Switch

SWITCH cloud - S3

Version 01, 13. März 2024

Agenda

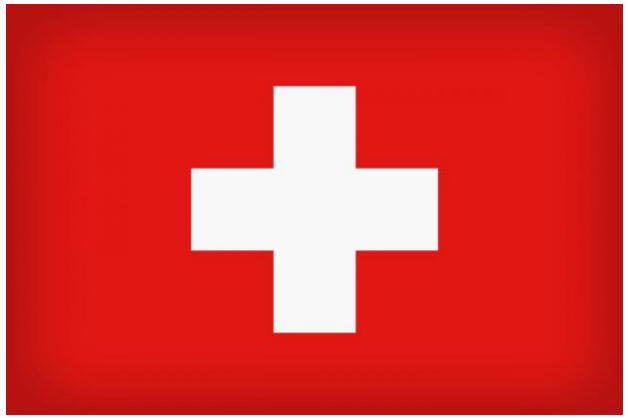
1. Introduction
2. Which S3 storage to use?
3. Not only Ceph
4. Where are we now?
5. Conclusion

Introduction

Switch

The Swiss NREN – 35 years.

Providing many services to our community.



Switch Cloud

The cloud built for the community.

“SWITCHEngines2”:

- Openstack
- Block storage
- S3 storage
- Networking services

openstack
CLOUD SOFTWARE

CURE – storage

Responsibilities:

- S3 clusters (~25PB raw)
- Block storage for Engines1
- Future S3 clusters

Which S3 to use?

- Workloads: a bit of everything
  - Backups: millions of small objects
  - Data science: very large files
  - Everything in the middle

- Our S3 clusters:
  - 2 * Ceph clusters
  - 2 * «off the shelf» S3 clusters

- Option 1: off the shelf S3
  - Bad experience => 40% of our total S3 capacity, 80% of our pain
  - Always significantly more expensive

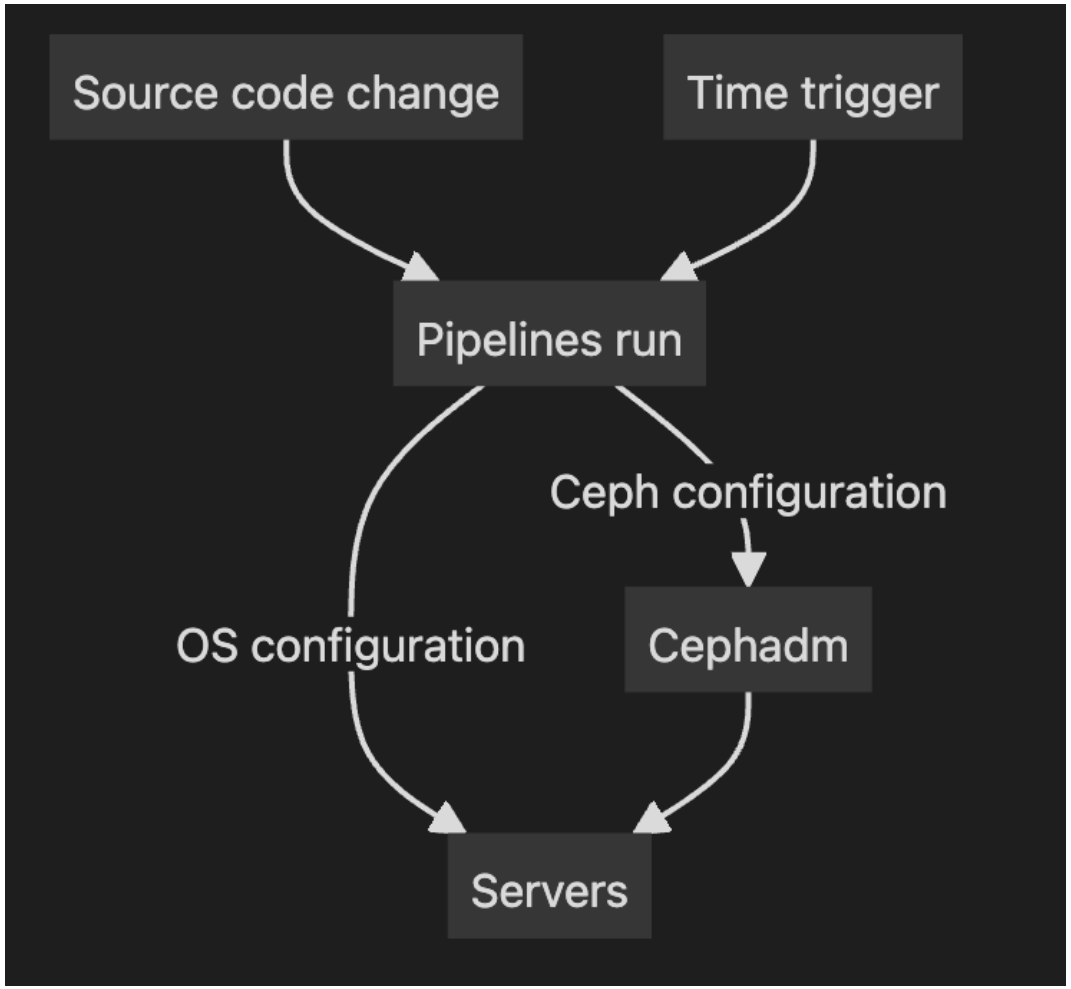
- Option 2: opensource based
  - Ceph does the trick. Services on top are very limited -> gaps to fill
  - MinIO gateway not an option

