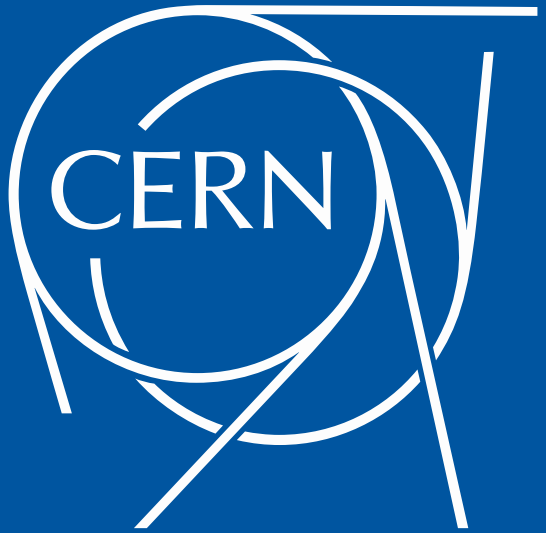
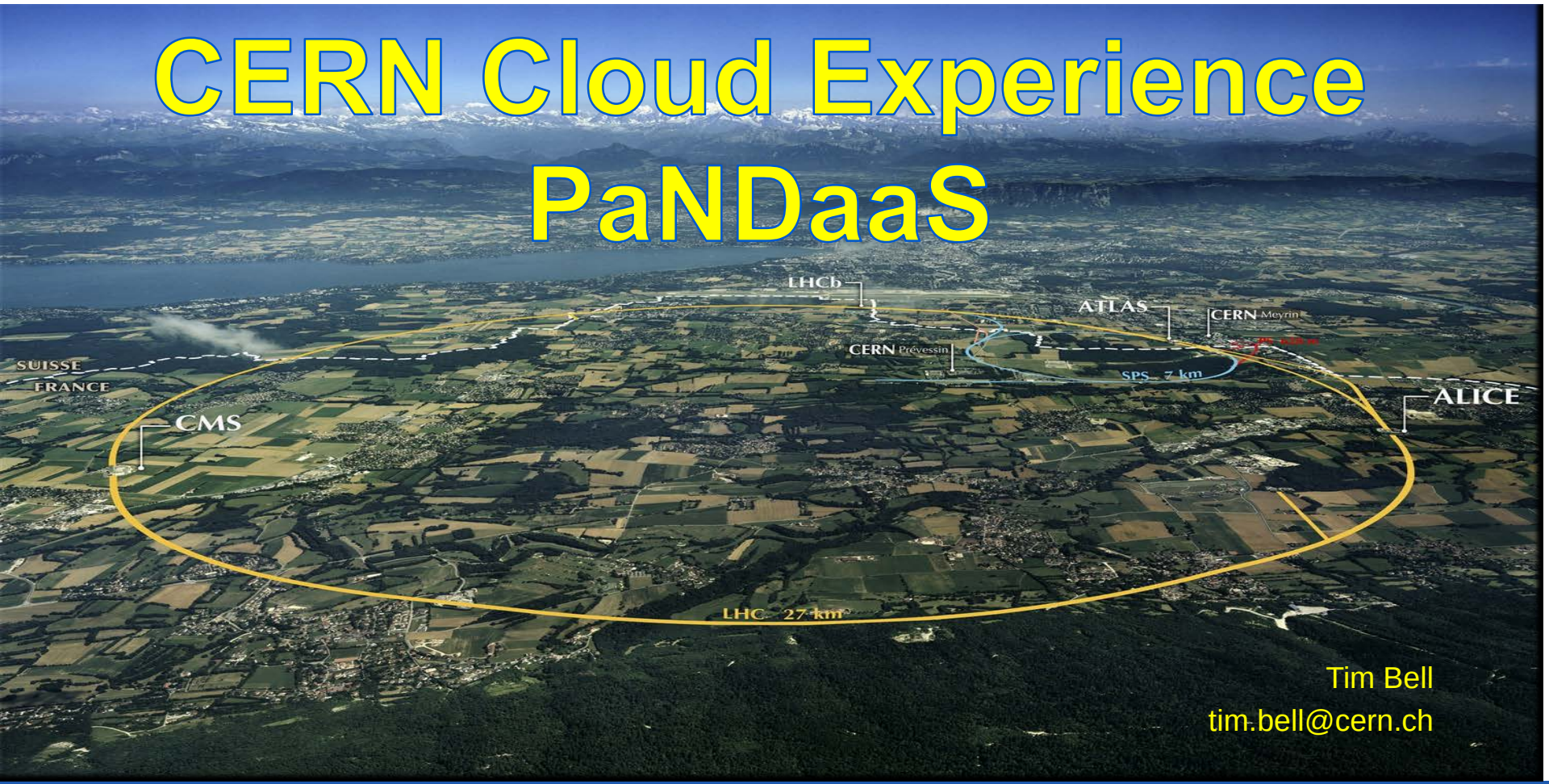

CERN

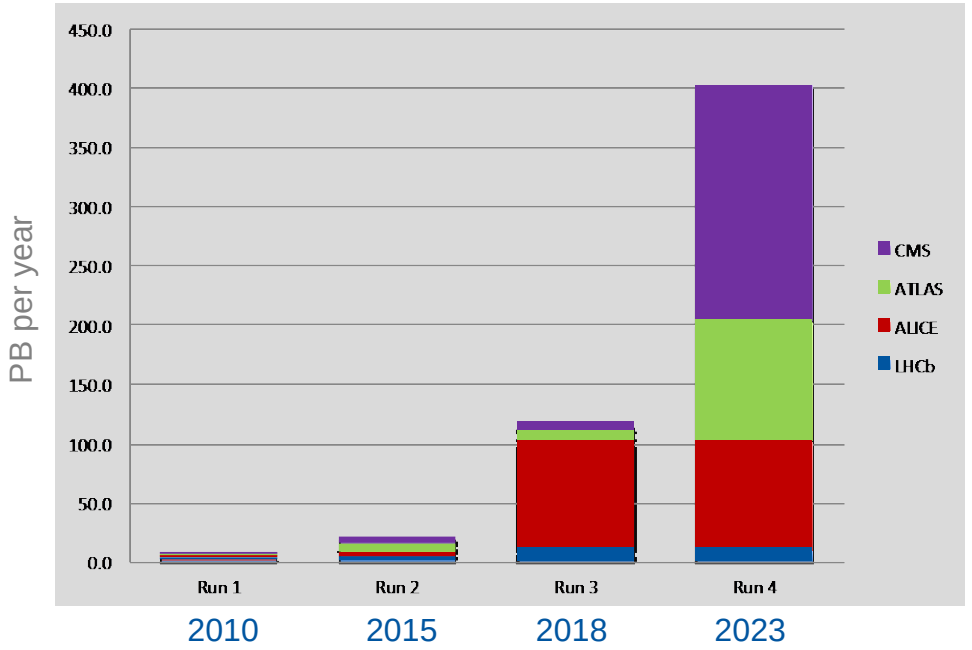
CERN Cloud Experience
PaNDaaS



Tim Bell
tim.bell@cern.ch

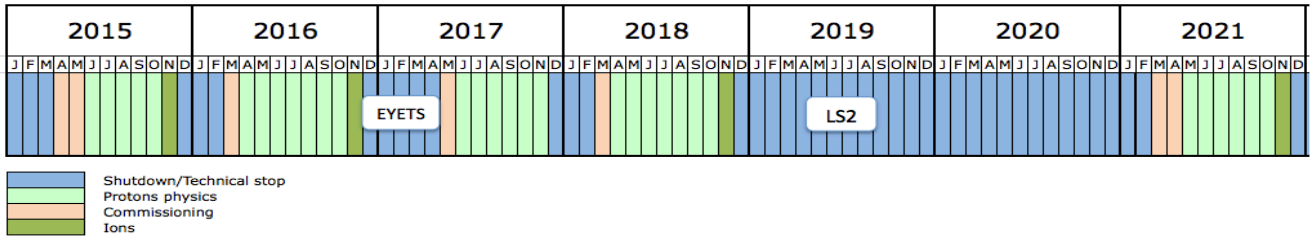
LHC Data Growth

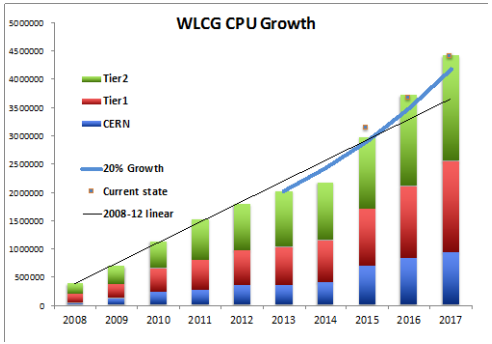
Expecting to record 400PB/year by 2023 with the High Luminosity LHC upgrade

