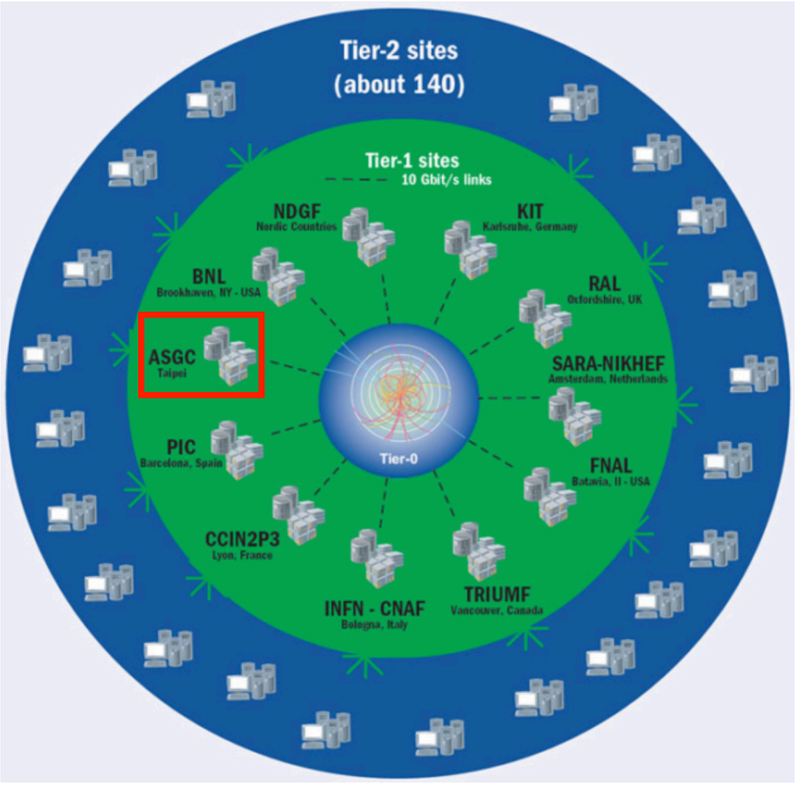
ML/AI Application Support of ASGC

Eric Yen and Felix Lee
Academia Sinica Grid Computing Centre (ASGC)

AS-IOP Machine Learning Workshop
22 April 2022

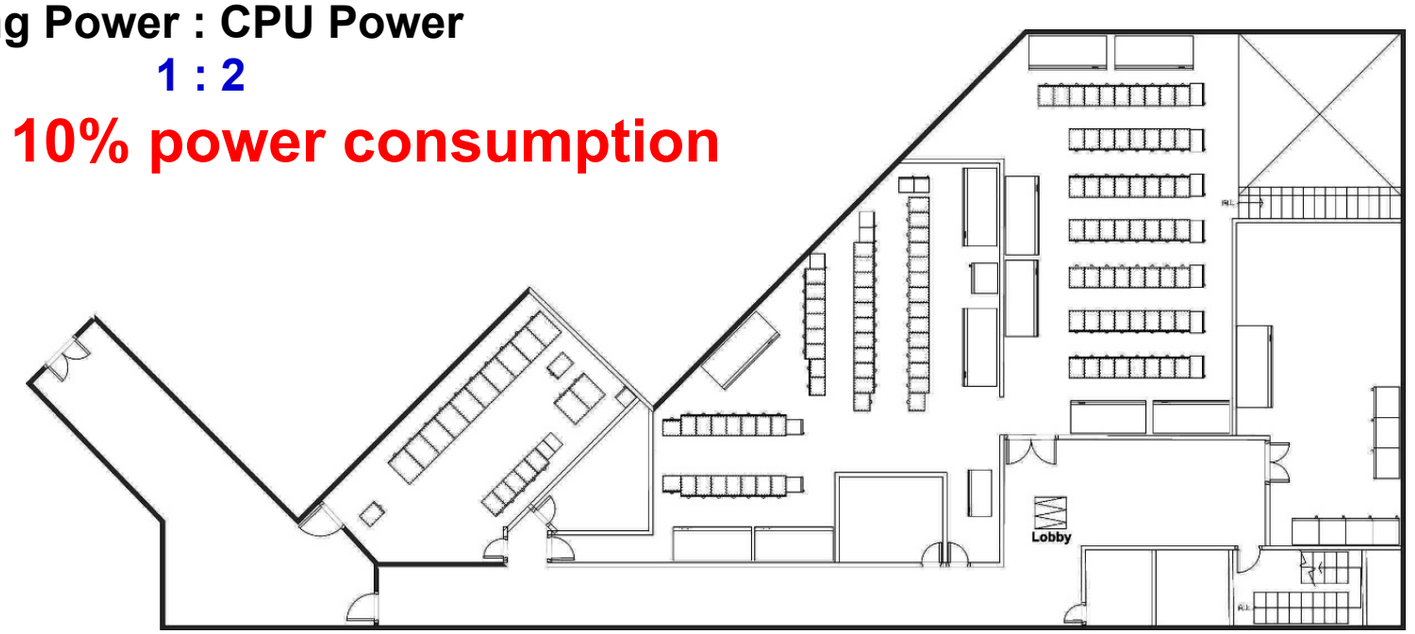
ASGC Mission

- Founded as a WLCG Tier-1 Centre in 2005 and jointly develop & deploy advanced, worldwide distributed cloud infrastructure for WLCG
- Enhancing science-enabling capabilities by advanced research infrastructure and big data analytics
  - Building capacity of large-scale distributed cloud for efficient big data analysis and scientific computing of Academia Sinica
  - The research infrastructure, application and services are improved progressively along with growing scientific applications of various disciplines
  - ML-enabled data analysis framework is also upgraded
- System efficiency optimization: power, thermal, application, operation, system, etc.
- ASGC will become a Core Facility for ML-enabled big data analysis and scientific computing from 2023

