



Cloud Resilience Unleashed: A Dual-Site Approach for HL-LHC era

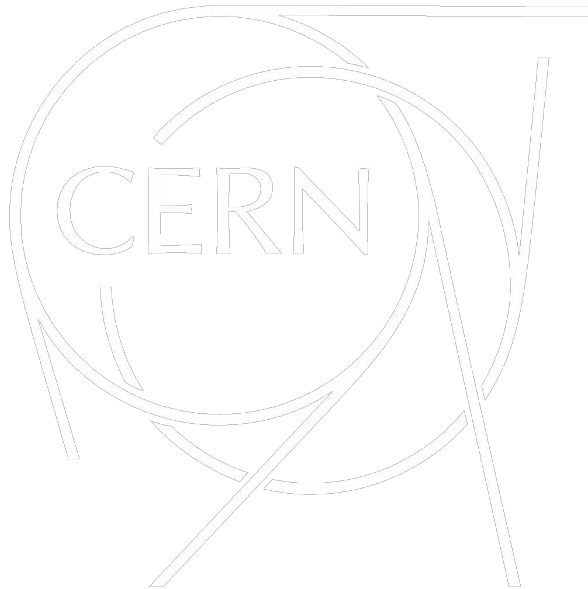
HEPiX Spring 2024

Daniel Failing, CERN



Outline

- CERN Cloud Overview
- Challenges and opportunities
- Preveessin Data Centre (PDC)
- Cloud Deployment in PDC
- Next Steps

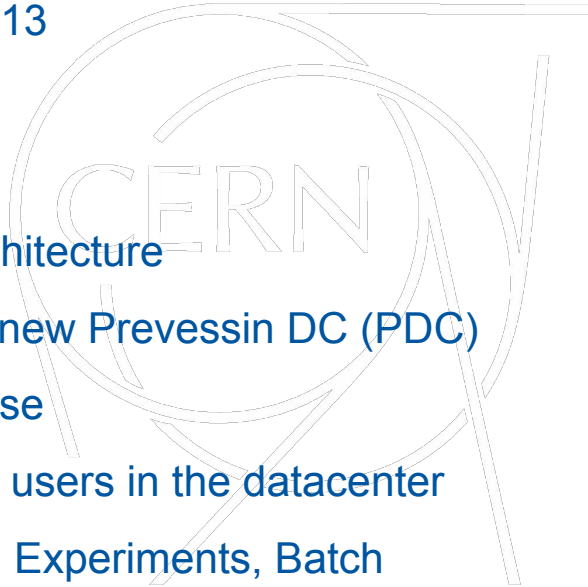


CERN Cloud Infrastructure



openstack®

- Infrastructure as a Service
 - Production since July 2013
- RHEL/ALMA9
 - Based on RDO
 - x86_64 and Aarch64 architecture
- Meyrin Data Centre (MDC) + new Preveessin DC (PDC)
- Currently running Yoga* release
- Providing VMs/Bare-metal for users in the datacenter
 - Services, Personal VMs, Experiments, Batch



Component	S	Train	U	V	W	X	Yoga	Z	A	B
OS	CC7	EL8		EL9						
Hypervisors										
Barbican										
Nova										
Neutron										
Placement										
Horizon										
Cinder										
Manila										
Glance										
Mistral										
Keystone										
Rally										
Ironic										
Octavia										

