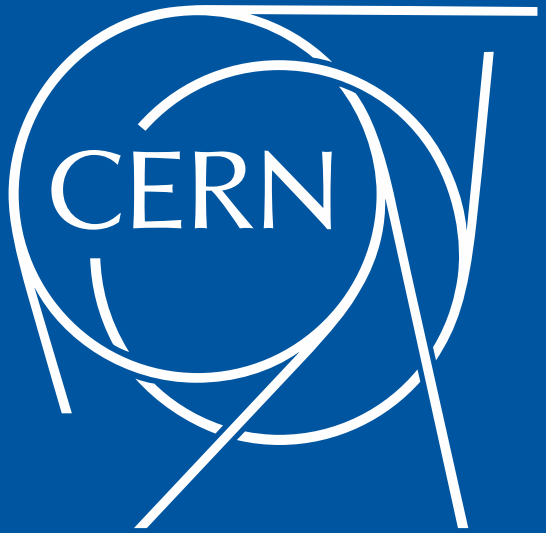
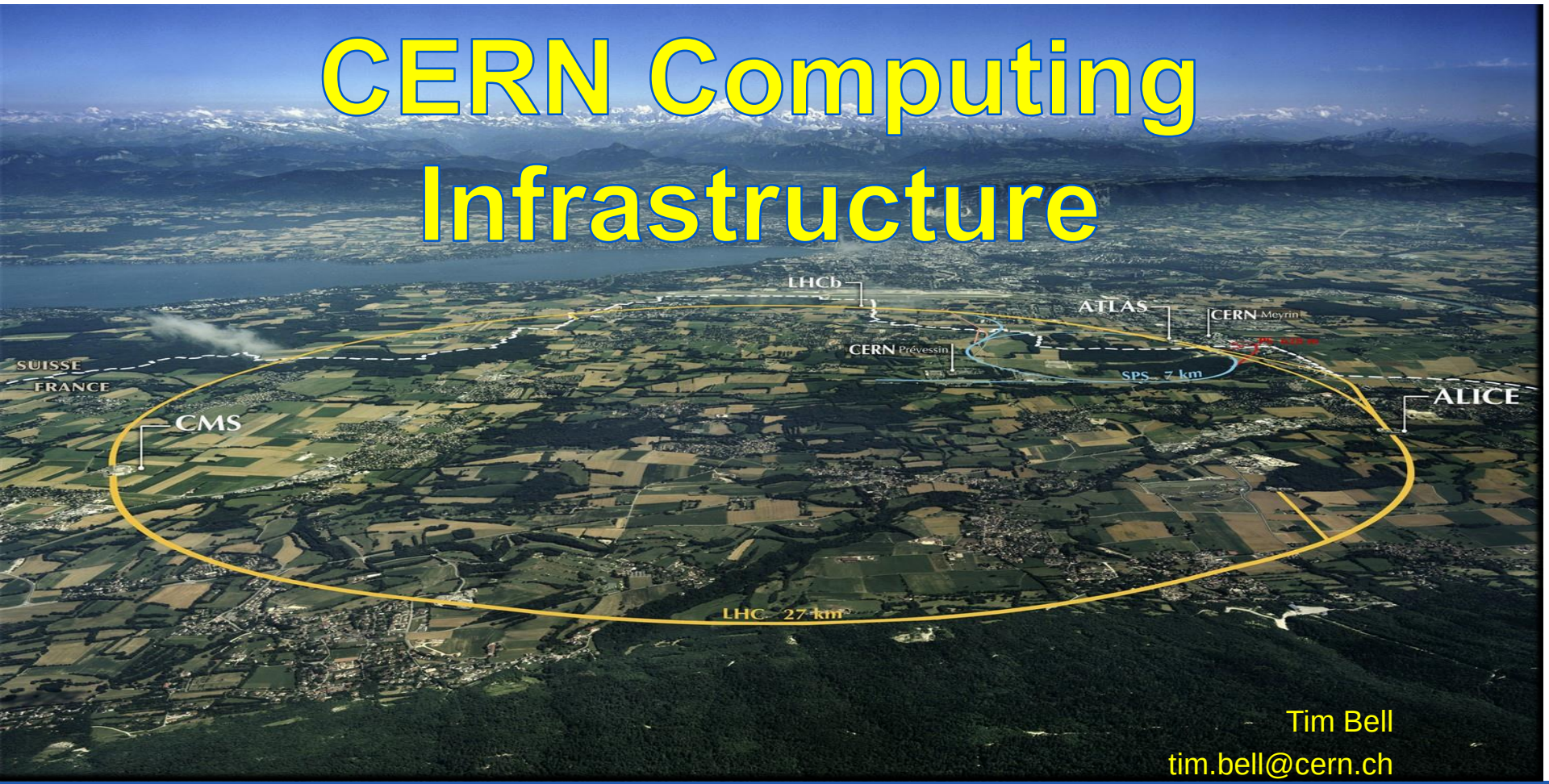
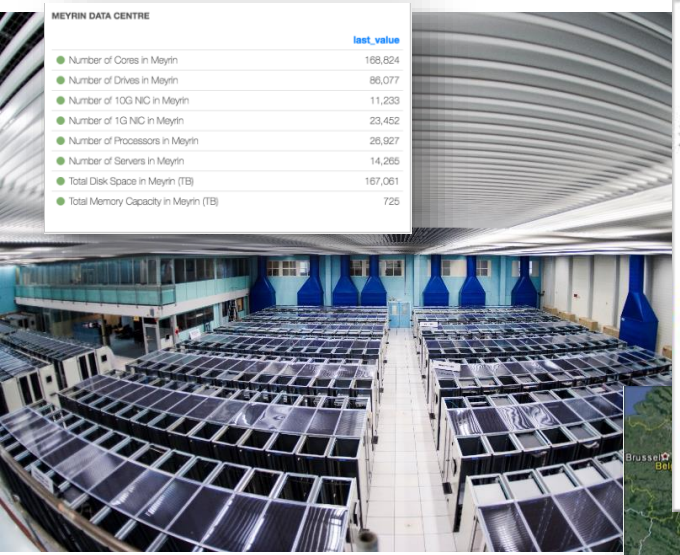

CERN

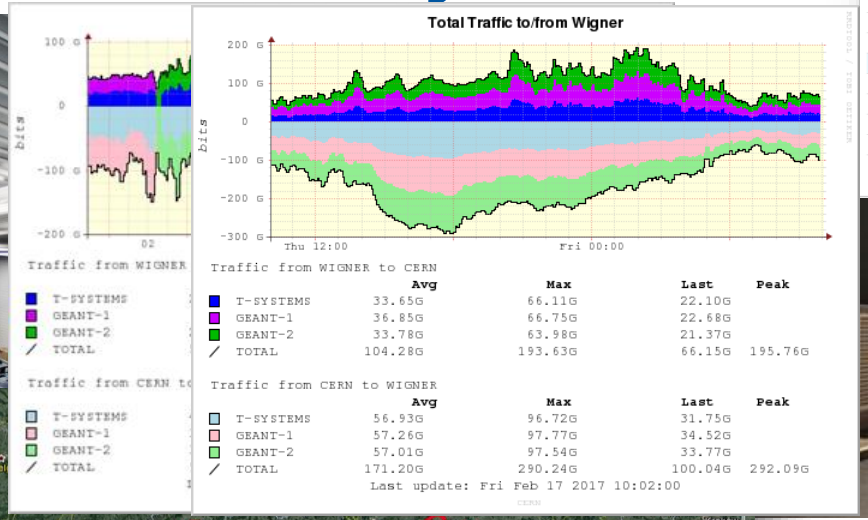
CERN Computing Infrastructure

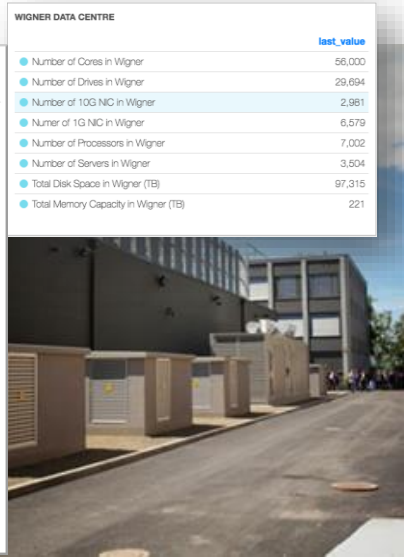


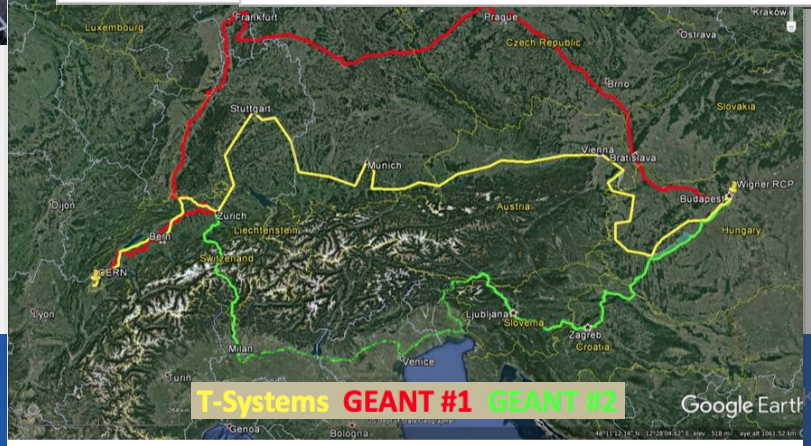
Tim Bell
tim.bell@cern.ch

CERN Facilities today









2017:

- 225k cores → 325k
- 150 PB raw → 250 PB

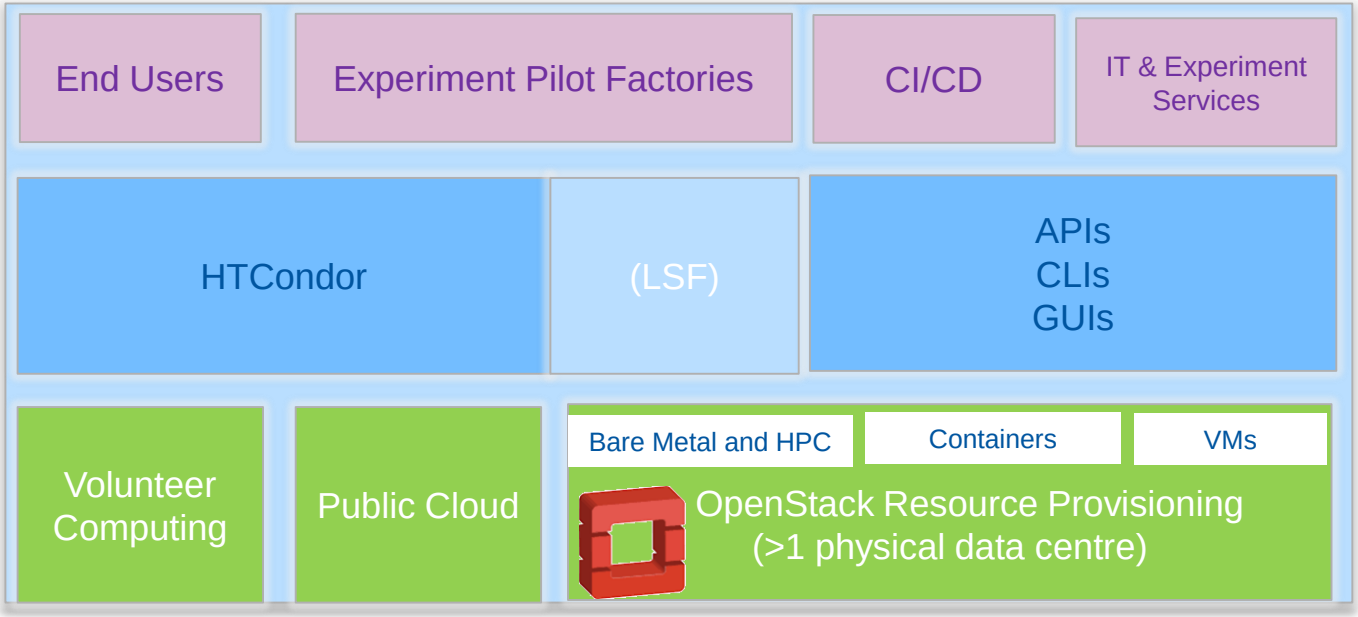
2017-18/19

- Upgrade internal networking capacity
- Refresh tape infrastructure

Common Computing Infrastructure

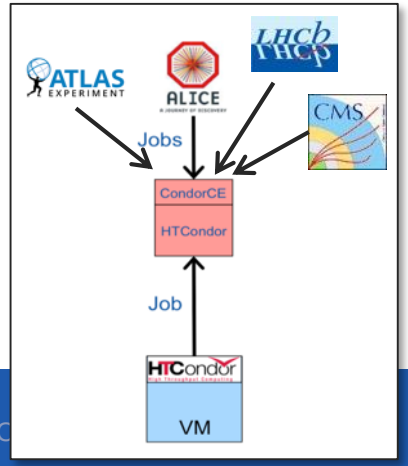
- Diverse computing services
  - Physics computing
  - IT and Experiment Services
  - Administrative Computing
- Target is for
  - Standardised procedures
  - Bulk purchasing

Provisioning services



Moving towards Elastic Hybrid IaaS model:

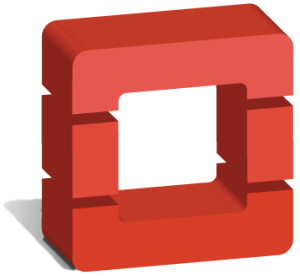
- In house resources at full occupation
- Elastic use of commercial & public clouds
  - Assume “spot-market” style pricing



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**

CERN Tool Chain



puppet labs

elasticsearch.

git

openstack™
CLOUD SOFTWARE

ceph

CentOS

hadoop

RDO

kibana

FOREMAN

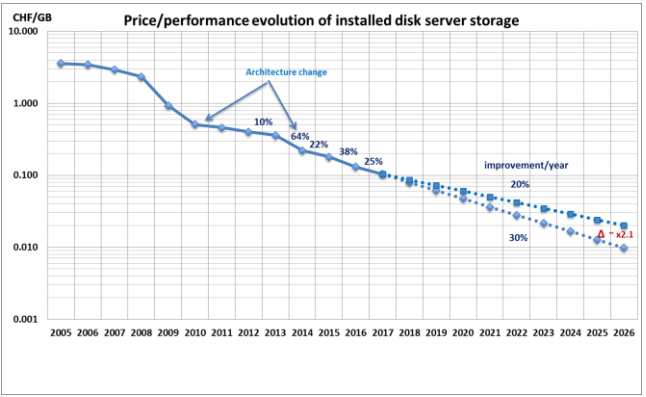
RUNDECK

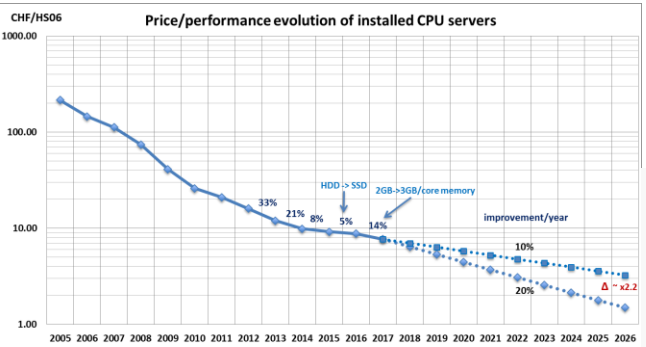
Jenkins

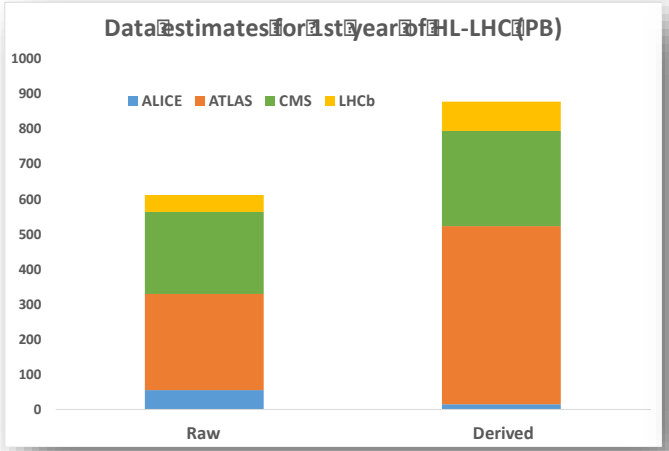
Future Challenges

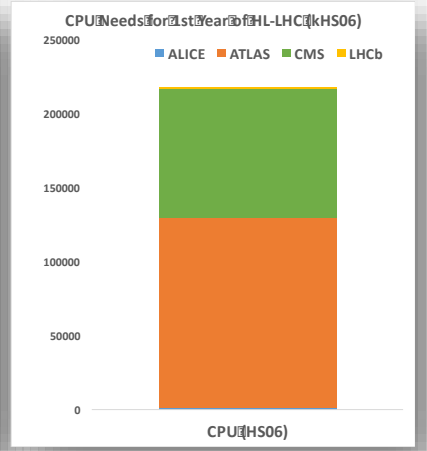
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 …… 2030?

