



Sample Project: Making the CERN notebook service a professional Big Data web analysis platform

Code	EP5822
Programme	FCT
Department	EP
Responsible	16919 - Dr. John Harvey
Created by	92892 - Dr. Danilo Piparo
Updated by	96245 - Mr. Vasco Miguel Chibante Barroso
Date Created	15-AUG-16
Date updated	17-AUG-16

Title

Making the CERN notebook service a professional Big Data web analysis platform

Description

SWAN (Service for Web based Data Analysis) is the jupyter notebook service of CERN. A software-as-a-service solution integrated with the major CERN IT services such as mass and synchronised storage, it allows scientists to perform the analysis of their data directly in the web browser, without the need of setting up software stacks or configuring data analysis frameworks. The technology chosen for the encapsulation of users is Docker containers which are scheduled on the SWAN backend, a distributed and elastic infrastructure hosted in the CERN computer centre.

SWAN is the cornerstone of the daily workflow of several scientists and is continuously evolving. This internship project aims to the improvement of the SWAN web platform, in particular of all the elements contributing to the user experience.

The trainee will be responsible for:

- The evolution of the distributed infrastructure acting as SWAN backend, for instance extending the computing capability of the SWAN service by interfacing it to external Big Data processing systems, such as Spark
- The integration with the CERN Docker Container Service
- Addition of new interfaces that are complementary to the notebooks, most notably remote desktop interfaces with a full windowsystem (e.g. Akhet).
- The synergy with code versioning services such as github and gitlab and their authentication mechanisms (e.g. OAuth).
- The amelioration and customisation of the Jupyter web interface via the integration of ipython extensions and widgets.
- The integration of SWAN with the most recent Jupyter releases.

The trainee will be given the opportunity to express her creativity in several ways, for example proposing strategies and features aiming to make SWAN a Big Data mining service even more competitive with modern industry standards.

The work will be carried out within the CERN Software group, EP-SFT, and in close collaboration with the CERN IT department, most notably the Storage group, IT-ST. The trainee will be involved in the everyday life of the Laboratory, contributing to group and project meetings.

Skills

Programming Languages: Javascript, Python, Shell Script <p>

- Docker

- Python

- Distributed systems and cluster management

- Jupyter, JavaScript and Apache Spark are not a requirement but certainly a plus</p>

Disciplines

Information Technologies

To edit this project go to https://hrapps.cern.ch/auth/f?p=131:4:::P4_ID:5822