Core Facility SCALE

Eric Yen and Felix Lee

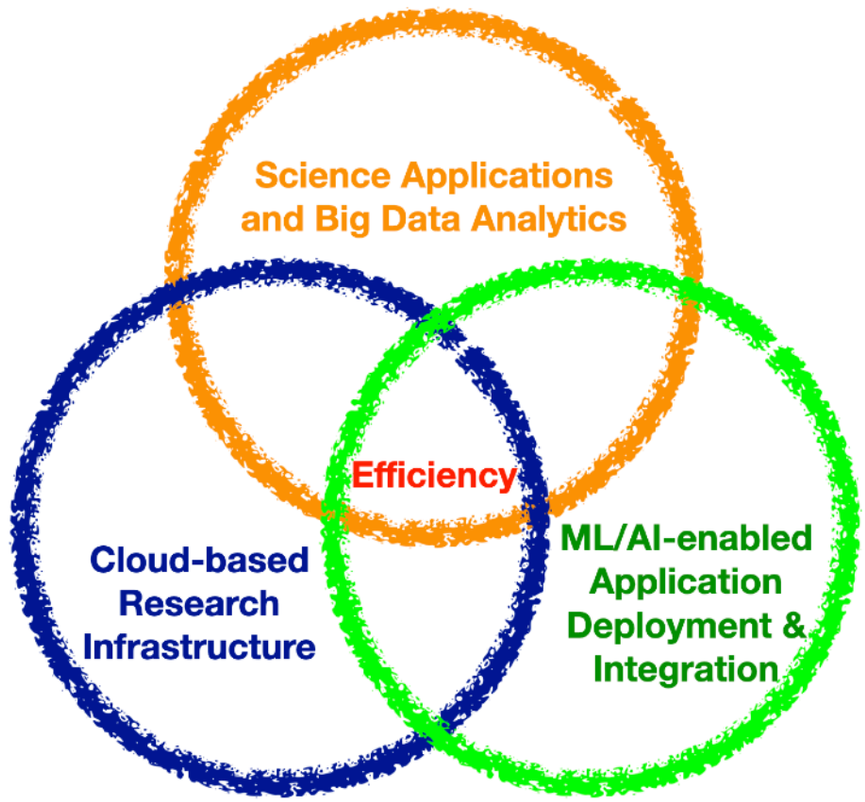
Academia Sinica Grid Computing Centre (ASGC)
Taiwan

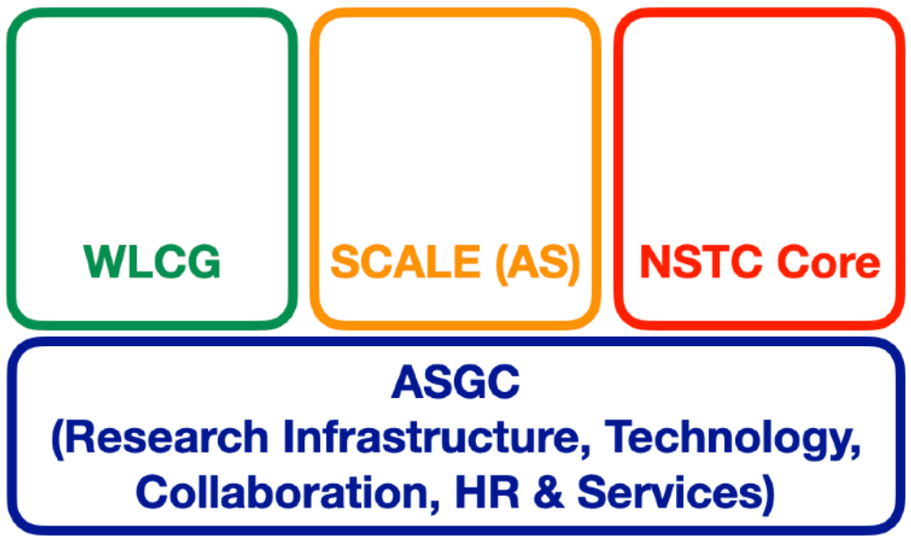
SCALE User Committee Meeting

23 Oct. 2023

ASGC Is Accelerating Discovery and Innovation

- 大數據分析與科學計算核心設施
  Scientific Computing and big data AnaLytics corE facility (SCALE)
- Schedule: Jan 2023 - Dec 2024
- Budget: 3.5M annually
  - All are functional expense (HR primarily)
  - Usage fee: > 6M
    - (2023 est.) Coll.HW: 3.7M, Usage: 3.2M
- Objectives
  - Upgrade of AS Research Infrastructure for Computing
- ASGC becomes a scientific computing core facility for Taiwan supported by NSTC from June 2023 (3-yr term)

