Site Report of IHEP

Xiaowei JIANG

On behalf of CC-IHEP, CAS

Outline

1. Overview of IHEP Computing Center
2. Computing Platform
3. LHCb Tier-1 Construction
4. Progress on R&Ds
5. Summary

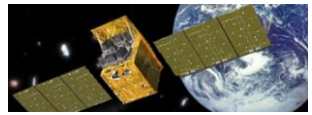
Overview of IHEP CC

- 58K CPU cores, 250 GPU cards to for more than 10 experiments
  - HTC cluster (42K CPU cores)
  - HPC cluster (10K CPU cores + 250 GPU)
  - Distributed computing, WLCG, DIRAC etc. (6K cores at IHEP)
- 97.4 PB disk storage, 80 PB tape storage
  - Lustre (39.4 PB, POSIX) and EOS (58 PB, XRootD)
  - EOSCTA for tape storage (80 PB, all have been migrated from Castor to EOSCTA)
- Network
  - IPV4/IPV6 dual stack
  - Ethernet/IB/ROCE protocols supported
  - WAN Bandwidth: 100 Gbps (LHCOPN and LHCONE 20Gbps)

Chinese located or IHEP driven experiments

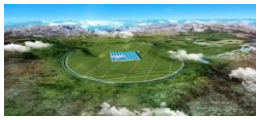
BESIII (Beijing Spectrometer III at BEPCII)

JUNO (Jiangmeng Underground Neutrino Observatory)



HXMT (Hard X-Ray Moderate Telescope)

CSNS (China Spallation Neutron Source)



LHAASO (Large High Altitude Air Shower Observatory)

HEPS (High Energy Photon Source)

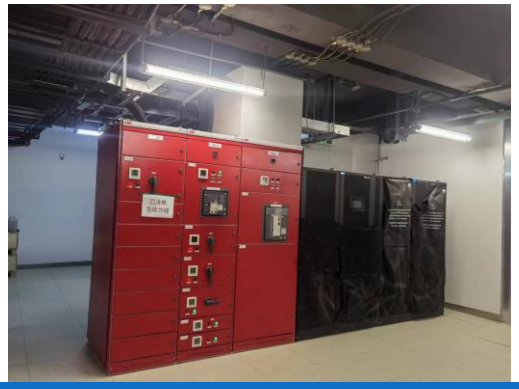
HERD (High Energy Cosmic Radiation Detection)

CEPC (Circular Electron Positron Collider)

International collaborated experiments

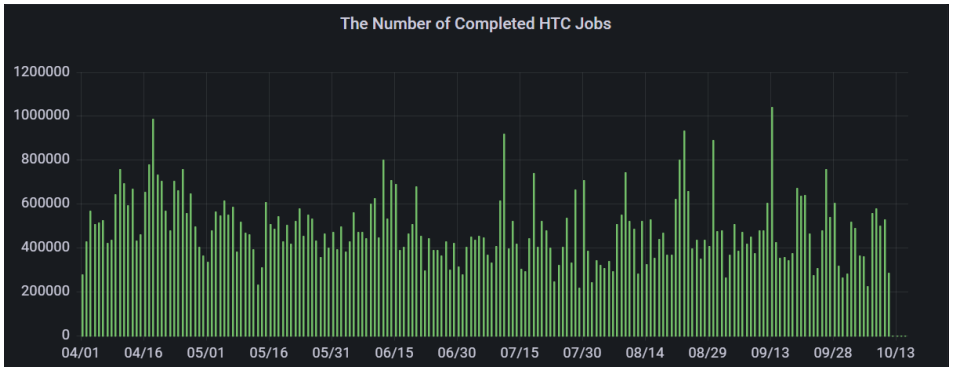
New Machine Room for HEPS

- **HEPS data center is located in the north of Beijing city**
  - The main machine room is 520㎡
  - 47 racks in Phase I: 20 for storage, 21 for computing and 6 for network
  - Power infrastructure
    - 2 transformers (2500kVA+2500kVA): backup for each other
    - Utility power supply and uninterruptible power supply
      - UPS capacity is 800kVA providing a backup time of half an hour
    - 15kW/rack for storage, 30kW/rack for computing (Utility power supply + UPS)
    - Cooling equipment (dual utility supply)
  - Wind Cooling system: Split air conditioner
- **Current Status**
  - Finished deployment of racks, power system and cooling system
  - The server devices are under procurement

