



Evaluation of Apache Superset

Supervisors: Miroslav Potocky, Lukas Gedvilas, Alberto Aimar

ELIF OHRI

15/09/2022

Project Description

Aim

In scope of exploring possible alternatives for license free BI/Data Visualization, evaluation of **Apache Superset** software to replace the enterprise version of **Pentaho**

Description & Scope



- Evaluation of the required **configuration** and **dependencies**
- Proof of concept **containerized installation** of Superset instance in **Data Integration & Reporting (DIR)** team's Kubernetes infrastructure
- Comparison of Superset's capabilities with existing Pentaho software stack

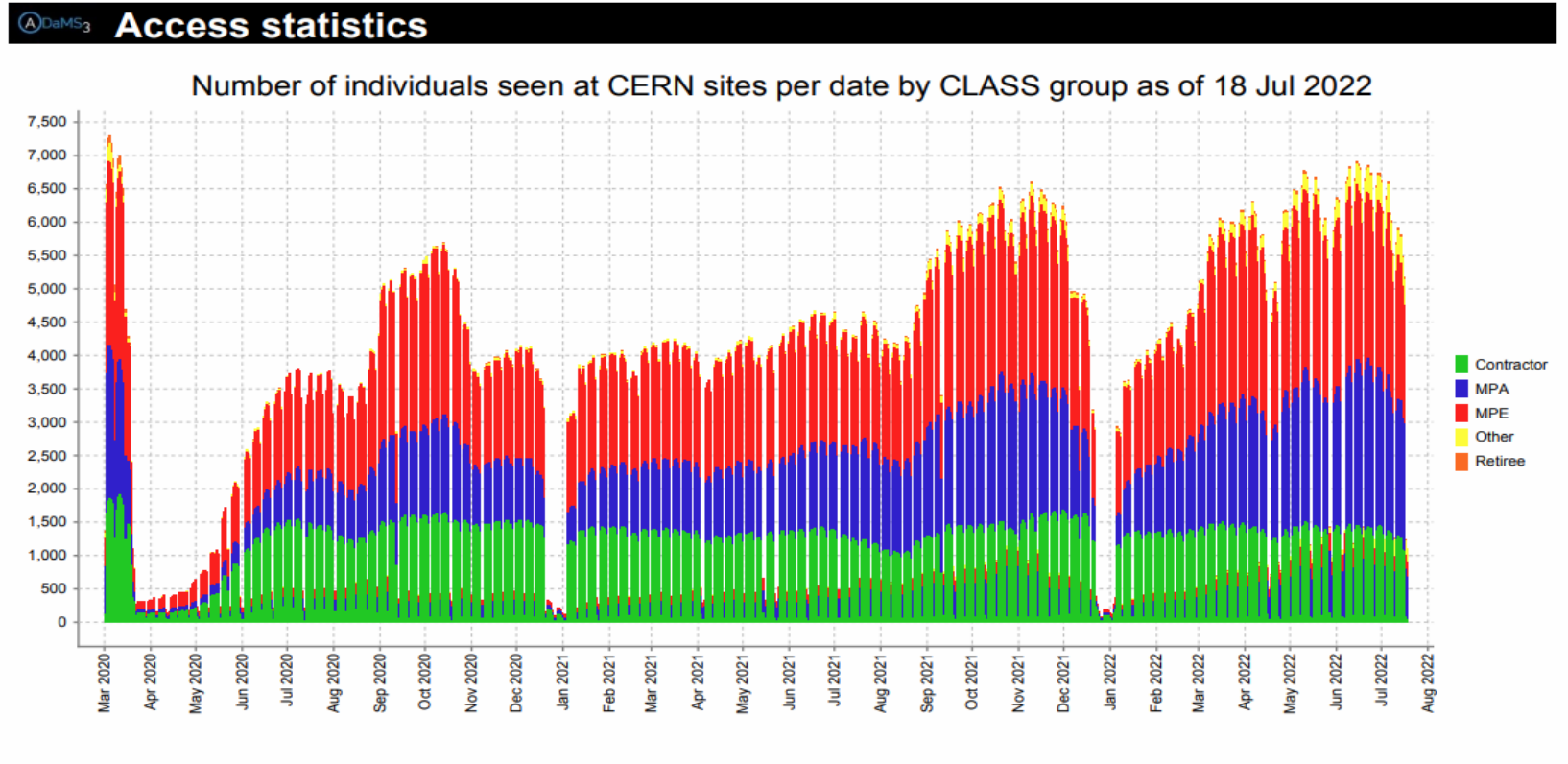
What is Pentaho?