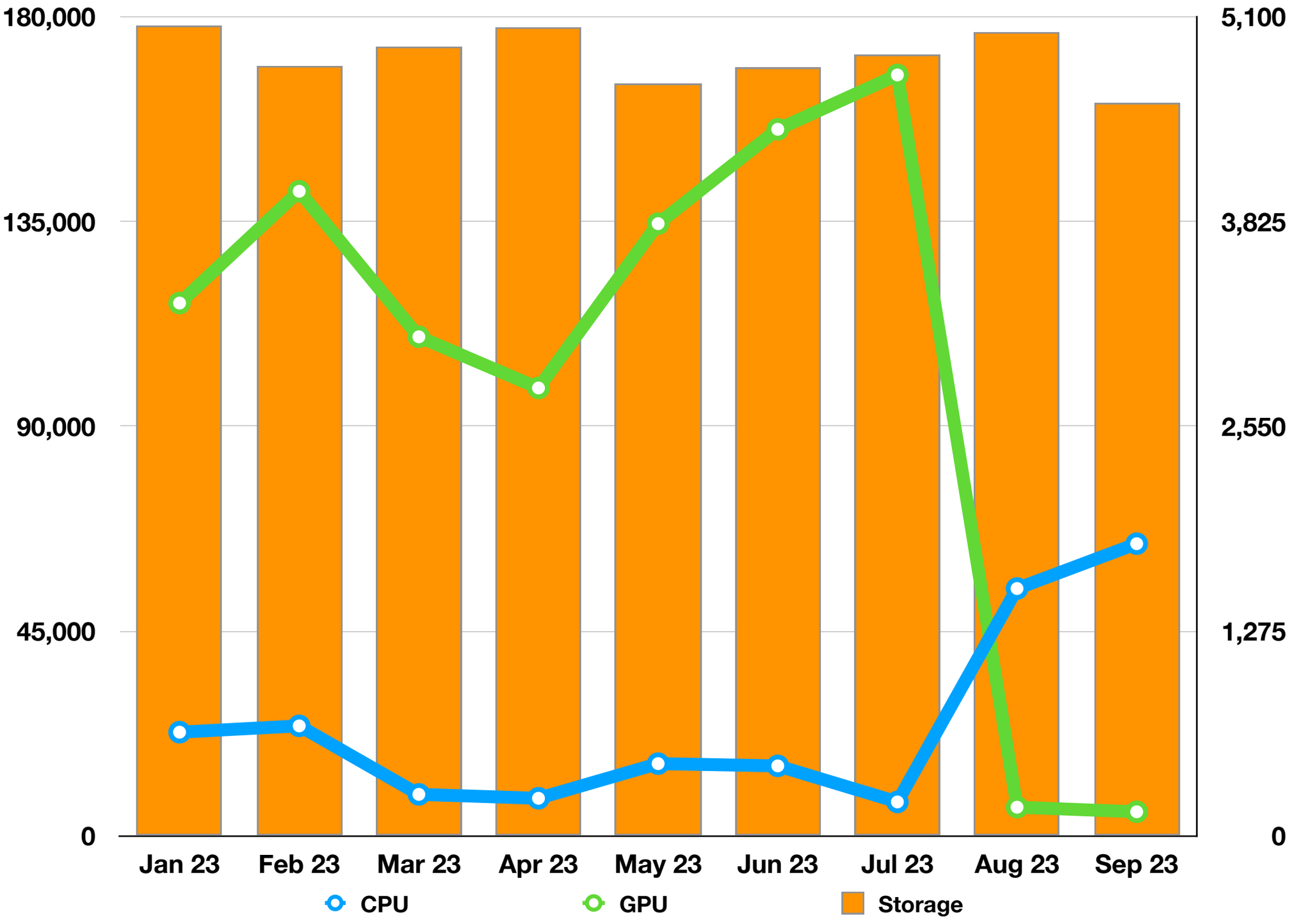




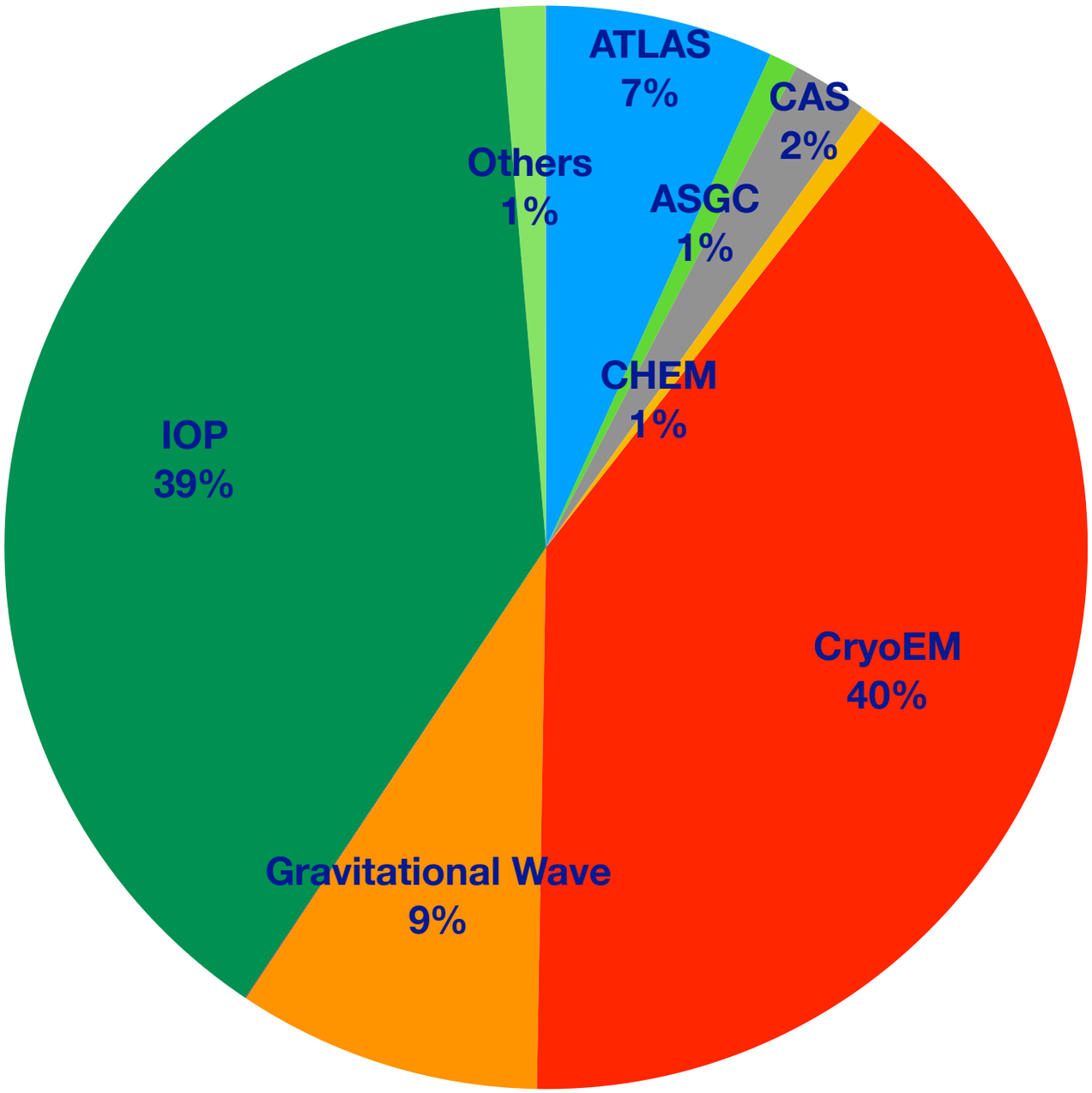
Enabling Innovations by Integrated Research Infrastructure - Connecting Instruments, Data, Minds, and Computing

- **ASGC is the primary computing arms in AS - by cloud-based research infrastructure**
  - **Integrating experiment/instruments and analysis facility**
  - **Batch and interactive job submission**
  - **Optimization of Data analysis pipeline and system efficiency**
  - **Collaborations: ATLAS, CMS, AMS, KAGRA, ICECube, Proton Therapy, CryoEM/Synchrotron Source, Astronomy, Condense Matter, Lattice QCD, NGS, Bioinformatics, Earth Science, Environmental Changes, etc.**
- **Resources: 20,090 CPU Cores; 236 GPU Cards; 30 PB Disk Storage**
- **Leverage the WLCG core technology and develop capacity to support broader scientific applications**
- **24/7/365 services since 2006**
  - **Data Center availability: 99%+**
  - **Scientific Computing Service reliability: 97%+**
  - **Daily average power consumption:10,326 KWH (2023), >20% reduction than 2022**
  - **Power saving efficiency: ~ 20% (cluster-based)**
  - **International Data Transmission (Inbound + Outbound, WLCG): > 21PB (2022)**
  - **Inside Data Center Traffic (Inbound + Outbound) > 1PB daily**
- **Reliability and Performance are the key objectives**
  - **User Scale：(#Groups, #Users) = (90, 350)**
  - **Finished #Jobs (2023 estimated): > 5,000,000 (40% for WLCG)**
  - **Supported research publications: >15 (15 in 2022, not including ATLAS & CMS)**
  - **Training and workshop: 5 events a year**

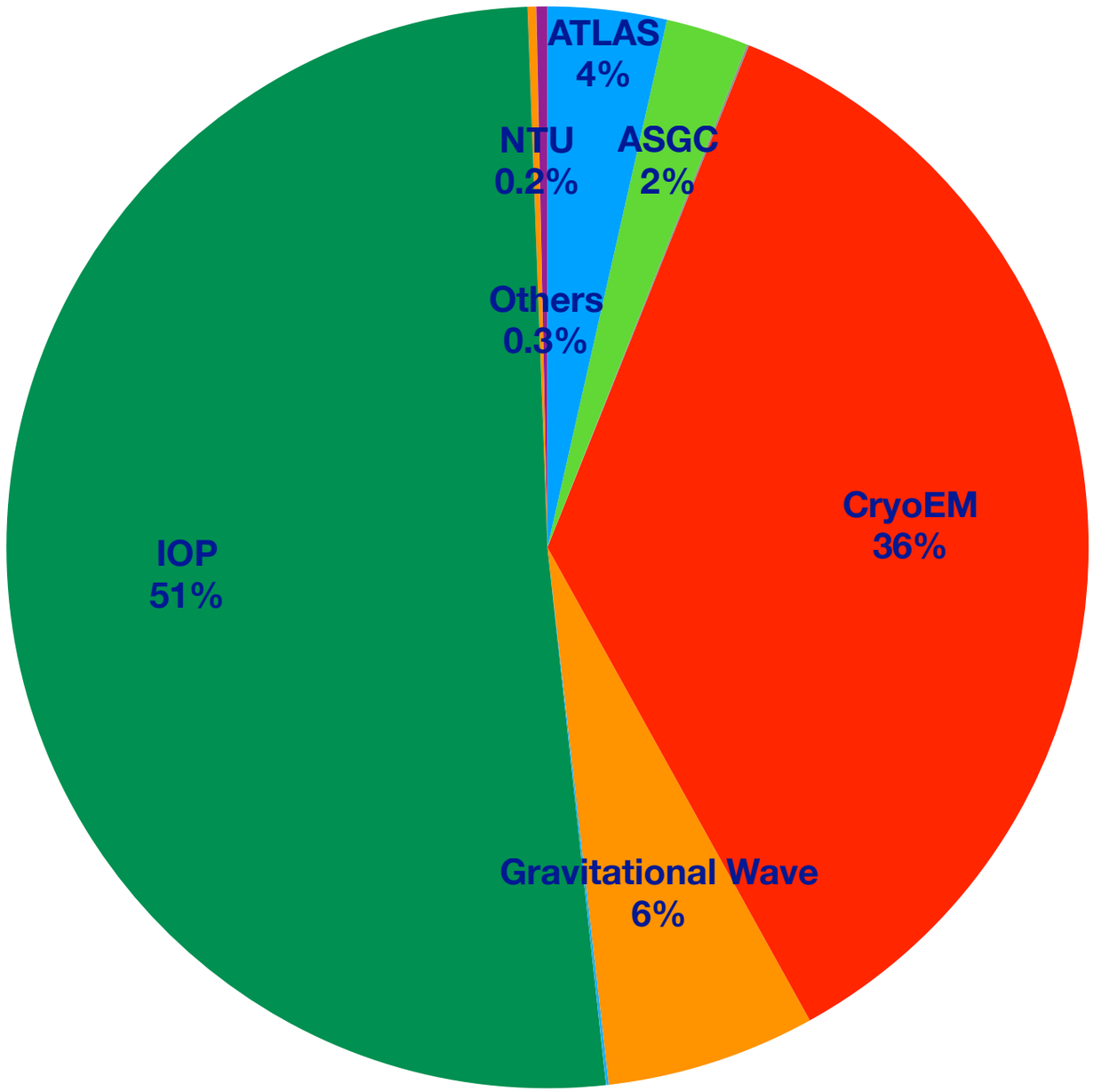
Resource & Services



ASGC Computing Resource Utilization in 2022 & 2023 (till June 2023)

