

# DISA: Data Intensive Systems and Applications

Prezentace laboratoře pro den otevřených dveří na FI MU

RNDr. Vladimír Míč

Fakulta informatiky Masarykovy univerzity, Brno

24. října 2019

# Big data $\approx$ big problem

Rok	2005	2010	2015	2020	2025
Množství digitálních dat	100	2 000	17 000	50 000	175 000

Tabulka: Digitální data, celosvětově, v milionech terabajtů (TB)<sup>1</sup>

- 175 miliard TB
  - Blu-ray disky<sup>2</sup>: sloupec výšky 222 krát obvod Země
  - Stahování 100 Mb/s<sup>3</sup>: 450 milionů let, tj. pětina stáří Země a 4 500 krát doba existence Homo sapiens sapiens
  - Na FI MUNI jen 52,3 milionů let!<sup>4</sup>

<sup>1</sup> „IDC data age study 2025“ sponzor: Seagate, listopad 2018

<sup>2</sup> „DVD s kapacitou 25 GB“

<sup>3</sup> slušná rychlost internetu v soukromém bytě v Brně 2018

<sup>4</sup> Rychlost stahování na FI MU, 14.12.2018 v 0:38: 860 Mb/s

## Čo bolo (to bolo)

---

**Malé** množství **datových typů**  
(*texty, jména, čísla, adresy ...* )

## Sílicí trendy

---

**Boom** datových typů  
(*obrázky, videa, pohybová data, zvuky, otisky prstů, časové řady, 3D grafy, modely, ...*)

---

**Strukturovaná** data

---

**Vytrácející se, dynamická** či **složitá** struktura, ... (*video*)

---

**Jasně vztahy a souvislosti dat**

---

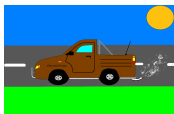
**Vágní, složitě, dynamické** vztahy,  
„big data se ovlivňují navzájem“  
(*internet, sociální sítě*)

---

---

- **Xindong Wu**: *On Big Wisdom* (ICDM 2018 Singapur)
  - „*HAO*“: **integrate** lidské (H), umělé (A) a firemní (O) **intelligence s analýzou big data**

- Nové nároky na zpracování dat
  - Detekce krádeží ve videu z bezpečnostní kamery
  - Doporučení relevantních položek (*reklama, vyhledávání*)
  - Identifikace lidí (*otisky prstů, duhovka, (tváře?)*)
  - Dynamické (kontextové) hledání souvislostí v datech (*setkaly se osoby A a B?*)
- Jak je řešit? **Podobnost!**



- **Laboratoř DISA:** podobnostní vyhledávání v multimédiích ve velkých datech
  - Základní úloha: „dotaz příkladem“  
Cílem je **nalézt nejpodobnější objekty k vybranému**