



# Cloudová řešení Business Intelligence

## Diplomová práce

*Studijní program:* N6209 – Systémové inženýrství a informatika

*Studijní obor:* 6209T021 – Manažerská informatika

*Autor práce:* **Bc. Vojtěch Lank**

*Vedoucí práce:* Ing. Vladimíra Zádová, Ph.D.



## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Vojtěch Lank**  
Osobní číslo: **E14000393**  
Studijní program: **N6209 Systémové inženýrství a informatika**  
Studijní obor: **Manažerská informatika**  
Název tématu: **Cloudová řešení Business Intelligence**  
Zadávací katedra: **Katedra informatiky**

### Z á s a d y p r o v y p r á c o v á n í :

1. Prostředky analýzy podnikových dat a podpory manažerského rozhodování
2. Trendy v BI - změna charakteristiky požadavků, dat i uživatelů
3. BI v Cloudu - specifika řešení, vyhodnocení komerčních nástrojů BI
4. Analýza průběhu návrhu a implementace cloudového řešení BI v konkrétním firemním prostředí
5. Zhodnocení přínosu řešení

Rozsah grafických prací:

Rozsah pracovní zprávy: **65 normostran**

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

**MAYER-SCHÖNBERGER, Viktor a Kenneth CUKIER. Big Data: Revoluce, která změní způsob, jak žijeme, pracujeme a myslíme. Brno: Computer Press, 2014. ISBN 978-80-251-4119-9.**

**KROENKE, David a David J AUER. Databáze. Brno: Computer Press, 2015. ISBN 978-80-251-4352-0.**

**GENDRON, Michael S. . Business intelligence and the cloud: strategic implementation guide. Hoboken: Wiley, 2014. ISBN 1118631722.**

**MARR, Bernard. Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance. Hoboken: Wiley, 2015. ISBN 1118965833.**

**Elektronická databáze článků ProQuest (knihovna.tul.cz).**

Vedoucí diplomové práce:

**Ing. Vladimíra Zádová, Ph.D.**

Katedra informatiky

Konzultant diplomové práce:

**Ing. Martin Matějka**

BI Consultant, Keboola s. r. o.

Datum zadání diplomové práce: **31. října 2015**

Termín odevzdání diplomové práce: **31. května 2017**



doc. Ing. Miroslav Žížka, Ph.D.  
děkan



doc. Ing. Jan Skrbek, Dr.  
vedoucí katedry

V Liberci dne 31. října 2015

## Prohlášení

Byl jsem seznámen s tím, 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 (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

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

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

# Poděkování

Rád bych touto cestou poděkoval své vedoucí diplomové práce, paní Ing. Vladimíře Zádové, PhD, za cenné rady, věcné připomínky a vstřícnost při konzultacích.

Chtěl bych také poděkovat firmě Keboola s.r.o. a především panu Ing. Martinu Matějkovi, za umožnění spolupráce, ochotu při konzultacích a cenné praktické rady, které mi během vypracování předal.

Děkuji i mé partnerce a rodině, za neustálou podporu, kterou mi během celého studia věnovali.

# **Anotace**

Diplomová práce je zaměřena na problematiku cloudových řešení Business Intelligence. Hlavním cílem práce je identifikovat trendy, kterým současné BI systémy podléhají, a poukázat na značnou rozdílnost komerčně nabízených cloudových BI platforem. Vedlejším cílem práce je návrh a implementace cloudového BI řešení v konkrétním firemním prostředí.

Teoretická část práce se nejprve zabývá obecným tématem Cloud Computingu a historickým vývojem systémů pro podporu rozhodování. Následně je objasněno postavení současných BI řešení v dnešním světě a shrnuto, jaký vliv měl na jednotlivé komponenty BI systémů jejich přesun do cloudového prostředí.

V praktické části práce se nachází vyhodnocení porovnání dvou cloudových BI platforem – GoodData a Microsoft Power BI – kterého bylo dosaženo na základě stanované testovací metodiky a pomocí sady kritérií. Poslední kapitola řeší návrh a implementaci cloudového BI v nejmenované společnosti „ABC“, jež bylo realizováno pomocí platformy GoodData a cloudového ETL nástroje Keboola Connection.

## **Klíčová slova**

Cloud, Business Intelligence, trend, Big Data, GoodData, Microsoft Power BI, Keboola Connection, data, analýza, podpora rozhodování

# **Annotation**

## **Business Intelligence Cloud Solutions**

This diploma thesis focuses on the topic of cloud-based Business Intelligence solutions. The main aim of this work is to identify trends that are subjected to current BI systems, and to highlight significant differences between proprietary BI platforms currently available on the market. The secondary objective is to design and implement a cloud-based BI solution in the environment of a particular company.

The theoretical part firstly deals with the general topic of Cloud Computing and the historical evolution of decision support systems. Later on, the position of current BI solutions in the world of today is clarified and the influence of the cloud environment's adoption on individual components of BI systems is summarized.

In the practical part, the comparison of two cloud-based BI platforms – GoodData and Microsoft Power BI – is evaluated based on pre-defined testing methodology and a set of criteria. The last chapter deals with the design and implementation of a cloud-based BI solution in the anonymous company "ABC". The project has been completed by using GoodData platform and cloud-based ETL tool – Keboola Connection.

## **Key words**

Cloud, Business Intelligence, trend, Big Data, GoodData, Microsoft Power BI, Keboola Connection, data, analysis, decision support

# Obsah

<b>Seznam ilustrací.....</b>	<b>13</b>
<b>Seznam tabulek.....</b>	<b>15</b>
<b>Seznam použitých zkratk .....</b>	<b>16</b>
<b>Úvod.....</b>	<b>18</b>
<b>Zhodnocení současného stavu .....</b>	<b>20</b>
<b>1 Cloud Computing .....</b>	<b>22</b>
1.1 Historie Cloud Computingu.....	22
1.2 Důvod vzniku.....	24
1.3 Distribuční modely .....	24
1.3.1 IaaS – infrastruktura jako služba (Infrastructure as a Service).....	25
1.3.2 PaaS – platforma jako služba (Platform as a Service).....	27
1.3.3 SaaS – software jako služba (Software as a Service).....	27
1.4 Modely nasazení .....	28
1.4.1 Veřejný cloud .....	28
1.4.2 Privátní cloud.....	29
1.4.3 Hybridní cloud.....	30
1.4.4 Komunitní cloud.....	30
1.5 Shrnutí charakteristik.....	31
1.5.1 Výhody Cloud Computingu .....	31
1.5.2 Nevýhody Cloud Computingu.....	32
<b>2 Podpora rozhodování a analýza dat .....</b>	<b>33</b>
2.1 Proces rozhodování.....	33
2.2 Struktura rozhodovacích problémů.....	34
2.3 Milníky v historii podpory rozhodování .....	34
2.4 Business Intelligence .....	36
2.5 BI 2.0 .....	38



2.6	Mobilní BI (3.0).....	40
<b>3</b>	<b>Trendy v BI .....</b>	<b>42</b>
3.1	Data-Driven Company .....	42
3.2	Uvést data do širšího kontextu.....	43
3.3	Kvalita dat .....	44
3.4	Data Mining .....	45
3.5	Nestrukturovaná data .....	47
3.6	Semi-strukturovaná data .....	48
3.7	Big Data .....	49
<b>4</b>	<b>BI v cloudu .....</b>	<b>52</b>
4.1	Zdroje dat.....	52
4.2	Datové transformace .....	54
4.3	Ukládání dat.....	55
4.4	Vizualizace dat.....	56
4.5	Shrnutí.....	57
<b>5</b>	<b>Srovnání vybraných cloudových BI.....</b>	<b>58</b>
5.1	Volba nástrojů.....	58
5.2	Modelová situace .....	59
5.3	Předmět hodnocení .....	61
5.3.1	Integrace dat .....	61
5.3.2	Datový model .....	61
5.3.3	Datumová dimenze .....	61
5.3.4	Metriky .....	62
5.3.5	Vizualizace .....	63
5.3.5.1	Mapy.....	63
5.3.5.2	KPI.....	63
5.3.5.3	Reporty a dashboardy .....	64
5.3.5.4	Změna detailu .....	64
5.3.6	Ad-hoc dotazy .....	64
5.3.7	Forecasting .....	64

5.3.8	Jedinečné vlastnosti .....	65
5.3.9	Export dat .....	65
5.3.10	Mobilní přístup .....	65
5.3.11	Sdílení a omezení přístupu .....	65
5.3.12	Rychlost .....	66
5.3.13	Cena a omezení.....	66
5.4	GoodData .....	67
5.4.1	Koncept.....	67
5.4.2	Integrace dat a datový model.....	69
5.4.3	Datumová dimenze .....	72
5.4.4	Metriky .....	73
5.4.5	Vizualizace .....	75
5.4.5.1	Mapy.....	75
5.4.5.2	KPI.....	77
5.4.5.3	Reporty a dashboardy .....	78
5.4.5.4	Změna detailu .....	79
5.4.6	Ad-hoc dotazy .....	80
5.4.7	Forecasting .....	81
5.4.8	Jedinečné vlastnosti .....	82
5.4.9	Export dat .....	82
5.4.10	Mobilní přístup .....	82
5.4.11	Sdílení a omezení přístupu .....	83
5.4.12	Rychlost.....	83
5.4.13	Cena a omezení.....	83
5.5	Microsoft Power BI .....	84
5.5.1	Koncept.....	84
5.5.2	Integrace dat .....	85
5.5.2.1	Power BI (online) .....	85
5.5.2.2	Power BI Desktop.....	86
5.5.2.3	Power BI Data Gateway .....	87

5.5.3	Datový model .....	88
5.5.4	Datumová dimenze .....	90
5.5.5	Metriky .....	90
5.5.6	Vizualizace .....	92
5.5.6.1	Mapy .....	94
5.5.6.2	KPI.....	96
5.5.6.3	Reporty a dashboardy .....	96
5.5.6.4	Změna detailu .....	98
5.5.7	Ad-hoc dotazy .....	99
5.5.8	Forecasting .....	99
5.5.9	Jedinečné vlastnosti .....	100
5.5.10	Export dat .....	102
5.5.11	Mobilní přístup .....	103
5.5.12	Sdílení a omezení přístupu .....	105
5.5.13	Rychlost .....	105
5.5.14	Cena a omezení.....	106
5.6	Souhrnné porovnání kritérií .....	106
5.6.1	Závěrečné hodnocení GoodData .....	110
5.6.2	Závěrečné hodnocení Power BI.....	111
5.7	Doporučení.....	111
<b>6</b>	<b>Implementace cloudového BI ve společnosti ABC.....</b>	<b>113</b>
6.1	Úvodní analýza .....	113
6.1.1	Prostředí.....	113
6.1.2	Předmět analýzy .....	113
6.1.3	Výchozí stav .....	114
6.1.4	Požadavky a cíle .....	115
6.2	Návrh řešení.....	115
6.3	Implementace.....	117
6.3.1	Zdroje dat.....	117
6.3.2	Extrakce dat z Google Drive .....	117

6.3.3	Transformace .....	118
6.3.4	Odeslání dat do GoodData.....	123
6.3.5	Návrh metrik.....	125
6.3.6	Dashboard.....	127
6.3.7	Zobrazení na mapě ČR .....	129
6.4	Zhodnocení výsledků a přínosu projektu.....	131
6.4.1	Dosažené cíle.....	131
6.4.2	Přínos pro firmu.....	132
6.4.3	Ekonomické zhodnocení.....	132
<b>Závěr .....</b>		<b>134</b>
<b>Seznam použité literatury .....</b>		<b>138</b>
	Citace .....	138
	Bibliografie.....	145
<b>Seznam příloh .....</b>		<b>146</b>
	Příloha A – Strom dotazů vygenerovaný v GoodData .....	147
	Příloha B – Vygenerovaný SQL skript reportu v GoodData.....	148
	Příloha C – Specifikace tabulek AF a CONT.....	150
	Příloha D – SQL skript výsledné transformace .....	151
	Příloha E – Datový model GoodData .....	154

## Seznam ilustrací

Obrázek 1-1: Porovnání distribučních modelů.....	25
Obrázek 2-1: Hierarchie podnikových informačních systémů .....	36
Obrázek 2-2: Generační vývoj BI.....	41
Obrázek 5-1: Gartner Magic Quadrant – únor 2016, s vyznačenými zvolenými nástroji...	59
Obrázek 5-2: Datový model pro testování cloudových BI nástrojů .....	60
Obrázek 5-3: GD – architektura platformy.....	68
Obrázek 5-4: GD Cloud Connect Designer – logický datový model .....	69
Obrázek 5-5: GD Cloud Connect Designer – ETL .....	71
Obrázek 5-6: GD – součet množství po dnech kalendářního měsíce a letech.....	73
Obrázek 5-7: GD – počet objednávek na mapě dle souřadnic obchodu.....	76
Obrázek 5-8: GD – počet objednávek na mapě dle státu zákazníka .....	76
Obrázek 5-9: GD – KPI dashboard.....	77
Obrázek 5-10: GD – návrh reportu.....	78
Obrázek 5-11: GD – dashboard .....	79
Obrázek 5-12: GD – prostředí pro ad-hoc dotazování .....	80
Obrázek 5-13: GD – forecasting pomocí regresní přímky .....	81
Obrázek 5-14: Power BI – schéma platformy .....	85
Obrázek 5-15: Power BI Desktop – dialog načtení souboru CSV .....	86
Obrázek 5-16: Power BI Desktop – datový model.....	88
Obrázek 5-17: Power BI Desktop – nastavení datové kategorie .....	89
Obrázek 5-18: Power BI – základní paleta vizualizačních nástrojů.....	93
Obrázek 5-19: Power BI – mapa dle státu zákazníka.....	94
Obrázek 5-20: Power BI – mapa dle souřadnic obchodu .....	95
Obrázek 5-21: Power BI – teplotní mapa Esri ArcGIS .....	95
Obrázek 5-22: Power BI – vizualizace KPI.....	96
Obrázek 5-23: Power BI – vzájemná interakce v rámci jedné stránky sestavy.....	97
Obrázek 5-24: Power BI – řídicí panel.....	98
Obrázek 5-25: Power BI – forecasting .....	100
Obrázek 5-26: Power BI – Q&A – dotazování v přirozeném jazyce .....	101
Obrázek 5-27: Power BI – návrh mobilního zobrazení řídicího panelu.....	103

Obrázek 5-28: Power BI pro Android – řídicí panel .....	104
Obrázek 6-1: Proces získávání nových zákazníků .....	113
Obrázek 6-2: Nový datový model .....	116
Obrázek 6-3: Schéma funkce celého řešení.....	116
Obrázek 6-4: KBC – extraktor Google Drive.....	118
Obrázek 6-5: SQL skript transformace pro nahrazení atributu partner_type .....	119
Obrázek 6-6: Schématické znázornění datové transformace.....	120
Obrázek 6-7: SQL skript transformace pro tabulku AF_CONT_EVENTS .....	121
Obrázek 6-8: KBC – nastavení transformace .....	122
Obrázek 6-9: KBC – writer GoodData .....	123
Obrázek 6-10: Výsledný logický datový model .....	123
Obrázek 6-11: KBC – writer GoodData – tabulka AF_CONT .....	124
Obrázek 6-12: KBC – writer GoodData – tabulka AF_CONT_EVENTS .....	125
Obrázek 6-13: GD – definice metriky „AF Created Count“ .....	126
Obrázek 6-14: GD – definice metriky „AF Created Count Month Growth“ .....	126
Obrázek 6-15: GD – formátování metriky „AF Created Count Month Growth“ .....	127
Obrázek 6-16: GD – definice metriky „AVG AF Processed Days“.....	127
Obrázek 6-17: GD – hlavní dashboard .....	128
Obrázek 6-18: GD – chybné zobrazení mapy na základě PSČ .....	130

## Seznam tabulek

Tabulka 1: Testovací metriky .....	62
Tabulka 2: Testovací metriky (GoodData – MAQL) .....	74
Tabulka 3: Testovací metriky (Power BI – DAX) .....	91
Tabulka 4: Souhrnné srovnání GoodData a Power BI .....	107
Tabulka 5: Souhrnné srovnání GoodData a Power BI dle kritérií.....	109

## Seznam použitých zkratek

Zkratka	Anglický termín	Český termín
<b>AF</b>	Active Form	Aktivní formulář
<b>API</b>	Application Programming Interface	Rozhraní pro programování aplikací
<b>BI</b>	Business Intelligence	Business Intelligence
<b>CC</b>	Cloud Computing	Cloud Computing
<b>CONT</b>	Contract	Smlouva
<b>CRM</b>	Customer Relationship Management	Řízení vztahů se zákazníky
<b>CSV</b>	Comma-separated values	Hodnoty oddělené čárkami
<b>DAX</b>	Data Analysis Expressions	Datový analytický výraz
<b>DM</b>	Data Mining	Dolování dat
<b>DSS</b>	Decision Support System	Systémy na podporu rozhodování
<b>DW</b>	Data Warehouse	Datový sklad
<b>EIS</b>	Executive Information System	Informační systém pro podporu řízení
<b>ERP</b>	Enterprise Resource Planning	Plánování podnikových zdrojů
<b>ESS</b>	Executive Support System	Informační systém pro podporu řízení
<b>ETL</b>	Extract, Transform and Load	Extrakce, transformace a nahrání
<b>GA</b>	Google Analytics	Google Analytics
<b>GB</b>	Gigabyte	Gigabajt
<b>GD</b>	GoodData	GoodData
<b>GPS</b>	Global Positioning System	Globální polohový systém
<b>IaaS</b>	Infrastructure as a Service	Infrastruktura jako služba
<b>IOT</b>	Internet of Things	Internet věcí
<b>JSON</b>	JavaScript Object Notation	JavaScriptový objektový zápis
<b>KBC</b>	Keboola Connection	Keboola Connection
<b>KPI</b>	Key Performance Indicator	Klíčový ukazatel výkonnosti
<b>LAU</b>	Local Administrative Units	Místní samosprávné jednotky
<b>LDM</b>	Logical Data Model	Logický datový model
<b>MAQL</b>	Multi-Dimension Analytical Query Language	Multi-dimenzionální analytický dotazovací jazyk
<b>MIS</b>	Management Information System	Manažerský informační systém
<b>MS</b>	Microsoft	Microsoft
<b>NLP</b>	Natural Language Processing	Zpracování přirozeného jazyka
<b>NUTS</b>	Nomenclature of Units for Territorial Statistics	Nomenklatura územních statistických jednotek



<b>Zkratka</b>	<b>Anglický termín</b>	<b>Český termín</b>
<b>OLAM</b>	Online Analytical Mining	Online analytické dolování dat
<b>OLAP</b>	Online Analytical Processing	Online analytické zpracování dat
<b>PaaS</b>	Platform as a Service	Platforma jako služba
<b>PB</b>	Petabyte	Petabajt
<b>PSČ</b>	Postal Code	Poštovní směrovací číslo
<b>REST</b>	Representational State Transfer	Reprezentativní stavový přenos
<b>RLS</b>	Row-Level Security	Možnosti specifikovat přístupová práva pro jednotlivé řádky
<b>SaaS</b>	Software as a Service	Software jako služba
<b>SCM</b>	Supply Chain Management	Řízení dodavatelského řetězce
<b>SOAP</b>	Simple Object Access Protocol	Protokol pro snadný přístup k objektům
<b>SQL</b>	Structured Query Language	Strukturovaný dotazovací jazyk
<b>TB</b>	Terabyte	Terabajt
<b>USD</b>	United States dollar	Americký dolar
<b>UX</b>	User Experience	Uživatelský prožitek
<b>XML</b>	Extensible Markup Language	Rozšiřitelný značkovací jazyk

# Úvod

V posledních několika letech zažívá obrovský rozmach Cloud Computing. Ve zjednodušené formě se dá říci, že se nejedná o nic jiného, než o využívání sdíleného výpočetního výkonu, respektive o prostý přesun serverů a lokálně instalovaných softwarových aplikací do sdíleného prostředí přístupného přes internet.

Jelikož je v dnešní době, vzhledem k vysokému pokrytí rychlým internetovým připojením, využívání cloudových aplikací velmi běžnou záležitostí, neunikly tomuto značnému technologickému trendu posledních let ani aplikace či nástroje z oblasti Business Intelligence (BI). A právě tímto tématem se tato diplomová práce zabývá. Konkrétně je zaměřena na analýzu současného stavu na trhu s cloudovými řešeními BI.

Práce má dva hlavní a jeden vedlejší cíl. Prvním hlavním cílem je identifikovat současné trendy v oblasti BI a druhým následně porovnat vybrané cloudové platformy BI a poukázat na jejich rozdílnost. Vedlejším cílem je návrh a implementace cloudového BI řešení v konkrétním firemním prostředí.

Práce je strukturována do šesti kapitol, přičemž první čtyři jsou zaměřeny na teoretickou stránku věci a poslední dvě se zabývají praktickými úlohami. Nejprve je vysvětlen pojem Cloud Computing a stručně probrána historie jeho vývoje. Dále je pojednáno o obecných principech, které jsou v souvislosti s tímto pojmem používány, jako jsou distribuční modely (IaaS, PaaS, SaaS) a modely nasazení (veřejný cloud, privátní cloud a hybridní cloud).

Druhá kapitola pojednává o problematice manažerského rozhodování a jeho podpoře. Kapitola si klade za cíl uvést do problematiky manažerského rozhodování a stručně nastínit historický vývoj oblasti počítačových systémů, která se podpory rozhodování bezprostředně týká. V chronologickém vývoji těchto systémů jsou zastoupeny jak rané implementace (EIS, MIS), tak pozdější vývojové éry Business Intelligence (BI 1.0, BI 2.0, BI 3.0). Jedná se o stručné shrnutí toho, jakým způsobem se systémy pro podporu rozhodování vyvinuly od svých raných počátků až do dnešní podoby.

Kapitoly číslo 3 a 4 jsou zaměřeny na naplnění prvního hlavního cíle, jímž je shrnutí trendů, kterým dnešní BI systémy podléhají. V první zmíněné kapitole jsou zastoupena například témata týkající se nestructurovaných nebo semi-structurovaných dat, kvality dat nebo také Big Dat. Cílem kapitoly je též objasnit, jakým způsobem se změnilы požadavky uživatelů na BI a jakým způsobem je na tyto požadavky reagováno.

Poslední část teoretického základu se zabývá změnami v architekturách jednotlivých komponent BI po jejich přesunutí do cloudového prostředí. Kapitola se zaměřuje na dnes používané koncepte a technologie a výhody vyplývající z jejich použití.

Kapitola číslo 5 obsahuje syntézu druhého hlavního cíle práce. Tím je poukázat na velikost rozdílů mezi komerčně nabízenými cloudovými BI nástroji. Cíl se odvolává na značnou rozličnost nabízených BI řešení, jejichž nesprávná volba může pro firmy znamenat zmařené investice. Cíle je dosaženo za pomoci srovnání dvou výrazně rozdílných plně cloudových BI platforem na základě 20 stanovených kritérií. Pro srovnání byly zvoleny platformy GoodData a Microsoft Power BI, kde výsledkem je objasnění jejich hlavních předností a nedostatků.

Poslední kapitola naplňuje vedlejší cíl práce, kterým je návrh a implementace cloudového BI v konkrétním firemním prostředí. Pro tento účel poskytla spolupráci společnost „ABC“, která si však přála zůstat v anonymitě. Projekt se skládá z úvodní analýzy, z návrhu řešení a z vytvoření funkčního prototypu, jež je postaven na technologiích GoodData a Keboola Connection.

Zdrojem prvních teoretických kapitol je odborná literatura zaměřená na problematiku Cloud Computingu a Business Intelligence a taktéž velké množství webových stránek souvisejících s toutéž problematikou. Praktická část se opírá především o dokumentaci jednotlivých softwarových nástrojů. Velmi cenným zdrojem byly také osobní konzultace, které poskytly spoustu věcných připomínek z praxe.

## Zhodnocení současného stavu

Přestože Cloud Computing započal svou éru již na začátku 21. století, k jeho masivnímu rozvoji došlo až kolem roku 2010, kdy se stal tento pojem velmi populárním a často diskutovaným. Proto bylo, především v posledních několika letech, vydáno poměrně velké množství odborných knih, vědeckých článků i bakalářských či diplomových prací zaměřených na tuto problematiku. V současné době je pojem „Cloud“ natolik populárním, že lze jen velmi těžko najít odborný časopis se zaměřením na informační technologie, ve kterém by se článek o Cloud Computingu nevyskytoval. Zdrojem pro tuto práci jsou jak níže uvedené publikace, tak velké množství webových příspěvků od renomovaných IT pracovníků z praxe.

Jedním z nosných děl této práce je kniha z roku 2014 od Michaela S. Gendrona (z nakladatelství Wiley) nesoucí název *Business Intelligence and the Cloud: Strategic Implementation Guide*.<sup>1</sup> Kniha je zpočátku zaměřena na vysvětlení obecných pojmů spojených s Cloud Computingem, dále pak na polemiku nad výhodami a nevýhodami plynoucími z přechodu do cloudového prostředí; v závěru autor poukazuje na to, jakým způsobem lze za pomoci technologií získat obchodní výhodu a jakým způsobem mohou být cloudové aplikace integrovány do podnikového BI. Tato publikace je jednou z nejpodstatnějších pro uchopení teoretického základu v rámci této práce.

Další knihou, která se daným tématem zabírá, je kniha nazvaná *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)* od Michaela Kavise, vydaná taktéž v roce 2014 nakladatelstvím Wiley.<sup>2</sup> Tato publikace je více zaměřena na elementární prvky cloudových technologií a konkrétní technická řešení. Přestože je větší část knihy směřována nad rámec této práce, v knize lze nalézt velké množství výstižných formulací, které jsou pro definování základních pojmů nezbytné.

---

<sup>1</sup> GENDRON, Michael. *Business intelligence and the cloud: strategic implementation guide*. Hoboken: Wiley, 2014. ISBN 1118631722.

<sup>2</sup> KAVIS, Michael. *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)*. 1st Edition. Hoboken, New Jersey: Wiley, 2014. Wiley CIO series. ISBN 978-1-118-82627-0.

Od českých autorů není žádná odborná kniha s touto tematikou dostupná, ale několik zahraničních publikací bylo do českého jazyka přeloženo. Je možné jmenovat například *Cloud Computing: praktický průvodce* od autorů Anthonyho T. Velta, Tobyho J. Velta a Roberta C. Elsenpetera.<sup>3</sup> Jak sám název napovídá, kniha tvoří průvodce světem Cloud Computingových technologií. Autoři se snažili přinést pohled na tento koncept ze všech stran, a to jak ze strany výhod a nevýhod, tak i ze strany technologické nebo legislativní.

Jako další česky publikovanou knihu je možné zmínit dílo s názvem *Big Data: Revoluce, která změní způsob, jak žijeme, pracujeme a myslíme* od autorů Viktora Mayer-Schönbergera a Kennetha Cukiera.<sup>4</sup> Zde se autoři zaměřují na nový trend ve zpracování velkých objemů dat, v němž jde namísto klasické analýzy vzorku spíše o analýzu všech dat pro získání co nejpřesnějšího výsledku. Téma Big Data je dnes s cloudovými aplikacemi typu Business Intelligence velmi úzce spojeno, právě kvůli analýze velkých objemů podnikových dat, pro získání určité obchodní výhody a zvýšení konkurenceschopnosti, které s dřívějšími technologiemi nebylo dost dobře možné.

Mezi kvalifikačními pracemi obhájenými v roce 2016 se vyskytly hned dvě s přímým zaměřením na cloudová řešení BI. První z nich je bakalářská práce nesoucí název *Business Intelligence v podnikové praxi s využitím cloud computingu* od Davida Huňky<sup>5</sup> a druhá je pak diplomová práce s názvem *Business Intelligence v prostředí Cloudu* od Bc. Tomáše Náhlovského<sup>6</sup>, obě od studentů Vysoké školy ekonomické v Praze. Zatímco bakalářská práce je zaměřena na praktickou implementaci cloudového řešení BI v organizaci, diplomová práce má za úkol pouze porovnat tradiční BI řešení oproti jejich cloudovým verzím a vystihnout odlišnosti obou přístupů. Obě práce jsou však velmi cennými prameny i mezi podklady pro tuto diplomovou práci.

---

<sup>3</sup> VELTE, Anthony, Toby VELTE a Robert ELSENPETER. *Cloud Computing: praktický průvodce*. 1. vyd. Brno: Computer Press, 2011. ISBN 978-80-251-3333-0.

<sup>4</sup> MAYER-SCHÖNBERGER, Viktor a Kenneth CUKIER. *Big Data: Revoluce, která změní způsob, jak žijeme, pracujeme a myslíme*. 1. vyd. Brno: Computer Press, 2014. ISBN 978-80-251-4119-9.

<sup>5</sup> HUŇKA, David. *Business Intelligence v podnikové praxi s využitím Cloud Computingu*. Praha, 2016. Bakalářská práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra informačních technologií. Vedoucí práce Doc. Ing. Jan Pour, CSc.

<sup>6</sup> NÁHLOVSKÝ, Tomáš. *Business Intelligence v prostředí Cloudu*. Praha, 2016. Diplomová práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra informačních technologií. Vedoucí práce Doc. Ing. Jan Pour, CSc.

# 1 Cloud Computing

Pojem Cloud Computing (CC), zastřešující koncept přístupu k aplikacím, platformám nebo infrastruktuře prostřednictvím internetu ve formě služby, má poměrně zajímavý původ. Název je odvozen od vyobrazení internetu v síťových schématech, kde se běžně kreslí ve formě obláčku (cloud [kloud]). Tato schématická značka internetu vznikla ve fázi, kdy začala být síťová infrastruktura internetu natolik složitá, že ji již nebylo možné vyobrazit, a tak se schéma internetu začalo zjednodušeně znázorňovat ve formě mraku.

Tato kapitola má za úkol nastínit historický vývoj CC, dále rozebrat členění dle distribučních modelů a modelů nasazení, a na závěr porovnat výhody a nevýhody tohoto konceptu.

## 1.1 Historie Cloud Computingu

Hlavní myšlenou CC je přístup k datům a aplikacím kdykoliv a odkudkoliv formou služeb. Jak uvádí Arif ve svém článku pro *ComputerWeekly*, vize Cloudových aplikací tu byla již od vzniku internetu samotného, ale nebyly takto nazývány.<sup>7</sup> Pozvolný vzestup tohoto typu aplikací, využívajících sdílených prostředků, započal teprve až v 90. letech (kdy se zvýšila rychlost internetového připojení natolik, že jej bylo možné pro tento účel využít) avšak v porovnání s dnešními cloudovými službami byly značně primitivní. Mezi prvními aplikacemi, které dnes řadíme do kategorie cloudových, byly emailové schránky. Mezi pionýry na českém internetu byl bezesporu Seznam.cz, který svou službu volně dostupné emailové schránky spustil již v roce 1998 (Hotmail – 1996, Gmail – 2004). Důležité je také dodat, že rychlý vzestup CC by nebyl vůbec možný bez masivního nasazení virtualizačních technologií, které umožnily jednodušší správu běhových prostředí a automatizaci mnoha, dříve ručně prováděných operací.

---

<sup>7</sup> ARIF, Mohamed. A history of cloud computing. In: *ComputerWeekly* [online]. London: ComputerWeekly.com, 2009 [cit. 2016-08-03]. Dostupné z: <http://www.computerweekly.com/feature/A-history-of-cloud-computing>

Prvním velkým milníkem v oblasti aplikací distribuovaných přes internet bylo spuštění CRM systému Salesforce<sup>8</sup> v roce 1999, který začal jako první nabízet profesionální podnikový systém přístupný přes jakýkoliv webový prohlížeč. Tato metoda distribuce softwaru se dnes nazývá Software as a Service (SaaS) – software jako služba. Jednalo se o revoluční krok, neboť do té doby byly pomocí internetu distribuovány pouze jednoduché aplikace, jako je například již zmíněný emailový klient.

Dalším milníkem v historii CC bylo uvedení služby Amazon Web Services (AWS) v roce 2006. Ta dnes poskytuje jak celkovou plně konfigurovatelnou infrastrukturu (IaaS), tak i předpřipravené platformy pro konkrétní použití (PaaS), vše dostupné z prostředí internetu. Ve stejném roce byly taktéž spuštěny do testovacího provozu Google Apps, které dnes nabízejí kompletní virtuální kancelářské prostředí ve webovém prohlížeči. Jednalo se o převratnou novinku, kdy uživatelé mohli začít používat běžné kancelářské nástroje (např. textový editor nebo tabulkový procesor) ve webovém prohlížeči a začít těžit výhody vyplývající z cloudové distribuce. Ostrého provozu se pak tento softwarový počín dočkal až v červenci 2009.<sup>9</sup>

V dalších letech se počet cloudových služeb nadále zvyšoval, neboť dnešní doba, ve které je internetové připojení dostupné takřka všude, tomuto trendu velice přispívá a celý softwarový průmysl se bude tímto směrem dozajista nadále ubírat. Ať se jedná o jednoduchého emailového klienta, pokročilejší Google Docs nebo komplexní vývojářské či podnikové aplikace, cloudové aplikace dnes využívají snad všichni uživatelé, kteří jsou připojení k internetu.

Tomuto značnému technologickému trendu v posledních letech neunikly ani aplikace nebo nástroje typu Business Intelligence a dnes se již běžné používají jejich cloudové verze.

---

<sup>8</sup> <https://www.salesforce.com/>

<sup>9</sup> GLOTZBACH, Matthew. Google Apps is out of beta (yes, really): Insights from Googlers into our products, technology, and the Google culture. In: *Google Official blog* [online]. Mountain View: Google, 2009 [cit. 2016-08-04]. Dostupné z: <https://googleblog.blogspot.cz/2009/07/google-apps-is-out-of-beta-yes-really.html>

## 1.2 Důvod vzniku

Důležitou otázkou je, jaký byl primární důvod vzniku CC. V předešlém textu bylo naznačeno, že myšlenka sdílení výpočetního výkonu tu byla již poměrně dávno, ale k většímu rozšíření došlo teprve v nedávné době.

Hlavním impulsem z podnikové sféry bylo především snižování nákladů. Firmy používající klasické „on-premises“<sup>10</sup> systémy přišly na to, že pořízení, provoz i údržba takových systémů je vysoce nákladná. A v případě, že nedokáží systém plně využít, zcela zbytečně plýtvají svými finančními prostředky. Dalším impulsem, který je ostatně se snižováním nákladů také spjatý, byla nutnost zaměstnávání vlastních IT pracovníků, kteří museli často absolvovat nákladná školení pro udržení znalosti aktuálních technologií, což opět tvořilo nemalou část ve firemních výdajích. Prvním krokem v duchu snižování nákladů tedy byl outsourcing celých IT oddělení. Firmy stále používaly vlastní servery, ale jejich údržba byla v rukou najaté společnosti, což v drtivém počtu případů bylo levnější než v případě údržby vlastními zaměstnanci.

Tento trend byl pouze malý krůček od outsourcingu celých IT infrastruktur, platform či celých softwarových řešení, což při masovém nasazení přineslo další snížení nákladů spojených s IT. Celkovým důsledkem outsourcingu je ve výsledku méně starostí s provozem IT a více času a soustředění na hlavní činnosti, které firmě přinášejí zisk, což je všeobecně považováno za hlavní důvod vzniku CC.<sup>11</sup>

## 1.3 Distribuční modely

Jak již bylo několikrát zmíněno, obecný pojem CC zastřešuje koncept přístupu k výpočetnímu výkonu formou služby. Dané služby ale mohou být různého charakteru, a proto se dělí do tří kategorií dle toho, co je v rámci dané služby nabízeno a jakou úroveň abstrakce nabízejí, respektive od čeho všeho je koncový uživatel odstíněn. Toto rozdělení se

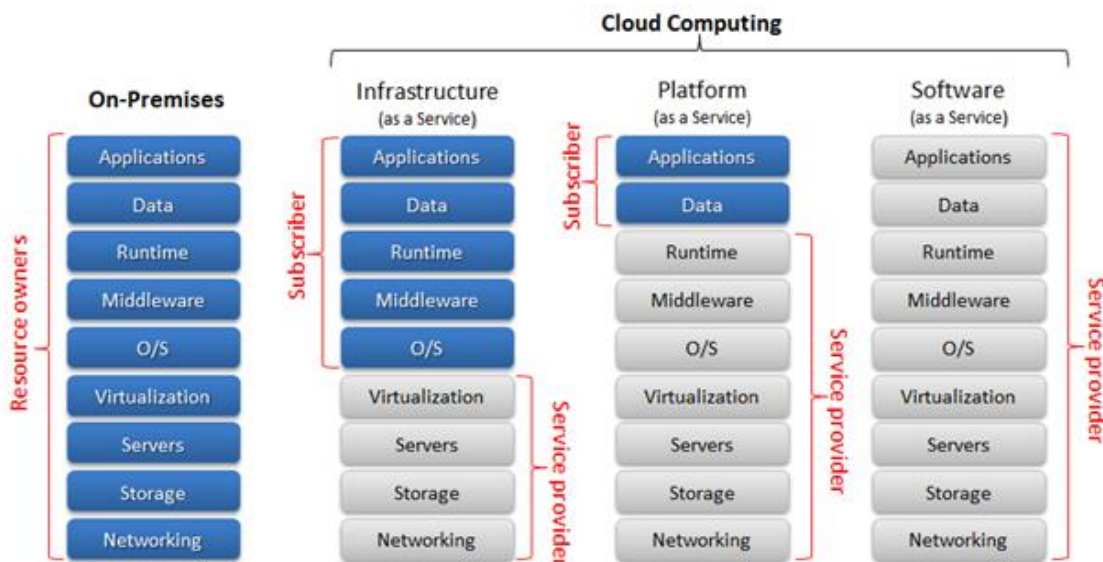
---

<sup>10</sup> Pojem „on-premises“ referuje k software nebo hardware, který je provozován přímo u zákazníka, konkrétně přímo v jeho prostorách/budově.

<sup>11</sup> VOLÉNA, Karel. *Cloud computing?*. Praha, 2013. Bakalářská práce. Bankovní institut vysoká škola Praha, Katedra matematiky, statistiky a informačních technologií. Vedoucí práce Ing. Vladimír Beneš. s. 10-11.



běžně nazývá distribučními modely (distribution models). Obrázek 1-1 graficky znázorňuje porovnání mezi klasickými „on-premises“ systémy a jednotlivými distribučními modely.



Obrázek 1-1: Porovnání distribučních modelů

Zdroj: Service Models. *Cloud Lighthouse* [online]. Kontich: cloudlighthouse.be, 2016 [cit. 2016-10-19]. Dostupné z: <http://cloudlighthouse.be/cloud/service-models>

### 1.3.1 IaaS – infrastruktura jako služba (Infrastructure as a Service)

Nejnižší úroveň abstrakce nabízí model označovaný jako IaaS neboli Infrastructure as a Service (infrastruktura jako služba). U klasických „on-premises“ systémů, neboli u systémů, nad kterými má uživatel (administrátor) plnou kontrolu od fyzického hardwaru až po instalovaný software a které jsou provozovány na půdě vlastnické organizace, jsou počítány jak vysoké pořizovací a provozní náklady, tak i vysoká časová náročnost spojená se zřízením a údržbou. Administrátor takového systému musí věnovat velké množství prostředků nejen na samotné pořízení, ale i na zabezpečení, zálohování, údržby, nemluvě o případné nutnosti rozšíření. Tyto systémy mají běžně velmi nízkou návratnost investice, a pokud nejsou na 100 % využívány, jedná se vždy o plýtvání zdroji.

Model IaaS představuje pro tuto situaci ideální řešení. Zákazník si totiž vždy pronajímá vysoce škálovatelnou platformu dle požadovaných specifikací a všechny operace nutné pro její chod na sebe bere poskytovatel služby. V rámci IaaS je tedy nabízena IT infrastruktura

tvořená konfigurovatelnými virtuálními servery, u kterých si zákazník volí potřebný procesorový výkon, velikost operační paměti a případně i operační systém. Dále si v rámci infrastruktury zákazník může zvolit typ a velikost úložiště dat, případně specifické síťové prvky, firewall, load balancing<sup>12</sup> nebo třeba zálohování. Zvolená konfigurace IT infrastruktury je běžně připravena k použití během několika málo minut od odeslání objednávky, čemuž klasické systémy nemohou absolutně konkurovat. Pokud chce zákazník na pronajaté infrastruktuře spustit vlastní aplikaci, dostane pouze virtuální server s vybraným operačním systémem a zbytek prostředí pro běh musí sám nainstalovat a nakonfigurovat.

Z faktu, že si zákazník pronajímá infrastrukturu formou služby, plynou nesporné výhody, které ostatně ze samotného obchodního modelu nabízení služeb přímo vyplývají. Výhoda využívání IaaS, resp. všech cloudových distribučních modelů, tkví v tom, že při vhodné konfiguraci zákazník platí pouze za to, co skutečně využívá, čímž minimalizuje vynaložené náklady a maximalizuje svůj zisk. Tomuto principu se anglicky nazývá „pay-as-you-go“ neboli „plat' pouze za to, co skutečně využíváš“. Dalším pojmem používaným v souvislosti s cloudovými službami je „on-demand“, neboli „na požádání.“ Tento pojem říká, že pronájem služeb je možný velmi rychle a flexibilně měnit. Ve velmi krátké době parametry infrastruktury zvýšit nebo snížit. Jak již bylo zmíněno, nasazení infrastruktury s požadovanými parametry je otázkou několika minut.

Mezi hlavní hráče na poli IaaS dnes patří Amazon se svými Web Services (AWS)<sup>13</sup>, konkrétně se službou Elastic Compute Cloud (EC2)<sup>14</sup>, Microsoft s Azure<sup>15</sup> nebo také Google Compute Engine<sup>16</sup> či Rackspace<sup>17, 18</sup>.

---

<sup>12</sup> Load balancing je technika pro rozložení zatížení mezi dva nebo více počítačů, síťových linek, procesorů, pevných disků nebo jiných zařízení, aby bylo dosaženo optimálního využití, prostupnosti nebo času odezvy. Použití více zařízení pro vyvážení zátěže může poskytnout i větší odolnost proti výpadkům.

<sup>13</sup> <https://aws.amazon.com/>

<sup>14</sup> <https://aws.amazon.com/ec2/>

<sup>15</sup> <https://azure.microsoft.com/>

<sup>16</sup> <https://cloud.google.com/compute/>

<sup>17</sup> <https://www.rackspace.com/>

<sup>18</sup> Best Infrastructure as a Service (IaaS). *G2 Crowd* [online]. Chicago: G2 Crowd, 2016 [cit. 2016-08-18]. Dostupné z: <https://www.g2crowd.com/categories/infrastructure-as-a-service-iaas>

### 1.3.2 PaaS – platforma jako služba (Platform as a Service)

Na vyšší úrovni abstrakce se nachází PaaS. Ta zaštiťuje abstrakci nad fyzickým hardwarem, operačním systémem, aplikačními mezivrstvami i běhovými knihovnami a poskytuje uživateli předpřipravené prostředí pro běh vlastních aplikací. Běžně jsou nabízeny různé varianty prostředí odvíjející se od programovacího jazyku spouštěných aplikací. Typicky se dnes jedná o PHP, .NET, Java, Ruby, Node.js, Python a další. V rámci poskytování platformy může mít dále zákazník přístup k různým serverovým aplikacím, mezi které lze zařadit například emailový server, různé druhy databázových serverů, úložiště, cachovací nástroje, nástroje pro logování, nástroje pro správu zabezpečení, vyhledávací nástroje nebo datové sklady a analytické nástroje.<sup>19</sup> Vše má za cíl minimalizovat zákaznicko úsilí pro spuštění vlastní aplikace. Poskytovatel v rámci služby také zajišťuje pravidelné aplikování záplat softwarových nástrojů pro zajištění vždy co nejnovějších a nejstabilnějších verzí. Tento distribuční model dále běžně automaticky řeší škálovatelnost platformy při vyšších požadavcích na výpočetní výkon nebo objem dat, zatímco v případě IaaS musí škálovatelnost a rozdělování zátěže (load balancing) řešit zákazník sám.

Jelikož jde o stejný princip využívání služeb jako v případě IaaS, zákazník opět platí pouze za to, co využívá. Cena se jasně odvíjí od velikosti a rychlosti úložiště, rychlosti databázového serveru a objemu dat na něm a samozřejmě od konfigurace samotného aplikačního serveru (procesor, paměť).

Nejvýznamnějšími poskytovateli PaaS dnešní doby jsou obdobní hráči jako v případě IaaS. Opět se jedná o Amazon, Microsoft Azure a dále také Google App Engine<sup>20</sup> nebo Salesforce s platformou Force.com.

### 1.3.3 SaaS – software jako služba (Software as a Service)

Na vrcholu pomyslné pyramidu cloudových distribučních modelů se nachází software jako služba (SaaS). Tento model nabízí nejvyšší úroveň abstrakce a zákazník si pronajímá již ucelený softwarový produkt v podobě služby. Poskytovatel zákazníka odstiňuje od veškeré

---

<sup>19</sup> KAVIS, Michael. *Architecting the cloud*, s. 50-51.

<sup>20</sup> <https://cloud.google.com/appengine/>

infrastruktury, aplikační logiky a všeho ostatního, co je pro chod aplikace zapotřebí. Zákazníci rovnou přistupují k již hotové aplikaci, kde běžně musí učinit pouze základní konfiguraci a spravovat vlastní uživatele, přičemž je služba běžně ihned připravena k provozu.

Mezi nejběžnější SaaS aplikace současnosti patří například již zmíněná emailová schránka (Seznam<sup>21</sup>, Gmail<sup>22</sup>, Outlook.com<sup>23</sup>), video servery (Stream<sup>24</sup>, YouTube<sup>25</sup>, Netflix<sup>26</sup>) nebo kancelářské aplikace v podobě Google Docs<sup>27</sup> nebo Microsoft Office Online<sup>28</sup>. Pro podnikové použití jsou taktéž dostupné aplikace typu ERP nebo CRM a pro vývojářské potřeby pak služby jako je GitHub<sup>29</sup>, Apiary<sup>30</sup>, Pingdom<sup>31</sup> a další. Přístup k SaaS aplikacím je nejčastěji poskytován přes webové rozhraní, ale může být i například ve formě lehkých klientských aplikací či přes pouhé aplikační rozhraní.

## 1.4 Modely nasazení

Předešlá kapitola byla zaměřena na distribuční modely, což bylo první segmentací cloudových služeb. Druhé rozdělení vychází z pohledu úrovně sdílení zdrojů. Zde se uplatňují čtyři skupiny: veřejný, privátní, komunitní a hybridní cloud.

