

**Univerzita Hradec Králové**  
**Fakulta informatiky a managementu**  
**Katedra informačních technologií**

**Analýza a zpracování dat z databázové aplikace**  
**Kalorické tabulky**  
Diplomová práce

**Autor:** Bc. Lukáš Kohl  
**Studijní obor:** Aplikovaná informatika

**Vedoucí práce:** Ing. Barbora Tesařová, Ph.D.

Hradec Králové

Listopad 2019

Prohlášení:

Prohlašuji, že jsem diplomovou práci na téma analýza a zpracování dat z databázové aplikace Kalorické tabulky zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 28.11.2019

Bc. Lukáš Kohl

**Poděkování:**

Děkuji vedoucí diplomové práce Ing. Barboře Tesařové, Ph.D. za metodické vedení, konzultace, rady a také čas, který mi věnovala. Děkuji firmě Dine4Fit, která provozuje aplikaci Kalorické tabulky, za instrukce a pokyny, vztahující se k technické stránce databáze, a možnosti pracovat na tomto tématu. Dále pak Zdeňku Svobodovi, který mi byl nápomocen při řešení technických problémů s připojením a aktualizací dat.

## **Anotace**

Tato diplomová práce se zabývá podrobným představením pojmů a technologií, které se používají v oblasti Business Intelligence. Budou zde zejména představeny pojmy jako jsou datové sklady a tržiště, také technologie OLAP, OLTP, dolování dat či ETL. Všechny uvedené pojmy a technologie souvisí s přípravou dat pro analýzy nad daty v databázích. V diplomové práci budou také v teoretické části vysvětleny i další důležité a nezbytné podpojmy výše zmíněných přístupů, technologií nebo mechanismů. Představeny budou programy, pomocí kterých lze data shromažďovat a využít je pro menší analytické účely. Praktickou částí této práce bude příprava a stažení dat z datového skladu Kalorických tabulek. Dále pak bude navržen datový sklad, při kterém se využije ETL proces, vytvoření faktových tabulek a tabulek dimenzí, a následně budou vypracovány základní dashboardy, které pomohou při zkoumání a analyzování změn stravovacích návyků a pohybových aktivity Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018. Výstupem práce budou připravené dashboardy pro možnost využití hlubších analýz prováděných nad daty a možná designová doporučení, jak rozšířit aplikaci. Aplikace je využívána uživateli, jak v mobilní podobě (iOS, Android), ale také pomocí desktopových prohlížečů na internetu. Kalorické tabulky jsou v dnešní době jedním z velmi významných nástrojů, které pomáhají uživatelům udržet fit postavu a mít přehled o všem, co konzumují a také o jejich aktivitách.

**Klíčová slova:** Big data, Business Intelligence, Data mining, databáze, analýza, kalorické tabulky, datová kostka, OLAP, transformace, ETL, databázové nástroje, databázové aplikace, JDBC

## **Annotation**

### **Title: Analysis and processing of data from database application named Kalorické tabulky**

This Diploma Thesis deals about a detailed presentation of concepts and technologies that are used in the wide field of Business Intelligence. In particular, terms such as data warehouses or marketplaces will be introduced as well as OLAP, OLTP, data mining and ETL technologies. All the above mentioned concepts or technologies relate to the preparation of data for analysis over the database.

The thesis will also explain in the theoretical part other important and necessary sub-concepts of the above mentioned approaches, technologies or mechanisms. The programs, which can be used to collect data and use it for smaller analytical purposes, will be presented. The practical part of this thesis will be to download and prepare data from database application Caloric tables. There will be designed a data warehouse, where the ETL process will be used and then will be developed basic dashboards, which will help in examining and analyzing changes in eating habits and physical activity of Czechs in individual socio-demographic groups in 2016-2018.

The output will be basic dashboards for deeper analytic usage and also possible design recommendations on how to extend the application. The application is used by users both in mobile form (iOS, Android) but also through desktop browsers on the Internet. Nowadays, calorie tables are one of the most important tools for keeping your body fit and keeping track of everything you eat and your activities.

Keywords: Big data, Business Intelligence, Data mining, database, analysis, caloric tables, data cube, OLAP, transformation, ETL, database tools, database applications, JDBC

# Obsah

1	Úvod.....	1
1.1	Cíl práce.....	2
1.2	Struktura práce .....	2
2	Business Intelligence .....	3
3	Architektura Business Intelligence .....	8
3.1	Datový sklad.....	8
3.1.1	Metody vzniku datových skladů .....	11
3.1.2	Architektura datových skladů .....	12
3.1.3	Schéma datových skladů.....	14
3.2	Datová tržiště.....	16
3.3	OLTP a OLAP .....	17
3.3.1	OLTP .....	17
3.3.2	OLAP .....	19
3.3.3	Multidimenzionální OLAP kostka.....	26
3.3.4	Operace datové kostky .....	27
3.4	ETL (Datová pumpa).....	29
3.4.1	Proces ETL.....	30
3.5	Data Mining.....	34
3.6	Data Culture.....	36
4	Analytické a reportovací aplikace.....	38
4.1	Analytické aplikace .....	38
4.2	Reportovací aplikace.....	39
5	Praktická část.....	40
5.1	Popis firmy.....	40
5.2	Aplikace Kalorické tabulky .....	41

5.3	Přístup k datům.....	44
5.3.1	Struktura dat .....	44
5.3.2	JDBC .....	45
5.3.3	Java databázový klient SQuirrel.....	47
5.3.4	Java databázový klient DBeaver.....	48
5.3.5	JetBrains DataGrip.....	49
5.3.6	GoodData.....	50
5.3.7	Microsoft Visual Studio .....	51
5.4	Schéma datového skladu .....	52
5.5	Tvorba dimenzí a faktových tabulek.....	54
5.6	Microsoft Power BI Desktop .....	59
5.6.1	Načtení dat .....	59
5.6.2	Vytvoření vazeb (relací).....	60
5.6.3	Tvorba dashboardů.....	60
6	Shrnutí výsledků.....	68
7	Závěry a doporučení .....	69
8	Seznam použité literatury a pramenů.....	71
8.1	Elektronické zdroje .....	71
8.2	Tištěné zdroje .....	76

## Seznam obrázků

Obrázek 1: Cyklus BI.....	4
Obrázek 2: Cyklus: data, informace, znalost, metaznalost.....	5
Obrázek 3: Graf operativního úložiště dat (ODS).....	6
Obrázek 4: Příklad scorecard dashboardu.....	7
Obrázek 5: Diagram datového skladu.....	8
Obrázek 6: Datová tržiště a sklady v kostce .....	10
Obrázek 7: Arhcitektura Bottom-Up podle R. Kimballa.....	12
Obrázek 8: Architektura Top-Down podle B. Inmona .....	12
Obrázek 9: Hvězdicové schéma datového skladu .....	14
Obrázek 10: Vločkové schéma datového skladu.....	15
Obrázek 11: Tržiště jako jednotlivé části skladu.....	16
Obrázek 12: Příklad OLTP databázového modelu .....	18
Obrázek 13: Řešení vztahů dimenzí M:N.....	23
Obrázek 14: Datová kostka OLAP.....	26
Obrázek 15: Datová kostka – operace slice .....	27
Obrázek 16: Datová kostka – operace drill-down, drill-up .....	28
Obrázek 17: Datová kostka – operace pivotace .....	28
Obrázek 18: ETL proces .....	33
Obrázek 19: Techniky dolování dat.....	35
Obrázek 20: Data Culture – sjednocení z různých pozic.....	37
Obrázek 21: SAP Crystal Report.....	39
Obrázek 22: Kalorické tabulky – webová platforma.....	42
Obrázek 23: Kalorické tabulky – mobilní platforma .....	43
Obrázek 24: Architektura JDBC.....	46
Obrázek 25: Prostředí Squirrel databázového klienta .....	47
Obrázek 26: Prostředí DBeaver databázového klienta .....	48
Obrázek 27: Výsledek dotazu v aplikaci DataGrid .....	49
Obrázek 28: Prostředí reportovacího nástroje GoodData .....	50
Obrázek 29: Schéma deníku jídel – faktová tabulka a dimenze.....	58
Obrázek 30: Schéma pohybového deníku – faktová tabulka a dimenze .....	58



Obrázek 31: Import – výčet tabulek z SQL serveru .....	59
Obrázek 32: Relace mezi tabulkami .....	60
Obrázek 33: Procentuální výsečový graf podle pohlaví.....	61
Obrázek 34: Procentuální výsečový graf podle zvoleného jazyka aplikace .....	61
Obrázek 35: Sloupcový graf věkového zastoupení .....	62
Obrázek 36: Dashboard – průměrně strávený čas aktivitami .....	63
Obrázek 37: Dashboard – průměrná hmotnost a vývoj hmotnosti v čase.....	64
Obrázek 38: Sloupcový graf – počet záznamů hmotnosti .....	64
Obrázek 39: Dashboard – komplexní přehled aktivity .....	65
Obrázek 40: Sloupcový graf stráveného času podle aktivity .....	66
Obrázek 41: Dashboard – komplexní přehled stravování .....	66
Obrázek 42: Sloupcový graf – specifické zadání.....	67

## Seznam tabulek

Tabulka 1: Porovnání charakteristik jednotlivých typů tabulek faktů .....	22
---	----

# 1 Úvod

Data jsou všude kolem nás, i my jako fyzické osoby vlastníme každý svá vlastní data, kde některá poskytujeme dále pro účely jako jsou například nákupy na internetu, při podání svědectví nebo také při evidenci svých návyků o stravě a aktivitě či stavbě těla. Data nemusejí být samozřejmě jen o nás, ale také o dalších různých objektech nebo činnostech. Firmy shromažďují tato data, kterých je dnes opravdu velké množství, ty následně ukládají do fyzických či virtuálních úložišť, a přitom nemusí ani vědět, že byt' jen jednoduchou analýzou těchto dat mohou přijít o vzácnosti v získaných datech, která mají k dispozici. Tyto vzácnosti nebo výkyvy, které lze získat analýzou právě daných dat, může společnost využít ke svému rozvoji nebo poskytnutí přehledu svým zákazníkům a pracovníkům. Aby byly výsledky zkoumání spolehlivé, je třeba analyzovat všechna data, a to do značné hloubky, což může být i pro specialisty značně obtížné. Analýzy jsou v dnešní době zjednodušeny pomocí již zhotovených specializovaných nástrojů a programů, které lze při zkoumání dat nasadit. Tyto nástroje jsou velkým přínosem i přes jejich vysoké pořizovací ceny. Správnou analýzou a optimalizací lze dosáhnout lepších výsledků než doposud. Firmy tím tak mohou očekávat velmi rychlou návratnost původně investované částky a také možnost vyšších zisků nebo úspěchů do budoucna. Základní pojmy, prvky či operace, které se starají o shromažďování, transformaci, dolování a analýzu dat, to vše zastřešuje systém Business Intelligence.

## **1.1 Cíl práce**

Cílem této diplomové práce je navrhnout datový sklad, při kterém bude využit ETL proces, vytvoření faktových tabulek a tabulek dimenzí, které budou následně využity při vypracování základních dashboardů. Dashboardy pomohou při zkoumání a analyzování změn stravovacích návyků a pohybové aktivity Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018.

## **1.2 Struktura práce**

Úvodem této práce je stručný popis využití, zpracování, sběr a uschovávání dat, následně se navazuje v teoretické části na jednotlivé technologie a metody, které se při zpracování, přípravě, ukládání dat a také jejich analýze využívají. Jde o podrobné představení tématu Business Intelligence. Praktickou částí této práce bude popis získání dat z databázové aplikace, následně jejich zpracování, návrh datového skladu a jejich připravení pro účely analýz. Příprava bude obsahovat již zmíněnou tvorbu datového skladu, tabulek dimenzí a faktových tabulek. Závěrem této práce budou vypracované základní dashboardy pro možnosti provádění analýz.

## 2 Business Intelligence

Business Intelligence je odborný termín pro určité kombinace, a to jak softwarových, tak i hardwarových nástrojů, analytických prostředků, metod a nástrojů, databází a příslušných aplikací. Všechny tyto vyjmenované pojmy slouží pro získávání znalostí a dat, které mohou být nějakým způsobem neviditelné při pouhém pohledu. Získané informace se týkají údajů reálného světa, které jsou všude kolem nás a spadají do oboru informačních technologií, ekonomiky či podnikání. (Žižka 2011)

Cílem Business Intelligence je zanalyzovat získaná data a cíleně odhalit trendy, vzorce či jakékoli poznatky zpozorované při zkoumání údajů a jejich povahy. Tato zjištění jsou brána jako přesné a promyšlené pohledy na zkoumanou společnost a také výsledky, které procesy přinášejí. Pomocí Business Intelligence lze odhalit například výkon zaměstnanců, lepší pohled na finance a zabezpečení po finanční stránce, celkovou spokojenost společnosti a také konverze. Tato technika se používá i pro předpověď budoucího výkonu firem, pokud je to jen z pohledu získaných údajů možné, sledují se tak v časové dimenzi data z minulosti, současnosti a za pomoci dalších nástrojů je možné odhalit nebo se jen přiblížit, jak budou data vypadat v budoucnosti. Všechny tyto postupy a procesy jsou velmi důležité a pokud nebude na toto téma kladena pozornost, ztrácí se tak schopnost vidět skutečně dosažené a kvantifikované výsledky a také možné dopady. V dnešní době je Business Intelligence velmi mocným nástrojem, který je využíván pro rozhodování firemních procesů. (Conrad 2019)

*„Proces business intelligence vychází z transformace vstupních dat na informaci, která je dále přeměněna na znalost použitelnou pro rozhodování zakončené vybraným typem akce.“ (Žižka 2011)*

Business Intelligence je využívána i pro její úžasnou vlastnost a to, že umožňuje pracovníkovi, který se této oblasti věnuje a pracuje v ni, přistupovat a dotazovat se na jakákoli data, která jsou k dispozici. Tehdy takto přistupovat k datům nebylo možné, bylo zapotřebí zkontaktovat odpovědnou osobu za uschování dat, nejčastěji se jednalo o pracovníky IT, ti data shromáždili na jedno místo, dále provedli další nutné kroky, aby byla data v čitelné podobě a následně je poskytly zájemcům, kteří s nimi potřebovali pracovat. Dnes se lze tedy dotazovat na přímo, kde prostředníkem je jen prostředí Business Intelligence, to data zpracuje za nás a vytvoří tak přehlednější formu a reprezentaci. Tento přístup je také několikanásobně rychlejší. Dalšími výhodami jsou možnost pracovat s opravdu velkými objemy dat, podpora Data Miningu (dolování dat), Data Warehouse (datových skladů) a také OLAP (online analytical processing).

(Laberge 2012)

Business Intelligence technologie by měly být podle Laberge (2012) s následujícími vlastnostmi:

- Rozšířená použitelnost a kompatibilita
- Rychlost reakce na podněty a požadavky
- Dostupnost a aktuálnost
- Vysoká kvalita a přesnost
- Významnost a užitek

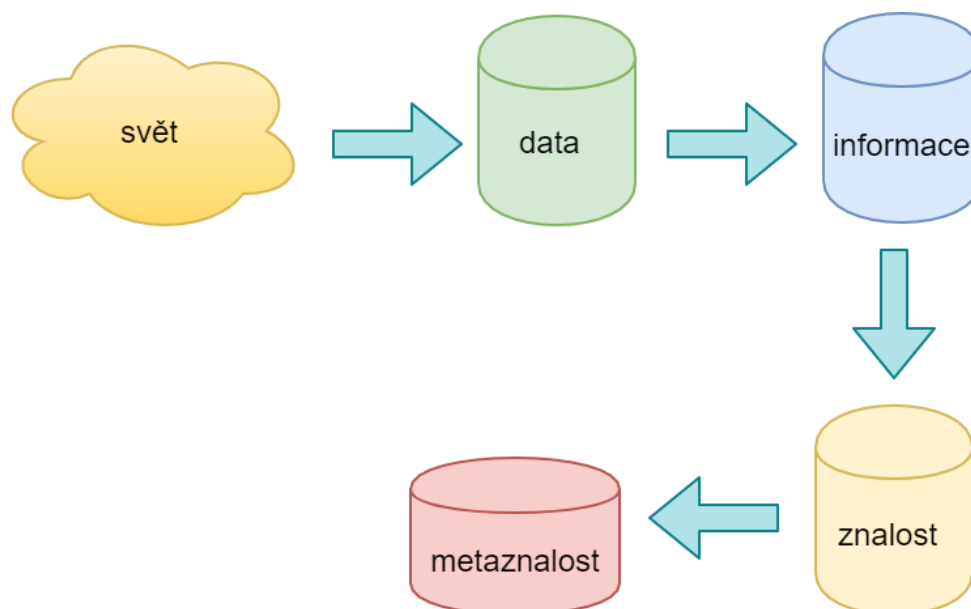
Jde jen o základní výčet nejdůležitějších vlastností.



**Obrázek 1: Cyklus BI**  
**Zdroj: NILORT IT GROUP 2019**

Data, znalost, informace, metaznalost, to vše jsou další prvky využívané v BI. Data jsou brána jako údaje, které jsou získávané pomocí měření či pozorování reálného světa kolem nás. Data mohou být numerická, nominální neboli vyjmenované povahy. Informace zde reprezentují část dat, která jsou k nějakému řešení problému relevantní. Metaznalost neboli takzvaná nadznanost se považuje za znalost o znalosti. A znalost je pouze informací zobecněnou. Znalosti a metaznalosti třídíme také jako induktivní a deduktivní. Induktivní vzniká zobecněním údajů konkrétních případů. Deduktivní znalost je od něčeho odvozena, zejména jde o odvození z nějakého vzorce. (Žižka 2011)

Cyklus těchto prvků lze vidět na obrázku (Obrázek 2) níže.



**Obrázek 2: Cyklus: data, informace, znalost, metaznalost**  
Zdroj: vlastní zpracování dle (Žižka 2011)

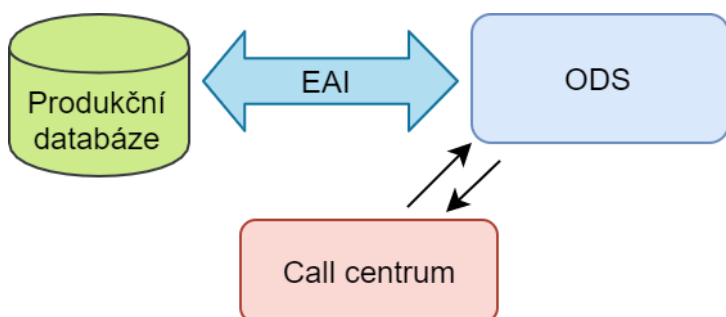
Prostředí BI se skládá z následujících komponent:

- BI aplikace
- Datový sklad
- ODS (Operation Data Store)
- Datové tržiště
- DSA (Data Staging Area)
- Systém řízení datového skladu
- Metadata datového skladu
- Datové zdroje
- A další

(Tesařová 2019)

Operativní úložiště dat, zkratkou ODS, představuje místo, kde je možné sledovat data, jak jsou agregována a konsolidována. Jde o integrovaná data z produkčních systémů. Jednou z výhod ODS je možnost modifikace a vkládání dat kdykoli. Této výhody je docíleno pomocí napojení na EAI (Enterprise Application Integration), jde o technologii, která má mnoho společného s ETL, jen je výhodnější v tom, že umožňuje obousměrný přenos dat. ODS oproti DSA disponuje díky transformačním operacím data, která jsou konzistentní, konsolidovaná a orientovaná, zároveň tato data mohou být doplněna o agregace.

(Novotný, Pour, Slánský 2005)



**Obrázek 3: Graf operativního úložiště dat (ODS)**

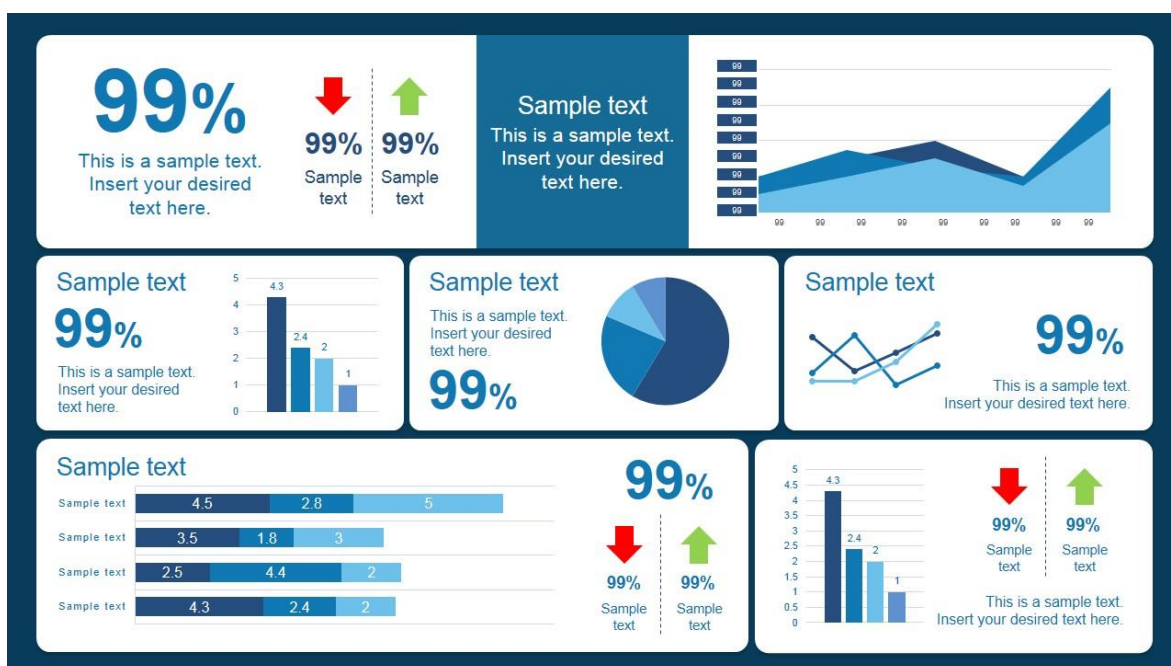
Zdroj: vlastní zpracování

DSA, dočasné úložiště dat, je zde bráno jako prostředník mezi zdrojem dat a datovým skladem, využívá se hlavně pro rychlé extrahování dat z různých zdrojů. Následně se získaná data uloží právě do tohoto dočasného úložiště, kde proběhne kombinace dat, transformace, validace a také očištění. Data jsou uspořádána podle hvězdicového schématu a začne následně proces nahrávání dat do datového skladu. (Christiansen 2019)

Oba zmíněné pojmy, ODS a DSA, jsou jedny z komponent Business Intelligence a zároveň jsou oba přítomné v datové vrstvě. Business Intelligence je využíván zejména pro účely napomáhání při rozhodování. Mezi typické příklady nástrojů a aplikací pro rozhodování uvádí B. Tesařová (2019):

- Dotazovací nástroje
- OLAP
- Data mining
- Scorecard-Dashboarding

Dotazovací nástroje pracují s předpřipravenými reporty z různých souborů nebo tabulek. Oproti marketingovému systému se zde mohou data čerpat z již očištěného skladu. (Tesařová 2019)



**Obrázek 4: Příklad scorecard dashboardu**  
**Zdroj: SlideModel 2019**

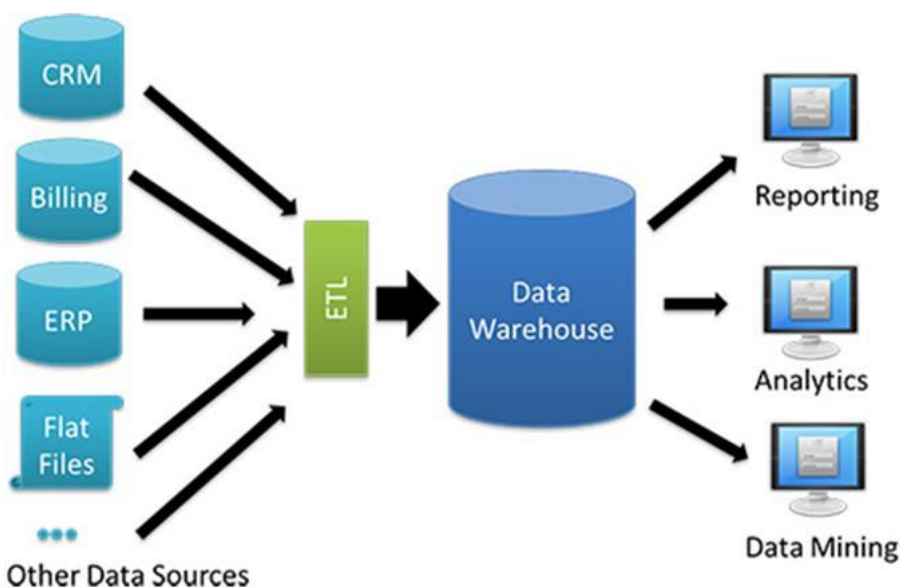


### 3 Architektura Business Intelligence

Jak již bylo uvedeno v kapitole výše, Business Intelligence je komplexní systém využívající několik komponent, které spolu kooperují, aby se dosáhlo požadovaného výsledku.