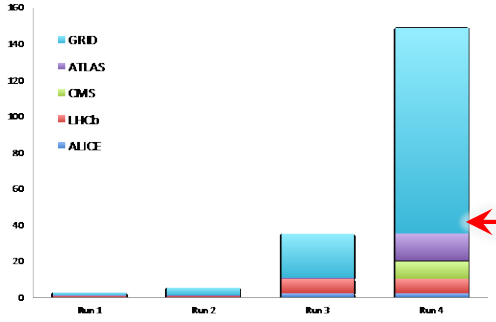


Where is x3 improvement ?

The outline LHC schedule out to 2035 presented by Frederick Bordry to the SPC and FC June 2015 can be found here







Compute: Growth > x50

What we think is affordable unless we do something differently


DANTE
100 GbE
T-Systems
100 GbE
Wigner RCP
CERN
Amsterdam
Netherlands
Niedersachsen
Berlin
Poznan
Warsaw
Poland
Antwerpen
Brussels
Belgium
Cologne
Germany
Luxembourg
Luxembourg
Prague
Czech Republic
Krakow
Paris
Slovakia
Munich
Vienna
Bratislava
Budapest
Austria
Hungary
Zurich
Liechtenstein
Berne
Switzerland
France
Slovenia
Zagreb
Croatia
Veneto
Milan
Torino
Belgrade

Good News, Bad News

- Additional data centre in Budapest now online
- Increasing use of facilities as data rates increase

But…

- Staff numbers are fixed, no more people
- Materials budget decreasing, no more money
- Legacy tools are high maintenance and brittle
- User expectations are for fast self-service

Public Procurement Cycle

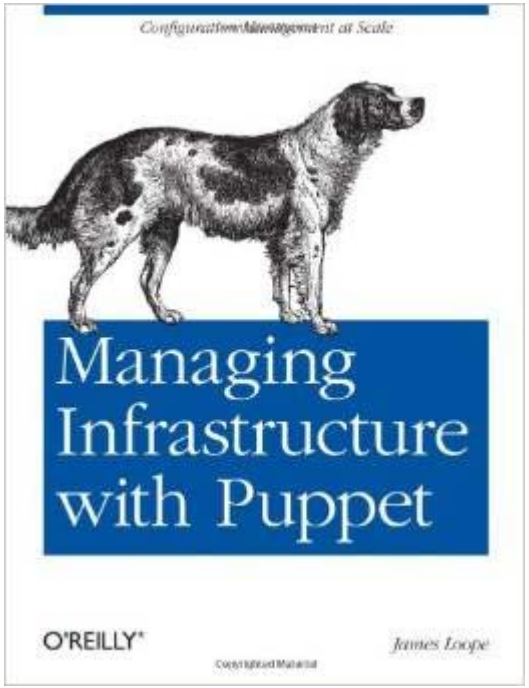
Step	Time (Days)	Elapsed (Days)
User expresses requirement		0
Market Survey prepared	15	15
Market Survey for possible vendors	30	45
Specifications prepared	15	60
Vendor responses	30	90
Test systems evaluated	30	120
Offers adjudicated	10	130
Finance committee	30	160
Hardware delivered	90	250
Burn in and acceptance	30 days typical with 380 worst case	280
Total		**280+ Days**

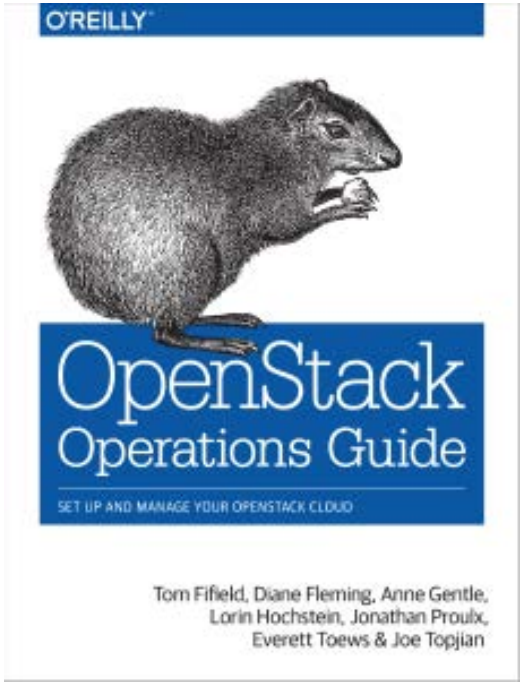
Innovation Dilemma

- How can we avoid the sustainability trap ?
  - Define requirements
  - No solution available that meets those requirements
  - Develop our own new solution
  - Accumulate technical debt
- How can we learn from others and share ?
  - Find compatible open source communities
  - Contribute back where there is missing functionality
  - Stay mainstream

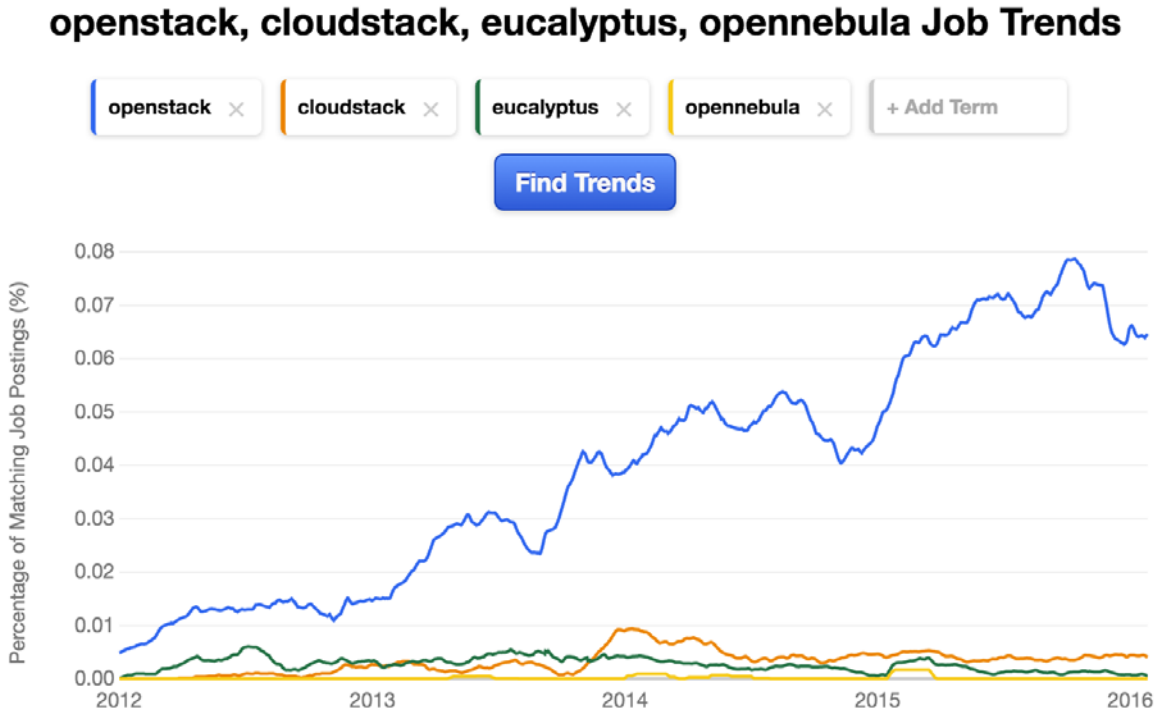
Are CERN computing needs really special ?

