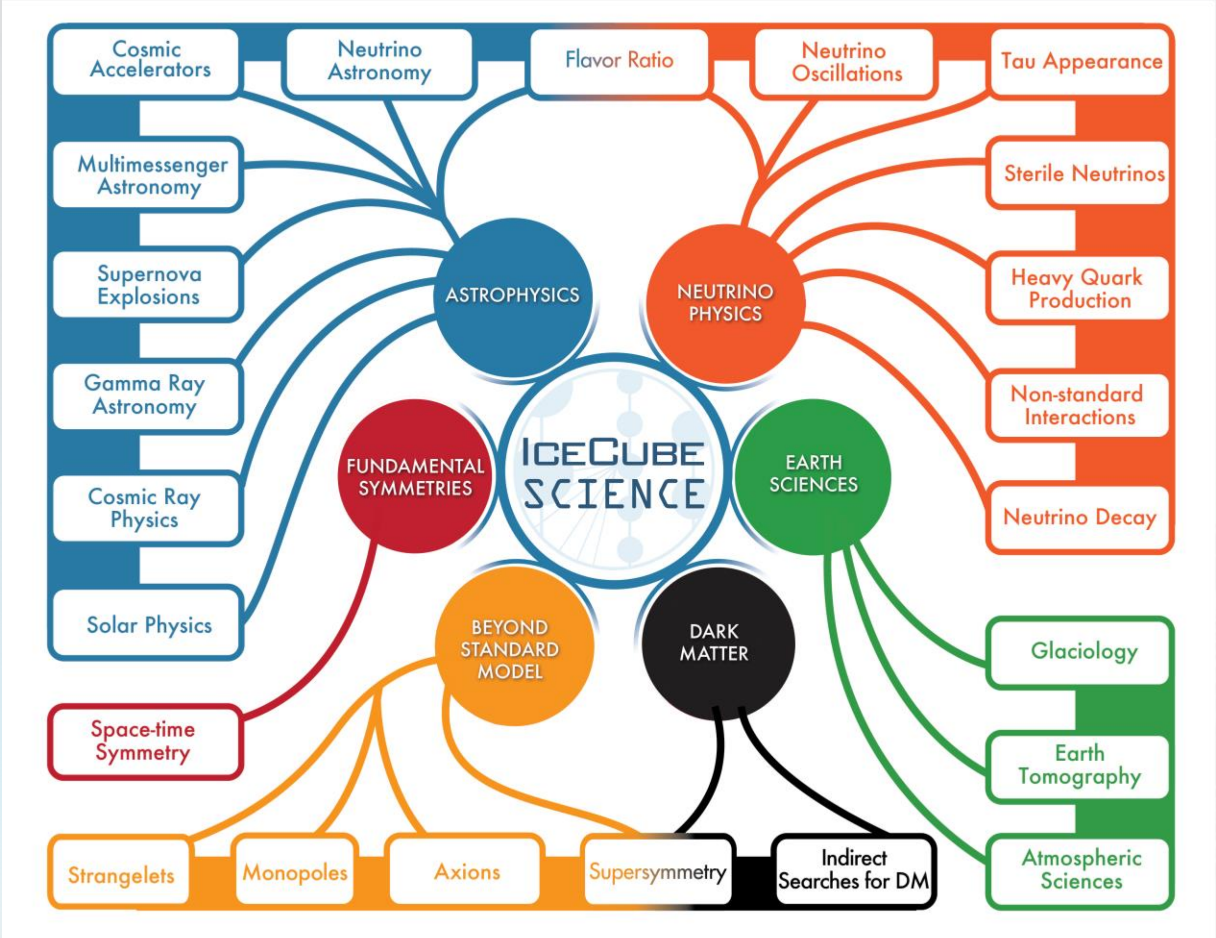
IceCube Computing

Benedikt Riedel

HTCondor Week 2019

May 21 2019

IceCube Computing – What drives us?