#### 3.1 Datový sklad

Datový sklad, anglicky Data Warehouse, se používá a je sám komponentou systému Business Intelligence. Datový sklad je jedno velké úložiště dat, kde se data ukládají z různých zdrojů a následně se při dalších operacích data z datového skladu čerpají, pracuje se s nimi. Datové sklady disponují objemnými daty, které mohou být ve velikostech až řádů terabajtů.



Obrázek 5: Diagram datového skladu  
Zdroj: Sharma 2018

Dle Kimballa a Rosse datové sklady musí splňovat nutná kritéria, ať už z hlediska interní či externí práce s ním, tak z důvodu bezpečnosti a konzistence. Musí usnadňovat přístup k informacím organizace, obsah datového skladu musí být srozumitelný, čitelný a jeho data zřetelná, srozumitelná, a to zejména pro ty, kteří s nimi pracují, ale také i pro další účastníky, kteří budou nahlížet na výstupy či budou přítomni při prezentaci výsledků. Dalším kritériem jsou nástroje, které využívají a přistupují k datovému skladu, ty musí být snadno použitelné, jednoduché a jejich přístupová doba musí být co nejmenší. Pomocí výše uvedených kritérií lze provádět různé kombinace a analýzy nad datovým skladem v rozumném časovém intervalu. Informace v datovém skladu jsou uloženy tak, aby byly konzistentní. Jelikož jsou data shromažďována z několika různých zdrojů, je nutné je před uložením očistit, a tak zaručit jejich důvěryhodnost. Je nutné také dbát na jejich kvalitu a přístupnost při potřebě uživatele. Informace, které pocházejí z jednoho business procesu, musí odpovídat informacím v jiném business procesu. Tento přístup také platí, pokud mají dva procesy stejné jméno, potom musí znamenat to samé, tzn. mají stejný význam. Nelze tedy mít dva a více stejných procesů se stejnými názvy, a tak je nutné je odlišit. To vše se týká právě konzistence dat. Adaptivnost a odolnost vůči jakýmkoli změnám, to je další kritérium pro datový sklad. I když se na začátku celého projektu řekne, že se nic měnit nebude, tak pro budoucí potřeby je vždy třeba udělat nějaké změny či rozšíření, je to jen otázka času. Také technologie se v průběhu času mění a využívají se lepší a novější. Z toho plyne, že datový sklad musí být připraven na případné změny menších i větších měřítek. Pokud je změna datového skladu nutná, nemělo by docházet k narušení stávajících dat nebo aplikací. Dalším důležitým faktorem a kritériem datového skladu je bezpečnost a ochrana dat proti úniku nebo neoprávněnému vstupu. Datový sklad musí účinně řídit důvěrné přístupy, které jsou uloženy. Je nutné, aby data v DS byla správná a podporovala základ pro zlepšené rozhodování a podporu při rozhodování. (Kimball, Ross 2011)

Podle J. Šarmanové (2007) je vznik datového skladu motivován následujícími aspekty:

- Umístění dat na více místech
- Odložená data, která nejsou využita (archivace)
- Lepší optimalizace řízení
- Možnost aplikace rozšířených statistik, analýz

Datové sklady, jejich vývoj a nasazení začaly vznikat především v době, kdy docházelo k masivnímu nasazování informačních systémů, které produkují obrovské balíčky dat. Následně se začaly používat analytické a také statistické nástroje nad uchovanými daty pro dosažení lepších výsledků. (Vrána 2001)

Při pohledu na obrázek (Obrázek 6) lze porovnat rozdíly mezi datovými sklady a datovými tržišti. Na první pohled lze vidět, že datová tržiště bývají používána pouze pro specifitější projekty.

	Data Mart	Data Warehouse
Size	< 100 GB	100 GB +
Subject	Single Subject	Multiple Subjects
Scope	Line-of-Business	Enterprise-wide
Data Sources	Few Sources	Many Source Systems
Data Integration	One Subject Area	All Business Data
Time to Build	Minutes, Weeks, Months	Many Months to Years

**Obrázek 6: Datová tržiště a sklady v kostce**  
**Zdroj: Tesařová 2019**

### 3.1.1 Metody vzniku datových skladů

Mezi další prvky, které jsou využity při tvorbě datových skladů zařazuje L. Lacko (2005) metody vytváření datových skladů. Mezi tyto metody patří metoda „velkého třesku“ a přírůstková metoda. Metoda „velkého třesku“ spočívá ve složení a kombinaci následujících etap:

- analýza požadavků
- návrh struktury
- vytvoření a naplnění datového skladu
- vytvoření přímého přístupu, případně přístup přes datové trhy

Metoda „velkého třesku“ je výhodná z hlediska toho, že je možné zanalyzovat a vypracovat projektový plán ještě před jeho realizací. Je jednodušší hlavně v tom, že se vytváří a implementuje pouze jeden návrh či projekt. Tato metoda má také ovšem nevýhody, kde jedna z nich je, že ačkoli lze vypracovat návrh bez předchozí realizace, tak tím uteče plno postřehů a nápadů, které se vyskytnou až při realizaci. Tato metoda špatně reaguje na změny. Čas, který je třeba vynaložit na projektový plán, je přímo úměrný velikosti datového skladu, tato nevýhoda také souvisí s tím, že přinášení benefitů je aplikováno až po dokončení realizace projektu.

(Lacko 2005)

Metoda přírůstková je opak metody „velkého třesku“, vzniká postupně, tak jak jdou po sobě jednotlivé etapy. Přírůstková metoda vyniká svojí elasticitou, model se postupem času upravuje dle potřeb. Tento model je nasazen dříve, případně jen jeho část, a také rychleji přináší užitek oproti předchozí metodě. (Vercellis 2009)

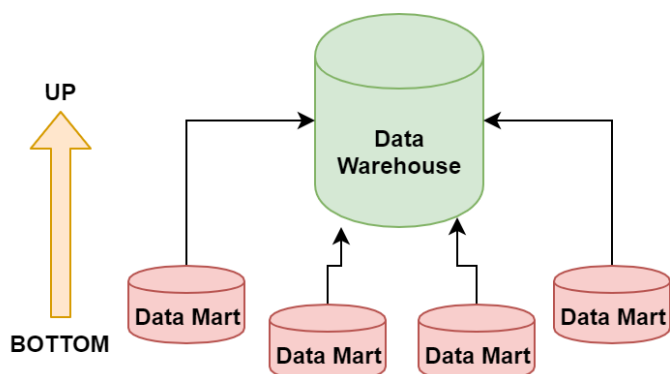
### 3.1.2 Architektura datových skladů

U tvorby datových skladů se prosadily a dodnes zůstaly nejpopulárnější dva koncepty, kterými jsou dvouvrstvá a třívrstvá architektura.

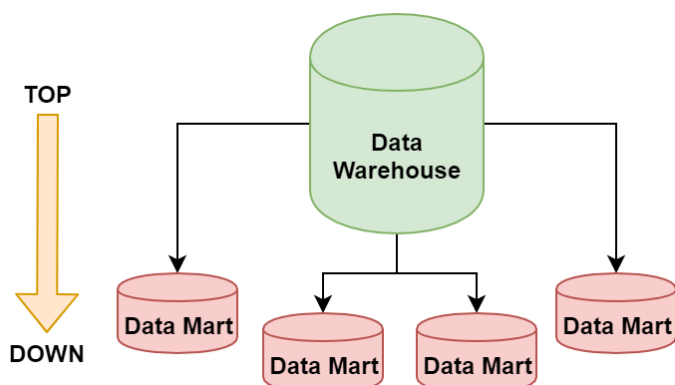
#### 3.1.2.1 Dvouvrstvá architektura

U dvouvrstvé architektury při budování datového skladu není nutné klást důraz na počáteční náklady a doby realizace, jelikož oproti třívrstvé architektuře je vývoj rychlejší a počáteční náklady na realizaci nižší. Strůjcem dvouvrstvé architektury je Ralph Kimball, který se snaží o to, aby tato architektura byla pouze jako samostatné datové tržiště a nikoli centrální úložiště dat. Oproti tomu je zde nevýhodou fakt, že se mohou data na některých místech nacházet duplicitně, což vede k větším paměťovým nárokům, a tím dojde ke zvýšení počtu potřebných datových tržišť. Při větším počtu datových tržišť se postupně vytrácí konsolidovanost a rapidně narůstají náklady na potřebnou údržbu skladů.

(Linstedt, Olschimke 2015)



Obrázek 7: Architektura Bottom-Up podle R. Kimballa  
Zdroj: vlastní zpracování dle Singh 2016



Obrázek 8: Architektura Top-Down podle B. Inmona  
Zdroj: vlastní zpracování dle Singh 2016

### 3.1.2.2 Třívrstvá architektura

Koncept třívrstvé architektury spočívá v řadě výhod oproti dvouvrstvé architektuře, mezi hlavní výhody patří flexibilita, podpora náročných analytických operací a také komplexnost. Pomocí zmíněných vlastností je možné v třívrstvé architektuře vybudovat nespočetně datových tržišť. Vývoj a investice oproti předchozí architektuře jsou několikanásobně vyšší, jde o jedinou nevýhodu mezi jmenovanými architekturami. (Danel 2010)

### 3.1.2.3 Vrstvy datového skladu

Datové sklady jsou tvořeny architekturou, která obsahuje následující vrstvy:

- **Vrstva zdroje dat** je místo, kde jsou původní data shromažďována z různých interních či externích zdrojů. Je umístěna v relační databázi.
- **Vrstva datového toku** slouží pro dělení dat mezi zdroji dat a datovým skladem. Dochází zde k extrahování dat z různých zdrojů a využívá se zde z mnoha nástrojů a technologií, protože extrahovaná data mohou mít různé formáty. Následně po extrakci proběhne kontrola kvality dat, kde výsledkem jsou kvalitní a očištěná data, která jsou nezbytná pro datový sklad. Tato vrstva obsahuje nástroje pro integraci dat (ETL).
- **Vrstva ukládání dat** je zde reprezentována právě jako datový sklad.
- **Vrstva prezentace dat** je vrstvou, která umožňuje uživatelům zobrazovat a dotazovat se na uložená data. Lze také provádět analýzy a využívat statistické nástroje.

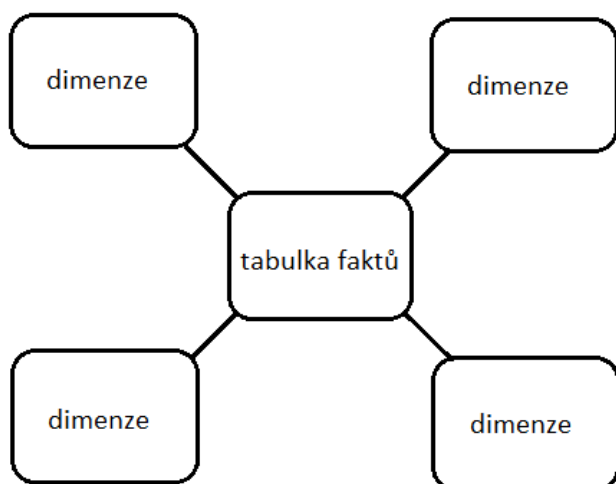
(Wainstein 2018)

### 3.1.3 Schéma datových skladů

Mezi schémata, která jsou využívána při tvorbě datových skladů, patří hvězda (star) a sněhová vločka (snowflake).

#### 3.1.3.1 Schéma hvězdy

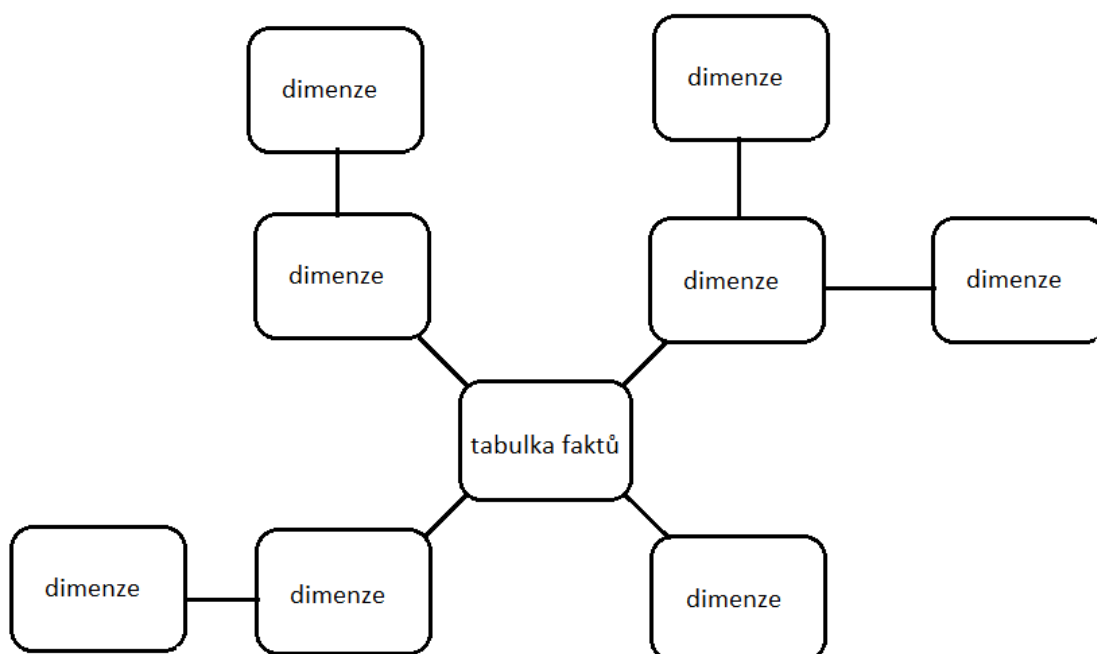
Schéma hvězdy umožňuje hvězdicové propojení dimenzí a faktů. Jde o nejčastější a nejjednodušší využívané schéma. Středem tohoto schématu je tabulka faktů, která je propojena s dalšími tabulkami dimenzí. Mezi jeho výhody patří jednoduchost nasazení, rychlost a odezva. Tabulky faktů mají cizí klíče, které jsou napojeny na primární klíče právě centrální tabulky faktů. (Žižka 2011)



**Obrázek 9: Hvězdicové schéma datového skladu**  
Zdroj: vlastní zpracování

### 3.1.3.2 Schéma vločky

Schéma vločky, jak uvádí J. Žižka (2011), umožňuje normalizovat dimenze více mezi sebou propojených tabulek. Liší se tedy od schéma hvězdy tím, že jsou dimenze normalizované. Z tohoto tvrzení vyplývá, že každá dimenze má právě svou vlastní tabulku. Vločkové schéma není natolik používané jako výše uvedené hvězdicové, a to z důvodu nutnosti propojit více tabulek mezi sebou. Tato propojení mají následně také vliv na celkový výkon a odezvu datového skladu. Výhodou je, že data nejsou redundantní, a tedy nároky na úložiště nejsou tak velké.



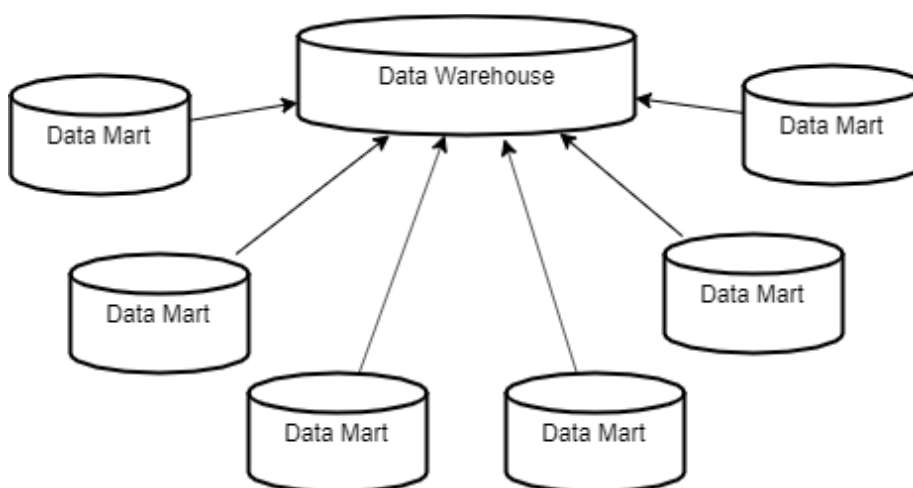
**Obrázek 10: Vločkové schéma datového skladu**  
Zdroj: vlastní zpracování



### 3.2 Datová tržiště

Datová tržiště, označována také jako Data Marts, jsou obdobou datových skladů, rozdílem je zde omezený okruh použití. I zde, jako tomu bylo ve vločkovém schématu, je podstatou decentralizovaný datový sklad. Výhodou je lepší orientace uživatelů v datech a rychlejší a levnější implementace. Datová tržiště jsou také využívána jako mezičlánek, který je využit při transformaci dat z produkčních databází. Datová tržiště pokrývají celou problematiku konkrétního okruhu. Důvody pro vytváření datových tržišť jsou snížení nákladů a zkrácení doby pro návratnost investice.

(Gála, Šedivá, Pour 2015)



**Obrázek 11: Tržiště jako jednotlivé části skladu**

Zdroj: vlastní zpracování

### **3.3 OLTP a OLAP**

Z hlediska IT můžeme rozdělit systémy na transakční a analytické, kde transakčními je OLTP a analytickými systémy OLAP. Pokud tento pohled zobecníme, vznikne tak OLTP jako „poskytovatel“ zdrojových dat, která se ukládají do datových skladů, a OLAP, který je analyzuje. Můžeme tedy říci, že OLTP poskytují zdrojová data do datových skladů, zatímco systémy OLAP je pomáhají analyzovat.

(Sodomka a Klčková 2018)

#### **3.3.1 OLTP**

OLTP (On-line Transaction Processing) je online transakční systém. Jde o systém, který se stará o ukládání dat, nad kterými se provádí online transakce. Tyto data jsou následně ukládána do datových skladů. U OLTP se provádí více ad-hoc čtení a zápisů v reálném čase. (Edureka 2019)

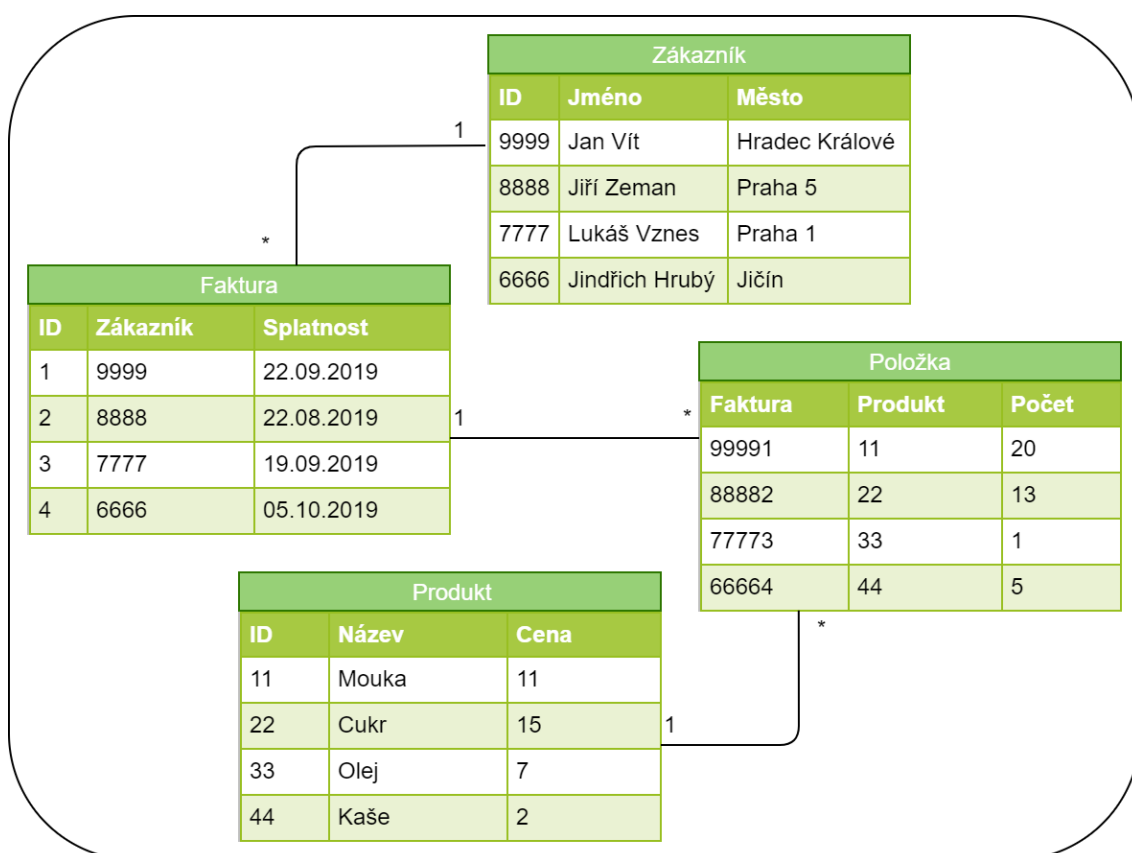
OLTP provádí a vyznačuje se velkými počty krátkých a jednoduchých transakcí. Ve většině případů jde o transakce typu aktualizovat, vložit, smazat. Transakční systém se stará o zachování integrity dat v prostředí s více přístupy a také řeší efektivitu, která je měřena transakcemi za sekundu. V OLTP databázích jsou ukládána aktuální podrobná data a také schéma jaké jsou v dané databázi využívány pro ukládání právě transakčních dat. Ve většině případů se u transakčních systémů využívá třetí normální forma (3NF). Třetí normální forma je, jak již bylo uvedeno schéma neboli návrh datové struktury databáze, 3NF je nutné dodržovat, a to z toho důvodu, aby bylo docíleno optimálního využití vlastností systému OLTP při tvorbě databází či databázových aplikací. (DW4U 2019)

Transakční systém OLTP lze také pojmut jako třídu softwarových nástrojů, které jsou schopné podporovat transakční orientované internetové aplikace. Za vznikem transakce je sled diskrétních výměn informací, kde transakce je základní jednotkou OLTP. Dále je nutné dodržovat dvě důležité charakteristiky transakčního systému OLTP, těmi jsou souběžnost či konkurence a atomicita. Souběžnost zakazuje vzájemné úpravy stejných dat více uživateli, jde o současnou změnu dat. Atomicita se stará o to, pokud se nevyvede nějaký krok transakce, tedy nedojde k jejímu správnému zakončení, neprovede se nic. Úspěšná transakce se vyznačuje

tím, že změna provedená nad databází musí být trvalá. OLTP systémy jsou také decentralizovány. (Rouse 2014)

Třetí normální forma, která je v mnoha případech využívána, musí splňovat podmínky druhé a druhá normální forma musí splňovat podmínky té první. Následuje tedy výčet podmínek a náležitostí 3NF, které uvádí M. Trakal (2009):

- Eliminace duplicitních sloupců v jednotlivých tabulkách
- Neexistuje závislost mezi neklíčovými sloupci v rámci tabulky
- Jednotlivé sloupce tabulky nesou jednoduché, dále nedělitelné informace
- Pro skupinu dat, které mají vymezeným význam se vytvoří zvláštní tabulka a řádky opatří unikátním primárním klíčem
- Podmnožinu dat, která je shodná pro určitý sloupec tabulky převede do samostatné tabulky a následně spoj se zdrojovou tabulkou cizím klíčem
- Odstranění sloupců tabulky, které jsou přímo závislé na jiné skupině sloupců tabulky než jen na primárním klíči



**Obrázek 12: Příklad OLTP databázového modelu**  
Zdroj: vlastní zpracování

### 3.3.2 OLAP

Online Analytical Processing (OLAP) je technologie, která obstarává ukládání dat do datových skladů či databázových aplikací. OLAP se používá pro velké objemy dat, může být až v řádech terabajtů, oproti OLTP, který je postavený na menší objemy dat a jednoduché operace. Uspořádává data tak, aby byla jednoduše přístupna pro analytické účely, které provádí specializovaní uživatelé, jde o uživatele, kteří se specializují na Business Intelligence. U technologie OLAP se využívají dotazy, které jsou oproti dotazům z OLTP složitější povahy, dále pak oproti OLTP nepoužívá ukládání dat ve třetí normalizované formě. U OLAP je důležitá rychlá odezva a přístup k datům, proto jsou data ukládána i duplicitně, jsou tedy redundantní, pro rychlejší odezvy je také využíváno více indexů. Pomocí této technologie lze zkoumat různé anomálie, které se v uložených datech nacházejí. Některé anomálie mohou být klíčovými pro další rozvoj společnosti. Zkoumají se také vztahy a závislosti mezi daty. U OLAP se využívají jako již bylo uvedeno složitější dotazy, kterými jsou například shlukování, agregace, filtrování či joinování. Pokládané dotazy mohou v závislosti na velikost databáze trvat až několik dní, urychlení těchto operací je závislé na efektivitě datového skladu. Urychlit průběh lze pomocí správné indexace v prostředí databáze či při využití materializovaných pohledů. (Wrembel 2006)

OLAP je specializován na řešení potřebných informací pomocí ad-hoc analýz podrobných dat. OLAP a OLTP spolu také spolupracují, jelikož OLAP pracuje s daty, které byly získány pomocí technologie OLTP a opačně tomu tak je u OLTP, které automatizuje podnikové procesy, které jsou řízené rozhodováním a rozhodování je právě podporováno nástroji OLAP. Hlavní strukturou analytického systému OLAP je datová kostka. O datové kostce bude psáno níže. (Žižka 2011)

Analytický systém pracuje s multidimenzionálními strukturami, které obsahují nenormalizované faktové tabulky a tabulky dimenzí. Fakta jsou numerické údaje o objektech, kterými se daná oblast zabývá. Tabulky faktů se vyznačují velkou velikostí. Dimenze jsou hierarchická uspořádání. Nejsou tak velké jako tabulky faktů a umožňují z různých pohledů zkoumat data. (Přichystal 2011)

### 3.3.2.1 Způsoby uložení dat v systémech OLAP

Rozlišují se způsoby ukládání dat pomocí nástrojů OLAP podle toho, nad jakými daty jsou analýzy prováděny. Existují tři způsoby ukládání dat.

