



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INFORMAČNÍCH SYSTÉMŮ

DEPARTMENT OF INFORMATION SYSTEMS

IMPLEMENTACE A NÁVRH DATOVÉ KOSTKY OLAP

DESIGN AND IMPLEMENTATION OF AN OLAP DATA CUBE

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

PAVEL HURDÁLEK

VEDOUcí PRÁCE

SUPERVISOR

prof. Ing. TOMÁŠ HRUŠKA, CSc.

BRNO 2023

Zadání bakalářské práce



143931

Ústav: Ústav informačních systémů (UIFS)
Student: **Hurdálek Pavel**
Program: Informační technologie
Specializace: Informační technologie
Název: **Implementace a návrh datové kostky OLAP**
Kategorie: Informační systémy
Akademický rok: 2022/23

Zadání:

1. Prostudujte teoretické základy teorie datových kostek. Seznamte se systémy SAP BW/Analysis.
2. Navrhněte zdroj dat, postup ETL a datovou kostku pro vybranou oblast dat.
3. Realizujte návrh struktur zdrojových dat, proces ETL a strukturu datové kostky k bodu 2.
4. Provedte ověření funkčnosti datové kostky pro zvolené analytické výstupy.

Literatura:

- přednášky Pokročilé informační systémy z magisterského studia (Hruška)
- webinář na téma Datové kostky (Hruška)
- Manuály SAP pro SAP BW a SAP Analysis

Při obhajobě semestrální části projektu je požadováno:

OLAP, SAP BW

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Hruška Tomáš, prof. Ing., CSc.**
Vedoucí ústavu: Kolář Dušan, doc. Dr. Ing.
Datum zadání: 1.11.2022
Termín pro odevzdání: 10.5.2023
Datum schválení: 21.10.2022

Abstrakt

Tato diplomová práce řeší tvorbu datové OLAP kostky, za účelem analýzy rozsáhlých dat. Tento proces je zobrazen při vytváření kostky obsahující údaje z přijímacího řízení univerzity Vysokého učení technického v Brně. Data jsou zpracovávána komerčním nástrojem Business Warehouse od německé společnosti SAP, samotná vizualizace kostky probíhá v programu Excel pomocí rozšíření Analysis. Výsledkem této práce je OLAP kostka, pomocí které lze vytvářet analýzy přijímacího řízení univerzity a získávat data pro vyplnění příslušné části výroční zprávy o činnosti.

Abstract

This diploma thesis deals with the creation of a data OLAP cube for the purpose of analyzing big data. The process is shown when creating a cube containing data from the university's admissions process (specifically Brno University of Technology). The data is processed by the commercial tool Business Warehouse from the German company SAP, the cube visualization itself takes place in the program Excel using the Analysis extension. The result of this work is an OLAP cube, which can be used to create analyses of the university's admission procedure and obtain data to fill in the relevant section of the annual activity report.

Klíčová slova

OLAP, datová kostka, multidimenzionální model databáze, datový sklad, datový trh, analýza dat, BI, Business intelligence, SAP, SAP BW, Business Warehouse, BEx, ETL, PSA, InfoPaket, InfoObjekt, ODS, InfoKostka, query, výroční zpráva o činnosti, přijímací řízení

Keywords

OLAP, data cube, multidimensional database model, data warehouse, data market, data analysis, BI, Business intelligence, SAP, SAP BW, Business Warehouse, BEx, ETL, PSA, InfoPacket, InfoObject, ODS, InfoCube, query, annual activity report, admissions process

Citace

HURDÁLEK, Pavel. *Implementace a návrh datové kostky OLAP*. Brno, 2023. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce prof. Ing. Tomáš Hruška, CSc.

Implementace a návrh datové kostky OLAP

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana prof. Ing. Tomáše Hrušky, CSc.

Další informace mi poskytl doc. Ing. Radek Kubásek Ph.D. (analytik VUT).

Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....
Pavel Hurdálek

25. dubna 2023

Poděkování

Mé díky patří panu prof. Ing. Tomáši Hruškovi, CSc., za jeho odborné vedení této práce. Dále nemohu opomenout pana doc. Ing. Radka Kubásku Ph.D., za jeho ochotu a rady z pohledu analytika.

Obsah

1	Úvod	6
1.1	Cíl práce	6
2	Úvod do datových skladů a OLAP technologií	7
2.1	Systémy OLTP	7
2.2	ETL (datová pumpa)	7
2.3	Datové sklady	9
2.3.1	Tvorba datového skladu	10
2.4	Multidimenzionální model databáze	11
2.4.1	Datová kostka	12
2.4.2	Implementace datové kostky	13
2.5	Analýza OLAP	13
2.5.1	Pravidla OLAP technologií	14
2.5.2	Druhy architektur serveru OLAP	14
2.5.3	Základní operace v OLAP analýze	15
2.6	Business Intelligence (BI)	16
2.7	Tok dat při použití OLAP	17
3	SAP	18
3.1	Komponenty, moduly a transakce	18
3.2	Technologie SAP NetWeaver	19
3.3	Komponenta SAP ERP	20
3.4	Modul SAP BW	20
3.4.1	PSA	21
3.4.2	InfoProvider	21
3.4.3	InfoObjekty	21
3.4.4	InfoKostka	23
3.4.5	Objekt datová schránka (ODS)	23
3.4.6	Rozdíl InfoKostky a ODS	23
3.4.7	Transformace a Proces přenosu dat	24
3.4.8	Tvorba datové kostky v SAP BW	24
3.5	Jazyk ABAP	25
3.6	Možnosti vizualizace datových objektů	25
3.6.1	Business Explorer (BEx)	26
3.6.2	SAP BEx Query Designer	26
3.6.3	SAP BEx Analyzer	26
3.6.4	SAP Analysis for Office	26

4	Zvolená oblast dat pro datovou kostku	27
4.1	Přijímací řízení univerzity	27
4.1.1	Přijímací řízení VUT	28
4.2	Výroční zpráva	29
4.2.1	Výroční zprávy vysokých škol	29
5	Návrh datové kostky	30
5.1	Zdroj dat	30
5.2	Volba použitých dat	31
5.2.1	Vybrané údaje z tabulek	32
5.2.2	Vybrané číselníky	32
5.3	Identifikace dimenzí a faktů	33
5.4	Volba vhodné reprezentace datové kostky	33
5.5	Složitější charakteristiky	33
5.5.1	Hierarchie PSČ	34
6	Implementace datové kostky	36
6.1	Databázový pohled	36
6.2	PSA	36
6.2.1	InfoPaket	38
6.3	InfoObjekty	38
6.3.1	Kmenové texty charakteristik	39
6.3.2	Hierarchie charakteristik	39
6.3.3	Atributy charakteristik	40
6.4	InfoKostka Přijímací řízení	40
6.4.1	Transformace	41
6.4.2	Proces přenosu dat	41
6.5	Hierarchie PSČ	43
6.5.1	ODS Část obce PSČ	43
6.5.2	ODS Hierarchie zdroj dat	43
6.5.3	Problémy identifikátorů uzlů hierarchie	44
6.5.4	Transformace	45
6.5.5	Výsledná hierarchie	46
6.6	Query	46
6.7	Transport na produkci	47
6.8	Řetězec procesů	48
7	Testování datové kostky	49
7.1	Způsob testování	49
7.2	Největší nalezené problémy	49
7.2.1	Číselník Návrh příjmu	50
7.2.2	PSČ začínající nulou	52
7.2.3	Zahraniční PSČ	52
7.2.4	Obor střední školy	53
8	Možné výstupy z datové kostky	54
8.1	Typ střední školy uchazečů	54
8.2	Způsob přijetí uchazečů	55
8.3	Zájem o studium	55

8.3.1	Zájem žen o studium na technické škole	56
8.4	Geografické analýzy	56
9	Závěr	57
	Literatura	58
A	Obsah přiloženého paměťového média	60
B	Seznam použitých polí z databáze VUT	61
C	Geografické grafy	63

Seznam obrázků

2.1	Znázornění průběhu ETL.	9
2.2	Přístupy k tvorbě datových skladů.	11
2.3	Rozdíl geometrické a datové kostky.	12
2.4	Porovnání schémat hvězdy a sněhové vločky.	13
2.5	Operace zavrtávání a vnořování.	15
2.6	Operace otočení.	15
2.7	Operace filtrování.	16
2.8	Operace řez.	16
2.9	Datový tok.	17
3.1	Komponenty, moduly a transakce systémů SAP.	19
3.2	Hierarchie geografických celků.	22
3.3	Tok dat v SAP BW.	24
5.1	Ukázka dokumentace jedné tabulky v Apollu.	30
5.2	Zjednodušené ERD oblasti příjímací řízení.	31
5.3	Zjednodušené ERD oblasti adres.	34
6.1	PSA – nastavení zdroje dat.	37
6.2	PSA – návrh polí podle zdroje dat.	37
6.3	Založení charakteristiky.	38
6.4	Základní nastavení charakteristiky.	38
6.5	Část nastavení kmenových textů.	39
6.6	Transformace dat do kmenových textů.	39
6.7	Nastavení atributů.	40
6.8	Transformace dat do atributů.	40
6.9	InfoKostka Příjímací řízení.	41
6.10	Chyba – neplatný znak při transformaci.	41
6.11	Transformace dat do InfoKostky Příjímací řízení.	42
6.12	Transformace do ODS Část obce PSČ.	43
6.13	ODS s více transformacemi a procesy přenosu dat.	44
6.14	Ukázka dat číselníku region.	44
6.15	Transformace struktury hierarchie.	46
6.16	Zobrazení hierarchie PSČ v SAP BW.	46
6.17	Query s vypočteným ukazatelem počet uchazečů.	47
6.18	Řetězec procesů aktualizující datovou kostku.	48
7.1	Vyhodnocené přihlášky Fakulty informačních technologií.	50
7.2	Vyhodnocené přihlášky Fakulty informačních technologií z roku 2020.	50

7.3	Příznak Přijato celkem u všech variant výsledků přijímacího řízení.	51
7.4	Opravený příznak Přijato celkem.	51
7.5	Ukázka některých chybně zadaných PSC.	52
7.6	Chybně zadaný záznam v číselníku Obor středních škol.	53
8.1	Střední školy uchazečů o bakalářské studium na FIT VUT od roku 2012. . .	54
8.2	Procentuální vývoj uchazečů o bakalářské studium na FIT VUT přijatých na základě přijímací zkoušky.	55
8.3	Počet podaných přihlášek na Ústav soudního inženýrství (ÚSI) VUT. . . .	55
8.4	Procentuální zastoupení žen na FIT VUT.	56
8.5	Kraj trvalého bydliště českých uchazečů VUT z roku 2022.	56
C.1	Státní příslušnost uchazečů VUT, rok 2022.	63
C.2	Stát trvalého bydliště uchazečů VUT, rok 2022.	64
C.3	Město trvalého bydliště českých uchazečů VUT, rok 2022.	65
C.4	Poloha středních škol českých uchazečů VUT, rok 2022.	66

Kapitola 1

Úvod

V dnešní době informačních technologií, kdy má téměř každá organizace obrovské množství údajů, je stále důležitější využívat tyto informace a získávat z nich nové poznatky, nápomocné při budoucím rozhodování a řízení. Pro analýzu firemních dat existuje mnoho přístupů. Jedním z nich jsou nástroje OLAP. Výhodou použití OLAP technologií je možnost interaktivního procházení, výběru a analýzy dat z různých perspektiv. OLAP využívá datových kostek, které umožňují rychlé a efektivní vyhledávání a agregaci dat v multidimenzionálním prostoru. Použití této technologie umožňuje rychlejší a přesnější rozhodování či plánování, což může vést ke zlepšení výkonnosti a konkurenceschopnosti organizací.

Tato práce vzniká ve spolupráci s organizací CVIS VUT v rámci rozjetého projektu MIS, který má za úkol vytvořit manažerský informační systém s daty univerzity Vysoké učení technické v Brně. Tento systém má umožnit analytikům snadný a na nikom nezávislý přístup k datům.

Pro implementaci datových kostek existuje mnoho nástrojů. V projektu MIS je využíván komerční software od společnosti SAP, a proto je použit i v této práci. Datové sklady a OLAP kostky jsou vytvářeny v programu SAP NetWeaver BW 7.50.

Ověřování správnosti dat a hledání případných odchylek v datové kostce již provádí analytici. V rámci této práce se spolupracuje s panem doc. Ing. Radkem Kubáskem Ph.D., analytikem rektorátu VUT.

1.1 Cíl práce

Cílem této bakalářské práce je nastudovat principy multidimenzionální analýzy dat a poté navrhnout a implementovat datovou kostku OLAP pro univerzitu Vysoké učení technické v Brně. V rámci práce budou použity komerční nástroje společnosti SAP pro správu a analýzu dat, a to včetně ETL procesů, multidimenzionálního modelování, vytváření datových skladů a jejich vizualizace. Výsledkem práce bude funkční datová kostka OLAP, která bude sloužit jako nástroj pro analýzu a reporting dat na vysoké škole. Její datový obsah bude zvolen podle aktuálních požadavků a potřeb.

Kapitola 2

Úvod do datových skladů a OLAP technologií

Datové sklady a OLAP technologie spolu úzce souvisí. V této kapitole si definujeme základní pojmy z dané oblasti a podíváme se jaké mají mezi sebou spojitosti. Závěrem si principiálně ukážeme, jakou cestou musí data projít, aby se dostala k cílovému uživateli.

2.1 Systémy OLTP

Transakční databázové systémy OLTP (*Online Transaction Processing*) jsou ve firemní praxi velice oblíbené a používané. Jejich cílem je automatizace každodenních činností našeho podnikání. Příkladem těchto aktivit může být vytváření objednávek a s tím související skladové hospodářství, nákupy a prodeje. [15]

OLTP databázové systémy se tedy vyznačují zejména neustále se měnícím stavem databáze, ke které přistupuje velké množství uživatelů (může se jednat jak o zaměstnance, tak o zákazníky). Tito uživatelé mohou data ze systému číst, ale případně i do něj zapisovat. Standardně mají ale možnost využít pouze předem definované databázové dotazy. [15]

2.2 ETL (datová pumpa)

Data organizací jsou často uloženy v mnoha nehomogenních a různorodých zdrojích. Cílem ETL je centralizace těchto údajů a uložení do jednoho cílového systému, ve kterém se s daty dále pracuje. Jedná se tedy o proces, jehož úkolem je extrahovat data ze všech zdrojů, upravit do požadované formy a nakonec nahrát do konečného úložiště. [15][22]

ETL se skládá ze tří částí:

- Extrakce (anglicky *Extraction*)
- Transformace (anglicky *Transformation*)
- Zavedení (anglicky *Loading*)

Extrakce

Jedná se o první krok procesu ETL, ve kterém dochází k samotnému extrahování dat. Tato data mohou být obecně uložena v libovolných zdrojích. Lišit se mohou hardwarem, operačním systémem či databázovým systémem. Důraz se klade jak na rychlost a efektivitu, tak na co možná nejmenší ovlivnění zdrojových systémů. [15]

Transformace

Jelikož se data při extrakci načítají z různých zdrojů, tak se mohou vyskytovat v rozdílných formátech. Dalším problémem může být kvalita údajů vyskytujících se v zdrojových systémech. Při použití takových dat snižujeme jakost výsledného produktu.

Podle [15] a [22] dochází nejčastěji k těmto problémům:

- Nejednoznačnost údajů – údaje mohou být uloženy několika způsoby, například stejné pohlaví být v různých zdrojových systémech vedeno pod jinou hodnotou (F, female, woman, žena atd.). Řeší se převodem do jednotného tvaru.
- Chybějící hodnoty – u malého množství se dá tento problém ignorovat. Jiným východiskem je doplnit hodnoty z jiných zdrojů, popřípadě vyplnit implicitní hodnotou.
- Duplicitní záznamy – v některých případech je náročné duplicitu vůbec objevit.
- Různé peněžní měny
- Formáty čísel a textových řetězců – čísla způsobují největší problémy, mohou být totiž uloženy v numerických i řetězcových datových typech, tedy jako numerická hodnota nebo posloupnost číslic a dalších znaků (mezery, tečky, čárky a další).
- Referenční integrita – v získaných datech jsou skryty i různé vztahy, hierarchické struktury a další údaje. Při změnách ve zdrojových databázích může dojít k narušení jejich integrity a tím i ovlivnit jakost extrahovaných údajů.

Úlohou transformace je tedy upravit extrahované údaje do cílového formátu a zvýšit jejich kvalitu na nejlepší možnou úroveň. [15]

Zavedení

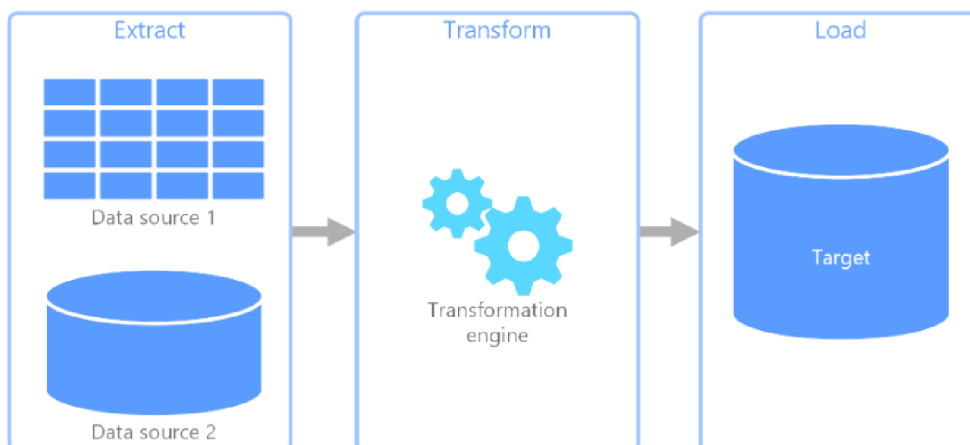
Poslední část zajišťuje přenos a uložení transformovaných dat do cílového systému. Obvykle se ETL provádí opakovaně v pravidelných časových obdobích, přičemž při prvotním průběhu může být přenášeno obrovské množství dat. Později se už zavádějí pouze nově vytvořené údaje. [15]

Chyby a testování ETL

V procesu ETL může dojít k chybám. Ty mohou být způsobeny například hardwarovou poruchou zdrojových systémů. Dle závažnosti chyby se určuje, zda se má začít znovu nebo pokračovat v procesu. [15]

Důkladné otestování je velice důležité, každopádně ani to nezaručí jistotu stoprocentní funkcionality. Stejně jako okolní svět se v praxi neustále mění i zdrojové systémy a jejich organizace dat. Z tohoto důvodu může do procesu kdykoliv přijít nečekaný vstup s následkem neobvyklého chování. [15]

Schéma procesu ETL



Zdroj: [11]

Obrázek 2.1: Znázornění průběhu ETL.

2.3 Datové sklady

V dnešní moderní době produkují organizace obrovské množství údajů. Obvykle se jedná o data z transakčních systémů (OLTP), které nesou informace o podnikových procesech. Příkladem by mohl být internetový obchod, ve kterém mají zákazníci možnost výběru zboží, umístění do košíku, volbu dopravy a vytvoření objednávky. Veškeré tyto akce produkují záznamy a ty se ukládají do operační databáze. Stav tohoto úložiště je tedy proměnný a každým okamžikem může obsahovat nové údaje. Každopádně v praxi firmy potřebují provádět na svých datech inspekce a analýzy, za účelem získání nových informací. V rámci našeho příkladu se může jednat o vývoj množství prodeje určitého zboží, ale i třeba marketingová data jako: ve kterém bodě před vytvořením objednávky zákazník nejčastěji ukončí proces nákupu (tedy kde unikají prodeje). Tyto údaje jsou zásadní pro následující rozhodnutí organizace a právě tyto důvody vedly k myšlence datových skladů. [9]

Samotný název datový sklad může vyvolávat hledání spojitostí s klasickým skladem, kde podniky přechodně uchovávají materiály, součástky či již hotové výrobky. U klasického skladu je ovšem snaha skladování co nejvíce zkrátit. Jinak řečeno použít materiální zdroje pro výrobu produktů a ty následovně prodat. Datové sklady fungují na úplně jiném principu. Zde se totiž nechceme zbavit jejich obsahu, ale právě naopak se snažíme uchovat informační bohatství za nejdelší možnou dobu. Mnohem lepší přirovnání se nabízí k depozitářům muzeí, kde se také snaží nasbírat co nejvíce historických exponátů. [15]

Definice datového skladu (anglicky *data warehouse*)

Asi nejznámější definice pochází od Billa Inmona, přezdívaného otce datových skladů.

„Datový sklad je podnikově strukturovaný depozitář subjektivě orientovaných, integrovaných, časově proměnlivých, historických dat použitých na získávání informací a podporu rozhodování. V datovém skladu jsou uložena atomická a sumární data.“ [15, strana 48]

Podle této definice jsou uchovávaná data v datovém skladu:

- subjektivě orientovaná – pohled na data je brán z hlediska subjektů (například zákazník, zaměstnanec, výrobek a podobně), nikoliv aplikací vytvářející dané údaje (fakturační aplikace, aplikace pro personalistiku a tak dále). [15]
- integrovaná – musí tedy být jednotná, to zahrnuje používání stejných formátů, konvencí, měn a jednotek. Jelikož data mohou pocházet z různých zdrojových systémů, musí projít nějakou úpravou, vyčištěním a sjednocením. [15]
- časově proměnlivá – v rámci podnikového procesu se mohou některé informace měnit, například stavy zúčastněných entit. Úkolem datového skladu je uchovávat data o těchto změnách, aby se na ně dalo nahlížet jak z aktuálního, tak historického kontextu. [9]
- historická – v některých zdrojích úváděno jako neměnná. V transakčních databázích je možné údaje nejen přidávat a číst, ale také upravovat a odstraňovat. Tyto změny by se v datovém skladu měly projevit pouze vytvořením nového záznamu, popisujícího danou událost, bez úprav již zavedených dat. V praxi je někdy ovšem potřeba opravit chybnou verzi záznamu, a proto se tento bod ne vždy plně dodržuje. [9]

Datový trh (anglicky *datamart*)

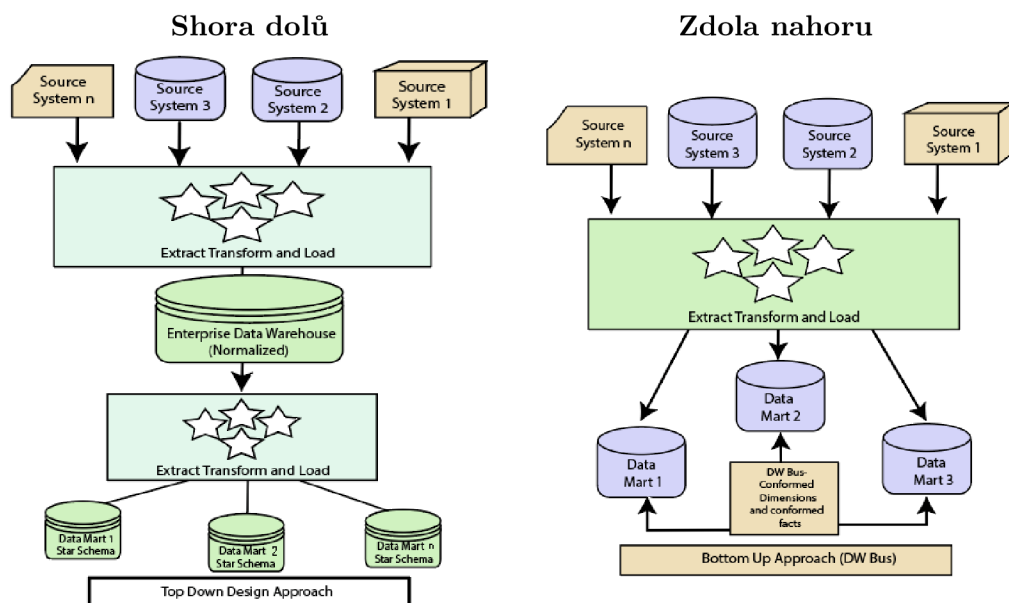
Tvorba datového skladu je velice náročný a rozsáhlý projekt, z toho důvodu se k jeho vývoji někdy přistupuje postupně (viz 2.3.1). Vytváří se tedy jeho podmnožiny s daty, která jsou relevantní pouze pro menší část organizace. Tato podmnožina datového skladu se nazývá datový trh.