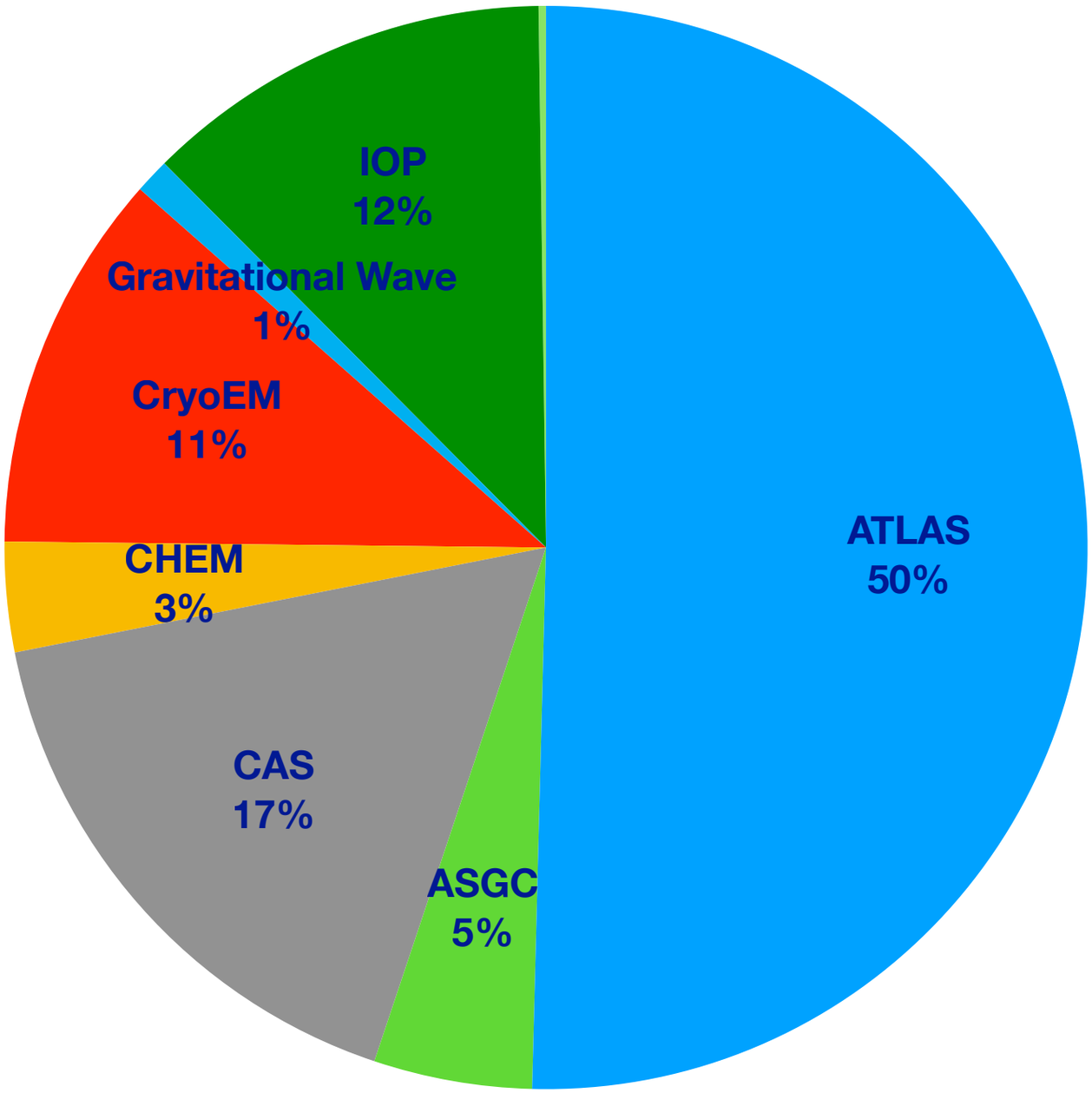


Monthly Average Computing Resource Usage of 2022
Total: 167,475 SRU

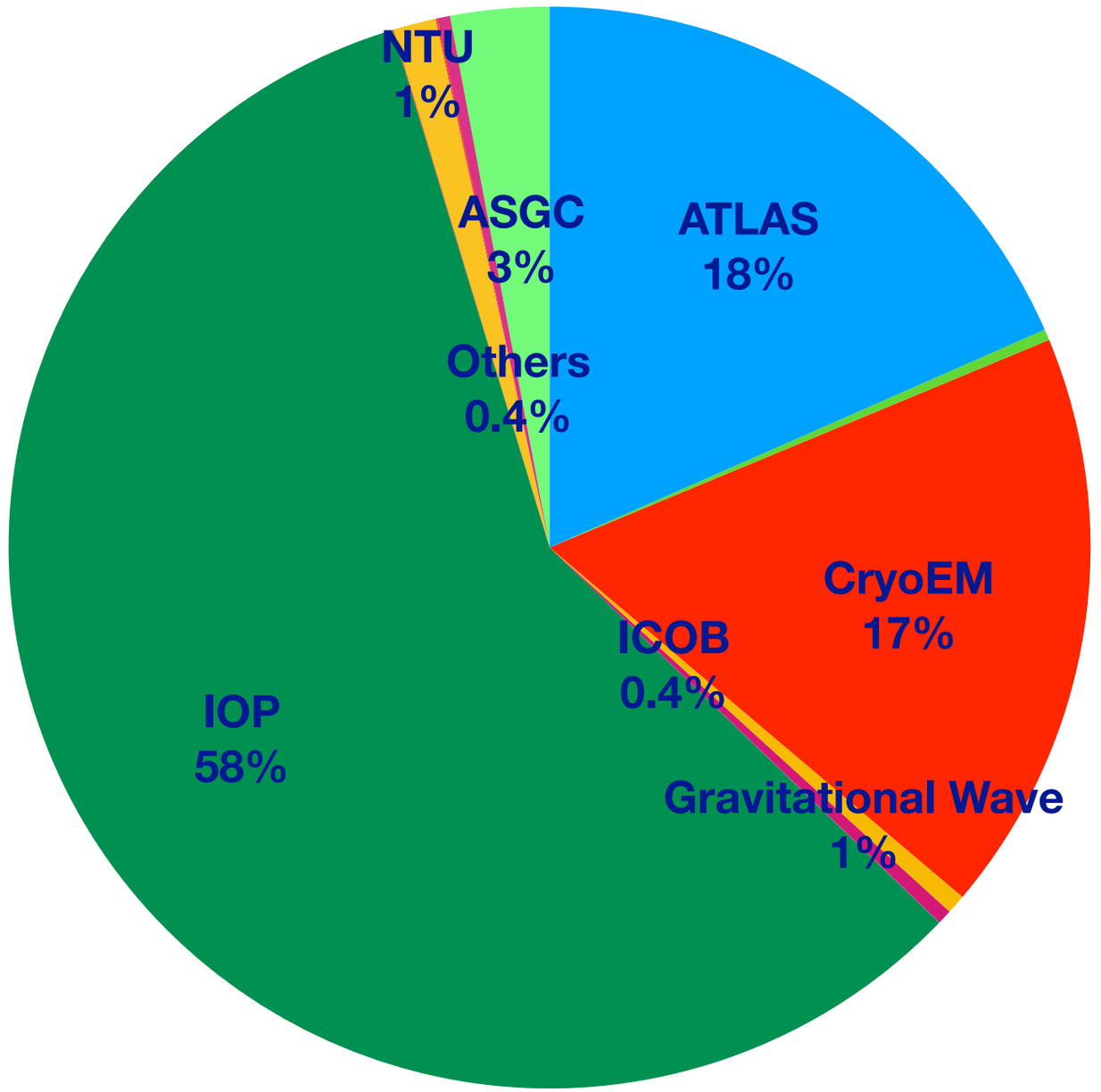


Monthly Average Computing Resource Usage of 2023
Total: 128,647 SRU

ASGC CPU Resource Utilization in 2022 & 2023 (till Sep. 2023)

