



INDUSTRY 4.0 for VET

3 BIG DATA

3.1 Téma

První seznámení

Dokážete si představit, že by běžný člověk potřeboval 181 miliónů let ke stažení veškerých dat z internetu? Toto obrovské množství všech dnes dostupných dat a způsob, jakým jsou zpracovány, se nazývá **big data** (česky **velká data**).

Jak se dozvíte v této kapitole, s velkými daty se setkáváme denně, často aniž bychom o tom sami věděli. Dozvíte se o **výhodách** i **nevýhodách** velkých dat a jak je často správné zacházení s tímto obrovským množstvím údajů důležitější než data sama.



Praktický význam – k čemu využijete znalosti a dovednosti

Nejen IT specialisté, ale téměř každý narazí na tzv. big data v **každodenních situacích**, například při návštěvě doktora, surfování po sociálních sítích jako je Facebook, Instagram anebo hledáním na Googlu. Znalost použití velkých dat a **možnosti i nástrahy**, které s tím mohou souviset, může být relevantní jak pro vaše **osobní používání internetu**, tak i pro váš **profesní život**, možná ve společnosti, která se analyzováním velkých dat zabývá.

Stručný přehled učebních cílů a kompetencí

Tato kapitola vám poskytne základní znalosti o velkých datech. Poznáte **3-V model** a zjistíte, jak je velké množství dat shromažďováno a analyzováno. Poté se dozvíte o **účelech**, pro které jsou využívány **znalosti** získané z velkého množství dat a rizika spojená s jejich zpracováním. Uvidíte, proč v posledních letech **ochrana dat** nabývá na **důležitosti**, a pochopíte, že **nakládání** s velkými daty představuje pro společnosti i jednotlivce velké výzvy.

Učební cíle

Pochopíte a dokážete popsat pojem big data.

Budete vědět, jak používat big data.

Pochopíte a dokážete vysvětlit, jak je velké množství dat sbíráno a analyzováno.

Dozvíte se, jaké výzvy a rizika velká data přinášejí.

3.2 Co jsou to big data?

Věděli jste, že 90 % všech dat, která jsou dnes na světě dostupná, bylo vytvořeno v posledních letech? Díky množství nových informací a komunikačních technologiích, celosvětový **objem dat** neuvěřitelně narostl a nabízí dříve neznámé možnosti. **Big data** je názvem pro tento **objem** strukturovaných a nestrukturovaných **dat**, které vzhledem ke své velikosti nemohou být zpracovány konvenčním softwarem nebo hardwarem.



Tyto objemy dat jsou, mimo jiné, vytvářeny s každým z našich **kliknutí na internetu**. To může být například jen platba na Amazonu, vyhledávací příkaz na Googlu, aktivita na sociálních sítích jako je Instagram nebo Facebook apod.

Velké množství dat však samo o sobě big data netvoří. Pouze **analýza a zpracování** těchto objemů dat, např. společností, big data charakterizuje. V roce 2001 analytik Doug Lane vytvořil definici velkých dat pomocí svého **3-V modelu**, který je dodnes uznávaný. Podle Lanea mají big data tyto tři následující charakteristiky:

- **Objem („Volume“):** Společnosti sbírají velký objem dat z různých zdrojů. Tyto zdroje zahrnují inteligentní nástroje (anglicky „intelligent devices“, IoT) jako jsou mobilní telefony, videa, sociální média atd. V minulosti by nebylo možné ukládat tyto obrovské objemy dat, dnes pro tento účel existují úložné platformy.
- **Rychlost („Velocity“):** Firmy jsou v dnešní době zaplaveny toky dat nebývalých rychlostí, které je třeba rychle zpracovat.
- **Různorodost („Variety“):** Sesbíraná data jsou různorodá a mají širokou škálu formátů: numerická data, která jsou k dispozici ve strukturované podobě a jsou uložena v běžných databázích, stejně jako nestrukturované textové dokumenty, data z finančních transakcí nebo e-maily.

Definice

Big data

...znamená **velké množství** dostupných **dat**, která jsou **analyzována a zpracována** pro konkrétní účel. Podle Douga Lanea, big data charakterizuje **objem, rychlost a různorodost**.