Pentaho is a Business Intelligence (BI) software suite used in CERN for different purposes.

Community Edition vs Enterprise Edition

Examples of usage:

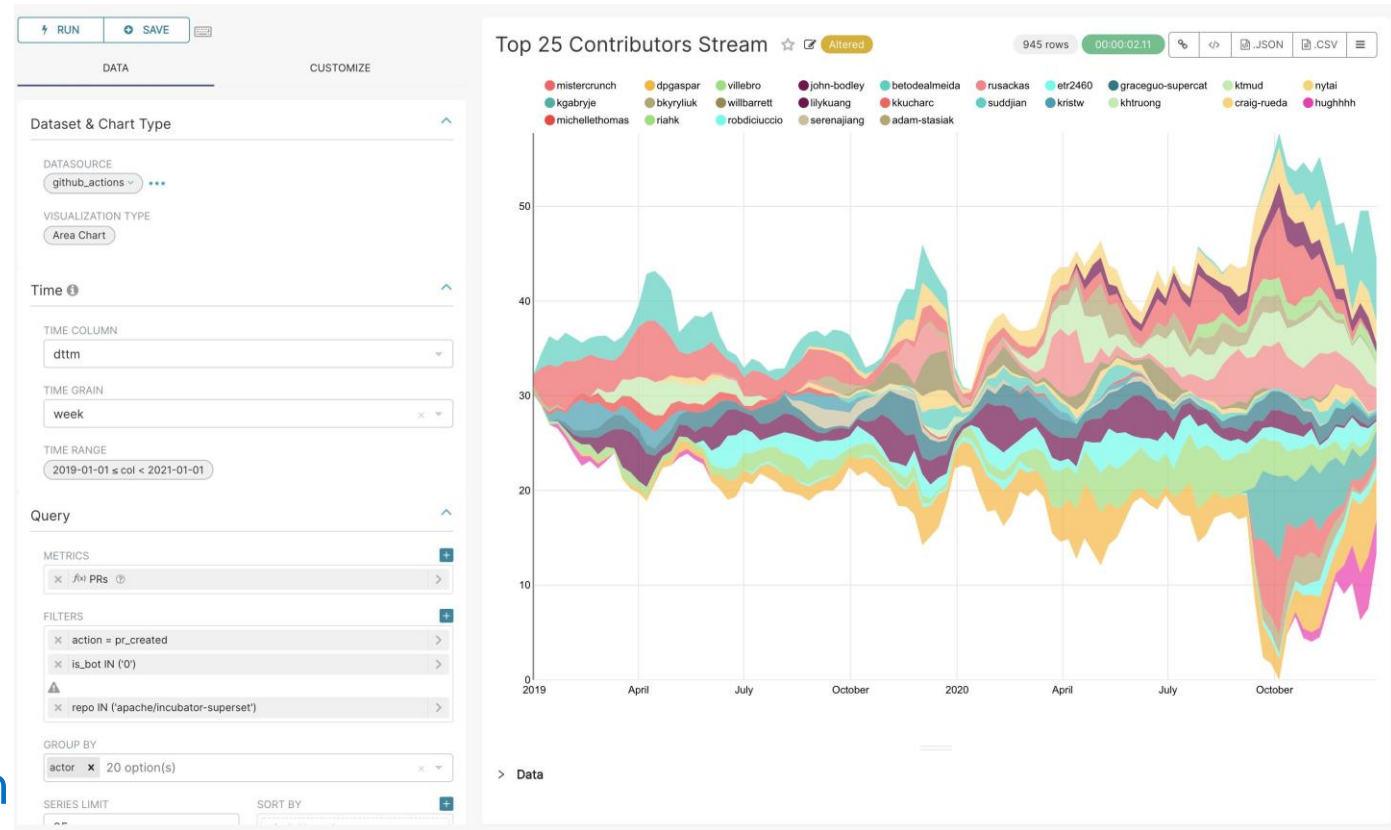
- Enterprise Asset Management reports
- Access Control reports
- CERN badge printing
- Invoices & Financial Reports



What does Apache Superset provide?

