INCD's new cloud infrastructures

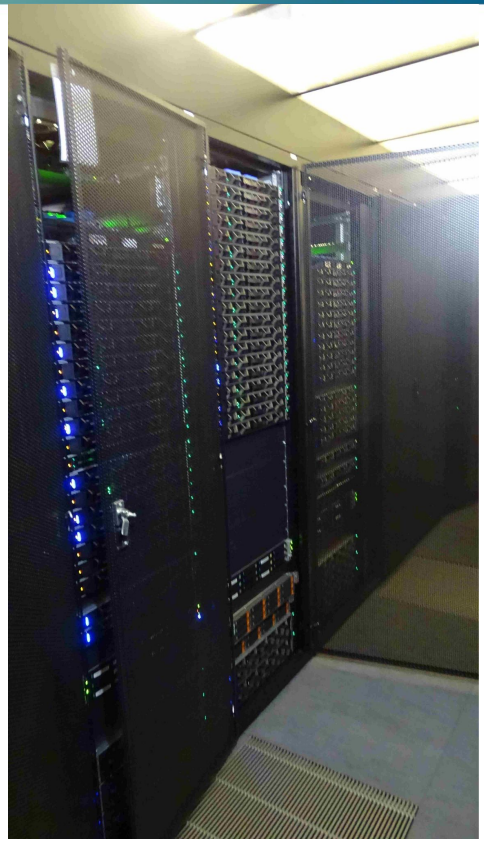
Mario David (david@lip.pt)

On behalf of LIP's Distributed Computing Group

Overview

Cloud infrastructure service:

- ➔ **INCD Lisbon: Stratus-A**
- ➔ **INCD UTAD: Stratus-D**



INCD new data centre: UTAD (Univ Trás os Montes e Alto Douro) in Vila Real:

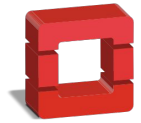
- https://www.incd.pt/?p=noticias/detalhes&id=41&lang=en

Provides:

- **Cloud computing**
- HPC
- **Data and storage services**

Cloud computing infrastructure I

- Based on Openstack Yoga:
- Underlying OS: Ubuntu 22.04 LTS:
  - Full support for 5 years.
  - “Dist-upgrade” between LTS versions (no reinstallation).

