dokumentů ve formátu XML, CSV či klasického textového výstupu. Může se jednat o nepřeberné množství kombinací databázových systémů (MS SQL Server, Oracle, IBM DB2 a další) a podnikových informačních systémů (SAP, Oracle, Microsoft Dynamics). Z toho vyplývá, že ETL ve fázi extrakce musí být schopné získat veškerá data nezávisle na platformě, ve které jsou implementovány zdrojové systémy. Extrakce dat by neměla ovlivňovat chod podnikových informačních systémů, ze kterých jsou čerpány datové zdroje. Z toho důvodu se pro extrakci volí čas od půlnoci do brzkých ranních hodin, kdy systémy nejsou využívány pro jejich primární účely. Toto je také důvod pro vznik dočasného prostoru pro uložení dat, který nazýváme Stage Area (viz obrázek 8). Následné transformace jsou prováděny ze Stage Area a není tím zatěžován zdrojový informační systém.

Typ replikace určuje, jestli se přenáší všechna data z datových zdrojů, nebo pouze nově získaná data od posledního provedení extrakce. Jedná se o offline replikaci, respektive o online replikaci: [27]

- **Offline replikace** – v jednom okamžiku jsou přenášena všechna data z datového zdroje. Pro zajištění konzistentnosti snímaných dat je odstaven podnikový informační systém a je zakázáno zapisovat do datového zdroje. Při snímání dat není prováděná žádná filtrace dat a veškerou úpravu dat je potřeba vykonávat v následujících transformační fázi. Implementace offline replikace je velmi jednoduchá. Aplikuje se

u datových zdrojů, které nepodporují čtení dat za chodu aplikace a při počátečním plnění datového skladu.

- **Online replikace** – za chodu podnikového informačního systému průběžně získává změny v operačních datech. Nutností je podpora tohoto druhu replikace na straně datového zdroje. Avšak prvotní naplnění Stage Area, případně datového skladu je téměř vždy potřeba provést pomocí offline replikace.

4.2 Transformace a čištění dat

Do fáze transformace a čištění dat vstupují přenesené soubory a tabulky, které jsou výstupem fáze extrakce. Z fáze transformace vznikají očištěné soubory a tabulky s konsolidovanými daty, které jsou připravené pro nahrání do datového skladu. Mezi další výstupy řadíme data, která nešla vyřešit automatickým procesem a je nutné, aby se jimi zabýval pověřený pracovník.

Při implementaci čištění a transformace dat je velmi důležité pamatovat na rovnováhu mezi automatizací veškerých anomálií vyskytujících se ve zdrojových datech a časem stráveným implementací. Často se vyskytují anomálie, které se ve zdrojových datech objeví pouze několikrát do roka a jejich automatizace je časově a tím pádem i finančně velmi náročná. V tomto případě je výhodnější, když úpravu dat provede odborný pracovník ručně.

Důvody vzniku anomálií ve zdrojových datech mohou být dvojího druhu. Prvním druhem způsobující chybná zdrojová data jsou změny v systémech, které data pořizují. Vývoj systému pořizujícího data je během na dlouhou trať a během jeho vývojového cyklu je nasazeno mnoho verzí softwaru. Obměňují se také platformy a operační systémy, na kterých se software vyvíjí. Příkladem může být přechod z operačního systému MS DOS na operační systém Windows. Což mělo za následek změnu kódování diakritiky, formátu času, oběživa a dalších. Druhý typ vzniku anomálií ve zdrojových datech je způsobuje lidský faktor. Příkladem mohou být pravopisné chyby, překlepy nebo nevyplněná pole. Čištěním dat docílíme sjednocení formátu dat, pořadí datových typů, jednotek a peněžních měn. [19]

Nejčastější problémy kvality zdrojových dat jsou popsány v následujících bodech: [19, 27]

- **Nejednoznačnost názvosloví dat** – je specifická tím, že data stejného typu jsou uložena v různých formách. Lze ji pozorovat v rámci jedné databáze nebo napříč databázemi, ze kterých čerpáme zdrojová data. Konkrétním případem může být základní atribut jako zápis pohlaví. Ve zdrojových datech jsou výskyty: žena, muž, M, Ž, ženské, mužské. Cílem transformace je tyto výskyty sjednotit do jednotného tvaru: M, Ž.
- **Chybějící hodnoty a překlepy** – sloupečky relačních databází, které obsahují NULL hodnotu. Chybějící data lze doplnit ručně z jiných zdrojů nebo se mohou v některých případech ignorovat. Dalšími chybami bývají uživatelské překlepy a psaní bez diakritiky.
- **Chybné datové typy** – záznam obsahuje hodnotu, která neodpovídá danému datovému typu.
- **Chybné klíče** – unikátní záznam obsažen v databázi se vyskytuje v téže databázi opakovaně, ale pouze s jiným primárním klíčem. V různých datových zdrojích nemusí být stejné primární klíče.
- **Chybějící metadata** – znemožňují zjištění struktury, která je v datových zdrojích zanesená. Nelze také určit jednoznačný význam polí, která se vyskytují ve struktuře datových zdrojů.

Výše uvedené anomálie a chyby ve zdrojových datech se řeší jejich transformací. Pojem transformace je definován jako soubor úkolů a úkonů, jejichž výsledkem je zvýšení kvality dat pořízených z podnikových informačních systémů. Pro jejich realizaci se používají následující metody: [27]

- **Selekce** – vybírá z dat pouze potřebné atributy k analýze
- **Filtrace** – výběr pouze určitých řádků z dat podle zvolené filtrační podmínky
- **Pivoting** – záměna řádků za sloupce nebo sloupců za řádky
- **Třídění** – setřídění dat podle konkrétního atributu sestupně nebo vzestupně

- **Spojení** – pomocí klíče spojit dva a více záznamů do jednoho uceleného záznamu. Realizováno může být vnějším (outer join), vnitřním (inner join) spojením a spojením pomocí tabulky lookup.
- **Agregace** – použitím statistické funkce (MIN, MAX, AVG, LAST a další) je z mnoha záznamů vytvořen pouze jeden. Důsledkem je snížení detailu a znemožnění zpětného výčtu dat, který byl agregován.
- **Merge** – sjednocení záznamů stejné entity, které pocházejí z různých datových zdrojů. Pro nalezení zmíněných záznamů se využívá metoda fuzzy porovnání.
- **Kombinace** – vhodné zkombinování dvou vstupních záznamů, jejichž výstupem jsou opět dva záznamy
- **Separace** – data s velkým množstvím sloupců se rozdělí do několika částí. Každá část bude obsahovat primární klíč, podle kterého lze následně data spojit. Toto se může vyskytovat u dat s vnitřní strukturou, kterou tímto rozdělíme na základní skupiny.
- **Konverze** – jedná se o transformaci hodnoty záznamu pomocí funkce, která může být matematická, časová, řetězová nebo s použitím lookup tabulky. Používá se například na změnu z malých na velká písmena a převod interních kódů výrobku na jejich názvy.

4.3 Nahrání dat

Do fáze nahrávání dat do datového skladu jsou vstupem očištěná a transformovaná data, které jsou výstupem předcházejícího procesu. Z fáze nahrání dat vzniká datový sklad s konsolidovanými a důvěryhodnými daty, které jsou podkladem manažerům při rozhodování v důležitých fázích vedení podniku. Cílem této fáze je přenos dat a jejich uložení do databázové struktury. Přenos by měl být plánovaný, automatizovaný a co nejvíce optimalizovaný. Ve fázi prvního plnění datového skladu lze očekávat, že může být nedostupný nebo jeho provoz bude omezen. Je to způsobené prováděním indexace tabulek, partitioningu, reorganizace a vytváření objektů z agregačních funkcí. K jednoznačné identifikaci každého řádku v tabulce se používají i uměle vytvořené klíče v datovém skladu.

Pro naplnění datového skladu je možné použít několik metod. Jejich výběrem ovlivňujeme dostupnost datového skladu a dobu jeho plnění. Při výběru metody nahrávání dat do datového skladu je vždy vhodné vzít v potaz, na jaké platformě je konkrétní datový sklad implementovaný. Existují následující metody nahrávání dat do datového skladu:

- **Přímé nahrávání dat** – v průběhu nahrávání nových dat musí být datový sklad odstaven, protože je realizovaný pouze na jednom počítači. Podle toho, na jaké technologické platformě je datový sklad implementován, je možné využívat jejich vlastností. Některé platformy umožňují nahrávat data do datového skladu přímo za chodu. Podmínkou však je, že v době plnění jsou pro uživatele data v datovém skladu dostupná pouze pro čtení. Kladem tohoto řešení je nízká cena. Nevýhodu lze spatřovat v možnosti výskytu technologických odstávek a také v tom, že při nahrávání dat do velkých datových skladů čas odstávky nemusí být dostačující.
- **Postupné nahrávání dat** – Změny provedené v informačních podnikových systémech jsou zpracovávány v malých objemech dat. Proces ETL je spouštěn v krátkých časových intervalech a částečné změny, které jsou provedené, označuje příznak neviditelnosti. Důvodem přidávání příznaku neviditelnosti je konzistence obchodních záznamů. Po provedení velké dávky ETL zpracování je příznak neviditelnosti změněn.
- **Oddělené ETL** – procesy extrakce a transformace dat jsou prováděny na zařízení, které je odděleno od stroje, na kterém je implementován datový sklad. Po dokončení procesů extrakce a transformace jsou data ověřena a mohou být přesunuta do datového skladu. Toto řešení dovoluje snížit potřebnou dobu odstávky datového skladu.
- **Flip-flop** – toto řešení vyžaduje existenci dvou shodných serverů. Jeden ze serverů je provozní a druhý je určen pro nahrávání dat. První server je dostupný i v době, kdy se nahrávají nová data na druhý server. Po ukončení nahrávání nových dat se z druhého serveru stane provozní server a naopak. Výsledkem je nedostupnost datového skladu pouze v době, kdy je prováděno přepínání mezi servery.

Nová data do datového skladu lze vložit více způsoby, které jsou popsány v následujících bodech:

- **Load Replace** – pokud při nahrávání dat tabulka v datovém skladu neexistuje, je nejprve vytvořena a poté jsou do ní nahrány nová data. V případě existence prázdné cílové tabulky, jsou do ní nahrána data bez předchozího zásahu. Jestliže se v tabulce nacházejí již určitá data, jsou nejprve smazána a poté jsou do ní nahrána nová data. Často se tento způsob uplatňuje při nahrávání prvních dat do datového skladu.
- **Load Append** – tímto způsobem se data do datového skladu pouze přidávají. Předchozí data v datovém skladu nejsou jakkoliv změněna nebo smazána. Avšak může nastat kolize s výskytem duplicitních klíčů. V případě shodných unikátních klíčů, by mohlo nastat to, že se nahrávaná nová data do datového skladu nevloží a systém nás na toto upozorní chybovou hláškou. Řešením je používání časových razítek v unikátních klíčích. Používá se pro vytváření událostních modelů historie a tabulek faktů.
- **Load Constructive Merge** – tímto způsobem se do tabulek přidají všechny nové řádky a je vytvořena historie nových řádků. Pro identifikaci změn mezi starými a nově přidávanými řádky musí být použit mechanismus, který, tyto změny rozpoznává. Tento způsob se používá pro inkrementální nahrávání dat v případech, kdy se datový sklad plní po částech.
- **Load Destructive Merge** – tímto způsobem jsou nová data nahrána do datového skladu a nahradí stará data, která jsou smazána. Z toho vyplývá, že se v datovém skladu neuchovává historie záznamů a proto tento způsob není vhodný pro použití v oblasti datových skladů.

5 DATOVÉ SKLADY

Datový sklad umožňuje shromažďovat velké množství dat umístěných v jednom centrálním úložném prostoru. Data uložená v datovém skladu jsou organizována do vhodných datových struktur. Datový sklad lze považovat za strukturované úložiště dat a představuje důvěryhodný zdroj pro čerpání dat určených k následným analýzám a reportům. Procesy, které zprostředkovávají plnění datového skladu ze zdrojových informačních systémů, jsou také součástí datového skladu. [19]

V následujících podkapitolách jsou detailně popsány přístupy k tvorbě datového skladu podle W. H. Inmona a R. Kimballa. Dále seznámení s dimenzionálním modelem ukládání dat a s jeho komponenty.

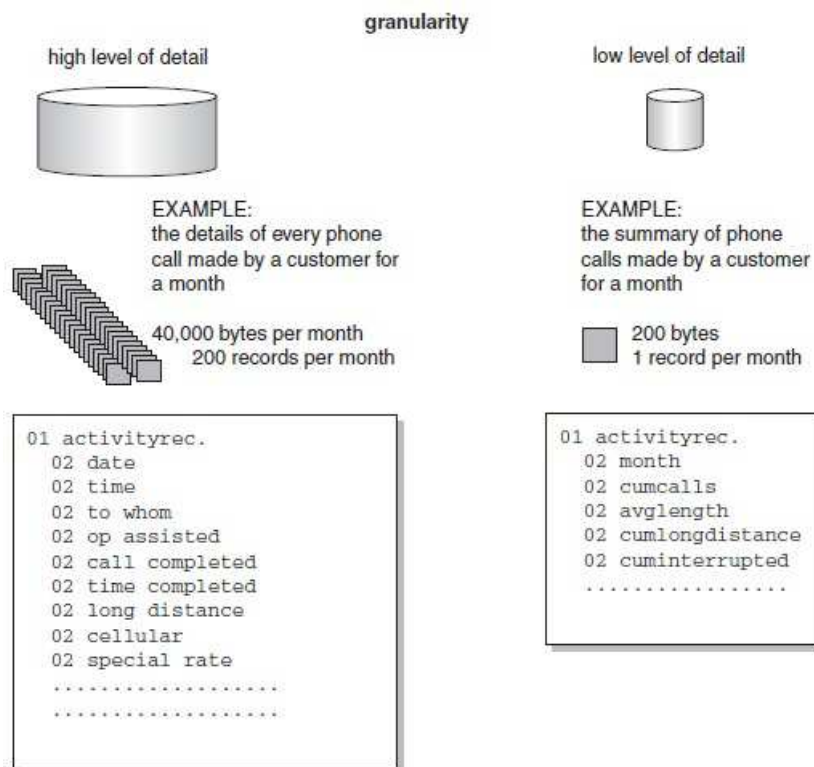
5.1 Pojem datový sklad podle W. H. Inmona

Pojem datový sklad (Data Warehouse) definoval v 80. letech 20. století William H. Inmon. Níže je uvedena jeho definice pojmu datový sklad.

„Datový sklad je podnikově strukturovaný depozitář subjektivě orientovaných, integrovaných, časově proměnných, historických dat použitých pro získávání informací a podporu rozhodování. V datovém skladu jsou uložena atomická a sumární data.“ [13]

Z definice je zřejmé, že pro datový sklad jsou důležité následující pojmy: subjektivá orientace, integrovatelnost, stálost a časová rozlišitelnost historických dat. Vyjmenované základní pojmy jsou detailně rozepsané v kapitole 2.2. Důležitým pojmem v souvislosti s datovým skladem je granularita.

Granularita v datovém skladu definuje maximální úroveň detailu dat, které mohou uživatelé analyzovat. Veškeré metriky, které jsou spojené s tabulkou faktů, musí být stejné granularity. Při úvodní části návrhu modelu datového skladu se používá maximální granularita, případně se vychází z konkrétních požadavků zákazníka. S detailními daty získanými díky požadavku na velkou granularitu, lze následně dále pracovat při vytváření agregací podle různých hledisek. Zvolením příliš malé granularity se může stát, že zákazník ze svých dat nezjistí důležité informace pro následné rozhodování. Naopak zvolením velké granularity zvyšujeme nároky na úložné kapacity a dobu vykonání dotazu. Příkladem může být obrázek 9, na kterém je znázorněn rozdíl v objemu dat a počtu záznamů o telefonních hovorech za jeden měsíc při velké úrovni granularity a nízké úrovni granularity. [13]



Obr. 9: Příklad volby úrovně granularity; Zdroj: [13]

Dále jsou uvedeny hlavní cíle a vlastnosti, na které by se při tvorbě ideálního datového skladu nemělo zapomenout: [13, 17, 27]

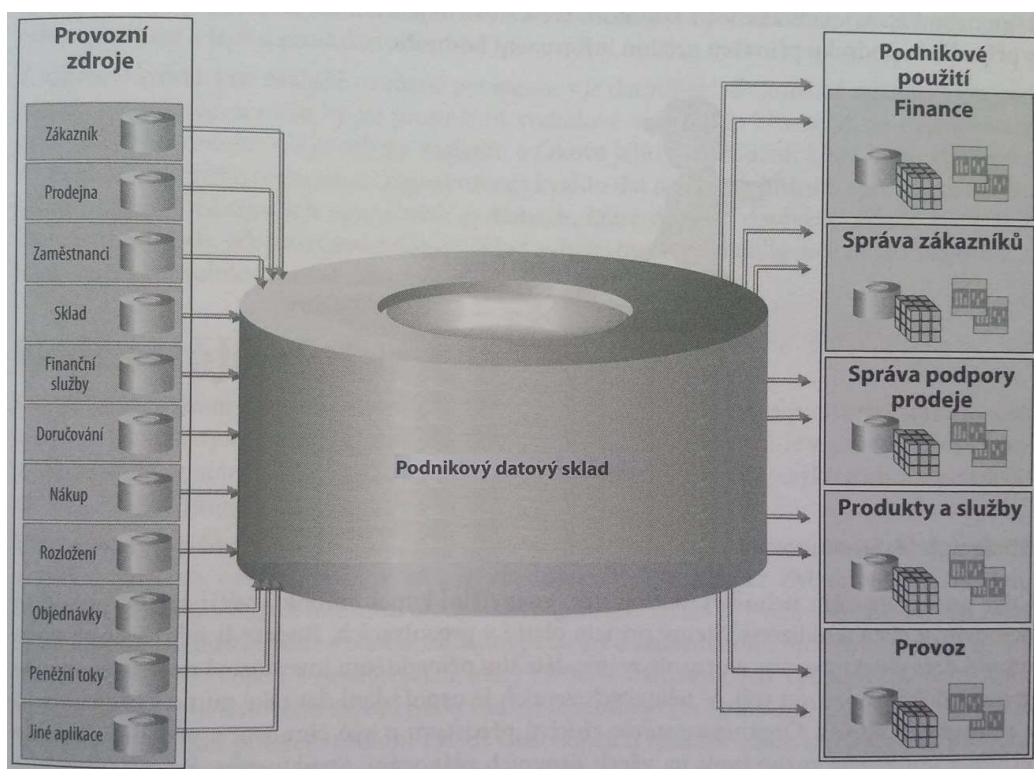
- V datovém skladu jsou obsažené veškeré informace z celé společnosti, které jsou následně využívány pro rozhodování a mají zásadní vliv na chod společnosti.
- Datový sklad musí obsahovat aktuální data, která uživatel je schopen použít pro tvorbu analýz a rozhodnutí. Z toho vyplývá, že je nutné provádět pravidelné plnění datového skladu novými daty.
- Data uložená v datovém skladu se nemění v závislosti na čase. Do datového skladu se data primárně ukládají a velmi zřídka se modifikují, případně mažou.
- Datový sklad musí být vytvořen podle reálných požadavků uživatele, který ho bude používat. Pokud by tomu tak nebylo, hrozilo by, že datový sklad nebude plně využíván a projekt by byl neúspěšný.
- Data vyskytující se v datovém skladu jsou konsolidovaná, integrovaná, pročištěná a chráněná. Před uložením dat do datového skladu je důležité

data upravit v ETL procesu. Výsledkem jsou věrohodná, kompletní, pročištěná a pravdivá data.