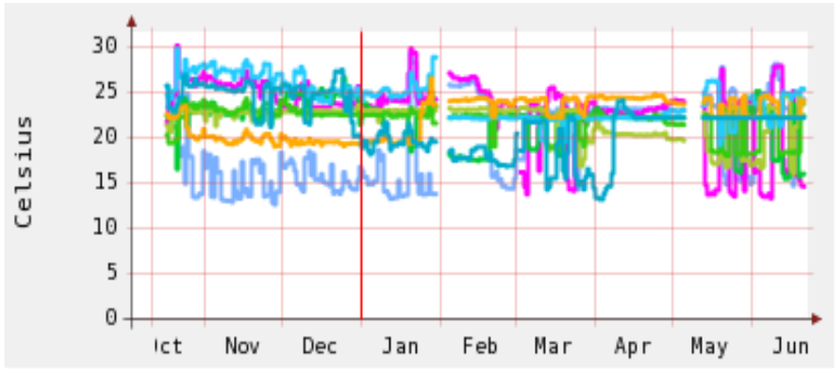


ASGC Resources

Cooling Power : CPU Power
1 : 2
No UPS to save 10% power consumption



- **Total Capacity**
  - 2MW, 400 tons AHUs
  - 112 racks in ~ 800 m²
- **Resources (2021)**
  - 13,652 CPU Cores
  - 192 GPU Cards
    - A100 Server *2 (8 GPU Card for each) are available in 2021
  - 24 PB Disk Storage
  - 2x10Gb links to CERN and primary NRENs worldwide
- WLCG Tier-1 Center since 2005
- Supporting HPC & HTC in Academia Sinica by distributed cloud operating system (DiCOS)
  - Usage > 1M CPUCore-Days in 2015
  - Usage > 2M CPUCore-Days in 2019
  - GPU usage is growing exponentially from 2017
- R&D on system efficiency optimization by intelligent monitoring & control



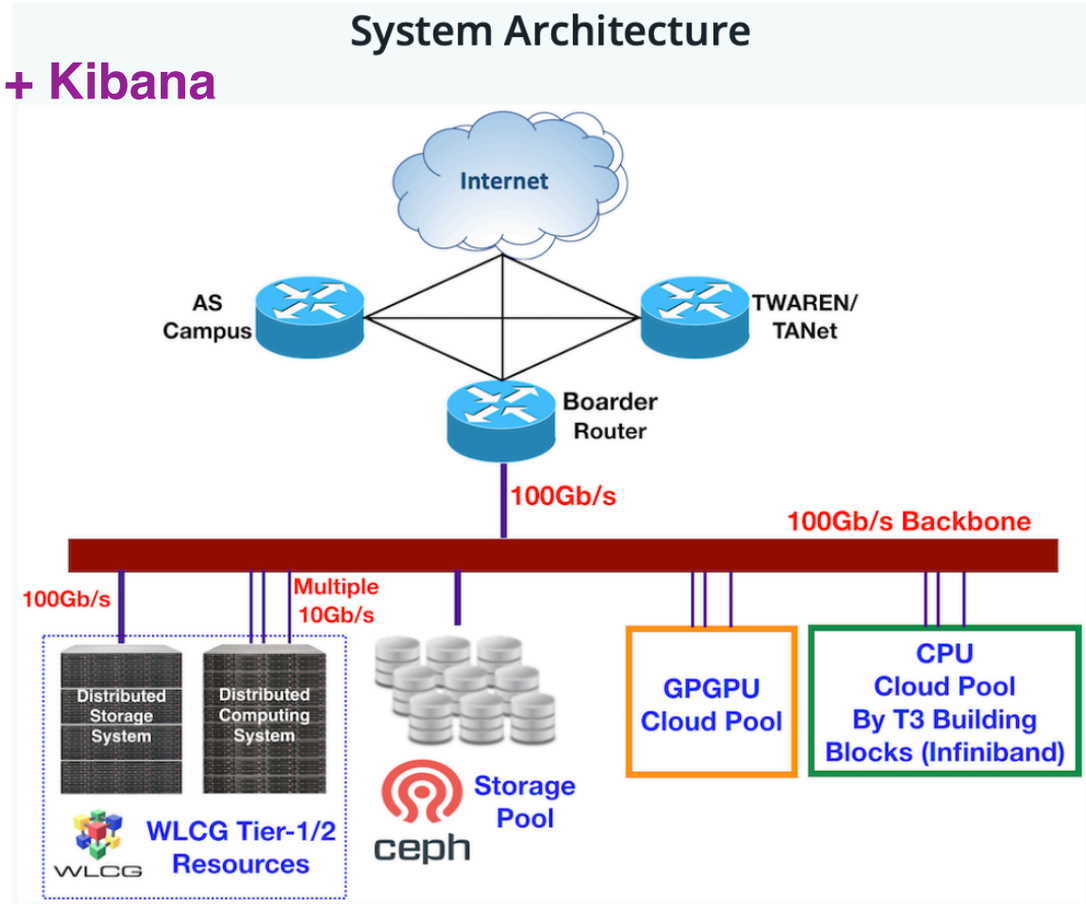
Monitoring the power consumption and temperature of every piece of equipment every 10 seconds.

All software used are open-source codes developed by ASGC and an international collaboration led by CERN

3

ASGC Science Cloud Infrastructure (DiCOS)

- OpenStack Cloud: for core services and on-demand worker nodes
  - Separate regions for scalability
  - Multiple cells/Region for various configurations and capabilities
    - e.g. GPU, Netron Compute, Nova Compute, …
  - Single hypervisor type: KVM
    - #hypervisors: 100+
    - #VMs: 500+, dynamic provisioning
  - Networking: flat and segmented
- Containerized Resources managed by Kubernetes framework
  - User cluster:
    - batch, interactive GUI jobs: remote Jupyterlab, virtual desktop
    - GPU Cloud
  - Core Services: distributed cloud cores; Elasticsearch + Kibana
- Storage system: Ceph-based
  - OpenStack services (RBD): 1.5PB
  - Users: CephFS (7+PB)
  - Cloud Storage (DropBox-like) services: 2TB/user
    - Accessible from user home directory
- UI: Web UI/Terminal; JupyterLab
- Operation and management
  - Source control: Gitlab
  - Puppet-based deployment of components
  - HELM

