

ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA STROJNÍ

Studijní program: N0715A270012 – Průmyslové inženýrství
a management

Studijní specializace: Bez specializace

DIPLOMOVÁ PRÁCE

Analýza a vyhodnocení dat z výroby

Autor: Bc. Petr MĚSTKA

Vedoucí práce: doc. Ing. Pavel RAŠKA, PhD.

Akademický rok 2023/2024

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta strojní

Akademický rok: 2023/2024

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Petr MĚSTKA**
Osobní číslo: **S22N0064P**
Studijní program: **N0715A270012 Průmyslové inženýrství a management**
Téma práce: **Analýza a vyhodnocení dat z výroby**
Zadávající katedra: **Katedra průmyslového inženýrství a managementu**

Zásady pro vypracování

- Úvod do řešené problematiky
- Analýza současného stavu
- Analýza řešení
- Implementace řešení
- Závěr a zhodnocení

Rozsah diplomové práce: **50 až 70 stran**
Rozsah grafických prací: **-**
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

1. DECKLER, Greg. *Learn power BI: a comprehensive, step-by-step guide for beginners to learn real-world business intelligence*. Second edition. Birmingham: Packt Publishing, 2021. ISBN 978-1-80181-007-4.
2. CLARK, Dan, *Beginning Microsoft Power BI: A Practical Guide to Self-Service Data Analytics*. APress, 2020. ISBN 9781484256190.
3. ESCOBAR, Miguel, PULS, Ken, *Master Your Data with Excel and Power BI*. Holy Macro! Books, 2021. ISBN 9781615470587.
4. SEAMARK, Philip, *Beginning DAX with Power BI*. APress, 2018. ISBN 1484234766.
5. FERRARI, Alberto, RUSSO, Marco, *Analyzing Data with Power BI and Power Pivot for Excel*. Microsoft Press, 2017. ISBN 9781509302765.
6. ALEXANDER, Michael, *Excel Power Pivot & Power Query For Dummies*. John Wiley & Sons Inc, 2022. ISBN 9781119844488.

Vedoucí diplomové práce: **Doc. Ing. Pavel Raška, Ph.D.**
Katedra průmyslového inženýrství a managementu

Konzultant diplomové práce: **Ing. Bc. Miroslav Malaga**
Katedra průmyslového inženýrství a managementu

Datum zadání diplomové práce: **16. října 2023**
Termín odevzdání diplomové práce: **24. května 2024**

L.S.

Doc. Ing. Vladimír Duchek, Ph.D.
děkan

Doc. Ing. Michal Šimon, Ph.D.
vedoucí katedry

Prohlášení o autorství

Předkládám tímto k posouzení a obhajobě diplomovou práci zpracovanou na závěr studia na Fakultě strojní Západočeské univerzity v Plzni.

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně, s použitím odborné literatury a pramenů uvedených v seznamu, který je součástí této diplomové práce.

V Plzni dne:

.....

podpis autora

Rád bych tímto poděkoval svému vedoucímu práce, panu doc. Ing. Pavlu Raškovi, PhD., za odborné vedení diplomové práce, rady, trpělivost, čas a možnost kdykoliv práci konzultovat. Také bych chtěl poděkovat panu Ing. Bc. Miroslavu Malagovi, PhD. za konzultace a pomoc při řešení problematiky této diplomové práce. V neposlední řadě patří veliké poděkování mé rodině, která mě po celou dobu studia neustále podporovala a bez které bych se tak daleko nedostal.

ANOTAČNÍ LIST DIPLOMOVÉ PRÁCE

AUTOR	Příjmení Městka	Jméno Petr	
STUDIJNÍ PROGRAM			
VEDOUCÍ PRÁCE	Příjmení (včetně titulů) Doc. Ing. Raška, PhD.	Jméno Pavel	
PRACOVIŠTĚ	ZČU - FST - KPV		
DRUH PRÁCE	DIPLOMOVÁ	BAKALÁŘSKÁ	Nehodící se škrtněte
NÁZEV PRÁCE	Analýza a vyhodnocení dat z výroby		

FAKULTA	strojní	KATEDRA	KPV	ROK ODEVZD.	2024
----------------	---------	----------------	-----	--------------------	------

POČET STRAN (A4 a ekvivalentů A4)

CELKEM	73	TEXTOVÁ ČÁST	65	GRAFICKÁ ČÁST	8
---------------	----	---------------------	----	----------------------	---

<p style="text-align: center;">STRUČNÝ POPIS (MAX 10 ŘÁDEK)</p> <p style="text-align: center;">ZAMĚŘENÍ, TÉMA, CÍL POZNATKY A PŘÍNOSY</p>	<p>Diplomová práce je zaměřena na analýzu a vyhodnocení dat z výroby. V první části se tato práce zabývá teorií spjatou s daty a databázemi. Jsou zde popsány jednotlivé problémy s daty, typy dat, možnosti jejich sběru, zpracování a ukládání. Zároveň jsou zde popsány základní operace v nástroji Microsoft Power BI. Druhá část práce je věnována analýze obdržených výrobních dat a popisu problémů s těmito daty. Následně proběhl návrh struktury databáze, který umožní smysluplně vizualizovat, a demonstrace funkčnosti tohoto návrhu vizualizací dat v nástroji Microsoft Power BI.</p>
<p style="text-align: center;">KLÍČOVÁ SLOVA</p> <p style="text-align: center;">ZPRAVIDLA JEDNOSLOVNÉ POJMY, KTERÉ VYSTIHUJÍ PODSTATU PRÁCE</p>	<p style="text-align: center;">Data, databáze, datová analýza, analýza výrobních dat, návrh struktury databáze, vizualizace dat, Power BI</p>

SUMMARY OF DIPLOMA SHEET

AUTHOR	Surname Městka	Name Petr	
STUDY PROGRAMME			
SUPERVISOR	Surname (Inclusive of Degrees) Doc. Ing. Raška, PhD.	Name Pavel	
INSTITUTION	ZČU - FST - KPV		
TYPE OF WORK	DIPLOMA	BACHELOR	Delete when not applicable
TITLE OF THE WORK	Analysis and evaluation of production data		

FACULTY	Mechanical Engineering	DEPARTMENT	KPV	SUBMITTED IN	2024
----------------	------------------------	-------------------	-----	---------------------	------

NUMBER OF PAGES (A4 and eq. A4)

TOTALLY	73	TEXT PART	65	GRAPHICAL PART	8
----------------	----	------------------	----	-----------------------	---

BRIEF DESCRIPTION TOPIC, GOAL, RESULTS AND CONTRIBUTIONS	The thesis is focused on the analysis and evaluation of production data. The first part of this thesis deals with the theory related to data and databases. It describes the different problems with data, types of data, possibilities of data collection, processing and storage. It also describes the basic operations in Microsoft Power BI. The second part of the thesis is devoted to the analysis of the received production data and the description of problems with these data. Subsequently, the design of a database structure that allows meaningful visualization is discussed, and the functionality of this design is demonstrated by visualizing the data in Microsoft Power BI.
KEY WORDS	Data, database, data analysis, production data analysis, database structure design, data visualization, Power BI

Obsah

Přehled použitých zkratk a symbolů.....	8
Seznam obrázků	9
Seznam tabulek	11
Úvod	12
1 Data	13
1.1 Typy dat.....	13
1.2 Ukládání dat	13
1.3 Problémy spojené s daty	16
1.4 Sběr dat.....	17
1.5 Zpracování dat.....	18
1.6 Normalizace dat.....	18
1.7 Datová analýza	19
1.8 Funkční analýza.....	22
2 Power BI.....	24
2.1 Součásti Power BI	24
2.2 Prostředí Power BI	26
2.3 Základní operace	28
2.3.1 Nahrání dat	28
2.3.2 Transformace dat.....	28
2.3.3 Tvorba relací	29
2.3.4 Vizualizace	30
3 Analýza současného stavu.....	31
3.1 Popis společnosti	31
3.2 Popis výrobních dat.....	32
3.3 Problémy s daty	39
3.4 Transformace dat.....	43
3.5 Tvorba míry a sloupce.....	46
3.6 Relace	47
3.7 Problémy s vizualizací dat.....	47
4 Návrh struktury databáze	49
4.1 Tabulky dimenzí.....	49
4.2 Faktové tabulky	54
4.3 Naplnění navržené struktury daty	55
4.4 Transformace dat.....	56

4.5	Relace	58
4.6	Vizualizace dat	59
4.6.1	Přehled.....	59
4.6.2	Stroj	60
4.6.3	Člověk	61
4.6.4	Operace.....	62
	Závěr.....	64
	Seznam použitých zdrojů	65
	PŘÍLOHA č. 1	i
	PŘÍLOHA č. 2.....	iii
	PŘÍLOHA č. 3	v
	PŘÍLOHA č. 4.....	vii

Přehled použitých zkratk a symbolů

IT	Informační technologie
BI	Business Intelligence
SQL	Structured Query Language
PHM	Pohonné hmoty
VBA	Visual Basic for Application
RTDB	Real Time Database
NoSQL	Not only Structured Query Language
E-R	Entity - Relationship
DFD	Data Flow Diagram

Seznam obrázků

Obr. 1 Základní prvky E-R diagramu [14].....	19
Obr. 2 Příklad ER diagramu	19
Obr. 3 Výskytový diagram [1]	20
Obr. 4 - Kardinalita – vztah 1:1 [1].....	21
Obr. 5 Kardinalita – Vztah 1: N [1]	21
Obr. 6 Kardinalita – Vztah M: N [1].....	22
Obr. 7 Základní grafické prvky DFD [1]	22
Obr. 8 Diagram datových toků [1]	23
Obr. 9 Součásti Power BI [17]	24
Obr. 10 Power BI Desktop [17]	25
Obr. 11 Služba Power BI [17].....	25
Obr. 12 Power BI Mobile [17]	26
Obr. 13 Typy zobrazení v Power BI desktop [17]	26
Obr. 14 Prostředí editoru Power Query [18].....	27
Obr. 15 Nahrání dat v Power BI Desktop	28
Obr. 16 Transformace dat v Power BI Desktop	29
Obr. 17 Tvorba relací v Power BI Desktop [20].....	29
Obr. 18 Typy vizuálů v Power BI Desktop [17]	30
Obr. 19 Vizualizace v Power BI Desktop [17].....	30
Obr. 20 Logo společnosti ALIMEX NEZVĚSTICE a.s. [22]	31
Obr. 21 WF_přehled_mechanizace – názvy strojů	39
Obr. 22 Problémy s daty - WF_IQ vnitro.....	42
Obr. 23 WF_přehled_mechanizace po transformaci dat.....	44
Obr. 24 spotřeba PHM po transformaci dat	45
Obr. 25 Míra – Pr_spotřeba.....	46
Obr. 26 Sloupec – NazevMesice	46
Obr. 27 Relace databáze současného stavu	47
Obr. 28 Problémy s vizualizací – spotřeba na stroj.....	48
Obr. 29 Problémy s vizualizací – spotřeba na řidiče.....	48
Obr. 30 Naplnění CF_cerven_zari – návrh struktury	55
Obr. 31 Návrh struktury – Míra – Průměrná spotřeba	56
Obr. 32 Návrh struktury – Míra – průměrná rychlost	57
Obr. 33 Návrh struktury – Míra – Náklady na vzdálenost.....	57
Obr. 34 Návrh struktury – Míra – Počet jízd.....	57

Obr. 35 Datový model návrhu struktury databáze	58
Obr. 36 Dashboard – Přehled	60
Obr. 37 Dashboard – Stroj.....	61
Obr. 38 Dashboard – Člověk.....	62
Obr. 39 Dashboard – Operace	63

Seznam tabulek

Tabulka 1 WF_přehled_mechanizace – struktura	32
Tabulka 2 WF_IQ vnitro – struktura	33
Tabulka 3 Spotřeba nafty – struktura	34
Tabulka 4 Spotřeba PHM - struktura	34
Tabulka 5 WF_spotřeba nafty – struktura	35
Tabulka 6 WF_spotřeba PHM – struktura	36
Tabulka 7 CF_cerven_zari_alimex – struktura	37
Tabulka 8 Plodiny_LPIS_alimex – struktura	38
Tabulka 9 VO_alimex – struktura	38
Tabulka 10 Neúplnost dat – WF_přehled_mechanizace	40
Tabulka 11 Neúplnost dat – CF_cerven_zari	41
Tabulka 12 Neúplnost dat – WF_IQ vnitro	42
Tabulka 13 WF_přehled_mechanizace – návrh struktury	49
Tabulka 14 Znacka_stroje – návrh struktury	50
Tabulka 15 Rada_stroje – návrh struktury	50
Tabulka 16 Skupina_stroje – návrh struktury	50
Tabulka 17 Druh_stroje – návrh struktury	51
Tabulka 18 Misto_zarazeni – návrh struktury	51
Tabulka 19 VO_alimex – návrh struktury	51
Tabulka 20 Plodiny_LPIS_alimex – návrh struktury	52
Tabulka 21 Pridavne_zarizeni – návrh struktury	52
Tabulka 22 Jmeno_ridice – návrh struktury	53
Tabulka 23 Parcela – návrh struktury	53
Tabulka 24 CF_cerven_zari – návrh struktury	54
Tabulka 25 spotřeba PHM – návrh struktury	55

Úvod

V dnešní době, kdy je digitalizace velkým trendem, téměř každý podnik pracuje s daty v digitální podobě. K tomu je zapotřebí datům porozumět, umět je zpracovat, ukládat, načítat, transformovat, čistit a vyhodnocovat. Firmy tak své pracovníky v datové analýze školí, a dokonce kvůli tomu vznikají i nové pozice. Existuje také možnost tyto činnosti outsourcovat firmami, povětšinou IT, které se těmito službami zabývají. Společnosti tímto způsobem, tedy daty v digitální podobě, předchází časovým ztrátám při hledání papírových záznamů, jejich ručnímu přepočítávání a vyhodnocování. Navíc se snáz předejde ztrátám záznamů a není zapotřebí fyzické místo pro jejich skladování.

Tato diplomová práce na téma Analýza a vyhodnocení dat z výroby se zaměřuje na analýzu dat, zjištění problémů s nimi spojených a následnou možností tato data vizualizovat. Práce se dělí na dvě části, teoretickou a praktickou. Teoretická část obsahuje informace, které jsou využity při řešení cíle diplomové práce. Cílem je navrhnout takovou strukturu databáze, aby bylo možné data vizualizovat. Pro jeho dosažení je tedy zapotřebí obdržená data náležitě analyzovat, odhalit jejich problémy a ty poté napravit.

1 Data

Data se v dnešní době nachází ve všech výrobních podnicích a celkově všude kolem nás. Jedná se o záznam lidského poznání, kterým mohou být údaje, vědomosti, poznatky, ale také výsledky pozorování různých procesů. Tyto záznamy jsou základem pro vytváření informací, které vznikají smysluplnou interpretací dat, a ze kterých člověk získává znalosti. [1]

1.1 Typy dat

Data se dělí do dvou kategorií, kvalitativní a kvantitativní, které odpovídají typům sledovaných veličin.

Kvalitativní data je možné rozdělit následovně:

- Binární data – reprezentují data, která mohou mít pouze dvě hodnoty. Nejčastěji se lze setkat s příkladem ano/ne, který se v číselné podobě kóduje pomocí 1 (ano) a 0 (ne). Dalšími příklady může být např. pohlaví (žena/muž), rodinný stav (ženatý/svobodný), zaměstnanost (zaměstnaný/nezaměstnaný) atd.
- Nominální data – u tohoto typu dat může být příkladem např. typ oddělení (IT, engineering, údržba, projektový management, manufacturing), národnost zaměstnance (česká/polská/belgická/...). Je to tedy typ, který obsahuje více skupin, u kterých neexistuje přirozené pořadí jejich hodnot a nelze je tedy seřadit. Zároveň tato data není možné porovnat v relaci větší/menší.
- Ordinální data – i tato skupina obsahuje více kategorií. Oproti skupině nominálních dat je zde však možnost tato data navzájem seřadit a lze je i porovnávat v relaci větší/menší. Jako příklad této kategorie lze uvést stupeň bolesti zranění pracovníka (mírná/střední/velká/nesnesitelná), míra automatizace výroby (manuální, poloautomatizovaná, automatizovaná, autonomní) (manuální stav produktu (nepoužitelný, opotřebovaný, zánovní, nový), a jiné.

Data kvantitativní již dělíme pouze do dvou skupin:

- Spojitá – pod spojitými daty si můžeme představit data taková, kterých je možné v určitém intervalu nabývat jakýchkoliv hodnot. Představitelem této kategorie může být například teplota, ale také jím mohou být biometrické údaje člověka jako výška, hmotnost.
- Diskrétní – druhá kategorie kvantitativních dat je omezena na spočetně mnoho hodnot. Pokud se tato data reprezentují, zobrazují se na reálné ose pomocí izolovaných bodů. Za příklad diskrétních dat lze uvést počet vyrobených produktů, počet zaměstnanců ve firmě, a další. [2]

1.2 Ukládání dat

Zaznamenaná data je možné ukládat několika způsoby, které budou popsány v této kapitole. Data by měla být uživateli vždy přístupná a uložena tak, aby byla i nadále použitelná.

Textové soubory

Běžným uložištěm dat jsou textové soubory, do kterých data nejčastěji proudí z protokolů, e-mailů, senzorů a transakcí. Využívají se kvůli tomu, že se snadno převádí do jiných formátů a snadněji se obnovuje jejich zpracování. Existuje několik formátů textových souborů. Nejčastěji se využívají formáty CSV, TXV, ve kterých se data oddělují čárkami, respektive tabulátorem. Dále se často užívá formátů XML a JSON. [3]

Soubory aplikace Excel

V tomto případě se bavíme o nástroji, který je v dnešní době hojně využíván. Jedná se o tabulkový procesor, ve kterém je možné data zaznamenávat, upravovat, počítat, ale také vizualizovat a vyhodnocovat pomocí kontingenčních grafů a tabulek. Navíc excelovský soubor může obsahovat kód v jazyku VBA, Visual Basic for Application, jehož pomocí lze vytvářet SQL dotazy na tabulky nebo externí databáze. Vytvořené soubory pomocí excelu ve formátu .xls (.xlsx), lze převádět potřeby na jiné textové formáty jako CSV, TSV, XML. [3]

Cloudové uložení

Cloud umožňuje uživatelům ukládat data mimo místní pevné disky a úložné sítě prostřednictvím veřejného internetu či připojením k soukromé síti. Tato přenesená data jsou spravována třetí stranou, která je za uložení dat zodpovědná. Tato strana hostuje a spravuje servery a zároveň poskytuje přístup ke svěřeným datům vždy když je potřeba.