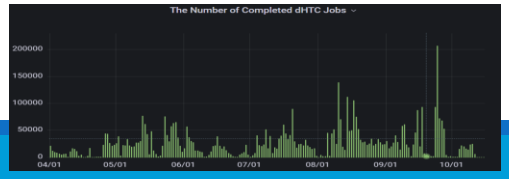

High Throughput Computing

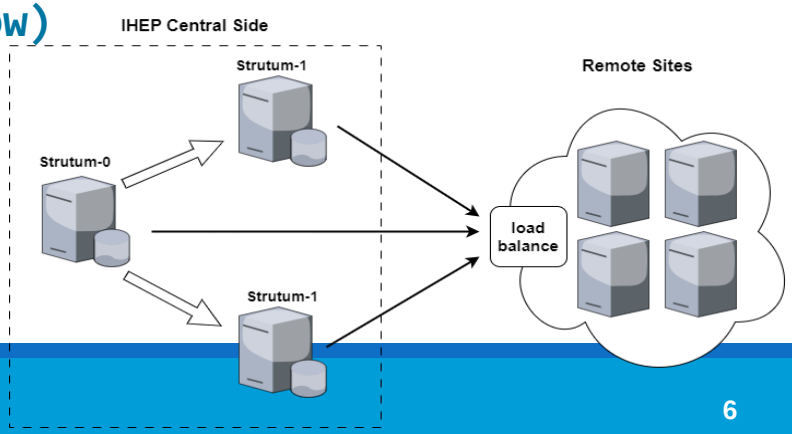
- **Upgrade the hardware of HTCondor servers**
  - Replace the central manager server with a new device
  - Replace the schedd server of BES experiment with a new device
- **Multiple negotiators**
  - One negotiator face pressure when massive short jobs are coming into the pool
  - Two Negotiators have been set up for the whole pool
    - Each negotiator is responsible for half of the worker nodes
  - Problem: the user priority settings are separated on each negotiator
- **HTC Job statistics**
  - 93,970,525 jobs completed
  - 154,318,720 CPU hours consumed



One platform, Multi Centers

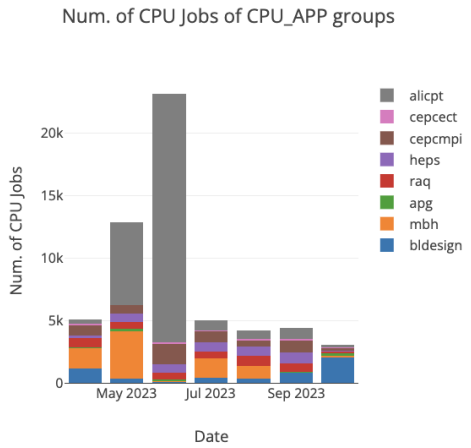
- **Distributed high throughput computing system**
  - Add a personal-software cache mechanism to reduce the redundant software file transfer
    - cache the software transferred by the first job on a worker node
- **Data access and transfer**
  - Updates on using CVMFS to share the common data (~100TB random trigger data used by BES experiment)
    - Deploy three CVMFS servers for BES experiment (each server covers 1500~2000 jobs)
    - Analyze and adjust the trunk size to match with the data size by each read of BES job
  - Updates on using XRootD to share the data stored in Lustre
    - Disabled caching KRB5 token on xrootd server
    - Grant the root permission to xrootd server
- **Network: add a new 10Gbps network link between north and south centers (totally 20 Gbps now)**
- **HTC Job statistics**
  - 4,977,002 jobs completed
  - 23,238,385 CPU hours consumed