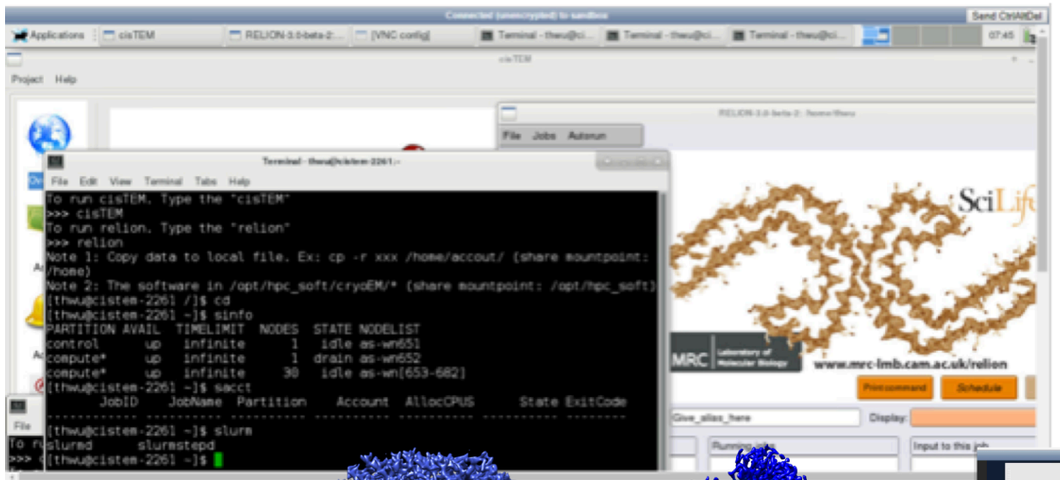


DiCOS for ML

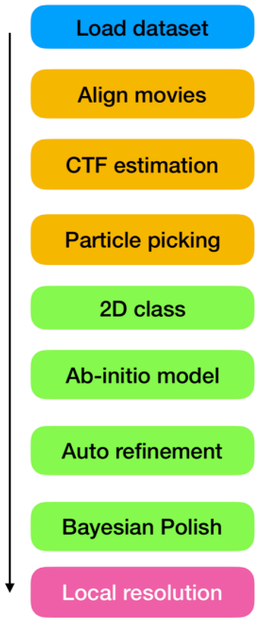
- DiCOSAPP (Interactive jobs) - user-oriented, Web-based computing environment or application portal
  - Jupyterlab (Jupyter notebooks)
    - PyTorch
    - Tensorflow2.x & Keras
  - Triton inference server
  - Cloud storage attached (2TB/user)
  - GPU: GTX-1080, RTX-3090, P100, V100(32G), A100(80G)
  - MPI/CPU Cluster resource is also accessible
- Batch jobs
  - through DiCOS command line UI or web terminal :
    - Queue: ANALY_TAIWAN_TWGRID_ML_V100
      - The system environment is the same as Jupyterlab
      - GPU: V100

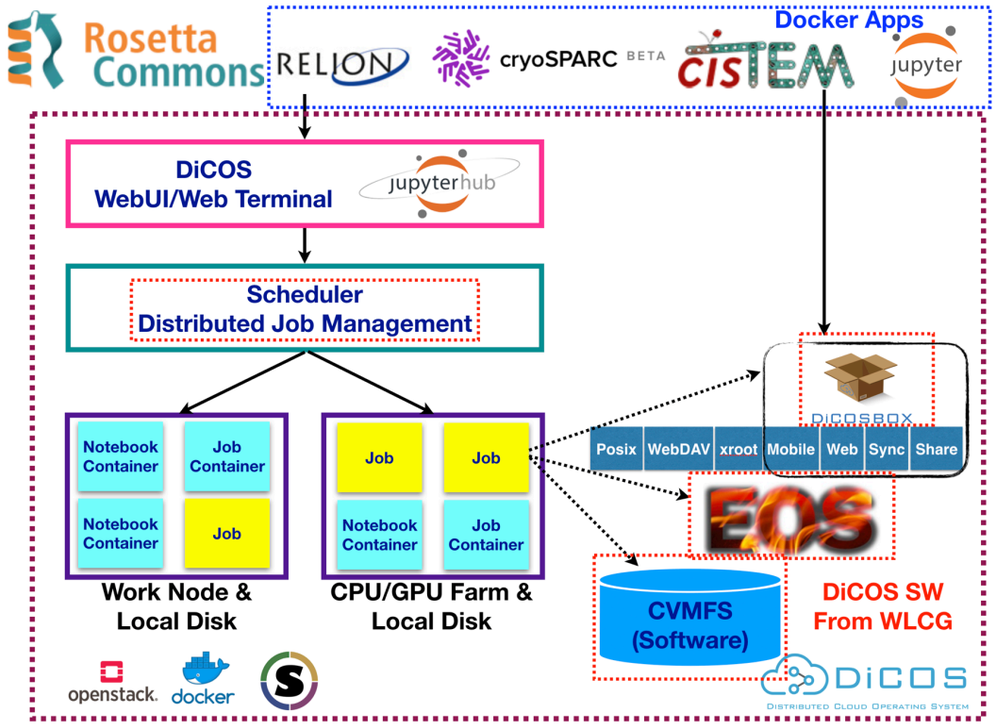
Supporting Cryo-EM Applications by DiCOS

- Primary GPU (single precision) and CPU (300-600 threads) users, ~5-10TB input/job
- ASGC Contributions: providing high-performance big data analysis and processing for multiple user groups concurrently in AS, with customized workflow
  - Integration of analysis pipeline over ASGC distributed cloud infrastructure, through Relion, CryoSparc and cisTEM
    - Computing (CPU, GPU, cluster), storage, networking, and web portal
    - Software and computing workflow setup, upgrade, maintenance and system operation
  - Shorten latency between ASCEM and analysis facility: on-the-fly processing & transmission
  - Web /JupyterLab UI development
  - Efficiency optimization: analysis workflow and applications
  - Extending from CryoEM to tomography
  - ASCEM IT & Security operation support
  - Data access from EMPIAR
  - New resource/tools provisioning: RosettaFold, AlphaFold

