國科會 高效能核心科學計算中心 (NSTCCORE)

https://nstccore.twgrid.org

NSTCCore Computing Service User Training Workshop in 2026: AI Applications With AMD

嚴漢偉 Eric Yen

中央研究院 物理所 網格中心

2026. 5. 27

Services are jointly provisioned by ASGC Core Facility (SCALE, 大數據分析與科學計算核心設施)

Program Today

9:15 AM → 9:35 AM **一、NSTCCore 高效能運算服務介紹 暨 AMD 受邀技術演講引言｜Introduction to NSTCCore Computing Services and Transition to the AMD Invited Technical Talk** 1F Audi...

- 計算服務架構 Introduction of Computing Service
- 運算資源、服務內容和計價模式 Resource, Service & Pricing
- 帳號申請與2FA Account Application & 2FA

Convener: Prof. Eric YEN (ASGC)

9:35 AM → 10:25 AM **AMD 技術專題：從資料中心到客戶端：以 AMD AI 軟體、ROCm 與 NPU 部署大型語言模型（LLM）｜Invited Technical Talk: From Datacenter to Client: Deploying LLMs with AMD AI Software, ROCm, and NPUs (Presented by AMD)** 1F ...

This session explores how AMD's AI software stack enables efficient LLM deployment from Datacenter to client across CPUs, GPUs, and NPUs. We will cover model development and optimization using ROCm in the data center, followed by seamless deployment to client platforms with AMD Ryzen™ AI NPUs. Using Lemonade, an open-source, OpenAI-compatible local LLM server, attendees will learn how to deploy LLMs with minimal code changes, leveraging heterogeneous execution across GPU and NPU to achieve low-latency, power-efficient, and private on-device generative AI experiences on PCs and edge devices.

Conveners: Mr Micky CHENG (AMD) , Mr Simon CHANG (AMD)

10:25 AM → 10:40 AM **休息 Break** 15m 1F Auditorium

10:40 AM → 11:00 AM **三、虛擬化SaaS雲端計算服務 - DiCOSApp｜Virtualized SaaS Cloud Computing Service – DiCOSApp** 1F Auditorium

- 計算資源 Computing Resources
- 服務架構 Service Overview
- 軟體部署 Software Deployment

Convener: Ms Ming-Jyuan YANG (ASGC)

11:00 AM → 11:30 AM **四、SLURM 計算服務｜SLURM Computing Service** 1F Auditorium

- SLURM執行工作操作 Running Jobs with SLURM
- SLURM參數介紹 Introduction to SLURM Parameters
- 多核心程式編程及操作 Multi-Core Job Programming and Execution

Convener: Ms Yi-Ru CHEN (ASGC)

11:30 AM → 12:00 PM **五、資料服務｜Data Service** 1F Auditorium

- 資料傳輸 Data Transfer
- 資料操作 Data Management

Convener: Mr Siang-Yu YANG (ASGC)

Data Service_Storag...

12:00 PM → 12:15 PM **六、綜合討論與交流（問卷填寫）｜Discussion & Exchange (Questionnaire)** 1F Auditorium

12:15 PM → 1:30 PM **午餐 Lunch @ P7A Meeting Room** 1h 15m P7A (Institute of Physics)

1:30 PM → 3:00 PM **交流（內部會議）｜Networking Session (Closed Meeting)**

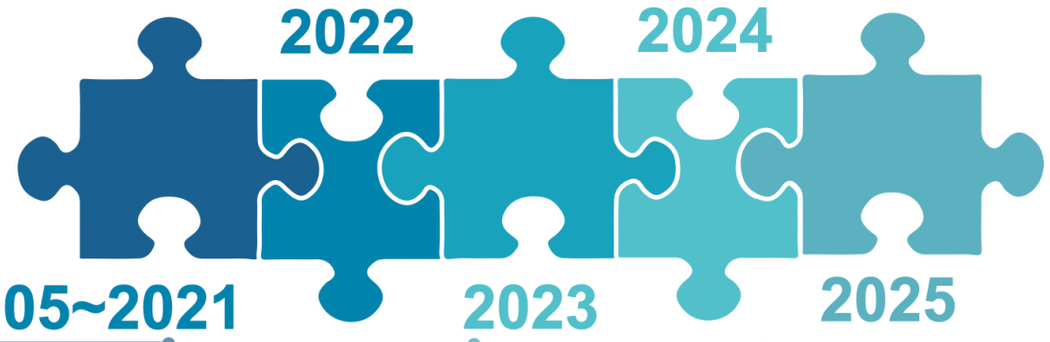
Mid-Scale Scientific Computing Services for Accelerating Discovery & Innovation

- **Funded by NSTC, June 2023 - May 2026**
  - **In association with the core facility supported by Academia Sinica**
- **Reliability & Efficiency are the top priority**
- **Serving as your computational research partner, so that scientific groups could focus on sciences**
- **Flexible and customized services: resource arrangement, workflow integration, performance improvement, etc.**
- **Resource scale: providing services utilizing O(10,000) CPUCores, O(100) GPUs and O(10)PB Disk storage resources**
  - **Able to support O(1,000) CPUCores parallel computing, multiple GPUs and big data analysis**
  - **Affordable cost and lower than operating by your own**
- **Aligning with NCHC, and enhancing the advanced computing ecosystem in Taiwan**

User-Oriented Services

24x7x365

中研院大數據分析與科學計算核心設施



2005~2021

WLCG Tier1 Centre for Global Scientific Collaboration
(CERN大強子對撞機)
(中研院物理所網格中心)

2023

國科會高效能核心科學計算資源

2025

- 2027 年前達成 10,000 CPUCores, O(10)PB storage 規模，並持續穩定運作
- 彈性的合作模式
- 協助有效利用設施資源、提升效能
- 訓練、科學計算技術推廣、支援國際合作

Research Infrastructure Keeps Advancing With Requirements of Multiple Disciplinary Scientific Communities

From 2006, we have been supporting user communities in multiple disciplines to make better use of available scientific computing resources with efficiency.

