Univerzita Pardubice

Fakulta ekonomicko-správní

Charakteristiky Big dat a jejich vliv na bezpečnost

Diplomová práce

2021 Ing. Irena Michalková

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2020/2021

ZADÁNÍ DIPLOMOVÉ PRÁCE
(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Bc. Irena Michalková**
Osobní číslo:	**E20685**
Studijní program:	**N0688A140007 Informatika a systémové inženýrství**
Studijní obor:	**Informační a bezpečnostní systémy**
Téma práce:	**Charakteristiky big dat a jejich vliv na bezpečnost**
Zadávající katedra:	**Ústav systémového inženýrství a informatiky**

Zásady pro vypracování

Cílem práce je popsat charakteristiky a vývoj big dat a na základě rešerše literatury identifikovat postupy při zohlednění charakteristik big dat a jejich vlivu na bezpečnost.
Osnova:
- Základní pojmy řešené problematiky.
– Charakteristiky big dat v kontextu bezpečnosti.
– Identifikace postupů při zohlednění charakteristik big dat.
– Formulace závěrů a doporučení.

Rozsah pracovní zprávy: **cca 50 stran**
Rozsah grafických prací:
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. 2015. Big Data a NoSQL databáze. Praha: Grada Publishing. ISBN 978-80-247-5938-8.
DEHGHANTANHA, Ali a Kim-Kwang Raymond CHOO. (Eds.). 2019. Handbook of Big Data and IoT Security. Springer International Publishing. ISBN 978-3-030-10543-3.
SIMON, Phil. 2013. Too Big to Ignore: The Business Case for Big Data. John Wiley & Sons. ISBN 978-1118638170.
TONIDANDEL, Scott, Eden B. KING a Jose M. CORTINA. (Eds.). 2016. Big Data at Work: The Data Science Revolution and Organizational Psychology. New York: Routledge, Taylor & Francis Group. ISBN 978-1-84872-581-2.
ZIKOPOULOS, Paul a Chris EATON. 2011. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media. ISBN 978-0-07-179053-6.

Vedoucí diplomové práce: **Ing. et Ing. Martin Lněnička, PhD.**
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **1. září 2020**
Termín odevzdání diplomové práce: **30. dubna 2021**

L.S.

prof. Ing. Jan Stejskal, Ph.D.
děkan

RNDr. Ing. Oldřich Horák, Ph.D.
vedoucí ústavu

V Pardubicích dne 1. září 2020

Prohlašuji:

Práci s názvem Charakteristiky big dat a jejich vliv na bezpečnost jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 20. 11. 2021.

Ing. Irena Michalková v. r.

PODĚKOVÁNÍ

Tímto bych chtěla poděkovat vedoucímu Ing. et Ing. Martinovi Lněničkovi, PhD. za veškerou pomoc, cenné rady, připomínky a čas, který mi věnoval při zpracování této diplomové práce. Dále ještě svým blízkým, kteří mi byli v té době velkou oporou.

ANOTACE

Big data je termín používaný pro velké datové soubory bez jednotné formy, které jsou vytvářeny velice rychle a nelze je snadno ukládat, analyzovat a vizualizovat, což ještě více ztěžuje jejich zabezpečení. Diplomová práce řeší tuto problematiku, když v teoretické části je nejdříve zpracován přehled existujících charakteristik, jimiž se big data definují. V další části je popsána metodika použitá k provedení systematické rešerše literatury. Praktická část je věnována analýze vybraných studií, kde se konkrétně řeší charakteristiky big dat (rychlost, objem, věrohodnost, zranitelnost atd.) a jejich vliv na bezpečnost. Nakonec jsou na základě analýzy představeny požadavky a doporučení na zpracování konkrétních problémů a rizik figurujících v rámci bezpečnosti a ochrany big dat, spolu s jejich kategorizací a odkazem na relevantní literaturu.

KLÍČOVÁ SLOVA

Big data, bezpečnost, charakteristiky, V, rešerše, analýza

TITLE

Characteristics of big data and their effect on security

ANNOTATION

Big data is a term used for large data sets without unified format that are created very fast and are hard to storage, analyze and visualize, which makes it even more difficult to secure them. The diploma thesis deals with this issue. The theoretical part of the thesis presents an overview of existing big data characteristics. The next part is dedicated to methodology used to conduct systematic literature review. The practical part focuses on analysis of relevant studies to identify the effects of big data characteristics (namely velocity, volume, veracity, vulnerability etc.) on security. Finally, based on the analysis, the requirements and recommendations for solving security and privacy issues of big data are presented together with their classification and references to the relevant literature.

KEYWORDS

Big data, security, characteristics, V, review, analysis

OBSAH

SEZNAM ILUSTRACÍ A TABULEK

SEZNAM ZKRATEK A ZNAČEK

APT	pokročilé trvalé hrozby (Advanced Persistent Threats)
BUG	generalizace zdola nahoru (bottom-up generalisation)
CSA	Cloud Security Alliance
DM	data mining
GDPR	obecné nařízení o ochraně osobních údajů
ICT	informační a komunikační technologie
IPS	systém prevence průniku (Inturison Prevention Systems)
IoT	internet věcí (Internet of Things)
PKE	šifrování veřejným klíčem (Public Key Encryption)
TDS	specializace shora dolu (top-down specialisation)

ÚVOD

Big data jsou pro většinu organizací a osob nový pojem. Ačkoliv tento termín vznikl již ve 20. století, zájem o big data se rozšířil teprve před pár lety. Mnoho z nás navíc ani v dnešní době netuší, co tento pojem vlastně znamená a domníváme se, že touto problematikou nejsme nijak ovlivněni. Big data jsou však nedílnou součástí našeho každodenního života a neustále přibývají spolu se vším, co na internetu, a v počítačových sítích obecně, děláme.

Big data se stala jednou z nejslibnějších a nejrozšířenějších technologií pro předpovídání budoucích trendů, a proto je třeba při jejich využívání vzít do úvahy i otázku bezpečnosti a ochrany dat.

Data musí mít určité charakteristiky, aby se dalo tvrdit, že se jedná o big data. V základu se jedná o 3V: volume (objem), variety (rozmanitost) a velocity (rychlost). Aby šlo o big data, musí být tyto vlastnosti splněny. V současnosti se produkuje ohromné množství dat denně, kdy se liší i jejich formát, a tvoří se tak rychle, že je velice složité je analyzovat v reálném čase. To vyvolává řadu problémů při jejich zpracování, především co se týká bezpečnosti, její implementace a řízení. U 3V to ale nekončí, dodnes je definováno téměř 62 charakteristik, kterými se big data definují a kategorizují, což ještě rozšiřuje prostor pro zranitelnost a rizika, a vyvolává potřebu pro nová opatření.

S rozšiřováním big dat ve světě se stále více organizací a jednotlivců zabývá jejich zpracováním a využitím, nebo to alespoň zamýšlí vzhledem k jejich potenciálu. Neustále se tak objevují nové výzvy a zkoumají nové příležitosti, které přichází spolu se sběrem big dat, jejich ukládáním, správou, analýzou, a také vizualizací, jež jsou způsobeny jejich charakteristikami. Big data navíc vyžadují nové a často výkonnostně náročnější postupy, které tradiční technologie nejsou schopny zajistit. Pro práci s big daty je proto nutné přizpůsobit infrastrukturu a přejít na nové technologie a komplexní systémy či rovnou vyvinout nové, které zvládnou pracovat s tak rozsáhlým objemem dat.

Cílem této práce je popsat charakteristiky a vývoj big dat a na základě rešerše identifikovat postupy při zohlednění charakteristik big dat a jejich vlivu na bezpečnost.

V současné době existuje obrovské množství studií a zdrojů, které se zabývají big daty, což přináší potíže pro výzkumníky i odborníky z praxe, kteří se snaží v této problematice zorientovat a najít témata v okruhu svých potřeb a zájmů. Tato práce se proto snaží přispět k vyjasnění, zpřehlednění, a především k vymezení současného stavu big dat a zároveň

nabídnout komplexní přehled existujících charakteristik a různých pohledů na problematiku bezpečnosti big dat.

V úvodu práce je možné se seznámit s definicí big dat a jejich charakteristikou, také známou jako V big dat, která se stále vyvíjí a nových vlastností neustále přibývá. Do roku 2021 existuje přes 56V. Druhá kapitola se zabývá metodikou a postupem řešení této práce. Jako hlavní metoda práce byla zvolena systematická literární rešerše, která se skládá z definování klíčových slov, výběru databází, postupů pro výběr relevantních studií až po přístup k analýze a prezentaci výsledků. Hlavní část je věnována analýze vybraných publikací s cílem identifikovat nejdůležitější charakteristiky big dat, které mají největší dopad na zabezpečení a ochranu big dat. Poslední kapitola shrnuje poznatky vědecké komunity na řešení problémů a zajištění bezpečnosti u big dat a klasifikuje je do kategorií.

Vytvořený přehled bezpečnostních problémů, jejich vymezení a doporučení pro teorii i praxi by proto mohl pomoci lépe porozumět změnám v požadavcích na bezpečnost, které vyplývají z různých charakteristik big dat, a zároveň nalézt nové směry k realizaci dalších studií věnujících se této problematice.

1 VYMEZENÍ ZÁKLADNÍCH POJMŮ

Big data v sobě v dnešní době skrývají jednu z potenciálně největších případných hodnot, co se týká přínosů pro rozvoj organizace a podporu rozhodování, a to nejenom v oblasti ICT, a jejich význam neustále roste. Jsou to taková data, která nejde běžnými postupy zpracovat či zachytit a v současnosti o nich lidé spíše jen mluví a píšou, než aby je skutečně používali. Takto se pak dá tvrdit, že se jedná o oblast, která do jisté míry na své zpracování a uchopení stále čeká.

1.1 Big data

Formální a přesnou definici big dat není možné nalézt, ale jde je poměrně přesně definovat podle jejich charakteristik, tzv. V-parametrů.

Ty se týkají technologií a postupů, které generují a obsahují data, jež jsou moc různorodá a rychle se měnící. Jinými slovy, jejich objem, rychlost a rozmanitost jsou příliš velké. Big data vyžadují nové technologie s takovou architekturou, která zajistí jejich zpracování, analýzu a vizualizaci výsledků, s čímž souvisí i následující tvrzení od Zikopoulos a Eaton (2011), že termín big data, nebo také česky velká data či veledata, se vztahuje na informace, které nejde zpracovat ani analyzovat pomocí tradičních nástrojů a procesů. Společnost Gartner je zase definuje jako: *„Data, jejichž objem, rychlost a různorodost vyžadují efektivní a inovativní formy pro zpracování informací k lepšímu pochopení a rozhodování“*. Otázkou však zůstává, jak velká data musejí být, aby se stala big daty.

Když se mluví o big datech, tak se uvažují digitální data, neboť se termín big data objevil až s příchodem nových technologií, kdy tyto technologie a aplikace generují každou vteřinou obrovský objem dat, která jsou potřeba uložit a účelně zpracovat. Holubová et al. (2015) ve své knize cituje společnost IBM, která vyjadřuje myšlenku *„V závislosti na odvětví a organizaci zahrnují big data informace z interních a externích zdrojů, jako jsou transakce, sociální média, podniková data, senzory a mobilní zařízení. Firmy mohou tato data využívat, aby lépe přizpůsobily své výrobky a služby potřebám zákazníka, dále optimalizovaly provoz a infrastrukturu a/nebo nalezly zcela nové zdroje příjmů“*.

Big data se nejčastěji dělí na strukturovaná, nestrukturovaná a často i semi-strukturovaná. Nestrukturovaná data tu byla pořád, ale až nedávno se začala v rámci big dat využívat, především k podpoře rozhodování.

1.2 Charakteristiky Big dat

Proces vyhledávání charakteristik big dat probíhal: 1) analýzou publikací tykající se big dat, tzn. Dehghantanha a Choo (2019), Holubová et al. (2015), Simon (2013), Tonidandel et al. (2016), Zikopoulos a Eaton (2011); 2) dohledání zdrojů a dalších charakteristik, které nebyly pokryty v těchto publikacích, především s využitím webových vyhledávačů.

Riahi (2018) tvrdí, že big data podléhají určitým charakteristikám, kdy odpovídají větším datovým souborům. Jsou různorodější, jelikož zahrnují strukturovaná i nestrukturovaná data a jsou přijímána mnohem rychleji než dříve – to jsou tzv. 3V:

3V – Volume, Velocity, Variety (2001)

Volume (objem) – představuje množství generovaných a uložených dat v systému stejně tak i dat, se kterými se pracuje. Zvětšující se objem je vysvětlován zvětšujícím se množstvím generovaných a uložených dat, ale i potřebou je využívat. Aby se dalo mluvit o big datech, tak velikost musí být větší či rovna 10 TB.

Velocity (rychlost) – charakterizuje frekvenci, na které jsou data generována, zachycována a sdílena. Data musí být analyzována v reálném čase. Jedná se o např. rychlost zpracování dat na sociálních sítích, vývoj na akciových trzích či senzory v autech. Dnes vše funguje stále v kratších intervalech. Data rostou a rychle se mění.

Variety (typ) – představuje různorodost typů dat, která vstupují do informačních systémů a databází. Spadají sem strukturovaná, semi-strukturovaná a nestrukturovaná data. Tedy data existují v různých formátech.

V nejužším pojetí se big data definují podle těchto tří charakteristik, ale časem se postupně přidávají i další.

4V – Volume, Velocity, Variety, Veracity (2011)

K původním charakteristikám (Volume, Velocity, Variety – viz definice 3V) se přidává další důležité V pro lepší definování big dat (Diaz, 2020):

Veracity (věrohodnost) – představuje stupeň kvality, přesnosti a nejistoty v datech a datových zdrojích. Jinak řečeno, informace má takovou váhu, z jakého zdroje je získána – příspěvek na sociálních sítích bude mít jinou důvěryhodnost než data získaná např. z databáze banky.

5V – Volume, Velocity, Variety, Veracity, Value (2012)

Volume, Velocity, Variety, Veracity – viz předchozí definice

+ Value (hodnota) – hodnota získaných dat. Jedná se vlastně o měřítko úspěšnosti využívání big dat. Je to *cíl*, neboť všichni chtějí, aby jejich organizace z dat něco získávala.

5V – Volume, Velocity, Variety, Viability, Value (2013)

Tato verze 5V je podle Biehna (2018) upravena o proveditelnost, která nahrazuje důvěryhodnost (Veracity) u předchozí verze.

Volume, Velocity, Variety, Value **–** viz předchozí definice

+ Viability (proveditelnost) – říká, že by se výsledky analýzy měly co nejvíce přibližovat realitě.

6V – Volume, Velocity, Variety, Veracity, Viability, Value

Fouad et al. (2015) ve své publikaci zase uvádí charakteristiky Big dat jako 6V, které se skládají z Volume, Velocity, Variety, Veracity, Viability a Value, jejichž definice je možné shlédnout v charakteristikách 5V.

7V – Volume, Velocity, Variety, Variability, Veracity, Visualization, Value (2013)

Volume, Velocity, Variety, Veracity, Value

+ Visualization (vizualizace) – se zaměřuje na schopnost vizualizace potřebných výsledků, neboť zpracovaná data bez vizualizace za pomocí např. grafů neposkytnou potřebné informace. Obtížnost vizualizace je jedním z parametrů pro analýzu big dat, neboť tradiční grafy nedokážou správně ukázat velké množství informací (Mcnulty, 2014).

+ Variability (proměnlivost) – nesmí být zaměněna za typ (Variety). Proměnlivost dat je způsobena změnou významu dat (např. slov v textu). Příkladem může být i dle Impactu (2016) káva z kavárny, kdy si ten samý typ budeme kupovat každý den, ale pokaždé bude chutnat o trochu jinak. To samé pak platí i pro data, pokud se budou pořád měnit, může to mít velký dopad na homogennost.

8V – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Value (2014)

Vorhies (2014) ve svém příspěvku definuje 8V charakteristiky big dat jako:

Volume, Velocity, Variety, Variability, Veracity, Value

+ Viscosity (viskozita) – termín používaný k popisu doby zpoždění a latence dat k popisované události. Vorhies tvrdí, že tento parametr je velmi podobný a chápaný jako parametr rychlosti (Velocity).

+ Virality (viralita) – definována jako rychlost šíření dat mezi subjekty, tj. jak často se data získávají a reprodukují dalšími uživateli a událostmi.

8V – Volume, Velocity, Variety, Variability, Veracity, Visualization, Validity, Value

Naopak Tech Blogger (2020) zastává názoru, že pod 8V-parametry Big dat se představují:

Volume, Velocity, Variety, Variability, Veracity, Visualization, Value – viz předchozí definice

+ Validity (platnost) – tato charakteristika je velmi podobná důvěryhodnosti dat (Veracity). Zaměřuje se na to, jak jsou data přesná a správná k cílovému použití. Firican (2017) uvádí, že 60 % času stráví výzkumníci nad čištěním dat, než nad nimi mohou provádět potřebné analýzy.

9V – Volume, Velocity, Variety, Variability, Veracity, Validity, Volatility, Visualization, Value (2014)

Tuto verzi doplněnou o volatilitu představila ve svém článku Mittal (2017).

Volume, Velocity, Variety, Variability, Veracity, Visualization, Validity, Value

+ Volatility (těkavost) – se zaměřuje na otázku, jak dlouho data uchovávat, než se stanou historickými a nepoužitelnými.

10V – Volume, Velocity, Variety, Variability, Veracity, Validity, Vulnerability, Volatility, Visualization, Value (2017)

V roce 2017 byla uvedena charakteristika Vulnerability jako další parametr big dat, čímž se 9V rozšířilo o další vlastnost.

Volume, Velocity, Variety, Variability, Veracity, Validity, Volatility, Visualization, Value

+ Vulnerability (zranitelnost) – big data s sebou přináší i nové bezpečnostní hrozby, koneckonců u big dat může uniknout více dat.

10V – Volume, Velocity, Variety, Variability, Veracity, Validity, Vocabulary, Venue, Value, Vagueness (2014)

Další verzi charakteristik Big dat známé jako 10V podle článku od Arockia et al. (2017) definoval Kirk Born v roce 2014, která sestává z Volume, Velocity, Variety, Variability, Veracity, Validity, Value

+ Vocabulary (slovník) – zahrnuje datovou terminologii, jako jsou datové modely a struktury.

+ Venue (místo) – zaměřuje se na odlišné umístění dat, kdy různé typy dat přicházejí z různých zdrojů a platforem jako je interní databáze či veřejný a soukromý cloud.

+ Vagueness (vágnost) – zaznamenává realitu, že význam nalezených dat je většinou nejasný nehledě na to, kolik dat je k dispozici, tj. data poskytují malé množství nebo žádnou informaci o tom, co ve skutečnosti znamenají.

14V + 1C – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Complexity (2017)

Arockia et al. (2017) ve svém článku uvádí, že se výzkum kolem big dat točí převážně kolem charakteristik, které lze nazvat jako 14V a 1C, ovšem taky dodává, že zkoumat jenom tyto charakteristiky není dostačující.

Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness – viz předchozí definice

+ Complexity (složitost) – data pocházejí z různých zdrojů, s čímž souvisí nutnost zjistit veškeré změny na datech (ať už malé nebo velké) s ohledem na dříve přijatá data. Data jsou při zpracování big dat potřeba porovnat, sloučit, provázat, vyčistit a transformovat a tím i zjistit vztahy mezi nimi.

17V – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Verbosity, Voluntariness, Versatility (2017)

Analýza dokumentů a literatury o big datech identifikovala další tři charakteristiky, které big data definují (Arockia et al., 2017).

Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness

+ Verbosity (výřečnost) – se vnímá jako nadbytečnost informací dostupných z různých zdrojů.

+ Voluntariness (dobrovolnost) – vzhledem k tomu, že big data dobrovolně pomáhají nespočetně organizacím, se dá dobrovolnost big dat chápat jako možnost využít plně dostupná big data k vybranému kontextu.

+ Versatility (všestrannost) – flexibilita big dat, tj. schopnost dat být použita k čemukoliv podle různých kontextů.

42V – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Viability, Virtuosity, Visibility, Valor, Varnish, Versed, Vault, Voodoo, Veil, Vulpine, Verdict, Vet, Vane, Vanilla, Victual, Vantage, Varmint, Vivify, Vastness, Voice, Vaticination, Veer, Voyage, Varifocal, Version control, Vexed, Vibrant, Vogue (2017)

Shafer (2017) ve svém příspěvku uvedl aktualizovaný seznam z roku 2017 ohledně počtu V-parametrů charakterizujících big data.

Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Viability

+ Virtuosity (virtuozita) – představuje dychtivost získávat další znalosti o big datech.

+ Visibility (viditelnost) – data science (věda o datech) poskytuje náhled do komplexních problémů big dat.

+ Valor (chrabrost) – znamená potřebu řešit velké problémy big dat čelem a statečně.

+ Varnish (pozlátko) – vzhledem k tomu, že záleží na tom, jak koncoví uživatelé přijímají a reagují na výslednou práci, se její přikrášlení počítá (Shafer, 2017).

+ Versed (zběhlost) – datoví výzkumníci big dat musí mít znalosti z různých disciplín (matematika, statistika, programování, databáze a další).

+ Vault (trezor) – představuje důležitost pro zabezpečení dat, když existuje mnoho aplikací datové vědy, které jsou založené na velkých a často citlivých souborů dat.

+ Voodoo – obrazně se jedná o otázku, jak přinést výsledky, které ovlivní reálný svět.

+ Veil (závoj) – data science poskytuje možnost nahlédnout za oponu a prozkoumat účinky skrytých proměnných v datech.

+ Vulpine (mazanost) – data vedou k novým technologiím pro ty, co chtějí uspět.

+ Verdict (rozhodnost) – rozhodování na základě modelů ovlivňuje stále větší počet lidí. Veracity (věrohodnost) a Validity (platnost) se tak stávají ještě důležitější

+ Vet (prověření) – data science umožňuje prověřit domněnky a posílit intuici o důkazy.

+ Vane (větrník) – věda o datech může ukázat směr pro správné rozhodování.

+ Vanilla (všednost) – i ty nejjednodušší modely mohou poskytnout hodnotu.

+ Victual (potravina) – jednoduše řečeno big data jsou krmivem pro data science.

+ Vantage (výhodnost) – big data umožňují nový pohled na složité systémy.

+ Varmint (havěť) – tím, jak se zvětšují big data, zvětšují se i softwarové chyby.

+ Vivify (oživení) – data science má potenciál poradit si se všemi problémy reálného života a inovovat všechny způsoby rozhodování a obchodních procesů, od reklamy po odhalování podvodů.

+ Vastness (rozsáhlost) – s příchodem internetu věcí se rozsah a velikost big dat zvyšuje.

+ Voice (vyjádření) – schopnost odborníků big dat pohybovat se v rámci různých témat.

+ Vaticination (předvídavost) – schopnost předpovídat. Ovšem přesnost těchto předpovědí závisí na náročnosti a složitosti problému.

+ Veer (obrat) – schopnost big dat orientovat se a měnit směr podle potřeb zákazníka

+ Voyage („dlouhá cesta") – big data nabízí možnost neustálého učení se a získávání nových znalostí při řešení problémů.

+ Varifocal (varifokální) – big data spolu s vědou o datech metaforicky umožňují pohled jak na celý les, tak i na každičký strom v něm.

+ Version control (kontrola verzí) – přehled o více verzích datových sad.

+ Vexed (spornost) – odpovídá potenciálu data science zvládnout složité problémy.

+ Vibrant (plnost) – data science poskytuje náhledy, postřehy, nápady a podporu ve všem snažení řešit big data.

+ Vogue (móda) – ze strojového učení, umělé inteligence a další pojmů souvisejících s big daty se stává módní trend.

51V – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Verbosity, Voluntariness, Versatility, Vagueness, Viability, Virtuosity, Visibility, Valor, Varnish, Versed, Vault, Voodoo, Veil, Vulpine, Verdict, Vet, Vane, Vanilla, Victual, Vantage, Varmint, Vivify, Vastness, Voice, Vaticination, Veer, Voyage, Varifocal, Version control, Vexed, Vibrant, Vogue, Verification, Vitality, Violation, Verve, Venturesomeness (2019)

Přes 50V identifikovali ve své práci Khan et al. (2019).

Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Verbosity, Voluntariness, Versatility, Vagueness, Virtuosity, Viability, Visibility, Valor, Varnish, Versed, Vault, Voodoo, Veil, Vulpine, Verdict, Vet, Vane, Vanilla, Victual, Vantage, Varmint, Vivify, Vastness, Voice, Vaticination, Veer, Voyage, Varifocal, Version control, Vexed, Vibrant, Vogue – viz předchozí definice (42V a 17V)

+ Verification (ověření) – big data je možné ověřit jejich skutečností, přesností nebo platností.

+ Vitality (vitalita) – popisuje dynamicky se měnící prostředí big dat, ve kterém se musí analytické a prediktivní modely neustále aktualizovat, aby nám neunikly příležitosti.

+ Violation (porušení ochrany) – big data pomáhají vládě a organizacím s odhalováním zločinů, podvodů a teroristických útoků, stejně tak i s posilováním veřejných služeb.

+ Verve (elán) – odpovídá zápalu pro věc, nadšení atd., vyplývající z big dat, když jsou přístupné prakticky všude a kdykoli pro všechny.

+ Venturesomeness (odvážnost) – big data s sebou přináší vždy něco nového

56V – Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Verbosity, Versatility, Vulnerability, Visibility, Valor, Varnish, Versed, Vault, Voodoo, Veil, Vulpine, Verdict, Vet, Vane, Vanilla, Victual, Vantage, Varmint, Vivify, Vastness, Voice, Vaticination, Veer, Voyage, Varifocal, Version control, Vexed, Vibrant, Vogue, Verification, Vitality, Valences, Virtual, Viability, Virility, Vendible, Vanity, Voracity, Visual, Vincularity, Veritable, Violable (2020)

Hussien (2020) publikoval článek o tom, jak se k původním 3V z roku 2001 postupem času přidávaly další charakteristiky, až vzniklo neuvěřitelných 56V.

Volume, Velocity, Variety, Variability, Veracity, Viscosity, Virality, Validity, Volatility, Visualization, Vocabulary, Venue, Value, Vagueness, Vulnerability, Viability, Visibility, Valor, Verbosity, Versatility, Varnish, Versed, Vault, Voodoo, Veil, Vulpine, Verdict, Vet, Vane, Vanilla, Victual, Vantage, Varmint, Vivify, Vastness, Voice, Vaticination, Veer, Voyage, Varifocal, Version control, Vexed, Vibrant, Vogue, Verification, Vitality – viz předchozí definice

+ Valences (mocenství) – míra určující hustotu daných dat.

+ Virtual (virtuální) – organizace a další skupiny mohou těžit z virtualizace big dat, protože je to opravňuje k použití všech shromážděných datových aktiv k dosažení různých cílů a záměrů.

+ Virility (mužnost) – u big dat se říká, že se tvoří sama, což znamená, že čím víc big dat máte, tím více big dat získává sílu a moc.

+ Vendible (prodejnost) – samotná existence zákazníka je velmi důležitá pro big data, což dokazují příklady některých známých způsobů obchodování s daty uživatelů, jak uvádí Hussien (2020).

+ Vanity (marnivost) – představuje spokojenost big dat, že mají efekt na ostatní jednotlivce.

+ Voracity (nenasytnost) – big data jsou potenciálně tak nenasytná, že mohou dosáhnout vlivu, správy a možnosti konzumovat sebe sama.

+ Visual (vizuálnost) – v současnosti žijeme v době prohlížení, sledování a výměny vizuálních výstupů, tzn. fotografií a videí, interaktivních výstupů a dalších objektů prostřednictvím internetu.

+ Vincularity (vinkularita) – ve svém přesném významu implikuje propojení nebo konektivitu. Tato myšlenka je především relevantní v dnešní éře, kdy je vše celosvětově propojeno prostřednictvím internetu.

+ Veritable (opravdovost) – data jsou pojmenovanou věcí, nejsou falešná, nereálná nebo imaginární.

+ Violable (porušitelnost) – big data obsahují i tzv. porušitelná data, které jsou náchylné k poškození.

Zmíněné charakteristiky byly přeloženy podle Shafer (2017), Farooqi et al. (2019), Hussien (2020) a Khan et al. (2019).

Po zpracování této části se zjistilo, že existuje celkem 62 charakteristik big dat do roku 2021, kdy se konkrétně jedná o 61V a 1C parametr, které byly pro lepší přehlednost zpracovány do Tabulky 1.

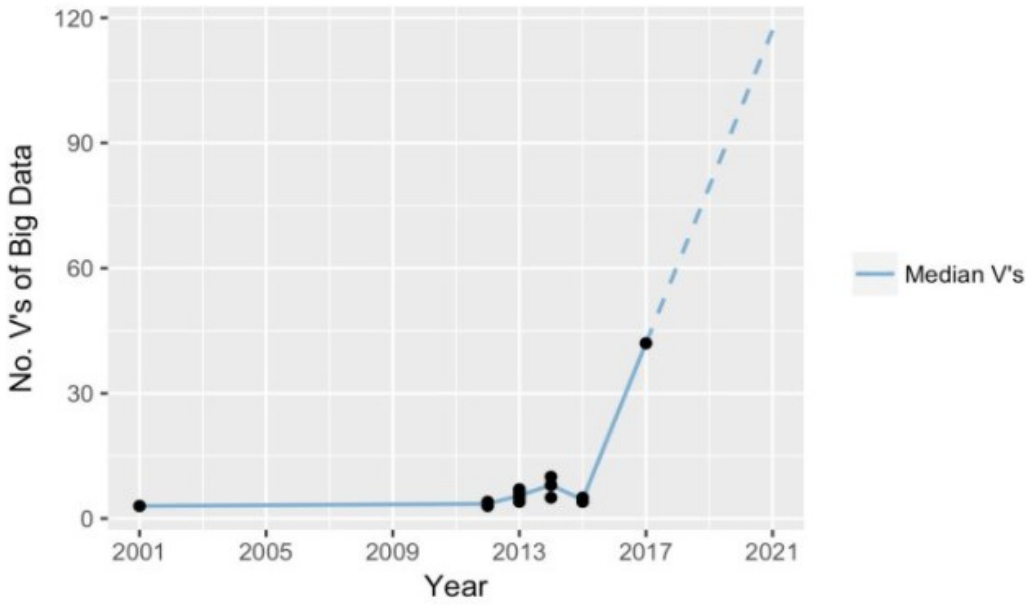
Tabulka 1: Kompletní přehled charakteristik big dat spolu s rokem jejich první zmínky

#	Charakteristika big dat	Vznik
1	Volume	2001
2	Velocity	2001
3	Variety	2001
4	Veracity	2011
5	Value	2012
6	Viability	2013
7	Visualization	2013
8	Variability	2013
9	Viscosity	2014
10	Virality	2014
11	Validity	2014
12	Volatility	2014
13	Vulnerability	2017
14	Vocabulary	2014
15	Venue	2014
16	Vagueness	2014
17	Verbosity	2017
18	Voluntariness	2017
19	Versatility	2017
20	Virtuosity	2017
21	Visibility	2017
22	Valor	2017
23	Varnish	2017
24	Versed	2017
25	Vault	2017
26	Voodoo	2017
27	Veil	2017
28	Vulpine	2017
29	Verdict	2017
30	Vet	2017
31	Vane	2017

#	Charakteristika big dat	Vznik
32	Vanilla	2017
33	Victual	2017
34	Vantage	2017
35	Varmint	2017
36	Vivify	2017
37	Vastness	2017
38	Voice	2017
39	Vaticination	2017
40	Veer	2017
41	Voyage	2017
42	Varifocal	2017
43	Version control	2017
44	Vexed	2017
45	Vibrant	2017
46	Vogue	2017
47	Verification	2019
48	Vitality	2019
49	Violation	2019
50	Verve	2019
51	Venturesomeness	2019
52	Virtual	2020
53	Valences	2020
54	Virility	2020
55	Vendible	2020
56	Vanity	2020
57	Voracity	2020
58	Visual	2020
59	Vincularity	2020
60	Veritable	2020
61	Violable	2020
62	*Complexity*	*2017*

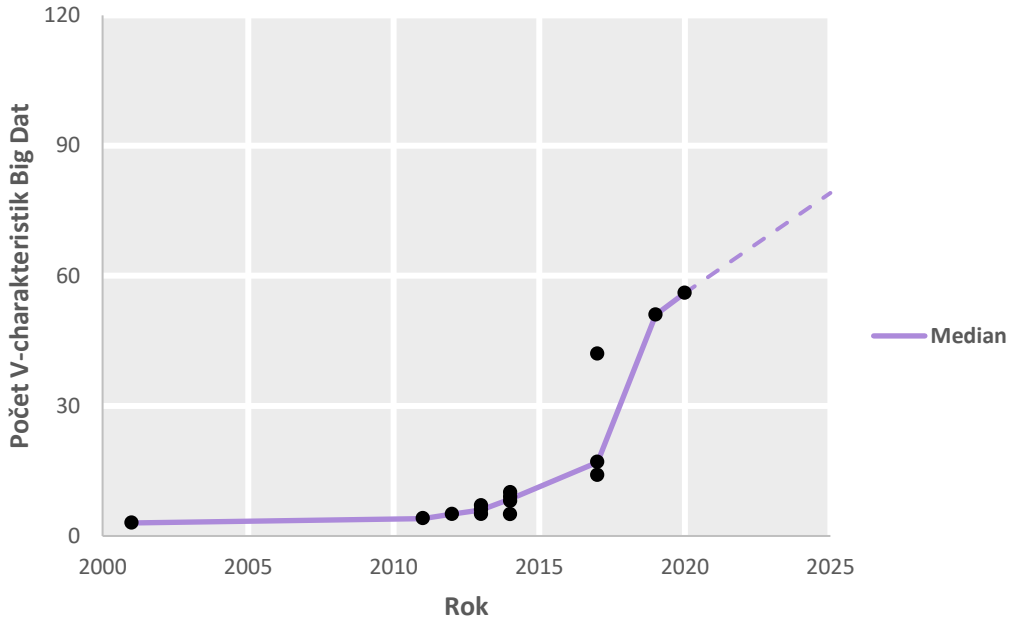
Zdroj: vlastní zpracování

Shafer (2017) ve svém článku uvedl vývoj charakteristik do té doby a s humorem uvedl, že pokud to takhle půjde dál, bude v roce 2021 už 120 charakteristik, viz predikce na Obrázku 1.



Obrázek 1: Vývoj charakteristik za roky podle Shafera (2017)

Ovšem pokud se přehled existujících charakteristik z této kapitoly dá dohromady, zjistí se, že jich je do roku 2021 jen 56 s tím, že celkový počet je rozšířen o dalších pět, tj. celkem 61V (viz Tabulka 1), což odpovídá polovině jeho predikce. Pokud by se ale ve stejném duchu udělal vývoj charakteristik big dat z tohoto přehledu, zjistilo by se, že jsme na cestě k 79V, kterých by se za stejného tempa dosáhlo v roce 2025. Tato úvaha je zachycena na Obrázku 2 přerušovanou čarou.

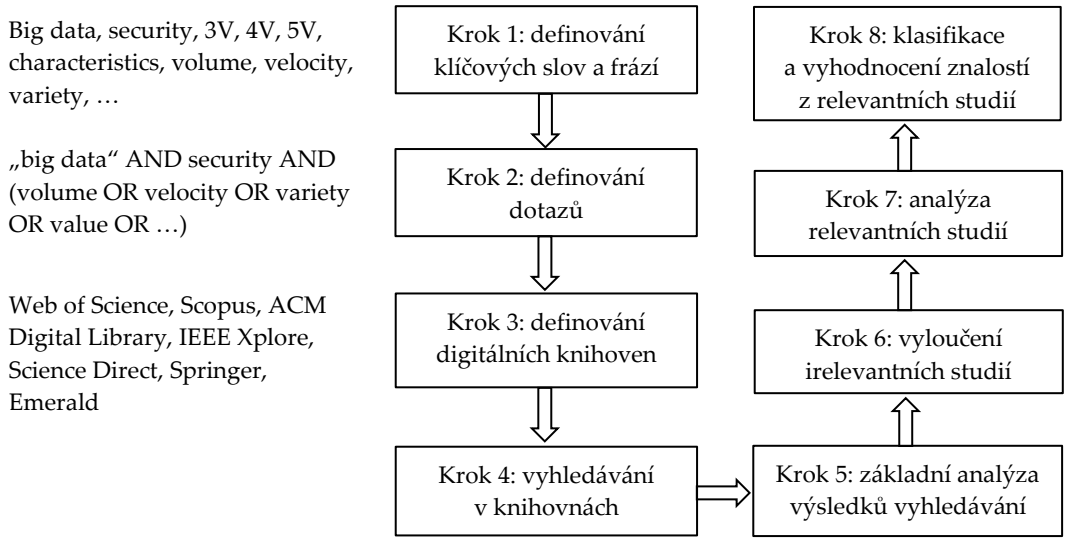


Obrázek 2: Vývoj charakteristik big dat

Zdroj: vlastní zpracování

2 METODIKA A POSTUP ŘEŠENÍ

Splnění cíle této práce bylo provedeno s využitím systematické rešerše literatury, jejíž kroky byly definovány na základě publikací od Webster a Watson (2002) a Timmins a McCabe (2005). Následně byly podrobně rozepsány metodické kroky práce, které se skládají z procesu vyhledávání a výběru literatury, její analýzy a poté syntézy výsledků. Celý postup je zachycen na Obrázku 3.



Obrázek 3: Metodické kroky práce

Zdroj: vlastní zpracování

2.1 Vyhledávání v knihovnách

Jako klíčová slova sloužily nalezené charakteristiky *V* a slova *big data* a *security*. Aby se dosáhlo správného výsledku, byly vytvořeny řetězce klíčových slov ve formě dotazů: *„big data“ AND security AND volume*, *„big data“ AND security AND velocity* atd., viz Tabulka 2 níže s celkovými počty výsledků v daných databázích.

Jako cílové databáze a knihovny byly zvoleny Scopus, Web of Science, ACM, IEEE Xplore a Science Direct. Důvodem pro výběr právě těchto knihoven bylo to, že obsahují velké množství zdrojů vhodných pro splnění cíle práce a usnadňují konkrétní vyhledávání. Dalšími databázemi, které byly zprvu vnímány jako vhodné možnosti, byly ještě Springer a Emerald, ale ty se nakonec pro finální vyhledávání nepoužily kvůli nedostatečné filtraci výsledků.

V knihovnách Web of Science, ACM a IEEE se prohledávala všechna pole pro požadované řetězce, zatímco u databází Scopus a Science Direct bylo možné vyhledávání zaměřit jen na název článku, abstrakt a klíčová slova.

Většina studií zabývající se bezpečností big dat ve vazbě na konkrétní charakteristiky byla především ve formě příspěvků z různých konferencí a článků z časopisů.

