HEPCloud: Provisioning 160,000 Compute Cores for Science

Dr. Burt Holzman, for the Fermilab HEPCloud Team

American Physical Society, Division of Particles and Fields

August 3, 2017

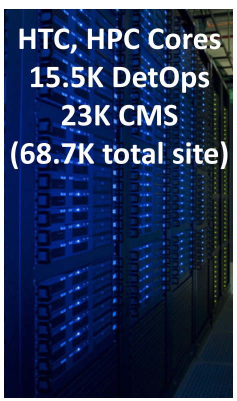
Disclaimer

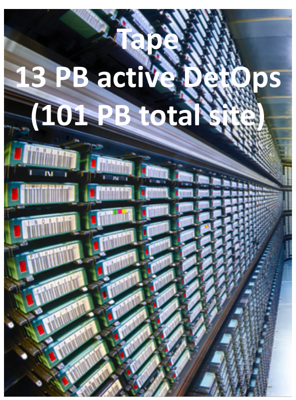
"Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof."

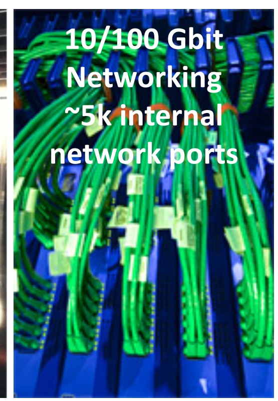
Changing Roles of HEP Facilities

- Strategic Plan for U.S. Particle Physics (P5 Report)

  Rapidly evolving computer architectures and increasing data volumes require effective crosscutting solutions that are being developed in other science disciplines and in industry. Mechanisms are needed for the continued maintenance and development of major software frameworks and tools for particle physics and long-term data and software preservation, as well as investments to exploit next-generation hardware and computing models.



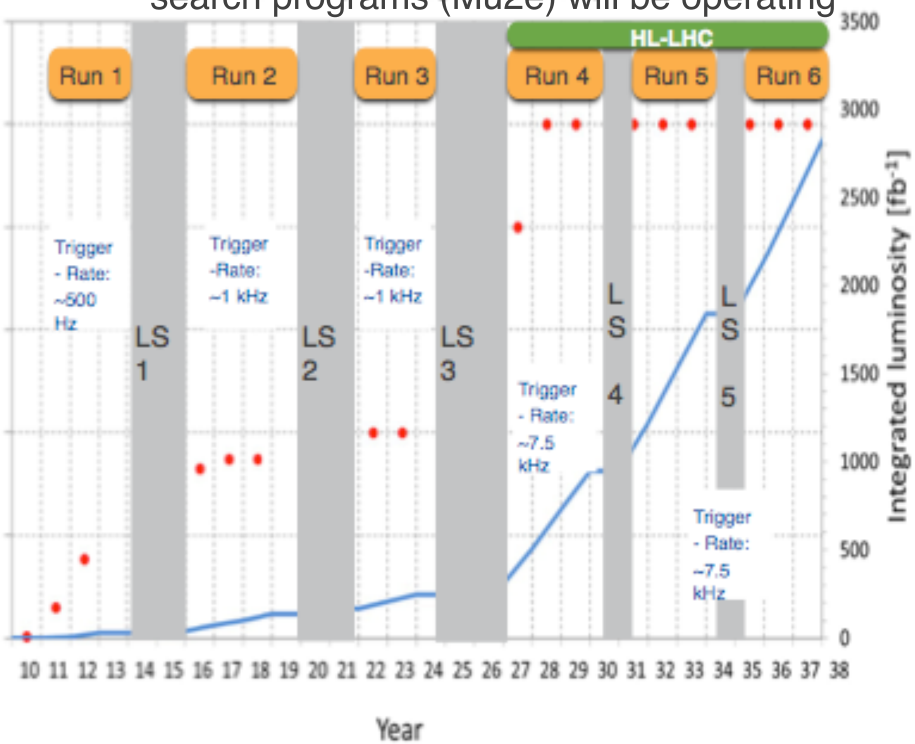




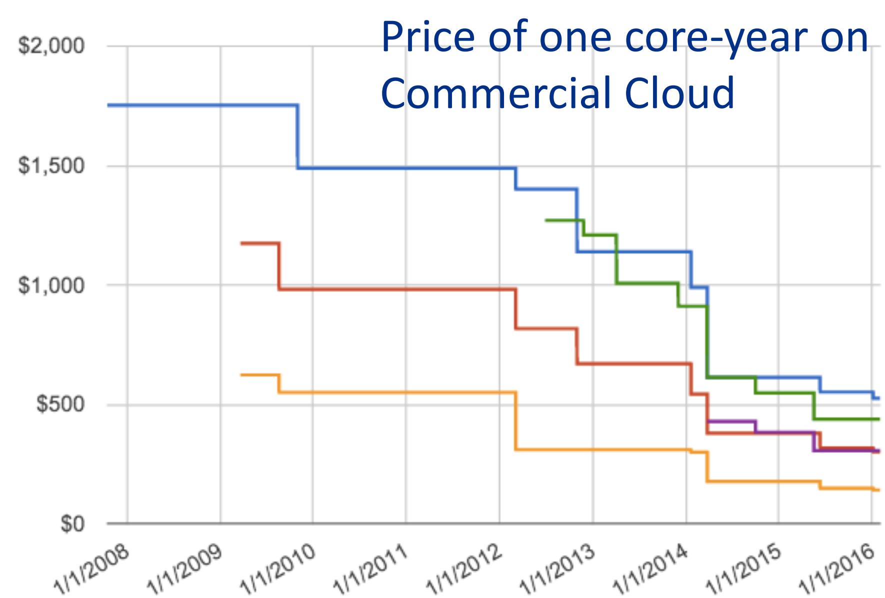
- Need to evolve the facility beyond present infrastructure

Drivers for Evolving the Facility: Capacity and Cost

- High Energy Physics computing needs will be 10-100x current capacity
  - Two new programs coming online (DUNE, High-Luminosity LHC), while new physics search programs (Mu2e) will be operating

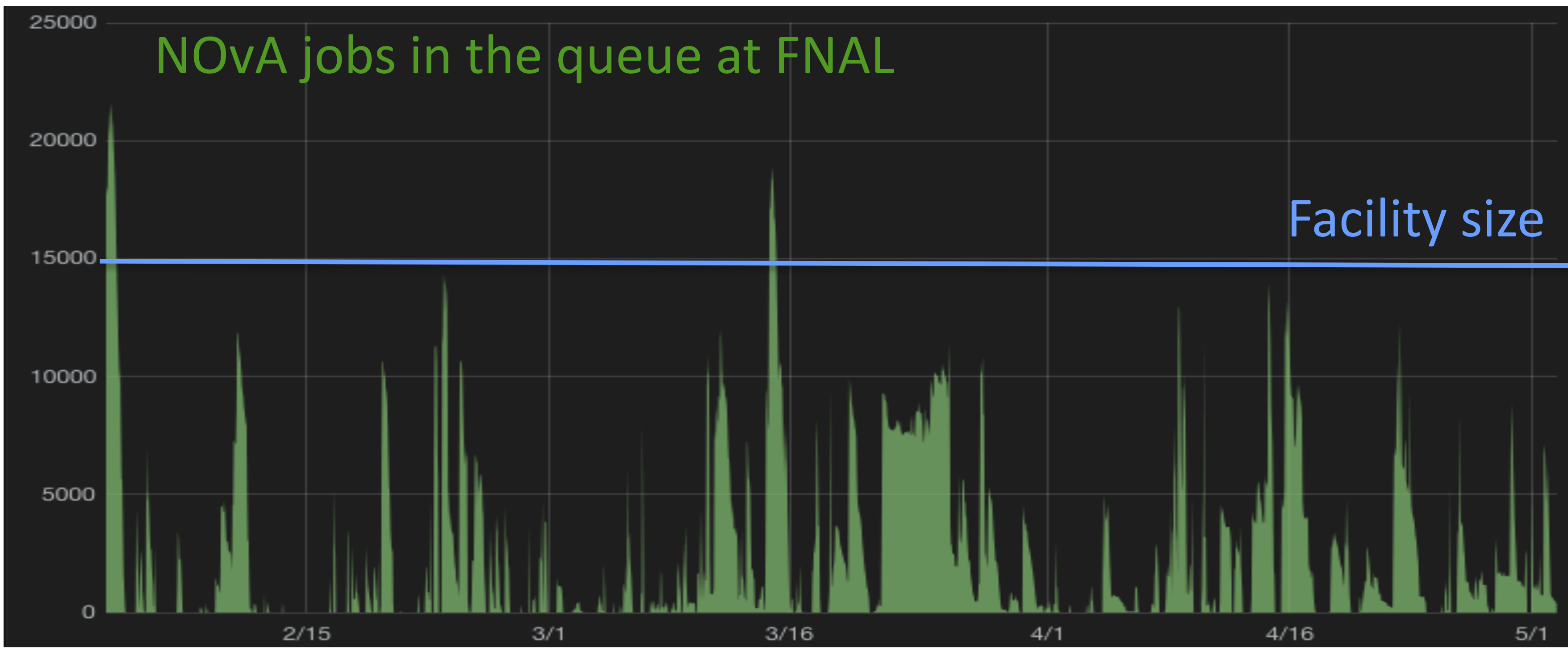


- Scale of industry at or above R&D
  - Commercial clouds offering increased **value** for decreased **cost** compared to the past



Drivers for Evolving the Facility: Elasticity

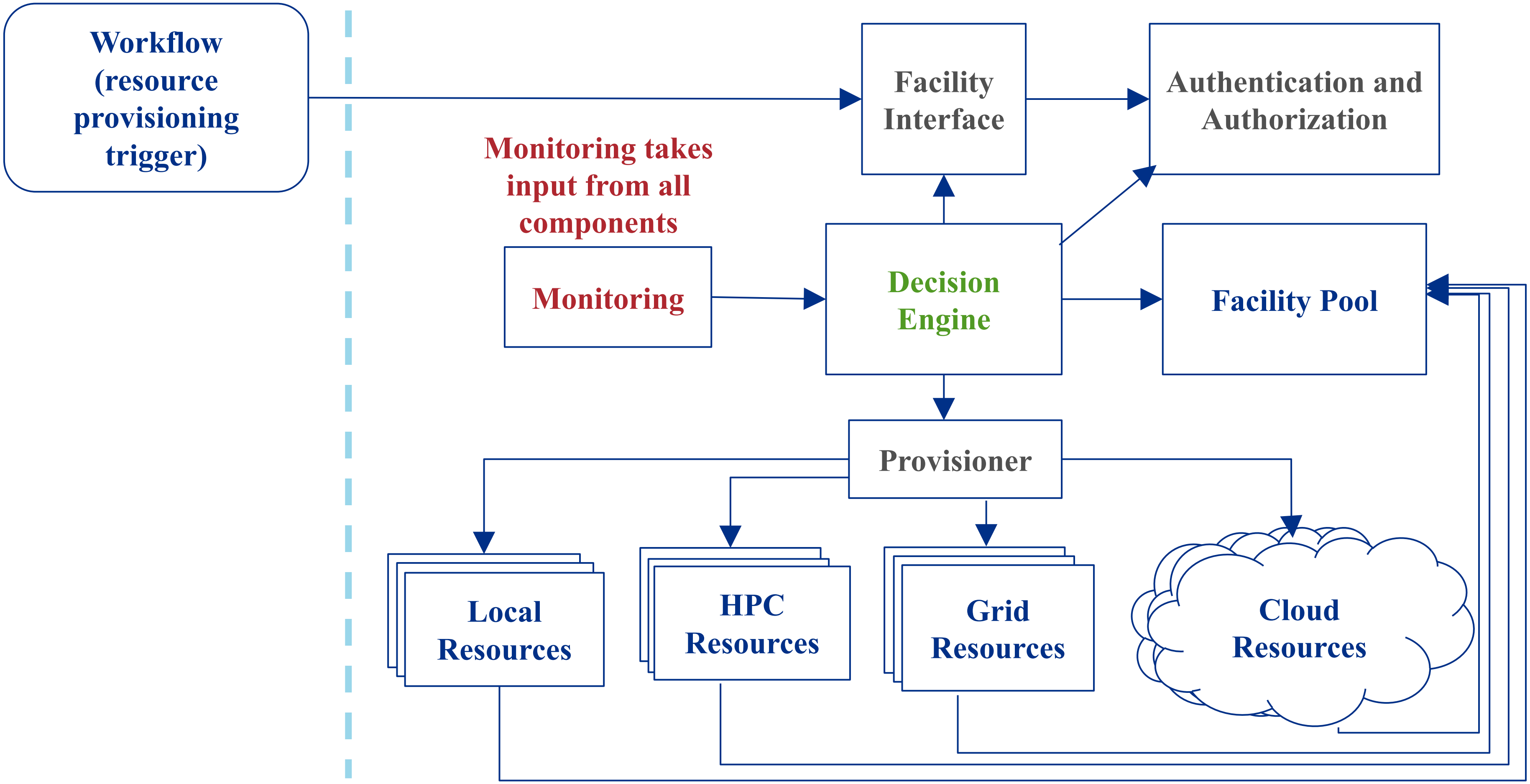
- Usage is not steady-state
- Computing schedules driven by real-world considerations (detector, accelerator, …) but also ingenuity – this is research and development of cutting-edge science



HEPCloud: the Evolved Facility

- Vision Statement
  - 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
  - The Facility routes to local or remote resources based on workflow requirements, cost, and efficiency of accessing various resources
  - Manages allocations of users to target compute engines
- Pilot project to explore feasibility, capability of HEPCloud
  - Goal of moving into production during FY18
  - Seed money provided by industry

HEPCloud Architecture



Early 2016 HEPCloud Use Cases - AWS

NoVA Processing

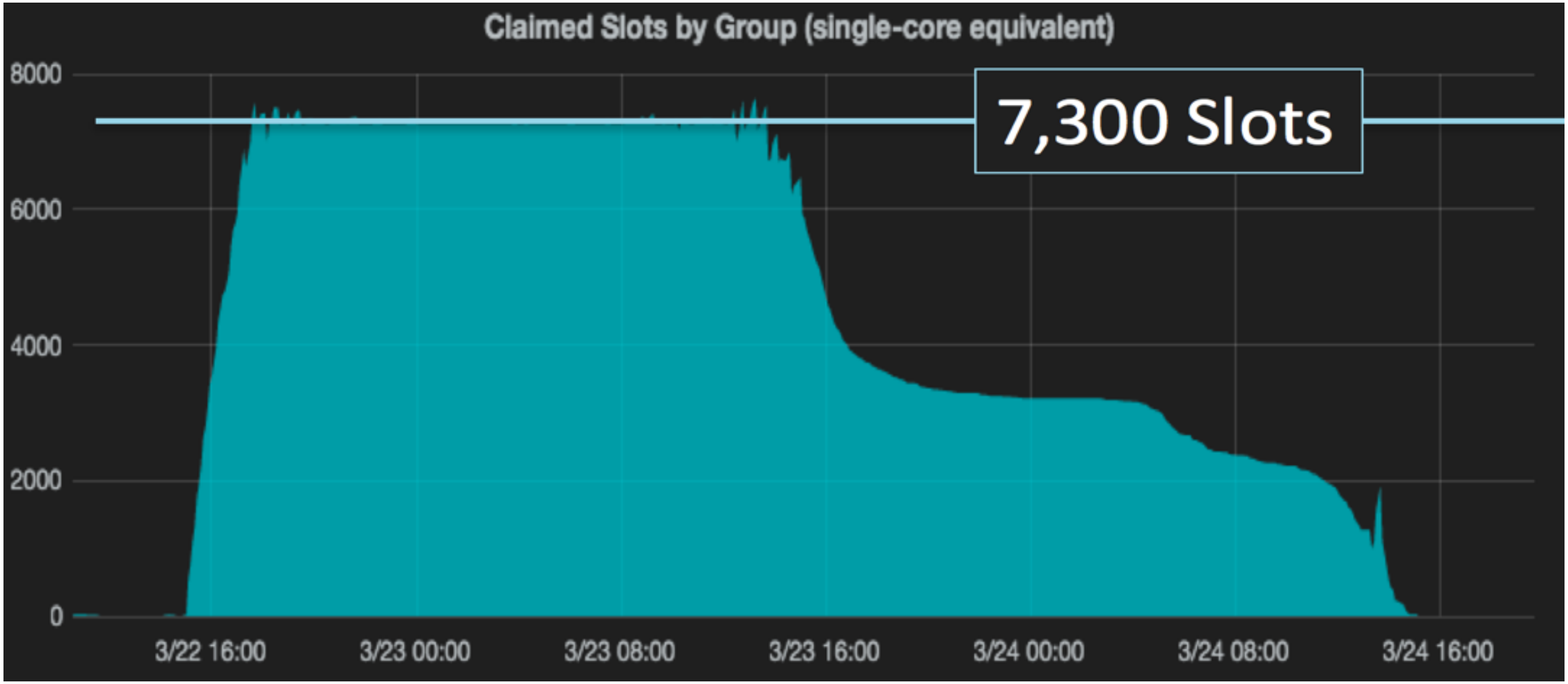
Processing the 2014/2015 dataset
16 4-day "campaigns" over one year
Demonstrates stability, availability, cost-effectiveness
Received AWS academic grant

