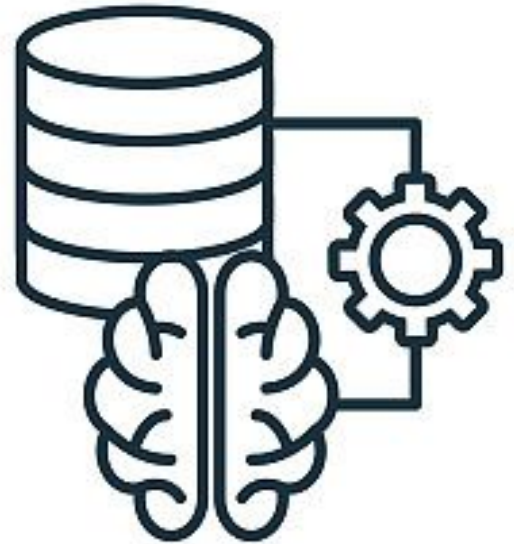


IT infrastruktura pro AI

Masarykova univerzita

Popis systému

- Budování infrastruktury na UVT MU pro uložení velkých dat a trénink AI modelů
- Systém by měl být nabízen v rámci univerzity i (potenciálně) firmám
- Klienty jsou
 - Vědecké týmy vyvíjející AI pro velká data
 - Firmy používající AI v rámci vývoje produktů



MU = Masarykova univerzita

UVT = Ústav výpočetní techniky na MU

Analýza situace - důvody ke změně

- Několik GPU serverů, nakoupeno ad hoc na základě požadavků jednotlivců
- Systém není srozumitelně nabízen mimo UVT
- Finanční ohodnocení služeb neexistuje, systém využíván interně na UVT, financován z infrastrukturních projektů

- Systém v podstatě spravují dva lidé, jen jeden je stálým zaměstnancem MU
- Nejasné řízení: Správci podléhají vedení UVT, vedení Cerit-SC, vedoucím pro jednotlivé oblasti (data, výpočetní infra), atd.

- Vývoj AI systémů na UVT je možný pouze experty na infrastrukturu nejasný interface mezi výzkumníky, AI vývojáři a správci infrastruktury

Cílový stav

- Datový a výpočetní systém pro AI vývoj na MU s dostatečnou kapacitou pro
 - Uložení dat výzkumných pracovišť
 - Trénink AI modelů nad těmito daty
- Jasná nabídka služeb pro klienty, zejména z řad vědců na MU i mimo MU
 - Škálované rozhraní pro experty i pro neexperty
 - Technické i lidské zdroje pro správu metadat a AI systémů
- Finanční ohodnocení služeb uvnitř MU i vně
- Stabilní a průhledné řízení
 - Řízení několika týmů: **Infrastruktura, data, AI vývoj**
 - Koordinace těchto týmů

Síly působící pro změnu a proti změně

Pro změnu:

- Výzkumné týmy na MU. Nutnost zapojení AI do výzkumu ve všech oblastech.
- Vedení UVT - zájem budovat funkční služby v AI

Proti změně

- Fakulta informatiky MU - vzájemná nedůvěra s UVT
- Správci systému na UVT - více práce, obavy z velké změny
- Vedení univerzity - finanční náklady

- Nejasná organizace na UVT - nejasný ownership a odpovědnost
- Ceny HW pro AI - rostoucí poptávka zvyšuje ceny a snižuje dostupnost

Plán změny řízení

Ustavení struktury týmů a řízení

- **Správa infrastruktury** - lze rozšířit aktuální pracoviště
- **Správa dat pro AI** - *datoví správci*, sjednocují přístup k datům
- **Správa AI systémů** - trénink modelů, jejich registrace a prezentace
- *Vedení nadřízené těmto týmům* - odpovědný koordinátor
- *Právní podpora* - ošetřování citlivých dat a AI modelů

Nutno zajistit škálování s počtem klientů!

Plán technologické změny 1

Datová a výpočetní infrastruktura

- Sběr informací o potenciálních uživatelích na MU
Objemy dat, nároky na výpočetní infrastrukturu
- Vyhodnocení informací a plán struktury IT zařízení
Rozšíření stávajícího cloudu, ne kompletní změna přístupu
- Výběrové řízení na dodavatele techniky
- Nákup techniky, instalace

Plán technologické změny 2

Datová správa

- Segmentace potenciálních uživatelů dle požadavků na datovou infrastrukturu
Typy dat, různé zdroje (fyzika, chemie, sociologie, medicína, atd.)
- Selektce software pro správu metadat pro jednotlivé segmenty
 - Rozhodnutí, zda jednotný systém nebo více systémů pro různé segmenty
 - Identifikace bodu spojení systémů datových toků
- Instalace SW a propojení s infrastrukturou

- Propojení správců dat se správou infrastruktury
Vymezení **hranic odpovědnosti** správců infrastruktury a správců dat!

Plán technologické změny 3

Správa AI systémů

- Zjištění požadavků uživatelů na AI systémy
Analýza malých dat vs trénink jednoúčelových neuronových sítí vs trénink velkých modelů
- Selektce softwarových nástrojů dle požadavků, nákup pokud je nutný
- Instalace SW a propojení s infrastrukturou
- Propojení správců AI systémů se správou infrastruktury
Vymezení **hranic odpovědnosti** správců infrastruktury a správců AI systémů!

Vyhodnocení změny

- *Ustavení základních týmů* - ukončená výběrová řízení, kick-off meeting
- *Datová a výpočetní infrastruktura* - IT benchmarking výkonu a spolehlivosti
- *Datová správa* - uživatelský test
čas a přívětivost ukládání dat a vyhledávání v datech
- *Správa AI modelů* - uživatelský test
dostupnost, spolehlivost a efektivita celého systému AI vývoje
- Dlouhodobé vyhodnocení funkčnosti v provozu



Financování

- Infrastrukturní projekty pro vybudování systému
- Rozpočet MU pro udržitelnost
MU musí akceptovat, že AI je nezbytná IT služba
- Příjmy z pronájmu AI infrastruktury mimo MU