





CERN Cloud Infrastructure service

José Castro León
CERN Cloud Infrastructure

Outline

- Introduction
- Datacentres
 - Meyrin and Preveessin
- Resources
- Operations



CERN Cloud Infrastructure



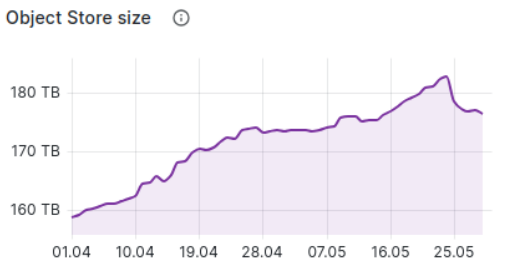
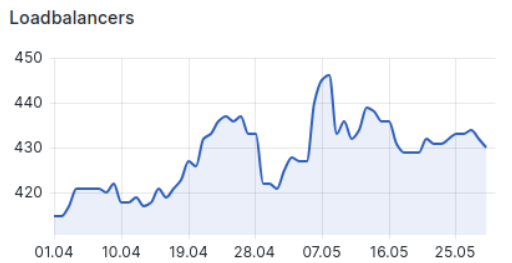
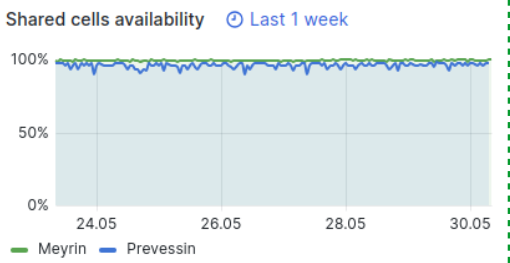
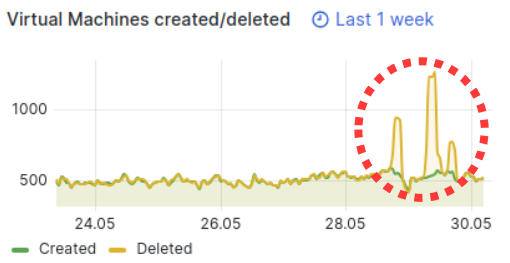
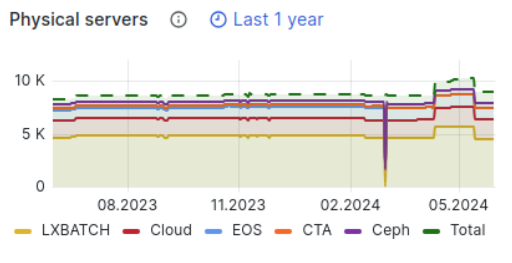
- Infrastructure as a Service
- Production since **July 2013**
- Running on **Redhat Enterprise Linux / AlmaLinux 9**
 - Based on Redhat Distribution of OpenStack (RDO)
- Meyrin and Preveessin Data Centres
- Currently running **Yoga+** release
 - Some services already in Zed release



Openstack services statistics

Users 3450	Projects 4768	Loadbalancers 430	Images 6569	Volumes 7480	Volumes size 4.60 PB	File Shares 4486	File Shares size 2.27 PB	Object Store b 601	Object Store si 175 TB			
Servers				Cores			RAM			Batch		
Physical 9145	Physical in use 8946	Hypervisors 1860	Virtual 17863	Physical 585 K	Hypervisors 482 K	Virtual 116 K	Physical 2.71 PB	Hypervisors 492 TB	Virtual 286 TB	Servers 4684	Cores 333259	RAM 1.38 PB

Time series



Initial offering

IaaS+

Web



horizon

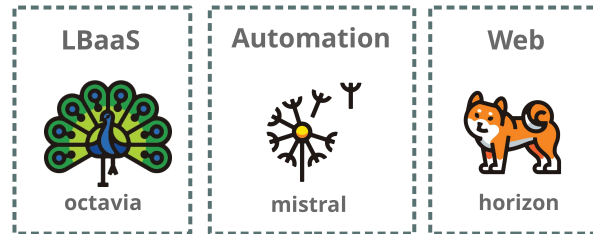
IaaS

<p>Compute</p>  <p>nova</p>	<p>Storage</p>  <p>glance</p>	<p>Identity</p>  <p>keystone</p>
--	--	---