Pokud jsou při OLAP analyzování dat využívána jen data, která jsou uložena čistě v relační databázi, hovoříme o způsobu **ROLAP**. Jedná se o relační OLAP, který se zabývá ukládáním dat multidimenzionální právě v relační databázi. Výhodou ROLAPu je flexibilita a lze jej využít v systému řazení báze dat, zkratkou SŘBD. (Tesařová 2019)

Další analýzy mohou být prováděny nad daty, které jsou v databázi či datovém skladu ukládány multidimenzionálně. O multidimenzionální ukládání dat a jejich analýzu se stará **MOLAP**. Provádí ukládání dat ve více dimenzích datových kostkách OLAPu. Rozdílem oproti ROLAPu je rychlost při generování výstupů z analýz, docíleno této rychlosti je pomocí nepotřebnosti provádět složitější SQL dotazy mezi propojenými tabulkami a čtení jejich dat. MOLAP má lepší strukturu ukládání dat, data jsou seskupována logicky a nemusí se zde provádět důraz na redundanci. Jedinou nutností, kterou je potřeba provést je převádět data z relačních databází pomocí vhodného ETL nástroje do datových skladů. Zde může dojít k menšímu zpoždění celého procesu, ale nedochází k významnému časovému omezení.

(Danel 2010)

A posledním způsobem ukládání dat v systémech OLAP je **HOLAP**. Jde o kombinaci výše zmíněných způsobů, tj. ROLAP a MOLAP, tím vzniká hybridní způsob OLAP, který nese již zmíněnou zkratku HOLAP. HOLAP zprostředkovává detailní data, která se nachází v relační databázi a také umožňuje provádět agregace v OLAP datových kostkách. U tohoto způsobu je jedinou nevýhodou udržování dat současně na dvou odlišných místech. Na druhou stranu umožňuje provádět přístup k velkému objemu dat, a to i s rychlou agregací. (Danel 2010)

### 3.3.2.2 Fakta

Fakta neboli faktová tabulka, jak již bylo zmíněno obsahuje velké množství cizích klíčů, které slouží pro napojení jednotlivých dimenzí. Faktová tabulka také vlastní ukazatele, které pro daný proces chceme sledovat. Při složení jednotlivých dimenzí dostaneme primární klíč dané tabulky, z toho plyne, že primární klíč není jednoduchého formátu nýbrž složitého. Skládání jednotlivých dimenzí zde reprezentuje vztahy M:N mezi jednotlivými dimenzemi. Ve faktové tabulce se dbá na granularitu, což znamená, že všechny hodnoty řádků musí mít stejnou úroveň. Každý řádek pak odpovídá jedné hodnotě ve faktové tabulce. (Kimball 2013)

Fakta lze dle Poura a kolektivu (2018) rozdělit podle agregace na:

- Adaptivní
- Neadaptivní
- Semiadaptivní
- Speciální typy – tabulky faktů bez ukazatelů (factless)

Adaptivní fakta je možné agregovat přes všechny dimenze, a to pomocí sčítání. U neadaptivních faktů není možné provádět sčítání, a tak nelze provést agregaci přes dimenze. Neadaptivní faktová tabulka obsahuje zpravidla textová či strukturovaná data. Při semiadaptivním faktu je možné agregovat pouze přes některé dimenze, aby bylo umožněno agregovat je nutné agregovat pomocí nějakých funkcí. (Kimball 2013)

Jak uvádí Pour a kolektiv (2018), v analytických databázích lze faktové tabulky podle jejich granularity rozdělovat na:

- Transakční
- Snímkové (periodické)
- Akumulované

Transakční tabulky obsahují detailní informace, které jsou vázané na jednotlivé transakce a jsou na nejvyšší možné granularitě. Transakční tabulky faktů jsou v praxi často využívány. Časová dimenze se zde rozděluje na den a čas, kde jde o čas podle jednotlivých transakcí v rámci onoho dne. Snímkové (periodické) tabulky faktů jsou z výše vyjmenovaných nejvíce využívány v praxi. A to díky tomu, že data se do datových skladů nahrávají v pevně daných časových intervalech (snímcích), mohou to být třeba dny. Vyjadřují také souhrnné hodnoty ukazatelů za celý časový snímek. Tento typ faktové tabulky se také používá pro predikci trendů vybraných ukazatelů. Disponují také uniformní hustotou, kde se snímky vytvářejí i když nedošlo k nějaké změně. Nevýhodou této vlastnosti je větší počet řádků a tím pádem i větší nároky na úložný prostor. Snímkové i transakční typy tabulek faktů mají data již neměnné, kdežto akumulovaný typ faktové tabulky umožňuje i jejich změnu. Akumulované (stavové) tabulky faktů, jako oba předešlé typy, jsou také závislé na výskytu transakcí, ale odlišností zde je, že jejich hodnoty se v průběhu času mění (aktualizují). Tyto tabulky se využívají pouze ve speciálních případech, jejichž uplatnění je takřka ojedinělé.

(Pour a kolektiv 2018)

V tabulce (Tabulka 1) níže lze porovnat rozdíl mezi hlavními charakteristikami jednotlivých typů faktových tabulek.

Charakteristika	Transakční	Snímková (periodická)	Akumulovaná
Časová perioda	Časový okamžik (čas transakce)	Pravidelné, určené intervaly	Nedefinováno
Granularita	1 záznam = 1 transakce	1 záznam = 1 časový interval	1 záznam (postupně aktualizovaný)
Plnění tabulky faktů	Přidávání záznamů	Přidávání záznamů	Přidávání záznamů a aktualizace
Aktualizace záznamů	Neprovádí se	Neprovádí se	Provede se vždy při změně
Časová dimenze	Datum transakce	Datum konce časového intervalu	Více kalendářních dat pro provádění změn
Fakta (ukazatele)	Obsah transakční aktivity	Obsah odpovídající def. čas. intervalu	Obsah odpovídající celému živ. cyklu dat

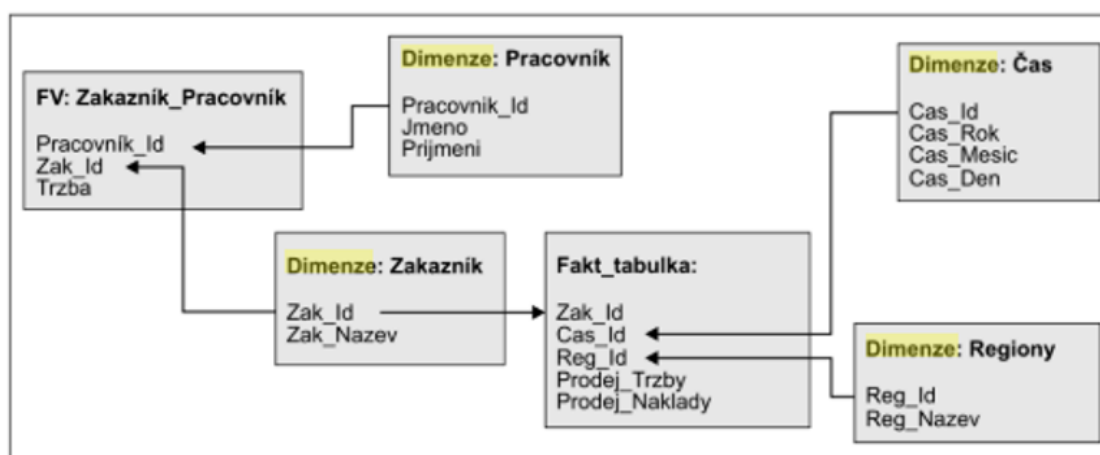
**Tabulka 1: Porovnání charakteristik jednotlivých typů tabulek faktů**

Zdroj: vlastní zpracování podle Pour a kolektiv 2018

### 3.3.2.3 Dimenze

Tabulky dimenzí obsahují popisná data, která jsou rozmístěna do velkého počtu atributů. Tato data následně slouží ke sledovaným faktům. Primární klíč má vždy tabulka dimenzí jen jeden, pomocí primárního klíče se propojuje s faktovými tabulkami, kde je klíč vedený jako cizí. Pomocí dimenzí lze reprezentovat vztahy hierarchie, jde o vztahy 1:M mezi jednotlivými atributy dimenze. Hierarchie je řešena pomocí normalizovaných modelů, kterými jsou sněhová vločka a hvězda. (Zádová 2014)

Dimenzionální modelování je založeno na vztazích 1:N, kde každý řádek pak odpovídá jedné hodnotě ve faktové tabulce. M:N vztahy, které nejsou nutností, vznikají i tak a lze je řešit pomocí přidání pomocné tabulky, která tento vztah upřesňuje. (Pour a kolektiv 2018)



Obrázek 13: Řešení vztahů dimenzí M:N  
Zdroj: Pour, Maryška, Šedivá, Stanovská 2018



Atributy dimenzí mají dle V. Zádové (2014) následující formy:

- Dimenzionální – úrovně dimenzí, určení agregační úroveň
- Nedimenzionální – vlastnosti atributů charakterizují blíže úrovně, jsou volitelné

Jak vysvětluje O. Zýka (2016), základem integrace datového skladu jsou konformní dimenze, kde podmínkou jsou alespoň dvě dimenze, kde jejich atributy jsou stejně pojmenované a také mají stejný význam. Konformní dimenze umožňují propojení různých tabulek faktů. Existují také specifické dimenze, mezi které patří:

- Přidané dimenze
- Dimenze času
- Degenerované dimenze
- Junk dimenze (seskupené/sběrné)
- Parent-child dimenze

Přidané dimenze vznikají, jakmile je nutné vytvořit vazbu 1:M napřímo mezi dimenzemi. Reprezentují se jako atributy vyčleněné z původní dimenze.

(Zádová 2014)

Dimenze času je vždy definovaná, a to kvůli tomu, že vždy se v praxi sleduje vývoj nějakých ukazatelů v čase. Vyjádření hodnoty dimenze je nejvíce využíváno ve formátu rok-měsíc-den. Dimenze času neobsahuje jen datum, ale také čas, z toho důvodu se dimenze času dělí na dvě dimenze, kde jednou je právě dimenze data (rok-měsíc-den) a druhou je dimenze času (hodina-minuta-sekunda). Z důvodu opakující se dimenze času každý den, a to kvůli její stejné struktuře, by nemělo dojít ke sloučení právě již zmíněných dimenzí. Pomocí těchto dimenzí je možné se dotazovat na různé časové intervaly. V dimenzi se také využívají atributy označující stanovené časové jednotky, které následně slouží pro reportování a výpočty. (Pour a kolektiv 2018)

Degenerovaná dimenze je dimenze, která nemá popis, to znamená, že nemá žádný obsah. Jediné, co tato dimenze vlastní je jen její primární klíč. Její využití je pouze v transakčních tabulkách, kde její atribut určuje řádek dané operace. (Zádová 2014) Junk neboli sběrné dimenze vznikají při analýze jako uskupení nesouvisejících atributů s nízkou kardinalitou, které nelze přiřadit k nějaké dimenzi. Jsou tedy vyjmuty z faktové tabulky a následně jsou vloženy do samostatné dimenze. (Pirkl 2004)

Dimenze dále podle O. Zýky (2016) rozlišujeme podle pohledu změn následovně:

- Statické
- Rostoucí
- Rychle rostoucí
- Pomalu se měnící

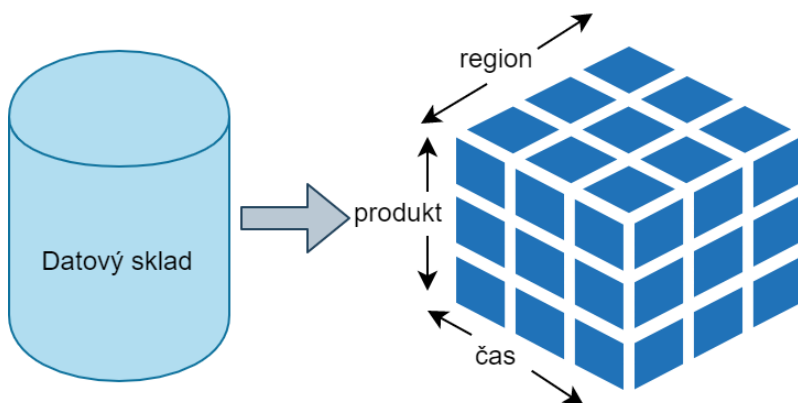
Statické nedisponují žádným ošetřením, takže v případě změny se starší záznamy přepíše. Statické dimenze nemají k dispozici ani historii. Rostoucí dimenze umožňují přidávat nové záznamy a další vlastnosti přebírá ze statické, tedy nemá historii a při změně se přepisuje starý záznam. Rychle rostoucí dimenze umožňují změny atributů častěji, třeba denní formou. (Zýka 2016)

Pomalou se měnící dimenze jsou dimenze, kde dochází ke změnám ve struktuře, jde o neobvyklé řídké změny. Provedené změny musí být sledovány a také musí být dodržena konzistentnost dat z časového (historického) hlediska. Řešení těchto změn přímo v dimenzionálních tabulkách je řešeno pomocí „typů“ v dimenzionálním modelování. Celkem existuje 8 typů, které se číslují od nuly, tedy typ 0 až typ 7. Typ 0 veškeré změny ignoruje a dále se nepromítají. Typ 1 je nejjednodušší formou, nahrazuje původní hodnoty atributů novou hodnotou, která je změněna ve stejném řádku dimenzionální tabulky. Následujících šest typů umožňuje větší změny dimenzionální tabulky, změny atributů s přidáním nové hodnoty, přidání dalšího atributu, přidání mini dimenze, kombinace již zmíněných vlastností a zdvojené dimenze typů. Provádění změn v jednotlivých dimenzích je časově náročné, z toho důvodu jsou využívány v každém záznamu databáze kontrolní součty, podle kterých se provádí verifikace. (Pour a kolektiv 2018)

### 3.3.3 Multidimenzionální OLAP kostka

OLAP kostka je multidimenzionální specifická metoda, která slouží pro ukládání velkého množství dat do datových skladů nebo datových tržišť. Pomocí kostky lze vytvářet pohledy, které se využívají následně při analýzách. Kostka také umožňuje na data nahlížet z různých stran a dimenzí. Datová kostka pracuje s daty uloženými v datových skladech, takže pracuje již s očištěnými daty. Reakce při práci s datovou kostkou jsou rychlé, a to díky předem provedeným výpočtům na datech, které se v kostce nachází. Není třeba se složitě dotazovat, uživatel má data k dispozici ihned, kdykoli je jen potřebuje. V této diplomové práci budeme také pracovat s datovou kostkou, která bude použita pro pozorování změn stravovacích návyků a také pohybových aktivit v průběhu času.

Kostka je tvořena tabulkou faktů a tabulkou dimenzí, kde fakta reprezentují hodnoty kostky a dimenze jednotlivé strany kostky. Nejčastější použití kostky je při zvolení třech dimenzí, jde tedy o tři strany kostky. Kostka může mít ale také více dimenzí, záleží jen na potřebě použití. (Lacko 2003, Berson 1997)

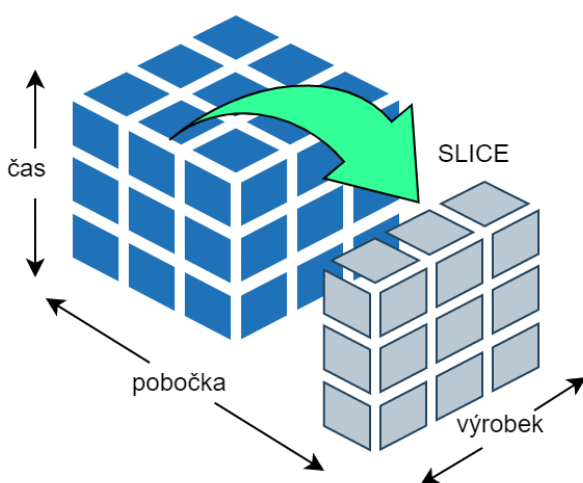


**Obrázek 14: Datová kostka OLAP**  
Zdroj: vlastní zpracování

### 3.3.4 Operace datové kostky

OLAP kostky umožňují několik základních operací.

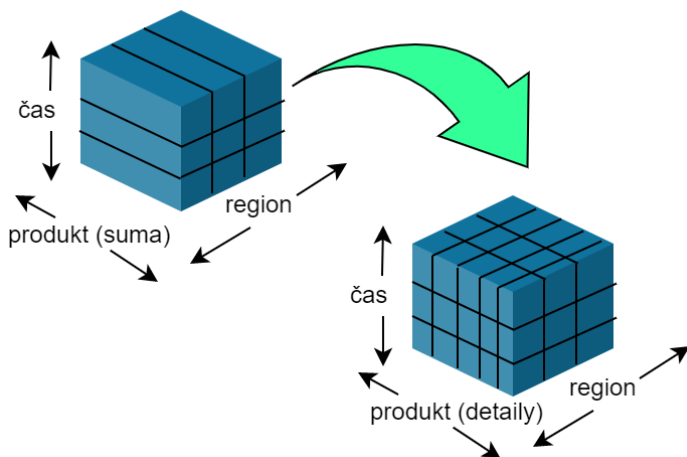
**Krájení a kostkování** (anglicky slicing and dicing) umožňuje omezit jednu či více dimenzí na podmnožinu o jednom i více prvcích. Typický příklad může být proveden na kostce viz. obrázek (Obrázek 15) níže, kde z modelové kostky odřízneme jednu či více poboček.



**Obrázek 15: Datová kostka - operace slice**  
Zdroj: vlastní zpracování

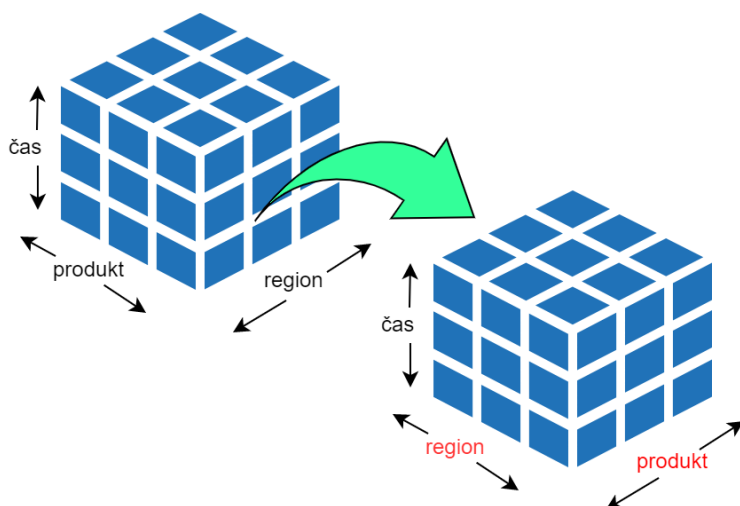
**Agregace** představuje operaci, která slouží ke konsolidaci dat pomocí určitého vzorce. Na našem příkladu se může jednat o součet prodaných produktů ve všech pobočkách.

**Roll up, drill down** operace se používají k postupné navigaci nahoru a dolu v hierarchii dat. Lze tedy hledat data do značné hloubky.



**Obrázek 16: Datová kostka – operace drill-down, drill-up**  
Zdroj: vlastní zpracování

**Pivotování**, česky otáčení, je proces, který umožňuje otáčet kostku k získání jiného pohledu, perspektivy na stejná data. Příkladem zde může být pohled, kde místo na počet prodaných produktů na pobočkách se podíváme na počet prodaných produktů na jednotlivých pobočkách.



**Obrázek 17: Datová kostka – operace pivotace**  
Zdroj: vlastní zpracování

### **3.4 ETL (Datová pumpa)**

ETL neboli datová pumpa je definována jako proces, při kterém se převádí data z různých provozních systémů, kde se vyskytují relační databáze, do datových skladů či tržišť. V podstatě jde o převod z relačních databází do analytických databází. ETL proces ještě před uložením dat data nejprve zpracuje a očistí.

(Dohnal, Pour 1997)

ETL není zdaleka jednoduchý proces, jak se může zdát. Vyžaduje aktivní vstupy od různých stran, kterými jsou vývojáři, testeři, manažeři a další, kteří s daty pracují. Tento proces se opakuje v pravidelných intervalech, denní, týdenní či měsíční, a pro správný chod musí být tento proces automatizován, dobře zdokumentován a naimplementován.

Mezi nejvýznamnější důvody, proč využívat proces ETL podle Guru99 (2019) patří:

- Možnost získávání a přesunu dat z různých zdrojů přímo do datového skladu
- Při změně zdrojů dat se datový sklad automaticky aktualizuje
- Možnost povolit ověřování pravidel pro transformaci, agregaci a výpočty dat
- ETL umožňuje porovnání vzorových dat mezi zdrojem a cílovým systémem
- Proces ETL může provádět složité transformace
- Pomoc při migrování dat do datového skladu a převod na různé formáty a typy – zaručení konzistentního systému
- Jde o přístup a manipulaci se zdrojovými daty do cílové databáze
- Nabízí hluboký historický kontext
- Zvýšení produktivity

### 3.4.1 Proces ETL

ETL proces je složen ze tří kroků, kterými jsou extrakce, transformace a nahrání.

#### 3.4.1.1 Extrakce

Extrakce je prvním krokem ETL procesu, kde se z stahují data z různých zdrojových systémů a následně jsou data ukládána do DSA (režijního úložiště). Je nutné dbát na správné použití filtrace obsahu dat a jaká data také číst. Datový sklad musí integrovat systémy, které mají různý hardware, operační systémy a komunikační protokoly. Dále se také musí přizpůsobit starším aplikacím, které mohou mít zdrojové systémy. Ještě před samostatnou extrakcí dat se provádí mapování. Pomocí mapování lze získat logické členění dat a popis vztahů mezi zdrojovými a cílovými daty. (Kimball 2013)

Guru99 (2019) popisuje, že existují tři metody pro extrakci dat a následně také popisuje jejich kroky pro ověření:

- Úplná extrakce
- Částečná extrakce
  - Bez oznámení aktualizace
  - S upozorněním na aktualizaci

Při správné implementaci by bez ohledu na použitou metodu extrakce neměla ovlivnit výkon ani dobu odezvy zdrojových systémů.