- 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

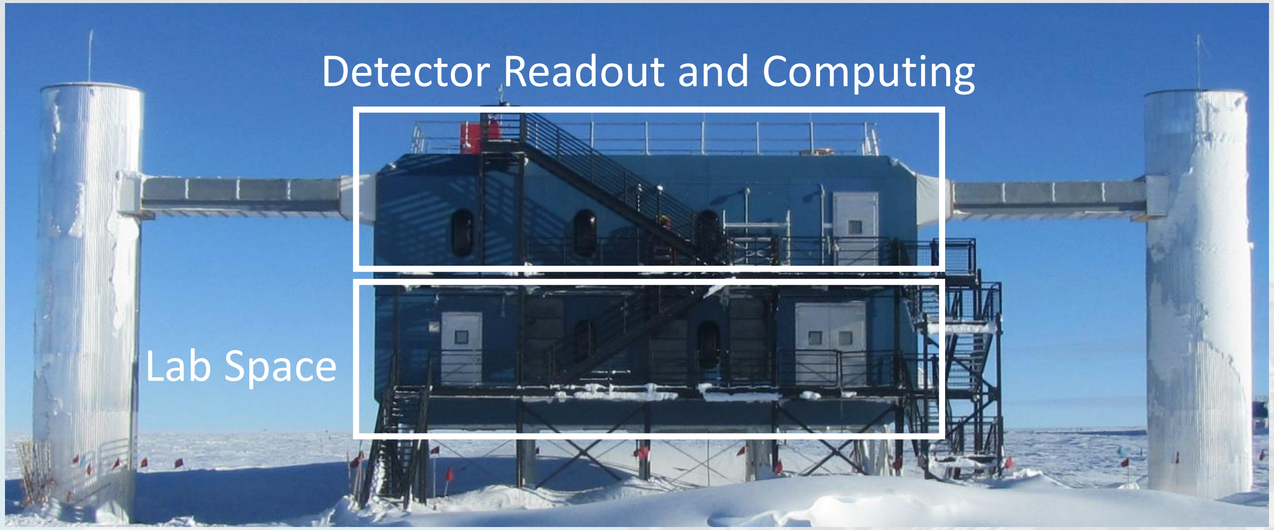
IceCube Computing – 30000 Foot View

- Classical Particle Physics Computing
  - Trivially/ingeniously parallelizable – Grid Computing!
  - "Events" - Time period of interest
  - Number of channels varies between events
  - Ideally would compute on a per event-basis
- Several caveats
  - No direct and continuous network link to experiment
  - Extreme conditions at experiment (-40 C is warm, desert)
  - Simulations require "specialized" hardware (GPUs)
  - In-house developed and specialized software required
  - Large energy range cause scheduling difficulties – Predict resource needs, run time, etc.

South Pole Cyberinfrastructure – Data Management

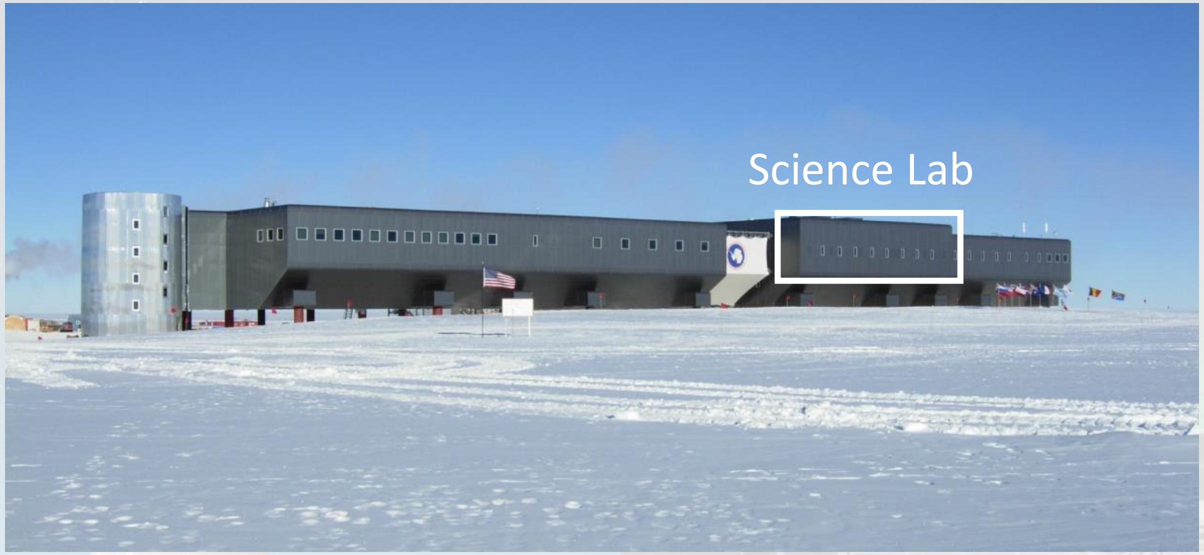
- Data Rate – 3 TB/day
- Using both data transfer options for data transfer – Drives/tapes and satellite
- Limited bandwidth from South Pole to Northern Hemisphere – 125 GB/day
- High bandwidth, high latency – Disks transfers every austral summer
- Need to filter data down to from ~3 TB to ~80 GB

