

Průzkum Business Intelligence řešení a jeho nasazení pro velká dopravní data

Bc. Dominika Pospíchalová

Diplomová práce
2019



Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

Univerzita Tomáše Bati ve Zlíně

Fakulta aplikované informatiky

akademický rok: 2018/2019

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Dominika Pospíchalová**

Osobní číslo: **A16571**

Studijní program: **N3902 Inženýrská informatika**

Studijní obor:

Forma studia: **kombinovaná**

Téma práce: **Průzkum Business Intelligence řešení a jeho nasazení pro velká dopravní data**

Téma anglicky: **A Survey of Business Intelligence Solutions and its Deployment for Large Transport Data**

Zásady pro vypracování

1. Seznamte se se softwarovými nástroji používaných v oblasti Business Intelligence.
2. V teoretické části práce popište vlastnosti konkrétních Business Intelligence řešení, vhodných pro zpracování a vizualizaci velkého množství dat.
3. Seznamte se se strukturou a obsahem reálných dat získaných monitorováním důležitých údajů ve městech.
4. Na základě vstupních dat a požadavků potřebných výstupů vyberte nejvhodnější softwarové řešení Business Intelligence.
5. Ve vybraném softwaru proveďte zpracování reálných dat a navrhnete vhodné vizuální výstupy tak, aby mohly být použité v průmyslové praxi.

Rozsah bakalářské práce:

Rozsah příloh:

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

Seznam doporučené literatury:

1. POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business intelligence v podnikové praxi. 1. vyd. Praha: Professional Publishing, 2012, 276 s. ISBN 978-80-7431-065-2.
2. NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. Business intelligence: jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada, 2005, 254 s. Management v informační společnosti. ISBN 80-247-1094-3.
3. LACKO, Ľuboslav. Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby. Vyd. 1. Brno: Computer Press, 2009, 456 s. ISBN 978-80-251-2887-9.
4. SCHEPS, Swain. Business intelligence for dummies. Hoboken, NJ: Wiley, c2008. ISBN 04-701-2723-6.
5. CUESTA, Hector. Analýza dat v praxi. Brno: Computer Press, 2015, 296 s. ISBN 978-80-251-4361-2.
6. HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. Big Data a NoSQL databáze. Praha: Grada, 2015, 281 s. Profes!onal. ISBN 978-80-247-5466-6.
7. MongoDB \matsymb{lbrack}online\matsymb{rbrack}. \matsymb{lbrack}cit. 2018-11-19\matsymb{rbrack}. Dostupné z: <https://www.mongodb.com/>
8. BI portal - ms business intelligence portal \matsymb{lbrack}online\matsymb{rbrack}. \matsymb{lbrack}cit. 2018-11-19\matsymb{rbrack}. Dostupné z: <https://biportal.cz/>

Vedoucí bakalářské práce:

Ing. Pavel Pokorný, Ph.D.

Ústav počítačových a komunikačních systémů

Datum zadání bakalářské práce:

26. července 2019

Termín odevzdání bakalářské práce:

26. srpna 2019

Ve Zlíně dne 2. srpna 2019

L.S.

doc. Mgr. Milan Adámek, Ph.D.
děkan

Ing. Miroslav Matýsek, Ph.D.
ředitel ústavu

Jméno, příjmení: Bc. Dominika Pospíchalová

Název diplomové práce: Průzkum Business Intelligence řešení a jeho nasazení pro velká dopravní data

Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou práci nebo poskytnout licenci k jejímu využití jen připouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně, dne 26. 8. 2019

Bc. Dominika Pospíchalová, v. r.

ABSTRAKT

Táto diplomová práca má za úlohu uviesť čitateľa do problematiky veľkých dát a možností ich spracovania. Jej hlavnou náplňou je popísať postup spracovania dát pochádzajúcich z rôznych zariadení – zvoliť správne štatistické metódy a vizualizovať vybrané súvislosti v spracovávaných dátach prostredníctvom vybraného Business Intelligence nástroja. Inštalácia všetkých ďalších softvérových nástrojov k tomu potrebných bude zahrnutá do postupu práce.

Kľúčové slová: Big Data, Business Intelligence, dátová analýza, štatistika, korelácie, doprava.

ABSTRACT

This Master's thesis purpose is to enlighten the meaning of the Big Data term in context of its use in city traffic and city infrastructure. Using the right Business Intelligence tool, data collected from various devices will be processed and visualised. The process of installation of software needed to complete this task will be included. Visualization of the data will provide correlation statement about chosen hypothesis.

Keywords: Big Data, Business Intelligence, data analytics, statistics, correlations, traffic.

POĎAKOVANIE

Na tomto mieste by som rada vyjadrila veľkú vďaku vedúcemu práce, Ing. Pavlovi Pokornému Ph. D., za poskytnutú príležitosť a cenné rady pri vedení tejto diplomovej práce.

Spoločnosti Incinity s. r. o. ďakujem za zverenie témy, sprostredkované dáta a odpovede na všetky moje otázky.

Osobitná vďaka patrí mojim najbližším za ich trpezlivosť a podporu.

OBSAH

ÚVOD.....	5
I TEORETICKÁ ČÁST.....	7
1 BIG DATA	8
1.1 CHARAKTERISTIKA VEĽKÝCH DÁT	8
1.1.1 Doplnujúce vlastnosti dát.....	9
1.2 ZAUJÍMAVOSTI Z HISTÓRIE POJMU.....	10
1.3 TECHNOLOGIE PRE PRÁCU S VEĽKÝMI DÁTAMI	12
1.3.1 Výber technológií.....	12
1.3.1.1 MongoDB	12
1.3.1.2 NoSQL.....	13
1.3.1.3 JSON.....	14
1.3.1.4 BSON.....	14
1.4 VEĽKOSŤ VEĽKÝCH DÁT.....	14
2 BUSINESS INTELLIGENCE.....	16
2.1 BUSINESS INTELLIGENCE VS. BUSINESS ANALYTICS.....	16
3 PRIESKUM TRHU BUSINESS INTELLIGENCE RIEŠENÍ.....	18
3.1 VÝBER PRODUKTOV	20
3.1.1 Google Data Studio	21
3.1.2 QlikView	21
3.1.3 Sisense.....	22
3.1.4 Tableau.....	23
3.1.5 Microsoft Power BI.....	24
4 ŠTRUKTÚRA A OBSAH SPRACOVÁVANÝCH DÁT.....	26
4.1 INVIPO STATUS.....	26
4.2 INVIPO EVENT	26
4.3 ZÁVERY Z PRESKÚMANIA DÁT	27
5 VÝBER BI NÁSTROJA PRE SPRACOVANIE DÁT	28
5.1 TABLEAU.....	28
5.2 MICROSOFT POWER BI.....	28
II PRAKTICKÁ ČÁST	29
6 INŠTALÁCIA POTREBNÝCH SOFTVÉROVÝCH PRVKOV.....	30
6.1 INŠTALÁCIA MONGODB	30
6.2 SPUSTENIE A OVLÁDANIE MONGODB.....	31
6.3 PRÍPRAVA A IMPORT DÁT	34
6.3.1 Import dát prostredníctvom príkazu insertMany().....	34
6.3.2 Import dát prostredníctvom príkazu mongorestore.....	35
7 IMPORT DÁT DO ZVOLENÉHO BI NÁSTROJA.....	37

7.1	IMPORT DÁT PROSTREDNÍCTVOM APLIKÁCIE STITCH.....	37
7.2	IMPORT DÁT PROSTREDNÍCTVOM ODBC.....	37
7.3	IMPORT DÁT VO FORMÁTE JSON.....	43
7.3.1	Postup exportu dát do formátu JSON z MongoDB prostredníctvom programu Studio 3T	43
8	SPRACOVANIE DÁT VO ZVOLENOM BI SYSTÉME	48
8.1	VÝPADKOVOSŤ ZARIADENÍ	48
8.1.1	Stanovenie vstupného súboru a čistenie dát.....	48
8.1.2	Transformácia počtu výpadkov na stupnicu výpadkovosti.....	50
8.1.3	Graf výpadkovosti zariadení	51
8.2	EXTRÉMNOŠŤ POČASIA	52
8.2.1	Čistenie dát o počasí.....	52
8.2.2	Rozdelenie roku na dve ročné obdobia	52
8.2.3	Transformácia údajov o počasí na stupnicu extrémnosti počasia	53
8.2.3.1	Extrémnosť počasia v letnom období	53
8.2.3.2	Extrémnosť počasia v zimnom období	53
8.2.4	Graf extrémnosti počasia.....	53
8.3	VÝPOČET KORELÁCIE VÝPADKOVOSTI ZARIADENÍ VOČI EXTRÉMNOŠTI POČASIA	54
8.3.1	Výsledné hodnoty korelačného koeficientu	55
	ZÁVER	58
	ZOZNAM POUŽITEJ LITERATÚRY	59
	ZOZNAM POUŽITÝCH SYMBOLOV A SKRATIEK.....	62
	ZOZNAM OBRÁZKOV	63
	ZOZNAM TABULIEK	65
	ZOZNAM PRÍLOH.....	66

ÚVOD

Hľadanie zaujímavých súvislostí v dátach už dávno nie je výsadou dátových analytikov. Dátová analýza a s ňou spojené štatistické postupy sú v dnešnej dobe dôležitou súčasťou každého väčšieho podnikania v akejkol'vek sfére. Rôzne systémy, dnes už pomerne bežne dostupné i pre nekomerčné použitie, umožňujú pomerne rozsiahlo spracovávať a vizuálne interpretovať dáta i pomerne neskúseným užívateľom v prípade, že sa jedná o jednoduché súbory dát s vopred jasne definovanými výskumnými cieľmi.

Analýza veľkých dát však vyžaduje o čosi odbornejší prístup. Veľké dáta pochádzajúce z rôznych zariadení sú často veľmi neštruktúrované, majú rôzny formát, podobu a ich spracovanie je často náročné i na hardvérové vybavenie.

Každý súbor veľkých dát je potenciálne bohatým zdrojom informácií, a každá dátová analýza má svoje rámcové pravidlá. Prvým krokom by malo byť zoznámenie sa so štruktúrou a obsahom dát za účelom odpovedať na relevantné otázky o nich, napríklad: „Aké parametre sú v dátovom súbore sledované?“ „Podľa akých parametrov je možné dáta agregovať?“ „Sú parametre a ich hodnoty všeobecne zrozumiteľné?“ „Aké vzťahy je možné v dátach pozorovať?“, a podobne. Tieto otázky o dátach a ich štruktúre by malo byť možné položiť a zodpovedať pri akomkoľvek súbore dát.

Odpovede na ne by mali predstavovať bránu k formulovaniu odborných otázok a výskumných cieľov týkajúcich sa konkrétnych dát. Tieto odborné otázky môžu mať akúkoľvek formu a podobu v závislosti na obsahu dát a aspoň rámcovo stanoveného cieľa analýzy. V praxi sa pri dátovej analýze totiž môže stať a často sa stáva, že v jej priebehu vyvstanú ďalšie otázky a potencionálne ciele, ktoré pred touto analýzou ani nebolo možné formulovať.

Dôležitou súčasťou spracovania a analýzy dát je ich čistenie. Vo veľkých súboroch dát býva častý výskyt extrémnych hodnôt. Na tieto je nutné sa pri analýze zamerať a vhodným spôsobom rozhodnúť o ich vylúčení či inom opatrení vedúcom k ich eliminácii. Bývajú k tomu používané rôzne štatistické metódy a postupy.

Táto práca má za úlohu poskytnúť jeden z možných prístupov k štatistickému spracovaniu dát - od určenia skúmanej hypotézy, cez konkrétne postupy čistenia dát až po výpočet korelácií vybraných parametrov.

K samotnej dátovej analýze v rámci tejto diplomovej práce budú využité softvérové nástroje určené pre použitie v podnikovom prostredí. Na jednoduchú analýzu menších súborov dát by pravdepodobne so svojou bohatou výbavou stačil i tabuľkový editor Excel; táto práca sa však bude zameriavať na zložitejšie spracovanie predovšetkým veľkých dát s predpokladom na pomerne vysoké nároky na ich vizualizáciu.

Postup použitý v tejto diplomovej práci by mal byť natoľko verzatilný, zrozumiteľný a univerzálny, aby mohol byť považovaný za návod pre spracovanie dát a mohol byť použitý pre overenie rôznych hypotéz; mal by teda predstavovať ďalší relevantný prínos tejto diplomovej práce.

I. TEORETICKÁ ČÁST

1 BIG DATA

Big Data (z angl. veľké dáta) je pojem označujúci veľké objemy dát, ktoré môžu byť potencionálnym zdrojom hodnotných informácií. Pre potreby tejto práce nasleduje niekoľko možných definícií tohto pojmu.

Cuesta [1] definuje Big Data ako „*dáta, ktoré presahujú výpočtový výkon bežnej databáze. Analýzu veľkých dát potrebujeme tam, kde sa dáta rýchlo rozrastajú, a kde je potrebné odhaliť skryté vzorce, neznáme vzťahy a ďalšie užitočné informácie.*“

Spoločnosť Gartner [2], uznávaná výskumná a poradenská firma pre oblasť informačných technológií so sídlom v USA, rozvíja túto definíciu nasledovne: „*Big Data je pojem, ktorý označuje dáta, ktorých veľkosť (volume), rýchlosť nárastu (velocity) a rôznorodosť (variety) neumožňujú spracovávanie pomocou doposiaľ známych a overených technológií v rozumnom čase.*“

Holubová a kol. [3] skúma Big Data predovšetkým z pohľadu ich vzniku; uvádza, že sa vyvíjajú s príchodom nových technológií, služieb a ich kombinácií. Príkladom môžu byť senzorové siete, vedecké prístroje skúmajúce prírodné javy, sociálne siete alebo mobilné technológie a súvisiace aplikácie. Tieto typy technológií a aplikácií, v interakcii s ich užívateľmi, používaním kontinuálne generujú množstvo dát, ktoré je potrebné efektívne uložiť a účelne spracovať.

Definícia, ktorú poskytuje server Techtarget.com [4], popisuje využitie veľkých dát v súvislosti s umelou inteligenciou: „*Big Data ako stále sa vyvíjajúci pojem označuje veľké množstvo štruktúrovaných, semi-štruktúrovaných i neštruktúrovaných dát, z ktorých je potencionálne možné ťažiť informácie a využívať ich v projektoch zameraných na strojové učenie a v iných analytických aplikáciách.*“

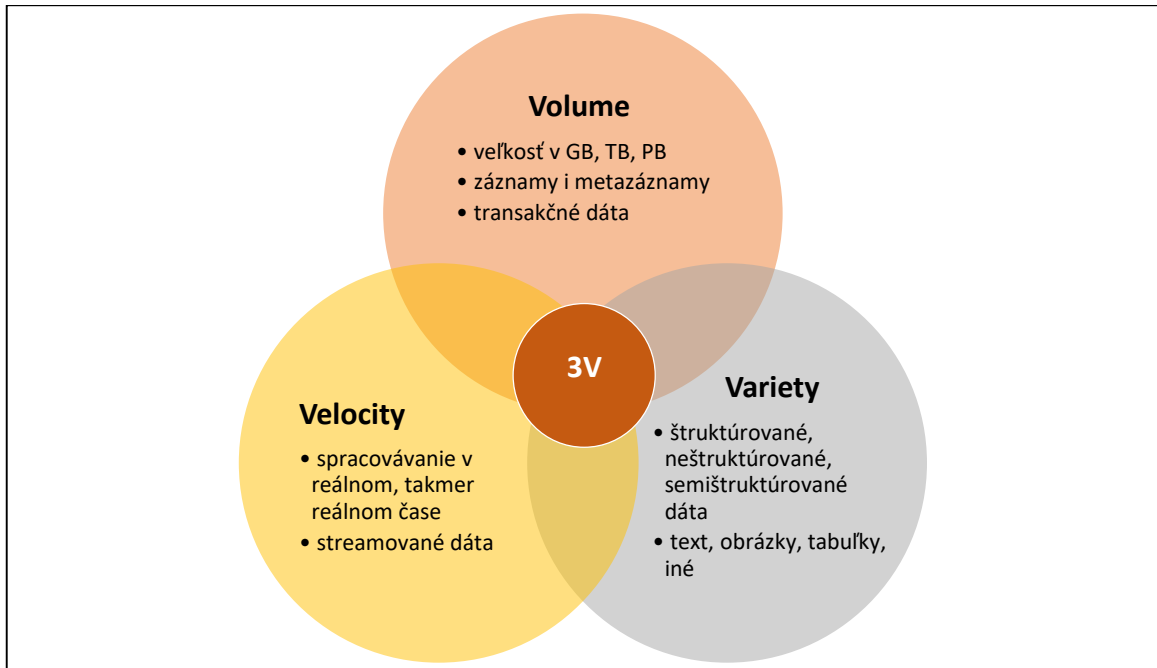
Napriek výstižnosti všetkých uvedených definícií je potrebné zdôrazniť, že formálna, presná a úplná definícia, rovnako ako pre mnoho pojmov z rôznych odborov informatiky ako pomerne mladej a neustále sa vyvíjajúcej vednej disciplíny, zatiaľ pre pojem Big Data neexistuje.

1.1 Charakteristika veľkých dát

Vlastnosti veľkých dát bývajú naprieč ich definíciami charakterizované takzvanými „3V“ podľa začiatkových písmen nasledujúcich anglických výrazov:

- Volume (objem),
- Velocity (rýchlosť),
- Variety (rôznorodosť).

Najdôležitejšie aspekty týchto vlastností priblíži obrázok (Obr. 1).



Obr. 1. Vlastnosti veľkých dát. [5]

Každou sekundu je z rôznych elektronických zariadení generované množstvo dát. Každé z týchto zariadení poskytuje iné údaje, v inom formáte a s inou štruktúrou, pričom je často potrebné tieto množstvá spracovávať v reálnom alebo takmer reálnom čase. [6]

V ďalších definíciách k týmto trom základným „V“ vlastnostiam pribúdajú aj ďalšie, najdôležitejšie z nich priblíži nasledujúca podkapitola.

1.1.1 Doplnujúce vlastnosti dát

Dôležitou vlastnosťou akýchkoľvek, nie len veľkých dát, je ich potencionálna vysoká hodnota (value) predovšetkým v podnikovej sfére pre spoločnosť, ktorá ich vlastní, dokáže spracovať a využiť pre generovanie zisku. [4] Aby však dáta prinášali hodnotu, mali by objektívne reflektovať realitu – platí tu pravidlo GIGO (z angl. „Garbage In, Garbage Out“) – chybné či neúplné vstupné dáta môžu viesť k nesprávnym manažérskym rozhodnutiam. [7]

S rôznorodosťou a početnosťou zdrojov veľkých dát môže klesať vierohodnosť (veracity). [4] V dátach, predovšetkým ak sú generované užívateľmi, je potrebné identifikovať nekvalitné, nepresné či neurčité dáta; takéto môžu výrazne znížiť hodnotu získaných informácií, a vhodným spôsobom eliminovať ich dopad. Jednou z možností je nastavenie tzv. váh, kedy sa vierohodnosť zvyšuje s narastajúcim počtom užívateľov, ktorí konkrétnu situáciu hlásia.¹

Vierohodnosť dát je taktiež možné zvýšiť zavedením metód čistenia dát. Zavedenie techník čistenia dát je preto kritickým krokom, ktorý by mal zaistiť, že výstupy analýzy dát budú relevantné a vierohodné. [4]

I čas hrá pri spracovaní veľkých dát často významnú rolu. V určitých oblastiach majú veľké dáta pomerne limitovanú dobu platnosti (validity). Konkrétne v doprave a vo veľkých mestách je často nutné rôzne situácie vyhodnocovať v reálnom čase. Dáta v takýchto prípadoch by síce stále mali spĺňať kritéria presnosti a správnosti, ale predovšetkým by mali byť dostupné v najkratšom možnom čase. Tieto dáta naopak často nebýva nutné dlho uchovávať, pretože po uplynutí určitého času či vyriešení dopravnej situácie prestávajú byť pre konkrétny účel relevantné – ich informačná hodnota je dočasná (volatility). [10]

Všetky tieto doplnujúce charakteristiky sa v závislosti na kontexte dajú aplikovať na prakticky akékoľvek, teda i „malé“ dáta a teda nemusia byť vždy pre charakteristiku veľkých dát relevantné.

1.2 Zaujímavosti z histórie pojmu

Pri snahe nájsť bod vzniku potreby termínu Big Data, čitateľ pomerne rýchlo objaví históriu siahajúcu „až“ do štyridsiatych rokov dvadsiateho storočia, k dielu knihovníka Fremonta Ridera s názvom „The Scholar and the Future of the Research Library“ [11]. V tomto diele Rider odhaduje potrebu každých šesťnásť rokov zdvojnásobiť

¹ Príkladom môže byť mobilná aplikácia Waze [8] - GPS navigačný softvér s prvkami sociálnej siete pre mobilné telefóny typu smartphone. Základnou odlišnosťou od komerčných produktov je tvorba mapových podkladov samotnými používateľmi a aktívne prispievanie k situácii na cestách, kedy systém Waze automaticky zbiera dáta o hustote premávky a zároveň umožňuje užívateľom aktívne hlásiť komplikácie v doprave. [9]

veľkosť amerických univerzitných knižníc. Na príklade knižnice univerzity Yale uvádza, že v roku 2040 bude musieť poskytnúť približne desaťtisíc kilometrov políc, aby bolo kam vtesnať dvesto miliónov výtlačkov kníh. Pre ich katalogizáciu predpovedá potrebu zamestnávať viac než šesť tisíc zamestnancov. [12]

Pravdepodobne ako jediný v tej dobe mohol pánu Riderovi ušetriť zbytočnú prácu s počítaním srbský vedec a vynálezca Nikola Tesla. Približne osemnásť rokov pred vydaním Riderovho diela, konkrétne 30. januára 1926, vyšiel v magazíne Colliers [13] rozhovor, v ktorom Tesla okrem iného predpovedá: *„Akonáhle dokážeme dokonale aplikovať bezdrôtové technológie, celý svet sa stane jedným veľkým mozgom. Budeme schopní komunikovať v reálnom čase bez ohľadu na vzdialenosť. Prostredníctvom televízie a telefónie sa budeme počuť a vidieť rovnako perfektne, akoby sme sa videli tvárou v tvár. A zariadenia, cez ktoré to všetko bude možné, budú vyzerat' prekvapivo jednoducho v porovnaní s terajším telefónom. Tieto zariadenia budú ľudia môcť nosiť vo vreckách svojich viest.“*

Slová Nikoly Tesly boli na svoju dobu síce odvážne, avšak i vtedy sa s trochou predstavivosti snád' dalo predpokladať, že v budúcnosti by mohlo byť všetko trochu inak a informácie sa nebudú schovávať v nekonečných uličkách knižníc; budú mať potenciál byť každému na dosah ruky.

Avšak späť do histórie; čitateľ, síce ohromený genialitou a predstavivosťou srbského vedca, však stále hľadájúc počiatok prvého náznaku práce s veľkými dátami, sa musí dostať v čase ešte o trochu ďalej. Ocitá sa na prelome 19. a 20. storočia a nahliada spolu s mladým inžinierom – úradníkom Hermanom Hollerithom na vtedajší problém Amerického úradu pre sčítanie ľudu, ktorý predpokladá, že spracovanie dát z aktuálneho, dosiaľ najrozsiahlšieho sčítania bude trvať minimálne 10 rokov. To by znamenalo, že informačná hodnota dát zozbieraných v tomto sčítaní bude takmer nulová, dáta totiž budú už neaktuálne; sčítanie v tej dobe prebiehalo každých 10 rokov. Než by sa teda úradníci dopočítali k výsledku, bolo by už potrebné pripravovať sčítanie nové. [14]

V roku 1881 však Hollerith vynaliezla tzv. tabulating machine, redukuje 10 rokov práce na 3 mesiace a zaujíma čestné miesto „otca moderného automatizovaného počítania“. Neskôr zakladá firmu, ktorú dnes poznáme pod názvom IBM.

1.3 Technologíe pre prácu s veľkými dátami

Podľa Cuesty [1] sú veľké dáta v súčasnosti najčastejšie ukladané a spracovávané pomocou architektúry MapReduce, ktorú definuje ako „*programovací model pre súbežné spracovanie veľkých dátových súrad s použitím distribuovaného clusteru*“. Ďalšie technológie pre prácu s veľkými dátami predstavujú dátové úložiská typu NoSQL (Not only SQL) a MPP (Massively Parallel Processing).

1.3.1 Výber technológií

Táto diplomová práca vznikla v spolupráci s firmou Incinity s. r. o. (Obr. 2), ktorá je dcérskou spoločnosťou zlínskej firmy CROSS a. s.. Táto patrí medzi lídrov v oblasti technológií a riešení pre cestnú dopravu.



Obr. 2. Logo spoločnosti Incinity, s. r. o. [15]

Spoločnosť Incinity s. r. o. [15] je mladá softwarová firma, ktorej hlavou činnosťou je vývoj, dodávanie a podpora vlastnej integračnej platformy pre chytré mestá s názvom Invipo [16].

V rámci tejto diplomovej práce budú z dôvodu zachovania kompatibility výstupov použité rovnaké technológie, ako pre chod platformy Invipo. Nasledujúce podkapitoly tieto technológie priblížia z teoretického hľadiska; v praktickej časti práce budú následne priblížené hlavné postupy práce s nimi.

1.3.1.1 MongoDB

MongoDB (Obr. 3) je databázové riešenie, ktoré implementuje technológie NoSQL. Databáze typu MongoDB ponúkajú dokumentovo-orientované úložisko, vysokú dostupnosť a flexibilitu agregácie spracovávaných dát. [1]



Obr. 3. Logo MongoDB. [17]

1.3.1.2 NoSQL

Podľa Holubovej a kol. [3] sú NoSQL databázy dôležitým pojmom v kontexte veľkých dát. Jedná sa o pomerne novú technológiu, ktorej vznik bol podmienený nedostatkami ich predchodcov – relačných SQL databáz. Tieto vyžadujú definovanie schém pred pridaním údajov, čo je takmer nezlučiteľné s moderným agilným prístupom vývoja aplikácií – každá pridaná funkcionálna môže vyžadovať zmenu štruktúry relačnej databázy, čo má za následok potrebu migrácie celej databázy do novej schémy. Pre relačné databázy je taktiež obtiažne adresovať údaje, ktoré sú úplne neštruktúrované alebo vopred neznáme. Prehľad ďalších najpodstatnejších rozdielov je spracovaný v tabuľke (Tab. 1).

Tab. 1. Relačné vs. NoSQL databázy – predpoklady o dátach. [3]

Relačné (SQL) databázy	NoSQL databázy
Integrita dát je zásadná.	Je postačujúce, ak je väčšina dát väčšinu času v poriadku.
Dátový formát je konzistentný a dobre definovaný.	Dátový formát nemusí byť známy ani konzistentný.
Predpokladá sa dlhá doba skladovania dát.	Vzhľadom k veľkému množstvu je často skladované iba „časové okno“, napr. posledný mesiac, posledný rok.
Často je potrebné dáta (obsah, nie štruktúru) aktualizovať.	Vkladané dáta typicky nie sú modifikované, dáta neustále pribúdajú v podobe nových záznamov. Nepotrebné záznamy bývajú mazané.
Nárast veľkosti dát je predvídateľný.	Nárast veľkosti dát je nepredvídateľný.
Je nutné vykonávať pravidelné zálohy dát.	Výpadky sú riešené replikáciou dát.

