Provisioning 160,000 cores with HEPCloud at SC17

Phil DeMar (FNAL)

LHCOPN/LHCONE Meeting – April 04, 2017

Fermilab
50 Years of Discovery

Challenge: **Doubling** CMS computing using Google Cloud Engine

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

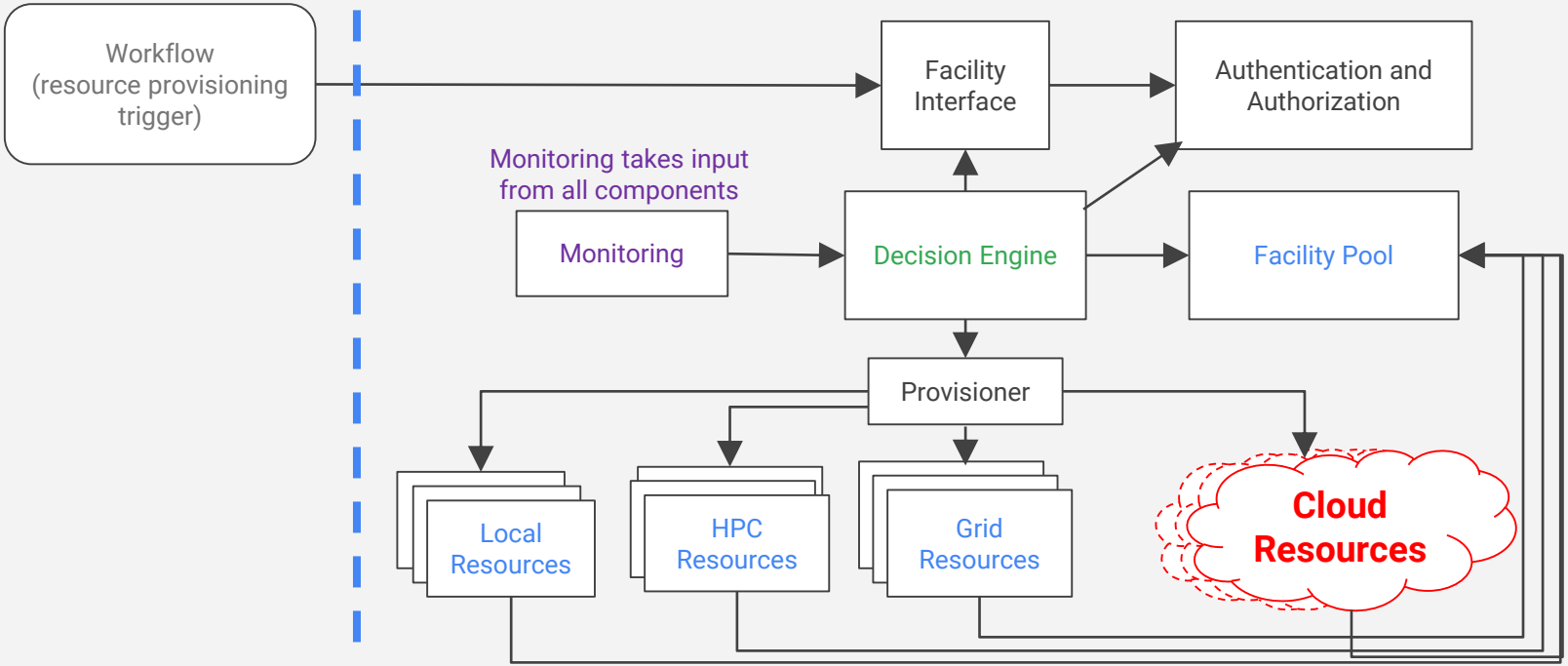
(B. Holtzman; FNAL)

HEPCloud

- 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. LCFs, NERSC)
- **Pilot project** to explore feasibility, capabilities of HEPCloud
  - Collaborative effort with industry, academia
  - Previously evaluated with AWS for NOvA computing
- Goal of moving into production by September 2018