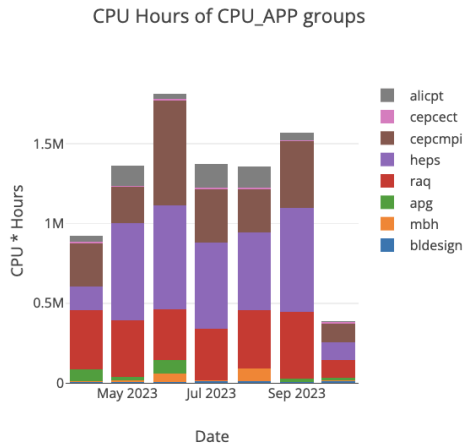




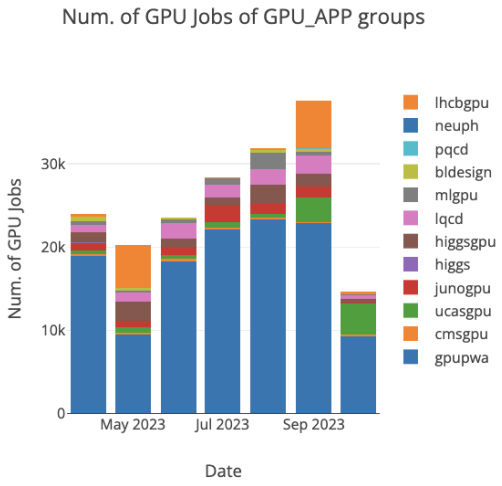
High Performance Computing

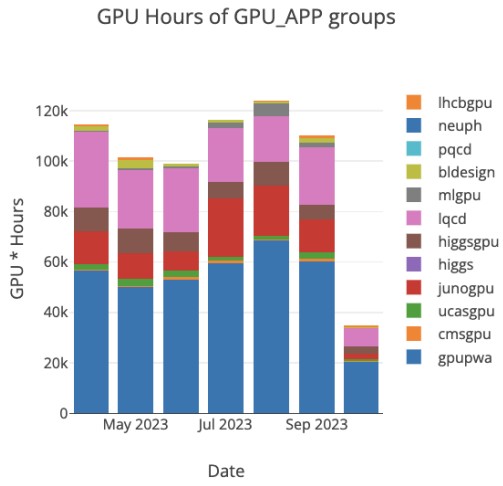
- **8 CPU apps, 57.8K jobs, 8.8M CPU hours**





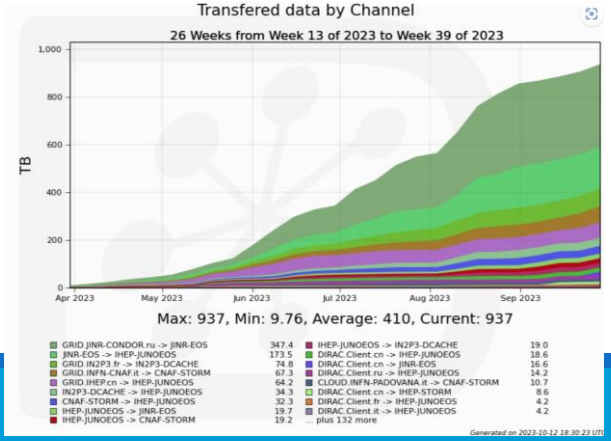
- **11 GPU apps, 189.2K jobs, 1.1M GPU hours**

