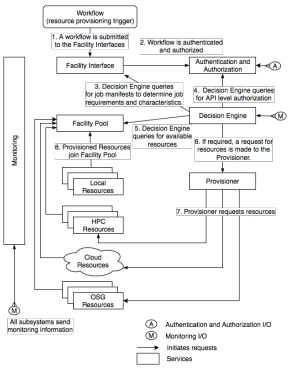
Virtual Machine Provisioning, Code Management, and Data Movement Design for the Fermilab HEP Cloud Facility

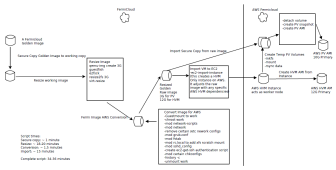
G. Cooper, S. Fuess, G. Garzoglio, D. Grassano, B. Holzman, R. Kennedy, H. Kim, A. Singh, S. Timm, A. Tiradani, Fermilab
R. Krishnamurthy, S. Vinayagam, I. Raicu, S. Ren, H. Wu, Illinois Institute of Technology; S. Y. Noh, KISTI

Fermilab

HEP Cloud Overview

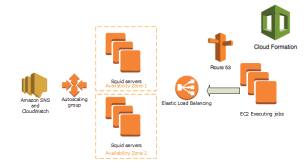


VM Image Management



Start with Scientific Linux Fermi 6 base install.
Add Open Science Grid worker node client.
Convert to "Raw" format.
Upload to AWS and declare as AMI.

On-demand Code and DB Caching

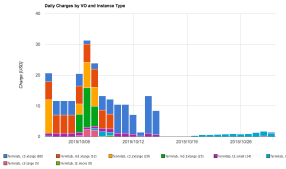


Squid server for
1. Code cache (CVMFS)
2. Database Cache (FronTier)

Squid on-demand components:
- Autoscaling Group
- Elastic Load Balancer
- AWS Route-53 (give a pre-determined name)

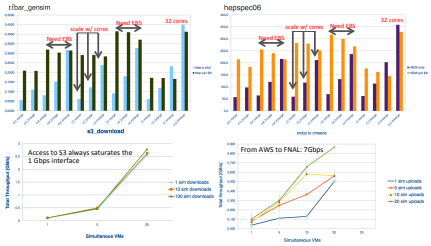
Squid available to AWS only. Peak: 8 per availability zone in 8 zones.

Accounting



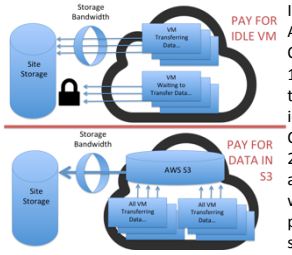
Write AWS Probe for existing Gratia accounting system.
How many VM's running each hour?
How much did we pay?
How many got pre-empted for bid price?

Benchmarking



Average AWS instance, 0.158 ttbar/s
Average CMS tier 1 bare metal instance, 0.163 ttbar/s

Data Movement



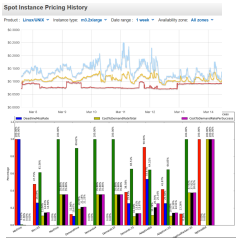
Input data was staged to AWS S3 in both cases
Output Decision:
1) Stage data directly back to Fermilab—more cost in idle VM's. CMS
OR
2) Write data to AWS S3 and stage back data later with dedicated file mover process – more cost in S3 storage. NOvA

AWS waived data egress fees up to 15% of total bill for scientific networks.

Monitoring

Use BOTO routines to query the state of AWS.
Custom scripts to query our batch pool.
Graphite backend.
Grafana dashboard.

Prediction and Decision

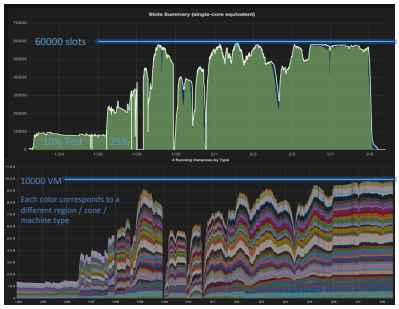


Prediction Engine calculates the probability that a VM will finish without preemption in a variety of bidding models.
We chose to bid 25% of on-demand price.
Estimated pre-emption rate ~10%.
Decision Engine picks VM with best price / performance at that time.

Acknowledgements

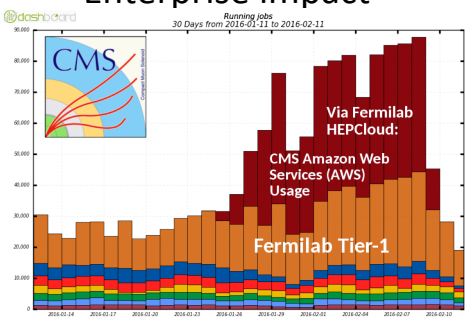
amazon web services™
Scientific Linux
KISTI www.kisti.re.kr
ILLINOIS INSTITUTE OF TECHNOLOGY
INFN
Fermilab
CMS
U.S. DEPARTMENT OF ENERGY | Office of Science
NOvA
HTCondor High Throughput Computing

CMS Full Scale Running



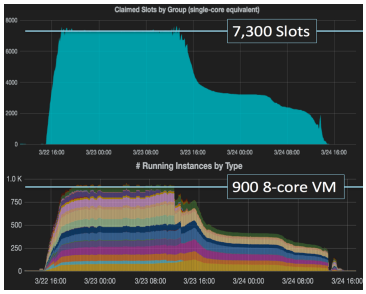
Sustained Peak—60000 job slots, 10000 AWS Instances.

Enterprise impact



The dark red shows HEP Cloud elasticity in terms of running jobs bursting to AWS. This corresponds to a net 25% increase of total CMS resources.

NOvA Full Scale Running



Data and memory intensive application.
Had to use bigger instances only (m3, m4, r3).

Fermilab is operated by the Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy. This work is also supported by KISTI and Fermilab under the joint Cooperative Research and Development Agreement CRADA-FRA 2015-001 / KISTI-C15005.