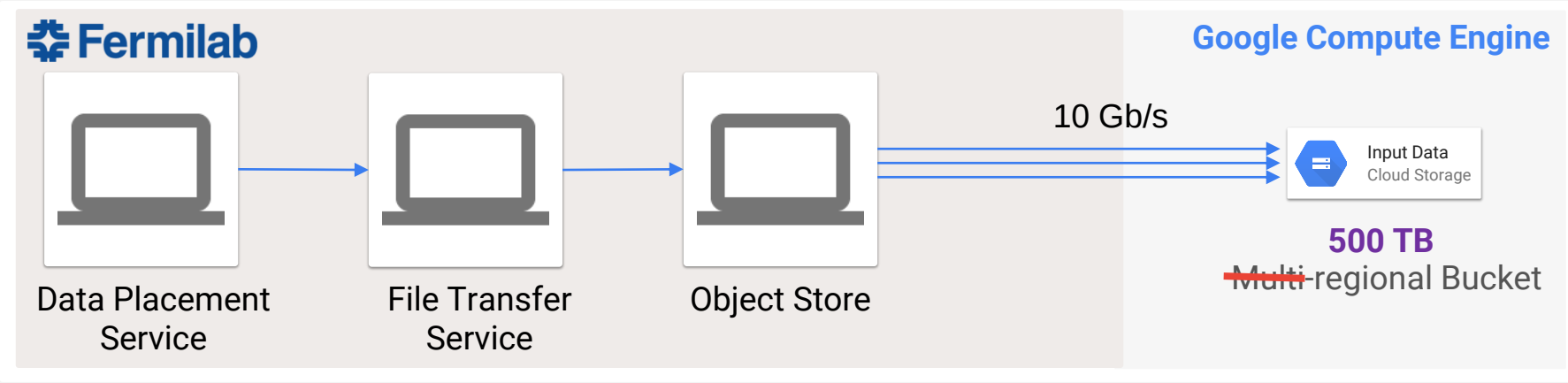
(B. Holtzman; FNAL)

HEPCloud Architecture



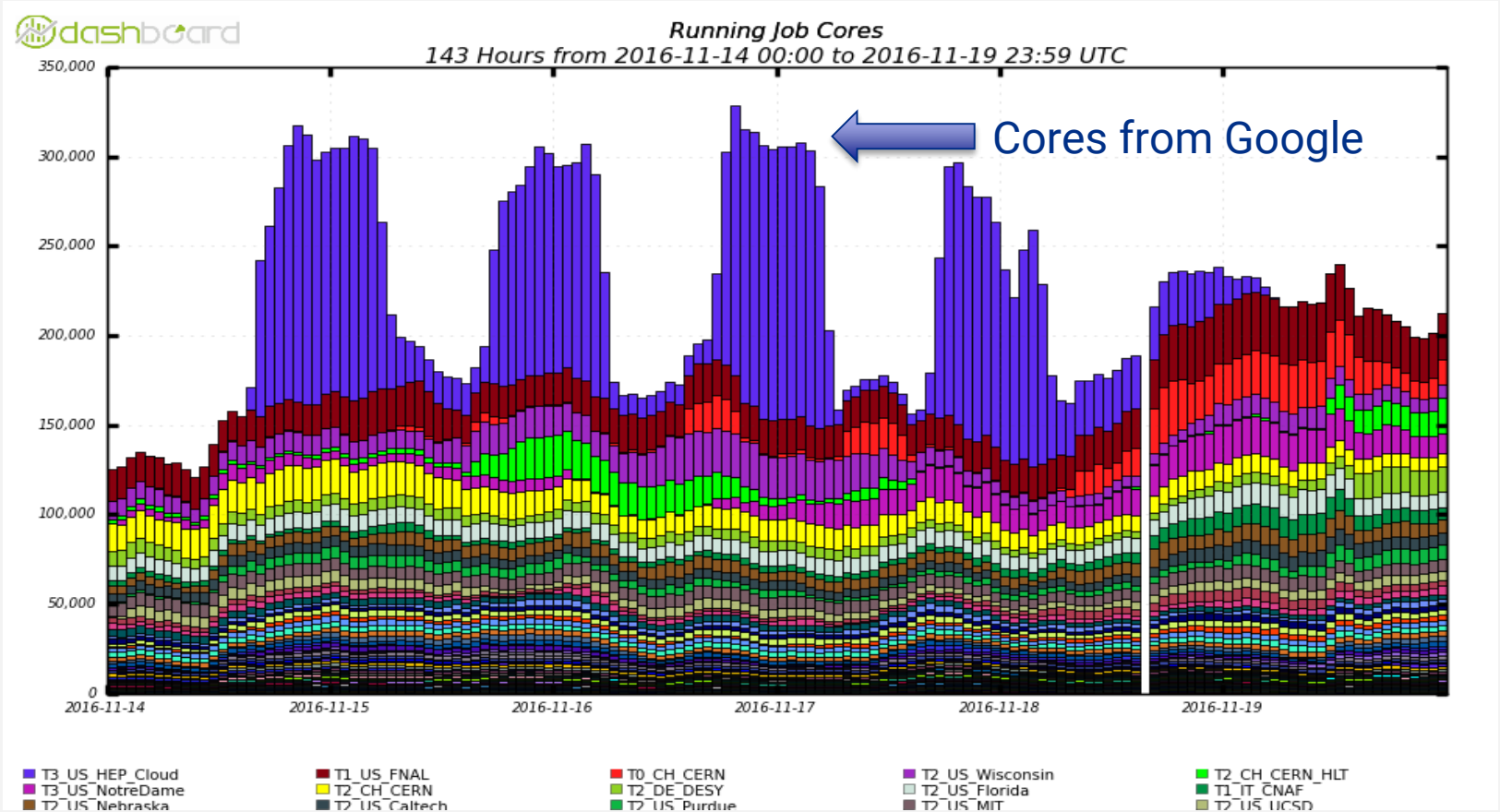
(B. Holtzman; FNAL)

