Site Report of IHEP

Xuantong Zhang

On behalf of CC-IHEP, CAS

Outline

1. **Brief Introduction**

2. **Computing Platform**

3. **LHCb Tier-1 Construction**

4. **Supports and R&Ds**

5. **Summary**

Brief Introduction

38 K CPU cores, 250 GPU cards to for more than 10 experiments

- HTCondor cluster runs for HTC jobs (38K CPU cores)
- Slurm cluster runs for HPC jobs (10K CPU cores + 250(+40) GPU)
- Distributed computing, WLCG, DIRAC etc., (~100 kHS06, 2K cores at IHEP)

89.4 PB disk storage, 55.1 PB tape storage

- Lustre (39.4(+9) PB, POSIX) and EOS (50(+4) PB, XRootD),
- Castor for tape storage, retiring, 0(-19) PB,
- EOSCTA for tape storage (37.5(+4.7) PB, castor still in migrating).

Network

- IPV4/ IPV6 dual stack,
- Ethernet / IB protocols supported,
- LHCONE member,
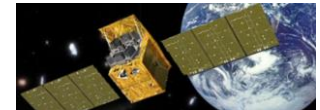- WAN Bandwidth: 100(+60) Gbps (LHCONE 20Gbps).

Chinese located or IHEP driven experiments

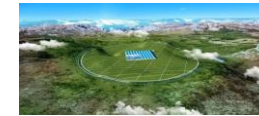**BESIII** (Beijing Spectrometer III at BEPCII)

JUNO (Jiangmeng Underground Neutrino Observatory)

HXMT (Hard X-Ray Moderate Telescope)

CSNS (China Spallation Neutron Source)

LHAASO (Large High Altitude Air Shower Observatory)
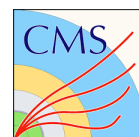
HEPS (High Energy Photon Source)

HERD (High Energy Cosmic Radiation Detection)

CEPC (Circular Electron Positron Collider)

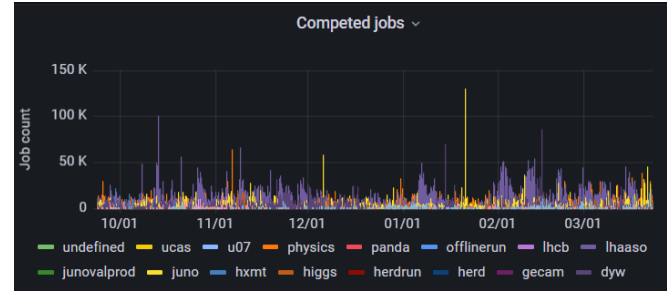### International collaborated experiments

Computing Status & Update

Computing:

◦ **Status:**
  ◦ HTCondor, 3 Sched, 1 Condor CM.
  ◦ Slurm, 233(+5) worker nodes, 9 CPU APPs, 11 GPU APPs.

◦ **Update:**
  ◦ A new admin portal added for Slurm.

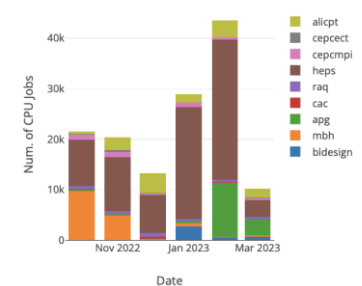## Distributed Computing:

◦ **Status:**
  ◦ Serving BESIII, JUNO, HERD, CEPC.
  ◦ DIRAC for computing and data management, v7.3

◦ **Update:**
  ◦ Rucio starts to support HERD with an API integrated to experiment software.
  ◦ IAM with Sci-token and VOMS AA is ready.
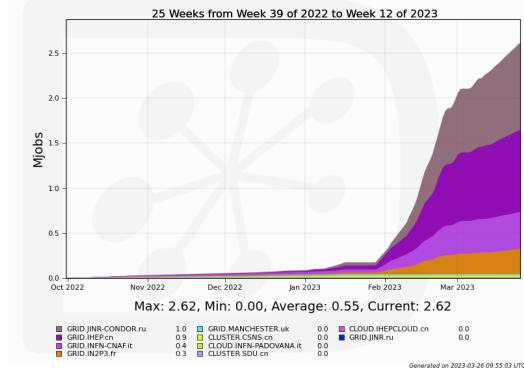