Datové trhy lze samozřejmě vytvořit i z kompletního datového skladu, ze kterého se vyberou pouze významné údaje pro danou část podniku, a proto mohou existovat jak samy o sobě, tak jako součást celého systému. [12]

2.3.1 Tvorba datového skladu

Při tvorbě datového skladu se využívá následujících přístupů:

- Centralizovaný přístup (budování shora dolů) – v počáteční fázi zahrnuje velké množství plánování a navrhování. Nejdříve se vytváří konceptuální model celého datového skladu, případně i jednotlivých datových trhů. Z toho důvodu se před samotnou implementací musí nasbírat a zafixovat všechny požadavky (pozdější změny jsou již problematické). Při použití této metody obvykle získáme definice dat a pravidla užití pro celou organizaci již v prvotní fázi. Zároveň se řeší bezpečnost a konzistence dat pouze nad datovým skladem. Nevýhodou jsou vysoké vstupní náklady a nízká flexibilita vývoje. [21]
- Decentralizovaný přístup (budování zdola nahoru) – jedná se o iterativní přístup, kde se postupně vytváří jednotlivé datové trhy pro určitá oddělení společnosti. Jedná se o používanější metodu, zejména kvůli nižším vstupním investicím a rychleji viditelným výsledkům. Nevýhodou je možná redundance dat a při nevhodném či nulovém návrhu datového skladu, nemusí být možné datové trhy sjednotit do datového skladu. [21]
- Kombinovaný přístup – jedná se o kombinaci předchozích metod.



Zdroj (upraveno): [10]

Obrázek 2.2: Přístupy k tvorbě datových skladů.

2.4 Multidimenzionální model databáze

Při budování datových skladů je nutné se zaměřit na to, jaký typ databáze zvolit. Standardní relační databáze používaná pro transakční OLTP systémy, kde jsou data uložena v mnoha atomických, relačně svázaných tabulkách, zde není vhodná. Toto řešení má minimální redundanci dat a je tedy velmi výhodné pro jednoduché transakce (například mazání či změny záznamů). Analýza velkého množství dat je však velmi neefektivní a pomalá, kvůli neustále se opakujícím výpočtům.

V datových skladech se již uložené údaje nemají měnit, mazat, ani jinak upravovat. Z toho důvodu není potřeba použít databázi optimalizovanou pro tento typ operací. Hlavním cílem je vytvořit úložiště, které bude poskytovat rychlé a nenáročné zpracování obrovského objemu dat. Tímto se dostáváme k multidimenzionálnímu (vícerozměrnému) modelu databáze.

Fakty a dimenze

Multidimenzionální model se skládá z dvou druhů údajů: faktů a dimenzí.

- **Fakty (ukazatele)** – jsou numerické měrné jednotky obchodování, které lze agregovat. Uloženy jsou v tabulce faktů, která je většinou v databázi pouze jedna a obsahuje velké množství dat. Primární klíč této tabulky je složený z cizích klíčů tabulek jednotlivých dimenzí. [18]
- **Dimenze (rozměr)** – jedná se o množinu hodnot diskrétního typu. Důležitou vlastností je uspořádatelnost (logická nebo hierarchická). Velmi časté je použití časových, produktových a geografických dimenzí. Tabulky, ve kterých jsou uloženy, jsou oproti tabulkám faktů menší a jejich hodnoty se nemění tak často. Většinou si je lze představit jako číselník, uchovávající textové informace o hodnotách faktů. [15][18]

Agregace dat

Multidimenzionální databáze se také často používají pro agregaci dat, což umožňuje uživatelům počítat a zobrazovat data na různých úrovních detailu.

Dle [8] je třeba rozlišovat tyto pojmy:

- Agregace – shrnutí či shlukování pomocí agregační funkce.
- Agregační funkce – funkce, která několik údajů sloučí do jediné výsledné hodnoty. Typické agregační funkce jsou:
 - Počet
 - Součet
 - Aritmetický průměr
 - Maximum a minimum
 - Modus (nejčastěji se vyskytující hodnota)
 - Medián (prostřední hodnota)
- Agregát – výsledná hodnota agregační funkce.

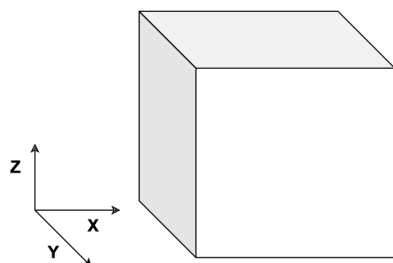
2.4.1 Datová kostka

Datová kostka (anglicky *data cube*) je datová struktura, která slouží k ukládání a analýze multidimenzionálních dat. Je to konceptuální model, který umožňuje ukládat data v mnoha různých úrovních sestavení a umožňuje rychlý přístup k datům pro různé druhy analýzy.

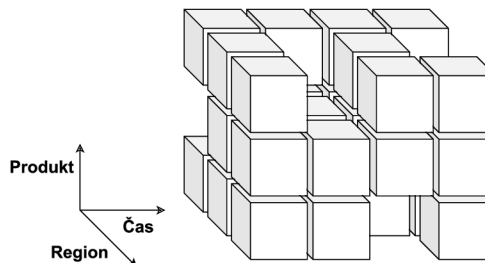
Datová kostka se skládá z libovolného počtu dimenzí, které určují samotnou polohu v datové struktuře. Při zvolení konkrétních hodnot všech použitých dimenzí se dostaneme na úroveň faktů. V jiném případě získáme agregát faktů, odpovídající zvolenému podprostoru pomocí hodnot dimenzí.

Vhodným přirovnáním třídimenzionální datové kostky je třírozměrná geometrická kostka. Při zvolení všech souřadnic dostáváme bod, pokud jsou ovšem určeny pouze dvě, získáme úsečku. V tomto případě nastoupí u datové kostky agregační funkce, která sloučí všechny zvolené fakty do jedné hodnoty. Dalším rozdílem je definovanost geometrické kostky v každém místě určeném souřadnicemi. U datové nejsou všechny body vyplněny fakty, jedná se tedy o řádkou n -rozměrnou matici. [8]

Geometrická kostka



Datová kostka



Zdroj: vlastní.

Obrázek 2.3: Rozdíl geometrické a datové kostky.

2.4.2 Implementace datové kostky

Přímé uložení datové kostky v databázi není možné. Pro její reprezentaci je proto nutné využít dimenzionálního modelu, který obsahuje určité topologické uspořádání, nazývaní se schéma. [15]

Nejčastěji se používají tato dvě schémata:

- Schéma hvězdy – jedna tabulka faktů obsahující cizí klíče, vytvářející vazbu na tabulky dimenzí. Dimenze nejsou normalizované, proto zde nejsou efektivně využity kapacity datového skladu. Na druhou stranu poskytuje rychlejší odezvu na dotazy. [15]
- Schéma sněhové vločky – dimenze jsou složeny z několika relačně spojených tabulek, které jsou již normalizované. Díky této vlastnosti dochází k rychlejším a jednodušším změnám v dimenzích. Tento způsob je ovšem časově náročný na realizaci. [15]

Schéma hvězdy

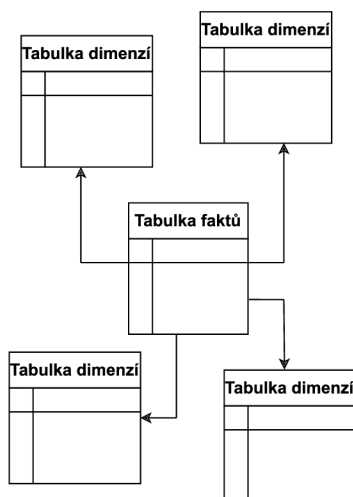
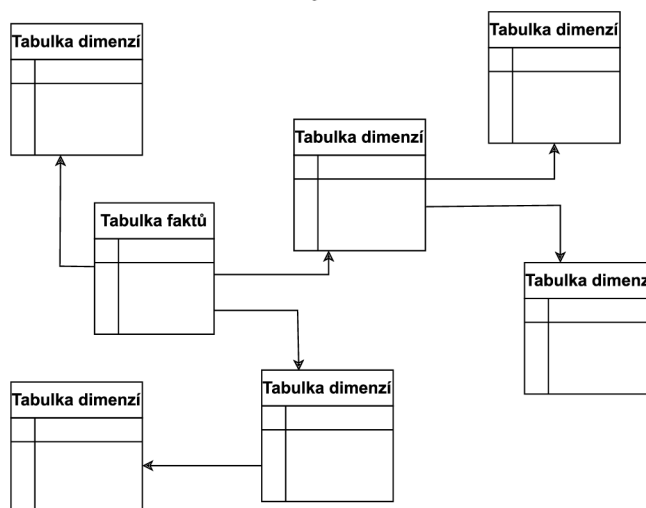


Schéma sněhové vločky



Zdroj: vlastní (podle [15]).

Obrázek 2.4: Porovnání schémat hvězdy a sněhové vločky.

2.5 Analýza OLAP

Pojmem OLAP (*Online Analytical Processing*) je nazývána technologie pro analýzu multidimenzionálních dat, která umožňuje uživatelům rychle a snadno vyhledávat a analyzovat data z různých úhlů pohledu. Často je používána například pro podnikové rozborů a reporting. OLAP se skládá z datového skladu, multidimenzionálního modelu dat a nástrojů pro analýzu dat. Tyto systémy umožňují rychle a efektivně vytvářet komplexní analýzy dat, jako je seskupování, filtrování a hloubkové procházení. Uživatel tak může lépe porozumět trendům, vzorcům a vztahům v datech.

Jednou z výhod OLAP je, že umožňuje svižný přístup k datům a snadnou interpretaci výsledků. Uživatelé mohou provádět analýzy rychle a efektivně, aniž by museli umět psát složité SQL dotazy. To umožňuje podnikovým analytikům a manažerům okamžitě reagovat na změny v podnikání a rychle rozhodovat.

2.5.1 Pravidla OLAP technologií

Dle [7] sepsal Edgar F. Codd (britsko-americký matematik a informatik) dvanáct pravidel týkajících se OLAP technologií:

1. Multidimenzionální pohled – analytik pohlíží na realitu ve více dimenzích, proto musí být data uložena pomocí modelu, který umožňuje takový pohled.
2. Transparentnost – pokud je OLAP součástí uživatelem používaného programu, mělo by to být zřejmé.
3. Dostupnost – schopnost přistupovat ke všem údajům potřebným pro danou analýzu, nezávisle na zdroji dat, ze kterého pochází, což ukazuje význam použití ETL.
4. Konzistentní výkon – zvýšení počtu dimenzí by nemělo mít zásadní vliv na výkon.
5. Klient-server architektura – OLAP server umožňující připojit různé klienty s minimálním úsilím.
6. Rovnoprávná dimenzionalita – každá dimenze by měla být ekvivalentní jak ve své struktuře, tak v operačních schopnostech.
7. Dynamická práce s řídkými maticemi – systém OLAP má být schopen dynamicky měnit své fyzické schéma, za účelem optimalizace zpracování řídkých matic.
8. Víceuživatelská podpora – poskytování souběžného načítání a aktualizací přístupu, integrity a zabezpečení.
9. Neomezené mezidimenzionální operace – výpočetní zařízení má umožňovat výpočty a manipulaci s daty napříč libovolným počtem datových dimenzí.
10. Intuitivní práce s údaji – uživatelské rozhraní by mělo nabízet jednoduché a intuitivní provádění základních OLAP operací.
11. Flexibilní vykazování – schopnost prezentovat informace jakýmkoli způsobem, který si uživatel přeje zobrazit.
12. Neomezený počet dimenzí a úrovní agregace – systémy OLAP by neměly zavádět žádná umělá omezení.

2.5.2 Druhy architektur serveru OLAP

Existují tři základní architektury, které lze při tvorbě OLAP serveru použít.

- Multidimenzionální OLAP (MOLAP) – dochází zde k výpočtu co nejvíce předběžných výsledků, které jsou následně uloženy ve vlastních datových strukturách (vícerozměrných). Hlavní výhodou je obrovský výkon vzhledem k dotazům uživatelů, na druhou stranu dochází k redundanci dat a tedy i vzniku velkých požadavků na výpočetní kapacitu. [15]
- Relační databázový OLAP (ROLAP) – údaje pochází z relačních datových skladů. Tyto údaje se uživateli překládají a jeví jako multidimenzionální. Data a metadata jsou ukládána jako záznamy v relační databázi, proto se zde nevyskytuje problém s redundancí. [15]
- Hybridní OLAP (HOLAP) – kombinuje MOLAP a ROLAP. Údaje jsou uloženy v relačních databázích, ale spočítané agregované hodnoty se ukládají do multidimenzionálních struktur. Snaží se tedy zachovat výhody obou a zároveň eliminovat nevýhody. [15]

2.5.3 Základní operace v OLAP analýze

Klíčovým prvkem v OLAP analýze je možnost uživatele pohlížet na data z mnoha různých úhlů. Z toho důvodu se používají základní operace nad OLAP kostkou, zajišťující tuto vlastnost. Jedná se o operace: zavrtávání, vynořování, otáčení, filtrování a tvorba řezů.

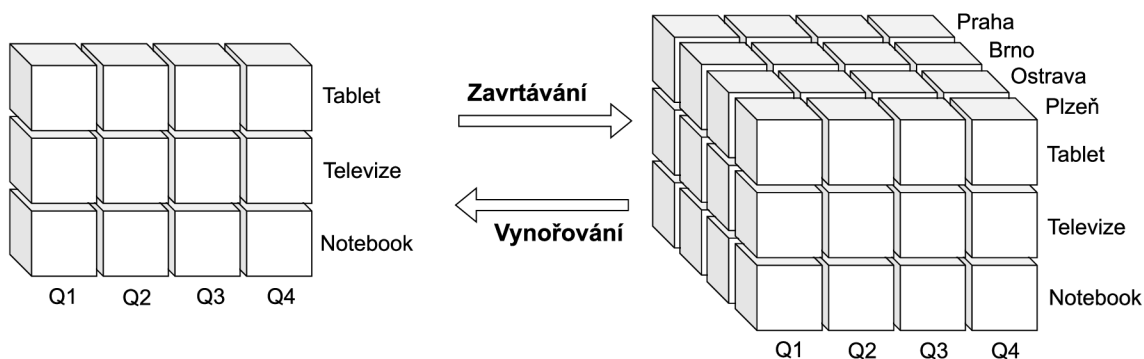
Zavrtávání (anglicky *Drill-down*)

V této operaci se nahradí obecná data detailnějšími. Implementace je možná buď přidáním další dimenze, nebo posunutím v hierarchii dimenzí o úroveň níže. [6]

Vynořování (anglicky *Roll-up*)

Inverzní operace k zavrtávání, tedy z detailnějšího popisu obecnější. Dochází ke zvýšení úrovně agregace, omezením počtu dimenzí nebo posunem v hierarchii dimenzí výše. [6]

Na obrázku 2.5 jsou graficky znázorněny operace zavrtávání a vynořování.

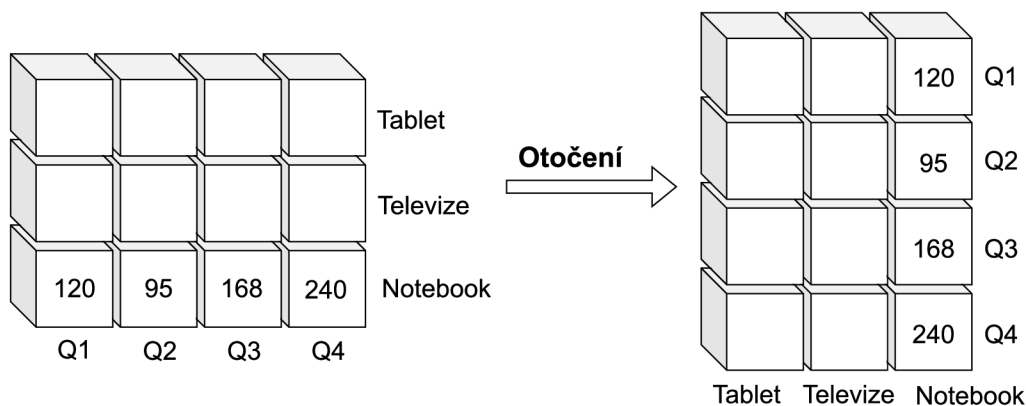


Zdroj: vlastní.

Obrázek 2.5: Operace zavrtávání a vynořování.

Otáčení (anglicky *Pivoting*)

Provádí změnu souřadnic datové kostky, za účelem získání alternativního pohledu na data. [6]

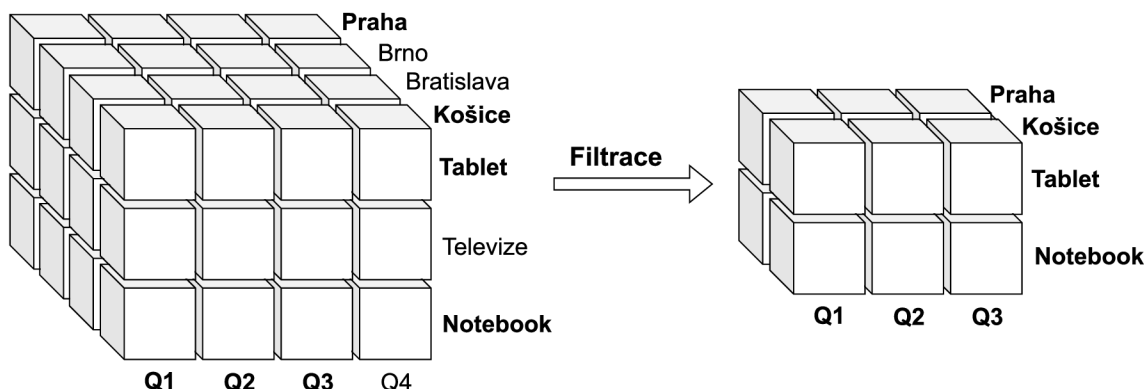


Zdroj: vlastní.

Obrázek 2.6: Operace otočení.

Filtrování (anglicky *Dicing*)

Operace vytvoří podkostku obsahující pouze vybrané hodnoty jednotlivých dimenzí. [6]

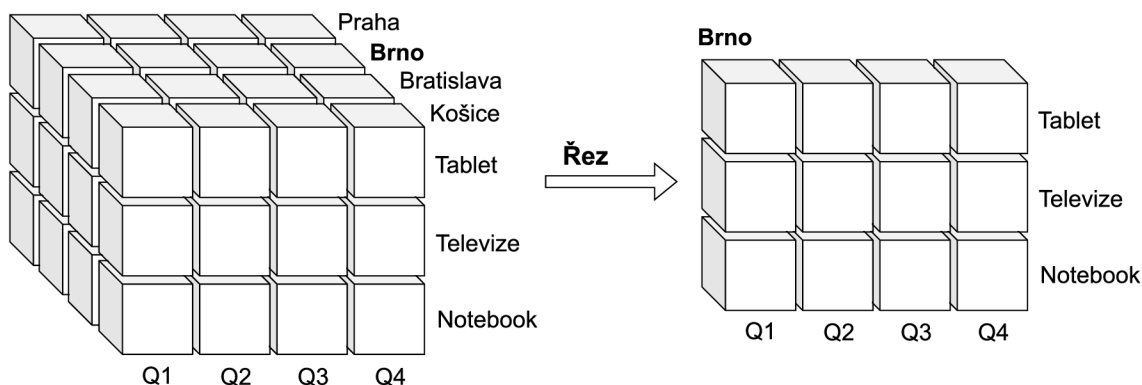


Zdroj: vlastní.

Obrázek 2.7: Operace filtrování.

Tvorba řezů (anglicky *Slicing*)

Jedná se o speciální případ filtrace, kdy se vybere pouze jedna konkrétní hodnota z dané dimenze kostky a dojde k vytvoření podkostky, která reprezentuje řez původní datové kostky. [6]



Zdroj: vlastní.

Obrázek 2.8: Operace řez.

2.6 Business Intelligence (BI)

Pro získání uceleného pohledu je nutné popsat pojem *Business Intelligence*. Jedná se o sadu procesů, aplikací a technologií zajišťující sběr, integraci, analýzu a prezentaci podnikových dat pro podporu řídicích aktivit v organizaci. BI zahrnuje širokou škálu nástrojů a metod, jako jsou reporty, analýzy (OLAP), dashboardy a datové sklady. [18]

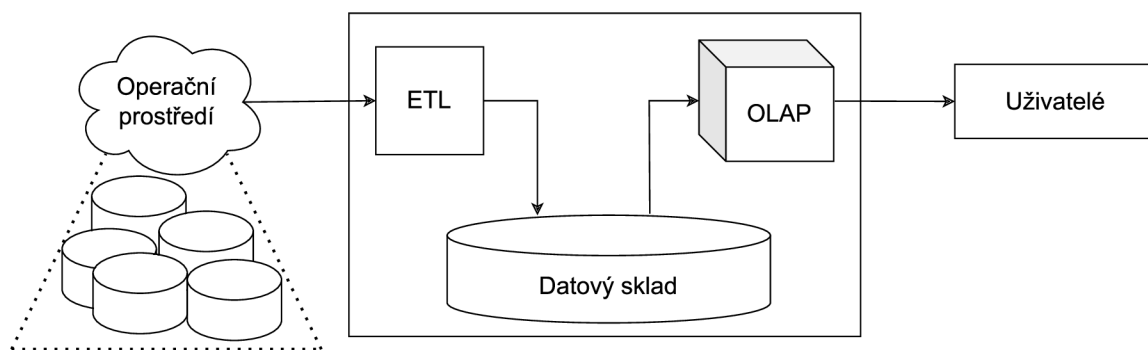
Cílem BI je poskytnout podnikovým uživatelům přehled o aktuálním stavu organizace a umožnit jim lépe porozumět. Díky tomu mohou manažeři rychle rozhodovat na základě

podnikových ukazatelů a jejich změn v čase, nebo předpovídat budoucí trendy a scénáře. BI také umožňuje identifikovat problémy a příležitosti, které by jinak zůstaly nezjištěny. [18]

Využití se najde v různých odvětvích a pro různé účely, například pro analýzu prodejních dat, finančních dat, dat o zákaznících a dalších. BI systémy jsou obvykle koncipovány tak, aby byly snadno pochopitelné a jednoduše použitelné. [14]

2.7 Tok dat při použití OLAP

Data musí při analýze projít dlouhou cestou. Nejdříve se pomocí procesu ETL získají všechny potřebné údaje z operačních databází, kterých může být větší množství. Dále dochází k čištění, sjednocení a zavedení do datového skladu, případně datových trhů. Nad těmito úložišti se provádí OLAP analýza, díky které jsou uživatelé schopni data dále zpracovávat. Celý tento průběh je principiálně zobrazen na obrázku 2.9.



Zdroj: vlastní (podle [15]).

Obrázek 2.9: Datový tok.