Big data vs. small data



“Let’s shrink Big Data into Small Data ...
and hope it magically becomes Great Data.”

Na rozdíl od velkých dat, se malá data (anglicky small data) vztahují k datům v objemu a formátu dostupnému lidem. Následující body ukazují, jak lze velká data odlišit od malých dat:

- **Cíle:** Malá data jsou používána pro určitý cíl, použití velkých dat se často vyvíjí nečekaně.
- **Umístění:** Malá data jsou běžně uchovávána na jednom místě, obvykle v jednom souboru na počítači, kdežto velká data se obvykle šíří do několika souborů na rozličných serverech umístěných v různých zemích.
- **Struktura dat:** Malá data jsou často strukturována v rovné linii, zatímco velká data mohou být nestrukturovaná a mohou být získána v různých formátech z různých oblastí.
- **Příprava dat:** Na přípravě malých dat se obvykle podílí pouze jeden koncový uživatel. V případě velkých dat se však často stává, že data připravuje jedna skupina lidí, jiná skupina je analyzuje, a nakonec je použije třetí skupina. Každá z těchto skupin může také sledovat jiné cíle.
- **Trvanlivost:** Malá data jsou obecně uchovávána po určitou dobu po dokončení projektu. V případě velkých dat však data zůstávají uložena po neomezenou dobu.
- **Původ:** Malá data jsou uložena na jednom místě po krátkou dobu a ve specifických měrných jednotkách. Velká data, na druhé straně, pochází z různých míst, zemí, společností, organizací atd.
- **Reprodukovatelnost:** Malá data lze obecně zcela reprodukovat. Naproti tomu velká data pochází z mnoha různých zdrojů a jsou k dispozici v mnoha podobách, takže reprodukce není možná.

- Jak můžete vidět z rozlišení velkých a malých dat, velká data jsou často doslova obtížně uchopitelná.

Deep Learning neboli „hluboké učení“ také používá big data: jedná se o speciální metodu **zpracování informací** a podoblast **strojového učení**. Stroj je „nahrnut“ velkým množstvím dat, která jsou analyzována a použita k trénování stroje. Stroj je schopen vzájemně propojit nové informace a na tomto základě může vytvářet předpovědi a přijímat svá vlastní rozhodnutí.



Nicméně, s velkými daty není vždy zacházeno **zodpovědně**. Některé společnosti nebo instituce nedodržují zásady ochrany dat, což znamená, že informace mohou unikat na veřejnost. Ač to možná vypadá jako drobnost, v některých případech může být únik velmi nebezpečný a může vést k **podvodům a vydíráním**.

Příklad

V roce 2015 se stal obětí hackerského útoku záletnický portál Ashley Madison, kde si profil mohli vytvořit lidé hledající mimomanželské dobrodružství. V důsledku toho byly informace o registrovaných uživateli portálu zveřejněny na internetu. Byly zveřejněny informace o záletech celebrit a osobní informace, jako jsou např. čísla kreditních karet. Poškození uživatelé byli navíc požádáni e-mailem o zaplacení výkupného, aby se jejich životní partneři nedozvěděli o jejich profilu na tomto portálu.

Zapamatujte si

Big data mohou být mimo jiné použita k následujícím účelům:

- Strategické směřování společnosti
- Deep Learning
- Boj proti zločinu a terorismu
- Vědecké zkoumání přírodních jevů (např. zemětřesení a změna klimatu)
- Protiprávní užití dat, které může vést k vydírání nebo podvodům

Rozhodujícím faktorem u velkých dat nejsou ani tak samotná data, ale to, co se s nimi stane.

Společnosti především těží z analýzy a vyhodnocení velkých dat. Vědomě i nevědomě dnes vytvářejí a ukládají obrovské množství dat. Nyní se dozvíte, jaké možnosti správná analýza velkých dat firmám nabízí.

Rozhodování

Analýzou velkého množství generovaných dat mohou společnosti identifikovat vzorce a odfiltrovat informace. To umožňuje firmám přijímat lepší obchodní rozhodnutí, která zvyšují úspěch společnosti. Vyhodnocením strojových dat je například možné vypočítat, v jakých intervalech se stroj porouchává. Společnost může tyto znalosti použít k opravě stroje dříve, než selže. Big data se také používají ve finančním a pojišťovacím průmyslu k lepšímu výpočtu rizik.