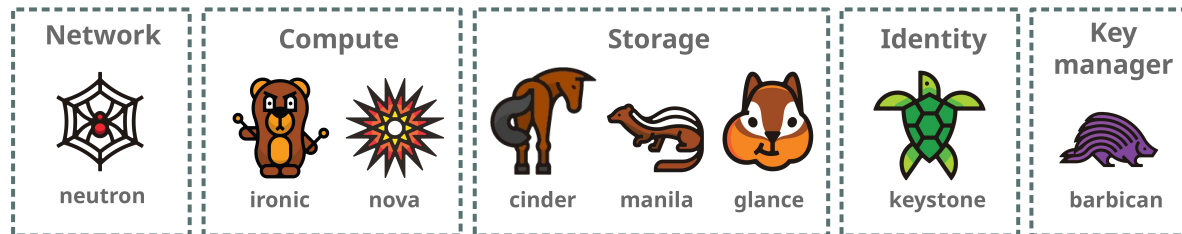
in Meyrin DC

CERN Cloud Infrastructure - now

IaaS+

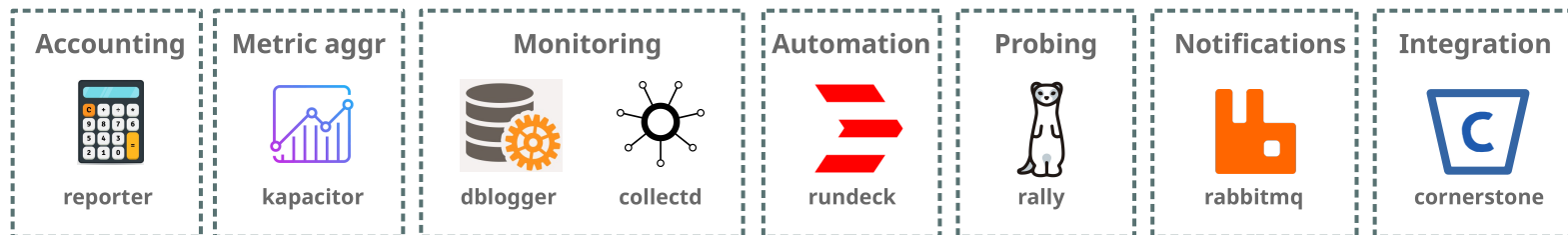


IaaS

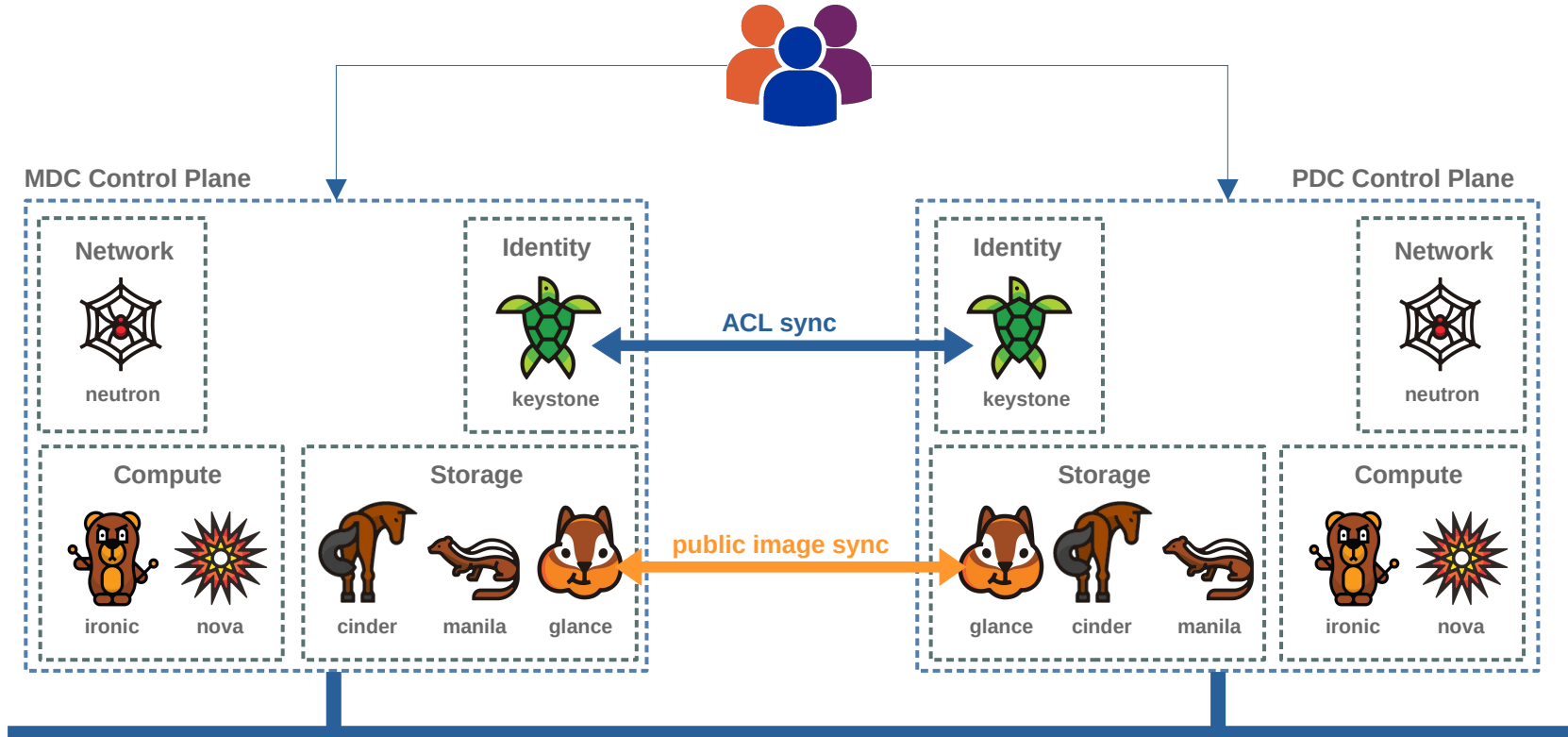


In MDC & PDC

Infra



Minimal interactions between sites



Differences between datacentres

Feature	Meyrin DC	Prevessin DC
OpenStack version	Yoga+	
OS version	RHEL9 / ALMA9	
Availability Zones	3 Compute 3 Storage	1 Compute & Storage
Number of Cells	34	1
Cross Zone attachments	YES	NO
Anti-/Affinity Filters	Host	Host, Rack, Room
Networks	Provider	Provider & Private
SDN Features	Load Balancers	Security Groups Load Balancers Floating IPs
Capacity (Memory on HV)	375TB	72TB
Capacity (on Diesel)	12TB	-
UPS expected lifetime	15min	5min



Available to all CERN users

- Also accessible on-demand to federated users (based on group membership)
- Types of projects (owned by a CERN primary account)

	Affiliation Expired	User Disabled	User Deleted
Shared	Promote	-	-
Personal	-	Stop	Delete

How to requests resources?

REQUEST NEW PROJECT

Create a new project

Please provide some details about the project like: name, description, owner, egroups and comments.

Details

Project Name *

Description *

Charge group *

Owner *

Administrator egroup(s) *

Set group as default responsible in landb

Set group as default mainuser in landb


Additional Comments

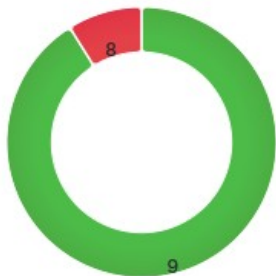
X CANCEL ← BACK NEXT → + CREATE NEW PROJECT



Service deployment

- From shared to “**per microservice**” architecture
- All deployed in VMs on our own infrastructure: “*eat our own dogfood*”
 - Bootstrap procedure and recovery methods
- Puppet managed running on RHEL/ALMA 8 and 9