Kapitola 3

SAP

SAP (zkratka anglických slov *Systems, Applications, and Products in Data Processing*) je mezinárodní softwarová společnost, která se zaměřuje zejména na vývoj podnikových aplikací. Tato společnost byla založena v roce 1972 v Německu a je jednou z největších firem v oboru podnikových informačních technologií. SAP nabízí širokou škálu produktů, které řeší určité potřeby. Například pro správu a řízení financí (SAP ERP), vedení vztahů se zákazníky (SAP CRM), podporu nákupů ve společnosti (SAP SRM) a další. Tento software je určen pro střední a velké podniky a pomáhá jim optimalizovat své obchodní postupy, zlepšit efektivitu a snižovat náklady. [2]

Společnost SAP poskytuje jak cloudové, tak i on-premise¹ řešení, které umožňují uživatelům pracovat s daty a informacemi z různých zdrojů a zařízení. V nabídce jsou i nástroje pro analytiku a reporting, díky kterým jsou organizace schopny získávat a rozebírat svá data, což jim pomáhá lépe porozumět svým obchodním procesům a učinit informovaná rozhodnutí. Softwarové aplikace od SAPu jsou vysoce flexibilní a lze je přizpůsobit potřebám jednotlivých firem. [2]

V této kapitole si popíšeme podstatné termíny pro lepší orientaci v rámci produktů této společnosti a zároveň rozebereme její implementaci nástrojů pro OLAP analýzu rozsáhlých dat.

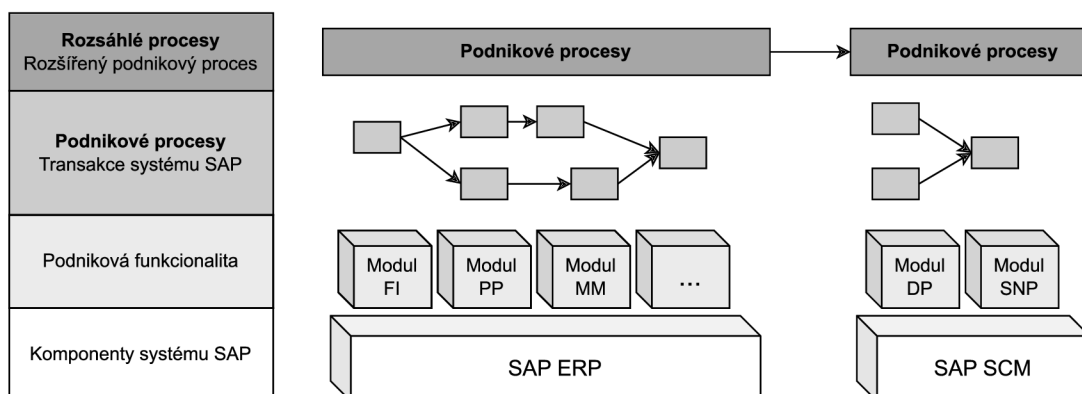
3.1 Komponenty, moduly a transakce

Problémem při popisu jednotlivých částí produktů od společnosti SAP je rozlišování pojmů komponenta, modul a transakce systému SAP.

- Komponenta – mnohdy nahrazována názvem podniková aplikace či pouze aplikace. Tyto komponenty se skládají z různých modulů a funkcí, které spolu dohromady tvoří kompletní řešení pro specifickou oblast obchodního řízení. [2]
- Modul – nabízí určitou funkcionalitu, která je nutná k řešení specifického obchodního problému. Příkladem mohou být moduly finančního účetnictví, plánování nebo výroby. Jejich spojováním vznikají samotné komponenty. Tyto moduly se také mohou nastavit a konfigurovat tak, aby vyhovovaly specifickým potřebám daného podniku. [2]

¹Jedná se o software, který je nainstalován a běží přímo na počítačích dané firmy nebo uživatele, nikoli na vzdáleném zařízení.

- Transakce – umožňují uživatelům provádět specifické úkony, jako je například zadávání objednávek, vytváření faktur, správa zdrojů a další. Dalo by se říct, že se jedná o jednotlivé kroky podnikového procesu, přičemž při jejich provedení ve správném pořadí je tento proces dokončen. SAP transakce se většinou spouštějí pomocí specifického kódu nebo názvu transakce. [2]



Zdroj: vlastní (podle [2]).

Obrázek 3.1: Komponenty, moduly a transakce systémů SAP.

3.2 Technologie SAP NetWeaver

SAP NetWeaver je technologickým základem většiny produktů společnosti SAP. Jejím cílem je redukování času nezbytného k integraci původně samostatných a oddělených aplikací, zkrácení doby potřebné na vývoj související s upgradem systému a tím i snížení nákladů. Je to výkonný a flexibilní systém, který umožňuje propojení a spolupráci různých řešení a informačních systémů v rámci podniku. Jedná se tedy o platformu pro integraci firemních aplikací. SAP NetWeaver také poskytuje široké možnosti pro vývoj a nasazení nových aplikací, jakož i pro řízení a správu stávajících aplikací. [2]

Z důvodu velkého množství produktů, skrývajících se pod názvem SAP NetWeaver, se podle [2] SAP rozhodl dělit tuto kolekci aplikací a nástrojů do šesti oblastí (někdy nazývaných doménami či tématy):

- Řízení základu (anglicky *Foundation Management*) – snaží se o unifikaci prostředí, obsahuje zejména aplikační server a dále například aplikace pro správu identit a aplikací.
- Middleware – označuje nástroje sloužící k propojení různých systémů takovým způsobem, aby mohli sdílet data a podporovat podnikové procesy probíhající ve vyšších vrstvách.
- Řízení informací (anglicky *Information Management*) – zahrnuje procesy usnadňující rozhodování tím, že zpřístupní všechny potřebné informace správným lidem ve správný čas. Patří sem mimo jiné i SAP Business Warehouse (datový sklad).
- Produktivita týmu (anglicky *Team Productivity*) – obsahuje nástroje pro zvýšení uživatelské přívětivosti a s tím související efektivity práce jak jednotlivců, tak celých týmů. Příkladem může být webové rozhraní pro sdílení informací.

- Kompozice (anglicky *Composition*) – jedná se o aplikace pro samotný vývoj, případně monitorování a řízení podnikových procesů.
- Řízení podnikových procesů (anglicky *Business Process Management*) – dalo by se říct, že se jedná o podmnožinu kompozice, která se zaměřuje pouze na řízení firemních procesů.

3.3 Komponenta SAP ERP

SAP ERP (zkratka anglických slov *Enterprise Resource Planning*) je podnikové řešení společnosti SAP, které umožňuje organizacím efektivně řídit a automatizovat své obchodní procesy. SAP ERP řeší řadu problémů, jako jsou finanční účetnictví, nákupy, prodej, lidské zdroje a řízení dodavatelského řetězce. Kromě toho poskytuje řešení pro analytiku a reporting, které umožňují společnostem získávat nebo analyzovat data a zlepšit tak svou schopnost rozhodování a plánování. [2]

Podle [3] má tato komponenta dva typy modulů:

- Funkční moduly
 - SAP FI – finanční data organizace.
 - SAP HR – správa lidských zdrojů.
 - SAP MM – správa materiálních zdrojů.
 - a další
- Technické moduly
 - SAP ABAP – programovací jazyk.
 - SAP Solution Manager – správa používaných aplikací.
 - SAP BW – datový sklad.
 - a další

Aplikace SAP ERP také poskytuje široké možnosti personalizace a lze ji přizpůsobit potřebám každé organizace. Celkově by se dalo říci, že SAP ERP je komplexním řešením pro řízení podnikání, které pomáhá společnostem optimalizovat své obchodní procesy, zlepšit efektivitu a snižovat náklady. [2]

3.4 Modul SAP BW

SAP BW (zkratka anglických slov *Business Warehouse*), někdy nazýván SAP NetWeaver BW, je modul komponenty SAP ERP. Slouží jako centralizované úložiště dat a informací pro podnikovou analytiku nebo rozhodování. Tento modul umožňuje organizacím získávat, integrovat i analyzovat různé typy dat z mnoha rozdílných zdrojů, včetně interních a externích systémů, které nemusí pocházet od společnosti SAP.

Jedná se tedy o implementaci problematiky datových skladů a OLAP analýzy, nevymáje nástrojů pro proces ETL. Pro uložení a reprezentaci multidimenzionálního modelu dat je využito upravené schéma hvězdy. [4]

3.4.1 PSA

Vstupním bodem dat do modulu SAP BW je PSA (zkratka anglických slov *Persistent Staging Area*). Je to persistentní úložiště sloužící k uchování dat v takové podobě, v jaké byla ze zdrojového systému dopravena. Výjimkou mohou být technické konverze, zdroj dat totiž nemusí nabízet stejné datové typy jako SAP. Může se jednat například o změnu formátu datumu nebo čísla. [4]

Údaje jsou uloženy v tabulkách, do kterých mají vývojáři přístup. Díky tomu mohou snadno nahlédnout do nezpracovaných, pouze extrahovaných dat a hledat případné speciální či chybné záznamy. [4]

3.4.2 InfoProvider

Pojem InfoProvider označuje všechny objekty, které lze použít jako zdroj dat pro analýzu a reportování. Pro organizaci v rámci projektu je možné InfoProvadery ukládat do složek, nazývaných InfoArea. [4]

Podle [4] lze InfoProvadery dělit na dva typy:

- Objekty reálně obsahující data – SAP BW ukládá údaje těchto objektů do databázových tabulek.
 - InfoObjekt
 - InfoKostka
 - Objekt datová schránka (ODS)
- Logické pohledy – přímo neobsahují data, ta jsou uložena v jiných objektech či systémech.
 - RemoteCube – získává data z jiných systémů.
 - Virtuální InfoKostka – data pochází z datových zdrojů (nemusí být produktem společnosti SAP) přes uživatelem definované funkční moduly.
 - InfoSet – sémantická vrstva nad InfoProvadery s daty, které lze spojovat pomocí databázových operací, jako je například operace JOIN.
 - MultiProvider – podobně jako InfoSet kombinuje data z různých InfoProviderů. Rozdílem ovšem je, že může využít i dříve uvedené virtuální InfoProvadery, zároveň ale již nepodporuje operaci JOIN, ale pouze operaci UNION.

3.4.3 InfoObjekty

InfoObjekty tvoří základ datového modelu v SAP BW. Definují datové typy a jejich skládáním se vyrábí jak InfoKostky, tak ODSky. Dělí se na ukazatele (anglicky *key figures*) a charakteristiky (anglicky *characteristics*). [4]

Ukazatel

Ukazatele nesou měrné hodnoty, které je možné agregovat zvolenou agregační funkcí. Jedná se tedy o implementaci faktů společností SAP (viz 2.4).

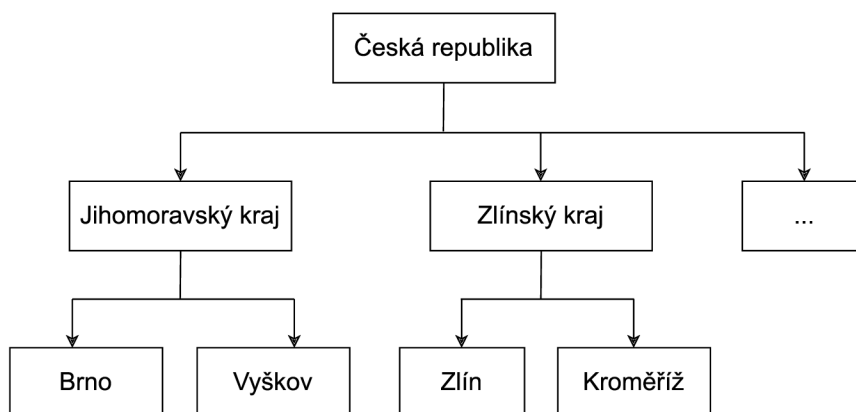
Při tvorbě ukazatele je nutné zvolit datový typ. Výběr je z desetinného a celého čísla, času, datumu a částky, u které je možné dodat její měnu, případně jednotku. Co se týče agregační funkce, můžeme vybírat mezi klasickou sumou, maximem a minimem.

Charakteristika

Charakteristiky implementují samotné dimenze datových kostek (viz 2.4). Jsou tvořeny diskrétními hodnotami, které lze filtrovat. Stejně jako u ukazatelů se musí definovat jejich datový typ. Tím může být čas, datum a znakový či numerický text s volitelnou délkou.

Tento typ InfoObjektu může obsahovat i další doplňující informace, poněvadž samotná jeho hodnota nemusí být vždy při analýze dostatečná. Pro jejich tvorbu je nutné vytvořit samostatnou transformaci, podle které se příslušná data nahrají. Na výběr je ze tří možností, přičemž každá se hodí v jiné situaci:

- Kmenové texty – tvoří rozšiřující popis dané hodnoty charakteristiky. Například pokud v charakteristice ukládáme identifikátory zboží, můžeme do těchto textů nahrát i název tohoto zboží. SAP BW dokáže pracovat i s časově či jazykově závislými popisy. Podle potřebné délky textu jsou nabízeny tři varianty, které lze libovolně kombinovat:
 - Krátký text – 20 znaků.
 - Středně dlouhý text – 40 znaků.
 - Dlouhý text – 60 znaků s možným rozšířením na 1333 znaků.
- Atributy – jsou to hodnoty charakteristik rozšiřující informace o charakteristice právě tvořené. Příkladem mohou být zaměstnanci, u kterých chceme vždy vědět i trvalé bydliště, telefonní číslo, email a další údaje. Pro jejich tvorbu už je nutné vytvořit další charakteristiky a ty následně použít jako atributy.
 - Navigační atribut – speciální typ atributu. Běžný atribut se při dalším použití chová pouze jako doplňující informace. Navigační atribut ovšem vypadá jako další dimenze a lze pomocí něj například i filtrovat.
- Hierarchie – nejsložitější varianta. Uspořádává hodnoty do hierarchické struktury. Je možné pro její tvorbu využít i více charakteristik reprezentující jednotlivé úrovně. Jako dobrý příklad mohou posloužit geografické dimenze, kde jako kořenový uzel figuruje samotný stát, přičemž se přes kraje dostaneme k obcím. Při následné analýze je možné se libovolně zanořovat nebo vynořovat v této struktuře.



Zdroj: vlastní.

Obrázek 3.2: Hierarchie geografických celků.

3.4.4 InfoKostka

InfoKostka (anglicky *InfoCube*) reprezentuje v SAP BW datovou kostku. Skládá se z několika charakteristik a alespoň jednoho ukazatele. Pro jednoznačné určení neagregovaného ukazatele jsou potřeba hodnoty všech charakteristik. Jinak řečeno všechny charakteristiky jsou klíčové. V praxi to ovšem data v objektu splňovat nemusí a primární klíč může tvořit jejich podmnožina.

Data jsou uložena podle speciálně upraveného schématu sněhové vločky a uživatelům jsou zpřístupněna ihned po zavedení. [4]

3.4.5 Objekt datová schránka (ODS)

Druhým způsobem implementace datové kostky v SAP BW je objekt datová schránka (anglicky *DataStore Object*), častěji nazýván svou zkratkou ODS. Stejně jako InfoKostka se skládá z charakteristik a ukazatelů, přičemž zde již není podmínka minimálně jednoho ukazatele. Ten je totiž vždy implicitní a ke každému vstupnímu záznamu přiřadí hodnotu 1, díky čemuž tvoří počítadlo záznamů.

Na rozdíl od InfoKostky je ODS uložena ve formě jednoduché tabulky, kde vývojář při její tvorbě určí, které charakteristiky jsou klíčové, ovšem jejich počet je značně limitován (16 klíčových polí). Tato reprezentace vede ke vzniku obrovských tabulek a proto jsou možnosti optimalizace výkonu značně více omezené než u InfoKostek. [4]

Data z ODS nejsou ihned po nahrání přístupná analytikům. Nejprve se dostanou do tabulky neaktivních údajů a pro jejich zpřístupnění se musí provést aktivace dat.

3.4.6 Rozdíl InfoKostky a ODS

InfoKostka i ODS reprezentují datovou kostku, jsou v nich ale určité rozdíly, které mohou být rozhodující pro volbu vhodného objektu v dané situaci. Proto jsou hlavní rozdíly shrnuty v následující tabulce:

Typ objektu	InfoKostka	ODS
Ukazatele	Minimálně jeden ukazatel.	Bez nutnosti explicitně přidaného ukazatele.
Klíčová pole	Všechny charakteristiky.	Zvolené charakteristiky.
Způsob uložení	Upravené schéma sněhové vločky.	Jednoduchá tabulka.
Přístup k datům	Dostupné ihned po nahrání.	Nutná aktivace dat.

Tabulka 3.1: Shrnutí rozdílů InfoKostky a ODS.

3.4.7 Transformace a Proces přenosu dat

Při zavádění dat z PSA do InfoProvederů dochází k transformaci. Jedná se o úpravy údajů, jako je kombinace, přejmenování sloupců, řazení, výpočty a další změny dat. Transformace je reprezentována speciálním objektem v SAP BW, který je nutné vytvořit a namodelovat. Důležitou vlastností je, že každý vstupní záznam je zpracován samostatně. Není možné tedy dělat výpočty z více řádků. Možným řešením by bylo přidat mezi PSA a cílový InfoProvider další ODS objekt, který by díky správně zvoleným klíčovým polím příslušnou agregaci provedl.

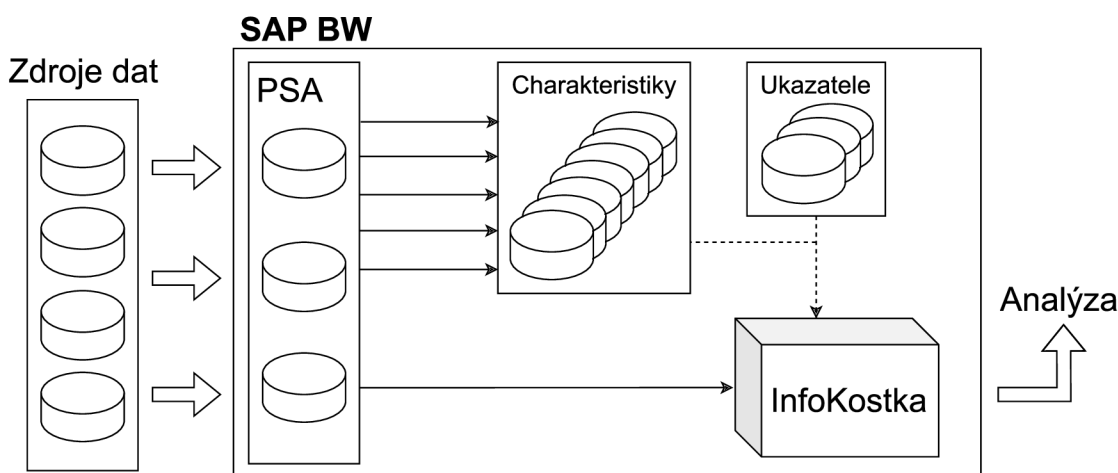
V transformaci je potřeba dělat různé úpravy s rozdílnou složitostí. Pro její tvorbu mají vývojáři na výběr z několika možností, jakým způsobem tento problém řešit. Mezi základní patří:

- Přímé přiřazení – zavedení bez úprav.
- Konstanta – do cílového pole se nahraje konstantní hodnota.
- Vzorec – slouží zejména pro snadnější úpravy. Využívá jednoduchých příkazů pro běžné konverze (získání podřetězce, matematické funkce, operace s daty a další).
- Rutina – nabízí více možností než vzorec. Úpravy jsou popisovány pomocí programovacího jazyka ABAP.

Ke každé transformaci se musí navíc vytvořit Proces přenosu dat, který určuje jakým způsobem jsou data přenášena. Definuje se například velikost paketů, způsob zpracování chyb či filtrace údajů.

3.4.8 Tvorba datové kostky v SAP BW

Celý proces tvorby zahrnuje mnoho objektů, transformací a procesů přenosu dat. Často je potřeba vytvořit i pomocné InfoProvedery, pro řešení konkrétních problémů. Zde se ale shrne pouze běžný případ a ukáže i samotný tok dat.



Zdroj: vlastní.

Obrázek 3.3: Tok dat v SAP BW.

Z obrázku 3.3 je zřejmé, že prvním krokem je vytvoření úložišť PSA pro uchování extrahovaných dat z různých zdrojů. Následují ukazatele a charakteristiky. U charakteristik je nutné vymodelovat i kmenové texty, atributy a hierarchie, spolu s příslušnými transformacemi, aby je bylo možné naplnit údaji z PSA. Tyto InfoObjekty se poté použijí pro sestavení InfoKostky (případně ODS). Posledním krokem před analýzou je zavedení dat do vytvářené reprezentace datové kostky.

3.5 Jazyk ABAP

ABAP (zkratka anglických slov *Advance Business Application Programming*) je programovací jazyk vyvinutý společností SAP zejména pro své databázové aplikace. Je možné použít rozšíření ABAP Objects, které přináší objektově orientované elementy, jako je zapouzdření, rozhraní či dědičnost. [13]

Tento jazyk umožňuje vývojářům používat širokou škálu nástrojů a funkcí, jako jsou například funkce pro práci s daty, tvorbu dotazů a reportů nebo integraci s dalšími aplikacemi. Díky tomu mohou vývojáři rychle a efektivně řešit specifické obchodní potřeby a vytvářet kvalitní reporty, formuláře či výstupy pro své zákazníky. [13]

Jeho syntaxe je podobná jazyku COBOL. Pro ukázkou jsou zde uvedeny dva příklady. První ukazuje implementaci programu Hello world v jazyku ABAP:

Algoritmus 1 Program *Hello World* v jazyku ABAP.

```
REPORT EXAMPLE.  
WRITE 'Hello World'.
```

Druhý principiálně zobrazuje větvení:

Algoritmus 2 Ukázka větvení v jazyku ABAP.

```
IF podmínka_1.  
  IF podmínka_2 AND podmínka_3.  
    příkazový_blok_1  
  ENDIF.  
ELSEIF podmínka_4.  
  příkazový_blok_2  
ELSE.  
  příkazový_blok_3  
ENDIF.
```

3.6 Možnosti vizualizace datových objektů

Aby bylo možné provádět analýzu, musí nám být poskytnuty i nástroje pro vizualizaci dat a případné OLAP operace nad kostkou. Společnost SAP nabízí několik produktů, poskytující potřebnou funkcionalitu. Příkladem mohou být:

- Business Explorer – viz 3.6.1
- SAP Analysis for Office – viz 3.6.4

- SAP Web Intelligence – umožňuje uživatelům vytvářet reporty a dashboardy pomocí webového prohlížeče, bez nutnosti instalace dalšího softwaru na uživatelově počítači.

3.6.1 Business Explorer (BEx)

Business Explorer (zkráceně BEx) je přímo součástí modulu SAP BW. Slouží k vytváření reportů a dashboardů pro analýzu dat uložených v datovém skladu. BEx neobsahuje pouze jeden produkt, ale zahrnuje několik samostatných nástrojů, mezi které patří například:

- BEx Query Designer – viz [3.6.2](#)
- BEx Analyzer – viz [3.6.3](#)
- BEx Web Application Designer – uživatelé mohou vytvářet jednoduché webové aplikace pro zobrazení reportů a dashboardů.

3.6.2 SAP BEx Query Designer

V programu BEx Query Designer se vytváří dotazy (anglicky *query*) nad InfoProvidery pomocí grafického uživatelského rozhraní, které usnadňuje tvorbu dotazů i uživatelům bez technického vzdělání. Tyto dotazy slouží jako další vrstva nad datovou kostkou, ve které je možné používat různé funkce, jako jsou filtrování, řazení a výpočty, pro získání potřebných výsledků. [5]

Mimo jiné obsahuje BEx i možnost definovat takzvané vypočtené ukazatele. Ty mohou nabývat vzorcem přepočtených hodnot jiných ukazatelů a zároveň je pro ně možné nastavit i jiné agregační funkce než u běžných ukazatelů.

