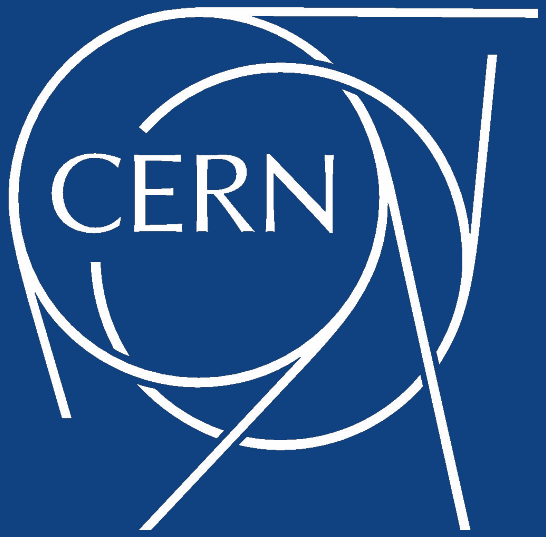

CERN

Openstack in production

Thomas Oulevey
for the CERN CM Team
IRC alphacc
@thomasnomas
RDO Day / Fosdem 2016
Bruxelles, Belgique
Jan 29, 2016

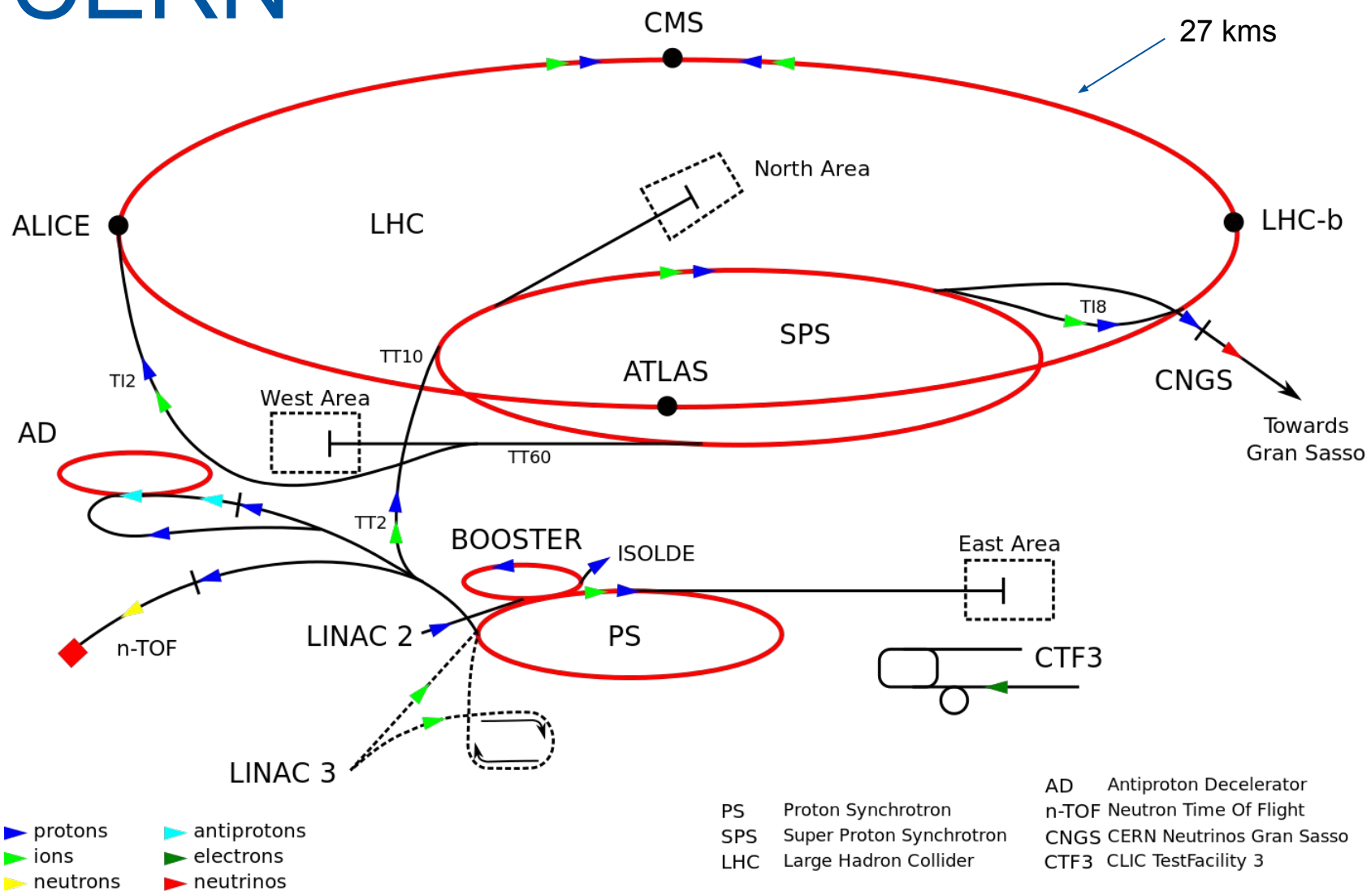
CERN

CERN Cloud

RDO

Summary

CERN


CMS
27 kms
North Area
ALICE
LHC
LHC-b
TI8
SPS
TT10
ATLAS
CNGS
TI2
West Area
Towards
Gran Sasso
AD
TT60
TT2
BOOSTER
ISOLDE
East Area
n-TOF
LINAC 2
PS
CTF3
LINAC 3
protons
ions
neutrons
antiprotons
electrons
neutrinos
PS Proton Synchrotron
SPS Super Proton Synchrotron
LHC Large Hadron Collider
AD Antiproton Decelerator
n-TOF Neutron Time Of Flight
CNGS CERN Neutrinos Gran Sasso
CTF3 CLIC TestFacility 3

CERN Cloud

- Based on RDO
  - Production service since July 2013
  - Performed four rolling upgrades since
  - Heterogenous hardware
  - KVM and Hyper-V
  - Ceph and NetAPP
  - Scientific Linux 6 and CentOS 7
  - In transition from Kilo to Liberty
  - Pets and Cattles
  - Nova, Glance, Keystone, Horizon, Cinder, Ceilometer, Rally, Heat, Neutron (WIP Magnum)

CERN Cloud Architecture (1)