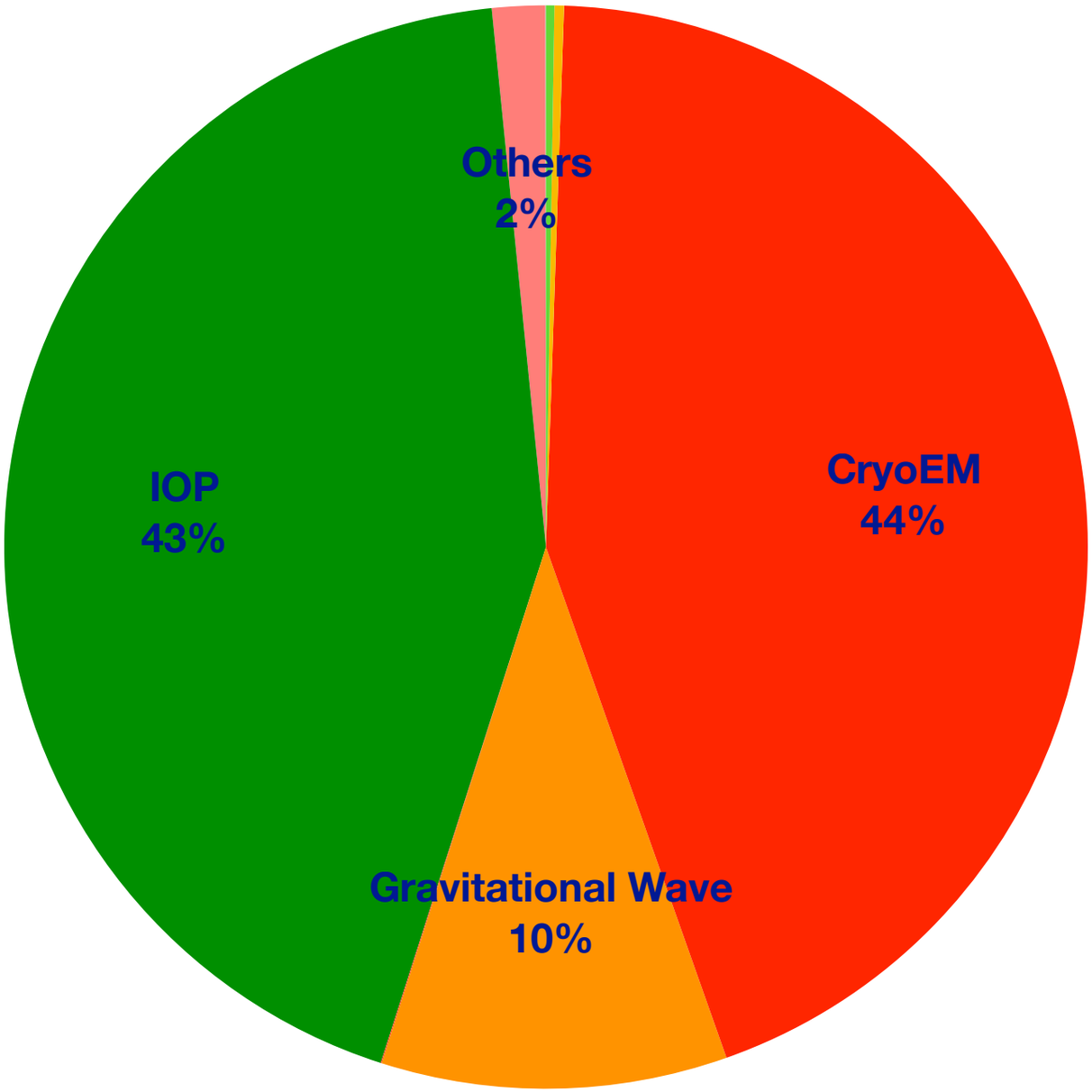


Monthly Average Computing Resource Usage of 2022: 22,592 SRU

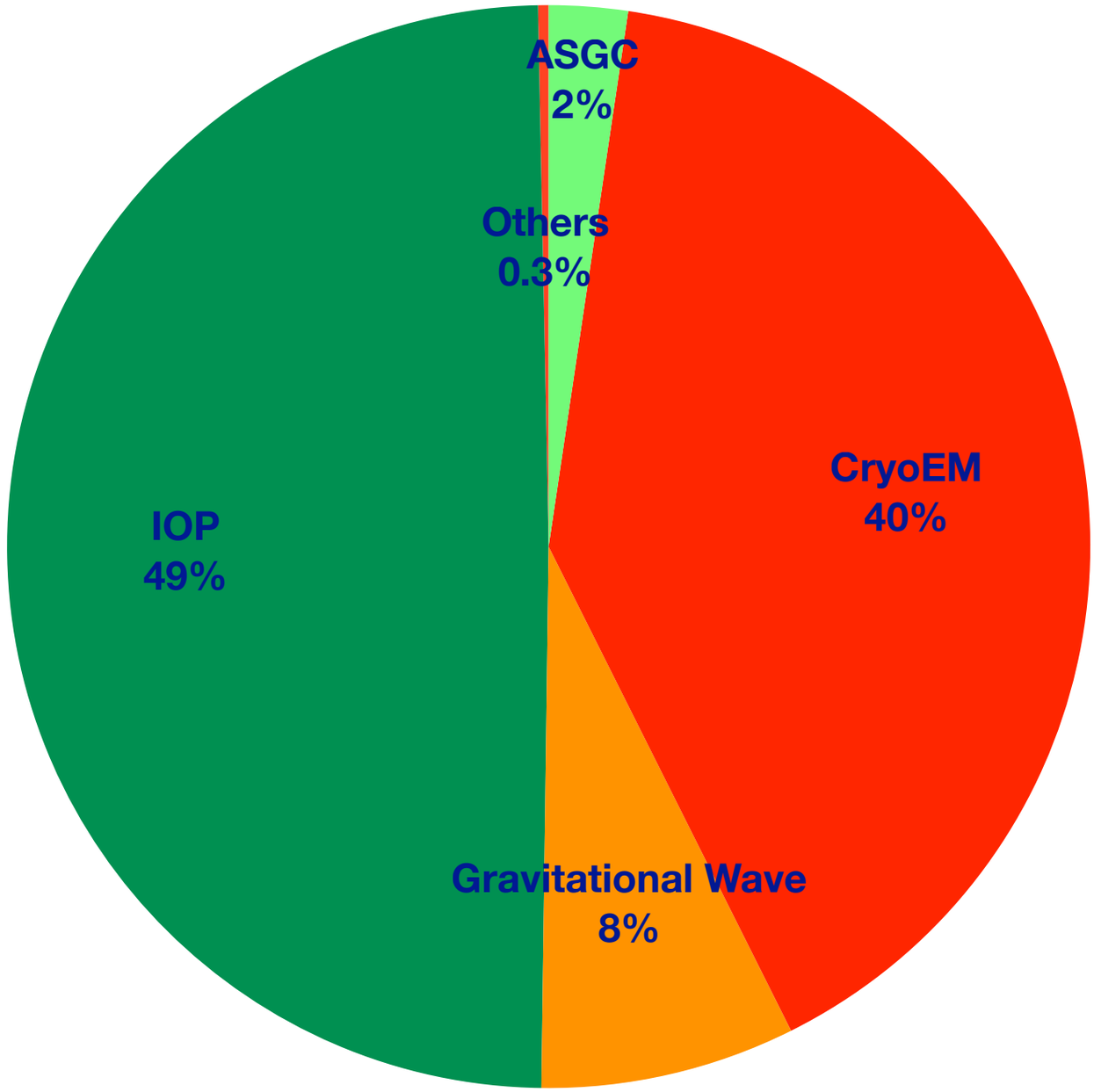


Monthly Average Computing Resource Usage of 2023: 24,670 SRU

ASGC GPU Resource Utilization in 2022 & 2023 (till Sep 2023)

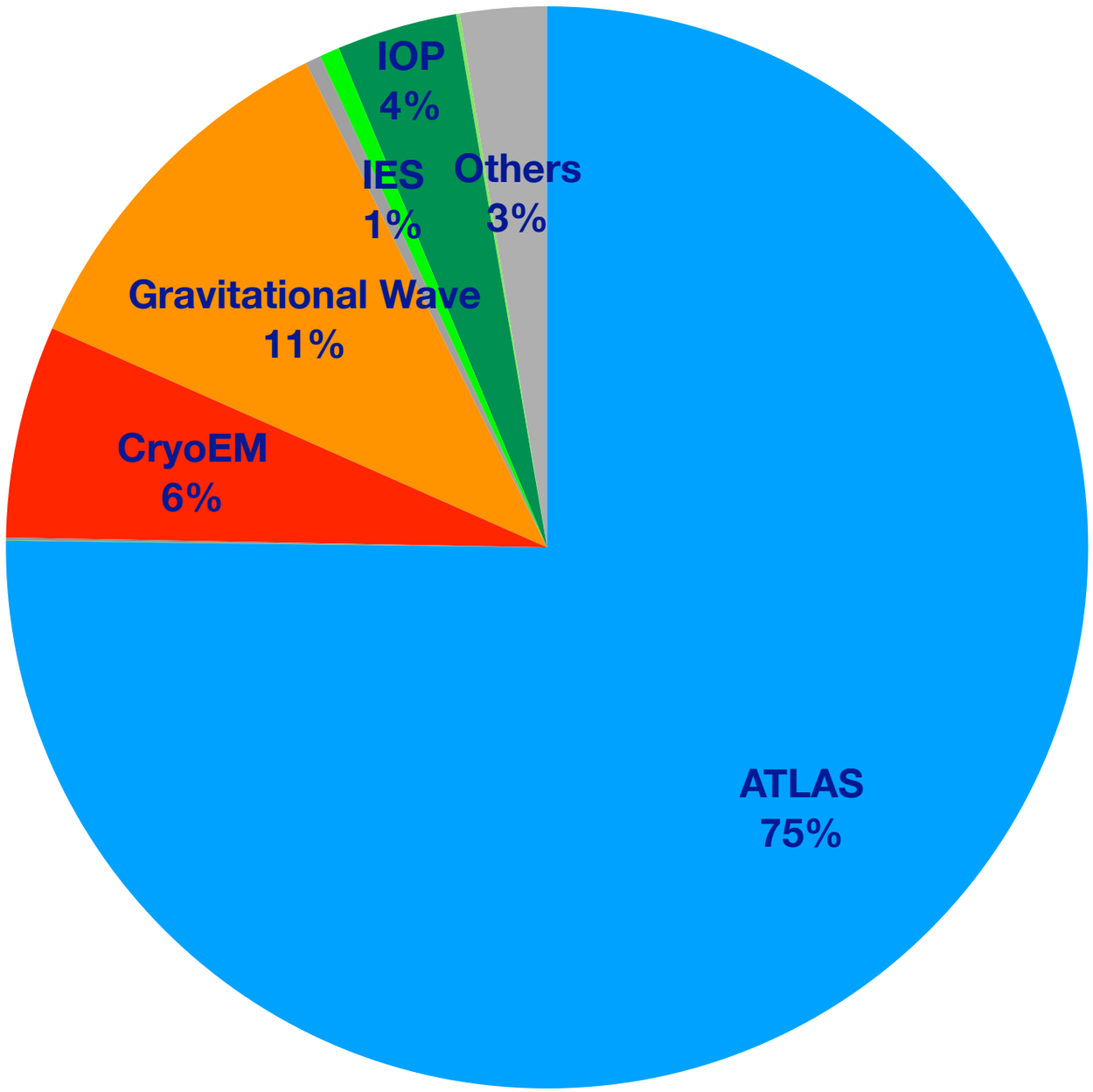


Monthly Average Computing Resource Usage of 2022: 144,987 SRU

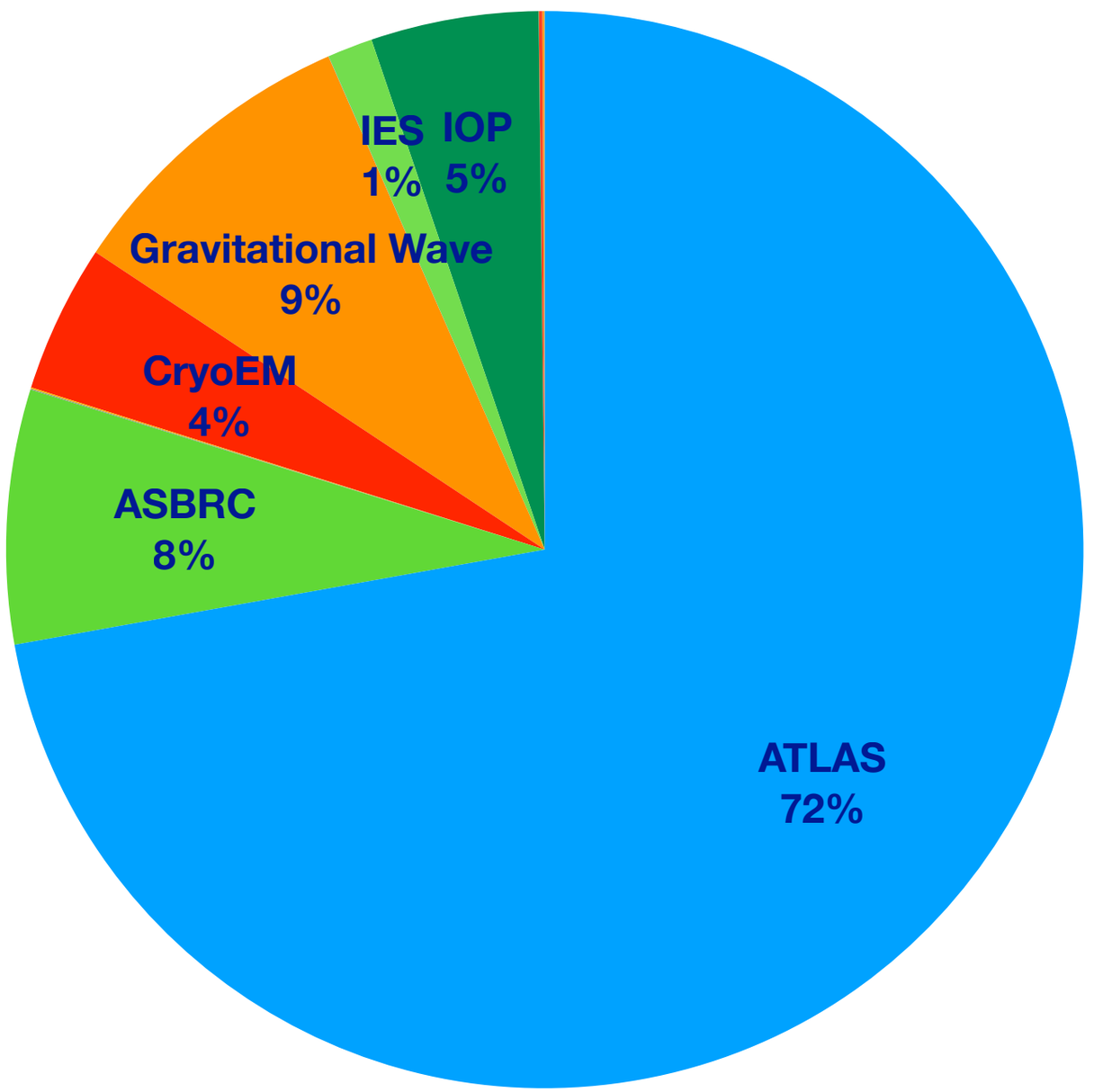


Monthly Average Computing Resource Usage of 2023: 103,977 SRU

ASGC Storage Resource Utilization in 2022 & 2023 (till Sep 2023)