Jedná se o variantu, která je škálovatelná a výhodná z hlediska nákladů. V případě narůstajícího objemu dat je tedy možné navýšit kapacitu cloudu, což pevné disky počítače se svým omezeným uloženištěm neumožňují. [3]

Relační databáze

Relačními databázemi se rozumí takové databáze, které obsahují položky s definovanými vazbami mezi nimi. Jednotlivé položky těchto databází jsou umístěny do tabulek tvořených sloupci a řádky, v nichž se nachází zaznamenaná data. Každý sloupec obsahuje jednotný druh záznamů a pole, která jsou uloženištěm hodnot atributů. Naopak řádky jsou souvislejícím souborem hodnot jednoho objektu. Tyto řádky bývají popsány primárním klíčem, který slouží jako jejich identifikátor. Dále mohou tyto řádky obsahovat tzv. cizí klíče, které jsou potřeba k propojování řádků z jiných tabulek. [5]

Relační databáze jsou uzpůsobeny pro práci s programovacím jazykem SQL. Ten se využívá k dotazování, manipulaci, načítání a ukládání dat. S jeho pomocí se mohou také vytvářet tabulky, jelikož obsahuje jazyk DDL. [6]

In-memory databáze

Další možností uložení dat jsou in-memory databáze, které se vyskytují i pod zkratkou RTDB (Real Time DataBase). Tyto databáze umožňují ukládání různých typů dat v jediném systému. Lze v nich ukládat jak strukturovaná, tak nestrukturovaná data, kterými jsou hlas, video, e-maily nebo volné dokumenty. Využívají se tam, kde je zapotřebí zpracovat velké množství dat a pro organizace, ve kterých dochází k rychlému rozšiřování dat, například v on-line bankovníctví, ve správě sítí atd. [5]

NoSQL databáze

NoSQL databáze jsou nerelační typy databází a lze je dotazovat více způsoby, proto se také nazývají „not only SQL“. K jejich dotazování se využívá idiomatičtých jazyků API a deklarativních strukturovaných dotazovacích jazyků. Výhodou těchto databází je jejich škálovatelnost a vysoká dostupnost. Z toho důvodu jsou velmi rozšířené při práci s Big daty a real-time weby. [6]

Existují čtyři základní typy NoSQL úložišť:

- **Úložiště klíčových hodnot** – V případě této databáze aplikace přebírá veškerou kontrolu nad ukládáním dat do polí hodnot, a to bez omezení. Na základě tohoto faktu se úložiště klíčových hodnot označuje jako nejflexibilnější typ NoSQL databáze. [6]
- **Úložiště dokumentů** – Tento typ lze také najít pod označením dokumentově orientované databáze. Slouží k ukládání, vyhledávání a správě polostrukturovaných dat a není zde nutná specifikace polí dokumentů. [6]
- **Úložiště grafů** – Tyto úložiště ukládají data ve formě uzlů a relací, které představují vazby mezi uzly. Tím dochází k podpoře komplexnějšího a obsáhlejšího zobrazení dat. Databáze grafů se využívají při odkrývání podvodů a v oblastech rezervačních systémů či sociálních sítí. [6]
- **Sloupcové úložiště** – Poslední typ databáze je ve formě tabulek, řádků a sloupců, do kterých data ukládá. Jsou tedy rozšířeny hlavně pro aplikace využívající sloupcový formát pro zapisování dat bez schématu. [6]

Datový sklad

Jedná se o systém ukládání dat, který je navržený pro podporu práce s Business Intelligence a datových analytiků. Jejich účel tedy spočívá hlavně ve vytváření dotazů a analýz. V těchto skladech se povětšinou nachází velké množství historických dat, která se získávají ze souborů aplikací a transakčních aplikací. Tyto sklady se skládají často z relačních databází, nástrojů pro extrakci, načtení a transformaci dat. Nechybí tomu však také možnosti jako statistické analýzy, reporting či data mining, ale také vizualizační a prezentační nástroje. [7]

Data Lakes

Pod pojmem Data lakes se skrývá centralizované úložiště. Toto úložiště pracuje s velkými objemy dat, které ingestuje, tedy načítá data z jednoho a více zdrojů, do určitého umístění a ukládá je v jejich prvotní formě. Data lakes mohou přijímat data jak strukturovaná, tak částečně strukturovaná i nestruturovaná, z kteréhokoliv zdroje. To je umožněné díky tomu, že Data lakes mají svoji architekturu otevřenou a škálovatelnou. S ohledem na široké spektrum uživatelů se datové soubory ukládají v nezpracovaných, vyčištěných a spravovaných zónách, aby jejich použití bylo bezproblémové v jakékoli formě. Tím jsou Data lakes dostupné pro různé typy uživatelů. Data lakes jsou vhodná pro analýzu velkého objemu dat, prediktivní analýzy nebo strojová učení, a proto se používají hlavně v IoT, streamovacích médiích, finančnictví, zdravotnictví atd. [8]

1.3 Problémy spojené s daty

Stejně jako u ostatních činností, tak i u práce s daty se vyskytují určité problémy, které narušují a přidělávají práci při jejich vhodné správě. Ať už je to redundance, či nekonzistentnost dat, je potřeba tyto problémy včas zachytit, věnovat se jim a také je správně vyřešit.

Nekonzistentnost dat

Tento problém s daty vzniká v případě, že se data sbírají z více zdrojů, ale dochází k němu také při slučování nebo přesunu společností. Tato nekonzistentnost se poté projevuje tak, že dochází k rozdílům ve formátu, jednotkách nebo ve způsobu zápisu těchto dat. Tyto nance je zapotřebí včas řešit, jelikož se mohou nahromadit, snižovat celkovou spolehlivost a kvalitu dat a ztížit následnou analýzu a vyhodnocení dat. [8]

Výpadky/prostoje dat

V dnešní době jsou data ve většině podniků velmi důležitým článkem řízení výroby, v některých případech dokonce celé společnosti. Může však nastat období, kdy data nejsou spolehlivá nebo k dispozici. Tyto výpadky ovlivňují chod podniku a mohou způsobit zmatky, přerušování výroby, služeb a z toho plynoucí nespokojenost zákazníků. Častým důvodem, kdy nastává výpadek, je migrace dat. Těmto problémům se lze vyvarovat včasnou aktualizací, údržbou a zajišťováním integrity datového toku, ale také jeho neustálým monitoringem a automatizací. [8]

Nejasnost dat

Pokud jsou posbíraná data nejasná, mohou způsobovat problémy s reportíngem a jejich analýzou. K tomu dochází především v masivních databázích a při značně rychlém proudění dat. Tyto chyby se mohou projevovat špatným formátováním, zavádějícími názvy sloupců, položek atd. [8]

Duplikace dat

Některé zdroje mají tendence duplikovat data a navzájem je překrývat. Pro příklad lze uvést duplikované kontaktní informace. V případě, že se duplikují tyto informace, dochází tak ke dvojímu kontaktování zákazníka a zbytečnému přidělování práce, nebo v horším případě k odrazování zákazníka. [8]

Nepřesnost dat

V některých oborech je přesnost dat naprosto klíčová. V případě nepřesnosti informací totiž dochází ke zkreslení skutečnosti, což vede ke nevhodným rozhodnutím na základě pomíjivého náhledu. Nepřesnosti mohou vznikat z důvodu lidského zaváhání, ale často vznikají kvůli degradaci dat. [8]

Velké množství dat

V případě nadbytečného množství informací může dojít ke snížení kvality dat, nepřehlednosti a z toho plynoucí jejich úplné ztrátě vlivem přehlédnutí. To může vést k chybějícím informacím v reportech a vyhodnocování. Z toho důvodu lidé v praxi věnují dost času hledání a organizování příslušných dat. [8]

Redundance dat

Redundance dat je spojena s jejich ukládáním na více jak dvou místech v databázi, nebo systému datového skladu. Pokud tento jev nastává vědomě, napomáhá společnosti zachovat nepřetržitý provoz výroby nebo služeb v případě poškození a ztráty dat. Využívá se tedy pro zálohování a obnovu dat. V případě, že k redundanci dochází nezáměrně, vznikají duplicitní data, která zaviňují nesrovnalosti a ovlivňují větší sady dat. Tento nárůst se proměňuje jeden z problému, který s daty řešíme. Ukládáním dat na několika místech dochází totiž nejen k zabírání úložného prostoru, ale dochází také ke vzniku nekonzistencí. [8]

1.4 Sběr dat

Pro efektivní řízení společnosti a identifikace problémů, je nutný sběr dat. Ten může probíhat za pomoci dvou typů zdrojů, kterými jsou zdroje interní a externí.

Interní zdroje

Jedná se o takové zdroje dat, které se uchovávají pouze uvnitř společnosti. Tyto zdroje poskytují data, která jsou jedinečná pro produkty, výrobky nebo služby dané firmy a jedná se tak o data s nejvyšší prediktivní schopností pro modelování. Nejčastějšími představiteli interních zdrojů jsou zákaznické databáze, transakční databáze, databáze historie nabídek, telefonické oslovení a datové sklady. [10]

Externí zdroje

Pod externími zdroji se často nachází prodejci a kompilátoři seznamů. Prodejci jsou firmy, které sbírají různé typy dat v rámci své hlavní činnosti a jejich prodej mají jako činnost vedlejší. Tato nashromážděná data se liší na základě hlavní činnosti a mohou jimi být například kontaktní údaje, tedy jména, adresy, telefonní čísla, ale také data demografického, behaviorálního či psychografického charakteru. Naopak pod pojmem kompilátoři si lze představit firmy prodávající různé seznamy, které jsou kompilovány z více databází, anebo založeny na jednom jediném seznamu. Tyto seznamy mohou vycházet například z telefonního seznamu a registračních dat z řidičských průkazů. Poté své seznamy zdokonalují výzkumem, nákupem dalších seznamů, jejich slučováním a doplňováním informací. [10]

1.5 Zpracování dat

V otázce zpracování dat se rozlišují dva typy, procedurální a databázové.

První ze zmíněných typů je procedurální, nebo také souborové, zpracování dat, které se zabývá otázkou „Jakým způsobem získám výsledek ze vstupních dat?“ Tento typ se dělí dále na zpracování klasických souborů a souborů indexsekvencních. V tomto případě je nevýhodou, že se nic neobejde bez vytváření kódu programátorem, což souvisí i s celkovou časovou náročností vývoje aplikace. Na druhou stranu je zde možnost programování i složitějších věcí a uživatel při koupi produktu nepotřebuje vývojové prostředí, ale pouze spustitelný program.

Druhý typ zpracování dat pracuje s daty za pomoci množinových operací a zabývá se čtyřmi základními otázkami:

- Jaké vlastnosti mají mít data, která chci vybrat?
- Jakou množinu dat chci vložit?
- Jakou množinu dat chci změnit?
- Za jakých podmínek chci data odstranit?

Oproti prvnímu případu, mívá databázové zpracování přívětivé grafické rozhraní pro uživatele, poskytuje víceuživatelský přístup, celková bezpečnost dat je na vyšší úrovni a vývoj aplikací je rychlejší díky pružnému řešení uživatelských změn. [1]

1.6 Normalizace dat

Pod pojmem normalizace schovává proces změn uspořádání databáze do takové podoby, aby zcela, nebo alespoň částečně, vymizela datová redundance. Dochází tak ke změnám tabulek, jejich atributů a relací mezi těmito tabulkami. Tyto úkony dost často vedou ke vzniku nových tabulek a jejich relací. Normalizace se rozlišuje v několika jejích formách. Pro účely této práce budou přiblíženy první tři z nich. [11]

- **První normální forma** – O první normální formě se u tabulek hovoří, pokud každý její atribut obsahuje pouze atomické neboli dále nedělitelné hodnoty. Je velmi důležitá pro funkčnost databáze, avšak k plnění této normální formy dochází podvědomě již při vytváření databáze. [11]
- **Druhá normální forma** – K druhému stupni normální formy dochází pouze tehdy když platí i první normální forma. Navíc musí platit závislost každého neklíčového atributu na primárním klíči, který se musí skládat z více hodnot. Pokud platí závislost pouze na části klíče, a ne na klíči celém, nejedná se o 2.NF. [12]
- **Třetí normální forma** – Ve chvíli, kdy platí druhá normální forma a v tabulkách nejsou obsaženy žádné atributy, které by podléhaly tranzitivní závislosti, jedná se o třetí normální formu. Tranzitivní závislostí se rozumí vzájemná závislost neklíčových atributů. [12]

1.7 Datová analýza

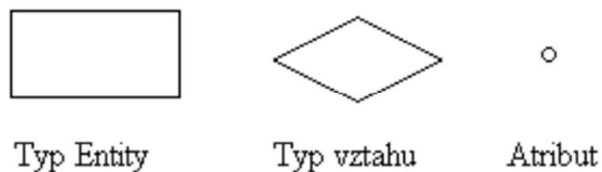
Datová analýza se zabývá zkoumáním sesbíraných dat z reálných objektů či procesů. Její jednotlivé metody se věnují vymezení obsahu datové základny a návrhu její struktury a realizace. K tomu se využívá nástrojů v podobě datových modelů. Těmito modely mohou být například E-R diagram a výskytový diagram.

Konceptuální modelování dat

Při konceptuálním modelování se vyvíjí snaha vytvořit tzv. konceptuální schéma, jehož účelem je vyjasnit strukturu dat, bez jakékoliv závislosti na skutečném uložení databáze. Tento popis je velmi úzce spjatý s realitou, jelikož jsou zde popisována fakta reálného světa. [1] K sestavování tohoto konceptuálního schématu se využívají celkem 4 typy konceptuálních modelů – Síťový, Hierarchický, Entitně-relační a Objektový. Pro účely této práce budou níže v kapitole popsány E-R diagram a Objektový diagram. [14]

E-R diagram

Jak již napovídá název, Entity-Relationship model popisuje vztahy mezi entitami a entity samotné. Tento popis může mít buď grafickou, nebo textovou podobu, přičemž grafické řešení je mnohem rozšířenější díky své názornosti. Pro vytvoření ER diagramu je zapotřebí znát jednotlivé typy entit, jejich atributy s daným klíčem a vztahy mezi entitami. Každý z těchto prvků má také svoje grafické vyobrazení, které lze vidět na obrázku níže. [14]



Obr. 1 Základní prvky E-R diagramu [14]

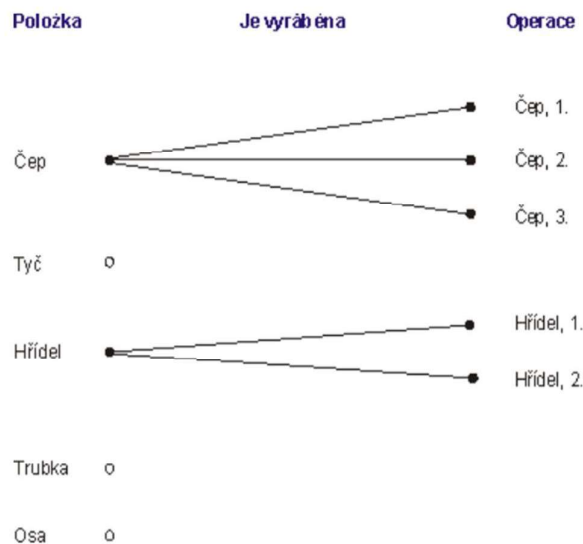
Příklad jednoduchého ER diagramu je vidět na obrázku (Obr. 2 Příklad ER diagramu) níže. Tento diagram popisuje vztah mezi dvěma entitami, konkrétně vztah „má zapsán“ mezi studentem a předmětem.



Obr. 2 Příklad ER diagramu

Výskytový diagram

Výskytový diagram se používá pro vyobrazení závislostí mezi jednotlivými entitami, zobrazuje jak kardinalitu, tak parcialitu. Zároveň je jednoduchou interpretací myšlenek analytika vůči uživateli. Příklad výskytového diagramu je uveden na následujícím obrázku (Obr. 3 Výskytový diagram). [1]



Obr. 3 Výskytový diagram [1]

Parcialita

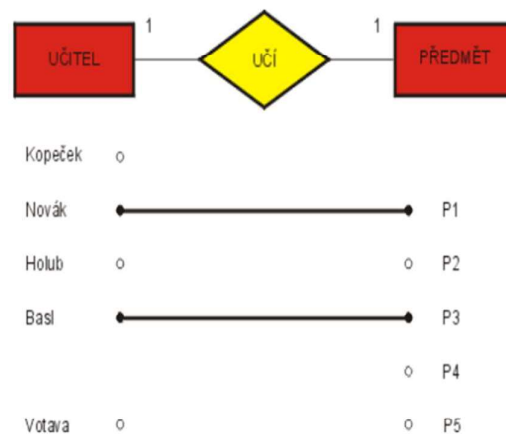
Parcialita sděluje, zda členství entity v relaci je povinné či nepovinné. Pokud se tedy jedná o členství povinné, je nutné danou entitu do vztahu vložit, v opačném případě to nutné není. Příklad lze vidět na obrázku výše (Obr.1 Výskytový diagram), kde položky čep a hřídel jsou položky vyráběné a mají tedy povinné členství ve vztahu s operací. Naopak položky tyč, osa, trubka se nevyrábí, a tudíž ve vztahu s operací zahrnuty nejsou. [1]

Kardinalita

Kardinalita zobrazuje počet entit vstupujících do vztahu. Rozeznáváme celkem tři typy, které jsou v této kapitole popsány a zobrazeny na příslušných obrázcích sestávajících se z ER a výskytového diagramů. [1]

- **Vztah 1:1**

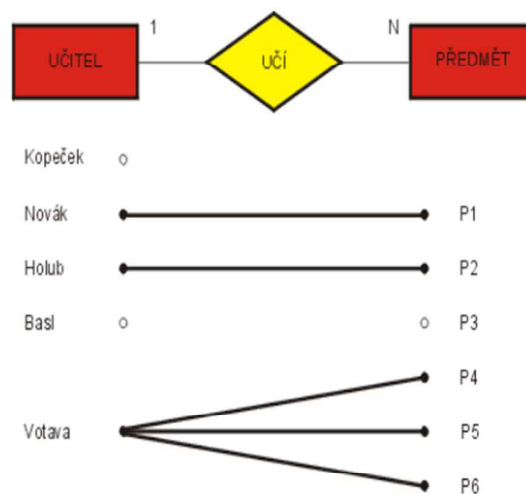
Případ vztahu 1:1 nastává v okamžiku, kdy první jedinečná entita odpovídá druhé entitě, která je též jedinečná. Jako příklad lze uvést učitele, který vyučuje pouze jediný předmět a tento předmět je zároveň vyučován pouze jedním učitelem. [1]



Obr. 4 - Kardinalita – vztah 1:1 [1]

- **Vztah 1: N**

Další typ vztah je 1: N, kdy první entita je jedinečná a odpovídá druhé entitě, která již jedinečná není a její množství výskytu je větší než 1. V takovém případě by učitel mohl již vyučovat více předmětů, ale tyto předměty mohou být stále vyučovány pouze jedním jediným učitelem. [1]

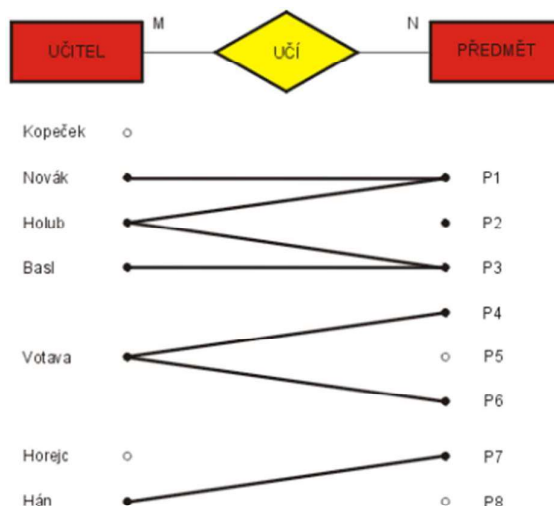


Obr. 5 Kardinalita – Vztah 1: N [1]

- **Vztah M: N**

Posledním zmíněným vztahem bude vztah M: N. Tento vztah nastává, když počet výskytů první entity neodpovídá počtu výskytů té druhé a naopak. Na příkladu učitele by to vypadalo následovně. Učitel má již na starosti výuku více než jednoho předmětu, ale tyto předměty vyučuje více učitelů.

Tento vztah je specifický tím, že ho ve velké části případů nelze v databázích použít. Je tedy třeba jeho dekompozice na dva vztahy 1: N pomocí průnikového typu, který je potřeba najít. Hledaný průnikový typ může být přirozený nebo umělý. V případě umělého průnikového typu je zapotřebí jeho vytvoření. Nejčastěji se tak děje zřízením nového pojmu. [1]



Obr. 6 Kardinalita – Vztah M: N [1]

1.8 Funkční analýza

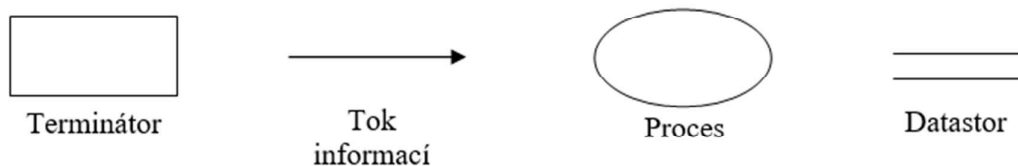
Při funkční analýze se řeší algoritmy zpracování dat a popisují datové toky. K tomu je nutné, aby existovala vstupní data. Pomocí funkční analýzy se vytváří specifikace požadavků na informační systémy. K tomu existuje metoda diagramu datových toků.