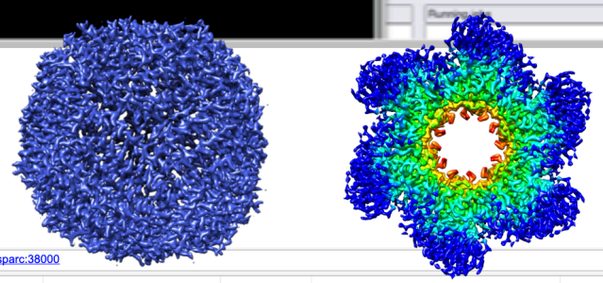


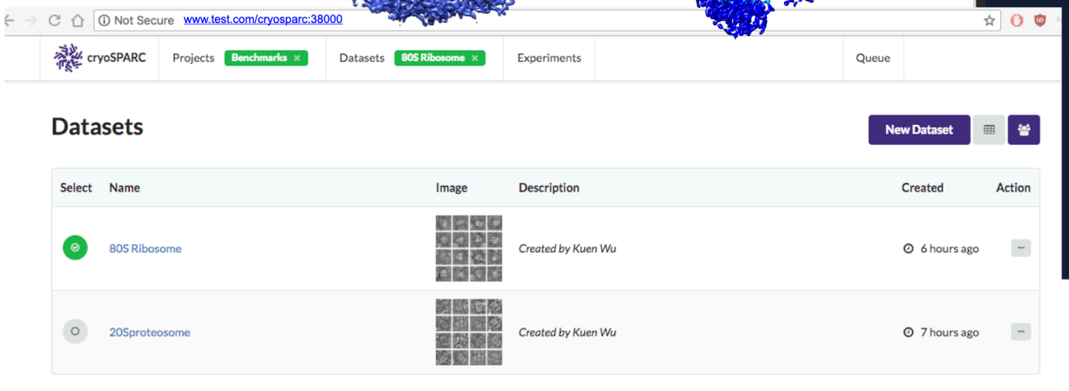
cisTEM

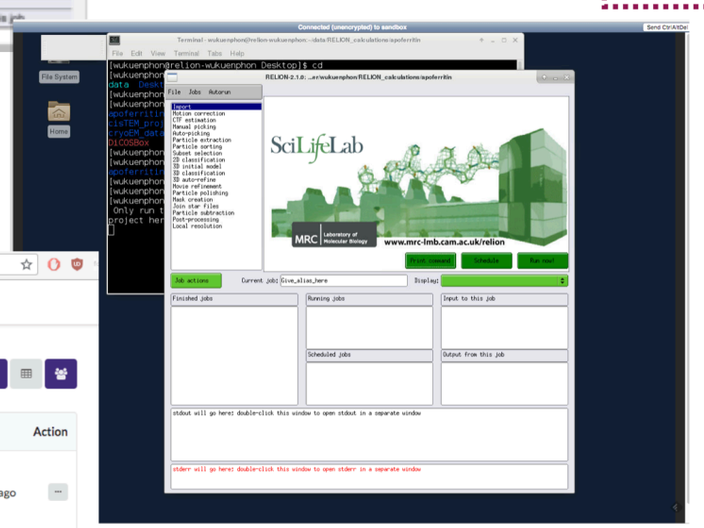




cryoSPARC







Relion

DiCOS
Distributed Cloud Operating System

AlphaFold2 Use Case

- AlphaFold: AI tool which could predict 3D shape of proteins from genetic sequence with accuracy
- Comparison of running on Google Colab and ASGC V100 server
  - 240min vs 15min for one simulation;
  - Length of amino acid: 900 vs 2000
- Research Applications
  - Large-scale screening for protein-protein interactions: ~15,000 protein
    - Could be finished in ONE month by ASGC V100. Or, shorter time by 2xV100
    - Only Top10 candidates shall be confirmed their structures through CryoEM experiments afterwards
    - ~200K NTD$ cost for consumables would be saved
- ASGC Support
  - Develop ease-of-use AlphaFold2 Web App
  - Support automation of large amount job submission when mutation varies
- RoseTTAFold is also available @ASGC - DL-enabled protein structure prediction service developed by U. Washington

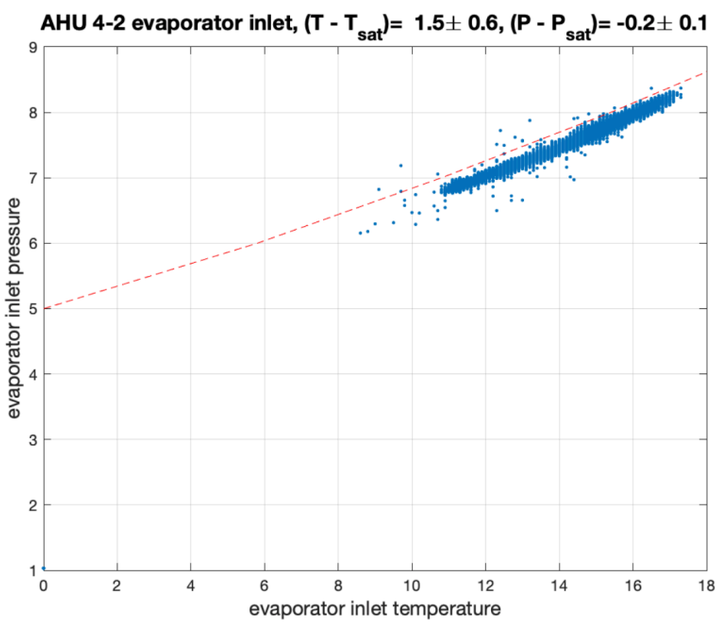
Other Use Cases

- **DeepMD-kit setup over GPU**
  - **Deep learning package for many-body potential energy representation and molecular dynamics using TensorFlow**
  - **Support interface with LAMMPS**
  - **ASGC also build up Web App for DeepMD**
- **NVIDIA Triton Inference Server**
  - **“streamlines AI inference by enabling teams to deploy, run and scale trained AI models from any framework on any GPU- or CPU-based infrastructure.”**
  - **Triton environment @ASGC is available**
- **Workflow, UI and analysis pipeline will be improved progressively according to users experiences and innovations**