Innovation and Collaboration are keys of scientific computing

- **Getting access to and making good use of NSTCCore Services**
- **Supporting innovation with flexible collaboration**

ATLAS EXPERIMENT · CMS · AMS-02 · PHENIX · STAR · CEPC · KAGRA · LIGO · The Electron-Ion Collider

TIDC — Taiwan Instrumentation And Detector Consortium (TIDC)

TWGrid ASGC

ASCEM

CryoEM

Neutrino/MHEP	CryoEM
Quantum Materials Physics	Bioimaging
Physics of Active & Living Matter	Drug Discovery
Astrophysics	Computational Chemistry
Earth Science	Biodiversity & Ecology

Computing & Storage Resources

CPU Cluster	Spec	CPU Cores	RAM(GB) / Core	Nodes	Performance (TFLOP/Core)	Network	Onboard Date	$ /Core-Day
EDR2	AMD Genoa 9645	1,920	8	10		100Gb/HDR	Mar 2026	1.2
EDR1	AMD Genoa 9654	3,840	8	20	1.1	100Gb	Mar 2024, Dec 2024	1.2
Intel-G4	Intel Xeon 6448H	512	8	4	1	100Gb	Sep 2024	1.4

GPU Cluster	Spec	GPU	Architecture	Nodes	CPU Cores	Server RAM (GB)	CPU	Network	Onboard Date	$/ GPU-Day
B6000	RTX 6000 96GB	16	Blackwell	2	32	1,536	Intel 6517Px2	200Gb	Mar 2026	127
L40S	L40S 48GB	4	Ada Lovelace	1	32	768	AMD 9374x1	10Gb	Dec 2024	70
4090	RTX 4090 24GB	16	Ada Lovelace	2	40	1536	Intel 4416x2	10Gb	Jul 2024	60
3090	RTX 3090 24GB	56	Ampere	7	32	1024	AMD 7302x2	100Gb	Dec 2020	40
A100	A100 NVL 80GB	24	Ampere	3	64	512	AMD 7542x2	100Gb, NVL	Jun 2020	120
V100	Tesla V100 32GB	48	Volta	6	24	768	Intel 6126x2	10Gb, NVL	Dec 2018	35
P100	Tesla P100 16GB	16	Pascal	4	24	128	Intel 2650x2	10Gb, IB-FDR	Dec 2017	8

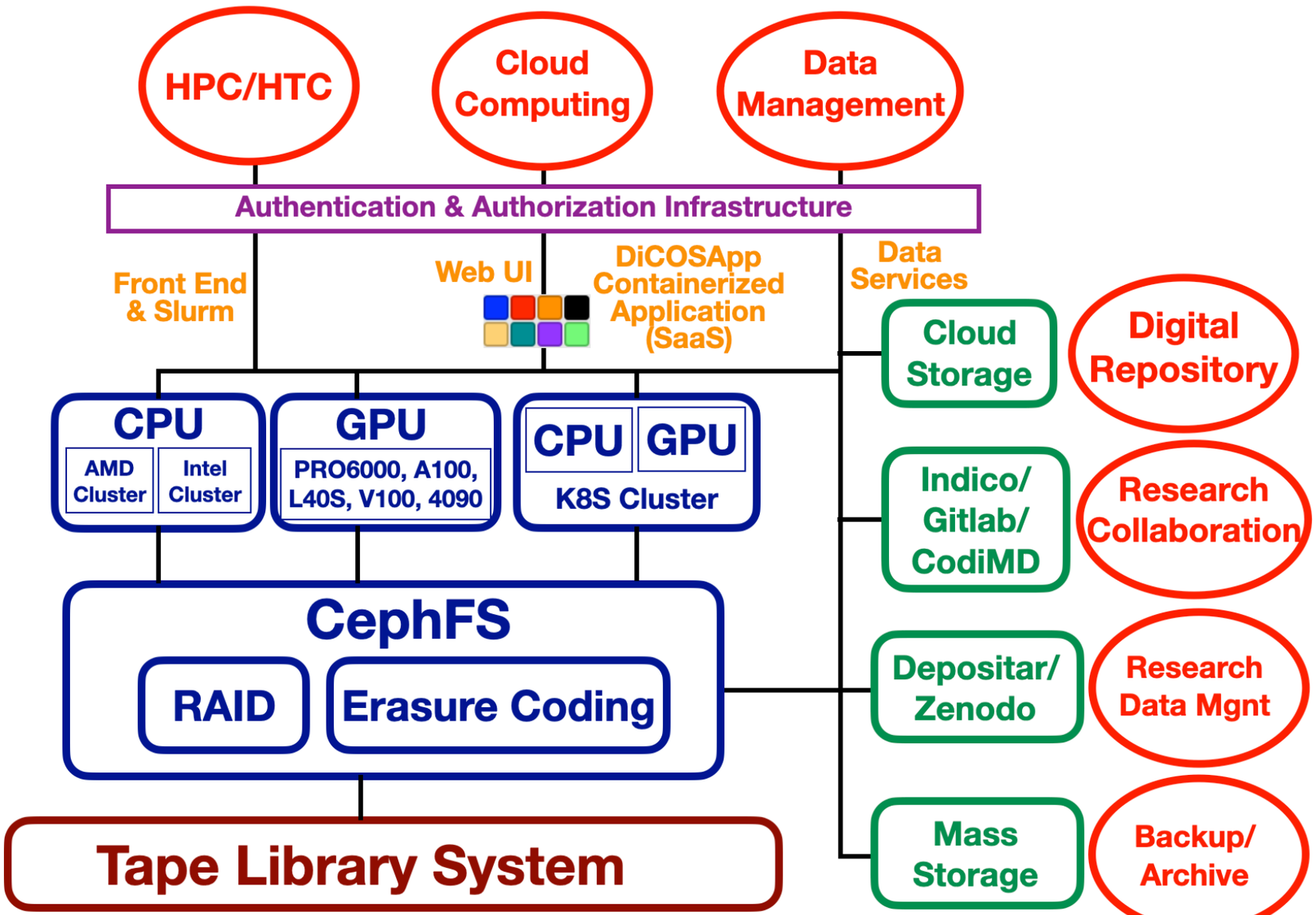
Storage	Capacity (TB)	System	Price	Remarks
Ceph File System	12,000	Online storage managed by Ceph distributed file system, with (8, 3) erasure coding	1,000 per TB-year	+3PB by end of Mar 2026
Tape Lib System	10,000	LTO9 (18TB/Tape)	300 per TB-year	Service from 2025