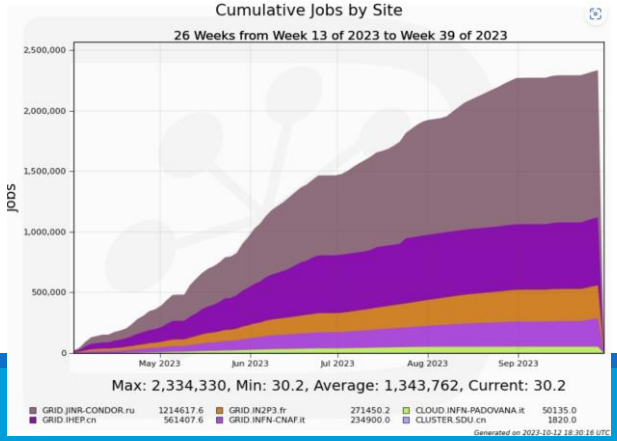




Distributed Computing

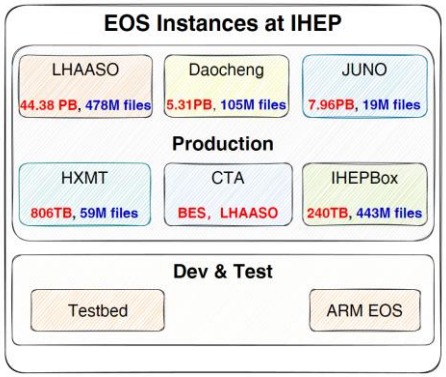
- **DIRAC at IHEP**
  - Serving BESIII, JUNO, HERD, CEPC
  - DIRAC for computing and data management, upgrade to v8.0.26 and move to distributed deployments since July 2023
  - Start to manage JUNO's First Data Challenge(DC1)
- **Rucio at IHEP**
  - Finished HERD Rucio API development and deployment, provided an integrated API to experiment software
- **Grid middle-ware services**
  - HERD IAM at IHEP deployed and in test
  - Service monitoring system and site monitoring system for distributed computing are under developing





Storage (Disk and Tape)

- **Disk storage - EOS**
  - 6 instances supporting 3 experiments, IHEPbox and CTA
  - Add 2 new instances for LHCb Tier-1 site (disk and tape)
- **Disk storage - Lustre**
  - 22 instances for BES, JUNO, HXMT, CEPC, HEPS, etc.

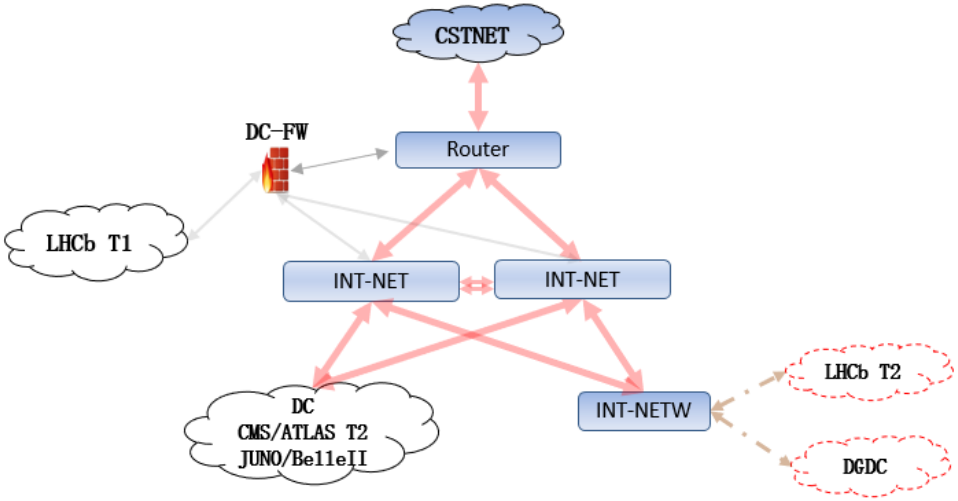


- **Tape storage – EOS-CTA**
  - Supporting 6 experiments including LHAASO, BESIII, JUNO, etc.
  - Upgrade all CTA&EOS to V5
  - Setup a tape buffer for LHCb Tier-1 site
  - Build a new tap library for HEPS