O’Reilly Consideration



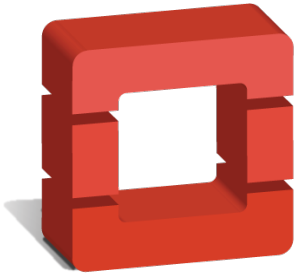


Job Trends Consideration

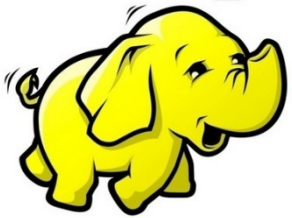


Source: Indeed.com

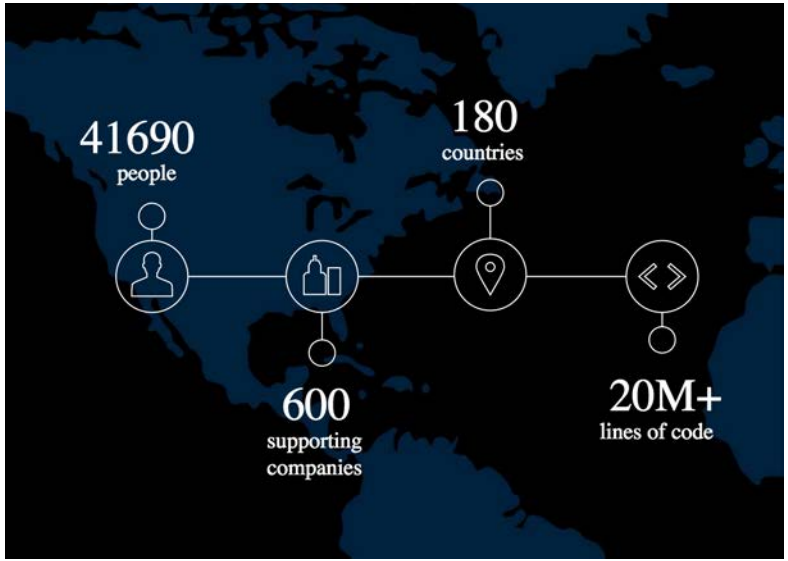
CERN Tool Chain



elasticsearch.



OpenStack Community



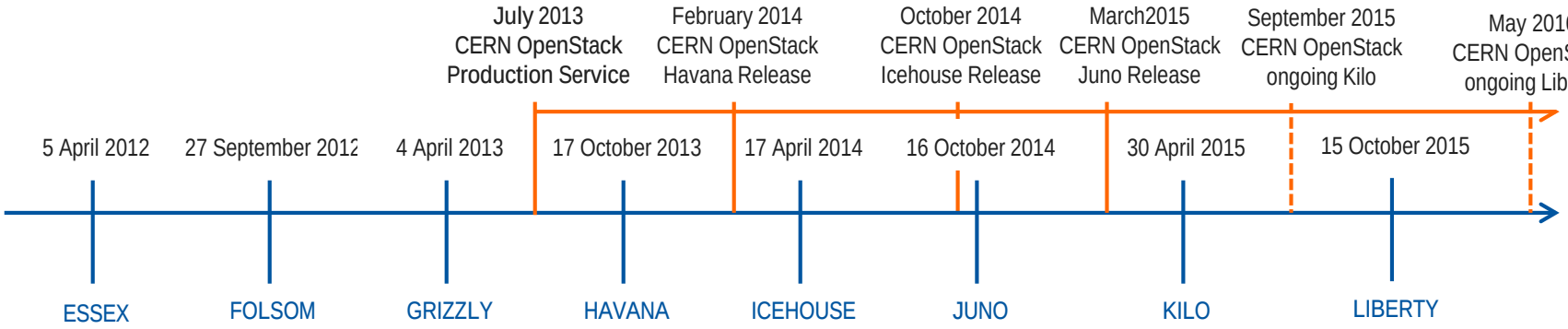
- One of the largest open source communities

- 90 user groups worldwide

OpenStack Governance

- Board consisting of 8 platinum companies, 8 gold members and 8 elected members of the community
- Technical committee guiding development direction, elected by the contributors
- User committee covering working groups such as Telecom NFV, Large Deployments and Scientific Working group
- Influence is a do-ocracy – contribute or accept destination

CERN OpenStack project



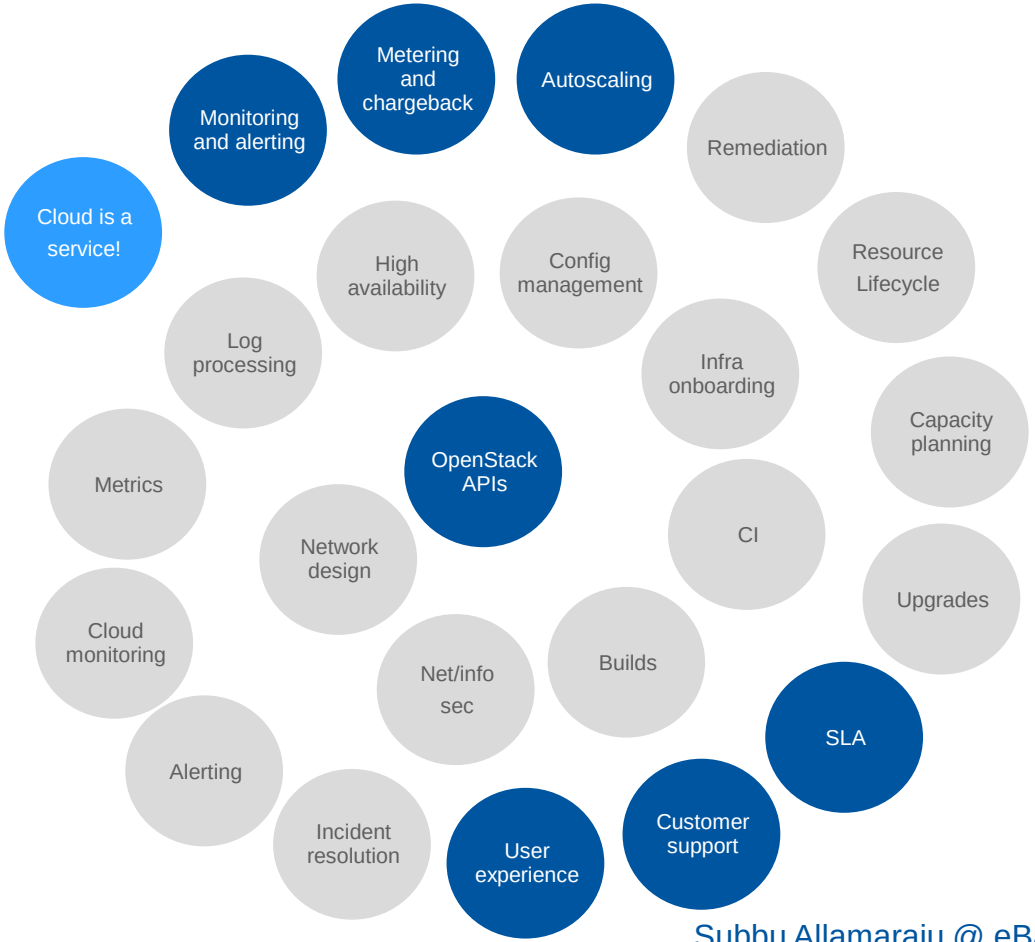
ESSEX	FOLSOM	GRIZZLY	HAVANA	ICEHOUSE	JUNO	KILO	LIBERTY
Nova	Nova	Nova	Nova	Nova	Nova	Nova	Nova
Swift	Swift	Swift	Swift	Swift	Swift	Swift	Swift
Glance	Glance	Glance	Glance	Glance	Glance	Glance	Glance
Horizon	Horizon	Horizon	Horizon	Horizon	Horizon	Horizon	Horizon
Keystone	Keystone	Keystone	Keystone	Keystone	Keystone	Keystone	Keystone
	Quantum	Quantum	Neutron	Neutron	Neutron	Neutron	Neutron (*)
	Cinder	Cinder	Cinder	Cinder	Cinder	Cinder	Cinder
		Ceilometer	Ceilometer	Ceilometer	Ceilometer	Ceilometer	Ceilometer
			Heat	Heat	Heat	Heat	Heat
					Rally	Rally	Rally
							Magnum (*)
							Barbican (*)