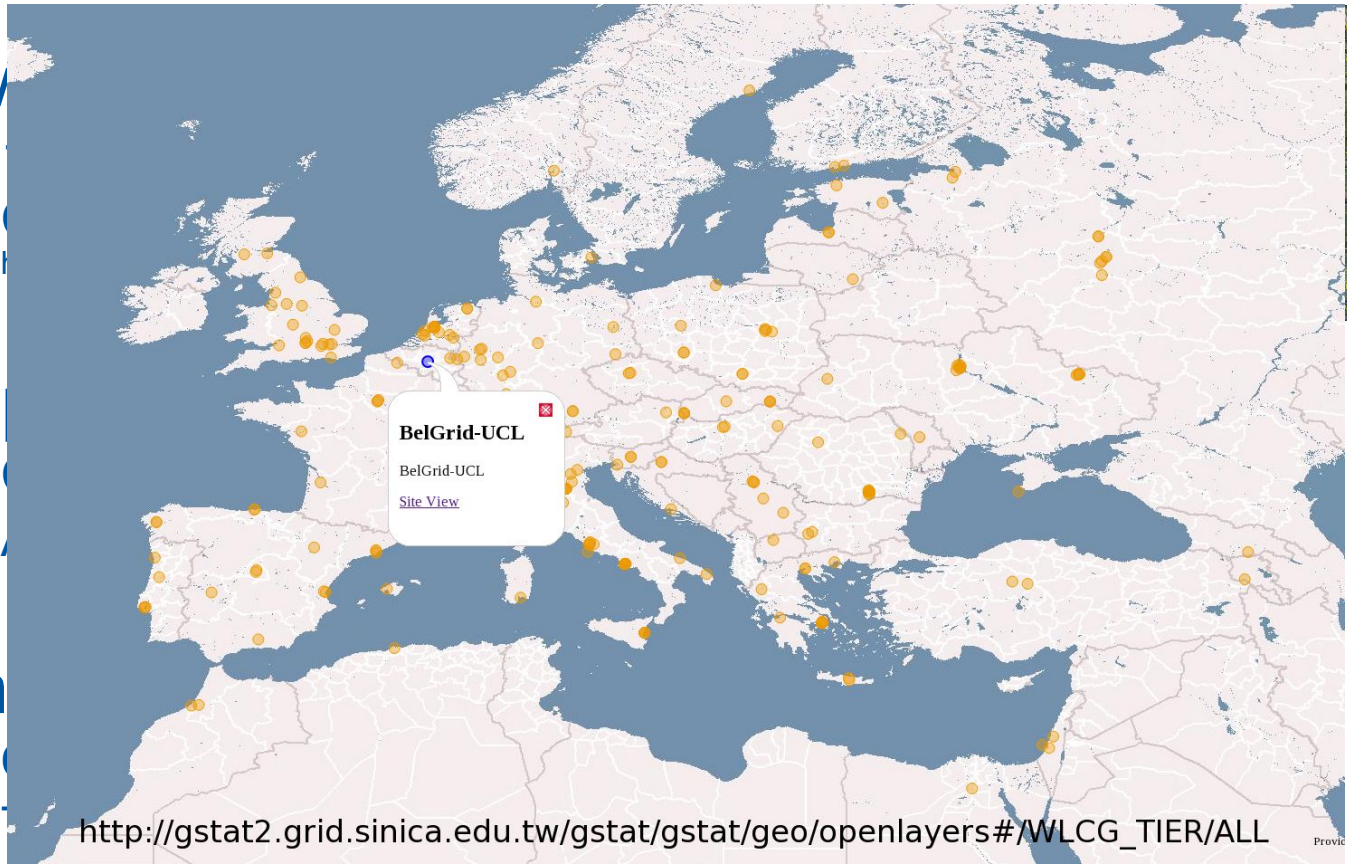
- Two data centers (Tier-0)
  - 1 region (1 API), 26 cells
  - Cells map use cases
    hardware, hypervisor type, location, users, …

- Top cell on several physical nodes in HA
  - Clustered RabbitMQ with mirrored queues
  - API servers are VMs in various child cells

- Child cell controllers are OpenStack VMs
  - **One** controller per cell
  - Tradeoff between complexity and failure impact

CERN Cloud Architecture (1)

