

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo : 103004/I/2022/36114651175673348

Trendy v oblasti veledát (big data)

Diplomová práca

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

Trendy v oblasti veledát (big data)

Diplomová práca

Študijný program: Informačný manažment

Študijný odbor: Ekonómia a manažment

Školiace pracovisko: Katedra aplikovanej informatiky

Vedúci záverečnej práce: doc. Ing. Martin Mišút, CSc.

Čestné vyhlásenie

Čestne vyhlasujem, že som záverečnú prácu vypracoval samostatne za pomoci školiteľa a dostupnej domácej a zahraničnej literatúry, ktorú som uviedol.

Dátum:

.....

Pod'akovanie

Touto cestou, by som chcel vyjadriť pod'akovanie hlavne môjmu školiteľovi doc. Ing. Martin Mišút, CSc., za jeho čas, prístup a návrhy, ktorými prispel k zhotoveniu tejto práce. Rovnako tak patrí vďaka aj mojej priateľke, rodine a kamarátom, ktorí mi dokázali vytvoriť vhodné pracovné prostredie a poskytli podporu a cenné rady potrebné pre úspešne zhotovenie práce.

ABSTRAKT

NEMEC, Patrik: *Trendy v oblasti veledát (big data)*. – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky, Katedra aplikovanej informatiky. – Vedúci záverečnej práce: doc. Ing. Martin Mišút, CSc. – Bratislava: FHI, 2022, počet strán 65.

Diplomová práca je rozdelená do štyroch kapitol. Obsahuje 18 grafov a 4 obrázky. Cieľom tejto diplomovej práce je identifikácia podstatných trendov v oblasti big data a poskytnutie odporúčaní, ako dané trendy implementovať v rámci rôznych sektorov. Prvá kapitola popisuje kľúčové pojmy z tejto oblasti, výhody, ktoré nám big data dokážu priniesť, ale aj výzvy, ktorým musia spoločnosti čeliť. V rámci tejto kapitoly je tiež uvedených a charakterizovaných niekoľko technológií, ktoré dokážu spoločnostiam uľahčiť prácu s big data. Druhá kapitola je tvorená cieľom tejto práce a stanovením ďalších čiastkových cieľov, ktoré bolo potrebné spraviť, aby sme dosiahli primárny cieľ. V tretej kapitole sa venujeme metodike práce a metódam skúmania, ktoré boli použité. Posledná kapitola zobrazuje identifikáciu výziev, ktorým musia spoločnosti čeliť pri transformácii na spoločnosti riadené dátami a taktiež sú tu identifikované podstatné trendy z viacerých oblastí, ako sú technologická oblasť, legislatívna oblasť a oblasť ľudského kapitálu. V rámci tejto kapitoly je zhrnutý stav implementácie daných trendov na Slovensku a v zahraničí. Táto diplomová práca tiež poskytuje aj naše odporúčania, ktoré by mali spoločnostiam poslúžiť na prekonanie výziev, ktoré sú spojené so súčasnými trendami v danej oblasti.

Kľúčové slová: Big data, dátová analytika, spoločnosti riadené dátami, trendy v oblasti big data, digitálna transformácia

ABSTRACT

NEMEC, Patrik: *Trends in big data*. – University of Economics in Bratislava. Faculty of Economic Informatics, Department of applied informatics. – Supervisor of the final thesis: doc. Ing. Martin Mišút, CSc. – Bratislava: FHI, 2022, pages: 65.

The diploma thesis is divided into four chapters. It contains 19 graphs and 4 figures. This thesis aims to identify significant trends in big data and provide recommendations on implementing these trends in various sectors. The first chapter describes the key concepts in this field, the benefits that big data can bring us, and the challenges that companies face. This chapter also lists and characterizes several technologies that can make it easier for companies to work with big data. The second chapter is the goal of this thesis and the setting of other sub-goals that need to be done to achieve the primary goal. The third chapter deals with the processing methodology and user research methods used. The last chapter shows the identification of the challenges companies face in the transformation to data-driven companies and the significant trends in several areas, such as technology, legislation, and human capital. This chapter summarizes the state of implementing these trends in Slovakia and abroad. This thesis also provides our recommendations, which should support companies to overcome the challenges associated with current trends in the field.

Keywords: Big data, data analytics, data-driven companies, big data trends, digital transformation

Obsah

Úvod.....	12
1. Súčasný stav riešenej problematiky doma a v zahraničí.....	14
1.1. Definícia Big data	14
1.2. Výhody a nevýhody Big data.....	15
1.3. História Big data	17
1.4. Vlastnosti Big data	18
1.4.1. Objem (Volume).....	18
1.4.2. Rýchlosť (Velocity).....	18
1.4.3. Rozmanitosť (Variety)	18
1.4.4. Hodnota (Value).....	19
1.4.5. Pravdivosť (Veracity).....	19
1.4.6. Vizualizácia (Visualization).....	20
1.4.7. Variabilita (Variability).....	20
1.5. Typy dát	20
1.5.1. Štruktúrované údaje.....	21
1.5.2. Neštruktúrované údaje.....	21
1.5.3. Polo štruktúrované údaje	22
1.6. Big data analytika.....	22
1.6.1. Zber dát.....	23
1.6.2. Ukladanie a spracovanie dát.....	23
1.6.3. Čistenie dát	24
1.6.4. Analýza dát.....	24
1.6.5. Vizualizácia dát	25
1.7. Big data technológie.....	25
1.7.1. NoSQL.....	26
1.7.2. Apache Hadoop	27
1.7.3. Apache Spark	28

1.7.4. Tableau	28
2. Cieľ práce.....	30
3. Metodika práce a metódy skúmania	31
4. Výsledky práce	32
4.1. Zhodnotenie aktuálnej situácie a identifikácie výziev pri transformácií spoločnosti na dátovo riadené spoločnosti.....	32
4.2. Identifikácia súčasných trendov	40
4.2.1. Cloud computing	42
4.2.2. Umelá inteligencia.....	45
4.2.3. Potreba nových zamestnancov a potreba nových pozícií.....	48
4.2.4. Bezpečnosť a ochrana osobných údajov	50
4.2.4.1. Bezpečnosť z technologického hľadiska.....	50
4.2.4.2. Bezpečnosť z pohľadu ochrany osobných údajov.....	51
4.3. Možnosti implementácie identifikovaných trendov v rámci vybraných sektorov vo svete a na Slovensku	52
4.3.1. Poistovníctvo	52
4.3.2. Bankovníctvo.....	54
4.3.3. Zdravotníctvo	55
4.4. Zhrnutie identifikovaných trendov, výziev a naše odporúčania	56
Záver	59
Zoznam použitej literatúry	61

Zoznam obrázkov

Obrázok 1- Znázornenie 3V – zdroj: [25].....	19
Obrázok 2- Znázornenie 7V -zdroj: [28]	20
Obrázok 3- Typy údajov – zdroj: [26]	22
Obrázok 4 - Rozdiely medzi SQL a NoSQL databázou – zdroj: [27]	27

Zoznam grafov

Graf 1-Podiel rozhodovacích prístupov na rozhodovaní v rámci spoločností – zdroj: vlastné spracovanie; údaje extrahované z [15]	33
Graf 2-Geografický pohľad na spoločnosti , ktoré uviedli že ich rozhodovanie je celé postavené na dátach – zdroj: vlastné spracovanie; údaje extrahované z [15].....	33
Graf 3-Sektorový pohľad na spoločnosti, ktoré uviedli že ich rozhodovanie je celé postavené na dátach – zdroj: vlastné spracovanie; údaje extrahované z [15].....	34
Graf 4-Prehľad sektorov, ktoré dokážu najefektívnejšie speňažiť svoje dáta a kvantifikujú hodnotu ich dát – zdroj :vlastné spracovanie; údaje extrahované z [15].....	34
Graf 5-Spokojnosť zamestnancov s kvalitou dát – zdroj: vlastné spracovanie; údaje extrahované z [15]	36
Graf 6-Výzvy ovplyvňujúce spoločnosti – zdroj: vlastné spracovanie; údaje extrahované z [16]	37
Graf 7-Predstava spoločností o čase potrebnom na transformáciu spoločnosti – zdroj vlastné spracovanie; údaje extrahované z [16]	38
Graf 8-Investičné priority spoločnosti pre rok 2022 – zdroj: vlastné spracovanie; údaje extrahované z [16].....	39
Graf 9-Stratégie spoločnosti v oblasti zavádzania big data – zdroj: vlastné spracovanie; údaje extrahované z [16].....	40
Graf 10- Množstvá dát v zettabajtoch, ktoré sú generované za rok a odhad do roku 2025 – zdroj: vlastné spracovanie; dáta extrahované z [17].....	40
Graf 11-Predpoklad využitia cloud computingu – zdroj: vlastné spracovanie; údaje extrahované z[18]	42
Graf 12-Predpokladaný vývoj SaaS na nasledujúcich 18 mesiacov – zdroj: vlastné spracovanie; údaje extrahované z [18].....	43
Graf 13-Zavedenie nových pracovných pozícií po prechode na cloud technológie – zdroj: vlastné spracovanie; údaje extrahované z [18].....	44
Graf 14-Najčastejšie prípady použitia AI- zdroj: vlastné spracovanie; údaje extrahované z [19]	46
Graf 15-Riziká ovplyvňujúce AI v roku 2020 – zdroj: vlastné spracovanie; údaje extrahované z [19]	47
Graf 16-Riziká ovplyvňujúce AI v roku 2021 – zdroj: vlastné spracovanie; údaje extrahované z [19]	47

Graf 17-Obsadenie pozície Chief data officer/ Chief data analytics officer- zdroj: vlastné spracovanie; údaje extrahované z [16]	49
Graf 18-Požiadavky na CDO/CDAO- zdroj: vlastné spracovanie; údaje extrahované z [16]	50
Graf 19-Najvyššie pokuty za porušenie GDPR v mil. eur do roku 2022 - zdroj: vlastné spracovanie; údaje extrahované z [21].....	51

Zoznam skratiek

ETL - extract, transform, load

ELT - extract, load, transform

IaaS - Infrastructure as a service

PaaS - Platform as a service

SaaS - Software as a service

AI - Artificial intelligence

CDO - Chief Data Officer

CDAO - Chief Data Analytics Officer

GDPR - General Data Protection Regulation

EU – Európska únia

Úvod

V ostatnom čase zaznamenávame rýchly rozvoj technológií a s týmto rozvojom technológií ide ruka v ruke aj rast množstva generovaných dát. Je to spôsobené najmä tým, že v dnešnej dobe, už skoro každé zariadenie odosiela množstvo dát, ktoré zaznamenáva. Tieto dáta sú však generované v takom množstve, tak rýchlo a v tak rôznych formách, či už ide o štruktúrované, neštruktúrované alebo polo štruktúrované dáta, že ich nedokážeme spracovávať tradičnými metódami. Preto pre nich vznikol vlastný pojem a to big data. Okolo nás je množstvo zariadení, ktoré sú denno-denne vyžívané ľuďmi a generujú veľké objemy dát. Takýmito zariadeniami sú napríklad smart hodinky, mobily, GPS lokátory a ďalšie rôzne senzory na internete vecí. Už niekoľko rokov pozorujeme trend využívania týchto big data v rámci súkromného sektora a snahu zahrnúť túto oblasť v rámci jednotlivých spoločností, aby z nej dokázali spoločnosti vyťažiť maximálny potenciál, ktorý v sebe skrýva. Nakoľko postupne prebieha digitalizácia aj vo verejnom sektore, tak vidíme postupný nárast záujmu o big data technológie aj v tejto oblasti. Sú to napríklad rôzne kamerové systémy pokrývajúce dopravu v rámci krajiny, zdravotné záznamy, úradne správy a rôzne ďalšie oblasti, kde sa prejavila digitalizácia a s ňou spojený nárast dát.

Avšak zber dát, je iba prvá časť, ktorá musí byť vykonaná, aby big data dokázali priniesť očakávaný prínos. Ak chceme získať pridanú hodnotu z týchto dát je potrebné tieto dáta spracovať, očistiť ich, vykonať analýzu a následne výsledky implementovať do rozhodnutí. Ale nie je to také jednoduché ako sa na prvý pohľad zdá. Analýza dát nám prináša aj ďalšie technologické výzvy, ktoré musia spoločnosti vyriešiť pred tým než získajú očakávané výsledky. Patria sem výzvy súvisiace s potrebnou infraštruktúrou, spôsoby ukladania dát a v neposlednom rade aj zabezpečenie kvalifikovaných dátových odborníkov. Výsledky z tejto analýzy nám dokážu pomôcť lepšie zvoliť stratégiu, správne nastaviť cieľ, včas identifikovať riziká a príležitosti, ktoré sme doteraz prehliadali a správnou implementáciou výsledkov vieme získať výhodu oproti konkurencii.

Ak sa bavíme o big data je nutné spomenúť nielen výhody, ktoré sme popísali, ale aj riziká, ktoré s týmto pojmom prichádzajú. Ako sme už vyššie spomínali, dáta sú generované množstvom zariadení, ktoré bežné používame a zachytávajú o nás údaje o polohe, nákupnom správaní a o rôznych ďalších veciach. Veľmi ľahko sa môže stať, že o nás spoločnosti získajú údaje, na ktoré sme im nedali povolenie. Preto v rámci big data je dôležité sa zaoberať aj témou ochrany osobných údajov. Táto téma sa najmä v posledných rokoch po prepuknutí viacerých káz do popredia a jednotlivé štáty sa snažia upraviť legislatívu tak, aby

sa šanca na práve takýto únik citlivých dát zminimalizovala. Nakoľko nás táto oblasť big data zaujala, a ide o tému, ktorá sa neustále vyvíja a mení, rozhodli sme sa túto oblasť bližšie popísať v rámci našej diplomovej práce.

Cieľom našej diplomovej práce bolo detailnejšie objasniť, nie len to, čo to vlastne big data sú a ako sa používajú v súčasnosti, ale aj to ako sa rozvíjajú a aké sú momentálne trendy v tejto oblasti, a ktoré z týchto trendov by mali spoločnosti implementovať do svojich riešení. Dané trendy sme si potom premietli do rôznych sektorov, v ktorých sú tieto trendy implementované, respektíve by sa dali implementovať v rámci Slovenska, a spísali naše odporúčania pre jednotlivé spoločnosti na Slovensku.

V prvej kapitole sme si bližšie popísali, čo to vlastne big data sú a zadefinovali sme si kľúčové pojmy v rámci tejto tematiky. Okrem toho sme si popísali, ako sa tento koncept vyvíjal v čase, prečo je dôležité sa zaoberať touto témou, aké výhody nám to môže priniesť a aké riziká sa s tým spájajú. Taktiež sme si predstavili zopár nástrojov, ktoré sa využívajú pre prácu s big data.

V rámci druhej a tretej kapitoly sme si charakterizovali metódy skúmania, ktoré sme použili pri vypracovaní diplomovej práce a bližšie sme si popísali hlavný cieľ a jeho jednotlivé čiastkové ciele, ktorých úspešným splnením sa nám podarilo naplniť hlavný cieľ.

V záverečnej kapitole, ktorá je rozdelená do dvoch okruhov sme si najskôr zhrnuli aktuálny stav a identifikovali rôzne výzvy, ktoré sú spojené s transformáciou spoločnosti na dátami riadené spoločnosti. Na základe výsledkov z prieskumov sme poukázali na to, čo predstavuje momentálne najväčšie výzvy pri transformácii spoločnosti a poskytli odporúčania, ako minimalizovať tieto výzvy. Okrem výziev sme odhadli aj smerovanie a investičné priority na ktoré by sa mali spoločnosti zamerať. V rámci výsledkov práce sme taktiež identifikovali trendy, ktoré sú v danej oblasti. Trendy sme rozdelili do troch oblastí a to trendy týkajúce sa technológií, trendy v oblasti bezpečnosti a regulácie a poslednou oblasťou boli trendy v oblasti pracovných pozícií. Z každej z týchto oblastí sme zvolili najvplyvnejší trend, ktorý sme popísali a zhodnotili aké výzvy sú s ním spájané a čo nám zavedenie môže priniesť. Následne sme si zvolili niekoľko sektorov a popísali ako sú na tom s implementáciou big data riešení a daných trendov. Taktiež sme sa pozreli na situáciu v rámci Slovenska a poukázali na možnosti implementácie a poskytli naše odporúčania.

1. Súčasný stav riešenej problematiky doma a v zahraničí

Pojem big data sa stále viac dostáva do popredia a to najmä kvôli exponenciálnemu nárastu dát, ktoré sú vytvárané. Technológie big data sú súčasťou architektúr väčšiny popredných spoločností v rôznych sektoroch, ktoré si uvedomujú aké výhody im to dokáže priniesť a akým veľkým stratám dokážu predchádzať, ak sa zaoberajú pojmom big data správnym spôsobom. Avšak nekončí všetko iba pri zozbieraní dát, je potrebné správne nielen zozbierať a očistiť dáta ale aj ich ďalej spracovať, analyzovať a výsledky implementovať do našich rozhodnutí, lebo len tak nám dokážu priniesť pridanú hodnotu. V ostatnom období sa začína tento pojem postupne udomácňovať v každom odvetví. Avšak aj napriek tomu ostáva stále množstvo dát, ktoré sú generované, a ostávajú nespracované. A práve preto je dôležité sa zaoberať touto témou. V nasledujúcich podkapitolách si zadefinujeme pojem big data, opíšeme jeho základne vlastnosti, v stručnosti si zhrnieme krátku históriu tohto pojmu a charakterizujeme technológie, ktoré nám môžu uľahčiť prácu.

1.1. Definícia Big data

Termín big data môžeme považovať za dôsledok neustále prebiehajúcej digitalizácie. V každom okamihu sú spoločnosti zaplavované množstvom dát, s ktorými najmä kvôli ich veľkosti a rozmanitosti nie je jednoduché manipulovať a spravovať ich. Pre manipuláciu s týmito dátami nie je možné použiť tradičné metódy, ale vyžaduje sa použitie špecializovaných technológií. V praxi sa stretávame s niekoľkými typmi definícií pojmu big data. Niektoré definície vychádzajú zo základných vlastností big data, iné zase poukazujú na neschopnosť spracovávať big data pomocou tradičných metód a niektoré sú kombináciou obidvoch spomínaných. [1]

Definícia od spoločnosti Gartner, konkrétne od analytika Doug Laney, ktorý ako prvý opísal tri základne vlastnosti big data znie : „*Big data sú veľkoobjemové, vysokorychlostne a rôznorodé informačné aktíva, ktoré si vyžadujú nákladovo efektívne, inovatívne formy spracovania informácií pre lepší prehľad a rozhodovanie.*“[2]

Táto definícia však časom, už neopisovala aktuálny stav a preto sa k trom základným vlastnostiam, postupne v čase pridávali aj ďalšie. V súčasnosti niektorí autori uvádzajú päť, sedem, respektíve až deväť základných vlastností takzvaných 9V. [3]

1.2. Výhody a nevýhody Big data

Ak sú big data používané správne, vedia nám priniesť veľa výhod, ale ako to už býva v oblasti technológií, výhody idú ruka v ruke s nevýhodami, respektíve s rizikami, ktoré na nás čakajú, v prípade že sa rozhodneme vstúpiť do tejto oblasti.