South Pole Cyberinfrastructure – IceCube Lab



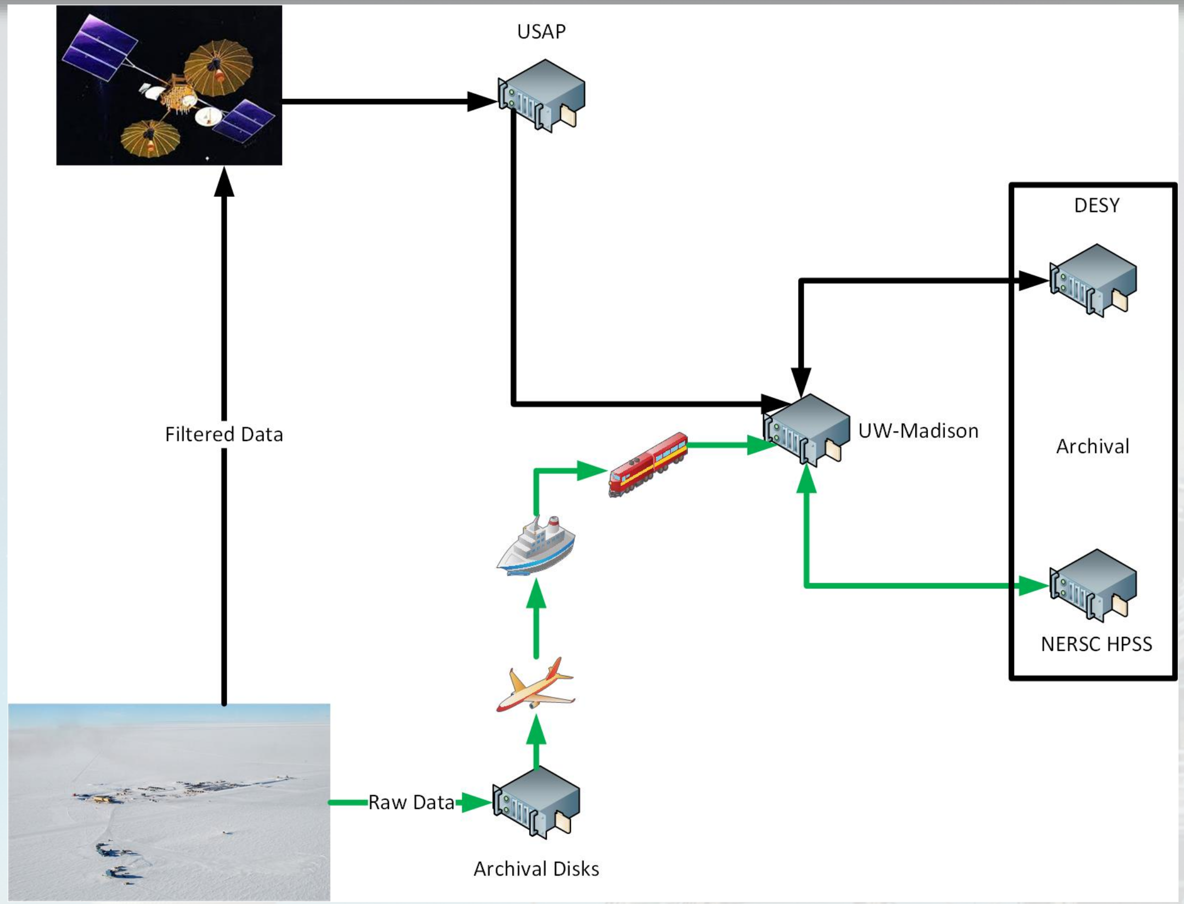
- ~500 core filtering cluster
- ~100 machines for detector readout
- Fiber connection to main station
- Data is triggered and filtered at the lab and shipped off to main station for "archival" and satellite transfer
- Cooling is an issue if air handlers freeze shut – Front of room freezes while back at 80 C
- Power can drop out randomly

South Pole Cyberinfrastructure – Station



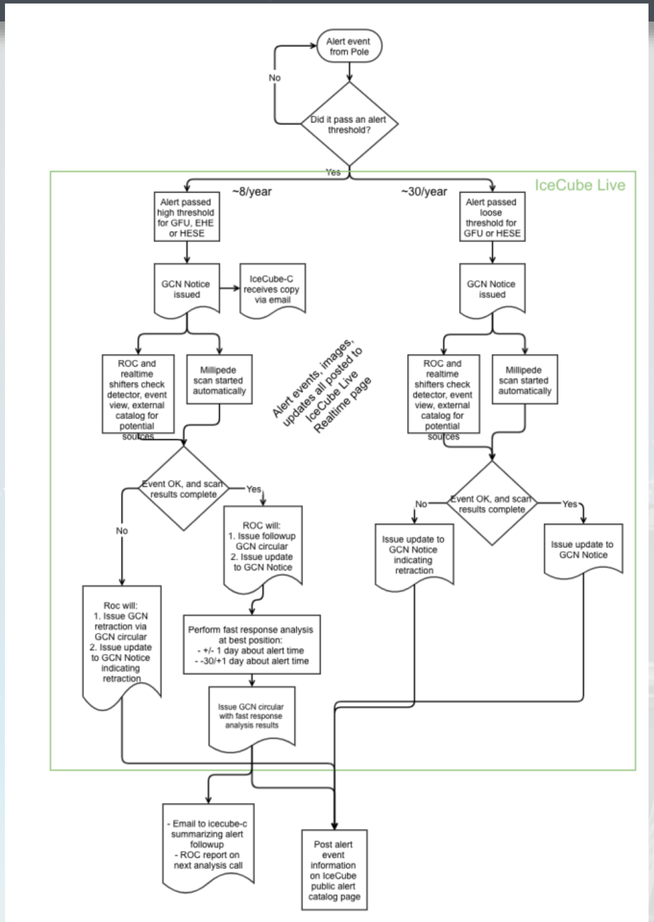
- Amundsen-Scott South Pole Station
- Lab with disk arrays for archival and servers to transfer data US Antarctic Program satellite transfer