CMS Monte Carlo Simulation

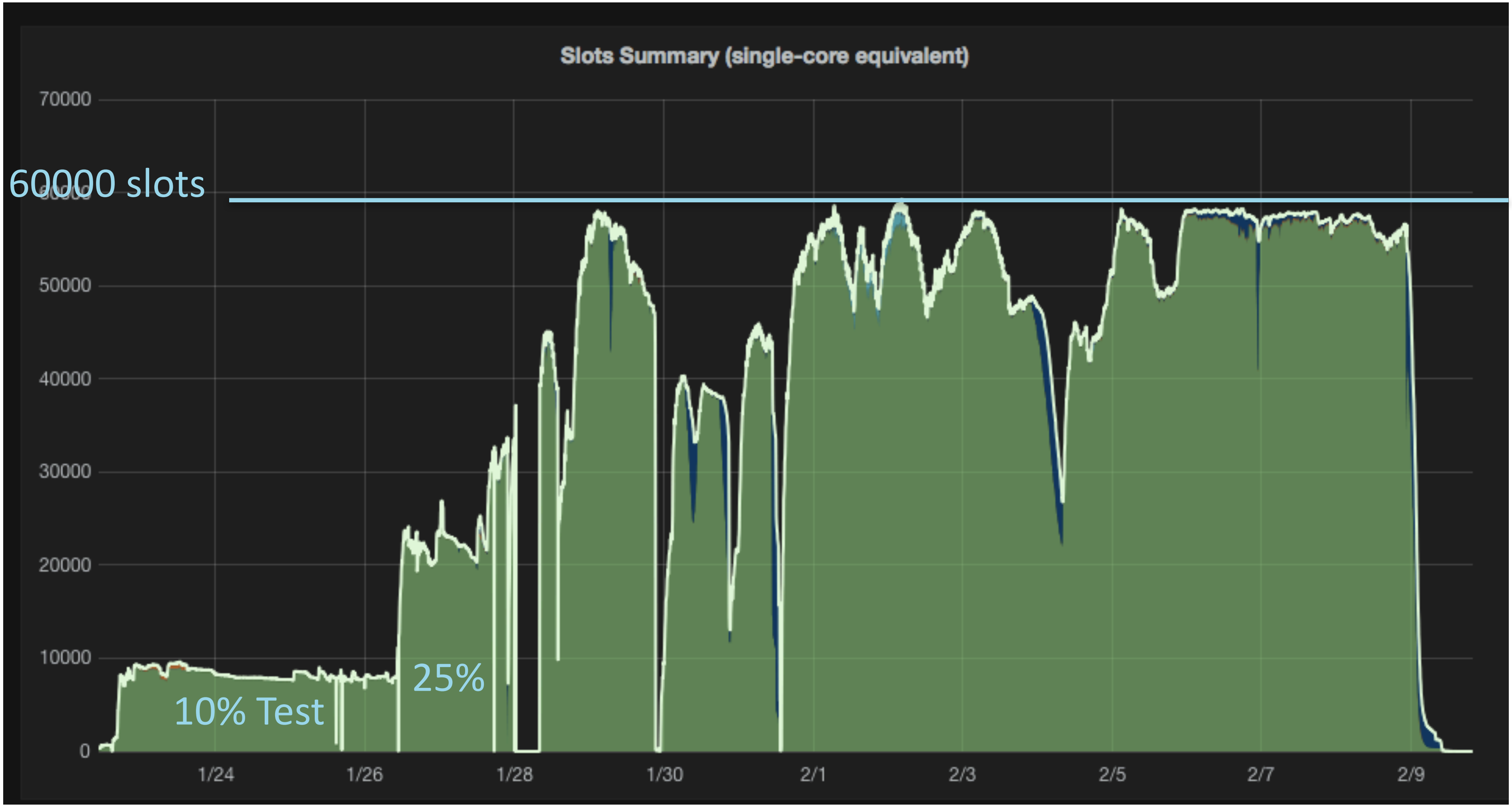
Generation (and detector simulation, digitization, reconstruction) of simulated events in time for Moriond16 conference
56000 compute cores, steady-state
Demonstrates scalability
Received AWS academic grant
(90% AWS funded, **10% US CMS**)

NOvA Use Case – running at 7300 cores

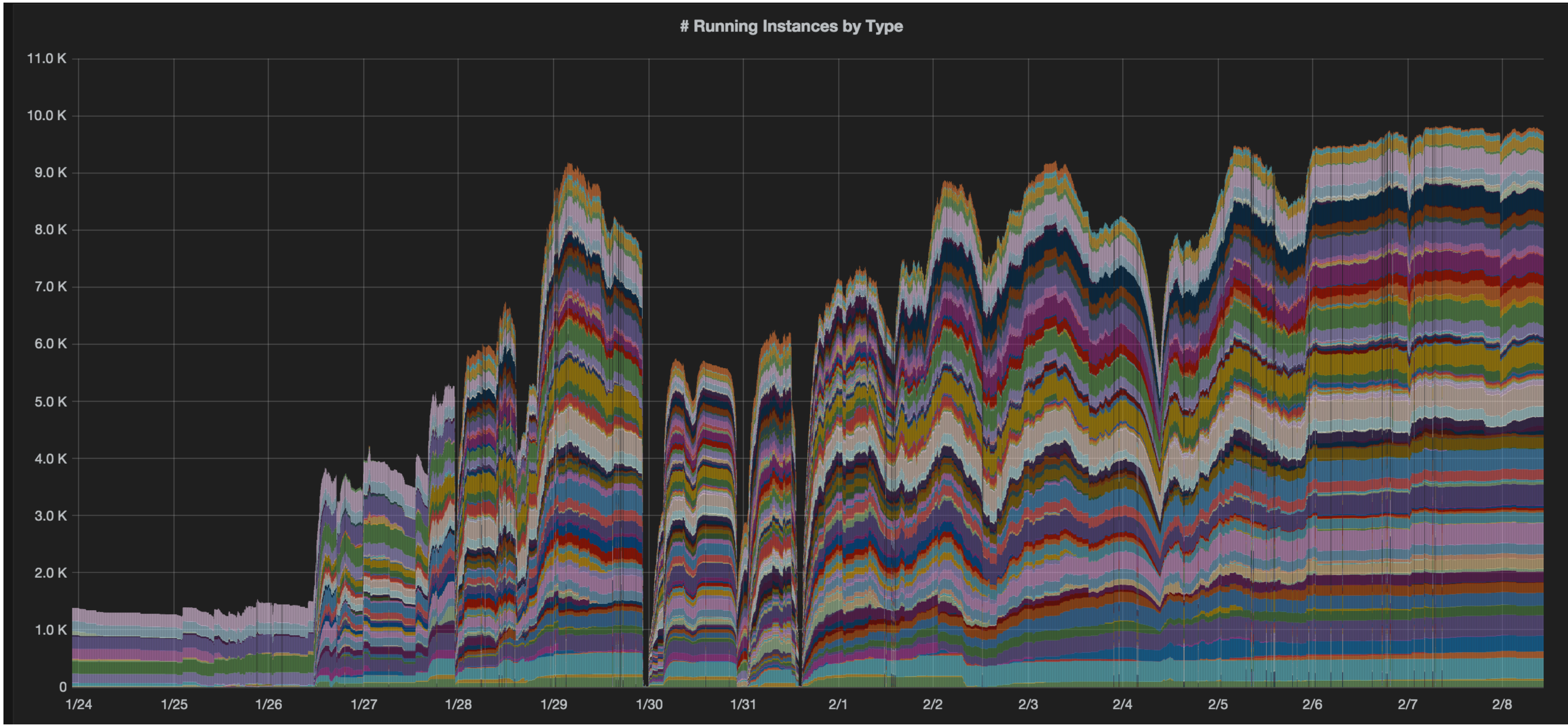
- Added support for general-purpose data-handling tools (SAM, IFDH, F-FTS) for AWS Storage and used them to stage both input datasets and job outputs



CMS Use Case – running at ~60,000 cores



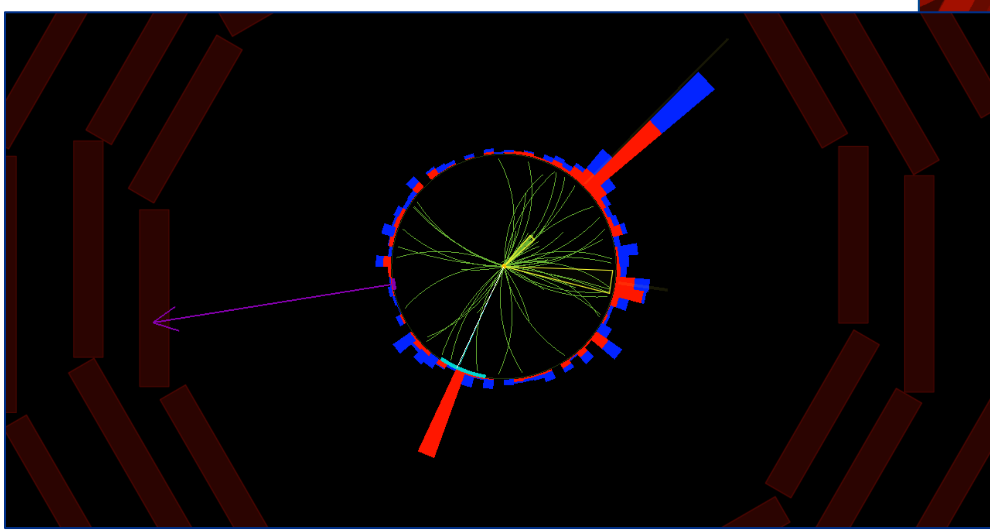
HEPCloud AWS slots by Region/Zone/Type

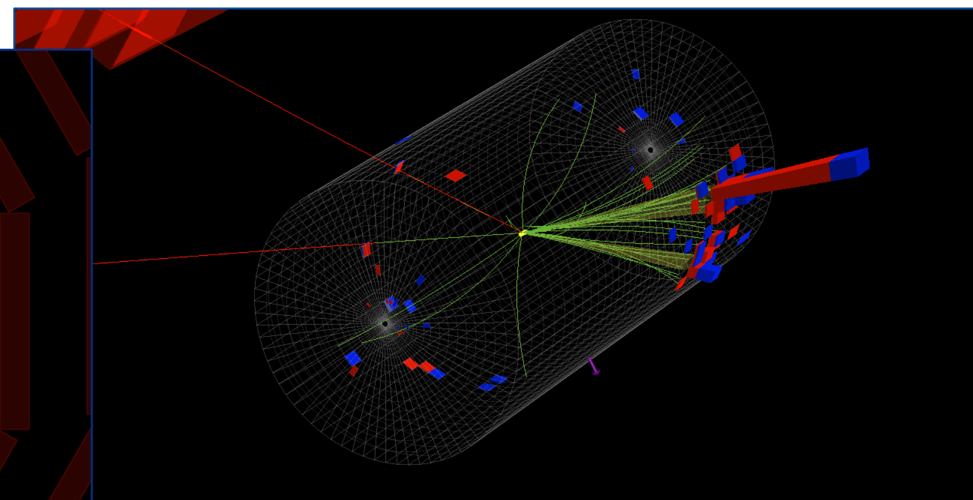


Each color corresponds to a different region+zone+machine type

Results from the Jan 2016 CMS Use Case

- **All CMS simulation requests fulfilled for conference**
  - **2.9 million jobs, 15.1 million wall hours**
    - 9.5% badput – including preemption
    - 87% CPU efficiency
  - **518 million events generated**





Late 2016 HEPCloud Use Cases - Google

NoVA Processing

Processing the 2014/2015 dataset
16 4-day "campaigns" over one year
Demonstrates stability, availability, cost-effectiveness
Received AWS academic grant

CMS Monte Carlo Simulation

Generation (and detector simulation, digitization, reconstruction) of simulated events in time for Moriond16 conference
56000 compute cores, steady-state
Demonstrates scalability
Received AWS academic grant

CMS Monte Carlo Simulation

Generation (and detector simulation, digitization, reconstruction) of simulated events in time for Moriond17 conference
160000 compute cores during Supercomputing 2016 conference (~48 h)
Demonstrates scalability, capability
Received Google Cloud Platform grant

mu2e Processing

Simulating cosmic ray veto detector and beam particle backgrounds
3M integrated core-hours
Demonstrates rapid on-boarding
Received Google Cloud Platform grant

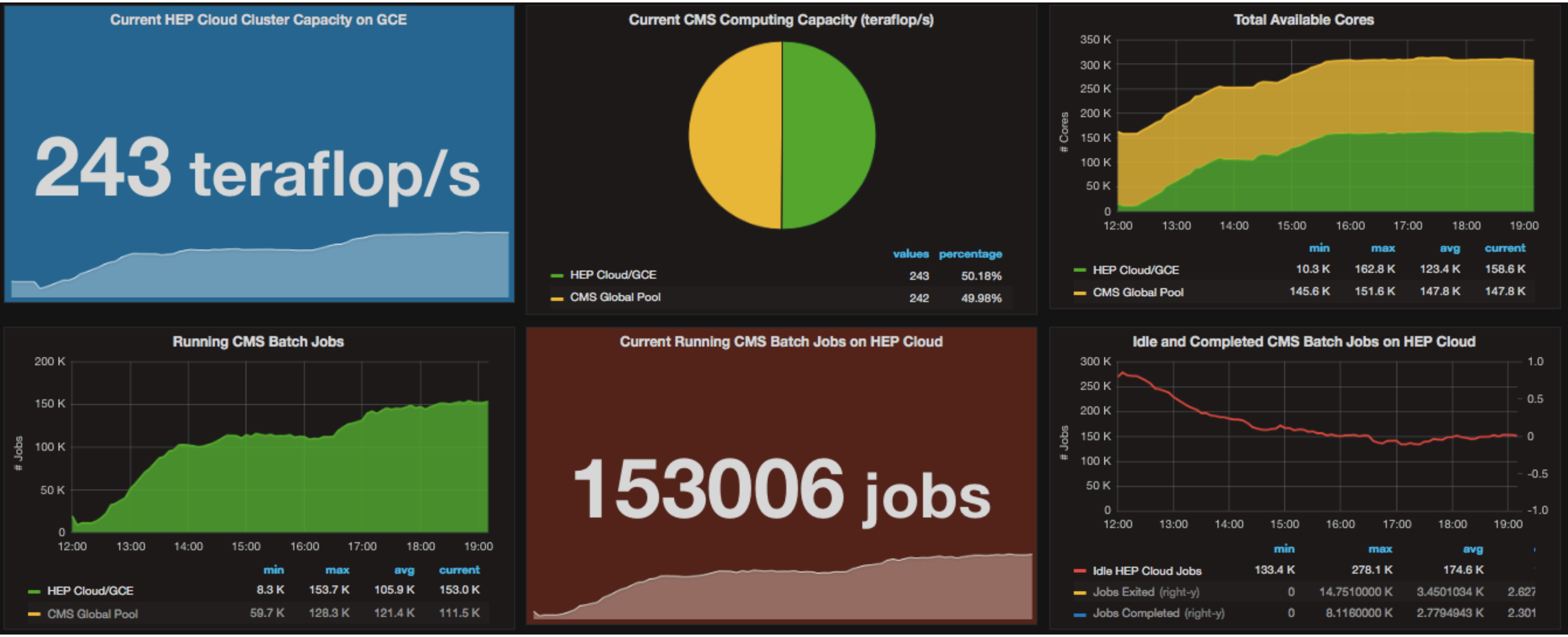
Results from the Jan 2016 CMS Use Case

- All CMS simulation requests fulfilled for conference
  - 2.9 million jobs, 15.1 million wall hours
    - 9.5% badput – including preemption
    - 87% CPU efficiency
  - 518 million events generated

