LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Developing a distributed and fault tolerant Dataverse architecture.

National Distributed Computing Infrastructure

Services: scientific computing, data processing and data storage

Target: scientific and academic community, infrastructures, R&I projects, SMEs

Promote: shared resources, advanced computing and data services for research

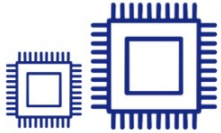
Interface: international digital infrastructures (EGI, IBERGRID, WLCG, EOSC)

Cloud Computing

cloud computing

HTC Computing

high throughput computing (GRID)



HPC Computing

high performance computing

More than 20 years of experience

2000

LIP participation in R&D projects focused on grid computing.

2006

The national grid initiative is established by the ministry.
IBERGRID joins PT and ES centres.
WLCG MoU is signed.

2008

Portuguese main node for grid computing, joining LIP, LNEC and FCCN.
Collaboration with U.Porto, U.Minho, U.Aveiro etc

2010

Construction of a datacenter for the main node for grid computing at LNEC.
Portuguese WLCG Tier-2 for ATLAS and CMS is established.

2013

Portugal joins EGI.
IBERGRID activities are boosted.
PT LHC Tier-2 starts.

Creation of INCD is proposed to the national roadmap of research infrastructures (RNIE).

2014

The proposal for the creation of INCD passes the scientific evaluation.

2015

The legal entity “Associação INCD” is established.

2016

The 1st project to fund INCD is submitted under the infrastructures roadmap.

2017

The project to establish INCD starts. Establishing the administrative and organizational basis for INCD.

2018

First INCD collaborators are contracted.
First international public tender for compute and storage.

2021

Infrastructure project is extended until October 2022.

Our challenges.

- Dataverse Service for National Scientific Computation Foundation;
- Fault tolerant service;
- No single point of failure;
- High availabilty.

Our issues.

- No recipe to elaborate on;
- Package installations not flexible enough;
- Expandable storage – local/S3/swift/minio;
- Distributed database;
- SOLR index – standalone vs distributed;
- Server role shifting.

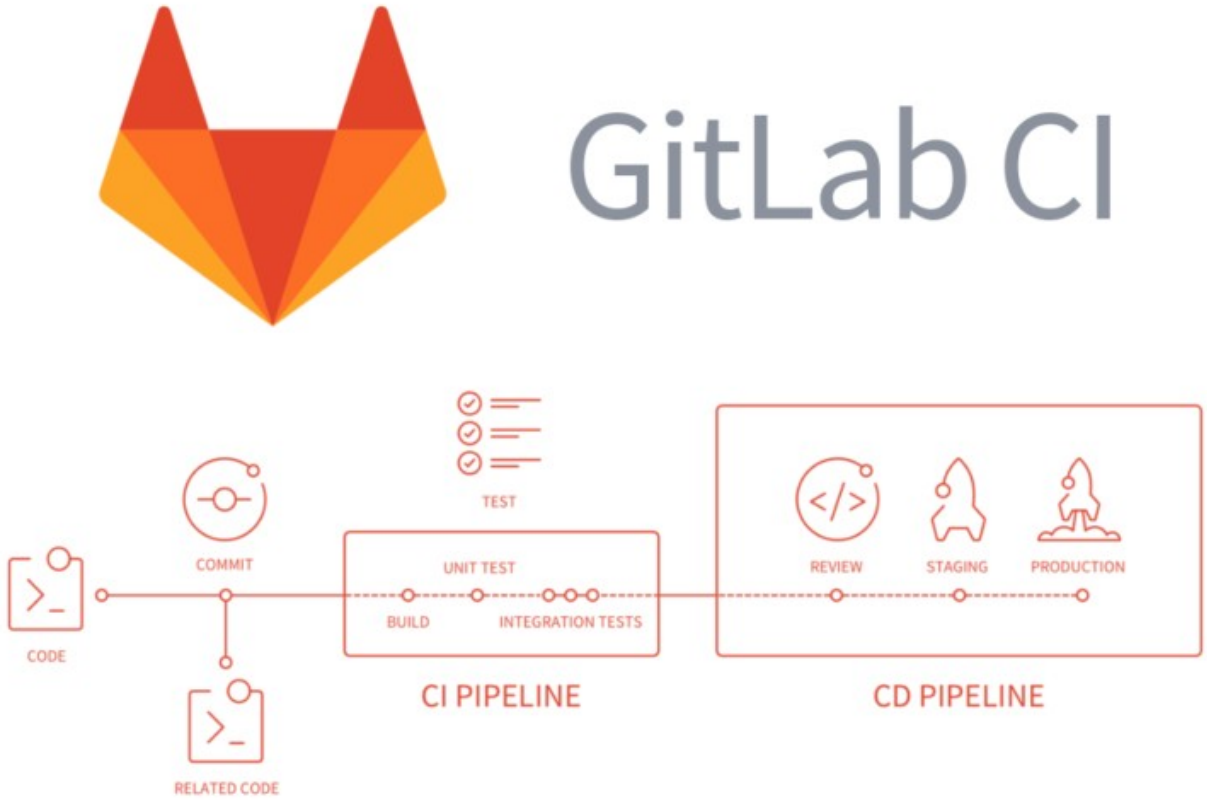
Our drawbacks.

