Computing Report

Benedikt Riedel
UW-Madison

NSF 2024 Mid-Term Review
29 April 2024

IceCube
South Pole Neutrino Observatory

Presenter Background

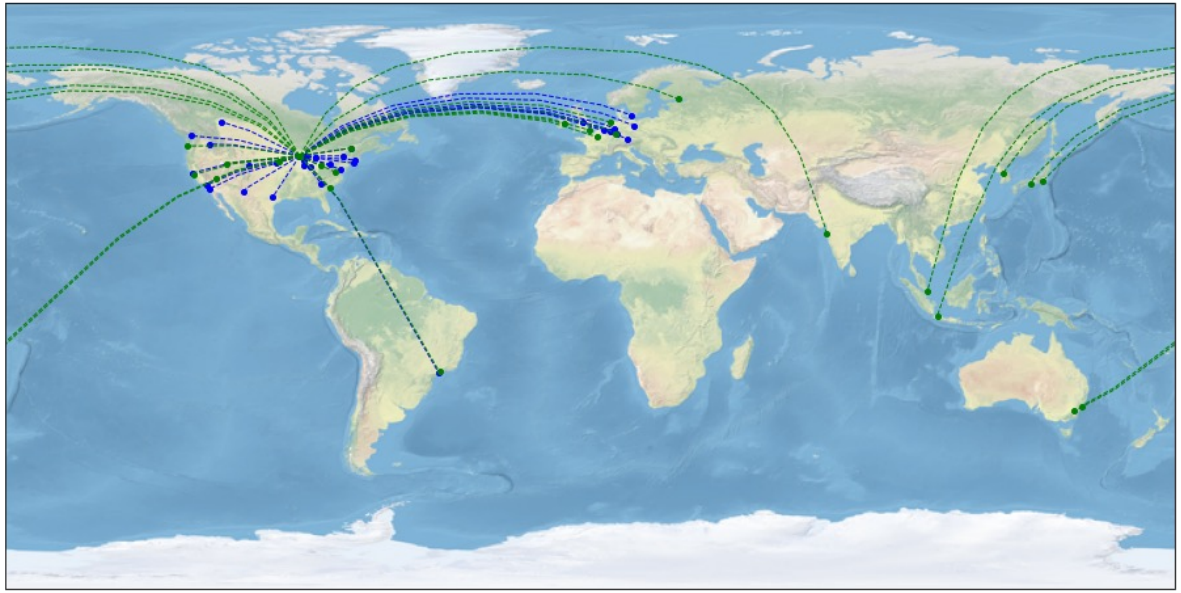
- Research Scientist at WIPAC
- IceCube M&O L2 Lead for Computing and Data Systems
- IceCube Coordination Committee Lead
- Heterogeneus Systems Lead for Accelerated AI Algorithms for Data-Driven Discovery
- Member of the Open Science Grid Council
- Active in IceCube 2011-2016; 2018-Present
- IceCube Ph.D. 2014

Outline

- Bird's Eye View
- Status of WBS 2.3 Performance Metrics
- Details for WBS 2.3
- PY1-3 Developments and Projects
- Additional Funding

IceCube Computing – Bird's Eye View

- Global heterogeneous resources pool
- Mostly shared and opportunistic resources
  - Collaboration institution resources, e.g. university HPC facilities
  - US: ACCESS, PATh, NSF Leadership class HPC
  - EU: EGI
- US- and collaboration-funded dedicated computing resources are centrally hosted at and operated out of UW-Madison and UMD
- Archival
  - Raw Data: At NERSC*, adding TACC**
  - Processed Data: DESY-Zeuthen
- Computing is continually evolving driven by changes in science goals (e.g. Multi-Messenger Astrophysics, Upgrade), analysis techniques (e.g. Artificial Intelligence), technology (e.g. Cloud Computing), and policy (e.g. NSPM-33)



* National Energy Research Scientific Computing Center (NERSC) – DOE-funded, in-kind
** Texas Advanced Computing Center

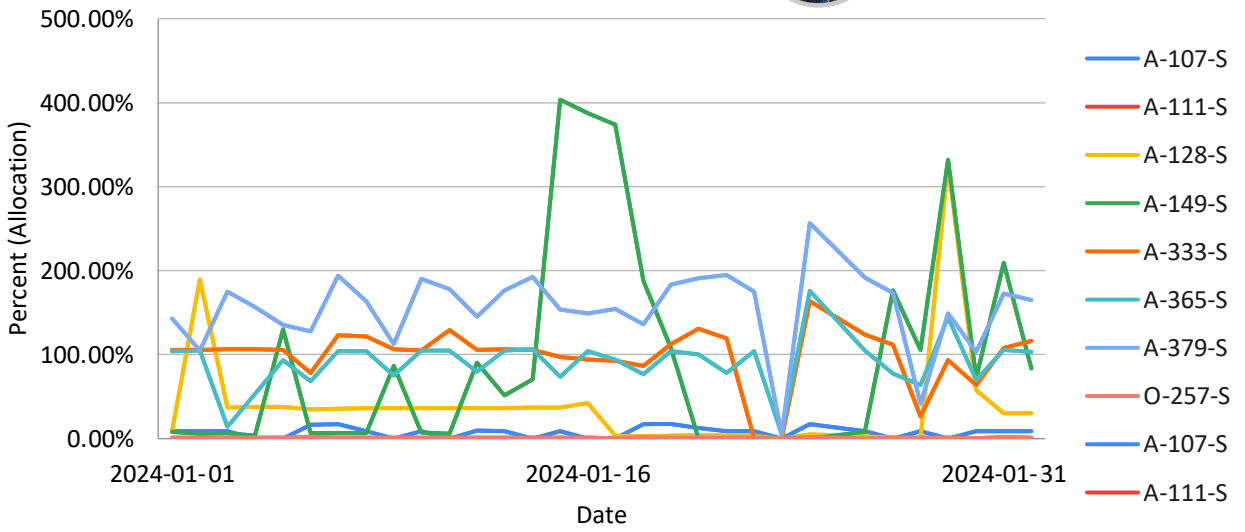
M&O WBS 2.3 Performance Metrics

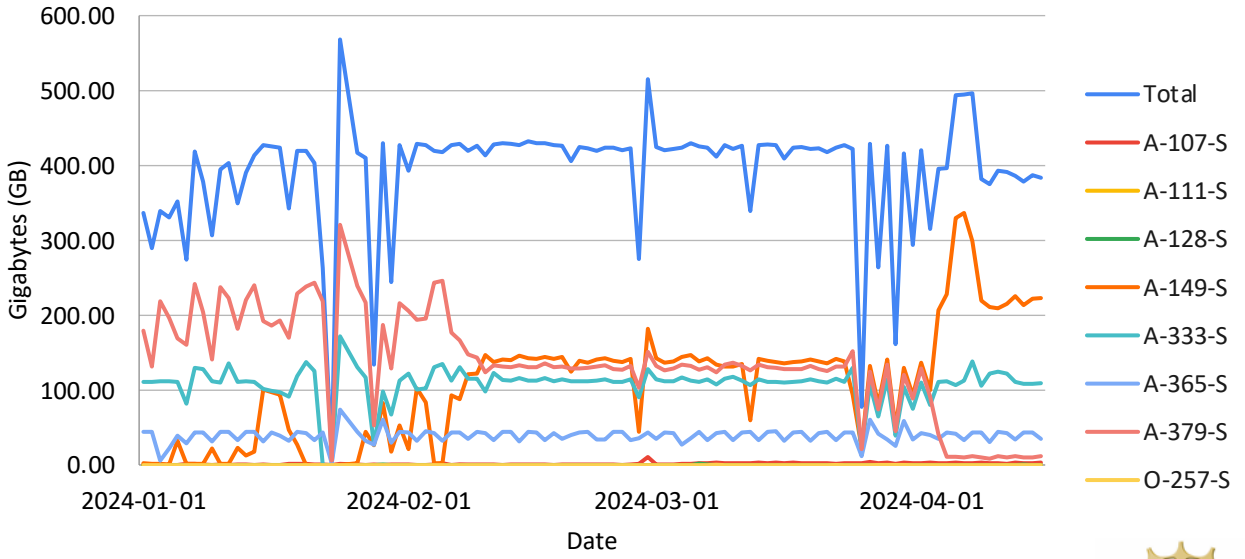
Milestone/Metric	Target	Status
Core Infrastructure Uptime	>=95%	PY1-3: ✅
Non-Core Infrastructure Uptime	>=90%	PY1-3: ✅
Data Transfer Delay	< 2 Days	PY1-3: ✅, PY3: 🚧/✅
Replication of Processed Data	<= 7 Days	PY1-2: ✅, PY3: 🚧
Replication of Raw Data	<= 90 Days	PY1-3: ✅
IceProd Uptime	>= 90%	PY1-3: ✅
Non-production IceProd users	20 Users	PY1-3: 🚧

- Data Transfer Delay: Changes in processing increasing satellite demands during austral summer reaching above our allocation; satellite bandwidth management a subject of discussion with USAP and other experiments at South Pole
- Replication of Processed Data: Retirement of staff member causing delay in online processed/satellite data replication in PY3, offline processed data is within metric in PY3
- Non-production IceProd users: Working on user interface improvements, workshops, etc. PY1: 3 Users, PY3: 10 Users

Data Transfer from Pole

- PY3: Overhauled processing at Pole
  - Increase data rate
  - Less biased sample
  - In preparation for Upgrade
- Satellite transmission delay during high-rate portion of year in first season of new filter
  - Requested increase was not granted previously
  - In discussion with USAP and other experiments to balance satellite needs and ensure fair allocation



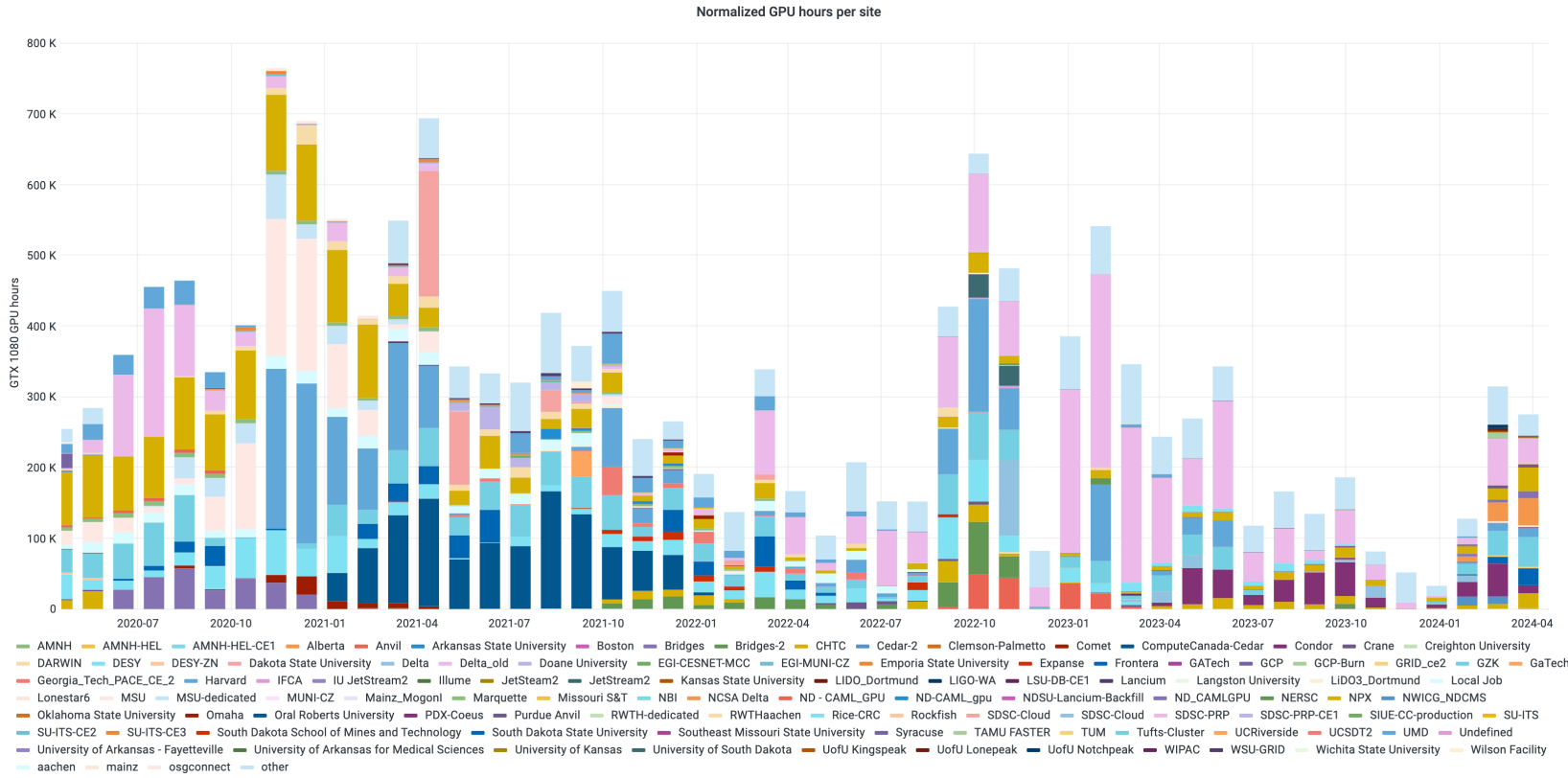


Dedicated CI – Compute

- Total dedicated:
  - UW: 4000 CPU cores, 300 GPUs
  - UMD: 300 CPU cores, 90 GPUs
- Significant deferred maintenance – Most hardware is >7 years old (a good portion is >10 years)
  - Average hardware lifetime is 3-6 years
  - Hardware refresh and/or enhanced support for external resources potential solutions
- Expecting dedicated resources to shrink in size in short-term
  - Affecting mostly researchers, external resources have a steep learning curve
  - Balancing cost of storage vs. compute – Storage wins
  - Consumer GPUs no longer available for servers – Need to invest in more expensive “workstation” GPUs
- Significant challenges ahead:
  - GPU competition between simulation production and AI training and inference
  - AI:
    - How much AI will we see?
    - How complex will AI models be?
    - How do we support AI training in the future?
    - High-End Hardware is very costly ($250-500k per server) and in high demand
  - Upgrade software resource needs, i.e. growth in RAM per core requirements, are an unknown – Working to clarify needs

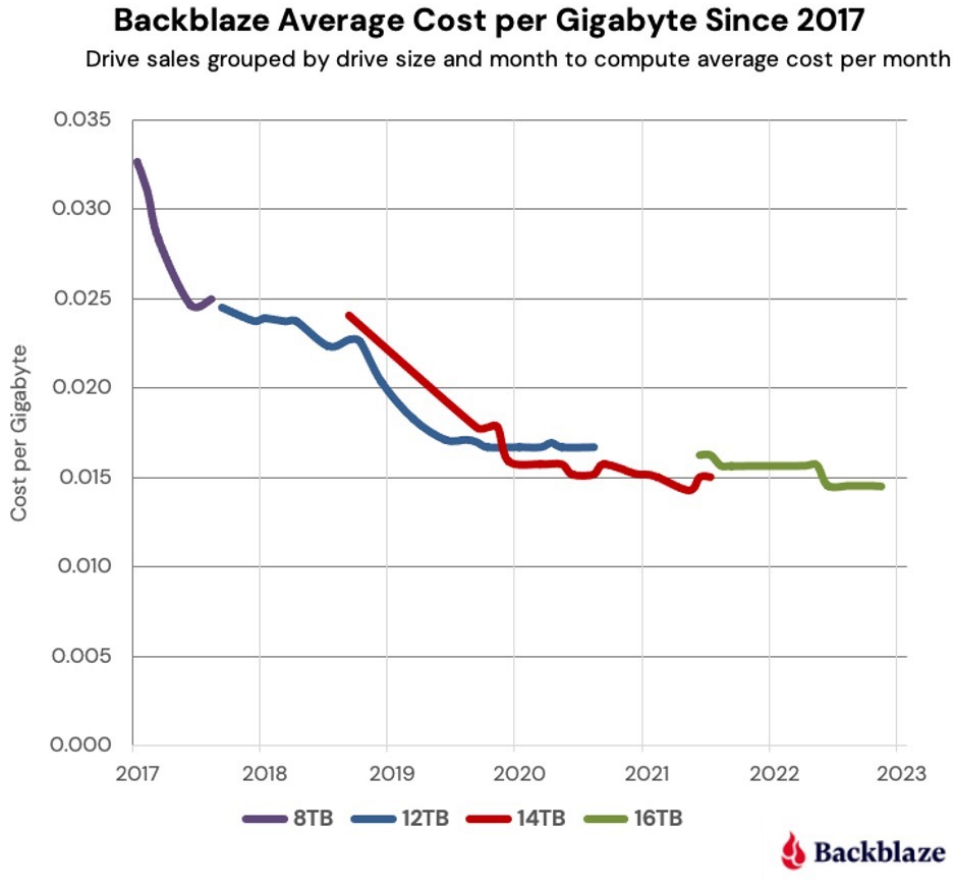
Contributed CI – Compute

- > 250 Sites used
- US:
  - ACCESS – NSF HPC resources
  - PATh – NSF HTC resources
  - National Research Platform (NRP)
  - NERSC
- EU
  - DESY
  - EGI
- Canada
  - Digital Alliance
- These resources are more plentiful, but relatively steep learning curve



Dedicated CI – Storage

- PY1-3 replacement of all distributed storage infrastructure, i.e. simulation, data, user, and analysis storage
  - Increased from ~10 to ~14 PB
  - All distributed storage converted to Ceph for increased reliability and resilience
  - Initial transition to Ceph created a two-week downtime because of issues with transition and lack of experience with new software and documentation
  - Second transition went without issues
- Long-term focus will be as data provider
  - Analysis will shift towards external resources
  - Cost of Storage is flattening
  - Data volume only increases
- Upgrade:
  - Depends heavily on data rate, current estimate is a 30% increase. Considered in most recent replacement
  - Reducing on disk data and simulation with new processing – Impact to be assessed



VM Farm

- Hosts support services, websites, etc. – Central to our operations
- Recently deployed new hardware after delays
- Delayed hardware upgrades to
  - Investigate alternatives to reduce cost
    - Hosting within UW-Madison's VM farm
    - Hosting storage within UW-Madison's VM farm
    - Exploring different VM hypervisor due to increase in prices of current one
    - Getting the most out of the current hardware
  - Work on other projects:
    - User management
    - Alternative virtualization infrastructure (Kubernetes)
    - Google Workspace
- Multiple O(hours) downtime due to hardware instability led us to replace hardware with an eye on future software replacement

User Management

- PY1 and 2 we replaced our user management system to be more self-service
  - Increased self-service – Reduced need for human support
  - More automation
  - Additional cybersecurity features
- Still working out “business” rules
  - What email lists are a “controlled” environment?
  - When does a user become emeritus/retired?
  - How to automate certain changes in personnel/leadership?

Google Workspace

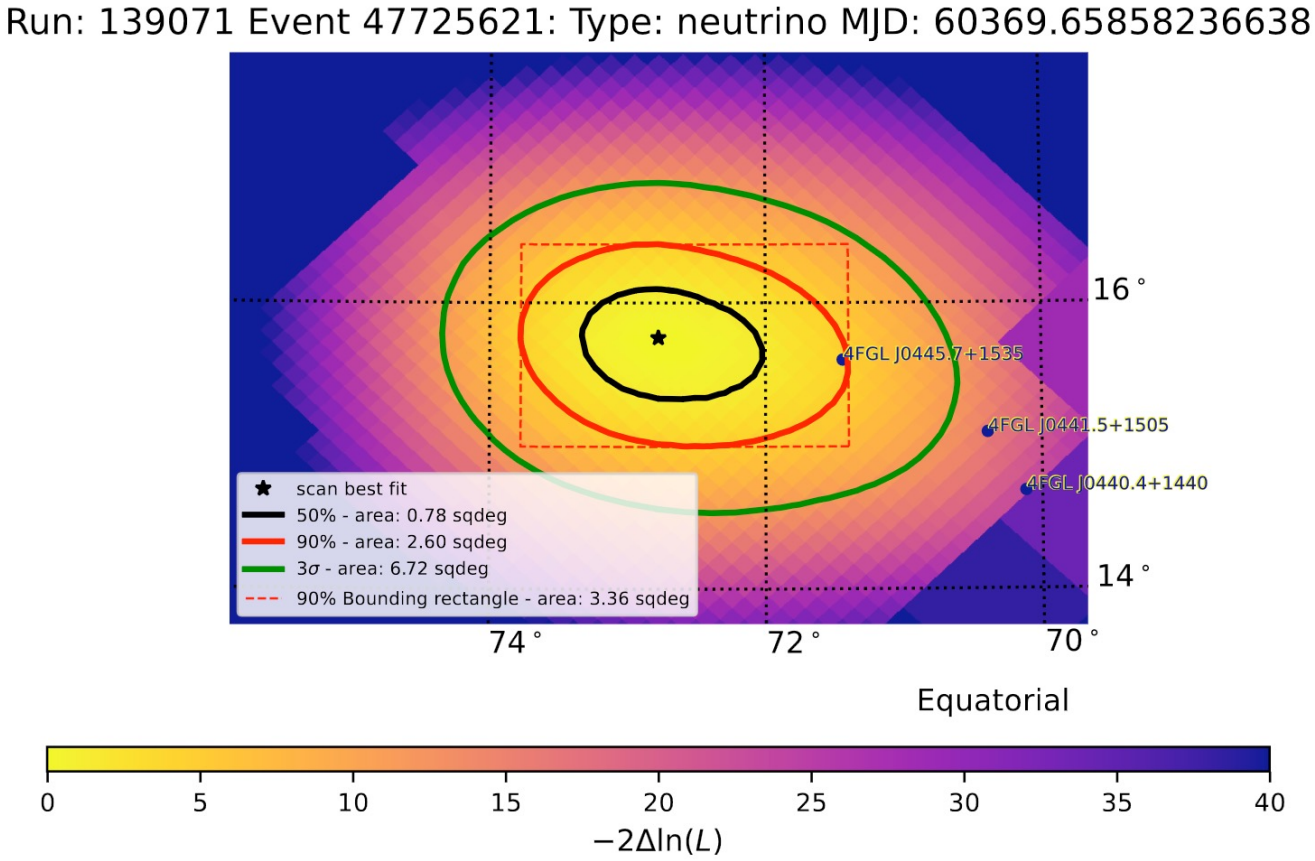
- Moving self-hosted “business” (email, doc. management, email lists, etc.) services to Google Workspace
- Why transition?
  - Self-hosted infrastructure needed an upgrade on both software and hardware side
  - Self-hosting has it advantages, but uses resources that can be used better
  - Much higher uptime guarantees and better cybersecurity than self-hosted
  - The various services don’t talk well to each other or are getting long in the tooth (DocuShare)
  - Cloud-based services are “cheaper” (in this case), provide integration between features, and modern features (search, calendaring)

Cybersecurity

- Significant changes are in the works
- Working on TrustedCI Framework implementation
- Multi-Factor Authentication
  - US Federal guidelines, UW cybersecurity, and general best practices are pushing us towards Multi-Factor Authentication
  - Exploring options within User Management Software (KeyCloak) or from UW (Duo)
- HTTP(S)-based File Transfer
  - Moving to modern JWT auth
  - Long-term industry support
- New guidance on the horizon (NSPM-33, updated RIG) – Waiting on specifics

SkyDriver – Reconstruction-a-a-S

- Replacement for Skymap Scanner to determine direction of events from Multi-Messenger Astrophysics Alerts
- Shifted resource pool to contributed rather than dedicated compute resources
  - Significantly larger resource pool
  - Less interruptions to analyzers
- Cloud computing a potential to reduce time to result further – Associated cost not budgeted for
- Development partially funded through a separate grant – NSF #2103963



ML Inference-as-a-Service

- We are expecting the use of ML/AI-based tools to increase, some examples:
  - Event Topology Classifier
  - DNNCascade selection
- “90% of an AI model’s lifetime is spent in inference”
- Competition between simulation and data analysis for GPUs
- 10x improvements on GPUs vs. CPU, but how well do we use the GPU?
- Other experiments are looking at dedicated GPUs for their ML/AI inference tasks
- ML Inference-as-a-Service to improve GPU utilization and availability
- Funded as part of Accelerated AI Algorithms for Data-Driven Discovery (A3D3) – NSF #2117997

Pass 3

- New filtering in development
  - Online/Pole filtering is in production – Includes new calibration
  - Offline/Northern filtering is under development
- Saw a significant improvement in analysis speed with “common” processing across different years of data
- Next data reprocessing (Pass 3) is in development
- Testing Pass 3 on NSF leadership-class HPC (TACC’s Frontera) as part of the NSF #2139536
  - Next NSF leadership-class HPC system (Vista and Horizon) will be ARM64
  - IceCube codes run on ARM64

Cloud Computing

- Since 2019 have been doing extensive and successful proof-of-concept using cloud resources in collaboration with SDSC
  - 2020: HPCWire Editors' Choice Awards for Best Use of HPC in the Cloud (Use Case)
  - 2022: HPCWire Readers' Choice Awards for Best Use of HPC in the Cloud (Use Case)
- Cloud resources are too expensive compared to self-hosting resources
  - Steady state usage: ~$8M/year
  - GPU and network cost are cost driver

Summary

- Leveraging non-M&O funding to help develop new projects and initiatives
- IceCube M&O computing shifting increasingly to external resources
- Providing storage and archival of IceCube data main priority
- Modernization of key systems is labor limited
- IceCube M&O computing adapting to a rapidly evolving environment