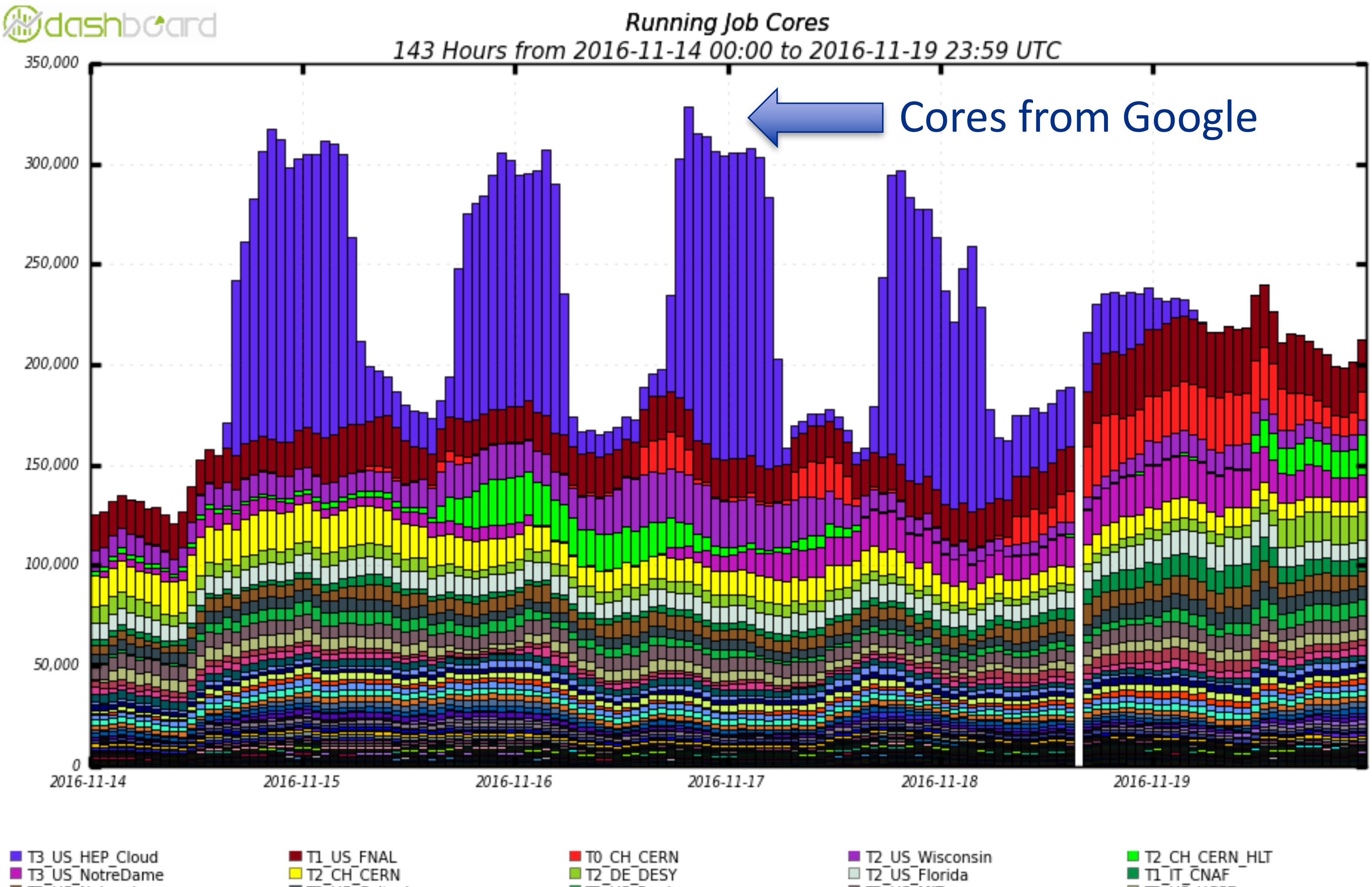
- **Supercomputing 2016**
  - **Aiming to generate* 1 Billion events <u>in 48 hours</u> during Supercomputing 2016**
  - **Double the size of global CMS computing resources**

* 35% filter efficiency – 380 million events staged out

Success! (Green = Google, Yellow = rest of world)



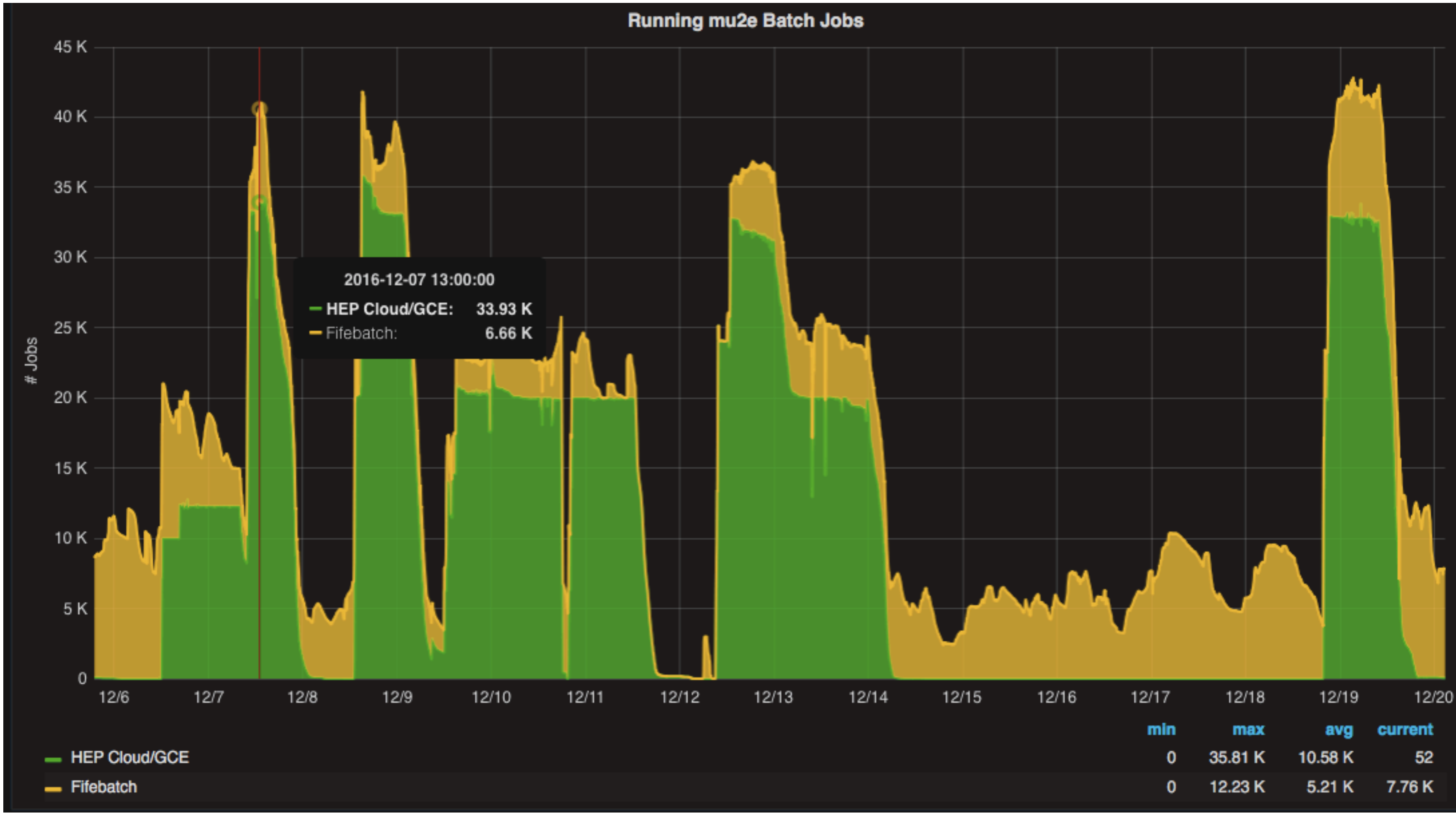
Doubling CMS compute capacity



CMS @ Google – overall numbers

- 6.35 M wallhours used; 5.42 M wallhours for completed jobs.
  - 730172 simulation jobs submitted; only 47 did not complete through the CMS and HEPCloud fault-tolerant infrastructures
  - Most wasted hours during ramp-up as we found and eliminated issues; goodput was at 94% during the last 3 days.
- Used ~$100k worth of credits on Google Cloud during Supercomputing 2016
  - $71k virtual machine costs
  - $8.6k network egress
  - $8.5k disk attached to VMs
  - $3.5k cloud storage for input data
- 205 M physics events generated, yielding 81.8 TB of data

Mu2e – executing on Google Cloud

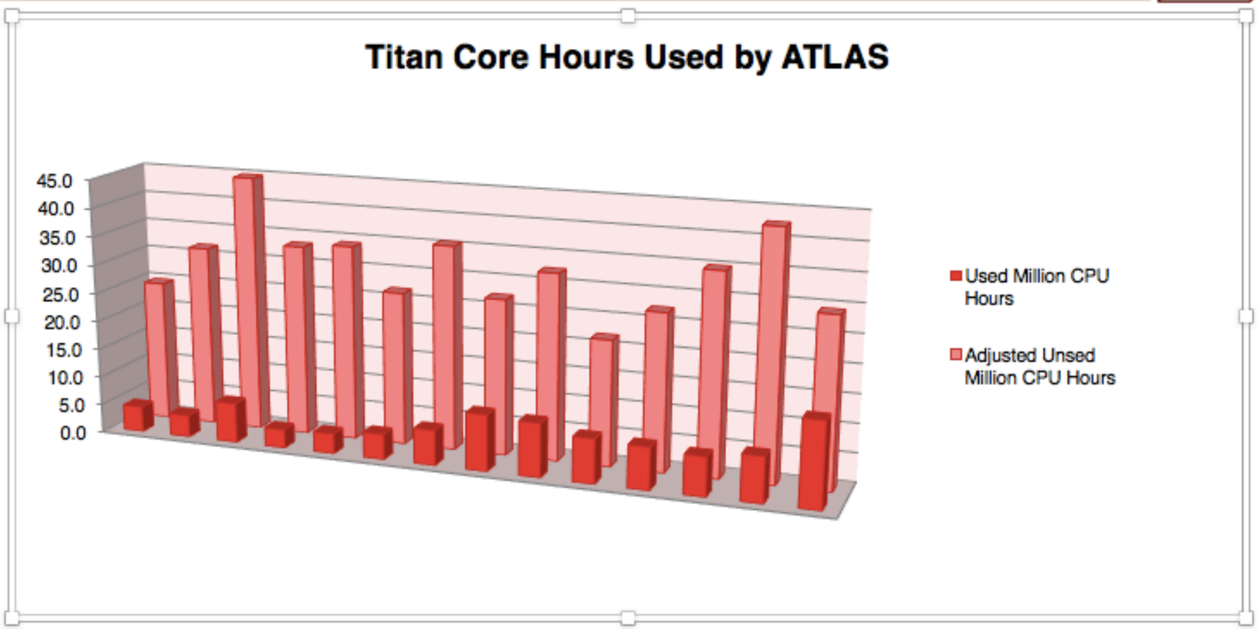


On-premises vs. cloud cost comparison - AWS

- Average cost per core-hour
  - On-premises resource: **.9** cents per core-hour
    - Includes power, cooling, staff
  - Off-premises at AWS: **1.4** cents per core-hour
    - Ranged up to 3 cents per core-hour at smaller scale
- Benchmarks
  - Specialized (“ttbar”) benchmark focused on HEP workflows
    - On-premises: **0.0163** (higher = better)
    - Off-premises: **0.0158**
- Raw compute performance roughly equivalent
- Cloud costs larger – but approaching equivalence
  - Google ~ **1.6** cents per core-hour

HPC ("Supercomputers"): Do they make sense for HEP jobs?

Titan Usage for MC Simulations



74 Million Titan Core Hours used in calendar year 2016
In backfill mode – no allocation
374 Million Hours remained unused.