### 1.4.1 Veřejný cloud

Pravděpodobně nejrozšířenější – a také veřejně nejznámější – model nasazení se nazývá veřejný cloud. Tento typ je obecně nejvíce používán, neboť k němu má přístup největší počet uživatelů. Do veřejného cloudu spadají všechny služby, které jsou veřejně přístupné široké veřejnosti, a to buď zdarma nebo za poplatek, běžně ve formě pevné měsíční taxy.

---

<sup>21</sup> <https://www.seznam.cz/>

<sup>22</sup> <https://mail.google.com/>

<sup>23</sup> <https://outlook.live.com/>

<sup>24</sup> <https://www.stream.cz/>

<sup>25</sup> <https://www.youtube.com/>

<sup>26</sup> <https://www.netflix.com/>

<sup>27</sup> <https://docs.google.com/>

<sup>28</sup> <https://www.office.com/>

<sup>29</sup> <https://github.com/>

<sup>30</sup> <https://apiary.io/>

<sup>31</sup> <https://www.pingdom.com/>

Veřejný cloud může být provozován jak soukromou organizací, tak i vládou nebo akademických sektorem.

Mezi největší výhody veřejného cloudu dozajista patří vysoká škálovatelnost, jasná cenová politika typu „pay-as-you-go“ a také vysoká míra samoobslužnosti, což znamená, že si uživatelé drtivou většinu nastavení mohou provádět sami bez jakékoliv pomoci ze strany poskytovatele služby. Klíčovým přínosem je absolutní odstínění uživatelů od správy pronajímané služby. To firmám přináší především méně starostí s dohledem – a tím více času na činnosti, které přinášejí zisk.

Nevýhodou veřejného cloudu je nulová kontrola nad celou službou. Zákazník musí ohledně výkonu a dostupnosti absolutně spoléhat na dodavatele služby. Veřejný cloud také nemusí být vhodný pro uchovávání vysoce citlivých dat, jimiž jsou například čísla kreditních karet, rodná čísla či jiné osobní údaje. Uchovávání takto důvěrných informací je ve velkém počtu států podmíněno legislativou. Ukládání těchto dat mimo dosah jejich majitele či jejich nechtěné úniky jsou trestány velmi vysokými sankcemi. V takových případech může být vhodným řešením tzv. hybridní cloud, v některých případech i možnost speciální certifikace dodavatele služby ve veřejném cloudu.<sup>32</sup>

Příkladem služeb z veřejného cloud mohou být rovněž již zmiňované servery Seznam, Gmail, YouTube, Stream, Hotmail, Microsoft Office Online, Google Docs nebo Facebook.

## **1.4.2 Privátní cloud**

Přesným opakem z hlediska sdílení zdrojů oproti veřejnému cloudu je tzv. privátní cloud. Ten může být do jisté míry připodobněn „on-premises“ systémům, neboť v rámci tohoto modelu nasazení je celá infrastruktura provozována pro výhradní použití jednou organizací nebo více organizačními jednotkami. Definice říká, že systém může být vlastněn, spravován nebo pouze provozován danou organizací, třetí stranou nebo kombinací obou možných a může tedy fungovat jak na principu „on-premises“, tak „off-premises“.

---

<sup>32</sup> KAVIS, Michael. *Architecting the cloud*, s. 53-55.

Z pohledu výhod a nevýhod je na tom privátní cloud opět zcela obráceně, než je tomu ve veřejném cloudu. Výhodou může být vysoká míra přizpůsobení přesným požadavkům daného zákazníka, splnění některých podmínek regulujících uchovávání osobních údajů či skutečnost, že zakoupené zdroje nejsou sdíleny s nikým jiným. Na druhou stranu, zákazník přichází o všechny výhody, které k sobě veřejný cloud váže. Zejména se jedná o vysokou škálovatelnost a flexibilitu a také princip účtování „pay-as-you-go“.<sup>33</sup>

### 1.4.3 Hybridní cloud

Mnoho organizací se snaží získat z použití cloudových technologií maximum a cesta většinou vede přes vhodnou kombinaci služeb z veřejného i privátního cloudu. Jako nejlepší praktika pro použití hybridního přístupu se jeví využití maxima služeb ze sektoru veřejného cloudu, které nabídnou vysokou flexibilitu škálovatelnosti a zajímavou cenovou politiku; přičemž k použití privátního cloudu by se mělo přistoupit pouze v případech, když se jedná o manipulaci s citlivými daty nebo když požadovaný produkt není ve formě veřejné služby k dispozici.<sup>34</sup>

Dle průzkumu trhu s cloudovými technologiemi zpracovaného agenturou Right Scale v roce 2016 se ukazuje, že právě hybridní cloud je velmi podstatnou součástí podnikové infrastruktury. Z 95 % respondentů, kteří odpověděli, že cloudové řešení využívají, odpovědělo 71 %, že jde o hybridní přístup. Čistě veřejný cloud využívá pouze 18 % dotázaných a naopak čistě privátní řešení využívá pouhých šest procent.<sup>35</sup>

### 1.4.4 Komunitní cloud

Některé zdroje dále uvádějí i čtvrtý model nasazení, který je označován jako komunitní cloud. Nejedná se v zásadě o další rozdílný model, ale spíše o jiný přístup v chápání této problematiky. Komunitní cloud se může nacházet v rovině veřejného cloudu i rovině cloudu

---

<sup>33</sup> KAVIS, Michael. *Architecting the cloud*, s. 55-56.

<sup>34</sup> KAVIS, Michael. *Architecting the cloud*, s. 56-57.

<sup>35</sup> WEINS, Kim. Cloud Computing Trends: 2016 State of the Cloud Survey. In: *RIGHT SCALE: CLOUD MANAGEMENT BLOG* [online]. Santa Barbara: RightScale Inc., 2016 [cit. 2016-08-29]. Dostupné z: <http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2016-state-cloud-survey>

privátního. Jde však o to, z jakého úhlu je na věc nahlíženo a pro jakou skupinu uživatelů daná služba slouží.

Společným atributem komunitních cloudů je myšlenka sdílení výpočetních zdrojů a dat mezi uživateli ze stejných zájmových skupin. Z pohledu veřejně-komunitního cloudu se může jednat například o veřejné aplikace provozované státem nebo o herní průmysl, kde je v současnosti vedoucím hráčem na trhu společnost Valve se svou platformou Steam<sup>36</sup>. V případě cloudu se spíše privátním charakterem se může jednat například o speciální bankovní aplikace (ať už jsou určeny přímo pro bankovní klienty nebo pro sdílení dat mezi bankami samotnými) nebo akademické služby.

## 1.5 Shrnutí charakteristik

Z výše uvedeného textu vyplývá hned několik klíčových výhod, které cloudové technologie činí tolik populárními. Závěrečná část této kapitoly je věnována shrnutí hlavních charakteristik a porovnání výhod a nevýhod.<sup>37</sup>

### 1.5.1 Výhody Cloud Computingu

- Výpočetní zdroje datového centra jsou sdíleny více zákazníky, což v důsledku vede ke snížení ceny za pronájem takovýchto zdrojů.
- Nízké bariéry vstupu – cloudové technologie se stávají dostupnými i pro malé podniky.
- Fungují na principu služby – uplatňuje se zde přístup „pay-as-you-go“ neboli „plat’ pouze za to, co skutečně využíváš“.
- Vysoká škálovatelnost – potřebný výkon pro chod služby je buď automaticky navyšován dle počtu zákazníků, nebo si zákazník sám volí konfiguraci dle potřeby.
- Flexibilita – zdroje lze velmi rychle zvýšit či snížit, přesně dle aktuálních požadavků (přístup „on-demand“).
- Samoobslužnost (self-service) – vše je možné konfigurovat přes přehledná grafická rozhraní bez jakékoliv nutnosti zásahu ze strany poskytovatele.

---

<sup>36</sup> <http://store.steampowered.com/>

<sup>37</sup> VELTE, Anthony, Toby VELTE a Robert ELSENPETER *Cloud Computing*, s. 32.

- Platformní nezávislost – díky webové technologii není uživatel výrazně omezen a může ke službě přistupovat z různých typů zařízení (PC, laptop, tablet nebo smartphone).
- Ochrana dat – data bývají v datových centrech replikována a zálohována kvůli zajištění maximální ochrany, která je mnohdy na nesrovnatelně vyšší úrovni, než by byla v případě vlastního řešení.

## 1.5.2 Nevýhody Cloud Computingu

Koncept CC s sebou ale nese pouze samé výhody. Tato technologie má i své stinné stránky a tak není v některých případech zcela vhodná.<sup>38</sup>

- Vysoká závislost na internetovém připojení. Ať se jedná o přístup přes webové rozhraní nebo formou API, vždy jde o komunikaci přes internet. Nastane-li chyba sítě, nebude logicky možné s aplikací pracovat.
- Síťová propustnost – některé vysoce datově náročné aplikace pro svůj rychlý chod potřebují odpovídající rychlost internetového připojení. V případě nesplnění této podmínky se může služba stát nepoužitelnou.
- Závislost na poskytovateli služby. Zákazník je zcela odkázán na důvěru v dodavatele řešení, kvalitu jeho infrastruktury a péči o data. Jestliže nastane na straně poskytovatele výpadek, zákazníci opět nebudou moci službu používat.
- Informační bezpečnost. Tím, že u veřejných cloudových řešení prakticky není možné získat kontrolu nad fyzickým uchováváním dat, není toto řešení vhodné pro ukládání vysoce citlivých informací, které mnohdy bývá regulováno úřady pod výhružkou vysokých pokut.

---

<sup>38</sup> VELTE, Anthony, Toby VELTE a Robert ELSERPETER *Cloud Computing*, s. 32.



## 2 Podpora rozhodování a analýza dat

Absolvování procesu rozhodování je rutinní záležitostí každého člověka. Na manažerské úrovni je ale nezbytné rozhodovat se na základě přesných a relevantních dat, neboť špatná rozhodnutí mohou vést k obrovským ztrátám.

Tato kapitola pojednává o podpoře manažerského rozhodování, jehož nedílnou součástí je analýza podnikových dat jakožto důležitý zdroj podpůrných informací nezbytných pro rozhodování. Kapitola si klade za cíl nejprve nastínit problematiku manažerského rozhodování a následně stručně shrnout historický vývoj oblasti počítačových systémů, která se jeho podpory týká. V závěru bude objasněno, jakým způsobem se systémy pro podporu rozhodování vyvinuly od svých raných počátků až do dnešní podoby.

### 2.1 Proces rozhodování

Nezbytnými předpoklady pro kvalitní rozhodnutí jsou především správné, včasné a hlavně úplné informace. Splnění těchto předpokladů není vždy jednoduchou záležitostí především ve větších firmách, které navíc používají větší množství informačních systémů.

Pro zajištění včasného dodání kvalitních informací, které pomohou řídicím pracovníkům na různých úrovních proces rozhodování usnadnit, slouží různé sady nástrojů souhrnně označované jako Decision Support Systems (DSS) neboli systémy pro podporu rozhodování. Přestože někdy bývá pojem DSS označován konkrétní úroveň v hierarchii podnikových informačních systémů, obecnější definice říká, že jde spíše o zastřešující pojem pro všechny systémy, které jakýmkoliv způsobem přispívají k podpoře rozhodování.<sup>39</sup>

---

<sup>39</sup> POWER, Daniel A Brief History of Decision Support Systems. *DSSResources.COM* [online]. Version 4.1. Iowa: DSSResources, 2007, 2007-03-10 [cit. 2016-08-30]. Dostupné z: <http://dssresources.com/history/dsshistory.html>

## 2.2 Struktura rozhodovacích problémů

Lidé na jednotlivých úrovních podnikového řízení se potýkají s odlišnými typy rozhodovacích problémů. Jelikož k podpoře každého z nich slouží různé systémy, je nejprve nutné definovat, jaké druhy existují. Základní klasifikaci představuje členění na rozhodovací problémy dobře a špatně strukturované.<sup>40</sup>

Dobře strukturované rozhodovací problémy mají tu vlastnost, že se zpravidla řeší opakovaně a existují pro ně rutinní postupy řešení. Pro tyto problémy je charakteristické, že se vyskytují na operativní úrovni řízení, přičemž proměnné, které se v nich vyskytují, lze dobře sledovat. Z toho vyplývá, že celý rozhodovací proces je možné velmi dobře vyhodnotit a dobře strukturované rozhodovací problémy tak zpravidla mívají pouze jediné optimální řešení.

Na druhé straně špatně strukturované rozhodovací problémy bývají zpravidla řešené na vyšších úrovních řízení a svým charakterem jsou do určité míry nové a neopakovatelné. Mívají vyšší počet proměnných, jejichž sledování je obtížnější, a také se u nich mohou objevit nepředvídatelné jevy, což jejich analýzu a vyhodnocení činí mnohem složitější.

Důležité je zdůraznit, že dobře a špatně strukturované problémy představují spíše určitou abstrakci. Jen málo problémů je možné označit výhradně za dobře nebo špatně strukturované. Většina z nich se nachází na pomezí obou těchto typů.

## 2.3 Milníky v historii podpory rozhodování

Historie analýzy podnikových dat pro podporu rozhodování sahá až do konce 60. let minulého století. Již v tomto raném stádiu vývoje počítačových systémů se výzkumní pracovníci snažili přijít s automatizovanými systémy, které by sloužily k podpoře rozhodování a plánování. Zpětným pohledem je dnes možné rozdělit vývoj DSS do několika skupin z pohledu konceptuální orientace. Nejprve se jednalo o modelově a datově orientované systémy (model-driven/data-driven); později na systémy komunikačně

---

<sup>40</sup> DONNELLY, James, James GIBSON a John IVANCEVICH. *Management*. 1. vyd. Praha: Grada, 1997. ISBN 978-80-7169-422-9.

orientované (communication-driven), dokumentově orientované (document-driven) a znalostně orientované (knowledge-driven). V roce 1970 doktor J. D. C. Little definoval čtyři základní kritéria, která musí moderní DSS systémy splňovat. Byla to robustnost, snadnost použití, jednoduchost a úplnost relevantního detailu.<sup>41</sup>

Mezi prvními takovými systémy byly tzv. expertní systémy (Expert Systems). Jejich vývoj započal v 70. letech a byly navrženy tak, aby dokázaly simulovat rozhodovací schopnost srovnatelnou s lidským expertem. Jednalo se o složité počítačové programy zaměřené na řešení komplexních úloh za pomoci zvažování možností z dostupné znalostní báze převzaté od expertů. Z dnešního pohledu se v podstatě jednalo o první formu fungující umělé inteligence.<sup>42</sup>

Jako další se objevily tzv. manažerské informační systémy (MIS). Jsou to datově orientované a účelově zaměřené systémy na generování strukturovaných periodických reportů na základě agregovaných účetních nebo jiných obchodních dat, avšak neumožňují žádnou uživatelskou interakci pro aktivní pomoc při rozhodovacím procesu. MIS patří do úrovně taktického rozhodování, kde slouží k podpoře dobře či částečně strukturovaných rozhodovacích problémů a jeho použití spadá do kompetencí středního až operativního managementu.<sup>43</sup>

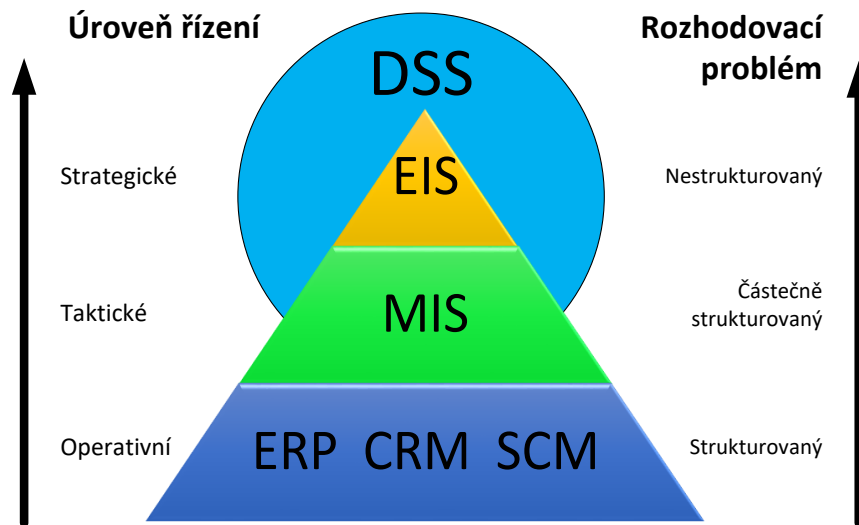
Pro potřeby nejvyššího managementu se objevily pokročilejší systémy EIS (Executive Information System) resp. ESS (Executive Support System), jejichž vývoj započal na začátku 80. let minulého století. Koncept těchto modelově orientovaných systémů byl primárně zaměřen na vrcholové manažery firem, kterým měl umožnit přehledné zobrazení klíčových indikátorů výkonnosti a pomoci tak utvářet kvalitnější strategická rozhodnutí. Systémy EIS se nacházejí na samotném vrcholu pomyslné pyramidy podnikových informačních systémů (viz obrázek 2-1), kde integrují data z různých podnikových systémů a poskytují tak dostatek informací potřebných pro řízení organizace jako celku.

---

<sup>41</sup> POWER, Daniel *A Brief History of Decision Support Systems* [online].

<sup>42</sup> LEONDES, Cornelius. *Expert systems: the technology of knowledge management and decision making for the 21st century*. 1st. San Diego: Academic Press, 2002. s. 1-22. ISBN 01-244-3880-6.

<sup>43</sup> KOPÁČKOVÁ, Hana a Markéta ŠKROBÁČKOVÁ. DECISION SUPPORT SYSTEMS OR BUSINESS INTELLIGENCE: WHAT CAN HELP IN DECISION MAKING?. In: *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration* [online]. Pardubice: University of Pardubice, 2006, s. 98-103 [cit. 2016-08-29]. ISSN 1211-555X. Dostupné z: <http://hdl.handle.net/10195/32436>



Obrázek 2-1: Hierarchie podnikových informačních systémů

Zdroj: vlastní

## 2.4 Business Intelligence

Všechny zmíněné systémy prošly až do konce 20. století značným vývojem. Ten byl ale v 90. letech jistým způsobem narušen příchodem nového trendu s názvem Business Intelligence (BI). Tento pojem byl původně zpopularizován analytikem společnosti Gartner, Howardem Dresnerem, který BI označil jako sjednocující pojem pro širokou škálu programů a řešení pro získávání, konsolidaci, analýzu a poskytování přístupu k datům takovým způsobem, aby umožňovala pracovníkům činit lepší obchodní rozhodnutí. BI znamenalo rychle a jednoduše porozumět velkému množství informací tak, aby bylo na jejich základě možné učinit co nejlepší rozhodnutí.<sup>44</sup>

Jednalo se o revoluční koncept, který zašitoval různé DSS systémy (MIS, EIS) a zpřístupňoval jejich možnosti prostřednictvím integrované platformy všem vrstvám podnikového řízení. Dnešní význam pojmu BI se ve své podstatě stále s Dresnerovou definicí shoduje, a z pohledu celkového vývoje DSS systémů se soudobé BI řadí mezi ty datově orientované.

<sup>44</sup> HEINZE, Justin. History of Business Intelligence. In: *Better Buys* [online]. Malvern: Better Buys, 2014 [cit. 2016-08-30]. Dostupné z: <https://www.betterbuys.com/bi/history-of-business-intelligence/>

Zajímavým paradoxem je fakt, že pojem „Business Intelligence“ je znám již od roku 1958<sup>45</sup> (tedy již před rozvojem prvních DSS), ale k jeho rozšíření došlo až na přelomu 21. století.

Současně s pojmem BI došlo i k rozvoji technologií datových skladů (DW) a OLAP (Online Analytical Processing), které se brzy staly stěžejní součástí BI řešení. Za zrodem datových skladů stál především W. H. Inmon, který DW charakterizoval jako „*subjektivě orientovanou, integrovanou, stálou a časově rozdílnou kolekci dat pro podporu rozhodování manažerů.*“<sup>46</sup>

Nástup datových skladů byl velmi rychlý, neboť díky centralizovanému úložišti umožňoval mnohem efektivnější a rychlejší přístup k datům, než bylo do té doby možné. Především umožnil data konsolidovat a tím poskytnout jednu verzi pravdy napříč celým podnikem. Vzniklý koncept BI složený ze tří základních komponent (ETL – Extract Transform and Load, DW a OLAP) umožnil integraci různorodých podnikových dat do centralizovaného datového skladu a díky datovým kostkám OLAP také rychlý přístup k agregovaným hodnotám. Díky multidimenzionálnímu modelování, jež zkomercializoval R. Kimball, bylo taktéž možné logicky oddělit jednotlivé datové fragmenty spadající pod různá oddělení, čímž byl zajištěn přístup jednotlivých pracovníků pouze k informacím, které byly pro jejich práci potřebné. OLAP taktéž umožňoval interaktivní ad-hoc dotazování, což ocenili především analytici, kteří mohli ověřovat své vlastní hypotézy a nebyli závislí pouze na předpřipravených reportech.

S narůstající známostí pojmu BI na přelomu 21. století se objevily nové firmy nabízející vlastní řešení. Zpočátku BI nástroje umožňovaly pouze přístup k datům prostřednictvím OLAP a základní reporting pro srozumitelnou vizualizaci dat (vše na bázi klient-server).

---

<sup>45</sup> V roce 1958 pojem Business Intelligence definoval vědecký pracovník společnosti IBM, Hans Peter Luhn, jako „*The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal.*“, tedy schopnost vnímat vzájemné vztahy prezentovaných faktů takovým způsobem, aby vedly dílčí kroky za požadovaným cílem.

<sup>46</sup> ZÁDOVÁ, Vladimíra. *Specifika postavení a návrhu datových skladů v rámci IS/ICT: Specifics of Data Warehouses' Status and Design in IS/ICT* [CD-ROM]. Liberec: Technická univerzita v Liberci, 2006. Disertační práce. Technická univerzita v Liberci. Vedoucí práce Prof. Ing. Jan Ehleman, CSc. s 24.

Objevil se také nový směr pro odhalování skrytých informací v datech, označovaný jako Data Mining (DM), čili „dolování dat“. Ten stál na samém vrcholu hierarchie BI komponent a jeho cílem bylo získávat z dat připravených pro analýzu potenciálně užitečné informace pomocí různých pokročilých statistických technik, kterými jsou například rozhodovací stromy nebo regresní či shluková analýza. DM může fungovat jak nad samotným DW, tak nad zpracovanými daty v OLAP kostkách. V případě kombinace DM a OLAP se výsledná fúze nazývá OLAM (Online Analytical Mining).

Koncepční model BI byl velmi slibný, ale jeho používání většinou zůstalo v rukou specializovaných analytiků, kteří managementu připravovali srozumitelné dashboardy a reporty. Základním problémem byl fakt, že BI řešení navrhovali experti pro experty. To činilo pokročilejší analýzy poměrně složitými, což bylo pro běžné pracovníky bez pokročilé znalosti dané problematiky nedosažitelné. Pouze IT pracovníci se znalostí daných analytických systémů a dotazovacích jazyků byli schopni zužítkovat nabízené možnosti pokročilých analýz a z nich vytvářet reporty. Nezbyvalo tedy nic jiného, než přenechat správu BI na IT pracovnících, což mělo bohužel za následek prodloužení komunikačních kanálů a kouzlo rychlých ad-hoc dotazů rázem pominulo. Doručování reportů klíčovými pracovníkům se tak značně zkomplikovalo a prodloužilo, což opět učinilo rozhodování bez včasných informací složitějším. Tato raná éra Business Intelligence se dnes označuje jako BI 1.0.<sup>47</sup>

## 2.5 BI 2.0

S postupem času rostl jak výpočetní výkon počítačů, tak i možnosti ukládání dat, což mělo výrazný vliv na vývoj BI. Současně s rozvojem „Webu 2.0“, jehož rozvoj je dnes datován k roku 2004, vznikaly také nové služby, které se brzy staly novým zdrojem podnikových dat. Právě nové zdroje různorodých dat byly jedním z impulsů na přeformulování požadavků na BI systémy, což vedlo k celkové změně filosofie, jak je na BI nahlíženo a jak je používáno.

---

<sup>47</sup> HEINZE, Justin. *History of Business Intelligence* [online].

Jelikož byly stávající BI nástroje navrženy pouze pro práci se strukturovanými daty, nebyly na nové zdroje semi-strukturovaných dat jednoduše adaptovatelné. Bylo tedy potřeba přijít s novými nástroji, které by nejen reflektovaly nové požadavky, ale také odstranily nedostatky stávajících BI systémů. Druhá vývojová éra Business Intelligence (BI 2.0) nástrojů se nesla v duchu „průzkumník dat“ (data-explorer), což znamenalo vyšší rychlost zpracování, vyšší objem zpracovávaných dat, vyšší flexibilitu datových struktur, širší možnosti analýzy a předpovědí, vyšší míru samoobslužnosti (self-service BI) a především vyšší uživatelský komfort.

Kvůli zcela nové koncepci používání BI nebylo příliš možné využít stávající technologie, proto přišli dodavatelé softwarových řešení s různými nástroji – zejména na cloudových bázích. Cloudové zázemí pomohlo posunout hranice zpracování dat na zcela novou úroveň: z původních předpočítaných OLAP datových kostek se stalo zpracování v reálném čase. To přineslo celkově rychlejší odezvy dotazů i nové možnosti ad-hoc dotazování, v nichž mohou samotní uživatelé vytvářet libovolné metriky, které jim mohou pomoci při objevování informací skrytých v jejich datech. Přejech do cloudového sdíleného prostředí s sebou dále přinesl nové zajímavé cenové politiky, což učinilo BI řešení dostupnější i pro malé a střední podniky.

Nově vzniklé BI platformy začaly nabízet příjemnější uživatelské prostředí přístupné přes webové rozhraní a jejich samoobslužnost se zvýšila natolik, že vytváření metrik a reportů i pokročilejší analýzy nyní po krátkém zaškolení zvládli i běžní pracovníci. Tento krok kupředu podpořil nejen výrazné rozšíření používání BI napříč organizacemi, ale pomohl změnit i uživatelské návyky. Uživatelé již nejsou pouhými konzumenty obsahu (reportů), ale aktivně vytvářejí obsah vlastní a tvoří dotazy na vlastní otázky ve vztahu k obchodním problémům.

Později s rozvojem Webu 2.0 a BI 2.0 se zrodila i nová myšlenka přístupu k budování datových skladů, která měla reagovat na vnější podněty o zvyšujícím se objemu nestrukturovaných dat. Tento přístup s názvem DW 2.0 představil W. H. Inmon v roce 2008

ve své knize *DW 2.0: The architecture for the next generation of data warehousing*<sup>48</sup>. Jeho cílem bylo definovat zcela nový koncept datového skladu, který by umožnil kromě využití strukturovaných dat, využívat i data nestrukturovaná. Takový datový sklad by měl, za podpory klasických i textových ETL, umět strukturovaná i textová data propojit pomocí metadat a doplnit tak jejich kontext. Koncept dále uvádí rozdělení DW na několik sektorů dle stáří dat a dle jejich struktury. To umožňuje data lépe kategorizovat a zároveň oddělit data historická od těch novějších což má zásadní vliv na výkon.

Jedná se ale pouze o konceptuální přístup, nikoliv o konkrétní návrh řešení. Datové skladování se stále vyvíjí a spíše než komplexních řešení se využívá kombinace různých nástrojů pro poskytnutí dílčích částí (funkcí), které dohromady dokáží nastíněné požadavky na moderní DW naplnit.

## 2.6 Mobilní BI (3.0)

Dalším krokem ve vývoji BI, někdy označovaným jako BI 3.0, je nástup mobilního přístupu. Mobilní BI v podstatě pouze doplňuje současná cloudová řešení o další úroveň uživatelského komfortu, který zcela odstraňuje závislost na platformě. Uživatelé tak mohou ke svým datům přistupovat kdykoliv, odkudkoliv a z jakéhokoliv zařízení. Nová generace BI si dále klade za cíl ještě vyšší uživatelský komfort a širší možnosti spolupráce uvnitř organizací.<sup>49</sup> Obrázek 2-2 graficky shrnuje jednotlivé vývojové etapy BI.

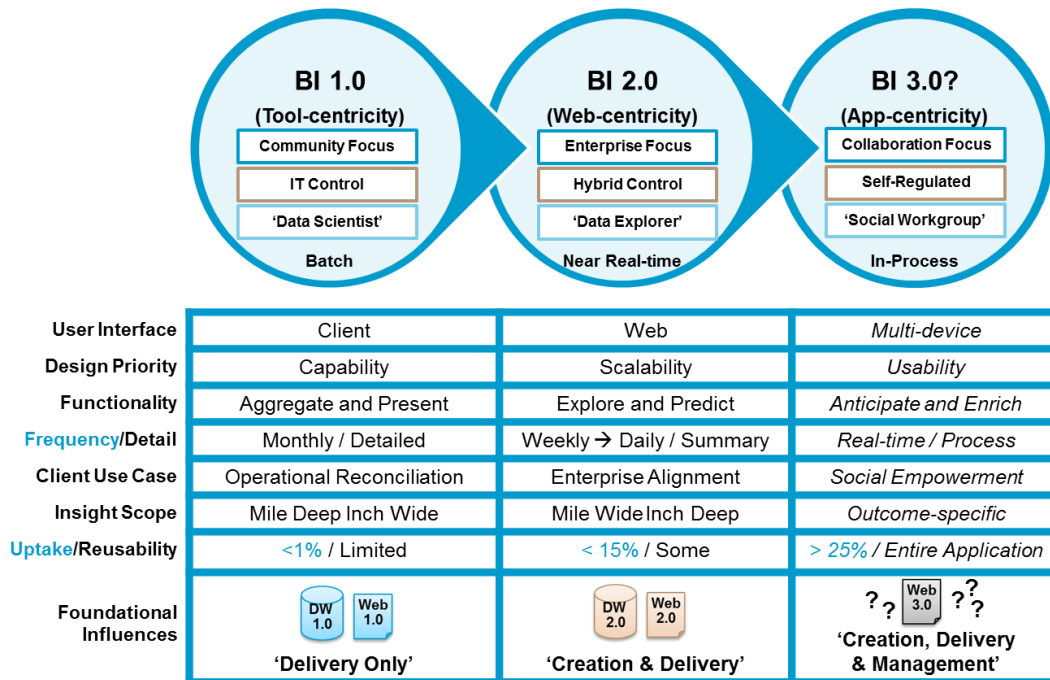
---

<sup>48</sup> INMON, William, Derek. STRAUSS a Genia. NEUSHLOSS. *DW 2.0: The architecture for the next generation of data warehousing*. 1st Edition. Boston: Morgan Kaufmann, 2008. ISBN 978-012-3743-190.

<sup>49</sup> LACHLAN, James. Defining Business Intelligence 3.0. In: *Yellowfin* [online]. Melbourne: Yellowfin International Pty Ltd, 2014 [cit. 2016-10-02]. Dostupné z: <http://www.yellowfinbi.com/YFCCommunityNews-Defining-Business-Intelligence-3-0-159445>



## BI 3.0 – The Journey to Business Intelligence in a Nutshell



Obrázek 2-2: Generační vývoj BI

Zdroj: LACHLAN, James. Defining Business Intelligence 3.0. In: *Yellowfin* [online]. Melbourne: Yellowfin International Pty Ltd, 2014 [cit. 2016-10-02]. Dostupné z: <http://www.yellowfinbi.com/YFCommunityNews-Defining-Business-Intelligence-3-0-159445>

## 3 Trendy v BI

Kromě zmíněných trendů ve formě self-service BI (vyšší samoobslužnosti) a možnosti analýzy dat na jakémkoliv zařízení (mobilní BI), se v současné době ve spojitosti s BI rozvíjejí i další trendy z různých oblastí. Tato kapitola je zaměřena na jejich identifikaci, což tvoří první hlavní cíl práce.

### 3.1 Data-Driven Company

Stále rychleji rozvíjející se trhy nutí obchodníky činit mnohem více rozhodnutí ve stále rychlejším sledu, než jak tomu bylo dříve. A kvalitní rozhodnutí musí být podložena relevantními daty. Tento trend tak tvoří velký tlak na sběr dat o téměř každé činnosti související s podnikáním. Zároveň dochází k celopodnikovému rozšíření BI nástrojů, přičemž přístup k určité části dat by měli mít téměř všichni zaměstnanci. Tento trend se dnes nazývá Data-Driven Company neboli „datově řízená firma“.<sup>50</sup>

Firmy se zaměřují na nový cíl, a to mít co nejširší přehled o tom, co se uvnitř podniku děje. Toho lze dosáhnout sledováním co nejvyššího možného počtu proměnných, jejichž vyhodnocováním lze získat poměrně přesný obraz o fungování vnitrofiremních procesů. Díky širokému záběru sledovaných metrik je možné rychle získat zpětnou vazbu na učiněné kroky a v případě nesprávného vývoje na tuto skutečnost ihned reagovat.<sup>51</sup>

V datově řízených firmách přestávají být BI nástroje pouze výsadou středního a vyššího managementu. Stále více se začínají objevovat i na operativních úrovních řízení a přehled o fungování vlastního oddělení může snadno a rychle získat takřka kterýkoliv pracovník.

---

<sup>50</sup> PEDERSEN, Ulrik. How to Become a Data-Driven Company. In: *TDWI* [online]. Renton: TDWI, 2015 [cit. 2016-12-17]. Dostupné z: <https://tdwi.org/articles/2015/08/04/How-to-Become-a-Data-Driven-Company.aspx>

<sup>51</sup> MITZNER, Dennis. What is a data-driven company?. In: *InfoWorld* [online]. San Francisco: IDG Communications, Inc., 2016 [cit. 2016-12-17]. Dostupné z: <http://www.infoworld.com/article/3074322/big-data/what-is-a-data-driven-company.html>

I v českém prostředí existuje případ, kdy majitel sítě papírnictví McPen<sup>52</sup>, Milan Petr, zpřístupnil obchodní data prodávacům ve všech prodejnách svého obchodního řetězce, což se v relativně krátkém časovém horizontu pozitivně odrazilo na prodejních výsledcích.

*"Mohli jsme vytvořit malý tým analytiků, kteří by každý den připravili pro naše prodejny reporty o aktuálních číslech, ale my se rozhodli, že datového analytika uděláme z každého zaměstnance. A hlavně z našich prodavaček, které mají k datům i zákazníkům nejbliž."*<sup>53</sup>

V citaci je zmíněno, že důvodem k takovému kroku byla domněnka, že pracovníci nejnižších úrovní mají k datům nejbliže. Tato domněnka se posléze ukázala jako více než správná, neboť obecně platí, že ti pracovníci, kteří jsou za vytváření dat zodpovědní (v tomto případě to jsou prodavači, kteří vytvářejí obchodní transakce), také nejlépe znají své zákazníky (obchodní partnery) a jejich zvyky. Proto jim správný pohled na uplynulá data může pomoci svou práci ještě zdokonalit.

Tento trend dnes následují mnohé další firmy, které se snaží svá data zpřístupnit mnohem širšímu poli spolupracovníků, než jsou pouze manažeři. Toto rozšiřování se týká zejména již zmíněných obchodníků a obchodních zástupců, ale výjimkou není například ani personální oddělení. Důležité je pouze to, aby byla ve srozumitelné formě zpřístupněna správná data správným lidem, kterým umožní poučit se ze svých chyb a v budoucnu se zdokonalit.<sup>54</sup>

### 3.2 Uvést data do širšího kontextu

Z hlediska koncepce dnešních BI je hlavním trendem uvedení dat do širšího kontextu, neboli „dát datům větší význam“. S nástupem sociálních sítí a webových analytických nástrojů

---

<sup>52</sup> <http://www.mcpen.cz/>

<sup>53</sup> Sítě Miloše Čermáka: Firma z každého zaměstnance udělala analytika dat. A záhy stouply prodeje. In: *IHNED.cz: Hospodářské noviny* [online]. Praha: IHNED.cz, 2015 [cit. 2016-07-13]. Dostupné z: <http://archiv.ihned.cz/c1-64900010-socialni-site-milose-cermaka-firma-primela-zamestnance-denne-sledovat-data-prodeje-zahy-stouply-prodeje>

<sup>54</sup> MARR, Bernard. Why We Must Rethink Self-Service BI, Analytics And Reporting. *Forbes* [online]. Jersey City: Forbes Media, 2016 [cit. 2016-12-10]. Dostupné z: <http://www.forbes.com/sites/bernardmarr/2016/10/25/why-we-must-rethink-self-service-bi-analytics-and-reporting/#14a3a83d4349>

(např. Google Analytics<sup>55</sup> – GA) se objevily zcela nové zdroje dat a s nimi i nové možnosti v odkrývání nevídaných informací.

Již nejde o pouhé sestavování reportů z izolovaných dat: mnohem větší důraz je kladen na propojenost dat a na jejich vzájemné korelace. Jde o odhalování skrytých souvislostí v souvisejících datech, ale i v datech, která spolu zdánlivě příliš nesouvisí. Propojením dat z účetního systému s daty z GA a například i s daty o počasí je možné zákazníky mnohem lépe segmentovat a pochopit jejich chování, což je možné zpětně zužitkovat formou lepšího cílení na potenciální zákazníky.

Cílem je umožnit firmám co nejlépe porozumět svým datům, poznat své zákazníky a poučit se z těchto znalostí za účelem dosažení lepších a účinnějších strategií. Všechna zmíněná fakta mohou firmám umožnit lepší kontrolu nákladů, což může přinést vyšší zisk nebo v případě vyšších investic vyšší konkurenceschopnost.

### **3.3 Kvalita dat**

Kvalitní rozhodnutí není možné činit bez podpory kvalitních dat. Během procesu shromažďování dat v nich běžně kvůli různým vlivům vznikají chyby, které kvalitu těchto dat snižují a z následné analýzy mohou vyplynout nesprávné informace. Největší vliv na kvalitu dat má dnes jednoznačně lidský faktor. Lidé jednoduše při ručním zadávání dělají různé chyby, ať už se jedná o neznalost nebo prostý překlep. I přes neustálé zdokonalování různých opatření pro minimalizaci chyb se nekonzistence dat objevuje stále a bývá také největším zdrojem problémů v kvalitě reportů.

Tlak na zvyšování kvality (resp. „čistoty“) dat, jenž logicky s počtem sledovaných proměnných stoupá – podpořil vznik specializovaných nástrojů, které pomocí různých technik dokáží data transformovat do standardizované podoby nebo je za podpory webových služeb doplnit o dodatečné informace a zvýšit tak jejich významovou hodnotu.

---

<sup>55</sup> <https://www.google.com/analytics/>

Jedním z nejpůvodnějších nástrojů poslední doby se stal Open source projekt OpenRefine<sup>56</sup> (původně Google Refine).<sup>57</sup> Jedná se o velmi mocný nástroj pro práci s datovými množinami, který pomocí široké palety textových, matematických a jiných funkcí umožňuje velmi efektivně transformovat data do požadované podoby. V datech je možné odstranit duplicity pomocí shlukování, rozdělovat, slučovat nebo standardizovat hodnoty či za pomoci webových služeb rozšířit data o další atributy.<sup>58</sup>

Samotná funkce na shlukování textu tvoří velmi podstatnou část celého systému. Za pomoci kombinace pokročilých algoritmů pro analýzu textu (tokenizační algoritmus n-gram / fonetický algoritmus + Levenshteinova vzdálenost / Kolmogorovova komplexita – PPM) je možné shlukování provádět automatizovaně a poměrně přesně. V praxi se shlukování stává cenným pomocníkem při standardizaci dat, která vznikla například sloučením z různých zdrojů a kde se významově stejné hodnoty objevují vícenásobně s různorodými odchylkami.<sup>59</sup>

Základ OpenRefine tvoří pracovní prostředí srovnatelné s tabulkovými procesory, kde je ale na rozdíl od nich každý krok transformace v rámci projektu zaznamenán. Výsledný postup je možné znovu aplikovat na další datové množiny. Nástroj je velmi oblíbený pro svou schopnost převést částečně strukturovaná data v textové formě do jasně strukturované podoby.

### 3.4 Data Mining

Přestože byl DM v určité formě v rámci BI relativně dostupný již dříve, byl pouze okrajovou záležitostí a byl v organizacích málo využíván. Dnes je ale situace jiná a jeho využití stále

---

<sup>56</sup> <http://openrefine.org/>

<sup>57</sup> *OpenRefine* [online]. OpenRefine, 2016 [cit. 2016-11-13]. Dostupné z: <http://openrefine.org/>

<sup>58</sup> GREGORY, Gwen. Putting Linked Data to Work. *Information Today*. Medford: Information Today, Inc, 2015, 32(7), 20. Dostupné také z: <http://search.proquest.com/docview/1724513127?accountid=17116>

<sup>59</sup> VERBORGH, Ruben, Max DE WILDE a Aniket SAWANT. *Using OpenRefine: The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web*. 1st Edition. Birmingham, England: Packt Publishing, 2013. Community experience distilled. s. 52-54. ISBN 978-1-78328-909-7.

více nabývá na významu, neboť oproti standardním analytickým postupům dokáže z dat „vydolovat“ mnohem více informací.

DM je další součástí BI systémů a přichází na řadu tehdy, když už standardní analytické metody nestačí. Slouží především k odhalování souvisejících jevů, které není jednoduché klasickými postupy identifikovat. Pro tyto účely je využíváno kombinací umělé inteligence, strojového učení a statistických metod. Mezi běžně používané metody lze zařadit například rozhodovací stromy, asociaci, klasifikaci, shlukování nebo prediktivní analýzu, ale i neuronové sítě. DM tak může být velmi užitečný při odhalování anomálií, segmentaci zákazníků nebo predikci časových řad.<sup>60</sup>

I přesto, že DM není zcela novodobým trendem – především kvůli stejné koncepci i stejným metodám, které byly známy již dříve – lze jej přesto do dnešních trendů zařadit, neboť dochází k jeho širšímu využívání. V návaznosti na ostatní trendy současnosti začínají firmy objevovat možnosti, které jim DM může přinést. Samozřejmě obecný trend CC neminul ani tuto oblast, a proto je (stejně jako mnoho jiných služeb) i DM možné využít ve formě cloudové služby. Mezi zástupce takových cloudových služeb se řadí například Microsoft Azure Machine Learning<sup>61</sup>. K většímu rozšíření tohoto segmentu došlo především kvůli jednoduššímu přístupu a jednoduššímu použití. Je tak možné využít hotových řešení bez nutnosti složitého nastavování nebo programování.

Použití předpřipravených modelů a algoritmů, které jsou běžně v rámci DM nástrojů dostupné, nemusí vždy vyhovovat specifickým podmínkám konkrétních trhů či firem. Z tohoto důvodu dnes značně roste poptávka po analytících a programátorech se znalostí jazyků R a Python, které jsou pro matematické analýzy hojně využívány.<sup>62</sup>

---

<sup>60</sup> FÜRNRANZ, Johannes, Dragan GAMBERGER a Nada LAVRAČ. *Foundations of rule learning*. 1st Edition. Heidelberg: Springer, 2012. s. 1-17. ISBN 978-354-0751-977.

<sup>61</sup> <https://azure.microsoft.com/ml>

<sup>62</sup> SMITH, David. Demand for R jobs on the rise, while SAS jobs decline. In: *Revolutions: Daily news about using open source R for big data analysis, predictive modeling, data science, and visualization since 2008* [online]. Redmond: Microsoft Corporation, 2013 [cit. 2016-12-11]. Dostupné z: <http://blog.revolutionanalytics.com/2013/08/job-trends-for-statistics-packages.html>

### 3.5 Nestrukturovaná data

Odhaduje se, že až 90 % dat v organizaci má charakter nestrukturovaných nebo semi-strukturovaných dat. Z tohoto důvodu logicky vzrůstají ze strany majitelů těchto dat požadavky na vývoj nástrojů, které taková data dokáží zpracovat a po zpracování je dokáží nabídnout přímo koncovým uživatelům nebo k dalšímu zpracování.<sup>63</sup> Jelikož nestrukturovaných dat je naprostá většina, očekává se, že v nich jsou ukryta obrovská množství velmi cenných informací, které pouze čekají na odhalení. Není proto divu, že zpracování nestrukturovaných dat je stále aktuálnější a diskutovanějším tématem.

Do nestrukturovaných dat, v nichž se skrývá nejvíce informací, patří především data, která jsou vyprodukována lidmi v běžném životě. Jedná se o emaily, zprávy, poznámky, dokumenty, příspěvky na sociálních sítích, blogy a mnohé další. Všechna tato data jsou nositeli informací o lidském chování, pocitech, problémech atd.: proto je má v souvislosti s firmami a v kontextu BI největší smysl analyzovat. Z analýzy příspěvků ze sociálních sítí či internetových diskusí lze například zjistit veřejné mínění o firmě a v případě negativních ohlasů se na zdroj problémů zaměřit a zjednat nápravu.<sup>64</sup>

Přímé zpracování tohoto druhu dat sice zatím není příliš běžné, ale do praktického využití se dostává čím dál více nástrojů pro převod či extrakci klíčových informací, které je možné dále zpracovávat jako běžná strukturovaná data. Tyto nástroje založené na počítačové lingvistice a umělé inteligenci se souhrnně označují jako programy pro zpracování přirozeného jazyka (Natural Language Processing – NLP).

V souvislosti se zmíněným konceptem DW 2.0, existuje nástroj zvaný Textual ETL™, za jehož zrodem stojí rovněž W. H. Inmon. Jak již název sám napovídá, jedná se o ETL nástroj, jehož cílem je převést data z různých textových zdrojů do strukturovaného formátu.

---

<sup>63</sup> GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi*. 3., aktualizované vydání. Praha: Grada Publishing, 2015. Management v informační společnosti. s. 81. ISBN 978-80-247-5457-4.

