

Towards a Serverless CernVM-FS

Thursday, 12 July 2018 17:20 (20 minutes)

The CernVM File System (CernVM-FS) provides a scalable and reliable software distribution and—to some extent—a data distribution service. It gives POSIX access to more than half a billion binary files of experiment application software stacks and operating system containers to end user devices, grids, clouds, and supercomputers. Increasingly, CernVM-FS also provides access to certain classes of data, such as detector conditions data, genomics reference sets, or gravitational wave detector experiment data. For most of the high-energy physics experiments, an underlying HTTP content distribution infrastructure is jointly provided by universities and research institutes around the world.

In this contribution, we will present recent developments, such as the integration into HPC environments and into XRootD networks, as well as future plans. For future developments, we put a focus on evolving the content distribution infrastructure and at lowering the barrier for publishing into CernVM-FS. Through so-called serverless computing, we envision cloud hosted CernVM-FS repositories without the need to operate dedicated servers or virtual machines. An S3 compatible service in conjunction with a content delivery network takes on data provisioning, replication, and caching. A chain of time-limited and resource-limited functions (so called “lambda function” or “function-as-a-service”) operate on the repository and stage the updates. As a result, any CernVM-FS client should be able to turn into a writer, possession of suitable keys provided. For repository owners, we aim at providing cost transparency and seamless scalability from very small to very large CernVM-FS installations.

Primary authors: BLOMER, Jakob (CERN); GANIS, Gerardo (CERN); POPESCU, Radu (CERN)

Presenter: BLOMER, Jakob (CERN)

Session Classification: Plenary

Track Classification: Track 7 –Clouds, virtualization and containers