Kaushik De *Mar 7, 2017* 3

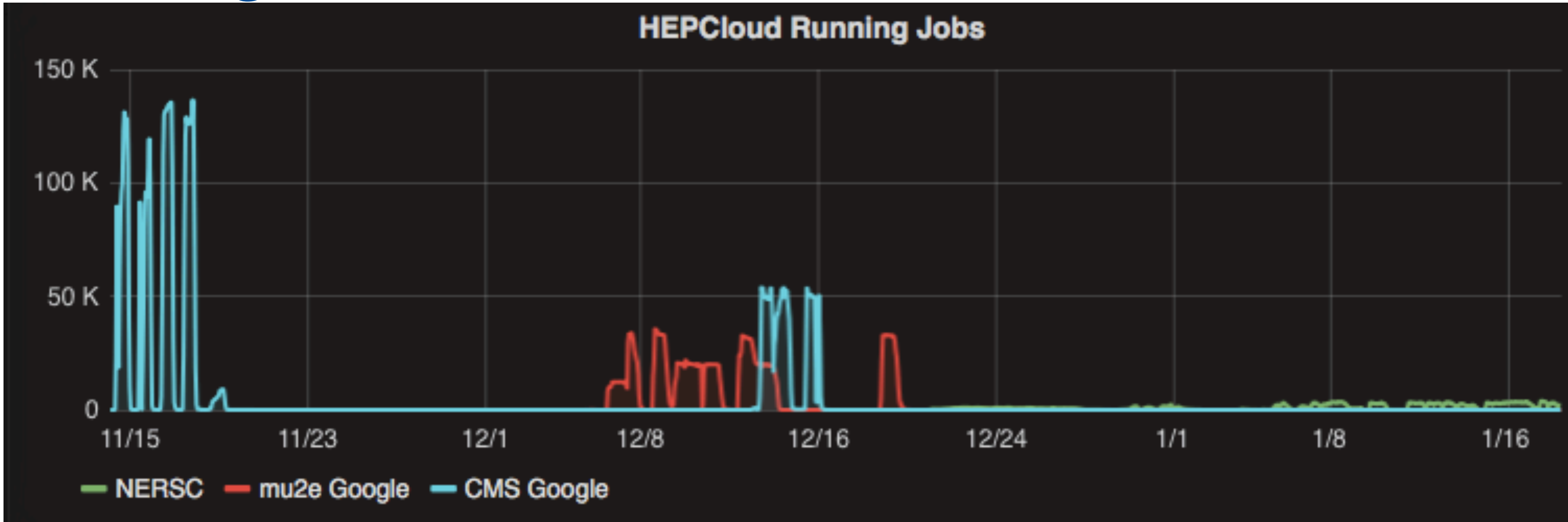
From Kaushik De @ UTA

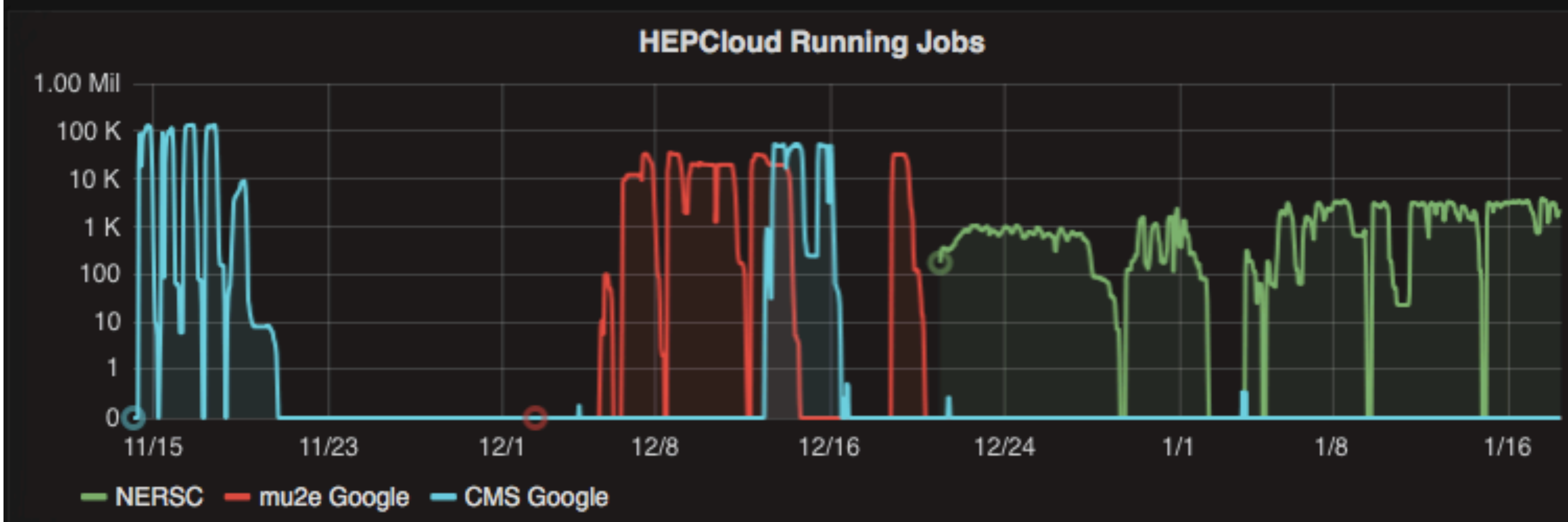
Fermilab

HEPCloud Compute and HPC

- **Plans for 2017:** HEPCloud provisioning @ NERSC
  - HEPCloud allocation granted for 28 million MPP-hours
    - 16 million MPP-hours for intensity frontier (mu2e, MicroBooNE, NOvA, …)
    - 12 million MPP-hours for CMS
      - CMS production is running on Knight's Landing; experiment is working to optimize and maximize efficiency
  - Leverage experience towards leadership computing facilities
    - Harden and test HTCondor-CE interface @ ALCF

Running at NERSC





Thanks

- The Fermilab team:
  - Joe Boyd, Stu Fuess, Gabriele Garzoglio, Hyun Woo Kim, Rob Kennedy, Krista Majewski, David Mason, Parag Mhashilkar, Neha Sharma, Steve Timm, Anthony Tiradani, Panagiotis Spentzouris
- The HTCondor and glideinWMS projects
- Open Science Grid
- Energy Sciences Network
- The Google team:
  - Karan Bhatia, Solomon Boulos, Sam Greenfield, Paul Rossman, Doug Strain
- The AWS team:
  - Sanjay Padhi, Jamie Baker, Jamie Kinney, Mike Kokorowski
- Resellers: Onix, DLT

http://hepcloud.fnal.gov

Backup

Classes of Resource Providers

Grid

- Virtual Organizations (VOs) of users trusted by Grid sites
- VOs get allocations ➜ **Pledges**
  - Unused allocations: opportunistic resources

"Things you borrow"

Trust Federation

Cloud

- Community Clouds - Similar trust federation to Grids
- Commercial Clouds - **Pay-As-You-Go** model
  - Strongly accounted
  - Near-infinite capacity ➜ **Elasticity**
  - Spot price market

"Things you rent"

Economic Model

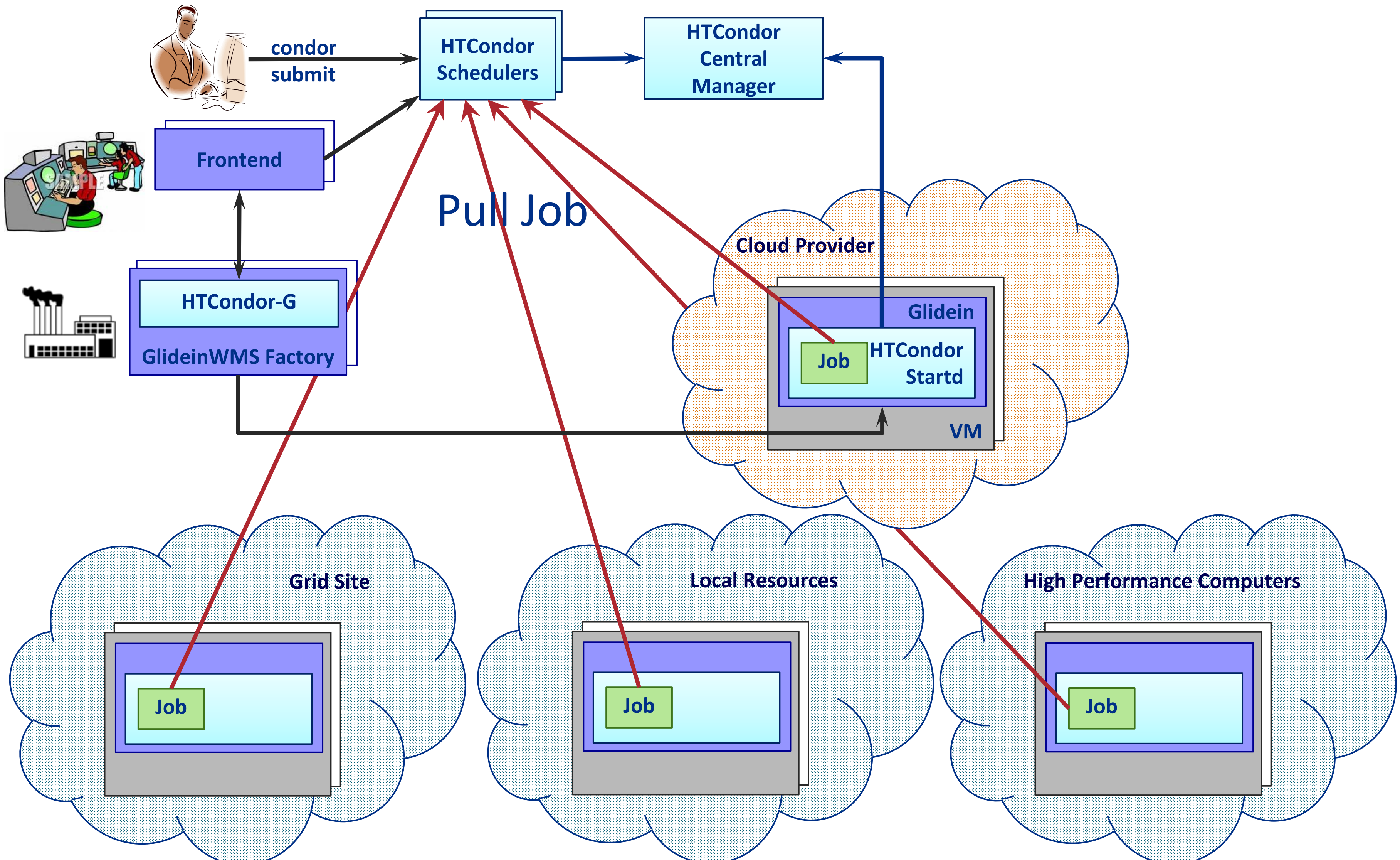
HPC

- Researchers granted access to HPC installations
- Peer review committees award **Allocations**
  - Awards model designed for individual PIs rather than large collaborations

"Things you are given"

Grant Allocation

HEPCloud – glideinWMS and HTCondor

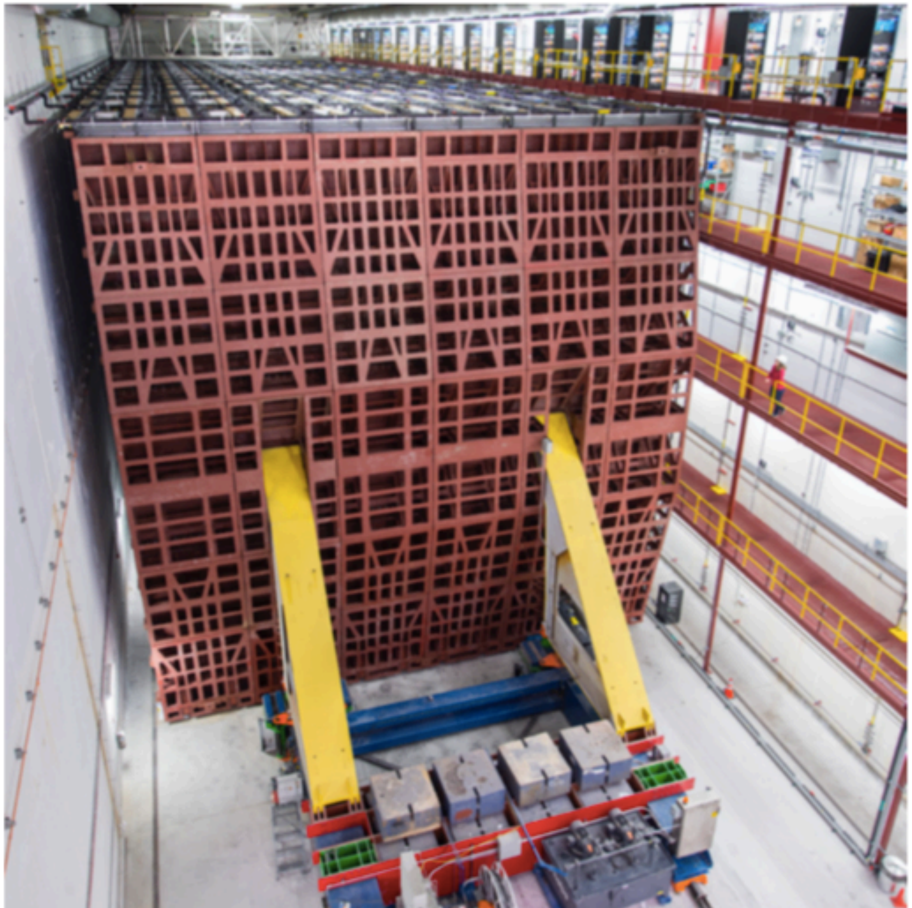


Lesson:
Non-technical challenges are still challenges

Legal, procurement, and other "fun" stuff

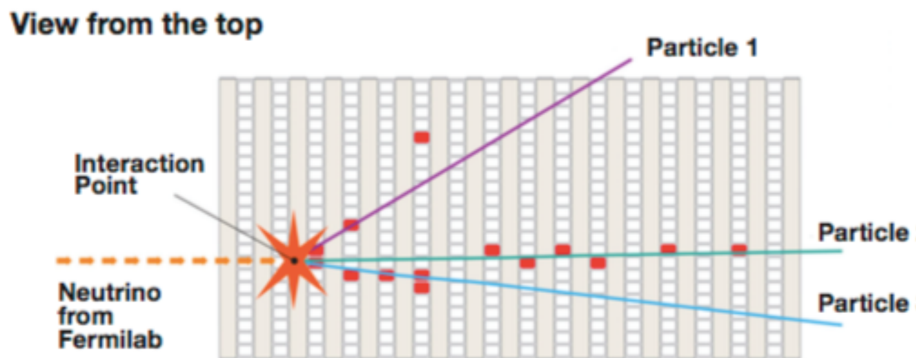
- Cloud accounts
  - Every cloud provider has a "**click-through**" **license** when you create an account (free tier, or otherwise)
  - **Terms and Conditions** of these accounts are generally incompatible with US Government statutes
- Procurement
  - **Cloud resources** look different than what we're used to buying
  - **Lack of fixed cost** can be confusing and set off red flags
- Security
  - Two of the big three providers are **FEDRAMP** certified
- Cost
  - Management fixates on dangers of **runaway costs**

NOvA: Neutrino Experiment



The NOvA detector in Minnesota occupies an area about the size of two basketball courts. It is 200 feet long and made of modules 50 feet high and 50 feet wide. The detector records particle tracks from neutrinos sent by a powerful accelerator at Fermilab. The construction of the NOvA detectors was completed in the fall of 2014, on time and under budget. The experiment is scheduled to collect information for six years.

Neutrino interaction recorded by NOvA



Neutrinos rarely interact with matter. When a neutrino smashes into an atom in the NOvA detector in Minnesota, it creates distinctive particle tracks. Scientists explore these particle interactions to better understand the transition of muon neutrinos into electron neutrinos. The experiment also helps answer important scientific questions about neutrino masses, neutrino oscillations, and the role neutrinos played in the early universe.

