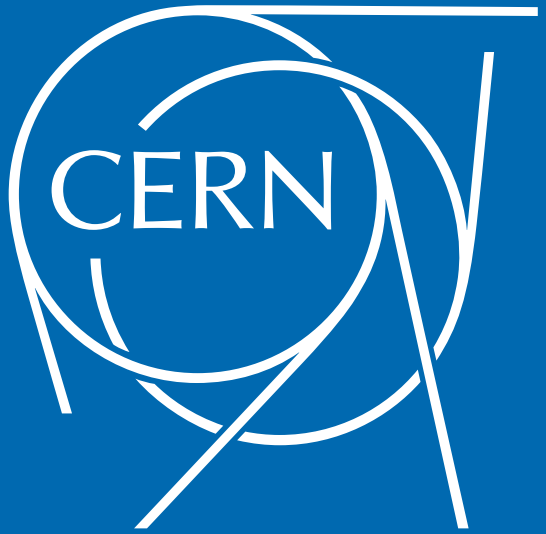

CERN

XXV International Symposium on Nuclear Electronics & Computing

28th Sep – 2nd Oct 2015, Budva, Becici, Montenegro

CERN LHC Run 2 on OpenStack

Sebastian Bukowiec

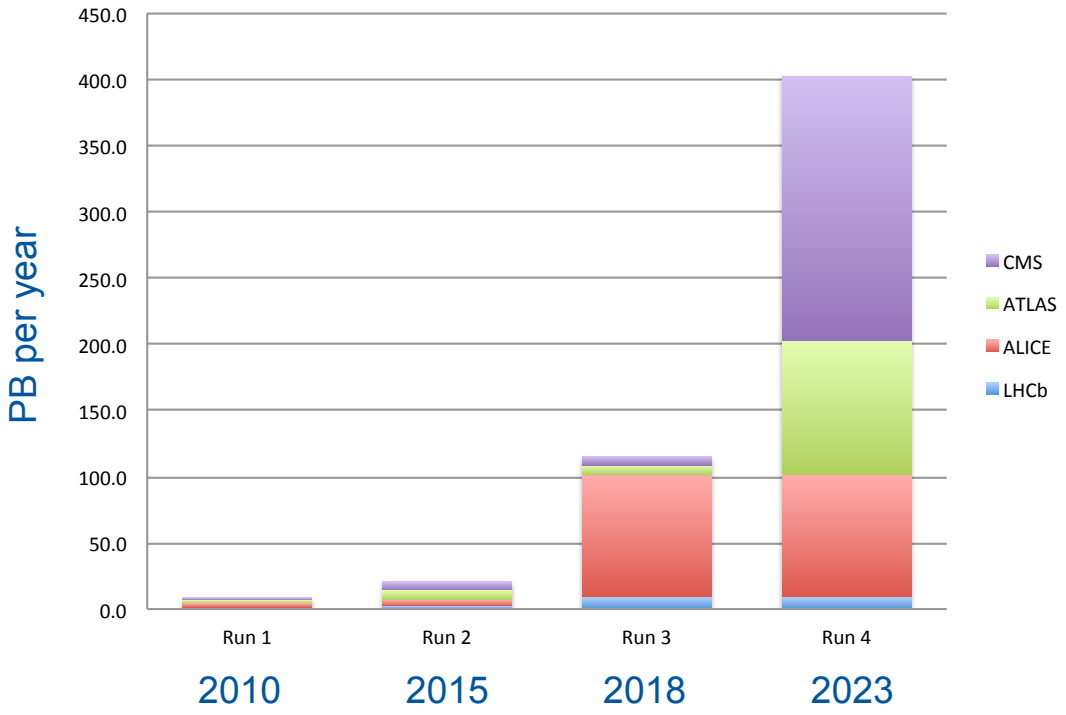
CERN Cloud Infrastructure

LHC Run 1 (2010-2013)

- The computing challenge resulted in a great success
- ~65 PB of data produced
- Analyzed in record time proving a theory proposed by Professors Higgs and Englert in 1960s

Source: http://www.uniovi.es/en/-/higgs-englert-y-el-cern-explican-en-la-universidad-de-oviedo-el-camino-hasta-el-descubrimiento-del-boson

Planned LHC data growth



- Estimating 400PB/ year by 2023
- Compute needs expected to be around 50x current levels if budget available

Infrastructure growth

- Large increase of computing resources required from 2015
- Constrained budgets
- The same amount of personnel

Enforce to evolve towards more dynamic, efficient and flexible systems.

CERN's Toolset

Agile Infrastructure project aimed to identify new tools needed to build CERN Cloud Infrastructure and enhance communication within IT department

Configuration Tools | **Cloud Software** | **Monitoring Tools** | **Storage Solution**

CERN's Toolset

Agile Infrastructure project aimed to identify new tools needed to build CERN Cloud Infrastructure and enhance communication within IT department

Configuration Tools | **Cloud Software** | **Monitoring Tools** | **Storage Solution**

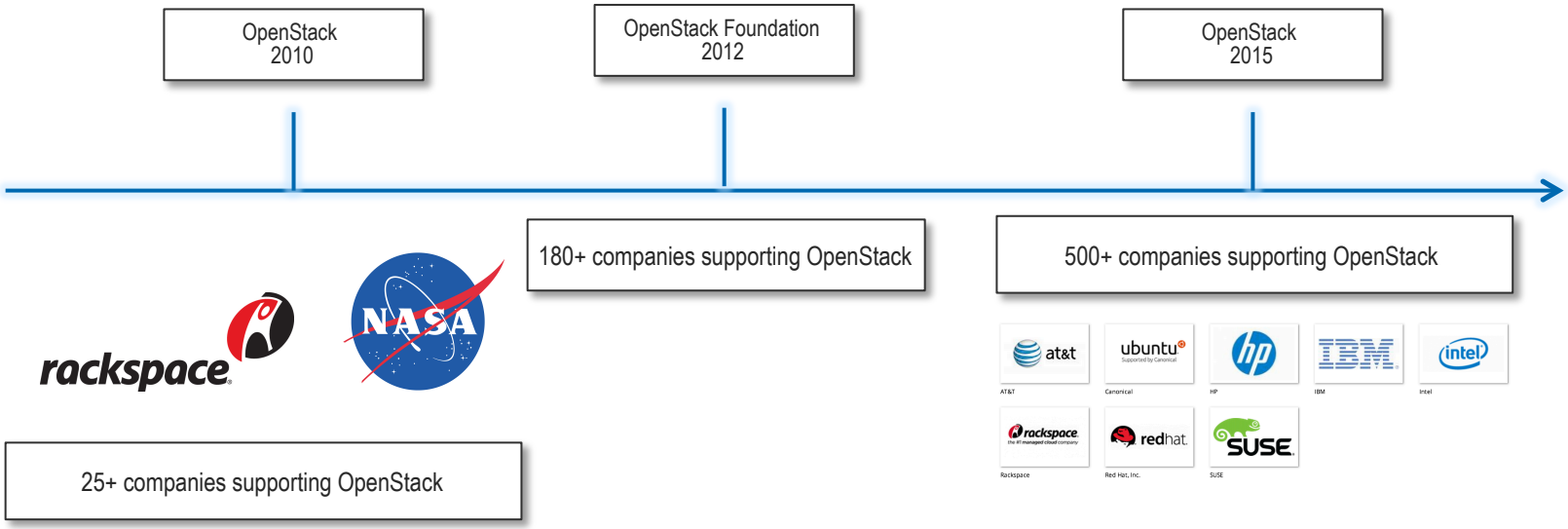
OpenStack

OpenStack software controls large pools of compute, storage, and networking resources throughout a datacenter, managed through a dashboard or via the OpenStack API. OpenStack works with popular enterprise and open source technologies making it ideal for heterogeneous infrastructure.

