



# Local Engagement of Network & Grid Computing in Taiwan



Yuan CHAO, Eric Yen  
(National Taiwan University, Taipei, Taiwan; ASGC)

AFAD 2024  
2024/04/17



# Engaged in Collider Physics

- Major research institutes on collider physics:
  - Academia Sinica,  
National Taiwan Univ. (Taipei)
  - National Center Univ. (Chung Li)
  - National Tsing-Tung Univ. (Hsin Chu)
  - National Chen-Kong Univ. (Tainan)
- Collaborations:
  - Atlas, AMS
  - CMS
  - Belle2
  - STAR
  - sPHENIX
  - EIC
  - ...
- With duplicated efforts and work





教育部

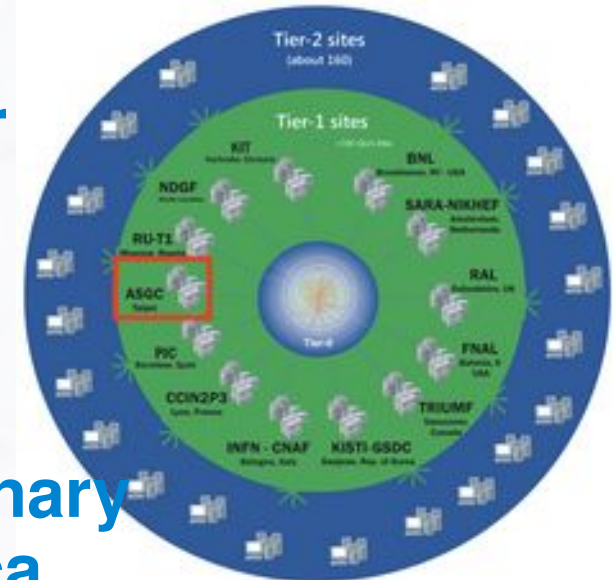


- **Taiwan Instrumentation Detector Consortium (TIDC)**
  - **Initiated in 2019**
  - **Approved by MOST(NSTC) in 2022**
- **Mission – producing novel devices, electronics, and related software for basics physics, interdisciplinary, as well as industrial applications**
  - **Cross-institutional and interdisciplinary**
  - **Training and education for younger generations**
- **Workshop, annual meetings and events: from 2021**
  - **10<sup>th</sup> anniversary of the Higgs boson discovery**
  - **Summer and Autumn School in High Energy Physics / electron-ion collider**



# Academia Sinica Grid Center

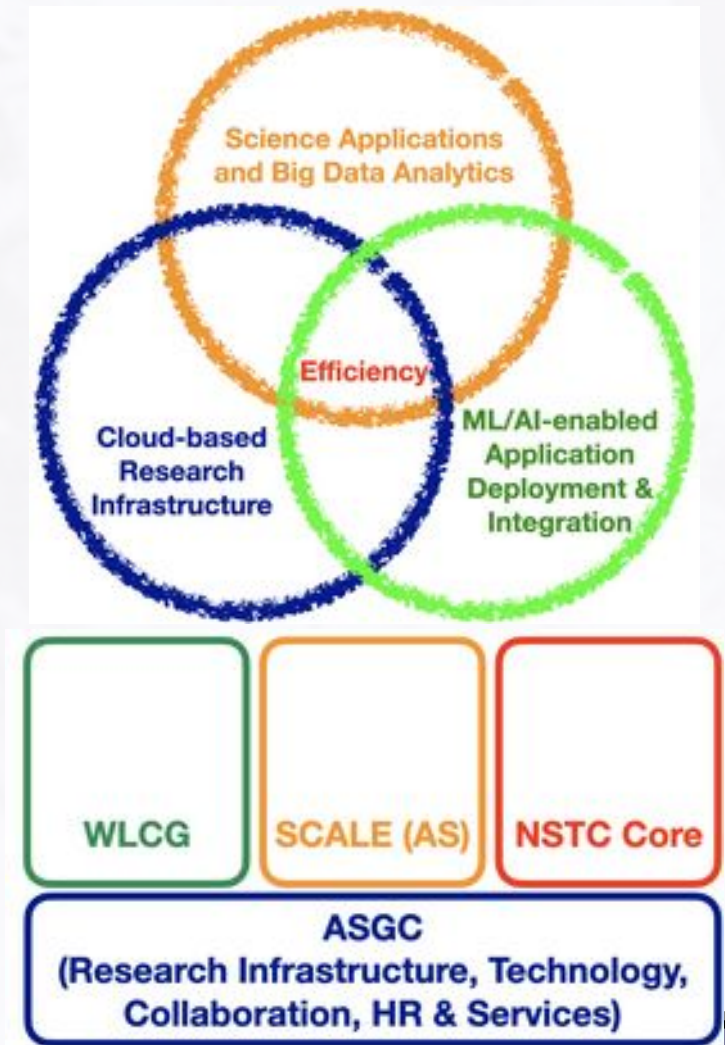
- **ASGC joined WLCG development and deployment for the Large Hadron Collider grand challenges since 2001**
  - **ASGC operates T1 and WLCG Asian Regional Operation Center has been operational from 2005**
- **ASGC has been supporting multi-disciplinary e-Science applications of Academia Sinica from 2006, based on WLCG core technologies**
  - **Building integrated research infrastructure and connecting high-performance / throughput computing facility with data-intensive applications.**
  - **Research infrastructure, platform and services are improved progressively with growing scientific applications of various disciplines.**