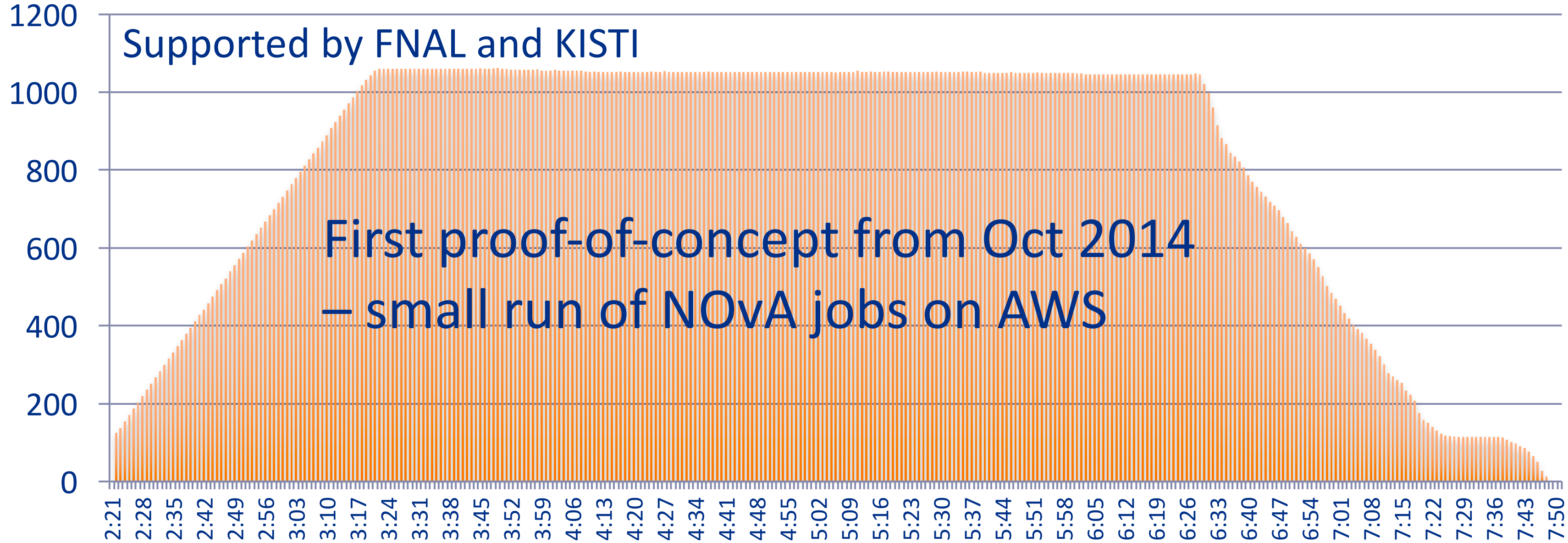
NOvA Use Case

NoVA Processing

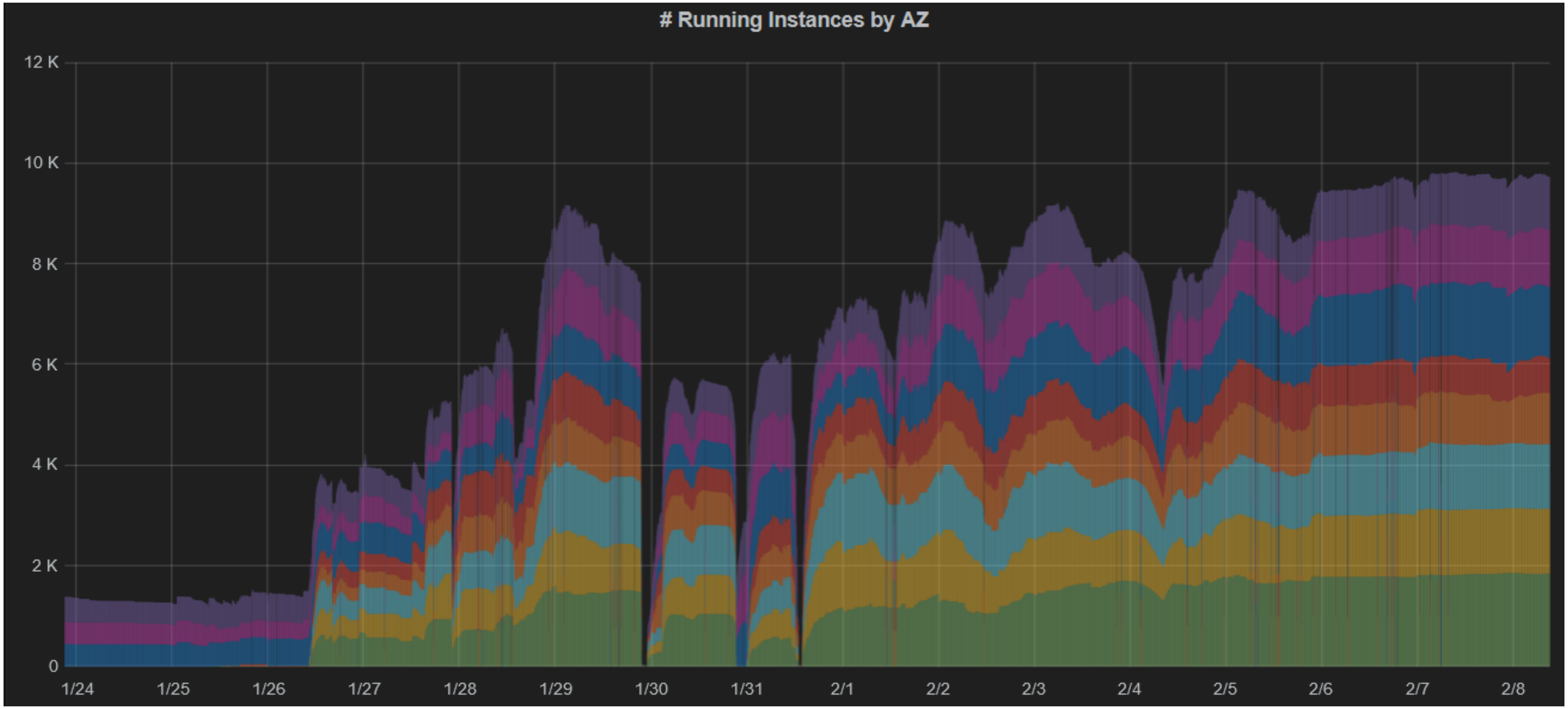
Processing the 2014/2015 dataset
16 4-day "campaigns" over one year
Demonstrates stability, availability, cost-effectiveness
Received AWS academic grant

CMS Monte Carlo Simulation

Generation (and detector simulation, digitization, reconstruction) of simulated events in time for Moriond16 conference
56000 compute cores, steady-state
Demonstrates scalability
Received AWS academic grant

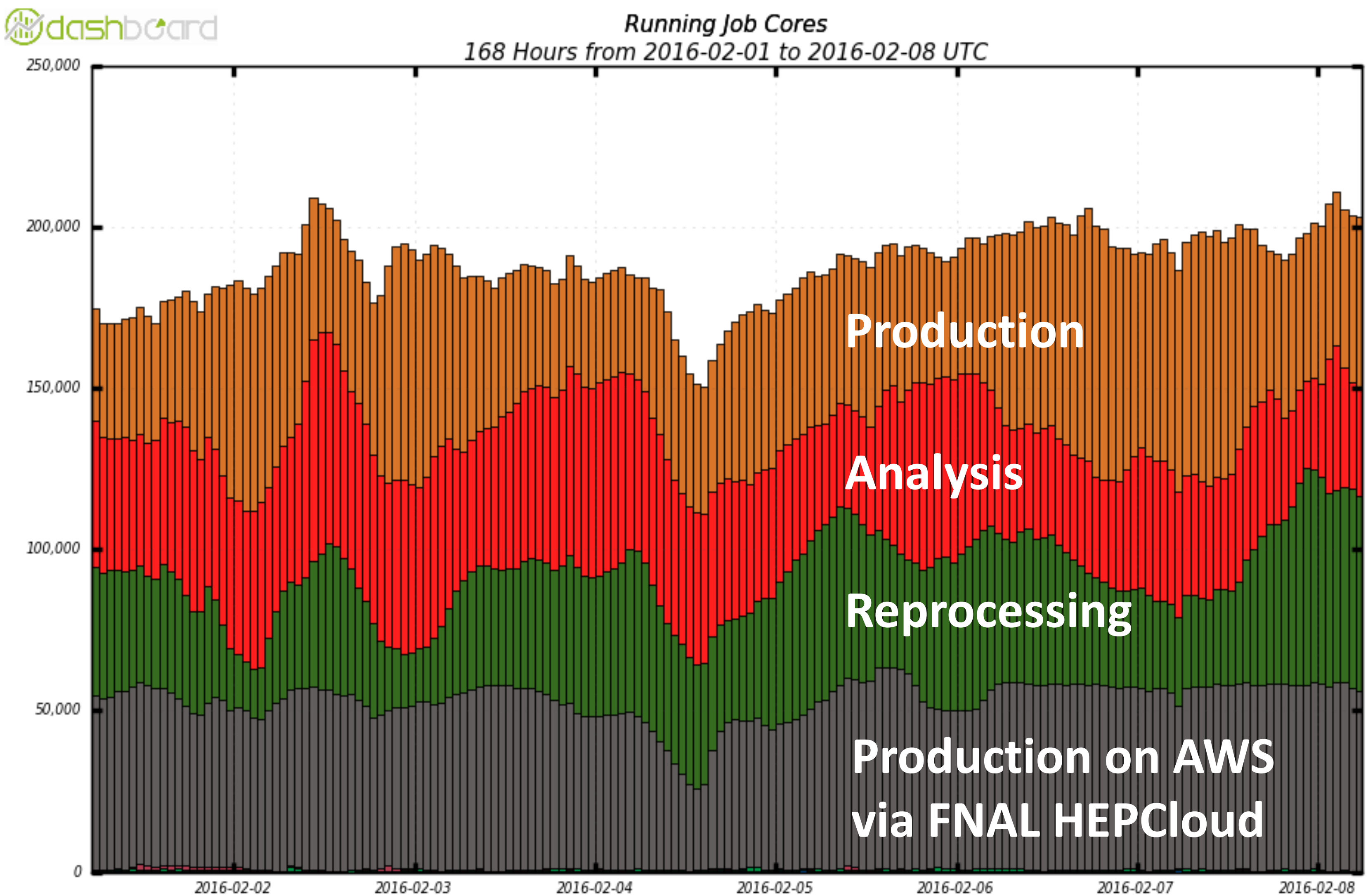


HEPCloud AWS slots by Region/Zone

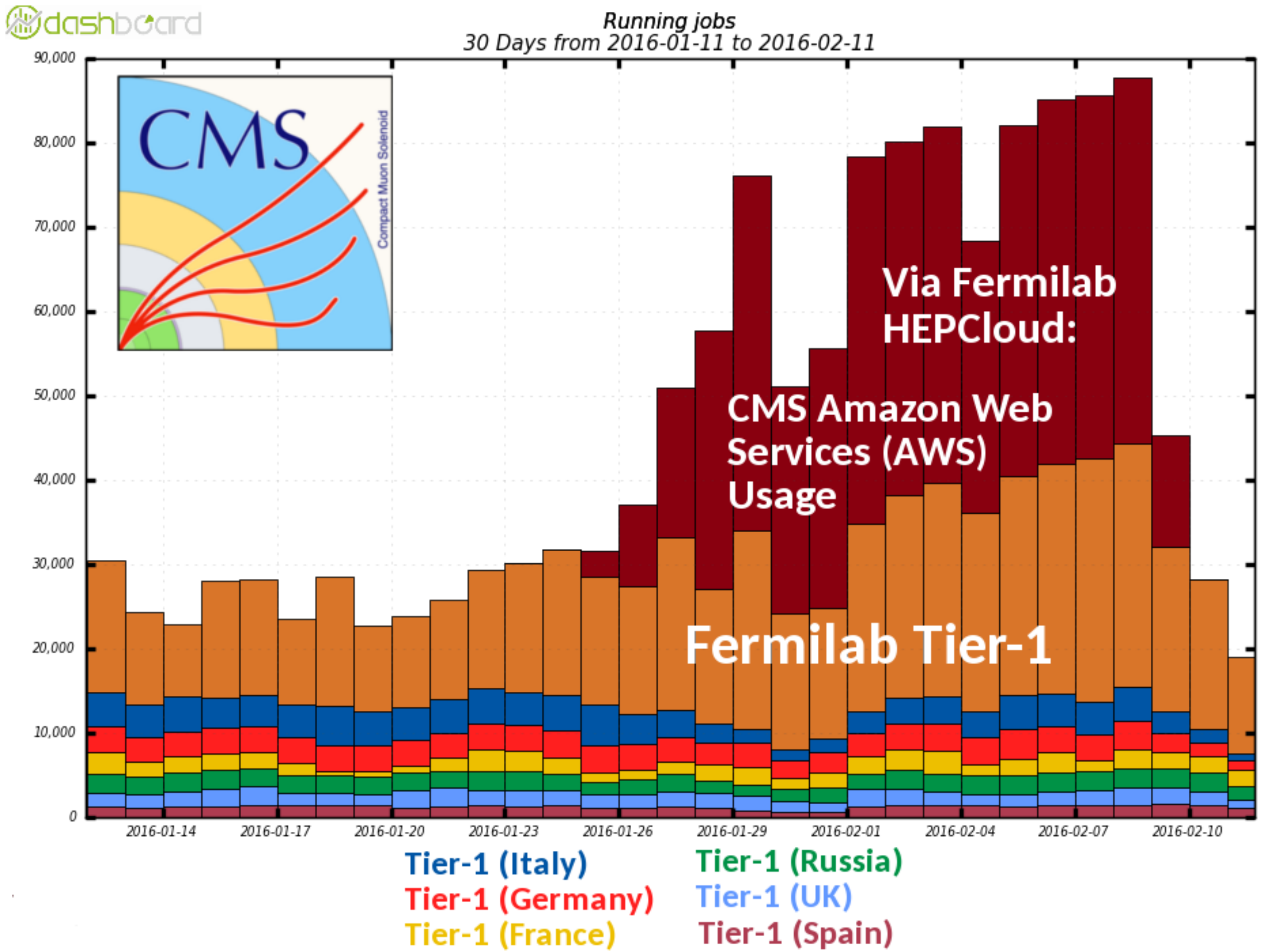


Each color corresponds to a different region+zone

HEPCloud AWS: 25% of CMS global capacity



Fermilab HEPCloud compared to global CMS Tier-1



Lesson:
You have amortized costs and don’t know it
(and you may find this out the hard way)

Fermilab

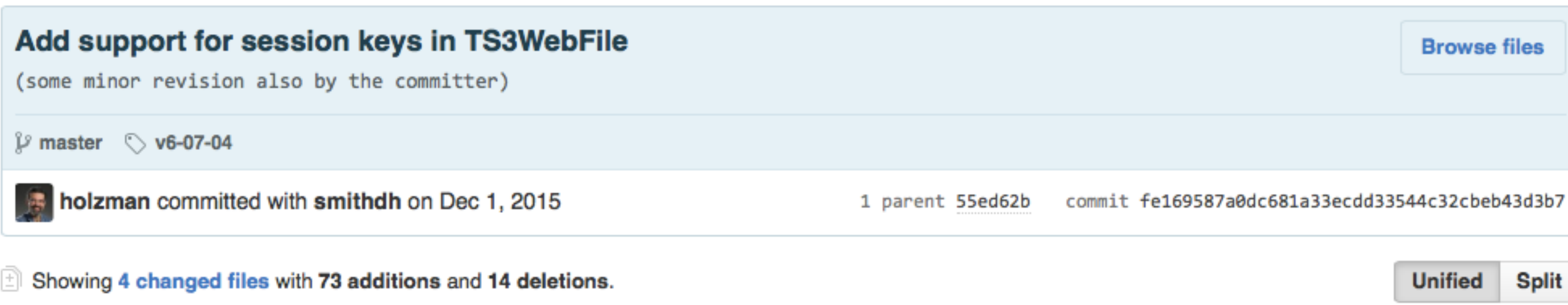
Description of CMS workflow

- Four chained steps (output of step N is input of step N+1)
  - Step 1 requires few GB input ("Gridpack") – same files per job
  - Step 3 requires additional input: **"pile-up" data** (simulating multiple events per bunch crossing), 5-10 GB

- Pile-up data is constructed on-the-fly by random seek and sequential reads into a 10 TB dataset
    - More than **150 pile-up events read** per simulated event – this step is extremely I/O intensive
    - Staged pile-up datasets to AWS S3 (storage service) ahead-of-time using WLCG FTS3

Reading pile-up from AWS S3 (storage)

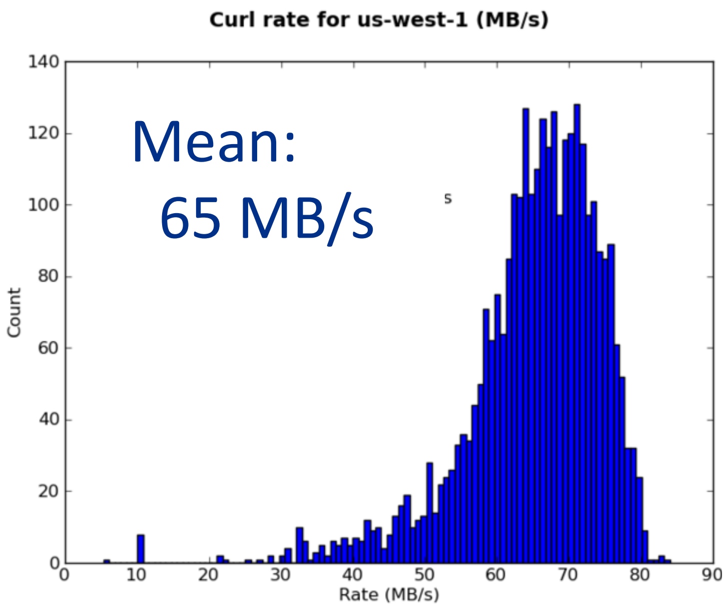
- AWS worker nodes granted permission to read from AWS S3 folder ("bucket") via AWS Security-Token-Service (STS)
- ROOT has a TS3WebFile class!
  – But session key support was missing (needed for STS!)