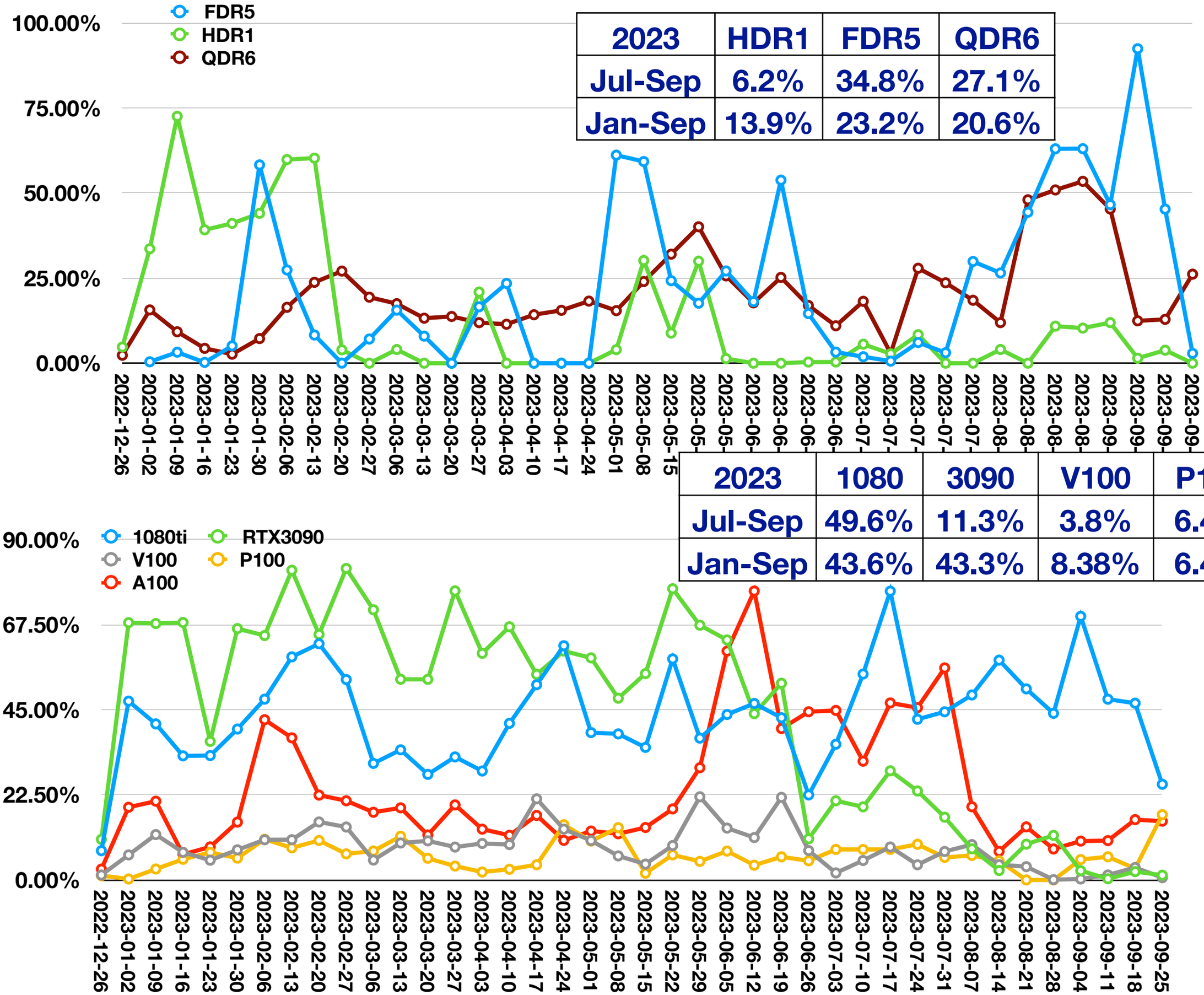


Max Monthly Disk Resource Usage of 2022 (TB)
Total: 4,314 TB
(excluding ATLAS usage 13,000 TB)



Max Monthly Disk Resource Usage of 2023 (TB)
Total: 5,036 TB
(excluding ATLAS usage 13,000 TB)

Utilization of CPU and GPU in 2023

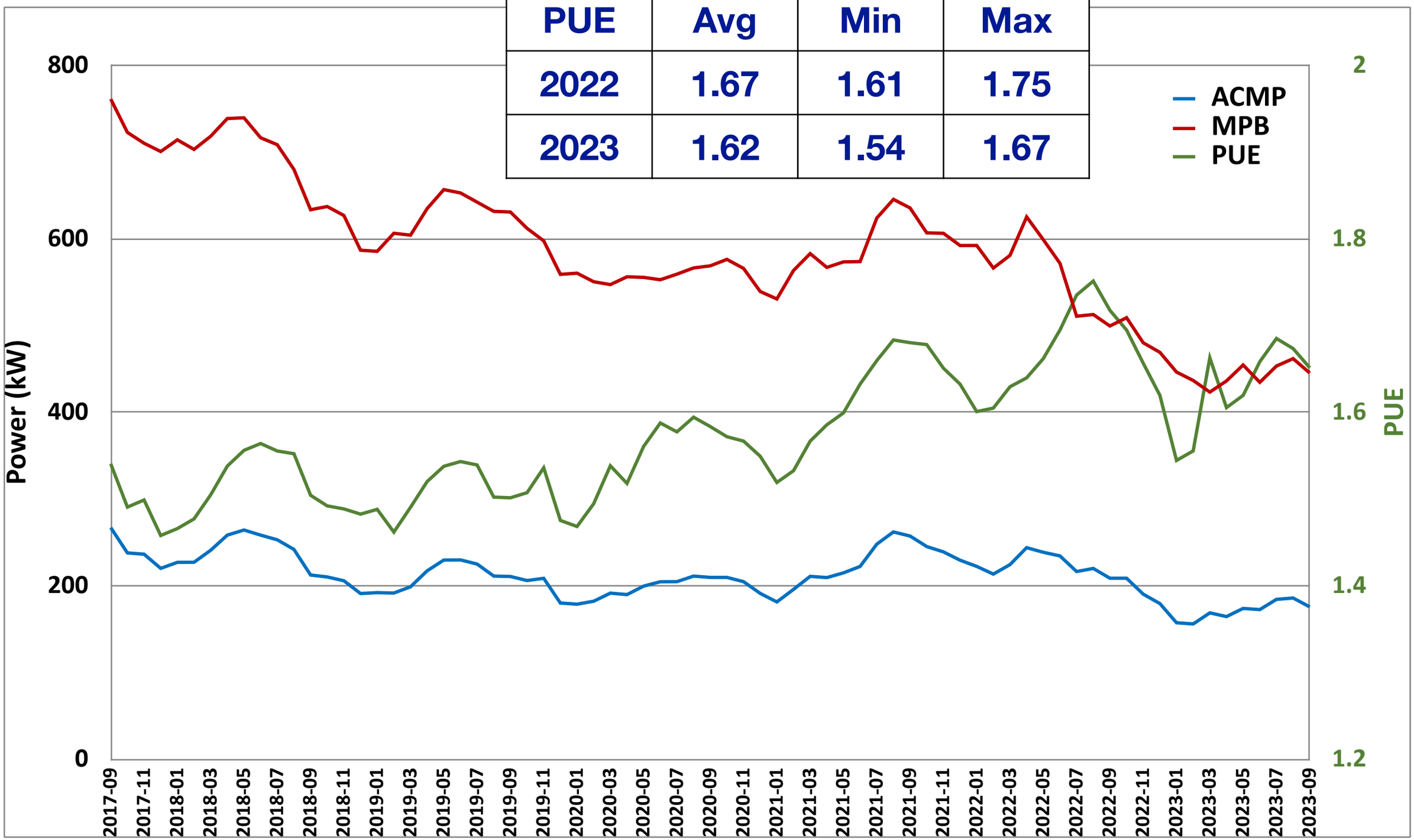


2023	HDR1	FDR5	QDR6
Jul-Sep	6.2%	34.8%	27.1%
Jan-Sep	13.9%	23.2%	20.6%

2023	1080	3090	V100	P100	A100
Jul-Sep	49.6%	11.3%	3.8%	6.4%	25.1%
Jan-Sep	43.6%	43.3%	8.38%	6.4%	23.8%