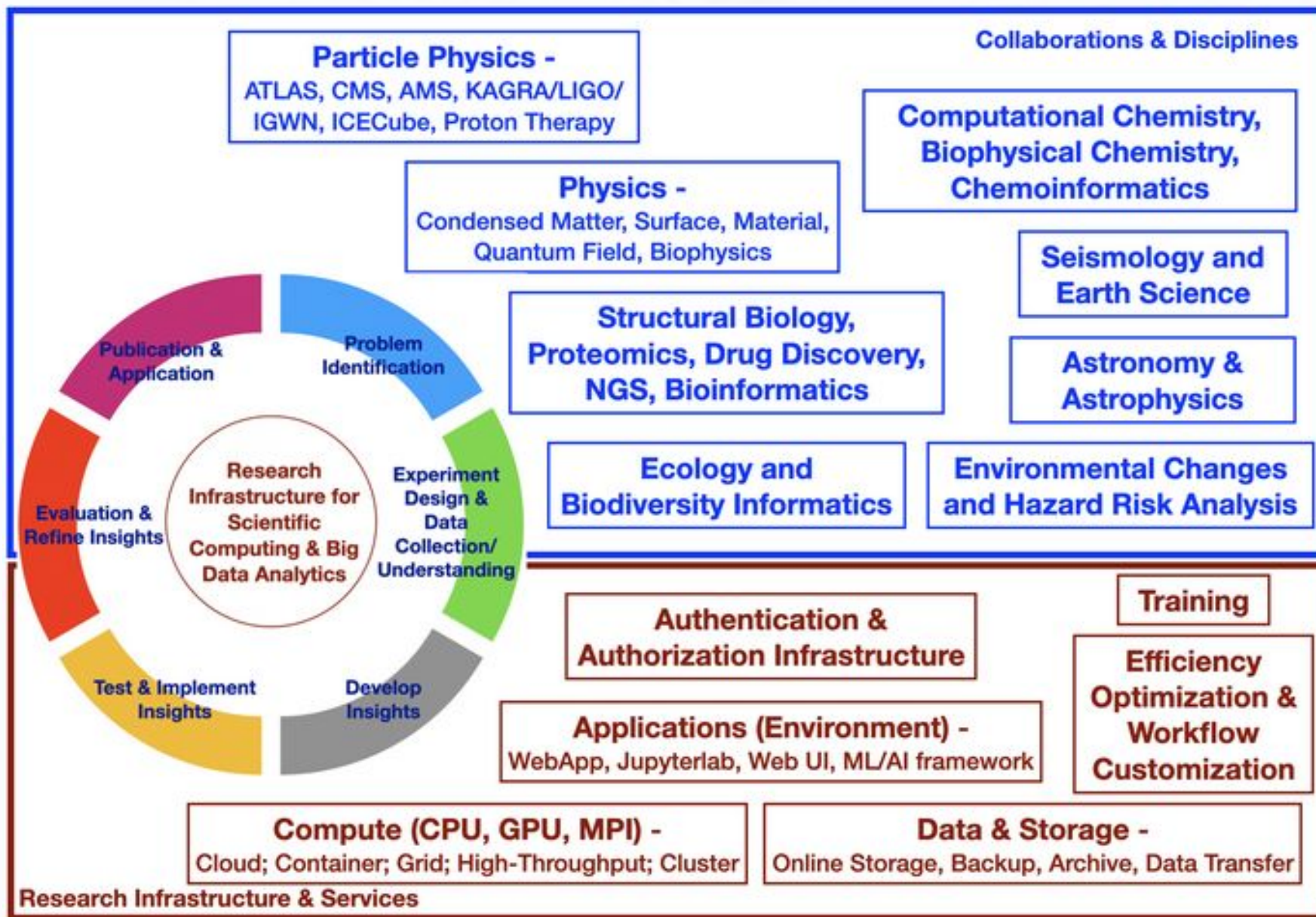
# Academia Sinica Grid Center

- ...
- **System efficiency optimization** (including power, thermal, system and applications, etc.) is also a strategic goal of ASGC aided by machine learning technologies
- **ASGC becomes the Core Facility for big data and scientific computing of AS and Taiwan from 2023 founded by AS and NSTC**





# Research oriented workflow





# WLCG Tier-2 @ASGC

- **Computing Resource – retirement of legacy hardware for energy saving (T2\_TW\_FTT)**
  - **Federated Taiwan T2 (2024):**
    - 17500 HEPSpec06 (1864 CPU cores)
    - GPU would be available after validation of new computing models (Atlas, CMS)
- **Storage Resource (2024): 5PB**
  - EOS available for ATLAS
- **Activities for ATLAS**
  - Finished 1300 billion jobs, 350PB in 2010-2022 (# of processed data and MC)
  - To join USATLAS activities
  - Development of High Granularity Timing Detector (HGTD) DB and backup system
  - Support Folding@Home for COVID-19 studies
- **Networking – connecting to BNL T1 directly**
- **Contribution to ATLAS Software and Computing**