Scientist Could Focus on Scientific Questions

- Utilizing high-performance computing service by affordable cost 筆電價格即可使用整年高規格計算資源
  - 100CPUCore-Yr < NTD$51,100
  - 1TB-Yr Disk = NTD$1,000
  - 1xA100 (80GB RAM) GPU-Yr = NTD$ 43,800
- Free quota for new every new user: NT$ 200
  - 新使用者可有 NT$ 200 免付費測試使用額度
- 若有特殊或急迫需求，請逕洽本團隊
- Storage System：
  - Free: 3TB shared space/group, 100GB/ user
  - Tape Storage: NT$300/TB-Year
- Welcomed to share your requirements and questions

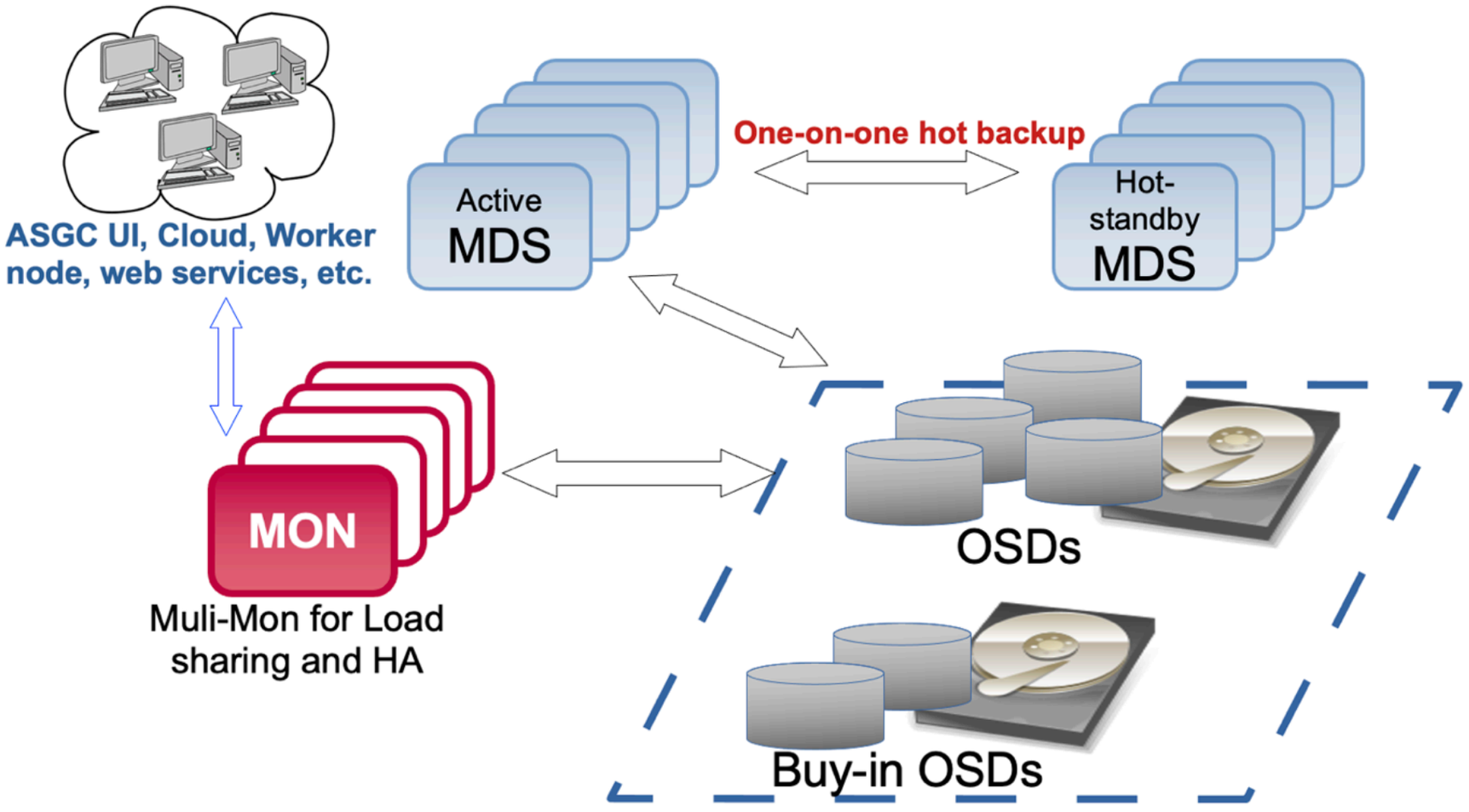
ASGC Service Architecture - Scalable Science Cloud

- Integrating data, algorithm and computing
- Scientific workflow implementation with integrated compute, storage and network resources
- Persistent services: Analysis facility, Web portals
- New resource and new services online yearly, based on users requirements



