MUNI
ICS

PV177 – DataScience seminář (Úvod do předmětu)

Tomáš Rebok & **Martin Macák**
Ústav výpočetní techniky MU
Fakulta informatiky MU

Představení vyučujících

RNDr. Tomáš Rebok, Ph.D. (ÚVT MU)

- senior výzkumný pracovník, vedoucí projektů a různých skupin
- dlouhodobá činnost v oblasti náročných výpočtů a zpracování dat
- v posledních letech orientace na oblast datové analytiky

Mgr. Martin Macák (FI MU, ÚVT MU)

- PhD. student
- diplomové i postgraduální téma v oblasti datové analytiky
- člen projektů orientujících se na oblast datové analytiky

MUNI ICS

Cíle kurzu

Seznámení s metodami/nástroji pro analýzy velkých objemů dat

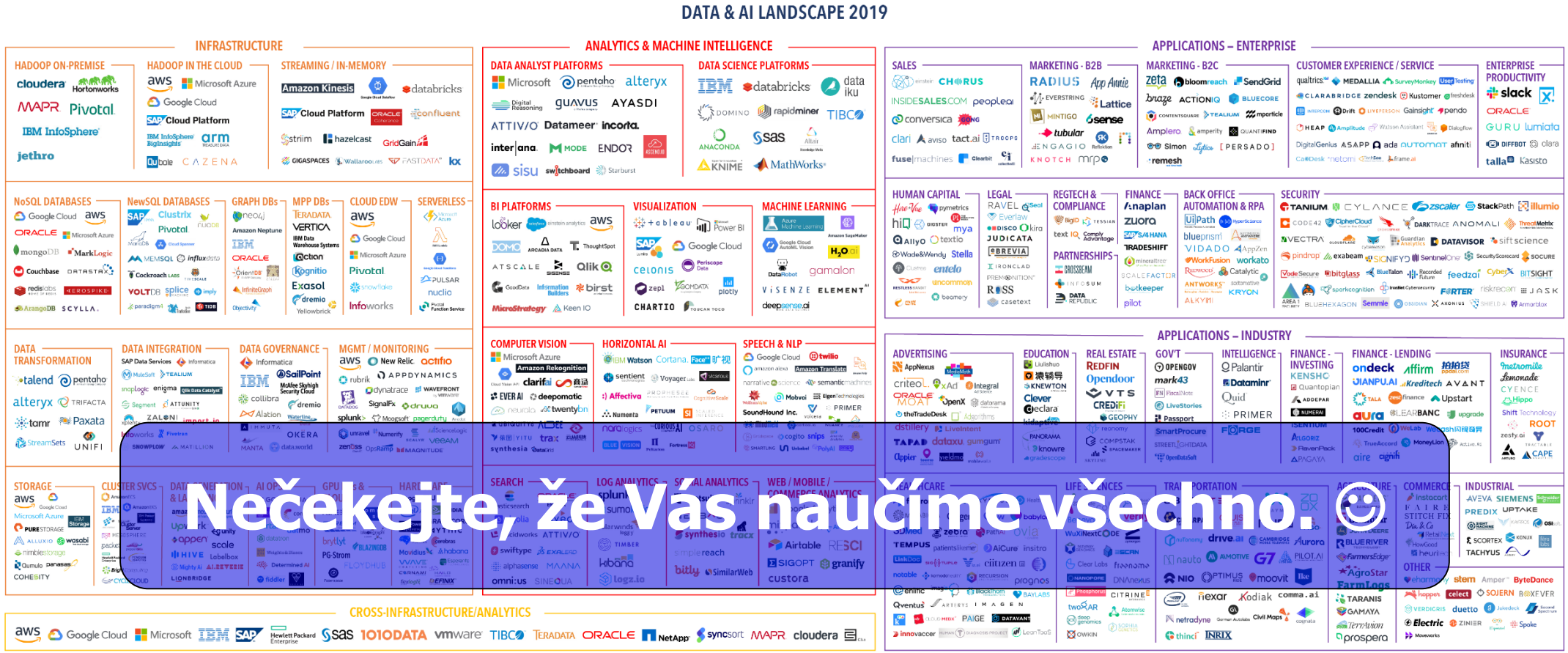
- tzv. *Big Data*
- oblast obrovská, proto jen vybrané nástroje

Praktické seznámení s dostupnými infrastrukturami pro náročné výpočty a analýzu dat

- superpočítačová a gridová centra v ČR a na MU
- založení účtu + realizace výpočtů na těchto infrastrukturách