<sup>64</sup> POZIN, Ilya. The Data You Ignore Is More Valuable Than You Thought. In: *Forbes* [online]. Jersey City: Forbes Media, 2014 [cit. 2016-12-10]. Dostupné z: <http://www.forbes.com/sites/ilyapozin/2014/06/09/the-data-you-ignore-is-more-valuable-than-you-thought/>

Tato technologie podporuje na vstupu různé formáty textových souborů, emaily, HTML stránky, výstupy ze sociálních sítí a další. Kromě jiného dokáže na výstupu určit sentiment obsahu nebo také sestavit ono zmíněné hodnocení veřejného mínění o firmě.<sup>65</sup>

Nástroj s obdobným zaměřením vyvíjí i česká firma GENEAA Analytics<sup>66</sup>. Jejich platforma je na rozdíl od Textual ETL plně cloudového charakteru a ze vstupních dat dokáže určit například jazyk, sentiment či téma. Zároveň dokáže extrahovat různé typy entit, jakými jsou jména nebo názvy míst či produktů či měřitelné údaje typu datum nebo číslo. Navíc umí kromě anglického jazyka pracovat i s jazykem českým. Analýza sentimentu je zde na velmi vysoké úrovni, je tedy možné poměrně jednoduše a kvalitně odhalit ve zdrojových datech pozitivní či negativní ohlasy.<sup>67</sup>

Zmíněné nástroje jsou ale stále spíše okrajovou a velmi specifickou záležitostí. Pro některé jednodušší operace, například přiřazování k článkům určitá předdefinovaná klíčová slova na základě shody s obsahem, lze použít značně rozšířenější Open Source projekt Elasticsearch<sup>68</sup>. Jedná se o fulltextový vyhledávač, který sice není pro tyto účely přímo navržen, ale některých jeho vlastností lze tímto způsobem využít.<sup>69</sup>

### 3.6 Semi-strukturovaná data

S příchodem nových datových zdrojů s nestálou strukturou dat bylo i s ohledem na datové skladování a zpracování potřeba na tyto požadavky reagovat. Tato data mohou běžně pocházet z cloudových služeb s velmi rychlým vývojem nebo ze strojů ve formě různých protokolů. Za nejběžnější formáty takových dat je možné označit třeba JSON nebo XML.<sup>70</sup>

---

<sup>65</sup> TextualETL™ Technology. *FOREST RIM TECHNOLOGY* [online]. Castle Rock: FOREST RIM TECHNOLOGY, INC., 2016 [cit. 2016-12-10]. Dostupné z: <http://www.forestrimtech.com/textual-etl/textual-etl-technology>

<sup>66</sup> <https://geneea.com/>

<sup>67</sup> *Geneea* [online]. Praha: GENEAA Analytics s.r.o, 2016 [cit. 2016-12-10]. Dostupné z: <https://geneea.com/>

<sup>68</sup> <https://www.elastic.co/>

<sup>69</sup> Elasticsearch. *Elastic* [online]. Amsterdam: Elasticsearch BV, 2016 [cit. 2016-12-10]. Dostupné z: <https://www.elastic.co/products/elasticsearch>

<sup>70</sup> HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. *Big Data a NoSQL databáze*. 1. vydání. Praha: Grada, 2015. s. 24-46. ISBN 978-80-247-5466-6.



Tradiční relační databáze ale nebyly na uchovávání takových dat nebo na časté změny schématu vůbec stavěné. Možností bylo převést data na pevnou strukturu, což ale většinou mělo za následek ztrátu souvisejících informací a zpomalení celého procesu zpracování, nebo uložení jako celého souboru, se kterým se později nedalo příliš pracovat.

Proto vznikly nové databázové nástroje, které si s flexibilním schématem dokáží poradit – často souhrnně označované jako NoSQL (Not Only SQL), tedy „nejen SQL“. Jako jejich zástupce je možné jmenovat například MongoDB<sup>71</sup>, Couchbase<sup>72</sup> nebo Snowflake<sup>73</sup>. Tyto databáze podporují nativní ukládání záznamů ve formátu JSON i jejich pohodlné zpracování.

Díky novým možnostem ukládání dat je možné flexibilněji reagovat na aktuální potřeby či požadavky trhu a nabídnout tak zákazníkům nové funkce rychleji, bez nutnosti složité změny struktury relační databáze nebo obav o ztrátu nebo nevyužitelnost uložených dat.

### 3.7 Big Data

Dalším trendem současné doby je růst objemu dat, který v posledních letech dosahuje téměř exponenciálního tempa a jeho pokračování je zcela jisté. V roce 2011 vydala americká organizace IDC<sup>74</sup> prognózu, dle které se má celosvětový objem dat do roku 2020 zvýšit na 50násobek<sup>75</sup>. Dle článku z roku 2013 je odhadováno, že v době vydání bylo 90 % všech dat vygenerováno v předcházejících dvou letech.<sup>76</sup>

Trend neustálého zvyšování objemu dat, který byl nastartován zrodem sociálních sítí, video-streamingových služeb, mobilních zařízení a internetu věcí (IOT), je dnes označován jako „BIG DATA“. K tomuto trendu přispívá též širší využívání různých typů senzorů

---

<sup>71</sup> <https://www.mongodb.com/>

<sup>72</sup> <http://www.couchbase.com/>

<sup>73</sup> <https://www.snowflake.net/>

<sup>74</sup> <https://www.idc.com/>

<sup>75</sup> MEARIAN, Lucas. World's data will grow by 50X in next decade, IDC study predicts. In: *COMPUTERWORLD* [online]. Framingham: Computerworld, Inc, 2011 [cit. 2016-10-03]. Dostupné z: <http://www.computerworld.com/article/2509588/data-center/world-s-data-will-grow-by-50x-in-next-decade--idc-study-predicts.html>

<sup>76</sup> Big Data, for better or worse: 90% of world's data generated over last two years. In: *Science Daily: Your source for the latest research news* [online]. Rockville: ScienceDaily, 2013 [cit. 2016-12-10]. Dostupné z: <https://www.sciencedaily.com/releases/2013/05/130522085217.htm>

a snímačů, které soustavně zaznamenávají nejrůznější informace. Všechny tyto systémy se staly zcela novým zdrojem dat, u kterých logicky vznikl i tlak na jejich zpracování.

Stále více se do analýz zahrnují data z různých nástrojů pro analytiku webu a chování zákazníků, které mnohdy generují obrovské množství dat. Analýza podnikových účetních dat je dnes pouhým začátkem na cestě za poznáním. Vyšší objemy dat ale znamenají ještě vyšší potřebný výkon pro jejich zpracování a pokročilejší analytické metody pro odhalování souvislostí.

Společnost Gartner<sup>77</sup> definovala základní charakteristiku Big Dat pomocí tzv. „3V“. Jedná se o „Volume“ (velikost) – objem dat je běžně větší než 10 TB a roste exponenciálně, „Velocity“ (rychlost nárůstu) – rychlý přírůstek dat, který je generován systémy nebo senzory neustále, a „Variety“ (různorodost) – nestálá nebo neexistující datová struktura.<sup>78</sup> Počet „V“ bývá někdy rozšiřován o „Veracity“ – (věrohodnost), „Value“ (hodnotu), „Validity“ (platnost) a „Volatility“ (krátkou dobu využitelnosti).<sup>79</sup> Společným atributem je fakt, že se jedná o data, která nelze jednoduše v rozumném čase zpracovat pomocí doposud známých a ověřených technologií.

Je důležité položit si otázku: „Jaký je vlastně vztah Big Dat k dnešnímu BI?“ Názory se v tomto ohledu různí, ale za pravděpodobně nejlogičtější a také nejjednodušší vysvětlení lze považovat tvrzení, že BI představuje sadu nástrojů, které pomáhají v Big Datech odhalovat skryté informace a ty interpretovat uživatelům. Z toho vyplývá, že dnešní cloudové BI platformy by měly být na zpracování Big Dat připraveny.<sup>80</sup>

---

<sup>77</sup> <https://www.gartner.com/>

<sup>78</sup> BEYER, Mark A. a Douglas LANEY. The Importance of 'Big Data': A Definition. *Gartner* [online]. Stamford: Gartner, Inc, 2012 [cit. 2016-12-10]. Dostupné z: <https://www.gartner.com/doc/2057415/importance-big-data-definition>

<sup>79</sup> NORMANDEAU, Kevin. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity. In: *InsideBIGDATA* [online]. Portland: InsideBIGDATA, 2013 [cit. 2016-12-10]. Dostupné z: <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>

<sup>80</sup> Big Data Vs. Business Intelligence: What's the Difference?. In: *Yurbi: bring your data to live* [online]. Henderson: 5000fish, Inc, 2016 [cit. 2016-12-10]. Dostupné z: <https://www.yurbi.com/blog/big-data-vs-business-intelligence-whats-the-difference/>

Situace na trhu je taková, že některá řešení jsou na zpracování Big Dat připravena více a jiná méně. Rozvoj technologií umožňující jejich základní skladování a nízko-úrovňové zpracování je stále na vzestupu. Vznikly nové technologie s novými možnostmi distribuovaného ukládání a paralelního zpracování dat, přičemž jedním z nejdiskutovanějších zástupců je Open Source projekt Hadoop<sup>81</sup> a jeho nadstavby.<sup>82</sup> Uvedení těchto technologií do praxe umožnilo posunout zpracování dat na zcela novou úroveň, kde se z analýzy pouze malého vzorku dat mohlo stát zkoumání celého souboru, což logicky přináší mnohem přesnější výsledky.

Téma Big Dat je sice stále velmi čerstvé, neuchopené a konkrétně nepopsané, ale jisté je, že je jednoznačným budoucím trendem, který se nikdy nezmění. Lidé budou generovat stále více dat, budou vznikat nové zdroje, firmy budou o svých zákaznících chtít vědět stále více informací a zpracování Big Dat bude jedinou cestou, jak tyto informace odhalit.

---

<sup>81</sup> <http://hadoop.apache.org/>

<sup>82</sup> DOLÁK, Ondřej. Big data: Nové způsoby zpracování a analýzy velkých objemů dat. In: *System On Line* [online]. Praha: CCB, spol. s r. o., 2011 [cit. 2016-12-11]. Dostupné z: <https://www.systemonline.cz/clanky/big-data.htm>

## 4 BI v cloudu

V návaznosti na generační vývoj BI systémů, nastíněný v kapitole 2, je pro úplné pochopení současného stavu na poli cloudových platform BI nezbytné osvětlení současného přístupu. Nástup cloudové éry se dotknul takřka všech oblastí podnikových informačních systémů. Oproti dříve používaným „on-premises“ systémům lze dnes provozovat veškerou firemní infrastrukturu a software v cloudu. V následujících částech této kapitoly bude rozebráno, jakým způsobem se dílčí části klasických BI řešení, které v základní formě zahrnují zdroje dat, datové transformace (ETL), datové sklady (DW) a vizualizaci či reporting, transformovaly do dnešní podoby a jaké výhody přinášejí.

Jak zmiňuje Náhlovský ve své diplomové práci, architektury cloudových BI řešení mohou nabývat tří podob. První variantou je, že se v cloudu nachází pouze prostředí pro vizualizaci a reporting; vrstva DW a ETL se nachází stále na půdě organizace. Druhá možnost obnáší přesun do cloudu DW i vizualizace, ale organizace pomocí „on-premises“ systému stále řeší ETL proces. Poslední kombinací je pak plně cloudový model, v němž jsou všechny komponenty BI přesunuty mimo organizaci.<sup>83</sup>

Přestože je možné provozovat jednotlivé komponenty BI systémů i v rámci privátního či hybridního cloudu, pro účely této práce bude zaměřeno pouze na ty poskytované jako veřejné služby. Konkrétně bude zacíleno na ty, které jsou svým charakterem nabízeny jako PaaS nebo SaaS.

### 4.1 Zdroje dat

Základním předpokladem budování všech BI řešení je potřeba analyzovat podniková data. Není proto od věci krátce shrnout, odkud dnes mohou data pocházet. Zdroje dat se mohou různit jak svým charakterem, tak i umístěním, ale v tomto případě bude řeč o zdrojích dat z pohledu obsahu a významu, nikoliv konkrétního umístění či formátu.

---

<sup>83</sup> NÁHLOVSKÝ, Tomáš. *Business Intelligence v prostředí Cloudu*, s. 51-54.

Mezi klasické zdroje patří především podnikové informační systémy, jejichž zástupci jsou například ERP, CRM, SCM apod. Mohou existovat jak ve formě „on-premises“, tak v cloudu. Tyto zdroje obsahují velmi dobře strukturovaná data a vyznačují se tím, že uchovávají největší množství dat o obchodních transakcích. Jejich podstata je proto zcela očividná. Tato data stále tvoří velmi podstatnou část analýzy a v některých organizacích jsou jediným podkladem pro rozhodování.

Ve firemním prostředí existují i další zdroje dat, většinou nestrukturovaných. Jedná se o různé dokumenty, smlouvy nebo emaily. Přestože mohou tyto zdroje obsahovat velké množství cenných informací, jak bylo dříve zmíněno, jejich analýza je stále poměrně složitá.

S rozvojem internetu, cloudu a webových služeb vznikly i v tomto segmentu nové zdroje dat. Za webové (cloudové) zdroje je možné považovat všechny, které jsou generovány nějakými webovými službami. Například měření návštěvnosti webu pomocí GA, statistiky z Facebooku, reklamní kampaně Google AdWords<sup>84</sup>, Sklik<sup>85</sup>, atd. Když se obohatí strohé účetní záznamy například o detailní statistiky reklamních kampaní, firmám to umožní nové úhly pohledu a tím i jednodušší a efektivnější vyhodnocování své činnosti.

Doplnění vlastních dat (a jejich uvedení do širšího kontextu) je dále možné pomocí různých typů externích dat, která mohou být buď shromažďována různými agenturami a zpřístupněna za poplatek nebo jsou veřejně dostupná. Příkladem takových zdrojů mohou být historická data o počasí z OpenWeatherMap<sup>86</sup> nebo o směnných kurzech z Open Exchange Rates<sup>87</sup>.

V neposlední řadě je (s ohledem na současný rozmach mobilních technologií) nezbytné zmínit velmi podstatný zdroj dat, kterým jsou mobilní zařízení. Firmy, které vlastní mobilní aplikace, jichž využívají jejich zákazníci, mohou sledovat údaje o pohybu těchto zákazníků a lépe například cílit nabízený obsah nebo reklamu – stejným způsobem, jakým to dnes praktikují mnohé firmy jako Facebook nebo Google. Bohužel etická stránka této věci je diskutabilní.

---

<sup>84</sup> <https://www.google.cz/adwords/>

<sup>85</sup> <https://www.sklik.cz/>

<sup>86</sup> <https://openweathermap.org>

<sup>87</sup> <https://openexchangerates.org/>

Vyjmenovat všechny možné zdroje prakticky není možné. V práci je proto uveden pouze výtah těch nejdůležitějších, které dnes mohou firmám pomoci činit lepší obchodní rozhodnutí. Všechny zmíněné zdroje mohou pocházet z rozdílných umístění, ať už se jedná o vnitřní prostředí nebo cloud. Data mohou být samozřejmě přístupná v odlišných formátech nebo prostřednictvím různých protokolů.

## 4.2 Datové transformace

Vytěžit maximum informací ze všech dostupných zdrojů nelze bez jejich vzájemné integrace. ETL nástroje umožňují sloučit data z různých zdrojů dohromady, aby nad nimi později mohly být prováděny analýzy – jsou tak vstupní branou do BI systémů.

Z pohledu své podstaty se ETL nástroje příliš nezměnily, ale jejich přesun do cloudového prostředí přinesl několik výhod. Výhodou běhu ETL procesu v cloudu je především to, že není potřeba zatěžovat vlastní infrastrukturu či pořizovat infrastrukturu novou.

Dnešní ETL nástroje lze obecně rozdělit do dvou kategorií. Ty, které jsou s plně grafickým uživatelským rozhraním, nabízejí transformace dostupné ve formě předpřipravených komponent, jejichž zřetězením lze dosáhnout požadovaného výsledku. Druhou kategorií tvoří ETL nástroje, v nichž jsou transformace založeny na určitém programovacím jazyku (SQL, R, Python). Mezi zástupce první kategorie lze zařadit například nástroje Informatica<sup>88</sup>, SnapLogic<sup>89</sup>, CloverETL<sup>90</sup> nebo Microsoft Azure Data Factory<sup>91</sup>. Z druhé skupiny je možné zmínit například český projekt Keboola Connection<sup>92</sup>. Grafické ETL nástroje jsou zpravidla jednodušší na použití, ale nástroje skriptové nabízejí mnohem větší volnost a uživatel není při jejich využívání závislý na připravených komponentách.

Pokud firma zvažuje zavedení nového cloudového BI řešení, není vždy nutné samostatný ETL nástroj pořizovat. Většina BI platform totiž nástroj na integraci dat již v rámci celého

---

<sup>88</sup> <https://www.informatica.com/>

<sup>89</sup> <https://www.snaplogic.com/>

<sup>90</sup> <https://www.cloveretl.com/>

<sup>91</sup> <https://azure.microsoft.com/en-us/services/data-factory/>

<sup>92</sup> <https://keboola.com/>

řešení nabízí. Tyto nástroje se ale svými schopnostmi mohou výrazně lišit – proto záleží na potřebách konkrétní firmy, zda je daný nástroj pro nutné operace dostačující nebo zda je nutné použít jiný.

### 4.3 Ukládání dat

Segment datového skladování byl již částečně zmíněn v předešlé kapitole, konkrétně v části zabývající se semi-strukturovanými daty. Pro potřeby doplnění kontextu je třeba uvést, čím konkrétně se dnešní DW vyznačují.

Společnými atributy moderních cloudových DW jsou téměř neomezená kapacita (řádově TB až PB) a oddělení datové a výpočetní vrstvy. Obě části jsou na sobě zcela nezávislé. Uživatel platí samostatně za uložení dat a jejich zpracování. Tento koncept umožňuje obě složky nezávisle škálovat, tudíž lze jejich parametry velmi rychle měnit – v případě potřeby je možné jednorázově několikanásobně zvýšit výpočetní výkon pro specifický dotaz a po dokončení jej opět snížit na standardní hodnotu. Díky dostatečně vysokému výpočetnímu výkonu v datových skladech není pro reporting nutné data předem nijak agregovat, proto se dnes již OLAP datové kostky příliš nepoužívají.

Za lídry trhu v tomto segmentu lze jednoznačně označit Snowflake<sup>93</sup> Google BigQuery<sup>94</sup> a Microsoft Azure SQL Data Warehouse<sup>95</sup>. Na rozdíl od Snowflake a BigQuery si řešení Microsoftu sice neporadí se semi-strukturovanými daty, ale v ohledu škálovatelnosti jsou na tom velmi podobně.

Důležité je zmínit, že do DW by měla přijít pouze data nezbytná pro budoucí analýzu, ačkoliv díky oddělené datové vrstvě a velmi nízkým cenám za uložená data (od cca 20 USD/TB/měsíc) je toto doporučení možné opomenout.

---

<sup>93</sup> <https://www.snowflake.net/>

<sup>94</sup> <https://cloud.google.com/bigquery/>

<sup>95</sup> <https://azure.microsoft.com/services/sql-data-warehouse/>

Kromě toho, že DW v konceptu cloudových BI figurují jako zdroj dat a výpočetního výkonu pro analýzu, existuje i další forma jejich využití. Pro jejich vysoký výkon a škálovatelnost jsou cloudové DW hojně využívány jako dočasné úložiště (staging area) pro ETL procesy. On-demand přístup k tomuto využití přímo nabízí.

## 4.4 Vizualizace dat

Vizualizační vrstvu lze dnes pořídit v zásadě ve dvou variantách. Buď se jedná o pouhý zobrazovač dat z DW nebo o nástroj s vlastním úložištěm a analytickým aparátem.

Oba přístupy mají své výhody i nevýhody. Při přístupu k externímu datovému skladu odpadá nutnost dalšího přenosu dat do analytického nástroje a zákazník má plnou kontrolu nad náklady spojenými s provozem DW. V případě dostatečného výkonu může být toto řešení i rychlejší. Na druhou stranu bývají omezeny možnosti datového modelování a někdy je nutné metriky vypočítat již v datovém skladu. Možnost přímého připojení k DW nabízí například nástroje Tableau<sup>96</sup>, QlikView<sup>97</sup>, BIME<sup>98</sup> nebo Microsoft Power BI<sup>99</sup>.

Přístup, kde je analytický nástroj plně integrovanou platformou s vlastním úložištěm, většinou nabízí přesný opak. Data je vždy nutné přenést do prostředí dané platformy (buď z vlastního DW nebo pomocí integrovaného ETL nástroje). Datové modelování bývá více flexibilní a umožňuje definování metrik za běhu z kombinace různých zdrojů. Výkon bývá vždy na prvním místě a uživatel tak platí pouze za fungování platformy jako celku. Krom několika výjimek nabízejí většinou dodavatelé cloudových BI obě možnosti přístupu k datům. Striktními zástupci této filozofie jsou ale například GoodData<sup>100</sup>, DOMO<sup>101</sup> nebo Sisense<sup>102</sup>.

---

<sup>96</sup> <https://www.tableau.com/>

<sup>97</sup> <https://www.qlik.com/>

<sup>98</sup> <https://www.bimeanalytics.com/>

<sup>99</sup> <https://powerbi.microsoft.com/>

<sup>100</sup> <https://www.gooddata.com/>

<sup>101</sup> <https://www.domo.com/>

<sup>102</sup> <https://www.sisense.com>



## 4.5 Shrnutí

Plně integrované platformy sice nabízejí komplexní systém se spoustou funkcí, ale kvůli omezením, která z distribučních modelů vyplývají (především nemožnost úpravy na míru), nemusí všechny části systému firmě vždy zcela vyhovovat. Cesta tak mnohdy vede přes použití různých nástrojů, přičemž každý z nich vyniká v jiných oblastech. Používání více systémů pro specifické účely také přispívá samotný koncept distribuce cloudových služeb. Pořízení je velmi rychlé, většinou i relativně levné, a v případě nespokojenosti lze od služby ve velmi krátkém čase odstoupit.

Zároveň díky rychlému spuštění projektu se dnes nejvíce využívá inkrementální přístup budování. Velká komplexní řešení, která by měla ihned integrovat všechna podniková data a řešit všechny obchodní otázky bývají vždy velmi složitá na analýzu i implementaci a také většinou nákladná. Existuje tak velké riziko, že takový projekt kvůli své složitosti ztroskotá. Je tak při budování mnohem větší snaha činit malé přírůstky, které řeší konkrétní problém.

## 5 Srovnání vybraných cloudových BI

Na současném trhu s cloudovými BI platformami operují desítky firem nabízejících produkty odpovídající jejich představám o Business Intelligence. I přes stejné obecné principy fungování či nabízenou (hláсанou) funkcionalitu – což je dnes možné souhrnně označit jako možnost jednoduše analyzovat data pomocí interaktivního grafického webového rozhraní – se skutečné vlastnosti jednotlivých systémů značně liší.

Mezi komerčně nabízenými cloudovými BI nástroji existují značné rozdíly, které však nemusí být na první pohled zřejmé. Jejich identifikace je proto obtížná a rozdíly se mnohdy projeví až po delší době používání. Pro podniky to znamená obtížnost volby ve výběru správného BI. Tato kapitola je zaměřena na naplnění druhého hlavního cíle práce, kterým je poukázat na značnou rozdílnost mezi komerčně nabízenými cloudovými BI platformami. Cíle je dosaženo srovnáním dvou vybraných platforem, u nichž je poukázáno na největší rozdíly.

### 5.1 Volba nástrojů

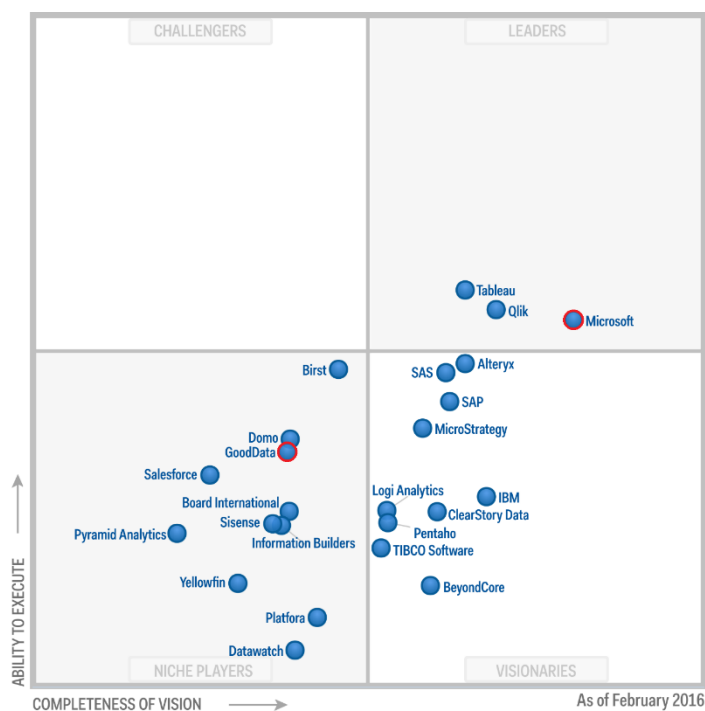
Pro kvalitní srovnání, ze kterého vyplynou zjevné rozdíly, je volba použitých BI nástrojů stěžejní. Při výběru byl kladen velký důraz na to, aby si byly zvolené platformy co možná nejvzdálenější a rozdíly se tak co nejvíce projevíly. K podpoře volby bylo použito hodnocení Gartner Magic Quadrant<sup>103</sup> z února roku 2016 (viz obrázek 5-1). Gartner dnes představuje přední výzkumnou organizaci zabývající se zkoumáním trhu s IS/ICT technologiemi.

Cílem výběru bylo zvolit z 24 hodnocených BI analytických platforem dva čistě cloudové nástroje z protichůdných kvadrantů. Jasně bylo jako první zacíleno na kvadrant „Leaders“, ve kterém se nacházejí výrobci s vedoucím postavením na trhu. Jako protichůdný kvadrant byl následně zvolen „Niche Players“, čili okrajoví hráči. Další rozhodování o volbě konkrétních produktů bylo výrazně zjednodušeno tím, že většina oslovených dodavatelů neumožňuje zkušební verze svých produktů. Mnozí dodavatelé také nebyli ochotni uvolnit

---

<sup>103</sup> Magic Quadrant for Business Intelligence and Analytics Platforms. *Gartner* [online]. Stamford: Gartner, Inc., 2016 [cit. 2016-11-23]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>

přístup ani pro akademické účely, jako je tento. Z oslovených i nabízených firem tak vzešla dvě řešení, která byla možná otestovat v režimu zkušební verze. V následujícím porovnání jsou zastoupeny platformy GoodData a Microsoft Power BI.



Obrázek 5-1: Gartner Magic Quadrant – únor 2016, s vyznačenými zvolenými nástroji

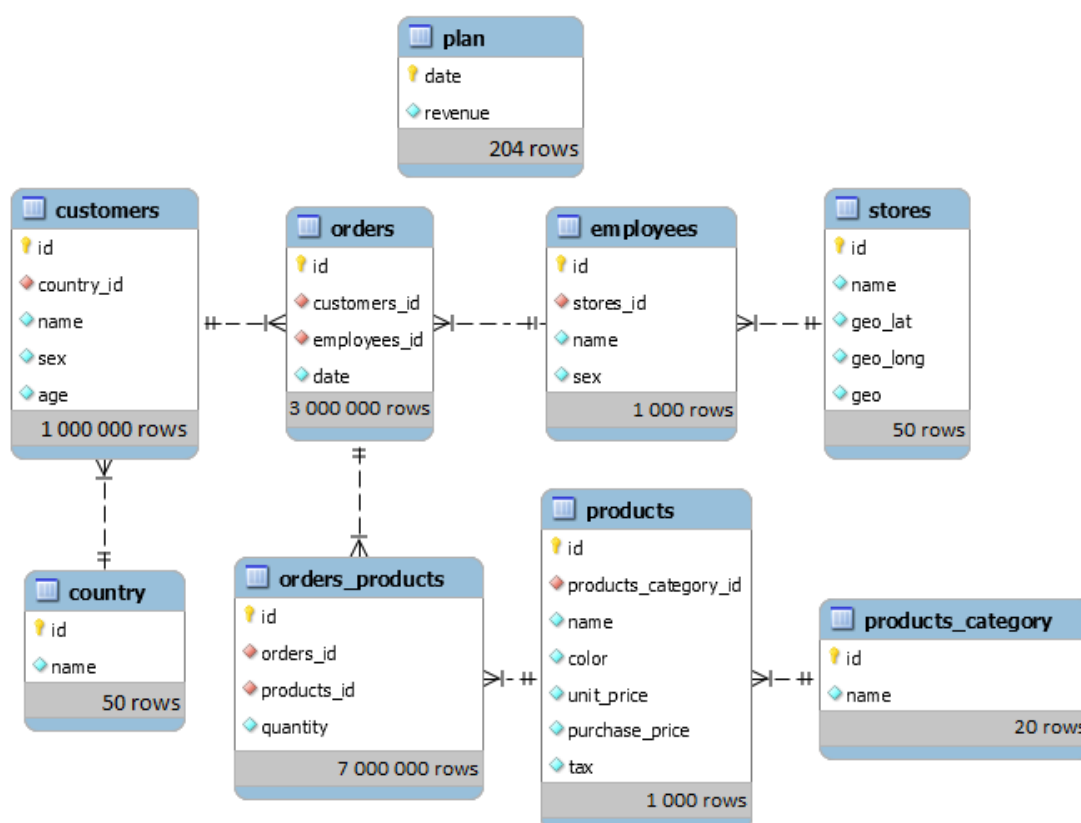
Zdroj: Magic Quadrant for Business Intelligence and Analytics Platforms. *Gartner* [online]. Stamford: Gartner, Inc., 2016 [cit. 2016-11-23]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>

## 5.2 Modelová situace

Pro testování systémových možností jednotlivých platform byl vytvořen jednoduchý datový model (viz obrázek 5-2) reprezentující fiktivní síť maloobchodních prodejen. Rámec modelu je zaměřen na klasickou analýzu obchodní činnosti jakožto jedné z nejdůležitějších částí analýzy podnikových dat.

Základ modelu tvoří objednávky (*orders*) specifikované datem objednání, zákazníkem a prodejcem. Pro každou objednávku existuje vazba na seznam objednaných produktů (*orders\_products*) obsahující objednané množství. Každý produkt (*products*) obsahuje různé atributy, jimiž jsou informace o nákupní a prodejní ceně, velikosti daně a také kategorie

produktů, do nichž produkty spadají (*products\_category*). Zaměstnanci (*employees*) obsahující kromě základních atributů i odkaz na prodejnu (*stores*), ve které pracují. Prodejny jsou determinována názvem a polohou ve formě zeměpisné šířky (*geo\_lat*) a délky (*geo\_long*). Podobně též zákazníci (*customers*) odkazují na zemi, ze které pocházejí (*country*). Pro jednodušší manipulaci obsahuje tabulka *country* pouze seznam 50 států USA. Pro porovnání dosažených obchodních výsledků s výsledky plánovanými figuruje v modelu tabulka *plan* obsahující měsíční plánované obraty.



Obrázek 5-2: Datový model pro testování cloudových BI nástrojů

Zdroj: vlastní

Na základě nastíněného datového modelu byla pro jednotlivé entity vygenerována náhodná data v předem definovaném objemu a s časovým rámcem let 2000-2016. Počet vygenerovaných řádků pro jednotlivé entity je uveden v zápatí každé tabulky datového modelu. Celkový objem nekomprimovaných dat ve formátu CSV čítá téměř 340 MB. Vezmeme-li v úvahu, že tento objem tvoří plně normalizovaný datový model, pro jeho denormalizovanou verzi by byl objem blížký 2 GB. Všechna data jsou dostupná v příloze DVD 1.

## 5.3 Předmět hodnocení

Pro vypracování hodnocení možností jednotlivých platforem bylo nutné nejprve stanovit, jakým způsobem bude testování probíhat a co bude samotným předmětem hodnocení. Z toho důvodu bylo stanoveno 20 oblastí kritérií (neboli dílčích testů), které mají za úkol komplexně prověřit funkcionalitu či možnosti systémů v různých případech užití, jež tvoří běžné požadavky na dnešní BI systémy.

### 5.3.1 Integrace dat

Napojení na datové zdroje je vstupní branou každé BI platformy. V rámci tohoto kritéria je hodnoceno, jaké možnosti integrace dat daný systém nabízí. Zaměřeno je především na absolutní počet datových zdrojů, ke kterým se lze pomocí BI nástroje připojit přímo a data v něm ihned zpracovat. Zdroje jsou dále hodnoceny i z hlediska kategorií, kterými jsou soubory (Excel, CSV, XML, JSON, ...), relační databáze (MySQL, Microsoft SQL Server, Oracle, ...), Big Data databáze (Google BigQuery, Mongo DB, ...), cloudové služby (GA, Facebook, ...) a jiné zdroje (např. data o počasí nebo směnných kurzech). Hodnoceno je také samotné prostředí pro integraci dat a jeho uživatelská přívětivost.

### 5.3.2 Datový model

V rámci této kategorie je hodnoceno, jakým způsobem probíhá práce s datovým modelem – zdali systém lpí na explicitní definici datových typů, významových typů a primárních či cizích klíčů. Další otázkou je, zda systémy podporují hierarchie atributů.

### 5.3.3 Datumová dimenze

Patrně jednou z nejdůležitějších částí datového modelu každého BI projektu je datumová dimenze. Přestože je kalendářní rok standardizovanou záležitostí, existují různé národní odchylky ve formě svátků nebo jiného chápání začátku pracovního týdne (Po-Ne vs. Ne-So). Samotnou kategorií jsou účetní neboli fiskální kalendáře, které mohou být pro konkrétní firmy zcela specifické. Celkově vznikají odlišné nároky na datumové dimenze, které jednotliví dodavatelé BI řešení zvládají různými způsoby. Během hodnocení datového

modelu je zacíleno i na problematiku datumových dimenzí, přičemž cílem je zhodnotit, jaké datumové dimenze systémy podporují a jak přívětivá je s nimi práce.

### 5.3.4 Metriky

Každá analýza v rámci BI je závislá na metrikách. Teprve jejich definicí odvozením od základních faktů za použití matematických či statistických funkcí je možné data důkladněji zkoumat. V tomto případě hraje největší roli výčet funkcí, které je možné využít, jejich reálné možnosti použití a jednoduchost. Hodnocení práce s metrikami je zaměřeno zejména na možnosti a přívětivost jazyka, kterým jsou definovány. Předmětem zkoumání jsou rovněž otázky, zdali je možné metriky nějakým způsobem kategorizovat a zda mohou uživatelé tvořit vlastní nebo zda jsou odkázány pouze na metriky definované správcem systému.

Pro potřeby testování bylo navrženo několik metrik, které mají za úkol prověřit jak možnosti daného jazyka, tak výkon systému. Jejich výčet včetně popisu výpočtu je uveden v následující tabulce.

Tabulka 1: Testovací metriky

Název metriky	Výpočet
<b># Orders</b>	Počet objednávek. Počet <i>orders.id</i> .
<b># Units</b>	Počet prodaných kusů. Součet <i>orders_products.quantity</i> .
<b>Revenue</b>	Obrat. Součet tržeb z jednotlivých produktů vypočtený jako násobek <i>orders_products.quantity</i> a <i>products.unit_price</i> .
<b>Revenue M</b>	Obrat zákazníků s pohlavím MUŽ.
<b>Revenue F</b>	Obrat zákazníků s pohlavím ŽENA.
<b>AVG Order Revenue</b>	Průměrná hodnota objednávky. Průměr z celkových hodnot objednávky, vypočtených jako součet obrátů jednotlivých produktů v objednávce.
<b>RUNAVG7 Order Revenue</b>	7 denní klouzavý průměr. Klouzavý průměr hodnoty objednávky počítaný ze 7 předcházejících hodnot ( $t - t-6$ ).

Zdroj: vlastní

### 5.3.5 Vizualizace

Nedílnou součástí každého BI je vizualizace dat, což při správném použití zvyšuje vypovídací hodnotu dat a usnadňuje jejich pochopení. Nejzákladnějším nástrojem pro prezentaci dat jsou tabulky. Ty jsou zkoumány z hlediska přehlednosti a možností nastavení. Dalším důležitým faktorem je otázka, zdali je možné data zobrazovat ve formě kontingenční tabulky (matice).

Dalším typem zobrazovačů jsou různé typy grafů. U nich je hodnocena rozmanitost i samotný vzhled. Důležitým kritériem je také interaktivita, čili to, jak grafy s uživatelem komunikují při kontaktu s kurzorem.

#### 5.3.5.1 Mapy

Oblíbenou metodou zobrazování je vykreslení geograficky vztažených dat na mapových podkladech. Zde se způsoby zobrazování obecně dělí na dvě kategorie. První z nich využívá k určení pozice zeměpisné souřadnice (šířka, délka). Data jsou pak zpravidla zobrazena jako kruhy. Druhá kategorie je založena na přesném názvu nebo normovaném kódu určitého územního celku (země/stát, region, kraj, apod.). V prvním případě bývá zobrazení běžně rozšířeno o kruhy s měnící se velikostí v závislosti na velikosti dané metriky; v druhém případě pak o měnící se odstín barvy daného územního celku, opět v závislosti na velikosti dané metriky. Vzrůstající oblíbenosti se těší také tzv. teplotní mapy (heatmaps).

Předmětem zkoumání je v tomto případě především kvalita mapových podkladů a možnosti zobrazení. Důležitým faktorem také je, jaká geografická data lze pro vykreslení použít.

#### 5.3.5.2 KPI

V souvislosti s vizualizacemi je taktéž hodnocena práce s klíčovými indikátory výkonnosti neboli Key Performance Indicator (KPI). Ty jsou specifickou kategorií vizualizací, jejichž cílem je signalizovat aktuální stav oproti ekvivalentu v minulosti nebo budoucí očekávané hodnotě. KPI se tedy používají například pro porovnání aktuální hodnoty oproti plánované (procentuální naplnění měsíčního plánu) nebo jak se hodnota změnila oproti minulému období. Tyto indikátory bývají nejčastěji zobrazovány buď v prosté textové formě (+23,5 %

apod.) nebo také ve formě semaforu či barometru signalizujícím aktuální stav fungování podniku. Do srovnání je opět zahrnuta uživatelská přívětivost tvorby KPI i jejich možnosti zobrazení.

### **5.3.5.3 Reporty a dashboardy**

Různé systémy využívají pro ukládání jednotlivých vizualizací či jejich shlukování do skupin odlišné přístupy. Většinou se jedná o přístup v podobě reportů a/nebo dashboardů. V rámci porovnání je zahrnuto hodnocení jednotlivých použitých přístupů s ohledem na míru jednoduchosti jejich použití či na jejich flexibilitu.

### **5.3.5.4 Změna detailu**

Velmi důležitou součástí vizualizačního aparátu je možnost jednoduché změny zobrazeného detailu dat, která v případě potřeby umožní rychle přejít na zobrazení s vyšší či nižší podrobností. Hodnotícím kritériem je tedy otázka, jakým způsobem je možné s operacemi drill-down/drill-in či drill-across pracovat.

### **5.3.6 Ad-hoc dotazy**

Někdy je potřeba rychle získat odpověď na zcela specifickou otázku, jež je založena na již existujících metrikách, avšak v reportech nebo dashboardech není zastoupena a ani se nepočítá s jejím dalším využitím. K takovým účelům běžně slouží prostředí pro tzv. ad-hoc dotazování. Jedná se o místo, kde je možné rychle klást dotazy (běžně pomocí drag&drop), přepínat mezi zobrazením v tabulce nebo různými typy grafů a pomocí filtrů se rychle dostat ke kýženému výsledku. Porovnání je opět zaměřeno především na uživatelský komfort, který je u této součásti BI nejdůležitějším atributem.

### **5.3.7 Forecasting**

Moderní dynamické firmy mohou mít na BI systémy specifické požadavky týkající se potřeby funkcí na odhad budoucího vývoje veličiny (forecasting). Jedná se o další přidanou hodnotu, které může být využito například jako pomocníka při plánování. Dalším



předmětem srovnání je i otázka, jakým způsobem si s předpovědí budoucího vývoje metriky jednotlivé nástroje poradí.

Obzvláště u této funkce je ale nutné zmínit, že se jedná o značně pokročilé nástroje, které však mohou při chybném použití nadělat více škody než užitku. Jejich používání by tak vždy mělo zůstat spíše v rukou odborníků na statistiku či ekonometrii.

### **5.3.8 Jedinečné vlastnosti**

V duchu diferenciacce se každý z výrobců logicky snaží do svého produktu integrovat nějakou jedinečnou vlastnost, která by jej od ostatních odlišila. Pokud se nějaké takové vlastnosti v testovaných nástrojích nacházejí, dojde i k jejich zmínění.

### **5.3.9 Export dat**

Důležitou součástí BI systému, a tedy dalším bodem srovnání, je možnost exportu dat. Uživatelé by měli mít možnost zobrazená data v tabulkách nebo grafech exportovat do různých datových formátů (běžně Excel či CSV) pro možnost dalšího zpracování. V případě grafů je běžně žádanou funkcí také export ve formě obrázku nebo jako dokument PDF.

### **5.3.10 Mobilní přístup**

Jak již bylo několikrát zmíněno, mobilní BI je jedním z hlavních trendů dnešní doby. Proto ani v tomto hodnocení nebude mobilní přístup opomenut. Zaměřeno je jak na samotnou existenci nativní mobilní aplikace, případně mobilní verze webového rozhraní, tak samozřejmě na možnosti zobrazování a mobilní analýzy.

### **5.3.11 Sdílení a omezení přístupu**

Pokud je dalším z dnešních trendů zpřístupňovat data širšímu spektru zaměstnanců, než tomu bylo dříve, je nezbytnou součástí BI platformy možnost sdílení obsahu s více uživateli napříč organizací. S tím také souvisí nutné řízení přístupových oprávnění pro konkrétní množiny dat. V odborné terminologii se tento přístup nazývá Row-Level Security (RLS)

neboli řízení přístupu na úrovni jednotlivých řádků. Důvodem omezení přístupu je především to, aby například jednotliví obchodní zástupci měli přístup pouze k datům vztahujícím se k jejich obchodům apod. Zkoumány jsou jak možnosti sdílení, tak omezování přístupu pro konkrétní uživatele či skupiny.

### **5.3.12 Rychlost**

Jedním z posledních hodnotících kritérií je rychlost odezvy systému. Přestože se webové aplikace většinou vyznačují bleskovými odezvami, zpracování většího objemu dat ji může ovlivnit. Bohužel tento test nemusí mít velkou výpovědní hodnotu, neboť zde vzniká několik problémů, které správné otestování rychlosti vylučují. Prvním problémem je fakt, že se odezva může s měnícím se objemem dat lišit, což v tomto případě (kdy je připraven fixní testovací objem dat) není možné změřit. Dalším limitujícím faktorem jsou zkušební verze BI nástrojů, které v tomto režimu povolují zpravidla nanejvýš 0,5-1 GB uživatelských dat. Tento limit byl vzat v úvahu během generování dat, kde výsledný objem surových dat činí oněch zmíněných 340 MB. Tento testovací objem dat lze označit na cloudové poměry za relativně malý, i když i takový objem může některým platformám činit potíže. Svou roli hraje rovněž normalizace datového modelu. Zpravidla plně denormalizované datové modely (jedna velká tabulka obsahující všechna data) jsou při zpracování mnohem rychlejší než plně normalizované datové modely. Vyplývá to obecně z technologie zpracování relačních dat i z dnes běžně používaných sloupcových databází, kde se při denormalizaci nemusí tabulky spojovat, což činí zpracování celkově rychlejší. Za poslední překážku před správným měřením lze označit nestabilitu internetového připojení, čímž je míněna nestálá rychlost a nerovnoměrné využívání linky, jejímž prostřednictvím je uživatel k systému připojen. I přes tyto překážky je v rámci porovnání alespoň snaha o subjektivní hodnocení rychlosti odezvy systému.

### **5.3.13 Cena a omezení**

Posledním bodem hodnocení je cena, avšak zde opět nastává několik problémů znemožňujících srovnání. Mezi nejzávažnější z nich patří zcela rozdílné cenové politiky či

podmínky licencování a také neveřejné ceny. Z toho důvodu nepůjde v případě ceny příliš o srovnání s konkurencí, ale spíše o kompilaci a objasnění konkrétního přístupu.

V souvislosti s cenou je zaměřeno i na omezení respektive limity systému. Zda jsou stanoveny nějaké pevné systémové limity, které nelze navýšit, či se vše odvíjí od potřeb zákazníka a limitujícím faktorem je pouze výsledná cena služby.

## 5.4 GoodData

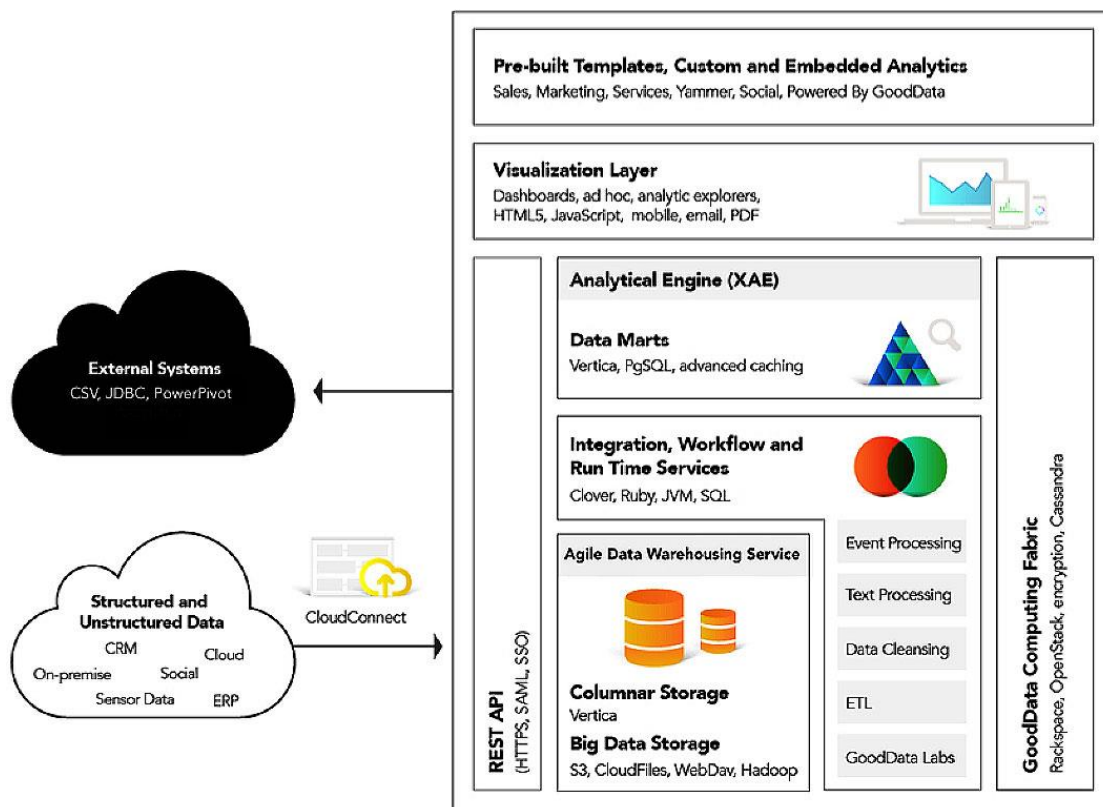
Tato technologická společnost se sídlem v San Franciscu byla založena v roce 2007 českým podnikatelem Romanem Staňkem. Proslavila se právě svou čistě cloudově orientovanou analytickou BI platformou. Podle dříve zmíněného hodnocení Gartner Magic Quadrant<sup>104</sup> z února 2016 byla GoodData (GD) zařazena do sekce „Niche Players“, tedy okrajoví hráči. Toto zařazení bylo odůvodněno sice silnou vizí pro odhalování skrytých informací v datech, ale postrádáním některých klíčových funkcí moderních BI a cílením na poměrně specifickou skupinu zákazníků vyžadující velmi silné analytické zázemí.