(*) Pilot

Upstream OpenStack on its own does not give you a cloud service

e.g.

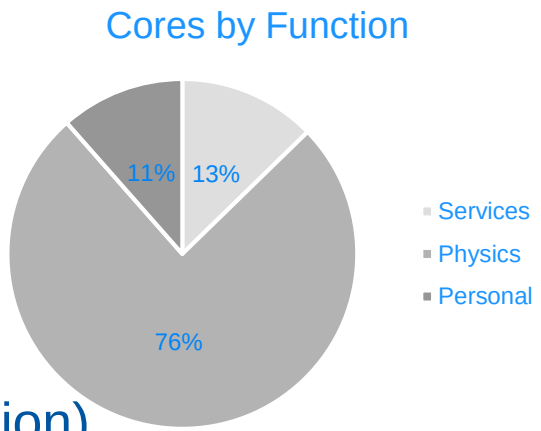
- 200 people arrive and leave CERN / month
- Skill levels vary widely
- Matching user requirements with capabilities



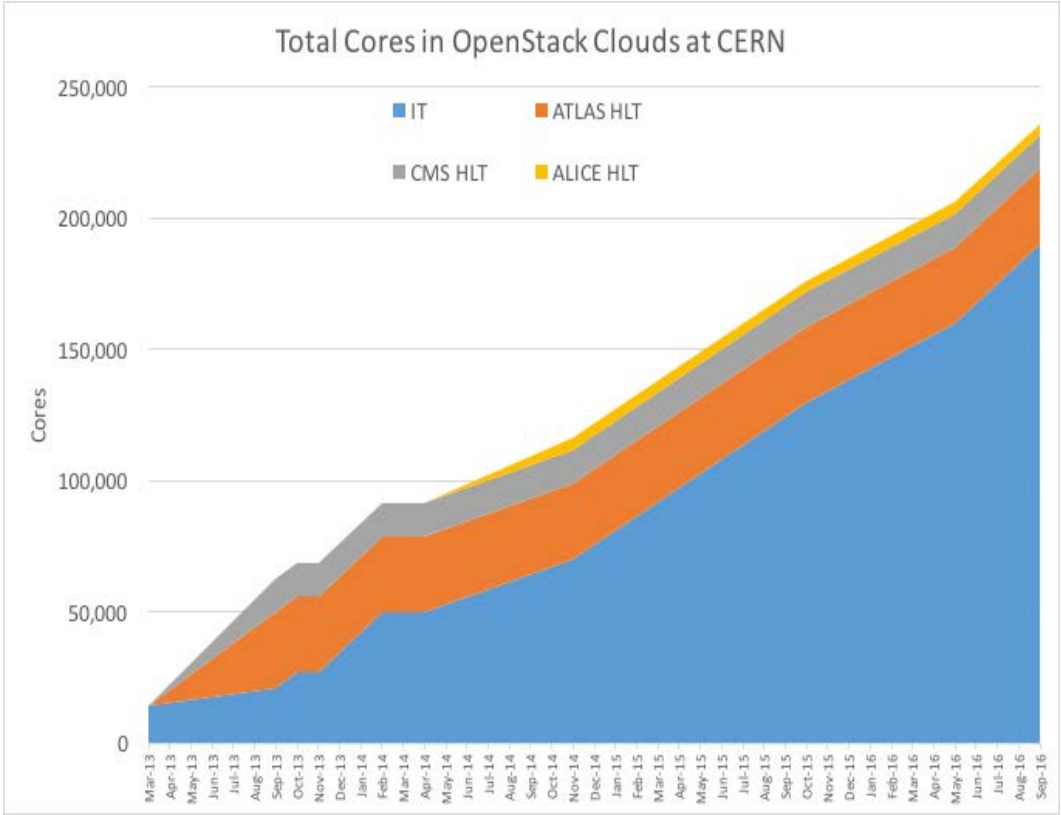
Subbu Allamaraju @ eBay

Workloads

- CERN's cloud provides a centralised resource portal
  - Project structures define roles
  - Accounting showback
  - Resource lifecycle
- For
  - Compute (e.g. virtual batch processing)
  - Production services (physics and administration)
  - Personal machines for build/test/…



OpenStack@CERN Status



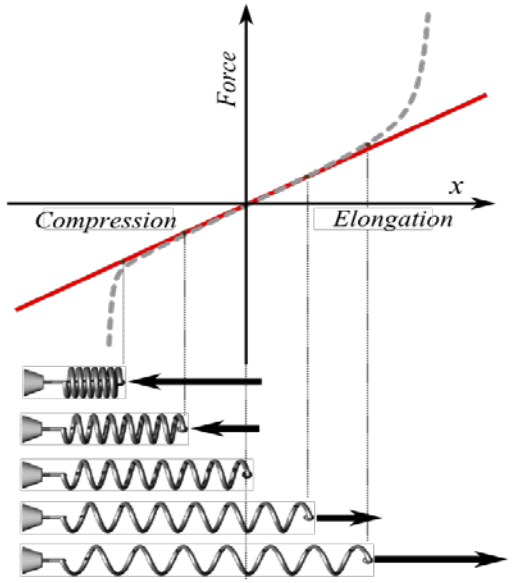
In production:

- 4 clouds
- >220K cores
- >9,000 hypervisors

~40,000 additional cores being installed in Q2 2016

90% of CERN's compute resources are now delivered on top of OpenStack

Hooke’s Law for Cultural Change

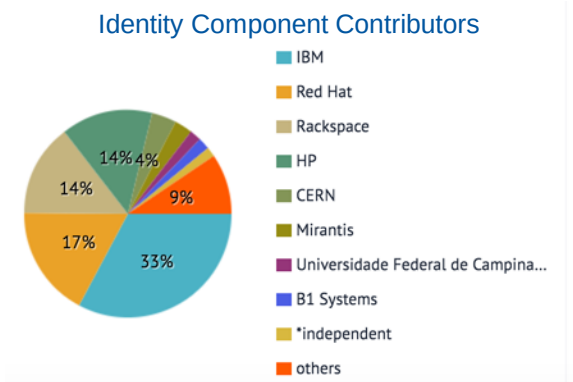


Under load, an organization can extend proportional to external force

Too much stretching leads to permanent deformation

Community Collaboration

- Open source collaboration sets model for in-house teams
- External recognition by the community is highly rewarding for contributors
- Reviews and being reviewed is a constant learning experience