- Data nacházející se v datovém skladu musí být srozumitelná i pro uživatele, který není odborníkem v oblasti informačních technologií. Proto by měly být součástí datového skladu metadata, která vhodně popisují vyskytující se tabulky v datovém skladu.
- Při návrhu datového skladu je nutné počítat se změnami uživatelských požadavků, organizačních struktur a technologií. Datový sklad musí být flexibilní vůči takovýmto změnám.
- K datům uloženým v datovém skladu musí mít přístup pouze oprávnění uživatelé, aby se důvěrné informace nedostaly k neoprávněným osobám či konkurenci.

Z výše uvedených vlastností a definice je možné konstatovat následující výhody a nevýhody tvorby datového skladu podle koncepce Williama H. Inmona. Největší výhodou tohoto řešení je uložení všech dat z provozních informačních systémů do centralizovaného úložiště. Datový sklad lze považovat za důvěryhodný zdroj informací pro tvorbu analýz, reportů a podkladů pro rozhodování. Další výhodou lze spatřovat v relativně jednoduchém a rychlém vytváření datových tržišť pro jednotlivá oddělení nacházející se ve společnosti. Datová tržiště si zachovávají konzistenci, protože jsou generována z jednotného datového skladu, jak je zobrazeno na obrázku 10.

Oproti tomu největší nevýhodou lze spatřovat ve velmi náročné realizaci. Jedná se o časově náročnou implementaci a je nutné na ni vynaložit nemalé finanční prostředky. Společnost na zhotovení datového skladu čeká velmi dlouho a to samozřejmě podněcuje pochybnosti o vložené investici do projektu.



Obr. 10: Centrální podnikový datový sklad; Zdroj: [18]

5.2 Pojem datový sklad podle R. Kimballa

Druhou významnou osobou, která se zasloužila o vývoj v této oblasti, je Ralph Kimball. Jako první definoval koncept tvorby datového skladu pomocí datových tržišť (Data marts). Následně popsal dimenzionální modely hvězda (Star) a vločka (Snowflake). Definice datového skladu podle R. Kimballa je velmi odlišná od předchozí uvedené, kterou zveřejnil William H. Inmon. Proto následuje její plné znění.

„Datový sklad je spojení všech datových tržišť v rámci společnosti. Informace jsou vždy uloženy v dimenzionálním modelu.“ [17]

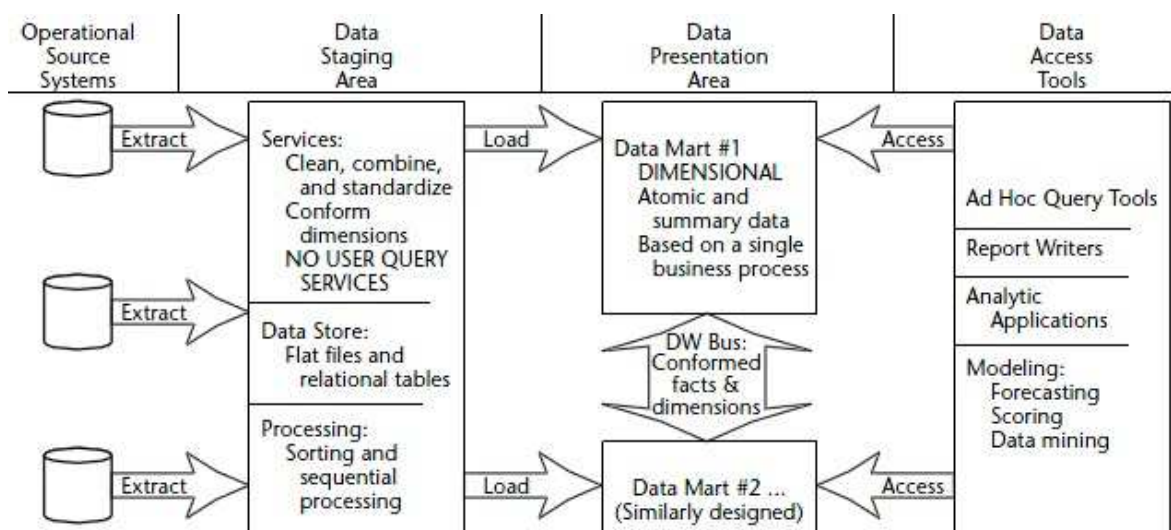
Z definice vyplývá, že pro tvorbu datového skladu podle R. Kimballa je nejdříve vytvořeno několik datových tržišť a jejich následným spojením vznikne komplexní datový sklad. Toto postupné budování datového skladu je pro společnosti velmi atraktivní. Časová náročnost na zhotovení datového tržiště je mnohem kratší a potřebné finanční náklady jsou nižší. Datové tržiště již okamžitě může fungovat v oblasti, pro kterou bylo navrženo a není závislé na ostatních datových tržištích nebo datovém skladu. V současné oblasti Business Intelligence je upřednostňován koncept

R. Kimballa. R. Kimball je právem nazýván otcem Business Intelligence, kdežto William H. Inmon je nazýván otcem datových skladů. [1]

Koncept tvorby datového skladu podle R. Kimballa znázorňuje obrázek 11, z kterého je patrné rozdělení do čtyř komponent. Jedná se o:

- **Zdrojové informační systémy** (Operational source systems)
- **Oblast transformace dat** (Data staging area)
- **Oblast prezentace dat** (Data presentation area)
- **Nástroje pro přístup k datům** (Data access tools)

Výše uvedené komponenty jsou následně podrobně popsány podle publikace R. Kimballa. [17]



Obr. 11: Koncept datového skladu podle R. Kimballa; Zdroj: [17]

Zdrojové informační systémy

Zdrojové informační systémy jsou určeny k zaznamenávání transakcí provedených v podnikovém prostředí. Ve svých strukturách udržují málo historických dat, a pokud je k dispozici dobře navrhnutý datový sklad, nemusí být po zdrojovém informačním systému vyžadováno udržování historických dat. Prováděné operace jsou krátké a jednoduché, očekává se rychlá interakce s velkým počtem uživatelů. Tudíž hlavními prioritami zdrojových informačních systémů je výkon a dostupnost.

Oblast transformace dat

V této oblasti se probírá sjednocení pořízených dat ze všech dostupných zdrojových informačních systémů. Zde data prochází procesem ETL a jsou nahrány do další oblasti prezentace dat. Podrobné informace o fungování ETL procesu jsou dostupné v předchozí kapitole 4. Hlavním požadavkem pro oblast transformace dat je její skrytí před koncovým uživatelem a nedovolení její používání pro poskytování dotazů a prezentaci služeb.

Oblast prezentace dat

V oblasti prezentace dat jsou data organizovaná, uložená a dostupná pro zpracovávání dotazů, které iniciují uživatelé nebo analytické aplikace. Typicky je tato oblast tvořena několika integrovanými datovými tržišti. Každé datové tržiště reprezentuje jednu ucelenou oblast nacházející se v podniku. Data v datových tržištích musí být prezentovaná, uchovávaná a přístupná v dimenzionálním schématu. Dimenzionální model obsahuje stejné informace jako normalizovaný model, který je používán ve zdrojových informačních systémech. Avšak se liší v tom, že je srozumitelnější pro uživatele a vhodný pro zpracovávání dotazů i odolný vůči změnám.

Data nacházející se v datovém tržišti musí být atomická. To znamená, že musí být dostupná v co nejnižším stupni detailu. To je nezbytné proto, aby datové tržiště bylo schopné odolávat ad hoc uživatelským dotazům. Pro zvýšení výkonu datového tržiště mohou být obsažena sumarizovaná nebo agregovaná data.

Všechna datová tržiště musí obsahovat společné dimenze a fakta, která označujeme jako odpovídající (conformed). Je to základ pro data warehouse bus architecture, ale bohužel pro tento pojem v českém jazyce neexistuje vhodný ekvivalent. Jedná se o propojení mezi jednotlivými datovými tržišti. Bez existence propojení mezi jednotlivými datovými tržišti by se zamezilo sdílení dimenzí a faktů, tudíž z každého datového tržiště by se stala samostatná aplikace. Propojení datových tržišť je zásadní princip, na kterém je založen koncept tvorby datového skladu podle R. Kimballa.

Nástroje pro přístup k datům

Poslední oblastí celého konceptu datového skladu jsou nástroje pro přístup k datům. V této fázi je možné poskytnout koncovým uživatelům analytické výstupy pro podporu v jejich rozhodování. Dotazování na informace je hlavním důvodem pro existenci datového skladu. Dotazy mohou být jednoduché, ad hoc, komplexní, nebo mohou být použity i aplikace pro dolování dat. Nástrojům pro tvorbu ad hoc dotazů rozumí a efektivně ho používá pouze malé procento uživatelů. Většina uživatelů přistupuje k informacím pomocí předpřipravených dotazů, ve kterých pouze uvedou požadované parametry.

Výhoda konceptu tvorby datového skladu podle R. Kimballa je v možnosti rychlého a postupného vývoje po jednotlivých datových tržištích, které jsou schopné zákazníkovi dodávat informace pro jejich určenou oblast. Také počáteční náklady na vývoj jsou nižší než při tvorbě centralizovaného datového skladu podle W. H. Inmona.

Oproti tomu nevýhodu lze spatřovat v nutnosti vytvořit větší množství rozhraní mezi zdrojovými informačními systémy a datovými tržišti. Taktéž pro vytvoření datového skladu je potřeba integrovat jednotlivá datová tržiště. Z toho vyplynou zvýšené nároky na správu datového skladu. V datovém skladu se také mohou objevovat redundance dat z důvodu spojování jednotlivých datových tržišť.

5.3 Dimenzionální model uložení dat

Data nacházející se ve zdrojových informačních systémech ve většině případů splňují třetí normální formu. Zde jsou uvedeny tři normální formy, kterých data v databázích mohou nabývat: [19]

- **První normální forma (1NF)** – tabulka splňuje první normální formu, pokud jsou všechny sloupce atomické a neobsahují žádnou vnitřní strukturu.
- **Druhá normální forma (2NF)** – tabulka pro zařazení do druhé normální formy musí splňovat podmínku první normální formy a zároveň každý atribut musí být úplně závislý na primárním klíči s výjimkou samotného primárního klíče.
- **Třetí normální forma (3NF)** – tabulka musí splňovat podmínky druhé normální formy a zároveň nesmí existovat závislosti mezi neklíčovými sloupci tabulky.

Ukládáním dat v třetí normální formě má za důsledek eliminaci duplicit. Tím je možné zefektivnit operace vkládání, mazání a upravování záznamů. Nevýhodou je to, že při komplexním dotazování, které vyžaduje spojení několika tabulek, se zvyšuje časová náročnost na vyhodnocení dotazu.

Z důvodu lepších vlastností při vyhodnocování komplexních dotazů je v datových tržištích a datových skladech používán **dimenzionální přístup k ukládání dat**. V porovnání s uložením dat v třetí normální formě se v dimenzionálním modelu objevuje redundance a již předpočítané agregované sloupce. Dimenzionální schéma je typicky složeno z tabulky faktů a k ní připojené dimenze.

5.3.1 Tabulka faktů

Tabulka faktů je základem dimenzionálního modelu. Data obsažená ve faktové tabulce musí být měřitelná a reprezentovat určitou míru nebo hodnotu. Z toho důvodu většina záznamů v tabulce faktů je reprezentována číselnou podobou. Příkladem může být např. cena objednávky, množství objednaného zboží nebo vlastní hodnotící metrika. Metriky jsou data, která pomáhají analytikům při hodnocení výkonu. Pro analýzu se používají následující typy metrik: [21, 27]

- **Aditivní** – jedná se o metriku získanou přirozeným sčítáním. Hodnoty se získávají použitím agregační funkce SUM.
- **Semiaditivní** – je to metrika, která lze sčítat pouze přes určité dimenze znázorňující statistický stav sledovaného atributu ke zvolenému datu. Příkladem může být uveden výpočet zisku v prodejnách, které se musí sčítat podle jednotlivých prodejů a ne podle času. Při agregaci podle časové dimenze musí být použity například následující funkce: FIRST, LAST, MIN, MAX, AVG a COUNT.
- **Neaditivní** – metriky, které nelze sčítat přes žádné dimenze. Příkladem mohou být procenta prodejů jednotlivých produktů. Pro agregace je možné použít funkce jako například: FIRST, LAST, MIN, MAX, AVG a COUNT.

Tabulka faktů obvykle zabírá až 90% úložného prostoru v dimenzionálních databázích. Objemná je převážně počtem řádků záznamů, avšak počet sloupců je nízký. [17]

Faktové tabulky by měly být navrhovány ve stejné úrovni granularity. V případě, že by tomu bylo opačně, mohly by se stát výstupní analýzy nepřehlednými a byl by možný výskyt chyb v interpretaci na různých úrovních granularity.

5.3.2 Dimenze

Dimenzionální tabulky jsou nezbytně nutné ve spojení s faktovými tabulkami. Dimenzionální tabulka obsahuje entity, které metrikám a faktům určují kontext a jejich prostřednictvím je co nejvíce popisuje. Obvykle jsou reprezentovány textovými popisy společnosti. Každá jednotlivá dimenze v návrhu dimenzionálního modelu může obsahovat velký počet sloupečků. Není výjimkou, že se může jednat o 50 až 100 atributů. Přesto je počet záznamů neboli řádků v tabulce velmi nízký oproti tabulce faktů. Z toho lze usuzovat, že dimenzionální tabulky v datovém skladu zabírají do 10% celkové velikosti úložných kapacit. [17]

V dimenzionálních tabulkách je možné často pozorovat výskyt hierarchie. Tímto způsobem lze definovat nadřazené a podřazené entity. V datovém modelu se může vyskytovat více paralelních hierarchií a je na volbě uživatele jakou hierarchie bude pro svoji analýzu využívat. Například se může jednat o dimenzi času v hierarchii: Den – Měsíc – Kvartál – Rok a nebo Den – Týden – Rok. [27]

Rozlišují se tři různé typy hierarchií: [27]

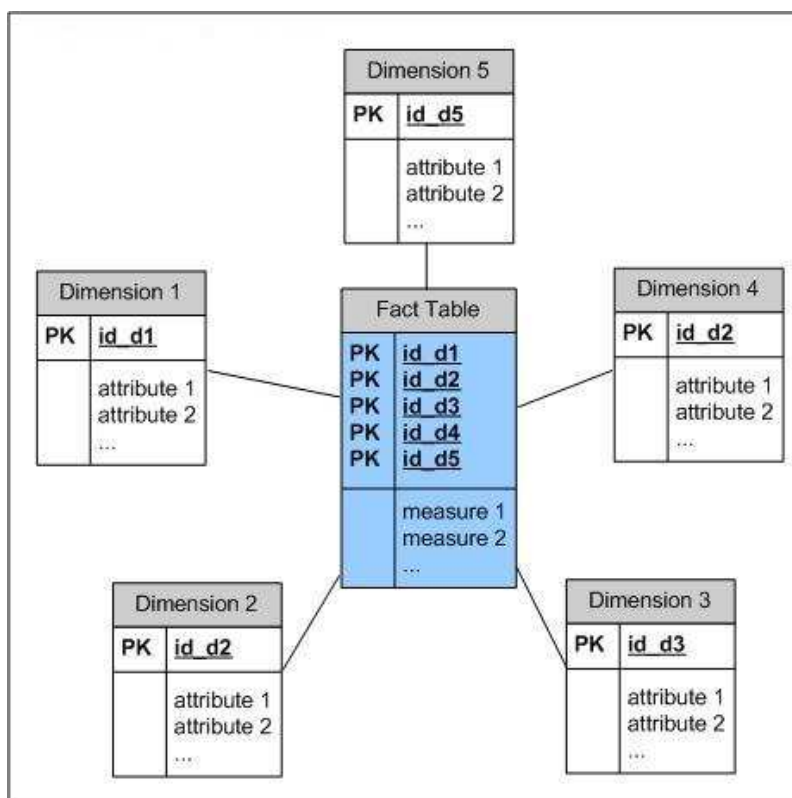
- **Klasifikační** – používá se k seskupování entit, například produktů do skupin a tříd.
- **Strukturální** – zobrazuje složení objektů z různých postupných částí.
- **Variační** – znázorňuje entity, které jsou odlišné verzemi, implementacemi a nebo mají různé modely.

Pokud již máme existující model ve třetí normální formě. Prvním krokem pro převedení do dimenzionálního modelu je rozdělení modelu ve třetí normální formě na jednotlivé podnikové procesy. Druhým krokem je vybrat vztahy M:N, které obsahují číselné a neklíčové hodnoty a označit jako entity faktové tabulky. Posledním krokem je denormalizace všech zbývajících tabulek do flat tabulek, které jsou následně propojené s tabulkou faktů. Z těchto tabulek se stanou dimenzionální tabulky.

Pro implementaci dimenzionálního modelu existují dvě základní schémata. Jedná se o schémata **Star (Hvězda)** a **Snowflake (Vločka)**.