3.6.3 SAP BEx Analyzer

SAP BEx Analyzer je nástroj pro analyzování informací získaných z datového skladu SAP BW pomocí rozhraní Microsoft Excel. BEx Analyzer umožňuje uživatelům vytvářet různé vizualizace dat, jako jsou grafy, tabulky nebo křivky. [5]

3.6.4 SAP Analysis for Office

SAP Analysis for Office je nástroj, který umožňuje provádění multidimenzionální analýzy v aplikacích Microsoft Office, konkrétně MS Excel a MS PowerPoint. Nabízí uživatelům funkce pro zobrazení dat ve formě grafů, tabulek a dalších vizuálních prvků. Zároveň se dokáže připojit k různým datovým zdrojům včetně SAP BW, SAP HANA či jiným systémům. [19]

SAP Analysis je často zaměňován s programem SAP BEx Analyzer. Hlavním rozdílem mezi těmito nástroji je v tom, že SAP BEx Analyzer je určen pro analýzu dat získaných z SAP BW, zatímco SAP Analysis for Office umožňuje analýzu dat z různých zdrojů.

Kapitola 4

Zvolená oblast dat pro datovou kostku

Aby tvorba datové kostky měla smysl, musí obsahovat data, jejichž analýzou dochází k zisku hodnotných informací. Jelikož tato práce vznikla v rámci rozjetého projektu organizace CVIS VUT (Centrum výpočetních a informačních služeb), který se zabývá vývojem manažerského informačního systému pro univerzitu VUT v Brně, je zřejmé, že data pochází z prostředí vysoké školy.

V době zahájení této práce byla aktuálně zpracovávána studijní agenda, přičemž již byly hotové či rozpracované datové kostky obsahující data z různých oblastí, jako studium, studijní změny, věda a výzkum, rozvrhy nebo předměty. Tyto tematické okruhy tedy nepřípadaly v úvahu.

Po diskuzi s analytiky VUT volba padla na přijímací řízení. Data z této oblasti jsou každoročně vyplňována do výroční zprávy o činnosti a zároveň podléhají opakovaným analýzám, za účelem zisku cenných informací, které mohou ukazovat například vývoj zájmu uchazečů o studium určitých oborů.

V této kapitole je nejdříve popsán celý proces přijímacího řízení. Následovně se zaměří na výroční zprávy, které tvořily prvotní signál pro volbu právě této oblasti.

4.1 Přijímací řízení univerzity

Přijímací řízení na vysoké škole je proces, pomocí kterého se určuje, zda kandidát splňuje požadavky pro přijetí na vybraný program, zda má dostatečné schopnosti a znalosti pro dané studium a tedy i předpoklady na jeho úspěšné dokončení. Každá univerzita má své vlastní přijímací kritéria a postupy, které většinou zahrnují následující kroky:

1. Přihlášení – kandidáti se musí zaregistrovat a vyplnit přihlášku k přijetí.
2. Požadavky na dokumenty – uchazeči musí poskytnout potřebné dokumenty. Příkladem mohou být vysvědčení, životopisy nebo doporučení.
3. Testy a zkoušky – některé vysoké školy požadují účast na zkouškách, jako jsou například přijímací testy, jazykové či talentové zkoušky nebo pohovory.
4. Hodnocení a rozhodnutí – přijímací komise hodnotí všechny dokumenty spolu s výsledky testů a zkoušek, aby rozhodla, zda kandidát splňuje požadavky na přijetí.

5. Oznámení o rozhodnutí – vysoká škola oznámí uchazečům své rozhodnutí o jejich úspěšnosti.
6. Zápis – přijatí uchazeči potvrzují účast ve studiu a stávají se z nich studenti dané školy.

Každá univerzita může klást důraz na jinou část přijímacího řízení. Některé školy se soustředí především na výsledky z přijímacích zkoušek, zatímco jiné kladou důraz na životopis a motivaci uchazeče, jiné používají standardizované testy, či vyžadují pohovory a talentové zkoušky. Avšak navzdory rozdílným přístupům mají všechny vysoké školy společný cíl: najít a přijmout ty nejlepší uchazeče. [20]

4.1.1 Přijímací řízení VUT

Na univerzitě VUT se uchazeči o studium na libovolné fakultě přihlašují přes systém nazývaný e-přihláška. Po registraci musí kandidáti vyplnit potřebné osobní údaje, jako je jméno a příjmení, státní příslušnost, pohlaví, místo a datum narození, rodné číslo, email a telefonní číslo. Dále se udává adresa trvalého bydliště a kontaktní adresa. Po dodání těchto dat je nutné doložit i předchozí studium, aby se ověřilo, zda má uchazeč dokončené vzdělání požadované úrovně.

Následuje výběr programů a specializací nabízených jednotlivými fakultami pro daný akademický rok. Samotná nabídka na VUT obsahuje zejména dvě formy studia:

- Prezenční studium
- Kombinované studium

Vysoká škola VUT poskytuje několik typů studia, které se liší vstupními požadavky, délkou i získaným titulem po úspěšném ukončení. Mezi hlavní typy studia patří:

- Bakalářské studium
- Magisterské navazující studium
- Doktorské studium

Po podání přihlášek začíná proces rozhodování o přijetí. Podmínky se zde velmi liší nejen mezi fakultami, ale někdy i mezi jednotlivými programy vyučovaných stejnou fakultou. Všeobecně by se ovšem dalo říct, že VUT často umožňuje prominutí přijímací zkoušky, spolu s přednostním přijetím, při splnění určitých kritérií, popsanych u daného studijního programu.

Předmětem těchto kritérií může být absolvování přípravného kurzu nabízeného fakultou a jeho následná klasifikace, výsledek Národní srovnávací zkoušky, hodnocení maturity, průměr známek z předchozího studia či hodnota naměřeného IQ. [1]

Pokud nedojde u uchazeče k přednostnímu přijetí, účastní se přijímací zkoušky, která se může lišit i pro různé programy. Na základě jejího výsledku jsou vybráni nejlepší kandidáti až do maximálního počtu přijatých uchazečů. [1]

Závěrečnou částí procesu je zápis, po kterém se z přijatých uchazečů stávají studenti. To s sebou nese řadu akcí, jako například tvorba VUT identity.

4.2 Výroční zpráva

Výroční zpráva je dokument, kterým podnik nebo organizace informuje o svých výsledcích, plánech a činnosti za uplynulé časové období (většinou jeden rok). Výroční zpráva je důležitým prvkem korporátní prezentace a propagace, ale zároveň slouží jako nástroj pro transparentnost a zodpovědnost vůči vlastníkům, zákazníkům, investorům, zaměstnancům a dalším zainteresovaným stranám. [16]

Výroční zpráva obvykle obsahuje informace o finančních výsledcích podniku, jako jsou příjmy, výdaje, zisky a ztráty, stejně jako další finanční ukazatele. Dále se v ní často nacházejí informace o organizační struktuře, zaměstnaneckých benefitech, environmentální a sociální odpovědnosti a dalších klíčových oblastech činnosti organizace. [16]

Některé společnosti mají takzvanou informační povinnost, která je daná zákonem, a musí tedy výroční zprávu vydávat. [16]

4.2.1 Výroční zprávy vysokých škol

Speciálním případem jsou výroční zprávy vysokých škol, které univerzity každoročně dodávají Ministerstvu školství, mládeže a tělovýchovy. U veřejných vysokých škol se jedná o zákonnou povinnost, kdežto u soukromých je to volitelné v případě, že neobdržela dotaci ze státního rozpočtu. [17]

Zprávy obsahují informace o výsledcích vzdělávání a výzkumu, plánech, projektech a aktivitách v oblasti akademického a společenského života školy. Získaná data ministerstvo zpracovává do souhrnných tabulek. [17]

Školy vždy vydávají dvě zprávy:

- Výroční zpráva o činnosti – zkráceně VZoČ. Soustředí se zejména na přehled a výsledky hodnocení činností vysokých škol. Další částí jsou změny vnitřních předpisů a změny v orgánech vysoké školy, které se udály v průběhu roku. Právě v této zprávě se vždy nachází i údaje z příjmacího řízení. [17]
- Výroční zpráva o hospodaření – nese informace o financích, tedy příjmy, výdaje, majetek a další. [17]

Kapitola 5

Návrh datové kostky

Návrh datové kostky je proces, při kterém se definují dimenze a fakty datové kostky, spolu s následným určením vhodné reprezentace v systému SAP BW. Dalším krokem je stanovení kmenových textů, atributů a hierarchií pro jednotlivé charakteristiky (dimenze).

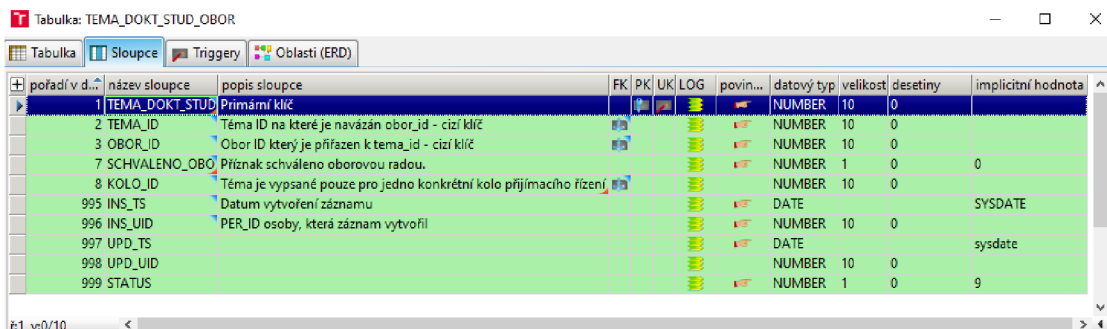
Tvorba datového skladu na VUT probíhá metodou zdola nahoru (viz 2.3.1), což znamená, že se postupně vytváří datové trhy, obsahující informace relevantní pouze pro daný účel. Nad těmito trhy jsou poté vyráběny datové kostky. Díky tomu se v rámci návrhu musí určit i které údaje budou v kostce požadovány, aby se mohly získat ze zdroje dat.

V této kapitole jsou popsány všechny tyto části procesu návrhu datové kostky, předcházejícímu samotné implementaci.

5.1 Zdroj dat

Data kostky jsou čerpána pouze z centrální databáze VUT, která obsahuje veškeré ukládané údaje o dění na univerzitě. Příkladem mohou být informace o studentech, zaměstnancích, předmětech nebo výzkumu. Díky tomu, že data jsou již centralizovaná, není nutné zkoumat více zdrojů dat.

Databáze je sémanticky dělená na mnoho tematických oblastí, za účelem lepší orientace. Její dokumentace je tvořena pomocí ER diagramů, mapující jednotlivé oblasti. Ke každému názvu sloupce je prostor i pro jeho popis, aby bylo v případě neúplně vysvětlujícího názvu možné přidat poznámku. Přístup k dokumentaci je řešen pomocí informačního systému VUT, nazývaného Apollo.



pořadí v d...	název sloupce	popis sloupce	FK	PK	UK	LOG	povin...	datový typ	velikost	desetiny	implicitní hodnota
1	TEMA_DOKT_STUD	Primární klíč						NUMBER	10	0	
2	TEMA_ID	Téma ID na které je navázán obor_id - cizí klíč						NUMBER	10	0	
3	OBOR_ID	Obor ID který je přiřazen k tema_id - cizí klíč						NUMBER	10	0	
7	SCHVALENO_OBOR	Průznak schváleno oborovou radou.						NUMBER	1	0	0
8	KOLO_ID	Téma je vypsané pouze pro jedno konkrétní kolo přijímacího řízení						NUMBER	10	0	
995	INS_TS	Datum vytvoření záznamu						DATE			SYSDATE
996	INS_UID	PER_ID osoby, která záznam vytvořil						NUMBER	10	0	
997	UPD_TS							DATE			sysdate
998	UPD_UID							NUMBER	10	0	
999	STATUS							NUMBER	1	0	9

Zdroj: vlastní.

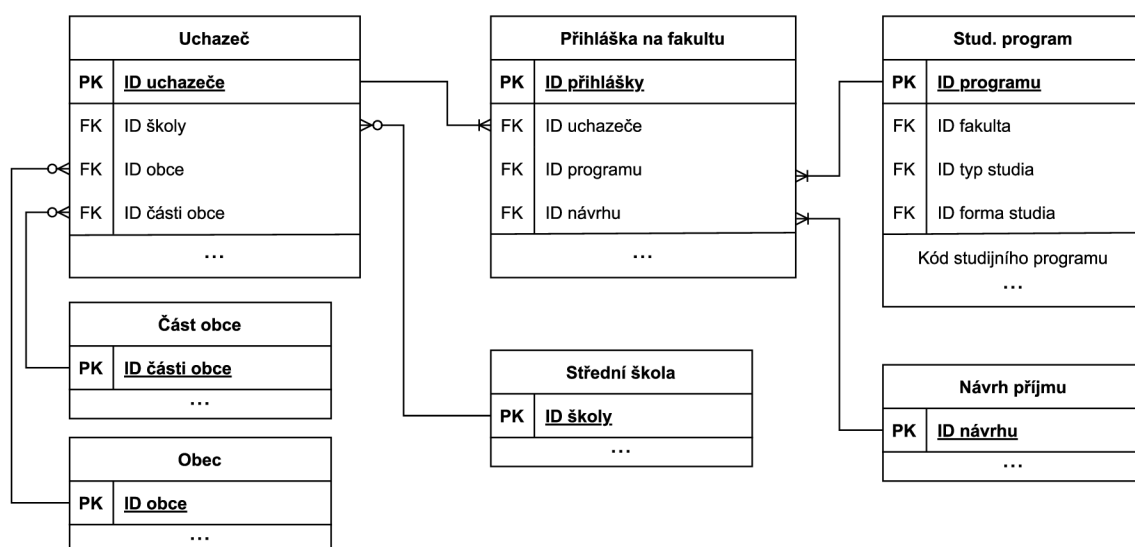
Obrázek 5.1: Ukázka dokumentace jedné tabulky v Apollu.

5.2 Volba použitých dat

Tabulky výročních zpráv o činnosti z minulých let nám daly první informace o potřebných údajích v datové kostce. Standardně se vyplňují počty uchazečů, přihlášek, přijetí a zápisů ke studiu podle fakult, typu studia a kódu ISCED-F¹. Další žádané údaje vplynuly z diskuze s analytiky. Jedná se zejména o bližší specifikaci programu, na který je podávána přihláška, informace o uchazeči a střední škole.

Data jsou čerpána ze tří oblastí databáze:

- Přijímací řízení (epřihláška)
- Studium
- Adresy



Zdroj: vlastní.

Obrázek 5.2: Zjednodušené ERD oblasti přijímacího řízení.

Nejvíce dat je získáno právě z oblasti Přijímacího řízení (epřihláška). ER diagram této části obsahuje více než 50 entitních množin a není ho zde možné prezentovat celý. Obrázek 5.2 znázorňuje nejzásadnější tabulky tohoto schématu.

Tabulka Uchazeč už sama o sobě obsahuje, mimo základní údaje o uchazeči, informace o střední škole, jako je její typ či IZO². Z tohoto důvodu se daná data mohou brát i z této tabulky. Návrh příjmu je číselník, obsahující jak textový popis výsledku přijímacího řízení, tak sadu příznaků, určujících na základě čeho bylo rozhodnuto.

Problematický je kód ISCED-F, který také není sám o sobě uložen. Ovšem je ho možné sehnat ze sloupce Kód studijního programu v relaci Stud. program.

Z oblastí Adresy a Studium se berou pouze číselníky pro tvorbu kmenových textů, atributů a hierarchií. Blíže je použití popsáno v 5.2.2 a 5.5.

¹Kód popisující obor studia.

²IZO je zkratka pro pojem identifikační znak organizace. Jedná se o identifikátor základních a středních škol v České republice, pod kterým je škola uvedena v rejstříku škol.

5.2.1 Vybrané údaje z tabulek

Tabulka	Údaj
Přihláška na fakultu	ID přihlášky, programu, uchazeče a návrhu děkana
	Datum vzniku přihlášky
	Kód oboru
	Rok příjmu
Uchazeč	Trvalé bydliště a místo narození
	Pohlaví a státní příslušnost
	ID střední školy a oboru střední školy, typ a IZO
	Věk uchazeče v době podání přihlášky
Návrh příjmu	Textový popis výsledku
	Příznaky, popisující výsledek. (Přijat bez/po přijímací zkoušce a další)
Stud. program	ID fakulty, typu a formy studia
	Rok platnosti
	Obor ISCED-F

Tabulka 5.1: Čerpané údaje z jednotlivých tabulek.

V tabulce 5.1 jsou shrnuty údaje brané z databáze VUT. Většinou se jedná o informace rozložené ve více sloupcích tabulek. Pro názornost jsou ovšem uvedeny po tematických celcích (celkem vybráno přes 50 sloupců, viz tabulky B.1 a B.2). Hodnoty identifikátorů se použijí pro tvorbu charakteristik s dalšími informacemi z číselníků.

Problematickým údajem je věk uchazeče v době podání přihlášky, který není nikde přímo uložen. Získání je možné výpočtem z datumů narození a podání přihlášky, datum narození je ovšem citlivý údaj a nesmí se dostat ven.

5.2.2 Vybrané číselníky

Oblast	Číselník
Přijímací řízení	Střední škola
	Obor střední školy
Přijímací řízení/Adresy	Obec
	Část obce
Adresy	Region
	Část obce PSČ
Studium	Typ studia
	Forma studia

Tabulka 5.2: Použité číselníky.

V tabulce 5.2 jsou uvedeny všechny číselníky, potřebné pro charakteristiky s kmenovými texty, atributy nebo hierarchiemi. Oblasti databáze nejsou nijak pevně oddělené, tvoří pouze tematické seskupování a z toho důvodu u nich dochází k občasnému překrývání.

5.3 Identifikace dimenzí a faktů

V případě našeho výběru údajů do datové kostky je jejich rozdělení na dimenze a fakta jednoduché. Selektce dat neobsahuje žádné numerické měrné jednotky, jedná se vždy o diskrétní hodnotu. Proto z databáze extrahujeme pouze dimenze.

Zároveň platí, že jsou všechny informace závislé na hodnotě identifikátoru přihlášky, která tvoří primární klíč našich dat. Faktem bude počítadlo přihlášek (záznamů), vytvořené přiřazením konstantní hodnoty 1 ke každému záznamu. Další číselné hodnoty, popisující například počty přijatých či zapsaných uchazečů, získáme filtrací pomocí příznaků, určujících tento stav.

5.4 Volba vhodné reprezentace datové kostky

V modulu SAP BW je možné datovou kostku modelovat pomocí objektů ODS a InfoKostky. Ačkoliv je ODS více omezující a jednodušší, dá se v tomto případě použít. Jediným klíčovým polem, definujícím všechny hodnoty je v tomto případě identifikátor přihlášky, díky čemu jsme limitováni omezením počtu klíčových polí ODS.

Z tohoto důvodu je možné pro kostku s daty přijímacího řízení VUT použít obě reprezentace. Po uvážení obou variant nakonec padlo rozhodnutí na InfoKostku. I přes to, že ODS má už zabudovaný ukazatel, počítající záznamy a v InfoKostce je nutné ho modelovat. Z uživatelského hlediska by totiž byl v ODS vždy pojmenován jako počet záznamů, i když vhodnější název je počet přihlášek. ODS tedy tuto výhodu ztrácí a ukazatel se musí vytvořit i zde.

Hlavním důvodem je ovšem jednoduchá rozšiřitelnost v budoucnosti. InfoKostka je obecnější a je tedy lépe upravitelná, protože nemá tolik omezení.

5.5 Složitější charakteristiky

Většina charakteristik v datové kostce budou pouze jednoduché datové typy bez dodatečných informací reprezentující jednotlivé dimenze. Některé z nich budou ovšem obsahovat pouze identifikátory, uživatelům většinou nic neříkající. K těmto identifikátorům se ale z použitých číselníků (viz tabulka 5.2) dodají další údaje ve formě kmenových dat, atributů nebo hierarchií, za účelem zvýšení uživatelské přívětivosti.

Mezi tyto charakteristiky patří následující:

- Fakulta – značí fakultu, na kterou je podávána přihláška. Tato charakteristika obsahuje všechny typy rozšiřujících informací, kde v textech jsou celé názvy útvarů VUT, v atributech názvy profit center či pracovních úseků a v hierarchii je celá struktura útvarů VUT. InfoObjekt je již v rámci běžícího projektu organizace CVIS VUT vytvořen a dojde tedy pouze k jeho použití.
- Státní příslušnost – identifikátorem je dvojpísmenná zkratka státu. Do kmenových textů bude přidán celý název státu.

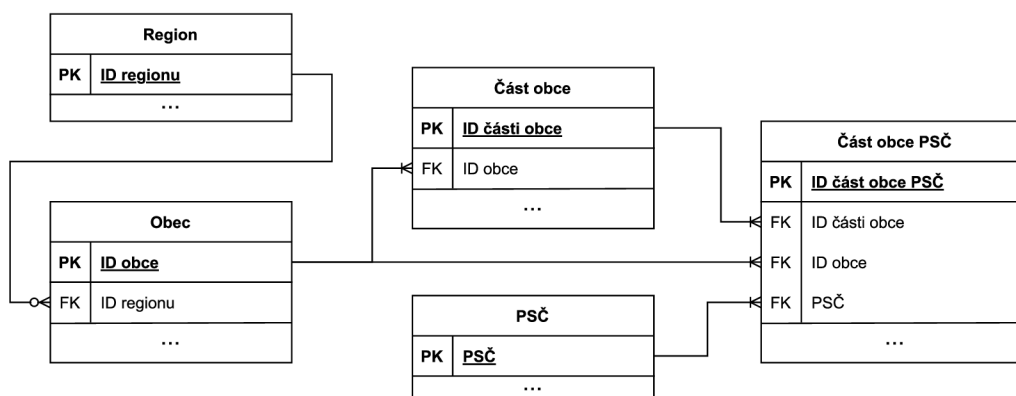
- Střední škola – zahrnuje jak zkrácený i plný název školy v kmenových textech. Do atributů je přidán IZO, rok, ve kterém byl záznam přidán, PSČ a obec, kde se škola nachází.
- Obor střední školy – v textech celý název oboru.
- Typ střední školy – data se berou z číselníku pro střední školy, kde jsou uvedeny i jednotlivé typy (gymnázium, SOŠ či jiné). Jejich IZO se skládá z devíti stejných číslic. InfoObjekt bude obsahovat pouze kmenové texty s popisem typu střední školy.
- Typ studia – charakteristika je již vytvořena, ale obsahuje pouze zkratku a popis v kmenových textech. Zároveň data nejsou brána přímo z číselníku a není tedy jisté, že zahrnují všechny možnosti. Při změně zdroje na číselník budou doplněny atributy s minimální a maximální dobou studia či požadovanou kvalifikací.
- Forma studia – obdoba typu studia. Záměna zdroje dat, přičemž v kmenových textech zůstává zkratka a popis.
- InfoObjekty geografických celků a PSČ – všechny tyto InfoObjekty mají společnou vlastnost, určují geografickou pozici. Pro jednodušší filtraci podle geografických dimenzí bude vytvořena hierarchie od kořenových států až po PSČ, která bude obsahovat všechny tyto InfoObjekty. Její návrh je blíže popsán v části 5.5.1.

InfoObjekty geografických celků zahrnují kromě hierarchie i kmenové texty a atributy. Jedná se o:

- Region – kromě hierarchie obsahuje pouze název regionu v kmenových textech. Pojmem region je myšleno vše, co popisuje číselník Region. Jedná se v hierarchii o několik úrovní, jako je stát, kraj nebo okres.
- Obec – mimo název obce, bude obsahovat i atribut kraj, ve kterém se nachází.
- Část obce – v kmenových textech název a v attributech obec, pod kterou patří.

5.5.1 Hierarchie PSČ

Cílem je vytvořit hierarchii pro umožnění filtrace uchazečů podle krajů, okresů, měst či jiných celků. Zdrojem dat jsou tabulky v oblasti adres databáze VUT, jejichž vzájemné vazby již částečně hierarchii tvoří.



Zdroj: vlastní.

Obrázek 5.3: Zjednodušené ERD oblasti adres.