Medzi výhody patrí:

- **Optimalizácia nákladov**

Asi každý manažér, chce riadiť svoju spoločnosť tak, aby dosahovala čo najvyššie zisky, avšak nie vždy sa im to darí. Pomocou nástrojov big data sa im to však môže podariť, nakoľko dokážu zefektívniť procesy a zlepšiť produktivitu, čo môže priniesť úspory. Dokážu taktiež identifikovať ďalšie príležitosti v spôsobe a rozvrhu práce, ktoré môže viesť k ďalšiemu zníženiu nákladov. [5]

- **Zlepšenie efektivity**

Pomocou nástrojov big data dokážeme zhromaždiť veľké množstvo údajov od zákazníkov, analyzovať ich a nájsť rôzne súvislosti, ako sú napríklad správanie zákazníkov pri nakupovaní, tendencie v správaní, preferencie zákazníkov a následným implementovaním výsledkov do rozhodnutí, môžeme v budúcnosti lepšie prispôbiť produkty, respektíve služby pre konkrétne cieľové skupiny. [6]

- **Šanca prijať lepšie rozhodnutia**

Nástroje big data môžu nájsť v množstve údajov, vzťahy, respektíve súvislosti, ktoré dovtedy neboli úplne známe. Na základe týchto skutočností, dokážu manažéri prijať lepšie obchodné rozhodnutia, upraviť dlhodobé stratégie spoločnosti, upraviť ceny a na základe analýzy konkurenčného prostredia dokážu lepšie prispôbiť produkty cieľovým skupinám, resp. začať vyvíjať nové produkty. [6]

- **Včasná identifikácia podvodov**

Táto výhoda sa týka najmä finančných spoločností ako sú banky a úverové spoločnosti ktoré pomocou algoritmov strojového učenia a umelej inteligencie dokážu zistiť podozrivé správanie, resp. transakcie. Tieto algoritmy dokážu na základe údajov z minulosti a z minulého správania zákazníka zistiť, že sa môže jednať o podvodnú transakciu a v dostatočnom predstihu ju zablokovať a informovať zákazníka. Práve v tom spočíva výhoda, že tieto spoločnosti dokážu lepšie slúžiť klientom, nakoľko ich dokážu včas varovať a ochrániť klienta pred možnými podvodmi. [5]

Medzi nevýhody patrí:

- **Nedostatočná kvalita údajov**

Nakoľko nástroje big data pracujú s obrovským množstvom údajov z rôznych zdrojov, nie vždy musia byť tieto zozbierané údaje presné resp. pravdivé. A práve to sa môže prejavovať na výsledkoch analýz z týchto údajov, kedy získané poznatky môžu byť nepresné a bezcenné pre spoločnosť a ich implementovaním do rozhodnutí nám môžu spôsobiť viac škody ako úžitku. [7]

- **Nedostatok odborníkov v danom obore**

Aby spoločnosť dokázala využiť všetok potenciál, ktorý big data ponúka, potrebuje odborníkov, ktorí dokážu spravovať špecializované nástroje, potrebnú infraštruktúru a majú vedomosti na vyhodnotenie výsledkov, ktoré dostanú. Nakoľko sa táto oblasť v ostatných rokoch udomácnila nielen v súkromnom, ale aj verejnom sektore a každý chce získať z tejto oblasti čo najviac, tak momentálne je na trhu nedostatkový stav dátových vedcov. Preto je veľmi zložitá nájsť vhodných kandidátov na zaplnenie týchto pozícií.[5]

- **Legislatíva**

S množstvom pribúdajúcich dát rastie aj množstvo zákonov, usmernení a nariadení, ktoré súvisia priamo so zbieraním a manipuláciou týchto dát. Týka sa to najmä spoločnosti, ktoré pracujú s osobnými údajmi napríklad rôzne poisťovne a banky, ktoré by si mali dať pozor na dodržiavanie práve týchto legislatívnych zmien, ak nechcú byť sankcionované od štátu za nedodržiavanie legislatívy a s pribúdajúcim množstvom dát sa bude množstvo tejto legislatívy len a len zväčšovať. [7]

- **Bezpečnostné riziká**

V množstve dát, ktoré sa zozbierajú sa nachádzajú často aj informácie s citlivým obsahom, preto je potrebné dostatočne zabezpečiť, aby boli tieto údaje chránené. To si vyžaduje dodatočnú pozornosť a údržbu najmä v oblasti kybernetickej bezpečnosti, pretože množstvo citlivých údajov pokope priťahuje pozornosť rôznych skupín, ktoré môžu reálne ohroziť spoločnosť kybernetickým útokom a zapríčiniť stratu citlivých údajov. A práve preto sa tieto úniky údajov považujú za jednu z najväčších, ak nie práve tu najväčšiu hrozbu, ktorá prichádza s big data.[7]

1.3.História Big data

V rámci histórie big data pred rokom 2005, môžeme hovoriť o čiastkových krokoch, ktoré viedli k zavedeniu tohto pojmu. K týmto krokom môžeme zaradiť napríklad vývoj počítačov, mobilov, rôznych senzorických zariadení a v neposlednom rade internetu.[1]

- Rok 1927 sa spája s menom Fritz Pfleumer, ktorý vyvinul jeden z prvých prostriedkov na magnetické ukladanie informácií [1]
- Počas druhej svetovej vojny, konkrétne v roku 1943 sa Britom podarilo vynájsť stroj, na prelomenie nemeckých kódov, ktorý bol založený na skenovaní rôznych vzorov v správach. Tento stroj dokázal znížiť čas potrebný na rozlúštenie z niekoľkých týždňov na pár hodín. Táto udalosť podnietila potrebu vyvinúť počítače do bodu, keď boli schopné zhromažďovať a spracovávať údaje.[8]
- Prvý záujem o ukladanie údajov sa spája s rokom 1965 a s vládou USA, ktorá postavila dátové centrum na uchovávanie súborov a daňových priznaní. Avšak nakoniec bol tento projekt neúspešný.[8]
- Rok 1977 sa spája s predstavením mikropočítačov, a považujeme ho za veľký míľnik v rozvoji internetu a veľkých dát.[8]
- Veľký prelom nastal v roku 1989, keď bol predstavený koncept world wide web a následne v roku 1999 koncept Internet of things.[8]
- S rozvojom osobných počítačov a internetu došlo k takému nárastu dát, že ich nebolo možné spracovávať tradičnými nástrojmi. Preto sa v roku 2005 definoval nový pojem big data. V tom istom roku bol tiež predstavený aj Hadoop, ako open source softvér pre spoľahlivé a distribuované výpočty.[8]
- V roku 2014 sa prvýkrát stalo, že pre prístup k digitálnym dátam použilo viac ľudí mobilné zariadenia ako domáce resp. kancelárske počítače. [1]
- Rok po prvom zedefinovaní pojmu big data a predstavení Hadoop, Amazon web services začali ponúkať službu známu ako cloud-computing. [8]
- V roku 2015 postupne začali Google a Microsoft masovo budovať nové dátové centrá a zaradili sa spolu s Amazonom k špičke v tomto odvetví.[8]
- Rok 2018 sa spája s postupnou integráciou do vývoja strojového učenia, a s prípravou na zvýšený dopyt po dátových rolách ako sú dátoví analytici a dátoví architekti, ale taktiež so zvýšeným záujmom o bezpečnosť údajov. [8]

1.4. Vlastnosti Big data

V rámci big data môžeme hovoriť o troch kľúčových vlastnostiach, ktoré nám pomôžu lepšie pochopiť tento pojem. Avšak aj tieto vlastnosti sa vyvíjali v čase a k trom základným V sa časom pridávali aj ďalšie V, a v niektorých prípadoch môžeme hovoriť o 5V, 7V, respektíve až o 9V. Podľa nášho názoru najlepšie opisuje big data definícia z roku 2013, kde bola pôvodná definícia, ktorá zahŕňala 3V rozšírená o ďalšie 4V a to konkrétne o veracity, variability, visualization and value. V ďalších podkapitolách sme si preto charakterizovali 3 základne vlastnosti a zhrnuli si aj ďalšie vlastnosti, ktoré rozšírili pôvodné 3V na 7V. [9]

1.4.1. Objem (Volume)

Najjasnejšou vlastnosťou je objem, nakoľko môžeme tvrdiť, že big data označujú obrovské množstvo údajov, ktoré sú generované. Keď sa na túto vlastnosť pozrieme v čase, tak môžeme vidieť, že napríklad kým v roku 2001, keď bol prvýkrát zadefinovaný pojem big data, sme hovorili priemerne o veľkosti týchto dát v MB a GB, tak od roku 2015 môžeme hovoriť už o zettabajtoch. Je to dôsledok najmä rýchleho rozvoja technológií, využívania sociálnych sietí a rôznych senzorov v rámci internetu, ktoré nám priniesli niekoľkonásobný denný prírastok nových údajov, ktoré vieme získavať a ďalej spracovávať. V roku 2020 bolo vytvorených 40 zettabajtov údajov, čo predstavuje 300 násobné zvýšenie oproti roku 2005. Do roku 2025 sa predpokladá, že bude na svete 175 zettabajtov údajov. Do roku 2003 bolo na celom svete iba 5 miliárd gigabajtov údajov. V súčasnosti 90% všetkých údajov bolo vygenerovaných v posledných rokoch.[9]

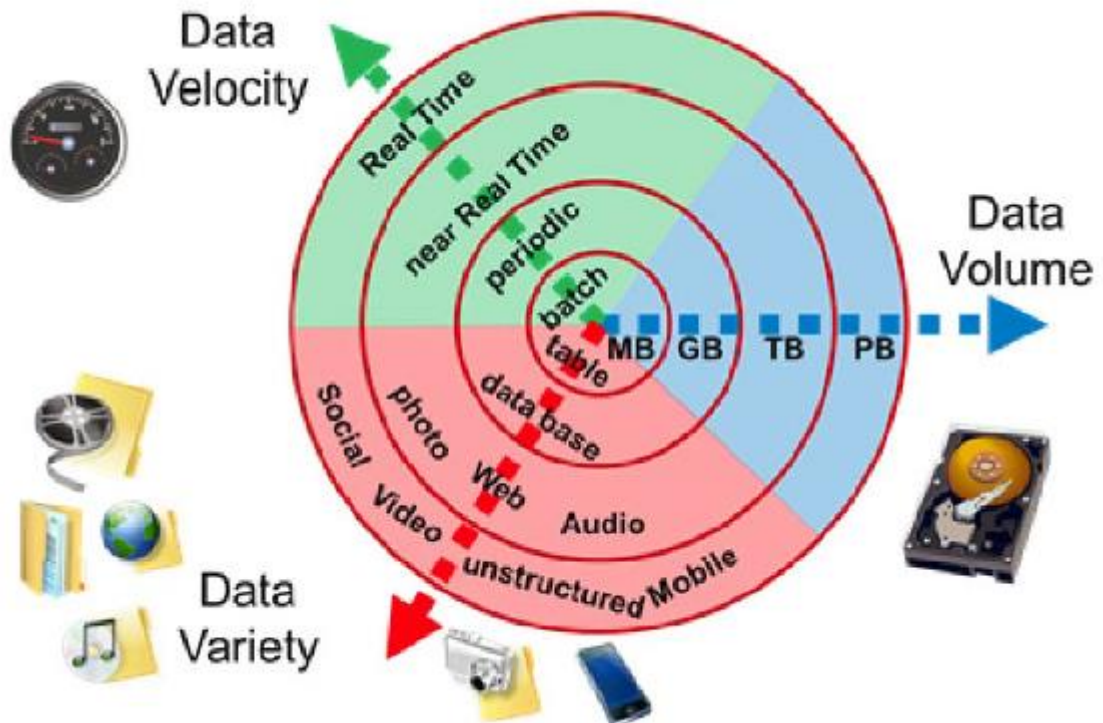
1.4.2. Rýchlosť (Velocity)

Pod touto vlastnosťou si môžeme predstaviť rýchlosť, akou sú dáta generované, analyzované a spracovávané. V dnešnej technologickej dobe sú údaje generované nielen na úrovni dennej alebo hodinovej, ale v každom okamihu sa vygeneruje obrovské množstvo dát, ktoré je potrebné zozbierať, spracovávať a ukladať. Je potrebné vykonávať tieto akcie v čo najkratšom čase, aby sme predišli prípadným rizikám, ktoré sme schopní odhaliť na základe analýzy dát.[9]

1.4.3. Rozmanitosť (Variety)

O rozmanitosti môžeme hovoriť v súvislosti s formou v akej sú tieto údaje generované. Kým v minulosti sme sa stretávali s tým, že údaje boli doručované v súboroch ako Excel alebo Access, v súčasnosti sa údaje medzi jednotlivými aplikáciami môžu veľmi líšiť. Údaje môžu byť v rôznych neštruktúrovaných formách ako sú videá, zvukové nahrávky

alebo fotografie. Na spracovanie takýchto údajov je potrebné mať pokročilejšie analytické schopnosti a technológie, ktoré sú schopné tieto údaje spracovať.[3]



Obrázok 1- Znáozornenie 3V – zdroj: [25]

1.4.4. Hodnota (Value)

Niektorí autori uvádzajú , že okrem týchto troch základných vlastností, môžeme pre opis big data použiť aj ďalšie vlastnosti, ako napríklad value teda hodnota údajov, na ktorej sa zhoduje viacero autorov. Avšak táto vlastnosť je subjektívna, nakoľko pre rôzne firmy, môžu byť hodnotné iné typy údajov, a preto je dôležité definovať konkrétne typy údajov, ktoré majú pre danú firmu najväčšiu hodnotu. Zozbierané údaje nám samé o sebe neposkytnú žiadnu pridanú hodnotu, potrebujeme vykonať viacero akcií, aby sme dokázali previesť tieto údaje na informácie a poznatky, ktoré nám pomôžu v budúcom rozhodovaní. [10]

1.4.5. Pravdivosť (Veracity)

Pravdivosť predstavuje vlastnosť, ktorá má najväčší priestor sa vyvíjať, respektíve zlepšovať. V dôsledku množstva dát, ktoré sú zozbierané z rôznych zdrojov je dôležité klásť dôraz nielen na kvalitu údajov, ale aj na dôveryhodnosť zdroja údajov. V rámci pravdivosti rozlišujeme dva typy údajov a to údaje s vysokou pravdivosťou, teda údaje ktoré vysokou mierou prispievajú k celkovým výsledkom, ako sú napríklad údaje z experimentov,

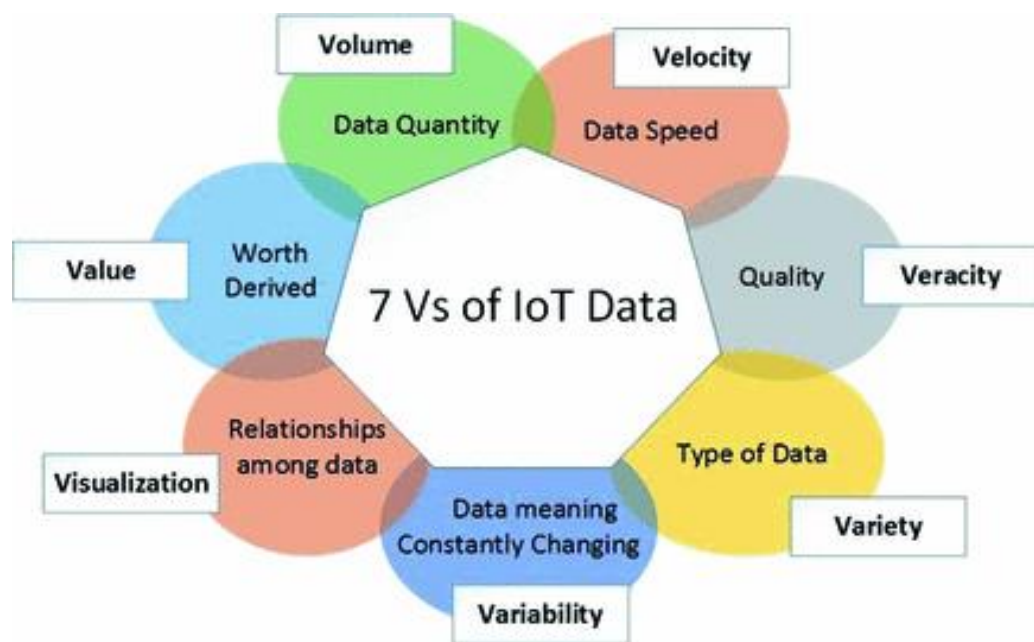
respektíve vedeckých pokusov. Na druhej strane máme údaje s nízkou pravdivosťou, ktoré obsahujú množstvo nezmyselných údajov, ktoré sa označujú ako šum. Preto je jedným z kľúčových krokov stanovenie pravdivosti údajov.[3]

1.4.6. Vizualizácia (Visualization)

Po zozbieraní a spracovaní údajov je dôležité vhodným spôsobom odprezentovať jednotlivé výsledky, a tým pádom sa dostávame k ďalšej charakteristike, ktorou je vizualizácia údajov. Pre manažment firiem je jednoduchšie na pochopenie, ak sú výsledky odprezentované formou rôznych grafov a dashboardov, ako keby mali dostať výsledky vo forme tabuliek s množstvom čísel, údajov a vzorcov.[3]

1.4.7. Variabilita (Variability)

Pri tejto charakteristike hovoríme najmä o neustále sa meniacom množstve údajov, zdrojov údajov, ich typu a rýchlosti, ktorou sú generované. Táto vlastnosť nám hovorí najmä o tom, že údaje sa neustále menia. Pri zbieraní údajov z rovnakého zdroja k nám každý deň prichádzajú rôzne údaje a to môže mať vplyv na kvalitu našich údajov. [10]



Obrázok 2- Znárodnenie 7V -zdroj: [28]

1.5. Typy dát

Jednou z vlastností big data je aj rozmanitosť, čo súvisí najmä tým, že v rámci big data sa stretávame s rôznymi formátmi údajov, ktoré sa spracovávajú z rôznych systémov. Ide napríklad o štruktúrované, neštruktúrované údaje, geografické alebo o údaje založené na nejakej udalosti, respektíve čase. Okrem toho, že v rámci big data sa môžeme stretávať s rôznym typom údajom, tiež záleží na tom, kto, resp. čo je zdrojom údajov, či sú to ľudia

alebo sú to údaje generované systémom. Pretože údaje vyprodukované ľuďmi sú častokrát menej presné a detailné. V nasledujúcich podkapitolách sme si charakterizovali rôzne formáty údajov s ktorými sa môžeme v rámci big data stretnúť.[1]

1.5.1. Štruktúrované údaje

Štruktúrované údaje sú údaje, o ktorých môžeme tvrdiť, že majú preddefinované pravidlá a štruktúru, pred tým než sú umiestnené do úložiska. Tieto údaje tvoria iba 20% zo všetkých údajov. Väčšinou v súvislosti so štruktúrovanými údajmi hovoríme o textových údajoch. Ako príklad, kde sú ukladané štruktúrované údaje môžeme uviesť napríklad relačnú databázu alebo Excel, kde sú údaje pred vložením presne naformátované, podľa preddefinovaných polí, ako je napríklad dátum, adresa, telefónne čísla a podobne, aby sa v nich jednoducho dalo vyhľadávať pomocou SQL kódu. [1]

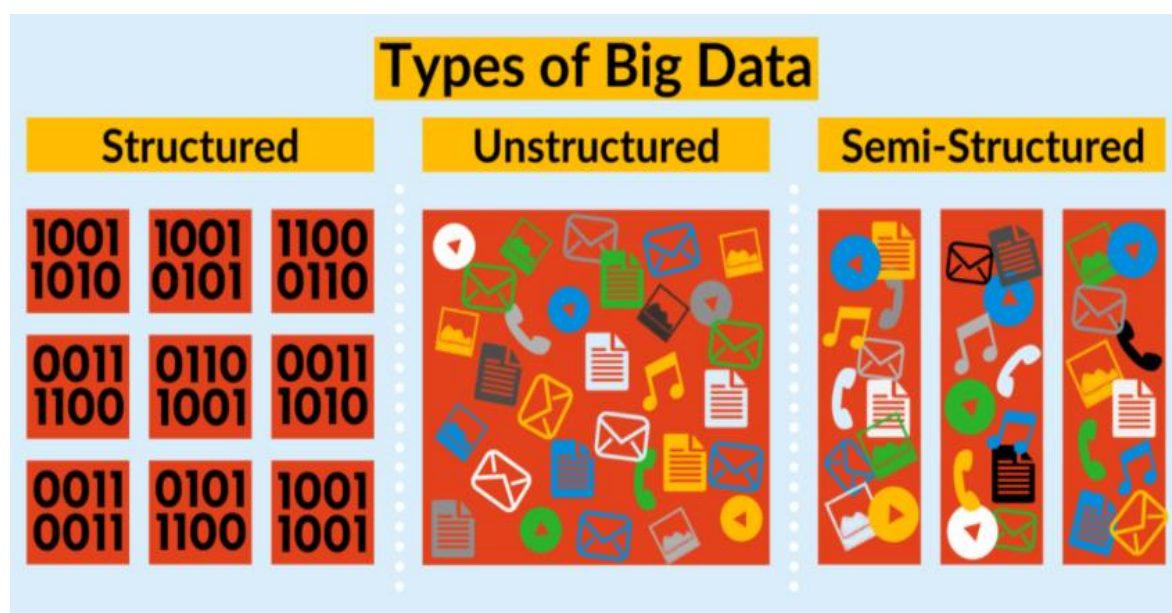
Medzi výhody štruktúrovaných dát môžeme zaradiť ich jednoduchosť pre používateľa, nakoľko na pochopenie týchto údajov, stačí ak používateľ aspoň trochu rozumie téme, ktorej sa dané údaje týkajú. Nepotrebuje vedieť bližšie podrobnosti o rôznych vzťahoch medzi jednotlivými údajmi ani im podrobne rozumieť. Ako o ďalšej výhode môžeme hovoriť o množstve nástrojov, ktoré sú dostupné pre používateľov na analýzu týchto údajov. Tieto štruktúrované údaje majú aj zopár nevýhod a to napríklad ich preddefinovanú štruktúru, nakoľko táto štruktúra je obmedzujúca len na použitie pre daný účel na ktorý bola vytvorená, čím je výrazne obmedzená flexibilita na rôzne prípady použitia.[1]

1.5.2. Neštruktúrované údaje

V prípade neštruktúrovaných údajov hovoríme o údajoch, ktorým chýba preddefinovaná štruktúra a preto nie sú vhodné pre relačnú databázu. Na ich ukládanie sa využívajú tzv. „dátové jazerá“. V týchto údajoch je zložitejšie sa orientovať a sú náročnejšie na vyhľadávanie. Ako neštruktúrované údaje môžeme označiť videá, obrázky, správy, poznámky, internetové stránky a rôzne iné. Na prvý pohľad nám tieto údaje nedokážu priniesť pridanú hodnotu. Ukládanie týchto údajov je náročnejšie ako pri štruktúrovaných, nakoľko tieto údaje poväčšine zaberajú veľa úložného miesta. Kvôli ich rôznorodnej štruktúre je zložité v nich vyhľadávať a aby sa naplno využil ich potenciál je potrebné mať odborné znalosti a ovládať prácu s rôznymi špecializovanými nástrojmi. Tieto údaje sú vhodné napríklad na predvídanie nákupných trendov na základe údajov zo sociálnych sietí a webových stránok alebo na určovanie efektivity marketingových kampaní.[4]

1.5.3. Polo štruktúrované údaje

Údaje, ktoré sú zozbierané nemusia byť iba štruktúrované alebo neštruktúrované. Na ich rozhraní stoja tzv. polo štruktúrované údaje, ktoré nemajú pevne danú štruktúru, ale obsahujú metadáta, ktoré uľahčujú ich analýzu. Ak si zoberieme ako príklad fotografiu, tá neobsahuje preddefinovanú štruktúru, ale má niektoré atribúty, vďaka ktorým môžeme tvrdiť, že je to polo štruktúrovaný údaj. Sú to napríklad ID zariadenia, dátum a čas, kedy bola fotografia vyhotovená a geolokáciu. Ako polo štruktúrované údaje môžeme označiť HTML kód, grafy, tabuľky a XML dokumenty.[26]