Příklad

Paní Schmidtová je 47 let a chtěla by uzavřít soukromé zdravotní pojištění. Při návštěvě svého pojišťovacího makléře je překvapena vysokými náklady a množstvím dotazů, které jsou jí pokládány. Ukazuje se, že její poskytovatel analyzuje velké množství dat, aby lépe vypočítal jednotlivé náklady na pojištění. Společnost například zjišťuje, jaká konkrétní zdravotní rizika nesou ženy tohoto věku, které jsou, stejně jako paní Schmidtová, kuřáky, nemají děti a nesportují.

Zvýšení efektivity

Konkurenceschopnost je pro společnosti velmi důležitá. Aby společnost udržela krok s konkurencí, musí navrhnout strategie, jak ušetřit náklady, aniž by došlo ke snížení výkonu. K tomu pomáhá analýza a propojení velkých dat.

Příklad

Už jste někdy slyšeli, že řidiči UPS téměř vždy odbočí doprava? Je to proto, že společnost UPS pomocí analýzy velkých dat zjistila, že tak může ušetřit skoro 10 milionů dolarů ročně. Nejspíše jste zvědaví, jak je to možné: sloučení různých datových souborů, jako jsou statistiky nehod, údaje o spotřebě paliva atd., ukázalo, že vozidla UPS jsou mnohem méně

zapojena do nehod, pokud neodbočí doleva. To může ušetřit spoustu peněz, a to i v případě, že v důsledku toho budou trasy komplikovanější.

Prognózy ve výzkumu a vývoji

Tím, že si stávající či potenciální zákazníci nebo klienti uvědomí, že preferují určité produkty, může výzkum identifikovat a předpovídat trendy, navrhovat vhodné marketingové strategie a vyvíjet produkty šité na míru. Pomocí vhodných analytických metod je také možné například předvídat bezpečnost výrobku při poškození, zatímco je ještě vyvíjen.

Příklad

Provozovatel internetového obchodu instaluje tzv. cookies a online sledování pohybu svých návštěvníků. Může zjistit, odkud návštěvníci pocházejí, na které produkty kliknou, jak často navštěvují stránky apod. Pomocí těchto údajů tak může provozovatel přizpůsobit obsah stránky a nabízené produkty preferencím návštěvníků a zvýšit tak svůj obrat.

Personalizovaný zákaznický servis

Uložením rozhodnutí zákazníků jsou společnosti schopné poskytovat personalizovaný zákaznický servis. Pokud například uživatel sleduje na Netflixu konkrétní film nebo seriál, systém jej uloží a při příštím přihlášení uživatele budou doporučení vycházet z filmů nebo seriálů, které už uživatel sledoval. Ne vždy se však tato přizpůsobená nabídka setká s uznáním:

Příklad

Když si pan Majer uvědomil, že jeho staré horské boty již dávno nejsou použitelné, zadal do Googlu dotaz na „nové horské boty pro muže“. Následně byl zahlcen mnoha různými nabídkami a také zjistil, že spousta produktů nelze dodat do jeho domovské země, Rakouska. Pan Majer se rozhodl získat osobní radu v specializovaném obchodě a také si pár horských bot zakoupil. V následujících dnech a týdnech však vidí na internetu více a více reklam na horské boty, jelikož jeho vyhledávací dotaz byl uložen a analyzován na Googlu. Pan Majer je popuzen a cítí se být sledován. Rozhodne se, že v budoucnu už nebude na Google vkládat žádné další vyhledávací dotazy.



Zopakujme si to znovu:

Zapamatujte si

Společnosti mají mnoho možností, jak využít big data ke zvýšení svého úspěchu. Ty zahrnuje:

- **Rozhodování**
Analýza velkých dat pomáhá společnostem dělat lepší obchodní rozhodnutí a lépe zvládat rizika.
- **Zvýšení efektivity**
Analýza a propojování dat (jako jsou například data o počasí a data s cenami pohonných hmot) pomáhá společnostem zefektivnit procesy.
- **Prognózy ve výzkumu a vývoji**
S pomocí velkých dat mohou být vytvořeny předpovědi týkající se trendů, vlastností produktu atd.
- **Personalizovaný zákaznický servis**
Ukládáním rozhodnutí zákazníků jim mohou společnosti při jejich další návštěvě nabídnout personalizovanou zákaznickou nabídku.