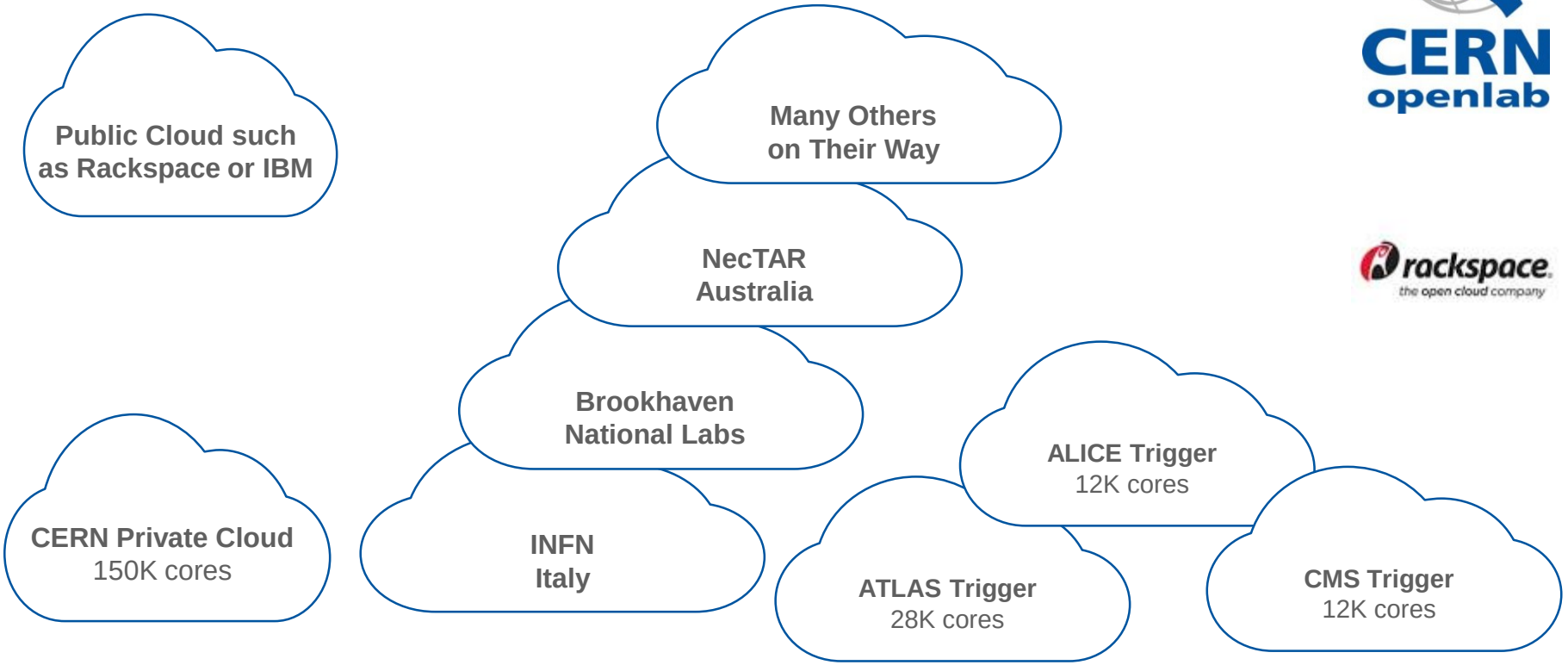
Private Cloud Experiences

- Cloud model has allowed us to scale within fixed man power
  - Emulate legacy environment
  - Enable new ways of working
  - But difficult to avoid divergence / segmentation – saying ’no’
- Skills / Culture needed to be adapted
  - Working with open source communities helps
  - Staff are enthusiastic
- Collaborations with industry covers technical debt
  - Future maintenance and testing
  - External mentoring avoids special solutions
  - Enhance staff job opportunities

Private Cloud Plans

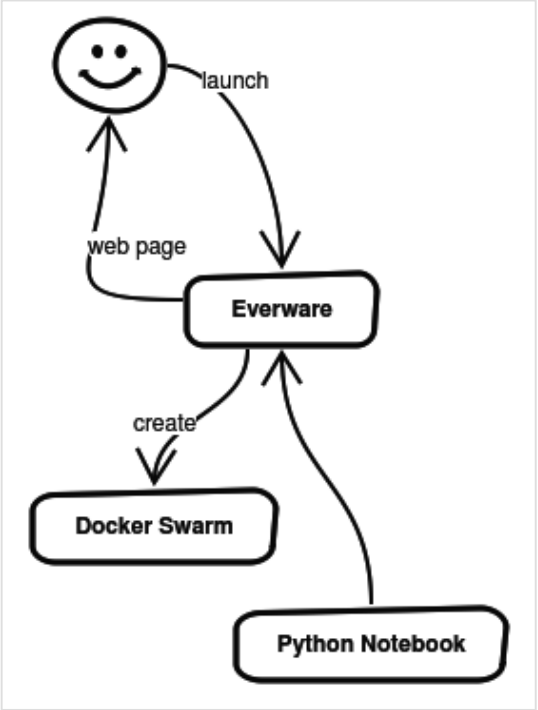
- Run the cloud
  - New releases every 6 months
  - Around 2,000 servers / year to renew
- Change the cloud
  - Containers (see later)
  - Bare metal management
  - Software Defined Networking
  - Fine grained accounting and quota
  - Resource partitioning for sensitive services

Onwards the Federated Clouds



Containers

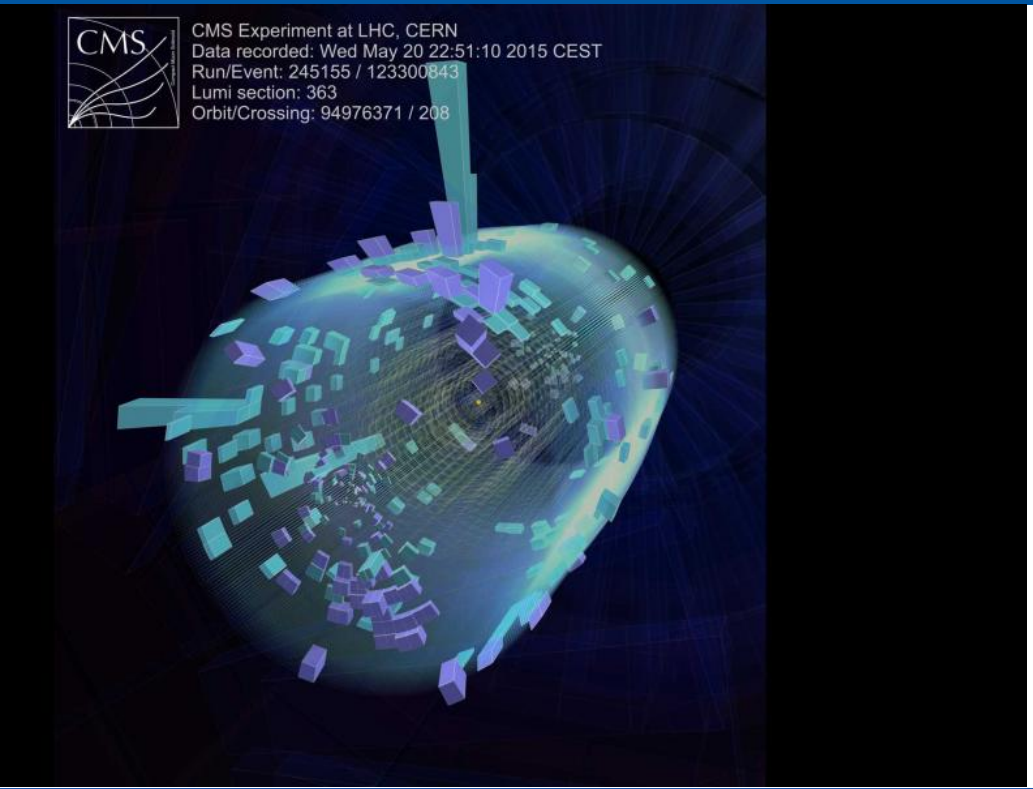
- Potentially disrupting
  - Lightweight
  - Dynamic
- Use cases
  - Indigo Dataclouds (H2020 project)
  - Jupyter / Everware notebooks
  - Batch submission including software
  - Platform-as-a-Service such as OpenShift