Competed jobs

Num. of GPU Jobs of GPU_APP groups
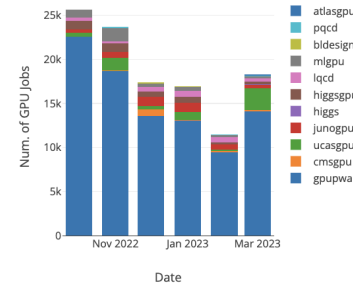
Num. of CPU Jobs of CPU_APP groups

Cumulative Jobs by Site

Storage Status & Update
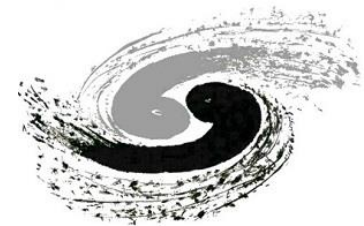
Disk:
- **Lustre:**
  - Status:
    - BES, JUNO, HXMT, CEPC, HEPS, etc.
    - 22(+1 for BES) instances.
- **EOS:**
  - Status:
    - 6 instances, support 3 experiment, IHEPBox and CTA.

Tape:
- **EOS-CTA:**
  - Status:
    - All experiments have adopted EOS-CTA for backup. Castor is retiring.
    - LHAASO, HXMT, BESIII, JUNO, HEPS.
  - Update:
    - CTA Transmission System(CTS) is used for backup from EOS/Lustre,
    - Integrated with Redis, Kafka and RabbitMQ.
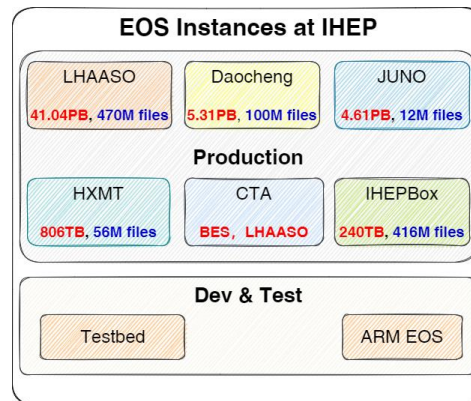
EOS CTA	LHAASO	YBJ	HXMT	DYB	BES3	TOTAL
Files	7M	2419	1.5K	1.3M	258K	8.5M
Used	9.25PB	185.28TB	25.17T	1.16PB	3.18PB	13.77PB

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

CPU (HEP-SPEC06X1000)

Year	Pledged	Offered
2016	11.5	16
2017	11.5	16
2018	11.5	36
2019	36	36
2020	36	56.8
2021	56.8	69
2022	56.8	92
2023	56.8	92
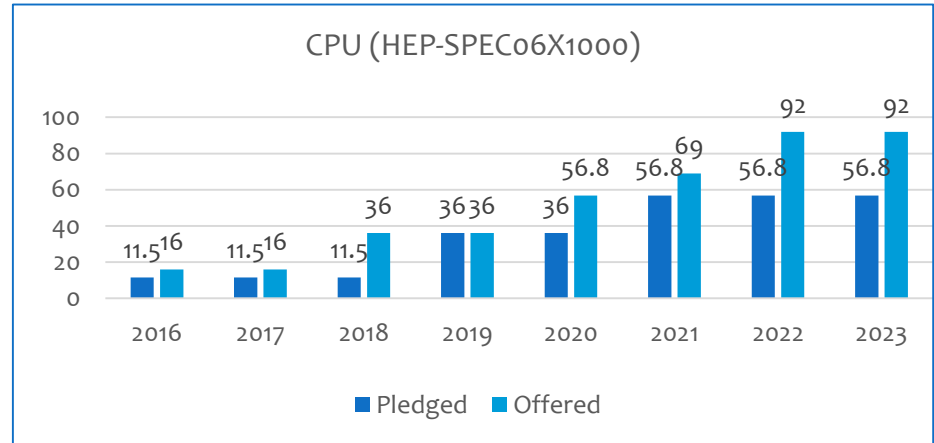
DPM: 1750TB