3.4 Jak jsou big data analyzována?

Již jste se naučili definici velkých dat a jaké možnosti nabízí jejich používání. V této kapitole půjdeme více do detailu a podíváme se na **analýzu** velkých dat. Tato specializovaná oblast je známá jako **big data analýza** neboli **analýza velkých dat**.



Analýza velkých dat – teorie:

Prvním krokem je shromáždění **velkého množství dat** z různých zdrojů v různých formátech, což se často provádí pomocí vyhledávacích příkazů. Poté jsou data **připravena** k dalšímu zpracování. Jedním z častých problémů je to, že je velké množství dat dostupné v nestrukturované podobě a ve zcela odlišných formátech, tudíž je nelze zachytit klasickým databázovým softwarem.

Analýza velkých dat proto používá **složitě procesy** k extrahování a získání dat. Data jsou poté **analyzována** pomocí big data softwaru. Nakonec jsou výsledky **zpracovány a prezentovány**.

Je důležité, aby byl použitý software schopen rychlého uskutečnění mnoha požadavků na vyhledávání a rychlého importu a zpracování různých datových záznamů. Za účelem zvýšit svoji výkonnost, spousta systémů nepoužívá **místo na pevném disku** (jako klasické databázové aplikace) ke zpracování dat, ale spíše obvykle mnohem rychlejší **hlavní paměť**. Díky tomu může být dosažená rychlost vyšší a analýza může být provedena téměř ihned.

Zapamatujte si

Analýza velkých dat lze v podstatě rozdělit na tři různé kroky:

- Získávání dat z mnoha různých zdrojů pomocí vyhledávacích dotazů
- Vyhodnocení a optimalizace shromážděných údajů
- Analýza dat a shrnutí a prezentace výsledků

Velmi důležitým je **výkonný a vhodný software**.

Analýza velkých dat v praxi

Je důležité upozornit, že analýza velkých dat je ve většině společností stále ještě v plenkách a **příležitostech**, které nabízí, **nejsou zdaleka vyčerpány**. V průměru společnosti analyzují pouze o něco více než třetinu dat generovaných digitálním kontaktem s jejich zákazníky (např. prostřednictvím online obchodů nebo webových stránek).

Důvodem jsou často přísná **pravidla ochrany údajů**, která stěžují analýzu velkých dat. Zákony a předpisy upravující ochranu údajů jsou podrobněji popsány v následující kapitole. Ve skutečnosti však společnosti ještě nejsou v mnoha ohledech připraveny efektivně využívat velké množství dat. Důležitou roli hrají následující faktory:

Především je vhodné výsledky správně distribuovat: **zdroje dat** by měly pocházet z různých oblastí, výsledky by měly být použity v několika sférách společnosti. Je vyžadována také vhodná **strategie**: společnost by měla vědět, za jakým účelem je velké množství údajů analyzováno. Vhodná **firemní kultura** je také velmi důležitá, například nové technologie by se v zásadě neměly odmítat, ale spíše realisticky zvažovat.



Většina firem nemá vlastní oddělení pro analýzu dat. Nicméně, někteří zaměstnanci by však měli s sebou do firmy přinášet potřebné odborné znalosti nebo je získat v odborných kurzech. V některých případech je potřeba najmout nové zaměstnance. V rámci společnosti musí být také jasně definovány odpovědnosti a oprávnění jednotlivých zaměstnanců.

K analýze velkých dat je nutná **efektivní technologie** ve formě vhodných nástrojů pro analýzu velkých dat. Jaké nástroje jsou však vhodné, závisí na dříve definované strategii nebo na definovaném účelu analýzy. V neposlední řadě je rovněž nezbytná vhodná strategie ochrany údajů, aby bylo zajištěno, že

osobní údaje jednotlivců nebudou zveřejňovány. Specializovaný odborník na ochranu údajů v rámci společnosti zajišťuje, aby analýza údajů byla v souladu s platnými zákony a předpisy.