South Pole Cyberinfrastructure – Data Flow



- Filtered Data comes via satellite
- Raw data is shipped once a year on disks – First on plane, then boat, finally

South Pole Cyberinfrastructure – Alerts

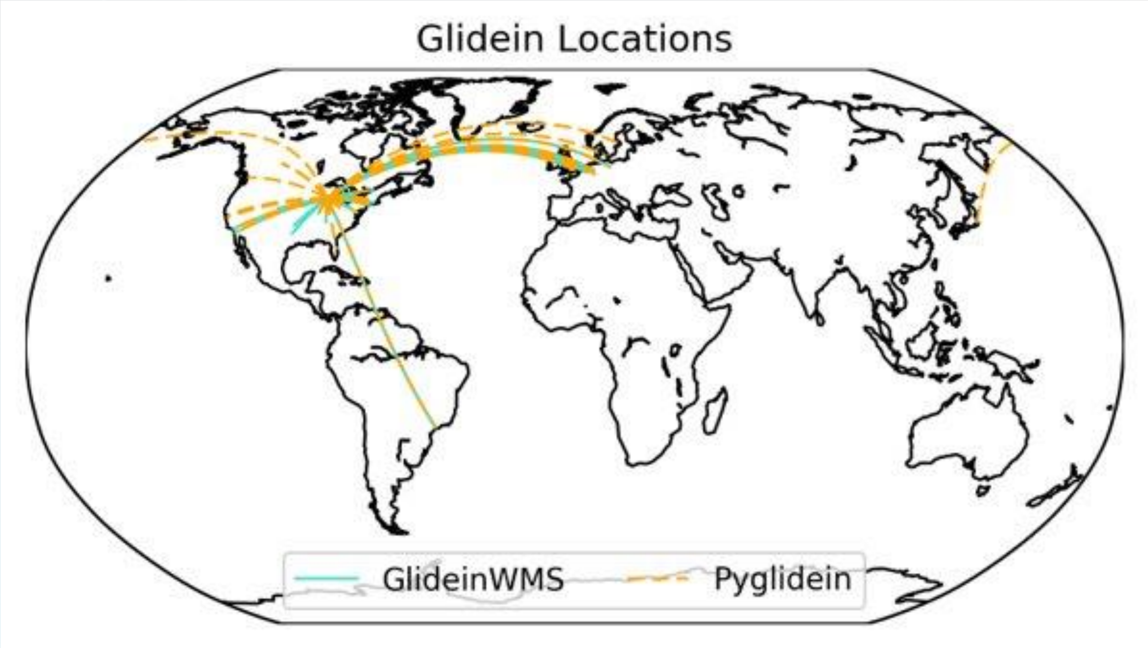


- Alerting the community about interesting events – Multimessenger Astrophysics (one of NSF's 10 Big Ideas)
- Want to alert the community at large about interesting events
- Fast event stream that is separate from main data frame
- Special filtering based on previous analyses
- Alerts are currently limited by
  - Knowledge about neutrino sources – Is it astrophysical?
  - Available CPUs for follow-up studies to improve error on direction on the sky – Very bursty usage, 12000 cores for 30 min once a month

