IBERGRID 2026

Some changes to Dataverse & upgrading the HA architecture

Stats · Integrations · High-availability redesign



OVERVIEW

Presentation structure

01

Some stats

02

Some integrations

03

Old architecture

04

Issues with old architecture

05

New architecture

06

HA inner workings

01

Stats

How the repository has grown

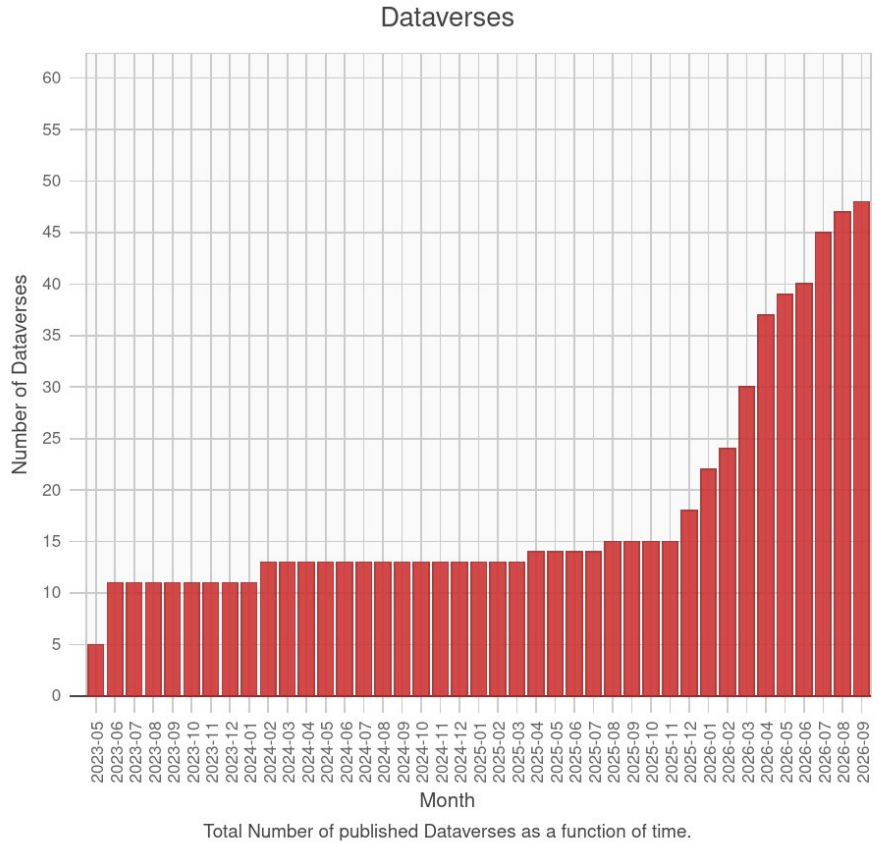


STATS

Number of dataverses

Dataverses over time

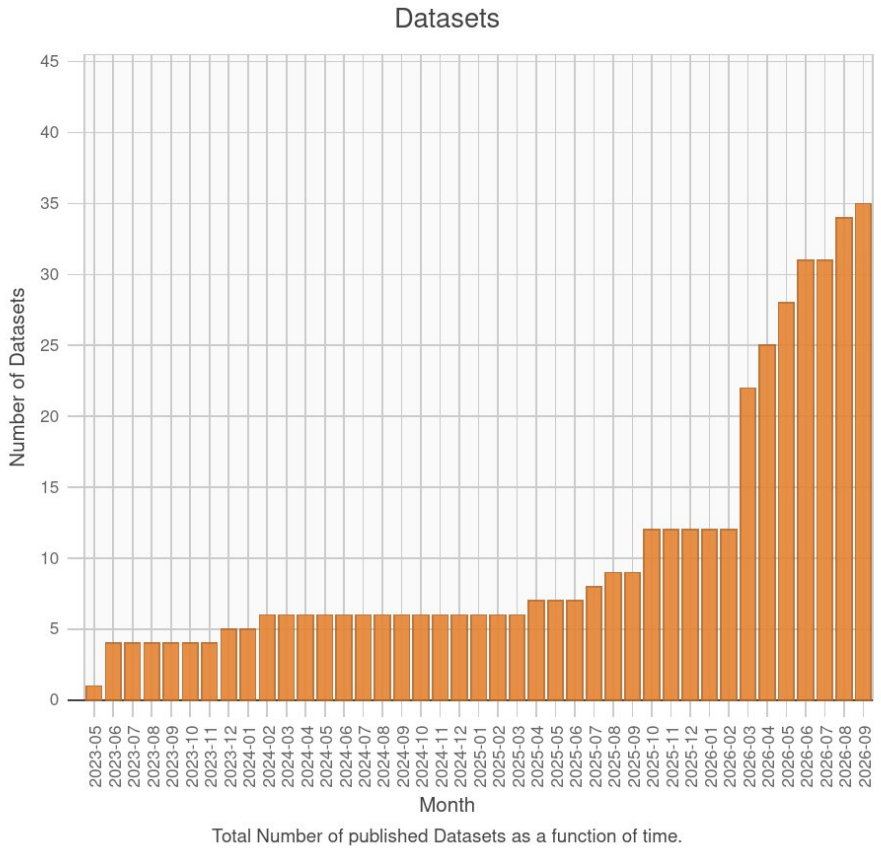
Total number of published dataverses, month by month.