Z obrázku 5.3 vyplývá, že tabulka Region obsahuje data nejvyšší úrovně. Sama o sobě ale také zahrnuje stát, oblast (Jihovýchod, Střední Morava, Praha a další), kraj a okres. Dále navazuje tabulka Obec, která se vhodně váže na Region a vytváří další stupeň pod okresy.

Problém nastává při další úrovni, kde to už není úplně jednoznačné. Tabulka Část obce PSČ je vazební tabulkou M:N vztahu mezi obcemi a PSČ, ale zároveň částí obcí a PSČ. Napojení částí obcí pod obce se ovšem jeví jako nejlepší první krok řešení.

Nyní se musí vyřešit propojení s PSČ. Ty totiž netvoří nejnižší stupeň, jedno PSČ může mít několik částí obcí, někdy i obcí. A naopak i část obce může mít více poštovních čísel. Zde se hodí vlastnost hierarchií SAP BW, kde stejný listový uzel může být pod několika odlišnými uzly. Navázání PSČ se tedy provede pod částí obce pouze pomocí vazební tabulky Část obce PSČ, která již obsahuje potřebný údaj (poštovní směrovací číslo) a tabulka PSČ proto není potřeba.

Kapitola 6

Implementace datové kostky

Po dlouhém zkoumání vstupních dat a návrhu datové kostky přichází na řadu samotná implementace v programu SAP NetWeaver BW 7.50 v transakci RSA1. Tvorba probíhá na vývojovém serveru, kde je kostka následně i testována a případné chyby se opraví. Po dokončení se všechny potřebné objekty transportují na produkční server a jsou zpřístupněny uživatelům.

V této kapitole je popsán celý vývoj datové kostky, společně s jejím zveřejněním a nastavením automatické aktualizace dat.

6.1 Databázový pohled

Prvním krokem implementace je extrakce dat ze zdroje dat. To se provádí pomocí databázového pohledu s požadovanými údaji, případně poskytnutím celých číselníků. Následně je nutné přidělit příslušná práva pro přístup z vývojového i produkčního serveru SAP BW. Celý tento proces provádí databázové oddělení organizace CVIS VUT.

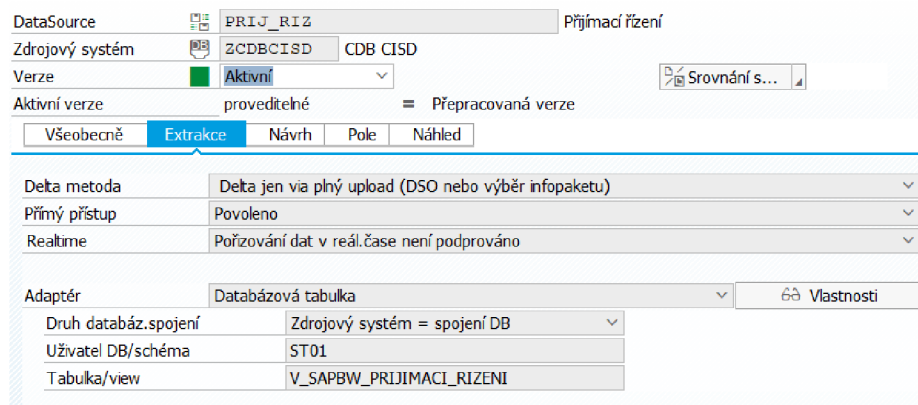
Pohled je vytvářen pomocí SQL dotazu, ve kterém jsou spojeny příslušné tabulky a z nich přidány požadované sloupce. V našem výběru dat do datové kostky figurovaly dva specifické údaje, přímo neuložené v databázi VUT. Jejich výpočet je proveden pomocí SQL v rámci databázového pohledu. Do SAP BW se tedy již nahrají a není potřeba je dále počítat. Jedná se o následující:

- Obor ISCED-F
- Věk uchazeče v době podání přihlášky – datum narození uchazeče není z databáze extrahován, proto není nutné se zabývat jeho utajením.

6.2 PSA

V SAP BW se po dokončení vývoje pohledu musí vytvořit PSA, aby se dala data nahrát do programu. Do tohoto úložiště mají vývojáři přístup pomocí možnosti administrace PSA. Díky prohlížení dat ve stavu, v jakém jsou dopravena se dá vyloučit případná chyba ve zdroji dat. Již na této úrovni je možné vyfiltrovat problematické či chybné záznamy.

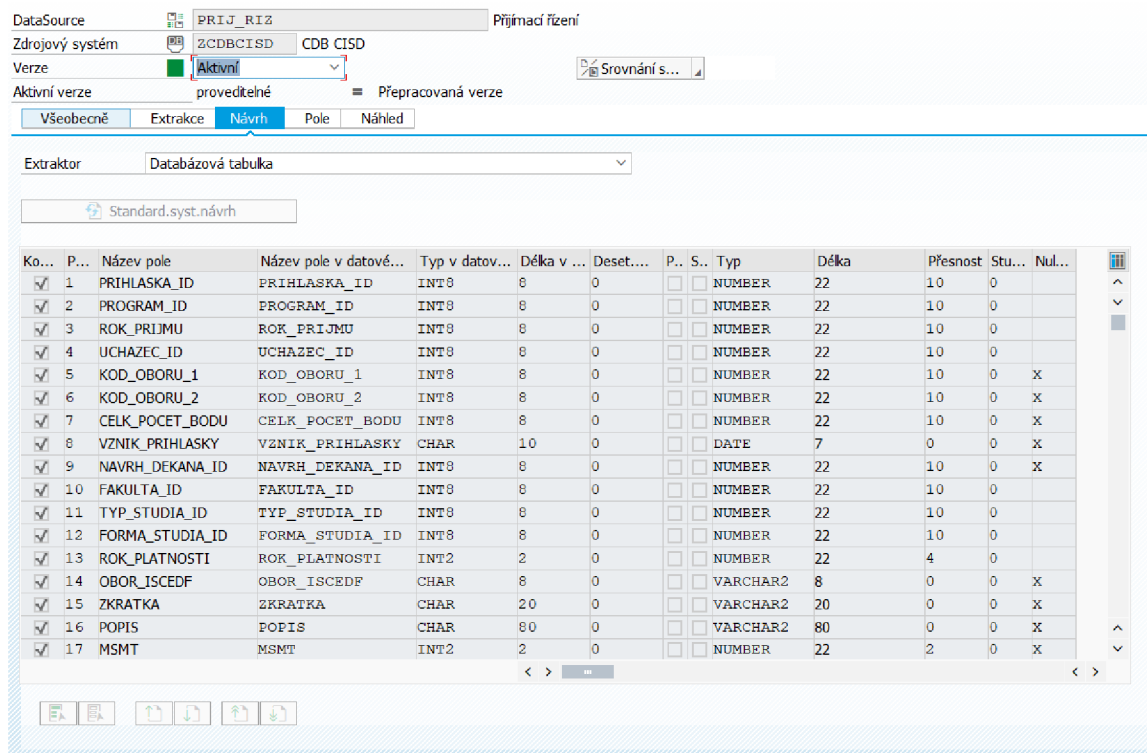
Při nastavování se nejdříve zadá druh spojení a identifikace zdroje dat. Ukázáno na obrázku 6.1.



Zdroj: vlastní.

Obrázek 6.1: PSA – nastavení zdroje dat.

Po nastavení zdroje dat jsou na základě jeho struktury navrženy pole s datovými typy SAPu pro alokaci prostoru (obrázek 6.2). Většinou není třeba návrh upravovat, ovšem častou výjimkou jsou datумы. Reprezentace datumů v databázi a SAPu nelze automaticky převést, proto se datумы v PSA reprezentují pomocí textového řetězce o délce 10 znaků.



Zdroj: vlastní.

Obrázek 6.2: PSA – návrh polí podle zdroje dat.

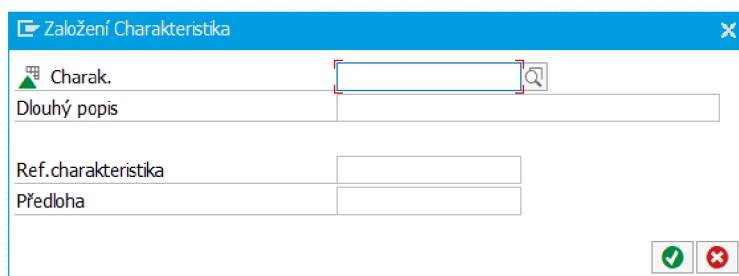
6.2.1 InfoPaket

InfoPaket je objekt v SAP BW, sloužící pro samotný přenos dat do PSA. Zde se nastavují filtry a způsob aktualizace dat. Tvorba těchto objektů je jednoduchá, téměř vše je vyplněno automaticky na základě PSA.

V rámci tohoto projektu se neprovádí žádné filtrace záznamů na této úrovni, kvůli snaze dodržet zavedené praktiky a přehlednost. Z tohoto důvodu jsou vždy přeneseny do PSA všechny údaje, což zajišťuje i základní nastavení.

6.3 InfoObjekty

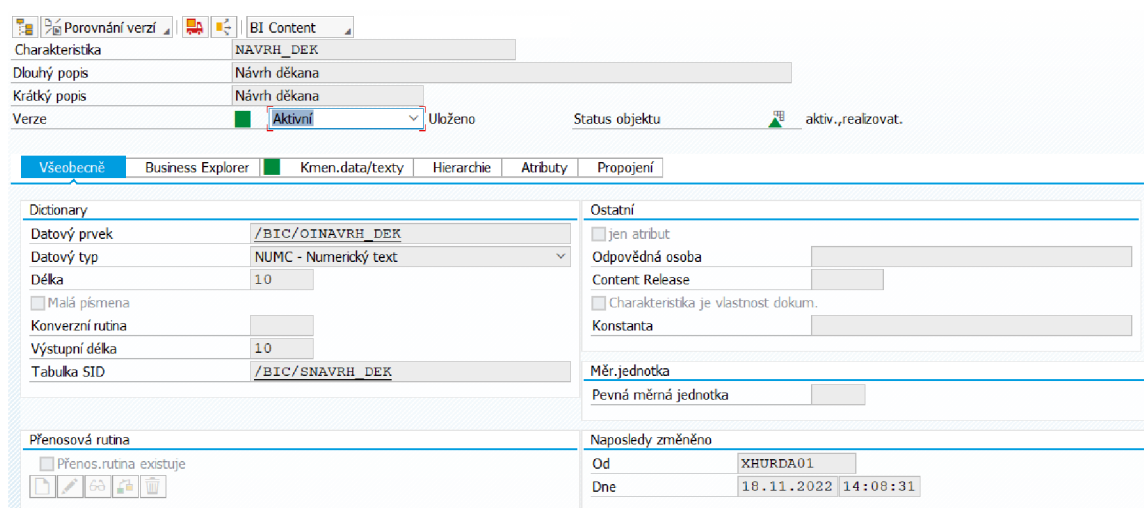
Pro každé pole databázového pohledu, které je potřeba přenést do datové kostky, je vytvořen InfoObjekt (charakteristika). InfoObjekt typu ukazatel je pouze jeden, pro počet podaných přihlášek.



Zdroj: vlastní.

Obrázek 6.3: Založení charakteristiky.

V případě použití již vytvořené charakteristiky s jiným názvem, zobrazeným uživatelem, je vytvořen referenční InfoObjekt, který obsahuje i všechny kmenové texty, atributy a hierarchie objektu referovaného.



Zdroj: vlastní.

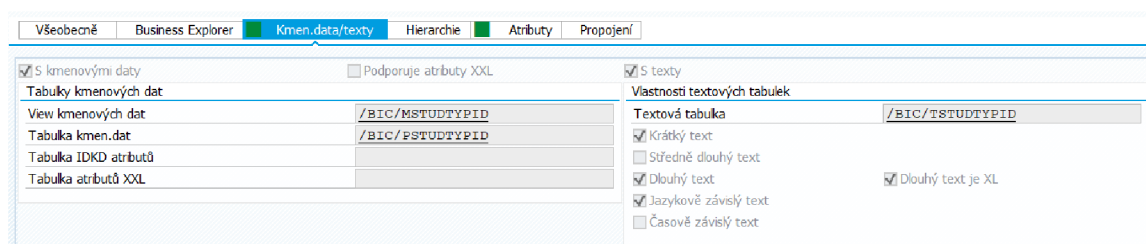
Obrázek 6.4: Základní nastavení charakteristiky.

Na obrázku 6.4 je zobrazeno všeobecné nastavení charakteristiky Návrh děkana. Mimo datový typ a délku lze vyčíst i kdo provedl poslední změny. V další záložce Business Explorer se nastavuje zejména zobrazování dat a popisů charakteristik v analytických nástrojích.

Každému typu dodatečné informace charakteristiky je také věnována záložka, ve které se specifikují jejich žádané vlastnosti. Poté je nutné vymodelovat transformaci a proces přenosu dat, obdobě jako u ostatních InfoProviderů.

6.3.1 Kmenové texty charakteristik

U kmenových textů se volí jejich délka a jazyková či časová závislost. V nastavení lze vyčíst i tabulky kmenových dat, do kterých lze nahlédnout v transakci SE16.



Zdroj: vlastní.

Obrázek 6.5: Část nastavení kmenových textů.

Velká část číselníků VUT neposkytuje jiné než české popisy (nachystané sloupce pro anglické popisy jsou prázdné), a proto pro ně zatím nejsou vytvořeny transformace.

Typ studia texty (TYP_ST_T)						
Poz	Klíč	Pole	Ik	Popis	Datov	Délka
1	TYP_STUDIA_ID	TYP_STUDIA_ID	INT8		000008	
2	POPIS	POPIS	CHAR		000200	
3	POPIS_EN	POPIS_EN	CHAR		000200	
4	ZKRATKA	ZKRATKA	CHAR		000020	
5	ZKRATKA_EN	ZKRATKA_EN	CHAR		000020	
6	POPIS_SR	POPIS_SR	CHAR		000200	
7	POPIS_SR_EN	POPIS_SR_EN	CHAR		000200	
8	POPIS_EP	POPIS_EP	STRG		000000	
9	POPIS_EP_EN	POPIS_EP_EN	STRG		000000	

Studium typ (STUDTYPID) Skup.praavidel: Standardní skupina									
Typ	Náz.prav.	Poz	Klíč	Infoobjekt	Ik	Popis	Datov	Délka	Int
	STUDTYPID	1	STUDTYPID			Studium typ	NUMC	000008	
	CS	2	OLANGU			Jazykový klíč	LANG	000001	
	OTXTSH	3	OTXTSH			Krátký popis	CHAR	000020	
	OTXTXL	4	OTXTXL			Extra dlouhý popis	SSTR	001333	

Zdroj: vlastní.

Obrázek 6.6: Transformace dat do kmenových textů.

6.3.2 Hierarchie charakteristik

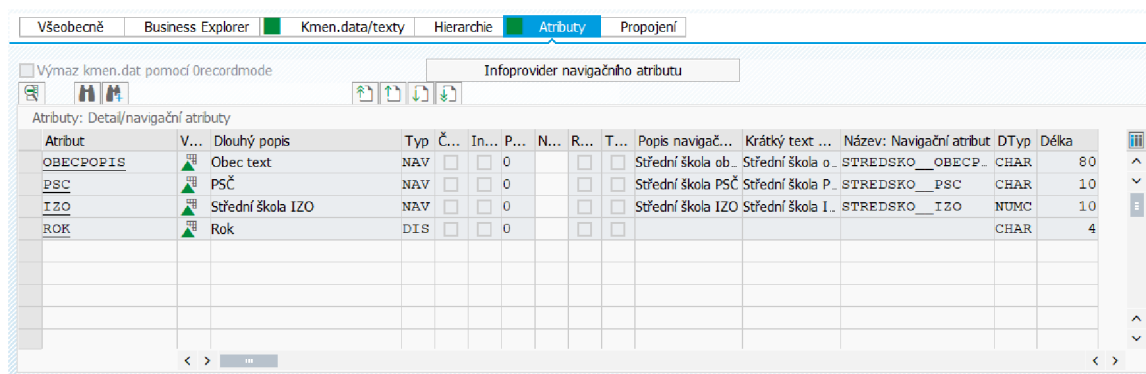
U hierarchií lze také nalézt nastavení tabulky, ve kterých je uložena. Volí se u ní i časová závislost. V případě reprezentace jednotlivých úrovní různými charakteristikami, je nutné přidat tyto externí charakteristiky do hierarchie. Tato vlastnost ovšem není povinná a je možné vytvořit hierarchii pouze s právě upravovaným InfoObjektem.

Její transformace je složitější. Skládá se z několika segmentů, popisující různé části. Jedná se například hlavičku, popisy uzlů nebo strukturu hierarchie.

U charakteristiky reprezentující fakultu VUT byla již hierarchie vytvořena a v kostce se používá její reference. Proto jedinou modelovanou hierarchií je hierarchie PSC, blíže popsanou v části 6.5.

6.3.3 Atributy charakteristik

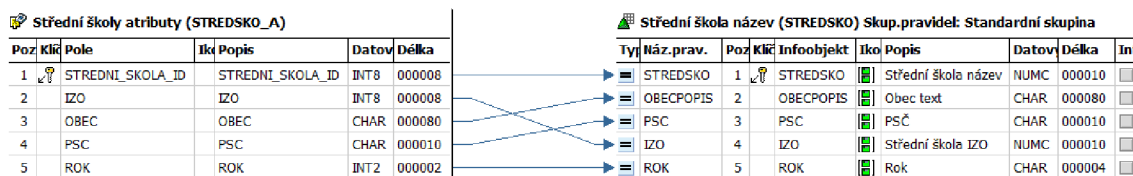
Atributy charakteristiky jsou také charakteristiky. Před přidáním mezi atributy je potřeba je vytvořit. Standardně jsou všechny přidání atributy typu DIS (zkratka anglických slov *display attributes*). V praxi to znamená, že uživatelé je vidí pouze jako doplňující informaci k charakteristice. Pokud je ale žádané podle něj i filtrovat či řadit, tedy udělat z něj v podstatě plnohodnotnou charakteristiku, musí se nastavit na navigační atribut (NAV, zkratka anglických slov *navigation attributes*). K těmto atributům se následovně doplní název, pod kterým budou v kostkách zobrazeny.



Zdroj: vlastní.

Obrázek 6.7: Nastavení atributů.

Transformace je podobná jako u kmenových textů. K identifikátoru modelované charakteristiky se dodávají údaje jednotlivým atributům (viz obrázek 6.8).



Zdroj: vlastní.

Obrázek 6.8: Transformace dat do atributů.

6.4 InfoKostka Přijímací řízení

InfoKostka Přijímací řízení je tvořena 50 InfoObjekty. Charakteristiky jsou pro lepší organizaci rozděleny do složek obecné charakteristiky, střední škola, uchazeč a přijetí (viz obrázek 6.9).

V kostce se používají i některé dostupné navigační atributy. Zajímavým je zejména Uchazeč kraj. Tento údaj není u uchazečů v databázi uveden, přičemž při vývoji přišla žádost dodání do kostky i nějaké možnosti filtrování uchazečů podle kraje, ve kterém mají trvalé bydliště. Vždy je ovšem uvedena obec a v číselníku obce je klíč příslušného kraje. Řešení tohoto problému je tedy vytvoření navigačního atributu, u charakteristiky obce trvalého bydliště uchazeče, reprezentující patřičný kraj.

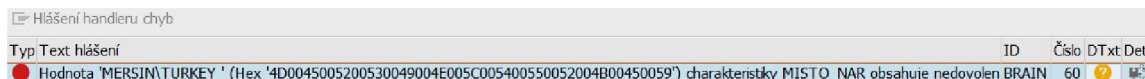
Infokostka	Techn.název / hodnota	Funk.	Z...	Název...	Dato...	D.	Typ u...	K...	N..	A...	A...	Refer.infoo...
▼ Příjímání řízení	PRIJ_RIZ											
> Informace o objektu												
> Nastavení												
▼ Dimenze												
> Dat.paket	PRIJ_RIZP											
> Čas	PRIJ_RIZT											
> Jednotka	PRIJ_RIZU											
> Obecné charakteristiky	PRIJ_RIZ1											
▼ Střední škola	PRIJ_RIZ2											
▪ Střední škola název	STREDSKO				NUMC	010						STREDSKO
▪ Střední škola typ	TYP_STR_5				CHAR	080						TYP_STR_5
▪ Střední škola IZO	IZO				NUMC	010						IZO
▪ Střední škola obor	OBOR_STRS				CHAR	020						OBOR_STRS
> Uchazeč	PRIJ_RIZ3											
> Přijetí	PRIJ_RIZ4											
▼ Navigační atributy												
▪ Studijní program kód	PROGID__STUDPROGK		<input type="checkbox"/>		CHAR	012						STUDPROGK
▪ Střední škola IZO	STREDSKO__IZO		<input type="checkbox"/>		NUMC	010						IZO
▪ Střední škola obec	STREDSKO__OBECPOPIS		<input checked="" type="checkbox"/>		CHAR	080						OBECPOPIS
▪ Střední škola PSC	STREDSKO__PSC		<input checked="" type="checkbox"/>		CHAR	010						PSC
▪ Uchazeč kraj	UCH_OBEC__KRAJ		<input checked="" type="checkbox"/>		CHAR	005						KRAJ
▼ Ukazatele												
▪ Počet přihlášek	POCPRIH				DEC	009 Počet (...)	<input checked="" type="checkbox"/> P	<input type="checkbox"/>		SUM	SUM	

Zdroj: vlastní.

Obrázek 6.9: InfoKostka Příjímání řízení.

6.4.1 Transformace

U transformace (obrázek 6.11) jsou využity všechny možnosti. Nejčastějším je přímé přiřazení, které je i nejrychlejší. Ukazatel je řešen přiřazením konstanty 1 záznamu a jeho agregace je nastavena na sumu. Častým problémem v transformaci jsou nepodporované znaky (obrázek 6.10). Ty se řeší pomocí ABAP rutiny, měnící tyto znaky na mezery.



Zdroj: vlastní.

Obrázek 6.10: Chyba – neplatný znak při transformaci.

Vzorce jsou využity při transformaci datumů do interní reprezentace SAPu, aby se zachovaly pro uživatele všechny možné operace nad datумы. Dále pro převod jednoho příznaku do správného formátu. Jako jediný totiž přichází ve formě písmen, značících ano a ne, kdežto ostatní mají již hodnoty 1 a 0.

Algoritmus 3 Vzorec pro převod datumu do formátu SAPu.