First run	LS1	Second run	LS2	Third run	LS3	HL-LHC	FCC?







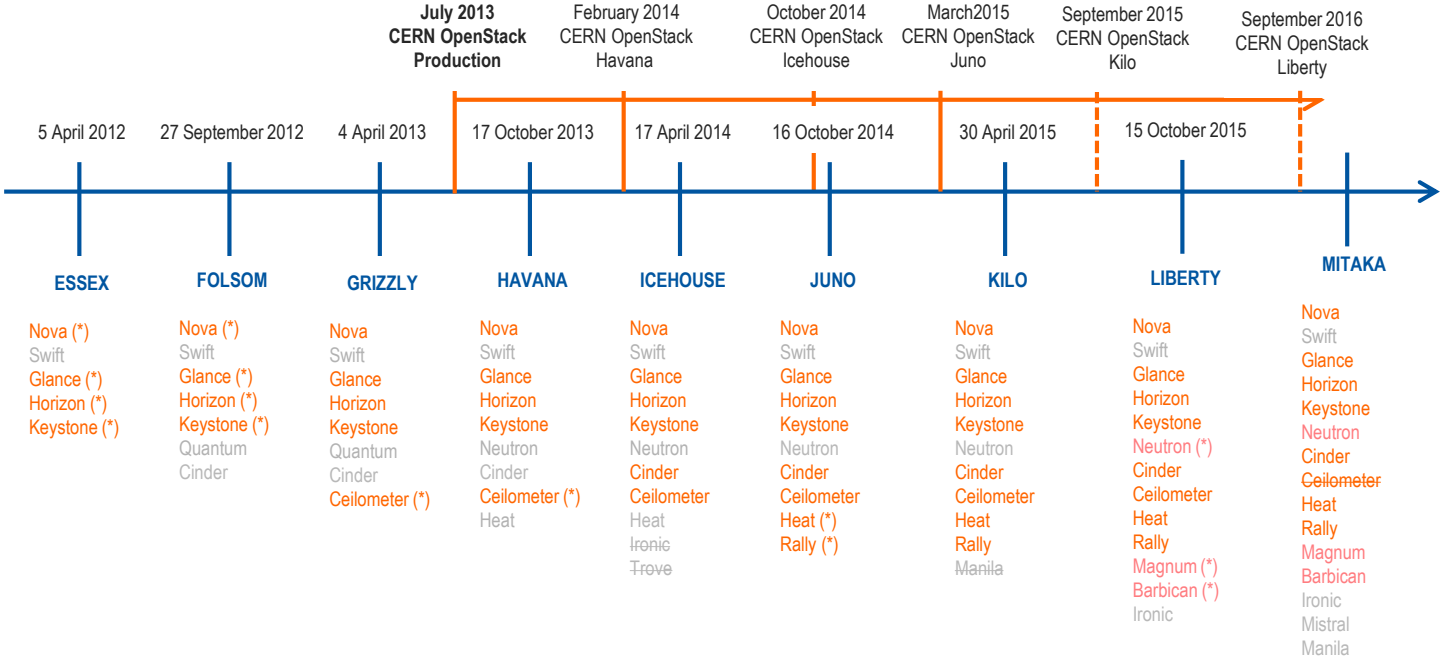


Data:
- Raw 2016: 50 PB → 2027: 600 PB
- Derived (1 copy): 2016: 80 PB → 2027: 900 PB

CPU:
- x60 from 2016

- Raw data volume for LHC increases exponentially and with it processing and analysis load
- Technology at ~20%/year will bring x6-10 in 10-11 years
- Estimates of resource needs at HL-LHC x10 above what is realistic to expect from technology with reasonably constant cost

CERN OpenStack Project



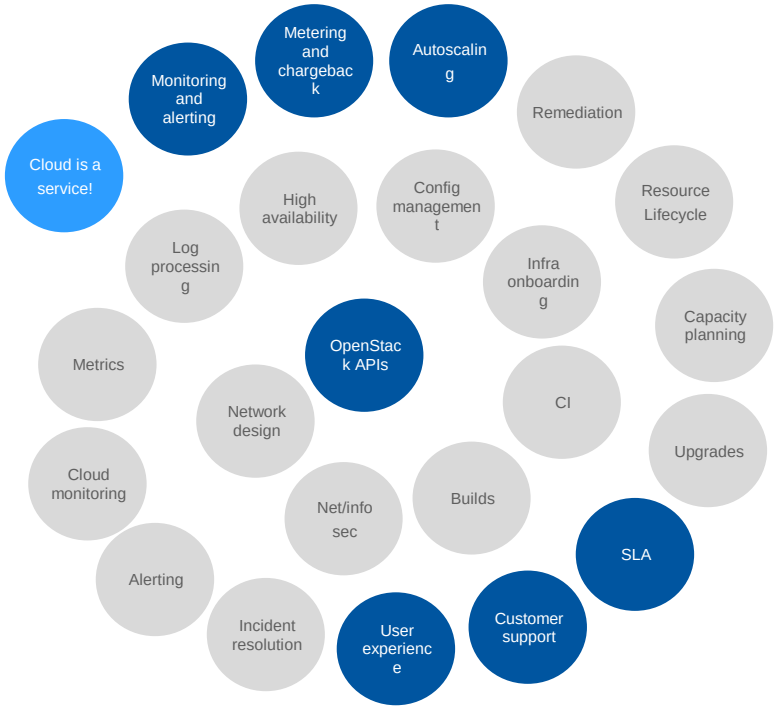
(*) Pilot Trial

Not Just the Software

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

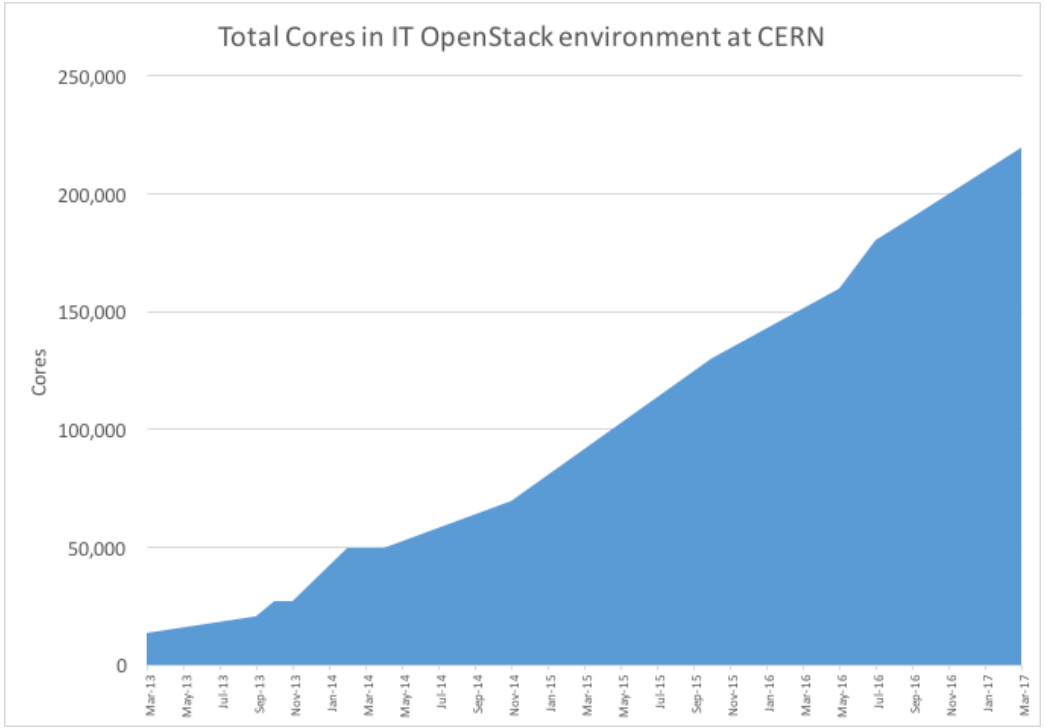
e.g.