- 4TB * 24 slots with Raid 6, 5 Array boxes
- DELL MD3860 8TB * 60 slots
- DELL ME4084 10TB * 42 slots
- DELL ME4084 12TB * 84 slots

Will migrate to EOS in May, 2023

DISK (Tbytes)

Year	Pledged	Offered
2016	640	940
2017	640	940
2018	940	1315
2019	940	1315
2020	1315	1500
2021	1315	1750
2022	1315	1750
2023	1315	1750
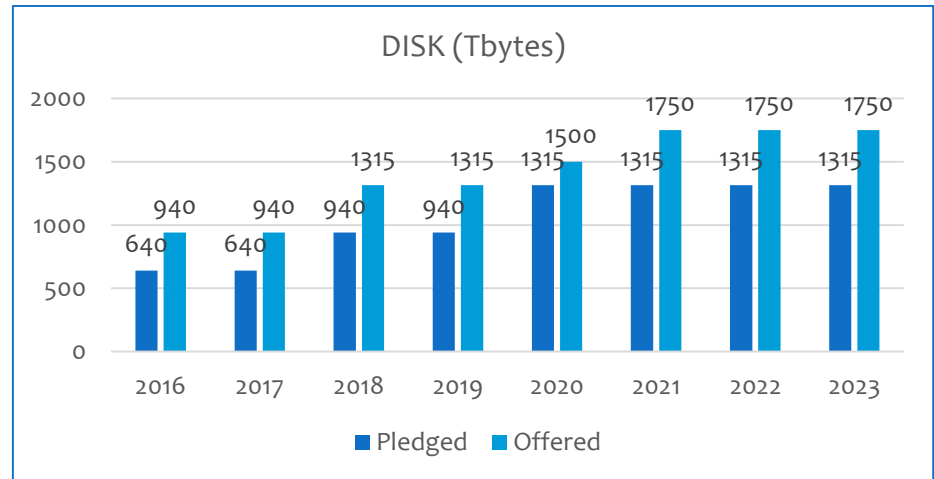
DPM – EOS Migration

Why EOS?
◦ **Most new HEP experiments use EOS as main storage,**
◦ **IHEP LHCb T1 will use EOS as its storage system,**
◦ **EOS is the main storage at CERN.**

New EOS configuration:
◦ **EOS SE and MGM are deployed on one server,**
◦ **ccsrm.ihep.ac.cn -> cceos.ihep.ac.cn,**
◦ **Each experiment will have a group in EOS.**

Time table:
◦ **Notification:**
  ◦ 17 March, 27 March, 14 April, 3 times notifications to users.
  ◦ 17 March, notification to VO admins.
◦ **Downtime for Migrations:**
  ◦ 8:00, 08 May – 23:30, 10 May (UTC+8:00).
◦ **EOS SE will be online before 23:30, 10 May (UTC+8:00).**

More details: ISGC report.

Network

Network bandwidth:
- Backbone: **80 GbE -> 200GbE** (July 2023),
- Internet: **40 GbE -> 100GbE** (April 2023) to CSTNET.

Data Center:
- **30% increased** in 25 G access switches, total 1152 ports,
- The proportion of 25 G hosts is 60%,
- Max throughput is 302 Gbps.

Experiments:
- **JUNO:**
  - 150 MbE new link from JUNO to CSNS,
  - Campus network upgraded.
- **HEPS:**
  - 10 GbE will be ready since July 2023.
- **Jinan Research Department:**
  - 500 MbE will be ready since July 2023.