CTA	LHAASO	YBJ	HXMT	DYB	BES3	TOTAL
Files	7M	2419	1.5K	1.3M	258K	8.5M
Used	9.25PB	185.28TB	25.17T	1.16PB	3.18PB	13.77PB

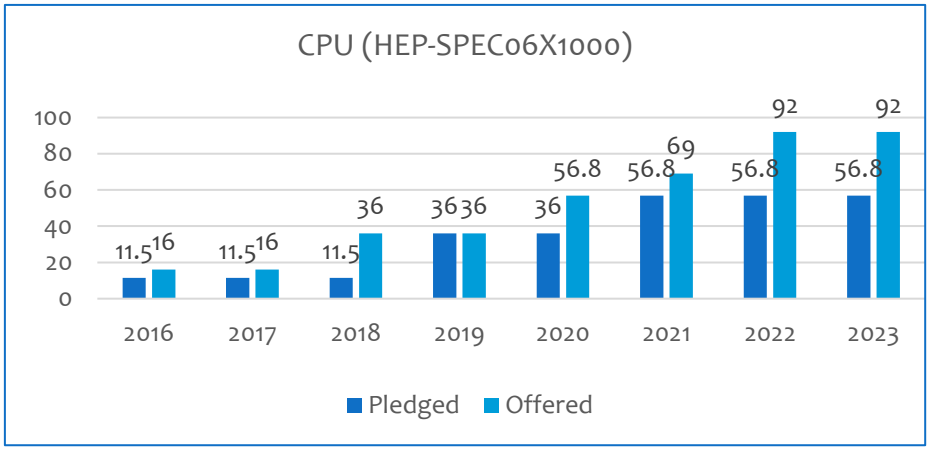
Network

- **Network Bandwidth**
  - Backbone: 200GbE (dual-machine redundancy) (July 2023)
  - Internet: 100GbE to CSTNET (Aug 2023)
- **Internal network status (inside IHEP)**
  - Max throughput is 233 Gbps
  - 21% increased in 25GbE access switches (total 1392 ports)
  - The proportion of 25GbE hosts is 62%
- **Experiment Supports**
  - HEPS (Sep 2023)
    - 100GbE to IHEP is ready
    - Backbone network is ready
  - LHCOPN
    - 20GbE LHCOPN and 20GbE LHCONE
    - Based on CSTNET-GEANT-100G

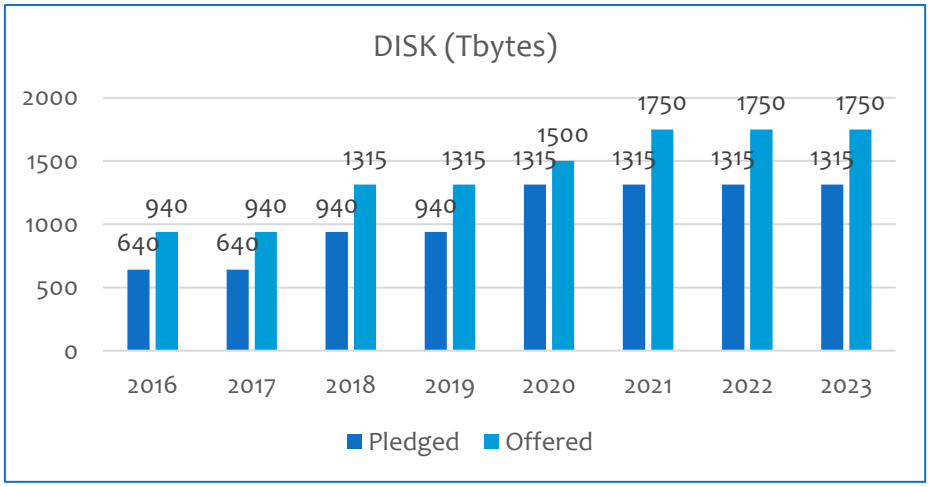


Grid Site Status

- **CPU: 4232 cores**
  - Intel Golden 6338: 1152 Cores
  - Intel Golden 6238R: 672 Cores
  - Intel Golden 6140: 2160 Cores
  - Intel E5-2680V3: 696 Cores
  - Intel X5650: 192 Cores
- **CE & Batch: HTCondorCE & HTCondor**
- **VO: ATLAS, CMS, LHCb, BelleII, JUNO, CEPC**

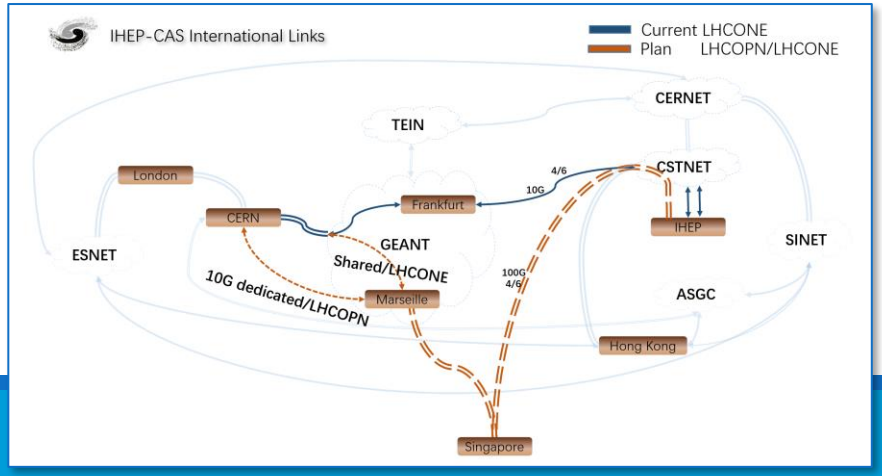


- **EOS: 1750TB**
  - 4TB * 24 slots with Raid 6, 5 Array boxes
  - DELL MD3860 8TB * 60 slots
  - DELL ME4084 10TB * 42 slots
  - DELL ME4084 12TB * 84 slots
- **EOS replaced DPM in this May**



