



Diplomová práce

Rozšíření zavedeného business intelligence řešení v podniku o data ze softwaru technické podpory

Studijní program:

N0688A140016 Systémové inženýrství a informatika

Autor práce:

Bc. Petra Nedvěďová

Vedoucí práce:

Ing. Athanasios Podaras, Ph.D.
Katedra informatiky

Liberec 2024



Zadání diplomové práce

Rozšíření zavedeného business intelligence řešení v podniku o data ze softwaru technické podpory

Jméno a příjmení:

Bc. Petra Nedvědová

Osobní číslo:

E22000447

Studijní program:

N0688A140016 Systémové inženýrství a informatika

Zadávající katedra:

Katedra informatiky

Akademický rok:

2023/2024

Zásady pro vypracování:

1. Zpracování teoretických termínů týkajících se oblasti business intelligence
2. Zmapování stavu firemního řešení a požadavků uživatelů
3. Návrh rozšíření stávajícího řešení
4. Zpracování dat a implementace návrhu
5. Posouzení implementovaného řešení

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování práce:

Jazyk práce:

65 normostran

tištěná/elektronická

čeština

Seznam odborné literatury:

- LEE, James; Tao WEI a Suresh Kumar MUKHIYA, 2018. *Hands-On Big Data Modeling: Effective Database Design Techniques for Data Architects and Business Intelligence Professionals*. 1st. Ed. Birmingham, UK: Packt Publishing. ISBN 978-1-78862-090-1.
- POUR, Jan; Miloš MARYŠKA; Iva STANOVSKÁ a Zuzana ŠEDIVÁ, 2018. *Self Service Business Intelligence: Jak si vytvořit vlastní analytické, plánovací a reportingové aplikace*. Praha: Grada Publishing. ISBN 978-80-271-0616-5.
- KUMAR V. K., Anoop, 2022. *Business Intelligence Demystified: Understand and Clear All Your Doubts and Misconceptions About BI*. 1st. Ed. India: BPB Publication. ISBN 978-93-91030-087.
- JACKSON, Peter a Caroline CARRUTHERS, 2019. *Data Driven Business Transformation: How to Disrupt, Innovate and Stay Ahead of the Competition*. Chichester: John Wiley & Sons. ISBN 978-1-119-54315-2.
- ZELENKA, Martin a Athanasios PODARAS, 2021. Increasing the Effectivity of Business Intelligence Tools via Amplified Data Knowledge. *Studies in Informatics and Control*, vol. 30, no 2, s. 67–77. ISSN 1220-1766.

Konzultant: Bc. Jakub Šebek – Datový Analytik, BizzTreat s.r.o.

Vedoucí práce:

Ing. Athanasios Podaras, Ph.D.

Katedra informatiky

Datum zadání práce:

1. listopadu 2023

Předpokládaný termín odevzdání: 31. srpna 2025

L.S.

doc. Ing. Aleš Kocourek, Ph.D.
děkan

doc. Ing. Klára Antlová, Ph.D.
garant studijního programu

V Liberci dne 1. listopadu 2023

Prohlášení

Prohlašuji, že svou diplomovou práci jsem vypracovala samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Jsem si vědoma toho, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má diplomová práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědoma následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

Rozšíření zavedeného business intelligence řešení v podniku o data ze softwaru technické podpory

Anotace

V této diplomové práci jsou zpracovávány pojmy týkající se business intelligence. Jedná se o BI koncepty, datovou architekturu, BI z pohledu podniků a trendy v BI. Na základě odborné rešerše a zmapování stavu řešení je následně navrženo jeho rozšíření o data z nového datového zdroje. Tento návrh je poté implementován rozšířením ETL procesů a vytvořením dashboardů ve vizualizačním softwaru GoodData.

Klíčová slova: business intelligence, ETL, BI řešení, datové vizualizace, datová architektura, SQL, zpracování dat, dashboard

Extending an established business intelligence solution in a company using data from technical support software

Anotation

In this thesis, concepts related to business intelligence are discussed. These are BI concepts, data architecture, BI from an enterprise perspective and trends in BI. On the basis of expert research and mapping of the state of the art, it is then proposed to extend the solution with data from a new data source. This proposal is then implemented by extending ETL processes and creating dashboards in GoodData visualization software.

Keywords: business intelligence, ETL, BI solution, data visualization, data architecture, SQL, data transformations, dashboard

Poděkování

Ráda bych zde poděkovala vedoucímu práce panu Ing. Athanasiosu Podarasovi, Ph.D za odborné vedení práce, vstřícný přístup a hodnotnou zpětnou vazbu.

Také bych ráda poděkovala všem kolegům z firmy BizzTreat za jejich pozitivní přístup, cenné rady a možnost tuto diplomovou práci zpracovávat právě v této firmě. Zejména bych pak chtěla poděkovat odbornému konzultantovi této práce Bc. Jakubovi Šebkovi za jeho odborné rady a podněty nejenom k této diplomové práci, ale i k ostatním BI řešením, na kterých jsem se mohla spolupodílet a mnohému se přiučit.

V neposlední řadě patří mé díky rodině, partnerovi, přátelům a spolužákům, kteří mi byli velkou oporou nejenom při psaní závěrečné práce, ale i po celou dobu mého studia.

Obsah

Seznam obrázků	13
Seznam zkratek	14
Úvod	15
1 Zpracování teoretických termínů týkajících se oblasti business intelligence	17
1.1 Úvod do business intelligence	17
1.1.1 Definice a význam BI	17
1.1.2 Historie a vývoj BI	18
1.1.3 Základní termíny v BI	19
1.2 Koncepty business intelligence	20
1.2.1 Datová úložiště a jejich role v BI	21
1.2.2 Datová analýza a reporting	21
1.2.3 Vizualizace dat v BI systémech	22
1.3 Architektura v business intelligence	27
1.3.1 ETL komponenta BI systémů	27
1.3.2 Architektura BI projektu	28
1.3.3 Datové modelování	28
1.4 Business intelligence v podnicích	30
1.4.1 Vývoj BI řešení v podniku	30
1.4.2 Faktory implementace BI v podniku	31
1.4.3 Role a funkce v BI: datoví inženýři, analytici a vědci	32
1.4.4 Vliv BI na rozhodovací procesy v podnicích	33
1.5 Trendy v business intelligence	33
1.5.1 Data storytelling	33
1.5.2 Data governance	34
1.5.3 Self-service BI	34
1.6 Software technické podpory	35
1.7 Shrnutí teoretické části	36
2 Zmapování stavu firemního řešení a požadavků uživatelů	37
2.1 Využívané technologie	37
2.2 Existující byznysové oblasti a datové zdroje	38
2.3 Datová architektura	39
2.4 Požadavky uživatelů na rozšíření	40

3	Návrh rozšíření stávajícího řešení	42
3.1	Prozkoumání vstupních dat ze softwaru Freshdesk	42
3.2	Návrh připojení datového toku k aktuálnímu řešení	43
3.3	Návrh dashboardu	44
4	Zpracování dat a implementace návrhu	46
4.1	Tvorba datových vrstev	46
4.1.1	Čistící vrstva	46
4.1.2	Transformační vrstva	50
4.2	Tvorba LDM	52
4.3	Tvorba metrik v softwaru GoodData	53
4.4	Tvorba dashboardu	54
4.4.1	Self-Servise BI	58
5	Posouzení implementovaného řešení	62
5.1	Problémové části při ETL procesu	62
5.2	Posouzení vizualizací	63
5.3	Posouzení naplnění uživatelských požadavků	63
5.4	Návrhy na rozšíření a vylepšení řešení	64
	Závěr	66
	Seznam použité literatury	67

Seznam obrázků

2.1	Schéma datové architektury, vlastní tvorba v softwaru Draw.io	40
3.1	Návrh rozšíření řešení o data ze softwaru Freshdesk, vlastní tvorba v softwaru Draw.io	44
4.1	Logický datový model vytvořený v softwaru GoodData	52
4.2	Ukázka prostředí pro tvorbu metriky v softwaru GoodData	54
4.3	Dashboard vytvořený v softwaru GoodData	57
4.4	Tabulka vytvořená v softwaru GoodData	59
4.5	Prstencový graf vytvořený v softwaru GoodData	59
4.6	Koláčový graf vytvořený v softwaru GoodData	60
4.7	Ukázkový dashboard vytvořený v softwaru GoodData	61

Seznam zkratek

API	application programming interface (rozhraní programové aplikace)
BI	business intelligence
CRM	customer relationship management (řízení vztahů se zákazníky)
ETL	extract, transform, load
ID	identification (unikátní identifikátor)
IT	informační technologie
KPI	key performance indicator (klíčové ukazatele výkonnosti)
LDM	logical data model (logický datový model)
MAQL	multi-dimension analytical query language
SLA	service-level agreement
SQL	structured query language
SSBI	self-service business intelligence

Úvod

V dnešní době digitalizace firem je velkým tématem nejenom zavádění nových softwarů do chodu firmy, ale i zpracovávání dat a rozhodování se na jejich základě. Digitální doba nabízí velké množství dat, které mají firmy k dispozici. S těmito daty mohou pracovat a dělat na jejich základě informovaná rozhodnutí. Jedním z klíčových hráčů v této oblasti firemní digitalizace se stává business intelligence (BI), která pomáhá firmám zorientovat se v jejich datech a vědomě je využít k vlastnímu prospěchu na trhu.

Firma, vybraná v této diplomové práci, již BI řešení využívá. Toto řešení ale v současné době nepokrývá všechny její požadavky. Klíčovým nedostatkem aktuálně využívaného BI řešení je absence jistého datového zdroje. Pro lepší rozhodovací schopnosti by vybraná firma ráda rozšířila své BI řešení o data ze softwaru, který využívá na procesy technické podpory svých klientů. Firma by uvítala zpracování a vizualizování těchto dat. Na jejich základě by poté mohla dělat informovaná rozhodnutí a řídit pomocí nich oddělení technické podpory. Případně by firma tato data mohla poskytnout jako podklady pro informační schůzky se svými klienty.

Tato diplomová práce má za cíl vypracovat návrh na rozšíření aktuálního BI řešení ve vybrané firmě a tento návrh následně implementovat. Vstupem do BI řešení jsou neočištěná data poskytnutá firmou a požadovaným výstupem jsou data vizualizovaná. Rozšíření BI řešení by mělo vycházet z teoreticky zpracovaných poznatků a termínů týkajících se business intelligence oblasti.

Práce se bude členit na dvě části, a to na teoretickou a praktickou část. V teoretické části práce bude v rámci odborné rešerše zpracována témata úvodu do BI, kde bude zkoumán význam BI, historie a základní termíny, se kterými se čtenář v rámci dalších částí práce setká. Zároveň se tato odborná práce bude zabývat koncepty tohoto oboru, budou zde vysvětlena datová úložiště, reporting či vizualizace dat. V rámci teoretické části nebude opomenuta ani architektura datových projektů, kde se práce zaměří na ETL komponentu či datové modelování. Předposlední kapitola teoretické části práce se bude věnovat problematice BI v podnicích. Zde bude uvedeno, jak se v podnicích k BI přistupuje, jaké faktory implementace se v podnicích objevují či jaké role se v BI objevují. Závěrem odborné rešerše bude uvedení aktuálních trendů v oblasti BI. V rámci této kapitoly se práce zaměří na oblasti jako Data Storytelling, Data Governance či Self-Servise BI řešení.

V druhém celku diplomové práce se bude práce zabývat praktickým přístupem k dané tematice. Dojde zde k zmapování firemního řešení za účelem přípravy

podkladů pro následný návrh rozšíření. Zmapování bude uvedeno z pohledu využívaných technologií ve firmě, existujících částí BI řešení, které jsou již plně funkční či aktuálně využívané datové architektury v podniku. Následovat bude návrh vytvořený na základě teoreticky získaných poznatků, zmapování stavu firemního řešení a možností, které budou poskytovat data ze softwaru technické podpory. V návrhu bude jak připojení datového toku do řešení, tak návrh vizualizací.

Po tvorbě návrhu bude přistoupeno k implementaci navrhovaného řešení. Zde budou uvedeny jednotlivé oblasti, které je nutné v rámci řešení zpracovat. Jedná se například o datové vrstvy, logický datový model či tvorbu výstupních datových vizualizací. Závěrečnou částí práce bude posouzení implementovaného řešení a případné navržení rozšíření tohoto řešení o další vhodné datové celky.

1 Zpracování teoretických termínů týkajících se oblasti business intelligence

V rámci první kapitoly budou zpracovány teoretické termíny spojené s business intelligence. K lepšímu pochopení praktické části práce je nezbytné nejprve definovat základní pojmy z této oblasti a popsat základní koncepty BI. Tato teoretická část diplomové práce má za cíl posloužit jako základní pilíř pro hlubší pochopení kapitol následujících. Tato kapitola se nejprve zaměřuje na historii, vývoj a význam BI a teoretické termíny, které se s nimi pojí. Poté jsou rozebírány klíčové koncepty datových úložišť, analýzy a datové vizualizace. Není zde opomenuta ani architektura BI projektů a pohled na BI z podnikového hlediska. Kapitola uzavírají aktuální trendy, které do jisté míry mohou určovat budoucnost tohoto odvětví.

1.1 Úvod do business intelligence

Tato kapitola se zaměřuje na základy oboru business intelligence a ukazuje roli BI jako disciplínu, která spojuje informační technologie a podnikové potřeby. Nejprve se kapitola zaměřuje na BI jako takovou, následně je probírána její historie a vývoj. Dále jsou rozebírány základní termíny, se kterými se setká pravděpodobně každý, kdo se o BI začne zajímat.

1.1.1 Definice a význam BI

Pro lepší pochopení konceptu BI budou v práci nejprve uvedeny definice, které odborníci v této oblasti pro BI využívají. Pour a kol. (2018, s. 17) uvádí: „*business intelligence (BI) představuje jednu z klíčových aplikací IT, která nejvýrazněji ovlivňuje celkovou kvalitu a výkonnost řízení podniků a jejich obchodních a dalších aktivit.*“ Fotr a kol. (2020, s. 50) uvádí: „*V současnosti se postupy analyzující podnikatelské prostředí metodicky rozvinuly do konceptů souhrnně nazývaných business intelligence.*“

Obecně je tedy možno říci, že je business intelligence oblastí, kde se protíná IT sektor s obchodními zájmy institucí. Cílem BI je dosáhnout efektivnějšího řízení podniků díky nalezení užitečných informací, které se v datech mohou vyskytovat. Investice do implementace BI řešení jsou v posledních letech rostoucím trendem. Často je to právě oblast BI, do které se při rozdělování financí pro IT přiděluje nejvíce finančních prostředků. (Pour et al., 2018), (Fotr et al., 2020)

Na BI je možno nahlížet jako na koncept, ve kterém nemusí nutně figurovat specifické nástroje, metodiky nebo technologie. Také se nedá říci, že by BI řešení bylo záležitostí pouze jednoho softwaru, který se dá na trhu sehnat, všechna data do něj naimportovat, počkat a dostat soubor reportů, které odpoví na všechny otázky firmy. BI je spíše proces, který by měl být unikátním pro každý podnik, měl by reagovat na potřeby a možnosti podniku a měl by se skládat z takových komponent, které dávají z pohledu dané instituce smysl. (Kumar V. K., 2022)

Hlavním významem BI je získávání dat a jejich následné zpracovávání za účelem přetvoření těchto dat do informací vhodných pro řízení podniku. Podniky mohou využít velkého množství dat, které se díky obecné světové digitalizaci tvoří. Může se jednat o interní data, která pochází ze softwaru podniků, konkrétně například data z účetních aplikací, personálních systémů, interních databází, dokumentů a dalších interních datových zdrojů. Zároveň mají firmy také možnost sledovat data externí, tedy data, která nesouvisí s podnikem jako takovým, ale s jeho okolím. V tomto případě se může jednat například o kurzovní pohyby měn, výsledky výzkumů veřejného mínění, volební výsledky či sledování veřejně dostupných informací o konkurenčních firmách. (Munoz, 2018)

Velké množství dat, takzvaná Big Data, nejsou dnes pouze výhodou představující velké množství zdrojů, ale i dvousečnou zbraní, která přináší také výzvu v podobě datového přehlcení. Výzvou tedy začíná být sběr pouze relevantních dat, jejich řízení, volba vhodné analýzy a interpretace dat. Tyto procesy, které je možno považovat za BI procesy, se snaží nástrahy spojené s velkými objemy dat minimalizovat. (Munoz, 2018)

BI řešení je schopno ukázat informace a pomoci identifikovat problém, neumí ho však vyřešit samo o sobě. BI řešení přináší shrnutí informací a vybrané informace, které jsou relevantní pro konkrétní skupiny nebo jednotlivce. BI řešení ale typicky neodpovídá na otázky, proč k daným problémům ve firmě dochází či jak tyto problémy řešit. Tyto závěry spočívají na odbornících v dané oblasti, kteří jsou uživateli reportů, například finanční ředitel firmy, vedoucí pobočky atd. BI přináší nalezení trendů, korelací, odhalení odlehlých hodnot a dalších pomocných ukazatelů, ale je pouze na konečných uživateli, jak tyto informace využijí. (Kumar V. K., 2022)

Klíčovou výhodou institucí, které úspěšně implementují BI řešení je konkurenční výhoda v podobě informovaných rozhodnutí. Pokud firma správně využívá BI řešení, její vedení se může spoléhat na faktické údaje, které má k dispozici v jasné formě a v pravidelných intervalech aktualizované. Díky této možnosti mají šanci se snadněji přizpůsobovat vývoji firmy jako takové a zároveň trhu. (Munoz, 2018)

1.1.2 Historie a vývoj BI

Přestože se může zdát, že BI je něco nového, co přišlo před několika málo lety společně s příchodem moderních technologií, opak je pravdou. Za předpokladu, že je o BI uvažováno více obecně, dá se říci, že potřeba být informovaný a dělat rozhodnutí na základě faktických dat byla v určitých oblastech vždy. Často se dokonce jednalo o klíčový aspekt daného odvětví. Příkladem může být politika,

věda, vojenství, právo, bankovníctví nebo podnikání. Všechny tyto oblasti pracovaly s informacemi a rozhodováním na jejich základě dávno předtím, než se objevila business intelligence dnešní doby. (Skyrius, 2021)

K pravděpodobně prvnímu využití pojmu business intelligence došlo v knize *Cyclopædia of Commercial and Business Anecdotes*, ve které toto slovní spojení využil autor R. M. Devens už v roce 1865. Popsal tak jistého Sira Henryho Furnese, který se živil jako bankéř. Tento člověk podle Devense zbohatl díky získávání informací o situaci na trhu, podle kterých se s předstihem zachoval. Díky tomu byl vždy o krok napřed před svými konkurenty. Využití pojmu BI v souvislosti s technologiemi tak, jak je tomu známo dnes použil nejspíše poprvé v roce 1958 vědec a počítačový odborník z IBM H. P. Luhn. Ten publikoval odborný článek, ve kterém se zabýval využitím technologií ke zpracování dat. (Mihiranga, 2022)

Po několika letech odborných debat, kdy zpracování dat bylo spíše otázkou velmi úzce specializovaných odborníků se v 80. letech 20. století začal šířit zájem o BI i mezi podniky. Tou dobou se díky zájmu ze strany byznysu začalo do BI investovat a rostl zájem o technologie, které by zvládaly pomáhat s organizací dat. Touto dobou začaly vznikat technologie jako datové sklady. Tyto technologie budou vysvětleny v následujících kapitolách. (Mihiranga, 2022)

1.1.3 Základní termíny v BI

V této kapitole budou rozebírány základní termíny z pohledu BI. Tyto termíny budou méně odborné, než termíny v následujících kapitolách a měly by přinést větší pochopení základních a nejdůležitějších aspektů BI. Mezi zmíněné termíny budou patřit:

- proces,
- data,
- informace,
- znalost a moudrost,
- data-driven.