- 200 people arrive and leave CERN / month
- Resource management and hardware lifecycles
- Integration into local inventory and network management



Subbu Allamaraju @ eBay

OpenStack@CERN Status



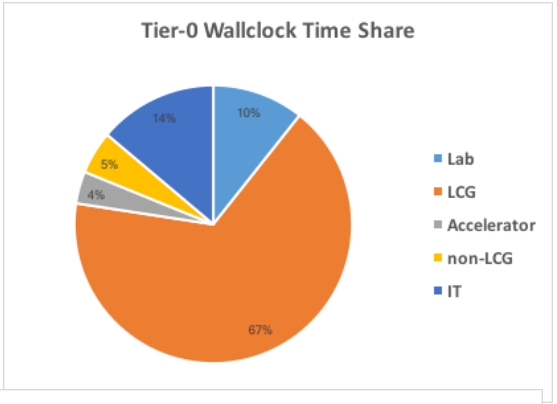
In production:

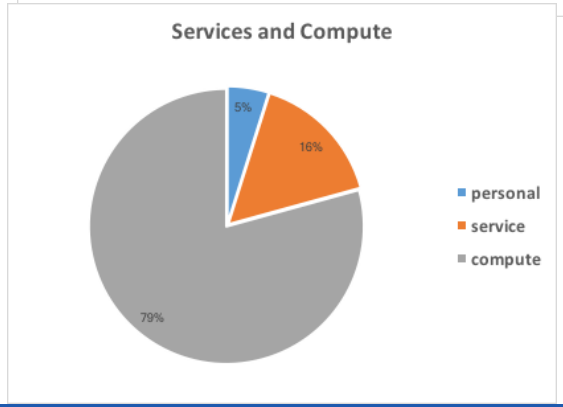
- ~230K cores
- ~7200 hypervisors

~70,000 additional cores being installed in next 6 months

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

Usage

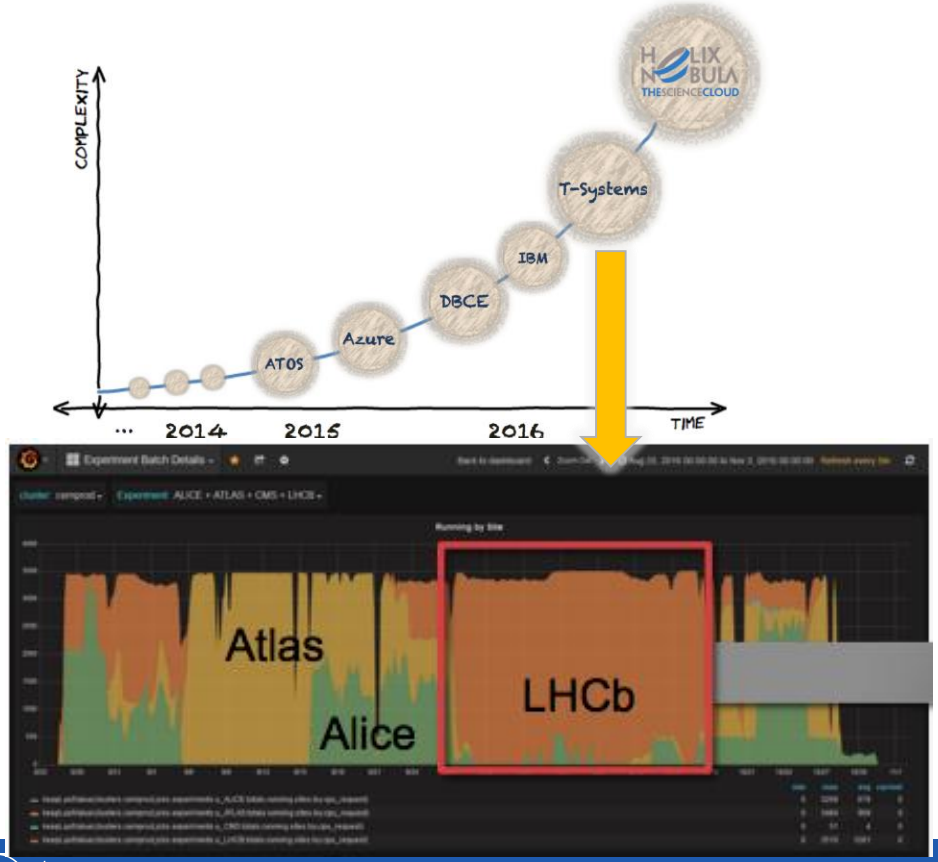


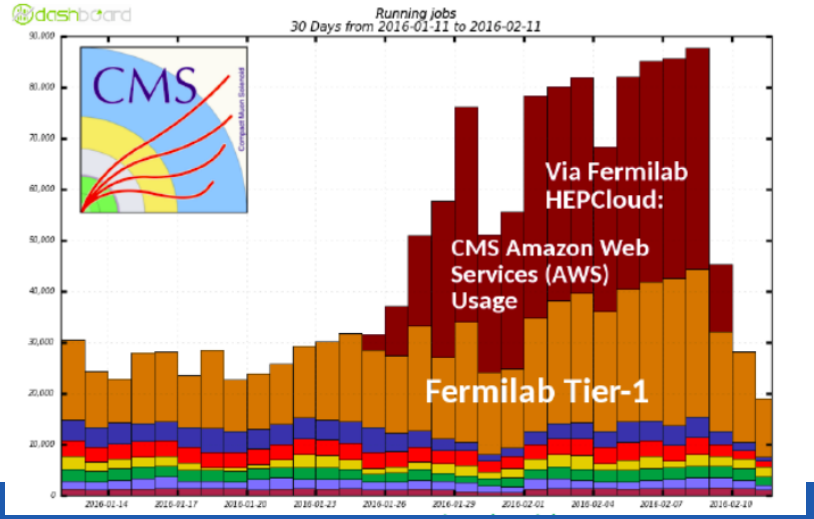


- Single resource pool for the lab infrastructure and the physics

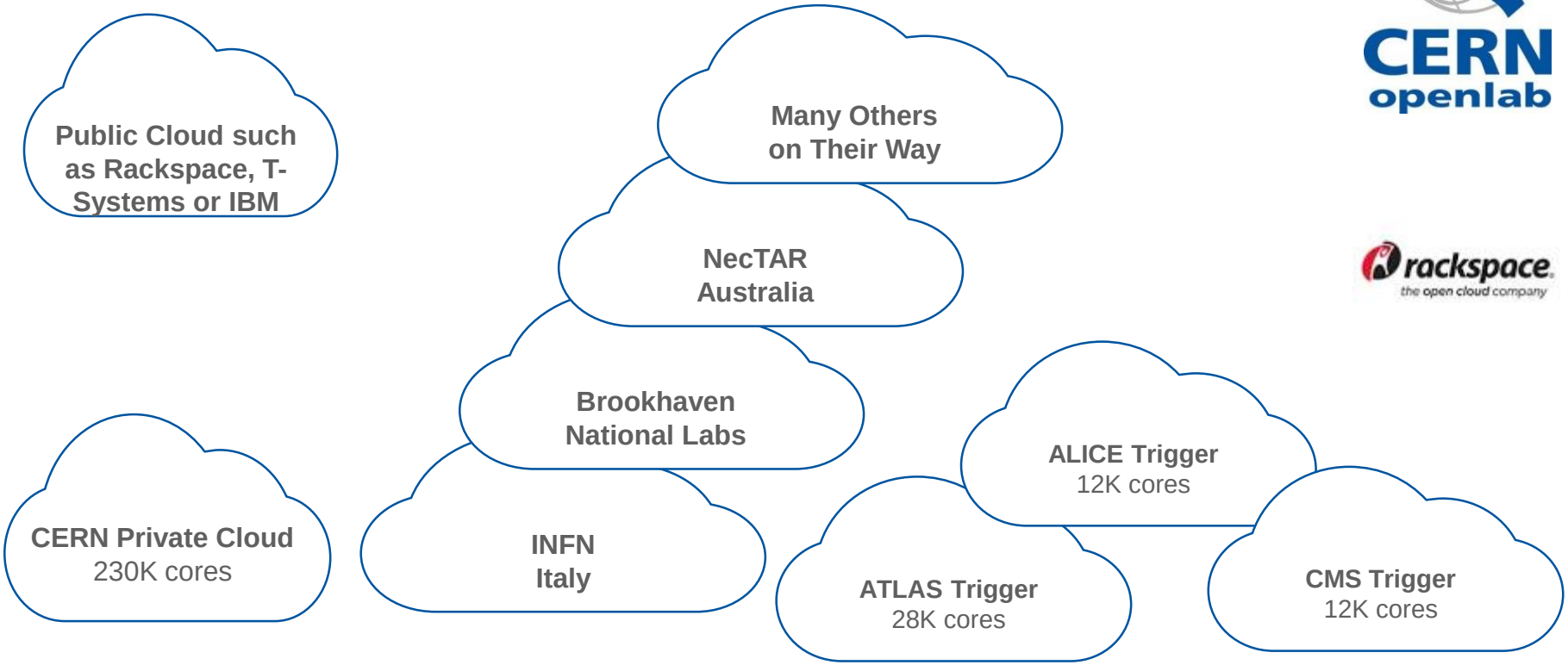
- Majority of the resources are allocated for physics compute workloads
- However, a significant share of resources are used for services which tend to be long running workloads