Tabulka 2: Výsledky vyhledávání ve vybraných knihovnách

Kombinace klíčových slov	Web of Science	Scopus	ACM Digital Library	Science Direct	IEEE Xplore
big data AND security AND volume	784	883	2 407	52	468
big data AND security AND velocity	198	245	650	21	144
big data AND security AND variety	614	608	2 767	61	322
big data AND security AND veracity	74	97	256	17	45
big data AND security AND value	1 409	1 089	5 587	87	626
big data AND security AND viability	29	18	575	14	14
big data AND security AND visualization	331	377	1 950	28	294
big data AND security AND variability	80	55	2 629	5	26
big data AND security AND viscosity	0	2	25	0	0
big data AND security AND virality	0	0	374	0	0
big data AND security AND validity	115	76	1 631	15	50
big data AND security AND volatility	33	27	333	4	13
big data AND security AND vulnerability	673	588	1 274	63	352
big data AND security AND vocabulary	21	13	454	0	21
big data AND security AND venue	17	12	325	3	4
big data AND security AND vagueness	6	5	231	2	1
big data AND security AND verbosity	0	0	83	0	0
big data AND security AND voluntariness	0	2	530	2	0
big data AND security AND versatility	13	11	250	1	5
big data AND security AND virtuosity	0	0	2	0	0
big data AND security AND visibility	56	44	1 110	3	26
big data AND security AND valor	0	0	37	1	0
big data AND security AND varnish	0	0	4	0	0
big data AND security AND versed	0	1	79	1	1
big data AND security AND vault	7	5	44	0	3
big data AND security AND voodoo	0	0	5	0	0
big data AND security AND veil	0	0	20	0	0
big data AND security AND vulpine	0	0	1	0	0
big data AND security AND verdict	2	2	41	0	0
big data AND security AND vet	33	0	118	0	2
big data AND security AND vane	0	1	5	0	0
big data AND security AND vanilla	6	3	98	1	2
big data AND security AND victual	0	0	2	0	0

Tabulka 2: Výsledky vyhledávání ve vybraných knihovnách (pokračování)

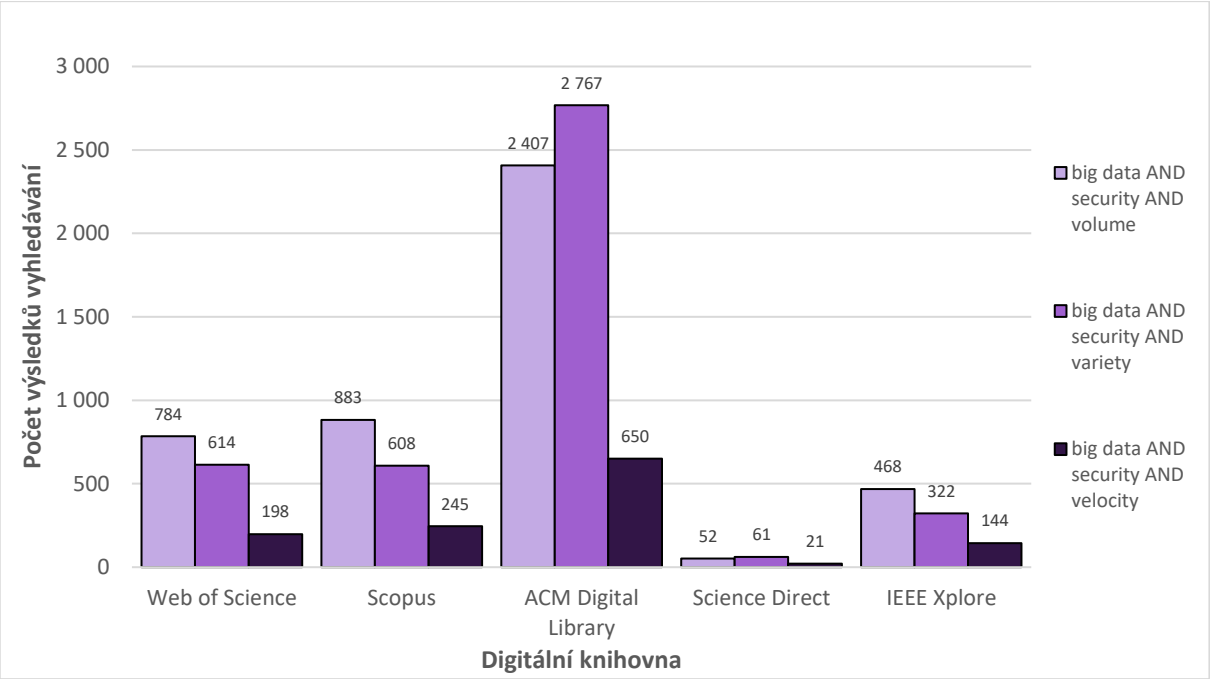
Kombinace klíčových slov	Web of Science	Scopus	ACM Digital Library	Science Direct	IEEE Xplore
big data AND security AND vantage	4	7	73	0	2
big data AND security AND varmint	0	0	1	0	0
big data AND security AND vivify	0	0	2	0	0
big data AND security AND vastness	1	1	1 009	0	2
big data AND security AND voice	69	62	645	2	34
big data AND security AND vaticination	0	0	1	0	0
big data AND security AND veer	1	0	11	0	1
big data AND security AND voyage	4	2	20	0	0
big data AND security AND varifocal	0	0	2	0	0
big data AND security AND version control	35	1	129	1	2
big data AND security AND vexed	1	0	28	0	1
big data AND security AND vibrant	4	2	264	1	1
big data AND security AND vogue	3	4	11	0	0
big data AND security AND verification	474	364	2 417	53	245
big data AND security AND vitality	5	8	893	1	1
big data AND security AND violation	100	79	1 037	3	50
big data AND security AND verve	0	0	7	0	0
big data AND security AND venturesomeness	0	0	3	0	0
big data AND security AND virtual	489	611	2 436	30	316
big data AND security AND valences	5	5	88	2	1
big data AND security AND virility	0	0	1	0	0
big data AND security AND vendible	0	0	799	0	0
big data AND security AND vanity	0	0	22	0	0
big data AND security AND voracity	1	1	5	0	0
big data AND security AND visual	290	253	1 222	17	139
big data AND security AND vincularity	0	0	1	0	0
big data AND security AND veritable	1	1	15	3	0
big data AND security AND violable	0	0	1 021	0	0
big data AND security AND complexity	852	602	4 459	64	506

Zdroj: vlastní zpracování [MS Excel]

Výsledky se dále zaznamenaly i za každou databázi zvlášť podle roku publikování. Tyto tabulky jsou připojeny k práci v příloze I až V.

2.2 Základní analýza výsledků

Tato část se blíže zaměřuje na statistickou analýzu výsledků získaných z provedených vyhledávání v jednotlivých databází, když nejprve budou podrobně porovnány výsledky vyhledávání všech zahrnutých digitálních knihoven.

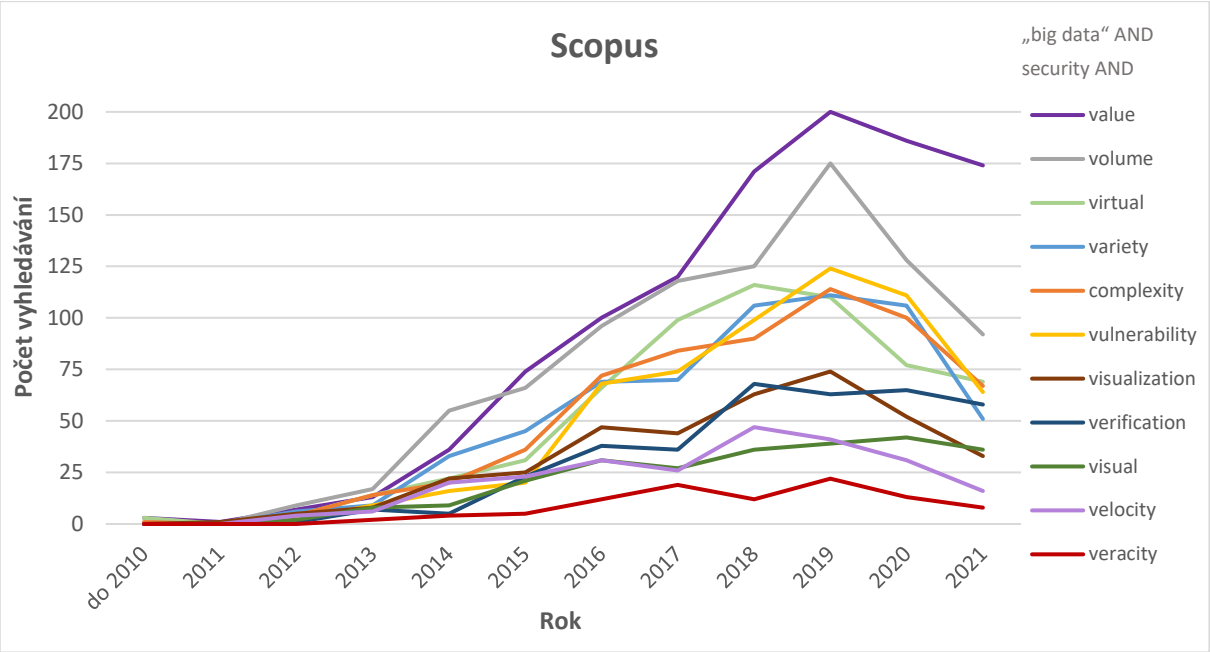


Obrázek 4: Výsledky vyhledávání v knihovnách pro základní 3V

Zdroj: vlastní zpracování [MS Excel]

Graf na Obrázku 4 znázorňuje výsledky hledání z jednotlivých elektronických archivů pro 3V big dat v souvislosti s bezpečností, čímž usnadňuje i jejich porovnání. Web of Science, Scopus a IEEE Xplore zahrnují do svých databází většinu stejných studií, zatímco Science Direct je svou skladbou zdrojů zřejmě zaměřen na jiné oblasti než big data. Databáze ACM je naopak zaměřená na oblast ICT, když obsahuje i jiná média jako například videa, čímž převyšuje ve výsledcích hledání ostatní knihovny. S ohledem na výsledek analýzy budou dále podrobněji rozebrány výsledky z databází Scopus a Web of Science. Databáze ACM není zobrazena z důvodu nepřehlednosti výsledného grafu.

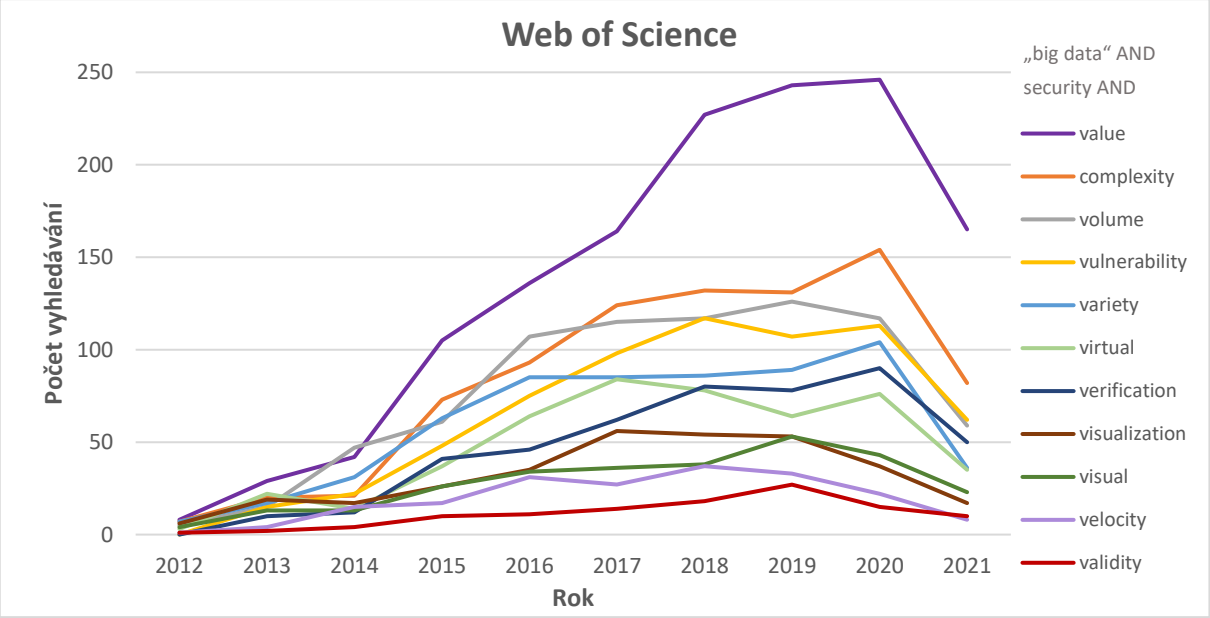
Při bližším průzkumu ročních tabulek (přílohy I–V) bylo možné zjistit, že většina studií vyšla až po roce 2012, kdy téma big dat nebylo tak rozšířené jako je dnes. Navíc databáze Scopus prakticky neobsahuje žádné publikace o bezpečnosti big dat a jejich charakteristikách do roku 2010, což potvrzuje tuto myšlenku, viz Obrázek 5.



Obrázek 5: Výsledky vyhledávání nejpočetnějších charakteristik big dat v souvislosti s bezpečností v letech 2012-2021 (Scopus)

Zdroj: vlastní zpracování [MS Excel]

V grafu jsou zobrazeny výsledky jen **jedenácti z šedesáti dvou** definovaných řetězců („big data“ AND security AND (nazev_charakteristiky)), které byly v knihovně **nejpočetnější.**



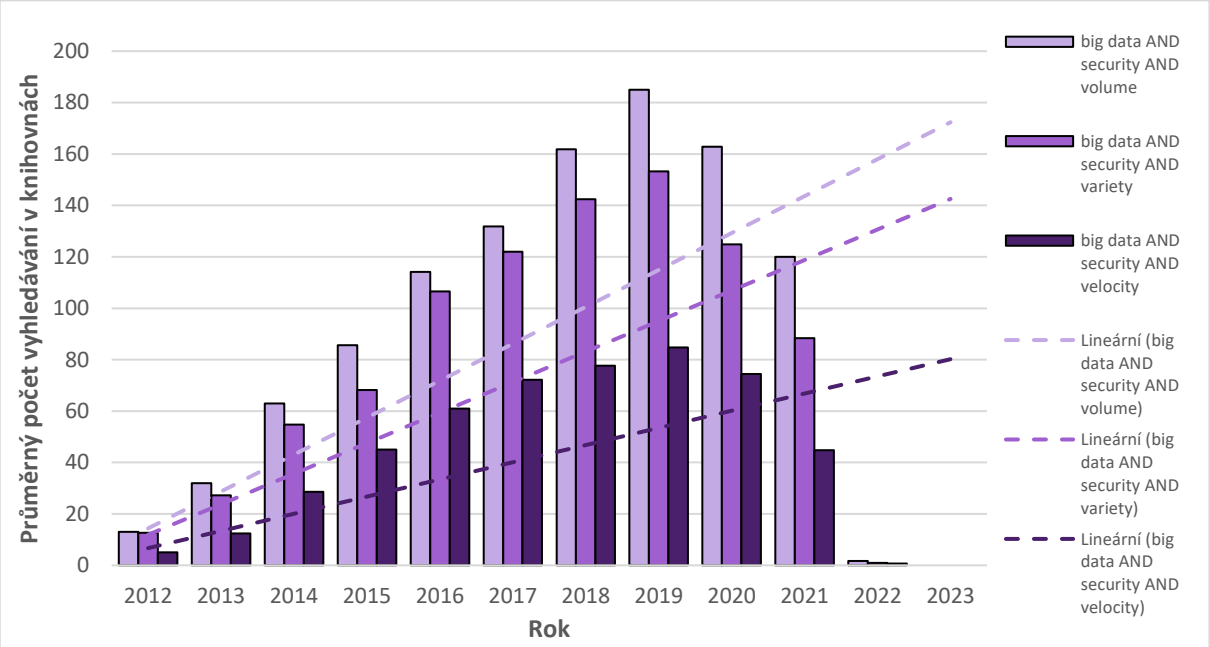
Obrázek 6: Výsledky vyhledávání nejpočetnějších charakteristik big dat v souvislosti s bezpečností v letech 2012-2021 (Web of Science)

Zdroj: vlastní zpracování [MS Excel]

Při porovnání vývoje na Obrázku 6 s předchozím Obrázkem 5 je zřejmé, že big data v souvislosti s bezpečností se začala opravdu výrazněji řešit až od roku 2013, kdy každý následující rok bylo publikováno stále více studií. Tím se dá potvrdit, že téma big dat, ač

vzniklo ve 20. století, se do té doby opomíjelo a hot topicem se stalo teprve před několika lety, především se vznikem nástrojů a platforem, které práci s big data usnadnily.

Na základě této informace je možné sestavit graf průměrných výsledků vyhledávání všech knihoven od roku 2012, a podívat se tak na meziroční vývoj existujících studií zaměřených na bezpečnost big dat. Zároveň slouží i k předpovědi možného vývoje, tj. v následujících letech se může předpokládat další rozvoj a přírůstek nových studií zabývajících se touto problematikou. Vše je zachyceno na Obrázku 7.



Obrázek 7: Průměrné výsledky vyhledávání za roky 2012-2022 [2023] pro 3V (Web of Science, Scopus, ACM, IEEE, Science Direct)

Zdroj: vlastní zpracování [MS Excel]

2.3 Identifikace relevantních studií

Díky provedené analýze se z prvotního vyhledávání vyřadily publikace, které vyšly před rokem 2012, kdy se teprve začala big data rozšiřovat ve světě, čímž se začaly řešit i výzvy, které s sebou přinášejí.

Další výběr a filtrace výsledků proběhla na základě:

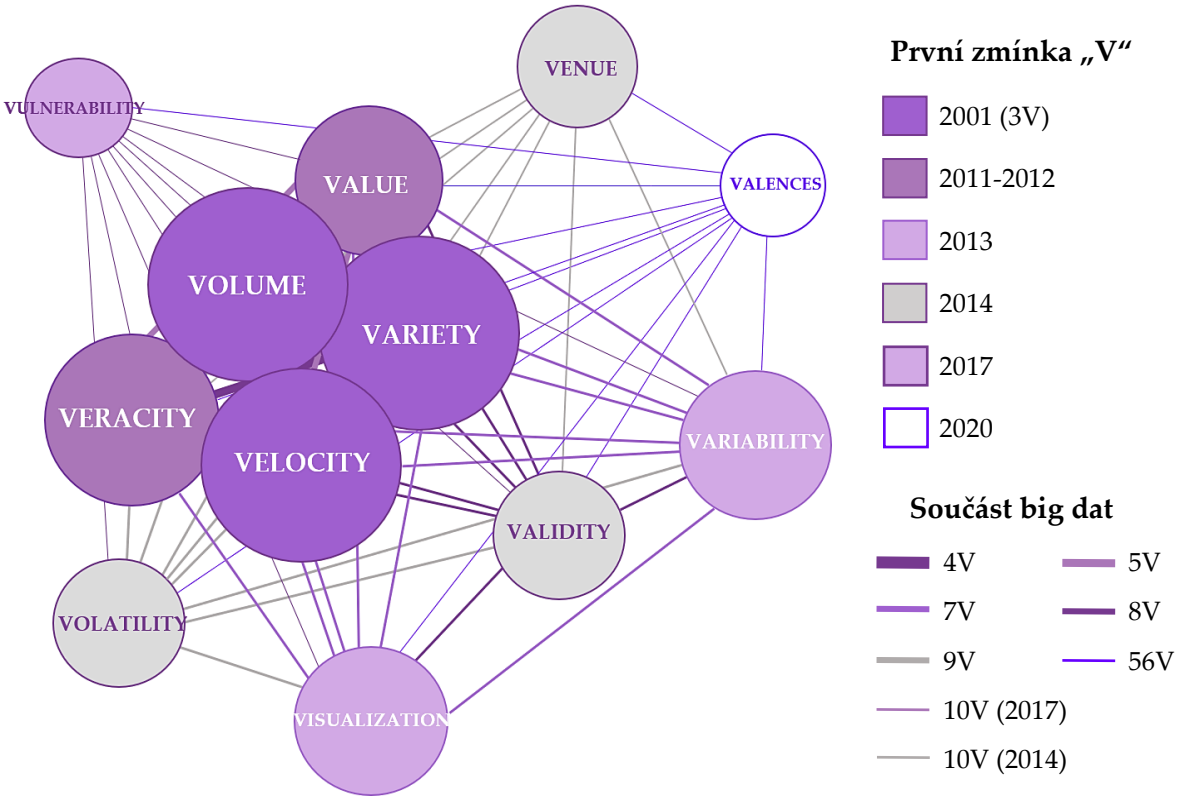
Jedna nebo více charakteristik big dat je uvedena v abstraktu nebo klíčových slovech, nýbrž většina nalezených publikací se nemusí těmito charakteristikami bezprostředně zabývat, jelikož se jedná o běžná slova vyskytující se v souvislosti s daty a bezpečností.

Byl vytvořen řetězec pro bližší vyhledávání, kdy se výběr zúžil ještě na publikace obsahující („big data“ AND security) mezi klíčovými slovy a zároveň (volume AND

variety AND velocity) v textu, jakož veškeré studie zmiňující alespoň 3V charakteristiky big dat v kontextu s bezpečností.

I takto vytříděné studie stále přesahovaly počty ve stovkách, a tak konečná selekce se provedla na základě abstraktu publikací, jejich citací a významnosti k cíli práce. Výběr se tím zúžil na 49 relevantních studií (převzatých především z Web of Science, Scopus a IEEE), které jsou dále podrobně analyzovány.