-- OpenStack

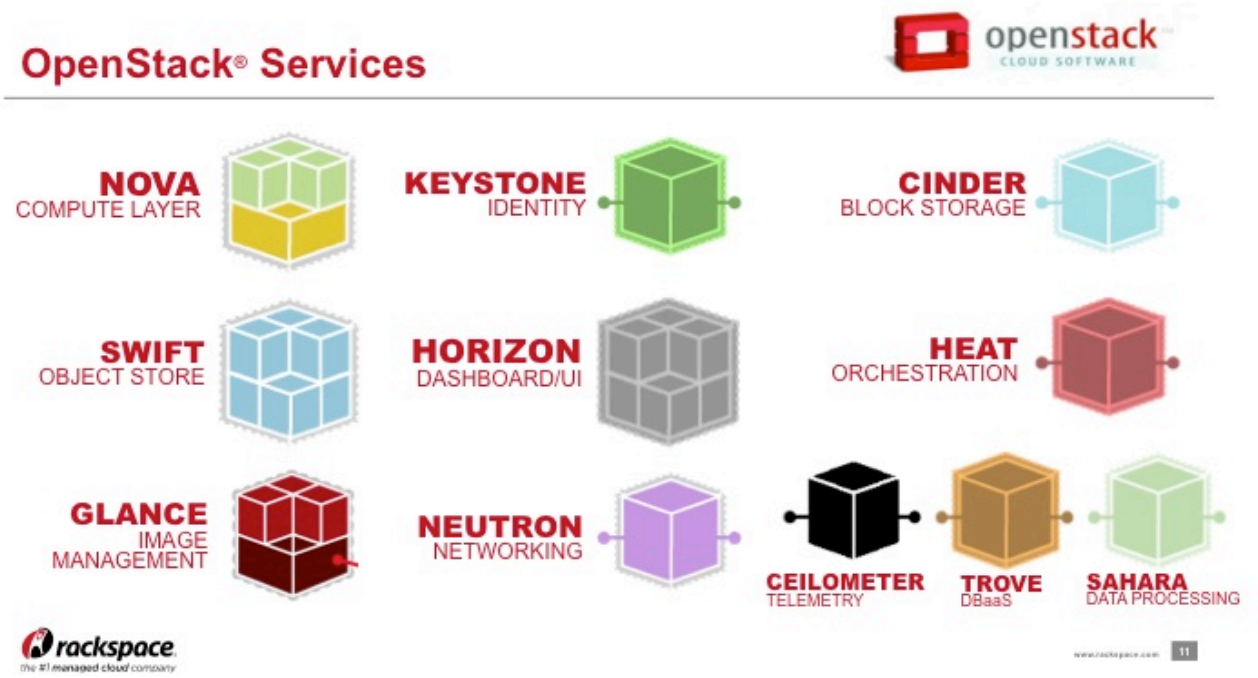
http://www.openstack.org

OpenStack beginning

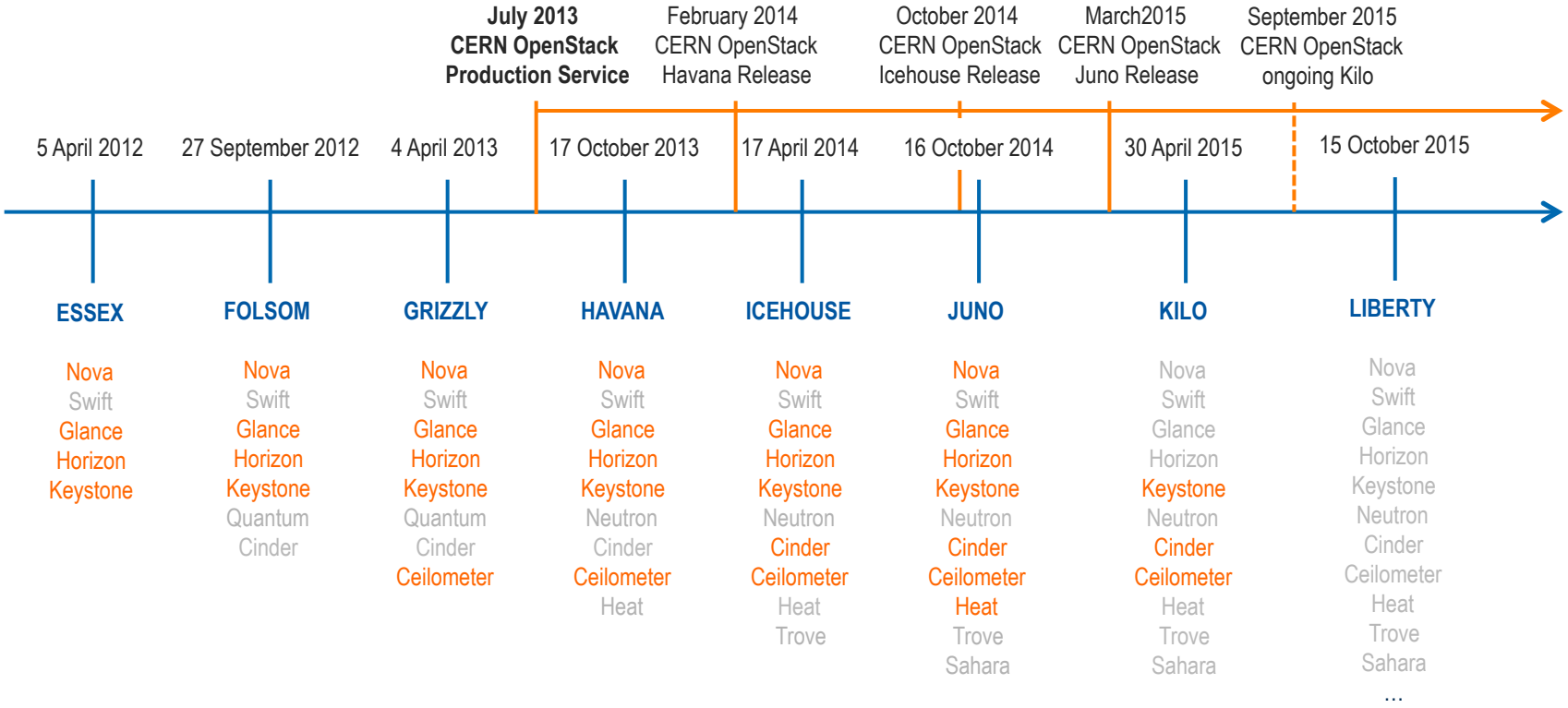



Gold Members
Corporate Sponsors
Aptira
CCAT
Cisco
Dell
DreamHost
DataCentred
Dualtec
EasyStack
Elastx AB
Emulex Corporation
Pure Storage
QCT (Quanta Cloud Technology)
SAP
Scality
eNovance
Enter Cloud Suite
F5 Networks
FalconStor
Flexiant
Servergy
Servosity
SI-TECH
SolidFire
Flextronics
Fujitsu
Fusion-io
GoDaddy
hastexo
StackIQ
StackLabs
StackStorm
StackVelocity
Infinidat
Infoblox
Inktank
Innova
Inspur
Stratus Technologies
Tata Communications
Tesora
teuto.net
A10 Networks
Accelerite
Adaptive Computing
Akanda
Alcatel Lucent
Internap
InWinSTACK
IO
iWeb
KIO Networks
Tintri
TransCirrus
Ultimum Technologies
UnitedStack Inc.
Anchor
Anuta Networks
AQORN
Arista Networks
AURO
Kontron
Lenovo
Linux Foundation
Midokura
NetScout
VEXXHOST, Inc.
Violin
Virtual Bridges
VMware
Avaya
Bit-isle
Blue Box
Breqwatr
Brocade
Nexenta
Nexus, a Dimension Data Co
Nokia Networks
Nubeliu
Numergy
Wind River
ZTE Corporation
Seagate
Thales Group
Bull
City Network
Cliqr
Cloud Enabled
Cloudwatt
