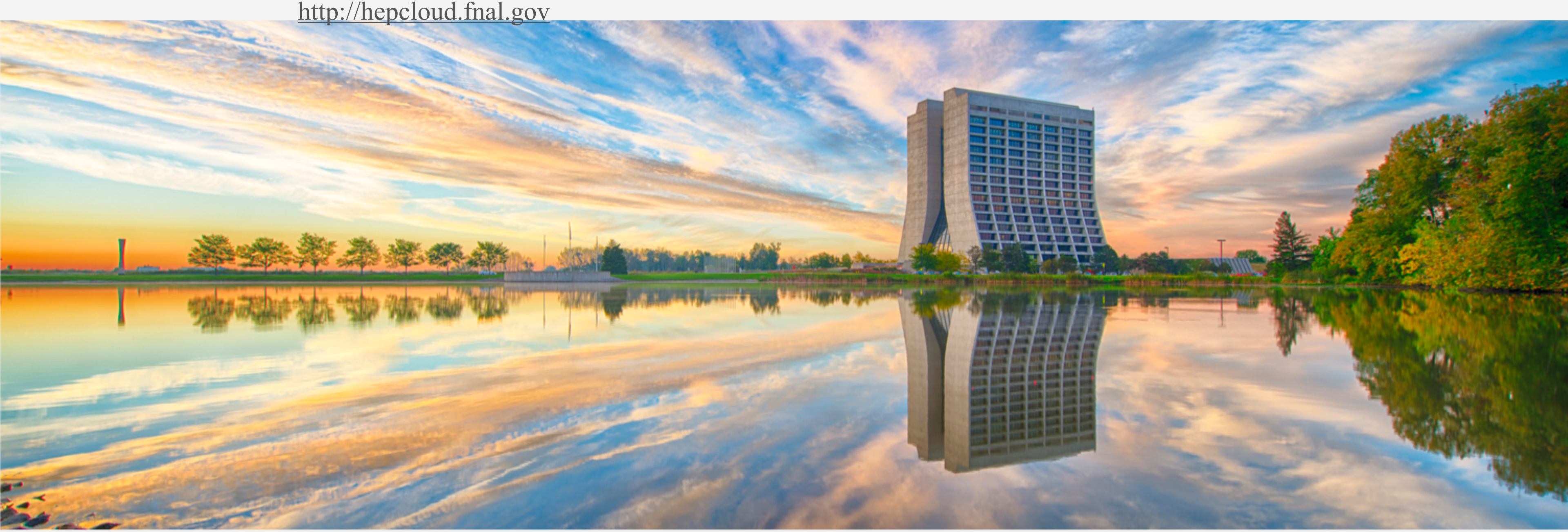
Provisioning 160,000 cores with HEPCloud

Steven Timm, Scientific Computing, Fermilab

http://hepcloud.fnal.gov



Fermilab

50 Years of Discovery

Disclaimer

> “Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Fermilab, the United States Government, or any agency thereof.”

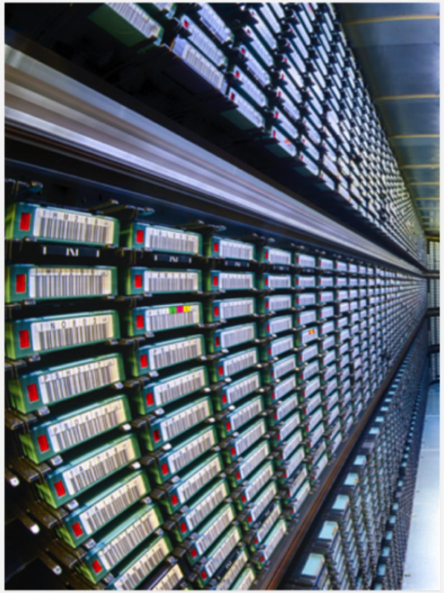
Fermi National Accelerator Laboratory

National Laboratory of Department of Energy
Specialized in High Energy Particle Physics
50 years of service

Located in Batavia, IL

Fermilab facility

48,000 cores
(plus 20,000 HPC)



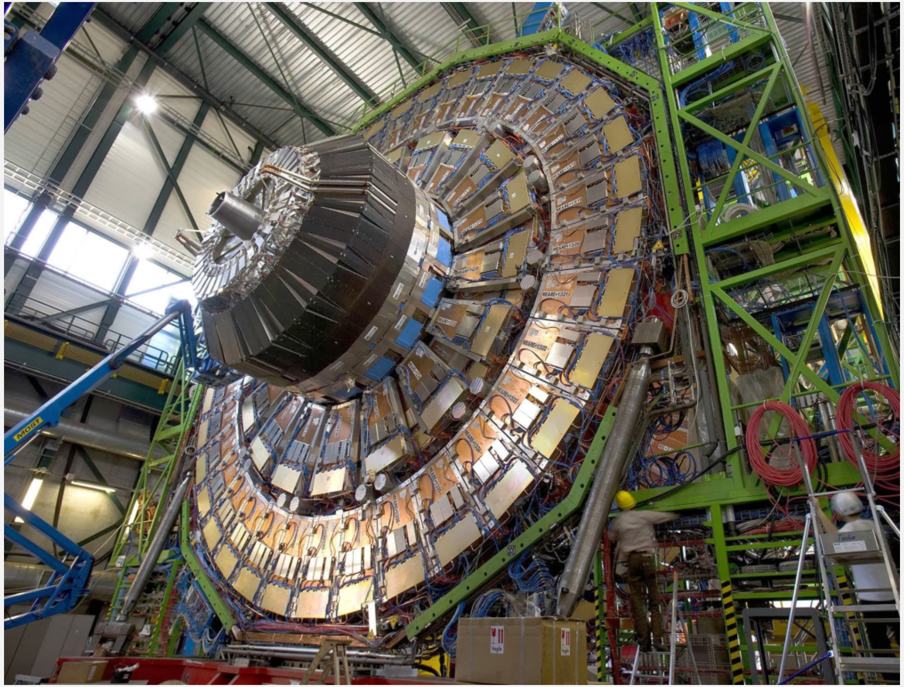
100 PB capacity
(90% full)

35 PB spinning disk in
dCache object store

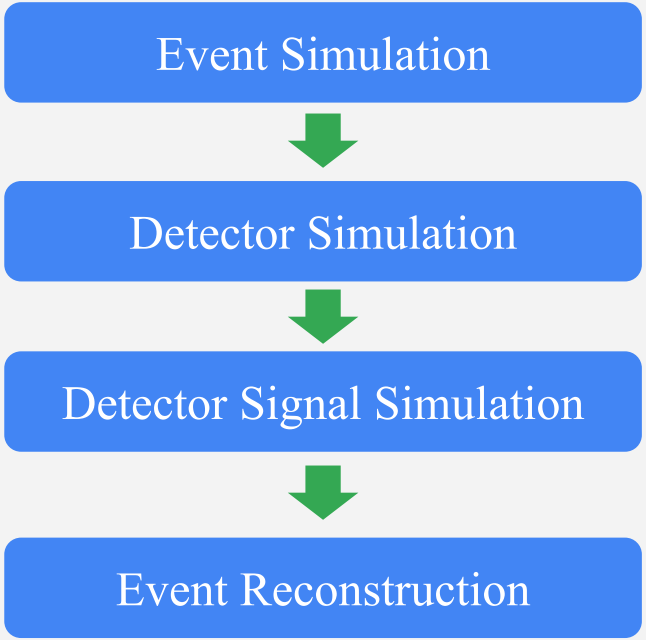
2x100 Gb/s offsite
100 Gb/s peering

Compact Muon Solenoid (CMS) experiment

- **Protons collide** at the LHC **14 million times per second**
- **100 Megapixel** “camera” captures energy, position
  - **1000 times per second**
- All measurements of a collision are called an “**event**”
- Typical ”event” contains many overlapping collisions
- More details were given in James Letts’ talk at this conference.