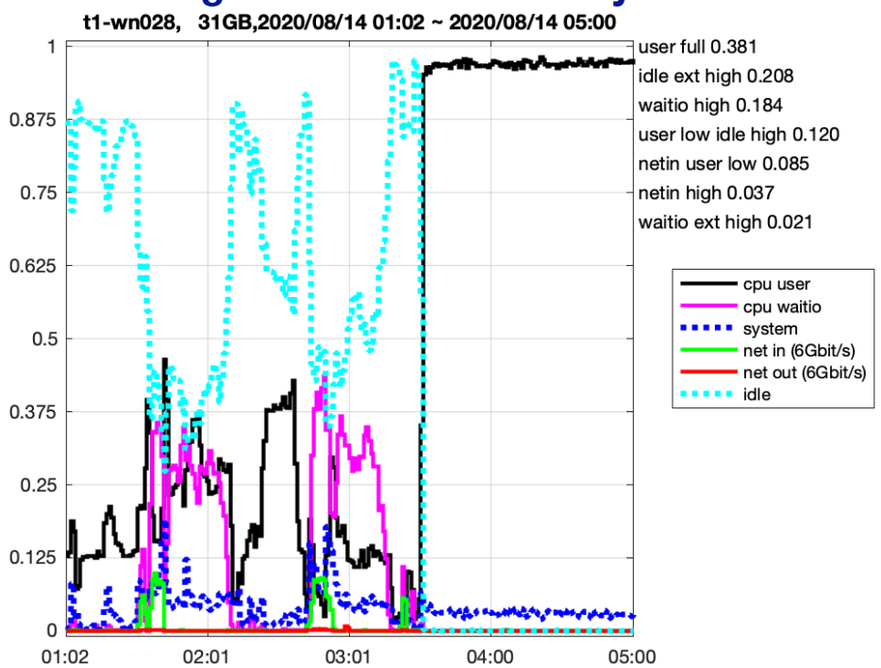
System Efficiency Optimization

- Goals: maximize application performance by available resources dynamically, in terms of power, thermal and system (Comp, Storage, Network, application) efficiency
- Scope: Power, Thermal and Distributed Cloud System management
- Strategy: intelligent monitory and control assisted by ML
- Example: Thermal management, Compute/storage/network anomaly detection, Power saving of work nodes
- AHU monitoring and control
  - Detection of refrigerant operating issues and abnormal components; Efficiency optimization
  - 13 sensors x 16AHU; 18K data points/day;
  - Realtime monitoring, adjustment and diagnostics: refrigerant operating issue; abnormal components detection; efficiency tuning; ML-based automatic detection of critical problems;
- System Anomaly Detection
  - Classify machine status into 5 clusters daily: based on CPU-user, CPU-wio, CPU-system, CPU-idle, Network In/Out
  - >30M records/day from all systems of ASGC are covered
  - 146 events in 14 types identified during March 2020 - March 2021