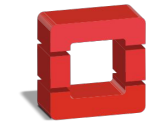


openstack
CLOUD SOFTWARE

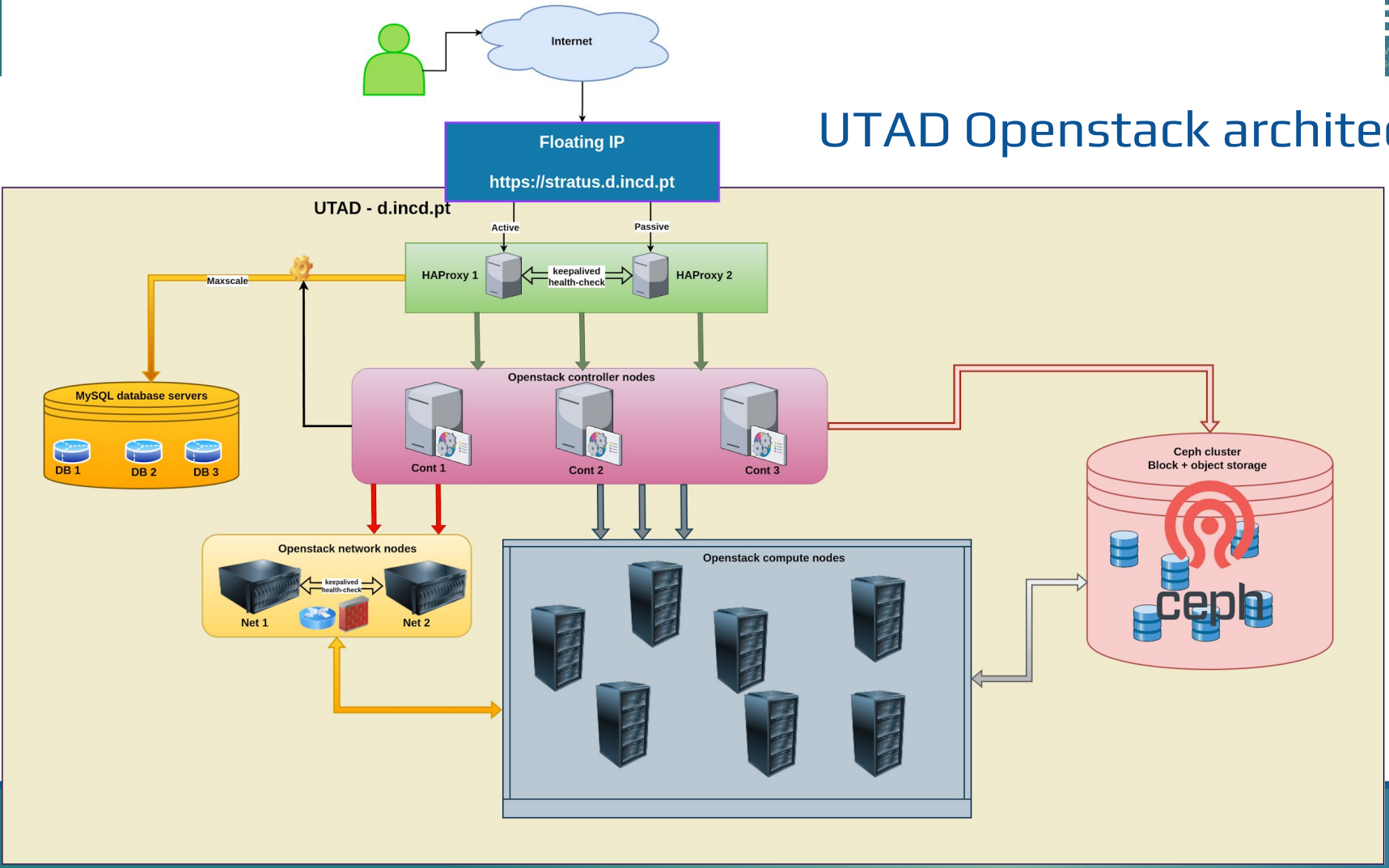
GitLab

Cloud computing infrastructure II

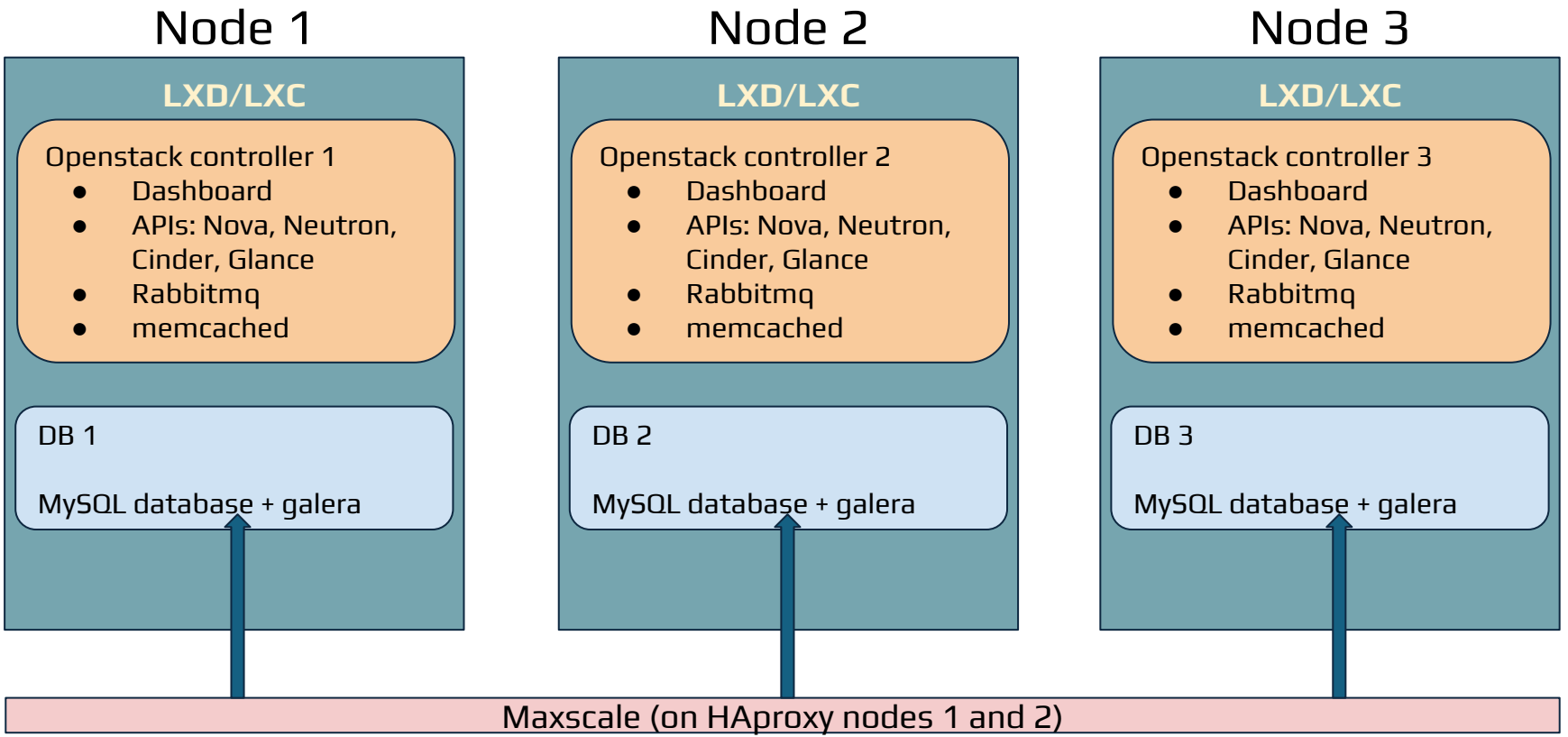
- Hardware:
  - 3 hosts with LXD/LXC - Openstack controllers and DBs.
  - 2 hosts for the Neutron agents.
  - 10 Nova compute nodes:
    - 1920 VCPUs - 2 x AMD EPYC 7643 48-Core Processor (192 VCPUs each).
    - Memory: 512 GB.
    - Local disks:
      - NVME for Operating System.
      - SATA3 7.3TB disk.
- Configuration management: Custom ansible playbooks
- VCS using gitlab private repository