Při průběhu extrakce se provádí nutná ověření:

- Porovnání záznamů se zdrojovými daty
- Detekce nevyžádaných dat
- Kontrola datových typů
- Odstranění duplikovaných a fragmentovaných dat
- Kontrola klíčů

(Guru99 2019)

### 3.4.1.2 Transformace

Po extrahování dat ze zdrojů následuje další krok extrakce, kterým je transformace. Jelikož data extrahovaná ze zdrojového serveru jsou nezpracovatelná a nepoužitelná v původní podobě, je třeba provést právě transformaci. Pokud se provádí transformace, dochází k provedení akce ve Staging oblasti, kde tím není výkon zdrojového systému zatěžován. Výhodou zde je, že pokud jsou zkopírována poškozená data přímo ze zdroje do databáze, tak následuje jejich vrácení zpět. Staging oblast poskytuje příležitost ověřit extrahovaná data ještě před jejich přesunem do datového skladu. (Sodomka a Klčková 2017)

Kimball (2013) uvádí, že fáze transformace dat je klíčovou, a to z hlediska kvality dat, kde z toho důvodu není již nadále nutné se kvalitě dat věnovat. Dále píše také o pěti hlavních činnostech, které se provádí u transformace:

- Čištění dat
- Sledování a evidence chyb
- Vytvoření auditní dimenze
- Deduplikace
- Harmonie dat

Nad daty se tedy provádí sady funkcí, které data připraví pro uložení do úložiště. V některých případech mohou být data ve správných formátech a není třeba na ně aplikovat transformace. Při transformaci se mohou provádět například změny sloupců nebo jejich zřetězení a další. Podle M. Schillera (2003) se mohou také při transformacích objevit leccjaké problémy:

- Hláskování, jazyky (Jan, Honza, Jenda)
- Různé formy názvů (IKEA vs IKEA Group)
- Různé formáty čísel
- Prázdné řádky tabulek
- Neplatná data



Guru99 (2019) zmiňuje možné operace, pomocí kterých lze zmíněné problémy od M. Schillera řešit a následně data validovat:

- Filtrace dat – možnost zobrazení určitého obsahu
- Použití pravidel na vyhledávání obsahu – docílení standardizace dat
- Problémy se znakovými sadami, převody jednotek a měn, data a času
- Povinná pole by neměla být prázdná
- Čištění a oprava dat – NULL jako 0 a pohlaví značit jako M a F
- Rozdělení či sjednocení sloupců
- Transponování řádků a sloupců
- Komplexní validace dat (globální pravidla)

#### **3.4.1.3 Načtení (Load)**

Poslední etapou ETL je načtení (uložení) dat. Zde se provádí již samotné nahrání upravených, konsolidovaných a očištěných dat do cílových datových skladů.

V typickém datovém skladu je nutné načíst obrovský objem dat v relativně krátké době. Proces načítání by proto měl být optimalizován pro výkon. (Kimball 2013)

Při načítacím procesu může dojít v některých případech také k selhání načtení. Z tohoto důvodu by měly být obnovovací mechanismy nakonfigurovány tak, aby se restartovaly z místa selhání bez ztráty integrity dat a následně se pokračovalo v načítacím procesu dále. (Guru99 2019)

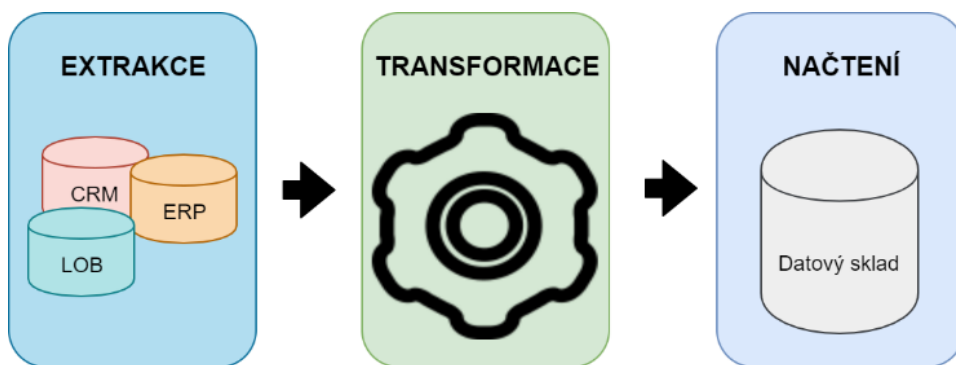
T. Jörg a S. Dessloch (2009) rozdělují načítací procesy podle jejich typů:

- Počáteční – dojde k naplnění všech tabulek v datovém skladu
- Přírůstkové – dochází k pravidelným změnám podle potřeb a aktualizací
- Celkové obnovení – vymaže se obsah jedné nebo více tabulek a dojde k opětovnému načtení nových dat

Při načítání a po jeho dokončení probíhají ještě kontrolní skripty, které se starají o

- Správné umístění klíčů
- Kontrola nulových dat
- Kontrola provedených kombinací a výpočtů
- Kontrola dimenzí a historie tabulek
- A další...

(Guru99 2019)



**Obrázek 18: ETL proces**  
Zdroj: vlastní zpracování

### **3.5 Data Mining**

Jde o termín, který je do českého jazyka překládán jako dolování dat. Data mining je nejednoduchý proces, který prohledává, vybírá, analyzuje a modeluje z velkých objemů dat. Jde v podstatě o dohledání doposud „neznámého“, co může být klíčové pro lepší úspěch například obchodní firmy nebo získání výhody oproti konkurenci. (Piatetsky-Shapiro 1996)

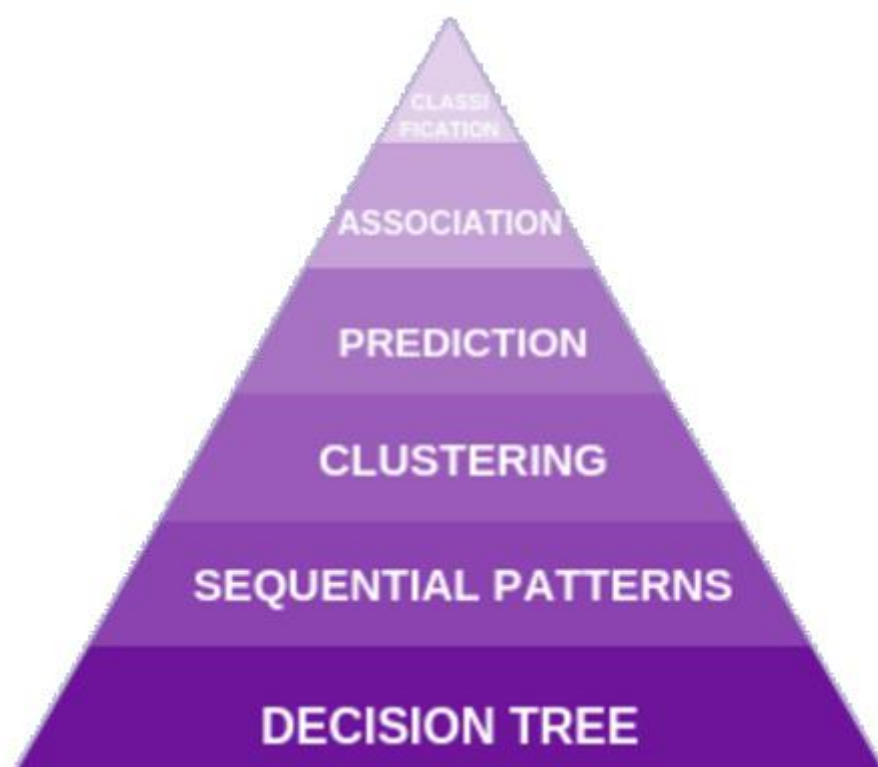
Při dolování dat by měl být vždy na začátku znám cíl, který specifikuje, čeho se má dosáhnout. Dolování lze také pojmově jako snahu najít v obrovském objemu dat nějaký vzor, který má vliv nebo vztah s ostatními daty ve stejné sféře.

(Mařík, Štěpánková, Lažanský 2003)

Proces získávání dat, jak uvádí J. Šarmanová (2002), se skládá z těchto následujících kroků:

- Specifikace a formulace problému
- Analýza z hlediska zkoumaného problému a dat
- Vyčlenění a výběr relevantních dat
- Fáze předzpracování (integrace, transformace)
- Dolování dalších hypotéz
- Interpretace dosažených výsledků
- Celkové vyhodnocení

Dolování dat se dnes v mnoha případech provádí nad datovým skladem, datové sklady obsahují velké objemy dat a také historická data, která jsou potřebná pro další analýzy. Do datových skladů jsou ukládána data, která jsou zkontrolována, jsou také vyšší kvality a jsou očištěny od ostatních nezajímavých předmětů. Dnes díky kladenému důrazu na rozvoj Big Data není v určitých případech možné využít v minulosti často používané metody dolování dat. Jednou z těchto klasických metod je právě statická metoda dolování dat, a to z důvodu ohromného nárůstu právě daných dat, pro které není tato metoda stavěná a nedokáže tak data zpracovat. (Witten 2016)



Obrázek 19: Techniky dolování dat  
Zdroj: Learntek 2019

Dle obrázku (Obrázek 19), jak uvádí Learnet (2019), se při dolování dat používají následující techniky:

- **Klasifikace** – přiřazení položek do cílových kategorií nebo tříd, které se používají k předpovídání toho, co se ve třídě přesně stane, klasifikuje každou položku v sadě dat do jedné z předdefinované sady tříd nebo skupin.
- **Asociace** – relační technika, rozpoznávání na základě vztahu položek v jedné transakci
- **Predikce** – hledání vztahů mezi nezávislými a závislými proměnnými
- **Shlukování** – seskupování souvisejících záznamů do jedné skupiny
- **Vzorkování** – technika identifikující pravidelné výskyty podobných událostí
- **Rozhodovací strom** – kategorizace nebo predikce dat, kořenem stromu je vždy jen jednoduchá otázka

### **3.6 Data Culture**

Data Culture, český překlad „Kultura dat“, je relativně nový koncept, který se v dnešním dění čím dále tlačí dopředu a stává se tak klíčovým. V podstatě jde o jedno velké pracovní prostředí, které využívá důsledný přístup k rozhodování, a to prostřednictvím důrazných a empirických údajů. Rozhodování se zde tedy dělá na základě striktně daných údajů. (Sciforge 2019)

Klíčovými koncepty, jak dále popisuje Sciforge (2019), jsou zde:

- Data Culture lze pojmu jako rozhodovací nástroj. Pomocí podrobných analýz dat a jejich základním cílům shromažďovat, zpracovávat a nasazovat data.
- Demokratizace dat zajišťuje přijetí lidí. Otevřený přístup k datům pomáhá lidem věřit, a tak se podílet na různých řešeních, které nevyžadují odbornou způsobilost. Jakmile je zaručena důvěra, lidé se začnou chovat úplně jinak, tento jev je právě cílem data culture.
- Řízení rizik je jednou z nejdůležitějších součástí datové kultury. Pomocí analýz klíčových procesů dojde k rozvoji společností a k pochopení toho, co je důležité. Na druhou stranu zde existuje riziko, že dojde k nesprávné analýze, kde společnosti na to musejí být připraveni a upozornit tak své okolí.
- Lidé jsou zde jedním z nejdůležitějších faktorů. Řízení lidských zdrojů je centrem pozornosti. Společnosti musejí dbát na kvalitu a dovednosti stávajících, ale také nově zaměstnaných pracovníků a také jim přiřadit odpovídající pozice.

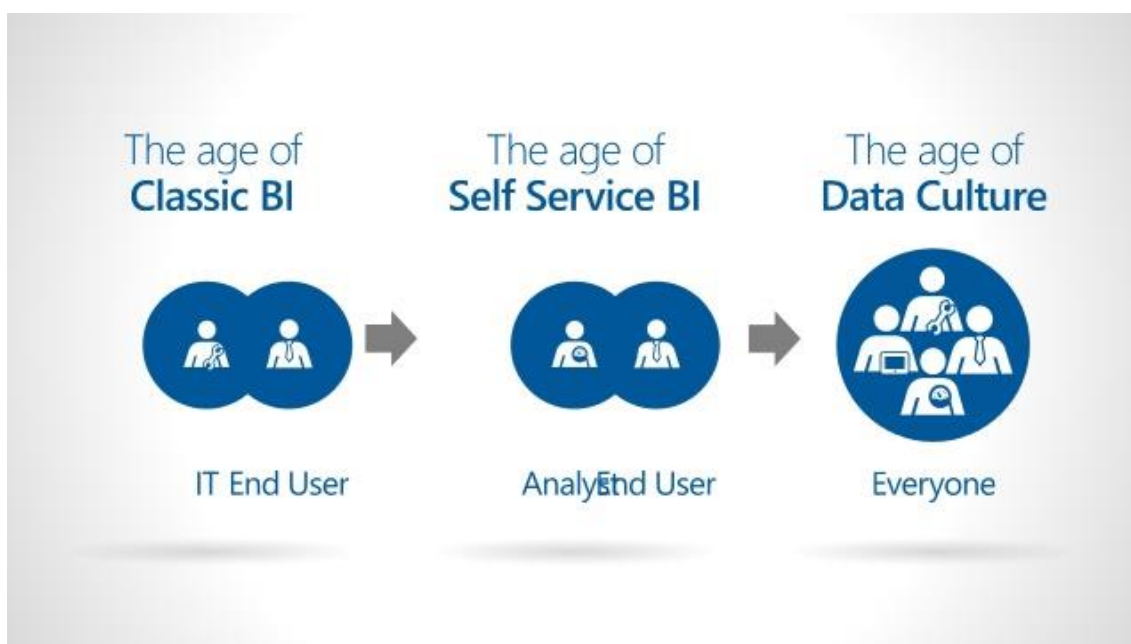
Kultura dat může být problémem nebo také řešením. Pokud jsou data organizace oddělena od obchodní strategie a základních operací, nemělo by zde být žádným překvapením, že výsledky analýz nebudou splňovat svá očekávání. Je nutné klást důraz na kvalitu zpracování podkladů pro zkoumání, a tak zaručit jejich úspěch.

(Díaz, Rowshankish, Saleh 2019)

Na otázku, proč vlastně budovat ve firmě kulturu dat zmiňuje DataTherapy (2017) pár klíčových vlastností:

- Data se nejčastěji používají pro zlepšení operací a jejich výsledků
- Větší využití dat k šíření zpráv, což vede k infografice – zobrazení dopadu programů
- Přivést a sblížovat lidi – umožnit tak sblížení lidí z odlišných sektorů a zajistit tak lepší komunikaci

*„Data Culture je postupný proces, který vyžaduje změnu návyků, postojů, zaměstnanců, a dokonce i zdrojů. Ale jakmile bude tento nový směr zaveden, bude mnohem snazší udržet a rozšířit danou organizaci.“ (Weiner 2019)*



**Obrázek 20: Data Culture – sjednocení z různých pozic**  
Zdroj: Cristian G. 2018

## 4 Analytické a reportovací aplikace

V Business Intelligence se využívá mnoho aplikací, které v dnešní době znatelně ulehčují vývoj a rozhodování v prostředí BI. Aplikace využívané v BI jsou ve většině případů pro komerční účely, ale lze také využít několik nekomerčních aplikací, takzvaných open source programů. Aplikace jsou děleny na dva základní typy, kterými jsou:

- Analytické (statistické) aplikace
- Reportovací (prezentační) aplikace

Většina aplikací všechny výše vyjmenované typy obsahují již v sobě jako jeden balíček.

### 4.1 Analytické aplikace

Analytické aplikace jsou klientské aplikace, které jsou navrženy speciálně pro poskytování informací ze získaných dat. Disponují dalšími funkcemi, kterými jsou prohlížení dat, agregování pohledu na data, sledování trendů, provádění analýz, identifikaci specifičností a další. (Gála, Pour, Šedivá 2009)

Analytické aplikace by měly být snadno použitelné a univerzální. Většina analytických aplikací disponuje grafickým rozhraním, ale existují i takové aplikace, které grafické rozhraní nemají.

Příklady aplikací, které obsahují analytické nástroje:

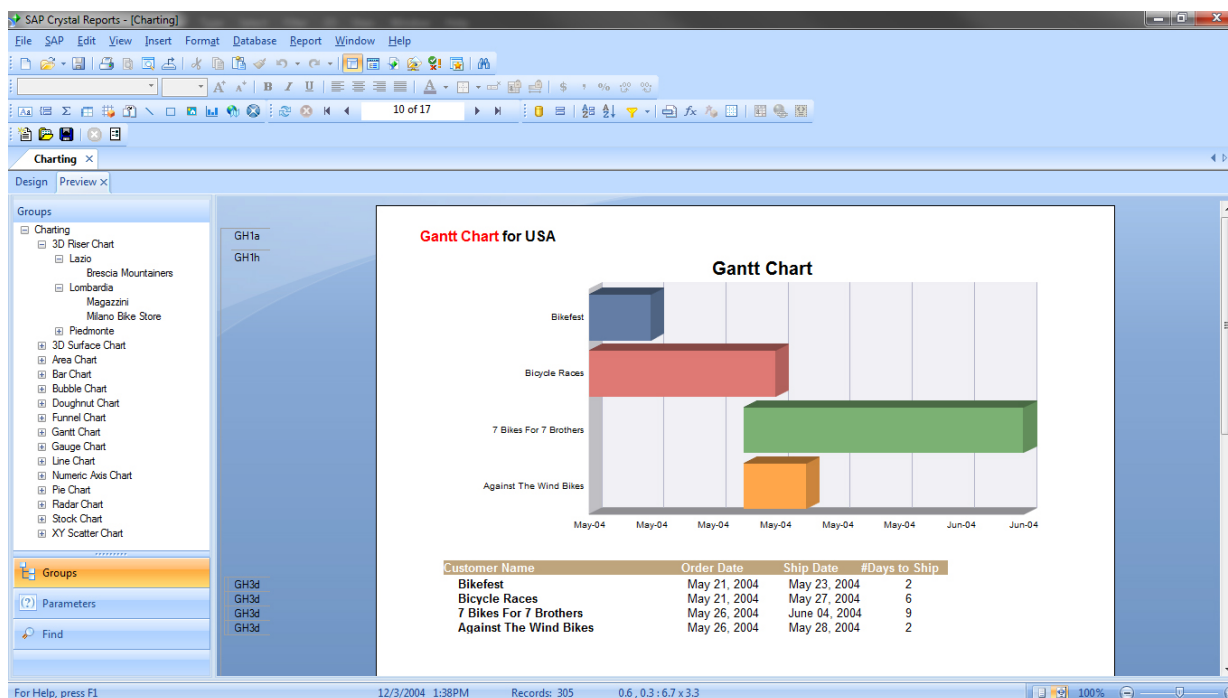
- SAP BusinessObjects BI
- SAP Analytics Cloud
- Microsoft Power BI
- Microsoft Access
- IBM Cognos BI
- Oracle Business Intelligence
- A další ...

## 4.2 Reportovací aplikace

Reportovací aplikace umožňují reprezentaci a vizualizaci dat, které jsou uloženy v datových skladech. Pro vizualizaci dat jsou využívány především grafy a tabulky. Je zde zpracováno velké množství dat do značné hloubky a následně po zpracování dat dojde k jejich vyhodnocení, kde pomocí výsledků lze přijímat strategická rozhodnutí s předvídatelnými dopady.

Aplikace, které umožňují přípravu a práci s reporty:

- GoodData
- MS Power BI
- Jet Reports
- SAP Crystal Report
- A další ...



Obrázek 21: SAP Crystal Report  
Zdroj: Sap 2016



## **5 Praktická část**

Praktickou částí této diplomové práce je rychlé představení firmy a databázové aplikace Kalorické tabulky, také popsání webové a mobilní aplikace, která je určená pro uživatele. Dále pak tvorba datového skladu, práce s ETL procesem, připojení k datům, jejich popis a možnosti stažení, a v konečné fázi pak tvorba základních dashboardů pro možnosti analýz.

### **5.1 Popis firmy**

Firma Dine4Fit provozuje a spravuje databázovou aplikaci s názvem Kalorické tabulky. Aplikace Kalorické tabulky umožňuje evidovat jídla, aktivity, tvořit recepty, spojit se s přáteli či dietními poradci. Aplikace dále poskytuje základní přehled o daném uživateli, kterým je vývoj hmotnosti a BMI, pitný režim, grafy denních příjmů jednotlivých složek jídel a další. Jde o středně velkou firmu, která se věnuje oblastem jako jsou informační technologie (databáze, programování), výživové poradenství a HR. Zaměstnanci se snaží neustále zapojovat do nových řešení aplikace, jak lze vidět podle aktualizací na Obchodu Play či AppStoru. Aplikace je tvořená pro uživatele a uživatelé zde mohou podávat zpětné vazby. Pracovníci si zpětných vazeb velice váží, a tak na ně rychle reagují a provádí patřičné změny či různá vylepšení na žádost uživatelů. Jde o velmi oblíbenou aplikaci, kde počet registrovaných účtů přesáhl již 3 miliony.

## **5.2 Aplikace Kalorické tabulky**

Aplikace je rozdělena do dvou platforem, kde jednou z nich je webová aplikace a druhou variantou je mobilní aplikace, která je určena pro zařízení s operačním systémem Android a iOS. Pro využití aplikace je nutné provést registraci, který je možná udělat pomocí Facebook či Google účtu nebo pomocí emailového účtu. Následně je uživatel vyzván pro vyplnění některých důležitých parametrů, se kterými aplikace nadále pracuje a provádí dle nich také výpočty.

Jednotlivé parametry osoby jsou:

- Pohlaví
- Výška
- Hmotnost
- Rok narození
- Nastavení osobních cílů
- Pitný režim
- A další

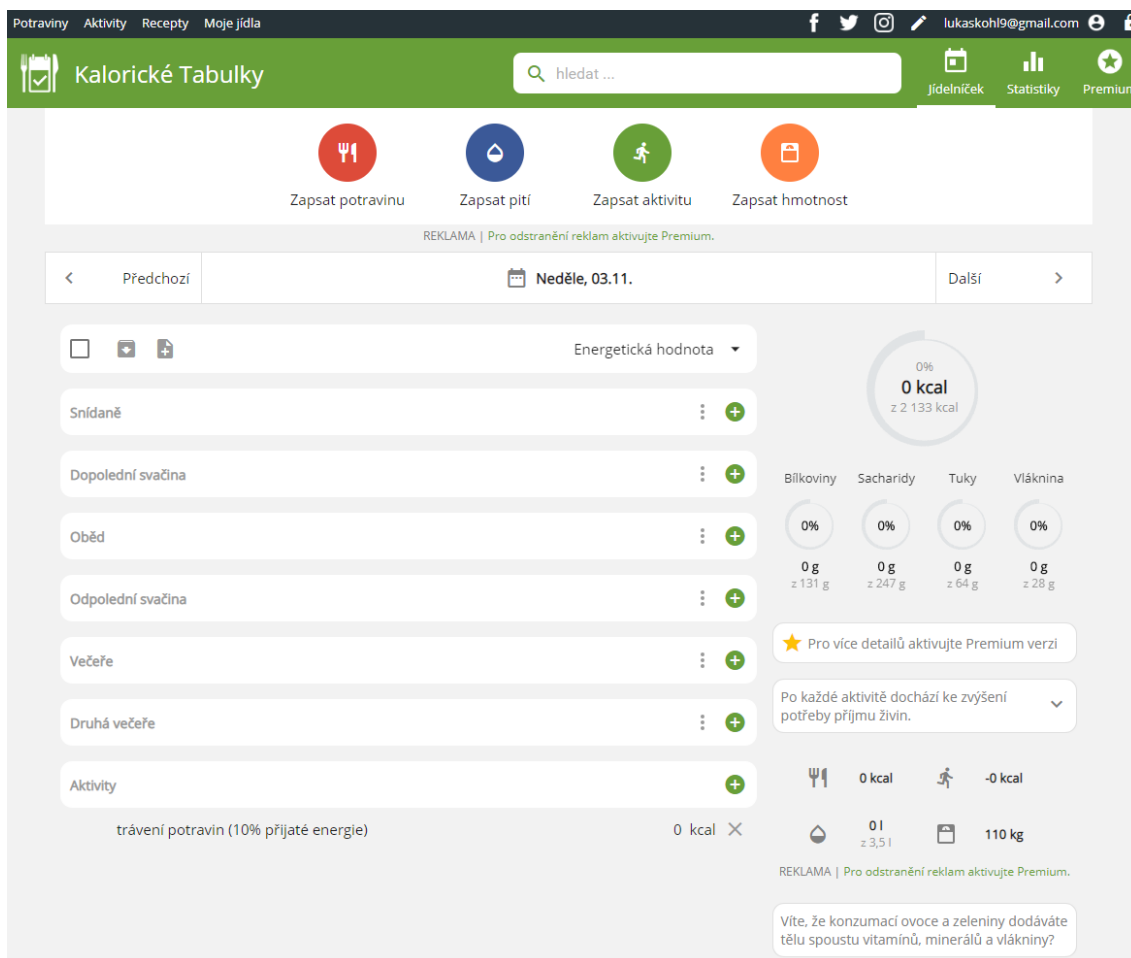
Výše uvedené údaje budou v další části využity jako sociodemografické charakteristiky jednotlivých osob.