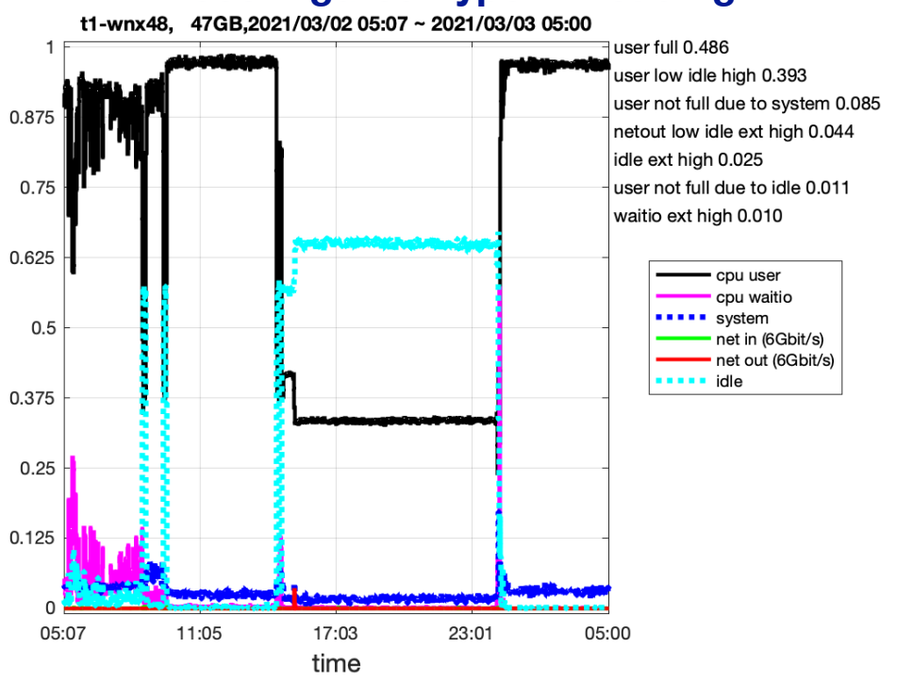
AHU Performance Monitoring



Worknode Monitoring: High ratio of WaitIO & System



Worknode Monitoring: Misconfigured hyper-threading

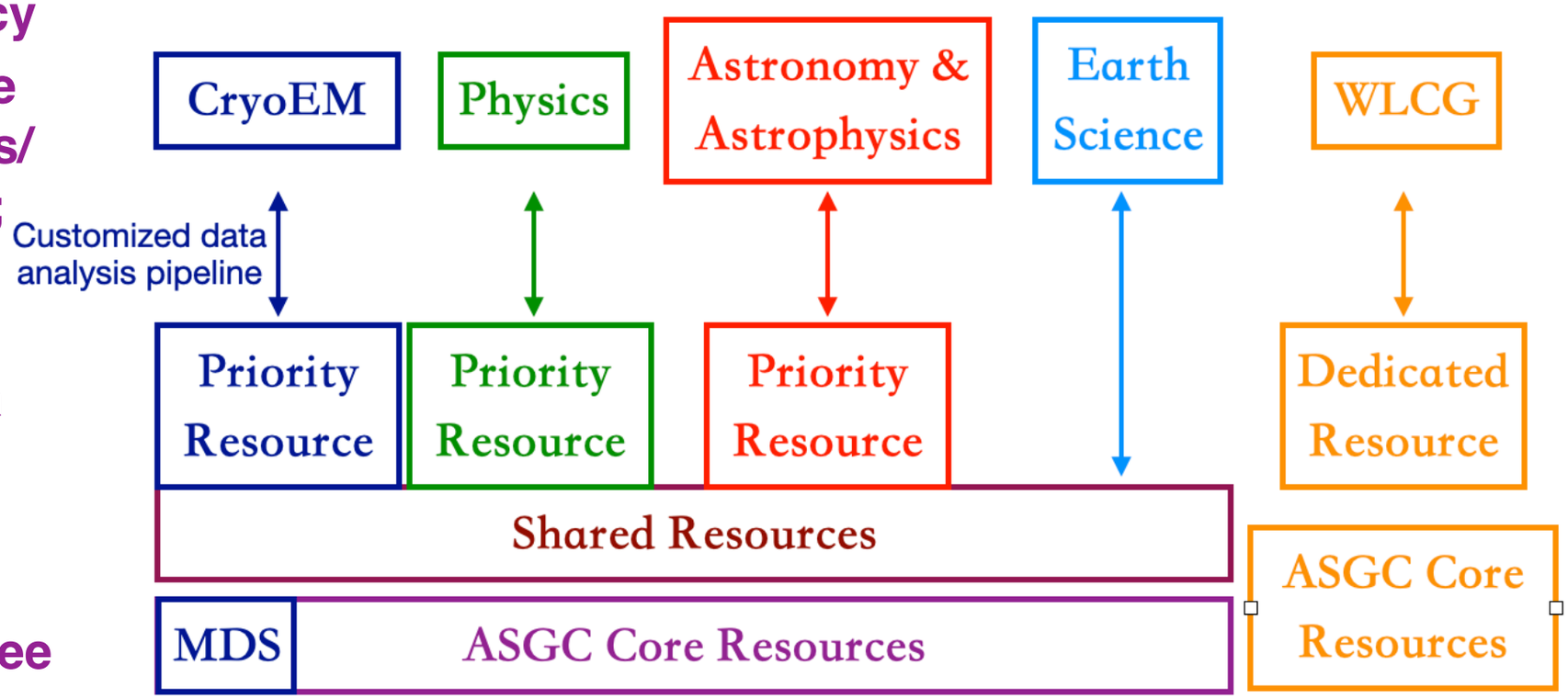


ML-Cloud

- Research groups could focus on scientific problems solving
- Application environment: Research Infrastructure based
  - HPC facility - computing + storage + network
  - Ease-to-use and customized analysis pipeline, including application and platform
  - Research data management (including tools, backup and archive)
- Use Cases collections: reusable, reproducible, training, learning and benchmarking
  - Based on FAIR principles
  - Coverage: data set, tools, execution environment (e.g, Jupyter Notebook), publication (e.g, documents)
- Leveraging HPC powers
  - To shorten training time, support pre-training by large amount of datasets, support (near) real-time applications, …
- Consulting services: ASGC + volunteers and invited experts
- Capacity building:
  - Providing ML resources based on IOP/users needs
    - Datasets, tools, codes, algorithms, case studies, …
    - Learning materials in all aspects
  - Training, workshop, GPU computing workshop, hackathon