Commercial Clouds





Onwards the Federated Clouds



Community

- OpenStack provides an open forum
  - Scientific Working Group
  - Open Research Cloud Declaration
  - HPC and Clouds (e.g. Cambridge/HPC)
- Increasing deployments in science
  - Many research labs already have an OpenStack service
- Sharing through the community and upstream avoids support dependencies
  - Industry Collaboration Opportunities (such as Rackspace and Huawei openlab)
  - Building on the community interest in science helps sustainability

Batch - Legacy

- Currently running Platform LSF from IBM
  - Around 45,000 running jobs
  - 400,000 jobs/day
  - ~5,000 virtual servers
- Grid and local submission

Batch - Future

- Moving to HTCondor
  - Improved scalability
  - Open source
  - Good collaboration with development teams
  - Widely deployed in HEP and other sciences
  - Better at handling dynamic resources
- Grown to over 50% of the batch capacity at CERN
- First user communities are now in production
- Aim to migrate before end of LS2

Containers

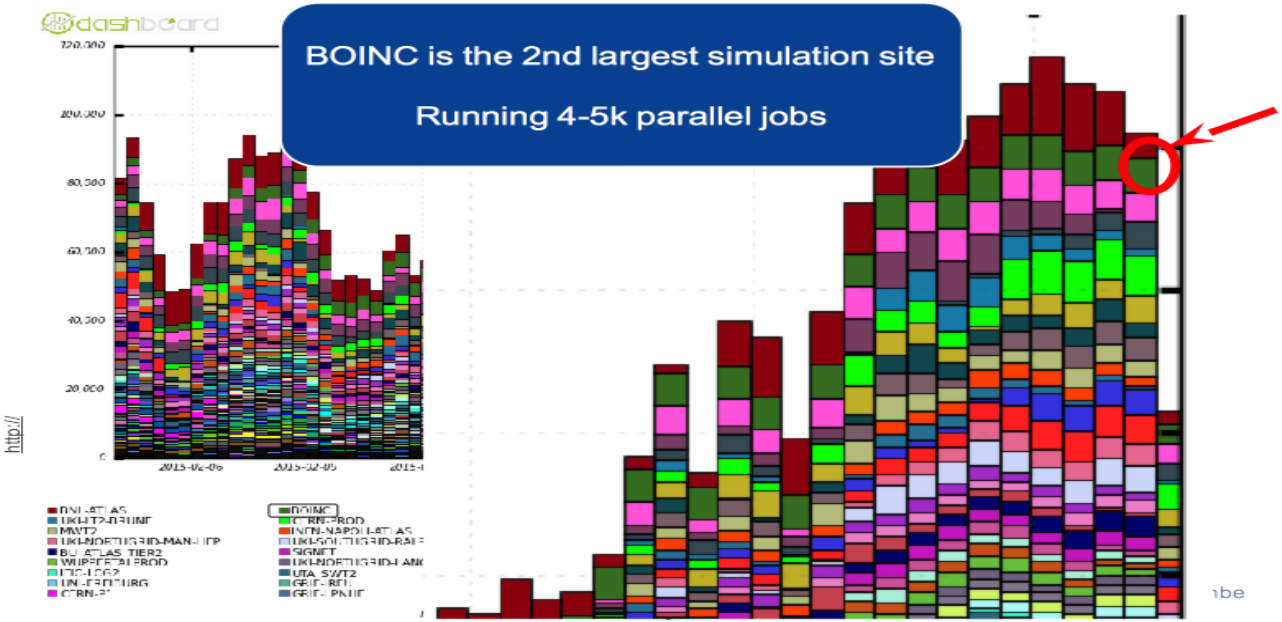
- A potentially disruptive technology
  - Microservices
  - Packaging all software in an object
  - Analysis (not just preservation)
- New Openlab collaboration project funded by Rackspace
  - Enhance OpenStack container service for scientific use cases
- Evaluating HTCondor container support

Volunteer / Opportunistic

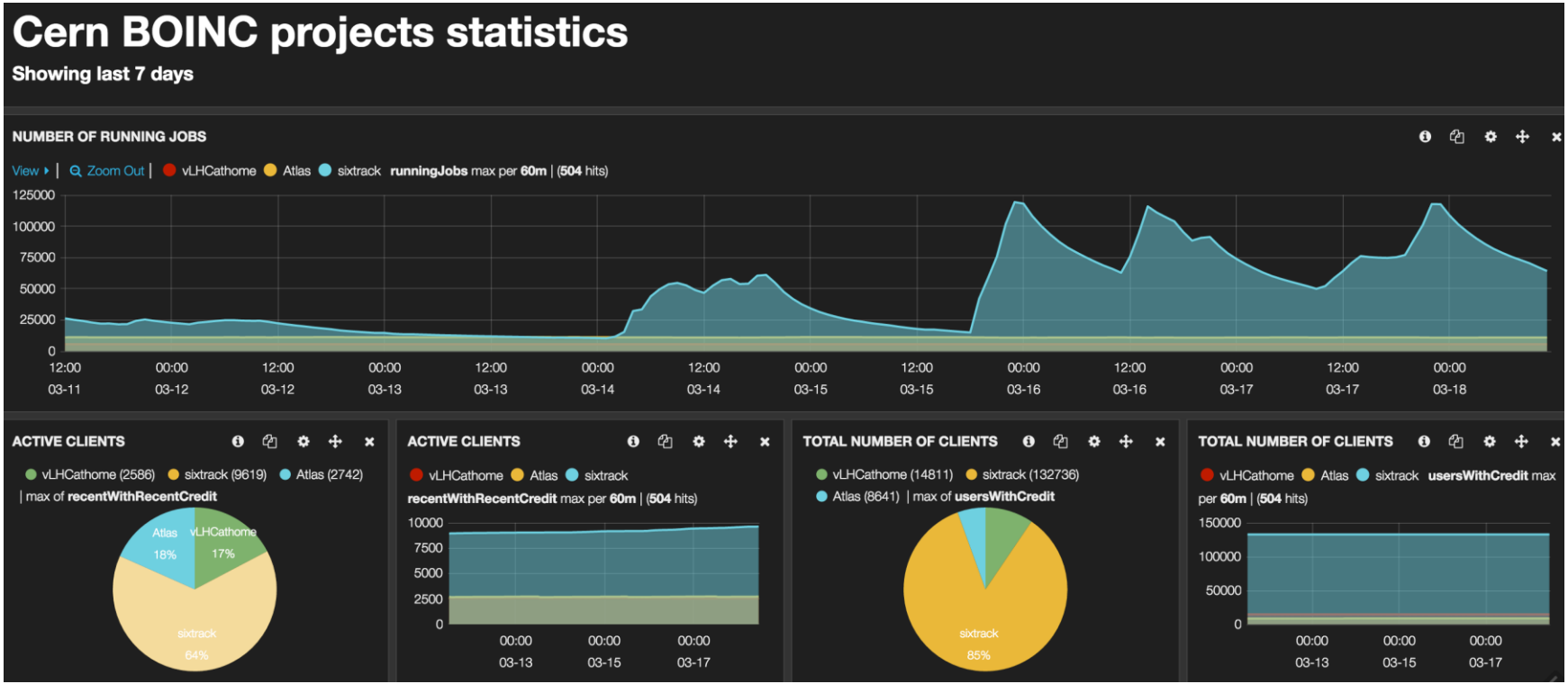
- Non-pledged scavenged resources
  - Volunteers (e.g. home PCs)
  - Institute desktops
  - Small farms with easy deployment
- Unpredictable but significant resources
  - Target CPU bound simulations
- Outreach benefits