Hvězda schéma

Jedná se o model, který má ve svém středu tabulku faktů a na ní připojené dimenzionální tabulky. V tabulce faktů se nacházejí cizí klíče, které jsou propojeny s primárními klíči jednotlivých dimenzí. Dimenze mohou obsahovat primární klíče k objektům, které nemají své cizí klíče v tabulce faktů, ale to je v souladu s pravidly dimenzionálního modelování. Ve schématu hvězdy je veškerá hierarchie dimenzí redukována pouze do jedné tabulky. Schéma je lépe pochopitelné pro běžného uživatele. Schéma hvězdy eliminuje spojování dimenzionálních tabulek a to má za následek možnost poskytnutí vyššího výkonu. Níže je uvedený příklad modelu ve schématu hvězda (obrázek 12).

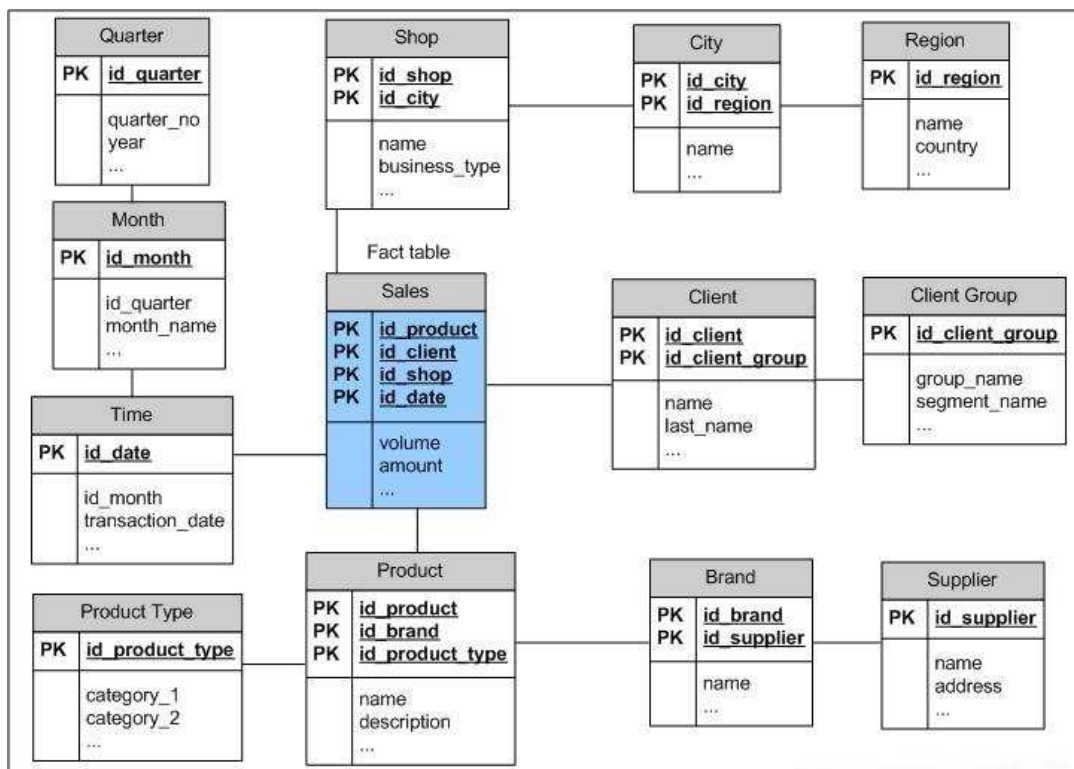


Obr. 12: Příklad modelu ve schématu hvězdy; Zdroj: [6]

Vločka schéma

Model schématu vločky vychází z předchozího schématu. Avšak umožňuje aplikování hierarchií v dimenzích. Mezi dimenzionálními tabulkami je definována referenční integrita.

Výhody tohoto modelu jsou spatřovány v zamezení redundance dat a tím pádem jsou menší nároky na velikost datového úložiště. Implementace je více stabilní při provádění změn v jednotlivých dimenzích a vkládání nových hierarchií do modelu. Její použití je vhodné v modelech, ve kterých se často mění struktura. Nevýhoda je v nutnosti spojování větších množství tabulek při analýze. Příliš složitá struktura modelu může být pro běžného uživatele nepochopitelná. [27] Na obrázku 13 je zobrazen příklad modelu vločkového schématu.



Obr. 13: Příklad modelu ve schématu vločky; Zdroj: [6]

6 TECHNOLOGICKÉ PLATFORMY BI

Cílem této kapitoly je seznámení s technologickými platformami, které primárně pro Business Intelligence řešení nevyužívají cloud technologie. Představeny budou tři zástupci technologických společností, kteří patří mezi lídry oboru Business Intelligence. Podkladem pro výběr je použit magický čtyřúhelník společnosti Gartner, který je zobrazen a blíže popsán v předcházející kapitole 3.2. Mezi lídry prakticky můžeme zařadit všechny velké softwarové společnosti. Někteří z nich vyvíjejí vlastní Business Intelligence produkt a někteří své portfolio v této oblasti obohatili akvizicemi s menšími společnostmi. V následujících podkapitolách jsou blíže představeny produkty společností Microsoft, IBM a SAP.

6.1 Microsoft

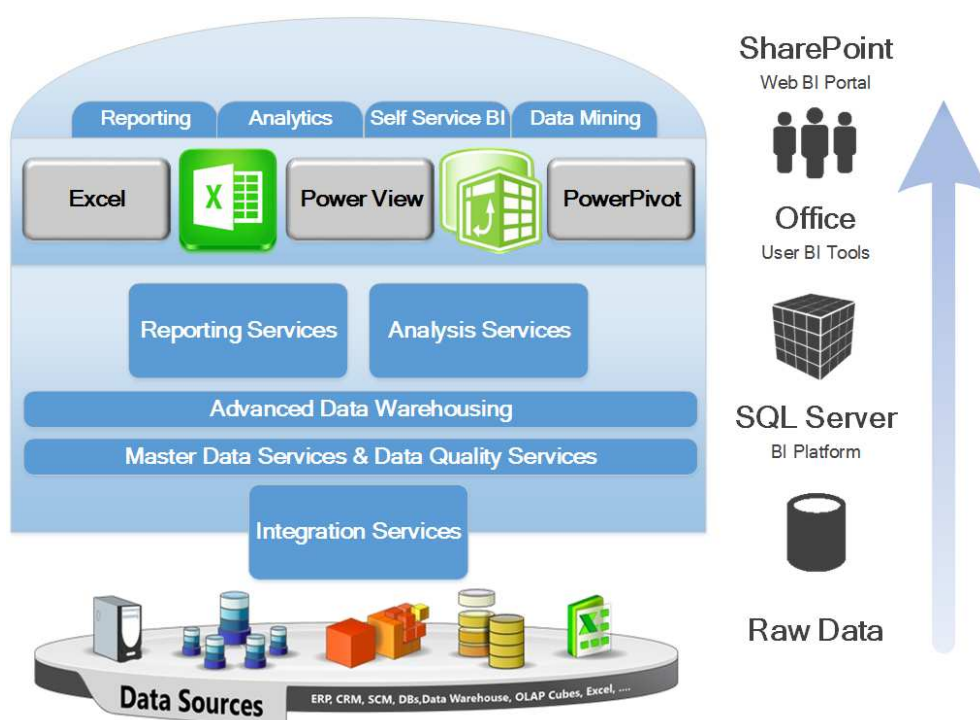
Společnost Microsoft je jednou z nejvýznamnějších v oblasti informačních technologií. Se svojí více než 40letou historií určuje trendy ve vývoji operačních systémů a kancelářských nástrojů Microsoft Office. Ve vývoji není pozadu ani v oblasti hardwaru, příkladem mohou být notebooky, mobilní telefony a herní konzole.

Řešení Business Intelligence od společnosti Microsoft má svůj základ v platformě Microsoft SQL Server. Získaná data jsou uživateli zobrazena pomocí kancelářského softwaru Microsoft Office. Pro komunikaci a sdílení obsahu mezi uživateli se využívá produkt Microsoft SharePoint. Výše uvedené produkty jsou zobrazeny na modelu struktury Business Intelligence podle společnosti Microsoft (obrázek 14).

Microsoft SQL Server představuje databázový systém, který zajišťuje datovou oblast pro Business Intelligence řešení. Jeho součástí je SQL Server Management Studio, jehož funkcí je zajištění komplexního integrovaného prostředí pro správu, databázového serveru. Další součástí MS SQL Serveru je SQL Server Data Tools, který obsahuje následující nástroje [19]:

- **Integrační služby** (Integration Services) – zde probíhá získávání dat z nehomogenních datových zdrojů a aplikuje se na data případně transformace a integrace.

- **Analytické služby** (Analysis Services) – v tomto nástroji jsou k datům připojeny výsledky analýz a predikce získaná z data miningu. Tímto se z relativně bezcenných dat stávají cenné informace pro následný rozhodovací proces.
- **Reportovací služby** (Reporting Services) – nástroj je určen pro prezentaci analytických výstupů pro uživatele ve vhodné formě, obsahu a rozsahu.



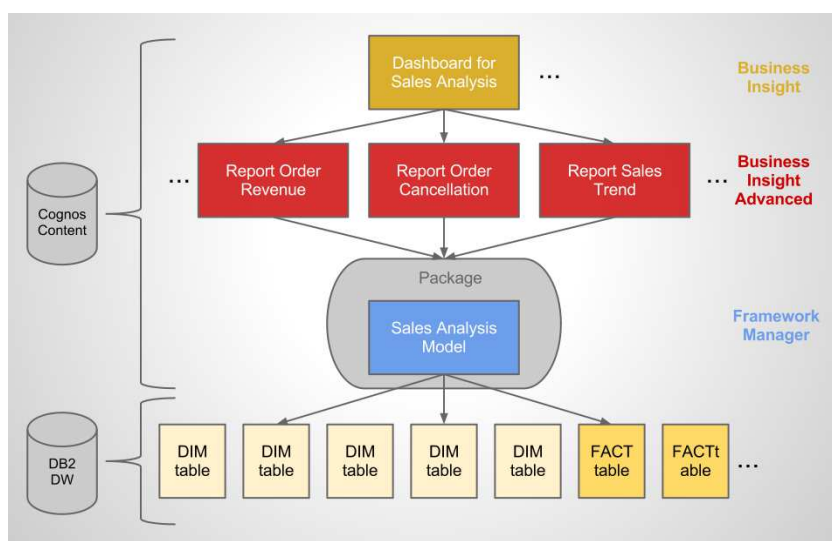
Obr. 14: Model BI podle společnosti Microsoft; Zdroj: [8]

Řešení od společnosti Microsoft je vhodné pro společnosti, které již využívají jakékoliv technologie od této společnosti. Produkt pro Business Intelligence je zahrnut v licenci pro MS SQL Server a proto nejsou nutné žádné další investice. Výstupy lze zpracovávat v kancelářském balíku MS Office, který je mezi uživateli všeobecně známý, a z toho důvodu nejsou nutná žádná další zaškolování uživatelů pro práci v novém programu.

6.2 IBM

Společnost IBM s 125letou historií je taktéž jedna z nejvýznamnějších v oblasti informačních technologií. Na začátku se IBM věnovala výrobě zařízení na zpracování dřevných štítků, které bylo využíváno převážně pro účely zpracování údajů ze sčítání lidu. Poté se společnost zaměřovala na osobní počítače, ale z důvodu velké konkurence tento segment opustila. Nyní se zaměřuje na oblast serverových řešení a ukládání dat. Zákazníkům poskytuje vlastní databázový systém DB2. Zpočátku se však IBM v oblasti Business Intelligence s vlastním vyvíjeným produktem neprosadila. K zařazení mezi lídry v této oblasti jí dopomohla akvizice s kanadskou společností Cognos, která vlastnila nejlepší Business Intelligence produkty. IBM pokračuje s vývojem tohoto produktu s ponechaným názvem Cognos. IBM Cognos je rozdělen do následujících tří úrovní (viz obrázek 15): [10]

- **Business Insight** – taktéž webový nástroj, který používá obsah Cognosu a externích zdrojů pro vytvoření interaktivních dashboardů, které usnadňují rozhodování.
- **Business Insight Advanced** – jedná se o webový nástroj, který umožňuje analyzovat data. Uživatelské prostředí je navrženo tak, aby pomohlo uživatelům získat vhled do jejich podnikatelské činnosti.
- **Framework Manager** – tato vrstva umožňuje ukládání dat v dimenzionálních a faktových tabulkách, které jsou následně použity pro analýzy a reportování.



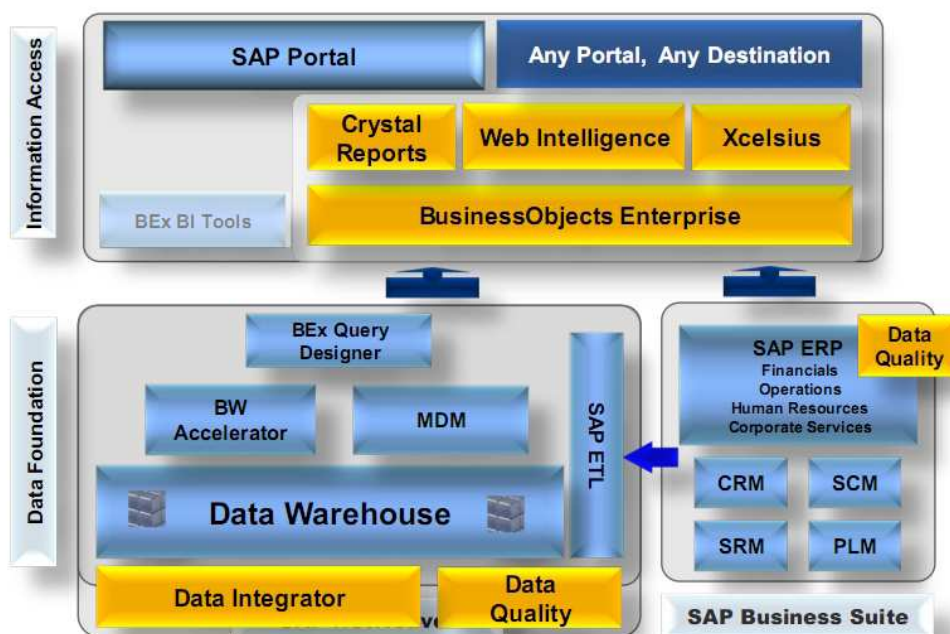
Obr. 15: Model BI podle společnosti IBM; Zdroj: [10]

6.3 SAP

SAP je jednou z největších softwarových společností a v oblasti podnikových informačních systémů (ERP) je lídrem na trhu. Produkty SAP jsou nasazovány ve společnostech všech velikostí od malých až po velké podniky. SAP NetWeaver je základní platforma, ze které vychází všechny SAP aplikace. Cílem této platformy je spolupráce s již existujícími řešeními, které je nasazené v zákaznickově společnosti. [33]

SAP své produkty v oblasti Business Intelligence rozšířil v roce 2007, kdy získal společnost Business Object. Proto produkty zaměřené na Business Intelligence nazývá SAP Business Object. Standardně získává data z různých podnikových informačních systémů v oblasti SAP Business Suite. Pro transformaci a integraci dat je používána oblast BusinessObjects Data Services. Následně je možné provádět analýzy, reporty v oblasti BusinessObjects Integrated Analytics Platform, ve které se například nachází Crystal Reports nebo Web Intelligence. Podrobné schéma modelu ilustruje obrázek 16.

Lze jmenovat jednu z výhod, která je způsobena vedoucím postavením společnosti SAP v oblasti informačních podnikových systémů. Jedná se o možnost aplikovat dotazy in-memory do SAP podnikového informačního systému, které urychlí zpracování dotazu, jak je ilustrováno na obrázku 16.



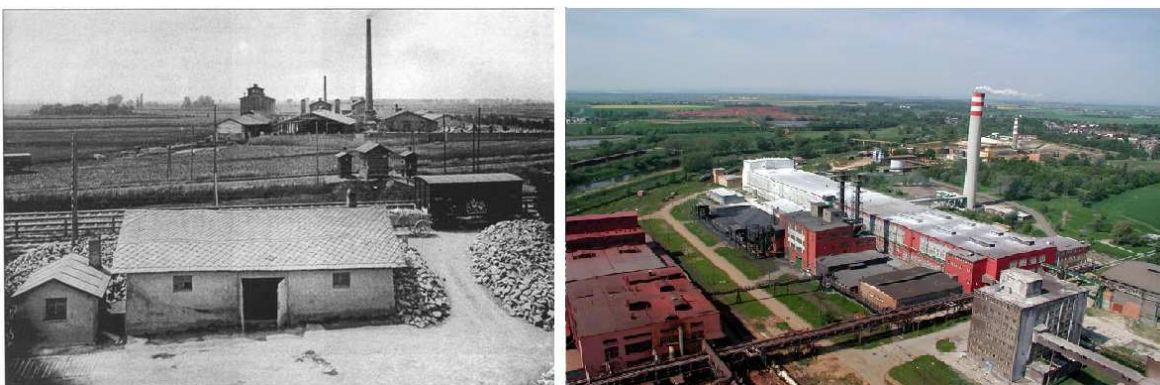
Obr. 16: Model BI podle společnosti SAP; Zdroj: [32]

7 IMPLEMENTACE BI V PODNIKU PRECHEZA A.S.

Cílem této kapitoly je seznámení se společností Precheza a.s. a současným stavem implementovaného Business Intelligence řešení od firem IBM a Microsoft. Jsou zde zaznamenány požadavky na nové Business Intelligence řešení od zaměstnanců společnosti Precheza. Následně je vytvořen návrh architektury s produkty společnosti Microsoft. Součástí kapitoly je seznámení se zdrojovými daty, které jsou uloženy v podnikovém informačním systému. Pro získání dat je nutné použít extraktor implementovaný v jazyce Java. V návrhu datového skladu je popsána tabulka faktů, jednotlivé dimenze a jejich atributy. V závěru je popsána implementace v SQL Server Integration Services a SQL Server Analysis Services. Výsledkem celého procesu je zobrazení dat v kontingenční tabulce produktu Microsoft Excel.