Diagram datových toků

Pomocí DFD (Data Flow Diagramu), který vychází z kontextového diagramu, se popisují funkce a toky dat informačních systémů. Lze jej však používat i k návrhu organizací nebo plánování podnikání. Tento diagram se zabývá zobrazením celkové vnitřní funkcionality systému z podkladů získaných pomocí analýzy dat interních funkcí systému s jeho okolím. Grafické zobrazení DFD pracuje se čtyřmi základními prvky

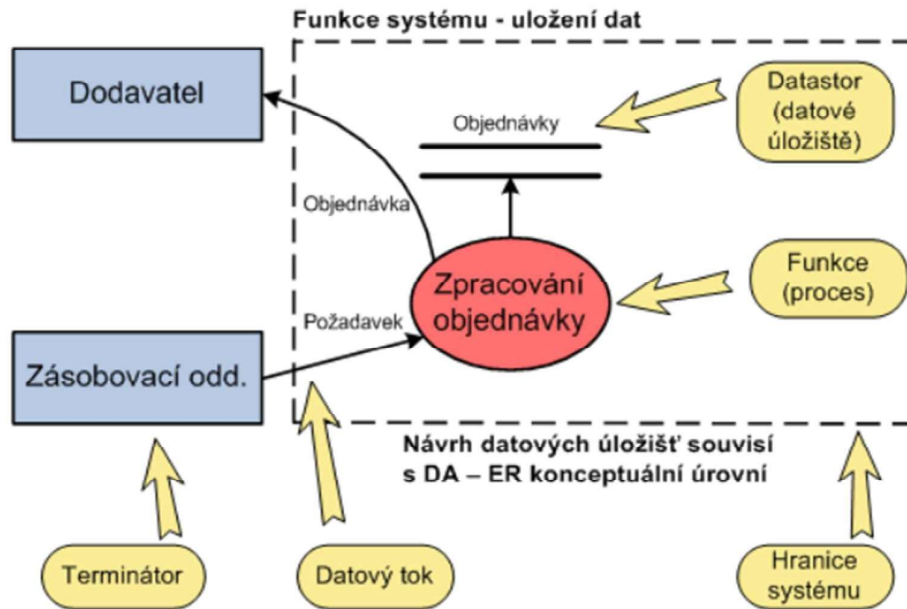
- Terminátor – zastupitel subjektu v okolí systému, př.: Dodavatel
- Tok informací – zobrazuje tok dat informací mezi terminátorem, procesem, a datastorem, př.: Požadavek
- Proces – funkce systému, který zpracovává informace, př.: Zpracování objednávky
- Datastor – zástupce databáze, př.: Objednávky

[1]



Obr. 7 Základní grafické prvky DFD [1]

Na následujícím obrázku lze vidět příklad diagramu datových toků. Tento diagram zobrazuje funkci systému uložení dat zpracované objednávky.



Obr. 8 Diagram datových toků [1]

2 Power BI

Power BI je nástroj od společnosti Microsoft, která je určená pro samoobslužné a podnikové funkce Business Intelligence. Tato platforma je škálovatelná a s její pomocí se propojují data, které je následně možné vizualizovat v aplikacích této platformy. Jedná se totiž o kolekci softwarových služeb. Tyto služby spolu kooperují a transformují nikterak související zdroje dat, ke kterým je velice snadné se připojit, na poutavé vizuály, díky kterým je možné odhalit, co je důležité. Nástroj nabízí nahrávat data z široké škály zdrojů. Mohou jimi být excelovské tabulky, ale také datové sklady. [17]

2.1 Součásti Power BI

Jak bylo zmíněno, Power BI se skládá z několika softwarových služeb, které spolu kooperují. Součásti, které je potřeba zmínit, jsou následující:

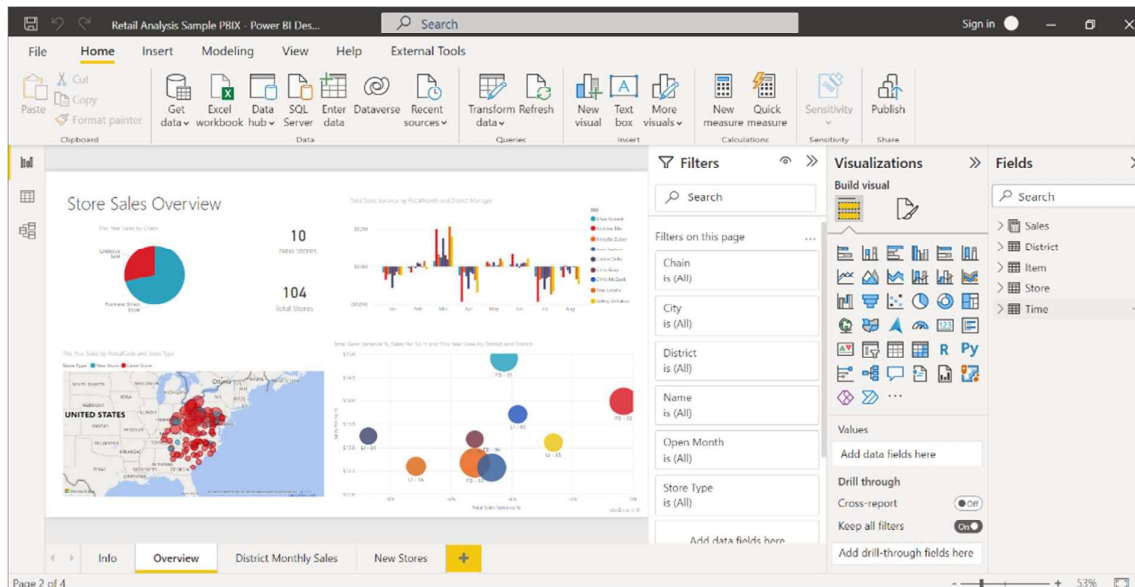
- **Power BI Desktop** – desktopová aplikace pro systém Windows
- **Služba Power BI** – online software jako služba
- **Power BI Mobile** – mobilní aplikace s podporou pro zařízení Windows, iOS, Android
- **Power BI Tvůrce sestav** – služba Power BI pro vytváření stránkových sestav
- **Server Sestav Power BI** – místní server pro sdílení sestav vytvořených v aplikaci Power BI Desktop [17]



Obr. 9 Součásti Power BI [17]

Power BI Desktop

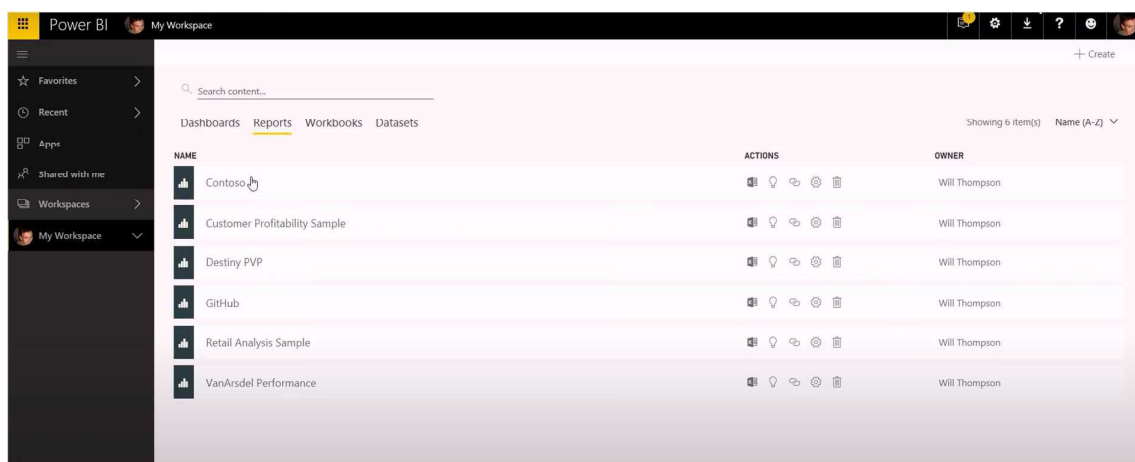
Jedná se o aplikaci, která je dostupná ke stažení zdarma pro každý počítač s operačním systémem Windows. S touto aplikací je možné se k datům jednoduše připojit, následně je transformovat a vizualizovat. Připojená data mohou být z různých zdrojů a aplikace je umožňuje mezi sebou kombinovat do datového modelu. Na základě vytvořeného datového modelu je možné vytvářet sestavy tvořené z vizuálů a jejich kolekcí. Tyto sestavy je poté možné sdílet s kolegy, zákazníky či jinak zainteresovanými osobami. [17]



Obr. 10 Power BI Desktop [17]

Služba Power BI

Tento prvek kolekce softwarů Power BI je možné dohledat i pod názvem Power BI online, který je k dispozici i na mobilních telefonech. V této službě je možné zobrazovat, sdílet a upravovat vytvořené sestavy. Ty se zde zobrazují ve formě dlaždic v řídicích panelech. V prostředí služby je k dispozici vytváření pracovních prostor pro kooperaci se spolupracovníky na jednotlivých sestavách a řídicích panelech. Tuto práci je možné poté sdílet i jako aplikaci. [17]



Obr. 11 Služba Power BI [17]

Power BI Mobile

Verzi Power BI pro mobilní zařízení je k dispozici ke stažení v mobilních obchodech s aplikacemi. V této aplikaci je možné zobrazovat vytvořené sestavy v Power BI Desktop skrze službu Power BI. Podporuje zařízení s operačními systémy iOS, Android a Microsoft. Může se tedy používat jak telefonech, tak tabletech, iPadech, a dokonce také na Apple Watch. Na těchto zařízeních jsou k dispozici nastavení pro notifikace, které se objeví v případě změny v datech. [17]

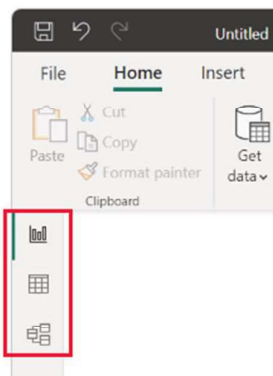


Obr. 12 Power BI Mobile [17]

2.2 Prostředí Power BI

Prostředí desktopové aplikace je možné přepnout celkem do tří zobrazení, které je možné vybrat po levé straně plochy pro vytváření sestav. Jedná se o následující možnosti:

- Sestava: Pro vytváření sestav a vizuálů
- Data: Náhled nahraných a upravených tabulek, vytvořených měř a dalších dat použitých v datovém modelu
- Model: Zobrazení pro tvoření a správu relací mezi tabulkami [17]



Obr. 13 Typy zobrazení v Power BI desktop [17]

Toto prostředí se využívá k modelování dat. Je zde možné vytvářet datové modely z nahraného zdroje dat, rozsáhlé relace a jednoduché či složitější výpočty za pomoci měř a jazyku DAX, následně také vizuály. [19]

Power Query

V tomto editoru je možné se připojit k datům z jednoho nebo více zdrojů a následně je upravovat podle své potřeby. Je možné se k němu přistoupit na kartě „Domů“ v poli „Transformovat data“. Pokud v něm nejsou načtena žádná data, zobrazí se pouze jeho prázdné podokno. Po jejich načtení se zde však objeví spousta možností pro práci s nimi. Power Query pak vypadá jako na Obr. 13 Prostředí editoru Power Query, na kterém jsou vyobrazeny body k jednotlivým částím editoru.

- Bod 1

První bod se váže k pásu karet, na kterém se nachází karty Domů, Transformace, Přidat sloupec, Zobrazení, Nástroje. Pod všemi kartami se nachází funkce potřebné k úpravě dat.

- Bod 2

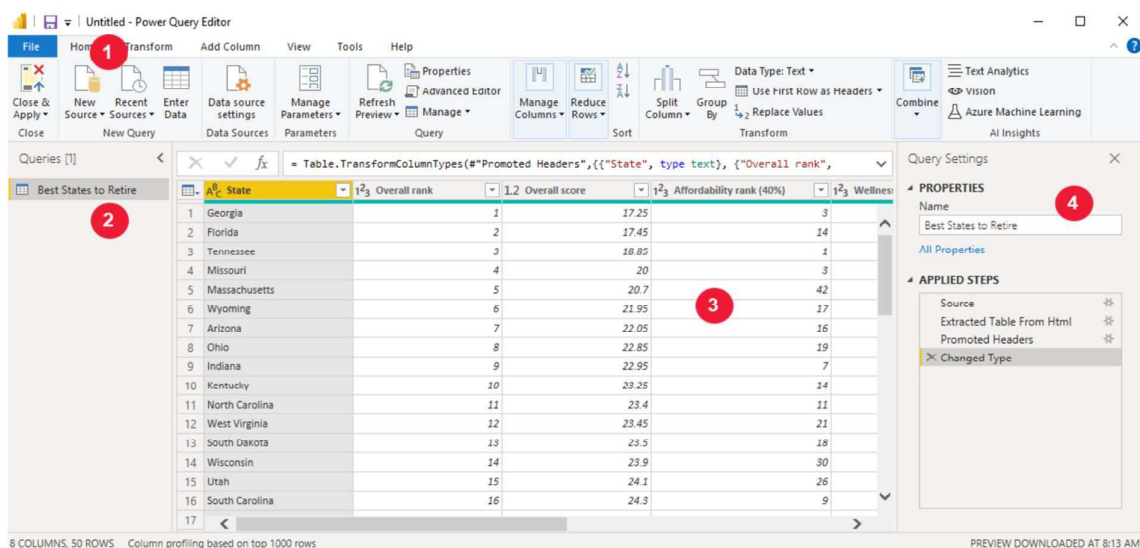
Po levé straně zobrazeného okna se nachází pás s dotazy. V tomto pásu se zobrazí dlaždice všech nahraných zdrojů, tabulek, se kterými se bude nadále pracovat.

- Bod 3

Prostor uprostřed okna je pak vymezený pro nahraná data vybraného zdroje, které jsou zde zobrazeny a k dispozici pro jejich úpravu.

- Bod 4

Po pravé straně okna se poté nachází podokno Nastavení. V tomto podokně se nachází vlastnosti dat a zobrazení provedených kroků, které se v tomto podokně dají odstraňovat nebo upravovat. [18]



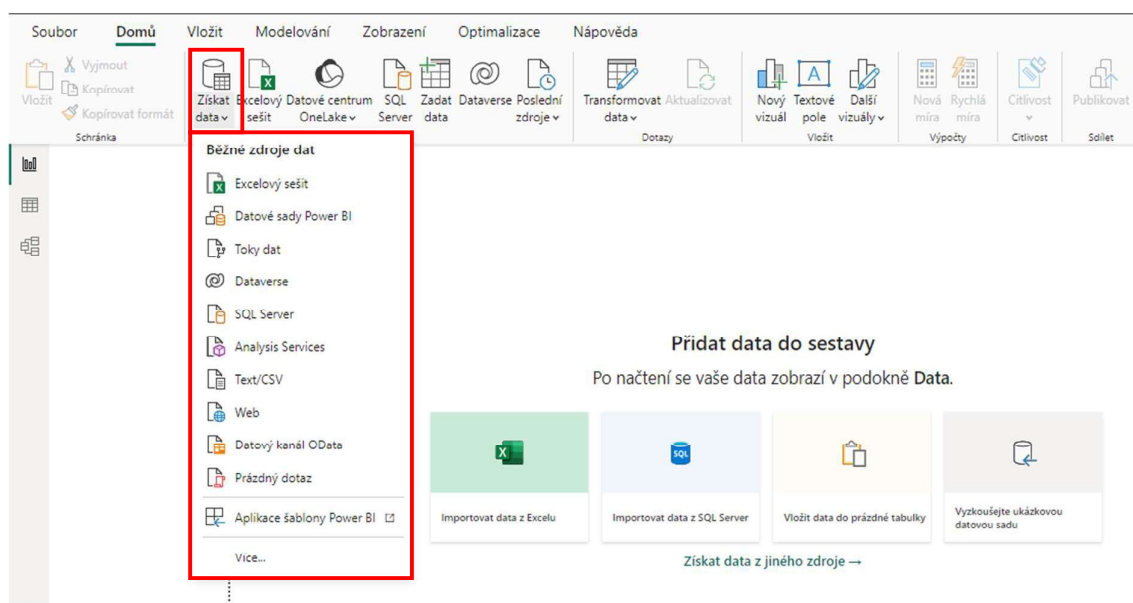
Obr. 14 Prostředí editoru Power Query [18]

2.3 Základní operace

Mezi základní operace při práci s daty v aplikaci Power BI desktop patří jejich nahrání, transformace, tvorba relací, tvorba měř a výpočtů a neposlední řadě jejich vizualizace. Všechny tyto úkony je možné provést v Power Query v prostředí aplikace Power BI Desktop. [13]

2.3.1 Nahrání dat

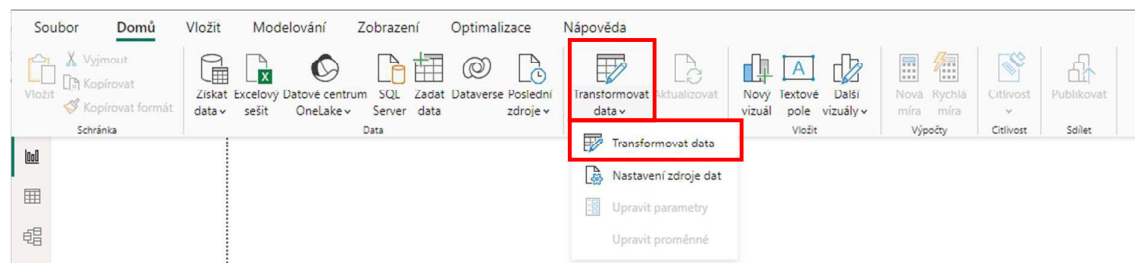
Power BI umožňuje import dat z mnoha zdrojů. Těmito zdroji mohou být například excelovský sešit a SQL Server, dále je možné nahrávat data z webu, CSV, JSON, ale i dalších. K tomuto nahrání dat je možné přistoupit ihned po otevření desktopové aplikace, kdy se zobrazí v pracovním okně hlavní možnosti nahrání dat a pod nimi odkaz „Získat data z jiného zdroje“. Tato možnost není jedinou, jak lze data vložit. K procesu nahrání dat je možné také přistoupit na kartě „Domů“ v záložce „Získat data“. Zde je možné provést výběr z běžných zdrojů nebo z rozšířené nabídky po kliknutí na možnost „Více...“. [13]



Obr. 15 Nahrání dat v Power BI Desktop

2.3.2 Transformace dat

Pod tímto pojmem si lze představit čištění dat a jejich úpravu. Nejčastějšími úkony spojených s transformací může být odstranění duplikací, chyb, sloupců, rozdělení sloupců, změna datového typu sloupce atd. Tyto úkony je možné provést v prostředí editoru Power Query po načtení dat. Tento editor je k nalezení na kartě „Domů“ po kliknutí na záložku „Transformovat data“. [13]

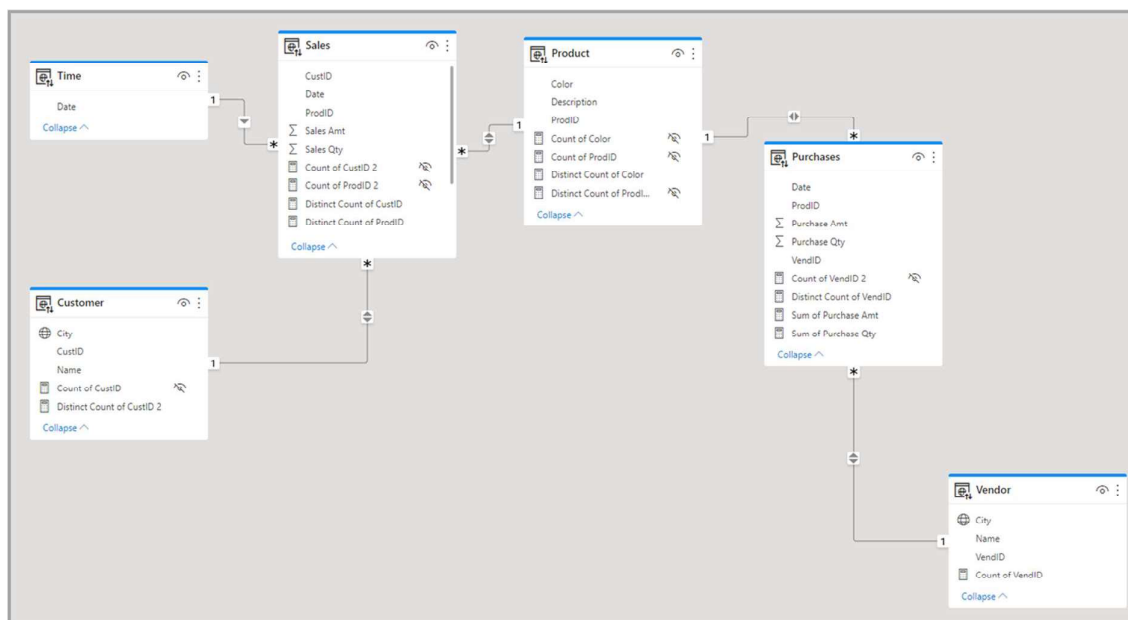


Obr. 16 Transformace dat v Power BI Desktop

2.3.3 Tvorba relací

Ve většině případů se pracuje s více tabulkami a je nutné je mezi sebou propojit. Tato propojení nazýváme relace a spojením tabulek pomocí relací vzniká datový model. Tento model se skládá z tabulek, sloupců, datových typů a již zmíněných relací. Správné vytvoření modelu je velmi důležité pro výslednou analýzu dat, která se tím může značně urychlit. [13]