DATECONV(VZNIK_PRIHLASKY, 'DD.MM.YYYY', ”)

6.4.2 Proces přenosu dat

Proces přenosu dat je vázán k dané transformaci. Může se zde provádět filtrace údajů z PSA, které se mají nahrát do InfoKostky. Toho bylo využíváno při debugování transformace, aby

Přijímací řízení (PRJ_RIZ)		Přijímací řízení (PRJ_RIZ) skup.převídel. standardní skupina	
Posř.čís.Pole	Popis	Datový Dělka	Typ.Náz.přev.
1	PŘIJLASKA_ID	INT8 000008	1 PŘIJLASKA
2	PROGRAM_ID	INT8 000008	2 PROGID
3	ROK_PŘIJMU	INT8 000008	3 OBORID
4	UCHAZEČ_ID	INT8 000008	4 DATECONV[VZNIK_PRI]
5	KOD_OBORU_1	INT8 000008	5 VZNIK_FR
6	KOD_OBORU_2	INT8 000008	6 STUDYPRD
7	CELK_POČET_BODU	INT8 000008	7 STUFORID
8	VZNIK_PŘIJLASKY	CHAR 000010	8 ROK_FLAT
9	NAVHR_DEKANA_ID	INT8 000008	9 OBSCEDF
10	FAKULTA_ID	INT8 000008	10 POCFRH
11	TP_STUDIA_ID	INT8 000008	11 NOVY_ST
12	FORMA_STUDIA_ID	INT8 000008	12 ROK_MAT
13	ROK_PLATNOSTI	INT2 000002	13 STREDSKO
14	OBOR_ISCEDF	CHAR 000008	14 TYP_STR_5
15	ZKRATKA	CHAR 000020	15 OBOR_STRS
16	POPS	CHAR 000080	16 UCHAZEC
17	MSMT	INT2 000002	17 ROK_PRIJM
18	NEPRIJAT_DO_SPLN	INT2 000002	18 CELK_Z_B
19	PRDIATO_CELUM	INT2 000002	19 NAVHR_DEK
20	PRDIAT_BEZ_FZ	INT2 000002	20 NPLR_DSPP
21	PRDIAT_PO_FZ	INT2 000002	21 PRL_CELK
22	NEPRIJAT_KAPACITA	INT2 000002	22 PRL_BEZPZ
23	NEPRIJAT_BEZ_NED	INT2 000002	23 PRL_POFZ
24	NEPRIJAT_MSMT	INT2 000002	24 NPLR_KAP
25	PRDIAT_DODATECNE	INT2 000002	25 PRL_DODAT
26	MSMT_POČET_ZVCASTHEN	INT2 000002	26 MSMT_POČZ
27	MSMT_POČET_SPLN	INT2 000002	27 MSMT_POCS
28	MSMT_POČET_NESPL	INT2 000002	28 MSMT_POCH
29	MSMT_PČT_PRIJ	INT2 000002	29 MSMT_POCP
30	MSMT_PČT_PRIJCEL	INT2 000002	30 MSMT_PRIC
31	MSMT_ZAPS	INT2 000002	31 MSMT_ZAPS
32	MSMT_FINAL	INT2 000002	32 MSMT_FIN
33	KA_CITY_CODE	CHAR 000010	33 UCH_POH
34	KA_PART_CITY_CODE	CHAR 000010	34 UCH_OBEC
35	POHLAVI	CHAR 000020	35 UCH_CORCE
36	MISTO_MAROZENI	CHAR 000080	36 UCH_SPRS
37	STATN_PŘISLUŠNO	CHAR 000002	37 MSMT_NPRI
38	NOVY_STUDENT	CHAR 000001	38 NPLR_BNED
39	STREDNÍ_SKOLA_ID	INT8 000008	39 UCH_ULICE
40	ROK_MAT_ZKOUSKY	INT2 000002	40 UCH_ULCTS
41	TYP_STREDNÍ_SKOL	CHAR 000080	41 UCH_KOZM
42	IZO	INT8 000008	42 UCH_VEK
43	MAZEV_OBORU_SS	CHAR 000020	43 UCH_MMAR
44	UCHAZEČ_VEK	INT2 000005	44 UCH_PSC
45	TP_CITY_CODE	CHAR 000010	45 STUD_ID
46	TP_PART_CITY_CODE	CHAR 000010	46 UCH_REG
47	TP_STREET_NAME	CHAR 000050	47 ZAPSAN
48	TP_STREET_NUMBER	CHAR 000020	48 UCH_O_TXT
49	TP_POSTAL_CODE	CHAR 000016	49 UCH_COTXT
50	TP_REGION_CODE	CHAR 000010	50 FAKSOUC
51	TP_COUNTRY_CODE	CHAR 000002	
52	TP_OBEC	CHAR 000080	
53	TP_CAST_OBCE	CHAR 000080	
54	STUDIUM_ID	INT8 000008	
55	ZAPSAN	INT2 000002	

Posř.čís.Pole	Infobojekt	Infobojekt	Datový Dělka	Inte
1	PŘIJLASKA	Přihláška	NUMC 000010	
2	PROGID	Studijní program ID	NUMC 000008	
3	OBORID	Studium obor	NUMC 000008	
4	DATECONV[VZNIK_PRI]	Vznik přihlášky	DATS 000008	
5	VZNIK_FR	Studium typ	NUMC 000008	
6	STUDYPRD	Studium forma	NUMC 000008	
7	STUFORID	Rok platnosti	CHAR 000004	
8	ROK_FLAT	Rok platnosti	CHAR 000004	
9	OBSCEDF	Obor ISCEDF	CHAR 000008	
10	POCFRH	Počet přihlášek	DEC 000017	
11	NOVY_ST	Nový student	CHAR 000001	
12	ROK_MAT	Rok maturitní zkoušky	CHAR 000004	
13	STREDSKO	Střední škola název	NUMC 000010	
14	TYP_STR_5	Střední škola typ	CHAR 000080	
15	OBOR_STRS	Střední škola IZO	NUMC 000010	
16	UCHAZEC	Střední škola obor	CHAR 000020	
17	ROK_PRIJM	Uchazeč ID	NUMC 000010	
18	CELK_Z_B	Celkový počet bodů	NUMC 000010	
19	NAVHR_DEK	Návrh děkana	NUMC 000010	
20	NPLR_DSPP	Nepřijat do splnění požadavků	CHAR 000001	
21	PRL_CELK	Přijato celkem	CHAR 000001	
22	PRL_BEZPZ	Přijato bez přijímací zkoušky	CHAR 000001	
23	PRL_POFZ	Přijato po přijímací zkoušce	CHAR 000001	
24	NPLR_KAP	Nepřijat kapacita	CHAR 000001	
25	PRL_DODAT	Přijetí dodatečně	CHAR 000001	
26	MSMT_POČZ	MSMT počet zúčastněných	CHAR 000001	
27	MSMT_POCS	MSMT počet úspěšných	CHAR 000001	
28	MSMT_POCH	MSMT počet nesplnil	CHAR 000001	
29	MSMT_POCP	MSMT počet přijatých	CHAR 000001	
30	MSMT_PRIC	MSMT počet přijatých celkem	CHAR 000001	
31	MSMT_ZAPS	MSMT zapsán	CHAR 000001	
32	MSMT_FIN	MSMT finální stav	CHAR 000001	
33	UCH_POH	Uchazeč pohlíví	CHAR 000010	
34	UCH_OBEC	Uchazeč obec kód	CHAR 000010	
35	UCH_CORCE	Uchazeč část obce kód	CHAR 000010	
36	UCH_SPRS	Uchazeč státní příslušnost	CHAR 000003	
37	MSMT_NPRI	MSMT nepřijat	CHAR 000001	
38	NPLR_BNED	Nepřijat bez nedostavených	CHAR 000001	
39	UCH_ULICE	Uchazeč ulice	CHAR 000020	
40	UCH_ULCTS	Uchazeč ulice číslo	CHAR 000020	
41	UCH_KOZM	Uchazeč stát bydliště	CHAR 000003	
42	UCH_VEK	Uchazeč věk	NUMC 000003	
43	UCH_MMAR	Uchazeč místo narození	CHAR 000080	
44	UCH_PSC	Uchazeč PSC	CHAR 000016	
45	STUD_ID	Studium ID	NUMC 000008	
46	UCH_REG	Uchazeč region	CHAR 000010	
47	ZAPSAN	Zapsal se	CHAR 000001	
48	UCH_O_TXT	Uchazeč obec text	CHAR 000080	
49	UCH_COTXT	Uchazeč část obce text	CHAR 000080	
50	FAKSOUC	Fakulta / součást	NUMC 000008	

Zdroj: vlastní.

Obrázek 6.11: Transformace dat do InfoKostky Přijímací řízení.

se nemuselo vždy čekat na celý proces zpracování všech dat. Nyní se již ale zavádí do kostky všechna data.

Dále je zde nastavené paralelní zpracování s velikostí paketů 50 000 záznamů, což se zdá z hlediska rychlosti zpracování jako optimální. Žádné sémantické skupiny v tomto případě nejsou potřeba, protože není nutné, aby se nahraovala data po skupinách se společnými vlastnostmi (nejso prováděny výpočty z více záznamů).

6.5 Hierarchie PSČ

Hierarchie PSČ je asi nejsložitější částí této práce. Zdrojem je několik číselníků, přičemž každý obsahuje data minimálně jedné úrovně ve výsledném stromu. Původní myšlenkou bylo vytvořit několik transformací do hierarchie, přičemž data by se nahrávala postupně podle toho, jaké úrovně jsou v daném zdroji.

Po vyzkoušení mnoha možností, jak nahrávat do stejné hierarchie z více zdrojů, se nakonec nepovedlo touto cestou uspět. Vypadá, že SAP takový přístup vůbec nepodporuje. Z toho důvodu je zvoleno řešení, které využívá dalších pomocných objektů typu ODS. Díky nim je možné sloučit data, již ve správném formátu pro hierarchii, do jediného objektu a ten následně použít jako zdroj dat pro hierarchii.

ODS jako typ pomocných objektů byl zvolen kvůli jeho jednoduché reprezentaci ve formě tabulky a možnosti navolit klíčová pole. Vytvořené objekty jsou dva, jedná se o následující:

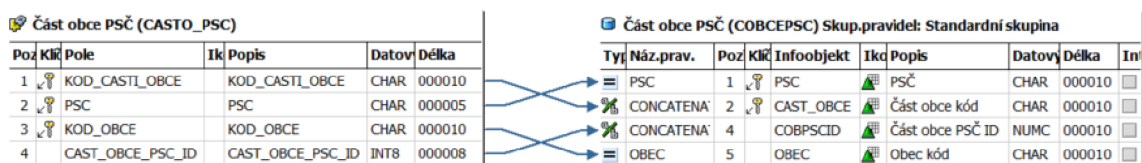
- Část obce PSČ
- Hierarchie zdroj dat

6.5.1 ODS Část obce PSČ

Data z číselníku Část obce PSČ (viz 5.5.1) slouží pro vytvoření nejnižší úrovně hierarchie. Jako primární klíč mají vytvořený identifikátor. Kandidátním klíčem je ovšem i trojice PSČ, spolu s identifikátory obce a části obce, z čehož lze vyvodit, že dvojice PSČ a identifikátor části obce se v tabulce opakuje. Pro tvorbu hierarchie SAP ovšem neumožňuje duplicitní záznamy.

ODS Část obce PSČ má za úkol převést data do formátu, ve kterém tvoří primární klíč pouze PSČ a identifikátor části obce. Díky možnosti definovat klíčová pole toho není problém dosáhnout. Duplicitní záznamy z pohledu ODS jsou sjednoceny a neklíčová pole sumarizovány podle pravidel SAPu.

Výsledná data v ODS obsahují pouze jedinečné dvojice hodnot PSČ a identifikátoru části obce. Původní identifikátor celého číselníku zůstává stále jedinečný. Přiřazená obec není pro výsledek důležitá.



Zdroj: vlastní.

Obrázek 6.12: Transformace do ODS Část obce PSČ.

6.5.2 ODS Hierarchie zdroj dat

ODS Hierarchie zdroj dat tvoří dočasné úložiště pro všechna data hierarchie, již ve finálním stavu. Z každého číselníku je vytvořena samostatná transformace s procesem přenosu dat. Ta převádí a filtruje data podle konkrétní potřeby daného číselníku.

▼ Hierarchie PSČ/část obce zdroj	HIER_M	=	Administrace
> RSDS COBCEPSC -> ODSO HIER_M	0BLLRZUNVG1OZDVJR7...	=	Změna
> RSDS CAST_OBCE ZCDBCISD -> ODSO HIER_M	0M707F170P2Y8PPCW...	=	Změna
> RSDS OBEC_H ZCDBCISD -> ODSO HIER_M	0FLH403JEEWBFTPX3K...	=	Změna
> RSDS REGION_H ZCDBCISD -> ODSO HIER_M	0CZZSVKOBXFIVLNM...	=	Změna
▼ Procesy přenosu dat	HIER_M		Založení procesu přenosu dat...
> CAST_OBCE / ZCDBCISD -> HIER_M	DTP_0002TIO35SXS17...		Změna
> OBCEPSC -> HIER_M	DTP_0002TIO35SXQB...		Změna
> OBEC_H / ZCDBCISD -> HIER_M	DTP_0002TIO35SXS17...		Změna
> REGION_H / ZCDBCISD -> HIER_M	DTP_0002TIO35SXS17...		Změna

Zdroj: vlastní.

Obrázek 6.13: ODS s více transformacemi a procesy přenosu dat.

Mimo změny z technických důvodů (popsané v 6.5.3) dochází k úpravám a filtraci údajů z tabulky region, kde se přeskakují a ignorují úrovně vedené v databázi VUT. Jedná se o odstranění duplikovaného státu a oblasti země (Jihovýchod, Střední Morava, Praha a další). Záznamy z této tabulky jako jediné neobsahují přímý odkaz na otcovský uzel, hierarchie je implicitně dána kódem regionu, kdy kód otce je kódem syna bez posledního znaku (viz obrázek 6.14). Přeskočení dvou úrovní je tedy provedeno odebráním tří znaků místo jednoho.

Status	Dat.paket	Dat.záznam	REGION_C	REGION_NAME	REGION_NAM	UPD_TS	UPD_UID	REGION_O	STATUS	NUTS_CLA
1	1	1	CZ	Česká republika	Czech Republic	01.01.1989	0		9	0
1	93	93	CZ0	Česká republika	Czech Republic	10.04.2002	999.997		9	1
1	94	94	CZ01	Praha	Prague	10.04.2002	999.997		9	2
1	95	95	CZ011	Hlavní město Praha		31.07.2007	88.520	3100	1	3
1	2	2	CZ0111	Praha 1		09.12.2014	88.520	3100	1	4
1	3	3	CZ0112	Praha 2		09.12.2014	88.520	3100	1	4
1	4	4	CZ0113	Praha 3		09.12.2014	88.520	3100	1	4
1	5	5	CZ0114	Praha 4		09.12.2014	88.520	3100	1	4
1	6	6	CZ0115	Praha 5		09.12.2014	88.520	3100	1	4
1	7	7	CZ0116	Praha 6		09.12.2014	88.520	3100	1	4
1	8	8	CZ0117	Praha 7		09.12.2014	88.520	3100	1	4
1	9	9	CZ0118	Praha 8		09.12.2014	88.520	3100	1	4
1	10	10	CZ0119	Praha 9		09.12.2014	88.520	3100	1	4
1	11	11	CZ011A	Praha 10		09.12.2014	88.520	3100	1	4
1	12	12	CZ011B	Praha 11		09.12.2014	88.520	3100	1	4
1	13	13	CZ011C	Praha 12		09.12.2014	88.520	3100	1	4
1	14	14	CZ011D	Praha 13		09.12.2014	88.520	3100	1	4
1	15	15	CZ011E	Praha 14		09.12.2014	88.520	3100	1	4
1	16	16	CZ011F	Praha 15		09.12.2014	88.520	3100	1	4
1	96	96	CZ02	Střední Čechy		10.04.2002	999.997		9	2
1	97	97	CZ021	Středočeský kraj		31.07.2007	88.520		1	3
1	17	17	CZ0211	Benešov		31.07.2007	88.520	3201	1	4
1	18	18	CZ0212	Beroun		31.07.2007	88.520	3202	1	4
1	19	19	CZ0213	Kladno		31.07.2007	88.520	3203	1	4
1	20	20	CZ0214	Kač...		31.07.2007	88.520	3204	1	4

Zdroj: vlastní.

Obrázek 6.14: Ukázka dat číselníku region.

ODS se skládá jak z InfoObjektů jednotlivých úrovní, tak z InfoObjektů reprezentujících samotnou strukturu hierarchie s datovými typy, které jsou ekvivalentní typům použitých organizací SAP při popisu hierarchií.

6.5.3 Problémy identifikátorů uzlů hierarchie

Uzly v hierarchii musí být jednoznačně určeny pomocí identifikátoru uzlu. Na to jsou využívány přímo identifikátory z jednotlivých číselníků, což vede k prvnímu problému. Primární klíč musí být jedinečný v rámci dané tabulky, ne mezi několika tabulkami. I v rámci použitých dat docházelo ke kolizím identifikátorů uzlů stromu. Problém je vyřešen přidáním prefixu před primární klíče číselníků pomocí transformací.

Druhým problémem je samotný datový typ identifikátoru uzlu hierarchie. Jedná se o numerický text, přičemž primární klíč číselníku Region je složen z alfanumerických znaků. Z tohoto důvodu je nutné provést převod písmen na číselný kód.

Nejdříve jsou nahrazeny prefixy států. V číselníku jsou vedeny pouze státy Česká a Slovenská republika, ostatní jsou vedeny pod jednou souhrnou hodnotou. Změny části kódu jsou následující:

- CZ → 01
- SK → 02
- OTHER → 03

Dále pokud je v rámci struktury regionů více synovských uzlů jak 9, je využito hexadecimálního značení. Znaky jsou převedeny na jejich dekadickou hodnotu.

Algoritmus 4 Část ABAP rutiny pro změnu kódu regionu.

DATA: source TYPE string,

 CZ_prefix TYPE string VALUE 'CZ',

 CZ_new_prefix TYPE string VALUE '01',

 SK_prefix TYPE string VALUE 'SK',

 SK_new_prefix TYPE string VALUE '02',

 OTHER_prefix TYPE string VALUE 'OTHER',

 OTHER_new_prefix TYPE string VALUE '03'.

MOVE SOURCE_FIELDS-REGION_CODE TO source.

REPLACE CZ_prefix WITH CZ_new_prefix INTO source.

REPLACE SK_prefix WITH SK_new_prefix INTO source.

REPLACE OTHER_prefix WITH OTHER_new_prefix INTO source.

"Replacing alpha characters

REPLACE 'A' WITH '10' INTO source.

REPLACE 'B' WITH '11' INTO source.

REPLACE 'C' WITH '12' INTO source.

REPLACE 'D' WITH '13' INTO source.

REPLACE 'E' WITH '14' INTO source.

REPLACE 'F' WITH '15' INTO source.

RESULT = source.

6.5.4 Transformace

Transformace do hierarchie se skládá z několika segmentů. Až na segment řešící strukturu stromu, se jedná pouze o textové popisy (uzlů, hlavičky či úrovní). Název celé hierarchie v hlavičce je dodán pomocí konstanty, ostatní části jsou přidány automaticky, díky použití vhodných InfoObjektů s texty.

Samotná struktura se bere z ODS Hierarchie zdroj dat, která obsahuje data již ve správném formátu a dojde tedy pouze k přímému přiřazení.

Hierarchie PSČ/část obce zdroj (HIER_M)					Struktura hierarchie PSČ (PSC) Skup.praavidel: Standardní skupina											
Pozí	Klíč	Infoobjekt	Ikó	Popis	Datov	Délka	Ty	Náz.prav.	Pozí	Klíč	Infoobjekt	Ikó	Popis	Datov	Délka	Int
1		NODEID	▲	Technická ID uzlu	NUMC	000008	▶	OBJECTID	1	▶	OBJECTID	▲	Identifikace objektu v segmentech	NUMC	000010	
2		IOBJNM	▲	Technický náz. charakteristiky	CHAR	000030	▶	OH_NODEID	2	▶	OH_NODEID	▲	Technická ID uzlu v hierarchii	NUMC	000008	
3		PARENTID	▲	Technická ID nadřazeného uzlu	NUMC	000008	▶	OH_IDBJNM	3		OH_IDBJNM	▲	Charakteristiky pro hierarchie	CHAR	000030	
4		NODENAME	▲	Název uzlu	CHAR	000032	▶	OH_PARENTID	4		OH_PARENTID	▲	Technický ID nadřazeného uzlu hier	NUMC	000008	
5		ORECORDMODE	▲	BW metoda delta: Režim aktualizace	CHAR	000001	▶	OH_CHILDDID	5		OH_CHILDDID	▲	Technický ID prvního podřízeného	NUMC	000008	
6		IOBJID	▲	ID infoobjektu	CHAR	000010	▶	OH_NEXTID	6		OH_NEXTID	▲	Technická ID sousedního uzlu hier	NUMC	000008	
7		REGION	▲	Region kód	CHAR	000010	▶	OH_TLEVEL	7		OH_TLEVEL	▲	Úroveň uzlu hierarchie	NUMC	000002	
8		OBEC	▲	Obec kód	CHAR	000010	▶	OH_LINK	8		OH_LINK	▲	Znak propojení uzlu hierarchie	CHAR	000001	
9		CAST_OBCE	▲	Část obce kód	CHAR	000010	▶	OH_HIERNODE	9		OH_HIERNODE	▲	Název textového uzlu	CHAR	000032	
10		PSC	▲	PSČ	CHAR	000010	▶	CAST_OBCE	10		CAST_OBCE	▲	Část obce kód	CHAR	000010	
							▶	OBEC	11		OBEC	▲	Obec kód	CHAR	000010	
							▶	PSC	12		PSC	▲	PSČ	CHAR	000010	
							▶	REGION	13		REGION	▲	Region kód	CHAR	000010	

Zdroj: vlastní.

Obrázek 6.15: Transformace struktury hierarchie.

6.5.5 Výsledná hierarchie

Vlivem odstranění dvou úrovní jsou některé kraje zdvojené, protože byly zároveň pod více oblastmi a jsou tedy uvedeny s více identifikátory. Tento stav ovšem není omezující, neboť v analytických nástrojích se zobrazují při rozboru kostek (obsahujících tuto hierarchii) pouze uzly, na které jsou navázány nějaké záznamy. Standardně se používá jeden ze dvou vzniklých podstromů a druhý se tedy implicitně nezobrazí.

STATY	Infoobjekt	Název uzlu	Z.
▼ Česká republika	REGION	CZ	<input type="checkbox"/>
▶ Hlavní město Praha	REGION	CZ010	<input type="checkbox"/>
▶ Hlavní město Praha	REGION	CZ011	<input type="checkbox"/>
▼ Středočeský	REGION	CZ020	<input type="checkbox"/>
▼ Benešov	REGION	CZ0201	<input type="checkbox"/>
▼ Vysoký újezd	OBEC	0000513482	<input type="checkbox"/>
▼ Větrov	CAST_OBCE	0000188425	<input type="checkbox"/>
• 25744	PSC	25744	<input type="checkbox"/>
▼ Vysoký újezd	CAST_OBCE	0000188433	<input type="checkbox"/>
• 25744	PSC	25744	<input type="checkbox"/>
▼ Benešov	OBEC	0000529303	<input type="checkbox"/>
▼ Mariánovice	CAST_OBCE	0000002259	<input type="checkbox"/>
• 25601	PSC	25601	<input type="checkbox"/>
▼ Radkovice	CAST_OBCE	0000002275	<input type="checkbox"/>
• 25601	PSC	25601	<input type="checkbox"/>
▼ Pomněnice	CAST_OBCE	0000002291	<input type="checkbox"/>
• 25601	PSC	25601	<input type="checkbox"/>

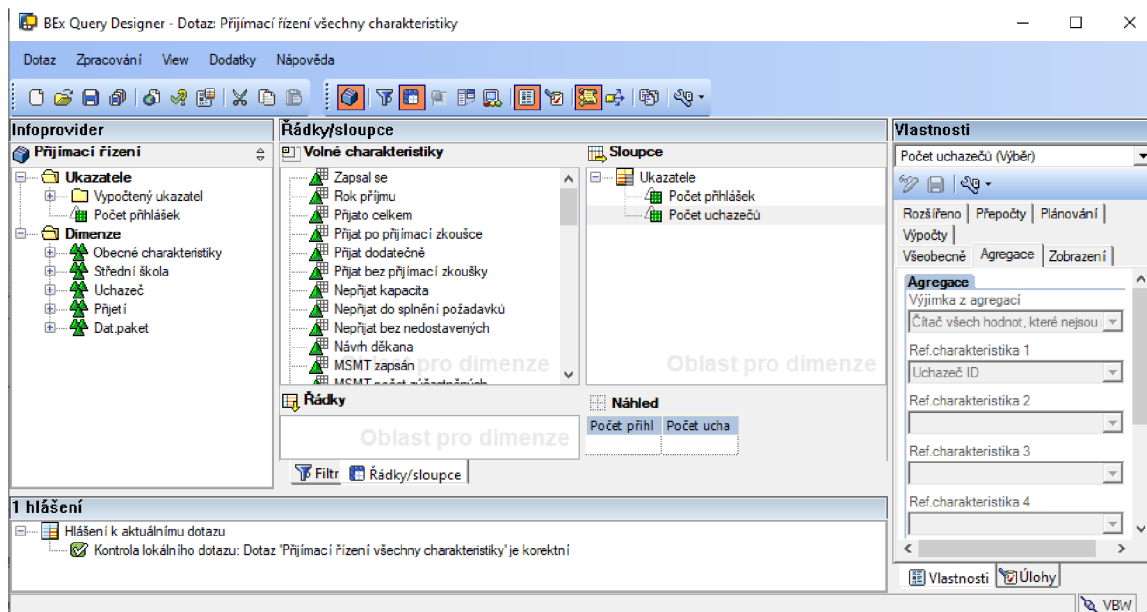
Zdroj: vlastní.