Simulations - detectors are complicated



Event Simulation

Detector Simulation

Detector Signal Simulation

Event Reconstruction

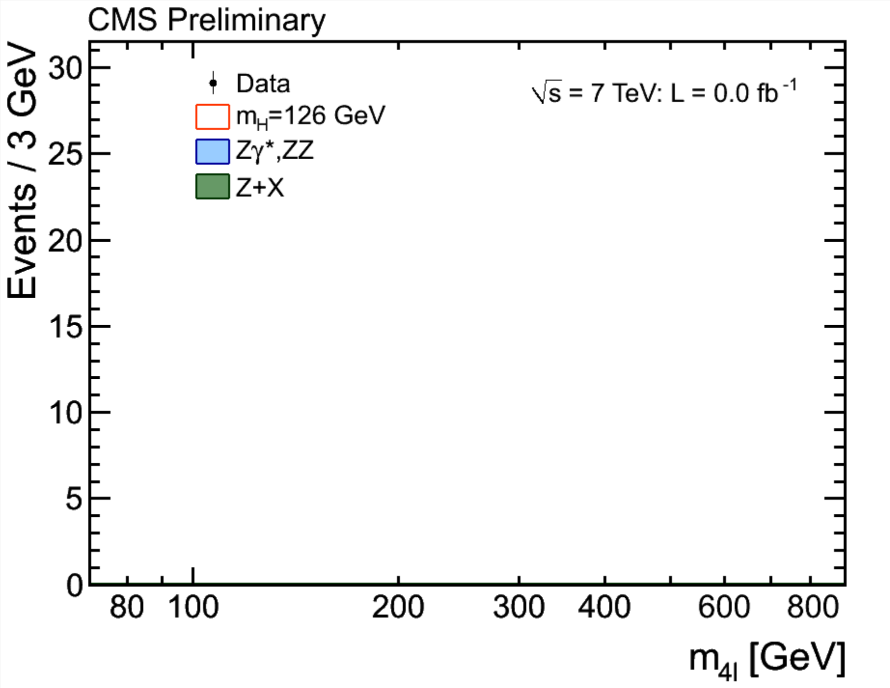
The rules of particle physics are governed by quantum mechanics

- Initial conditions cannot be controlled precisely
- Recorded particle collisions sample a large space of possibilities

We are using **probabilistic** techniques to sample this space in simulation

Analysis of selected corners of the space allow us to compare experiment with simulation and extract physics results

Discovering the Higgs



Detect particle interactions and **compare** to Standard Model

Black dots: measurement

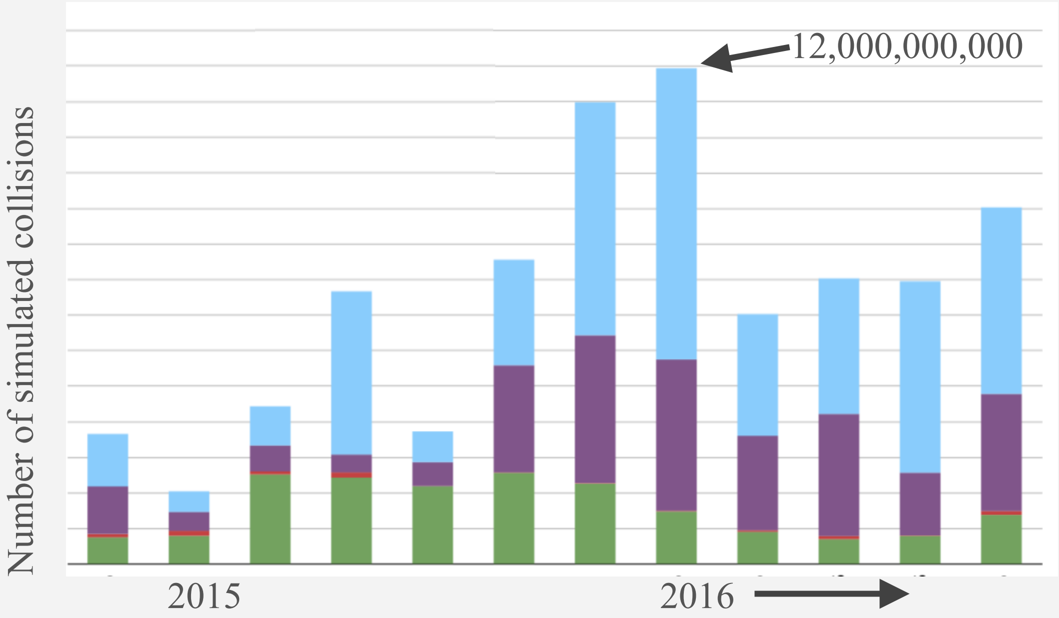
Blue shape: simulation of Standard Model

Red shape: simulation of new theory (in this case the Higgs)

Simulation contains everything we know: the Standard Model and much more

Simulations are vital for Particle Physics

Scale of simulations for the CMS experiment



Each experiment is simulating Billions of collisions per year

Huge thirst for simulation

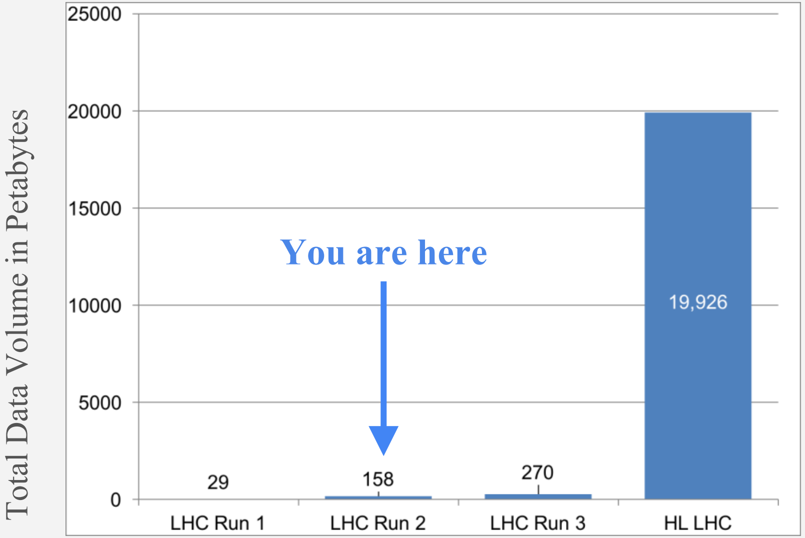
LHC experiments simulate billions of proton-proton collisions per year

Workflows are ~~Embarrassingly~~ **Pleasingly** parallel

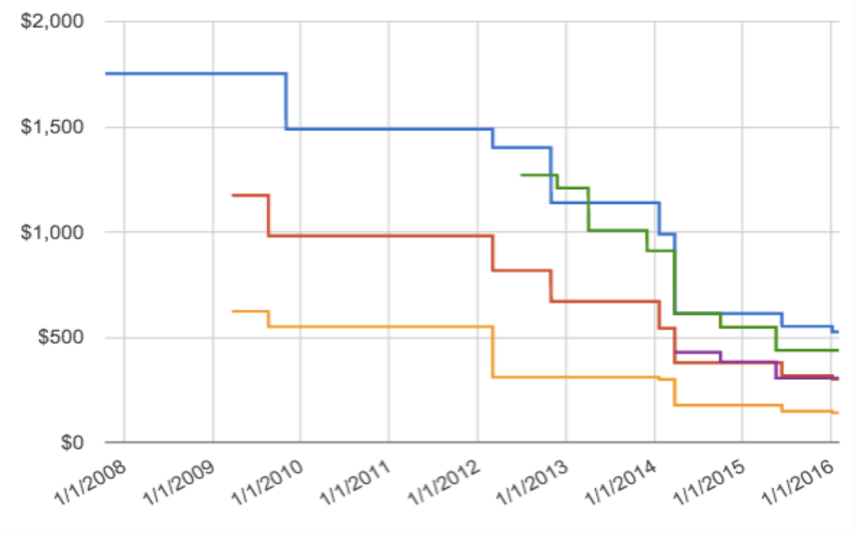
Every event/collision can be simulated separately on its own core

150,000 cores is not enough… !