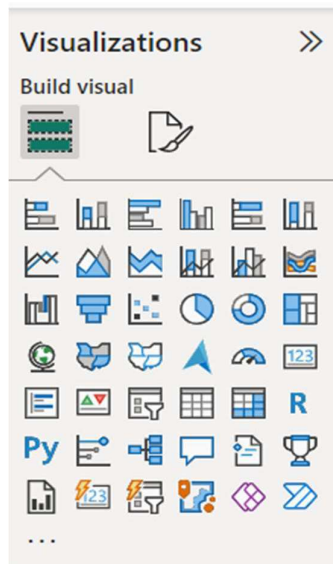
Power BI Desktop se v případě načtení dvou a více tabulek pokusí vytvořit relace automaticky na základě dotazovaných tabulek a jejich názvů sloupců. Nastaví možnosti kardinality, směr křížového filtru a aktivitu relace. Tyto automatické návrhy je možné poté v prostředí modelu upravovat. Pokud se však neprokáže vysoká míra spolehlivosti, aplikace relace nevytvoří a bude nutné relace vytvořit ručně. To je možné provést na kartě „Modelování“>„Spravovat relace“>„Nové“. [20]



Obr. 17 Tvorba relací v Power BI Desktop [20]

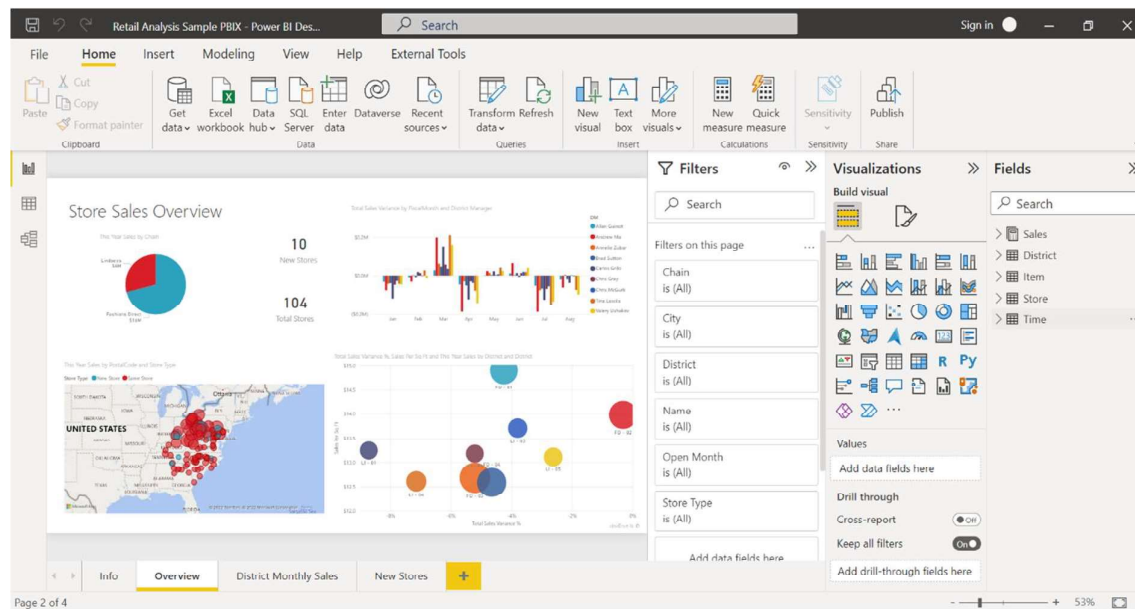
2.3.4 Vizualizace

Posledním krokem je vytvoření sestav. Sestavy se skládají z kolekcí, tedy z jednoho či více vizuálů, ve kterých jsou různé pohledy na získaná data. Vytvořené sestavy je možné vytvářet na jednu i více stránek, podle rozsahu získaných dat a informací nutných k zobrazení. V aplikaci Power BI Desktop existuje pro tuto operaci sada s různými možnostmi předdefinovaných vizuálů. Jsou k nalezení při tvorbě sestavy v podokně „Vizualizace“. Na výběr je zde z několika typů grafů, ale také tabulek, map, filtrů atd. [17]



Obr. 18 Typy vizuálů v Power BI Desktop [17]

Po vybrání konkrétního typu se vizuál přenesne na plochu sestavy. V jeho vlastnostech pod nabídkou vizuálů jsou pole, do kterých se vkládají vymodelovaná data, po jejichž vložení se daný vizuál těmito daty zaplní. [17]



Obr. 19 Vizualizace v Power BI Desktop [17]

3 Analýza současného stavu

Diplomová práce je zpracována pro společnost ALIMEX NEZVĚSTICE a.s., která sídlí nedaleko města Plzeň a zabývá se zemědělskou výrobou.

3.1 Popis společnosti

ALIMEX NEZVĚSTICE a.s. je společností, která je členem koncernu AGROFERT řízeného společností AGROFERT a.s., vznikla v roce 1999.

Společnost se zabývá zemědělskou výrobou. Obhospodařuje zemědělskou půdu v 39 katastrálních územích o celkové rozloze 4500 m² a k této potřebě zaměstnává více než 106 zaměstnanců. Vede živočišnou a rostlinnou výrobu, jež jsou oblastí hlavního zdroje příjmu. Dalším zdrojem je bioplynová stanice Žákava, ve které dochází k výrobě elektrické energie.

Společnost ALIMEX NEZVĚSTICE a.s. využívá uzavřený obrat stáda skotu k hospodaření v oblasti živočišné výroby. Tento okruh výroby má celkem k dispozici cca 1100 kusů dojnic k produkci mléka z Holštýnského plemena. Tato produkce se skládá z provozu v VKK Číčov s kapacitou 500 kusů dojnic a se stejným počtem dojnic z provozu v VKK Žákava. Společnost dále provozuje odchovny mladého dobytka v Planinách a Nezvěsticích s celkovým počtem 620 kusů a stádo krav bez tržní hodnoty produkce mléka ve středisku v Louňové s přibližným počtem 160 kusů.

Rostlinná výroba se sestává z 4500 ha pozemků, z kterého cca 3100 ha tvoří orná půda. Tato společnost využívá svých vlastních strojů a další techniky pro sklizeň jak polních plodin, tak obilovin, luk a krmných plodin. Disponuje také prostory pro čištění, sušení a skladování produkce obilovin a olejnin, včetně dopravní techniky. Součástí rostlinné výroby je provoz ovocného sadu o rozloze areálu 35 ha v katastru obce Těnovice. Jsou zde pěstovány jabloně letních, podzimních a zimních odrůd. Dále jsou zde pěstovány rané až pozdní odrůdy švestek a pološvestek. Pro sesbírané ovoce jsou využívány termosklady s ideálními podmínkami pro jeho skladování. [22]



Obr. 20 Logo společnosti ALIMEX NEZVĚSTICE a.s. [22]

3.2 Popis výrobních dat

Společnost ALIMEX NEZVĚSTICE a.s. dodala k analýze data za období červen až září, která jsou uložena v tabulkách do devíti excelovských souborů pod názvy:

- WF_přehled mechanizace
- WF_IQ vnitro
- WF_spotřeba nafty
- WF_spotřeba PHM
- spotřeba nafty
- spotřeba PHM
- CF_cerven_zari_alimex
- plodiny_LPIS_alimex
- VO_alimex

Jednotlivé soubory jsou níže slovně popsány a vizualizovány pomocí jednoduchých tabulek, které obsahují název sloupce a jeho datový typ.

WF_přehled mechanizace

V tomto souboru se nachází přehled o využívaných strojích k výrobě. Je zde celkem 23 sloupců obsahující podrobnosti o jednotlivých zařízeních jako jsou název, kód, klíč, druh zařízení, palivo, SPZ a informace o GPS jednotce. Vyskytuje se zde několik nepotřebných sloupců, které mohou odstraněny. Tyto sloupce jsou totiž nevypovídající, z velké části obsahují prázdné hodnoty (null), nebo po dohodě se společností byly vyřazeny jako nepotřebné.

Jedná se o následující sloupce: n5803uzit_, n5803celk_, n5803prohm, n5803pnapr, n5803kateg, n5803motty, n5803ccm_, n5804koddv, n5804nazev, n5803vykon, n5803paliv, n5802nazev

Tabulka 1 WF_přehled_mechanizace – struktura

WF_přehled mechanizace	
Atribut	Datový typ
n5802kod__	string
n5802nazev	string
komb_klic	string
c5809nazev	string
n5803druh	string
n5803paliv	string
n5803spz__	string
n5803nazev	string
n5803uzit_	float
n5803celk_	float
n5803prohm	integer
n5803pnapr	float

n5803kateg	string
n5803motty	string
n5803ccm__	float
n5804koddv	integer
n5804nazev	string
n5803zaber	float
n5803podvo	string
n5803kodje	string
n5803telje	string
n5803vykon	integer
n5803id_wd	integer

WF_IQ vnitro

Tato tabulka je účetní výkaz, který obsahuje data ve vztahu má dáti, dal. Záznamy jsou zapisovány do dvanácti sloupců. Nachází se zde čísla a názvy účtů má dáti, dal. Následně se zde najde ke každému účtu kombinace klíčů, která je rozdělena do tří sloupců vždy s názvem konkrétní části. Vzhledem k obsahu dat se nebude tato tabulka při tvorbě vizuálů využívat. Může však posloužit při tvorbě číselníků, pro které by se dalo využít názvů obsažených v jednotlivých částech klíče.

Tabulka 2 WF_IQ vnitro – struktura

WF_IQ vnitro	
Atribut	Datový typ
kniha	string
ucet_md	string
naz_uctu_md	string
ucet_dal	string
naz_uctu_dal	string
cklic_2r1_md	string
cklic_3r2_md	string
cklic_4r3_md	string
cklic_2r1_dal	string
cklic_3r2_dal	string
cklic_4r3_dal	string
castka	float

Spotřeba nafty

Tabulka zobrazující množství spotřebované pohonné hmoty v jednotlivých měsících konkrétními stroji. Jelikož se zde zaznamenává jednotková cena pohonné hmoty, lze dopočítat finanční náklady stroje v daném období. Tabulka obsahuje kombinaci klíčů, který se skládá ze tří částí a odkazuje tak na místo stroje, skupinu stroje a konkrétní stroj.

Tabulka 3 Spotřeba nafty – struktura

Spotřeba nafty	
Atribut	Datový typ
polozka	string
nazev	string
pohyb	string
mesic	date
doklad	string
mnozstvi	integer
mj	string
skladova_cena	float
skl_cena_jedn	float
kombinace_kl	string

Spotřeba PHM

Jedná se o tabulku podobnou tabulce „spotřeba nafty“. Rozdíl je zde v tom, že tato tabulka obsahuje podrobně rozepsanou kombinaci klíčů, kdy ke každé číselné podmnožině je přiřazen její název.

Tabulka 4 Spotřeba PHM - struktura

Spotřeba PHM	
Atribut	Datový typ
polozka	string
nazev	string
pohyb	string
mesic	date
doklad	string
mnozstvi	integer
mj	string

skladova_cena	float
skl_cena_jedn	float
cklic_2r1	string
klic_2r1	string
naz_klice_2r1	string
cklic_3r2	string
klic_3r2	string
naz_klice_3r2	string
cklic_4r3	string
klic_4r3	string
naz_klice_4r3	string
sp1_klice_3	string
ns1_klice_3	string
kombinace_kl	string

WF_spotřeba nafty a WF_spotřeba PHM

Soubory WF_spotřeba nafty a WF_spotřeba PHM jsou obsahově stejné jako soubory spotřeba nafty a spotřeba PHM. Jedná se pouze o vyexportované tabulky z jiného zdroje dat. Konkrétně ze systému Win Fas.

Tabulka 5 WF_spotřeba nafty – struktura

WF_spotřeba nafty	
Atribut	Datový typ
polozka	string
nazev	string
pohyb	string
mesic	date
doklad	string
mnozstvi	integer
mj	string
skladova_cena	float
skl_cena_jedn	float
kombinace_kl	string

Tabulka 6 WF_spotřeba PHM – struktura

WF_spotřeba PHM	
Atribut	Datový typ
polozka	string
nazev	string
pohyb	string
mesic	date
doklad	string
mnozstvi	integer
mj	string
skladova_cena	float
skl_cena_jedn	float
cklic_2r1	string
klic_2r1	string
naz_klice_2r1	string
cklic_3r2	string
klic_3r2	string
naz_klice_3r2	string
cklic_4r3	string
klic_4r3	string
naz_klice_4r3	string
sp1_klice_3	string
ns1_klice_3	string
kombinace_kl	string

CF_cerven_zari_alimex

Tabulka CF_cerven_zari je knihou zaznamenaných jízd a jedná se tedy o tzv. faktovou tabulku. V této tabulce jsou zaznamenány jednotlivé jízdy, které byly provedeny. Je zde uvedeno datum a trvání jízdy s konkrétním řidičem a strojem. U každé jízdy je také uvedeno její místo, typ operace, název výrobní operace a ujetá vzdálenost v rámci akce.

Tabulka 7 CF_cerven_zari_alimex – struktura

CF_cerven_zari_alimex	
Atribut	Datový typ
Datum	date
Od	time
Trvání	time
Jméno řidiče	string
Farma řidiče	string
Typ operace	string
Výrobní operace	string
Uživatel	string
Blok	string
Parcela	string
Rozloha parcely	float
Plocha operace	float
Ujetá vzdálenost	float
Bonus	integer
Název stroje	string
SPZ	string
Název přídatného zařízení	string
Farma přídatného zařízení	string
Plodina	string
Středisko	string

Plodiny_LPIS_alimex

Přehled plodin ve vazbě na určité místo. Obsahuje jedenáct sloupců, ve kterých se nachází informace o katastrálních územích, obhospodařovaných společností ALIMEX NEZVĚSTICE a.s., s jejich rozpadem na okres a konkrétní obec. K jednotlivým obcím jsou poté přiřazeny plodiny, které jsou zde pěstovány a výměra ploch, na kterých jsou zasety. Každé katastrální území má svoje pořadové číslo, kód KÚ, čtverec, kód čtverce.

Tabulka 8 Plodiny_LPIS_alimex – struktura

Plodiny_LPIS_alimex	
Atribut	Datový typ
Poř.č.	string
Čtverec	string
Kód	string
Kat.území	string
Kód KÚ	string
Okres	string
Obec	string
Plod. ID	string
Plodina	string
Výměra	float
Pěstování od	date

VO_alimex

VO_alimex je číselníkem výrobních operací. Obsahuje celkem tři sloupce datového typu string, kde je uvedené ID výrobní operace, její název a měrná jednotka.

Tabulka 9 VO_alimex – struktura

VO_alimex	
Atribut	Datový typ
id_vyr_op	string
nazev_vo	string
kod_mj	string

3.3 Problémy s daty

V datech, která byla společností poskytnuta, se vyskytují určité problémy, které zabraňují jejich přehledné vizualizaci. Nejčastěji se naráží na problém spojený s konzistencí dat, který je způsoben ručním zadáváním záznamů a dochází tak k různým překlepům, zkratkám nebo dokonce záměnám. Tato nekonzistentní data se ve vizualizaci promítnou ve více záznamech místo jednoho unikátního, což ovlivní výsledný výstup.

Dále se zde naráží na problém s propojením jednotlivých tabulek mezi sebou. To je způsobeno tím, že jednotlivé tabulky neobsahují unikátní identifikační klíče z jiných tabulek. Bez tohoto propojení není možné následně vytvořit v Microsoft Power BI vizuály, které by obsahovaly data a výpočty napříč těmito tabulkami.

WF_přehled_mechanizace

Prvním problémem v této tabulce je její druhý řádek, který musí být odstraněn. Jedná se o pomocnou položku pro ukládání tankovacích záznamů. Dále se v této tabulce vyskytují sloupce s velkou neúplností dat. Některé ze sloupců však obsahují nepovinné údaje a pro analýzu nebudou využity.

Za velký problém se považují sloupce s názvy strojů. Položky v těchto sloupcích totiž obsahují více hodnot, přičemž neexistuje jejich jednotný formát. Dochází například k tomu, že název stroje má několik různých podob. Zároveň tyto položky obsahují název stroje, jeho typ, SPZ, středisko, ale ne vždy jsou všechny tyto hodnoty zastoupeny. Východiskem z tohoto problému je atomizace dat neboli rozdělení těchto hodnot do více sloupců a jasná definice názvu a formátu jednotlivých položek.



n5802nazev
Škoda Oct.com 9P2 2399 Voříšek
Škoda Yeti 1UM 9753 p. Hajný
škoda Karoq 2,0 TSI 7P3 3449
škoda octavia 1,4TSI 6P3 86-76
Škoda Fabia-4Z5 2793-Červinkov
Škoda Octavia-6P0 0330-p.Bečv
ŠKODA FABIA 1P2 9458-RV Vlková
Š.FABIA COMBI-1SZ8041-RV Vlkov
ŠKODA FABIA COMBI-1SZ7931-Ober

Obr. 21 WF_přehled_mechanizace – názvy strojů

Sloupce SPZ poté obsahují nestandardní označení jako „BOBŽák“. Jedná se o SPZ strojů, které nejsou určeny k provozu na veřejných komunikacích. Takové stroje by neměly obsahovat SPZ.

Problémy se také vyskytují ve sloupci s kombinací klíčů. Klíče se mají skládat ze tří částí, ale některé se sestávají pouze ze dvou, např.: 960-100, přičemž u některých není klíč vůbec uveden. Není možné je poté propojit s klíčem stroje v jiných tabulkách. Problém s klíči byl poté u evidence tzv. kelímků, které jsou evidovány dvěma různými způsoby a je potřeba je sjednotit. Další problémy se váží k zařazení více strojů na jeden klíč jako tomu je například u kombinace klíče 890-100, který se váže ke čtyřem strojům.

V tabulce se navíc vyskytuje poměrně vysoká neúplnost dat. To se děje například ve sloupci n5803paliv, který obsahuje typ paliva daného stroje. V tomto sloupci se vyskytuje téměř stoprocentní neúplnost záznamů. Další problémy s neúplností dat jsou vyobrazeny v tabulce níže.

Tabulka 10 Neúplnost dat – WF_přehled_mechanizace

Název sloupce	% Neúplnost
n5803paliv	99
n5803spz	80
n5803uzit	100
n5803celk	98
n5803prohm	100
n5803pnapr	99
n5803kateg	99
n5803motty	100
n5803ccm	99
n5804nazev	97
n5803zaber	1
n5803podvo	1
n5803kodje	77
n5803telje	80
n5803vykon	85
n5803id_wd	34

CF_cerven_zari_alimex

Prvním problémem vyskytující se v této tabulce jsou položky, které obsahují více hodnot a měly by tak být rozděleny do více sloupců. Konkrétně se jedná o sloupec:

- Výrobní operace
- Název stroje
- Název přídatného zařízení

Výrobní operace vždy obsahují předponu VO, která tím pádem může být odtržena a následně odstraněna. Za touto předponou je umístěna pomlčka, za kterou se již nachází název operace. Dochází zde také k tomu, že název operace se skládá z několika z několika částí. Příkladem může být položka „VO-Doprava-ZELENÁ-Dovoz a odvoz z pole, odvoz hnoje na pole“, „VO-Doprava-NE ZELENÁ-Dovoz materiálu, převozy“. Takovéto záznamy se dají rozložit do více sloupců, které by rozlišovaly, zda se jedná o Dopravu, zda je ZELENÁ/NEZELENÁ a v posledním sloupci by byl umístěn název operace.

Obdobné řešení by se vztahovalo na sloupec Název stroje a Název přídatného zařízení. V obou případech se vyskytují záznamy, které se dají dále rozebrat na více položek. Jako příklad lze uvést název stroje „NEW HOLLAND, T5.115“, který se dá rozdělit na název výrobce stroje „NEW HOLLAND“ a řadu stroje „T5.115“. Stejný postup by se aplikoval u názvu přídatného zařízení.

Druhým problémem, kterého si lze všimnout, je poměrně velká neúplnost dat. Vyskytují se zde dva sloupce, Farma řidiče a Farma přídatného zařízení, které neobsahují jediný záznam a nejsou tak vůbec využity. Dále je zde Název přídatného zařízení, které nemusí být vždy potřeba a nevyžaduje tak nutně záznam. Ovšem se zde vyskytují jiné sloupce bez záznamu, obzvláště Jméno řidiče, který by byl ve vizualizovaných přehledech zapotřebí, ale jejich neúplnost je příliš vysoká. Přehled procentuální neúplnosti dat je k náhledu v následující tabulce.