ATLAS@HOME

BOINC ATLAS contribution



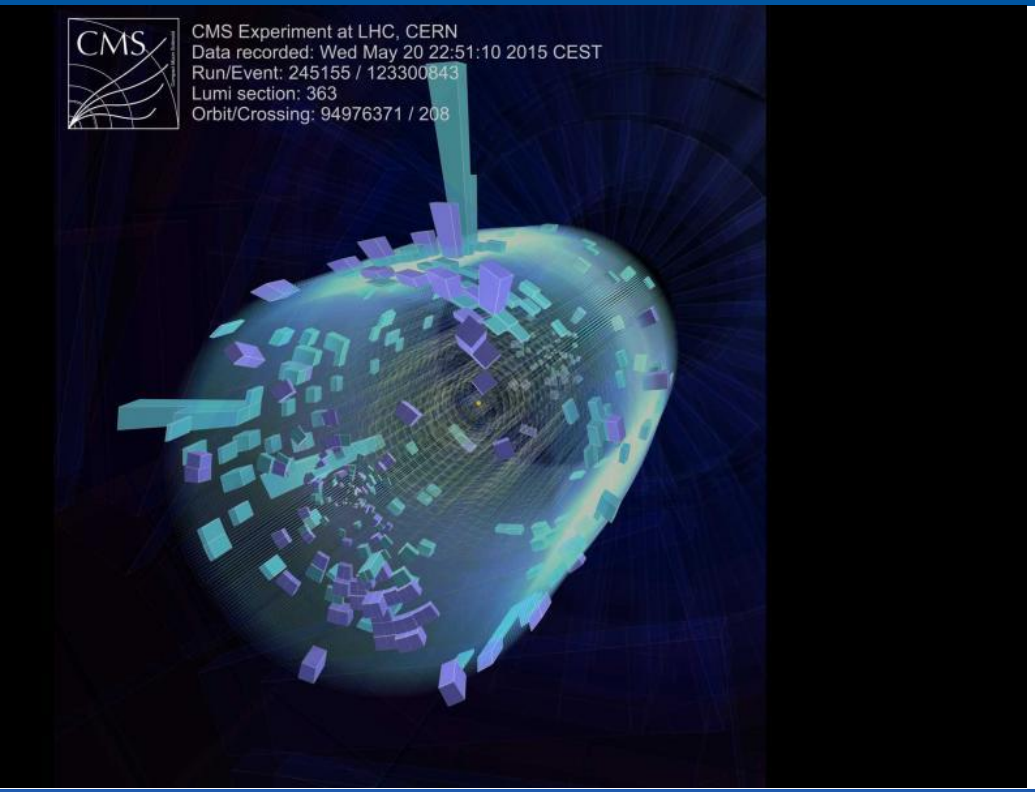
Multiple Projects Running



Summary

- Computing infrastructure is rapidly evolving
  - Clouds
  - Batch
  - Opportunistic
- Challenges likely to continue in
  - Sustainability
  - Scale
  - Efficiency

For Further Information



- Technical details at http://openstack-in-production.blogspot.fr
- Helix Nebula Initiative at http://www.helix-nebula.eu/
- Scientific Working Group at https://wiki.openstack.org/wiki/Scientific_working_group
- Open Research Cloud documents at https://goo.gl/yG7Mzp

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

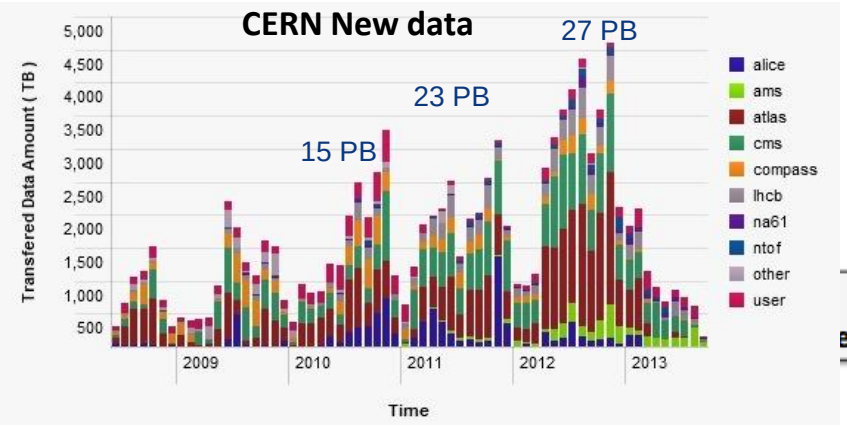
For comparison:
1990's: Total LEP data set ~few TB
Would fit on 1 tape today

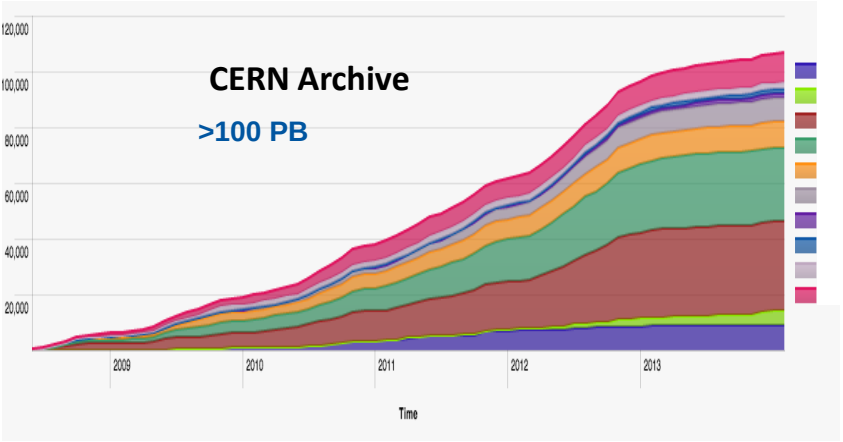
Today: 1 year of LHC data ~27 PB

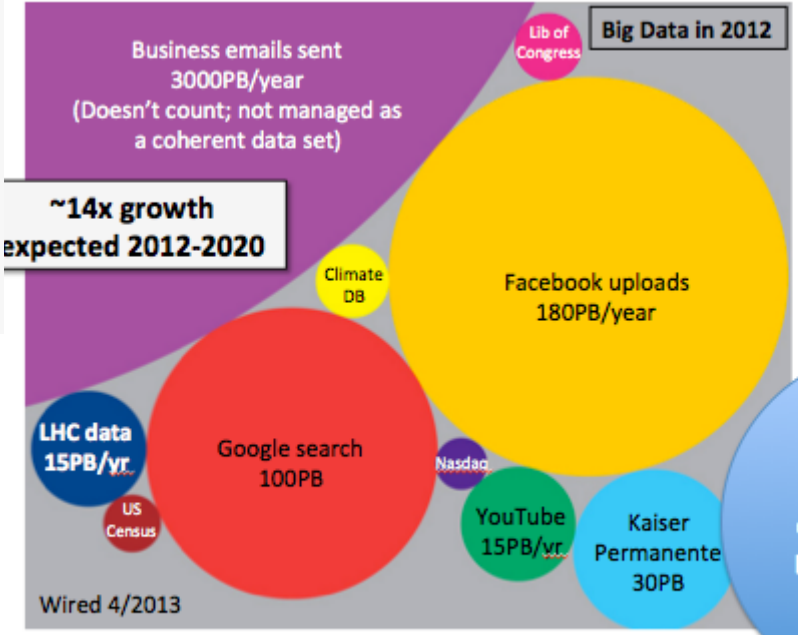
Innovation Dilemma

- 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 ?