Pre-staging input data to Google Cloud Storage

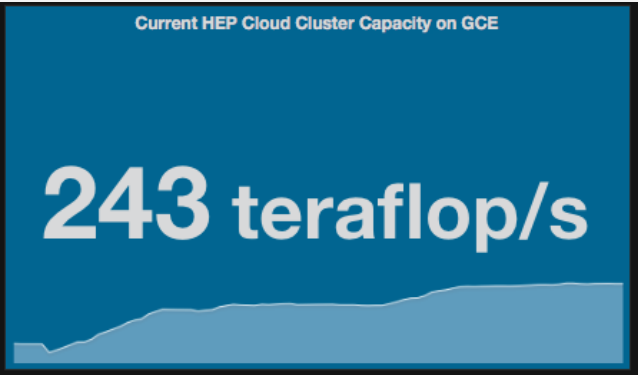


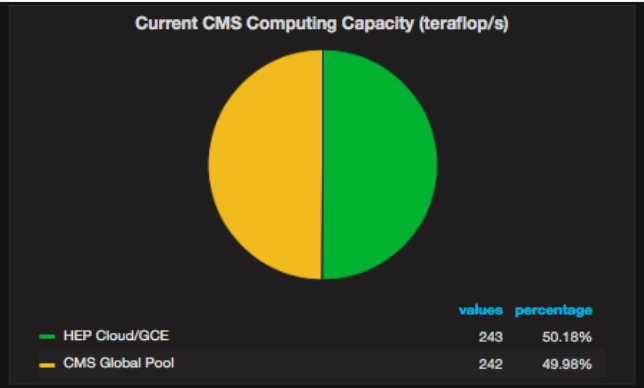
- Experiment-specific data placement service (“PhEDEx”) tracks datasets, schedules transfers
- File Transfer Service supports **S3-compatibility** mode (gfal-copy, davix)
- Google Cloud Storage mounted into preemptible VMs using **gcsfuse** via startup scripts
- Google to ESNet peering (via Equinix @SL) upgraded to **100 Gb/s** capacity (but staging used 10Gb/s…)
- Converted multi-regional to regional bucket overnight: resulted in 30% less cost

