



TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies



BIG DATA

Lukáš Matějů

21.2.2024 | DPB





ČÁST I.: ORGANIZACE PŘEDMĚTU DATABÁZE PRO BIG DATA



ORGANIZACE

- základní informace
 - přednášející i cvičící
 - Lukáš Matějů
 - lukas.mateju@tul.cz
 - rozsah předmětu 2+2
 - veškeré materiály zveřejňovány na [elearningu FM](#)
- přednášky
 - každý čtvrtok od 08:50
 - budova A, místnost A-A0110
 - účast nepovinná, ale vítaná



ORGANIZACE

- cvičení
 - každý čtvrtok od 12:30 a 14:20
 - budova A, místnost A-A0304
 - samostatné práce volně doplňující přednášky
 - 12 povinných a 12 bonusových úloh
 - každá bonusová úloha je za **0,5** bodu
 - na vypracování a odevzdání úloh je **1** týden
 - odevzdává se **výhradně** na cvičeních (s osobním vysvětlením)
 - za každý týden opožděného odevzdání je **-0,5** bodu
 - finální počet úloh může být ovlivněn odpadnutím výuky
 - 2 povolené absence
 - každá další absence je za **-3** body
 - absence automaticky prodlužuje dobu odevzdání o **1** týden



ORGANIZACE

- zápočet
 - odevzdané a správně vyřešené povinné úlohy ze cvičení
- zkouška
 - prezenční
 - písemná
 - 20 bodů
 - 10 otázek po 2 bodech
 - body ze cvičení jsou přenášeny ke zkoušce
 - zaměřená na základní koncepty probírané v rámci předmětu



ORGANIZACE

- hodnocení
 - dvě varianty
 - jen za bonusové body ze cvičení...
 - 6,0 bodů -> 1
 - 5,5 bodů -> 2
 - 5,0 bodů -> 3
 - povinná docházka na přednášky i cvičení (2 + 2 povolené absence)
 - v případě absolvování písemné zkoušky
 - maximum 26 bodů (20 + 6)
 - ≥ 22 bodů -> 1 ≥ 20 bodů -> 1-
 - ≥ 18 bodů -> 2 ≥ 16 bodů -> 2-
 - ≥ 14 bodů -> 3 < 14 bodů -> 4
 - v případě odpadnutí výuky budou potřebné body adekvátně upraveny



ČÁST II.: BIG DATA



„Až 3,3 miliard uživatelů smartphonů.“
[2019]

„Každou minutu je na YouTube nahráno 300 hodin videí.“ [2017]

„Denně je posláno 140 milionů tweetů.“ [2019]

„Google zaznamenává více jak 63 000
vyhledávání každou vteřinu.“ [2019]

„Uživatelé každý den nahrají na Facebook
více jak 300 milionů fotek.“ [2017]

„Reddit – 25 milionů hlasů každý den.“ [2019]

„PornHub přenese každou vteřinu 147 GB dat.“ [2018]



„Denně je odesláno 294 miliard emailů.“ [2019]

„Instagram zaznamenává denně 4,2 miliardy liků.“ [2018]

„NASA má k dispozici 32 PB dat pozorování a simulací klima.“ [2015]

„Za sekundu je odhadem vytvořeno 1,7 MB dat pro každého člověka.“ [2019]

„Za dva dny vytvoříme tolik informací jako bylo vygenerováno od počátku věků do roku 2003.“ [2014]

„Jen 0,5 % big data je ve skutečnosti analyzováno.“ [2014]



DEFINICE BIG DATA

„Soubory dat, jejichž velikost je mimo schopnosti zachycovat, spravovat a zpracovávat data běžně používanými softwarovými prostředky v rozumném čase.“ [Gartner]

„Big data jsou tam, kde jsou potřeba paralelní výpočty pro zpracování dat.“ [2018]

„Nový nástroj pro vyhledávání relevantních dat a jejich analýzu.“ [Forbes]

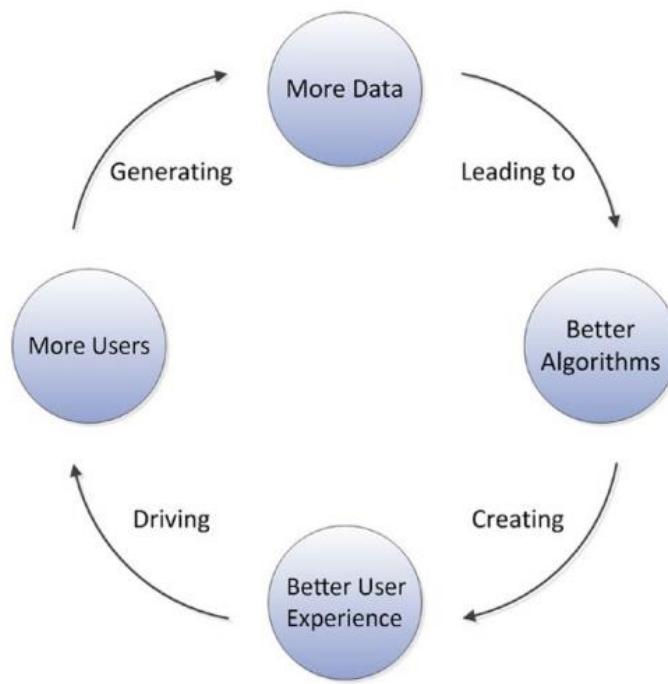


ÉRA BIG DATA

- dnes žijeme v éře big data
 - výrazně ovlivňuje nás každodenní život
- vychází ze dvou základních předpokladů
 - existuje stálý (rostoucí) a rychlý příspun nových dat
 - v roce 2017 proteklo internetem 1.5 ZB (10^{21}) dat
 - pro rok 2022 byl odhad 4.8 ZB
 - inovace ve způsobu zpracování
 - cloud computing
 - propůjčení výpočetních serverů
 - umožňuje provádět výpočty kdekoliv a odkudkoliv
- umožňuje dynamickou a škálovatelnou analýzu dat



ÉRA BIG DATA



HARRISON, Guy. *Next generation databases: NoSQL, NewSQL, and Big Data*. ISBN 978-1484213308.