Northern Hemisphere Cyberinfrastructure

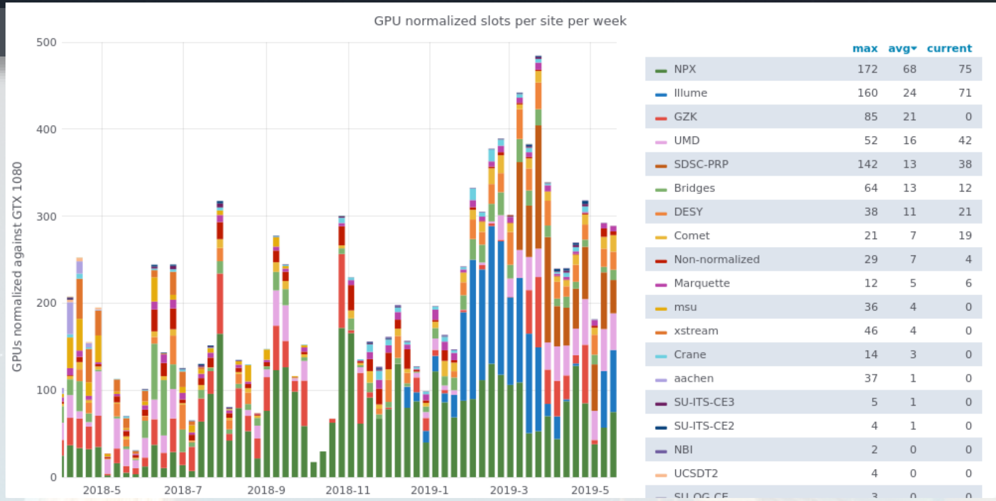
- Central Data Processing and Analysis Facility at UW-Madison
  - ~6500 core, ~300 GPU cluster
  - ~10 PB storage – Roughly even split between data, simulation, analysis output, user data
  - Connected to SciDMZ through Starlight – ESNet for connection to DOE facilities
  - End user analysis infrastructure
  - Access to IceCube Grid, OSG, and EGI
- Every group has respective campus-based resources, e.g. campus cluster
- Pledge system to contribute CPU and GPU
- Use XSEDE (and DOE) resources – Mostly for GPU, scavenge allocated CPU, DOE resources (Titan) hard to use or just added (NERSC)
- Use CVMFS to distribute software

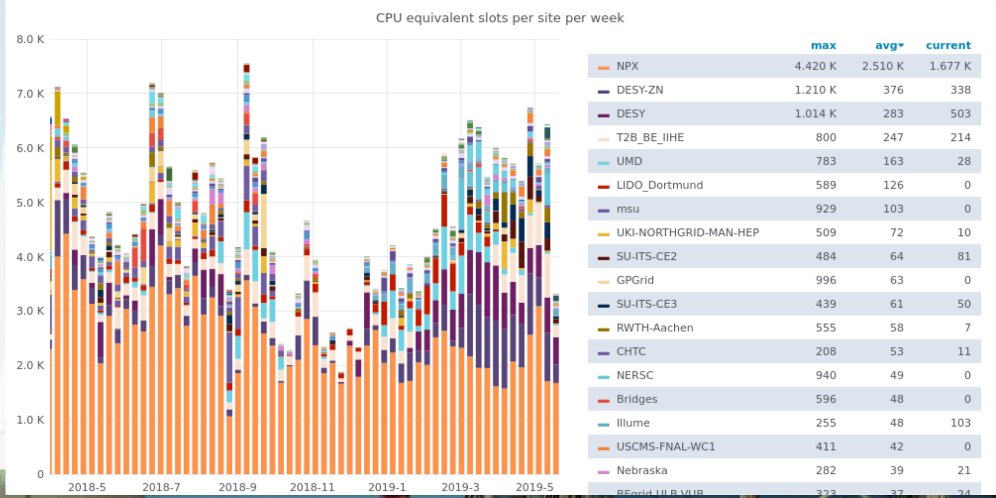
Northern Hemisphere Cyberinfrastructure – IceCube Grid



- IceCube has computing allocations at campus facilities, national facilities (XSEDE), and uses opportunistic computing
- Resources are a mix of both CPU and GPU
- Depending on facility the usage ranges from few hours to ~55M hours per year
- In-house developed software to tie resources together and workload management

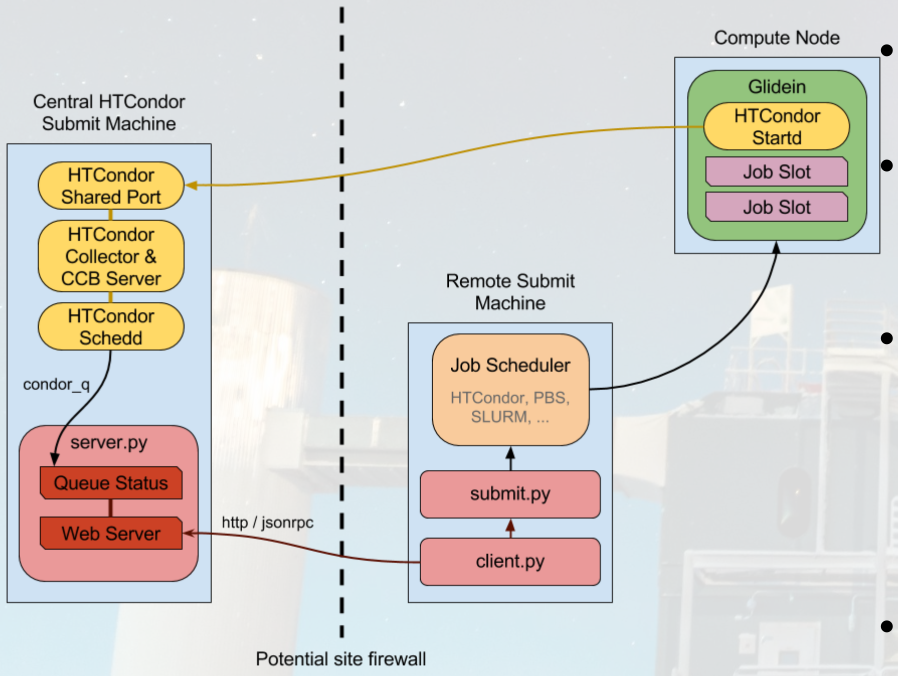
Northern Hemisphere Cyberinfrastructure – IceCube Grid