### 5.4.1 Koncept

Jak o sobě GD tvrdí, jedná se o cloudovou otevřenou analytickou platformu (PaaS) skládající se ze základních komponent BI, umožňující extrakci dat, uložení, analýzu a vizualizaci. Jednotlivé komponenty je možné využít zcela nezávisle nebo jako kompletní, plně integrovanou službu (SaaS), na kterou je tento výzkum zaměřen. Schématické znázornění fungování platformy GD – viz obrázek 5-3.

---

<sup>104</sup> *Magic Quadrant for Business Intelligence and Analytics Platforms* [online]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>.



Obrázek 5-3: GD – architektura platformy

Zdroj: *GoodData Open Analytics Platform Overview* [online]. San Francisco: GoodData Corporation, 2014 [cit. 2016-12-20]. Dostupné z: <http://info.gooddata.com/rs/gooddata/images/GoodData%20Platform%20Technical%20Brief.pdf>

Tím, že se jedná o samostatné komponenty, lze je současně použít i v kombinaci s jinými systémy. Proto není vždy nezbytné pořizovat zcela nové celkové řešení. Pokud je ale cílem na budování nového BI řešení, platforma poskytuje všechny komponenty nezbytné pro jeho provoz. Zároveň škálovatelnost datové a analytické vrstvy hlásá použitelnost i během zpracování Big Data či dat nestrukturovaných.

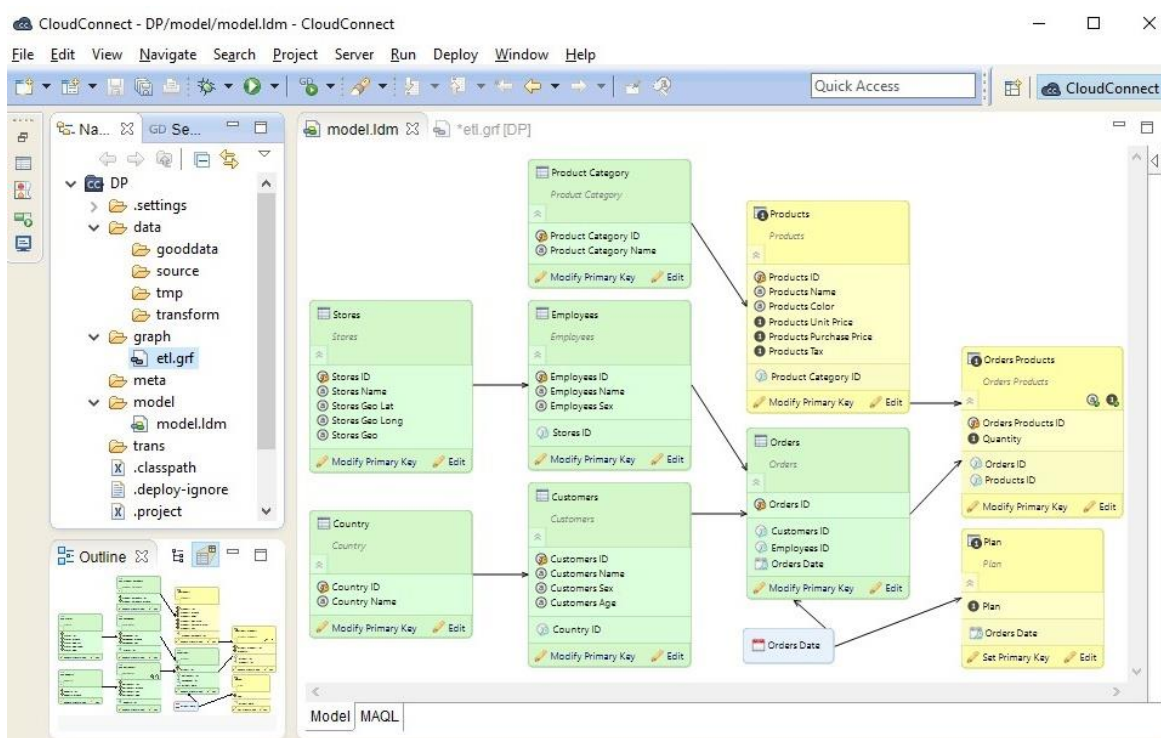
Otevřenost platformy je zajištěna především tím, že všechny dílčí komponenty umožňují přístup přes otevřené komunikační rozhraní REST ve formátu JSON. Z tohoto důvodu je snadné integrovat jednotlivé části s téměř jakýmkoliv jiným systémem.<sup>105</sup>

<sup>105</sup> *GoodData Open Analytics Platform Overview* [online]. San Francisco: GoodData Corporation, 2014 [cit. 2016-12-11]. Dostupné z: <http://info.gooddata.com/rs/gooddata/images/GoodData%20Platform%20Technical%20Brief.pdf>

## 5.4.2 Integrace dat a datový model

GD nabízí pro integraci dat vlastní ETL nástroj, zvaný Cloud Connect Designer<sup>106</sup>, postavený na základech CloverETL. Toto je také jediný způsob<sup>107</sup>, jak bez pomoci aplikací třetích stran či použití API do GD data odeslat.

V prostředí této desktopové aplikace (Windows, Mac, Linux) je nejprve potřeba definovat logický datový model tzv. LDM (viz obrázek 5-4). Ten může být připodobněn k relačnímu modelu dat a v GD slouží pro jasné definování významu datových polí a vazeb mezi datovými objekty, díky čemuž systém následně určí, jakým způsobem mají být data zpracována. Model nemusí odpovídat klasickým star-schématům, čímž umožňuje lepší a jednodušší zachycení „business vztahů“.



Obrázek 5-4: GD Cloud Connect Designer – logický datový model

Zdroj: vlastní

<sup>106</sup> Data Loading / ETL. *GoodData Developers* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-23]. Dostupné z: <https://developer.gooddata.com/docs/data-loading>

<sup>107</sup> GoodData sice umožňuje přímé nahrání dat do projektu prostřednictvím odeslání CSV souborů, avšak z těchto souborů nelze vytvořit samostatný datový model, pouze je možné je připojit na existující datovou dimenzi v existujícím datovém modelu.

Při vytváření LDM jsou v prostředí Cloud Connect Designer uplatňovány dva významové typy, a to atribut a fakt. Datové typy nemusejí být explicitně specifikovány, neboť systém atributům automaticky přiřadí typ Text (128) a faktům Decimal (12,2), avšak tyto výchozí hodnoty mohou být změněny. Poslední povinnou složkou tvorby LDM je určení primárních klíčů a vazeb na související objekty. Datumová dimenze je nabízena jako hotový objekt s denní granularitou, bohužel bez jakékoliv možnosti úprav (více viz 5.4.3 - Datumová dimenze).

Hierarchii atributů je možné tvořit později ve webovém rozhraní, kde lze každému atributu přiřadit podřízený atribut, na který se má při požadavku o vyšší detail reportu přesunout. Výhodou je, že propadové (drill-in) atributy lze mezi sebou nastavit zcela libovolně v rámci celého datového modelu.

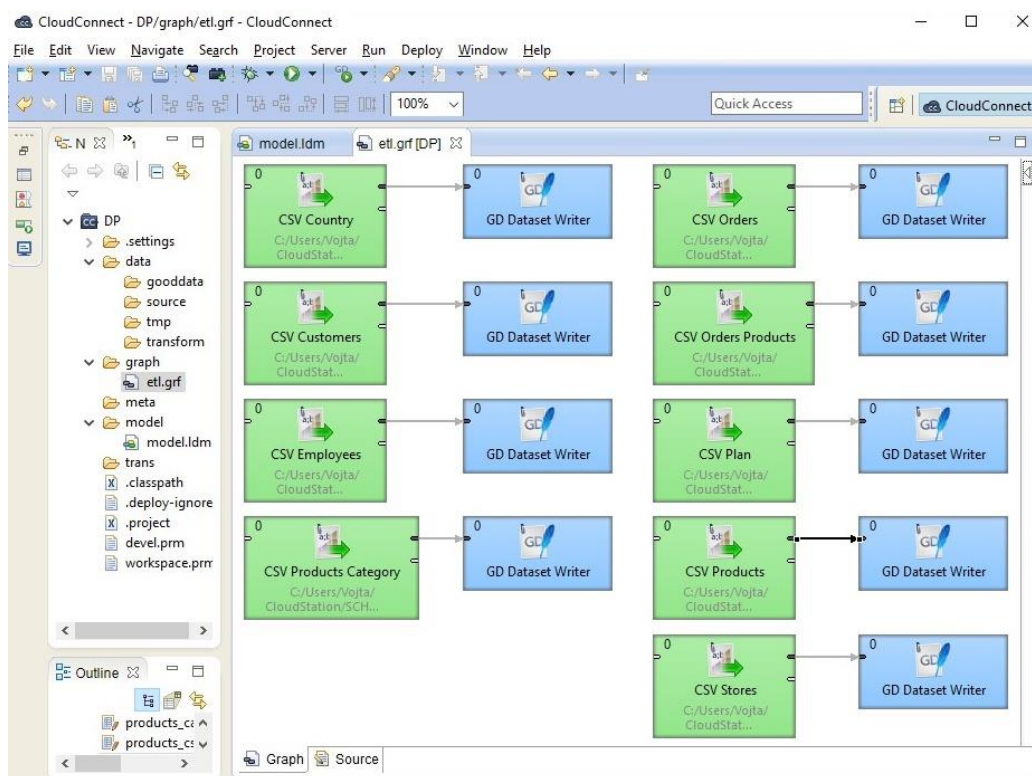
Po definování LDM a odeslání do GD je teprve možné model naplnit daty. Tento proces probíhá prostřednictvím vytvořeného ELT procesu, který je v daném prostředí označován jako „graph“. Připojení na testovací CSV soubory nebylo příliš složité, ale také nijak výrazně uživatelsky přívětivé. Systém stále vyžadoval více uživatelských znalostí než jiné ETL nástroje (například v porovnání s Microsoft SQL Server Integration Services – SSIS) a nebyl dostupný ani žádný intuitivní průvodce.

Přestože je v dokumentaci uvedeno, že pomocí Cloud Connect Designer lze do GD integrovat prakticky libovolný zdroj dat, není zřejmé, jakým způsobem toho může být docíleno. Během testování bylo sice v prostředí Cloud Connect Designer dostupných několik databází, protokolů a dalších strukturovaných či nestruturovaných zdrojů, ale pouze 4 cloudové služby (GA, Facebook, Salesforce a ExactTarget). Existuje také dokument z roku 2014<sup>108</sup>, v němž jsou přímo specifikovány konkrétní datové zdroje (cca 60), které je možné do GD integrovat. O možnosti rozšíření škály (zejména služeb) se bohužel nepodařilo zjistit více. Lze se pouze domnívat, že tvůrci poukazují spíše na vysokou flexibilitu a rozšiřitelnost možných datových zdrojů pomocí podporovaných komunikačních protokolů

---

<sup>108</sup> *GoodData Integrations* [online]. In: . San Francisco: GoodData Corporation, 2014, s. 1-1 [cit. 2016-11-23]. Dostupné z: [http://info.gooddata.com/rs/gooddata/images/GoodData\\_DataSources.pdf](http://info.gooddata.com/rs/gooddata/images/GoodData_DataSources.pdf)

(jako jsou REST<sup>109</sup> nebo SOAP<sup>110</sup>) a zpracování formátů XML či JSON. V takovém případě by sice bylo připojení k jakékoliv službě teoreticky možné, nikoliv však uživatelsky ani implementačně vlné.



Obrázek 5-5: GD Cloud Connect Designer – ETL

Zdroj: vlastní

Prostředí pro tvorbu ETL procesu nabízí řadu připravených transformačních funkcí, pomocí nichž lze data různými způsoby slučovat, třídít či formátovat, a tím je dostat do požadované podoby vhodné k následné analýze.

Výsledná data je velmi jednoduché zapsat do připraveného datového modelu, a to pomocí tzv. GD Dataset Writeru. Na vstup této komponenty stačí napojit výstup poslední z řady transformací či přímo výstup datového zdroje, a nastavit mapování atributů. Testovací ETL proces zachycuje obrázek 5-5. Projekt Cloud Connect Designer – viz příloha DVD 2.

<sup>109</sup> Aplikační rozhraní REST je architektura rozhraní, navržená pro distribuované prostředí, které je použitelné pro jednotný a snadný přístup ke zdrojům (datům nebo stavům aplikace). Jedná se o datově orientované rozhraní, kde za pomoci jednoduchých HTTP volání lze jednoduše vytvořit, číst, editovat nebo smazat informace ze serveru.

<sup>110</sup> SOAP je protokolem pro výměnu zpráv založených na XML přes síť, hlavně pomocí protokolu HTTP.

Celý proces integrace dat v GD funguje tím způsobem, že je projekt vytvořený v prostředí Cloud Connect Designer odeslán na server GD (Data Integration Console), kde je spuštěn v definovaných intervalech. Výhodou běhu ETL procesu v cloudu je, že není nutné zatěžovat vlastní infrastrukturu či pořizovat infrastrukturu novou. Existují zde ale i některá omezení, například fakt, že maximální doba běhu ETL procesu je 5 hodin, nebo že všechny datové zdroje musejí být přímo dostupné v rámci sítě internet. Zdroje proto musejí být přímo vystaveny do internetu, což může znamenat vyšší bezpečnostní riziko, než když jsou uzavřeny ve vnitřní síti bez přímého vnějšího přístupu. Power BI má v tomto ohledu navrch, neboť poskytuje nástroj pro vytvoření virtuálního spojení mezi uzavřenou sítí a svým datovým centrem. To na rozdíl od GD umožňuje přenést data z libovolného počítače připojeného k síti internet.

GD nabízí též možnost odesílání dat přímo pomocí určitého rozhraní, avšak jeho implementace pro vlastní projekt nemusí být zdaleka tak jednoduchá jako použití aplikace Cloud Connect Designer a Data Integration Console. Je zřejmé, že komunikace pomocí rozhraní je určena spíše aplikacím třetích stran, které mají za úkol integraci dat do GD zprostředkovat. Jedním z takových nástrojů je například cloudové ETL – Keboola Connection.

### 5.4.3 Datumová dimenze

Jak bylo v předešlé části nastíněno, v prostředí Cloud Connect Designer je možné zvolit pouze jednu statickou předpřipravenou datumovou dimenzi. Pokud si uživatel přeje nějaké odchylky v podobě hodinové granularity nebo struktury kalendáře (fiskální), může se pokusit o vlastní implementaci, jak radí například Tobolka ve svém příspěvku<sup>111</sup>, nebo o tuto službu přímo požádat konzultanty GD. Jelikož ale může být zavedení vlastní datumové/časové dimenze pro většinu uživatelů velmi obtížné, lze proto očekávat, že bude nutné kontaktovat zástupce GD, kteří ji za nemalý poplatek<sup>112</sup> implementují. Na druhou

---

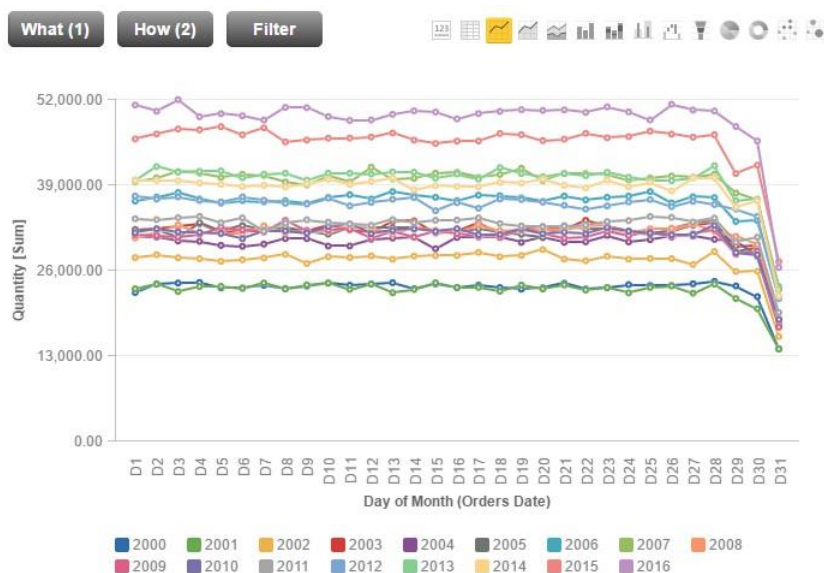
<sup>111</sup> TOBOLKA, Jiří. How to load time in GoodData?. *Stack Overflow* [online]. Prague: Stack Exchange Inc, 2014 [cit. 2016-11-24]. Dostupné z: <http://stackoverflow.com/questions/20964162/how-to-load-time-in-gooddata>

<sup>112</sup> Dle nejménovaného zdroje jde o částku v řádu desítek tisíc Kč.



stranu výchozí dimenze obsahuje vše potřebné pro práci s kalendářním rokem – čtvrtletí, měsíce i týdny, které jsou zastoupeny jak v pojetí evropském (po-ne), tak americkém (ne-so). Více o struktuře a hierarchiích této dimenze je možné najít v dokumentaci<sup>113</sup>.

Pozdější práce s datovou dimenzí je v GD na prvotřídní úrovni zejména díky tomu, že s každým atributem lze pracovat zcela nezávisle. Například zobrazení hodnoty veličiny pro jednotlivé dny měsíce, které jsou zároveň zobrazeny po jednotlivých letech (obrázek 5-6), je proto zcela triviální záležitostí. Pohodlí práce také přispívá přítomnost několika funkcí a klíčových slov (THIS, PREVIOUS, FOR Next(), ...), které umí s nabízenými časovými obdobími automaticky pracovat a dokáží tak pohyb po časové ose velmi zpříjemnit.



Obrázek 5-6: GD – součet množství po dnech kalendářního měsíce a letech

Zdroj: vlastní

#### 5.4.4 Metriky

Metriky se (na rozdíl od datového modelu) tvoří přímo v online prostředí GoodData. Základ každé metriky tvoří název, definice v jazyce MAQL (Multi-Dimension Analytical Query Language) a syntaxe řídicí formátování čísel při zobrazení. Dotazovací jazyk MAQL je

<sup>113</sup> Date Attribute Hierarchies. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-24]. Dostupné z: <https://help.gooddata.com/display/doc/Date+Attribute+Hierarchies>

vlastním produktem společnosti GoodData, která jej vyvinula pro potřeby jednoduché a intuitivní práce s daty. Jazyk může být kvůli některým použitým klíčovým sloům zdánlivě připodobněn k jazyku SQL, ale fundamentálně se jedná o datový dotazovací jazyk mnohem vyšší úrovně. Oproti SQL se uživatel pohybuje pouze na vrstvě logického modelu dat, čímž je zcela odstíněn od spojování tabulek. Je tak možné velmi jednoduše kombinovat různé fakty a atributy napříč LDM (samozřejmě pouze ty související) a systém veškeré operace nutné pro získání dat z fyzického modelu automaticky zajistí.

V následující tabulce jsou uvedeny definice testovacích metrik, stanovených dříve (viz tabulka 1). Díky abstrakci na úrovni LDM a jednoduchým agregacím pomocí klíčového slova „BY“ je syntaxe velmi čistá a jednoduchá. Díky přítomnosti spousty vestavěných matematických či statistických funkcí<sup>114</sup> jazyk umožňuje data analyzovat důkladně, ale zároveň velmi jednoduše.

Tabulka 2: Testovací metriky (GoodData – MAQL)

Název metriky	MAQL
# Orders	SELECT COUNT( <b>Orders ID</b> )
# Units	SELECT SUM( <b>Orders Products Quantity</b> )
Revenue	SELECT SUM( <b>Orders Products Quantity</b> * <b>Products Unit Price</b> )
Revenue M	SELECT <b>Revenue</b> WHERE <b>Customers Sex</b> = <b>M</b>
Revenue F	SELECT <b>Revenue</b> WHERE <b>Customers Sex</b> = <b>F</b>
AVG Order Revenue	SELECT AVG(SELECT <b>Revenue</b> BY <b>Orders ID</b> )
RUNAVG7 Order Revenue	SELECT RUNAVG(SELECT <b>Revenue</b> BY <b>Orders ID</b> ) ROWS BETWEEN 6 PRECEDING AND CURRENT ROW WITHOUT PF
<b>Legenda:</b> <b>Fakt</b> , <b>Metrika</b> , <b>Atribut</b> , <b>Hodnota atributu</b>	

Zdroj: vlastní

Ze zápisu metrik v předešlé tabulce je patrné, jaké výhody MAQL oproti SQL přináší. V případě metriky *Revenue* zcela odpadá spojování tabulek *Orders Products* a *Products*. U *AVG Order Revenue* dále odpadá potřeba dalšího spojení s tabulkou *Orders* a v případě *RUNAVG7 Order Revenue* pak další složitější spojení na základě datového intervalu.

<sup>114</sup> Getting Started with MAQL. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-24]. Dostupné z: <https://help.gooddata.com/display/doc/Getting+Started+with+MAQL>

Pro představu o míře zjednodušení práce s MAQL oproti SQL se v příloze A se nachází grafický strom dotazů, který je v GD na pozadí vytvořen na základě definice reportu obsahující metriku AVG Order Revenue a RUNAVG7 Order Revenue. V příloze B se pak nachází samotný vygenerovaný SQL dotaz pro získání dat daného reportu z datového skladu.

V GD existují dva typy metrik, globální a lokální. Globální metriky jsou dostupné napříč celým systémem (reporty, dashboardy, ad-hoc, KPI) a je možné je třídit do složek; zatímco lokální metriky jsou definovány uvnitř konkrétního reportu, kterému také výhradně slouží. Metriky mohou, i díky definování až v online prostředí, vytvářet rovněž samotní uživatelé.

### **5.4.5 Vizualizace**

GoodData poskytuje poměrně standardní sadu vizualizačních nástrojů. Data je možné zobrazit ve formě tabulky i ve formě matice. U grafů je umožněn výběr ze všech základních běžně používaných typů. Z těch méně běžných stojí za zmínku vodopádový graf, bublinový graf nebo typ „bullet“. Grafy je možné poměrně dobře přizpůsobovat (minima/maxima os, jednotky, barvy, popisky), u tabulek lze nastavovat pouze formátování čísel. Některé uživatele by mohla odradit nepřítomnost grafů typu budík či barometr, které jsou, zejména u vyššího managementu, velmi oblíbené pro vizualizaci KPI.

Co se interaktivity týče, při styku kurzoru s datovým bodem/oblastí je ve formě bubliny zobrazena hodnota včetně jejího vymezení. U většiny vizualizací je také možné nastavit propad (drill-in) na detailní atribut.

Z pohledu moderního grafického designu, kterým se lídři trhu ubírají, se zdá, že GD v tomto směru poměrně zaostává, neboť vizuální stránka přímých konkurentů (Tableau, Chartio, Power BI) vypadá mnohem lépe.

#### **5.4.5.1 Mapy**

Zobrazení dat na mapě je v GD možné, nabízí se ale pouze dva typy vizualizací geografických dat s minimálními možnostmi nastavení. Prvním typem je zobrazení na základě zeměpisných souřadnic (obrázek 5-7), kde se data na mapě zobrazují jako kruhy

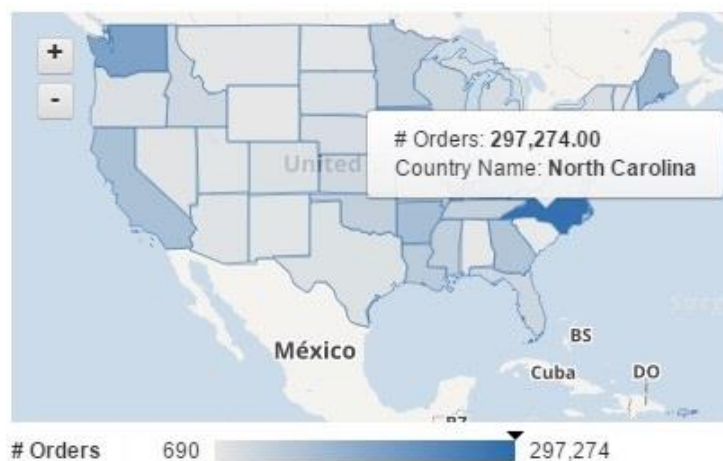
s různou světlostí zvolené barvy. Bohužel chybí množnost celkového zvětšení kruhů nebo měnící se velikost v závislosti na velikosti zobrazované veličiny. Stejně tak není možné každý bod odlišit jinou barvou.



Obrázek 5-7: GD – počet objednávek na mapě dle souřadnic obchodu

Zdroj: vlastní

Druhou možností je zobrazení na základě přesně specifikovaného názvu nebo normovaného kódu určitého územního celku (země/stát/region) – viz obrázek 5-8.



Obrázek 5-8: GD – počet objednávek na mapě dle státu zákazníka

Zdroj: vlastní

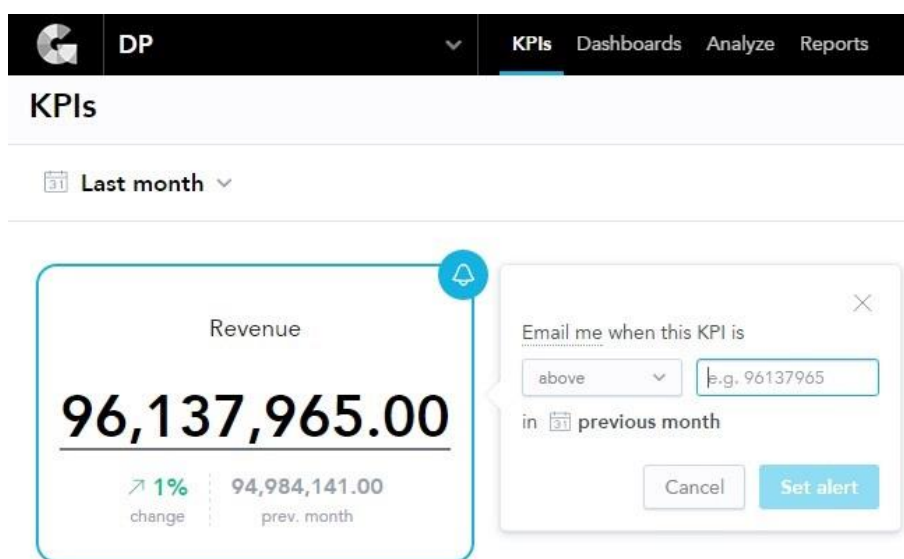
Podpora těchto územních celků je v GD poměrně dobrá. Podporovány jsou samozřejmě všechny státy světa, ale také jednotlivé státy USA a Austrálie, a co více, také okresy ČR, což je s nejvyšší pravděpodobností zásluhou českého zakladatele firmy GoodData i díky

umístění vývojových center na našem území. Bohužel ani v tomto případě není možné kromě barvy zobrazení nijak upravit.

Mapy jsou sice podrobné, ale celkově působí velmi jednoduše. Jistě by nebylo od věci, kdyby bylo možné zobrazovanou bublinu doplnit o další údaje, které by informační hodnotu ještě pozvedly. Poslední výtkou je, že uvedené mapy lze pouze vložit do dashboardu a nelze z nich vytvořit samostatný report.

### 5.4.5.2 KPI

Pro klíčové indikátory výkonnosti existuje (na rozdíl od jiných platform, kde se KPI umisťují přímo na společný dashboard) v GD speciální místo. Na vyhrazený KPI dashboard (obrázek 5-9) je možné umístit libovolný počet panelů signalizující vývoj zvolené metriky oproti předešlému období nebo stejnému v předešlém roce. U každého panelu lze nastavit emailové upozornění v případě překročení určité hodnoty. Bohužel tyto mezní hodnoty je nutné zadat ručně a nelze je brát například z metriky určující plán. Pro celý dashboard je dostupný globální časový filtr, díky němuž je jednoduché přepnout zobrazení KPI pro minulý týden, měsíc, čtvrtletí nebo jiné období. Dle ústního sdělení je KPI dashboard stále ve vývoji, tudíž se v budoucnu dají očekávat další vylepšení, avšak vývoj je velmi pomalý, čímž konkurence získává stále větší náskok.



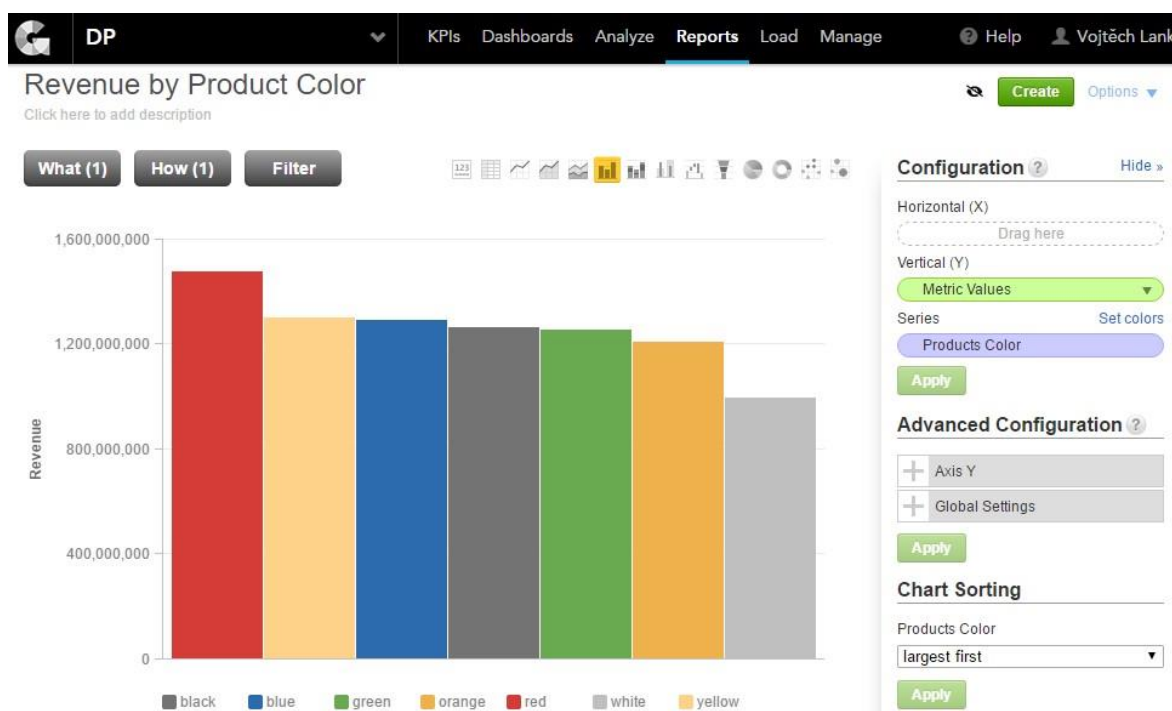
Obrázek 5-9: GD – KPI dashboard

Zdroj: vlastní

Pro porovnání plánovaných hodnot oproti dosaženým je možné využít vizualizaci typu bullet chart, v jejímž rámci lze za použití metrik nastavit i rozsahy pro určení stavu jako „špatný“, „přiměřený“ a „dobrý“.

### 5.4.5.3 Reporty a dashboardy

GD používá pro zobrazování dat reporty i dashboardy. Základním principem práce s vizuály je, že jednotlivé vizualizace jsou nejprve vytvořeny v rámci konkrétních reportů (každý report může obsahovat pouze jeden typ vizualizace). Ty pak mohou být skupinově umístěny na dashboard. Obrázek 5-10 zachycuje prostředí pro návrh reportů. Proces návrhu spočívá ve zvolení globálních metrik nebo vytvoření lokálních – „What“, vybrání řezových atributů – „How“, případné nastavení požadovaných filtrů – „Filter“ a zvolení vhodného typu vizualizace.

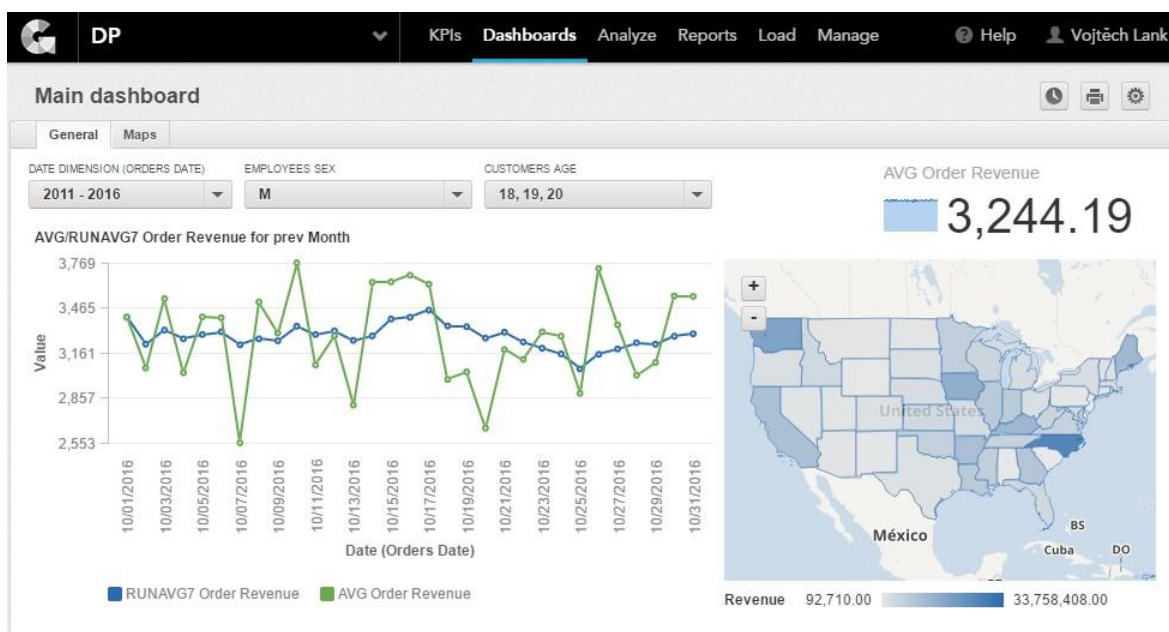


Obrázek 5-10: GD – návrh reportu

Zdroj: vlastní

Takto vytvořený report je následně uložen (může být ve složce) a může být společně s dalšími připnut na dashboard – viz obrázek 5-11. Návrh reportů je poměrně intuitivní záležitostí a s dostatečným záběrem potřebných metrik ho zvládne prakticky každý.

Dashboards je možné vytvořit několik (např. dle divizí nebo oddělení společnosti). Každý z nich může být dodatečně strukturován pomocí záložek. Kromě připravených reportů lze přímo na dashboard umístit také jednočíselné metriky a již zmíněné mapy. Dashboard je taktéž možné doplnit o filtr datumové dimenze nebo jakéhokoliv jiného atributu. Celé prostředí a koncept práce s reporty a dashboardy tedy působí velmi komplexním a propracovaným dojmem.



Obrázek 5-11: GD – dashboard

Zdroj: vlastní

Celkově dobré hodnocení kazí poměrně složitá individualizace dashboardů. Panel nástrojů umožňuje pouze vložení textu nebo vertikální či horizontální čáry. Pokud chce firma dashboard doplnit například o logo společnosti, nelze jej přímo nahrát, ale musí na něj být odkázáno pomocí URL adresy, která musí být zabezpečená (HTTPS). Vniká tak nutnost pro hostování vlastních zdrojů (obrázků), což je další uživatelskou překážkou. Tímto způsobem je sice možné vložit libovolné obrázky či iframe rámy s embedovaným obsahem typu YouTube video nebo Google Maps, ale uživatelsky se nejedná o příliš přívětivé řešení.

#### 5.4.5.4 Změna detailu

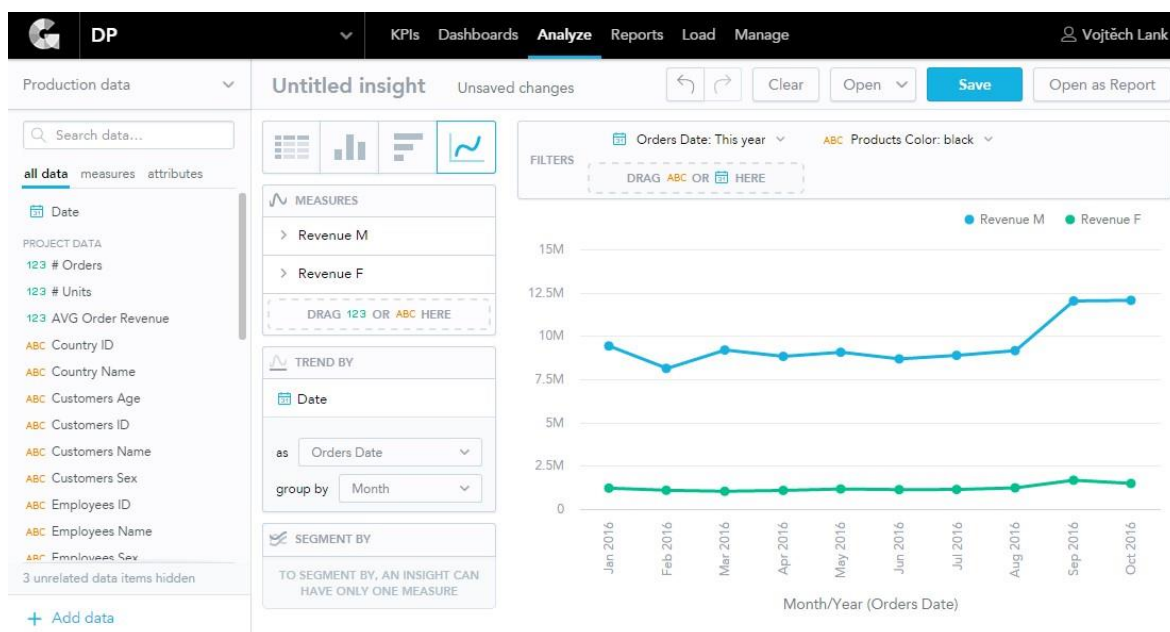
Změna zobrazeného detailu reportu je v GD možná několika způsoby. Kliknutím na název kategorie v rámci osy dochází k filtrování výsledku dle vybrané kategorie a propadu



na podřízený atribut v rámci globální hierarchie. U některých reportů lze nastavit i propad při kliknutí na samotnou datovou oblast kategorie (bod/sloupec/výseč). V takovém případě dochází ke stejné akci, ale pro jiný dril-in atribut, definovaný konkrétně po daný report. V rámci dashboardů lze chování zcela změnit a nastavit libovolné akce pro kategorie i datové oblasti kategorií, jimž je možné nastavit propad na atribut nebo přímo na jiný report.

## 5.4.6 Ad-hoc dotazy

Prostředí pro ad-hoc dotazování (obrázek 5-12) se sice v podání GD nevyznačuje velkým množstvím funkcí, ale je zcela jednoduché na používání. Prostým drag&drop přetažením atributů a metrik na požadovaná místa získá uživatel ihned výsledek. Filtrování je možné pomocí libovolného množství atributů. Krájení pomocí datové dimenze je bohužel z nepochopitelných důvodů omezeno pouze na hierarchii den, týden (Ne-So), měsíc, čtvrtletí a rok. Nevýhodou této implementace ad-hoc prostředí je přítomnost pouze spojnicového a dvou sloupcových grafů; další nevýhodou je fakt, že tabulková data nelze zobrazit jako matici. Pro potřeby důkladnějšího zkoumání je však možné právě zobrazený výsledek ihned otevřít v editoru reportů, a tam s ním dále pracovat.



Obrázek 5-12: GD – prostředí pro ad-hoc dotazování

Zdroj: vlastní



## 5.4.7 Forecasting

Předpovídání budoucího vývoje veličiny je v prostředí GD možné několika způsoby. Prvním nejjednodušším a také jediným prostředkem, pomocí kterého lze hodnoty předpovídat na více následujících období, je prostá lineární regrese. Ta je v GD k dispozici ve formě funkce FORECAST<sup>115</sup>. Na základě zadaných dat s určitým časovým horizontem funkce vrátí lineární regresní přímku pro až 10 následujících období.

Práce s touto funkcí je velmi jednoduchá, neboť stačí zadat metriku, se kterou se má pracovat, a počet následujících období, pro která má být hodnota spočítána. Výhodou je rovněž fakt (díky vlastnostem jazyka MAQL), že není potřeba explicitně specifikovat granularitu období, se kterými se bude pracovat. Vše je odvozeno až na základě zvolených časových řezových atributů. Obrázek 5-13 zachycuje příklad použití funkce FORECAST pro odhad vývoje obratu pro následujících 6 měsíců na základě údajů z posledních 12 uplynulých měsíců.



Obrázek 5-13: GD – forecasting pomocí regresní přímky

Zdroj: vlastní

Jako další možnost predikce lze využít jednoduchého nebo váženého klouzavého průměru s libovolnou šířkou plovoucího okna. Použití této metody ale již z principu výpočtu omezuje

<sup>115</sup> Time Series Forecasting. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-28]. Dostupné z: <https://help.gooddata.com/display/doc/Time+Series+Forecasting>

předpověď pouze na jedno následující období a její použití také obnáší mírně složitější zápis. V dokumentaci<sup>116</sup> je ale vše poměrně dobře popsáno.

### 5.4.8 Jedinečné vlastnosti

Z pohledu unikátních nebo značně pokročilých vlastností nabízí GD oproti konkurenci velmi silnou základnu pokročilých statistických funkcí<sup>117</sup>. Ať už se jedná o jednodušší průběžné agregační funkce (RUNSUM, RUNMAX, RUNAVG, ...) nebo pokročilejší funkce na výpočet rozptylu (VAR/VARP), směrodatné odchytky (STDEV/STDEVP), kovariance (COVAR/COVARP), korelace (CORREL) nebo koeficientu determinace známým jako R-squared (RSQ). Tuto širokou škálu nástrojů je dále možné zužitkovat například při výpočtu Pearsonova korelačního koeficientu nebo testování hypotéz.

Další zajímavou funkcí je možnost nahrání vlastního souboru pro rozšíření analýzy. Každý uživatel s dostatečným oprávněním může do projektu nahrát libovolné množství souborů s vlastními daty ve formátu CSV. Bohužel data lze s produkčními daty spojit pouze na základě kalendářní datumové dimenze, což použitelnost značně omezuje.

### 5.4.9 Export dat

U každého reportu je nabízena možnost exportu do několika formátů. Vybrat lze z dokumentu PDF, obrázku PNG, sešitu Excel (XLS i XLSX) nebo prostého CSV. Nabízen je také tisk reportu přímo z okna prohlížeče nebo uložení celého dashboardu do PDF.

### 5.4.10 Mobilní přístup

Trend mobilního BI se zdá být v GoodData zcela opomenut. Web nemá podporu mobilního zobrazení, a ani nativní mobilní aplikace, která je dostupná pouze pro Apple iOS, není příliš propracovaná. Aplikace umožňuje pouze pohodlnější přepínání reportů a dashboardů,

---

<sup>116</sup> *Time Series Forecasting* [online]. Dostupné z: <https://help.gooddata.com/display/doc/Time+Series+Forecasting>.

<sup>117</sup> Statistical Functions. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2016 [cit. 2016-11-28]. Dostupné z: <https://help.gooddata.com/display/doc/Statistical+Functions>

jejichž rozložení a vzhled jsou však stejné jako v případě rozložení na PC. Mobilní ad-hoc dotazování ani KPI v aplikaci dostupné nejsou. Z pohledu aktuálních trendů v tomto segmentu GD absolutně nestíhá.

### **5.4.11 Sdílení a omezení přístupu**

Sdílení obsahu je v podání GD poměrně jednoduché. Do projektu je možné přizvat libovolný počet uživatelů, kterým mohou být přiděleny role administrátora, editora nebo diváka. Dále je poskytnuta možnost pravidelného zasílání reportů nebo celých dashboardů emailem ve formátu PDF.

Z pohledu omezení přístupu je tato problematika mnohem složitější. GD sice RLS podporuje, nicméně pro jeho nastavení neexistuje žádné grafické uživatelské rozhraní – existuje pouze prostředí pro přímé zadávání příkazů do GoodData API ve formátu JSON.<sup>118</sup> Takovýto způsob nastavování je pro běžné uživatele absolutně nekomfortní a zcela odporuje trendu samoobslužnosti, neboť nutí uživatele přenechat tyto operace administrátorům z řad IT pracovníků nebo přímo konzultantům GD.

### **5.4.12 Rychlost**

I přes výše uvedené možné problémy spojené s měřením rychlosti byly očekávané bleskové odezvy poměrně zklamáním. Výpočet některých reportů trval (i při takto relativně malém objemu dat) kolem 10 sekund; některé dotazy při ad-hoc dotazování i déle. Z uživatelského pohledu se jedná o jasně negativní dopad na UX.

### **5.4.13 Cena a omezení**

Cenu tohoto produktu bohužel nelze příliš posoudit, neboť GD nemá žádný veřejně dostupný ceník služeb. Vše se odvíjí od konkrétních požadavků daného zákazníka a od výsledku jednání s obchodním zástupcem.

---

<sup>118</sup> Let's get started with Data Permissions. *GoodData Developers* [online]. San Francisco: GoodData Corporation, 2016 [cit. 2016-11-29]. Dostupné z: <https://developer.gooddata.com/article/lets-get-started-with-data-permissions>

Nicméně dle webových archivů se podařilo zjistit cenovou politiku z roku 2010. Ta začínala na 500 amerických dolarech za měsíc pro neomezený počet uživatelů s neomezeným počtem reportů, ale s omezením pouze na 1 dashboard pro každý ze dvou možných projektů (objem každého projektu byl omezen na 1 GB).<sup>119</sup> Dle nejmenovaného zdroje platí podobné podmínky dodnes.

Dle dostupných informací z dokumentace systém nemá žádná výrazná omezení úložné kapacity ani výpočetního výkonu a pravděpodobně je tak limitujícím faktorem pouze cena, kterou je zákazník za službu ochoten zaplatit.

## 5.5 Microsoft Power BI

Společnost Microsoft se již od 80. let minulého století řadí mezi nejvýznamnější technologické firmy světa. Přestože měl Microsoft v portfoliu „on-premises“ BI řešení již od roku 1998, kdy představil své OLAP Services jako součást SQL Server 7, podstatné modernizace a cloudové implementace se široká veřejnost dočkala až v polovině roku 2015.