APLIKACE BIG DATA

- obecná aplikace
 - big data oproti klasickým modelům umožňují
 - kvalitnější modely
 - přesnější výsledky
 - aplikace cílené na jednotlivce
 - aplikace cílené na skupinu
- snaha o pochopení a vytěžení informací z velkého množství dat
 - big data analytics
 - vlády vnímají big data jako prostředek pro vylepšení měst
 - firmy jako způsob porozumění a pochopení potřeb svých zákazníků
 - a získání výhod nad svými rivaly
 - vědci jako možnost uložení a zpracování obrovského množství dat



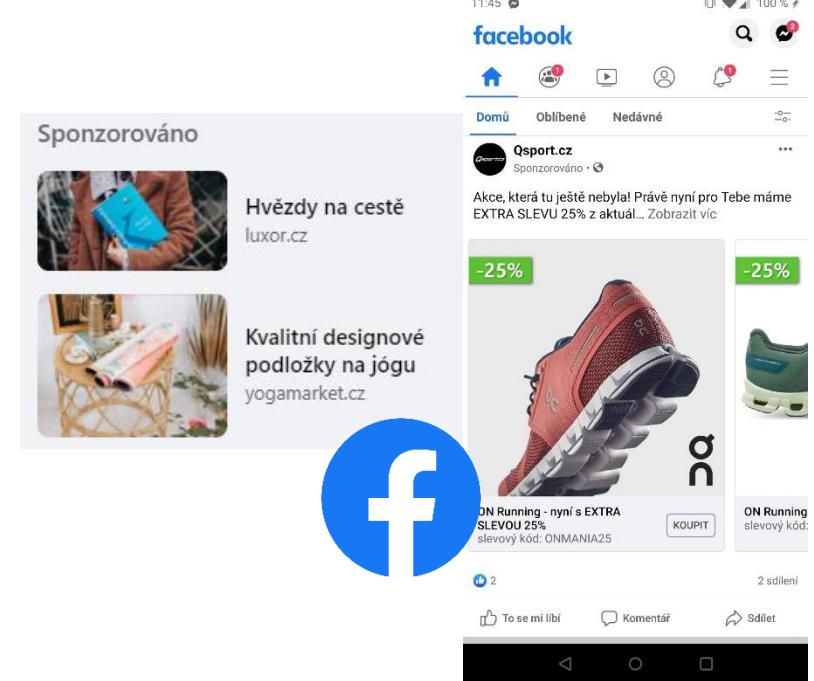
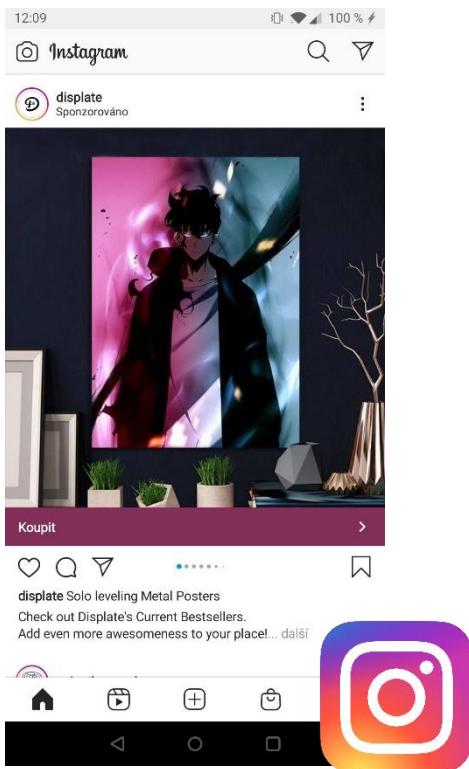
CÍLENÝ MARKETING

- cílené doporučení zboží
- cílená reklama
- určeno na základě aktivity uživatele
 - navštívené stránky
 - historie vyhledávání
 - historie objednávek
 - zhlédnutá videa
 - social media...





CÍLENÝ MARKETING



CÍLENÝ MARKETING



Google Ads

Example of your image ad at 160x600



Acme Business Webinars

Webinars covering every facet of B2B marketing - Sign up for a free intro webinar today

Example of your native ad at 480x120



Anywhere USA's Top B2B Webinar Provider

Webinars covering every facet of B2B marketing - Sign up for a free intro webinar...

Acme Webinar

Visit Site

Example of your image ad at 300x250



Acme Business Webinars

Webinars covering every facet of B2B marketing - Sign up for a free intro webinar today

Acme Webinar

Example of your text ad at 300x250



Acme Business Webinars

Webinars covering every facet of B2B marketing - Sign up for a free intro webinar today

Acme Webinar

Example of your image ad at 728x90

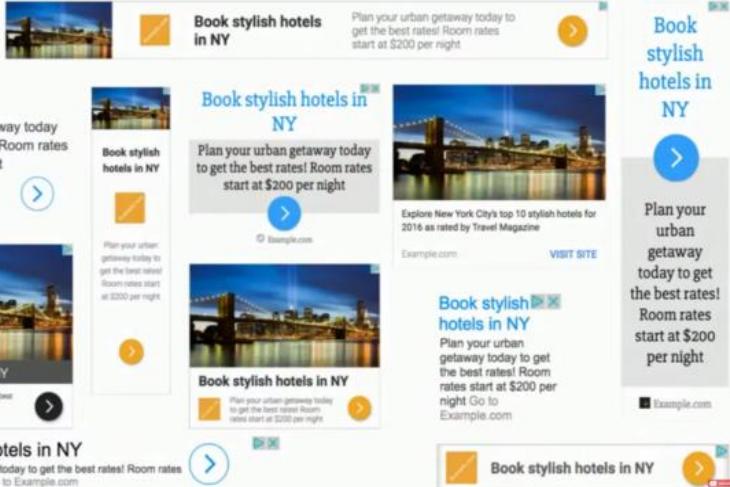


Acme Business Webinars

Acme Webinar

Webinars covering every facet of B2B marketing - Sign up for a free intro webinar today

Acme Webinar



Book stylish hotels in NY

Plan your urban getaway today to get the best rates! Room rates start at \$200 per night

Example.com

Book stylish hotels in NY

Plan your urban getaway today to get the best rates! Room rates start at \$200 per night

Example.com

Book stylish hotels in NY

Plan your urban getaway today to get the best rates! Room rates start at \$200 per night

Example.com

Book stylish hotels in NY

Plan your urban getaway today to get the best rates! Room rates start at \$200 per night

Example.com

Book stylish hotels in NY

Plan your urban getaway today to get the best rates! Room rates start at \$200 per night

Example.com

<https://www.growth-hackers.net/wp-content/uploads/2018/11/Google-Programmatic-Display-Ads-Sample-Example.jpg>

