Contribution ID: **33**   Type: **not specified**

Distribution of container images: From tiny deployments to massive analysis on the grid

Wednesday 17 March 2021 08:50 (25 minutes)

In recent years, containers became the de-facto standard to package and distribute modern applications and their dependencies. A crucial role in the container ecosystem is played by container registries (specialized repositories meant to store and distribute container images) which have seen an ever-increasing need for additional storage and network capacity to withstand the demand from users. The HEP community also demonstrates an increasing interest, with scientists encapsulating their analysis workflow and code inside a container image. The analysis is first validated on a small dataset and minimal hardware resources to then run at scale on the massive computing capacity provided by the grid.

CERN IT offers a centralized GitLab Container Registry based on S3 storage. This registry is tightly integrated with code repositories hosted on CERN GitLab and allows for building and publishing images via CI/CD pipelines. Plans are to complement the GitLab Registry with Harbor, the Open Cloud Initiative container registry, which provides advanced capabilities including security scans of uploaded images, non-blocking garbage collection of unreferenced blobs, and proxying/replication from/to other registries.

In this context of HEP, the CernVM File System (CVMFS) has recently introduced the support for ingestion and distribution of container images. It implements file-level deduplication and an optimized distribution and caching mechanism that overcome the limitations of the push-pull model used by traditional registries, ultimately making the distribution of containers more efficient across the WLCG resources. A prototype integration between Harbor and CVMFS has been developed to provide the end-users with a unified management portal for their container images while supporting the large-scale analysis scenarios typical of the HEP world.

Speaker release

Yes

Desired slot length

20

Author: BOCCHI, Enrico (CERN)

Presenter: BOCCHI, Enrico (CERN)

Session Classification: Storage & File Systems

Track Classification: Storage & Filesystems