Nástroj zvaný Power BI je založen na cloudovém přístupu a částečně také desktopovém rozhraní (Power BI Desktop). Gartner jej v rámci svého Magic Quadrantu<sup>120</sup> řadí mezi absolutní lídry trhu, a to především díky silnému vzestupu díky poslední vydané verzi Power BI, velkému tržnímu podílu, čisté vizi budoucích produktů a také přítomnosti dotazování v přirozeném jazyce (anglickém).

### 5.5.1 Koncept

Na rozdíl od GD je Power BI prezentována jako cloudová BI služba (SaaS) s vysokou mírou samoobslužnosti, která se skládá ze všech komponent nezbytných pro integraci, analýzu a vizualizaci dat, ale nenabízí jednotlivé části k samostatnému použití. Jedinou

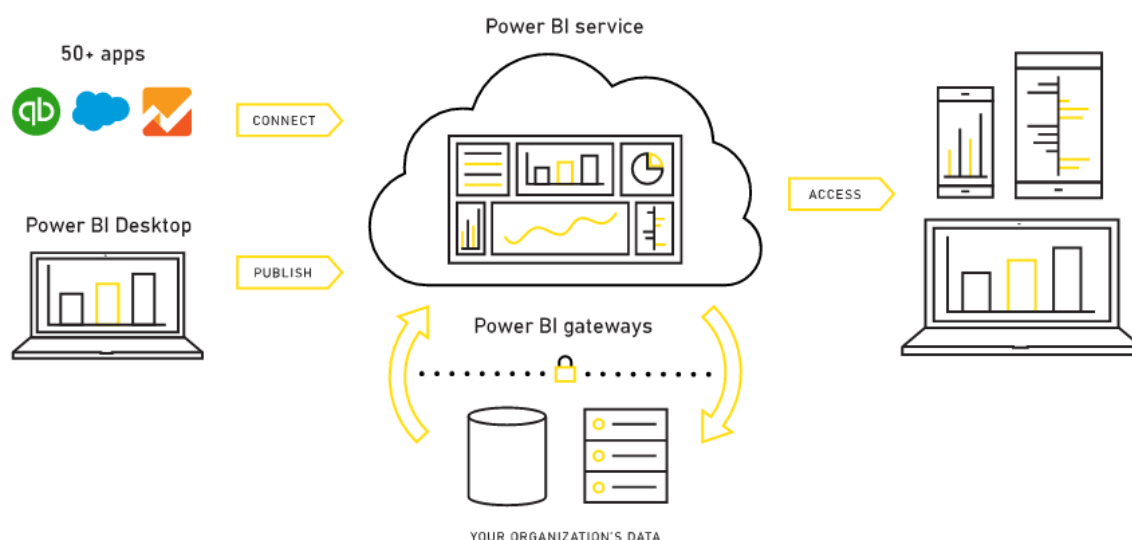
---

<sup>119</sup> Pricing. *GoodData* [online]. San Francisco: GoodData Corporation, 2010 [cit. 2016-12-07]. Dostupné z: <http://web.archive.org/web/20100811023135/http://www.gooddata.com/pricing/>

<sup>120</sup> *Magic Quadrant for Business Intelligence and Analytics Platforms* [online]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>.

komponentou, kterou lze využít samostatně, je vizualizační/prezentační vrstva, a to při přímém napojení na vlastní databázi nebo datový sklad.

Přestože není nikde přesné fungování platformy Power BI jasně uvedeno, je možné z některých zdrojů vydedukovat, že je celý systém postaven na standardní analytické službě Microsoft SQL Server Analysis Services Tabular v rámci programu Azure. Datové skladování je v rámci licencování služby zásadně omezené – takové možnosti, jaké nabízí GD pro ukládání Big Dat, zde dostupné nejsou. Jednoduché schéma viz obrázek 5-14.



Obrázek 5-14: Power BI – schéma platformy

Zdroj: What is Power BI? *Microsoft Power BI* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-20]. Dostupné z: <https://powerbi.microsoft.com/en-us/what-is-power-bi/>

## 5.5.2 Integrace dat

Koncepce Power BI je v porovnání s GD značně rozdílná. Microsoft totiž Power BI nabízí ve dvou formách: buď každá z nich může fungovat zcela nezávisle, nebo spolu obě dokáží spolupracovat. Jelikož lze získat celkově použitelnou (a především cloudovou) BI platformu pouze při současném použití obou, je nezbytné se zabývat každou z nich.

### 5.5.2.1 Power BI (online)

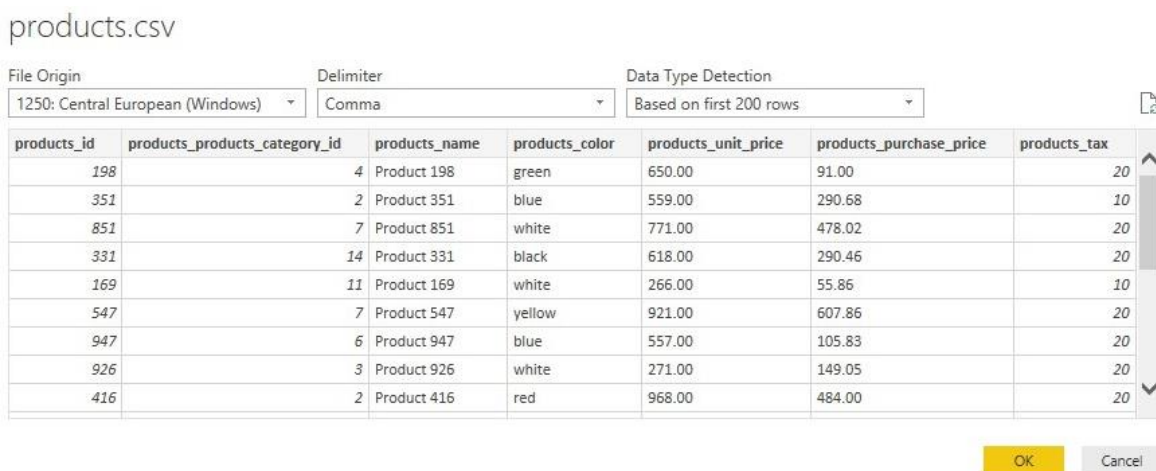
Hlavním bodem zájmu z pohledu moderních cloudových BI je online verze Power BI, někdy také nazývaná Power BI Service. Toto webové rozhraní má být považováno za stěžejní z celé

analytické platformy, ale v celém ekosystému slouží spíše pouze jako místo pro konzumaci a sdílení dat.

Navzdory tomu je přímo z webového rozhraní nabízeno načtení dat do platformy z cca 70 cloudových služeb, místních či vzdálených souborů v úložišti OneDrive nebo některých databázových služeb poskytovaných v rámci Microsoft Azure. To se zdá být poměrně solidním spektrem datových zdrojů pro poskytnutí potřebných dat k následné analýze, avšak webové rozhraní trpí zásadním nedostatkem. Připojená data nelze nikterak upravovat, modelovat ani spojovat. V tomto případě je celý nástroj možné použít pouze pro zkoumání samotných surových dat z daného zdroje bez širšího kontextu.

### 5.5.2.2 Power BI Desktop

Pro spojení dat z různých zdrojů je nutné začít projekt vytvářet v desktopové aplikaci Power BI Desktop. Ta umožňuje připojení k několika typům místních souborů, ale oproti online prostředí je zde výrazně rozšířena podpora databází. Naopak razantně omezena je nabídka cloudových služeb (cca 30); podpora těch nejznámějších (Facebook, GA, Salesforce) však zůstala zachována. Zajímavým zdrojem dat je obyčejná webová stránka, kde při zadání URL adresy stránky s nějakými tabulkami v HTML, program automaticky všechny tabulky na stránce rozpozná a nabídne je k načtení do datového modelu. Lze tak například jednoduše načíst různé seznamy či statistiky třeba ze stránek Wikipedie.



Obrázek 5-15: Power BI Desktop – dialog načtení souboru CSV

Zdroj: vlastní

Načtení testovacích CSV souborů bylo velmi intuitivní. Jak dokládá obrázek 5-15, systém automaticky rozpoznal jak kódování souboru, tak použitý oddělovač i uvozovací znaky (ty lze ručně změnit). V rámci testování proběhla i zkouška načtení dat z cloudové služby GitHub. I v tomto případě bylo uživatelské prostředí na prvotřídní úrovni. Po zadání přihlašovacích údajů systém ihned nabídne tabulky, které je možné využít; po potvrzení vybraných jsou ihned zvolené tabulky přidány do datového modelu.

Stejně jako GoodData Cloud Connect Designer i Power BI Desktop umožňuje data před použitím upravit. K tomuto účelu zde slouží nástroj zvaný Power Query, původně vyvinutý jako doplněk pro MS Excel. Jedná se o poměrně jednoduchý a uživatelsky opět velmi přívětivý nástroj, který pomocí řady běžných funkcí dokáže data základním způsobem transformovat. Sloupce je možné slučovat, rozdělovat, vytvářet odvozené nebo je agregovat, či spojovat celé tabulky atd. Na výrazně složitější transformace ale pravděpodobně nebude zcela vhodným nástrojem.

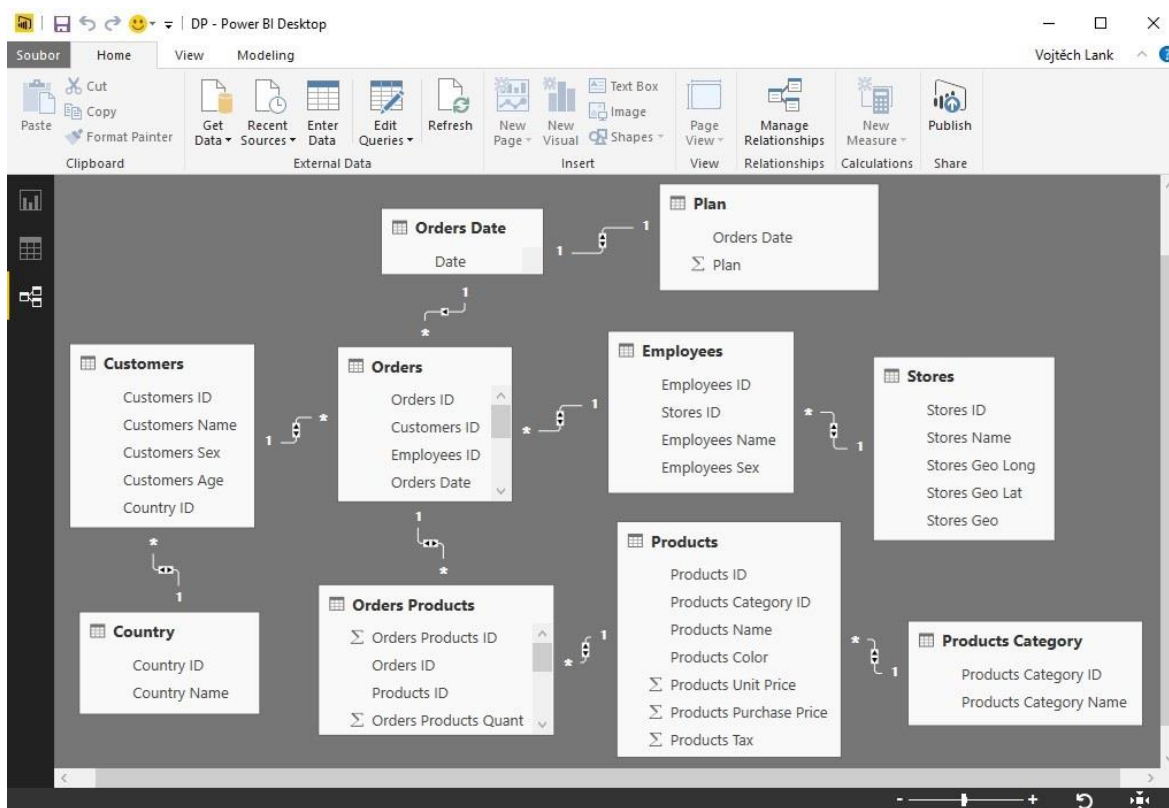
Po vytvoření všech datových připojení a transformaci dat jsou všechna data načtena do datového modelu přímo v aplikaci. Při práci s malými objemy dat není co vytknout; pokud však obsahují zdrojové databáze několik GB dat, bude práce s takovým objemem na běžném počítači velmi obtížná, ne-li nemožná. Z tohoto pohledu se zdá být zvolený koncept dosti limitujícím. Jistě by nebyla od věci možnost práce pouze s datovým modelem bez dat nebo načtení malého vzorku dat, který by tvořil jen zlomek z celkového objemu. Zpracování by se tak značně urychlilo.

### **5.5.2.3 Power BI Data Gateway**

Pozdější automatické aktualizace dat jsou možné pouze pomocí další podpůrné aplikace, Power BI Data Gateway. Tuto aplikaci je nutné spustit na serveru Windows, který má přístup ke všem místním zdrojům dat. Aplikace vytvoří virtuální privátní spojení mezi daným serverem a sítí Microsoft, kam jsou data z daných zdrojů odeslána ke zpracování. Výhodou tohoto řešení je, že všechny místní datové zdroje mohou zůstat zabezpečené v rámci lokální uzavřené sítě. Vzniká tak nižší riziko útoku než v případě přímého vystavení, jako je tomu u GD.

### 5.5.3 Datový model

Přístup Microsoftu se k datovému modelování v BI v posledních letech výrazně změnil, neboť začal oproti dříve značně rozšířenému multidimenzionálnímu OLAP datovému modelu více prosazovat vlastní tzv. tabulární model (Tabular). Jako důvod MS uvádí jednodušší definici a sestavení datového modelu, zejména pokud v pozadí stojí normalizovaná relační databáze. Data není potřeba transformovat do podoby star-schémat složených z klasických faktových tabulek a dimenzí, jak je tomu u multidimenzionálního modelování.<sup>121</sup> Pro vytvoření tabulárního datového modelu není nutné původní relační model nijak měnit – koncept se podobá tomu u LDM v případě GoodData.



Obrázek 5-16: Power BI Desktop – datový model

Zdroj: vlastní

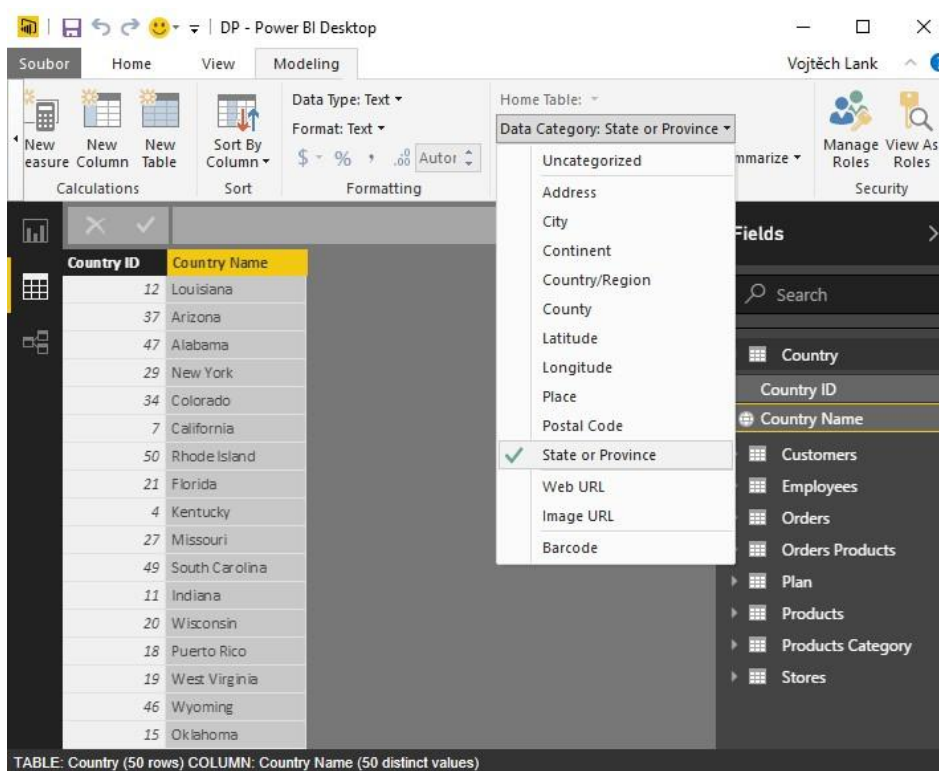
Po přidání datového zdroje do modelu proběhne automatická kontrola na možné vazby (funguje na základě stejných názvů atributů). Pokud jsou nějaké vazby detekovány, systém

<sup>121</sup> Tabular Model Databases (SSAS Tabular). *Microsoft Developer Network* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-01]. Dostupné z: <https://msdn.microsoft.com/en-us/library/hh230906.aspx>



je automaticky vytvoří. Jelikož byly během vstupní transformace standardizovány názvy atributů a referenční atributy (např. *Customers – Country ID*) se tak shodovaly s těmi klíčovými (*Country – Country ID*), byly během testování všechny vazby automaticky detekovány. Načtený datový model vč. vazeb zachycuje obrázek 5-16.

Další práce s datovým modelem je značně podobná práci s aplikací MS Excel. Panel nástrojů i chování se drží stejných zásad. U každého atributu je dále nutné určit datový typ (pokud nebyl automaticky správně rozpoznán), výchozí agregační funkci, formátování při zobrazení a v některých případech i tzv. kategorii dat. Ta se používá k označení atributů majících určitý speciální význam, který je využíván při zobrazení. Příkladem datové kategorie je název země, státu nebo zeměpisná šířka či délka. Nastavení kategorie dat v rámci testování znázorňuje obrázek 5-17.



Obrázek 5-17: Power BI Desktop – nastavení datové kategorie

Zdroj: vlastní

Hierarchie atributů jsou i zde podporovány, ale oproti GD existuje v Power BI omezení pouze na atributy v rámci jedné tabulky, což využitelnost při operacích drill-down citelně omezuje.

## 5.5.4 Datumová dimenze

Power BI v rámci vytváření datového modelu nenabízí vložení žádné datumové dimenze. Umí totiž později pracovat jako s datumovou dimenzí s každým atributem označeným typem datum. Práce s touto automatickou dimenzí je sice poměrně jednoduchá, ale v základní formě dosti omezená. Dimenze nabízí pouze kalendářní rok dělený na kvartály, měsíce a dny, což tvoří jednu hierarchii. Kalendářní týdny nejsou podporovány. Pokud je požadována specifická forma datumové dimenze, je nutné vložit ji do datového modelu ručně a následně z atributů sestavit požadované hierarchie. Na některé prvky vlastní datumové dimenze pak ale nelze aplikovat časové funkce dostupné v dotazovacím jazyce.

Dalším omezením je nemožnost současného řezu dat z více tabulek automatickou datumovou dimenzí. Příkladem z testovacího modelu je porovnání plánu s dosaženými tržbami. Obě tabulky obsahují datumový atribut *Orders Date*, ale data nelze současně zobrazit na jedné časové ose. Pro tyto případy je nutné vytvořit samostatnou tabulku obsahující seznam dat, která bude sloužit jako společná datumová dimenze pro obě zmíněné tabulky. V tomto případě byla vytvořena tabulka *Orders Date* s atributem *Date*. Potom je možné využít funkce automatické dimenze pro více faktových tabulek.

## 5.5.5 Metriky

Pro vytváření metrik používá Power BI dotazovací jazyk DAX (Data Analysis Expressions), který byl opět původně vyvinut v rámci doplňku pro MS Excel – PowerPivot a dnes je nejpodporovanějším analytickým jazykem napříč produkty firmy Microsoft.

V následující tabulce jsou opět uvedeny definice testovacích metrik, stanovených dříve (viz tabulka 1). Jazyk DAX se snaží částečně těžit z definovaných relací v datovém modelu, ale u některých operací je stále nutné specifikovat, že jde o atribut z jiné tabulky. Příkladem je výpočet metriky *Revenue*, kde cena produktu musí být do výpočtu přidána pomocí klíčového slova *RELATED*. Z pohledu nutnosti odkazovat na jiné tabulky se tak jazyk MAQL jeví jako dotazovací jazyk mnohem vyšší úrovně.

Tabulka 3: Testovací metriky (Power BI – DAX)

Název metriky	DAX
# Orders	# Orders = COUNT(Orders[Orders ID])
# Units	# Units = SUM(Orders Products[Orders Products Quantity])
Revenue	Revenue = SUMX(Orders Products,Orders Products[Orders Products Quantity] * RELATED(Products[Products Unit Price]))
Revenue M	Revenue M = CALCULATE([Revenue];Customers[Customers Sex] = "M")
Revenue F	Revenue F = CALCULATE([Revenue];Customers[Customers Sex] = "F")
AVG Order Revenue	AVG Order Revenue = AVERAGEX(Orders,[Revenue])
RUNAVG7 Order Revenue	RUNAVG7 Order Revenue = CALCULATE([AVG Order Revenue];DATESINPERIOD(Orders[Orders Date],[Date];LASTDATE(Orders[Orders Date],[Date]);-7;DAY))
Legenda: Fakt, Metrika, Atribut, Hodnota atributu, Tabulka	

Zdroj: vlastní

Jako velmi složitý a nepřehledný se zdá být aparát pro práci s časovou osou. I přes dostupnost široké škály funkcí, souhrnně označovaných jako „Time Intelligence Functions“,<sup>122</sup> jsou některé implementace matoucí a neintuitivní. Týká se to zejména funkcí pro výběr dat v určitém časovém období (DATESINPERIOD a DATESBETWEEN), které jsou natolik komplikované, že bez různých pomocných opatření není možné vytvořit tak elementární dotaz, jako například vyfiltrovat data k měsíci, který se nachází 2 období před obdobím aktuálním. Pro porovnání, v jazyce MAQL by takováto metrika vypadala následovně: *SELECT Revenue WHERE Month/Year (Orders Date) = THIS - 3*. V jazyce DAX tuto metriku za pomoci standardních funkcí s klíčových slov sestavit nelze.

Zajímavá je i definice metriky pro výpočet klouzavého průměru. Zatímco v GD pro klouzavý průměr přímo existuje funkce, zde je jeho výpočet nutné konstruovat pomocí základních časových funkcí. Kdyby vzniknul požadavek na tříměsíční klouzavý průměr, v GD by tato změna znamenala pouze změnu časového intervalu a řezového atributu s vyšší agregací (měsíc oproti dnu). V Power BI takovou metriku s použitím dostupných funkcí sestavit nelze.

<sup>122</sup> Time Intelligence Functions (DAX). *Microsoft Developer Network* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-02]. Dostupné z: <https://msdn.microsoft.com/en-us/library/ee634763.aspx?f=255&MSPPErrror=-2147217396>

Rozdílný zápis se týká též agregačních funkcí. Pokud je požadavek na nucenou agregaci faktu k některému atributu, v GD pro to slouží klíčové slovo „BY“ (*AVG(SELECT Revenue BY Orders ID)*), zatímco DAX pro podobnou konstrukci používá odlišné agregační funkce, které mají na konci názvu písmeno „X“. Stejný výsledek tak DAX vyprodukuje při použití funkce „AVERAGEX“, která však přijímá jako první argument konkrétní tabulku, nikoliv libovolný atribut, jako je tomu v GD, k níž je agregace vztažena (*AVERAGEX(Orders;[Revenue])*).

Z hlediska statistických funkcí DAX standardně neumožňuje možnosti výpočtu korelace či kovariance, naproti tomu však nabízí několik pravděpodobnostních rozdělení (beta, chí kvadrát, ...).

Metriky v Power BI nelze třídit do složek, určuje se pouze tzv. „mateřská tabulka“ (čili informace, v rámci seznamu atributů které tabulky má být metrika zobrazena). Existuje zde také pouze jeden typ metrik (globální), ale oproti GD je zde možné vytvářet tzv. vypočtené sloupce či celé tabulky (které mají stejný zápis jako metriky, ale v okamžiku vytvoření jsou ihned vypočteny a zapsány do datového modelu). Pozitivem může být vyšší rychlost, negativem nižší flexibilita než u „živých metrik“. Vypočtené sloupce také zabírají prostor.

Celkovou nevýhodou Power BI je, že metriky se dají tvořit pouze v desktopovém návrhovém prostředí. Je tím znatelně omezena samoobslužnost, neboť při jakémkoliv požadavku na novou metriku musí uživatel kontaktovat správce modelu. Avšak při zmíněné složitosti syntaxe DAX jde pravděpodobně o opodstatněný přístup.

Teprve po vytvoření všech potřebných metrik je vhodné projekt odeslat do cloudového prostředí Power BI, kde může být dále sdílen a využíván širším spektrem uživatelů. Soubor projektu Power BI Desktop s metrikami se nachází v příloze DVD 3.

## 5.5.6 Vizualizace

Tvorba vizualizací je možná jak v desktopovém návrhovém prostředí, tak v tom webovém, kde jsou obě téměř zcela identická (vzhledem i dostupnými funkcemi). Nabídka grafů je

znamení a nechybí žádná z běžně známých vizualizačních technik – viz obrázek 5-18. Paletu lze dále rozšířit pomocí doplňkových vizualizací, které jsou dostupné ke stažení na webu Power BI<sup>123</sup>; eventuálně si lze dle dokumentace vytvořit i vlastní.



Obrázek 5-18: Power BI – základní paleta vizualizačních nástrojů

Zdroj: vlastní

Power BI oproti GD disponuje integrovaným prostředím pro zpracování skriptů v jazyce R<sup>124</sup>, což otevírá mnohem širší možnosti statistického zkoumání. Již hotové skripty nebo vizualizace založené na jazyce R je možné získat z uvedené knihovny volně stažitelných doplňkových vizualizací (jakými jsou například korelační tabula, shluková analýza či predikční modely). Skripty je samozřejmě možné psát i vlastní (za podpory více jak 350 standardních knihoven). Bohužel R-skripty jsou v Power BI Desktop podporovány pouze po dodatečné instalaci běhového prostředí a v online verzi, následně kvůli bezpečnosti jen doménově ověřeným uživatelům.<sup>125</sup>

Práce s výchozími zobrazovači je obecně na prvotřídní úrovni. Výběr zobrazovaných faktů i řezových či filtračních atributů probíhá vždy formou drag&drop ze seznamu tabulek obsahující všechny dostupné atributy, hierarchie a metriky. Vytvoření požadovaného grafu je proto mírně rychlejší než v GD.

Z pohledu moderního grafického designu je na tom Power BI o mnoho lépe než GD. Grafy působí čistším a svěžejším dojmem. Širší jsou také možnosti nastavení či volby barev.

---

<sup>123</sup> Visuals library. *Power BI* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-03]. Dostupné z: <https://app.powerbi.com/visuals/>

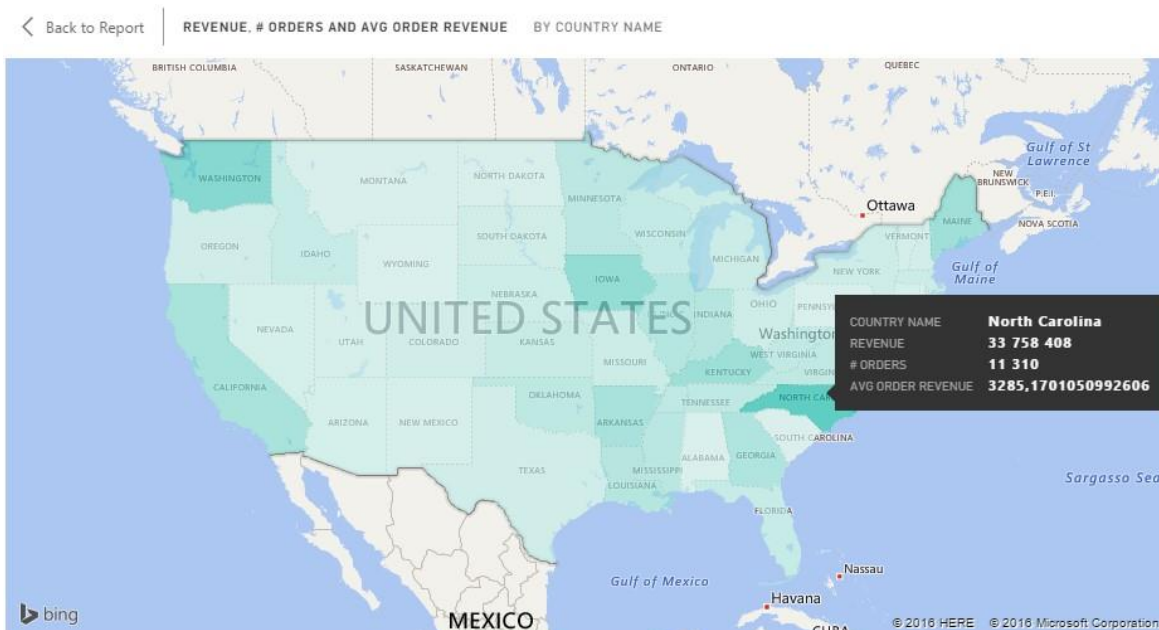
<sup>124</sup> R je volně šiřitelný, multiplatformní, programovací jazyk a prostředí určené pro statistickou analýzu dat a jejich grafické zobrazení. Dnes je jedná o faktický standard v řadě oblastí statistiky a analytiky. Více viz <https://www.r-project.org/>

<sup>125</sup> Creating R visuals in the Power BI service. *Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-03]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-r-visuals/>

U tabulek je možné vybírat i z více grafických stylů, které lze dále libovolně upravovat. Možnosti změny grafického vzhledu jsou tak celkově na opravdu vysoké úrovni.

### 5.5.6.1 Mapy

V rámci vizualizace geograficky vztažených dat má Power BI opravdu hodně co nabídnout. Dostupné jsou dvě základní zobrazení a také profesionální mapové podklady ArcGIS od společnosti Esri. Obrázek 5-19 zachycuje základní zobrazení počtu objednávek dle státu zákazníka. Zobrazení probíhá na mapách Bing, které jsou přehledné a podrobné.

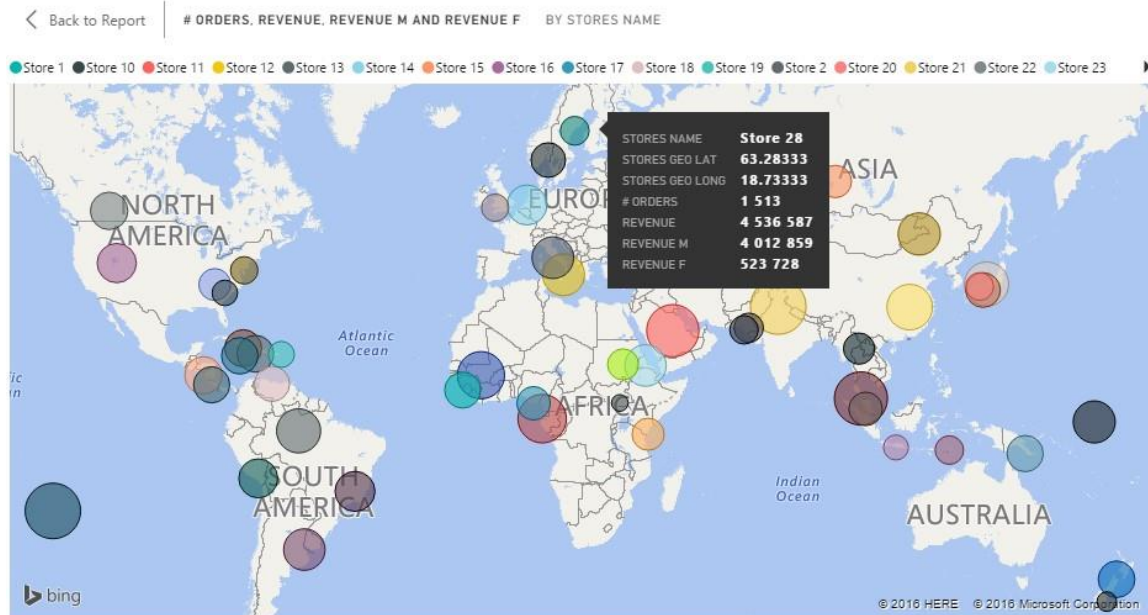


Obrázek 5-19: Power BI – mapa dle státu zákazníka

Zdroj: vlastní

Na následujícím obrázku se nachází druhé ze základních typů zobrazení – zobrazení podle geografických souřadnic. Na výběr je ze zobrazení kruhů s měnící se světlostí, velikostí nebo z obou současně. Data lze přímo kategorizovat, což se projeví barevnou odlišností každé kategorie.

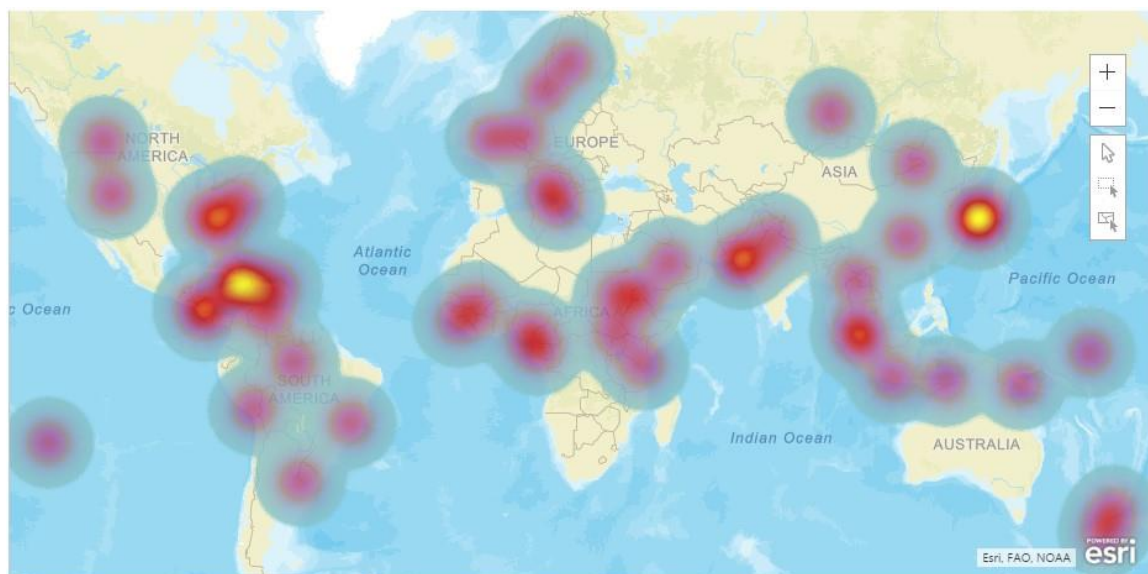




Obrázek 5-20: Power BI – mapa dle souřadnic obchodu

Zdroj: vlastní

Poslední možností jsou profesionální mapy ArcGIS. Ty nabízí několik typů podkladových map (světlá, tmavá, barevná) či typů zobrazení (pouze bod, velikost, barva, teplotní mapa, shlukování). Kromě teplotních map (obrázek 5-21) nebo shlukování lze zobrazení doplnit o některou ze široké nabídky referenčních vrstev, jimiž jsou např. světová města, celosvětová demografie či náboženská vyznání. Vzhled map lze také velmi široce upravovat.



Obrázek 5-21: Power BI – teplotní mapa Esri ArcGIS

Zdroj: vlastní

### 5.5.6.2 KPI

Klíčové indikátory výkonnosti lze v Power BI zobrazit dvěma způsoby. Prvním z nich je přímo určená vizualizace s názvem „KPI“, která dokáže zobrazit vývoj dané metriky v porovnání s metrikou představující cílové (plánované) hodnoty – viz obrázek 5-22. Bohužel tato komponenta nedokáže s hierarchií datumové dimenze pracovat na nižší úrovni než je rok, což její použitelnost značně omezuje. Pro zobrazení hodnot pro aktuální čtvrtletí či měsíc je proto potřeba použít statické filtry či vytvořit samostatné metriky.



Obrázek 5-22: Power BI – vizualizace KPI

Zdroj: vlastní

Jako druhý způsob se nabízí použití vizualizace měřidla (angl. gauge). Ta je více než k zobrazení odchylky od plánu vhodná ke grafickému znázornění jeho naplnění. Nepodporuje ale osu trendu, tudíž zde je použití vyhrazených metrik nutností.

V porovnání s KPI dashboardem v GD jsou možná vizualizace v podání Power BI (díky trendovému grafu v pozadí) více atraktivní, ale zobrazení vývoje metriky oproti předcházejícímu období je zde mnohem složitější. GD naopak neumožňuje přímé porovnání oproti metrice představující plán, které je možné pouze pomocí grafu typu „bullet“.

Emailová upozornění v případě překročení hraniční hodnoty jsou i zde podporována, ale opět je nutné meze nastavit manuálně.

### 5.5.6.3 Reporty a dashboardy

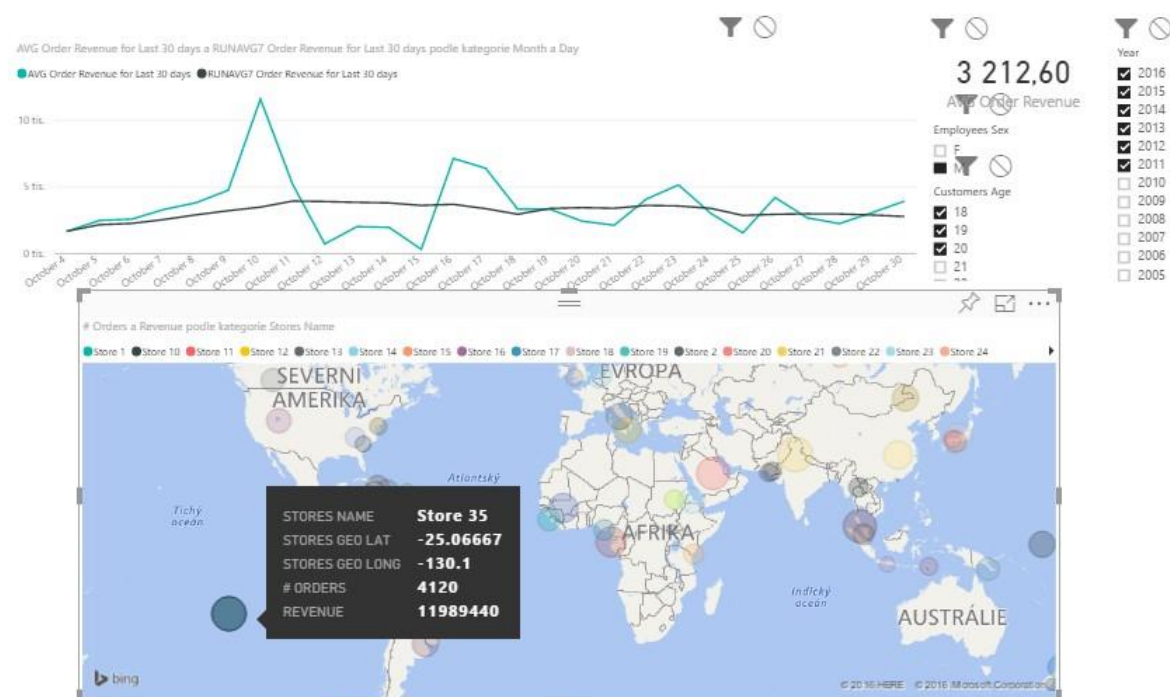
Power BI přistupuje k otázce reportů a dashboardů svébytným způsobem. Existují zde dvě místa pro ukládání a zobrazování vizualizací – sestavy a řídicí panely.



Základním prostředím jsou tzv. sestavy. Jedná se pracovní prostor, který je dělen pomocí stránek (záložek), přičemž na každou z nich je možné umístit libovolné vizuály. Dále je možné vkládat filtry, popisné texty, předdefinované obrazce a jiné. Vlastní obrázky lze vkládat pouze v případě, že je sestava vytvářena pomocí Power BI Desktop. Ve webovém rozhraní je umožněno pouze nastavit obrázek jako pozadí stránky sestavy.

V rámci sestav jsou ctěny tři typy filtrů podle úrovně, dle které jsou aplikovány: úroveň sestav, úroveň stránek a úroveň jednotlivých vizuálů. Jednotlivé filtry mohou být vytvořeny z libovolných atributů. Nabízí se opět také velmi široké vizuální nastavení.

Zajímavým prvkem sestav, resp. jejich jednotlivých stránek, jsou vzájemné automatické filtry. Pokud uživatel klikne na nějakou oblast jednoho z vizuálů, kde na příkladu – viz obrázek 5-23 – je na mapě vybrána kategorie „Store 25“, tak jsou podle daného výběru automaticky filtrovány výsledky všech okolních vizuálů. Toto chování lze upravit, pokud je u některé požadován statický výsledek.

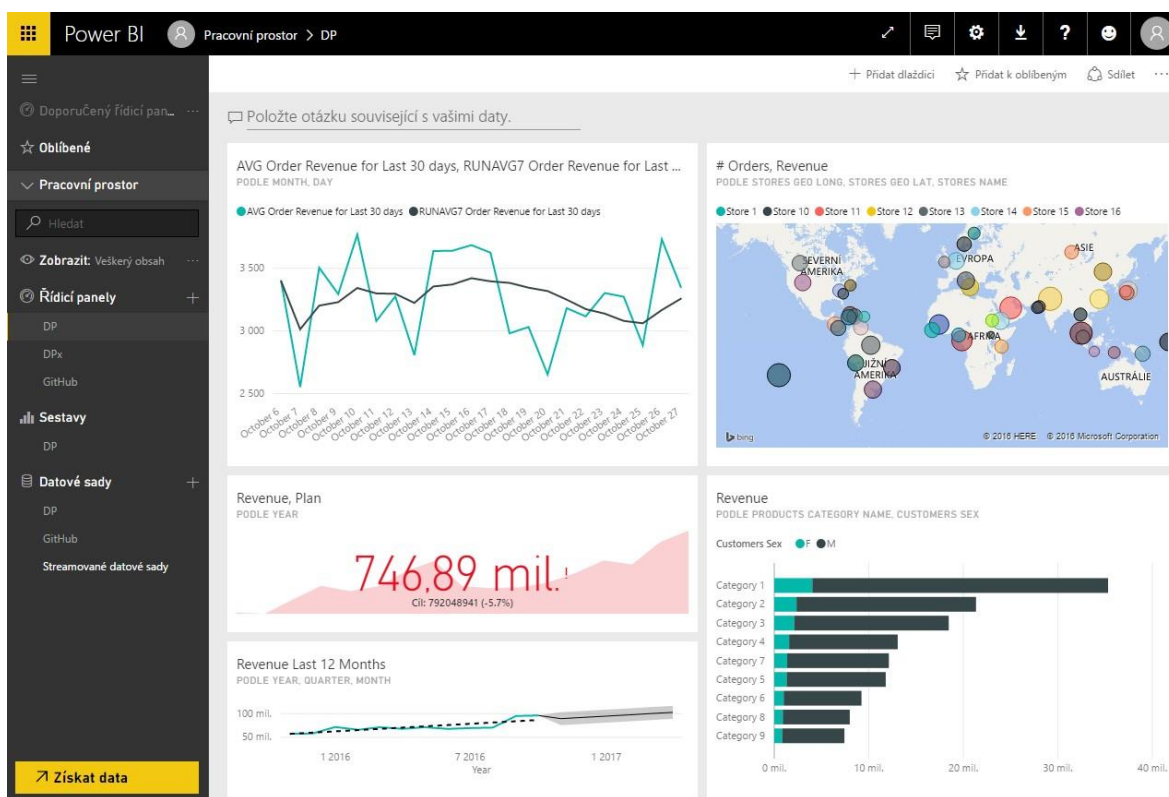


Obrázek 5-23: Power BI – vzájemná interakce v rámci jedné stránky sestavy

Zdroj: vlastní

Je vhodné také zmínit, že v desktopovém prostředí lze vytvořit pouze jednu sestavu, která je publikována společně s projektem. Naproti tomu ve webovém rozhraní je jich možné uložit hned několik.

Jednotlivé vytvořené vizualizace nebo celé stránky sestav lze poté umístit na souhrnný řídicí panel – viz obrázek 5-24. Jednotlivé objekty pak tvoří samostatné dlaždice, které je možné libovolně organizovat a měnit jejich velikost (obdobně jako u nové nabídky Start představené ve Windows 8). Panel lze dále doplnit o různý webový obsah, obrázek či video.



Obrázek 5-24: Power BI – řídicí panel

Zdroj: vlastní

#### 5.5.6.4 Změna detailu

Navzdory omezení hierarchií atributů pouze v rámci dané tabulky mají grafy podporu operací drill-down velmi dobře implementovanou. Buď je možné použít předdefinovanou hierarchii, nebo lze na hlavní osu pod sebe poskládat libovolné atributy (v rámci celého datového modelu), podle kterých lze následně měnit granularitu zobrazení třemi různými způsoby. Grafy nabízejí celkový propad na nižší úroveň (rok → měsíc), rozpad aktuální

úrovně dle nižší úrovně (rok → rok/měsíc) nebo propad na nižší úroveň dle vybrané kategorie aktuální úrovně.

### **5.5.7 Ad-hoc dotazy**

Otázku ad-hoc dotazování není v prostředí Power BI příliš jednoduché zodpovědět. Celý systém je totiž de-facto na principu ad-hoc dotazování postaven. Pro získání odpovědi na požadovanou otázku stačí zvolit datovou sadu, které se konkrétní úloha týká a během několika málo vteřin vytvořit libovolnou vizualizaci stejným způsobem jako při vytváření sestav.

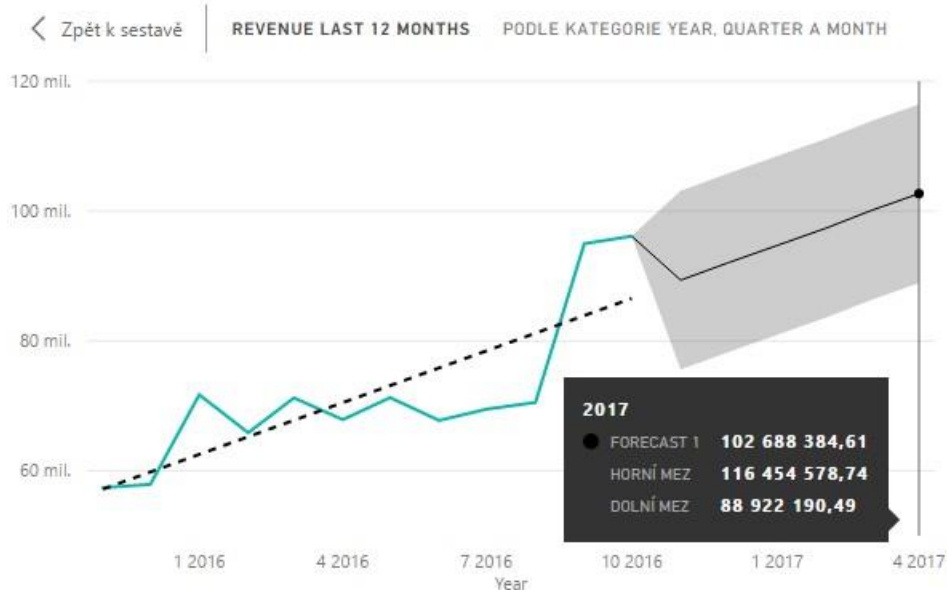
Díky přítomnosti drag&drop pro vytváření vizuálů je prostředí svou přívětivostí srovnatelné s GD a navíc na rozdíl od tohoto konkurenta není dotazování nijak omezeno.