# CMS T3

- Operated by ASGC since 2022 (T3\_TW\_TIDC)
- In collaboration with TIDC and local CMS groups (NTU, NCU)
- Analysis facility
  - Both Condor/UI (internal) and CRAB/ARC-CE available
    - CephFS shared file-system: 3TB/group by default
    - EOS by XROOTD and fuse: 1PB
  - Condor cluster
    - 768 cores (AMD EPYC 7713) + 768 cores (Intel E5-2650 v4)
  - Supporting user's access to CMS data or CRAB jobs
  - JupyterLab
- User training and support are also provided
  - E.g., using Coffea with template for columnar analysis



Coffea = Column Object Framework For Effective Analysis,  
<https://github.com/CoffeaTeam/coffea>





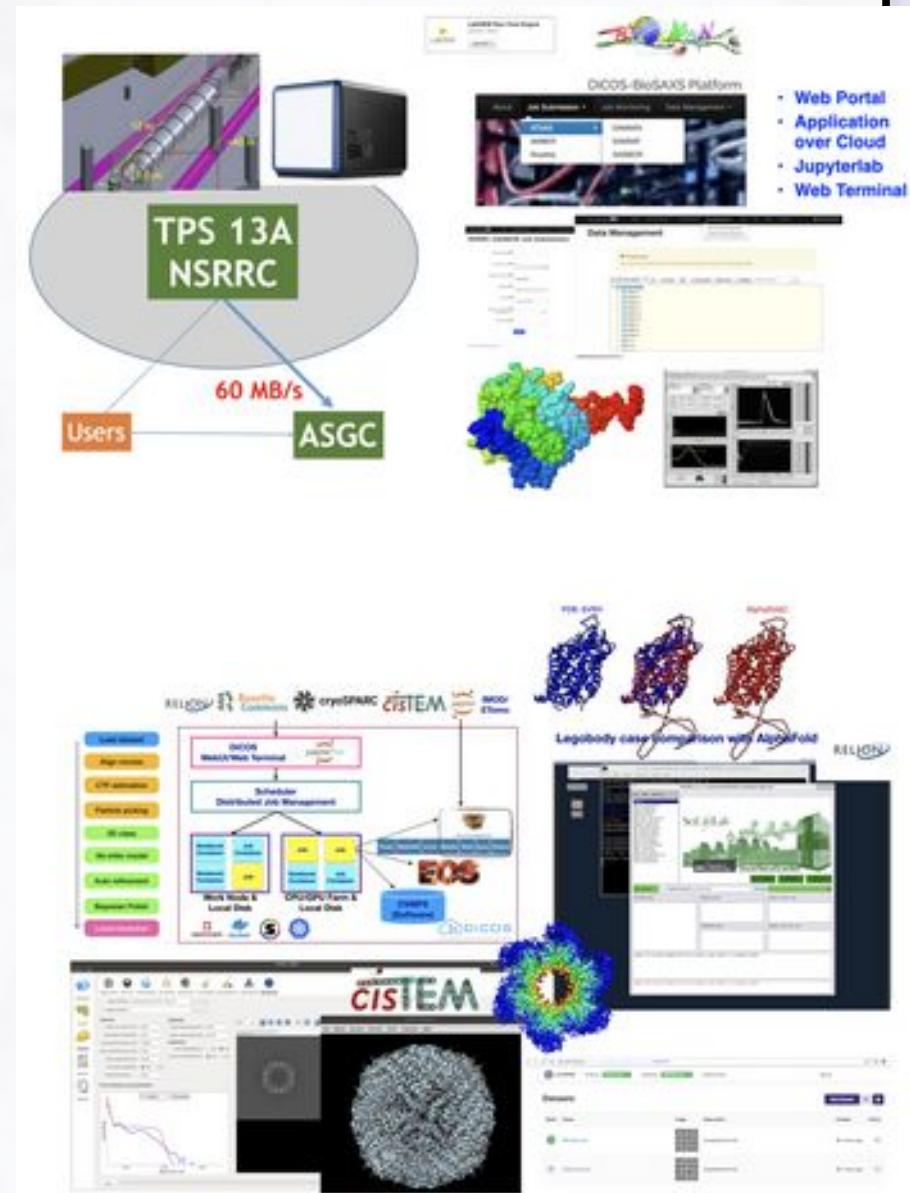
# Cross-facility Data Processing

- **Customized automated data analysis pipeline**
  - **Cloud service: better scalability, efficiency, performance**
  - **Integration & Optimization: data pipeline and workflow (compting model)**
    - **Reduced latency between experiment facility (source) and data analysis facility**
    - **Integrating required software, application framework, storage and analysis workflow**
  - **Web service: developing web app, web portal or science gateway**
  - **Generalization and new service creation**