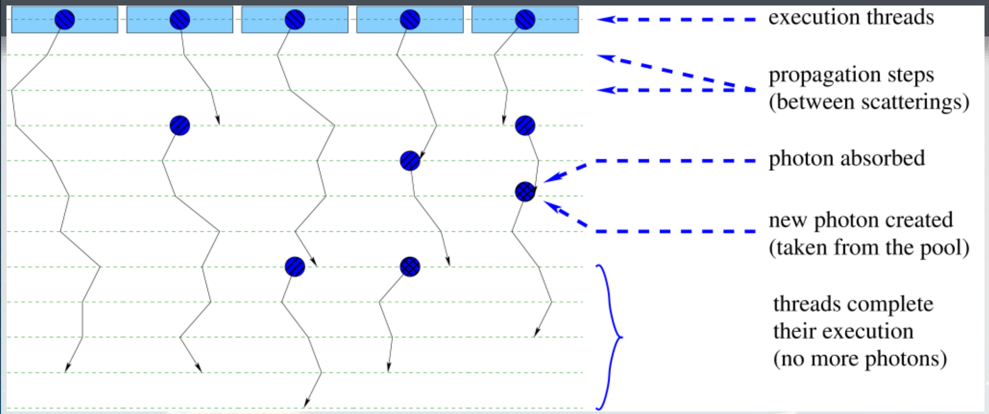
- Steadily expanding resources
- Fairly continuous use
- Slow transition to the "grid" for users – Biggest pain points are data access and job failure
- Big issue – Lots of scavenging of resources and transition between CPU and GPU resources means a lot of data movement

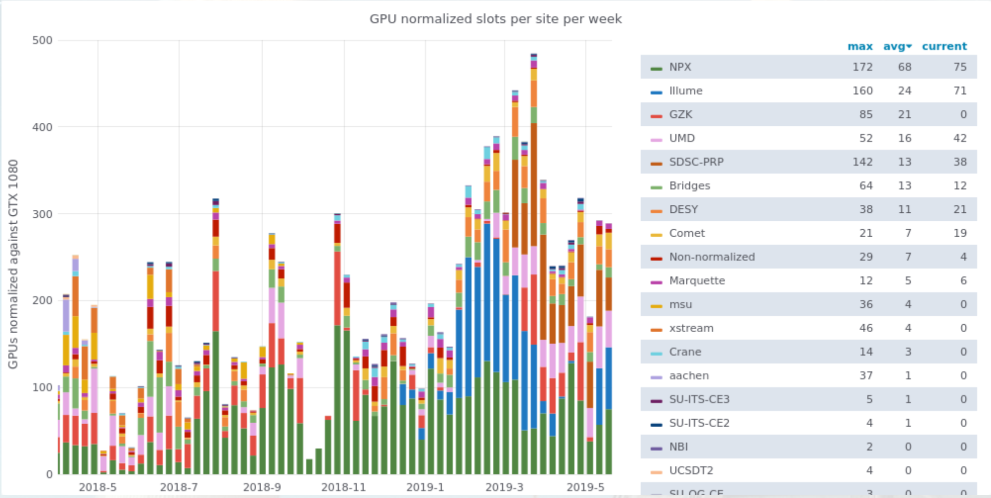
Northern Hemisphere Cyberinfrastructure – Pyglidein



- Pyglidein – In-house developed Python library that starts jobs on remote sites - Pull jobs to remote site
- Lightweight as possible – Knows how to query server and submit to local scheduler
- Server-side
  - Server reads a HTCondor queue
  - Determines job requirements
- Client-side
  - Client periodically queries server for jobs
  - If jobs match site-specific requirements, submit a job
  - Job will execute a HTCondor `startd` and connect back to global pool
- No advanced logic
  - No on limit number of a times a task is submitted – Will be used by other jobs or die quickly
  - No job routing