- High Energy Physics computing will need 10-100x current capacity



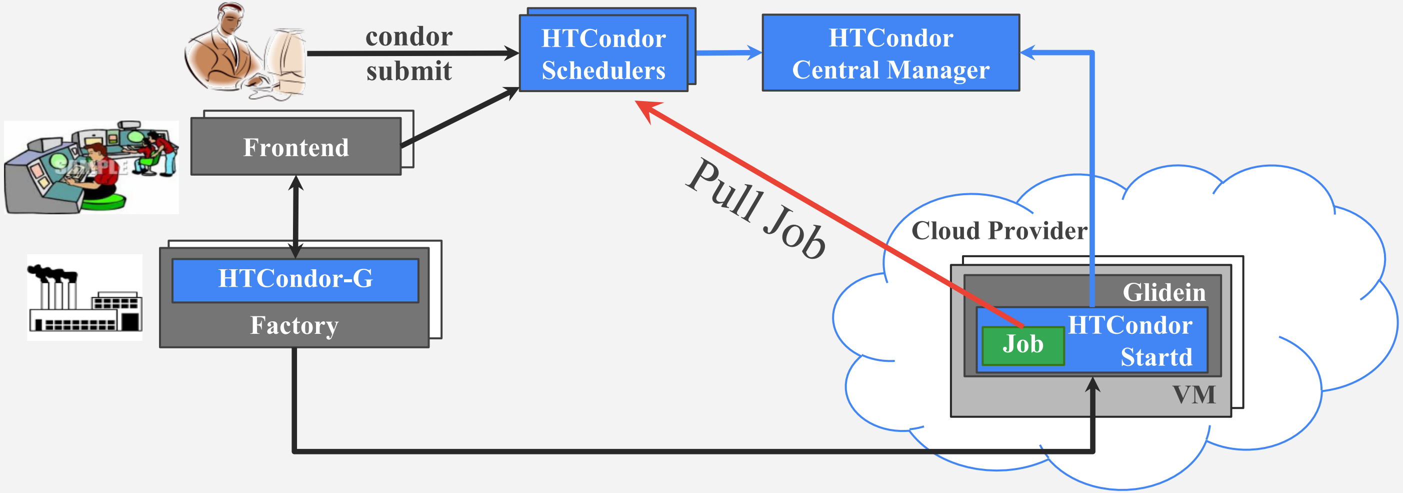
- Scale of industry at or above R&D
  - Commercial clouds offering increased **value** for decreased **cost** compared to the past



Challenge: can we **double** CMS computing?

- **Live demo** during Supercomputing 2016
  - Four days, 12 hours a day
- Expand the Fermilab facility to an additional **160,000** cores
- Use **HEPCloud technology** to do this as transparently as possible to the application

Provisioning remote resources via glideinWMS

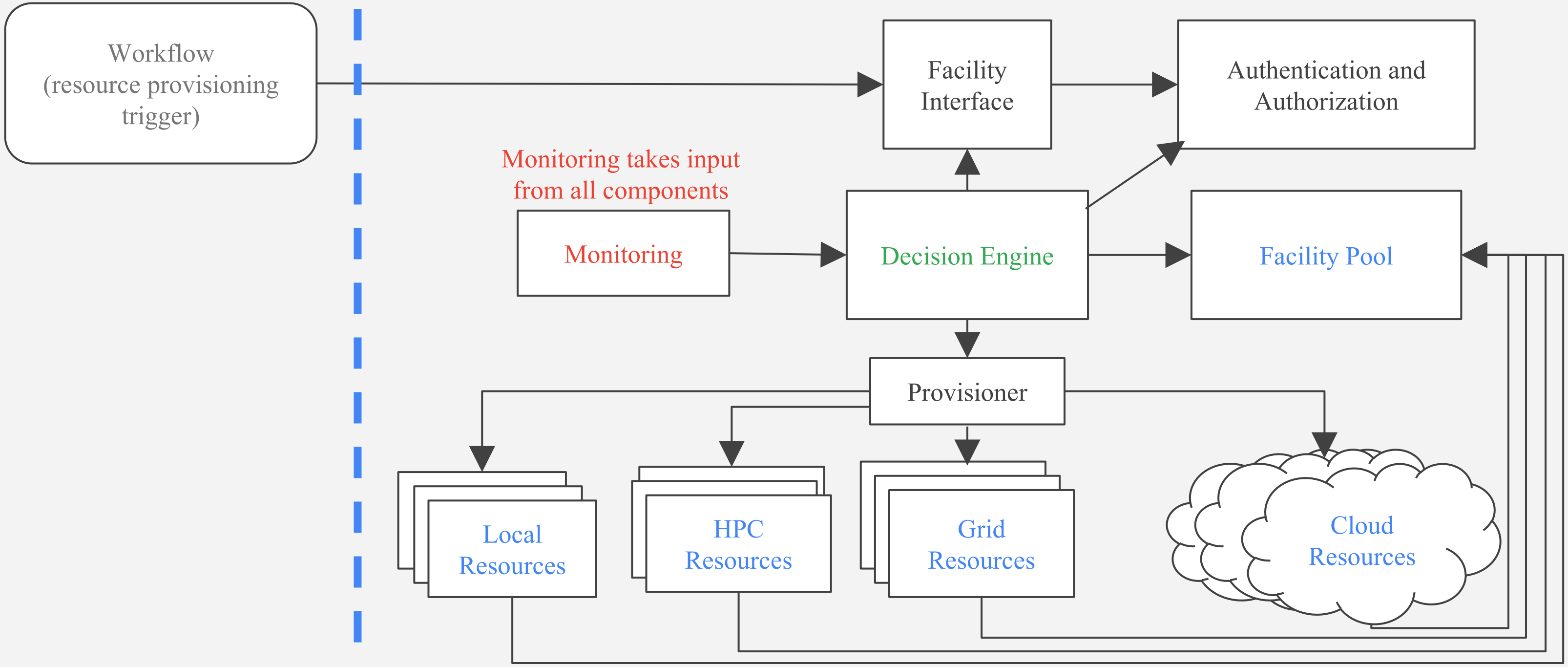


- GlideinWMS submits “**pilot jobs**” to compute resources based on demand
- Pilot jobs execute on the resource and fetch user jobs from a queue
  - Pilot jobs **hide heterogeneity** of compute from the user and **validate environment** (will not start user jobs on bad resource)
  - Goal is to grab resources from wherever possible, as needed.