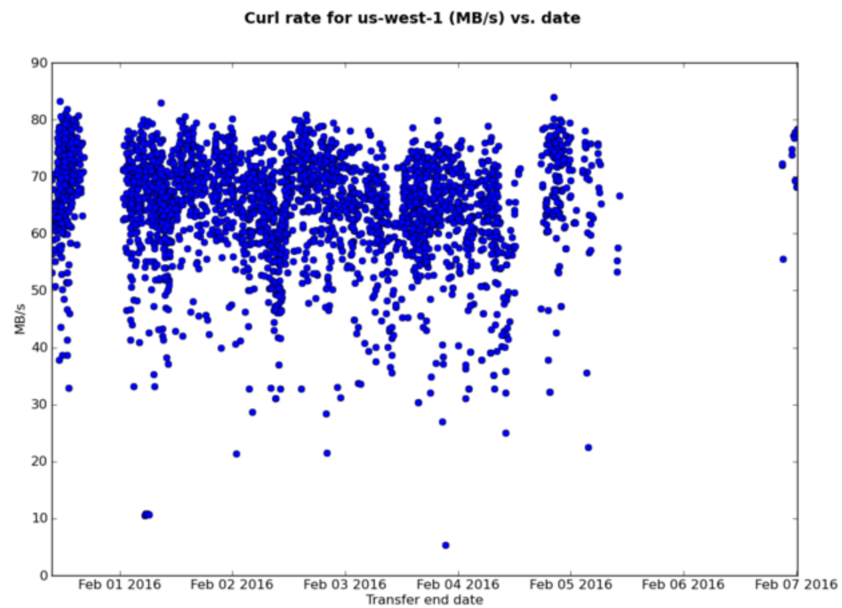


- This worked great!
  – Except…

Reading pile-up from AWS S3 (storage)

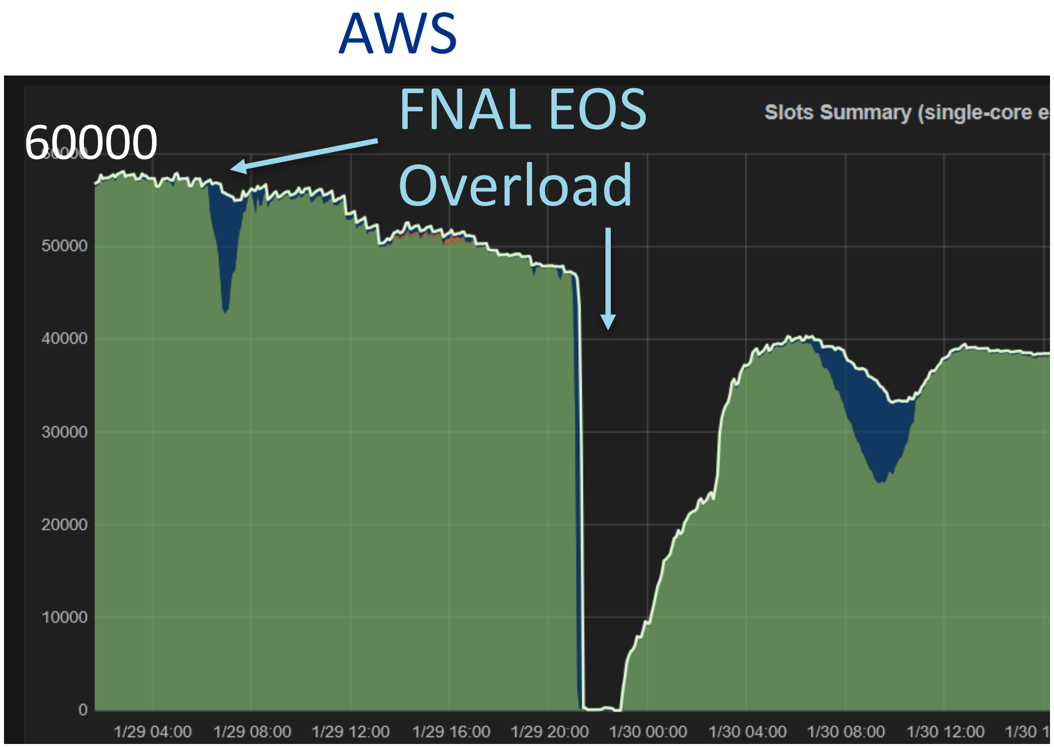
- **Cost of data access was 30% of compute costs**
  - AWS microcharges per HTTP GET, and **150 million** GETs per hour is a lot!
- Wrote a curl wrapper to provide the custom AWS authentication headers
  - (Not often I can say I reduced costs by 5 orders of magnitude!)





Lesson:
Just because the cloud provider scales doesn't mean that you will

Overloading FNAL storage with stage-out



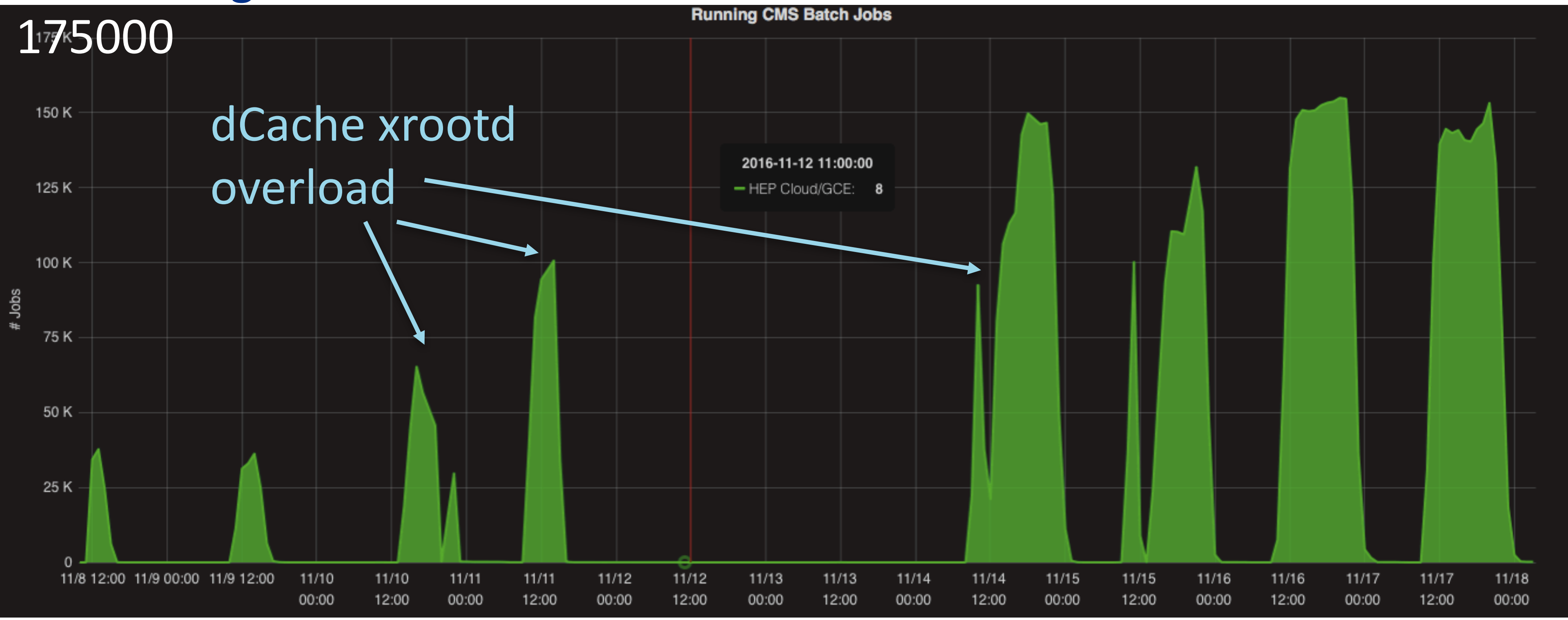
All jobs using SRM to stageout to Fermilab EOS

BeSTMan component could not keep up!

Switched to **xrootd** protocol and all problems are solved, right?

Overloading FNAL storage with stage-out

Google

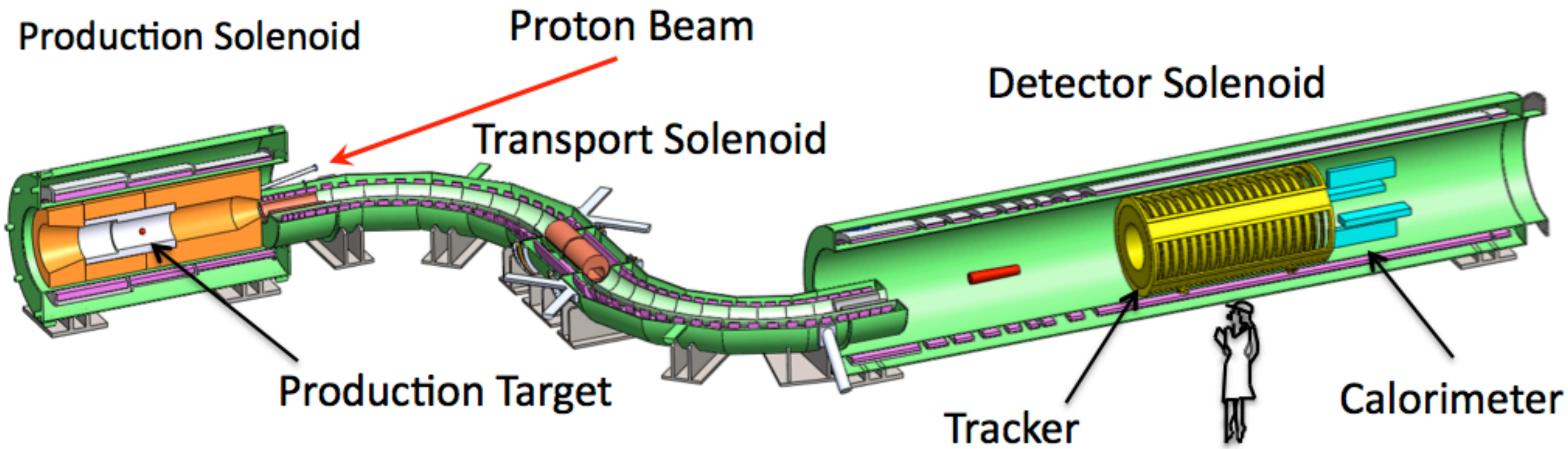


All jobs using xrootd to stageout to Fermilab dCache

Single xrootd door could not keep up!

Added “load balancing” (create more doors and randomly choose!)

Mu2e experiment



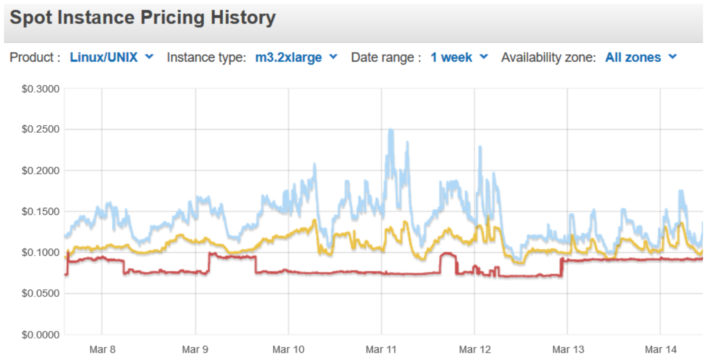
- Charged Lepton Flavor Violation is a near-universal feature of extensions to the Standard Model of particle physics
- Rare muon processes offer the best combination of new physics reach and experimental sensitivity
- **Search for muon (in bound state) converting to an electron ("mu" to "e")**

Commercial Cloud Pricing

- Significant costs on Commercial Cloud
  - Compute charges over time (per hour)*
  - Persistent storage for large input data sets
  - Network egress (per GB)
  - Ancillary support services (persistent scalable web caches)
  - Per-operation API charges

VM Pricing: using the AWS "Spot Market"

- AWS has a fixed price per hour (rates vary by machine type)
- Excess capacity is released to the free ("spot") market at a fraction of the on-demand price
  - End user chooses a bid price
  - If (market price < bid), you pay only market price for the provisioned resource
    - If (market price > bid), you don't get the resource
  - If the price fluctuates while you are running and the market price exceeds your original bid price, you may get kicked off the node (with a 2 minute warning!)



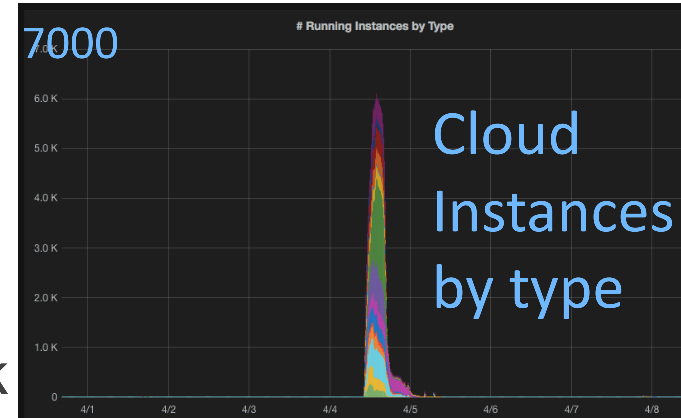
VM Pricing: using Google preemptible VMs

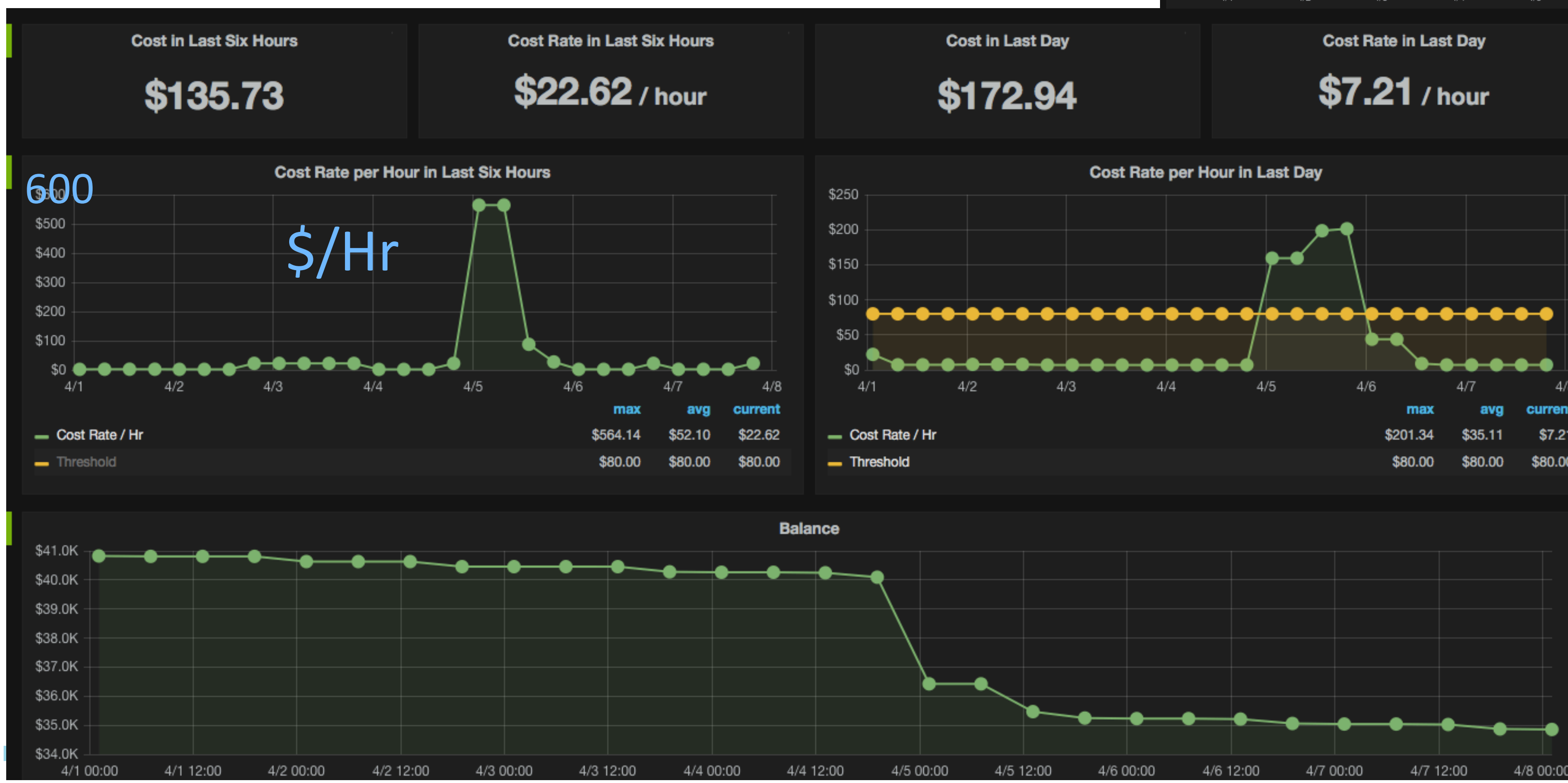
- Google VMs have a fixed cost (varies by machine types)
- Preemptible Google VMs are available at a significantly smaller fixed cost – 1 cent per core hour for a “standard candle”
  – We saved a few percent on cost by using custom VMs (2 GB per core instead of the standard 3.75 GB per core)

Lesson:
Cost monitoring is crucial

HEPCloud: Orchestration

- Monitoring and Accounting
  - Synergies with FIFE monitoring projects
    - But also monitoring **real-time** expense
  - Cloud providers **lack real-time** feedback



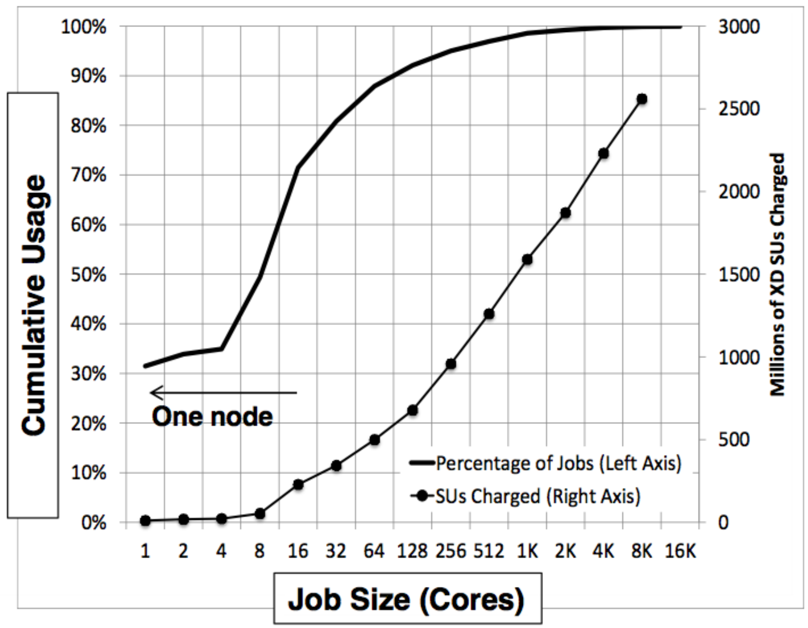


Fermilab

HPC: does it makes sense for our jobs?