Tabulka 11 Neúplnost dat – CF_cerven_zari

Název sloupce	% Neúplnost
Jméno řidiče	15
Farma řidiče	100
Uživatel	78
Blok	78
Parcela	81
Název přídatného zařízení	74
Plodina	81
Středisko	81

WF_IQ vnitro

V této tabulce se vyskytuje celkem šest sloupců obsahující klíč a název tohoto klíče, např.: 890-traktory a mechanizace. Konkrétně se tento stav týká sloupců cklic_2r1_md, cklic_3r2_md, cklic_4r3_md, cklic_2r1_dal, cklic_3r2_dal a cklic_4r3_dal. Lze si tedy povšimnout, že se zde nachází v jedné buňce více než jeden záznam a mělo by zde dojít k rozdělení na dva atributy, který by mohlo být číslo a název klíče.

Největší problémy se týkají atributů cklic_4r3_md a cklic_4r3_dal. Zde dochází ke sloučení klíče, který je šestimístný, označení stroje, názvu stroje, geografické zařazení a jeho SPZ a zároveň dochází v záznamu k proměnlivému zastoupení těchto hodnot. Na rozdíl od předešlých sloupců, které se vyplňovaly ve formátu klíč, pomlčka, název, např.: 100 – Rostlinná výroba, jsou tyto atributy spojené bez pomlčky nebo jakékoliv mezery mezi číselným označením klíče a následným názvem, např.: 119899kol.tr.NH T7.315 HD P02 6143, což působí zmatečně.

Dalším problémem v posledně zmiňovaných sloupcích je nejednotnost dat, kdy dochází k několika různým vyjádřením jednoho a toho samého názvu stroje. To se například týká názvu stroje New Holland, který se kromě již zmíněné podoby vyskytuje také jako NEW HOLLAND nebo NH. Pokud by se tyto záznamy nesjednotily a ponechaly v tomto stavu, Power BI by s nimi počítalo a zobrazovalo je jako tři různé hodnoty.

cklic_2r1_dal	cklic_3r2_dal	cklic_4r3_dal
100-Rostlinná výroba	890-traktory a mechanizace	119898-New Holland Žákava - P00 2518
100-Rostlinná výroba	890-traktory a mechanizace	119893-NEW HOLLAND-1107 P013466
100-Rostlinná výroba	890-traktory a mechanizace	119898-New Holland Žákava - P00 2518
100-Rostlinná výroba	890-traktory a mechanizace	119893-NEW HOLLAND-1107 P013466
100-Rostlinná výroba	890-traktory a mechanizace	119915-New Holland P02 0655
100-Rostlinná výroba	890-traktory a mechanizace	119892-JOHN DEERE P00 2166/07
100-Rostlinná výroba	890-traktory a mechanizace	119893-NEW HOLLAND-1107 P013466
100-Rostlinná výroba	890-traktory a mechanizace	319894-JOHN DEE.7530-P002222/08
100-Rostlinná výroba	890-traktory a mechanizace	119890-kol.tr.NH Genesis P029848
100-Rostlinná výroba	890-traktory a mechanizace	119915-New Holland P02 0655

Obr. 22 Problémy s daty - WF_IQ vnitro

Neúplnost dat se v tomto souboru týká pouze sloupců cklic_4r3_md a cklic_4r3_dal, ve kterých je procentuální neúplnost 83 a 11 procent.

Tabulka 12 Neúplnost dat – WF_IQ vnitro

Název sloupce	% Neúplnost
cklic 4r3_md	83
cklic 4r3_dal	11

Spotřeba PHM

V sešitu spotřeba PHM se vyskytují dva listy, List1 a spotřeba PHM. Z těchto dvou listů se bude využívat pouze list spotřeba PHM, který obsahuje navíc názvy jednotlivých klíčů. A právě tyto názvy jsou u strojů opět problém, jelikož se znovu skládají z několika hodnot (klíče, označení stroje, názvu stroje, geografické zařazení, SPZ) a navíc je jejich zastoupení opět proměnlivé. Tyto záznamy je tedy potřeba atomizovat a rozdělit do více sloupců.

Firmou zasláná data obsahovala soubor pod názvem WF spotřeba PHM, který byl totožný se souborem Spotřeba PHM, a tudíž se v něm vyskytují stejné problémy s daty. Dále se zde vyskytovaly obdobné excelovské sešity pod názvem WF spotřeba nafty a Spotřeba nafty, které jsou podobné souborům WF spotřeba nafty a Spotřeba nafty, ale neobsahují názvy jednotlivých částí klíče. Struktura těchto tabulek byla popisována v kapitole Popis výrobních dat.

VO_alimex

V souboru VO_alimex se vyskytují výrobní operace, které je potřeba, podobně jako názvy strojů v předchozích tabulkách, rozdělit na více sloupců. V tomto případě se ve sloupci pod názvem nazevo vyskytují záznamy složené z předpony VO, názvu skupiny operace a názvem konkrétní operace, např.: VO-Doprava-Doprava vody. Takovéto záznamy se opět musí atomizovat a rozdělit na více atributů.

Plodiny_LPIS_alimex

V tomto číselníku se vyskytuje prázdný sloupec Plod.ID, který by měl fungovat jako identifikační klíč a případně ho propojit s jinými tabulkami. Tím, že se zde nevyskytují žádná

data, není možné tuto tabulku možné s jinými propojit. Pro možnost propojení s jinými tabulkami je tedy nutné zavést identifikační klíč.

3.4 Transformace dat

K vizualizaci dat je potřeba nahrát jejich zdroje do Microsoft Power BI a poté je vyčistit a transformovat. Následně se úpravy dat uloží a bude s nimi možné vizuálně pracovat.

WF_přehled_mechanizace

Jako první byla transformována tabulka WF_přehled_mechanizace, která je číselníkem strojů. Zde se vyskytuje několik nepotřebných sloupců, které jsou odstraněny. Tyto sloupce jsou totiž nevypovídající, z velké části nulové, tedy nevyskytují se zde žádné záznamy, nebo po dohodě se společností byly vyřazeny jako nepotřebné.

Jedná se o sloupce:

- n5802kod
- n5803uzit_
- n5803celk_
- n5803prohm
- n5803pnapr
- n5803kateg
- n5803motty
- n5803ccm__
- n5804koddv
- n5804navez
- n5803vykon
- n5803paliv
- n5802navez
- n5803zaber
- n5803podvo
- n5803telje
- n5803id_wd

Vzhledem k tomu, že se jedná o číselník, je zapotřebí, aby se zde vyskytoval vždy pouze jeden záznam stroje. Z toho důvodu byly odebrány duplicitní kombinace klíčů a SPZ stroje, čímž počet záznamů klesl z 346 na 89. Následně proběhlo nahrazení chyb vyskytujících se v záznamech jako „Error“ hodnotou null, čímž se umožní alespoň částečná práce s těmito záznamy. Ve sloupci s kombinacemi klíčů se vyskytují položky, které klíč vůbec neobsahují. Takové položky jsou odebrány, jelikož jsou neidentifikovatelné, není možné je propojit s jinými tabulkami a následně použít při tvorbě vizuálů. Úplně stejně se odfiltrují klíče služebních vozů externích pracovníků, které se skládají pouze ze dvou částí namísto tří. Tyto klíče v jiných tabulkách nelze dohledat, neobsahují tedy žádné záznamy, a není možné je dále využít. Zároveň je nutné změnit formát kombinace klíčů v této tabulce. Ostatní tabulky totiž obsahují tuto hodnotu ve formátu xxx-xxx-xxxxxx, zatímco zde se vyskytuje v jeho zrcadlové podobě, tedy xxxxxx-xxx-xxx. Tato náprava proběhla rozdělením sloupce na základě konkrétního znaku, kterým je zde pomlčka. Touto operací se sloupec rozdělí na tři a následně

se sloučí zpět v již preferovaném pořadí. Další vhodnou úpravou je rozdělení názvu značky stroje a jeho řady ve sloupci n5803navez. Opět se tak využije rozdělení sloupce, tentokrát však pomocí čárky. Vzniknou dva nové sloupce, které se pojmenují Znacka_stroje a Rada_stroje. Úplně na závěr se opraví překlepy v těchto názvech a dojde tak k jejich sjednocení. Konkrétně se jedná o překlep NEW HOLLNAD na NEW HOLLAND, Žákavá na Žákava a Škoda na ŠKODA.

Po výše zmíněných úpravách se v tabulce budou vyskytovat sloupce:

- Komb_klicu
- n5803spz__
- Znacka_stroje
- Rada_stroje
- c5809navez
- n5803druhv

Tyto sloupce jsou pro další práci s daty nejdůležitější, jelikož obsahují všechna potřebná data. Kombinace klíčů se využije k tvorbě relací a ostatní sloupce především při tvorbě vizuálů, ve kterých bude možné zobrazovat analýzy na základě názvu stroje, jeho skupiny či druhu.

Komb_klic	n5803spz__	c5809navez	n5803druhv	Znacka_stroje	Rada_stroje
201-861-112743	1S27931	Osobní automobil	Energetické	ŠKODA	Fabia
100-861-132744	1S28041	Osobní automobil	Energetické	ŠKODA	Fabia
100-861-112740	1P29458	Osobní automobil	Energetické	ŠKODA	Fabia
101-890-199943	P013751	Traktor 100 kW a méně	Energetické	FENDT	329,T2
100-892-114055	P002760	Sklízecí řezačka	Samojízdné	NEW HOLLAND	FR 600
100-891-114607	P029303	Sklízecí mlátička	Samojízdné	NEW HOLLAND	CX8.80 SL
202-892-227539	BOBČřř	Zastýlací vůz	Samojízdné	Comfort	Čřřov
201-890-217543	BOBŽák	Zastýlací vůz	Samojízdné	Promax	Žákava
100-891-114610	P026864	Sklízecí mlátička	Samojízdné	NEW HOLLAND	CX8.80
202-892-217544	P033478	Krmný vůz	Samojízdné	STRAUTMANN	Verti-mix
201-892-217817	UNC060Žák	Kolový nakladač	Samojízdné	UNC-060	
202-892-227812	UN050Čřř	Kolový nakladač	Samojízdné	UN-050	
100-892-137819	P013455	Manipulátor	Samojízdné	KRAMER	4107,400
100-892-317820	P013473	Manipulátor	Samojízdné	KRAMER	4507
207-892-277821	P013481	Kolový nakladač	Samojízdné	SCHAFFER	690T
100-890-139899	P002519	Traktor 100 kW a méně	Energetické	NEW HOLLAND	JHT50.70
100-890-119898	P002518	Traktor 100 kW a méně	Energetické	NEW HOLLAND	T5070
100-890-139900	P013764	Traktor 300 kW a více	Energetické	NEW HOLLAND	T8.360
100-890-119901	P013768	Traktor 101 až 200 kW	Energetické	NEW HOLLAND	T7.270
100-892-317821	P002588	Manipulátor	Samojízdné	NEW HOLLAND	LM 5080,KA
100-860-312518	5P38592	Nákladní vozidlo	Energetické	LIAZ	150 809

Obr. 23 WF_přehled_mechanizace po transformaci dat

WF_IQ vnitro

Vzhledem ke skutečnosti, že tabulka WF_IQ vnitro je účetním výkazem, není pro účely této práce relevantní a může být odstraněna. Pokud by s ní mělo být operováno, bylo by nutné zde vytvořit kombinaci klíčů. Ta se dá vytvořit ze sloupců cklic_2r1_md, cklic_3r2_md, cklic_4r3_md a cklic_2r1_dal, cklic_3r2_dal, cklic_4r3_dal odstraněním číselných předpon před názvem střediska, skupiny stroje a názvem stroje, a jejich následným sloučením do jednoho

sloupce, který se pojmenuje například „Kombinace klíčů“. Vytvořením tohoto sloupce lze poté tuto tabulku propojit s jinými tabulkami. Ostatní sloupce, tedy sloupce s názvy jednotlivých částí klíče, se mohou odstranit, jelikož tyto údaje jsou již obsaženy v číselníku strojů, který lze připojit přes kombinaci klíčů.

spotřeba nafty

Z důvodu duplikace s tabulkou spotřeba PHM je tato tabulka spotřeba nafty nepotřebná a dále se bude pracovat pouze se spotřeba PHM, která obsahuje navíc sloupce s názvy středisek a skupin zařazení stroje.

spotřeba PHM

V této tabulce je potřeba vytvořit kombinaci klíčů, přes kterou se bude spotřeba PHM propojovat s ostatními tabulkami. Využije se pro to sloupců klic_2r1, klic_3r2 a klic_4r3, které se sloučí v jeden s odřázkami mezi jednotlivými částmi klíče. Dále se odebere sloupec mj, který obsahuje v každém poli písmeno „l“ jako „litry“. Po této úpravě dojde k přejmenování sloupce množství na množství v litrech, aby bylo stále jednoznačné, ve kterých jednotkách je spotřeba zadávána. K další práci nebudou zapotřebí sloupce obsahující část klíče a název střediska, skupiny stroje nebo stroje samotného. Jedná se o sloupce cklic_2r1, cklic_3r2, cklic_4r3, naz_klice_2r1, naz_klice_3r2, naz_klice_4r3. Společně s nimi se odstraní sloupce pohyb, sp1_klice_3 a ns1_klice_3, které jsou nepotřebné. Nakonec se v atributu položky odstraní „Error“ a vytvoří takový filtr, aby se zde zobrazovaly pouze hodnoty „MOTOROVÁ NAFTA“ a „MOTOROVÝ BENZÍN NATURAL 95“.

polozka	nazev	mesic	mnozstvi v litrech	skladova_cena	skl_cena_jedn	kombinace_kl
500002	MOTOROVÁ NAFTA	6	139	3417,22	24,584	100-890-139888
500002	MOTOROVÁ NAFTA	6	793	19495,36	24,584	100-890-139900
500002	MOTOROVÁ NAFTA	6	43	1057,13	24,584	100-890-119882
500002	MOTOROVÁ NAFTA	6	324,63	7980,81	24,584	100-890-119910
500002	MOTOROVÁ NAFTA	6	217	5334,8	24,584	100-890-319894
500002	MOTOROVÁ NAFTA	6	907	22297,97	24,584	100-890-119935
500002	MOTOROVÁ NAFTA	6	129,62	3186,62	24,584	100-890-139899
500002	MOTOROVÁ NAFTA	6	1352,14	33241,44	24,584	100-890-119891
500002	MOTOROVÁ NAFTA	6	1014,98	25363,39	24,989	100-890-139907
500002	MOTOROVÁ NAFTA	6	171,01	4204,16	24,584	100-890-339890
500002	MOTOROVÁ NAFTA	6	304,72	7491,33	24,584	100-890-119889

Obr. 24 spotřeba PHM po transformaci dat

CF_cerven_zari_alimex

V této tabulce se vyskytují dva nulové sloupce, které je možné odstranit. Konkrétně se jedná o Farma řidiče a Farma přídatného zařízení. Dále je odstraněn sloupec Bonus, který není pro vizualizaci a práci s daty potřebný, a Středisko, které obsahuje pouze jednu hodnotu a v analýze se neprojeví. Ve sloupci Výrobní operace se nachází název, který obsahuje předponu VO v každém záznamu. Tato předpona je odebrána rozdělením sloupce na dvě, kdy se v první vyskytuje samotné VO a ve druhé název výrobní operace. Jelikož první část není žádána, je celý tento sloupec odstraněn a vzniká tak název bez předpony.

VO_Alimex

Tabulka VO_Alimex nebyla upravována, jelikož se nebude využívat. Tabulka slouží jako číselník výrobních operací, které jsou zastoupeny v CF_cerven_zari. V současném stavu není potřeba tyto tabulky mezi sebou propojovat, jelikož CF_cerven_zari obsahuje názvy výrobních operací. Pokud by bylo tabulku upravit, odstranila by se předpona VO před názvem výrobní operace podobně jako tomu bylo v tabulce CF_cerven_zari. Bylo by tedy nutné opět sloupec rozdělit na dva a ten, který by obsahoval předpony VO, následně odstranit.

Plodiny_LPIS_alimex

Tato tabulka neobsahuje žádné kombinace klíčů či jiné identifikační záznamy, které by ji mohli propojit s jinou tabulkou. Plodiny_LPIS_alimex navíc neobsahují žádné hodnoty, se kterými by se dalo pracovat, a tak je jako nepotřebná odstraněna.

3.5 Tvorba míry a sloupce

Funkce míra se v Power BI využívá, pokud se v přehledech nenachází vše, co uživatel potřebuje vidět. Lze jejich pomocí provádět běžné operace jako sčítání odečítání, či pokročilé výpočty.

V případě této práce je zapotřebí použít tuto funkci při výpočtu průměrné spotřeby stroje. K tomuto výpočtu se využijí hodnoty sloupce mnozstvi v litrech, který se nachází v tabulce spotřeba PHM, a sloupce Ujetá vzdálenost z tabulky CF_cerven_zari. Podílem hodnot v těchto sloupcích a následným vynásobením číslem sto lze získat průměrnou spotřebu na sto kilometrů. Tato míra je v souboru uložena pod názvem Pr_spotreba. Další míry jsou vytvořeny kvůli zjištění počtu strojů a řidičů. K tomu se využije funkce DISTINCTCOUNT, která vrátí počet jedinečných hodnot ve sloupci.

```
1 Pr_spotreba = (SUM('spotřeba PHM'[mnozstvi v litrech])/SUM(CF_cerven_zari[Ujetá vzdálenost]))*100
```

Obr. 25 Míra – Pr_spotreba

V tabulce spotřeba PHM jsou zadávány měsíce, ve kterých bylo spotřebované určité množství pohonných hmot. Problémem je, že se vyskytují ve formě číslice a pro vizualizaci je vhodnější využít textového formátu. Z toho důvodu je vytvořen nový sloupec NavezMesice, který číslo měsíce převede na text pomocí funkce SWITCH.

```
1 MesicNavez = SWITCH( TRUE(), 'Vývoj cen PHM'[mesic] = 6, "červen", 'Vývoj cen PHM'[mesic] = 7, "červenec", 'Vývoj cen PHM'[mesic] = 8, "srpen", 'Vývoj cen PHM'[mesic] = 9, "září")
```

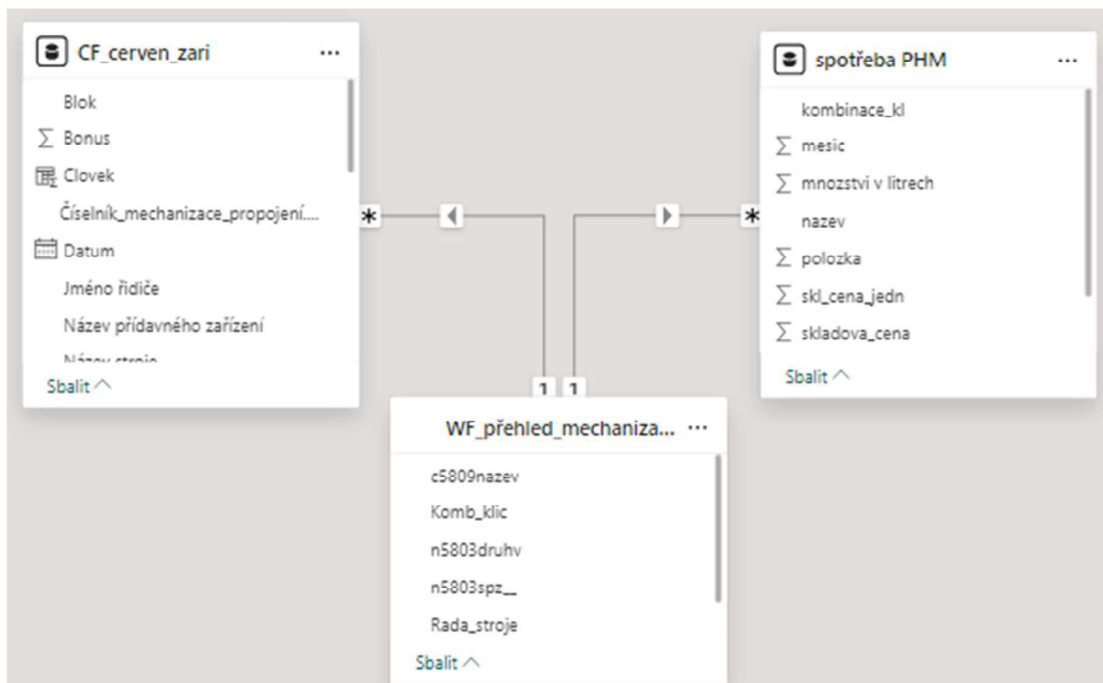
Obr. 26 Sloupec – NavezMesice

3.6 Relace

V současném stavu lze pracovat pouze se třemi tabulkami, kterými jsou WF_přehled_mechanizace, CF_cerven_zari a spotřeba PHM. Vzhledem k této skutečnosti budou existovat dvě relace.