Onyx CCS
Oracle
Orange
OVH Group
Pactera
Solinea
VCE
Comcast
Coraid
Cray
Cumulus Networks
Cyan Inc.
Parallels
PayPal
Platform9 Systems, Inc
PLUMgrid
Produban
Stratoscale
Western Digital Corporation

OpenStack services



CERN OpenStack project timeline



Strategy to deploy OpenStack at CERN

- Configuration infrastructure based on Puppet
- Community Puppet modules for OpenStack
- RDO - RPM Packages
- CERN CentOS 7 (CC7), Microsoft Windows 2012 R2 Operating Systems
  - Scientific Linux CERN 6 (SLC6) during phase out

Strategy to deploy OpenStack at CERN

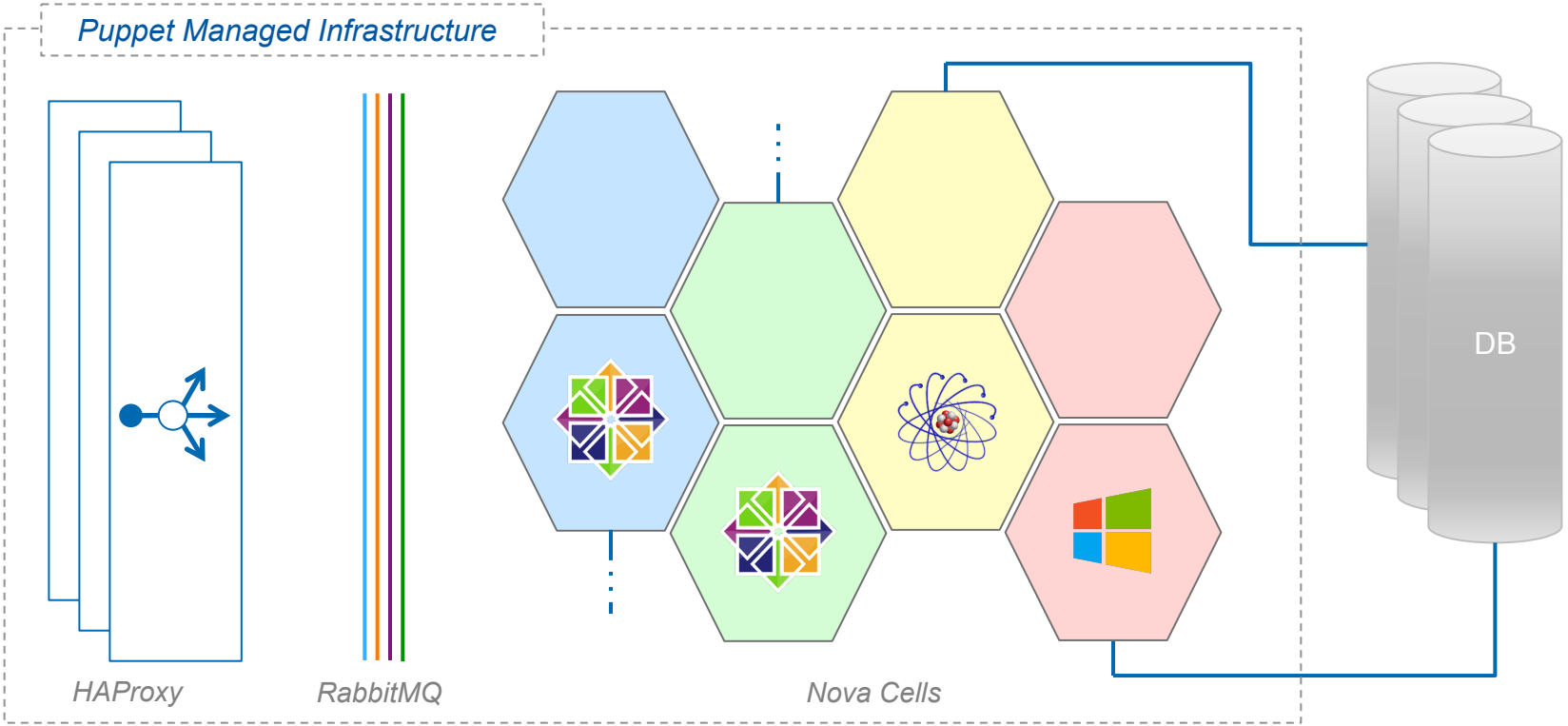
- Wigner - Budapest Computer Center hardware deployed as OpenStack compute nodes

Size of CERN Cloud Infrastructure

Compute Nodes	5000
Virtual Machines	15 000
Cores	130 000
Tenants	2216

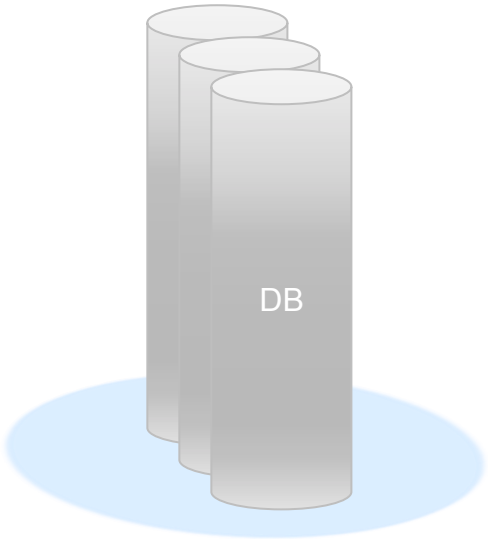
continuously growing …

Infrastructure Overview



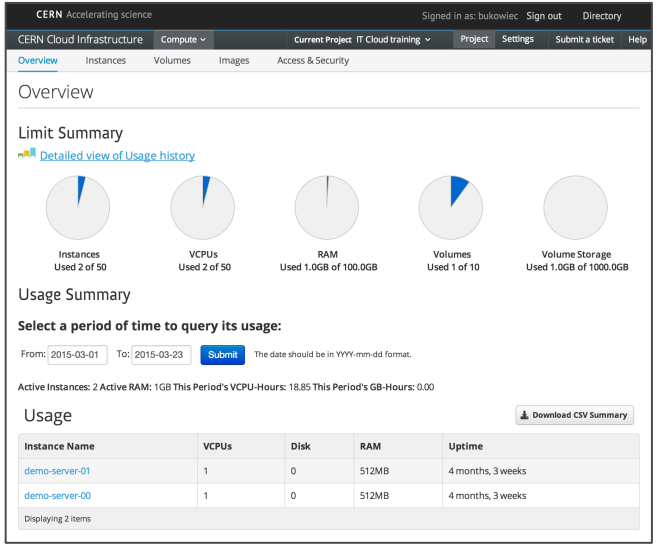
Infrastructure Overview

- mySQL instance per Cell
  - provided by CERN DB team (DB on demand service)
  - Running on top of Oracle CRS
  - NetApp storage backend
  - Backups every 6 hours