7.1 O společnosti Precheza a.s.

Historie chemické společnosti Precheza a.s. se datuje do roku 1894, kdy byla založena jako První akciová továrna na soustředěná hnojiva a lučebniny v Přerově. V té době byla zaměřena na výrobu kyseliny sírové a superfosfátu. Společnost mnohokrát změnila svůj název a až v roce 1991 po transformaci ze státního podniku Přerovské chemické závody byl uveden současný název Precheza. V roce 2000 byl do společnosti zaveden provoz informačního podnikového systému SAP R/3. Dalším důležitým milníkem je rok 2003, ve kterém se majoritním vlastníkem stal Agrofert holding a.s. založený Andrejem Babišem. [28]



Obr. 17: Vlevo z počátku 19. století a vpravo z počátku 21. století; Zdroj: [28]

V průběhu let společnost vyráběla velké množství chemických produktů. V současnosti jejími hlavními produkty jsou: [29]

- **Titanová běloba** – světovým prodejem okolo 5 milionů tun ročně se řadí mezi nejdůležitější a nejrozšířenější anorganické pigmenty. Jedná se o univerzální bílý pigment, který se využívá v průmyslu papírenském, gumárenském, potravinářském, kosmetickém, farmaceutickém a v mnoha dalších průmyslových odvětvích.
- **Železité pigmenty** – pro jejich výrobu je hlavní surovinou zelená skalice, která vzniká při výrobě titanové běloby. Železité pigmenty ve finální podobě mohou být v barvách: černé, hnědé, červené, zelené, oranžové a žluté. Jsou užity při pigmentaci tmelů, gumy, keramiky, skla, papíru a dalších. Nejvíce je používán ve stavebním průmyslu na pigmentaci střešních krytin, cementů, dlaždic a cihel.
- **Monosal** – taktéž vzniká při výrobě titanové běloby. Používán jako přísada do krmných směsí pro dobytek a pro zlepšení hygienických vlastností cementu.
- **Kyselina sírová technická** – bezbarvá olejovitá kapalina, která má své využití v textilním a papírenském průmyslu.

Hlavním prodejním produktem je titanová běloba, které se prodá přibližně 86% z celého obratu společnosti. Obrat společnosti také posiluje z 8% prodej železitých pigmentů a následují další chemické výrobky v řádu jednotek procent. Více jak 90% vyrobených chemických výrobků je určeno pro export. Do západní Evropy (51%), střední a východní Evropy (19%), severní a střední Ameriky (13%), Asie (15%) a procentuální zbytek připadá na oblast Afriky a Střední východ. Použití výrobků společnosti Precheza lze hledat v oblastech nátěrových hmot, plastů, stavebnictví, papírenství, farmacie, potravinářství a zemědělství. [30]

7.2 Analýza a požadavky na BI řešení

Společnost Precheza v současné době využívá Business Intelligence řešení pro zobrazení skutečného objemu vyrobených, nakoupených a prodaných produktů. Kromě skutečných výsledků jednotlivých produktů a jejich skupin se také zobrazuje očekávání a plán. Plán je určován s ročním předstihem a očekávání vychází z ročního plánu, avšak každý měsíc je postupně zpřesňováno. Při vytváření sestavy je vstupním

parametrem aktuální měsíc. Pro měsíce, které jsou před aktuálním měsícem, se zobrazují skutečné hodnoty a pro budoucí měsíce se zobrazují očekávané hodnoty.

Každá sestava obsahuje pro konkrétní výrobek nebo skupinu výrobků předem definované metriky, které svojí numerickou hodnotou poskytují význam dimenzím. V sestavě (obrázek 18) se vyskytují následující metriky:

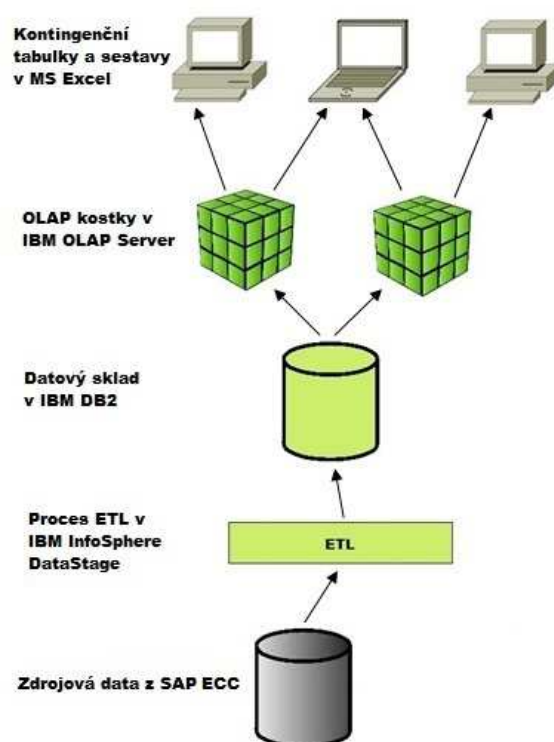
- **Zásoba počáteční** - konečná zásoba z předchozího měsíce
- **Zásoba počáteční v Kč** - konečná zásoba v Kč z předchozího měsíce
- **Výroba** – objem vyrobeného produktu
- **Nákup** – objem nakoupeného produktu
- **Spotřeba** – spotřebované množství daného produktu
- **Prodej** – objem prodaného produktu
- **Prodej Kč** – objem prodeje prodaného produktu vyjádřený Kč
- **Ostatní** – metrika představuje rozdíly mezi počáteční a konečnou zásobou, které nejsou zahrnuty ve výrobě, spotřebě, nákupech ani prodejích
- **Zásoba konečná** - je rovna Počáteční zásoba + Výroba + Nákup – Prodej – Spotřeba + Ostatní
- **Zásoba konečná v Kč** – konečná zásoba vyjádřená v Kč
- **Prodejní cena měsíční** – je rovna Prodej Kč / Prodej
- **Prodejní cena kumulativní** – je rovna \sum Prodej Kč od začátku roku / \sum Prodej od začátku roku

	A	B	C	D	E	F	G	H	I	J	K	L	M
1									duben				
2			1/2015	2/2015	3/2015	1.Q/2015	4/2015	5/2015	6/2015	2.Q/2015	3.Q/2015	4.Q/2015	2015
3			Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost	Skutecnost
4	10004	Zasoba poc											
5		Zasoba poc KC											
6		Vyroba											
7		Nakup											
8		Spotreba											
9		Prodej											
10		Ostatni											
11		Zasoba											
12		Zasoba KC											
13		Prodej KC											
14		PC mes											

Obr. 18: Metriky agregující se v čase

Tento projekt je nazván ZPL2 a slouží pro vytváření podkladů a reportů. Na základě dodaných podkladů a reportů z oddělení controllingu se plánovací komise rozhoduje o budoucím objemu výroby, prodeje a nákupu jednotlivých produktů.

Současné řešení je implementováno v produktech společnosti IBM a Microsoft. Za datový zdroj je považován podnikový informační systém SAP ECC, který zaznamenává a ukládá data systému do databází. Pro relační databázový systém je používán moderní IBM DB2. Následuje fáze ETL, ve které je také používán software společnosti IBM, konkrétně InfoSphere DataStage. Data pro fázi ETL se získávají zpracováním SQL dotazů do zdrojových dat. Pro datový sklad je použita databáze IBM DB2. Výpočet olap kostek je prováděn v produktu IBM OLAP Server. Jako nástroj pro zobrazování kontingenčních tabulek, sestav a reportů je používán Microsoft Excel. Současnou architekturu výše popsaných komponent ilustruje obrázek 19.



Obr. 19: Komponenty architektury současné implementace BI

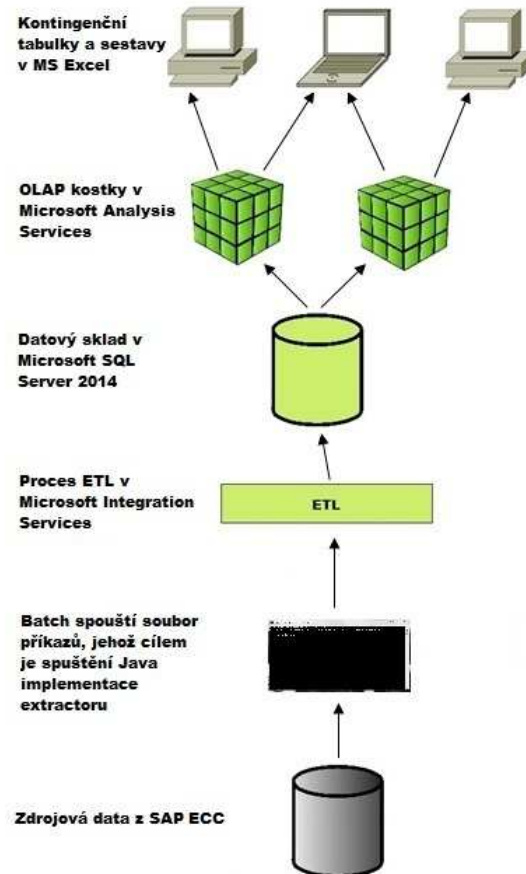
Úvodní schůzky společnosti Precheza se zúčastnili zaměstnanci z oblasti controllingu a informačních technologií. Cílem úvodní schůzky bylo definování požadavků a cílů jednotlivých oddělení na budoucí řešení Business Intelligence. Následují hlavní požadavky, které byly výsledkem jednání s představiteli společnosti Precheza.

- použít technologické produkty od společnosti Microsoft a to z důvodu sjednocování používaných platforem ve společnosti Precheza a již vlastněných licencí pro potřebné produkty
- použití moderních softwarových prostředků pro tvorbu Business Intelligence řešení
- podoba sestavy v MS Excel by měla být zachována totožná se současně používanou sestavou

7.3 Návrh BI řešení

Na základě požadavků společnosti Precheza byly pro návrh Business Intelligence řešení vybrány produkty společnosti Microsoft. Konkrétně se jedná o produkty SQL Server 2014 Standard Edition, v jejíž licenci je zahrnut i další potřebný produkt pro implementaci Business Intelligence a to SQL Server Data Tools. Posledním produktem bude kancelářský balík MS Office (MS Excel).

V návrhu nové implementace budou zdrojová data čerpána z totožného datového zdroje jako současná implementace, která reprezentuje podnikový informační systém SAP ECC. O přesun uložených zdrojových dat generovaných programem SAP ECC se stará program extraktor, který je implementován v jazyce Java za použití knihoven produktu SAP Connector. Spouštění tohoto programu je realizováno .bat souborem, který vykoná sérii příkazů v příkazovém řádku. Získaná data vstupují do fáze ETL, ve které jsou zpracovávána programem Microsoft SQL Server Integration Services. Tato fáze se vyznačuje velkou časovou náročností, obvykle se jedná o 70% z celkového času implementace. Po vyčištění a potřebné transformaci dat jsou nahrány do datového skladu, který je realizován produktem Microsoft SQL Server 2014. Dále jsou vypočteny olap kostky, které umožní pohled na danou problematiku z více dimenzí a jejich realizace je provedena produktem Microsoft SQL Server Analysis Services. MS Excel je nástroj pro zobrazování finálních sestav, kontingenčních tabulek a reportů, jehož funkcionalit se využívá již v současné implementaci. Obrázek 20 znázorňuje navrhovanou architekturu za použití produktů společnosti Microsoft.



Obr. 20: Komponenty navrhované architektury

7.3.1 Zdrojová data

Pro práci se zdrojovými daty je nutná znalost, v jakých tabulkách a sloupcích jsou data v podnikovém informačním systému uložena. Taktéž je potřeba znát jejich datový typ a základní popis, který informuje o tom, co data představují. Následuje přehled tabulek nacházejících se v SAP ECC ve společnosti Precheza pro ZPL2.

Tabulka SAPR3.PGMI

Obsahem tabulky PGMI je hierarchie hotových výrobků pro ZPL2. Hierarchie je tvořena pomocí rekurzivních odkazů v rámci tabulky PGMI. Všechny základní výrobky musí patřit do 4 skupin, pokud tomu tak není, jedná se o nevhodná data.

Název sloupce	Datový typ	Popis
PRGRP	Varchar (18)	Jméno přímé nadřazené skupiny
NRMIT	Varchar (18)	Jméno skupiny nebo výrobku. Výrobky se rozlišují podle začínajícího znaku 0.

Tab. 1: Tabulka PGMI

Tabulka SAPR3.MAKT

Tabulka MAKT obsahuje textové popisky k jednotlivým výrobkům v ZPL2. Pro volbu jazyka se ve sloupci SPRAS používá filtr. Extrahují se pouze výrobky, které se nacházejí v tabulce PGMI.

Název sloupce	Datový typ	Popis
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
MAKTX	Varchar (40)	Obsahuje celé jméno výrobku nebo skupiny.
SPRAS	Char (1)	Znak určující jazyk záznamů, používán je znak C pro český jazyk.

Tab. 2: Tabulka MAKT

Tabulka SAPR3.MARA

Obsahem tabulky MARA je základní měrná veličina pro konkrétní výrobek. Extrahují se pouze výrobky, které se nacházejí v tabulce PGMI.

Název sloupce	Datový typ	Popis
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
MEINS	Varchar (3)	Obsahuje měrnou veličinu konkrétního výrobku.

Tab. 3: Tabulka MARA

Tabulka SAPR3.MARM

Tabulka obsahuje pro každý výrobek a měrnou veličinu převodní koeficient, který převádí výrobek na základní jednotku materiálu. Tato tabulka je důležitá pro správný výpočet hodnot v tabulce faktů.

Název sloupce	Datový typ	Popis
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
MEINH	Varchar (3)	Základní měrná jednotka materiálu, do které se hodnota převádí.
UMREZ	Decimal (5)	Konstanta pro převod ve vzorci je vždy na pozici dělence.
UMREN	Decimal (5)	Konstanta pro převod ve vzorci je vždy na pozici dělitele.

Tab. 4: Tabulka MARM

Tabulka SAPR3.S031

Obsahem tabulky S031 jsou fyzické pohyby zásob ve skladech. Pro výpočet stavu zásob ve skladech je nutné sumarizovat zásoby ve skladech od počátku používání informačního podnikového systému SAP.

Název sloupce	Datový typ	Popis
VRSIO	Varchar (3)	Znaky, které specifikují verzi plánu.
SPMON	Char (6)	Měsíc a rok ve formátu RRRRMM, který slouží pro výběr níže uvedených hodnot v daném měsíci konkrétního roku.
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
BASME	Varchar (3)	Obsahuje měřnou veličinu konkrétního výrobku. Shodné se sloupcem MEINS v tabulce MARA.
MZUBB	Decimal (15,3)	Množství přírůstku oceněné zásoby
WZUBB	Decimal (15,2)	Hodnota přírůstku oceněné zásoby vyjádřená v Kč
MAGBB	Decimal (15,3)	Množství úbytku oceněné zásoby
WAGBB	Decimal (15,2)	Hodnota úbytku oceněné zásoby vyjádřená v Kč

Tab. 5: Tabulka S031

Tabulka SAPR3.MBEW

Tabulka MBEW obsahuje prodejní cenu pro výrobky, které se nacházejí v tabulce PGMI. Tabulka je určena pro výpočet počátečních zásob ve skladech, které jsou vyjádřené v Kč.

Název sloupce	Datový typ	Popis
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
VPRSV	Char (1)	Znak řízení ceny.
STPRS	Decimal (11,2)	Standardní cena výrobku.
VERPR	Decimal (11,2)	Jedná se o pohyblivou průměrnou cenu / periodickou zúčtovací cenu
PEINH	Decimal (5)	Počet jednotek v ceně.
BWKEY	Char (4)	Konstanta, která je filtrována na hodnotu 1000

Tab. 6: Tabulka MBEW

Tabulka SAPR3.S578

Obsahem tabulky S578 jsou plány a očekávání pro prodej, výrobu, nákup a spotřebu. Plán pro daný rok se počítá jako VRSIO = aktuální rok – 1700. Například pro rok 2016 je hodnota sloupce VRSIO rovna 316.

Název sloupce	Datový typ	Popis
VRSIO	Varchar (3)	Znaky, které specifikují verzi plánu.
SPMON	Char (6)	Měsíc a rok ve formátu RRRRMM, který slouží pro výběr níže uvedených hodnot v daném měsíci konkrétního roku.
PMNUX	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI a sloupcem MATNR.
BASME	Varchar (3)	Obsahuje měrnou veličinu konkrétního výrobku. Shodné se sloupcem MEINS v tabulce MARA.
ABSAT	Double	Plánovaný prodej
PRODU	Double	Plánovaná výroba
MENGE	Decimal (15,3)	Plánovaný nákup
MGVBR	Decimal (15,3)	Plánovaná spotřeba
NETWR	Decimal (15,2)	Plánovaný prodej vyjádřen v Kč

Tab. 7: Tabulka S578

Tabulka SAPR3.CE1PRE1

Tabulka CE1PRE1 obsahuje skutečné prodeje výrobků společnosti Precheza.

Název sloupce	Datový typ	Popis
PERIO	Char (8)	Měsíc a rok ve formátu RRRRMMM, který slouží pro výběr níže uvedených hodnot v daném měsíci konkrétního roku.
ARTNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI a sloupcem MATNR.
ABSMG	Decimal (15,3)	Obsahuje objem provedených nákupů.
ERLOS	Decimal (15,2)	Obsahuje objem provedených nákupů vyjádřený v Kč.
ABSMG_ME	Varchar (3)	Obsahuje měrnou veličinu konkrétního výrobku. Shodné se sloupcem MEINS v tabulce MARA.
VTWEG	Varchar (2)	Jedná se o příznak cest odbytu.
VV999	Decimal (15,2)	Obsahuje objem prodeje.
BUKRS	Char (4)	Jedná se o účetní příznak.
KOKRS	Char (4)	Jedná se o účetní příznak.
VRGAR	Char (1)	Jedná se o účetní příznak.

Tab. 8: Tabulka CE1PRE1

Tabulky SAPR3.MSEG a SAPR3.MKPF

Záznamy v tabulce představují skutečnou výrobu společnosti Precheza. Data jsou získávána z tabulky MSEG a datum je extrahováno z tabulky MKPF, která obsahuje hlavičku dokladu. Tabulky MSEG a MKPF jsou spojovány přes sloupce MJAHR a MBLNR. Vybírány jsou pouze výrobky, které se nacházejí v tabulce PGMI. Do výběru spadají pouze data výroby pro aktuální rok, který je určen tabulkou ZPL2_ETL.ROK_MESIC a pouze data s konkrétními konstantami ve sloupci BWART.

