VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ
FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY
INSTITUTE OF INFORMATICS

DOLOVÁNÍ DAT JAKO NÁSTROJ PRO NALEZENÍ OBCHODNÍCH PŘÍLEŽITOSTÍ
DATA MINING AS A TOOL FOR PREPARING BUSINESS OPPORTUNITIES

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE Adam Křiva
AUTHOR

VEDOUCÍ PRÁCE Ing. Jiří Kříž, Ph.D.
SUPERVISOR

BRNO 2021

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA PODNIKATELSKÁ

Zadání bakalářské práce

Ústav:	Ústav informatiky
Student:	**Adam Křiva**
Studijní program:	Systémové inženýrství a informatika
Studijní obor:	Manažerská informatika
Vedoucí práce:	**Ing. Jiří Kříž, Ph.D.**
Akademický rok:	2020/21

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává bakalářskou práci s názvem:

Dolování dat jako nástroj pro nalezení obchodních příležitostí

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Cílem práce bude analýza nákupního koše od roku 2015 do roku 2020. Tedy určení závislostí mezi produkty, které zákazníci nakoupili a kategoriemi, do kterých produkty spadají, s využitím vybraných metod dolování dat.

Základní literární prameny:

HAN, Jiawei, Micheline KAMBER a Jian PEI. Data mining: concepts and techniques. 3rd ed. Waltham: Morgan Kaufmann, 2012. ISBN 9780123814791.

LABERGE, Robert a Jakub GONER. Datové sklady: agilní metody a business intelligence. Brno: Computer Press, 2012. 350 s. ISBN 978-80-251-3729-1.

LACKO, Ľuboslav. Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby. Brno: Computer Press, 2009. 456 s. ISBN 978-80-251-2887-9.

NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. Business intelligence: jak využít bohatství ve vašich datech. Praha: Grada, 2005. 254 s. ISBN 80-247-1094-3.

PYLE, Dorian. Data preparation for data mining. San Francisco: Morgan Kaufmann, 1999. 540 s. ISBN 1-55860-529-0.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2020/21

V Brně dne 28.2.2021

L. S.

Mgr. Veronika Novotná, Ph.D. ředitel	doc. Ing. Vojtěch Bartoš, Ph.D. děkan

Abstrakt

Předmětem bakalářské práce je získání vědomostí z dat společnosti PREMO s.r.o. pomocí metod dolování dat, které by měly napomoci k nalezení obchodních příležitostí. Proces přípravy dat pro tyto analýzy je založen na teoretických poznatcích, které jsou shrnuty v bakalářské práci. Získané vědomosti by měly podpořit vrcholový management ve správném rozhodování a přispět tak k vyšší konkurenceschopnosti na trhu.

Klíčová slova

Dolování dat, Datový sklad, Datové trhy, Příprava dat, Business Intelligence

Abstract

The subject of this barchelor thesis is to gain knowledge from the data of the PREMO Ltd company by use of data mining methods. That should help to find business opportunities. Data preparation process for these analyzes is based on theoretical information, which is summarized in this bachelor´s thesis. The gained knowledge should support top management in the right decisions and increase competitiveness in the market.

Key words

Data mining, Data warehouse, Data marts, Data preparation, Business Intelligence

Bibliografická citace práce:

KŘIVA, Adam. *Dolování dat jako nástroj pro nalezení obchodních příležitostí.* Brno, 2021. Dostupné také z: https://www.vutbr.cz/studenti/zav-prace/detail/135465. Bakalářská práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce Ing. Jiří Kříž Ph.D.

Čestné prohlášení

Prohlašuji, že předložená bakalářská práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 16. května 2021 .....

Adam Křiva

Poděkování

Tímto bych chtěl poděkovat vedoucímu mé bakalářské práce panu Ing. Jiřímu Kříži Ph.D. za cenné rady a připomínky, které mi poskytl při vypracování mé bakalářské práce. Dále bych chtěl poděkovat za pomoc panu Ing. Vladimíru Bartíkovi Ph.D. Zároveň také děkuji společnosti PREMO s.r.o. za poskytnuté informace.

Obsah

Úvod 12

1 Cíle práce, metody a postupy zpracování 13

1.1 Vymezení problému 13

1.2 Cíle práce 13

1.3 Postup a metody řešení 13

2 Teoretická východiska práce 14

2.1 Business intelligence 14

2.2 Zdroje dat 15

2.3 Atribut 16

2.3.1 Kvalitativní 16

2.3.2 Kvantitativní 17

2.3.3 Diskrétní a Spojitá hodnota 17

2.4 Příprava dat 18

2.4.1 Čištění 19

2.4.2 Integrace 21

2.4.3 Redukce 21

2.4.4 Transformace 21

2.5 Datový sklad 24

2.5.1 Typy datových skladů 26

2.5.2 Budování datového skladu 26

2.5.3 Tříúrovňová architektura 28

2.6 Multidimenzionální datový model 29

2.6.1 Datová kostka 30

2.6.2 Schéma uspořádání 31

2.6.3 Uložiště OLAP 33
2.7 Dolování dat 33
2.7.1 Decision Trees 34
2.7.2 Clustering 34
2.7.3 Association Rules 35
2.7.4 Time Series 36
2.7.5 Neural Network – neuronové sítě 36
3 Analýza současného stavu 38
3.1 Představení společnosti 38
3.1.1 Základní údaje společnosti 39
3.1.2 Milníky společnosti 40
3.1.3 Struktura společnosti 40
3.2 Produktové portfolio 41
3.3 Cílová skupina 42
3.4 Hardware 42
3.4.1 Položky 43
3.5 Software 43
3.5.1 Windows 10 Pro 43
3.5.2 Windows Server 44
3.5.3 Microsoft SQL Server 44
3.5.4 Microsoft Office 2016 44
3.5.5 Informační systém I6 44
3.5.6 Money S3 45
3.5.7 Kaspersky Endpoint Security 45
3.6 Placené služby 45

3.7 Cybersoft I6 ... 46
3.7.1 Založení odběratele / dodavatele ... 46
3.7.2 Přidání produktu ... 48
3.7.3 Vytvoření objednávky ... 49
3.8 Databáze a tabulky ... 50
3.8.1 Invoice ... 51
3.8.2 Company ... 51
3.8.3 Department ... 52
3.8.4 InvItem ... 52
3.8.5 StoItem ... 52
3.8.6 SCategory ... 53
3.9 Sestavy a Reporty ... 53
3.10 Zdroje dat ... 57
3.11 Zhodnocení současného stavu ... 58
4 Vlastní návrhy řešení ... 59
4.1 Nutná rozšíření softwaru ... 59
4.2 Datový sklad ... 61
4.3 Příprava dat ... 61
4.3.1 Import zdrojů ... 62
4.3.2 Tabulky dimenzí a faktu ... 63
4.4 Propojení zdroje s Visual Studiem ... 69
4.5 Datová kostka ... 70
4.6 Dolování dat ... 74
5 Závěr ... 81
Seznam použité literatury ... 82

Seznam Obrázků 84

Úvod

V současné době se trend práce s daty – Business Intelligence a dolování dat dostává do podvědomí i menších společností, které o toto téma začínají projevovat zájem. Získávání vědomostí z dat může být podstatnou částí při rozhodování a vytváření budoucích strategií.

Důležitou částí pro hodnotné výstupy analýz jsou kvalitní data, která se v databázích bez předchozích úprav nachází jen zřídka. Proto je výběr, příprava a čištění dat nejzásadnější a zároveň nejpodceňovanější částí dolování dat.

Téma bakalářské práce „Dolování dat jako nástroj pro nalezení obchodních příležitostí“ jsem si vybral z důvodu vzniklého zájmu o toto téma v průběhu studia.

V práci je zahrnuta teoretická část, analýza společnosti a návrh vlastního řešení.

1 Cíle práce, metody a postupy zpracování

1.1 Vymezení problému

Za nevyhovující považuji aktuální stav využívání dat ve společnosti PREMO s.r.o. Společnost nyní nevyužívá naplno svá interní data, která má nashromážděné a uložené v databázi za celou dobu svého působení na tuzemském trhu. Ve společnosti chybí lidské zdroje, které by rozuměly problematice práce s velkým množstvím dat. Jednatelé společnosti mají v podvědomí pojem big data, ale nevědí, co si pod tímto pojmem představit. Proto je i velký zájem o využití dat společnosti ve větším měřítku než doposud, ale chybí časový prostor, který by mohli věnovat tomuto tématu a zároveň je pro ně samotné téma technicky velmi náročné.

Uvedený nevyhovující stav dle mého názoru omezuje jednatele společnosti menším počtem možností při strategickém rozhodování a snižuje šanci na nalezení prvotřídní varianty, čímž může dojít ke snížení konkurenceschopnosti na trhu. Dále by řešení tohoto problému mohlo podpořit prodeje obchodních zástupců a pomohlo produktovému manažerovi při plánování nákupu většího množství produktů, které společnost prodává.

1.2 Cíle práce

Cílem práce bude analýza nákupního koše od roku 2015 do roku 2020. Tedy určení závislostí mezi produkty, které zákazníci nakoupili a kategoriemi, do kterých produkty spadají, s využitím vybraných metod dolování dat.

1.3 Postup a metody řešení

Proces získávání znalostí začíná výběrem klíčových dat pro zadaný cíl. Pokračuje jejich přípravou a čištěním, tak aby bylo dosaženo, co nejvyšší kvality. Poté přichází na řadu nahrávání dat do datového trhu a spuštění metod dolování dat.

2 Teoretická východiska práce

2.1 Business intelligence

Business intelligence nebo zkráceně BI je soubor činností, které poskytují zásadní informace pro vedení společnosti. Tyto informace mohou vést ke kvalitnějšímu a snadnějšímu rozhodování o směru a budoucnosti podniku. Hlavní zdroje, které tato technologie využívá, jsou interní informace společnosti, které mohou plynout z historicky získaných dat, ale i z aktuálních procesů. Výstup je využit k podpoře interních a externích analýz, které podniku pomáhají ve strategickém plánování. [1]

Hodnota

Přidaná hodnota je v možnosti individuálního dotazovaní, která zásadně zkracuje dobu pro získání výstupu. Eliminuje možnost nepochopení požadavku mezi žadatelem a oddělením, které před vznikem této technologie vyřizovalo požadavky na sestavy. Tím vedení získává schopnost se často dotazovat. Zvýšení přehledu o podniku umožňuje rychlou reakci na situaci na trhu a konkurenceschopnost. Business intelligence rozšiřuje podniku možnosti, jak přímo, rychle a uživatelsky přívětivě využívat vlastní strategický zdroj, který je kvalitní, aktuálně dostupný a užitečný. [1]

Prvky

Pevným bodem pro rozhodování je kvalitně zpracovaný systém umožňující získávat spolehlivé informace. Integrita je stěžejním pojmem, který shrnuje význam vysoce kvalitních dat, která jsou základem podkladů, která přispívají ke správným rozhodnutím. [1]

Faktory

Úspěšný projekt business intelligence je zaměřen na potřeby podniku a přináší mu přidanou hodnotu. Aby hodnota byla co nejvyšší, je nutné jasně definovat cíl, na který se má projekt zaměřit. [1]

Funkce

Podniky pravidelně vytváří široké spektrum sestav, které jsou vedením zdlouhavě analyzovány a vyhodnocovány. Technologii můžeme použít k mnoha účelům, různými

způsoby a provést ji velkým množstvím metod. Musíme však vytvořit systém, který řeší konkrétní požadavky a splňuje definované cíle. Tím dokážeme zvolit vhodné prostředí, které nejlépe zobrazí různé hodnotné sestavy z různých úhlů pohledu. Výběrem podstatných analýz a patřičných výstupů business intelligence můžeme zefektivnit celý systém. [1]

Prezentace

Prezentace výstupů business intelligence je důležitá část, která může být předložena mnoha způsoby. Mezi nejstarší patří vytváření sestav. Ty slouží jako statické přehledy informací, které jsou vytvářeny pravidelně a informují vedení podniku o jeho stavu. Dotazy naopak využívají strukturovaného dotazování pomocí jazyka SQL. Tímto způsobem může uživatel zjistit podrobnosti a souvislosti, které nejsou zřetelné z obecných sestav. Online analytical processing, zkráceně OLAP je technologie, která umožňuje uživatelům online a interaktivně pracovat se sestavami a vidět tak data z globálního nebo detailního pohledu. Ovládací panely a přehledy výsledků představují vizuální přehled o situaci podniku. [1]

2.2 Zdroje dat

Databázové soubory

Databáze v SQL Serveru umožňuje ukládat data do tří formátů. Přičemž musí obsahovat hlavní dva typy. Maximálně jeden primární datový soubor s příponou **MDF** a jeden či více souborů s logy transakcí, které mají příponu **LDF**. Soubor **NDF** tzv. sekundární, databáze obsahovat nemusí a záleží jen na nastavení databáze.

- Master Database File – Primární datový soubor je v databázi vždy jen jeden. Nese hlavní informace.
- Secondary Database File – Sekundární datový soubor může a také nemusí být v databázi zastoupen. Záleží na konfiguraci při vytváření. Obsahuje data, která nejsou uložena přímo v primárním datovém souboru.
- Log Database File – Soubor transakčních logů. V databázi musí být minimálně jeden. Obsahuje transakční data. Informace o změnách dat, které lze využít při obnovování databází. [7][8]

Formát XLSX

Soubor s touto příponou je výchozím formátem programu Microsoft Excel. Významem XLSX je sešit aplikace Excel. Základem souboru je jazyk XML, který zlepšuje bezpečnost a snižuje velikost pomocí komprimace ZIP. Zároveň tento formát neumožňuje využívání makra v jazyce VBA a XLM. [6]

Formát CSV

K importování a exportování textových souborů se často využívá soubor CSV, který umožňuje uložení jednoho listu Excelu a zanechává pouze formát u čísel. Buňky sousedící na jednom řádku jsou odděleny pomocí středníků, což umožňuje importování dat do většího množství aplikací a databází. [6]

2.3 Atribut

Atribut je pole, které obsahuje informaci o datovém objektu. Toto pole může mít více názvů podle různé teorie. V dolování dat a v databázích se běžně setkáváme s pojmem atribut. V oblasti datových skladů je využíván zaměnitelný pojem dimenze. Atribut popisuje objekt jeho vlastnostmi a může nabývat různých hodnot. Je možné rozdělit tyto znaky na kvalitativní a kvantitativní. Mezi kvalitativní patří nominální, binární a ordinální. Kvantitativní jsou číselné atributy. [2]

2.3.1 Kvalitativní

Nominální

Hodnoty nominálního atributu jsou nejčastěji názvy nebo symboly. Hodnoty, které není možné logicky seřadit, zařazuje atribut do určité kategorie podle názvu. Příkladem použití můžou být barvy toneru do tiskárny. Lze ji vyjádřit dvěma způsoby. Černou barvu můžeme napsat znakem 0, B nebo popisem „černá“. Červená bude 1, R nebo „červená“. [2]

Binární

Tento atribut může nabývat pouze dvě hodnoty a to 0 a 1. Nejčastěji se využívá pro přítomnost, kdy 0 naznačuje nepřítomnost a 1 přítomnost daného popisu. Jako příklad si můžeme představit tiskárnu. Pokud tiskána bude mít binární atribut barevná a hodnota u

tiskárny bude rovna 1 znamená to, že tiskárna bude tisknout barevně. Pokud bude hodnota rovna 0 tiskne pouze černobíle.

- Symetrický – Značí, že obě hodnoty jsou stejně důležité.
- Asymetrický – Hodnota je pro nás zásadní. Podstatnější informace je definována hodnotou 1. [2]

Ordinální

Ordinální atribut je nejčastěji definován názvem nebo symbolem. Hodnota, kterou je možné logicky seřadit a zařazuje nám atribut do kategorie. Ukazatel není možné objektivně ohodnotit a vyjadřuje nám interní hodnocení objektu. Příkladem použití může být zařazení zákazníka podle důležitosti. Kdy nejdůležitější odběratel může být označen jako „klíčový“ další jako „důležitý“ až po „náhodný“. [2]

2.3.2 Kvantitativní

Číselný

Kvantitativní atribut se vyznačuje celočíselným datovým typem integer nebo reálnou hodnotou. Rozdíl v zobrazování číselných atributů záleží na jeho vztahu k nule. Může být reprezentován dvěma způsoby:

- Intervalovou stupnicí – Číslo může klesnout do záporných čísel. Příkladem této stupnice je teplota ve stupních Celsia, kdy hodnota může být plusová i mínusová, tedy nad nulou i pod nulou.
- Poměrovou stupnicí – Hodnota této stupnice nikdy neklesne pod nulu. Vzorem je váha. Váha jakéhokoli tělesa nemůže klesnout pod nulu. [2]

2.3.3 Diskrétní a Spojitá hodnota

Je spoustu možností, jak třídit atributy podle typu, které se navzájem nevylučují. Klasifikační algoritmy vyšly ze strojového učení, které rozlišuje na diskrétní a spojité.

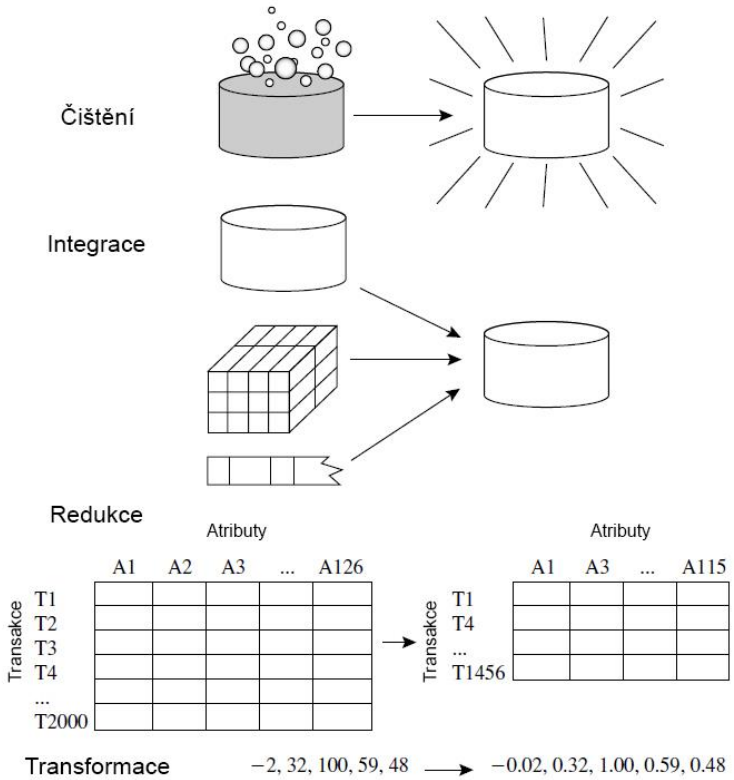
- Diskrétní atribut – Má konečný počet hodnot, které nemusí být znázorněny pouze číslem, ale také type či statusem. Jako je například barva nebo hodnost.
- Spojitý atribut – Příkladem je desetinné číslo, které může nabývat nekonečného počtu možností. [2][5]

2.4 Příprava dat

Díky velikosti a práci většího množství uživatelů obsahují databáze často poškozená, chybějící, nebo neúplná data. Špatná kvalita může vést k nekvalitním výsledkům analýz a dolování. Existují různé techniky, které zvyšují kvalitu a tím zvyšují i hodnotu výstupů. Faktory, které ovlivňují kvalitu dat:

- Přesnost – data neobsahují žádné chyby ani nečekané hodnoty
- Úplnost – nechybí žádné záznamy
- Konzistence – nejsou nesrovnalosti mezi hodnotami z různých zdrojů
- Časová osa – uživatelé zadávají data v čas
- Věrohodnost – uživatelé musí věřit používaným datům
- Interpretovatelnost – schopnost uživatelů vyznat se v datech

Kvalita dat může být snížena z různých důvodů. Data můžou být špatně použita. Může být nastaven špatný formát pro vstup. Před vložením mohou být data špatně nahrána uživatelem nebo chybou v systému. Uživatel zadá schválně na vstupu nepravdivou hodnotu, protože nechce udávat své osobní údaje. [2]



Obrázek 1 - Proces čištění [Vlastní zpracování podle zdroje 2]

2.4.1 Čištění

V procesu čištění dat je hlavní detekování nesrovnalostí. Ty mohou vzniknout různými způsoby. Lidskou chybou, špatným návrhem pro vstupní data, neochotou uživatelů sdílet své osobní informace atd. [2]

V prvním kroku je důležité se pozastavit nad informacemi, které již znám. Metadata nám mohou poskytnout důležité vstupní informace o našich datech. Jaké typy dat můžeme očekávat a jak jsou popsané atributy pomocí domén. [2]

Chybějící hodnoty

Je důležité si nejprve uvědomit, že chybějící hodnota neznamená vždy chybu v datech. To může být způsobeno softwarem, který v určitých případech dovoluje nechat hodnotu nevyplněnou. Je důležité předem definovat hodnotu, která by mohla nahradit tu prázdnou, jako například „nevím“ nebo „nic“. [2]

První možností, jak se vypořádat s chybějící hodnotou, je ignorovat celou n-tici, ve které není informace uvedena. Tato varianta není nejvhodnější, pokud řádek neobsahuje více polí s chybějící hodnotou, protože se okrádáme o zbylé atributy, které jsou uvedeny. [2]

Chybějící pole můžeme také vyplnit ručně. Tato metoda je ale časově vysoce náročná až neproveditelná podle množství dat, které nejsou uvedeny. [2]

Můžeme také vyplnit chybějící části dat, která nahradí prázdné místo hodnotou „neznámá“ nebo „prázdná“ a tím můžeme oklamat dolovací programy. Tato možnost je jednoduchá, ale není spolehlivá. [2]

Prázdné místo lze nahradit průměrem nebo mediánem hodnot. Základ pro výběr hodnot těchto operací můžeme změnit podle ostatních atributů, které jsou společné pro více objektů. [2]

Poslední a nejpoužívanější způsob pro doplnění hodnot je využití dolovacích nástrojů. Pomocí rozhodovacích stromů nebo Bayesovou klasifikací můžeme předvídat chybějící hodnotu. [2]

Poškozená nebo zkreslená data

Takzvaný datový šum je náhodná chyba nebo odchylka od ostatních hodnot. Na tento výjimečný rozptyl můžeme narazit pomocí technik vizualizace dat. Zobrazení nám ukáže proměnné, které vystupují mimo standardní hodnoty. [2]

K vyhlazení šumu může použít metodu sloučení dat tzv. **binning**:

- Pomocí průměrů

  Vytvoříme skupiny o 3 hodnotách a nahradíme všechny hodnoty průměrnou hodnotou, kterou jsme vypočítali z daných hodnot.
- Pomocí mediánu

  Všechny hodnoty ve skupině nahradíme mediánem.
- Pomocí ohraničení

  Skupina hodnot bude ohraničena minimem a maximem a každá hodnota bude zaměněna za tu bližší. [2]

Tříděné hodnoty: 4, 8, 15, 21, 21, 24, 25, 28, 34

Vytvořené skupiny hodnot:

1. skupina: 4, 8, 15
2. skupina: 21, 21, 24
3. skupina: 25, 28, 34

Sloučení pomocí průměru:

1. skupina: 9, 9, 9
2. skupina: 22, 22, 22
3. skupina: 29, 29, 39

Sloučení pomocí mediánu:

1. skupina: 8, 8, 8
2. skupina: 21, 21, 21
3. skupina: 28, 28, 38

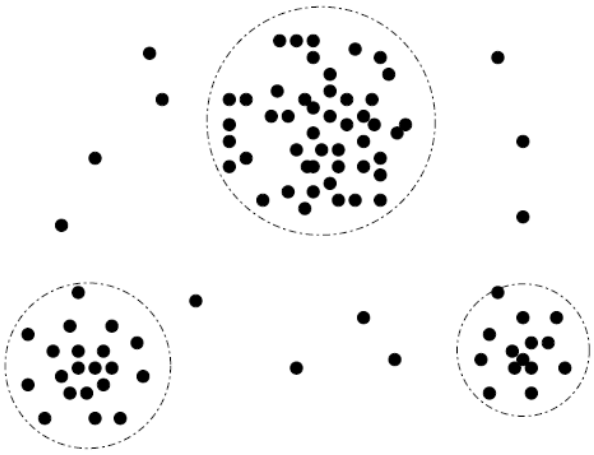
Sloučení pomocí ohraničení:

1. skupina: 4, 4, 15
2. skupina: 21, 21, 24
3. skupina: 25, 25, 34

Obrázek 2 - Metoda binning [Vlastní zpracování podle zdroje 2]

Další možností, jak provést vyhlazení dat, je využití **regrese**. Tato metoda přizpůsobuje hodnoty funkci. Lineární regrese vytváří nejvhodnější přímku mezi hodnotami tak, že atribut může předvídat polohu druhého atributu. [2]

Outlier analýza detekuje hodnoty pomocí shlukování a vytváří skupiny hodnot. Data, která nepatří do žádné vytvořené skupiny jsou nalezené odchylky. [2]



Obrázek 3 - Metoda outlier [2]

2.4.2 Integrace

Integrací můžeme rozšířit možnosti analýz o data, která nejsou ukládána do databáze nebo je nelze získat z interních systémů. Využitím externích zdrojů je možné dosáhnout obohacení o informace, které nejsou sbírány. [4]

Integrace má silný vliv na jednotnou úpravu významu dat. [10]

Datové sklady se většinou skládají z více zdrojů. Neopatrné spojení těchto zdrojů do jednoho sebou může přinést redundanci a nekonzistenci dat. Následkem může být značné zpomalení analýz a dolovacích procesů. [2]

2.4.3 Redukce

Výběrem celé databáze bychom získali obrovské množství dat, což by způsobilo extrémně dlouhé až nemožné provádění neefektivních analýz a dolování. Proto na řadu přichází redukce. Výběrem a odstraněním nepodstatných atributů snížíme objem dat. [2]

2.4.4 Transformace

Reprezentuje vhodnou úpravu dat pro jejich snadnější pochopení a zvyšuje efektivnost dolování dat. Představíme si techniky, které tuto patřičnou transformaci ztělesňují. [2]

Standardizace

Sjednocuje různé datové formáty ze všech zdrojů dle předdefinovaných pravidel. [4]

Agregace

K vyššímu výkonu přispívá agregace. Jedná se o předem provedené a uložené operace, které sumarizují data na nejdetailnější (nejnižší) úrovni hierarchie. Tím, že sníží počet záznamů nutných ke zpracování zásadně zrychluje dotazování na vyšších úrovních. [10]

Při přenosu dochází k provedení agregačních funkcí, mezi které patří součet, průměr, nejnižší a nejvyšší hodnota. [8]

Normalizace

Výběr jednotek, ve kterých jsou hodnoty udávány, může mít velký vliv na výsledek analýzy. Normalizace má snahu hodnotám přiřazovat stejnou váhu, což je zásadní pro určité algoritmy dolování dat. Data jsou změněna do intervalů mezí [-1;1] nebo [0.0;1.0]. Pro získání normalizované hodnoty existují metody. [2]

- Min-max

  V případě, že nová hodnota bude mimo původní nejvyšší a nejnižší hodnotu, narazíme na chybu mimo meze. To znamená, že nová hodnota se nevleze do původního rozsahu. [2]

  Nejvyšší hodnotu z atributu cena je 96000 a nejnižší je 1800. Chceme, aby nové hodnoty byly v rozsahu [0.0;1.0]. Cena, kterou budeme převádět je 47000.

$$X_i = \frac{47000 - 1800}{96000 - 1800} * (1.0 - 0) + 0$$

- Z-score

  Tato metoda je vhodná, pokud neznáme hranice rozsahu nebo máme část hodnot mimo rozsah, což v případě normalizace metodou min-max způsobuje chybu. [2]

  Průměrná hodnota ceny je rovna 36000 a směrodatná odchylka je 12000. Hodnota, kterou chceme přepočítat je 28000.

$$X_i = \frac{28000 - 36000}{12000}$$

- Desetinná stupnice

  Jedná se v podstatě o posouvání desetinného místa. Vybereme takovou mocninu čísla deset, která je větší než minimum a maximum.[2]

  Pokud máme hodnoty 96 a –91 stačí tyto hodnoty vydělit 100.