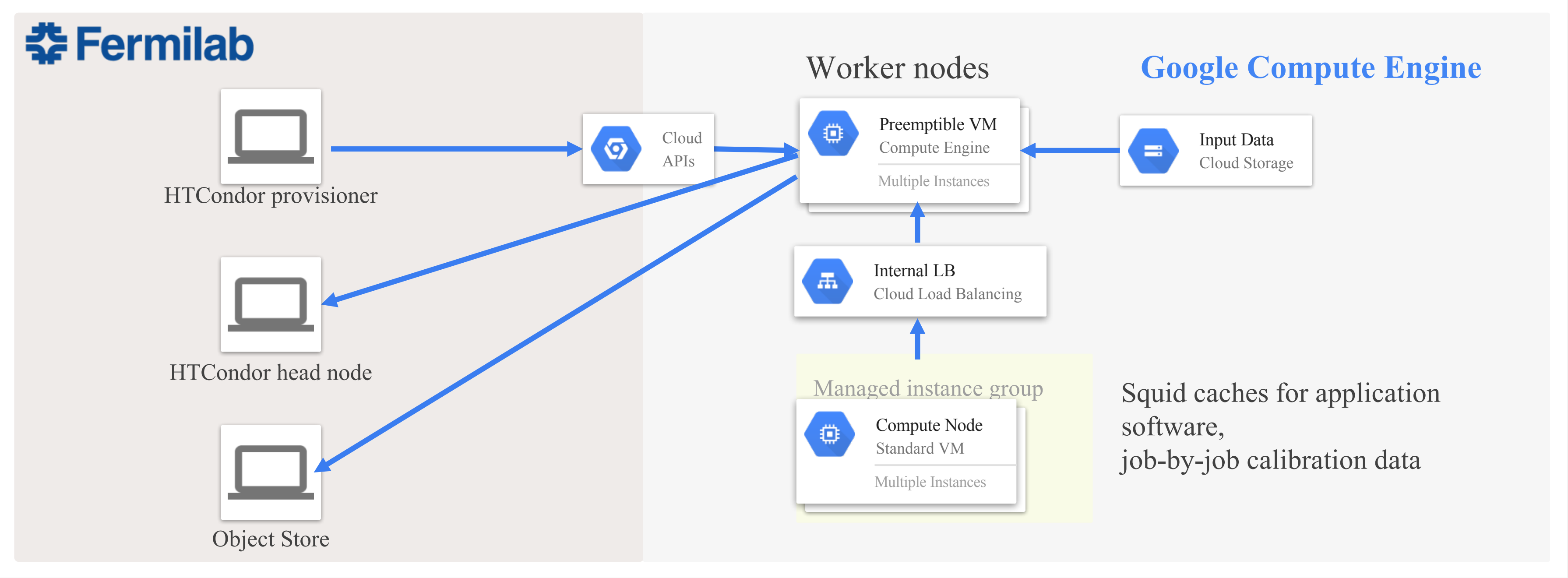
HEPCloud Vision

- HEPCloud is envisioned as a **portal** to an ecosystem of **diverse computing resources**, commercial or academic
  - Provides "complete solutions" to users, with agreed-upon levels of service
  - Routes to **local or remote** resources based on workflow requirements, cost, and efficiency of accessing various resources
  - Manages allocations of users to supercomputing facilities (e.g. NERSC, Argonne, Oak Ridge, …)
- Pilot project to explore feasibility, capabilities of HEPCloud
  - Collaborative effort with industry, academia
  - Goal of moving into production by September 2018

HEPCloud Architecture

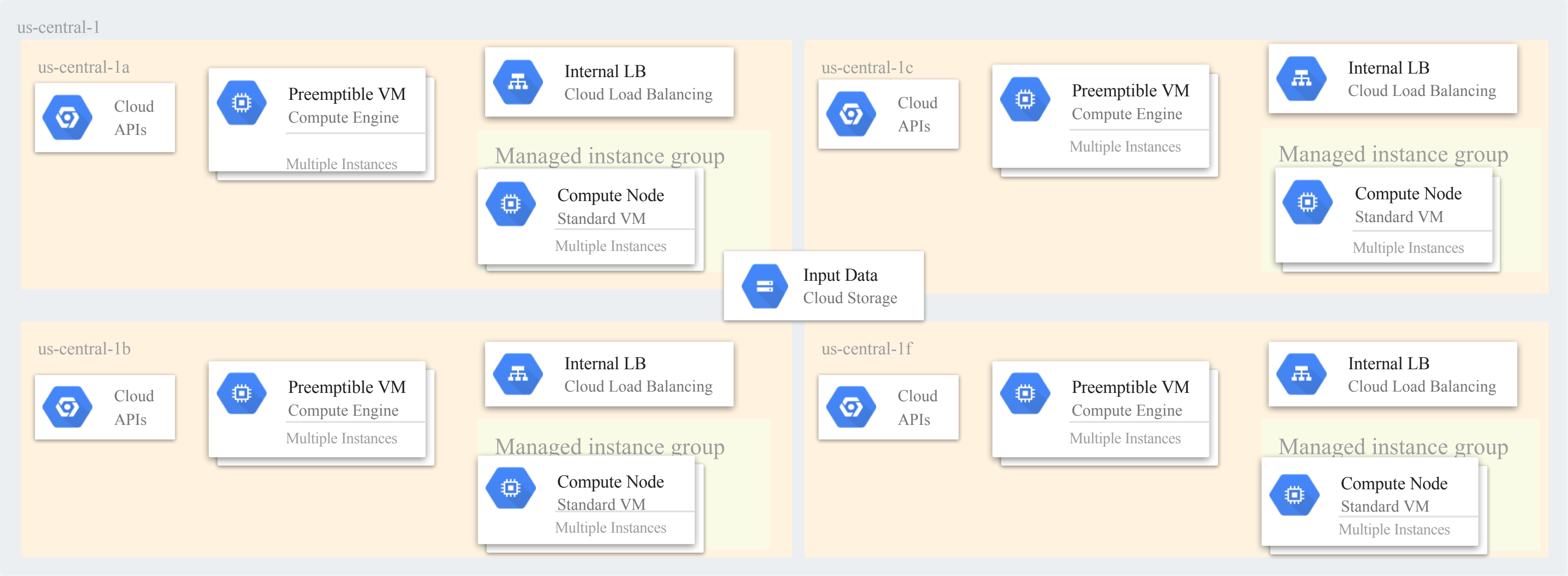


Architecture inside a single zone

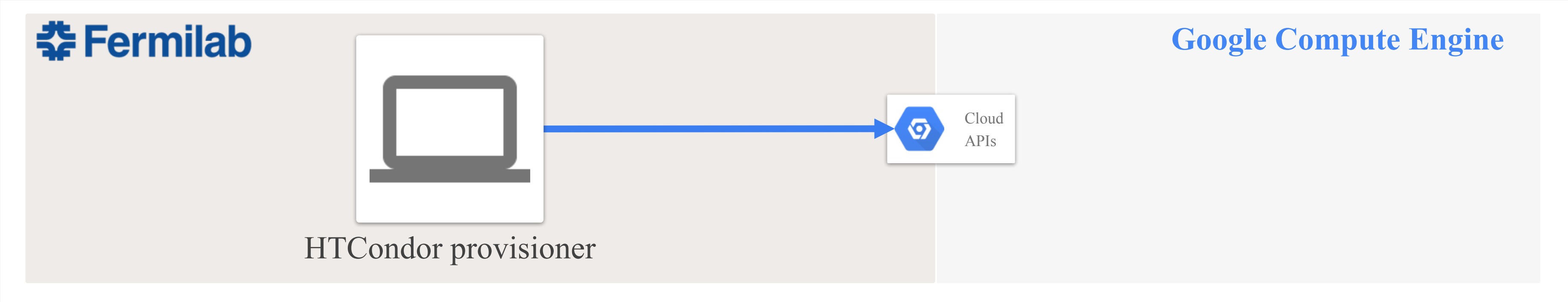


Squid caches for application software,
job-by-job calibration data

Using 4 zones in us-central-1

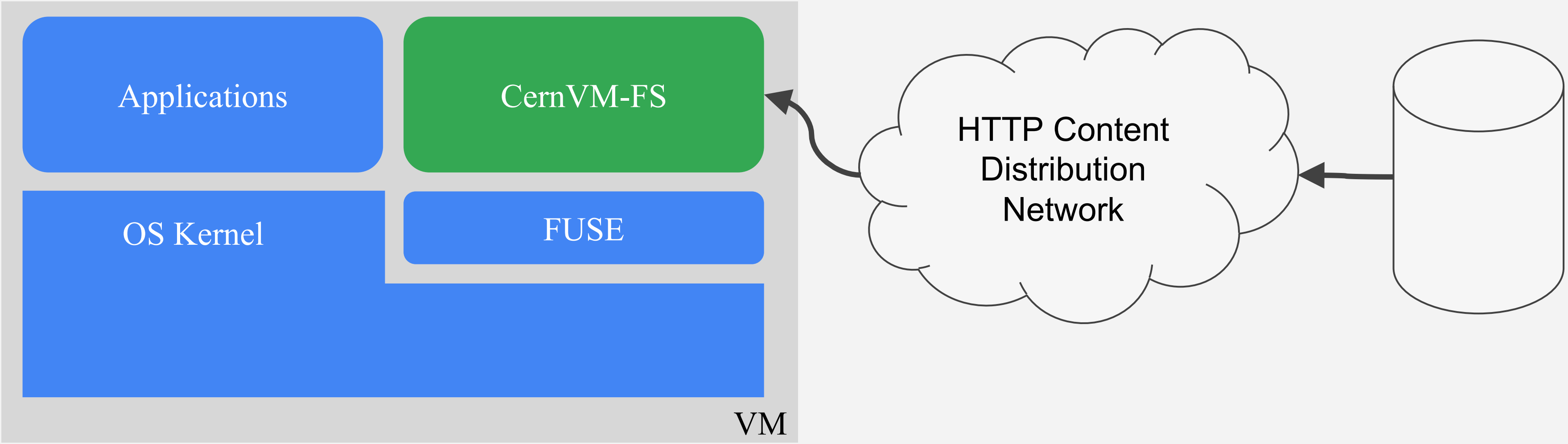


HTCondor: speaking Cloud APIs



- HTCondor provisioner initially written by HTCondor team @ UW-Madison
- Google contributed to the Open Source HTCondor project
  - Added support for **preemptible VMs** and service accounts
  - Note that Google preemptible VM's are a fixed price, last for up to 24hr.
  - Fixed **critical bug** to address scaling