- Tw
  -
  -
- To
  -
  -
- Ch
  -
  -



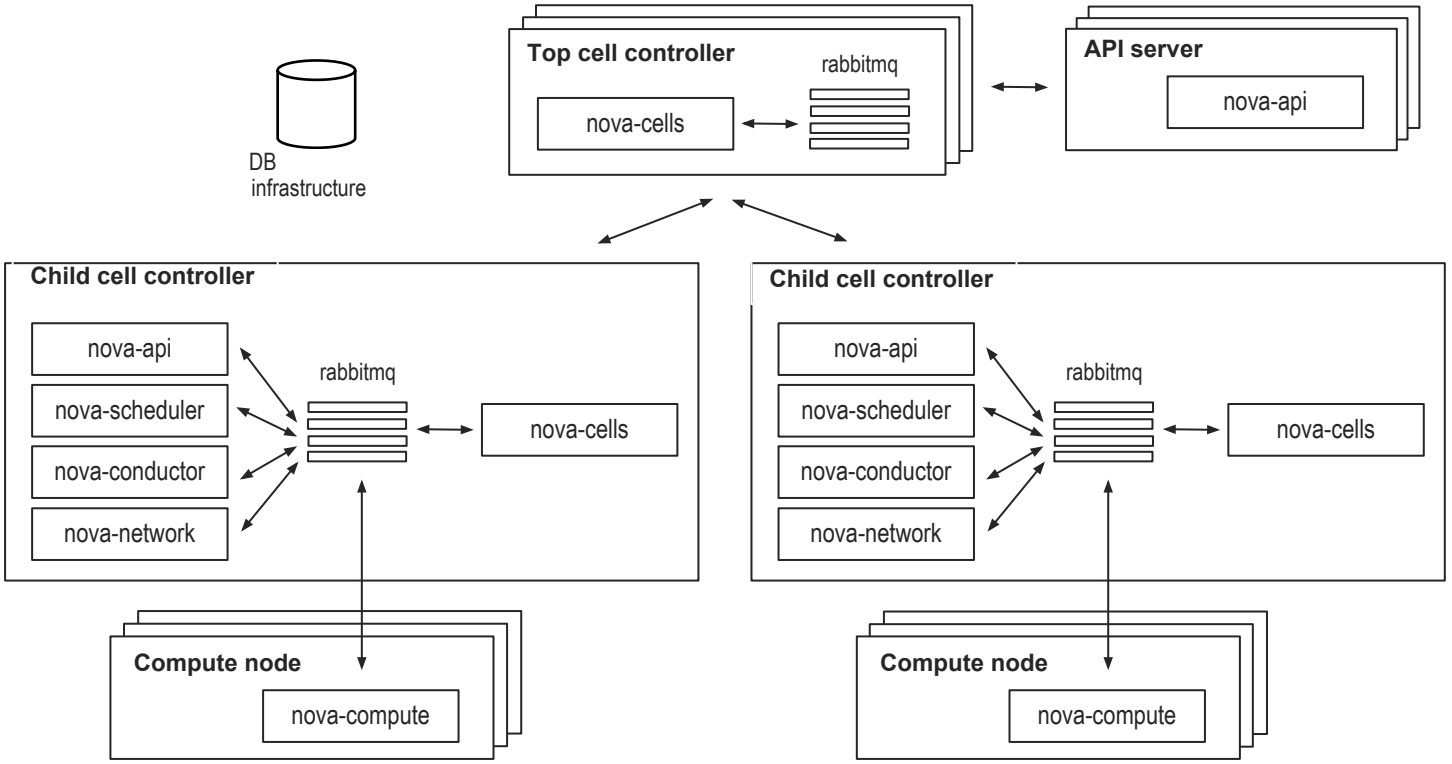
CERN Cloud Architecture (1)

- Two data centers (Tier-0)
  - 1 region (1 API), 26 cells
  - Cells map use cases
    hardware, hypervisor type, location, users, …

- Top cell on several physical nodes in HA
  - Clustered RabbitMQ with mirrored queues
  - API servers are VMs in various child cells

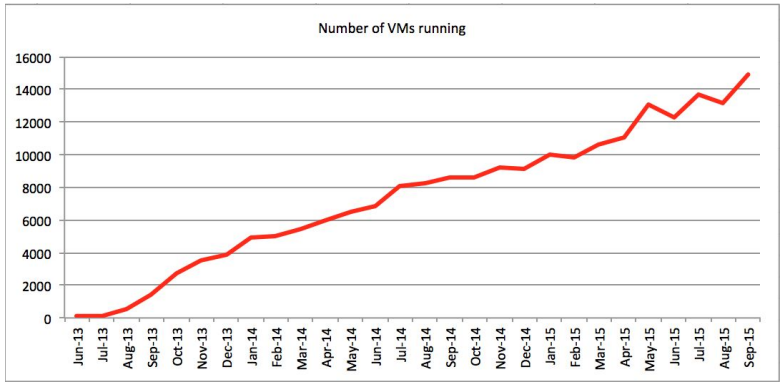
- Child cell controllers are OpenStack VMs
  - **One** controller per cell
  - Tradeoff between complexity and failure impact