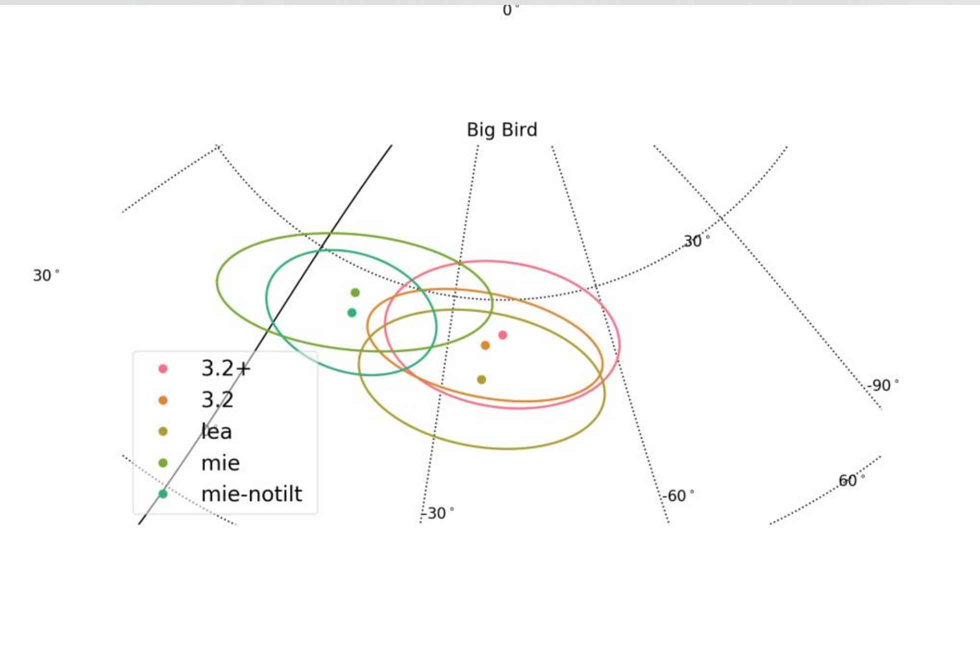
Northern Hemisphere Cyberinfrastructure – GPUs





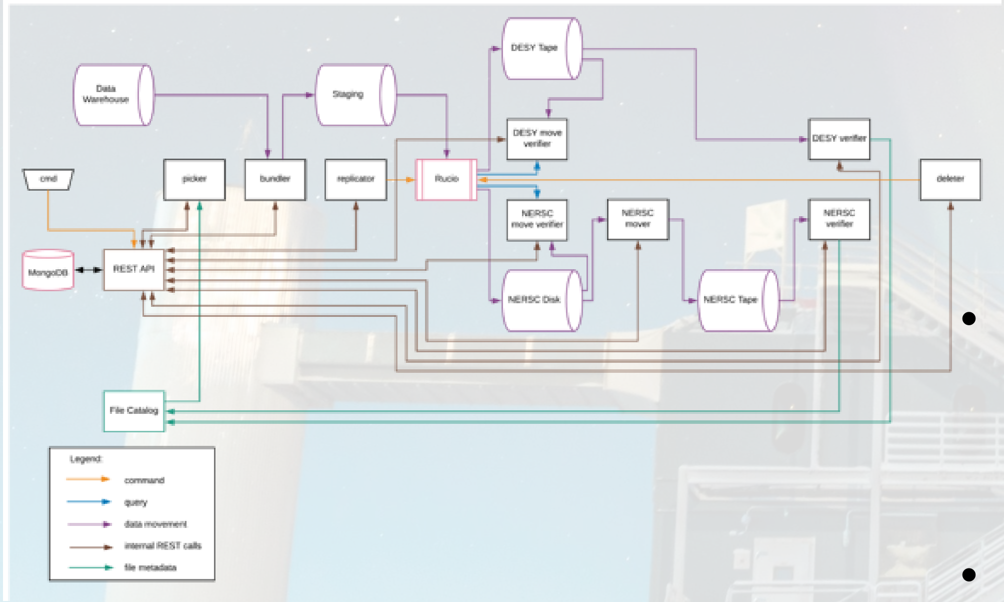
- Why does IceCube need GPUs? – Propagating photons produced by neutrino interaction products in the ice
- Calibration has to be all done *in-situ* – Little information about optical properties are beforehand
- Previously statistically modelled
  - Could not account for all optical properties of the ice
  - Discovered new optical features in the ice
- GPUs provide 100-200x speed up compared to CPUs
- Still a scarce resources – Most GPUs are bought by member institutions
- Currently ~300 GPUs dedicated, another ~500 GPUs pledged
- Biggest bottleneck – Resource contention

Northern Hemisphere Cyberinfrastructure – Ice Model



- Modelling the ice is very important – Esp. In era of Multimessenger Astrophysics
- Want to alert the community at large about interesting events
- Need to inform telescopes where to point
- Ice model can shift the location of event on sky significantly
- Optical telescopes have a minute area of the sky they cover
- Need to be as precise as possible, else wasting valuable telescope time or will miss source (transient sources)