Dáta sú obvykle uložené na jednom serveri.	Dáta sú rozmiestnené na viacerých serveroch (clusteroch).
--	---

Databázy NoSQL sú vytvorené tak, aby umožňovali vkladanie dát a vyhľadávanie údajov bez preddefinovanej schémy. Vďaka tomu je možné vykonávať významné zmeny aplikácií v reálnom čase, bez odstávok služieb. [17]

1.3.1.3 JSON

Formát JSON (Javascript Object Notation) predstavuje podmnožinu jazyka Javascript, Je to otvorený štandard a spolu s XML je hlavným formátom používaným v moderných webových technológiách pre výmenu údajov. JSON podporuje všetky základné typy údajov: čísla, reťazce, logické hodnoty, polia i hash. JSON pozostáva z objektov a polí, pričom objekt predstavuje kolekcia párov kľúč-hodnota a pole predstavuje zoznam hodnôt. [17]

Dokumentové databázy, ako napríklad MongoDB, používajú formát JSON predovšetkým pre prenos dát. Predpokladom je vlastnosť formátu JSON – je zrozumiteľný pre človeka. [17]

1.3.1.4 BSON

Formát BSON (Binary JSON) umožňuje databáze MongoDB ukladať, transformovať a exportovať dokumenty JSON v binárne kódovanom formáte. MongoDB tak poskytuje užívateľom jednoduchosť používania a flexibilitu dokumentov formátu JSON spolu s rýchlosťou a úspornosťou binárneho formátu BSON. [17]

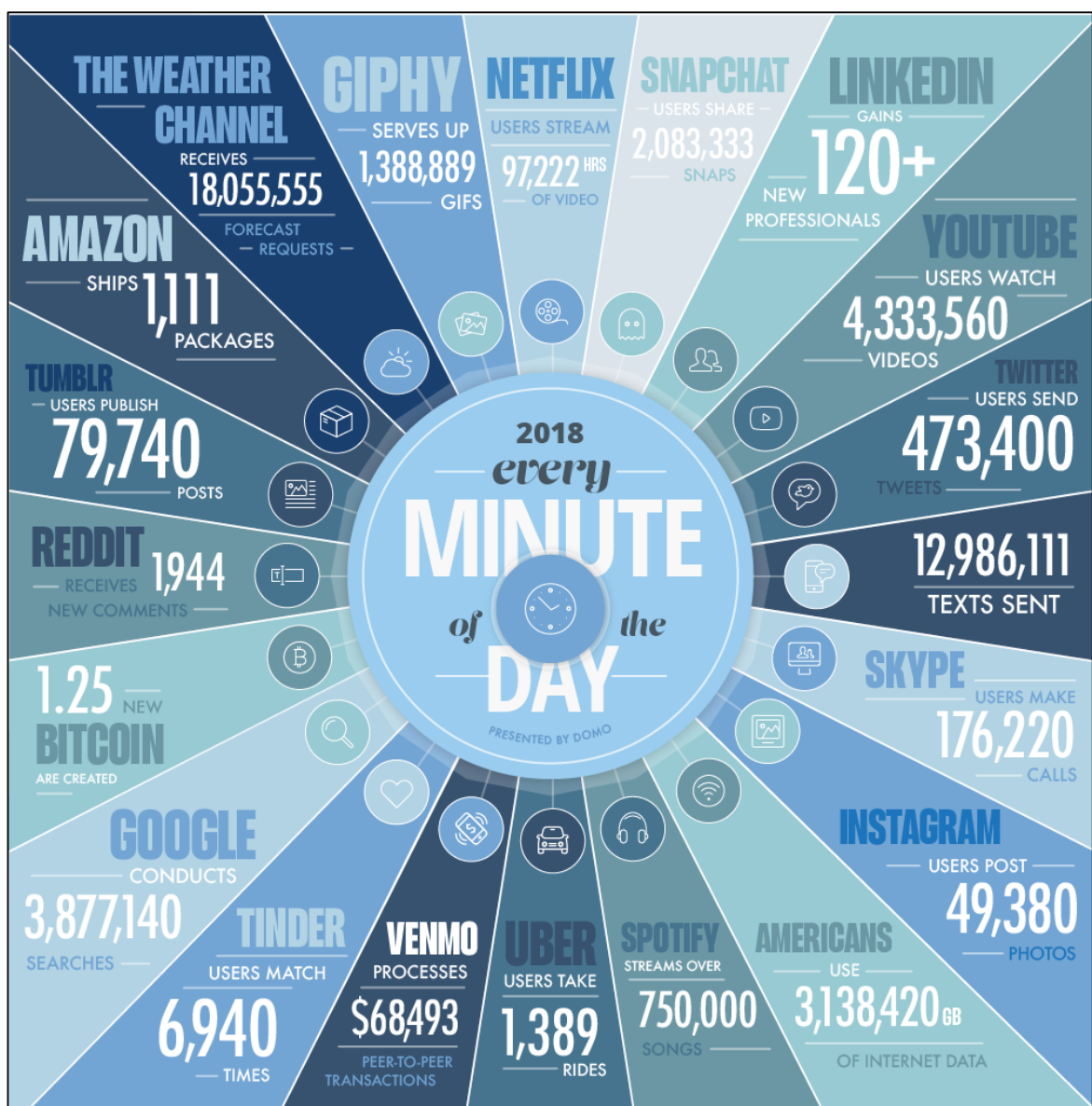
BSON je rozšírením formátu JSON – ponúka viac dátových typov, využíva indexovanie a vďaka tomu je dáta vo formáte BSON jednoduchšie spracovávať, triediť a porovnávať. [17]

1.4 Veľkosť veľkých dát

Na záver tejto kapitoly, a pre predstavu ako „veľké“ sú veľké dáta – v roku 2008 servery na celom svete podľa dostupných údajov spracovávali 9,57 zetabajtov ($1 \text{ ZB} = 10^{21} \text{ B}$) dát denne. Je to ekvivalent približne 12 gigabajtov na osobu – každého vtedajšieho obyvateľa planéty na deň. [14]

V roku 2018 to bolo $2,5 \times 10^{30}$ bajtov. [18] Približne 0,3 ZB na osobu na deň. [19]

Nasledujúca infografika (Obr. 4) poskytuje prehľad množstva dát vygenerovaných či sprostredkovaných priemerne v rámci jednej minúty roku 2018, a to najznámejšími webovými službami v ich „jednotkách“. [20] Napríklad známou americkou službou na streamovanie seriálov a filmov Netflix, bolo každú minútu odvysielaných 97 222 hodín videa, z čoho vyplýva, že táto služba beží v každý okamžik na takmer šiestich miliónoch zariadení paralelne. Popri tom bolo americkou obchodnou spoločnosťou Amazon v roku 2018 každú minútu odoslaných viac než tisíc balíčkov, do vyhľadávača Google zadaných bežmála 4 000 000 požiadaviek a na sociálnu sieť Instagram nahraných takmer 50 000 fotografií.



Obr. 4. Množstvo dát vygenerovaných za jednu minútu roku 2018 známymi internetovými službami. [20]

2 BUSINESS INTELLIGENCE

Business Intelligence (BI) je ďalší špecifický anglický pojem, pochádzajúci pôvodne z podnikového prostredia; v dnešnej dobe je úzko spojený predovšetkým s informačnými technológiami. Z angličtiny sa obvykle neprekladá, jeho význam by sa však dal vyjadriť pomocou slovného spojenia „podnikové spravodajstvo“. V tomto kontexte podľa Novotného a kol. [21] „*termín Business Intelligence označuje celý komplex činností, úloh a technológií, ktoré dnes stále častejšie tvorí bežnú súčasť riadenia podnikov a ich informačných systémov*“.

V podnikovej sfére sú tieto nástroje najčastejšie využívané predovšetkým manažérmi, a to pre zjednotenie údajov z rôznych systémov, ktoré podniky v dnešnej dobe v akejkol'vek ich forme nevyhnutne využívajú – napríklad ERP, SCM či CRM systémy, ale i špecializované systémy pre podporu personálnych či finančných oddelení. Zdrojom dát však môžu byť i externé systémy, napríklad výstupy štatistických úradov či iné verejne dostupné dáta. [22]

Z infromatického pohľadu je Business Intelligence disciplínou zloženou z niekoľkých navzájom súvisiacich aktivít, ako je vyťažovanie dát (datamining), OLAP (On-line Analytical Processing), vyhľadávanie v databáze (querying) a vytváranie zostáv (reporting) s možnou grafickou vizualizáciou dát (dashboarding). [23]

Odhliadnuc od primárneho určenia BI nástrojov v podnikovom prostredí, bude vďaka týmto možnostiam spracovania dát vybraný BI nástroj pre potreby tejto práce použitý pre vyťaženie a vizualizáciu informácií z veľkých dát z oblasti dopravy v mestách. Pri výbere konkrétneho riešenia budú posudzované predovšetkým možnosti vizualizácie dát a schopnosť spracovávať pomerne veľké objemy dát.

Pre tento účel definuje pojem Business Intelligence najvýstižnejšie server Microsoft.com [24]: „*Nástroje Business Intelligence (BI) sú typom aplikačného softvéru, ktorý zhromažďuje a spracováva veľké objemy neštruktúrovaných dát z interných aj externých systémov. Tieto nástroje tiež pomáhajú oddelovať dáta pre účely analýzy, takže môžete vytvárať zostavy, riadiace panely a vizualizácie dát.*“

2.1 Business Intelligence vs. Business Analytics

Pojem Business Intelligence (BI) je možné zameniť s pojmom Business Analytics (BA) – podnikové analytické nástroje. [25] Každý z týchto pojmov však označuje mierne odlišný

typ systémov. Obecne sú BI systémy označované ako deskriptívne – popisujúce reálny stav, a BA systémy ako prediktívne – popisujúce tendenciu možného budúceho vývoja. Podnikové analytické nástroje bývajú dimenzované na väčšie objemy dát a poskytujú komplexnejšie možnosti pre ich spracovanie; preto bývajú obsluhované skúsenými dátovými analytikmi a/alebo informatikmi – ich ovládanie totiž nie je natoľko intuitívne, ako je tomu zvyklosťou u BI nástrojov.

Najpodstatnejšie odlišnosti sú zhrnuté v tabuľke (Tab. 2).

Tab. 2. Business Intelligence vs. Business Analytics riešenia. [26]

	Business Intelligence	Business Analytics
Odpovedá na otázky	<ul style="list-style-type: none"> • Čo sa stalo? • Kedy? • Kto? • Koľko? 	<ul style="list-style-type: none"> • Prečo sa to stalo? • Stane sa to znovu? • Čo sa stane, ak zmeníme x?
Umožňuje	<ul style="list-style-type: none"> • vytváranie zostáv, • automatizované monitorovanie, • tvorbu násteniek, • OLAP, • ad hoc vyhľadávanie. 	<ul style="list-style-type: none"> • možnosti štatistickej analýzy, • vyt'azovanie dát, • prediktívne modelovanie, • analýzu veľkých dát.

Mnoho dnešných BI nástrojov však implementuje možnosti kedysi typické iba pre BA nástroje, i keď iba v jednoduchších či obmedzených formách. [26]

3 PRIESKUM TRHU BUSINESS INTELLIGENCE RIEŠENÍ

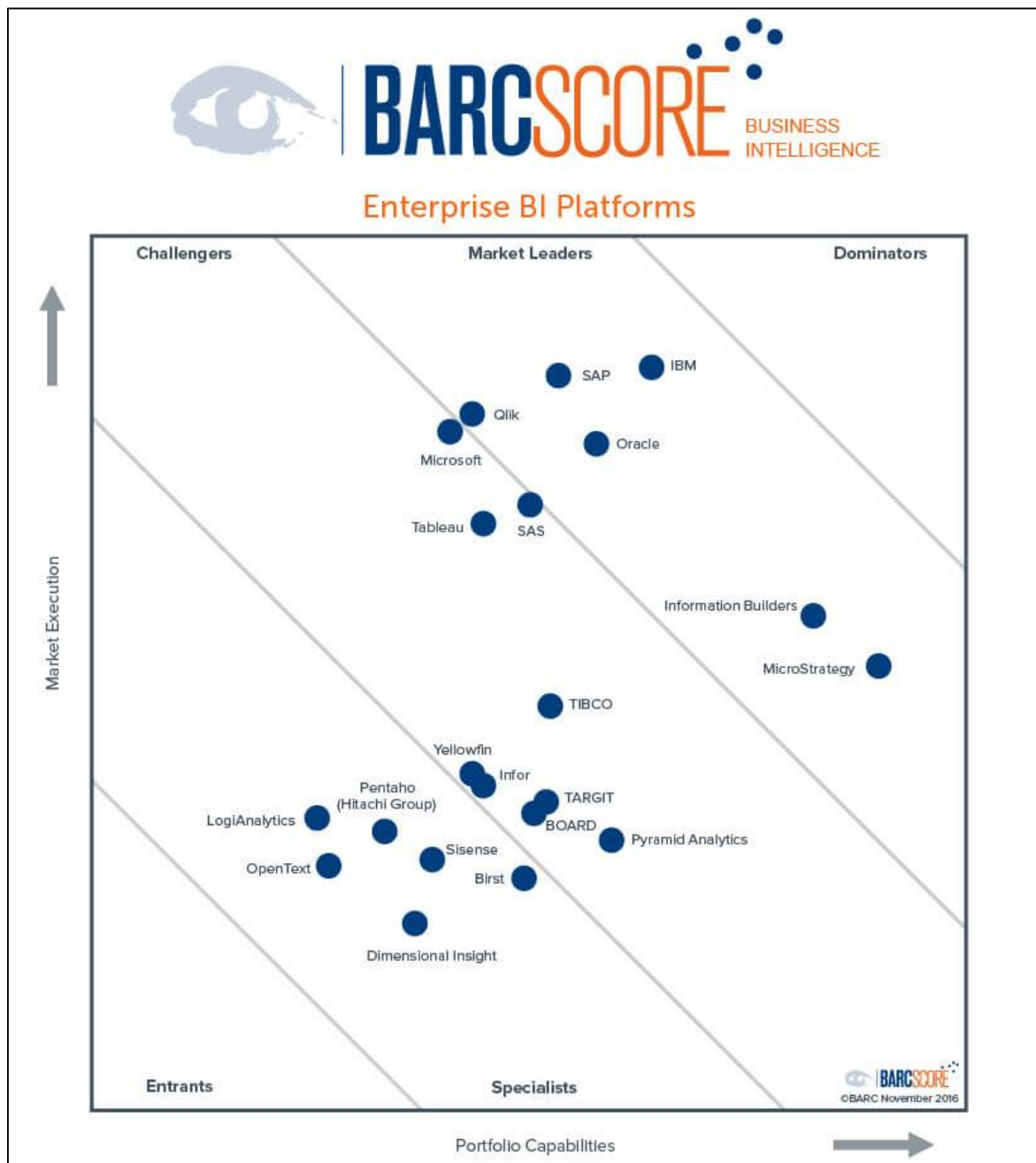
V súčasnosti existuje na trhu pomerne veľké množstvo BI systémov. Orientačný počet pre každý rok je možné získať napríklad z rôznych špecializovaných prieskumov trhu. Napríklad americká výskumná a poradenská firma Gartner a. s. [27], každoročne publikuje rozsiahly prieskum trhu rôznych technológií, v ktorých hodnotí a porovnáva poskytovateľov jednotlivých systémov v rámci jednej kategórie. Jednou z približne stovky hodnotených kategórií pre rok 2019 je i kategória „Business Intelligence and Analytics Platforms“, v ktorej bolo detailne hodnotených 21 poskytovateľov týchto nástrojov.



Obr. 5. Porovnanie poskytovateľov BI riešení spoločnosťou Gartner. [27]

Nemecký výskumný a poradenský inštitút Business Application Research Center, skrátene BARC s. r. o., uviedol k hodnoteniu prostredníctvom portálu Bi-survey.com v roku 2015 celkový počet 37 BI nástrojov. [28] Pre zaradenie do zoznamu detailne analyzovaných nástrojov bolo potrebné, aby mal nástroj minimálne 30 užívateľských recenzií. Túto podmienku splnilo i v tomto prípade 21 z existujúcich 37 nástrojov.

Výsledky porovnaní BI nástrojov pre rok 2019 od tohto inštitútu zatiaľ nie sú k dispozícii. Výsledky pre roky 2016, 2017 a 2018 sú spoplatnené pomerne vysokou čiastkou, preto ich nebolo možné v rámci tejto práce zohľadniť. Výsledky z roku 2015 teda budú slúžiť ako ukážka pokroku jednotlivých riešení na trhu za posledné 4 roky.



Obr. 6. Porovnanie poskytovateľov BI riešení inštitútom BARC. [28]

3.1 Výber produktov

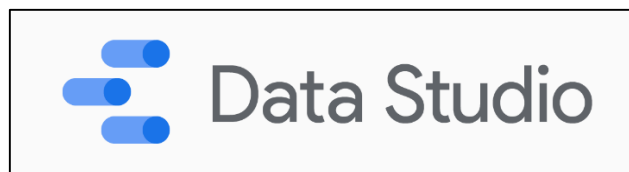
Výber konkrétnych riešení prebiehal na základe rozdelenia a hodnotenia spoločnosťami Gartner a BARC, avšak i mimo tieto rebríčky bolo nájdené jedno potencionálne vhodné riešenie.

Náplňou tejto kapitoly bude predstavenie hlavných funkcií vybraných riešení a výpočet výhod a nevýhod jednotlivých riešení v kontexte tejto práce. Na základe výsledkov týchto

porovnaní bude nasledovať výber jednej alebo viacerých platforiem, ktoré budú v praktickej časti tejto práce použité pri spracovaní zadaných úloh z oblasti veľkých dát.

3.1.1 Google Data Studio

Google Data Studio [29] (Obr. 7) je spomedzi vybraných riešení novinkou, preto sa nevyskytuje v štatistikách spoločnosti BARC ani Gartner. K vyskúšanju je zatiaľ dostupný iba v Beta verzii prostredníctvom webového rozhrania.



Obr. 7. Logo Google Data Studio. [29]

Nespornou výhodou je bezproblémové prihlásenie prostredníctvom Google účtu a napojenie na ostatné služby Google, a to v rámci balíčku s názvom Google Analytics Solution, ktorého je súčasťou.

Google Data Studio je silne zamerané na tímovú prácu viacerých užívateľov simultánne – je v ňom veľmi jednoduché sledovať zmeny ostatných užívateľov a vytvárať tímy.

Vyniká jednoduchým a moderným dizajnom a intuitívnym ovládaním. Nevýhodou však je, že poskytuje veľmi málo možností tvorby vlastných vizuálov – je v ňom prakticky možná iba úprava preddefinovaných šablón.

Najväčšou a kritickou nevýhodou sú však jeho obmedzené možnosti prípravy a modelovania dát. Inak povedané, dáta, ktoré majú byť do Google Data Studia importované, musia byť „čisté“, naformátované a pripravené na vizualizáciu. [30]

V súčasnosti je jeho distribúcia zdarma.

3.1.2 QlikView

Qlik [31] je jedným zo štyroch poskytovateľov BI aplikácií zo skupiny lídrov hodnotenia spoločnosti Gartner, jeho celkové skóre je však zo všetkých lídrov v tomto kvadrante najnižšie. V hodnotení spoločnosti BARC sa však umiestnil v kategórii lídrov trhu už v roku 2015, jedná sa teda o silného a stabilného poskytovateľa BI riešení.



Obr. 8. Logo QlikView. [31]

QlikView (Obr. 8) je jedno z najflexibilnejších desktopových riešení s množstvom možností úprav grafov a násteniiek. Vďaka technológii spracovania dát in-memory umožňuje vytvárať robustné, vizuálne interaktívne aplikácie, komplexné dátové modely a vykonávať komplexné výpočty, pričom je podporovaných viacero zdrojov vstupných dát. Je teda vhodný pre spracovávanie rozsiahlych analýz neštruktúrovaných dát.

Vďaka funkcionalite s názvom Cognitive Engine dokáže rozoznať známe vzťahy v dátach a umožňuje tak automatické generovanie grafov, prípadne poskytne nápovedu pri určovaní dôležitých pohľadov na dáta. Aj napriek týmto výhodám však nie je označovaný ako samoobslužný a jeho nasadenie by vždy malo byť konzultované s IT oddelením firmy. [27]

Jeho kritickou nevýhodou v kontexte tejto práce je komplikovaná cenová a produktová politika. K samotnému QlikView je potrebný doplnok Qlik Data Catalyst, ktorý umožňuje pokročilú prípravu dát. Pre možnosti reportingu je potom potrebná inštalácia ďalšej platformy s názvom NPrinting.

3.1.3 Sisense

Sisense [32] (Obr. 9) je jediný zástupca nižšej skupiny vizionárov z matice hodnotenia spoločnosti Gartner. V matici spoločnosti BARC z roku 2015 sa umiestnil taktiež najnižšie spomedzi systémov testovaných v rámci tejto práce; v kategórii špecialistov.



Obr. 9. Logo Sisense. [32]

Dáta spracúva technológiou in-chip, ktorá zabezpečuje rýchlosť spracovania vyhľadávani na veľkých dátach. Zaujímavosťou je súčasť desktopovej aplikácie s názvom Sisense Hunch – funkcionality umelej inteligencie, ktorá umožňuje zredukovať terabajty vstupných dát na neurónovú sieť o veľkosti niekoľkých megabajtov. Na tejto môže byť následne spustené vyhľadávanie, ktoré by malo byť rýchlejšie a efektívnejšie. [27]

Nevýhodou je netransparentná cenová politika na vyžiadanie, slabá užívateľská komunita a pomerne neintuitívne ovládanie.

3.1.4 Tableau

Tableau [33] (Obr. 10) – ďalší z lídrov podľa spoločnosti Gartner, avšak s obrovským náskokom pred nástrojom Qlik. V roku 2015 bola jeho pozícia podľa štatistiky serveru BARC o niečo horšia, jeho pokrok za dané obdobie je teda zrejmý a vypovedá o snažení jeho poskytovateľa.



Obr. 10. Logo Tableau. [33]

Tableau poskytuje široké spektrum vstupných zdrojov dát. Poskytuje dva druhy pripojenia dát. Prvá z možností sa nazýva Live Connect – dáta neukladá do pamäte, iba sa k nim pripája a v dátach vyhľadáva on-line. Druhá možnosť s názvom Hyper pracuje na princípe in-memory, vyžaduje načítanie dát a využíva sa pre pomalé alebo „zaneprázdnené“ databáze, prípadne pre prácu offline. [28]

Možnosti násteniek sú flexibilné a poskytujú dostatočné množstvo možností. I keď sú dobre optimalizované pre digitálne spracovanie, neprijemnosťou by mohli byť veľmi obmedzené možnosti ich exportu pre tlač.

Tableau je považovaný za jeden z najjednoduchších BI systémov na ovládanie i pre netechnicky zameraných užívateľov, k príjemnému užívateľskému zážitku prispievajú kontextové nápovedy, podpora širokej komunity užívateľov, množstvo tutoriálov, blogov a fór, na ktorých je pomerne jednoduché nájsť návod či odpoveď na konkrétnu otázku. [28]

Nevýhoda Tableau spočíva v nutnosti prípravy dát pred ich samotným načítaním. Po tomto kroku sú už úpravy štruktúry importovaných dát veľmi obmedzené. Pre načítanie neštruktúrovaných dát, prípadne dát s neznámou štruktúrou je preto pomerne nevhodný – do modelu dát je totiž potrebné načítať i časti záznamov, ktoré nemusia byť pre analýzu dôležité. V súvislosti s veľkými dátami toto môže predstavovať výrazné spomalenie práce. Pre čistenie dát je nutné inštalovať balíček s názvom Tableau Prep.

Skúšobná verzia je k dispozícii iba na 14 dní a má veľmi obmedzené možnosti importu dát.

3.1.5 Microsoft Power BI

Power BI [34] (Obr. 11) je produkt od Microsoftu; ako všetky produkty z balíčku Microsoft teda disponuje známym prostredím a princípmi práce. Umožňuje jednoduchú registráciu a prihlásenie firemným či školským účtom, prepojenosť s ostatnými nástrojmi od Microsoftu je taktiež samozrejmosťou. Neprekvapí tak výsledok prieskumu trhu spoločnosti Gartner – Microsoft týmto nástrojom pre rok 2019 vyhráva najlepšie umiestnenie v sekcii lídrov a teda i celej matici trhu BI systémov. V roku 2015 len tesne zaostával za nástrojom Qlik.



Obr. 11. Logo Microsoft Power BI. [34]

Balíček Power BI pozostáva z odľahčenej web verzie ktorá sa zameriava predovšetkým na tvorbu a zdieľanie vizuálov, pričom ich možnosti nie sú nijako obmedzené oproti desktopovej aplikácii. Táto však rozširuje tieto možnosti o Power Query – nástroj pre

čistenie a normalizáciu dát, a má omnoho väčšie portfólio formátov vstupných dát. Súčasťou balíčku je i mobilná aplikácia pre Windows, iOS i Android. [35]

Poskytuje bezplatnú zjednodušenú verziu bez časového obmedzenia, pre webovú aplikáciu je navyše možnosť vyskúšať verziu Pro až na 60 dní.

Veľkou výhodou je možnosť spracúvania streamovaných dát. V bezplatnej verzii je limit nastavený na 10 000 záznamov za hodinu, v platenej verzii je API programu schopné spracovávať maximálne 1 milión záznamov za hodinu. [30]

Ďalšou výhodou je silná komunita užívateľov, množstvo tutoriálov a blog, ktorý je kvôli vágnosti chybových hlások veľmi užitočný. [35]

Príjemnou zmenou oproti ostatným poskytovateľom BI systémov je krištáľovo jasná cenová politika Microsoftu, ktorý okrem bezplatnej verzie ponúka Pro a Premium verzie, pričom na webových stránkach uvádza ich cenu a prehľadne porovnáva rozdiely v ich funkcionalitách. [30]

4 ŠTRUKTÚRA A OBSAH SPRACOVÁVANÝCH DÁT

Invipo je produktom firmy Incinity, s.r.o, predstavenej v kapitole 1.3.1. Nasledujúci popis platformy Invipo pochádza z produktových stránok vývojára [16]:

„Invipo je pružná a otvorená platforma pre integráciu technológií, systémov a služieb v mestách a na cestách. Invipo prepája dáta z rôznych technológií a systémov v jeden celok, ponúka tak prehľadné sledovanie výstupov a efektívne riadenie smart projektov“.



Obr. 12. Logo platformy Invipo. [16]

4.1 Invipo Status

Invipo Status je stavový objekt, ktorý oznamuje technický stav zariadenia v systéme Invipo. Každý Invipo Status objekt obsahuje unikátny identifikátor zariadenia, ktorého sa stav týka, časovú značku zmeny statusu objektu, variabilný tok dát, ktorý bližšie popisuje stav objektu, a najdôležitejšiu informáciu o aktuálnom stave objektu. Tento stav môže nadobúdať nasledujúcich hodnôt: Ok, Warn, Error, Disconnected.

Príklad záznamov z kolekcie Invipo Status je k nahliadnutiu v prílohe (P I).

4.2 Invipo Event

Invipo Event je udalosť, ktorá je zaregistrovaná v rámci Invipo systému a oznamuje, čo sa na danom zariadení stalo alebo zmenilo. Každý Invipo Event obsahuje unikátny identifikátor zariadenia, ktorého sa stav týka, typ vzniknutej udalosti, časovú značku udalosti na objekte, variabilný blok dát, ktorý bližšie popisuje udalosť, a informáciu o závažnosti udalosti, ktorá môže nadobúdať hodnôt: Info, Warn, Error, Fatal.

Príklad záznamov z kolekcie Invipo Status je k nahliadnutiu v prílohe (P II).

4.3 Závěry z preskúmania dát

Pre potreby tejto práce bola s jej zadávateľom, firmou Incinity, s. r. o. na základe možností dostupných dát popísaných v predchádzajúcich kapitolách, k overeniu stanovená hypotéza z oblasti veľkých dopravných dát v nasledujúcom znení: „Pôsobenie faktoru počasia priamo ovplyvňuje počet výpadkov exteriérových elektronických zariadení monitorujúcich dopravu v meste Zlín.“

MongoDB kolekcia Invipo Events vo formáte JSON mala veľkosť 81,2 GB a bola tvorená viac než 76 miliónmi záznamov. Pre overenie stanovenej hypotézy nebolo nutné s týmto súborom dát pracovať; pravdepodobne by to ani s dostupným výpočtovým vybavením pre túto diplomovú prácu nebolo možné. V prípade potreby by tento súbor bolo možné spracovávať po častiach, čo by však bolo mimoriadne časovo náročné.

MongoDB kolekcia Invipo Status mala vo formáte JSON veľkosť 274,2 MB, čo predstavovalo viac než 852 tisíc záznamov. V týchto dátach bolo možné získať informácie o výpadkovosti zariadení – konkrétne boli relevantné záznamy s hodnotou „Disconnected“ kľúča „status“. Táto kolekcia bola preto zvolená pre ďalší postup.

5 VÝBER BI NÁSTROJA PRE SPRACOVANIE DÁT

Na základe zoznámenia s vstupnými dátami a predstavenia vybraných, potencionálne vhodných BI nástrojov v predchádzajúcich kapitolách, boli pre svoje relevantne kladné charakteristiky vybrané k užšiemu hodnoteniu dve z predstavených riešení. Práca s nimi bola otestovaná niekoľkými testovacími súborami.

5.1 Tableau

Pri prvom spustení Tableau a otestovaní malým testovacím súborom s vopred naformátovanými a „čistými“ dátami o počasí vyzerala práca s týmto softvérom nádejne. Pri pokuse importu dát z platformy Invipo sa Tableau „tváril“, že si poradí so súborom o veľkosti niekoľkých desiatok GB, niekoľko desiatok minút sa ho snažil importovať a nakoniec oznámil, že veľkosť súboru prekračuje limit, ktorý nebol špecifikovaný. Maximálna veľkosť jedného úspešne nahraného súboru bola približne 100 MB. Kvôli týmto obmedzeniam bol pre potreby tejto práce Tableau označený za nevyhovujúci.

5.2 Microsoft Power BI

Práca s Power BI sa už na prvý pohľad javí intuitívne. Otestovanie tohto softvéru s prvým testovacím súborom s dátami o počasí prebehlo bez problémov, pre inšpiráciu prvých krokov bolo možné vybrať z mnoho návodov a ilustračných videí poskytnutých internetovým vyhľadávačom. Import dát z platformy Invipo síce neprebehol na prvý pokus úspešne (podrobnosti v kapitole 7.2), avšak sila Power BI sa ukázala i v rôznorodosti možností importu dát, ktorý nakoniec iným spôsobom prebehol úspešne. V bezplatnej verzii je možné pracovať až s 1 GB dát na jeden dátový zdroj. Pre potreby tejto práce je tento dátový limit akceptovateľný.