Z bližší analýzy těchto studií dále vyplynulo, že charakteristiky Volume, Variety, Velocity, Veracity, Value, Venue, Validity, Volatility, Variability, Visualization, Valences a Vulnerability se nejčastěji objevují v souvislosti s bezpečností big dat. Na Obrázku 8 je zachycen vztah mezi těmito charakteristikami (součást 3V, 4V atd.), kde velikost kruhu odpovídá době jejich existence/definice, ergo i výskytům ve výsledcích vyhledávání s výjimkou vulnerability, která mimo názvu charakteristiky odpovídá anglickému slovu úzce spjatému s bezpečností. Podrobnější přehled lze nalézt rovněž v příloze VI.



Obrázek 8: Přehled řešených charakteristik big dat

Zdroj: vlastní zpracování (inspirováno Shaferem, 2017)

3 ANALÝZA RELEVANTNÍCH STUDIÍ

Tato kapitola analyzuje vybrané studie s cílem identifikovat a popsat charakteristiky big dat v kontextu bezpečnosti. Nejprve se zaměřuje na bezpečnost obecně a na výzvy související s big daty, a následně vlivům na bezpečnost, které s sebou přináší charakteristiky big dat.

3.1 Bezpečnost big dat

Navzdory všem výhodám, které potenciálně přináší používání big dat, problémy se zabezpečením představují pro tuto problematiku skutečnou výzvu, a to mnohem více, když organizace zpracovává a ukládá velké množství dat. Pokud tedy organizace chce relevantní data dlouhodobě uchovávat, musí počítat s tím, že problematika bezpečnosti dat vyžaduje dlouhodobou koncepci (Chitransh et al., 2017; Moreno et al., 2016; Sánchez a Urquiza, 2019). Tato data se totiž mohou stát cílem útočníků a zločinců, často i v rámci konkurenčního boje (Guo et al., 2020; Hashmani et al., 2018; Sánchez a Urquiza, 2019; Su, 2019), vůči čemuž je třeba bojovat a najít konkrétní řešení (Jadon a Mishra, 2018). Hlavní problémy a řešení týkající se bezpečnostních rizik a ochrany soukromí však nejsou v oblasti big dat stále plně prozkoumány (Venkatraman a Venkatraman, 2019; Ye et al. 2016).

Ochrana soukromí dat je výsadou mít určitou kontrolu nad tím, jak jsou osobní údaje shromažďovány a používány. Obecně to odpovídá schopnosti kontrolovat, co se z dat získá oprávněnými osobami (Jain et al., 2016; Zhu et al., 2019). Je potřeba si to neplést s důvěrností dat. Zveřejnění nějaké informace obsažené v big datech nebo její části, neovlivní její důvěrnost, ale určitě to zasáhne její ochranu (Zhu et al., 2019). Jedním z vážných problémů ochrany soukromí uživatelů je identifikace osobních informací během šíření přes internet (Jain et al., 2016).

Na rozdíl od toho **zabezpečení big dat** odpovídá praxi ochrany informací a informačních aktiv pomocí technologií, procesů a školení před neoprávněným přístupem, prozrazením, narušením, modifikací, kontrolou, záznamem a zničením (Jain et al., 2016).

Zhu et al. (2019) dále uvádí čtyři formy ochrany big dat: *otevřenost*, kdy jsou data veřejně dostupná; *úschova*, kdy se část získaných informací skryje před zbytkem; *anonymita*, data jsou de-identifikována (anonymizována), aby je nešlo identifikovat z dalšího počtu dat; a *izolace*, kdy big data nesmí zahrnovat žádné citlivé údaje, a tak jsou zcela znepřístupněna ostatním.

Většina lidí a společností není připravena se vypořádat se složitostmi, které přichází spolu s big daty, jako je jejich bezpečnost a správa dat obecně. Mimo to mají mnoho různých zdrojů dat, které mohou mít různé zásady zabezpečení a správy (Paryasto et al., 2014). Tradiční mechanismy zabezpečení a ochrany soukromí jsou navíc pro analýzu big dat nedostatečné (Terzi et al., 2015).

Obecně se bezpečnost dat skládá z důvěrnosti a ochrany soukromí, integrity, dostupnosti, autentizace, autorizace, nepopiratelnosti a řízení přístupu (Himthani et al., 2020). To vše by mělo být aplikováno rovněž na big data, samozřejmě s využitím odpovídajících nástrojů, metod, technik atd. (Holubová et al., 2015).

Big data s sebou přináší velkou zodpovědnost (Moreno et al., 2016). Mohou stát v cestě ochraně, bezpečnosti a důvěře, kdy je vláda a organizace dokážou využívat ze špatných důvodů (Suresh, 2014). I přesto, že se většina organizací snaží eliminovat identifikovatelné informace z dat, tak to nemá na těžení identity z dat a ochranu soukromí moc velký vliv. V dnešní éře big dat, kdy nejen vláda, banky, nemocnice, školy a další společnosti mají data obsahující osobní údaje jednotlivců, ale i prakticky celá celosvětová síť, neboť sdílíme data na internetu pomocí sociálních sítí, vyhledávání, online nákupech atd. Citlivé informace mohou být získány ze všech stop, které na internetu zanecháme, od odhalení identity, koníčků, osobnosti až po stravovací návyky (Matturdi, 2014; Su, 2019). Analýza big dat však většinou pomáhá komerčním společnostem získat užitečné informace pro jejich vlastní rozvoj, případně získání konkurenční výhody (Bouleghlimat a Hacini, 2018).

S příchodem big dat se vláda může dovědět o svých občanech téměř cokoliv, což ne vždy přispívá ke spokojenosti veřejnosti. Organizace je zase mohou využít k tomu, aby se dozvěděly více o vlastní pracovní síle, kdy své zaměstnance sledují a měří jejich výkon atd. Big data tímto vylučují soukromí jednotlivce, což není vždy příznivé pro celou naši společnost (Suresh, 2014). Bezpečnost dat souvisí jak s ochranou osobních údajů, tak i s veřejnou informační bezpečností. Zajištění bezpečnosti dat a informací je tak důležitou částí nejen pro další rozvoj počítačových sítí, ale i pro zachování sociální stability a také důvěryhodnosti ze strany zákazníků i občanů (Guo et al., 2020).

Big data nejen zvyšují rozsah výzev souvisejících s ochranou soukromí a bezpečností, ale také vytvářejí nové, ke kterým je třeba přistupovat novým způsobem. Vzhledem k tomu, že organizace či vlády ukládají a analyzují více dat, je k řešení těchto problémů zapotřebí více předpisů. Dosažení bezpečnosti u big dat se proto stalo jednou z nejdůležitějších překážek, které mohou zpomalit šíření technologií, neboť bez odpovídajících

bezpečnostních záruk nedosáhnou big data požadované úrovně důvěry (Chen a Yan, 2016; Moreno et al., 2016; Zhang, 2018).

Bezpečnostní rizika neustále narůstají, neboť organizace čelí vnitřním i vnějším hrozbám, a tak je nezbytné chránit informace považované za kritická aktiva. Jejich únik, poškození nebo ztráta by mohla přerušit činnost organizace, paralyzovat služby nebo způsobit vážné ekonomické škody (Sánchez a Urquiza, 2019, Su, 2019).

Situace v oblasti bezpečnosti a soukromí se ještě zhoršuje objemem (**volume**), rychlostí (**velocity**) a rozmanitostí / typem (**variety**) big dat (Khan et al., 2018; Khan et al., 2019; Tiwari et al., 2015; Zhu et al., 2019), které navíc každou chvílí narůstají a vznikají tak stále nové požadavky na bezpečnost (Singh, 2014).

Organizace Cloud Security Alliance (CSA) uvádí deset problémů v oblasti zabezpečení big dat, jimiž jsou (CSA, 2013): 1) zabezpečení výpočtů v distribuovaných programovacích rámcích; 2) bezpečnostní best practices pro nerelační databáze dat; 3) zabezpečení úložiště dat a transakčních protokolů; 4) ověření / filtrace vstupů; 5) monitorování zabezpečení v reálném čase; 6) škálovatelné a zároveň modulární architektury pro zpracování dat a analytika zachovávající soukromí; 7) kryptograficky vynucená kontrola přístupu a bezpečná komunikace; 8) řízení přístupu; 9) pravidelné nezávislé audity a 10) původ dat. Další problémy a nevýhody lze nalézt ve studiích od Jha et al. (2017) a Su (2019).

Tyto problémy lze podle pracovní skupiny organizace CSA, 2013 zabývající se big daty rozdělit na čtyři aspekty: 1) zabezpečení infrastruktury; 2) ochrana dat; 3) správa dat; a 4) integrita a reaktivní zabezpečení (CSA, 2013; Moreno et al., 2016).

3.2 Vliv jednotlivých charakteristik na bezpečnost

Tato část práce je zaměřena především na charakteristiky volume, variety, velocity, veracity, value, venue, validity, volatility, variability, visualization, valences a vulnerability, jelikož, jak bylo na základě prvních kroků rešerše zjištěno, právě tyto charakteristiky jsou nejčastěji diskutovány v souvislosti s riziky a bezpečnostními problémy.

3.2.1 Volume

Je velmi obtížné zajistit bezpečnost big dat kvůli jejich velkému objemu a rychlosti, kdy se v současnosti produkuje přes 40 zettabajtů dat denně (ZB = 1 099 511 627 776 GB) a ročně se tato čísla ještě zvyšují (Hussien, 2020; Paryasto et al., 2014; Singh, 2014; Venkatraman a

Venkatraman, 2019; Yang et al., 2016). A právě kvůli tomu se zvětšuje i pravděpodobnost jejich úniku (Yang et al., 2016). Tradiční bezpečnostní modely se setkávají s problémy při řešení tak rozsáhlých dat (Sahafizadeh a Nematbakhsh, 2015, Terzi et al., 2015).

Dá se říct, že se velké množství dat generuje každou sekundou, s čímž souvisí i bezpečnostní problém posuzování a ověřování zdroje dat a zajištění integrity, že data nebyla změněna neautorizovaným jedincem, a autentizace, že zdroj je opravdu tím, za koho se vydává, což také souvisí i s charakteristikou důvěryhodnosti (veracity). Velký objem dat taky může zpochybnit jejich efektivitu a dostupnost (Jha et al., 2017). Navíc představuje problémy pro analýzu a jejich zpracování, především s ohledem na rostoucí složitost a časové požadavky (Chandra et al., 2017).

Pro poskytovatele dat je téměř nemožné dohlížet nebo kontrolovat všechna data, která poskytují. Navíc velký objem dat zvyšuje riziko úniku informací a jejich používáním lze identifikovat nebo předvídat chování lidí, což může narušit soukromí jednotlivců. Kromě toho stávající bezpečnostní strategie, jako je pravidelné sledování, monitorování, audit nebo bezpečnostní skenování, už nejsou dostatečné a jejich implementace pro big data jsou nákladné a komplikované (Ye et al., 2016).

Charakteristika *volume* u big dat dále ovlivňuje bezpečnost a soukromí v dalších dvou aspektech, kdy: 1) jsou uloženy na více místech (serverech, uzlech, clusterech, cloudu atd.), kde standardní databázové systémy a softwarové nástroje nejsou schopny nepřetržitě monitorovat a prosazovat standardizované bezpečnostní protokoly; 2) jakékoli selhání clusteru nebo uzlu může ovlivnit datové transakce a výkon v rámci tolerančních časových lhůt a jsou náchylné ke zranitelnosti (Venkatraman a Venkatraman, 2019).

Big data se také mohou stát nositelem pokročilého viru (Yang et al., 2016), kdy například skrytý malware v takovémto velkém množství dat představuje obrovskou překážku pro jeho odhalení a tím i pro celkovou bezpečnost (Su, 2019).

3.2.2 Velocity

Velocity způsobuje ještě závažnější problémy pro bezpečnost a ochranu soukromí big dat (Ye et al. 2016) a jejich analýzu, kdy je rychlost tvorby dat mnohem rychlejší než tempo analýzy (Chandra et al., 2017; Paryasto et al., 2014), a je třeba ji sladit s rychlostí zpracování a schopnostmi počítačových systémů v reálném čase (Venkatraman a Venkatraman, 2019; Yang a Jia, 2016). Pro mnoho organizací je rychlá odezva a dosažitelnost výsledků výzvou (Quasim a Meraj, 2017).

Ještě důležitější než objem úložiště, které se dá zvýšit, je rychlost generování nových dat. I když jsou data dostupná, jestliže je nelze zpracovat v reálném čase, obchodní příležitosti mohou být ztraceny. Pokud se například předpovědi počasí zpozdí kvůli pomalejší rychlosti zpracování, jež se nemůže rovnat rychlosti přijatých dat, má to vliv na správnost rozhodnutí, které jsou třeba učinit v daný čas. Velocity big dat má tak dopad na bezpečnost a ochranu, kdy se vyžadují rychlejší kryptografické algoritmy, které mají držet krok s tempem zpracování transakcí v reálném čase (Venkatraman a Venkatraman, 2019).

Navíc rychle rostoucí a opakující se data vyžadují nerelační databáze, kdy je potřeba vyvíjet distribuované programovací rámce s ohledem na bezpečnost a soukromí (Ye et al., 2016). Kromě toho může hacker snáze spouštět pokročilé trvalé hrozby (Advanced Persistent Threats; APT), zatímco jej je obtížné odhalit tradičními strategiemi ochrany (Venkatraman a Venkatraman, 2019; Ye et al., 2016). Velocity s sebou rovněž přináší i bezpečnostní výzvy týkající se efektivity a dostupnosti dat a výsledků, kdy je nutné řešit celý životní cyklus těchto dat (Jha et al., 2017; Xiaorong et al., 2018).

3.2.3 Variety

Dalším problémem, který se může v analýze a bezpečnosti big dat objevit je, že se data liší v jejich formátu a obsahu a většina z nich je nestrukturovaných, tzn. zvuk, videa, obrázky, data ze sensorů či třeba logy ze sociálních sítí, satelitů a jiných zdrojů (Himthani et al., 2020; Paryasto et al., 2014; Venkatraman a Venkatraman, 2019; Yang a Jia, 2016). Navíc každý zdroj má vlastní zásady ochrany soukromí a zabezpečení, kdy například data ze sociálních sítí mohou obsahovat citlivé informace (Venkatraman a Venkatraman, 2019).

Parametr různorodosti se nevztahuje pouze na různé reprezentace dat, ale odkazuje také na prostředky a způsoby, kterými jsou tato data přenášena. Takováto rozmanitost big dat vnáší požadavky na vhodnou klasifikaci dat a řízení přístupu pro různé zdroje, typy a formáty dat pro zajištění soukromí a bezpečnosti (Venkatraman a Venkatraman, 2019; Xiaorong et al., 2018).

Další překážkou této vlastnosti je, že zvyšuje složitost rozhodování a efektivního využívání big dat, protože je nejdříve nutné řešit sjednocení a standardizaci dat (Khan et al., 2018). Z bezpečnostního hlediska pak tato vlastnost big dat přináší rizika týkající se dostupnosti vhodných dat pro zpracování a efektivity celého procesu (Jha et al., 2017). Zároveň je nutné zmínit i to, že zabezpečení různých zdrojů, ze kterých big data proudí, opět zvyšuje náročnost celé bezpečnostní politiky organizace. A rozličné typy a zdroje dat

navíc vyžadují různou ochranu, takže většina tradičních zabezpečení je nevhodná. (Hussien, 2020).

Obecně mají 3V vliv na bezpečnost a ochranu big dat v důsledku různých zdrojů dat, jejich architektuře a konfiguraci, prostředí pro získávání dat a přesunech dat mezi různými úložišti, ať už interními nebo cloudovými řešeními. Ve všech těchto případech je správné nastavení bezpečnostních mechanismů klíčové (Tiwari et al., 2015).

3.2.4 Value

U big dat se dále liší i způsob jejich využití a potenciální přidaná hodnota, což opět může představovat překážku pro analýzu a jejich zabezpečení, zejména s ohledem na stanovení priorit toho, co bude řešeno dříve a má potenciálně větší význam (Paryasto et al., 2014).

Potencionálně vysoká hodnota obecně přitahuje hackery, kteří pokud úspěšně zaútočí na databázi, získají velké množství dat a potenciálně hodnotné nebo i citlivé informace, čímž snižují náklady na vyřešení problému. To může vést k vyšší pravděpodobnosti útoku na big data, kde se očekává vyšší hodnota při jejich úniku (Ye et al., 2016). Při úniku dat tak hrozí ztráta jejich hodnoty. Hackeři také mohou změnit data nebo získat utajené informace, čímž opět klesá jejich hodnota, ale i další parametry jako důvěryhodnost (Jha et al., 2017).

Účelem data miningu (DM) je analýza dat a extrakce užitečných informací z datových sad, což přispívá k předpovědi budoucnosti a rozhodování. Z prediktivní analýzy big dat mohou jednotlivci a organizace získávat výhody, ale na druhou stranu budou snadněji identifikovatelní (Ye et al., 2016). Navíc vzhledem ke zpracovávání velkého objemu big dat vystavují tyto analýzy proces dolování znalostí z databází a dalších datových úložišť diskuzím, jak zachovat soukromí cílových datových souborů (Quasim a Meraj, 2017). Měl by proto být zvážen kompromis mezi zachováním soukromí a výhodami plynoucími z využívání dat (Ye et al., 2016).

Big data obecně představují zdroj hodnoty pro organizace, ale její uvědomění si může často vést ke zvýšení dalších rizik. Například kombinování zdrojů osobních údajů může odhalit citlivé informace, u nichž hrozí jejich zveřejnění. Podobně současný trend nabízení zákazníkům personalizované produkty a služby založené na analýze big dat vyvolává mnoho obav souvisejících s ochranou soukromí, krádeží identity, nezákonnou diskriminací, nespravedlivou klasifikací, a dokonce i s vykořisťováním zranitelných osob (Günther et al., 2017; Tonidandel et al., 2016).

Z bezpečnostního hlediska *value* přináší rizika týkající se důvěrnosti, autenticity, dostupnosti a integrity dat (Jha et al., 2017).

3.2.5 Veracity

Aby na big datech mohla být provedena smysluplná analýza, je důležité mít správná a přesná data, která lze zpracovat v požadovaném množství a v pravý čas. Jakákoli data, která jsou nadbytečně neúplná nebo mají chyby, nemohou vést k správným a přínosným analytickým výsledkům (Su, 2019; Venkatraman a Venkatraman, 2019). Čím více big dat máme, tím se veracity (věrohodnost) snižuje, což vede ke slabší důvěře a spolehlivosti dat (Jha et al., 2017; Venkatraman a Venkatraman, 2019), kdy i analýza se stane problémem (Chandra et al., 2017). Navíc, jak bylo zmíněno výše, i zde se vyskytuje problém se zajištěním integrity a autenticity (Jha et al., 2017; Su, 2019).

Dosažení důvěryhodnosti dat je obtížné vzhledem k tomu, že se musí kombinovat různé techniky jako je sémantická integrita, elektronické podpisy a různé techniky kvality. Předpoklad důvěryhodnosti dat může vyžadovat inteligentní kontrolu procesu správy dat, která má zase dopad na soukromí (Quasim a Meraj, 2017).

Výsledky DM se obvykle využívají pro komerční nebo veřejná rozhodování, takže hlavním problémem zůstává, zda jsou výsledky DM důvěryhodné. Veracity prostupuje celým řetězcem big dat, od pravosti zdrojových dat přes integritu zpracovávaných až k důvěryhodnosti publikovaných dat (Ye et al., 2016).

Pokud jde o citlivost dat, organizace by měla uplatňovat účinné strategie k ochraně dat a dodržení regulačních požadavků (Khan et al., 2018). Zlepšením pravdivosti big dat by bylo možné kontrolovat obchodní rizika spojená s rozhodováním, což by mělo dopad na zásady zabezpečení a ochrany osobních údajů s ohledem na prosazování kvalitních dat prostřednictvím vhodné aplikace zajišťující ověření vlastnictví a zdrojů dat a metod pravidelné kontroly přístupu (Venkatraman a Venkatraman, 2019).

Data získaná ze sociálních sítí mohou mít vliv i na národní bezpečnosti, kdy vláda čerpá ze sociálních sítí názory lidí, které mohou pomoci předejít různým útokům a posílit důvěryhodnost vlády. Navíc pomáhá krizovému managementu při katastrofách, kdy je cílem detekovat abnormální chování a zajistit v reálném čase vyhodnocování (monitoring) potenciálních hrozeb. Avšak i když sledování jednotlivců vládami a jinými organizacemi může vést ke zlepšení veřejné kontroly a bezpečnosti, ubírá to jednotlivcům pocity svobody a soukromí (Günther, 2017).