Number of datasets

Datasets over time

Total number of published datasets, month by month.



Total Number of published Datasets as a function of time.

STATS

Dataverses & datasets by subject

Social Sciences make up the largest share of the collection.



Area of each tile is proportional to its share of content.

STATS

Users and data

181

Users

5

Superusers

170

Datasets

78

Dataverses

People · **Content** — figures from the Dataverse dashboard

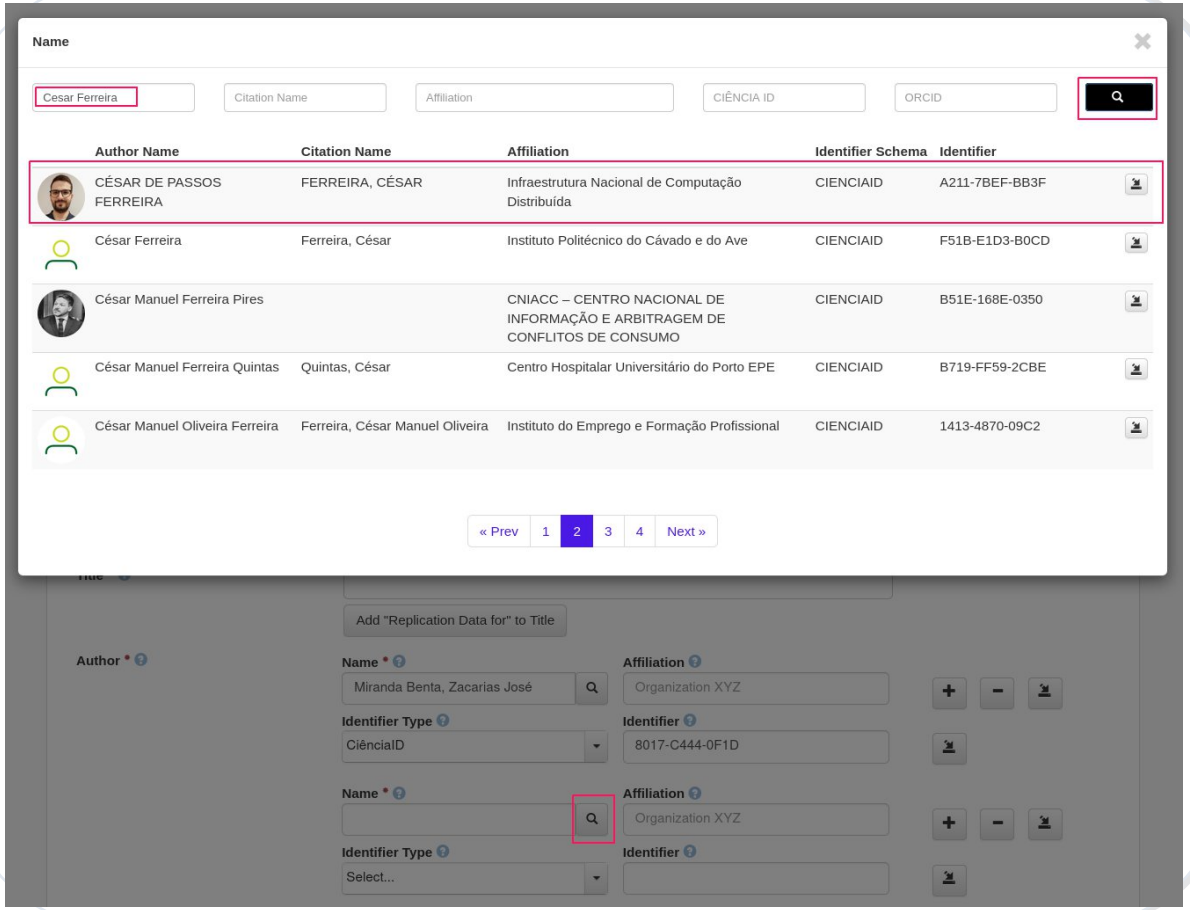
02

Integrations

New features in the deposit workflow

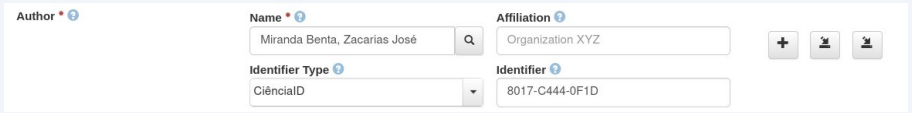


Author lookup with CiênciaID



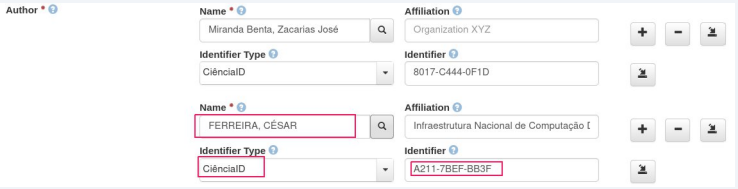
1 Search

Look up an author by name from the deposit form.