IHEP Network Topology 2023
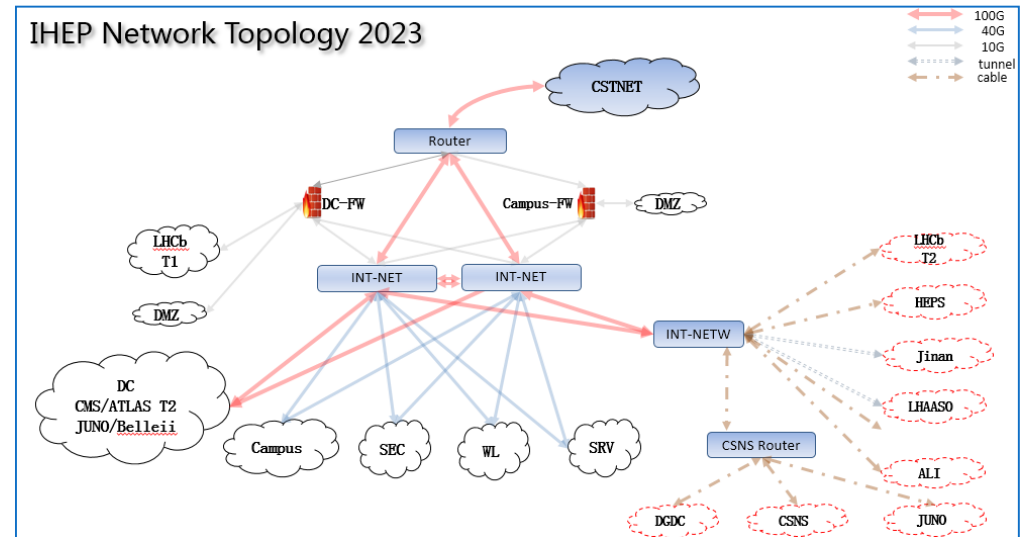
LHCb T1 (IHEP) & T2 (LZU)

Tier1 at IHEP	Tier2 at Lanzhou University (LZU)

Upgraded from Beijing LHCb Tier2

Initial resources will be ready in the first half year of 2023.

- ~3000 CPU cores, Intel Xeon Platinum 8352Y,
- ~3.2PB disk storage, DELL PowerVault ME484,
- Lenovo TS4500 Tape Library, LTO9 Drives and Tapes.

A 100 Gbps new link (CSTNET-GEANT).

A 10 Gbps dedicated link from IHEP to CERN.

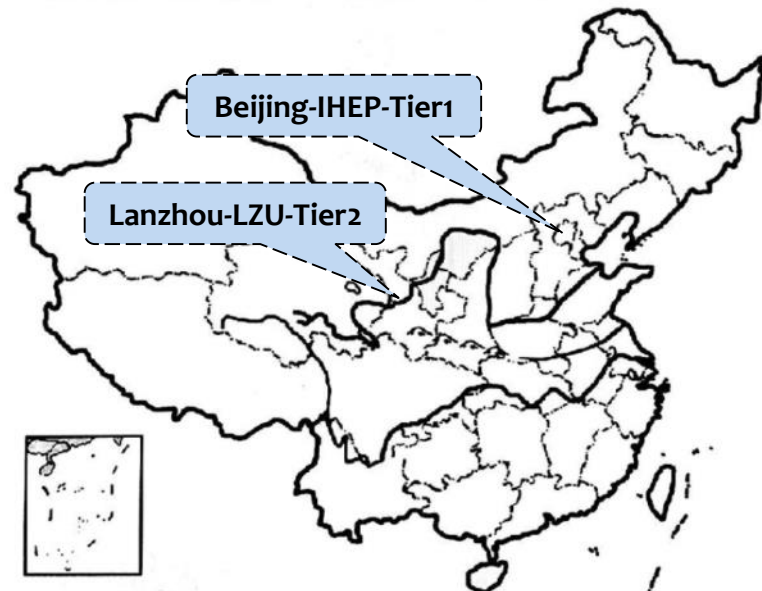**Additional manpower can be added if needed, when Tier-1 is established.**

New Tier-2 site will be built at LZU in 2023.

- ~3500 CPU cores,
- ~3PB Disk Storage,
- A dedicated 2Gbps link between IHEP and LZU.

CC-IHEP will be responsible for the site establish and maintenance.

Beijing-IHEP-Tier1

Lanzhou-LZU-Tier2

Tier Upgrade Construction Status

A proto-T1 is constructed at IHEP.
- Plan to finish T1 construction before June 2023.
- Construction is in smooth progress.

Computing:
- CPU (HTCondor CE):
  - 40% is ready, the rest has finished procurement, waiting for delivering.

Storage:
- Disk (EOS SE):
  - In bidding, will be ready in May 2023.
- Tape (EOS CTA):
  - Disk buffer is ready in March 2023.
  - Tape will be ready in April 2023.