Data & Storage Services - CephFS

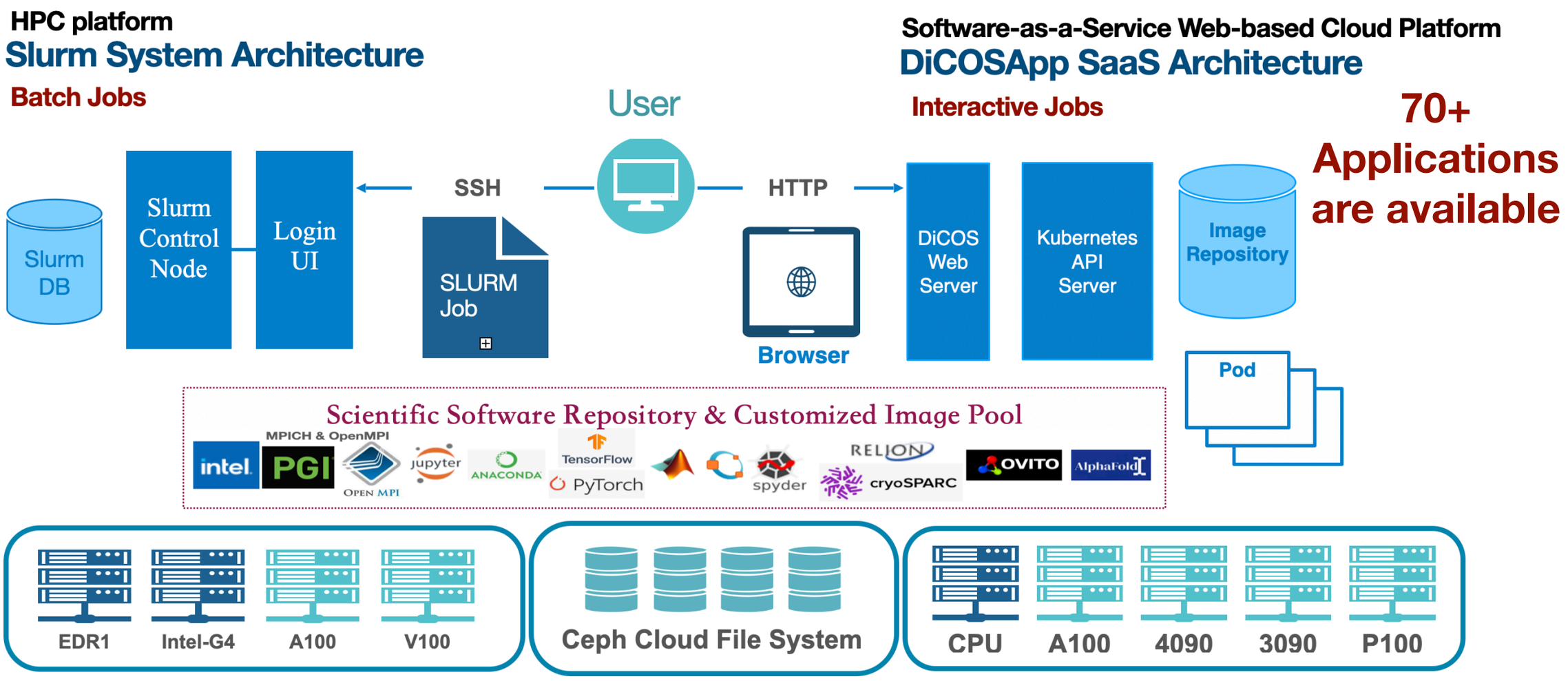
- Ceph distributed storage system - O(10)PB scalable data pool
- High performance - 4.2GB/s throughput, 1.5K op/s rd, 2K op/s wr
- Reliability - Erasure coding (8, 3), with 462 OSDs, 51 hosts and multiple MDS & MONs
- Scalability - from O(10)TB in 2012 to 17PB in 2025
- Data transmission and local backup supported
- New Services in 2026: data transfer, cloud storage and tape backup



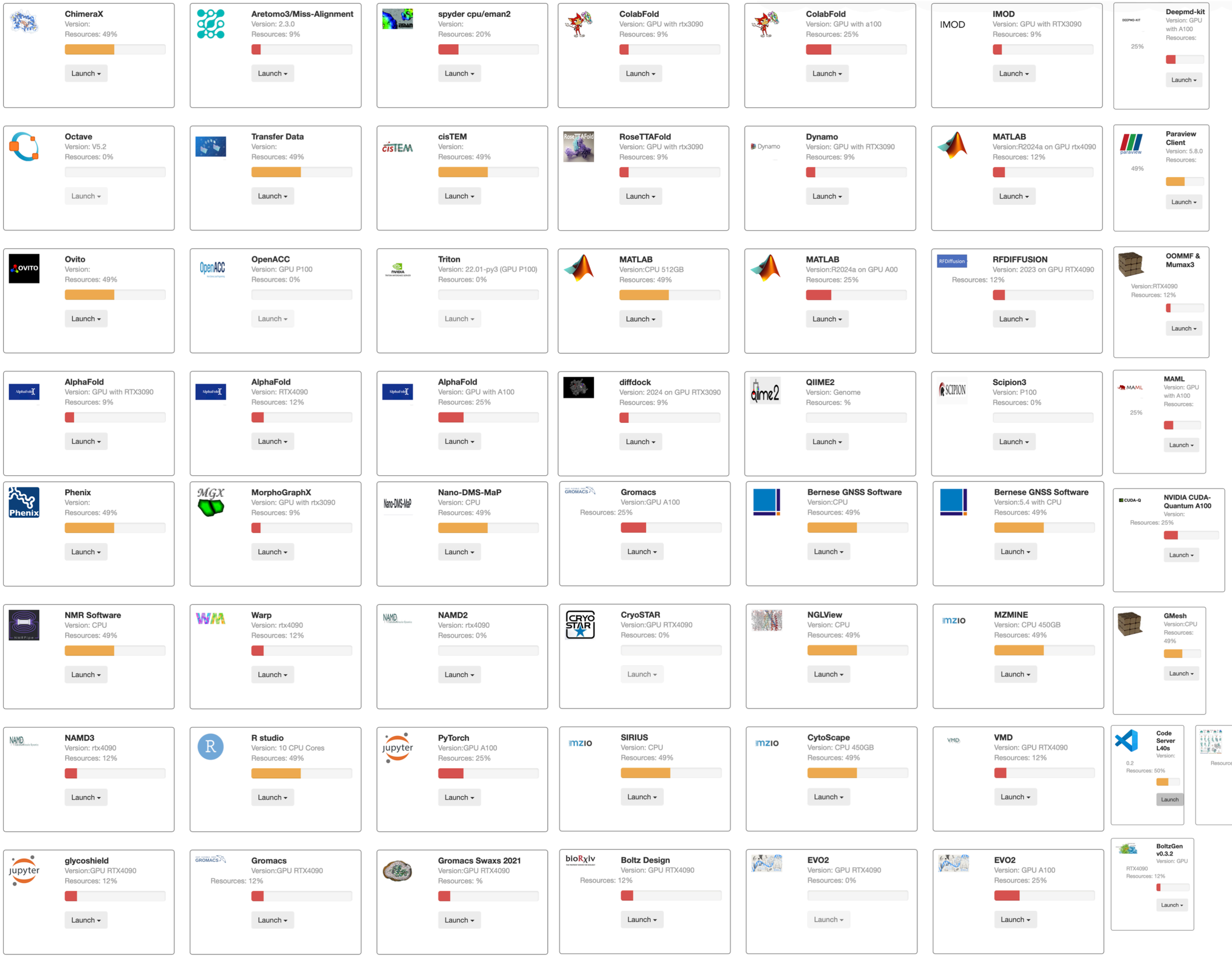
Computing Services:
Optimization of Application and System Efficiency