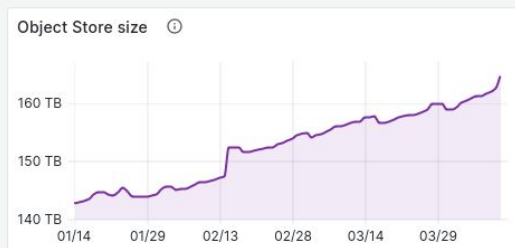
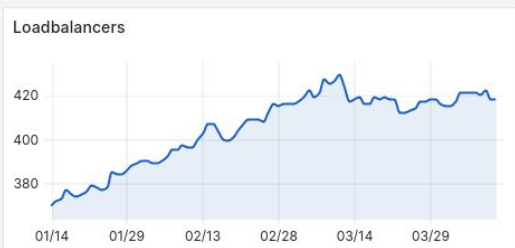
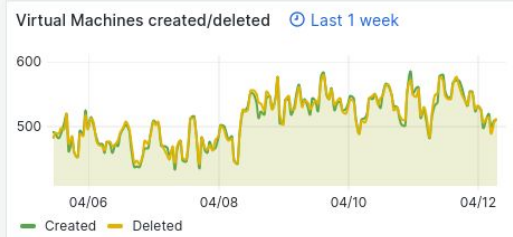
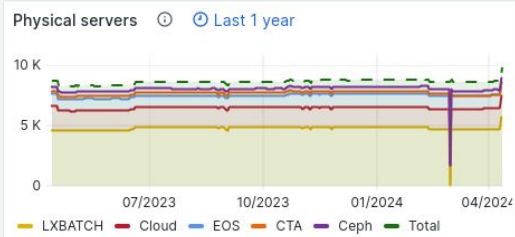


CERN Cloud Overview

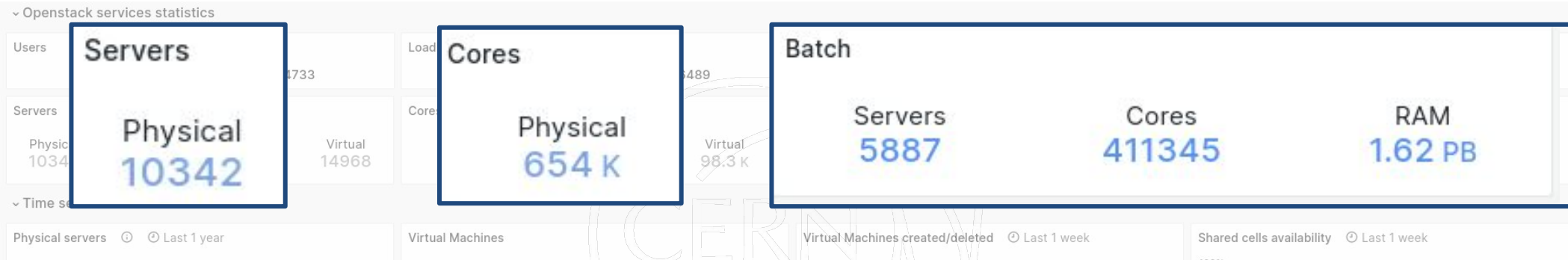
Openstack services statistics

Users 3426	Projects 4733	Loadbalancers 418	Images 6489	Volumes 7437	Volumes size 4.55 PB	File Shares 4418	File Shares size 2.24 PB	Object Store ... 592	Object Store ... 163 TB			
Servers				Cores			RAM			Batch		
Physical 10342	Physical in use 9794	Hypervisors 1767	Virtual 14968	Physical 654 K	Hypervisors 119 K	Virtual 98.3 K	Physical 2.88 PB	Hypervisors 418 TB	Virtual 222 TB	Servers 5887	Cores 411345	RAM 1.62 PB