Aplikace umožňuje evidovat jídelníček (potraviny) a také aktivity. Dalšími možnostmi je využívání receptů a také sdílení jídelníčků s uživateli, kteří jsou vedeni jako výživový poradci nebo jen s přáteli. Pro odemčení některých funkcí, jako je právě sdílení nebo podrobná analýza jídelníčku za krátké i delší období, je nutná aktivace premium účtu.

Hlavními prvky ať už ve webové či mobilní verzi jsou:

- Zapsání potravin
- Zapsání pití
- Zapsání aktivity
- Zapsání hmotnosti

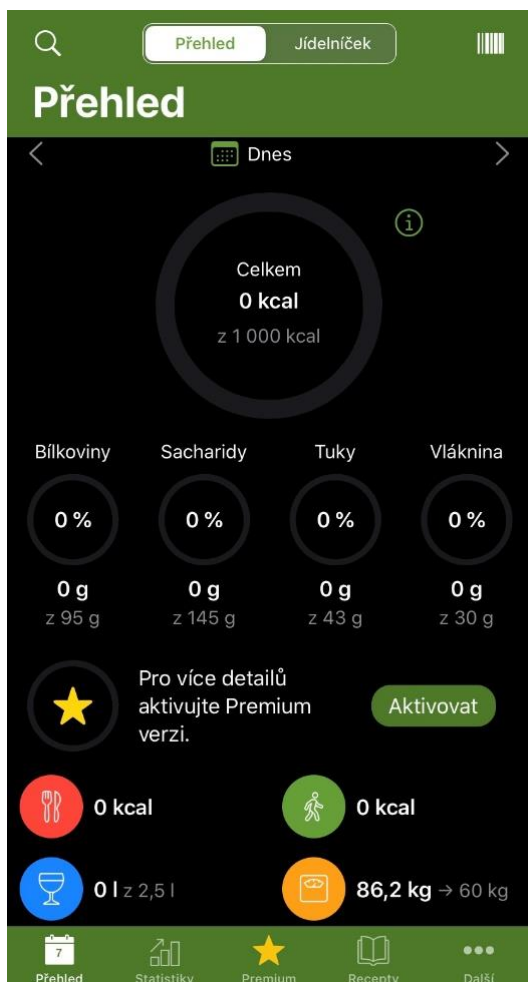
Jídelníček je dělen do šesti etap, kterými jsou: snídaně, dopolední svačina, oběd, odpolední svačina, večeře, druhá večeře. Toto rozdělení na jednotlivé etapy je ideální pro správné a častější stravování. Aplikace<sup>1</sup> disponuje nahlížením do historie jídelníčku či aktivit a také statistik, které jsou prováděné na pozadí. V modulu statistik lze vidět graf hmotnosti vyvíjející se v časovém horizontu, dále pak denní a týdenní sumarizaci příjmu, pitného režimu a aktivit.



**Obrázek 22: Kalorické tabulky – webová platforma**  
**Zdroj: Kaloricketabulky.cz**

<sup>1</sup> Webová aplikace Kalorické tabulky - <https://www.kaloricketabulky.cz/>

Na mobilní verzi aplikace je zachováno stejné uspořádání prvků jako na webové verzi, jedná se o pravou spodní část webové verze. Uživatel má tak ihned přehled pomocí moderního designu s grafy v podobě kruhů o svých příjmech, pitném režimu či aktivitách. Viz obrázek (Obrázek 23) níže.



**Obrázek 23: Kalorické tabulky – mobilní platforma**  
**Zdroj: Mobilní aplikace Kalorické tabulky**

## 5.3 Přístup k datům

Přístup k datům byl udělen na základě smlouvy o mlčenlivosti, kde se pomocí přihlašovacích údajů přihlašuje do datového skladu Kalorických tabulek, který je skladován v GoodData.<sup>2</sup>

Pro přístup k datům je nutné využít Java SQL klienta, kde se následně využívá možnosti připojení pomocí JDBC ovladače<sup>3</sup>. V praktické části jsou pro nahlížení a stahování dat využívány programy Squirrel, DBeaver a DataGrip.

### 5.3.1 Struktura dat

Data jsou v datovém skladu GoodData uspořádána do třech typů tabulek, které jsou rozpoznatelné pomocí prefixů v jejich názvu. Jde o prefixy s\_\*, p\_\* a o\_\*.

Prefix s\_\* jsou staging tabulky, do kterých se nahrávají vstupní data buď inkrementálně nebo všechna data.

Prefix p\_\* jsou persistentní tabulky, ve kterých se data mergují ze staging tabulek.

Prefix o\_\* jsou output stage tabulky, kam se data dostávají po provedení transformací nad daty z persistentních tabulek.

Pro účely této diplomové práce budou využívány pouze persistentní tabulky. Persistentní data obsahují aktuální data z produkce.

Následující výčet důležitých tabulek tvořících databázi Kalorických tabulek, se kterými bude nadále pracováno:

- Tabulky aktivit a potravin
- Tabulka typ aktivit a potravin
- Tabulka deníků aktivit a potravin
- Číselníková tabulka
- Tabulka uživatelů
- Tabulka měření hodnot uživatel

---

<sup>2</sup> GoodData - <https://www.gooddata.com/>

<sup>3</sup> GoodData JDBC - <https://secure.gooddata.com/downloads.html>

### 5.3.2 JDBC

JDBC je technologie, která je základem J2EE a slouží pro práci a připojení k databázím. Zkratka JDBC je v plném znění Java Database Connectivity, jde tedy o Java databázový konektor. (Šeda 2003)

Může být také označován jako Java API, které se stará o připojení k databázi, dotazy a příkazy a také zpracování výsledků. JDBC vznikl v roce 1997, byl veden jako součást JDK 1.1., kde byl v persistentní vrstvě. Jelikož je JDBC součástí Javy, lze s ním pracovat na různých operačních systémech jako je Windows, MacOS a další UNIXově založené systémy. (Tyson 2019)

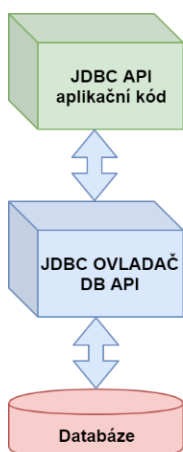
Novější verze jako je JDBC 4.0 obsahují dva balíčky, kde jedním z nich je balíček na straně klienta, označován jako `java.sql`, a druhý balíček je na straně serveru, který je označován jako `javax.sql`. Tato verze také obsahuje mnohá vylepšení a nové funkce, kterými jsou dle Tutorialspoint (2019):

- Automatické načítání databázového ovladače
- Vylepšení na straně zpracování výjimek
- Vylepšené BLOB / CLOB funkcionality
- Vylepšení rozhraní pro připojení a příkazy
- Podpora národní znakové sady
- SQL ROWID přístup
- Podpora datového typu XML 2003
- A také anotace

JDBC navíc není určen jen pro přístup k relačním databázím, ale umožňuje se připojit k jakémukoli formátu, který je sloupcové podoby. Jedná se o spreadsheetsy nebo textové soubory. (Šeda 2003)

### 5.3.2.1 Architektura JDBC

JDBC byl vyvinut na základě rozhraní ODBC a stal se její alternativou. JDBC je tvořeno pomocí dvou vrstev, kde první je JDBC API, která slouží pro komunikaci mezi aplikací a uživatelem, a druhou vrstvou je zde JDBC ovladač, která daný aplikační kód umožňuje integrovat. (Tyson 2019)



**Obrázek 24: Architektura JDBC**

**Zdroj: vlastní zpracování podle Šeda 2003**

### 5.3.2.2 JDBC operace

Pomocí JDBC lze provádět také několik operací, jako například transakce, kde následně lze vypnout auto commit a provádět změny manuálně pomocí příkazu `connection.commit()`; JDBC dokáže pracovat dávkově, což zvyšuje rychlost a výkon. Lze také zvýšit flexibilitu programu pomocí nahrazení Statement tříd za PreparedStatement třídy, tímto se také zvýší znovupoužitelnost kódu. (Tyson 2019)

### 5.3.2.3 Rozdělení JDBC

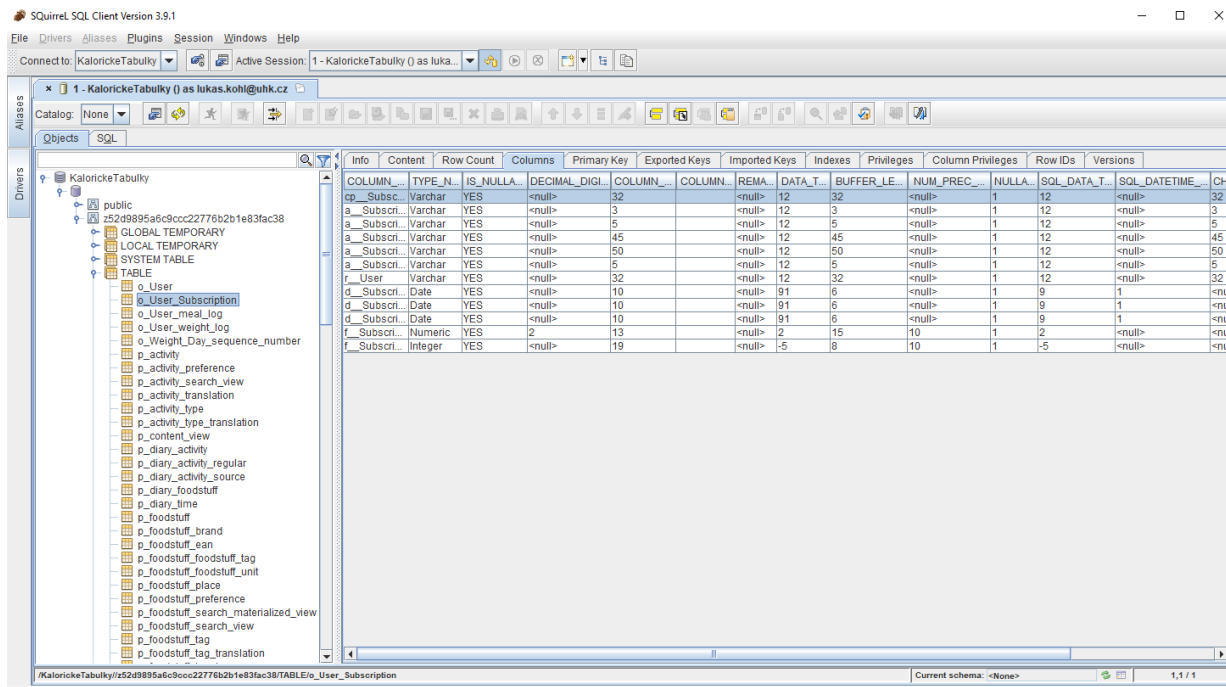
JDBC lze tak, jak popisuje Šeda (2003), rozdělit do čtyř typů ovladačů. Kde prvním typem je myšlen lokální ODBC ovladač, ke kterému se přistupuje pomocí JDBC-ODBC mostu. Druhým typem je ovladač, který se stará o překládání požadavků z JDBC do jiného nativního ovladače, který je daný pro příslušnou databázi. Třetí typ je čistě založený Java ovladač, který umožňuje komunikovat přes síťový protokol. A posledním čtvrtým typem je zde ovladač v jazyce Java, který poskytuje plnohodnotnou podporu pro optimalizace dané databáze. (Šeda 2003)

### 5.3.3 Java databázový klient Squirrel

Squirrel<sup>4</sup> je databázový grafický klient založený na Java. Pomocí tohoto programu lze zobrazit struktury databáze, které jsou s JDBC kompatibilní. Dále lze vytvářet a spouštět SQL dotazy a příkazy, nebo se procházet ve struktuře a datech jednotlivých tabulek. Požadavkem pro práci s tímto SQL klientem je Java 1.8 a vyšší. Squirrel je výhodný v tom, že lze program rozšiřovat pomocí několika set pluginů, ať pro vytváření složitějších dotazů nebo procházení a vytváření datového skladu, ale také mnohem více jiných věcí. Squirrel je původem vydán na GNU General Public License, který byl následně od verze 1.1 nahrazen GNU Lesser General Public License. (Squirrel 2019)

Pomocí Squirrelu byla procházena struktura databáze Kalorických tabulek a následně provedeny jednoduché dotazy pro zobrazení několika řádků s daty pro vytvoření přehledu.

Tento program byl také využíván pro stahování a přenos dat některých dat. Data lze exportovat v CSV souboru nebo v SQL formě, která vytvoří sled SQL příkazů, nebo pomocí kopírování tabulek a následného vložení do nové databáze.



**Obrázek 25: Prostředí Squirrel databázového klienta**  
Zdroj: výstřížek z aplikace Squirrel

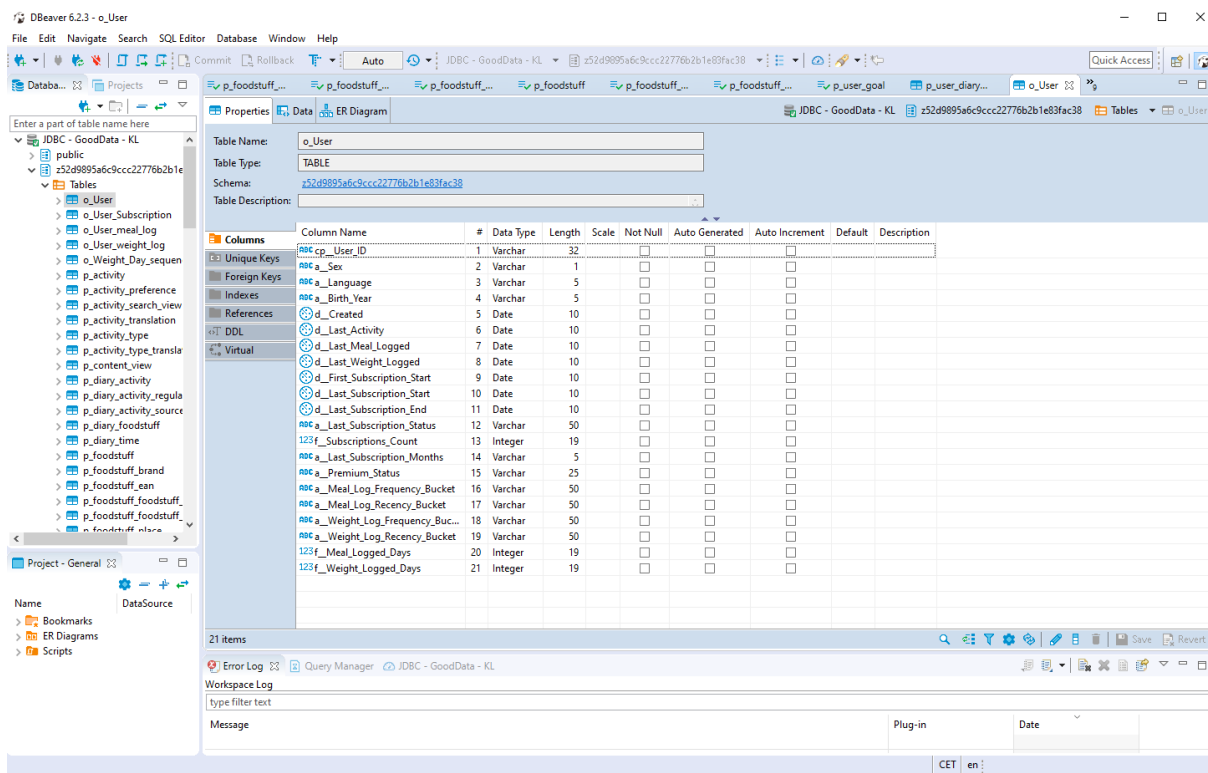
<sup>4</sup> SquirrelL - <http://squirrel-sql.sourceforge.net/>



### 5.3.4 Java databázový klient DBeaver

DBeaver<sup>5</sup> je stejně jako Squirrel bezplatný a otevřený databázový klient založený na Java, který je zaměřen pro správce a vývoj databází. Jeho grafický design je oproti Squirrelu uživatelsky přívětivější a nabízí více integrovaných funkcí a rychlejší reakce.

Tento databázový klient, jak již bylo zmíněno, je open-source, je multiplatformní a umožňuje využití různých pluginů, jelikož je založen na frameworku opensource. Podporuje všechny databáze, které mají ovladač JDBC, což je jedním z důvodů, proč byl zvolen a využíván. Pro DBeaver existuje mnoho sad pluginů pro různé databáze (MySQL, Oracle, DB2, SQL Server, PostgreSQL, Vertica, Informix, MongoDB, Cassandra, Redis, InfluxDB ve verzi 5.x). Klient obsahuje také různé nástroje pro správu databází jako je: ERD, přenos dat, porovnat, exportovat či importovat data nebo simulovat generování dat. (DBeaver 2019)



Obrázek 26: Prostředí DBeaver databázového klienta  
Zdroj: výstřižek z aplikace DBeaver

<sup>5</sup> DBeaver - <https://dbeaver.io/>

### 5.3.5 JetBrains DataGrip

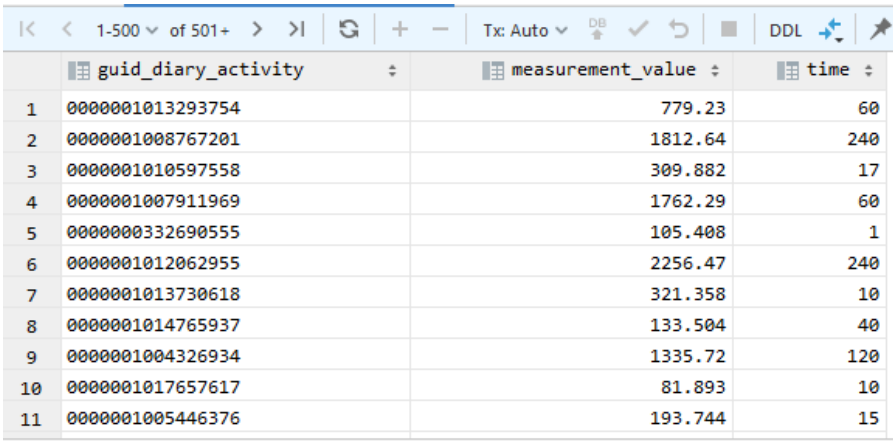
V diplomové práci byl také využit nástroj DataGrip od firmy JetBrains<sup>6</sup>. DataGrip je prostředí pro správu databází pro vývojáře. Je určen jak k dotazování, tak i k vytváření a správě databází. Databáze mohou pracovat lokálně, na serveru nebo lze pracovat i s cloudovými databázemi. DataGrip podporuje MySQL, PostgreSQL, Microsoft SQL Server, Oracle a další. (JetBrains 2019)

Jedním z důležitých vlastností, kterou DataGrip disponuje je podpora JDBC, stejně jako u dvou předchozích databázových klientů. DataGrip je zde využíván také pro jeho další vlastnost, a to možnost zafiltrování dat a následný export podle zvolených filtrů. Tabulky deník aktivit a deník jídel obsahují desítky až stovky milionů řádků, a z toho důvodu je není možné zpracovat pomocí předchozích databázových klientů. Filtrace záznamů zde probíhala pomocí specifikování rozmezí datumu.

Příklad dotazu s aplikovaným filtrem na datum v rozmezí od 1.6.2017 do 30.6.2017

```
SELECT guid_diary_activity, created_date, kj, time, time_unit,  
guid_activity, guid_user,distance, steps, activity_type  
FROM p_diary_activity  
WHERE created_date >= '2017-06-01' AND created_date <= '2017-06-30'
```

Následně po vyhotovení dotazu byl celý výsledek dotazu s filtrovanými daty vyexportován do CSV souboru a následně v rámci této aplikace také vložen do faktové tabulky. Pomocí tohoto dávkování bylo docíleno stažení celé databáze.



	guid_diary_activity	measurement_value	time
1	0000001013293754	779.23	60
2	0000001008767201	1812.64	240
3	0000001010597558	309.882	17
4	0000001007911969	1762.29	60
5	0000000332690555	105.408	1
6	0000001012062955	2256.47	240
7	0000001013730618	321.358	10
8	0000001014765937	133.504	40
9	0000001004326934	1335.72	120
10	0000001017657617	81.893	10
11	0000001005446376	193.744	15

Obrázek 27: Výsledek dotazu v aplikaci DataGrid

Zdroj: vlastní zpracování pomocí aplikace DataGrid

<sup>6</sup> DataGrip - <https://www.jetbrains.com/datagrip/>

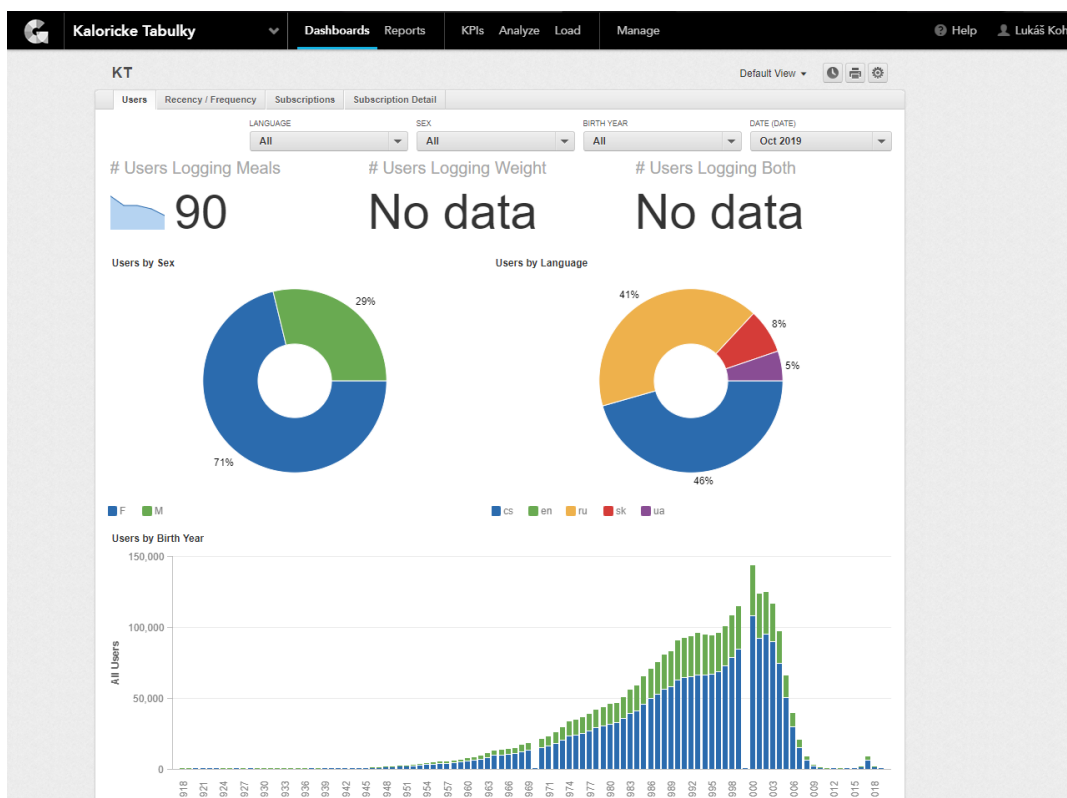
### 5.3.6 GoodData

Je webový reportovací nástroj, ale jde zároveň i o datový sklad Kalorických tabulek, ze kterých jsou čerpána data do této diplomové práce.