BI je samo o sobě proces. Jedná se o proces zpracování dat. Vstupem do tohoto procesu jsou data a výstupem z něj jsou informace. BI proces je proces, který má řadu podprocesů, mezi ty se dají zařadit technologické procesy, procesy související s architekturou projektu, procesy související s lidským faktorem. Tyto zde uvedené podprocesy by se daly také dále členit. (Jackson a Carruthers, 2019)

Data jsou v BI zdrojem. Jsou extrahována, skladována a transformována. Za data jsou považována čísla, texty, obrázky, datumy a další. Příkladem interních dat mohou být adresy zaměstnanců, informace o zákaznících, hodnocení kvality služeb, využívání webových stránek a mnoho dalších. Zároveň jsou zde chápána data i jako data externí či data třetích stran, která jsou pro podnik nějakým

způsobem relevantní. Data jsou typicky ve formátech .csv, .xml, .json nebo jako zdroj z databází. Mezi používáním pojmu data a informace v běžné mluvě a v IT sféře je lehká nuance. Data z technického pohledu jsou tzv. raw data, tedy data, která nebyla ještě nijak zpracovaná a obsahují velké množství datových nečistot. V laické mluvě se termíny data a informace běžně zaměňují. (Kumar V. K., 2022), (Jackson a Carruthers, 2019)

Informace jsou data rozšířená o kontext či data s přidanou informativní hodnotou. Tedy z dat se mohou při zpracování stát informace. Kontext je datům dodán díky vysvětlení jejich významu, propojení dat mezi sebou, obohacení o další data a vysvětlení k čemu a jak se daná data dají použít. V této chvíli se z dat stávají informace. (Kumar V. K., 2022)

Znalost je v pořadí po informacích další stupeň na pomyslném žebříčku. Je o ní možno mluvit v případě, kdy na scénu přicházejí odborníci na danou problematiku. Ti zanalyzují informace a jsou schopni říci, co znamenají. V teorii je dalším stupněm po znalosti moudrost. Tento pojem si můžeme v datové sféře představit jako správné vyhodnocení situace a reagování na ni správným a logickým způsobem. Dále bude uveden zjednodušený příklad problematiky rozdílu mezi daty, informacemi, znalostmi a moudrostí (někdy také uváděné jako zkušenost). Data se dají představit jako slovo „červené světlo“ - bez kontextu to může znamenat spoustu věcí. Informace jsou chápány jako: „Červené světlo je výše než oranžové a zelené světlo. Pouze červené světlo svítí.“ Znalost znamená, že někdo, kdo zná problematiku dopravy vyhodnotí, že tato tři světla znamenají semafor a moudrost znamená, být si vědom souvislosti, že na červenou se má zastavit. (Jackson a Carruthers, 2019)

Data-driven je pojem, který by se dal přeložit jako datově řízeno. To znamená, že se typicky firma rozhoduje a řídí na základě dat, nikoliv na základě dohadů či pocitů. Existují určitá rozhodnutí, která manažeři z logiky věci musí udělat na základě pocitů nebo emocí, protože byznys je také o lidech. Každopádně jisté formy rozhodování by se měly dít právě na základě dat a informací z nich získaných. Například pokud z dat manažer vyčte, že největší počet úrazů se stává v pobočkách na jihu země, je racionálním krokem zaplatit dodatečné školení cíleně právě zde. (Kumar V. K., 2022)

1.2 Koncepty business intelligence

V této kapitole jsou probírány klíčové koncepty oboru BI. Nejprve se kapitola zabývá datovými úložišti a jejich rolí při ukládání podnikových dat. Následně je probírána datová analýza a reporting. Jedná se o klíčové nástroje pro zpracování dat a podporu manažerských rozhodnutí. V poslední a nejvíce rozsáhlé kapitole je pojednáváno o důležitosti datových vizualizací a interpretaci dat pro uživatele BI systémů.

1.2.1 Datová úložiště a jejich role v BI

Datová úložiště patří mezi hlavní komponenty BI řešení. Obecně slouží k ukládání dat. Podle odborné literatury je možno dělit datová úložiště na několik skupin podle principů, kterými data ukládají. V této kapitole bude rozebíráno několik typů datových úložišť: datový sklad, datové tržiště a dočasné úložiště dat. (Pour et al., 2018)

Datový sklad, anglicky Data Warehouse, se dá přirovnat k relační databázi. Oproti ní jsou ale data v datovém skladu jinak uložena. Jedná se například o to, že v datovém skladu jsou data ze všech aplikací skladována na jednom místě neohledně na zdroj, ze kterého pochází. Dalším rozdílem je to, že datové sklady jsou ve většině případů navrženy tak, aby se z nich pouze četlo. Tedy ve většině případů se v datových skladech žádná data nevytvářejí ani nepřepisují. Datové sklady jsou místa, kam se skladují data z celé firmy, neprobíhá zde žádné dělení podle středisek nebo oddělení. Datový sklad je také místem, kde se shromažďuje i historie dat. Data kvůli tomu musí mít jasně definovanou časovou dimenzi, aby bylo možné určit, z které doby pocházejí. V datových skladech se typicky drží normalizovaná data, k denormalizaci dochází až v datových tržištích. (Pour et al., 2018)

Datová tržiště, anglicky Data Marts, se podobají ve většině ohledech datovým skladům. Na rozdíl od datových skladů zde ale začíná rozdělování dat podle potřeb uživatelů. Datová tržiště mohou být buď decentralizovanými datovými sklady. Tedy postupně se jednotlivá datová tržiště přidávají podle aktuálních potřeb podniku. Pokud jsou finance a prostor na vývoj může se přidat další datové tržiště pro konkrétní skupinu uživatelů (finance, výroba, marketing, pobočka A, pobočka B) a toto tržiště se začlení do decentralizovaného datového skladu. Tento přístup je známý jako Bottom-up a jeho výhodou je snížení rizik při implementaci a zkrácení doby finanční návratnosti investice. Druhým přístupem k datovým tržištím je přístup top down, tedy existuje celofiremní centralizovaný datový sklad, který obsahuje všechna data a ta se přerozdělují do jednotlivých datových tržišť. Výhodou tohoto přístupu je možnost analýzy nad konkrétními daty, a to v menších objemech. (Pour et al., 2018)

Dočasná úložiště dat mají za cíl připravit data do potřebné formy předtím, než budou nahrána do datových skladů. Do těchto úložišť jsou extrahována data z produkčních databází. Dočasná úložiště dat, jak už plyne z jejich názvu, jsou pouze mezistupněm v datovém zpracování. Poté, co je dosaženo dostačující datové kvality, se data pošlou do datového skladu a z dočasného úložiště se odstraní. (Pour et al., 2018)

1.2.2 Datová analýza a reporting

Datová analýza je kombinací statistických metod, počítačových technologií a byznysového porozumění firemním datům. Datová analýza odpovídá na otázky typu: „Co se ve firmě právě děje?“, „Co by se mohlo stát?“ nebo „Co se dělo v minulosti?“. Datová analýza je schopná odhalit trendy, zákaznické preference a skryté byznysové vzorce, a to dnes už často i prakticky v reálném čase. Analýzu

můžeme dělit na tři typy: deskriptivní, prediktivní a preskriptivní. (A. Azevedo a Santos, 2021), (Cheng et al., 2023)

Deskriptivní analýza popisuje, co se stalo a co se aktuálně děje. Pracuje tedy s již existujícími daty a jejím cílem je co nejpřesněji prezentovat jejich stav. Díky tomu je možné odhalit trendy, které by bez použití analýzy nebyly patrné. Prediktivní analýza má za cíl vytvářet predikce skutečností, které by se mohly v budoucnu stát i s jejich pravděpodobnostmi vzniku. Preskriptivní analýza kombinuje dva předchozí typy analýz, tedy řeší aktuální stav věcí a zároveň předvídá nejpravděpodobnější scénář vývoje v budoucnu. Tyto tři typy analýz se mohou leckdy překrývat a doplňovat, nelze tedy říci, která z nich je nejvhodnější pro řízení byznysu. Deskriptivní analýza se většinou pojí s business intelligence a preskriptivní analýza je něco, z čeho čerpá oblast data mining. (A. Azevedo a Santos, 2021)

Cílem datové analýzy je získat užitečné informace, na jejichž základě je možné podpořit manažerská rozhodnutí. Nástroje, díky kterým tohoto cíle dosahuje, jsou modelování dat či reporting. Právě reporting je podporou pro manažerská rozhodnutí. (Löffler, Štětínová a Bernat, 2021) V dnešní době se začínají objevovat čím dál tím více personalizovaná řešení a zároveň řešení přestávají být konkurenční výhodou pouze pro velké firmy a začínají se dostávat i k menším podnikům, které mohou využít datové analýzy a reportingu k řízení svých procesů. K těmto změnám dochází díky snižujícím se nákladům datových skladů a velkému zájmu o BI řešení. (Cheng et al., 2023)

Reporting je systémem, který obsahuje všechny informace potřebné k fungování společnosti. Je možno mluvit o značně komplexním systému, protože by měl nabízet informace pro všechny konečné uživatele, tedy ideálně pro všechna oddělení, management a další uživatele BI řešení, a to včas a ve formě, kterou si uživatelé přejí. Předání informací je dosahováno díky reportům v podobě tabulek a grafů. Jejich různorodost bude detailněji popsána v následující kapitole o datových vizualizacích. (Pour et al., 2018)

1.2.3 Vizualizace dat v BI systémech

Datová vizualizace je jedním z nejvíce stěžejních konceptů BI, protože jsou to právě datové vizualizace, kterými jsou data reprezentována koncovým uživatelům řešení. Pro lidské oko není snadné zorientovat se v nečleněném textu či větším množství číselných hodnot. Vizualizace jsou formou komunikace, díky které se složitější datové koncepty dají pochopit rychle a jednoduše. Na první pohled je například viditelné, jestli je firma v kladných, nebo záporných číslech, což by z pročítání účetních záznamů sice bylo viditelné také, ale trvalo by to spoustu času. (Machiraju a Gaurav, 2018)

Vizualizace by měly usnadňovat prezentování dat napříč celou organizací, pochopení chování zákazníků, predikování finančních výsledků, porovnání nabízených produktů a spoustu dalších datově orientovaných faktorů. (Machiraju a Gaurav, 2018) Na trhu existuje spousta softwarů, které nabízí tvorbu datových vizualizací. Mezi nejznámější z nich patří bezesporu Microsoft Power BI či

Tableau. Základní datové vizualizace se dají tvořit i v Excelu, ale z pohledu BI to není vhodný nástroj pro práci s velkým množstvím dat, které je potřeba pro BI řešení typicky zpracovávat. (Machiraju a Gaurav, 2018), (Ryan, 2018)

Tvorba datových vizualizací se může zdát snadná, ale ve skutečnosti je to komplexní obor. Datové vizualizace by se měly řídit pravidly, která v této části kapitoly budou detailněji rozvedena. Prvním pravidlem, kterým by se měl datový analytik řídit, je odstranění nepodstatného. U vizualizací platí pravidlo „méně je někdy více“. Nadbytečné prvky ve vizualizacích odvádí pozornost od podstatného a ruší konzumenta. Je doporučeno za graf nepřidávat žádné tematické obrázky či texty, nepoužívat příliš mnoho barev, nestínovat ani nevylepšovat graf použitím jeho 3D verze. (Pour et al., 2018)

Dalším pravidlem je používat pro různé typy dat vhodné grafické vizualizace. Tedy v prvním kroku je vhodné se zamyslet, která data jsou vizualizovaná. Také je dobré zvážit, co je cílem jejich vizualizace. Podle typu dat je pak volen typ grafu. Existují tři základní typy dat:

- kvantitativní,
- kategorická,
- ordinální.

Kvantitativní data, jsou data číselná. Jsou to data, se kterými je možné provádět nejvíce matematických operací, dají se mezi sebou porovnávat, odčítat, sčítat, násobit a dělit. Z těchto dat je možno vypočítat průměr, medián či procenta. Obecně nad nimi lze provádět matematické a statistické operace. Tato data bývají nazývána fakty, protože číselně (fakticky) vyjadřují jisté hodnoty. Příkladem takových dat mohou být: výše mezd, zisk firmy nebo počet dní ve službě. (Pour et al., 2018)

Kategorická data, také známá jako data nominální, se dají popsat jako vlastnosti. Na rozdíl od dat kvantitativních nejsou měřitelné, nedají se zprůměrovat ani se na nich nedá z logiky věci provádět většina matematických operací. Dá se z nich ale vyčíst nejčastější hodnota. Příkladem je jméno, bydliště, pohlaví, barva (výrobku) a další vlastnosti. (Pour et al., 2018)

Třetím typem dat jsou data ordinální. Ta se nedají měřit, stejně jako nominální data, ale na rozdíl od nominálních dat se dá určit jejich pořadí. Hodnoty se tedy dají porovnat a dá se určit, jestli jsou větší/menší, lepší/horší nebo rychlejší/pomalejší vůči ostatním záznamům v souboru. Příkladem těchto dat je umístění v závodech, dosažené vzdělání nebo věk. (Pour et al., 2018)

Datové vizualizace také pracují s velikostí detailu dat (granularitou). Data v nejmenším detailu zobrazují často pouze jedno číslo. V těchto případech je třeba zdůraznit pouze toto číslo, protože má samo o sobě vypovídající hodnotu. Střední granularita pokrývá případy, kdy jde o více než jedno číslo a kdy není detail příliš velký, takže se dá bez problémů zobrazit v grafech. V případech, že je nutné zobrazovat příliš velký detail, který by v grafu zanikl, se využívají graficky upravené tabulky, ve kterých může být číselná hodnota v největším detailu. (Evergreen, 2020)

V případech, kdy dává smysl vizualizovat pouze jedno číslo, je možné využít jednu ze čtyř níže zmíněných možností. První možností je vizualizovat toto číslo zvětšené na velikost ostatních vizualizací v dashboardu a zobrazit ho samostatně. Ať už se jedná typově o kvantitativní, kategorické či ordinální data, tato možnost je funkční pro všechny typy dat. Samostatně stojící číslo či slovo dostatečné velikosti upoutá pozornost datového konzumenta. Druhou možností je tzv. Icon array, což by se dalo přeložit jako pole ikon. Jedná se o grafické znázornění procent z celku. V příkladě „Devět z deseti zubařů doporučuje zubní pastu XY“ je v této grafice zobrazeno deset osob, z nichž je devět vybarvených výraznou barvou, zatímco poslední osoba (ta co zubní pastu nedoporučuje) je znázorněna nějakou nevýraznou, typicky šedou, barvou. Třetí možností je využití koláčového grafu, kdy je opět podstatné číslo zobrazeno v rámci koláčového grafu nějakou výraznou barvou. V ideálním případě je graf popsán i číselnou či procentuální hodnotou. Zbytek grafu není popsán ani označen, aby vynikla informace, která má být vizualizována. Čtvrtou možností je sloupcový graf, kde opět pouze cílové číslo zůstává barevně zvýrazněno, zatímco zbytek sloupců je šedý a bez dalších popisků, tím vynikne pouze tato důležitá hodnota. (Evergreen, 2020)

Základní typy grafů a jejich využití

Typickým příkladem vizualizace jsou grafy. Tato kapitola rozebere základní typy grafů a jejich vhodné použití z pohledu BI. Kapitola se bude věnovat sloupcovým, výsečovým a spojnicovým grafům, protože jsou to nejčastěji využívané grafy.

Sloupcový graf, je hojně využíván v byznyse, protože pro lidské oko je snadno čitelný. Skládá se ze svislých sloupců, na kterých se znázorňuje porovnávání jednotlivých hodnot, a to podle výšky sloupce. V rámci jednoho sloupcového grafu může být vizualizováno i více proměnných, pak je každá taková proměnná typicky znázorněna jinou barvou. Znázorňované hodnoty by spolu měly logicky souviset. Například množství financí investovaných do marketingové kampaně v prvním sloupci a počet zákazníků ve sloupci druhém. Oba tyto sloupce se v grafu opakují například v rámci času, aby bylo možno vyzorovat, jestli při zvýšení finančních prostředků na marketingové výdaje se také zvyšuje počet zákazníků. Sloupcový graf má i skládanou formu, v níž se v rámci jednoho sloupce objevuje více kategorií, které se na sebe skládají. Tímto skládáním je tvořena výsledná hodnota sloupce. Hodnoty sloupcových grafů se dají vyjadřovat v absolutních i relativních hodnotách. Pruhovým grafem je nazývána obdoba sloupcového grafu, která je prakticky stejně koncipovaná, ale její rozložení je horizontální. (Pour et al., 2018)

Při používání sloupcového grafu, nebo některé z jeho obdob, je dobrým zvykem hodnoty setřídít. Toto setřídění pomáhá v rychlé orientaci a je lépe čitelné pro konzumenta. Setřídění dat musí dávat smysl. Pokud graf zobrazuje vývoj v jednotlivých měsících, je logické ho seřadit podle posloupnosti měsíců. Pokud graf zobrazuje nejdělečnější pobočky, měl by být seřazen od nejvíce výdělečných po ty nejméně výdělečné. Pokud má graf naopak odhalovat prodělečné pobočky, bude seřazen od nejmenších hodnot po ty nejvyšší. Vždy je dobré seřazení přizpůsobit konkrétní situaci. Sloupcový graf je sice oblíbeným, ale má jisté limity.

Při velkém množství zobrazovaných hodnot se stává nečitelným. V těchto případech je doporučované nahradit sloupcový graf grafem spojnicovým, kde není velké množství hodnot problém. (Pour et al., 2018)

Druhým velmi často využívaných grafem je graf koláčový či výsečový. Ten by se měl používat při zobrazení vztahu části k celku. Pro procentuální vyjádření je tedy velmi vhodný, přestože se dají na výsečovém grafu zobrazit i absolutní hodnoty. Výsečový graf je sice oblíbený mezi uživateli, ale méně oblíbený mezi datovými analytiky. Tento rozpor pramení z možného zkreslení, ke kterému může při jeho použití snadněji docházet. Lidský mozek totiž mnohem hůře na kruhu vnímá nuance. Tedy hodně těžko rozpozná například rozdíl mezi 29 % a 32 %. Protože se tyto hodnoty zdají pohledem lidského oka na koláčovém grafu prakticky totožné, je v těchto případech volen spíše graf sloupcový, kde jsou tyto tři procenta rozdílu jasně patrná. Dalším problémovým použitím výsečového grafu je zobrazení většího množství kategorií než pěti, protože v těchto případech opět dochází opět k nečitelnosti.

Nevhodné je ve výsečových grafech také zobrazování hodnot, které jsou velmi malé, a to v řádu jednotek procent. V odborné literatuře je uváděno, že se dá mluvit o hodnotách, které mají velikost do 5 % z celku. V těchto případech může být řešením tyto hodnoty uvádět jako kategorii *Ostatní* a zobrazit je v případě potřeby v úplně samostatném grafu, který bude sloučit pouze pro *Ostatní* hodnoty. Alternativou výsečového grafu je prstencový graf, kterému (podle jeho názvu) chybí prostřední část, takže vypadá jako prsten. Protože tyto typy grafu přináší velké množství rizik, je opravdu potřeba dobře promyslet, jestli jsou vhodnou variantou při tvorbě reportů. (Pour et al., 2018)

Posledním grafem, který zde bude uveden, protože je opravdu typickým, je graf spojnicový. Jeho podoba je přímkou mezi body, které znázorňují průsečíky hodnot os x a y. Tento typ grafu ve většině případů využívá na ose x čas, je to zároveň jeho vhodné využití, protože je v něm velmi dobře vidět i velké množství hodnot, které se často právě v rámci časového období vyskytují. Vhodným použitím tohoto grafu může tedy být například zobrazení vývoje veličiny v čase. Spojnicový graf, anglicky Line chart, stejně dobře poslouží pro zobrazení několika proměnných zároveň. (Pour et al., 2018)

Dashboard

Dashboard je chápán jako kombinace reportů, kterými je předávána informace. Dashboards jsou hojně využívány, protože pro lidské oko jsou daleko přívětivější než tabulky plné nestrukturalizovaných dat. Data na dashboardu by měla být strukturalizovaná a logicky na sebe navazovat. Dashboard je celkem, který pomáhá uživatelům hledat odpovědi na určité byznysové otázky. Dashboard se skládá z grafů, KPI či tabulek, obecně tedy z vizualizací. Často také může nabízet určité typy filtrací, díky kterým se uživatelé snadněji orientují v datech, například filtry pro časové období. Obecně také platí, že by měl pomáhat s rychlou orientací v datech, jasně zvýrazňovat důležité informace a třeba i poskytovat novou perspektivu pro uživatele dashboardu. (J. Azevedo, Duarte a Santos, 2022)

Dashboard má za úkol manažera informovat o dění. Také by ho měl upozornit na výkyvy od normálních stavů sledovaných hodnot. Dashboards by neměly obsahovat nadbytečné množství informací, naopak by měly poskytovat shrnutí a upozornit v případě výchylek. Tvorba dashboardu má jistá pravidla, která by měla být dodržována pro co nejlhůdnější práci s dashboardem. Jedním z nejdůležitějších pravidel tvorby dashboardů je dělení na tři vrstvy. Jedná se o vrstvy rozdělené podle hloubky datového detailu. (Pour et al., 2018)

První vrstva obsahuje hodnoty velmi obecného rázu. Jedná se o agregované hodnoty, často jednotlivá čísla a jejich porovnání například s předchozím rokem. Tato první vrstva pomáhá datovému konzumentovi rychle najít případný problém. Například se v této první vrstvě mohou objevovat KPI, tedy klíčové ukazatele výkonnosti. Pokud se dashboard zabývá prodejem, v této sekci by se měly objevit KPI jako *Celkový objem prodeje oproti plánu*, *Počet prodaných kusů* s porovnáním oproti minulému období či *Průměrné prodeje na zákazníka*. Druhá vrstva či sekce by měla jít do středně velkých detailů. Zde se typicky objevují grafy. Tato část vysvětluje trendy ve větším detailu, pomáhá nalézt, kde dochází k problému. První vrstva totiž odhalí například informaci o tom, že podnik je oproti minulému roku ve ztrátě, ale nic víc. Druhá vrstva pomáhá najít, kde ke ztrátě dochází. Například v grafu je vidět rozpad dat přes jednotlivá střediska a začíná být viditelné, že několik středisek dlouhodobě prodělává. Přesně tyto informace jsou dostupné v prostřední vrstvě.

