

# Superset

## Popis řešení

### Co je Apache Superset

Jedná se o moderní BI řešení postavené na otevřené technologii. Umožňuje vizualizovat data v mnoha podobách - od jednoduchých koláčových grafů po komplexní vizualizace využívající mapové vrstvy. Hlavní vlastnosti:

- Intuitivní rozhraní pro vizualizaci datasetů a vytváření interaktivních dashboardů
- Množství vizualizačních prvků
- Nástroj pro vytváření vizualizací, který nevyžaduje znalost programovacího jazyka
- SQL IDE pro přípravu dat
- Vrstvu pro definici dimenzí a metrik
- Podporu pro většinu SQL databázových řešení
- Asynchronní cachování a dotazy
- Rozsáhlé možnosti nastavení oprávnění přístupu k datasetům a funkcionalitám
- Integrace s většinou autentikačních řešení
- Rozšiřitelnost o vizualizační pluginy
- API pro přizpůsobení pomocí vlastního vývoje
- Škálovatelnost

**Zdroj:** <https://superset.apache.org/docs/intro>

### K čemu lze využít Apache Superset pro piráty

Disponujeme množstvím systémů, které generují data všeho druhu - tato data však nejsou nijak konsolidována pro využití v operativním, či dokonce strategickém rozhodování. Využití řešení, které nám umožní s minimálními náklady (finančními i časovými) tato data snadno vizualizovat, ale především **umožnit jejich uživatelům si vizualizace samostatně vytvářet** otevřeme dveře k tomu, aby rozhodování ve vnitrostranických záležitostech bylo méně opřené o odhady a zkušenost jednotlivců a více o reálná data o naší organizaci a našem chování na různých úrovních

### Příležitosti

Apache Superset je možné integrovat s drtivou většinou běžně používaných technologií, včetně těch využívaných námi. Z toho vyplývá vysoká pravděpodobnost, že zprovoznění a integrace nebude nepřiměřeně náročné a přidaná hodnota převáží

Možnost vytvářet vizualizace bez nutnosti psát kód může přinést kreativní pohledy na naše data, které sice vzniknou pro konkrétní úkol či konkrétní úroveň řízení, ale budou mít natolik vysokou přidanou hodnotu, že přinesou benefit i na dalších úrovních a v dalších týmech.

### Rizika

Úvodní zprovoznění a integrace se systémy, jakož i úvodní definování rolí a z nich vyplývajících přístupů k různým úrovním dat a funkcí zabere nějaký čas a bude vyžadovat kapacity TO. Stejně tak provoz a aktualizace. Pokud vzniknou požadavky, vyžadující zásah ze strany TO, bude nutné je provádět v rozumném časovém horizontu, jinak hrozí že nástroj bude uživateli opuštěn (BI nástroje s sebou obecně často nesou vysoké nároky jejich uživatelů na kvalitu podpory a rychlost odezvy na požadavky).

Nastavení přístupu k datům bude vyžadovat velkou pečlivost, aby pomocí vhodně formulovaných dotazů (a pomocí vhodně zvolených filtrů) nebylo možné deanonymizovat data nebo k již dostupná data z konkrétního systému obohatit o další data, která na úrovni jednotlivých záznamů dostupná být nemají.

## Proof of Concept (POC)

### Úvod

Pro ověření smysluplnosti nasazení a provozu tohoto řešení (a tedy smysluplnosti vynakládání vzácných zdrojů v podobě času a financí TO) je žádoucí provést menší projekt jako proof of concept. Zde uvedené navržené role, datasey a reporty vycházejí z vlastní potřeby autora dokumentu jako člena PKS a z konzultace s pověřeným členem PO (jmenovitě Jakubem Dedkem). V případě, že z diskuse vyplynou jiné, vhodnější role, datasey a vizualizace, je třeba mít na paměti, že POC by měl být spíše menšího rozsahu. Není tedy vhodné seznam v takovém případě seznam rozšiřovat - vhodnější je existující položky seznamu nahradit jinými a zachovat tak rozsah projektu víceméně nezměněný.

### Návrh rolí konzumentů

- Člen PKS
- Člen PMS
- Člen PO

### Zdůvodnění:

Členové PKS a PMS vykonávají obdobné činnosti a k ní mohou využít obdobná data - ale v různém rozsahu. V rámci POC jsou navrženy tyto dvě role najednou, takže zároveň pokrývají podporu rozhodování pro „běžnou operativu“ a zároveň umožní vyzkoušet různou úroveň řízení přístupu k datům (na úrovni konkrétního MS vs. na úrovni celého kraje, tedy všech jeho MS)

Členové PO pak mají z podstaty činnosti odboru i své činnosti přirozený zájem mít přehled o členech, příznivcích a některých aspektech jejich chování pro efektivní organizaci vlastní práce a podporu dosahování našich cílů. Zároveň mají největší rozsah přístupu k citlivým údajům, lze tedy předpokládat největší přidanou hodnotu z vizualizace vhodných dat a zároveň nejmenší potřebu omezovat přístup k datům. Všechny datasey pro PKS a PMS tak budou dávat smysl i pro PO, ale bez omezení přístupu k datům. Kromě toho PO bude mít zájem využívat i datasey dle vlastní definice

## Datasety

Datasety jsou navrženy tak, aby maximálně využívaly již existujících a propojitelných dat, jejichž využití v praxi je buď vysoce pravděpodobné nebo dokonce opakovaně poptávané a aktuálně řešené workarouny.