Snáď jedinou relevantnou, avšak nie kritickou nevýhodou bolo relatívne pomalé načítanie väčších datasetov, a to nie len pri inicializácii, ale i pri úprave a čistení dát a aplikovaní zmien prostredníctvom Power Query. Naopak neoceniteľnou výhodou Power Query je za cenu trochu vyššej trpezlivosti možnosť v každom okamihu sa vrátiť do ktoréhokolvek kroku vykonaných úprav späť, pretože program si všetky tieto kroky ukladá a dokáže ich zopakovať pri každej aktualizácii reportu.

Microsoft so svojím Power BI riešením vďaka veľkorysým možnostiam neplatenej verzie a intuitívnosti prostredia i napriek drobným nedostatkom toto porovnanie vyhráva.

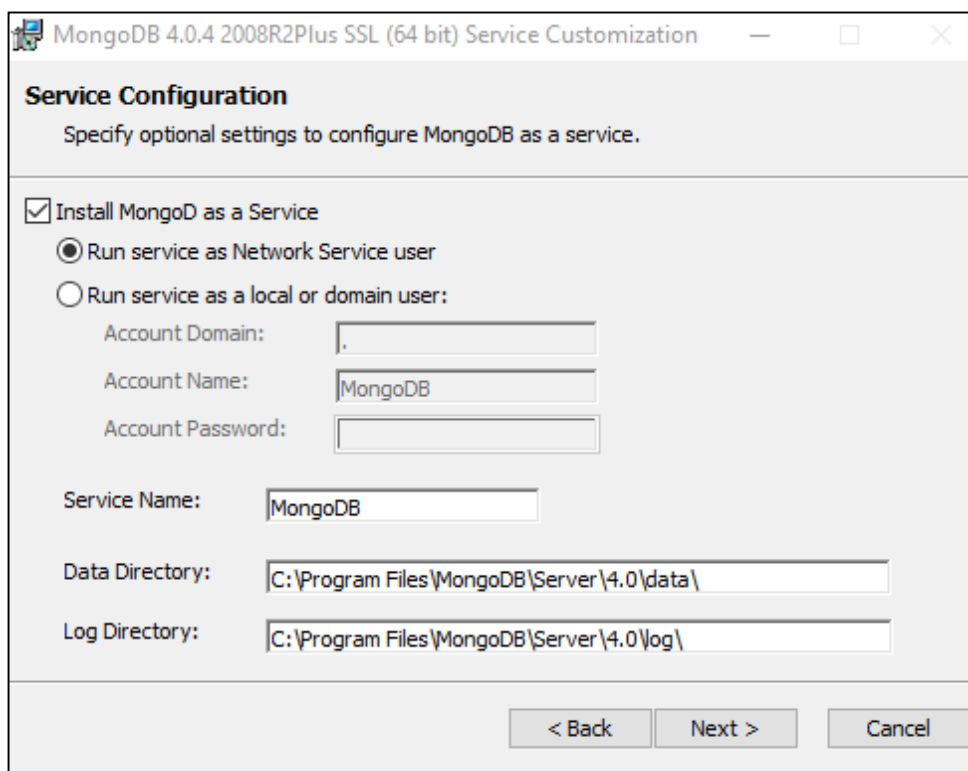
II. PRAKTICKÁ ČÁST

6 INŠTALÁCIA POTREBNÝCH SOFTVÉROVÝCH PRVKOV

Dáta pre vypracovanie tejto diplomovej práce boli spoločnosťou Incinity, s. r. o. poskytnuté vo formáte BSON. Pre prácu s dátami v tomto formáte bolo doporučené nainštalovať desktopovú verziu MongoDB. Inštaláciu popisujú nasledujúce kapitoly.

6.1 Inštalácia MongoDB

Na stránkach mongodb.com [17] je možné stiahnuť inštalačný súbor aktuálnej verzie MongoDB. Inštalácia je pomerne štandardná, je potrebné nainštalovať MongoDB ako službu (Obr. 13) a v ďalšom kroku zvoliť kompletnú inštaláciu.



Obr. 13. Inštalácia MongoDB. Zdroj: vlastná tvorba.

Po dokončení inštalácie sa vytvorí priečinok s názvom MongoDB, ktorého obsah by mal vyzeráť približne ako na obrázku (Obr. 14).

Název	Datum změny	Typ	Velikost
bsondump.exe	06.11.2018 19:49	Aplikace	8 608 kB
InstallCompass.ps1	06.11.2018 20:15	Windows PowerSh...	2 kB
libeay32.dll	03.04.2018 19:58	Rozšíření aplikace	2 405 kB
mongo.exe	06.11.2018 20:11	Aplikace	17 967 kB
mongod.cfg	21.11.2018 18:36	Soubor CFG	1 kB
mongod.exe	06.11.2018 20:17	Aplikace	31 722 kB
mongod.pdb	06.11.2018 20:17	VisualStudio.pdb.1...	348 284 kB
mongodump.exe	06.11.2018 19:53	Aplikace	10 674 kB
mongoexport.exe	06.11.2018 19:51	Aplikace	8 862 kB
mongofiles.exe	06.11.2018 19:51	Aplikace	8 793 kB
mongoimport.exe	06.11.2018 19:52	Aplikace	8 961 kB
mongorestore.exe	06.11.2018 19:52	Aplikace	11 760 kB
mongos.exe	06.11.2018 20:05	Aplikace	16 423 kB
mongos.pdb	06.11.2018 20:05	VisualStudio.pdb.1...	181 796 kB
mongostat.exe	06.11.2018 19:50	Aplikace	8 950 kB
mongotop.exe	06.11.2018 19:54	Aplikace	8 758 kB
ssleay32.dll	03.04.2018 19:58	Rozšíření aplikace	350 kB

Obr. 14. Obsah priečinka MongoDB. Zdroj: vlastná tvorba.

6.2 Spustenie a ovládanie MongoDB

MongoDB sa v prostredí Windows ovláda prostredníctvom príkazového riadku.

Najprv je teda potrebné spustiť príkazový riadok (cmd.exe), a v ňom aplikáciu mongod.exe (Obr. 15). Ešte predtým ale bude pravdepodobne nutné vytvoriť priečinok s cestou C:\data\db\. Tento priečinok je defaultne nastavený pre ukladanie dát z databáz vytvorených v MongoDB.

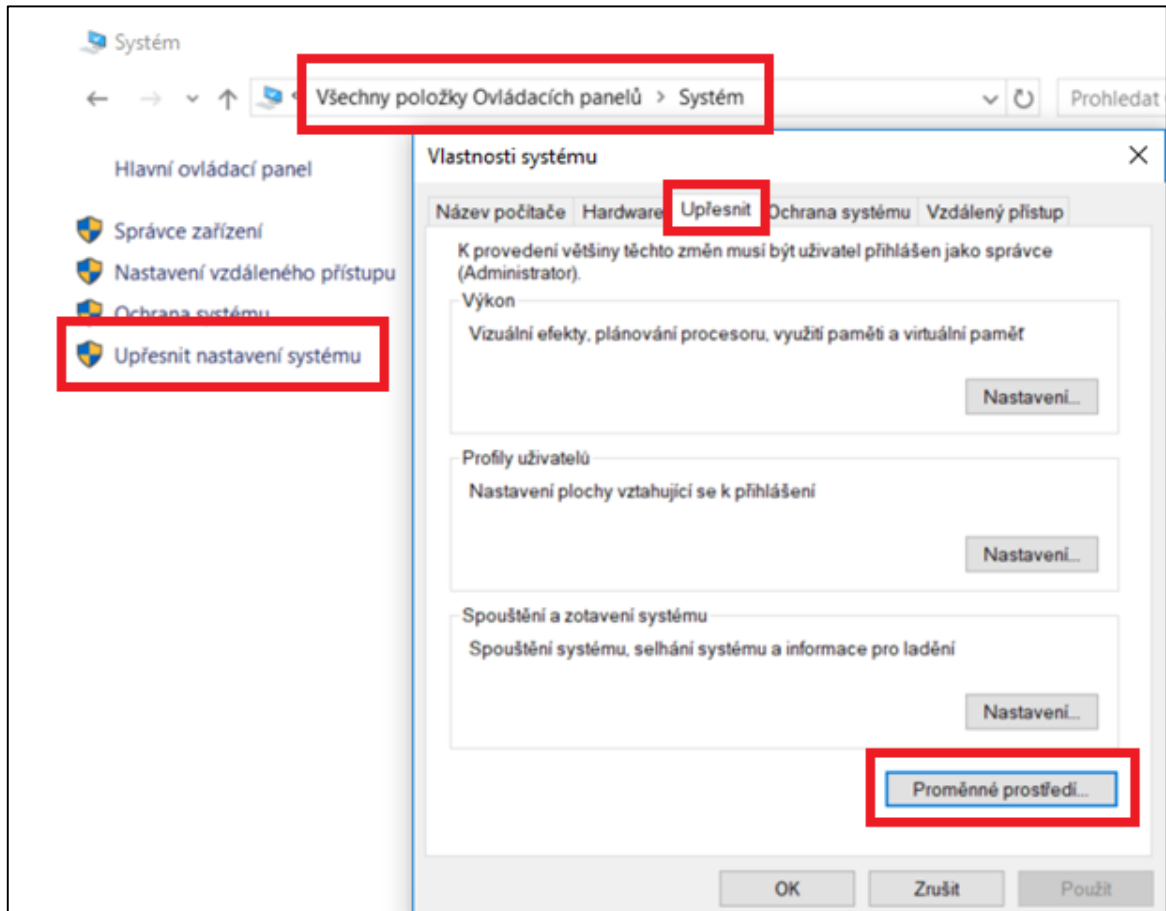
```

Microsoft Windows [Version 10.0.17134.765]
(c) 2018 Microsoft Corporation. Všechna práva vyhrazena.

C:\Users\Uzivatel>cd C:\Program Files\MongoDB\Server\4.0\bin
C:\Program Files\MongoDB\Server\4.0\bin>mongod
  
```

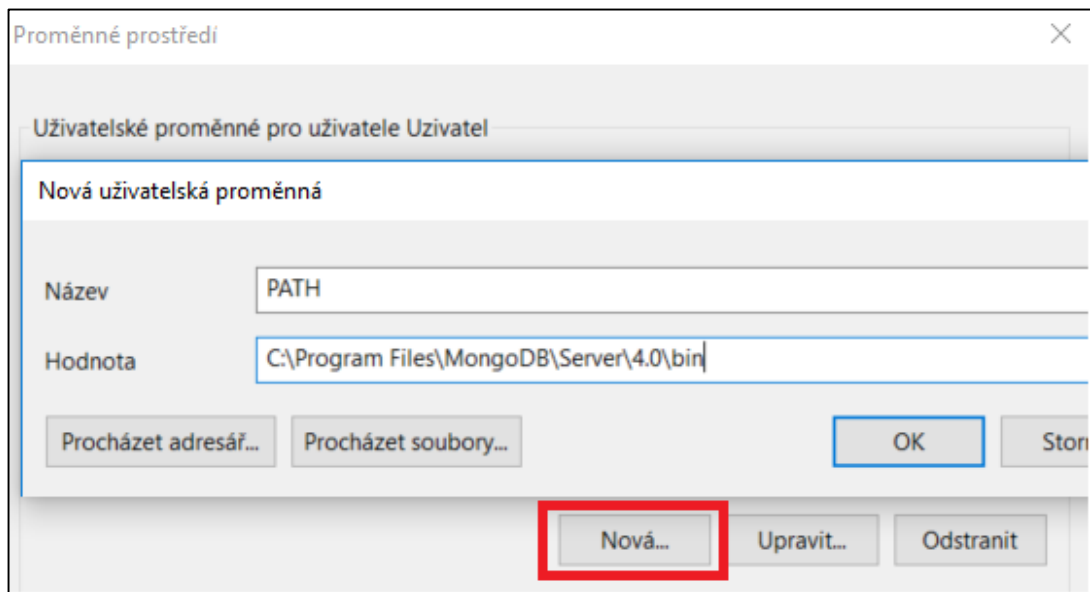
Obr. 15. Navigácia do priečinka MongoDB a spustenie mongod.exe. Zdroj: vlastná tvorba.

Následně je pro zjednodušení spouštění vhodné vytvořit systémovou skratku (uživatelskou proměnnou prostředí Windows) prostřednictvím Ovládacího panelu. V položce Systém, po kliknutí na možnost Upřesnit nastavení systému na karte Upřesnit je potřebné zvolit ponuku Proměnné prostředí (Obr. 16).



Obr. 16. Cesta k nastavení nové uživatelské proměnné. Zdroj: vlastní tvorba.

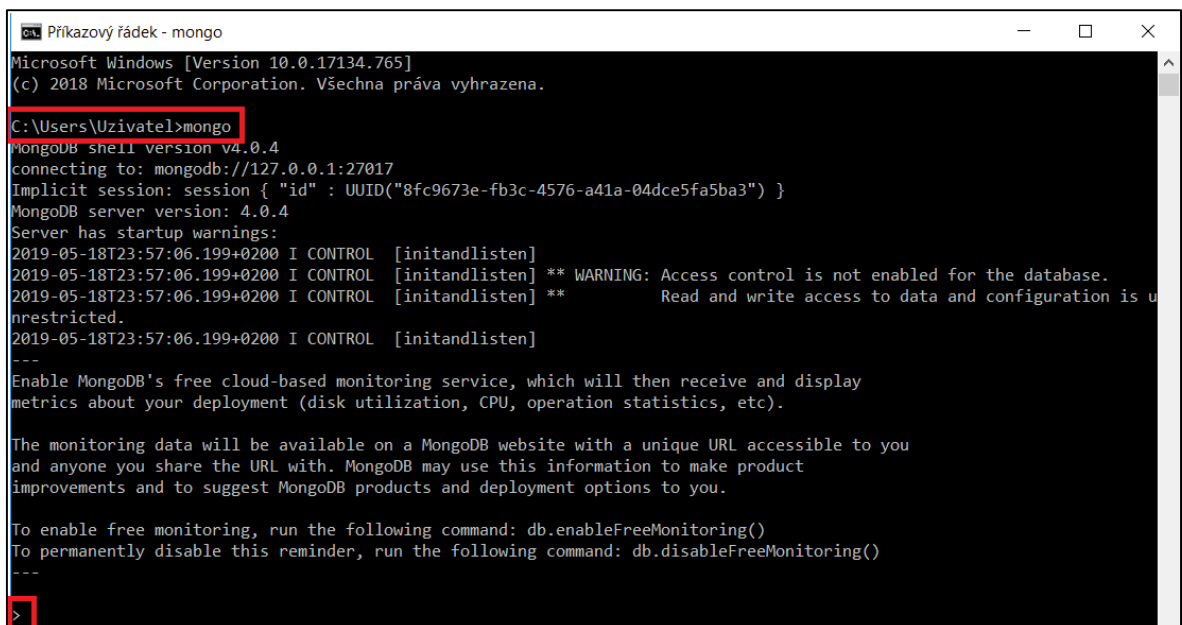
Následuje založení nové proměnné s názvem PATH. Hodnotu bude představovat cesta k souboru mongod.exe (Obr. 17).



Obr. 17. Založení novej užívateľskej premennej. Zdroj: vlastná tvorba.

Po uložení je možné spustiť ktorýkoľvek súbor z priečinku v zadanej ceste prostredníctvom príkazového riadku priamo, bez nutnosti túto cestu uvádzať.

Spustením programu mongod.exe bolo spustené prostredie MongoDB. Je to však zatiaľ iba neaktívne pripojenie. Pre spustenie prostredia, v ktorom je možné vykonávať v tejto databáze zmeny, je potrebné spustiť program mongo.exe.



```
Microsoft Windows [Version 10.0.17134.765]
(c) 2018 Microsoft Corporation. Všechna práva vyhrazena.
C:\Users\Uzivatel>mongo
MongoDB shell version v4.0.4
connecting to: mongodb://127.0.0.1:27017
Implicit session: session { "id" : UUID("8fc9673e-fb3c-4576-a41a-04dce5fa5ba3") }
MongoDB server version: 4.0.4
Server has startup warnings:
2019-05-18T23:57:06.199+0200 I CONTROL [initandlisten]
2019-05-18T23:57:06.199+0200 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-05-18T23:57:06.199+0200 I CONTROL [initandlisten] **           Read and write access to data and configuration is u
nrestricted.
2019-05-18T23:57:06.199+0200 I CONTROL [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---
```

Obr. 18. Spustenie programu mongo.exe. Zdroj: vlastná tvorba.

Možnosť úprav databázy je naznačená znakom vyznačeným na obrázku (Obr. 18).

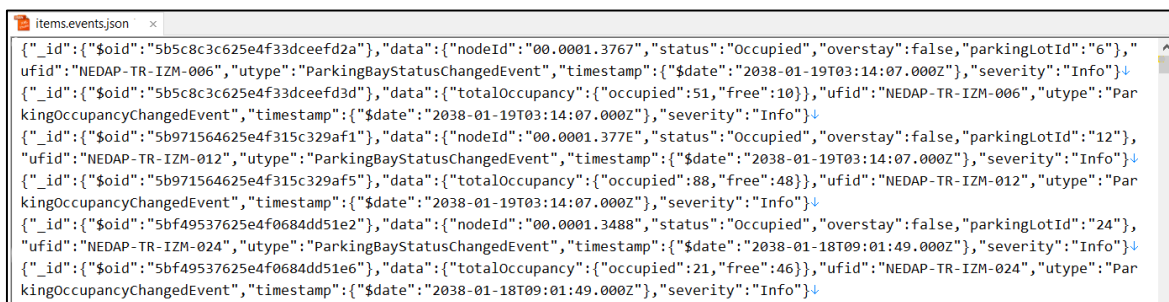
6.3 Příprava a import dát

Existuje viacero možností, ako dáta importovať do lokálnej databázy MongoDB. Nasledujúce podkapitoly poskytujú návod pre import dát nasledujúcimi spôsobmi:

- vo formáte JSON prostredníctvom príkazu `insertMany()`,
- vo formáte BSON prostredníctvom príkazu `mongorestore`.

6.3.1 Import dát prostredníctvom príkazu `insertMany()`

Práca s dátami často vyžaduje ich prípravu. „Surové“ dáta vygenerované z logu systému Invipo vo formáte JSON môžu mať napríklad nasledujúcu podobu (Obr. 19). Prehľadná stromová štruktúra dát je vložená do príloh (P I, P II).



```
{ "_id": {"$oid": "5b5c8c3c625e4f33dceefd2a"}, "data": {"nodeId": "00.0001.3767", "status": "Occupied", "overstay": false, "parkingLotId": "6"}, "ufid": "NEDAP-TR-IZM-006", "utype": "ParkingBayStatusChangedEvent", "timestamp": {"$date": "2038-01-19T03:14:07.000Z"}, "severity": "Info"}
{ "_id": {"$oid": "5b5c8c3c625e4f33dceefd3d"}, "data": {"totalOccupancy": {"occupied": 51, "free": 10}}, "ufid": "NEDAP-TR-IZM-006", "utype": "ParkingOccupancyChangedEvent", "timestamp": {"$date": "2038-01-19T03:14:07.000Z"}, "severity": "Info"}
{ "_id": {"$oid": "5b971564625e4f315c329af1"}, "data": {"nodeId": "00.0001.377E", "status": "Occupied", "overstay": false, "parkingLotId": "12"}, "ufid": "NEDAP-TR-IZM-012", "utype": "ParkingBayStatusChangedEvent", "timestamp": {"$date": "2038-01-19T03:14:07.000Z"}, "severity": "Info"}
{ "_id": {"$oid": "5b971564625e4f315c329af5"}, "data": {"totalOccupancy": {"occupied": 88, "free": 48}}, "ufid": "NEDAP-TR-IZM-012", "utype": "ParkingOccupancyChangedEvent", "timestamp": {"$date": "2038-01-19T03:14:07.000Z"}, "severity": "Info"}
{ "_id": {"$oid": "5bf49537625e4f0684dd51e2"}, "data": {"nodeId": "00.0001.3488", "status": "Occupied", "overstay": false, "parkingLotId": "24"}, "ufid": "NEDAP-TR-IZM-024", "utype": "ParkingBayStatusChangedEvent", "timestamp": {"$date": "2038-01-18T09:01:49.000Z"}, "severity": "Info"}
{ "_id": {"$oid": "5bf49537625e4f0684dd51e6"}, "data": {"totalOccupancy": {"occupied": 21, "free": 46}}, "ufid": "NEDAP-TR-IZM-024", "utype": "ParkingOccupancyChangedEvent", "timestamp": {"$date": "2038-01-18T09:01:49.000Z"}, "severity": "Info"}
```

Obr. 19. Ukážka dát vo formáte JSON. Zdroj: vlastná tvorba.

Pri pokuse o import vzorky dát do databázy MongoDB v uvedenom formáte dôjde k syntaktickej chybe. Pri exporte totiž MongoDB nahrádza natívne funkcie ako `ObjectId` a `ISODate` znakom doláru, ktorý pri importe nie je podporovaný, pre potreby ďalšieho spracovania je ho však postačujúce zo záznamov odstrániť.

Následne je kvôli dodržaniu syntaxe importu nevyhnutné pridať na koniec každého záznamu oddeľovač (čiarku) a všetky záznamy vložiť do tzv. poľa hodnôt (array), ktoré je značené hranatými zátvorkami. Takto pripravené pole je potrebné vložiť ako argument funkcie `insertMany()`. Import vzorky o počte 10 000 riadkov záznamov o veľkosti takmer 3 MB týmto spôsobom trvá približne 30 minút, v závislosti od výkonu konkrétneho zariadenia. Ukážka konečnej podoby kódu je na obrázku (Obr. 20).

```

items.events - funkčný insert,json * ×
db.testdata.insertMany([
{"_id":{"oid":"5b5c8c3c625e4f33dceefd2a"},"data":{"nodeId":"00.0001.3767","status":"Occupied","overstay":false,"parkingLotId":"6"},"u
fuid":"NEDAP-TR-IZM-006","utype":"ParkingBayStatusChangedEvent","timestamp":{"date":"2038-01-19T03:14:07.000Z"},"severity":"Info"},
{"_id":{"oid":"5b5c8c3c625e4f33dceefd3d"},"data":{"totalOccupancy":{"occupied":51,"free":10}},"ufuid":"NEDAP-TR-IZM-006","utype":"Park
ingOccupancyChangedEvent","timestamp":{"date":"2038-01-19T03:14:07.000Z"},"severity":"Info"},
{"_id":{"oid":"5b971564625e4f315c329af1"},"data":{"nodeId":"00.0001.377E","status":"Occupied","overstay":false,"parkingLotId":"12"},"
ufuid":"NEDAP-TR-IZM-012","utype":"ParkingBayStatusChangedEvent","timestamp":{"date":"2038-01-19T03:14:07.000Z"},"severity":"Info"},
{"_id":{"oid":"5b971564625e4f315c329af5"},"data":{"totalOccupancy":{"occupied":88,"free":48}},"ufuid":"NEDAP-TR-IZM-012","utype":"Park
ingOccupancyChangedEvent","timestamp":{"date":"2038-01-19T03:14:07.000Z"},"severity":"Info"},
{"_id":{"oid":"5bf49537625e4f0684dd51e2"},"data":{"nodeId":"00.0001.3488","status":"Occupied","overstay":false,"parkingLotId":"24"},"
ufuid":"NEDAP-TR-IZM-024","utype":"ParkingBayStatusChangedEvent","timestamp":{"date":"2038-01-18T09:01:49.000Z"},"severity":"Info"},
{"_id":{"oid":"5bf49537625e4f0684dd51e6"},"data":{"totalOccupancy":{"occupied":21,"free":46}},"ufuid":"NEDAP-TR-IZM-024","utype":"Park
ingOccupancyChangedEvent","timestamp":{"date":"2038-01-18T09:01:49.000Z"},"severity":"Info"}])

```

Obr. 20. Ukážka upravených dát vo formáte JSON. Zdroj: vlastná tvorba.

Práve kvôli nutnosti úprav zdrojových dát a dlhému času importu je tento postup pre daný účel nevhodný. Je však pomerne nenáročný a zvládne ho i menej skúsený užívateľ a pre malý súbor dát by mohol postačovať.

6.3.2 Import dát prostredníctvom príkazu mongorestore

Príkaz `mongorestore` sa používa pri importovaní dát vo formáte BSON. V tomto formáte sú dáta takmer nečitateľné, nemajú žiadnu zrozumiteľnú štruktúru (Obr. 21). [17]

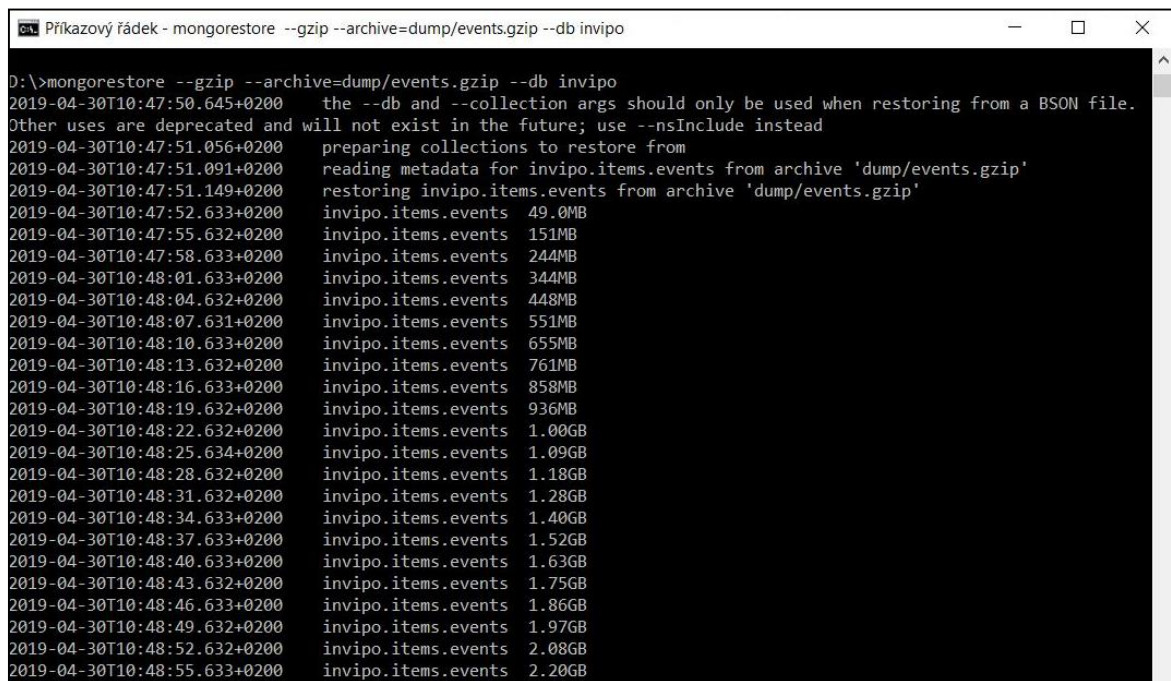
```

events ×
6D E2 99 81 65 00 00 00 10 63 6F 6E 63 75 72 72 65 6E 74 5F 63 6F 6C 6C 65 63 74 69 6F 6E 73 00 m...e...concurrent_collections.
04 00 00 00 02 76 65 72 73 69 6F 6E 00 04 00 00 00 30 2E 31 00 02 73 65 72 76 65 72 5F 76 65 72 .....version....0.1..server_ver
73 69 6F 6E 00 06 00 00 00 33 2E 32 2E 37 00 02 74 6F 6F 6C 5F 76 65 72 73 69 6F 6E 00 07 00 00 ta.5...{"options":{},"indexes":[
00 72 33 2E 32 2E 37 00 00 7E 02 00 00 02 64 62 00 07 00 00 00 69 6E 76 69 70 6F 00 02 63 6F 6C .r3.2.7...db....invipo..col
6C 65 63 74 69 6F 6E 00 00 00 00 69 74 65 6D 73 2E 65 76 65 6E 74 73 00 02 6D 65 74 61 64 61 lection....items.events..metada
74 61 00 35 02 00 00 7B 22 6F 70 74 69 6F 6E 73 22 3A 7B 7D 2C 22 69 6E 64 65 78 65 73 22 3A 5B ta.5...{"options":{},"indexes":[
7B 22 76 22 3A 31 2C 22 68 65 79 22 3A 7B 22 5F 69 64 22 3A 31 7D 2C 22 6E 61 6D 65 22 3A 22 5F {"v":1,"key":{"_id":1},"name":_
69 64 5F 22 2C 22 6E 73 22 3A 22 69 6E 76 69 70 6F 2E 69 74 65 6D 73 2E 65 76 65 6E 74 73 22 7D id"},"ns":"invipo.items.events"}
2C 7B 22 76 22 3A 31 2C 22 68 65 79 22 3A 7B 22 74 69 6D 65 73 74 61 6D 70 22 3A 31 7D 2C 22 6E ,{"v":1,"key":{"timestamp":1},"n
61 6D 65 22 3A 22 74 69 6D 65 73 74 61 6D 70 5F 31 22 2C 22 6E 73 22 3A 22 69 6E 76 69 70 6F 2E ame":"timestamp_1","ns":"invipo.
69 74 65 6D 73 2E 65 76 65 6E 74 73 22 7D 2C 7B 22 76 22 3A 31 2C 22 68 65 79 22 3A 7B 22 75 74 items.events"},{"v":1,"key":{"ut
79 70 65 22 3A 31 7D 2C 22 6E 61 6D 65 22 3A 22 75 74 79 70 65 22 2C 22 6E 73 22 3A 22 69 6E 76 ype":1},"name":"utype","ns":"inv
69 70 6F 2E 69 74 65 6D 73 2E 65 76 65 6E 74 73 22 2C 22 62 61 63 6B 67 72 6F 75 6E 64 22 3A 74 ipo.items.events","background":t
72 75 65 7D 2C 7B 22 76 22 3A 31 2C 22 68 65 79 22 3A 7B 22 75 66 69 64 22 3A 31 7D 2C 22 6E 61 rue),{"v":1,"key":{"ufid":1},"na
6D 65 22 3A 22 75 66 69 64 22 2C 22 6E 73 22 3A 22 69 6E 76 69 70 6F 2E 69 74 65 6D 73 2E 65 76 me":"ufid","ns":"invipo.items.ev
65 6E 74 73 22 2C 22 62 61 63 6B 67 72 6F 75 6E 64 22 3A 74 72 75 65 7D 2C 7B 22 76 22 3A 31 2C ents","background":true},{"v":1,

```

Obr. 21. Ukážka dát vo formáte BSON. Zdroj: vlastná tvorba.