Summary

- Private clouds provide a flexible base for resource allocation building on open source technologies
  - Moving users to on-premise clouds can aid evolution
- API driven infrastructure has allowed significant automation to reduce manual work
  - And share tools with the community
- Finding the sweet spot with fixed budget, staff and procurement rules remains difficult

For Further Information



Technical details at http://openstack-in-production.blogspot.fr

Scientific Working Group at https://wiki.openstack.org/wiki/Scientific_working_group

CERN Containers at https://indico.cern.ch/event/506245/

Other links at http://clouddocs.web.cern.ch/clouddocs/additional/index.html

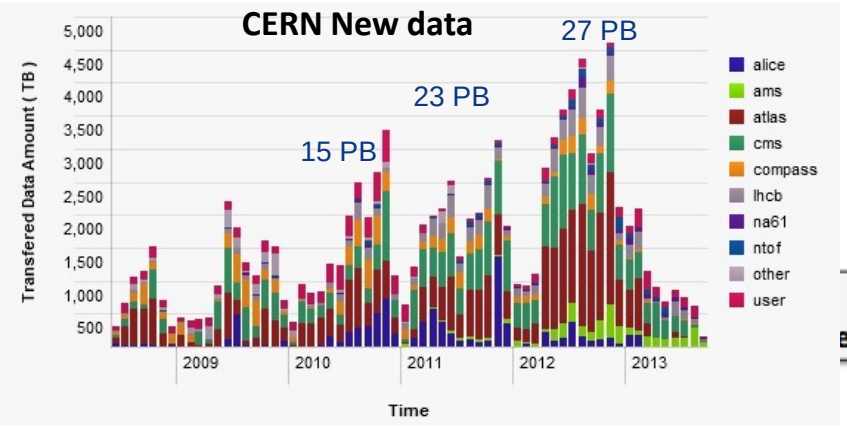
CERN tools at http://github.com/cernops

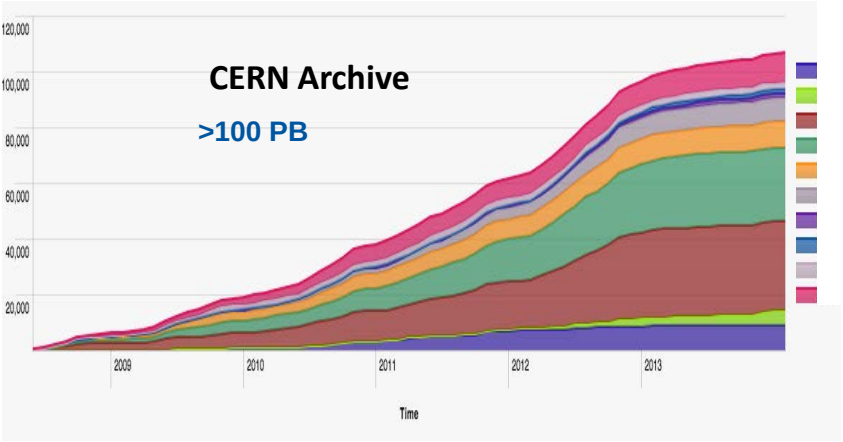
Some history of scale…

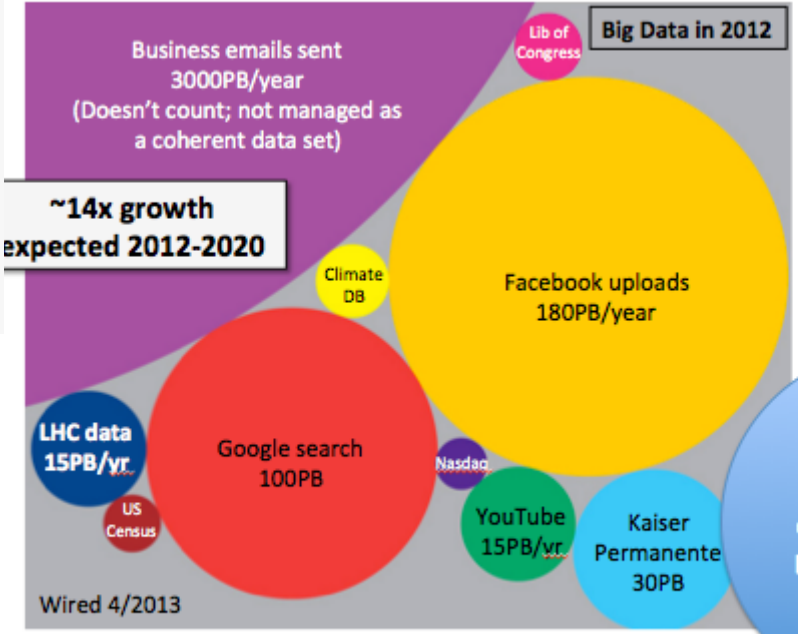
Date	Collaboration sizes	Data volume, archive technology
Late 1950's	2-3	Kilobits, notebooks
1960's	10-15	kB, punchcards
1970's	~35	MB, tape
1980's	~100	GB, tape, disk
1990's	~750	TB, tape, disk
2010's	~3000	PB, tape, disk