Název sloupce	Datový typ	Popis
BUDAT	Char (8)	Jedná se o datum zapsání záznamu ve formátu RRRRMMDD.
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
MENGE	Decimal (13,3)	Množství vyrobeného výrobku.
MEINS	Varchar (3)	Obsahuje měrou veličinu konkrétního výrobku.
SHKZG	Char (1)	Účetní příznak má dáti / dal
MJAHR	Varchar (4)	Rok vydání materiálového dokladu.
MBLNR	Varchar (10)	Číslo materiálového dokladu.
BWART	Varchar (3)	Druh pohybu (vedení zásob).
KZBEW	Char (1)	Příznak pohybu.

Tab. 9: Tabulky MSEG a MKPF

Tabulka SAPR3.S012

Obsahem tabulky S012 jsou skutečné nákupy společnosti Precheza. Používají se konstantní filtry na sloupce VRSIO = '000' a SPMON <> '00000'. Filtry jsou používány pro odstranění nerelevantních dat, které se vyskytují v datovém zdroji. Vybírány jsou pouze výrobky, které se nacházejí v tabulce PGMI. S podmínkou uskutečněného nákupu v aktuálním roce a neprázdného sloupce BASME.

Název sloupce	Datový typ	Popis
SPMON	Char (6)	Měsíc a rok ve formátu RRRRMM, který slouží pro výběr níže uvedených hodnot v daném měsíci konkrétního roku.
MATNR	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI.
BASME	Varchar (3)	Obsahuje měrou veličinu konkrétního výrobku. Shodné se sloupcem MEINS v tabulce MARA.
VRSIO	Varchar (3)	Znaky, které specifikují verzi plánu.
REMNG_B	Decimal (15,3)	Fakturované množství k datu účtování.

Tab. 10: Tabulka S012

Tabulka SAPR3.S026

Tabulka S026 obsahuje skutečnou spotřebu společnosti Precheza. Používají se konstantní filtry na sloupce VRSIO =' 000' a SPTAG <>' 00000'. Vybírány jsou pouze výrobky, které se nacházejí v tabulce PGMI. S podmínkou provedené spotřeby v aktuálním roce a neprázdného sloupce MCOMP.

Název sloupce	Datový typ	Popis
SPTAG	Char (8)	Jedná se o datum zapsání záznamu ve formátu RRRRMMDD.
MCOMP	Varchar (18)	Jedná se o číslo výrobku, které je shodné se sloupcem NRMIT v tabulce PGMI a sloupcem MATNR.
BASME	Varchar (3)	Obsahuje měrnou veličinu konkrétního výrobku. Shodné se sloupcem MEINS v tabulce MARA.
ENMNG	Decimal (15,3)	Představuje množství výrobku, které bylo odebráno.

Tab. 11: Tabulka S026

7.3.2 Extraktor zdrojových dat

Data, která jsou uložena v podnikovém informačním systému SAP R3 je nutné extrahovat do textových souborů. Umožňuje to s daty dále pracovat ve fázích čištění a transformací dat. K procesu extrakce dat se používá program napsaný v jazyce Java. Extraktor se spouští pomocí bat souboru, který vykoná sérii příkazů (viz obrázek 21) v příkazovém řádku. Dávkový soubor nastaví cestu do adresáře, kde se nachází samotný class soubor a připojí nezbytné sapjco3 knihovny. Poslední příkaz již spouští samotný program s doplněnými parametry.

```
1 @echo off
2 set EXPORT_PATH=C:\SAP_Extractor
3 set PATH=%PATH%;%EXPORT_PATH%
4 set CLASSPATH=%CLASSPATH%;%EXPORT_PATH%\sapjco3.jar;.
5 cd %EXPORT_PATH%
6 "C:\Program Files\jre1.8.0_66\bin\java" sap_extract **
```

Obr. 21: Série příkazů v dávkovém souboru

Po spuštění extraktoru bez zadaných parametrů je v konzoli vypsán informační výstup o povinných vstupních parametrech a v hranatých závorkách uvedených parametrech, které jsou volitelné. Následně jsou popsány jednotlivé parametry a na obrázku 22 je zobrazen konzolový výstup a příklad extrakce tabulky S012.

- **-s sap_connection_file** – soubor s parametry pro připojení do podnikového informačního systému SAP. Například se jedná o název serveru, uživatelské jméno, heslo a jazyk.
- **-t table** – název tabulky, který má být ze SAP exportován
- **-o file** – pojmenování výstupního textového souboru
- **[-d delimiter]** – zvolení oddělovače sloupců ve výstupním souboru, jako výchozí oddělovač se používá znak '|'
- **[-c columns]** – zvolení pouze vybraných sloupců dané tabulky, výběr sloupců je oddělován znakem ','
- **[-w condition]** – možnost zakomponování sql podmínky pomocí klauzule WHERE

```

C:\SAP_Extractor>"C:\Program Files\Java\jre1.8.0_74\bin\java" sap_extract
Error in syntax: -t table name is required
sap_extract syntax:
=====
-s sap_connection_file - file with SAP connection parameters
-t table              - SAP table for exporting
-o file               - output text file
[-d delimiter]       - delimiter in output file
[-c columns]         - columns of exported table, delimited by ,
[-w condition]       - SQL WHERE condition

C:\SAP_Extractor>"C:\Program Files\Java\jre1.8.0_74\bin\java" sap_extract -s des
tination -t S012 -o C:\ZPL2\TXT\s012.txt -c URSIO,SPMON,MATNR,BA$ME,REMNG_B
Progress: 2016-03-21 10:14:37.02
Source: Extrakce S012 Execute Process Task
Operation Complete: 100% complete
End Progress
DTExec: The package execution returned DTSER_SUCCESS (0).
Started: 10:14:00 AM
Finished: 10:14:37 AM
Elapsed: 36.875 seconds

```

Obr. 22: Konzolový výstup s parametry a příklad extrakce

7.3.3 Datový sklad

Na základě schůzky se zástupci společnosti Precheza byly definované požadavky na fakta a dimenze, jejichž sledování je pro společnost důležité. Po provedení potřebných transformací a čištění extrahovaných dat ze zdrojového systému SAP budou data přenášena do datového skladu. Tabulka faktů bude obsahovat metriky zobrazené v tabulce 12.

Název metriky	Popis
Zasoba_poc	Počáteční zásoba neboli konečná zásoba z předchozího měsíce konkrétního výrobku nebo skupiny výrobků.
Zasoba_poc_Kc	Počáteční zásoba neboli konečná zásoba z předchozího měsíce konkrétního výrobku nebo skupiny výrobků, která je vyjádřena v Kč.
Vyroba	Množství vyrobeného produktu nebo jeho skupiny.
Nakup	Množství nakoupeného produktu.
Prodej	Prodej vyrobeného produktu.
Spotreba	Spotřebované množství konkrétního produktu nebo jeho skupiny.
Ostatni	Rozdíl mezi počáteční a koncovou zásobou, který nepochází z výroby, spotřeby, nákupu nebo prodeje.
Prodej_Kc	Prodej vyrobeného produktu, který je vyjádřen v Kč.
Zasoba	Množství výrobku, který byl zhotoven do zásoby. Výpočet je proveden podle vzorce: $Zasoba_poc + Vyroba + Nakup - Prodej - Spotreba + Ostatni$
Zasoba_Kc	Množství výrobku v Kč, které bylo zhotoveno do zásoby.
PC_mes	Prodejní cena za měsíc, která je spočtená podle vzorce: $Prodej_Kc / Prodej$
PC_mes_kum	Prodejní cena kumulativní, která je spočtena podle vzorce: $\sum Prodej_Kc \text{ od začátku roku} / \sum Prodej \text{ od začátku roku}$.

Tab. 12: Metriky obsažené v tabulce faktů

Následují návrhy dimenzí, které budou použity v datovém skladu. Současný návrh je koncipován s dimenzí času, sestavy a výrobku. Při návrhu dimenze času jsou uvažovány tyto charakteristiky: **jméno dimenze**, **popis**, **úroveň**, **nejnižší úroveň** a **hierarchie**. Aplikované charakteristiky pro dimenzi času jsou znázorněny v tabulce 13. Hlavním přínosem časové dimenze je možnost sledování vývoje všech metrik a možnost srovnání v jednotlivých časových obdobích.

Charakteristika	Dimenze čas
Jméno dimenze	čas
Popis	Dimenze je používána pro výběr faktů podle času.
Úrovně	měsíc, kvartál a rok
Nejnižší úroveň	měsíc
Hierarchie	měsíc – kvartál – rok měsíc – rok

Tab. 13: Dimenze času

Každá úroveň časové dimenze obsahuje atributy, které jsou uvedeny v souhrnné tabulce 14.

Název atributu	Popis	Hodnoty
ÚROVEŇ MĚSÍC		
MESIC_PORADI_KVARTAL	Pořadí měsíce v daném kvartálu.	Přípustné hodnoty: 1,2,3
MESIC_PORADI_ROK	Pořadí měsíce v daném roce.	Přípustné hodnoty: 1-12
MESIC_JMENO	Jméno měsíce v češtině bez diakritiky.	Např.: zari
MESIC_JMENO_CS	Jméno měsíce v češtině.	Např.: září
MESIC_ZKRATKA_CS	Zkratka měsíce v češtině.	Např.: zař
MESIC_JMENO_EN	Jméno měsíce v angličtině.	Např.: September
MESIC_ZKRATKA_EN	Zkratka měsíce v angličtině.	Např.: Sep
MESIC_PO CET_DNI	Počet dní v měsíci.	Přípustné hodnoty: 28 - 31
ROK_MESIC	Řetězec YYYYMM.	Např.: 201609
ÚROVEŇ KVARTÁL		
ROK	Rok kvartálu	Např.: 2016
KVARTAL_PORADI_ROK	Pořadí kvartálu v roce	Přípustné hodnoty: 1 - 4
KVARTAL_JMENO	Jméno kvartálu	Např.: 2016/3Q
ÚROVEŇ ROK		
ROK	Rok kvartálu	Např.: 2016
PRESTUPNY_ROK	Je rok přestupný?	Přípustné hodnoty: 0 - 1

Tab. 14: Atributy časové dimenze

Druhou navrhovanou dimenzí je sestava. Cílem této dimenze je rozlišit fakta v jednotlivých sestavách (skutečnost, plán, očekávání yyyy/mm). Vytvoření a naplnění tabulky je nutné provést ručně před spuštěním projektu. Charakteristika dimenze sestava je uvedena v tabulce 15.

Charakteristika	Dimenze sestava
Jméno dimenze	sestava
Popis	Dimenze určuje fakta v jednotlivých sestavách.
Úrovně	sestava
Hodnoty	<ul style="list-style-type: none"> • Skutečnost – metriky reprezentují skutečné prodeje, nákupy, výrobu a další • Plán – metriky představují roční plán, který se v průběhu roku nemění • Očekávání yyyy/mm – každý rok disponuje 12ti řadami očekávání v závislosti na vstupním parametru. Očekávání 2017/04 – metrika počáteční zásoba pro duben se přebírá ze skutečné koncové zásoby v únoru. Očekávání 2017/03 – metrika počáteční zásoba pro duben se přebírá z očekávané koncové zásoby v březnu.

Tab. 15: Dimenze sestava

Dimenze sestava obsahuje pouze dva atributy, které lze spatřit v tabulce 16.

Název atributu	Popis	Hodnoty
SESTAVA	Jméno sestavy.	Např.: Plán, Skutečnost, Očekávání 2016/06, Očekávání 2016/10
POPIS	Podrobnější popis dané sestavy.	

Tab. 16: Atributy dimenze sestava

Poslední navrhovaná dimenze jsou výrobky. Jejím cílem je umožnit sledovat skupiny výrobků nebo konkrétní výrobky v definovaných metrikách. Hierarchie skupin výrobků je převzata z informačního podnikového systému SAP. Charakteristika dimenze výrobky je uvedena v tabulce 16.

Charakteristika	Dimenze výrobky
Jméno dimenze	výrobky
Popis	Dimenze obsahuje hierarchii Hotových výrobků, která je zavedená v podnikovém informačním systému SAP.
Úrovně	skupina4, skupina3, skupina2, skupina1, výrobek
Nejnižší úroveň	výrobek (NRMIT)
Hierarchie	výrobek – skupina1 – skupina2 – skupina3 – skupina4

Tab. 17: Dimenze výrobky

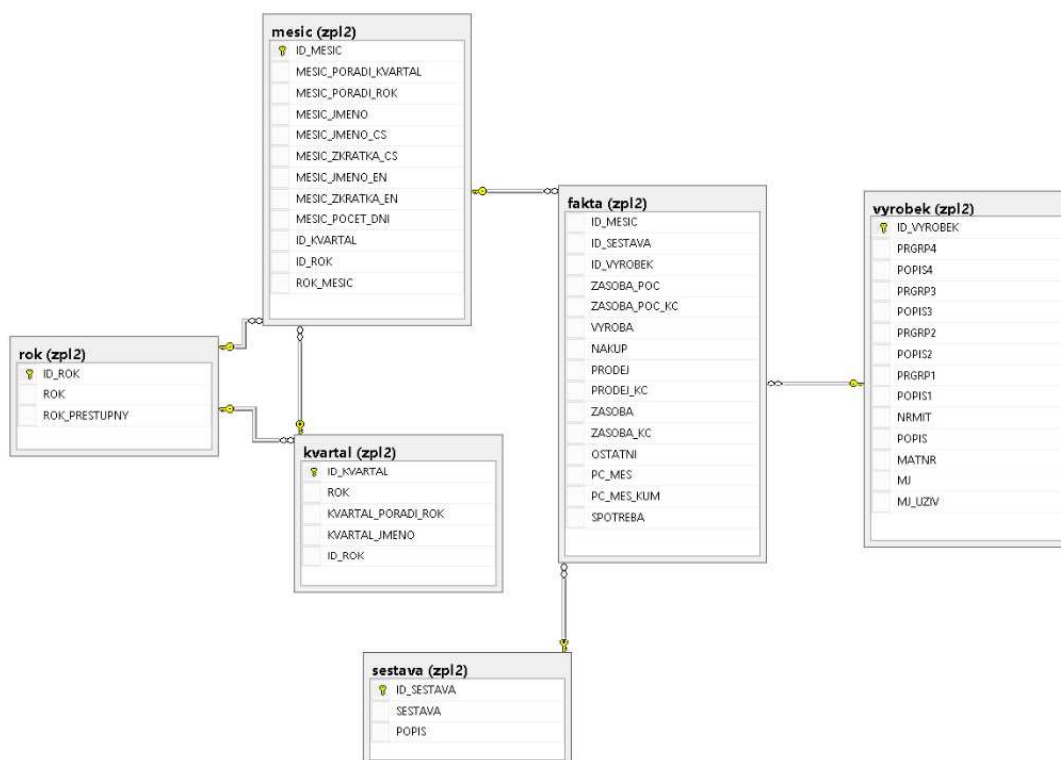
Atributy dimenze výrobky obsahují 4 skupiny, které ve výstupní kontingenční tabulce budou představovat stromovou strukturu skupiny nebo konkrétního výrobku. Mimo jiné obsahuje detailní informace k přesné identifikaci vybraného výrobku. Veškeré atributy jsou popsány v tabulce 18 s konkrétním příkladem hodnot pro výrobek 10001.

Název atributu	Popis	Příklad výrobku 10001
PRGRP4	Skupina 4.	HOTOVE VYROBKY
POPIS4	Popis skupiny 4.	Hotové výrobky
PRGRP3	Skupina 3.	TITANOVA BELOBA
POPIS3	Popis skupiny 3.	Titanová běloba
PRGRP2	Skupina 2.	ANATAS
POPIS2	Popis skupiny 2.	Anatas
PRGRP1	Skupina 1.	AV01
POPIS1	Popis skupiny 1.	AV01
NRMIT	Jméno výrobku.	10001
POPIS	Popis výrobku.	AV01 25 kg pytel rozpuštěné 52x58
MATNR	Interní číslo výrobku v SAPu.	000000000000010001
MJ	Měrná jednotka.	TO
DELENEC	Dělenec převodního vzorce do správné měrné jednotky.	1
DELITEL	Dělitel převodního vzorce do správné měrné jednotky.	1

Tab. 18: Atributy dimenze výroby

Návrh datového skladu

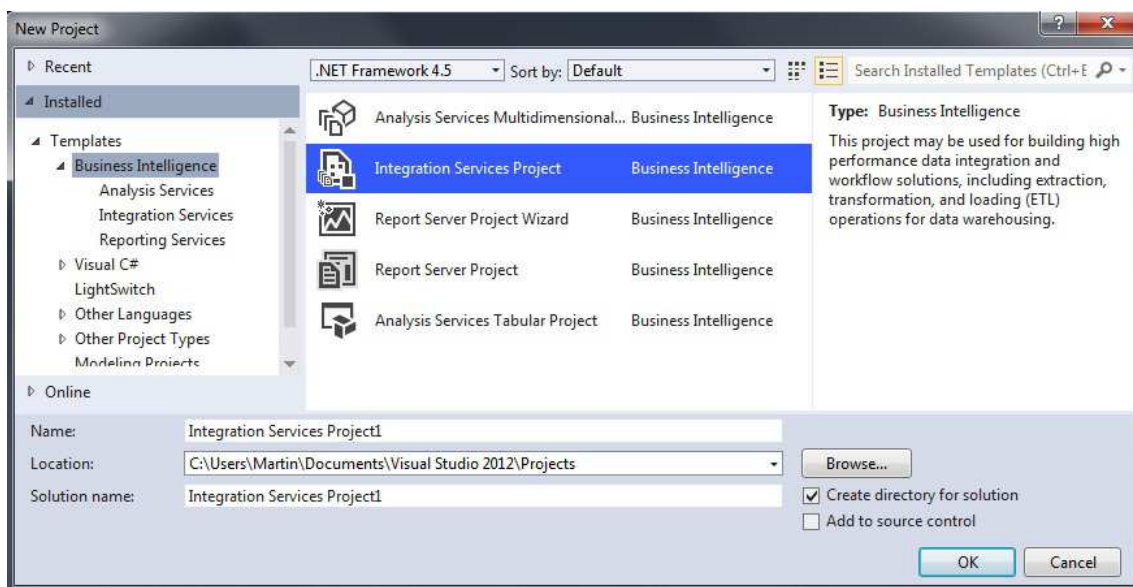
Na základě předcházejících dimenzí a faktů byl navržený dimenzionální model pro realizaci datového skladu. Datový sklad je kombinací schématu vločky, kterou reprezentuje dimenze času a schématu hvězdy pro dimenze sestava a výrobky. Kompletní návrh datového skladu ilustruje obrázek 23.