Tento postup vyextrahuje dáta z formátu BSON do čitateľného formátu JSON. Zaujímavosťou je, že dáta tak zväčšili svoj objem približne stonásobne. Tento postup je však niekoľkonásobne rýchlejší, ako je možné vidieť na obrázku (Obr. 22), import 2 GB dát trval na rovnakom zariadení jednu minútu.



```
Příkazový řádek - mongorestore --gzip --archive=dump/events.gzip --db invipo
D:\>mongorestore --gzip --archive=dump/events.gzip --db invipo
2019-04-30T10:47:50.645+0200 the --db and --collection args should only be used when restoring from a BSON file.
Other uses are deprecated and will not exist in the future; use --nsInclude instead
2019-04-30T10:47:51.056+0200 preparing collections to restore from
2019-04-30T10:47:51.091+0200 reading metadata for invipo.items.events from archive 'dump/events.gzip'
2019-04-30T10:47:51.149+0200 restoring invipo.items.events from archive 'dump/events.gzip'
2019-04-30T10:47:52.633+0200 invipo.items.events 49.0MB
2019-04-30T10:47:55.632+0200 invipo.items.events 151MB
2019-04-30T10:47:58.633+0200 invipo.items.events 244MB
2019-04-30T10:48:01.633+0200 invipo.items.events 344MB
2019-04-30T10:48:04.632+0200 invipo.items.events 448MB
2019-04-30T10:48:07.631+0200 invipo.items.events 551MB
2019-04-30T10:48:10.633+0200 invipo.items.events 655MB
2019-04-30T10:48:13.632+0200 invipo.items.events 761MB
2019-04-30T10:48:16.633+0200 invipo.items.events 858MB
2019-04-30T10:48:19.632+0200 invipo.items.events 936MB
2019-04-30T10:48:22.632+0200 invipo.items.events 1.00GB
2019-04-30T10:48:25.634+0200 invipo.items.events 1.09GB
2019-04-30T10:48:28.632+0200 invipo.items.events 1.18GB
2019-04-30T10:48:31.632+0200 invipo.items.events 1.28GB
2019-04-30T10:48:34.633+0200 invipo.items.events 1.40GB
2019-04-30T10:48:37.633+0200 invipo.items.events 1.52GB
2019-04-30T10:48:40.633+0200 invipo.items.events 1.63GB
2019-04-30T10:48:43.632+0200 invipo.items.events 1.75GB
2019-04-30T10:48:46.633+0200 invipo.items.events 1.86GB
2019-04-30T10:48:49.632+0200 invipo.items.events 1.97GB
2019-04-30T10:48:52.632+0200 invipo.items.events 2.08GB
2019-04-30T10:48:55.633+0200 invipo.items.events 2.20GB
```

Obr. 22. Ukážka behu príkazu mongorestore. Zdroj: vlastná tvorba.

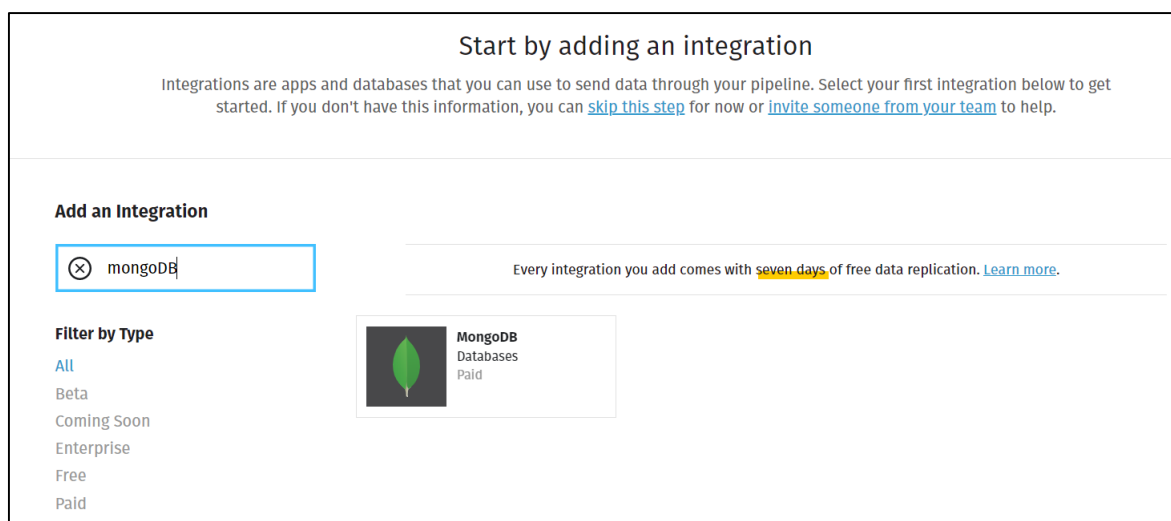
Import súboru s dátami pre potreby tejto práce o veľkosti približne 82 GB trval takmer celú hodinu. Akonáhle MongoDB dosiahol posledného riadku oznamujúceho import, na niekoľko minút sa „zasekol“ a nevykazoval žiadnu činnosť. Toto bolo potreba prečkať a nepokúšať sa znovu program rozbehnúť. Nakoniec program vypísal hlášku potvrdzujúcu dokončenie operácie.

7 IMPORT DÁT DO ZVOLENÉHO BI NÁSTROJA

Nasledující podkapitoly poskytnú detailný postup práce, ktorého výsledkom by mali byť dáta naimportované do BI nástroja, ktorý bol zvolený v kapitole 5.

7.1 Import dát prostredníctvom aplikácie Stitch

Stránky MongoDB [17] priamo ponúkajú možnosť pripojenia na Microsoft Power BI prostredníctvom nástroja Stitch (Obr. 23).



Obr. 23. Ukážka pripojenia MongoDB na Power BI prostredníctvom nástroja Stitch. [36]

Tento ETL nástroj môže byť veľmi užitočný pre technicky nezdatných užívateľov. Poskytuje veľké množstvo vstupných formátov a služieb pre import dát do najznámejších BI a analytických nástrojov. [36]

Bohužiaľ, jeho použitie pre MongoDB je platené, preto nebolo v rámci tejto práce možné vyskúšať jeho funkcionality.

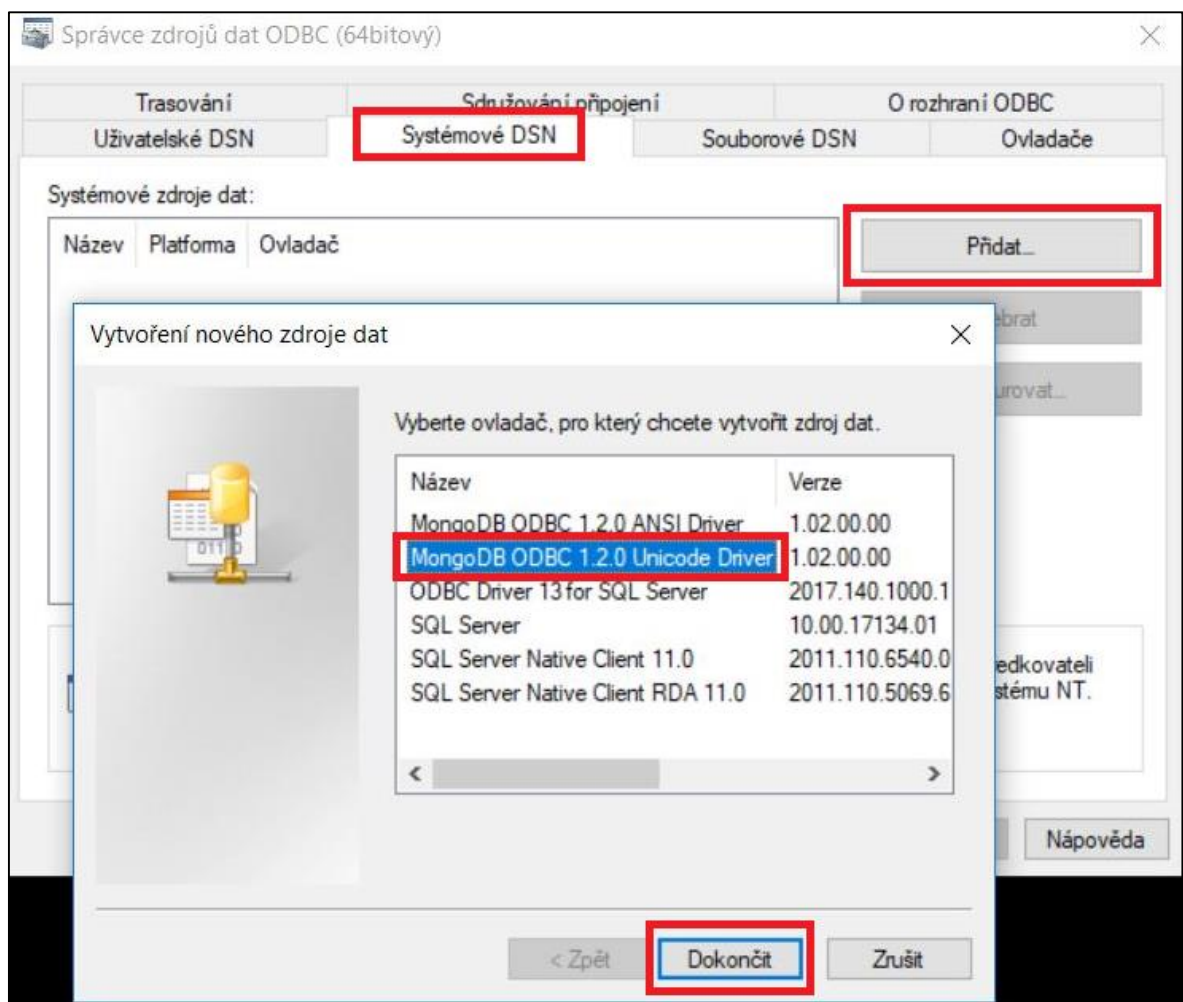
7.2 Import dát prostredníctvom ODBC

Pre import dát prostredníctvom ODBC (Open Database Connectivity) rozhrania je potrebné nainštalovať dva ovládače. Tieto ovládače sú dostupné na stránkach mongodb.com [17].

Prvý ovládač (*mongodb-bi-win32-x86_64-v2.11.0-beta1.msi*) umožní komunikovať systému Windows s databázou MongoDB. Druhý ovládač (*mongodb-connector-odbc-1.2.0-win-64-bit.msi*) umožní databáze MongoDB komunikovať s nástrojom Power BI, prípadne s programom Studio 3T. Bitová verzia ovládačov musí byť zhodná s bitovou verziou Power BI, Studio 3T a Windows.

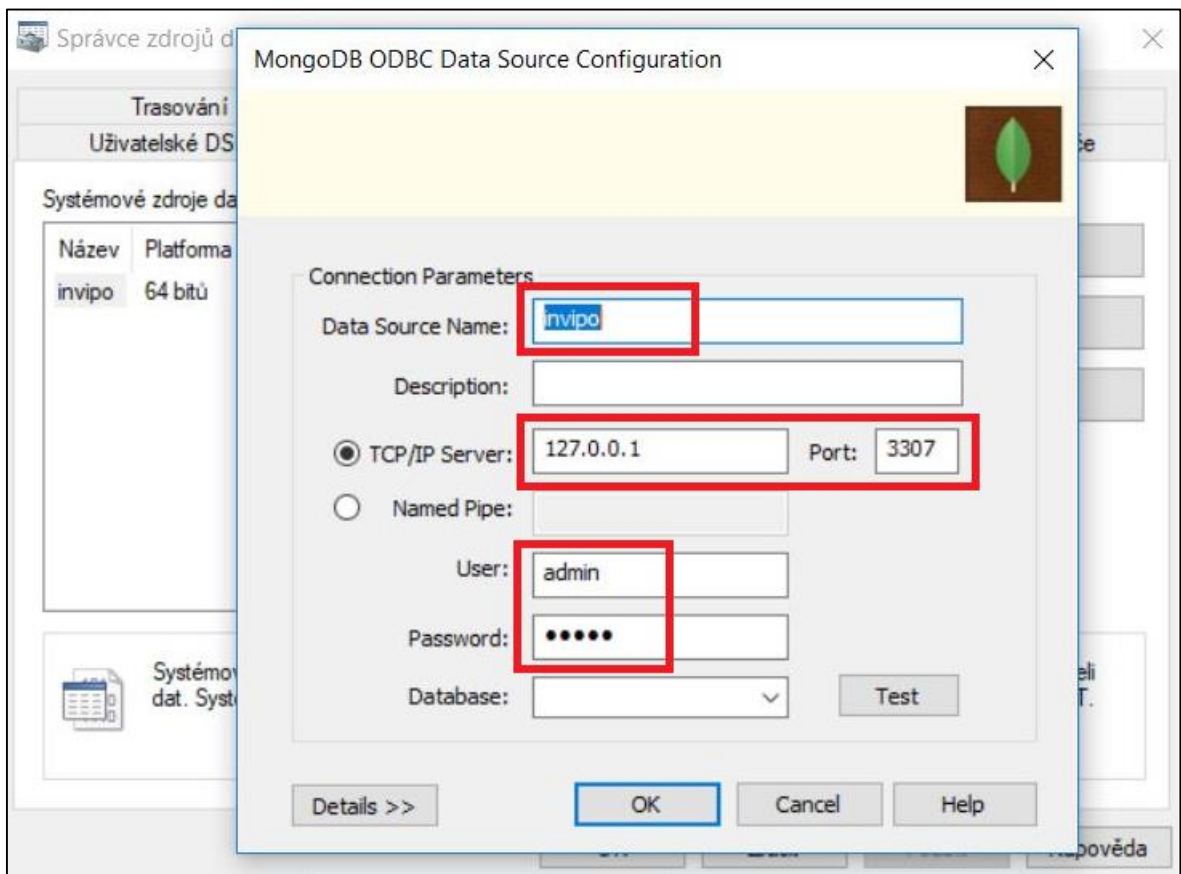
Nainštalovanie prvého ovládača vytvorí na disku zložku, v ktorej sa nachádza spustiteľný program mongosqld.exe. V tomto programe bude možné kontrolovať funkčnosť pripojenia na MongoDB databázy z akéhokoľvek programu, v tomto prípade z programov Power BI a Studio 3T.

Po inštalácii prvého ovládača je možné spustenie programu Správca zdrojov dát ODBC (Obr. 24). Na karte Systémové DSN (Data Source Name) je potrebné definovať nový zdroj dát – odkaz na lokálnu MongoDB databázu. Po inštalácii druhého ovládača by mala byť v zozname typov možných ovládačov dostupná možnosť „MongoDB ODBC Unicode Driver“.



Obr. 24. vytvorenie nového systémového zdroja dát. Zdroj: vlastná tvorba.

Data Source Name by sa mal zhodovať s názvom databázy. Pre správnu funkčnosť ODBC pripojenia je taktiež potrebné zadať akékoľvek užívateľské meno a heslo (Obr. 25).



Obr. 25. Konfigurácia systémového zdroja dát. Zdroj: vlastná tvorba.

Nasleduje kontrola prostredníctvom tlačidla Test. Kontrola odozvy by mala byť v programe mongosqld.exe viditeľná ako registrácia požiadavky na pripojenie (Obr. 26).

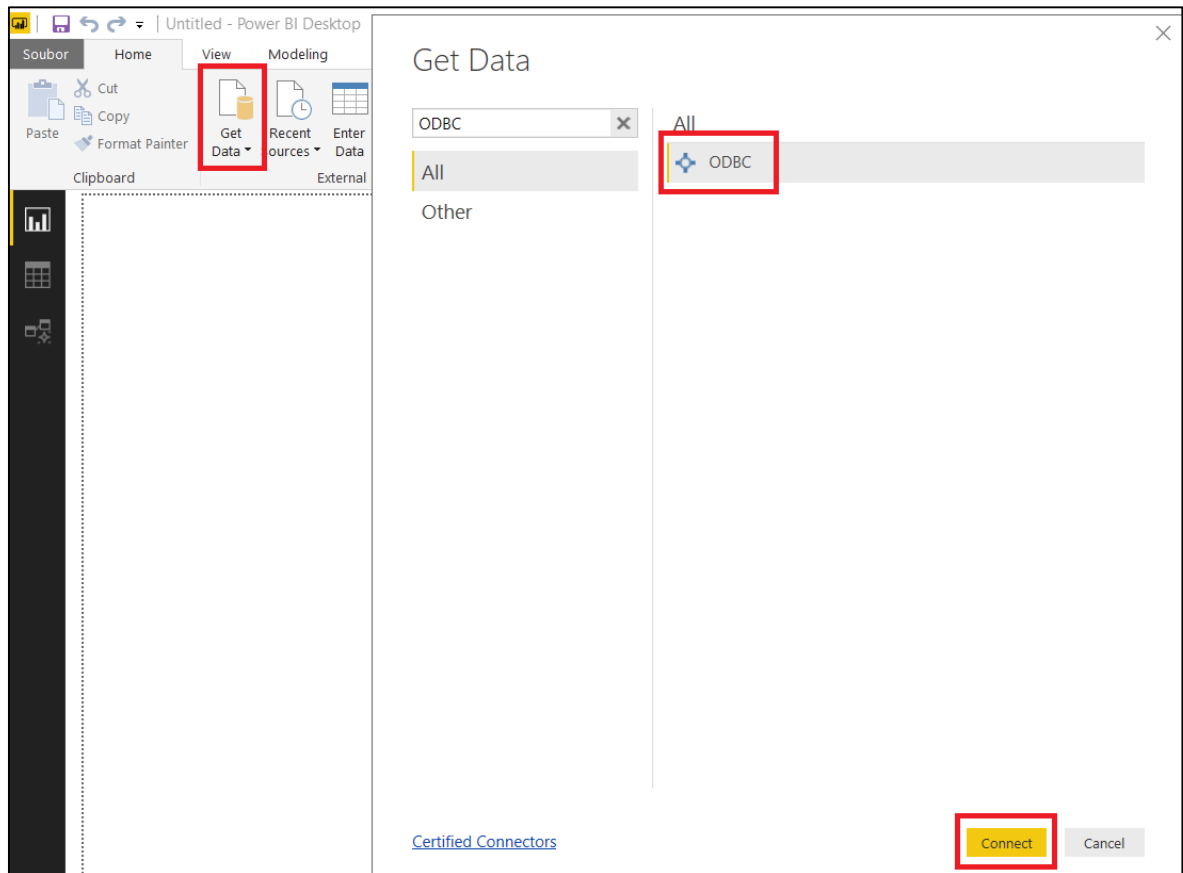
```

C:\Program Files\MongoDB\Connector for BI\2.1\bin\mongosqld.exe
2019-04-30T11:05:06.499+0200 I CONTROL [initandlisten] mongosqld starting: version=v2.11.0-beta1 pid=5996 host=C5-PRO
450G5-5
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] git version: f796585be06b4de34c95905ce20cc61e8532836c
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] OpenSSL version OpenSSL 1.0.2n-fips 7 Dec 2017 (built with Op
enSSL 1.0.2o 27 Mar 2018)
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] options: {}
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] ** NOTE: This is a development version (v2.11.0-beta1) of mong
osqld.
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] ** Not recommended for production.
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for mongosqld.
2019-04-30T11:05:06.529+0200 I CONTROL [initandlisten]
2019-04-30T11:05:06.534+0200 I NETWORK [initandlisten] waiting for connections at 127.0.0.1:3307
2019-04-30T11:05:06.534+0200 I SAMPLER [schemaDiscovery] initializing schema
2019-04-30T11:05:09.597+0200 I SAMPLER [schemaDiscovery] sampling MongoDB for schema...
2019-04-30T11:05:10.706+0200 I SAMPLER [schemaDiscovery] mapped schema for 2 namespaces: "invipo" (1): ["items.events
"]; "testdata" (1): ["testdata"]
2019-04-30T11:22:48.251+0200 I NETWORK [conn1] connection accepted from 127.0.0.1:53172 #1 (1 connection now open)
2019-04-30T11:22:50.356+0200 I NETWORK [conn1] end connection 127.0.0.1:53172 (0 connections now open)
2019-04-30T11:24:33.266+0200 I NETWORK [conn2] connection accepted from 127.0.0.1:53183 #2 (1 connection now open)

```

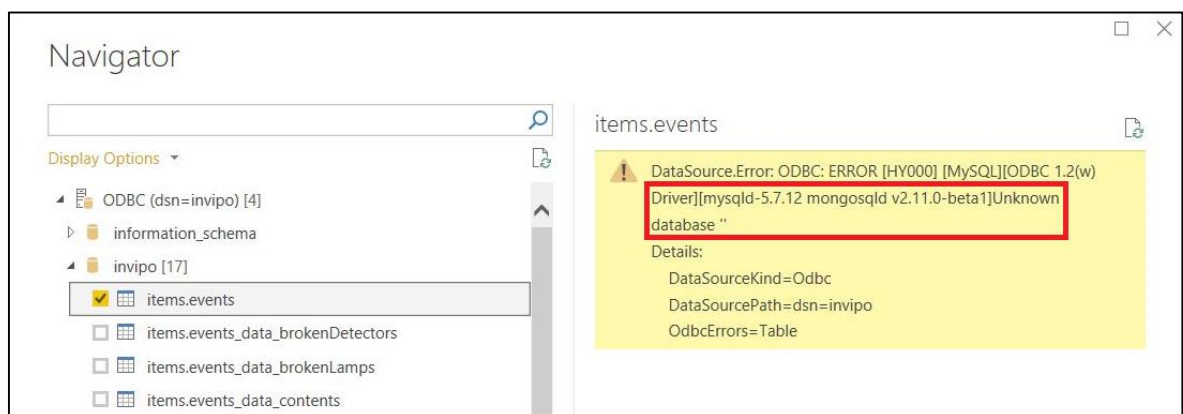
Obr. 26. Odozva programu mongosqld.exe na pokus o pripojenie MongoDB ODBC ovládača. Zdroj: vlastná tvorba.

V prípade, že program mongosqld.exe odpovedá, nasleduje potvrdenie a ukončenie programu Správca zdrojov. Ďalším krokom je spustenie Microsoft Power BI. Z možností kontextového okna je potrebné zvoliť zdroj typu ODBC (Obr. 27).



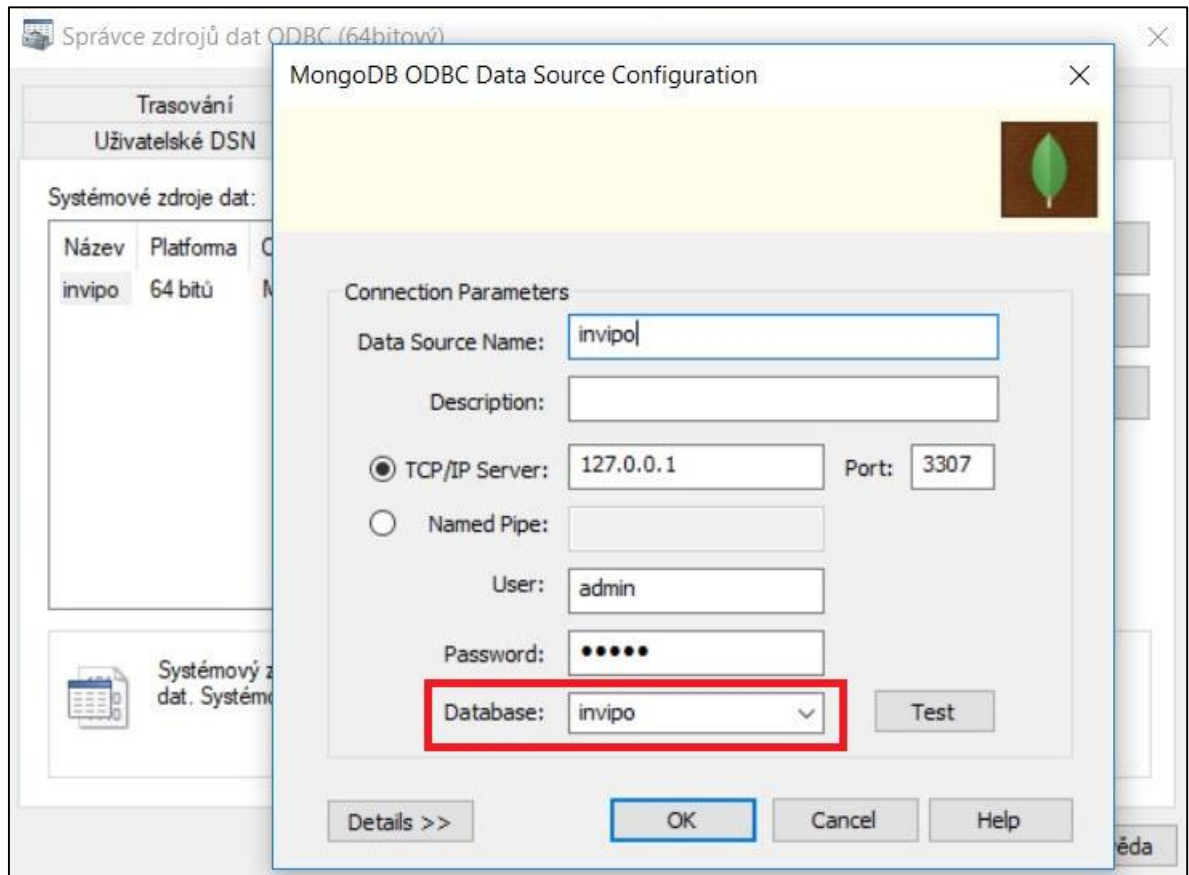
Obr. 27. Pripojenie na ODBC zdroj dát v prostredí Power BI. Zdroj: vlastná tvorba.

Navigátor zobrazí dostupné databázy pripojené v konektore Správca dát ODBC. Nasleduje výber požadovanej databázy, v tomto prípade databázy invipo items.events. Pri pokuse o pripojenie však nastáva chyba „Unknown database“ (Obr. 28).



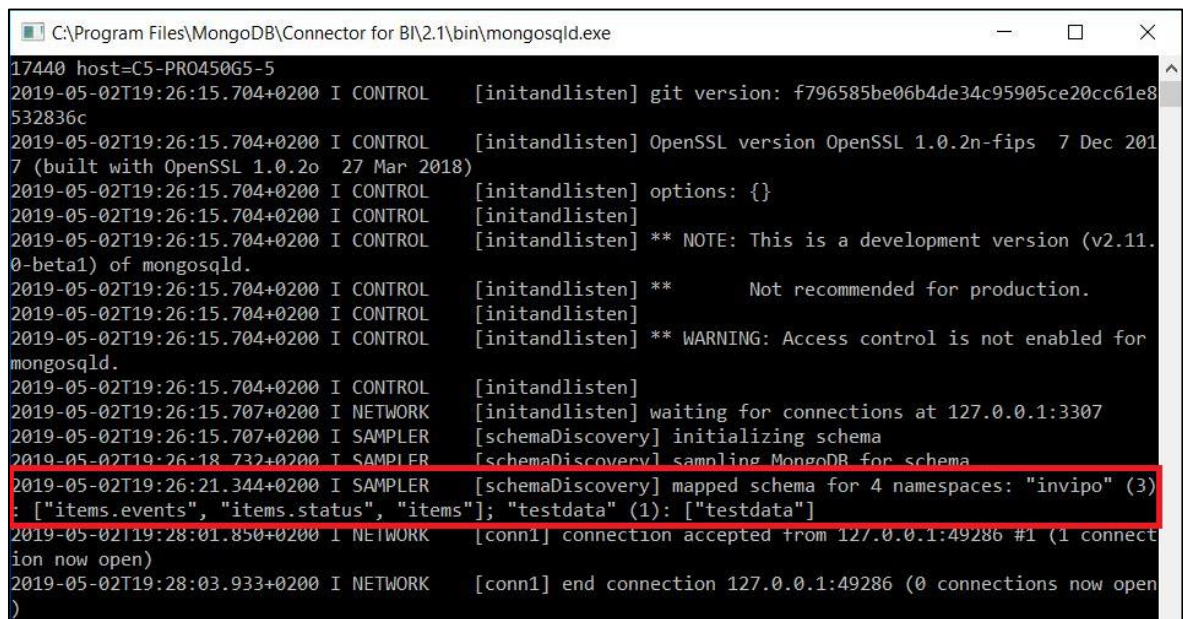
Obr. 28. Chyba pripojenia ODBC zdroja dát. Zdroj: vlastná tvorba.