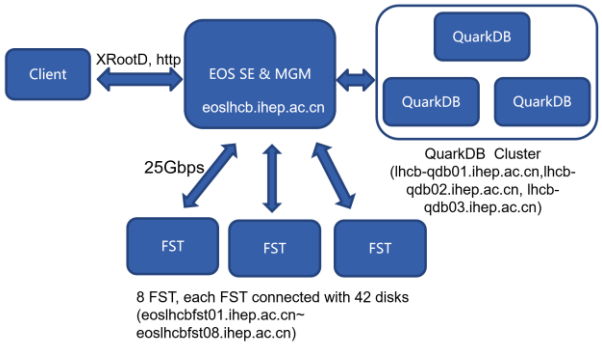
LHCb Tier1 Site Construction

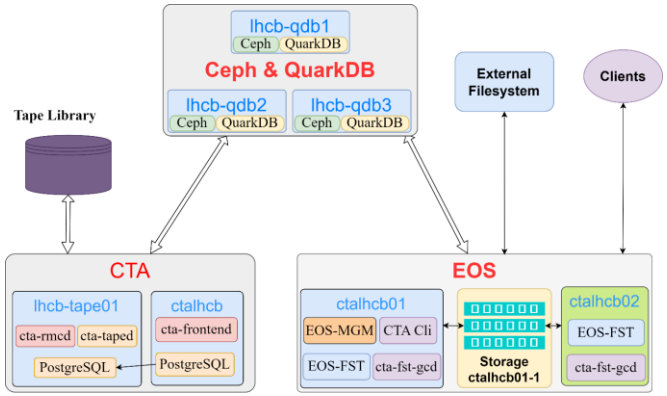
- All the hardware is ready
  - Computing: 3216 CPU cores, 40 worker nodes (Intel & AMD)
  - Disk storage: ~3.2PB, 4 sets of storage array
  - Tape storage: ~3PB, 170 tapes, 4 drivers (IBM)
  - Network equipment: 6 switches, 1 router, 2 band cards
  - Management servers: 10 servers
- Currently Tier1 site is also reusing part of existing hardware
  - Firewall device, tape library, CA system, …
- The network capacity for LHCOPN and LHCONE is ready
  - 20 Gbps bandwidth for LHCOPN

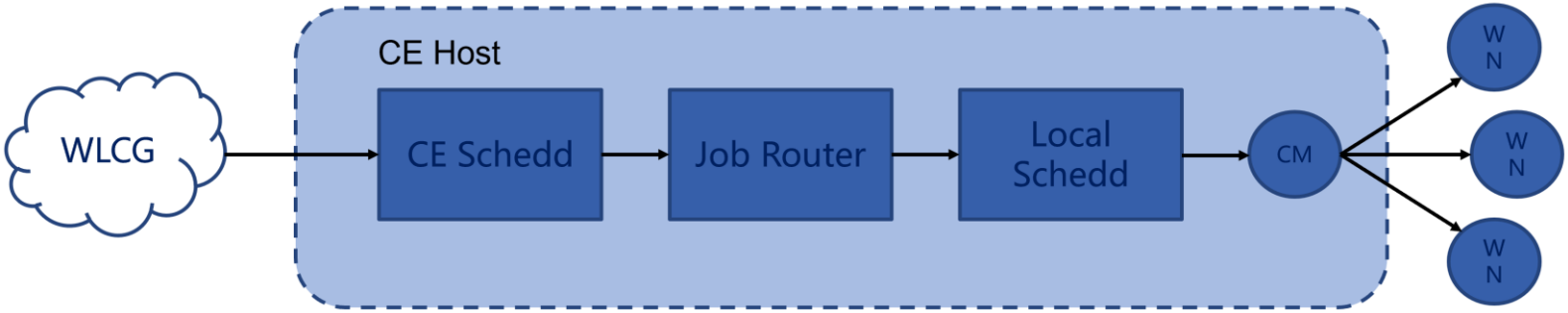


LHCb Tier1 Site Construction

- **Disk storage: EOS**
  - services: QuarkDB, MGM, FST
  - protocol: xrootd and http
- **Tape strorage: EOS & EOS-CTA**
  - Protocols: xrootd and http
- **CE: HTCondor-CE & HTCondor**
  - Support for SCIToken and GSI
- **Other middle software**
  - Argus, BDII, APEL