Three interfaces

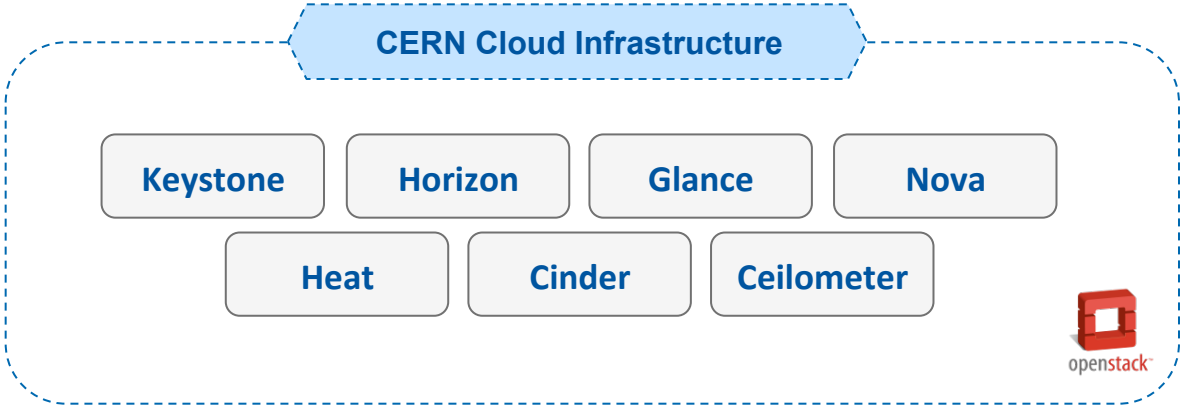
```
$ openstack server list
+--------------------------------------+----------------+--------+-------------------------------------------------------+
| ID                                   | Name           | Status | Task State | Power State | Networks                         |
+--------------------------------------+----------------+--------+-------------------------------------------------------+
| b30d60b3-e4a2-43a1-8889-e022a0c12f20 | demo-server-00 | ACTIVE | CERN_NETWORK=188.XXX.XX.XXX, 2001:1458:XXX:XX::XXX:XXX  |
| f79a4a90-9c96-4af8-af94-e89cc35f2fe7 | demo-server-01 | ACTIVE | CERN_NETWORK=188.XXX.XXX.XXX, 2001:1458:XXX:XX::XXX:XXX |
+--------------------------------------+----------------+--------+-------------------------------------------------------+
```




```
$ curl -s -X POST http://8.21.28.222:5000/v2.0/tokens \
               -H "Content-Type: application/json" \
               -d '{"auth": {"tenantName": "", "passwordCredentials":
               {"username": "'"$OS_USERNAME"'", "password": "'"$OS_PASSWORD"'"}}}' \
               | python -m json.tool


$ curl -i -H "X-Auth-Token:token" http://8.21.28.222:8774/v2/tenant_id/servers
```

CERN OpenStack projects

- Modular architecture
- Designed to easily scale out

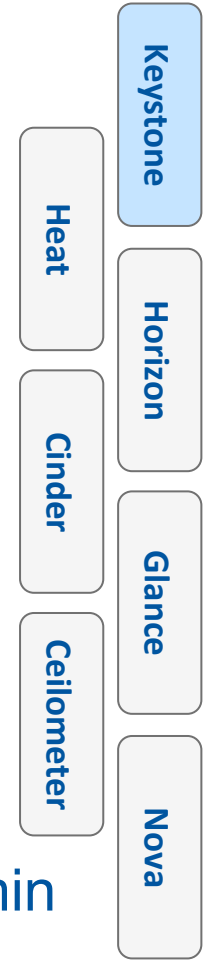


Keystone (Identity)

- Main functions:
  - Identity: user authentication
    Keystone is integrated with CERN AD via LDAP backend

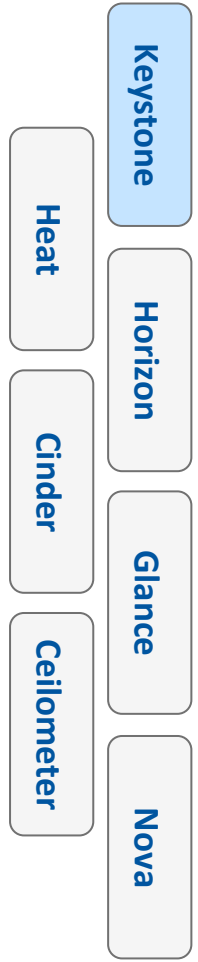
    CERN Active Directory infrastructure:
    - unified identity across the site
    - 44 000 users
    - ~200 arrivals/departures per month

  - Policies: enforces different user levels: Member, Admin



Keystone (Identity)

- Main functions:
  - Token: Provides token to communicate with other OpenStack components
  - Service catalog: service endpoints registered with keystone



Keystone (Identity)

- CERN user subscribes the "cloud service”
  - Created "Personal Tenant" with limited quota
  - ~ 2000 users
- Shared projects created by request
  - ~300 shared
- Project life cycle integrated with Microsoft Forefront Identity Manager (FIM) to be compliant with standard CERN policies

Nova (Compute)

Boots and manages the lifecycle of virtual machine on CERN supported hypervisors:

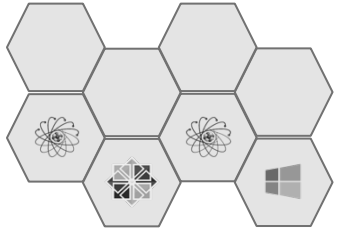
- KVM (Linux OS VMs)
- Hyper-V (Windows OS VMs)

Hypervisor selection is based on “Image” properties.

Nova (Compute)

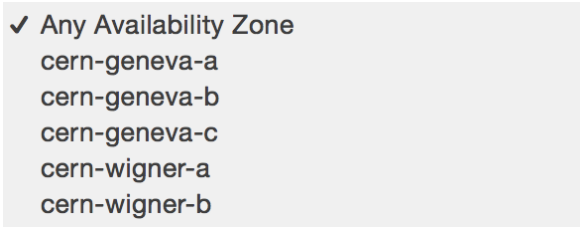
- Set of compute nodes equipped with hybrid storage for IO intensive workflow
- Various flavors matching ~20 different types of Hardware
- ~5000 compute nodes
  - vast majority QUEMU/KVM (~150 Hyper-V)
  - ~220 HVs on critical power
  - ~800 HVs at Wigner in Hungary
  - ~2000 HVs used by batch - “cattle” use case
  - rest shared by users, services, experiments - “pets” use case

Nova Cells

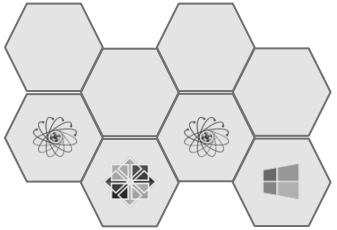


Cells are used in order to scale the infrastructure, better fault tolerance and for project distribution.

