

**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

**MASARYKŮV ÚSTAV VYŠŠÍCH STUDIÍ**



**BAKALÁŘSKÁ PRÁCE**

**Business Intelligence: Zpracování podnikových dat**

**Business Intelligence: Business data processing**

**2023**

**Ernest Török**

**Studijní program:** Podniková ekonomika a management

**Vedoucí práce:** doc. Ing. Tomáš Kubálek, CSc.

TÖRÖK, Ernest. *Business Intelligence: Zpracování podnikových dat*. Praha: ČVUT 2023. Bakalářská práce. České vysoké učení technické v Praze, Masarykův ústav vyšších studií.



**MASARYKŮV ÚSTAV  
VYŠŠÍCH STUDIÍ  
ČVUT V PRAZE**

## Prohlášení

Prohlašuji, že jsem svou bakalářskou práci vypracoval samostatně. Dále prohlašuji, že jsem všechny použité zdroje správně a úplně citoval a uvádím je v přiloženém seznamu použité literatury.

Nemám závažný důvod proti zpřístupnění této závěrečné práce v souladu se zákonem č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění.

V Praze dne: 20. 4. 2023

Podpis:

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Török** Jméno: **Ernest** Osobní číslo: **485058**  
Fakulta/ústav: **Masarykův ústav vyšších studií**  
Zadávací katedra/ústav: **Institut manažerských studií**  
Studijní program: **Ekonomika a management**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Business intelligence - Zpracování podnikových dat**

Název bakalářské práce anglicky:

**Business Intelligence - Business Data Processing**

Pokyny pro vypracování:

Bakalářská práce se zabývá problematikou zpracování firemních dat, jejich využitím při určování a naplňování klíčových indikátorů výkonu.

V teoretické části se práce zabývá tím, jaké možnosti BI přináší. Praktická část se zabývá zpracováním firemních dat a jejich analýzou v prostředí aplikací Power BI. Struktura práce: 1. Stanovení cílů práce. 2. Definice, teorie, možnosti BI. 3. Praktická část - analýza dat.

Metody: deskripce, analýza.

Seznam doporučené literatury:

MOSS, Larissa T., Shaku ATRE a Edward YOURDON. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley Professional, 2003. ISBN 978-0201784206.  
SHERMAN, Rick. Business Intelligence Guidebook: From Data Integration to Analytics. USA: Morgan Kaufmann, 2014. ISBN 978-0124114616.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

**doc. Ing. Tomáš Kubálek, CSc. Masarykův ústav vyšších studií ČVUT v Praze**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **09.12.2022**

Termín odevzdání bakalářské práce: **27.04.2023**

Platnost zadání bakalářské práce: \_\_\_\_\_

doc. Ing. Tomáš Kubálek, CSc.  
podpis vedoucí(ho) práce

Ing. Dagmar Skokanová, Ph.D.  
podpis vedoucí(ho) ústavu/katedry

prof. PhDr. Vladimíra Dvořáková, CSc.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

\_\_\_\_\_ Datum převzetí zadání

\_\_\_\_\_ Podpis studenta

## Poděkování

Rád bych věnoval toto poděkování všem, kteří mi jakýmkoliv způsobem pomohli s realizací této práce. Především chci poděkovat svému vedoucímu práce za jeho dohled a pomoc s odbornou a formální stránkou práce. Svému nadřízenému z firmy, který mi vytvořil možnost zpracovávat konkrétní data a pomáhal při komunikaci s vedením firmy. V neposlední řadě chci věnovat toto poděkování svým rodičům, spolužákům a kamarádům za jejich čas a snahu o porozumění a podporu při psaní této práce.

## Abstrakt

Bakalářská práce se zabývá problematikou zpracování firemních dat pomocí business intelligence, jejich využitím při určování a naplňování klíčových indikátorů výkonu a jiných strategických cílů. V teoretické části se práce zabývá základy BI – strukturou, jednotlivými komponenty a nástroji, a tím, jaké možnosti BI pro firmy přináší. Praktická část se zabývá zpracováním firemních dat a jejich analýzou v prostředí aplikací Power BI.

## Klíčová slova

Business Intelligence, analýza dat, datový sklad, ETL, OLAP, dolování dat, KPI, Microsoft Power BI

## Abstract

The bachelor's thesis deals with the issue of processing company data with the use of business intelligence, its use in determining and fulfilling key performance indicators as well as other strategic objectives of the company. In the theoretical part the thesis describes the BI basics – including the structure of BI, individual components, and tools, as well as the possibilities of usage for the companies. In practical part is theory applied to real business data which are processed and analysed in Microsoft Power BI app.

## Keywords

Business Intelligence, data analysis, data warehouse, ETL, OLAP, data mining, KPI, Microsoft Power BI

# Obsah

Úvod .....	9
1 Business Intelligence .....	12
1.1 Z historie .....	12
2 Obecná architektura BI .....	13
2.1 Datová vrstva .....	13
2.1.1 ETL – Extract, transform, load .....	14
2.2 Databázová vrstva .....	15
2.2.1 Datový sklad, jezero, tržiště .....	16
2.3 Analytická vrstva .....	16
2.3.1 OLAP – Online Analytical Processing .....	16
2.3.2 Data mining .....	18
2.4 Uživatelská vrstva .....	18
3 Využití BI .....	20
3.1 KPI .....	20
3.2 Využití ve firmách .....	21
3.2.1 Veřejná informovanost .....	22
4 Nástroje BI .....	23
4.1 Microsoft Power BI .....	24
4.1.1 Power Query .....	24
4.1.2 Jazyk M nebo DAX .....	25
4.2 Alternativy k Power BI .....	26
5 Definice problému .....	28
5.1 Cíle praktické části .....	28
5.1.1 Použití zařízení .....	29
6 Implementace řešení .....	30
6.1 Získání dat .....	30
6.1.1 Popis dat .....	31
6.1.2 Tvorba datového modelu .....	31
6.1.3 Relace .....	33
6.1.4 Kalendář .....	33
6.1.5 Úpravy dat pomocí DAX .....	34
6.2 Vizualizace a analýza dat .....	35
6.2.1 Hlavní stránka .....	35
6.2.2 Informace o zpracovaných datech .....	36

6.2.3	Maticе použitі	37
6.2.4	Sledování údržby	39
6.2.5	Mapa klientů a zařízení	40
6.2.6	Režimy	41
	Závěr.....	43
	Seznam použité literatury.....	44
	Seznam obrázků.....	47



# Úvod

Informace jsou všude okolo nás. V dnešní době se však informační tok ještě zvýšil díky novým technologiím a vynálezům, mezi které patří například internet, mobilní zařízení, cloudové technologie, umělá inteligence, big data, IoT (Internet věcí) a mnoho dalších. Tyto technologie a vynálezy umožňují rychlejší a snadnější výměnu informací mezi lidmi a zařízeními, což má velký vliv na podnikání a ekonomiku obecně.

V teoretické části mé práce se budu zabírat teorií a základy Business Intelligence (BI), což je soubor technologií, nástrojů a metodik, které umožňují sběr, zpracování, analýzu a prezentaci podnikových dat. BI umožňuje získávat užitečné informace, které pomáhají vylepšovat a optimalizovat různé procesy v podniku. K tomu dochází po několika krocích, jako je extrakce, transformace a nahrávání dat (ETL), analýza dat pomocí OLAP krychlí a technologií jako třeba data mining. Existuje mnoho nástrojů pro BI, známé jsou například Tableau, Mode, Datapine nebo Microsoft Power BI, které umožňují vizualizovat a analyzovat data prostřednictvím BI dashboardů a reportingu. Tyto nástroje jsou klíčové pro efektivní zpracování a prezentaci dat a jsou nezbytné pro úspěšné využití BI v podnikovém prostředí.

V praktické části mé práce se zaměřím na konkrétní případovou studii. Analyzuji data firmy pracovně nazvané XY pomocí nástroje Microsoft Power BI, což je jeden z nejpoužívanějších nástrojů pro BI. S pomocí Power BI budu schopen vizualizovat a analyzovat data dodaná firmou v rámci Projektu X. Díky tomu budu moci poskytnout podrobnou analýzu a doporučení pro zlepšení podnikových procesů a rozhodování při dalším vývoji v rámci tohoto projektu.

Je důležité mít na paměti, že množství informací a dat z těchto zdrojů může být obrovské a těžko srozumitelné. Proto je třeba mít vyspělé nástroje a technologie, jako jsou BI dashboardy a reporting, které umožní rychlé a efektivní zpracování a prezentaci dat. S využitím takovýchto nástrojů bude práce s daty mnohem snazší a efektivnější.

## Seznam použitých zkratek

B2B - Business to business

BI - Business Intelligence

CFO - Chief Financial Officer

CRM - Customer Relation Management

DAX - Data Analysis Expressions

DSA - Data Staging Area

DSS - Decision Support System

EAI - Enterprise Application Integration

ER - Entity relation (diagram)

ERP - Enterprise Resource Planning

ESS - Executive Support System

ETL - Extract, Transform, Load

IaaS - Infrastructure as a Service

IoT - Internet of Things

KPI - Key Performance Indicator

MIS - Management Information System

OLAP - Online Analytical Processing

PaaS - Platform as a Service

SaaS - Software as a Service

# TEORETICKÁ ČÁST

# 1 Business Intelligence

Definice Business Intelligence najdeme mnoho, protože se jedná o zastřešovací pojem, kterým lze označit nástroje, metody, osvědčené postupy a další. Cíl BI je však vždy stejný: dodat správné informace správnému člověku ve správný čas [1 p. 62].

Podle IBM Business Intelligence je software, který přijímá různé typy firemních dat a prezentuje je uživatelům srozumitelným způsobem pomocí reportů, dashboardů, schémat a grafů. [2]

Podle Microsoftu Business Intelligence je sada nástrojů analyzujících data, které prezentují ve vizuálně atraktivní podobě, čímž podporují strategická rozhodnutí. [3]

Trochu komplexnější a lepší definici najdeme v publikaci Petra Suchánka: „*Business Intelligence (BI) je množina metod, přístupů a aplikací informačních systémů (IS) a informačních a komunikačních technologií (ICT – Information and Communication Technology) určených pro podporu řídicích, analytických a plánovacích činností podniků umožňujících nahlížet na danou problematiku z různých úhlů pohledu – multidimenzionálně.*“ [4]

## 1.1 Z historie

Business Intelligence začala vznikat až ve 20. století, kdy tento pojem v roce 1958 zavedl v povědomí zaměstnanec IBM Hans Peter Luhn, i když jako první použil tento pojem Richard Millar Devens už v roce 1865. [5; 2]

V 60. a 70. letech 20. století začaly vznikat první systémy pro správu dat a tzv. Decision support systems (DSS) pro lepší správu rychle rostoucího objemu dat, díky vynálezu hard disků, které byly revoluční technologií na poli ukládání dat. [5; 2]

V roce 1988 proběhla mezinárodní konference v Římě, jejíž cílem bylo sjednotit procesy zpracování a ukládání dat. Po této události začala éra moderní BI. O rok později se Howard Dresner, analytik společnosti Gartner Group, zasloužil o zavedení pojmu *business intelligence* do běžné řeči jako souhrnný pojem zahrnující datová úložiště, datové analýzy a další pojmy jako DSS nebo EIS (Executive information system). [5]

Společně s vývojem datových úložišť, který se dále ubíral směrem datových skladů, vznikly i nástroje jako ETL (Extract, Transform, Load) viz kapitola 2.1.1 nebo OLAP (Online Analytical Processing) viz kap. 2.3.1, které jsou dnes již základem BI. Tyto systémy již jsou postaveny na multidimenzionálních databázích a umožňují nové metody dolování dat (angl. data mining). [5; 4]

Koncem 20. století BI sestávalo ze dvou základních funkcí: tvorbou dat a jejich vizualizací v prezentovatelné podobě. Tuto fázi dnes označuje jako Business Intelligence 1.0. [5]

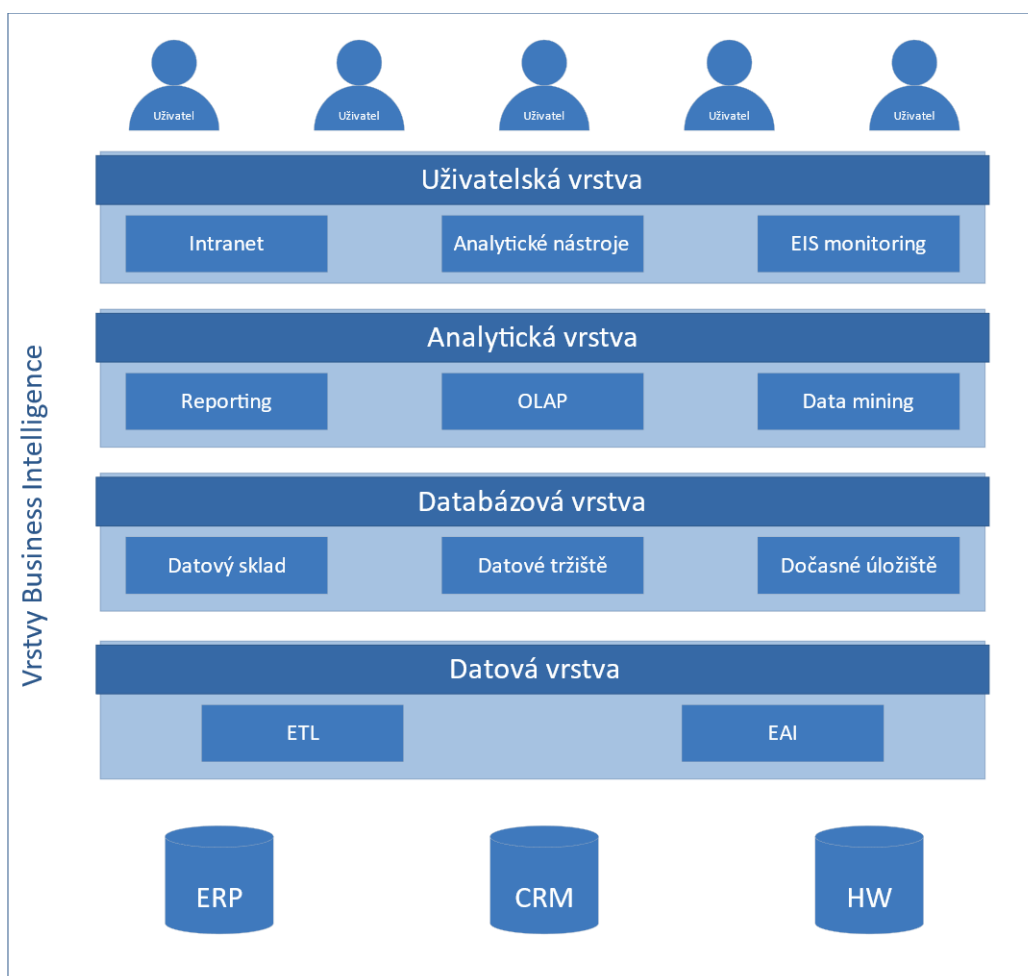
Business Intelligence 2.0 začíná s rozvojem technologií v začátcích 21. století, kdy výrazně pomáhají s problémy složitosti a rychlosti, především rozvoj cloudových technologií. [5]

## 2 Obecná architektura BI

Při navrhování řešení BI je potřeba vycházet z potřeb podniku. Je tedy důležité nejdříve znát požadavky, které firma má, a až poté lze navrhovat BI systém – nejdříve se navrhne logická část, tedy co by se mělo dělat, poté se bude řešit fyzická část, tedy jak se to bude dělat. [6 p. 1]

Aplikace BI se může různě lišit v implementaci konkrétního řešení, vždy ale bude vycházet z obecné architektury BI, která se skládá následujících vrstev:

- datová vrstva,
- databázová vrstva,
- analytická vrstva,
- uživatelská vrstva.



OBR. 2.1 ARCHITEKTURA BUSINESS INTELLIGENCE. PŘEVZATO Z [4]

Jednotlivé vrstvy jsou dále složeny z komponent BI, kterých je velmi mnoho, například transformační nástroje ETL, datové sklady, OLAP, reporting, manažerské aplikace a další. [4]

### 2.1 Datová vrstva

Na začátku všeho se nachází aplikační software a/nebo zařízení, která data generují ať už automaticky nebo rukou člověka. Tato data mají nejrůznější podobu, často jsou nestructurovaná, bez rozlišení v čase. V datové vrstvě BI přebírá vstupní data z různých zdrojů ve strukturované

i nestrukturované formě, nejčastěji pomocí metody ETL. Data jsou jednoduše vyextrahovaná ze zdrojů, transformována do požadované podoby a předána další vrstvě k uložení. [3; 4]

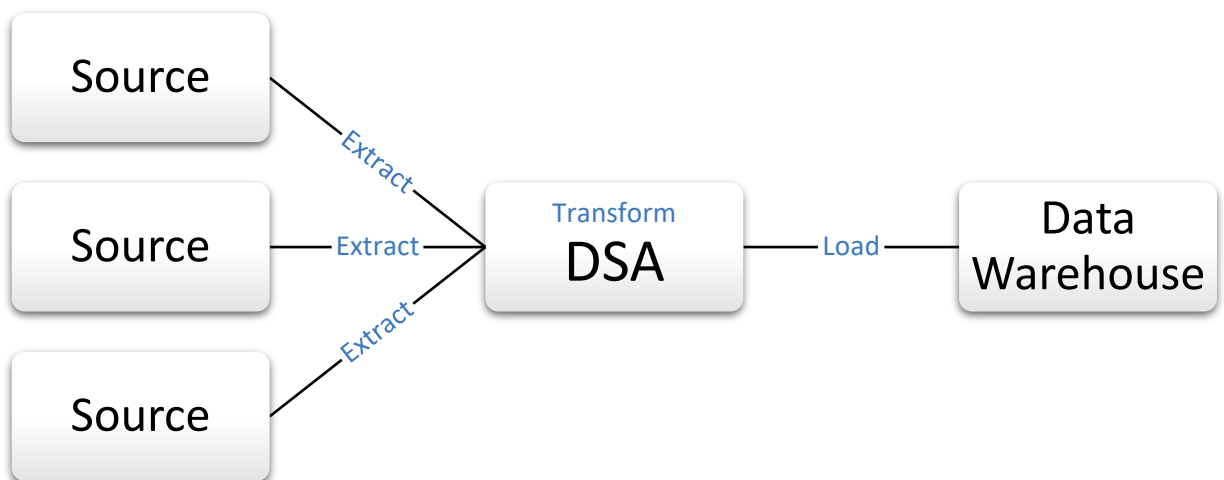
### 2.1.1 ETL – Extract, transform, load

Data je potřeba nejprve získat – extrahovat (*Extract*) – z původních zdrojů, které můžeme dále dělit na vnitřní (CRM, ERP, ...) a vnější (státní správy, statistické údaje, veřejné rejstříky, ...). Tato data se nahrávají do tzv. dočasných úložišť (angl. *Data staging area*), kde poté dochází k druhé fázi – transformace. [4]

Proces transformace (*Transform*) dat zahrnuje očištění dat – je potřeba odstranit data, která jsou nekompletní nebo jinak poškozená a mohly by narušit integritu datové sady. Dalším častým problémem je diverzita dat, kdy sbíráme stejná data z různých zdrojů v různém formátu. Standardně jeden zákazník může mít v různých systémech jiná ID (např. abcd-efgh-ijkl-wxyz nebo 200524) nebo každý zdroj může používat jiný formát pro datum (např. americký: MM/DD/YYYY nebo mezinárodní: YYYY-MM-DD). [4; 7]

Připravená data poté nahrajeme (*Load*) do datového skladu, kde jsou uschována a připravena k další práci. Tento proces je náročný, protože se často jedná o přesun velkého množství dat. Můžeme volit ze dvou přístupů:

- Plné nahrání dat: původní data v databázi se smažou a nahradí se novým datasetem, který v sobě zahrnuje data již dříve zahrnutá v původním datasetu. Jedná se o nejjednodušší způsob nahrání dat.
- Postupné nahrání dat: V případě, že je dat obrovské množství, jsou rozdělena do menších částí a do databáze nahrávána postupně. [7]



OBR. 2.2 ETL JE PROCES POSTUPNÉHO ZÍSKÁVÁNÍ DAT Z RŮZNÝCH ZDROJŮ, JEJICH TRANSFORMACE ULOŽENÍ DO DATOVÝCH ÚLOŽIŠŤ

## 2.2 Databázová vrstva

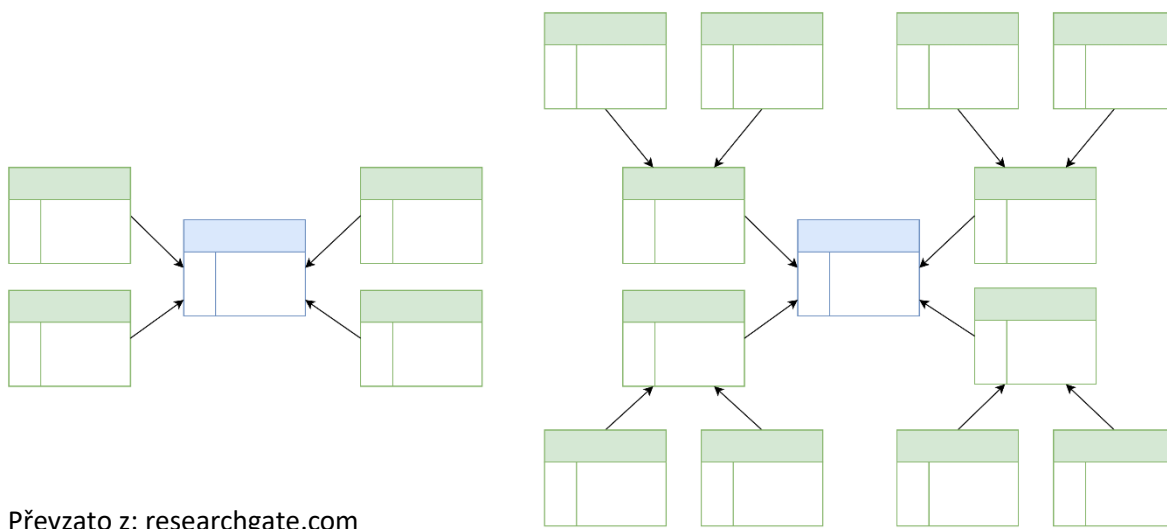
Databázová vrstva se stará o uložení dat, abychom k nim měli v budoucnu přístup. Data jsou ukládána do datových skladů a tržišť anebo do krátkodobých úložišť – operativní datová úložiště nebo dočasná datová úložiště.

Obecně uznávaným způsobem ukládání dat pro potřeby BI je pomocí dimenzionálního modelu. Tento model sestává ze tří částí: dimenzí, faktů a vazeb mezi nimi. Toto uspořádání umožňuje dodávat uživatelům snadněji pochopitelná data a zvyšuje rychlost dotazování. Tabulka faktů obsahuje události a fakta, která sledujeme, zatímco tabulka dimenzí je složena z hierarchicky uspořádaných informací, které dodávají kontext jednotlivým faktům. [6; 4]

Například tabulka objednávky může být faktovou tabulkou obsahující referenční číslo produktu, datum objednávky a počet objednaných kusů. Dimenze související s touto tabulkou jsou dimenze času a produktu, kde čas je hierarchicky uspořádán jako roky, měsíce, dny a produkty mohou být dále členěny do jednotlivých kategorií zboží. Tabulka objednávek tedy obsahuje fakta – informace – o tom, kdy, kolik a čeho bylo objednáno. Z dimenzionální tabulky produktů můžeme dohledat konkrétní vlastnosti prodaného zboží, například název, hmotnost nebo jednotkovou cenu.

Dimenzionální model je také velmi důležitý pro následnou OLAP analýzu, viz str. 16.

U dimenzionálních schémat rozlišujeme několik typů struktur. Základní struktura se jmenuje *star* a její podobu lze vidět na obr. 2.3 vlevo. Základem je jedna faktová tabulka, na kterou je napojeno několik dimenzí. V případě, že jsou dimenzionální tabulky dále děleny do menších celků (na obr. 2.3 vpravo) jedná se o typ struktury, který se nazývá *snowflake*. [4; 8]



Převzato z: researchgate.com

OBR. 2.3 SROVNÁNÍ RŮZNÝCH SCHÉMÁT DATOVÝCH STRUKTUR. STAR (VLEVO) A SNOWFLAKE (VPRAVO)

Faktové tabulky tedy obsahují transakce – informace o tom, co se stalo (objednávka, změřená teplota, známka z testu) a cizí klíče<sup>1</sup> odkazující na odpovídající záznamy v dimenzionálních tabulkách. Dimenzionální tabulky mají popisný/deskriptivní charakter. Obsahují data, která by se ve faktových tabulkách často opakovala (například o produktech, lokacích, kde jsou umístěny teploměry, studentech nebo testových otázkách, ze kterých student získal výslednou známku). [8]

<sup>1</sup> Cizí klíč (angl. *Foreign Key*) je pojem z oblasti databází. Jedná se o identifikátor, pomocí kterého dokážeme přiřadit záznamy z jedné tabulky k záznamům v druhé tabulce.

Dimenzionální tabulky nám pomáhají při filtrování a seskupování dat. Podle těchto filtrů poté agregujeme data z faktových tabulek. Například můžeme počítat počet prodaných fotoaparátů podle jednotlivých značek, průměrnou teplotu v Praze nebo medián známky z testu z matematiky. Tyto agregované hodnoty nazýváme míry.

*„V návrhu schématu star je mírou sloupec tabulky faktů, který ukládá hodnoty, které mají být shrnuty.“ [8]*

Při návrhu datového modelu podle schématu star (nebo snowflake) je důležité zajistit, aby relace mezi dimenzionální a faktovou tabulkou měla kardinalitu 1:N. To znamená, že jeden záznam v dimenzionální tabulce může odpovídat více záznamům ve faktové tabulce (student může napsat více testů) a současně každý záznam ve faktové tabulce může odpovídat maximálně jednomu ze záznamu z tabulky dimenze (každý test psal pouze jeden student).

### 2.2.1 Datový sklad, jezero, tržiště

Z mnoha důvodů je vhodné mít všechna uložena na jednom místě, především pro snadný a rychlý přístup nebo minimalizaci redundantních informací. Podle typu strukturalizace dat a účelu ukládáme tyto data do datových skladů, tržišť a jezer.

Datové jezero je úložiště, které bychom velmi laicky mohli přirovnat ke složce „Stažené soubory,“ ve které se nachází mnoho různých, navzájem nesouvisejících souborů. Tyto soubory – data – mohou být již strukturovaná, částečně strukturovaná nebo i nestruturovaná. Mezi daty také nemusejí být vazby (relace). Zdrojem dat pro datová jezera bývají například sociální média, IoT nebo jiné velké objemy dat. Tato data bývají používána k real-time analýzám, prediktivním analýzám a strojovému učení. [9]

Datový sklad je úložiště dat, které umožňuje úschovu strukturovaných a částečně strukturovaných dat. Tyto data lze nahrávat z více různých zdrojů, například z SQL databází, firemních systémů, souborů typu Excel, XML, JSON a dalších. Tato data jsou připravena k další analýze. Data do datových skladů získáváme například z různých podnikových aplikací a využíváme je k základnímu generování sestav a BI. [9]

Datové tržiště je podmnožinou datového skladu. Představuje předpřipravenou sadu dat, která jsou připravená pro konkrétní skupinu lidí, oddělení, divize atd. [4]

## 2.3 Analytická vrstva

Do analytické vrstvy spadají tři části: OLAP, data mining a tvorba reportů. Pro tvorbu kvalitních analytických závěrů je potřeba mít správně zpracovaná data vstupující do fáze analýzy. Tato data by měla být:

- přesná a správná,
- konzistentní,
- kompletní,
- neredundantní. [4]

### 2.3.1 OLAP – Online Analytical Processing

OLAP je zkratka pro Online analytical processing, a jak lze pochytit z názvu, jedná se o procesy analyzování dat. OLAP provádí vícerozměrné analýzy dat za pomoci tzv. vícerozměrných kostek a poskytuje uživateli například komplexní výpočty, analýzy nebo forecasting. [10]

Zdrojová data pro OLAP, jedná se standardně o databáze uložené v datových skladech, jsou hierarchicky uspořádaná do datových kostek, které výrazně urychlují práci s daty a dotazováním.

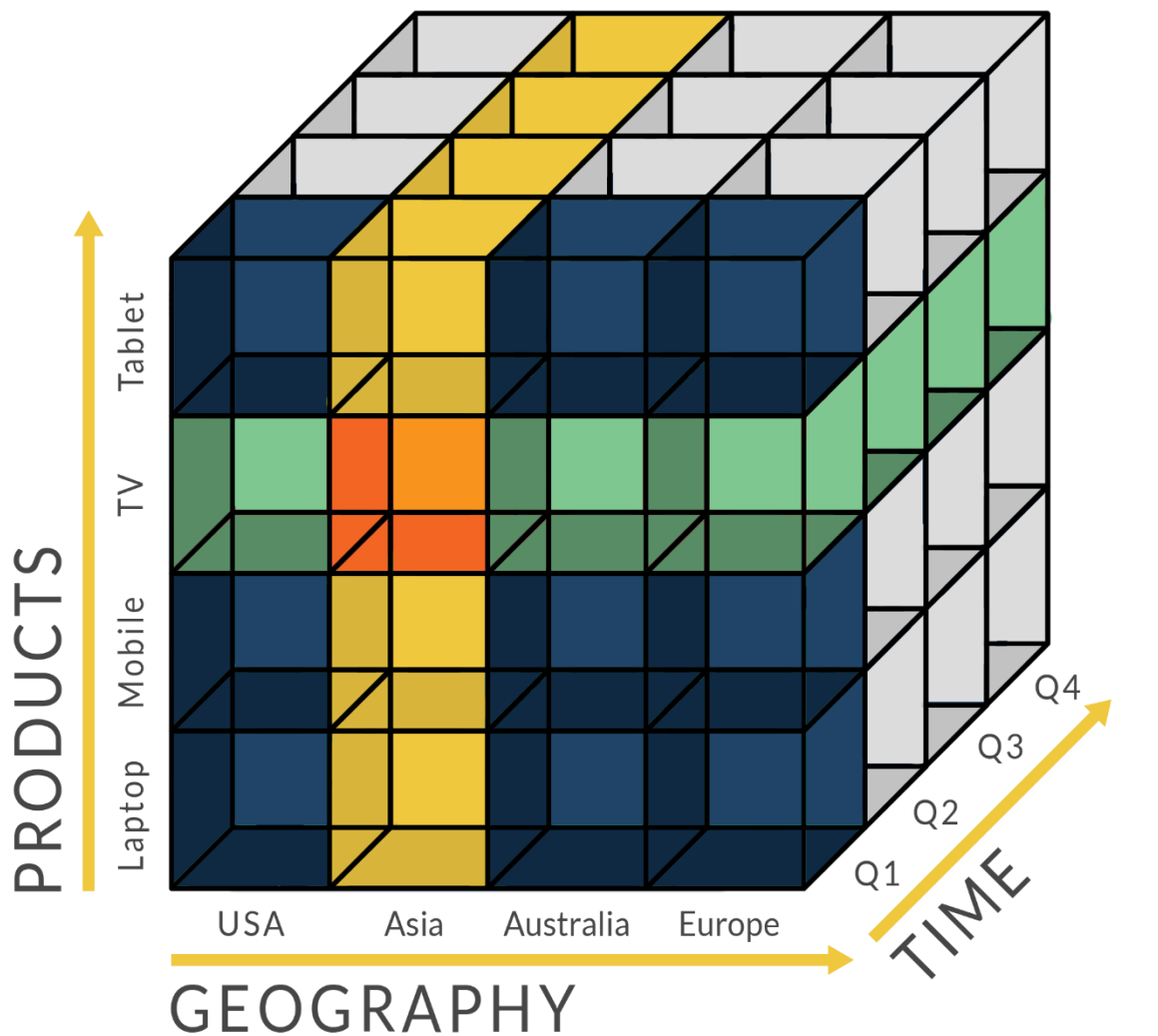


Multidimenzionální kostka nebo též krychle (z angl. cube) je označení pro datovou strukturu, s kterou OLAP pracuje. Je složena z faktových a dimenzionálních tabulek.

Microsoft multidimenzionální krychli definuje následně:

„Datová struktura, která agreguje míry podle úrovní a hierarchií jednotlivých dimenzí, které chcete analyzovat. Datové krychle kombinují několik dimenzí, jako je čas, zeměpis a produktové řádky, se souhrnnými daty, jako jsou údaje o prodeji nebo zásobách. Krychle nejsou "krychle" v striktně matematickém smyslu, protože nemusí mít nutně stejné strany. Je to ale výstižná metafora pro složitý koncept.“ [11] Datová krychle může kromě různých stran mít také více než 3 dimenze (strany).

Dimenzionální tabulky obsahují jednotlivé dimenze – souhrn členů uspořádaných do hierarchického uskupení, pomocí kterých uživatel analyzuje data. Na obrázku obr. 2.4 jsou dimenze *Products*, *Geography*, *Time*.



Dostupné z [olap.com](http://olap.com)

OBR. 2.4 MULTIDIMENZIONÁLNÍ KOSTKA OLAP

V jednotlivých kostkách vytvořených průsečíky jednotlivých dimenzí jsou míry, které představují nějakou agregovanou hodnotu, typicky součet, minimum, maximum nebo průměr.

### 2.3.2 Data mining

Dolování dat, známé též jako objevování znalostí v datech, je proces odhalování vzorců ve velkých data setech pomocí matematické analýzy. S pomocí nástrojů určených pro data mining, například Apache Spark od IBM nebo SSaaS od Microsoftu, můžeme popisovat obrovské objemy dat, nebo v návaznosti na moderní algoritmy strojového učení předvídat výsledky. Data mining najde své uplatnění na mnoha různých místech ať už jde o různá oddělení ve firmách (obchodní, marketingové) či například při vzdělávání, optimalizaci procesů nebo odhalování podvodů. [12; 13]

Proces dolování dat je obdobný jako při tvorbě BI. Na začátku je třeba stanovit cíle nebo potřeby firmy. Toto bývá často náročný a podceňovaný krok. Datové modely pro data mining mohou být použity například k předpovídání (forecasting), analýze rizik nebo pravděpodobnosti, hledání sekvencí, každá z těchto možností bude využívat jiný datový model. [12; 13]

V dalším kroku přichází na řadu sběr dat, jejich čištění, kontrola odstranění vad. Stejně jako při OLAP analýze aj. i zde mohou data pocházet z mnoha různých zdrojů, které je třeba sjednotit, vyčistit od nekompletních dat, nebo mohou obsahovat špatné informace. Tato činnost by byla velmi náročná při kontrole dat záznam po záznamu, proto je vhodné využít speciální nástroje a filtry. Na rozdíl od BI se zde může úmyslně snižovat počet dimenzí pro lepší výkon. [13; 12]

V třetí fázi se provádí analýza dat. Podle způsobu lze volit různé metody hledání vztahů mezi daty: hledání sekvencí, asociace nebo korelace dat patří mezi ty nejčastěji užívané. Zajímavá jsou také data, která se naopak velmi odlišují od zbytku data setu. Tímto způsobem lze například odhalovat podvody. [12]

Podle požadovaného cíle můžeme vybírat algoritmy podle typu:

- Klasifikační algoritmy
- Regresní algoritmy
- Segmentační algoritmy
- Asociativní algoritmy
- Algoritmy pro hledání sekvencí [14]

V poslední fázi je třeba interpretovat získané výsledky a aplikovat je do praxe. Finální výsledky by měly být srozumitelné validní a použitelné. Na jejich základě může firma dále vymýšlet nové strategie a dosahovat svých cílů. [12]

## 2.4 Uživatelská vrstva

Jedná se o vrstvu, které umožňuje uživatelům přistupovat k datům pomocí různých aplikací a nástrojů podporujících správu reportů a dashboardů. V rámci uživatelské vrstvy vystupují do popředí různé systémy, pomocí kterých mohou uživatelé sledovat a dále interagovat s výstupy datových analýz, například:

- EIS (Executive information system),
- ES (expertní systémy),
- DSS (Decision support system).

EIS je souhrnné označení pro aplikace podporující rozhodování manažerů ve firmách a podávají přehled o celkovém stavu firmy a efektivitu měřených dat, nejčastěji v podobě reportů. Tyto aplikace nabízí jednoduché a srozumitelné uživatelské rozhraní, které dovoluje uživatelům měnit parametry dotazů a získat tím chtěné výsledky. Ačkoli dříve byly tyto aplikace zaměřené pouze na vyšší management, v dnešní době se dostávají do rukou také nízkému (operativnímu) managementu. [4]

Obdobné jsou také systémy DSS, které slouží primárně pro střední a nižší management a operativní rozhodování. DSS prošlo časem pěti generacemi vývoje. První generace byla orientovaná na konkrétní problém, druhá již pracovala s lepšími návrhy databází, třetí nabízela podporu rozhraní mezi uživatelem a systémem. Čtvrtá generace nabídla modulární řešení a pátá generace umělou

inteligenci. [4] Tyto systémy lze využít například pokud bychom chtěli uvést nový produkt na trh, pomocí DSS můžeme provést několik různých scénářů s modelem propojujícím například cenu výrobku, náklady na výrobu a náklady na marketing. Změnou ceny produktu lze sledovat změny v očekávaných výsledcích a vybrat tak tu nejlepší. [15]

Expertní systémy jsou aplikace, které pomocí umělé inteligence simulují rozhodnutí na základě poskytnuté báze znalostí a interferenčního mechanismu. Často se jedná o velmi úzce zaměřené úkoly, při jejichž řešení je vyžadován odborník na danou problematiku. První expertní systém byl vyvinut v roce 1965 Edwardem Feigenbaumem a Joshua Lederbernem ze Stanfordské univerzity v Kalifornii. Znalostní báze obsahuje informace o tom, jak jsou určité situace řešeny za standardních podmínek a interferenční mechanismus představuje soubor algoritmů schopných dané situace řešit pomocí poskytnuté báze. Mezi typické úkoly interferenčních mechanismů patří klasifikace, diagnózy, sledování nebo plánování specializovaných činností. [4; 16]

## 3 Využití BI

Jak již bylo zmíněno dříve, smyslem BI je podpora firemních procesů a rozhodování, a to napříč všemi odděleními podniku. Svého využití nalezne BI v různých odděleních a nabízí oporu v různých situacích, ať už jde o řešení problému nebo hledání nových příležitostí. V této kapitole se pokusím demonstrovat některé ze standardních aplikací BI.

### 3.1 KPI

Klíčový ukazatel výkonnosti (angl. *Key performance indicators*) jednoduše označuje indikátor (metriku), kterým lze sledovat průběh plnění nějakého cíle. KPI se výrazně liší nejen podle toho, v jaké oblasti chcete dané metriky sledovat, ale také podle toho, kdo s danou informací bude pracovat. Jiné KPI bude sledovat vedení R&D oddělení a jiné KPI bude sledovat CFO. [17]

KPI vychází ze strategických cílů firmy a jsou jednou z pomůcek při jejich dosahování. Při nastavování KPI je tedy prvním krokem vytyčit si cíle, kterých chceme dosáhnout. Při nastavování strategických cílů je osvědčeným postupem nastavení tzv. SMART<sup>2</sup> cílů. [17]

V druhém kroku nastavujeme samotné KPI, tedy vlastnosti, kterými chceme sledovat průběh naplňování nastaveného cíle v prvním kroku. Pokud jsme si v prvním kroku nastavili jako cíl navýšení tržeb o 20 %, relevantní metriky budou určitě transakce, jejich objem a hodnota. [17]

Třetím krokem je určení časového intervalu vyhodnocování KPI. Ten bývá standardně měsíc, čtvrtletí nebo rok. [17]

KPI můžeme rozdělit podle časové relevantnosti na

- Lagging KPI: Jedná se o indikátor minulého/současného stavu; sledujeme tím, jak jsme si doposud vedli,
- Leading KPI: Jedná se o indikátor umožňující předpovídání budoucího stavu; počet prodaných vstupenek nás informuje o počtu lidí, kteří pravděpodobně dorazí na koncert. [17]

V balíčku nastavených KPI bychom měli mít vyvážený počet obou druhů KPI, tedy jak lagging, tak leading.

KPI lze dále kategorizovat podle jejich typu:

- vstupy,
- výstupy,
- procesy,
- výsledky,
- projektové KPI. [18]

---

<sup>2</sup> SMART cíle jsou jednou z metod nastavování cílů. Jedná se o akronym složený ze slov: *Specific* (konkrétní), *Measurable* (měřitelný), *Attainable* (dosažitelný), *Relevant* (relevantní), *Time-bound* (časově omezený) [35]

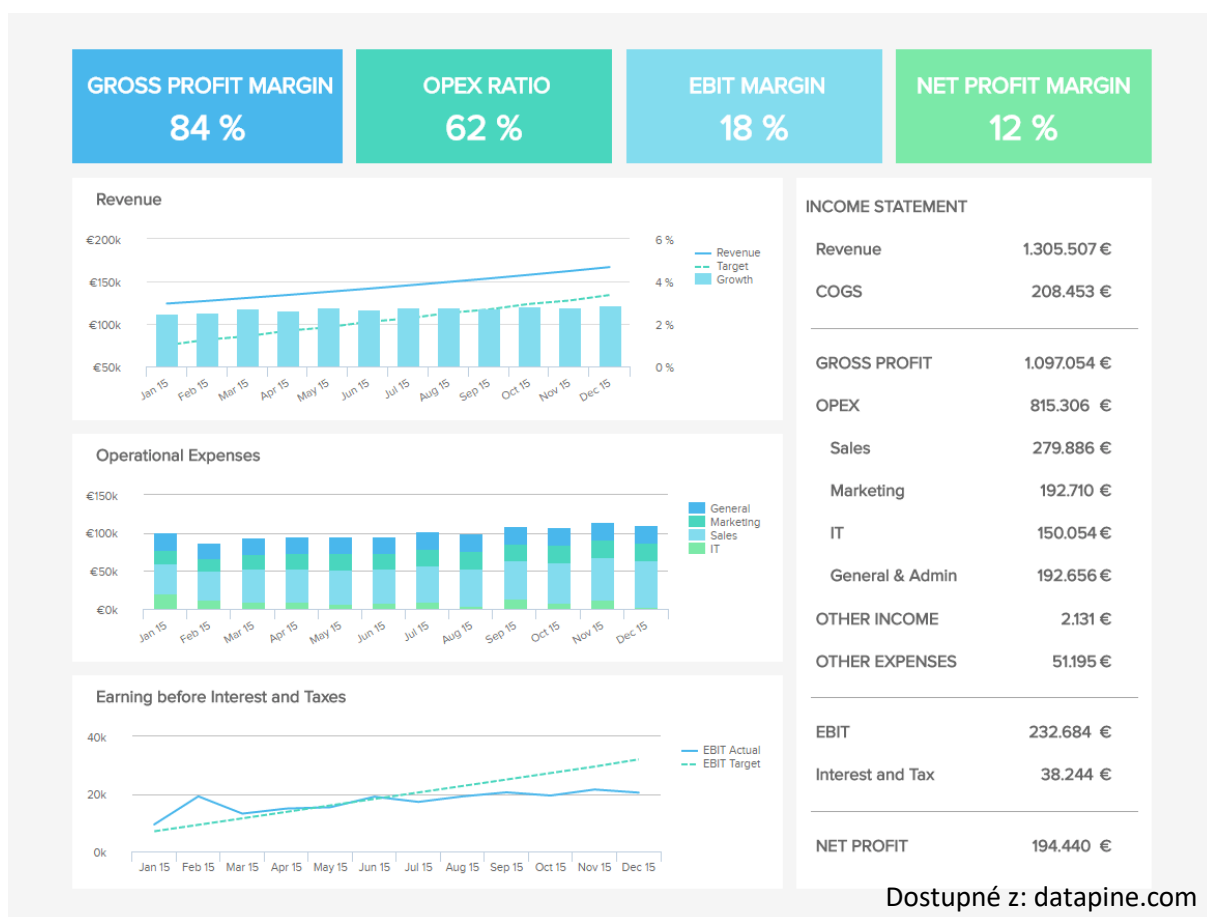
## 3.2 Využití ve firmách

### Management

Jedním z velmi důležitých odvětví, kde BI najde své uplatnění, je beze sporu management. Využití business intelligence v managementu má mnoho výhod. Pomáhá manažerům při rozhodování na základě datových analýz a predikcí. Díky těmto informacím mohou efektivně řídit svůj tým, sledovat klíčové ukazatele výkonnosti a přizpůsobit své rozhodnutí aktuálním podmínkám. [19]

### Finance

Další odvětví, kde business intelligence sehrává svoji roli, jsou bezesporu finance. Management tak může snadno sledovat finanční ukazatele v reálném čase. Na obr. 3.1 jsou například: marže hrubého zisku (GP), poměr operačních nákladů (OPEX) nebo zisk před zdaněním a úroky (EBIT). Na grafech lze sledovat vývoj tržeb, nákladů a EBIT. Po pravé stránce je zjednodušený výkaz zisků a ztrát. [19]



OBR. 3.1 PŘÍKLAD FINANČNÍHO DASHBOARDU. DOSTUPNÉ Z [19]

Zpracovávat data pomocí BI lze i z opačného pohledu, tedy například chceme-li spravovat nejen naše manažery, ale i investory o vývoji finanční situace stavu firmy. Zde můžeme sledovat

například ukazatel rentability aktiv – ROA<sup>3</sup> (angl. Return on Asset), rentability vlastního kapitálu – ROE<sup>4</sup> (angl. Return on Equity), míru zadlužení firmy nebo jednotkovou cenu akcií společnosti. [19]

### 3.2.1 Veřejná informovanost

Velké využití našlo BI také během pandemie COVID-19, kdy WHO, ministerstva i publikační společnosti mohly v reálném čase prezentovat shromažďovaná data pomocí interaktivních map, grafů, které se samy aktualizovaly a čtenář tak vždy viděl současný stav. Mezi typické grafy patřily grafy zobrazující počty nálezů a úmrtí, ale také stupeň různých opatření pro školy, firmy nebo nošení roušek. Ze stránek WHO můžeme zjistit, že od začátku pandemie do 16. dubna 2023 bylo evidováno 763,74 milionů případů nakažení a celkem na následky nemoci zemřelo přes 6,9 milionu lidí. Na jejich webu si lidé také mohou sami prohlédnout kdy, v kterých částech světa/kontinentu/země bylo kolik smrtí/nakažení. Tyto výstupy BI si může čtenář sám vyzkoušet a prohlédnout na oficiálních stránkách WHO: <https://covid19.who.int/> [20]

---

<sup>3</sup> ROA představuje poměr čistého zisku s úroky vůči celkovým aktivům, tedy jak dobře firma se svými aktivy hospodaří.

<sup>4</sup> ROE představuje poměr výsledku hospodaření vůči vlastnímu kapitálu.

## 4 Nástroje BI

V dnešní době existuje řada různých poskytovatelů BI služeb od jednodušších nástrojů po komplexní řešení. V této kapitole představím některé z možností a co nabízí.

### IaaS, PaaS, SaaS, Self-service

S pojmy self-service a SaaS se objevují velmi často napříč materiály o BI, jedná se také o moderní, důležité vlastnosti, které pomohly rozšířit svět BI, proto jim věnuji krátké vysvětlení:

IaaS = Infrastructure as a Service,

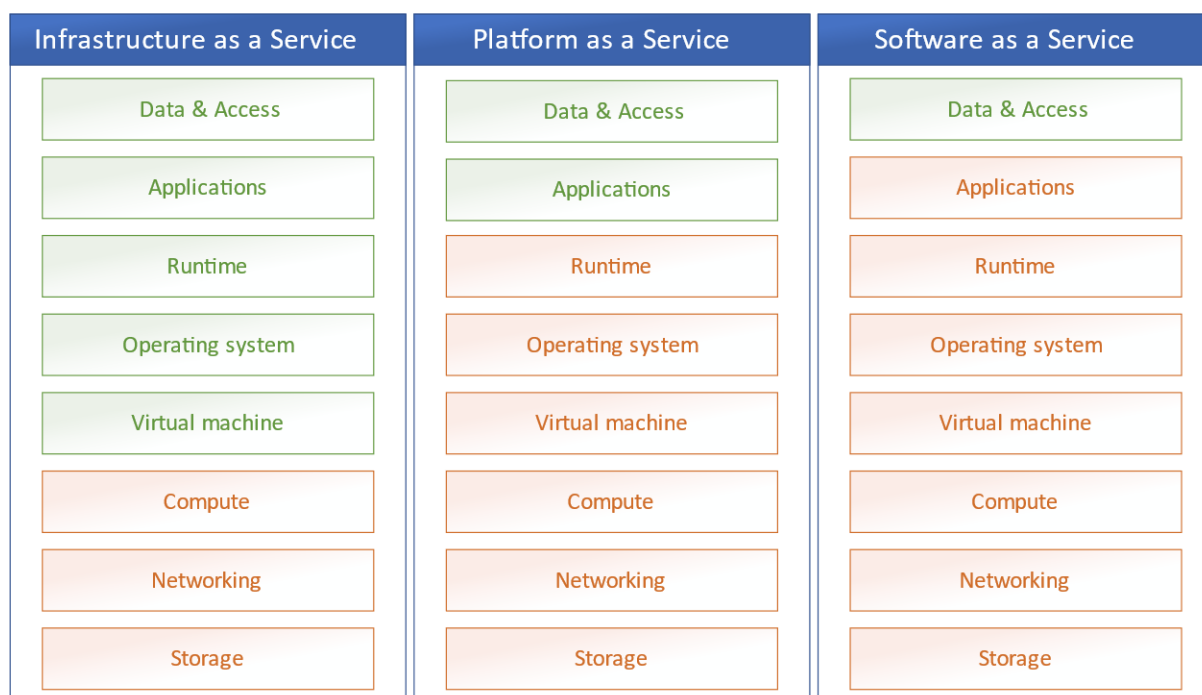
PaaS = Platform as a Service,

SaaS = Software as a Service.

Služby typu „as a service“ jsou cloudová řešení nabízená klientům za paušální cenu nabízející online přístup k určitým produktům. Rozdíl mezi jednotlivými službami je dobře vyobrazen na obr. 4.1.

Self-service označuje novou generaci nástrojů BI, které jsou navrženy tak, aby je dokázal používat jediný uživatel bez hlubších znalostí. Opakem je tzv. tradiční BI, pod kterým se skrývají dedikované IT oddělení expertů specializovaných právě na práci s daty a jakékoliv změny v reportech znamenaly mnoho práce navíc. I tento problém eliminují moderní nástroje velmi efektivně. [21],

SaaS BI označuje nástroje BI, které jsou uloženy v Cloudu poskytovatele a uživatel je používá online. Často i data, ke kterým se připojuje pomocí BI nástrojů bývají uložena u poskytovatele BI nástrojů jako SaaS.



OBR. 4.1 SROVNÁNÍ IAAS, PAAS A SAAS. O ČERVENÉ BLOKY SE STARÁ POSKYTOVATEL SLUŽBY, UŽIVATEL MÁ PŘÍSTUP K ZELENÝM BLOKŮM. PŘEVZATO Z [33]

## 4.1 Microsoft Power BI

Microsoft Power BI (dále jen Power BI<sup>5</sup>) je komplexní služba od společnosti Microsoft. K dispozici je jako cloudová aplikace SaaS, desktopová aplikace nebo jako aplikace pro mobilní zařízení. Power BI je dále rozdělena na dvě části: Power BI report builder – pro tvoření reportů a vizuálů a Power BI report server – pro sdílení vytvořených reportů. Power BI umožňuje spojování různých zdrojů dat, vizualizovat je a analyzovat. [9; 22]

Pro vypracování praktické části jsem si zvolil právě řešení Microsoft Power BI, a to proto, že mi na rozdíl od ostatních nástrojů není úplně cizí, a proto, že firma vlastní data již v současné době využívá cloudových služeb od Microsoftu, včetně Power BI, což výrazně zjednodušuje následnou implementaci řešení. Z tohoto důvodu je kapitola rozsáhlejší.

Základní stavební kameny Power BI, kterými jsou:

- dashboard,
- report,
- workbook,
- dataset
- dataflow.

Jsou rozděleny do kapacit a pracovních prostorů. Jako kapacitu Microsoft označuje koncept rozdělování hardwarových prostředků, jako je čas procesoru, operační paměť nebo velikost úložiště. Pracovní prostor je souhrn základních kamenů (dashboardů, reportů, datasetů, ...), který je dále rozdělen na vlastní a ostatní. [23]

Dataset je kolekce vstupních dat do modelu z různých zdrojů, ať už importovaných nebo připojených pomocí Direct Query. Dataflow je speciální zdroj dat uložený na serverech Azure, a do Power BI desktop se připojí pomocí speciálního konektoru. Workbook je také zvláštní případ datasetu připojeného do sešitu Excelu. [23]

Report je základní stránka, na kterou přidáváme jednotlivé vizuály. Tyto vizuály mají nejčastěji podobu grafů a tabulek. Reporty (sestavy) nebo jejich části mohou být připnuty na takzvaný dashboard (řídící panel), který je dostupný z webového rozhraní Power BI. V rámci jednoho dashboardu je možné vizualizovat data z mnoha různých datasetů a reportů.

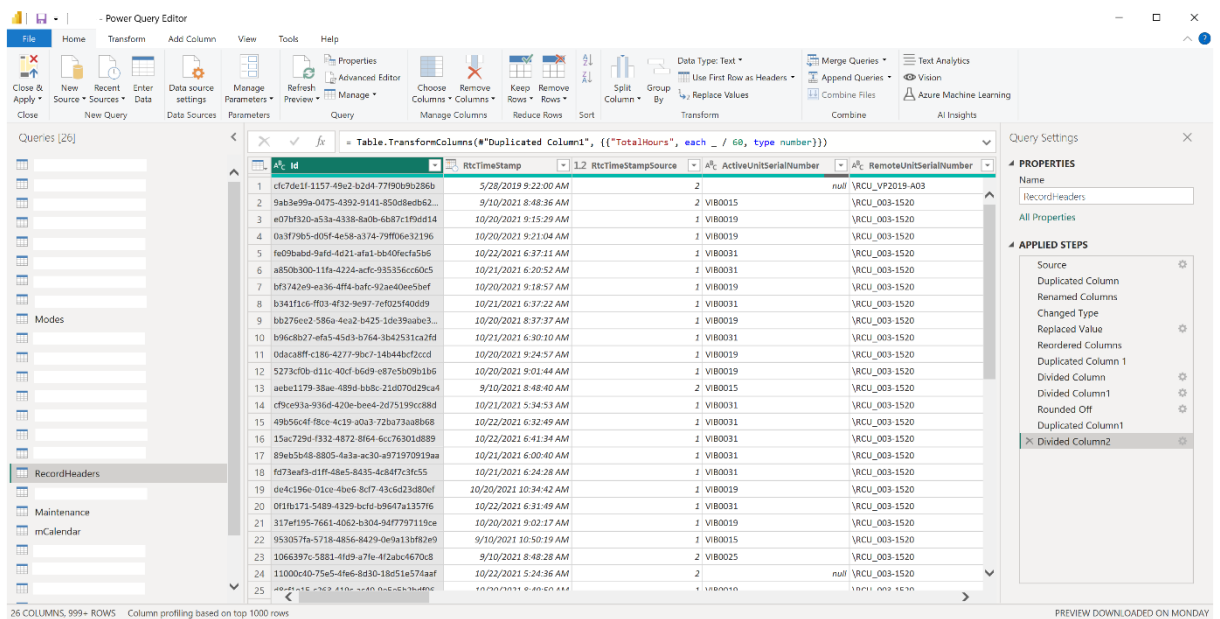
### 4.1.1 Power Query

Power Query editor, který lze nalézt také v Microsoft Excel, je vestavěný nástroj pro připojení, transformaci a úpravu dat před jejich načtením do datového modelu Power BI. Nabízí možnost připojení stovek různých zdrojů dat například MS Excel, SharePoint, Azure Data Lake, nebo běžných databázových systémů (MySQL, PostgreSQL, Oracle), přes 300 různých možností transformace (například změny datových typů, odstranění prázdných hodnot, změna hodnot, dopočítání nových hodnot) včetně moderních transformací pomocí umělé inteligence, a to vše bez znalosti programovacích jazyků. Kromě zmíněných aplikací Power BI a Microsoft Excel je integrováno do mnoha dalších produktů společnosti Microsoft, je podporováno v prostředí Power Automate a datové výstupy lze opětovně využít napříč celou škálou produktů. Pro pokročilejší uživatele je k dispozici jazyk M. [24]

---

<sup>5</sup> Pokud nebude uvedeno jinak, mám na mysli desktopovou aplikaci Power BI Desktop.





## OBR. 4.2 POWER QUERY

### 4.1.2 Jazyk M nebo DAX

Power BI pro transformaci a zpracování dat využívá dvou jazyků: **DAX** (Data analysis expressions) a **jazyk M** (dále jen M). Ač se to může zdát matoucí či nelogické, má to svůj význam.

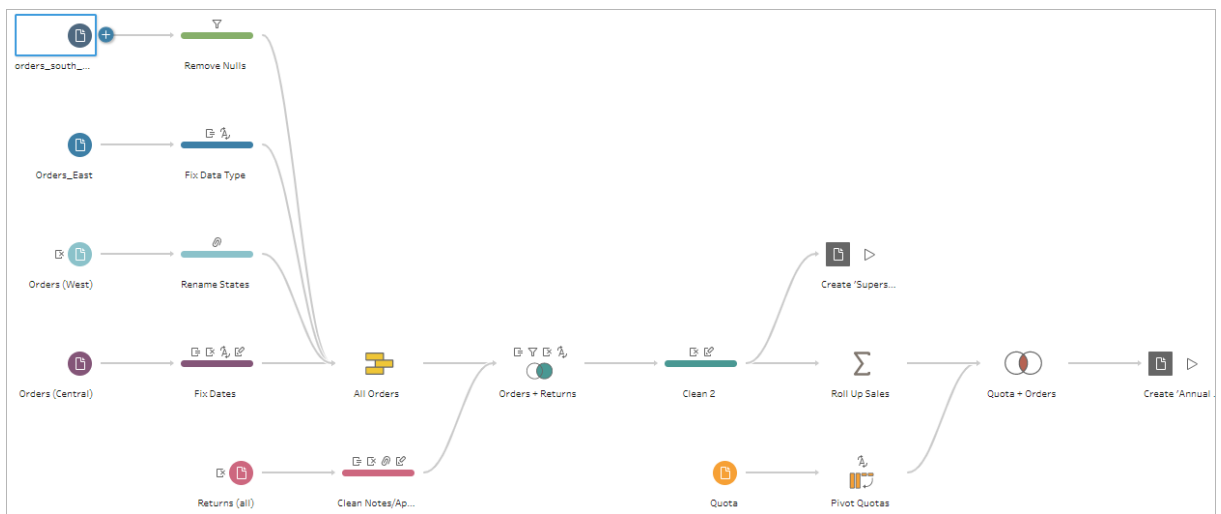
M je dotazovacím jazykem pro Power Query, který nám umožňuje filtrovat, kombinovat a jinak připravovat data ještě předtím, než se importují do datového modelu v Power BI. V editoru Power Query si po načtení zdrojů dat můžeme například vybírat, které sloupce chceme nebo nechceme do modelu načíst. Můžeme přejmenovat tabulky i sloupce nebo dopočítat další hodnoty, které ve zdrojových datech nejsou (např. kvůli normalizaci). [25; 26]

DAX je analytický jazyk pro počítání a kalkulace, kterým můžeme komplexně analyzovat datový model, který jsme si předtím vytvořili v Power Query editoru. Jedná se tedy o dva jazyky ve dvou různých fázích práce s daty, které spolu nejsou nijak propojeny. [26; 25]

## 4.2 Alternativy k Power BI

### Tableau

Tableau je další ze společností nabízející nástroje BI, existuje od roku 2013 a je jedním z hlavních hráčů na trhu nástrojů BI. Je známé především pro svoji schopnost vizualizovat data. V nabídce je mnoho různých produktů, například: Tableau Desktop, Tableau Server, Tableau Prep nebo Tableau Cloud, které společně nabízí komplexní řešení ukládání, zpracování, analýzy a vizualizace dat. Od roku 2019 je Tableau součástí portfolia společnosti Salesforce. Mezi silné stránky Tableau patří především vizuálně příjemná editace a transformace dat a převod dotazů v přirozeném jazyce na vizuály pomocí umělé inteligence. [27; 28]



OBR. 4.3 TABLEAU PREP – FLOW PANE PRO SNADNOU ÚPRAVU A TRANSFORMACI DAT. OBDOBA K POWER QUERY. PŘEVZATO Z [34]

### Mode

Mode patří k novějším nástrojům BI, vznikl teprve v roce 2013 a rychle přišel k oblíbě firem. V roce 2022 zaujímal Mode největší podíl na trhu BI [29]. Patří k velmi volným BI službám, kde si uživatel sám může vytvořit prakticky cokoli. Mode nabízí zabudovaný editor SQL dotazů, ve kterých lze postavit veškeré potřebné dotazy pro analýzy, které lze dále zpracovat v notebookech pomocí Python skriptů, pro které taktéž nabízí online editor. Výsledné reporty a vizualizace lze poté sdílet mnoha různými způsoby, včetně embedded<sup>6</sup> objektů. [30]

### Datapine

Datapine je další alternativou mezi BI nástroji. Nabízí nástroje pro napojení dat, nástroje pro vlastní analýzu dat a tvorbu dashboardů, to vše jako cloudovou službu (SaaS). Nabízí online SQL Editor, nástroje pro online vizualizaci dat a nástroj pro tvoření a editaci dashboardů. Nabízí také dedikovaný nástroj pro analýzu prodejů. Tvorba vizuálů se provádí v editoru pomocí bloků, tažením myši (drag & drop). [19]

<sup>6</sup> Embedded je objekt typu iframe (většinou je to velmi dlouhý kód) který se vkládá na webové stránky do souborů HTML nebo v případě web builderů do polí k tomu určených.

# **PRAKTICKÁ ČÁST**

## 5 Definice problému

V této části definuji cíle praktické části práce a implementaci řešení pro sledování využívání zařízení. Tato zařízení spadají pod projekt, v této práci dále označený jako Projekt X. Praktická část v Power BI byla zpracována v anglickém jazyce, protože vlastník projektu, který bude s vytvořenými vizuály dále interagovat je mezinárodní společnost s hlavním komunikačním jazykem angličtinou. Kvůli anonymizaci klíčových pojmů zde vypisuji slovník klíčových pojmů, které budu v následující části používat a co znamenají, včetně jejich anglického překladu, protože některé pojmy jsou použity v obou jazycích v jiných souvislostech.

**Zařízení (*Device*):** Zařízení složené z ovladače a kontrolní jednotky, umístěné u klienta, generující záznamy o aktivitách.

**Vlastník:** Firma vlastnící, spravující zařízení.

**Klient (*Client*):** Ten, kdo si zařízení pronajímá. Zařízení je během pronajímaného období standardně umístěno u klienta.

**Uživatel:** Podle kontextu je uživatel zařízení interagující s ovládací jednotkou nebo uživatel Power BI výstupů.

**Použití zařízení (*Usage, Event*):** Na sebe navazující soubor aktivit v různých intervalech ne delších než 6 hodin. V datovém modelu a vizuálech je použití zařízení označováno jako Usage nebo Events.

**Událost (*Header*):** Záznam vznikající spuštěním zařízení a končící jeho vypnutím. V datovém modelu je reprezentována tabulkou Headers.

**Aktivita (*Action*):** Záznam konkrétní aktivity během události.

**Režim/mód (*Mode*):** Druh aktivity. Zařízení disponuje několika různými režimy (módy), v kterých může být používáno.

**Preprocesor:** Jedná se o externí aplikaci, která má na starost ETL proces, během kterého získá data vygenerovaná zařízením a zpracuje do podoby, ve které je poté nahraje do datového úložiště.

### 5.1 Cíle praktické části

Primárním cílem této datové analýzy je zjistit, jak efektivně a spravedlivě fakturovat klientům používání pronajatých zařízení. Toho dosáhneme přípravou reportů, které budou vykreslovat již analyzovaná data pro klienty. Nabízí se dvě možnosti fakturace: fixní měsíční poplatek – dnes využívaný drtivou většinou poskytovatelů služeb, např. Netflix, Adobe Cloud, Microsoft Office 365 nebo poplatek per-use, který se zatím používá spíše v B2B sféře a používá ho například Microsoft Azure nebo moderní umělá inteligence DALL-E 2. Firma si přeje klienty fakturovat způsobem *per-use* a prvním úkolem této práce tedy je zjistit, kdo a jak moc zařízení používá. V další části poté analyzuji ostatní data dostupná v datasetu a navrhuji možné rozšíření využití BI. V druhém úkolu analyzuji kvalitu a efektivitu zpracovaných vstupních dat, které jsou generovány v zařízení a extrahovány preprocesorem; údržbu jednotlivých zařízení; využití jednotlivých režimů zařízení a zpracuji informace o klientech.

### 5.1.1 Použití zařízení

Před zpracováním dat je třeba porozumět problému. Musíme tedy vědět, kdo je klient, co přesně znamená „použití“ a v jakém časovém období se pohybujeme. Za použití zařízení se počítá různá sekvence různě dlouhých událostí zařízení, mezi kterými mohou být také časové rozestupy, kdy je zařízení vypnuto, ne však déle než 6 hodin. Sekvence s rozestupem delším jak 6 hodin, s celkovým součtem času používání kratším než 20 minut do výsledku nebudou započítány. Tato krátkodobá použití mohou být způsobena různými faktory od kontroly či údržby po zvědavého uživatele.

## 6 Implementace řešení

### 6.1 Získání dat

Data vznikají v zařízeních, která je průběžně ukládají do textových souborů. Tyto soubory jsou poté periodicky načítány a zpracovávány preprocesorem, který je očišťuje a převádí do takové podoby, aby data v nich obsažená mohla být uložena do databáze. Tato data z databáze jsou poté konektorem připojena do Power BI, kde jsou transformována pomocí Power Query tak, abychom s nimi mohli dále pracovat.

První část dat jsou data interní, tedy data, která jsou generována zařízením. Jedná se o technická data vypovídající o provozu, používání, chybách atd. Druhá polovina dat jsou data externí, tedy data doplněná zaměstnanci, stažená z veřejně dostupných zdrojů nebo vygenerovaná aplikacemi. Tato data rozšiřují kontext k danému zařízení, například o které zařízení se jedná, kde se nachází nebo kdy bylo uvedeno do provozu.



OBR. 6.1 PROCES ZÍSKÁVÁNÍ A ZPRACOVÁNÍ DAT FIREMNÍCH DAT

Firemní data jsou uschována v databázi PostgreSQL na platformě Azure. Jsou však přístupná pouze z interní firemní sítě. Pro připojení do databáze je třeba kromě nastavení konektoru v Power BI nastavit také VPN na počítači.

Během připojování k databázi je třeba vybrat režim připojení. První možností je *Import*, která načte data z databáze a uloží je do vyrovnávací paměti, takže další práce s daty není závislá na kvalitě připojení a vytíženosti. Velikost importovaného datasetu je omezena na 1 GB [31]. Tato možnost je pro tyto účely dostačující, a proto jsme zvolili tuto. Alternativou k *Import* režimu je *Direct Query*.

Za účelem zpracování a anonymizace práce byla vytvořena umělá data, která jsou přidána do datového modelu ke skutečným datům pro dostatečnou demonstraci řešení. Tato data pomáhají lépe vizualizovat některé části řešení a zároveň částečně zkreslují výsledky analýzy dat. Mnou vygenerovaná data jsou připojena k datovému modelu pomocí souborů Excel, mají přidanou vlastní značku uměle vytvořeného záznamu a z reportů je tedy lze velmi snadno vyfiltrovat nebo dokonce úplně odebrat. Účelem těchto dat je kromě anonymizace skutečného stavu dat také lepší demonstrace možností analýzy těchto dat.

Id	ActiveControlUnit	RemoteControlUnit	From
b145fb09-32c9-4af9-99f9-9acd6c716966	VIB0009	\RCU_024-1520	6/28/2021 3:05:00 AM
13e975d2-d8d8-48a6-b356-9c2ea9eae700	VIB0024	\RCU_022-1520	10/25/2021 12:00:00 AM
8b21db01-8cb9-40d7-a759-c181420a1a76	VIB0012	\RCU_011-1520	10/25/2021 12:00:00 AM
c601655f-5526-4031-8958-4112f6a6dcf0	VIB0029	\RCU_010-1520	10/25/2021 12:00:00 AM

OBR. 6.2 UKÁZKA DAT NAČTENÝCH Z DATABÁZE. NA OBRÁZKU JSOU ÚDAJE IDENTIFIKUJÍCÍ KONKRÉTNÍ ZAŘÍZENÍ Z TABULKY DEVICES

### 6.1.1 Popis dat

Pro lepší představu popíšu ve stručnosti jednotlivé tabulky načtené do datového modelu. ER diagram datového modelu lze vidět na obr. 6.4.

#### *Clients*

Tabulka klientů obsahující údaje o klientech. V této tabulce najdeme například informace o tom, z které země a města klient pochází, jeho označení, úřední jazyky a datum, kdy byla s klientem podepsaná smlouva.

#### *Devices*

Tabulka zařízení je složena ze záznamů různých kombinací Klientů, RCU a ACU a časového údaje, od kdy je tato kombinace platná. Tento záznam je platný, dokud není vložený časový údaj o jeho terminaci. Kromě těchto informací je u každého zařízení veden záznam o tom kdy a kdo zařízení vytvořil (přidal do databáze), případně kdy a kým došlo k nějaké změně. V této tabulce je také abstraktní ID, kterým lze seskupovat jednotlivé variace zařízení, pokud vyměníme jednu z komponent, vytvoří se nový záznam, ale de facto se jedná o to samé zařízení.

#### *Headers*

Tabulka událostí. Sleduje informace o zařízení. Jsou zde informace o konkrétních komponentách, časové razítko, verze softwaru, ve kterém záznam vznikl a název souboru, z kterého událost pochází. Protože zařízení nemají povědomí o tom, jak jsou vedena v databázi (neví, že existuje tabulka *Devices*), ukládají informace pouze o sobě a informace z této tabulky je třeba ručně spojit se záznamy v tabulce *Devices*.

#### *Events*

Tabulka *Events* je vyextrahovaná z tabulky *Headers* a slouží k dopočítání informací o používání zařízení. Více informací o této tabulce píše v kapitole 6.1.5 Úpravy dat pomocí DAX.

#### *Actions*

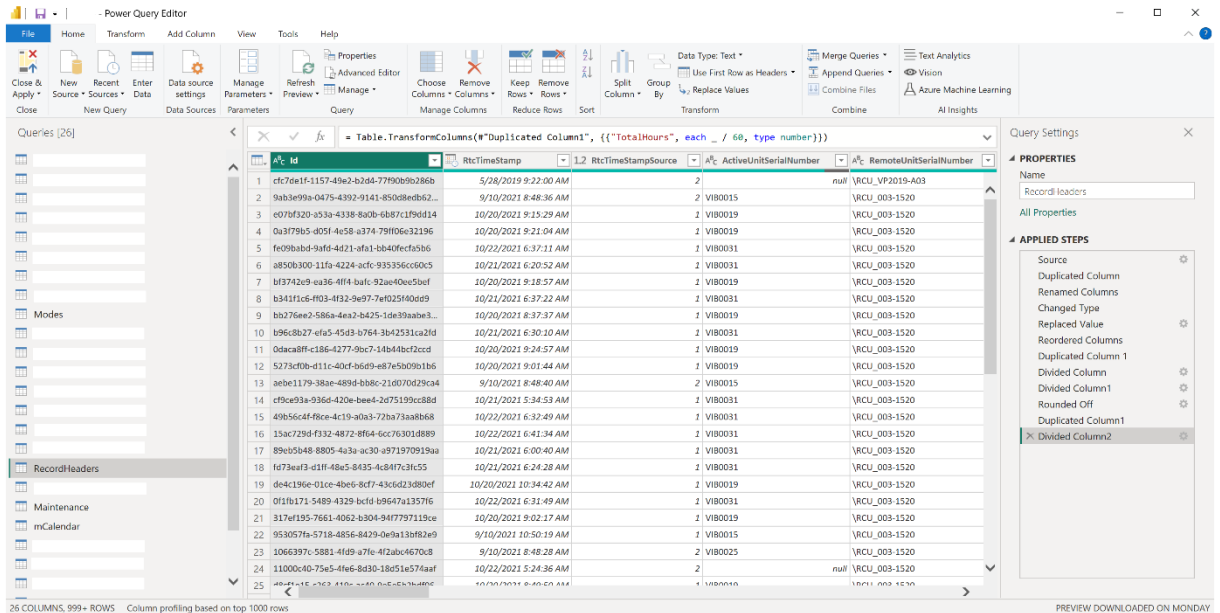
Tato tabulka obsahuje záznamy o jednotlivých činnostech zařízení. Obsahují informace od kdy do kdy, co, dodatečné parametry a *Header ID*, pomocí kterého lze záznamy přiřadit jednotlivým událostem.

### 6.1.2 Tvorba datového modelu

Načítání dat do Power BI se dělá pomocí datových konektorů. Nám postačí dva: jeden databázový a jeden pro sešit aplikace Excel. Načtená data lze dále upravovat, spojovat, filtrovat atd. v editoru Power Query.

Po načtení dat ze zdroje je potřeba zkontrolovat strukturu dat především shodující se názvy sloupců a datové typy. Příjemná výhoda je, že pro spojení dvou tabulek není třeba mít sloupce seřazené ve stejném pořadí. To nám dovolí spojit stejná data (tabulky) z různých zdrojů, tedy připojit naše data k informacím z databáze.

V dalším kroku odstraníme z modelu informace, o kterých můžeme bezpečně říct, že je nebudeme potřebovat k další analýze. Jedná se především o záznamy, kde chybí důležité informace, například cizí klíče nebo časové údaje. U záznamů, které neobsahují veškeré informace, ale chybí jim pouze takové informace, které můžeme dopočítat, provedeme výpočty a tyto chybějící informace do modelu doplníme. V této fázi může probíhat tzv. denormalizace dat<sup>7</sup>. Je důležité upravit data do takové podoby, aby se s nimi v následné analýze dobře pracovalo.



### OBR. 6.3 POWER QUERY JE BUILT-IN EDITOR PRO TRANSFORMACI DAT

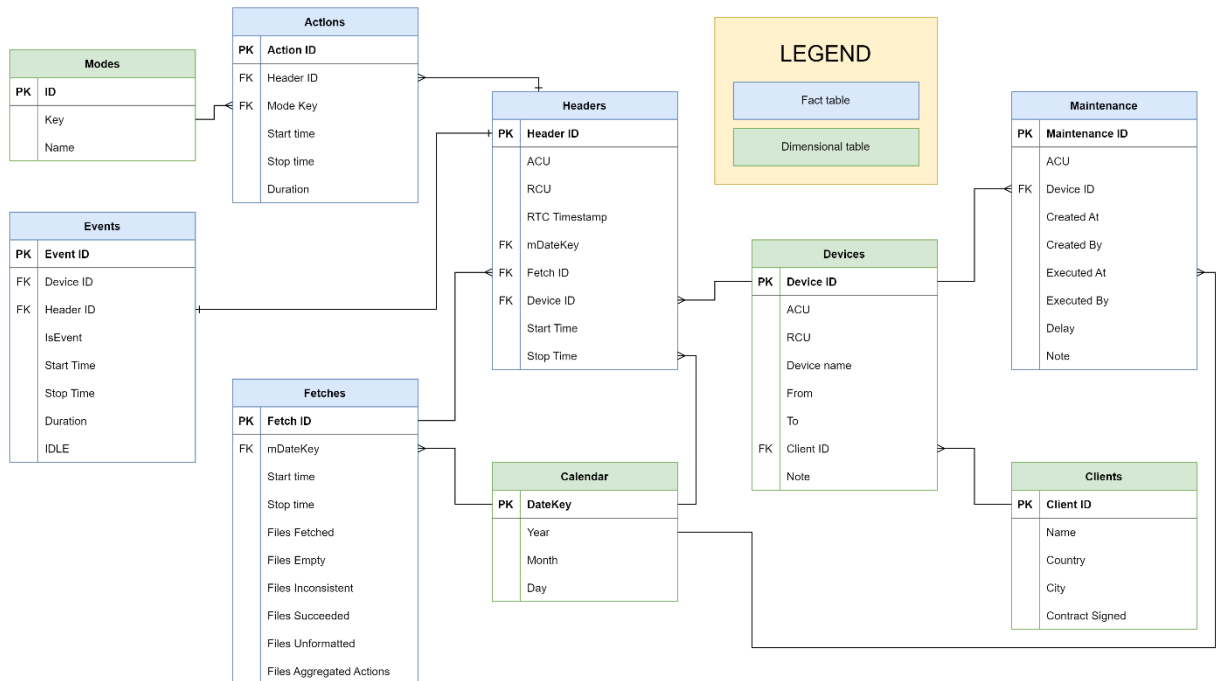
Tabulka *Devices* reprezentuje fyzické zařízení pronajímané klientům. Každé zařízení je složeno z dvou komponent: *Active unit controller* (dále ACU) a *Remote unit controller* (dále RCU). ACU je řídicí jednotka motorů, spíná motory, reguluje rychlost a směr pohybu. RCU je ovladač, pomocí kterého může uživatel zařízení ovládat.

<sup>7</sup> Denormalizace dat velmi jednoduše znamená, že přidáváme informace, vytváříme redundantní data. Opakem je normalizace dat, která je fundamentální při návrhu databázových systémů, jejíž cílem je minimalizace nároků na paměť pomocí eliminace redundantních informací.



### 6.1.3 Relace

Nově vytvořený model bylo třeba propojit relacemi, aby Power BI vědělo, jak spolu data souvisí. K tomu jsme využili nativní prostředí Power BI. Na obrázku obr. 6.4 je základní relační schéma<sup>8</sup> tabulek načtených z databázového serveru.



OBR. 6.4 ZJEDNODUŠENÝ ER DIAGRAM OBSAHUJÍCÍ ENTITY, PRIMÁRNÍ A CIZÍ KLÍČE A DALŠÍ DŮLEŽITÉ ATRIBUTY. DIAGRAM DÁLE ZOBRAZUJE RELACE MEZI ENTITAMI. TABULKY JSOU BAREVNĚ ROZLIŠENY NA DIMENZIONÁLNÍ A FAKTOVÉ.

Ze struktury tabulek lze vypočítat, že data jsou rozdělena do dvou částí, které spolu nelze propojit. Jedná se o záznamy ze zařízení, tj. provozní záznamy a o záznamy o zařízeních, tj. reprezentaci fyzických zařízení u klientů. Zařízení samo o sobě neví, které konkrétní zařízení je. Je však složeno z několika komponent, které ví, kdo jsou. Pomocí kombinace těchto komponent a časových záznamů lze dopočítat, které záznamy odpovídají jakému zařízení, a data tím propojit.

Druhá chybějící vazba je mezi tabulkou údržby a tabulkou zařízení. Protože sledované zařízení pro údržbu je pouze jedna komponenta celku, data z údržby jsou propojena s tabulkou zařízení nejednoznačným identifikátorem vytvářející vazbu M:N. Stejně jako u předchozího případu je třeba pomocí dostupných údajů identifikovat, ke kterému zařízení záznam patří. Tato vazba nám nevyhovuje, proto ji později upravím pomocí jazyka DAX.

### 6.1.4 Kalendář

Microsoft Power BI nabízí balík funkcí Time Intelligence, které, pokud jsou zapnuty, automaticky vytváří na pozadí datové tabulky pro všechny sloupce typu datum nebo datum a čas. [32] Alternativou k Time Intelligence je vytvoření vlastní datové tabulky. Tuto alternativu aplikujeme v implementaci této práce.

<sup>8</sup> ER diagram je nakreslený v aplikaci Diagrams.net, protože skutečný diagram vygenerovaný v Power BI je špatně čitelný, obsahuje nedůležité informace, kritické pojmy a je monochromatický. Zjednodušený znamená, že neobsahuje atributy, které jsou prázdné nebo je nemá smysl v rámci této práce uvádět.

Pro lepší analýzu dat, které obsahují datумы je potřeba si v Power BI hotovou tabulku, která obsahuje data od začátku do konce, respektive se snažíme najít nejnižší a nejvyšší hodnotu v datech, které budeme dále zpracovávat. Tuto tabulku vytváříme na úrovni Power Query, tedy ještě předtím, než načteme data do datového modelu. Začneme tím, že si najdeme první a poslední datum, které chceme v naší datové tabulce mít. Poté si vygenerujeme seznam jednotlivých dní od prvního do posledního dne a ten si převedeme na tabulku. Toto datum slouží jako klíč. Do tabulky dál můžeme doplnit jednotlivé sloupce konkrétních hodnot, které nás zajímají: roky, měsíce, dny či názvy měsíců a dnů. Pokud bychom potřebovali, můžeme také přidat čas, například hodiny. Je však dobré vědět, že s každým dalším dělením na menší část kalendáře vytváříme mnohem víc záznamů. Pokud bychom tedy chtěli mít kalendář po hodinách, znamená to, že pro každý jeden den budeme mít v tabulce 24 záznamů.

Abychom nově vytvořenou tabulku mohli použít jako hlavní kalendářovou tabulku, je potřeba v nastavení Power BI vypnout funkci inteligentního času, která pro každou tabulku, která obsahuje sloupec typu datum, implicitně vygeneruje vlastní kalendářovou tabulku. Toto řešení je sice rychlé, ale zabírá mnohem více paměti a nedovoluje nám vytvářet vlastní formáty dat.

### 6.1.5 Úpravy dat pomocí DAX

Protože datový model nebyl kompletní a ve stavu, ve kterém s ním můžu dále pracovat, bylo třeba dopočítat některé hodnoty až na úrovni DAX. V tabulkách, kde existují časové značky start a stop jsem pomocí DAX dopočítal sloupce trvání (duration) v sekundách, minutách a hodinách. V rámci vizuálů jsou vždy zobrazeny časy v minutách, není-li u grafu uvedeno jinak.

V tabulce *Headers* bylo potřeba dopočítat `Device ID`, protože zařízení svoje ID nezná. Z kombinace informací `ACU`, `RCU` a `RTC Timestamp` bylo třeba dohledat odpovídající záznam v tabulce *Devices* a záznamy spolu propojit. Obdobným způsobem bylo potřeba postupovat také u tabulky *Maintenance*.

V dalším kroku jsem pomocí funkce `SELECTCOLUMNS` vytvořil kopii tabulky *Headers*, kterou jsem pojmenoval *Events*. Do této tabulky jsem přesunul čtyři informace:

- Header ID
- Device ID
- Start Time
- Stop Time

V dalším kroku jsem dopočítal čas, po který zařízení bylo zapnuto. Tuto hodnotu jsem uložil do sloupců `TimeOnSeconds`, respektive `TimeOnMinutes` pro čas v sekundách a v minutách. Pro každé `Device ID` bylo potřeba v každém řádku dopočítat, kdy skončil předchozí záznam. Tato informace nám pomáhá zjistit rozestup jednotlivých zapnutí/vypnutí. Časový rozestup mezi těmito událostmi jsem označil do sloupce `IDLE`.

Když byl spočítaný časový rozdíl mezi jednotlivými sekvencemi, bylo potřeba každý záznam očíslovat. Číslování se provádí pro každé `Device ID` zvlášť, chronologicky. Záznam dostal přidělenou hodnotu, která se inkrementuje pokaždé, když je na řádku ve sloupci `IDLE` hodnota větší než 6 (hodin). Tato přidělená hodnota seskupuje jednotlivé záznamy jednoho použití. Tedy, když sečteme unikátní hodnotu pro každé `Device ID`, získáme počet použití na daném zařízení. Toto číslo označíme jako `Rank`.

Toto číslo stále není konečné, protože musíme ještě odstranit aktivity, které nespĺňují minimální časovou hranici pro používání. Proto je třeba dopočítat celkový čas spolu souvisejících událostí. Ty mohou být způsobeny například testováním zařízení nebo jeho údržbou a za použití se tedy nepočítají. V dalším kroku chceme detekovat aktivity, jejichž celková délka je kratší nebo rovna 20 minutám. Toho jsem dosáhl přidáním booleovské proměnné `IsEvent` s hodnotami pravda/nepravda.

Pro správné počítání použití je potřeba vytvořit nový klíč, který jsem vygeneroval kombinací polí Device ID a Rank, kterou jsem pojmenoval Event ID. Pro tuto hodnotu jsem vytvořil explicitní míru, která poté bude počítat unikátní hodnoty tohoto klíče podle vybraných dimenzí/filtrů a současně pouze ty, které splňují, že hodnota ve sloupci IsEvent je pravda.

V rámci výpočtů DAX jsem dále potřeboval vytvořit další explicitní míry, které se vkládají do vizuálů. Mezi ty důležité míry patří:

- Total Usage: počet *použití zařízení*.
- Total False Usage: *počet použití* zařízení s hodnotou nepravda ve sloupci IsEvent.
- Total Usage LM: počet použití v minulém měsíci.
- Total Usage LY: počet použití v minulém roce.
- Total Usage MoM: o kolik se změnila hodnota *Total Usage* od minulého měsíce.
- Running total: kumulativní součet použití od začátku do aktuálního data.
- Average Usage Duration: průměrná doba trvání jednoho *použití zařízení*.

V Power BI rozlišujeme dva typy měř (measures): implicitní a explicitní. Implicitní míry vytváří Power BI automaticky na základě datového typu, u čísel jsou to sumy. Pokud chceme složitější míru nebo jiný typ míry, můžeme si vytvořit vlastní – explicitní – míru, v rámci které můžeme dále upřesnit výpočet dané míry. Typicky můžeme upravit či změnit vstupní data nebo aplikovat různé filtry.

## 6.2 Vizualizace a analýza dat

V předchozí kapitole jsme analyzovali datový model, dopočítali jsme potřebné hodnoty, které jsme potřebovali pro zjištění použití zařízení. Nyní chceme tyto zjištěné hodnoty vizualizovat pomocí tabulek a grafů a předat managementu takovým způsobem, aby pro ně bylo snadné data číst a filtrovat a pomocí nich dělat další rozhodnutí.

### 6.2.1 Hlavní stránka

Aby uživatel věděl, kde začít, a měl důležité informace ihned po ruce, chceme vytvořit hlavní stránku. Ta slouží kromě rychlého přehledu také k navigaci napříč sestavami.



OBR. 6.5 HLAVNÍ STRÁNKA REPORTU

Na hlavní stránce (obr. 6.5) je v levém sloupci umístěna navigace napříč celou sestavou. V pravé části shora jsou údaje o *klientech*, počtu *zařízení*, zaznamenaných *událostí* a detekovaných *použití*. Dále je KPI pro počet použití s cílem nastaveným na předchozí rok. Z obrázku lze vyčíst, že v minulém roce 2021 bylo zaevidováno 98 použití zařízení, zatímco v roce 2022 jich bylo 339<sup>9</sup>. Poslední zeleno-růžový graf ukazuje do jaké míry jsou události počítány jako použití zařízení. Z obrázku vidíme, že 67,4 % všech záznamů bylo registrováno jako *použití* a zbylých 32,6 % představuje neregistrované používání. Cílem vlastníka by měla být minimalizace těchto neregistrovaných událostí, což povede k efektivnímu využívání zařízení a v neposlední řadě také maximalizaci zisku.

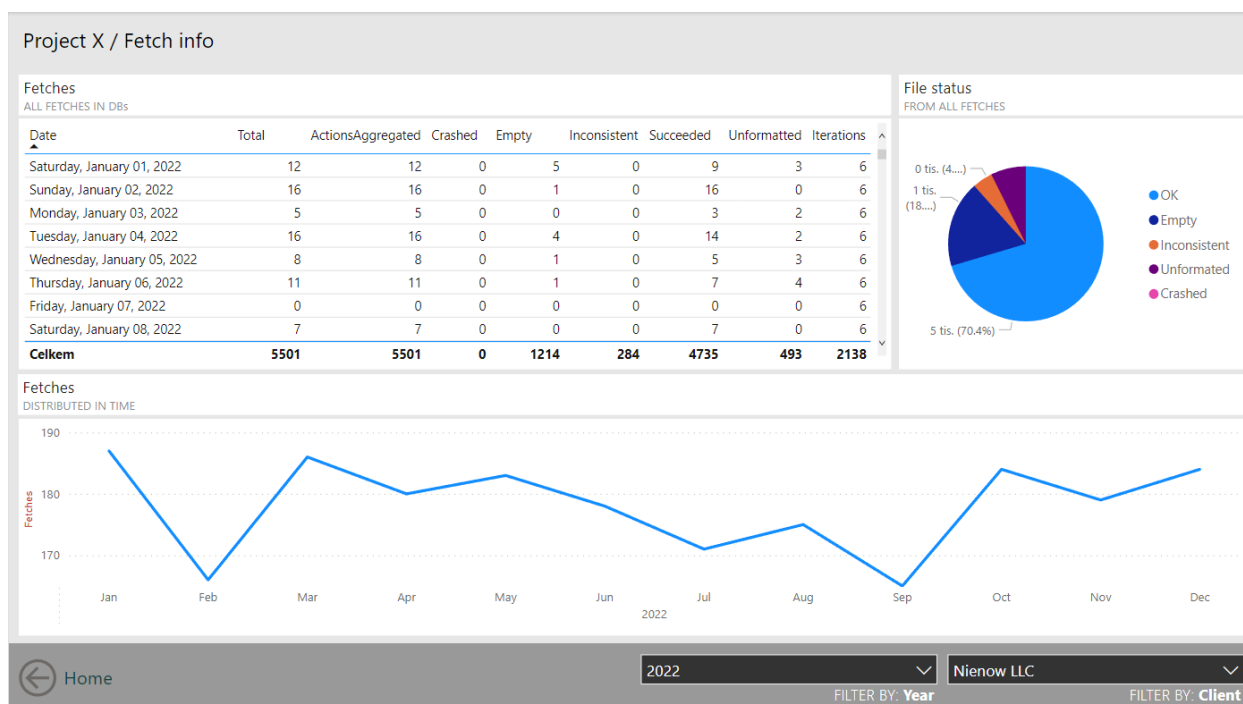
Navigace na hlavní stránce nabízí pět dalších sestav, kterými jsou:

- matice použití,
- informace o zpracování dat,
- režimy a jejich využívání,
- údržba,
- mapa klientů.

Na následujících stránkách popisují jednotlivé sestavy.

## 6.2.2 Informace o zpracovaných datech

Následující sestava zobrazuje souhrnné informace o tom, jak byla zpracována podkladová data. Pro potřeby další analýzy ukládá datový procesor, který má za úkol vyextrahovat surová data z textových záznamů, dodatečné informace o své činnosti do databáze. Mezi tyto informace patří například počet zpracovaných souborů, počet chyb a zpracované záznamy o chybách a dalších hlášeních systému. Na stránce je souhrnná tabulka zobrazující počet zpracovaných souborů pro každou



OBR. 6.6 SESTAVA O VÝSLEDKU A KVALITĚ ZPRACOVÁNÍ PŮVODNÍCH DAT

<sup>9</sup> Rozdíl mezi hodnotou KPI 339 a hodnotou v předchozím vizuálu 329 je ten, že v předchozím vizuálu je navíc zapnutý filtr, který nepočítá zařízení, která byla zrušena. To mohlo vzniknout například přesunutím zařízení nebo výměnou jedné z komponent RCU/ACU. Takovouto úpravou původní zařízení zaniká a vzniká záznam o novém zařízení.

iteraci zpracování dat. Vpravo se nachází graf zobrazující podíl různých výstupů zpracování. Ve spodní části se nachází počet volaných zpracování na časové ose.

V roce 2021 proběhly první vlny zpracování souborů a převod informací do databáze. Nejvíce v jeden den proběhlo 25 iterací, během kterých bylo zpracováno 10 137 souborů, celkem pak bylo v roce 2021 zpracováno 13 258 souborů v průběhu 175 iterací. Vyjma inicializační fáze během prvních 3 dní probíhá extrakce dat přibližně 6x denně, a to i v následujících letech. Úspěšně bylo zpracováno 11,8 tisíc souborů (82 % z celkového počtu), 8 % bylo prázdných a 8 % neobsahovalo správně formátovaný obsah souboru. 191 souborů z bylo neúplných (přibližně 2 %).

Oproti tomu v roce 2022 bylo zpracováno 5 501 souborů napříč 2 138 iteracemi, z toho 70 % úspěšně, dalších 18 % bylo prázdných.

I když průměrně vychází počet iterací na 178 za měsíc (tj. cca 6 denně), na spodním grafu můžeme vidět, že počet iterací v jednotlivých měsících kolísá. Například v únoru a přes léto proběhlo méně (166) iterací. Naopak v lednu, od března do května a v říjnu jich bylo více.

## 6.2.3 Matice použití

Jedná se o souhrnný report, složený z několika filtrů, navigačních tlačítek a maticové tabulky zobrazující počty použití konkrétního zařízení podle měsíce ve vybraném roce. Pokud není vybraný rok, ukazuje tabulka součet všech použití v konkrétním měsíci pro všechny roky v záznamech.

Project X / Usage matrix

Month Name	January		February		March		April		May		June		July		August		September		October		November		December		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
A033/R018 Demo Devices					1								1		3	200%			1						6	0%
▢ Dickinson LLC	2	100%	2	0%	5	150%	7	40%	13	86%	9	-31%	10	11%	12	20%	10	-17%	11	10%	11	0%	9	-18%	97	8%
A012/R011 Dickinson LLC	2	100%	2	0%	2	0%	4	100%	5	25%	4	-20%	4	0%	5	25%	4	-20%	10	150%	5	-50%	7	40%	52	11%
A017/R001 Dickinson LLC					3		1	-67%	3	200%	4	33%	2	-50%			2		1	-50%	5	400%	1	-80%	21	5%
A029/R010 Dickinson LLC							2		5	150%	1	-80%	4	300%	7	75%	4	-43%			1		1	0%	24	4%
▢ Ernestova společnost, s.r.o.					1																				1	0%
A904/R914 Ernestova společnost, s.r.o.					1																				1	0%
▢ Huel-Dooley			1		5	400%	3	-40%	3	0%	4	33%	3	-25%	6	100%	5	-17%	6	20%	1	-83%	4	300%	41	0%
A032/R007 Huel-Dooley			1		5	400%	3	-40%	3	0%	4	33%	3	-25%	6	100%	5	-17%	6	20%	1	-83%	4	300%	41	11%
▢ Nienow LLC	33	-20%	25	-24%	1...	-28%	1...	6%	21	11%	23	10%	33	43%	37	12%	41	11%	34	-17%	32	-6%	26	-19%	339	-5%
A004/R020 Nienow LLC													3		8	167%	8	0%	7	-13%	10	43%	3	-70%	39	8%
A004/R029 Nienow LLC	5	-38%	5	0%																					10	-44%
A010/R030 Nienow LLC	10	-9%											11		8	-27%	8	0%	11	38%	6	-45%	6	0%	59	-8%
A011/R031 Nienow LLC	9	29%	12	33%	8	-33%	8	0%	10	25%	10	0%	5	-50%	1	-80%	7	600%	7	0%	11	57%	6	-45%	93	-1%
A013/R013 Nienow LLC	5	-29%	3	-40%	5	67%	4	-20%	7	75%	6	-14%	5	-17%	9	80%	5	-44%					1		50	-11%
A015/R032 Nienow LLC	4	-50%	5	25%	5	0%	7	40%	4	-43%	7	75%	9	29%	11	22%	13	18%	9	-31%	5	-44%	10	100%	88	1%
<b>Total</b>	<b>35</b>	<b>-27%</b>	<b>28</b>	<b>-20%</b>	<b>31</b>	<b>11%</b>	<b>32</b>	<b>3%</b>	<b>37</b>	<b>16%</b>	<b>37</b>	<b>0%</b>	<b>59</b>	<b>59%</b>	<b>61</b>	<b>3%</b>	<b>56</b>	<b>-8%</b>	<b>52</b>	<b>-7%</b>	<b>44</b>	<b>-15%</b>	<b>42</b>	<b>-5%</b>	<b>507</b>	<b>-2%</b>

Navigation: Home, Detail, 2022, All, FILTER BY: Year, FILTER BY: Client

OBR. 6.7 SOUHRNNÝ REPORT S PŘEHLEDEM VYUŽÍVÁNÍ ZAŘÍZENÍ

Na obr. 6.7 se vpravo nahoře nachází název reportu, který uživatele informuje o tom, kde se právě nachází. Pod tímto záhlavím se pak nachází velká maticová tabulka, kde v řádcích jsou jednotlivá zařízení seskupená podle klientů a v sloupcích měsíce od ledna do prosince. Hodnoty tabulky jsou rozděleny na dva podsloupce **N** a **%**. První sloupec **N** označuje konkrétní počet *použití zařízení* v daném měsíci, zatímco druhý sloupec **%** označuje procentuální změnu přírůstku/poklesu oproti měsíci předchozímu. Pokud v předchozím měsíci je pole **N** prázdné, v aktuálním měsíci je i **%** pole prázdné. Tmavě šedá políčka označují měsíce, kde nebyla zaregistrována žádná událost (nebo přírůstek z minulého měsíce). Ve spodní části jsou další ovládací prvky vizuálů. Po pravé straně jsou připraveny

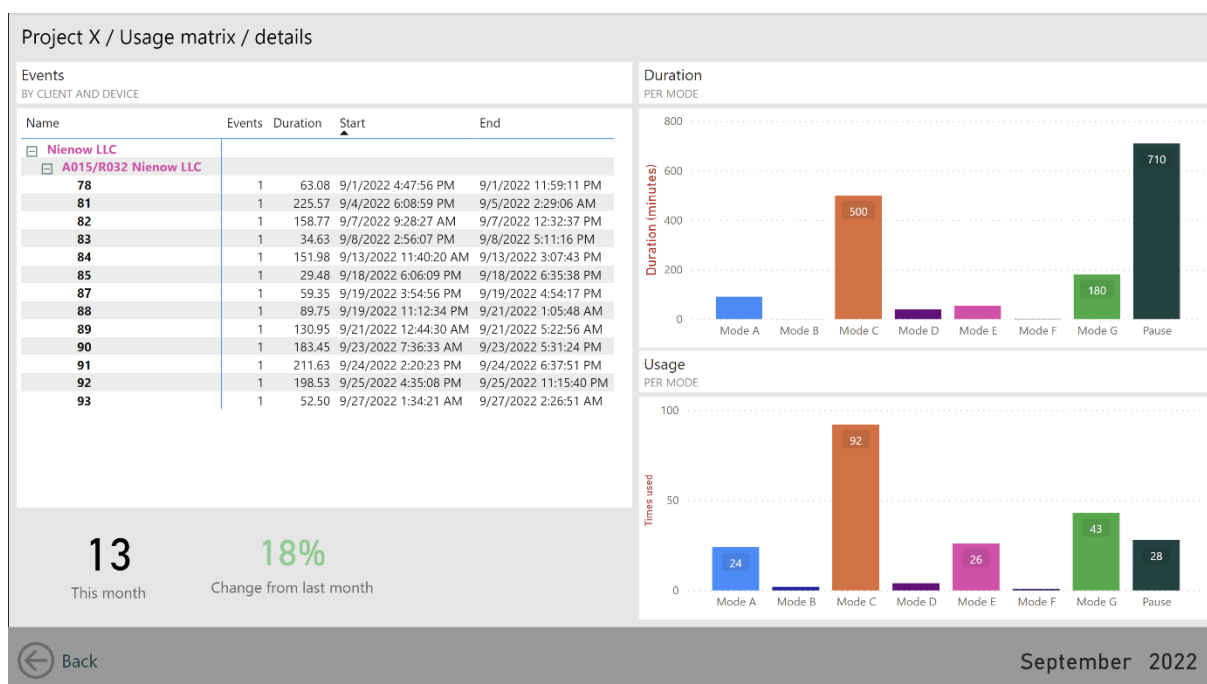
filtry, kde si uživatel může vybrat rok nebo roky, které chce filtrovat, a klienta, jehož zařízení chceme zobrazit. Vlevo se nachází tlačítko pro zobrazení detailu. Detail zobrazuje skladbu jednotlivých použití podle vybraných zařízení nebo podle měsíce. Pro aktivaci tlačítka je potřeba v tabulce vybrat konkrétní buňku nebo sloupec.

Z tabulky hned vidíme, že využívání jednotlivých zařízení se pohybuje v jednotkách až nižších desítkách použití za měsíc. Celkem bylo zaregistrováno 669 *použití*, od roku 2021 postupně 148, 507 a 14 v roce 2023, s průměrnou dobou použití 133, 126 a 110 minut. Uvažujme dobu použití 2 hodiny + 6 hodin rozestup mezi jednotlivými událostmi, teoretická kapacita zařízení je 3 hodiny denně, respektive 90 použití za měsíc. Máme-li 19 zařízení, 507 registrovaných použití a 12 měsíců, využitá kapacita jednoho zařízení vychází na **2,4 %**. Tento výpočet je ovšem velmi teoretický. Délka 1 použití může trvat různě dlouho – i několik hodin, stejně tak rozestupy mezi jednotlivými událostmi mohou být delší. Do výpočtu je také započítáno 5 testovacích zařízení, která nejsou reálně využívána klienty a jejich využití je minimální (zatímco jejich potenciál je stejný jako u klientů, tj. 90 použití za měsíc). Po odečtení těchto zařízení dostaneme výsledek **3,2 %**.

Přírůstky v jednotlivých měsících ukazují na relativní změnu oproti předchozímu období, tedy přírůstek o 1 může vykazovat jiné číslo napříč různými políčky. Vzhledem k malému počtu vzorků není tento údaj tolik vypovídající. Důležitější ovšem je barevné značení této změny. Tam kde je kladný přírůstek, je číslo zelené. Tam, kde je přírůstek záporný, je číslo růžové. Na první pohled tedy můžeme říct, jestli se situace od posledního měsíce zlepšila, či zhoršila. U většiny zařízení se trend střídá nahoru a dolů, což může být způsobeno nízkým využitím zařízení. Z dlouhodobého hlediska však využívání zařízení roste, jak možno vidět na obr. 6.5.

Doplňující informace k jednotlivým událostem lze zobrazit vybráním konkrétního pole (kombinace zařízení a měsíce) a stisknutím tlačítka Detail. Takto se přesuneme do detailního rozložení pro konkrétní měsíc a zařízení. (Lze vybrat také celý sloupec, tj. celý měsíc. Karta detailu poté zobrazí události na všech zařízeních ve vybraném měsíci.).

Stránka detailního zobrazení je z výchozího nastavení přepnuta do tzv. režimu drill-through. Tato karta se zobrazí po kliknutí na tlačítko Detail na kartě matice použití. Tento report zobrazuje podrobnější informace k souhrnným údajům z předchozího reportu. V horní části reportu jsou důležité



OBR. 6.8 DETAILNÍ POHLED UKAZUJE INFORMACE O JEDNOTLIVÝCH POUŽITÍCH V DANÉM MĚSÍCI

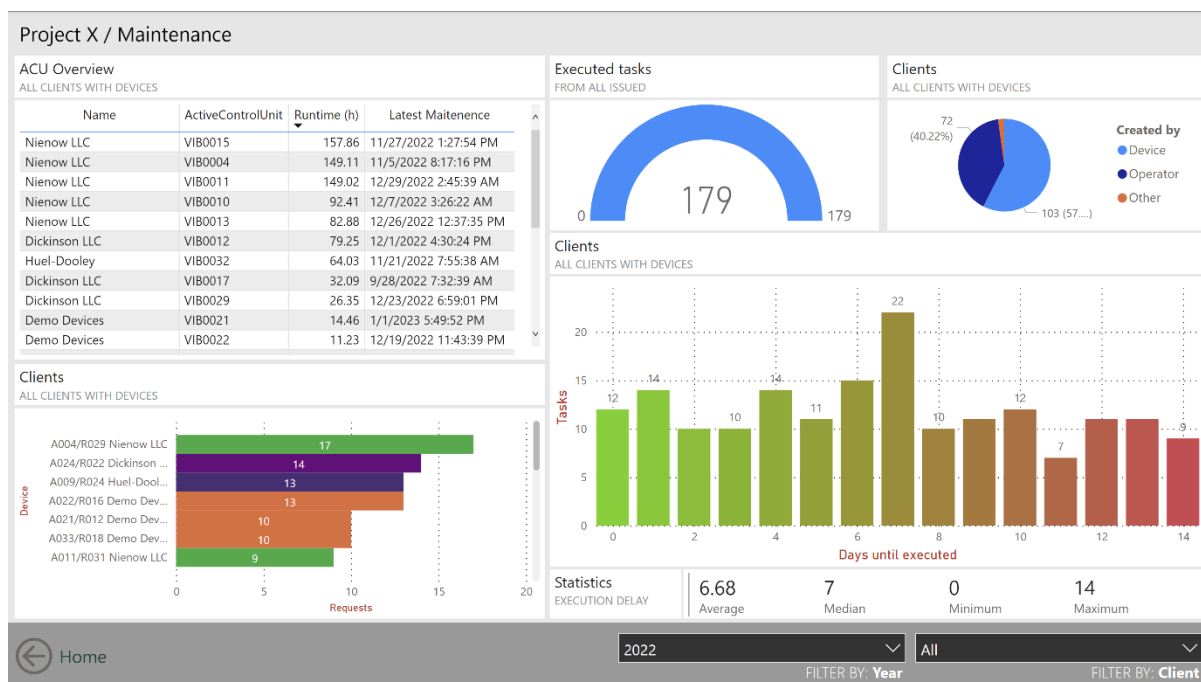
informace o tom, kde se uživatel nachází, pro jaké datum (měsíc, rok) je vizuál filtrován a součet jednotlivých použití v daném období. Dále zobrazuje pro každé zařízení konkrétní případy použití, přehled jednotlivých módů používaných během procesu a vizualizaci trvání a frekvence použití jednotlivých módů.

Na obr. 6.8 je vidět rozbor událostí ze září 2022. U každé události je uvedeno její identifikační číslo (pomocný identifikátor pro Power BI), doba trvání události v minutách a časy začátku a konce. Ve spodní části jsou umístěny hodnoty z tabulky, tedy celkový počet událostí a relativní změna z minulého měsíce. Na pravé straně se nachází dva grafy znázorňující využití jednotlivých módů během těchto použití. V horním grafu jsou hodnoty v minutách, ve spodním grafu je počet zapnutí konkrétního módu.

Z vyobrazených dat vidíme, že nejvíce časově využitými módy jsou mód C, mód G a mód A. Režim pauzy je velmi důležitý při práci se zařízeními, přesto při analýze režimů budu režimy pauzy a chyb a testovacích módů vypisovat zvlášť. Více informací o jednotlivých módech a jejich využití rozepisují v kapitole Režimy na straně 41.

## 6.2.4 Sledování údržby

Používáním se věci opotřebovávají, to je známá věc. Obzvláště pak ty složité a drahé věci chceme udržovat v co nejlepším stavu, aby fungovaly správně co nejdelší dobu. Z dostupných dat lze vytvořit přehledný report o údržbě jednotlivých zařízení. Na tomto reportu je přehledně vyobrazený stav údržby jednotlivých částí zařízení.



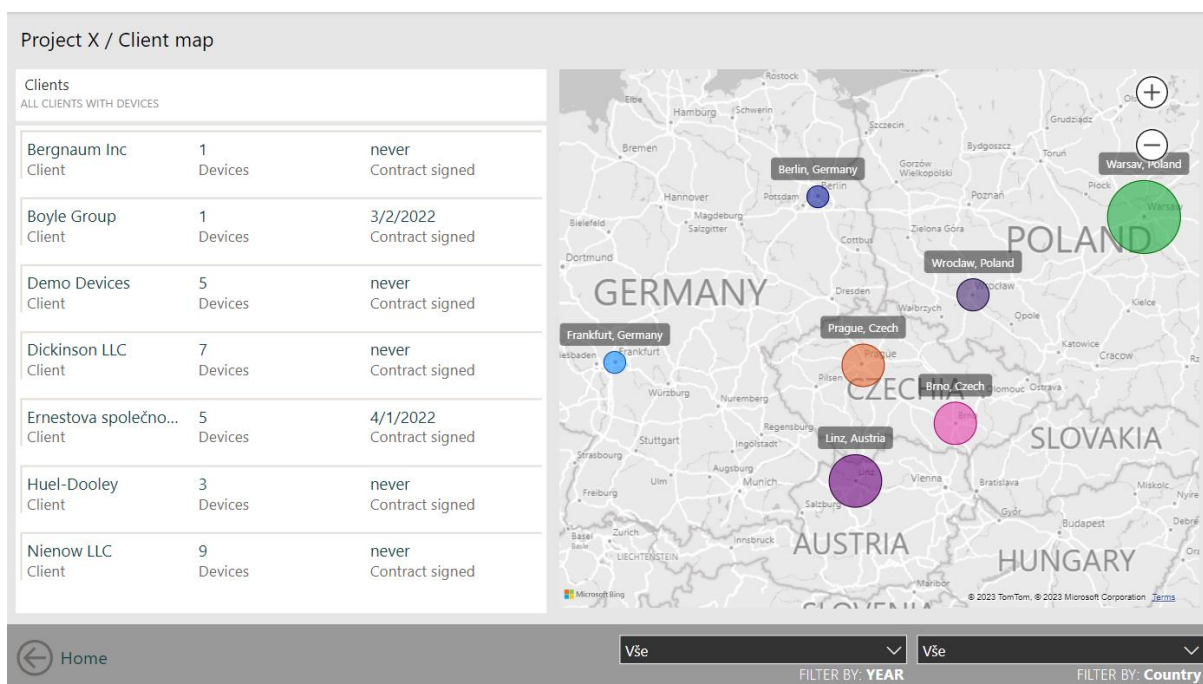
OB. 6.9 SESTAVA ÚDRŽBY ZAŘÍZENÍ

Na této sestavě se nachází tabulka klientů a komponent *Active Control Unit*, pro každou dvojici pak vypočítaný celkový čas běhu a poslední provedená údržba. Nahoře uprostřed se nachází půlkruhový vizuál měřidlo znázorňující, kolik oprav bylo již vykonáno z celkového počtu úkolů, které jsou zadány do systému. Vpravo se nachází graf původce žádosti o údržbu, resp. kým byla žádost vytvořena. Těmito původci jsou nejčastěji sama zařízení, která při překročení hlídání hranice času běhu motorů vygenerují žádost o kontrolu nebo požadavek vytvoří operátor na výzvu klienta či z jiného důvodu. Ve spodní části je graf zařízení (na vertikální ose) a počet vygenerovaných žádostí. Na posledním grafu vpravo dole vidíme, kolik žádostí bylo vyřešeno podle délky trvání, tedy podle

rozdílu vykonání údržby od jejího zadání do systému (ve dnech). Dole jsou čtyři doplňující údaje ke grafu zpoždění: průměrný počet dní k vykonání údržby, medián, nejrychlejší a nejdelší čas od založení do vykonání údržby.

V roce 2021 bylo založeno 34 žádostí, z nichž polovina byla vytvořena zařízením a druhá polovina operátorem. Za rok 2022 bylo zadáno do systému celkem 179 žádostí o údržby z nichž 103 bylo zadáno do systému automaticky, 72 žádostí bylo vytvořeno operátorem a čtyři žádosti byly do systému zadány jiným způsobem. 57 % z celkového počtu je tedy tvořena primárně žádostmi o kontrolu. V roce 2023 bylo vygenerováno 42 žádostí, z nichž 3 v době psaní tohoto textu ještě nejsou vyřízené. Z těchto žádostí je 55 % žádostí zadáno do systému operátorem, zbylých 45 % automaticky generuje zařízení.

Nejrychleji vyřízené žádosti, tj. žádosti vyřešené ve stejný den mají dobu trvání 0. To je minimální hodnota a z logiky věci nemůže být nižší, může ovšem být vyšší. Okamžitě vyřešených žádostí je v roce 2021 dvě, v roce 2022 dvanáct a v roce 2023 zatím čtyři. Průměrná doba provedení kontroly, respektive údržby je 6,68 dne, střední hodnota je rovna 7. Můžeme tedy říci, že minimálně polovina žádostí je vyřízena do jednoho týdne. Nejdelší doba na vyřízení žádosti byla v roce 14 dní. V předchozím roce byla průměrná doba 5,65 dne a medián 6 dní. I roce 2023 zůstává průměrná doba vyřízení na 6,36 dnech a střední hodnotě 6. Vzhledem k relativně malým vzorkům v letech 2021, kdy projekt začal, a roku 2023, kdy v době psaní této práce uběhly necelé 4 měsíce, by srovnání dat v letech mohlo vést k nekompletním zjištěním a mylným závěrům.



OBR. 6.10 VIZUÁL ZOBRAZUJÍCÍ GEOGRAFICKÉ ROZLOŽENÍ KLIENTELY SPOLEČNOSTI A POČET PRONAJÍMANÝCH ZAŘÍZENÍ

## 6.2.5 Mapa klientů a zařízení

Tento report nám nabízí rychlý přehled všech klientů a zařízení, která jsou u nich umístěna, a to jak vizuálně, tak deskriptivně. Lze tak sledovat pokrytí v jednotlivých státech, což je velmi důležité pro další expanzi produktu. Nalevo se nachází seznam klientů, počet zařízení, která mají k dispozici ve svých prostorách a datum, kdy byla podepsaná smlouva, pokud je o tom uveden záznam v databázi. Vpravo nahoře se nachází filtr, kterým lze vybrat klienty v konkrétních zemích.



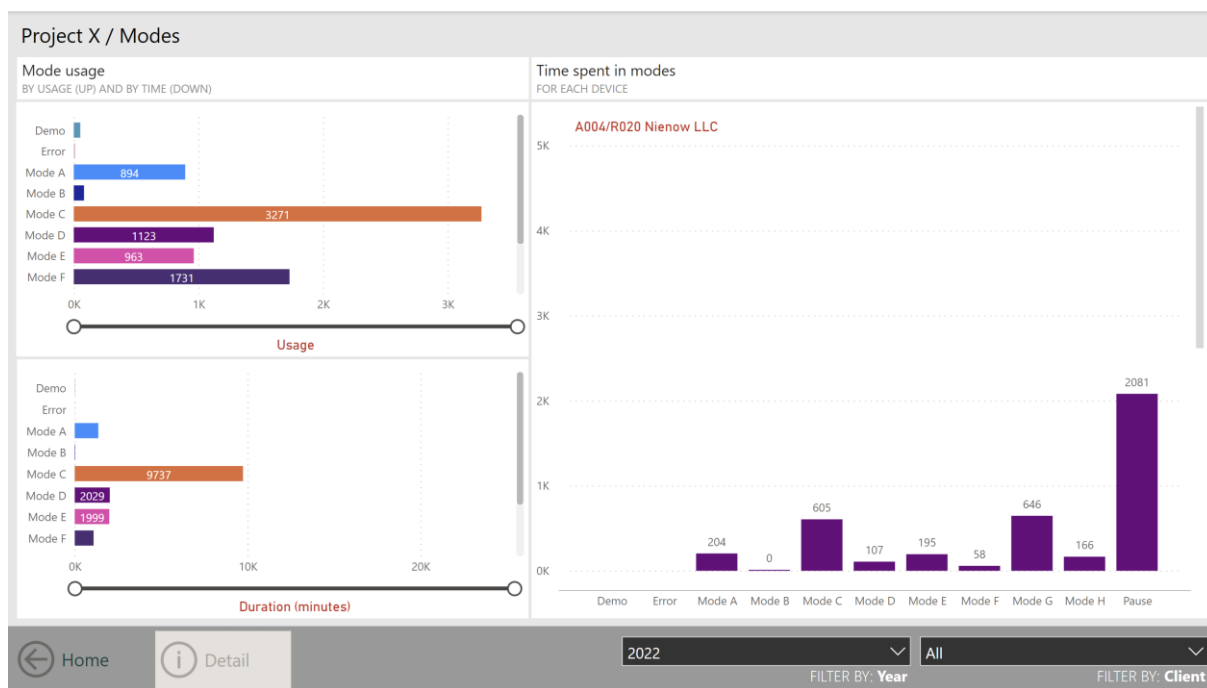
Z dostupných dat vidíme, že firma aktuálně spolupracuje se sedmi klienty ve čtyřech různých zemích, u kterých má umístěna svá zařízení, všechna jsou ve střední Evropě. Nejvíce zařízení se nachází ve společnosti Nienow LLC v Polsku, která disponuje 9 zařízeními. Dalších 10 zařízení se nachází na území České republiky, z čehož 5 zařízení je testovacích, 7 v Rakousku a 2 v Německu.

Celkem firma disponuje 31 zařízeními.

## 6.2.6 Režimy

Další z vlastností, které můžeme sledovat během používání zařízení, jsou režimy, v kterých zařízení pracuje. Zařízení nabízí několik různých režimů, mezi kterými může uživatel libovolně přepínat. Zařízení umí rozpoznat, pokud nastane chyba a přepne se do chybového režimu (error). Dalším důležitým režimem mimo ty standardní je režim pauza (pause), který je velmi důležitý pro bezpečné a efektivní používání zařízení. Poslední kategorie mimo standardní režimy je kategorie demo – ta zahrnuje veškeré umělé režimy, které se používají při simulaci používání bez skutečného používání zařízení.

Ve standardní nabídce režimů je celkem 8 režimů, v práci označených slovem „Mode“ a písmenem A–H. Mezi doplňující režimy patří pauza, error a demo.



OBR. 6.11 PŘEHLED VYUŽITÍ JEDNOTLIVÝCH REŽIMŮ

Na sestavě na obr. 6.11 se po levé straně nachází dvojice grafů využití jednotlivých módů, z nichž horní graf je v jednotkách použití (spuštění) konkrétního režimu a spodní v minutách, po kterou byl daný mód aktivní. Na pravé straně je skupina grafů pro jednotlivá zařízení<sup>10</sup> zobrazující, kolik času ve kterém režimu zařízení strávilo.

Pro lepší přehled o využití konkrétního režimu stačí daný režim označit v kterémkoli vizuálu a kliknout na tlačítko Detail, kterým zobrazíme podrobnou analýzu pro vybraný režim, která je vyobrazena na obr. 6.12. V horní části je opět graf znázorňující využití v čase a počtu spuštění,

<sup>10</sup> Jednotlivé grafy jsou umístěny pod sebou a lze mezi nimi procházet posouváním kolečka myši (nebo posuvníku), jako při procházení dokumentu nebo webové stránky.

tentokrát ale podle konkrétního zařízení. Ve spodní části se nachází graf používání v jednotlivých dnech. Na horizontální ose jsou jednotlivé dny v měsíci. Sloupce indikují celkový počet minut strávený v aktuálním režimu pro všechna zařízení.



OBR. 6.12 DETAILNÍ POHLED NA VYUŽITÍ KONKRÉTNÍHO REŽIMU

## Závěr

V první části práce jsem popsal různé významy pojmu business intelligence, včetně její historie, která se začíná přibližně v druhé polovině 20. století. Zmínil jsem důvody, proč a k čemu je BI vhodné využít, její teoretickou architekturu, která se skládá ze čtyř vrstev: datové, databázové, analytické a uživatelské. Prošel jsem také technologie, které se přímo týkají nebo úzce navazují na tematiku BI, mezi které patří například proces ETL, datové sklady a jezera, OLAP nebo data mining. Zároveň jsem v průběhu tohoto textu vysvětlil některé pojmy z databázových systémů, neboť ty jsou důležitou součástí základních vědomostí každého, kdo pracuje s daty (podobně jako každý matematik umí derivace i když s nimi nepočítá přímo).

V praktické části této práce jsem se zaměřil na kontrolu zpracovaných dat, detekci *použití zařízení*, způsob, jakým jsou zařízení během používání využívána a jejich údržbu. Jako další krok bych doporučil zaměřit se na druhou část dat obsažených v datovém modelu – tj. na data o použití, které nesplňují podmínky *použití zařízení*. V těchto datech se nachází cenné informace, a hlavně provozní čas, který vlastníka zařízení stojí peníze, protože jej klientům nelze fakturovat. Analýzou a postupnou minimalizací vzniku těchto nevyužitých časů by vlastník mohl dojít nižších nákladů na provoz a současně maximalizovat zisk. Z dostupných dat jsem také zjistil, že využitá kapacita zařízení rozhodně nedosahuje maxima, potenciál růstu je tedy stále velmi vysoký.

Sestavy popsané v praktické části jsou připraveny k publikaci do firemního intranetu firmy, kde mohou být dále z části nebo vcelku napojeny do již existujících dashboardů nebo aplikací *Power Apps*<sup>11</sup>, které mohou mít manažeři stále po ruce ve svých mobilních zařízeních.

Business Intelligence nabízí možnost porozumět shromážděným datům, vizualizovat jejich vývoj v čase i současný stav, porovnávat různé možnosti a s pomocí regresních algoritmů, strojového učení a umělé inteligence vytvářet prognózy, s jejíž pomocí mohou firmy usměrňovat svá současná rozhodnutí a dosahovat tak svých budoucích cílů. Velký Bratr by byl hrdý.

*„Kdo ovládá minulost, ovládá budoucnost. Kdo ovládá přítomnost, ovládá minulost.“*

– George Orwell, 1984

---

<sup>11</sup> Produkt z rodiny Microsoft 365 zaměřený na tvorbu toků.

## Seznam použité literatury

- [1] **Withee, Ken.** *Microsoft Business Intelligence for Dummies*. Indianapolis : Wiley Publishing, Inc., 2010. ISBN: 978-0-470-52693-4.
- [2] **IBM.** What is Business Intelligence? *IBM.com*. [Online] IBM. [Cited: February 28, 2023.] <https://www.ibm.com/topics/business-intelligence>.
- [3] **Microsoft.** What is business intelligence? *Microsoft.com*. [Online] Microsoft. [Cited: February 28, 2023.] <https://powerbi.microsoft.com/en-us/what-is-business-intelligence/>.
- [4] **Suchánek, Petr.** *Business Intelligence: Distanční studijní text*. Karviná : Obchodně podnikatelská fakulta v Karviné, 2018. ISBN 978-80-7510-308-6.
- [5] **Heinze, Justin.** History of Business Intelligence. *BetterBuys.com*. [Online] 5 27, 2020. [Cited: February 20, 2023.] <https://www.betterbuys.com/bi/history-of-business-intelligence/>.
- [6] **Kimbal, Ralph and Ross, Margy.** *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd edition. Indianapolis : Wiley, 2013. ISBN 978-1-118-53080-1.
- [7] **Sudan, Aditya.** What Is ETL And How the ETL process works? *datachannel.co*. [Online] DataChannel Technologies. [Cited: March 14, 2023.] <https://www.datachannel.co/blogs/what-is-etl-and-how-the-etl-process-works>.
- [8] **Microsoft.** Understand star schema and the importance for Power BI. *Microsoft.com*. [Online] Microsoft. [Cited: March 20, 2023.] <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema>.
- [9] **Microsoft.** Co je datový sklad? *azure.microsoft.com*. [Online] Microsoft. [Citace: 20. březen 2023.] <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-a-data-warehouse/#data-warehouse-definition>.
- [10] **OLAP.com.** What is OLAP? *OLAP.com*. [Online] Paris Technologies. [Cited: March 14, 2023.] <https://olap.com/olap-definition/>.
- [11] **Microsoft.** Přehled technologie OLAP (Online Analytical Processing). *Microsoft.com*. [Online] Microsoft. [Citace: 14. březen 2023.] <https://support.microsoft.com/cs-cz/office/p%2599ehled-technologie-olap-online-analytical-processing-15d2cdde-f70b-4277-b009-ed732b75fdd6>.
- [12] **IBM.** What is data mining? *IBM.com*. [Online] IBM. [Cited: February 25, 2023.] <https://www.ibm.com/topics/data-mining>.
- [13] **Microsoft.** Data mining concepts. *Microsoft.com*. [Online] Microsoft, December 10, 2022. [Cited: April 16, 2023.] <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>.

- [14] **Microsoft.** Data mining algorithms (Analysis Services - Data mining). *Microsoft.com*. [Online] Microsoft, December 10, 2022. [Cited: April 16, 2023.] <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>.
- [15] **Zwass, Vladimir.** Information system. *Britannica.com*. [Online] Encyclopedia Britannica, April 3, 2023. [Cited: April 19, 2023.] <https://www.britannica.com/topic/information-system>.
- [16] **Zwass, Vladimir.** Expert system. *britannica.com*. [Online] Encyclopedia Britannica, February 10, 2016. [Cited: April 19, 2023.] <https://www.britannica.com/technology/expert-system>.
- [17] **myTimi.** KPI ukazatele: Co to je a jak je vytvořit? *myTimi.cz*. [Online] myTimi.cz, 7. leden 2021. [Citace: 20. březem 2023.] <https://www.mytimi.cz/kpi-ukazatele-co-to-je-a-jak-je-vytvorit/>.
- [18] **KPI.org.** What is a Key Performance Indicator (KPI)? *kpi.org*. [Online] KPI.org, a Strategy Management Group company. [Cited: March 20, 2023.] <https://www.kpi.org/KPI-Basics/>.
- [19] **Datapine.** Management Dashboard Examples. *Datapine.com*. [Online] Datapine, 2023. [Cited: March 27, 2023.] <https://www.datapine.com/dashboard-examples-and-templates/management>.
- [20] **World Health Organization.** WHO COVID-19 Dashboard. *who.int*. [Online] World Health Organization, 2020. [Cited: 16 April, 2023.] <https://covid19.who.int/>.
- [21] **Datapine.** Learn about the benefits of self-service BI tools. *datapine.com*. [Online] Datapine. [Cited: February 27, 2023.] <https://www.datapine.com/articles/self-service-bi-tools>.
- [22] **Microsoft.** What is Power BI? *Microsoft.com*. [Online] Microsoft, January 19, 2023. [Cited: February 20, 2023.] <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>.
- [23] **Microsoft.** Basic concepts for designers in the Power BI service. *Microsoft.com*. [Online] Microsoft, February 7, 2023. [Cited: February 20, 2023.] <https://learn.microsoft.com/en-us/power-bi/fundamentals/service-basic-concepts>.
- [24] **Microsoft.** Microsoft Power Query. *Microsoft.com*. [Online] Microsoft.com, 2023. [Cited: March 28, 2023.] <https://powerquery.microsoft.com/en-us/>.
- [25] **Richardson, Ben.** Differences between the M Language and DAX in Power BI. *sqlshack.com*. [Online] SQLShack, May 14, 2020. [Cited: March 1, 2023.] <https://www.sqlshack.com/differences-between-the-m-language-and-dax-in-power-bi/>.
- [26] **Microsoft.** Power Query M formula language. *Microsoft.com*. [Online] Microsoft. [Cited: March 1, 2023.] <https://learn.microsoft.com/en-us/powerquery-m/>.
- [27] **Dambrosio, Laura.** Looker vs. Tableau: A Comparison. *Mode.com*. [Online] Mode, December 22, 2022. [Cited: February 26, 2023.] <https://mode.com/blog/looker-vs-tableau-vs-mode/>.

- [28] **Tableau.** What is Tableau. *Tableau.com*. [Online] Tableau. [Cited: February 26, 2023.] <https://www.tableau.com/why-tableau/what-is-tableau>.
- [29] **Datanyze.** Leading vendors' share of the business intelligence (BI) software market worldwide in 2022. *Statista.com*. [Online] Datanyze, September 13, 2022. [Cited: February 26, 2023.] <https://www.statista.com/statistics/590031/worldwide-business-analytics-software-vendor-market-share/>.
- [30] **Mode.** Decisions powered by data. *mode.com*. [Online] Mode. [Cited: March 23, 2023.] <https://mode.com/ad-hoc-analysis/>.
- [31] **Microsoft.** Data sources for the Power BI service. *Microsoft.com*. [Online] Microsoft, February 7, 2023. [Cited: February 27, 2023.] <https://learn.microsoft.com/en-us/power-bi/connect-data/service-get-data>.
- [32] **Rešl, Štěpán.** Time Intelligence. <https://www.jaknapowerbi.cz/time-intelligence/>. [Online] jaknapowerbi.cz. [Cited: 7. duben 2023.] <https://www.jaknapowerbi.cz/time-intelligence/>.
- [33] **Microsoft.** Describe different cloud services. *Microsoft.com*. [Online] Microsoft. [Cited: March 1, 2023.] <https://learn.microsoft.com/en-us/training/modules/fundamental-azure-concepts/categories-of-cloud-services>.
- [34] **Tableau.** About Tableau Prep. *Tableau.com*. [Online] Tableau. [Cited: February 26, 2023.] [https://help.tableau.com/current/prep/en-us/prep\\_about.htm#using-tableau-prep](https://help.tableau.com/current/prep/en-us/prep_about.htm#using-tableau-prep).
- [35] **Microsoft 365 Team.** Klíčové ukazatele výkonnosti (KPI): Co jsou a jak se používají? *Microsoft*. [Online] Microsoft, 9. říjen 2019. [Cited: 2. března 2023.] <https://www.microsoft.com/cs-cz/microsoft-365/business-insights-ideas/resources/what-are-kpis-and-how-to-use-them>.

## Seznam obrázků

obr. 2.1 Architektura Business Intelligence. Převzato z [4] .....	13
obr. 2.2 ETL je proces postupného získávání dat z různých zdrojů, jejich transformace uložení do datových úložišť .....	14
obr. 2.3 Srovnání různých schémat datových struktur. Star (vlevo) a Snowflake (vpravo) .....	15
obr. 2.4 Multidimenzionální kostka OLAP .....	17
obr. 3.1 Příklad finančního dashboardu. Dostupné z [19] .....	21
obr. 4.1 Srovnání IaaS, PaaS a SaaS. O červené bloky se stará poskytovatel služby, uživatel má přístup k zeleným blokům. Převzato z [33] .....	23
obr. 4.2 Power Query .....	25
obr. 4.3 Tableau Prep – flow pane pro snadnou úpravu a transformaci dat. obdoba k power query. Převzato z [34].....	26
obr. 6.1 Proces získávání a zpracování dat firemních dat .....	30
obr. 6.2 Ukázka dat načtených z databáze. Na obrázku jsou údaje identifikující konkrétní zařízení z tabulky Devices .....	30
obr. 6.3 Power Query je built-in editor pro transformaci dat.....	32
obr. 6.4 Zjednodušený ER diagram obsahující entity, primární a cizí klíče a další důležité atributy. Diagram dále zobrazuje relace mezi entitami. Tabulky jsou barevně rozlišeny na dimenzionální a faktové.....	33
obr. 6.5 Hlavní stránka reportu .....	35
obr. 6.6 Sestava o výsledku a kvalitě zpracování původních dat .....	36
obr. 6.7 Souhrnný report s přehledem využívání zařízení .....	37
obr. 6.8 Detailní pohled ukazuje informace o jednotlivých použitích v daném měsíci .....	38
obr. 6.9 Sestava údržby zařízení.....	39
obr. 6.10 Vizualizace zobrazující geografické rozložení Klientely Společnosti a počet pronajímaných zařízení .....	40
obr. 6.11 Přehled využití jednotlivých režimů.....	41
obr. 6.12 Detailní pohled na využití konkrétního režimu .....	42