Platforma GoodData<sup>7</sup> je cloudový analytický nástroj end-to-end typu, který umožňuje shromažďovat a načítat data z četných a rozmanitých zdrojů dat, a poté uživatelům umožňuje vytvářet metriky, zprávy či dashboardy. Lze vytvářet také svůj datový kanál, navrhnout vlastní logický datový model a využívat nástroje pro analýzy. S platformou GoodData máte přístup k výkonným nástrojům SDK a bohaté knihovně REST API. (GoodData 2019)

Hlavní nabídka je tvořena dashboardy a reporty, dalšími vedlejšími prvky jsou zprávy a metriky. GoodData nabízí přehledné a jednoduché uživatelské rozhraní. Většina prvků je navržena tak, aby uživatelé nepotřebovali vývojářské dovednosti.

(GoodData 2019)



**Obrázek 28: Prostředí reportovacího nástroje GoodData**  
Zdroj: GoodData – Kalorické tabulky

<sup>7</sup> GoodData - <https://www.gooddata.com/>

### 5.3.7 Microsoft Visual Studio

Pomocí Microsoft Visual Studia<sup>8</sup> byly v této diplomové práci vytvořeny faktové tabulky a tabulky dimenzí, který následně byly taktéž v tomto nástroji převedeny do vizuální formy a propojeny pomocí primárních a cizích klíčů. Program byl využit také pro ilustraci datových kostek, se kterými je pracováno následně v Microsoft Power BI.

Visual Studio je program od společnosti Microsoft, který pracuje s různými jazyky, kterými jsou například: C, C++, C#, J, ASP .NET, Java script, HTML, CSS a další.

Aktuální verzí je nyní Visual Studio 2019.

Novinkami v této verzi, jak uvádí Microsoft (2019) jsou:

- Zlepšení výkonu a okamžité vyčištění kódu
- Spolupráce pomocí Git-First, úpravy a ladění v reálném čase
- Možnost zvýrazňování a rychlý přechod mezi označenými prvky

Microsoft nyní usilovně pracuje na nové verzi, které je pod názvem Visual Studio Online, jde zatím jen o předpremiéru tohoto vývojářského prostředí, které bude dostupné v rámci cloudových služeb platformy Azure. Pomocí spravovaných prostředí, které jsou k dispozici na vyžádání lze pracovat na dlouhodobých projektech či vývoji nových funkcí. Do nového Visual Studia Online se lze dostat prostřednictvím standardního webového prohlížeče a také přes již známé Visual Studio 2019 nebo také přes Visual Studio Code. Tato novinka disponuje vývojářským prostředím včetně zdrojových kódu, debuggeru a běhového prostředí. Microsoft se zaměřuje na aktuální trendy, a to pracovat ve větším týmu na dálku, a také big data. Prostředí jsou přístupná odkudkoliv a hlavními výhodami jsou rychlost přístupu, škálovatelnost a personalizace. (VT 2019)

---

<sup>8</sup> Microsoft Visual Studio - <https://visualstudio.microsoft.com/cs/vs/>

## 5.4 Schéma datového skladu

Schéma datového skladu, které bylo vytvořeno pro účely této práce je níže popsáno relačním způsobem.

Databáze se skládá ze třech hlavních tabulek, kterými jsou uživatelé, deník jídla a deník aktivit. Tyto tabulky jsou dále propojeny s dalšími tabulkami, které model rozšiřují a specifikují.

**Tabulka Uživatel** obsahuje atributy – guid\_u<sup>9</sup> (varchar), rok\_narození (int), jazyk (varchar), pohlaví (varchar). Primárním klíčem je atribut guid\_u.

**Tabulka Záznam\_uživatele** obsahuje atributy – guid\_zu (varchar), název\_měření (varchar). Primárním klíčem je atribut guid\_zu a cizím klíčem je guid\_u (uživatel)

**Tabulka Hodnota\_měření\_uživatele** obsahuje atributy – guid\_hmu (varchar), hodnota\_měření (varchar). Primárním klíčem je atribut guid\_hmu a cizím klíčem je guid\_u (uživatel) a guid\_zu (záznam uživatele).

**Tabulka Deník\_aktivit** obsahuje atributy – guid\_da (varchar), datum\_aktivity (date), kj (float), čas (float), jednotka\_času (varchar), aktivita (varchar), kroky (float), vzdálenost (float). Primárním klíčem je atribut guid\_da a cizími klíči jsou zde guid\_u (uživatel), guid\_a (aktivita).

**Tabulka Aktivita** obsahuje atributy – guid\_a (varchar), kj (float), typ\_aktivity (varchar). Primárním klíčem je atribut guid\_a a cizím klíčem je guid\_ta (typ\_aktivity)

**Tabulka Typ\_aktivity** obsahuje atributy – guid\_ta (varchar), nazev (varchar). Primárním klíčem je atribut guid\_ta.

---

<sup>9</sup> Guid = Globální jedinečný identifikátor (Globally Unique Identifier)

**Tabulka Deník\_jídel** obsahuje atributy – guid\_dj (varchar), kalcium (float), cholesterol (float), datum\_zápisu (date), tuk (float), vláknina (float), kj (float), protein (float), sůl (float), cukr (float), voda (float), jednotky (varchar), typ\_jídla (varchar).

Primárním klíčem je atribut guid\_dj a cizími klíči jsou zde guid\_dt (denní\_doba\_stravy), guid\_f (jídlo) a guid\_u (uživatel).

**Tabulka Jídlo** obsahuje atributy – guid\_j (varchar), kalcium (float), cholesterol (float), datum\_zápisu (date), tuk (float), vláknina (float), kj (float), protein (float), sůl (float), cukr (float), voda (float). Primárním klíčem je atribut guid\_j a cizím klíčem je guid\_u (uživatel) a guid\_tj (typ\_jídla).

**Tabulka Typ\_jídla** obsahuje atributy – guid\_tj (varchar). Primárním klíčem je atribut guid\_tj.

**Tabulka Název\_typu\_Jídla** obsahuje atributy – guid\_ntj (varchar), jazyk (varchar), název (varchar). Primárním klíčem je atribut guid\_ntj a cizím klíčem je guid\_tj (typ\_jídla).

**Tabulka Denní\_doba\_stravy** obsahuje atribut – guid\_dds (varchar), název (varchar). Primárním klíčem je atribut guid\_dds.

## 5.5 Tvorba dimenzí a faktových tabulek

Pro účely zkoumání a provádění analýz jsou zde vytvořeny dimenze a tabulky faktů, které se využívají následně při operacích v datové kostce. Zkoumanou otázkou je změna stravovacích návyků a pohybových aktivit Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018, z toho důvodu potřebujeme pracovat s následujícími tabulkami:

- Jídla a deník jídel, jednotky,
- Aktivity a deník aktivit, typ aktivit
- Uživatelé
- Uživatelské záznamy z měření (hmotnosti)
- Čas stravování (snídaně, svačina, oběd, ...)

Faktovou tabulkou pro stravovací návyky je deník jídla, kde sledujeme, jaký má uživatel příjem tuků, vlákniny, kalcia, cholesterolu, kJ (energetický příjem), proteinů, cukrů, soli a také sledujeme hmotnost uživatele.

--Diary\_Food (Fact\_Diary\_Food)

```
SELECT fat AS measurement_value, guid_diary_foodstuff, guid_user,
guid_foodstuff_unit, created_date, 1 AS measurement_type_food, NULL AS
guid_user_track, guid_foodstuff, guid_diary_time INTO Fact_Diary_Food FROM
p_diary_foodstuff
UNION
SELECT fiber, guid_diary_foodstuff, guid_user, guid_foodstuff_unit, created_date,
2, NULL, guid_foodstuff, guid_diary_time FROM p_diary_foodstuff
UNION
SELECT calcium, guid_diary_foodstuff, guid_user, guid_foodstuff_unit,
created_date, 3, NULL, guid_foodstuff, guid_diary_time FROM p_diary_foodstuff
UNION
SELECT cholesterol, guid_diary_foodstuff, guid_user, guid_foodstuff_unit,
created_date, 4, NULL, guid_foodstuff, guid_diary_time FROM p_diary_foodstuff
UNION
select kj, guid_diary_foodstuff, guid_user, guid_foodstuff_unit, created_date, 5,
NULL, guid_foodstuff, guid_diary_time FROM p_diary_foodstuff
UNION
SELECT t1.value, NULL, t2.guid_user, NULL, t1.created, 9, t2.guid_user_track,
NULL, NULL FROM p_user_track_value t1, p_user_track t2
WHERE t2.guid_user_track_type = '71415a6215e54f4eb20e85c9c668a1d8' AND
t2.guid_user_track = t1.guid_user_track
```

...

Příkaz dále zahrnuje složky protein, sůl a cukr.

K faktové tabulce připojíme potřebné dimenze, přes které se budeme ptát.

Jde o následující dimenze:

Dimenze uživatele, kde máme k dispozici rok narození, jazyk, a pohlaví. Tyto atributy jsou právě brány jako sociodemografické charakteristiky.

```
--User (Dim_User)
SELECT guid_user, birth_year, lang, sex INTO Dim_User
FROM p_user_statistics_view;
```

Dimenze jednotek potravin, které nám pomohou určit porce uživatelů.

```
--Foodstuff_unit (Dim_Food_Unit)
SELECT guid_foodstuff_unit, multiplier, unit INTO Dim_Food_Unit
FROM p_foodstuff_unit;
```

Dimenze času (denní doby) stravování, jde o zařazení jídla, kdy bylo konzumováno, zde je na výběr snídaně, svačina, oběd, svačina, večeře a pozdní večeře.

```
--Diary_time (Dim_Diary_Time)
SELECT guid_diary_time, api_key INTO Dim_Diary_Time
FROM p_diary_time;
```

Dimenze s názvy jídel, které jsou konzumovány.

```
--Food_Name (Dim_Food_Name)
SELECT guid_foodstuff, title AS food_name, title_ascii, lang AS food_name_lang
INTO Dim_Food_Name FROM p_foodstuff_translation;
GO
ALTER TABLE Dim_Food_Name
ALTER COLUMN guid_foodstuff CHAR(32) NOT NULL;
```

Dimenze pro možnost určený typů jídel.

```
--Food_Type (Dim_Food_Type)
SELECT t1.guid_foodstuff, t1.guid_foodstuff_type,
t2.guid_foodstuff_type_translation, t2.lang AS food_type_lang,
t2.title AS food_type, t2.title_ascii INTO Dim_Food_Type
FROM p_foodstuff t1, p_foodstuff_type_translation t2
WHERE t2.guid_foodstuff_type = t1.guid_foodstuff_type;
```



Pro správné určení, o jakou sledovanou hodnotu ve faktové tabulce s jídly jde, nám slouží dimenze s typem sledování. Tato dimenze obsahuje pouze klíčové odkazy, které jsou obsaženy i ve faktové tabulce a také obsahují název sledovaného parametru (tuky, cukry, hmotnost, ...).

```
--Measurement_Type_Food (Dim_Measurement_Type_Food)
CREATE TABLE Dim_Measurement_Type_Food (guid_measurement_type INT NOT NULL
PRIMARY KEY, measurement_name VARCHAR(255) NOT NULL)
INSERT INTO Dim_Measurement_Type_Food VALUES ('1', 'Fat')
INSERT INTO Dim_Measurement_Type_Food VALUES ('2', 'Fiber')
INSERT INTO Dim_Measurement_Type_Food VALUES ('3', 'Calcium')
...
```

Následují odkazy na cholesterol, kJ, protein, cukr, sůl a hmotnost

Pro možnost pracování s daty, tak jak je uvedeno ve zkoumané otázce, tedy v letech 2016-2018 je zde vytvořena dimenze času, která je postavena na granularitě YYYY-MM-DD.

```
-- Date_Dimension
CREATE TABLE [dbo].[Dim_Date] (
    [Date_Key] [int] NOT NULL,
    [Date] [date] NOT NULL,
    [Day] [tinyint] NOT NULL,
    [Month] [tinyint] NOT NULL,
    [Year] [int] NOT NULL,
    [WeekOfYear] [tinyint] NOT NULL,
    PRIMARY KEY CLUSTERED ([Date_Key] ASC)
)
SET NOCOUNT ON
TRUNCATE TABLE DIM_Date
DECLARE @CurrentDate DATE = '2016-01-01'
DECLARE @EndDate DATE = '2018-12-31'

WHILE @CurrentDate < @EndDate
BEGIN
    INSERT INTO [dbo].[Dim_Date] (
        [Date_Key],
        [Date],
        [Day],
        [Month],
        [Year],
        [WeekOfYear]
    )
    SELECT Date_Key = YEAR(@CurrentDate) * 10000 + MONTH(@CurrentDate) * 100 +
DAY(@CurrentDate),
        DATE = @CurrentDate,
        Day = DAY(@CurrentDate),
        [Month] = MONTH(@CurrentDate),
        [Year] = YEAR(@CurrentDate),
        [WeekOfYear] = DATEPART(wk, @CurrentDate)

    SET @CurrentDate = DATEADD(DD, 1, @CurrentDate)
END
```

Druhou částí otázky je změna v pohybových aktivitách, z toho důvodu je nutné vytvořit faktovou tabulku deník aktivit. V této faktové tabulce jsou sledovány spálené kJ, kroky, vzdálenost a čas strávený aktivitou.

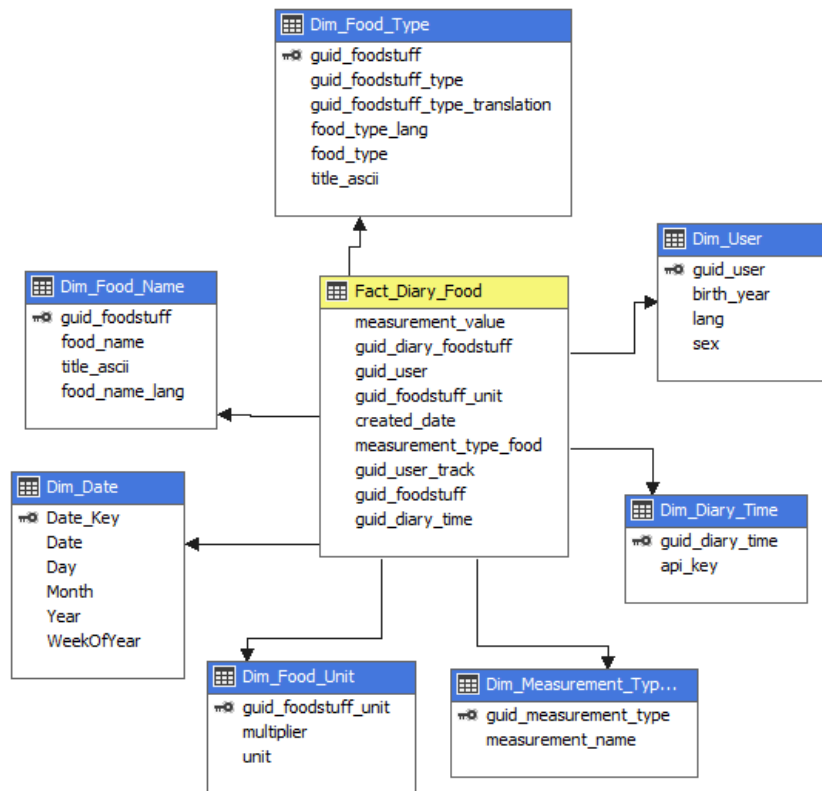
```
-- Diary_Activity (Fact_Diary_Activity)
SELECT t1.kj AS measurement_value, t1.time, t2.kj*66 AS
average_kj_per_minute, t1.guid_diary_activity, t1.guid_user, t1.created_date,
t1.created_time, t1.activity_type AS guid_activity_type, 1 AS
measurement_type_activity INTO Fact_Diary_Activity FROM p_diary_activity t1,
p_activity t2
WHERE t2.guid_activity = t1.guid_activity AND t1.kj <> 0 AND t1.kj != 0
UNION
SELECT t1.steps, t1.time, NULL, t1.guid_diary_activity, t1.guid_user,
t1.created_date, t1.created_time, t1.activity_type, 2 FROM p_diary_activity t1
WHERE t1.steps <> 0
UNION
SELECT t1.distance, t1.time, NULL, t1.guid_diary_activity, t1.guid_user,
t1.created_date, t1.created_time, t1.activity_type, 3 FROM p_diary_activity t1
WHERE t1.distance <> 0
UNION
SELECT t1.time, NULL, NULL, t1.guid_diary_activity, t1.guid_user, t1.created_date,
t1.created_time, t1.activity_type, 4 FROM p_diary_activity t1, p_activity t2
WHERE t2.guid_activity = t1.guid_activity AND t1.time <> 0 AND t1.time != 0;
GO
ALTER TABLE Fact_Diary_Activity
ALTER COLUMN guid_activity_type CHAR(32) NULL;
```

Druhá faktová tabulka využívá dimenzi uživatele a časovou dimenzi, které jsou uvedeny výše. Také využívá dimenzi typu aktivity, která říká, o jaký typ aktivity jde.

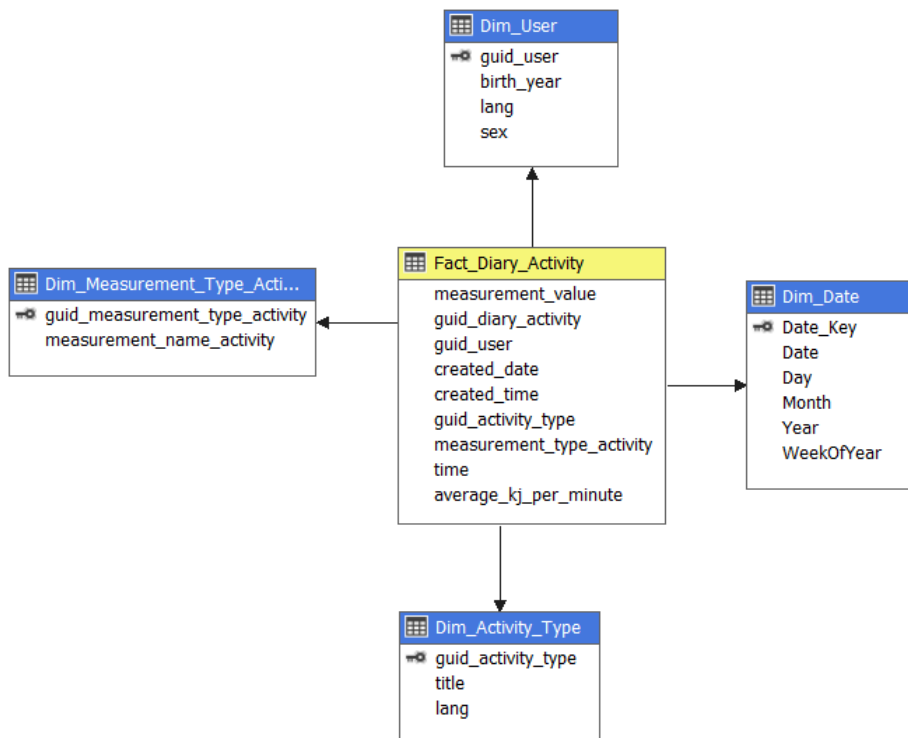
```
--Activity_Type (Dim_Activity_Type)
SELECT guid_activity_type, title, lang INTO Dim_Activity_Type FROM
p_activity_type_translation
```

A také, jako předchozí faktová tabulka, využívá dimenzi s typem sledování pro správné určení sledované aktivity pomocí klíčových odkazů.

```
--Measurement_Type_Activity (Dim_Measurement_Type_Activity)
CREATE TABLE Dim_Measurement_Type_Activity (guid_measurement_type_activity INT
NOT NULL PRIMARY KEY, measurement_name_activity VARCHAR(255) NOT NULL)
INSERT INTO Dim_Measurement_Type_Activity VALUES ('1', 'kJ')
INSERT INTO Dim_Measurement_Type_Activity VALUES ('2', 'Steps')
INSERT INTO Dim_Measurement_Type_Activity VALUES ('3', 'Distance')
INSERT INTO Dim_Measurement_Type_Activity VALUES ('4', 'Time')
```



**Obrázek 29: Schéma deníku jídel – faktová tabulka a dimenze**  
 Zdroj: vlastní zpracování v programu MS Visual Studio



**Obrázek 30: Schéma pohybového deníku – faktová tabulka a dimenze**  
 Zdroj: vlastní zpracování v programu MS Visual Studio

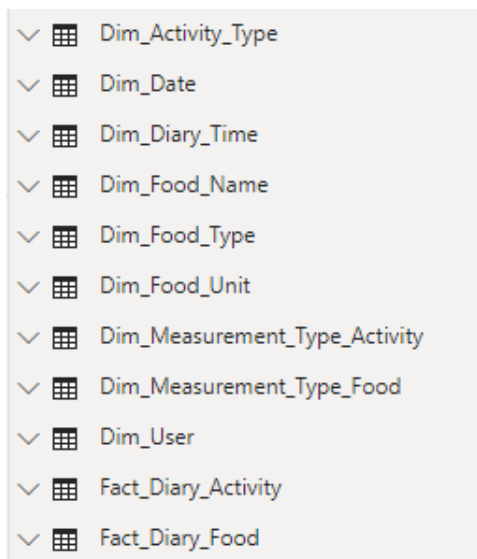
## 5.6 Microsoft Power BI Desktop

Pro vytvoření základních dashboardů, pro analýzu změn stravovacích návyku a pohybových aktivit Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018, je v této práci využit program Power BI Desktop<sup>10</sup> od společnosti Microsoft.

Pomocí Power BI Desktop lze vytvářet kolekce dotazů, připojovat různé datové zdroje a sestavy, které lze také sdílet s ostatními. Aplikace disponuje výkonným dotazovacím jádrem, modelováním a vizualizací dat. Datové analytiku programu Power BI Desktop jsou účinné a flexibilní, jde o dostupný nástroj umožňující připojení k velkému množství dat a jejich zpracování. (Microsoft 2019)

### 5.6.1 Načtení dat

Jak již bylo řečeno, Power BI umožňuje připojit jako zdroj dat skoro cokoli – soubory, databáze, cloudový zdroj, či jiné online služby jako zdroje dat. V této práci jsou data umístěna lokálně na MS SQL Serveru. Po vyplnění nutných údajů se vybírají soubory (tabulky), které požadujeme importovat. Pro účely v této práci byly načteny pouze připravené tabulky dimenzí a faktové tabulky. Na obrázku (Obrázek 31) níže lze vidět o které tabulky se jedná.



**Obrázek 31: Import – výčet tabulek z SQL serveru**  
Zdroj: vlastní zpracování

<sup>10</sup> Microsoft Power BI - <https://powerbi.microsoft.com/cs-cz/desktop/>

## 5.6.2 Vytvoření vazeb (relací)

Po úspěšném připojení dat je nutné tabulky propojit pomocí klíčů. V Power BI lze tuto operaci udělat velmi snadno, stačí se přepnout do modulu „Model“ a tam jednoduše sloupce z načtených tabulek propojovat přetažením, kde se následně vytvoří vazba. Vazby lze také vytvářet přes správce vazeb. Výsledné propojení viz. Obrázek 32.