- Allow to scale transparently between different Computer Centers
- Mapped to the availability zones

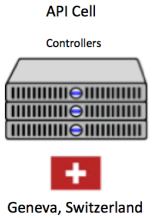


Nova Cells



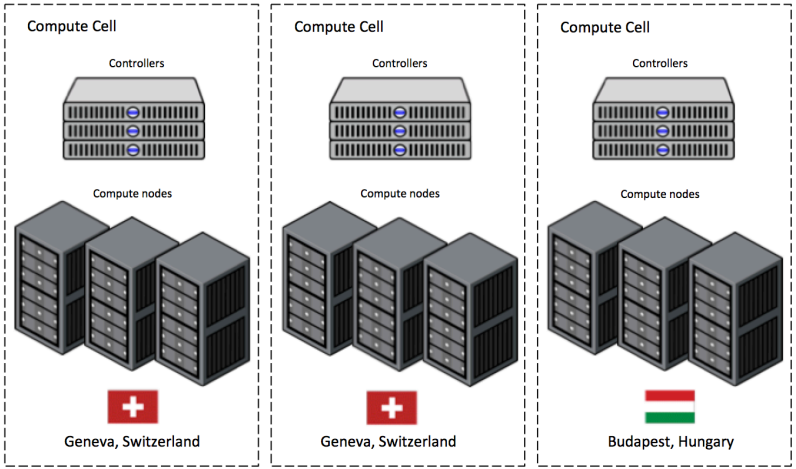
Top level cell

- Runs API service
- Top cell scheduler



Child cells run:

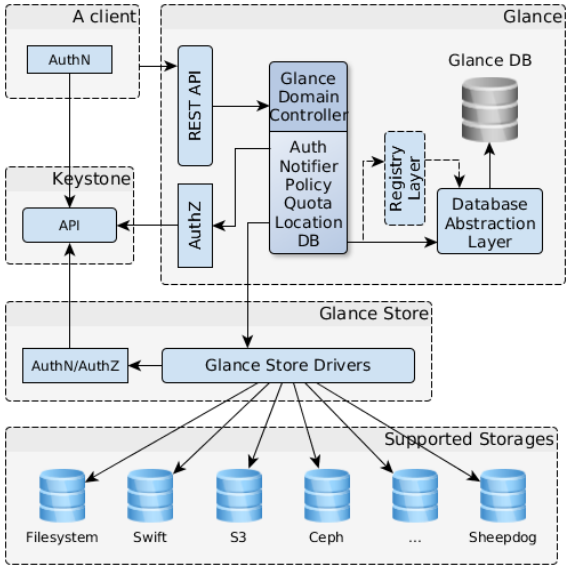
- Compute nodes
- Nova network
- Scheduler
- Conductor



Glance (Image service)

Stores and retrieves VMs images (virtual machines templates)

- Glance backend - Ceph
  (currently ~28 TB of images)
- Snapshot capabilities

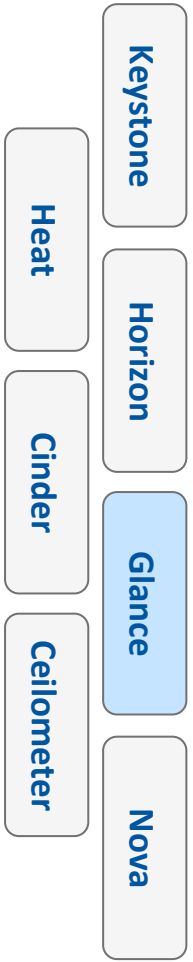


Source: http://docs.openstack.org/developer/glance/architecture.html

Keystone
Horizon
Glance
Nova
Cinder
Ceilometer
Heat

Glance (Image service)

- Maintain a small set of CERN images as default
  - Difficult to offer only the most updated set of images
  - Resize and Live Migration not available if image is deleted from Glance
- Users can upload their own images
  - Users don’t pay for storage
  - No quotas per Tenant