Northern Hemisphere Cyberinfrastructure – Current Projects

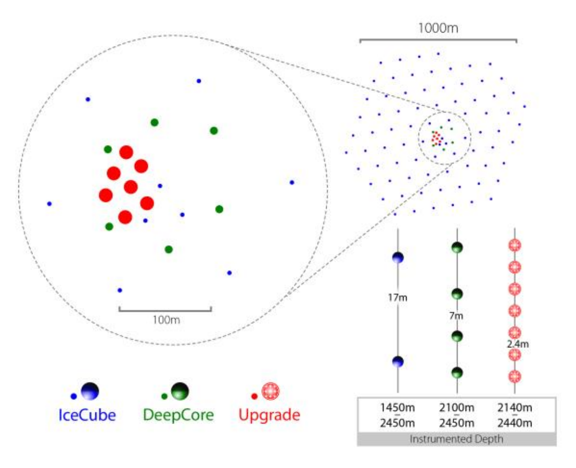


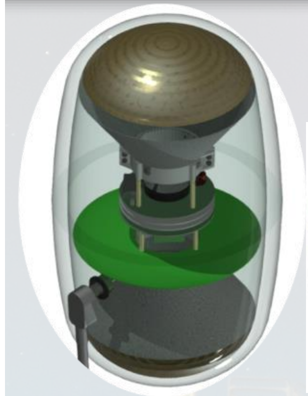
- Cloud Computing – E-CAS award from Internet2
- Machine Learning
  - Machine learning becoming more popular
  - Building first test infrastructure – Already have experience with running and using GPUs
  - First results are promising – Needs more study before deployment in production
- Backups
  - Refactoring code that modes data to tape backups at DESY and NERSC
  - Part of CESER grant
- Expanding resources – More XSEDE resources and campus resources
- Automated and user CVMFS builds

Northern Hemisphere Cyberinfrastructure – Future Projects

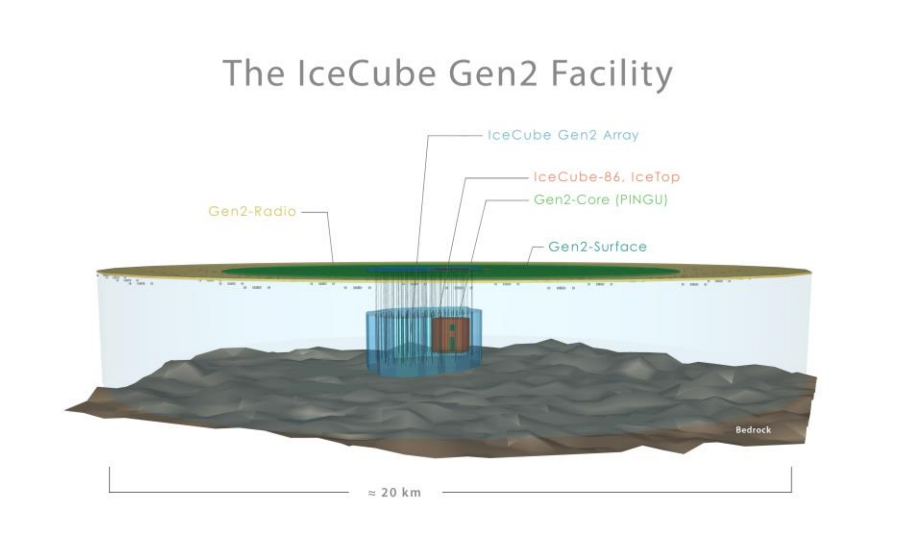
- Re-thinking data organization, management, and access
  - Xrootd-based solution?
  - Spreading data across multiple locations?
  - Ceph-based solution?
  - www-based solution?
- Other resources
  - Cloud
    - Bursting into cloud for multimessenger studies?
    - Using cloud GPUs?
    - Cloud machine learning resources?
  - Resource sharing in multimessenger astronomy
- Continuous integration/deployment
  - Starting with production software
  - Science software – How to test properly?

Future of IceCube





- IceCube Upgrade
  - Deploying next generation detector modules in an in-fill
  - Lower energy threshold
  - Test new technology and designs for future expansions
- IceCube-Gen2
  - Much larger detector focused on high energies
  - Including several ways to do astroparticle physics at the South Pole – Radio detection of neutrinos, air Cherenkov detectors, etc.
- Will need to rethink computing



Summary

- Globally distributed, heterogenous resources pool
- Atypical usage model, resources requirements and software stack
  - Mostly opportunistic and shared usage
  - Accelerators (GPUs)
  - Broad physics reach – Lots of physics to simulate
  - Data flow includes leg across satellite
  - “Analysis” software is produced in-house
    - “Standard” packages, e.g. GEANT4, don’t support everything or don’t exist
    - Niche dependencies, e.g. CORSIKA (air showers)
- Detector up time at 99+% level
- Significant changes of requirements over the course of experiment – Accelerators, Multimessenger Astrophysics, alerting, etc.

Thank you!

Questions?