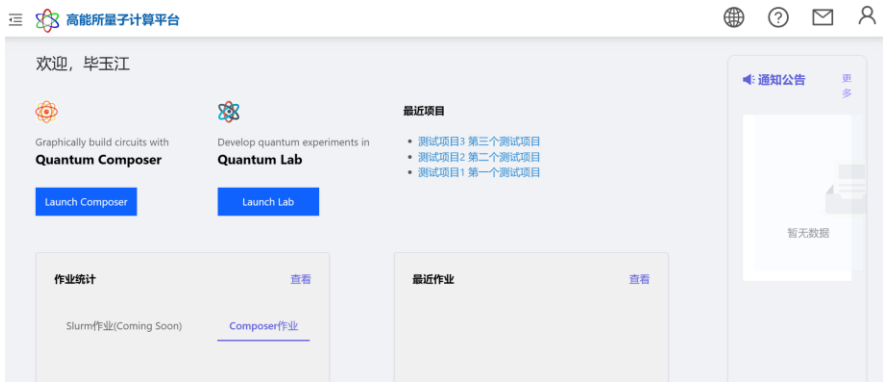






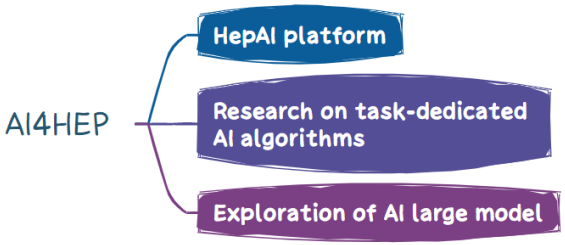
Quantum Computing

- **QuIHEP**
  - A distributed heterogeneous interactive developing platform
  - Facilitate the explorations of quantum algorithms in HEP experiments
    - LQCD, CEPC, BESIII, etc.
  - Connect IHEP HPC cluster to QuIHEP platform
    - Provide more GPU resources
- **Qiskit simulation on AMD Platform**
  - Ported qiskit-aer from CUDA to ROCm platform



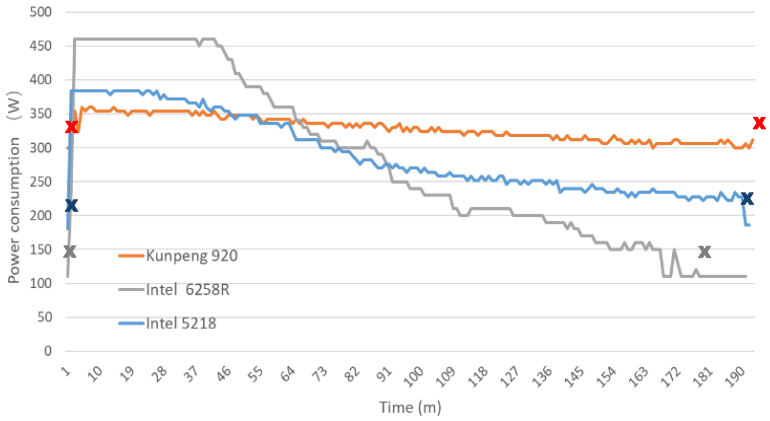
AI Platform

- **HepAI platform**
  - The Distributed, cross-system, high-concurrency Deployment Framework (HepAI DDF) have been developed and deployed
  - The portal webui is developed and deployed (https://ai.ihep.ac.cn)
  - Serveral AI models (LLMs, SAM, PointNet, ParticleNet) are integrated into the platform
  - A annotation tool based on HepAI GF for HEPS image labeling has been developed
- **Task-dedicated AI algorithms**
  - An AI algorithm for fast reconstruction of Ptycho-graphy is under development
  - An AI algorithm for intelligent analysis of microscopic defects for X-ray additive manufacturing images is under development
- **Large Lanuage Model**
  - Xiwu, a large language model boasting 13B params with just-in-time learning for HEP has been developed
  - The research on enhancing the language model's capabilities and exploring the feasibility of rediscovering Zc(3900) is currently underway