Realizace vybraného praktického projektu v oblasti zpracování dat

- případně v oblasti rozvoje nástrojů určených pro zpracování dat
- **vlastní témata vítána!**



Nečekejte, že Vás naučíme všechno. ☺

Zázemí kurzu

Centrum CERIT-SC – výzkumné centrum vybudované na ÚVT MU

- původně Superpočítačové centrum Brno (SCB)

Poskytovatel HW a SW zdrojů (5500+ jader)

- SMP uzly
- HD uzly (2624 jader)
- SGI UV uzel 384 jader, 6 TB paměti
- SGI UV uzel 504 jader, 10 TB paměti)
- Xeon Phi cluster
- úložné kapacity (~ 3,5 PB)

Služby nad rámec „běžného“ HW centra

- **zázemí pro kolaborativní výzkum**

MUNI ICS

Zázemí kurzu

Hlavní cíle Centra CERIT-SC:

- flexibilní infrastruktura, vlastní výzkum v infrastrukturních oblastech
- dva hlavní **výzkumné směry**
  - *High-performance computing* – akcelerace výpočtů, GPU computing, …
  - ***Big Data analytics + AI***
- úzká spolupráce mezi informatiky a partnery centra
  - výpočetní a úložné kapacity jsou pouze nástrojem
  - zaměření na inteligentní a nové použití těchto nástrojů
  - synergický posun informatiky a spolupracujících věd (kolaborativní výzkum)

Zázemí kurzu

Snaha o maximální zapojení studentů

- **bakalářského -> magisterského -> doktorského studia**
  - vedení závěrečných prací v praktických a užitečných oblastech
  - možnost zapojení studentů do řešených projektů
    - možná podpora finančními granty

Zázemí kurzu

CERIT-SC – vybrané příklady spoluprací (datová analytika)

- spolupráce s Policií ČR
  - vývoj nástrojů pro datovou analytiku kriminálních činů
    - uživateli jsou policejní analytici
  - příležitostná spolupráce na analýze dat reálných kauz
- spolupráce s výzkumnými partnery (uvnitř i vně MUNI)
  - mnoho spoluprací na pomezí IT a spolupracujících věd
  - Ústav výzkumu globální změny, bioinformatika a analýzy genomu, Ústav fyziky materiálů AV ČR, …
- spolupráce s komerčními subjekty a státními organizacemi
  - aktuálně se rozvíjející spolupráce s RedHat a.s.
  - dlouhodobá spolupráce se společnostní MycroftMind a.s.

Zázemí kurzu

Zázemí kurzu nabízí možnost skloubení výuky s pokročilou praxí

- můžete podpořit svou přípravu do budoucí (komerční) praxe či akademické kariéry
  - případně rozvíjet se i v non-IT oblasti, která je pro Vás zajímavá
- spolupracemi již prošlo mnoho studentů
  - dlouhodobé zkušenosti (cca 8 let)
  - velmi pozitivní zpětná vazba
    - získaná praxe pro hledání zaměstnání, nabídka pracovních pozic u partnerů, dlouhodobější spolupráce s ÚVT/CESNETem, …
- **Vaše dosavadní znalosti a zkušenosti nejsou podmínkou, <u>důležitá je vlastní motivace</u>**

MUNI ICS

Průběh kurzu

aneb *„O zajímavý obsah se podělíme“*

Teoretické přednášky

- úvodních cca 6 týdnů
- doplníme o zvanou/é přednášku/y dalších kolegů
  - specialistů na oblast analýzy/zpracování dat
  - máte nějaké podněty, co byste chtěli slyšet?
- slidy budou dostupné v IS MU (po proběhlém kurzu)

Zadání praktického projektu

- práce ve skupinách (3 studenti)
- témata nejen zajímavá, ale i užitečná (Vaše výsledky budou prospěšné)

Průběžné konzultace při zpracování projektů

Prezentace výsledků, závěrečné shrnutí

Podmínky úspěšného ukončení

- účast na přednáškách (na většině)
- realizace praktického projektu a prezentace výsledků
- **vlastní krví stvrzená celoživotní oddanost datové analytice** a **budoucí spolupráce** ☺ ☺ ☺

Historické praktické projekty – jaro 2019

Ústav výzkumu globální změny AV ČR (CzechGlobe)

- příprava analytických pohledů pro analýzu dat meteorologických měření
  - v nástroji Kibana (+ ElasticSearch)
- návaznost běžící diplomovou prací

Policie ČR

- příprava nástroje pro nahrávání dat do skladu rozsáhlých heterogenních dat
  - podpora běžícího projektu
- 4 studenti zapojeni do běžícího projektu