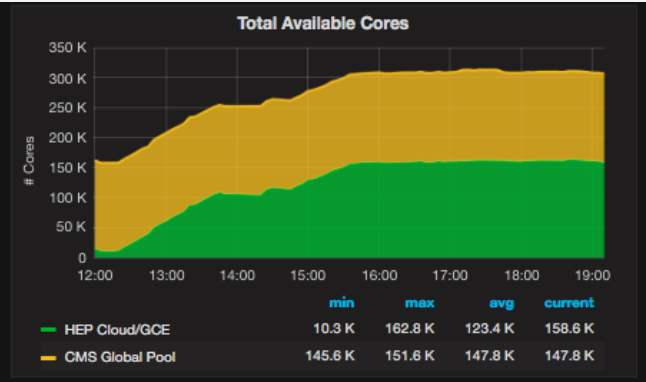
(B. Holtzman; FNAL)

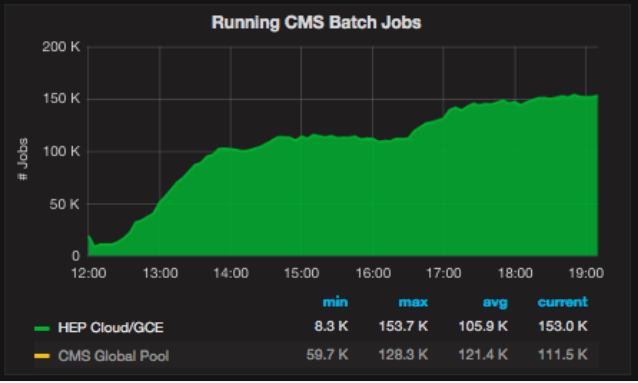


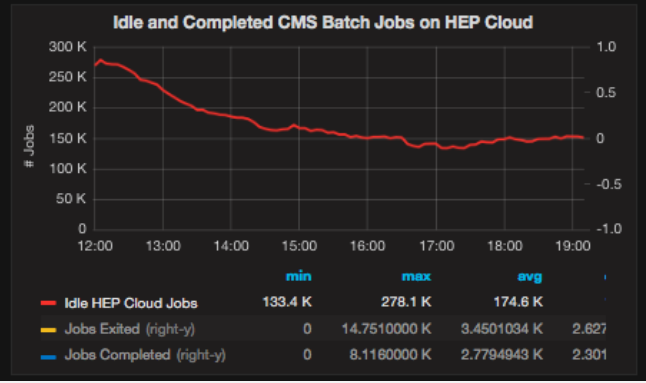
(B. Holtzman; FNAL)











(B. Holtzman; FNAL)

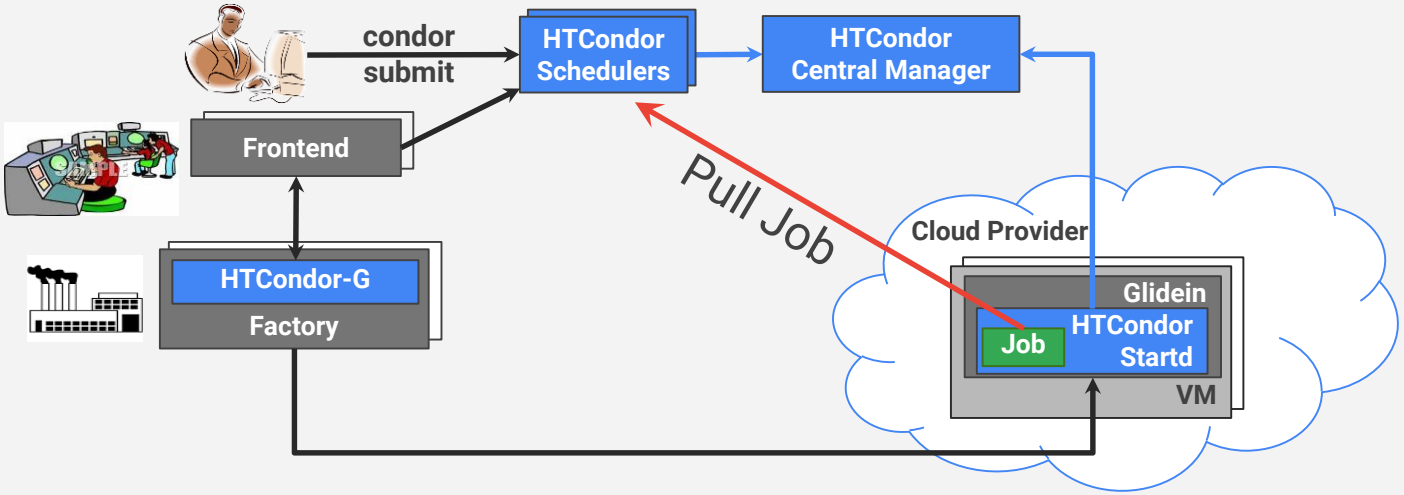
Tale of the tape

- **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)