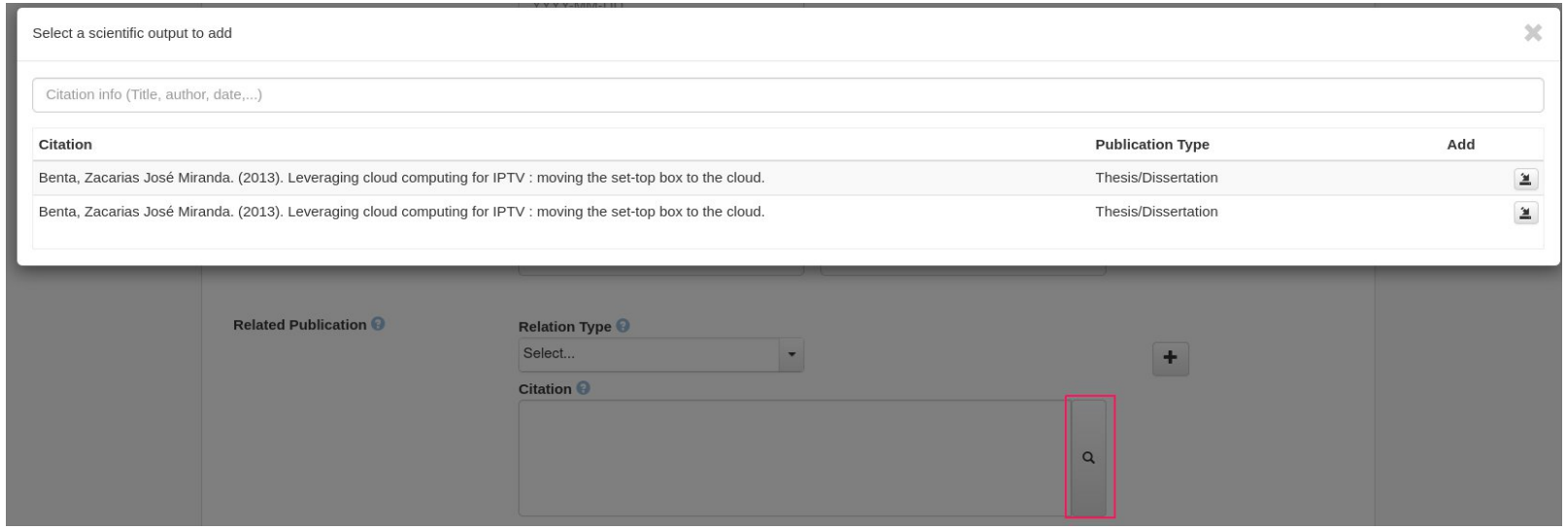


2 Auto-fill

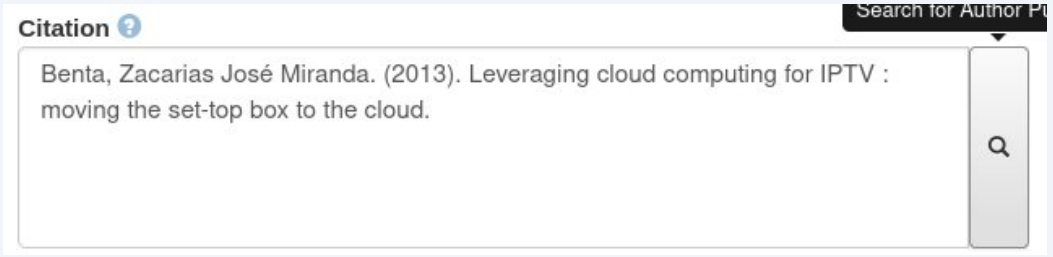
Name, affiliation and identifier are filled in for you.



Related publications lookup

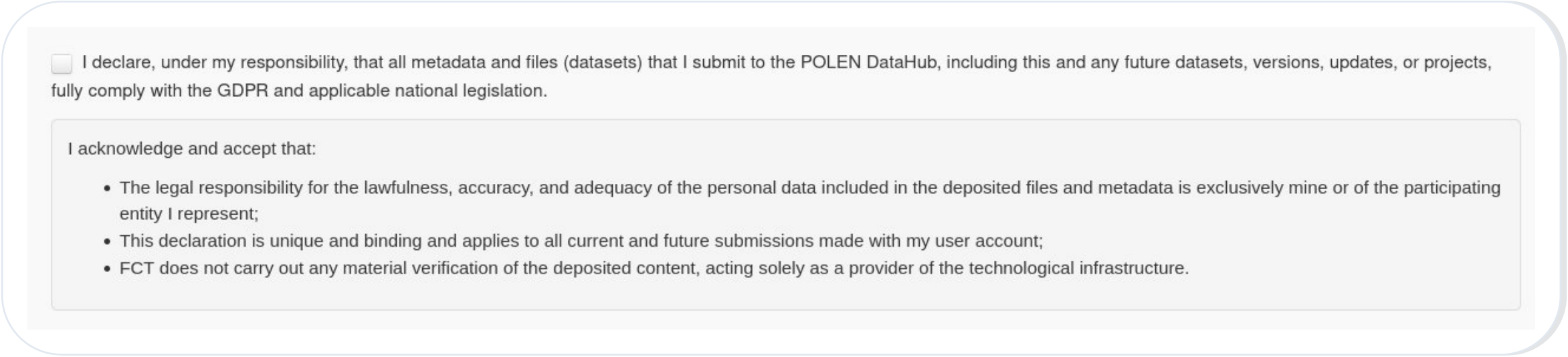


Pick a scientific output
and the citation is filled in automatically.



INTEGRATIONS

GDPR declaration on deposit



Depositor is responsible

Lawfulness and accuracy of personal data lie with the depositor or their entity.

One binding declaration

Applies to all current and future submissions from the account.

Infrastructure only

FCT does not verify deposited content; it provides the platform.