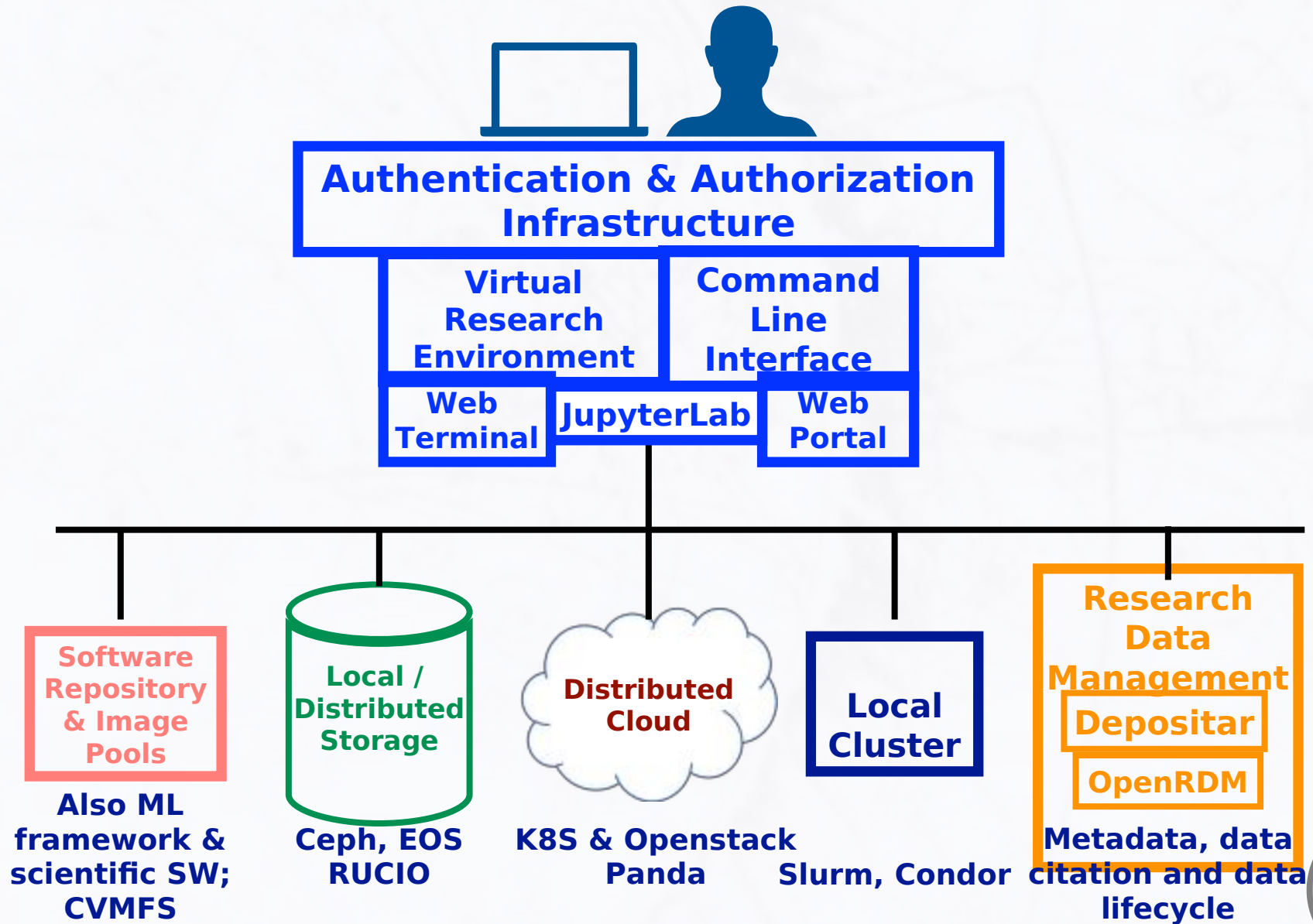
# Cross-facility Data Processing

- Ex. Structural biology: NSRRC (BioSAXS) and Taiwan Protein Project (CryoEM core facility)
- ASGC Science Cloud
  - Application platform + Computing infrastructure + workflow integration + efficiency optimization
  - AlphaFold, RosettaFold, RosettaFold Diffusion, DiffDock