Je patrné, že normalizace mění původní hodnoty, a proto je nutné uložit normalizační parametry. V případě metody z-score je to průměr a směrodatná odchylka. Pokud použijeme desetinnou stupnici musíme uložit hodnotu, kterou dělíme.[2]

Diskretizace

Diskretizace rozděluje číselné hodnoty do intervalů, které mohou mít číselný rozsah nebo jsou označeny popisujícím štítkem pro daný interval. Jedná se o formu redukce velkého množství dat, která jsou nahrazena nižším počtem intervalů. [2]

Tento proces mění spojité hodnoty na diskrétní a může být prováděn různými technikami a dolovacími algoritmy. Příkladem je slučování dat (binning), shlukování (clustering) a rozhodovací stromy (decision tree). [2]

Příklad diskretizace hodnot věku

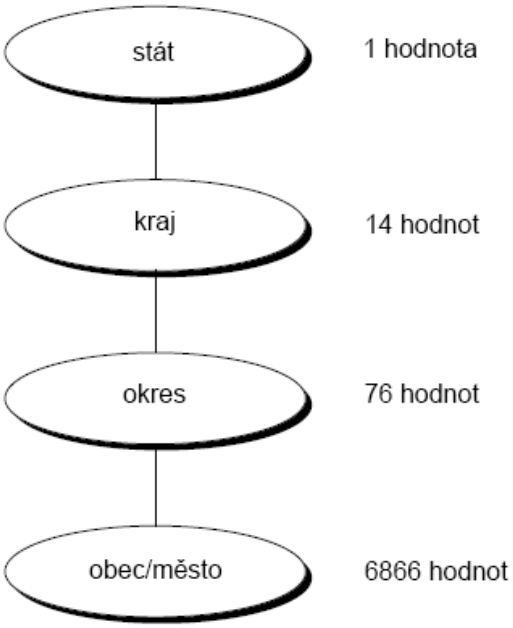
Hodnoty	Interval	Štítek
0,1,2,3,4,5,6,7,8,9,10	(0 - 10>	Dítě
10,11,12,13,14,15,16,17,18,19,20	(10 - 20>	Mladiství
21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,...	(20 - 30> (30 - 40> (40 - 50> (50 - 60> (60 - 70>	Dospělý
71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,...	(70 - 80> (80 - 90> (90 - 100> (100 - ∞)	Senior

Obrázek 4 – Diskretizace [Vlastní zdroj]

Generalizace

Určité druhy atributů je možné generalizovat. Což znamená, že lze definovat jejich vzájemnou hierarchii. Ta může nahrazovat nižší úroveň dat vyšší úrovní. Příkladem může být:

- Geografická poloha: město – okres – kraj – země
- Časová osa: den – měsíc – kvartál – rok
- Jiné: produkt – podkategorie – kategorie [2]



Obrázek 5 – Generalizace [Vlastní zpracování podle zdroje 2]

2.5 Datový sklad

Datový sklad poskytuje strukturu a nástroje pro vedení společnosti, které systematicky organizuje, chápe a využívá data podniku ke strategickým rozhodnutím. Datový sklad je důležitý nástroj, který zvyšuje konkurenceschopnost podniku v dnešní silné konkurenci a rychle se vyvíjejícím světě. Pomáhá nám získat zákazníka tím, že zjistíme víc o jeho potřebách. Datový sklad je volně řečeno datové uložiště, které je odděleno od operačních dat společnosti. Sjednocuje zdroje dat z různých systémů. Poskytuje kvalitní místa pro sloučené zdroje historických dat a podporuje zpracovávání informací. [2]

Podle vedoucího projektanta systémů datových skladů Williama H. Inmona je datový sklad místo, které uchovává souhrn dat, jejichž cílem je podpora vedení podniku v rozhodování. Musí ale splňovat čtyři podmínky. [2]

Subjektová orientace

Data jsou organizována v datovém skladu podle hlavních subjektů. Hlavním subjektem může být například produkt, zákazník, oddělení a jiné. Toto rozdělení podporuje rozhodování více, než data organizovaná dle konkrétních dnů a podnikových procesů. Odstraněním nepodstatných dat umožňuje snadnější pohled na konkrétní otázky. [2]

Integrovanost

Většinou vzniká datový sklad spojením různorodých zdrojů, u kterých je potřeba nejprve sjednotit význam. Proto je využívání integračních technik a čištění dat důležitou součástí, která zajišťuje sjednocení dat. [2]

Časová variabilita

Skladovaná data poskytují informace z dlouhodobého časového hlediska. Proto musí každý důležitý prvek datového skladu přímo nebo nepřímo obsahovat časový záznam, aby jej bylo možné zařadit. [2]

Neměnnost

Datové sklady jsou fyzicky odděleny od operačních transakčních databází a na rozdíl od nich neumožňují již nahraná data jakkoli upravovat, nebo mazat, což zásadně usnadňuje práci. Jsou možné pouze dvě možnosti, jak s nimi pracovat, a to nahrávání a zobrazování dat. [2][3]

Dá se říct, že datový sklad je významné uložiště dat, které slouží k podpoře rozhodování. K vybudování datového skladu je nutné data očistit a sjednotit. Datový sklad má od operační databáze odlišnou strukturu, obsah a jiné využití. Proto je důležité mít oddělený datový sklad od operační databáze, abychom dosáhli co nejvyššího výkonu. [2][3]

2.5.1 Typy datových skladů

Podnikový sklad

Tento typ skladu sbírá pro podnik data o všech důležitých subjektech z více operačních systémů a dalších externích zdrojů. Obsahuje detailní i sumarizační data. Může dosahovat velikosti od jednotek gigabyte po terabyte a výš. Vyžaduje důkladný podnikový model, který se navrhuje a tvoří roky. [2]

Datový trh

Obsahuje část dat podnikového skladu, které jsou podstatné pro konkrétní okruh uživatelů. Je zaměřený na důležitý subjekt podniku. Subjekt a obsažená data se mohou měnit podle oddělení podniku. Oddělení prodeje bude mít odlišný datový trh než oddělení marketingu. Tento typ skladu je provozován na nízkonákladových serverech založených na Unix/Linux nebo Windows a jeho vytvoření probíhá v řádu týdnů. [2]

Podle zdroje dat lze trhy rozdělit na závislé a nezávislé. Závislé trhy mají zdroj dat přímo v podnikovém datovém skladu oproti nezávislým trhům, které mohou mít různé zdroje z operačních systémů a externích zdrojů. [2]

Virtuální sklad

Z důvodu efektivního dotazování jsou vytvářeny pohledy nad operační databází. Těmto pohledům můžeme říkat virtuální sklad. Pohledy lze snadno vytvořit a nejsou nijak náročné na finance. Velkou nevýhodou je zatěžování operační databáze dotazy, které jsou náročné a zabírají tak důležitý výpočetní výkon pro firemní procesy. [2]

2.5.2 Budování datového skladu

Zásadní částí při vytváření datových skladů je zvolení vhodné metody budování. Každá metoda může obnášet jiné nástrahy, na které lze v průběhu implementace narazit. [3]

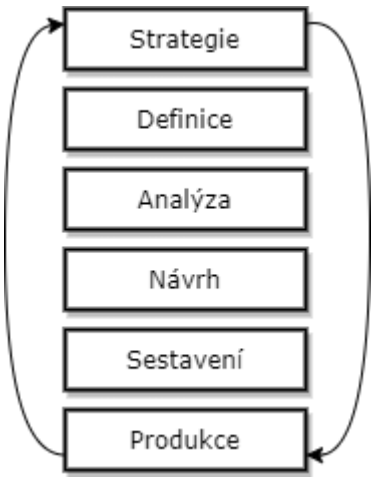
Metoda „Velkého třesku“

První variantou je implementace datového skladu pomocí jednorázového projektu. Tato metoda je časově náročná a hrozí zde riziko, že se v průběhu budování vyvine technologie, která bude efektivnější, nebo nastane změna požadavků ze strany

zadavatele, na kterou bude náročné reagovat z hlediska času i nákladů, které sebou přinese tato změna. Velký třesk je založen na třech hlavních částech, ve kterých se provádí analýza požadavků, vytváří se datový sklad a následně přístupy uživatelů. [3]

Přírůstková metoda

Zde je datový sklad budován postupně ve vytyčených etapách, které jsou stanoveny dle architektury datového skladu. Postupně se vytváří dílčí datové trhy, které jsou implementovány a poskytnuty uživatelům. Výhodou tohoto řešení je rychlé uspokojení zákazníka, který může začít využívat obsažené informace a poskytovat zpětnou vazbu na projekt a jeho částečné zpracování. Pokud je spokojený, nebo byly nedostatky odstraněny pokračuje se v implementaci dalších menších částí, dokud není datový sklad vytvořen kompletně. Jedná se tedy o opakování procesů, které vedou k postupné výstavbě. [3]



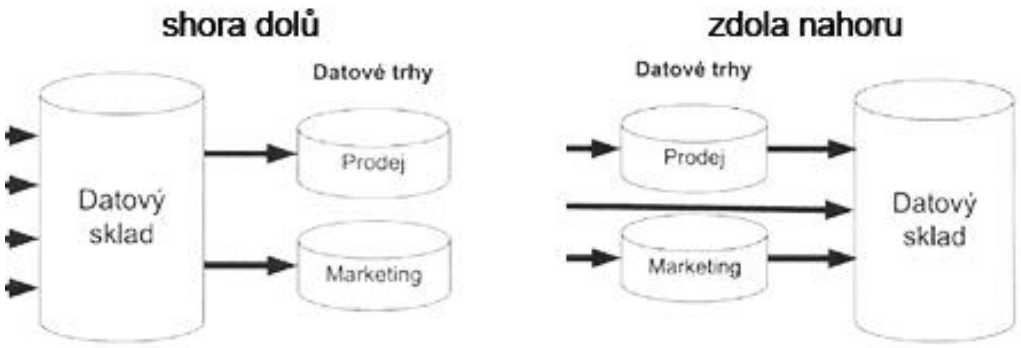
Obrázek 6 - Cyklus procesů přírůstkové metody [Vlastní zpracování podle zdroje 3]

Celý cyklus se může v přírůstkové metodě opakovat několikrát, než dojde ke kompletnímu dokončení. Vždy se začíná fází strategie, ve které se definují cíle podniku, krátkodobá i dlouhodobá strategie, účel datového skladu, dokumentace a školení uživatelů. Definice zastřešuje rozsah konkrétního přírůstku, dokumentaci datových zdrojů a kvalitu dat. Analýza má za cíl získat informace o uživatelích a vstupních datech obchodních analýz. V této fázi se rozhoduje o nástrojích a komponentech datového skladu. Fáze návrhu transformuje výstupy z analýz do podmínek návrhu a dokončí architekturu. V dalším kroku jsou sestaveny a testovány moduly správy skladu a přístupy k datům, sestavám a dotazům. Závěrem cyklu je spuštěný datový sklad ve

zkušební verzi a následně v ostrém provozu. Fáze produkce se také stará o růst a údržbu. [3]

Přírůstková metoda má dva směry budování. Shora dolů a zdola nahoru.

- Shora dolů – Podle požadavků uživatelů se navrhne model datového skladu a následně jsou vytvářeny jednotlivé datové trhy. Výhodou je rychlá implementace a návratnost investice, která nese podstatně nižší riziko než metoda velkého třesku. Nevýhodou jsou však vstupní náklady, které vzniknou dřív, než dokážeme předvídat návratnost investice.
- Zdola nahoru – Tato metoda je podobná metodě shora dolů. Prioritou jsou však v tomto případě data. Začíná se budováním datových trhů pro oblasti analýz. [3]



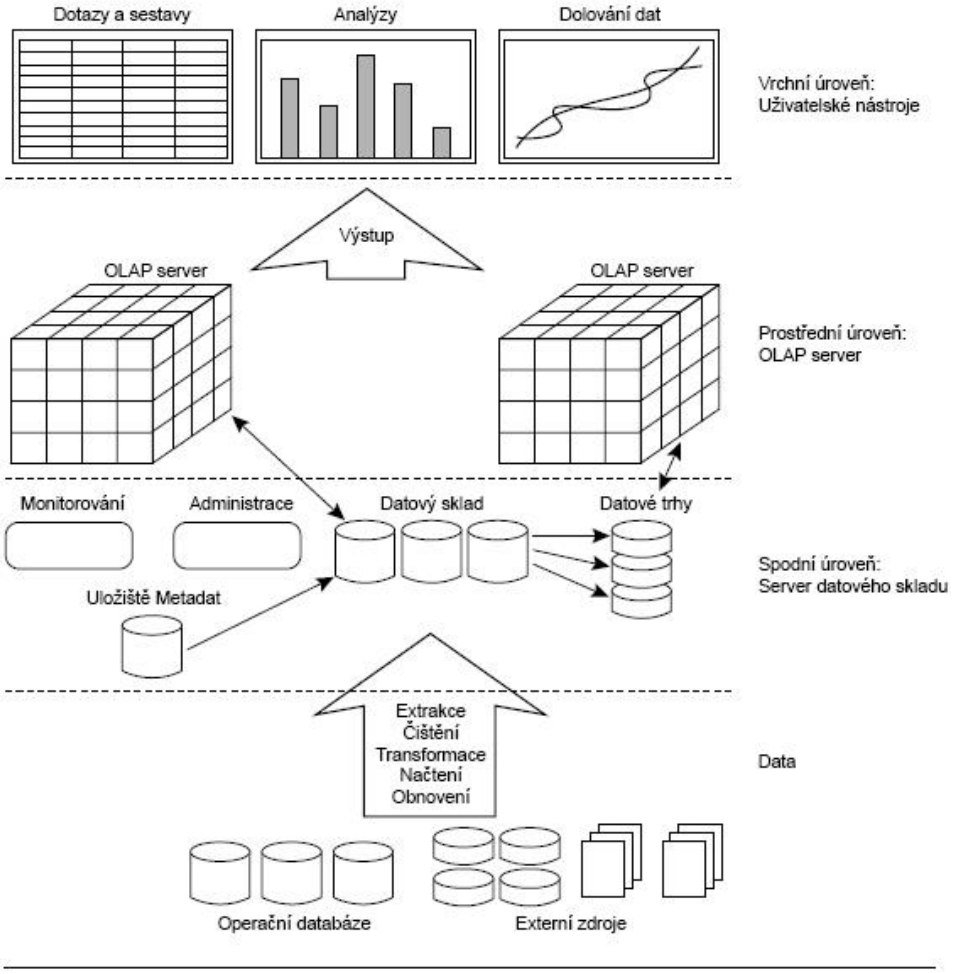
Obrázek 7 - Přírůstkové metody [Vlastní zpracování podle zdroje 3]

2.5.3 Tříúrovňová architektura

Nejnižší vrstvou tříúrovňové architektury je server datového skladu. Ten má nejčastěji podobu relační databáze, která je plněna daty z uživatelských a jiných programů. Tato data se vytváří na pozadí programů. Jsou ukládána do operačních databází a dalších externích zdrojů. Data musí projít procesem extrakce, čištění a úpravy, aby se z více zdrojů sjednotila do stejného formátu před tím, než jsou nahrána do datového skladu. V této vrstvě se také nachází uložiště metadat, která uchovávají informace o datovém skladu a jeho obsahu.

Druhou vrstvou je OLAP server, jenž zobrazuje uživatelům multidimenzionální data z datových skladů a trhů bez jakýchkoli informací o tom, kde a jak jsou data uložena. Ten však řeší problémy s ukládáním a implementací skladu pro OLAP analýzy nejčastěji metodami ROLAP nebo MOLAP. [2]

Nejvyšší vrstva zahrnuje výstupy pro koncového uživatele pomocí dotazování, reportovacích a analytických nástrojů nebo nástrojů pro data mining. [2]



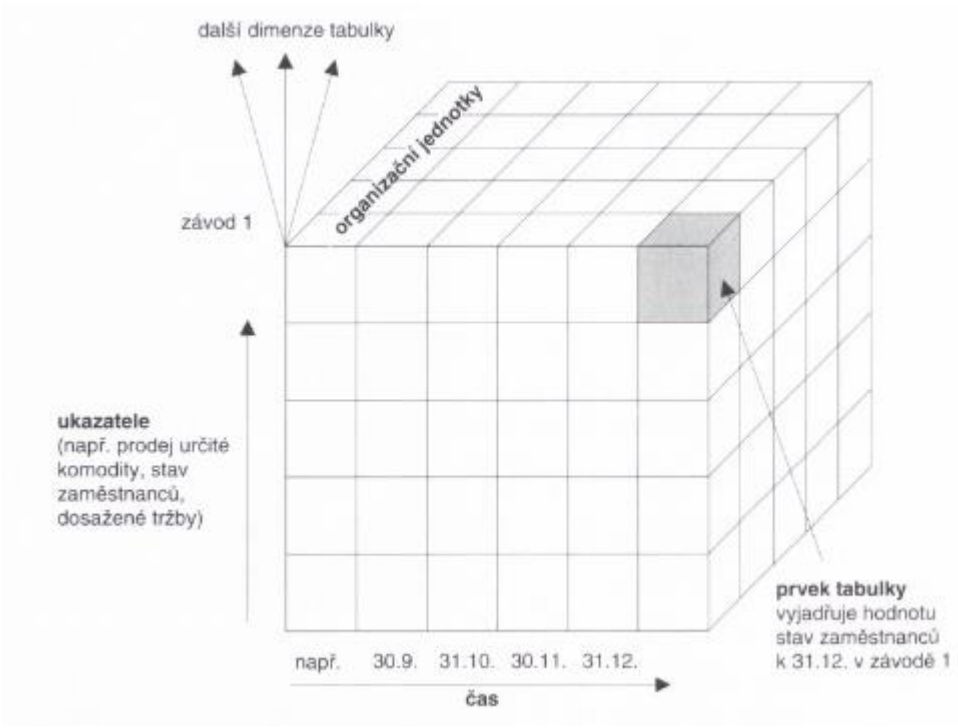
Obrázek 8 - Tříúrovňová architektura [Vlastní zpracování podle zdroje 2]

2.6 Multidimenzionální datový model

Multidimenzionální databáze nebo datový model je základem k provádění komplexních analýz. Umožňuje rychle a souhrnně využívat objemná data bez nutnosti psaní velkého množství dotazů. Pomocí nich je možné čerpat zásadní informace. Tyto databáze obsahují již vybraná a čistá data. [3]

2.6.1 Datová kostka

Datová kostka nahrazuje dvourozměrné tabulky a umožňuje zobrazovat data ve více dimenzích. Je tvořena dvěma typy tabulek. Dimenzemi a fakty. [2][3]



Obrázek 9 - Datová kostka [4]

Tabulka faktu

Centrem modelu je hlavní tabulka, která obsahuje zásadní informace ke zkoumanému tématu. Jedná se o nejrozsáhlejší tabulku v modelu, která shromažďuje fakta. Ty znázorňují důležitá data a číselné měřitelné jednotky, která jsou podstatné například pro prodeje podniku. Obsahují také vazby pro tabulky dimenzí. [2][3]

Tabulka dimenze

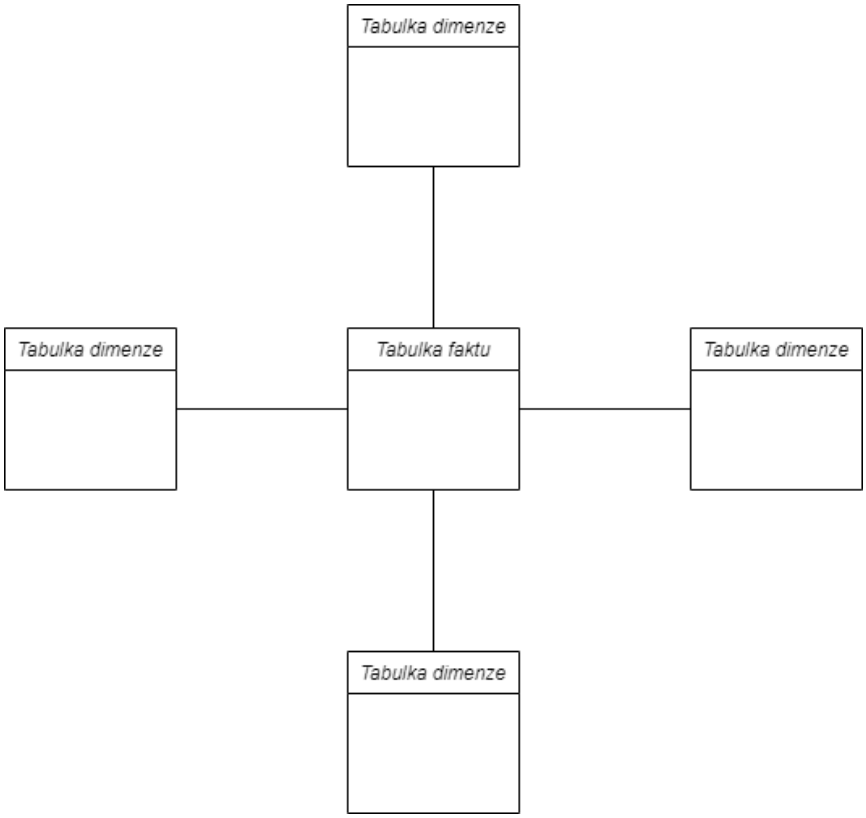
Dimenze pomáhají ucelit pohled na tabulku faktu a doplňují k ní informace. Rozdělují data a skládají je do pořadí podle úrovně vnoření nebo detailnosti. Nejčastější je časová dimenze, která rozděluje datum do roku, kvartálu, měsíce a dne. [2][3]

2.6.2 Schéma uspořádání

Topologie uspořádání jsou využívána v různých částech datových uložišť a pro danou část mají své výhody. Hvězda a vločka se často využívá v datových trzích, kdežto souhvězdí je využíváno pro datový sklad. [2]

Schéma hvězdy

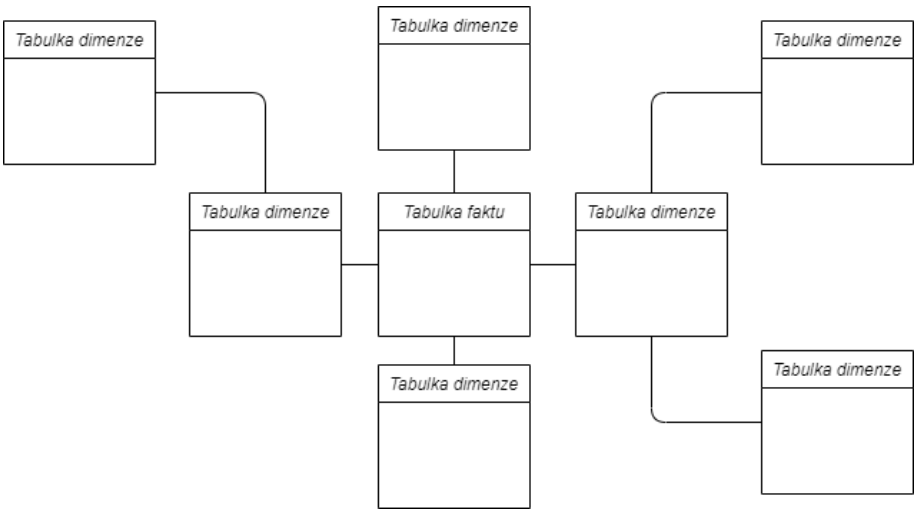
Výhodou tohoto schéma je na první pohled srozumitelnost. Tabulky dimenzí jsou soustředěny kolem centrální tabulky faktu. Takto navržený datový model může sloužit podniku k analýze konkrétního procesu. Tabulka faktu může obsahovat informace ohledně prodeje a jednotlivé dimenze můžeme nastavovat tak, že budou zobrazovat prodeje podle oddělení a kategorie produktů. Hvězdicové schéma je založeno na vztahu k centrální tabulce, ve které jsou obsaženy cizí klíče, které propojují tabulky dimenzí. Model je pomalejší z důvodu nenormalizovaných dimenzí. [1]



Obrázek 10 - Hvězdicové schéma [Vlastní zpracování podle zdroje 1]