03

Architecture

From manual failover to self-healing HA

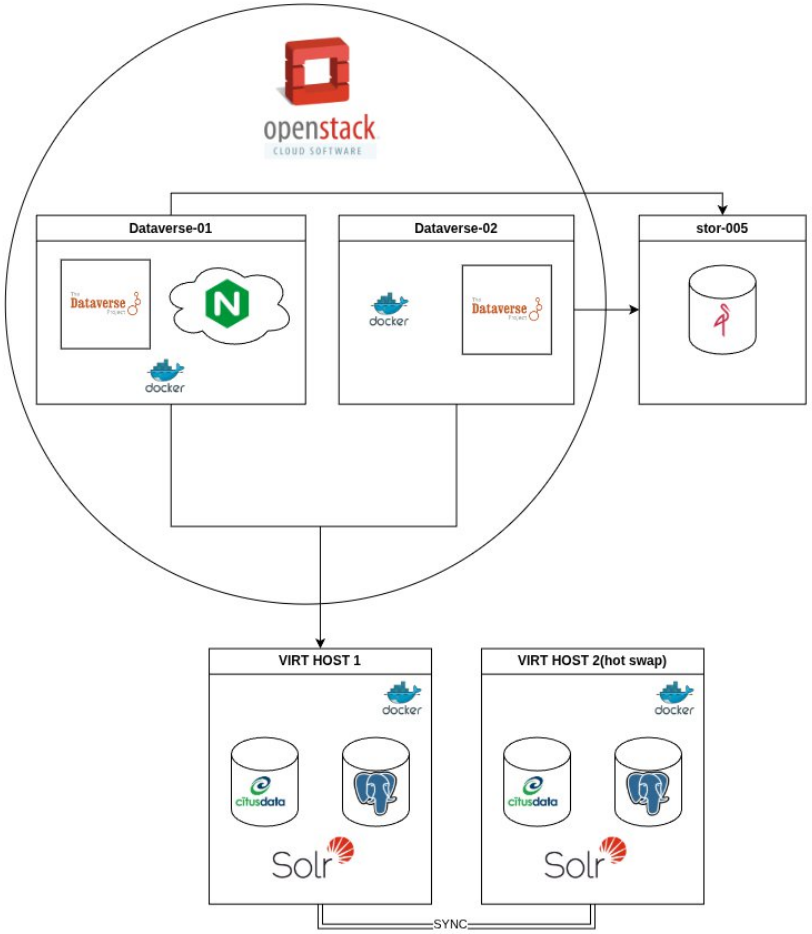


Old architecture

Where we started

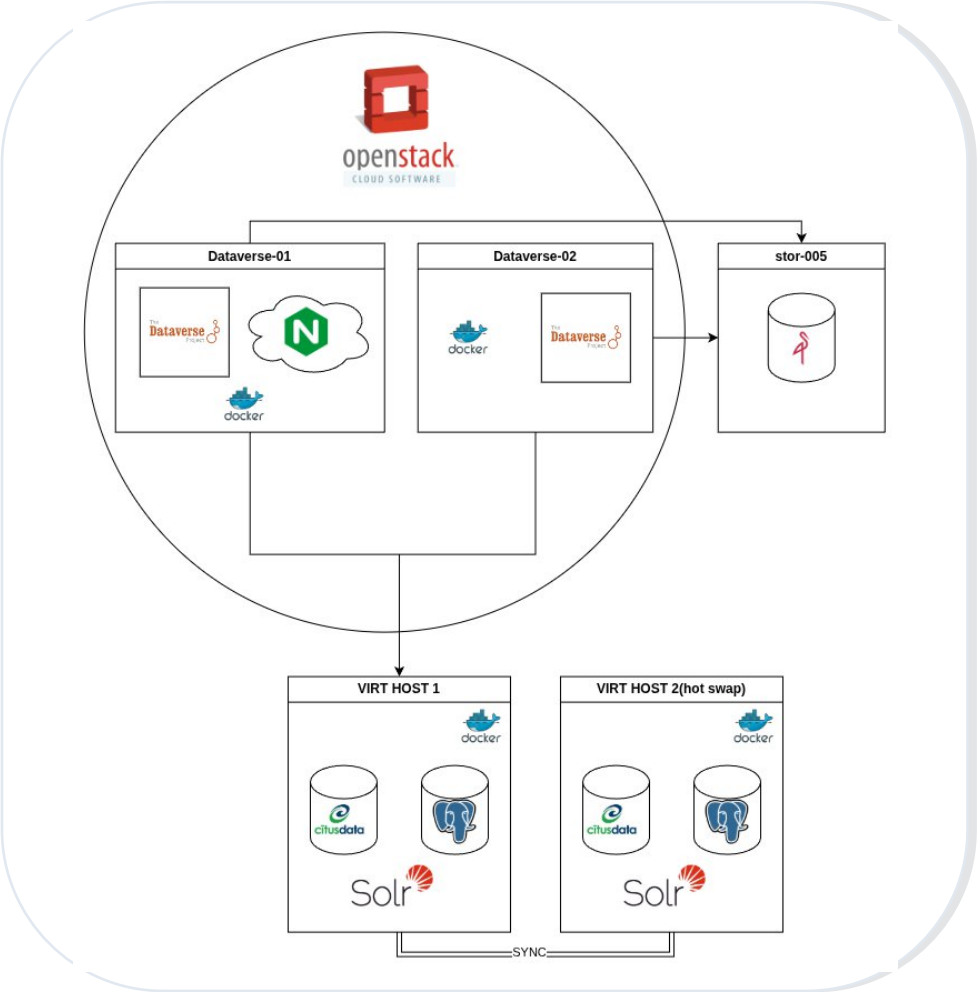
Dataverse services ran on OpenStack, with the database and SOLR on separate hosts — and failover handled by hand.