Service VMs by OS (version)  Last 1 day



	Value	Percent
9	546	91%
8	53	9%

Hypervisors by OS (version)  Last 1 day

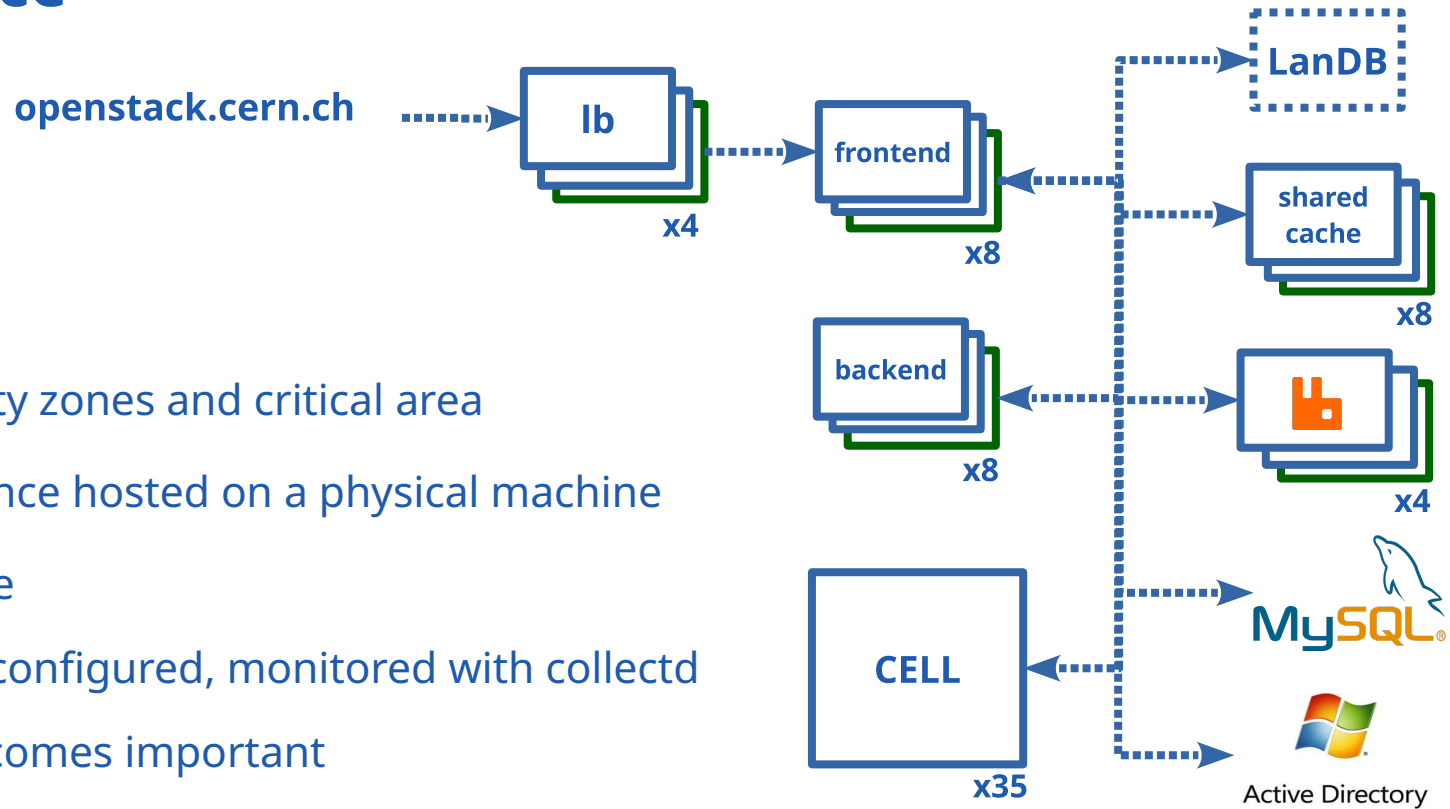


	Value	Percent
9	1863	100%
7	2	0%

Service operations

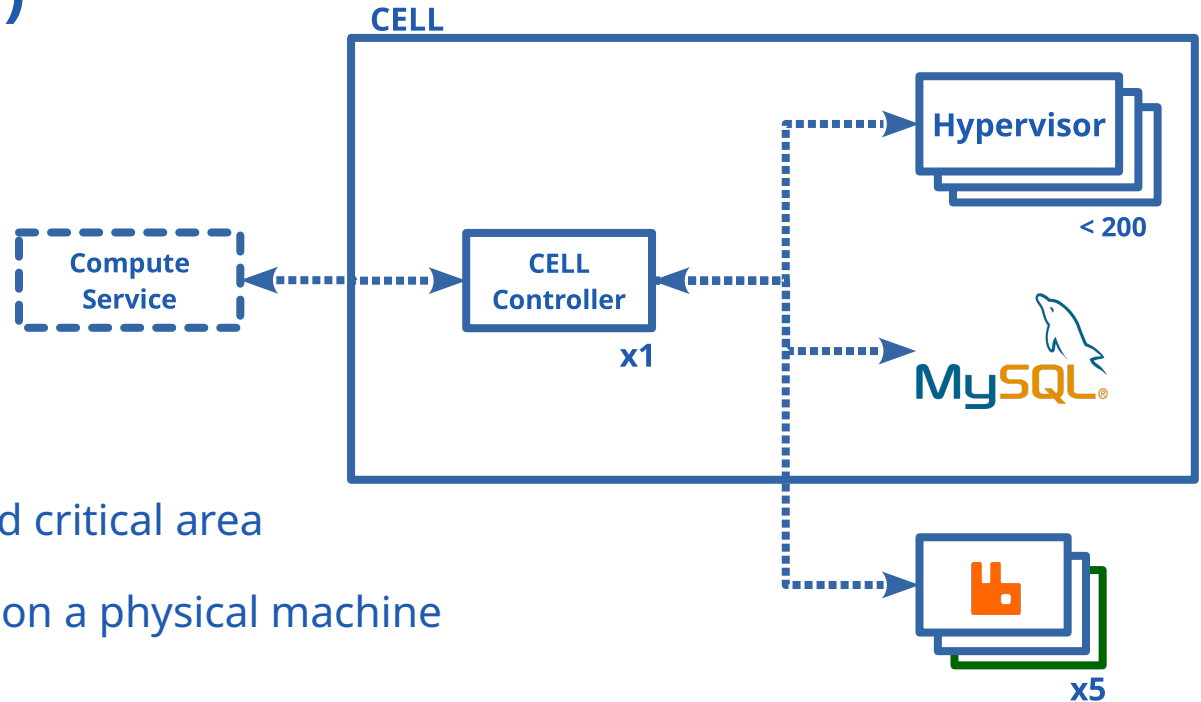
- Meyrin Deployment upgraded since **July 2013**
- Per-service upgrade model (A/B or in place)
- Compute + Storage availability zones
- Huge investment on **automation**:
 - Delegate as much as possible administrative tasks (repair team, quota mgmt, end-user)