UTAD Openstack architecture



UTAD Openstack architecture: Details I



UTAD Openstack architecture: Details II

Network node 1

Neutron agents:

- Openvswitch
- DHCP
- L3
- Metadata

L3 HA - High Availability for provider networks

Network node 2

Neutron agents:

- Openvswitch
- DHCP
- L3
- Metadata

Nova compute nodes:

- nova-compute
- libvirtd
- neutron-openvswitch
- **neutron-sriov**

Direct access to eth interface

High Availability (HA) and Load Balancing (LB)

- HA + LB with:
  - haproxy + keepalived: Openstack Dashboard (Horizon) and APIs.
  - maxscale + keepalived: MySQL mariadb with galera cluster.
  - Neutron L3 agent with HA: for provider networks.
- Tests:
  - **Shutdown 1 Neutron node**: the subnetwork has transitioned to the other neutron node. Verified when logged into a VM with a provider network.
  - **Shutdown 1 DB node**: access to Dashboard and use of APIs continued to work.
  - **Shutdown 1 controller node**: access to Dashboard and use of APIs continued to work.

Openstack Dashboard

Hypervisor Summary

VCPU Usage

Used 144 of 1,920

Memory Usage

Used 293GB of 4.9TB

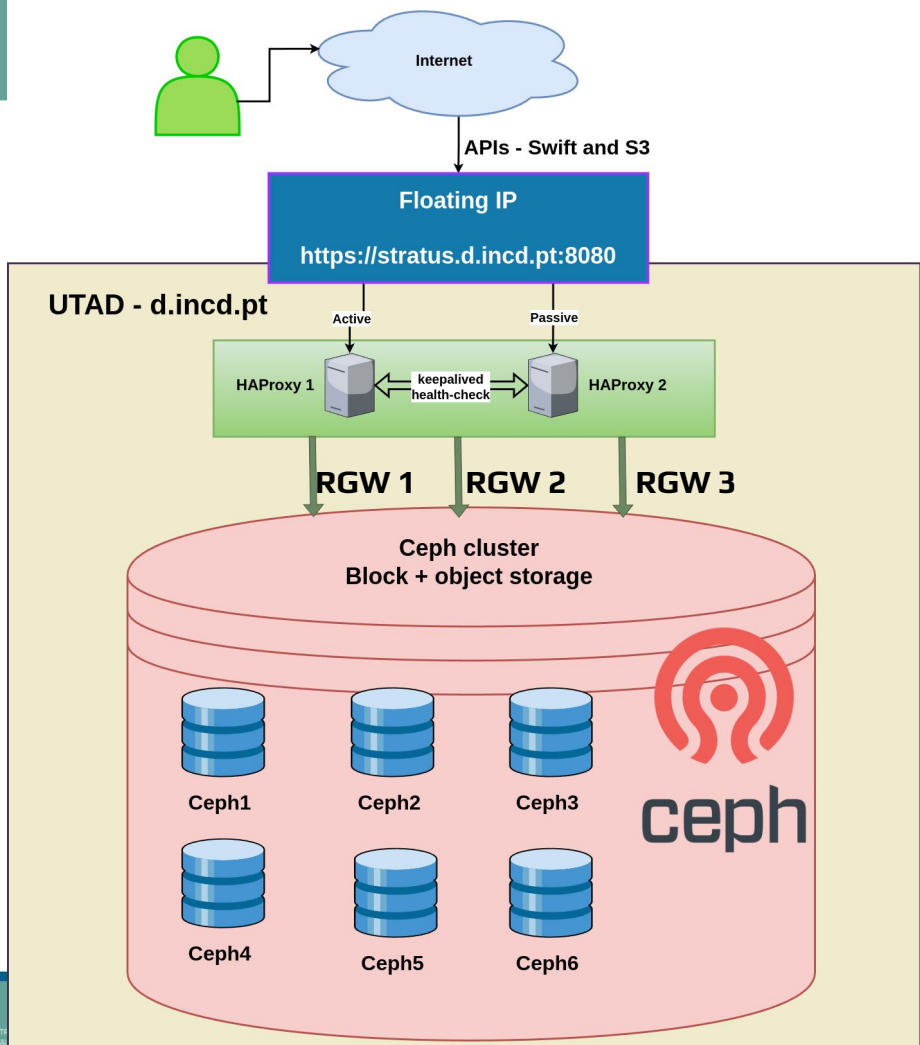
Hypervisor | Compute Host

Exibindo 10 itens

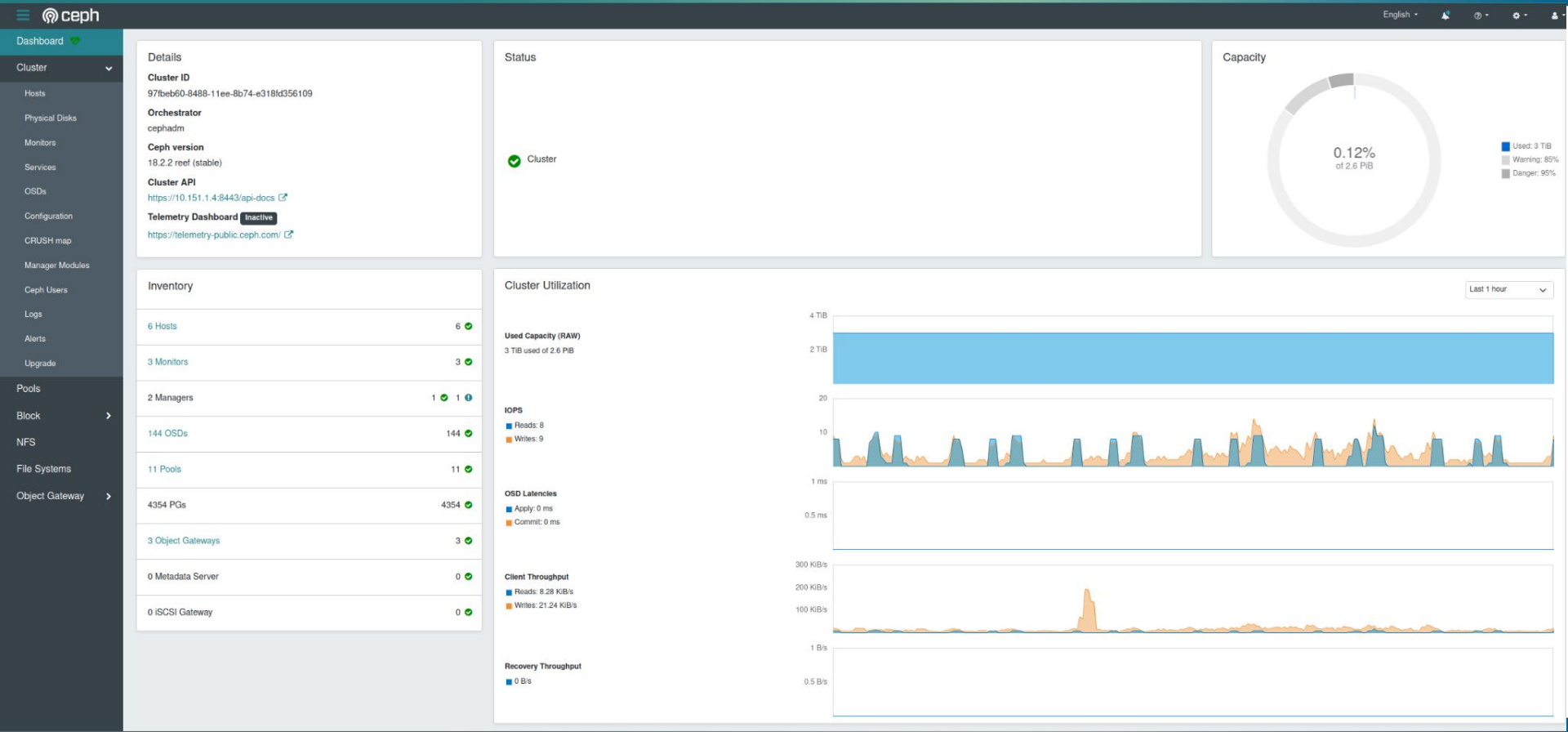
Hostname	Type	VCPUs (used)	VCPUs (total)	RAM (used)	RAM (total)	Local Storage (used)
comp-001.d.incd.pt	QEMU	0	192	512MB	503.5GB	0B
comp-002.d.incd.pt	QEMU	0	192	512MB	503.5GB	0B

