IceCube in the Clouds

By Benedikt Riedel

The March Toward the Clouds: MF Perspectives

March 2, 2022
CI Compass Cyberinfrastructure for NSF Major Facilities Workshop
IceCube in the Clouds

Benedikt Riedel
UW-Madison

CI Compass
March 2, 2022
IceCube Science

- Novel instrument in multiple fields
- Broad science abilities, e.g. astrophysics, particle physics, and earth sciences
- Lots of data that needs to be processed in different ways
- Lots of simulation that needs to be generated
IceCube Computing

• Particle Physics Computing
  • Ingeniously Parallel!
  • Globally distributed computing
• Full Stack IT services for 400+ people across 53 institutions in 13 countries
• Basic business IT (email, document management) to large-scale distributed computing
• Cloud touches IceCube in multiple ways
  • Transitioning business services from on-prem to cloud services
  • Already using cloud-based services – Zoom
• Cloud resources could easily compensate for lack of resources – But can we “pay” for them? (Answer: No)
GPU Cloudburst Experiments

- Original Goal: Create an ExaFLOP compute pool in the cloud (80,000 NVIDIA V100) and address review panel recommendations

- First Experiment – On Nov 16 2019 we bought all GPU capacity that was for sale in Amazon Web Services, Microsoft Azure, and Google Cloud Platform worldwide - Creating The Largest GPU Cloud Pool in History
  - 51k NVIDIA GPUs in the Cloud
  - 380 Petaflops for 2 hours (90% of DOE’s Summit, No. 1 in Top 500)
  - Distributed across, US, EU, and Asia-Pacific
  - Cost: $50-150k (under NDA)

- Second Experiment – More realistic test
  - Most cost-efficient GPUs for 8 hours
  - Achieve 1 ExaFLOP-hour of compute
  - Distributed across, US, EU, and Asia-Pacific
  - Cost: ~$60k

- Third Experiment – Network Peering between clouds and UW
  - 100 Gbps between UW and clouds with the right setup – Multiple clients on both sides, Internet2 peering
PARTS OF THE ELEPHANT
IN THE ROOM

© John Atkinson, Wrong Hands

reluctance  denial  ignorance

avoidance  diversion  silence

awkwardness  trunk

© John Atkinson, Wrong Hands • gocomics.com/wrong-hands • wronghands1.com
Institutional Hesitance/Social Engineering

- UWisc took 18 months to negotiate a contract with GCP
  - 4 years after contract with AWS
  - Standard NSF grant is 3 years
  - How would a smaller school or program do this? – Something like NSF’s Cloudbank looks like the solution
  - What about vendors like Oracle? DigitalOcean? Linode?
  - Need to make sure we don’t spend over budget

- Inconsistent overhead policy among instructions
  - 50+% potential price swing

- Internet2-UW direct connections needed to make cloud viable for large scale science
Institutional Hesitance/Social Engineering

• Campus-wide SaaS – Between a rock and a hard place
  • Campus-based SaaS is not negotiated for collaborative cross-institutional science
  • Typically limited to affiliated people (students, staff, faculty) – Collaborators excluded
  • Science is becoming more collaborative – IceCube is no longer an outlier
  • Providing business services to non-UW collaborators can be expensive – Example: Groups are paying more per user than UW

• Billing – Unknown cost can have issues with funding agencies and campus administration

• Push towards “free” existing resources – On campus, NSF/DOE HPC, etc.