(B. Holtzman; FNAL)

Additional/Backup Slides

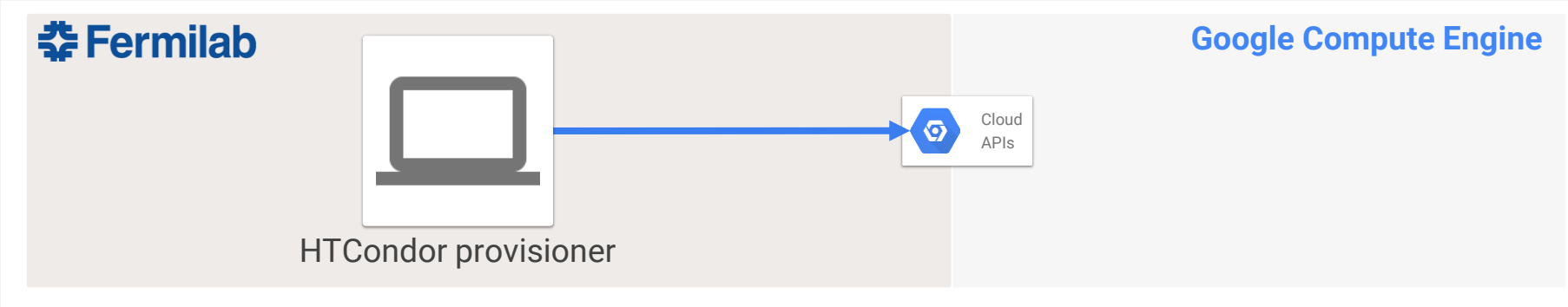
Provisioning remote resources via glideinWMS



- 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 resources)

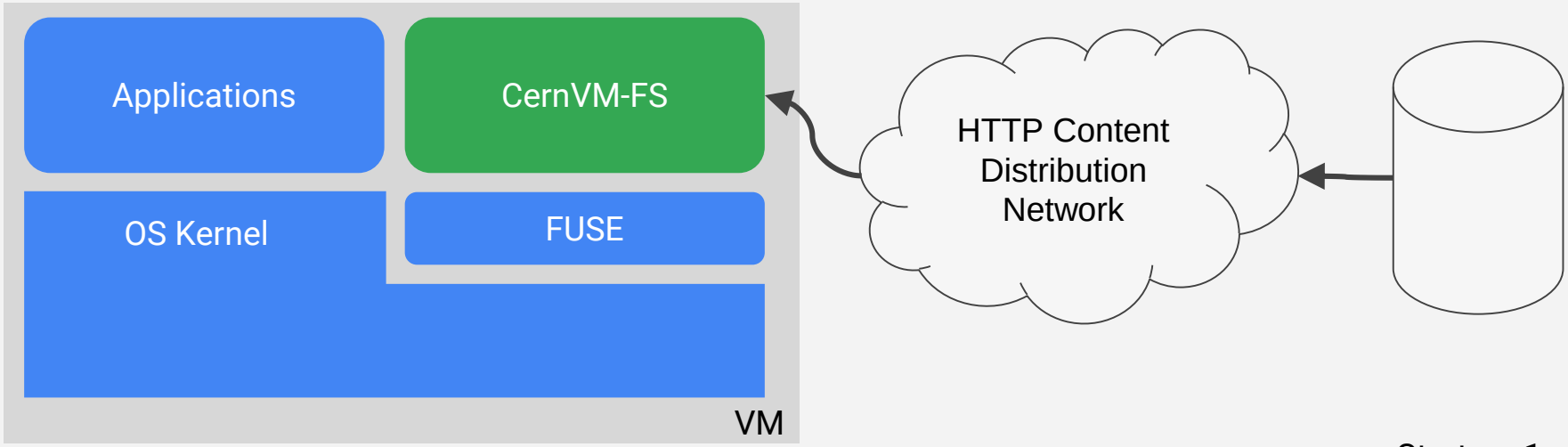
(B. Holtzman; FNAL)