Schéma sněhové vločky

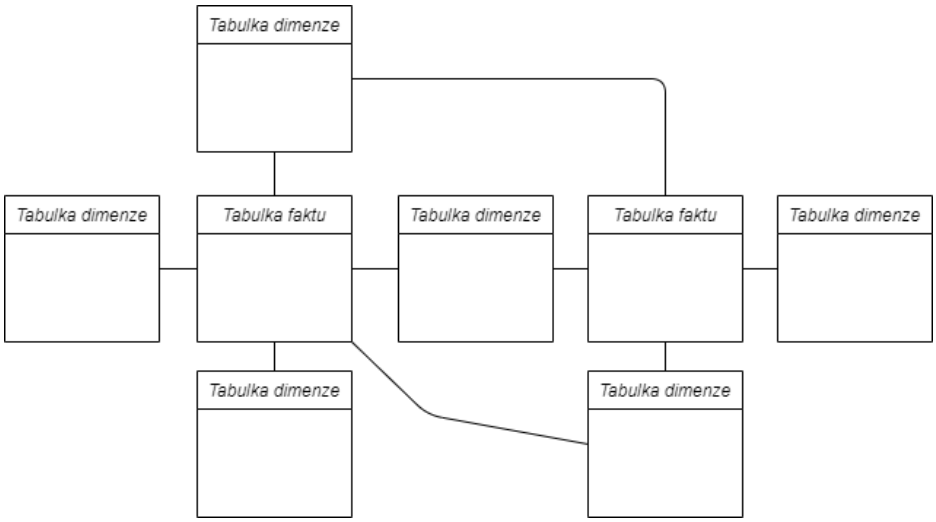
Sněhová vločka má stejný základ jako hvězdicové schéma, ale oproti němu jsou rozšířeny některé dimenze o další relační tabulky, se kterými jsou svázány. Tento model má více normalizované dimenze, což umožňuje rychlejší zavedení dat, ale na úkor dotazovacího výkonu, který se sníží.[1][3]



Obrázek 11 - Schéma vločky [Vlastní zpracování podle zdroje 1]

Schéma souhvězdí

Toto schéma zahrnuje více tabulek faktu. Tyto tabulky mají společné dimenze, které mezi sebou navzájem sdílejí. [2]



Obrázek 12 - Schéma souhvězdí [Vlastní zpracování podle zdroje 2]

2.6.3 Uložiště OLAP

Online analytical processing server poskytuje multidimenzionální data uživatelům z datových trhů a datového skladu bez toho, aniž by uváděl, jak a kde jsou data uložena. Data ale musí být někde fyzicky skladována. [2]

- MOLAP – Multidimenzionální OLAP ukládá data v multidimenzionálních datových kostkách. Hlavní výhoda je maximalizace výkonu. Nevýhodou je redundance dat, což zvyšuje požadavky na velikost uložiště.
- ROLAP – Relační OLAP ukládá data do relačních databází, díky níž se lze vyhnout problému s redundancí dat.
- HOLAP – Hybridní OLAP kombinuje obě výše zmíněné technologie. Přičemž se snaží využít výhod obou technologií. Agregované hodnoty jsou uloženy v datových kostkách a detailní data v relační databázi.
- DOLAP – Desktopový OLAP se připojuje k hlavnímu uložišti a stahuje do lokálního uložiště vybranou část dat, nad kterou bude prováděna analýza. Operace jsou později prováděny lokálně a není vyžadováno připojení k serveru. [3][4]

2.7 Dolování dat

Dolování dat nebo také data mining lze definovat mnoha způsoby, protože se s tímto tématem můžeme setkat napříč různými obory. Některé kratší názvy, jako dolování vědomostí, dostatečně nevystihují podstatu této disciplíny, a proto se musíme uchýlit k těm delším. Dolování vědomostí z dat neboli data mining je proces, který hledá drobné skupiny drahocenných informací ve velkém množství dat. Další populární definicí je nalezení vědomostí z dat. [2]

Dolování dat nebo také data mining objevuje strategické informace ve velkém objemu dat pomocí algoritmů. Je to proces, který slouží k objevování předem nedefinovaných, ale zásadních informací. [4]

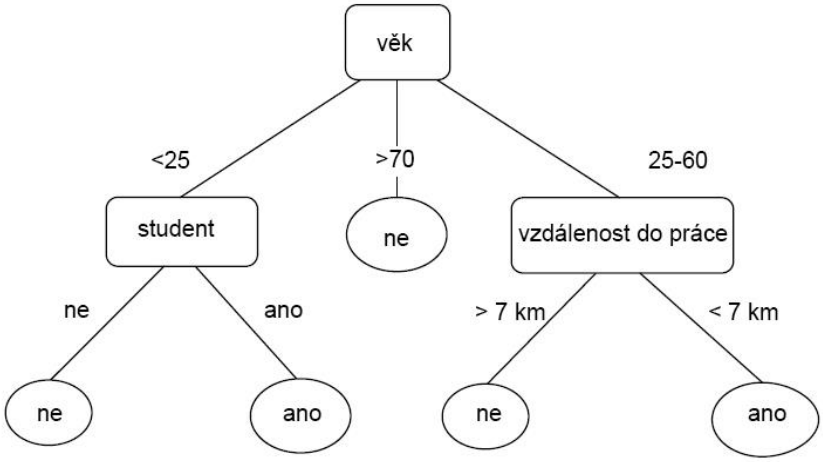
Denně shromažďujeme obrovské množství dat, ze kterých získáváme velmi malé, nebo žádné množství informací. Data mining nám poskytuje nástroje, které nám pomocí

složitých metod umožňují sofistikovaně analyzovat data a získávat tak důležité vědomosti. [2]

2.7.1 Decision Trees

Rozhodovací stromy vytváří modely, které předvídají rozhodování podle vlastností a závislostí atributů. Výhodou je jednoduché a srozumitelné zobrazení výstupů. Cílem je zjistit a předvídat chování skupiny na základě vlastností, které má. [3]

Příkladem může být rozhodovací strom, který podle věku, studentského statusu a vzdálenosti do práce predikuje, zda si zákazník koupí kolo. [3]

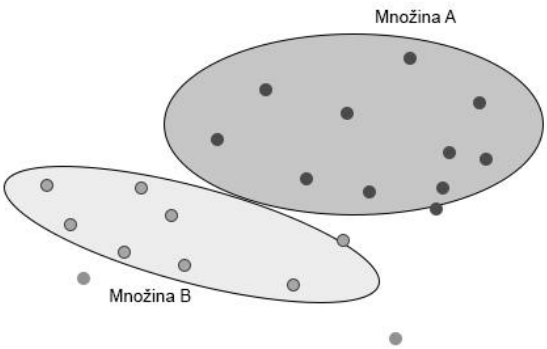


Obrázek 13 - Rozhodovací strom [Vlastní zpracování podle zdroje 2 a 3]

2.7.2 Clustering

Shlukování spojuje objekty, které mají podobné vlastnosti a vytváří z nich skupiny. Podmínky těchto podobností nejsou předem definovány, ale jsou zjištěny z dat. Tyto skupiny se mohou navzájem překrývat. To znamená, že určitá skupina objektů může mít společnou vlastnost s jinou skupinou. [3]

Vhodný cíl pro využití algoritmu je segmentování zákazníků. Tím můžeme rozdělit například zákazníky ze segmentu B2B do skupin podle obratu, počtu zaměstnanců, právní formy a hlavní činnosti podniku. [3]



Obrázek 14 - Shlukování [Vlastní zpracování podle zdroje 3]

2.7.3 Association Rules

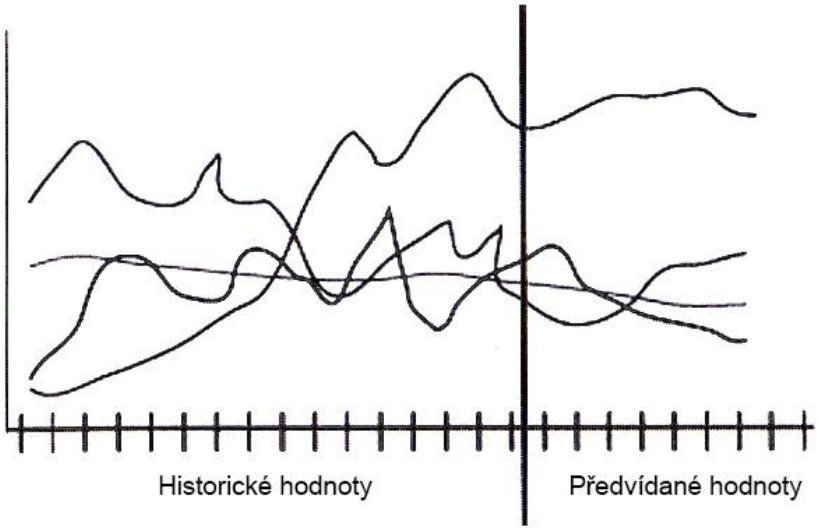
Nejčastější způsob využití tohoto algoritmu je analýza nákupního koše. Odpovídá na otázku, které produkty jsou nejčastěji nakupovány společně. Objevuje vzájemné vztahy mezi subjekty. Tento vztah může být pozitivní, ale i negativní. Pozitivní korelace říká, že pokud koupím produkt X s velkou pravděpodobností koupím i produkt Y. Negativní korelace uvádí, že pokud vyberu produkt X je malá pravděpodobnost, že vyberu i produkt Y. Výsledek algoritmu může být využit při sestavování produktového katalogu, rozložení zboží v regálu nebo při doporučování dalšího zboží v e-shopu. [2][3]

Obrázek 15 - Asociační pravidla [Vlastní zpracování podle zdroje 2]

2.7.4 Time Series

Časové řady předvídají chování v následujícím období podle historických dat. Ty se musí skládat z rozsáhlých číselných dat, která jsou zaznamenávána ve stejných časových intervalech. [2]

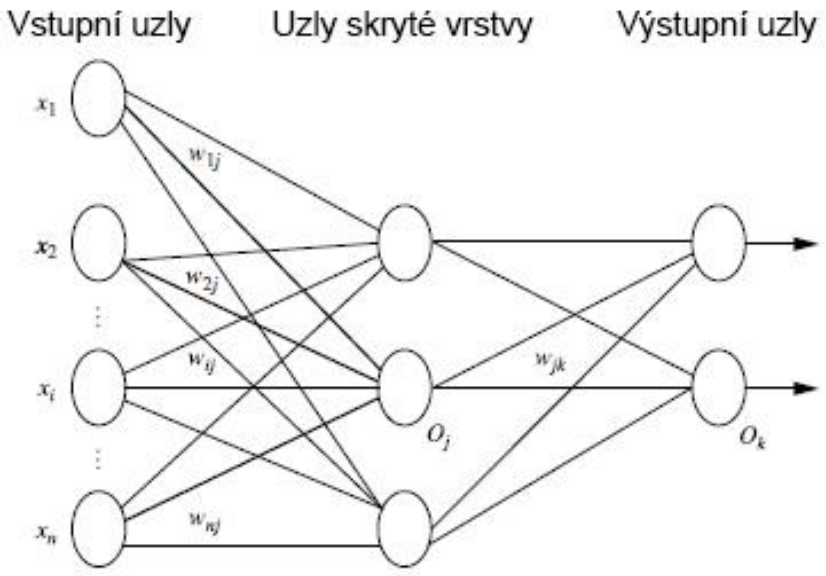
Data pro tento algoritmus nejčastěji představují zisk, náklady, obrat a podobné hodnoty na základě nichž lze předvídat trend proměnné. [3]



Obrázek 16 - Časové řady [Vlastní zpracování podle zdroje 3]

2.7.5 Neural Network – neuronové sítě

Zpracování dat funguje podobně jako v lidském mozku. Neuronové buňky získávají vstupní informace, které zpracovávají a vyhodnocují z nich závěry. Na tomto principu funguje i dolování pomocí neuronových sítí. Algoritmus přijímá informace a z každé přijaté informace získá zkušenosti, ze kterých se učí. Nejprve jsou data rozdělena na trénovací a testovací část. Následně jsou vstupy opakovaně zpracovávány do uzlů, které mají své vlastnosti a pravidla. Ty jsou porovnávány se skutečnými hodnotami a na základě vyhodnocení chybovosti se upraví váha vazby mezi těmito dvěma uzly. Čím vyšší váha vazby, tím vyšší je na daný uzel. [2][3]



Obrázek 17 - Neuronové sítě [Vlastní zpracování podle zdroje 2]

3 Analýza současného stavu

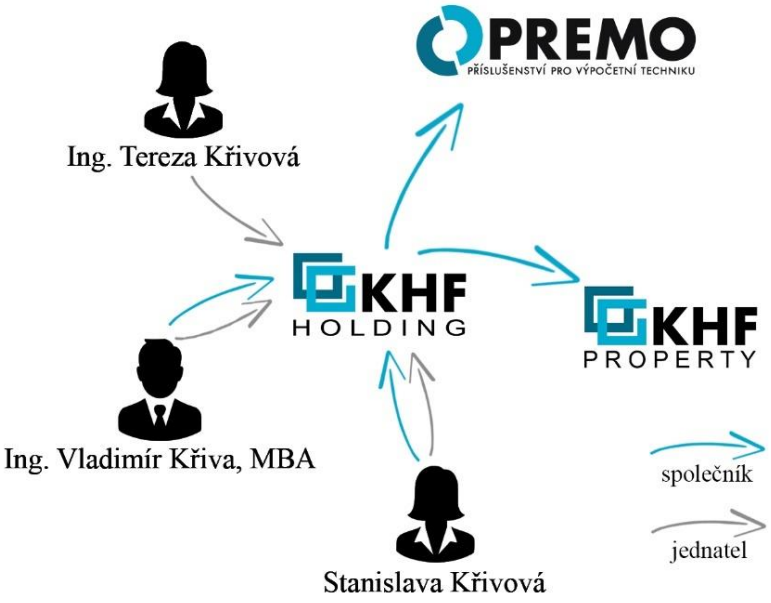
V této části práce představím společnost PREMO s.r.o. a uvedu nejdůležitější události, které se udály za celou dobu působení na tuzemském trhu od jejího vzniku. Popíši služby a produkty, které společnost poskytuje. Vymezím cílovou skupinu společnosti, která tvoří pravidelný zisk. Uvedu hardware, na kterém pracují data společnosti a software, který podporuje rychlou administrativu. Provedu prvotní analýzu databáze, popíši tabulky, jejich atributy a vazby mezi nimi. Představím práci s daty, analýzy a reporty, které jsou ve společnosti již prováděny a na závěr shrnu celou kapitolu.

3.1 Představení společnosti

Společnost PREMO s.r.o. byla založena v roce 2001, tehdy ještě pod prvotním názvem HORN INTERNATIONAL, s.r.o. dvěma společníky. Panem Petrem Hornem, který chtěl v České republice vybudovat pobočku své německé firmy HORN GmbH., jehož obchodní podíl činil 51 %. a panem Ing. Vladimírem Křivou, MBA., který se stal společníkem pana Petra Horna s obchodním podílem 49 %.

V roce 2008 prodal pan Petr Horn svůj podíl panu Ing. Vladimíru Křivovi, MBA, který se stal jediným vlastníkem s obchodním podílem 100 %. Podmínkou při prodeji podílu byla změna obchodního jména společnosti, a tak byla společnost HORN INTERNATIONAL přejmenována v září téhož roku na společnost PREMO s.r.o. podle latinského slova premo, které v překladu do českého jazyka znamená tisknout.

Společnost PREMO s.r.o. v roce 2017 odkoupila 100 % obchodního podílu společnosti VODO - REGULA, s.r.o., která byla následně přejmenována na KHF PROPERTY s.r.o. a oba obchodní podíly obou společností PREMO s.r.o. a KHF PROPERTY s.r.o. ve výši 100% byly následně prodány nově vzniklé společnost KHF HOLDING s.r.o., kterou vlastní dva společníci. Pan Ing. Vladimír Křiva, MBA s obchodním podílem 60 % a paní Stanislava Křivová s obchodním podílem 40 %.



Obrázek 18 - Struktura jednatelů [Vlastní zpracování]

3.1.1 Základní údaje společnosti

Název společnosti:	PREMO s.r.o.
Sídlo:	Brněnská 474, 686 03 Staré Město u Uherského Hradiště
IČO:	26251531
DIČ:	CZ26251531
Webová stránka:	www.premocz.eu
E-shop:	eshop.premocz.eu
Telefon:	+420 572 433 860
E-mail:	premo@premocz.eu

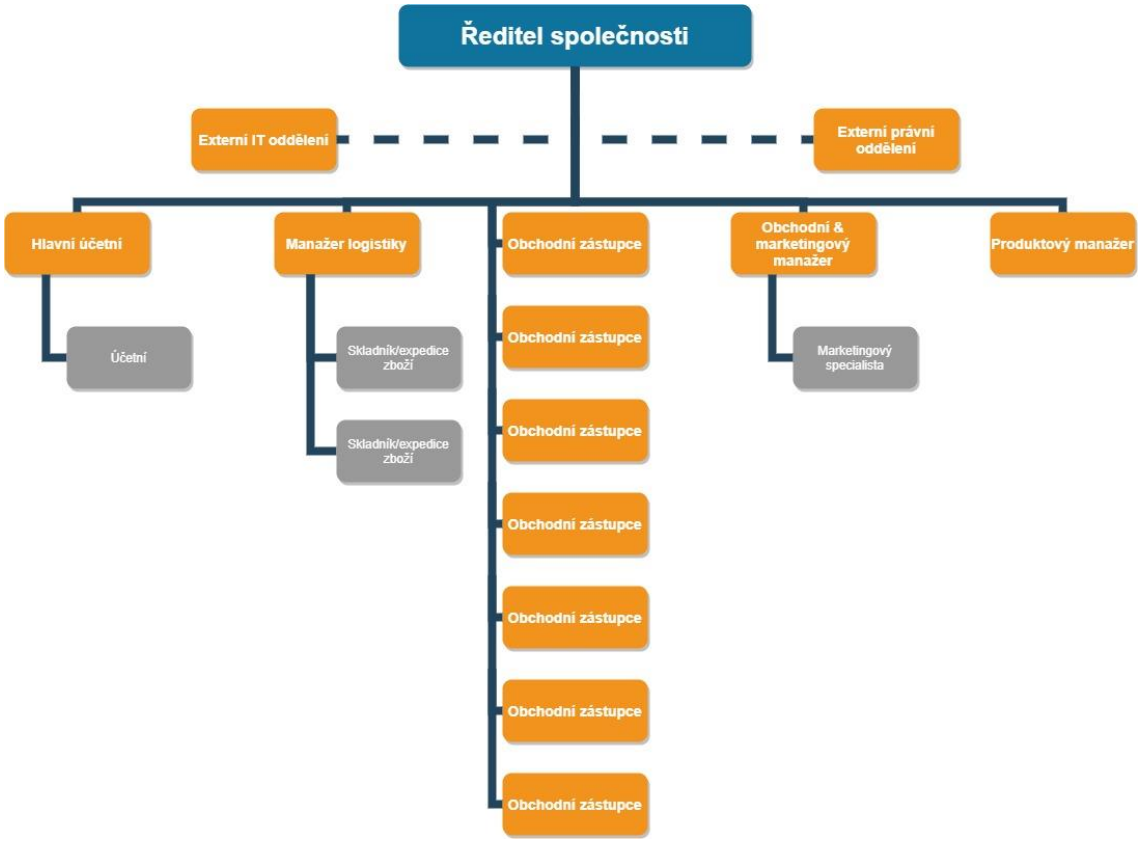
Obrázek 19 - Logo společnosti [PREMO s.r.o. © 2021]

3.1.2 Milníky společnosti

- Zavedení ČSN EN ISO 9001:2001 v roce 2004
- Získání ČSN EN ISO 14001:2005 v roce 2007
- Osvědčení EKO-KOM v roce 2011
- Získání partnerství se společností HP Inc. v roce 2013
- Zařazení do programu „HP Supplies Return and Recycling Programme“ v roce 2015

3.1.3 Struktura společnosti

Společnost má jednoduchou hierarchii řízení. Organizační struktura společnosti je liniová. Pomocí liniové struktury jsou jasné vztahy nadřízenosti a podřízenosti ve společnosti. Řízení formální i neformální povahy je ve společnosti zavedeno od začátku podnikatelské činnosti. Zaměstnanci jej akceptují a podřizují se tomuto způsobu řízení.



Obrázek 20 - Struktura společnosti [Vlastní zpracování]

3.2 Produktové portfolio

Společnost má v obchodním rejstříku uvedený předmět podnikání výrobu, obchod a služby neuvedené v přílohách 1 až 3 živnostenského zákona. Obchoduje s příslušenstvím pro výpočetní techniku. Prodává tiskárny a veškerý spotřební materiál k nim potřebný, jako jsou tonerové a inkoustové kazety, válce, odpadní nádobky, třídící podavač papíru a další produkty spojené s tiskárnami. Nabízí také spotřební materiál pro výpočetní techniku. Obchodní oblastí společnosti je celá Česká republika, kterou pokrývá díky oblastním obchodním zástupcům a zaměstnancům společnosti.

Výpočetní technika

Mobily, tablety, počítače, notebooky, servery, datová uložiště, chytré hodinky, GPS navigace

Komponenty pro výpočetní techniku

Procesory, grafické karty, základní desky, operační paměti, zdroje, chladiče, disky, flash disky, paměťové karty, skříně

Příslušenství pro výpočetní techniku

Tiskárny, multifunkční tiskárny, kopírky, skenery, monitory, klávesnice, myši, sluchátka, reproduktory, přenosné reproduktory, obaly

Video, foto

Televize, kamery, fotoaparáty, objektivy, domácí kina

Spotřební materiál pro tiskárny

- Originální spotřební materiál

Kazety vyráběné výrobcem tiskárny. Používáním originálních náplní u nových tiskáren nedochází ke ztrátě záruky a zaručuje kvalitu tisku.

- Renovovaný spotřební materiál

Renovované, nebo také repasované inkousty a tonery, jsou šetrné k životní prostředí díky opětovnému využití originální kazety. Vyráběná pouze certifikovanými výrobci, kteří splňují mezinárodní normu ISO 9001.

- Kompatibilní spotřební materiál

Kompatibilní inkoustová, nebo tonerová kazeta je vyrobena výrobcem třetí strany z těla originální kazety. V této originální kazetě jsou vyměněny opotřebované nebo poškozené části za nové a je do ní vložena nová náplň.

3.3 Cílová skupina

Společnost rozděluje své zákazníky podle tří následujících vztahů - B2C, B2B a B2G. Obchoduje ve všech zmíněných kategoriích, ale její primární cílová skupina, o kterou se nejvíce zajímá je B2B segment. O tuto skupinu chce společnost pečovat svými kvalitními produkty, profesionálními službami a individuálním vztahem mezi obchodním zástupcem a zákazníkem. B2G segment je také důležitou součástí obratu a zisku společnosti, ale v této kategorii se hraje podle přesně stanovených pravidel zadavatele veřejných zakázek a soutěží se primárně podle nejnižší ceny na daný spotřební koš. Zakázku vyhraje účastník s nejnižší nabídkou a není možné ji ovlivnit žádným jiným faktorem. Proto primární zájem společnosti směřuje k B2B segmentu.

3.4 Hardware

Společnost v roce 2017 investovala do nového hardwaru, který snížil čas odezvy na požadavky ze strany zaměstnanců a obchodních zástupců na informační systém. Z původního serverového racku se stal datový rozvaděč, ke kterému byl přidán firewall a nová racková skříň byla osazena serverem a datovým uložištěm. Páteřní spojení mezi těmito dvěma racky bylo propojeno optickým kabelem z důvodu zvýšení rychlosti přenosu dat. Byl také zakoupený mikrovlnný datový spoj pro rychlejší a spolehlivější přístup k internetu.

Zaměstnanci a obchodní zástupci, kterým je umožněno pracovat i mimo sídlo společnosti, jsou vybaveni notebooky z důvodu mobility a konektivity pro případ, kdy nemají možnost se dostavit na pracoviště nebo jim to neumožňuje jejich aktuální situace. Každý zaměstnanec a obchodní zástupce má v sídle společnosti své vlastní pracovní místo, které je vybaveno externím monitorem, klávesnicí a myší. Svým notebookem se pomocí dokovací stanice rychle a snadně připojí k veškerému příslušenství.

3.4.1 Položky

Rack 1

- Switch ZyXEL XGS2210-52
- Uložiště Synology DS416 DiskStation
- Server HP ProLiant DL380 G9
- Záložní zdroj APC Smart UPS 1500

Rack 2

- Firewall ZyXEL ATP500 ZyWALL ATP Firewall
- Patchpanel ZyXEL GS-1548
- Telefonní ústředna Panasonic KX-NS500

HW k dispozici pro zaměstnance

- HP EliteDisplay E243i
- HP EliteBook 850 G5
- Dokovací stanice HP Ultraslim pro EliteBook

3.5 Software

Nedílnou součástí pro efektivní fungování společnosti je software, který zaměstnancům usnadňuje práci a zkvalitňuje komunikaci se zákazníkem.

3.5.1 Windows 10 Pro

Společnost má na všech počítačích nainstalovaný operační systém Windows 10 ve verzi Pro z důvodu využívání funkcí, které nižší verze Windows 10 nepodporují.

Nejvyužívanější funkcí je použití vzdálené plochy, díky níž se zaměstnanci mohou odkudkoli připojit, jak k informačnímu systému společnosti, tak i k datovému uložišti.

Druhá důležitá funkce je služba Active Directory, která umožňuje připojení k doméně společnosti a tím také přístup ke sdíleným souborům a tiskárnám. Správci sítě umožňuje snadno a rychle spravovat oprávnění pro každého uživatele zvlášť a tím zvyšuje zabezpečení celé sítě.

3.5.2 Windows Server

Jedná se o serverový operační systém, který poskytuje veškeré technologie důležité pro síťové, aplikační a online funkce. Umožňuje veškerou správu pomocí textových i grafických nástrojů. [9]

3.5.3 Microsoft SQL Server

Microsoft vytvořil databázový systém, který představuje ucelené řízení podnikových dat. Umožňuje využívat široké spektrum funkcí od relačních databází, přes správu a manipulaci s daty až po náročné analytické úkony. [8]

3.5.4 Microsoft Office 2016

Mezi nejdůležitější části balíčku patří Outlook, Word a Excel. Nepostradatelné programy, které umožňují rychlou komunikaci se zákazníkem, zpracování nabídek do veřejných zakázek a přípravu smluv se zákazníky a dodavateli.

3.5.5 Informační systém I6

Společnost používá od roku 2005 nájem softwaru od společnosti CYBERSOFT, s.r.o., která od roku 2000 vyvíjí informační systém I6, jehož databáze běží na databázovém MS SQL serveru a využívá jeho transakční zpracování. Tento informační systém patří do kategorie ERP/CRM (Enterprise Resource Planning / Customer Relationship Management). [12]

Informační systém I6 umožňuje automatizované importování a aktualizaci záznamů od distributorů. Podporuje import nových produktů. Aktualizuje cenu a množství produktů, které má distributor skladem. Přehledně a systematicky uvádí všechny důležité informace pro obchodní zástupce, kterým usnadňuje zpracování cenové nabídky pro zákazníky.