<https://www.wordstream.com/blog/ws/2018/07/23/responsive-display-ads>



DOPORUČOVACÍ SYSTÉMY

- doporučení produktu na základě předchozí aktivity uživatele
 - Netflix navrhuje uživateli filmy / seriály
 - Steam doporučuje hry
 - Amazon zboží
 - cílem je využít znalosti o uživateli k doporučení ideálního produktu
 - udržení pozornosti uživatele
 - prodej produktu



DOPORUČOVACÍ SYSTÉMY

PROTOŽE JSTE HRÁLI ROCKET LEAGUE®



Fotbalové (kopačá) Kompetitivní S online kooperací S nekonečným běháním

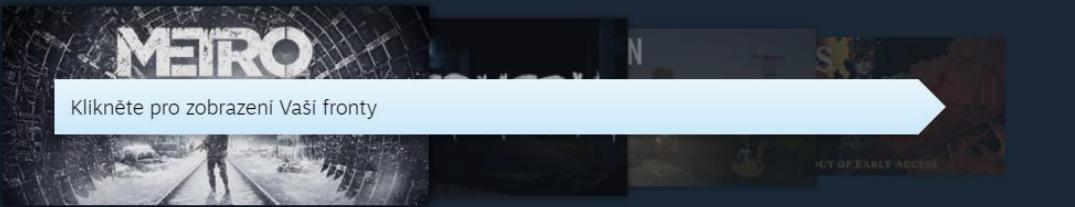
Fotbalové (kopačá) S online kooperací Realistické E-sportovní

Fotbalové (kopačá) Kompetitivní Závodní S důrazem na týmovou hru

Fotbalové (kopačá) S online kooperací Zábavné Kompetitivní

< >

VAŠE FRONTA DOPORUČENÍ



Klikněte pro zobrazení Vaši fronty

ZJISTIT VÍCE



CO BY SE VÁM MOHLO LÍBIT



NA ZÁKLADĚ HER, KTERÉ JSTE HRÁLI PROCHÁZET A PŘIZPŮSOBIT

< >

DOPORUČOVACÍ SYSTÉMY



Top Picks for Lukáš



Doporučené playlisty



Posluněný folk
Playlist • YouTube Music



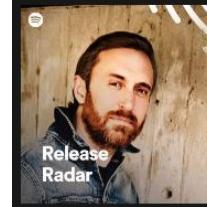
Presenting Ed Sheeran
Playlist • YouTube Music



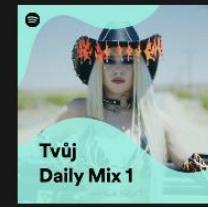
Presenting Škwar
Playlist • YouTube Music

Speciálně pro uživatele Lukáš Matějů

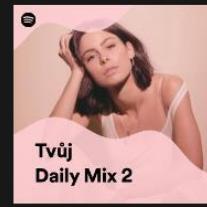
Čím víc budeš poslouchat, tím lepší budou naše doporučení.



Release Radar
Nejnovější hudba umělců, které sleduješ, a nové singly vybrané speciálně...



Daily Mix 1
Ava Max, Sia, Alessia Cara a další žánry

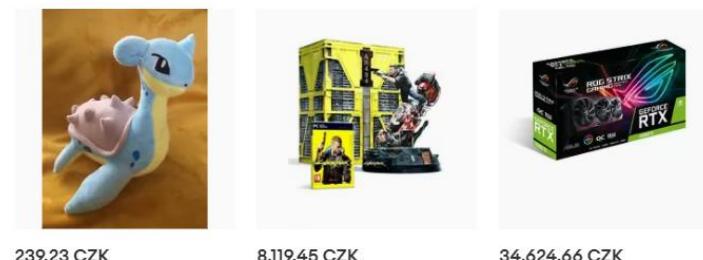


Daily Mix 2
Lena, Ana Kohler, Mike Williams a další žánry



DOPORUČOVACÍ SYSTÉMY

Your Recently Viewed Items | See all →



239.23 CZK 8,119.45 CZK 34,624.66 CZK

Pokémon Go | See all →
Recommended for you



Doporučeno přímo pro Vás



5★★★★ Datacom 10-pack RJ45,
CAT5E, UTP, 8p8c, na drát
59,- od 31,-



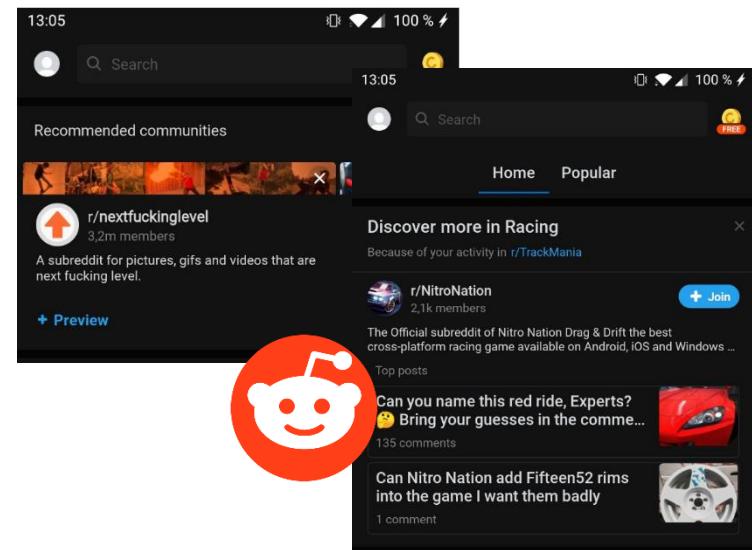
5★★★★ AKASA AK-CBFA01-30 0.3m
89,- od 66,-



499,- 359,-
Philips QP100/51



5★★★★ STX bavlněná rouška bez
potisku - černá (5 ks)
199,- 175,-

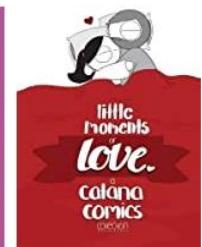


DOPORUČOVACÍ SYSTÉMY

Related to items you've viewed [See more](#)



More recommendations for you [See more](#)



Inspired by your browsing history



amazon.com

ANALÝZA SENTIMENTU

- založená na hodnocení / komentářích uživatelů
 - pozitivní / negativní (NLP)
 - analýza zboží na základě všech recenzí
 - negativní
 - nedoporučovat
 - analýza uživatele na základě jeho recenzí
 - návrh zboží
 - mínění veřejnosti o firmě
 - je potřeba něco změnit?
 - reakce veřejnosti na nastalou událost
 - je potřeba reakce?
 - reakce veřejnosti na nově uvedený produkt
 - je o něj zájem?