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
  - Fixed **critical bug** to address scaling

Providing application software in a distributed world



Worker Node
Memory Buffer

Worker Node
Disk Cache

Stratum 1
Web Server
Entire Software Stack

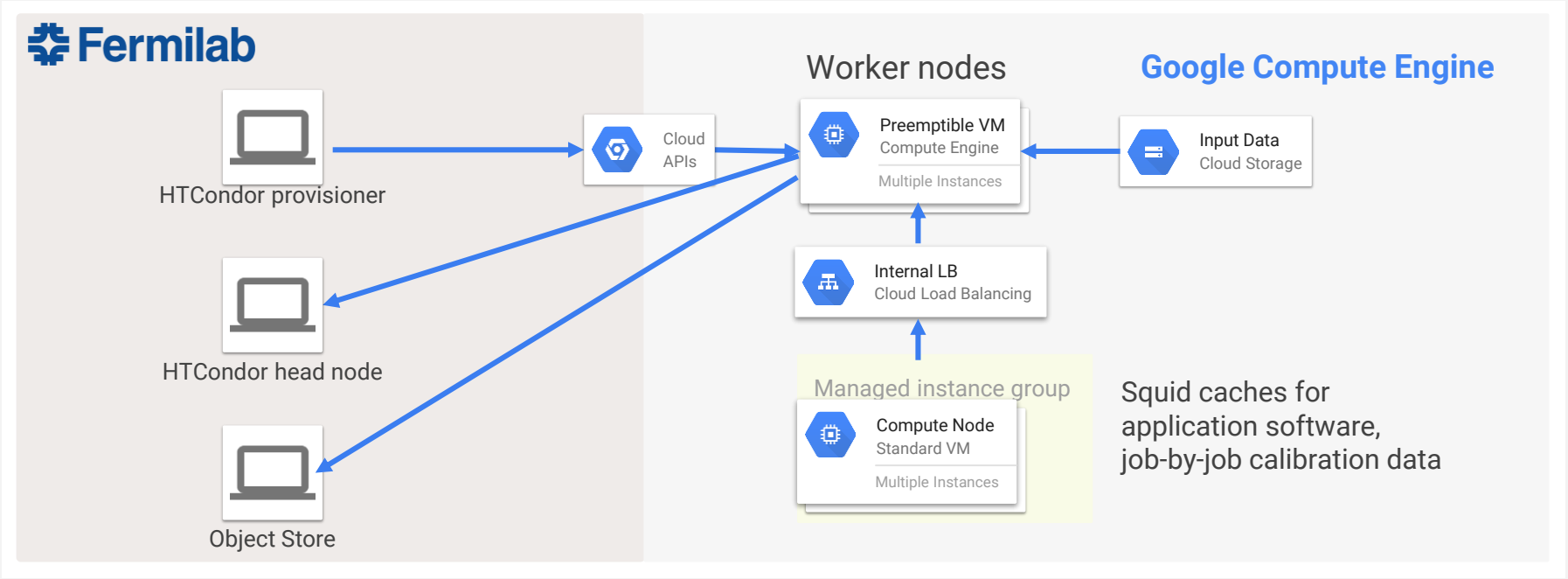
Megabytes

Gigabytes

Terabytes

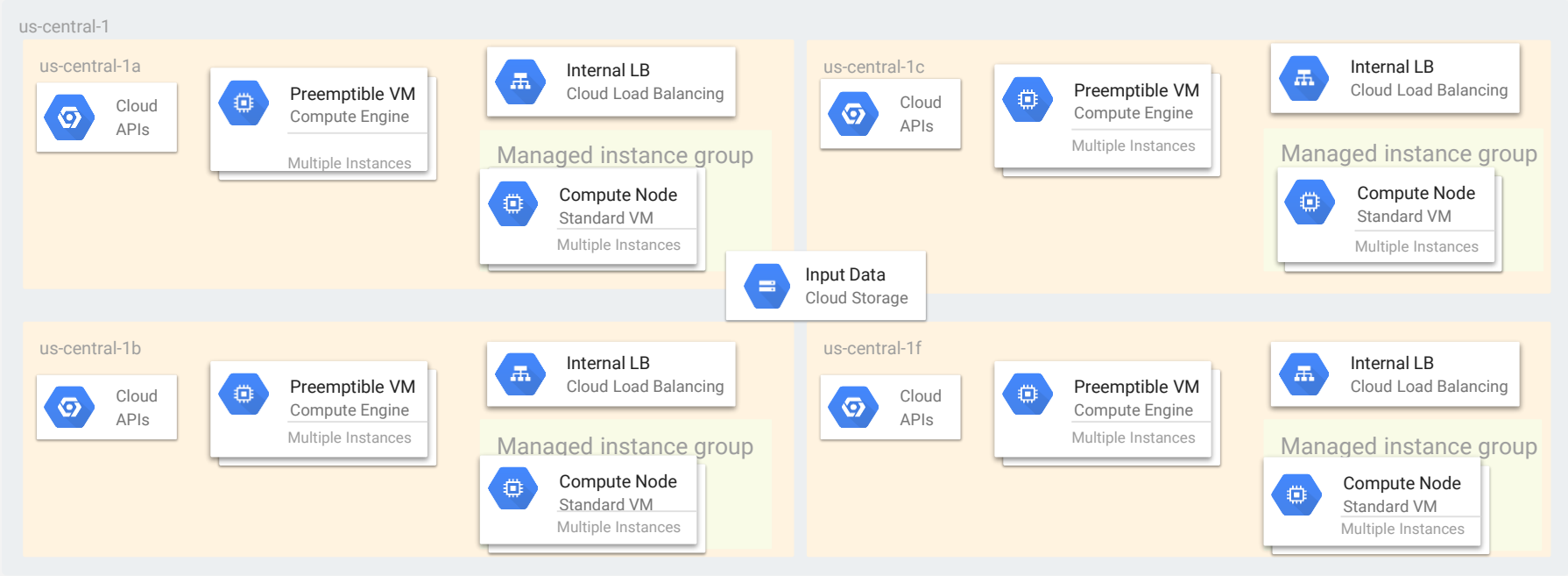
(B. Holtzman; FNAL)

Architecture inside a single zone