Poslední vrstva, vrstva nejnižší, ukazuje největší detail v datech a poskytuje data k detailní analýze. Detail může být například rozpad ve dnech, po konkrétních produktech či v jednotlivých městech. V této vrstvě dashboardu se většinou objevují tabulky, protože grafy od určitého bodu už nedokážou velké množství dat zobrazit dostatečně kvalitně. Příkladem tohoto limitu může být příklad zmíněný v předchozí kapitole, kdy sloupcový graf od příliš vysokého množství kategorií není čitelný. Zmíněné vrstvy se v dashboardu objevují buď pod sebou, tedy první vrstva je na stránce první a čím níže se uživatel dostává, tím větší detail dat mu dashboard poskytuje, nebo je dashboard tvořený zleva doprava a nejvíce agregované hodnoty čtenář nalezne vlevo, čím více se dostává doprava, tím větší detail konzumuje. Ať už je zvolena jakákoliv z těchto dvou možností, je potřeba zůstat konzistentní v její aplikaci napříč celým projektem, aby nedocházelo ke zmatení uživatelů. (Pour et al., 2018)

Mezi další pravidla tvorby dashboardu patří jednoduchý design. Na to je potřeba myslet už od začátku. Už při návrhu dashboardu by se mělo myslet na co největší zjednodušení grafického zobrazení. Měly by se omezit různé doprovodné texty a grafiky, které nepřinášejí faktickou informaci. Vynechat či minimalizovat by se měla loga nebo popisky, pokud nejsou nezbytně nutné ke správné interpretaci. Vynechat by se měl také přílišný detail či zobrazení KPI v grafu místo zobrazení pouze jednoho vypovídajícího čísla. Tomuto zobrazení jednoho čísla se říká návěstí s číslem či headline. Pro dosažení jednoduchosti by se také mělo zamezit používání 3D grafů a volit jednoduchou paletu barev. Základní barvy dashboardu by se měly pohybovat v modrých či šedých odstínech, tedy neutrálních barvách. K výrazným barvám, jako je červená, by se mělo přistupovat pouze

v případech, že je žádoucí zvýraznění. Dále je také dobré nepoužívat za grafy mřížku, pokud by vedla ke zhoršení čitelnosti. (Pour et al., 2018)

Celý dashboard by se měl ideálně dát zobrazit na jedné obrazovce. Protože dashboard by měl obsahovat pouze ty nejdůležitější informace, měly by se všechny vejít na jednu stránku. Pokud je potřeba zobrazit více informací je dobrým zvykem vytvořit další dashboard a přidat mezi tyto dva dashboards interakci přepínání či možnost se z jednoho dashboardu dostat do druhého. V BI se tato možnost nazývá drill-down. Drill-down může být odkaz na jiný dashboard, vyskakovací okno či zobrazující se poznámka. Drill-down se ale vždy zobrazí pouze při jasně vyjádřeném zájmu ze strany uživatele, tedy například při kliknutí do příslušného pole. (Pour et al., 2018)

1.3 Architektura v business intelligence

V této kapitole je detailně nahlíženo na vybrané komponenty BI systémů a na jejich význam z pohledu architektury BI projektu. První část kapitoly se zaměřuje na ETL procesy, které zajišťují extrakt, transformaci a nahrání dat, tedy cestu dat od zdroje k výstupům. Druhá část se věnuje samotnému konceptu architektury projektu: jak se architektura navrhuje, znázorňuje či jak se pracuje s byznysovými potřebami uživatelů. Poslední částí kapitoly je datové modelování jako klíčový prvek při práci s velkými objemy dat. (Pour et al., 2018)

1.3.1 ETL komponenta BI systémů

Komponenty jsou části BI řešení, které na sebe navazují, spolupracují spolu a dohromady tvoří architekturu BI řešení. Mezi základní komponenty BI řešení se řadí datové sklady, ETL a reporting. Koncept datových skladů už byl popsán výše, o ostatních komponentách bude pojednávat právě tato kapitola. (Pour et al., 2018)

První komponentou uvedenou v této kapitole bude nejdůležitější a nejčastěji zmiňovaná komponenta v souvislosti s BI a to ETL. Písmena v této zkratce zastupují slova Extract, Transform a Load. Pod pojmem extract je získání dat ze zdrojů. Transform zastupuje úpravy dat, jejich obohacení a uspořádání. Nahráním (Load) se rozumí nahrání dat do datových skladů nebo tržišť. Tato komponenta pomocí zmíněných tří procesů zajišťuje přenos dat mezi dvěma místy. Těmito místy mohou být databáze nebo datové soubory. ETL procesy jsou typicky opakujícího se rázu, podle potřeb firmy jsou data tímto procesem pouštěna v denních, týdenních, měsíčních či čtvrtletních intervalech. Pokud se jedná o nižší časový úsek, například každou hodinu či každých pět minut, dá se mluvit o Real-Time BI řešení. Čím častější aktualizace dat, tím vyšší náklady na provoz vznikají. Nejvyšší náklady vznikají při transformaci dat, ty přibližně představují 60 % nákladů spojených s provozem infrastruktury, na které ETL proces probíhá. (Pour et al., 2018)

O ETL je možné mluvit na dvou úrovních, může to být ETL jako koncept a ETL jako nástroj. ETL jako koncept je soubor procesů, zatímco ETL jako

nástroj je software, který pomáhá vývojářům a analytikům s budováním BI řešení. Součástí ETL jsou tzv. jobs, pipelines či flows, což jsou výrazy pro procesy spojené s tokem dat. Tyto datové procesy (jobs) mohou být vyvíjeny pomocí kódu typicky SQL, Python nebo Java. Zároveň ale existují ETL nástroje jako Informatica, IBM DataStage či Pentaho. Používání ETL nástrojů je v dnešní době mnohem typičtější než vytváření celého ETL procesu pouze pomocí kódu. (Kumar V. K., 2022)

Alternativou k ETL, je ELT přístup. Proces začíná stejně, datových extraktem, ale poté jsou data rovnou nahrána do cílového nástroje, kde jsou později zpracována. Tento přístup se využívá při použití no-sql databází či cloudových řešení. Toto řešení přináší nevýhody v podobě složitější implementace či nedostatečného množství odborníků, protože koncept je poměrně nový. Na druhou stranu může přinášet výhody jako například lepší časovou dostupnost dat (mezi nahráním dat není transformační vrstva, která může zpomalit dostupnost) nebo menší náklady na hardwarové vybavení. (Nordeen, 2020)

1.3.2 Architektura BI projektu

BI architektura je soubor komponent, které dohromady tvoří BI řešení odpovídající byznysovým požadavkům. BI architektura, tak jako další softwarové architektury, se dají dobře znázornit diagramem či schématem. Tento diagram by měl odpovídat detailu podle potřeb čtenářů a jejich odbornosti, existují schémata architektury, která jsou velmi málo detailní a popisují spíše koncept samotného BI řešení než jeho provedení. Na druhé straně architektura projektu může být popsána do velmi malých detailů, vždy závisí na skupině osob, pro kterou je náčrt architektury projektu zamýšlen. (Kumar V. K., 2022)

Obecně se v datové architektuře objevují komponenty jako datové zdroje, datová úložiště, datové vrstvy (čistící, přípravné, sémantické), nástroje pro datovou vizualizaci a uživatelé. Množství komponent se odvíjí od byznysových požadavků a obecně by měla BI architektura přinášet efektivní řešení, které v rámci poskytnutého rozpočtu bude schopné tyto požadavky naplnit. Při její tvorbě je potřeba vzít v potaz cíl BI řešení, firemní limity, finanční stránku věci, požadavky jednotlivých skupin uživatelů i technické nástroje dostupné na trhu a jejich limity. (Kumar V. K., 2022)

1.3.3 Datové modelování

Big data, tedy obrovské datové objemy, už nejsou v byznysu žádným novým pojmem. Firmy se kvůli jejich zpracování setkávají s velkým množstvím překážek, ať už se jedná o jejich sběr, uchovávání či analýzu. Datové modelování je pro manipulaci s big daty prakticky standard. Datové modelování je způsob, který pomáhá lidem získat vhled do dat. Díky datovému modelování data získávají lepší strukturalizaci a kategorizaci a zároveň se v nich dá lépe orientovat. Datové modely mohou vizualizovat jak základní koncepty fungování firmy, tak detailní úrovně datové logiky. Jejich hlavním přínosem je zjednodušení, kterého se dosahuje

díky primitivním vizualizačním prvkům například šipkám pro znázornění vztahu mezi objekty. (Lee, Wei a Mukhiya, 2018)

Existuje dělení na tři základní typy datových modelů:

- konceptuální model,
- fyzický model,
- logický model.

Konceptuální model znázorňuje podstatu projektu, tedy jaké výchozí stavy existují. Jeho úkolem je zmapovat koncepty byznysu, pro který se datový projekt tvoří. Jedná se o velmi obecné pojetí projektu a často tento model tvoří datoví architekti. Fyzický model oproti tomu řeší technickou stránku řešení, popisuje, jak bude řešení provedeno, zabývá se technickými požadavky, velikostí řešení a na základě těchto potřeb navrhuje provedení a nástroje, které budou pro řešení nejvhodnější. Logický model řeší, jak bude řešení provedeno neohledě na technické požadavky. Mapuje potřeby ze strany byznysu, ale mnohem detailněji než konceptuální model. Zabývá se návrhem konkrétních postupů, pravidel a datových procesů. Často na těchto modelech pracují datoví analytici. (Nordeen, 2020)

Pro logické datové modelování je typické, že popisuje data sety a vztahy mezi nimi. Příkladem může být data set *Studenti*, ve kterém každý student má jméno a ID, tedy unikátní označení. Druhým data setem může být *Škola*, ke které student patří. Tato škola má opět atributy jako název a unikátní ID. Pro nalezení vztahu mezi školou a studentem, musí existovat vazby, které tyto vztahy popisují. Tyto vazby se definují pomocí primárních a cizích klíčů. Primární klíč je specifická hodnota, například ID, nebo kombinace více hodnot. Pro primární klíče musí platit jejich unikátnost v data setu, tedy každá škola by měla mít svůj jedinečný primární klíč, který ji odliší od ostatních, stejně tak student. Cizí klíč, je hodnota primárního klíče v jiném data setu, například ID školy v data setu *Studenti* značí vazbu studenta k dané instituci. (Lee, Wei a Mukhiya, 2018)

Dimenzionální model datového skladu

Při skladování dat v datových skladech je běžné využívat datové modely, aby data byla strukturovaná a lépe se v datech orientovalo. Existuje několik doporučených přístupů při tvorbě datových modelů. Mezi základní techniky při tvorbě modelů patří star (hvězda), tedy hvězdicové schéma a snowflake (sněhová vločka), tedy schéma sněhové vločky. (Eagar, 2023)

První rozebíranou technikou bude architektura hvězdy. Jak název napovídá, jedná se o model, kde je jedna centrální tabulka, obsahující fakta, tedy číselně vyjádřené hodnoty například cenu či množství. Okolo této faktické tabulky jsou rozmístěny dimenze, tedy tabulky obsahující informace nečíselného typu. Tyto informace dodávají kontext, tedy například tabulka zaměstnanci, která obsahuje údaje o zaměstnancích či tabulka produkty, která může obsahovat typ produktu nebo jeho kategorii. Tyto informace se pomocí primárních klíčů v dimenzi a cizích klíčů ve faktové tabulce na sebe navážou. Mezi dimenzemi a faktovou tabulkou

platí vztahová vazba 1:N, tedy pro jeden záznam v dimenzionální tabulce může existovat více záznamů ve faktové tabulce. Ve hvězdicovém schématu jsou většinou normalizovány faktové tabulky a denormalizovány tabulky dimenzionální, a to z praktických důvodů. Díky této denormalizaci mohou všechny atributy související s danou dimenzí být pohromadě v jedné tabulce. To je praktičtější při pozdějším dotazování do těchto tabulek. Díky denormalizaci je totiž značně zkrácen čas dotazování a pochopení modelu analytikem je také snazší. Při větších datových objemech se ale s touto denormalizací může snadno objevit nežádoucí množství duplicit a některé atributy mohou být nekonzistentní. (Eagar, 2023)

Právě v případech, kdy se začínají objevovat tyto nežádoucí problémy, je načase přejít k druhému schématu, a to schématu sněhové vločky. Toto schéma také drží uprostřed faktovou tabulku, ale narozdíl od hvězdicového schématu jsou všechny dimenze normalizované. Každá dimenze se tedy rozdělí na několik dalších dimenzí. Například původní dimenze Sklad Adresa, kde byl uvedený název skladu a jeho kompletní adresa, tedy jedna tabulka obsahující sloupce Město, Region, Země je normalizovaná a rozdělená do několika dimenzí, kde první obsahuje pouze id města, název města a cizí klíč pro region, druhá obsahuje primární klíč regionu, název regionu a cizí klíč pro zemi, přičemž poslední obsahuje primární klíč země a její název. Tedy je kladen důraz na to, aby všechny tabulky obsahovaly pouze atributy, které s nimi logicky přímo souvisí. Schéma sněhové vločky většinou vzniká právě rozšířením hvězdicového schématu, ale může být vytvořeno i rovnou na začátku projektu, pokud datový architekt usoudí, že by díky typu skladovaných dat mohlo docházet k výše zmíněným problémům. Schéma sněhové vločky minimalizuje duplicitní hodnoty, čímž šetří kapacitu úložiště. Na druhou stranu však tvoří složitou hierarchii, kde se při dotazování musí vytvářet velké množství „joinů“, tedy všechny dimenze se na sebe musí napojit, aby byla získána kompletní informace. To může vést k většímu výpočetnímu výkonu a dotazy tedy budou časově náročnější. (Eagar, 2023)

1.4 Business intelligence v podnicích

Jak je již patrné z názvu kapitoly, tato kapitola se zabývá vývojem a implementací BI, ale tentokrát z pohledu podnikového prostředí. V první části je pojednáváno o procesech vývoje BI řešení jako o procesu vývoje jakéhokoliv jiného podnikového softwaru. Je zde kladen důraz na plánování, porozumění byznysovým požadavkům a volbu technických nástrojů. Poté následuje kapitola řešící faktory úspěšné implementace BI v podniku, kde jsou rozebírány předpoklady a překážky úspěšné implementace. Kapitulu zakončuje zamyšlení se nad vlivem BI na rozhodovací procesy v podnicích.

1.4.1 Vývoj BI řešení v podniku

Tato kapitola shrnuje postup při vývoji BI řešení, tedy jeho vytvoření a kroky, které bývají jeho součástí. Vývoj BI řešení je, stejně jako vývoj jiného softwaru,

projekt a je potřeba k němu jako k projektu přistupovat. Před jeho zahájením je potřeba si jednotlivé části naplánovat, dohodnout se na výstupech a cílech a zároveň si potvrdit, že podmínky ve firmě, ve které se tento vývoj bude konat, jsou adekvátní a nebudou vývoji bránit. (J. Azevedo, Duarte a Santos, 2022)

Před zahájením samotného vývoje je nutné pochopit, co má být jeho cílem. Tedy jaké jsou byznysové požadavky firmy. Do těch se zahrnuje jak pochopení potřeb budoucích uživatelů, tak pochopení struktur a firemní kultury. K těmto krokům je vhodné znát například firemní činnosti, cíle, rozpočet či vztah k inovacím ve firmě. Podle těchto informací se dále volí například komplexita a složitost řešení, technické nástroje či vizualizační nástroje. (J. Azevedo, Duarte a Santos, 2022)

Poté co jsou zvoleny všechny technické nástroje, sepsány požadavky a unifikovaná očekávání na obou stranách - tedy na straně zadavatelů i tvůrců řešení, začíná se řešit datový aspekt projektu. Pokud nejsou posbíraná potřebná data, začne se s jejich sběrem a ukládáním. Posbíraná data se analyzují a zjistí se, jestli jsou dostačující, nebo je potřeba zlepšení datové kvality, či sběru dat jiných. Dalším krokem je ETL proces, který je teoreticky popsán v dřívějších kapitolách. Po vytvoření ETL procesu následuje tvorba datových vizualizací, tedy dashboardů ve vybraném vizualizačním nástroji. V průběhu všech předchozích kroků je ideální, aby všechny zainteresované strany spolupracují průběžně a dochází k pravidelným konzultacím a validaci. Ty jsou dobré k tomu, aby se projekt neodkláněl od požadavků zadavatelů, a aby BI odborníci měli všechna potřebná data a chápali je správně v kontextu dané firmy. (J. Azevedo, Duarte a Santos, 2022)

1.4.2 Faktory implementace BI v podniku

V dnešní době je kladen velký důraz na efektivitu podniků a s ní spojené inovace, díky kterým se podniky snaží této efektivitě dosahovat. Úspěšná implementace BI řešení může být jednou z těchto inovací. Vytvoření BI řešení, ale není kolikrát samo o sobě dostačující a existuje několik faktorů, které musí fungovat v symbióze s BI řešením. V této kapitole budou probírány některé podmínky pro úspěšnou implementaci BI řešení v podniku. (Zelenka a Podaras, 2021)

Prvním z faktorů vypovídajícím o výsledku implementace je kvalita vstupních dat. Tedy výsledný výstup je závislý na vstupech. Při nekvalitních vstupech - nekonzistentních datech, chybějících hodnotách, duplicitních hodnotách a nedodržování jednotných postupů při vyplňování dat se firma musí potýkat s nižší kvalitou datových výstupů. K dobré kvalitě dat přispívá porozumění datům, tedy správné pochopení jejich významu v souvislostech. Pro zvýšení kvality vstupních dat tedy podniky mohou zapracovat na zlepšení porozumění datům. S problémy týkající se datové kvality může pomoci Data Governance, která má za úkol jasně přerozdělit zodpovědnost za datovou kvalitu mezi konkrétní zaměstnance. (Zelenka a Podaras, 2021)

Druhým z faktorů úspěšné implementace je kvalita procesu adopce. Proces adopce spočívá ve spolupráci mezi osobami, které BI řešení vyvíjejí, a uživateli. Adopce vede k lepšímu pochopení BI procesu ze strany koncových uživatelů. Může pro ně sloužit i jako jistá forma edukace a přípravy pro pozdější užívání, často jsou

právě oni v roli osob s rozhodovací pravomocí a jejich požadavky jsou zapracovávány. Správně nastavený proces adopce také přináší snadnější řízení projektu nebo snazší nadefinování datových standardů. Na BI řešení mohou spolupracovat byznys analytici, IT odborníci, BI odborníci a stakeholders, tedy osoby jako zaměstnanci, zákazníci, manažeři a další budoucí uživatelé výstupů. Je tedy poměrně složité, aby se všichni shodli, ale vzájemná spolupráce BI projekty také velmi posouvá kupředu. (Skyrius, 2021)

Naopak překážkou v implementaci může být neochota uživatelů, tedy typicky zaměstnanců, se s novým BI řešením naučit pracovat na uživatelské úrovni. Procesu seznámení zaměstnanců s BI řešením se říká adopce uživateli. Neochota se může týkat jak jednotlivců, tak celých oddělení. Bohužel pro tento případ jsou uživatelé nedílnou součástí implementace a bez jejich spolupráce a ochoty není prakticky možné implementaci považovat za úspěšnou. Možností pro zmírnění této neochoty adopce může být zmírnění nedůvěry vůči datům díky validování dat a začleňováním uživatelů do procesu průběžně. Také je dobré dostatečně proškolení všechny zaměstnance a nabídnout jim nejen BI řešení, které jim opravdu pomáhá s jejich každodenními úkoly, ale i dostatečné vysvětlení tohoto faktu. (Mihiranga, 2022)

1.4.3 Role a funkce v BI: datový inženýři, analytici a vědci

Pro každý větší datový projekt je potřeba několik specialistů na konkrétní části řešení. Podle velikosti projektu je pak možné se bavit buď o jednotlivých odbornících nebo o celých odděleních zaměřujících se na dané datové obory. Pro zjednodušení se zde bude bavit pouze o jednotlivcích, ale stejné role se mohou ve větších firmách vztahovat na jednotlivé týmy či dokonce oddělení. Mezi základní role patří datový inženýr, datový analytik a data science odborník, tedy volně přeloženo jako datový vědec. Tito odborníci by spolu měli spolupracovat pro dosažení co nejlepších datových výsledků. (Eagar, 2023)

Role datového inženýra spočívá ve vytvoření datových toků, které zajišťují přesun surových dat do datových skladů, přípravu těchto surových dat pro datové analytiku a vytvoření datových toků tak, aby tyto toky pro jednotlivé datové konzumenty poskytovaly právě ta data, která potřebují. Zároveň musí datový inženýr znát bezpečnostní a právní předpisy, které se k jeho práci s daty vztahují a nástroje, které jsou na trhu pro jeho práci dostupné a které je vhodné využít. Je tedy zodpovědný za datovou infrastrukturu projektu. Jeho práce je většinou technického rázu a obsahuje minimální přesahy do byznysu. (Eagar, 2023)

Funkce datového analytika spočívá ve zkoumání datových souvislostí mezi data sety za účelem nalezení datových trendů. Primárním cílem datových analytiků je poskytnout byznysově orientovaným osobám informace, na jejichž základě poté tyto osoby mohou dělat datově podložená rozhodnutí. Datový analytici úzce spolupracují s lidmi z byznysu a snaží se pomocí dat zjistit odpovědi na jejich otázky. Dosahují toho pomocí komplexního dotazování do databází či jiných datových úložišť. (Eagar, 2023)

Role datového vědce je ovládat oblasti jako machine learning, umělá inteligence, pokročilé statistické metody, matematika či programování. Pomocí těchto odborných analýz je schopen predikovat složitější trendy a odkrývat na první pohled skryté datové souvislosti. Datoví vědci trénují machine learning modely také za účelem podpoření byznysových rozhodnutí, stejně jako datoví analytici. Na rozdíl od nich ale využívají spíše nástrojů prediktivní analýzy, zatímco datoví analytici se často zaměřují na deskriptivní analýzu podniku. (Eagar, 2023)

1.4.4 Vliv BI na rozhodovací procesy v podnicích

BI má mnoho využití, v této kapitole budou její role na rozhodovací procesy rozděleny na tři kategorie: využití BI při procesech strategického rozhodování, využití BI při provozních procesech a využití BI při procesech souvisejících s datovou integrací. Schopnost poskytnout důležité informace při strategických rozhodnutích znamená, že je BI schopná připravit informace pro vytváření firemních strategií a podpořit rozhodovací procesy. Konkrétně může jít o vyhodnocování rizik, identifikaci nových příležitostí či nalezení tržních trendů. Všechny tyto podklady pak může rozhodující osoba (decision-maker) využít. Pokud je řešena optimalizace zavedených firemních procesů či sdílení informací napříč podnikem jedná se o využití BI při operativních procesech. Posledním zmíněným okruhem je využití BI pro datovou integraci, kdy se podniky snaží o sjednocení jednotlivých datových zdrojů mezi sebou. K těmto procesům dochází za účelem nalezení spojitostí. Protože více oddělení může pracovat se stejnými daty, je pro firemní komunikaci podstatné, aby existovala celopodniková shoda a jedno místo pravdy. (Alsaad et al., 2022), (Mudau, Cohen a Papageorgiou, 2024)

1.5 Trendy v business intelligence

Následující kapitola se zaměřuje na aktuální vývoj a trendy v oblasti BI. Jsou zde rozebírány trendy jako Data Storytelling, Data Governance a Self-Service BI, které začínají hrát stále významnější roli v podnikovém prostředí.