Obrázok 3- Typy údajov – zdroj: [26]

1.6. Big data analytika

Analytika big data predstavuje proces, v ktorom sa z množstva bezcenných údajov stanú kľúčové poznatky, ktoré môžu niekedy predstavovať práve ten chýbajúci kus, ktorý rozhodne o tom či bude splnený cieľ, ktorý si spoločnosť dala alebo nie. Tento proces sa snaží odhaliť rôzne súvislosti, vzorce a trendy medzi nespracovanými údajmi, ktoré sme zozbierali. Avšak tento proces zahŕňa aj niekoľko ďalších fáz ako zber, zhromažďovanie, spracovanie, čistenie a následnú analýzu dát. S rozvojom výpočtovej techniky a výrazným nárastom údajov, ktoré sa v každom okamihu generujú, sa začali vyvíjať aj prvé projekty zamerané na uľahčenie tohto procesu, nakoľko sa tu tradičné metódy použiť nedali. Patria sem projekty ako Hadoop, Spark a NoSQL databázy. V nasledujúcich podkapitolách si v krátkosti charakterizujeme kroky, ktoré sa skrývajú pod pojmom big data analytika.[11]

1.6.1. Zber dát

Dnešné pokročilé technológie umožňujú organizáciám zber rôznych údajov, či už ide o štruktúrované, polo štruktúrované alebo neštruktúrované údaje. Tieto údaje môžu byť z rôznych zdrojov, či už sú to rôzne senzory na internete, mobilné aplikácie, alebo cloudové úložiska. Rozlišujeme niekoľko základných typov údajov, ktoré môžu byť zozbierané :

- Údaje v reálnom čase – vznikajú prostredníctvom senzorov, ktoré generujú údaje v každom okamihu či už o našom správaní, resp. sú to údaje zachytávané v rôznych sektoroch prostredníctvom senzorov alebo rôzne aplikácie, ktoré poskytujú množstvo údajov
- Transakčné údaje – vznikajú po uskutočnení online nákupu, ide o údaje ako čas nákupu, spôsob nákupu a podobne
- Geografické údaje – každé dnešné smart zariadenie obsahujú GPS lokátory a pomocou satelitov sú tieto údaje o polohe neustále zachytávané
- Sieťové údaje – ide o údaje z rôznych typov sietí, ako sú sociálne médiá, internetu, mobilných a informačných sietí
- Údaje závisle na čase - ide o údaje, odohrávajúce sa práve v tomto okamihu, ako údaje o teplote, úroveň znečistenia alebo miera úmrtnosti
- Údaje z hlasového vyhľadávania - tieto údaje sa zhromažďujú z hlasového vyhľadávania zo zariadení pripojených na internet [1]

Proces zbieranie údajov sa líši v každej spoločnosti. Zber údajov prebieha v reálnom čase, aby sa zabezpečilo spracovanie v čo najkratšom čase. Súčasná technológia, ktoré sú dostupné nám umožňujú zhromažďovať nielen štruktúrované ale aj neštruktúrované dáta z rôznych zdrojov ako sú mobilné zariadenia, webové stránky, senzory a rôzne databázy. Tieto údaje, ktoré sú zozbierané vstupujú do ďalšieho procesu, v ktorom prechádzajú viacerými činnosťami ako sú extrakcia, transformácia a načítanie, tak aby sa dostali až do centralizovaných úložísk. Rozlišujeme dva prístupy presunu týchto údajov do cieľových úložísk, a to ETL a ELT. Ak je vyžitý ETL prístup, tak k transformácii dát musí dôjsť ešte predtým, než sa dostanú dáta do cieľového úložiska. Ak je použitý prístup ELT, tak je možné upravovať dáta aj po nahraní do cieľového úložiska. [1]

1.6.2. Ukladanie a spracovanie dát

Údaje je možné ukladať do úložísk ako sú napríklad rôzne cloudové dátové sklady alebo „dátové jazerá“, kde k nim jednoducho môžu pristupovať nástroje business

intelligence. V súčasnej dobe je v ponuke viacero cloudových riešení, ktoré v sebe obsahujú komponenty úložnej výpočtovej a klientskej infraštruktúry. V niektorých prípadoch sa stretávame aj s takzvanou klientskou vrstvou, kde prebiehajú činnosti správy údajov.

Spracovanie týchto údajov môžeme rozdeliť na základe počtu zariadení na ktorom sú dáta spracovávané. Rozlišujeme centralizované a distribuované spracovanie. Ak sa rozprávame o centralizovanom spracovaní, tak hovoríme, že všetko spracovanie sa odohráva v rámci jedného systému, na vyhradenom serveri. Slabou stránkou tohto prístupu je, že ak nastane nejaká chyba, tak zlyhá celý systém. Ak hovoríme o distribuovanom spracovaní, ktoré sa dostáva čoraz viac do popredia, tak hovoríme o množinách údajoch, ktoré sú príliš veľké nato, aby ich bolo schopné spracovávať na jednom zariadení a preto je táto množina údajov rozdelená na menšie časti a sú uložené na viacero miest na viacerých serveroch. Hlavnou výhodou tohto prístupu je v tom, že ak jeden server zlyhá je možné presunúť danú množinu údajov na iný server.[1]

Na základe času spracovania máme k dispozícii dva prístupy a to prístup v reálnom čase a dávkový prístup. V rámci spracovania v reálnom čase sa údaje spracovávajú nepretržite. Stále sú poskytované aktuálne údaje, všetko prebieha v rámci krátkeho časového rozpätia. Je najlepšou voľbou pre rýchle rozhodovanie, ale je zložitejšie na implementáciu. Na rozdiel od prístupu spracovania dát v reálnom čase, kde boli dáta spracovávané okamžite, pri dávkovom spracovaní, sa dáta zbierajú a spracovávajú iba vo vopred zadaných časoch. Táto možnosť je dobrou voľbou, najmä ak nám záleží viac na presnosti, než na rýchlosti.

1.6.3. Čistenie dát

Nakoľko údaje pochádzajú z rôznych zdrojov, vždy je potrebné dáta očistiť, aby sa zachovala kvalita údajov a dosiahli sa výsledky, ktoré sa vyžadujú. V tomto kroku sa zabezpečí, aby boli odstránené všetky duplicity, nezrovnalosti, nadbytočnosti a nepresné údaje. Rovnako tak sa zabezpečí aby boli všetky údaje správne naformátované. Nakoľko nečisté údaje môžu byť nepresné, a ak by neboli očistené môžu zmeniť výsledok celej analýzy. Preto je vhodné použiť nástroje kvality údajov, ktoré sú dostupné .[11]

1.6.4. Analýza dát

Zabezpečenie použiteľného stavu všetkých dát, si vyžaduje určitý čas. Kým sa nenachádzajú všetky dáta v požadovanom stave, nemôžu byť spustené ani pokročilé analytické procesy, ktoré dokážu premeniť dáta na pokročilé štatistiky. V rámci tohto kroku hovoríme o premene dát na použiteľné poznatky. Medzi tieto analytické procesy môžeme zahrnúť:

- Dolovanie dát (data mining)– popisuje proces hľadania významu a súvislosti v obrovskom množstve údajov. Spoločnosti využívajú tento proces, hlavne na lepšie spoznanie svojej cieľovej skupiny. [11]
- Strojové učenie (Machine learning)- podľa MIT[30]: „*Strojové učenie je podoblasť umelej inteligencie, ktorá je široko definovaná ako schopnosť stroja napodobňovať inteligentné ľudské správanie.*“ Keďže spoločnosti získavajú čoraz väčšie množstvo dát, stáva sa pre nich čoraz zložitejším využívať tieto dáta efektívne. Práve tú im môže pomôcť strojové učenie, nakoľko pre tieto systémy sú dáta prínosom a čím viac dát sme schopní im zabezpečiť, tým lepšie výsledky dokážu poskytovať.[11]
- Prediktívna analytika – táto technológia využíva algoritmy a techniky strojového učenia, aby dokázala identifikovať pravdepodobnosť budúcich výsledkov na základe dát z minulosti. [11]

1.6.5. Vizualizácia dát

V prípade vizualizácie údajov hovoríme o interpretácií spracovaných údajov do takej podoby, aby bolo na prvý pohľad jasné, čo nám dané dáta hovoria. Medzi techniky vizualizácie môžeme zaradiť napríklad grafy, mapy, pohyblivú grafiku, časové osy, regresné čiary a podobne. Tieto techniky umožňujú spoločnostiam lepšie prijímať rozhodnutia, nakoľko dokážu odprezentovať množstvo dát zrozumiteľným spôsobom. Vizualizácia big data dokáže popísať dáta skoro všetkých typov. Prináša nám oveľa viac možností, ako iba obyčajné grafy, histogramy a koláčové grafy, čo umožňuje spoločnostiam preskúmať údaje na identifikáciu rôznych neočakávaných vzorov, ktoré by sa v nich mohli vyskytnúť. Pre spoločnosti je dôležité zachovávať dobrý spôsob vizualizácie dát, pretože dokážu zachytiť rôzne nové trendy, odhaliť nové vzorce a umožniť kľúčovým osobám správne a včasné pochopenie a vykonanie rozhodnutí. Medzi najpoužívanejšie nástroje vizualizácie dát patria Power BI, Tableau, Qlik.[11]

1.7. Big data technológie

Big data sa udomácnili v množstve spoločností, a tým ako rýchlo sa navyšuje objem dát, je neustále potrebný aj vývoj nových technológií a nástrojov, ktoré nám dokážu uľahčiť prácu s nimi. Pod pojmom technológie big data , môžeme rozumieť softvérové nástroje, ktoré sú vyvinuté na vyhodnocovanie, spracovanie a extrahovanie informácií z množstva údajov. Tieto technológie sú spájané aj s inými oblasťami ako sú deep learning, machine

learning, umelá inteligencia a internet vecí. V rámci ďalších častí sme si charakterizovali niektoré obľúbené big data technológie.

Technológie big data rozdeľujeme do dvoch základných skupín :

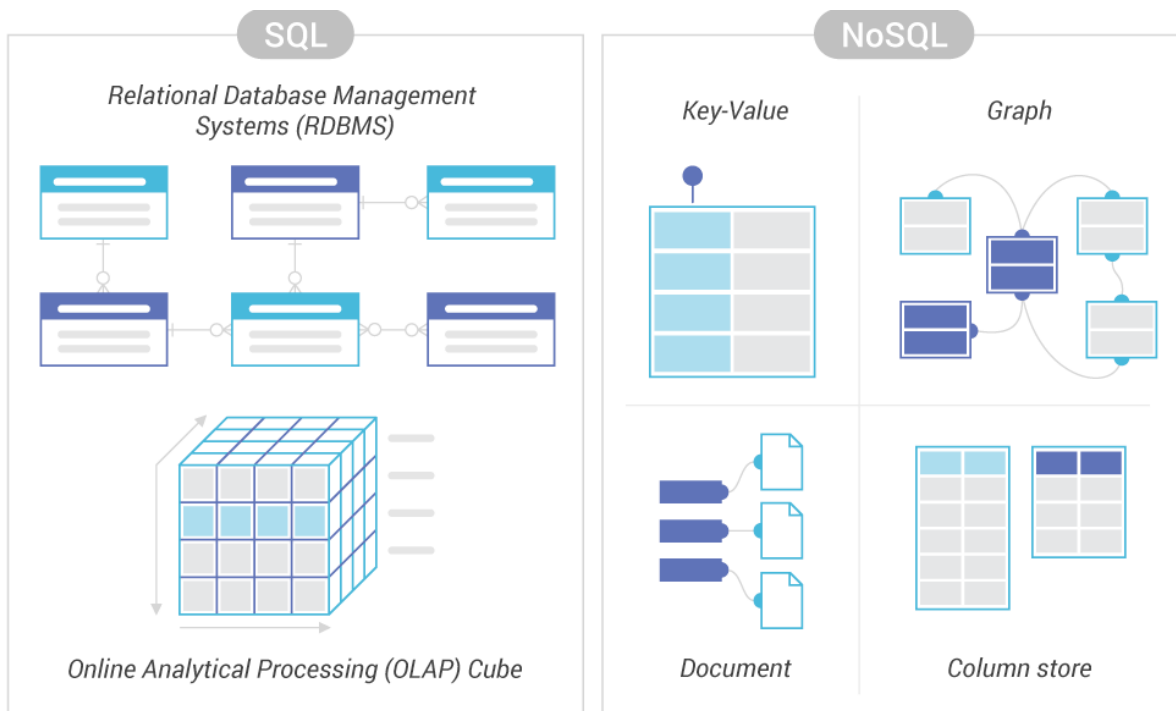
- Operatívne – tieto technológie poskytujú možnosti pre správu údajov, ktoré sú zozbierané na dennej báze, ide napríklad o informácie zhromaždené prostredníctvom sociálnych sietí alebo transakcie uskutočnené v online priestore
- Analytické – ide o pokročilejšie technológie, ktoré spracúvajú údaje, ktoré sú kľúčové pre obchodné rozhodnutia, napríklad predpoveď počasia alebo prognóza akciového trhu

1.7.1. NoSQL

Databázy, ktoré sú označené ako NoSQL, vystihuje už priamo ich názov, a teda „bez SQL“ alebo v niektorých prípadoch aj „not only SQL“. V týchto databázach sa pre prácu s dátami používajú iné prostriedky ako klasické schémy tabuliek, ktoré sú použité v rámci klasických relačných databáz typu SQL. Údaje, ktoré sú vkladane do klasických relačných databáz, majú presne definovaný formát a musia spĺňať presne stanovené požiadavky. Pre manipuláciu s týmito údajmi sa používa dopytovací jazyk SQL. Nerelačné databázy na rozdiel od relačných nepoužívajú žiadne tabuľkové schémy. Pre tento typ databáz nie je štandardizovaný ani žiadny dopytovací jazyk.[13]

NoSQL databázy sú obľúbené najmä pre ich možnosť manipulácie s neštruktúrovanými údajmi. Vyznačujú sa jednoduchým dizajnom, flexibilitou a škálovateľnosťou. Tento druh databáz nám predstavuje ďalšiu možnosť ukladania dát. Aj tieto NoSQL databázy sa vyvíjali v čase, a momentálne môžeme hovoriť o 4 základných typoch databáz NoSQL. Patria medzi nich databáza dokumentov, databáza grafov, databázy typu kľúč-hodnota a databáza stĺpcov. [27]

NoSQL databázu je vhodné použiť najmä v prípadoch, ak v rámci projektu potrebujeme ukladať štruktúrované a neštruktúrované údaje, pracujeme s obrovskými množstvami údajov alebo požadujeme rýchly agilný vývoj a škálovateľnú architektúru. Medzi najpopulárnejšie NoSQL databázy patria MongoDB, Apache Cassandra DB, Apache Hbase a Oracle NoSQL database.[13]



Obrázok 4 - Rozdiely medzi SQL a NoSQL databázou – zdroj: [27]

1.7.2. Apache Hadoop

Apache Hadoop je považovaný za jeden z najlepších nástrojov, ktorý možno použiť na ukládanie alebo spracovanie veľkých súborov naprieč jednotlivými klastrami počítačov za využitia jednoduchých programovacích modelov. Zaradzujeme ho medzi open source frameworky a je vyvíjaný, aby poskytoval distribuované, spoľahlivé a škálovateľné výpočty. Tento softvér je navrhnutý tak, aby sa škáloval od jednotlivých serverov až po tisíce zariadení, pričom každý z nich ponúka úložisko a lokálny výpočet. Samotná knižnica tohto softvéru je navrhnutá tak, aby dokázala zistiť popri prípade vyriešiť zlyhanie na aplikačnej vrstve, čím sa zabezpečuje vysoká dostupnosť služby.[12]

Skladá sa zo štyroch modulov:

- Hadoop Distributed File System – je to distribuovaný systém, bežiaci na štandardnom hardvéri, v porovnaní s tradičnými súborovými systémami nám zabezpečuje lepšiu dátovú priepustnosť.
- MapReduce – tento framework zabezpečuje pomoc pre jednotlivé programy pri vykonávaní distribuovaného a paralelného spracovania. Názov ukrýva dve samostatné úlohy a to Map a Reduce. Prvá úloha spočíva v preberaní a načítaní vstupných údajov ku ktorým sa následne vygenerujú páry kľúč-hodnota. Následne sa tieto páry privedú do redukčnej úlohy, kde sa agregujú údaje párov do menšej sady hodnôt, z ktorej je produkovaný výstup.

- Yet Another Resource Negotiator (Yarn) – Zabezpečuje správu a monitorovanie uzlov klastra a rovnako tak aj využitie jednotlivých zdrojov. Taktiež plánuje jednotlivé úlohy.
- Hadoop Common – rôzne knižnice, ktoré sú použiteľné v rámci všetkých modulov

Od prvého predstavenie softvéru Hadoop, sa celý tento ekosystém rozrástol a to najmä vďaka svojej ľahkej rozšíriteľnosti. Pod pojmom Hadoop si dnes vieme predstaviť celé množstvo rôznych nástrojov a aplikácií, ktoré nám dokážu pomôcť so zhromažďovaním, ukladaním, spracovaním a analyzovaním big data. Ako príklad si môžeme uviesť napríklad Apache Spark, HBase, Presto, Zeppelin a mnoho ďalších.[12]

1.7.3. Apache Spark

Apache Spark je rovnako ako Hadoop open source softvér. Je schopný bežať samostatne alebo na cloude. Na ukladanie a spracovanie údajov používa pamäť RAM a je navrhnutý tak, aby dosahoval čo najvyšší výkon. Spark dokáže vykonávať rôzne typy úloh, či je to dávkové spracovávanie, spracovanie v reálnom čase, výpočty, grafy a rôzne iné interaktívne dopyty. Cieľom vývoja Apache Spark bolo zlepšenie efektivity MapReduce. Skladá sa z piatich hlavných komponentov :

- Apache Spark Core – zabezpečuje plánovanie, odosielanie, vstupné a výstupné operácie, stoja na ňom všetky ďalšie funkcionality.
- Spark Streaming – pomocou tohto komponentu je zabezpečené spracovávanie živých dátových tokov.
- Spark SQL – tento komponent je využívaný najmä na zhromažďovanie informácií o štruktúrovaných údajoch.
- MLlib(Knižnica strojového učenia) – v tejto knižnici sa nachádzajú rôzne algoritmy strojového učenia.
- GraphX – používa sa na uľahčenie riešenia úloh v rámci grafovej analýzy. [14]

1.7.4. Tableau

Tableau patrí medzi jeden z popredných nástrojov na vizualizáciu údajov, ktorý je využívaný na analýzu údajov a business intelligence. Spoločnosť Gartner ho v minulosti klasifikovala ako jedného z lídrov vo svojom obore. Aj keď túto pozíciu pomaly stráca, stále patrí k jedným z najlepších možností. Slúži hlavne na vytváranie rôznych grafov, máp, tabuliek a analýzu údajov a následnú vizualizáciu. Medzi základné výhody, ktoré tento nástroj ponúka patrí najmä jeho jednoduchosť a intuitívnosť najmä pre nových

používateľov. Nie je potrebné mať žiadne predchádzajúce skúsenosti s programovaním. Medzi ďalšiu výhodu môžeme zaradiť schopnosť pripojiť sa k viacerým zdrojom dát, ktoré iné nástroje nepodporujú. [11]

V rámci prvej kapitoly sme zhrnuli základné poznatky o big data, ich výhodách a nevýhodách. V krátkosti sme si zhrnuli históriu a nástroje, ktoré je možné použiť v rámci konceptu big data. Po zhrnutí všetkých týchto tém, ktoré sme charakterizovali, sme usúdili, že tento koncept stále ešte nedosiahol svoje maximum, stále sa vyvíja a preto sme sa rozhodli si za cieľ našej práce zvoliť identifikáciu podstatných trendov a výziev, ktorým budú spoločnosti čeliť v prípade ak sa rozhodnú pre transformáciu svojej spoločnosti na dátami riadenú spoločnosť.

2. Cieľ práce

V rámci našej diplomovej práce sme si zvolili za primárny cieľ práce identifikáciu trendov v oblasti big data a zhodnotenie možnosti implementácie v rámci spoločnosti vo svete a na Slovensku. Aby sme dokázali naplniť hlavný cieľ práce, museli sme najprv splniť viacero čiastkových cieľov, ktoré sme si stanovili.

Prvým takýmto čiastkovým cieľom bolo získanie prehľadu v danej oblasti. V rámci tohto čiastkového cieľa sme si zhrnuli situáciu na Slovensku a vo svete, aby sme boli schopní lepšie pochopiť danú problematiku. Popísali sme čo to vlastne big data sú, ako sa odlišujú od normálnych dát a aké rôzne typy dát poznáme. Taktiež sme si zhrnuli možné výhody, ktoré nám dokáže táto oblasť priniesť v rámci spoločností a tiež s akými problémami a výzvami sa jednotlivé spoločnosti musia zapodievať v súvislosti s big data.

Ako ďalší čiastkový cieľ sme si zvolili prieskum a zanalyzovanie niektorých najpoužívanejších nástrojov v danej oblasti.

Po naplnení týchto čiastkových cieľov sme sa pozreli na situáciu, v ktorej sa momentálne nachádzajú spoločnosti a identifikovali sme výzvy, ktoré ich najviac ovplyvňujú pri transformácii svojej spoločnosti na dátami riadenú spoločnosť.