První propojení bude mezi tabulkou CF_cerven_zari a WF_přehled_mechanizace pomocí SPZ strojů. Konkrétně se jedná o sloupec SPZ z CF_cerven_zari a sloupec n5803spz__ z tabulky WF_přehled_mechanizace. Tímto se vytvoří jednoduchá vazba, jejíž kardinalita je N:1.

Druhé propojení bude též kardinality N:1. Tentokrát se však bude jednat o vazbu mezi tabulkami spotřeba PHM a WF_přehled_mechanizace. Tímto propojením se budou přiřazovat názvy strojů k jednotlivým záznamům spotřeby pomocí sloupců s kombinací klíčů. V tabulce WF_přehled_mechanizace se tento sloupec najde pod názvem Komb_klic, zatímco ve spotřeba PHM je veden jako kombinace_kl.

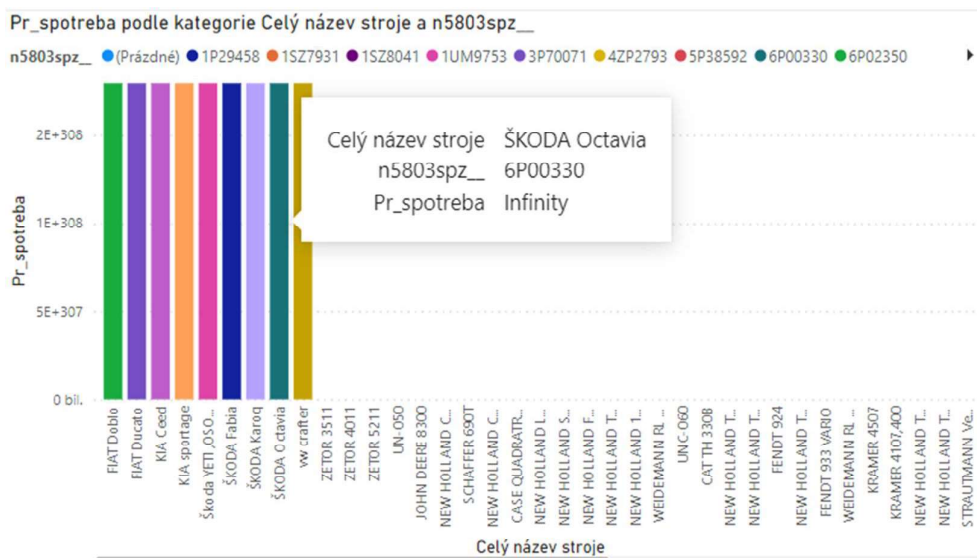


Obr. 27 Relace databáze současného stavu

3.7 Problémy s vizualizací dat

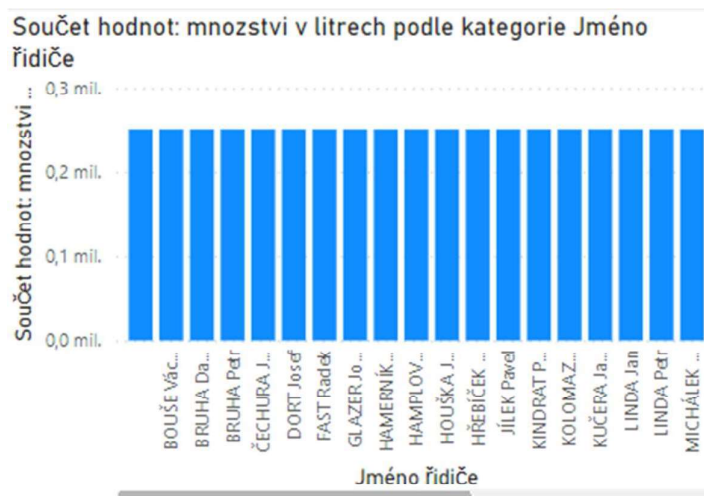
Spousta strojů, které mají uvedenou spotřebu ve sloupci mnozstvi v litrech, nemají jediný záznam v knize jízd, čímž se u nich nevyskytuje žádná ujetá vzdálenost. Tím vzniká problém s následnou vizualizací dat pomocí míry Pr_spotreba, která s ujetou vzdáleností počítá. Ve vizuálech se proto u strojů, které tuto hodnotu nemají, objevuje hodnota nekonečno a zkresluje tak datové přehledy. Příkladem je Škoda Octavia, která má SPZ 6P00030 a kombinaci klíčů 100-861-112737. Celková spotřeba tohoto stroje za sledované období je 578,69 litrů, přičemž v tabulce CF_cerven_zari není součástí jediného záznamu. Dále se zde vyskytují stroje, které mají vysokou spotřebu pohonných hmot, jenže na toto množství mají zaznamenané velmi malé množství kilometrů, čímž výpočet průměrné spotřeby vychází příliš vysoký. Příkladem lze uvést stroj UN-050 s SPZ UN050Čič a kombinací klíčů 207-892-277821, který na ujetých 80

metrů spotřeboval 370 litrů nafty a jeho průměrná spotřeba tak činí 42 833,33 litrů na 100 kilometrů. Na níže přiloženém obrázku lze vidět, jak moc je takový vizuál nevyhovující.



Obr. 28 Problémy s vizualizací – spotřeba na stroj

Podobně nelze zobrazit ani spotřebu pohonných hmot na jméno řidiče. Power BI pouze udělá sumu sloupce množství v litrech a následně tuto sumu přiřadí ke každému jménu, které se v tabulce CF_cerven_zari nachází. Výsledný vizuál opět nemá žádnou vypovídající hodnotu a je nutné nastavit správné propojení mezi tabulkami již v návrhu jednotlivých tabulek. Vizuál je opět k náhledu na obrázku níže.



Obr. 29 Problémy s vizualizací – spotřeba na řidiče

4 Návrh struktury databáze

Současná struktura databáze není vhodná. Vyskytuje se zde spousta chyb, které byly rozebírané v kapitole Problémy s daty a v kapitole Problémy s vizualizací. Tyto nejasnosti je potřeba opravit, nastavit identifikační klíče a zaměřit se na správnou možnost propojení tabulek mezi sebou.

Pro správné nastavení databáze je zapotřebí vytvořit tabulky dimenzí, které poté vstupují do tabulek faktů. Jak již bylo zmíněno, takové tabulky musí obsahovat unikátní identifikační klíče, přes které lze poté tabulky dimenzí a faktů propojit.

4.1 Tabulky dimenzí

Tabulky dimenzí obsahují číselníky. V případě řešené databáze se jedná o tabulky WF_přehled_mechanizace, VO_alimex, Plodiny_LPIS_alimex. Dále je zapotřebí několik dimenzí vytvořit, jedná se o tabulky Přídavné zařízení, Jméno řidiče, Parcela.

WF_přehled_mechanizace

První tabulka dimenzí, WF_přehled_mechanizace, je číselníkem využívaných strojů. V současné podobě se zde vyskytovaly problémy s atomizací dat, tedy se zaznamenáváním více hodnot do jedné buňky. Navrhovaný formát obsahuje sloupce, ve kterých se má vyskytovat pouze jeden údaj. Byly tedy vytvořeny sloupce Znacka_stroje, Rada_stroje, SPZ_stroje. K těmto sloupcům byly zachovány sloupce Skupina_stroje, Druh_stroje, Palivo a GPS jednotka. Dále přibyl sloupec Id_stroje, který by měl být pouze z jedné části, namísto předešlé kombinace klíčů sestávající se ze tří částí klíče.

Tabulka 13 WF_přehled_mechanizace – návrh struktury

WF_přehled_mechanizace	
Název sloupce	Datový typ
Id stroje	string
Id_znacka_stroje	string
Id_rada_stroje	string
SPZ_stroje	string
Id_skupina_stroje	string
Id_druh_stroje	string
Id_misto_zarazeni	string
Palivo	string
GPS_jednotka	string

Znacka_stroje

Vzhledem k problémům se zápisem značky stroje v nynějším stavu bylo zapotřebí vytvořit číselník pro tuto kategorii. Tento číselník obsahuje klíč příslušící k jednotlivým názvům značek. Tato tabulka má za úkol eliminovat záměny názvů výrobce stroje. Pro příklad byl v popisu problémů s daty uváděn zápis značky New Holland, který se vyskytoval také v podobě NEW HOLLAND, NH nebo s překlepem na NEW HOLLNAD.

Tabulka 14 Znacka_stroje – návrh struktury

Znacka_stroje	
Název sloupce	Datový typ
Id_znacka_stroje	string
Nazev	string

Rada_stroje

S řadou stroje byl obdobný problém, co se značkou stroje. V datech byly k nalezení chyby v těchto záznamech, například se jednalo o zápis fabia, který se objevoval také jako fabie. V této tabulce se vyskytují, stejně jako v tabulce Znacka_stroje, pouze dva sloupce, jimiž jsou Id_rada_stroje a Nazev.

Tabulka 15 Rada_stroje – návrh struktury

Rada_stroje	
Název sloupce	Datový typ
Id_rada_stroje	string
Nazev	string

Skupina_stroje

Atribut skupina stroje je v momentálním stavu používán jako jedna část tříčlenné kombinace klíčů a je uváděna společně s názvem stroje a oblastí, do které stroj spadá. Z toho důvodu je skupina uváděna i v navrhované struktuře a je pro ni vytvořena vlastní dimenze pod názvem Skupina stroje. Záznamy v této tabulce obsahují identifikační klíč, ke kterému vždy přísluší konkrétní název.

Tabulka 16 Skupina_stroje – návrh struktury

Skupina_stroje	
Název sloupce	Datový typ
Id_skupina_stroje	string
Nazev	string

Druhstroje

Druh stroje v obdržených datech popisoval, zda se jedná o stroj energetický, samojízdný nebo přípojný. I v tomto případě je pro odstranění rizika chybného zadání vytvořen číselník, který obsahuje již zmíněné tři hodnoty s přiřazeným identifikačním klíčem.

Tabulka 17 Druhstroje – návrh struktury

Druhstroje	
Název sloupce	Datový typ
Id_druhstroje	string
Nazev	string

Mistozarazeni

Tabulka Mistozarazeni vstupuje do WF_prehled_mechanizace. V současné struktuře tabulky WF_prehled_mechanizace se vyskytuje v podobě jedné části kombinace klíče a nemá vlastní dimenzi. Ta vzniká v tomto návrhu, kde obsahuje Id_mistozarazeni, který slouží pro propojení, a k němu příslušný název místa.

Tabulka 18 Mistozarazeni – návrh struktury

Mistozarazeni	
Název sloupce	Datový typ
Id_mistozarazeni	string
Nazev	string

VOalimex

Další tabulkou je VOalimex, která byla oproti původnímu formátu obohacena o jeden sloupec, jímž je Druhvo. Tento sloupec byl vytvořen z důvodu rozdělení názvu a druhu operace, které se vyskytovaly v předešlém formátu v jedné buňce. Celá tabulka tedy obsahuje sloupce Idvo, Druhvo, Nazevvo a mj.

Tabulka 19 VOalimex – návrh struktury

VOalimex	
Název sloupce	Datový typ
Idvo	string
Druhvo	string
Nazevvo	string
mj	string

Plodiny_LPIS_alimex

Tento soubor, který společnost zaslala, se v původním formátu nedal vůbec využít. V této tabulce chyběli identifikační klíče a nebylo možné ji připojit k dalším tabulkám. V souboru tak musel být zaveden sloupec obsahující identifikační klíč, ke kterému je vždy přiřazen název plodiny

Tabulka 20 Plodiny_LPIS_alimex – návrh struktury

Plodiny_LPIS_alimex	
Název sloupce	Datový typ
Id_plodina	string
Nazev	string

Pridavne_zarizeni

Stejně jako přehled strojů, tedy tabulka WF_přehled_mechanizace, je potřeba zavést číselník přídatných zařízení, aby nedocházelo k zaznamenávání více údajů do jedné kolonky. Tato tabulka dimenzí obsahuje Id_prid_zarizeni, Znacka_prid_zarizeni, Rada_prid_zarizeni a SPZ_prid_zarizeni.

Tabulka 21 Pridavne_zarizeni – návrh struktury

Pridavne_zarizeni	
Název sloupce	Datový typ
Id_prid_zarizeni	string
Znacka_prid_zarizeni	string
Rada_prid_zarizeni	string
SPZ_prid_zarizeni	string

Jmeno_ridice

V obdržení souborech se nevyskytovala žádná evidence řidičů, proto byla nově zavedena tabulka Jmeno_ridice. V této tabulce se nachází Id_jmeno_ridice, pomocí kterého je možné tabulku propojit s tabulkou CF_cerven_zari, která obsahuje záznamy a údaje o jednotlivých jízdách. Dále se zde nacházejí sloupce Prijmeni, Jmeno, Vek, Rok nastupu. Sloupce Vek a Rok nastupu jsou zavedeny z důvodu následných vizualizací dat, ve kterých díky těmto záznamům je možné sledovat efektivitu jízd řidičů na základě věku a odpracovaných let.

Tabulka 22 Jmeno_ridice – návrh struktury

Jmeno_ridice	
Id_jmeno_ridice	string
Prijmeni	string
Jmeno	string
Vek	integer
Rok nastupu	integer

Parcela

Podobně jako jméno řidiče, tak se pouze v evidenci jízd objevily názvy parcel. Je tedy vytvořena další dimenzní tabulka, která obsahuje údaje o obsluhované parcele. Kvůli propojení s evidencí jízd je nutné zavést sloupec Id_parcela. V tomto identifikátoru jsou obsaženy informace jako název parcely, její blok a rozloha, které byly předtím také vepsány v jízdách.

Tabulka 23 Parcela – návrh struktury

Parcela	
Název sloupce	Datový typ
Id_parcela	string
Nazev	string
Blok	string
Rozloha	float

4.2 Faktové tabulky

Na rozdíl od tabulek dimenzí obsahují faktové tabulky kvantitativní data (měřitelné údaje). V případě této práce se jedná o tabulky spotřeba PHM a CF_cerven_zari. První tabulka obsahuje záznamy o spotřebovaných litrech pohonných hmot v určitém měsíci a určitým strojem. Také je zde uváděná celková a jednotková cena. V druhé tabulce, CF_cerven_zari, lze nalézt ujeté kilometry konkrétním řidičem. Tímto jsou splněny podmínky o měřitelných datech a lze považovat tyto tabulky za faktové.

V současném stavu však není možné tyto dvě tabulky mezi sebou propojit, jelikož nejsou na sebe nijak navázané a není možné určit řidiče a jeho spotřebu.

CF_cerven_zari

Jak bylo zmíněno výše, tato tabulka obsahuje záznam o ujeté vzdálenosti konkrétním řidičem na konkrétním stroji při konkrétní výrobní operaci a v konkrétní den s uvedeným časovým začátkem jízdy. Obsahuje navíc sloupce Farma řidiče a Farma přídatného zařízení, které jsou bez jediného záznamu, a tak se s nimi v návrhu nebude počítat. Odstraněn byl také sloupec Bonus, o kterém společnost na společném sezení rozhodla, že není zapotřebí. Dále není třeba zobrazovat sloupce typ operace, blok, rozloha parcely SPZ, jelikož tyto údaje jsou nově uvedeny v tabulkách příslušných dimenzí. Důležité je zavést sloupec Id_jizdy, který se v současné verzi nevyskytuje a je potřeba k propojení s tabulkou spotřeba PHM. Struktura navrhované tabulky je uvedena v následující tabulce.

Tabulka 24 CF_cerven_zari – návrh struktury

CF_cerven_zari	
Název sloupec	Datový typ
Id_jizdy	string
Datum	date
Začátek jízdy	time
Trvání	time
Ujetá vzdálenost	float
Id_jmeno_ridice	string
Id_vo	string
Id_stroje	string
Id_prid_zarizeni	string
Id_parcela	string
Id_plodina	string

spotřeba PHM

Druhou faktovou tabulkou je spotřeba PHM. Zde se nachází informace o spotřebovaném množství pohonných hmot v litrech. Tato spotřeba se momentálně váže pouze na stroj a nelze nijak propojit s databází o jízdách. Z toho důvodu byl vytvořen v návrhu tabulky CF_cerven_zari identifikační klíč pod názvem Id_jizdy a tento identifikační klíč se vyskytuje v databázi spotřeba PHM jako tzv. cizí klíč. Tímto výskytem lze mezi sebou tyto dvě tabulky propojit a je možné přiřadit spotřebované množství pohonných hmot ke konkrétnímu člověku.

Tabulka 25 spotřeba PHM – návrh struktury

spotřeba PHM	
Název sloupce	Datový typ
ID	string
Datum	date
Id_jizdy	string
Spotrebovane_mnozstvi	float
Skladova_cena	float
Skladova_cena_jednotka	float

4.3 Naplnění navržené struktury daty

Do navržené struktury jednotlivých tabulek byla vložena experimentální data tak, aby se s těmito zdroji mohlo dále pokračovat v diplomové práci a využít je v Microsoft Power BI. Z původní databáze byly pro naplnění tabulek dimenzí využity názvy strojů, operací, plodin, parcel, míst a přídatných zařízení. Naplněna byla také tabulka Jméno řidiče, ve které byla všechna jména smyšlená a neodkazují na žádné skutečné osoby. Další experimentální data se vyskytují ve faktových tabulkách, ve kterých byly vytvořeny záznamy jízd, jejich daty, ujeté vzdálenosti, doby trvání, spotřebované litry atd. K těmto faktům byly poté náhodně přiřazeny údaje z tabulek dimenzí. Příklad naplněné databáze lze shlédnout na následujícím obrázku.

Id_jizdy	Datum	Zacatek_jizdy	Trvani	Ujeta_vzdalenost	Id_jmeno_ridice	Id_vo	Id_stroje	Id_prid_zarizeni	Id_parcela	Id_plodina
1	01.06.2023	12:56:26	02:13:55	20,05	6	4	7	3	14	8
2	01.06.2023	12:18:36	01:52:12	5,50	12	7	2	9	12	13
3	01.06.2023	13:50:59	03:25:53	9,37	15	2	11	4	13	6
4	01.06.2023	13:56:25	03:32:15	17,14	12	5	8	3	15	11
5	02.06.2023	11:21:44	03:38:18	1,65	12	4	14	5	4	9
6	02.06.2023	14:30:53	01:42:15	18,96	14	14	7	9	8	13
7	02.06.2023	15:05:32	01:42:49	3,79	10	7	3	4	11	7
8	02.06.2023	15:19:47	04:11:51	27,11	14	9	12	4	12	5
9	02.06.2023	14:23:00	02:04:38	29,31	12	11	9	4	13	11
10	02.06.2023	07:10:58	02:34:15	22,48	9	4	3	6	13	5

Obr. 30 Naplnění CF_cerven_zari – návrh struktury

4.4 Transformace dat

Vzhledem k navržené struktuře není potřeba rozsáhlé transformace dat a jedná se pouze o malé úpravy. Tyto úpravy se týkají především typu dat, který po načtení jednotlivých tabulek může být nepřesný a je potřeba ho upravit do požadované podoby. Dále byly zaokrouhleny všechna čísla na jednotný počet tří desetinných míst a vytvořeny dotazy a míry důležité pro vizualizaci dat. Poslední úpravou byla duplikace a sloučení sloupců obsahující příjmení a jména řidiče, aby se ve vizuálech zobrazovalo celé jméno.

Změna typu dat

Pro demonstraci funkčnosti navržené struktury byla tato databáze vytvořena v programu Microsoft Excel a naplněna experimentálními daty. Po načtení tohoto souboru však dochází ke změně typů dat v jednotlivých sloupcích, což je způsobeno snahou Power BI jednotlivé formáty dat rozpoznat a přiřadit. Všechny sloupce v jednotlivých tabulkách tak byly nastaveny do formátu dat uvedených v návrhu struktury databáze.