Táto chyba má zjavnú príčinu – pri konfigurácii dátového zdroja v programe Správca zdrojov dát nebol špecifikovaný názov databáze – názov bol zadaný iba pre zdroj dát. Je preto potrebné stávajúci záznam upraviť a doplniť názov databázy (Obr. 29).



Obr. 29. Úprava konfigurácie systémového zdroja dát. Zdroj: vlastná tvorba.

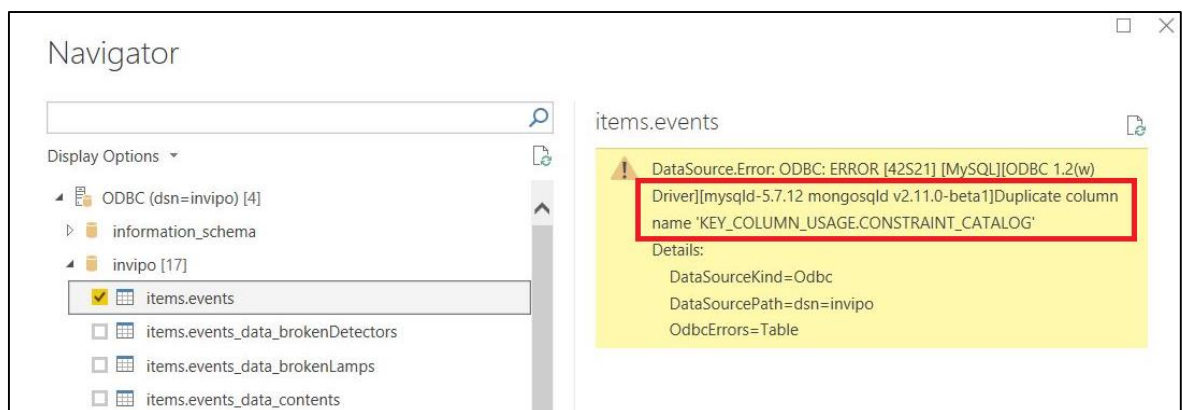
Opäť by malo nasledovať najprv otestovanie. Z obrázku (Obr. 30) je zrejmé, že nastala zmena – v zozname pribudli i zvyšné dve kolekcie items.status a items, ktoré bude takisto potrebné pripojiť do Power BI. Po uložení a potvrdení operácie tlačidlom OK je znovu možné zavrieť Správcu dát ODBC.



```
C:\Program Files\MongoDB\Connector for BI\2.1\bin\mongosqld.exe
17440 host=C5-PR0450G5-5
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] git version: f796585be06b4de34c95905ce20cc61e8532836c
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] OpenSSL version OpenSSL 1.0.2n-fips 7 Dec 2017 (built with OpenSSL 1.0.2o 27 Mar 2018)
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] options: {}
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] ** NOTE: This is a development version (v2.11.0-beta1) of mongosqld.
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] ** Not recommended for production.
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for mongosqld.
2019-05-02T19:26:15.704+0200 I CONTROL [initandlisten]
2019-05-02T19:26:15.707+0200 I NETWORK [initandlisten] waiting for connections at 127.0.0.1:3307
2019-05-02T19:26:15.707+0200 I SAMPLER [schemaDiscovery] initializing schema
2019-05-02T19:26:18.732+0200 I SAMPLER [schemaDiscovery] sampling MongoDB for schema
2019-05-02T19:26:21.344+0200 I SAMPLER [schemaDiscovery] mapped schema for 4 namespaces: "invipo" (3): ["items.events", "items.status", "items"]; "testdata" (1): ["testdata"]
2019-05-02T19:28:01.850+0200 I NETWORK [conn1] connection accepted from 127.0.0.1:49286 #1 (1 connection now open)
2019-05-02T19:28:03.933+0200 I NETWORK [conn1] end connection 127.0.0.1:49286 (0 connections now open)
```

Obr. 30. Odozva programu *mongosqld.exe* na zmenu konfigurácie MongoDB ODBC ovládača. Zdroj: vlastná tvorba.

Medzi zdrojmi dostupných ODBC databáz sa v rámci databázy *invipo* objavili všetky kolekcie (*items.events*, *items.status* a *items*). Nasleduje voľba požadovanej kolekcie a potvrdenie výberu.



Obr. 31. Chyba pripojenia ODBC zdroja dát. Zdroj: vlastná tvorba.

Z obrázku (Obr. 31) je zjavné, že znovu nastáva chyba. Bohužiaľ, bolo overené, že v danej kolekcii importovaných dát sa nenachádza žiadny stĺpec s názvom „CONSTRAINT_CATALOG“. Kvôli vágnosti tejto chybovej hlášky sa nepodarilo dohľadať riešenie, avšak podľa odpovedí na podobné otázky na fóre Microsoft Power BI je pravdepodobné, že sa jedná o nerelevantnú chybovú hlášku a v skutočnosti by mohlo ísť o chybu spôsobenú príliš veľkým množstvom záznamov vo vybranej kolekcii.

7.3 Import dát vo formáte JSON

Ďalšia možnosť importu dát do Power BI nástroja je import dát vo formáte JSON. Dáta vo formáte JSON je možné z databázy vyexportovať prostredníctvom programu Studio 3T [37] (Obr. 32).



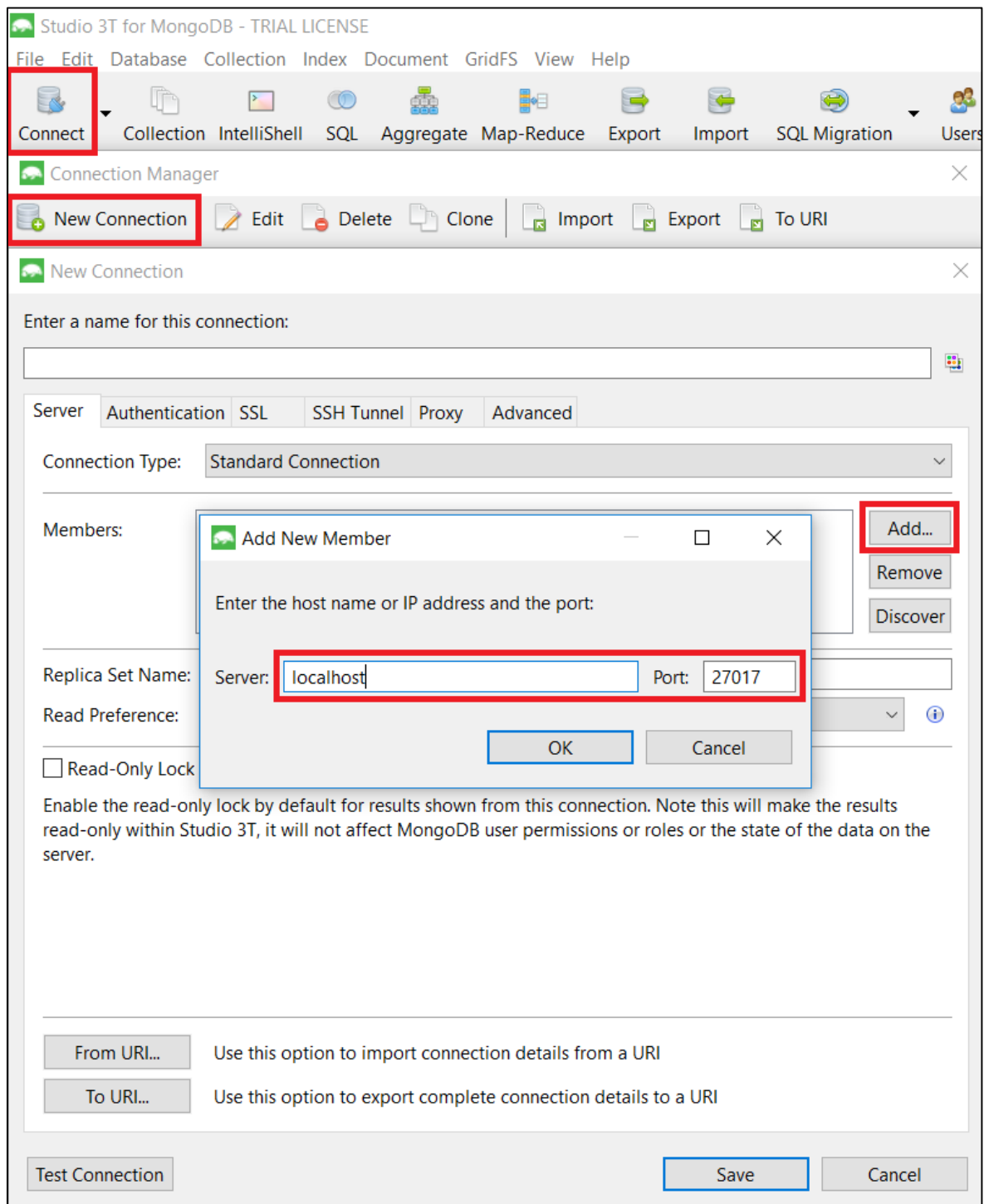
Obr. 32. Logo Studio 3T. [37]

7.3.1 Postup exportu dát do formátu JSON z MongoDB prostredníctvom programu Studio 3T

Studio 3T je IDE a GUI rozhraním pre MongoDB. Poskytuje 30 dňovú bezplatnú verziu, ktorá pre potreby tejto práce postačuje. Práca s programom Studio 3T je pomerne intuitívna a jednoduchá, nevyžaduje žiadne zvláštne predchádzajúce skúsenosti.

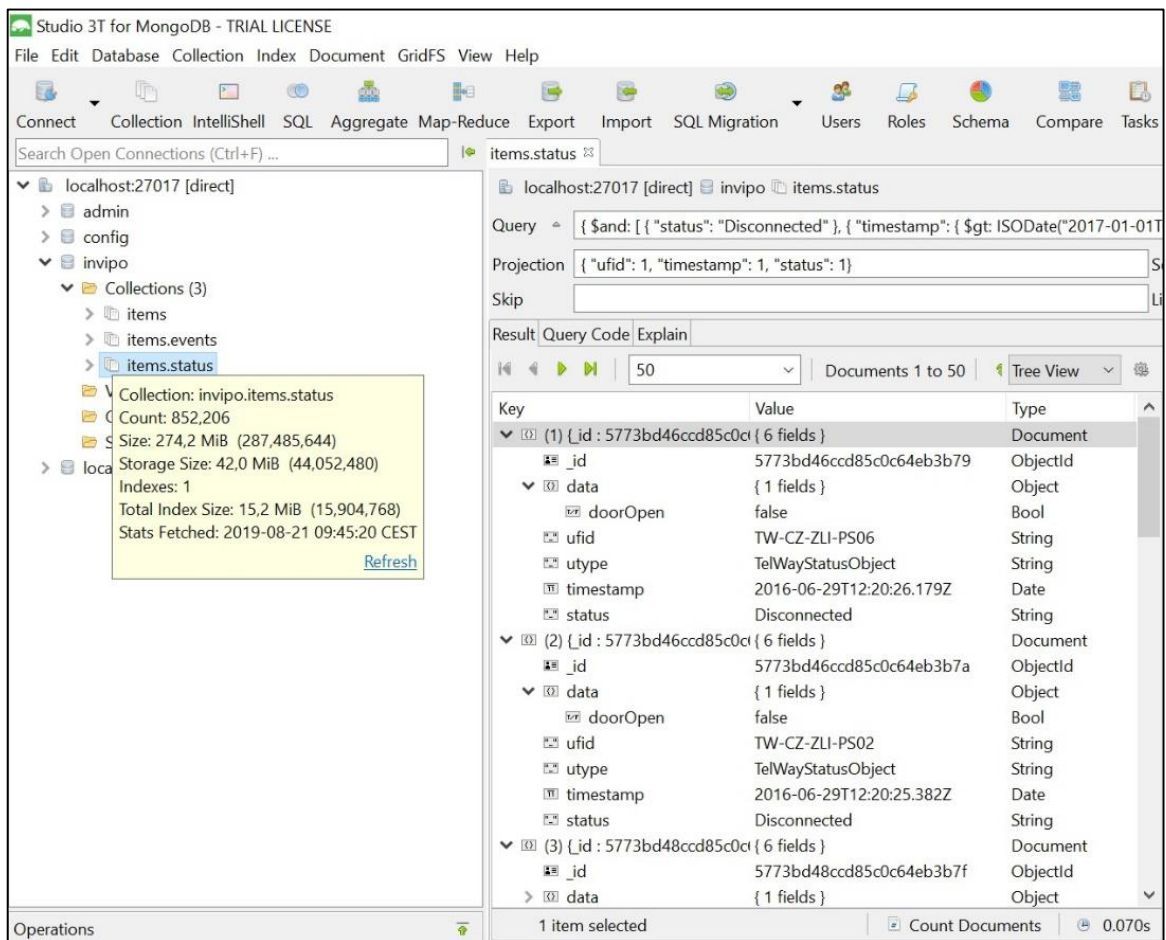
Po zvolení možnosti Connect sa zobrazí dialógové okno Connection Manager, v ktorom je možné pridať nové pripojenie. Názov pripojenia je nepovinný, je potrebné nastaviť jedine adresu serveru, v tomto prípade localhost a číslo portu, v tomto prípade 27017 (Obr. 33).

Pokus o toto pripojenie by mal byť takisto zaznamenaný v programe mongosqld.exe.



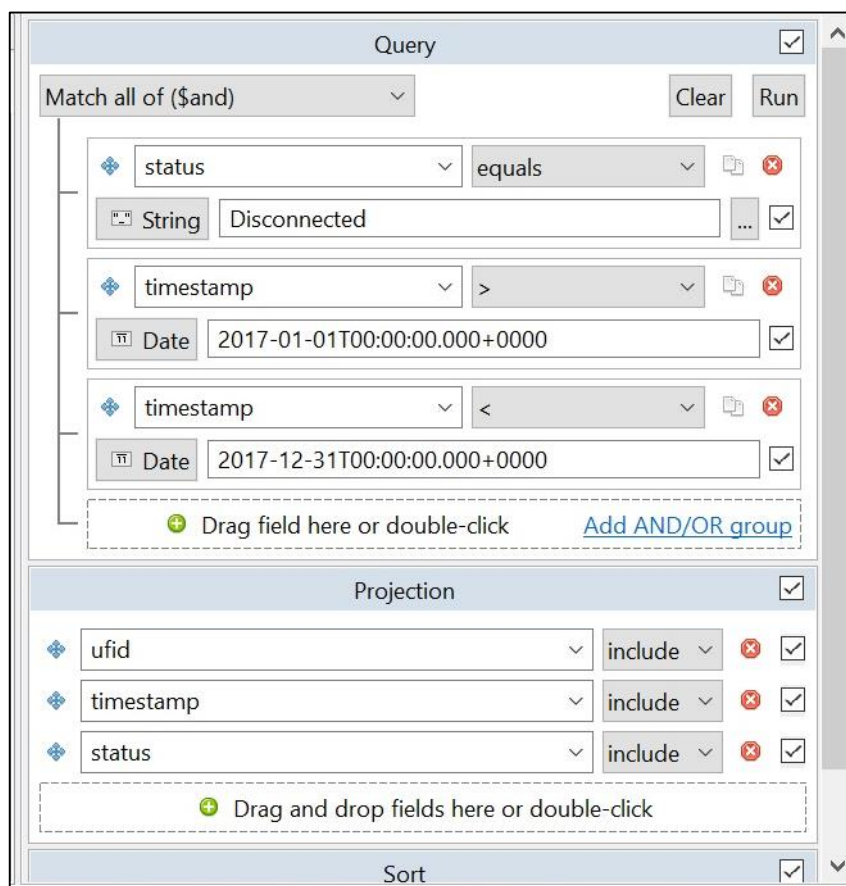
Obr. 33. Nové pripojenie databázy MongoDB v programe Studio 3T. Zdroj: vlastná tvorba.

V prípade, že pripojenie prebehlo v poriadku, po uložení je nové pripojenie v kontextovej ponuke programu Studio 3T k dispozícii (Obr. 34).



Obr. 34. Prostredie programu Studio 3T. Zdroj: vlastná tvorba.

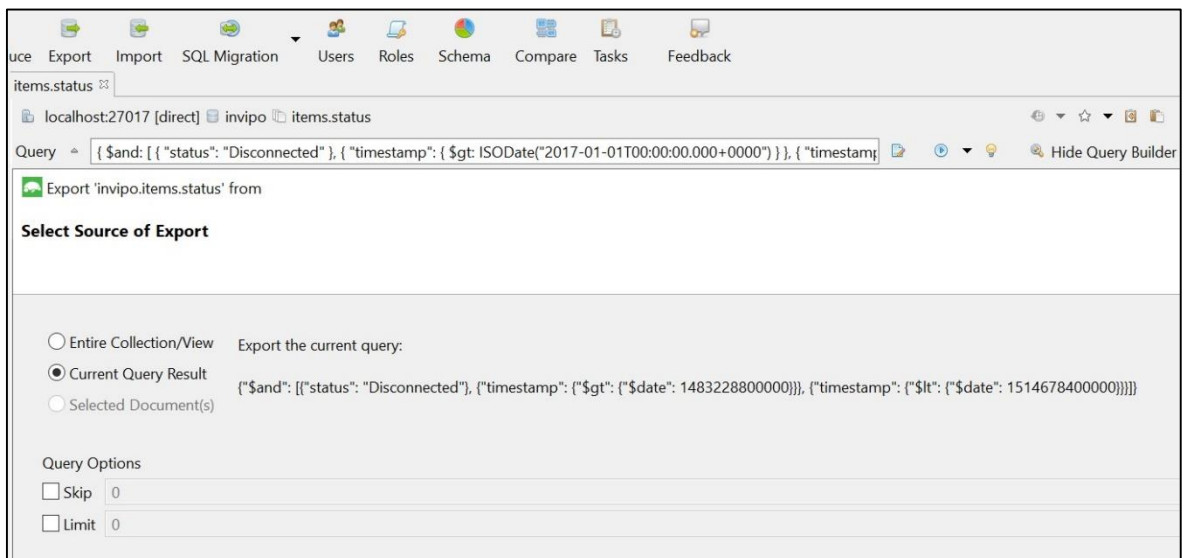
Voľbou príslušnej kolekcie je možné prezerať stránkované dáta, vytvárať a spúšťať vyhľadávania, pričom na základné vyhľadávanie nie je potrebné ovládať žiadny programovací jazyk a dajú sa vyvolať vďaka šikovnému GUI programu s názvom Visual Query Builder, ktorý umožňuje technológiou Drag and Drop jednoducho vytvárať pomerne zložité dotazy. (Obr. 35)



Obr. 35. Visual Query Builder programu Studio 3T.

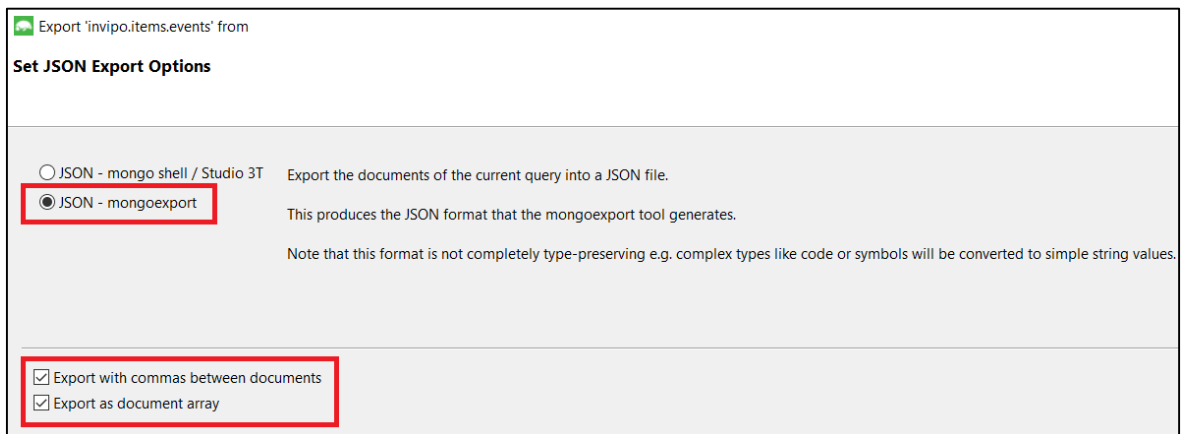
Zdroj: vlastná tvorba.

Pre potreby tejto práce bolo postačujúce vyexportovať dáta, ktoré sú voči stanovenej hypotéze relevantné. Takisto nie je nutné exportovať všetky pole nerelačnej databázy – postačia pole „ufid“, „timestamp“ a „status“. (Obr. 35) Výhodou je tak následná práca s menším súborom dát. V kontextovom okne exportu zvolíme možnosť „Current Query Result“. (Obr. 36)



Obr. 36. Možnosti exportu dát v programe Studio 3T. Zdroj: vlastná tvorba.

Ďalšie dôležité nastavenia exportu sú zobrazené na obrázku (Obr. 37).



Obr. 37. Možnosti exportu dát v programe Studio 3T. Zdroj: vlastná tvorba.

8 SPRACOVANIE DÁT VO ZVOLENOM BI SYSTÉME

Požiadavkou zadávateľa diplomovej práce, spoločnosti Incinity s. r. o. [15] bolo nájsť v dátach korelácie výpadkov rôznych typov zariadení monitorujúcich dopravu v meste Zlín voči extrémnym výkyvom počasia (oficiálna hypotéza je formulovaná v úvode tejto práce). Dáta o počasí boli čerpané z verejne dostupnej štatistiky Českého hydrometeorologického ústavu. [38]

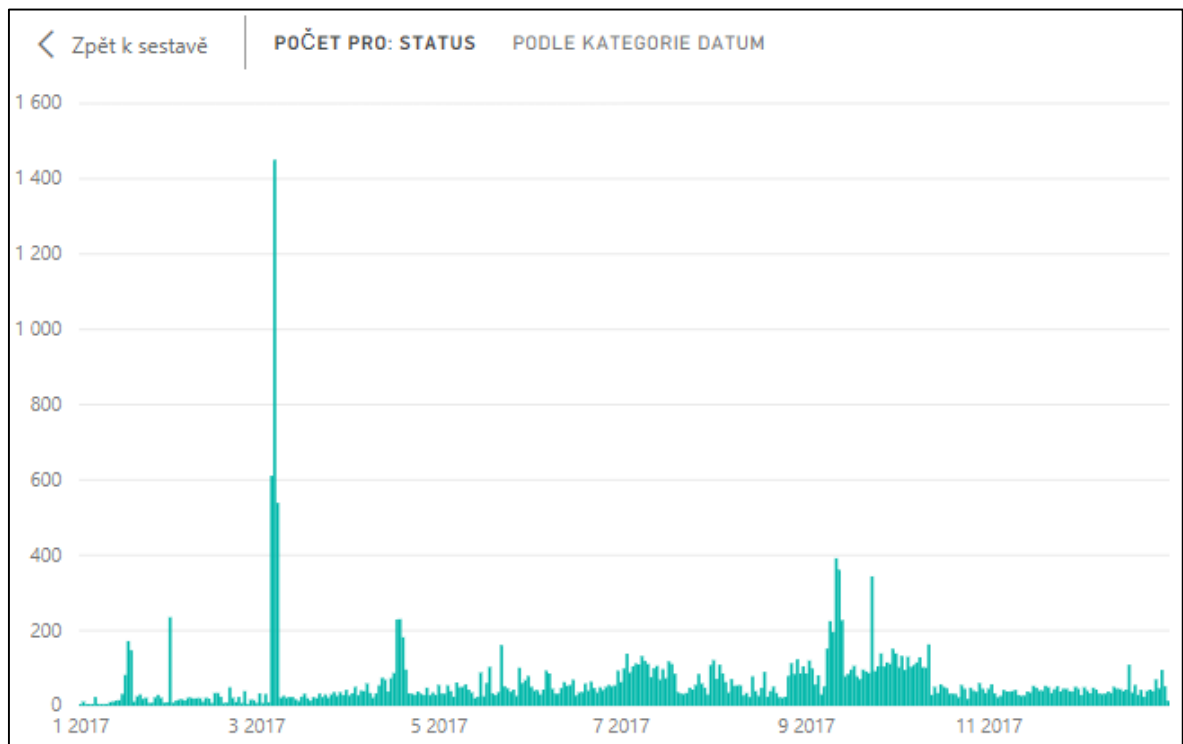
8.1 Výpadkovosť zariadení

Údaje o výpadkovosti zariadení je možné získať z kolekcie Invipo Status. Vyexportované boli všetky záznamy z tejto kolekcie, ktoré obsahovali pre kľúč „status“ hodnotu „Disconnected“.

8.1.1 Stanovenie vstupného súboru a čistenie dát

Prvým krokom každej dátovej analýzy by malo byť počiatočné preskúmanie dát za účelom odhalenia extrémnych hodnôt. Tieto majú vplyv na spoľahlivosť výsledkov väčšiny pokročilých štatistických techník, a preto je žiaduce ich zo vstupného súboru dát vhodným spôsobom vyradiť. [39] Priamo totiž ovplyvňujú všetky dôležité štatistické ukazatele (aritmetický priemer, medián, smerodajná odchýlka a pod.).

Vstupný súbor dát predstavoval údaje o výpadkovosti zariadení v roku 2017. Na obrázku (Obr. 38) je možné pozorovať výskyt extrémnych hodnôt počtu hlásených výpadkov pre určité dni.



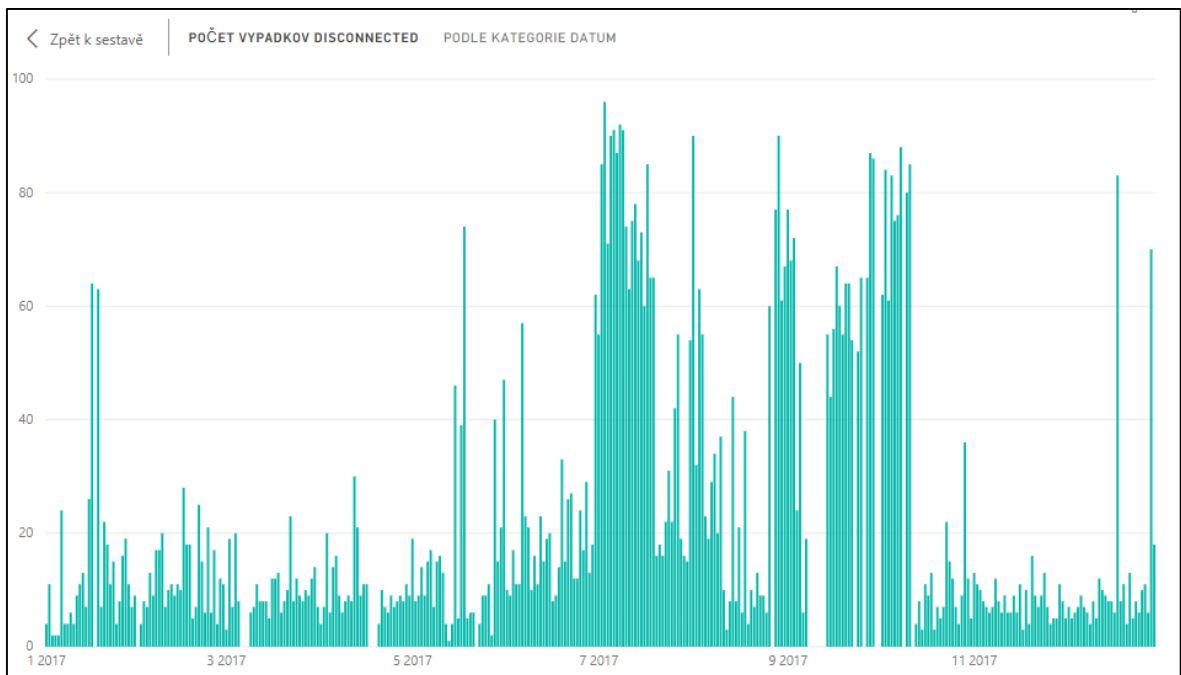
Obr. 38. Graf počtu hlásených výpadkov (status: Disconnected) pre jednotlivé dni roku 2017. Zdroj: vlastná tvorba.