3.2.6 Validity

Před jakoukoli analýzou většina organizací stráví hodně času nad čištěním dat, k čemuž jsou vyžadovány řádné postupy přípravy a správy dat, aby byla zachována jejich platnost v procesu kontroly kvality. To vyžaduje správné řízení dodavatelů a partnerů, kteří prosazují ochranu datového dodavatelského řetězce (Venkatraman a Venkatraman, 2019). Tuto vlastnost big dat lze tedy řešit nastavením a monitorováním požadavků, kterou na datové sady mají organizace a jednotlivci, např. frekvence aktualizací, porovnání s jinými daty (v čase), správnost dat atd. (Simon, 2013).

3.2.7 Volatility

Další výzvou pro big data je jejich nestálost (*volatility*), neboť data, která jsou soukromá dnes, nemusí být zítra. Data také postupem času zastarávají a většinou se neví, jak dlouho jsou určitá data důležitá pro dané případy užití v organizaci (Khan et al., 2018; Simon, 2013). Je tak potřeba stanovit pravidla a regulovat dostupnost dat pro plynulý chod procesů organizace, aby se zabránilo tomu, že na data bude pohlíženo jako na nestálá (Khan et al., 2018). Opět i u této charakteristiky je nutné pracovat s dlouhodobou strategií a vlivem času na big data.

3.2.8 Variability

Všech sedm výše zmíněných charakteristik může být ovlivněno variabilitou big dat, která odkazuje na nekonzistenci zdrojů při načítání dat do datového úložiště v různých rychlostech, formátech a typech, což může představovat problémy pro bezpečnost a soukromí (Venkatraman a Venkatraman, 2019). Tato vlastnost big dat se stala výzvou především kvůli stále se zvyšujícímu používání digitálních médií, což představuje hlavní důvod nárůstu zatížení dat (Khan et al., 2018).

Na druhou stranu však může detekovat anomálie a odlehlé hodnoty, které mohou organizaci prospět, neboť lze zachycené informace o variabilitě spojit s daty v úložišti a využít je pro získání informací z big dat (Firican 2017; Venkatraman a Venkatraman, 2019.) Odpovědné osoby by proto měly zohledňovat variabilitu big dat ve svých postupech pro zabezpečení v rámci různých protokolů auditu a metod monitorování (Venkatraman a Venkatraman, 2019).

3.2.9 Visualization

Vizualizace big dat pomáhá k efektivnímu rozhodování (Venkatraman a Venkatraman, 2019), ale vizualizační nástroje čelí podle Firicana (2017) technickým obtížím jako je omezená paměť, špatná funkčnost, čas odezvy atd., kdy tradiční grafy nedokážou správně ukázat bilióny informací, kterými big data oplývají. Navíc manuální vytahování informací z mnoha různých zdrojů big dat do těchto nástrojů bývá časově náročné a mnohdy i stresující (Hussien, 2020). Proto se dnes mnoho sofistikovaných vizualizačních nástrojů integruje s modely pro analýzu dat, aby bylo možné vytvářet smysluplnější grafické interpretace big dat.

S tím však přichází další výzva pro zachování soukromí, kdy není snadné grafy úplně zprostit všech potenciálně citlivých informací (Quasim a Meraj, 2017). Na druhou stranu může vizualizace pomoci analyzovat stav či předpovídat trend vývoje bezpečnostních incidentů (Yang a Jia, 2016).

K výstupům z vizualizačních nástrojů by proto měly být stanoveny zásady ochrany soukromí spolu s přiřazením kontroly přístupu a oprávnění podle uživatelských rolí a odpovědností (Venkatraman a Venkatraman, 2019).

3.2.10 Venue

Big data se dále ukládají na různé platformy, jak je definováno charakteristikou *venue*, s čímž vzniká další problém bezpečnosti, kdy jsou big data většinou ukládána a přenášena napříč mnoha cloudy a dalšími systémy, což zvyšuje rizika zabezpečení – sdílení dat mezi mnoha sítěmi (Quasim a Meraj, 2017).

Ovšem sdílení dat je důležité, neboť se jedná o přístup k datům pro partnery, klienty, poskytovatele a zaměstnance, ale to s sebou přináší hrozby jako odhalení důvěrných informací (jako jsou procesy a metody výroby), nebo nezákonný přístup do síťového provozu. Příkladem může být zdravotní sektor, kdy jsou data sdílena mezi několika nemocnicemi a farmaceutickými laboratořemi kvůli výzkumu a analýze, což však může ovlivnit pacientovo soukromí i přesto, že jsou data anonymní, a to tak, že se najdou závislosti mezi zdravotními záznamy a vzájemnými pojištěními (Benjelloun a Lahcen, 2015).

3.2.11 Valences

Valences přeneseně odpovídá propojení mezi atomy v molekule, kdy se jedná o míru určující poměr mezi skutečně připojenými datovými položkami a počtem spojení, která se mohou v rámci sběru dat vytvořit. Odhalení nepřímých souvislostí mezi datovými položkami je sice obtížné, ale přinášejí organizaci potenciální přidanou hodnotu (Venkatraman a Venkatraman, 2019). To však stejně jako v předchozím případě může vyústit v prolomení ochrany soukromí dat.

Venkatraman a Venkatraman (2019) proto navrhují, že by řízení bezpečnosti a ochrany mělo zachovat stejnou úroveň výkonu jak pro současný, tak pro budoucí růst systémů big dat.

3.2.12 Vulnerability

Vulnerability je tou nejdůležitější vlastností big dat související s bezpečností, ochranou soukromí a technologickými riziky, která vyplývá z potenciálu big dat, jež jsou shromažďovány z různých produktů a služeb internetových aplikací, sociálních sítí a zařízeních využívající internet věcí (IoT) (Dehghantanha a Choo, 2019).

Zranitelnost big dat souvisí s technologiemi, procesy a řízením big dat, kdy vznikají skuliny v zabezpečení a ochraně. Vulnerability pak prochází všemi zmíněnými charakteristikami big dat, které vyžadují nepřetržité monitorování (Venkatraman a Venkatraman, 2019). Pravidelné kontroly zranitelnosti a penetrační testy by měly být vyvinuty se zohledněním na jedinečné vlastnosti big dat. Identifikace zranitelných míst úniku citlivých dat je nezbytná spolu s vhodnými opatřeními pro kontrolu důvěrnosti, integrity a dostupnosti systémů big dat (Simon, 2013; Venkatraman a Venkatraman, 2019).

4 POŽADAVKY A NÁVRHY POSTUPŮ PRO ZAJIŠTĚNÍ BEZPEČNOSTI BIG DAT

Tato kapitola se blíže věnuje identifikaci požadavků a návrhům postupů při zohlednění charakteristik big dat.

Pro implementaci zabezpečení a ochrany soukromí big dat je nejlepším řešením samotná legislativa než jakékoli bezpečnostní technologie (Matturdi et al., 2014). Bohužel ale zákony nedokážou držet krok s vývojem technologií a liší se v jednotlivých zemích, což ještě ztěžuje problematika cloud computingu, kdy jsou servery s daty také umístěny v různých zemích, kde se na ně vztahuje jiná legislativa.

Typickými technikami, které se běžně používají pro zabezpečení jakýchkoliv dat jsou kontroly přístupu, heslování a dvoufázová autentizace. Ty ale přináší pouze nízkou úroveň bezpečnosti, pokročilejším řešením je šifrování (Matturdi et al., 2014). Tradiční řešení jsou však často nedostatečná, zvlášť pro big data, jelikož lze prolomit šifrovací schémata, přístupová oprávnění, firewally atd., dokonce lze znovu identifikovat anonymizovaná data. Kvůli tomu dochází k vývoji pokročilých technik a technologií pro zabezpečení, monitorování a audit procesů big dat v rámci infrastruktury, aplikací a dat (Dehghantanha a Choo, 2019; Terzi et al., 2015).

Zabezpečení a ochrana soukromí u big dat je velmi složitý úkol pro jakoukoli organizaci, jelikož dat je opravdu mnoho a jejich zdrojem může být prakticky cokoli. Kvůli tomu může lehce dojít k prolomení v zabezpečení (Chitransh et al., 2017) a data mohou přenášet APT, které jsou těžko odhalitelné v reálném čase (Hussien, 2020). Proto je vždy nutné využívat adekvátní bezpečnostní technologie a další metody vytvořené pro potřeby big dat (Matturdi et al., 2014), např. pokud se jedná o tzv. Apache Hadoop ekosystém (Zikopoulos a Eaton, 2011; Zhao et al., 2014), kde je zajištěna kompatibilita mezi jednotlivými platformami, nástroji a službami.

Pro zajištění bezpečnosti se dále provádí různé výzkumy a studie, které mají ověřit dané postupy v praxi (Chitransh et al., 2017).

CSA (2013) rozděluje bezpečnost big dat do čtyř kategorií: 1) zabezpečení infrastruktury, 2) ochrana dat, 3) správa dat a 4) integrita a reaktivní zabezpečení. Význam těchto aspektů se liší podle zkoumaného oboru, například pro big data ve zdravotnictví je ochrana citlivých údajů jedním z nejdůležitějších cílů (Ye et al., 2016).

Řešení problémů v oblasti zabezpečení a ochrany soukromí big dat obvykle vyžaduje: 1) *modelování* – formalizace modelu hrozby, který pokrývá většinu scénářů kybernetických útoků nebo úniků dat; 2) *analýza* – nalezení řešitelných řešení na základě modelu hrozby a 3) *implementace* – implementace řešení do stávajících infrastruktur (CSA, 2013).

Následující Tabulka 3 shrnuje řešené problémy v bezpečnosti a ochrany soukromí big dat do kategorií a zaměření a odkazuje na související studie, které se touto problematikou zabývají. Dále následují návrhy postupů, jak danou problematiku řešit.

Tabulka 3: Seznam kategorií řešených v rámci bezpečnosti big dat

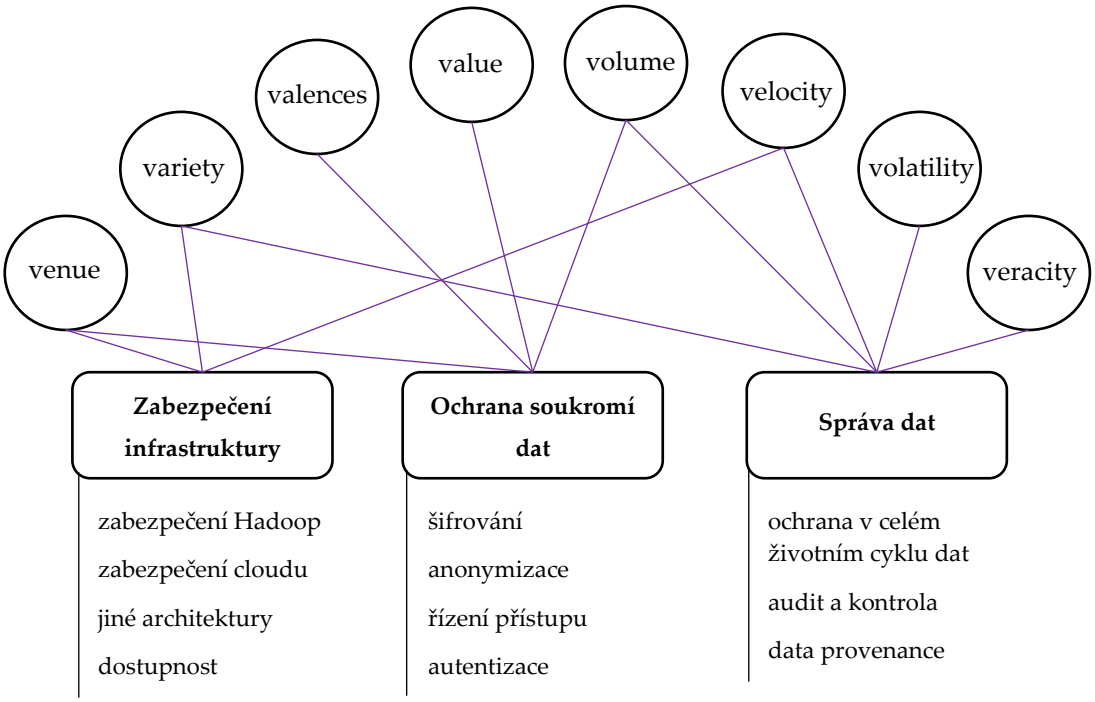
Kategorie, zaměření	**Stručný popis**	**Reference**
zabezpečení infrastruktury a architektury	zabývá se bezpečností přenosových kanálů a prostředí, kde jsou data uložena a zpracovávána, každý typ architektury vyžaduje jiný přístup	
zabezpečení distribuovaných výpočtů		(Bouleghlimat a Hacini, 2018), (CSA, 2013), (Chandra et al., 2017), (Jadon a Mishra, 2018) (Sahafizadeh a Nematbakhsh, 2015)
zabezpečení Apache Hadoop		(Dupré a Demchenko, 2016), (Chandra et al., 2017), (Chen a Yan, 2016), (Jadon a Mishra, 2018), (Jain et al., 2016), (Moreno et al., 2016), (Sahafizadeh a Nematbakhsh, 2015), (Terzi et al., 2015), (Zhao et al. 2014)
zabezpečení cloudu		(Bouleghlimat a Hacini, 2018), (Dasoriya, 2017), (Chen a Yan, 2016), (Cheng et al., 2015), (Chitransh et al., 2017), (Jain et al., 2016), (Sahafizadeh a Nematbakhsh, 2015), (Terzi et al., 2015)
jiné architektury		(Bouleghlimat a Hacini, 2018), (Dasoriya, 2017), (Dupré a Demchenko, 2016), (Farooqi et al., 2018), (Chen a Yan, 2016), (Jadon a Mishra, 2018), (Sahafizadeh a Nematbakhsh, 2015), (Singh, 2018)
zabezpečení komunikace		(Benjelloun a Lahcen, 2015), (Dupré a Demchenko, 2016), (Moreno et al., 2016), (Sánchez a Urquiza, 2019)
zajištění dostupnosti systémů big dat		(Bouleghlimat a Hacini, 2018), (Moreno et al., 2016)
ochrana soukromí dat	velikost big dat zvyšuje riziko úniku informací, pro jejich ochranu je třeba zajistit integritu, autentičnost, nepopiratelnost, důvěrnost a dostupnost	
technika anonymizace, de-identifikace dat		(Benjelloun a Lahcen, 2015), (Bouleghlimat a Hacini, 2018), (Hamthani et al., 2020), (Chen a Yan, 2016), (Irudayasamy a Arockiam, 2015), (Jain et al., 2016), (Jha et al., 2017), (Koo et al., 2020), (Matturdi et al., 2014), (Paryasto et al., 2014), (Terzi et al., 2015)
tokenizace		(Paryasto et al., 2014)
ovládací prvky cloudu		(Paryasto et al., 2014)

šifrovací algoritmy		(Benjelloun a Lahcen, 2015), (Bouleghlimat a Hacini, 2018), (Dasoriya, 2017), (Dupré a Demchenko, 2016), (Hamthani et al., 2020), (Hussien, 2020), (Chandra et al., 2017), (Chen a Yan, 2016), (Chen et al., 2021), (Jain et al., 2016), (Jha et al., 2017), (Koo et al., 2020), (Matturdi et al., 2014), (Mustafa et al., 2018), (Sahafizadeh a Nematbakhsh, 2015), (Sánchez a Urquiza, 2019), (Su, 2019), (Venkatraman a Venkatraman, 2019), (Yang et al., 2016), (Yu, 2016), (Zhu et al., 2019)
autentizace		(Benjelloun a Lahcen, 2015), (Bouleghlimat a Hacini, 2018), (Dasoriya, 2017), (Jadon a Mishra, 2018), (Moreno et al., 2016), (Mustafa et al., 2018), (Sahafizadeh a Nematbakhsh, 2015), (Zhu et al., 2019)
zlepšení povědomí o bezpečnosti		(Benjelloun a Lahcen, 2015), (Dongpo, 2018), (Guo et al., 2020), (Moreno et al., 2016)
řízení přístupu		(Benjelloun a Lahcen, 2015), (Hussien, 2020), (Jha et al., 2017), (Koo et al., 2020), (Matturdi et al., 2014), (Sahafizadeh a Nematbakhsh, 2015), (Venkatraman a Venkatraman, 2019), (Yang et al., 2016), (Zhu et al., 2019)
správa dat	velký objem a komplexnost big dat vede k novým výzvám, kdy je obtížné je bezpečně získávat, přesouvat, skladovat, spravovat a analyzovat	
ochrana celého životního cyklu dat	životní cyklus se skládá z různých fází od sběru dat, uložení, distribuce, používání, archivaci a odstranění	(CSA, 2013), (Dupré a Demchenko, 2016), (Hussien, 2020), (Koo et al., 2020), (Ma a Wu, 2014), (Mathur a Gupta, 2020), (Matturdi et al., 2014), (Moreno et al., 2016), (Ye et al., 2016), (Sánchez a Urquiza, 2019), (Su, 2019), (Terzi et al., 2015), (Yang et al., 2016), (Xiaorong et al., 2018), (Zhu et al., 2019)
data mining		(Hussien, 2020), (Chandra et al., 2017), (Venkatraman a Venkatraman, 2019), (Ye et al., 2016), (Yu, 2016)
datová analytika		(Alsuhibany, 2016), (Dasoriya, 2017), (Hussien, 2020), (Chitransh et al., 2017), (Jain et al., 2016), (Koo et al., 2020), (Terzi et al., 2015)
cloud computing	velice úzce souvisí s big daty	
- přístup k datům		(Benjelloun a Lahcen, 2015), (Chandra et al., 2017), (Chen a Yan, 2016), (Cheng et al., 2015), (Jadon a Mishra, 2018), (Sahafizadeh a Nematbakhsh, 2015)
- audit pro ukládání dat		(Bouleghlimat a Hacini, 2018), (Chandra et al., 2017), (Chen a Yan, 2016), (Koo et al., 2020) (Su, 2019), (Terzi et al., 2015)
zabezpečení datových uložišť		(CSA, 2013), (Chen a Yan, 2016), (Guo et al., 2020), (Jain et al., 2016), (Moreno et al., 2016), (Sahafizadeh a Nematbakhsh, 2015), (Su, 2019), (Venkatraman a Venkatraman, 2019)

technologie data provenance (původ a změny dat)		(Alguliyev a Imamverdiyev, 2014), (Bouleghlimat a Hacini, 2018), (CSA, 2013), (Hussien, 2020), (Matturdi et al., 2014), (Venkatraman a Venkatraman, 2019)
key management		(Jadon a Mishra, 2018), (Terzi et al., 2015)
integrita a reaktivní bezpečnost	big data pochází z různých zdrojů, a tak je třeba zajistit jejich integritu, aby mohla být správně využita, a to i pro kontrolu, zda není systém napaden	
monitoring v reálném čase		(Benjelloun a Lahcen, 2015), (CSA, 2013), (Guo et al., 2020), (Terzi et al., 2015)
následné zpracování incidentů		(Alguliyev a Imamverdiyev, 2014), (Zhu et al., 2019)
detekce škodlivého URL		(Hashmani et al., 2018), (Lin et al., 2013)
integrita dat		(Bouleghlimat a Hacini, 2018), (Jain et al., 2016), (Mathur a Gupta, 2020), (Mustafa et al., 2018)
síťové zabezpečení		(Guo et al., 2020), (Matturdi et al., 2014), (Su, 2019), (Terzi et al., 2015), (Yang a Jia, 2016),
technologie umožňující zpracování big dat		(Alguliyev a Imamverdiyev, 2014), (Janssen a Grady, 2013), (Suresh, 2014)
kybernetická bezpečnost v souvislosti s big daty	big data jsou častým terčem kybernetických útoků díky svým charakteristikám (převážně jejich hodnotě, objemu a rychlosti)	
obrana před útoky		(Mustafa et al., 2018), (Sagiroglu a Sinanc, 2013), (Sánchez a Urquiza, 2019), (Terzi et al., 2015), (Yang et al., 2016), (Zhu et al., 2019)
identifikace hrozeb a rizik		(Terzi et al., 2015), (Tiwari et al., 2015),
machine learning (strojové učení)	může pomoci s vývoji modelů pro zajištění bezpečnosti a ochrany big dat	(Himthani et al., 2020), (Singh, 2014), (Yu, 2016)
předpisy, zásady a zákony		(Benjelloun a Lahcen, 2015), (Dasoriya, 2017), (Dongpo, 2018), (Koo et al., 2020), (Ma a Wu, 2014), (Moreno et al., 2016), (Yang et al., 2016)

Zdroj: vlastní zpracování

Zjištění prezentovaná v tabulce výše lze dále využít pro rozšíření modelu od Ye et al. (2016). Jeho aktualizace je zachycena na Obrázku 9, když propojuje klíčová témata v bezpečnosti spolu s hlavními charakteristikami big dat, jež je ovlivňují. Samotná charakteristika vulnerability (zranitelnosti) pak prochází všemi zobrazenými.