UTAD CEPH architecture

- CEPH REEF (18.2.2)
- 6 storage nodes:
  - 24 SATA3 disks - 18.2TB each.
  - Total 2.6PB raw.
  - Replica 3 ~870 TB available.
- Deployment:
  - cephadm and podman:
    - Official docker container images.
  - 3 Rados GW services (SWIFT, S3).
    - Under haproxy.
  - AuthN/AuthZ → keystone.

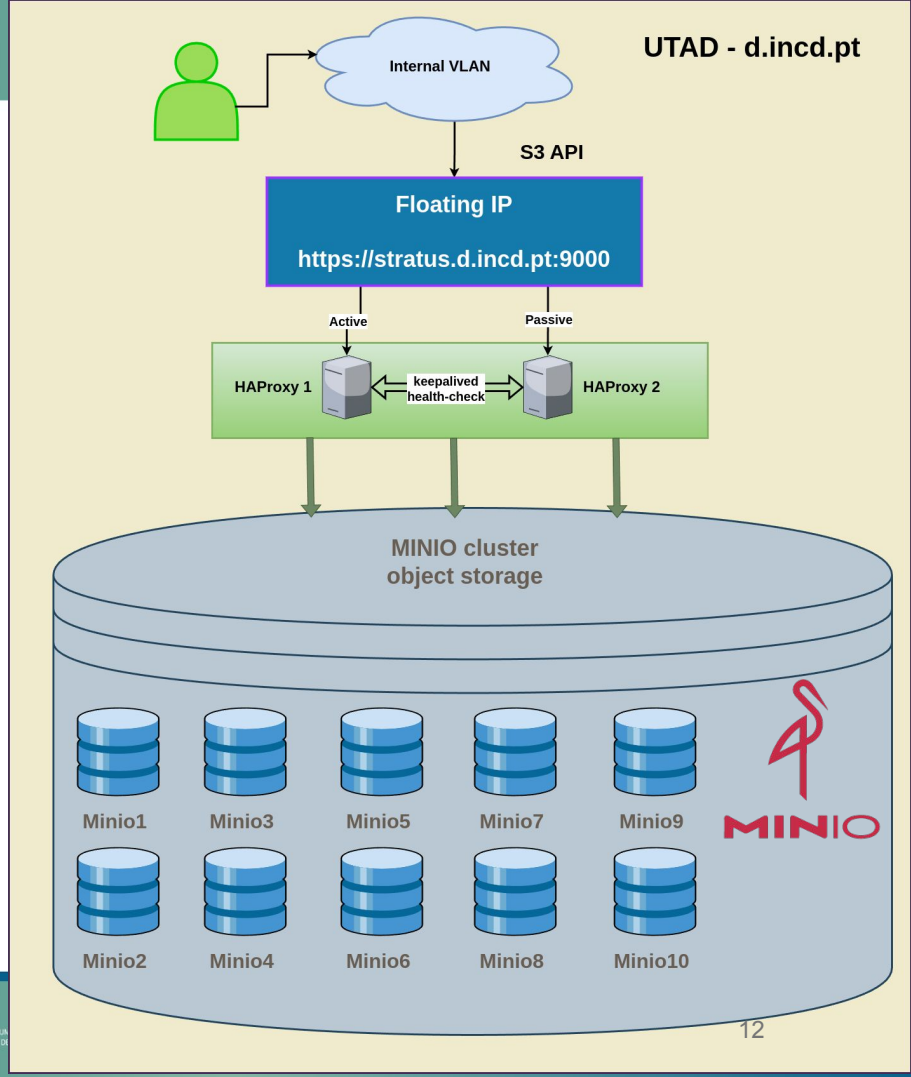


CEPH Dashboard

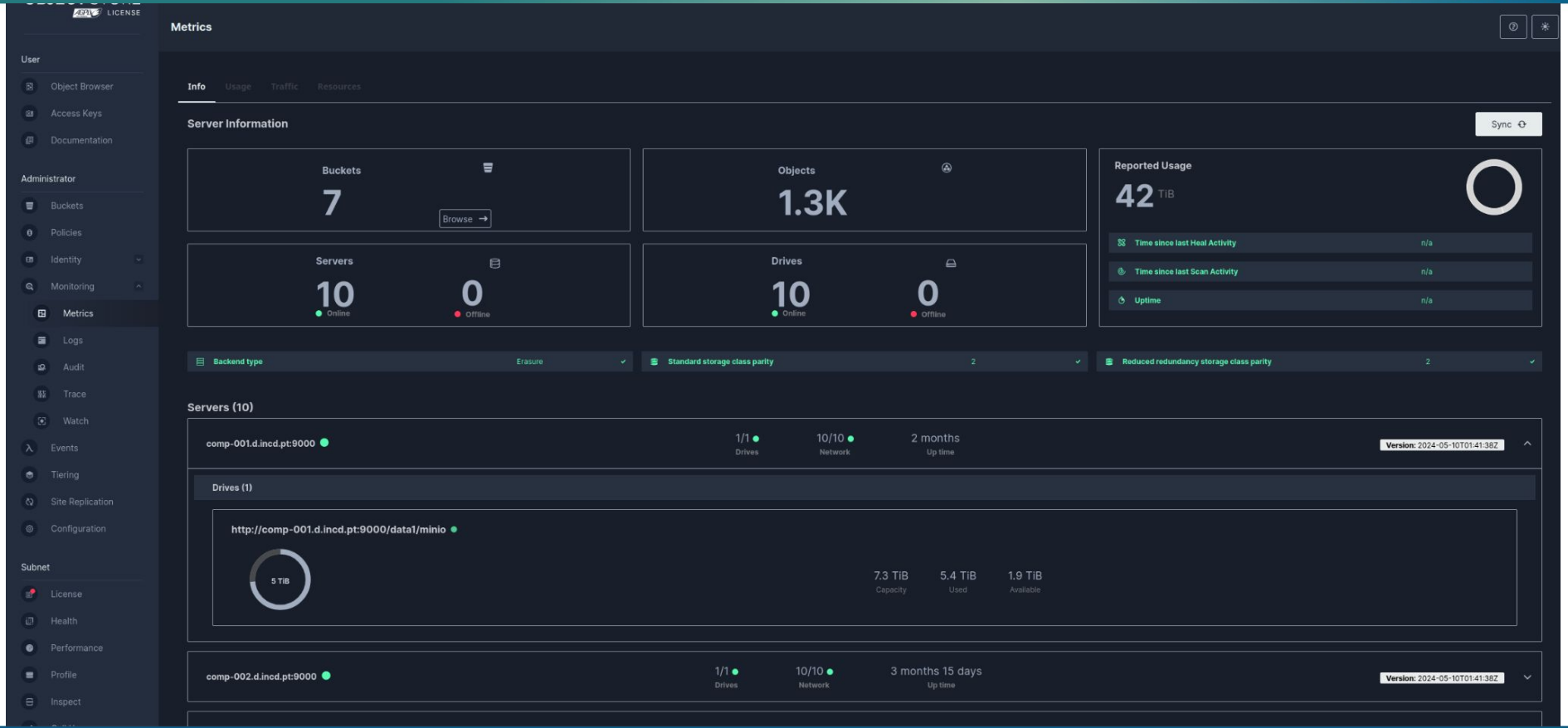


UTAD MINIO architecture

- MINIO 2024-05-10
- Deployed on the 10 compute nodes:
  - 1 SATA3 disk - 7.3TB each.
  - Total 73TB raw.
  - Erasure code 2 parity disks ~58 TB available.
- Deployment:
  - Custom ansible playbooks.
  - 10 minio S3 API services:
    - Under haproxy.
- **Only for internal/INCD usage.**