 - Detect and fix known issues
 - User communication
- Quite some big campaigns:
 - KVM consolidation, Spectre/Meltdown and L1TF, Cold Migration, Migration to 8/9, ...

Compute service



- Using all availability zones and critical area
- Single DBoD instance hosted on a physical machine
- Standard IT service
 - VMs, puppet configured, monitored with collectd
 - Bootstrap becomes important

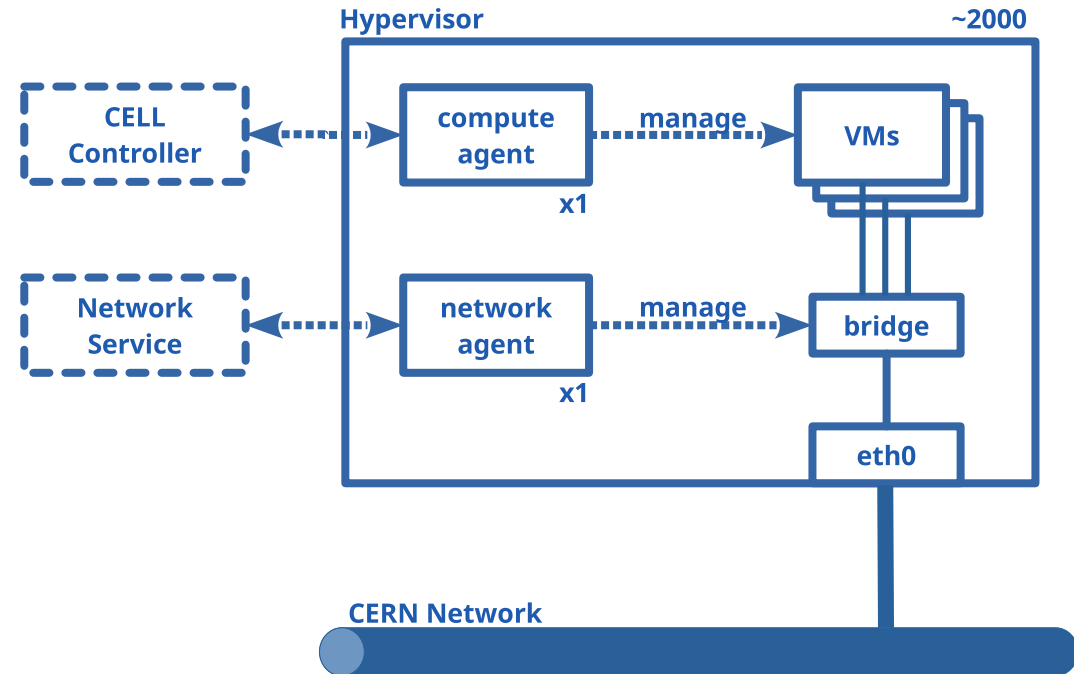
Compute service (Cell)