HPC for the 99%

- 99% of jobs run on NSF's HPC resources in 2012 used <2,048 cores
- And consumed >50% of the total core-hours across NSF resources



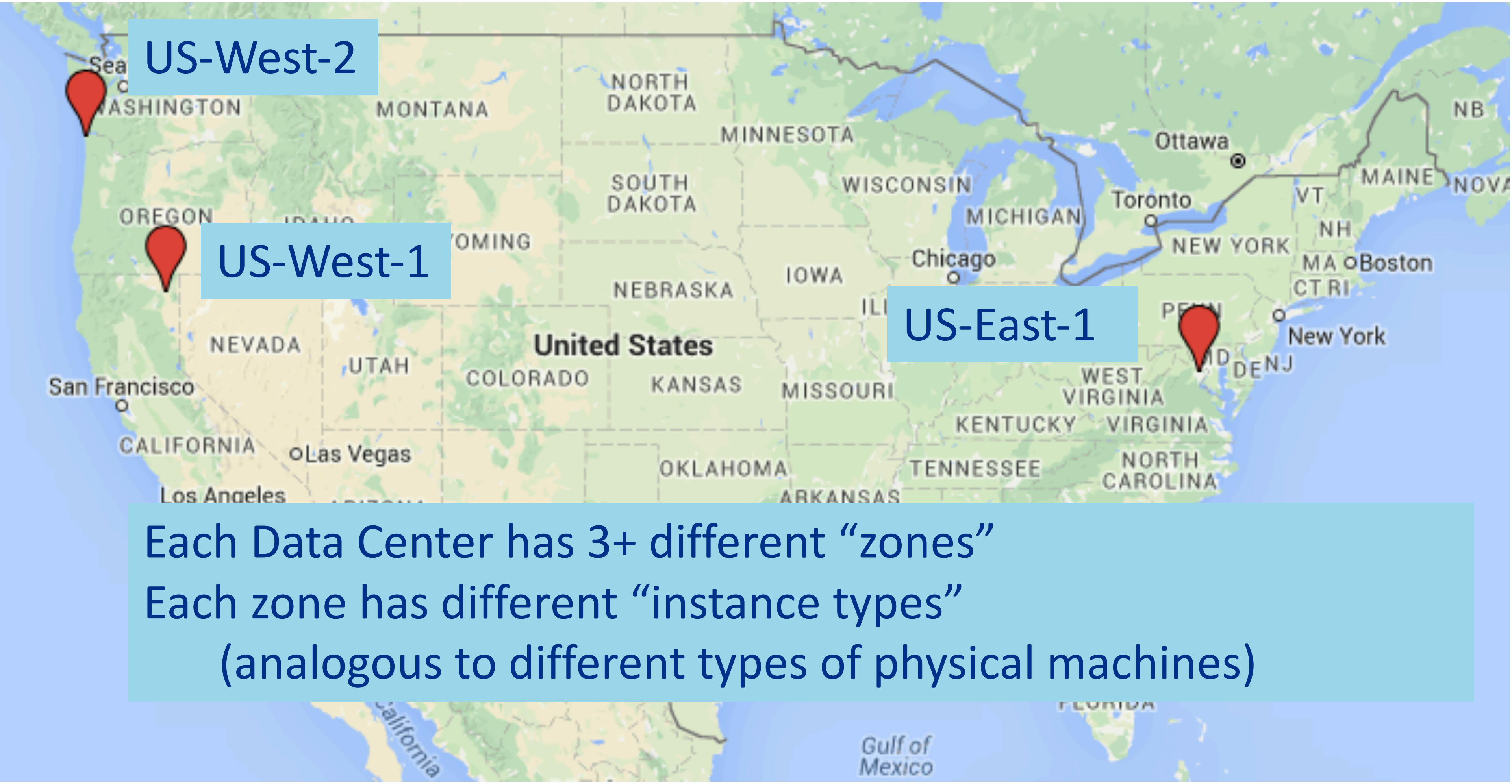
SDSC SAN DIEGO SUPERCOMPUTER CENTER

at the UNIVERSITY OF CALIFORNIA; SAN DIEGO

UCSD

From Michael Norman @ SDSC

AWS topology – three US data centers (“regions”)

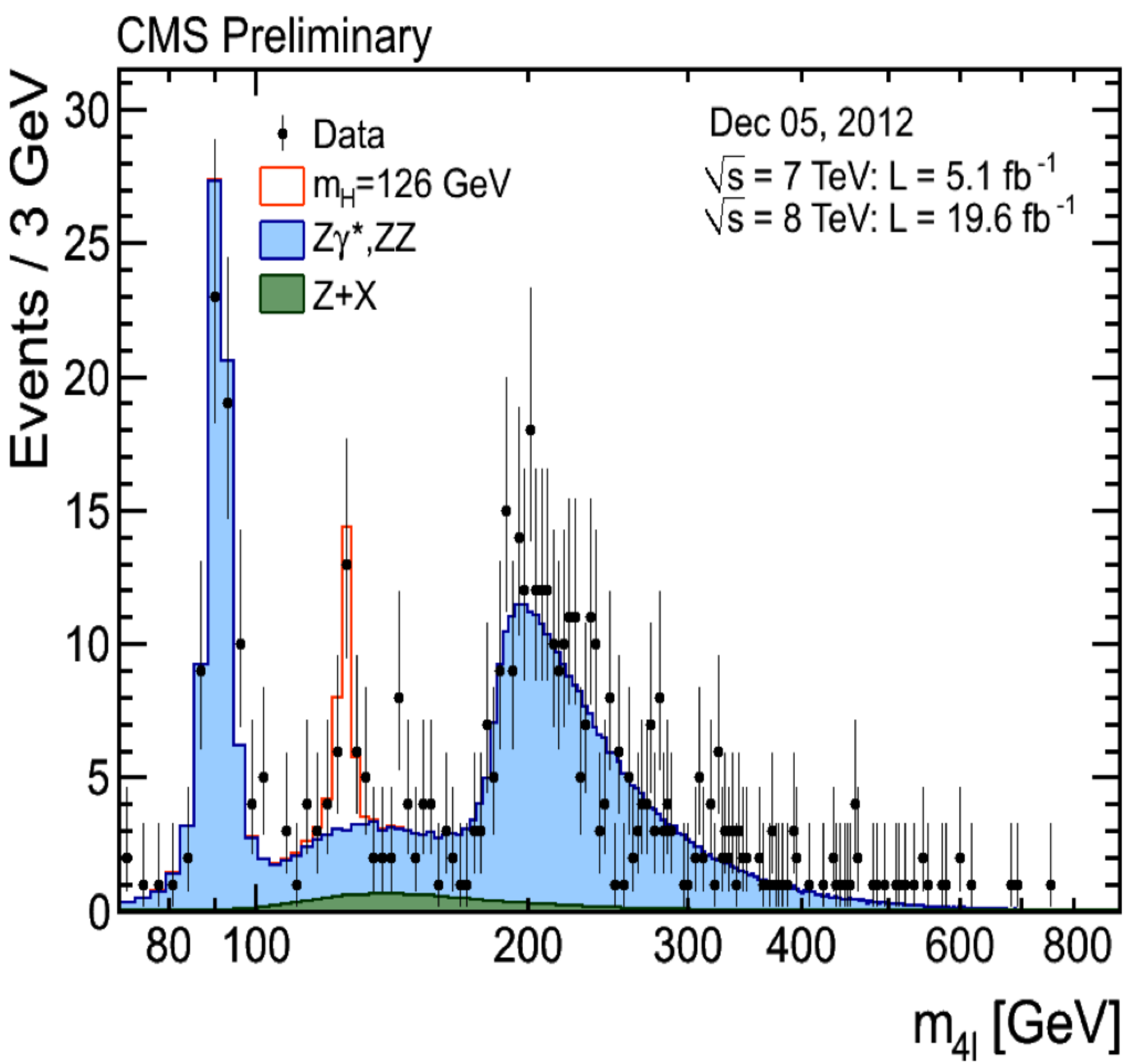


Each Data Center has 3+ different “zones”
Each zone has different “instance types”
(analogous to different types of physical machines)

Running on Google Cloud - Google Services

- Distributing experiment code (many versions and codes)
  - CVMFS: caching layer using squid web caches
  - Scalable, easy-to-manage software distribution
  - Good fit for **Google Load Balancing**
- Reading input data
  - Staged 500 TB of input data to **Google Cloud Storage**
    - Standard HEP and CMS data management tools now speak http!
    - **Thanks to ESNet and Google** for upgraded (100 Gbit+) peering!
  - Mounted data using **gcsfuse**
    - Good for big serial reads
- Monitoring
  - **Stackdriver logging**
    - Splunk-like functionality – a big help for troubleshooting

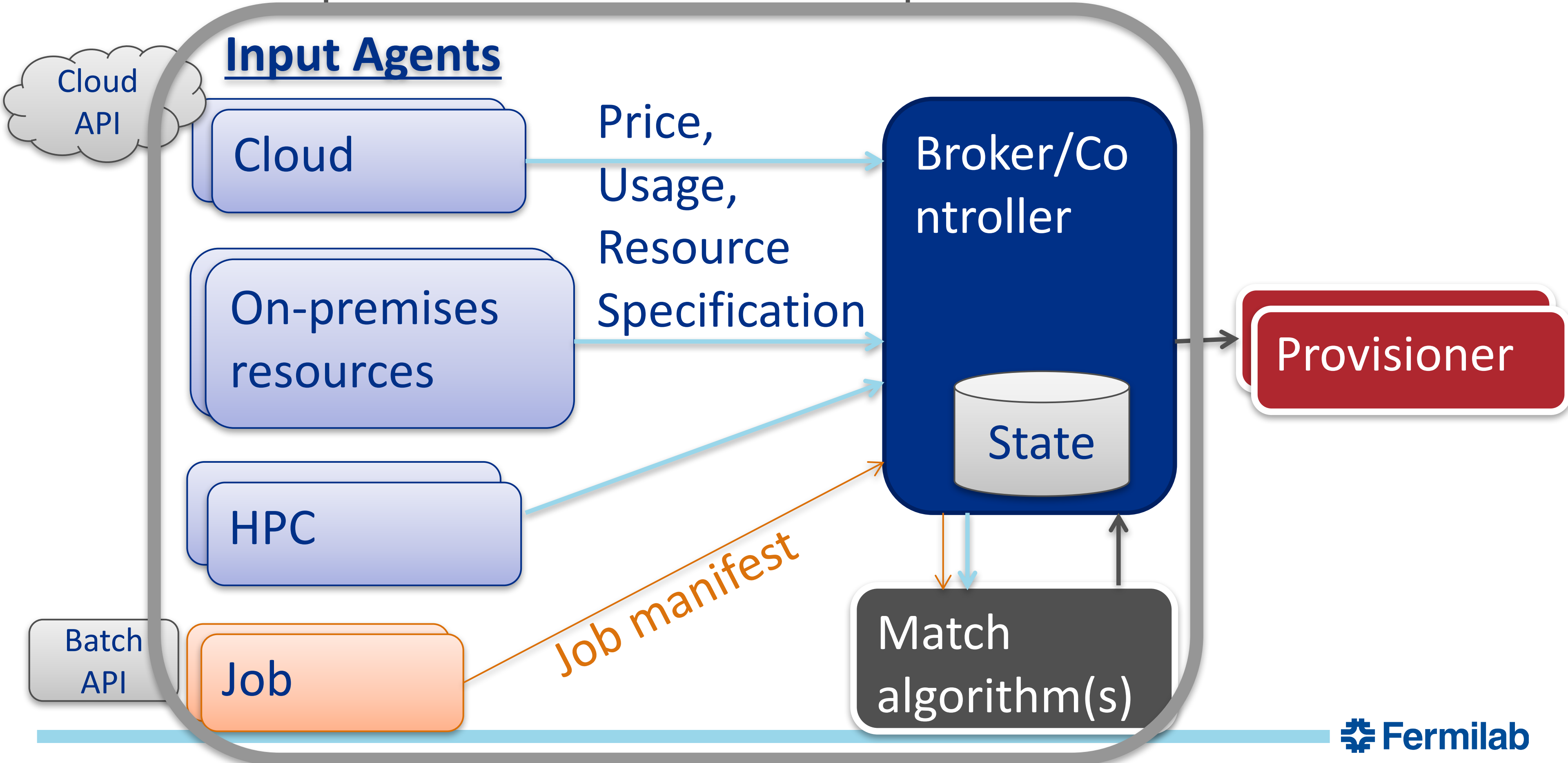
How is the science done?



- Particle Physics: **Statistical Science**
- Comparison with what we know (Standard Model)
- Analyze all data and look for deviations ➜ Needle in the Haystack

Decision Engine – design & architecture

- Decision Engine chooses what to provision next
  - v1.5 implementation: Strict matching based on processing type
  - v2.0 implementation: Zeroth-order prioritization based on cost



Pythia on Mira – Details

- We incorporated MPI into the main routines, using scatter and broadcast to send out unique parameters. The plan is to start one process on each node, running 64 threads, each with an instance of the pythia-based analysis. We will do this in chunk of 128 nodes, where each chunk is a gather collection point for writing to disk.

- Things were running on our x86 cluster - the porting to power PC of the build tools was the challenging part.