- Trial and error advanced installation;
- Existing docker recipe not what we intended;
- Storage direct upload/download issues with S3/swift;
- Postgres documentation lacking solutions;
- Distributed SOLR index was cumbersome;
- Server role shifting demands intervention.

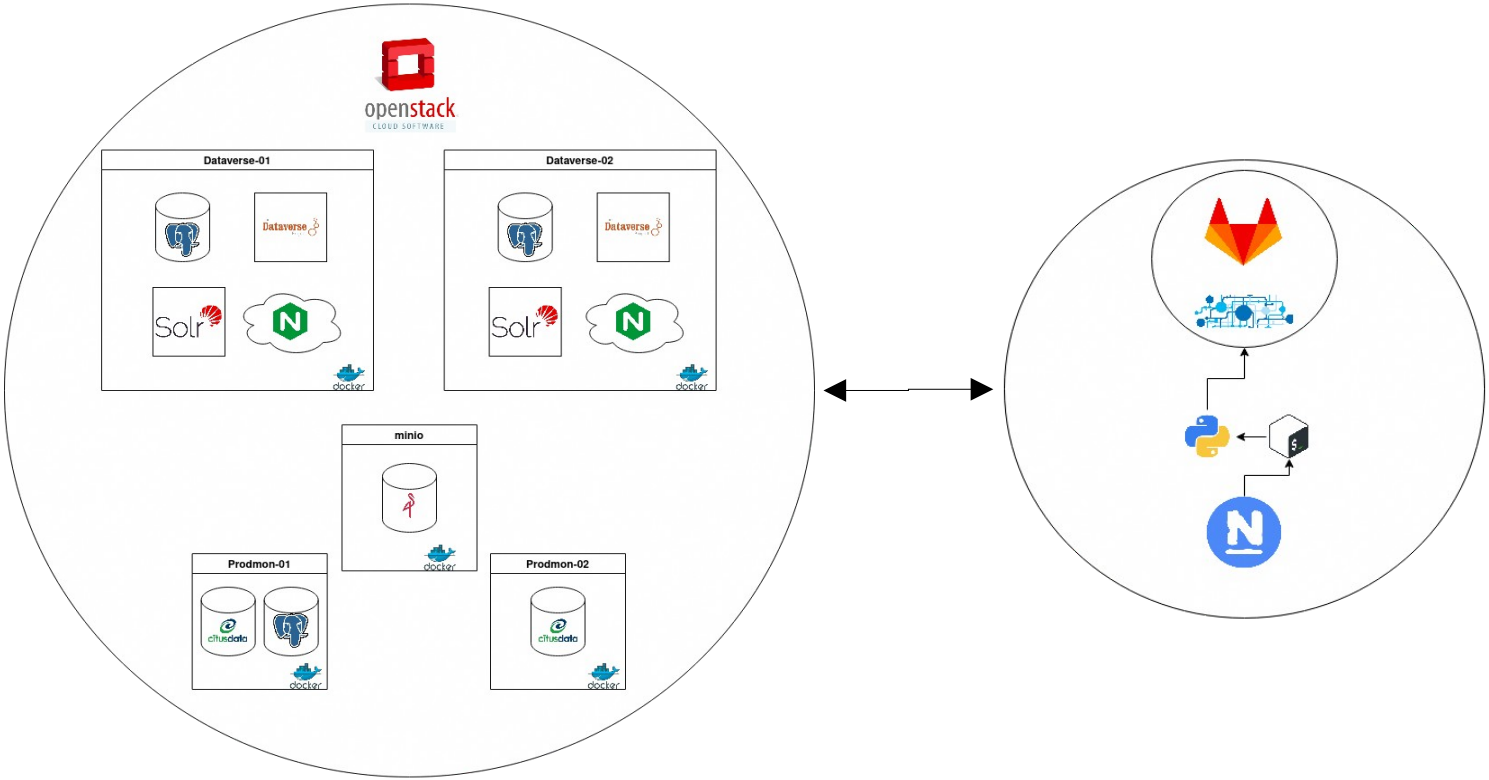
Our solutions.