Zapamatujte si

V souhrnu jsou pro úspěch analýzy velkých dat ve společnosti důležité následující body:

- Strategie velkých dat – definovat účel analýzy
- Vhodná firemní kultura – otevřenost vůči novým technologiím
- Personál s nezbytným know-how – školení nebo nábor
- Výkonná technologie – nástroje vhodné pro analýzu velkých dat
- Vhodné zásady ochrany osobních údajů – soulad s platnými zákony a předpisy

3.5 Výzvy a rizika velkých dat

V předchozích kapitolách jsme byli svědky toho, jak složité je big data analyzovat a používat.

Minimálně stejně složité jsou i výzvy a rizika spojená s užitím velkého množství dat. Pravděpodobně největší **výzvou** pro společnosti v souvislosti s velkými daty je **ochrana dat**:



Přestože společnosti v posledních letech věnují více pozornosti ochraně údajů, stále se objevují problémy s tím spojené. Například osobní údaje uživatelů internetu mohou být užity bez jejich souhlasu a dotyčné osoby mohou být identifikovány, kontrolovány a v nejhorším případě vydírány.

Definice

Osobní údaje

...jsou **údaje** které se vztahují k určité osobě a umožňují vyvozovat závěry o její osobnosti. Patří mezi ně například číslo poznávací značky automobilu, datum narození nebo výpis z bankovního účtu.

Příkladem **narušení ochrany osobních údajů** ve spojitosti s velkými daty je případ portálu Ashley Madison, který již byl zmíněn v kapitole 2. V tomto případě se zveřejněné **osobní údaje** použily k **vydírání** vlastníků těchto dat.

Předpisy a zákony o ochraně osobních údajů pomáhají chránit spotřebitele před zneužitím. **Základem obecného zákona o ochraně údajů** v Evropské unii je **obecné nařízení o ochraně údajů (anglicky „General Data Protection Regulation“)**, které vstoupilo v platnost dne 25. května 2018.



Exkurz

Obecné nařízení o ochraně osobních údajů

Obecné nařízení o ochraně osobních údajů (anglicky General Data Protection Regulation, zkráceně GDPR) se nazývá v celém rozsahu „Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob při zpracování osobních údajů a o volném pohybu těchto údajů o zrušení směrnice 95/46 / ES ". Je přímo aplikovatelné v České republice a je doplněno zákonem o zpracování osobních údajů.

Toto nařízení umožňuje občanům EU lépe kontrolovat shromažďování a používání jejich osobních údajů. To by mělo posílit důvěru spotřebitelů v jednotlivých společnostech. Stávající práva občanů EU jsou konsolidována v GDPR a jsou rovněž zavedena nová práva v souladu s obecným nařízením, která zahrnují:

- **zjednodušený přístup k osobním údajům** – to zahrnuje poskytování komplexních, jasných a srozumitelných informací o zpracování údajů
- nové **právo na přenositelnost dat** – osobní údaje budou předávány zjednodušeným způsobem
- jasnější **právo na výmaz („právo být zapomenut“)** - údaje se vymažou, pokud občan nesouhlasí se zpracováním svých údajů a neexistuje legitimní důvod k jejich uchování
- právo na **informace o zneužitých osobních údajích** – společnosti a organizace neprodleně informují dotčené osoby o závažném porušení ochrany osobních údajů. Rovněž musí být informován odpovědný kontrolní orgán pro ochranu údajů.

Pro společnosti GDPR představuje možnost vytvořit více obchodních příležitostí a podporovat inovace pomocí řady opatření, které zahrnují:

- vytvoření **jednotných pravidel pro celou EU**, což povede k velkým úsporám
- **jmenování referenta pro ochranu údajů** v rámci orgánů a společností, které pracují s velkým množstvím dat
- **určení jediného kontaktního místa** ve své zemi, na které se podniky musí obrátit
- vytvoření **pravidel EU pro společnosti ze třetích zemí**, které musí dodržovat při nabídce zboží nebo služeb nebo sledování chování lidí
- vytvoření **pravidel, která podporují inovace** a zajišťují, aby se pravidla ochrany údajů zohledňovala v rané fázi vývoje služeb nebo produktů
- **používání technik slučitelných s ochranou údajů**, jako je pseudonymizace (nahrazení pasáže v datovém záznamu, které umožňují identifikaci přidružené osoby) a šifrování (data jsou šifrována tak, aby je mohla číst pouze oprávněná osoba)

- odstranění povinnosti společností **podávat zprávy** s cílem podpořit volný pohyb osobních údajů v rámci Evropské unie
- provádění **posouzení výše dopadů**, kdy je pravděpodobné, že zpracování údajů ohrožuje práva a svobody jednotlivců

Úplné znění obecného nařízení o ochraně osobních údajů je k dispozici na <https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A32016R0679>.