Pro opravdu rychlé ad-hoc dotazování existuje v online verzi unikátní modul pro dotazování v přirozeném jazyce, o kterém bude více pojednáno v části 5.5.9 – Jedinečné vlastnosti.

### **5.5.8 Forecasting**

U standardních grafů existuje volba zobrazení doplňkových čar – minima, maxima, průměr nebo medián. V případě spojnicového grafu je dostupná i možnost spojnice trendu a předpovědi budoucího vývoje.

Prognózu je možné určit na libovolný počet budoucích období (třeba 100 let dopředu), a to i s určitým intervalem spolehlivosti, který je možné zvolit mezi 75 a 99 procenty. Předvolby dále umožňují vynechat několik posledních hodnot (např. aktuální měsíc) nebo nastavit počet období určujících sezónnost veličiny.



Obrázek 5-25: Power BI – forecasting

Zdroj: vlastní

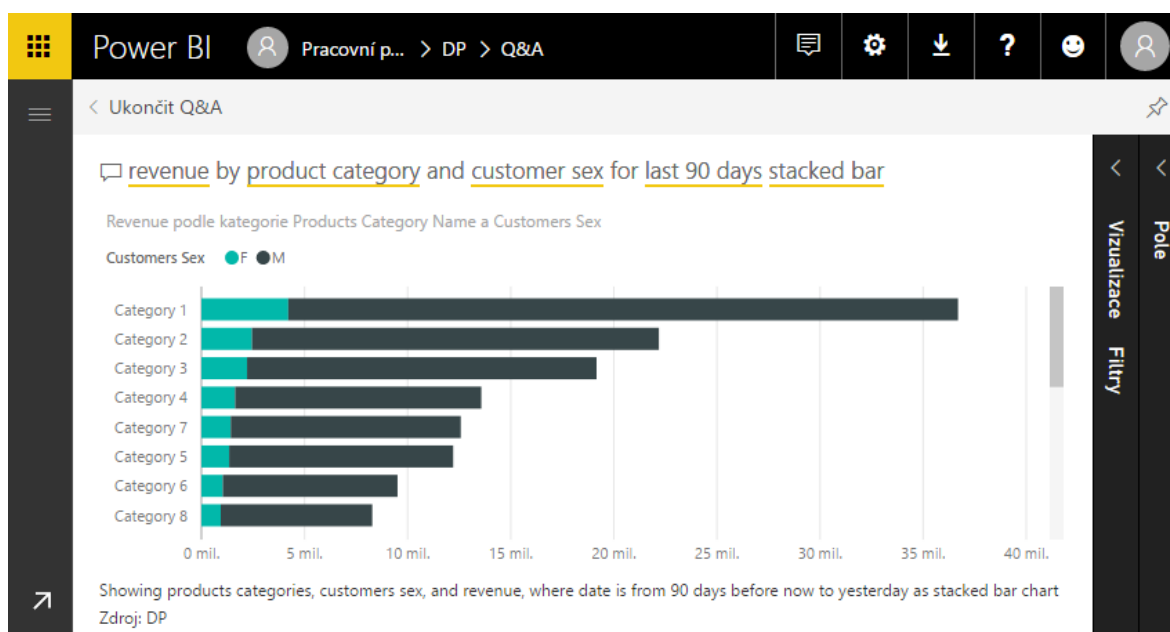
Obrázek 5-25 zachycuje předpověď vývoje tržeb na následujících 6 měsících v závislosti na 12 přecházejících. Sezónní složka byla ponechána na automatické detekci, což se ukázalo jako hodnota 1 a interval spolehlivosti činí 95 %. Graf obsahuje také spojnicí trendu, kterou však nelze prodloužit do budoucnosti.

Výpočet probíhá pouze na úrovni grafu, ani metoda výpočtu není z dokumentace patrná. Není tak možné získat predikované budoucí hodnoty ve formě tabulky, ani s těmito hodnotami jakkoliv dále počítat, což opět snižuje výslednou použitelnost této funkce.

### 5.5.9 Jedinečné vlastnosti

Microsoft do své cloudové BI platformy integroval několik zajímavých funkcí, které jsou na trhu buď ojedinělé, nebo zcela unikátní. Dozajista tou nejzajímavější je možnost přímého dotazování v přirozeném jazyce, zvané Q&A, založené na Microsoft Cortana. Do dotazového pole, umístěného na každém řídicím panelu, stačí napsat správně položený dotaz odkazující se na dostupné atributy, metriky, vizualizace či časová období, a ihned je zobrazen výsledek. Vše funguje překvapivě dobře a velmi rychle. Pokud není uživatel s výsledkem spokojen, může zobrazený výstup upravit do kýžené podoby pomocí nástrojů a datových polí stejným způsobem jako při vytváření standardní vizualizace.

Intuitivnost je na velmi vysoké úrovni, ale stále je potřeba, aby si uživatel osvojil celkové fungování, klíčová slova či konstrukce. Získání rychlého výsledku bez toho, aby uživatel trávil dlouhý čas vymýšlením, jak se na daný výsledek správně zeptat, je jinak poměrně obtížné. Prozatím je tato funkce sice dostupná pouze v anglickém jazyce, ale rozšíření podpory dalších bude jistě následovat.



Obrázek 5-26: Power BI – Q&A – dotazování v přirozeném jazyce

Zdroj: vlastní

Obrázek 5-26 ukazuje okno s dotazem na obrat dle kategorie produktu a pohlaví zákazníka za posledních 90 dní ve formě skládaného pruhového grafu. Takto vytvořený vizuál je následně možné připnout na řídicí panel.<sup>126</sup>

Další zajímavou funkcí, kterou Power BI disponuje, je možnost přímého dotazování do zdrojových databází, označované jako „DirectQuery“. Podporovány jsou jak databázové služby z dílen Microsoftu (SQL Server, Azure SQL, ...), tak i Oracle, Teradata nebo cloudové Amazon Redshift či Snowflake. Následná analýza probíhá tím způsobem, že při každém požadavku je odeslán nový dotaz do zdrojové databáze. Tím je sice zajištěna absolutní aktuálnost dat, ale dotazy mohou zdrojové databáze značně zatěžovat, což může

<sup>126</sup> Q&A in Power BI. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a/>

způsobovat pomalejší odezvy. Kromě tohoto problému má toto řešení i další výrazná omezení. Konkrétně na úrovni datového modelu, kde v tomto režimu se nelze dotazovat na data z více zdrojů současně.<sup>127</sup>

V souvislosti s datovými zdroji existuje další funkce umožňující vyšší aktuálnost dat. Jedním z podporovaných datových zdrojů je dokument Excel uložený v cloudovém úložišti OneDrive for Business nebo podnikovém serveru SharePoint. U takovýchto datových zdrojů může docházet k automatické aktualizaci, neboť server Power BI naslouchá ohledně změn v daných souborech. Jakmile je změna detekována, systém dokáže automaticky data aktualizovat. Produktové materiály tvrdí, že by aktualizace měla proběhnout maximálně do jedné hodiny od provedené změny.<sup>128</sup>

Power BI umí též naslouchat streamovaným datovým sadám v reálném čase. Do platformy tak lze přímo napojit výstupy z různých senzorů pomocí rozhraní REST/JSON nebo PubNub a zobrazovat touto cestou změny veličin v reálném čase.

V porovnání s testovanou platformou GoodData stojí opět za zmínku podpora jazyka R, což výrazně zvyšuje celkový analytický potenciál platformy.

### 5.5.10 Export dat

Ani v případě Power BI není možnost exportu dat opomenuta. Zobrazená data lze exportovat ze všech vizualizací, ale oproti GD je dostupný pouze formát aplikace Excel (XLSX) a CSV. Naopak zde nechybí možnost tisku panelu či sestavy, navíc se zde nachází funkce pro export celé sestavy (všech listů) ve formátu prezentace PowerPoint.

---

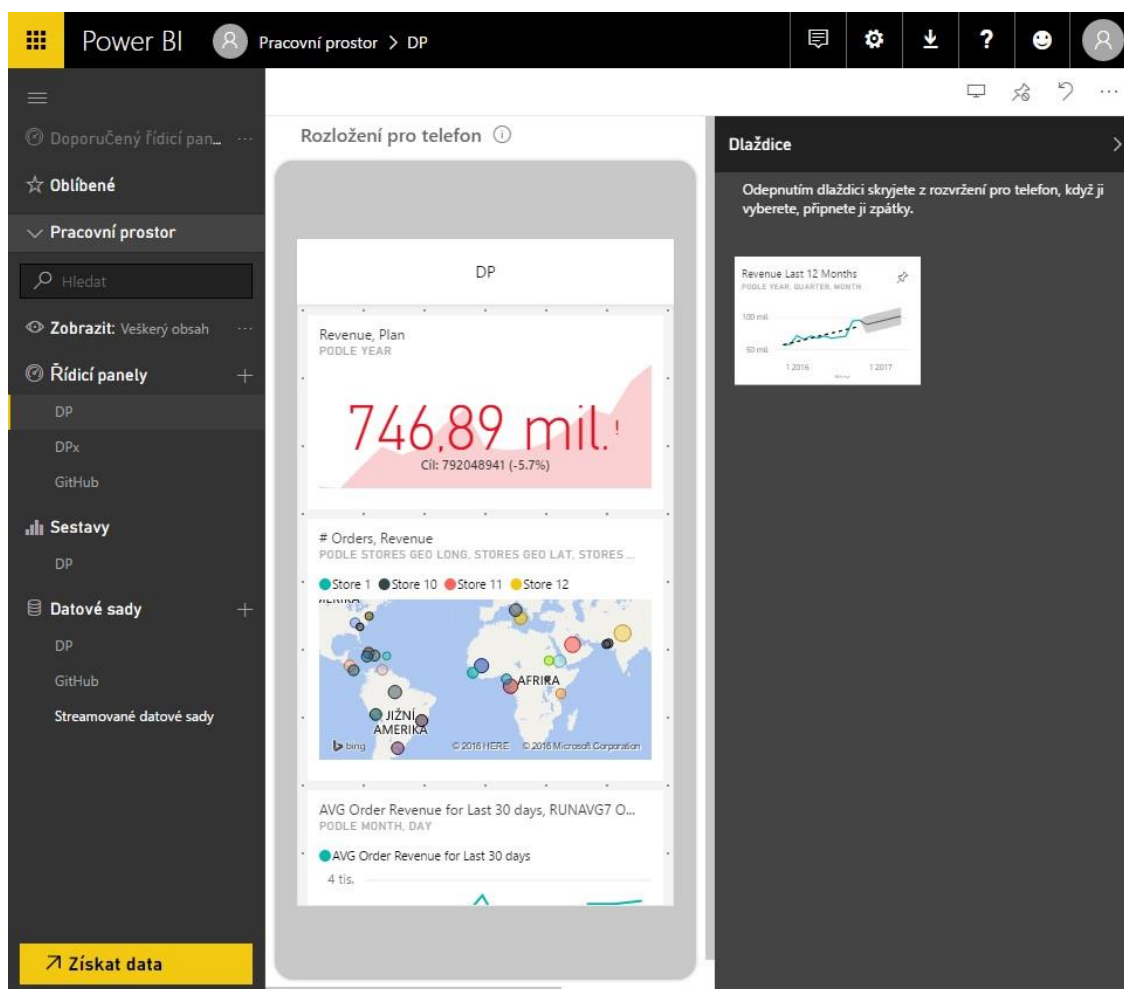
<sup>127</sup> Use DirectQuery in Power BI Desktop. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-use-directquery/>

<sup>128</sup> Refresh a dataset created from an Excel workbook on OneDrive, or SharePoint Online. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-refresh-excel-file-onedrive/>

## 5.5.11 Mobilní přístup

Vše nasvědčuje tomu, že trend mobilního BI je Microsoftu dobře znám a je na něj velmi dobře připraven. Webové rozhraní sice není pro použití na mobilním telefonu absolutně uzpůsobeno, ale zato jsou dostupné aplikace pro Android, Apple iOS i mobilní Windows 10.

Inhned zpočátku je nutné zmínit, že i přes jednoduchost návrhu mobilního zobrazení je právě tento proces značně nekonzistentní. V desktopové aplikaci lze pro mobilní aplikace navrhovat vzhled jednotlivých stránek sestav, avšak to samé není možné ve webovém prostředí. Tam je naopak možné tvořit pouze rozložení prvků řídicích panelů. Lze očekávat, že se tato nesourodost prostředí dočká opravy v některé z příštích aktualizací, ale prozatím to působí zmatečně.

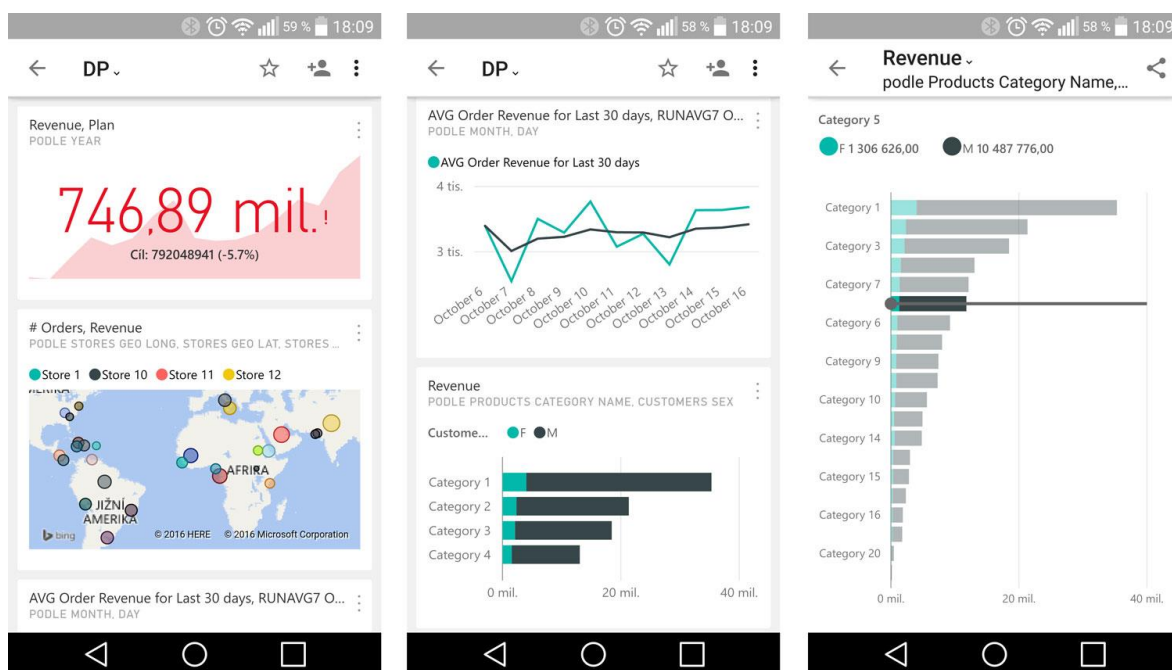


Obrázek 5-27: Power BI – návrh mobilního zobrazení řídicího panelu

Zdroj: vlastní

Samotný návrh je na obou místech velmi jednoduchý a intuitivní. Jak dokládá obrázek 5-27, na pomyslnou plochu mobilního telefonu stačí přetáhnout dostupné vizuály, změnit jejich velikost a uspořádat je do požadovaného vzhledu.

Následující obrázek zachycuje výsledné zobrazení řídicího panelu na mobilním telefonu s operačním systémem Android. Je patrné, že zobrazení je téměř absolutně shodné s návrhem, což je pozitivní zjištění. Dle dostupných informací jsou si aplikace pro všechny platformy velmi podobné, tudíž by se neměly funkcemi ani vzhledem zásadně lišit. Je však nutné dodat, že toto zobrazení se uplatňuje pouze pro mobilní telefony. Zobrazení sestav a panelů je na zařízeních typu tablet i přes rozhraní aplikace identické s tím webovým.



Obrázek 5-28: Power BI pro Android – řídicí panel

Zdroj: vlastní

Ovládání je v aplikaci velmi příjemné a pohyb po osách či kategoriích je taktéž zvládnutý na vysoké úrovni. Aplikace umožňuje pohodlné sdílení konkrétní vizualizace, kterou je díky vestavěnému poznámkovému režimu možné ještě před sdílením doplnit o popis či kresbu.



### 5.5.12 Sdílení a omezení přístupu

Sdílení obsahu uvnitř organizace je v Power BI založeno na uživatelských účtech a skupinách spojených s Office 365. Administrátor, který vytvoří prvotní projekt, musí před jeho sdílením nejprve vytvořit tzv. balíček obsahu (content pack), který se může skládat z různých projektů, zdrojů, panelů i sestav. Tento balíček je teprve pak možné sdílet buď v rámci celé organizace, nebo konkrétním uživatelům. Na rozdíl od GD, kde se po sdílení projekt ihned ostatním objeví v seznamu přiřazených projektů, si ho v Power BI musejí jednotliví spolupracovníci sami aktivně přidat z knihovny balíčků obsahu do svého prostředí. Tato ideologie činí každého uživatele sice více nezávislého a „svobodného“, ale zároveň činí sdílení obtížnějším.

Podpora RLS je zde oproti GD na prvotřídní úrovni. Filtry je možné definovat pro konkrétní uživatele nebo celé skupiny (role) uživatelů, a to pohodlně v prostředí Power BI Desktop. Zároveň je (díky možnosti spojení účtů Office 365 s adresářovou službou Active Directory) správa rolí a uživatelských přístupů kompaktní a může být řízena centrálně v rámci celé podnikové infrastruktury.

### 5.5.13 Rychlost

Po mírném zklamání z občasného čekání v GD bylo Power BI příjemným překvapením. S testovacím objemem dat si systém skvěle poradil a práce v online prostředí probíhala zcela hladce. Rychlost odezvy i vykreslování grafů či panelů tak nezbyvá než hodnotit jako výbornou.

Práce na datovém modelu a některých vizuálech v desktopovém návrhovém prostředí byla sice značně pomalejší, ale stále na přijatelné úrovni (práce probíhala na notebooku s průměrným výkonem<sup>129</sup>). Nicméně vytváření datového modelu z vícera objemnějších zdrojů (kde se všechna data musejí do modelu nejprve načíst – viz 5.5.2.2) pravděpodobně není zcela příjemné. Nemožnost tvorby datového modelu bez dat je značným nedostatkem, který naopak svědčí ve prospěch GD.

---

<sup>129</sup> Intel Core i5, 6 GB RAM, SSD

### 5.5.14 Cena a omezení

Posledním hodnotícím kritériem je cena. Power BI je nabízeno ve dvou verzích, a to „FREE“ (zdarma) a „PRO“ (placená). Neplacená verze sice obsahuje 1 GB datový limit a omezení některých vlastností, ale pro osobní či mikro-podnikové účely dokáže i přesto nabídnout velmi mnoho.

Verze „PRO“ stojí 9,99 amerických dolarů (cca 250 Kč<sup>130</sup>) za uživatele za měsíc při ročním předplatném, které tak činí cca 3 000 Kč. Tato licence obsahuje všechny dostupné funkce, avšak velmi znepokojující je limit 10 GB dat, který je uplatňován jak na jednotlivé uživatele, tak v rámci celé organizace. Power BI sice využívá pro dosažení co nejvyšší efektivity velmi vysoké komprese dat, ale i přes relativně nízkou cenu tak může být toto řešení pro některé firmy nepoužitelné z důvodu relativně nízkého datového limitů.<sup>131</sup> Mezi další omezení verze „PRO“ patří aktualizace dat pouze jednou za hodinu a 1 milion řádků/hodinu u streamovaných datových sad.

Je sice možné využít funkce DirectQuery, díky které je možné se připojit přímo do zdrojové databáze s libovolným objemem, ale to může znamenat výrazně pomalejší odezvy systému a (kvůli množství omezení) také celkově nižší užitek z jinak velmi propracované platformy. Budování nebo pronájem samostatného výkonného centrálního datového skladu, který by pro Power BI sloužil jako zdroj dat prostřednictvím DirectQuery, by bylo jistě možným řešením, ale celkové náklady na implementaci i provoz by se zvýšily.

## 5.6 Souhrnné porovnání kritérií

V rámci porovnání vybraných platforem proběhlo velmi detailní zkoumání téměř ze všech možných úhlů pohledu. Toto srovnání je možné strukturovaně shrnout pomocí následující tabulky, která obsahuje záznamy o hodnocení všech zmíněných kritérií.

---

<sup>130</sup> Dle kurzu ČNB k datu 8. 12. 2016

<sup>131</sup> Pricing. *Power BI* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-04]. Dostupné z: <https://powerbi.microsoft.com/en-us/pricing/>

Tabulka 4: Souhrnné srovnání GoodData a Power BI




Kritérium		Platforma	
		GoodData	Power BI
Integrace dat	Nástroj	Cloud Connect Designer	Power BI Service/ Power BI Desktop
	ETL	Ano, na bázi CloverETL	Ano, Power Query
	Soubory	Excel, CSV, XML, JSON	Excel, CSV, XML, Text, JSON
	Relační databáze (počet)	11	Service – 2, Desktop – 12
	Big Data databáze (počet)	4	Service – 1, Desktop – 3
	Cloudové služby (počet)	4 (v dokumentu zmíněno 40?)	Service – 70, Desktop – 30
	Jiné	Email, LDAP	AD, OData, Web, ...
Datový model	Označení	LDM	Tabular
	Hierarchie	Ano, napříč datovým modelem	Ano, ale pouze v rámci tabulky
Datumová dimenze	Kalendářní	Ano (R, Q, M, T, D)	Ano, ale pouze automatická (R, Q, M, D)
	Týdny	EU i US	Ne
	Vlastní (Fiskální)	Ano, ale složitá implementace nebo za pomoci GD	Ano, nutno vyrobit vlastní, funkce ale pouze pro kalendářní dimenzi
	Pohodlnost práce	Velmi pohodlné	Méně pohodlné
Metriky	Jazyk	MAQL	DAX
	Složitost	Jednoduchý	Složitější
	Funkce	Komplexní	Jednouúčelové
	Použitelnost	Velmi vysoká	Nižší, spousta věcí nejde jednoduše
	Uživatelské metriky	Ano	Ne
	Typy metrik	globální/lokální	globální/vypočtené sloupce
	Definice metrik	Správce, oprávněný uživatel	Správce (Power BI Desktop)
Vizualizace	Tabulka/Matice	Ano/Ano	Ano/Ano
	Grafy	Standardní + bulet chart	Nadstandardní, lze rozšířit z knihovny nebo vlastní, podpora R
	Vzhled	Průměrný	Moderní
	Úprava vzhledu	Základní	Pokročilá
	Interaktivita	Drill-in, tooltip	Drill-in, tooltip, vzájemné filtrování
Mapy	Podklad	Základní (neznámý)	Pokročilé (Bing), profesionální Esri ArcGIS
	Určení pozice	Souřadnice, stát, region ČR: KNOK, NUTS 4 (LAU 1)	Souřadnice, kontinent, stát, město, místo, adresa, ...
	Zobrazení	Region, bod	Region, bod, kruh, výsečový graf, teplotní mapa, shlukování, podkladové mapy
	Vzhled	Jednoduchý	Moderní
KPI	Použití	KPI dashboard, nebo pomocí vlastních metrik a vizualizace	Vyhrazené vizualizace
	Definice	Jednoduchá	Jednoduchá, ale omezení pouze na den nebo rok
	Vzhled	Velmi dobrý	Jednodušší

Kritérium	Platforma		
	GoodData	Power BI	
Práce s časovou osou	Práce s časovou osou	Jednoduchá	Složitá, vyhrazené metricky
	Plán vs. skutečnost	Bullet-chart	KPI, měřidlo (gauge)
	Upozornění emailem	Ano	Ano
Dashboard	Název	Dashboard	Řídící panel
	Záložky	Ano	Ne
	Způsob použití	Umístění reportů na dashboard	Vizualizace nebo celé listy sestav jako plovoucí dlaždice
	Úprava vzhledu	Libovolná ale složitá	Žádná
Reporty	Způsob práce	Pouze jeden vizuál na report	Ano (sestavy), více vizuálů na sestavu i více stran sestav
	Úprava vzhledu	Základní (viz vizualizace)	Pokročilá
Změna detailu	Drill-down	Ne	Ano
	Změna detailu	Ano, dle hierarchie nebo definice v rámci reportu či dashboardu	Ano, dle hierarchie nebo definice v rámci vizualizace (3 typy rozpadu)
Ad-hoc		Ano, vlastní prostředí	Ano, celý systém je de-facto ad-hoc
Forecasting		Formou metricky – lineární regrese, jednoduchý a vážený klouzavý průměr	Pouze neznámým výpočtem v rámci spojnicového grafu, nebo pomocí R
Jedinečné vlastnosti		Statistické funkce, import vlastních dat, dashboard emailem	Q&A, DirectQuery, R, streamované datové sady, automatická aktualizace z OneDrive
Export dat	Data z vizualizace	PDF, PNG, XLS, XLSX, CSV	XLSX, CSV
	Tisk	Tisk reportu, dashboard do PDF	Tisk panelu i sestavy, celá sestava jako PowerPoint prezentace
Mobilní přístup	Mobilní verze webu	Ne	Ne
	Android	Ne	Ano
	Apple iOS	Ano, ale zobrazení je stejně jako na PC	Ano
	Windows Phone	Ne	Ano
Sdílení		Ano, účty GD	Ano, účty Microsoft (Office 365)
Omezení přístupu (RLS)		Ano, ale nastavení jen přes API	Ano, jednoduché nastavení
Rychlost		Občas pomalejší	Velmi rychlé
Omezení		Nepodstatná	10 GB dat Aktualizace 1x za hodinu 1M ř./hodinu u streamovaných dat
Cena	Zdarma	Ne, pouze zkouška na 30 dní	Ano, omezené funkce, 1 GB
	Placená verze	Neznámá (dříve od \$ 500/měsíce/1 GB)	\$ 9.99/uživatel/měsíc/ 10 GB

Zdroj: vlastní

Z uvedené tabulky i z předcházejícího textu je patrné, že předpoklad o značné rozdílnosti platforem se zcela potvrdil. Jednotlivé systémy se více či méně odlišují téměř ve všech oblastech. Kvůli tolika rozdílným parametrům není zcela možné určit, který ze systémů je lepší či horší. Takovéto závěry je možné aplikovat pouze v rámci jednotlivých samostatných hodnotících kategorií (jejichž souhrnné hodnocení obsahuje tabulka 5), ale nikoliv na celkový výzkum.

Tabulka 5: Souhrnné srovnání GoodData a Power BI dle kritérií

Kritérium	Platforma	
	GoodData	Power BI
Integrace dat	✓	✓
Datový model	✓	–
Datumová dimenze	✓	✗
Metriky	✓	✗
Vizualizace	–	✓
Mapy	✗	✓
KPI	–	–
Dashboard	✓	✓
Reporty	✓	✓
Změna detailu	–	✓
Ad-hoc	–	✓
Forecasting	–	–
Jedinečné vlastnosti	–	✓
Export dat	✓	–
Mobilní přístup	✗	✓
Sdílení	✓	✓
Omezení přístupu (RLS)	✗	✓
Rychlost	–	✓
Omezení	✓	✗
Cena	✗	✓
<p>  Převyšuje konkurenta nebo na stejné dobré úrovni  <b>Legenda:</b>  Průměrné, uspokojivé   Neuspokojivé, chybějící         </p>		

Zdroj: vlastní

Zeleným symbolem jsou označeny oblasti, ve kterých daný nástroj buď výrazně převyšuje konkurenta, nebo jsou oba na stejně vysoké (dobré) úrovni. Modrým (neutrálním) symbolem jsou pak označeny oblasti, které jsou v rámci daného systému zpracovány na průměrné úrovni, a červeným ty, v nichž byly výsledky velmi neuspokojivé.

### 5.6.1 Závěrečné hodnocení GoodData

GoodData je skvělým nástrojem na důkladnou a zároveň velmi jednoduchou analýzu dat, které je možné dosáhnout za pomoci velkého množství statistických funkcí. Jazyk MAQL je unikátní především díky jednoduchosti své syntaxe, abstrakce na úrovni LDM a také díky zcela bezkonkurenčním možnostem práce s datumovou dimenzí, která je pro práci s kalendářním rokem zcela dostačující. Definování i extrémně složitých metrik je proto velmi jednoduchou záležitostí, což je asi nejsilnější doménou GD.

System ale zcela zaostává v podpoře mobilního BI, omezení přístupu pomocí RLS je bez uživatelského rozhraní a celkový vzhled vizualizací i grafického prostředí není příliš moderní. Mapy sice podporují české územní celky, ale svým vzhledem a možnostmi nastavení jsou oproti konkurenci naprosto tristní.

Naopak během testování vše fungovalo bez sebemenších problémů a kromě občasného čekání a pomalých načítání reportů byla práce se systémem velmi intuitivní a příjemná.

Z pohledu všech silných a slabých stránek je možné GD doporučit uživatelům resp. firmám, které mají velká analytická očekávání, neboť GD je dozajisté dokáže naplnit. Tito uživatelé nesmí být příliš nároční, co se nejnovějších designových trendů týče, dokáží si občas poradit i s obsluhou REST API a neodradí je relativně vysoká cena – i když hodnocení ceny bez její přesné znalosti a bez konkrétní specifikace prostředí (počet uživatelů, objem dat, ...) je velmi orientační. GoodData by si rozhodně měli pořizovat pouze ti, kteří opravdu vědí, co chtějí a co za své peníze dostanou.

## 5.6.2 Závěrečné hodnocení Power BI

Power BI vyniká ve všech oblastech, ve kterých GD ztrácí. Především se jedná o oblasti mobilního BI, celkového vzhledu prostředí i vizualizací, rychlosti, omezení přístupu k datům a moderních doplňkových funkcí.

Poměrně nepohodlná práce s metrikami (jež musejí být definovány v desktopovém návrhovém prostředí), nepřiliš jednoduchý a méně použitelný dotazovací jazyk DAX a nekompletní datumová dimenze činí z Power BI přesný opak toho, co je GD.

Zjevné je zaměření Power BI na mnohem širší skupinu uživatelů, kteří preferují moderní, skvěle vypadající nástroj se zcela intuitivním rozhraním, který zároveň nabízí nadčasové funkce (kde Q&A je ukázkou jedné z nich). Zároveň se od cílové skupiny uživatelů neočekává, že by horlivě vyžadovali pokročilé analytické vlastnosti a spokojí se s jednoduššími metrikami, které definuje správce systému.

Největším omezením je u Power BI 10GB datový limit, který nelze nijak navýšit. To může potenciální zákazníky i přes velmi příznivou cenu odradit. Microsoft však zjevně sází na široké použití funkce DirectQuery a využití platformy pro pouhou konzumaci již předpřipravených dat. Pravděpodobně se jedná o taktiku, jak zákazníkům prodat i cloudové databáze v rámci programu Microsoft Azure.

## 5.7 Doporučení

Z provedeného průzkumu vyplynulo, že navzdory tomu, že oba nástroje spadají do skupiny cloudových BI analytických platforem, se jejich možnosti použití zásadně liší. Tato skutečnost podporuje fakt, že organizace musejí být při výběru takového řešení velmi obezřetné a zpočátku provést velmi důkladnou analýzu všech požadavků, které budou na později implementovaný systém kladeny. Nesprávná volba by totiž mohla znamenat buď zbytečně vynaložený kapitál na nástroj, u kterého je využit pouze zlomek jeho potenciálu, nebo naopak poddimenzované prostředky, které nebudou schopny vysoká očekávání naplnit.

Správný výběr analytické platformy nemusí být vůbec jednoduchou záležitostí, a to především z důvodu existence velkého počtu výrobců takových řešení a jejich vzájemných rozdílů. Z tohoto důvodu dnes na trhu figurují výzkumné organizace (jimž dominuje americká společnost Gartner<sup>132</sup>), které se zkoumáním celého trhu s IS/ICT technologiemi zabývají a dokáží poskytnout kvalitní konzultační služby a provést zákazníka celým procesem volby platformy. Bohužel tyto konzultační služby nejsou vůbec levnou záležitostí. Náklady na implementaci mohou rázem přesáhnout stovky tisíc až miliony korun, ale za kvalitu poskytnutých služeb a garanci správné volby tyto vynaložené prostředky jistě stojí.

---

<sup>132</sup> About Gartner. *Gartner* [online]. Stamford: Gartner, Inc, 2016 [cit. 2016-12-09]. Dostupné z: <http://www.gartner.com/technology/about.jsp>



## 6 Implementace cloudového BI ve společnosti ABC

Dílním cílem práce byla implementace cloudového BI v konkrétním firemním prostředí. Pro tento účel poskytla spolupráci nejmenovaná společnost, která bude nadále označována jako firma „ABC“.

### 6.1 Úvodní analýza

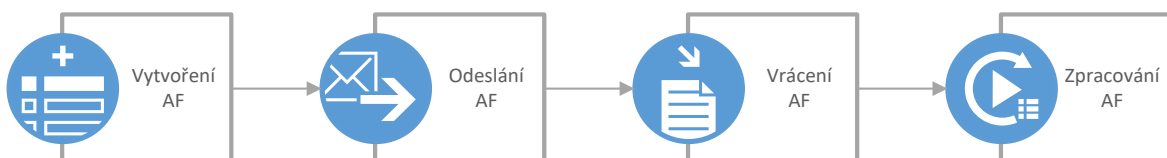
Prvním krokem při návrhu řešení byla nutná analýza původního stavu, která slouží jako výchozí bod pro celý projekt a bez níž by bylo o mnoho náročnější problematiku pochopit.

#### 6.1.1 Prostředí

Během úvodní analýzy bylo zjištěno, že firma již disponuje cloudovým řešením BI, konkrétně se jedná o kombinaci nástrojů Keboola Connection (KBC) a GoodData. Firma tyto nástroje již nějakou dobu celopodnikově aktivně používá a úkolem tohoto projektu tedy bylo řešení doplnit a rozšířit. Tato výchozí situace se ukázala jako velmi výhodná, neboť proces rozšíření současného řešení je mnohem méně finančně i časově náročný než implementace zcela nového.

#### 6.1.2 Předmět analýzy

Spolupráce s firmou ABC pro potřeby této práce se týkala pouze jedné části jejich celkově komplexního BI systému. Tou byla analýza procesu získávání nových zákazníků neboli tzv. akvizičního procesu. Celý proces je znázorněn následující ilustrací a dále bude blíže specifikován.



Obrázek 6-1: Proces získávání nových zákazníků

Zdroj: vlastní

Prvním krokem v procesu získávání nového zákazníka je vytvoření tzv. „aktivního formuláře“ (AF), který je možné připodobnit k jisté formě poptávky. Tento formulář, obsahující pouze základní informace, je v následujícím kroku poštovně odeslán zákazníkovi, který do něho musí vyplnit doplňující informace, podepsat a odeslat zpět. Třetí krok procesu sleduje navrácení AF firmě, kde v podobě kroku číslo 4 dojde k jeho zpracování a zaevidování. Dále pak již dochází k podpisu smlouvy, avšak to se zkoumaného předmětu týká pouze okrajově.

Podklad pro analýzu nastíněného procesu tvoří jedna základní zdrojová tabulka (AF) obsahující aktivní formuláře. Informace v této tabulce ale nemusí být vždy správné, neboť do podpisu smlouvy není vynaložena dostatečná snaha na zajištění 100% správných informací. Jakmile je smlouva podepsána, vytváří se o daném zákazníkovi nový záznam v druhé tabulce (CONT), obsahující údaje o podepsané smlouvě. Pro kvalitní analýzu je potřeba použít kombinaci údajů z obou zmíněných tabulek. Specifikace tabulek AF a CONT se nacházejí v příloze C.

### 6.1.3 Výchozí stav

Pro analýzu procesu získávání nových zákazníků jsou v tabulce AF nejdůležitějšími atributy časová razítka událostí, která identifikují jednotlivé kroky v procesu. Jedná se o razítka *af\_created\_date* (datum vytvoření AF), *af\_send\_date* (datum odeslání AF zákazníkovi), *af\_returned\_date* (datum navrácení AF od zákazníka) a *af\_processed\_date* (datum zpracování AF).

Firma má analýzu nastíněného procesu v rámci BI již zpracovanou, ale vznikl požadavek na jeho změnu. V původní verzi řešení jsou data z obou tabulek (AF i CONT) jistým způsobem zpracována: výsledkem je jedna tabulka sjednocující údaje z obou tabulek. Tato výsledná tabulka je dále nahrána do GD, kde jsou její data analyzována a zobrazena.

Důležitým faktem je, že každý řádek výsledné tabulky reprezentuje jeden AF a časová razítka zůstávají zachována v podobě atributů. Toto řešení je specifické tím, že každému časovému razítku (atributu) je v GD přiřazena vlastní časová dimenze.

Charakteristickým rysem původního řešení je díky vícero časovým dimenzím velmi jednoduchá filtrace a sledování postupujících kroků procesu. Uživatelé mohou například pohodlně zobrazit AF, které byly vytvořeny v minulém měsíci, ale k aktuálnímu dni doposud nebyly zpracovány. Naopak bez nutnosti použití pomocných atributů je nemožné sledování času mezi jednotlivými kroky procesu (např. jak dlouho průměrně trvá, než dojde ke zpracování AF od jeho návratu apod.).

#### **6.1.4 Požadavky a cíle**

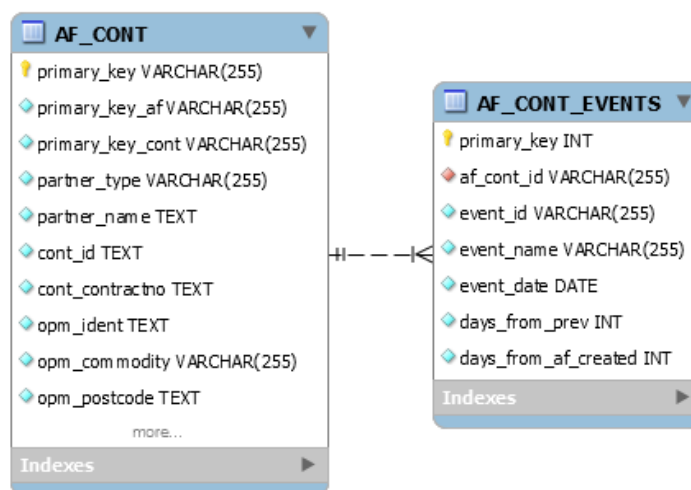
Z důvodu nepohodlné práce s několika časovými dimenzemi a nemožností sledování délky akvizičního procesu si firma vyžádala návrh nového řešení, které by tyto neduhy řešilo. Cílem projektu bylo vytvoření nového datového modelu, který bude umožňovat pohodlné sledování délky jednotlivých kroků v procesu získávání nových zákazníků a zároveň k tomu bude sloužit pouze jedna časová dimenze. Jako další cíle firma stanovila možnost zobrazení počtu událostí na mapových podkladech České republiky a celkovou optimalizaci procesu zpracování dat, který je v aktuální podobě kvůli častým úpravám velmi nepřehledný a složitý. Jedinou podmínkou zpracování bylo použití již fungujícího ekosystému v podobě cloudového ETL nástroje Keboola Connection a BI platformy GoodData.

## **6.2 Návrh řešení**

Z analýzy původního řešení a nových požadavků vyplynulo, že pro dosažení pouze jedné časové dimenze bude nutné změnit datový model a rozdělit výstupní data do dvou tabulek. První tabulka bude obsahovat pouze informace vztahující se k jednotlivým AF a v podřízené tabulce budou jen časová razítka jednotlivých událostí ve formě samostatných řádků. Toto řešení umožní při analýze a vizualizaci jednoduše filtrovat všechny události v určitém časovém období, což poslouží jako cenný artikl při souhrnném zobrazení počtu jednotlivých událostí v jednotlivých dnech měsíce.

Během pozdějšího vývoje a testování se ukázalo, že pro sledování délky akvizičního procesu je stejně nutné pomocné atributy použít. Proto byly do nového modelu do podřízené tabulky *AF\_CONT\_EVENTS* přidány atributy *days\_from\_prev* a *days\_from\_af\_created*. První

z uvedených obsahuje vypočtený počet dní od předcházející události a druhý atribut pak počet dní od vytvoření AF. Zjednodušené schéma výsledného datového modelu ilustruje obrázek 6-2.

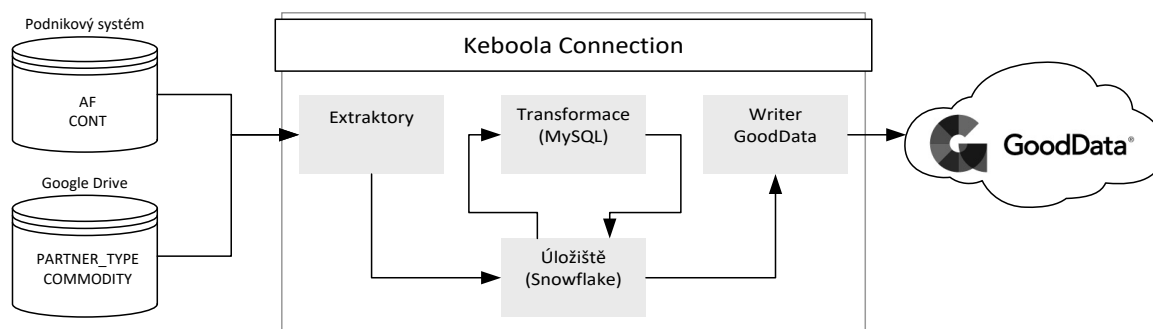


Obrázek 6-2: Nový datový model

Zdroj: vlastní

Tabulka *AF\_CONT* bude během procesu transformace dat vznikat obdobným způsobem jako v původním řešení. Bude se jednat o podmíněčné slučování dat z tabulek *AF* a *CONT* s případným připojením dalších doplňujících tabulek s popisnými atributy.

Zobrazení dat na mapě České republiky bude probíhat na základě zákaznickova poštovního směrovacího čísla (PSC), dle kterého se budou data seskupovat do správních oblastí odpovídajících NUTS 4 (nyní také známé jako LAU 1).



Obrázek 6-3: Schéma funkce celého řešení

Zdroj: vlastní

Obrázek 6-3 znázorňuje funkci celkového řešení od získání dat až po jejich zápis do GD. Tento proces bude v následující části detailněji osvětlen.

## 6.3 Implementace

Po úvodní analýze a vytvoření návrhu řešení bylo možné přistoupit k samotné implementaci a testování. V následujících podkapitolách jsou uvedeny postupné kroky, které vedly k výslednému funkčnímu prototypu.

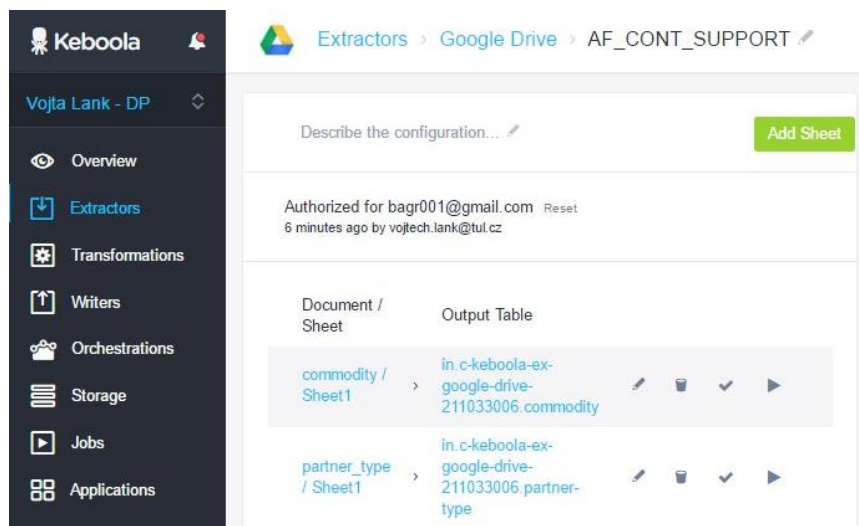
### 6.3.1 Zdroje dat

Jak již bylo zmíněno, základem jsou tabulky *AF* a *CONT*. Tyto tabulky jsou produktem podnikového informačního systému a jsou do prostředí KBC přenášeny přímo ze strany firmy ABC prostřednictvím vnějšího rozhraní.

Firma si dále vyžádala použití vlastních číselníků pro jednodušší orientaci v datech během vizualizace. Tyto číselníky proto budou sloužit pro nahrazení některých atributů během procesu transformace dat a jako jejich zdroj byly definovány tabulky Google Sheets přístupné přes cloudové úložiště Google Drive. Konkrétně se jedná o tabulky *PARTNER\_TYPE* a *OPM\_COMMODITY*.

### 6.3.2 Extrakce dat z Google Drive

K extrakci dat z cloudové úložiště Google Drive slouží v prostředí KBC stejnojmenný extraktor „Google Drive“. Konfigurace je velmi jednoduchá. Stačí zadat přihlašovací údaje k účtu Google, který disponuje právy na čtení daného dokumentu, v dialogovém okně vybrat požadovaný dokument Google Sheet a zvolit list dokumentu, na kterém jsou data umístěna. KBC již požadované tabulky automaticky stáhne do vnitřního úložiště. Konfiguraci tohoto extraktoru je možné vidět na následujícím obrázku.



Obrázek 6-4: KBC – extraktor Google Drive

Zdroj: vlastní

### 6.3.3 Transformace

Všechna data jsou po extrakci uložena v úložišti KBC, které pro tyto účely využívá cloudové databázové služby Snowflake. V této fázi je možné načít uložená data dále pracovat pomocí datových transformací. KBC sice svým zákazníkům nabízí poměrně širokou škálu různorodých předpřipravených aplikací, jakými jsou například analýza nákupního košíku nebo vytvoření histogramu četností, ale pro potřeby tohoto projektu bylo nezbytné vytvoření zcela individuální transformace.