Z: Tabulka (Sloupec)	Do: Tabulka (Sloupec)
Dim_Activity_Type (guid_activity_type)	Fact_Diary_Activity (guid_activity_type)
Dim_Food_Name (guid_foodstuff)	Fact_Diary_Food (guid_foodstuff)
Dim_Food_Type (guid_foodstuff)	Fact_Diary_Food (guid_foodstuff)
Fact_Diary_Activity (created_date)	Dim_Date (Date)
Fact_Diary_Activity (guid_user)	Dim_User (guid_user)
Fact_Diary_Activity (measurement_type_activity)	Dim_Measurement_Type_Activity (guid_measurement_type_activity)
Fact_Diary_Food (created_date)	Dim_Date (Date)
Fact_Diary_Food (guid_diary_time)	Dim_Diary_Time (guid_diary_time)
Fact_Diary_Food (guid_foodstuff_unit)	Dim_Food_Unit (guid_foodstuff_unit)
Fact_Diary_Food (guid_user)	Dim_User (guid_user)
Fact_Diary_Food (measurement_type_food)	Dim_Measurement_Type_Food (guid_measurement_type)

Obrázek 32: Relace mezi tabulkami

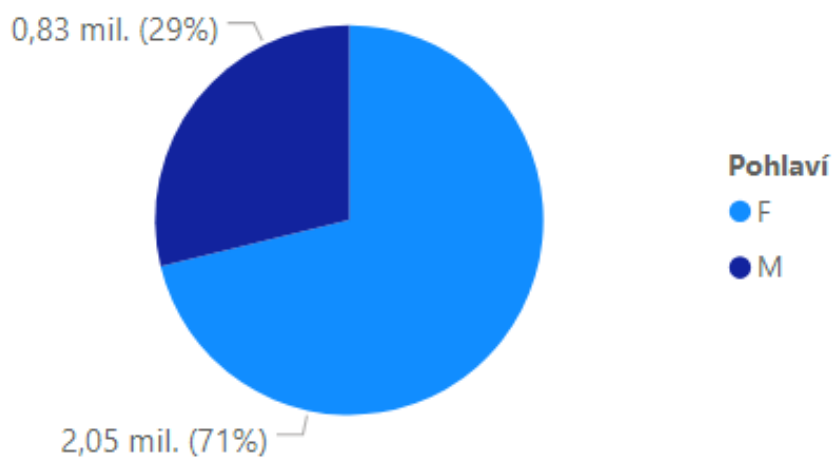
Zdroj: vlastní zpracování

## 5.6.3 Tvorba dashboardů

Nyní po vytvoření relací je možné vytvářet dashboardy a vizualizovat data. Na prvním vytvořeném dashboardu jsou vizualizovány základní charakteristiky:

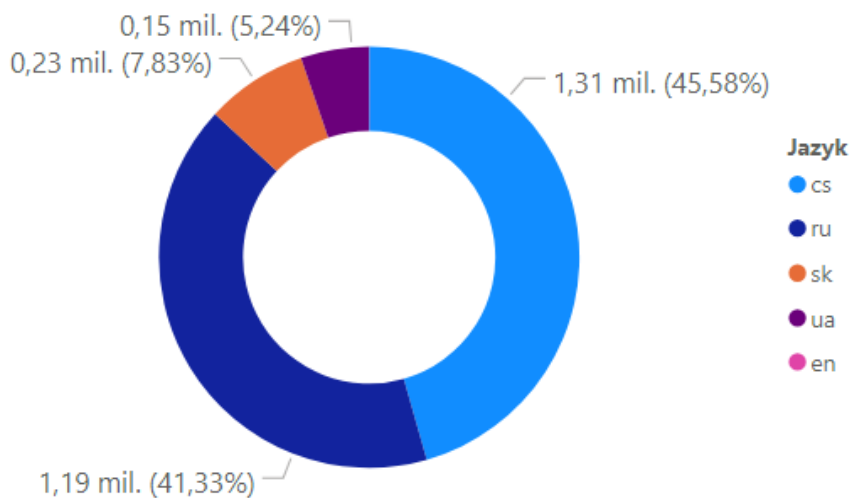
- Procentuální graf podle pohlaví (Obrázek 33)
- Procentuální graf podle zvoleného jazyka aplikace (Obrázek 34)
- Sloupcový graf věkového zastoupení (Obrázek 35)

### Procentuální graf podle pohlaví



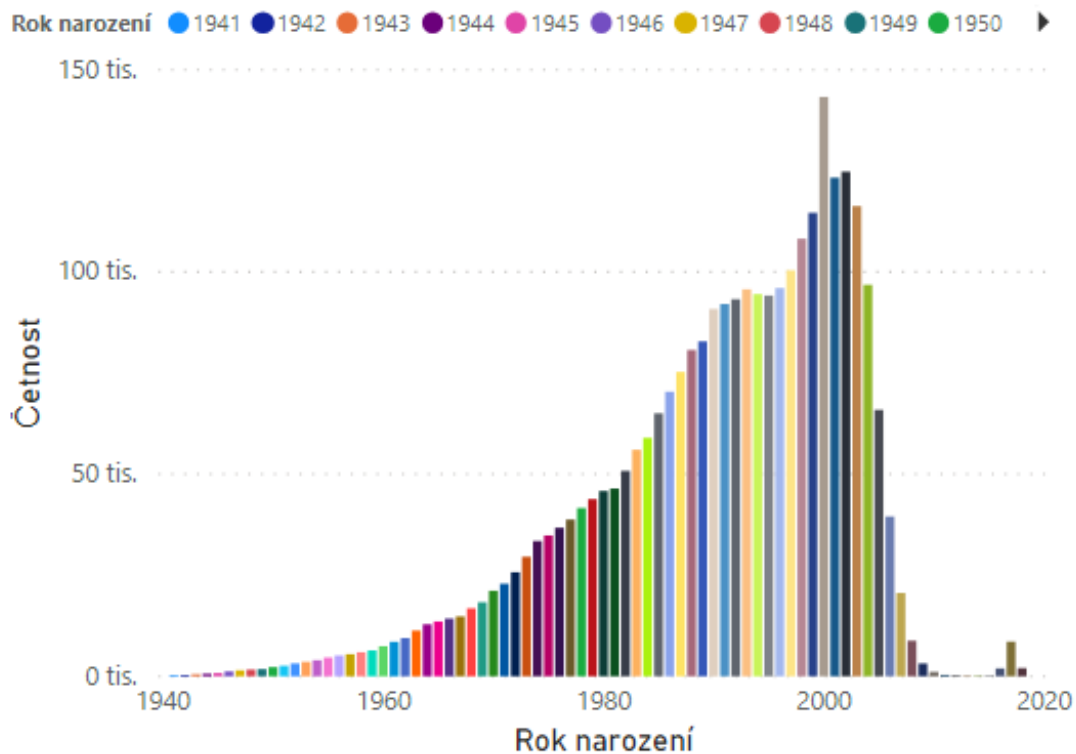
**Obrázek 33: Procentuální výsečový graf podle pohlaví**  
Zdroj: vlastní zpracování pomocí aplikace MS Power BI

### Procentuální graf zvoleného jazyka aplikace



**Obrázek 34: Procentuální výsečový graf podle zvoleného jazyka aplikace**  
Zdroj: vlastní zpracování pomocí aplikace MS Power BI

## Věkové zastoupení

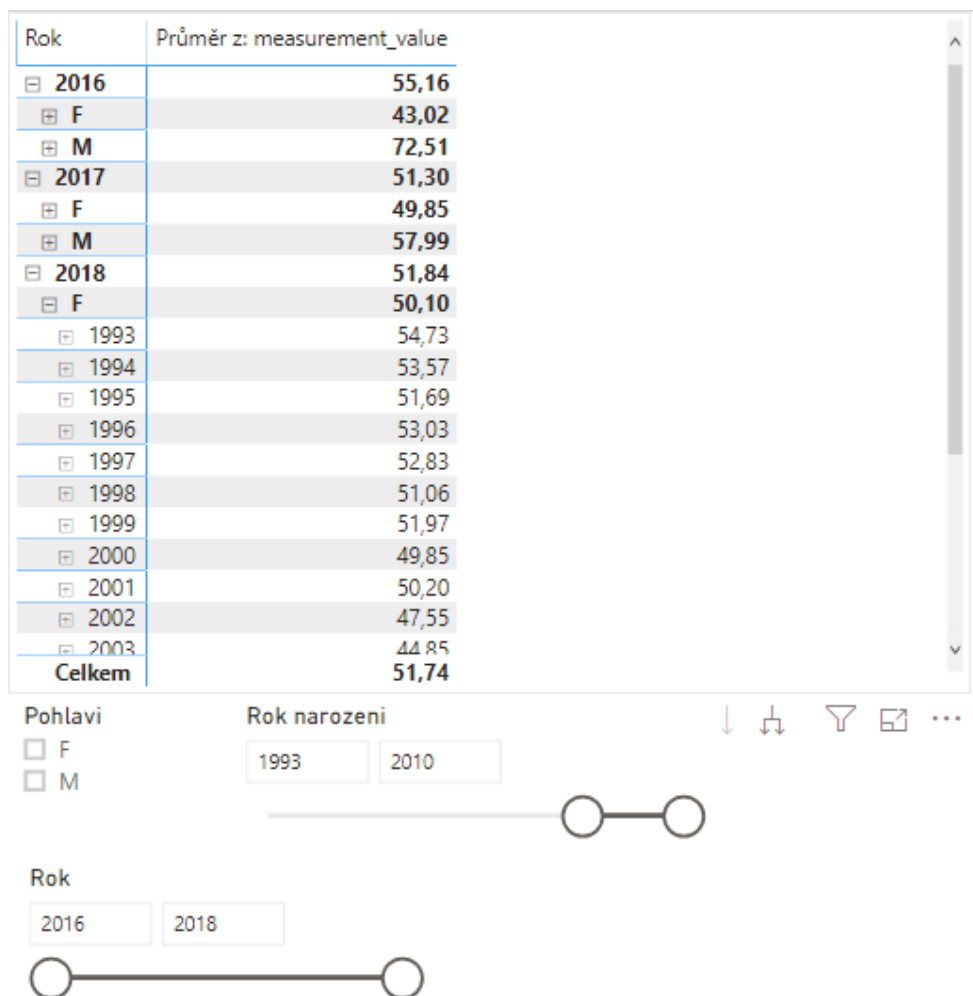


**Obrázek 35: Sloupcový graf věkového zastoupení**  
**Zdroj: vlastní zpracování pomocí aplikace MS Power BI**

Podle výsledků z jednotlivých grafů můžeme říci, že aplikaci Kalorické tabulky nejvíce využívají ženy okolo roku narození 2000. Zajímavým jevem je ze 41% nastavený ruský jazyk aplikace.

Následující dashboardy pomohou při zkoumání a analyzování změn stravovacích návyků a pohybových aktivit Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018. Sociodemografickými skupinami jsou v této práci rok narození a pohlaví.

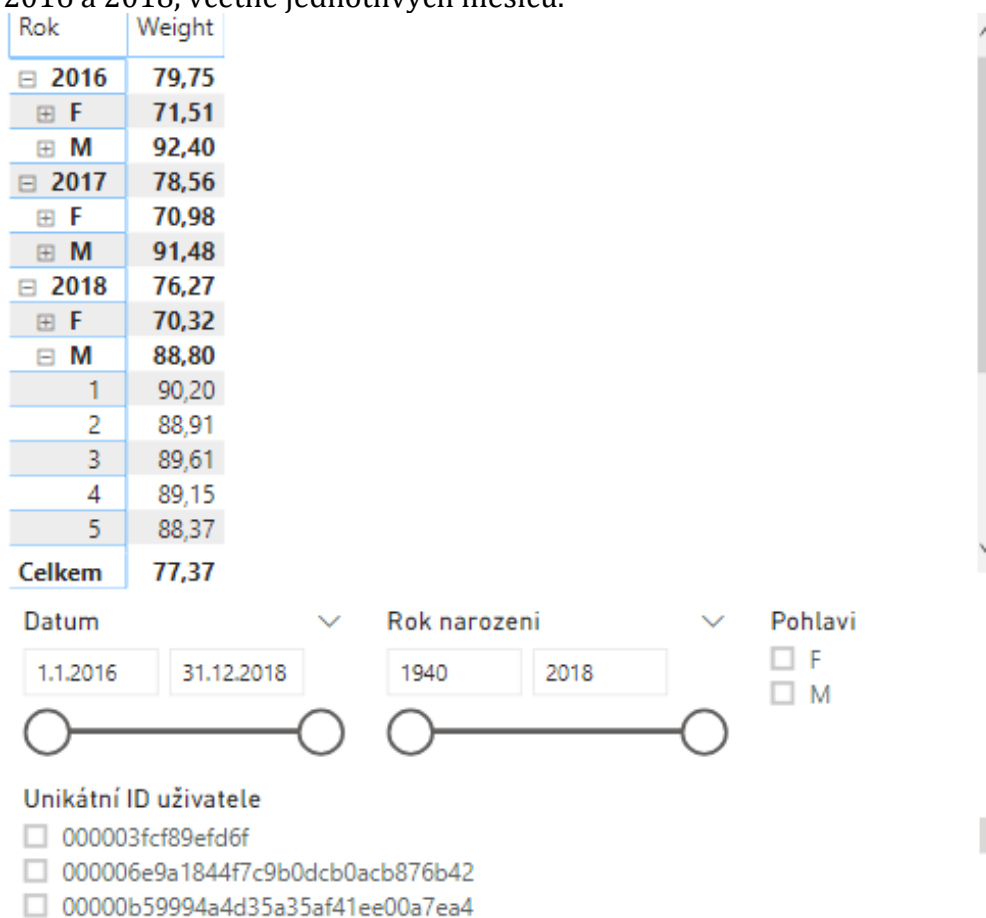
Prvním dashboardem je průměrně strávený čas u aktivit v rozmezí roků 2016 a 2018.



**Obrázek 36: Dashboard – průměrně strávený čas aktivitami**  
**Zdroj: vlastní zpracování pomocí aplikace MS Power BI**

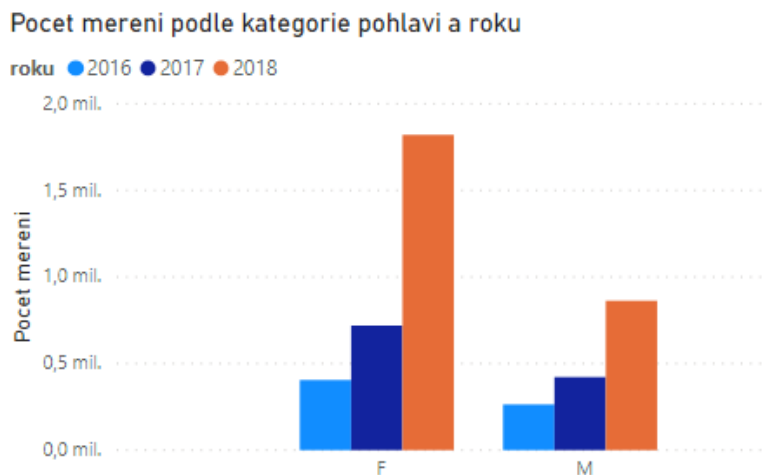


Druhým dashboardem je průměrná hmotnost a vývoj hmotnosti v rozmezí roků 2016 a 2018, včetně jednotlivých měsíců.



**Obrázek 37: Dashboard – průměrná hmotnost a vývoj hmotnosti v čase**  
**Zdroj: vlastní zpracování pomocí aplikace MS Power BI**

Na grafu (viz. Obrázek 38) lze vidět, že nejvíce záznamů hmotností mají ženy, kde rok 2018 je nejsilnějším ročníkem pro zázpisy hmotností.



**Obrázek 38: Sloupcový graf – počet záznamů hmotnosti**  
**Zdroj: vlastní zpracování pomocí aplikace MS Power BI**

Následující dashboard je komplexnější, jde o dashboard, kde lze sledovat vzdálenost, spálené kJ, kroky a čas. Tyto veličiny lze dále rozkládat do jednotlivých typů aktivity jako chůze, těžká práce aj. Opět jsou zde možnosti filtrovat podle datumu aktivit, pohlaví, typu měření, typu aktivity a také roku narození uživatele.

title	Distance	kJ	Steps	Time
[-] Běh		258 054 779 536,44		2 083 932 407,06
F		145 082 694 706,74		1 232 472 272,99
M		112 972 022 449,17		851 454 001,07
		59 721,36		5 893,00
		2 659,18		240,00
[-] Pobyt ve společnosti	1 584 504 094,00		107 066 212 088,00	
F	1 583 202 544,00		56 923 129 241,00	
M	1 301 550,00		50 143 082 847,00	
[+] Chůze		56 948 936 055,78		298 389 654,76
[-] Pobyt na zahradě, chalupě a v přírodě	7 085 372 352,79		913 535 882,00	
M	6 901 014 425,24		632 236 716,00	
F	184 357 927,55		281 299 166,00	
[+] Pobyt doma	2 884 487 016,54		3 591 974 677,00	
[+] Ostatní sportovní odvětví	3 411 776 593,97		764 414 712,00	
[+] Velmi těžká práce	1 072 817,00		1 961 411 199,00	
[+] Jízda na kole (cyklistika)			957 094 941,00	
[+] Těžká práce			647 304 355,00	
[+] Vyčerpávající práce			438 271 149,00	
[+] Jiné	874 637,81		13 131,54	
<b>Celkem</b>	<b>14 968 087 512,11</b>	<b>315 003 715 592,22</b>	<b>116 340 232 134,54</b>	<b>2 382 322 061,83</b>

Datum: 1.1.2016 31.12.2018

Pohlaví:  F  M

Typ měření:  Distance  kJ  Steps  Time

Typ aktivity:  Běh  Chůze  Jiné  Jízda na kole (cyklistika)  Ostatní sportovní odvětví  Pobyt doma

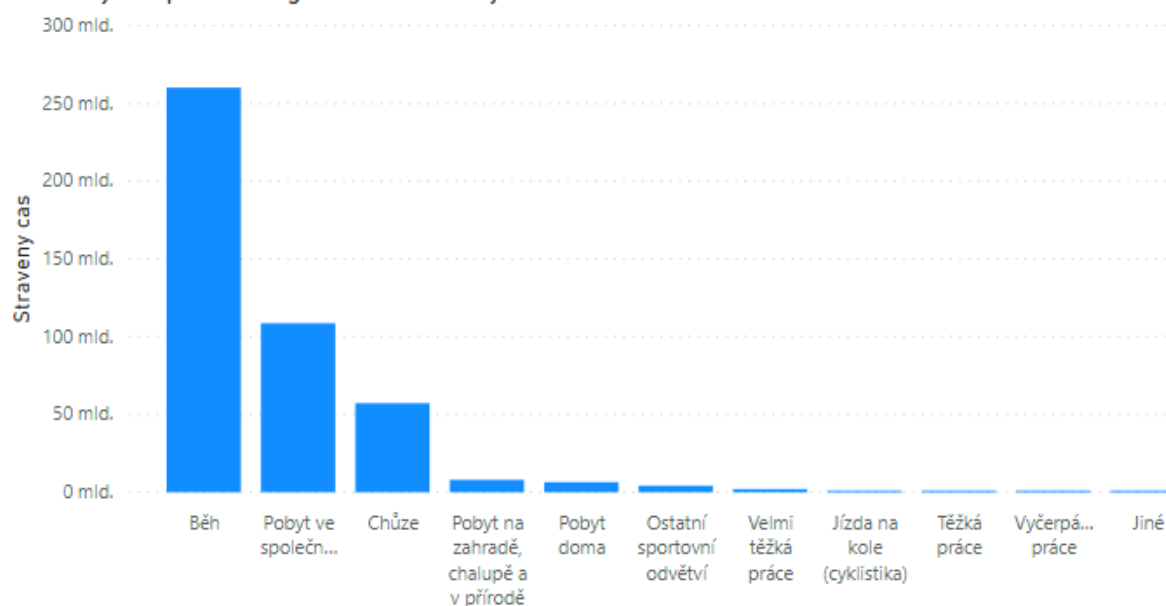
Rok narození: 1940 2002

**Obrázek 39: Dashboard – komplexní přehled aktivity**

**Zdroj: vlastní zpracování pomocí aplikace MS Power BI**

Z výše uvedeného dashboardu (Obrázek 39) při vizualizaci na sloupcovém grafu lze zpozorovat, že běh je nejvíce aplikovaným sportem.

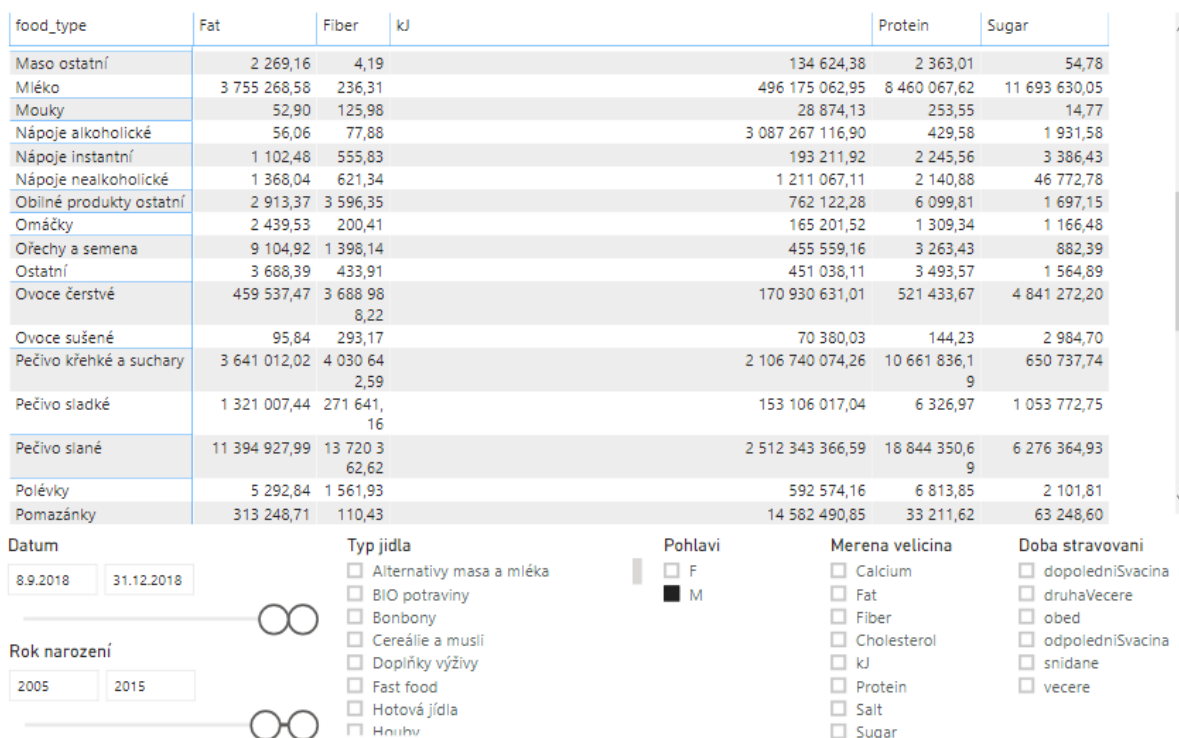
Stravený čas podle kategorie Druh aktivity



Obrázek 40: Sloupcový graf stráveného času podle aktivity

Zdroj: vlastní zpracování pomocí aplikace MS Power BI

Jako poslední dashboard je zde zpracován komplexní dashboard týkající se stravování. V tomto dashboardu je možné pomocí filtrů opět volit mezi rokem narození a pohlavím, měřenou veličinou, typem jídla, datumem, kdy bylo jídlo zapsáno, a denní dobou stravování.



Obrázek 41: Dashboard – komplexní přehled stravování

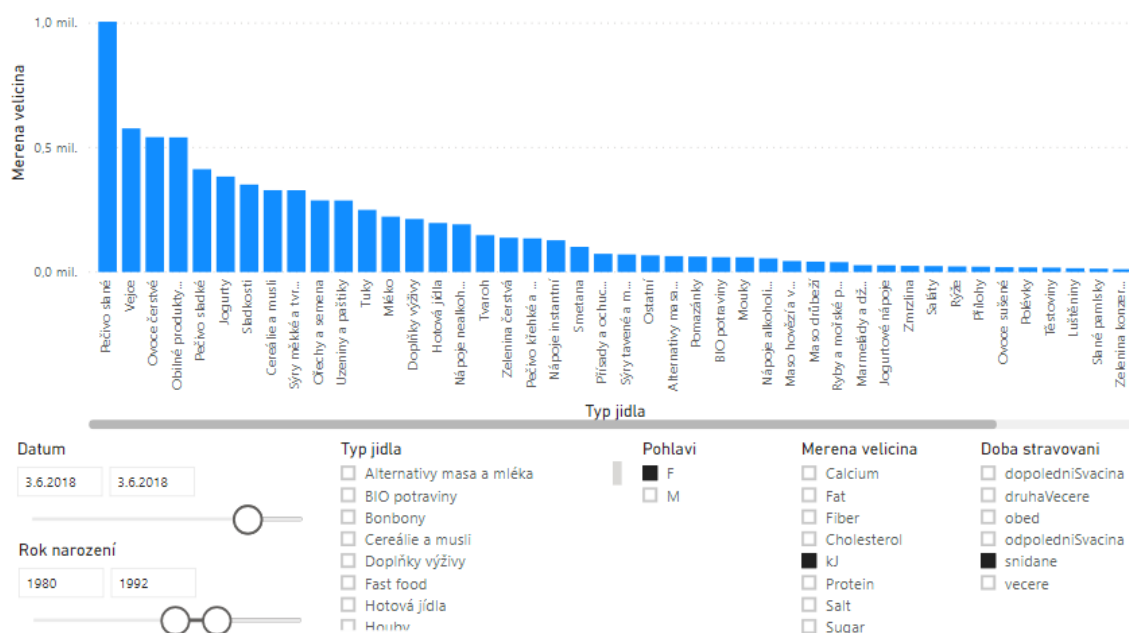
Zdroj: vlastní zpracování pomocí aplikace MS Power BI

Příklad využití tohoto dashboardu může být například následující nastavení:

- Datum stravování: 3.6.2018
- Pohlaví: ženy
- Rok narození: 1980–1992
- Měřená veličina: kJ
- Doba stravování: snídaně

Kde výsledkem viz. (Obrázek 42) je nejvíce zastoupené slané pečivo, na druhém místě vejce a na třetím místě pak čerstvé ovoce.