Obrázek 9: Kategorizace výzev u bezpečnosti big dat

Zdroj: vlastní zpracování, inspirováno Ye et al. (2016)

Vazby vyplývají z definic charakteristik a předchozí kapitoly, kde se tyto charakteristiky zpracovávaly z pohledu bezpečnosti.

4.1 Zabezpečení infrastruktury

Pro zabezpečení infrastruktury systémů big dat musí být zabezpečeny především distribuované výpočty a datová úložiště (CSA, 2013). Při řešení otázky bezpečnosti big dat z hlediska **zabezpečení infrastruktury** bylo v posledních letech publikováno mnoho článků představujících různé návrhy a možnosti, jak to zajistit.

Hlavním tématem u zabezpečení infrastruktury bývá zajištění bezpečnosti pro Apache Hadoop, což je rámec vyvinutý společností Apache, který umožňuje distribuované zpracování velkých datových sad napříč propojenými výpočetními zdroji pomocí programovacích modelů a algoritmů. Bezpečnostní problémy související s touto technologií jsou tak často diskutovány výzkumníky a někteří dokonce navrhují i metody pro lepší zabezpečení Hadoop systému s názvem G-Hadoop (Moreno et al., 2016). Více o tomto modelu je možné si přečíst v publikaci od Zhao et al. (2014) s názvem „A security framework in G-Hadoop for big data computing across distributed Cloud data centres“.

V Hadoopu navíc není autentizace a data se nešifrují, takže útočník může ukrást jaká data chce. S tím pak souvisí vylepšení autentizace, která má za úkol identifikovat subjekt a udělit přístup jen autorizovaným a neautorizované subjekty nevpustit do systému.

Společnost Hortonworks vyvinula také Apache Ranger, který slouží k zajištění ochrany dat, auditu, autorizace a autentizace pro centrální správu bezpečnostních politik systému (Chen a Yan, 2016). Mezi další projekty řešící bezpečnost v Apache Hadoop ekosystému patří Apache Knox, Apache Sentry nebo Apache Accumulo.

Další překážkou bývá dostupnost systémů big dat. V tomto případě se navrhuje mít najednou aktivních více namenodes (master serverů) nebo vytvořit novou infrastrukturu úložného systému za účelem zlepšení dostupnosti a odolnosti proti poruchám (Moreno et al., 2016).

Dále se doporučuje zabezpečení distribuovaných výpočtů pomocí protokolu Kerberos či jiných podobných metod. Důležité je také zajistit bezpečnost nerelačních databází NoSQL aj. pomocí šifrovacích algoritmů. Více návrhů a možností zmiňují také další studie, viz Tabulka 3.

4.2 Ochrana dat

Po zajištění bezpečné infrastruktury je důležité **ochránit soukromí** samotných dat**.** Pro jejich zabezpečení se musí nejprve zabezpečit informace o nich, např. umístění, velikost, vazby, hodnocení atd., a citlivá data musí být chráněna pomocí kryptografie a podrobných kontrol přístupů (CSA, 2013). Co se však týká bezpečnosti big dat, tak tradiční zabezpečení jako je šifrování atd. zpomalují výkon a jsou časově náročné v kontextu s big daty (Benjelloun a Lahcen, 2015; Dupré a Demchenko, 2016), navíc nejsou ani účinné právě kvůli jejich charakteristikám. Z čehož plyne, že opravdu jen malá část dat je testována a zpracována pro bezpečnostní účely, takže útoky jsou většinou zaznamenány až po způsobení a rozšíření škody (Benjelloun a Lahcen, 2015).

Šifrování je však klíčovým prvkem adekvátní bezpečnostní strategie, ale přináší s sebou několik překážek a výzev, které se těžko překonávají (Dupré a Demchenko, 2016). Autoři dále zpracovali čtyři různé strategie, jak ochránit systém, data či oboje naráz.

Quasim a Meraj (2017) zmiňují iniciativu, která sleduje myšlenku využití šifrovacího mechanismu pro zabezpečení dat na cloudu a tím zabránění narušování soukromí a bezpečnosti tak, že navrhne inteligentní algoritmy pro dotazování na šifrovaná data, aniž by ovlivnila přesnost odpovědi. Chen a Yan (2016) dále uvádí užitečná schémata

a techniky pro zabezpečení big dat a ochranu soukromí, jako je homomorfní šifrování a schémata bezpečného ukládání a správy dat v cloudu. Chen et al. (2021) ve své studii zmiňuje další dvě cesty, jak efektivně zašifrovat big data – technologii bloom filter a šifrovací algoritmus na základě redundance dat.

Aby byla zachována bezpečnost a soukromí dat, měla by být zajištěna jejich anonymita bez vlivu na výkon systému a kvalitu dat. Ovšem to není tak snadné, neboť běžné techniky anonymizace jsou založeny na časově náročných výpočtech a iteracích, což může ovlivnit konzistenci dat a zpomalit systém, zvlášť při práci s velkým množstvím dat. Navíc je obtížné zpracovávat a analyzovat anonymizovaná data, neboť to vyžaduje složitější algoritmy a náročnější výpočty (Benjelloun a Lahcen, 2015).

Zhang et al. (2014) a Irudayasamy a Arockiam (2015) však navrhují metodu anonymizace, která kombinuje dvě nejčastěji používaná schémata: specializaci shora dolů (TDS) a generalizaci zdola nahoru (BUG).

Mezi další návrhy, jak ochránit soukromí údajů, patří zavedení nových právních předpisů a posílení smýšlení veřejnosti, zvlášť v dnešní době, kdy každý chodí na sociální sítě a sdílí své osobní údaje bez obav z toho, jak s takovými daty bude naloženo (Guo et al., 2020; Moreno et al., 2016).

4.3 Správa dat

Správa obrovského objemu dat vyžaduje škálovatelná a distribuovaná řešení, a to jak pro zabezpečení datových úložišť, tak pro umožnění efektivních auditů a ověřování dat.

Vzhledem k velkému objemu big dat, který navíc exponenciálně roste, si uložiště big dat vyžádaly automatické vrstvení. Toto řešení však nesleduje, kde jsou jaká data uložena, což představuje výzvu pro bezpečnost daného datového uložiště. Nové mechanismy jsou nezbytné pro řešení neoprávněných přístupů a udržení konstantní dostupnosti dat (CSA, 2013). Nejen, že je důležité najít řešení pro bezpečné ukládání dat, ale je nutné i vědět, jak správně data získat a spravovat (Moreno et al., 2016).

Pro řešení těchto problémů výzkumníci navrhují mechanismus, který vytváří parametr pro měření přijatelné úrovně soukromí, čímž může zajistit ochranu vlastníků dat. Cheng et al. (2015) dále zmiňují přístup, kdy je možné zabezpečit uložiště rozdělením dat, která jsou uložena v databázích a systémech big dat, do sekvenčních částí a uložit je k různým poskytovatelům cloudů. Jiní autoři zase navrhují šifrování dat přímo v datovém skladu, kdy tak budou data chráněna i během přenosu, nahrávání a stahování (Yang et al., 2016).

Základním požadavkem na systém ukládání big dat je ochrana soukromí jednotlivce. Pro splnění tohoto požadavku existují určité mechanismy. Odesílatel může například zašifrovat svá data pomocí šifrování veřejným klíčem (PKE) takovým způsobem, že data může dešifrovat pouze platný příjemce (Jain et al., 2016). Autoři také zmiňují přístupy k ochraně soukromí uživatele při ukládání dat do cloudu: homomorfním šifrování, využitím hybridních cloudů, šifrováním atributů a cest úložiště.

Data se dále zpracovávají pro získání znalostí, k čemuž slouží metody data miningu. To sice umožňuje objevovat cenné informace z big dat, ale může to také vést k extrakci citlivých informací, které by měly zůstat skryté. Pro řešení tohoto problému se vyvíjejí postupy aplikovatelné v data miningu, které by měly zajistit i ochranu soukromí (Ye et al., 2016). DM se také využívá k identifikaci zranitelnosti big dat a rizik, a k předvídání potenciálních útoků (Hussien, 2020).

V této oblasti bezpečnosti big dat figuruje i tzv. data provenance. To je termín používaný pro metadata obsahující původ dat, která big data produkují. Jedná se o jiný druh dat, který vyžaduje vlastní ochranu. Venkatraman a Venkatraman (2019) a Hussien (2020) navrhují přizpůsobit tuto technologii k efektivnímu využití analytických metod pro detekci anomálií ve fázi získávání big dat. Několik algoritmů pro detekci anomálií je zmíněno v článku od Hussien (2020). Metadata původu big dat také násobí složitosti práce s nimi. Analýzy pro identifikování závislostí metadat kvůli zabezpečení u aplikací big dat jsou rovněž výpočetně náročné (CSA, 2013).

4.4 Integrita a reaktivní zabezpečení

Integrita a reaktivní zabezpečení v sobě zahrnují ověřování, filtrování a monitorování v reálném čase (CSA, 2013; Terzi et al., 2015). Integrita obecně chrání data před neoprávněnými změnami během jejich životního cyklu a spolu s důvěrností a dostupností tvoří tři základní dimenze zabezpečení. Zajištění integrity je u big dat zásadní, ale často těžce dosažitelné (Moreno et al., 2016).

Proudící data, pocházející z různých zdrojů, musí být zkontrolována z hlediska integrity (CSA, 2013; Chandra et al., 2017), kdy je třeba vědět, že nepochází ze škodlivého zdroje. Tím i vyfiltrovat data, která mohou představovat rizika (Chandra et al., 2017). Data mohou být použity i k analýze bezpečnostních incidentů v reálném čase, aby byla zajištěna dostupnost a bezpečnost infrastruktury (CSA, 2013).

Zabezpečení je třeba monitorovat v reálném čase, aby bylo možné zjistit, zda je systém napaden (Moreno et al., 2016). Je to vyžadováno i pro dohled nad infrastrukturou big dat a také pro detekci předem definovaných parametrů v datech. Kromě toho ale i pro rozhodování o datech proudících do systému. To by mohlo potenciálně odhalit problémy s infrastrukturou a snížit falešně pozitivní a skutečně negativní hodnoty dat (Chandra et al., 2017), které jsou ignorovány kvůli omezené kapacitě lidí pro analýzu. To se ještě znásobuje u big dat kvůli jejich velkému objemu a rychlosti datových toků. Na druhou stranu však technologie big dat umožňuje rychlé zpracování a analýzu různých typů dat, čímž je možné je využít k detekci anomálií v reálném čase (CSA, 2013).

Jedna z navržených metod pro řešení této problematiky zabezpečení je detekce a filtrování škodlivých URL adres v aplikacích big dat. Více o tom ve studii „Malicious URL Filtering – A Big Data Application“ od Lin et al. (2013)

5 NÁVRHY A DOPORUČENÍ

Tato kapitola se věnuje návrhům a doporučením vyplývajícím z provedené analýzy. Na úvod je nutné zmínit, že problematika big dat je velmi široká, což ostatně dokazuje i počet nalezených a diskutovaných charakteristik, a proto je nutné zvolit konkrétní modely, resp. situace, pro které budou návrhy a doporučení relevantní.

Prvním krokem pro ochranu, který by se měl zajistit ve všech organizacích, je zabránění neoprávněného přístupu např. správnou autentizací těch, kteří budou s daty pracovat.

Malé organizace a firmy bývají častějším terčem kybernetických útoků, protože nemají tak pokročilé zabezpečení a zkušené pracovníky. Navíc big data jsou stále nová, a ne plně uchopená, takže většina systémů přichází bez vlastního zabezpečení a firmy je tím musí dovybavit. Protože riziko útoku je zde větší než u jiných organizací, a také s ohledem na nízké rozpočty na bezpečnost, tak by se firmy měly zaměřit na zabezpečení ochrany samotných dat. Nejlepším řešením je šifrování, které se pro zabezpečení dat využívá již dlouho a lze jej využít i pro big data. Ovšem s big daty je to trochu jinak, neboť kvůli jejich charakteristikám jsou tradiční techniky nepoužitelné, takže si firmy musí udělat průzkum ohledně nových algoritmů, které jsou pro jejich potřeby vhodné. Různorodost big dat pak toto komplikuje, kdy je potřeba různé typy dat šifrovat a chránit jinými způsoby. Nicméně u malých organizací se nepředpokládá, že budou mít tak různorodá data a jejich hlavním problémem bude spíše velikost dat, kam je bezpečně uložit a dále spravovat, a jak nastavit postupy pro jejich analýzu. Pokud budou big data šifrována, tak je lze poměrně efektivně dlouhodobě ukládat v cloudových úložištích a přistupovat k nim podle potřeby, včetně využívání dalších distribučních modelů cloud computingu, jako např. infrastruktura jako služba pro analýzy big dat. I zde je však nutné zmínit podmínku vyjednání a uzavření smlouvy s poskytovatelem cloudu.

Samozřejmě pokud má organizace vyšší rozpočet pro bezpečnost, tak si může dovolit vybudovat takové zabezpečení infrastruktury a řízení přístupu pro odrážení samotných pokusů o útoky tak, že to může potenciální útočníky odradit a tím snížit i samotné riziko vniknutí.

Velké organizace a firmy by se měly starat o správu dat, tj. mít komplexní přehled o tom, jaká data mají, kde je mají, v jakých formátech, jaká je jejich potenciální hodnota atd. Díky tomu mohou snadněji identifikovat, popsat a zejména monitorovat toky dat a potenciální zranitelná místa a tím i správně rozhodovat o jejich zabezpečení. Z předchozí kapitoly však víme, že toto není snadný úkol, zvlášť co se týká big dat vzhledem k jejich velkému

objemu, kdy není možné veškerá data sledovat v reálném čase. Organizace využívající Apaxhe Hadoop či jinou architekturu pro zdravotní a finanční záznamy by se měly zajímat i o jeho zabezpečení. To samé platí i pro cloud, kde big data ukládají. Koneckonců se jedná o velké množství dat, a tak je zapotřebí nepodceňovat jejich bezpečnost. Dále lze opět doporučit šifrování dat.

Jelikož u těch organizací lze předpokládat vyšší rozpočty na bezpečnost a zároveň také pracovníky s odpovídajícími znalostmi, tak by zde vždy měly existovat bezpečnostní politiky, které jasně definují postupy pro práci s daty, včetně implementace charakteristik big dat – minimálně těch základních 3V a 4V, a určení odpovědnosti za související činnosti. Kromě citlivých údajů, na které se při manipulaci vztahují odpovídající právní předpisy, lze přepokládat, že i potenciální hodnota dat v těchto organizacích je vysoká a tomu je nutné přizpůsobit zabezpečení a ochranu big dat. Zatímco u malých organizací lze při omezených rozpočtech doporučit zaměření se především na fázi ukládání a archivace dat pro potenciální analýzy, tak u velkých organizací je nutné brát v úvahu celý životní cyklus big dat. Jak je patrné i z Tabulky 3, tak zabezpečení každé fáze je důležité, protože právě v nezabezpečené fázi může dojít k úniku dat nebo k útoku.

Veřejná správa by svou pozornost měla určitě zaměřit na zabezpečení infrastruktury vzhledem ke všem státním tajemstvím a citlivým informacím o všech občanech státu. Je proto důležité útočníky držet mimo od samého začátku a nepustit je dovnitř a držet tak riziko úniku informací na minimu. Například speciálně vyvinutým firewallem pro zabránění útoků a nežádaných vniknutí. Zde je ale potřeba mít na paměti, že i speciálně nastavené firewally představují jen jednovrstevní ochranu, takže při prolomení už nic útočníkovi nebrání. Dalším příkladem může být Zettaset Orchestrator, který byl vyvinut na základě malé ochrany firewallů a přináší dvouvrstvou ochranu pro dat s myšlenkou, že je zapotřebí zabezpečení přenést blíž k samotným datům pro jejich ochranu. Ovšem je i tak dobré mít záložní plán, kdyby náhodou prvotní zabezpečení selhalo z jakéhokoliv důvodu, a tak šifrování samotných dat v systému by nebylo na škodu.

Veřejná správa a její orgány by se dále měla snažit i o de-identifikaci informací z dat a také anonymizaci kvůli ochraně soukromí občanů a sama tyto informace nezneužívat. Nutno však podotknout, že přestože se jedná o adekvátní doporučení, tak není vždy garancí pro bezpečnost. Existuje totiž i tzv. reidentifikace, která stejným způsobem zase získává identifikační údaje zpět, čímž podkopává zajištění maximální ochrany touto technikou. To je i jedna z nevýhod big dat, tzn., že co jde provést jedním směrem k vyššímu zabezpečení, tak mohou útočníci zneužít i opačným směrem pro narušení zabezpečení. Jelikož s big

daty úzce souvisí i problematika cloud computingu, tak pro organizace veřejné správy je nutné zmínit i centralizované řešení cloudu veřejné správy, kam mohou jednotlivé instituce bezpečně ukládat svá data a zároveň využívat služby tohoto cloudu ve formě dalších distribučních modelů, jako např. software jako služba.

Jednotlivec (běžný občan) by se měl ze všeho nejvíc snažit pochopit big data a na základě toho se chovat na internetu a v prostředí počítačových sítí. Ač sám neovlivní, jak jeho data budou zabezpečena a zpracována v organizacích jako jsou nemocnice, pojišťovny či vládní orgány, mimo daných regulací a zákonných pravidel jako je GDPR atd., může své bezpečí zajistit na síti omezením či změnou sdílení na sociálních sítí, prohlížení, nakupování, nahrávání videí a obrázků a veškerých dalších aktivit. Přestože se říká, že co se jednou dostane do sítě tak tam zůstane na vždy, určitě by nebylo na škodu i své aktivity občas promazat a pročistit.

Poslední doporučení, které vyplývá i z podstaty této práce, je směrem k výzkumníkům a dalším osobám, které dále pomáhají rozvíjet tuto oblast. I oni při své práci generují velké objemy různorodých dat, která mají často tu úplně největší hodnotu, protože přispívají k dalšímu rozvoji dané oblasti i života lidí. Proto i oni by měli dodržovat bezpečností opatření a chránit svoje data, aby nebyla ukradena nebo zneužita.

Zabezpečení v prostředí big dat se především musí pravidelně aktualizovat. Protože jak tato práce ukázala, tak počet charakteristik a tím i počet potenciálních zranitelností pořád roste, a zároveň stále více organizací implementuje související postupy a metody big dat do svých strategií rozvoje, což opět může zvětšovat okruh útočníků, kteří mají o cenná data dané organizace zájem.

ZÁVĚR

V této práci byly popsány charakteristiky a vývoj big dat a na základě rešerše literatury identifikovány požadavky a postupy při zohlednění charakteristik big dat a jejich vliv na bezpečnost.

Cíl práce ohledně přehledu vývoje big dat a popisu charakteristik byl splněn v první kapitole. Druhý cíl práce, kde se měly identifikovat postupy při zohlednění charakteristik big dat a jejich vlivu na bezpečnost, je obsažen v rámci kapitol 3-5.

V úvodu práce lze nalézt definici a charakteristiky, které musí existovat, aby se dalo tvrdit, že se jedná právě o big data. Obecně se mluví o 3V, kdy big data musí mít masivní objem, rychlost vytváření a rozmanitost, kdy navíc většina z nich pochází z nestrukturovaných dat, které se těžko analyzují, ukládají a zabezpečují. Po porovnání vývoje bylo zjištěno, že do roku 2020 existuje celkem 62 charakteristik big dat, jež sestávají konkrétně z 61V a 1C parametrů.