- HTC/HPC & Cloud Computing Platform
- Integration of computing model and software

- Research data workflow and management
- Information security
- Application efficiency
- Customized IaaS, Paas, and SaaS



70+ Web Applications Provided (Software-As-a-Service)



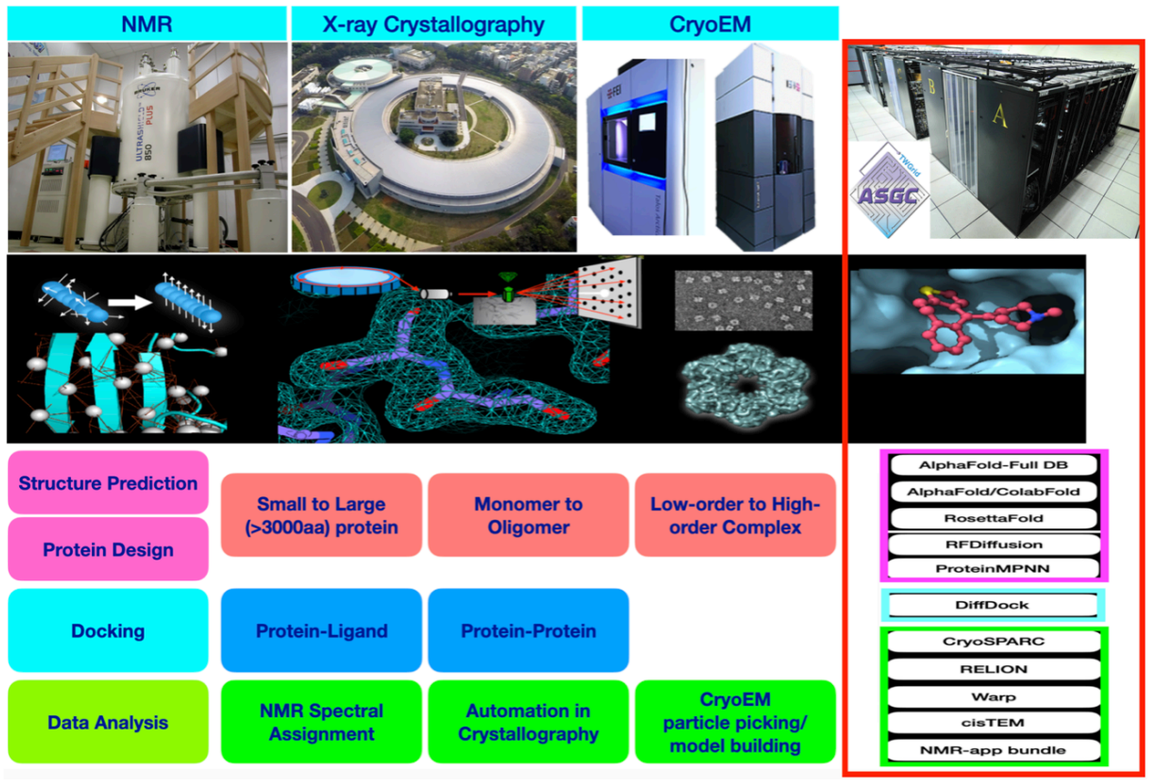
Flexible Collaboration Model: User-Driven, Service Oriented

- Pay-as-you-go
- Resource Integration: Collocation & Buy-In
- Service Hosting
- Reservation
- Experiment & Collaboration support

- Advanced Service: Web Application and portal, efficiency tuning, workflow optimization, customization, hands-on training
- User Engagement：weekly User Meeting, User Committee, dissemination & outreach