(B. Holtzman; FNAL)

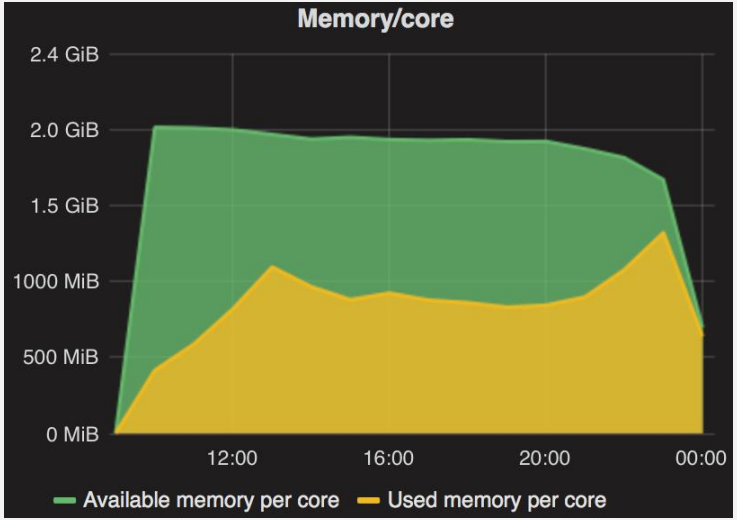
Using 4 zones in us-central-1



(B. Holtzman; FNAL)

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



(B. Holtzman; FNAL)

Next steps

- HEPCloud moves into production in September 2018
  - Decision engine (when and how much to provision) is in R&D
- 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

(B. Holtzman; FNAL)