ARM Architecture

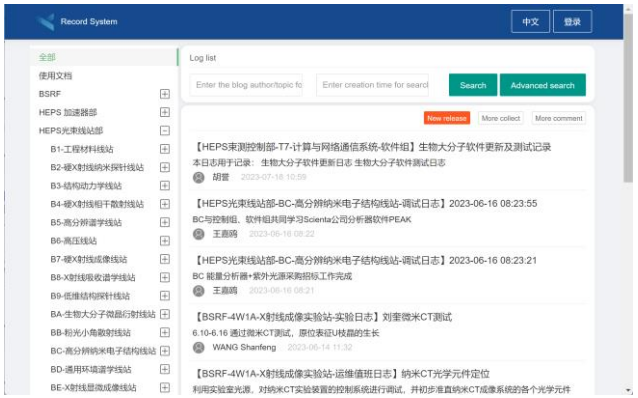
- **Port LHAASO-WFCTA&KM2A to ARM**
  - Corsika-V77420 and G4KM2A-4.10
- **Port HERD software to ARM**
  - HERDOS and simulation software
- **Performance test using WFCTA-Corsika**
  - Test conditions
    - Kunpeng920 (ARM)
    - Intel 6258R and 5218 (X86)
  - Test Results
    - The ARM server based on the Kunpeng 920 architecture has certain power consumption advantages when running Corsika simulation jobs



CPU type	Number of jobs	Running time (m)	Average running time per job	Electricity consumption (W • H)	Jobs electricity consumption (W • H)
ARM-920	96	4h6m	103.75m	1355.51	125.51
X86-6258R	56	2h54m	77.17m	933.57	614.57
X86-5218	32	3h20m	83.09m	967.36	367.36

HEPS Experiment

- **High Energy Photon Source (HEPS)**
  - Plan to start service in 2025.
- **Computing & Communication system (HEPSCC):**
  - Network, Computing, Storage, Data analysis framework, Data management, Database & Public Service, Monitoring, Security.
- **Data analysis framework (in developing):**
  - Integrate methods and algorithms: Liquid Diffract, DM
  - Developed multi-threaded software for parallel reading and writing of TIFF files.
  - Developed a distributed parallel CT reconstruction program based on Spark and K8s
  - Developed the CI/CD system: an automated pipeline for software repository compilation and deployment, an automated pipeline for container image packaging and distribution
- **Data management (development finished):**
  - Developed the logbook, release to the HEPS user
  - The entire system is beginning to be deployed and debugged on the HEPS site.
- **User service system (development finished):**
  - Has completed system design and development
  - Including beamline management, proposal submission and review, beamtime reservation and allocation, and user visits.



HERD Experiment

- **The High Energy cosmic Radiation Detection facility (HERD)**
  - Installed on the China Space Station, plans to launch in 2027
- **Distributed Computing System**
  - Rucio: HERD-Policy for pre-study data is deployed in production
  - DIRAC: multi-vo DIRAC instance at IHEP is ready
  - IAM at IHEP: already deployed and in test
  - Other grid services: multi-vo FTS3 instance, StoRM over Lustre (will be replaced by EOS)
- **Data Management**
  - Simulation data management system has been designed and developed
  - Simulation data processing workflow is implemented
    - Data generation→ temporary storage→ validation→ data transfer(distributed sites)→ metadata extraction→ catalogue
    - Monitor the running state of any node in the data workflow

Summary

- The platform runs without big problem in last 6 months
  - Add more resources and optimize the performance on HTC and EOS
- A new machine room is built in HEPS data center
  - Finished deployment of racks, power system and cooling system
- LHCb Tier1 site construction is close to be done
  - All resource devices and services are ready
  - Start feature test and data challenge with LHCb
- Some R&D work are progressed as plan
  - Quantum Computing Platform
  - AI platform
  - ARM Porting and Application
  - Software and computing system for HEPS and HERD

Thanks!
Q&A