Tvorba dotazu

V navržené struktuře se nevyskytuje kompletní název stroje. V tabulce WF_přehled_mechanizace se vyskytují pouze identifikační klíče jednotlivých částí názvu a z toho důvodu je nutné připojit dotaz z propojených tabulek a tím vytvořit název stroje, který se využije ve vizuálech. Power BI pro tuto operaci obsahuje funkci sloučit dotazy. V tomto případě se tato funkce využije na tabulky WF_přehled_mechanizace, Znacka_stroje a Rada_stroje, kdy ke sloupcům v tabulce WF_přehled_mechanizace obsahujícím identifikační klíče ze zmíněných tabulek, budou připojeny hodnoty odpovídající těmto klíčům. Posledním krokem je sloučení sloupců, které vznikly tímto postupem. Výsledný sloupec tak bude obsahovat jak název značky stroje z tabulky Znacka_stroje, tak i jeho řadu z tabulky Rada_stroje v tomto pořadí a oddělené čárkou, např.: FENDT, 933 Vario.

Tvorba měr

Součástí Power BI je možnost vytvářet míry a následně je aplikovat při vytváření vizuálů. Pro tuto práci byly vytvořeny celkem čtyři míry, kterými se počítá průměrná spotřeba, průměrná rychlost, počet jízd a náklady na vzdálenost.

Průměrná spotřeba se spočítá obdobně jako tomu bylo při výpočtu v původní verzi. Využijí se záznamy o spotřebovaném množství pohonných hmot z tabulky spotřeba PHM a zaznamenané ujeté vzdálenosti z tabulky CF_cerven_zari. Následně se tyto hodnoty mezi sebou podělí a výsledek vynásobí stem, čímž se získá průměrná spotřeba na sto kilometrů.

```
1 Průměrná spotřeba = (SUM('spotřeba PHM'[Spotrebovane_mnozstvi])/
2 | SUM('CF_cerven_zari'[Ujeta_vzdalenost])
3 )*100
```

Obr. 31 Návrh struktury – Míra – Průměrná spotřeba

Pro tvorbu výsledného dashboardu bylo využito také výpočtu průměrné rychlosti jízdy. Tento výpočet se počítá opět z ujeté vzdálenosti zaznamenané v tabulce CF_cerven_zari a z času trvání jízdy. Obě hodnoty se mezi sebou podělí, čímž se obdrží požadovaný výstup.

```
Průměrná rychlost = SUM(CF_cerven_zari[Ujeta_vzdalenost])/SUM(CF_cerven_zari[Čas číslo])
```

Obr. 32 Návrh struktury – Míra – průměrná rychlost

Třetí míra, která byla vytvořena pro zobrazení nákladů na ujetou vzdálenost, je výsledkem poměru skladové ceny a ujeté vzdálenosti. Ujetá vzdálenost se opět nachází v tabulce CF_cerven_zari, kde je uvedena u každé jízdy, zatímco skladová cena je zaznamenána v návaznosti na každou jízdu v tabulce spotřeba PHM.

```
Náklady na vzdálenost = SUM('spotřeba PHM'[Skladova_cena])/SUM(CF_cerven_zari[Ujeta_vzdalenost])
```

Obr. 33 Návrh struktury – Míra – Náklady na vzdálenost

Poslední míra počítá počet zaznamenaných jízd pomocí příkazu COUNT, který využívá data z tabulky CF_cerven_zari ve sloupci Id_jizdy.

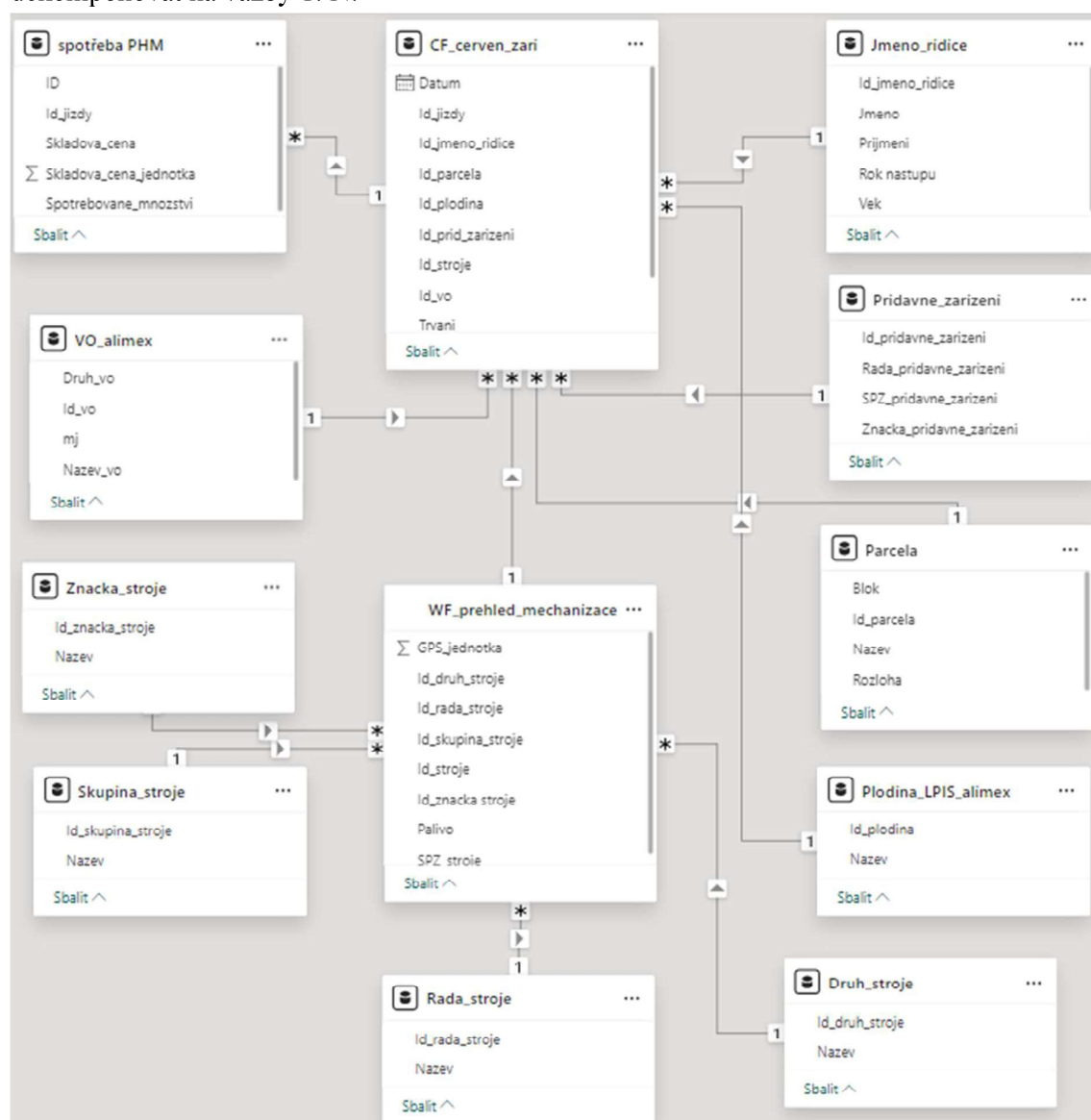
```
Počet jízd = COUNT(CF_cerven_zari[Id_jizdy])
```

Obr. 34 Návrh struktury – Míra – Počet jízd

4.5 Relace

Datový model je při tomto návrhu, oproti původnímu stavu, na první pohled složitější. Vyskytuje se zde větší množství tabulek dimenzí, která předtím byla pouze jedna, přičemž nyní jich je deset. Počet faktových tabulek zůstává zachován v počtu dvou.

Tento datový model lze pomyslně rozdělit na tři úrovně. V první úrovni se vyskytují tabulky dimenzí vstupující do WF_přehled_mechanizace, jenž je taky dimenzí. Jedná se o tabulky Znacka_stroje, Skupina_stroje, Rada_stroje a Druh_stroje, které jsou s koncovou tabulkou propojeny ve vazbě 1: N pomocí svých identifikačních klíčů. Druhou úrovní jsou tabulky, které zásobují první tabulku faktů, tedy CF_cerven_zari. Tato tabulka je již propojena s WF_přehled_mechanizace a dále s VO_alimex, Parcela, Pridavne_zarizeni a Jmeno_ridice opět za pomoci svých identifikačních klíčů. Do třetíce se propojí CF_cerven_zari s tabulkou spotřeba PHM. Výhodou tohoto datového modelu je, že díky vytvořeným číselníkům neobsahuje žádné vazby M: N, které mohou být problematické a museli by se následně dekomponovat na vazby 1: N.



Obr. 35 Datový model návrhu struktury databáze

4.6 Vizualizace dat

Pro potvrzení funkčnosti navržené struktury databáze, která byla zároveň naplněna experimentálními daty, byly vytvořeny celkem čtyři dashboardy, kterými jsou přehled, stroj, člověk a operace. Tyto dashboardy se zaměřují každý na jinou oblast odpovídající jeho názvu. Oproti původní struktuře databáze bylo možné data nyní bez problémů vizualizovat a jak bylo již zmíněno v kapitole o transformaci dat, nebylo k tomu zapotřebí rozsáhlých úprav a transformací dat po načtení tabulek do nástroje Microsoft Power BI.

4.6.1 Přehled

Na první straně vizuálů, která je pojmenována jako Přehled, se nachází informace o spotřebě pohonných hmot a finančních nákladech s nimi spojených.

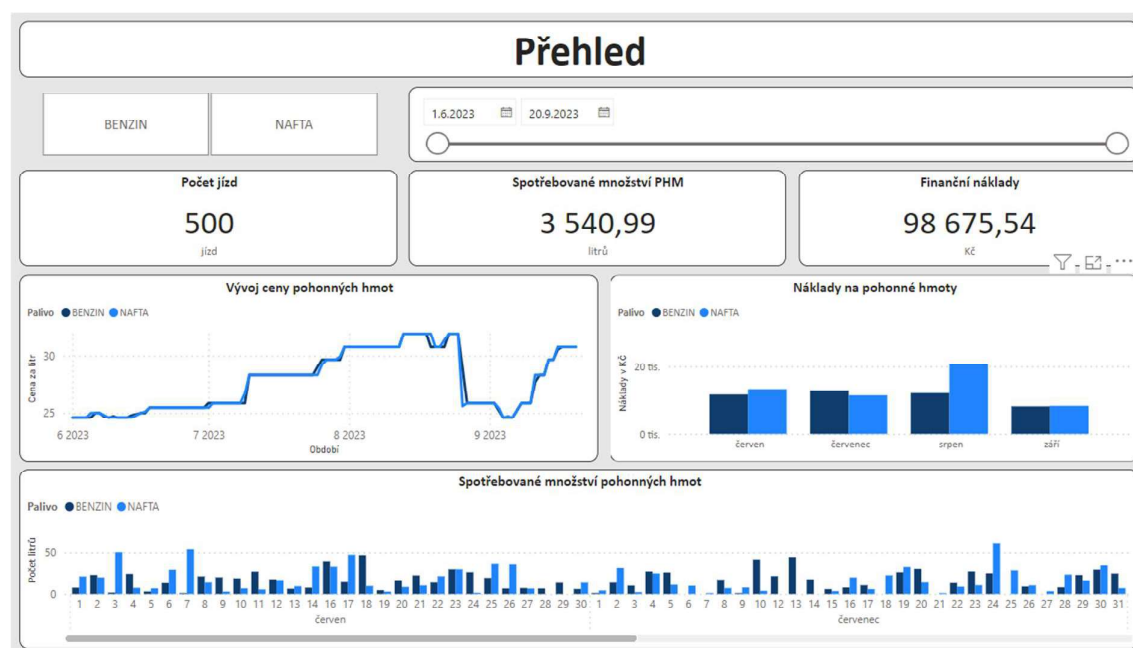
V první řadě se zde nachází vizuály, které slouží pro filtrování obsahu této strany. V levém rohu se pomocí kliku vybere zda, uživatel chce vidět záznamy o benzínu či naftě. Pokud bude chtít vidět oboje zároveň, nevybere nic. V případě, že je však jedna z možností již zvolená, deaktivuje ji opětovným klikem na tuto možnost. Poté se ve vizuálech promítnou opět obě možnosti. Po pravé straně těchto dlaždic určených pro filtrování, se nachází posuvník, kterým lze ovlivňovat období, které bude zobrazeno. Lze tak provést zúžení, či roztažení posuvníku, ale také vybráním konkrétního datumu kliknutím na ikonu kalendáře a zvolením požadovaného data. Poslední možností je ruční přepsání datumu.

Druhá řada vizuálů jsou takzvané karty. V těchto kartách se nachází údaje o počtu jízd, průměrné spotřebě pohonných hmot, celkové spotřebě pohonných hmot a finanční nákladech na pohonné hmoty. První karta tedy zobrazuje součet absolvovaných jízd, který je počítán pomocí míry Počet jízd. Na další kartě následuje využití míry Průměrná spotřeba, která ukazuje průměrné spotřebované množství litrů pohonných hmot na sto kilometrů. Třetí karta sčítá všechno spotřebované palivo za sledované období a poslední, čtvrtá karta, uvádí finanční náklady na tuto spotřebu.

V další řadě se již vyskytují dva grafy. První graf je spojnicový a týká se vývoje ceny jak benzínu, tak nafty. Ve vizuálu jsou tedy zastoupeny dvě křivky, kdy každá znázorňuje jiný druh pohonné hmoty. V ose y nabývá graf hodnot ceny za litr, zatímco v ose x je zobrazené období. Druhý graf je již skupinový sloupcový a je na něm vidět náklady za palivo v průběhu sezóny. V ose x jsou zobrazeny čtyři měsíce ze sledovaného období, kterými je červen, červenec, srpen a září, a v ose y náklady na pohonné hmoty udávané v Kč. Podobně jako v předešlém grafu, tak se zde vyskytují hodnoty jak za benzín, tak za naftu. Tentokrát je to však ve formě složených dvou sloupců, které jsou barevně odlišeny. Legenda v levém horním rohu tuto barevnou odlišnost popisuje.

Spodní oblast dashboardu je využita pro skupinový sloupcový graf, který svojí rozlohou zabírá celou řadu. Jeho velikost je dána především obsahem dat na ose x, která obsahuje dny za měsíc červen až září. Do osy y jsou poté vyneseny údaje o spotřebovaném množství pohonných hmot v jednotlivých dnech. Jak bylo již zmíněno, graf v ose x je poměrně rozsáhlý a všechna data nejsou tak vyobrazena v jednom náhledu. Vzhledem k této skutečnosti je ve spod grafu umístěn posuvník, kterým lze volně přejíždět do stran a prohlížet tím všechny zaznamenané údaje od června až do září.

Příložený obrázek tohoto dashboardu slouží pro ukázkou rozložení jednotlivých vizuálů a použitých grafů. Čitelná verze je poté k dispozici viz. PŘÍLOHA č. 1.



Obr. 36 Dashboard – Přehled

4.6.2 Stroj

Druhá strana vizuálů je zaměřená na stroj. Dashboard je rozdělený do čtyř částí, ve kterých jsou data zobrazena za pomoci karet, grafů a tabulek.

V první zobrazované oblasti se nachází filtr, který slouží k výběru sledovaného období. Za tímto filtrem je poté umístěno rozevírací okno pro výběr stroje. Po výběru stroje se v kartě umístěné za rozevíracím oknem uživateli zobrazí druh paliva, které daný stroj spotřebovává.

Pod výběrovými okny se nachází celkem šest karet se základními informacemi, které zobrazují souhrny o provedených jízdách vybraným strojem. První karta informuje o celkovém počtu ujetých jízd, přičemž na následujících dvou kartách je zaznamenáno, jak dlouho tyto jízdy celkově trvaly a jaká vzdálenost byla v jejich rámci ujeta. Poslední tři karty podávají informace o průměrné spotřebě pohonných hmot při těchto jízdách, jejich celkovou spotřebu udávanou v litrech a náklady na pohonné hmoty.

Další řada obsahuje celkem tři vizuály. Prvním je sloupcový graf, který zobrazuje průměrnou rychlost jízd stroje v jednotlivých dnech užívání. Následuje prstencový graf s údaji o průměrné spotřebě vázaných k operacím, v jejichž rámci byly jízdy vykonávány. Posledním vizuálem zde byla zvolena tabulka, která jízdy přiřazuje k jednotlivým řidičům a zároveň k nim uvádí informace o ujeté vzdálenosti a průměrné spotřebě. Jak tato tabulka, tak prstencový graf zároveň slouží k dalšímu filtrování v tomto dashboardu. Pokud uživatel klikne v tabulce na jméno řidiče, ostatní vizuály se přizpůsobí tomuto výběru a zobrazí se v nich informace závislé na vybraném stroji a řidiči. Obdobně lze pracovat s prstencovým grafem, kde je možné zvolit jednu ze tří operací kliknutím na příslušnou část grafu.

Ve spodní oblasti se podobně, jako tomu bylo v přehledu vyskytuje sloupcový graf, který zobrazuje průměrnou spotřebu v jednotlivých dnech jízd. Zde je graf obohacen o čáru konstanty, která představuje celkovou průměrnou spotřebu, a uživatel tak může sledovat, ve kterých dnech byla překročena hodnota průměrné spotřeby.

Na přiloženém obrázku, Obr. 37 Dashboard – Stroj, lze shlédnout rozmístění popisovaných vizuálů. PŘÍLOHA č. 2 je čitelnější verzi tohoto dashboardu.



Obr. 37 Dashboard – Stroj

4.6.3 Člověk

Další dashboard se zaměřuje na analýzu výrobních dat spojených s člověkem, tedy řidičem. Jedná se o podobnou šablonu, která byla vytvořena při analýze dat s ohledem na stroj, ale vyskytuje se zde rozdíl v podobě skládaného pruhového grafu, který je v levé části umístěn přes třetí a čtvrtou řadu vizuálů.

První řada je opět věnována filtrování a obsahuje posuvník pro upravování sledovaného období, který se ovládá obdobně jako tomu bylo na straně Přehled. Před ním se nachází rozevírací okno, ve kterém se vybírá jméno řidiče, kterého potřebuje uživatel zobrazit.

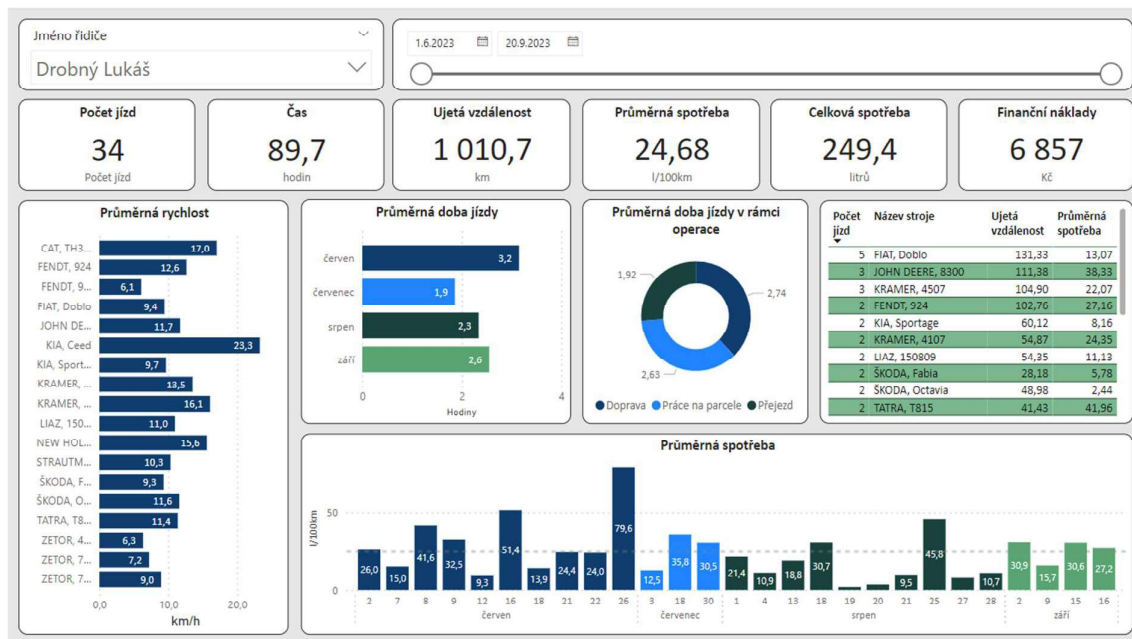
V další řadě se vyskytují karty, které obsahují základní informace o jízdách vykonaných zvoleným řidičem. Uživatel zde uvidí přehled, který se skládá z počtu jízd řidiče, času stráveného jízdou a celkové ujeté vzdálenosti. Následují karty zobrazující údaje o průměrné spotřebě jízd řidiče, jeho celkové spotřebě a finančních nákladech na spotřebu pohonných hmot.