For comparison:
1990's: Total LEP data set ~few TB
Would fit on 1 tape today

Today: 1 year of LHC data ~27 PB





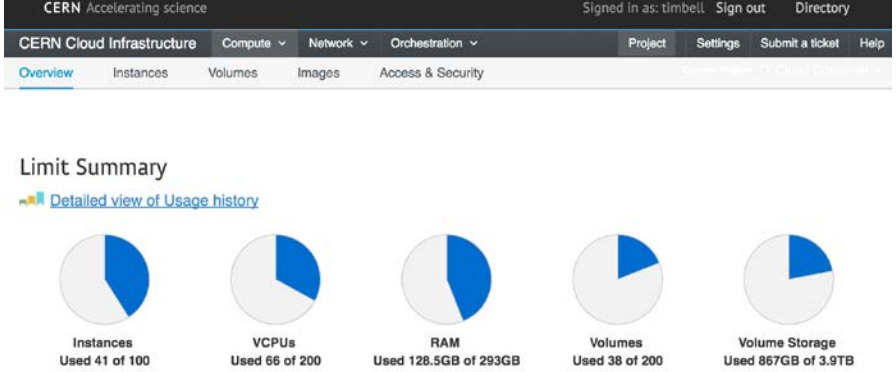


In 2012: 2800 exabytes created or replicated
1 Exabyte = 1000 PB

http://www.wired.com/magazine/2013/04/bigdata/

Functions

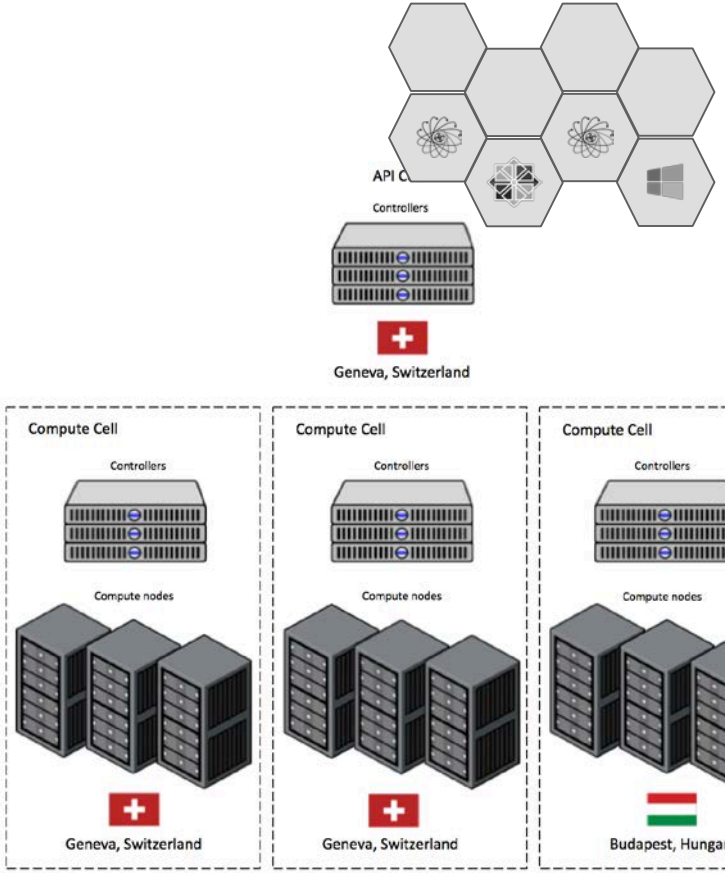
- Infrastructure as a Service
  - Virtual machines on demand
  - Volumes
  - Images
  - Orchestration
  - Identity and Authorisation
- User Interfaces
  - GUI – Novice users
  - CLI – Power users
  - API – Experiment frameworks




```
-bash-4.1$ openstack server show mv-6hzvkx2p3u-1-x7m23732t3l4-kube-minion-3vl46vzrkpxe
+--------------------------------------+--------------------------------------------------------------------------------+
| Field                                | Value                                                                          |
+--------------------------------------+--------------------------------------------------------------------------------+
| OS-DCF:diskConfig                    | MANUAL                                                                         |
| OS-EXT-AZ:availability_zone          | nova                                                                           |
| OS-EXT-STS:power_state               | 1                                                                              |
| OS-EXT-STS:task_state                | None                                                                           |
| OS-EXT-STS:vm_state                  | active                                                                         |
| OS-SRV-USG:launched_at               | 2016-04-06T15:53:08.000000                                                     |
| OS-SRV-USG:terminated_at             | None                                                                           |
| accessIPv4                           |                                                                                |
| accessIPv6                           |                                                                                |
| addresses                            | CERN_NETWORK=137.138.6.125                                                     |
| config_drive                         |                                                                                |
| created                              | 2016-04-06T15:49:00Z                                                           |
| flavor                               | m1.small (2)                                                                   |
| hostId                               | 3ecc756ca2ffabde9e2dd04a14fc7e5ab05a182d768ae06989bfdc9f                       |
| id                                   | d13bce14-04f9-4d55-9cdd-af69c2645f56                                           |
| image                                | fedora-atomic-23 (f2cdb7d1-5f97-4950-ae7c-8b75f01a46c8)                        |
| key_name                             | cern-desktop                                                                   |
| name                                 | mv-6hzvkx2p3u-1-x7m23732t3l4-kube-minion-3vl46vzrkpxe                          |
| os-extended-volumes:volumes_attached | [{u'id': u'892ce5c6-ea87-4895-9894-8cd686a89e6f'}]                             |
| progress                             | 0                                                                              |
| properties                           |                                                                                |
| security_groups                      | [{u'name': u'default'}]                                                        |
| status                               | ACTIVE                                                                         |
| tenant_id                            | 81a10d61-0561-489f-9c35-421310c0b7b6                                           |
| updated                              | 2016-04-06T15:53:08Z                                                           |
| user_id                              | mvelten                                                                        |
+--------------------------------------+--------------------------------------------------------------------------------+
```

Cells for scaling

- Small (~100s of servers) units
- Allows OpenStack to scale
- Distribute geographically
- Used by
  - Rackspace
  - Yahoo!
  - GoDaddy
  - NecTAR
  - Walmart



Why Public Cloud?

- Flexibility?
  - Procurement time vs user needs
- Cost?
  - Building on larger providers economies of scale
- Skills?
  - Focus fixed staff on value added functions

Public Cloud Approaches

- Compute is ephemeral
  - Technical change of supplier could be done in O(1 week)
  - Can tolerate variable service levels
- Data is our key resource
  - Keep at least one copy at CERN
  - Cache locally in cloud if performance improves
- Networks are variable costs
  - How to manage the usage ?
  - How to optimise use of academic networking ?

HELIX NEBULA THESCIENCECLOUD

Strategic Plan

- Establish multi-tenant, multi-provider cloud infrastructure
- Identify and adopt policies for trust, security and privacy
- Create governance structure
- Define funding schemes

CERN

To support the computing capacity needs for the ATLAS experiment

EMBL

Setting up a new service to simplify analysis of large genomes, for a deeper insight into evolution and biodiversity

esa

cnes

To create an Earth Observation platform, focusing on earthquake and volcano research

PIC port d'informació científica

To improve the speed and quality of research for finding surrogate biomarkers based on brain images

Additional Users: ECMWF, ESO, ESRF, ILL NEUTRONS FOR SCIENCE

Suppliers

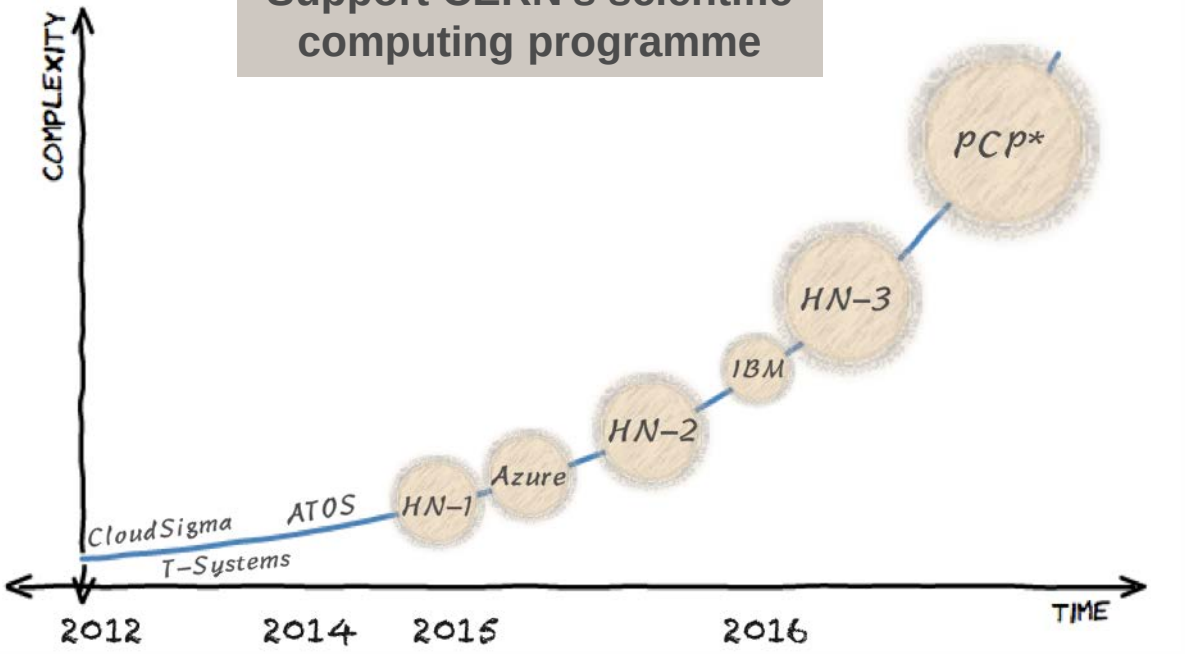
Atos, CLOUDEO, CloudSigma, Cloudwatt, DataCentred, Deutsche Börse Cloud Exchange, EGI, exoscale, GÉANT, interoute, Prologue, sixsq, SWITCH, T-Systems, the SERVER LABS, ultimum technologies