Obrázek 6.16: Zobrazení hierarchie PSČ v SAP BW.

6.6 Query

Nad InfoKostkou Příjímáací řízení je velmi užitečné vytvořit query (dotaz) pomocí SAP BEx Query Designeru. Specifické omezení charakteristik nebo filtrování záznamů nad kostkou je vlastnost, kterou uživatelé také využijí při tvorbě svých dotazů, každopádně zásadním přínosem jsou vypočtené ukazatele.

InfoKostka poskytuje pouze ukazatel počítající přihlášky, častým požadavkem je ovšem počet uchazečů, přičemž tato čísla se nemusí rovnat, protože uchazeči mohou podat více přihlášek. Vytvořit takový ukazatel jde pomocí nastavení referenční charakteristiky na identifikátor uchazeče v agregaci, čímž se vytvoří čítač všech rozdílných hodnot těchto identifikátorů, tedy uchazečů. Pro výroční zprávu bylo toto query vytvořeno analytikem VUT.



Zdroj: vlastní.

Obrázek 6.17: Query s vypočteným ukazatelem počet uchazečů.

6.7 Transport na produkci

Po ukončení vývoje dochází k přesunu na produkční server. Tento proces začíná v transakci RSA1, kde se do transportního požadavku natáhnou všechny objekty. Důležité je správně navolit režim seskupení. Na výběr je například tok dat předtím, potom, či předtím a potom. Při nevhodné volbě není možné přidat požadované objekty.

Opačným problémem je přidání objektů, které se transportovat nemají. K tomu dochází díky obrovské míře závislosti na jiných objektech. Proto SAP někdy nabídne přesun i téměř celého serveru. Při pokusu o transport této práce v jednom požadavku v něm bylo velké množství objektů. Nebylo tedy možné zkontrolovat, zda neobsahuje i nepatřičné objekty. Nakonec bylo nutné transport provádět po datových kostkách, tedy ODSkách nebo InfoKostkách.

Po tvorbě požadavku a přidání objektů se v případě žádných kolizí mezi požadavky zvolí paket pro přenos. Pokud dojde ke kolizi musí se vymazat všechny aktivní požadavky, kde jsou objekty již obsaženy.

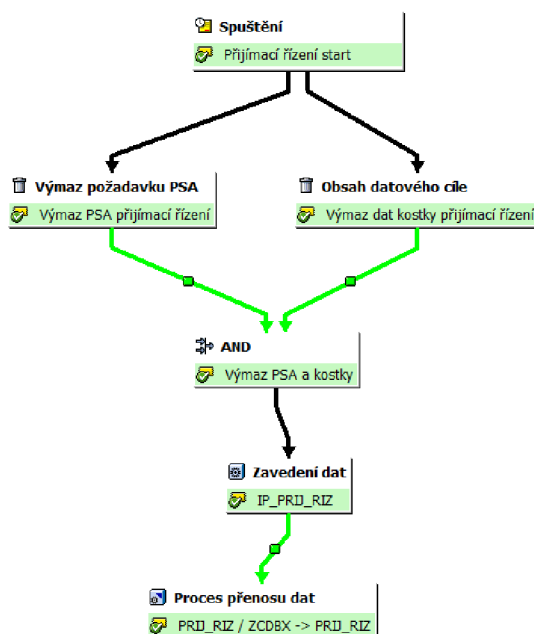
V transakci SE10 se nově vytvořený transportní požadavek uvolní, čímž je připraven k transportu. Na závěr se v transakci STMS spustí uvolněný požadavek a tím se spustí i samotný transport.

6.8 Řetězec procesů

Řetězce procesů se v projektu vytváří až na produkci. Jedná se definovatelnou posloupnost automaticky spouštěných procesů. Využití je například u aktualizace dat. Datová kostka Příjímání řízení je aktualizována na týdenní bázi, vždy v neděli.

Využití není pouze u aktualizace kostek, ale i u InfoObjektů. Řetězce procesů jsou tedy vytvořeny i pro všechny texty, atributy a hierarchie. U hierarchie PSČ je nutné v rámci řetězce provést i aktualizaci údajů v pomocných objektech typu ODS a jejich aktivaci, aby bylo možné z nich čerpat data.

Řetězec běžně začíná výmazem starých dat v PSA a kostkách, poté je provedena extrakce nových dat a transformace do InfoProviderů. Procesy je možné pouštět paralelně pro urychlení zpracování na sobě nezávislých částí.



Zdroj: vlastní.

Obrázek 6.18: Řetězec procesů aktualizující datovou kostku.

Kapitola 7

Testování datové kostky

Stejně jako každý jiný softwarový produkt se musí i datová kostka před zveřejněním otestovat, za účelem eliminace co možná nejvíce chyb. Bezchybovost u ní každopádně nelze zaručit. Vždy se počítá s určitou podobou příchozích dat, ale ve výsledku mohou dojít v libovolné podobě, kterou nelze očekávat, a tím způsobit nekonzistenci či jiné problémy.

V této kapitole je popsán přístup k testování datové kostky implementované v rámci této práce. Zároveň jsou ukázány zajímavé problémy, které byly díky testování odhaleny.

7.1 Způsob testování

První část testování přichází již v průběhu vývoje. V SAP BW je v rámci vytváření pravidel transformací možnost otestovat toto pravidlo s námi vloženými vstupními hodnotami. Při dosazení krajních či speciálních hodnot je možné vyzkoušet chování a zkontrolovat výstup u takových vstupů.

Při prvotním nahrávání dat do kostky přes transformaci často dochází k chybám vlivem neočekávaných hodnot. Celý proces zavádění dat je poté zastaven. Díky chybovým hláškám lze pravidla transformace upravit i pro tyto nově objevené případy.

Pro následující testování je již potřeba data zobrazit. Nejjednodušším způsobem je využít SAP Analysis v programu Microsoft Excel. Nejdříve se zkoumá samotná podoba dat, zda dochází k přenosu do všech charakteristik, případně zda jsou dostupné veškeré rozšiřující informace složitějších InfoObjektů.

Samotnou validitu dat poté zkoumají analytici. Pomocí datové kostky se vytváří sestavy s cílem získání stejných hodnot, jako v analýzách z minulosti. Pro tento účel se využijí i výroční zprávy z minulých let. Při rozdílných údajích se hledá příčina. Díky možnosti nahlédnout do PSA je možné rychle identifikovat, zda chybná data pochází již z databázového pohledu, nebo vznikají úpravami v SAP BW.

7.2 Největší nalezené problémy

Za pomoci testování se odhalilo hodně problémů. Veškerým chybám vzniklých nepodporovanými znaky, či transformacemi v SAP BW se není třeba věnovat, objevily se ale i takové, které indikovaly chybné záznamy v databázi VUT. Díky kostce bylo možné tyto závažnější problémy lokalizovat a případně i vyřešit.

7.2.1 Číselník Návrh příjmu

Na základě porovnávání hodnot z datové kostky analytici zjistili, že nesedí počty přijatých uchazečů. Problém se týkal zejména Fakulty informačních technologií (zkráceně FIT). Odchyly se zvyšovaly hlavně v posledních letech, kdy se jednalo už o rozdíl několika set přihlášek.

Díky upozornění analytiků na tento problém byla nalezena jeho příčina pomocí analytického systému. FIT přijímá studenty na základě výsledků SCIO testů a poté vyhodnocuje výsledek přijímacího řízení jako: *přiját na základě SCIO testu*. Tato varianta ovšem měla v číselníku Návrh příjmu chybná zdrojová data. Ve sloupci, nesoucí příznak Přijato celkem, obsahovala hodnotu, která značí, že uchazeč nebyl přijat.

	A	B	C	D
1	Fakulta ID	Návrh děkana	Přijato celkem	Počet přihlášek
2	FIT	Bude přijat po doručení maturitního vysvědčení	ne	339
3		Nepřijet dokud nesplní formální požadavky	ne	280
4		Nepřijet pro nedostačující kapacitu oboru/studijního programu	ne	5 658
5		Nepřijet pro neprospěch u přijímací zkoušky	ne	1 931
6		Nepřijet pro nesplnění formálních náležitostí ve stanoveném termínu	ne	66
7		Nepřijet pro nesplnění minimální úrovně požadavků	ne	203
8		Nepřijet pro nesplnění podmínek pro přijetí	ne	132
9		Nepřijet pro nesplnění podmínek přijímací zkoušky	ne	48
10		Nepřijet pro obsazenost tématu	ne	2
11		Přijet bez přijímací zkoušky	ano	9 489
12		Přijet dodatečně (po odvolání, po přezkumném řízení, apod.)	ano	6
13		Přijet mimo přijímací řízení	ano	40
14		Přijet na základě SCIO testu	ne	2 318
15		Přijet na základě přijímací zkoušky	ano	8 737
16		Přijet přednostně	ano	2 284
17		Přijímací řízení zastaveno	ne	3 878
18		Rozhodnutí odloženo	ne	1 208
19		Výsledek		36 619

Zdroj: vlastní.

Obrázek 7.1: Vyhodnocené přihlášky Fakulty informačních technologií.

Z obrázku 7.1 na řádku 14 je vidět, že celkově více než 2300 přihlášek má výsledek s chybně nastaveným příznakem. Při vyfiltrování pouze přijatých uchazečů pomocí tohoto příznaku se všechny tyto záznamy odeberou.

	A	B	C	D	E
1	Fakulta ID	Rok platnosti	Přijato celkem	Návrh děkana	Počet přihlášek
2	FIT	2020	ano	Přijet bez přijímací zkoušky	423
3				Přijet na základě přijímací zkoušky	29
4				Přijet přednostně	434
5				Výsledek	886
6			ne	Nepřijet pro nedostačující kapacitu oboru/studijního programu	144
7				Nepřijet pro neprospěch u přijímací zkoušky	2
8				Nepřijet pro nesplnění minimální úrovně požadavků	29
9				Nepřijet pro nesplnění podmínek pro přijetí	4
10				Přijet na základě SCIO testu	417
11				Přijímací řízení zastaveno	736
12				Výsledek	1 332
13			Výsledek		2 218

Zdroj: vlastní.

Obrázek 7.2: Vyhodnocené přihlášky Fakulty informačních technologií z roku 2020.

Z obrázku 7.2 řádku 10 lze vyčíst, že v roce 2020 byla chyba u 417 přihlášek, což tvořilo téměř třetinu všech přijatých studentů.

Po určení příčiny problému, se zkontroval celý číselník Návrh děkana. V příznaku Přijato celkem se našly i další chyby u méně používaných vyhodnocení přijímacího řízení.

	A	B	C
1	Návrh děkana	Přijato celkem	Počet přihlášek
2	Bude přijat po doručení bakalářského diplomu	ne	2
3	Bude přijat po doručení maturitního vysvědčení	ne	1 607
4	Může pokračovat v další části přijímacího řízení	ne	321
5	Nedostavil se k přijímacímu řízení	ne	784
6	Nepřijet dokud nesplní formální požadavky	ne	4 091
7	Nepřijet pro nedostačující kapacitu oboru/studijního programu	ne	22 387
8	Nepřijet pro neprospěch u přijímací zkoušky	ne	40 910
9	Nepřijet pro nesplnění formálních náležitostí ve stanoveném termínu	ne	971
10	Nepřijet pro nesplnění minimální úrovně požadavků	ne	227
11	Nepřijet pro nesplnění podmínek pro přijetí	ne	999
12	Nepřijet pro nesplnění podmínek přijímací zkoušky	ne	9 454
13	Nepřijet pro obsazenost tématu	ne	33
14	Přijet (přestup na jiný obor)	ano	2
15	Přijet bez přijímací zkoušky	ano	122 066
16	Přijet do bakalářského studia	ano	1 897
17	Přijet do profesních oborů	ne	41
18	Přijet dodatečně (po odvolání, po přezkumném řízení, apod.)	ano	1 516
19	Přijet mimo přijímací řízení	ano	1 976
20	Přijet na základě SCIO testu	ne	2 741
21	Přijet na základě přijímací zkoušky	ano	98 021
22	Přijet na základě přípravného kurzu	ne	27
23	Přijet přednostně	ano	2 297
24	Přijímací řízení zastaveno	ne	13 320
25	Rozhodnutí odloženo	ne	2 069
26	Celkový výsledek		327 759

Zdroj: vlastní.

Obrázek 7.3: Příznak Přijato celkem u všech variant výsledků přijímacího řízení.

Celkem byly chyby ve třech záznamech. Na obrázku 7.3 jsou vidět na řádcích 17, 20 a 22. Po přesné identifikaci všech chybných údajů se podal požadavek na databázové oddělení CVIS VUT, které provedlo změnu příslušných hodnot. Po aktualizaci dat se oprava přenesla i do datové kostky (viz obrázek 7.4).

	A	B	C
1	Návrh děkana	Přijato celkem	Počet přihlášek
2	Přijet do profesních oborů	ano	41
3	Přijet na základě SCIO testu	ano	2 741
4	Přijet na základě přípravného kurzu	ano	27
5	Celkový výsledek		2 809

Zdroj: vlastní.

Obrázek 7.4: Opravený příznak Přijato celkem.

7.2.2 PSČ začínající nulou

Slovenská PSČ mohou začínat nulou, což způsobilo problém kvůli automatickému nastavení charakteristik v SAP BW. Ačkoliv je u charakteristiky PSČ nastaven datový typ na textový řetězec, je u ní implicitně prováděna takzvaná Alpha konverze. Ta v případě, že je celá hodnota složena pouze z číslic, doplní zleva nuly do maximální délky a poté všechny prefixové nuly odstraní.

Účelem této konverze je namapovat všechna stejná čísla do stejné hodnoty, bez ohledu na předcházející nuly. V případě PSČ je to však chybná úprava a vznikající údaje bez počátečního prefixu nedávaly smysl. Po vypnutí automatického provádění konverze v nastavení charakteristiky se problém vyřešil a PSČ již přichází v neupraveném stavu.

7.2.3 Zahraniční PSČ

Zahraniční PSČ způsobují další problém. Česká poštovní čísla jsou při zadávání do systému vybírána z číselníku adres, kdežto zahraniční jsou vypisována ručně, což vede ke vzniku nejrůznějších hodnot uložených v databázi VUT.

	A	B	C
1	Uchazeč PSČ	Počet přihlášek	Počet uchazečů
2	_	1	1
3	XXXXX	1	1
4	X	3	3
5	UKRAJINA	1	1
6	01010_	1	1
7	000000000	2	2
8	000	5	5
9	0 13 11 ŽILINA	1	1
10	0	5	5
11	-----	4	4
12	-----	2	2
13	-	13	13
14	+23321	1	1
15	+211	1	1

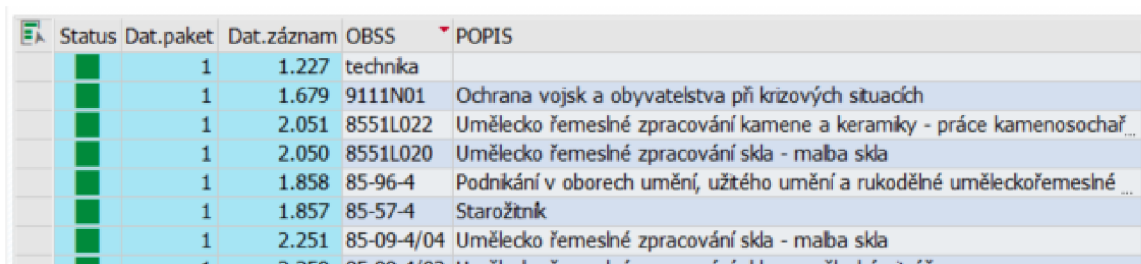
Zdroj: vlastní.

Obrázek 7.5: Ukázka některých chybně zadaných PSČ.

Byly navrhovány nejrůznější možnosti řešení pomocí transformací na úrovni SAP BW, nakonec se ale rozhodlo nechat tyto záznamy zatím beze změny, přičemž úpravy by se měly provést již v databázi VUT. Počet těchto záznamů není velký a tak prozatím tvoří zanedbatelnou část.

7.2.4 Obor střední školy

Číselník Obor střední školy obsahuje jeden zvláštní záznam, kdy je místo identifikátoru uveden text *technika* a popis je prázdný (viz obrázek 7.6). Chyba se objevila při nahrávání kmenových textů do charakteristiky. Vzhledem k tomu, že na tento záznam neodkazuje žádná přihláška, není přenášen do datové kostky a není nikde použit.



Status	Dat.paket	Dat.záznam	OBSS	POPIS
1	1	1.227	technika	
1	1	1.679	9111N01	Ochrana vojsk a obyvatelstva při krizových situacích
1	2.051	8551L022	Umělecko řemeslné zpracování kamene a keramiky - práce kamenosochař...	
1	2.050	8551L020	Umělecko řemeslné zpracování skla - malba skla	
1	1.858	85-96-4	Podnikání v oborech umění, užitého umění a rukodělné uměleckořemeslné ...	
1	1.857	85-57-4	Starožitník	
1	2.251	85-09-4/04	Umělecko řemeslné zpracování skla - malba skla	
1	2.250	85-09-4/02	Umělecko řemeslné zpracování skla - umělecké vřezá...	

Zdroj: vlastní.

Obrázek 7.6: Chybně zadaný záznam v číselníku Obor středních škol.

Kapitola 8

Možné výstupy z datové kostky

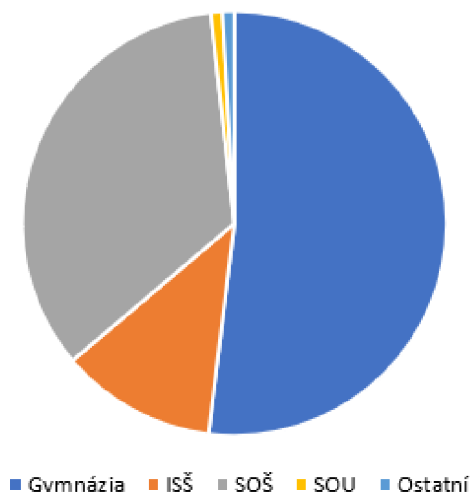
Po dokončení vývoje je datová kostka připravena k praktickému použití. Pomocí analytických nástrojů pro práci s OLAP kostkami lze rychle a efektivně získávat nové informace z již dostupných dat.

V této kapitole je ukázáno několik možných výstupů kostky Příjímací řízení. Data v kostce jsou ve všech analýzách aktualizována dne 2.4.2023. Z tohoto důvodu není v datech kompletně dořešené přijímací řízení za rok 2023 a počty se mohou ještě měnit. Analýzy a grafy jsou vytvořeny pomocí rozšíření SAP Analysis v programu Microsoft Excel.

8.1 Typ střední školy uchazečů

Složení uchazečů o studium je pro vysoké školy důležité a to jak kvůli propagaci, tak kvůli samotné náplni učiva ve studijních programech. U studentů z různých středních škol se mohou očekávat rozdílné vstupní znalosti.

Na VUT se tento údaj začal sledovat v roce 2012. Pomocí kostky je možné sledovat složení na jednotlivých fakultách, oborech i programech.

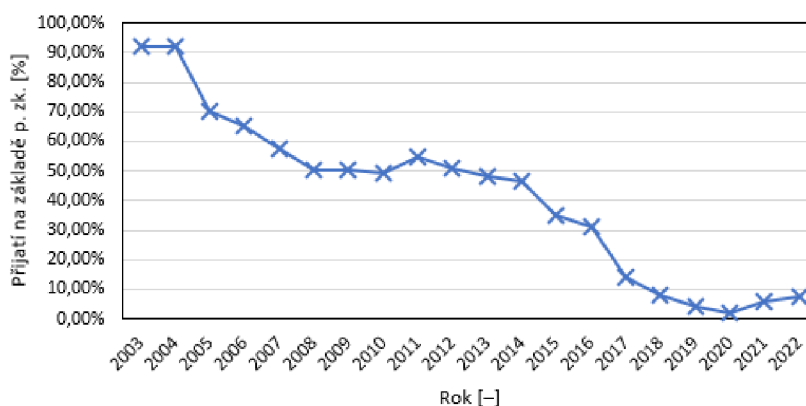


Zdroj: vlastní.

Obrázek 8.1: Střední školy uchazečů o bakalářské studium na FIT VUT od roku 2012.

8.2 Způsob přijetí uchazečů

Řada vysokých škol nabízí možnosti, jak se vyhnout psaní přijímací zkoušky. Velká část uchazečů se poté snaží splnit stanovené podmínky, aby si usnadnili přijetí na školu. Poté je možné například zkoumat, jaké procento uchazečů je přijato přednostně či na základě přijímací zkoušky.

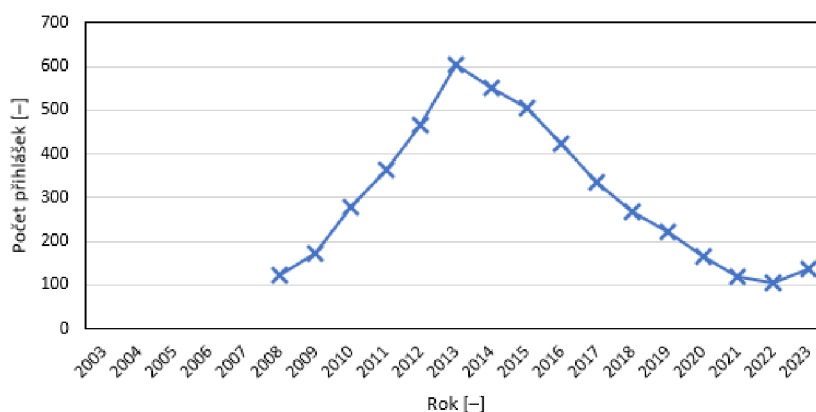


Zdroj: vlastní.

Obrázek 8.2: Procentuální vývoj uchazečů o bakalářské studium na FIT VUT přijatých na základě přijímací zkoušky.

8.3 Zájem o studium

Zájem o studium lze měřit několika způsoby. Jedním je počet podávaných přihlášek na danou fakultu, nebo studijní program. Podle těchto údajů se dá rozhodnout o smyslu určitých programů, případně o zvýšení jejich propagace.



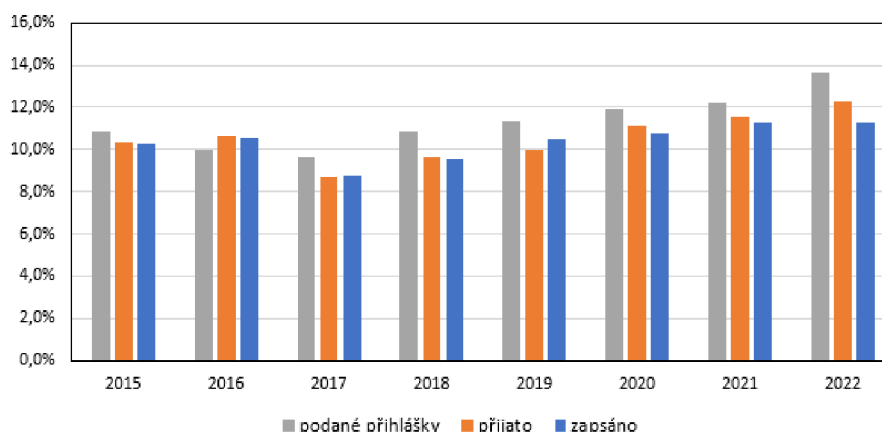
Zdroj: vlastní.