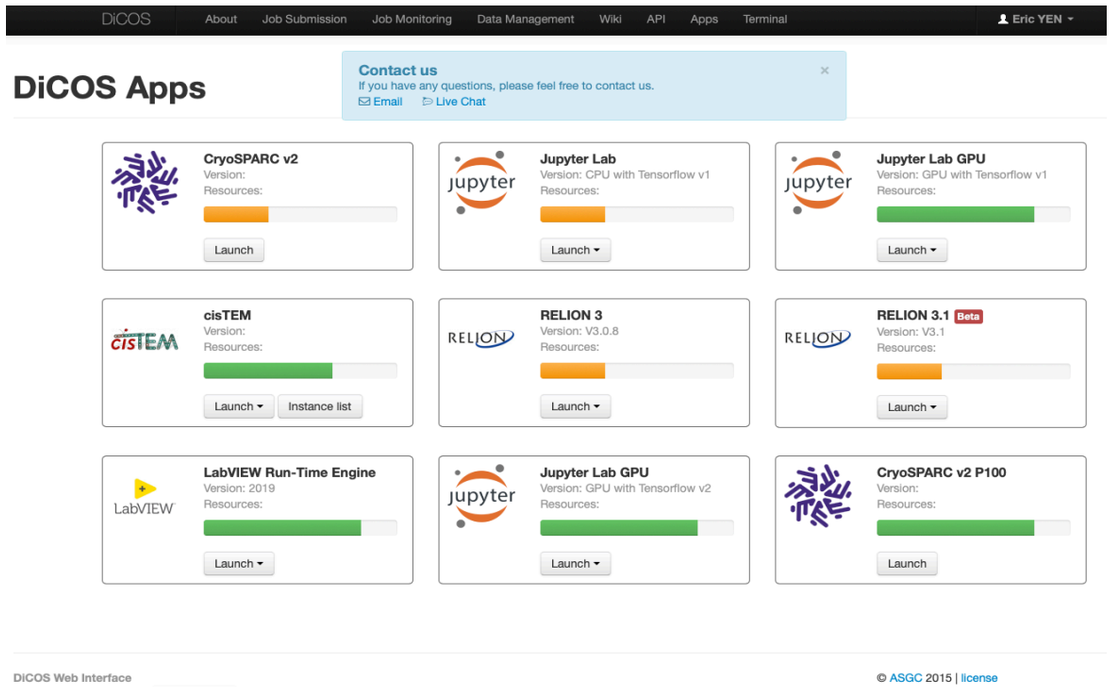
Resource Policy

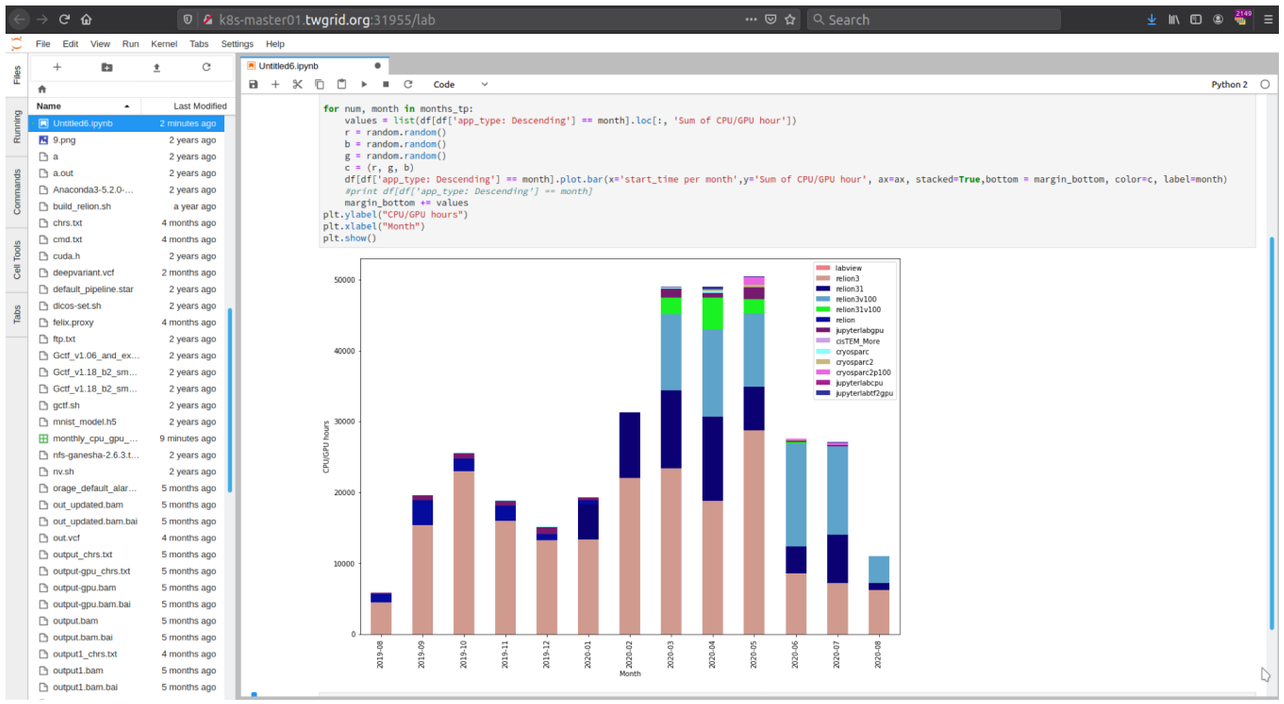
- **Type of resource**
  - Core resource: used for providing services
  - Shared resource: shared by all users
  - Priority resource: invested by user communities. Will be shared to other users whenever facility is free
- **Allocation Policy**
  - Once resource is not insufficient: 1) request funding support from AS (申請核心設施設備擴充); 2) invest by user communities
  - Any user community needs more resource: 1) <= 50% of shared resource; 2) <= 75% of total resource; 3) <= 25% growth rate
- **Rules**
  - AUP, Privacy, Security policy
  - Shared resource: fare share basis; limit of #running jobs/ PI group; limit of time (job); limit (used CPU/GPU)/user
  - Storage: quota/user; dark data purge reminder; quota review regularly (every 3 month); quota increase request
  - Reviewed by User Committee