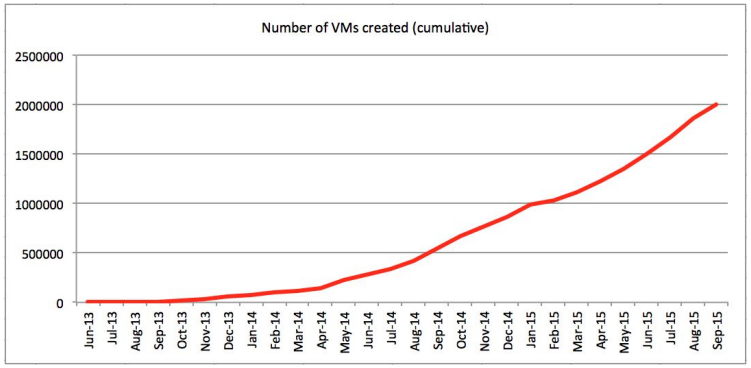
CERN Cloud Architecture (2)



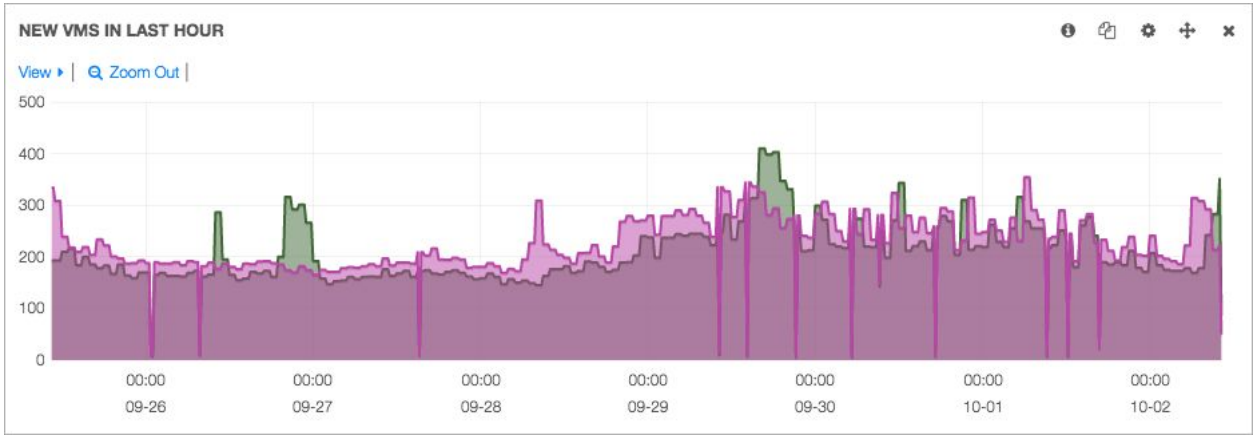
CERN Cloud in Numbers (1)

- 4’600 hypervisors in production (1y ago: 3000)
  - Majority qemu/kvm now on CC7 (~150 Hyper-V hosts) (SLC6)
  - ~2’000 HVs at Wigner in Hungary (batch, compute, services) (batch)
  - 250 HVs on critical power
- 145k Cores (64k)
- 250 TB RAM (128TB)
- ~15’000 VMs (8’000)
- To be increased in 2016!
  - +65k cores until spring





CERN Cloud in Numbers (2)



Every 10s a VM gets created or deleted in our cloud!