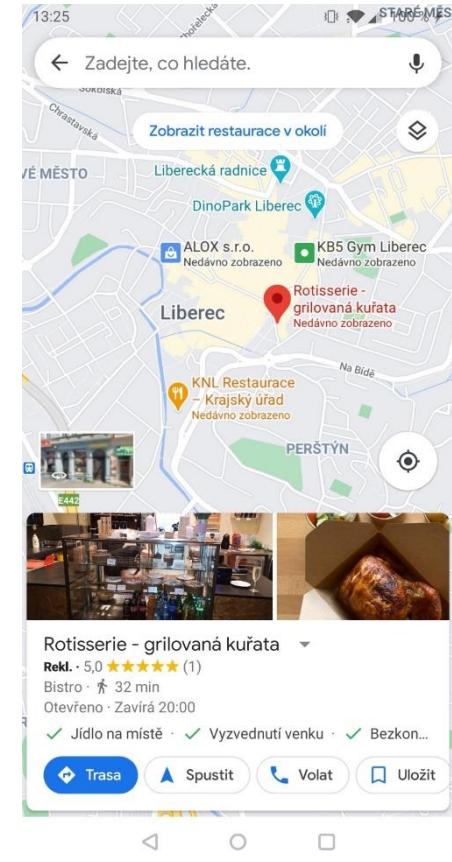
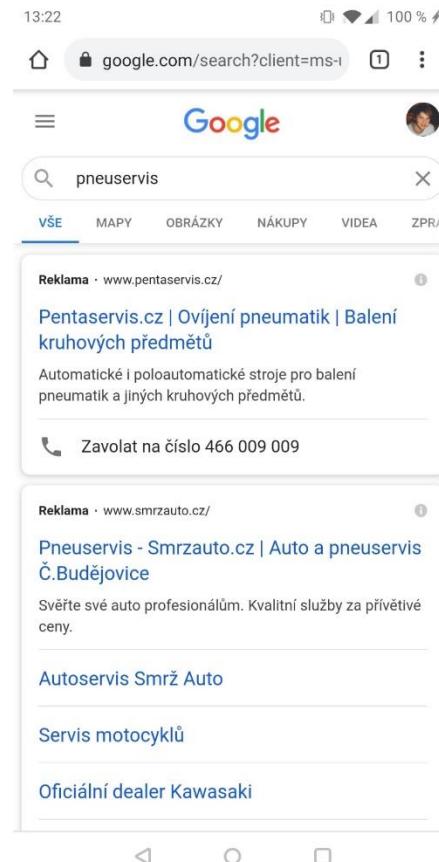
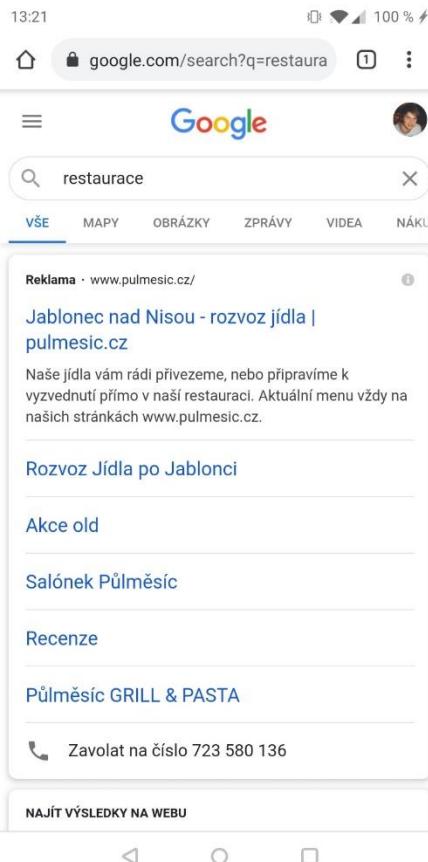
MOBILNÍ REKLAMY

- kombinují informace o uživateli se senzory v telefonu
 - primárně GPS
 - slevové akce určené přímo pro zákazníka
 - doporučení obchodu / restaurace v okolí
 - na základě předchozího vyhledávání zboží
 - na základě stravovacích návyků
 - ...



Google

MOBILNÍ REKLAMY



CHOVÁNÍ SKUPINY

- úprava služeb podle chování cílové skupiny
 - letecké společnosti
 - ranní lety kompletně vyprodány
 - večerní lety s poloviční kapacitou
 - přesun některých večerních letů na ráno
 - platí v USA, v Japonsku je ale situace opačná
 - cílení pro konkrétní skupinu / trh



BIOMEDICÍNA

- obrovské množství dat (lidský genom)
 - až 40 EB v roce 2025
- využití
 - výzkum
 - léčebních postupů, léků, atd.
 - cílená léčba
 - určená přímo na míru pacientovi
 - léčba rakoviny



BIG DATA MĚSTA

- využití veškerých senzorů umístěných ve městě
 - dopravní senzory, satelity, kamerová síť, ...
 - analýza v reálném čase umožňuje
 - zlepšování životních podmínek obyvatel
 - redukci znečištění
 - redukci dopravních zácp
 - úsporu energie
 - potenciálně predikci kriminality

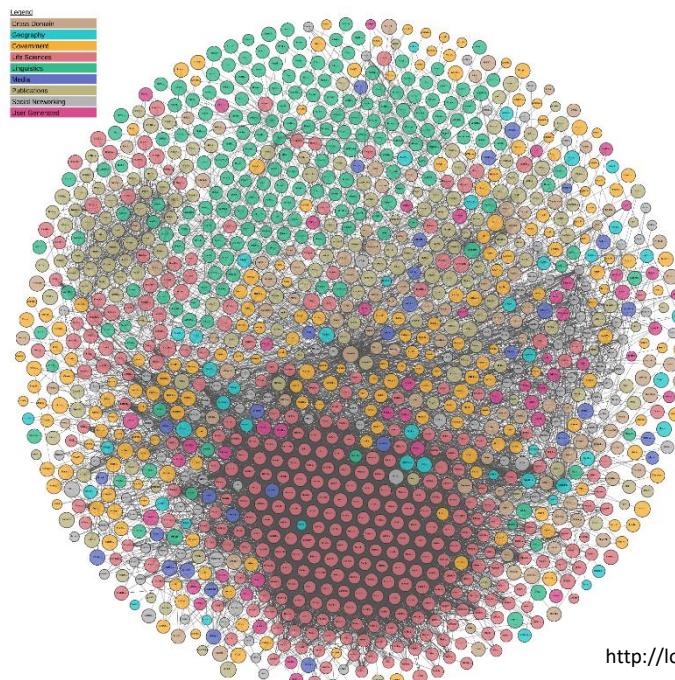


<https://datafloq.com/read/how-create-smart-city-iot-big-data/4101>



ODKUD BIG DATA POCHÁZÍ?