Zároveň všechny navržené datasety jsou takové povahy, že všechny tři navržené role je mohou vytěžit - a každá jiným způsobem. Lze tak ověřit, zda dosáhneme rozmanitého a pro uživatele cenného využívání stejných dat k různým účelům v různých rolích. Jako bonus, datasety pro takto definované role mohou být v jisté míře zajímavé i pro členy RP a RV - v takovém případě lze relativně snadno POC rozšířit a zavést role RP a RV využívající stejné datasety s širším přístupem k datům.

- Aktivita uživatelů v Heliosu
  - **Obsah:** Dataset bude obsahovat anonymizovaná data o provedení hlasování uživatelů
  - **Účel:** Bude sloužit k ověření relativní účasti na hlasování (za jednotlivá MS, za KS jako celek) jak na hlasování na úrovni oblastních sdružení (MS, KS) tak na úrovni celostátní
  - **Poznámky:** Data musí být obohacena o organizační strukturu (aktuálně na fóru) a zohlednit vývoj členské základny v čase
- Aktivita uživatelů na systémech strany sloužících k diskusi
  - **Obsah:** Dataset bude obsahovat anonymizovaná data o provedení akcí na fóru a zulipu, případně dalších diskusních systémech
  - **Účel:** Bude sloužit k ověření aktivity členů ve vnitrostranické diskusi (za jednotlivá MS, za KS jako celek) jak na hlasování na úrovni oblastních sdružení (MS, KS) tak na úrovni celostátní
  - **Poznámky:** Data nesmí obsahovat detail o typu akce, aby nemohla být zneužita. Palcování, příspěvky, zakládání vláken apod. bude mít stejnou váhu a bude agregováno do jediné hodnoty
- Data operativního rozhodování z redmine (podání, žádosti)
  - **Obsah:** Dataset bude obsahovat základní data z redmine, která umožní vytvářet přehledné dashboardy a upozornění
  - **Účel:** Bude sloužit k zjednodušení kontroly úkolů na úrovni předsednictva oblastního sdružení, primárně s cílem dosáhnout systematického a včasného odbavování podání, sekundárně pak k vyhodnocení používání redmine členy oblastního sdružení (od řadových členů po členy PKS)
  - **Poznámky:** Není třeba obsah konkrétních tiketů, ale musí být k dispozici všechno, co lze použít k řazení a značkování - tak aby si uživatel mohl sestavit dashboard dle své potřeby a dle toho i případná upozornění. Zejména je tedy třeba Název, stav tiketu timestamp poslední změny, Termín vyřešení, typ issue a obdobná pole. Dále pak propojení na základní data o uživatelích, tak aby bylo možné vyhodnocovat aktivitu členů a předsednictev, ideálně s vazbou na „míru nalodění“ dotyčného (tedy např. absolvovaná školení)
- Data o členech a příznicích z chobotnice
  - **Obsah:** Dataset bude obsahovat data z chobotnice, která umožní sledovat události (eventy) s časovou, geografickou a procesní dimenzí
  - **Účel:** Bude sloužit k vyhledávání vzorců chování a slabých míst procesů pro členy PO. Zejména pro sledování statistik o nalodování zájemců o členství - dobách trvání jednotlivých kroků, vyhodnocování rozdílů mezi oblastními

sduženými nebo mezi jednotlivými osobami odpovědnými za nalodění nováčka, vlivu proškolení členů na rychlost a míru jejich zapojení (zaevidovaného v systémech strany) apod.

- **Poznámky:** Chobotnice je stále ve vývoji, proto kromě využití stávajících dat bude třeba plánovat s dalším rozvojem chobotnice i rozšiřování tohoto datasetu

## Vizualizace

### Úvodní vizualizace

V rámci POC budou vytvořeny úvodní jednoduché vizualizace „1:1“, budou nabídnuty uživatelům v daných rolích.

### Podpora

Bude uspořádáno základní školení, jak si vytvářet vlastní vizualizace a vytvořen jasně definovaný kanál pro podporu těch, co toto řešení budou chtít vyzkoušet.

## Vyhodocení

S odstupem 3 měsíců bude užívání Apache Supersetu vyhodnoceno. TO zajistí následující následující podklady

- Základní data o využití aplikace (počet uživatelů, počet vizualizací a jejich využití, počet požadavků na podporu/pomoc od uživatelů a vývoj všeho uvedeného v čase)
- Zpětnou vazbu uživatelů
  - prostřednictvím follow-up setkání s aktivními uživateli řešení
  - Prostřednictvím dotazníku s jednoduchým individuálním hodnocením užitečnosti aplikace a jejího využití respondenty v budoucnu
- Náklady na zprovoznění a provoz řešení, náklady na vytvoření existujících datasetů (jako indikátor nákladů dalšího rozvoje), časovou náročnost podpory

Na základě těchto podkladů pak bude možné rozhodnout o dalším osudu projektu - jeho dalším rozvoji, zakonzervování nebo ukončení

Připravil J. Hora

Konzultace: Matěj Kubíček, Jakub Dedek, ...

PKS Praha