Contribution ID: **66** Type: **not specified**

Onedata - Eventually Consistent Virtual Filesystem for Multi-Cloud Infrastructures

Tuesday 31 January 2017 08:30 (20 minutes)

Onedata - Eventually Consistent Virtual Filesystem for Multi-Cloud Infrastructures

Onedata [[1]] is a global high-performance data management system, that provides easy and unified access to globally distributed storage resources and supports wide range of use cases from personal data management to data-intensive scientific computations. Due to its fully distributed architecture, Onedata enables creation of complex hybrid-cloud infrastructure deployments, including private and commercial cloud resources. It allow susers to share, collaborate and publish data as well as perform high performance computations on distributed data.

Onedata system comprises zones (Onezone) which enable establishment of federations of data centers and users, storage providers (Oneprovider) who expose storage resources and clients (Oneclient), who can access their data via a virtual POSIX file system. Onedata manages all operations on files at the level of variable sized blocks,ensuring highly efficient data access to files available remotely and giving the users an eventually consistent view of the filesystem from anywhere. In order to efficiently propagate local changes to other storage providers, who support specific user spaces, we employ tree propagation algorithm, which means that each storage provider sends out the local modifications events only to a subset of all providers who can be affected by this change. Onedata introduces the concept of space, a virtual volume, owned by one or more users, where the data is stored. Each space can be supported by a dedicated amount of storage supplied by one or multiple storage providers.Storage providers deploy Oneprovider instance near the storage resources, register it in selected Onezone service to become part of a federation and expose those resources to users. By supporting multiple types of storage backends, such as such as POSIX, S3, Ceph and OpenStack Swift,Onedata can serve as a unified virtual filesystem for multi-cloud environments.

For flexible collaboration and data sharing, Onedata provides fine-grained management of access rights, including POSIX-like access permissions and access control lists (ACLs), that allow users to share entire spaces,directories or files with individual users or user groups. Onedata allows integration with several identity providers,by means of OpenID Connect protocol, enabling users to login using their existing accounts, while all authorization decisions within Onedata are based on bearer tokens (Macaroons) generated by Onezone service.

Currently Onedata is used in INDIGO-DataCloud [[2]] project asa federated data access solution, aggregating computing centres and infrastructures; and in EGI-Engage [[3]], as the basis of EGI Open Data Platform, support-ing various open science use cases such as open data curation(metadata editing), publishing (DOI registration)and discovery (OAI-PMH protocol).

Acknowledgements: This work has been partially funded under Horizon 2020 EU projects: INDIGO-DataCloud(Project ID: 653549) and EGI-Engage (Project ID: 654142).

REFERENCES

1. Onedata project website.
2. INDIGO-DataCloud (Integrating Distributed data Infrastructures for Global Exploitation).
3. EGI-Engage (Engaging the Research Community towards anOpen Science Commons).

Corresponding author: Łukasz Dutka (lukasz.dutka@cyfronet.pl)

Primary author: Dr DUTKA, Łukasz (AGH University of Science and Technology, Academic Computer Centre Cyfronet AGH, Krakow, Poland)

Co-author: Dr KRYZA, Bartosz (AGH University of Science and Technology, Academic Computer Centre Cyfronet AGH, Krakow)

Session Classification: Storage technologies