Contribution ID: **185**   Type: **Presentation**

Scalable Metadata Management Using Onedata and OpenFaaS

Wednesday 27 January 2021 14:10 (10 minutes)

Onedata [1] is a global high-performance, transparent data management system, that unifies data access across globally distributed infrastructures and multiple types of underlying storages, such as NFS, Amazon S3, Ceph, OpenStack Swift, WebDAV, XRootD and HTTP and HTTPS servers, as well as other POSIX-compliant file systems.
Onedata allows users to collaborate, share, and perform computations on data using applications relying on POSIX compliant data access. Thanks to a fully distributed architecture, Onedata allows for the creation of complex hybrid-cloud infrastructure deployments, including private and commercial cloud resources.

Onedata comprises the following services: Onezone - authorisation and distributed metadata management component that provides access to Onedata ecosystem; and Oneprovider - provides actual data to the users and exposes storage systems to Onedata and Oneclient - which allows transparent POSIX-compatible data access on user nodes. Oneprovider instances can be deployed, as a single node or an HPC cluster, on top of high-performance parallel storage solutions with the ability to serve petabytes of data with GB/s throughput.

Onedata introduces the concept of Space, a virtual volume, owned by one or more users, where they can organize their data under a global namespace. The Spaces are accessible to users via a web interface, which allows for Dropbox-like file management, a Fuse-based client that can be mounted as a virtual POSIX file system, a Python library (OnedataFS [2]), or REST and CDMI standardized APIs. As a distributed system Onedata can take advantage of modern scalable solutions like Kubernetes and thanks to a rich set of REST APIs and OnedataFS library it can process at scale data and metadata alike using FaaS systems like OpenFass.

Currently Onedata is used in European Open Science Cloud Hub [2], PRACE-5IP [3], EOSC Synergy [4], and Archiver [5] project, where it provides data transparency layer for computation deployed on hybrid clouds.

[1] Onedata project website. http://onedata.org.
[2] OnedataFS - PyFilesystem Interface to Onedata Virtual File System. https://github.com/onedata/fs-onedatafs.
[3] European Open Science Cloud Hub (Bringing together multiple service providers to create a single contact point for European researchers and innovators.). https://www.eosc-hub.eu.
[4] Partnership for Advanced Computing in Europe - Fifth Implementation Phase. http://www.prace-ri.eu.
[5] European Open Science Cloud - Expanding Capacities by building Capabilities. https://www.eosc-synergy.eu.
[6] Archiver - Archiving and Preservation for Research Environments). https://www.archiver-project.eu.

Authors: DUTKA, Lukasz; ORZECHOWSKI, Michał (AGH University of Science and Technology, Academic Computer Centre Cyfronet AGH, Krakow, Poland); KRYZA, Bartosz (ACC Cyfronet-AGH)

Presenter: DUTKA, Lukasz

Session Classification: Scalable Storage

Track Classification: Main session: Scalable Storage Backends for Cloud, HPC and Global Science