- většina zdrojů existovala již dříve
- změnil se ale náš přístup a množství analyzovaných dat



<http://lod-cloud.net/>





ODKUD BIG DATA POCHÁZÍ?

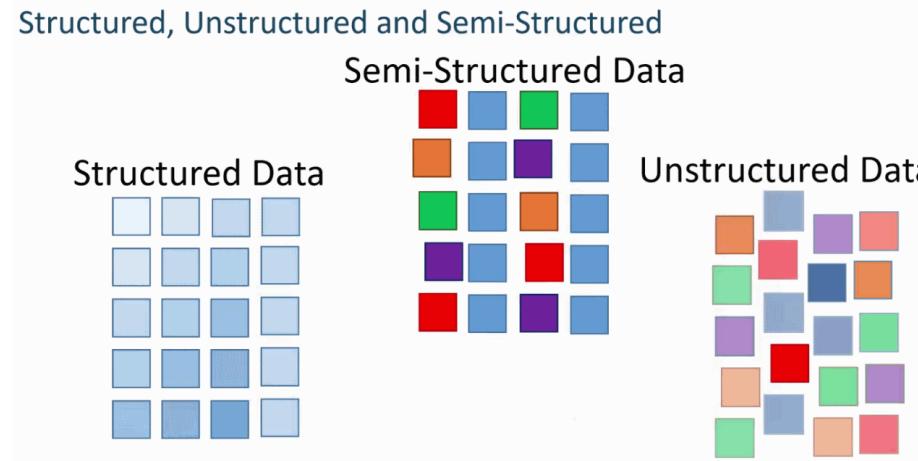
- tři hlavní zdroje dat
 - stroje
 - senzory sbírající data v reálném čase v průmyslu, automobilech, ...
 - environmentální senzory
 - zařízení sledující zdravotní stav
 - lidí
 - social media
 - články, blogy, ...
 - organizace
 - transakce v databázích
- data mohou být
 - strukturovaná, částečně strukturovaná nebo nestrukturovaná
- hlavní přínos spočívá v jejich kombinaci

ROZDĚLENÍ DAT

- strukturovaná data
 - informace formátované a převedené do definovaného datového modelu
 - uložené v definovaných polích umožňujících snadný přístup a čtení
 - strojová, lidská i organizační data
 - snadno analyzovatelná
 - např. relační databáze – relace, atributy
- nestrukturovaná data
 - data v nezpracované podobě, nemají žádný specifický formát
 - obtížné na zpracování kvůli komplexnosti a složitosti
 - flexibilní, mnoho podob
 - social media příspěvky, chaty, satelitní snímky, prezentace...
 - ukládána a analyzována v datových skladech

ROZDĚLENÍ DAT

- částečně strukturovaná data
 - na pomezí mezi strukturovanými a nestrukturovanými daty
 - příkladem může být digitální fotografie
 - samotný obraz nemá pevně definovanou strukturu
 - obsahuje ale strukturované atributy typu datum a místo pořízení snímku, ID zařízení, ...



<https://www.astera.com/type/blog/structured-semi-structured-and-unstructured-data/>





STROJOVÁ DATA

- největší zdroj big data
 - velký hadronový urychlovač generuje 40 TB dat každou vteřinu experimentů
 - Boeing 787 produkuje 0,5 TB dat během každého letu
- chytré zařízení
 - zařízení schopná měřit a produkovat big data (pomocí senzorů)
 - proč chytré?
 - schopnost připojení k dalším zařízením / sítím
 - autonomní sběr a analýza dat
 - mají povědomí o prostředí
 - internet věcí (internet of things)
 - např. chytré hodinky
 - měření teploty, tepu, kvality spánku, nachozené vzdálenosti, počtu schodů, atd.
 - co když je bude mít každý?
 - co když jich každý bude mít více?



STROJOVÁ DATA

- zpracování v reálném čase
 - vztahy se zákazníky
 - detekce podvodů
 - monitorování systému
 - okamžitá analýza a reakce
- zpracování se přesouvá za daty
 - In-Situ
- potřeba škálovatelných výpočetních systémů
- SCADA
 - vzdálený monitoring a ovládání průmyslových procesů



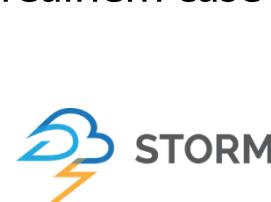
LIDSKÁ DATA

- lidé vytváří obrovské množství dat na internetu každý den
 - Facebook, Twitter, LinkedIn
 - Instagram, YouTube
 - blogy, komentáře
 - vyhledávání
 - textové zprávy, emaily
 - osobní dokumenty
- většina dat je textových a nestrukturovaných
 - složité zpracování
 - nelze použít předdefinované datové modely (NE relační databáze)
 - komplikace
 - množství formátů
 - množství dat a jejich rychlý růst
 - potvrzení je časově náročné (sběr, uložení, těžba, čištění a zpracování)



LIDSKÁ DATA

- zpracování – několik základních open source frameworků
- nástroje pro zpracování a analýzu jsou vyvíjeny od nuly
 - většina založena na Hadoop
 - zpracování velkého množství dat v distribuovaném výpočetním prostředí
 - často potřeba zpracování dat v reálném čase
 - aktualizace na social media
 - tržní data
 - Apache Storm, Spark, Flink
 - ukládání dat
 - NoSQL databáze
 - ukládání dat typicky na výpočetním cloudu
 - zpracování po vrstvách
 - těžba a uložení, předzpracování, analýza



ORGANIZAČNÍ DATA

- důvěryhodná a užitečná data
- liší se výrazně organizaci od organizace
 - transakce, kreditní karty, bankovnictví, akcie, zdravotní záznamy, senzory, atd.
- současné a budoucí použití ale i analýza minulosti
 - predikce prodejů / úspěchů na základě dat a dění ve světě
- vysoce strukturovaná data
 - datový model
 - transakce, referenční tabulky, vazby + metadata doplňující kontext
 - ukládána v relačních databázích (+ SQL)
- riziko
 - datová sila

ORGANIZAČNÍ DATA

- hlavní užitek spočívá v kombinaci s ostatními big daty
 - efektivita provozu
 - zvýšené prodeje
 - vylepšený marketing
 - zlepšená bezpečnost
 - vyšší zákaznická spokojenost
 - příklad: UPS, Walmart
- firmy dnes mohutně investují do big data
 - výsledky ve všech sektorech
 - prvotní firmy získaly náskok nad svými konkurenty