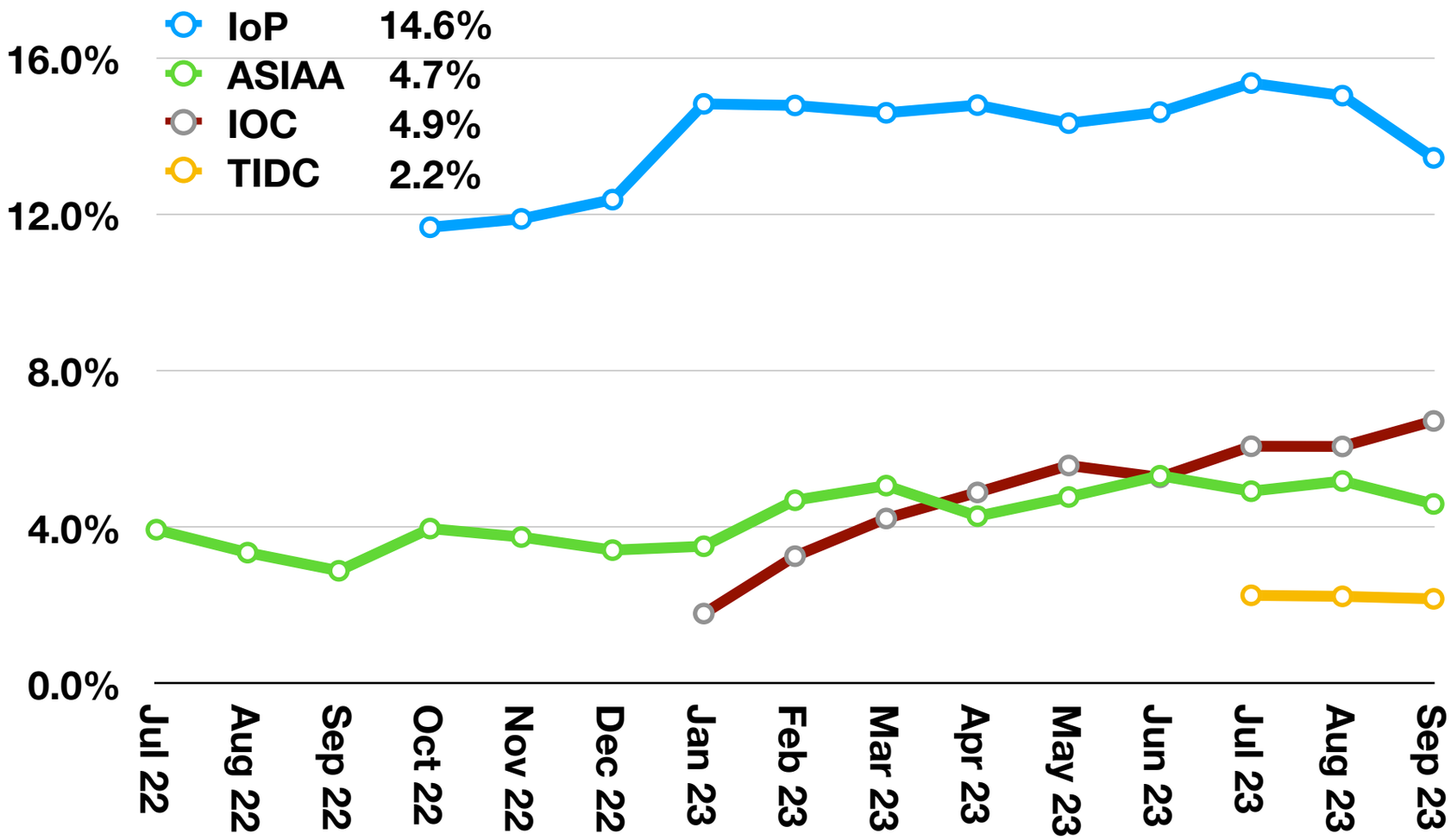
DC Power Consumption (2017.9-2023.9)

PUE	Avg	Min	Max
2022	1.67	1.61	1.75
2023	1.62	1.54	1.67



Power Consumption of Collocated Resources

ASGC DC	Monthly Average (KWH)	Unit Cost (NTD/KWH)	Annual Electricity Cost
2023	309,769.8	3.205	11,975,221
2022	380,966.0	2.710	12,279,072
Comparison	81.31%	118.26%	97.53%



- **ASGC DC Operation (2023Q1&2): $822,756**

Collaboration Model

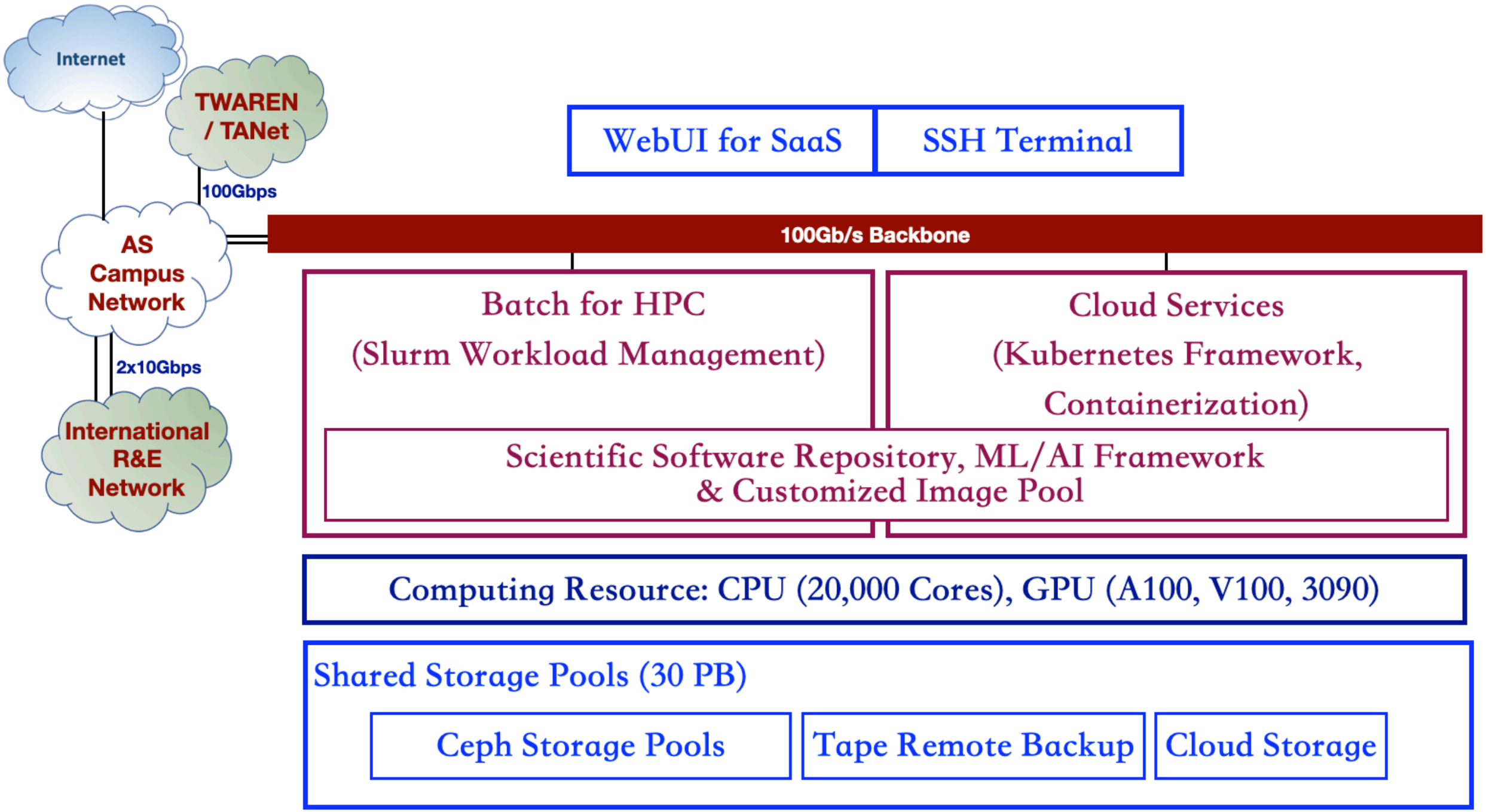
- Pay-as-you-go: using existing resource and services
- Buy-In: joint procurement with ASGC by PI's fund and the procured hardware is shared to all NSTCCore users managed by ASGC
  - Contributor would have priority to the resource and deducted usage fee
  - Examples: IOP, TIDC
- Facility collocation without sharing to all users at this moment (for AS institute only)
  - Example: ASIAA, IOC
- Service Collocation: providing shared services developed by user group(s)
  - Examples: IIS
- Engagement with other Core facility: e.g., ASCEM
- Advanced or Customized services:
  - Shared or common service first
  - CryoEM
- Exclusive Mode (for least-performance CPU nodes): by request, monthly basis › <= 50% capacity. For now, only FDR5 and QDR6 are applicable.

Pricing Scheme and Strategy

- **Pricing Policy (pricing scheme was revised when NSTCCore commenced in Aug. 2023)**
  - **Has to be lower than NCHC (what's the price of NCHC ?), why ?**
  - **Target to cover power cost (utilization matters most)**
  - **Disk storage: 100GB Home and 3TB Group space are default free space; PI could pay for extended storage space, $3/TB-day or $1000/TB-yr**
  - **Tape storage: $300/TB-yr**
- **Collaboration model with CryoEM/ASCEM: further discount is applied after Buy-in reduction**
- **IOP, TIDC: Buy-In model**