Pod kartami s údaji o jízdách řidiče jsou umístěny grafy. Po levé straně jsou pomocí skládaného pruhového grafu zobrazena data s průměrnými rychlostmi řidiče na jednotlivých obsluhovaných strojích. Vedle něj se nachází v řadě dva grafy a jedna tabulka. První graf, skládaný pruhovaný, zobrazuje výkonnost řidiče ve sledovaných měsících pomocí průměrné doby jízdy. Za ním následuje prstencový graf, který informuje o výkonnosti řidiče v rámci operace obdobně jako v předchozím vizuálu, pomocí průměrné doby jízdy. Posledním vizuálem v této řadě je tabulka, která uchovává informace o počtu provedených jízd řidičem na konkrétním stroji, celkové ujeté vzdálenosti v rámci těchto jízd a průměrnou spotřebu pohonných hmot. V případě potřeby je v této tabulce možné vyfiltrovat konkrétní stroj a přizpůsobit tak tento dashboard sledování výkonů řidiče na tomto stroji. Pro tuto činnost lze

využit i prvně zmiňovaný skládaný pruhový graf, který též obsahuje názvy strojů. Prstencovým grafem je poté možné filtrovat druh výrobní operace.

Pod zmiňovanými grafy se vyskytuje poslední vizuál této strany. Jeho obsahem jsou data o průměrné spotřebě pohonných hmot v jednotlivých dnech sledovaného období. Pro vizualizaci těchto dat bylo využito míry Průměrná spotřeba a datumu jízdy z tabulky CF_cerven_zari.

Na obrázku umístěném níže je vyobrazeno rozložení zmíněných vizuálů. Pro čitelný náhled tohoto dashboardu je zapotřebí přejít na PŘÍLOHA č. 3.



Obr. 38 Dashboard – Člověk

4.6.4 Operace

Poslední strana vizuálů se zabývá provedenými jízdami v rámci operace. Šablona této strany je téměř totožná se stranou Stroj. Rozdíl je v uspořádání čtvrté řady, kde se nyní vyskytují dva skládané sloupcové grafy, a v obsahu vizuálů.

V horní oblasti dashboardu jsou umístěny filtry. V prvním filtru uživatel zvolí, který druh operace chce zobrazit, zda se jedná o práci na parcele, přejezd nebo dopravu. Následně za ním pomocí posuvníku zvolí sledované období. Jak bylo zmíněno dříve, sledované období je možné upravit také po kliknutí na ikonu kalendáře a následně vybráním data, ale zároveň ručním přepsáním datumu.

Stejně jako tomu bylo v předešlých dashboardech, tak i zde druhou řadu vizuálů pokrývají karty se souhrnnými informacemi. Tyto karty opět poskytují uživateli pohled na počet zaznamenaných jízd a jejich dobu trvání v rámci operace s doplněním o celkové ujeté vzdálenosti, průměrné a celkové spotřebě pohonných hmot. Podobně jako v předchozích případech je tato řada uzavřena kartou s finančními náklady na spotřebované pohonné hmoty.

Dashboard Operace je k dispozici jako PŘÍLOHA č. 4. Přiložený obrázek slouží pouze jako náhled.

Závěr

Diplomová práce se zabývá analýzou a vyhodnocením výrobních dat. Cílem této práce je analyzovat obdržená data od společnosti ALIMEX NEZVĚSTICE a.s. a následně navrhnout takovou strukturu databáze, aby bylo možné v nástroji Microsoft Power BI smysluplně vizualizovat data uložená v této databázi za účelem zjištění, kde dochází k největší spotřebě pohonných hmot. Data obsahovala informace o uskutečněných jízdách na jednotlivých strojích, včetně dalších dodatečných informací.

Teoretická část obsahuje dvě kapitoly pod názvem Data a Power BI. V první kapitole se nachází informace o práci s daty, jaké jsou jejich druhy, typy jejich sběru, ukládání, zpracování a možné problémy s nimi spojené. Zabývá se také normalizací dat společně s datovou a funkční analýzou. V další kapitole je přiblížen nástroj Microsoft Power BI. Jsou zde popsány součásti, ze kterých se tento nástroj skládá, jeho pracovní prostředí a následně základní operace, které v něm lze provést.

Společnost ALIMEX NEZVĚSTICE a.s. zaslala několik souborů s daty, které potřebovala vizualizovat, čímž se zabývá praktická část této práce. Vzhledem k rozsáhlým problémům s těmito daty nebylo možné tato data mezi jednotlivými tabulkami propojit a následně smysluplně vizualizovat. Identifikované problémy při analýze současného stavu byly náležitě popsány a současně bylo pomocí několika vizuálů demonstrováno, proč s touto strukturou databáze nelze dále pracovat. Na základě tohoto nedostatku byla navržena nová struktura databáze, která odstraňuje předešlé nedostatky, umožňuje jednotlivé tabulky mezi sebou jednoduše propojit a tím bez problémů vytvářet datové vizuály v nástroji Microsoft Power BI.

Kvůli zachování anonymity jednotlivých pracovníků, byla změněna jména pracovníků v databázi.

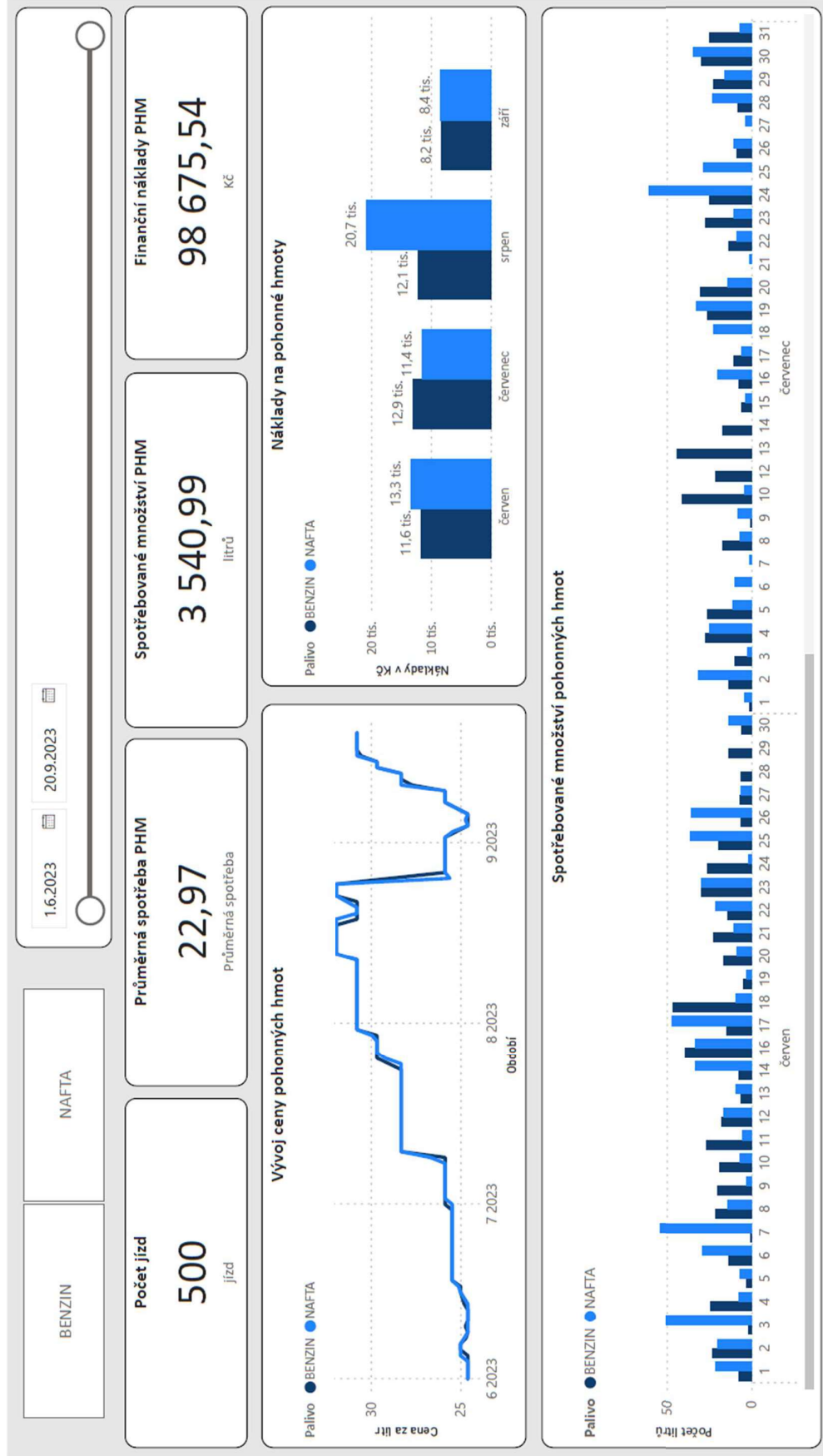
Seznam použitých zdrojů

- [1] C. doc. Ing. Pavel Kopeček. [Online]. Available: <http://home.zcu.cz/~kopecek/DBCsyl/Prednasky/Přednášky%20PPVS%20nové%20-%20v3.pdf>.
- [2] [Online]. Available: <https://portal.matematickabiologie.cz>.
- [3] H. CUESTA, Analýza dat v praxi, Computer Press, 2015.
- [4] „OptoNet,“ [Online]. Available: <https://optonet.cz/datove-centrum-vysocina/cloudove-uloziste.html>.
- [5] „AWS,“ [Online]. Available: <https://aws.amazon.com/relational-database/>.
- [6] „Microsoft Azure,“ [Online]. Available: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-a-relational-database>.
- [7] „SAP,“ [Online]. Available: <https://www.sap.com/cz/insights/in-memory-database.html>.
- [8] „Oracle,“ [Online]. Available: <https://www.oracle.com/cz/database/nosql/what-is-nosql/>.
- [9] „Oracle,“ [Online]. Available: <https://www.oracle.com/cz/database/what-is-a-data-warehouse/>.
- [10] „Microsoft Azure,“ [Online]. Available: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-a-data-lake/>.
- [11] „TechTarget,“ [Online]. Available: <https://www.techtarget.com/searchstorage/definition/redundant>.
- [12] O. P. RUD, Data Mining, Computer Press, 2001.
- [13] [Online]. Available: https://ki.ujep.cz/opory/Aplikovana_Informatika/Bc/Uvod_do_relacnich_databazi.pdf.
- [14] [Online]. Available: https://projekty.fs.vsb.cz/463/edubase/VY_01_044/Databázové%20systémy/02%20Text%20pro%20e-learning/Databázové%20systémy%2005.%20Normalizace%20dat.pdf.
- [15] [Online]. Available: [https://projekty.fs.vsb.cz/463/edubase/VY_01_044/Databázové%20systémy/02%20Text%20pro%20e-learning/Databázové%20systémy%2003.%20Databázové%20\(datové\)%20modely%20a%20modelování.pdf](https://projekty.fs.vsb.cz/463/edubase/VY_01_044/Databázové%20systémy/02%20Text%20pro%20e-learning/Databázové%20systémy%2003.%20Databázové%20(datové)%20modely%20a%20modelování.pdf).
- [16] „Microsoft Learn,“ [Online]. Available: <https://learn.microsoft.com/cs-cz/power-bi/fundamentals/power-bi-overview>.
- [17] „Microsoft Support,“ [Online]. Available: <https://support.microsoft.com/cs-cz/office/power-pivot-přehled-a-výuka-f9001958-7901-4caa-ad80-028a6d2432ed>.

- [18] „Microsoft Learn,“ [Online]. Available: <https://learn.microsoft.com/cs-cz/power-bi/transform-model/desktop-query-overview#power-query-editor>.
- [19] D. CLARK, Beginning Microsoft Power BI: A Practical Guide To Self-Service Data Analytics., APress, 2020.
- [20] „Microsoft Learn,“ [Online]. Available: <https://learn.microsoft.com/cs-cz/power-bi/transform-model/desktop-create-and-manage-relationships>.
- [21] „ALIMEX a.s.,“ [Online]. Available: <https://www.alimex-as.cz>.
- [22] P. K. ESCOBAR Miguel, Master Your Data with Excel and Power BI, Holy Macro! Books, 2021.
- [23] P. SEAMARK, Beginning DAX with Power BI, APress, 2018.
- [24] M. R. Alberto FERRARI, Analyzing Data with Power BI and Power Pivot for Excel, Microsoft Press, 2017.
- [25] M. ALEXANDER, Excel Power Pivot & Power Query For Dummies, John Wiley & Sons Inc, 2022.
- [26] „simplilearn,“ [Online]. Available: <https://www.simplilearn.com/what-is-data-collection-article>.
- [27] „Microsoft Azure,“ [Online]. Available: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-a-relational-database>.

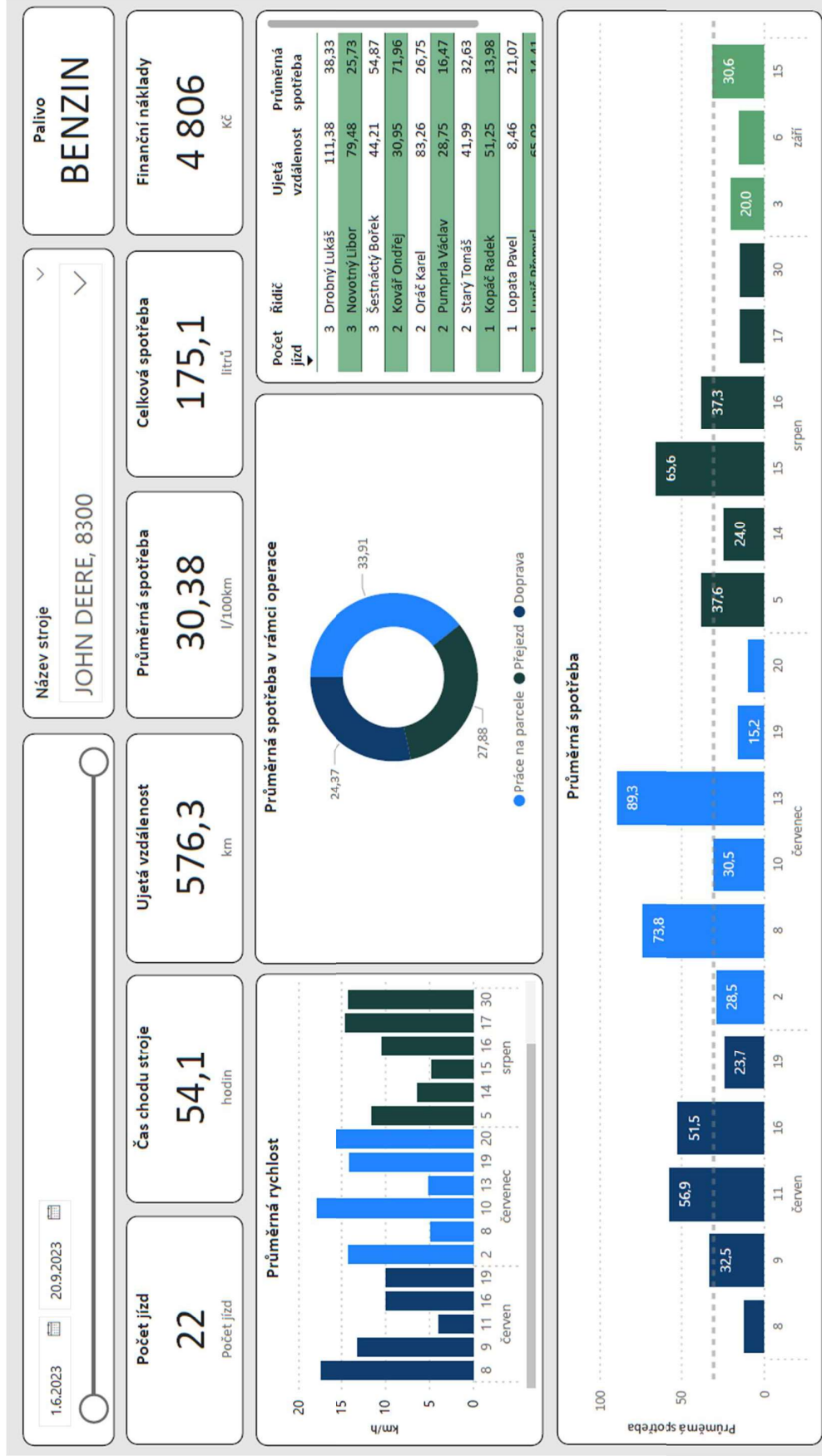
PŘÍLOHA č. 1

Dashboard – Přehled



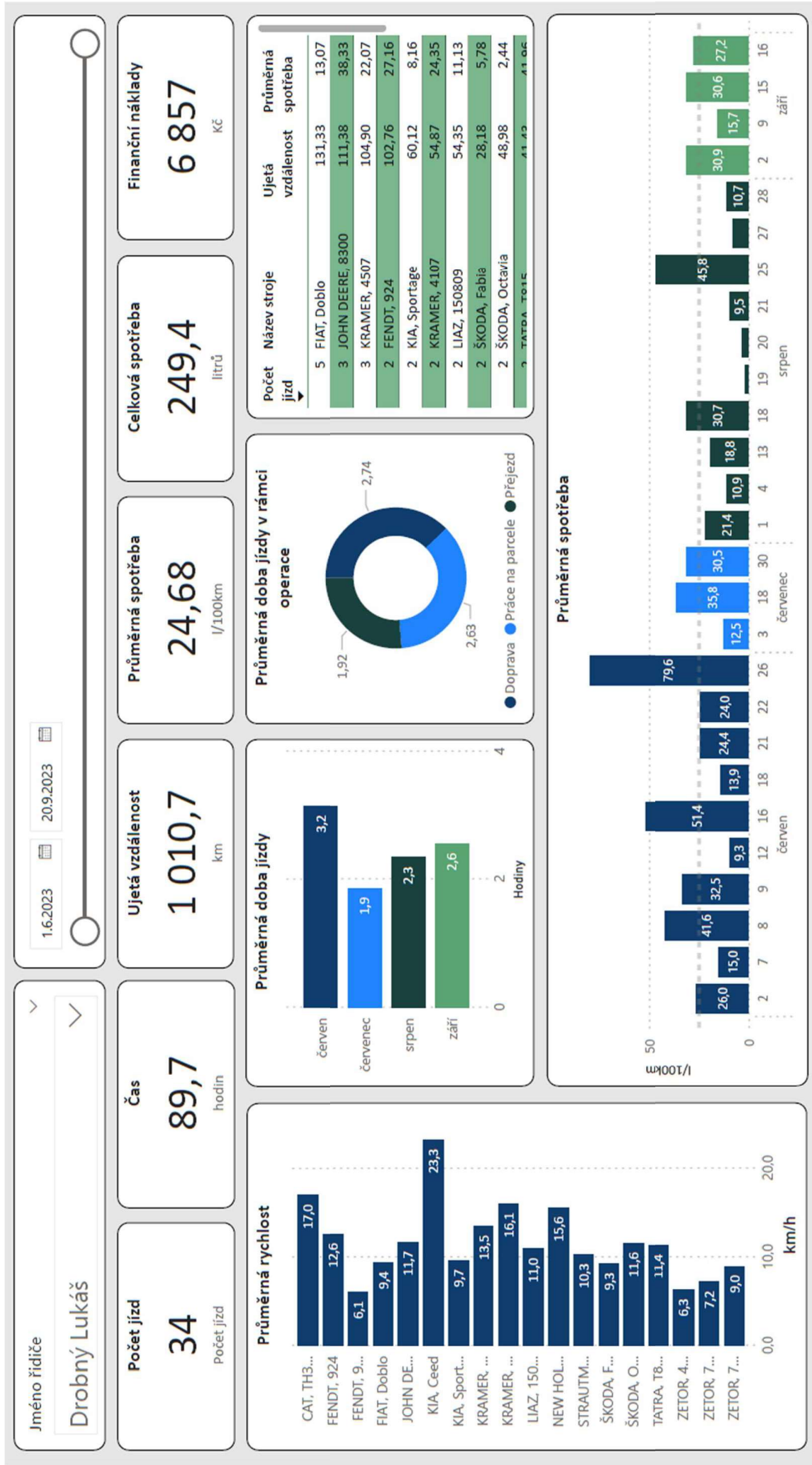
PŘÍLOHA č. 2

Dashboard – Stroj



PŘÍLOHA č. 3

Dashboard – Člověk



PŘÍLOHA č. 4

Dashboard – Operace