INTEGRACE ROZLIČNÝCH DAT

- hlavní přínos big data je v integraci rozličných dat
 - zpracování, analýza a vytvoření závěrů
- integrace dat = znalost
 - přeměna dat z různých zdrojů na jednotnou a užitečnou informaci
 - správa a přeměna různých dat na něco, co je programově zpracovatelné
- skládá se z mnoha kroků
 - objevení, zpřístupnění, monitorování dat
 - modelování, transformace dat
- integrace zvyšuje hodnotu big dat
 - snižuje komplexitu dat
 - zvyšuje dostupnost dat
 - sjednocuje datový systém



CHARAKTERISTIKA BIG DATA

„Soubory dat, jejichž velikost je mimo schopnosti zachycovat, spravovat a zpracovávat data běžně používanými softwarovými prostředky v rozumném čase.“ [Gartner]

- 3 základní V's
 - objem (volume)
 - množství dat generovaných každou vteřinu
 - různorodost (variety)
 - neustále rostoucí počet forem dat
 - rychlosť (velocity)
 - rychlosť generování dat
 - rychlosť přesouvání dat z jednoho bodu do druhého



CHARAKTERISTIKA BIG DATA

- další často uváděná V's
 - věrohodnost (veracity)
 - zaujatost, šum, abnormality v datech
 - nejistota v pravdivost a věrohodnost dat
 - valence (valence)
 - propojovatelnost dat formou grafů
- a nezapomenout na
 - hodnotu (value)
 - srdce a lepidlo všech ostatních V's
 - jak z dat získat jejich pravou hodnotu?



<https://suryagutta.medium.com/the-5-vs-of-big-data-2758bfcc51d>

OBJEM (VOLUME)

- množství dat generovaných každou vteřinu
 - PB, EB, ZB (10^{21} bytů)
- zpracování a ukládání velkého objemu dat přináší výzvy
 - škálování
 - horizontální / vertikální / kombinace
 - zajištění kapacity úložiště a výkonu na zpracování dat
 - dostupnost
 - přístup k datům a možnost jejich zpracování
 - bandwidth a výkon
 - přístup k datům v potřebný okamžik
- cílem firem je analýza dat
 - zlepšení poskytovaného produktu / služby
 - náskok před konkurencí

Volume



<https://www.youtube.com/watch?v=bAyrObI7TYE>



OBJEM (VOLUME)

data / doména	popis
social media (Twitter)	12+ TB (10^{12}) tweetů každý den (rostoucí trend)
ChatGPT 3	45 TB trénovacích dat
Velký hadronový urychlovač	13-15 PB (10^{15}) v roce 2010
internetová komunikace (CISCO)	667 EB (10^{18}) v roce 2013
Digital Universe	7,9 ZB (10^{21}) v roce 2015

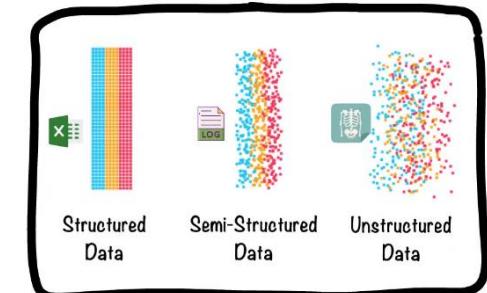
NEVES, Pedro Caldeira a Jorge BERNARDINO. *Big data in the Cloud: A Survey*. ISSN 2365-029X.



RŮZNORODOST (VARIETY)

- rostoucí různorodost forem dat
 - textová data, obrazová data, síťová data, geografické mapy, simulace, ...
- různá různorodost
 - strukturovaná různorodost
 - rozdíl ve struktuře dat (EKG vs. novinový článek)
 - nosičová různorodost
 - forma, ve které jsou data získána (audio nahrávka vs. text)
 - sémantická různorodost
 - jak data interpretovat
 - různorodost dostupnosti
 - data generovaná v reálném čase (senzory) nebo uložená (záznamy)
 - dostupná neustále (kamery) nebo jen příležitostně (sondy, satelity)
- hybridní data – např. emaily

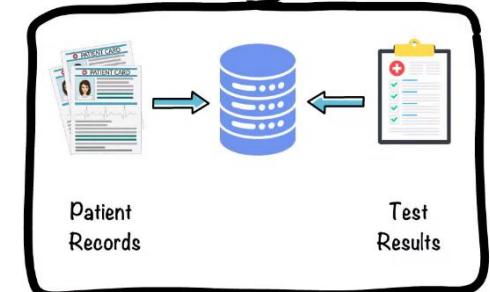
Variety



<https://www.youtube.com/watch?v=bAyrObI7TYE>



RYCHLOST (VELOCITY)



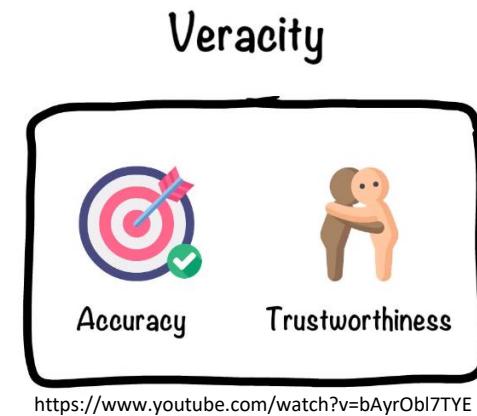
<https://www.youtube.com/watch?v=bAyrObI7TYE>

- rostoucí rychlosť vytvárení big data
 - z rôznych zdrojov sa liší (aktualizace jednou za den vs. každou sekundu)
- rostoucí rychlosť ukladania a analýzy big data (zpracovanie)
- cílem je zpracovanie v reálnom čase
 - vytvorenie reklamy na základe užívateľskej historie a zobrazenie pri hľadaní
 - pomalá reakcia vede k ztrátenej príležitosti
 - napr. kempovanie – zajíma mň dnešné počasie, ne počasie, ktoré bolo pred rokom
 - napr. neštěstí – okamžitý zásah záchranných složiek
- potreba zvážiť rychlosť vytvárenia i rychlosť zpracovania dat
 - zpracovanie môže čakať na data
 - data môžu čakať na zpracovanie
 - rovnováha



VĚROHODNOST (VERACITY)