Network:
- Tier1:
  - 100 GbE link (CSTNET to GEANT): will be ready in April 2023.
  - 10 GbE link (IHEP to CERN-LHCOPN): will be ready in April 2023.
- Tier2:
  - 2 GBE link (LZU to IHEP): ready.

One Platform, Multi-Centers

Porting more applications to ARM-Processor machines:
- Experiment HERD has running simulation software (within HERDOS) with 6000 jobs in parallel,
- Experiment LHAASO has tested running Corsika software on ARM,
  - the large-scale testing is under development.

Transparently run BES application in the distributed site:
- Before that, BES has no distributed element in her computing platform: only a local cluster without WMS, DMS and Power authentication,
- Scale up the DB system (launch the dedicated sub-DB on remote site),
- Package non-transparent stuffs with boss.condor and its plugins,
- Optimize the data access with CVMFS (used by the reconstruction job).

Job statistics:
- 3,545,643 job completed,
- 23,189,756 CPU hours consumed.

Details about LHASSO data process will be at HEPiX report.
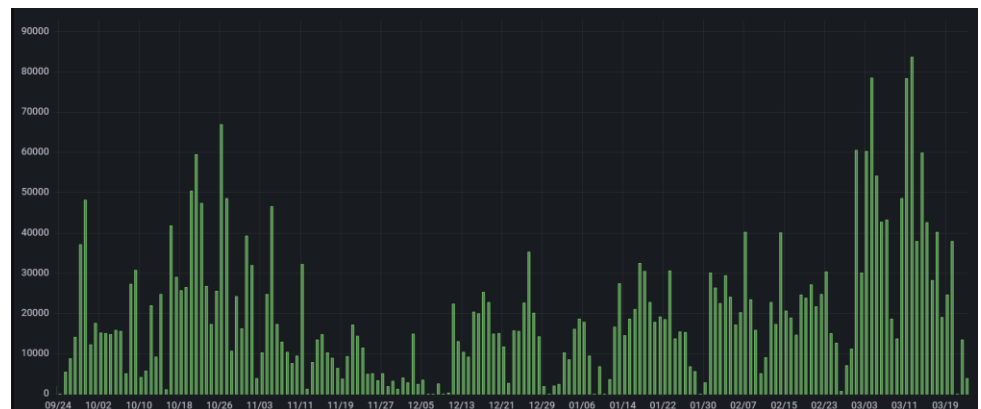
R&D: Quantum Computing Developing Platform

A distributed heterogeneous interactive developing platform:
◦ **Facilitate the explorations of quantum algorithms in HEP experiments,**
  ◦ LQCD, CEPC, BESIII, etc.

A simple user dashboard:
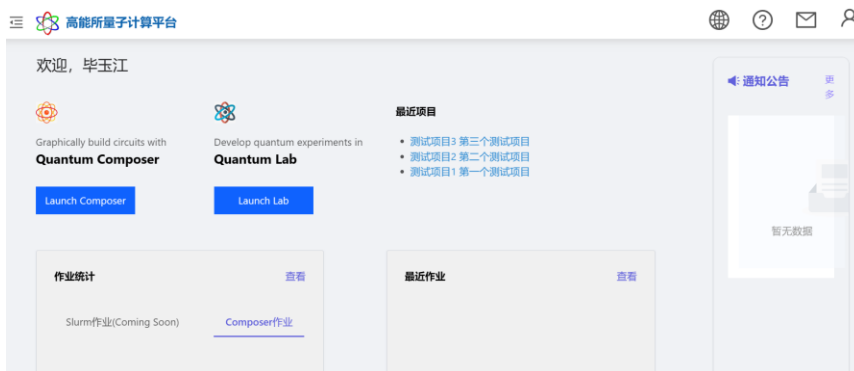◦ **Entry-points of programing interfaces and information.**

Jupyter-based interactive developing & analysis platform:
◦ **Support CPU and GPU environment,**
◦ **Integrated with HPC cluster via PySlurm plugin.**

Drag-and-drop programming interface:
◦ **Support basic quantum gates, and QASM translation.**

R&D: AI Platform

HAI Platform:

- **Integrated with HEP experiment:**
  - Integrate four deep learning algorithms for particle physics jet tagging task.
  - Integrate an algorithm for atmospheric neutrino reconstruction and identification for JUNO experiments.
  - Integrate three particle physics datasets.
- **Other features:**
  - Provide a NPU 8 card computing power.
  - Free trail version of ChatGPT online: https://ai.ihep.ac.cn.
  - Sources: https://code.ihep.ac.cn/zdzhang/hai
- **A DL GUI framework (HaiGF) based on Hai backend is being developed.**
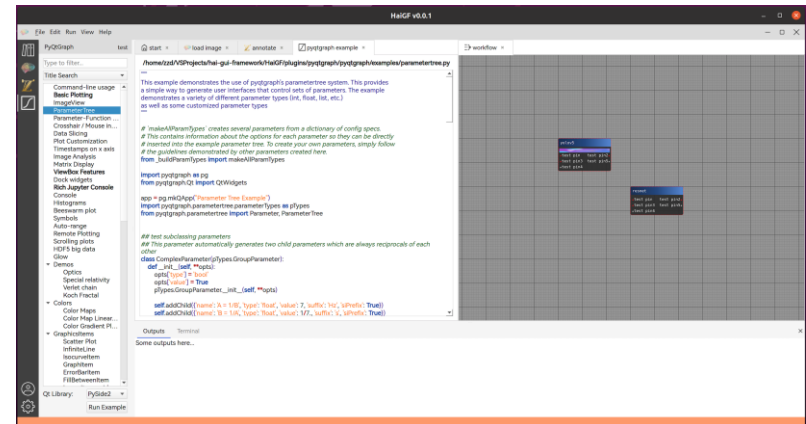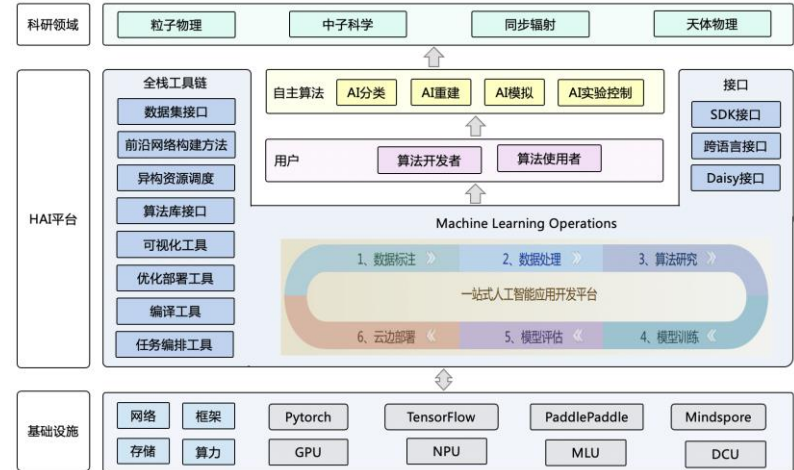
R&D: Computational Storage

Developed the first ARM and FPGA-based computational storage server and began deployment (~3PB).

◦ **Equipped with built-in compression algorithm acceleration and increasing compression speed by 3 times.**

Develop a computational storage software: XkitS

◦ **Processing local data on the EOS storage server FST to reduce data transfer.**
◦ **Fully configurable and compatible with existing data access interfaces.**

Source: ACAT'2022
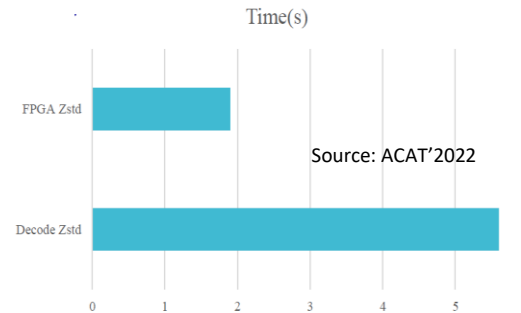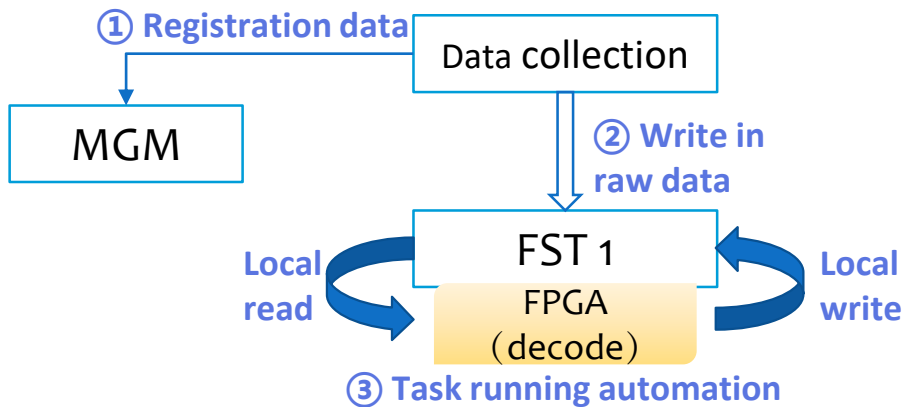
The compression speed of FPGA zstd is about 3 times that of software zstd