V druhé kapitole byla představena metodika a postup řešení práce. Nejprve byly vybrány databáze, v nichž proběhlo vyhledávání. Zpočátku se uvažovalo o sedmi knihovnách (ACM, Springer, Web of Science, IEEE, Scopus, Science Direct a Emerald), ale po bližším náhledu do výsledků vyhledávání byly vyloučeny databáze Springer a Emerald, protože z nalezených výsledků nebylo možné identifikovat počty relevantních publikací a trendy.

Dále byla definována klíčová slova pro vyhledávání, která byla složena z řetězce („big data“ AND security) plus název definovaných charakteristik – volume, variety, value atd. Výsledky hledání byly zaznamenány do šesti tabulek, kdy jedna sloužila pro celkový přehled, a dalších pět obsahovalo záznamy z jednotlivých databází za jednotlivé roky. Na základě těchto tabulek bylo možné vypracovat základní analýzu, která přispěla k lepší specifikaci výběru relevantních studií. Výsledkem bylo 49 tematicky vhodných publikací.

Praktická část práce tyto studie analyzovala. Proběhla identifikace charakteristik big dat, kdy se z rešerše zjistilo, že konkrétně 12 z 62 se nejčastěji řeší v rámci problematiky zabezpečení big dat. Jednalo se o základních 9V (volume, velocity, variety, value, veracity, validity, volatility, variability a visualization) plus tří z jiných verzí, jimiž byly venue, valences a vulnerability, které byly definovány mnohem později. Nejčastěji se však ve studiích v souvislosti s bezpečností probíralo 5V, jelikož se uvažovaly i starší publikace z roku 2012, kdy do té doby jiné definice neexistovaly.

Zjistilo se, že právě velká rychlost tvoření dat představuje velký problém pro zabezpečení, neboť předčítá i rychlost analýzy. Dále, že hodnota big dat je tak velká, že se o ně hackeři často zajímají a hrozí tak i jejich únik. Navíc protože je big dat opravdu mnoho a nelze je všechny v reálném čase monitorovat, není většinou možné včas útoku zabránit a útočníka, případně virus v datech, odhalit. Mimo to je obecně těžké zajistit důvěryhodnost dat, zvlášť když jich je tolik. A čím menší je jejich věrohodnost, tím se zvyšují rizika bezpečnosti a ochrany soukromí, neboť zdroje, a s tím i data, mohou být škodlivé.

V neposlední řadě se identifikovaly další problémy se zabezpečením big dat spolu s návrhy na jejich řešení, na které se dá pohlížet ze 4 různých aspektů. Nejprve se jedná o *zabezpečení infrastruktury*, kde je potřeba existující architektury dovybavit zabezpečením, neboť např. Apache Hadoop nebyl vyvinut se žádným zabezpečením, například novými rámci či architekturami nebo speciálně vyvinutými firewally. Dále je důležité zajistit *ochranu soukromí dat* např. šifrováním či de-identifikací. Třetím hlediskem je zabezpečení *správy dat*, jelikož je třeba chránit data během celého jejich životního cyklu od sběru, uložení, přenosu, využívání a zpracování až po jejich archivaci a odstranění. Čtvrtým a posledním aspektem pro zajištění bezpečnosti big dat je *integrita a reaktivní zabezpečení*, která chrání data před neoprávněnými změnami během jejich životního cyklu a umožňuje monitoring systému v reálném čase, zda nedošlo např. k napadení.

Většina výzkumníků a dalších odborníku se nicméně shodne na tom, že je zapotřebí hlavně zlepšit povědomí veřejnosti o soukromí, neboť lidé sdílejí spoustu osobních věcí na internetu, aniž by se pozastavili nad skutečností, že v této éře big dat se z toho dá zjistit mnoho informací o nich, které by jinak zveřejnit nechtěli.

Také se došlo k závěru, že kombinace technologie s příslušnými opatřeními a předpisy může pomoci vyřešit problém zabezpečení big dat a ochrany soukromí. Po bližší analýze studií se dále vyfiltrovaly kategorie, na které je třeba se zaměřit pro posílení bezpečnosti.

Výsledkem je, že autoři analyzovaných prací došli převážně k podobným zjištěním, a to, že se v rámci tématu bezpečnosti a ochrany big dat stále nachází dost prostoru k dalšímu výzkumu.

Mezi přínosy diplomové práce patří:

- komplexní přehled existujících V (charakteristik) big dat s jejich významem a vymezením,
- vymezení celkového obrazu klíčových problémů souvisejících s bezpečností big dat spolu s návrhy na jejich řešení,

- klasifikace předchozího bodu do bezpečnostních kategorií pro zdůraznění a pochopení důležitých oblastí v rámci zabezpečení big dat,
- doporučení a návrhy na zajištění bezpečnosti big dat pro konkrétní modelové situace vyplývající z analyzované literatury.

Ochrana soukromí, bezpečnost a zabezpečení big dat představují hlavní problémy, o kterých se bude i nadále diskutovat, neboť je to téma, které se stále ještě vyvíjí, takže je třeba vyvinout nové techniky, technologie a řešení.

POUŽITÁ LITERATURA

ALGULIYEV, Rasim a Yadigar IMAMVERDIYEV. 2014. Big Data: Big Promises for Information Security. *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*. DOI: 10.1109/icaict.2014.7035946

ALSUHIBANY, Suliman A. 2016. A space-and-time efficient technique for big data security analytics. *2016 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT)*. DOI: 10.1109/kacstit.2016.7756065

AROCKIA, Panimalar, VARNEKHA Shree.S a Kathrine.A VENESHIA. 2017. The 17 V's of Big Data. *International Research Journal of Engineering and Technology*, 4(9), 329-333.

BENJELLOUN, Fatima-Zahra a Ayoub Ait LAHCEN. 2015. Big Data Security: Challenges, Recommendations and Solutions. *Handbook of Research on Security Considerations in Cloud Computing*, pp. 25-38. DOI: 10.4018/978-1-4666-8387-7.ch014

BIEHN, Neil. 2018. *The Missing V's in Big Data: Viability and Value*. WIRED [online]. c Condé Nast. [cit. 2021-09-30]. Dostupné z: https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/

Big Data Working Group; Cloud Security Alliance. 2013. *Expanded Top Ten Big Data Security and Privacy*. [online] c CSA [cit. 2021-11-20]. Dustupné z: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf

BOULEGHLIMAT, Imene a Salima HACINI. 2018. *Big Data Processing Security Issues in Cloud Environment.* Lecture Notes in Networks and Systems, pp. 27-36. DOI: 10.1007/978-3-319-98352-3_4

CHANDRA, Sudipta, RAY, Soumya a R. T. GOSWAMI. 2017. Big Data Security: Survey on Frameworks and Algorithms. *2017 IEEE 7th International Advance Computing Conference (IACC)*, pp. 48-54. DOI:10.1109/iacc.2017.0025

CHEN, Hanlu a Zheng YAN. 2016. Security and Privacy in Big Data Lifetime: A Review. *Security, Privacy and Anonymity in Computation, Communication and Storage*, pp. 3-15. DOI: 10.1007/978-3-319-49145-5_1

CHEN, Weijie, CHEN, Guodong, ZHAO, Yanheng a Jinghua ZHANG. 2021. Security vulnerability and encryption technology of computer information technology data under

big data environment. *Journal of Physics: Conference Series*, 1800(1), pp. 1-8. DOI: 10.1088/1742-6596/1800/1/012012

CHENG, Hongbing, RONG, Chunming, HWANG, Kai, WANG, Weihong, a Yanyan LI, Y. 2015. Secure big data storage and sharing scheme for cloud tenants. *China Communications*, 12(6), pp. 106-115. DOI: 10.1109/cc.2015.7122469

CHITRANSH, Nayan, MEHROTRA, Chitvan a Ajay Shanker SINGH. 2017. Risk for big data in the cloud. *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 277-282. DOI: 10.1109/ccaa.2017.8229815

DASORIYA, Rayan. 2017. A review of big data analytics over cloud. *2017 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pp. 1-6. DOI: 10.1109/icce-asia.2017.8307833

DEHGHANTANHA, Ali a Kim-Kwang Raymond CHOO. (Eds.). 2019. *Handbook of Big Data and IoT Security*. Springer International Publishing. ISBN 978-3-030-10543-3.

DIAZ, Alba. 2020. *The Four V's of Big Data.* OpenSistemas [online]. c OpenSistemas [cit. 2021-10-18]. Dostupné z: https://opensistemas.com/en/the-four-vs-of-big-data/

DUPRÉ, Lionel a Yuri DEMCHENKO. 2016. Impact of information security measures on the velocity of big data infrastructures. *2016 International Conference on High Performance Computing & Simulation (HPCS)*, pp. 492-500. DOI: 10.1109/HPCSim.2016.7568374

FAROOQI, Muhammad, Munam SHAH, Abdul WAHID, Adnan AKHUNZADA, Faheem KHAN, Noor AMIN a Ali IHSAN. 2019. Big Data in Healthcare: A Survey. *Applications of Intelligent Technologies in Healthcare*, pp. 143-152. DOI: 10.1007/978-3-319-96139-2_14

FIRICAN, George. 2017. *The 10 Vs of Big Data*. Transforming Data with Intelligence. [online]. c TDWI [cit. 2021-08-25]. Dostupné z: https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

FOUAD, Mohamed Mostafa, OWEIS, Nour E., GABER, Tarek, AHMED, Maamoun, a SNASEL, Vaclav. 2015. Data mining and fusion techniques for WSNs as a source of the big data. *Procedia Computer Science*, 65, 778-786. DOI: 10.1016/j.procs.2015.09.023

GUO, Yunyao, ZHANG, Baofu, a WEICHAO Miao. 2020. Research on Network Information Security Protection Technology Based on Big Data. *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pp. 19-22. DOI: 10.1109/cibda50819.2020.00013

GÜNTHER, Wendy Arianne, MEHRIZI, Mohammad H. Rezazade, HUYSMAN, Marleen a Frans FELDBERG. 2017. Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191-209. DOI: 10.1016/j.jsis.2017.07.003

HASHMANI, Manzoor Ahmed, JAMEEL, Syed Muslim, IBRAHIM Aidarus M., ZAFFAR, Maryam a Kamran RAZA. 2018. An Ensemble Approach to Big Data Security (Cyber Security). *International Journal of Advanced Computer Science and Applications,* 9(9). DOI: 10.14569/IJACSA.2018.090910

HIMTHANI, Puneet, DUBEY, Ghanshyam Prasad, SHARMA, Brij Mohan a Ankur TANEJA. 2020. Big Data Privacy and Challenges for Machine Learning. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 707-713. DOI: 10.1109/i-smac49090.2020.9243527

HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. 2015. *Big Data a NoSQL databáze*. 1. vyd. Praha: Grada Publishing, ISBN 978-80-247-5938-8.

HUSSIEN, Abou_el_ela Abdou. 2020. Fifty-Six Big Data V's Characteristics and Proposed Strategies to Overcome Security and Privacy Challenges (BD2). *Journal of Information Security*, **11**, 304-328. DOI: 10.4236/jis.2020.114019

IRUDAYASAMY, Amalraj a L. AROCKIAM. 2015. Scalable multidimensional anonymization algorithm over big data using map reduce on public cloud. *Journal of Theoretical & Applied Information Technology*, 74(2), 221-231.

JADON, Priyanshu a Durgesh Kumar MISHRA. 2018. Security and Privacy Issues in Big Data: A Review. *Emerging Trends in Expert Applications and Security*, pp. 659–665. DOI: 10.1007/978-981-13-2285-3_77

JAIN, Priyank, GYANCHANDANI, Manasi a Nilay KHARE. 2016. Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 1-25. DOI: 10.1186/s40537-016-0059-y

JANSSEN, Terry a Nancy GRADY. 2013. Big data for combating cyber attacks. *CEUR Workshop Proceedings*, 1097, 158-161.

JHA, Anupama, DAVE, Meenu a Supriya MADAN. 2017. Big Data Security and Privacy: A Review on Issues, Challenges and Privacy Preserving Methods. *International Journal of Computer Applications*, 177(4), 23-28. DOI: 10.5120/ijca2017915713

KHAN, Nawsher, ALSAQER, Mohammed, SHAH, Habib, BADSHA, Gran, ABBASI, Aftab Ahmad a Soulmaz SALEHIAN. 2018. The 10 Vs, issues and challenges of big data. *ICBDE '18: Proceedings of the 2018 International Conference on Big Data and Education*, pp. 52-56. DOI: 10.1145/3206157.3206166

KHAN, Nawsher, NAIM, Arshi, HUSSAIN, Mohammad Rashid, NAVEED, Quadri Noorulhasan, AHMAD, Naim a Shamimul QAMAR. 2019. The 51 V's Of Big Data: Survey, Technologies, Characteristics, Opportunities, Issues and Challenges. *In Proceedings of ACM Omni-layer Intelligent Systems Conference (COINS'19)*, pp. 19-24. DOI: 10.1145/3312614.3312623

KOO, Jahoon, KANG, Giluk a Young-Gab KIM. 2020. Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges. *Sustainability*, 12(24), 10571. DOI: 10.3390/su122410571

LIN, Min-Sheng, CHIU, Chien-Yi, LEE, Yuh-Jye a Hsing-Kuo PAO. 2013. Malicious URL filtering — A big data application. *2013 IEEE International Conference on Big Data*, pp. 589-596. DOI: 10.1109/bigdata.2013.6691627

MA, Xiaoxing a Di WU. 2014. Research on Information Security Issues Facing the Era of Big Data. *Applied Mechanics and Materials*, 651-653, 1913-1916. DOI: 10.4028/www.scientific.net/amm.651-653.1913

MATHUR, Akanksha a C.P. GUPTA. 2020. *Big Data Challenges and Issues: A Review*. *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2018)*, pp. 446-452. DOI: 10.1007/978-3-030-24643-3_53

MATTURDI, Bardi, ZHOU, Xianwei, LI, Shuai a Fuhong LIN. 2014. Big Data security and privacy: A review. *China Communications*, 11(14), 135-145. DOI: 10.1109/cc.2014.7085614

McNULTY, Eileen. 2014. *Understanding Big data: The seven V'S.* Dataconomy [online]. [cit. 2021-09-18]. Dostupné z: http://dataconomy.com/2014/05/seven-vs-big-data/.

MITTAL, Mamta. 2017. Study of Big Data Architecture and Tools. *International Journal of Current Engineering and Technology*, 7(1), 1-4.

MORENO, Julio, SERRANO, Manuel A. a Eduardo FERNÁNDEZ-MEDINA. 2016. Main Issues in Big Data Security. *Future Internet.* 8(3), 44. DOI: 10.3390/fi8030044

MUSTAFA, Ghulam, ASHRAF, Rehan, MIRZA, Muhammad Ayzed, JAMIL, Abid a MUHAMMAD. 2018. A review of data security and cryptographic techniques in IoT based

devices. *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems - ICFNDS '18*, pp. 1-9. DOI: 10.1145/3231053.3231100

PARYASTO, Marisa, ALAMSYAH Andry, RAHARDJO Budi a KUSPRIYANTO. 2014. Big-data security management issues. *2014 2nd International Conference on Information and Communication Technology (ICoICT)*, pp. 59-63. DOI: 10.1109/ICoICT.2014.6914040

QUASIM, Mohammad a Mohammad MERAJ. 2017. Big data security and privacy: a short review. *International Journal of Mechanical Engineering and Technology (IJMET)*, 8(4), 408-412.

RIAHI, Youssra. 2018. Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering*, 5(9), 524-528. DOI: 10.21276/ijre.2018.5.9.5

SAGIROGLU, Seref a Duygu SINANC. 2013. Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 42-47. DOI: 10.1109/cts.2013.6567202

SAHAFIZADEH, Ebrahim a Mohammad Ali NEMATBAKHSH. 2015. A Survey on Security Issues in Big Data and NoSQL. *Advances in Computer Science: an International Journal*, 4(4), 68-72.

SÁNCHEZ, Marco a Luis URQUIZA. 2019. Security Enhancement through Effective Encrypted Communication using ELK. *Proceedings of the 2019 International Conference on Big Data and Education – ICBDE'19*, pp. 88-92. DOI: 10.1145/3322134.3322154

SHAFER, Tom. 2017. *The 42 V's of Big Data and Data Science*. Machine Learning, Data Science, Big Data, Analytics, AI [online]. [cit. 2021-09-28]. Dostupné z: https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html.

SIMON, Phil. 2013. *Too Big to Ignore: The Business Case for Big Data*. John Wiley & Sons. ISBN 978-1118638170.

SINGH, Jainendra. 2014. Real time BIG data analytic: Security concern and challenges with Machine Learning algorithm. *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, pp. 1-4. DOI: 10.1109/csibig.2014.7056985

SU, Chunli. 2019. Big Data Security and Privacy Protection. *2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, pp. 87-89. DOI: 10.1109/icvris.2019.00030

SURESH, Jetti. 2014. Bird's eye view on "big data management". *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, pp. 1-5. DOI: 10.1109/CSIBIG.2014.7056930

TERZI, Duygu Sinanc, TERZI, Ramazan a Seref SAGIROGLU. 2015. A survey on security and privacy issues in big data. *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 202-207. DOI: 10.1109/icitst.2015.7412089

The 7 V's of Big Data – Impact. Partnership Automation: Key to Partnership Success [online]. 2016 c Impact [cit. 2021-09-28]. Dostupné z: https://impact.com/marketing-intelligence/7-vs-big-data/

TIMMINS, Fiona a Catherine MCCABE. 2005. How to conduct an effective literature search. *Nursing standard*, 20(11), 41-47. DOI: 10.7748/ns2005.11.20.11.41.c4010

TIWARI, Ankit Kumar, CHAUDHARY, Hemlata a Surendra YADAV. 2015. A review on Big Data and its security. *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1-5. DOI: 10.1109/iciiecs.2015.7193110

TONIDANDEL, Scott, Eden KING a Jose M. CORTINA. 2016. *Big data at work: the data science revolution and organizational psychology*. New York: Routledge, Taylor & Francis Group. ISBN 978-1-84872-581-2.

VENKATRAMAN, Sitalakshmi a Ramanathan VENKATRAMAN. 2019. Big data security challenges and strategies. *AIMS Mathematics*, 4(3), 860-879. DOI: 10.3934/math.2019.3.860

VORHIES, William. 2014. *How Many "V's" in Big Data? The Characteristics that Define Big Data*. Data Science Central [online]. c 2021 [cit. 2021-10-18]. Dostupné z: https://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data

WEBSTER, Jane a Richard T. WATSON. 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), 13-23.

XIAORONG, Feng, JIA, Shizhun a Mai SONGTAO. 2018. The research on industrial big data information security risks. *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, pp. 19-23. DOI: 10.1109/icbda.2018.8367644

YANG, Mengke, ZHOU Xiaoguang, ZENG, Jianqiu a Jianjian XU. 2016. Challenges and solutions of information security issues in the age of big data. *China Communications*, 13(3), 193-202. DOI: 10.1109/cc.2016.7445514

YANG, Tingting a Shuwen JIA. 2016. Research on Network Security Visualization under Big Data Environment. *2016 International Computer Symposium (ICS)*, pp. 660-662. DOI: 10.1109/ics.2016.0135

YE, Haina, CHENG, Xinzhou, YUAN, Mingqiang, XU, Lexi, GAO, Jie a Chen CHENG. 2016. A survey of security and privacy in big data. *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, pp. 268-272. DOI: 10.1109/ISCIT.2016.7751634.

YU, Shui. 2016. Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data. *IEEE Access*, 4, 2751-2763. DOI: 10.1109/access.2016.2577036

ZHANG, Dongpo. 2018. Big Data Security and Privacy Protection. *8th International Conference on Management and Computer Science (ICMCS 2018)*, pp. 275-278. DOI: 10.2991/icmcs-18.2018.56.

ZHANG, Xuyun, LIU, Chang, NEPAL, Surya, YANG, Chi, DOU, Wanchun a Jinjun CHEN. 2014. A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud. *Journal of Computer and System Sciences*, 80(5), 1008-1020. DOI: 10.1016/j.jcss.2014.02.007

ZHAO, Jiaqi, WANG, Lizhe, TAO, Jie, CHEN, Jinjun, SUN, Weiye, RANJAN, Rajiv, KOŁODZIEJ, Joanna, STREIT, Achim a Dimitrios GEORGAKOPOULOS. 2014. A security framework in G-Hadoop for big data computing across distributed Cloud data centres. *Journal of Computer and System Sciences*. 80(5), 994-1007. DOI: 10.1016/j.jcss.2014.02.006

ZHU, Yan, ZHANG, Yi, WANG, Jing, SONG, Weijing, CHU, Cheng-Chung a Guowei LIU. 2019. From Data-Driven to Intelligent-Driven: Technology Evolution of Network Security in Big Data Era. *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, pp. 103-109. DOI: 10.1109/compsac.2019.10191

ZIKOPOULOS, Paul a Chris EATON. 2011. *Understanding Big Data*: *Analytics for Enterprise Class Hadoop and Streaming Data*, 1. vyd. McGraw-Hill Osborne Media. ISBN 978-0-07-179053-6.