- ~150 core test

Description of CMS workflow

- Four chained steps (output of step N is input of step N+1)
  - Step 1 requires few GB input (“Gridpack”) – same files per job
  - Step 2 requires additional input: “pile-up” data (simulating multiple events per bunch crossing), 5-10 GB

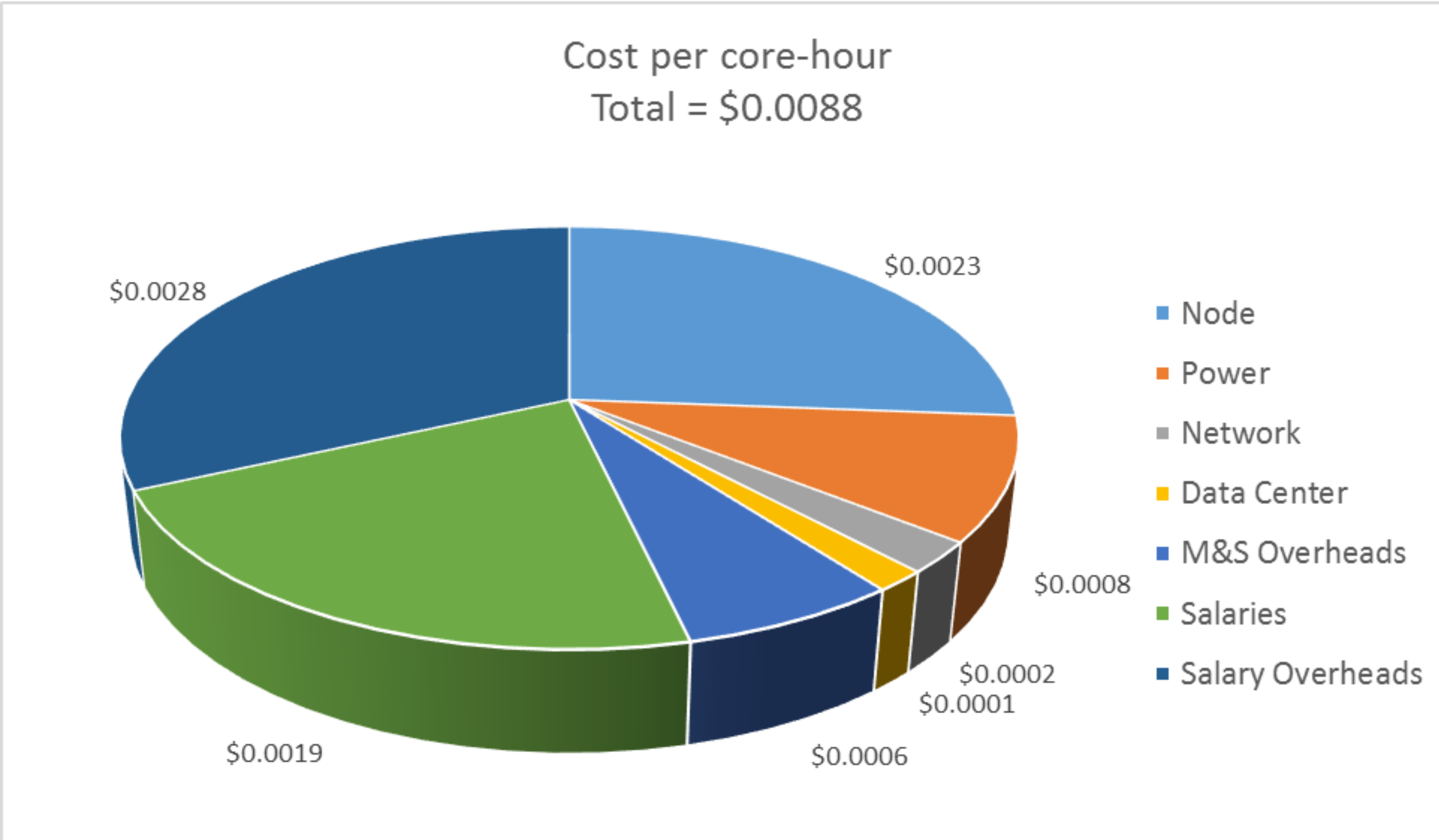
- Pile-up data is constructed on-the-fly by random seek and sequential reads into a 500 TB dataset
    - Staged pile-up datasets to Google Cloud Storage (storage service) ahead-of-time using FTS3 and PhEDEx – standard HEP grid tools and CMS data placement service

Reading pile-up from Google Cloud Storage

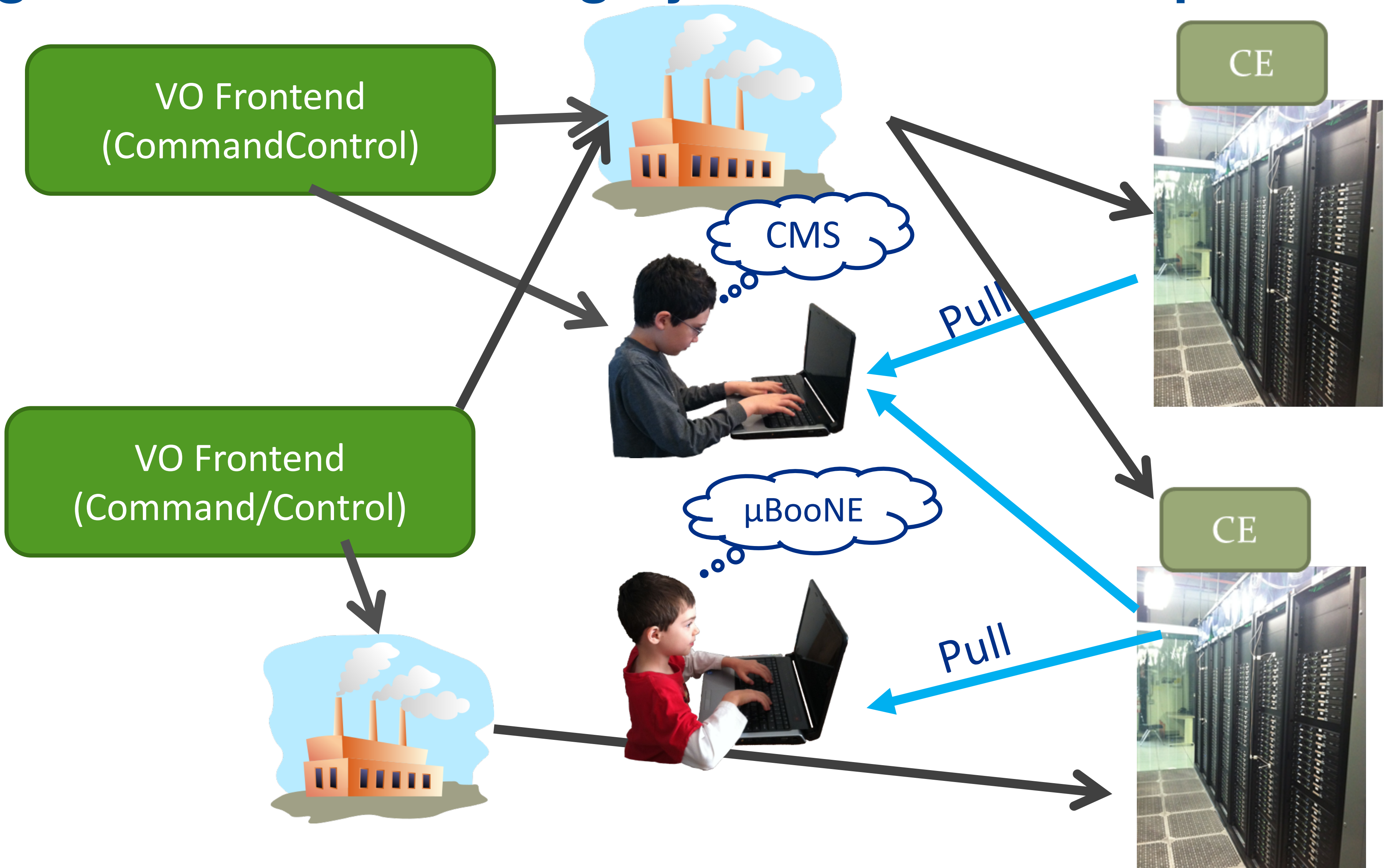
- Mounted regional bucket via gcsfuse on glide-in startup to /gcsfuse
- Used HTCondor “additional_json_file” functionality to specify role tied to image

Elements of the cost per core-hour

Based on Fermilab CMS Tier-1



glideinWMS – Building dynamic HTCondor pools



HEPCloud: Networking

- All models of distributing computing rely on the performance of the underlying (local and wide-area) network
- Fermilab is approaching **1 Terabit** data center – connect to Energy Sciences Network (ESNet) at 4*100 Gigabit
  - ESNet enables distributed computing beyond ESNet sites: 100 Gigabit peering points with other networks

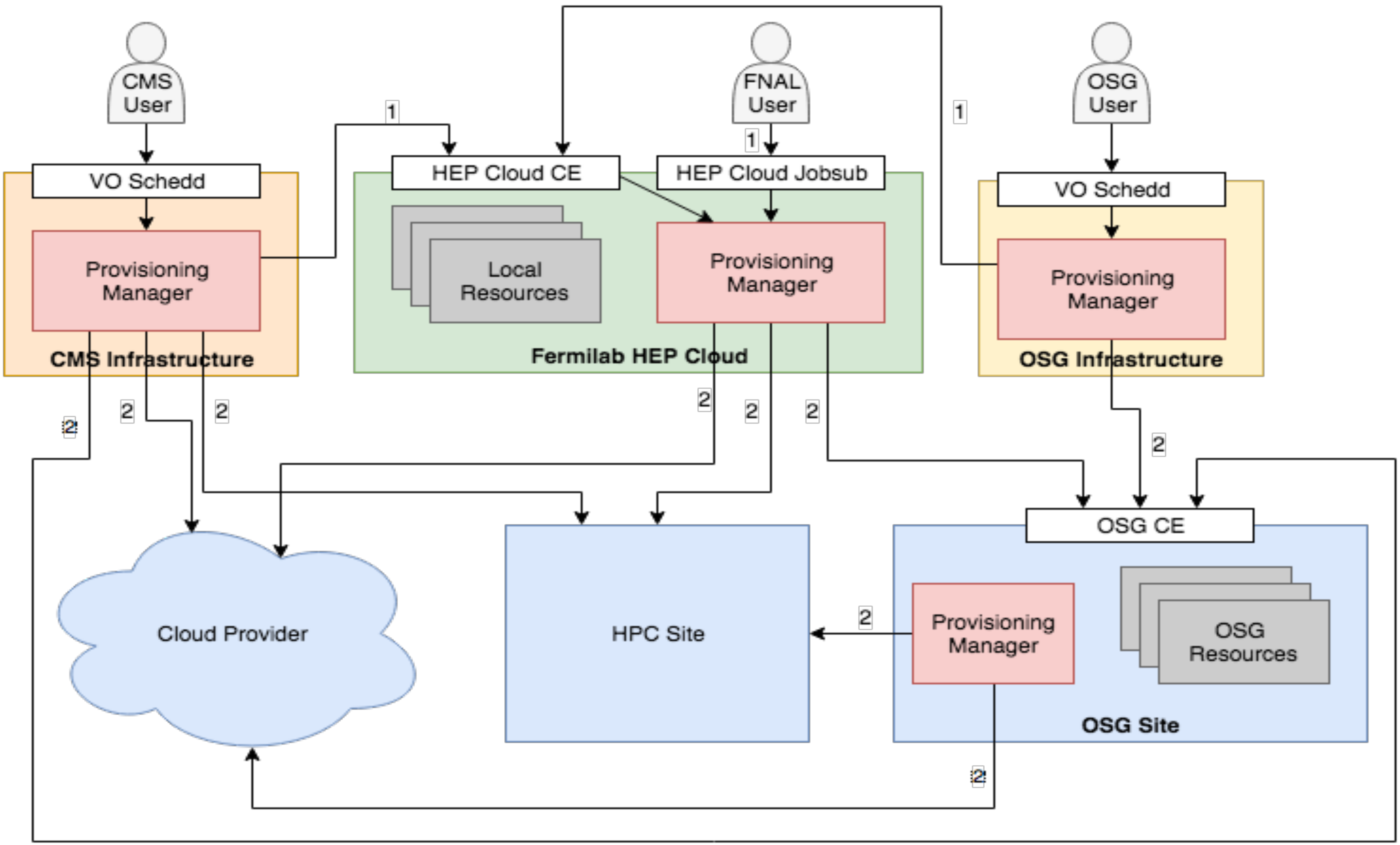
- Zone-based security protection of network resources
- On-demand (**Software Defined Network**-based) traffic controls
- Virtualization of network resources

Fermilab

HEPCloud: Storage

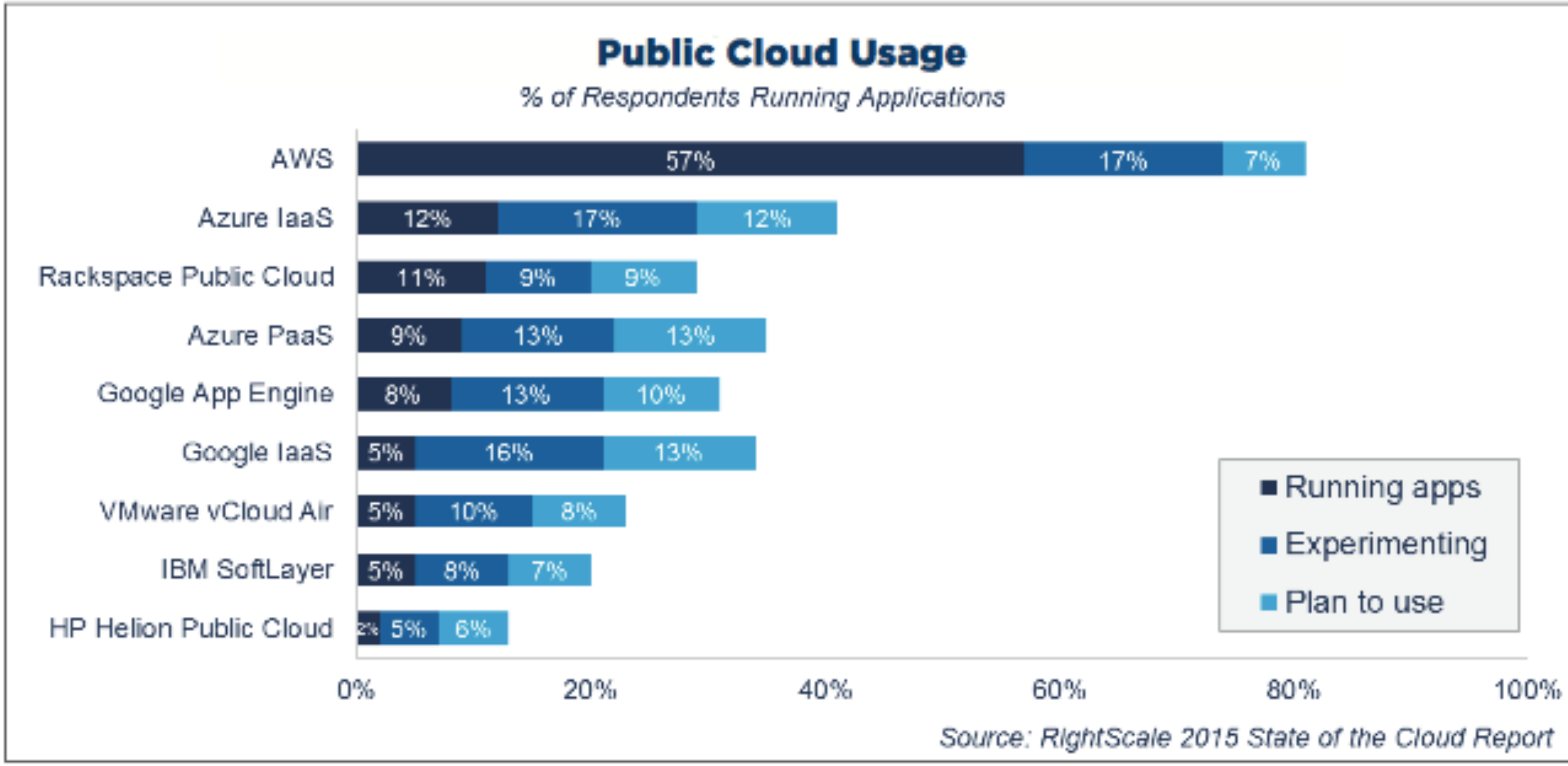
- Data is the lifeblood of science
  - HEP experiments generate it by the station-wagon-load
  - Fermilab is a leader in the field in storing and serving petabytes of data to the world
- We are working with industry and other collaborators to modernize our services
  - Data storage and retrieval
  - Data cataloging
  - Support multiple-layer storage infrastructure approach
- One part of HEPCloud is to understand how to integrate all of these components – always driven by the experiment needs, both present and future

User's View of HEPCloud



Fermilab

Fermilab HEPCloud: expanding to the Cloud



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 United States Government or any agency thereof.