- Created our own recipe / pipeline;
- Docker containers;
- Minio;
- pg_auto_failover;
- Standalone SOLR
- Keepalived/VIP.

The gitops approach

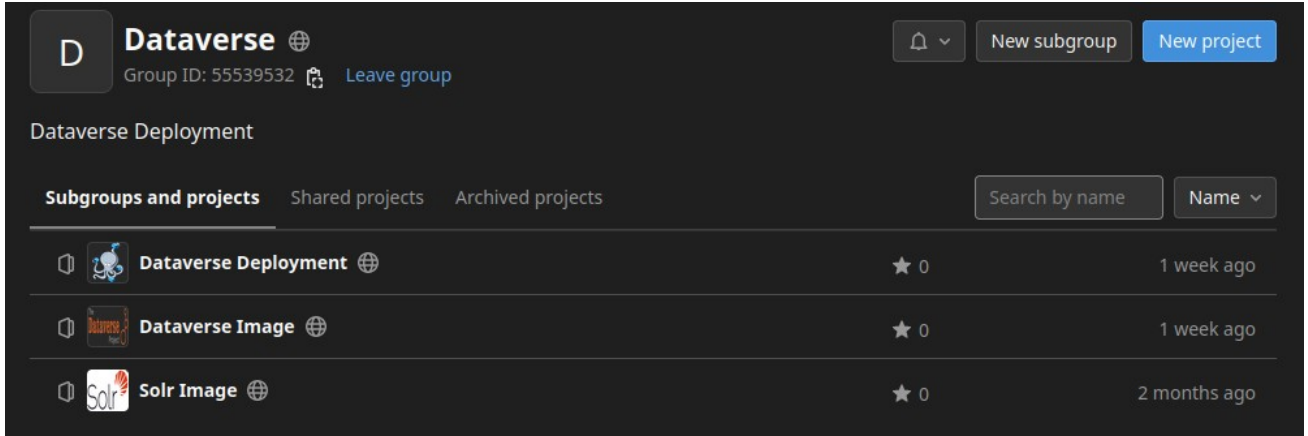


Dataverse Distributed Architecture



Our one stop shop solution

https://gitlab.com/lip-computing/dataverse



Thanks

Any questions?

You can find me at zacarias@lip.pt