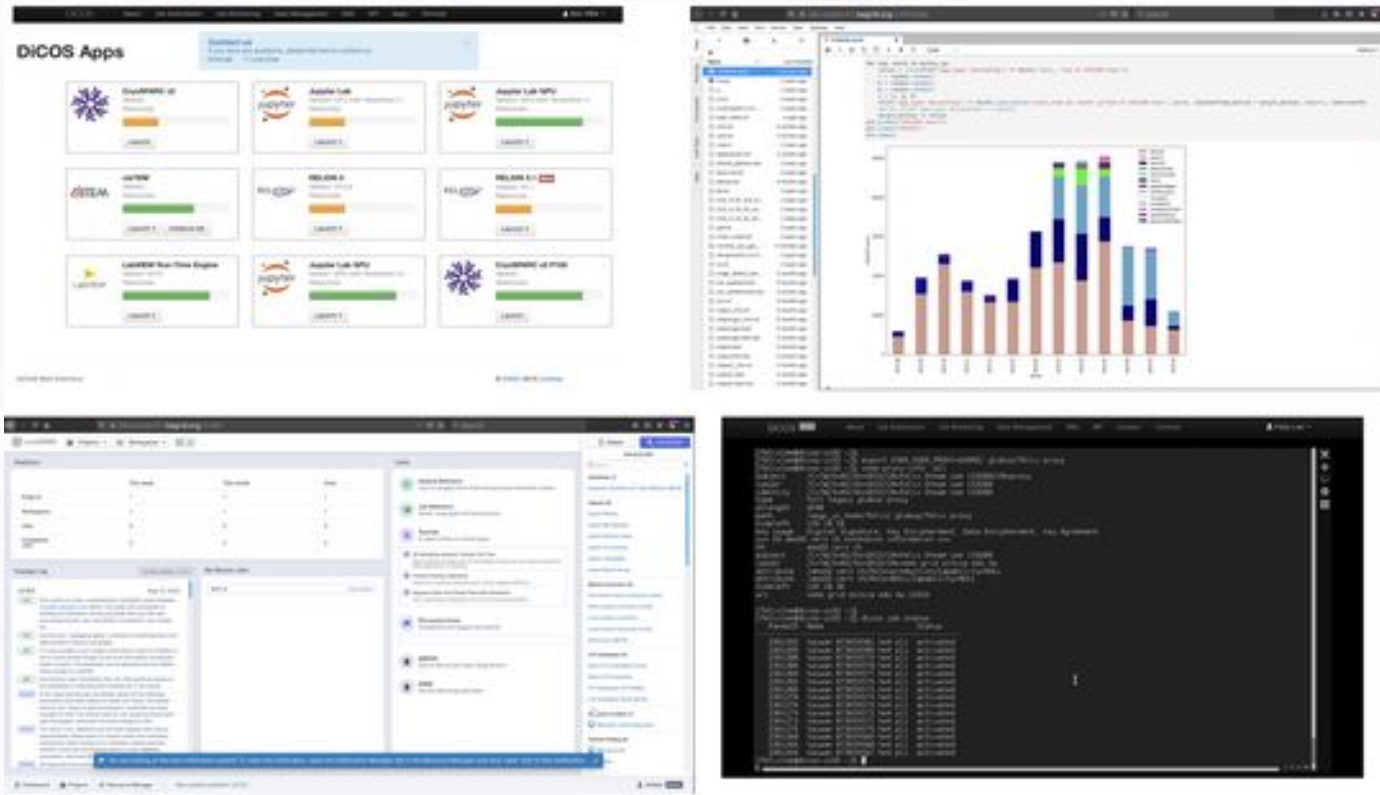


# Collaboration Infrastructure





# Supporting Big Data and AI



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

Application-specific/  
Generic Learning Engines    learn    ROOT  
Data Analysis Framework

Deep Learning  
Engines/Frameworks    NVIDIA  
TRITON INFERENCE SERVER    PyTorch    TensorFlow    Keras

Computing Resource  
(Cloud/Grid/Slurm)

Storage Resource  
(Ceph/EOS)

Distributed Data Management  
& Cloud Storage Services

Network & Data  
Transmission Services



# 60+ Web Application Provided

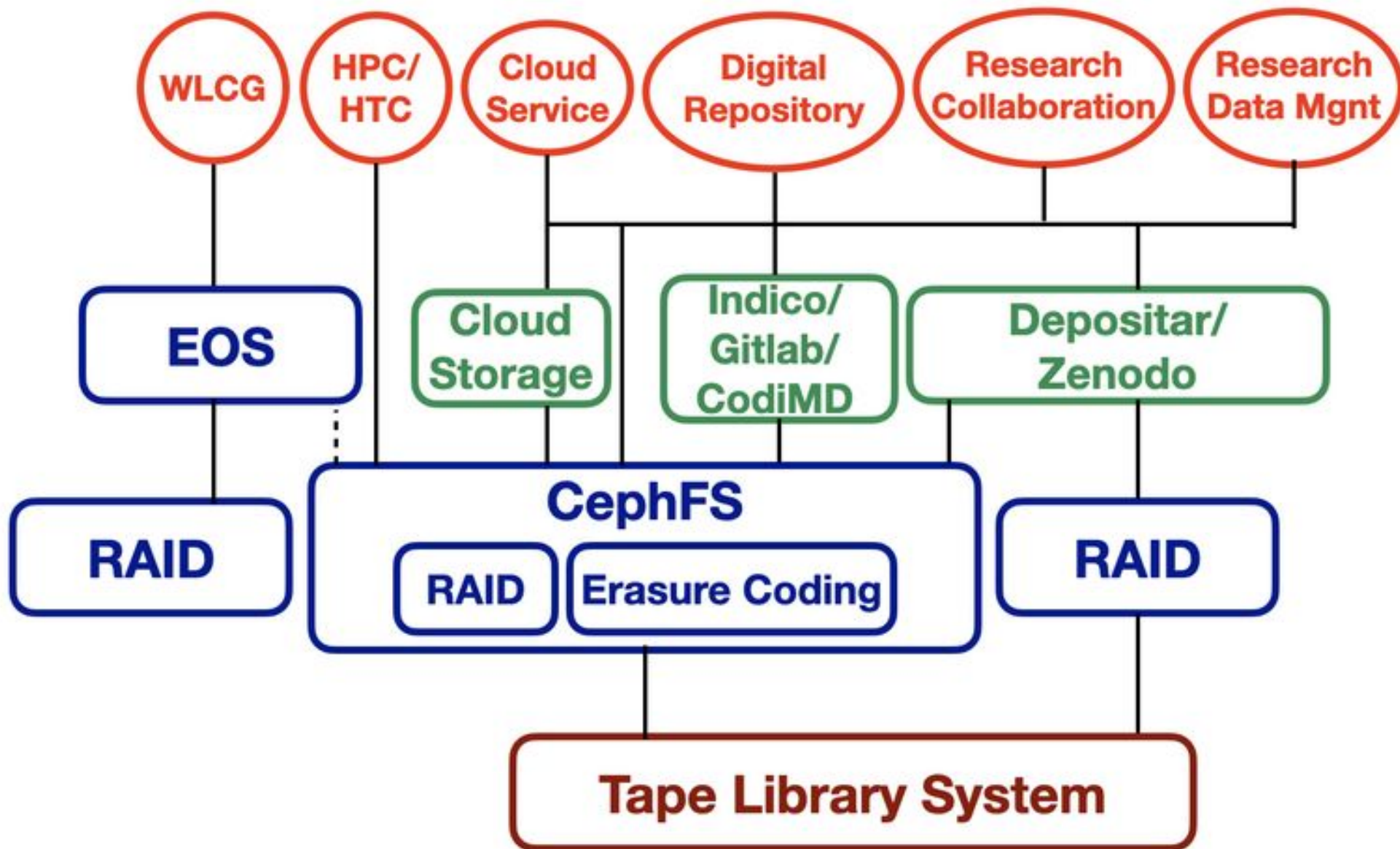
The image displays a grid of 60+ web application launchers. Each launcher card includes a logo, the application name, version number, resource requirements (CPU, GPU, RAM), and a 'Launch' button. The applications are organized into several rows and columns. Some applications have progress bars indicating their status. The applications include:

- CryoSPARC 3.2 P100, CryoSPARC 1000i, CryoSPARC RTX3090 (multiple instances)
- AlphaFold, AlphaFold (Full DB), IMOD, Triton, AlphaFold (GPU with RTX3090), AlphaFold (GPU with A100)
- RosettaFold, DynaMO, MATLAB, Jupyter Lab (multiple instances)
- RFDIFFUSION, dRDock, EvoDiff, Jupyter Lab GPU V100, Jupyter Lab GPU A100, Jupyter Lab Cryovue GPU
- QIME2, Sculptor2, Phenix, Jupyter Lab GPU A100
- MGX MorphoGraphX, Deepnd-kit (multiple instances), MAME, MAME Run-Time Engine
- MAME, Pyserver, Pyserver Client
- PyRoot, qikit

The screenshot shows the DICOS-BioSAXS Platform interface. It features a navigation menu with the following options: About, Job Submission, Job Monitoring, and Data Management. Below the menu, there are several application icons: ATLAS, AMBER, Rosetta, DAMMIN, DAMIF, and GASSOR. The background of the interface shows a molecular structure visualization.



# ASGC Storage System





# ML-enabled Analytics by ASGC

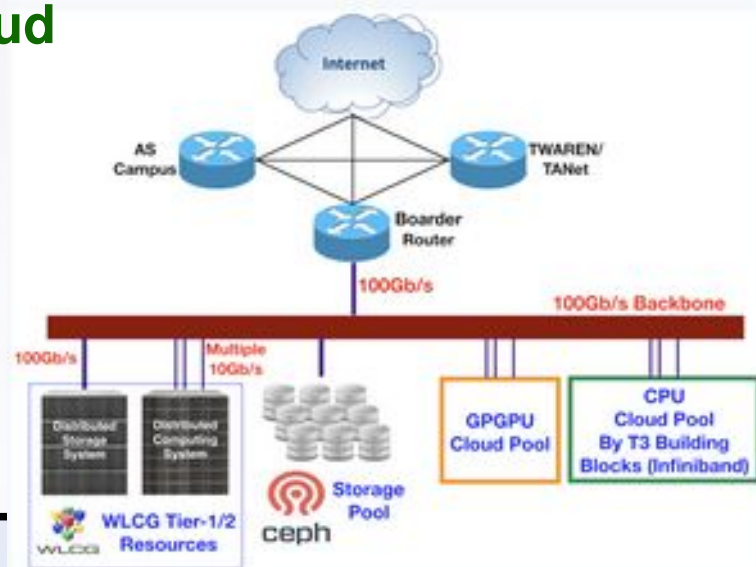
- **ML/AI application platform service is available NOW - SW library, HW integration and application**
  - Build up customized ML platforms for user specific projects – Deploy **ML packages ready** environment in order to help ML development smoothly and provide on-demand compt. power
  - Upkeep the application framework
  - Workflow and data pipeline integration
  - Efficiency improvements
- **Potential use cases**
  - Users who bring exists source code – ASGC could help to setup a virtual environment and confirm code running normally
- **Approaches**
  - Supporting Kubernetes / Jupyter lab for development purpose
    - Create env. with user specific ML package ready
    - Support on-demand scalable CPU / GPU computing power
  - Supporting containerized env. For dev. Purpose
    - Create takeout images in docker format as an option for user
    - Docker images could be d/l from ASGC and deployed locally



# Science Cloud Infrastructure DiCOs

- **OpenStack cloud: for core services and on-demand worker nodes maintained by openstack-ansible**
  - **Multiple cells / region for various configurations and capabilities**
    - e. g., GPU, Neutron Compute, Nova Compute, ...
  - **Single hypervisor type: KVM**
    - # of hypervisors: 100+
    - # of VMs: 500+, dynamic provisioning
  - **Networking: flat and segmented**
- **Containerized resources managed by Kubernetes framework – for software on-demand services and port of core services**
  - **User cluster:**
    - Batch, interactive GUI jobs; GPU cloud
    - Remote JupyterLab, virtual desktop
  - **Core services: distributed cloud cores, ElasticSearch + Kibana**
  - **High availability enabled (by HELM)**
- **UI: web UI / terminal; JupyterLab**
- **OS migration strategy**
  - **To AlmaLinux from CentOS7 in 2024**