Providing application software in a distributed world



Worker Node
Memory Buffer

Megabytes

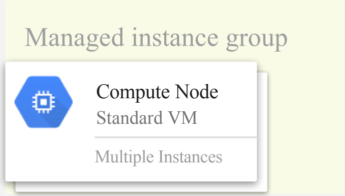
Worker Node
Disk Cache

Gigabytes

Stratum 1
Web Server
Entire Software Stack

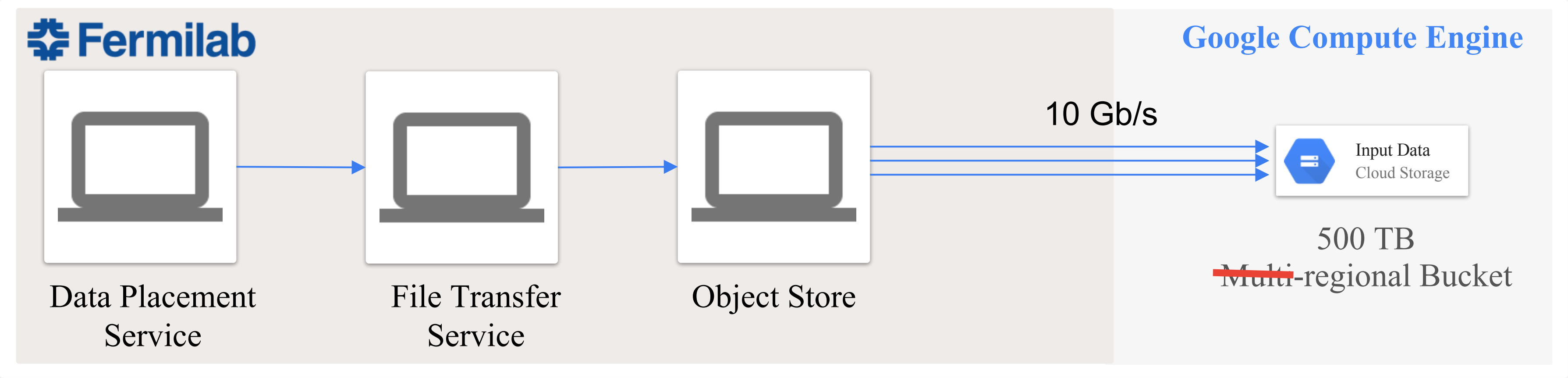
Terabytes

Managed instance group - squid web-cache



- Used for caching both code and remote database queries.
- Internal-facing web-cache
- Internal Load Balancer service
  - Autoscaling when instance/network/sent_bytes > 9 MB/s
- Health checks
  - **Problem**: health checks execute GET /path/to/file and require a leading /, but **squids are proxies** and execute GET http://mysite.com/foo/bar/baz instead
  - **Solution** (hack?): provide squid internal URI /squid-internal-static/icons/anthony-c.gif

Pre-staging input data to Google Cloud Storage

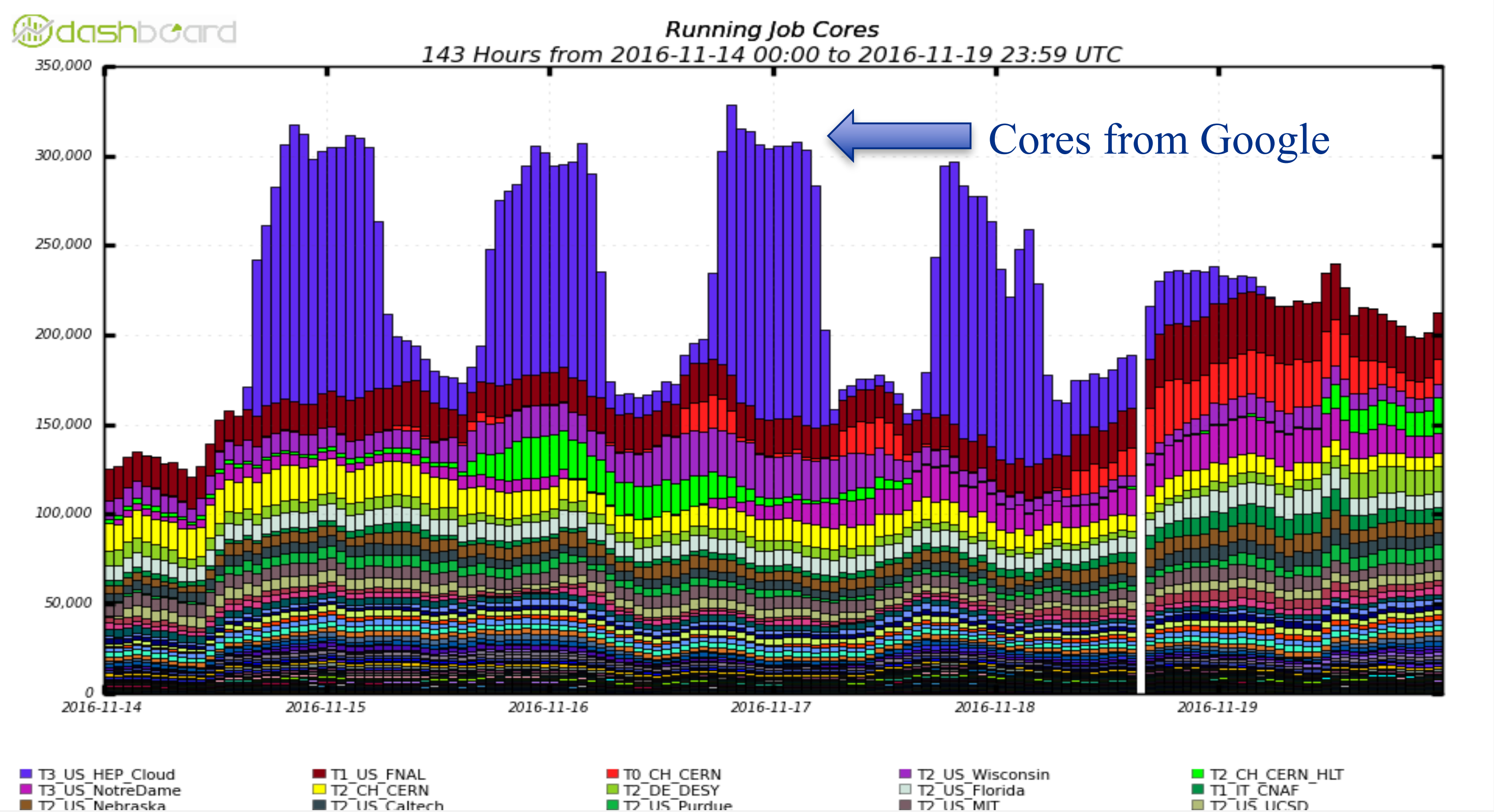


- Experiment-specific data placement service (“PhEDEx”) tracks datasets, schedules transfers
- File Transfer Service supports **S3-compatibility** mode (gfal-copy, davix)
- Google Cloud Storage mounted into preemptible VMs using **gcsfuse** via startup scripts
- Google to ESNet peering (via Equinix) upgraded to **100 Gb/s** capacity