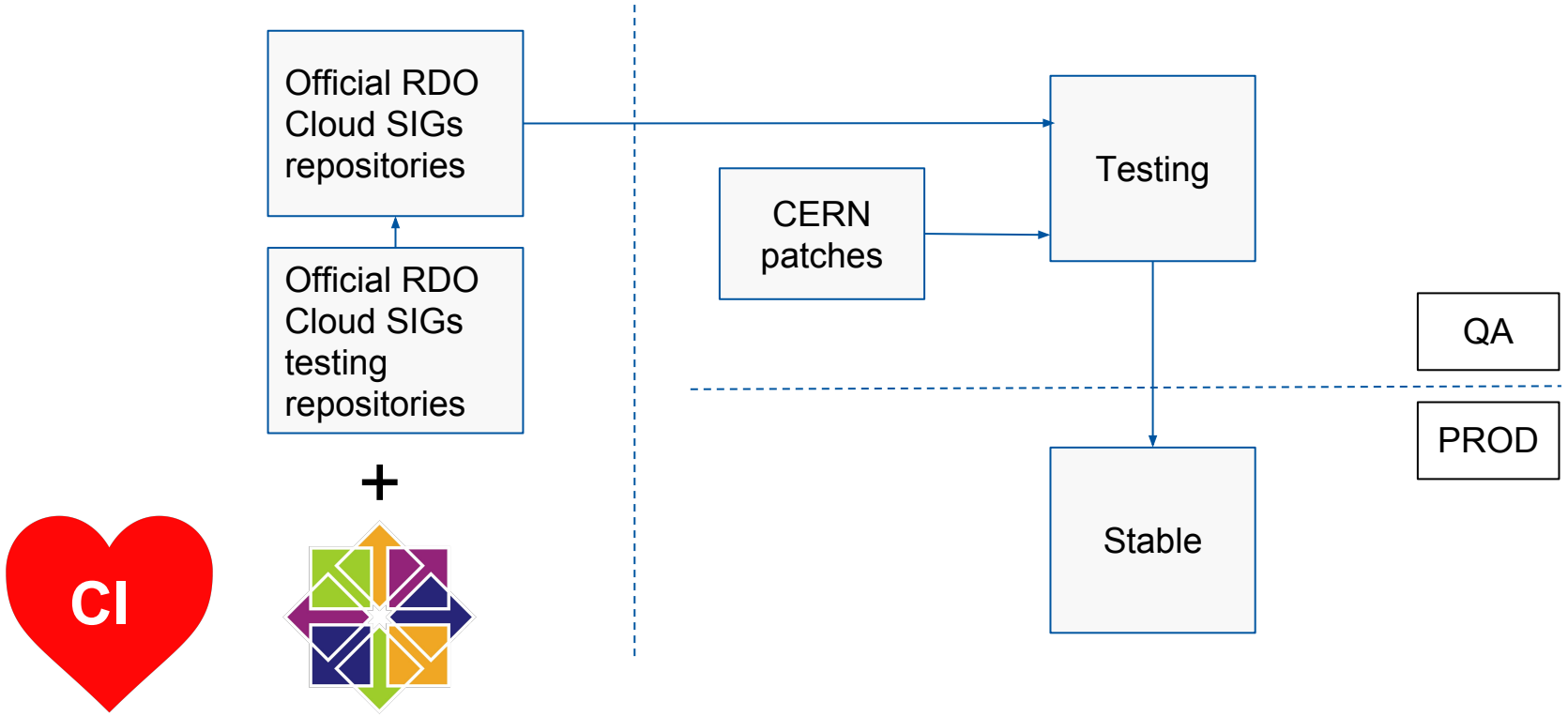
- 2’000 images/snapshots (1’100)
  - Glance on Ceph
- 1’500 volumes (600)
  - Cinder on Ceph (& NetApp)

RDO (1)