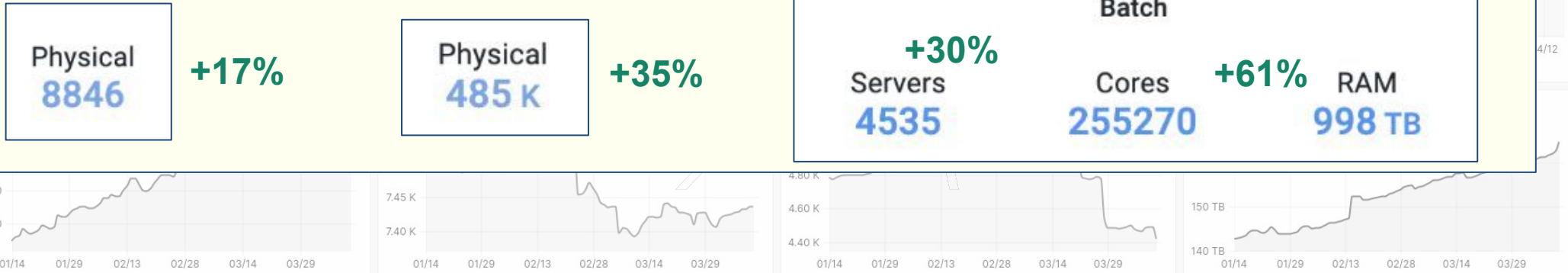
Time series



CERN Cloud Overview

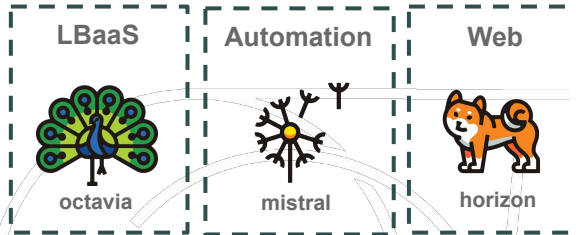


Compared to last year:

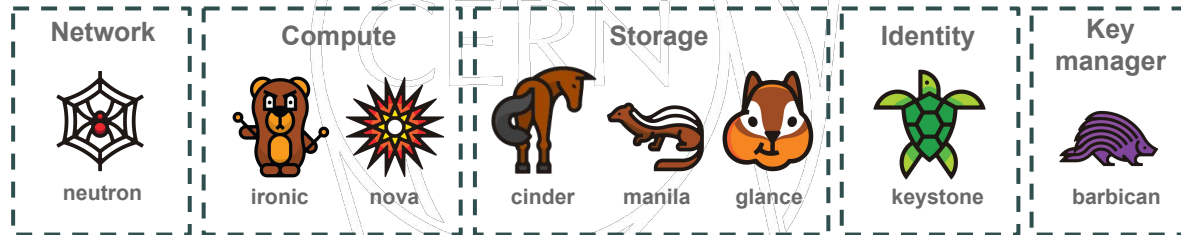


Cloud Infrastructure APIs

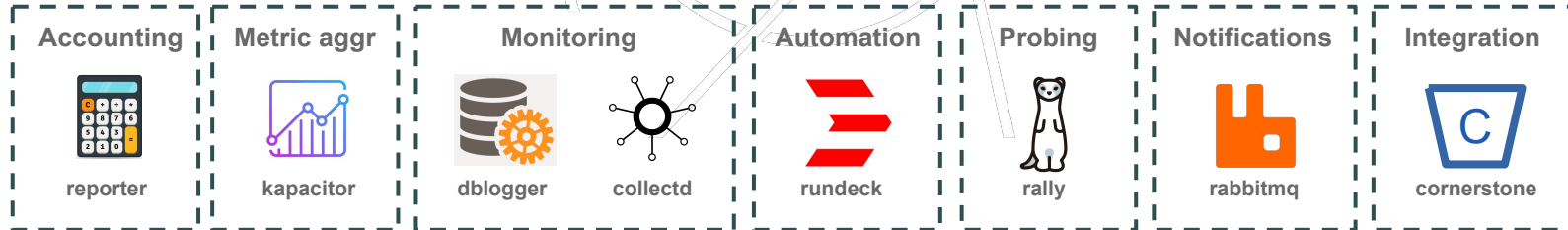
IaaS+



IaaS



Infra



User Visible

Challenges/Opportunities for the next years

Challenges - BC/DR

- CERN IT Strategy as Provider include:
 - Recognise operational risks
 - Definition of policies for BC/DR
 - Enable business continuity and disaster recovery (BC/DR)
 - Resources + training for services
- Cold recovery tests for all services planned 2024
- New resources for Active/Active and Active/Standby operations