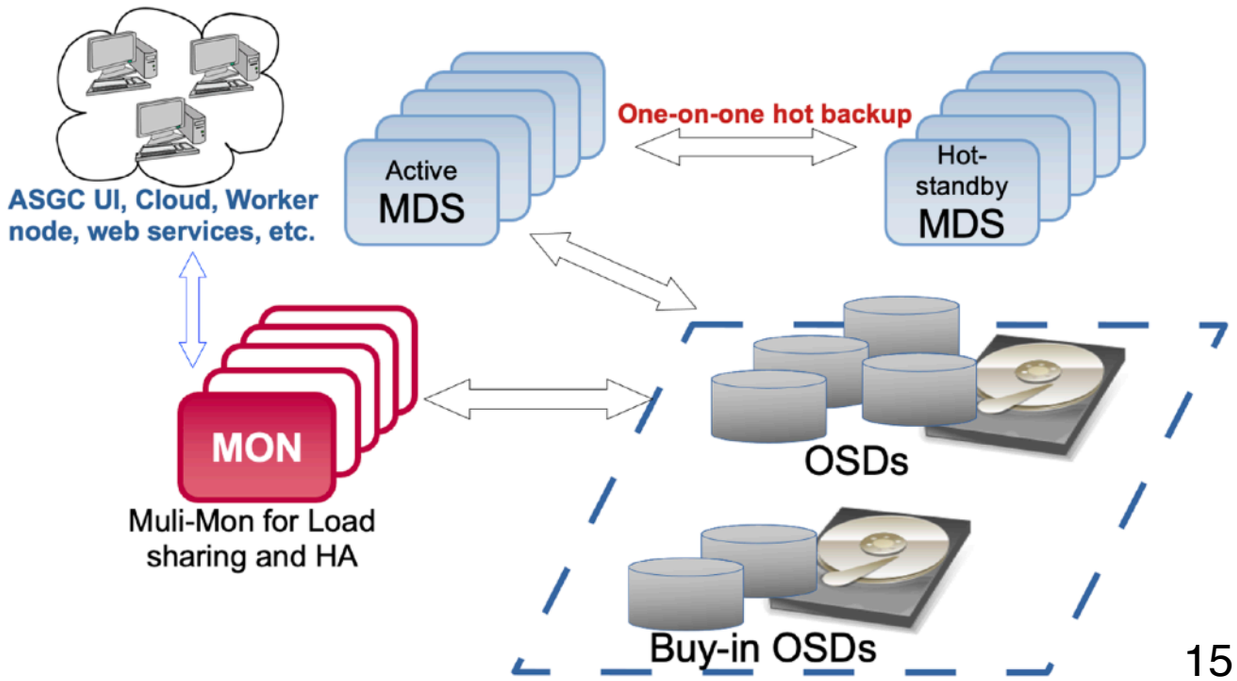
	NSTCCore	SCALE	**New (Aligned)**	Power cost coverage cap
CPU ($/Core-Day)				
AMD Genoa(1792)			**1.5**	Need more info
HDR1(768)	1.1	1.34	**1.1**	Utilization needs 60%+
FDR5 (2209)	0.24	0.48	**0.24**	Very low
QDR6	0.15	0.30	**0.15**	Very low
GPU ($/Board-Day)				
A100 (24)	346	346	**173**	Good if Util > 35%
V100 (48)	140	140	**70**	Good if Util > 40%
3090 (32)	158	158	**79**	Good if Util > 40%
P100 (8)		94	**20**	
1080 (64)		2	**1**	
Storage ($/TB-Yr)				
Disk	1000	1500	**1000**	
Tape	300		**300**	

ASGC Science Cloud System Diagram

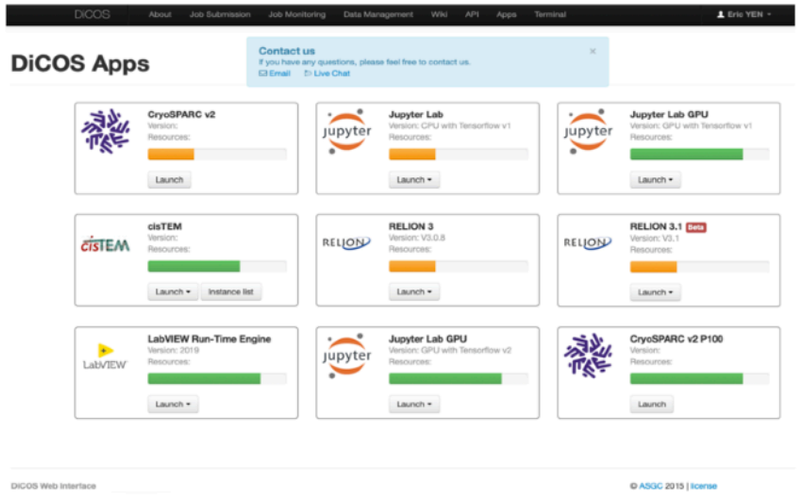


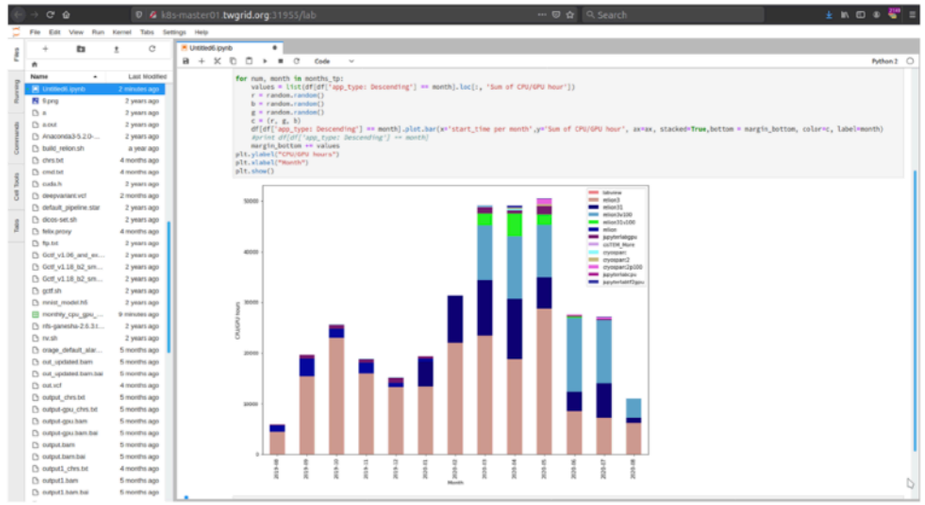
ASGC Science Cloud Storage Architecture

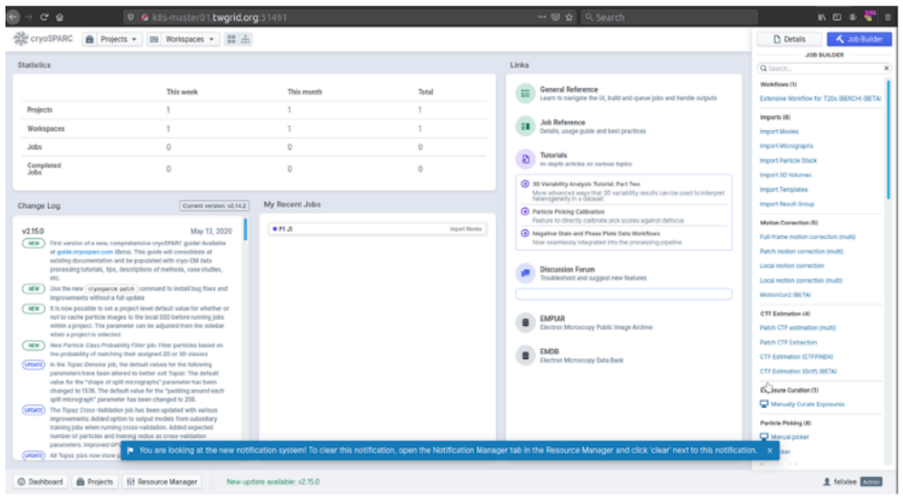
- Scalable and reliable online storage system based on Ceph mainly
- Ceph Configurations: ~9PB
  - 6 MDS + 6 hot-standby (one-on-one backup); 7 MONs
  - 462 OSDs, 51 hosts.
- Services
  - 3 TB/PI Group setup by default; PI could extend the space through management UI flexibly
- Reached 2GB/s R/W throughput so far
- Tape-based remote backup system (4PB) will be established and integrated in late 2023, supported by EOS
- Providing big pool for HPC, HTC, AI and various applications concurrently
- Capacity will be growing to 13PB by end of 2023
  - Plan to procure new 4PB disk servers for Ceph System in 2024 and 2025 respectively

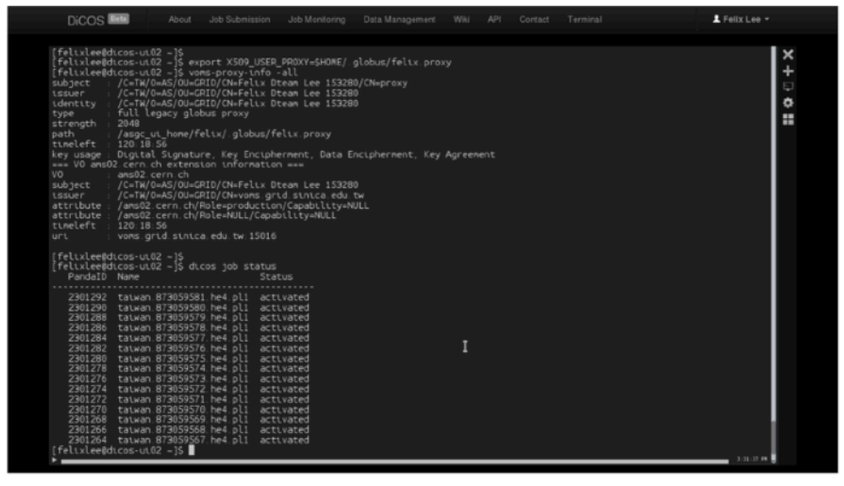


Supporting Big Data & AI in Innovations









CLI | **Web Portal** | **DiCOS APP** | **Jupyter Notebook** | **Science Portal** | **Web Browser/ Terminal**

Application-specific/ Generic Learning Engines

Deep Learning Engines/Frameworks

Computing Resource (Cloud/Grid/Slurm)

Storage Resource (Ceph/EOS)

Distributed Data Management & Cloud Storage Services

Network & Data Transmission Services

50+ Web Applications Provided

- Web Portal
- Application over Cloud
- Jupyterlab
- Web Terminal

PHYS

Deepmd-kit
Version: GPU with A100
Resources: 12%
Launch

Deepmd-kit
Version: GPU with V100
Resources: 80%
Launch

MAML
Version: GPU with A100
Resources: 12%
Launch

MAML
Version: GPU with V100
Resources: 80%
Launch

PVserver
Version: 5.8.0 (GPU 1080Ti)
Resources: 66%
Launch

Paraview Client
Version: 5.8.0
Resources: 97%
Launch

PyRoot
Version: GPU with 1080ti
Resources: 66%
Launch

Jupyter

Jupyter Lab
Version: CPU with Tensorflow v1
Resources: 97%
Launch

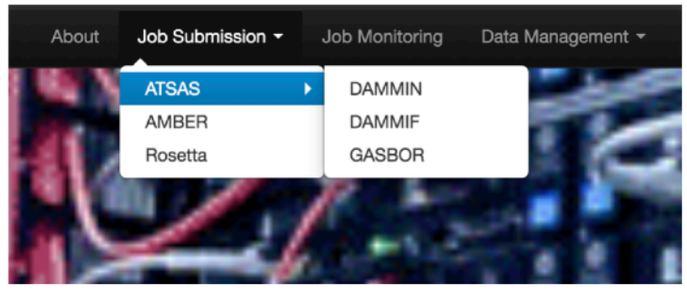
Jupyter Lab gpu 3090
Version: GPU with Tensorflow 3090
Resources: 51%
Launch