Its limitations are on the next slide.

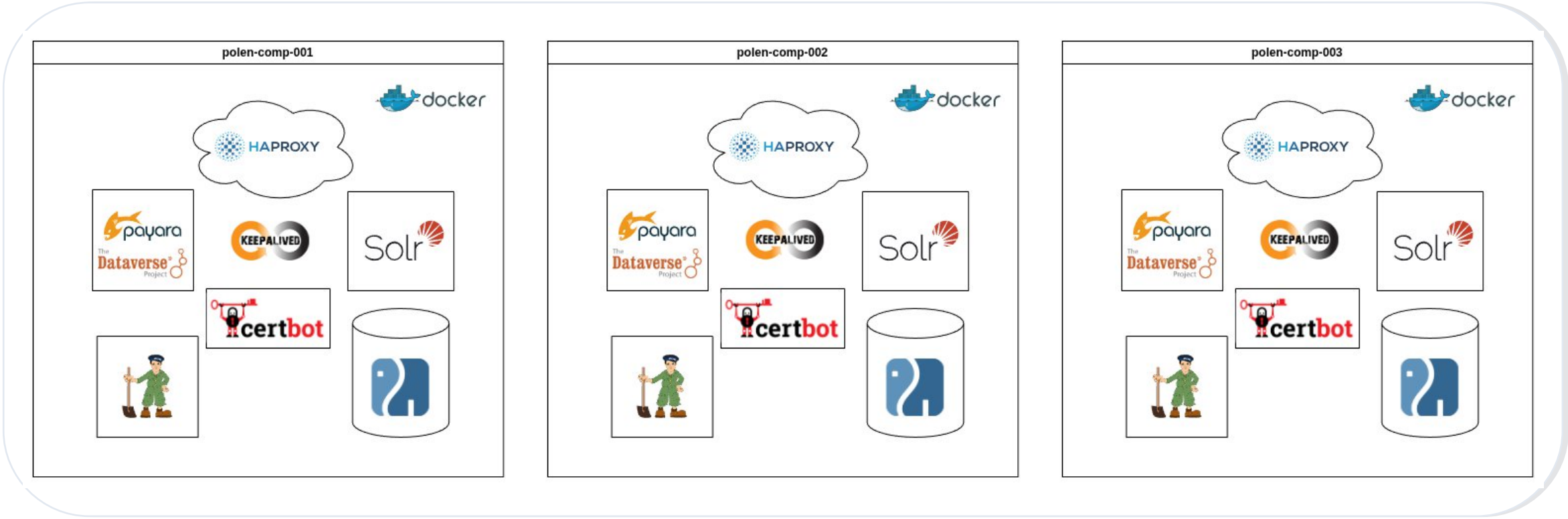


Old architecture issues

- No self-healing high availability
- Manual switchover on database failures
- Manual switchover on SOLR failures
- No keepalived for the public IP
- No OpenStack redundancy



New architecture



3 identical nodes

Containerised services

No single point of failure

ARCHITECTURE

HA inner workings

01 Same FQDN

All hosts respond to the same name

02 DNS round robin

Our DNS server spreads requests across hosts

03 Sticky sessions

HAProxy uses JSESSIONID to pin connections to a backend

04 SOLR Cloud

Indexes are synced between hosts

05 Patroni

Keeps PostgreSQL in sync across hosts

06 Zookeeper

Orchestrates both SOLR and PostgreSQL



Result: **there is no single point of failure**

Thank you!

Questions?

Have more questions after the session? Get in touch:

Cesar Ferreira
cesar@lip.pt

Zacarias Benta
zacarias@lip.pt