AS-IS
External
Assessment



Recognise operational risks

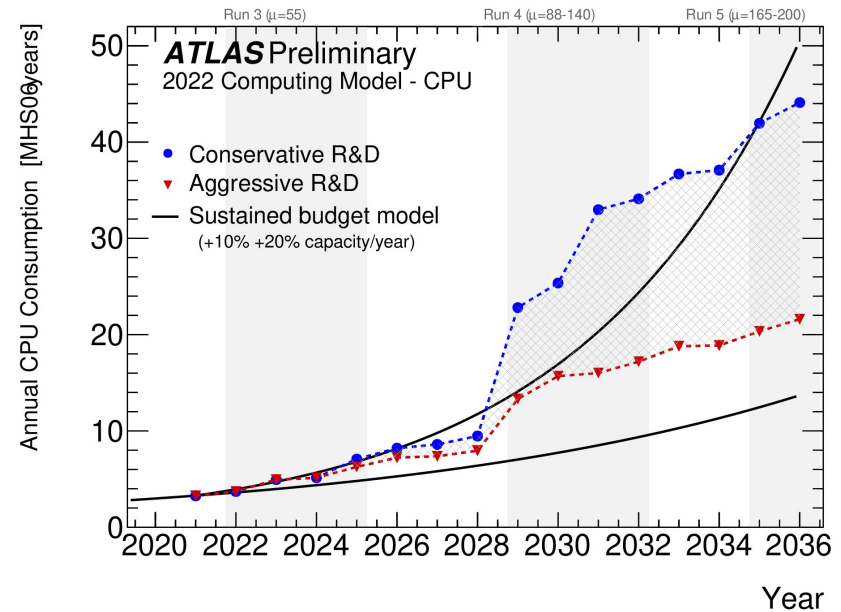
Today: IT disaster recovery and business continuity procedures are not adequate. Although failures are limited, the risk is significant to ongoing operations

'We don't have a proper disaster recovery and business continuity plan'

[Presentation about BC/DR](#)

Challenges - Computing

- LHC imposes increasing computing needs
 - More data throughput to our data centers
 - Expected CPU capacity growth not enough
- Physicists exploring new paradigms
 - Machine Learning and other techniques leveraging hardware accelerators (e.g. GPU)
 - New architectures like ARM64
- Moving towards a more heterogeneous cloud
 - Bring new features to the Cloud (e.g. MIG)
 - Adapt our operations (e.g. resource tracking)



Locations



Meyrin Data Centre

- 4 MW capacity
- Main room with UPSs
- Critical room on diesel generators



Containers Point8

- 0.5 MW without UPSs

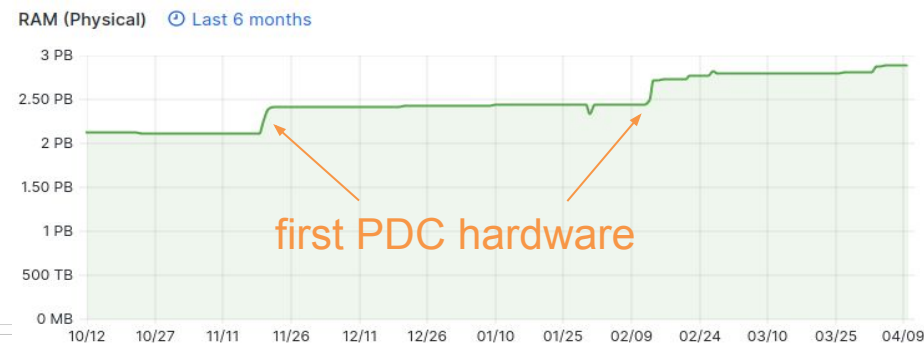


Preveessin Data Centre

- 3 floors with up to 12 MW
- 75% in non-UPS racks
- Inaugurated [Q1/2024](#)

Preveessin Data Centre (PDC)

- Inaugurated [23 February 2024](#)
- Most hardware installed Q4/2023 + Q1/2024
- Compute
 - Additional capacity for next run(s) of the LHC and HL-LHC
 - Currently one floor: 4 MW
 - Extendable to 3 floors / 12 MW
- Services
 - Enable BC/DR recovery scenarios
 - Use remaining space for services



Context – PDC Deployment Cloud - Goals

- Focus is to **enable** users to deploy their BC/DR scenarios
- Base building blocks should **facilitate** and enable those
 - Fabric, Network, Storage, Compute and Database
- Keep in mind **existing** use-cases and **future** ones
- Use of a **green field deployment**
 - Review existing shortcomings
 - Opportunity for new features
 - No legacy constraints
- **Remove dependencies** between data centers to enable own BC/DR