KBC v současné době nabízí několik různých prostředí (backend), ve kterých je možné transformace provádět. Uživatel může aktuálně volit mezi databázovými servery MySQL, Amazon Redshift a Snowflake, nebo mezi programovacími (skriptovacími) jazyky R a Python. Jelikož nebyly v rámci projektu očekávány žádné velké objemy dat ani pokročilé matematické operace, bylo zvoleno nejjednodušší prostředí, a to relační databáze MySQL.

Pro testovací účely byla v KBC nejprve vytvořena tzv. sandbox<sup>133</sup> databáze, ve které vývoj transformace probíhal. KBC na vyžádání automaticky vytvoří databázi s požadovanými

<sup>133</sup> Sandbox, neboli v doslovném překladu „pískoviště“, je v informatické terminologii pojem pro označení libovolného testovacího prostředí, ve kterém je možné provádět vývoj, či někdy i nebezpečné pokusy, bez nutnosti obav z ohrožení použitých zdrojů.

tabulkami, které budou v rámci transformace využívány a sdělí uživateli všechny údaje nezbytné pro přístup. Připojení ke vzdálené databázi je poté možné provést z libovolného terminálu MySQL.

Nyní bude vysvětlen celý proces fungování výsledné transformace, kterou schematicky znázorňuje obrázek 6-6 a jejíž SQL kód se nachází v příloze D i příloze DVD 4. Prvním úkolem transformace je vytvoření hlavní výstupní tabulky *AF\_CONT*, kterou je v další fázi potřeba naplnit daty.

Jak již bylo zmíněno, tabulka *AF* nemusí vždy obsahovat správné údaje. Ty je ale možné v případě, kdy již došlo k podepsání smlouvy, převzít z tabulky *CONT*. Proto je druhým krokem transformace proces, který spojí tabulky *AF* a *CONT* na základě rovnosti atributů *contract\_no* a *opm\_ident* a v případě, že existuje neprázdná hodnota některého ze zúčastněných atributů v tabulce *CONT*, zvolí tuto hodnotu a výsledný řádek zapíše do cílové tabulky.

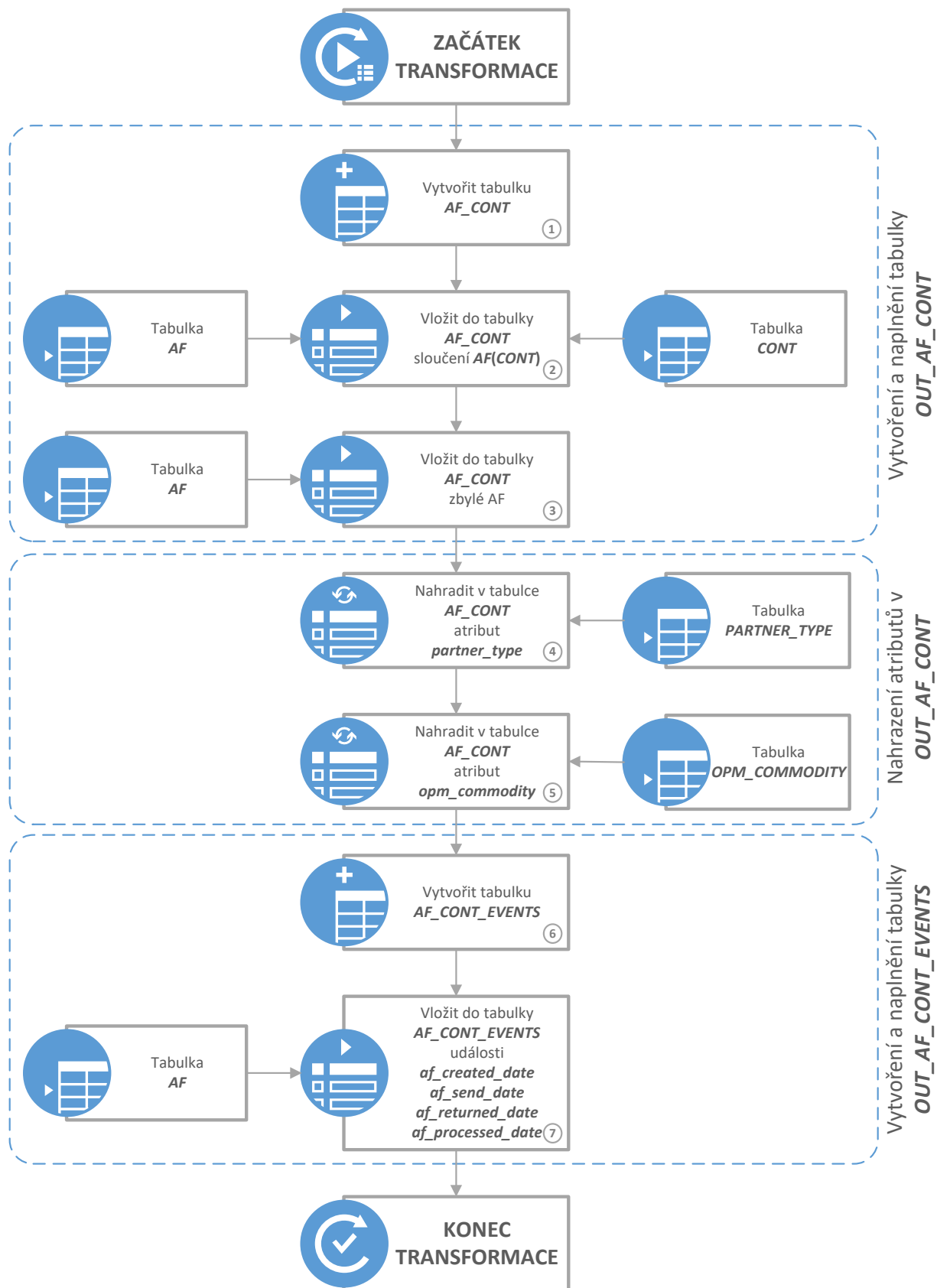
Druhý krok transformace do tabulky *AF\_CONT* zapíše pouze část ze zdrojových řádků. Proto je potřeba z tabulky *AF* vybrat zbývající řádky (i přestože není správnost údajů zaručena) a zapsat je opět do požadované cílové tabulky. Tento úkol splňuje třetí krok transformace.

Tímto krokem je ukončen blok transformace, který má za úkol generovat data pro tabulku *AF\_CONT*. Nyní je potřeba ještě v této tabulce nahradit hodnoty některých atributů kvůli lepší orientaci při reportingu. Tento úkon je ve schématu reprezentován kroky 4 a 5, které v tabulce *AF\_CONT* nahrazují hodnoty atributů *partner\_type* a *opm\_commodity* na základě hodnot v tabulkách *PARTNER\_TYPE* a *COMMODITY*. Následující obrázek zachycuje SQL skript transformace nahrazující hodnoty atributů *partner\_type*.

```
UPDATE `AF_CONT`  
JOIN `PARTNER_TYPE` ON `PARTNER_TYPE`.`hodnota` = `AF_CONT`.`partner_type`  
SET `AF_CONT`.`partner_type` = `PARTNER_TYPE`.`popis`;
```

Obrázek 6-5: SQL skript transformace pro nahrazení atributu *partner\_type*

Zdroj: vlastní



Obrázek 6-6: Schématické znázornění datové transformace

Zdroj: vlastní



Posledním blokem transformace je vytvoření a naplnění podřízené tabulky *AF\_CONT\_EVENTS*, což je reprezentováno kroky číslo 6 a 7. První ze zmíněných má za úkol vytvoření tabulky a druhý pak naplnění daty. Je důležité podotknout, že krok číslo 7, ve vztahu k celkovému SQL skriptu transformace, v sobě kvůli jednoduššímu schématickému znázornění integruje několik procesů dohromady. Generování dat pro tuto výstupní tabulku je založeno na výběru požadovaných dat z tabulky *AF* a spojení s tabulkou *AF\_CONT* pro potřeby generování unikátního primárního klíče. Spojení obou zmíněných tabulek probíhá na základě rovnosti atributů *AF\_CONT.primary\_key\_af* a *AF.primary\_key*, a unikátní klíč vzniká spojením atributů *AF\_CONT.primary\_key*, “\_” a daného datumového atributu z tabulky *AF* (např. *af\_processed\_date*).

Jelikož jsou ve zdrojové tabulce datumové hodnoty uloženy v českém formátu (tedy D.M.RRRR), bylo je potřeba pomocí vestavěné funkce *STR\_TO\_DATE* převést do standardizovaného datumového formátu MySQL ve formátu RRRR-MM-DD<sup>134</sup>. Do výsledné tabulky jsou dále zapsány pravé názvy událostí a vypočtené rozdíly časů mezi danou událostí a dnem vytvoření AF. Zapsány jsou pouze záznamy s neprázdnými nebo ne-NULL-ovými hodnotami. Pro ilustraci, obrázek 6-7 zobrazuje SQL skript této části transformace, konkrétně pak pro událost *af\_processed\_date*, neboli datum zpracování AF.

```

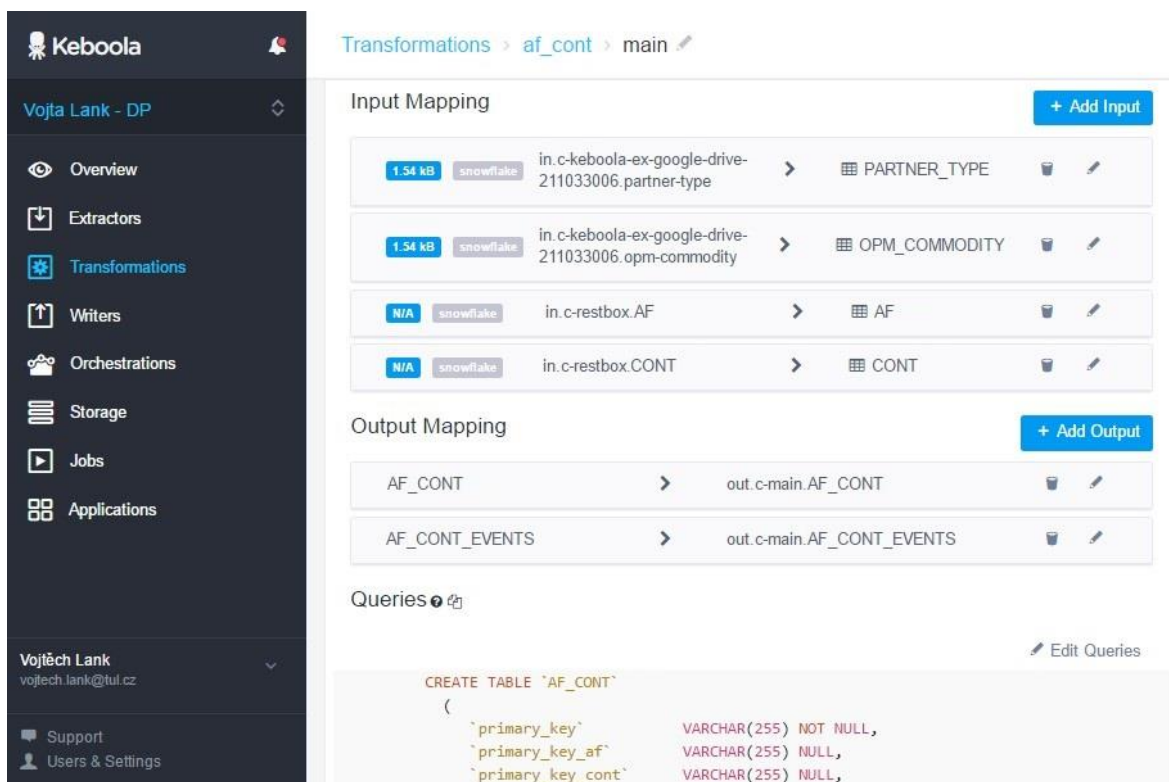
INSERT INTO `AF_CONT_EVENTS`
SELECT CONCAT(`AF_CONT`.`primary_key`, '_', 'af_processed_date'),
       `AF_CONT`.`primary_key`,
       'af_processed_date',
       'Datum zpracování AF',
       STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'),
       DATEDIFF(
           STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'),
           STR_TO_DATE(`AF`.`af_returned_date`, '%d.%m.%Y')
       ),
       DATEDIFF(
           STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'),
           STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y')
       )
FROM `AF`
JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF`.`af_processed_date` != ''
AND `AF`.`af_processed_date` IS NOT NULL ;

```

Obrázek 6-7: SQL skript transformace pro tabulku *AF\_CONT\_EVENTS*

Zdroj: vlastní

<sup>134</sup> The DATE, DATETIME, and TIMESTAMP Types. *MySQL* [online]. Redwood: Oracle Corporation, 2016 [cit. 2016-10-23]. Dostupné z: <http://dev.mysql.com/doc/refman/5.7/en/datetime.html>



Obrázek 6-8: KBC – nastavení transformace

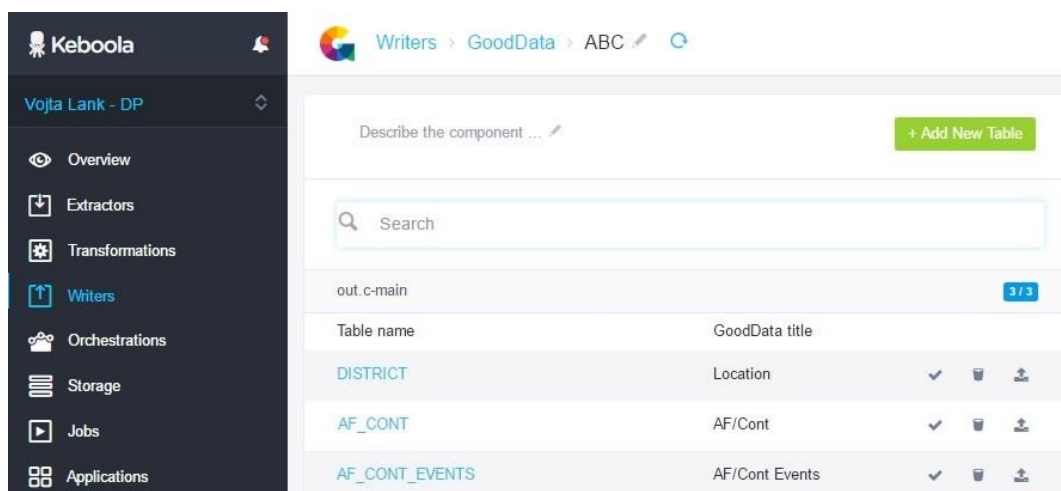
Zdroj: vlastní

V prostředí KBC, jež zachycuje obrázek 6-8, je závěrečné nastavení transformace snadnou záležitostí. Nejprve je potřeba zvolit požadované vstupní tabulky (v tomto případě se jedná o *PARTNER\_TYPE* a *OPM\_COMMODITY* z Google Drive a *AF* a *CONT* z podnikového systému) a také nastavit mapování výstupních tabulek (*AF\_CONT* a *AF\_CONT\_EVENTS*). Dále stačí pouze zadat vytvořený zdrojový kód transformace a vše je hotovo. Po spuštění a úspěšném vykonání transformace jsou data výstupních tabulek uložena v úložišti KBC, připravena pro další zpracování.

V úvodní analýze bylo zmíněno, že jedním z požadavků na výstup je možnost vizualizace dat na mapových podkladech České republiky na základě zákazníkova PSČ. Pro tyto účely byla za použití kombinace volně dostupných geografických a poštovních dat vytvořena tabulka *DISTRICT*. Jelikož se jedná o data s poměrně statickým charakterem, která nemusí být při každé aktualizaci dat nijak transformována, byla pouze přímo importována do úložiště KBC ze souboru CSV.

### 6.3.4 Odeslání dat do GoodData

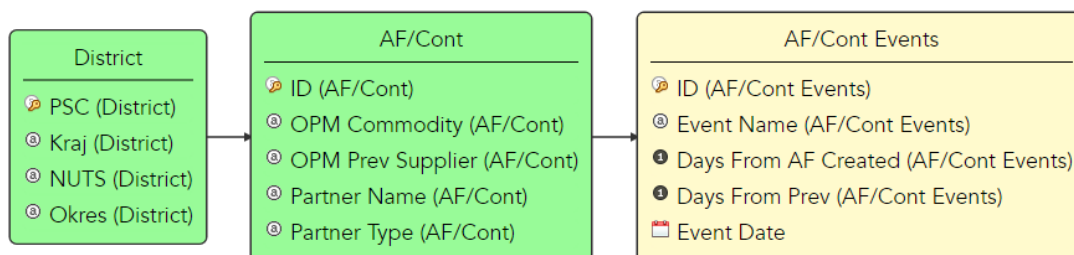
Po zpracování vstupních dat do formy vhodné pro analýzu bylo možné přejít k dalšímu kroku v procesu implementace, čímž je odeslání dat do analytické platformy GD. Pro výstup dat a jejich odesílání do dalších služeb slouží v KBC tzv. writery (zapisovače), které v terminologii ETL nástrojů zastupují fázi „Load“.



Obrázek 6-9: KBC – writer GoodData

Zdroj: vlastní

V tomto případě byl použit stejnojmenný writer (GoodData). Jeho konfigurace spočívá v zadání přihlašovacích údajů a identifikátoru stávajícího projektu a dále nastavení přenášených tabulek a jejich atributů. V návaznosti na předešlé kroky implementace projektu se jedná o tabulky shodné s výstupem transformace (*AF\_CONT* a *AF\_CONT\_EVENTS*) a také tabulku s geografickými údaji (*DISTRICT*). Souhrnnou konfiguraci writeru ukazuje obrázek 6-9.



Obrázek 6-10: Výsledný logický datový model

Zdroj: vlastní

Data přenášená do GD je nejprve potřeba uspořádat do LDM (více viz 5.4.2 – Integrace dat a datový model). Obrázek 6-10 dokumentuje výsledný LDM, který byl vygenerován aplikací GoodData LDM Visualizer.

Při konfiguraci LDM musí mít každý z atributů specifikovaný popisný název (GoodData title), který bude zobrazen uživateli v prostředí GD a také typ (Type) atributu. Typ CONNECTION\_POINT určuje primární klíč tabulky a společně s typem REFERENCE, který značí cizí klíč, zajišťují tvorbu vazeb 1:N.

Při konfiguraci atributů tabulky *AF\_CONT* (viz obrázek 6-11) byly kromě dvou všechny nastaveny na typ ATTRIBUTE, což v oblasti multidimenzionálního modelování odpovídá dimenzionálnímu atributu. Výjimku tvořil pouze atribut *primary\_key*, který slouží jako primární klíč, a proto byl nastaven na typ CONNECTION\_POINT, a také atribut *opm\_postcode*, který odkazuje na atribut *psc* v tabulce *DISTRICT*, a proto byl nastaven na typ REFERENCE.



Column	GoodData title	Type	Reference
primary_key	ID (AF/Cont)	CONNECTION_POINT	
partner_type	Partner Type (AF/Cont)	ATTRIBUTE	
opm_postcode		REFERENCE	out.c-main.DISTRICT

Obrázek 6-11: KBC – writer GoodData – tabulka *AF\_CONT*

Zdroj: vlastní

Tabulka *AF\_CONT\_EVENTS*, jejíž nastavení zachycuje obrázek 6-12, byla konfigurována obdobným způsobem, avšak kvůli přítomnosti faktových atributů muselo být nastavení náležitě upraveno. V této situaci odkazuje atribut *af\_cont\_id* do nadřazené tabulky *AF\_CONT*. Dále byly atributy *days\_from\_prev* a *days\_from\_af\_created* označeny jako fakty (typ FACT). Poslední důležité změny se týkaly atributu *event\_date*. Tento atribut zde totiž slouží jako reference do datumové dimenze, proto byl označen typem DATE. Tato datumová dimenze s názvem Event Date byla pro účely tohoto projektu systémem GoodData automaticky vygenerována ve své běžně používané podobě.

Column	GoodData title	Type	Reference
primary_key	ID (AF/Cont Events)	CONNECTION_POINT	
af_cont_id		REFERENCE	out.c-main.AF_CONT
event_id		IGNORE	
event_name	Event Name (AF/Cont Events)	ATTRIBUTE	
event_date		DATE	yyyy-MM-dd Dimension: Event Date
days_from_prev	Days From Prev (AF/Cont Events)	FACT	
days_from_af_created	Days From AF Created (AF/Cont Events)	FACT	

Obrázek 6-12: KBC – writer GoodData – tabulka AF\_CONT\_EVENTS

Zdroj: vlastní

Po dokončení všech nastavení je možné tlačítkem „Upload project“ odeslat vytvořený datový model společně s daty do GD, kde je možné začít s daty ihned dále pracovat. Vizuální reprezentace datového modelu přeneseného do prostředí GD s výčtem nejdůležitějších atributů se nachází v příloze E.

### 6.3.5 Návrh metrik

Čistý projekt v GD obsahuje pouze „holé“ fakty, se kterými se dá pracovat jen ve velmi omezené míře. Pro rozšíření pracovních možností bylo nejprve potřeba vytvořit odvozené metriky, které dokáží nejen celou práci výrazně zjednodušit, ale budou také sloužit jako podklad pro odvozování dalších složitějších metrik. Více základních informací o tvorbě metrik a jazyku MAQL se nachází v části 5.4.4 – Metriky.

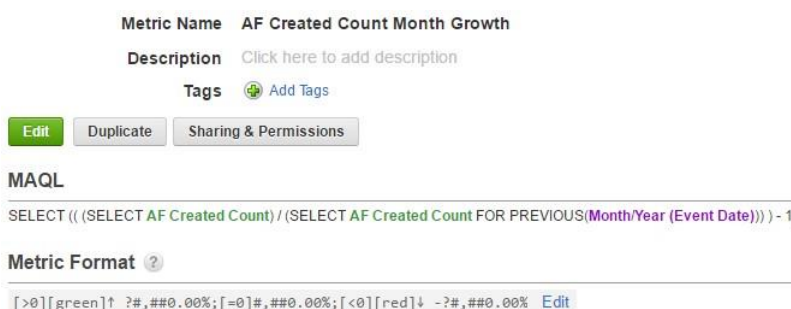
Jako první byly vytvořeny metriky určující počet jednotlivých událostí. Následující obrázek zachycuje proces vytváření metriky s názvem „AF Created Count“ obecně určující počet událostí vytvoření AF. Tato metrika je založena na počtu řádků dle klíčového atributu tabulky AF\_CONT\_EVENTS a konkrétním názvu události. Stejným způsobem byly následně vytvořeny stejné metriky pro všechny dříve diskutované události.



Obrázek 6-13: GD – definice metriky „AF Created Count“

Zdroj: vlastní

Tyto metriky budou dále sloužit pro vytváření různých typů přehledů o počtu jednotlivých událostí, avšak pro složitější reporty je potřeba vytvořit další odvozené metriky. Dalším ukazatelem, který má být v projektu sledován, je procentuální rozdíl oproti počtu jednotlivých událostí v předešlém měsíci. Pro tyto potřeby byla vytvořena sada odpovídajících metrik.



Obrázek 6-14: GD – definice metriky „AF Created Count Month Growth“

Zdroj: vlastní

Obrázek 6-14 ukazuje návrhové okno metriky s názvem „AF Created Count Month Growth“. Tato metrika vychází z již dříve vytvořené metriky „AF Created Count“ a jejím výstupem je podíl počtu vytvořených AF ve zvoleném měsíci oproti počtu v měsíci předcházejícím. Jelikož je produktem metriky číslo reprezentující procentuální hodnotu, je potřeba při jeho prezentaci dbát na správné zobrazení. Z toho důvodu bylo nastaveno formátování (viz obrázek 6-15) tak, aby byla hodnota vždy zobrazena jako procento (včetně patřičného znaku) a také aby se při kladné hodnotě text zbarvil do zelena a před číslem se zobrazila šipka vzhůru a znak „+“. V případě záporné hodnoty bylo formátování nastaveno tak, aby byl text červený, šipka směřovala dolů a zobrazil se znak „-“.

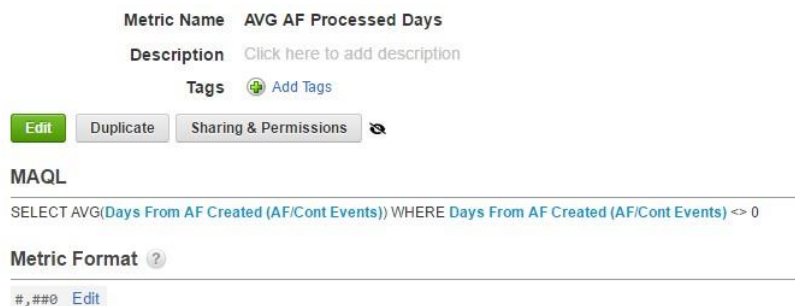


```
[<0][red]↓ -?#,##0.00%; -1 ↓ -100.00%
[=0]#,##0.00%; 0 0.00%
[>0][green]↑ +?#,##0.00% 1 ↑ +100.00%
```

Obrázek 6-15: GD – formátování metriky „AF Created Count Month Growth“

Zdroj: vlastní

Jedním z požadavků na výstup byla také možnost sledování průměrné délky jednotlivých kroků i celkové délky akvizičního procesu. Proto bylo nutné vytvořit další metriky vyhovující těmto potřebám. Jako příklad zde poslouží metrika „AVG AF Processed Days“, jejíž návrh dokumentuje obrázek 6-16. Tato metrika slouží pro výpočet průměrné délky celého procesu získávání nového zákazníka (tedy od vytvoření AF až po jeho zpracování) a její základ tvoří průměrování dříve vypočtené hodnoty *days\_from\_af\_created*, jež je reprezentována faktem „Days From AF Created (AF/Cont Events)“. Výpočet je dále podmíněn průměrováním nenulových hodnot, které by jinak výsledek zkreslovaly. Ostatní průměrné metriky jsou obdobné, pouze s rozdílem použití faktu „Days From Prev“.

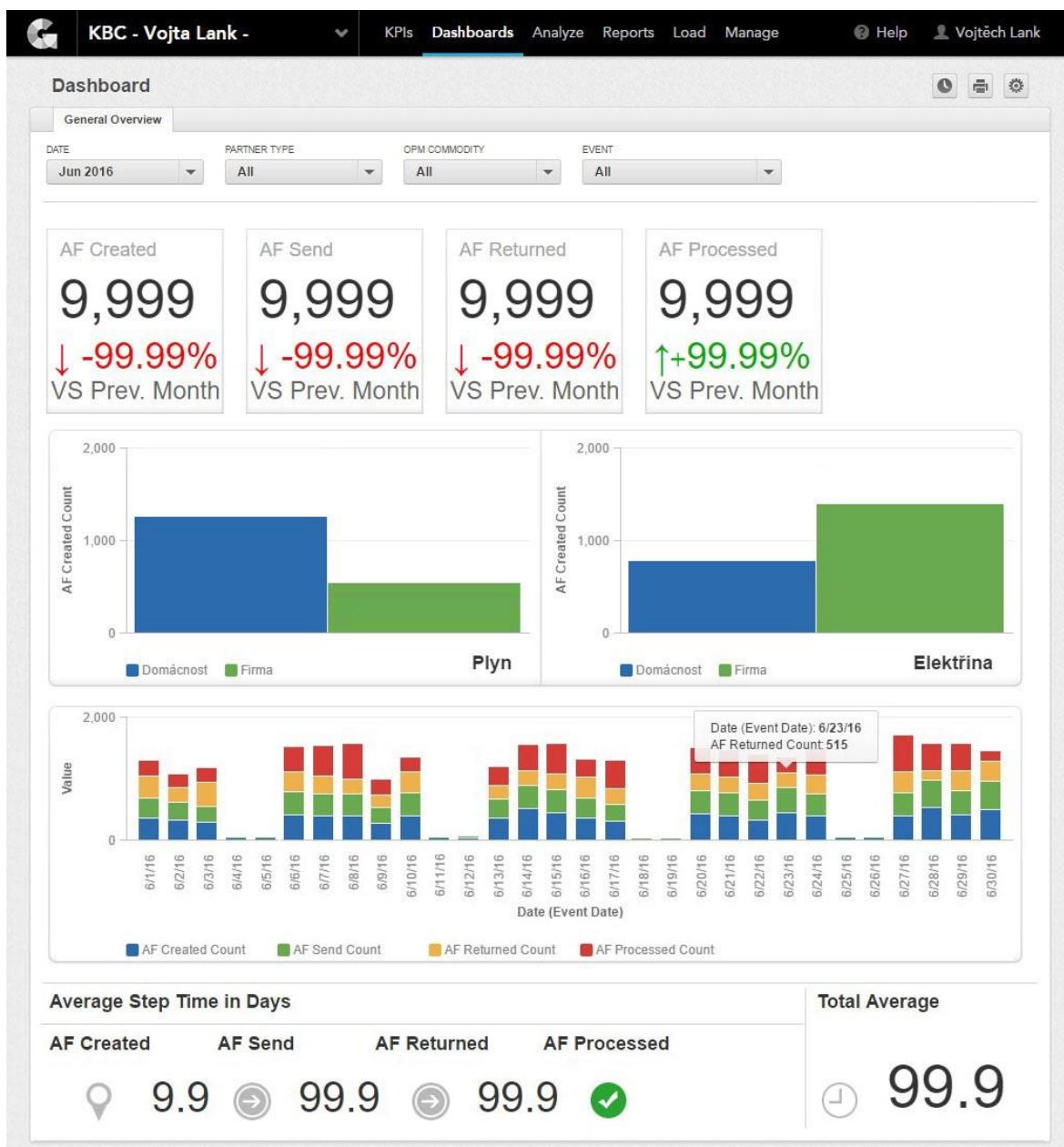


Obrázek 6-16: GD – definice metriky „AVG AF Processed Days“

Zdroj: vlastní

### 6.3.6 Dashboard

Hlavním výsledkem celého projektu je souhrnný dashboard (obrázek 6-17) sloužící k přehledné vizualizaci různých informací na jednom místě. Je zde zobrazen výběr z možných přehledů či statistik tak, aby byla pro uživatele orientace v celém dashboardu co možná nejjednodušší a aby si uživatel mohl při jeho otevření udělat rychlý obrázek o aktuální situaci.



Obrázek 6-17: GD – hlavní dashboard

Zdroj: vlastní

V souvislosti s výše uvedeným dashboardem je potřeba zmínit, že obsažené informace musely být zkresleny z důvodu potenciálního úniku informací – proto má pouze demonstrativní charakter.

V záhlaví vytvořeného dashboardu se nachází oblast filtrů, pomocí kterých je možné rychle změnit zobrazovaná data dle vybraných hodnot nabízených atributů. Ve výchozím stavu je



pro datumovou dimenzi zvolen aktuální kalendářní měsíc a ostatní filtry jsou nastaveny na hodnotu „All“, což znamená, že všechny hodnoty jsou vybrány.

Přímo pod záhlavím figuruje oblast s přehledem o počtu jednotlivých událostí, které se odehrály ve zvoleném měsíci, společně s porovnáním oproti měsíci předešlému. Uživatel má již při vstupu na dashboard ihned přehled o aktuálním stavu.

Další sekce dashboardu je věnována porovnání počtu vytvořených AF pro různé typy zákazníků a různé nabízené komodity. Jsou zde uvedeny dva grafy, každý pro jinou komoditu, přičemž každý z nich si bere za úkol porovnat rozdíl mezi jiným typem zákazníka.

Ve třetí části se nachází skládaný sloupcový graf zobrazující počty jednotlivých událostí pro každý den zvoleného období. Pokud uživatel zvolí u filtru EVENT pouze jednu hodnotu, získá sloupcový graf pouze pro konkrétní zvolenou událost.

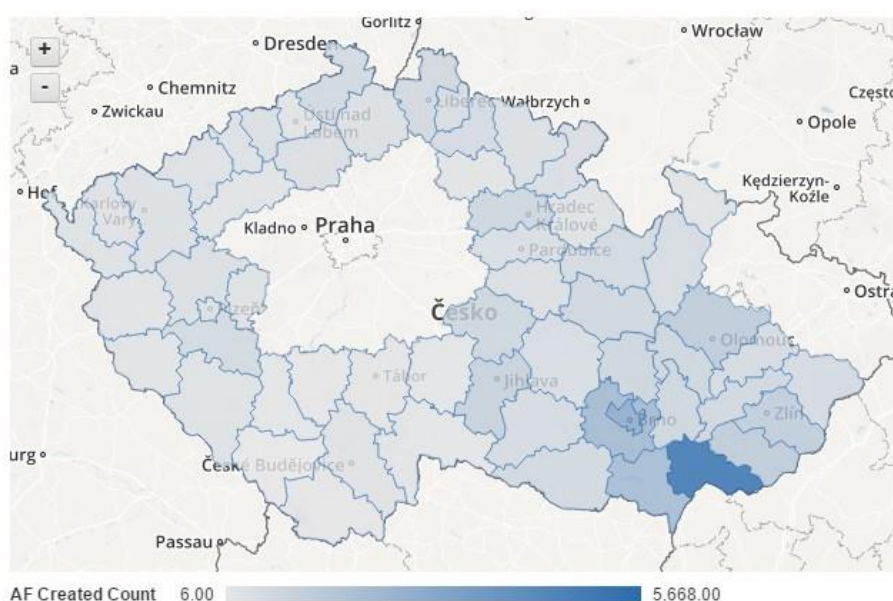
Poslední sekce se zabývá průměrným časem jednotlivých kroků akvizičního procesu. Číslo mezi jednotlivými schématickými ikonami označuje průměrnou délku mezi danými událostmi a poslední číslo vpravo s označením „Total Average“ referuje o průměrném čase trvání celkového procesu.

Vytvořený dashboard je z velké části tvořen předem vytvořenými samostatnými reporty, které byly do prostředí dashboardu pouze zakomponovány. Více o tvorbě reportů bylo zmíněno v části 5.4.5.3 – Reporty a dashboardy.

### **6.3.7 Zobrazení na mapě ČR**

Posledním z cílů tohoto projektu bylo zobrazení dat na mapě České republiky. Pro tyto účely byla vytvořena tabulka *DISTRICT*, která umožňuje propojení různých územních celků na základě zákaznickova poštovního směrovacího čísla. GoodData umí zobrazovat data v různých úrovních územních celků, ze kterých byla pro potřeby tohoto projektu zvolena jednotka NUTS 4 (LAU 1).

Dle teoretického rozboru mělo vše fungovat, avšak při praktické aplikaci se data zobrazovala zcela nelogicky chybně. Po důkladném prozkoumání celé situace se ukázalo, že poštovní směrovací čísla, jejichž správcem je Česká pošta, nenáleží výlučně pouze jedné správní jednotce LAU 1, ba či dokonce ani samotnému kraji. Jinak řečeno, jedno PSČ může patřit vícero obcím, z nichž každá může ležet v jiném kraji. Pro příklad je možné uvést PSČ 250 63, jež patří jak obci Čakovičky (okres Mělník), tak i obci Veleň (okres Praha-východ). Tato skutečnost celou situaci značně komplikuje a kvůli tomuto nečekanému zjištění je výsledná vizualizace zcela bezcenná. Chybné zobrazení mapy zachycuje obrázek 6-18.



Obrázek 6-18: GD – chybné zobrazení mapy na základě PSČ

Zdroj: vlastní

Řešení tohoto problému se nabízí ve formě párování příslušného kódu NUTS přímo s názvem obce. To ale naráží na možné chyby v datech ve formě překlepů, což opět párování znesnadňuje. Dalším možným řešením je použití KBC extraktoru Geocoding Augmentation. Ten umožňuje za pomoci různých mapových služeb (Google Maps, Bing Maps, OpenStreetMap, aj.) získání unifikovaných adres či přesných GPS souřadnic na základě zadaných různě formátovaných adres či názvů známých míst. Značnou výhodou této služby je, že podporované mapové služby mají většinou velmi vospělý systém pro rozpoznávání zadaných adres v chybném nebo neúplném formátu, což činí tento extraktor skvělým nástrojem pro obecné čištění geografických dat i vhodným řešením vzniklé situace v rámci tohoto projektu.

Bohužel použití extraktoru Geocoding Augmentation by v tomto stádiu projektu přineslo poměrně velké změny v transformační proceduře. Z důvodu další časové náročnosti na přepracování projektu byl tedy tento návrh řešení pouze doporučen jako budoucí rozšíření.

## **6.4 Zhodnocení výsledků a přínosu projektu**

Poslední část této kapitoly je věnována zhodnocení dosažených výsledků a také přínosu pro firmu.

### **6.4.1 Dosažené cíle**

V projektu se podařilo splnit téměř všechny stanovené cíle. Podařilo se vytvořit nový datový model, který umožňuje práci pouze s jednou datumovou dimenzí a zároveň je s jeho podporou možné efektivně sledovat jednotlivé kroky akvizičního procesu (včetně jejich délek).

Díky jasným cílům bylo možné celý projekt navrhnout čistě, což mělo pozitivní vliv na celkovou přehlednost (nejen na přehlednost zdrojového kódu transformace). Jediným cílem, který se nepodařilo splnit, bylo zobrazení dat na mapových podkladech České republiky. Tento cíl ztroskotal až při finálním testování, kdy se ukázalo, že pro úroveň krajů (NUTS 3) či okresů (NUTS 4) nelze data seskupovat na základě poštovních směrovacích čísel, což způsobovalo chybné zobrazení dat. Byla sice navržena některá možná řešení, avšak z časových důvodů byla pouze doporučena jako budoucí rozšíření.

Fungující prototyp byl následně po dokončení představen firmě, která měla dále nějaký čas na otestování celého řešení a vznesení námitek. Po otestování zástupci firmy uvedli, že navržené řešení zadané cíle splňuje, i když v některých ohledech je původní řešení lepší. Jde především o to, že s jednou datumovou dimenzí nelze zobrazit aktuální stav AF, které byly vytvořeny v určitém období.

Možným řešením vzniklého problému s jednou datumovou dimenzí by mohlo být ponechání jak starého, tak nového modelu. Avšak existence rozdílných datových modelů se stejnými daty by mohla způsobovat zmatky a nejasnosti. Vhodným řešením by tak mohla být určitá kombinace obou datových modelů současně, což by obnášelo pouze přidání nově vytvořené tabulky *AF\_CONT\_EVENTS* do stávajícího datového modelu. Vytvoření jiného návrhu řešení ohledně spojení obou modelů ale bude pravděpodobně otázkou další komunikace s firmou.

### **6.4.2 Přínos pro firmu**

Zjevným přínosem nově navrženého modelu je možnost sledování délek jednotlivých kroků akvizičního procesu. Díky vysoké podrobnosti dat je možné velmi detailně sledovat celkový průběh uzavírání smluv a v případě negativních odchylek na ně ihned reagovat. Nově navržený model umožní zjistit příliš dlouho trvající dílčí kroky procesu, které je možné dále a detailněji zkoumat a odhalit tak přesnou příčinu problému.

Firma může například vyhodnotit, že některý zaměstnanec zpracovává přijaté AF příliš pomalu, nebo že problém může být v doručovací společnosti. Všechny zjištěné příčiny pak mohou vést k personálním nebo jiným změnám a ve výsledku k celkovému zkvalitnění poskytovaných služeb. To vše může v budoucnu znamenat snížení nákladů či vyšší spokojenost zákazníků, což se opět může dále projevit všeobecně lepším veřejným míněním o společnosti a přeneseně i růstem tržního podílu a tržeb.

### **6.4.3 Ekonomické zhodnocení**

Ekonomické vyhodnocení zpracovaného projektu je poměrně nelehkou záležitostí. Firma již vlastnila licence k nástrojům GD i KBC a autor na projektu pracoval v rámci studia bez finančního ohodnocení. Jelikož byla spolupráce s firmou založena na pouhých konzultacích, celkové náklady na samotný projekt byly pouze v rámci práce interních pracovníků, kteří se občasnými otázkami ohledně projektu zabývali. Celkové náklady na tento projekt tak lze označit za zcela zanedbatelné.

V případě, že by byl projekt zpracován zaměstnancem firmy ABC či externí organizací, náklady by byly zcela jiné. Dle statistických údajů ze serveru *Platy.cz* je průměrný hrubý měsíční plat IT pracovníků se zaměřením na analýzu či databázové technologie přibližně 40 000 Kč<sup>135</sup>, což tvoří celkové náklady na zaměstnance ve výši 53 600 Kč. Pokud by byl řešitel projektu zaměstnancem firmy ABC, při uvažovaném platu a odhadované časové náročnosti cca 100 hodin by náklady na zpracování projektu tvořily právě 33 500 Kč.

Pokud by se jednalo o formu externího zpracování další organizací, u níž je obecně známo, že se hodinové sazby prací pohybují běžně od 1 000 Kč výše, cena by se v takovém případě vyšplhala minimálně na 100 000 Kč. Je však nezbytné přihlídnout k tomu, že zpracovatel nebyl s použitými technologiemi před zpracováním obeznámen, tudíž se čas strávený studiem negativně odrazil na celkové výši časové náročnosti. Pokud by se jednalo o zkušeného pracovníka, stejný projekt by pravděpodobně mohl být dokončen za třetinový čas, kde by výsledné náklady mohly být srovnatelné s předcházející variantou.

---

<sup>135</sup> Platy v kategorii: Informační technologie. *Platy.cz* [online]. Praha: Profesia CZ, spol. s r.o., 2016 [cit. 2016-12-17]. Dostupné z: <http://www.platy.cz/platy/informacni-technologie>

## Závěr

Přesun nástrojů a aplikací kategorie Business Intelligence do cloudového prostředí umožnil zcela nové možnosti analýzy dat a zároveň změnil způsob, jakým jsou tyto aplikace používány. S ohledem na výhody, které z cloudové distribuce vyplývají, je možné nástroje nasadit mnohem rychleji a mnohdy i s nižšími náklady, než tomu bylo dříve. Díky velmi nízkým vstupním bariérám se cloudové technologie stávají dostupnými i pro malé podniky, které si mohou dovolit vhodnější a mnohem lépe zabezpečený software, než jaký by si mohli dovolit bez existence těchto řešení. Díky webovému přístupu se zároveň smazává závislost na platformě a k BI je tak možné přistupovat takřka kdykoliv, odkudkoliv a z jakéhokoliv zařízení.

V práci byl shrnut historický vývoj systémů pro podporu rozhodování a bylo ujasněno postavení těchto systémů v dnešním světě. V rámci prvního hlavního cíle práce došlo k identifikaci několika zjevných trendů, kterým dnešní BI systémy podléhají. Současná generace BI systémů se vyznačuje vysokou mírou samoobslužnosti a možností kolaborace, což bylo považováno za hlavní nedostatek první generace BI. Důraz je dále kladen na širší rozšíření mezi řadové zaměstnance, kteří tak s BI nástroji mohou aktivně pracovat a klást ad-hoc dotazy ve vztahu k obchodním problémům. Uživatelé již nejsou pouhými konzumenty obsahu (reportů), ale aktivně vytvářejí obsah vlastní a tvoří dotazy na vlastní otázky, které jim mohou přinést rychlou zpětnou vazbu či podklady pro zlepšení své činnosti.

Všechny zmíněné trendy de facto vycházejí z podstaty datově řízených firem, které si uvědomily, že k udržení konkurenceschopnosti je nezbytné důkladně sledovat a analyzovat své počínání, aby se ze svých dosavadních kroků mohly poučit. Firmy proto začínají sledovat takřka každý proces uvnitř i vně organizace, který s jejich činností souvisí. Vznikly nové zdroje dat, interní i externí, které je třeba do BI systémů integrovat, přičemž nutností je zpracovávat tato data ve vzájemném kontextu. Z hlediska koncepce dnešních BI je tak hlavním trendem uvedení dat do širšího kontextu, neboli v přeneseném slova smyslu „dát datům větší význam“.

Byly odhaleny i další trendy, které spolu úzce souvisí. Například větší počet zdrojů dat a nároky na jejich propojenost logicky zvyšují i nároky na kvalitu dat, kterou je nezbytné pro kvalitní výstupy zajistit. Došlo i k širšímu použití Data Miningu. Ačkoliv byly používané techniky DM známy již dříve, firmy dnes možnosti DM znovu objevují a začínají je více využívat. Oproti základním statistickým metodám dokáží v datech odhalit mnohem více skrytých souvislostí, jejichž odhalování je jinak v dnešních vysoce propojených datech velmi náročné. Posunulo se i zpracování semi-strukturovaných dat: některé dnešní cloudové datové sklady již nabízejí plnou podporu formátů JSON nebo XML. Zpracování nestrukturovaných dat je sice na vzestupu a objevují se nástroje, které z prostých textů dokáží vytěžit klíčové informace, ale stále se jedná o velmi okrajovou záležitost. Stejně tak téma Big Dat je dnes hojně diskutováno, ale stále není dostatečně uchopeno. Pojem není ustálen – proto nelze přesně definovat, co Big Data jsou a co nejsou.

Dnešní BI systémy je možné pořídit jak ve formě plně integrované platformy, tak jako jednotlivé komponenty. Vše záleží na tom, co zákazník od řešení očekává. Kompletní platformy nabízejí jednodušší integraci, optimalizaci výkonu a jsou rychlejší na spuštění. Naopak sestavení BI systému ze samostatných komponent může znamenat větší volnost a naplnění určitých specifických požadavků. Z hlediska přístupu k budování BI řešení je dnes cílem spíše do projektu postupně přidávat malé přírůstky, obsahující řešení konkrétních problémů, než velký projekt, který ihned integruje všechna podniková data s vizí vysoce komplexního řešení.

Druhým hlavním cílem práce bylo poukázat na značnou rozdílnost mezi komerčně nabízenými cloudovými BI nástroji. Cíle bylo dosaženo srovnáním dvou záměrně vybraných platforem, GoodData a Microsoft Power BI. Srovnání prokázalo, že přestože mají obě platformy umožňovat analýzu dat pomocí interaktivního grafického webového rozhraní – jsou natolik rozdílné, že každá z testovaných platforem je vhodná pro zcela odlišnou skupinu uživatelů. Nejpodstatnější rozdíly se týkaly datového modelu a práce s ním, respektive možností definování metrik, a také samotného vzhledu uživatelského rozhraní. Zatímco Power BI vyniká vzhledem, použitý jazyk DAX je velmi neohrabaný a těžkopádný. Naproti tomu GoodData umožňuje díky jazyku MAQL velmi jednoduché definování i složitých

metrik, avšak z pohledu vzhledu za konkurencí zaostává. Z tohoto pohledu obě platformy vynikají v jiných oblastech a nadneseně lze říci, že jsou svými přesnými protiklady.