Ďalším čiastkovým cieľom, ktorý sme si zvolili bola identifikácia trendov v oblasti big data, v rámci tohto čiastkového cieľa sme identifikovali trendy v rôznych oblastiach, ako napríklad v oblasti technológie, ochrany osobných údajov a potreby nových zamestnancov a nových pozícií na základe výsledkov z prieskumov.

Za posledný čiastkový cieľ sme si zvolili zmapovanie situácie v rôznych sektorov. V rámci tohto čiastkového cieľa sme si zmapovali stav a možnosti implementácie identifikovaných trendov vo svete a na Slovensku.

Splnením týchto čiastkových cieľov sme dokázali naplniť náš primárny cieľ a na základe trendov a výziev, ktoré sme identifikovali, sme boli schopní ponúknuť naše odporúčania, pre spoločnosti.

3. Metodika práce a metody skúmania

Na splnenie nášho primárneho cieľa a teda identifikáciu trendov v oblasti big data, ktorý sme si rozdelili na viacero čiastkových cieľov, sme použili zahraničnú literatúru dostupnú v online knižniciach. Nakoľko táto oblasť sa rýchlo vyvíja a je aktuálnou, je k dispozícii dostatok literatúry. V niektorých prípadoch má tento rýchly rozvoj v danej oblasti za následok neaktuálnosť niektorých informácií, ktoré boli zverejnené v nedávnej minulosti. Ako ďalší dôležitý zdroj sme použili prieskumy trhov a rozhovory s poprednými odborníkmi v danej oblasti. Taktiež sme sa zamerali aj na štúdie, ktoré boli vykonané priamo v spoločnostiach a poskytli nám dôležité dáta pre splnenie nášho primárneho cieľa. Okrem tejto literatúry sme použili aj dáta z prieskumov, ktoré boli vykonané a to konkrétne z prieskumu od spoločnosti Capgemini „The data-powered enterprise“ [15] a od spoločnosti NewVantagePartners „Data and AI leadership executive survey 2022. Dáta z týchto prieskumov sme spracovali do grafov, porovnali s dátami z prieskumov z minulých rokov a snažili sa odhadnúť, ako sa bude vyvíjať daná oblasť v budúcnosti. Na dosiahnutie našich čiastkových cieľov a následného naplnenie primárneho cieľa sme použili komparačné metódy, kde sme sa snažili porovnávať aktuálny stav a stav danej problematiky v roku 2018 a na základe dedukciou zistených skutočností sme predikovali budúci vývoj.

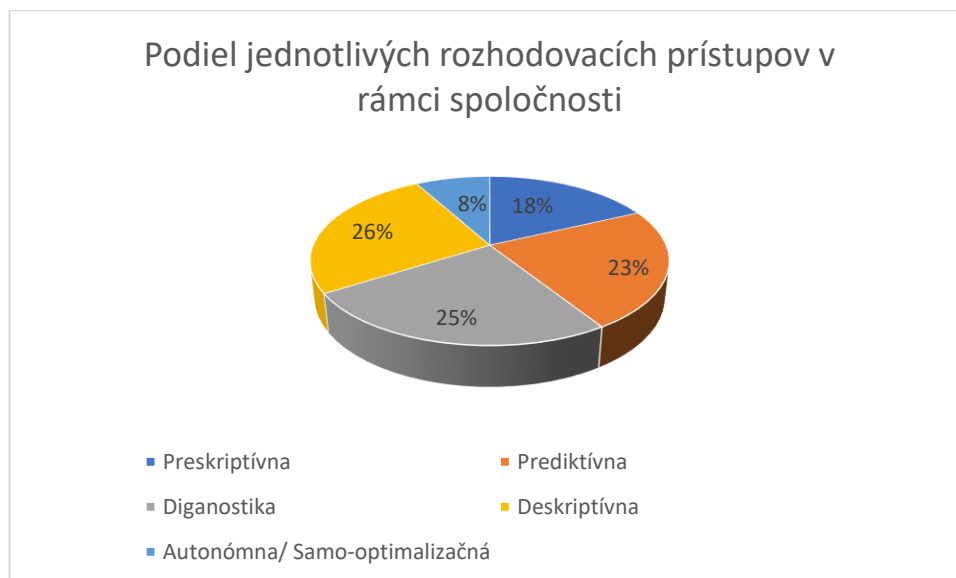
4. Výsledky práce

4.1. Zhodnotenie aktuálnej situácie a identifikácie výziev pri transformácií spoločnosti na dátovo riadené spoločnosti

V rámci tejto podkapitoly sme zhodnotili výzvy, ktoré spoločnosti musia prekonať, v prípade ak sa chcú stať spoločnosťami založenými na dátach. Na identifikáciu týchto výziev sme spracovali výsledky z dvoch prieskumov a to konkrétne z prieskumu od spoločnosti Capgemini „The data-powered enterprise“[15] a od spoločnosti NewVantagePartners „Data and AI leadership executive survey 2022“[16]. Jednotlivé výsledky prieskumov sme spracovali do grafov a identifikovali sme situáciu a výzvy, ktorým podniky čelia pri implementácii big data riešení.

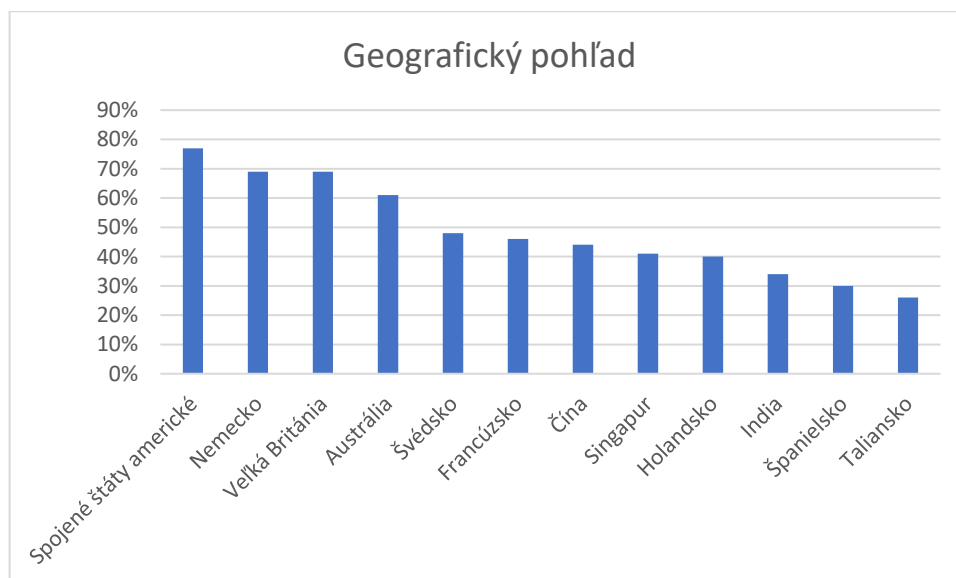
Spoločnosti po celom svete kladú čoraz väčší dôraz, aby sa jednotlivé ich rozhodnutia konali na základe údajov. Vzhľadom na rýchlo vyvíjajúce sa IT prostredie, rastúci počet zákazníkov a aktívne konkurenčné prostredie, musia jednotlivé spoločnosti dbať na každý krok, ktorý spravia. V tomto prostredí je už momentálne nevyhnuté využívať rôzne nástroje na analýzu údajov nakoľko práve to sa stáva kľúčom k úspechu.

Podľa prieskumu z roku 2018 [15], môžeme vidieť, že iba 38 % spoločnosti sa rozhodovalo skôr na základe dát ako na základe intuícii. V roku 2020 nastal prvý väčší zlom a pokrok, ktorý jednotlivé spoločnosti urobili, pretože pochopili, že práve dátová analytika im môže priniesť úplne nový pohľad, ktorý dovtedy prehliadali. Ak chcú spoločnosti vychádzať z presnejších predpovedí, ktoré dostanú na základe dát, je nutné aby sa transformovali. Najmä v roku 2020, keď sa prvýkrát objavila pandémia COVID-19 si spoločnosti začali uvedomovať, že potrebujú čoraz viac prijímať inovatívne spôsoby, aby boli schopní vytážiť maximum z externých a interných údajov, ktoré majú k dispozícii. V rámci prieskumu v roku 2020 sa vyjadrilo až 50 % spoločností, že sa riadia na skôr základe dát, ktoré majú k dispozícii ako na základe intuície [15]

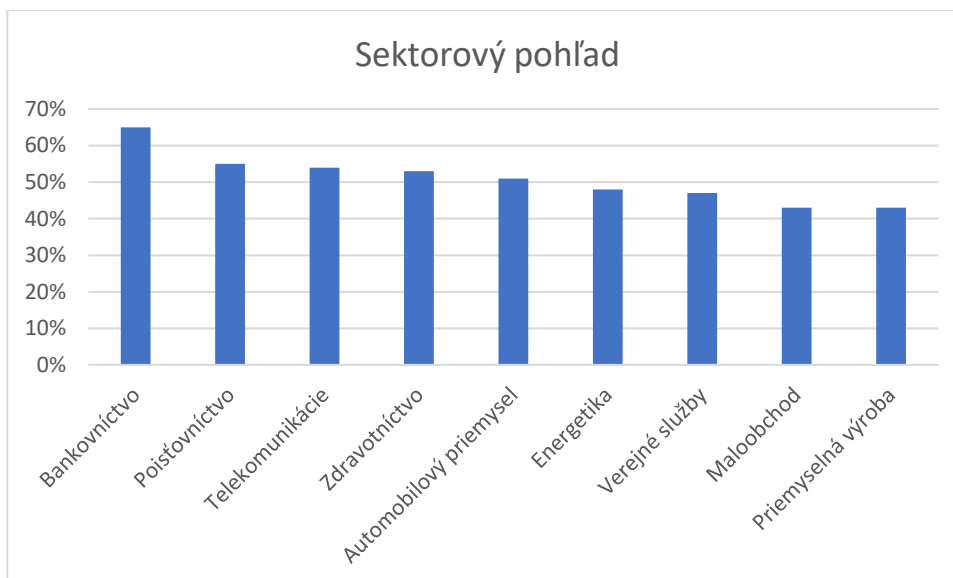


Graf 1-Podiel rozhodovacích prístupov na rozhodovaní v rámci spoločností – zdroj: vlastné spracovanie; údaje extrahované z [15]

V otázke, ktorá sa zaoberala podielom jednotlivých rozhodovacích prístupov v rámci spoločností sa nám potvrdilo, že väčšia časť rozhodovania zostáva stále reaktívna, teda vychádza z toho, čo sa stalo v minulosti, respektíve prečo sa to stalo. Na druhej strane spoločnosti uviedli, že 23% času sú využívané prediktívne prístupy, teda to čo sa môže stať v budúcnosti. Preskriptívne prístupy sú využívané 18% z času, ide o rôzne odporúčania na zlepšenie a len 8% času sú používané autonómne prístupy, ako sú systémy a procesy, ktoré dokážu pomôcť používateľom s ich rozhodnutiami na základe vopred zvoleného cieľa.

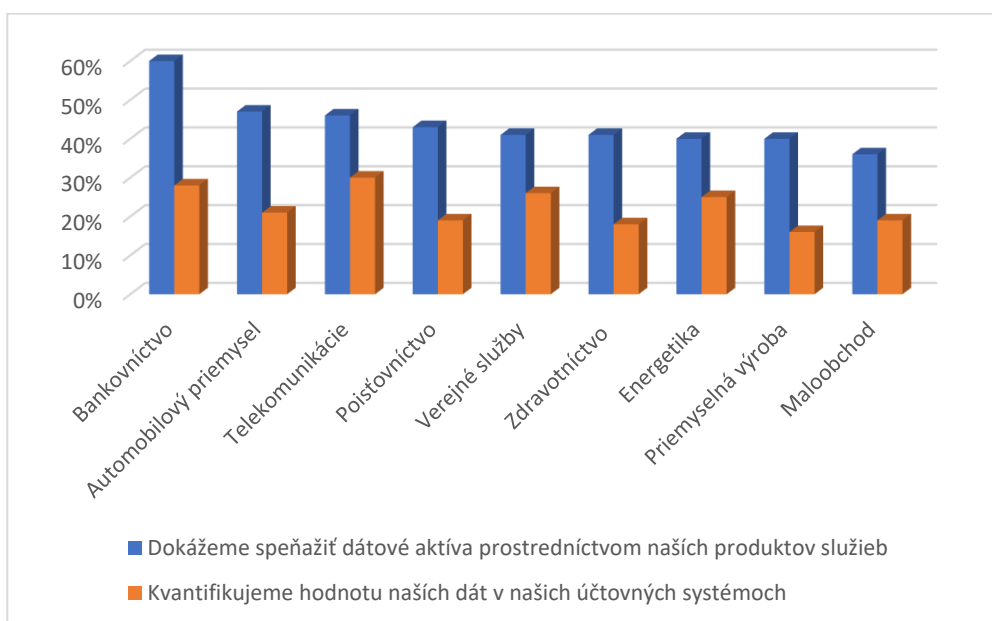


Graf 2-Geografický pohľad na spoločnosti, ktoré uviedli že ich rozhodovanie je celé postavené na dátach – zdroj: vlastné spracovanie; údaje extrahované z [15]



Graf 3-Sektorový pohľad na spoločnosti, ktoré uviedli že ich rozhodovanie je celé postavené na dátach – zdroj: vlastné spracovanie; údaje extrahované z [15]

V rámci prieskumu spoločnosti Capgemini [15], ktorého sa zúčastnili spoločnosti z rôznych sektorov a rôznych geografických oblastí, sme vyhodnotili odpovede na otázku, či sú ich spoločnosti riadené dátami. Na základe grafov, ktoré sme vytvorili na základe odpovedí, môžeme konštatovať, že hlavne spoločnosti z bankovníctva, poisťovníctva a telekomunikácií, sú riadené dátami a už implementovali pokročilé riešenia big data. Podľa nášho názoru to nie je vôbec prekvapením, ale stále vidíme veľký priestor na zlepšenie aj v iných odvetviach. V rámci spoločnosti je dôležité, aby jednotliví manažéri začali brať dáta ako podnikové aktíva, nakoľko nimi aj sú. Tu však nastáva dôležitá otázka, ktorú je potrebné si položiť a tou je: Akú hodnotu majú tieto údaje a ako sa dajú speňažiť?



Graf 4-Prehľad sektorov, ktoré dokážu najefektívnejšie speňažiť svoje dáta a kvantifikujú hodnotu ich dát – zdroj :vlastné spracovanie; údaje extrahované z [15]

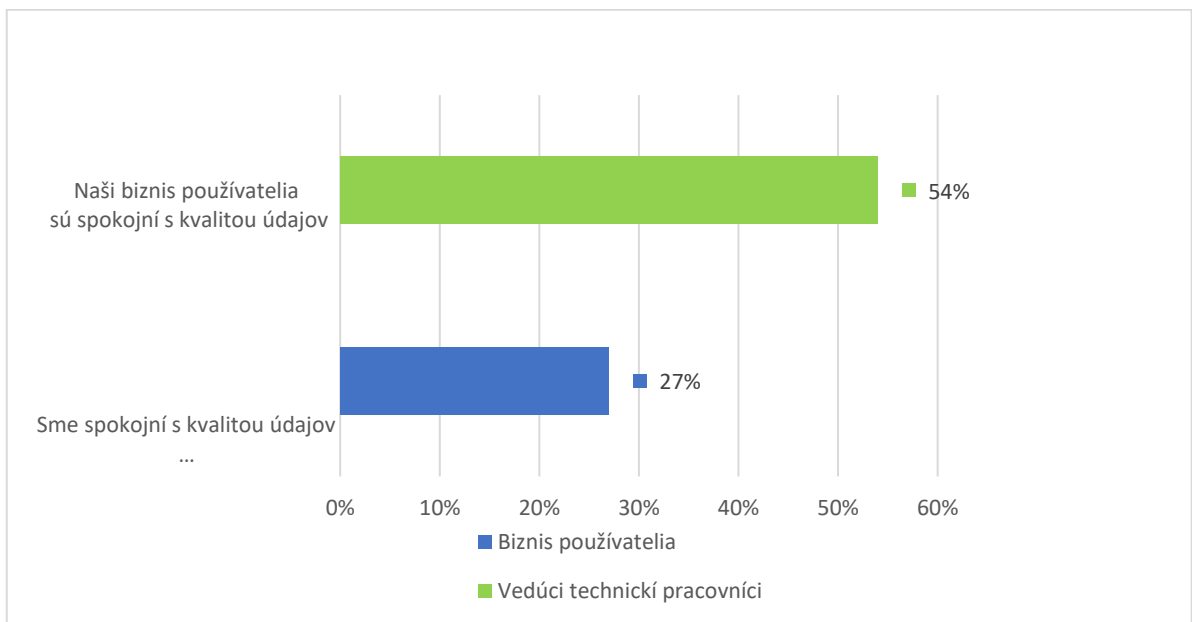
Ak sa pozrieme na výsledky, ktoré vyplývajú z odpovedí na otázku či dokážu spoločnosti speňažiť svoje dátové aktíva prostredníctvom svojich produktov a služieb, respektíve či dokážu kvantifikovať hodnotu údajov vidíme, že stále množstvo spoločností dáta nevníma ako aktívum a nedokážu kvantifikovať ich hodnotu vo svojich účtovných systémoch. Nie je to také jednoduché ako sa na prvý pohľad môže zdať, pretože ak chceme získať skutočnú hodnotu údajov, musíme sa zamerať na celkový pohľad, akým sú údaje aplikované priamo v praxi. Až vtedy dokážeme určiť ich skutočnú hodnotu. Na základe toho si vieme odvodiť jednoduchý vzorec, ktorý môžeme popísať ako „*dáta plus použitie rovná sa hodnota*“.

Ako príklad si môžeme uviesť spoločnosť Siemens, ktorá pochopila, že plánovaná údržba je častokrát omnoho efektívnejšia a menej nákladnejšia ako riešenie nečakanej poruchy, ktorá nastala. Všetko to prebieha na základe analytického modelu, ktorý dodáva predpovede potreby údržby, čím sú schopní eliminovať zbytočné neplánované prestoje. Na základe odpovedí z prieskumu spoločnosti Capgemini[15] vidíme, že len 43% spoločností, ktoré sa zúčastnili tohto prieskumu, dokáže prostredníctvom svojich produktov speňažiť dáta. Z tohto hľadiska môžeme vidieť, že vedie bankovníctvo a automobilový sektor. V praxi sa stretávame hneď s niekoľkými príkladmi ako sa bankovníctvo spája s maloobchodom, aby im dokázali speňažiť obrovské množstvo dát, ktoré má maloobchod k dispozícii. Napríklad klienti bánk Lloyds a Santander, môžu získať radu výhod u množstva maloobchodníkov. Táto spolupráca vznikla na základe pripojenia sa týchto bánk do digitálnej vernostnej schémy, ktorú prevádzkuje spoločnosť Cardlytics na základe údajov, ktoré získava od spoločnosti v USA. Táto spoločnosť využíva údaje o mňaní zákazníkov a na základe týchto údajov poskytuje týmto zákazníkom zľavy do týchto obchodov. Cardlytics takto získava dáta o správaní spotrebiteľov, ktoré má pomôcť prispôbiť ponuku. Banky na druhej strane dostanú z každého obchodu percentá z predaja. Na základe výsledkov konštatujeme, že iba malá časť spoločností dokáže speňažiť svoje poznatky a údaje.

Jedným z riešením tejto situácie, teda aby spoločnosti dokázali speňažiť svoje údaje, je včasný a odborný prístup k navrhovaniu ich procesov a produktov, ktoré budú schopné zachytávať nové údaje a premieňať ich na pridanú hodnotu, ktoré tieto údaje poskytujú. Pretože práve tieto nové údaje, môžu byť pre našu spoločnosť novým zdrojom príjmov.

Ako jednu z výziev pri transformácii považujeme pre spoločnosti kvalitu údajov. Kvalita údajov je jednou z ďalších kľúčových vlastností údajov. Ak sa zaoberáme kvalitou

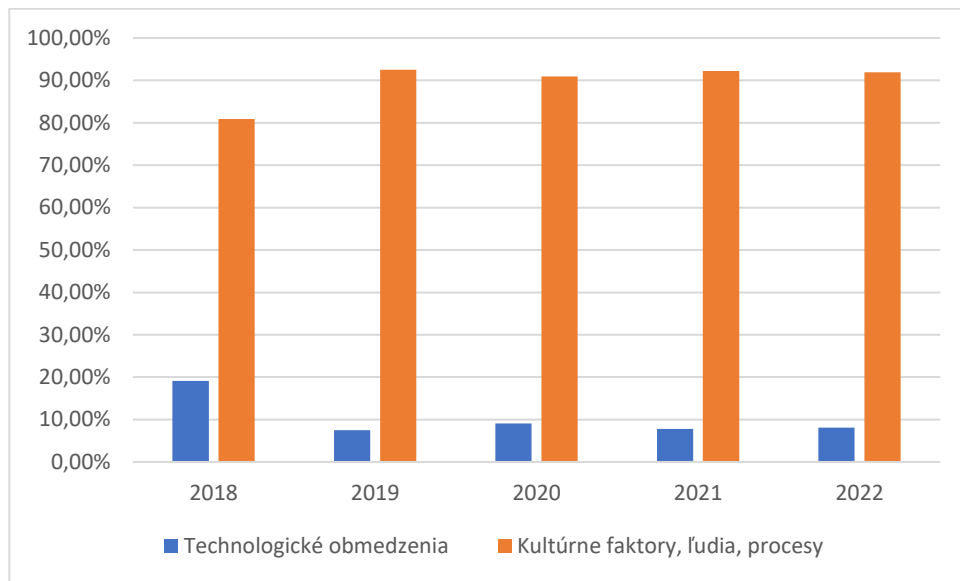
údajov, tak je potrebné tu hovoriť nielen o zdroji údajov, ale aj o problémoch, ktoré prichádzajú od biznis používateľov, ako sú napríklad problémy s presnosťou údajov resp. s pravdivosťou údajov. Je potrebné si dať pozor práve na túto kvalitu údajov, nakoľko podľa výsledkov štúdie MIT Sloan Management Review [31] dokážu spoločnosti stratiť 15 až 25 % svojich príjmov, tým, že pracujú s nepresnými údajmi. Tieto náklady vznikajú z dôsledku potreby opraviť rôzne chyby spôsobené zlou kvalitou údajov, stratou času na opätovné potvrdenie pravdivosti údajov a vysporiadaním sa ďalších chýb, ktoré sa často objavujú až časom. Preto by sa spoločnosti mali zaoberať tým, aby bola kvalita týchto údajov čo najvyššia, aby sa zbytočne nepredišlo rôznym problémom, ktoré môžu byť práve touto zlou kvalitou spôsobené. Ako vhodný spôsob navrhujeme hlbšiu analýzu zdrojových dát z ktorých jednotlivé spoločnosti zbierajú dáta a zlepšenie a zefektívnenie celého procesu dát ako takého.