- Using all availability zones and critical area
- Single DBoD instance hosted on a physical machine
- Standard IT service
 - VMs, puppet configured, monitored with collectd
 - Bootstrap becomes important

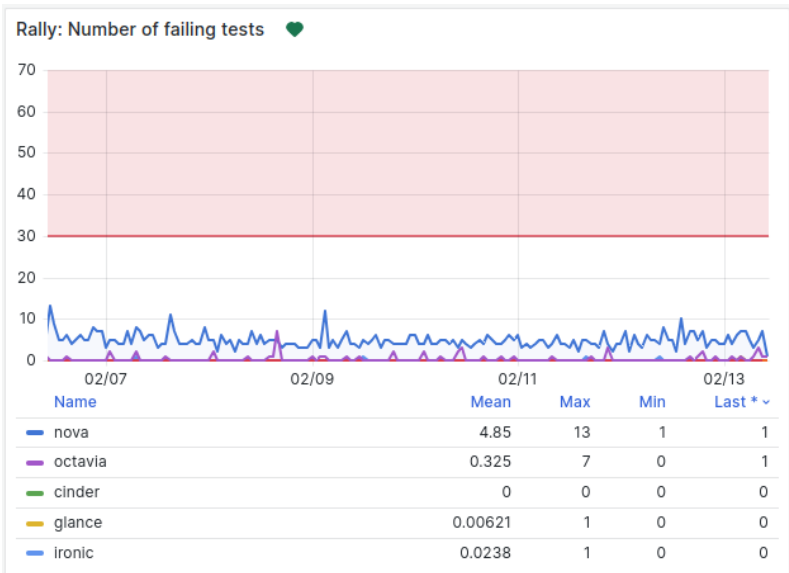
Compute service (Hypervisor)

- Standard RHEL9 / Alma9 server
- For VMs stored on local disk
 - RAID1 or RAID0 depending on profile
- Several agents that operate resources
 - Configuration untouched if APIs are down
- Bootstrap procedure exercised on every maintenance operation



Continuous probe the Cloud APIs

- Extensive use of automated probe system
- Focus on infrastructure wide issues



Passing % in time frame per availability zone

availability zone	attach-volume	boot-attach-port-d	boot-linux	boot-linux-with-dns	boot-with-attached	boot-with-ipv6read	boot-without-ipv6r	cold-migrate
cern-geneva-a	100%	100%	100%	99%	100%	100%	100%	
cern-geneva-b	100%	100%	100%	100%	100%	100%	100%	
cern-geneva-c	100%	100%	100%	100%	100%	100%	100%	
gva-critical	100%	99%	100%	100%	100%	99%	100%	

Global actions: Passing % in time frame

deployment	authenticate	boot-from-snapshot-lin	boot-from-volume-linu	boot-with-landb-prope	create-and-delete-ima	list-images
global	100%	100%	100%	87%	99%	100%
global_next	100%		98%	100%	98%	100%
global_poc	100%		96%	99%	97%	100%

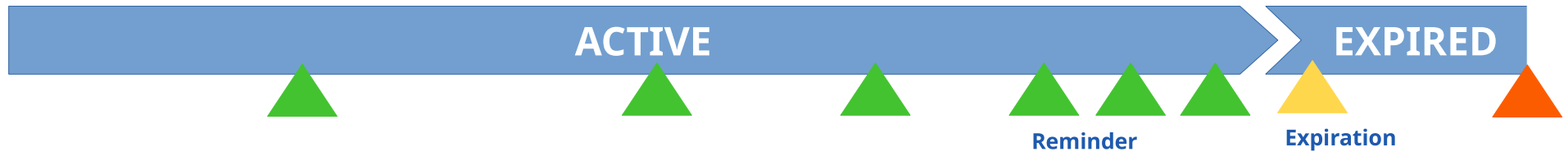
Cinder actions: Passing % in time frame

deployment	create-and-delete-snapshot	create-and-delete-volume	create-and-extend-volume	list-volumes
cinder	100%	100%	100%	100%



Optimize resource availability - Expiration

- Each VM in a personal project has an expiration date
- Set shortly after creation and evaluated daily
- Configured to 180 days and renewable
- Reminder mails starting 30 days before expiration
- Implemented on a Workbook in Mistral



Task delegation

- Rely on Rundeck for offloading tasks to different teams
 - Repair Team
 - Resource coordinator
 - Cloud operations
- Example: disk replacement