Obr. 23: Návrh datového skladu

7.3.4 Implementace v SQL Server Integration Services

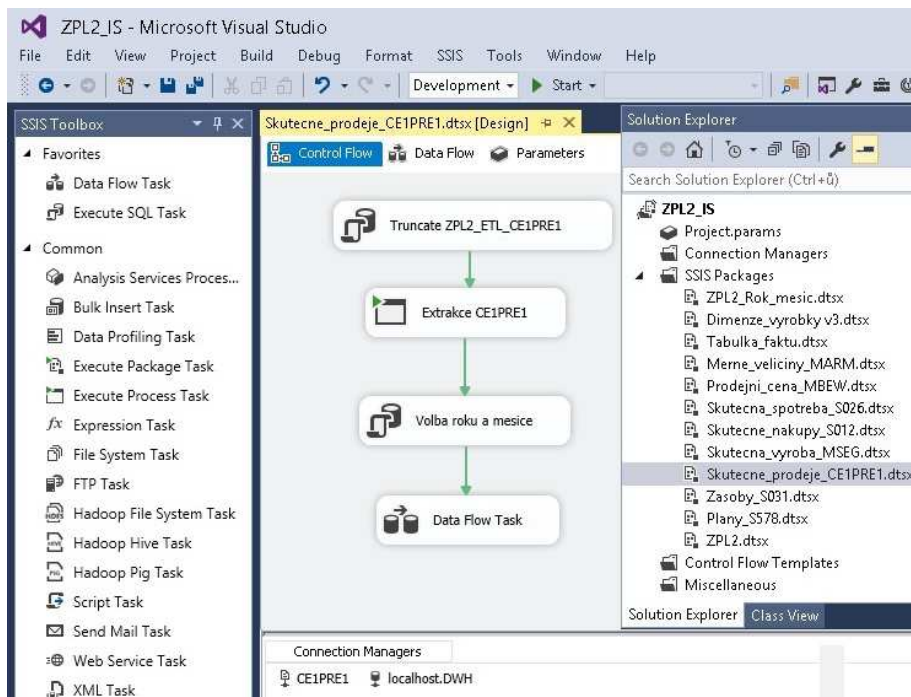
Tvorba fáze ETL se realizuje v programu Microsoft Visual Studio 2012, který je pro tyto účely součástí instalace SQL Server Data Tools. Při jeho spuštění je založen nový projekt, pro který je vybrána nabídka Integration Services viz obrázek 24. Tento typ projektu je přizpůsoben pro čištění a transformování zdrojových dat, která jsou poté nahrána do datového skladu.



Obr. 24: Založení nového projektu pro tvorbu ETL procesu

Po založení projektu se zobrazí hlavní stránka, na které jsou dvě základní záložky pro tvorbu ETL procesu. Jedná se o Control Flow a Data Flow.

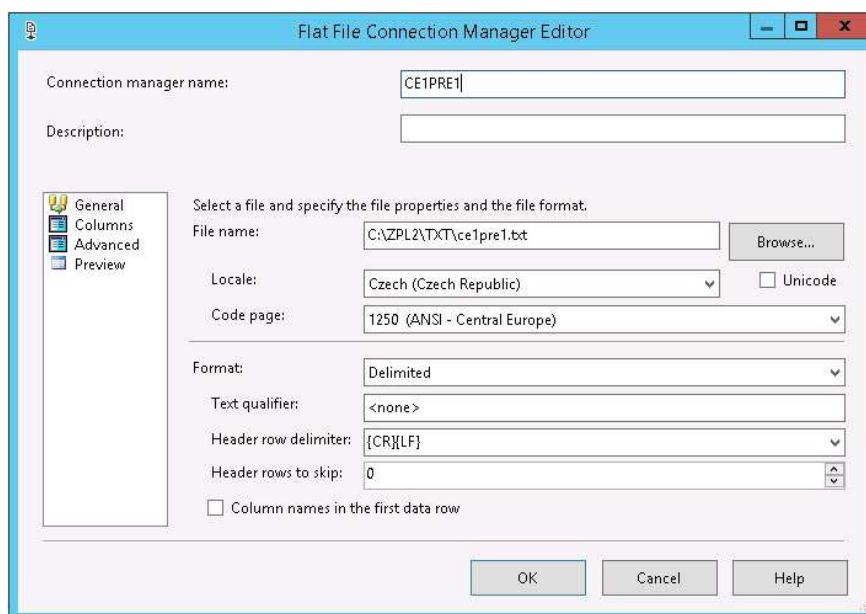
Control Flow definuje průtok úkolů, které mají být provedeny. Například se může jednat o provedení cyklů, spuštění SQL skriptů, zavolání COM rozhraní, provedení .NET komponent nebo odeslání emailu. Taktéž zde může být uveden pouze odkaz do Data Flow, ve kterém je umožněno pracovat s daty. Pro názornost je vybrána implementovaná package Skutecne_prodeje_CE1PRE1, ve které je použito spuštění SQL skriptu (Truncate ZPL2_ETL_CE1PRE1), jenž vyčistí pomocnou tabulku (zpl2.etl_ce1pre1) v databázi. Následuje provedení procesového úkolu (Extrakce CE1PRE1), který spustí dávkový soubor sap_extract.bat se zadanými parametry. Výsledkem jsou data v textovém souboru pro použití v Data Flow. Pro filtraci dat podle měsíce a roku je nutné tyto data SQL skriptem získat z databáze do parametrů v Integration Services. Poslední úkol je odkaz do Data Flow záložky. Celý proces je zobrazen na obrázku 25.



Obr. 25: Control Flow pro Skutecne_prodeje_CE1PRE1.dtsx

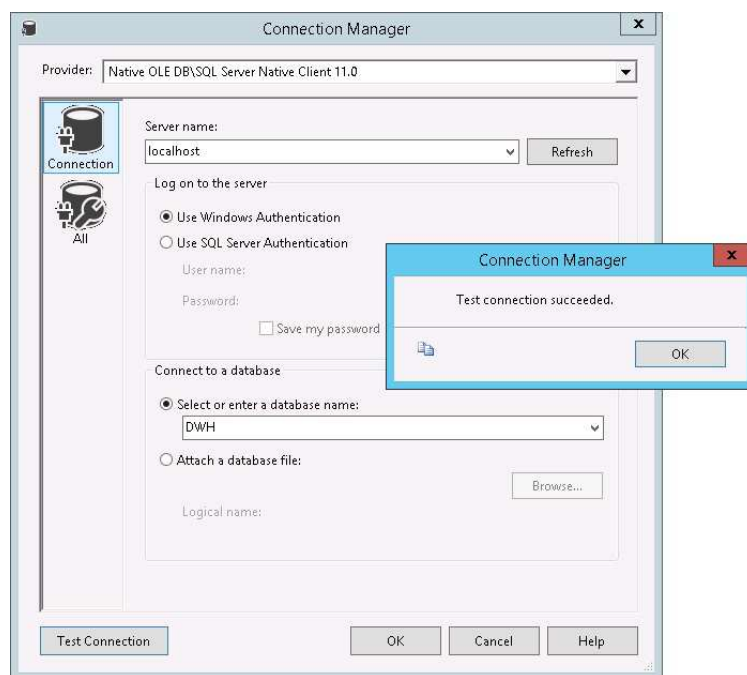
Data Flow definuje tok dat ze zdrojového souboru do cílové destinace. Data mohou pocházet z více zdrojových souborů a bohatá nabídka funkcí umožňuje jejich spojování, transformování, filtrování, třídění a mnoho dalších úkonů. Pro názornost je uvedeno pokračování implementace package Skutecne_prodeje_CE1_PRE1.

Data Flow pro Skutecne_prodeje_CE1_PRE1 začíná nahráním zdrojových dat v textovém formátu, které je realizované úlohou Flat File Source (CE1PRE1). Pro její správné fungování je nutné nastavit nový Flat file connection manager, ve kterém se specifikuje zdrojový textový soubor, místní umístění, kódování textu a oddělovací prvek záznamů (viz obrázek 26). Následně je nutné jednotlivé sloupce pojmenovat, protože zdrojový soubor v prvním řádku neobsahuje jméno příslušných sloupců. V samotném Flat File Source se vybere pouze odpovídající connection manager a jsou vybrány sloupce, které budou procházet procesem čištění a transformace.



Obr. 26: Nastavení Flat File Connection Manageru pro CE1PRE1

Druhou vstupní úlohou je OLE DB Source (Zpl2 výrobek), kterou je získáván aktuální seznam výrobků požadovaných pro analýzu zpl2. Tento proces je zpracováván v package Dimenze_vyroby v3.dtsx, jehož výstup je ukládán do databázové struktury a který je nyní použit jako jeden z vstupů do package Skutecne_prodeje_CE1_PRE1.dtsx. Taktéž jako v předchozím případě je nutné nastavit nový OLE DB connection manager. V první položce je vybráno jméno serveru, ke kterému je prováděno připojení. Následuje výběr druhu autentifikace, pro účely projektu je používána Windows Authentication. Posledním bodem je výběr databáze, ve které se nachází požadovaná tabulka. Pro ověření připojení k serveru lze využít tlačítko Test Connection, které informuje o výsledku pokusu připojení k vybranému serveru. Connection manager a výsledek testu připojení ilustruje obrázek 27.



Obr. 27: Nastavení OLE DB Connection Manageru pro Zpl2 výrobek

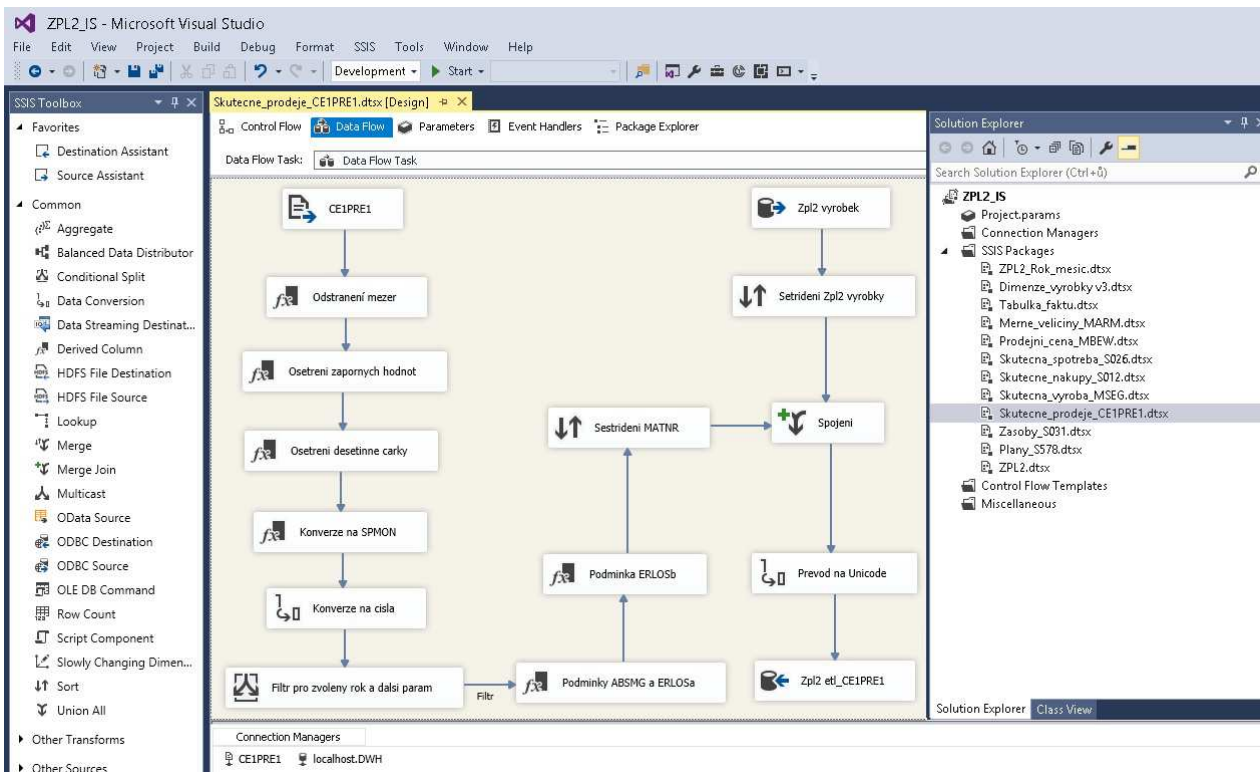
Po načtení dat ze zdrojového textového souboru je nutné data upravit v následujících pěti funkčních úlohách:

- **Odstranění mezer** – z důvodu možnosti výskytu bílých znaků na začátku a konci dat v jednotlivých sloupcích je ve funkci použit výraz `LTRIM(RTRIM("jméno sloupce"))`, který odstraní zmíněné bílé znaky.
- **Ošetření záporných hodnot** – sloupce `ABSMG`, `VV999` a `ERLOS` obsahují číselné hodnoty, které záporné hodnoty rozlišují přidáním znaku `"-"` za číselnou hodnotu. Pro účely analýzy je nutné znak přesunout na začátek číselné hodnoty. Ve funkci se používá výraz, který zjistí, zda-li se na konci číselné hodnoty nachází znak `"-"` a pokud ano, tak ho přesune před číselnou hodnotu.
- **Ošetření desetinné čárky** – z důvodu toho, že všechny zdrojová data jsou datového typu `string`, je nutný jejich převod na datový typ `numeric`. Pro umožnění tohoto převodu je vyžadováno nahrazení desetinné tečky za desetinnou čárku. Funkce používá následující výraz pro zmíněné nahrazení `REPLACE(LTRIM(RTRIM("jméno sloupce")),",",".")`.

- **Konverze na SPMON** – sloupec PERIO, který je ve formátu RRRRMMM, je nutné převést na formát RRRRMM. Používá se výraz pro výběr prvních čtyř znaků a posledních dvou znaků, který je zapsán následovně: SUBSTRING(SPMON,1,4) + SUBSTRING(SPMON,6,2)
- **Konverze na čísla** – všechny datové zdroje jsou při vstupu do Data Flow v datovém typu string, jak již bylo zmíněno u ošetření desetinné čárky. Proto hodnoty, které obsahují číselný formát, jsou převáděny na datový typ numeric.

Další použitou úlohou je Conditional Split (Filtr pro zvoleny rok a dalsi param), který má funkci filtrování dat. Filtrace je prováděna podle uživatelem zvoleného roku, vybraných ekonomických příznaků a nenulové hodnoty sloupce ERLOS. Data, která splnila všechny podmínky, pokračují do následujících dvou funkčních úloh. V těchto funkcích jsou jednotlivým sloupcům na základě podmínky přiřazována data z jiných sloupců. Bohužel podrobnější informace nemohou být zveřejněny z důvodu citlivosti použitých dat v podmínkách.

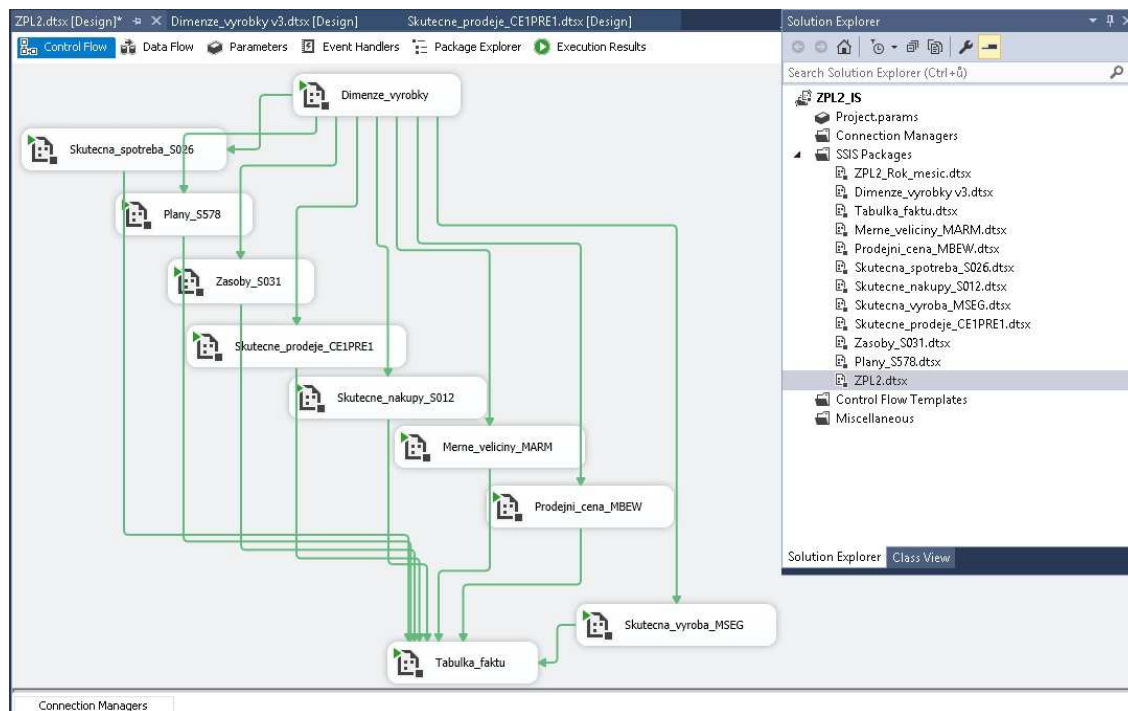
Před provedením úlohy Merge Join (Spojení) jsou vykonány úlohy Sort (Sestřidení MATNR a Setřidení Zpl2 výrobky), které setřídí data podle sloupce MATNR. Klíčem pro spojení tabulek je sloupec MATNR a výsledkem spojení je výběr sloupců nutných pro analýzu zpl2. Finální úlohou je nahrání vyčištěných a transformovaných dat do databázové struktury, které je realizováno úlohou OLE DB Destination (Zpl2 etl_CE1PRE1). Celé schéma procesu v Data Flow pro package Skutecne_prodeje_CE1_PRE1.dtsx je na obrázku 28.