- odpovídá kvalitě dat
- big data jsou často zašuměná, nejistá nebo nepřesná
 - analýza big data může být jen tak dobrá jako vstupní data
 - hlavní problém u nestrukturovaných dat
- kvalita big data závisí na
 - přesnosti dat
 - spolehlivosti zdroje
 - způsobu vygenerování dat
 - analýze kontextu
- potřeba monitorovat posbíraná data, jejich původ a jak byly dříve analyzovány (x Google Flu Trends)

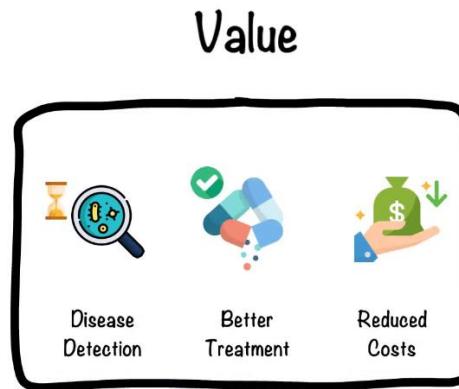


VALENCE (VALENCE)

- propojení dat
 - přímé spojení
 - město – stát
 - zaměstnanec - zaměstnání
 - nepřímé spojení
 - dva vědci jsou propojeni, protože jsou oba fyzici
- poměr množství propojených dat ku možnému počtu propojení
- postupně se časem zvyšuje
 - může vést ke vzniku nových vzorů skupin -> změna
 - potřeba komplexnější analýzy
 - potřeba modelovat a analyzovat valenci
 - detekce skupin, událostí

HODNOTA (VALUE)

- srdce a lepidlo všech ostatních V's
- pravá hodnota dat
 - resp. potenciální hodnota dat z pohledu informací, které obsahují
 - jak ji získat?
- big data ztrácí význam, pokud nenesou hodnotu pro toho, kdo je analyzuje



<https://www.youtube.com/watch?v=bAyrObI7TYE>

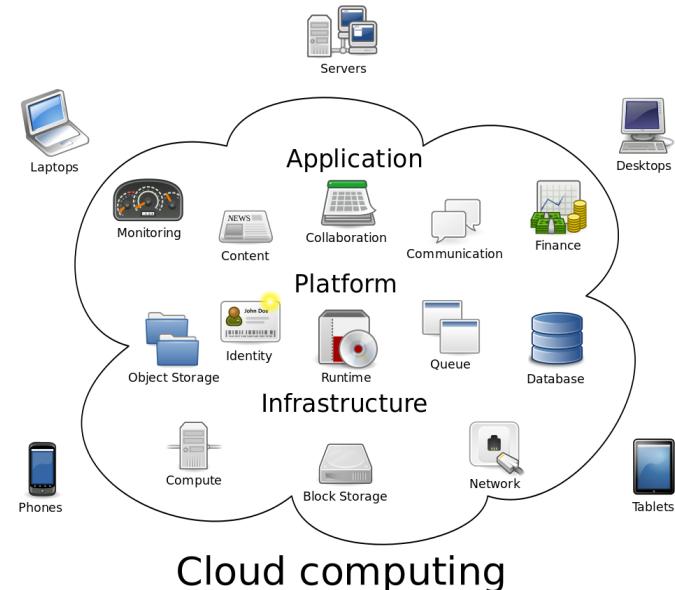


ČÁST III.: CLOUD COMPUTING



CLOUD COMPUTING

- internetový model vývoje a používání počítačových technologií
 - dostupnost systémových prostředků na vyžádání bez aktivní správy klienta
 - nejčastěji datová (cloudová) úložiště a výpočetní výkon
 - systémové prostředky hostované na vzdálených datových centrech
 - spravované cloudovými poskytovateli služeb
 - uživatelé k nim přistupují přes internet
 - prostředky jsou uživatelům pronajímány
 - charakteristika
 - více nájmů – zdroje sdíleny všemi uživateli
 - škálovatelnost – změna prostředků za běhu
 - pay as you go – platba za využité prostředky
 - aktuálnost – zajištěna poskytovatelem
 - přístup přes internet

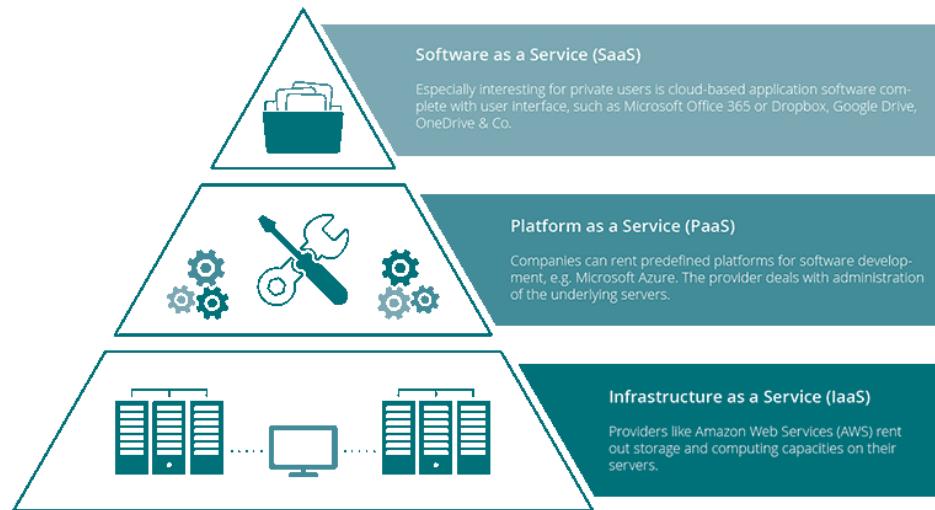


https://en.wikipedia.org/wiki/Cloud_computing



CLOUD COMPUTING

- distribuční model
 - vyjadřuje, co je v rámci služeb nabízeno
 - v základu tři modely s rostoucí mírou abstrakce
 - definice od NIST
 - často znázorňovány jako vrstvy

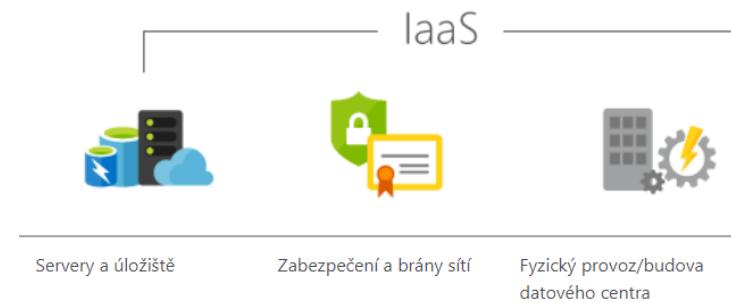


<https://doublehorn.com/cloud-computing-service-models/>



CLOUD COMPUTING

- distribuční model
 - IaaS
 - infrastruktura jako služba (Infrastructure as a Service)
 - předem připravená výpočetní infrastruktura poskytovaná a spravovaná přes internet
 - virtualizované komponenty
 - virtuální stroje, servery, úložiště, síť, ...
 - umožňuje rychle navyšovat kapacitu podle aktuální potřeby
 - Amazon Web Services, Rackspace, Windows Azure