Informační systém I6 pokrývá všechny činnosti společnosti mimo personalistiku a mzdy.

3.5.6 Money S3

Jednatel společnosti v softwaru Money S3 zpracovává pouze mzdy z důvodu, že informační systém I6 nemá modul pro zpracování mezd.

3.5.7 Kaspersky Endpoint Security

Společnost klade důraz na bezpečnost a ochranu všech svých interních dat, a proto nakoupila licenci pro každou koncovou stanici, aby snížila riziko neoprávněného přístupu k datům.

3.6 Placené služby

Cybersoft, s.r.o.

PREMO s.r.o. platí měsíční pronájem licencí ERP/CRM systému I6 od společnosti CyberSoft, s.r.o., čím získává pravidelný bezplatný update, který aktualizuje systém a zapracovává změny v legislativě. V pronájmu je zahrnuta podpora Hot line a HelpDesk v časovém rozsahu čtyř hodin za měsíc, za které společnost PREMO s.r.o. v případě řešení individuálních požadavků neplatí.

Credit Check, s.r.o.

Společnost Credit Check, s.r.o. prověřuje bonitu společností v České republice a tím pomáhá chránit podnikání před rizikem z neplacení pohledávek. Tato služba umožňuje prověřit obchodní partnery v reálném čase v mnoha informačních systémech, na webové stránce, nebo pomocí pravidelných reportů zasílaných přímo do e-mailu. Společnost prochází data z insolvenčního, živnostenského a obchodního rejstříku, data pojišťoven a dalších registrů, která obsahují informace o hospodaření a schopnostech dostát svým závazkům. [11]

Při vyhledávání různých ekonomických subjektů podle IČO, můžeme vidět tři stavy semaforu, který rozlišuje aktuální bonitu. Zelený semafor představuje spolehlivého obchodního partnera, u kterého není třeba se aktuálně obávat platební neschopnosti. Oranžové upozornění ve tvaru vykřičníku nabádá k opatrnosti a případné komunikaci se zákazníkem, zda v budoucnu nenastanou problémy. Červený kříž je znamení vysokého rizika obchodování, kdy obchodování není vůbec doporučeno z důvodu zvýšené šance

na nesplacení pohledávky nebo pouze se silným opatřením, jako platba za zboží předem nebo poskytnutí zálohy na zboží.

Obrázek 21 - Semafor rizika [Vlastní zpracování podle zdroje XY]

3.7 Cybersoft I6

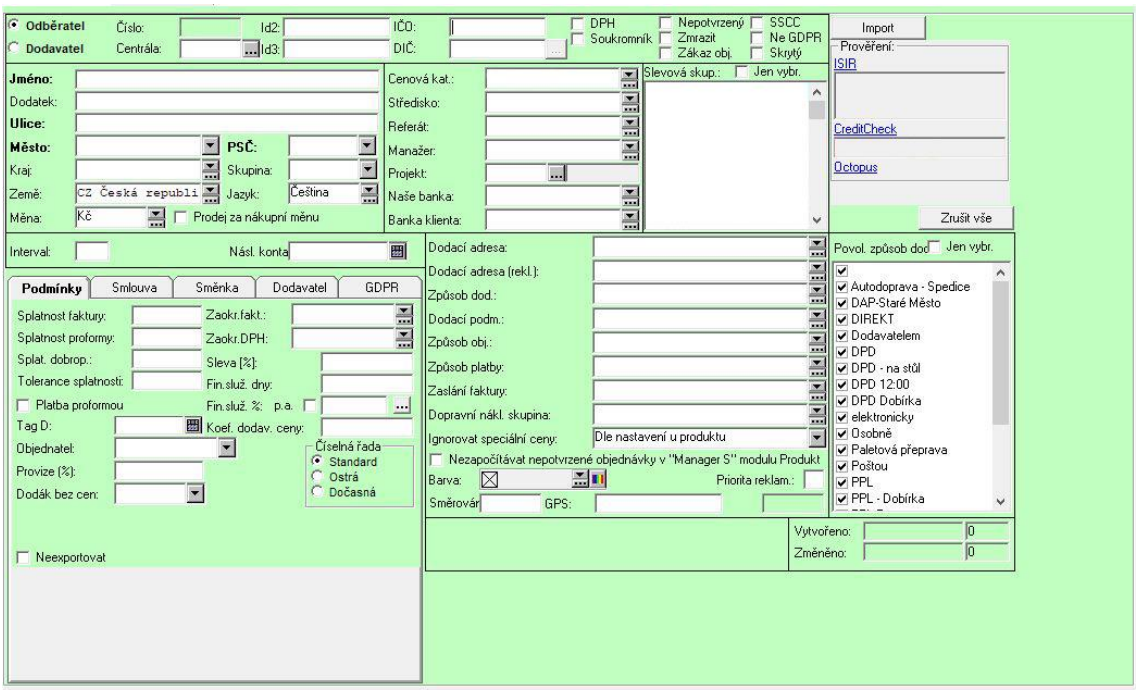
I6 je informační systém kategorie ERP/CRM, který je vyvíjen od roku 2000 a podporuje internetové objednávání pomocí e-shopu. Datová komunikace probíhá za využití technologie klient – server. Systém je založen na databázovém systému Microsoft SQL Server a programovacím jazyku Microsoft Visual Basic. Dokumentace umožňuje propojení produktů třetích stran se systémem I6. Podporuje aplikace balíku Microsoft Office především Word a Excel. Výstup umožňuje exportování do formátu PDF nebo vytištění na tiskárně. [12]

3.7.1 Založení odběratele / dodavatele

V případě zakládání dodavatele či odběratele, který není ještě založený a má své IČO, systém umožňuje rychlé importování dat o subjektu. I6 nahrává data z administrativního registru ekonomických subjektů tzv. ARES a vyplní veřejně dostupné informace jako jméno subjektu a adresu.

V případě, že se nejedná o ekonomický subjekt, který nemá přiřazené IČO. Je nutné vyplnit informace ručně.

V obou případech je nutné přidat středisko, ke kterému obchodnímu zástupci bude zákazník patřit a nastavit další formality jako splatnost faktury, dodací adresu, pokud se dodací adresa liší od fakturační adresy a vybrat možné způsoby dopravy.



Obrázek 22 - Zakládací formulář [Vlastní zpracování podle společnosti PREMO s.r.o.]

Prověřování kredibility

Po vložení do systému se otevře karta zákazníka/dodavatele a můžeme vidět aktuální bonitu. Šedý pruh značí, že obchodování je bezpečné a momentálně nehrozí riziko. Pokud je barva oranžová či červená, je vhodné kliknout na nápis CreditCheck. Odkaz nás přesune na webovou stánku společnosti Credit Check, s.r.o., kde se dozvíme konkrétní důvod, který je za označením danou barvou.

Obrázek 23 - Prověření kredibility [Vlastní zpracování podle společnosti PREMO s.r.o.]

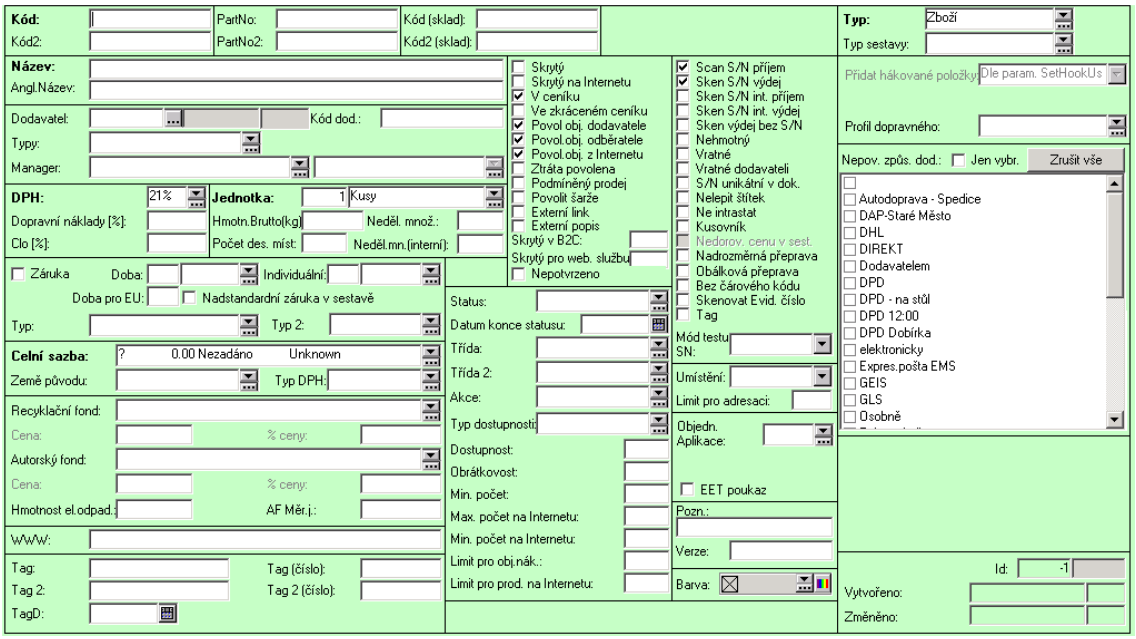
3.7.2 Přidání produktu

Nejdůležitějším bodem společnosti při hledání dodavatelů je důraz na pravidelnou a automatizovanou datovou výměnu. Tato podmínka zefektivňuje objednávání zboží produktového manažera a usnadňuje orientaci v dostupnosti a ceně zboží obchodnímu zástupci při nabízení produktu zákazníkovi.

Společnost navázala přidávání produktů na jednoho ze svých hlavních dodavatelů. Pokud se u tohoto dodavatele objeví nový produkt, je automaticky naimportován do databáze společnosti.

Ruční přidání produktu

Může nastat situace, kdy hlavní dodavatel daný produkt neprodává. V tomto případě projde produktový manažer webové stránky ostatních dodavatelů a pokud nalezne tento produkt u některého z nich, vloží informace do systému pomocí připraveného formuláře.



Obrázek 24 - Vložení produktu [Vlastní zpracování podle společnosti PREMO s.r.o.]

Nejdůležitější při zakládání produktu je přidat správný part number, na který se později párují ceny dodavatele. Pokud by nebyl vložen správně, mohlo by dojít k načtení chybné ceny a společnosti by mohl způsobit ztrátu. Dále musíme přidat název

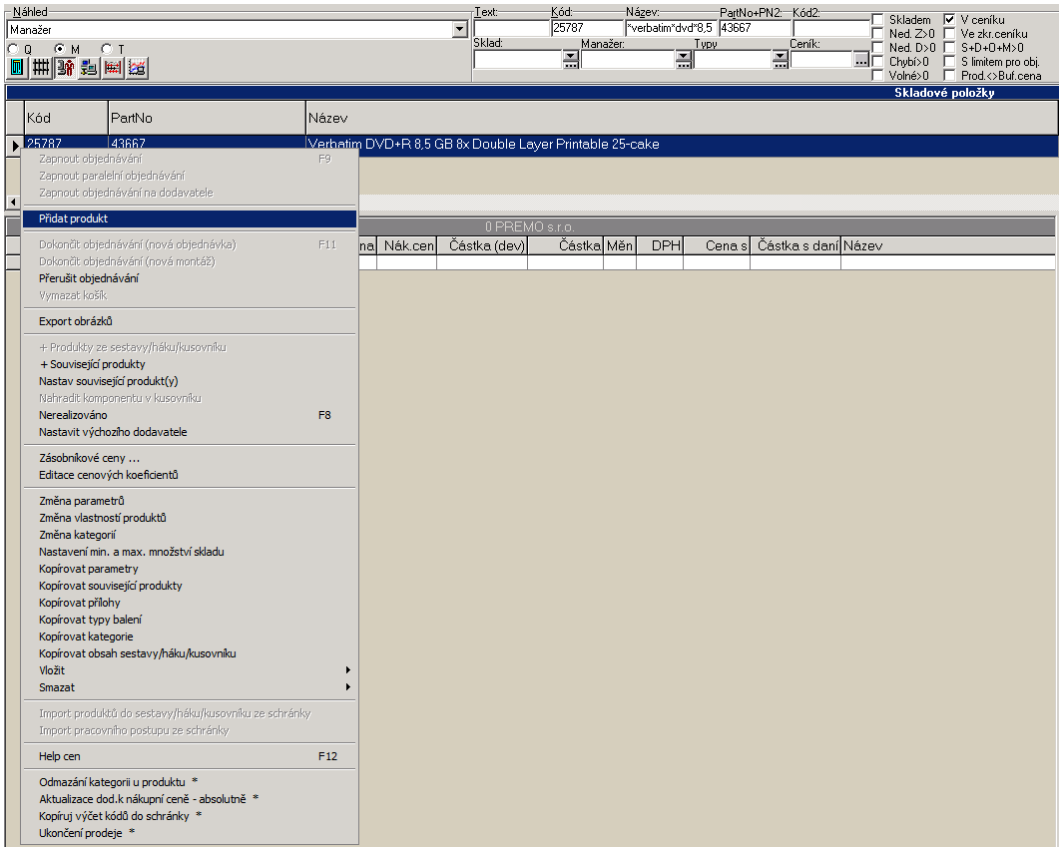
produktu, přiřadit dodavatele, vybrat kategorii, sazbu daně, nastavit možnosti doručování. V případě zahraničního dodavatele vybereme zemi původu a celní sazbu.

3.7.3 Vytvoření objednávky

Pokud je produkt zakoupen pomocí e-shopu, objednávka je založena automaticky. Pokud je zboží skladem, tak pomocí interní funkce systému se odešle požadavek a ve skladu je vytisknuta faktura. Podle produktů na faktuře je nachystáno zboží a v hlavičce jsou uvedeny informace, které rozlišují, jak bude zboží doručeno zákazníkovi.

Objednávka může proběhnout také přes telefon nebo pomocí e-mailové komunikace s obchodním zástupcem, který poté objednávku vytváří ručně v systému.

Jestli zákazník objednal více druhů produktů najednou je nejsnazší pro obchodního zástupce vyhledat jeden z produktů pomocí part numberu. Kliknutím přímo na vybraný produkt lze zapnout objednávací modul, který po výběru zákazníka ze systému umožňuje přidávat produkty do nákupního košíku.

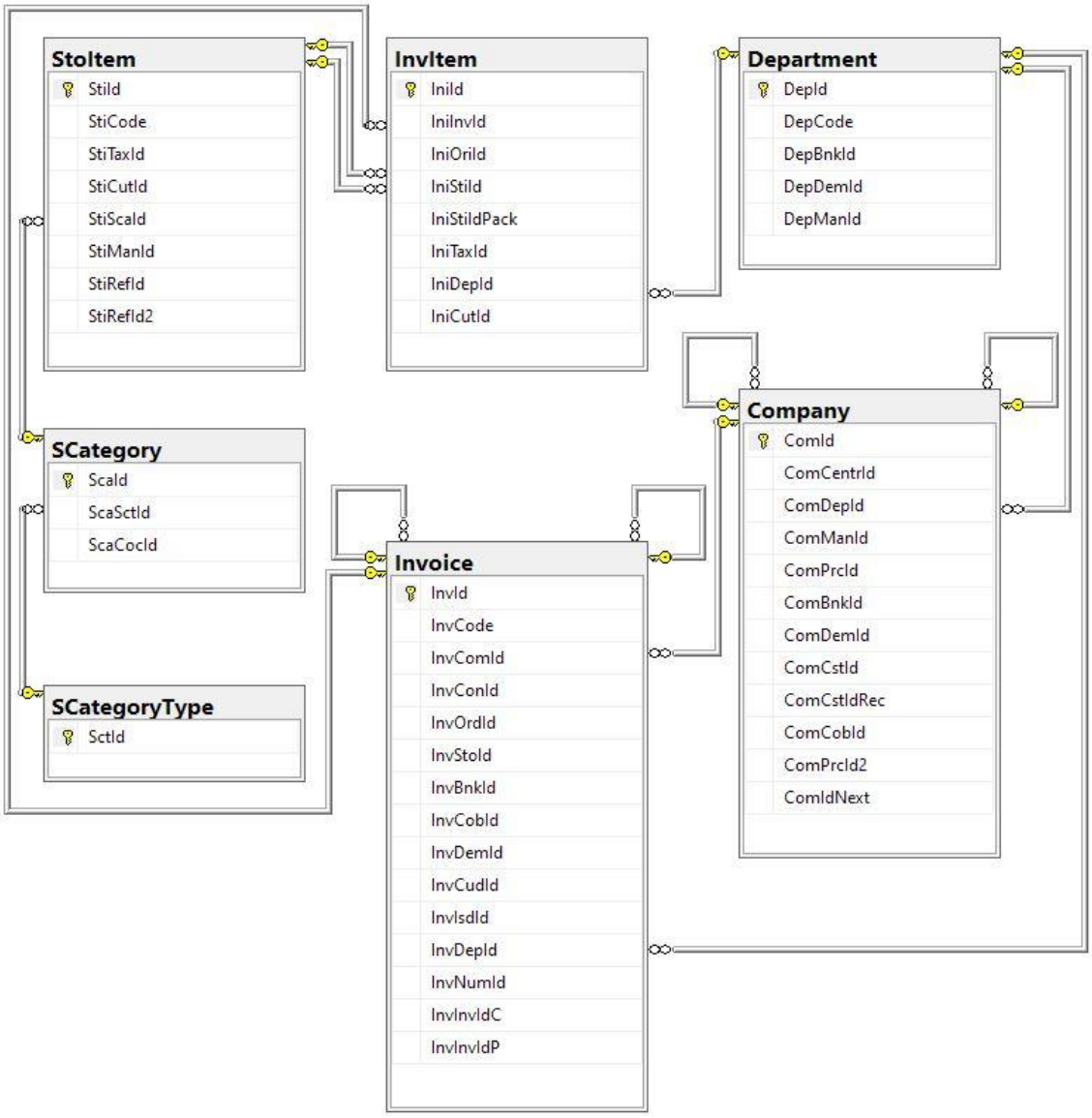


Obrázek 25 - Přidání produktu do košíku [Vlastní zpracování podle společnosti PREMO s.r.o.]

3.8 Databáze a tabulky

Společnost od zavedení informačního systému I6 v roce 2005 ukládá všechny své záznamy do jedné databáze, která aktuálně obsahuje 349 tabulek, jejichž počet se stále zvyšuje s rostoucími požadavky na systém. Jak kvůli uživatelům, tak i kvůli legislativním nařízením přijímaným vládou České republiky. Data jsou uložená v tabulkách, jejichž záznamy přesahují desítky milionů o velikosti 72 GB.

Informační systém I6 má robustní databázi, a proto budu v následující části popisovat jen ty tabulky, které obsahují nejdůležitější data pro splnění cíle této práce.



Obrázek 26 - Diagram DB [Vlastní zpracování]

3.8.1 Invoice

Tabulka zaznamenává každou fakturu, která je ve společnosti vystavena, nebo obdržena od jiné společnosti a obsahuje informace, které ji spojují se zákazníkem a oddělením, které fakturu vystavilo. Obsahuje 119 atributů, jejichž velká část je vyplněna hodnotou NULL.

Důležité atributy:

- InvId – Hodnota je Primárním klíčem a jednoznačným identifikátorem.
- InvCode – Interní označení faktury, které pomáhá rozlišovat účetní doklady.
- InvComId – Odkazuje na společnost, která je spojena se zákazníkem nebo dodavatelem.
- InvConId – Kontakt na osobu, která danou společnost zastupuje (nemusí být nutně uvedena).
- InvOrdId – Spojení s nabídkou, která byla vystavena zákazníkovi,
- InvStoId – Udává sklad, se kterým souvisí položky.
- InvDateAcc – Datum vystavení, nebo přijetí faktury.
- InvDateDue – Datum splatnosti faktury.
- InvVal – Cena faktury s DPH.
- InvDepId – Odkazuje na oddělení, které fakturu vytvořilo, nebo přijalo.

3.8.2 Company

Tabulka obsahuje informace o všech odběratelích a dodavatelích. Záznamy obsahují informace jako název společnosti nebo jméno fyzické osoby. U společnosti ukládá IČO, DIČ a kontakt na osobu, která zastupuje danou společnost. Obsahuje 131 atributů, jejichž velká část je vyplněna hodnotou NULL.

Důležité atributy:

- ComId – Primární klíč a jednoznačný identifikátor každého záznamu.
- ComRegId – IČO
- ComTaxNum – DIČ
- ComDepId – Oddělení, k němuž daný zákazník patří.
- ComName – Název společnosti, institutu nebo jméno fyzické osoby.

- ComStreet – Ulice s číslem popisným.
- ComCity – Název obce nebo města.
- ComPostCode – Poštovní směrovací číslo.
- ComC – Datum vytvoření záznamu o zákazníkovi.
- ComU – Datum posledního vzniklého záznamu s daným zákazníkem.

3.8.3 Department

Záznamy nesou informaci o jménu obchodního zástupce nebo názvu střediska společnosti a jeho kódovém značení. V tabulce nalezneme 36 atributů.

Důležité atributy:

- DepId – Primární klíč a jednoznačný identifikátor oddělení.
- DepCode – Označení oddělení kódem.
- DepName – Jméno obchodního zástupce nebo název oddělení.

3.8.4 InvItem

Seznam produktů a prodaných kusů s cenou za kus ke každé vystavené faktuře. Tabulka obsahuje 65 atributů

Důležité atributy:

- IniId – Primární klíč a jednoznačný identifikátor záznamu o položce faktury.
- IniInvId – Cizí klíč tabulky Invoice k přiřazení faktury.
- IniStiId – Cizí klíč k tabulce StoItem k přiřazení produktu.
- IniQty – Počet prodaných kusů produktu.
- IniPrc – Cena za 1 kus bez DPH.
- IniDepId – Cizí klíč pro spojení se střediskem nebo obchodním zástupcem.
- IniScaId – Cizí klíč k zařazení produktu do kategorie.
- IniC – Datum nacenění položky.

3.8.5 StoItem

Tabulka zaznamenává název kategorie, do které jsou produkty zařazeny. Obsahuje 19 atributů z nichž nejdůležitější jsou vypsány níže.

Důležité atributy:

- StiId – Primární klíč a identifikátor.
- StiCode – Jednoznačný kód produktu v informačním systému.
- StiPartNo – Značení produktu výrobcem.
- StiName – Název produktu.
- StiQty – Počet prodaných kusů daného produktu.
- StiTaxId – Sazba daně z přidané hodnoty.
- StiScaId – Identifikátor kategorie, do které je produkt přiřazený.
- StiComId – Id dodavatele produktu.

3.8.6 SCategory

Tabulka zaznamenává název kategorie, do které jsou produkty zařazeny. Obsahuje 19 atributů z nichž nejdůležitější jsou vypsány níže.

Důležité atributy:

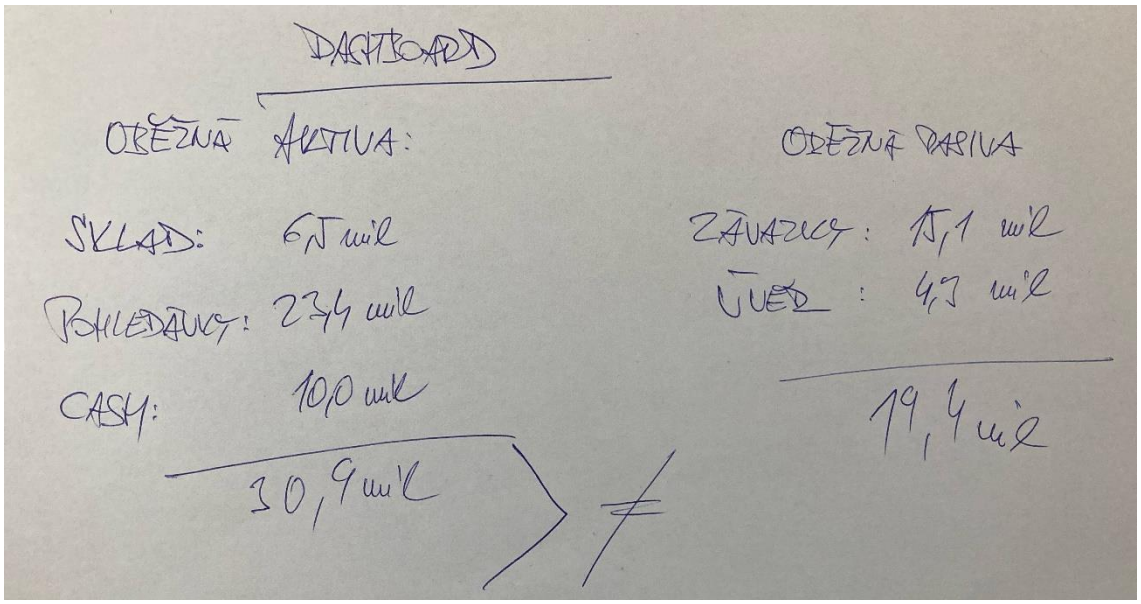
- ScaId – Primární klíč a identifikátor kategorie.
- ScaName – Název kategorie.

3.9 Sestavy a Reporty

Informační systém I6 nabízí možnost vytváření vlastních sestav, jejichž zobrazení je z časového hlediska náročnější na přípravu. Prostředí je přiměřeně uživatelsky příznivé a intuitivní.

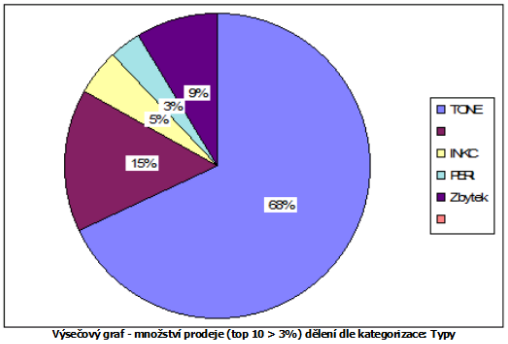
Vedení společnosti se zajímá na týdenní bázi o oběžná aktiva a oběžná pasiva. K tomu, aby vedení získalo tyto potřebné informace, musí projít 5 různých sestav z nichž si vypisuje sumarizované hodnoty na papír a tím získává aktuální přehled.