CERN Images

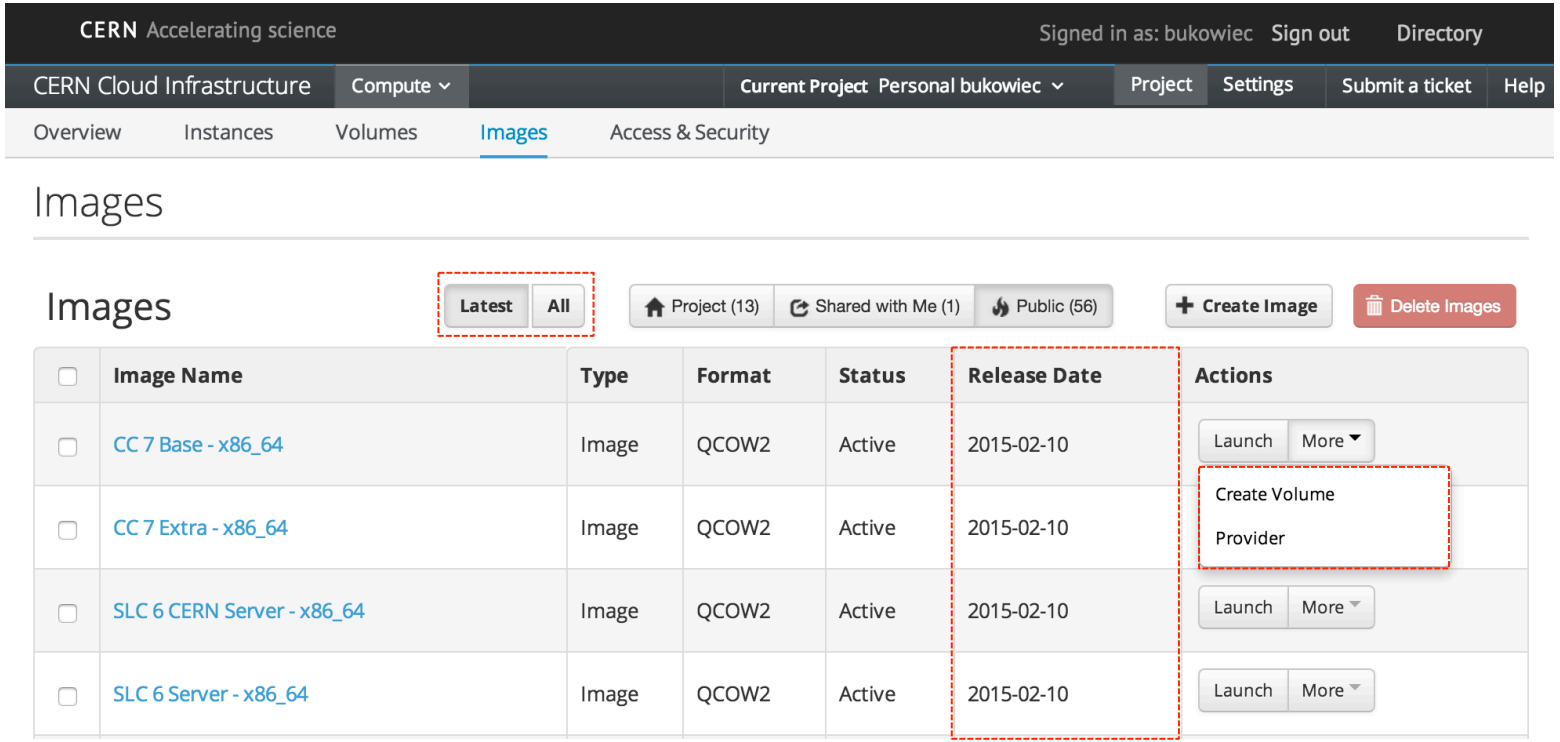
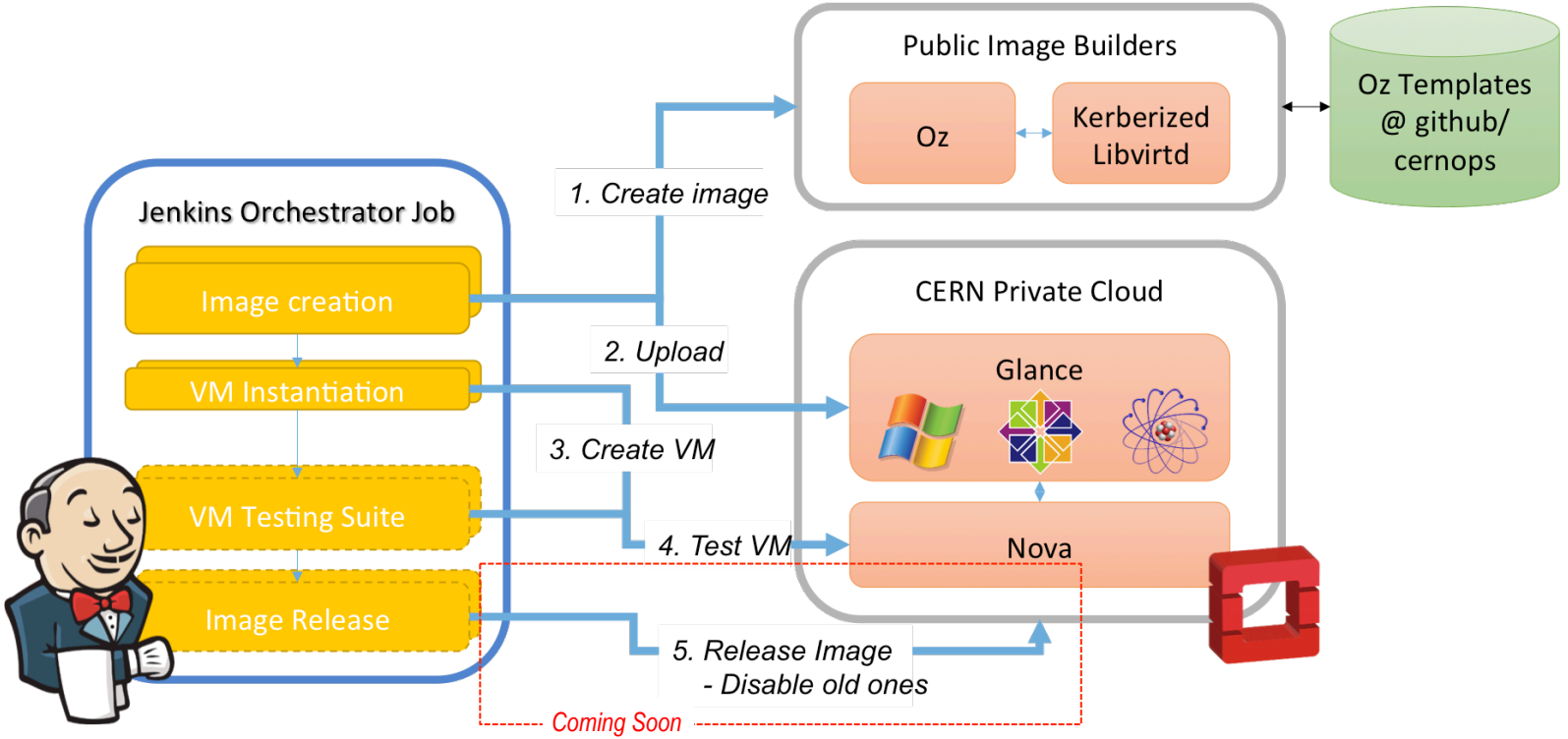

CERN Accelerating science
Signed in as: bukowiec Sign out Directory
CERN Cloud Infrastructure
Compute
Current Project Personal bukowiec
Project
Settings
Submit a ticket
Help
Overview
Instances
Volumes
Images
Access & Security
Images
Images
Latest
All
Project (13)
Shared with Me (1)
Public (56)
Create Image
Delete Images
Image Name
Type
Format
Status
Release Date
Actions
CC 7 Base - x86_64
Image
QCOW2
Active
2015-02-10
Launch
More
CC 7 Extra - x86_64
Image
QCOW2
Active
2015-02-10
Create Volume
Provider
SLC 6 CERN Server - x86_64
Image
QCOW2
Active
2015-02-10
Launch
More
SLC 6 Server - x86_64
Image
QCOW2
Active
2015-02-10
Launch
More

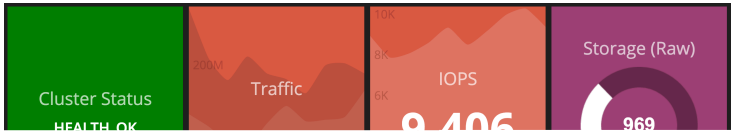
Image Lifecycle

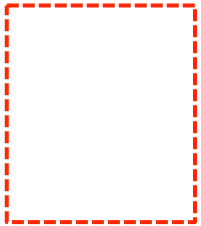


Cinder (Block Storage)

Provides persistent disks for virtual machines

- Volumes for Linux and Windows VMs
- Utilizes Ceph and NetApp as storage backends
- Various volume types