Ref: talk of INFN/Alessandro Costantini

Not only Ceph

Based on past experience:

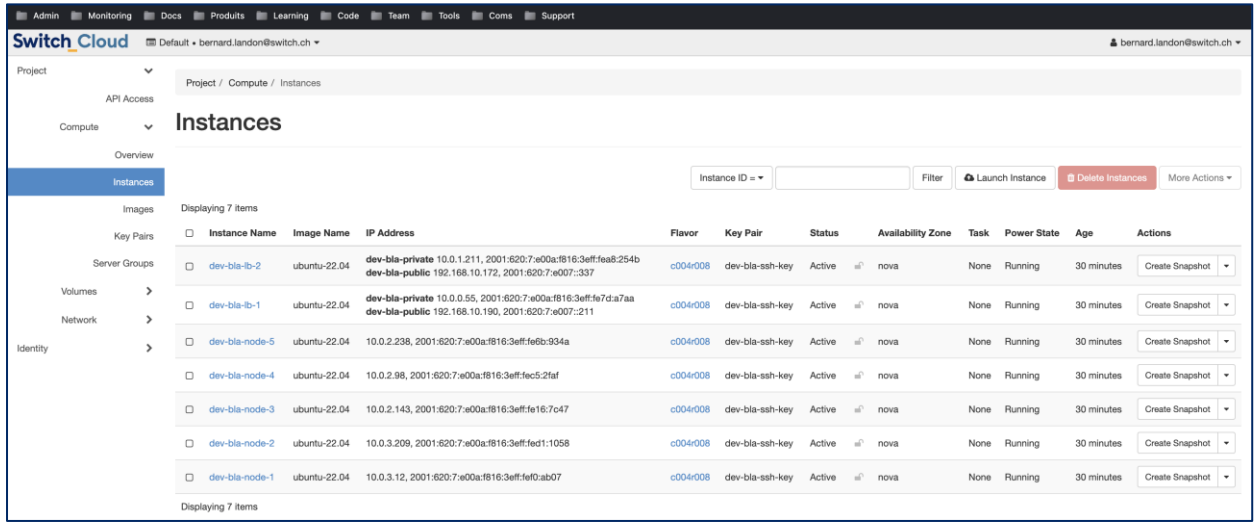
- Manual maintenance can be very tedious => orchestrator needed
- A lot of configuration on the servers => Ansible
- Test environnements
- Configuration drift