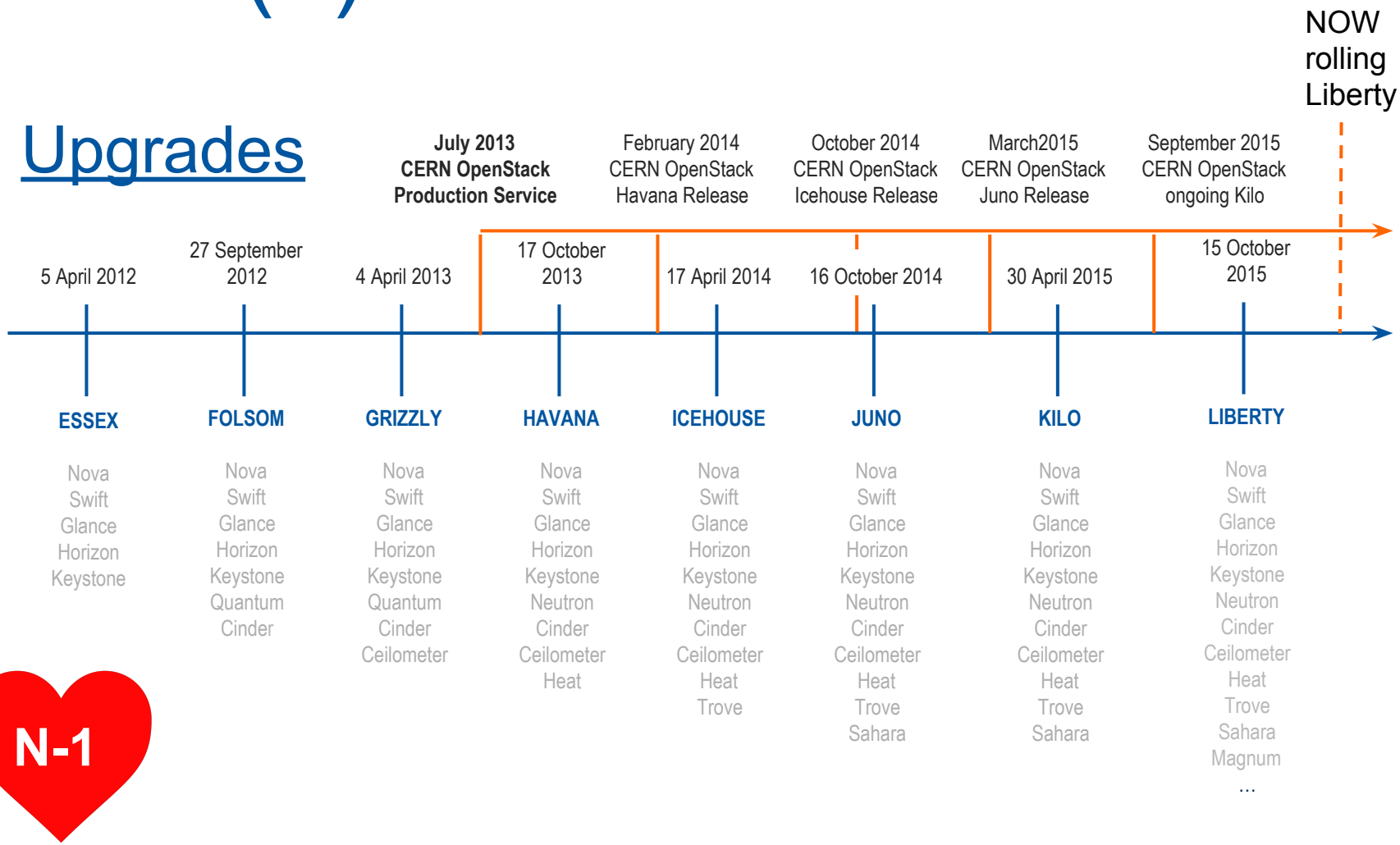
Development

RDO (2)

Software distribution



RDO (3)


NOW
rolling
Liberty
Upgrades
July 2013
CERN OpenStack
Production Service
February 2014
CERN OpenStack
Havana Release
October 2014
CERN OpenStack
Icehouse Release
March2015
CERN OpenStack
Juno Release
September 2015
CERN OpenStack
ongoing Kilo
5 April 2012
27 September 2012
4 April 2013
17 October 2013
17 April 2014
16 October 2014
30 April 2015
15 October 2015
ESSEX
Nova
Swift
Glance
Horizon
Keystone
FOLSOM
Nova
Swift
Glance
Horizon
Keystone
Quantum
Cinder
GRIZZLY
Nova
Swift
Glance
Horizon
Keystone
Quantum
Cinder
Ceilometer
HAVANA
Nova
Swift
Glance
Horizon
Keystone
Neutron
Cinder
Ceilometer
Heat
ICEHOUSE
Nova
Swift
Glance
Horizon
Keystone
Neutron
Cinder
Ceilometer
Heat
Trove
JUNO
Nova
Swift
Glance
Horizon
Keystone
Neutron
Cinder
Ceilometer
Heat
Trove
Sahara
KILO
Nova
Swift
Glance
Horizon
Keystone
Neutron
Cinder
Ceilometer
Heat
Trove
Sahara
LIBERTY
Nova
Swift
Glance
Horizon
Keystone
Neutron
Cinder
Ceilometer
Heat
Trove
Sahara
Magnum
...
N-1

RDO (4)

Operations

- Rundeck for daily operation and automation
- EL6 support for clients ; support until 2020, main OS version for experiments users.