Obrázek 8.3: Počet podaných přihlášek na Ústav soudního inženýrství (ÚSI) VUT.

Z obrázku 8.3 je zřejmé, že po oddělení od rektorátu a zřízení samostatné fakulty Ústavu soudního inženýrství, rostl i zájem o studium na této škole. Od roku 2014 se počty podaných přihlášek snižují a zájem o studium upadá.

8.3.1 Zájem žen o studium na technické škole

Na technických školách je častým problémem malé zastoupení žen. Důvody mohou být různé, od mínění společnosti až po rozpaky samotných dívek, zda tam nebudou jediné. Fakulta informačních technologií se proti tomu v posledních letech snaží bojovat a prosazovat zájem o techniku u žen. Výsledky je možné také pozorovat na údajích z datové kostky.



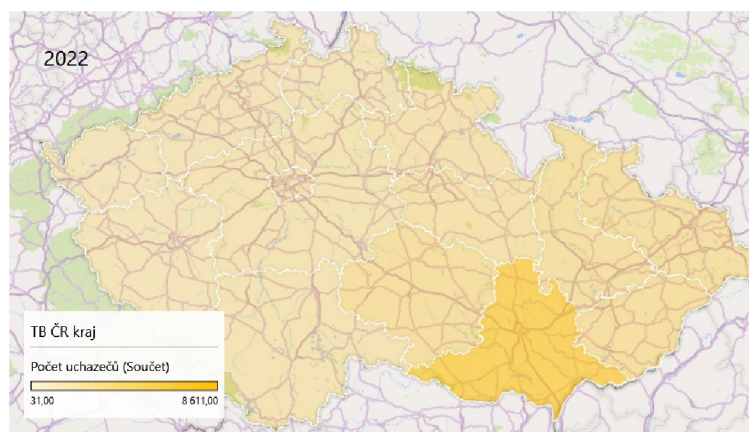
Zdroj: vlastní.

Obrázek 8.4: Procentuální zastoupení žen na FIT VUT.

Podle obrázku 8.4 se procento uchazeček na FIT VUT zvyšuje. Z dat ovšem také vyplývá, že nebývají tolik úspěšné v přijímacím řízení. Poslední tři roky si také dívky i po přijetí častěji studium rozmyslely a nezapsaly se.

8.4 Geografické analýzy

Díky informacím o trvalém bydlišti a státní příslušnosti je možné údaje o uchazečích zanášet i do map. Vznikají pak různé geografické grafy (případně i animace, ukazující časový průběh) znázorňující místa, odkud pochází budoucí studenti. Mimo obrázek 8.5 jsou další ukázky uvedeny v příloze C. Do mapy se dají zanášet i střední školy uchazečů, protože kostka obsahuje jejich lokalitu, tedy obec a PSČ.



Zdroj: vlastní.

Obrázek 8.5: Kraj trvalého bydliště českých uchazečů VUT z roku 2022.

Kapitola 9

Závěr

Cílem této práce bylo nastudovat teorii multidimenzionální analýzy a její implementaci společností SAP. Za pomoci jejich komerčních nástrojů poté navrhnout a implementovat datovou kostku OLAP pro zvolenou oblast dat z univerzity Vysokého učení technického v Brně.

Předem vytyčené cíle se podařilo splnit. Principy vícerozměrné analýzy jsou využity při návrhu a modelování datové kostky v systému SAP BW 7.50, přičemž pro kostku byla zvolena oblast dat z přijímacího řízení vysoké školy. Pro udržování aktuálních dat jsou nastaveny automatické aktualizace dat na týdenní bázi.

Ihned po dokončení začala být datová kostka používána v praxi, za účelem vyplnění částí výroční zprávy o činnosti, týkajících se přijímacího řízení. Na základě požadavků analytiků v průběhu používání byl do kostky přidán příznak Zaplacen, aby byla možnost vyfiltrovat nezaplacené přihlášky, se kterými se v analýzách často nepracuje.

Díky této práci jsem se naučil teorii OLAP analýzy a její využití v praxi, spolu s pochopením principů pokročilejšího použití databázových systémů a multidimenzionální reprezentace dat. Zároveň jsem si mohl vyzkoušet práci s komerčními nástroji, které nejsou běžně dostupné.

Datová kostka je nyní ve stavu, kdy není nutný žádný další zásadní vývoj. V budoucnu je potřeba vyřešit problém zahraničních PSČ. Změny budou ale nejspíše provedeny na úrovni databáze VUT, bez zásahu do samotné datové kostky. Dále se bude řešit přenos datové kostky na novější technologii SAP HANA, kvůli končící podpoře aktuálně použitého softwaru. Tento proces by již neměl způsobovat žádné problémy a měl by být zvládnutelný za krátký čas.

Literatura

- [1] *Směrnice děkana č. 3 / 2022 – Pravidla přijímacího řízení a podmínky pro přijetí ke studiu v bakalářských studijních programech pro akademický rok 2023/24* [online]. Brno: FEKT VUT Brno, 2022 [cit. 2023-03-02]. Dostupné z: <https://www.vut.cz/uredni-deska/vnitřni-legislativa-fekt/-d232000/pravidla-prijimaciho-rizeni-bsp-2023-24-p228965>.
- [2] ANDERSON, G. W. *Naučte se SAP za 24 hodin*. 1. vyd. Brno: Computer Press, 2012. ISBN 978-80-251-3685-0.
- [3] CAVENDISHPROFESSIONALS. *SAP and ERP: A Beginners Guide* [online]. 2020 [cit. 2023-02-11]. Dostupné z: <https://www.cavendishprofessionals.com/sap-and-erp-a-beginners-guide/>.
- [4] EGGER, N. *SAP BW data modeling*. 1st ed. Fort Lee: Galileo Press, 2005. SAP press. ISBN 1-59229-043-4.
- [5] EGGER, N. *SAP BW reporting and analysis*. 1st ed. Boston: Galileo Press, 2006. SAP press. ISBN 978-1-59229-045-1.
- [6] GALAKTIKASOFT. *OLAP Operations in Data Mining* [online]. Leden 2018 [cit. 2023-01-29]. Dostupné z: <https://galaktika-soft.com/blog/olap-operations-in-data-mining.html>.
- [7] GEEKSFORGEEKS. *OLAP Guidelines (Codd's Rule)* [online]. Srpen 2021 [cit. 2023-01-28]. Dostupné z: <https://www.geeksforgeeks.org/olap-guidelines-codds-rule/>.
- [8] HRUŠKA, T. *Pokročilé informační systémy - Online Analytical Processing & Business Intelligence*. Brno: FIT VUT Brno, 2022 [cit. 2023-01-22].
- [9] JANEČKA, P. *Použití agilních metodologií při návrhu datových skladů* [online]. Brno, 2014. [cit. 2023-01-20]. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce BARTÍK, V. Dostupné z: <https://theses.cz/id/d82fpy/>.
- [10] JAVATPOINT. *Data Warehouse Design* [online]. 2020 [cit. 2023-01-22]. Dostupné z: <https://www.javatpoint.com/data-warehouse-design>.
- [11] JHAWAR, R. a TEJADA, Z. *Extract, transform, and load (ETL)* [online]. Microsoft, 2022 [cit. 2023-01-20]. Dostupné z: <https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>.

- [12] KISS, T. *Aplikace pro zobrazování OLAP datových kostek* [online]. Brno, 2020. [cit. 2023-01-21]. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce BARTÍK, V. Dostupné z: <https://theses.cz/id/4a59oz/>.
- [13] KÜHNHAUSER, K.-H. *ABAP : výukový kurz*. Vyd. 1. Brno: Computer Press, 2009. ISBN 978-80-251-2117-7.
- [14] LABERGE, R. *Datové sklady : agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.
- [15] LACKO, L. *Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle*. 1. vyd. Brno: Computer Press, 2003. ISBN 8072269690.
- [16] MIROVÁ, E. *Výroční zpráva* [online]. WikiKnihovna, 2015 [cit. 2023-03-02]. Dostupné z: https://wiki.knihovna.cz/index.php?title=V%C3%BDro%C4%8Dn%C3%AD_zpr%C3%A1va.
- [17] MŠMT. *Výroční zprávy vysokých škol* [online]. Ministerstvo školství, mládeže a tělovýchovy, 2022 [cit. 2023-03-02]. Dostupné z: <https://www.msmt.cz/vzdelavani/vysoke-skolstvi/vyrocni-zpravy-o-cinnosti-vysokych-skol>.
- [18] POUR, J. *Business intelligence v podnikové praxi*. 1. vyd. Praha: Professional Publishing, 2012. ISBN 978-80-7431-065-2.
- [19] S PEERS. *Intuitive data processing with SAP Analysis for Office* [online]. 2022 [cit. 2023-02-17]. Dostupné z: <https://s-peers.com/en/sap-analytics/reporting-mit-sap/analysis-for-office/>.
- [20] VAŠKOVÁ, M. *Přijímací řízení na vysoké školy* [online]. Brno, 2013. [cit. 2023-03-02]. Diplomová práce. Masarykova univerzita, Právnická fakulta. Vedoucí práce SMUTNÁ, V. Dostupné z: <https://theses.cz/id/fsffjy/>.
- [21] VESELÝ, I. *Datový sklad se zaměřením na optimalizaci ETL procesu* [online]. Brno, 2011. [cit. 2023-01-22]. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce ŠEBEK, M. Dostupné z: <https://theses.cz/id/81xdh8/>.
- [22] ŠIMŮNEK, J. *Procesy ETL* [online]. Hradec Králové, 2019. [cit. 2023-01-20]. Bakalářská práce. Univerzita Hradec Králové, Fakulta informatiky a managementu. Vedoucí práce TESAŘOVÁ, B. Dostupné z: <https://theses.cz/id/vpbszo/>.

Příloha A

Obsah přiloženého paměťového média

```
/ - - . - - analysis/ - - analýzy v Microsoft Excel
|
. - - imgs/ - - vektorové obrázky ve formátu .svg
|
. - - src/ - - . - - ABAP/ - - . - - cube/ - - zdrojové kódy ABAP rutin InfoKostky
|                                     |
|                                     | Příjímací řízení
|                                     |
|                                     | . - - hierarchy/ - - zdrojové kódy ABAP rutin pomocných
|                                     | ODS objektů pro tvorbu hierarchie
|                                     |
|                                     | . - - formulas/ - - zdrojové kódy vzorců
|
. - - text/ - - . - - src/ - - zdrojový kód textu bakalářské práce
|
|                                     |
|                                     | . - - text.pdf (text bakalářské práce)
|
. - - view_instructions.xlsx (zadání pro databázové oddělení na databázový pohled)
```


Příloha B

Seznam použitých polí z databáze VUT

Tabulka	Sloupec
ST01.PRIHLASKA_NA_FAKULTU	PRIHLASKA_ID
	PROGRAM_ID
	ROK_PRIJMU
	UCHAZEC_ID
	KOD_OBORU_1
	KOD_OBORU_2
	CELK_POCET_BODU
	VZNIK_PRIHLASKY
	NAVRH_DEKANA_ID
ST01.PROGRAM	FAKULTA_ID
	TYP_STUDIA_ID
	FORMA_STUDIA_ID
	ROK_PLATNOSTI
	KOD_STUD_PROG (obor ISCEDF)
ST01.C_NAVRH_PRIJM	ZKRATKA
	POPIS
	MSMT
	NEPRIJAT_DO_SPLNENI_POZADAVKU
	PRIJATO_CELKEM
	PRIJAT_BEZ_PZ
	PRIJAT_PO_PZ
PRIJAT_DODATECNE	

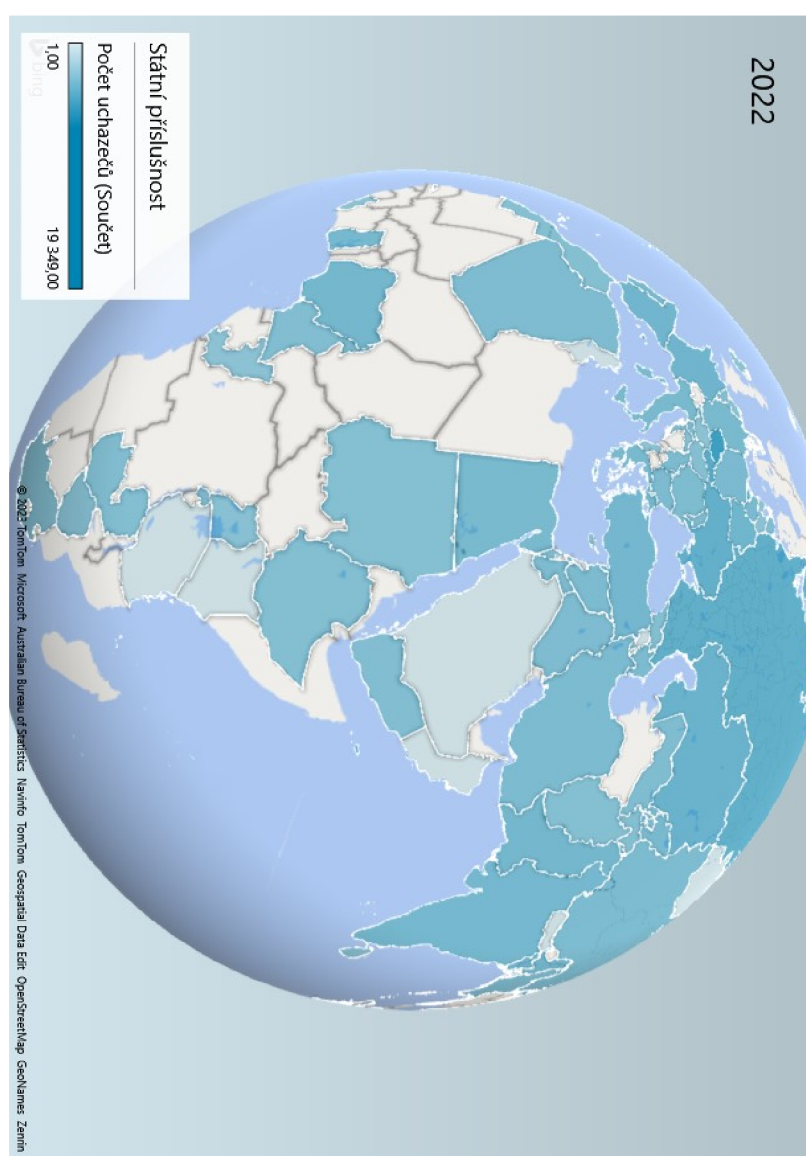
Tabulka B.1: Seznam použitých polí z databáze VUT, 1. část.

Tabulka	Sloupec
ST01.C_NAVRH_PRIJM	NEPRIJAT_KAPACITA
	NEPRIJAT_BEZ_NEDOSTAVENYCH
	NEPRIJAT_MSMT
	MSMT_POCET_ZUCASTNENYCH
	MSMT_POCET_SPLNIL
	MSMT_POCET_NESPLNIL
	MSMT_POCET_PRIJATYCH
	MSMT_POCET_PRIJATYCH_CELKEM
	MSMT_ZAPS
	MSMT_FINAL
ST01.UCHAZEC	TP_CITY_CODE
	TP_PART_CITY_CODE
	TP_COUNTRY_CODE
	TP_POSTAL_CODE
	TP_REGION_CODE
	TP_STREET_CODE
	POHLAVI
	MISTO_NAROZENI
	STATNI_PRISLUSNOST
	NOVY_STUDENT
	STREDNI_SKOLA_ID
	ROK_MAT_ZKOUSKY
	TYP_STREDNI_SKOLY
	IZO
	TYP_STREDNI_SKOLY
	NAZEV_OBORU_SS
DATUM_NAROZENI (věk uchazeče)	

Tabulka B.2: Seznam použitých polí z databáze VUT, 2. část.

Příloha C

Geografické grafy



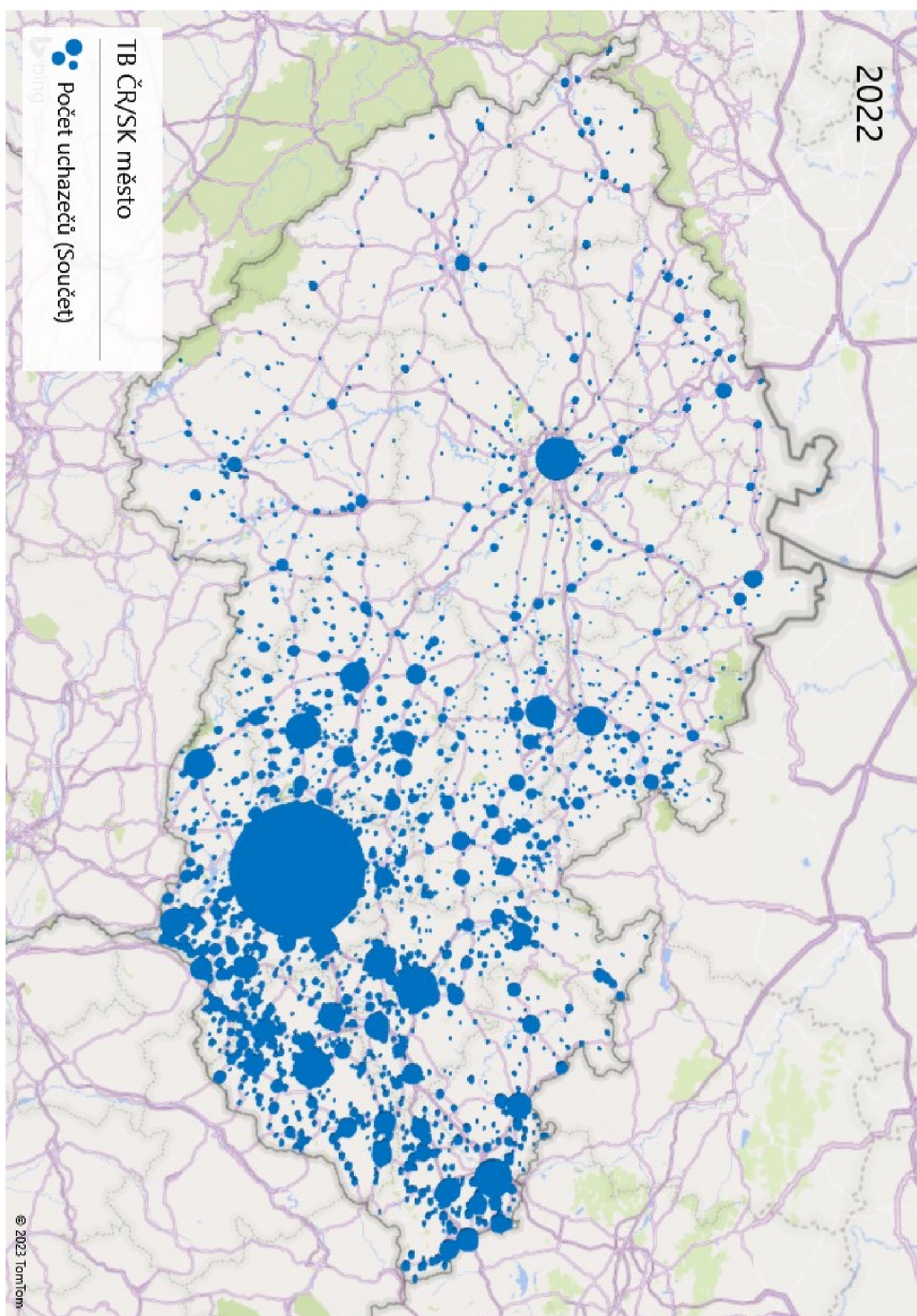
Zdroj: vlastní.

Obrázek C.1: Státní příslušnost uchazečů VUT, rok 2022.



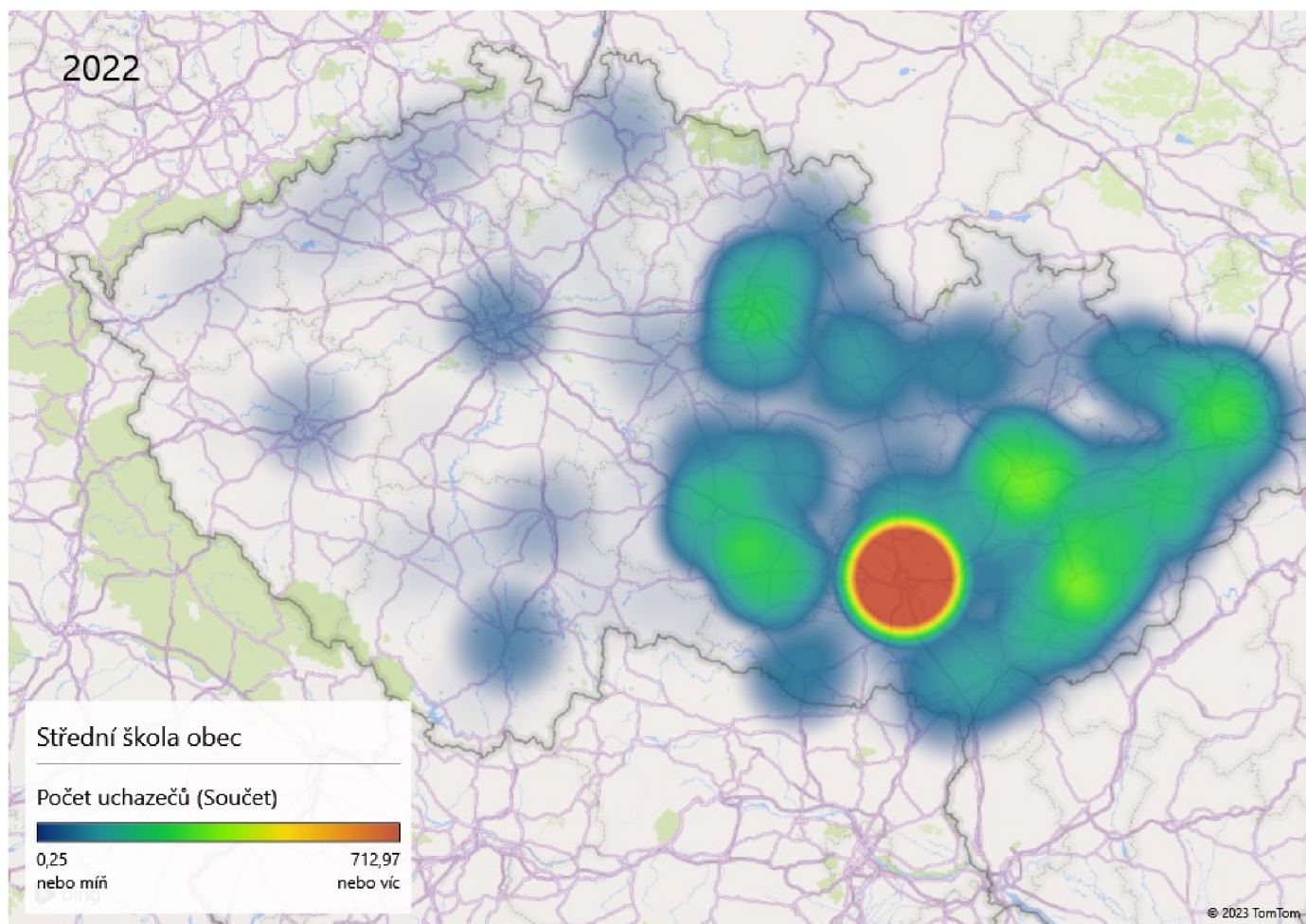
Zdroj: vlastní.

Obrázek C.2: Stát trvalého bydliště uchazečů VUT, rok 2022.



Zdroj: vlastní.

Obrázek C.3: Město trvalého bydliště českých uchazečů VUT, rok 2022.



Zdroj: vlastní.
Obrázek C.4: Poloha středních škol českých uchazečů VUT, rok 2022.