Modern, opensource, enterprise-ready, easy-to-use BI tool

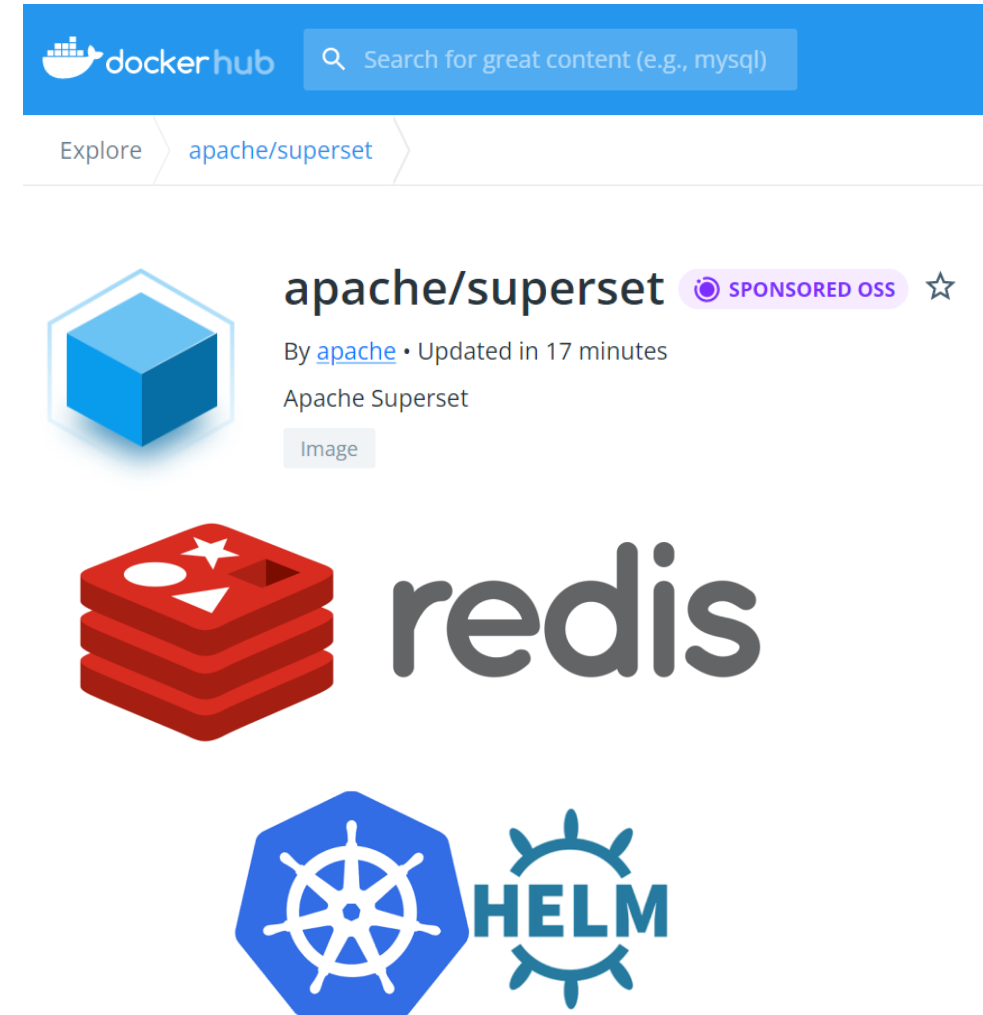
- Intuitive web interface
- Wide variety of data visualizations
- Charts and interactive dashboards
- Lightweight semantic layer
- No coding skills needed
- SQL editor and querying
- Database connectivity
- Cloud native architecture
- Highly scalable
- Integration with major authentication providers
- Data security and roles



Configuration & Deployment

Easy to configure

- Official Superset Docker image as a base image
- «**superset_config.py**» added to «**PYTHONPATH**»
 - web server
 - metadata database engine
 - message queue
 - results backend
 - caching layer
- Extended image with custom configuration
- Automation with GitLab CI
- Deployment in K8s cluster in a separate namespace
- Created Helm chart for easier deployment in K8s



Conclusion

Key notes on comparison:

- A rich set of data visualizations with a more user-friendly interface
- Integration with most of DBs
- Provides SQL IDE for querying data
- Capable of creating interactive reports and dashboards
- Based on roles and permissions, Superset provides access control to:
 - Datasources
 - Charts and dashboards
- Table view of data and sorting is possible with Pivot Tables
- Possible to add filters to charts and dashboards while viewing ...





QUESTIONS?

Thank you for your attention!