Graf 5-Spokojnosť zamestnancov s kvalitou dát – zdroj: vlastné spracovanie; údaje extrahované z [15]

Ako však ďalej odhalil prieskum[15], vedúci technickí pracovníci si častokrát nie sú ani vedomí toho, že ich biznis používatelia nie sú spokojní s kvalitou údajov. Na základe výsledkov tohto prieskumu[15], ktoré sme popísali, môžeme konštatovať, že spoločnosti sa už začali zaoberať dátovou kultúrou a rok čo rok sledujeme ich väčší záujem rozvíjať túto dátovú oblasť priamo v ich spoločnosti. Až 92% spoločností uviedlo v rámci prieskumu spoločnosti Wavestone [16], že sa im investície do dát a umelej inteligencie vrátili, čo môžeme považovať za významný skok oproti roku 2017, kedy návratnosť týchto investícií dosahovalo len 48,1% opýtaných spoločností.[16]. Na základe týchto čísel konštatujeme, že spoločnosti si dávajú záležať na implementácii nových nástrojov a technológií v rámci

svojich organizácií, tak aby sa im investície dokázali vrátiť, avšak stále je pred nimi niekoľko výziev. Medzi najvýznamnejšie výzvy, ktorým ešte stále čelia jednotlivé spoločnosti patria hlavne kultúrne prekážky, nakoľko vykonávať zmeny v rámci spoločností nebýva nikdy jednoduché. Obzvlášť zložité to býva v rámci spoločností, ktoré majú dlhodobú históriu a nie vždy chcú prijať nové zmeny. Ako nám však už história viackrát ukázala, nakoniec ich toto odmietnutie častokrát stálo vedúce postavenie na trhu, a preto sú spoločnosti viac menej nútené sa transformovať.



Graf 6-Výzvy ovplyvňujúce spoločnosti – zdroj: vlastné spracovanie; údaje extrahované z [16]

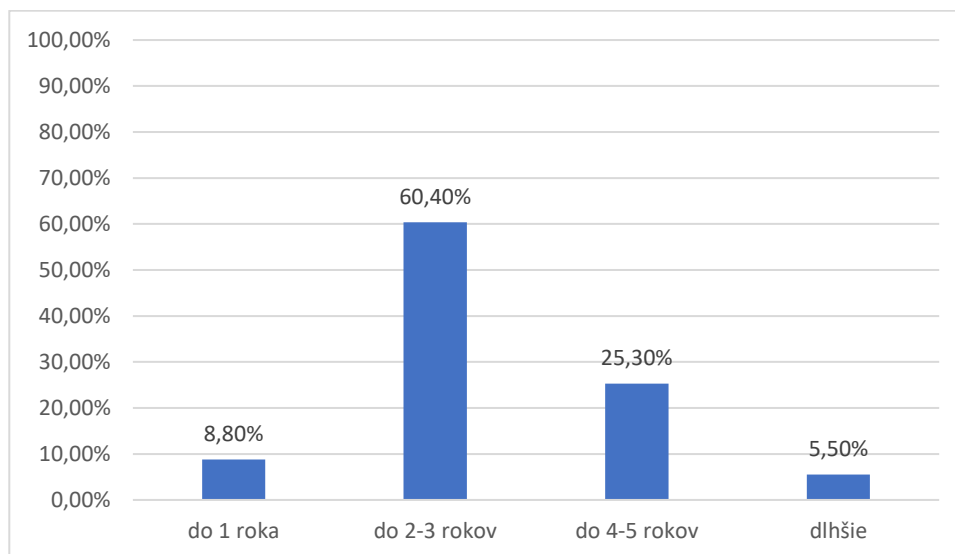
Na základe odpovedí na otázku, ktoré faktory najviac ovplyvňujú transformáciu spoločnosti, konštatujeme, že kým technologické obmedzenia sa v posledných rokoch minimalizovali a už nie sú až tak kľúčovou výzvou, kvôli ktorej by sa spoločnosti nemohli stať dátami riadenými spoločnosťami, tak kultúrne faktory a ľudia sú stále výzvou, ktorú iba veľmi pomaly minimalizujú. Medzi takéto kultúrne faktory patria napríklad zastaralé organizačné štruktúry, prirodzený rešpekt človeka voči novým zmenám a podobne. Preto najväčšou výzvou nie je ani tak implementácia rôznych riešení, ktoré sú v súčasnosti k dispozícii, ale ľudia ako takí. Kým tomuto konceptu neveria ľudia, nikdy nebude úspešný. Môžeme každý rok zavádzať najnovšie technológie, ale ak nebudú mať v nich ľudia dôveru, tak sa stráca ich pridaná hodnota, ktorú môžu spoločnostiam poskytnúť.

Ako ďalšie výzvy, s ktorými sa jednotlivé spoločnosti budú musieť vysporiadať, zaradujeme potrebu nových manažérov pre dáta a umelú inteligenciu. Títo manažéri však potrebujú mať okrem skúsenosti s technológiou a analytikou aj skúsenosti s organizačnými zmenami, ktoré mnohokrát treba spraviť, aby sa dátová kultúra dokázala udomáčniť v rámci celej spoločnosti a zvýšilo sa percento kľúčových používateľov, ktorí jej začnú dôverovať.

Z toho nám vyplýva aj ďalšia výzva a tou je, potreba prijímania takých ľudí, ktorí sa nebránia zmenám, podporujú rozhodovanie založené na dátach a nemajú strach zavádzať nové praktiky.

A ako poslednú výraznú zmenu, ktorú v rámci spoločností treba vykonať, by sme navrhovali väčšiu propagáciu úspechov, ktoré prinieslo rozhodovanie na základe dát. Pretože mnoho vedúcich pracovníkov im neverí najmä preto, že nemajú dostatok informácií, respektíve úspechy, ktorú boli dosiahnuté na základe takéhoto rozhodovania, sú iba na papieri a nie sú odprezentované vhodnou formou vedúcim pracovníkom danej spoločnosti. Po ústupe pandémie, predpokladáme na nasledujúce roky zrýchlený pokrok, ktorý musia spoločnosti vykonať najmä čo sa týka dátovej, digitálnej a AI transformácie, aby sa mohli stať plnohodnotnými spoločnosťami, ktorých rozhodovanie je založené na dátach.

Stať sa spoločnosťou, ktorá je založená na dátach, nie je otázkou pár týždňov, respektíve mesiacov, ale je to dlhodobý proces. Na základe odpovedí na otázku Ako dlho trvá transformácia na dátami riadenú spoločnosť z prieskumu [16], konštatujeme, že stále veľká časť spoločností má predstavu, že to dokáže zvládnuť v rámci horizontu 2-3 rokov. Nie je to nemožné, ale je k tomu potreba súhra viacerých faktorov. Medzi najdôležitejšie faktory patria kvalifikovaní ľudia a dostatok vyčlenených financií. Ak sa aj spoločnostiam podarí zabezpečiť tieto faktory stále ešte nemajú záruku, že ich transformácia bude úspešná. Nakoľko to nie je iba nejaké miesto, kam sa chceme dostať, ale je to dlhodobá cesta, po ktorej treba kráčať, aby sme sa dostali k vytúženému cieľu. Iba 30,8% spoločností má reálnejšie odhady, a predpokladá, že to zaberie viac času.

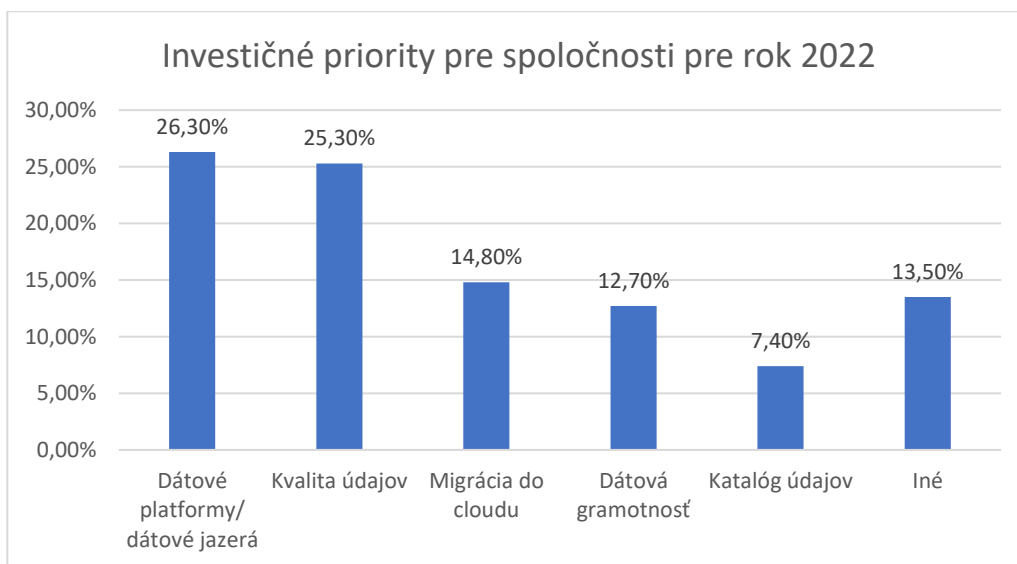


Graf 7-Predstava spoločností o čase potrebnom na transformáciu spoločnosti – zdroj vlastné spracovanie; údaje extrahované z [16]

Na základe zhodnotenia aktuálnej situácie s implementáciou dátovej stratégie musíme na záver tejto podkapitoly konštatovať jednotlivé výzvy, ktorým budú spoločnosti čeliť alebo už čelia pri transformácii ich spoločnosti na spoločnosť riadenú dátami.

- Spoločnosti musia pochopiť, že dáta sú pre nich podnikové aktíva a začať hľadať odpovede na otázku Akú hodnotu majú pre nich tieto údaje a ako sa dajú speňažiť?
- Zabezpečenie vysokej kvality údajov
- Odstránenie kultúrnych faktorov, úprava procesov a presvedčenie ľudí o dôležitosti kroku stať sa spoločnosťou založenou na dátach
- Pochopenie spoločností, že proces transformácie na spoločnosť založenú na dátach nie je krátkodobý, ale je potrebné mu venovať dostatok času

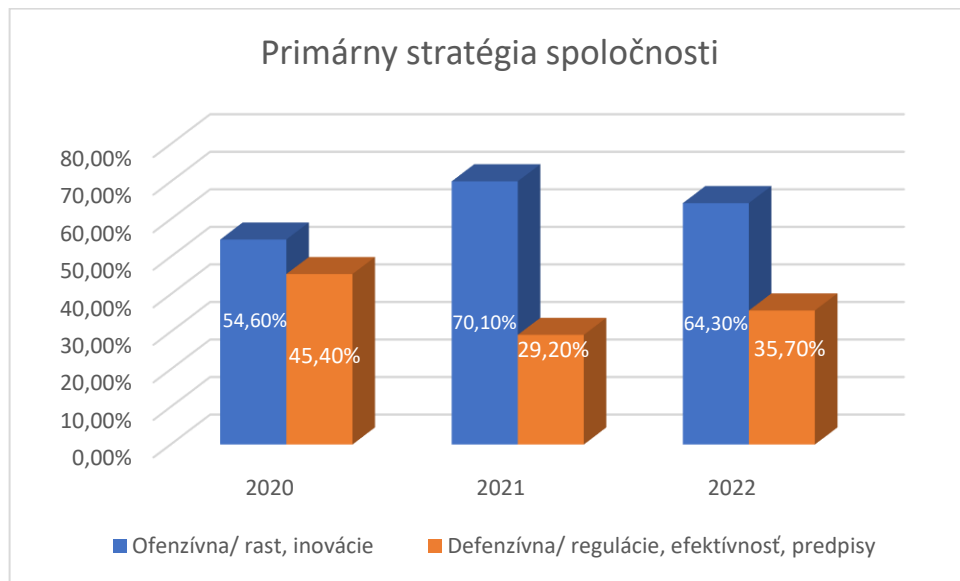
Ak sa pozrieme na stratégie spoločností a smer, ktorým by sa chceli uberať konštatujeme, že v rámci odpovedí, ktoré uviedli spoločnosti na otázku, čo bude tvoriť ich najväčšia investícia do dát v roku 2022, je ich prioritou rozvoj dátových platforiem a dátových jazier. A to najmä z dôvodu možnosti ukladania a spracovania väčšieho množstvo údajov. Avšak netreba zabúdať ani na kvalitu údajov, nakoľko nepresné údaje dokážu spôsobiť nemalé problémy. Treťou prioritou je migrácia do cloudu, nakoľko spoločnosti, nedisponujú takou výpočtovou kapacitou, ktorá by stačila na spracovania dnešného množstva údajov.



Graf 8-Investičné priority spoločnosti pre rok 2022 – zdroj: vlastné spracovanie; údaje extrahované z [16]

Ak sa pozrieme na ciele, ktoré spoločnosti chcú dosiahnuť v najbližších rokoch, v porovnaní s cieľmi, ktoré si určovali v minulých rokoch, tak z ich odpovedí[16] vyplýva jedna zaujímavá tendencia, a to hlavne v povahe cieľov, ktoré si stanovujú. Kým v minulosti to bolo viac vyrovnané, a spoločnosti dost' dbali aj na defenzívne stratégie, kde si za ciele

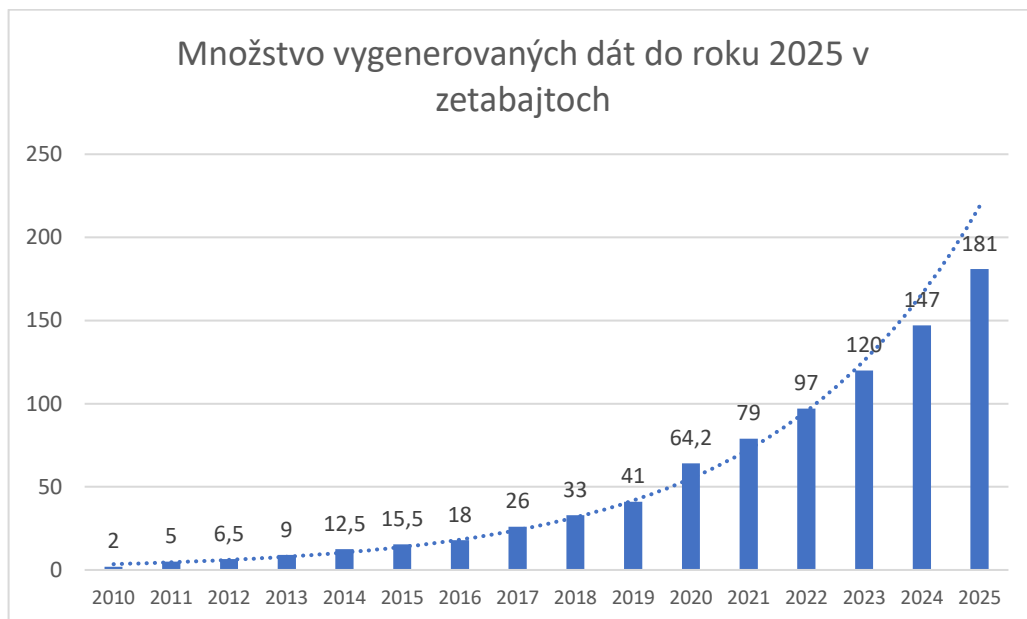
dávali reguláciu, efektívnosť a dodržiavanie predpisov, tak v súčasnosti si dávajú skôr ofenzívne ladené ciele ako rast, inovácie a nové praktiky.



Graf 9-Stratégie spoločnosti v oblasti zavádzania big data – zdroj: vlastné spracovanie; údaje extrahované z [16]

4.2. Identifikácia súčasných trendov

Ako sme si už opísali v prvej časti, množstvo dát, ktoré je denno-denne generované rastie z roka na rok. Na grafe si môžeme pozrieť vývoj množstva dát, ktoré sú generované. Na grafe sme si zobrazili štatistiku nárastu týchto dát od roku 2010 až po súčasnosť.



Graf 10- Množstvá dát v zetabajtoch, ktoré sú generované za rok a odhad do roku 2025 – zdroj: vlastné spracovanie; dáta extrahované z [17]

Na základe údajov ktoré sme spracovali v grafe sme zistili, že tento nárast je exponenciálny a množstvo dát, ktoré bolo vygenerované pred 10 alebo viac rokmi sa zvýšil niekoľkonásobne. V roku 2020 môžeme pozorovať prvý väčší nárast dát. Predpokladáme, že to bolo spôsobené príchodom pandémie COVID 19, nakoľko sa väčšina spoločnosti začala presúvať čoraz viac do online priestoru, či už to bolo v rámci práce, voľného času alebo štúdia. Ak sa pozrieme na trendovú krivku, tak môžeme vidieť, že do roku 2025 sa odhaduje nárast dát až na hodnotu 181 zettabajtov. Preto sa spoločnosti musia čoraz viac zaoberať analýzou týchto dát, ktoré im dokážu zaistiť výhodnejšiu pozíciu oproti konkurencii. Avšak zavádzanie takéhoto celého procesu, ktorý vykonáva zber, spracovanie a interpretáciu výsledkov trvá dlhšie obdobie, a je potrebné ho nastaviť priamo pre konkrétnu spoločnosť.

Podľa nášho názoru je najvyšší čas, aby sa analýzou dát začali spoločnosti zaoberať, ak chcú v najbližších rokoch čerpať výhody, ktorú im dátová analytika dokáže poskytnúť. S týmto nárastom dát sa však vynára aj jedna z ďalších výziev a tou je či je infraštruktúra v podnikoch pripravená na takéto navýšenie dát. Rovnako tak môžeme hovoriť aj o potrebe zabezpečenia nielen hardvéru, ale aj kvalitného softvéru, ktorý dokáže pomôcť spoločnostiam získať maximum z dostupných dát. Ak hovoríme o kvalitnom softvéri, tak spoločnosti by si mali zanalyzovať situáciu na trhu, nakoľko je dostupných množstvo možností od rôznych spoločností a vybrať tú, ktorá najviac vyhovuje ich požiadavkám. Netreba však zabúdať na náklady spojené s kúpou, resp. prevádzkou tohto softvéru.. Niekoľko takýchto nástrojov, ktoré môžu byť nápomocné sme si popísali v prvej kapitole.

Taktiež nesmieme zabúdať aj na nových kvalifikovaných zamestnancov, so znalosťami dátovej vedy. Na základe prieskumu spoločnosti QuantHub[32], z roku 2020, môžeme tvrdiť že je väčší dopyt po týchto kvalifikovaných zamestnancoch ako ponuka. Z tohto prieskumu vyplynulo, že mnohé pracovné portály ako LinkedIn, Glasdoor, Dice a Harnham, každoročne zverejňujú správy o raste dopytu a raste plátov práve pre dátových vedcoch. Na základe údajov z týchto pracovných portálov sa predpokladá, že v roku 2020 chýbalo na pracovnom trhu viac ako 250 000 dátových vedcov. Na základe týchto skutočností konštatujeme, že v najbližších rokoch môžeme očakávať významne investície do oblasti big data od spoločnosti, ktoré budú chcieť vyťažiť z nárastu dostupných dát maximum.