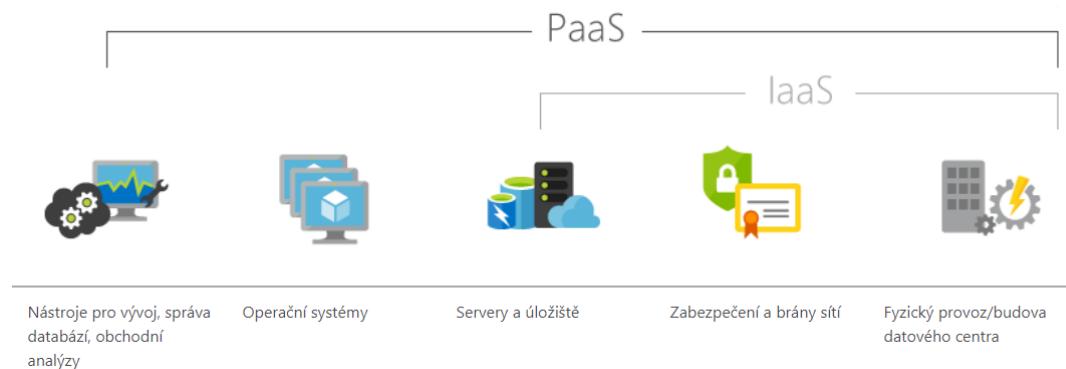


<https://azure.microsoft.com/cs-cz/overview/what-is-iaas/>



CLOUD COMPUTING

- distribuční model
 - PaaS
 - platforma jako služba (Platform as a Service)
 - úplné prostředí pro vývoj a nasazení v cloudu
 - poskytuje prostředky umožňující dodat jednoduché cloudové aplikace ale i propracované aplikace s podporou cloudu
 - zahrnuje nejen IaaS infrastrukturu, ale i další nástroje
 - operační systém, middleware, vývojářské nástroje, systémy správy databází, ...
 - celý životní cyklus webové aplikace – sestavení, testování, nasazení, správa a aktualizace
 - Google App Engine, Heroku, Azure App Services, AWS Elastic Beanstalk

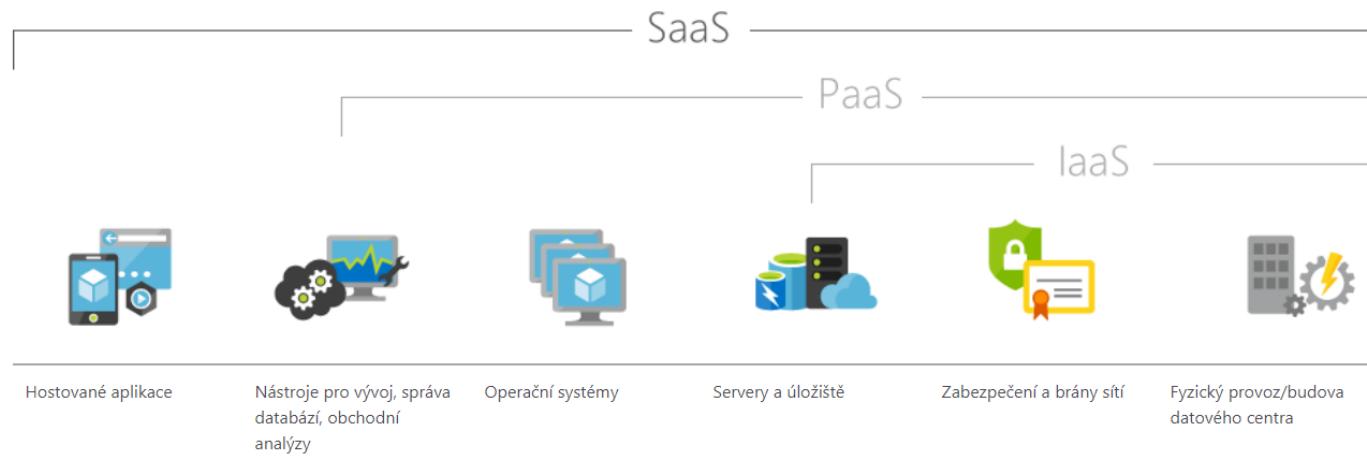


<https://azure.microsoft.com/cs-cz/overview/what-is-paas/>



CLOUD COMPUTING

- distribuční model
 - SaaS
 - software jako služba (Software as a Service)
 - umožňuje uživatelům připojení ke cloudovým aplikacím a jejich používání přes internet
 - úplné softwarové řešení
 - včetně IaaS a PaaS funkcionality
 - Dropbox, Office 365, Google G Suite, Slack



<https://azure.microsoft.com/cs-cz/overview/what-is-saas/>





CLOUD COMPUTING

- distribuční model
 - existují ale i další modely nabízených služeb
 - DBaaS a BDaaS
 - databáze jako služba (Database as a Service)
 - big data jako služba (Big Data as a Service)
 - často SaaS umožňující uživatelům pronajmout databázové služby
 - big data
- cloud virtualizuje zdroje distribuované v clusterech / datacentrech
 - vhodný framework na ukládání a zpracování big data
 - virtualizace velkého množství strojů přináší velká datová úložiště a velkou výpočetní sílu
 - schopnost pojmut, zpracovávat a analyzovat big data



ONLINE KURZY

- big data

<https://www.coursera.org/specializations/big-data>

<https://www.udemy.com/course/the-ultimate-hands-on-hadoop-tame-your-big-data/>

LITERATURA

- SIMON, Phil. *Too Big to Ignore: The Business Case for Big Data*. ISBN 978-1119217848.
- MAYER-SCHÖNBERGER, Viktor. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. ISBN 978-0544227750.
- MARZ, Nathan a James WARREN. *Big data: principles and best practices of scalable real-time data systems*. ISBN 978-1617290343.



A PŘÍŠTĚ?

- kam ukládat big data?
 - databáze pro big data
 - NoSQL databáze





Děkuji za pozornost.
Otázky?