Mezi trendy, které se v oblasti BI začínají stále více objevovat, se dá řadit:

- Data Storytelling,
- Data Governance,
- Self-Service BI

1.5.1 Data storytelling

Data storytelling je metoda komunikace, díky níž lze předávat informace ve formě narativů tak, aby si příjemce tyto informace snáze zapamatoval. Cílem data storytellingu je předat data relevantním a dostatečným způsobem, ideálně v rámci příběhu, protože díky tomu je na jejich základě možné stavět racionální

rozhodnutí. Člověk, který se má na základě těchto dat rozhodovat nebo se jimi řídit, musí znát jejich význam a příklady použití, aby jeho kroky vedly k smysluplným rozhodnutím. (Mihiranga, 2022), (Ryan, 2018)

Pokud na data storytelling bude nahlíženo nejenom ze strany konzumenta, ale i ze strany zadavatele řešení, je i v jeho zájmu, aby se naučil správně interagovat se svým dodavatelem a dokázal správně vydefinovat svoje datové potřeby. Tato schopnost je úzce spojená s poskytováním narativů. Zadavatel by měl prioritizovat svoje potřeby a nechat dodavatele řešení pochopit přidanou hodnotu, kterou pro něj BI řešení poskytuje. Každý zákazník totiž profituje z lehce jiných datových výsledků a správná komunikace je základem pro předání těchto informací. Uvádění příkladů z praxe a ukázky reálného používání výstupů ve firmě na straně zadavatele jsou velmi přínosné pro obě strany, protože jenom díky správnému pochopení může dodavatel vytvořit řešení přesně na míru. (Boldosova, 2020), (Ryan, 2018)

1.5.2 Data governance

Pojmem data governance jsou chápány procesy, které hlídají dostatečnou datovou kvalitu a řeší formality spojené se zpracováním dat napříč celou společností, například sjednocená pravidla pro vyplňování dat do datových zdrojů, a zajišťují spolehlivost dat. Cílem data governance je dosažení vysoké datové kvality, datové konzistence a datové spolehlivosti. Také se zabývá bezpečností, která je stále palčivějším tématem. K dosažení těchto cílů může vést například jasné rozdělení zodpovědnosti za výše zmíněné datové standardy, zajištění minimálního opakování vstupů a zpracování. Tedy pokud ve firmě existují čtyři zdroje dat týkající se produktových skupin, bude potřeba skupiny sjednotit, najít případné rozdíly, pochopit je, odstranit a určit jednu produktovou skupinu, která bude tzv. místem pravdy, a ze které bude možné spolehlivě čerpat. Z logiky věci je data governance nejvíce řešená ve velkých často i mezinárodních společnostech, finančních institucích a ve státní správě. (Pour et al., 2018)

1.5.3 Self-service BI

Self-service BI, často uváděné pod zkratkou SSBI, je bezesporu dalším trendem objevujícím se v odborné literatuře. Jedná se o BI řešení, které funguje bez trvalé přítomnosti IT odborníků a BI specialistů. Principem tohoto přístupu je převést méně technické úkony na běžné (netechnické) uživatele a maximum procesů automatizovat. Prerekvizitou jsou tedy jednoduché BI nástroje, které nevyžadují přílišné technické znalosti od jejich běžných uživatelů. Tato forma BI přináší mnohá pozitiva, ale i jisté překážky. O obojím bude pojednááno níže. (M. Pałys a A. Pałys, 2023)

Mezi pozitiva SSBI je řazeno menší zatížení technických a odborných pracovníků z IT a BI oblastí. Neznamena to, že by tito lidé již nebyli potřeba, ale značně to může zmenšit objem práce, která je od nich v současné době vyžadována. Tito odborníci se mohou díky řešení, které funguje samo o sobě, soustředit na složitější a více technické úkoly. Hlavním přínosem této automatizace

BI řešení, je dodávání reportů pravidelně a bez zpoždění ze strany BI/IT týmů. Díky SSBI se BI odborníci stávají spíše konzultanty a podporou např. při tvorbě náročnějších reportů. Tím, že se určité úkony přenáší na běžné uživatele se zrychlí tvorba reportů, protože se práce rozprostře z původních pár BI odborníků na všechny běžné uživatele a každý z nich si může vytvořit reporty na míru. Díky tomu se také může zvýšit přesnost reportů, uživatelé lépe ví, co potřebují a informace se neztrácí při předávání požadavku na další osoby. Začleněním běžných uživatelů do datového procesu se v podnicích zvyšuje datová maturita a zaměstnanci více chápou důležitost dat, jejich rozhodnutí mohou začít být více datově podložená a zároveň se může začít dít, že budou lépe zpracovávat zdrojová data, protože pochopí důležitost jejich konzistence. (M. Pałys a A. Pałys, 2023)

Na druhé straně s sebou ale SSBI nese i jistá rizika a překážky. Pokud s daty pracují zkušení odborníci snadno si poradí i s daty, které nejsou v ideální datové kvalitě či obsahují nestandardně vyplňované hodnoty. Data jsou schopni na rozdíl od běžných uživatelů vyčistit a obohatit. V případech, kdy to není možné, jsou také schopni tuto skutečnost vyhodnotit a data nepoužívat k tvorbě analýz, čímž se vyvarují utvoření závěrů, které mohou být zkreslené kvůli nízké datové kvalitě. Řešením tohoto problému může být nastavení jistých podmínek, podle kterých běžný uživatel pozná, jestli se jedná o data set, který je datově až příliš nekvalitní nebo se na něm nedá provádět žádná smysluplná analýza. Oko zkušeného datového analytika zde ale vždy bude chybět. (M. Pałys a A. Pałys, 2023)

SSBI řešení musí být jednoduchá a uživatelsky přívětivá. Běžný uživatel musí dobře chápat řešení, které využívá. Při práci s komplexnějšími řešeními, není tedy SSBI vhodné, protože běžný uživatel může nevědomky špatně interpretovat data. Běžní uživatelé preferují vizuální uživatelské prostředí před formou přístupu k informacím takovým způsobem, u kterého by museli psát kód. Nejdůležitější je důsledné proškolení uživatelů, protože pokud nedostanou dostatek informací a BI řešení odpovídající jejich schopnostem, snadno se stane, že zahltí BI/IT oddělení ještě více dotazy než doposud a SSBI se stane neefektivním. Uživatelé by také měli chápat BI jako pozitivní inovaci, která jim může usnadnit jejich práci. Pokud dostatečně nechápu přínosy nebo z řešení žádné přínosy neplynou, začne u nich nechuť či dokonce odpor vůči BI řešení, pravděpodobně se vrátí k jejich původním zvykům, např. opět začnou vytvářet tabulky v Excelu. (M. Pałys a A. Pałys, 2023)

1.6 Software technické podpory

Protože data, se kterými se v praktické části práce bude pracovat jsou ze softwaru, který je využíván pro technickou podporu BI projektů, tato kapitola bude rozebírat některé pojmy týkající se právě softwaru pro technickou podporu, někde uvážených také jako systémy generující požadavky (tikety) - tiketovací systémy či systémy IT podpory. Tyto IT systémy se zavádí v případě většího množství požadavků. Klienti mohou těmito požadavky pokládat dotazy či stížnosti. Tyto požadavky mohou být buď vytvářeny samotnými klienty nebo být automaticky generovány, a to zejména v případě poruchy či chyby. (Zangari et al., 2023)

Samotné tikety jsou přiřazeny jednotlivci či týmu, který je začne řešit a komunikuje se zadavatelem tohoto tiketu. Těmto týmům se říká většinou Support týmy, do češtiny se výraz buď nepřekládá nebo se mluví o podpoře. Klienti, kteří využívají tuto službu jako servis mají většinou smluvně podchycené podmínky jejího používání a podmínek, za kterých se bude tiket řešit. Těmto smluvním podmínkám se říká SLA (Service-Level Agreements), což by se volně dalo přeložit jako dohoda o úrovni poskytovaných služeb. (Zangari et al., 2023)

1.7 Shrnutí teoretické části

V teoretické části této diplomové práce byly zpracovány termíny a koncepty business intelligence. V úvodu kapitoly bylo pojednáváno o tom, co to BI je a kdy se prvně začala objevovat. Také zde byly zmíněny primární termíny, týkající se dat a informací z pohledu business intelligence. V druhé podkapitole byly rozebírány koncepty BI, a to v posloupnosti podle toho, jak se data přes datovou pipeline dostávají, tedy první byly rozebírány koncepty datových skladů, poté datová analýza a reporting a poslední část této podkapitoly se věnovala datovým vizualizacím, jejich typům, použitím a dashboardům. Poté, co byly zpracovány jednotlivé koncepty a vysvětleny termíny týkající se BI, které s nimi souvisí, nastal prostor na rozebrání architektury v BI projektech, která se týkala základů datového modelování, rozebírány byly typy modelů a schémat dimenzionálních modelů v datových skladech. Předposlední podkapitola se věnovala BI v podnicích z pohledu jejich vývoje a implementace. Zde byly probrány podmínky těchto procesů, doporučené postupy a termíny, které se k těmto procesům vážou, například adopce uživateli. Také zde byly vysvětleny jednotlivé BI funkce osob v podniku. Poslední kapitola se věnovala aktuálním trendům v BI, jakými je data storytelling, data governance a self-service BI.

V této části byly kromě vysvětlení teoretických termínů popsány i best practices: jak skladovat data, analyzovat, tvořit datové modely, vizualizovat i provádět předání datových výstupů uživatelům. Tedy vhodné postupy při tvorbě end-to-end řešení od dat až k smysluplným vizualizacím. Tyto poznatky budou v další části využity při zpracovávání části BI řešení v praxi. Tyto znalosti získané při zpracování teoretické části se budou v práci využívat i v následujících kapitolách, kde budou použity při rozšíření aktuálního BI řešení o nový datový zdroj. V následující části bude z teoretických poznatků konkrétně využita datová architektura, primárně tvorba datových vrstev a datového logického modelu, také bude zpracována analýza a vizualizace dat, návrh dashboardu podle doporučených teoretických kritérií a jeho reálné vytvoření ve vizualizačním softwaru.

2 Zmapování stavu firemního řešení a požadavků uživatelů

Tato kapitola se bude zabývat popisem stavu aktuálně fungujícího interního BI řešení ve firmě BizzTreat. Aby bylo možné v dalších kapitolách navázat na toto řešení a rozšířit ho o nový datový zdroj, je prerekvizitou zmapování stavu současně fungujícího. V rámci této kapitoly bude na zmapování firemního stavu pohlíženo z několika různých pohledů. Nejprve se kapitola bude zabírat využívanými technologiemi, které jsou v tomto řešení využívány. Dále zde bude popisován aktuální stav z pohledu byznysových oblastí, se kterými už BI řešení úspěšně pracuje. Třetím úhlem pohledu na věc bude pohled z hlediska datové architektury, tedy jak se přistupovalo v minulosti k datovým vrstvám u ostatních datových toků či jak datoví architekti vymysleli ETL.

2.1 Využívané technologie

V kapitole bude rozebíráno BI řešení z pohledu technologií, se kterými se v rámci interního projektu pracuje. První část řešení je BigQuery. BigQuery je platforma, která funguje jako datový sklad. Na této platformě jsou uchovávána všechna data, která jsou pro potřeby interního řešení využívána, žádný další datový sklad zde nefiguruje.

Druhou důležitou komponentou BI řešení je ETL. ETL je zde řešeno interně vytvořenou aplikací, která se nazývá BizzFlow. BizzFlow je nástroj napsaný v programovacím jazyce Python a kooperující se softwarem Apache Airflow. BizzFlow nabízí velké množství datových extraktorů pro různé datové zdroje. Tyto extraktory extrahují data z nejrůznějších databází, aplikací či dokumentů a zapisují je právě do datového skladu. BizzFlow je kompatibilní s větším množstvím datových skladů, ale v rámci tohoto řešení, jsou zapisována data právě do BigQuery. Pro přidání nového datového zdroje se přidává konfigurační soubor do repozitáře řešení a nastavují se parametry pro napojení na konkrétní end pointy. BizzFlow také pomáhá analytikům s pravidelným spouštěním jednotlivých SQL skriptů, kterým se říká transformace či transformační skripty, protože tvoří právě transformační část ETL. Aby se dosáhlo pravidelného spouštění těchto transformačních skriptů, dotazuje se BizzFlow právě do BigQuery, kde tyto skripty pravidelně spouští. BizzFlow podporuje větší škálu cloudových služeb. Infrastruktura interního projektu je postavena na infrastruktuře Google Cloud.

Poslední klíčovou technologií, která je v řešení využívána, je platforma GoodData. GoodData je vizualizační nástroj. Tento nástroj poskytuje tvorbu základních vizualizací i datové modelování či spravování uživatelských práv. Je zde možná tvorba dashboardů a jednotlivých vizualizací, kterým se říká insighty. Také platforma GoodData nabízí možnost vytvořit metriky. Metriky jsou jednoduché funkce pro zpracování dat, data agregují, nabízí jednoduché podmínky či filtrace. Složitější obohacování dat je tedy efektivnější provádět v transformační části, ještě před nahráním dat do GoodData, ale pro jednodušší úpravy jsou tyto metriky zajímavou možností. Metriky se píšou v dotazovacím jazyce, který vychází z SQL a je nazýván MAQL. Detailnější ukázka těchto metrik bude uvedena v rámci kapitoly o implementaci řešení.

2.2 Existující byznysové oblasti a datové zdroje

V této kapitole bude BI řešení zmapováno z pohledu již existujících a funkčně nasazených podnikových částí, které jsou datově podchyceny. Tato kapitola bude na řešení pohlížet z pohledu datových vizualizací, protože z byznysového pohledu jsou to právě primárně reporting, díky kterému může být podnik data-driven. Aktuálně se sledované řešení skládá z 33 produkčních dashboardů, tedy dashboardů, ke kterým uživatelé mají přístup a nějakým způsobem je využívají v rámci svých činností. Dashboardy jsou roztřizeny do skupin podle byznysových oblastí a ty jsou poté nasdíleny cílovým uživatelům. Alespoň k jednomu dashboardu má přístup každý z firmy BizzTreat. Dále je zde přibližně 5 dashboardů ve vývoji, pro tyto dashboardy se buď sbírají či zpracovávají data nebo na nich interní tým postupně pracuje. Dále se v interním řešení nachází určité množství osobních dashboardů, u těch ale není možné určit množství, protože jsou z logiky věci přístupné pouze pro jejich tvůrce. Jedná se zde tedy o část řešení, která je Self-Service BI, protože je tvořena vyloženě datovými konzumenty, pro čistě jejich konkrétní potřeby a tyto dashboardy či reporty si vytváří sami bez zásahu interního analytického týmu. Právě proto není možné určit jejich počet.

Největší část řešení se zabývá daty, které se týkají dodávání hlavní služby, kterou BizzTreat poskytuje, tedy právě datových projektů na míru. Zde jsou datově řízeny plánované projekty, interní kapacity, množství dodaných služeb, časové odhady na vývoj či další informace týkající se právě dodávání těchto služeb. Datovými zdroji pro tyto informace jsou datové soubory z následujících aplikací:

- Google Sheets,
- Paymo,
- Trello,
- Hubspot (CRM).

Z Google Sheets jsou získávána data týkající se plánování kapacit, a to jak kapacit alokovaných na projektech, tak kapacit, které se přerozdělují mezi jednotlivé týmy.

Tyto data jsou ručně vyplňovaná do předem nadefinovaných struktur, aby nedocházelo k chybám způsobených lidským faktorem. To může být například přeskočení sloupce, či záměna datového typu. V softwaru Paymo jsou sledovány odpracované hodiny na klientských projektech i na podpůrných činnostech. Trello pro firmu slouží jako nástroj projektového řízení, jsou z něj sbírána data týkající se odhadů, seznamy dílčích částí projektů a jejich pokrok v časovém období. Tato data pomáhají identifikovat situace, kdy by mohlo dojít k nesouladu mezi plány z Google Sheets a odhady jednotlivých dílčích částí v rámci Trello kartiček. Hubspot je CRM software, ze kterého jsou extrahována data týkající se zakázek a klientů. Posledním datovým zdrojem pro data řídicí dodání jsou finanční data z účetního softwaru, tato data jsou vysoce citlivá, a proto se dostávají do projektu v agregované podobě bez zbytečně detailních informací.

Další velmi rozsáhlou oblastí, která je pokryta bezmála šesti dashboardy je prodej. Pro obchodní potřeby jsou zde sbírána data ze softwaru Hubspot, hlídá se zde hodnota jednotlivých obchodních zakázek, k dotvoření pohledu se využívají informace o aktivitách, které obchodníci a account managers generují. To pomáhá analyzovat výkonnost firemního prodejního týmu. Zároveň jsou zde monitorována i marketingová data, která jsou sbírána ze sociální sítě LinkedIn a dotvářejí obrázek o dosazích, které obchodníci na těchto sítích mají.

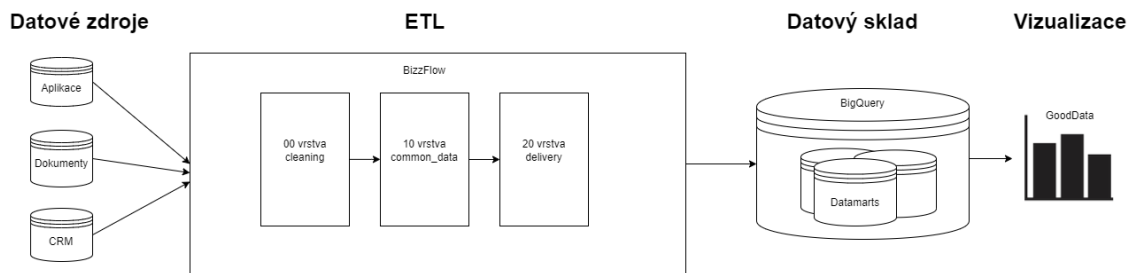
Firma si také bedlivě hlídá datovou kvalitu a existují vyhrazené dashbordy právě pro tyto účely. Tyto dashboardy jsou interně nazývány Sanity dashboardy a slouží k automatizované kontrole konzistence dat. Reporty zde dokáží snadno identifikovat případné chyby v datech a upozorní osobu, která je za ně zodpovědná, že je žádoucí oprava těchto dat. Tímto způsobem je ošetřena datová kvalita a případné nesrovnalosti jsou odchyceny.

2.3 Datová architektura

V rámci datové architektury jsou data rozdělena do několika skupin na: raw data, vstupní data a výstupní data. Raw data jsou nezpracovaná data, která ještě neprošla čištěním. Jedná se tedy o ta data, která byla pouze extrahována a jsou uložena v tabulkách s prefixem `raw_ex_`. Pro každou vrstvu (vrstvy jsou probrány v rámci této kapitoly níže) existuje vstupní a výstupní tabulka. Vstupní tabulky jsou zdrojem pro danou vrstvu a označují se prefixem `in_`. Výstupní tabulky jsou vždy poslední částí transformační vrstvy a označují se prefixem `out_`. Dále je možné se v projektu setkat s pomocnými tabulkami. Existují zde ještě dva typy pomocných tabulek. Tabulky s prefixem `cln_`, u nichž se jedná o čistící tabulky a tabulky s prefixem `tmp_`, což jsou pomocné tabulky, které tvoří přípravu pro výstupní tabulky a probíhá v nich příprava logických dílčích celků.