Obr. 28: Data Flow pro Skutecne_prodeje_CE1PRE1.dtsx

Aktualizace dat v ZPL2

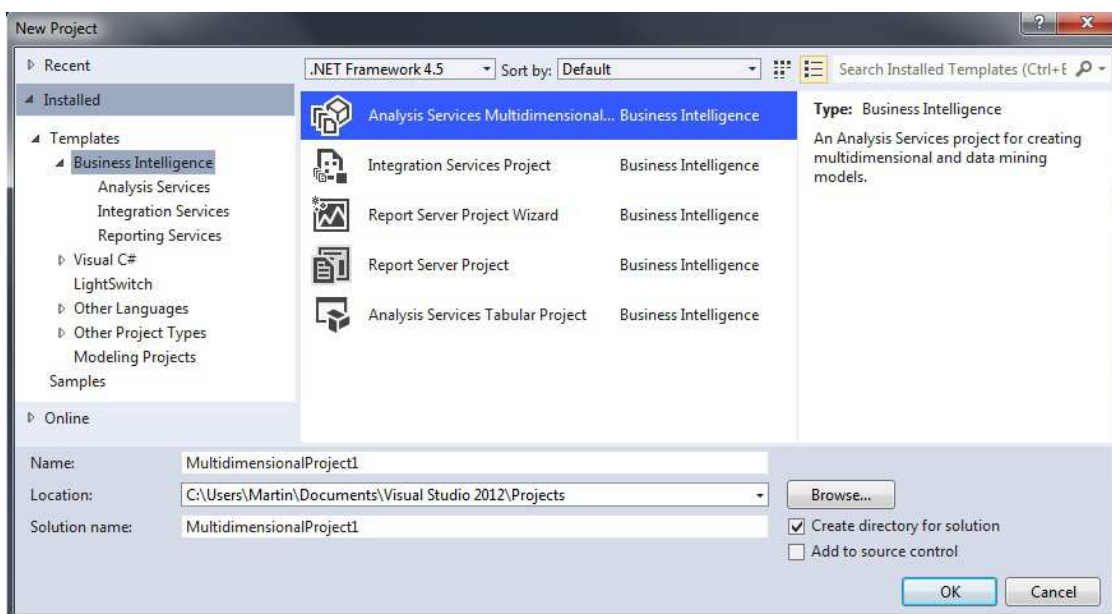
Z důvodu rozdělení projektu do jednotlivých package je nutné vytvoření package, která spustí kompletní proces ETL pro celý projekt ZPL2. Tento problém řeší package s názvem ZPL2.dtsx, která ve svém Control Flow využívá úlohy Execute Package pro odkazování se k jednotlivým package v projektu. Hierarchické schéma procesu aktualizace ilustruje obrázek 29. Spuštění package ZPL2.dtsx je prováděno automatickým procesem z prostředí Microsoft SQL Server Management Studio. Konkrétně ho spouští SQL Server Agent, ve kterém je vytvořen nový job pro spuštění aktualizace datových zdrojů. Podle požadavků zákazníka je tento job spuštěn každý den v nočních hodinách, kdy je nejmenší zatížení serverů.



Obr. 29: Package ZPL2.dtsx pro proces aktualizace

7.3.5 Implementace v SQL Server Analysis Services

Tvorba OLAP kostek je taktéž realizovaná pomocí programu Microsoft Visual Studio 2012, ve kterém byl založen nový projekt s názvem ZPL2_AS. Při zakládání projektu byla vybrána nabídka Analysis Services Multidimensional and Data Mining viz obrázek 30. Tento typ projektu je přizpůsoben pro tvorbu multidimenzionálních modelů, které poté pomáhá sledovat fakta z různých dimenzí.



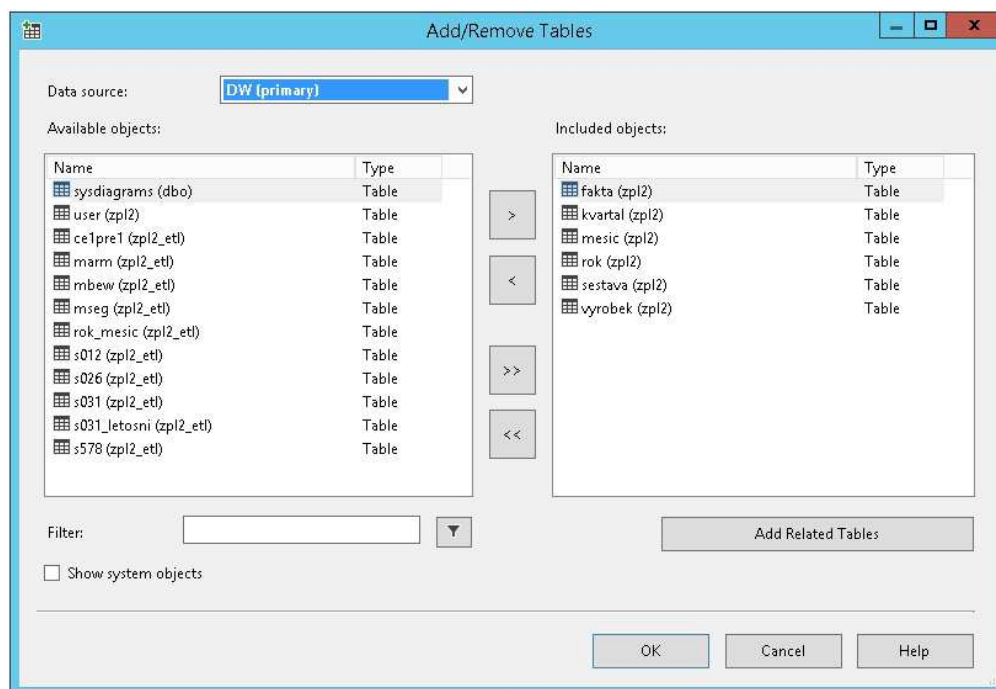
Obr. 30: Založení nového projektu pro tvorbu OLAP kostky

Datové zdroje

Prvním krokem po vytvoření projektu je definice připojení ke zdroji dat, které je umístěné po pravé straně v Solution Explorer, konkrétně v záložce Data Sources. Zdrojová data jsou čerpána z databázové struktury, kam se byla uložena po provedení ETL procesu. Připojení do databázové struktury je shodné jako v případě vytváření Connection Manageru k tabulce Zpl2 vyrobek, který je na obrázku 27.

Pohled na datové zdroje

V Solution Explorer se nachází záložka Data Source Views, pomocí které je vytvořeno zobrazení zdrojových dat. Prostřednictvím průvodce jsou do schématu vloženy požadované tabulky z databázové struktury spolu se vzájemnými relacemi. Pro tvorbu OLAP kostky jsou vybrány následující tabulky: mesic, kvartal, rok, sestava, vyrobek a fakta. Výběr tabulek je graficky znázorněn na obrázku 31.

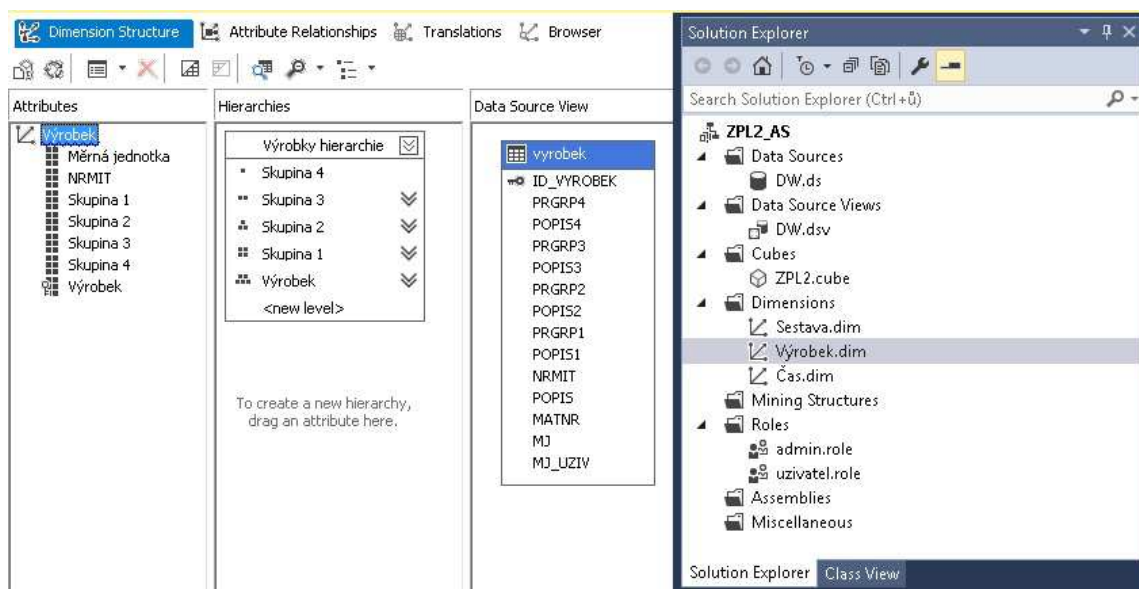


Obr. 31: Výběr tabulek pro OLAP kostku z relační databáze

Dimenze

V Solution Explorer se nachází Dimensions, ve které se definují dimenze budoucí OLAP kostky. Pro projekt ZPL2 jsou vytvořené dimenze času, sestavy a výrobku. V záložce Dimension Structure jsou zobrazeny atributy, které dimenze obsahuje. Prostřední část obrazovky je určena pro vytváření hierarchických struktur

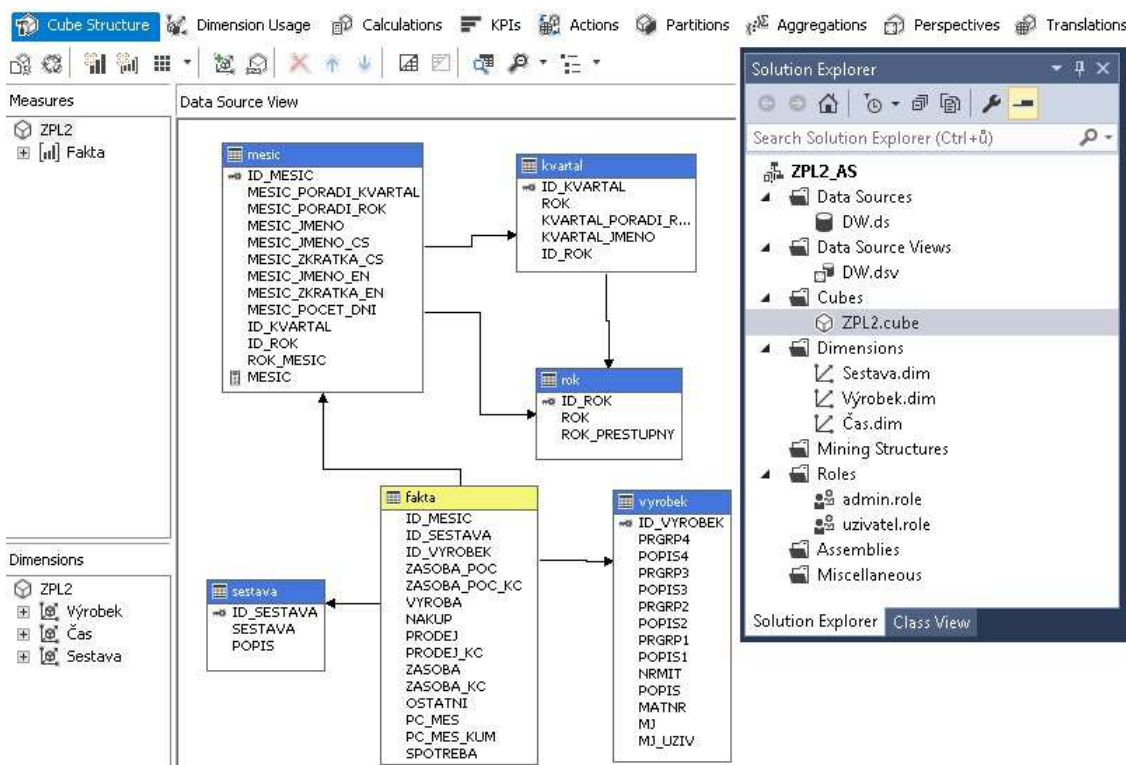
z dostupných atributů. Pravá část obrazovky zobrazuje pohled na datové zdroje. Názorným příkladem je obrázek 32, na kterém je implementace dimenze Výrobek.



Obr. 32: Implementace dimenze Výrobek

OLAP kostka

Vytvoření nové OLAP kostky probíhá v části Cubes, která se nachází v Solution Explorer. Po iniciaci vytvoření nové kostky se spustí průvodce jejím vytvářením. V prvním kroku je vybrána možnost použití existujících tabulek pro tvorbu kostky. Druhý krok obsahuje výběr faktové tabulky, pro projekt ZPL2 je to tabulka fakta. Následuje výběr dimenzí, které byly vytvořeny v předcházejícím kroku a poté je zobrazena finální rekapitulace. Vytvořená OLAP kostka v záložce Cube Structure na levé straně informuje o obsažených faktech a dimenzích. Ve středu schématu je zobrazen pohled na datové zdroje, které jsou součástí OLAP kostky, jak je znázorněné na obrázku 33.

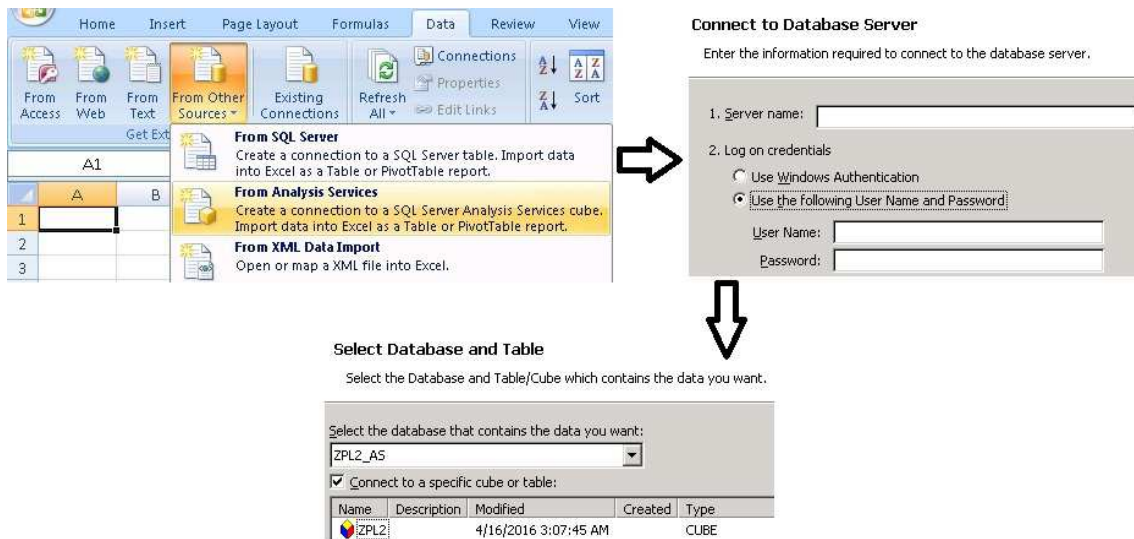


Obr. 33: Implementace OLAP kostky ZPL2

7.3.6 Výstup v Microsoft Excel

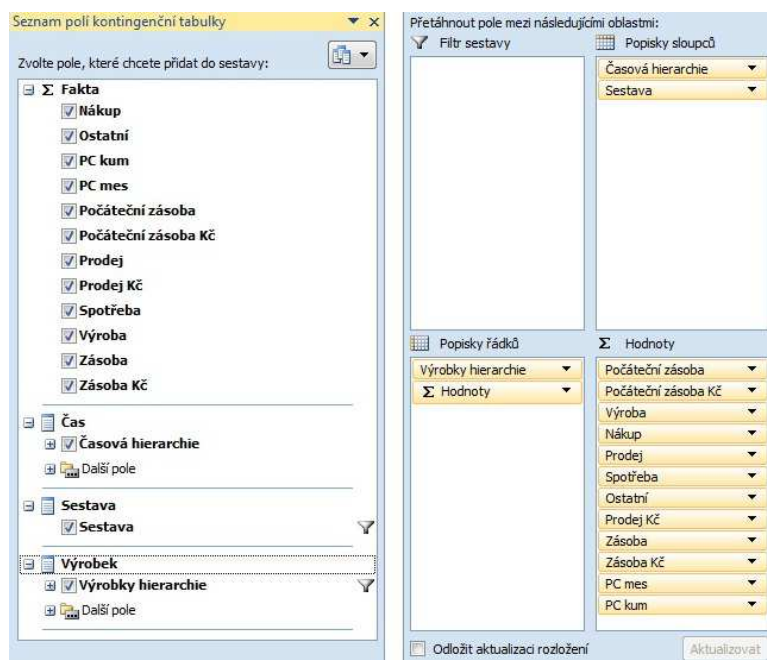
Implementovaná OLAP kostka ZPL2, která je naplněná vyčištěnými a transformovanými daty z procesu ETL je dostupná pro analýzu. Analýzu dat umožňuje zpracovávat mnoho softwarových produktů. Jedním z požadavků společnosti Precheza na projekt ZPL2 je provádění analýzy v programu Microsoft Excel. Důvodem je dlouhodobá spolupráce se společností Microsoft a zkušenosti zaměstnanců controllingu s tímto produktem.

Nahrání OLAP kostky do prostředí programu Microsoft Excel je prováděno přes záložku Data. V oblasti Get External Data je zvolena položka From Other Sources. Po kliknutí se rozbálí nabídka s možnostmi datových zdrojů pro zobrazení Microsoft Excel. Data budou čerpána z SQL Server Analysis Services, proto je pro další postup zvolena položka From Analysis Services. V průvodci připojení k databázovému serveru, kde se OLAP kostka nachází, vyplníme jméno zdrojového serveru a přihlašovací jméno s heslem. Po úspěšném přihlášení je nabídnuta dostupná OLAP kostka ZPL2. Schéma celého postupu připojení je znázorněno na obrázku 34.



Obr. 34: Postup nahrání OLAP kostky do Microsoft Excel

OLAP kostka je v prostředí Microsoft Excel interpretována pomocí kontingenční tabulky. Pravá část programu je vyhrazena pro výběr faktů a dimenzí zobrazovaných v kontingenční tabulce. Nastavení umožňuje vybraná fakta či dimenze zobrazovat v řádcích nebo sloupcích. Příkladem je nastavení kontingenční tabulky projektu ZPL2, kde se ve sloupcích zobrazuje dimenze času a sestavy, v řádcích se zobrazuje dimenze výrobek a fakta (viz obrázek 35).