- Stav pokladen a bankovních účtů
- Cena pohledávek
- Cena skladových zásob
- Závazky
- Úvěry

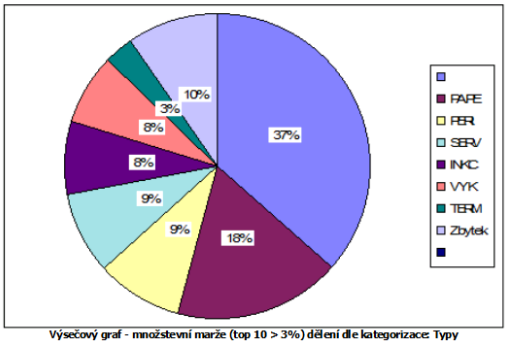


Obrázek 27 - Ručně vytvořená sestava [Vlastní zpracování]

Vytvoření sestavy statistiky prodeje umožňuje nastavit určité období, které chceme sledovat. Vybrat kategorii produktů, výrobce produktů, zákazníka u kterého nás zajímá počet prodaných kusů, které od společnosti koupil. Veškeré hodnoty vidíme pouze staticky k danému zákazníkovi a nemáme možnost interaktivních úprav sestavy. Pokud nás zajímá detailnější sestava, musíme zadat nový filtr a vytvořit novou konkrétnější sestavu.



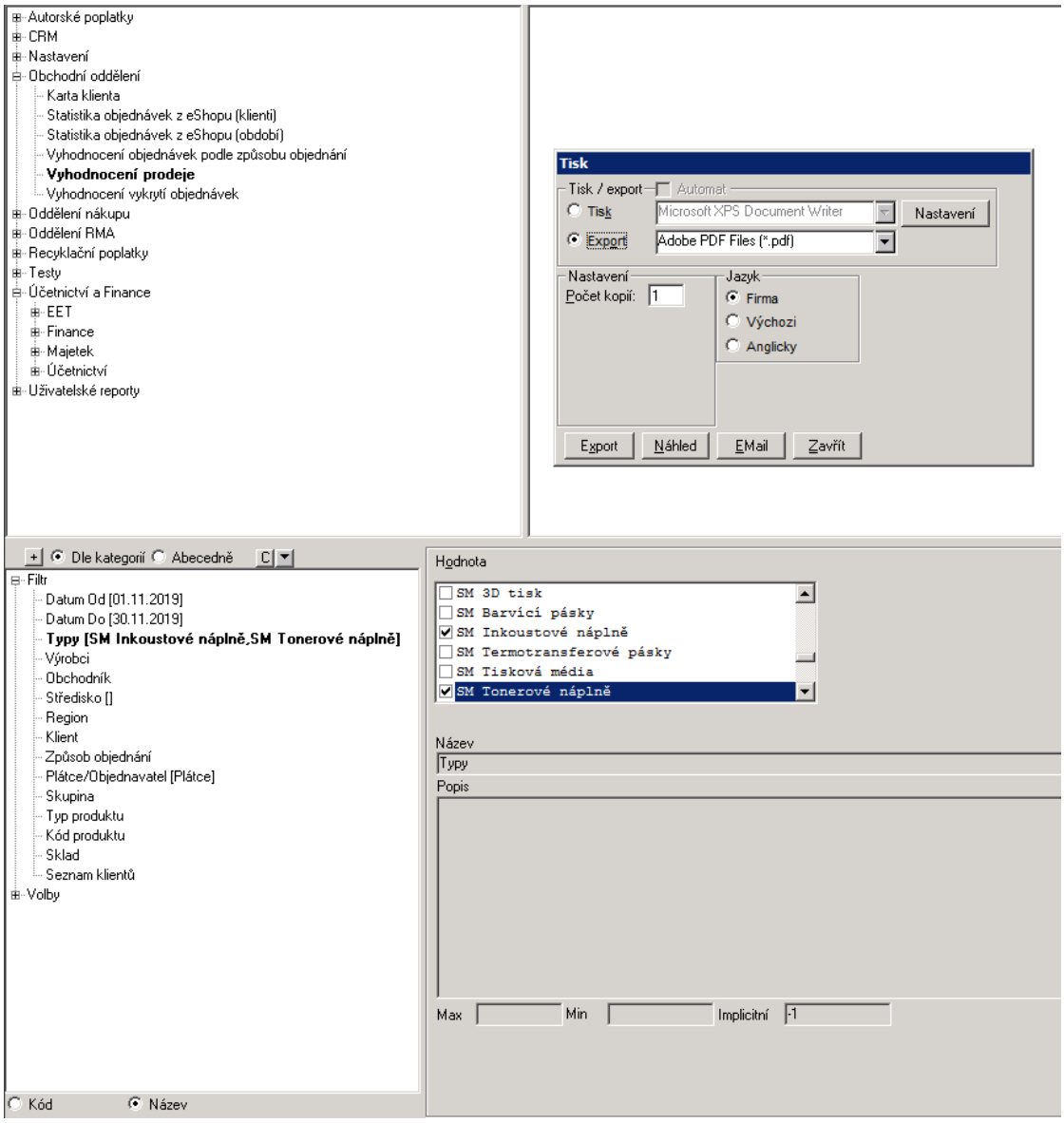
Výsečový graf - množství prodeje (top 10 > 3%) dělení dle kategorizace: Typy



Výsečový graf - množstevní marže (top 10 > 3%) dělení dle kategorizace: Typy

Kód	Název	Počet	Hodnota	H[%]	ΔH[%]	Marže	M[%]	ΔM[%]	M1[%]
1.	**Příslušenství k IT**	1 026	3 067 452	15.18	15.18	358 123	165.11	36.55	11.67
2. PAPE	**SM Tisková média**	22 344	556 841	2.76	2.76	172 788	79.66	17.63	31.03
3. PERI	**Tiskárny, kopírky, faxy, MFC**	246	663 500	3.28	3.28	89 065	41.06	9.09	13.42
4. SERV	**Služby**	4 409	85 528	0.42	0.42	85 528	39.43	8.73	100.00
5. INKC	**SM Inkoustové náplně**	1 555	975 747	4.83	4.83	75 982	35.03	7.75	7.79
6. VYK	**Prázdné kazety**	6 211	225 845	1.12	1.12	74 151	34.19	7.57	32.83
7. TERM	**SM Termotransferové pásky**	1 227	130 415	0.65	0.65	29 906	13.79	3.05	22.93
8. SW	**Software**	74	101 888	0.50	0.50	17 484	8.06	1.78	17.16
9. KABR	**Kabely, strukturovaná kabeláž**	683	69 149	0.34	0.34	17 023	7.85	1.74	24.62
10. NET	**Síťové prvky**	34	126 805	0.63	0.63	13 909	6.41	1.42	10.97
11. POKO	**PC komponenty**	98	121 954	0.60	0.60	9 366	4.32	0.96	7.68
12. OTHE	**Ostatní**	591	14 277	0.07	0.07	8 929	4.12	0.91	62.54
13. RIBO	**SM Barvicí pásky**	412	96 412	0.48	0.48	8 353	3.85	0.85	8.66
14. SPE	**Spotřební elektronika**	252	100 364	0.50	0.50	6 935	3.20	0.71	6.91
15. PAK	**Paměťové karty, čtečky**	15	59 243	0.29	0.29	4 425	2.04	0.45	7.47
16. KANT	**Kancelářské potřeby a zařízení**	46	7 790	0.04	0.04	1 658	0.76	0.17	21.29
17. BARC	**Brašny, pouzdra**	22	24 446	0.12	0.12	1 485	0.68	0.15	6.07
18. BAZ	**Bazar**	1	6 090	0.03	0.03	1 419	0.65	0.14	23.30
19. ARCH	**Baterie, nabíječky baterií**	13	6 441	0.03	0.03	981	0.45	0.10	15.24
20. CLEA	**Čisticí prostředky**	18	2 375	0.01	0.01	916	0.42	0.09	38.55
21. BACK	**Záložní zdroje, UPS**	3	3 282	0.02	0.02	683	0.31	0.07	20.81
22. SEPA	**Servisní díly**	2	8 419	0.04	0.04	403	0.19	0.04	4.79
23. DATC	**Datová média, CD, DVD,..**	100	1 510	0.01	0.01	295	0.14	0.03	19.54
24. PUBL	**Reklamní a provozní materiály**	329	8 208	0.04	0.04	-110 972	-51.16	-14.55	-1 352.00
25. TONE	**SM Tonerové náplně**	7 313	13 741 498	68.01	68.01	-651 942	-300.58	-85.45	-4.74
	Celkem +:		**20 205 479**			**979 809**			
	Celkem -:		**0**			**-762 914**			
	Celkem:	**47 023**	**20 205 479**			**216 895**	**=> 1.07%**		

Obrázek 28 - Sestava prodeje dle kategorií [Vlastní zpracování podle společnosti PREMO s.r.o.]



Obrázek 29 - Výběr sestav a filtrů [Vlastní zpracování podle společnosti PREMO s.r.o.]

Reporty

Společnost poskytuje každý týden report o prodeji všech produktů společnosti GfK Czech, s.r.o., která je dceřinou společností německé organizace. Ta se zabývá průzkumem trhu a výměnou za poskytovaná data zasílá kvartálně analýzu trhu.

Report je poskytován v podobě .xls souboru. Zahrnuje part number, interní kód produktu, název, kategorii produktu, výrobce, prodané kusy a průměrnou cenu.

	A	B	C	D	E	F	G
1	PartNo	Kod	Typ	Vyrobce	Nazev	KS	Prum_Cena
2	003R94641	51081	SM Tisková média	Xerox	Kanc.papír Xerox Colotech,A4,500 listů,90g	60	187,55
3	003R94642	51082	SM Tisková média	Xerox	Kanc.papír Xerox Colotech,A3,500 listů,90g	40	355,74
4	003R94671	07653	SM Tisková média	Xerox	Kanc.papír Xerox Colotech,A4,250 listů,250g	1	346,06
5	006R01461	45574	SM Tonerové náplně	Xerox	Xerox WorkCentre 7120v, toner black 22k	2	1935,4
6	006R01462	45577	SM Tonerové náplně	Xerox	Xerox WorkCentre 7120v, toner yellow 15k	1	2611,18
7	006R01463	45575	SM Tonerové náplně	Xerox	Xerox WorkCentre 7120v, toner magenta 15k	1	3465,44
8	006R01464	45576	SM Tonerové náplně	Xerox	Xerox WorkCentre 7120v, toner cyan 15k	2	2627,52
9	006R01693	152783	SM Tonerové náplně	Xerox	Xerox Black Toner Cartridge pro DocuCentre SC2020 (9000 str.)	4	1128,63
10	006R01694	152785	SM Tonerové náplně	Xerox	Xerox Cyan Toner Cartridge pro DocuCentre SC2020 (3000 str.)	3	884,91
11	006R01695	152784	SM Tonerové náplně	Xerox	Xerox Magenta Toner Cartridge pro DocuCentre SC2020 (3000 str.)	2	897,22
12	006R01696	152786	SM Tonerové náplně	Xerox	Xerox Yellow Toner Cartridge pro DocuCentre SC2020 (3000 str.)	1	883,3
13	008R12941	70298	SM Tonerové náplně	Xerox	Xerox Sešívací svorky do finišeru, 3 kazety x 5000 kusů celkem 15000 kusů	6	1754,5
14	008R13088	107948	SM Tonerové náplně	Xerox	Xerox WorkCentre 7120, fuser unit 100k (R8)	1	5263,5
15	008R13089	74111	Tiskárny, kopírky, faxy, MFC	Xerox	Xerox odpadní nádoba pro WC7120 (33.000 str) (R5)	9	936
16	008R13177	210910	SM Tonerové náplně	Canon	Xerox WorkCentre 7970 - 5000 sponky - svorka kazety - pro AltaLink B8145, B8155, B8170, C8130, C8135, C8145, C8155, C817	1	2393,38
17	00XL199	202181	PC komponenty	Lenovo	Lenovo RS232 cable pro ThinkCentre FRU 00XL199	1	958,32
18	01247401	121078	SM Tonerové náplně	OKI	OKI ES8430 8451 8461 8460, válec yellow 20k/30k	1	2778,16
19	01247402	121077	SM Tonerové náplně	OKI	OKI ES8430 8451 8461 8460, válec magenta 20k/30k	1	2778,16
20	01247403	121076	SM Tonerové náplně	OKI	OKI ES8430 8451 8461 8460, válec cyan 20k/30k	1	2778,16
21	01247404	131817	SM Tonerové náplně	OKI	OKI ES8430 8451 8461 8460, válec black	2	2164,69
22	02300BK08945	144279	SM Termotransferové pásky	Zebra	Zebra páska 2300 Wax. šířka 89mm. délka 450m	4	220,83
23	02300BK10245	113444	SM Termotransferové pásky	Zebra	ZEBRA páska 2300 Wax. šířka 102mm. délka 450m	12	291,61
24	02300GS06407	29731	SM Termotransferové pásky	Zebra	Zebra páska 2300 Wax. šířka 64mm. délka 74m	41	47,6
25	02300GS11007	12922	SM Termotransferové pásky	Zebra	ZEBRA páska 5319 WAX - 110x74m pro TLP2844	15	67,6
26	0264B002	09469	SM Tonerové náplně	Canon	Canon Laserbase MF-65xx, Toner Cartridge CRG706	5	3023,79
27	03200BK10245	50495	SM Termotransferové pásky	Zebra	Zebra 3200 wax/resin. 102mmx450m 580080000015	36	606,48
28	03200GS08407	32781	SM Termotransferové pásky	Zebra	Zebra páska 3200 Wax/Resin. š 84. návin 74m 580080000010	1032	91,03
29	0332C005	152723	SM Inkoustové náplně	Canon	Canon CLI-571XL CMYK ink valuepack + 50xPP201 fotopapír 10x15	2	1282,6
30	0452B002	22875	SM Tonerové náplně	Canon	Canon iR-C2380 2880, toner black CEXV21	2	884,51
31	0453B002	22876	SM Tonerové náplně	Canon	Canon iR-C2380 2880, toner cyan CEXV21	2	1680,69
32	0454B002	22877	SM Tonerové náplně	Canon	Canon iR-C2380 2880, toner magenta CEXV21	2	1680,69
33	0455B002	22878	SM Tonerové náplně	Canon	Canon iR-C2380 2880, toner yellow CEXV21	2	1680,69
34	0481C002	138563	SM Tonerové náplně	Canon	Canon iR C5535i, C-EXV51 Toner BK	2	3380,14
35	0482C002	138564	SM Tonerové náplně	Canon	Canon iR C5535i, C-EXV51 Toner Cyan	3	7882,75
36	0483C002	138565	SM Tonerové náplně	Canon	Canon iR C5535i, C-EXV51 Toner Magenta	1	6669,52
37	0484C002	138566	SM Tonerové náplně	Canon	Canon iR C5535i, C-EXV51 Toner Yellow	1	6669,52
38	04Y2660	210912	Příslušenství k IT	Lenovo	Lenovo CZ klávesnice pro ThinkPad Edge E531 E540 FRU 04Y2660 04Y2697	1	2032,8
39	050100000	208736	SM 3D tisk	Filament PM/Plasty Mladec	Filament PM tisková struna (filament) PLA 1.75mm,1kg,oranžová	1	517,88
40	050180002	208735	SM 3D tisk	Filament PM/Plasty Mladec	Filament PM tisková struna (filament) PLA 1.75mm,2kg,černá	2	975,26
41	05095GS06407	92808	SM Termotransferové pásky	Zebra	Zebra páska 5095 Resin. šířka 64mm. délka 74m	10	135,52
42	0615B043	92417	SM Inkoustové náplně	Canon	Canon iP 1600 2200, MP 150, multipack black+color PG40+CL41	1	896,61
43	079214	211140	Síťové prvky	oem	Legrand 079214	2	196,02
44	09002308	00439	SM Barvící pásky	OKI	Oki ML 3410, páska	4	968

Obrázek 30 - Report Gfk [Vlastní zpracování podle společnosti PREMO s.r.o.]

Další týdenní report je generován společnosti HP Inc., která tím provádí kontrolu svých smluvních partnerů, zda nakupují produkty z oficiální distribuce a nepodílí se tak na rozšiřování a podpoře šedého trhu. PREMO s.r.o. zasílá atributy jako produktové číslo HP, interní kód partnera, název produktu, prodané kusy, zůstatek kusů na skladě, název zákazníka a další.

Reporty vyžaduje také stát. Celní správa České republiky dostává pravidelně informace o obchodu mezi členskými státy Evropské Unie.

Seznam | Položky

Náhled: Sumárně | Intrastat: | Rok: 2021 | Období: 3 Březe | Číslo/Verze: | Vše | Přijetí | Odeslání

Intrastat - položky sumárně

Kód (Celní	Jednotka	Země odesl.	Země pův./Země urč.	Kraj určení/původu	Fakt. hodn.	Stat.hodn.	Hmotn.hodn.	Počet	Typ transakce	Dod. podm.	Typ přepravy	Název (Celní sazba)	Speci	Číslo/Verze	DIČ moje
32159020	ks	DE	CN	ZLI	7 801,00	7 801,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	ID	ZLI	1 532,00	1 532,00	0,110	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	JP	ZLI	322 886,00	322 886,00	140,000	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	MX	ZLI	12 064,00	12 064,00	0,765	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	MY	ZLI	18 132,00	18 132,00	3,000	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	SG	ZLI	3 410,00	3 410,00	0,360	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	DE	XU	ZLI	854,00	854,00	0,041	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	EE	CN	ZLI	38 113,00	38 113,00	15,000	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	PL	CN	ZLI	1 333,00	1 333,00	0,499	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	PL	JP	ZLI	4 287,00	4 287,00	0,720	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
32159020	ks	PL	MY	ZLI	7 304,00	7 304,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Inkoustové kazety (bez			CZ26251531
37079020	ks	DE	TH	ZLI	2 515,00	2 515,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Vývojky,ustalovače pro			CZ26251531
37079090	ks	DE	XU	ZLI	1 308,00	1 308,00	0,455	0	11-Transakce zahrnující	M	3-silniční doprava	Citlivé emulze - Toner E			CZ26251531
37079090	ks	LT	CN	ZLI	62 700,00	62 700,00	32,000	0	11-Transakce zahrnující	M	3-silniční doprava	Citlivé emulze - Toner E			CZ26251531
84439990	ks	AT	CN	ZLI	45 829,00	45 829,00	54,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	BE	CN	ZLI	40 106,00	40 106,00	22,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	BE	JP	ZLI	10 073,00	10 073,00	7,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	BE	KR	ZLI	446,00	446,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	BE	XU	ZLI	2 097,00	2 097,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	CN	ZLI	275 568,00	275 568,00	211,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	JP	ZLI	2 004 764,00	2 004 764,00	1 165,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	MX	ZLI	2 261,00	2 261,00	0,855	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	MY	ZLI	4 835,00	4 835,00	3,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	PL	ZLI	23 568,00	23 568,00	6,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	DE	XU	ZLI	33 042,00	33 042,00	13,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	EE	CN	ZLI	11 434,00	11 434,00	5,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	EE	JP	ZLI	113 351,00	113 351,00	164,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	LT	CN	ZLI	18 053,00	18 053,00	16,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	LT	JP	ZLI	28 956,00	28 956,00	21,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	LT	XU	ZLI	11 913,00	11 913,00	2,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	NL	XS	ZLI	107 351,00	107 351,00	141,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	CN	ZLI	42 273,00	42 273,00	26,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	CH	ZLI	4 363,00	4 363,00	1,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	IN	ZLI	2 221,00	2 221,00	4,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	JP	ZLI	286 878,00	286 878,00	128,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	KR	ZLI	19 496,00	19 496,00	9,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84439990	ks	PL	PL	ZLI	13 975,00	13 975,00	4,000	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
84713000	ks	SK	SK	ZLI	4 671,00	4 671,00	3,000	1	11-Transakce zahrnující	M	3-silniční doprava	Počítače do 10 Kg vtill			CZ26251531
84716070	ks	DE	CN	ZLI	3 277,00	3 277,00	2,000	13	11-Transakce zahrnující	M	3-silniční doprava	Ostatní - Vstupní nebo			CZ26251531
84719000	ks	DE	MX	ZLI	1 206,00	1 206,00	0,567	1	11-Transakce zahrnující	M	3-silniční doprava	Ostatní výrobky pro IT			CZ26251531
84733080	ks	DE	CN	ZLI	3 199,00	3 199,00	0,900	0	11-Transakce zahrnující	M	3-silniční doprava	Části, součásti a přísluš			CZ26251531
85181000	ks	DE	CN	ZLI	7 157,00	7 157,00	1,000	0	11-Transakce zahrnující	M	3-silniční doprava	Mikrofony a jejich stoja			CZ26251531
85183000	ks	DE	CN	ZLI	498,00	498,00	0,088	0	11-Transakce zahrnující	M	3-silniční doprava	Sluchátka všech druhů			CZ26251531
85183000	ks	DE	VN	ZLI	758,00	758,00	0,036	0	11-Transakce zahrnující	M	3-silniční doprava	Sluchátka všech druhů			CZ26251531
85232915	ks	DE	JP	ZLI	38 133,00	38 133,00	12,000	45	11-Transakce zahrnující	M	3-silniční doprava	Pásky,disky magnet,ne			CZ26251531
85444290	ks	DE	CN	ZLI	492,00	492,00	0,086	0	11-Transakce zahrnující	M	3-silniční doprava	Elektrické vodič - ostat			CZ26251531
96121010	ks	DE	BE	ZLI	7 155,00	7 155,00	6,000	0	11-Transakce zahrnující	M	3-silniční doprava	Pásky do psacích stroj			CZ26251531
96121010	ks	DE	CN	ZLI	2 231,00	2 231,00	1,000	0	11-Transakce zahrnující	M	3-silniční doprava	Pásky do psacích stroj			CZ26251531
96121080	ks	DE	BE	ZLI	319,00	319,00	0,051	0	11-Transakce zahrnující	M	3-silniční doprava	Pásky do tiskáren EUF			CZ26251531
96121080	ks	DE	CN	ZLI	5 177,00	5 177,00	1,000	0	11-Transakce zahrnující	M	3-silniční doprava	Pásky do tiskáren EUF			CZ26251531
96121080	ks	SK	FR	ZLI	1 030,00	1 030,00	1,000	0	11-Transakce zahrnující	M	3-silniční doprava	Pásky do tiskáren EUF			CZ26251531

Obrázek 31 - Report pro Intrastat [Vlastní zpracování podle společnosti PREMO s.r.o.]

3.10 Zdroje dat

PREMO s.r.o.

Hlavním zdrojem jsou interní data, která společnost vytváří každodenními procesy. Tato data jsou ukládána do relační databáze a jsou pravidelně zálohována. Záloha byla poskytnuta správcem po vyžádání a souhlasu vedení společnosti ve formátu .bak.

- Sleva na zboží HP Inc.

Společnost je partnerem nadnárodní korporace, což nese určité výhody i povinnosti. Výhodou pro společnost je možnost získání speciálních produktových cen pro zákazníky v B2B a B2G segmentu a tím zvýšit svou konkurenceschopnost. Po schválení kontraktu společnost HP Inc. zpětně doplácí část ceny. Tyto informace a konkrétní částky už nejsou v systému doplňovány položkově, ale pouze globálně. Což v některých detailních sestavách způsobuje záporné marže.

Každý měsíc je vytvářena detailní tabulka v excelu ve formátu .xlsx. Pro doplacení částek chystá produktový manažer podklady, které obsahují číslo faktury, produktové číslo, kusy, částku k doplacení a další informace.

01/06/2021	3897847	CF279A	1	2922008469	12/10/2020	5,96
01/06/2021	3897839	CF364A	2	2922008270	12/15/2020	76,72
01/06/2021	3897845	CF364A	1	2922008718	12/21/2020	38,36
01/06/2021	3897863	CE278AC	2	2922008558	12/11/2020	45,12

Obrázek 32 - Dorovnání cen [Vlastní zpracování podle společnosti PREMO s.r.o.]

Český statistický úřad

ČSÚ vede registr ekonomických subjektů tzv. RES, u kterého je zaznamenána každá organizační složka státu, právnická i fyzická osoba, která je účetní jednotkou. Informace jsou veřejné a na vyžádání jsou dostupné každému. Kompletní databáze je zaslána ve formátu .csv a poskytnutí může být zpoplatněno. Obsahuje informace jako IČO, adresu, právní formu, kategorii počtu pracovníků, činnost, název a další. [13]

3.11 Zhodnocení současného stavu

Společnost denně produkuje velké množství dat, se kterými následně pracuje dle potřeby na základní úrovni exportů a sestav.

Tabulky v databázi obsahují obrovské množství atributů, které jsou prázdné nebo nenesou žádnou hodnotnou informaci. Nevyužitelná data tak zbytečně komplikují orientaci v databázi a zvyšují nároky na uložiště. Pro analýzy a dolování dat, tak vzniká zbytečné množství atributů, se kterými pracuje.

Před zahájením dolování dat je z hlediska efektivity a kvality výstupu nutné provést přípravu a čištění dat.

4 Vlastní návrhy řešení

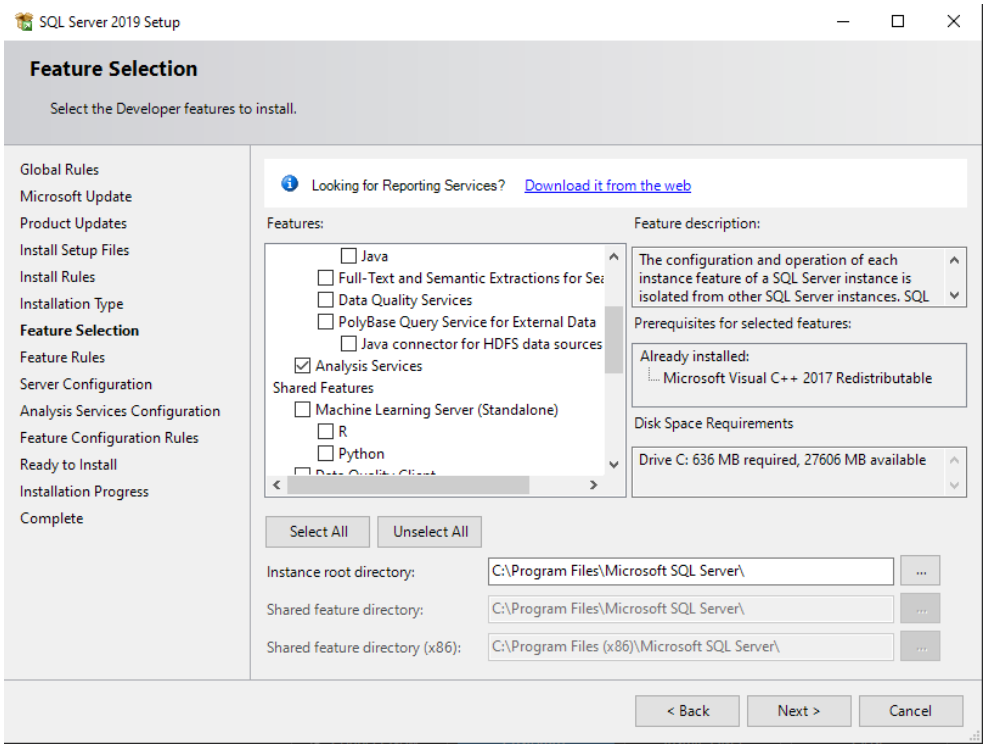
V této části se budu zabývat vlastním návrhem řešení.