V rámci datové architektury projektu jsou využívány datové vrstvy pro lepší přehlednost kódu. V rámci interního projektu se jedná o tři typy vrstev pojmenované interně podle číselné logiky, a to na nula nula vrstvu, desítkovou vrstvu, dvacítkovou vrstvu a třicítkovou vrstvu. Vrstvy jsou zde členěné tak, že všechny složky mají prefix podle toho, ve které vrstvě se nacházejí. V nula nula vrstvě (00) jsou data čištěna a je

zde filtrován počet sloupců, které je potřeba zpracovávat. V desítkové vrstvě (10) se odehrávají transformace, pro tabulky, které jsou společné pro více datových zdrojů v rámci projektu. Příkladem tedy jsou hlavně dimenze typu projekty, týmy a další podobné tabulky, které jsou pro všechny datové zdroje sjednoceny. V této vrstvě tedy dochází k jejich zpracování, aby se nemusely pro každý datový zdroj zpracovávat odděleně. Vrstvou, která zde obsahuje největší množství skriptů je dvacítková vrstva (20), zde se totiž tvoří hlavní byznysová logika a obohacení dat.



Obrázek 2.1: Schéma datové architektury, vlastní tvorba v softwaru Draw.io

Po zpracování dat v rámci transformačních vrstev se data přelévají do datamartů. Datamarty jsou zde rozděleny podle toho, jak je s nimi dále nakládáno. Jedná se zde primárně o datamarty *Delivery*, *Dokumentace* či *Sample data*. *Delivery* je právě datamart, který se poté nahrává do platformy GoodData. Do tohoto datamartu budou přidávána i data ze softwaru Freshdesk. V datamartu *Sample data* se nacházejí randomizovaná data z interního projektu. Tato data jsou identická data, která jsou do datamartu *Delivery*, ale dochází zde ještě k mezikroku při němž se randomizují a nahrají do datamartu *Sample Data*. Tato data slouží ke zkušebním a prezentačním účelům. Obecně se tedy dá říci, že se jedná o demo verzi BI řešení. Datamart *Dokumentace* je datamart, ze kterého se data nevizualizují, ale používají se jako podklad pro tvorbu automatizované dokumentace projektu.

2.4 Požadavky uživatelů na rozšíření

Hlavním požadavkem ze strany BizTreatu je nahradit základní analytické řešení, které poskytuje aplikace pro technickou podporu, konkrétně se jedná o software Freshdesk. V rámci jejich služeb je jisté analytické řešení poskytováno jako zabudované rozšíření pro Freshdesk. Toto řešení není ale bohužel pro firmu v aktuálním stavu dostačující, aby mohla efektivně řídit Support oddělení, které s tímto softwarem denně pracuje. Jak bylo zmíněno, tato služba nepřináší pro firmu dostačující informace, proto je jedním z hlavních požadavků vytvořit interní BI řešení pro data z této platformy. Data nejsou často dostatečně detailní. Často jsou také k vidění pouze v předem vytvořených vizualizacích, které se nedají příliš měnit. Data mají sloužit pro interní využití tak, aby se lépe sledoval a řídil počet odbavených požadavků, systém služeb pro osoby, které v tomto oddělení pracují či doba odbavení požadavků (incidentů).

Při interní analýze těchto dat a vytvoření vlastní datové pipeline, si firma může tyto data více přizpůsobit vlastním potřebám, rozšířit nabízené analýzy o další pohledy, grafy a KPI. Tedy přinést ještě větší výpovědní hodnotu pro konkrétní firmu než původně nabízené obecně pojaté řešení. Zároveň je žádoucí, aby všechna data byla na jednom místě, tedy v softwaru GoodData a aby nebylo třeba pro datové konzumenty přepínat mezi více softwary. Dalším důvodem je tedy i sjednocení několika míst, kde se data vyskytují, do jednoho. To by mělo uživatelům nejen usnadnit přístup k datům z časového hlediska, ale také vytvořit více uživatelsky přívětivé grafické znázornění, které je v souladu se zbytkem firemních dashboardů.

Dalším požadavkem je větší přístupnost k datům. Přímý přístup k datům je samozřejmě podmíněný účtem v softwaru Freshdesk, tento účet zároveň ale spoustu osob nepotřebuje a přeposílání dat těmto osobám je časově nevýhodné, nepraktické a zároveň je nutné při nahlížení do dat mít u sebe osobu, která je vlastníkem účtu či přístupy k jednomu z účtů. Lepší přístupnost se týká manažerů, kteří nejsou součástí Support oddělení, ale rádi by měli větší přehled o jeho fungování. Také se tento problém týká jednotlivých členů Support oddělení, kteří pokud nemají admin práva, nemají ani práva na přístup k datům. Toto také není vždy žádoucí, protože není možné udělovat admin práva všem osobám, které chtějí data vidět a pracovat s nimi. Z požadavků budoucích uživatelů tedy plyne, že je při tvorbě BI řešení žádoucí:

- poskytnout i detailní data,
- přidat větší škálu vizualizací,
- sjednotit místa výskytu datových vizualizací,
- poskytnout přístup k datům většímu množství osob.

3 Návrh rozšíření stávajícího řešení

Tato část diplomové práce se bude zabývat návrhem řešení. Nejprve zde budou rozebrána vstupní data, která jsou novým vstupem do řešení. Zde bude primárně zkoumáno, jaká data byla poskytnuta, a které jejich části jsou použitelné pro tvorbu BI řešení. Data se zde budou rozklíčovávat ještě před zahájením jakéhokoli návrhu. Tento bod je při tvorbě BI řešení stěžejní, protože by měl minimalizovat situace, kdy dojde k návrhu takového řešení, které není proveditelné s dostupnými daty. Po zmapování těchto dat se bude následující kapitola věnovat návrhu připojení těchto dat ke stávajícímu řešení. Dojde zde primárně k návrhu rozšíření dat tak, aby se udrželo původně vytvořené řešení a zároveň došlo k plynulému navázání dat nových. Poslední částí této kapitoly bude návrh vizualizací a jejich umístění na dashboard. V rámci návrhu se budou objevovat nejenom postupy, kterých by se měl dashboard držet, ale také se zde budou navrhovat informace, které by měly výsledné vizualizace obsahovat. Všechny tyto části je vhodné provést před zahájením práce na samotném rozšíření BI řešení, aby bylo možné ještě konzultovat jednotlivé části se zadavateli řešení a doladit případné detaily, které by bylo potřeba změnit.

3.1 Prozkoumání vstupních dat ze softwaru Freshdesk

Vstupní data ze softwaru Freshdesk byla poskytnuta firmou BizzTreat. V této kapitole bude rozebráno, v jakém stavu data jsou. Zároveň se zde budou nacházet informace o tabulkách, které byly poskytnuty a jejich obsah. Na základě těchto dat bude navržený dashboard pro potřeby firmy.

Poskytovaná data se nachází v raw stavu, tedy nebyla ještě nijak zpracovávána. Při zkoumání dat autorka narazila na typické problémy, které se takovýchto dat týkají. Mezi tyto problémy patřily například: neošetřené NULL hodnoty, mezery před a za textem či nesjednocené formáty datumových typů. Všechny tyto problémy budou muset být odstraněny v čistící datové vrstvě při zpracování dat.

Data se nacházejí celkem v osmi tabulkách. První podstatnou tabulkou je tabulka Agents, která obsahuje data o pracovnících, kteří v Support týmu pracují. Tato tabulka nemá velké množství záznamů, protože pracovníků není velké množství. Také se zde pro každého z nich nachází v jednom řádku všechny informace o nich. Nejdůležitější zde bude sloupec ID, kde se nachází unikátní primární klíč agenta a zároveň sloupec Contact, který obsahuje velké množství informací o agentech. Tento sloupec bude potřeba ještě rozdělit do několika dalších

sloupců, protože aktuálně obsahuje jeden textový řetězec, ve kterém jsou všechny informace, což je z pohledu datové analýzy nepřijatelné. Z tohoto sloupce bude nutné ještě v rámci transformací dostat data jako jméno agenta či jeho emailovou adresu. Další sloupce jako zahájení jeho činnosti či poslední aktivita zatím považuje autorka za data, která nebudou potřeba zpracovávat. Zůstanou tedy pouze v čisticí vrstvě pro případ, že by se do budoucna o tyto data řešení rozšiřovalo.

Druhou velmi důležitou a datově nejnáročnější tabulkou, je tabulka Tickets. V této tabulce jsou informace o jednotlivých požadavcích, které jsou vygenerované buď automaticky, díky nastaveným upozorněním v případech, kdy dojde k technickým problémům na některém z projektů, nebo tyto požadavky nahlásí sám klient. Tyto incidenty agenti denně prověřují. Tyto problémy či požadavky budou dále v textu uváděny jako incidenty. Protože těchto incidentů je velké množství i tato tabulka je velikostně největší. Aktuálně obsahuje data o více než pěti tisících požadavků, což odpovídá velikosti 76 MB. Tato tabulka bude hlavním datovým zdrojem pro toto řešení. Obsahuje také potřebné cizí klíče na ostatní tabulky, takže bude hlavní tabulkou, nad kterou se budou napočítávat metriky. Zároveň i z logiky věci, je to právě počet odbavených incidentů a data o jejich odbavení, jako primární téma tohoto řešení.

Poslední poměrně podstatnou tabulkou je tabulka Companies, kde se nachází data o firmách. Pro tyto firmy jsou nastaveny upozornění a jejich datová řešení jsou monitorována a technicky podporována právě pomocí softwaru Freshdesk. Další tabulky již zde nebudou detailně zkoumány, protože budou sloužit už jenom pro detailnější dobarvení kontextu pro některou z výše zmíněných tří tabulek. Jedná se například o detailnější informace o agentech, aktivní hodiny, v rámci kterých jsou incidenty odbavovány, či typ technické podpory, kterou firmy požadují.

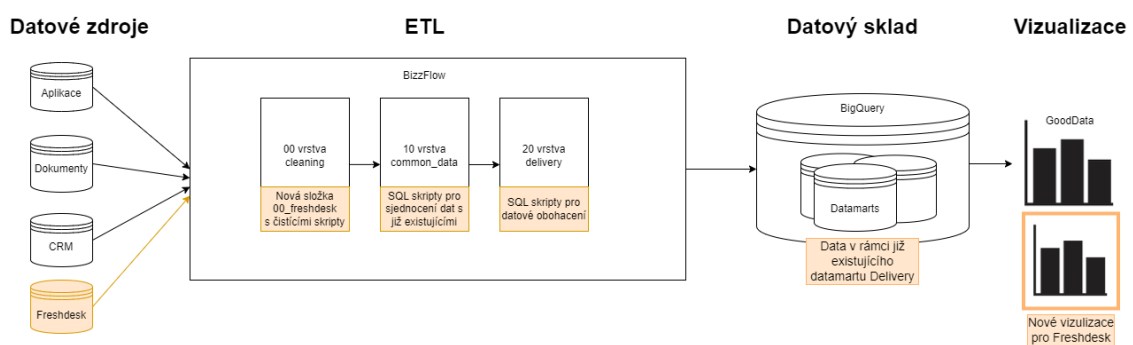
3.2 Návrh připojení datového toku k aktuálnímu řešení

Tato kapitola se bude zabývat návrhem na rozšíření aktuálně využívaného řešení. Celý zde sepsaný proces je pro lepší pochopení na konci kapitoly graficky znázorněn schématem. Prvním krokem v rámci rozšíření je přidání datového zdroje Freshdesk do řešení. Pro tuto aplikaci firma poskytla v rámci interně vyvíjeného ETL komponentu Freshdesk extraktoru. Tento extraktor data přes aplikační programové rozhraní (API) dostane do ETL BizzFlow, kde jsou data k dispozici pro další zpracování. Tato data jsou v surové podobě tak, jak je poskytuje přímo platforma Freshdesk a nejsou tedy ve stavu vhodném k využití v rámci analytického řešení.

Při transformaci je navrženo nejprve přidat novou čisticí složku do čisticí 00 vrstvy. Tato vrstva se bude jmenovat 00_freshdesk a bude sloužit jako místo pro uchování SQL čisticích skriptů, také z této složky bude při každém běhu ETL tyto skripty spouštět, jak tomu dochází i u ostatních skriptů. Další částí bude přidání SQL transformací do 10 vrstvy. K tomu dojde v případě, že budou existovat

dimenze, které by bylo vhodné zpracovat ještě, než se začnou data obohacovat. V rámci této vrstvy mohou být data pouze protažena nebo transformována. Pokud poskytnutá data nebudou obsahovat takové dimenze, které by bylo potřeba připravit pro sjednocení, dojde pouze k protažení dat beze změny touto vrstvou. Poslední částí v rámci transformačního procesu je tvorba SQL skriptů v rámci 20 vrstvy, která data obohatí a připraví k nahrání do datového skladu.

Data nahraná do datového skladu se přidávají do již existujícího datamartu Delivery, který obsahuje data týkající se dodávání služeb. Podle názoru autorky tato nově přidaná data mají své místo právě v tomto datamartu, protože se jedná o ucelený celek zabývající se poskytováním jak služeb na projektech, tak služeb týkajících se technické podpory. Data se z datamartu nahrají přímo do vizualizačního nástroje GoodData. Zde pro ně budou vytvořeny vhodné vizualizace, o kterých bude pojednávat následující kapitola.



Obrázek 3.1: Návrh rozšíření řešení o data ze softwaru Freshdesk, vlastní tvorba v softwaru Draw.io

3.3 Návrh dashboardu

Po prozkoumání dat a nalezení tří primárních tabulek, ze kterých budou vizualizace primárně vycházet, byly navrženy dva dashboardy. Na základě odborné literatury, která doporučuje držet jeden logický celek v rámci jednoho dashboardu a zároveň tvořit dashboardy co nejjednodušší, se autorka rozhodla raději vizualizace rozdělit na dva celky. První celek se bude zabývat vizualizacemi z pohledu klientů a druhý celek bude řešit vizualizace z pohledu pracovníků technické podpory, tedy agentů.

Na dashboardu určeném pro analýzu práce těchto agentů je navrženo rozložení do dvou sekcí pro lepší orientaci. První sekce, kterou dashboard bude obsahovat v horní části bude sekce s klíčovými výkonnostními indikátory (KPI). Tato sekce bude obsahovat vizualizace, které budou v co možná nejmenším detailu. Tato sekce by měla obsahovat následující informace jako: počet odbavených incidentů za měsíc v porovnání s minulým měsícem, počet incidentů, které byly agentům přiřazeny za tento měsíc v porovnání s minulým měsícem, doba, za kterou agent stihl odpovědět na tikety, rychlost, v rámci které stihl agent tikety vyřešit. Druhá sekce tohoto dashboardu by podle návrhu měla obsahovat rozpad těchto dat

do většího detailu a měly by se zde začít objevovat grafická znázornění. Například by se zde měl objevit sloupcový graf, který bude ukazovat rozpad těchto incidentů podle typu jejich hlavních příčin.

Druhý dashboard bude také pracovat s daty ohledně incidentů, ale bude na ně nahlíženo z jiné perspektivy, a to z perspektivy klientů. Tento dashboard navrhuji využívat pro evidenci incidentů pro jednotlivé klienty jak interně, tak zároveň jako podklad pro pravidelné schůzky s klienty, které se konají za účelem řešení stavu technické podpory na BI projektech. Díky těmto přehledným vizualizacím by měly mít obě strany lepší přehled o fungování této podpory. Autorka zde tedy opět navrhuje umístit dvě sekce jako pro předchozí dashboard. První část by obsahovala následující KPI:

- počet odbavených incidentů za měsíc v porovnání s minulým měsícem,
- počet dní v měsíci, kdy na projektu nastal incident s porovnáním oproti předešlému měsíci,
- průměrná doba do první reakce agenta na incident.

Druhá část by obsahovala graficky znázorněný počet incidentů přes jednotlivé kategorie příčin, tedy vhodným grafem zde bude opět sloupcový graf. Tento graf by měl být seřazený od nejvíce častých po méně časté typy incidentů. Zároveň autorka navrhuje, aby z tohoto grafu vedl drill-down do tabulky, která by obsahovala detailní informace o jednotlivých incidentech.

4 Zpracování dat a implementace návrhu

V následující kapitole diplomové práce bude rozebírán postup, který byl využit při implementaci návrhu pro rozšíření BI řešení o nová data. První část bude popisovat, jak bylo rozšířeno řešení z pohledu datových vrstev projektu, budou zde uváděny ukázky SQL kódu a popisovány části transformací jako je čistící či transformační datová vrstva. Po datovém zpracování bude následovat kapitola, která bude rozebírat tvorbu logického datového modelu (LDM). Tato kapitola se bude věnovat vazbám mezi tabulkami a konkrétně bude uveden příklad datového modelování v softwaru GoodData. Následně budou z nahraných dat vytvořeny metriky, tedy data budou zpracovávána jistými funkcemi tak, aby bylo možné z nich snadno v softwaru vytvářet vizualizace. Poslední část této kapitoly se bude věnovat právě tvorbě dashboardu. Budou zde popsány jednotlivé vizualizace, které se na dashboardu objevují a bude uvedeno jejich zakomponování do dashboardů.

4.1 Tvorba datových vrstev

V rámci této kapitoly budou probírány jednotlivé datové vrstvy a jejich obsah. Bude zde sepsáno, které kroky bylo nutné udělat pro zpracování dat do formy, kterou je vhodné nahrát do vizualizačního softwaru GoodData. Bude zde popisován postup při tvorbě jednotlivých datových vrstev, tedy konkrétně SQL skriptů, které tyto vrstvy obsahují. Tato tvorba bude zahrnovat čištění a zpracování dat. Při obou procesech dochází k obohacení surových dat o nové informace. Kapitola bude ještě rozčleněna na transformace týkající se pouze čištění dat a transformace týkající se datového zpracování z pohledu byznysové logiky v dotazech a rozšíření dat o nové sloupce obsahující nově získané informace.

4.1.1 Čistící vrstva

Datová kvalita je stěžejním předpokladem dobrého BI řešení. Právě proto existuje proces čištění dat, který je prvním krokem při vstupu dat do datových transformací. Jak bylo zmíněno v předchozích kapitolách tato vrstva se v rámci architektury projektu značí jako 00 vrstva. Tato 00 vrstva obsahuje jednotlivé složky pro každý datový zdroj oddělené. Tyto složky obsahují čistící skripty pro jednotlivé tabulky z daného zdroje. Pro nově připojený datový zdroj Freshdesk byla vytvořena složka 00_freshdesk_support.

V této složce se nachází celkem šestnáct skriptů, dva skripty pro každou vstupní tabulku. První z dvojice skriptů obsahuje základní datové očistění, které ošetří hodnoty typu NULL, nadefinuje pro každý sloupec datový typ a data zároveň pomocí funkce TRIM zbaví přebytečných mezer na začátku a na konci řetězců. Tyto přebytečné mezery mohou vzniknout například manuálním zadáváním vstupu od uživatele a jsou velkým problémem jak u cizích a primárních klíčů, tak u atributů. U primárních a cizích klíčů mohou nežádoucí mezery způsobit, že nedojde k propojení hodnot, které k sobě patří právě kvůli nadbytečným mezerám. U atributů zase mohou mezery způsobit rozdělení do redundantních kategorií.

NULL hodnoty jsou nahrazeny podle datového typu hodnotami –empty–, 0.00 či 1970-01-01. Toto nahrazení je důležité z několika důvodů. Na NULL hodnoty se v SQL nedotazuje nejlépe, často vypadávají z různých operací, i přestože je to nežádoucí, mohou vypadnout při tvorbě vizualizací a zároveň není jasně viditelné, že k tomu došlo. Díky tomuhle nahrazení za předem nadefinované hodnoty se lépe rozpoznávají. Je určitě důležité tyto hodnoty volit tak, aby nezkreslovali výpočty a být si vědom jejich existence v průběhu práce s daty. Příkladem může být výpočet aritmetického průměru stáří populace, do kterého se poté nemůžou hodnoty pro 1970-01-01 započítávat. Pokud by ale v datech byla opravdu osoba s tímto datem narození, z dat by tento údaj vypadl. V rámci tohoto řešení, z logiky věci dat ohledně incidentů za posledních několik jednotek let to nebude závažný problém, ale autorka přesto na toto úskalí chtěla upozornit.

Příkladem skriptu, který tuto problematiku vyřeší je například základní čistící skript pro tabulku Agents, který je k nahlédnutí níže. Byl značně zjednodušen, pro každý datový typ je zde uvedena právě jedna čistící podmínka. První z podmínek je pro číselný datový typ, druhá pro datum a poslední pro textový datový typ.

```
CREATE OR REPLACE TABLE `cln_agents` AS
SELECT
  CASE WHEN `id` IS NULL OR TRIM(`id`) = ''
    THEN 0 ELSE CAST(TRIM(`id`) as Int64)
  END as `id`,
  CASE WHEN `created_at` IS NULL OR TRIM(`created_at`) = ''
    THEN CAST('1970-01-01 00:00:00' as TIMESTAMP)
    ELSE CAST(TRIM(`created_at`) as TIMESTAMP)
  END as `created_at`,
  CASE WHEN `type` IS NULL OR TRIM(`type`) = ''
    THEN '--empty--' ELSE TRIM(`type`)
  END as `type`
FROM `in_agents`
;
```

Základně vyčištěny byly všechny tabulky. Tři nejdůležitější tabulky (Tickets, Agents, Companies), ve kterých jsou nejvíce relevantní data pro toto řešení, byly vyčištěny ještě detailněji. K pokročilemu čištění dat docházelo v druhém skriptu z výše zmíněné dvojice skriptů pro každou tabulku. Zbytek tabulek byl v rámci druhého skriptu pouze protažen do výstupních tabulek, tedy těch s prefixem out_.