Cloud Deployment in PDC

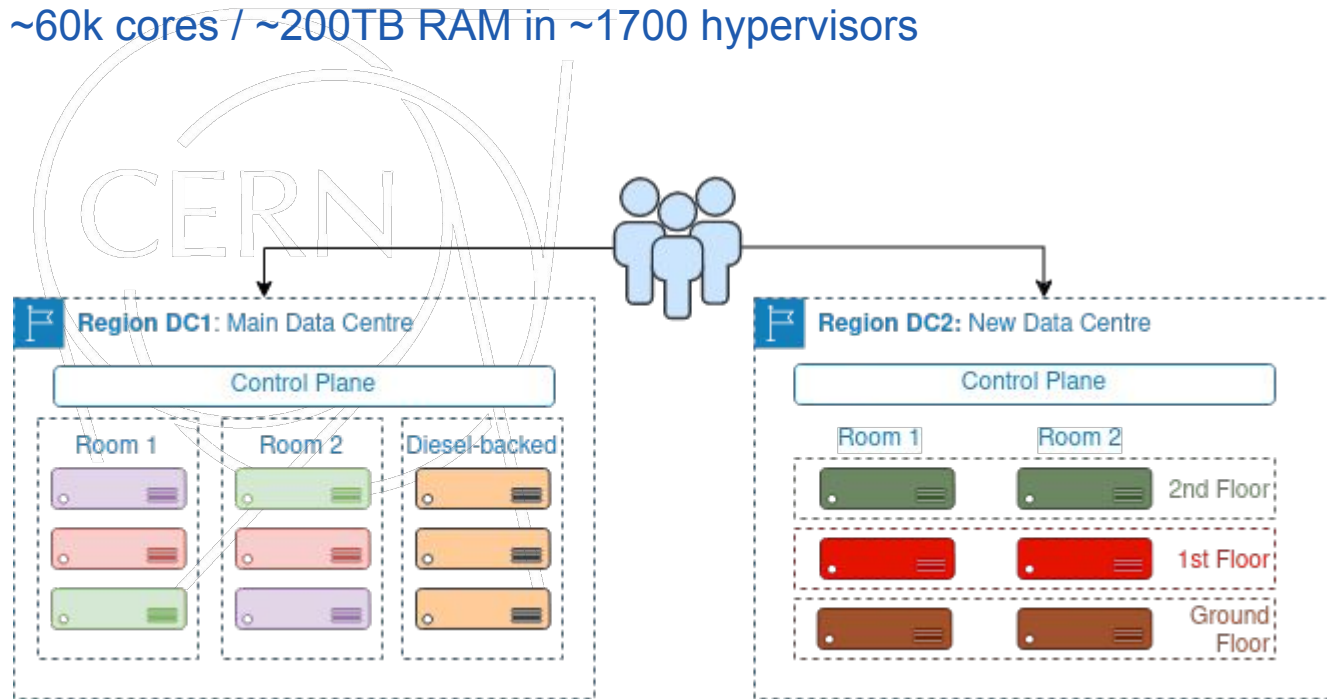
Fully independent region

- Complete break from existing deployment
 - Users will see a complete new set of APIs to talk with
 - Keep same project structure and ACLs in place
 - **No** need for **data replication** across sites
- Pros:
 - Enforces **cross-site restrictions** on applications and/or services
 - Allows **upgrade rollout** on sites
 - Covers a full stop of any of the DCs and any disconnection in between
- Cons:
 - Increased **support effort** on different service versions

2 independent regions (physical layout)

- Hypervisor Resources PDC: ~9000 cores / ~72TB RAM in ~100 hypervisors
 - Compared to old DC: ~60k cores / ~200TB RAM in ~1700 hypervisors

- Separated Control Plane
- Availability Zones:
 - match physical failure domain (power, net)
 - One per floor