Postup orientačného vylúčenia extrémnych hodnôt bol nasledovný [40]:

1. stanovenie podozrivej (extrémnej) hodnoty y ,
2. výpočet aritmetického priemeru \bar{x} zo súboru bez podozrivej hodnoty y ,
3. výpočet trojnásobku smerodajnej odchýlky s zo súboru bez podozrivej hodnoty y ,
4. výpočet absolútnej hodnoty $|z|$ rozdielu podozrivej hodnoty y od aritmetického priemeru \bar{x} ,
5. porovnanie absolútnej hodnoty $|z|$ s trojnásobkom smerodajnej odchýlky s :
 - a. ak $|z| \geq 3s$ – podozrivá hodnota je extrémna a celý postup opakujeme s ďalšou nižšou podozrivou hodnotou;
 - b. ak $|z| < 3s$ – podozrivá hodnota nie je extrémna a zahrnieme ju do ďalšieho štatistického spracovania.

Samotné čistenie dát spočívalo vo vyradení dní, v ktorých bolo zaznamenaných extrémne mnoho výpadkov zariadení. Jednalo sa konkrétne o 22 z 365 dní – 17. 1., 31. 1., 6. 3., 7. 3., 8. 3., 17. 4., 18. 4., 19. 4., 22. 5., 27. 8., 8. 9., 10. 9., 11. 9., 12. 9., 13. 9., 23. 9., 26. 9., 30. 9., 1. 10., 9. 10., 12. 10. a 31. 12. 2017.

Po vyčistení dát (odstránení všech hodnôt vyšších alebo rovných podozrivej hodnote), má štatistika počtu výpadkov v jednotlivé dni nasledujúci tvar (Obr. 39)



Obr. 39. Graf počtu hlásených výpadkov (status: Disconnected) pre jednotlivé dni roku 2017 po vyčistení od extrémnych hodnôt. Zdroj: vlastná tvorba.

8.1.2 Transformácia počtu výpadkov na stupnicu výpadkovosti

Vyčistené hodnoty výpadkovosti pre jednotlivé dni bolo následne potrebné previesť na škálu od 1 do 10, pričom stupeň 1 by značil najmenšiu výpadkovosť a stupeň 10 výpadkovosť najväčšiu. Postup bol nasledovný (Obr. 40):

1. stanovenie najvyššej hodnoty zo súboru hodnôt x_{max} ,
2. stanovenie najnižšej hodnoty zo súboru hodnôt x_{min} ,
3. výpočet rozptylu t odpočítaním najnižšej hodnoty x_{min} od najvyššej hodnoty x_{max} ,
4. výpočet normalizačnej hladiny h vydelením rozptylu t číslom, ktoré predstavuje žiadaný počet stupňov - indexov i ,
5. výpočet hraničných hodnôt jednotlivých stupňov:
 - a. hraničná hodnota pre prvý stupeň i_1 je vypočítaná pripočítaním normalizačnej hladiny h k najnižšej hodnote x_{min} ,
 - b. hraničná hodnota pre každý ďalší stupeň i_{n+1} je vypočítaná pričítaním normalizačnej hladiny h k hodnote stupňa i_{n-1} .

The screenshot shows the Power BI Desktop interface. The ribbon is set to 'Modelování' (Modeling). The formula bar contains the following DAX formula:

```
1 Index výpadkovosti = IF('Výpadky'[Počet výpadkov Disconnected]<10,5;1;IF('Výpadky'[Počet výpadkov Disconnected]<20;2;IF('Výpadky'[Počet výpadkov Disconnected]<29,5;3;IF
```

The data table below shows the results of the calculation for the month of January 2017:

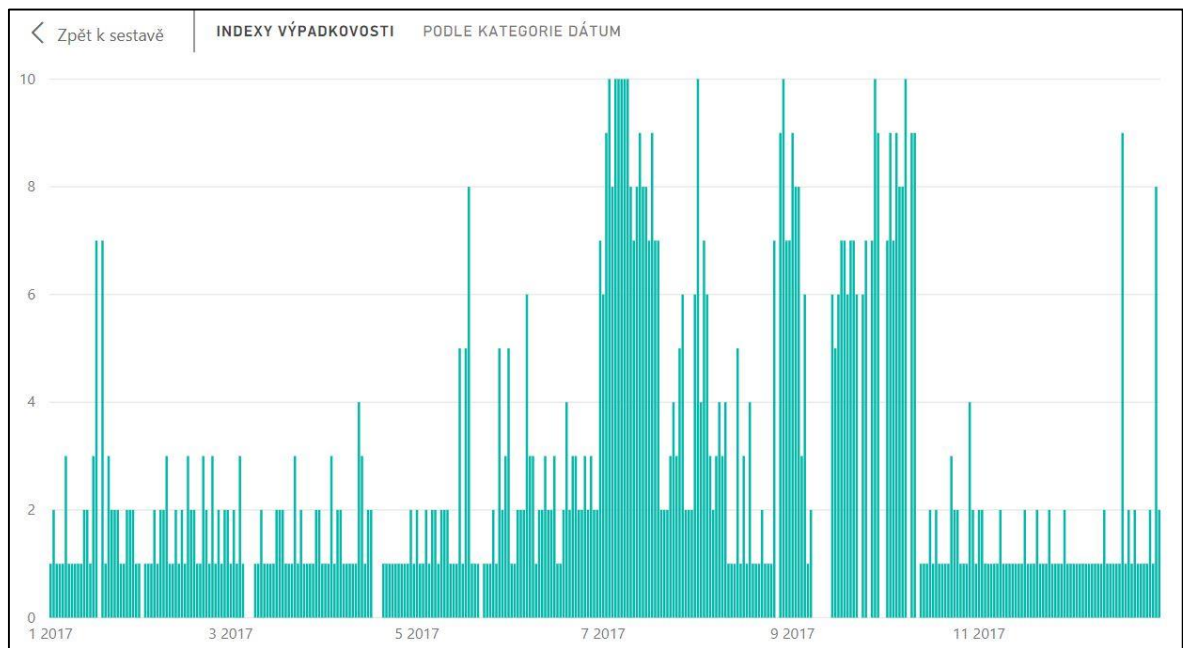
Dátum	Počet výpadkov Disconnected	Index výpadkovosti
neděle 1. ledna 2017	4	1
pondělí 2. ledna 2017	11	2
úterý 3. ledna 2017	2	1
středa 4. ledna 2017	2	1
čtvrtek 5. ledna 2017	2	1
pátek 6. ledna 2017	24	3
sobota 7. ledna 2017	4	1
neděle 8. ledna 2017	4	1
pondělí 9. ledna 2017	6	1
úterý 10. ledna 2017	4	1
středa 11. ledna 2017	9	1
čtvrtek 12. ledna 2017	11	2
pátek 13. ledna 2017	13	2
sobota 14. ledna 2017	7	1
neděle 15. ledna 2017	26	3
pondělí 16. ledna 2017	64	7
středa 18. ledna 2017	63	7
čtvrtek 19. ledna 2017	7	1
pátek 20. ledna 2017	22	3
sobota 21. ledna 2017	18	2

At the bottom of the table, it says: TABULKA: Výpadky (počet řádků: 342) SLOUPEC: Index výpadkovosti (počet odlišných hodnot: 10)

Obr. 40. Postup prevodu počtu výpadkov na stupnicu 1-10 v programe Power BI. Zdroj: vlastní tvorba.

8.1.3 Graf výpadkovosti zariadení

Štatistiku počtu výpadkov v jednotlivé dni s normalizovanými hodnotami do stupnice 1-10 opisuje graf výpadkovosti zariadení (Obr. 41), ktorý znázorňuje stupne výpadkovosti pre jednotlivé dni roku 2017 po vyčistení od dní, v ktorých dosahovala výpadkovosť extrémnych hodnôt.



Obr. 41. Stupne výpadkovosti zariadení (status: Disconnected) pre jednotlivé dni roku 2017 po vyčistení od extrémnych hodnôt a po prevedení na stupnicu 1-10. Zdroj: vlastná tvorba.

8.2 Extrémnosť počasia

Dáta o počasi z Českého hydrometeorologického ústavu [38] bolo potrebné vyčistiť, normalizovať a naimportovať spoločne s vyčistenými a normalizovanými dátami z platformy Invipo do Power BI, aby bolo možné zobrazit' a spracovať v rámci jednej vizualizácie dáta z oboch zdrojov.

8.2.1 Čistenie dát o počasi

Aby bolo možné porovnávať údaje o výpadkovosti s hodnotami počasia, bolo najprv potrebné z tabuľky s údajmi o počasi pre jednotlivé dni odstrániť rovnaké dni, ktoré boli odstránené pri čistení dát súvisiacim s výpadkovosťou zariadení.

8.2.2 Rozdelenie roku na dve ročné obdobia

Pre potreby tejto práce bol skúmaný rok 2017 rozdelený na dve ročné obdobia – leto a zimu, pričom „zima“ trvala do polovice jari a od polovice jesene, a leto od polovice jari do polovice jesene 2017 [41].

8.2.3 Transformácia údajov o počasí na stupnicu extrémnosti počasia

Skúmanými parametrami bola maximálna a minimálna teplota pre daný deň, percentuálna vlhkosť vzduchu, úhrn zrážok a počet hodín slnečného svitu za deň. Tieto údaje sú k nahliadnutiu v prílohe (P III).

Rovnakým postupom ako v kapitole 8.1.2 prevedieme hodnoty pre jednotlivé parametre počasia na stupnicu 1-10, pričom 1 označuje najmiernejšie a 10 najextrémnejšie hodnoty voči danému ročnému obdobiu. Extrémnosť počasia v lete a extrémnosť počasia v zime závisí od rozličných parametrov. Tieto boli pre potreby skúmanej hypotézy experimentálne označené a boli im stanovené experimentálne váhy.

8.2.3.1 Extrémnosť počasia v letnom období

V letnom období bola výpadkovosť zariadení skúmaná voči kombinácii vysokých teplôt a intenzívneho slnečného svitu, pričom v prípade výskytu zrážok sa predpokladalo mierne ochladenie zariadení. Preto bol pre stanovenie indexu extrémnosti počasia v lete pre každý letný deň stanovený nasledujúci pomer:

$$0,45 * \text{index maximálnej teploty} + 0,45 * \text{index slnečného svitu} + 0,1 * \text{index úhrnu zrážok}$$

Výsledné hodnoty bolo následne nutné previesť na stupnicu 1-10 rovnakým postupom ako v kapitole 8.1.2.

8.2.3.2 Extrémnosť počasia v zimnom období

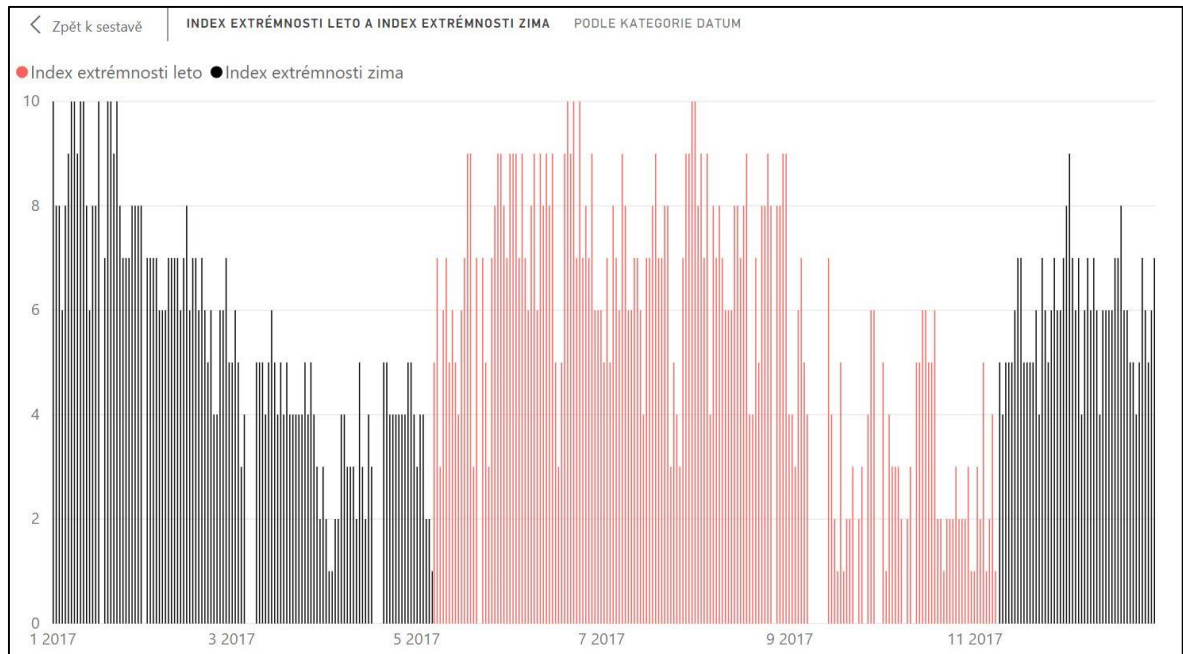
V zimnom období bola výpadkovosť zariadení skúmaná voči kombinácii nízkych teplôt a vysokej vlhkosti vzduchu, pričom v prípade výskytu slnečného svitu sa predpokladalo mierne oteplenie zariadení. Preto bol pre výpočet hodnoty extrémnosti počasia v zime pre každý zimný deň stanovený nasledujúci pomer:

$$0,8 * \text{index minimálnej teploty} + 0,1 * \text{index vlhkosti vzduchu} - 0,1 * \text{index slnečného svitu}$$

Výsledné hodnoty bolo následne nutné previesť na stupnicu 1-10 rovnakým postupom ako v kapitole 8.1.2.

8.2.4 Graf extrémnosti počasia

Graf extrémnosti počasia (Obr. 42) znázorňuje stupne extrémnosti počasia pre jednotlivé dni roku 2017 po vyčistení od dní, v ktorých dosahovala výpadkovosť extrémnych hodnôt.



Obr. 42. Stupne extrémností počasia pre jednotlivé letné a zimné dni roku 2017. Zdroj: vlastná tvorba.

8.3 Výpočet korelácie výpadkovosti zariadení voči extrémnosti počasia

Pre výpočet korelácií zimných a letných hodnôt bol použitý Pearsonov korelačný koeficient (Obr. 43).

$$r = \frac{\sum_i (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} * \sqrt{\sum_i (y_i - \bar{y})^2}}$$

Obr. 43. Vzorec pre výpočet Pearsonovho korelačného koeficientu [42].

Hodnoty x v uvedenom vzorci (Obr. 43) predstavujú hodnoty výpadkovosti a hodnoty y predstavujú hodnoty extrémnosti počasia. Farebné rozdelenie jednotlivých častí vzorca Pearsonovho korelačného koeficientu slúži k označeniu medzivýpočtov, pretože Microsoft Power BI nemá definovanú funkciu pre jeho výpočet – preto bolo potrebné každú z farebne označených častí vypočítať pre každý deň vo forme nových stĺpcov v Power BI (Obr. 44).

Datum	Index extrémnosti	Index výpadkovosti	Leto_Medzivýpočet1	Leto_Medzivýpočet2	Leto_Medzivýpočet3
neděle 7. května 2017	5	2	1,66762439156301	4,02328961600865	0,691218226068145
pondělí 8. května 2017	7	1	-3,51260816657653	9,0349175229854	1,36563683071931
úterý 9. května 2017	3	2	5,67925229853975	4,02328961600865	8,01679962141698
středa 10. května 2017	6	2	-0,338189561925365	4,02328961600865	0,0284275283937263
čtvrtek 11. května 2017	7	2	-2,34400351541374	4,02328961600865	1,36563683071931
pátek 12. května 2017	5	1	2,49901974040022	9,0349175229854	0,691218226068145
sobota 13. května 2017	6	1	-0,506794213088155	9,0349175229854	0,0284275283937263
neděle 14. května 2017	5	1	2,49901974040022	9,0349175229854	0,691218226068145
pondělí 15. května 2017	4	6	-3,65214305029746	3,97677798810168	3,35400892374256
úterý 16. května 2017	6	1	-0,506794213088155	9,0349175229854	0,0284275283937263
středa 17. května 2017	7	4	-0,00679421308815525	3,3802055164949E-05	1,36563683071931
čtvrtek 18. května 2017	9	8	12,6559964845863	15,9535221741482	10,0400554353705
pátek 19. května 2017	9	1	-9,52423607355327	9,0349175229854	10,0400554353705

Obr. 44. Medzivýpočty pre výpočet Pearsonovho korelačného koeficientu v prostredí Microsoft Power BI. Zdroj: vlastná tvorba.

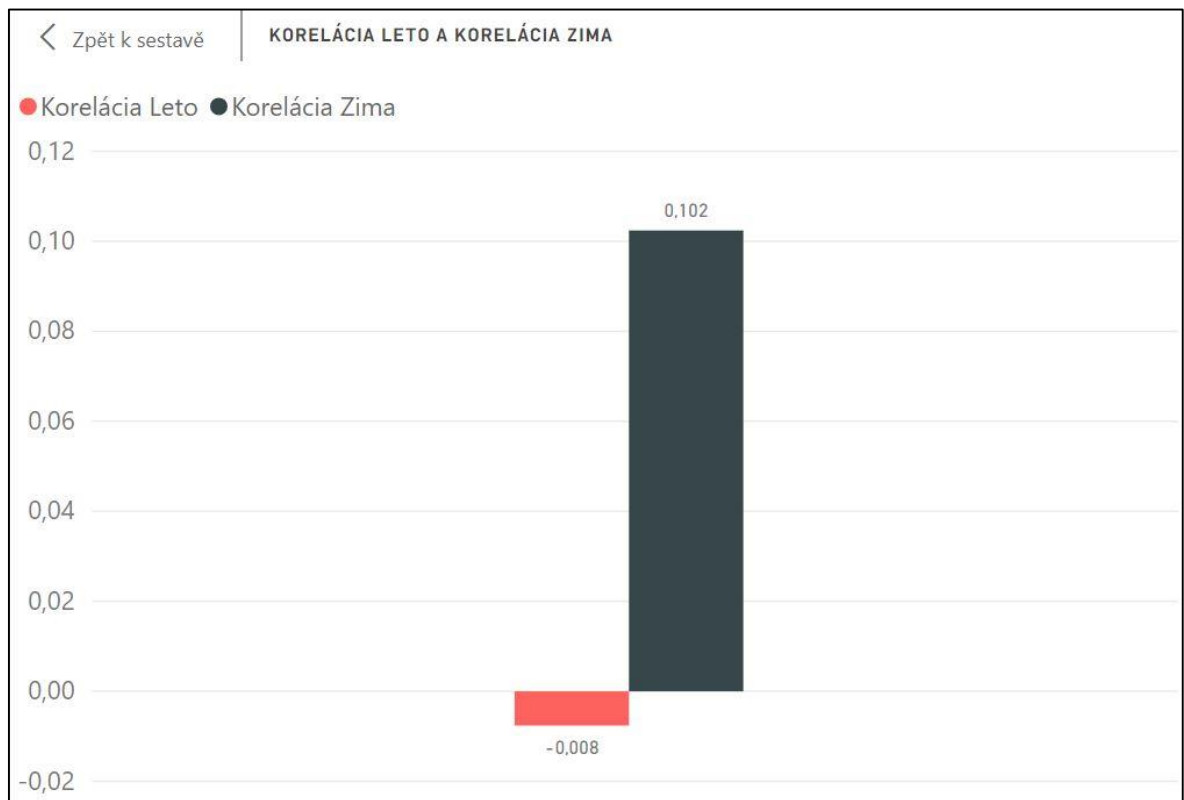
Korelačný koeficient bol následne v Power BI vypočítaný ako miera. Korelačný koeficient môže nadobúdať hodnôt z intervalu $<-1, 1>$, pričom „ak je korelačný koeficient rovný 0, znamená to, že medzi skúmanými veličinami nie je žiadny vzťah. Ak je rovný jednej, tak sú veličiny navzájom priamo závislé, nárast jednej veličiny spôsobuje nárast druhej, ak je rovný -1 , nárast jednej veličiny spôsobuje pokles druhej veličiny“ [43]. Hinkle a kol. uvádzajú interpretáciu jednotlivých hladín hodnôt korelačného koeficientu (Tab. 3).

Tab. 3. Interpretácia hodnôt Pearsonovho korelačného koeficientu [44].

Hodnota korelačného koeficientu	Interpretácia
0,90 do 1,00 (-1,00 do -0,90)	veľmi vysoká korelácia
0,70 do 0,90 (-0,90 do -0,70)	vysoká korelácia
0,50 do 0,70 (-0,70 do -0,50)	stredná korelácia
0,30 do 0,50 (-0,50 do -0,30)	nízka korelácia
0,00 do 0,30 (-0,30 do 0,00)	veľmi nízka príp. žiadna korelácia

8.3.1 Výsledné hodnoty korelačného koeficientu

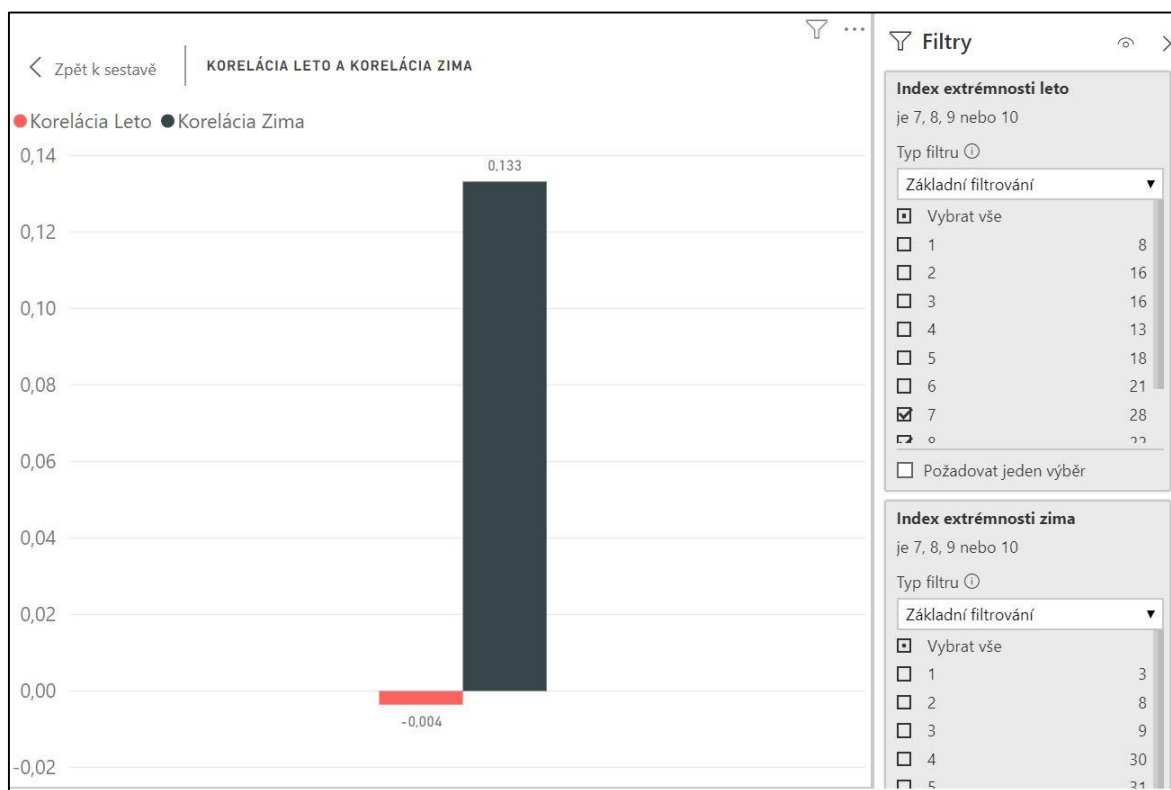
Korelačné koeficienty boli spočítané pre obe obdobia – leto a zimu – zvlášť. Ich hodnoty sú graficky znázornené na obrázku (Obr. 45).



Obr. 45. Hodnoty Pearsonovho korelačného koeficientu pre vyčistený nefiltrovaný vstupný súbor dát. Zdroj: vlastná tvorba.

Korelačný koeficient pre zimné obdobie dosiahol pre dané vstupné dáta hodnoty 0,102. Korelačný koeficient pre letné obdobie dosiahol hodnoty -0,008. Obe hodnoty spadajú do oblasti veľmi nízkej, prípadne žiadnej korelácie podľa Hinkle a kol. [44].

Obrázok (Obr. 46) znázorňuje hodnoty korelačného koeficientu pre dni, v ktorých bola extrémnosť počasia stanovená na vyššiu alebo rovnú stupni 7. Koeficienty sa iba mierne zlepšili smerom k potvrdeniu hypotézy – pre zimné obdobie dosiahol hodnoty 0,133 a pre letné obdobie hodnoty -0,004 – avšak vypočítaná závislosť je stále veľmi nízka a preto nie je možné hypotézu „pôsobenie faktoru počasia priamo ovplyvňuje počet výpadkov exteriérových elektronických zariadení monitorujúcich dopravu v meste Zlín.“ pre dané vstupné dáta potvrdiť.



Obr. 46. Hodnoty Pearsonovho korelačného koeficientu pre filtrovaný vstupný súbor dát.

Zdroj: vlastná tvorba.

ZÁVER

Pre potreby tejto práce bola s jej zadávateľom, firmou Incinity, s. r. o. k overeniu stanovená hypotéza na základe dostupného vzorku dát z oblasti dopravy v nasledujúcom znení: „Pôsobenie faktoru počasia priamo ovplyvňuje počet výpadkov exteriérových elektronických zariadení monitorujúcich dopravu v meste Zlín.“

Potvrdenie, prípadne nepotvrdenie tejto hypotézy môže byť podkladom pre dôležité manažérske rozhodnutia, posvätením alebo spochybnením správnosti technológií a postupov používaných pri inštalácii zariadení monitorujúcich dopravu v mestách, prípadne známkou kvality inštalovaných monitorovacích zariadení.

Jedným z hlavných cieľov práce bolo poskytnúť dôkaz, ktorý by viedol k potvrdeniu, prípadne nepotvrdeniu tejto hypotézy. Postup obsahujúci konkrétne štatistické metódy a postupy viedol k záveru, že počasia nemá na výpadkovosť týchto zariadení žiadny vplyv.

Najzaujímavejšou a najkritickejšou časťou procesu bolo pravdepodobne čistenie dát – práca s nerelevantnými a extrémne odľahlými hodnotami mohla nepriaznivo ovplyvniť ďalší proces spracovania dát a v konečnom dôsledku i výsledný korelačný koeficient.

Pre vierohodnejšie výsledky by bolo vhodné zaniest', prípadne dodatočne zaznamenávať do monitorovaných údajov i informáciu o dôvode výpadku, ak je známy. Dočasne zvýšená výpadkovosť by napríklad mohla byť zaznamenaná v súvislosti s úkonmi, ako je inštalácia, kontrola či servis zariadenia, prípadne úprava softvéru monitorujúceho tieto výpadky. Taktiež by bolo možné opatrit' vybrané zariadenia meteorostanicami, ktoré by monitorovali meteorologické údaje priamo v danom mieste inštalácie zariadenia.

Postup analýzy dát využitý v tejto diplomovej práci bol zvolený tak, aby ho bolo v budúcnosti možné aplikovať pri skúmaní závislostí iných potencionálne závislých veličín z oblasti dopravných dát.