Při pokročilém čištění dat došlo k několika dalším krokům. Sloupce byly prozkoumány jednotlivě a byly odstraněny ty, které z nějakého důvodu byly nadbytečné. Jednou z možností, jak určit nadbytečné sloupce je to, že obsahovaly pouze prázdné hodnoty. Jednalo se tedy o sloupce, které v aplikaci uživatelé nikdy nevyplňují. K rychlejší orientaci v datech byl využit vždy následující skript.

```
SELECT DISTINCT
    `nazev_sloupce`
FROM `out_tickets`
;
```

Tento skript vždy našel unikátně vyskytující se hodnoty ve sloupci, pokud se zde objevovala pouze prázdná hodnota či hodnoty, které o ničem nevypovídaly (byly vyplněny například pouze jednorázově a nejednalo se o dostatečné množství dat, které by se dlouhodobě sbíralo). V případě nejistoty ohledně relevance některého ze sloupců, zadavatel také ochotně určil, jestli se jedná o sloupec důležitý či pouze pokusný.

Dále zde byly identifikovány sloupce, které obsahovaly ve velkém množství informací totožné hodnoty. Jedním z takových příkladů byly sloupce `due_by` a `fr_due_by` v tabulce `Tickets`. Po detailním prozkoumání a komunikaci se zadavatelem bylo zjištěno, že jeden ze sloupců obsahuje hodnotu automaticky generovanou a druhý hodnotu upravenou agentem. Pro naši analýzu je relevantnější právě druhá hodnota, proto byly v takovýchto případech prakticky totožné sloupce odstraněny a vybrány ty, které byly již upraveny agenty. Poslední úpravou týkající se úplného odstranění sloupců byla normalizace tabulek. Docházelo zde k případům, kdy například tabulka `Tickets` obsahovala jak firemní ID, tak firemní název, protože údaje o firmě se nachází z tabulce `Companies`, je nadbytečné tuto informaci mít i v tabulce `Tickets`, tyto sloupce byly také odstraněny. Byla zde ponechána pouze vazba na tabulku `Companies`, tedy cizí klíč ve sloupci `company_id`.

Poslední částí čištění dat byla identifikace klíčů a rozdělení tabulek podle typu sloupců. Tento krok je podstatný hlavně u tabulek s větším množstvím sloupců, ale hodí se ho využívat vždy pro rychlejší orientaci. Sloupce se v tabulce seřadí do následujících kategorií:

- primární klíč (či všechny jeho části v případě, že se jedná o více než jeden sloupec),
- cizí klíče,
- datумы,
- atributy,
- fakta.

Každá kategorie se označí komentářem na jejím začátku. Toto velmi usnadní orientaci ve sloupcích. V případě, že zde tato kategorie není, autorka ponechala

toto pomyslné pole prázdné pro případné rozšíření do budoucna. Zároveň je v rámci tohoto uspořádání určeno, jestli tabulka obsahuje nějaké cizí klíče a jestli má nějaký primární klíč, nebo jestli bude nutné tento primární klíč vytvořit. Příklad zjednodušeného kódu pro rozřazení do kategorií:

```
CREATE OR REPLACE TABLE `out_companies` AS
SELECT
--PK
  `id`,
--FK
--dates
  `created_at`,
  `updated_at`,
--attributes
  `name`,
  `delivery_team`,
  `description`
--fact
FROM `cln_companies`
;
```

Poslední záležitostí, která byla v rámci čistících skriptů nutná vyřešit, bylo ošetření několika sloupců napříč tabulkami, které neobsahovaly jednotlivé hodnoty, ale data v .json formátu. Tento textový formát je bohužel nemožné analyzovat, protože obsahuje nejenom data uzavřená v závorkách a s dalšími specifickými znaky jako jsou dvojtečky a apostrofy, ale zároveň jsou tato data v několika kategoriích. Níže je příklad takového .json formátu, který se objevoval v jedné buňce. Toto je příklad textového řetězce ze sloupce Stats, tedy informace o stavu tiketu/incidentu v jedné buňce.

```
{
'agent_responded_at': '2023-12-08T13:28:36Z',
'requester_responded_at': None,
'first_responded_at': '2023-12-08T13:28:36Z',
'status_updated_at': '2023-12-11T06:17:56Z',
'reopened_at': None,
'resolved_at': '2023-12-11T06:17:56Z',
'closed_at': '2023-12-11T06:17:56Z',
'pending_since': None
}
```

Jak je z tohoto příkladu zřejmé, nelze zde bez očištění dat nic analyzovat a je nutné tento textový řetězec rozdělit do několika dalších sloupců. Protože ale není potřeba využít všechna data, byly z tohoto textového řetězce vytvořeny dva nové sloupce: `first_responded_at` a `closed_at`. Konkrétně zde k tomuto účelu byla využita funkce `REGEXP_EXTRACT`, která hledá v daném řetězci klíčové slovo, poté zapisuje znaky, dokud se nedostane na nejbližší určený znak, zde se jedná o apostrof. V níže uvedeném příkladu se jedná o extrahování `first_response`, což je

velmi důležitý údaj, na jehož základě se určuje, jestli agent zvládá reagovat ve smluvně dohodnutém čase. Druhý nově vytvořený sloupec z příkladu poskytuje čas uzavření incidentu.

```
COALESCE(CAST(
  REGEXP_EXTRACT(stats,
    r"first_responded_at '\s*:\s*'([\^']*)"
  AS TIMESTAMP),
TIMESTAMP '1970-01-01 00:00:00') as `first_response`,
```

```
COALESCE(CAST(
  REGEXP_EXTRACT(stats,
    r"resolved_at '\s*:\s*'([\^']*)"
  AS TIMESTAMP),
TIMESTAMP '1970-01-01 00:00:00') as `closed_at`,
```

4.1.2 Transformační vrstva

Po očištění dat je dalším logickým krokem v BI řešení datová transformace. V této fázi se data obohacují o nové sloupce, které se napočítávají v již existující tabulce a také vznikají kombinací mezi jednotlivými tabulkami. Pokud by data obsahovala všechny informace, které jsou pro byznysové účely potřeba, nebylo by nutné žádná data přidávat. Toto se ale v praxi neděje, a proto je tato vrstva často velmi obsáhlá. Právě zde také v průběhu času kód nejvíce nabývá na obsahu, protože se zde s každým dalším požadavkem od byznysového zadavatele dopočítávají další a další datové pohledy.

V rámci tohoto řešení bylo v transformační vrstvě, té označené prefixem 20_, nutné vymyslet fakta obsahující časové rozdíly mezi dvěma časovými otisky. Existovaly zde sloupce obsahující informaci o vytvoření tiketu a sloupce obsahující data o jeho uzavření a tzv. First response. To je informace o přesném čase, kdy agent napíše k tiketu zprávu s určitými informacemi, například proč k problému došlo, čeho se týká, do kdy bude přibližně vyřešen. Většina těchto tiketů chodí automaticky, nejedná se tedy o tikety, které by zadal člověk. Na druhou stranu tyto informace už si klient čte a dostane v rámci nich ucelenější obraz o tom, která část jeho BI řešení je nefunkční, proč a kdy může očekávat aktualizovaná data. Problém, který bylo nutné vyřešit v rámci transformací, byl ten, že čas vytvoření tiketu se nerovná času, od kdy je agentům počítán čas odpovědi. Jak již bylo řečeno, tiket se automaticky vytvoří ve chvíli, kdy k problému opravdu dojde, což často bývá v brzkých ranních hodinách. Odpočet by měl ale začínat, až ve chvíli, kdy pro agenta začíná denní služba, tedy konkrétně v devět hodin dopoledne. Proto bylo potřeba najít způsob, jak se k tomuto času dostat.

V datech v tabulce Tickets byl nalezen sloupec 'due_by_first', který obsahoval informaci o čase, do kterého by měl agent zareagovat. Pokud by se od této časové značky odečetl počet hodin podle smluvních podmínek, jednalo by se o vždy o čas, kdy incident vznikl. Po investigaci v rámci poskytnutých tabulek byla tato

informace nalezena v tabuce SLA Policies. Zároveň zde existovala datová vazba na firmu, schovaná ve sloupci Applicable_to, který držel informace o tom, na které firmy se tyto smluvní podmínky aplikují. Tento sloupec posloužil jako cizí klíč tabulky Companies. Protože i v tabulce Tickets tento cizí klíč do tabulky Companies existoval, byla tato vazba využita pro spojení těchto tabulek pomocí LEFT JOIN. Níže uvedeno konkrétní použití:

```
LEFT JOIN `in_sla_policies` sla
  ON t.`company_id` =
     CAST(sla.`applicable_to` as INT64)
```

Po nalezení vazby mezi těmito tabulkami byl pomocí několika funkcí vytvořen nový sloupec. Vytvoření tohoto sloupce vyžadovalo větší dávku testování a bylo zde nutné využít několika podmínek. Tyto podmínky byly ošetřeny za pomoci SQL funkce na tvorbu podmínky: CASE WHEN. Bylo totiž nutné vyřešit situace, které se týkaly rozložení vytvoření tiketu v rámci jednotlivých denních hodin i dní v týdnu. Bylo potřeba ošetřit případy, kdy byl tiket vytvořený před zahájením pracovních hodin, v takových případech se jeho zahájení posunulo až na devět hodin dopoledne daného dne. V případě vytvoření tiketu v rámci dne zůstal čas stejný. V případě, že byl tiket vytvořen po ukončení pracovních hodin agentů jeho zahájení se posunulo na následující den také na devátou hodinu dopolední. Zároveň v případech, kdy se jednalo o víkend bylo zahájení odpočtu u tiketu posunuto vždy na pondělí.

Takto vytvořený nový sloupec obsahuje informaci o začátku incidentu z pohledu pracovní doby agenta. Tato informace je důležitá pro byznysový pohled, aby noční tikety nezkreslovali statistiky a nepůsobily umělé prodloužení času, ve kterém agent na tiket stihne odpovědět. Aktuálně neexistuje klient, který by potřeboval technickou podporu i během noci či víkendů. Pokud by k tomuto případu došlo, bylo by potřeba tato data vyčlenit a neaplikovat na ně výše zmíněné podmínky. Aktuálně ale takové služby firma neposkytuje. Po vytvoření sloupce 'incident_time_start' v přípravné tabulce s prefixem 'tmp_tickets', je tato pomocná tabulka využita jako zdroj výstupní tabulky, kde dochází k výpočtu vteřin, které jsou mezi nově vypočítaným zahájením tiketu a první odpovědí na tiket. Pomocí funkce TIMESTAMP_DIFF se odečtou hodnoty mezi dvěma časovými značkami. Na základě jejich rozdílu ve formě počtu sekund, je vytvořen nový sloupec obsahující čas mezi těmito událostmi. Tyto hodnoty jsou spočítány pro první odpověď, jak je uvedeno v SQL níže, tak pro celkový čas vyřešení tiketu.

```
TIMESTAMP_DIFF(`first_response`,
  `incident_time_start`, SECOND)
as `first_response_seconds_diff`,
```

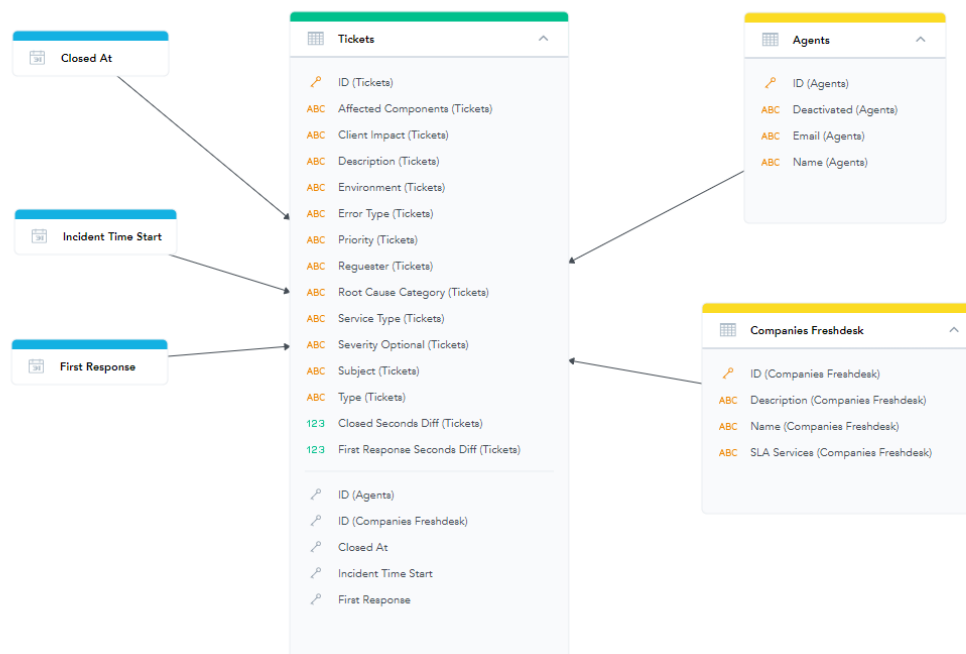
Na závěr jsou obohacená data pomocí ETL nahraná do datamartu. V datamartu jsou uložena a pomocí zabudované ETL komponenty na nahrání dat do jednotlivých vizualizačních nástrojů nahraje data do vizualizačního nástroje GoodData. Toto nahrávání je nastaveno tak, aby probíhalo automaticky každý den v ranních hodinách. Datoví konzumenti tedy mohou již ráno vidět aktuální data z předchozího dne, což by jim mělo usnadnit jejich každodenní práci.

4.2 Tvorba LDM

Nyní jsou data připravená tak, aby se z nich daly vytvářet vizualizace, není nad nimi ale stále ještě postavený logický datový model. V rámci čistící vrstvy byly sice vazby mezi tabulkami odhaleny, ale byly v transformacích používány pouze pro tvorbu dotazovacích skriptů. Teď se nadefinují vazby v rámci vizualizačního softwaru tak, aby s nimi tento software mohl pracovat a nabízet pro vizualizace jenom ty sloupce a tabulky, které spolu souvisí.

Do datamartu byly nahrány tři nejdůležitější tabulky Tickets, Agents a Freshdesk Companies. Bylo zde postaveno hvězdicové schéma. V tomto schématu je primární tabulkou tabulka Tickets. Všechny ostatní tabulky mají vůči této tabulce vazbu 1:N, tedy pro jeden výskyt v ostatních tabulkách existuje několik výskytů v tabulce Tickets. Například jeden agent má více tiketů, ale tiket je přiřazen právě jednomu agentu. Zároveň každá firma může mít více tiketů, ale jeden tiket je přiřazen právě jedné firmě. Takto jsou data koncipována z byznysového hlediska.

Tabulka Tickets je faktická tabulka, proto je v modelu označena zeleně. Tato tabulka sice obsahuje i atributy týkající se právě tiketů, ale pokud obsahuje alespoň jeden fakt, je softwarem vyhodnocena jako faktová tabulka. Žlutě označené tabulky jsou dimenze, tedy tabulky, které obsahují pouze atributy. Modrou barvou jsou zde pak označeny datумы, ty jsou v datamartu v tabulce Tickets, formálně tedy rozdělené do jednotlivých tabulek nejsou. Software však pro přehlednost a specifickou práci s těmito hodnotami datумы takto odděluje. Celý LDM je ukázán na obrázku níže.



Obrázek 4.1: Logický datový model vytvořený v softwaru GoodData

Každý sloupec je v tomto softwaru vytvořen a je mu přidělen datový typ. Sloupce se zde rozdělují podle datových typů na text, číslo a datum. Dalším dělením podle typu sloupce, již ne datového typu, ale spíše logického typu, je dělení na atributy, fakta, datумы, primární klíče a cizí klíče. Primární klíče jsou zde označeny oranžovou ikonkou klíče a v tabulkách se přiřazují na začátek tabulky. Dále následují abecedně seřazené atributy, které jsou značeny také oranžovou značkou. Ta tentokrát zastupuje první tři písmena v abecedě, tedy značkou ABC. Předposledním typem sloupce v modelu jsou fakta, která jsou značena zelenou značkou prvních tří číslic tedy 123. Poté se v tabulce objevuje sotva viditelná šedá linka, která odděluje cizí klíče od zbytku tabulky. Cizí klíče zde mají značku šedého klíče a mají vazbu na primární klíče ostatních tabulek v uvedeném modelu a také na datumové dimenze.

4.3 Tvorba metrik v softwaru GoodData

Po vytvoření LDM a nahrání dat do softwaru GoodData, jsou data připravena tak, aby se z nich daly vytvořit metriky. Metriky v GoodData jsou objekty, které obsahují jednoduché agregace faktů, matematické operace, lze zde formátovat výstupy a obecně se jedná o první krok pro vytvoření dashboardu či jednotlivých reportů. V ideálním případě se nejprve všechny metriky vytvoří právě pomocí zabudovaného dotazovacího jazyka MAQL, který tato platforma poskytuje. Z těchto vytvořených metrik jsou poté poskládány jednotlivé vizualizace.

GoodData nabízí i možnost tvořit vizualizace bez prvotního vytvoření metrik, ale metriky zde mají více možností a jsou pro dlouhodobou údržbu projektu lepší volbou. Důvodů je hned několik. Prvním z nich je kaskádová úprava metrik, tedy v případě, že je potřeba nějakou část metriky upravit, dojde k této úpravě pouze na jednom místě, nikoli na všech, k čemuž by muselo při použití prostých faktů dojít. Za druhé fakty neposkytují takové množství funkcí, kterými se dají v analytické části platformy upravit. Metriky nabízí možnosti podmínek, práce s kombinací již vzniklých metrik a obecně širší škálu úprav. Zároveň se metriky mnohem lépe spravují, a to protože se dá v softwaru dohledat, na kterých všech místech jsou použity, kdo je vytvořil či kdy je kdo upravil.

V rámci dashboardů, které se pro tento projekt zhotovily, bylo potřeba vytvořit sedm nových metrik, které se na dashboardech objevují. Jedná se o metriky, které pracují v první řadě s počtem tiketů a jejich kategoriemi a možnými filtracemi. Druhou pomyslnou skupinou metrik jsou metriky, které pracují s časem, a to buď časem odbavení tiketu od jeho vytvoření po jeho vyřešení nebo časem od jeho vytvoření po první odpověď od agenta. První odpověď je z byznysového pohledu důležitá a její hodnoty se sledují, jedná se o čas, v němž jsou agenti průměrně schopni odpovědět klientovi. Odpověď obsahuje informace o vzniklém problému, řeší se zde čeho se problém týká, do kdy bude pravděpodobně vyřešen, jestli má dopad na jeho data či jaké další kroky budou následovat k jeho odstranění.

Níže uvedený příklad jedné z vytvořených metrik počítá počet tiketů na základě primárního ID. Díky tomu, že ID je unikátním identifikátorem, je možné

na jeho základě vypočítá i přesný počet těchto tiketů. MAQL se podobá SQL, ale nevyužívá JOIN. Spojování tabulek se řídí podle pravidel, které se nadefinují při stavbě LDM. Stejně jako u SQL je zde podmínka WHERE, která vybírá atribut obsahující informaci o dopadech na klienta a vybírá pouze takové tikety, které měly nějaký dopad na fungování klienta. Pro lepší orientaci v BI řešení byly všechny metriky týkající se dashboardů pro technickou podporu odděleny do složky s názvem Support, což je vidět na ukázce na pravé horní straně, kde se na platformě vyskytují informace o umístění vybrané metriky.

The screenshot displays the GoodData metric configuration interface. At the top, the breadcrumb 'Data | Metric' is visible. The main configuration area includes:

- Metric Name:** # Tickets Client Impact
- Description:** Click here to add description
- Tags:** Add Tags
- Actions:** Edit, Duplicate, Sharing & Permissions
- MAQL Query:** SELECT COUNT(ID (Tickets), ID (Tickets)) WHERE Client Impact (Tickets) = Yes
- Metric Format:** #,##0 (with an Edit button)
- Metric Comments:** (empty field)

On the right side, the **Metric Location** panel shows:

- Metric is placed in the Metrics Support folder.
- Change Folder button
- Delete button
- Warning: Metric will be removed from the list of predefined metrics. Existing reports will not be affected.
- Metric Statistics:** Last updated by You on Today.
- Review Affected Reports button

Obrázek 4.2: Ukázka prostředí pro tvorbu metriky v softwaru GoodData

4.4 Tvorba dashboardu

Dashboards byly vytvářeny na základě informací získaných z odborné literatury. Autorka se tedy držela pravidla, že obecné informace mají být na začátku dashboardu a níže se mají objevovat informace více detailní. Zároveň, aby se dashboard zbytečně neprotahoval a byl co nejkratší na délku a nejjednodušší, byly ke každé graficky zpracované vizualizaci přidány interakce spustitelné kliknutím, vedoucí na detailnější informace. Dashboards byly pro větší přehlednost rozděleny do sekcí a každá z nich byla popsána komentářem.