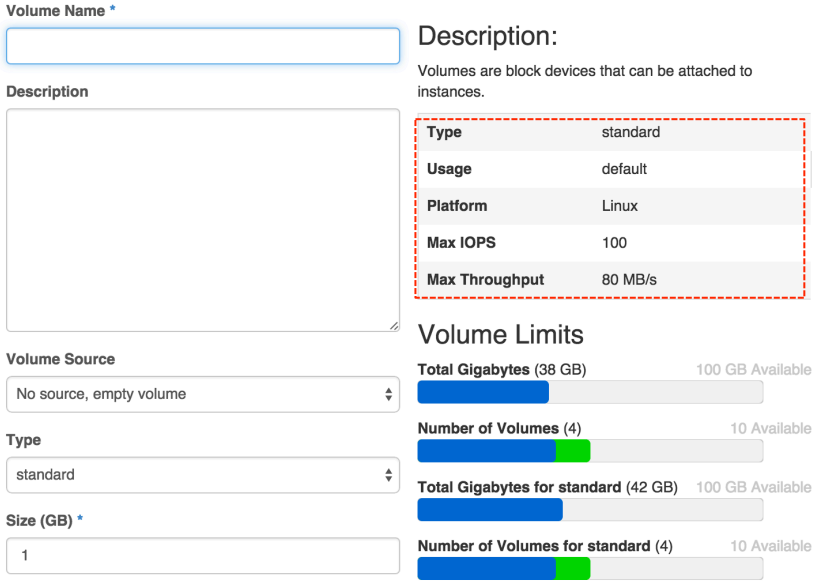


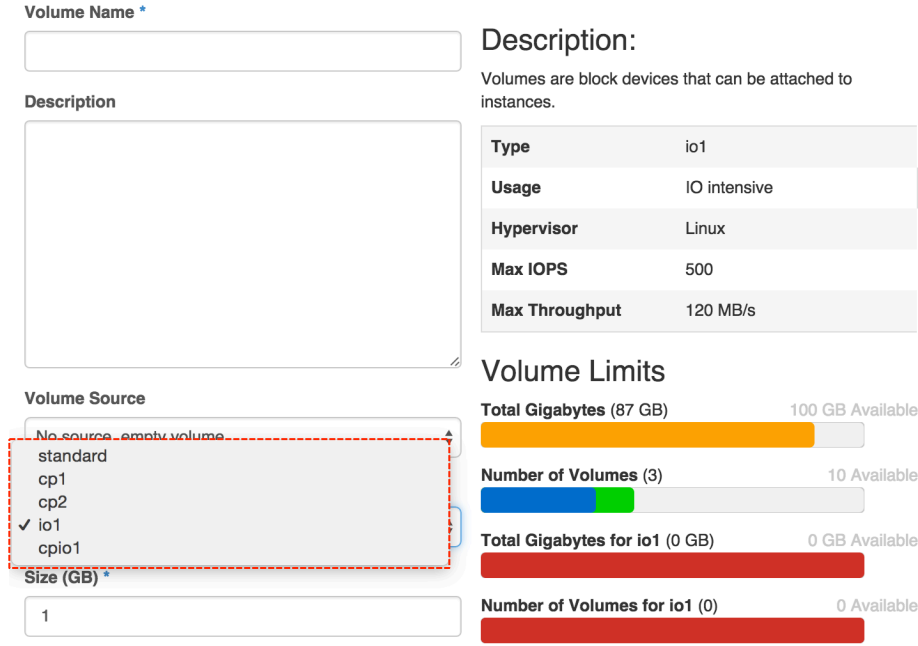


24/09/2015 09:50

Keystone
Heat
Horizon
Cinder
Glance
Ceilometer
Nova

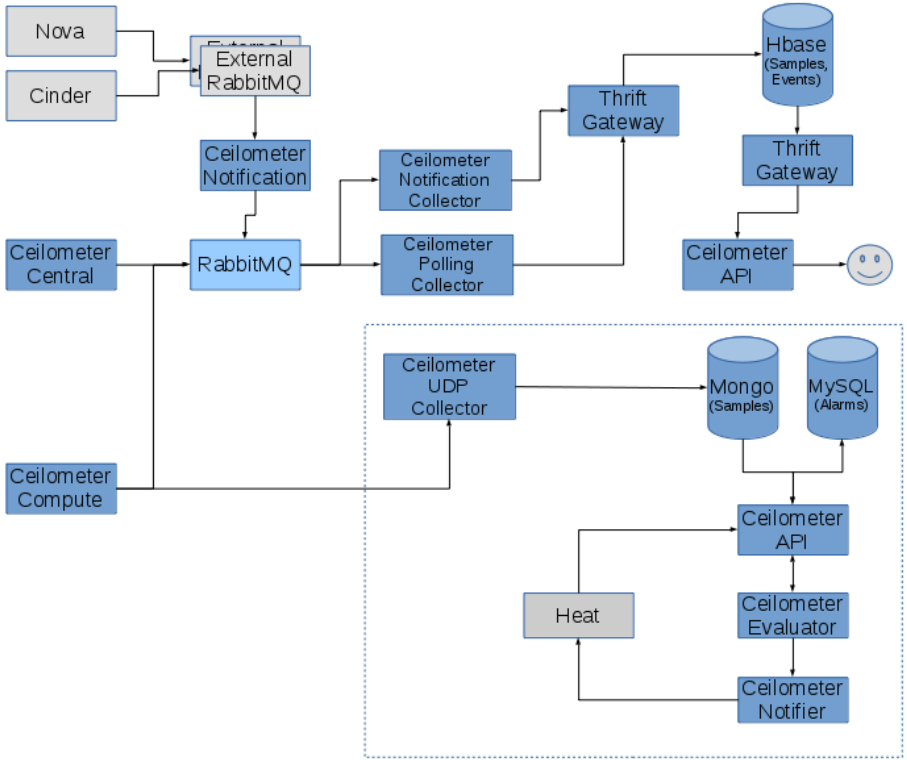
Cinder Volumes





Ceilometer Architecture

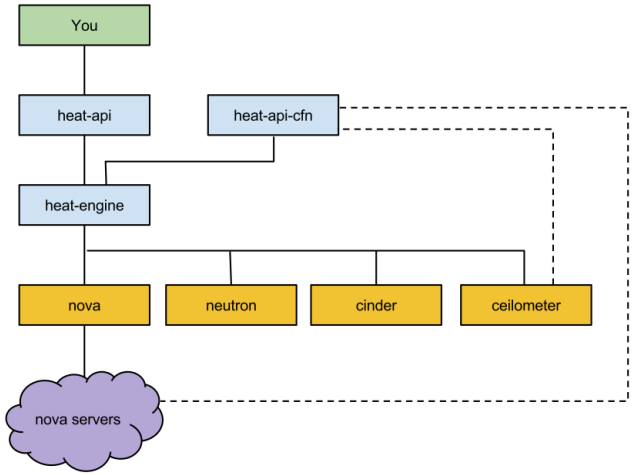
- Monitors, collects and stores usage data for all OpenStack Infrastructure
- Provides data used in accounting
- Used for Heat auto scaling scenarios



Keystone
Heat
Horizon
Cinder
Glance
Ceilometer
Nova

Heat (Orchestration)

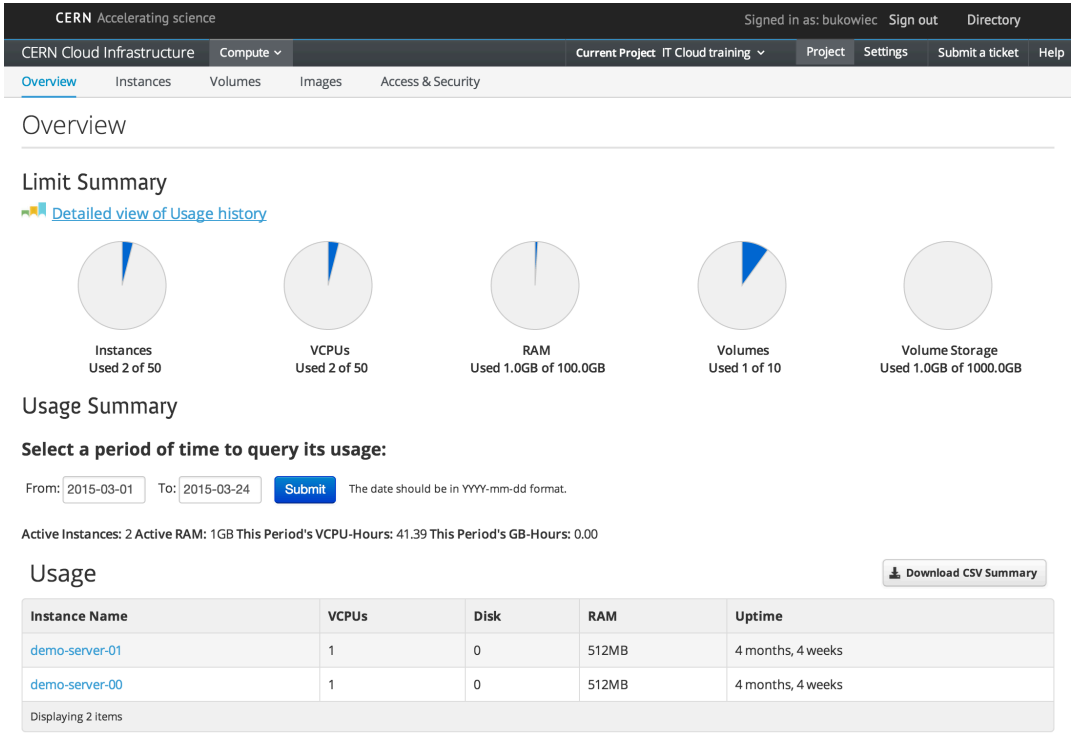
Heat provides a mechanism for orchestrating OpenStack resources through templates




```
heat_template_version: 2013-05-23

description: Simple template to deploy a single compute instance

parameters:
  key_name:
    type: string
    label: Key Name
    description: Name of key-pair to be used for compute instance
  image_id:
    type: string
    label: Image ID
    description: Image to be used for compute instance
  instance_type:
    type: string
    label: Instance Type
    description: Type of instance (flavor) to be used

resources:
  my_instance:
    type: OS::Nova::Server
    properties:
      key_name: { get_param: key_name }
      image: { get_param: image_id }
      flavor: { get_param: instance_type }
```