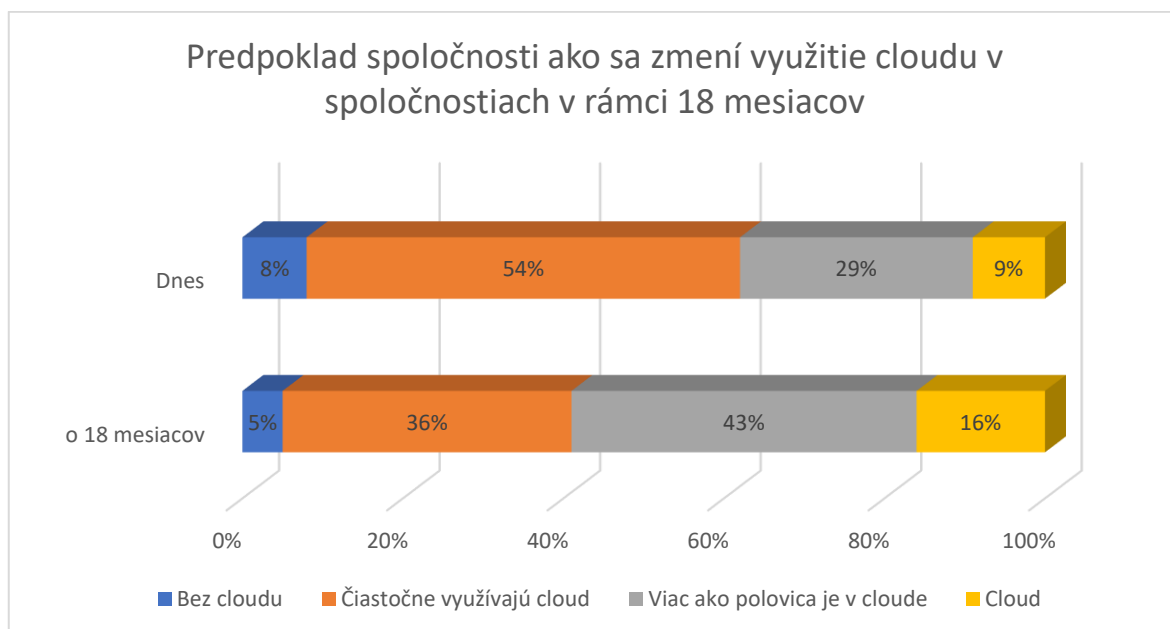
Spoločnosť Gartner predpokladá, že rozpočty IT porastú najrýchlejšie za posledných 10 rokov. Na rok 2022 sa predpokladá nárast výdavkov na IT o 5,5% oproti roku 2021. Nakoľko sa po pandémii COVID-19 začala presadzovať práca a vzdelávanie na diaľku a

očekávame, že tento nastavený trend bude ďalej pokračovať, spoločnosti budú potrebovať inovovať resp. investovať do nových zariadení na podporu takéhoto pracovného nastavenia.

S takýmto veľkým nárastom množstva dát, sa spoločnostiam neotvarujú iba nové možnosti z ktorých vedia niečo získať, ale čakajú ich aj výzvy, ktorým sa budú musieť postaviť a vyriešiť ich.

4.2.1. Cloud computing

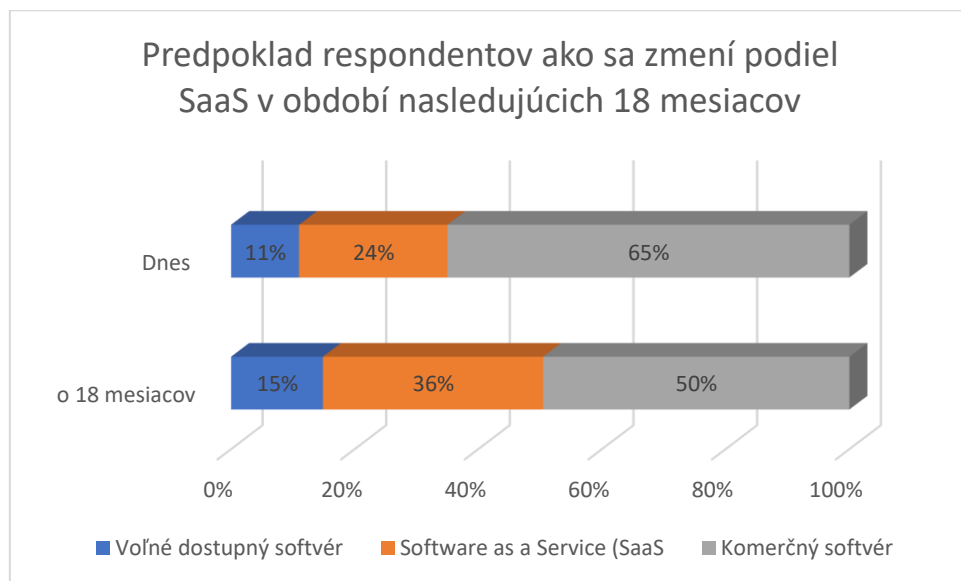
Predpokladáme, že na množstvo dát, ktoré bude k dispozícii už nebude stačiť iba infraštruktúra priamo v spoločnosti, nakoľko výpočtová kapacita nebude schopná spracovávať takéto množstva dát, ktoré sa predpokladajú v najbližších rokoch resp. prevádzkové náklady na takúto infraštruktúru, by boli pre spoločnosti neúnosne. Ako jedno z riešení, by podľa nášho názoru v najbližších rokoch malo čoraz viac spoločnosti prechádzať na cloudové možnosti, ktoré sú k dispozícii. Tieto riešenia dokážu spoločnostiam ušetriť nemalé prostriedky, či už na správu, prevádzkové náklady, a v neposlednom rade aj na zamestnancov, ktorých predtým potrebovali. Rovnako tak nesmieme zabúdať ani na čas, ktorý tým spoločnosti dokážu ušetriť, nakoľko im odpadne množstvo povinností ohľadom inštalácií. Aj keď mnoho spoločností, už začalo s touto technológiou, stále je tu dosť takých, ktorí sa vstupu do tejto oblasti bránia. Na základe prieskumu spoločnosti IDG [18], ktorý bol vykonávaný v roku 2020, a bolo do neho zapojených 551 IT spoločností., vieme povedať, že viac ako tretina výdavkov na IT predstavuje výdavky na cloud.



Graf 11-Predpoklad využitia cloud computingu – zdroj: vlastné spracovanie; údaje extrahované z[18]

V rámci tohto prieskumu[18] sa vyjadrilo až 59%, že plánujú mať väčšinu spoločnosti alebo celú spoločnosť v cloudu v priebehu nasledujúcich 18 mesiacov. Je to

nárast o 21 % oproti súčasnému stavu. Zaujímavým údajom je aj 92 % podiel spoločnosti, ktoré využívajú aspoň do nejakej miery cloud. V prieskume [18] sa tiež pýtali respondentov, koľko plánujú ich spoločnosti investovať v priebehu nasledujúcich 12 mesiacov na cloud computing. Respondenti uviedli priemernú investíciu na úrovni 73,8 milióna dolárov, čo predstavuje nárast o 59% oproti roku 2018. Z toho nám jasné vyplýva, že kľúčové spoločnosti, ktoré sa rozhodli ísť týmto smerom so sebou ťahajú aj ostatné, ktoré nechcú zaspáť a stratiť kontakt s najlepšimi. Priemerne to predstavovalo 32% celkového rozpočtu na IT. V rámci cloudu majú spoločnosti možnosť použiť platformu IaaS, ako napríklad Google Cloud Platform, Amazon Web Services alebo majú možnosť využiť SaaS priamo u poskytovateľa.

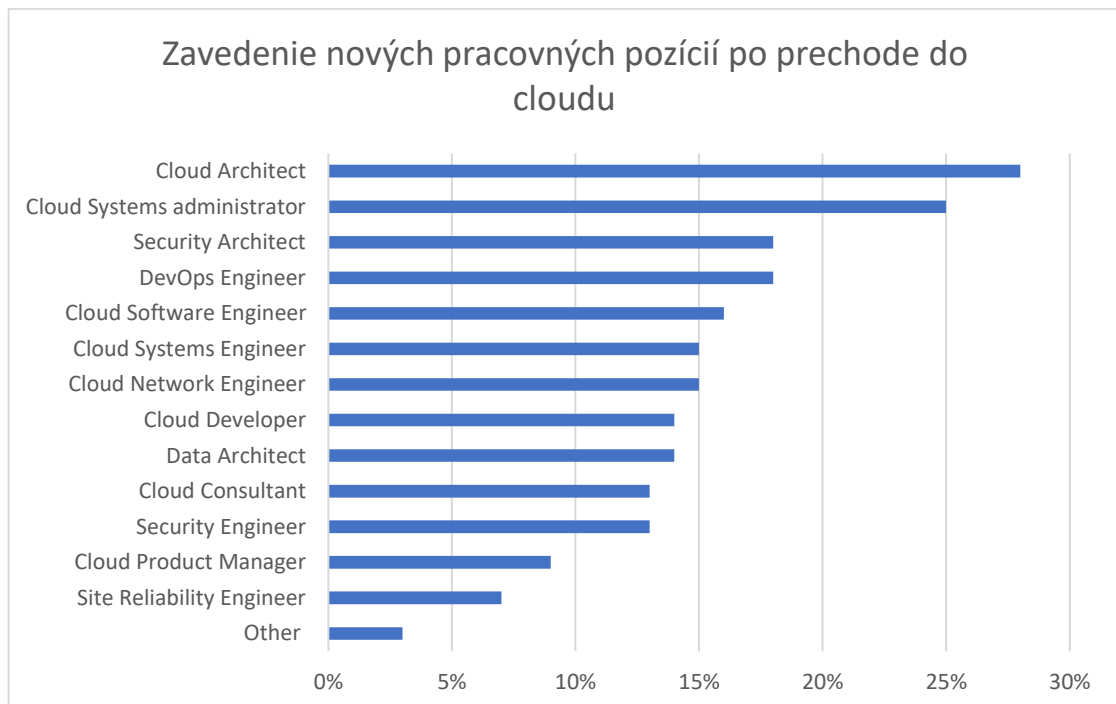


Graf 12-Predpokladaný vývoj SaaS na nasledujúcich 18 mesiacov – zdroj: vlastné spracovanie; údaje extrahované z [18]

Jednou z otázok, na ktorú odpovedali spoločnosti v rámci prieskumu[18] bolo aj to ako predpokladajú podiel využívania SaaS v rámci spoločnosti. Z výsledkov prieskumu môžeme vyvodit' tendenciu nárastu prechodu spoločnosti do cloud, či už ide o SaaS, kde môžeme vidieť nárast o 12% alebo, ak hovoríme o IaaS, kde bol nárast o 6% až na 48% oproti súčasným 42%. Ako výhodu použitia IaaS v rámci spoločnosti vidíme aj to, že aplikácia nemusí byť nutne vytvorená od nuly na cloudovej infraštruktúre, ale je tu možnosť premigrovať už existujúce aplikácie z vlastnej infraštruktúry. Ak sa pozrieme na výsledky prieskumu[18], tak až 54% aplikácií, ktoré momentálne bežia na cloude boli premigrované z vlastnej infraštruktúry. Avšak nie pre každú aplikáciu je vhodná migrácia na cloud, hovoríme hlavne o starších aplikáciách, ktoré neustále zaťažujú výpočtové úložisko. Tieto aplikácie nie sú väčšinou vhodnou voľbou na migráciu. Preto je potrebné si pred vstupom

do cloudu skutočne zväžiť, čo od danej technológie očakávame a či je skutočne pre našu spoločnosť cloud prínosom. Ak však naša spoločnosť je pripravená na prechod do cloudu a má dostatok kvalifikovaných ľudí, ako sú napríklad vývojári využívajúci výhody modernej aplikačnej architektúry, tak by spoločnosť mala rozmýšľať nad možnosťami, ktoré jej ponúka cloud. Hlavná výhoda je hlavne používanie mikroslužieb, ktoré sú následne zostavené do aplikácií. Súčasne je trendom spúšťanie týchto mikroslužieb vo vlastných kontajneroch. Kontajnery požadujú iba zlomok zdrojov a sú mimoriadne prenosné. Avšak nie je to len o výhodách, ktoré nám cloud prináša.

Jednou z hlavných výziev pre jednotlivé spoločnosti je podľa nášho názoru dostatočná kontrola nákladov vynaložených na cloud. Je potrebné mať prehľad o tom aké služby využívame v rámci cloudu a čo skutočne potrebujeme. Či náhodou nevyužívame niečo, čo je pre nás zbytočné, respektíve nemá pre nás význam a prínos alebo je už obsiahnuté v nejakej inej službe a zbytočne platíme dvakrát za to isté. Je potrebné sa zaoberať aj prieskumom jednotlivých cloudových služieb, ktoré sú ponúkané, aby sme dokázali zabezpečiť čo najlepšie cloudové služby pre našu spoločnosť. Netreba tiež zabúdať ani na nové pozície, ktoré je nutné obsadiť v rámci spoločnosti, čo nám môže spôsobiť ďalšie dodatočné náklady.



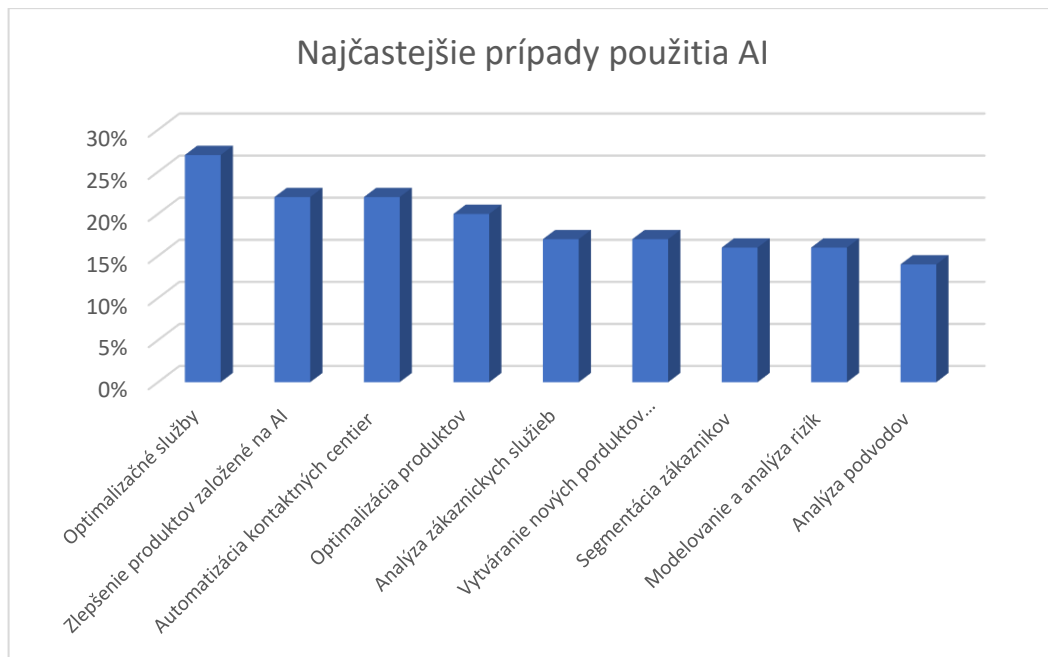
Graf 13-Zavedenie nových pracovných pozícií po prechode na cloud technológie – zdroj: vlastné spracovanie; údaje extrahované z [18]

V rámci prieskumu [18] tiež spoločnosti odpovedali na otázku, aké nové pozície boli nútené zaviesť v súvislosti s prechodom do cloudu. Na základe týchto odpovedí môžeme

konštatovať, že pre dosiahnutie čo najväčšieho prínosu z cloudu potrebujeme obsadiť aj niektoré kľúčové pozície, ako je napríklad cloud architekt, ktorý musí mať dostatok skúsenosti, aby dokázal zladať všetky potrebné poznatky a zabezpečil pre spoločnosť čo najvyššiu pridanú hodnotu. Ďalšou takouto kľúčovou pozíciou je bezpečnostný architekt s ktorým súvisí aj ďalšia výzva cloud computingu a to ochrana osobných údajov a celková bezpečnosť. Cloud je pred hrozbami oveľa menej zraniteľnejší ako klasické podnikové dátové centra. Dôležité je však aj správna konfigurácia a zabezpečenie cloudu. Ak sa na to pozrieme z celkového hľadiska a zvažíme si pre a proti stále nám cloud má čo ponúknuť, a teda náklady stým spojené stoja za pridanú hodnotu, ktorú nám ponúka. V rámci big data riešení nám cloud poskytuje možnosť mať miesto, kde sa získané dáta dokážu ukladať, spracovávať a analyzovať v reálnom čase. Veľkou výhodou cloudu je škálovateľnosť nakoľko zákazníkovi poskytuje možnosť priebežného platenia za využívanie. Bez využívania tohto cloud computingu by bolo len veľmi ťažké využiť celý potenciál, ktorý nám big data ponúkajú, nakoľko súčasné hardvérové vybavenie by nedokázalo spracovať také množstvo údajov. Big data tiež zohrávajú veľkú rolu pri vývoji cloud computingu, pretože hlavne ich zásluhou, stúpol dopyt po týchto cloudových riešeniach.

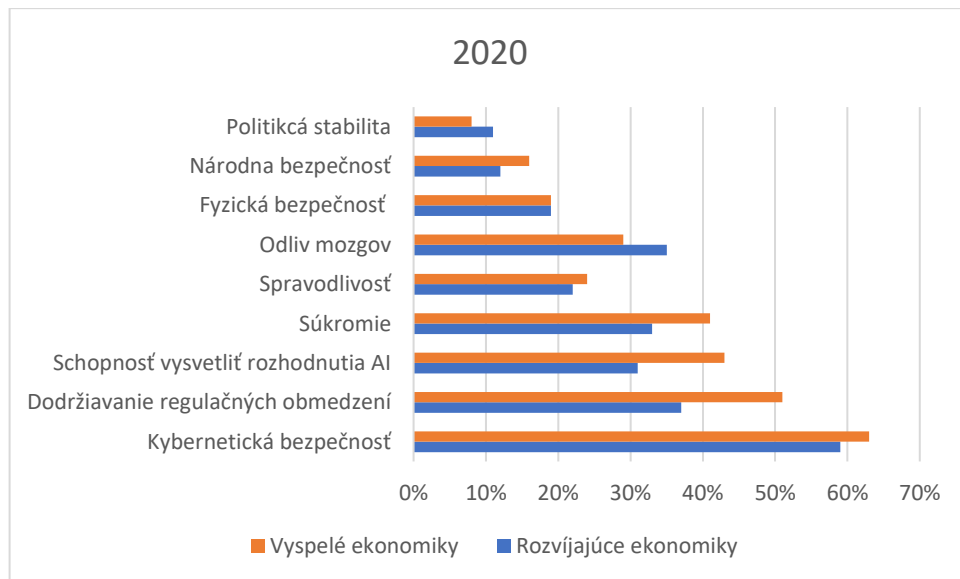
4.2.2. Umelá inteligencia

Po rokoch pandémie, kedy sme zaznamenali zníženie výdavkov do tejto oblasti, postupne zaznamenávame znovu trend nárastu investícií. Na základe prieskumu „State of AI in 2021“, ktorý bol vykonaný spoločnosťou McKinsey Analytics[26], vidíme, že aj tento rok pokračuje nárast prijímania umelej inteligencie, nakoľko až 56 % z respondentov vyjadrilo názor, že už používajú umelú inteligenciu v rámci svojej spoločnosti, čo je nárast oproti minuloročným 50 %. Ak sa pozrieme na geografické rozdelenie tak najvyššia miera využívania je v indických spoločnostiach a v Ázii.

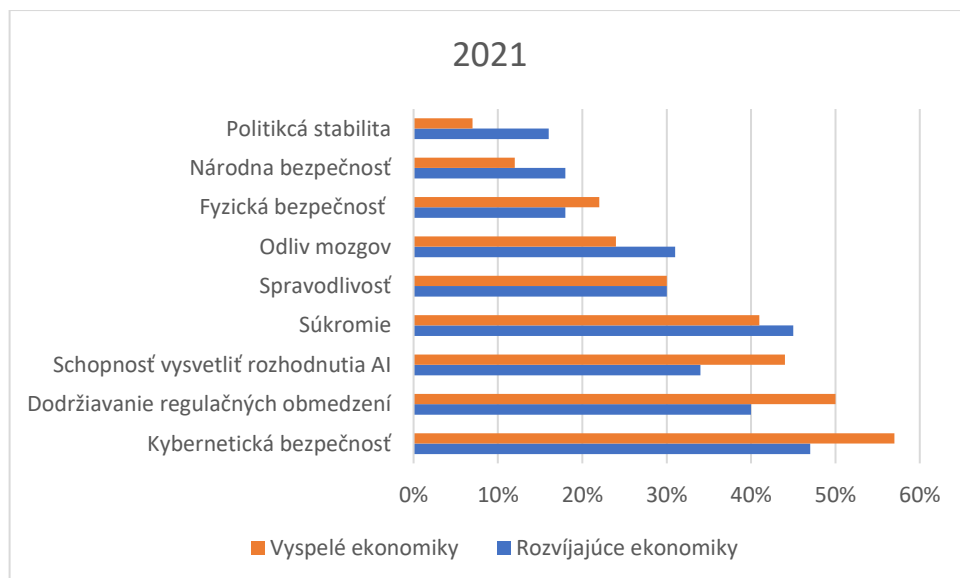


Graf 14-Najčastejšie prípady použitia AI- zdroj: vlastné spracovanie; údaje extrahované z [19]

Z odpovedí na otázku, ako v rámci ich spoločnosti využívajú umelú inteligenciu uviedlo najviac respondentov, že najväčšiu časť, kde využívajú umelú inteligenciu tvoria optimalizačné služby, vývoj produktov a služieb a automatizácia kontaktných centier. Hoci výnosy z AI v posledných rokoch nerástli respektíve klesali, hlavne čo sa týka oblasti dodávateľského reťazca, ktorá bola ovplyvnená hlavne pandémiou, tak respondenti uvádzajú, že AI im dokázala ušetriť nemalé náklady v každej z oblastí, ktoré uviedli v predchádzajúcej otázke. V rámci výziev pre oblasť umelej inteligencie, ktorými by sa mali spoločnosti zaoberať tu zaradujeme riadenie rizík, nakoľko aj väčšina respondentov uviedla, že stále majú v tejto oblasti priestor na zlepšenie. Medzi jedno z najväčších rizík, ktorým musia spoločnosti čeliť zaradili respondenti kybernetickú bezpečnosť.