MINIO Dashboard



Monitorization

Nagios core 4.5.6

Nagios®

stratus-001.d.incd.pt	Disk free /	OK	10-23-2024 10:55:39	3d 21h 23m 4s	1/1	DISK OK - free space: / 63059 MiB (81.78% inode=97%):
	Disk free /home	OK	10-23-2024 10:57:01	3d 21h 22m 13s	1/1	DISK OK - free space: / 63059 MiB (81.78% inode=97%):
	Disk free /var	OK	10-23-2024 10:57:51	3d 21h 21m 23s	1/1	DISK OK - free space: / 63059 MiB (81.78% inode=97%):
	Load per CPU	OK	10-23-2024 10:58:02	3d 21h 21m 12s	1/1	OK - load average per CPU: 0.02, 0.02, 0.02
	Service cinder-scheduler	OK	10-23-2024 10:56:29	3d 20h 47m 45s	1/1	Active: active (running) since Thu 2024-09-26 16:18:05 WEST; 3 weeks 5 days ago
	Service cinder-volume	OK	10-23-2024 10:58:29	3d 20h 50m 45s	1/1	Active: active (running) since Thu 2024-09-26 16:18:03 WEST; 3 weeks 5 days ago
	Service glance-api	OK	10-23-2024 10:58:09	3d 20h 51m 6s	1/1	Active: active (running) since Thu 2024-09-26 16:18:04 WEST; 3 weeks 5 days ago
	Service httpd	OK	10-23-2024 10:58:59	3d 20h 50m 15s	1/1	Active: active (running) since Thu 2024-09-26 16:18:05 WEST; 3 weeks 5 days ago
	Service memcached	OK	10-23-2024 10:54:49	3d 20h 49m 25s	1/1	Active: active (running) since Thu 2024-08-01 10:13:46 WEST; 2 months 22 days ago
	Service neutron-server	OK	10-23-2024 10:55:39	3d 20h 48m 35s	1/1	Active: active (running) since Thu 2024-09-26 16:18:04 WEST; 3 weeks 5 days ago
	Service nova-api	OK	10-23-2024 10:56:29	3d 20h 47m 45s	1/1	Active: active (running) since Thu 2024-09-26 16:18:03 WEST; 3 weeks 5 days ago
	Service nova-conductor	OK	10-23-2024 10:58:33	3d 20h 50m 41s	1/1	Active: active (running) since Thu 2024-09-26 16:18:04 WEST; 3 weeks 5 days ago
	Service nova-novncproxy	OK	10-23-2024 10:58:07	3d 20h 51m 5s	1/1	Active: active (running) since Thu 2024-09-26 16:18:04 WEST; 3 weeks 5 days ago
	Service nova-scheduler	OK	10-23-2024 10:58:59	3d 20h 50m 15s	1/1	Active: active (running) since Thu 2024-09-26 16:18:03 WEST; 3 weeks 5 days ago
	Service rabbitmq-server	OK	10-23-2024 10:54:49	3d 20h 49m 25s	1/1	Active: active (running) since Tue 2024-07-09 15:31:40 WEST; 3 months 14 days ago
	ntp time	OK	10-23-2024 10:55:39	3d 21h 19m 37s	1/1	NTP OK: Offset -0.0008652210236 secs, stratum best:3 worst:3
	ro mounts /	OK	10-23-2024 10:56:29	3d 21h 23m 53s	1/1	RO_MOUNTS OK: No ro mounts found
	ro mounts /home	OK	10-23-2024 10:55:50	3d 21h 23m 3s	1/1	RO_MOUNTS OK: No ro mounts found
	ro mounts /var	OK	10-23-2024 10:57:01	3d 21h 22m 13s	1/1	RO_MOUNTS OK: No ro mounts found
	ssh check	OK	10-23-2024 10:57:51	3d 21h 21m 23s	1/1	SSH OK - OpenSSH_8.9p1 Ubuntu-3ubuntu0.10 (protocol 2.0)