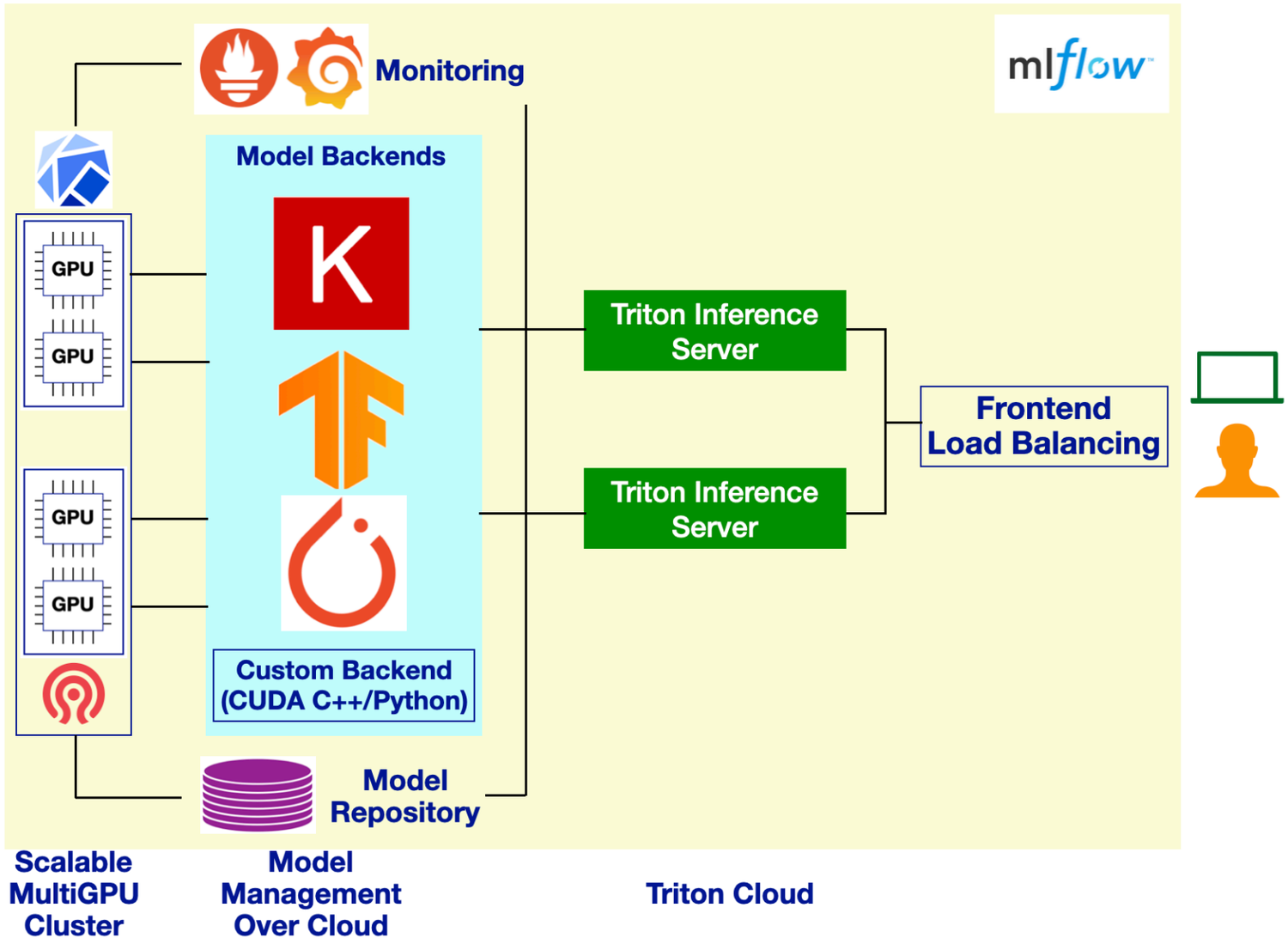
ATLAS EXPERIMENT · CMS · AMS-02 · PHENIX · STAR · CEPC · KAGRA · LIGO · The Electron-Ion Collider

TIDC — Taiwan Instrumentation And Detector Consortium (TIDC)