4.1 Nutná rozšíření softwaru

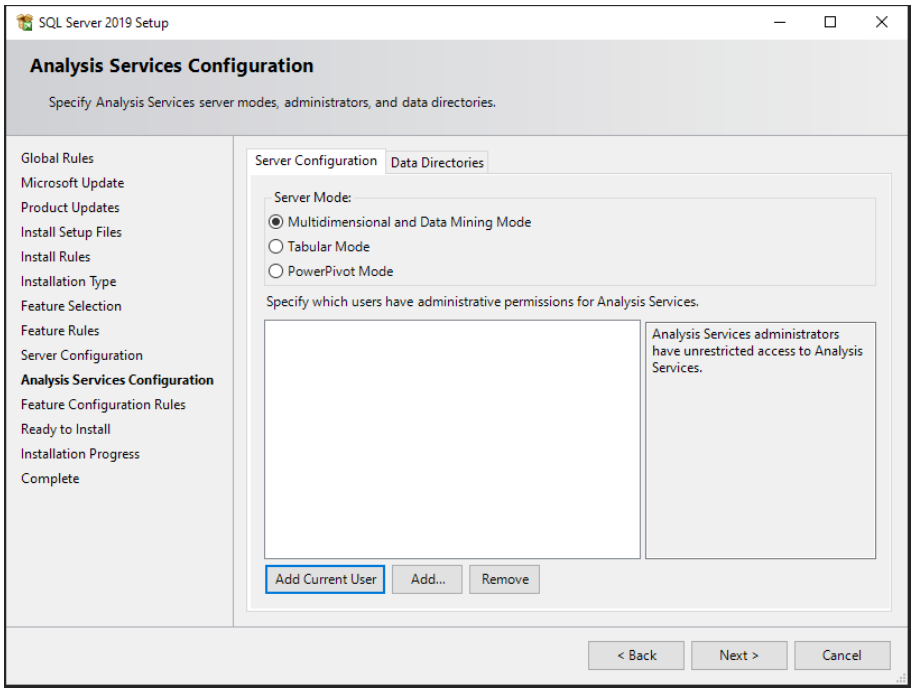
V této části chci zmínit softwarová rozšíření, která je nutné přidat po dokončení instalace základních částí programů Microsoft SQL Server, Microsoft SQL Server Management Studio a Microsoft Visual Studio.

Analytické služby v SQL Serveru

Microsoft SQL Server v základní instalaci nepodporuje analytické služby. Ty je nutné po dokončení přidat pomocí přizpůsobení vlastního výběru funkcí. Je taky potřeba povolit přístup uživatelům a nastavit server na multidimenzionální režim a režim dolování dat. Bez analytických služeb SQL Serveru není možné využívat nástroje, které podporují dolování dat a vytvoření datové kostky ve Visual Studiu.



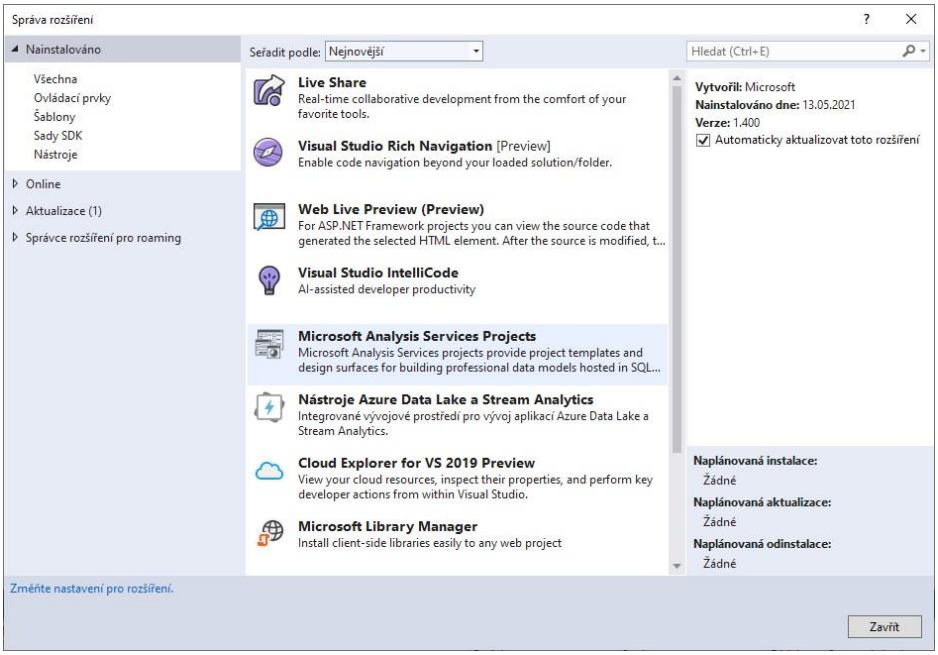
Obrázek 33 - Přidání analytických služeb [Vlastní zpracování]



Obrázek 34 – Výběr režimu a práv uživatelů [Vlastní zpracování]

Microsoft Analysis Services Projects ve Visual Studiu

Důležitým rozšiřujícím nástrojem je projektová šablona pro vytváření multidimenzionálních datových modelů a algoritmů dolování dat v Microsoft Visual Studio. To je možné nainstalovat po spuštění programu. Výběrem rozšíření z horního menu a následným kliknutím na správu rozšíření.

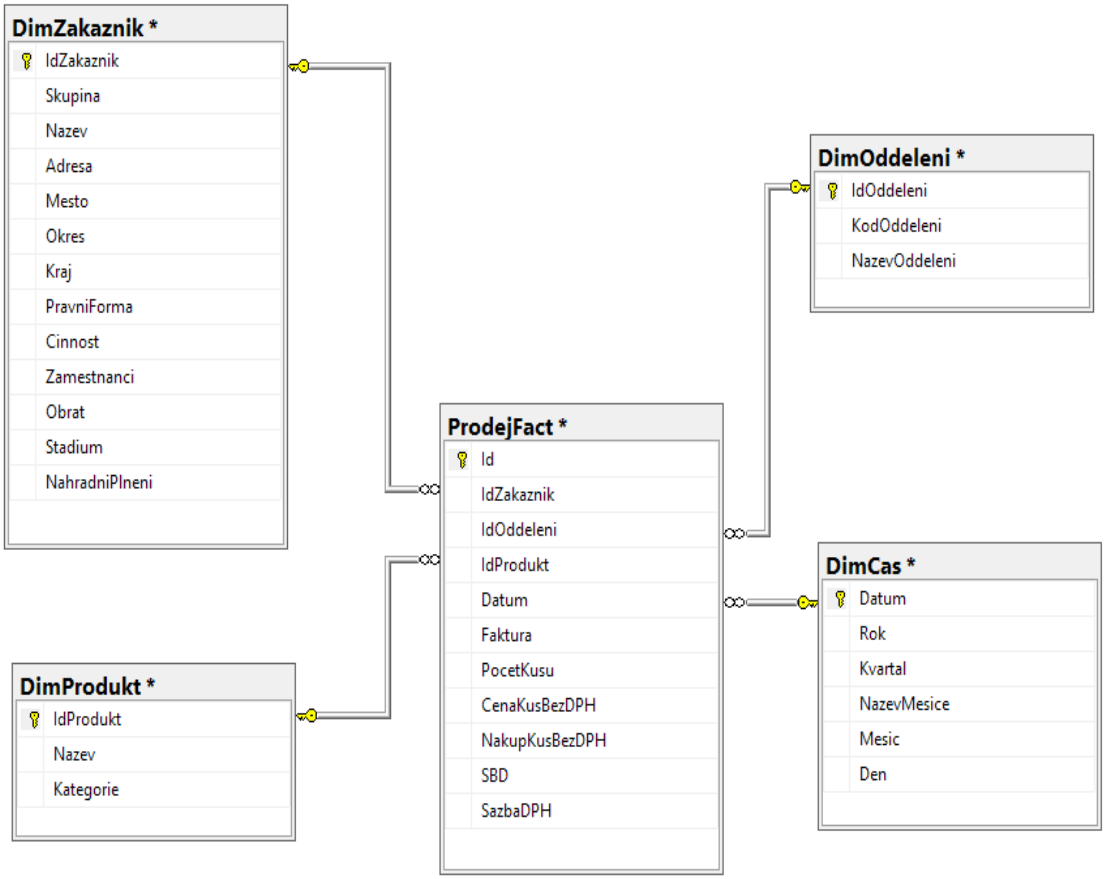


Obrázek 35 - Rozšíření Visual Studia [Vlastní zpracování]

4.2 Datový sklad

Datový sklad je navržen podle schématu hvězdy z důvodu přehlednosti. Centrální částí topologie je tabulka faktů, která obsahuje podstatné informace o prodeji. Okolní tabulky dimenzí poskytují rozšiřující nebo doplňující informace.

Budování skladu proběhne metodou velkého třesku.



Obrázek 36 - Schéma datového skladu [Vlastní zpracování]

4.3 Příprava dat

Příprava je nejdůležitější a časově nejnáročnější částí celého procesu dolování dat. V následující části jsou rozepsány kroky, které zahrnují integraci dat z různých zdrojů do databáze. Vytvořím nové tabulky, do kterých nahraji redukovaná a upravená data. Přidám nové odvozené atributy, které umožní více pohledů v datové kostce.

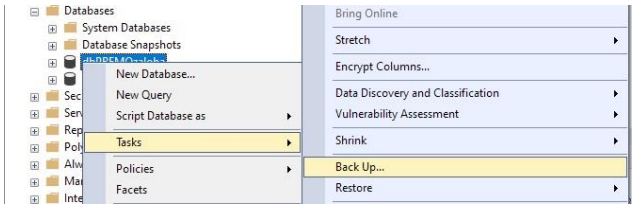
4.3.1 Import zdrojů

Prvním krokem je naplnění dat do databáze, kterou si vytvořím v programu Microsoft SQL Server Management Studio. Novou databázi nazvu „dbPREMOzaloha“, do které obnovím poskytnutou zálohu databáze od společnosti PREMO s.r.o. Poté nahraji další získané zdroje dat z českého statistického úřadu a data společnosti uložené mimo databázi informačního systému, ve kterých je zaznamenáno dorovnání cen výrobcem HP Inc.

Příkazem SQL vytvořím novou databázi, do které obnovím zálohu dat pomocí grafického prostředí od Microsoftu.

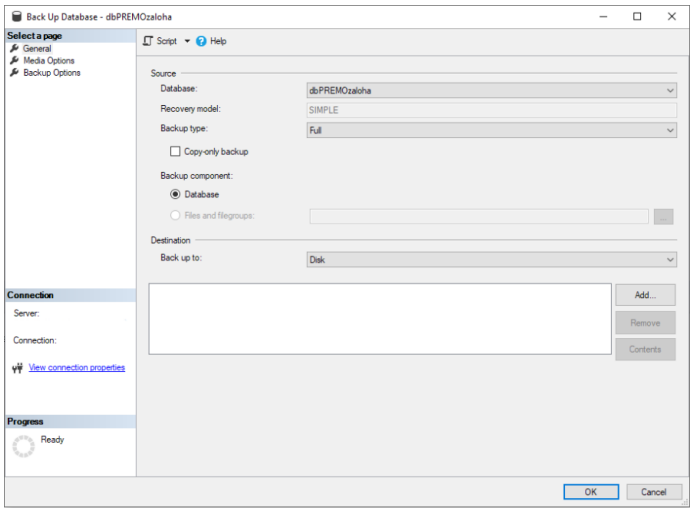
```
CREATE DATABASE dbPREMOzaloha
```

Kliknutím na vytvořenou databázi se otevře menu, ve kterém vybereme záložku „Tasks“ a následně „Back UP…“, který nám otevře nové okno.



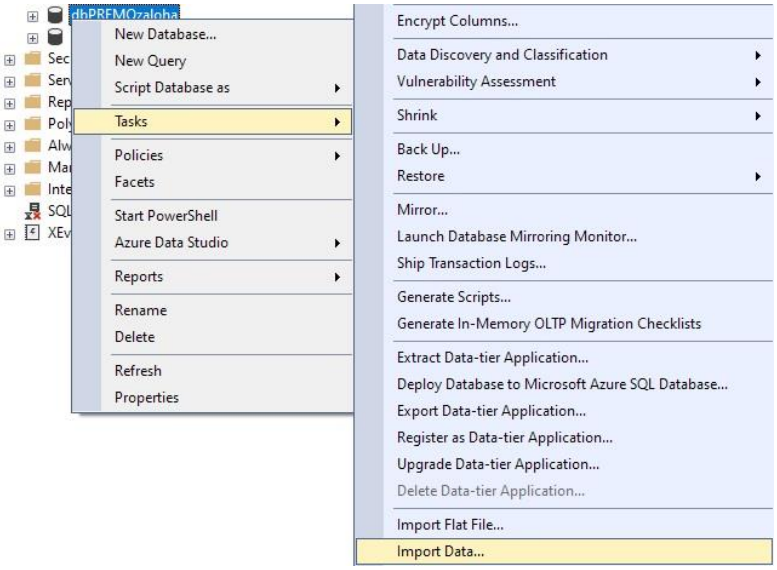
Obrázek 37 – Menu back up [Vlastní zpracování]

Na stránce „General“ v části destination přidáme cestu k souboru kliknutím na tlačítko „Add…“. Následně přepneme stránku okna na „Media Options“, kde v sekci „Overwrite media“ zaklikneme „Over write all existing backup sets“ a dáme „OK“.



Obrázek 38 - Obnovení dat ze zálohy [Vlastní zpracování]

Po obnovení zálohy můžeme začít importovat data z dalších zdrojů. K tomuto slouží „SQL Server Import and Export Wizard“, který provede uživatele importem dat do databáze. Ten umožňuje nahrát datové zdroje typu textový soubor, soubor Excelu, databázi z programu Microsoft Access a další.



Obrázek 39 - Import dat [Vlastní zpracování]

4.3.2 Tabulky dimenzí a faktu

Tabulky jsou vytvořeny do nové databáze s názvem PREMO.

Tabulka faktu ProdejFact

```
CREATE TABLE ProdejFact(
    IdZakaznik int
    ,IdOddeleni smallint
    ,IdProdukt int
    ,Datum date
    ,Faktura char(10)
    ,PocetKusu int
    ,CenaKusBezDPH money
    ,NakupKusBezDPH money
    ,SBD money
    ,SazbaDPH tinyint);

ALTER TABLE ProdejFact
ADD Id int NOT NULL IDENTITY (1,1),
CONSTRAINT PK_ProdejFact PRIMARY KEY (Id)
```

Tabulka faktu je vytvořena pomocí jazyka SQL a obsahuje identifikační čísla k tabulkám dimenzí a podstatné informace spojené s prodejem. Následně je přidán primární klíč, který se vyplňuje automaticky.

```sql
INSERT INTO [PREMO].[dbo].[ProdejFact]

SELECT
      i.InvComId AS IdZakaznik
      ,i.InvDepId AS IdOddeleni
      ,ii.IniStiId AS IdProdukt
      ,CONVERT(DATE,i.InvDateAcc) AS Datum
      ,i.InvCode AS Faktura
      ,ii.IniQty AS PocetKusu
      ,ii.IniPrc AS CenaKusBezDPH
      ,st.SopPrcP AS NakupKusBezDPH
      ,ISNULL(ks.CZSlevaKS,0) AS SBD
      ,t.TaxRate AS SazbaDPH

  FROM Invoice i
       join InvItem ii on i.InvId = ii.IniInvId
       join StoItem si on ii.IniStiId = si.StiId
       join Tax t on ii.IniTaxId = t.TaxId
       join StoOper st On ii.IniId = st.SopIniId
       full join KurzSBD ks On ks.ProductNr = si.StiPartNo
                               AND ks.InvoicePackNr = i.InvCode
  WHERE InvDateAcc >='2015-01-01 00:00:00'
        and InvDateAcc <='2020-12-31 23:59:59' and InvCode like '29%'
```

Data tabulky faktů jsou nahrána výběrem z tabulek Invoice, InvItem, Tax, StoOper a KurzSBD. StoItem je využita pouze pro vytvoření propojovací vazby k tabulce KurzSBD.

Z Invoice jsou přejaty hodnoty Id zákazníka, Id oddělení, Id produktu, číslo faktury a datum. Ten je ořezaný o časový údaj z původní varianty YYYY-MM-DD HH:MM:SS.

Následující tabulka InvItem poskytuje rozšiřující údaje o počtu prodaných kusů daného produktu a cenu, za kterou byl prodán zákazníkovi.

StoItem obsahuje cenu za jakou byl produkt pořízen od dodavatele. Daňová sazba je uložena v tabulce Tax.

K dopočítání přesné nákupní ceny nám u produktů, u kterých je pro zákazníka získána lepší cena, slouží tabulka KurzSBD v níž jsou uloženy částky, které výrobce zpětně vrací. Pro umožnění správného počítání je u ostatních produktů, na které se dohoda nevztahuje vložena nula pomocí funkce ISNULL.

Podmínka při nahrávání dat ošetřuje vystavené faktury jejichž interní označení začíná vždy hodnotou „29“ a byly vystaveny mezi lety 2015 a 2020.

Dimenze čas

```
CREATE TABLE DimCas (
    Datum date NOT NULL PRIMARY KEY
    ,Rok int NOT NULL
    ,Kvartal nchar(10) NOT NULL
    ,NazevMesice varchar(10) NOT NULL
    ,Mesic int NOT NULL
    ,Den int NOT NULL);
```

Tabulka obsahuje primární klíč datum. Datum je následně rozdělen do dalších atributů, které společně vytváří časovou hierarchii.

```
INSERT INTO [PREMO].[dbo].[DimCas]

SELECT DISTINCT
    CONVERT(DATE,InvDateAcc) AS Datum
    ,YEAR(InvDateAcc) AS Rok
    ,CASE WHEN MONTH(InvDateAcc) = 1 THEN '1. kvartál'
            WHEN MONTH(InvDateAcc) = 2 THEN '1. kvartál'
            WHEN MONTH(InvDateAcc) = 3 THEN '1. kvartál'
            WHEN MONTH(InvDateAcc) = 4 THEN '2. kvartál'
            WHEN MONTH(InvDateAcc) = 5 THEN '2. kvartál'
                        . . .
    END AS Kvartal
           ,CASE WHEN MONTH(InvDateAcc) = 1 THEN 'Leden'
            WHEN MONTH(InvDateAcc) = 2 THEN 'Únor'
            WHEN MONTH(InvDateAcc) = 3 THEN 'Březen'
            WHEN MONTH(InvDateAcc) = 4 THEN 'Duben'
                        . . .
    END AS NazevMesice
    ,MONTH(InvDateAcc) AS Mesic
    ,Day(InvDateAcc) AS Den

FROM [dbPREMOzaloha].[dbo].[Invoice]

WHERE InvDateAcc >='2015-01-01 00:00:00'
      and InvDateAcc <='2020-12-31 23:59:59' and InvCode like '29%'
```

SQL kód je zkrácen o CASE varianty, které stejným způsobem doplňují zbylé kvartály a měsíce.

Při vytváření časové dimenze je využitá jediná tabulka Invoice a její atribut InvDateACC, který označuje den a čas vystavení faktury. Tato hodnota je následně zkrácena na datum bez časového údaje a pomocí vnitřních funkcí jsou z ní získávány hodnoty rok, měsíc a den. Za využití CASE je měsíc rozdělen do čtyř kvartálů a číslo měsíce je pojmenováno.

Každý den je uveden jen jednou díky příkazu distinct.

Dimenze produkt

```
CREATE TABLE DimProdukt (
    IdProdukt int NOT NULL PRIMARY KEY
    ,Nazev varchar(255)
    ,Kategorie varchar(255));
```

V tabulce je uložen primární klíč produktu, název produktu a kategorie, do které produkt patří.

```
INSERT INTO [PREMO].[dbo].[DimProdukt]

SELECT DISTINCT
    si.StiId AS IdProdukt
    ,si.StiName AS Nazev
    ,sc.ScaName AS Kategorie

FROM InvItem ii join StoItem si on ii.IniStiId = si.StiId
     join SCategory sc on si.StiScaId = sc.ScaId

WHERE InvDateAcc >='2015-01-01 00:00:00'
      and InvDateAcc <='2020-12-31 23:59:59' and InvCode like '29%'
```

Z tabulky StoItem nahráváme informace o produktu jako je id a název. K produktu je vždy přiřazen název kategorie, do které patří díky tabulce SCategory. Tabulka InvItem je zde využita pouze v podmínce.

Dimenze oddělení

```
CREATE TABLE DimOddeleni (
    IdOddeleni int NOT NULL PRIMARY KEY
    ,KodOddeleni nchar(6)
    ,NazevOddeleni varchar(50));
```

Obsahuje primární klíč oddělení, interní kód a název.

```
INSERT INTO [PREMO].[dbo].[dimOddeleni]

SELECT DISTINCT
    d.DepId AS IdOddeleni
    ,d.DepCode AS KodOddeleni
    ,d.DepName AS NazevOddeleni

FROM Department d
    join Invoice i ON d.DepId=i.InvDepId

WHERE InvDateAcc >='2015-01-01 00:00:00'
    and InvDateAcc <='2020-12-31 23:59:59' and InvCode like '29%'
```

Z tabulky department je využito id oddělení, jeho interní kód a název oddělení. Podle toho, které oddělení fakturu vytvořilo.

Dimenze zákazník

```
CREATE TABLE DimZakaznik (
    IdZakaznik int NOT NULL PRIMARY KEY
    ,Skupina nchar(3)
    ,Nazev varchar(255)
    ,Adresa varchar(255)
    ,Mesto varchar(50)
    ,Okres varchar(50)
    ,Kraj varchar (50)
    ,PravniForma varchar(255)
    ,Cinnost varchar(255)
    ,Zamestnanci varchar(255)
    ,Obrat varchar (255)
    ,Stadium varchar (255)
    ,NahradniPlneni nchar(1) NOT NULL);
```

Dimenze ukládá důležité informace o zákazníkovi. Název společnosti nebo jméno zákazníka, geografické údaje, rozděluje jej dle interních směrnic do skupin a další.

```
INSERT INTO [PREMO].[dbo].[DimZakaznik]

SELECT DISTINCT
    c.InvComId AS IdZakaznik
    ,CASE WHEN c.ComRegId ='' then 'B2C'
          WHEN c.ComRegId !=''
                  AND cc.pravniForma ='Státní podnik'
                  OR cc.pravniForma ='Organizační složka státu'
                  OR cc.pravniForma ='Kraj'
                  OR cc.pravniForma ='Vysoká škola'
                  OR cc.pravniForma ='školská právnická osoba'
                              . . .
                  OR cc.pravniForma ='Veřejné výzkumné instituce' then 'B2G'
          ELSE 'B2B'
    END AS Skupina
    ,c.ComName AS Nazev
    ,c.ComStreet AS Adresa
    ,c.ComCity AS Mesto
    ,g.okres AS Okres
    ,g.kraj AS Kraj
    ,cc.pravniForma AS PravniForma
    ,cc.nace AS Cinnost
    ,cc.pocetZam AS Zamestnanci
    ,cc.obrat AS Obrat
    ,CASE WHEN c.ComRegId !='' AND cc.semafor = 3 then 'zanikla'
          WHEN c.ComRegId !='' AND cc.semafor != 3 then 'funguje'
    END AS Stadium
    ,CASE WHEN c.ComRegId = '25590511' then '1'
          ELSE '0'
    END AS NahradniPlneni

FROM Company c
     full join CSU cc on cc.ICO = c.ComRegId
     left join CSUg g on g.psc = c.ComPostCode
```

SQL kód je zkrácen o CASE varianty, které stejným způsobem doplňují zbylým institucím podle právní formy název skupiny B2G.

Data dimenze DimZakaznik jsou výběrem ze dvou zdrojů. Interní databáze společnosti a českého statistického úřadu. Jsou využity atributy z tabulek Company a přejmenované tabulky registru ekonomických subjektů RES na CSU a CSUg.

Z tabulky Company je využito Id zákazníka, název, adresa a město.

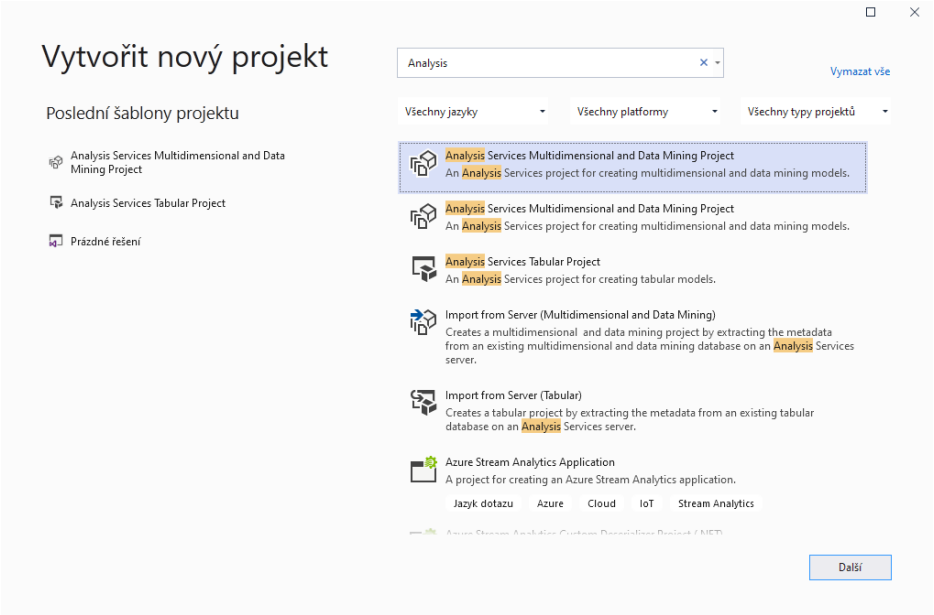
Pomocí atributu právní forma z dat registru ekonomických subjektů českého statistického úřadu a hodnoty IČO z databáze společnosti vznikla podmínka rozdělující zákazníky do skupin. Tato podmínka zařazuje do skupiny označené B2C pokud zákazník nemá IČO. Příznak skupiny B2G přidává zákazníkům, kteří vlastní IČO a jejich právní forma je jedna z uvedených v podmínce. Ostatní zákazníci získávají status skupiny B2B.

