



Contribution ID: 469

Type: Oral

Science Box: Converging to Kubernetes containers in production for on-premise and hybrid clouds for CERNBox, SWAN, and EOS

Tuesday, November 5, 2019 5:00 PM (15 minutes)

Container technologies are rapidly becoming the preferred way by developers and system administrators to package applications, distribute software and run services. A crucial role is played by container orchestration software such as Kubernetes, which is also the natural fit for microservice-based architectures. Complex services are re-thought as a collection of fundamental applications (each of these hosted in a separate container), while the final service is achieved by executing multiple containers at the same.

The Storage Group of the IT department at CERN has been successfully exploiting containers technology as a basis of Science Box, a self-contained Docker-based version of EOS (the CERN storage technology for LHC data and users' files), CERNBox (cloud synchronization and sharing for science), and SWAN (Service for Web-based ANalysis). Science Box has been successfully deployed on multiple cloud providers, including commercial platforms such as Amazon or Open Telekom Cloud.

In 2018, Science Box was at the core of a project investigating Big Data tools to analyze data from the TOTEM experiment at the LHC. In this context, a Kubernetes-managed instance of EOS, CERNBox, and SWAN has been deployed on the infrastructure provided by the Helix Nebula Science Cloud, an initiative targeting procurement of cloud resources from commercial providers and publicly funded science clouds. The infrastructure consisted of 400 CPUs, 1.5TB of memory, and 22TB of block storage. In addition, SWAN has been interfaced with a dedicated Spark cluster of ~2,000 cores to boost its computational capabilities. The deployment has been maintained for approximately 6 months during which it has been actively used by the TOTEM scientists to exploit a new interface for declarative analysis (called RDataFrame and now part of the ROOT analysis framework), which enables interactive processing of large datasets. The system provided validated physics results and achieved considerable speed-ups, effectively allowing the physicists to perform complex analysis tasks in quasi-interactive response times.

We are currently investigating the feasibility of running critical production storage services at CERN in containers. We leverage the experience gained with the development of Science Box and plan to evolve our service deployment model by combining Kubernetes to orchestrate containers execution and Helm to manager their configuration. In addition, we plan to use the cloud container orchestration service provided by the Computing and Monitoring group of CERN-IT, which employs OpenStack-provided resources and embeds centralized monitoring and auto-scaling capabilities.

Consider for promotion

No

Primary authors: BOCCHI, Enrico (CERN); GONZALEZ LABRADOR, Hugo (CERN); CASTRO, Diogo (CERN); MOSCICKI, Jakub (CERN); KOTHURI, Prasanth (CERN); CANALI, Luca (CERN); MROWCZYNSKI, Piotr (CERN); MALAWSKI, Maciej (AGH University of Science and Technology)

Presenter: BOCCHI, Enrico (CERN)

Session Classification: Track 7 – Facilities, Clouds and Containers

Track Classification: Track 7 – Facilities, Clouds and Containers