System Architecture

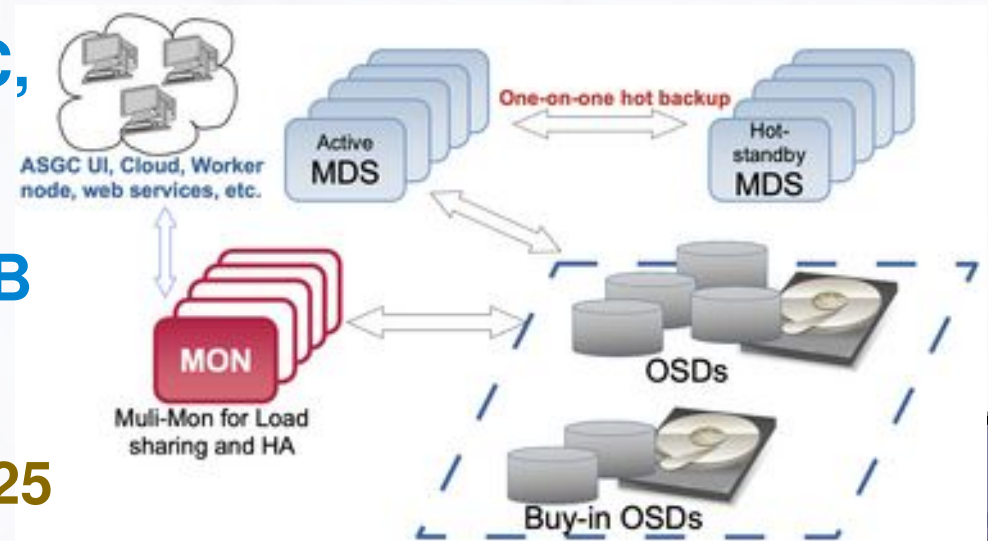






# ASGC Science Cloud Infrastructure

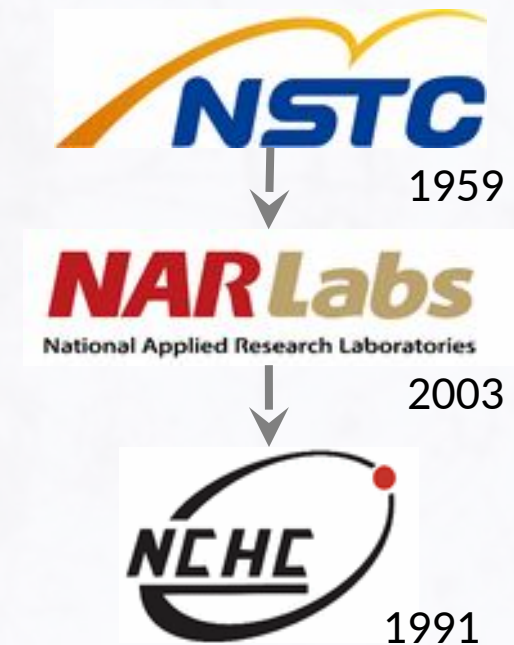
- Scalable and reliable online storage based on Ceph mainly
- Ceph Configurations: ~10PB
  - 6 MDS + 6 hot-standby (one-on-one backup); 7 MONs
  - 462 OSDs, 51 hosts
- Services
  - 3 TB per PI group setup by default
  - PI could extend the space through management UI flexible
- Reached 2GB/s Read / Write throughput so far
- Tape-based remote backup system (12PB) will be established and integrated in 24Q2.
- Providing big pool for HPC, HTC, AI and various applications concurrently
- Capacity will be growing to 13PB by the end of 2024
  - Plan to procure new 4PB disk servers for Ceph in 2024 & 2025





# About NARLabs

- GONGO R&D platform to support academic, to foster talent, to promote frontier S&T.
  - Hope to support evidence-based policy making for innovation economy.
  - Seek to respond challenges on **Sustainability** and **Humanities**.
- 7 research institutes:
  - Animal Laboratory Technology
  - Earthquake Engineering
  - Instrument Technology
  - **Network and HPC**
  - Ocean Research
  - Semiconductor Technology
  - S&T Policy Research





# NCHC milestones



250+ employees, ~25% PhD

- ✓ ISO 9001+
- ✓ ISO 20000
- ✓ ISO 27001/27701
- ✓ ISO 27017/8
- ✓ ISO 500000
- ✓ DCOS 2021



Hsinchu HQ (1993)



Taichung (2008)



Tainan (2005)



Tainan IDC (~2025)



# Taiwania series

- 30k+ users. Usage from
  - ~90% Academia
  - ~10% Gov
  - <1% Industry
- ~2k projects per year
  - ~ 50 % NSTC, ~ 20% Gov,
  - ~ 20 % Industry, ~ 10% Others
- Usage by NSTC:
  - ~< 70 % Math/Phy/Chem
  - ~< 20 % Engineering
  - ~ 10 % Biomedic
  - < 4 % Others
- Varies by machine but overall trend holds

Model	Year	Computing Power	TOP500 Rank	Green500 Rank
台灣杉一號 TAIWANIA 1	2017/11	1.7 PFLOPS	#95	#31
台灣杉二號 TAIWANIA 2	2018/11	9 PFLOPS	#20	#10
台灣杉三號 TAIWANIA 3	2020/11	2.7 PFLOPS	#181	#69
Next Generation	-	3.5 PF	62k-core	5.7 GF/Watt, PUE~1.35