Další výzvou pro společnosti je to, že stávající zaměstnanci nemají vždy potřebné **odborné znalosti** a nejsou vždy otevření možnostem, které analýza velkých dat firmám nabízí.

Často je plýtváno časem a prostředky, protože zúčastněným stranám není jasný cíl projektu velkých dat nebo jaká infrastruktura je k tomu zapotřebí. Najít a udržet si **kvalifikované zaměstnance** je pro společnosti obvykle obtížné, jelikož je po nich obvykle velká poptávka.



Navíc, technologie velkých dat může být pro začátečníky rozsáhlá a matoucí. Slyšeli jste někdy o Spark, Hadoop MapReduce, Cassandra nebo Hbase? Jedná se o big data technologie s různými nástroji a výhodami.

Technologie se navíc vyvíjejí rychlým tempem, takže společnosti mají často problém držet krok s rychlostí osvojování. Pro společnosti, které zvažují zavedení analýzy velkých dat, tak může být užitečná **rada odborníků**.

Dále je důležité zmínit, že projekty velkých dat jsou velmi **nákladné**. To platí jak pro společnosti, které si zvolí **on-premise model**, tak i pro ty, co preferují **cloudový model**. Rozdílem je to, že s on-premise modelem společnost používá big data software ve svém vlastním datovém centru a je zodpovědná za jeho provoz a údržbu. Naopak v cloudovém modelu je software firmou pouze pronajat a data zůstávají u poskytovatele.

Definice

On-Premise model

...označuje řešení, kdy si firma **koupí** nebo **pronajme** big data software a nastaví jej ve **vlastním datovém centru**. Společnost se musí starat o hardware sama, je zodpovědná za použití softwaru a data zůstávají uložena ve společnosti.

Definice**Cloud model**

...označuje řešení, kdy si firma hradí software velkých dat jako službu; poskytovatel přebírá zodpovědnost za údržbu a použití. Společnost platí pronájem, který zahrnuje náklady na hardware, provoz a údržbu. V případě tohoto řešení jsou data uložena u poskytovatele.

Pokud se společnost rozhodne pro používání on-premise modelu, musí zainvestovat do nového hardwaru a přijmout nové zaměstnance na provoz a údržbu. V případě cloudového řešení musí firma pouze najmout zaměstnance na provoz a údržbu a musí platit za cloudové služby.

Kvalita dat ostatně bývá často nízká a firmy čelí výzvám sladit údaje z různých zdrojů v různé kvalitě. Například, jeden online obchodník může chtít analyzovat data současně ze sociálních sítí, protokolů webových stránek, call center a webů, které mají různé formáty.

Ale i v případě, že všechny výše zmíněné problémy a otázky jsou vyřešeny, není často pro firmy tak jednoduché získat z velkého množství dat **užitečné informace**. Naopak může být velmi nebezpečné, pokud jsou informace vzájemně propojeny a dochází tak k nesprávným závěrům. Například, člověk může být považován bankou, která provádí analýzu velkých dat, za nedůvěryhodného, jelikož žije ve stejném sousedství jako spousta nedůvěryhodných lidí a řídí stejné auto jako lidé, kteří nejsou podle banky kredibilní. Následující příklad také ukazuje, proč je správné využití velkých dat zásadní:

Příklad

Online prodejce se spoléhá na analýzu velkých dat, která je založena na historických údajích o chování zákazníků. Ukáže se, že lidé, kteří kupují v jeho obchodě černé tenisky, často přidávají k nákupu i pár černých ponožek. Maloobchodník odpovídajícím způsobem upraví svoji nabídku na jaro. Nicméně těsně před začátkem jara známý rapper zveřejní na Instagramu fotografii sebe s černými teniskami a žlutými ponožkami. Spousta mladých lidí se tak začne dívat po žlutých ponožkách, které by šly s jejich černými teniskami. Bohužel prodejce nebyl připraven na vysokou poptávku, a tak mu žluté ponožky brzy dojdou. Prodejce jednoduše použil špatnou strategii velkých dat, spoléhal se pouze na historické výsledky a ignoroval jiné důležité zdroje dat, jako jsou například sociální média, konkurenční obchody apod.