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

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

OpenStack Collaborations

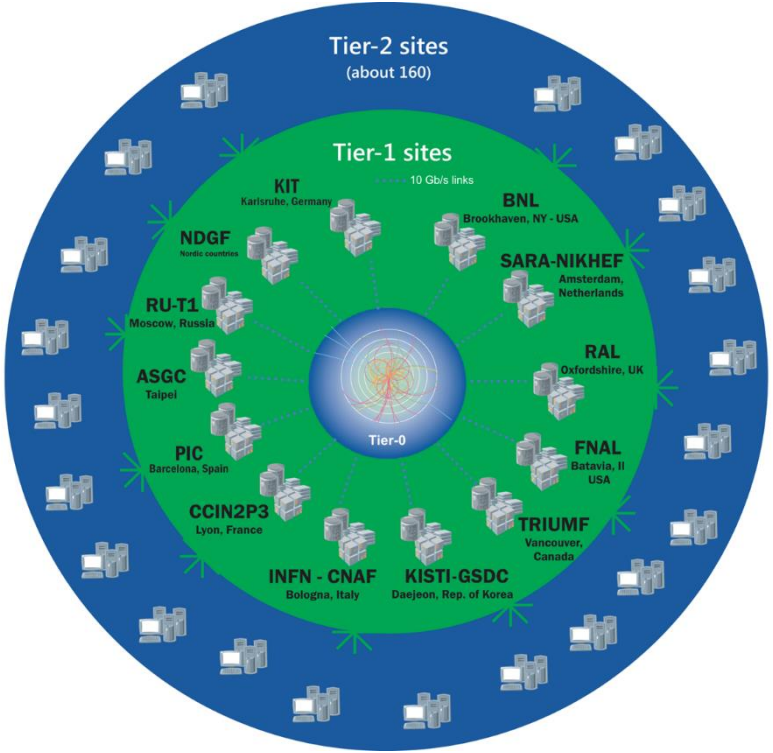
- Large Deployment Team
  - Walmart, Yahoo!, Rackspace, eBay, Paypal, …
- Containers
  - Rackspace, Red Hat
- OpenStack Scientific Working Group
  - Not just academic
  - High Performance and High Throughput

The Worldwide LHC Computing Grid

TIER-0 (CERN):
data recording, reconstruction and distribution

TIER-1:
permanent storage, re-processing, analysis

TIER-2:
Simulation, end-user analysis



nearly 170 sites, 40 countries

~350’000 cores

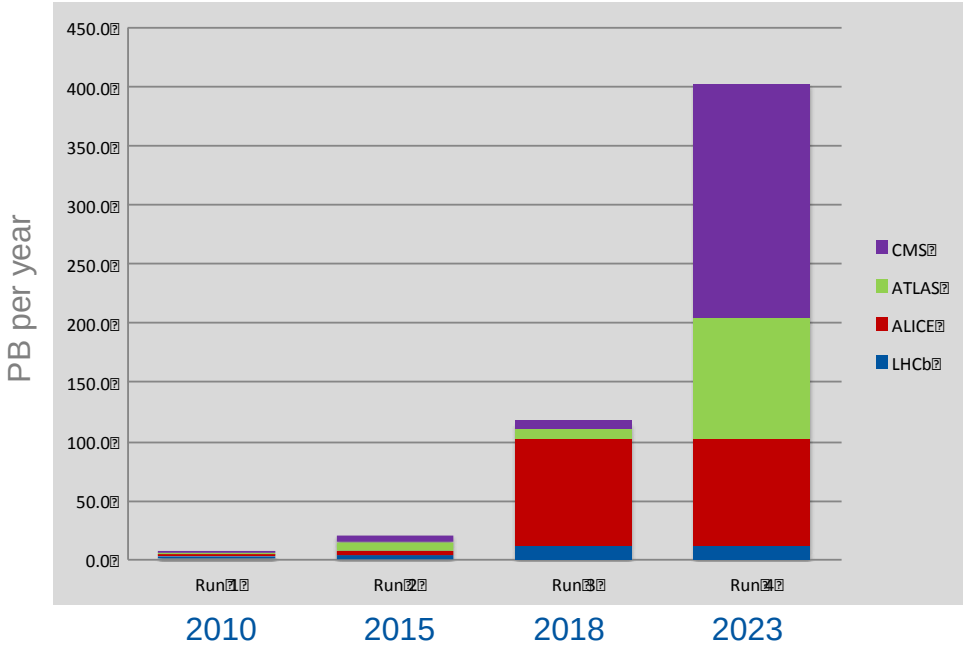
500 PB of storage

> 2 million jobs/day

10-100 Gb links

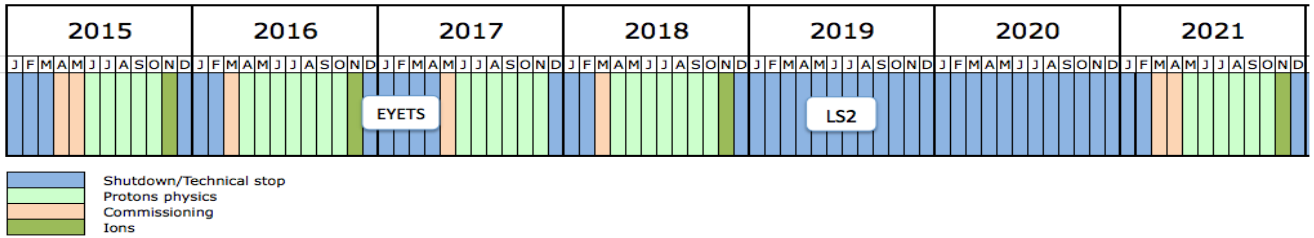
LHC Data Growth

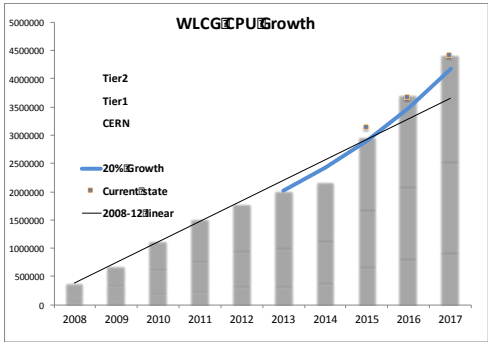
Expecting to record 400PB/year by 2023 with the High Luminosity LHC upgrade

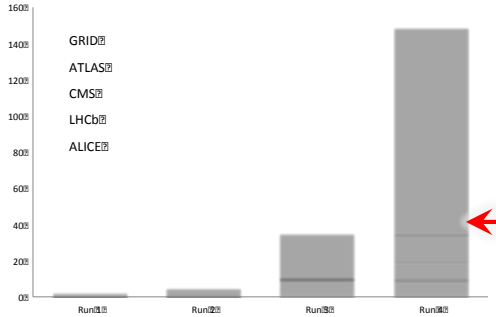


Where is x3 improvement ?