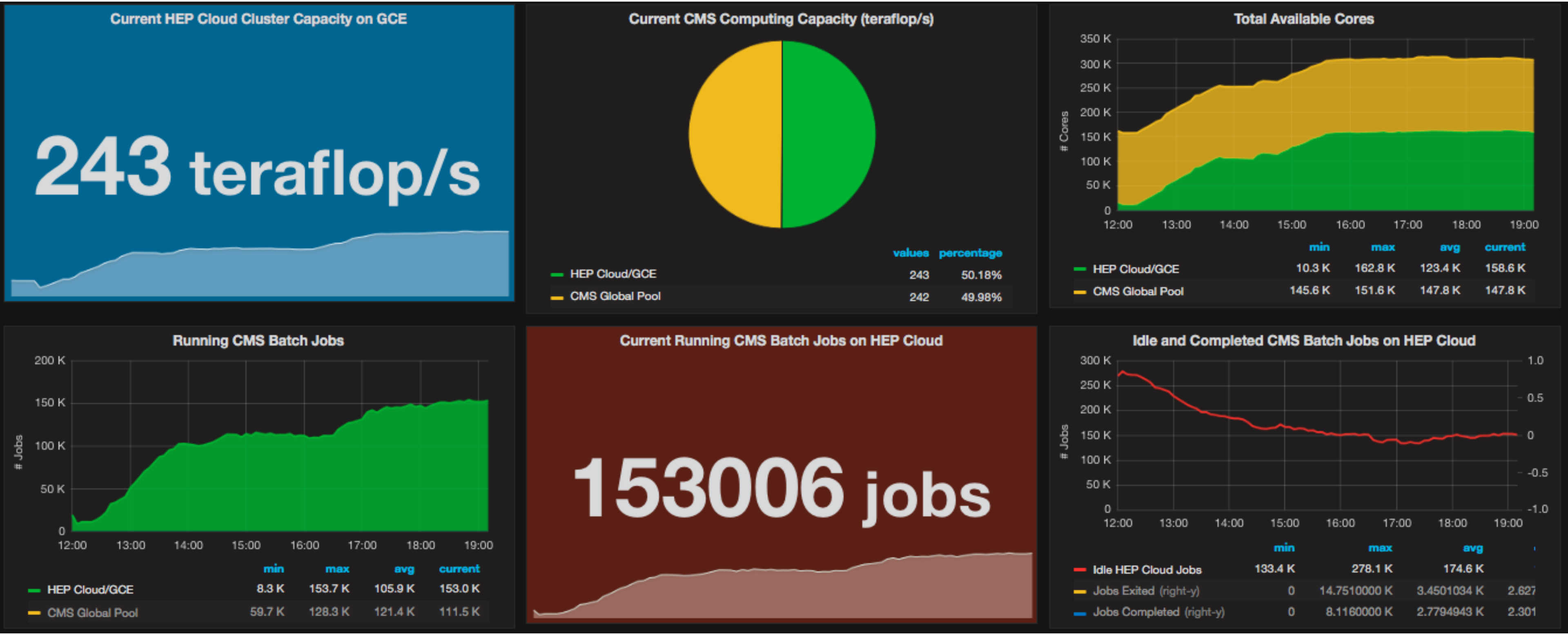
- Converted multi-regional to regional bucket overnight: resulted in 30% less cost

Challenge: can we **double** CMS computing?

How did we do?



Cores from Google



Current HEP Cloud Cluster Capacity on GCE

243 teraflop/s

Current CMS Computing Capacity (teraflop/s)

	values	percentage
🟩 HEP Cloud/GCE	243	50.18%
🟨 CMS Global Pool	242	49.98%

Total Available Cores

	min	max	avg	current
🟩 HEP Cloud/GCE	10.3 K	162.8 K	123.4 K	158.6 K
🟨 CMS Global Pool	145.6 K	151.6 K	147.8 K	147.8 K

Running CMS Batch Jobs

	min	max	avg	current
🟩 HEP Cloud/GCE	8.3 K	153.7 K	105.9 K	153.0 K
🟨 CMS Global Pool	59.7 K	128.3 K	121.4 K	111.5 K

Current Running CMS Batch Jobs on HEP Cloud

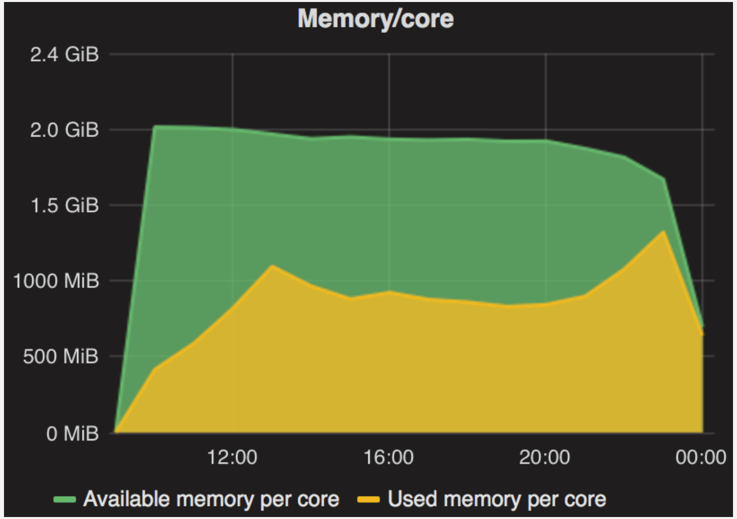
153006 jobs

Idle and Completed CMS Batch Jobs on HEP Cloud

	min	max	avg	
🟥 Idle HEP Cloud Jobs	133.4 K	278.1 K	174.6 K	
🟨 Jobs Exited (right-y)	0	14.7510000 K	3.4501034 K	2.627
🟦 Jobs Completed (right-y)	0	8.1160000 K	2.7794943 K	2.301

Some lessons learned at scale

- Standard VM (3.75 GB) had more memory than the applications need
  - **Custom machine type** with 2 GB
  - 20% cost savings
- Bug in HTCondor provisioning code
  - Ignoring the pagination API
  - Only triggered above **500 VMs**!
  - **Patch provided by Google**
- Expanded subnet from **4096** to **16384 IPs** gcloud compute networks subnets expand-ip-range
  - But had firewall rule on the squid caches: Allow-internal-squid 10.128.0.0/20 tcp:3128



Observations in a high-turnover dynamic pool

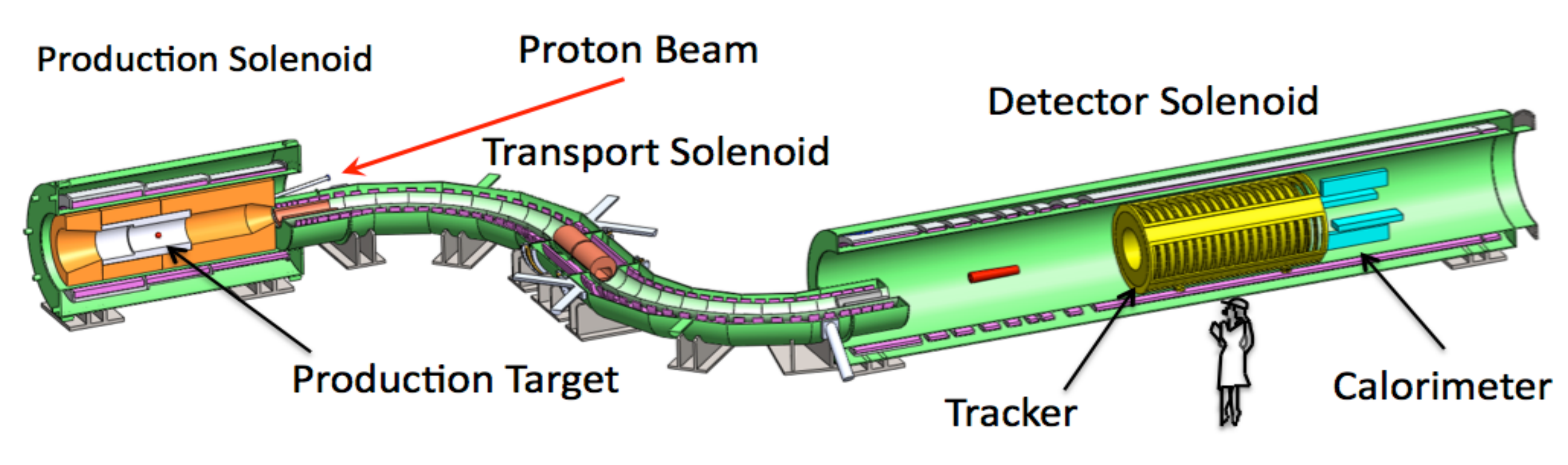
- These are single-core jobs matched to single-core dynamic slots
- condor_status takes 2 minutes to come back! (8.4.x version collector)
- Had 8 schedd’s at peak each running 20000 jobs.
- On average we are matching 9000 slots per negotiation cycle, more if there was a pre-emption burst.
- That’s more than any one schedd can start during that time.
    - Observed some matches time out and get rematched
    - Tuning our autoclustering would have helped this.
- Accesses to storage tend to be peaked in time.
- One CMS workflow “Madgraph” uncompresses a 500MB tarball to ~9GB, 1M files
    - Try that 32 times on same node synchronously, see what happens
    - Troublesome for any local disk, bare metal @FNAL or on any cloud.

