



Título da proposta para apresentação *Flash Talk*

Zacarias Benta

LIP, zacarias@lip.pt

Resumo

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 cataloged, 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. Our solution was implemented on top of the Openstack cloud management framework, the authentication is performed by using single sign on provided by severall IdPs. 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..

Palavras-chave: Dataverse, distributed, fault tolerant

1. Designação do projeto/infraestrutura/iniciativa

Dataverse FCCN..

2. Público-alvo

Data centers, gestores de ciência, curadores de dados, especialistas de informática.

3. Ligações web úteis

<https://gitlab.com/lip-computing/dataverse/dataverse-deployment>