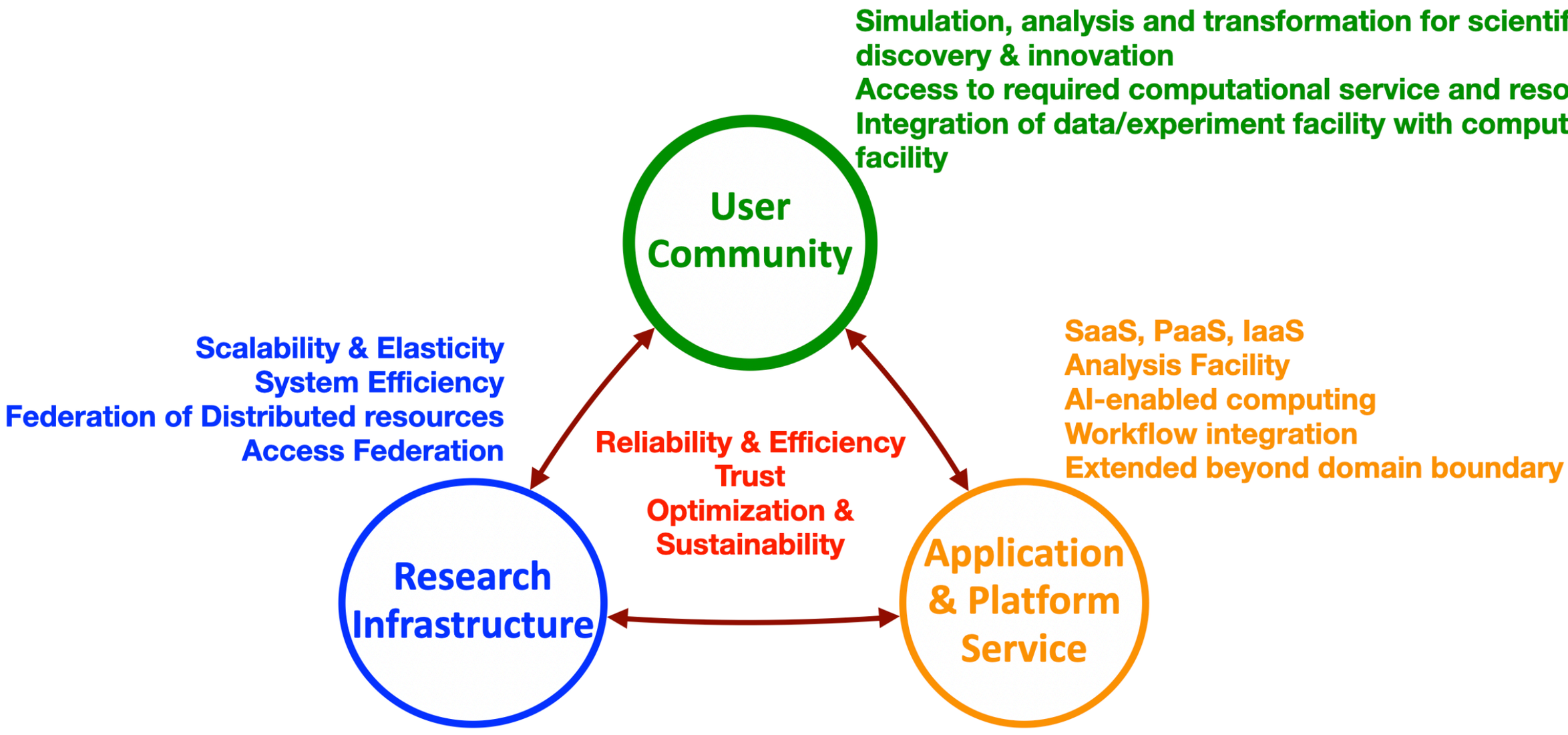


Services of ML-Enabled Data Analytics

- **ML/AI application platform service - SW library, HW, integration and application development management (e.g, MLflow, Triton)**
  - **Build up customized ML platforms for user specified projects - Deploy ML packages ready environment in order to help ML development smoothly and provide on-demand computing power**
  - **Upkeep of the application framework**
  - **Workflow and data pipeline integration: tracking, model serving, packaging, model registry, evaluation & tracing**
  - **Efficiency Improvement**
  - **Inference-as-a-Service (incl. Model repository): e.g., Triton-based services**
- **Potential use cases**
  - **Users who bring existing source code - ASGC could help to setup a virtual environment and confirm source code running normally**
  - **Share and reuse of trained/verified models**
  - **MLOps over the cloud**
- **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.**



Capacity Building

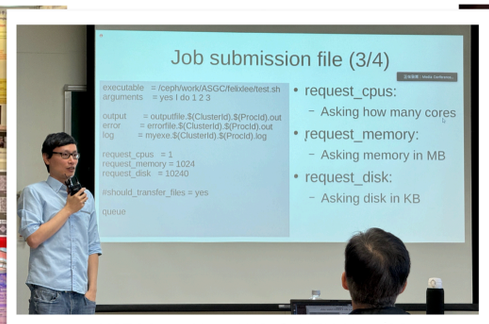


Enhancing User Community Engagement

- Training, dissemination and outreach
- Improving communication and response
- Supporting user access and application efficiency
- Understanding user requirements and experiences
- Facilitating capacity building of users and NSTCCore
- Training and workshop/conference: 250+ participants in 2025
- Weekly user meeting: 2:30pm, Wednesday

Basic	•30min – full day •Get access, details of using compute, data and SaaS services, w/o hands-on, User Forum
Thematic	•1 hrs – full day •Discipline-based Data analysis •ICT topics: GPU Computing, AI, CPU Trends
Domain-Specific Workshop	•30min – 3hrs •Supporting hands-on, w/o ASGC service introduction
Conference	•e-Science, community-specific workshop, core technology •ISGC (International Symposium on Grids & Clouds, since 2002) •HEPiX, CHEP (Computing in High Energy Physics) •APAN (Asia Pacific Advanced Network)

TPS 2026 @ Chiayi Jan/26



NSTCCore User WS @ NTU Nov/23

AI/ML WS @ NCKU Jun/24

AI/ML WS @ ISGC 2025, AS Mar/25

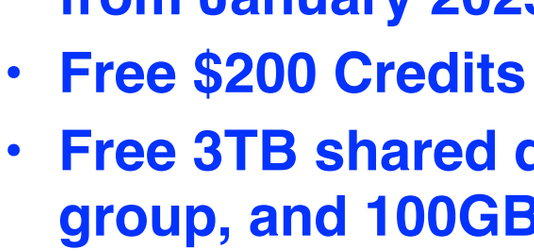
NSTCCore User WS @ Taichung Jan/26

Access to NSTCCore Services

- Two-Factor Authentication enabled from January 2025
- Free $200 Credits for 1-month Trial
- Free 3TB shared disk storage per group, and 100GB per user