Adopters

AWST, Capgemini, CSA cloud security alliance, deac, Ifremer, MEMSET, orange Business Services, RHEA, terradue 2.0, THALES, vito, VU, Yandex

Past, ongoing & future commercial activities@CERN

Support CERN's scientific computing programme



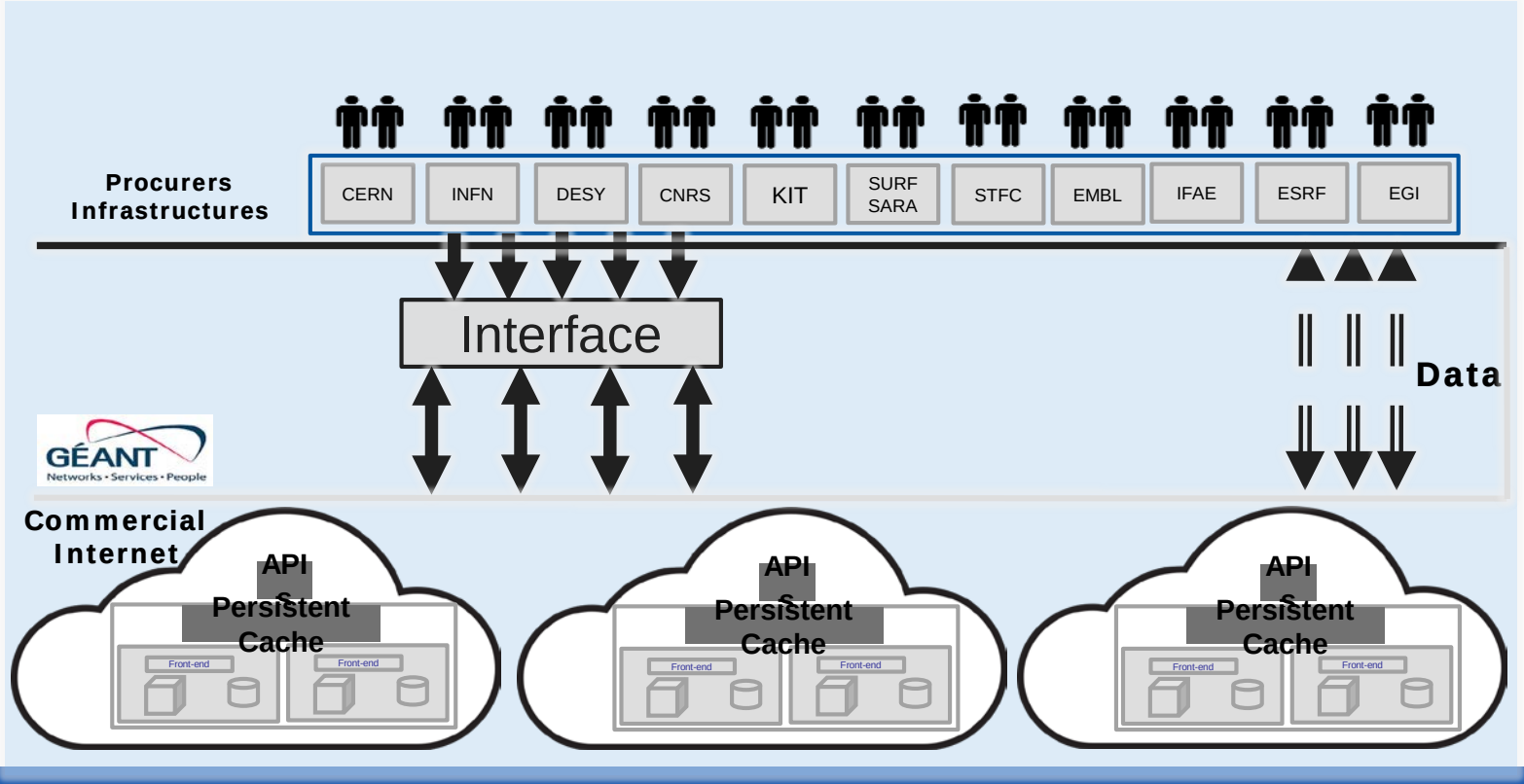
HN - Helix Nebula

- Partnership between research organization and European commercial cloud providers

* EC co-funded joint Pre-Commercial Procurement (PCP) project: https://indico.cern.ch/event/319753

** Other work has been conducted outside CERN, such as the Amazon Pilot project at BNL for ATLAS

HNSciCloud Architecture

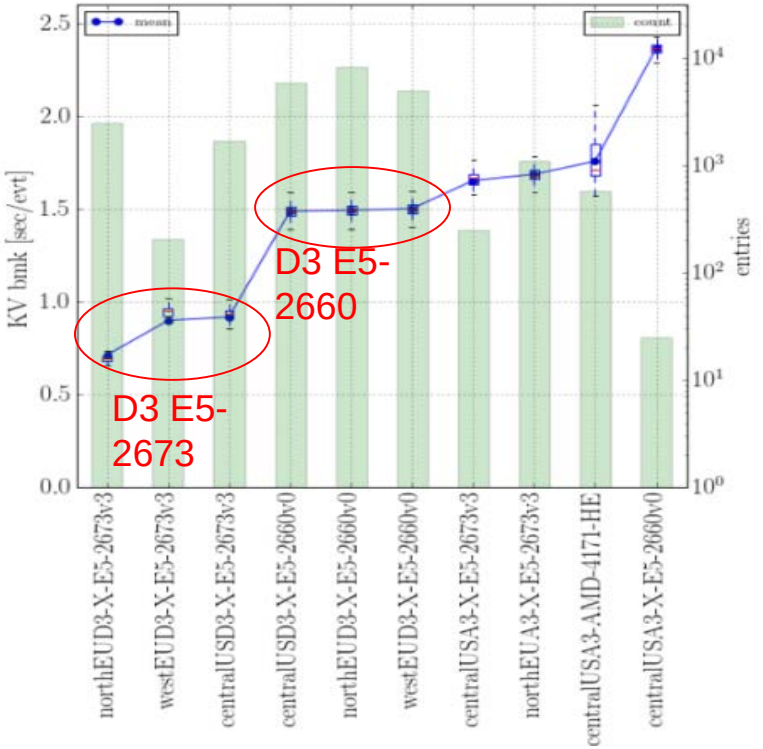


Public Cloud Experiences

- Workloads
  - CPU intensive workloads successfully run
    - With some supplier issues
  - I/O intensive workloads now being tested
    - Network remains a concern for bandwidth and control
    - Some suppliers have no public IPv4 addresses!
- Cost management
  - Differences in bills and utilisation not easy to resolve
  - Comparison with on-premise needs clear identification of sunk costs
- Procurement
  - Technical offerings are not standardised so care needed on specifications
  - Hyperscale contracts and penalty clauses may not be compatible
  - Do we need more contract managers and less system administrators ?

Benchmarking

- Procurement
  - How fast are your cores?
- Running
  - Have things got worse ?
- Accounting
  - What work was done ?



Resource models

- On Demand
  - Self-service
  - Elastic
- Reserved Instance
  - Commitment
  - Upfront payment?
  - **PREDICTABLE!**
- Spot Market
  - Bid
  - No guarantee
  - **CHEAP!**
- Bare-metal
  - Cost significant if commitment << warranty

Expansion Options

- Options to improve Meyrin CC are limited
- Wigner contract will end in 2019
- What next ?
  - On-premise ?
  - Find a replacement centre like Wigner ?
  - Public cloud reserved instance procurements ?
  - Long term leasing of bare metal ?
  - Spot?
  - Combined procurement tests ongoing with HNSciCloud