Jupyter Lab GPU V100
Version: GPU with Tensorflow V100
Resources: 80%
Launch

Jupyter Lab GPU A100
Version: GPU with Tensorflow A100
Resources: 12%
Launch

LabVIEW Run-Time Engine
Version: 2019
Launch

DiCOS-BioSAXS Platform



Other

spyder cpu/eman2
Version:
Resources: 97%
Launch

Octave
Version: V5.2
Resources: 66%
Launch

Transfer Data
Version:
Resources: 97%
Launch

cisTEM
Version:
Resources: 100.0%
Launch

Ovito
Version:
Resources: 97%
Launch

OpenACC
Version: GPU P100
Resources: 50%
Launch

Triton
Version: 22.01-py3 (GPU P100)
Resources: 50%
Launch

AlphaFold
Version: GPU with V100
Resources: 80%
Launch

AlphaFold
Version: GPU with A100
Resources: 12%
Launch

IMOD
Version:
Resources: 66%
Launch

RoseTTAFold
Version:
Resources: 51%
Launch

Dynamo
Version:
Resources: 66%
Launch

Continuous Improvement and Evolution

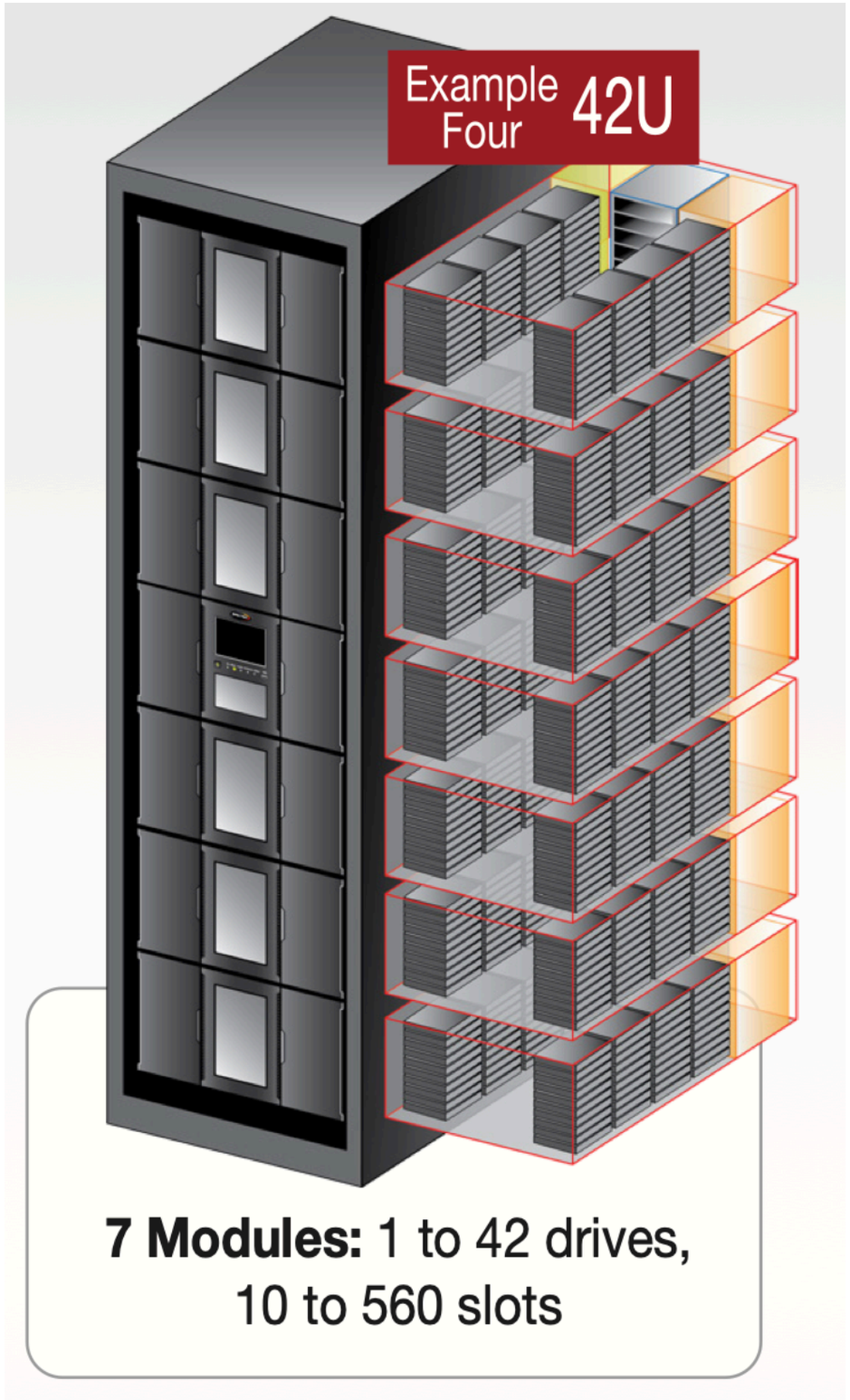
- Reliability and efficiency are the primary target
- Improvement from incident, collaboration and IT evolution
  - Ceph has been in a very stable state since recovery from May 2023: 100% availability
    - Automatic recovery again HDD failures works very well: almost zero impact to users and services
  - DC network reliability is 100% since the replacement of legacy network devices in March 2023 (and Nov 2022)
  - Harness the performance of AMD CPU
  - ML-enabled data analysis support is available, started from platform and containerization
  - Increasing power cost is a big challenge: 20% power usage reduction in 2023, but the cost is 20% growth
- Capacity development for the research infrastructure and services: core technology, talent and efficiency
  - Service quality upgrade by R&D. Collaboration relies on trust.
  - Next goal is to win the 2nd 2-yr term core facility fund from AS (2025-2026, proposal dues in mid March 2024)

NSTC Core Facility Resource Plan

Resource Plan	PY1 (2023.6-2024.5)	PY2 (2024.6- 2025.5)	PY3 (2025.6- 2026.5)	Remark
Shared ASGC Resource	CPU, GPU (A100, V100), 1PB Disk	CPU, GPU (A100, V100)	CPU, GPU (A100, V100)	
CPU (#Cores)	1,792	1,792	1,792	Plan to buy AMD Genoa (96Cores/CPU) in PY1
GPU	0	0	0	Able to share A100/V100 of ASGC with lower priority
Storage (PB)	3	3	3	Ceph-based; 3TB/group free space for computing; 1000/TB-yr
Tape (PB)	4	4	4	Integrated with Ceph; ease-to-access workflow; 300/TB-yr
Core Service	192 CPUCores 0.25PB Disk	192 CPUCores, 0.25PB Disk	192 CPUCores, 0.25PB Disk	For Cloud, Ceph, UI, Monitoring etc.

Tape Remote Backup System Will be setup by end of 2023

- **Serving as 2nd layer remote backup system**
  - **Plan to setup at a separate data center**
  - **For cold data, or 2nd-copy backup**
  - **For backup of users' core data on Ceph**
  - **Reliability of ASGC services will be increased**
- **Scalability: capacity on demand**
  - **Max 7 modules x 6u, 42 drives, 560 tape slots**
  - **LTO-9 tape: 1.44PB (native)/3.6PB (compressed) per module (80x 18TB/tapes)**
- **Tape drive performance: 300 - 400 MB/s using fibre network**
- **Integration and services: based on EOS and CTA**
- **Tape-related data services should be operational by end Q1 2024.**
- **Will be extended to 12PB capacity in 2024**



Support and Service of ML-Enabled Data Analytics by ASGC

- **ML/AI application platform service is available NOW - SW library, HW, integration and application**
  - **Build up customized ML platforms for user specified projects - Deploy ML packages ready envirnonment in order to help ML development smoothly and provide on-demand computing power**
  - **Upkeep of the application framework**
  - **Workflow and data pipeline integration**
  - **Efficiency Improvement**
- **Potential use cases**
  - **Users who bring existing source code - ASGC could help to setup a virtual environment and confirm source code running normally**
- **Approaches**
  - **Supporting Kubernetes/Jupyter lab for development purpose**
    - **Create Kubernetes/Jupyter lab environment with user specified ML packages ready.**
    - **Support on-demand scalable CPU/GPU computing power.**
  - **Supporting containerized environment (e.g, Docker image) for deployment purpose**
    - **Create takeout images in Docker format as an option for user who wants to train/predict model**
    - **Docker images could be downloaded from ASGC server and deployed on users' Docker Desktop on Windows/Linux.**

Available Hardware, Software & Use Cases

- GPU Servers (with local SSD enhanced)
  - A100 (8xboards/server, 80GB RAM/board) * 3
  - V100 (8xboards/server, 24GB RAM/board) * 6
  - 3090 (8xboards/server, 11GB RAM/board) * 4
- ML related framework and tools
  - TensorFlow, PyTorch, Keras, NVIDIA Triton, Scikit Learn
- Large-scale storage /file system
  - 9 Petabyte+ disk-based storage system managed by CephFS
  - Tape-based backup storage will be available by end of 2023
- Use Cases
  - CryoEM - ML-enabled bioimage processing
  - Deployment of ML-enabled protein simulation tools - AlphaFold, RosettaFold & Diffusion, DiffDock
  - Deployment of ML-enabled packages (by IOP PABS group): DeepMD-kit (with interface with LAMMPS)
  - AMS & KAGRA - programs developed by local groups
  - Data Center intelligent monitoring & control (ASGC projects): Air Handler, power saving, etc.

Welcome To ISGC2024 in Taipei

- Schedule: 24-29 March 2024
- Venue: Academia Sinica, Taipei, Taiwan
- Call for Abstract/ Session will be open on 20 Oct. until 30 Nov 2023
- Event Web site: https://indico4.twgrid.org/event/33/
- Contact: ISGC Secretariat
  - vic@twgrid.org

ASGC Services

- Weekly User Meeting: 1:20pm, Wed
- ASGC Web Site: https://www.twgrid.org
- Access to ASGC Resources
  - https://dicos.grid.sinica.edu.tw/
- Contact point: DiCOS-Support@twgrid.org

Your Advice and Support Are Indispensable to ASGC

- **Comments & Discussion**