Masarykova univerzita – CESNET

- analýza dat jednotného přihlášení
  - odhalování abnormálního chování uživatele
- předběžná domluva na zapojení studentů zapojených do tohoto projektu

Historické praktické projekty – podzim 2019

Sběr a analýza dat využití IT infrastruktury MU

- MU disponuje rozsáhlou IT infrastrukturou, jejíž využití je nezbytné sledovat pro účely rozhodování o dalších investicích do ní
  - disková úložiště, výpočetní kapacity (servery), síťová infrastruktura, služby, …
- cíle projektu: příprava infrastruktury a produkčního systému pro sběr a vyhodnocení dat + příprava typových datových analýz

Analýza/sběr/vizualizace dat z energetického sektoru

- ve spolupráci se společností MycroftMind a.s.
- cíl projektu: analýza dat topologie smartmeterů energetické soustavy
  - a odhalování abnormalit v nich
- pokračující spolupráce s 1 studentem

Průzkum a dokumentace (pokročilých) vlastností analytického nástroje Kibana (ElasticSearch) a rozvoj integrující platformy CopAS

- s primárním zaměřením na analýzu síťových toků
- pokračující spolupráce se 2 studenty

MUNI ICS

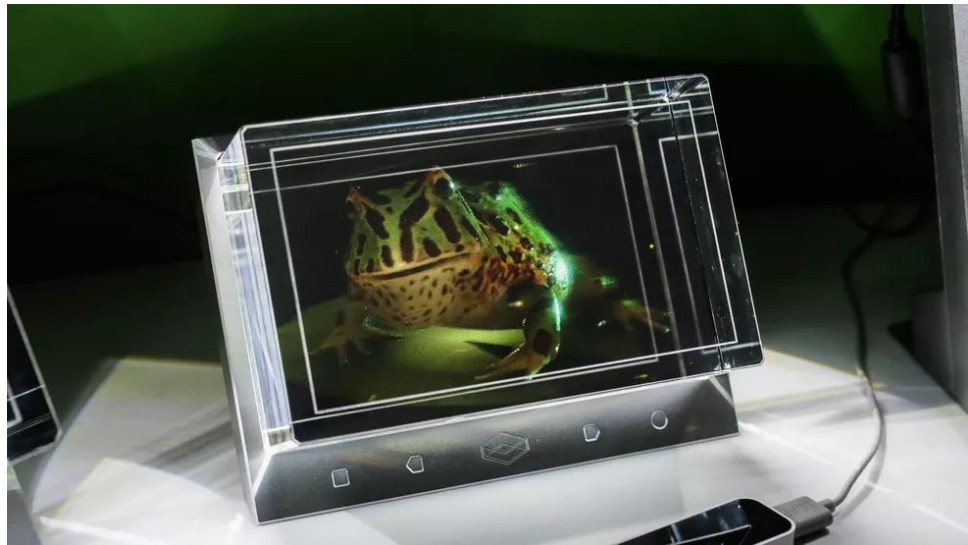
Aktuální praktické projekty – jaro 2020

Prozatimní témata (budou dopřesněna)

- škálovatelnost grafových databází – až na samou hranici jejich schopností
  - výsledky budou využity pro reprezentaci a analýzy proteinů
- analýza obrazových dat ve spolupráci s Moravskou zemskou knihovnou
  - identifikace obrázků v digitalizovaných dokumentech a jejich klasifikace
  - vyhledávání podobných obrázků
  - identifikace osob na obrázcích
  - vyhledávání knih podle obálky
- zpracování leteckých dat Ústavu výzkumu globální změny
- …

a možná i překvapení …

- praktické využití
  největšího holografického displeje v ČR
  - včetně ovládání Kinectem
  - možné využití vícero týmy




thank you
danke
謝謝
рахмат
ngiyabonga
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
vinaka
mersi
barka
welalin
tack
spas
kia ora
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
gracias
blagodaram
спасибі
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
köszönöm
hvala
gràcie
bayarlalaa
bedankt
enkosi
dziękuję
akun dankon ačiū
хвала
asante
manana
obrigada
tenki
murakoze
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
sobodi
děkuji
sagolun
sukriya
kop khun krap
taiku
obrigado
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
grazie
arigatô
takk
dakujem
trugarez
kam sah hamnida
rahmat
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
mercé
мерси
감사합니다
xièxie
ευχαριστώ
merci

MUNI
ICS

MUNI

ICS

MUNI

ICS