Horizon (Dashboard)

Provides
a self-service
web based
user interface
to OpenStack
services
for end users.



Keystone
Heat
Horizon
Cinder
Glance
Ceilometer
Nova

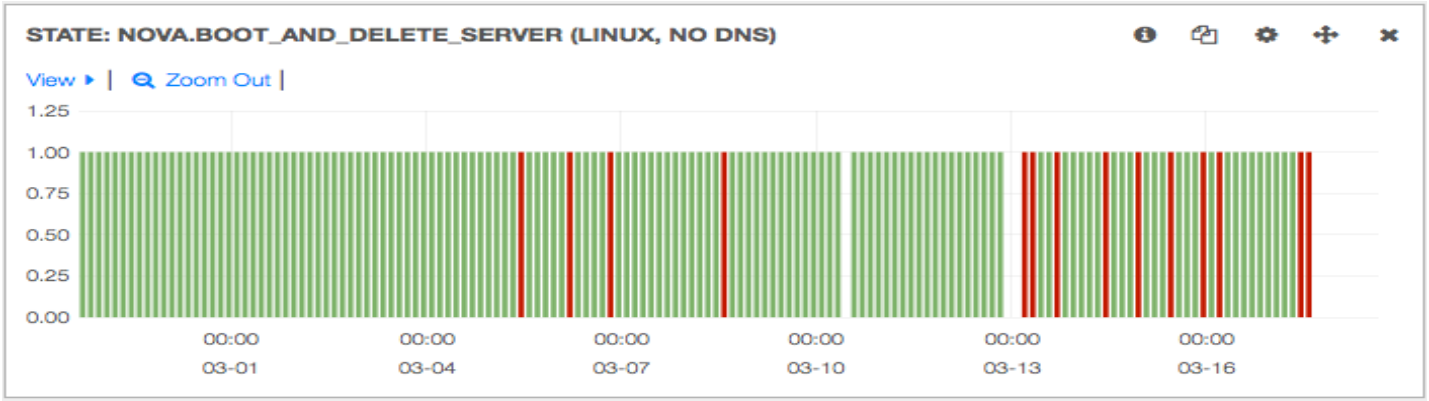
Automation

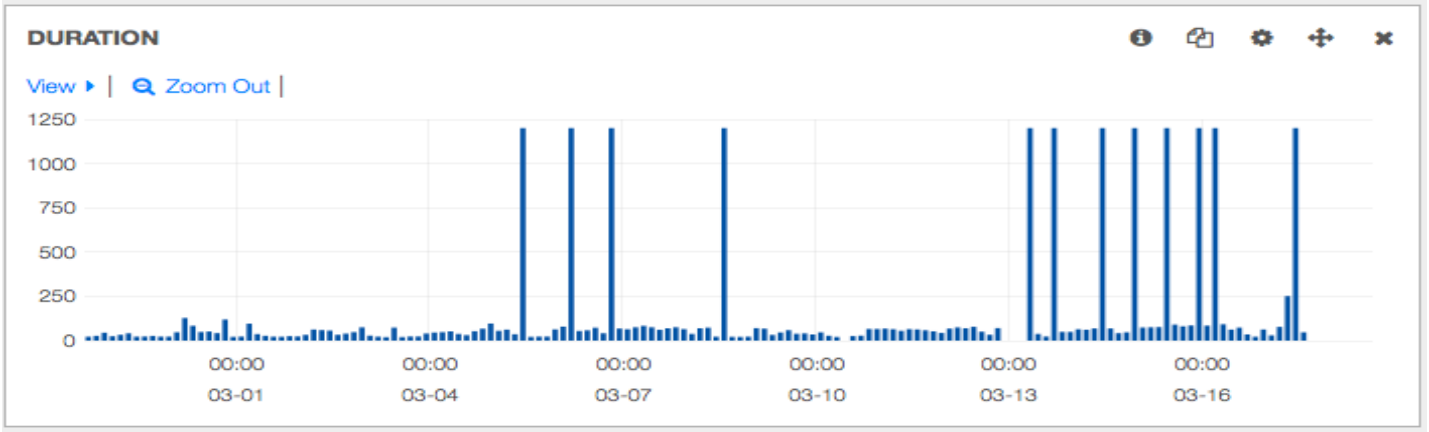
RUNDECK

- Friendly and easy interface from where we can organize and launch jobs on our hosts
- Sharing of sensitive tasks to other groups without exposing credentials or procedures
- Operations delegation:
  - **SysAdmins**: Workflows related to hypervisor maintenance (h/w intervention, notify users…)
  - **Cloud-Operations**: Project creation, Health reports, Quota update

Operations - Rally

- Benchmarking tool for OpenStack
- Performance test
- Cloud verification
- Used for OpenStack Continuous Integration
- Check if services work correctly
  - Rally runs against QA and Production environments regularly
  - We can compare results between the environments.





Heat
Keystone
Cinder
Horizon
Ceilometer
Glance
Rally
Nova

OpenStack Federations

Evaluating cloud identity federation available in OpenStack ecosystem that allows for on premise bursting into remote clouds with use of local identities (i.e. domain accounts).

A way to Hybrid Cloud architectures - virtualized infrastructures layered across independent private and public clouds.

Challenges

- Rapidly growing infrastructure
- Hardware lifecycle
- Neutron deployment
- Federations
- Keeping pace, rolling new updates every six months
- Integration of LXC containers into infrastructure – internal compute use cases

OpenStack at CERN – Summary

- OpenStack scales and meets our needs for LHC Run 2
- Geneva and Budapest Computer Centers run on OpenStack
- Cells deployment allows us to scale out easily adding new groups of nodes
- Ceph backend for images and volumes, NetApp backend for Windows volumes
- Working with upstream for new features and bug fixes

OpenStack at CERN

- CERN code publicly available on GitHub:
  https://github.com/cernops

- Hints and tips from the CERN OpenStack cloud team
  http://openstack-in-production.blogspot.com

www.cern.ch