Tale of the tape

- **6.35 M** wallhours used; **5.42 M** wallhours for completed jobs.
  - **730172** simulation jobs submitted; only 47 did not complete
  - Most wasted hours during ramp-up as we found and eliminated issues; goodput was at **94%** during the last 3 days.
- Costs on Google Cloud during Supercomputing 2016
  - **$71k** virtual machine costs
  - $8.6k network egress
  - $8.5k magnetic persistent disk (attached to VMs)
  - $3.5k cloud storage for input data
- **205 M** physics events generated, yielding **81.8 TB** of data
- Cost: ~**1.6** cents per core-hour (on-premises: 0.9 cents per core-hour assuming 100% utilization)

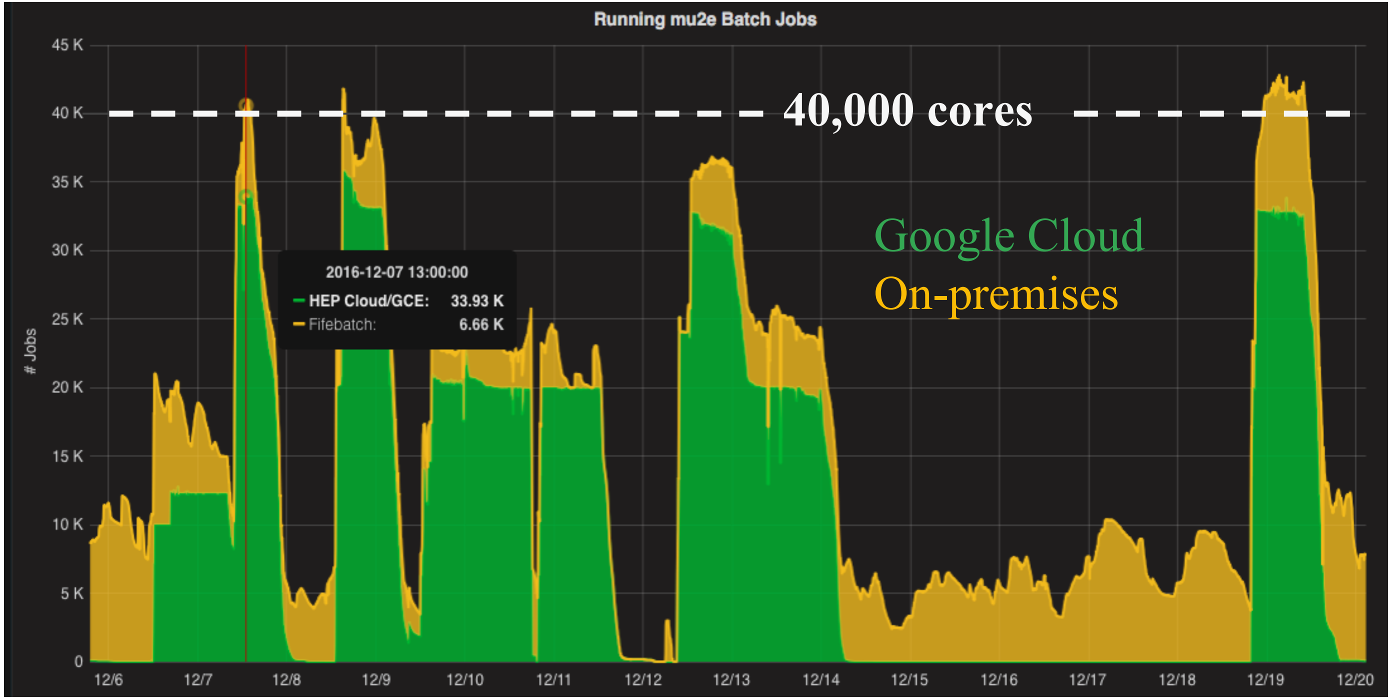
How quickly can we on-board a new user?

"**Mu2e**" experiment being designed: measure **rare decays** of muons to electrons



Simulating different placement and geometry of detector components

Mu2e on-boarded in less than a day



Next steps

- HEPCloud moves into production in September 2018
    - Decision engine (when and how much to provision) is in development
        - Data structures—how to store the information the Decision Engine needs
        - Rule-based engines—what is best one to use, how to set it up.
- Supercomputers at Department of Energy Facilities
    - Already provisioning cycles on Edison, Cori at NERSC
- Additional commercial cloud providers
    - Done: Google Cloud Platform, Amazon Web Services
    - Next: Microsoft Azure, ?
- Non-pleasingly parallel problems
    - Deep learning
    - New architectures