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

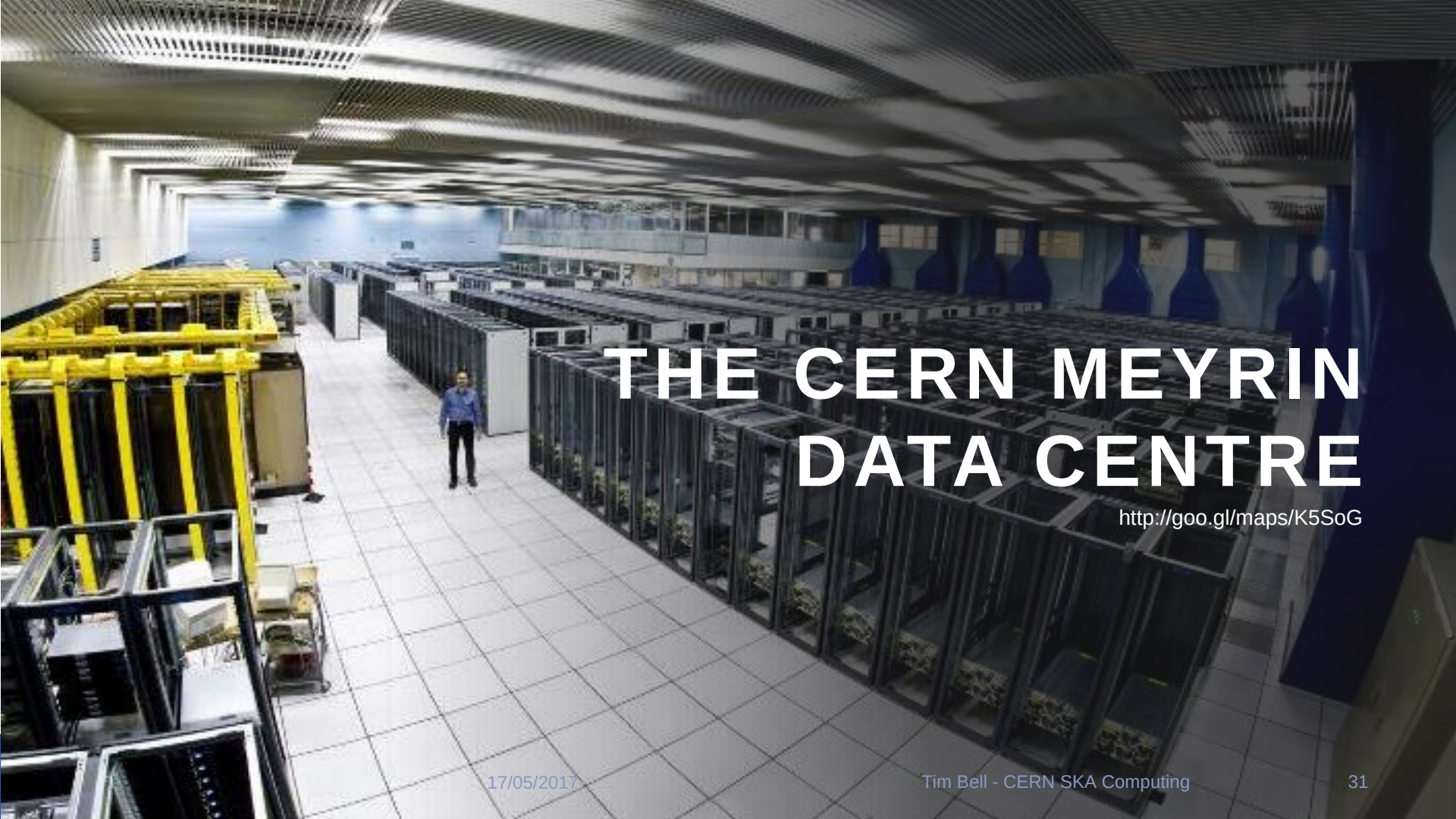






Compute: Growth > x50

What we think is affordable unless we do something differently

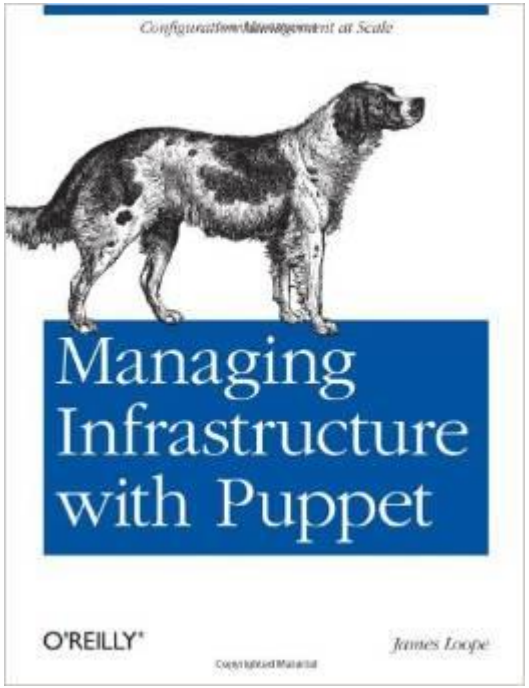


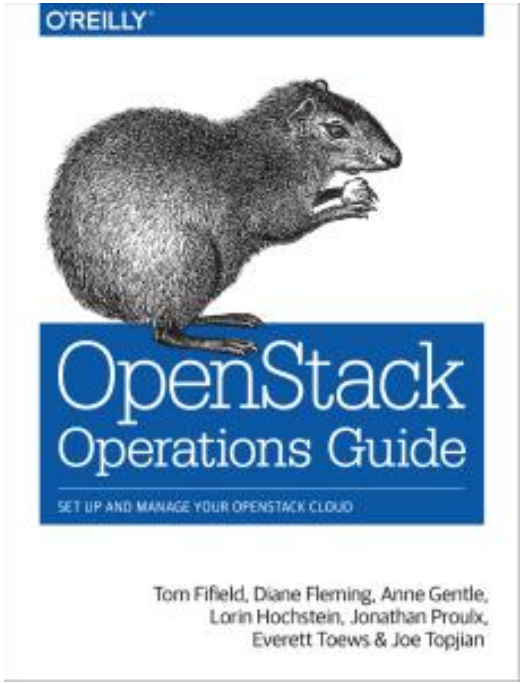
THE CERN MEYRIN DATA CENTRE

http://goo.gl/maps/K5SoG

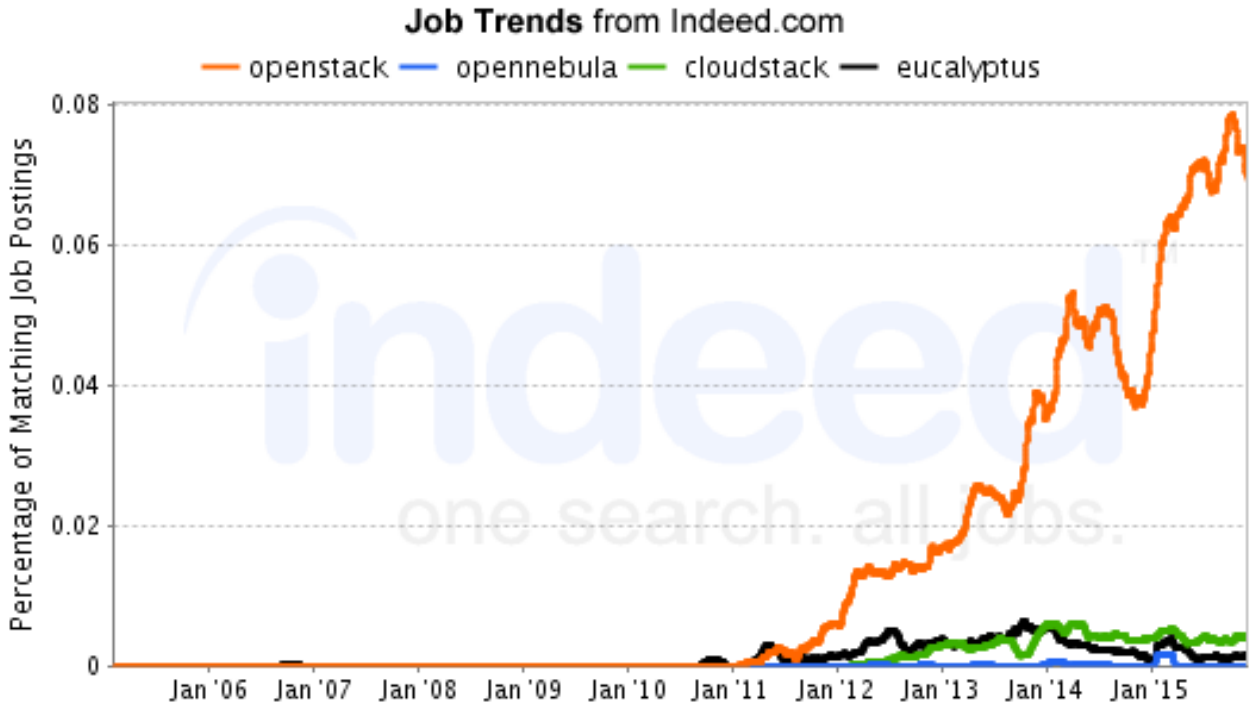

DANTE
100 GbE
T-Systems
100 GbE
Wigner RCP
CERN
Amsterdam
Netherlands
Niedersachsen
Berlin
Poznan
Warsaw
Poland
Antwerpen
Brussels
Belgium
Cologne
Germany
Prague
Czech Republic
Krakow
Luxembourg
Luxembourg
Paris
Slovakia
Munich
Vienna
Bratislava
Budapest
Austria
Zurich
Liechtenstein
Bern
Switzerland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

O'Reilly Consideration





Job Trends Consideration



Open Research Clouds I

- OpenStack Scientific Working Group initiated a discussion at the Barcelona summit in Fall 2016
- “Open Research Cloud” workshop in May 2017 in Boston
- Wide range of participants
  - Universities and research labs (US, Asia, Europe)
  - NRENs / EGI / NSF / Internet 2
  - Public cloud providers

Global Scientific Clouds II

- Challenges
  - Reference architecture
  - Security Considerations (digital rights management, IP protection, privacy sensitive data management, etc…)
  - Authorized Shared Use Facilitation
  - Data Federation Resource federation (network/storage/compute federation)
  - Sharing and Business Facilitation
  - Interoperability and Shared-use Policy
  - Avoiding duplication with other initiatives (e.g. RDA, EGI,…)