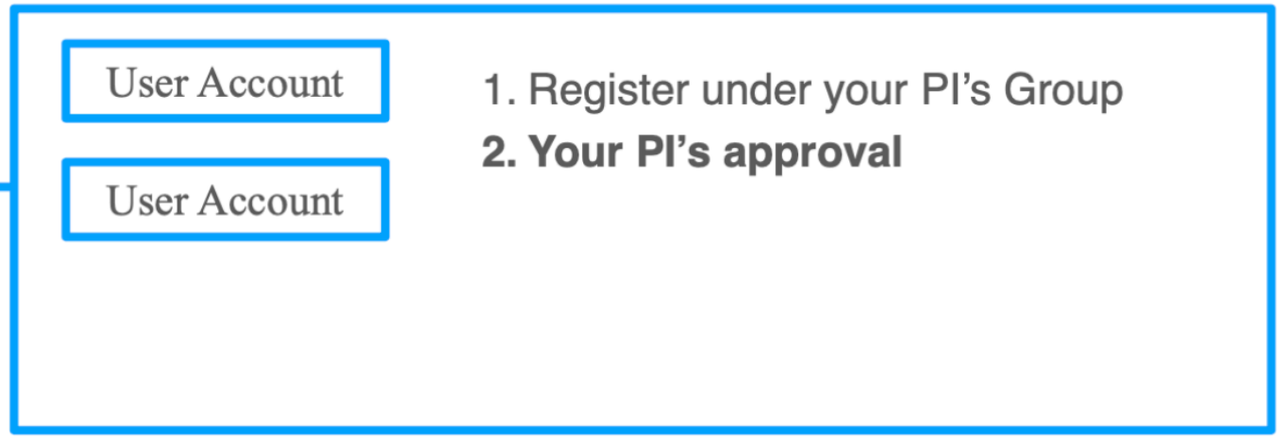
- Group Resource Usage and Payment Management
- Member accounts management support
  - Members list
  - Abnormal member usage report
  - Member's usage review

PI



- Group Accounts
  - https://canew.twgrid.org/ApplyAccount/groupcreate.php

Users



- User Accounts
  - https://canew.twgrid.org/ApplyAccount/ApplyAccount.php

Group Management

- **Members management**
  - **Members list**
  - **Abnormal member usage report**
  - **Member's usage review**

使用者 (username)	姓名(name)	Email	Expired Date	Active	Joined Date	Last Login	Storage Usage (UI Home)
chiong	CHAN-HIN IONG	chiong@me.com	Jun. 13, 2024, 00:00 AM	True	Aug. 22, 2018, 00:00 AM	Jul. 31, 2023, 03:41 AM	**33.4G/100G** Latest Update: 2023-07-31 00:50:03
dickie	Dickie Chang	dickie.chang@twgrid.org	Oct. 12, 2023, 00:00 AM	True	Dec. 15, 2022, 07:16 AM		
eric	嚴漢偉 嚴漢偉	Eric.Yen@twgrid.org	Oct. 18, 2023, 00:00 AM	True	Mar. 11, 2019, 09:13 AM	Jul. 06, 2023, 04:37 AM	**0.0G/100G** Latest Update: 2023-07-31 00:50:03
ericyen	Eric YEN	Eric.Yen@twgrid.org	Dec. 08, 2023, 00:00 AM	True	Apr. 23, 2013, 00:00 AM	Jul. 24, 2023, 03:22 AM	**0.0G/100G** Latest Update: 2023-07-31 00:50:03
etomo	etomo etomo	etomo@twgrid.org	Aug. 09, 2025, 00:00 AM	True	Dec. 23, 2020, 09:19 AM	Jan. 22, 2021, 10:16 AM	**0.0G/100G** Latest Update: 2023-07-31 00:50:03
felixlee	Felix Lee	felix@twgrid.org	Oct. 24, 2023, 00:00 AM	True	Mar. 05, 2013, 00:00 AM	Jul. 27, 2023, 07:29 AM	**436.3G/100G** Latest Update: 2023-07-31 00:50:03

使用統計摘要(TOTAL SUMMARY TABLE) | 使用明細(DETAILED USAGE ACCOUNTING)

使用明細(DETAILED USAGE ACCOUNTING)

使用者(username): chiong (CHAN-HIN IONG)

用量異常回報(submit issue)

Alert		#Instance	A100		FDR5		P100		RTX3090		V100	
			CPU (SRU)	GPU (SRU)	CPU (SRU)	GPU (SRU)	CPU (SRU)	GPU (SRU)	CPU (SRU)	GPU (SRU)	CPU (SRU)	GPU (SRU)
☐	relion311rtx3090	3							151	6,320		
☐	jupyterlabtf24gpu3090	2							11	474		
☐	matlab	2									0	6
☐	jupyterlabgpu26a100	1	1	519								
☐	openaccp100	1					11	282				
☐	FDR5 slurm	2			0	0						
☐	STORAGE USAGE (0.0 TB)											

使用統計摘要(TOTAL SUMMARY TABLE) | 使用明細(DETAILED USAGE ACCOUNTING)

- 計算資源使用統計(COMPUTING USAGE OF GROUP) – 預估使用費(INITIAL COST): NT $15,878

使用者(username)	姓名(name)	CPU (SRU)	GPU (SRU)	預估費用 (Initial Cost Estimation)
chiong	CHAN-HIN IONG	175	7,601	15,552
jyou	Jingya You	1	54	110
rudy	陳侑廷	3	0	6
thwu	Tsung-Hsun Wu	4	101	210
總計	單價SRU=NT$2	183	7,756	(183+7,756)*2=15,878

- 儲存資源使用統計(STORAGE USAGE OF GROUP) – 預估使用費(INITIAL COST): NT $417

帳號(ID)	名稱(name)	儲存空間用量(storage(TB))	預估費用(Initial Cost Estimation)	說明(note)
ASGC	Group	0.0	0.0	Group使用空間
chiong	CHAN-HIN IONG	0.0	0.0	User使用空間
jyou	Jingya You	0.0	0.0	User使用空間
rudy	陳侑廷	0.0	0.0	User使用空間
thwu	Tsung-Hsun Wu	4.9	408.3	User使用空間

NSTCCore Services 聯絡資訊

- Core Facility Services
  - https://nstccore.twgrid.org
  - https://scale.grid.sinica.edu.tw/index.php
- ASGC Web Site: https://www.twgrid.org
- Access to ASGC Resources
  - https://dicos.grid.sinica.edu.tw/
- Contact point: DiCOS-Support@twgrid.org

- Gentle Reminder
  - Please include ASGC (Academia Sinica Grid-computing Center) in the acknowledgement when research outcomes that relied on ASGC resources, services or expertise are presented in your research.