Graf 15-Riziká ovplyvňujúce AI v roku 2020 – zdroj: vlastné spracovanie; údaje extrahované z [19]



Graf 16-Riziká ovplyvňujúce AI v roku 2021 – zdroj: vlastné spracovanie; údaje extrahované z [19]

Z prieskumu[19] vyplýva, že kybernetická bezpečnosť je stále považovaná za jednu z veľkých výziev, ktorej sa musia spoločnosti postaviť, aj keď ako vidíme z grafov oproti roku 2020 má táto oblasť klesajúci charakter. Na základe údajov z prieskumu, môžeme konštatovať, že vo vyspelých ekonomikách sa razantne nezmenili riziká medzi jednotlivými rokmi, počas ktorých bol prieskum vykonávaný. Ak sa pozrieme na dôvody, prečo spoločnosti nezmierňujú všetky relevantné riziká, tak môžeme konštatovať, že nemajú dostatok kapacít, aby dokázali pokryť všetky riziká, ktoré ich ovplyvňujú.

Ak sa na to pozrieme podľa typu ekonomík, tak môžeme konštatovať, že kým na najväčšom riziku, ktoré ich ovplyvňuje sa zhodnú, tak v prípade niektorých ďalších rizík, už môžeme vidieť nesúlad. Je to napríklad v prípade zabezpečenia súkromia, ktoré považujú

spoločnosti z rozvíjajúcich sa ekonomík za väčší problém ako spoločnosti z vyspelých štátov. Na základe výsledkov z tohto prieskumu[19] predpokladáme, že sa umelou inteligenciou začnú v najbližších rokoch zaoberať aj sektory, ktoré ju doteraz viac menej ignorovali a to aj napriek rizikám, ktoré so sebou prináša.

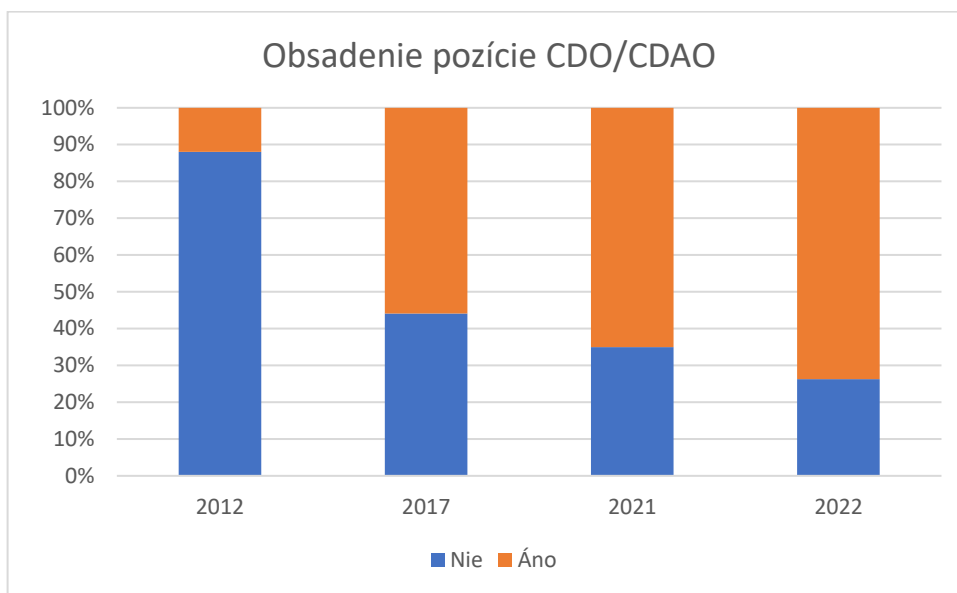
V rámci sektora dopravy predpokladáme ďalší rozvoj cieľa, na ktorom mnohé spoločnosti pracujú už dnes a už prešli dlhú cestu a to je autonómna doprava a schopnosť prepraviť ľudí z miesta na miesto. V prípade výroby vidíme cestu ešte vo väčšej miere zapájať robotov s rôznymi senzormi, ktoré dokážu zabezpečovať hladký chod zariadenia. Jednou z oblastí, kde predpokladáme najväčší prínos umelej inteligencie je zdravotníctvo a pomoc pri diagnostikovaní chorôb.

Je ťažké predpovedať čo bude o pár rokov, pretože táto oblasť sa vyvíja rýchlym tempom, nakoľko nielen spoločnosti ako Amazon, Google a Microsoft, ale aj vlády sa snažia investovať do tejto oblasti za účelom získania čo najväčšieho množstva výhod, ktoré im z toho môžu vyplývať.

4.2.3. Potreba nových zamestnancov a potreba nových pozícií

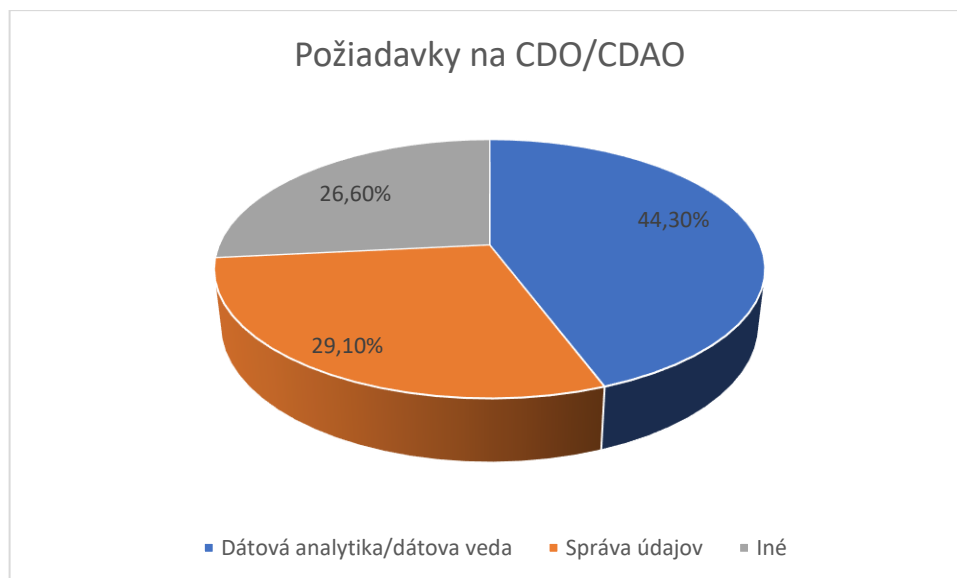
Za jeden z ďalších trendov môžeme pokladať stúpajúci dopyt po zamestnancoch s odborným vzdelaním v odbore dátová veda. Nakoľko v posledných rokoch sa dátovou analytikou zaoberá čoraz viac spoločností, je potreba aj množstvo kvalifikovaných odborníkov. Takýchto odborníkov však na trhu nie je dostatok.[32] Jednotlivé spoločnosti vytvárajú nové pozície, na základe nových potrieb, ktoré sú spojené najmä s implementáciou riešení big data v rámci podniku. Kým v minulosti neboli spoločnosti nútené obsadzovať tieto pozície a dokázali túto agendu pokryť aj internými zamestnancami, v súčasnosti sa táto oblasť vyvíja takým rýchlym tempom, že je nevyhnutné mať odborné vzdelanie a skúsenosti na to, aby naša spoločnosť dokázala dosiahnuť maximálne výsledky. V posledných rokoch sledujeme trend, že mnoho vysokých škôl aj v rámci Slovenskej republiky zaradilo medzi svoje programy dátovú analytiku, respektíve dátovú vedu, alebo aspoň rôzne predmety podporujúce tieto oblasti. V rámci tejto výzvy odporúčame spoločnostiam budovať si vlastné akadémie priamo v spoločnostiach, kde si dokážu vychovať zamestnancov priamo na mieru už počas vysokoškolského štúdia, a následne by z tejto akadémie prechádzali absolventi rovno do plnohodnotného pracovného pomeru. Nakoľko vidíme množstvo absolventov, ktorí aj po skončení školy majú problém nájsť si vhodnú pracovnú pozíciu, pretože nespĺňajú požiadavky, ktoré sú často prehnané. Aj na Slovensku sledujeme tento trend, že niektoré spoločnosti sa v posledných rokoch zamerali práve na tieto akadémie a rok čo rok sú schopní

produkovat' nových a nových ľudí vhodných priamo do praxe, ale stále je tu mnoho takých spoločností, ktoré to nepraktizujú a pred ktorými stojí výzva ako tento problém vyriešiť.



Graf 17-Obsadenie pozície Chief data officer/ Chief data analytics officer- zdroj: vlastné spracovanie; údaje extrahované z [16]

Jednou z kľúčových pozícií, ktoré musia spoločnosti obsadiť, ak chcú vyťažiť maximum z dát je pozícia Chief Data Officer. Ak sa pozrieme na výsledky prieskumu [16] na otázku obsadenie pozície CDO, tak vidíme že v roku 2012, malo túto pozíciu obsadených iba 12% spoločností. Ak to porovnáme v čase, tak môžeme vidieť prudký nárast obsadenia tejto pozície a to hlavne kvôli dôležitosti pri implementovaní a riadení celého procesu spojeného s dátovou analytikou v rámci spoločnosti. Výsledky z tohto roku len potvrdili tento trend v obsadzovaní tejto pozície, nakoľko až tri štvrtiny spoločností uviedli, že už majú obsadenú túto pozíciu. Pozícia Chief data analytics officer zahŕňa širší súbor zručností a vedomostí, kde kľúčovú rolu zohráva najmä znalosť dátovej vedy, ale taktiež nemenej dôležité je aj správa údajov. Avšak aj napriek potvrdzujúcej sa potrebe obsadiť aj túto novú pozíciu CDAO, viac ako polovica spoločností ešte vyčkáva s obsadením tejto pozície, nakoľko podľa nich je potrebné, aby táto pozícia preukázala to, že dokáže priniesť spoločnosti pridanú hodnotu.



Graf 18-Požiadavky na CDO/CDAO- zdroj: vlastné spracovanie; údaje extrahované z [16]

4.2.4. Bezpečnosť a ochrana osobných údajov

Ako sme už v predchádzajúcich kapitolách spomínali, big data nám dokážu priniesť rôzne výhody, dokážu spoločnostiam uľahčiť rozhodovanie, ale dokážu pomôcť vyriešiť aj zložitejšie problémy, ako sú rôzne prognózy chorôb a podobne. Napriek týmto výhodám, ktoré dokážu prinášať, je potrebné sa pozrieť aj na jednu z najväčších výziev, ktoré sa týkajú najmä súkromia a ochrany osobných údajov. Ako sa už v minulosti potvrdilo, spoločnosti si musia dávať pozor, nakoľko po posledných škandáloch, ako napríklad škandál spoločnosti Cambridge analytical spoločne s Facebookom počas volieb v USA, alebo škandál spoločnosti Netflix, keď poskytli údaje o svojich zákazníkoch, síce iba identifikačné číslo a hodnotenia, ktorí hodnotili rôzne filmy a poskytli ich ďalším účastníkom súťaže, ktorá bola zorganizovaná za účelom zlepšenia odporúčacieho algoritmu. Avšak tieto poskytnuté údaje o zákazníkoch podliehali ochrane osobných údajov.

4.2.4.1. Bezpečnosť z technologického hľadiska

Ak sa pozrieme na bezpečnosť z technologického hľadiska tak môžeme konštatovať, že najväčším problémom je objem dát a nedostatok technológií, ktoré by dokázali zaistiť bezpečnosť pre také množstvo dát. Preto je výzvou pre organizácie nájsť také riešenie, ktoré bude schopné zabezpečiť bezpečnosť takeého množstva dát a vysporiadať sa aj s ich rozmiestnením po celej sieti. Proces šifrovania dát, už nie je takým jednoduchým, nakoľko sa tento proces stáva zložitým, kvôli obrovskému objemu. Niektoré metódy šifrovania, ktoré vykazovali dobrý výkon pri práci s menšími objemami dát, nie je možné použiť pri práci s big data.[33]

4.2.4.2. Bezpečnosť z pohľadu ochrany osobných údajov

Spoločnosť je ovplyvňovaná viacerými vplyvmi z jej prostredia, ktoré na ňu vplyvajú zakaždým, keď sa pokúsi implementovať nové technológie. Každá spoločnosť, ktorá chce použiť množstvo dát, ktoré boli zozbierané na základe ich interakcie so zákazníkmi, musí zvážiť toto použitie aj z hľadiska prostredia, ktoré na ňu pôsobí a sem patrí aj vláda a rôzne vládne nariadenia. Nakoľko v poslednom období sa nahromadilo viacero škandálov, tak vlády jednotlivých krajín začali významne rozvíjať legislatívu, ktorou sa snažia ochrániť bežných spotrebiteľov, pred nechceným použitím dát o ich súkromí bez toho, aby o tom vedeli. V súčasnosti takou najvýznamnejšou legislatívou je GDPR, ktoré je nariadením EU o ochrane osobných údajov. Týmto nariadením sa vymedzili pravidla pre spracovanie osobných údajov fyzických osôb. GDPR sa vzťahuje na všetky podniky, ktoré nejakým spôsobom spracovávajú osobné údaje, či už manuálne alebo automatizovane. Toto nariadenie je platné pre všetky podniky spracúvajúce osobné údaje so sídlom v EU alebo pre spoločnosti, ktoré ponúkajú produkty, respektíve služby v rámci územia EU.



Graf 19-Najvyššie pokuty za porušenie GDPR v mil. eur do roku 2022 - zdroj: vlastné spracovanie; údaje extrahované z [21]

Na grafe si môžeme pozrieť najvyššie pokuty, ktoré boli udelené za porušenie GDPR od roku 2018 do Februára 2022. Na základe týchto údajov môžeme predpokladať, že aj keď je tu toto nariadenie už dlhšiu dobu, tak stále evidujeme dosť porušení tohto nariadenia. Podľa statistika [21] je to najmenej 946 porušení. Ak sa nato pozrieme tak väčšina vysokých pokút bola udelená v priebehu posledných 2 rokov a to z dôvodu, že spoločnosti mali dostatok času na to si uvedomiť, že to, čo bolo ticho tolerované v minulosti už im ďalej tolerované nebude a budú musieť zapracovať na dodržiavaní tohto nariadenia a nesnažiť

hľadať cesty ako toto nariadenie obchádzať. Pre spoločnosti môže byť prínosom tohto nariadenia najmä zjednotenie legislatívy v rámci celej Európskej únie, nakoľko všade v EU bude platiť jednotný súhrn pravidiel, čím ušetrí spoločnosti náklady na zapracovaní viacerých zákonov skrz jednotlivých krajín v ktorých pôsobia. Okrem ušetrných nákladov sa spoločnostiam zjednoduší aj komunikácia s kontrolnými úradmi nakoľko v rámci EU, bude nutné komunikovať iba s úradom na ochranu údajov. Navyše GDPR podporuje inovácie, pretože nie je technologicky závislý. Nakoľko všetko je o dôvere, a tú mnohé spoločnosti najmä v posledných rokoch svojimi opakovanými škandálmi stratili, nie je prekvapením, že z prieskumu, ktorý bol vykonaný v rámci EU až osem z desiatich ľudí má pocit, že nemá úplnú kontrolu nad svojimi údajmi. Šesť z desiatich ľudí neverí online podnikom a preto pokladáme GDPR ako vhodnú príležitosť pre spoločnosti, aby získali späť dôveru svojich zákazníkov. [20]

4.3. Možnosti implementácie identifikovaných trendov v rámci vybraných sektorov vo svete a na Slovensku

4.3.1. Poistovníctvo

Ak sa pozrieme na súčasný stav v rámci poisťovacieho sektora, tak momentálna situácia je taká, že v rámci tohto odvetvia sa formuje množstvo spoločnosti, ktoré sa pokúšajú transformovať podnikanie v tejto oblasti prostredníctvom využitia big data, AI a strojového učenia. Toto riešenie v rámci poisťovacieho sektora je nazývané Insurtech. Spoločnosti, ktoré sú na trhoch už niekoľko rokov, využívajú poväčšine klasické prístupy na rozdiel od nových spoločností, ktoré využívajú množstvo dostupných dát a prostredníctvom AI a strojového učenia dokážu poskytovať čo najviac personalizovanú ponuku pre svojich zákazníkov. Podľa správy spoločnosti GlobalData [29], ktorá bola zverejnená v apríli 2021, predpokladá, že príjmy platformy AI v rámci sektora poisťovníctva porastú do roku 2024 až o 24%. Spoločnosť Traffk, ktorá sa radi k jedným z lídrov v oblasti Insurtech si za cieľ na najbližšie roky dáva lepšie pochopenie rizík a modernizovanie procesov v rámci ohodnotenia poistenia, využitím moderných technológií a analytických nástrojov. Väčšina tradičných spoločností v rámci poisťovníctva, ešte stále používa klasické prístupy a spoliehajú sa na rôzne demografické údaje staré niekoľko rokov. Postupné zavádzanie AI spolu s využitím big data však v nasledujúcich rokoch dokážu vytvoriť novú éru, ktorá bude mať vplyv od distribúcie cez upisovanie až po konečnú cenotvorbu. Najmä čo sa týka distribúcie a upisovania, môžeme tvrdiť, že už aj v súčasnosti sú do veľkej miery ovplyvňované týmito novými technológiami. Spoločnosť Traffk vyvinula distribučnú a

upisovaciú platformu, ktorá dokáže využívať obrovské množstvo údajov a funkcií, aby sa im podarilo optimalizovať spotrebiteľské správanie, reklamácie, rôzne signály od spotrebiteľov a odporúčania produktov. Počas pandémie, sme mohli pozorovať zrýchlenie digitálnej transformácie, nakoľko množstvo spotrebiteľov bolo nútených využiť digitálne kanály. Ďalším takým významným cieľom, ktoré si spoločnosti stanovujú na najbližšie roky je vybudovať systém robo-poradenstva, ktoré by dokázalo poskytovať automatické poradenstvo s cenovými ponukami. Umožnilo by to osloviť aj skupiny, ktoré nemajú možnosti prístupovať k takémuto finančnému poradenstvu a bolo by to pre nich aj nákladovo efektívnejšie.

Odhaľovanie poistných podvodov

Na opísanie poistného podvodu sa najčastejšie používa definícia, že je to správanie, ktoré je vedené za účelom obohatenia sa na úkor poisťovne. Poisťovne majú svoje rokmi overené prístupy a metódy, ktorými sa snažia predchádzať týmto podvodom, ale v posledných rokoch sa čoraz viac zameriavajú na rôzne špeciálne analytické nástroje, ktoré im dokážu s týmto problémom pomôcť. Kľúčom k tomu aby dokázali tieto nástroje dosahovať čo najlepšie výsledky, je zabezpečenie prístupu k čo najväčšiemu množstvu údajov, či už interným ako sú dáta generované z uzatvárania zmlúv, likvidácie poistných udalostí alebo z iných poistných podvodov. Stále viac a viac sa do popredia dostávajú verejne dostupné dáta či už ide o oficiálne dáta ako sú rôzne obchodné registre a podobne alebo neoficiálne údaje ako sú napríklad údaje zo sociálnych sietí, diskusných fór a blogov.

Stanovenie podmienok na mieru

V rámci budúcich rokov predpokladáme, že sa spoločnosti v danej oblasti budú snažiť skrátiť čas, ktorý je potrebný na vybavenie poistenia. Jednou z možností ako to dokážu urobiť je implementovanie algoritmov AI, ktoré budú vytvárať rizikové profily zákazníkov pre spoločnosti, nakoľko big data poskytujú nové údaje, z ktorých budú spoločnosti lepšie schopné identifikovať tieto rizikové profily. Ako výhody tejto lepšej klasifikácie rizika vidíme najmä zlepšenie vzťahu s potenciálnymi zákazníkmi, ktorí by nemuseli byť ochotní zaplatiť za cenu vyššieho rizikového profilu, ktorý by odrážal širšiu populáciu. Bez využitia veľkých dát by nebolo možné vytvárať takúto klasifikáciu rizika. Každý jedinec za sebou v súčasnosti zanecháva nemalú digitálnu stopu, a preto by bola škoda tieto dáta nevyužiť. Na základe analýzy dát o potenciálnom klientovi, môže byť upravená konkrétna zmluva pre daného zákazníka. Množstvo ľudí v dnešnom svete využíva

rôzne nástroje na meranie svojich životných funkcií ako sú smart hodinky a rôzne iné zariadenia, a preto by poisťovne mohli zväziť poskytovanie rôznych výhod z poistenia za poskytnutie týchto údajov o ich zdravotnom stave.

4.3.2. *Bankovníctvo*

Umelá inteligencia patrí medzi jeden z kľúčových trendov, ktorý sa banky snažia implementovať. Svoje miesto nachádza nielen pri vyhodnocovaní údajov o zákazníkoch, ale aj priamo pri komunikácii a budovaní vzťahu so zákazníkom. V posledných rokoch, sledujeme trend znižovania počtu pracovísk, nakoľko banky už nepotrebujú prevádzkovať také množstvo malých pobočiek a zamestnancov. V prípade ak potrebuje zákazník pomôcť so správou svojich financií, tak vo väčšine prípadov už nie je nútený navštíviť priamo pobočku danej banky, ale dokáže si pomôcť prostredníctvom aplikácie. Medzi prvé banky, ktoré poskytli takúto možnosť patrí Bank of America so svojim poradcom nazývaným Erica, ktorý využíva dátovú analytiku a umelú inteligenciu. Prostredníctvom tohto poradcu sú zákazníci schopní lepšie pochopiť svoje financie, možnosti investovania a šetrenia a pomáha im to s lepším manažovaním svojich financií.[23]

Podobnú aplikáciu už používa viacero bánk aj na Slovensku ako napríklad Slovenská sporiteľňa, ktorá prevádzkuje svojho poradcu George. Avšak nie je to len o procesoch smerujúcich k zákazníkovi, ale umelá inteligencia je využívaná aj v rámci vnútorných procesov ako napríklad v singapurskej banke DBS, ktorá predstavila ako prvá agenta na prijímanie nových zamestnancov. K tomuto kroku ju viedlo to, že až 20 % času strávili zamestnanci zbieraním údajov o možných kandidátoch a rôznymi činnosťami týkajúcich sa výberového konania. Týmto ušetrila banka čas svojich zamestnancov, ktorý mohli venovať priamo rozhovorom a výberom z kandidátov.[24]

Aj na Slovensku môžeme v posledných rokoch sledovať práve takéto aktivity, ktorými banky chcú znížiť svoje náklady a na druhej strane zlepšiť svoj vzťah so zákazníkmi a získať konkurenčnú výhodu. Avšak stále majú pred sebou veľa oblastí, v ktorých sa potrebujú zlepšiť. Aj priamo Národná banka Slovenska, sa podieľa na tomto procese a to zriadením inovačného hubu a regulačného sandboxu. Je to z dôvodu množstva sťažností od spoločnosti, ktoré prichádzali s rôznymi inováciami a narážali na problémy. V rámci inovačného hubu je umožnené Fintech spoločnostiam, priamo komunikovať a viesť dialóg s expertami z rôznych oblastí NBS, ktorí im dajú lepšiu predstavu, ako správne implementovať svoje riešenia a vyhnúť sa rôznym problémom týkajúcich sa regulácií. Avšak nie každému je umožnené, použiť tento inovačný hub. Kľúčové otázky, ktoré rozhodujú o

zapojení do tohto hubu sú či bude daná služba odlišná od súčasnej ponuky iných účastníkov finančného trhu a či ponuka daná služba nejaké priame, respektíve nepriame výhody pre koncového spotrebiteľa .

Od prvého januára 2022 Národná banka Slovenska spustila aj takzvaný regulačný sandbox, ktorý umožňuje nastavenie finančnej inovácie v súlade s rôznymi reguláciami a následne ju otestovať priamo v praxi.[22] Okrem toho NBS, sa snaží prepojiť Fintech spoločnosti so slovenskými spoločnosťami, ktoré majú záujem o rôzne inovácie a preto vytvorila NBS databázu záujemcov o spoluprácu. Aj keď táto služba od NBS bola spustená iba od januára, už sa v databáze nachádza niekoľko takýchto subjektov, ktoré majú záujem o spoluprácu. Na základe údajov z tejto databázy je vidieť, že najväčší záujem je o oblasť umelej inteligencie, big data a cloudu. Výhodou je, že väčšina bánk je súčasťou väčších organizácií, ktoré pôsobia v rámci celej Európy a tak je zabezpečené, že aj v rámci Slovenska sa dostaneme k moderným riešeniam oveľa skôr ako v minulosti, kedy tieto riešenia neboli vždy samozrejmosťou na našom území a častokrát sa k nám dostali až po rokoch.

4.3.3. Zdravotníctvo

Big data a umelá inteligencia sú jednými zo základných pilierov, ktoré prispievajú k rýchlejšej transformácii. V poslednom období rastie dôležitosť internetu vecí, najmä čo sa týka množstva pripojených lekárskejších a nositeľných zariadení. Čoraz viac sa do popredia dostáva telemedicína, ktorá tu už niekoľko rokov je s nami, ale až pandémie COVID-19, ju dostala do takého stavu, aký sa od nej očakával už v začiatkoch. Pacienti, ale aj poskytovatelia zdravotnej starostlivosti nedôverovali veľmi tejto inovácii, stále radšej uprednostňovali priamu návštevu v nemocnici a strávili svoj čas radšej v preplnených čakárňach. Avšak, práve táto telemedicína dokáže jednak pacientom, ale aj poskytovateľom zdravotnej starostlivosti priniesť úsporu času, nákladov a zvýšenie pohodlia, najmä ak hovoríme o ľuďoch z odľahlých miest a obcí a starších pacientov. Táto telemedicína dokáže rapídne znížiť počet návštev u lekára a počet prípadov na pohotovosti. Rovnako tak je možné jej použitím sa skontaktovať s pacientom po prepustení a zmonitorovať jeho stav prostredníctvom video hovoru.

Ak sa pozrieme na túto telemedicínu v podmienkach Slovenska, stále vidíme veľké medzery, ktoré sú spôsobené jednak nedostatkom dôvery, keďže práve starší ľudia, ktorým by to dokázalo pomôcť najviac stále radšej navštívia lekára priamo. Ďalšou výzvou je zvýšenie digitálnych zručností najmä čo sa týka, týchto nových technológií, nakoľko na Slovensku je ešte stále dosť veľká časť populácie bez základných digitálnych zručností.

Slovenská spoločnosť Globallogic Slovakia sa podieľala na vytvorení digitálnej platformy na komunikáciu lekára s pacientom, ktorá je dostupná v rámci celého sveta. Celý systém je navrhnutý tak, aby lekár dokázal v čo najkrajšom čase a za použitia rôznych integrovaných medicínskych zariadení vyhodnotiť stav pacienta prostredníctvom audio alebo video hovoru.

Za jeden z ďalších trendov v oblasti zdravotníctva môžeme považovať aj zavádzanie digitálnych dotazníkov, ktoré si získali veľkú popularitu aj počas pandémie, kde si ľudia vedeli na základe rôznych príznakov vyplniť či sú zdraví alebo by mali radšej podstúpiť testovanie. Tieto dotazníky sa však čoraz viac začínajú využívať aj na kontrolu zdravotného stavu, kde dokáže pacient priamo z domu odpovedať na otázky lekára a na základe týchto odpovedí dokáže lekár monitorovať a reagovať na vývoj stavu, respektíve urobiť nevyhnutné zmeny v navrhnutej liečbe.

Do budúcnosti predpokladáme ešte väčšie využitie takzvaných wearables, teda nositeľných medicínskych zariadení. Čoraz viac zariadení v sebe obsahuje elektroniku, ako napríklad taký inhalátor, čo uľahčuje vyhodnocovanie správneho používania. IT je už teraz dôležitou súčasťou zdravotníctva a predpokladáme, že tento trend bude aj ďalej pokračovať. Technológie, ktoré prichádzajú do zdravotníctva, tu nie sú za účelom aby nahrádzali lekárov, ale aby im uľahčovali prácu, zjednodušili diagnostiku a poskytovali väčší komfort pre pacientov.

4.4. Zhrnutie identifikovaných trendov, výziev a naše odporúčania

Na základe vykonanej analýzy, vyhodnotenia jednotlivých prieskumov a identifikácií konkrétnych trendov sme dospeli k záveru, že pred spoločnosťami, ktoré sa chcú pretransformovať na spoločnosti riadené dátami, stojí niekoľko výziev, ktorým musia čeliť. Medzi takéto základné výzvy, ktoré sme identifikovali na základe výsledkov prieskumov patria:

- Spoločnosti musia pochopiť, že dáta sú pre nich podnikové aktíva a začať hľadať odpovede na otázku Akú hodnotu majú pre nich tieto údaje a ako sa dajú speňažiť?
- Zabezpečenie vysokej kvality údajov
- Odstránenie kultúrnych faktorov, úprava procesov a presvedčenie ľudí o dôležitosti kroku stať sa spoločnosťou založenou na dátach
- Pochopenie spoločností, že proces transformácie na spoločnosť založenú na dátach nie je krátkodobý, ale je potrebné mu venovať dostatok času

Aby spoločnosti dokázali naplniť tieto výzvy, odporúčali by sme im urobiť niekoľko rozhodnutí ako sú :

- Dosadenie do kľúčových manažérskych pozícií, ľudí, ktorí majú záujem a skutočne veria v úspešnú transformáciu spoločnosti.
- Realizácia osvedčenej už pred samotnou realizáciou takejto transformácie, robiť osvetu medzi ľuďmi, ukázať im jednotlivé prínosy a vybudovať novú kultúru spoločnosti založenú na dátach, aby aj staršie generácie pochopili, že to aj pre nich bude prospešné.
- Vyčleniť si dostatok kapacít, či už finančných, časových alebo ľudských na realizáciu takejto transformácie spoločnosti a detailne naplánovanie celej tejto transformácie s jasne vytýčenými a merateľnými cieľmi, ktoré sa budú snažiť dosiahnuť.

V rámci diplomovej práce sme tiež odhadli trendy big data z technologického oblasti, z oblasti ochrany osobných údajov a nakoniec aj trendy z oblasti ľudských faktorov. Ako kľúčové trendy sme identifikovali :

- Cloud computing a umelú inteligenciu
- Ochranu osobných údajov a s ňou spojené dodržiavanie legislatívy
- Nedostatok kvalifikovaných zamestnancov v oblasti dátová veda

Na základe týchto trendov sme zostavili niekoľko odporúčaní pre spoločnosti na Slovensku:

- Väčší tlak zo strany spoločností, na zavedenie viacerých vysokoškolských programov, čo sa týka dátovej vedy, respektíve pridanie predmetov v oblasti dátovej vedy a dátovej analytiky
- Zavádzanie interných akadémií, v rámci ktorých si spoločnosti dokážu vychovávať vlastných odborníkov zo študentov už počas ich štúdia a tým z časti zmiernia nedostatok odborníkov na trhu.
- Postupné prechádzanie na cloud riešenia, či už ako IaaS alebo SaaS, čím dokážu ušetriť náklady a čas potrebný na riešenie inštalácií, opravu chýb a v ďalšom rade šetrenie ľudskej kapacity
- Vytvorenie podobného priestoru ako je inovačný hub, ktorý už je využívaný v rámci bankového sektora, jeho rozvoj a zavedenie aj do iných sektorov

- Naše hlavné odporúčanie je v zlepšení celkových digitálnych zručností, a vo väčšej miere plnenie stratégie akčného plánu digitálnej transformácie Slovenska, ktorý bol schválený, ale stále ešte nevidíme aspoň nejaké čiastkové výsledky.

Záver

Neustály rozvoj inteligentných zariadení, ktoré nám zvyšujú komfort má za následok aj prudký nárast dát, ktoré tieto zariadenia generujú. Ale dáta, ktoré nie sú spracované nám nedokážu nič priniesť, a práve preto sa v tejto oblasti snaží rozvíjať nielen súkromný, ale aj verejný sektor. Ako sme demonštrovali v tejto diplomovej práci, big data sú tu už nejaký čas a aj spoločnosti si uvedomujú, aké dôležité je rozvíjať sa v oblasti dátovej analytike. S tým súvisí aj celková zmena kultúry každej spoločnosti, aby sa dokázali transformovať na moderné dátami riadené spoločnosti a využili maximálny potenciál, ktorý nám táto oblasť ponúka.

Nakoľko oblasť big data je už spoločnostiam známa, v rámci diplomovej práce sme sa snažili identifikovať možné trendy a výzvy, ktoré sú spájané s touto oblasťou. Náš hlavný cieľ sme si rozdelili na niekoľko čiastkových cieľov. Na základe týchto čiastkových cieľov sme našu prácu rozdelili na dve časti. V prvej časti sme sa zamerali na popísanie aktuálnej situácie, vlastnosti a celkový proces dátovej analytiky. Taktiež sme opísali niektoré obľúbené technológie, ktoré nám dokážu uľahčiť prácu s big data. Táto kapitola nám má slúžiť na lepšie pochopenie celého tohto konceptu, aby sme si uvedomili, čo si predstaviť pod pojmom big data. Po získaní prehľadu o big data sme sa zamerali na druhú časť.

V druhej časti sme sa snažili naplniť viacero čiastkových cieľov. Jedným z týchto čiastkových cieľov, ktorý sme si zvolili bolo aj zhrnutie aktuálnej situácie v rámci spoločnosti a ich transformácie, na dátami riadené spoločnosti. Na základe dát získaných z prieskumov, ktoré sme spracovali do grafov, sme identifikovali niektoré základne výzvy, ktoré pred spoločnosťami stoja a s ktorými sa musia vyrovnávať ak chcú, aby ich transformácia prebehla úspešne. Taktiež sme sa zamerali na stratégie, ktoré by spoločnosti mali zvoliť a priority na ktoré by sa mali zamerať.

Ďalším čiastkovým cieľom, ktorý sme si stanovili bola identifikácia trendov, ktoré sú v oblasti big data. Tieto trendy sme rozdelili do troch oblastí a to technologická oblasť, legislatívna oblasť a oblasť ľudského kapitálu. Na základe vyhodnotenia údajov, ktoré sme spracovali sme identifikovali niekoľko trendov, ktoré sú podľa nášho názoru kľúčové pre to, aby sa spoločnosti dokázali stať dátami riadenými, respektíve aby zlepšili svoju efektivitu a znížili náklady.

V rámci posledného čiastkového cieľa sme sa pozreli na situáciu v jednotlivých sektoroch, a porovnali situáciu v rámci týchto sektorov na Slovensku a v zahraničí.

Výsledkom práce je záverečné zhrnutie identifikovaných výziev a súbor našich odporúčaní pre jednotlivé spoločnosti v rôznych sektoroch na Slovensku.

Zoznam použitej literatúry

- [1] KASWAN, Kuldeep Singh – DHATTERWAL, Jagjit Singh. *Big data: An Introduction* [online]. India: Shashwat Publication, 2020. 277 s. [cit.10.12.2021]. ISBN 978-93-90290-31-4. Dostupné na : https://books.google.sk/books?id=2Iv-DwAAQBAJ&pg=PA26&dq=big+data+history&hl=sk&sa=X&ved=2ahUKEwj3sMG8ibb0AhU_SvEDHbZNBqkQ6AF6BAgDEAI#v=onepage&q=big%20data%20history&f=false
- [2] GARTNER. *Information Technology Gartner Glossary* [online]. [cit. 20.12.2021]. Dostupné na: <https://www.gartner.com/en/information-technology/glossary/big-data>
- [3] ALI, H M Sohaib – ARSHAD, Muhammad Jameel. 7, *Vs of Big Data: A Survey* [online]. Engineering science and technology international research journal, vol.3, 2019. [cit.20.12.2021] Dostupné na: <http://www.estirj.com/Volume.3/No.4/12Sohaib34.pdf>
- [4] EBERENDU, Adanma Cecilia. *Unstructured Data: an overview of the data of Big Data* [online]. International Journal of Computer Trends and Technology (IJCTT), Volume 38 Number 1, 2016. [cit. 28.12.2021]. Dostupné na: https://www.researchgate.net/profile/Adanma-Eberendu/publication/309393428_Unstructured_Data_an_overview_of_the_data_of_Big_Data/links/5bc89b5c458515f7d9c65beb/Unstructured-Data-an-overview-of-the-data-of-Big-Data.pdf
- [5] GODDARD, William. *Pros and Cons of Big Data* [online]. 2021. [cit. 28.12.2021]. Dostupné na : <https://itchronicles.com/big-data/pros-and-cons-of-big-data/>
- [6] EURÓPSKY PARLAMENT. *Aké výhody a výzvy prinášajú so sebou veľké dáta* [online]. 2021.[cit.28.12.2021]. Dostupné na : <https://www.europarl.europa.eu/news/sk/headlines/priorities/umela->

inteligencia/20210211STO97614/big-data-ake-vyhody-a-vyzvy-prinasaju-so-sebou-velke-data-infografika

[7] SEDKAOUI, Soraya – KHELFAOUI, Mounia – KADI, Nadjat. *Big data analytics: Harnessing Data for New Business Models* [online]. Apple Academic Press Inc, 2022. [cit.28.12.2021]. Dostupné na : https://edu.anarcho-copy.org/other/Big%20Data/Big%20Data%20Analytics_Harnessing%20Data.pdf#page=36

[8] VAN RIJMENAM, Mark. *A Short History Of Big Data* [online]. 2013. [cit.03.01.2022]. Dostupné na: <https://datafloq.com/read/big-data-history/>

[9] RIVERA, Eduardo. *The V's of Big Data* [online]. 2020. [cit. 21.12.2021]. Dostupné na: <https://miuc.org/vs-big-data/>

[10] SHAFER, Tom. *The 42 V's of Big Data and Data Science* [online]. 2017. [cit. 20.12.2021]. Dostupné na : <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>

[11]TABLEAU. *Big Data Analytics: What It Is, How It Works, Benefits, And Challenges* [online]. [cit.14.12.2021]. Dostupné na : <https://www.tableau.com/learn/articles/big-data-analytics>

[12]APACHE hadoop. *Apache Hadoop 3.3.2* [online]. [cit. 14.12.2021]. Dostupné na : <https://hadoop.apache.org/docs/stable/>

[13]MongoDB. *What is NoSQL?* [online]. [cit. 14.12.2021]. Dostupné na : <https://www.mongodb.com/nosql-explained>

[14]APACHE spark. *Spark Overview* [online]. [cit. 15.12.2021]. Dostupné na : <https://spark.apache.org/docs/latest/>

[15]CAPGEMINI. *The data-powered enterprise* [online]. 2020 [cit 25.3.2022]. Dostupné na :https://www.capgemini.com/wp-content/uploads/2020/11/Data-powered-enterprise_Digital_Report.pdf

[16]NEWVANTAGE PARTNERS. *Data and AI Leadership executive survey* [online]. 2022. [cit. 26.3.2022]. Dostupné na: <https://www.wavestone.com/app/uploads/2022/01/Wavestone-2022-Data-and-AI-Leadership-Executive-Survey-Report-1.pdf>

[17] STATISTA. *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025* [online]. 2022. [cit.26.3.2022]. Dostupné na: <https://www.statista.com/statistics/871513/worldwide-data-created/>

[18] KNORR, Erick. *The 2020 IDG Cloud Computing Survey* [online]. 2020. [cit. 22.3.2022]. Dostupné na : <https://www.infoworld.com/article/3561269/the-2020-idg-cloud-computing-survey.html>

[19] MCKINSEY. *The state of AI in 2021* [online]. 2021. [cit. 28.3.2022]. Dostupné na: https://www.mckinsey.com/~/_media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Global%20survey%20The%20state%20of%20AI%20in%202021/Global-survey-The-state-of-AI-in-2021.pdf

[20] EUROPEAN COMMISSION. *The GDPR: new opportunities, new obligations* [online]. Luxembourg: Publications Office of the European Union, 2018. [cit. 28.3.2022]. ISBN 978-92-79-79430-8. Dostupné na: https://ec.europa.eu/info/sites/default/files/data-protection-factsheet-sme-obligations_en.pdf

[21] STATISTA. *Largest fines issued for General Data Protection Regulation (GDPR) violations as of February 2022* [online]. 2022. [cit.20.3.2022]. Dostupné na: <https://www.statista.com/statistics/1133337/largest-fines-issued-gdpr/>

- [22] NBS. *Inovačný hub a pravidlá jeho fungovania* [online]. [cit. 15.3.2022]. Dostupné na: <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech#ih>
- [23] SCHWARTZ, Eric Hal. *Bank of America's Virtual Assistant Erica Explodes in Popularity* [online]. 2021. [cit.25.3.2022]. Dostupné na : <https://voicebot.ai/2021/04/21/bank-of-americas-virtual-assistant-erica-explodes-in-popularity/>
- [24] AHUJA, Sudhanshu . *JIM* [online].2018.[cit.25.3.2022].Dostupné na: <https://www.dbs.com/startupxchange/stories-jim.html>
- [25] ORMANDJIEVA, Olga – TRUDEL, Sylvie. *Measuring the 3V's of Big Data: A Rigorous Approach* [online].2020.[cit. 20.12.2021]. Dostupné na: <http://ceur-ws.org/Vol-2725/paper5.pdf>
- [26] ALLEN, Richard. *What are the Types of Big Data?* [online]. [cit.21.12.2022]. Dostupné na : <https://www.selecthub.com/big-data-analytics/types-of-big-data-analytics/>
- [27] SCYLLADB. *nosql-vs-sql* [online]. [cit. 28.12.2021]. Dostupné na: <https://www.scylladb.com/learn/nosql/nosql-vs-sql/>
- [28] MULAY, Preeti – CHAUDHARI, Archana. *SCSI: Real-Time Data Analysis with Cassandra and Spark* [online]. 2019. [cit. 03.01.2022]. Dostupné na : https://www.researchgate.net/figure/The-proposed-dimensions-of-Internet-of-Thing-IoT-Data_fig1_325813017
- [29] BEAN, Rendy. *Transforming The Insurance Industry With Big Data, Machine Learning And AI* [online]. 2021. [cit. 28.03.2022]. Dostupné na: <https://www.forbes.com/sites/andybean/2021/07/06/transforming-the-insurance-industry-with-big-data-machine-learning-and-ai/?sh=6aa7c4faf8a6>

[30] BROWN, Sara. *Machine learning, explained* [online]. 2021. [cit.22.3.2022]. Dostupné na : <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

[31] REDMAN, Thomas C.. *Seizing Opportunity in Data Quality* [online]. 2017.[cit. 27.3.2022]. Dostupné na : <https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/>

[32] DUBOIS, Jen. *The Data Scientist Shortage in 2020* [online]. 2020. [cit. 03.04.2022]. Dostupné na: <https://quanthub.com/data-scientist-shortage-2020/>

[33] ENCRYPTIONCONSULTING. *Data Protection in Big Data/Data Lakes* [online]. 2019. [cit.03.04.2022]. Dostupné na: <https://encryptionconsulting.com/data-protection-in-big-data-using-encryption/>