Tabulka CSUg přiděluje okres a kraj podle poštovního směrovacího čísla uvedeného v kartě zákazníka.

Atributy právní forma, činnost, zaměstnanci a obrat jsou přidávány na základě IČO. Toto se netýkají zákazníků skupiny B2C u nichž je každý z výše zmíněných atributů vyplněn hodnotou NULL. Činnost je hlavní předmět podnikání zákazníků, který je uveden v obchodním rejstříku.

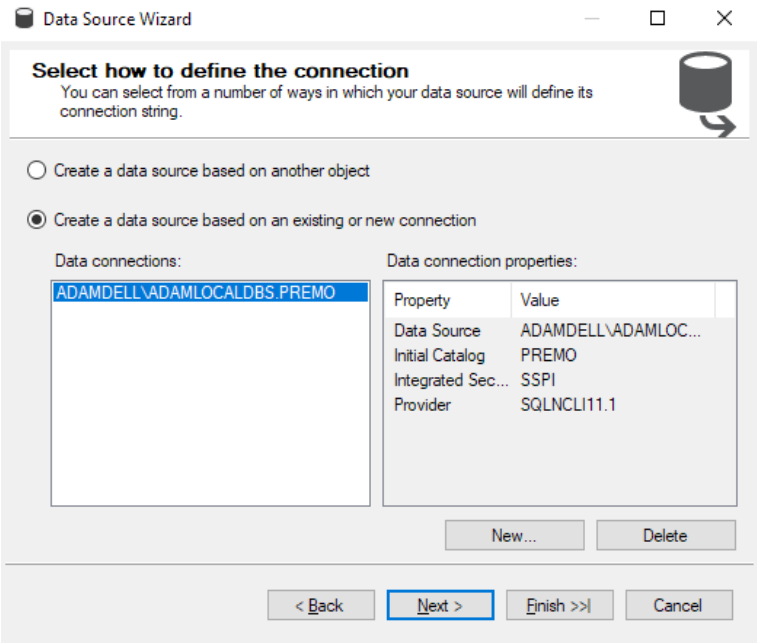
4.4 Propojení zdroje s Visual Studiem

V programu Visual Studio vytvoříme nový projekt podle šablony Analysis Services Multidimensional and Data Mining Project.



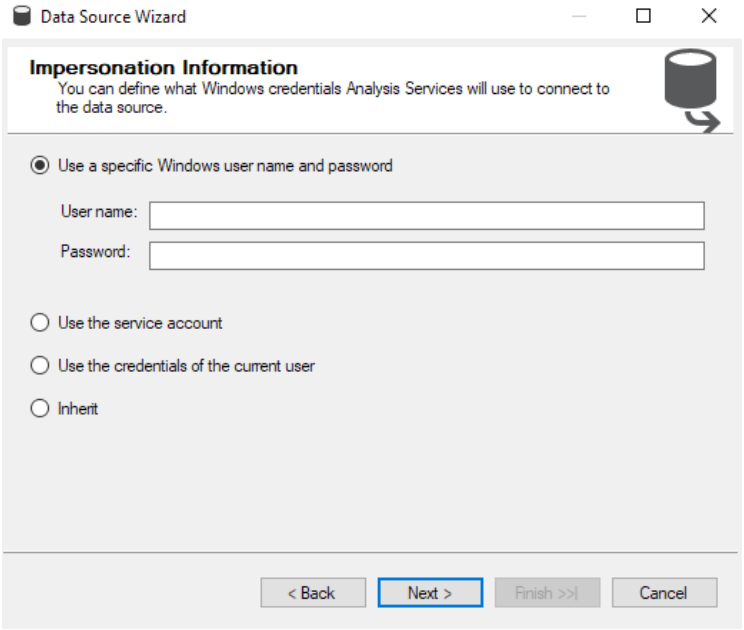
Obrázek 40 - Vytvoření projektu [Vlastní zpracování]

Vybereme připojení ke zdroji dat. Spojení již máme vytvořené z předchozích projektů, proto jej stačí vybrat a pokračovat. Pokud máme nový zdroj dat spojení musíme teprve vytvořit.



Obrázek 41 - Spojení se zdrojem [Vlastní zpracování]

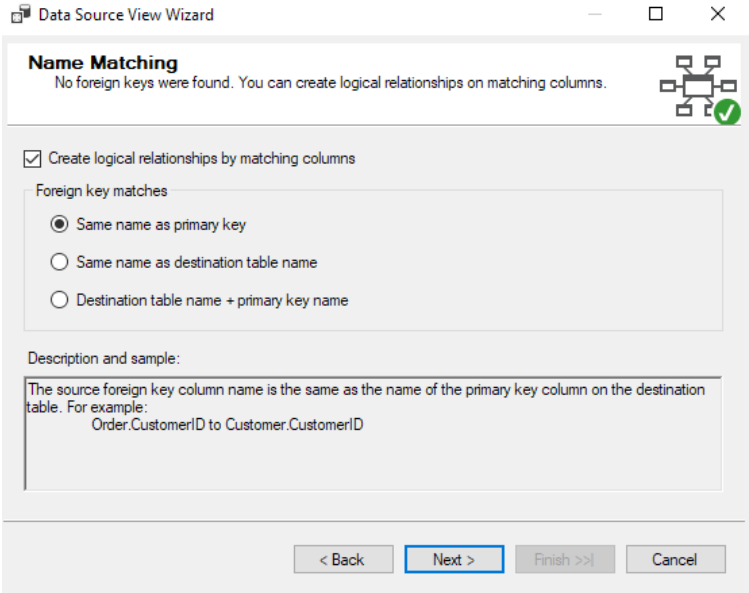
Po výběru musíme projít ověřením, které nám povolí přístup k datovému zdroji. Po úspěšném ověření jsme získali přistup k datovému zdroji.



Obrázek 42 - Ověření přístupu [Vlastní zpracování]

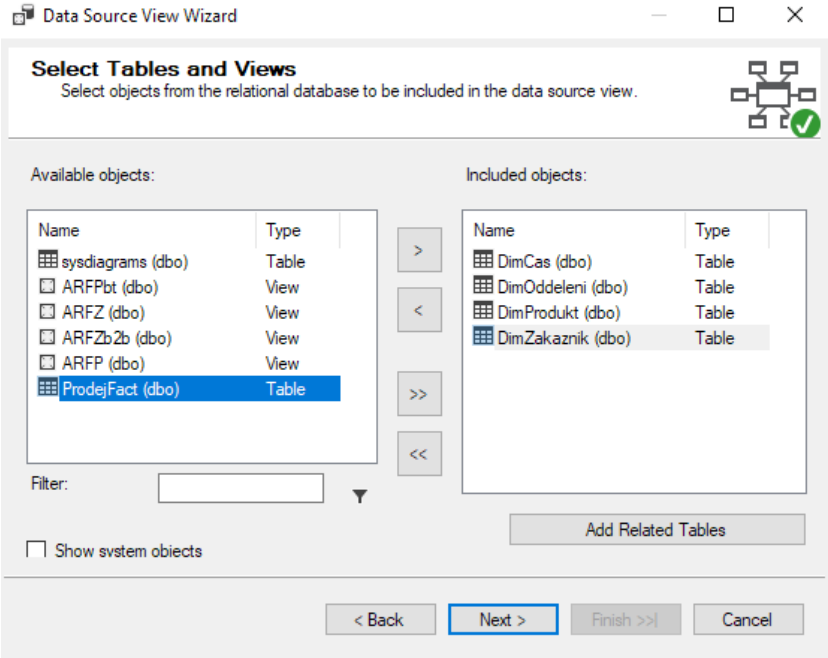
4.5 Datová kostka

Před vytvořením datové kostky musíme přidat pohled datového zdroje. Ten nám umožňuje pomoc automatickým vytvořením vazeb mezi tabulkami.



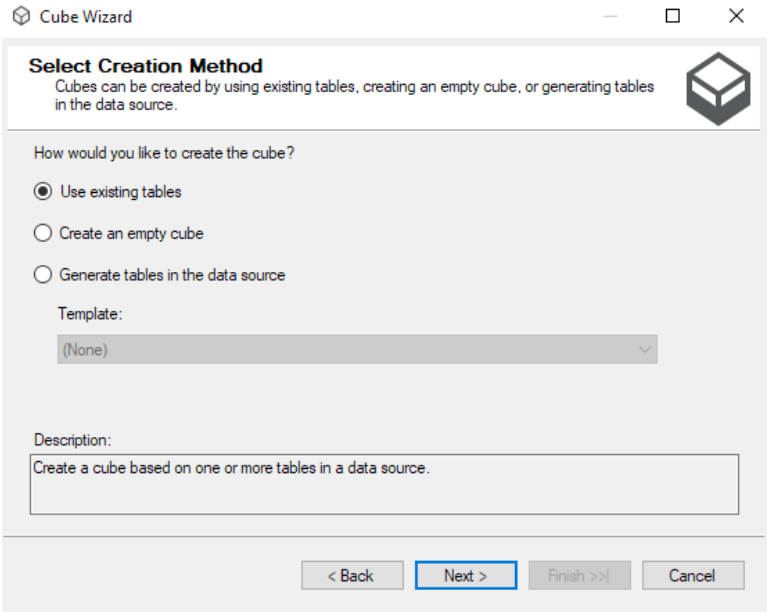
Obrázek 43 - Automatické vytvoření vazeb pohledu zdroje [Vlastní zpracování]

Pokračujeme výběrem tabulek ze zdroje. Přidáváme podle toho, které chceme využívat v datové kostce. Přesunutím z levé strany, které ukazuje všechny tabulky ze zdroje na stranu pravou, které budeme používat v datové kostce.



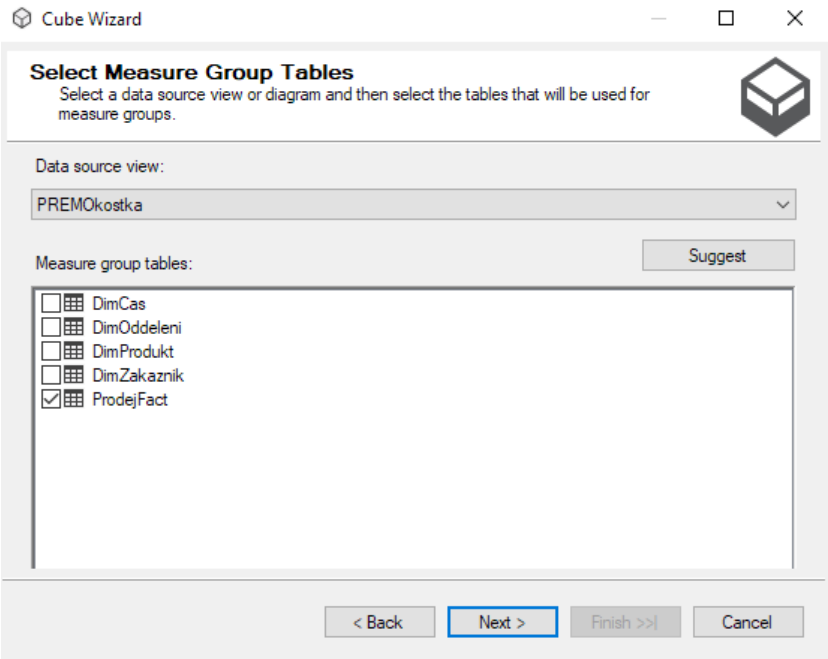
Obrázek 44 - Výběr tabulek pohledu zdroje [Vlastní zpracování]

Přicházíme k vytvoření datové kostky. Vybereme možnost „Use existing tables“, která nám n přidá existují tabulky. Pokud chceme vše vytvářet ručně zaklikneme možnost „Create an empty cube“.



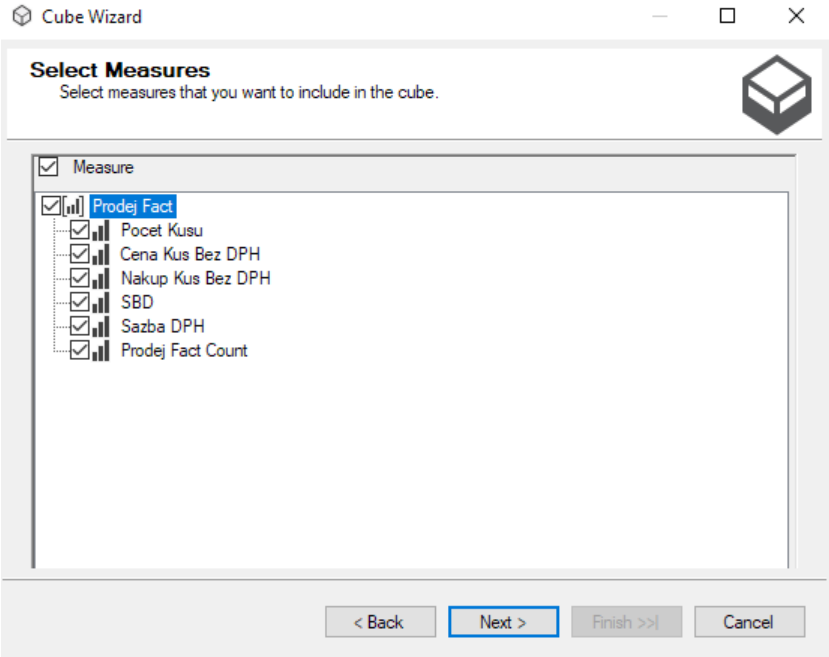
Obrázek 45 - Metoda vytvoření datové kostky [Vlastní zpracování]

Při vytváření můžeme upřesnit, ze kterého pohledu datového zdroje chceme tabulky vybrat. V našem případě se jedná o pohled PREMOkostka. Označíme tabulky, které obsahují fakta tabulky. Musí být vybrána minimálně jedna.



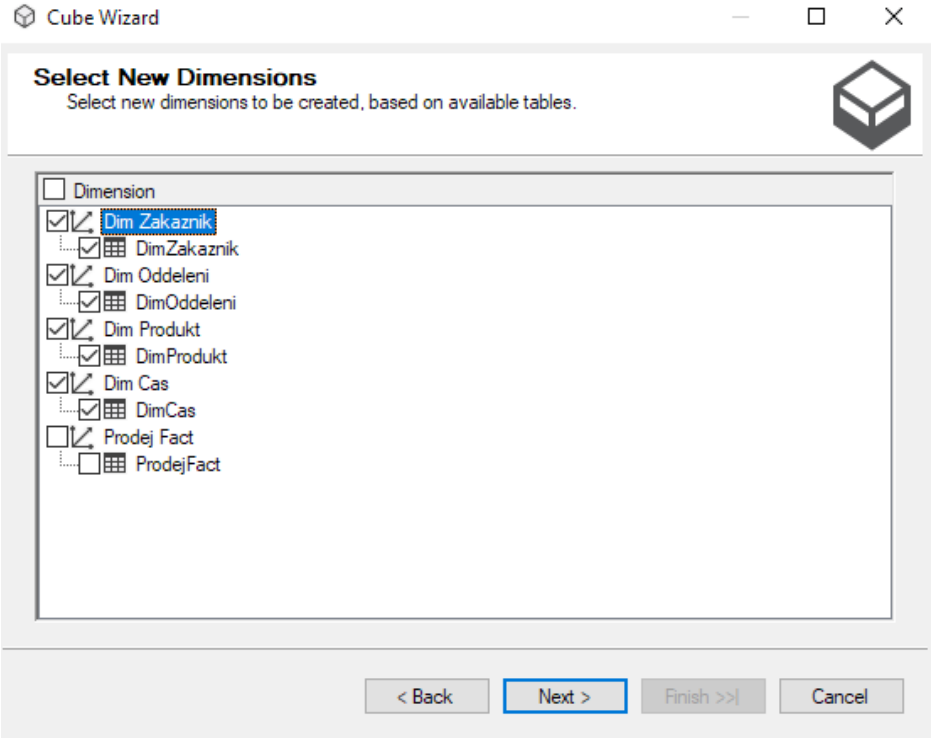
Obrázek 46 - Výběr tabulky faktu [Vlastní zpracování]

Označíme atributy, které chceme, aby se z tabulky faktu zobrazovaly v datové kostce.



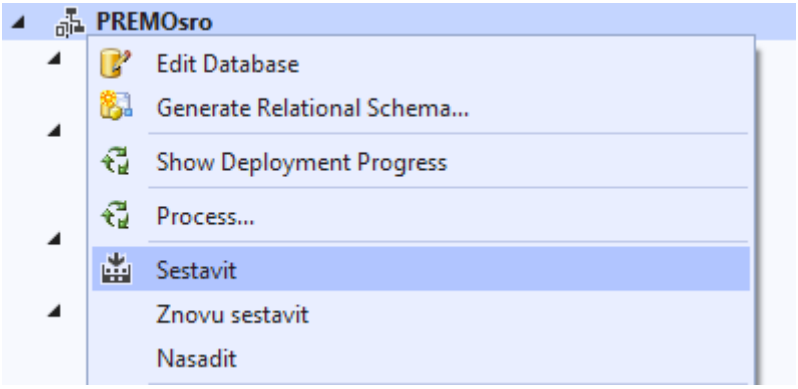
Obrázek 47 - Výběr atributů tabulky faktu [Vlastní zpracování]

Následně vybereme dimenze, které doplňují pohled k tabulce faktu a můžeme si podle nich měnit pohledy. Pohledy si měníme podle různých filtrů a detailnosti hierarchie.



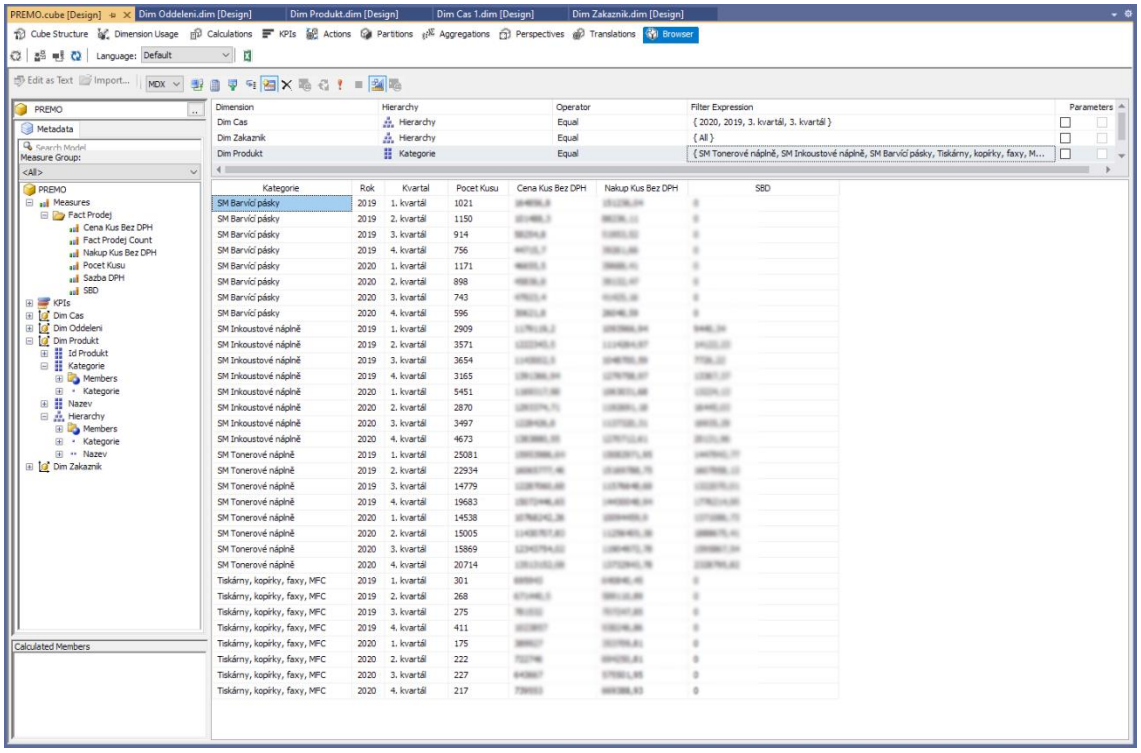
Obrázek 48 - Výběr dimenzí [Vlastní zpracování]

Takto vytvořenou datovou kostku musíme následně sestavit a nasadit. Tento proces provádíme vždy před použitím, pokud se nějak změní zdroj dat. Klikneme na projekt a nejprve provedeme „Sestavit“ a poté „Nasadit“, nebo můžeme oba úkony nechat proběhnout pomocí „Process…“



Obrázek 49 - Sestavení a nasazení [Vlastní zpracování]

Výsledkem je prázdný náhled datové kostky, do kterého můžeme libovolně vkládat atributy dimenzí. Ty nám dále dovolují filtrovat podle toho, co nás zrovna zajímá. Vybíráme od nejdetailnějších informací až po obecná shrnutí.



Obrázek 50 - Datová kostka [Vlastní zpracování]

4.6 Dolování dat

Pro dolování dat pomocí asociačních pravidel si vytvoříme tři nové pohledy dat.

Pohled ARFZb2b

```
CREATE VIEW ARFZb2b AS
SELECT DISTINCT
    Faktura
    ,f.IdZakaznik
FROM ProdejFact f
join DimZakaznik z ON z.IdZakaznik = f.IdZakaznik
WHERE z.Skupina = 'B2B'
```

Primárním segmentem společnosti je segment B2B, proto budeme provádět dolování dat podle zákazníků z tohoto segmentu. Segment B2G by nám zbytečně zkresloval výstup asociačních pravidel z důvodu velkého počtu položek na jedné faktuře, které jsou v tomto množství v segmentu B2B spíše výjimečné.

Pohled obsahuje číslo faktury a Id zákazníka.

Pohled ARFP

```
CREATE VIEW ARFP AS
WITH CTE(Faktura,
    IdProdukt,
    RadekFaktury)
AS (SELECT f.Faktura,
              f.IdProdukt,
           ROW_NUMBER() OVER(PARTITION BY Faktura
           ORDER BY IdProdukt) AS RadekFaktury
FROM ProdejFact f)
SELECT f.Faktura, f.IdProdukt, CTE.RadekFaktury, p.Nazev, p.Kategorie
FROM CTE
join DimProdukt p ON p.IdProdukt = CTE.IdProdukt
join ProdejFact f ON f.IdProdukt = p.IdProdukt
```

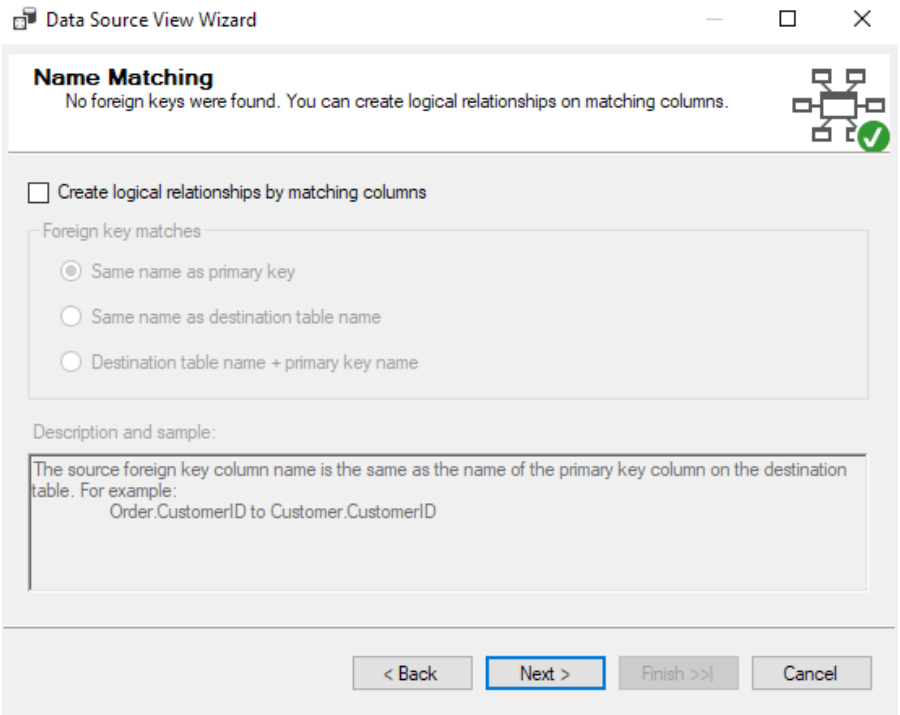
Druhý pohled v zásadě zjednodušuje tabulku faktu a propojuje ji s tabulkou v níž jsou informace o produktu. Obsahuje číslo faktury, Id zákazníka, přidává pořadové číslo produktu na faktuře, název produktu a kategorii produktu.