Merena veličina podle kategorie Typ jídla



Obrázek 42: Sloupcový graf – specifické zadání

Zdroj: vlastní zpracování pomocí aplikace MS Power BI

Pomocí uvedených dashboardů lze vyzorovat rozdíly stravovacích návyků a pohybových aktivity Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018.

## 6 Shrnutí výsledků

Shromažďování dat je v dnešní době jednou z klíčových činností, které ať už malé firmy nebo velké firmy, které s daty pracují, provádějí. Ačkoli je nutné investovat větší částku, a to jak do úložišť či serverů, kam se data nahrávají, tak i do oblasti jejich zpracování, tedy výkonných komerčních analytických aplikací. Pokud firma zaměstnává odborného či vyškoleného pracovníka, který se věnuje analytice, lze pomocí analýz, které jsou následně prováděny nad uloženými daty, získat výhody do budoucna. Firmy mohou docílit rychlejší návratnosti původně investované částky, podpořit jejich prodej nebo pomoci při výběru svým stávajícím nebo budoucím zákazníkům a uživatelům. V této práci bylo shrnuto v teoretické části téma Business Intelligence i včetně podtémat, které s daným tématem souvisí. Business Intelligence je v dnešní době velmi oblíbené téma, bez kterého se již úspěšné podniky nedokáží obejít. Poslední částí teoretické části bylo uvedení analytických i reportovacích (prezentačních) aplikací, které jsou na dnešním trhu využívány jak pro sběr, vyhodnocování tak i pro vizualizaci dosažených výsledků. Následně v praktické části byly nejprve popsány přístupy k datům Kalorických tabulek a postupy, jak byla data získávána. Nad získanými daty byly provedeny transformace, pomocí kterých se docílilo jemného očištění a větší srozumitelnosti dat. U ETL procesů nebyl kladen takový důraz, zde by bylo vhodné provést větší investigaci dat, která zaručí kvalitní očištění dat, tedy důsledné odstranění nulových hodnot, ošetření maximálních a minimálních hodnot, dále pak nahradit nereálné hodnoty průměry. Praktická část také obsahuje tvorbu faktových tabulek a tabulek dimenzí, které byly následně využity v programu MS Power BI, kde byly vytvořeny ze získaných dat náhledy, dashboardy a grafy, které budou využity pro budoucí analytické účely a analýzy.

## 7 Závěry a doporučení

Cílem této práce bylo nejprve získat data z datového skladu, který je provozován v GoodData. Pro stahování dat byly využity programy SQuirrel, DBeaver a DataGrip od JetBrains, které podporují možnost připojení k datovému skladu pomocí JDBC ovladače. Nejvhodnějším nástrojem byl DataGrip, který poskytl možnost filtrace dat před prováděnými exporty, kde možnost filtrace podle data byla právě tím klíčovým, pomocí kterého se docílilo stažení všech potřebných dat. Následně data pomocí Business Intelligence principů a databázových nástrojů, kterými jsou MS SQL Server Management Studio 18 a MS Visual Studio 2019 zpracovat do potřebného formátu, vytvořit tak datový sklad a následně jednotlivé dimenze a faktové tabulky, které v sobě obsahují data pro potřeby při zkoumání a analyzování změn stravovacích návyků a pohybových aktivity Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018. Ačkoli se naskytly při získávání dat mnohonásobné problémy, tak i přes to byla všechna potřebná data stažena. Omezením zde byl jak výkon lokálního počítače, tak slabší internetové připojení, a také natvrdo nastavené maximální časové limity dotazů přímo v JDBC ovladači. Faktové tabulky a tabulky dimenzí byly následně exportovány z lokálního MS SQL Serveru do analytického programu MS Power BI, kde bylo nutné jednotlivé tabulky propojit pomocí klíčů, vytvořit tím tak schéma a následně bylo možné vytvářet jednotlivé dashboardy týkající se zkoumaných otázek. Jak na otázku ohledně změny stravování v období 2016-2018, tak i ohledně změny pohybových aktivit v období 2016-2018 byly vytvořeny komplexní dashboardy s možnostmi filtrování podle zadaných kritérií. Z programu Power BI bylo získáno několik zajímavých charakteristik, které popisují data Kalorických tabulek, a to, že aplikace je využívána hlavně osobami ženského pohlaví, které jsou nejvíce četné okolo roku narození 2000. Další zajímavostí je fakt, že ze 41 % je aplikace nastavená do ruského jazyka. Doporučením v závěru této práce je pro Kalorické tabulky ošetření vstupů dat, a tak se vyvarovat prázdně ukládaných polí (nevyplněné hodnoty pro pole pohlaví, hmotnost, výška, rok narození) a také doporučuji hodnotově omezit či udělat minimální a maximální hranice či pravidla pro zadávání hodnot, které souvisí s poli, které obsahují číselné nebo časové jednotky, zde jde zejména o rok narození, dobu aktivity a spálených

kalorií. Nejedná se o závažné problémy, nýbrž je možné, že jde o nějaká historická či testovací data, jelikož při kontrole jak webové, tak mobilní platformy jsou již výše popsané vstupy ošetřeny.

## 8 Seznam použité literatury a pramenů

### 8.1 Elektronické zdroje

- 1) BUSINESS INTELLIGENCE. In: *NILORT IT GROUP* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://www.nilort.be/services/business-intelligence/>
- 2) CONRAD, Alainia. BI vs Big Data vs Data Mining: A Comprehensive Comparison of the Difference Between Them. In: *SelectHub: Consifende In Software* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://selecthub.com/business-intelligence/bi-vs-big-data-vs-data-mining/>
- 3) DANEL, Roman. Datový sklad. In: *HomeL – VŠB* [online]. 2010 [cit. 2019-10-23]. Dostupné z: [http://homel.vsb.cz/~dan11/is\\_skripta/IS%202010%20-%20Danel%20-%20Datovy%20sklad.pdf](http://homel.vsb.cz/~dan11/is_skripta/IS%202010%20-%20Danel%20-%20Datovy%20sklad.pdf)
- 4) DATA THERAPY. You Don't Need a Data Scientist, You Need a Data Culture. In: *Data Therapy* [online]. 2017 [cit. 2019-10-27]. Dostupné z: <https://datatherapy.org/2017/12/06/building-a-data-culture/>
- 5) DBEAVER. DBeaver 6.2.4. In: *DBeaver: Community Free Universal Database Tool* [online]. 2019 [cit. 2019-11-08]. Dostupné z: <https://dbeaver.io/>
- 6) DÍAZ, Alejandro, Kayvaun ROWSHANKISH a Tamim SALEH. Why data culture matters. In: *McKinsey & Company* [online]. 2019 [cit. 2019-10-27]. Dostupné z: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/why-data-culture-matters>
- 7) G., Cristian. What is Data-Culture and Why should I care about it? In: *Data Resources: Data Culture and Literacy, Traditional BI, reporting tools and other Data related resources* [online]. 2018 [cit. 2019-10-23]. Dostupné z: <http://www.dataresources.net/blog/what-is-data-culture-and-why-should-i-care-about-it/>
- 8) GOODDATA. GoodData Platform Overview. In: *GoodData* [online]. 2019 [cit. 2019-11-08]. Dostupné z: <https://help.gooddata.com/doc/en/reporting-and-dashboards>



- 9) GOODDATA. Reporting and Dashboards. In: *GoodData* [online]. 2019 [cit. 2019-11-08]. Dostupné z: <https://help.gooddata.com/doc/en/reporting-and-dashboards>
- 10) GURU99. ETL (Extract, Transform, and Load) Process. In: *GURU99* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://www.guru99.com/etl-extract-load-process.html>
- 11) CHRISTIANSEN, Thomas. What is a Data Staging Area? In: *Time X Tender* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://support.timextender.com/hc/en-us/articles/210438083-What-is-a-Data-Staging-Area->
- 12) JETBRAINS. Getting started with DataGrip. In: *DataGrip: 2019.2* [online]. 2019 [cit. 2019-11-24]. Dostupné z: <https://www.jetbrains.com/help/datagrip/meet-the-product.html>
- 13) JÖRG, Thomas a Stefan DESSLOCH. Formalizing ETL jobs for incremental loading of data warehouses. In: *GI: Gesellschaft Für Informatik* [online]. 2009 [cit. 2019-10-27]. Dostupné z: <https://dl.gi.de/handle/20.500.12116/20456>
- 14) KALORICKÉ TABULKY. Kalorické Tabulky [online]. 2019 [cit. 2019-11-14]. Dostupné z: <https://www.kaloricketabulky.cz/user/diary>
- 15) LEARNTEK. Data Mining Examples and Techniques. In: *LEARNTEK* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://www.learntek.org/blog/data-mining-examples-and-techniques/>
- 16) MICROSOFT. Novinky v sadě Visual Studio 2019. In: *Microsoft: Dokumenty pro Visual Studio* [online]. 2019 [cit. 2019-11-24]. Dostupné z: <https://docs.microsoft.com/cs-cz/visualstudio/ide/whats-new-visual-studio-2019?view=vs-2019>
- 17) MICROSOFT. Začínáme s Power BI Desktopem. In: *Microsoft: POWER BI* [online]. [cit. 2019-11-24]. Dostupné z: <https://docs.microsoft.com/cs-cz/power-bi/desktop-getting-started>
- 18) OLTP vs OLAP. In: *Edureka* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://www.edureka.co/blog/oltp-vs-olap/>
- 19) OLTP vs. OLAP. In: *DW4U: Datawarehouse4u.info* [online]. 2019 [cit. 2019-10-23]. Dostupné z: <https://www.datawarehouse4u.info/OLTP-vs-OLAP.html>

- 20)PIRKL, David. Datové sklady a BI aplikace: MFF – část 2. In: *Tiscali* [online]. 2004 [cit. 2019-10-27]. Dostupné z: [http://home.tiscali.cz/dedekj/DW\\_Prezentace/DW\\_MFF\\_cast2.ppt](http://home.tiscali.cz/dedekj/DW_Prezentace/DW_MFF_cast2.ppt)
- 21)PŘICHYSTAL, Jan. Datové sklady. In: *Akela.mendelu.cz: studentský server* [online]. 2011 [cit. 2019-10-23]. Dostupné z: <https://akela.mendelu.cz/~jprich/predn/db2.pdf>
- 22)ROUSE, Margaret. OLTP (online transaction processing). In: *TechTarget: Search Data Center* [online]. 2014 [cit. 2019-10-23]. Dostupné z: <https://searchdatacenter.techtarget.com/definition/OLTP>
- 23)SAP. SAP Crystal Reports – Popular Design Features. In: *SAP: Community* [online]. 2016 [cit. 2019-11-24]. Dostupné z: <https://archive.sap.com/documents/docs/DOC-29561>
- 24)SCIFORGE. How to Build Data Culture and Make Data Your Friend. In: *Medium* [online]. 2019 [cit. 2019-10-27]. Dostupné z: <https://medium.com/sciforce/how-to-build-data-culture-and-make-data-your-friend-764183b6786e>
- 25)SHARMA, Saanvi. The Difference Between Data Warehouses and Data Marts. In: *DZone: Database Zone* [online]. 2018 [cit. 2019-10-23]. Dostupné z: <https://dzone.com/articles/difference-between-data-warehouse-and-data-mart>
- 26)SCHILLER, Martin. Co se skrývá pod zkratkou ETL: Jak zpracovat informace uložené v různých podnikových systémech. In: *SystemOnLine: S přehledem ve světě informačních technologií* [online]. 2003 [cit. 2019-11-02]. Dostupné z: <https://www.systemonline.cz/clanky/co-se-skryva-pod-zkratkou-etl.htm>
- 27)SINGH, Mukesh. Inmon vs Kimball Data Models Approaches. In: *Microsoft Business Intelligence (Data Tools): Easy way to learn and implement the Microsoft technologies* [online]. 2016 [cit. 2019-11-24]. Dostupné z: <http://www.sql-datatools.com/2016/09/inmon-vs-kimball-data-warehouse-Architecture-approaches.html>

- 28)SLIDEMODEL. Scorecard Dashboard PowerPoint Template. In: *SlideModel* [online]. 2019 [cit. 2019-11-02]. Dostupné z: <https://slidemodel.com/templates/scorecard-dashboard-powerpoint-template/>
- 29)SODOMKA, Petr a Hana KLČOVÁ. Využití nástrojů a aplikací Business Intelligence v podnicích ČR. In: *SystemOnLine: S přehledem ve světě informačních technologií* [online]. 2018 [cit. 2019-11-02]. Dostupné z: <https://www.systemonline.cz/business-intelligence/vyuziti-nastroju-a-aplikaci-bi-v-podnicich-cr.htm>
- 30)SQUIRREL. Universal SQL Client Version 4.0.0: Overview. In: *Squirrel SQL: Universal SQL Client Version 4.0.0* [online]. 2019 [cit. 2019-11-08]. Dostupné z: <http://squirrel-sql.sourceforge.net/>
- 31)ŠARMANOVÁ, Jana. Informační systémy a datové sklady. In: *Vysoká škola báňská: Technická univerzita Ostrava* [online]. 2007 [cit. 2019-10-23]. Dostupné z: <http://gisak.vsb.cz/wikivyuka/images/2/20/Ins.pdf>
- 32)ŠARMANOVÁ, Jana. Metody dolování znalostí z dat. In: *ReserchGate* [online]. 2002 [cit. 2019-10-23]. Dostupné z: [https://www.researchgate.net/profile/Jana\\_Sarmanova/publication/228647572\\_Metody\\_dolovani\\_znalosti\\_z\\_dat/links/02bfe5137376007769000000.pdf](https://www.researchgate.net/profile/Jana_Sarmanova/publication/228647572_Metody_dolovani_znalosti_z_dat/links/02bfe5137376007769000000.pdf)
- 33)ŠEDA, Jan. Úvod do JDBC. In: *Interval* [online]. 2003 [cit. 2019-11-04]. Dostupné z: <https://www.interval.cz/clanky/uvod-do-jdbc/>
- 34)TESAŘOVÁ, Barbora. Komponenty BI, OLAP, úvod do DW. In: *Oliva UHK: KIKM-BIN Business Intelligence I* [online]. 2019 [cit. 2019-10-23]. Dostupné z: [https://oliva.uhk.cz/webapps/blackboard/execute/content/file?cmd=view&content\\_id=\\_95885\\_1&course\\_id=\\_719\\_1](https://oliva.uhk.cz/webapps/blackboard/execute/content/file?cmd=view&content_id=_95885_1&course_id=_719_1)
- 35)TRAKAL, Matěj. Databázové systémy I. - II. 2009/2010. In: *Materiály by Trtkal: Nejen pro studenty FEI na UPCE* [online]. 2009 [cit. 2019-10-23]. Dostupné z: [http://fei.mtrakal.cz/materialy\\_public/\\_szz/databaze.pdf](http://fei.mtrakal.cz/materialy_public/_szz/databaze.pdf)
- 36)TUTORIALSPPOINT. JDBC Tutorial. In: *Tutorialspoint: simplyeasylearning* [online]. 2019 [cit. 2019-11-04]. Dostupné z: <https://www.tutorialspoint.com/jdbc/index.htm>

- 37) TYSON, Matthew. What is JDBC? Introduction to Java Database Connectivity: Get to know Java's low-level API for making database connections and handling SQL queries and responses. In: *JAVAWORLD: FROM IDG* [online]. 2019, 11 April 2019 [cit. 2019-11-04]. Dostupné z: <https://www.javaworld.com/article/3388036/what-is-jdbc-introduction-to-java-database-connectivity.html>
- 38) VRÁNA, Jan. Datové sklady a jejich optimalizace. In: *SystemOnLine: S přehledem ve světě informačních technologií* [online]. 2001 [cit. 2019-10-23]. Dostupné z: <https://www.systemonline.cz/clanky/datove-sklady-a-jejich-optimalizace.htm>
- 39) VT. Microsoft nabízí Visual Studio nově i v cloudové variantě. In: *Computerworld: FROM IDG* [online]. 2019 [cit. 2019-11-24]. Dostupné z: <https://computerworld.cz/software/microsoft-nabizi-visual-studio-nove-i-v-cloudu-55686>
- 40) WAINSTEIN, Limor. Data Warehouse Architecture — An Overview. In: *Techburst.io* [online]. 2018 [cit. 2019-10-23]. Dostupné z: <https://techburst.io/data-warehouse-architecture-an-overview-2b89287b6071>
- 41) WEINER, George. How to build a data culture — and why it matters for nonprofits. In: *Whole Whale* [online]. 2019 [cit. 2019-10-27]. Dostupné z: <https://www.wholewhale.com/tips/data-culture-nonprofits/>
- 42) ZÁDOVÁ, Vladimíra. Datový sklad. In: *MultiEdu – Technická Univerzita v Liberci* [online]. 2014 [cit. 2019-10-27]. Dostupné z: [http://multiedu.tul.cz/~vladimira.zadova/multiedu/Pokrocile\\_databazove\\_systemy/PDS\\_DW\\_navrh\\_2014.pdf](http://multiedu.tul.cz/~vladimira.zadova/multiedu/Pokrocile_databazove_systemy/PDS_DW_navrh_2014.pdf)
- 43) ZÝKA, Ondřej. Pattern Datový sklad. In: *PROFINIT* [online]. 2016 [cit. 2019-10-27]. Dostupné z: [https://profinit.eu/wp-content/uploads/2016/05/05\\_Pattern\\_DWH.pdf](https://profinit.eu/wp-content/uploads/2016/05/05_Pattern_DWH.pdf)
- 44) ŽIŽKA, Jan. Podnikové informační systémy: Business Intelligence. In: *Vysoká škola managementu a ekonomie* [online]. 2011 [cit. 2019-10-23]. Dostupné z: [https://www.vsem.cz/data/data/sis-texty/studijni-texty-bc/st\\_pis\\_bi\\_zizka.pdf](https://www.vsem.cz/data/data/sis-texty/studijni-texty-bc/st_pis_bi_zizka.pdf)

## 8.2 Tištěné zdroje

- 1) BERSON, Alex a Stephen J SMITH. *Data warehousing, data mining, and OLAP*. New York: McGraw-Hill, 1997, 612 s. ISBN 0070062722.
- 2) DOHNAL, Jan; POUR, Jan. *Architektury informačních systémů v průmyslových a obchodních podnicích*. 1. Vyd. Praha: Ekopress, 1997, 301 s. ISBN 8086119025.
- 3) GÁLA Libor, ŠEDIVÁ Zuzana, POUR Jan. *Podniková informatika: Počítačové aplikace v podnikové a mezipodnikové praxi - 3*. Grada Publishing a.s., 2015. 240 s. ISBN 8024799189, 9788024799186
- 4) GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika2., přepracované a aktualizované vydání*, Grada Publishing a.s., 2009, 496 s. ISBN 9788024726151.
- 5) KIMBALL, Ralph, Joe Caserta. *The data warehouse ETL toolkit practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis: John Wiley, 2004, 491 s. ISBN 0764579231
- 6) KIMBALL, Ralph, Margy Ross. *The data warehouse toolkit: the definitive guide to dimensional modeling*. 3. vydání Indianapolis: John Wiley, 2013, 601 s. ISBN 1118530802
- 7) KIMBALL, Ralph, Margy Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Vydání 2. John Wiley & Sons, 2011. 1118082141, 9781118082140
- 8) LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*; přeložil Jakub Goner. 1. vyd.. Brno: Computer Press, 2012. 350 s. brož. ISBN:978-80-251-3729-1
- 9) LACKO, Luboslav. *Business Intelligence v SQL Serveru 2005*. Brno, Computer Press, 2006. 388 s. ISBN 80-251-1110-5
- 10) LACKO, Luboslav. *Datové sklady, analýza OLAP a dolování dat*. Vyd. 1. Brno: Computer Press, 2003, 486 s. ISBN 8072269690
- 11) LINSTEDT, Daniel; OLSCHIMKE, Michael. *Building a scalable data warehouse with data vault 2.0*. Morgan Kaufmann, 2015. 684 s. ISBN 0128025107

- 12) MAŘÍK, V.; ŠTĚPÁNKOVÁ, O.; LAŽANSKÝ, J. a kol: *Umělá inteligence (4)*. 1. vyd. Praha: Academia, 2003. 476 s. ISBN 80-200-1044-0
- 13) NOVOTNÝ, Ota, POUR, Jan a SLÁNSKÝ, David. *Business Intelligence: Jak využít bohatství ve vašich datech*. Praha: Grada, 2005. 256 s. ISBN 80-247-1094-3
- 14) PIATETSKY-SHAPIRO, Gregory, et al. *Advances in knowledge discovery and data mining*. Menlo Park: AAAI press, 1996. 160 s. ISBN 0262560976
- 15) POUR, Jan, Miloš MARYŠKA, Zuzana ŠEDIVÁ a Stanovská IVA. *Self Service Business Intelligence: Jak si vytvořit vlastní analytické, plánovací a reportingové aplikace*. Grada Publishing a.s., 2018, 352 s. ISBN 9788027108237
- 16) SODOMKA, Petr a Hana KLČKOVÁ. *Informační systémy v podnikové praxi*. Computer Press, Albatros Media, 2017. ISBN 9788025145432
- 17) VERCELLIS, Carlo. *Business Intelligence: Data mining and optimization for decision making*. 1. vydání 2009. 420 s. ISBN 9780470511381
- 18) WITTEN, Ian H., et al. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016. 654 s. ISBN 9780123748560
- 19) WREMBEL, Robert (ed.). *Data Warehouses and OLAP: Concepts, Architectures and Solutions: Concepts, Architectures and Solutions*. Igi Global, 2006. 360 s. ISBN 1599043661

## Podklad pro zadání DIPLOMOVÉ práce studenta

Jméno a příjmení: **Bc. Lukáš Kohl**  
Osobní číslo: **I1700488**  
Adresa: **Hrubínova 703, Jaroměř – Pražské Předměstí, 55101 Jaroměř 1, Česká republika**  
Téma práce: **Analýza a zpracování dat z databázové aplikace Kalorické tabulky**  
Téma práce anglicky: **Analysis and processing of data from database application named Kalorické tabulky**  
Vedoucí práce: **Ing. Barbora Tesařová, Ph.D.**  
**Katedra informatiky a kvantitativních metod**

### Zásady pro vypracování:

- Úvod
  - Struktura a vymezení cílů práce
- Teoretická část (Business Intelligence)
- Praktická část (Kalorické Tabulky)
- Výsledky porovnání a rozbor výsledků
- Shmutí a závěr
- Seznam literatury
- Přílohy
- Zadání práce – kopie

Cílem práce je analyzovat data z databázové aplikace Kalorické tabulky, navrhnout datový sklad, ETL proces a základní dashboardy pro analýzu změn stravovacích návyků a pohybových aktivit Čechů v jednotlivých sociodemografických skupinách v letech 2016-2018.

### Seznam doporučené literatury:

WREMBEL, Robert. *Data Warehouses and OLAP: Concepts, Architectures and Solutions: Concepts, Architectures and Solutions*. Idea Group Inc (IGI), 2006. 360 s. ISBN 1599043661, 9781599043661

LABERGE, Robert. *Dataové sklady: agilní metody a business intelligence*; přeložil Jakub Goner. 1. vyd.. Brno: Computer Press, 2012. 350 s. brož. ISBN:978-80-251-3729-1

WITTEN, I. H., FRANK, Eibe a HALL, Mark A.. *Data mining [online]: practical machine learning tools and techniques*. Ian H. Witten, Eibe Frank, Mark A. Hall. 2011. ISBN 9780123748560.)

MARÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá inteligence (4)*. Praha: Academia, 2003. ISBN 80-200-1044-0

KIMBALL, Ralph, ROSS, Margy. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Vydání 2. John Wiley & Sons, 2011. 1118082141, 9781118082140

LACKO, Luboslav. *Business Intelligence v SQL Serveru 2005*. Brno, Computer Press, 2006. 388 s. ISBN 80-251-1110-5.

Podpis studenta:

Datum:

Podpis vedoucího práce:

Datum: