# New Institutional Resources at BNL

Tony Wong

(HEPIX Spring 2021)

BROOKHAVEN NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY
BROOKHAVEN SCIENCE ASSOCIATES

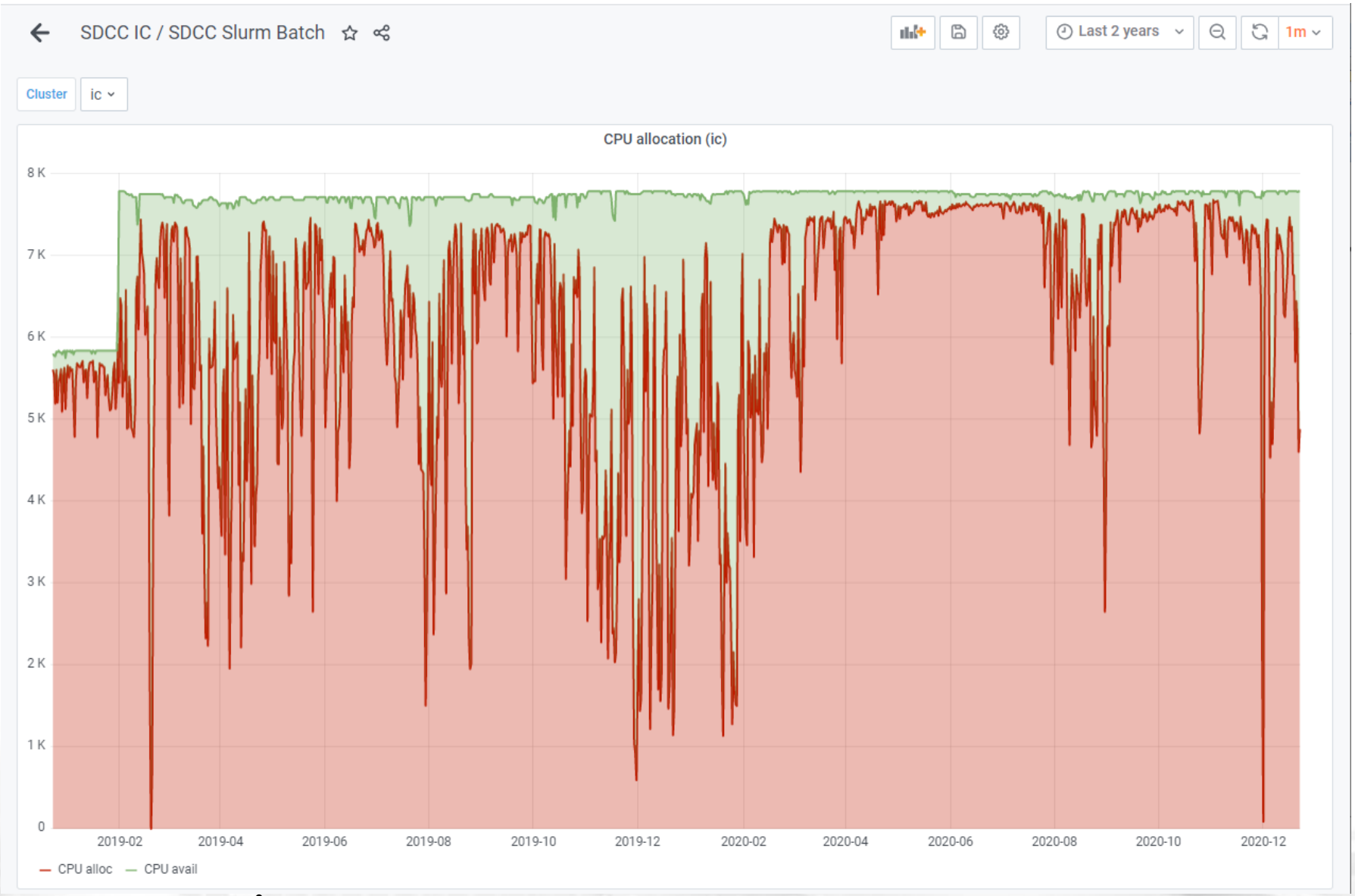
# Background

- Computational Science Initiative (CSI) was formed in 2015 to consolidate and optimize BNL computing  infrastructure
- The Institutional Cluster (IC) is a Lab-wide HPC resource
  - Operated by the Scientific Data & Computing Center (SDCC)
  - regulated by MOU's between CSI and stakeholders
  - IC is meant as a ramp to Exascale computing systems at Leadership Class Facilities (LCF) such as NERSC, Argonne LCF, etc
- IC arrived in 4 batches. First one (50% of cluster) was in Fall 2016
- A 1 PB high-bandwidth (up to 24 GB/s) GPFS storage system was connected to IC
- Augmented in subsequent years with KNL and SL (Skylake) clusters
- New data center available to host new institutional resources beginning in Summer 2021
  - See Alex Zaytsev presentation later this week

# User Communities

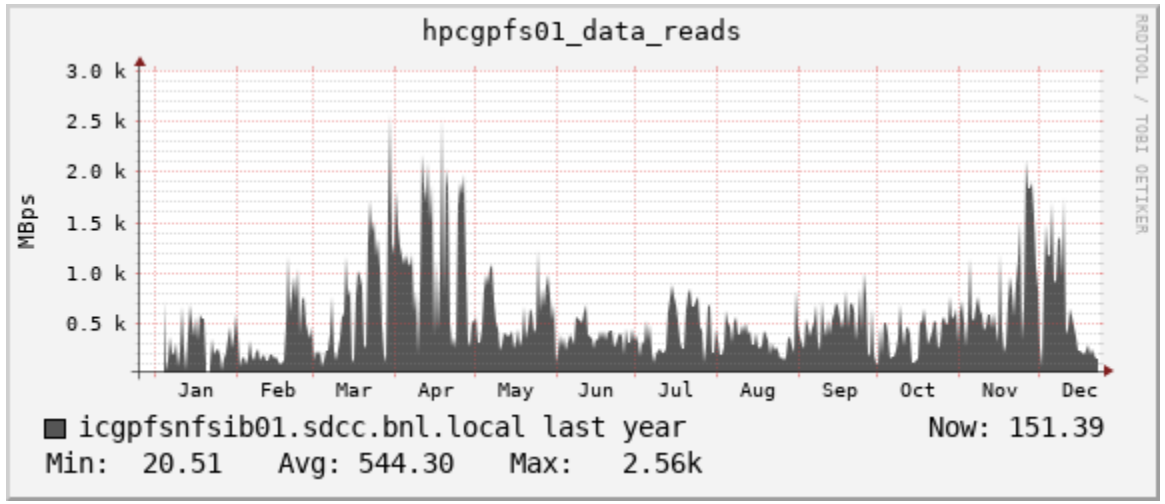
- Approximately 500 users since inception (Fall of 2016) of support for HPC activities
- Diverse user background
  - LQCD (U.S. based community)
  - Center for Functional Nanomaterials
  - Condensed Matter Physics and Material Sciences
  - US ATLAS
  - Computational Science

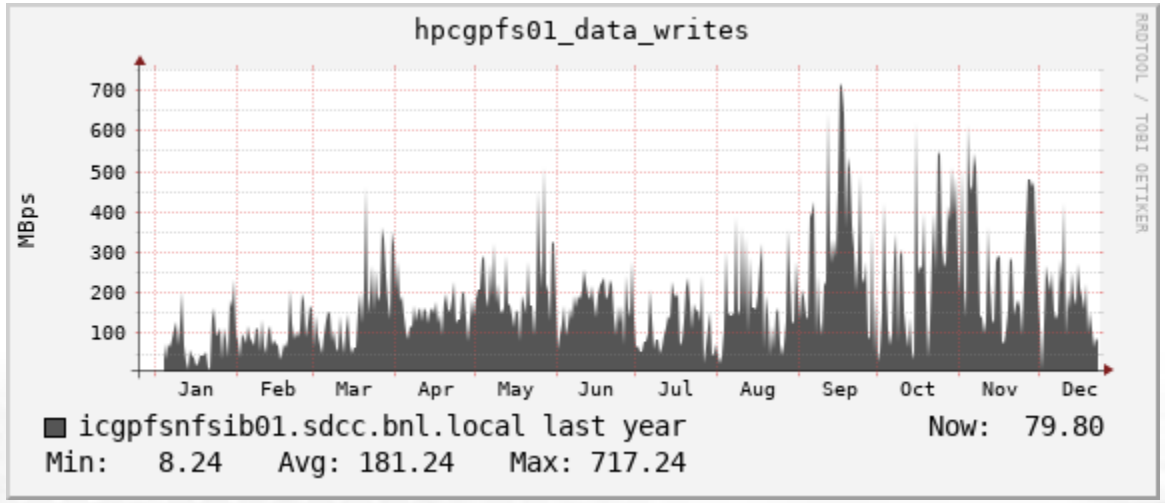
# Recent IC Utilization




- Optimizations increased usage to ~95% for past year

# GPFS Storage



# HPC Resources

- Current resources includes 3 distinct clusters
  - Institutional cluster (216 nodes with dual Intel Xeon cpu's + gpu's and 256 GB RAM) – 810 TFlops
  - KNL cluster (142 nodes with single Intel Xeon Phi cpu and 192 GB RAM) – 200 TFlops
  - Skylake cluster (64 nodes with dual Intel Xeon cpu's and 192 GB RAM) – 130 TFlops
  - All interconnected with Infiniband EDR and accessible via Slurm for batch management
  - Further details available here https://www.sdcc.bnl.gov/resources/hpc
- A large portion of the current institutional resources is ~5 years old and reaching operational end-of-life
- In January 2021, the SDCC began a series of meeting with current and potential future stakeholders to discuss the next generation of institutional resources

# General Strategy

- Start procurement in late FY21 for installation in new data center in 2022
- Comparable capabilities to existing IC
- SDCC will provide technology overview and solicit user feedback
- SDCC will contact potential suppliers and set up testbeds
- Stakeholders will be asked to assist with benchmarking during evaluation process
- Expect evaluation to be an iterative process over several months
- Series of meetings to discuss evaluation findings and narrow technology choices

# Other Considerations

- Operate existing IC until FY23 when data center migration is completed—availability overlap with ramp up of new resources
- Open to operating a heterogeneous cluster, if that's the optimal solution for user applications—stakeholders will steer architectures to be deployed
- Existing GPFS storage can operate up to an aggregate I/O rate of 24 GB/s—in actuality, peak usage ~3 GB/s (~11 GB/s on shorter 1-2h spurts)
  - Plan to replace with a cost-effective solution to better match users' needs
- Will address updates to MOU's after procurement plans are understood

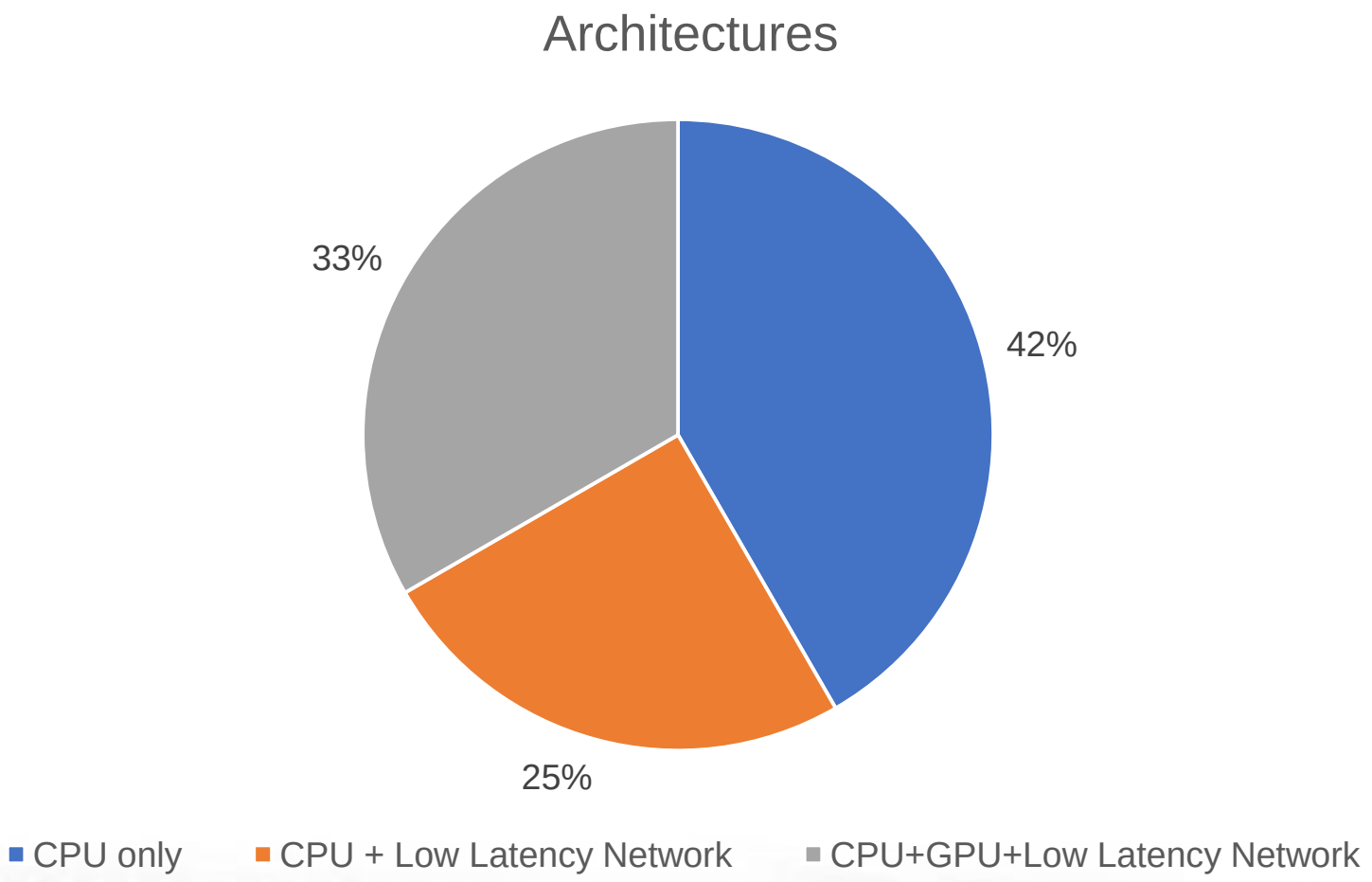
# (Approximate) Timeline

- Initial meeting (Jan 2021)
- Set up test beds (Feb-Jun)
- Evaluation (Mar-Aug)
- Decision (Sep)
- Procurement (Sep-Nov)
- Availability (Dec-Jan 2022)

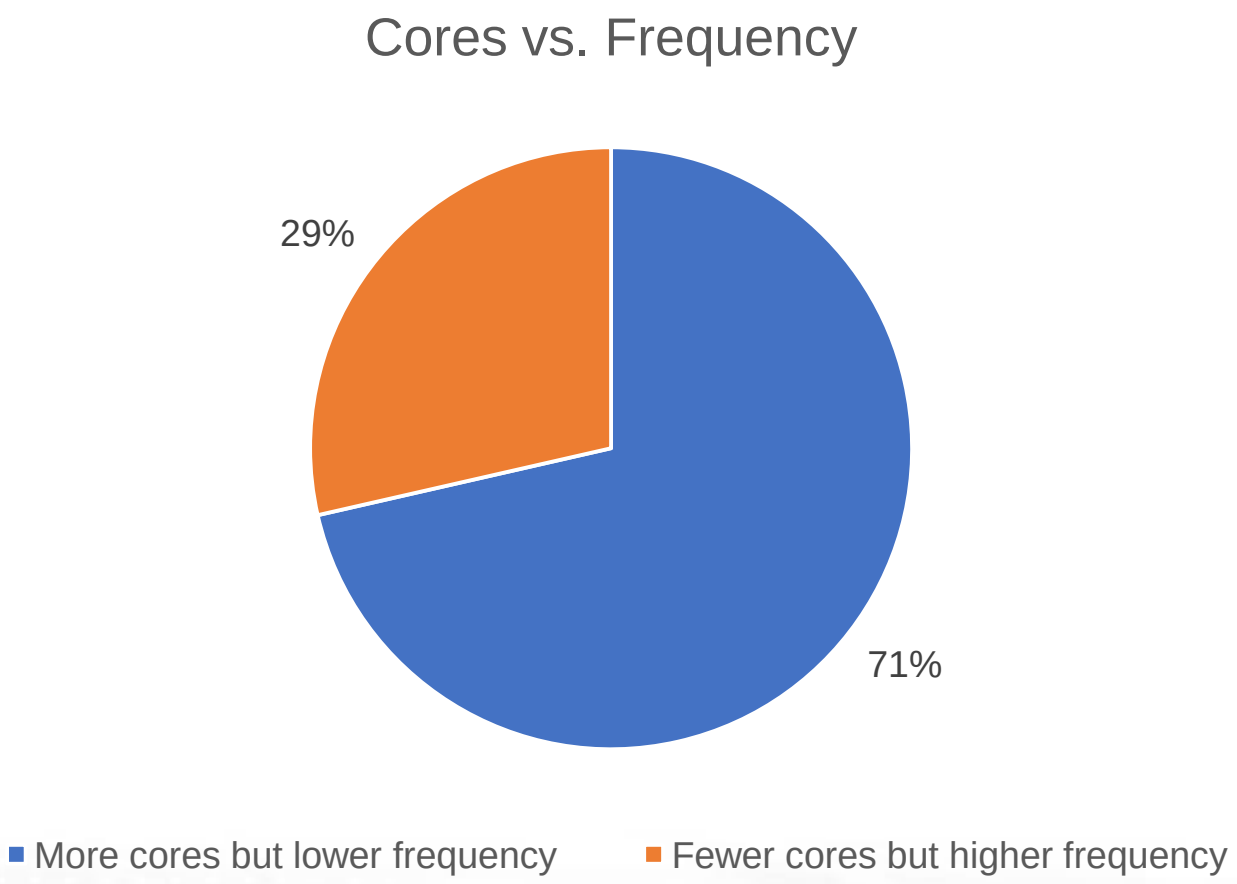
# First Step—User Feedback

- Brief questionnaire sent to stakeholders to survey:
  - Architecture(s) to be evaluated
  - Estimated needs (core or node-hours, TB of storage, storage modes, etc) and timeline
  - Availability to assist in evaluation process
- Obtained responses from 7 communities (new communities in green)
  - USQCD (US Quantum Chromo Dynamics) collaboration
  - Neutrino (DUNE) physics community
  - Environmental and Climate Sciences
  - CFN (Center for Functional Nanomaterials)
  - Collider Accelerator Division
  - CSI (Computational Science Initiative)
  - ATLAS

# Computing Architecture



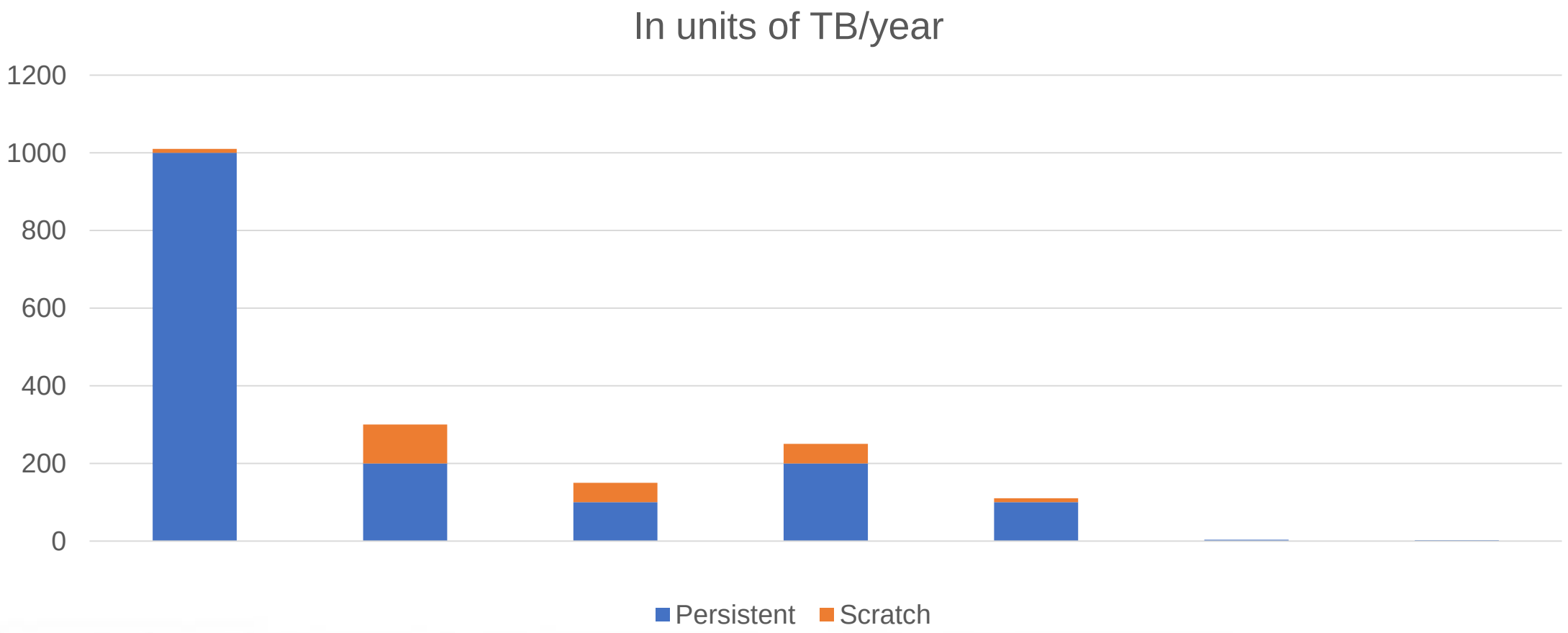
# Computing Architecture



# Computing Needs

- HPC-like resources
  - Estimates vary from 1,000 core-hours to 1 million core-hours per year
- HTC-like resources
  - HEP collaborations needs up to 30k core-years per year
- Modest disk storage requirements
  - 4 TB/node
  - HDD okay for disk I/O needs
- Wide range in memory requirements
  - From 2 to 256 GB/core
  - 32 GB per GPU (standard for V100 gpu's)

# Disk Storage Needs



# Other Survey Comments

- Interest in additional services (BNLBox, Git, web, email, chat)
  - Already available to HEPN user communities
  - Encourage other communities to make use of SDCC services
- Hardware accelerators
  - Nvidia is the default choice
  - Willingness to try other suppliers (Intel and AMD)
  - Desire to explore other solutions (ie, FPGA)
- Interest on access to GPU resources for AI/ML usage

# Next Steps

- Set up testbeds
  - 2 x Intel Cascade Lake Gold, 192 GB RAM, 8 TB SSD's and 2 x 1 Gb connectivity (available now)
  - 2 x Intel Ice Lake, 256 GB RAM, Infiniband HRD connectivity (up to 8 nodes available in early April)
  - Gpu-based system (availability TBD)
  - Other testbeds TBD
- Organize evaluation
  - Staff
  - Identified community "volunteers" in survey
- Upcoming meetings
  - Gauge progress on evaluation of testbeds
  - Experience with non-BNL recent HPC-like cluster acquisitions as a sanity check