ceph001.d.incd.pt	Disk free /	OK	10-23-2024 10:53:03	5d 22h 14m 25s	1/1	DISK OK - free space: / 72361 MiB (76.10% inode=96%):
	Disk free /home	OK	10-23-2024 10:50:18	5d 22h 13m 25s	1/1	DISK OK - free space: / 72362 MiB (76.10% inode=96%):
	Disk free /var	OK	10-23-2024 10:51:08	5d 22h 12m 25s	1/1	DISK OK - free space: / 72361 MiB (76.10% inode=96%):
	Load per CPU	OK	10-23-2024 10:53:21	5d 22h 11m 25s	1/1	OK - load average per CPU: 0.01, 0.02, 0.02
	SmartMon sdc	OK	10-23-2024 10:52:46	4d 0h 2m 22s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3WN: no SMART errors detected.
	SmartMon sdd	OK	10-23-2024 10:52:25	4d 0h 1m 22s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3QZ: no SMART errors detected.
	SmartMon sde	OK	10-23-2024 10:53:24	4d 0h 0m 21s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4AT: no SMART errors detected.
	SmartMon sdf	OK	10-23-2024 10:52:35	3d 22h 16m 11s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y45T: no SMART errors detected.
	SmartMon sdg	OK	10-23-2024 10:53:24	3d 22h 15m 21s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y33Y: no SMART errors detected.
	SmartMon sdh	OK	10-23-2024 10:49:15	3d 22h 14m 31s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4C1: no SMART errors detected.
	SmartMon sdi	OK	10-23-2024 10:48:37	3d 22h 13m 41s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3ZB: no SMART errors detected.
	SmartMon sdj	OK	10-23-2024 10:49:28	3d 22h 12m 50s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3VP: no SMART errors detected.
	SmartMon sdk	OK	10-23-2024 10:52:33	3d 22h 16m 13s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y37J: no SMART errors detected.
	SmartMon sdl	OK	10-23-2024 10:52:35	3d 22h 16m 11s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4GJ: no SMART errors detected.
	SmartMon sdm	OK	10-23-2024 10:53:26	3d 22h 15m 21s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4SV: no SMART errors detected.
	SmartMon sdn	OK	10-23-2024 10:49:15	3d 22h 14m 31s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT2J739: no SMART errors detected.
	SmartMon sdo	OK	10-23-2024 10:50:19	3d 22h 13m 40s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y434: no SMART errors detected.
	SmartMon sdp	OK	10-23-2024 10:52:36	3d 22h 12m 50s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3JM: no SMART errors detected.
	SmartMon sdq	OK	10-23-2024 10:51:08	3d 22h 12m 31s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y45R: no SMART errors detected.
	SmartMon sdr	OK	10-23-2024 10:52:35	3d 22h 16m 11s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4K9: no SMART errors detected.
	SmartMon sds	OK	10-23-2024 10:53:26	3d 22h 15m 20s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4QN: no SMART errors detected.
	SmartMon sdt	OK	10-23-2024 10:49:17	3d 22h 14m 30s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y491: no SMART errors detected.
	SmartMon sdu	OK	10-23-2024 10:51:45	3d 22h 13m 40s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4AC: no SMART errors detected.
	SmartMon sdv	OK	10-23-2024 10:53:25	3d 22h 12m 50s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y3JP: no SMART errors detected.
	SmartMon sdw	OK	10-23-2024 10:52:38	3d 22h 16m 6s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT3Y4QJ: no SMART errors detected.
	SmartMon sdx	OK	10-23-2024 10:52:36	3d 22h 16m 10s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT2JC64: no SMART errors detected.
	SmartMon sdz	OK	10-23-2024 10:53:25	3d 22h 15m 20s	1/1	OK: Drive ST20000NM008D-3DJ133 S/N ZVT2JC42: no SMART errors detected.
	ntp time	OK	10-23-2024 10:52:03	5d 22h 10m 25s	1/1	NTP OK: Offset -0.0008407831192 secs, stratum best:3 worst:3
	ro mounts /	OK	10-23-2024 10:53:03	5d 22h 14m 25s	1/1	RO_MOUNTS OK: No ro mounts found
	ro mounts /home	OK	10-23-2024 10:52:49	5d 22h 13m 25s	1/1	RO_MOUNTS OK: No ro mounts found
	ro mounts /var	OK	10-23-2024 10:52:47	5d 22h 12m 25s	1/1	RO_MOUNTS OK: No ro mounts found
	ssh check	OK	10-23-2024 10:48:38	5d 22h 11m 25s	1/1	SSH OK - OpenSSH_8.9p1 Ubuntu-3ubuntu0.10 (protocol 2.0)

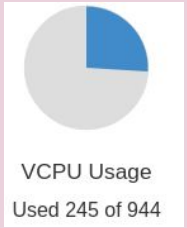
New Openstack infrastructure @INCD Lisbon

- Same architecture, deployment, versions and configuration as of UTAD.
- Same custom ansible playbooks - only change the hosts/inventory and variables.
- Started migration of projects, VMs and storage from the old Openstack.
  - In most cases the procedure is to instantiate new VMs and copying data from the old ones.
- Migrate compute nodes and CEPH storage nodes from old to new.

Stratus - A

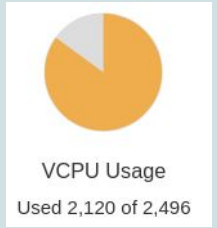
New infra:

VCPUs: **944**
CEPH Storage RAW/Avail: **520/175 TB**



Old infra:

VCPUs: **2 496**
CEPH Storage RAW/Avail: **520/175 TB**



Summary

- Use of custom ansible playbooks at INCD-UTAD, allowed easier and faster deployment/configuration of the new INCD-Lisbon Openstack.
- Use of Ubuntu LTS releases will allow upgrade of the Operating System in place (no re-installation).
- Use of LXD/LXC containers for the Openstack controllers and Databases will allow smooth upgrade of Openstack by deployment in new containers, with easy rollback if something goes wrong.
- CEPH deployment with cephadm and podman (Docker containers) provide the ability of live upgrading the system.
- The total amount of resources: **5360** VCPUs, **3.6/1.2 PB** Raw/avail storage.