Differences between regions

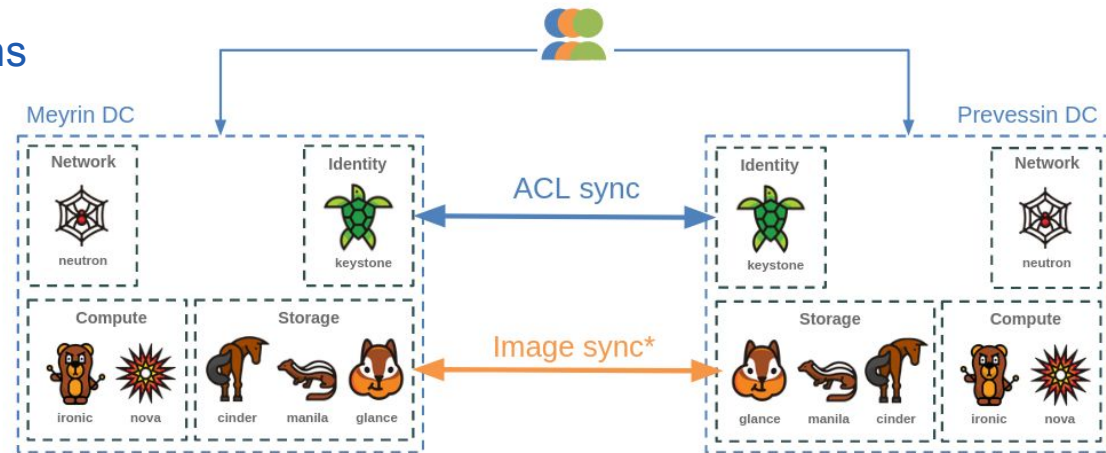
Feature	Meyrin DC	Prevessin DC
Availability Zones	3 Compute 3 Storage	1 Compute & Storage
Anti-/Affinity Filters	Host	Host, Rack, Room
Networks	Provider	Provider & Private
SDN Features	Load Balancers	Security Groups Floating IPs Load Balancers
Capacity (RAM)	250TB	72TB
Capacity (on Diesel)	12TB	-
UPS expected lifetime	15min	5min

Wiring the Future: Open Virtual Networking and Beyond *Daniel Failing*
 Amphithéâtre Buffon, Laboratoire Astroparticule et Cosmologie (APC) de l'Université Paris-Cité 13:30 - 13:55

[separate presentation in the afternoon](#)

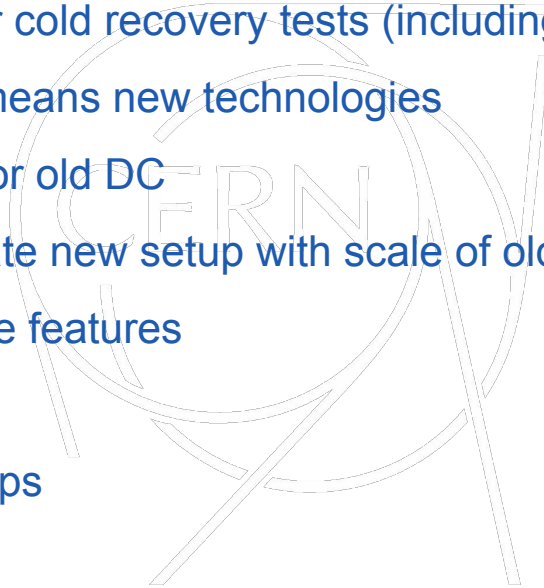
Tooling

- Custom scripts / automation
- ACL sync
 - Synchronize Projects and assignments across regions
 - Time based sync and on project creation
- Image sync
 - Copy public images between regions
 - Heavy operation, on-demand



Next Steps

- Foundations for BC/DR Use-cases are there
 - Services to perform their cold recovery tests (including us)
- Greenfield deployment also means new technologies
 - Expertise to be gained for old DC
 - Scalability tests to validate new setup with scale of old DC
 - Extend old DC with those features
- Storage
 - Cross DC volume backups
 - Cross DC file shares
 - [Talk from OpenInfra Summit Vancouver](#)



Thank You!



All our open source code is available on:
<https://gitlab.cern.ch/cloud-infrastructure>

Multiple contributors over the years helped to achieve this.

For contact:
Daniel Failing - [daniel.failing](mailto:daniel.failing@cern.ch) <at> cern.ch