8 characteristics of big data – know the 8 V's of big data. 2020. Tech Blogger [online]. [cit. 2021-10-19]. Dostupné z: https://contenteratechspace.com/blogs/8-characteristics-of-big-data/

SEZNAM PŘÍLOH

PŘÍLOHY

Příloha I: Výsledky hledání v knihovně Web of Science

Web Science	*Počet vyhledaných publikací v letech*																					
Kombinace klíčových slov	**do 2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
big data AND security AND volume	1	0	1	0	1	1	0	1	2	4	2	1	5	15	47	61	107	115	117	126	117	59
big data AND security AND velocity	1	0	0	0	0	0	0	1	1	0	0	0	1	4	15	17	31	27	37	33	22	8
big data AND security AND variety	3	0	0	1	0	1	0	1	1	1	1	1	7	17	31	63	85	85	86	89	104	36
big data AND security AND veracity	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	5	9	11	16	11	10	5
big data AND security AND value	2	2	2	1	1	2	7	3	4	8	7	3	8	29	42	105	136	164	227	243	246	165
big data AND security AND viability	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	3	7	4	4	4	2
big data AND security AND visualization	0	0	0	0	0	2	1	3	2	1	0	2	6	19	17	26	35	56	54	53	37	17
big data AND security AND variability	1	0	0	0	0	0	0	0	0	1	1	1	1	3	2	3	6	10	14	9	18	10
big data AND security AND viscosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virality	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND validity	0	0	1	0	0	0	1	0	1	0	0	0	1	2	4	10	11	14	18	27	15	10
big data AND security AND volatility	0	0	0	0	0	0	0	0	1	2	0	0	0	1	1	2	4	3	4	3	7	5
big data AND security AND vulnerability	0	1	0	0	1	0	1	1	2	6	2	2	0	15	22	48	75	98	117	107	113	62
big data AND security AND vocabulary	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	1	2	2	3	4	1	4
big data AND security AND venue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3	3	1	4	3

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND vagueness	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2	0	2	0
big data AND security AND verbosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voluntariness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versatility	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2	3	3	1	2
big data AND security AND virtuosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visibility	1	0	0	0	0	0	0	1	0	0	0	0	1	1	2	6	5	4	5	13	9	8
big data AND security AND valor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varnish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vault	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	2	0	1
big data AND security AND voodoo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vulpine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND verdict	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
big data AND security AND vet	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	2	2	6	6	5	7	2
big data AND security AND vane	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanilla	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	3
big data AND security AND victual	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND vantage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1
big data AND security AND varmint	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vivify	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vastness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND voice	0	0	1	1	0	1	0	2	0	0	0	1	0	2	0	5	9	3	15	9	14	6
big data AND security AND vaticination	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND voyage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
big data AND security AND varifocal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND version control	0	0	0	0	0	1	0	0	0	0	3	0	0	1	2	3	4	1	3	8	7	2
big data AND security AND vexed	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
big data AND security AND vibrant	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2
big data AND security AND vogue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
big data AND security AND virtual	1	0	1	0	1	1	0	1	1	0	4	2	3	22	14	37	64	84	78	64	76	35
big data AND security AND verification	0	0	0	1	0	0	0	0	0	0	1	3	0	10	12	41	46	62	80	78	90	50
big data AND security AND vitality	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1	0
big data AND security AND violation	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	8	19	20	11	8	14	15
big data AND security AND verve	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND venturesomeness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND valences	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	1	1	0
big data AND security AND virility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vendible	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voracity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND visual	0	0	0	0	0	1	2	1	1	0	2	0	4	13	13	26	34	36	38	53	43	23
big data AND security AND vincularity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veritable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND violable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND complexity	1	0	0	0	0	1	0	1	0	4	5	3	7	20	21	73	93	124	132	131	154	82

Zdroj: vlastní zpracování [MS Excel]

Příloha II: Výsledky hledání v knihovně Scopus

Scopus	*Počet vyhledaných publikací v letech*											
Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND volume	0	0	9	17	55	66	96	118	125	175	128	92
big data AND security AND velocity	0	0	4	6	20	23	31	26	47	41	31	16
big data AND security AND variety	0	0	6	9	33	45	69	70	106	111	106	51
big data AND security AND veracity	0	0	0	2	4	5	12	19	12	22	13	8

Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND value	3	1	7	13	36	74	100	120	171	200	186	174
big data AND security AND viability	0	0	1	0	0	0	2	4	2	5	3	1
big data AND security AND visualization	0	1	5	8	22	25	47	44	63	74	52	33
big data AND security AND variability	0	0	0	1	2	1	6	4	9	10	16	6
big data AND security AND viscosity	0	0	0	0	0	1	0	0	0	0	0	1
big data AND security AND virality	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND validity	0	0	0	1	3	1	11	5	11	18	10	16
big data AND security AND volatility	0	0	0	0	0	1	3	2	3	4	8	6
big data AND security AND vulnerability	0	0	2	9	16	20	68	74	99	124	111	64
big data AND security AND vocabulary	0	0	0	0	1	2	0	1	2	5	0	2
big data AND security AND venue	0	0	0	0	0	2	0	3	1	3	2	1
big data AND security AND vagueness	0	0	0	0	0	0	0	0	2	0	3	0
big data AND security AND verbosity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voluntariness	0	0	0	0	1	0	0	0	0	1	0	0
big data AND security AND versatility	0	0	0	0	0	2	1	2	0	1	3	2
big data AND security AND virtuosity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visibility	0	0	0	2	2	3	4	3	8	8	6	8
big data AND security AND valor	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varnish	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versed	0	0	0	0	0	0	0	0	0	0	1	0
big data AND security AND vault	0	0	0	0	1	0	0	0	2	1	1	0
big data AND security AND voodoo	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veil	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vulpine	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND verdict	0	0	0	0	0	0	1	0	0	0	0	1
big data AND security AND vet	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vane	0	0	0	0	1	0	0	0	0	0	0	0
big data AND security AND vanilla	0	0	0	0	0	0	0	1	0	0	1	1

Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND victual	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vantage	0	0	0	0	1	0	1	1	0	2	0	2
big data AND security AND varmint	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vivify	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vastness	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND voice	0	0	0	0	2	2	7	4	13	9	12	12
big data AND security AND vaticination	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veer	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voyage	0	0	0	0	0	0	0	1	0	0	1	0
big data AND security AND varifocal	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND version control	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND vexed	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vibrant	0	0	0	0	0	0	0	0	1	0	0	1
big data AND security AND vogue	0	0	0	0	0	0	1	1	1	0	0	1
big data AND security AND verification	0	0	1	7	5	23	38	36	68	63	65	58
big data AND security AND vitality	0	0	0	0	0	1	0	2	0	0	2	3
big data AND security AND violation	0	0	0	1	1	9	11	16	11	9	10	9
big data AND security AND verve	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND venturesomeness	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virtual	3	0	4	14	22	31	66	99	116	110	77	69
big data AND security AND valences	0	0	0	0	0	0	1	0	2	1	1	0
big data AND security AND virility	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vendible	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voracity	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND visual	0	0	2	8	9	21	31	27	36	39	42	36
big data AND security AND vincularity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veritable	0	0	0	0	0	0	0	0	0	0	0	1

Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND violable	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND complexity	1	0	4	14	20	36	72	84	90	114	100	67

Zdroj: vlastní zpracování [MS Excel]

Příloha III: Výsledky hledání v knihovně IEEE Xplore

IEEE Xplore	*Počet vyhledaných publikací v letech*											
Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND volume	0	0	2	13	41	45	60	56	78	82	63	28
big data AND security AND velocity	0	0	1	5	18	16	24	16	22	24	15	3
big data AND security AND variety	0	0	2	9	23	28	40	33	62	51	49	25
big data AND security AND veracity	0	0	0	2	6	2	6	3	13	9	3	1
big data AND security AND value	0	0	3	22	20	48	54	56	130	113	110	70
big data AND security AND viability	0	0	1	0	0	0	0	4	5	0	3	1
big data AND security AND visualization	0	0	4	10	15	18	33	46	58	59	34	17
big data AND security AND variability	0	0	0	1	1	3	3	2	3	4	7	2
big data AND security AND viscosity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virality	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND validity	0	0	0	0	2	1	1	4	12	12	11	7
big data AND security AND volatility	0	0	0	0	0	1	0	0	1	4	4	3
big data AND security AND vulnerability	0	0	1	6	11	4	31	42	74	90	60	33
big data AND security AND vocabulary	0	0	0	1	2	0	3	2	6	4	1	2
big data AND security AND venue	0	0	0	0	0	1	0	1	2	0	0	0
big data AND security AND vagueness	0	0	0	0	0	0	0	0	0	0	1	0
big data AND security AND verbosity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voluntariness	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND versatility	0	0	0	0	1	0	0	2	0	0	2	0
big data AND security AND virtuosity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visibility	0	0	0	1	1	1	3	2	3	6	3	6
big data AND security AND valor	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varnish	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versed	0	0	0	0	0	0	1	0	0	0	0	0
big data AND security AND vault	0	0	0	0	0	0	0	0	1	0	0	2
big data AND security AND voodoo	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veil	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vulpine	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND verdict	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vet	0	0	0	0	0	0	0	1	0	1	0	0
big data AND security AND vane	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanilla	0	0	0	0	0	0	0	0	0	1	1	0
big data AND security AND victual	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vantage	0	0	0	0	0	0	0	0	0	1	0	1
big data AND security AND varmint	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vivify	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vastness	0	0	0	0	0	0	0	0	1	1	0	0
big data AND security AND voice	0	0	0	0	1	2	4	1	6	2	12	6
big data AND security AND vaticination	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veer	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND voyage	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varifocal	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND version control	0	0	0	0	0	0	0	1	0	1	0	0
big data AND security AND vexed	0	0	0	1	0	0	0	0	0	0	0	0
big data AND security AND vibrant	0	0	0	0	0	0	0	1	0	0	0	0
big data AND security AND vogue	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND verification	0	0	1	8	7	11	17	24	61	42	44	30
big data AND security AND vitality	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND violation	0	0	1	2	0	3	8	11	7	8	5	4
big data AND security AND verve	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND venturesomeness	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virtual	0	0	3	12	18	22	39	49	55	42	45	31
big data AND security AND valences	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND virility	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vendible	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voracity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visual	0	0	3	5	9	13	16	15	24	26	17	11
big data AND security AND vincularity	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veritable	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND violable	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND complexity	0	0	3	12	24	30	43	61	93	108	75	57

Zdroj: vlastní zpracování [MS Excel]

Příloha IV: Výsledky hledání v knihovně ACM Digital Library

ACM Digital Library	***Počet vyhledaných publikací v letech***																					
Kombinace klíčových slov	**do 2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
big data AND security AND volume	1	0	1	1	0	0	2	1	2	4	6	8	48	105	193	202	271	312	326	394	308	222
big data AND security AND velocity	1	0	0	0	0	0	2	0	0	0	0	0	8	20	47	68	91	100	103	92	72	46
big data AND security AND variety	6	0	0	1	0	0	0	1	2	5	5	3	52	98	180	219	312	355	410	468	371	279

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND veracity	0	0	0	0	0	0	0	0	0	0	1	0	1	5	18	31	31	37	47	39	32	14
big data AND security AND value	4	0	0	3	0	0	3	2	3	10	12	14	83	189	330	389	554	685	846	1013	849	598
big data AND security AND viability	0	0	0	0	0	0	1	0	1	4	1	3	9	17	36	39	78	85	81	93	74	53
big data AND security AND visualization	3	0	0	1	0	0	0	1	2	3	4	5	27	65	127	155	191	236	304	322	289	215
big data AND security AND variability	1	0	0	1	0	0	3	1	2	3	7	6	37	81	159	182	275	348	370	436	397	320
big data AND security AND viscosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	5	0	6	3	2	5
big data AND security AND virality	0	0	0	0	0	0	0	0	1	0	0	1	6	14	23	25	35	58	42	57	63	49
big data AND security AND validity	2	0	0	1	0	0	2	1	1	2	3	3	13	44	99	123	172	208	234	289	237	197
big data AND security AND volatility	0	0	0	0	0	0	0	0	0	1	0	1	3	9	18	30	46	35	53	61	48	28
big data AND security AND vulnerability	0	0	0	0	0	0	3	0	1	0	0	2	16	33	59	56	115	160	192	260	209	168
big data AND security AND vocabulary	0	0	0	0	0	0	1	0	1	2	0	0	11	11	26	28	51	58	74	65	50	76
big data AND security AND venue	0	0	0	0	0	0	0	1	1	1	0	0	6	16	21	19	42	38	59	33	36	52
big data AND security AND vagueness	0	0	0	0	0	0	1	0	0	0	1	0	5	2	12	21	26	18	35	38	33	39
big data AND security AND verbosity	0	0	0	0	0	0	0	0	0	0	0	1	3	2	7	4	11	7	23	11	7	7
big data AND security AND voluntariness	0	0	0	0	0	0	2	0	1	2	0	1	9	20	33	39	61	66	85	89	57	65
big data AND security AND versatility	1	0	0	1	0	0	0	1	0	0	0	0	2	5	14	20	33	27	25	54	41	26
big data AND security AND virtuosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
big data AND security AND visibility	1	0	0	0	0	0	1	1	2	3	6	5	17	48	81	77	130	132	156	170	155	125

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND valor	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	7	1	1	6	9	6	6
big data AND security AND varnish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	0
big data AND security AND versed	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	11	9	4	17	15	8	9
big data AND security AND vault	0	0	0	0	0	0	0	0	0	0	0	2	0	5	1	5	5	3	2	9	5	7
big data AND security AND voodoo	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1
big data AND security AND veil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	2	2	1	4	4	4
big data AND security AND vulpine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND verdict	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	6	3	9	5	6	4
big data AND security AND vet	1	0	0	0	0	0	0	0	0	0	0	0	2	4	13	5	16	13	21	16	13	14
big data AND security AND vane	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	2
big data AND security AND vanilla	0	0	0	0	0	0	0	0	0	2	1	1	0	2	4	7	15	7	11	14	17	17
big data AND security AND victual	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
big data AND security AND vantage	0	0	0	0	0	0	0	0	0	0	0	0	1	2	6	5	10	11	12	15	3	8
big data AND security AND varmint	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND vivify	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
big data AND security AND vastness	2	0	0	1	0	0	1	0	0	3	2	3	26	49	57	77	104	148	137	173	125	101
big data AND security AND voice	1	0	0	0	0	0	2	0	1	0	0	1	7	13	27	46	62	75	95	112	113	90
big data AND security AND vaticination	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND veer	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	2	2	0	1	2	1
big data AND security AND voyage	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	1	3	1	4	2	3
big data AND security AND varifocal	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
big data AND security AND version control	0	0	0	0	0	0	0	0	1	0	0	0	1	12	10	8	17	23	13	14	18	12
big data AND security AND vexed	0	0	0	0	0	0	0	0	0	0	0	0	1	3	3	3	3	3	4	2	1	5
big data AND security AND vibrant	0	0	0	0	0	0	0	1	0	0	0	3	6	9	12	16	28	40	31	38	25	55
big data AND security AND vogue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	2	2	2	1
big data AND security AND verification	3	0	0	2	0	0	2	1	1	4	4	3	24	76	127	154	237	279	350	464	407	279
big data AND security AND vitality	0	0	0	0	0	0	0	1	1	1	0	0	7	21	40	57	88	99	122	163	158	135
big data AND security AND violation	0	0	0	0	0	0	3	1	0	0	1	3	17	37	59	65	118	128	136	187	147	134
big data AND security AND verve	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	2	3	0	1	0	0
big data AND security AND venturesomeness	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0
big data AND security AND virtual	1	0	1	1	0	0	2	1	3	4	3	10	37	93	134	152	241	304	336	401	363	349
big data AND security AND valences	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	12	10	17	19	16	7	6
big data AND security AND virility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND vendible	6	0	0	1	0	0	1	0	2	1	4	3	15	39	46	60	86	130	113	119	100	73
big data AND security AND vanity	2	0	0	0	0	0	1	1	0	0	0	0	0	0	3	1	1	1	2	7	2	1
big data AND security AND voracity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	2

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND visual	2	0	0	0	0	0	0	1	3	1	4	3	20	38	73	81	114	161	169	211	184	157
big data AND security AND vincularity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND veritable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	3	2	3	2	1
big data AND security AND violable	0	0	0	0	0	0	3	1	0	0	1	3	17	37	59	65	118	128	136	187	147	119
big data AND security AND complexity	6	0	0	2	0	0	3	2	2	6	9	10	64	163	281	323	454	558	646	802	626	502

Zdroj: vlastní zpracování [MS Excel]

Příloha V: Výsledky hledání v knihovně Science Direct

Science Direct	***Počet vyhledaných publikací v letech***																					
Kombinace klíčových slov	**do 2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
big data AND security AND volume	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	4	7	7	6	11	11
big data AND security AND velocity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	3	6	2	3	1
big data AND security AND variety	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	8	8	7	11	10	10
big data AND security AND veracity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	3	4	0	3	3
big data AND security AND value	0	0	0	1	0	0	0	1	0	0	1	0	1	1	3	3	6	9	17	13	15	14
big data AND security AND viability	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	2	2	1	1	5
big data AND security AND visualization	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	4	1	5	8	4
big data AND security AND variability	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND viscosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virality	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND validity	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	1	4	1	4
big data AND security AND volatility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0
big data AND security AND vulnerability	0	1	0	2	0	0	0	0	0	0	1	0	0	2	3	3	2	4	4	13	15	12
big data AND security AND vocabulary	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND venue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
big data AND security AND vagueness	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND verbosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voluntariness	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versatility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
big data AND security AND virtuosity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visibility	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1
big data AND security AND valor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
big data AND security AND varnish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND versed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND vault	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voodoo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND veil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vulpine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND verdict	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vane	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanilla	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND victual	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vantage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varmint	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vivify	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vastness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
big data AND security AND vaticination	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voyage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND varifocal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND version control	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
big data AND security AND vexed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Kombinace klíčových slov	do 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
big data AND security AND vibrant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND vogue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND verification	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	2	10	6	4	12	15
big data AND security AND vitality	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
big data AND security AND violation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0
big data AND security AND verve	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND venturesomeness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND virtual	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	4	4	3	5	5	6
big data AND security AND valences	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
big data AND security AND virility	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vendible	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND vanity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND voracity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND visual	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	2	0	3	1	7
big data AND security AND vincularity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND veritable	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
big data AND security AND violable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
big data AND security AND complexity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	3	11	10	12	14	9

Zdroj: vlastní zpracování [MS Excel]

Příloha VI: Přehled řešených charakteristik bezpečnosti a jejich součástí

Název V	Vznik	Součást (ANO / NE)									Průměr výsledků v knihovnách
		3V	*4V*	*5V*	*7V*	*8V*	*9V*	*10V (2017)*	*10V (2014)*	*56V*	
Volume	2001	1	1	1	1	1	1	1	1	1	919
Velocity	2001	1	1	1	1	1	1	1	1	1	252
Variety	2001	1	1	1	1	1	1	1	1	1	874
Value	2012	0	0	1	1	1	1	1	1	1	98
Veracity	2011	0	1	1	1	1	1	1	1	1	1760
Validity	2014	0	0	0	0	1	1	1	1	1	377
Volatility	2014	0	0	0	0	0	1	1	0	1	82
Variability	2013	0	0	0	1	1	1	1	1	1	559
Visualization	2013	0	0	0	1	1	1	1	0	1	596
Venue	2014	0	0	0	0	0	0	0	1	1	72
Valences	2020	0	0	0	0	0	0	0	0	1	20
Vulnerability	2017	0	0	0	0	0	0	1	0	1	590

Zdroj: vlastní zpracování [MS Excel]