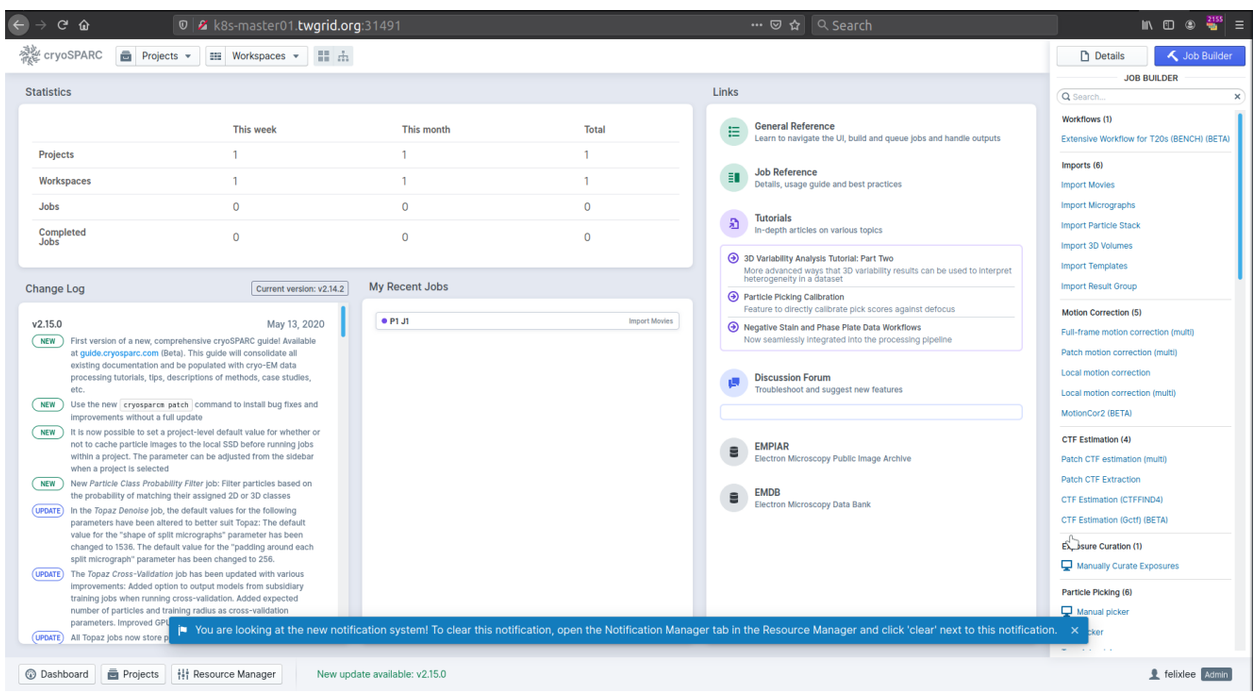


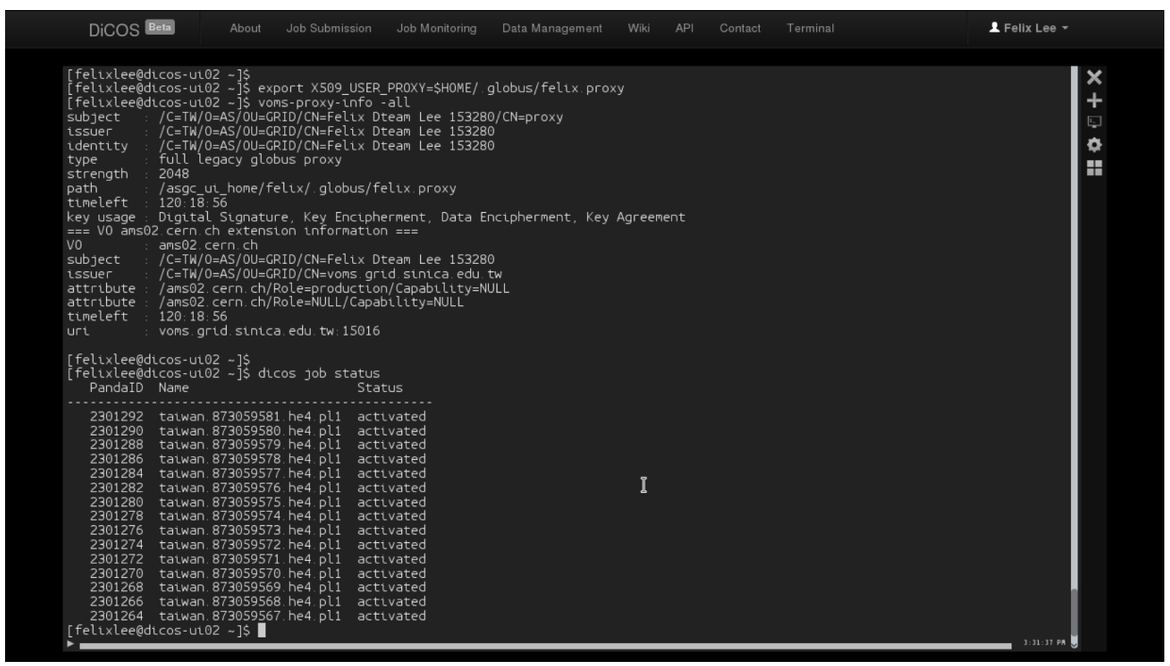
Interactive UI

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









ASGC Federated Resource

GPU Cluster	GPU	#Nodes	GPU RAM (GB)	CPU	Sys RAM (GB)	Inter-Connection
IOC-GPU-A100	A100*4/node	2	80	AMD		Eth: 100Gbps Shall be online in late 2022
GPU-A100	A100*8/node	2	80	AMD EPYC 7302 16C @3GHz*2	1024	IB: EDR; Eth: 100Gbps
GPU-V100	V100*8/node	6	24	Xeon 6126 12C @2.6GHz*2	768	IB: EDR; Eth: 10Gbps
GPU-P100	P100*4/node	2	11	Xeon E5-2650L v4 8C@2.2G1GHz *2	503	IB: FDR; Eth: 10Gbps
GPU-3090	3090*8/node	4+4*4	11	AMD EPYC 7302 16C @3GHz*2	1024	IB: FDR; Eth: 10Gbps
GPU-1080	1080*8/node	8	11	Xeon E5-2650L v4 8C@2.2G1GHz *2	768	IB: QDR; Eth: 10Gbps

Cluster	CPU	#Nodes	#Cores	RAM(GB)/Core	Inter-connection	Remark
IOC	AMD 7H12	12	768		Eth: 100Gbps	Shall be online in late 2022
ASIAA	AMD Rome 7763 @2.5GHz *2/node	32	1,792		Eth: 10Gbps	Online in May 2022
AMD	AMD Rome 7662 @2.0GHz *2/node	6	768	6.0	IB: EDR; Eth: 100Gbps	Online in May 2022
FDR5	Intel Xeon E5-2650L v4@2.2GHz *2/	92	2,208	5.2	IB: FDR; Eth: 10Gbps	50 TFLOPS;
FDR4	Intel Xeon E5-2650L v3@1.8GHz *2/	39	936	3.9	IB: FDR; Eth: 10Gbps	11 TFLOPS;
QDR4	Intel Xeon E5-2650L v2@1.7GHz *2/	50	1,000	6.3	IB: QDR; Eth: 10Gbps	6 TFLOPS;
QDR2	Intel Xeon E5-2650L @1.8GHz *2/node	96	1,536	3.9	IB: QDR; Eth: 10Gbps	
FDR1&2	Intel Xeon E5-2650L @1.8GHz *2/node	50	1,200	3.9	Eth: 10Gbps	WLCG T1;
WLCG	Quad-Core Xeon L5408/L5500/E5-2609	360	3,172	2.0	Eth: 10Gbps	WLCG T1/T2;

ASGC Resource + IAA & IOC HW

Disk 儲存資源	Raw Space (TB)	Usable Space (TB)	說明
WLCG 全球網格 T1 & T2	13,080	9,810	2019/2020 購置。9PB 為 WLCG 2021 承諾額度。720TB 為 EOS 轉換用
科學計算+1000 (CryoEM)	8,628	6,471	提供使用者登入、計算、儲存與備份等空間
Cloud Storage	2,370	1,777.5	提供核心服務、公共服務與測試環境
總計	24,078	18,058.5	

Summary

- Collaboration model with ASGC
  - Upkeep of scientific computing and big data analysis system
  - Workflow integration, customization and efficiency improvement
  - Resource federation for extension of the research infrastructure
    - Facility owner has priority usage but resource has to share with AS users
    - Bring your own hardware or investment, etc.
  - Distributed cloud operation
    - Capacity building: consulting, training, workshop, hackathon, etc.
- ASGC is turning requests into services
- ML-Cloud services
- ASGC ongoing operations are funded by AS and collaborators.
  - Cost sharing model: pay-as-you-go
- Test of the most advancement AMD/NVIDIA GPU (MI250x4 and Hopper)
  - AMD provides trail accounts for AS during 13-27 May. Longer term of trial might be possible if more demands call by end of April

ASGC Services

- **ASGC Web Site: https://www.twgrid.org**
- **Access to ASGC Resources**

- **Contact point: DiCOS-Support@twgrid.org**