Zapamatujte si

V souhrnu, toto jsou hlavní **výzvy**, kterým musí firmy čelit při používání velkých dat:

- **zajistit ochranu dat** – soulad s obecným nařízením o ochraně osobních údajů (GDPR)
- **odborná způsobilost** zaměstnanců – **zdatné používání** rozmanité a rychle se rozvíjející technologie velkých dat
- **vysoké náklady** big data projektů (hardware a software nebo náklady na pronájem, údržbu atd.)
- **nízká kvalita** dat, standardizace dat v různých formátech a s různou kvalitou
- **správná interpretace** výsledků

Jak jste si jistě všimli, big data nabízejí obrovské možnosti a příležitosti, které ještě firmy zdaleka plně nevyužívají. Nicméně, velké množství dat je také spojeno s výzvami a riziky, která by neměla být podceňována a které mohou být pro spoustu lidí znepokojující. Rozhodujícím faktorem, který zajišťuje úspěšné využití dat bez způsobení škody jiným lidem, je proto **zodpovědné a odborné** nakládání s velkými daty.



3.6 Shrnutí

Big data označují velké množství dat, která již nemohou být zpracována klasickým softwarem nebo hardwarem a jejichž zpracování a analýza jsou provedeny pro konkrétní účel. Na rozdíl od **velkých dat**, **malá data** označují data, která jsou díky jejich objemu a formátu lidem dostupná.

S velkými daty se setkáváme v každodenních situacích, například při surfování na sociálních sítích nebo při vyhledávání na Googlu. Pro lepší vymezení velkých dat, analytik Doug Lane navrhl **3-V model**, který uvádí, že big data jsou charakterizována svým **objemem**, **rychlostí** a **různorodostí**.

Big data mohou být použita mimo jiné pro zlepšení strategického směřování společnosti, pro **deep-learning** systémy, v **boji proti zločinu a terorismu**, k **vědeckému zkoumání přírodních jevů** (např. zemětřesení nebo změny klimatu), ale také pro **nezákonná vyhodnocení**, které mohou vést k vydírání nebo podvodům. Rozhodujícím faktorem nejsou ani tak velké objemy dat samy o sobě, ale to, co se s nimi následně stane.

Firmy mohou použít big data ke **zvýšení svého obchodního úspěchu**. Analytici velkých dat umožňují společnostem, kromě jiného, **zlepšovat jejich obchodní rozhodnutí a vyhodnocovat rizika** s větší přesností. Navíc, **efektivnost** obchodních procesů může být **zvýšena** díky analýze, vyhodnocení a propojení dat. Big data pomáhají společnostem ve výzkumu a vývoji vytvářet **předpovědi** o trendech, charakteristikách produktů atd. V poslední řadě, informace získané z velkých dat mohou být použity k nabídnutí **personalizovaného zákaznického servisu**.

Pro úspěšnou analýzu velkých dat je nutná vhodná **strategie velkých dat**, vhodná **podniková kultura**, **personál** s nezbytným **know-how**, **efektivní technologie** a v neposlední řadě **vhodná strategie ochrany údajů**. Analýza a zpracování velkých dat však nejen nabízejí příležitosti a šance, ale nesou s sebou také výzvy a rizika.

Hlavní výzvou pro společnosti je **zajištění bezpečnosti dat** a dodržování **obecného nařízení o ochraně osobních údajů (GDPR)**. Pro firmy je navíc často obtížné najít a udržet si **vhodné odborníky**, kteří zvládnou **komplexní technologii velkých dat**. Projekty velkých dat jsou často spojené s **velkými náklady** a **kvalita dat** je často **nízká**. Nakonec je třeba z výsledků analýzy údajů vyvodit správné **závěry** a musí být učiněna správná **rozhodnutí**.