Pohled ARFPbt

```
CREATE VIEW ARFPbt AS
WITH CTE(Faktura,
    IdProdukt,
    RadekFaktury)
    AS (SELECT f.Faktura,
                  f.IdProdukt,
           ROW_NUMBER() OVER(PARTITION BY Faktura
           ORDER BY IdProdukt) AS RadekFaktury
 FROM ProdejFact f)
SELECT f.Faktura, f.IdProdukt, CTE.RadekFaktury, p.Nazev, p.Kategorie
FROM CTE
join DimProdukt p ON p.IdProdukt = CTE.IdProdukt
join ProdejFact f ON f.IdProdukt = p.IdProdukt
join DimZakaznik z ON z.IdZakaznik = f.IdZakaznik
 WHERE p.Kategorie != 'SM Tisková média'
    AND p.Kategorie !='SM Tonerové náplně'
    AND p.Kategorie !='SM Barvící pásky'
    AND p.Kategorie !='SM Inkoustové náplně'
    AND p.Kategorie !='Reklamní a provozní materiály'
    AND p.Kategorie !='Prázdné kazety'
    AND z.Skupina = 'B2B'
```

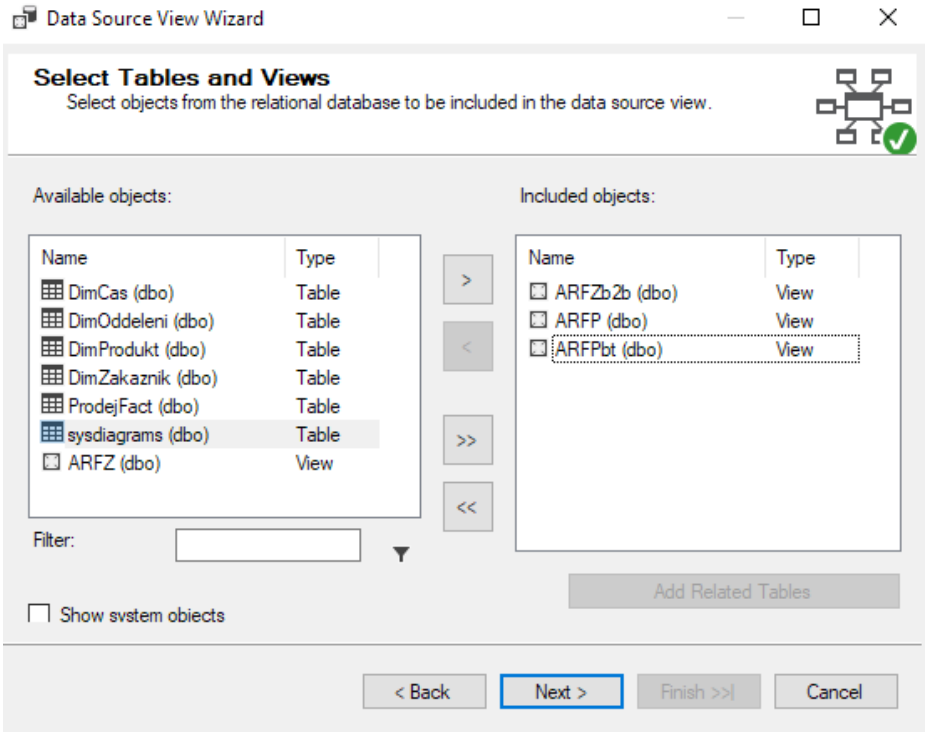
Poslední pohled je v zásadě stejný jako druhý. Pouze je očištěn o produktové kategorie, které dosahují svým zastoupením velkých rozdílů oproti zbylým kategoriím.

Po přidání těchto pohledů do datového skladu musíme přidat nový datový zdroj i do projektu. V tomto případě nevybereme automatické vytvoření vztahů mezi tabulkami.



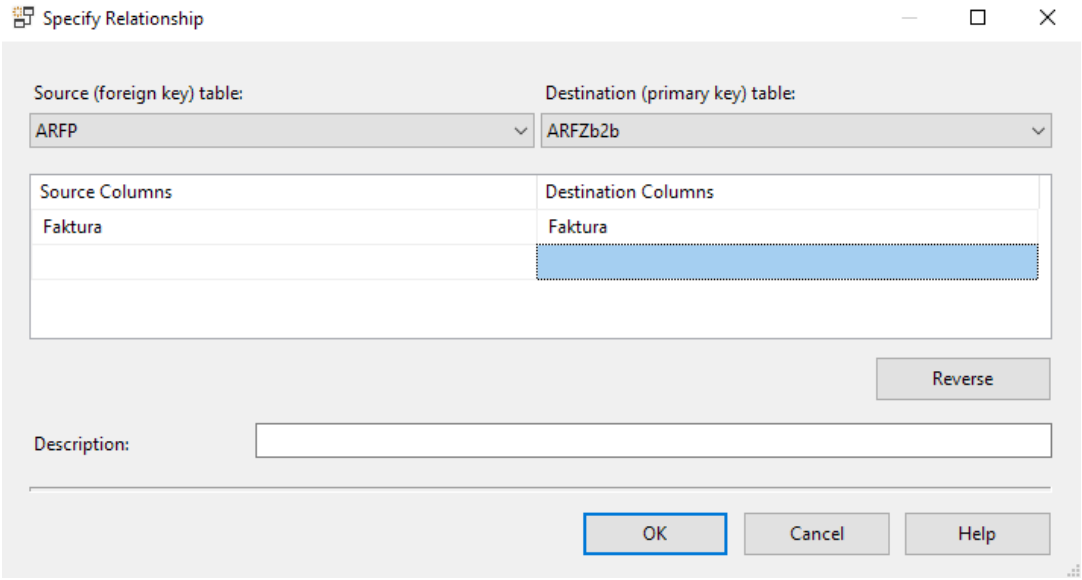
Obrázek 51 - Přidání zdroje bez vazeb [Vlastní zpracování]

Stejným způsobem, jako při přidávání zdroje u datové kostky vybereme tři pohledy, které jsem vytvořil speciálně pro využití na dolování dat pomocí asociačních pravidel.



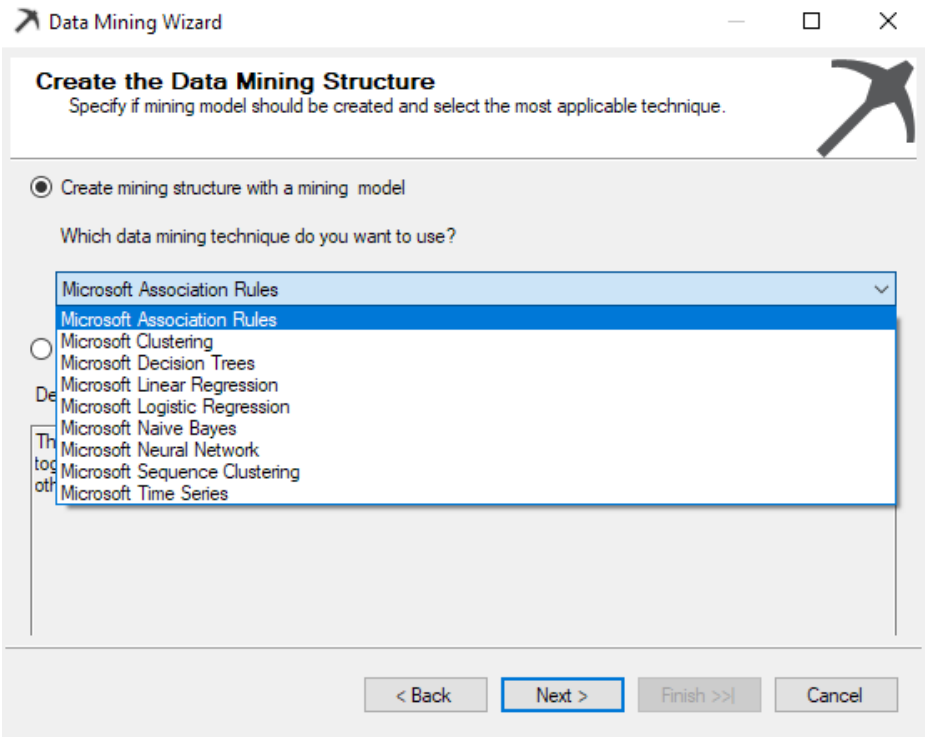
Obrázek 52 - Výběr pohledů ze zdroje [Vlastní zpracování]

Ručně vytvoříme vazbu mezi pohledem ARFP a ARFZb2b a stejná vazba bude i mezi pohledem ARFPbt a ARFZb2b.



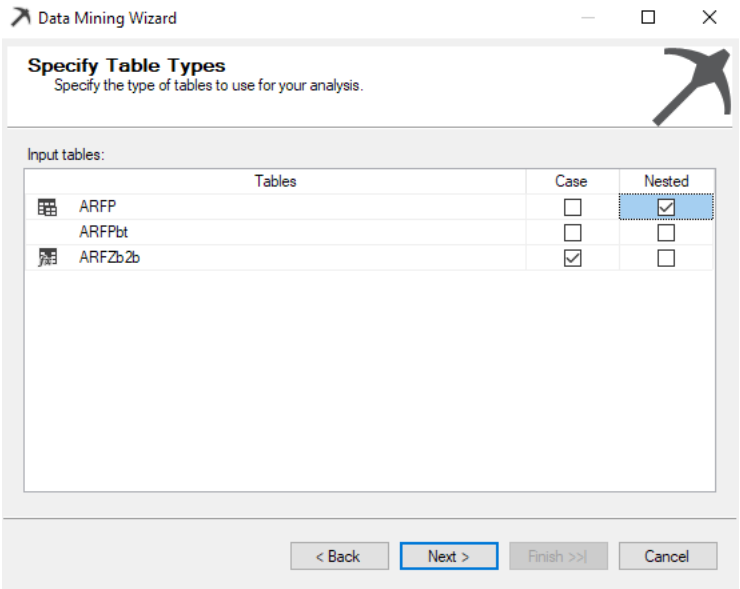
Obrázek 53 - Vytvoření vazby mezi pohledy [Vlastní zpracování]

Vybereme metodu dolování dat. Pro analýzu nákupního koše je nejvhodnější dolování dat metodou asociačních pravidel.



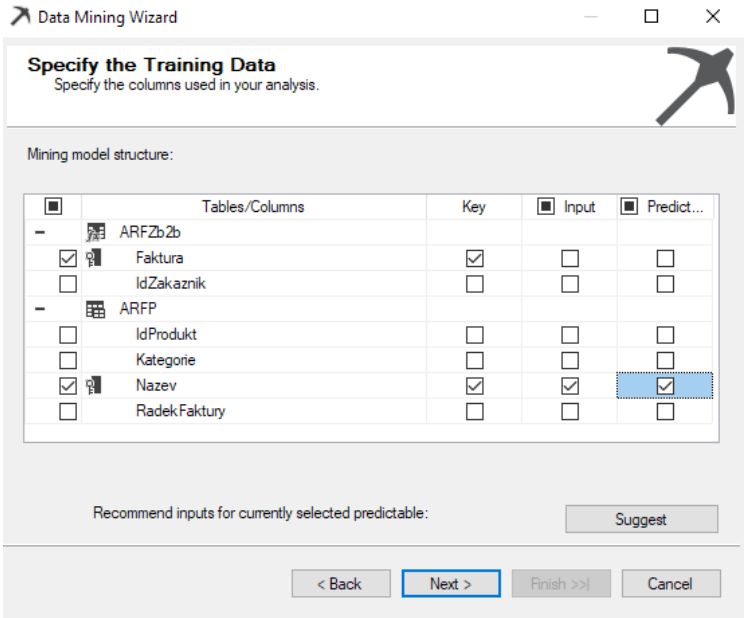
Obrázek 54 - Výběr metody dolování dat [Vlastní zpracování]

Z vytvořeného datového zdroje vybereme pohledy. Těmto pohledům přiřadíme typ, podle obsažených dat. Tabulka **case** obsahuje informace o faktuře. V našem případě Id zákazníka a číslo faktury. Tabulka **nested** obsahuje detailní informace o produktech, které jsou v konkrétní faktuře obsaženy. Tedy název a kategorii produktu. Spojení mezi tabulkami je zprostředkováno na základě čísla faktury.



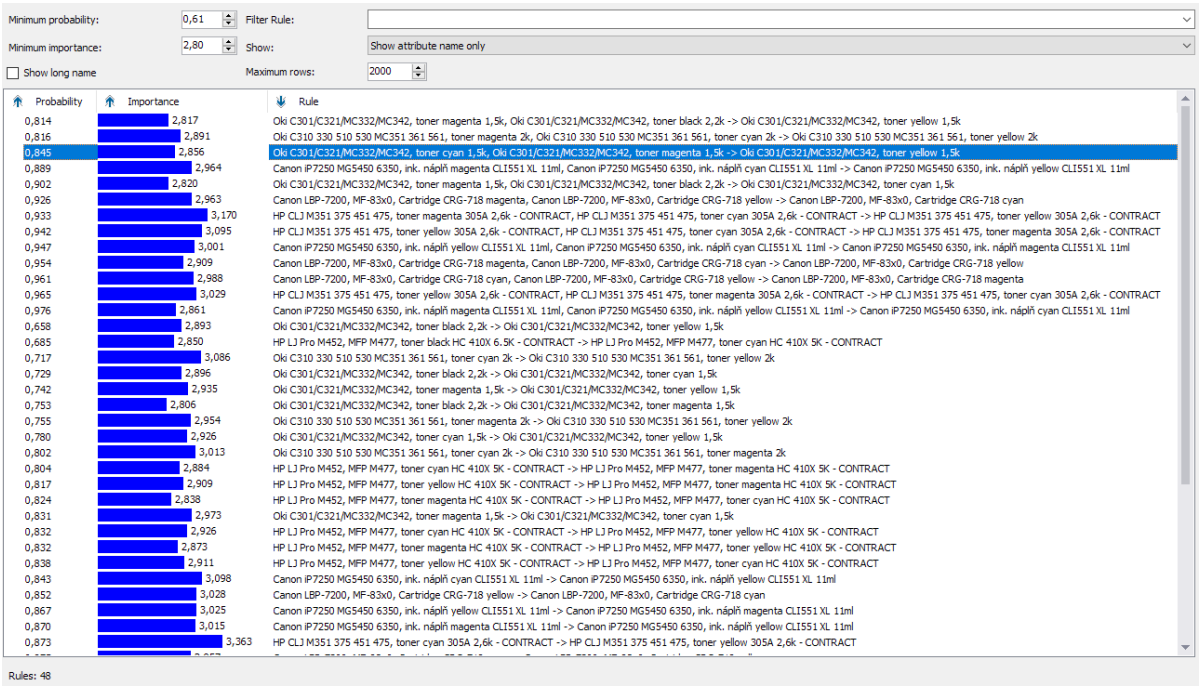
Obrázek 55 - Výběr case a nested [Vlastní zpracování]

V následujícím okně vybereme klíč, vstupní hodnotu a hodnotu, podle která mají být vyhodnoceny výstupy algoritmu.



Obrázek 56 - Key, input a predictable [Vlastní zpracování]

Na Obrázku 57 je výstup asociačních pravidel podle produktu. Na třetím vyznačeném řádku můžeme vidět pravidlo, jehož význam zní - pokud je ve faktuře obsažen produkt „Oki C301/C321/MC332/MC342, toner **cyan** 1,5k“ a produkt „Oki C301/C321/MC332/MC342, toner **magenta** 1,5k“ je pravděpodobně obsažen i produkt „Oki C301/C321/MC332/MC342, toner **yellow** 1,5k“. Sloupec „Probability“ nám znázorňuje spolehlivost s jakou je obsažen produkt „Oki C301/C321/MC332/MC342, toner **yellow** 1,5k“. V tomto případě je hodnota 0,845. „Importance“ udává užitečnost tohoto pravidla. Čím vyšší je hodnota, tím vyšší je i váha tohoto pravidla. Zde se hodnota rovná 2,856. Můžeme říct, že se jedná o silné pravidlo, které se ve fakturách vyskytuje podle produktu.



Obrázek 57 - Výstup asociačního pravidla produktu [Vlastní zpracování]

Výsledkem dolování dat pomocí asociačních pravidel podle produktu je 48 pravidel, které mají hodnotu „probability“ minimálně 0,61 a hodnota „Importance“ je zároveň vyšší než 2,8.

Na Obrázku 58 je výstup asociačních pravidel podle kategorie produktu. Na prvním vyznačeném řádku můžeme vidět pravidlo, jehož význam zní - pokud je ve faktuře obsažena kategorie „**Brašny, pouzdra**“ a zároveň je obsažena i kategorie „**Software**“ pravděpodobně je obsažena i kategorie „**Notebooky, netbooky**“. Sloupec „Probability“ nám znázorňuje spolehlivost s jakou je obsažena kategorie „Notebooky, netbooky“. V tomto případě je hodnota 0,894. „Importance“ udává užitečnost tohoto pravidla. Čím vyšší je hodnota, tím vyšší je i váha tohoto pravidla. Můžeme říct, že se jedná o nejsilnější pravidlo, které se ve fakturách vyskytuje podle kategorií.

Rules | Itemsets | Dependency Network

Minimum probability: 0,40 Filter Rule:

Minimum importance: 1,02 Show: Show attribute name and value

Show long name Maximum rows: 2000

Probability	Importance	Rule
0,894	1,358	Brašny, pouzdra = Existing, Software = Existing -> Notebooky, netbooky = Existing
0,808	1,195	Baterie, nabíječky baterií = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,798	1,194	Síťové prvky = Existing, Telefony = Existing -> Kabely, strukturovaná kabeláž = Existing
0,795	1,183	Baterie, nabíječky baterií = Existing, Monitory a projekční zařízení = Existing -> Kabely, strukturovaná kabeláž = Existing
0,791	1,193	Spotřební elektronika = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,782	1,095	Spotřební elektronika = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Příslušenství k IT = Existing
0,781	1,090	Brašny, pouzdra = Existing, Síťové prvky = Existing -> Příslušenství k IT = Existing
0,774	1,176	Servery, storage, racky = Existing, Příslušenství k IT = Existing -> Kabely, strukturovaná kabeláž = Existing
0,762	1,169	Spotřební elektronika = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Kabely, strukturovaná kabeláž = Existing
0,750	1,277	Software = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Notebooky, netbooky = Existing
0,748	1,161	Baterie, nabíječky baterií = Existing, Telefony = Existing -> Kabely, strukturovaná kabeláž = Existing
0,745	1,080	Spotřební elektronika = Existing, Síťové prvky = Existing -> Příslušenství k IT = Existing
0,744	1,155	Telefony = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Kabely, strukturovaná kabeláž = Existing
0,743	1,069	Software = Existing, Síťové prvky = Existing -> Příslušenství k IT = Existing
0,743	1,158	Baterie, nabíječky baterií = Existing, Notebooky, netbooky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,741	1,162	Síťové prvky = Existing, Monitory a projekční zařízení = Existing -> Kabely, strukturovaná kabeláž = Existing
0,740	1,151	Brašny, pouzdra = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,735	1,067	Brašny, pouzdra = Existing, Monitory a projekční zařízení = Existing -> Příslušenství k IT = Existing
0,731	1,065	Baterie, nabíječky baterií = Existing, Síťové prvky = Existing -> Příslušenství k IT = Existing
0,726	1,151	Servery, storage, racky = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,720	1,056	Telefony = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Příslušenství k IT = Existing
0,718	1,055	Brašny, pouzdra = Existing, Software = Existing -> Příslušenství k IT = Existing
0,716	1,138	Software = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,714	1,051	Brašny, pouzdra = Existing, Spotřební elektronika = Existing -> Příslušenství k IT = Existing
0,713	1,145	Spotřební elektronika = Existing, Notebooky, netbooky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,712	1,050	Baterie, nabíječky baterií = Existing, Monitory a projekční zařízení = Existing -> Příslušenství k IT = Existing
0,710	1,142	Síťové prvky = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Kabely, strukturovaná kabeláž = Existing
0,708	1,048	Software = Existing, Tiskárny, kopírky, faxy, MFC = Existing -> Příslušenství k IT = Existing
0,703	1,249	Software = Existing, Síťové prvky = Existing -> Notebooky, netbooky = Existing
0,698	1,048	Síťové prvky = Existing, Telefony = Existing -> Příslušenství k IT = Existing
0,692	1,135	Baterie, nabíječky baterií = Existing, PC komponenty = Existing -> Kabely, strukturovaná kabeláž = Existing
0,689	1,127	Brašny, pouzdra = Existing, PC komponenty = Existing -> Kabely, strukturovaná kabeláž = Existing
0,682	1,118	Záložní zdroje, UPS = Existing, Síťové prvky = Existing -> Kabely, strukturovaná kabeláž = Existing
0,680	1,150	Síťové prvky = Existing, Příslušenství k IT = Existing -> Kabely, strukturovaná kabeláž = Existing

Rules: 105

Obrázek 58 - Výstup asociačního pravidla kategorie [Vlastní zpracování]

Výsledkem dolování dat pomocí asociačních pravidel podle kategorie produktu je 105 pravidel, které mají hodnotu „Probability“ minimálně 0,40 a hodnota „Importance“ je zároveň vyšší než 1,02.

5 Závěr

Cílem bakalářské práce bylo provedení analýzy nákupního koše společnosti PREMO s.r.o. v letech 2015 – 2020 pomocí metody dolování dat. Výstup z této analýzy může přispět společnosti ke zvýšení prodeje tím, že bude doporučovat na e-shopu další produkty zákazníkovi na základě produktů, které již sám zákazník vložil do košíku a vztahů, které mezi sebou tyto produkty mají. Vazby těchto produktů a kategorií, do kterých patří, jsou navázány na základě historický prodejů ve sledovaném období.

Bakalářská práce je rozdělena do čtyř částí. První část se zabývá formulací cíle a shrnutí nevyhovujícího stavu. Ve druhé části jsou uvedeny teoretické informace, které shrnují oblast Business Intelligence, zdroje a přípravu dat, datové sklady, multidimenzionální datové modely a dolování dat. Třetí částí je analýza současného stavu, ve které uvádím základní informace o společnosti PREMO s.r.o. Představuji hlavní předmět podnikání, cílovou skupinu, hardware, software a data v databázi. Poslední částí je návrh vlastního řešení. Zde popisuji, jak jsem postupoval v přípravě a dolování dat.

Dle mého názoru bylo dosaženo stanoveného cíle bakalářské práce. Pomocí metody dolování dat jsem analyzoval nákupní koš zákazníka.

Seznam použité literatury

[1] LABERGE, Robert a Jakub GONER. *Datové sklady: agilní metody a business intelligence.* Brno: Computer Press, 2012, 350 s. : il., grafy, tab. ISBN 978-80-251-3729-1.

[2] HAN, Jiawei, Micheline KAMBER a Jian PEI. *Data mining: concepts and techniques.* 3rd ed. Waltham: Morgan Kaufmann, c2012. Morgan Kaufmann series in data management systems. ISBN 9780123814791.

[3] LACKO, Ľuboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby.* Brno: Computer Press, 2009, 456 s. : il. ISBN 978-80-251-2887-9.

[4] NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business intelligence: jak využít bohatství ve vašich datech.* Praha: Grada, 2005, 254 s. ISBN 80-247-1094-3.

[5] PYLE, Dorian. *Data preparation for data mining.* San Francisco: Morgan Kaufmann, 1999, xix, 540 s. + 1 CD-ROM. ISBN 1-55860-529-0.

[6] DODGE, Mark a Craig STINSON. *Mistrovství v Microsoft Office Excel 2007.* Brno: Computer Press, 2008. Mistrovství. ISBN 978-80-251-1980-8.

[7] *Microsoft SQL Server 7.0.* Brno: UNIS Publishing, c2000. Kompendium. ISBN 80-86097-41-2.

[8] HOTEK, Mike. *Microsoft SQL Server 2008: krok za krokem.* Brno: Computer Press, 2009. Krok za krokem (Computer Press). ISBN 978-80-251-2466-6.

[9] STANEK, William R. *Mistrovství v Microsoft Windows Server 2008: [kompletní informační zdroj pro profesionály].* Brno: Computer Press, 2009. ISBN 978-80-251-2158-0.

[10] HUMPHRIES, Mark. *Data warehousing: návrh a implementace.* Praha: Computer Press, 2002. Databáze. ISBN 80-7226-560-1.

Internetové zdroje

[11] *Credit Check, s.r.o.* [online]. CreditCheck: ©2021 [cit. 6.5.2021]. Dostupné z: https://www.creditcheck.cz/

[12] *Cybersoft, s.r.o.* [online]. Cybersoft: ©2021 [cit. 6.5.2021]. Dostupné z: https://www.cybersoft.cz/

[13] *Český statistický úřad* [online]. ČSÚ: ©2021 [cit. 6.5.2021]. Dostupné z: https://www.czso.cz/csu/res/registr_ekonomickych_subjektu

Seznam Obrázků

Obrázek 1 - Proces čištění 18

Obrázek 2 - Metoda binning 20

Obrázek 3 - Metoda outlier 21

Obrázek 4 – Diskretizace 23

Obrázek 5 – Generalizace 24

Obrázek 6 - Cyklus procesů přírůstkové metody 27

Obrázek 7 - Přírůstkové metody 28

Obrázek 8 - Tříúrovňová architektura 29

Obrázek 9 - Datová kostka 30

Obrázek 10 - Hvězdicové schéma 31

Obrázek 11 - Schéma vločky 32

Obrázek 12 - Schéma souhvězdí 32

Obrázek 13 - Rozhodovací strom 34

Obrázek 14 - Shlukování 35

Obrázek 15 - Asociační pravidla 35

Obrázek 16 - Časové řady 36

Obrázek 17 - Neuronové sítě 37

Obrázek 18 - Struktura jednatelů 39

Obrázek 19 - Logo společnosti 39

Obrázek 20 - Struktura společnosti 40

Obrázek 21 - Semafor rizika 46

Obrázek 22 - Zakládací formulář 47

Obrázek 23 - Prověření kredibility 47

Obrázek 24 - Vložení produktu 48

Obrázek 25 - Přidání produktu do košíku 49

Obrázek 26 - Diagram 50

Obrázek 27 - Ručně vytvořená sestava 54

Obrázek 28 - Sestava prodeje dle kategorií 54

Obrázek 29 - Výběr sestav a filtrů 55

Obrázek 30 - Report Gfk 56

Obrázek 31 - Report pro Intrastat 56

Obrázek 32 - Dorovnání cen 57

Obrázek 33 - Přidání analytických služeb 59

Obrázek 34 – Výběr režimu a práv uživatelů 60

Obrázek 35 - Rozšíření Visual Studia 60

Obrázek 36 - Schéma datového skladu 61

Obrázek 37 – Menu back up 62

Obrázek 38 - Obnovení dat ze zálohy 62

Obrázek 39 - Import dat 63

Obrázek 40 - Vytvoření projektu 69

Obrázek 41 - Spojení se zdrojem 69

Obrázek 42 - Ověření přístupu 70

Obrázek 43 - Automatické vytvoření vazeb pohledu zdroje 70

Obrázek 44 - Výběr tabulek pohledu zdroje 71

Obrázek 45 - Metoda vytvoření datové kostky 71

Obrázek 46 - Výběr tabulky faktu 72

Obrázek 47 - Výběr atributů tabulky faktu 72

Obrázek 48 - Výběr dimenzí 73

Obrázek 49 - Sestavení a nasazení 73

Obrázek 50 - Datová kostka 74

Obrázek 51 - Přidání zdroje bez vazeb 76

Obrázek 52 - Výběr pohledů ze zdroje 76

Obrázek 53 - Vytvoření vazby mezi pohledy 77

Obrázek 54 - Výběr metody dolování dat 77

Obrázek 55 - Výběr case a nested 78

Obrázek 56 - Key, input a predictable 78

Obrázek 57 - Výstup asociačního pravidla produktu 79

Obrázek 58 - Výstup asociačního pravidla kategorie 80