- Forerunner (formerly Twnia-4, CPU-based w/ASUS)
  - 3.5 PF, 62k-core, 5.7 GF/Watt, PUE~1.35
  - Top222 and Green 92 @ Nov 2023
  - Cost ~15 mn USD
  - To be commissioned in early 2024



# Taiwan Advanced Research Network

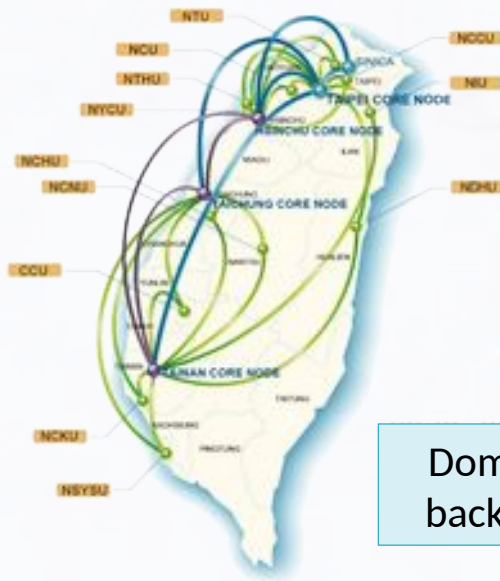
- Optical, dedicated bandwidth

- ✓ 200 Gbps
- ✓ 12 GigaPOPs
- ✓ 94 universities & research institutes
- ✓ 500 k users

- ✓ Sharing network fiber with **TANet** (4000 schools, 4.5 mn users)

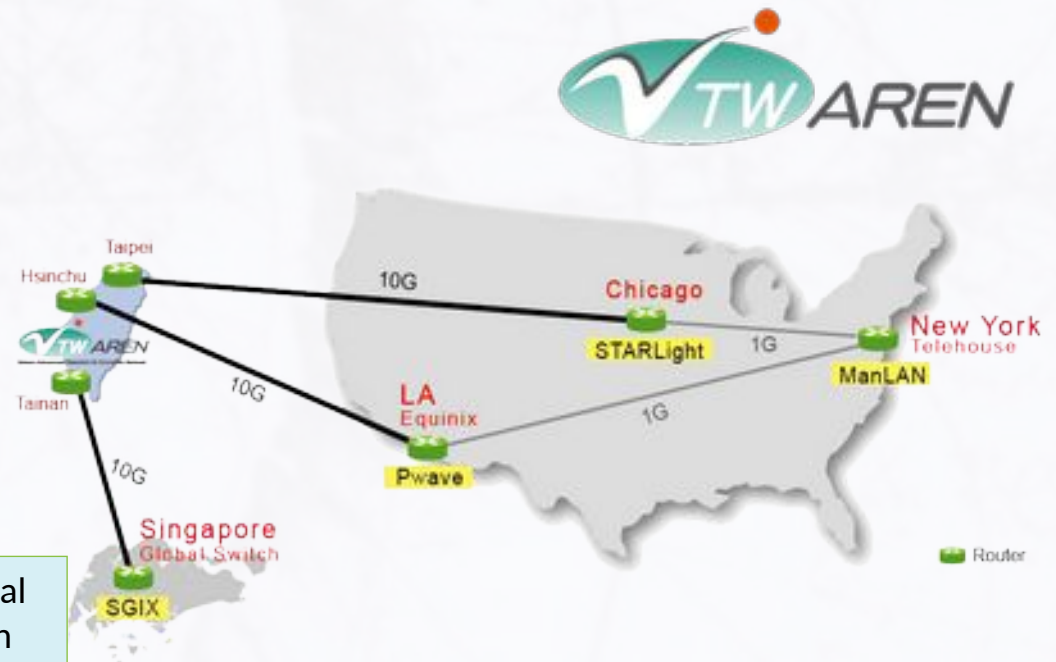
- ✓ Availability: 99.99% ↑

- ✓ 30 Gbps links to **Los Angeles, Chicago, New York, and Singapore** (thru which peers with other RENs)



Domestic backbone

International connection





# WLCG CMS Tier-2 @NCHC

- **Computing Resource**
  - CMS Taiwan T2 since 2015 (T2\_TW\_NCHC)
  - 576 CPU cores @ Tainan
  - Same scale in Hsin-Chu in 2020 (Taiwania Gen 1)
  - No UI, only accepting Grid jobs (gLite, ARC-CE+Condor)
- **Storage Resource:**
  - DPM ~0.5PB (2015)
  - Migrated to dCache ~1PB (2020)
- **OS Migration**
  - Starting with Scientific Linux CERN 6
  - Moved to SLC 7 as SLC6 support ended
  - Will move to AlmaLinux following central
- **Networking**
  - Using 100Gbe to TWAREN backbone
  - LHCONE through ASGC network
- **Contribution to CMS Software and Computing**
  - NTU, NCU



# Summary

- **Multiple Research institutes and Multi-disciplinary about accelerators and detectors in Taiwan**
- **Multiple collaborations for historical reasons**
- **Grid technologies help to support needed computing and network**
- **Thanks to ASGC and NCHC on various supports**



**以上**

**Thank YOU!**

**謝謝**

**Merci de votre  
attention!**