Dashboards byly vytvořeny dva. První dashboard byl sestaven z pohledu klientských technicky podporovaných projektů a pojmenován *Support / Clients*. Druhý dashboard byl vytvořen z pohledu agentů technické podpory, kteří v rámci tohoto oddělení pracují a byl pojmenován *Support / Agents*. Dashboards sdílí některé grafické vizualizace, protože co se týče základních metrik vychází z metrik stejných, jen se na ně dívají přes odlišné kategorie a filtry. Dashboards jsou rozděleny do tří sekcí, první a třetí sekce je u obou dashboardů totožná. Díky podobnostem mezi těmito dvěma dashboardy zde bude uveden konkrétní popis dashboardu *Support / Clients* a bude přidán jeho vyexportovaný obrázek, který je k nahlédnutí závěrem této kapitoly. Jedná se o obrázek 4.3 *Dashboard vytvořený v softwaru GoodData*. Níže poté budou uvedeny pouze rozdíly mezi dashboardy. Uvádět detailně popis obou z nich by bylo zbytečné.

Dashboard *Support / Clients* obsahuje tři sekce. První sekci tvoří klíčové metriky zobrazené jako jednoduché vizualizace s porovnáním. První čtyři KPI obsahují informace o množství tiketů. První vizualizace obsahuje počet tiketů v rámci časového období a jeho porovnání s obdobím předchozím. Bylo zde zvoleno měsíční období. Druhá vizualizace v řadě obsahuje informaci o tom, kolik z uvedených tiketů mělo nějaký dopad na klienta, tedy odpovídá na otázku, jestli bylo jeho BI řešení zasaženo incidentem natolik, že by omezilo jeho chod. Zároveň je zde toto číslo zobrazeno procentuálně, tedy o jakou část z celkového množství tiketů se jedná. Třetí a čtvrtá vizualizace v první řadě se věnuje rozdělení tiketů. Tikety se dělí do kategorií podle jejich typu, nejčastějším typem je incident, ke kterému došlo v rámci BI řešení, to zobrazuje třetí vizualizace. Čtvrtá vizualizace podává informaci o upozorněních. Tato upozornění mohou předejít vzniku incidentů a jsou to často automaticky nastavená upozornění, která se odešlou při nestandardním vývoji v rámci BI řešení. Příkladem takového upozornění může být zaplnění paměti disku na více než devadesát procent. V takové situaci přijde automaticky upozornění na oddělení technické podpory. Obě tyto vizualizace obsahují také porovnání s množstvím za minulé časové období.

V rámci druhého řádku první sekce se objevují dvě klíčové metriky. První z nich sleduje střední hodnotu rychlosti odpovědi na příchozí tiket. Jedná se zde o medián, který byl zvolen před aritmetickým průměrem kvůli potenciálním odlehklým hodnotám, které by mohly zkreslovat tyto statistiky. Metrika je naformátována tak, aby se číslo zobrazovalo v časovém formátu dobře čitelném pro člověka. Data za metrikami v této sekci vchází do vizualizačního nástroje ve vteřinách a je zde nastaveno formátování, které přepočítává tyto vteřiny na hodiny, minuty a vteřiny pro lepší čitelnost. Tato metrika zobrazuje čas, v rámci kterého je přibližně tiket zaregistrován, je provedena investigace a klientovi je odpovězeno. Druhá vizualizace obsahuje informaci o vyřešení tiketů. Tedy jak dlouho trvá, než je tiket úspěšně vyřešen a uzavřen.

Druhá sekce dashboardu se zabývá pohledy přes jednotlivé klienty a množství tiketů, které bylo vytvořeno za daný měsíc. Zároveň jsou zde tikety rozděleny do dvou kategorií na ty, které se podařilo vyřešit tak rychle, že neměly žádný dopad na klienta a ty, které dopad na klienta měly. Zeleně se zde zobrazují tikety bez dopadu na klienta, červeně ty s dopadem. Jsou zde k nalezení dva sloupcové grafy. První obsahuje množství tiketů přes jednotlivé klienty a druhý množství tiketů podle jejich kategorií. Názvy klientů jsou na níže uvedeném obrázku anonymizovány. V rámci druhého grafu, kde jsou vidět kategorie příčin, je možné z grafu vyčíst, že nejčastější příčinou výpadku BI řešení jsou chyby třetích stran. Tato kategorie obsahuje chyby serveru, výpadky databází, nefunkční krátkodobá připojení do aplikací a podobné chyby, které se netýkají ani poskytovatele BI, ani klienta. Naopak lze z grafu vyčíst, že za sledovaný měsíc bylo minimum tiketů, které se týkaly vývoje, a to buď na straně poskytovatele BI řešení, nebo na straně klientů, kteří si vývoj řeší sami. Oba grafy v této sekci poskytují drill-down na detail tiketů. Tedy tabulku obsahující informace o ID tiketů, jeho příčině, klientovi, kterému tiket náleží a stručnému popisu tiketů.

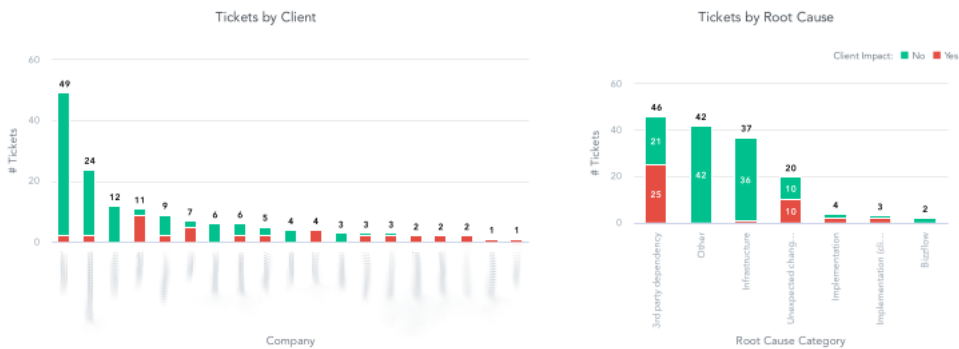
Poslední sekce dashboardu se opět věnuje časům odbavení tiketu a první odpovědi na tiket. Obě tyto informace vedle sebe vkládá do spojnicového grafu. Jsou zde rozřazeny přes jednotlivé týdny v měsíci a je zde opět pozorován medián těchto hodnot. Tato vizualizace také obsahuje drill-down na detailnější data. Dashboard obsahuje možnost filtrace na jakékoliv časové období, největší smysl zde samozřejmě dávají jednotlivé měsíce, ale dynamicky se dá pohybovat i v jiných časových periodách. Zároveň dashboard poskytuje i filtr s jednotlivými klienty. Je tedy možné, například při schůzce s konkrétním klientem, zobrazit pouze jeho data, a to vybráním jeho firmy ve filtru.

Tickets & Response Time



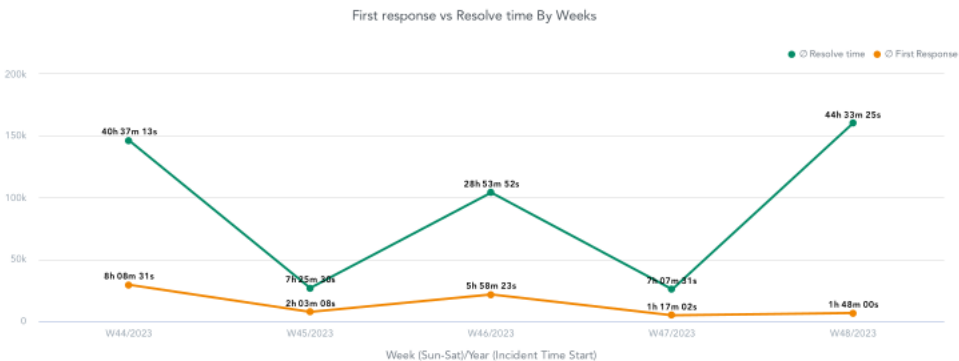
Tickets Detail

drill-downs available



Response Time Details

drill-downs available



Obrázek 4.3: Dashboard vytvořený v softwaru GoodData

Jak už bylo zmíněno druhý dashboard *Support / Agents* obsahuje stejné vizualizace, pouze druhá sekce je odlišná. V rámci té je umístěna pouze jedna vizualizace místo dvou a tou je horizontální sloupcový graf. Ten je rozdělen do sloupců podle jednotlivých agentů a opět je rozdělen na tikety podle dopadu na klienta. Kvůli citlivosti těchto dat zde daná vizualizace nebude uvedena na obrázku. Druhý dashboard obsahuje také odlišný filtr. Místo filtrace přes zákazníky se zde filtruje přes jednotlivé agenty. Zbytek vizualizací, filtrací i interakcí zůstává stejný pro oba dashboardy.

4.4.1 Self-Servise BI

Jak bylo popsáno v teoretické části této práce, trendem dnešní doby jsou BI řešení, která poskytují uživatelům větší možnost tvorby vlastních pohledů a vizualizací na základě dat, která byla zpracovaná pomocí ETL. U tohoto řešení tomu není jinak, data nahraná do platformy GoodData poskytují uživatelům možnost si vytvářet mimo poskytnutého dashboardu i vlastní typy jak jednotlivých vizualizací, tak celých vlastně navržených dashboardů. Jak vypadá prostředí v softwaru na tvorbu vlastních vizualizací, bude uvedeno níže v textu.

Mimo data použitá v uvedených dashboardech jsou do BI řešení nahraná i další data, která mohou uživatelé využít při tvorbě vlastních vizualizací. Díky těmto datům mohou dosahovat vizualizace různé granularity, tedy datového detailu. Uži

vatelé mohou sledovat data přes jednotlivé dimenze a atributy. Na uvedeném příkladu na následující stránce je možné vidět datovou granularitu tiketů a jejich filtrování. Při tvorbě této vizualizace byla zafiltrována pouze data týkající se tiketů, které vznikly kvůli chybě nebo problému s extrahováním dat. Dále byly vybrány čtyři příčiny díky filtraci v horní liště softwaru, jak je možné vidět na přiloženém obrázku níže. Data byla agregovaná do skupin a byla zobrazena jejich četnost. Takto vytvořená vizualizace se dá dále filtrovat či naopak rozšiřovat. Díky tomu mohou uživatelé vidět pouze data, která jsou pro ně relevantní a tuto relevanci si mohou určit sami díky SSBI.

Pokud by si uživatel položil nějakou konkrétní otázku například: Jaké bylo rozložení typu tiketů v rámci posledního čtvrtletí roku 2023? Mohl by si snadno vytvořit vizualizaci, která by toto rozložení zobrazila například pomocí prstencového grafu a odpověděla by na jeho otázku. Na níže uvedeném příkladu je zobrazena časová filtrace grafu pro čtvrté čtvrtletí roku 2023. Zároveň je zde filtr, který zobrazuje pouze tikety, kterým agenti přiřadili tuto informaci, tedy jejich hodnota pro typ tiketu byla vyplněna.

FILTERS ABC Affected Co...s (Tickets) Extract ▾ ABC Root Cause C...ry (Tickets) 3rd party depende... (4) ▾ DRAG ABC OR HERE

Affected Components (Tickets)	Root Cause Category (Tickets)	Type (Tickets)	Severity Optional (Tickets)	# Tickets	
Extract	3rd party dependency	Incident	1 - service is unavailable, not possible to log in, more modules are broken	4	
			2 - critical functionality is unavailable and there is no workaround	298	
			3 - service does not work as specified, however there is a workaround; application res...	58	
			4 - defects which does not affect service functionality (typos, documentation mistakes,...	20	
			Rollup (Total)	380	
Implementation		Incident	1 - service is unavailable, not possible to log in, more modules are broken	2	
			2 - critical functionality is unavailable and there is no workaround	27	
			3 - service does not work as specified, however there is a workaround; application res...	9	
			4 - defects which does not affect service functionality (typos, documentation mistakes,...	4	
			Rollup (Total)	43	
Infrastructure	Alert		2 - critical functionality is unavailable and there is no workaround	3	
			4 - defects which does not affect service functionality (typos, documentation mistakes,...	1	
	Incident		1 - service is unavailable, not possible to log in, more modules are broken	1	
			2 - critical functionality is unavailable and there is no workaround	159	
			3 - service does not work as specified, however there is a workaround; application res...	4	
			4 - defects which does not affect service functionality (typos, documentation mistakes,...	1	
	Problem		3 - service does not work as specified, however there is a workaround; application res...	1	
			Rollup (Total)	170	
	Unexpected change on the client side	Alert		4 - defects which does not affect service functionality (typos, documentation mistakes,...	2
				Incident	2 - critical functionality is unavailable and there is no workaround
Incident			3 - service does not work as specified, however there is a workaround; application res...	18	
			4 - defects which does not affect service functionality (typos, documentation mistakes,...	17	
Problem			3 - service does not work as specified, however there is a workaround; application res...	1	
Rollup (Total)	115				
Rollup (Total)				708	

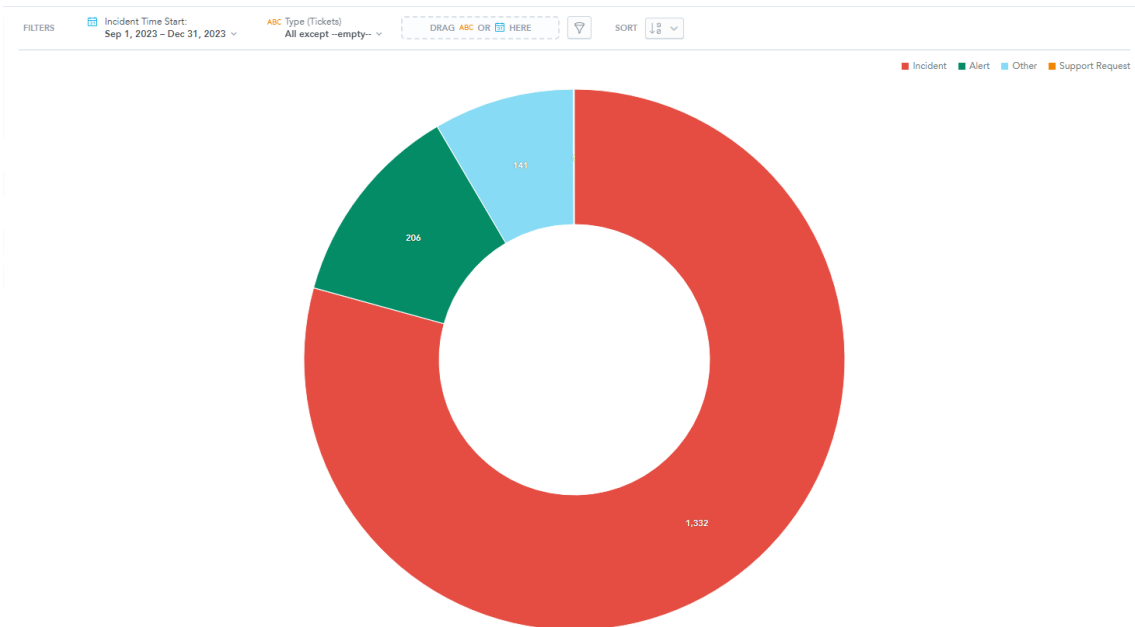
METRICS (IN COLUMNS)
Tickets

ROWS
ABC Affected Components (Ti...
ABC Root Cause Category (Tic...
ABC Type (Tickets)
ABC Severity Optional (Tickets)

COLUMNS
DRAG ABC OR HERE

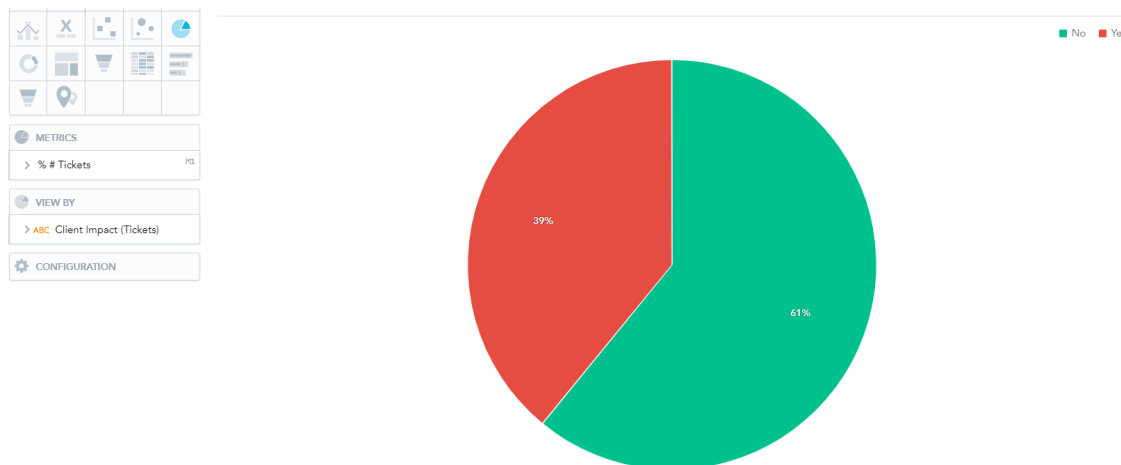
CONFIGURATION

Obrázek 4.4: Tabulka vytvořená v softwaru GoodData



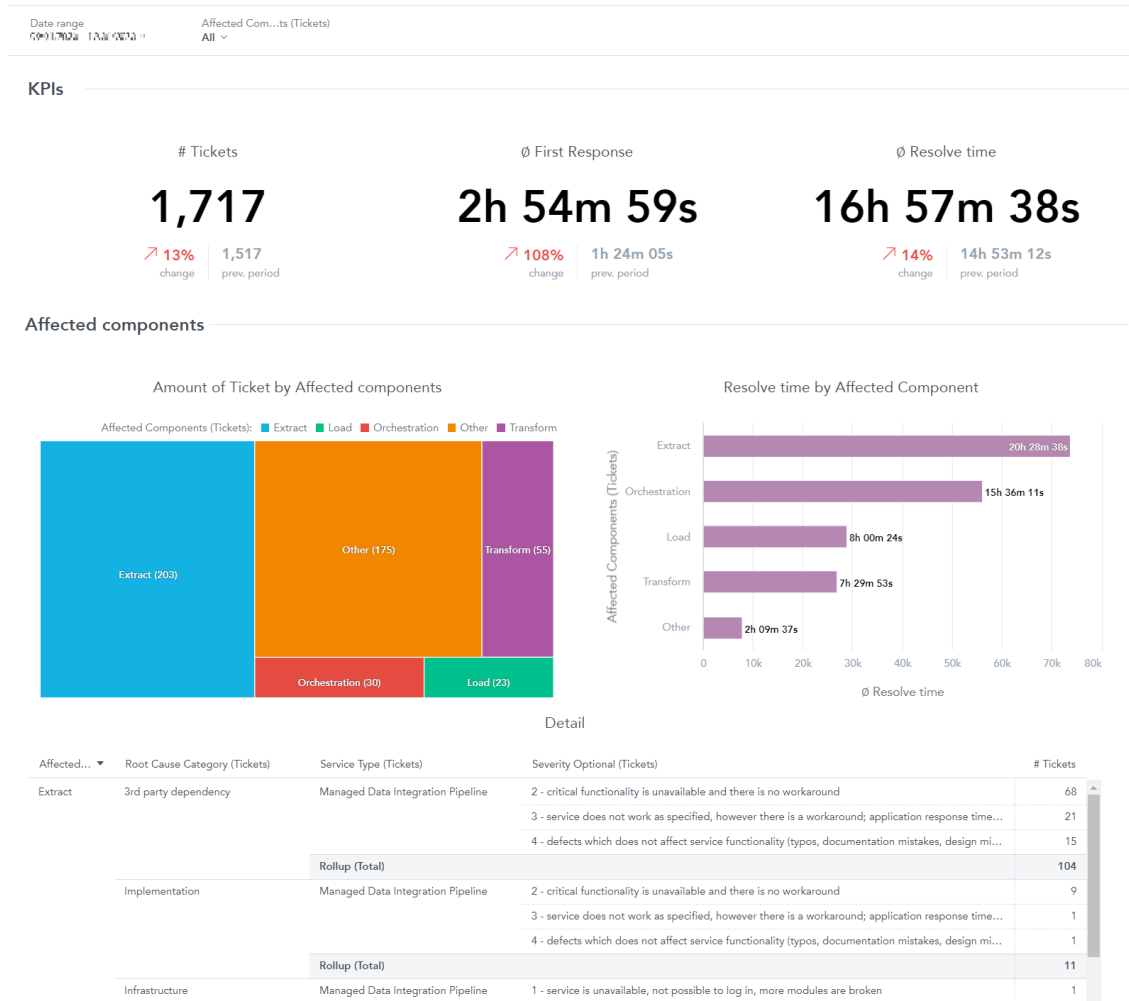
Obrázek 4.5: Prstencový graf vytvořený v softwaru GoodData

V horní části vizualizací se zobrazuje legenda pro lepší orientaci v grafu. Takto vytvořená vizualizace se může umístit na dashboard či se k ní může přistupovat i pouze v prostředí pro tvorbu vizualizací. Uživatelé mohou vybírat z větší škály vizualizací, sami si vizualizace upravovat a tvořit. Je proto dobré dostatečně je proškolit a při předávání řešení jim ukázat všechny možnosti, které software nabízí, aby ho mohli využívat efektivně. Dalším příkladem může být koláčový graf uvedený níže, který uvádí rozložení tiketů přes jejich dopad na klienta. Data jsou zde uvedena bez časového období, jde čistě o příklad pro demonstrování možností softwaru.