Thanks

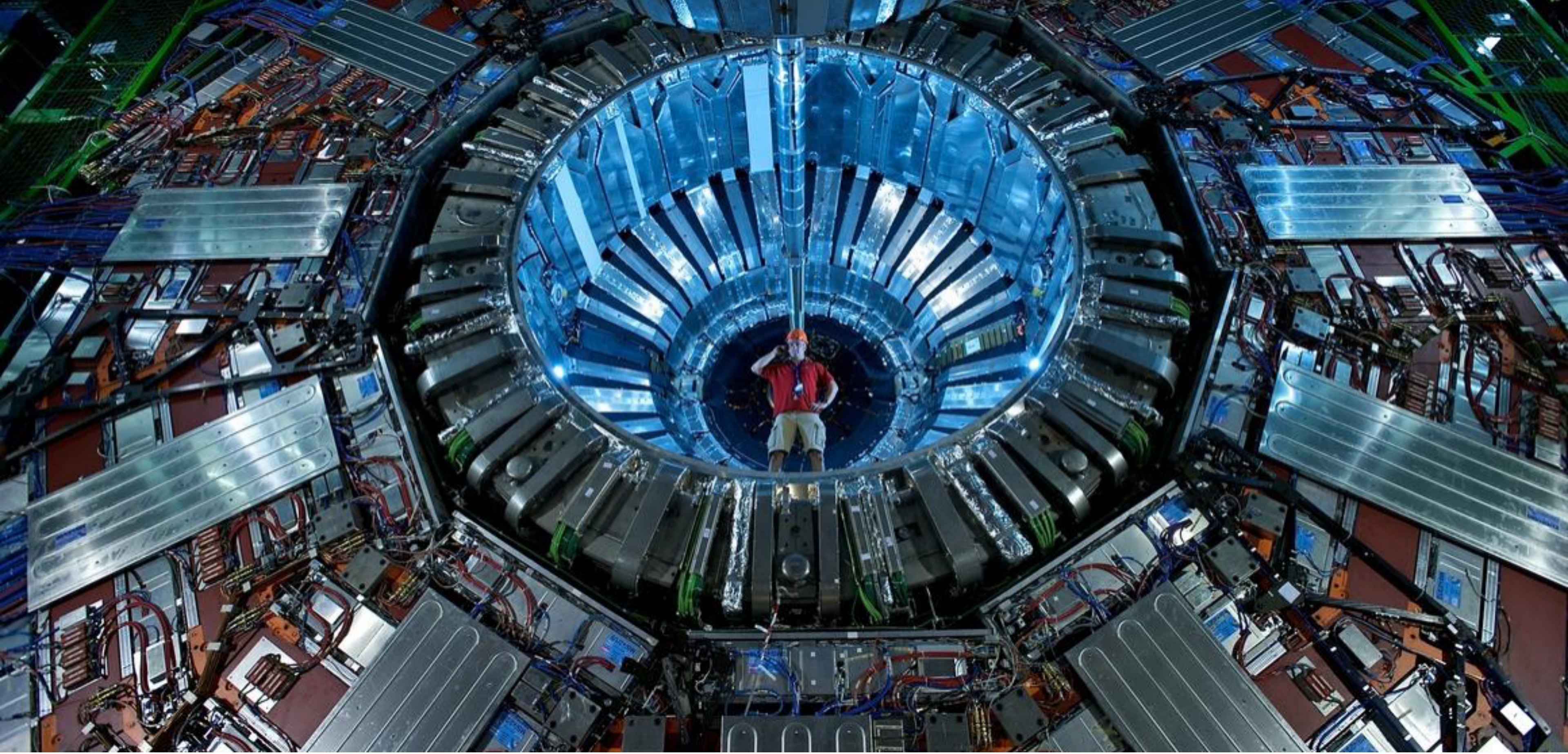
- **The Fermilab team**: Joe Boyd, Stu Fuess, Gabriele Garzoglio, Dirk Hufnagel, Hyun Woo Kim, Rob Kennedy, Krista Majewski, David Mason, Parag Mhashilkar, Neha Sharma, Panagiotis Spentzouris, Steve Timm, Anthony Tiradani, Burt Holzman
- The **HTCondor** and **glideinWMS** projects
- **Open Science Grid**: they provide the software packaging and tooling underneath distributed computing
- **Energy Sciences Network**
- **The Google team**: Michael Basilyan, Karan Bhatia, Solomon Boulos, Sam Greenfield, Paul Nash, Paul Rossman, Doug Strain
- **Resellers**: Onix

2017 is Fermilab's 50th Anniversary!

Visit http://50.fnal.gov/ for anniversary-related events and content

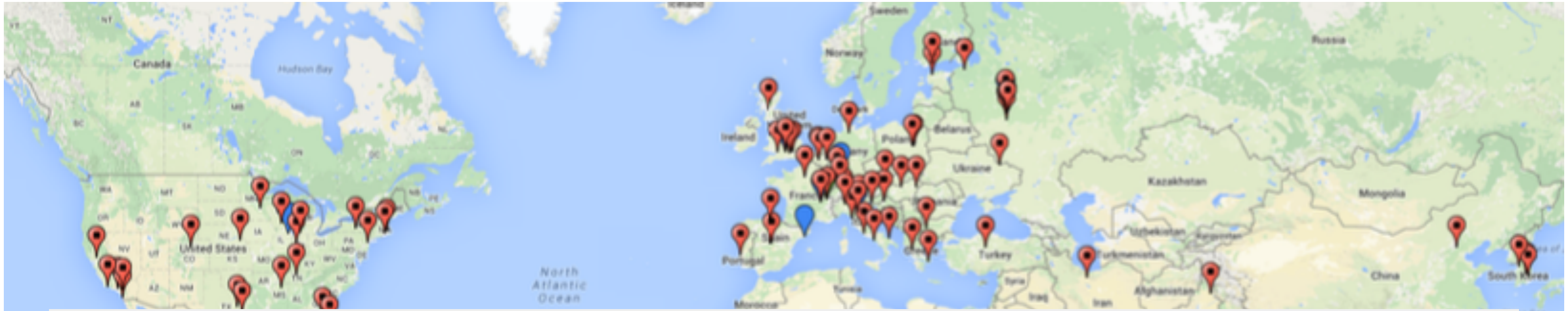
- **June 7**: 50th Anniversary Symposium
- **June 15**: Social media birthday celebration
- **September 23**: Public Open House and Innovation Fair
- …and much more!

Backup slides



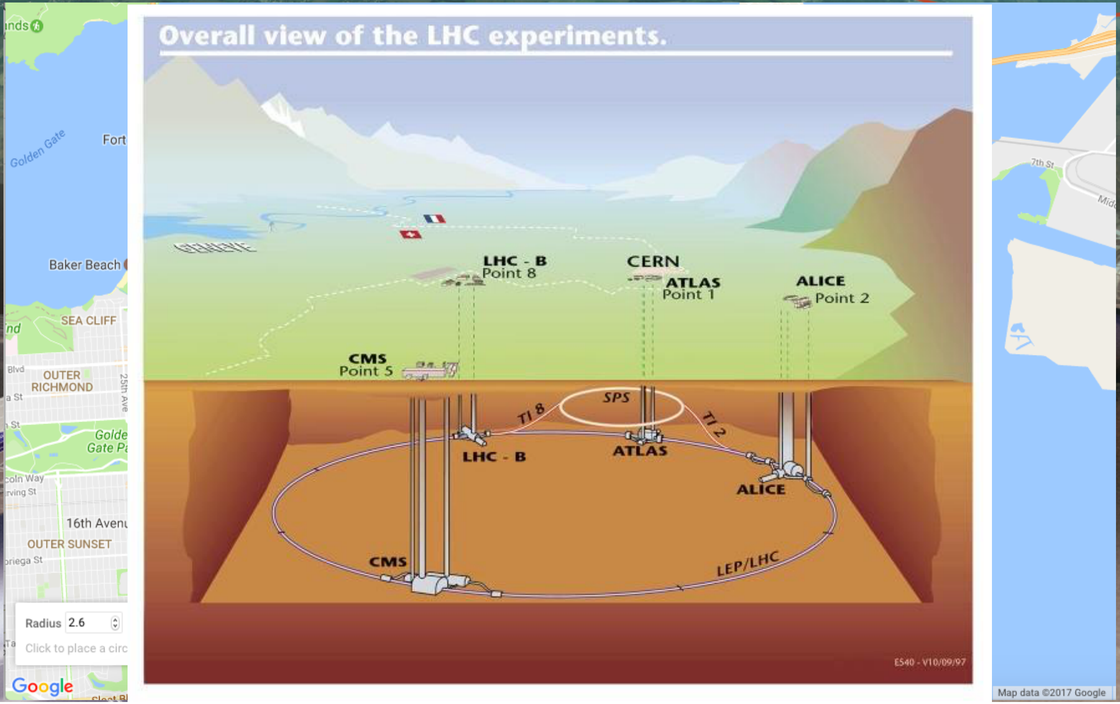
Fermilab
50 Years of Discovery

Global computing for CMS



- 70+ compute clusters (**Open Science Grid** and **Worldwide LHC Computing Grid**)
  - 150,000 cores
  - ~75 Petabyte Disk
  - ~100 PB used tape space
- Strong networks connecting the individual sites
  - Weekly transfer volume between all sites: 4-6 Petabyte
  - Total LHC Trans-Atlantic network capacity: 340 Gigabits per second

Large Hadron Collider in Geneva, Switzerland


Overall view of the LHC experiments.
GENEVE
LHC - B
Point 8
CERN
ATLAS
Point 1
ALICE
Point 2
CMS
Point 5
SPS
TI 8
TI 2
LHC - B
ATLAS
ALICE
CMS
LEP/LHC
E540 - V10/09/97
Golden Gate
Fort
Baker Beach
SEA CLIFF
OUTER RICHMOND
25th Ave
Golden Gate Pa
16th Avenu
OUTER SUNSET
7th St
Radius 2.6
Click to place a circ
Google
Map data ©2017 Google