RDO (5)

Operations

- Ru
  an
- EL
  ma

```
linuxsoft.cern.ch updaters statistics for: 2016-01-25
------------------------------------------------------------------------------
------------------------------------------------------------------------------
slc5/i386:        249   [int:  103 +    ext:  143 +     pup:    3]      snap:    0
slc5/x86_64:      888   [int:  399 +    ext:  389 +     pup:  100]      snap:   57
slc5/all    :    1128   [int:  502 +    ext:  524 +     pup:  102]      snap:    0

slc6/i386:        248   [int:  158 +    ext:   86 +     pup:    4]      snap:    0
slc6/x86_64:    11940   [int: 3779 +    ext: 1503 +     pup: 6658]      snap: 7543
slc6/all    :   12165   [int: 3931 +    ext: 1575 +     pup: 6659]      snap:    0

cc7/i386:           0   [int:    0 +    ext:    0 +     pup:    0]      snap:    0
cc7/x86_64:      5633   [int:  752 +    ext:  125 +     pup: 4756]      snap:    0
cc7/all    :     5633   [int:  752 +    ext:  125 +     pup: 4756]      snap:    0

------------------------------------------------------------------------------
All         :   18896   [int: 5175 +    ext: 2180 +     pup:11541]      snap: 7644
------------------------------------------------------------------------------
```

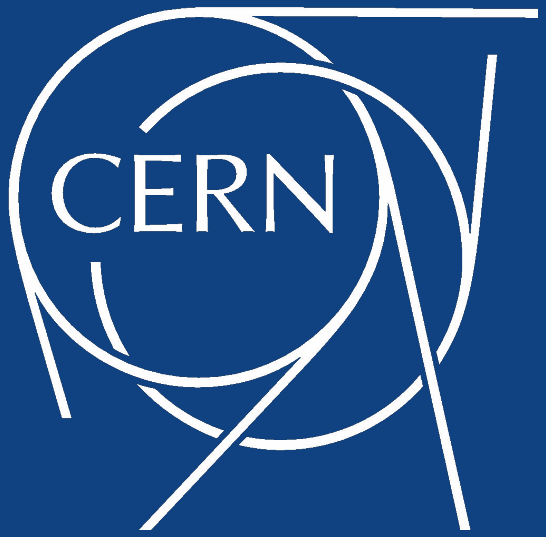
RDO (6)

Contributions

- Openstack Juno for el6
- Several packages (ec2-api, Murano, etc...)
- Cloud SIG workflow on the CentOS side.
- https://github.com/cernops

Summary

- The CERN OpenStack Cloud based on RDO evolved into a rapidly growing but very stable service
  - More than doubled the resources since 2013
  - Will enable significant growth 2016
- We moved new OpenStack projects into production and have some more under evaluation
- http://openstack-in-production.blogspot.com


CERN

WIP: Container integration

- Started to look into integration of containers with our OpenStack deployment
  - Initially triggered by the prospect of low performance overheads
  - LXC due to the lack of an upstream Docker driver (not suitable for general purpose)
- We've setup a test cell
  - Performance looks good
  - OpenStack patches for AFS & CVMFS done
  - AFS in containers: kernel access, multiple containers, tokens, …
- Started to look into OpenStack Magnum
  - Container orchestration via Docker or Kubernetes become first class OpenStack resources
  - More details probably already at next workshop

WIP: Life-cycle management

- Hardware in former cell01 will soon reach EOL
  - VMs are mostly pets and run services
  - Users would like to keep their VMs
- Service nodes left in SLC6 → CC7 migration
  - Kilo on RDO RHEL6 was difficult, but Liberty ?
- The service needs to support live-migration!
  - Not used in daily operations: resources & network constraints
  - VMs booted from volume: unproblematic, fast
  - VMs on ephemeral disks: **block** live-migration seems to work (from SLC6 to CentOS 7 out-of-box, from CentOS 7 **after qemu version update**)
  - VMs with volumes: needs volume detach
- We need tools to do this at scale so that live-migration can be become part of our daily operations.