Obrázek 4.6: Koláčový graf vytvořený v softwaru GoodData

Z takto vytvořených vizualizací si uživatelé mohou sami vytvořit vlastní dashboard. Tento dashboard mohou využívat nejenom oni, ale mohou ho i nasdílet svým kolegům. Následně mohou na tvorbě dashboardů i spolupracovat. Zároveň se v rámci těchto dashboardů dá využít kombinace už vytvořených vizualizací s vizualizacemi novými. Uživatelé musí pouze dávat pozor, aby již vytvořené vizualizace nedopatřením nezměnili. Vždy je tedy žádoucí v případě, že uživatelé chtějí dělat změny na již vytvořených vizualizacích uložit ty nové jako kopie původních vizualizací a až poté začít dělat změny. Níže uvedený dashboard je příkladem kombinace již existujících KPI s nově vytvořenými vizualizacemi, které si uživatelé mohou během několika desítek minut vytvořit. KPI v první sekci jsou používané na dashboardu, který byl v textu uveden již dříve, stačí je tedy přepoužít. Zbytek vizualizací byl pro ukázkou vytvořen nově.



Obrázek 4.7: Ukázkový dashboard vytvořený v softwaru GoodData

5 Posouzení implementovaného řešení

Tato kapitola se bude věnovat posouzení implementovaného řešení. Bude zde rozebíráno, jak se podařilo řešení provést či jak moc odpovídá konečné řešení původním návrhům. Zároveň se v této kapitole objeví doporučení na rozšíření a vylepšení řešení. Toto řešení byl první návrh a stejně tak, jak se vyvíjí byznys, vyvíjí se i BI řešení. Je tedy potřeba vědět, že žádné BI řešení nepokrývá sto procent toho, co by teoreticky mohlo. Množství dat je dnes tak obrovské, že je není v lidských možnostech pokrýt. V rámci BI řešení je podle autorčina názoru dnes největším úspěchem vystihnout právě ta data, která jsou podstatná, a která přinesou pro datové konzumenty největší kvalitu. Detailní zpracování všech datových toků v podniku, by v teorii bylo určitě zajímavé, ale v praxi není reálné i z toho důvodu, že se ve firmách procesy často optimalizují a tím pádem i mění. Najít taková data, která budou vystihovat to, co datoví konzumenti potřebují vidět, není jednoduché a není možné jejich potřeby zjistit lepší cestou než formou přímé komunikace s nimi. Analytici mohou pokládáním těch správných otázek za účelem pochopení byznysu zjistit právě tyto stěžejní informace. Komunikace by ideálně měla probíhat po celou dobu tvorby BI řešení, a to se v rámci tohoto řešení podařilo. Primárně to bylo díky vstřícnému přístupu manažera i agentů pracujících na oddělení technické podpory.

5.1 Problémové části při ETL procesu

Nejvíce problémovou částí v rámci ETL řešení byla příprava dat pro časové metriky, která se odehrávala v rámci transformační datové vrstvy. Bohužel se zde poskytnutá data neseťkávala úplně ideálně s byznysovou realitou a bylo tedy nutné vytvořit tyto metriky složitou sadou podmínek.

Neexistoval zde totiž údaj o tom, od kdy se začíná odpočítávat doba odpovědi na tiket, protože doba jeho vytvoření to podle smluvních podmínek není. Například tikety, které se vytvoří v noci, čekají na zahájení odpočtu do deváté hodiny ranní. To znamenalo otestovat data, která se týkala různých časových úseků. Příkladem může být jiný výpočet v případě, že se tiket objeví před zahájením oficiálních pracovních hodin, jiný výpočet pro tiket příchozí v rámci pracovní doby či tiket příchozí po jejím ukončení. Tento problém bylo zároveň nutné otestovat pro všechny typy SLA, které firma poskytuje a myslet i na víkendy.

Zde se stále mohou do budoucna nacházet drobné nesrovnalosti v případě, že zde bude docházet k nějakým nestandardním, a tedy neošetřeným přístupům k zahájení

odpočtu tohoto času. Metriky tvořené z těchto dat nejsou tedy na minutu přesné, ale jsou spíše orientační. Zbytek ETL procesu nepřinášel žádné další větší problémy, které by stálo za to zmiňovat. Data byla v dobré kvalitě a hezky se s nimi pracovalo.

5.2 Posouzení vizualizací

Vizualizace byly vytvořeny podle návrhu dashboardů. Byly vytvořeny dva dashboardy, jak bylo navrženo. Oba dashboardy byly rozděleny do sekcí, tak jak bylo uvedeno v návrhu. V rámci návrhu bylo vždy uvedeno, které klíčové metriky mají být na dashboardu zpracovány. Všechny tyto metriky byly do dashboardu zapracovány úspěšně. Zároveň byly oproti návrhu některé metriky přidány, a to hlavně v rámci sjednocení dashboardu z pohledu klientů a dashboardu z pohledu agentů. První sekce obsahující KPI je pro oba tyto dashboardy tvořena stejně, byly tedy využity nápady na jednotlivé metriky z obou návrhů. Tyto metriky byly uvedeny pro větší přehlednost a pro větší škálu uvedených informací v obou dashboardech.

Jak bylo navrženo, druhá sekce se věnuje více detailním datům. Na dashboardu agentů byl oproti původnímu návrhu lehce změněn obsah druhé sekce. Vizualizace týkající se příčin incidentů byla nakonec z tohoto dashboardu odstraněna. O tento pohled ale datoví konzumenti rozhodně nepřijdou, protože byl umístěn na druhý dashboard, kde dává podle autorky práce větší smysl, a to z pohledu logického datového celku, kterým se tento dashboard zabývá. Místo této vizualizace je na dashboardu s tématem agentů umístěná vizualizace, která doplňuje pohled na výkonnost agentů, protože tomuto tématu se dashboard má věnovat primárně.

Také je oproti návrhu přidána třetí sekce, která obsahuje detailní informace o časech první odpovědi od agenta a vyřešení tiketů. Zároveň byly do všech sekcí oproti návrhu přidány drill-downy. Celkově tedy byly naplněny všechny požadavky a zároveň byl původní návrh rozšířen a doplněn o výše zmíněné vizualizace a interakce.

5.3 Posouzení naplnění uživatelských požadavků

Podarilo se vytvořit nový datový tok, data očistit, obohatit a vytvořit z nich vizualizace ve vizualizačním softwaru, ve kterém už má většina firemních datových zdrojů své místo. Teď mezi ně přibyl i Freshdesk, což bylo ve firmě velmi oceněno. Dashboard byl předán manažerovi oddělení a nasdílen vybraným uživatelům. Zároveň manažer oddělení dashboardy začal využívat v rámci firemního chodu a začlenil je jako podklady pro pravidelné schůzky s klienty. Výsledkem byl potěšen a má v plánu s těmito daty nejen nadále pracovat, ale zároveň řešení do budoucna rozšířit.

Z původně vytvořených požadavků od budoucích uživatelů byly splněny všechny čtyři body. Data byla poskytnuta v detailnější formě, a to díky interakcím ve formě drill-downů, které jsou vytvořeny pro všechny grafické vizualizace. Také byla přidána větší škála vizualizací, původní sloupcové grafy byly rozšířeny

o spojnicové grafy a samostatně vizualizované metriky (KPI). Zároveň se zde objevují nové pohledy na data, příkladem může být pohled na čas potřebný k uzavření tiketu, který dříve sledován nebyl či pohled přes rychlost a počet tiketů, které zvládnou odbavit jednotliví agenti. Přesunem vizualizací na platformu GoodData bylo docíleno sjednocení míst, kde se vizualizace objevují. V rámci tvorby dashboardu byl zároveň kladen důraz na sjednocení stylu s již existujícími dashboardy, které se zde objevují. To bylo dosaženo držením stejné struktury sekcí, komentáři, názvy vizualizací i využívanou barevnou škálou.

Čtvrtým bodem bylo poskytnout přístup k datům většímu množství osob. Dashboardy byly zpřístupněny manažerovi oddělení, jednotlivým agentům technické podpory i vedení firmy. Zvýšil se tedy počet osob, které mají k datům přístup oproti původním osobám, které musely mít administrátorský přístup do softwaru Freshdesk. Zároveň je také možné, aby si uživatelé vytvářeli vlastní pohledy nad zpracovanými daty sami. Další vizualizace na míru si tak mohou vytvořit sami, což rozšiřuje jejich možnosti v práci s daty i nad rámec jejich původních požadavků, což bylo přijato velmi kladně.

5.4 Návrhy na rozšíření a vylepšení řešení

Autorka navrhuje vytvořit dashboard, který by kontroloval datovou kvalitu vstupních dat. Při tvorbě řešení byly nalezeny tikety, kterým chyběla určitá data. Protože toto je první datový výstup s daty ze softwaru technické podpory ve firmě, je přirozené, že může docházet k prázdným místům v datech. Z pohledu autorky práce se nejednalo o časté případy a jednalo se o jednotky tiketů, ale doporučovala by vytvořit reporting, který by pokrýval právě tyto jednotky tiketů. Pokud tiketům chybí stěžejní informace, musí být z analytického řešení vyjmuty, například pokud nejsou vyplněné smluvní podmínky, ve kterých se firma pohybuje, nelze ji zařadit do výpočtu průměrné odpovědi, protože k té je potřeba znát pravidla, od kterých se tento výpočet odvíjí, a které jsou nadefinované podle poskytovaných služeb.

Takový dashboard by firmě pomohl identifikovat chybějící data a díky tomu je snadno doplnit. V případě, že by se docílilo pravidelné kontroly datové kvality, dalo by se mluvit o řešení, které by bylo zase o něco přesnější. Jak již bylo zmíněno, jednalo se spíše o výjimky, datová kvalita byla podle autorčina názoru dobrá, ale tato kontrola by pomohla dosáhnout ještě lepších výsledků. Protože se nejedná o kritická data, ale spíše o data orientační, navrhovala by autorka v aktuální situaci kontrolu datové kvality pověřenou osobou nastavit maximálně jednou za měsíc.

Dalším návrhem na vylepšení řešení je propojení incidentů s jejich konkrétním obrazem ve firemním ETL řešení. Tedy v případě, že dojde k pádu nějaké části BI řešení, a že se na základě této události vygeneruje tiket v aplikaci Freshdesk, bylo by možné do budoucna možné najít vazbu podle času incidentu a dalších metadat tiketu a propojit již vzniklou analytickou část o detailnější informace týkající se problému. Po diskuzi s firmou byla zjištěna informace, že tato data z interního ETL BizFlow by bylo možné do budoucna také sbírat a připojit jako další zdroj řešení. Díky těmto datům by se daly investigovat případné trendy týkající se technické

podpory. Nevýhodou by však bylo to, že by se tato data dala pravděpodobně sbírat v první řadě pouze z interního ETL, které nevyužívají všichni zákazníci. Firma nabízí technickou podporu i pro externí ETL softwary dostupné na trhu. Takže by tento detailní pohled vznikl pravděpodobně primárně pro klienty, kteří využívají pro BI řešení software BizzFlow.

Protože byla data nasdílena většímu množství uživatelů a byla jim poskytnuta možnost s daty samostatně pracovat, navrhuje autorka řešení také jistou formu školení pro tyto uživatele. Předání dashboardu proběhlo úspěšně, ale zároveň by se v případě zájmu dalo vytvořit krátké školení pro vysvětlení všech možností, která data nabízí. Díky tomuto školení by se dalo zlepšit povědomí o možnostech vlastních datových pohledů. Zároveň by toto školení mohlo přinést i jistou formu diskuze a poznatků o tom, jak uživatelé rozšíření BI řešení využívají, a jestli mají nějaké nápady či připomínky. Na základě této diskuze s uživateli by bylo možné vyhodnotit jejich zpětnou vazbu na poskytnuté řešení a případné připomínky či náměty do řešení zabudovat.

Závěr

Cílem této diplomové práce bylo navržení BI řešení pro data ze softwaru Freshdesk, jejich začlenění do již existujícího firemního BI řešení a následná implementace tohoto řešení. Dosažení tohoto cíle pomůže firmě vidět tato data na denní bázi a ve srozumitelné formě, která je uživatelsky přívětivá. Díky výstupům z tohoto řešení může firma dělat informovaná rozhodnutí na základě vlastních dat a lépe se orientovat ve fungování technické podpory, které se tato data týkají.

Teoretická část této práce uvedla čtenáře do problematiky business intelligence z technického a byznysového pohledu. Nejprve zde byly popsány základní termíny, koncepty a komponenty BI, které jsou více technického rázu. Následně se teoretická práce věnovala naopak byznysovému pohledu, a to na vývoj BI řešení ve firmách, její implementaci v podnicích nebo vlivu na rozhodovací procesy firem. Nechyběly zde ani trendy, které se tohoto odvětví v poslední době týkají.

Druhá část práce se věnovala praktickému návrhu řešení a jeho implementaci. Návrh i implementace se týkala využití vybraných dat, jejich zpracování, očištění a obohacení o další data. Poté se tato obohacená data nahrála do vizualizačního softwaru, kde byly na základě návrhu vytvořeny dva dashboardy s vizualizacemi. Tyto dashboardy sledují KPI týkající se počtu tiketů technické podpory, a to z nejrůznějších pohledů. Dashboardy přinášejí informace o kategoriích tiketů, jejich rozřazení přes příčiny, klienty i přes agenty technické podpory, kteří se zabývali jejich vyřešením. Mimo množství těchto tiketů se dashboard věnuje také časovému hledisku. Je zde sledována rychlosti prvních reakcí na nově přichozí incidenty i rychlost celkového vyřešení těchto problémů.

Takto vytvořené výstupy přinášejí firmě možnost větší informovanosti o oddělení technické podpory. Tyto informace může manažer využít pro efektivní řízení. Také samotní členové technické podpory mohou vidět data o svých výkonech, což může vést k větší pracovní motivaci. Zároveň mají tyto informace přínos i pro firmu jako takovou, protože díky nim může docházet i k lepšímu dlouhodobému plánování. V neposlední řadě jsou tyto vizualizace přínosným podkladem pro schůzky s klienty, kterým je technická podpora firmy zajišťována. Obě strany tak mohou mít lepší přehled o kvalitě poskytovaných služeb.

Závěrem této práce autorka navrhuje do budoucna přidat k řešení dashboard, který by kontroloval datovou kvalitu. Ten by mohl pomoci při odhalování nevyplněných tiketů, které neobsahují všechny základní informace. Druhým návrhem pro budoucí vylepšení je kombinace datového zdroje Freshdesk s informacemi z ETL softwaru BizzFlow. Toto rozšíření by mohlo do dat doplnit detailnější informace o vzniku incidentů.

Seznam použité literatury

- Alsaad, Abdallah et al. (2022). „Linking business intelligence with the performance of new service products: Insight from a dynamic capabilities perspective“. In: *Journal of Innovation Knowledge* 7.4, s. 100262. ISSN: 2444-569X. DOI: <https://doi.org/10.1016/j.jik.2022.100262>. URL: <https://www.sciencedirect.com/science/article/pii/S2444569X2200097X>.
- Azevedo, Ana a Manuel Filipe Santos (2021). *Integration Challenges for Analytics, Business Intelligence, and Data Mining*. 1. vyd. Hershey, PA: IGI Global. ISBN: 978-1-79985-781-5.
- Azevedo, João, Júlio Duarte a Manuel Filipe Santos (2022). „Implementing a business intelligence cost accounting solution in a healthcare setting“. In: *Procedia Computer Science* 198. 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, s. 329–334. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2021.12.249>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050921024881>.
- Boldosova, Valeriia (2020). „Telling stories that sell: The role of storytelling and big data analytics in smart service sales“. In: *Industrial Marketing Management* 86, s. 122–134. ISSN: 0019-8501. DOI: <https://doi.org/10.1016/j.indmarman.2019.12.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0019850118303353>.
- Eagar, G. (2023). *Data Engineering with AWS: Acquire the skills to design and build AWS-based data transformation pipelines like a pro*. Packt Publishing. ISBN: 978-1-80461-313-9. URL: <https://books.google.cz/books?id=1dffEAAAQBAJ>.
- Evergreen, Stephanie D. H. (2020). *Effective data visualization. the right chart for the right data*. Second edition. Los Angeles: Sage. ISBN: 978-1-5443-5088-2.
- Fotr, Jiří et al. (2020). *Tvorba strategie a strategické plánování. teorie a praxe*. 2., aktualizované a doplněné vydání. Praha: Grada Publishing. ISBN: 978-80-271-1632-4. URL: <https://www.bookport.cz/e-kniha/tvorba-strategie-a-strategicke-planovani-1619889/> (cit. 19. 01. 2024).
- Cheng, Jie et al. (2023). „The impact of business intelligence, big data analytics capability, and green knowledge management on sustainability performance“. In: *Journal of Cleaner Production* 429, s. 139410. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2023.139410>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652623035680>.

- Jackson, Peter a Caroline Carruthers (2019). *Data Driven Business Transformation. How to Disrupt, Innovate and Stay Ahead of the Competition*. První vydání. Chichester: John Wiley & Sons, Ltd. ISBN: 978-1-119-54315-2.
- Kumar V. K., Anoop (2022). *Business Intelligence Demystified. Understand and Clear All Your Doubts and Misconceptions About BI*. 1. vyd. India: BPB Publication. ISBN: 978-93-91030-087.
- Lee, James, Tao Wei a Suresh Kumar Mukhiya (2018). *Hands-On Big Data Modeling. Effective database design techniques for data architects and business intelligence professionals*. 1. vyd. Birmingham, UK: Packt Publishing Ltd. ISBN: 978-1-78862-090-1.
- Löffler, Vladimír, Barbora Štětinová a Lukáš Bernat (2021). *BIG DATA a umělá inteligence pro manažery*. 1. vyd. Vladimír Löffler. ISBN: 978-80-908226-3-4. URL: https://books.google.cz/books?id=U15cEAAAQBAJ&printsec=frontcover&dq=BIG+DATA+a+um%C4%9B1%C3%A1+inteligence+pro+mana%C5%BEery&hl=cs&newbks=1&newbks_redir=0&sa=X&redir_esc=y#v=onepage&q=BIG%20DATA%20a%20um%C4%9B1%C3%A1%20inteligence%20pro%20mana%C5%BEery&f=false (cit. 20. 01. 2024).
- Machiraju, Suren a Suraj Gaurav (2018). *Power BI Data Analysis and Visualization*. 1. vyd. Walter de Gruyter Inc. ISBN: 978-1-5474-1678-3.
- Mihiranga, Nisal (2022). *POWER BI DATA MODELING build interactive visualizations, learn dax, power query, and... develop bi models*. 1. vyd. BPB PUBLICATIONS. ISBN: 978-93-89328-837.
- Mudau, Thanyani Norman, Jason Cohen a Elmarie Papageorgiou (2024). „Determinants and consequences of routine and advanced use of business intelligence (BI) systems by management accountants“. In: *Information Management* 61.1, s. 103888. ISSN: 0378-7206. DOI: <https://doi.org/10.1016/j.im.2023.103888>. URL: <https://www.sciencedirect.com/science/article/pii/S0378720623001362>.
- Munoz, J. Mark (2018). *Global business intelligence*. 1. vyd. New York: Routledge, Taylor & Francis Group. ISBN: 978-1-138-20368-6.
- Nordeen, A. (2020). *Learn Data Warehousing in 24 Hours*. Guru99. URL: <https://books.google.cz/books?id=wgf9DwAAQBAJ>.
- Pałys, Marcin a Andrzej Pałys (2023). „Benefits and Challenges of Self-Service Business Intelligence Implementation“. In: *Procedia Computer Science* 225. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2023), s. 795–803. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2023.10.066>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050923012243>.
- Pour, Jan et al. (2018). *Self service business intelligence. jak si vytvořit vlastní analytické, plánovací a reportingové aplikace*. Praha: Grada Publishing. ISBN: 978-802-7108-169.
- Ryan, Lindy (2018). *Visual data storytelling with Tableau*. 1. vyd. Boston: Addison-Wesley. ISBN: 978-0-13-471283-3.

- Skyrius, Rimvydas (2021). *Business Intelligence. A Comprehensive Approach to Information Needs, Technologies and Culture*. 1. vyd. Switzerland: Springer International Publishing. ISBN: 978-3-030-67031-3.
- Zangari, Alessandro et al. (2023). „Ticket automation: An insight into current research with applications to multi-level classification scenarios“. In: *Expert Systems with Applications* 225, s. 119984. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2023.119984>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417423004864>.
- Zelenka, Martin a Athanasios Podaras (2021). „Increasing the Effectivity of Business Intelligence Tools via Amplified Data Knowledge“. In: *Studies in Informatics and Control* 30.2, s. 67–77. ISSN: 1220-1766.