Obr. 35: Nastavení kontingenční tabulky pro projekt ZPL2

Samotná kontingenční tabulka umožňuje filtrování požadovaných výrobků, času nebo sestavy. Kontingenční tabulka se všemi výrobky bude zaměstnancům společnosti

Precheza sloužit jako podklad pro zhotovování plánů výroby. Na obrázku 36, je zobrazená stromová struktura dimenze výrobku, na jejímž konci se nachází konkrétní výrobek 10004. Ve sloupcích je zobrazena struktura času, která končí nejmenší jednotkou měsíce. Pro každý měsíc jsou vypsány tři ukazatele z dimenze sestava. Číselné hodnoty v kontingenční tabulce musely být vymazány z důvodu ochrany citlivých dat společnosti Precheza.

	A	B	C	D	E	F
1	Popisky sloupců					
2	2016					
3	2016/1Q					
4	1/2016			2/2016		
5	Popisky řádků	Skutečnost	Plán	Očekávání 2016 03	Skutečnost	Plán
6	HOTOVÉ VÝROBKY					
7	TITANOVÁ BĚLOBA					
8	ANATAS					
9	AV01FG					
10	10004					
11	Počáteční zásoba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
12	Počáteční zásoba Kč	xxxxxx	x: Počáteční zásoba			
13	Výroba	xxxxxx	x: Hodnota: xxxxxx			
14	Nákup	xxxxxx	x: Řádek: HOTOVÉ VÝROBKY - TITANOVÁ BĚLOBA - ANATAS - AV01FG - 10004 - Počáteční zásoba			
15	Prodej	xxxxxx	x: Sloupec: 2016 - 2016/1Q - 1/2016 - Plán			
16	Spotřeba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
17	Ostatní	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
18	Prodej Kč	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
19	Zásoba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
20	Zásoba Kč	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
21	PC mes	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
22	PC kum	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
23	AV01FG Počáteční zásoba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
24	AV01FG Počáteční zásoba Kč	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
25	AV01FG Výroba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
26	AV01FG Nákup	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
27	AV01FG Prodej	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
28	AV01FG Spotřeba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
29	AV01FG Ostatní	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
30	AV01FG Prodej Kč	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
31	AV01FG Zásoba	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
32	AV01FG Zásoba Kč	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
33	AV01FG PC mes	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
34	AV01FG PC kum	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx

Obr. 36: Zobrazení kontingenční tabulky s nastaveným filtrem pro výrobek 1004

8 ZÁVĚR

Cílem diplomové práce bylo vytvořit Business Intelligence řešení pro společnost Precheza a.s. Projekt byl zpracováván ve společnosti MF Servis s.r.o. v rámci pracovního úvazku. První verze implementovaného řešení byla předána společnosti Precheza, která v současné době provádí testování a ověřování správnosti získávaných dat.

Teoretická část práce seznamuje čtenáře s obecnými principy technologie Business Intelligence. Ukazuje moderní trendy Business Intelligence, které jsou zpracovávány pomocí technologie cloud computing. Tyto trendy jsou srovnány s klasickým přístupem tvorby Business Intelligence. Z porovnání vyplynulo, že každý přístup má své výhody a nevýhody a bez znalosti potřeb zákazníka nelze rozhodnout, který přístup je pro tvorbu Business Intelligence projektů vhodnější. Podrobně jsou popsány jednotlivé části procesu ETL a přístupy k tvorbě datových skladů podle W. H. Inmona a R. Kimballa. Následuje seznámení se třemi technologickými společnostmi a jejich produkty pro tvorbu Business Intelligence. Podle preferencí a požadavků představitelů společnosti Precheza bylo vytvořeno Business Intelligence řešení, které využívá produkty společnosti Microsoft.

Praktická část diplomové práce se zabývá implementací projektu pro společnost Precheza. Projekt zahrnoval analýzu stávajícího řešení, požadavky na nové řešení od zaměstnanců společnosti, popis a extrakce zdrojových dat nacházejících se v podnikovém informačním systému. Následoval proces čištění a transformace získaných dat, jehož výstup byl nahrán do navrhnutého datového skladu. Datový sklad se stal důvěryhodným zdrojem pro čerpání dat při tvorbě OLAP kostky. Výsledným produktem je kontingenční tabulka, která čerpá data z vytvořené OLAP kostky.

Zaměstnanci v oblasti controllingu výsledný produkt používají jako zdroj informací pro zhotovování plánů výroby, prodeje, spotřeby a dalších metrik společnosti Precheza pro nadcházející roky. Budoucí vývoj projektu bude vycházet z nových požadavků a možnosti zpracování dalších oblastí zdrojových dat.

Zpracování diplomové práce splnilo má veškerá počáteční očekávání a přineslo mi novou zkušenosti z podnikového prostředí. Zároveň jsem nabyl mnoho nových informací a znalostí z oblasti Business Intelligence. Této problematice bych se i nadále rád věnoval.

LITERATURA

- [1] ANUPINDI, NAGESH .V. *Inmon vs. Kimball - An Analysis*[online]. 2005 [cit. 2016-02-24]. Dostupné z: <http://www.nagesh.com/publications/technology/173-inmon-vs-kimball-an-analysis.htm>
- [2] ATTUNITY.COM. *Amazon Redshift & Transformations* [online]. 2016 [cit. 2016-01-28]. Dostupné z: <http://www.attunity.com/amazon-redshift-transformations>
- [3] AUSTIN, Benny. Kimball and Inmon DW Models. *bennyaustin.wordpress.com* [online]. 2.5.2010 [cit. 2016-01-08]. Dostupné z: <https://bennyaustin.wordpress.com/2010/05/02/kimball-and-inmon-dw-models/>
- [4] BIERE, Mike. *Business intelligence for the enterprise*. London: Prentice Hall PTR, 2003, x, 222 p. ISBN 0131413031.
- [5] CODD E.F., CODD S.B., and SALLEY C.T.. Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. *Hyperion Solutions*. [online]. 1993 [cit. 2016-01-11]. Dostupné z: http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf
- [6] DATAWAREHOUSE4U. *Star and Snowflake schema* [online]. [cit. 2016-02-28]. Dostupné z: <http://datawarehouse4u.info/>
- [7] DATAWAREHOUSING AND OLAP TECHNOLOGY. *International Journal of Engineering Research and Applications* [online]. Mar-Apr 2012, 955-960 [cit. 2016-01-11]. ISSN: 2248-9622. Dostupné z: http://www.ijera.com/papers/Vol2_issue2/FD22955960.pdf
- [8] ESUASIVE. *Enterprise BI and Data Warehouse*[online]. [cit. 2016-03-07]. Dostupné z: <http://www.esuasive.co.uk/bi-analytics/enterprise/>
- [9] GOODDATA.COM. *GoodData platform* [online]. 2016 [cit. 2016-01-28]. Dostupné z: <http://www.gooddata.com/platform>
- [10] HAVLENA Matouš. *Business Intelligence in IBM Cognos 10* [online]. [cit. 2016-03-08]. Dostupné z: <http://www.havlana.net/en/business-analytics-intelligence/business-intelligence-in-ibm-cognos-10/>
- [11] HAYNES, George. *IaaS, PaaS, SaaS, & the Cloud 101*[online]. 2014 [cit. 2016-01-18]. Dostupné z: <https://www.linkedin.com/pulse/20140907071547-305726885-iaas-pass-saas-the-cloud-101>
- [12] INEKON-SYSTEMS.CZ. *Tableau Partner* [online]. 2016 [cit. 2016-01-25]. Dostupné z: <http://www.inekon-systems.cz/tableau/>

- [13] INMON, William H. *Building the data warehouse*. 3rd ed. New York: J. Wiley, 2002. ISBN 0471081302.
- [14] INMON, William H. *Building the operational data store*. 2nd ed. New York: John Wiley, 1999, xix, 315 p. ISBN 047132888x.
- [15] JOSYULA, Venkata, Malcolm ORR a Greg PAGE. *Cloud computing: automating the virtualized data center*. Indianapolis, IN: Cisco Press, 2012, xix, 371 p. ISBN 1587204347.
- [16] KIMBALL, Ralph a Joe CASERTA. *The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis, IN: Wiley, 2004, xxxiv, 491 p. ISBN 0764567578.
- [17] KIMBALL, Ralph a Margy ROSS. *The data warehouse toolkit: the complete guide to dimensional modeling*. 2nd ed. New York: Wiley, 2002, xxiv, 436 p. ISBN 0471200247.
- [18] LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012, 350 s. ISBN 978-80-251-3729-1.
- [19] LACKO, Ľuboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Vyd. 1. Brno: Computer Press, 2009, 456 s. ISBN 978-80-251-2887-9.
- [20] MELL, Peter, GRANCE, Tomothy. *The NIST Definition of Cloud Computing* [online]. In: National Institute of Standards and Technology, 2011 [cit. 2016-01-18]. Dostupné z: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- [21] MUNDY, Joy, Warren THORNTHWAITE a Ralph KIMBALL. *The Microsoft data warehouse toolkit: with SQL Server 2005 and the Microsoft Business Intelligence toolset*. Indianapolis: Wiley Publishing, 2006. ISBN 978-0-471-26715-7.
- [22] NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business intelligence: jak využít bohatství ve vašich datech*. 1. vyd. Praha: Grada, 2005, 254 s. Management v informační společnosti. ISBN 80-247-1094-3.
- [23] PEDERSEN, Torben Bach; S. Jensen, Christian (December 2001). "Multidimensional Database Technology". *Distributed Systems Online* (IEEE): 40–46. ISSN 0018-9162
- [24] PETERKA, Miloslav. Seznamte se s BI. In: *DAQUAS* [online]. 9.6.2010 [cit. 2016-01-09]. Dostupné z: <http://www.daquas.cz/articles/379-seznamte-se-s-bi>
- [25] POWER, D.J. A Brief History of Decision Support Systems. *DSSResources.COM*. [online]. 10.3.2007 [cit. 2016-01-07]. Dostupné z: <http://dssresources.com/history/dsshistory.html>

- [26] POWERBI.MICROSOFT.COM. *Power BI Support* [online]. 2016 [cit. 2016-01-25].
Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/>
- [27] POZDÍLEK, Martin. *Studijní materiály předmětu BIN* [online]. 2015 [cit. 2016-02-03].
Dostupné z: <http://www.oliva.uhk.cz>
- [28] PRECHEZA - HISTORIE. *Precheza* [online]. [cit. 2016-03-15]. Dostupné z:
<http://www.precheza.cz/historie/>
- [29] PRECHEZA - PRODUKTY. *Precheza* [online]. [cit. 2016-03-15]. Dostupné z:
<http://www.precheza.cz/produkty/>
- [30] PRECHEZA - PROFIL. *Precheza* [online]. [cit. 2016-03-16]. Dostupné z:
<http://www.precheza.cz/profil/>
- [31] SALLAM, Rita L., Bill HOSTMANN, Kurt SCHLEGEL, Joao TAPADINHAS, Josh PARENTEAU a Thomas W. OESTREICH. *Magic Quadrant for Business Intelligence and Analytics Platforms* [online]. 2015 [cit. 2016-01-18]. Dostupné z:
<http://www.gartner.com/technology/reprints.do?id=1-2ACLP1P&ct=150220&st=sb>
- [32] SAP BI. *Saptanet* [online]. [cit. 2016-03-08]. Dostupné z:
<http://www.saptanet.com/Saptanet/SapBi>
- [33] SAP COMMUNITY NETWORK. *SAP NetWeaver Technology Platform* [online]. [cit. 2016-03-08]. Dostupné z: <http://scn.sap.com/community/netweaver>
- [34] SOSINSKY, Barrie A. *Cloud computing bible*. Indianapolis, IN: Wiley, 2011, xxviii, 497 p. ISBN 978-0-470-90356-8.
- [35] SWISSFORTKNOX.COM. *Swiss Fort Knox* [online]. 2016 [cit. 2016-01-25]. Dostupné z:
<http://www.swissfortknox.com/>
- [36] TABLEAU.COM. *Tableau* [online]. 2016 [cit. 2016-01-25]. Dostupné z:
<http://www.tableau.com/>
- [37] THOMSEN, Erik. *OLAP solutions: building multidimensional information systems*. 2nd ed. New York: Wiley Computer Pub., 2002, xxiv, 661 p. ISBN 0471400300.
- [38] WEBLOG.9C.CZ. *Co je ta Keboola a GoodData?* [online]. 2013 [cit. 2016-01-28].
Dostupné z: <http://weblog.9c.cz/2013/03/novinky/co-je-ta-keboola-a-gooddata/>

SEZNAM OBRÁZKŮ

Obr. 1: Zobrazení integrovanosti datového sklade dle Inmona; Zdroj: [14].....	12
Obr. 2: Schéma modelu Ralpa Kimballa; Zdroj: [3].....	13
Obr. 3: Obecný koncept architektury Business Intelligence; Zdroj: [24]	14
Obr. 4: Definice cloud computing podle NIST; Zdroj: [34]	20
Obr. 5: Hierarchie modelu služeb a konkrétní příklady; Zdroj: [11]	22
Obr. 6: Datové centrum Swiss Fort Knox; Zdroj: [35]	24
Obr. 7: Magický kvadrant pro oblast Business Intelligence; Zdroj: [31].....	25
Obr. 8: ETL schéma zpracování dat; Zdroj: [2].....	28
Obr. 9: Příklad volby úrovně granularity; Zdroj: [13].....	36
Obr. 10: Centrální podnikový datový sklad; Zdroj: [18].....	38
Obr. 11: Koncept datového skladu podle R. Kimballa; Zdroj: [17].....	39
Obr. 12: Příklad modelu ve schématu hvězdy; Zdroj: [6].....	44
Obr. 13: Příklad modelu ve schématu vločky; Zdroj: [6].....	45
Obr. 14: Model BI podle společnosti Microsoft; Zdroj: [8].....	47
Obr. 15: Model BI podle společnosti IBM; Zdroj: [10].....	48
Obr. 16: Model BI podle společnosti SAP; Zdroj: [32]	49
Obr. 17: Vlevo z počátku 19. století a vpravo z počátku 21. století; Zdroj: [28].....	50
Obr. 18: Metriky agregující se v čase	52
Obr. 19: Komponenty architektury současné implementace BI.....	53
Obr. 20: Komponenty navrhované architektury.....	55
Obr. 21: Série příkazů v dávkovém souboru.....	60
Obr. 22: Konzolový výstup s parametry a příklad extrakce.....	61
Obr. 23: Návrh datového skladu	65
Obr. 24: Založení nového projektu pro tvorbu ETL procesu	66
Obr. 25: Control Flow pro Skutecne_prodeje_CE1PRE1.dtsx	67
Obr. 26: Nastavení Flat File Connection Manageru pro CE1PRE1	68
Obr. 27: Nastavení OLE DB Connection Manageru pro Zpl2 vyrobek.....	69
Obr. 28: Data Flow pro Skutecne_prodeje_CE1PRE1.dtsx.....	71
Obr. 29: Package ZPL2.dtsx pro proces aktualizace.....	72
Obr. 30: Založení nového projektu pro tvorbu OLAP kostky.....	72
Obr. 31: Výběr tabulek pro OLAP kostku z relační databáze.....	73
Obr. 32: Implementace dimenze Výrobek	74
Obr. 33: Implementace OLAP kostky ZPL2.....	75

Obr. 34: Postup nahrání OLAP kostky do Microsoft Excel.....	76
Obr. 35: Nastavení kontingenční tabulky pro projekt ZPL2.....	76
Obr. 36: Zobrazení kontingenční tabulky s nastaveným filtrem pro výrobek 1004.....	77

SEZNAM TABULEK

Tab. 1: Tabulka PGMI	55
Tab. 2: Tabulka MAKT.....	56
Tab. 3: Tabulka MARA	56
Tab. 4: Tabulka MARM.....	56
Tab. 5: Tabulka S031	57
Tab. 6: Tabulka MBEW	57
Tab. 7: Tabulka S578	58
Tab. 8: Tabulka CE1PRE1	58
Tab. 9: Tabulky MSEG a MKPF	59
Tab. 10: Tabulka S012	59
Tab. 11: Tabulka S026	60
Tab. 12: Metriky obsažené v tabulce faktů	62
Tab. 13: Dimenze času	63
Tab. 14: Atributy časové dimenze.....	63
Tab. 15: Dimenze sestava.....	64
Tab. 16: Atributy dimenze sestava	64
Tab. 17: Dimenze výrobky.....	64
Tab. 18: Atributy dimenze výrobky	65

Podklad pro zadání DIPLOMOVÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
Urban Martin	E.Beneše 1553, Hradec Králové - Moravské Předměstí	11301666

TÉMA ČESKY:

Implementace Business Intelligence řešení v podnikovém prostředí.

TÉMA ANGLICKY:

Implementation of Business Intelligence solution in the business environment

VEDOUcí PRÁCE:

Ing. Karel Mls, Ph.D. - KIT

ZÁSADY PRO VYPRACOVÁNÍ:

Cíl: Analyzovat vlastnosti cloudových a non-cloudových přístupů k BI. Navrhnout a implementovat BI řešení pro konkrétní firmu s využitím technologií a nástrojů MS.

Osnova:

Úvod

Business Intelligence

Nejnovejší trendy BI - výhody a problémy

ETL

Datové sklady

Technologické platformy BI

Implementace v SQL Server Data Tools

SEZNAM DOPORUČENÉ LITERATURY:

OBEIDAT, Muhammad, et al. Business Intelligence Technology, Applications, and Trends. International Management Review, 2015, 11.2: 47.

HOWSON, Cindi; HAMMOND, Mark. Successful Business Intelligence: Unlock the Value of BI & Big Data. McGraw-Hill Education, 2014.

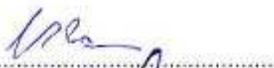
CHANG, Victor. The business intelligence as a service in the cloud. Future Generation Computer Systems, 2014, 37: 512-534.

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 2012, 36.4: 1165-1188.

CHAUDHURI, Surajit; DAYAL, Umeshwar; NARASAYYA, Vivek. An overview of business intelligence technology. Communications of the ACM, 2011, 54.8: 88-98.

LEBLANC, Patrick, et al. Which Analysis and Reporting Tools Do You Need?. Applied Microsoft? Business Intelligence, 2015, 1-20.

Podpis studenta:



Datum: 8.12.2015

Podpis vedoucího práce:



Datum: 15.12.2015