



Contribution ID: 34

Type: **Lightning Talk 8 mins**

Developing a distributed and fault tolerant Dataverse architecture.

Thursday, 22 September 2022 11:15 (8 minutes)

Dataverse is an open source data repository solution with increased adoption by research organizations and user communities for data sharing and preservation. Datasets stored in Dataverse are catalogued, described with metadata, and can be easily shared and downloaded. However, despite all its features, Dataverse is still missing an architecture that ensures a distributed, fault tolerant, highly available and out-of-the-box service deployment.

In this presentation we will report the efforts by the Portuguese Distributed Computing Infrastructure (INCD), to address these current limitations by creating a dataverse deployment architecture that is easy to set-up, portable, highly available and fault tolerant.

We tackled this objective, following a DevOps approach, resorting to a wide range of open software tools such as Linux containers, source code repositories, CI/CD pipelines, keepalived in conjunction with Virtual IPs (VIPs), pg_auto_failover for database replication and high availability object storage as scalable data storage backend. The solution is implemented on top of the Openstack cloud management framework, while the authentication is performed through the egi-checkin.

This architecture, is therefore capable of providing a stable and fault tolerant Dataverse installation, while keeping a flexible enough set-up to allow for the expansion of the storage and facilitate the upgrade to new versions.

The deployment architecture is currently under testing and will be used to support a catchall data repository for the Portuguese research and academic community. Furthermore, we expect that this solution can be deployable in EGI fedcloud resources to support FAIR data both for thematic services and generic use.

Any relevant links

Topic

Data Spaces

Primary author: BENTA, Zacarias

Co-authors: GOMES, Jorge (LIP); DAVID, Mário (LIP); FERREIRA, César (LIP); PINA, João (LIP)

Presenter: BENTA, Zacarias

Session Classification: Lightning talks: Data Spaces & Data Lakes

Track Classification: Data Spaces