Tato skutečnost – značná rozdílnost obou platform – poukazuje na velmi nesnadnou volbu správného BI. Firmy musejí být při výběru velmi obezřetné, neboť nesprávná volba by mohla znamenat zmařené investice. Kvůli velkému množství výrobců na trhu ale není snadné se v situaci zorientovat. Z tohoto důvodu dnes na trhu figurují výzkumné organizace, které se zkoumáním celého trhu s IS/ICT technologiemi zabývají a dokáží zákazníka celým procesem volby systému provést. Přestože takovéto poradenské služby nejsou vůbec levné, mnohdy za garanci správné volby tyto vynaložené prostředky jistě stojí.

Vedlejším cílem práce byla implementace cloudového BI v prostředí nejmenované společnosti „ABC“. Jelikož firma cloudové BI nástroje Keboola Connection a GoodData již měla, úkolem projektu bylo řešení doplnit a rozšířit. Projekt se zabíral analýzou akvizičního procesu, tedy procesu získávání nových zákazníků. Firma již analýzu tohoto procesu prováděla, ale původní datový model neumožňoval některé úhly pohledu. Cílem proto bylo navrhnout nový datový model, který by především snížil počet datumových dimenzí z původních čtyř na jednu, umožnil sledování délek jednotlivých kroků akvizičního procesu a umožnil zobrazit data na mapě ČR.

V rámci projektu byl navržen nový datový model, který neduhy stávajícího řešení napravil. Fungující prototyp dashboardu v GoodData byl představen zástupcům firmy, kteří následně uvedli, že navržené řešení zadané cíle splňuje, i když v některých ohledech je původní řešení lepší. Konkrétně projektu vytkli, že s jednou datumovou dimenzí nelze zobrazit aktuální stav AF, které byly vytvořeny v určitém období. Vhodným řešením by mohla být určitá kombinace obou datových modelů současně. Jediným cílem projektu, který se nepodařilo splnit, bylo zobrazení dat na mapě ČR, neboť se ukázalo, že zobrazení pouze na základě PSC není možné.

Zjevným přínosem nového přístupu je možnost sledování délek jednotlivých kroků akvizičního procesu. Analýza absolutních i průměrných hodnot může firmě pomoci v odhalování negativních odchylek, tak aby na ně firma mohla ihned reagovat. To může



v budoucnu přispět k optimalizaci jednotlivých kroků, vyšší spokojenosti zákazníků a tím i k vyšší ziskovosti firmy.

Z pohledu budoucího vývoje BI systémů je zřejmé, že se vývoj bude ubírat směrem mobilního BI, vyšší propojenosti dat, zpracování většího objemu dat, integrace nestrukturovaných dat, používání sofistikovanějších analytických metod a ještě lepších možností spolupráce a sdílení. Vše má přispět k tomu, aby lidé na správných místech dostávali úplné, aktuální a především správné informace, důležité pro rozhodování, což je ostatně hlavní podstatou celého Business Intelligence.

# Seznam použité literatury

## Citace

About Gartner. *Gartner* [online]. Stamford: Gartner, Inc, 2016 [cit. 2016-12-09]. Dostupné z: <http://www.gartner.com/technology/about.jsp>

ARIF, Mohamed. A history of cloud computing. In: *ComputerWeekly* [online]. London: ComputerWeekly.com, 2009 [cit. 2016-08-03]. Dostupné z: <http://www.computerweekly.com/feature/A-history-of-cloud-computing>

Best Infrastructure as a Service (IaaS). *G2 Crowd* [online]. Chicago: G2 Crowd, 2016 [cit. 2016-08-18]. Dostupné z: <https://www.g2crowd.com/categories/infrastructure-as-a-service-iaas>

BEYER, Mark A. a Douglas LANEY. The Importance of 'Big Data': A Definition. *Gartner* [online]. Stamford: Gartner, Inc, 2012 [cit. 2016-12-10]. Dostupné z: <https://www.gartner.com/doc/2057415/importance-big-data-definition>

Big Data Vs. Business Intelligence: What's the Difference?. In: *Yurbi: bring your data to live* [online]. Henderson: 5000fish, Inc, 2016 [cit. 2016-12-10]. Dostupné z: <https://www.yurbi.com/blog/big-data-vs-business-intelligence-whats-the-difference/>

Big Data, for better or worse: 90% of world's data generated over last two years. In: *Science Daily: Your source for the latest research news* [online]. Rockville: ScienceDaily, 2013 [cit. 2016-12-10]. Dostupné z: <https://www.sciencedaily.com/releases/2013/05/130522085217.htm>

Creating R visuals in the Power BI service. *Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-03]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-r-visuals/>

Data Loading / ETL. *GoodData Developers* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-23]. Dostupné z: <https://developer.gooddata.com/docs/data-loading>

Date Attribute Hierarchies. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-24]. Dostupné z: <https://help.gooddata.com/display/doc/Date+Attribute+Hierarchies>

DOLÁK, Ondřej. Big data: Nové způsoby zpracování a analýzy velkých objemů dat. In: *System On Line* [online]. Praha: CCB, spol. s r. o., 2011 [cit. 2016-12-11]. Dostupné z: <https://www.systemonline.cz/clanky/big-data.htm>

DONNELLY, James, James GIBSON a John IVANCEVICH. *Management*. 1. vyd. Praha: Grada, 1997. ISBN 978-80-7169-422-9.

Elasticsearch. *Elastic* [online]. Amsterdam: Elasticsearch BV, 2016 [cit. 2016-12-10]. Dostupné z: <https://www.elastic.co/products/elasticsearch>

FORECAST - Add a Trend Line to a Report. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-28]. Dostupné z: <https://help.gooddata.com/display/doc/FORECAST+++Add+a+Trend+Line+to+a+Report>

FÜRNKRANZ, Johannes, Dragan GAMBERGER a Nada LAVRAČ. *Foundations of rule learning*. 1st Edition. Heidelberg: Springer, 2012. ISBN 978-354-0751-977.

GÁLA, Libor, Jan POUR a Zuzana ŠEDIVÁ. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi*. 3., aktualizované vydání. Praha: Grada Publishing, 2015. Management v informační společnosti. ISBN 978-80-247-5457-4.

GENDRON, Michael. *Business intelligence and the cloud: strategic implementation guide*. Hoboken: Wiley, 2014. ISBN 1118631722.

*Geneea* [online]. Praha: GENEAA Analytics s.r.o, 2016 [cit. 2016-12-10]. Dostupné z: <https://geneea.com/>

Getting Started with MAQL. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-24]. Dostupné z: <https://help.gooddata.com/display/doc/Getting+Started+with+MAQL>

GLOTZBACH, Matthew. Google Apps is out of beta (yes, really): Insights from Googlers into our products, technology, and the Google culture. In: *Google Official blog* [online]. Mountain View: Google, 2009 [cit. 2016-08-04]. Dostupné z: <https://googleblog.blogspot.cz/2009/07/google-apps-is-out-of-beta-yes-really.html>

*GoodData Integrations* [online]. In: . San Francisco: GoodData Corporation, 2014, s. 1-1 [cit. 2016-11-23]. Dostupné z: [http://info.gooddata.com/rs/gooddata/images/GoodData\\_DataSources.pdf](http://info.gooddata.com/rs/gooddata/images/GoodData_DataSources.pdf)

*GoodData Open Analytics Platform Overview* [online]. San Francisco: GoodData Corporation, 2014 [cit. 2016-12-11]. Dostupné z: <http://info.gooddata.com/rs/gooddata/images/GoodData%20Platform%20Technical%20Brief.pdf>

GREGORY, Gwen. Putting Linked Data to Work. *Information Today*. Medford: Information Today, Inc, 2015, 32(7), 20. Dostupné také z: <http://search.proquest.com/docview/1724513127?accountid=17116>

HEINZE, Justin. History of Business Intelligence. In: *Better Buys* [online]. Malvern: Better Buys, 2014 [cit. 2016-08-30]. Dostupné z: <https://www.betterbuys.com/bi/history-of-business-intelligence/>

HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. *Big Data a NoSQL databáze*. 1. vydání. Praha: Grada, 2015. ISBN 978-80-247-5466-6.

HUŇKA, David. *Business Intelligence v podnikové praxi s využitím Cloud Computingu*. Praha, 2016. Bakalářská práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra informačních technologií. Vedoucí práce Doc. Ing. Jan Pour, CSc.

INMON, William, Derek. STRAUSS a Genia. NEUSHLOSS. *DW 2.0: The architecture for the next generation of data warehousing*. 1st Edition. Boston: Morgan Kaufmann, 2008. ISBN 978-012-3743-190.

KAVIS, Michael. *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)*. 1st Edition. Hoboken, New Jersey: Wiley, 2014. Wiley CIO series. ISBN 978-1-118-82627-0.

KOPÁČKOVÁ, Hana a Markéta ŠKROBÁČKOVÁ. DECISION SUPPORT SYSTEMS OR BUSINESS INTELLIGENCE: WHAT CAN HELP IN DECISION MAKING?. In: *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration* [online]. Pardubice: University of Pardubice, 2006, s. 98-103 [cit. 2016-08-29]. ISSN 1211-555X. Dostupné z: <http://hdl.handle.net/10195/32436>

LACHLAN, James. Defining Business Intelligence 3.0. In: *Yellowfin* [online]. Melbourne: Yellowfin International Pty Ltd, 2014 [cit. 2016-10-02]. Dostupné z: <http://www.yellowfinbi.com/YFCCommunityNews-Defining-Business-Intelligence-3-0-159445>

LEONDES, Cornelius. *Expert systems: the technology of knowledge management and decision making for the 21st century*. 1st. San Diego: Academic Press, 2002. ISBN 01-244-3880-6.

Let's get started with Data Permissions. *GoodData Developers* [online]. San Francisco: GoodData Corporation, 2016 [cit. 2016-11-29]. Dostupné z: <https://developer.gooddata.com/article/lets-get-started-with-data-permissions>

Magic Quadrant for Business Intelligence and Analytics Platforms. *Gartner* [online]. Stamford: Gartner, Inc., 2016 [cit. 2016-11-23]. Dostupné z: <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>

MARR, Bernard. Why We Must Rethink Self-Service BI, Analytics And Reporting. *Forbes* [online]. Jersey City: Forbes Media, 2016 [cit. 2016-12-10]. Dostupné z: <http://www.forbes.com/sites/bernardmarr/2016/10/25/why-we-must-rethink-self-service-bi-analytics-and-reporting/#14a3a83d4349>

MAYER-SCHÖNBERGER, Viktor a Kenneth CUKIER. *Big Data: Revoluce, která změní způsob, jak žijeme, pracujeme a myslíme*. 1. vyd. Brno: Computer Press, 2014. ISBN 978-80-251-4119-9.

MEARIAN, Lucas. World's data will grow by 50X in next decade, IDC study predicts. In: *COMPUTERWORLD* [online]. Framingham: Computerworld, Inc, 2011 [cit. 2016-10-03]. Dostupné z: <http://www.computerworld.com/article/2509588/data-center/world-s-data-will-grow-by-50x-in-next-decade--idc-study-predicts.html>

MITZNER, Dennis. What is a data-driven company?. In: *InfoWorld* [online]. San Francisco: IDG Communications, Inc., 2016 [cit. 2016-12-17]. Dostupné z: <http://www.infoworld.com/article/3074322/big-data/what-is-a-data-driven-company.html>

NÁHLOVSKÝ, Tomáš. *Business Intelligence v prostředí Cloudu*. Praha, 2016. Diplomová práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra informačních technologií. Vedoucí práce Doc. Ing. Jan Pour, CSc.

NORMANDEAU, Kevin. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity. In: *InsideBIGDATA* [online]. Portland: InsideBIGDATA, 2013 [cit. 2016-12-10]. Dostupné z: <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>

*OpenRefine* [online]. OpenRefine, 2016 [cit. 2016-11-13]. Dostupné z: <http://openrefine.org/>

PEDERSEN, Ulrik. How to Become a Data-Driven Company. In: *TDWI* [online]. Renton: TDWI, 2015 [cit. 2016-12-17]. Dostupné z: <https://tdwi.org/articles/2015/08/04/How-to-Become-a-Data-Driven-Company.aspx>

Platy v kategorii: Informační technologie. *Platy.cz* [online]. Praha: Profesia CZ, spol. s r.o., 2016 [cit. 2016-12-17]. Dostupné z: <http://www.platy.cz/platy/informacni-technologie>

POWER, Daniel A Brief History of Decision Support Systems. *DSSResources.COM* [online]. Version 4.1. Iowa: DSSResources, 2007, 2007-03-10 [cit. 2016-08-30]. Dostupné z: <http://dssresources.com/history/dsshhistory.html>

POZIN, Ilya. The Data You Ignore Is More Valuable Than You Thought. In: *Forbes* [online]. Jersey City: Forbes Media, 2014 [cit. 2016-12-10]. Dostupné z: <http://www.forbes.com/sites/ilyapozin/2014/06/09/the-data-you-ignore-is-more-valuable-than-you-thought/>

Pricing. *GoodData* [online]. San Francisco: GoodData Corporation, 2010 [cit. 2016-12-07].  
Dostupné z:

<http://web.archive.org/web/20100811023135/http://www.gooddata.com/pricing/>

Pricing. *Power BI* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-04].  
Dostupné z: <https://powerbi.microsoft.com/en-us/pricing/>

Q&A in Power BI. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a/>

Refresh a dataset created from an Excel workbook on OneDrive, or SharePoint Online. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-refresh-excel-file-onedrive/>

Service Models. *Cloud Lighthouse* [online]. Kontich: cloudlighthouse.be, 2016 [cit. 2016-10-19]. Dostupné z: <http://cloudlighthouse.be/cloud/service-models>

Sítě Miloše Čermáka: Firma z každého zaměstnance udělala analytika dat. A záhy stouply prodeje. In: *IHNED.cz: Hospodářské noviny* [online]. Praha: IHNED.cz, 2015 [cit. 2016-07-13]. Dostupné z: <http://archiv.ihned.cz/c1-64900010-socialni-site-milose-cermaka-firma-primela-zamestnance-denne-sledovat-data-prodeje-zahy-stouply-prodeje>

SMITH, David. Demand for R jobs on the rise, while SAS jobs decline. In: *Revolutions: Daily news about using open source R for big data analysis, predictive modeling, data science, and visualization since 2008* [online]. Redmond: Microsoft Corporation, 2013 [cit. 2016-12-11]. Dostupné z: <http://blog.revolutionanalytics.com/2013/08/job-trends-for-statistics-packages.html>

Statistical Functions. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2016 [cit. 2016-11-28]. Dostupné z: <https://help.gooddata.com/display/doc/Statistical+Functions>

Tabular Model Databases (SSAS Tabular). *Microsoft Developer Network* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-01]. Dostupné z: <https://msdn.microsoft.com/en-us/library/hh230906.aspx>

TextualETL™ Technology. *FOREST RIM TECHNOLOGY* [online]. Castle Rock: FOREST RIM TECHNOLOGY, INC., 2016 [cit. 2016-12-10]. Dostupné z: <http://www.forestrimtech.com/textual-etl/textual-etl-technology>

The DATE, DATETIME, and TIMESTAMP Types. *MySQL* [online]. Redwood: Oracle Corporation, 2016 [cit. 2016-10-23]. Dostupné z: <http://dev.mysql.com/doc/refman/5.7/en/datetime.html>

Time Intelligence Functions (DAX). *Microsoft Developer Network* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-02]. Dostupné z: <https://msdn.microsoft.com/en-us/library/ee634763.aspx?f=255&MSPPErr=-2147217396>

Time Series Forecasting. *GoodData Help* [online]. San Francisco: GoodData Corporation, 2015 [cit. 2016-11-28]. Dostupné z: <https://help.gooddata.com/display/doc/Time+Series+Forecasting>

TOBOLKA, Jiří. How to load time in GoodData?. *Stack Overflow* [online]. Prague: Stack Exchange Inc, 2014 [cit. 2016-11-24]. Dostupné z: <http://stackoverflow.com/questions/20964162/how-to-load-time-in-gooddata>

Use DirectQuery in Power BI Desktop. *Microsoft Power BI Documentation* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-06]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-use-directquery/>

VELTE, Anthony, Toby VELTE a Robert ELSENPETER. *Cloud Computing: praktický průvodce*. 1. vyd. Brno: Computer Press, 2011. ISBN 978-80-251-3333-0.

VERBORGH, Ruben, Max DE WILDE a Aniket SAWANT. *Using OpenRefine: The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web*. 1st Edition. Birmingham, England: Packt Publishing, 2013. Community experience distilled. ISBN 978-1-78328-909-7.

Visuals library. *Power BI* [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-03]. Dostupné z: <https://app.powerbi.com/visuals/>



VOLENA, Karel. *Cloud computing?*. Praha, 2013. Bakalářská práce. Bankovní institut vysoká škola Praha, Katedra matematiky, statistiky a informačních technologií. Vedoucí práce Ing. Vladimír Beneš.

WEINS, Kim. Cloud Computing Trends: 2016 State of the Cloud Survey. In: *RIGHT SCALE: CLOUD MANAGEMENT BLOG* [online]. Santa Barbara: RightScale Inc., 2016 [cit. 2016-08-29]. Dostupné z: <http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2016-state-cloud-survey>

What is Power BI? Microsoft Power BI [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-20]. Dostupné z: <https://powerbi.microsoft.com/en-us/what-is-power-bi/>

ZÁDOVÁ, Vladimíra. *Specifika postavení a návrhu datových skladů v rámci IS/ICT: Specifics of Data Warehouses' Status and Design in IS/ICT* [CD-ROM]. Liberec: Technická univerzita v Liberci, 2006. Disertační práce. Technická univerzita v Liberci. Vedoucí práce Prof. Ing. Jan Ehleman, CSc.

## **Bibliografie**

GoodData Help [online]. San Francisco: GoodData Corporation, 2016 [cit. 2016-12-13]. Dostupné z: <https://help.gooddata.com/>

Power BI Documentation [online]. Redmond: Microsoft Corporation, 2016 [cit. 2016-12-13]. Dostupné z: <https://powerbi.microsoft.com/en-us/documentation/>

ŠIMEČEK, Petr. KEBOOLA S.R.O. Padakův deníček [online]. Praha: Petr Šimeček, 2016 [cit. 2016-12-13]. Dostupné z: <http://padak.keboola.com/>

Elektronická databáze článků ProQuest (knihovna.tul.cz)

# Seznam příloh

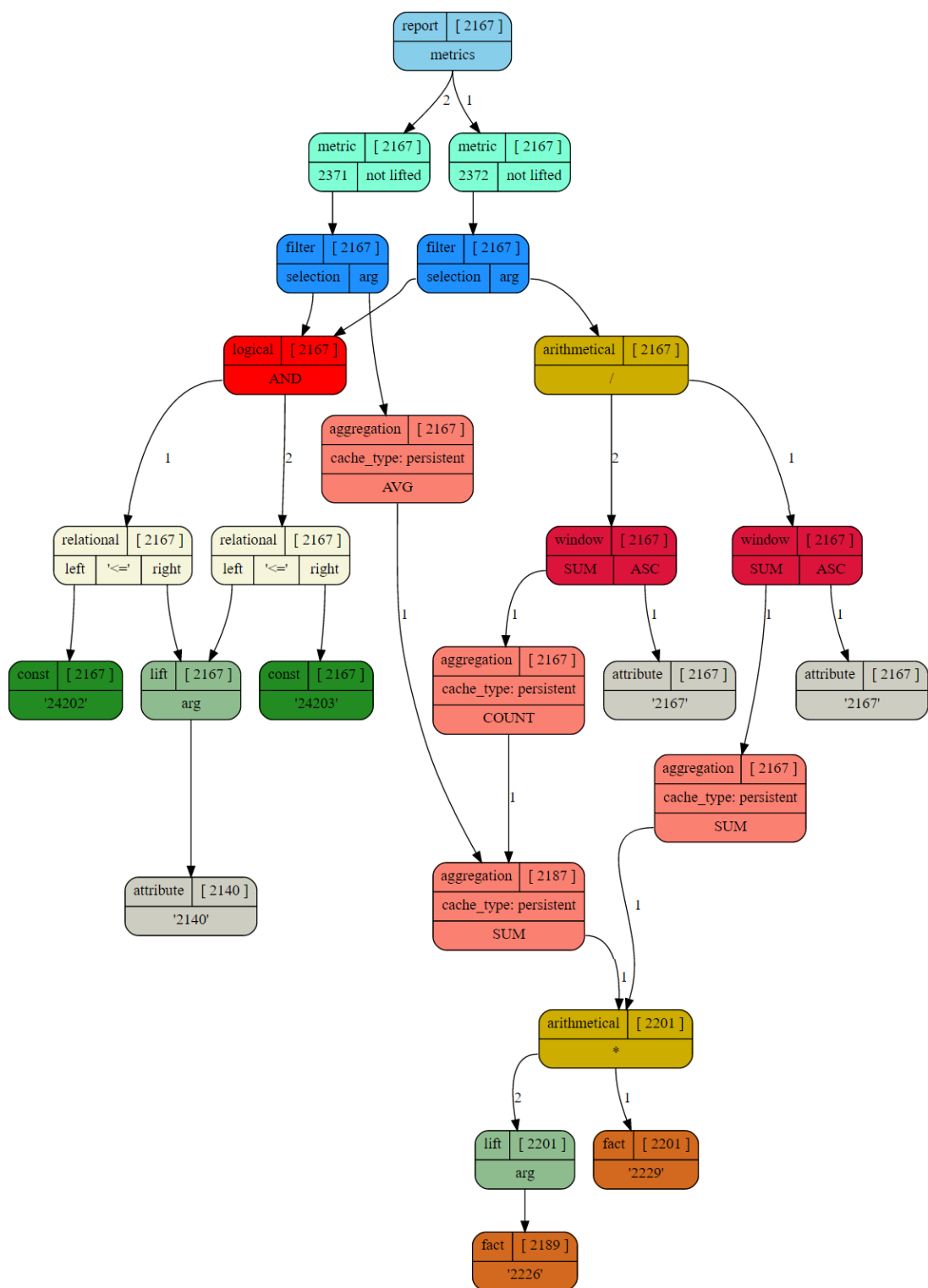
## Tištěné přílohy

Příloha A – Strom dotazů vygenerovaný v GoodData .....	147
Příloha B – Vygenerovaný SQL skript reportu v GoodData.....	148
Příloha C – Specifikace tabulek AF a CONT.....	150
Příloha D – SQL skript výsledné transformace .....	151
Příloha E – Datový model GoodData .....	154

## Přílohy na DVD

Příloha DVD 1 – Data vygenerovaná pro testování ve formátu CSV
Příloha DVD 2 – Projekt GoodData Cloud Connect Designer
Příloha DVD 3 – Projekt Microsoft Power BI Desktop
Příloha DVD 4 – SQL skript výsledné transformace

## Příloha A – Strom dotazů vygenerovaný v GoodData



Obrázek A-1: Strom dotazů vygenerovaný v GoodData

Zdroj: vlastní

## Příloha B – Vygenerovaný SQL skript reportu v GoodData

```
CREATE TABLE "qt_7d346c409963c8f53d477f2728e425eb"
AS(
  SELECT "f_ordersproducts_aa5j20uh3jac6gy".
    "orders_id"
  AS "a_2187",
  SUM(("f_ordersproducts_aa5j20uh3jac6gy".
    "f_quantity" * "f_products_aa7p2tbu41alkfd".
    "f_products")) AS "c_f988907fdf84bc34",
  TRUE AS "def_f988907fdf84bc34"
  FROM("f_ordersproducts_aa5j20uh3jac6gy"
    INNER JOIN "f_products_aa7p2tbu41alkfd"
    ON("f_ordersproducts_aa5j20uh3jac6gy".
      "products_id" = "f_products_aa7p2tbu41alkfd".
      "id")) GROUP BY 1)
```

```
ANALYZE "qt_7d346c409963c8f53d477f2728e425eb"
```

```
CREATE TABLE "qt_038b67135ef4f7b911c8f1489c5624bb"
AS(
  SELECT "f_orders_abbd2ufun6ajtjs".
    "dt_ordersdate_id"
  AS "a_2167",
  AVG("qt".
    "c_f988907fdf84bc34") AS "c_aef6fc8047fb2c90",
  COUNT("qt".
    "c_f988907fdf84bc34") AS "c_f7f52570cc1e36d5",
  TRUE AS "def_aef6fc8047fb2c90",
  TRUE AS "def_f7f52570cc1e36d5"
  FROM("qt_7d346c409963c8f53d477f2728e425eb"
    AS "qt"
    INNER JOIN "f_orders_abbd2ufun6ajtjs"
    ON("qt".
      "a_2187" = "f_orders_abbd2ufun6ajtjs".
      "id")) WHERE "qt".
    "def_f988907fdf84bc34"
  GROUP BY 1)
```

```
ANALYZE "qt_038b67135ef4f7b911c8f1489c5624bb"
```

```
CREATE TABLE "qt_cf81b8988f2bed094e18265400a750eb"
AS(
  SELECT "f_orders_abbd2ufun6ajtjs".
    "dt_ordersdate_id"
  AS "a_2167",
  SUM(("f_ordersproducts_aa5j20uh3jac6gy".
    "f_quantity" * "f_products_aa7p2tbu41alkfd".
    "f_products")) AS "c_b92bb541f78dcd1c",
  TRUE AS "def_b92bb541f78dcd1c"
  FROM(("f_ordersproducts_aa5j20uh3jac6gy"
    INNER JOIN "f_products_aa7p2tbu41alkfd"
    ON("f_ordersproducts_aa5j20uh3jac6gy".
      "products_id" = "f_products_aa7p2tbu41alkfd".
      "id")) INNER JOIN "f_orders_abbd2ufun6ajtjs"
    ON("f_ordersproducts_aa5j20uh3jac6gy".
      "orders_id" = "f_orders_abbd2ufun6ajtjs".
      "id")) GROUP BY 1)
```

```
ANALYZE "qt_cf81b8988f2bed094e18265400a750eb"
```

```
CREATE TABLE "qt_rc_5622e2bc995686c0066032bbeda0c939"
AS(
  SELECT "dwh_dm_aa2z2tiykkabo71".
    "id"
  AS "a_2167", (
    CASE WHEN(COALESCE("sub".
      "def_64dfa65c579c123",
      FALSE) AND COALESCE("sub".
      "def_a2d3496307c36903",
      FALSE)) THEN(CAST("sub".
      "c_64dfa65c579c123"
      AS double precision) / NULLIF("sub".
      "c_a2d3496307c36903",
      0)) ELSE CAST(NULL AS INT) END) AS "m_1", (
    CASE WHEN COALESCE("qt".
      "def_aef6fc8047fb2c90",
      FALSE) THEN "qt".
```

```

"c_aef6fc8047fb2c90"
ELSE CAST(NULL AS INT) END) AS "m_2"
FROM(((
    SELECT COALESCE("qt".
        "a_2167",
        "qt_1".
        "a_2167") AS "a_2167",
    SUM((
        CASE WHEN COALESCE("qt".
            "def_b92bb541f78dcd1c",
            FALSE) THEN "qt".
            "c_b92bb541f78dcd1c"
        ELSE CAST(NULL AS INT) END)) OVER(
            ORDER BY COALESCE("qt".
                "a_2167",
                "qt_1".
                "a_2167") ASC NULLS FIRST ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS
"c_64dfa65c579c123",
    SUM((
        CASE WHEN COALESCE("qt_1".
            "def_f7f52570cc1e36d5",
            FALSE) THEN "qt_1".
            "c_f7f52570cc1e36d5"
        ELSE CAST(NULL AS INT) END)) OVER(
            ORDER BY COALESCE("qt".
                "a_2167",
                "qt_1".
                "a_2167") ASC NULLS FIRST ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS
"c_a2d3496307c36903",
    COALESCE("qt".
        "def_b92bb541f78dcd1c",
        FALSE) AS "def_64dfa65c579c123",
    COALESCE("qt_1".
        "def_f7f52570cc1e36d5",
        FALSE) AS "def_a2d3496307c36903"
    FROM("qt_cf81b8988f2bed094e18265400a750eb"
        AS "qt"
        FULL OUTER JOIN "qt_038b67135ef4f7b911c8f1489c5624bb"
        AS "qt_1"
        ON("qt".
            "a_2167" = "qt_1".
            "a_2167")) WHERE(COALESCE("qt".
            "def_b92bb541f78dcd1c",
            FALSE) OR COALESCE("qt_1".
            "def_f7f52570cc1e36d5",
            FALSE))) AS "sub"
    FULL OUTER JOIN "qt_038b67135ef4f7b911c8f1489c5624bb"
    AS "qt"
    ON("sub".
        "a_2167" = "qt".
        "a_2167")) INNER JOIN "dwh_dm_aa2z2tiykkabo71"
    ON(COALESCE("sub".
        "a_2167",
        "qt".
        "a_2167") = "dwh_dm_aa2z2tiykkabo71".
        "id")) WHERE((24202 <= "dwh_dm_aa2z2tiykkabo71".
        "id_month") AND("dwh_dm_aa2z2tiykkabo71".
        "id_month" <= 24203) AND(COALESCE("qt".
        "def_aef6fc8047fb2c90",
        FALSE) OR(COALESCE("sub".
        "def_64dfa65c579c123",
        FALSE) AND COALESCE("sub".
        "def_a2d3496307c36903",
        FALSE))))))
ANALYZE "qt_rc_5622e2bc995686c0066032bbeda0c939"

```

## Příloha C – Specifikace tabulek AF a CONT

AF	CONT
primary_key	primary_key
PARTNER_TYPE	PARTNER_TYPE
PARTNER_NAME	PARTNER_NAME
PARTNER_DN	PARTNER_DN
PARTNER_IC	PARTNER_IC
PARTNER_LEADID	PARTNER_LEADID
PARTNER_ID	PARTNER_ID
CONT_CONTRACTNO	CONT_CONTRACTNO
CONT_SIGNATURE_DATE	CONT_SIGNATURE_DATE
OPM_IDENT	OPM_IDENT
OPM_COMMODITY	OPM_COMMODITY
OPM_CONS_MWH	OPM_CONS_MWH
OPM_REGION	OPM_REGION
OPM_DISTRICT	OPM_DISTRICT
OPM_CITY	OPM_CITY
OPM_POSTCODE	OPM_POSTCODE
OPM_DISTRIBUTOR	OPM_DISTRIBUTOR
OPM_PREV_SUPPLIER	OPM_PREV_SUPPLIER
OPM_DEALER_NAME	OPM_DEALER_NAME
AF_CREATED_USER	PARTNER_NO
AF_CREATED_DATE	ACCOUNT_NO
AF_STATUS	OPM_VALID_FROM
AF_VALID	OPM_VALID_TO
AF_STATUS_DETAIL	OPM_SUPPLY_CONFIRMED
AF_PRICELIST_NAME	OPM_CANCEL_TYPE
OPM_TDD_CODE	OPM_CANCEL_DATE
OPM_DISTR_TARIF	OPM_CANCEL_REASON
OPM_CONS_MWH_VT	OPM_BILLING_PERIOD
OPM_CONS_MWH_NT	OPM_ID
AF_SOURCE	CONT_ID
PARTNER_SEGMENT_CODE	ACCOUNT_ID
DEPARTMENT_NAME	OPM_PRICELIST_NAME
DISCOUNT_NAME	OPM_TDD_CODE
AF_SEND_METHOD	OPM_DISTR_TARIF
AF_SEND_DATE	OPM_CONS_MWH_VT
AF_RETURNED_STATE	OPM_CONS_MWH_NT
AF_RETURNED_DATE	PARTNER_SEGMENT_CODE
AF_PROCESSED_DATE	TRADE_ALG_VALID_FROM
AF_UNPROCESSED_REASON	TRADE_ALG_VALID_TO
TRADE_PRICELIST_ID	TRADE_ALG_DISCOUNT_TYPE
PARTNER_MOBILE_NUMBER	TRADE_ALG_DISCOUNT_VALUE1
	TRADE_ALG_DISCOUNT_VALUE2
	TRADE_ALG_DISCOUNT_VALUE3
	TRADE_ALG_DISCOUNT_FLATFEE
	TRADE_PRICELIST_ID
	NEW_SUPPLIER_ID
	NEW_SUPPLIER_NAME

Obrázek C-1: Specifikace tabulek AF a CONT

Zdroj: vlastní

## Příloha D – SQL skript výsledné transformace

```
CREATE TABLE `AF_CONT`
(
  `primary_key`          VARCHAR(255) NOT NULL,
  `primary_key_af`      VARCHAR(255) NULL,
  `primary_key_cont`    VARCHAR(255) NULL,
  `partner_type`        VARCHAR(255) NULL,
  `partner_name`        TEXT NULL,
  `partner_no`          TEXT NULL,
  `partner_ic`          TEXT NULL,
  `partner_id`          TEXT NULL,
  `partner_mobile_number` TEXT NULL,
  `cont_id`             TEXT NULL,
  `cont_contractno`     TEXT NULL,
  `opm_ident`           TEXT NULL,
  `opm_commodity`       VARCHAR(255) NULL,
  `opm_cons_mwh`        TEXT NULL,
  `opm_postcode`        TEXT NULL,
  `opm_distributor`     TEXT NULL,
  `opm_prev_supplier`   TEXT NULL,
  `opm_dealer_name`     TEXT NULL,
  `opm_tdd_code`        TEXT NULL,
  `opm_distr_tarif`     TEXT NULL,
  `opm_cons_mwh_vt`     TEXT NULL,
  `opm_cons_mwh_nt`     TEXT NULL,
  `opm_supply_confirmed` TEXT NULL,
  `opm_cancel_type`     TEXT NULL,
  `opm_billing_period`  TEXT NULL,
  `opm_id`              TEXT NULL,
  `opm_pricelist_name`  TEXT NULL,
  `af_created_user`     TEXT NULL,
  `af_status_detail`    TEXT NULL,
  `af_pricelist_name`   TEXT NULL,
  `af_send_method`      TEXT NULL,
  `af_unprocessed_reason` TEXT NULL,
  `discount_name`       TEXT NULL,
  `trade_pricelist_id`  TEXT NULL,
  `account_id`          TEXT NULL,
  `new_supplier_id`     TEXT NULL,
  `new_supplier_name`   TEXT NULL,
  `opm_cacel_before`    TEXT NULL,
  PRIMARY KEY (`primary_key`),
  INDEX `i_af_cont_primary_key_af` (`primary_key_af`),
  INDEX `i_af_cont_primary_key_cont` (`primary_key_cont`),
  INDEX `i_af_cont_partner_type` (`partner_type`),
  INDEX `i_af_cont_opm_commodity` (`opm_commodity`)
) ENGINE=INNODB DEFAULT CHARSET=UTF8
;

-- af + cont

INSERT INTO `AF_CONT`
SELECT CONCAT(`AF`.`primary_key`, '-', `CONT`.`primary_key`) AS `primary_key`,
       `AF`.`primary_key` AS `primary_key_af`,
       `CONT`.`primary_key` AS `primary_key_cont`,
       COALESCE(NULLIF(`CONT`.`partner_type`, ''), `AF`.`partner_type`) AS `partner_type`,
       COALESCE(NULLIF(`CONT`.`partner_name`, ''), `AF`.`partner_name`) AS `partner_name`,
       `CONT`.`partner_no`,
       COALESCE(NULLIF(`CONT`.`partner_ic`, ''), `AF`.`partner_ic`) AS `partner_ic`,
       COALESCE(NULLIF(`CONT`.`partner_id`, ''), `AF`.`partner_id`) AS `partner_id`,
       `AF`.`partner_mobile_number`,
       `CONT`.`cont_id`,
       COALESCE(NULLIF(`CONT`.`cont_contractno`, ''), `AF`.`cont_contractno`) AS `cont_contractno`,
       COALESCE(NULLIF(`CONT`.`opm_ident`, ''), `AF`.`opm_ident`) AS `opm_ident`,
       COALESCE(NULLIF(`CONT`.`opm_commodity`, ''), `AF`.`opm_commodity`) AS `opm_commodity`,
       COALESCE(NULLIF(`CONT`.`opm_cons_mwh`, ''), `AF`.`opm_cons_mwh`) AS `opm_cons_mwh`,
       COALESCE(NULLIF(`CONT`.`opm_postcode`, ''), `AF`.`opm_postcode`) AS `opm_postcode`,
       COALESCE(NULLIF(`CONT`.`opm_distributor`, ''), `AF`.`opm_distributor`) AS `opm_distributor`,
       COALESCE(NULLIF(`CONT`.`opm_prev_supplier`, ''), `AF`.`opm_prev_supplier`) AS `opm_prev_supplier`,
       COALESCE(NULLIF(`CONT`.`opm_dealer_name`, ''), `AF`.`opm_dealer_name`) AS `opm_dealer_name`,
       COALESCE(NULLIF(`CONT`.`opm_tdd_code`, ''), `AF`.`opm_tdd_code`) AS `opm_tdd_code`,
       COALESCE(NULLIF(`CONT`.`opm_distr_tarif`, ''), `AF`.`opm_distr_tarif`) AS `opm_distr_tarif`,
       COALESCE(NULLIF(`CONT`.`opm_cons_mwh_vt`, ''), `AF`.`opm_cons_mwh_vt`) AS `opm_cons_mwh_vt`,
       COALESCE(NULLIF(`CONT`.`opm_cons_mwh_nt`, ''), `AF`.`opm_cons_mwh_nt`) AS `opm_cons_mwh_nt`,
       `CONT`.`opm_supply_confirmed`,
       `CONT`.`opm_cancel_type`,
       `CONT`.`opm_billing_period`,
```

```

`CONT`.`opm_id`,
`CONT`.`opm_pricelist_name`,
`AF`.`af_created_user`,
`AF`.`af_status_detail`,
`AF`.`af_pricelist_name`,
`AF`.`af_send_method`,
`AF`.`af_unprocessed_reason`,
`AF`.`discount_name`,
CAST(COALESCE(NULLIF(`CONT`.`trade_pricelist_id`, ''), NULLIF(`AF`.`trade_pricelist_id`, '')) AS
DECIMAL(10,0)) AS `trade_pricelist_id`,
`CONT`.`account_id`,
`CONT`.`new_supplier_id`,
`CONT`.`new_supplier_name`,
CASE WHEN `CONT`.`opm_valid_from` != ''
      AND `CONT`.`opm_valid_to` != ''
      AND `CONT`.`opm_valid_from` = `CONT`.`opm_valid_to`
      THEN 'Ano' ELSE 'Ne'
END
FROM `CONT`
JOIN `AF` ON (`AF`.`cont_contractno` = `CONT`.`cont_contractno` AND `AF`.`opm_ident` = `CONT`.`opm_ident`)
WHERE `AF`.`af_valid` = 'A'
      AND `AF`.`cont_contractno` != ''
      AND `AF`.`opm_ident` != ''
;

-- zbytek af

INSERT INTO `AF_CONT`
SELECT `AF`.`primary_key`,
`AF`.`primary_key`,
NULL,
`AF`.`partner_type`,
`AF`.`partner_name`,
NULL,
`AF`.`partner_ic`,
`AF`.`partner_id`,
`AF`.`partner_mobile_number`,
NULL,
`AF`.`cont_contractno`,
`AF`.`opm_ident`,
`AF`.`opm_commodity`,
`AF`.`opm_cons_mwh`,
`AF`.`opm_postcode`,
`AF`.`opm_distributor`,
`AF`.`opm_prev_supplier`,
`AF`.`opm_dealer_name`,
`AF`.`opm_tdd_code`,
`AF`.`opm_distr_tarif`,
`AF`.`opm_cons_mwh_vt`,
`AF`.`opm_cons_mwh_nt`,
NULL,
NULL,
NULL,
NULL,
NULL,
`AF`.`af_created_user`,
`AF`.`af_status_detail`,
`AF`.`af_pricelist_name`,
`AF`.`af_send_method`,
`AF`.`af_unprocessed_reason`,
`AF`.`discount_name`,
CAST(NULLIF(`AF`.`trade_pricelist_id`, '') AS DECIMAL(10,0)) AS `trade_pricelist_id`,
NULL,
NULL,
NULL,
'Ne'
FROM `AF`
LEFT JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF_CONT`.`primary_key_af` IS NULL
      AND `AF`.`af_valid` = 'A'
      AND `AF`.`cont_contractno` != ''
      AND `AF`.`opm_ident` != ''
;

-- update partner_type

UPDATE `AF_CONT`
JOIN `PARTNER_TYPE` ON `PARTNER_TYPE`.`hodnota` = `AF_CONT`.`partner_type`
SET `AF_CONT`.`partner_type` = `PARTNER_TYPE`.`popis`
;

```



```

-- update opm_commodity

UPDATE `AF_CONT`
JOIN `OPM_COMMODITY` ON `OPM_COMMODITY`.`hodnota` = `AF_CONT`.`opm_commodity`
SET `AF_CONT`.`opm_commodity` = `OPM_COMMODITY`.`popis`
;

CREATE TABLE `AF_CONT_EVENTS`
(
  `primary_key`          VARCHAR(255) NOT NULL,
  `af_cont_id`          VARCHAR(255) NOT NULL,
  `event_id`            VARCHAR(255) NOT NULL,
  `event_name`         VARCHAR(255) NOT NULL,
  `event_date`         DATE NOT NULL,
  `days_from_prev`    INT(11) NULL,
  `days_from_af_created` INT(11) NULL,
  PRIMARY KEY (`primary_key`),
  index `i_AF_CONT_EVENTS_af_cont_id` (`af_cont_id`),
  index `i_AF_CONT_EVENTS_event_id` (`event_id`)
) ENGINE=INNODB DEFAULT CHARSET=UTF8
;

-- eventy

INSERT INTO `AF_CONT_EVENTS`
SELECT CONCAT(`AF_CONT`.`primary_key`, '_', `af_created_date`),
       `AF_CONT`.`primary_key`,
       `af_created_date`,
       'Datum vytvoření AF',
       STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y'),
       NULL,
       NULL
FROM `AF`
JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF`.`af_created_date` != ''
      AND `AF`.`af_created_date` IS NOT NULL
;

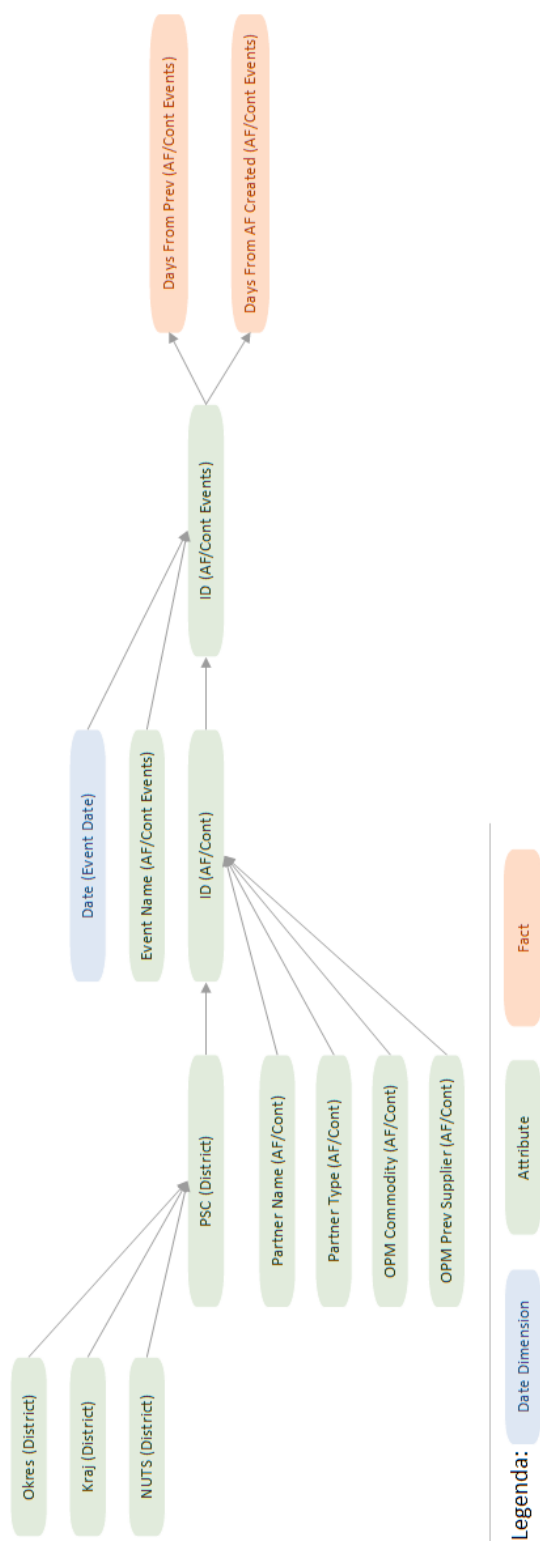
INSERT INTO `AF_CONT_EVENTS`
SELECT CONCAT(`AF_CONT`.`primary_key`, '_', `af_send_date`),
       `AF_CONT`.`primary_key`,
       `af_send_date`,
       'Datum odeslání AF',
       STR_TO_DATE(`AF`.`af_send_date`, '%d.%m.%Y'),
       DATEDIFF(STR_TO_DATE(`AF`.`af_send_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y')),
       DATEDIFF(STR_TO_DATE(`AF`.`af_send_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y'))
FROM `AF`
JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF`.`af_send_date` != ''
      AND `AF`.`af_send_date` IS NOT NULL
;

INSERT INTO `AF_CONT_EVENTS`
SELECT CONCAT(`AF_CONT`.`primary_key`, '_', `af_returned_date`),
       `AF_CONT`.`primary_key`,
       `af_returned_date`,
       'Datum vrácení AF',
       STR_TO_DATE(`AF`.`af_returned_date`, '%d.%m.%Y'),
       DATEDIFF(STR_TO_DATE(`AF`.`af_returned_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_send_date`, '%d.%m.%Y')),
       DATEDIFF(STR_TO_DATE(`AF`.`af_returned_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y'))
FROM `AF`
JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF`.`af_returned_date` != ''
      AND `AF`.`af_returned_date` IS NOT NULL
;

INSERT INTO `AF_CONT_EVENTS`
SELECT CONCAT(`AF_CONT`.`primary_key`, '_', `af_processed_date`),
       `AF_CONT`.`primary_key`,
       `af_processed_date`,
       'Datum zpracování AF',
       STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'),
       DATEDIFF(STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_returned_date`, '%d.%m.%Y')),
       DATEDIFF(STR_TO_DATE(`AF`.`af_processed_date`, '%d.%m.%Y'), STR_TO_DATE(`AF`.`af_created_date`, '%d.%m.%Y'))
FROM `AF`
JOIN `AF_CONT` ON `AF_CONT`.`primary_key_af` = `AF`.`primary_key`
WHERE `AF`.`af_processed_date` != ''
      AND `AF`.`af_processed_date` IS NOT NULL
;

```

## Příloha E – Datový model GoodData



Obrázek E-1: Datový model GoodData

Zdroj: vlastní