ZOZNAM POUŽITEJ LITERATURY

- [1] CUESTA, Hector. Analýza dat v praxi. Brno: Computer Press, 2015, 296 s. ISBN 978-80-251-4361-2.
- [2] BEYER, Mark a Douglas LANEY. The Importance of 'Big Data': A Definition [online]. [cit. 2019-05-21]. Dostupné z: <https://www.gartner.com/en/documents/2057415>
- [3] HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. Big Data a NoSQL databáze. Praha: Grada, 2015, 281 s. Profes!onal. ISBN 978-80-247-5466-6.
- [4] ROUSE, Margaret. What is big data? [online]. [cit. 2019-05-21]. Dostupné z: <https://searchdatamanagement.techtarget.com/definition/big-data>
- [5] How many V's are in big data? [online]. [cit. 2019-05-21]. Dostupné z: <https://www.quora.com/How-many-V%E2%80%99s-are-in-big-data>
- [6] KOSEČEK, Miroslav. Čo je to big data? [online]. [cit. 2019-05-21]. Dostupné z: <https://dennikn.sk/blog/278593/co-je-to-big-data/>
- [7] SCHEPS, Swain. Business intelligence for dummies. Hoboken, NJ: Wiley, c2008. ISBN 978-0-470-12723-0.
- [8] Waze [online]. [cit. 2019-05-21]. Dostupné z: <https://www.waze.com/sk/waze>
- [9] Waze. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-05-21]. Dostupné z: <https://cs.wikipedia.org/wiki/Waze>
- [10] NORMANDEAU, Kevin. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity [online]. 12.9.2013 [cit. 2019-05-21]. Dostupné z: <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
- [11] RIDER, Fremont. The Scholar And the Future of the Research Library: a Problem And Its Solution. New York city: Hadham press 1944.
- [12] PRESS, Gil. A Very Short History Of Big Data [online]. 9.3.2013 [cit. 2019-05-21]. Dostupné z: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#2ca87cb165a1>
- [13] KENNEDY, John. When Woman is Boss. Colliers [online]. 30.1.1926 [cit. 2019-05-21]. Dostupné z: <http://www.tfcbooks.com/tesla/1926-01-30.htm>

- [14] MARR, Bernard. A brief history of big data everyone should read [online]. 25.2.2015 [cit. 2019-05-21]. Dostupné z: <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/>
- [15] Incinity s. r. o. [online]. [cit. 2019-05-21]. Dostupné z: <https://www.incinity.cz/>
- [16] Invipo [online]. [cit. 2019-05-21]. Dostupné z: <https://www.invipo.com/sk/>
- [17] MongoDB [online]. [cit. 2019-05-21]. Dostupné z: <https://www.mongodb.com/>
- [18] MARR, Bernard. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read [online]. 21.5.2018 [cit. 2019-05-21]. Dostupné z: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6f005fb960ba>
- [19] The World: Population [online]. [cit. 2019-05-21]. Dostupné z: <http://www.geoba.se/>
- [20] Data Never Sleeps 6.0 [online]. [cit. 2019-05-21]. Dostupné z: <https://www.domo.com/learn/data-never-sleeps-6>
- [21] NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. Business Intelligence: jak využít bohatství ve vašich datech. 1. Vyd. Praha: Grada, 2005, 254 s. Management v informační společnosti. ISBN 80-247-1094-3.
- [22] POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business Intelligence v podnikové praxi. 1. Vyd. Praha: Professional Publishing, 2012, 276 s. ISBN 978-80-7431-065-2.
- [23] LACKO, Ľuboslav. Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby. Vyd. 1. Brno: Computer Press, 2009, 456 s. ISBN 978-80-251-2887-9.
- [24] Co jsou nástroje business intelligence (BI)? [online]. [cit. 2019-05-21]. Dostupné z: <https://azure.microsoft.com/cs-cz/overview/what-are-business-intelligence-tools/>
- [25] PRETT, Mary. What is BI? Business intelligence strategies and solutions [online]. 1.9.2017 [cit. 2019-05-21]. Dostupné z: <https://www.cio.com/article/2439504/business-intelligence-definition-and-solutions.html>
- [26] ROUSE, Margaret. What is business intelligence (BI)? [online]. [cit. 2019-05-21]. Dostupné z: <https://searchbusinessanalytics.techtarget.com/definition/business-intelligence-BI>
- [27] Gartner Magic Quadrant & Critical Capabilities [online]. [cit. 2019-05-21]. Dostupné z: <https://www.gartner.com/en/research/magic-quadrant>
- [28] Finding the Right Business Intelligence Tool [online]. [cit. 2019-05-21]. Dostupné z: <https://bi-survey.com/business-intelligence-software-comparison>

- [29] Google Data Studio [online]. [cit. 2019-05-21]. Dostupné z: www.datastudio.google.com
- [30] BOBRIAKOV, Igor. A Comparative Analysis of Top 6 BI and Data Visualization Tools in 2018 [online]. [cit. 2019-05-21]. Dostupné z: <https://www.kdnuggets.com/2018/02/comparative-analysis-top-6-bi-data-visualization-tools-2018.html/>
- [31] QlikView [online]. [cit. 2019-05-21]. Dostupné z: <https://www.qlik.com/us/>
- [32] Sisense [online]. [cit. 2019-05-21]. Dostupné z: <https://www.sisense.com/>
- [33] Tableau [online]. [cit. 2019-05-21]. Dostupné z: <https://www.tableau.com/>
- [34] Microsoft Power BI [online]. [cit. 2019-05-21]. Dostupné z: <https://powerbi.microsoft.com/en-us/>
- [35] Biportal [online]. [cit. 2019-05-21]. Dostupné z: <https://biportal.cz/power-bi-free-10-duvodu-proc-ho-pouzivat/>
- [36] Stitch [online]. [cit. 2019-05-21]. Dostupné z: <https://www.stitchdata.com/>
- [37] Studio 3T [online]. [cit. 2019-05-21]. Dostupné z: <https://studio3t.com/>
- [38] Český hydrometeorologický ústav [online]. [cit. 2019-05-21]. Dostupné z: <http://portal.chmi.cz/>.
- [39] ULDRICH, Miloš. Čistota dat – z provozních na analytická [online]. [cit. 2019-08-21]. Dostupné z: <https://www.systemonline.cz/business-intelligence/cistota-dat-z-provoznich-na-analyticka.htm>
- [40] Centrum informačních systémů VFU Brno [online]. [cit. 2019-08-21]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn2/extremy.htm>
- [41] Kalendář Beda [online]. [cit. 2019-08-21]. Dostupné z: <https://kalendar.beda.cz/kdy-zacina-jaro-letopodzim-zima>
- [42] Excel Town [online]. [cit. 2019-08-21]. Dostupné z: <https://exceltown.com/en/tutorials/power-bi/powerbi-com-and-power-bi-desktop/dax-query-language-for-power-bi-and-power-pivot/pearson-correlation-coefficient/>
- [43] Wikipedie [online]. [cit. 2019-08-21]. Dostupné z: https://sk.wikipedia.org/wiki/Bravaisov-Pearsonov_korela%C4%8Dn%C3%BD_koeficient
- [44] HINKLE, Dennis E., William WIERSMA a Stephen G. JURIS. Applied statistics for the behavioral sciences. 5th ed. Boston: Houghton Mifflin, c2003. ISBN 978-0618124053.

ZOZNAM POUŽITÝCH SYMBOLOV A SKRATIEK

3V	Volume, Variety, Veracity
API	Application Programming Interface
B	Byte
BA	Business Analytics
BARC	Business Application Research Centre
BI	Business Intelligence
BSON	Binary JSON
CRM	Customer Relation Management
DSN	Data Source Name
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
GB	Gigabyte
GIGO	Garbage In, Garbage Out
GPS	Global Positioning System
GUI	Graphic User Interface
IBM	International Business Machines
IDE	Integrated Development Environment
IT	Information Technology
JSON	JavaScript Object Notation
MB	Megabyte
MPP	Massive Parallel Processing
NoSQL	Not only SQL
ODBC	Open Database Connectivity
OLAP	On-line Analytical Processing
PB	Petabyte
SCM	Supply Chain Management
SQL	Structured Query Language
TB	Terabyte
XML	eXtensible Markup Language
ZB	Zettabyte

ZOZNAM OBRÁZKOV

<i>Obr. 1. Vlastnosti veľkých dát. [5]</i>	9
<i>Obr. 2. Logo spoločnosti Incinity, s. r. o. [15]</i>	12
<i>Obr. 3. Logo MongoDB. [17]</i>	13
<i>Obr. 4. Množstvo dát vygenerovaných za jednu minútu roku 2018 známymi internetovými službami. [20]</i>	15
<i>Obr. 5. Porovnanie poskytovateľov BI riešení spoločnosťou Gartner. [27]</i>	18
<i>Obr. 6. Porovnanie poskytovateľov BI riešení inštitútom BARC. [28]</i>	20
<i>Obr. 7. Logo Google Data Studio. [29]</i>	21
<i>Obr. 8. Logo QlikView. [31]</i>	22
<i>Obr. 9. Logo Sisense. [32]</i>	23
<i>Obr. 10. Logo Tableau. [33]</i>	23
<i>Obr. 11. Logo Microsoft Power BI. [34]</i>	24
<i>Obr. 12. Logo platformy Invipo. [16]</i>	26
<i>Obr. 13. Inštalácia MongoDB. Zdroj: vlastná tvorba.</i>	30
<i>Obr. 14. Obsah priečinka MongoDB. Zdroj: vlastná tvorba.</i>	31
<i>Obr. 15. Navigácia do priečinka MongoDB a spustenie mongod.exe. Zdroj: vlastná tvorba.</i>	31
<i>Obr. 16. Cesta k nastaveniu novej užívateľskej premennej. Zdroj: vlastná tvorba.</i> ...	32
<i>Obr. 17. Založenie novej užívateľskej premennej. Zdroj: vlastná tvorba.</i>	33
<i>Obr. 18. Spustenie programu mongo.exe. Zdroj: vlastná tvorba.</i>	33
<i>Obr. 19. Ukážka dát vo formáte JSON. Zdroj: vlastná tvorba.</i>	34
<i>Obr. 20. Ukážka upravených dát vo formáte JSON. Zdroj: vlastná tvorba.</i>	35
<i>Obr. 21. Ukážka dát vo formáte BSON. Zdroj: vlastná tvorba.</i>	35
<i>Obr. 22. Ukážka behu príkazu mongorestore. Zdroj: vlastná tvorba.</i>	36
<i>Obr. 23. Ukážka pripojenia MongoDB na Power BI prostredníctvom nástroja Stitch. [36]</i>	37
<i>Obr. 24. vytvorenie nového systémového zdroja dát. Zdroj: vlastná tvorba.</i>	38
<i>Obr. 25. Konfigurácia systémového zdroja dát. Zdroj: vlastná tvorba.</i>	39
<i>Obr. 26. Odozva programu mongosqld.exe na pokus o pripojenie MongoDB ODBC ovládača. Zdroj: vlastná tvorba.</i>	39
<i>Obr. 27. Pripojenie na ODBC zdroj dát v prostredí Power BI. Zdroj: vlastná tvorba.</i>	40

Obr. 28. Chyba pripojenia ODBC zdroja dát. Zdroj: vlastná tvorba.	40
Obr. 29. Úprava konfigurácie systémového zdroja dát. Zdroj: vlastná tvorba.	41
Obr. 30. Odozva programu mongosql.exe na zmenu konfigurácie MongoDB ODBC ovládača. Zdroj: vlastná tvorba.	42
Obr. 31. Chyba pripojenia ODBC zdroja dát. Zdroj: vlastná tvorba.	42
Obr. 32. Logo Studio 3T. [37]	43
Obr. 33. Nové pripojenie databázy MongoDB v programe Studio 3T. Zdroj: vlastná tvorba.	44
Obr. 34. Prostredie programu Studio 3T. Zdroj: vlastná tvorba.	45
Obr. 35. Visual Query Builder programu Studio 3T. Zdroj: vlastná tvorba.	46
Obr. 36. Možnosti exportu dát v programe Studio 3T. Zdroj: vlastná tvorba.	47
Obr. 37. Možnosti exportu dát v programe Studio 3T. Zdroj: vlastná tvorba.	47
Obr. 38. Graf počtu hlásených výpadkov (status: Disconnected) pre jednotlivé dni roku 2017. Zdroj: vlastná tvorba.	49
Obr. 39. Graf počtu hlásených výpadkov (status: Disconnected) pre jednotlivé dni roku 2017 po vyčistení od extrémnych hodnôt. Zdroj: vlastná tvorba.	50
Obr. 40. Postup prevodu počtu výpadkov na stupnicu 1-10 v programe Power BI. Zdroj: vlastná tvorba.	51
Obr. 41. Stupne výpadkovosti zariadení (status: Disconnected) pre jednotlivé dni roku 2017 po vyčistení od extrémnych hodnôt a po prevedení na stupnicu 1-10. Zdroj: vlastná tvorba.	52
Obr. 42. Stupne extrémnosti počasia pre jednotlivé letné a zimné dni roku 2017. Zdroj: vlastná tvorba.	54
Obr. 43. Vzorec pre výpočet Pearsonovho korelačného koeficientu [42].	54
Obr. 44. Medzivýpočty pre výpočet Pearsonovho korelačného koeficientu v prostredí Microsoft Power BI. Zdroj: vlastná tvorba.	55
Obr. 45. Hodnoty Pearsonovho korelačného koeficientu pre vyčistený nefiltrovaný vstupný súbor dát. Zdroj: vlastná tvorba.	56
Obr. 46. Hodnoty Pearsonovho korelačného koeficientu pre filtrovaný vstupný súbor dát. Zdroj: vlastná tvorba.	57

ZOZNAM TABULIEK

<i>Tab. 1. Relačné vs. NoSQL databázy – predpoklady o dátach. [3]</i>	13
<i>Tab. 2. Business Intelligence vs. Business Analytics riešenia. [26]</i>	17
<i>Tab. 3. Interpretácia hodnôt Pearsonovho korelačného koeficientu [44].</i>	55

ZOZNAM PRÍLOH

- P I Invipo Status – ukážka dát
- P II Invipo Event – ukážka dát
- P III Meteorologické údaje

PRÍLOHA P I: INVIPO STATUS – UKÁŽKA DÁT

```
{
  "_id" : ObjectId("58e3381cccd85c05fcf9961c"),
  "data" : {
    "controllerId" : NumberInt(18),
    "ocitErrorCode" : "Success",
    "manufacturerErrorCode" : "",
    "ocitOperatingState" : "On",
    "controllerDateTime" : "2017-04-04T08:07:29+02:00",
    "modifiedOn" : "0001-01-01T00:00:00",
    "planNo" : NumberInt(26),
    "isEmergencyPlan" : false,
    "planModifications" : [
      {
        "modificationType" : "TA",
        "state" : "On",
        "projectModificationId" : NumberInt(0),
        "id" : NumberInt(0),
        "no" : NumberInt(0)
      },
      {
        "modificationType" : "PT",
        "state" : "On",
        "projectModificationId" : NumberInt(0),
        "id" : NumberInt(0),
        "no" : NumberInt(0)
      },
      {
        "modificationType" : "PM",
        "state" : "On",
        "projectModificationId" : NumberInt(0),
        "id" : NumberInt(0),
        "no" : NumberInt(0)
      }
    ],
    "sectionsState" : [
      ],
    "controlMode" : "LocalSchedule"
  },
  "ufid" : "RS4-CR-ZLI-129",
  "utype" : "CrossRs4StatusObject",
  "timestamp" : ISODate("2017-04-04T06:07:24.291+0000"),
  "status" : "Ok"
}
// -----
{
  "_id" : ObjectId("58e33862ccd85c05fcf996ab"),
  "data" : {
    "doorOpen" : false
  },
  "ufid" : "TW-CZ-ZLI-PS05",
  "utype" : "TelWayStatusObject",
  "timestamp" : ISODate("2017-04-04T06:07:33.650+0000"),
  "status" : "Disconnected"
}
// -----
```

```

{
  "_id" : ObjectId("58e3389eccd85c05fcf9971e"),
  "data" : {
    "doorOpen" : false
  },
  "ufid" : "TW-CZ-ZLI-PS05",
  "utype" : "TelWayStatusObject",
  "timestamp" : ISODate("2017-04-04T06:09:33.713+0000"),
  "status" : "Ok"
}
// -----
{
  "_id" : ObjectId("58e34096ccd85c05fcf9a7e0"),
  "data" : {
    "doorOpen" : false
  },
  "ufid" : "TW-CZ-ZLI-PS05",
  "utype" : "TelWayStatusObject",
  "timestamp" : ISODate("2017-04-04T06:42:33.982+0000"),
  "status" : "Disconnected"
}
// -----
{
  "_id" : ObjectId("58e34096ccd85c05fcf9a7eb"),
  "data" : {
    "doorOpen" : false
  },
  "ufid" : "TW-CZ-ZLI-PS05",
  "utype" : "TelWayStatusObject",
  "timestamp" : ISODate("2017-04-04T06:43:34.560+0000"),
  "status" : "Ok"
}
// -----
{
  "_id" : ObjectId("58e34b9bccd85c05fcf9c088"),
  "data" : {
    "controllerId" : NumberInt(18),
    "ocitErrorCode" : "InternalErrorWithoutShutdown",
    "manufacturerErrorCode" : "",
    "ocitOperatingState" : "On",
    "controllerDateTime" : "2017-04-04T09:30:45+02:00",
    "modifiedOn" : "0001-01-01T00:00:00",
    "planNo" : NumberInt(23),
    "isEmergencyPlan" : false,
    "planModifications" : [
      {
        "modificationType" : "TA",
        "state" : "On",
        "projectModificationId" : NumberInt(0),
        "id" : NumberInt(0),
        "no" : NumberInt(0)
      },
      {
        "modificationType" : "PT",
        "state" : "On",
        "projectModificationId" : NumberInt(0),
        "id" : NumberInt(0),
        "no" : NumberInt(0)
      }
    ],
  },
}

```



```
    {
      "modificationType" : "PM",
      "state" : "On",
      "projectModificationId" : NumberInt(0),
      "id" : NumberInt(0),
      "no" : NumberInt(0)
    }
  ],
  "sectionsState" : [

  ],
  "controlMode" : "LocalSchedule"
},
"ufid" : "RS4-CR-ZLI-129",
"utype" : "CrossRs4StatusObject",
"timestamp" : ISODate("2017-04-04T07:30:35.925+0000"),
"status" : "Warn"
}
```

PRÍLOHA P II: INVIPO EVENT – UKÁŽKA DÁT

```
{
  "_id" : ObjectId("58b68bddccd85c15286db07e"),
  "data" : {
    "contents" : [
      {
        "content" :
"iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAYAAABZY7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
v8YQUAAAAJcEhZcwAADsMAAA7DAcdvqGQAAAEsSURBVfH7Y5RDsIwDEO5/6WBT1i1XCdhtPzF0l0c1
23a451nk2JlM7GymVh5wXE0cR7Pu/fURzvh3R14TPZi5Ef4Jl1i5vLSzc89uUecdU4HH1F7dEqxcPrCz
c49umOrdzj27cVeym2Dl8oFvdo7eoh3h03y0c89u3BWN+e+aDlctLd3bu1Y60qb7aMdW7G5PdBCuXD+z
s3LNb1LMdU/3dW4KVFxx3u7sj6qOdw8/wvXKcyI/wTbCymVjZTKxsJlZug1S0b5WPutsPYuUW/LPomN
ozp177iLsdxspj4McxTbMbyZz2EXc7jJVH4J+OurpoRl3nH7ByG/1h3rmi2bVR7Afxsot3M+y4+7cL
30E+2Gs3EKjnp/lW+Wj7vaDWNlMrGwmVjYTK5uLx/MFFD02SWhvbXsAAAAASUVORK5CYII=",
        "duration" : NumberInt(3)
      },
      {
        "content" :
"iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAYAAABZY7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
v8YQUAAAAJcEhZcwAADsMAAA7DAcdvqGQAAAEQSURBVfH7Y9RCsMwDEN7/0t3y8BECM10u+TPgkfk5
5aQ65u7SZGymUjZTKQswTg/gp6/i50jvPpXzdgR5xeQMoUvwxk7EsGZ96sz9mpX+QWkTMkuc7sIzrxf
nbGr3QjvR9A9QMqU7DI0enXyHmcM79wc3Z0vkDI1u8zt0Efcnmfs1RzdnS+QMoUvwxk7gj7i9tWMPdu
5/hApSzD0j6Dn7/6dI+xxrvwCUjYTKZuJlM1Eyi1w1J4de9fVfAgpt1A9w03Rcx9Ru4NIuQVMts8c9x
G104iU21GPCed01/k8jJTbUY8J586qj8R8ECm3gA/ArtybPhLzQaTcRsTtKu+6mg8hZTORsplI2Uykb
H5c9weLycGvoainSQAAAABJRU5ErkJggg==",
        "duration" : NumberInt(3)
      }
    ]
  },
  "ufid" : "TW-CZ-ZLI-PS07",
  "utype" : "VmsNewContentRequestEvent",
  "timestamp" : ISODate("2017-03-01T08:52:45.516+0000"),
  "severity" : "Info"
}
// -----
{
  "_id" : ObjectId("58b68bddccd85c15286db080"),
  "data" : {
    "contents" : [
      {
        "content" :
"iVBORw0KGgoAAAANSUHEUgAAAGAAAAAgCAYAAADtwH1UAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
v8YQUAAAAJcEhZcwAADsMAAA7DAcdvqGQAAAKBSURBVGHd7ZELaiQxEENz/0tn17CCZ42q7G4vkzC0Q
FTpqdzk8/VX349/1BE+fp8jfPw+R/h44SH0A79CKjEpcd2nd86rTKWeFtck1yo+x06qT9//8wz8oyeZ
e9dV070mW1zT91X3CzwD/wFPMveq03SeMveu4+7uus6UsyqTDZHDM/DDnUx5V2WJvXiVuXcdd7cr3ST
zVrvmTt94Bv7oSua+yto1na+yvpPU0V3Xme+0a2qnnCsHz8CPTzL3rqv2Lms6v9rtmu+0a/qe3PSvkE
rd1S5rzLFG/YrRlV8iF1niqy7oXhnjvCxefFHPHGEj83PP+Bz/Qqpig+R+52mK/H0NmXudMU7842Ld
+qd0a/2lM0z8GNm7rTE7P1u5r7qVryyVHU7zLnvQ6kLnkF3XHUSs/e7mXvqhrwfIttxep0Y+FDHfB9K
XfAMumMXeZreM1PeVV17Na84vUmMvJrV7rPwDLrjqioXqt4z91XWXs0rTm8SI6/mah9SDp6BHnzp8m
lq191711X7bv2N55pdnf2IeXgV0hVfIjC706z5Jx5xTv7G8901ZFXe8rmCB+/zxEuvfivTtat5o51q+
n8gxxh6zt/hP/xxv0HeAb+S3rumFTxoZ30Ra791PyWi3fqntmv9pTNM5CYve94t6+6ZN5wP7HEzN5d9
eS+d6UueAYSs/cd7/ZV5/be84n5LYo33nfM96HUBc9Axx59T53kN9SVjrlid119K3Gxa1a7z8Iz4LHk
mWJ3ZU+ZT11id119K3Gxaq72IeXgGfBYck6La3b7qrvL75jfqbE7uxDysEz8GN1Tbc41fqUXTtc+6n
9WxIZuxWv9pTNET5+i7++/wC8G1qjWYJz8gAAAABJRU5ErkJggg=="
      }
    ]
  },
}
```

```

    "ufid" : "TW-CZ-ZLI-PS03",
    "utype" : "VmsNewContentRequestEvent",
    "timestamp" : ISODate("2017-03-01T08:52:45.595+0000"),
    "severity" : "Info"
}
// -----
{
    "_id" : ObjectId("58b68bddccd85c15286db082"),
    "data" : {
        "contents" : [
            {
                "content" :
                "iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAAABZy7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
                v8YQUAAAAJcEhZcwAADsMAAA7DAcdivqGQAAAEuSURBVFH7ZPBDSIwDEP3/z896IRVy3VSxsItlp6aP
                HeoF453zibFymZiT0x8oLjHOI87rvv1Ec7x/UMPE72S0RHuB0sXD56sv0cddHM004FHqf0uy7ByuUH
                nuw8Rx109W7n0et4VrJ0sHL5gW92jnbRjnAPH+08Zx3Pisbd+WD18tGdnfdjhmn+t2OU73rmKwTrFx
                +4Mn0c9ZFc7bjVH+3S7Dygu06uzuiPto5fIf7neNEfoQ7wcpmYmUzsbKZWFkC/7c1fA+90vXR7PZCrH
                wMwjv3StSz13nEdcVYwQI/msN3tM+czioUk8bKEqJH0w8XndGs5x+wsoTo0c7DReduHsFejJU18K0j2
                blf5hHsxVhZgj4aYcfdzkez2wuxsplY2UysbCZWNhFH+QLADSpVZ9YByQAAAABJRu5ErkJggg==",
                "duration" : NumberInt(3)
            },
            {
                "content" :
                "iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAAABZy7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
                v8YQUAAAAJcEhZcwAADsMAAA7DAcdivqGQAAAEuSURBVFH7ZBBCsMwDATz/0+ndWHxIkY2bmTIwYJBu
                6OLk+s792EIyJLa+Kbby0FZCv0Ici8F5RjXzL43ypnz71m9TXTeH4ByifgYde3oRz1z3ilrR18AyiXi
                Y9S1ox/lzHlX1p75h6BcIs7M60Y5c941fvcb+QegXCJ71HvPsxs578raDrmHoFwie1T03r0c0e9Zp14
                AyiWyR0XvPWYf8nLyWfdcBMpDB+Whg/LQVmgJrvNfJapbwJlCf4Bnkcu+pjb0G0jKEvwD/Dsrs3Ixd
                yGbhtBWYYmu412luPeDMoS/AM8R5ftW6jvhGUJfgHeCb3T26jvhGUZWiy28xnmfomUB46KA8d1IcOy
                sOP6/4AEhGI6N35yMIAAAAASUVORK5CYII=",
                "duration" : NumberInt(3)
            }
        ]
    },
    "ufid" : "TW-CZ-ZLI-PS08",
    "utype" : "VmsNewContentRequestEvent",
    "timestamp" : ISODate("2017-03-01T08:52:45.829+0000"),
    "severity" : "Info"
}
// -----
{
    "_id" : ObjectId("58b68bddccd85c15286db084"),
    "data" : {
        "contents" : [
            {
                "content" :
                "iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAAABZy7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw
                v8YQUAAAAJcEhZcwAADsMAAA7DAcdivqGQAAAFuSURBVFH7ZPRbsMwDAP7/z/dzQU4E8zJqTcN24MFH
                ESegiB96ONznoc1KcY4/sddp79Y1Bus/ODd579B6C84D8qsw89k92zo/Hu98pnJ6eunX4Bygv+Is+O
                e011UyanrJ24V9Z0n107/QKUF/xFniunobvn1SNfde301JW1b0B5wV/mmbq73JkdTbqqe6ZbTvXsDSg
                v+As9U0+noRuRd3Vt9z1+y2dXfQHKF/6Sd7LjXkM3cp69a2d25LXTV30Byi806fKuoWey3Y01W3V02
                unr/oClIcJysME5WGCsgX/n+f4c7qnS19l6o2g/DEa735Pqrv7zGPo1gzKFvyjffyZvK9c5jF0awZ1C
                9VhK5erdpVz/wIoW6g+mrxcte/yGPVmULBgH11lct/JY9SbQd1CfrTGnd/ufJWpN4LyMEF5mKA8TFAe
                XjyeH5c7BnnBF02LAAAAAE1FTkSuQmCC",
                "duration" : NumberInt(3)
            },
            {
                "content" :
                "iVBORw0KGgoAAAANSUHEUgAAAEgAAAAAYCAAABZy7uwAAAAAXNSR0IArs4c6QAAARnQU1BAACxjw

```

```

v8YQUAAAAJcEhZcwAADsMAAA7DAcdvqGQAAAEcSURBVFhH7ZNBgMxDATz/09n1wvCTV0SnYn2pobCr
dIE+5LXb95DCcphg3LYoDyiyfyKev8uTg95+i3N2pXMX4CyxC/TwbsS0dn3t7P20+7kL0BZUI2W7SI6
+/521k67Fd+vqPsAlCXVZR71dPpeZ43vsjl6dj4AZU11WbZTH8n2Pms/zdGz8wEoS/wynbUr6iPZ/jR
rr3ZZ/xCURzSZX1Hv3307R9zrfPIXoBw2KIcNymGDsgX/30fU6e7ks05zIyi/JqIz9cq5975Cu2ZQtp
A9mnykct5XaNcMyhbo0eTUZ2fW/fwHULaQPZp8uOw89ZWYm0HZgj466+Se9JWYm0HZgj86ok53J591m
htBOWxQDhuUwwb18Mfr/QN/w9CgRB97vwAAAABJRU5ErkJggg==",
    "duration" : NumberInt(3)
  }
]
},
"ufid" : "TW-CZ-ZLI-PS01",
"utype" : "VmsNewContentRequestEvent",
"timestamp" : ISODate("2017-03-01T08:52:45.892+0000"),
"severity" : "Info"
}
// -----
{
  "_id" : ObjectId("58b68bdfccd85c15286db087"),
  "data" : {
    "rule" : "vms-start-scenarios",
    "pid" : NumberInt(8876),
    "started" : "2017-02-20T13:51:51.782Z",
    "timestamp" : "2017-03-01T08:52:47.439Z",
    "runtime" : {
      "receivedEvents" : NumberInt(1046248),
      "runs" : NumberInt(1)
    },
    "process" : {
      "memory" : {
        "rss" : NumberInt(49168384),
        "heapTotal" : NumberInt(32542720),
        "heapUsed" : NumberInt(26888520)
      },
      "cpu" : {
        "user" : NumberInt(226993000),
        "system" : NumberInt(36581000)
      }
    },
    "os" : {
      "memory" : {
        "free" : NumberInt(1477988352),
        "total" : NumberLong(6441914368)
      },
      "uptime" : 6023656.3145583
    }
  },
  "ufid" : "SERVICE-INVIPO-RE",
  "utype" : "RuleWorkerStatusEvent",
  "timestamp" : ISODate("2017-03-01T08:52:47.439+0000"),
  "severity" : "Info"
}