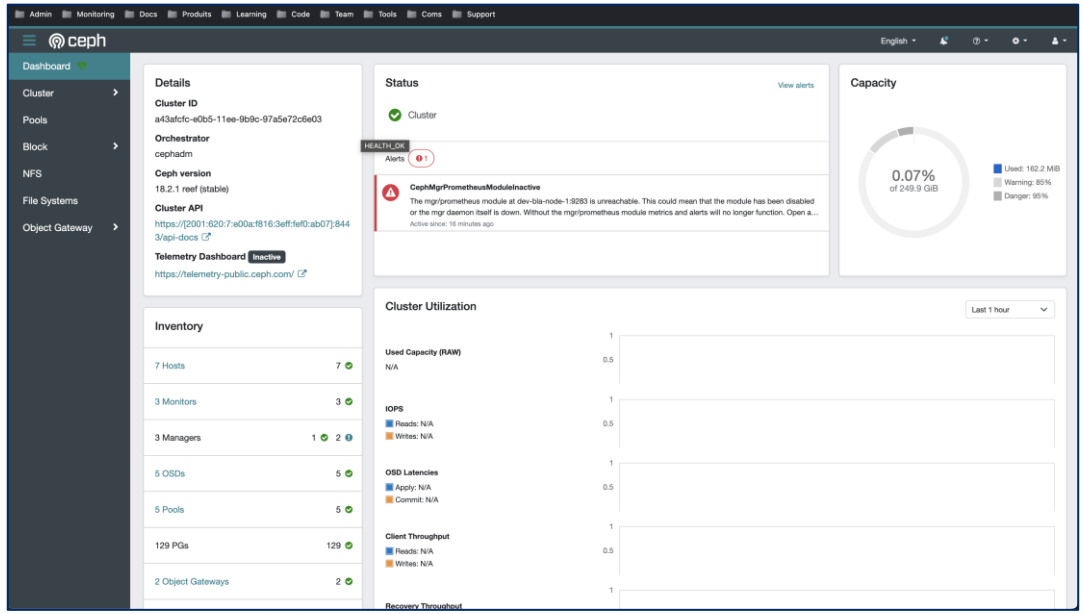
Where are we now? (1/2)

Current status

- Deployment of our latest cluster:
  520 OSD ~ 2.5PB raw done in ~2h
- Ingress service (Haproxy/Keepalived) works
- Deploying/scraping clusters for testing
  from nothing to Ceph in ~16 mins => super powerful!




```
√ ubuntu@we:~/we/git/storage/osv3-cm > export CLUSTER_NAME=dev-bla
√ ubuntu@we:~/we/git/storage/osv3-cm > make bootstrap
```



Surprises

- Lack of IPv6 / dual stack support. In 2024. Really?
- Cephadm IS an orchestrator => need time to learn
- Working with Ansible can be surprisingly frustrating!

Where are we now? (1/2)

Work in progress

- Workflow not yet fully set up
- Network setup not as we’d like. Maybe using a custom container service to deploy BGP?
- Monitoring

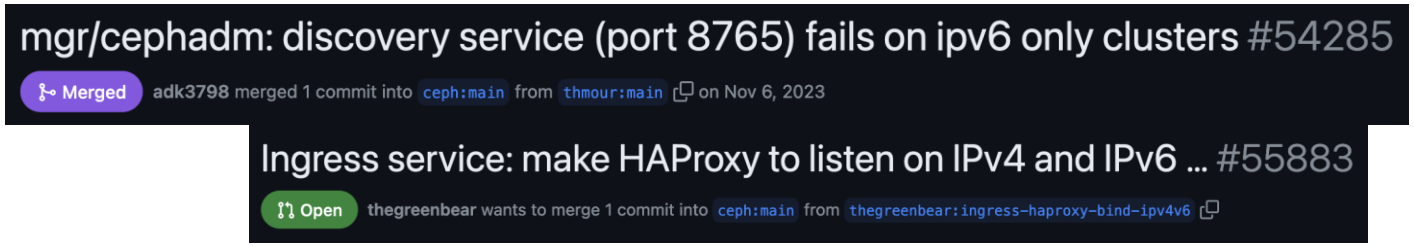
Main challenge

- Many workflows not easily supported (groups, ro, etc)

To wrap up

- We have an alpha release!
- We try to patch issues as we go…
- Solid foundations to spawn / maintain more clusters.
- Still a lot of work / a lot to learn !
- Not a promise – but we hope to open-source our code! Any interest?

α



Q&A

Bernard Landon
Cloud Engineer
CURE Team

bernard.landon@switch.ch
EPFL Innovation Park
Bâtiment I
1015 Lausanne

Switch_