This factsheet is a result of the second Wageningen University & Research (WUR) Data Science and AI Fellowship program. With this program we aim to increase and integrate our expertise in DS/AI throughout the entire organisation. The variety of projects highlights the potential for DS/AI across the WUR domains.

Author: Philippe Debie and Marjolein Verhulst

Bringing CERN technologies to WUR

deploying ScienceBox

Objective

The goal is to successfully implement ScienceBox at the Social Sciences Group (SSG), and extend it throughout the whole organization. ScienceBox is a containerized web service of the data science platform at the European Organization for Nuclear Research (CERN). It consists of a Jupiter Notebook-based web interface, the EOS bulk storage system, and a personalized storage environment named CERNBox (ScienceBox, 2022). Together, they contain all the main components to design and execute a data analysis, data science, or an artificial intelligence research objective.

Method

ScienceBox will be deployed at WUR with the help of FB-IT and CERN. It will be released to a limited set of peers at the Social Sciences Group. Feedback will be collected, systems will be optimized and an introductory workshop will be organized. After a succesful pilot, the number of concurrent users will be increased to accommodate more WUR researchers.

Results

The deployment of ScienceBox deemed more challenging than initially understood from the gathered information of CERN and the IT-department. The underlying technology of ScienceBox is ready and in full production at CERN, however the technical difficulties of migrating this service to a Kubernetes-cluster, and attachment to foreign storage and login systems are not thoroughly tested. ScienceBox is production ready as a standalone system, but not to be attached to the extensive and rich infrastructure that is offered by WUR. The current deployment allows users to program in the four provided programming languages, using all preinstalled packages, including CERN's data analysis framework ROOT (Antcheva et al., 2011). However, it is still a work in progress, a few key elements are still under construction.

Impact

This project had an impact on the following components:

1 The WUR IT-department had high interest in deploying this complex Kubernetes cluster. ScienceBox is one of the most complex deployments they have deployed so far, and they have learned a great deal of new skills in this field.
2 The ScienceBox development team has shown gratitude for the commitment to try and deploy ScienceBox at WUR. The feedback that we were able to give them will improve ScienceBox significantly.
3 It strengthened the collaboration between WUR and CERN, which reflects well on the image of WUR as an important partner in the Data Science community.
4 It helped to establish a collaboration in the form of an Innovation Center for Artificial Intelligence (ICAI) lab.

Future plans

The deployment of ScienceBox is still in progress. We are replicating the core mechanisms on the High-Performance Cluster of WUR. Next, a workshop will be organized for early adopters within SSG. Feedback will be collected and systems will be optimized. Last, ScienceBox will be made available for all interested WUR researchers.

Further information

This project is executed by Philippe Debie, Marjolein Verhulst, and Serdar Demirel. If you are interested in joining the ScienceBox pilot, contact philippe.debie@wur.nl.

Antcheva, I., Ballintijn, M., Bellenot, B., Biskup, M., Brun, R., Buncic, N., ... & Moneta, L. (2011). ROOT—A C++ framework for petabyte data storage, statistical analysis and visualization. Computer Physics Communications, 182(6), 1384-1385.

ScienceBox. (2022). ScienceBox: *Run CERN Services on Kubernetes*. Retrieved from https://sciencebox.web.cern.ch/.