```

PRÍLOHA P III: METEOROLOGICKÉ ÚDAJE

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
01.01.2017	-6,0	1	-11,2	9	92	9	0,0	1	0,0	1
02.01.2017	-4,7	1	-7,1	7	94	10	0,2	1	0,0	1
03.01.2017	1,7	1	-6,6	7	84	8	0,9	1	0,5	1
04.01.2017	3,3	1	-0,8	5	85	8	0,0	1	0,6	1
05.01.2017	1,9	1	-6,2	7	64	4	0,0	1	1,6	2
06.01.2017	-5,5	1	-9,1	8	63	4	0,0	1	2,8	2
07.01.2017	-7,2	1	-17,2	10	65	5	1,2	1	5,7	4
08.01.2017	-7,5	1	-11,3	9	69	5	0,0	1	2,1	2
09.01.2017	-3,6	1	-10,1	8	73	6	0,0	1	0,2	1
10.01.2017	-6,5	1	-12,3	9	74	6	0,0	1	1,3	1
11.01.2017	-4,4	1	-14,8	10	68	5	3,5	2	7,3	5
12.01.2017	4,5	1	-6,1	7	80	7	1,9	1	2,8	2
13.01.2017	4,9	1	-0,4	5	73	6	0,0	1	0,2	1
14.01.2017	2,4	1	-6,2	7	85	8	1,3	1	0,1	1
15.01.2017	2,0	1	-7,5	7	82	8	0,0	1	3,1	3
16.01.2017	-2,5	1	-15,0	10	83	8	0,0	1	5,0	4
18.01.2017	-1,7	1	-8,1	7	75	6	0,0	1	6,7	5
19.01.2017	-5,9	1	-13,3	9	81	7	0,0	1	5,5	4
20.01.2017	-3,8	1	-16,7	10	77	7	0,0	1	5,6	4
21.01.2017	-1,6	1	-11,2	9	80	7	0,0	1	7,7	6
22.01.2017	-0,9	1	-11,3	9	75	6	0,0	1	5,8	4
23.01.2017	-3,0	1	-7,9	7	83	8	0,0	1	0,0	1
24.01.2017	-3,9	1	-5,1	6	76	7	0,0	1	0,0	1
25.01.2017	-0,1	1	-4,1	6	78	7	0,0	1	0,0	1
26.01.2017	0,9	1	-4,6	6	75	6	0,0	1	0,0	1
27.01.2017	-1,6	1	-10,1	8	79	7	0,0	1	7,5	6
28.01.2017	-6,1	1	-7,8	7	91	9	0,0	1	0,0	1
29.01.2017	-6,0	1	-7,1	7	93	9	0,0	1	0,0	1
30.01.2017	-6,9	1	-7,8	7	90	9	0,6	1	0,0	1
01.02.2017	-0,2	1	-4,4	6	94	10	3,5	2	0,0	1
02.02.2017	0,5	1	-1,5	5	95	10	0,0	1	0,0	1
03.02.2017	2,3	1	-0,4	5	99	10	0,4	1	0,0	1
04.02.2017	2,3	1	0,6	5	98	10	0,0	1	0,0	1
05.02.2017	5,7	1	-0,4	5	90	9	0,5	1	3,2	3
06.02.2017	3,2	1	-0,9	5	87	8	1,0	1	0,0	1
07.02.2017	2,3	1	-1,5	5	73	6	0,0	1	1,4	1
08.02.2017	-1,2	1	-4,8	6	73	6	0,0	1	0,0	1
09.02.2017	-2,7	1	-4,9	6	75	6	0,0	1	0,0	1

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
10.02.2017	1,8	1	-4,3	6	81	7	0,0	1	0,1	1
11.02.2017	1,7	1	-2,4	6	72	6	0,0	1	0,0	1
12.02.2017	2,8	1	0,2	5	69	5	0,0	1	0,0	1
13.02.2017	0,9	1	-5,0	6	78	7	0,0	1	8,1	6
14.02.2017	3,5	1	-9,0	8	77	7	0,0	1	8,9	6
15.02.2017	7,5	1	-4,9	6	68	5	0,0	1	9,4	7
16.02.2017	5,9	1	-3,9	6	83	8	0,0	1	7,2	5
17.02.2017	1,6	1	-2,4	6	98	10	2,2	2	0,0	1
18.02.2017	6,0	1	-0,9	5	87	8	0,0	1	1,9	2
19.02.2017	3,3	1	-1,3	5	94	10	0,0	1	0,5	1
20.02.2017	5,8	1	-1,9	5	89	9	2,5	2	1,0	1
21.02.2017	9,8	2	3,0	4	87	8	1,3	1	1,0	1
22.02.2017	12,2	2	-1,2	5	91	9	1,4	1	0,0	1
23.02.2017	12,4	2	5,9	3	77	7	0,0	1	0,0	1
24.02.2017	11,0	2	3,1	4	52	2	0,0	1	2,6	2
25.02.2017	7,3	1	-3,0	6	60	4	0,0	1	9,0	7
26.02.2017	7,0	1	0,2	5	66	5	0,0	1	1,4	1
27.02.2017	11,5	2	-2,5	6	72	6	0,0	1	6,6	5
28.02.2017	10,6	2	3,3	4	69	5	0,0	1	2,7	2
01.03.2017	10,8	2	2,2	4	65	5	1,1	1	2,3	2
02.03.2017	13,9	3	0,0	5	62	4	0,2	1	4,1	3
03.03.2017	12,6	2	-1,5	5	64	4	0,0	1	9,7	7
04.03.2017	18,6	5	4,4	3	54	3	0,0	1	8,0	6
05.03.2017	14,1	3	5,9	3	68	5	4,1	2	0,9	1
09.03.2017	12,0	2	1,8	4	84	8	3,3	2	0,8	1
10.03.2017	8,8	1	3,5	4	78	7	0,0	1	1,6	2
11.03.2017	10,7	2	2,2	4	63	4	0,1	1	4,4	3
12.03.2017	9,2	1	0,7	4	62	4	0,0	1	4,7	4
13.03.2017	11,7	2	0,4	5	62	4	0,0	1	10,1	7
14.03.2017	13,3	3	-0,8	5	65	5	0,0	1	4,3	3
15.03.2017	13,7	3	3,5	4	66	5	0,0	1	1,2	1
16.03.2017	15,1	3	2,2	4	55	3	0,0	1	7,6	6
17.03.2017	17,0	4	-0,7	5	65	5	4,1	2	7,7	6
18.03.2017	12,1	2	4,2	3	82	8	3,5	2	0,2	1
19.03.2017	9,7	1	1,6	4	81	7	3,1	2	1,1	1
20.03.2017	16,3	4	6,1	3	64	4	0,0	1	2,3	2
21.03.2017	16,0	4	5,0	3	70	6	0,6	1	5,6	4
22.03.2017	12,4	2	5,3	3	83	8	0,4	1	0,0	1
23.03.2017	11,4	2	4,7	3	84	8	0,0	1	0,7	1
24.03.2017	13,5	3	5,1	3	68	5	0,0	1	1,1	1

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
25.03.2017	13,4	3	0,5	5	57	3	0,0	1	7,1	5
26.03.2017	11,7	2	2,4	4	53	3	0,0	1	11,8	8
27.03.2017	15,8	4	-2,2	5	50	2	0,0	1	11,9	8
28.03.2017	20,2	5	0,8	4	49	2	0,0	1	12,0	9
29.03.2017	21,6	6	5,2	3	55	3	0,0	1	8,8	6
30.03.2017	17,3	4	12,4	1	63	4	0,0	1	1,0	1
31.03.2017	20,4	5	4,7	3	61	4	0,0	1	11,1	8
01.04.2017	21,8	6	9,3	2	50	2	0,0	1	11,7	8
02.04.2017	23,4	6	10,9	1	41	1	0,0	1	11,8	8
03.04.2017	20,8	5	9,5	1	60	4	0,0	1	7,0	5
04.04.2017	19,0	5	7,7	2	67	5	0,0	1	8,9	6
05.04.2017	15,3	3	9,8	1	70	6	0,6	1	4,4	3
06.04.2017	11,1	2	6,3	3	56	3	0,6	1	3,4	3
07.04.2017	10,0	2	4,5	3	78	7	1,6	1	0,1	1
08.04.2017	14,7	3	7,8	2	69	5	0,0	1	0,2	1
09.04.2017	17,9	4	5,2	3	61	4	0,0	1	8,0	6
10.04.2017	21,0	5	4,3	3	61	4	4,5	3	12,0	9
11.04.2017	16,6	4	7,2	2	50	2	0,0	1	6,5	5
12.04.2017	14,4	3	3,3	4	62	4	0,1	1	1,6	2
13.04.2017	13,0	3	8,3	2	65	5	1,4	1	3,6	3
14.04.2017	14,9	3	7,1	2	58	3	0,0	1	6,6	5
15.04.2017	15,4	3	4,3	3	67	5	0,4	1	0,5	1
16.04.2017	12,6	2	4,7	3	60	4	0,1	1	6,7	5
20.04.2017	5,0	1	1,7	4	57	3	0,0	1	0,7	1
21.04.2017	11,6	2	-2,0	5	54	3	0,0	1	7,5	6
22.04.2017	12,6	2	5,3	3	64	4	0,0	1	1,4	1
23.04.2017	10,6	2	3,3	4	60	4	0,2	1	7,8	6
24.04.2017	13,8	3	1,5	4	57	3	0,0	1	11,5	8
25.04.2017	16,0	4	4,7	3	65	5	0,7	1	1,1	1
26.04.2017	12,6	2	6,2	3	79	7	5,1	3	0,0	1
27.04.2017	8,0	1	4,4	3	86	8	10,2	5	0,0	1
28.04.2017	5,4	1	2,9	4	89	9	2,8	2	0,0	1
29.04.2017	12,7	3	0,6	5	69	5	0,2	1	7,3	5
30.04.2017	14,7	3	3,5	4	63	4	0,0	1	9,2	7
01.05.2017	15,6	4	5,6	3	56	3	0,2	1	7,2	5
02.05.2017	16,4	4	6,0	3	73	6	0,0	1	8,2	6
03.05.2017	18,8	5	3,9	3	79	7	13,6	7	3,8	3
04.05.2017	18,8	5	11,1	1	75	6	0,0	1	5,9	4
05.05.2017	19,3	5	10,1	1	80	7	1,7	1	5,6	4
06.05.2017	20,9	5	11,0	1	72	6	0,0	1	9,9	7

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
07.05.2017	19,2	5	10,1	1	66	5	6,5	3	7,2	5
08.05.2017	18,7	5	7,1	2	61	4	1,4	1	10,8	8
09.05.2017	11,9	2	5,0	3	59	4	0,1	1	3,1	3
10.05.2017	13,5	3	-1,3	5	58	3	0,0	1	13,4	9
11.05.2017	20,8	5	2,5	4	56	3	0,0	1	13,4	9
12.05.2017	22,1	6	12,3	1	71	6	4,0	2	5,6	4
13.05.2017	23,7	6	10,8	1	74	6	5,7	3	8,0	6
14.05.2017	20,8	5	9,4	2	73	6	0,0	1	6,7	5
15.05.2017	20,3	5	12,9	1	85	8	2,5	2	2,2	2
16.05.2017	22,9	6	9,6	1	65	5	0,0	1	8,2	6
17.05.2017	23,0	6	11,9	1	52	2	0,0	1	11,8	8
18.05.2017	25,0	7	10,8	1	56	3	0,0	1	14,5	10
19.05.2017	26,0	7	11,2	1	51	2	0,0	1	14,8	10
20.05.2017	19,2	5	13,6	1	69	5	0,0	1	0,0	1
21.05.2017	22,1	6	13,7	1	47	2	0,0	1	11,0	8
23.05.2017	25,3	7	10,4	1	62	4	2,9	2	7,8	6
24.05.2017	18,9	5	11,8	1	62	4	0,0	1	6,1	5
25.05.2017	19,3	5	8,5	2	61	4	0,1	1	1,3	1
26.05.2017	22,9	6	9,0	2	54	3	0,0	1	10,9	8
27.05.2017	24,0	6	11,1	1	53	3	0,0	1	12,5	9
28.05.2017	25,9	7	9,3	2	49	2	0,0	1	15,1	10
29.05.2017	28,7	8	11,7	1	47	2	0,0	1	13,3	9
30.05.2017	31,8	9	16,8	1	57	3	5,8	3	10,0	7
31.05.2017	27,0	7	16,6	1	64	4	0,0	1	10,3	7
01.06.2017	25,3	7	16,4	1	46	1	0,0	1	13,8	10
02.06.2017	27,7	8	8,7	2	42	1	0,0	1	13,3	9
03.06.2017	28,7	8	10,2	1	43	1	0,0	1	14,8	10
04.06.2017	28,0	8	15,1	1	63	4	0,0	1	7,2	5
05.06.2017	24,3	7	11,0	1	57	3	0,0	1	14,2	10
06.06.2017	28,7	8	14,7	1	65	5	3,4	2	8,2	6
07.06.2017	20,7	5	12,3	1	55	3	1,9	1	8,4	6
08.06.2017	23,7	6	8,6	2	45	1	0,0	1	15,2	10
09.06.2017	26,8	7	9,6	1	45	1	4,9	3	14,4	10
10.06.2017	24,5	7	14,4	1	67	5	0,5	1	7,3	5
11.06.2017	26,8	7	9,6	1	55	3	0,0	1	14,9	10
12.06.2017	30,5	9	12,8	1	44	1	0,0	1	8,1	6
13.06.2017	25,0	7	14,3	1	46	1	0,0	1	15,0	10
14.06.2017	23,7	6	12,4	1	51	2	0,0	1	13,6	10
15.06.2017	27,5	8	9,4	2	44	1	1,3	1	15,2	10
16.06.2017	25,1	7	14,9	1	81	7	8,8	5	4,4	3

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
17.06.2017	17,5	4	12,3	1	61	4	0,3	1	1,6	2
18.06.2017	24,1	6	12,8	1	57	3	0,0	1	5,9	4
19.06.2017	28,7	8	10,7	1	54	3	0,0	1	15,5	10
20.06.2017	33,0	10	14,0	1	49	2	0,0	1	15,4	10
21.06.2017	30,2	9	19,9	1	49	2	0,0	1	12,3	9
22.06.2017	33,1	10	13,9	1	46	1	13,0	6	13,7	10
23.06.2017	28,0	8	17,1	1	68	5	0,0	1	7,7	6
24.06.2017	31,9	9	14,7	1	46	1	0,0	1	13,8	10
25.06.2017	28,4	8	16,3	1	62	4	0,9	1	6,7	5
26.06.2017	29,2	8	16,3	1	55	3	0,0	1	10,3	7
27.06.2017	30,2	9	17,3	1	55	3	0,0	1	6,4	5
28.06.2017	33,3	10	19,1	1	47	2	0,0	1	11,2	8
29.06.2017	26,9	7	19,2	1	52	2	0,0	1	6,8	5
30.06.2017	26,9	7	15,6	1	52	2	0,0	1	4,7	4
01.07.2017	25,9	7	12,8	1	50	2	0,0	1	5,6	4
02.07.2017	25,3	7	14,1	1	60	4	0,3	1	2,8	2
03.07.2017	24,2	6	16,1	1	49	2	0,0	1	11,0	8
04.07.2017	25,2	7	11,7	1	51	2	0,0	1	3,1	3
05.07.2017	30,6	9	13,0	1	50	2	0,0	1	9,1	7
06.07.2017	28,1	8	15,1	1	54	3	7,6	4	8,8	6
07.07.2017	29,0	8	15,4	1	74	6	12,9	6	6,0	5
08.07.2017	30,5	9	16,1	1	56	3	0,0	1	11,8	8
09.07.2017	30,4	9	17,3	1	61	4	0,0	1	8,3	6
10.07.2017	31,7	9	19,0	1	70	6	6,8	4	5,2	4
11.07.2017	27,1	7	18,9	1	62	4	0,0	1	6,7	5
12.07.2017	27,6	8	14,9	1	60	4	0,2	1	7,3	5
13.07.2017	21,6	6	15,4	1	50	2	0,0	1	11,6	8
14.07.2017	22,5	6	8,7	2	51	2	0,3	1	7,3	5
15.07.2017	18,8	5	13,2	1	74	6	6,7	4	4,0	3
16.07.2017	24,4	7	10,9	1	52	2	0,0	1	10,2	7
17.07.2017	27,3	8	12,0	1	55	3	0,0	1	6,9	5
18.07.2017	29,7	8	15,2	1	54	3	0,0	1	10,5	8
19.07.2017	31,2	9	15,6	1	49	2	0,0	1	10,7	8
20.07.2017	32,5	9	16,1	1	56	3	22,0	10	8,7	6
21.07.2017	28,9	8	18,8	1	68	5	0,0	1	7,8	6
22.07.2017	30,6	9	18,2	1	60	4	0,0	1	7,7	6
23.07.2017	27,7	8	16,9	1	66	5	4,8	3	10,5	8
24.07.2017	22,3	6	16,9	1	78	7	15,1	7	0,1	1
25.07.2017	22,9	6	14,1	1	62	4	0,0	1	4,3	3
26.07.2017	20,3	5	12,7	1	76	7	3,0	2	3,7	3

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
27.07.2017	19,5	5	14,3	1	80	7	0,9	1	0,0	1
28.07.2017	27,3	8	15,2	1	65	5	0,0	1	6,5	5
29.07.2017	27,3	8	12,6	1	54	3	0,0	1	12,9	9
30.07.2017	31,3	9	14,7	1	57	3	0,0	1	11,4	8
31.07.2017	33,3	10	19,0	1	56	3	0,0	1	13,6	10
01.08.2017	35,0	10	21,2	1	48	2	0,0	1	13,9	10
02.08.2017	33,1	10	22,9	1	52	2	0,0	1	8,8	6
03.08.2017	36,1	10	19,3	1	52	2	0,0	1	9,7	7
04.08.2017	30,1	9	19,3	1	73	6	1,8	1	5,6	4
05.08.2017	32,2	9	21,1	1	48	2	0,0	1	11,3	8
06.08.2017	25,2	7	18,2	1	71	6	2,3	2	0,3	1
07.08.2017	25,0	7	15,6	1	52	2	0,0	1	12,3	9
08.08.2017	26,0	7	13,6	1	61	4	0,0	1	9,0	7
09.08.2017	31,8	9	18,5	1	57	3	12,7	6	10,1	7
10.08.2017	32,5	9	19,1	1	69	5	8,7	4	7,3	5
11.08.2017	29,9	8	17,5	1	79	7	0,1	1	3,9	3
12.08.2017	21,8	6	14,1	1	62	4	0,0	1	7,2	5
13.08.2017	24,6	7	12,3	1	64	4	0,0	1	7,0	5
14.08.2017	25,7	7	14,7	1	61	4	0,0	1	11,7	8
15.08.2017	29,3	8	14,9	1	53	3	0,0	1	11,1	8
16.08.2017	27,9	8	15,7	1	61	4	0,0	1	6,1	5
17.08.2017	28,8	8	17,3	1	61	4	0,0	1	9,9	7
18.08.2017	32,4	9	16,7	1	52	2	0,0	1	13,1	9
19.08.2017	25,5	7	18,0	1	79	7	1,9	1	0,0	1
20.08.2017	23,3	6	15,7	1	64	4	0,1	1	2,8	2
21.08.2017	21,9	6	10,5	1	56	3	0,0	1	9,7	7
22.08.2017	21,8	6	8,9	2	57	3	0,0	1	5,8	4
23.08.2017	22,8	6	8,3	2	55	3	0,0	1	12,6	9
24.08.2017	27,5	8	8,8	2	48	2	0,0	1	11,1	8
25.08.2017	30,7	9	13,6	1	52	2	0,0	1	11,0	8
26.08.2017	32,4	9	16,9	1	52	2	0,0	1	10,3	7
28.08.2017	25,6	7	16,5	1	47	2	0,0	1	12,5	9
29.08.2017	26,1	7	10,3	1	50	2	0,0	1	13,0	9
30.08.2017	28,5	8	11,7	1	50	2	0,0	1	12,7	9
31.08.2017	31,4	9	14,5	1	43	1	0,0	1	11,8	8
01.09.2017	25,7	7	13,5	1	81	7	6,5	3	0,2	1
02.09.2017	17,5	4	11,7	1	69	5	0,6	1	3,3	3
03.09.2017	16,2	4	12,3	1	78	7	1,3	1	1,3	1
04.09.2017	20,4	5	9,5	1	60	4	0,0	1	8,0	6
05.09.2017	23,8	6	7,8	2	54	3	0,0	1	9,9	7

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
06.09.2017	22,6	6	10,2	1	73	6	0,3	1	3,0	3
07.09.2017	20,1	5	11,6	1	58	3	0,0	1	2,6	2
14.09.2017	24,8	7	12,0	1	63	4	1,8	1	8,4	6
15.09.2017	19,1	5	9,3	2	61	4	1,3	1	3,1	3
16.09.2017	13,9	3	10,8	1	94	10	2,6	2	0,0	1
17.09.2017	12,6	2	11,1	1	92	9	7,7	4	0,0	1
18.09.2017	18,4	4	4,4	3	76	7	0,0	1	6,2	5
19.09.2017	14,6	3	8,2	2	85	8	11,5	6	0,2	1
20.09.2017	13,0	3	10,2	1	85	8	2,2	2	0,0	1
21.09.2017	15,3	3	9,2	2	76	7	0,5	1	0,8	1
22.09.2017	17,6	4	10,1	1	78	7	0,0	1	1,9	2
24.09.2017	14,6	3	11,0	1	89	9	0,5	1	0,0	1
25.09.2017	17,2	4	11,4	1	89	9	0,1	1	1,3	1
27.09.2017	19,9	5	9,9	1	82	8	0,0	1	3,4	3
28.09.2017	19,9	5	9,1	2	73	6	0,0	1	9,5	7
29.09.2017	18,6	5	7,6	2	70	6	0,0	1	8,4	6
02.10.2017	17,6	4	4,9	3	57	3	9,1	5	10,0	7
03.10.2017	13,1	3	7,7	2	97	10	13,7	7	0,0	1
04.10.2017	16,6	4	4,2	3	75	6	0,0	1	4,9	4
05.10.2017	19,5	5	10,0	1	59	4	0,0	1	0,4	1
06.10.2017	13,3	3	6,2	3	72	6	0,6	1	3,8	3
07.10.2017	14,5	3	5,6	3	75	6	1,9	1	4,2	3
08.10.2017	14,3	3	8,3	2	85	8	1,8	1	0,4	1
10.10.2017	15,2	3	5,1	3	79	7	0,2	1	0,1	1
11.10.2017	18,2	4	9,3	2	81	7	0,0	1	1,4	1
13.10.2017	17,8	4	6,1	3	77	7	0,0	1	6,3	5
14.10.2017	19,6	5	5,0	3	82	8	0,0	1	7,1	5
15.10.2017	21,0	5	6,7	2	80	7	0,0	1	9,9	7
16.10.2017	19,5	5	8,0	2	86	8	0,0	1	9,7	7
17.10.2017	20,5	5	7,6	2	87	8	0,0	1	5,4	4
18.10.2017	19,0	5	8,5	2	92	9	0,0	1	5,0	4
19.10.2017	18,9	5	8,5	2	82	8	0,2	1	9,1	7
20.10.2017	12,0	2	7,7	2	96	10	0,1	1	0,0	1
21.10.2017	13,2	3	9,3	2	94	10	3,7	2	0,0	1
22.10.2017	12,4	2	9,5	1	89	9	3,5	2	0,0	1
23.10.2017	10,9	2	8,2	2	80	7	0,0	1	0,0	1
24.10.2017	13,5	3	8,1	2	75	6	0,0	1	1,2	1
25.10.2017	14,6	3	6,2	3	85	8	0,0	1	1,2	1
26.10.2017	17,2	4	6,5	3	84	8	0,0	1	1,8	2
27.10.2017	12,3	2	7,5	2	81	7	1,7	1	1,0	1

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
28.10.2017	10,9	2	7,9	2	69	5	0,7	1	0,0	1
29.10.2017	11,8	2	5,7	3	66	5	9,4	5	3,3	3
30.10.2017	8,4	1	3,0	4	70	6	0,4	1	6,3	5
31.10.2017	7,1	1	0,7	4	87	8	0,1	1	0,1	1
01.11.2017	9,0	1	-0,8	5	85	8	0,0	1	0,9	1
02.11.2017	13,1	3	3,3	4	79	7	0,0	1	3,5	3
03.11.2017	12,3	2	2,4	4	81	7	0,0	1	2,5	2
04.11.2017	13,0	3	1,2	4	79	7	0,0	1	7,6	6
05.11.2017	9,5	1	5,7	3	81	7	0,0	1	0,0	1
06.11.2017	10,5	2	7,2	2	79	7	0,0	1	0,2	1
07.11.2017	12,0	2	6,4	3	76	7	0,0	1	7,3	5
08.11.2017	6,9	1	5,1	3	87	8	1,9	1	0,0	1
09.11.2017	7,8	1	5,1	3	94	10	0,0	1	0,0	1
10.11.2017	8,0	1	5,7	3	84	8	0,5	1	0,0	1
11.11.2017	8,2	1	2,8	4	75	6	0,8	1	3,6	3
12.11.2017	7,0	1	0,8	4	87	8	8,8	5	3,8	3
13.11.2017	6,2	1	2,4	4	77	7	0,0	1	0,0	1
14.11.2017	6,3	1	-0,9	5	79	7	0,0	1	3,3	3
15.11.2017	7,1	1	-4,3	6	84	8	0,0	1	8,4	6
16.11.2017	4,0	1	-2,0	5	96	10	0,0	1	0,0	1
17.11.2017	5,6	1	2,6	4	86	8	1,7	1	0,0	1
18.11.2017	4,9	1	1,8	4	83	8	0,0	1	0,6	1
19.11.2017	4,8	1	1,6	4	80	7	2,6	2	1,5	2
20.11.2017	5,8	1	0,8	4	73	6	1,9	1	0,4	1
21.11.2017	6,5	1	1,3	4	96	10	1,6	1	0,0	1
22.11.2017	11,5	2	3,6	3	86	8	0,0	1	5,6	4
23.11.2017	5,5	1	-1,4	5	97	10	0,0	1	1,3	1
24.11.2017	5,2	1	3,5	4	96	10	0,0	1	0,0	1
25.11.2017	7,8	1	4,0	3	94	10	0,0	1	0,0	1
26.11.2017	5,6	1	-0,7	5	73	6	0,0	1	2,0	2
27.11.2017	6,2	1	-3,6	6	76	7	0,0	1	2,6	2
28.11.2017	4,4	1	-0,6	5	78	7	0,0	1	0,0	1
29.11.2017	2,6	1	0,1	5	90	9	4,8	3	0,0	1
30.11.2017	2,0	1	0,1	5	94	10	8,6	4	0,0	1
01.12.2017	1,9	1	-5,5	7	91	9	0,0	1	7,1	5
02.12.2017	0,4	1	-8,6	8	88	9	0,0	1	7,3	5
03.12.2017	1,2	1	-3,4	6	75	6	0,0	1	0,0	1
04.12.2017	2,8	1	-1,4	5	85	8	0,0	1	0,0	1
05.12.2017	4,9	1	-2,4	6	84	8	0,0	1	0,0	1
06.12.2017	4,9	1	4,1	3	71	6	0,0	1	0,0	1

Dátum	T max	Index T max	T min	Index T min	Vlhkosť vzduchu	Index vlhkosť	Úhrn zrážok	Index zrážok	Sl. svit	Index svit
07.12.2017	4,4	1	-0,8	5	84	8	0,0	1	6,2	5
08.12.2017	0,2	1	-2,1	5	98	10	1,1	1	0,0	1
09.12.2017	1,8	1	-2,2	5	75	6	0,0	1	0,0	1
10.12.2017	1,8	1	-4,8	6	72	6	0,0	1	5,4	4
11.12.2017	11,0	2	0,0	5	77	7	0,0	1	0,0	1
12.12.2017	12,3	2	3,7	3	70	6	0,0	1	0,0	1
13.12.2017	4,5	1	-1,6	5	76	7	0,0	1	7,5	6
14.12.2017	6,9	1	-1,6	5	75	6	0,8	1	1,6	2
15.12.2017	4,2	1	0,1	5	87	8	0,3	1	0,1	1
16.12.2017	3,6	1	-2,2	5	84	8	0,0	1	0,1	1
17.12.2017	3,1	1	-3,8	6	88	9	0,0	1	0,8	1
18.12.2017	1,4	1	-3,3	6	83	8	0,3	1	2,5	2
19.12.2017	-0,5	1	-6,9	7	84	8	0,0	1	1,7	2
20.12.2017	1,3	1	-2,0	5	84	8	0,0	1	0,0	1
21.12.2017	3,1	1	-1,2	5	93	9	5,1	3	0,0	1
22.12.2017	6,5	1	1,0	4	77	7	0,0	1	0,1	1
23.12.2017	8,3	1	2,9	4	73	6	0,0	1	0,0	1
24.12.2017	9,3	1	6,3	3	73	6	0,0	1	0,0	1
25.12.2017	7,8	1	0,8	4	88	9	0,0	1	6,3	5
26.12.2017	2,5	1	-4,3	6	96	10	0,0	1	0,0	1
27.12.2017	8,8	1	-0,6	5	84	8	1,3	1	0,1	1
28.12.2017	8,2	1	2,9	4	82	8	0,9	1	0,0	1
29.12.2017	3,2	1	-1,9	5	73	6	0,0	1	3,2	3
30.12.2017	1,0	1	-3,8	6	88	9	2,6	2	1,3	1
31.12.2017	6,9	1	0,1	5	95	10	2,8	2	0,4	1