R&D: 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: PyNx,**
◦ **Develop PCA&LCF absorption spectrum components analysis interface of Daisy workbench based on PyQt5**
◦ **Develop the pair distribution function pipeline GUI based on jupyterlab**

Data management (development finished):
◦ **Develop data management module for 3-Tiers data storage (beamline storage → Central storage → Tape)**
◦ **Data transfer module moves data between different storage media automatically**
◦ **Design HDF5 data format for five beamline stations**
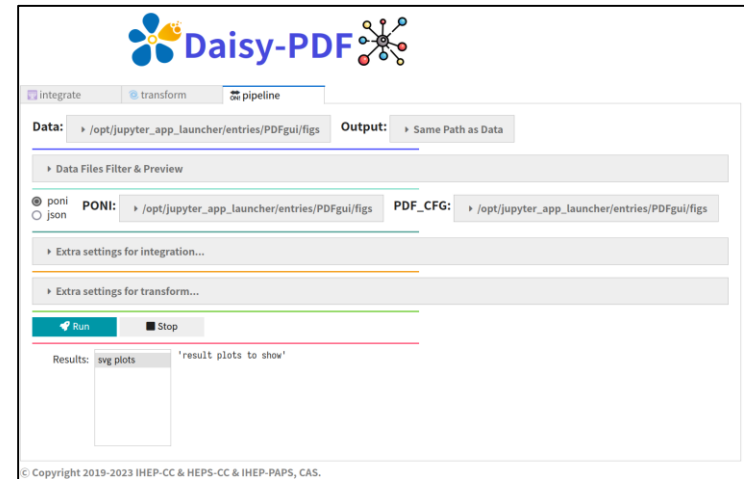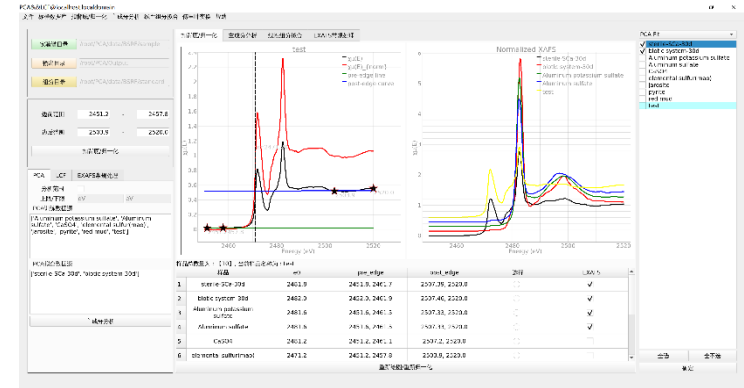
R&D: HERD Experiment

The High Energy cosmic Radiation Detection facility (HERD).
- It will be installed on the China Space Station, plans to launch in 2027.
- Intended to detect dark matter, cosmic ray and high-energy gamma rays.

Two work groups are set up for HERD.
- **Distributed Computing System.**
  - Serve data production and data insert checking.
  - Use DIRAC-Rucio Integration Model as computing model.
- **Data Management System.**
  - Whole lifecycle management of simulated data and experiment data.
  - Organize and manage calibration data and parameters of detectors.
  - Monitor the operation status of payloads, trigger alarms when an error occurs.
  - The workflow is used to implement the whole data processing flow.
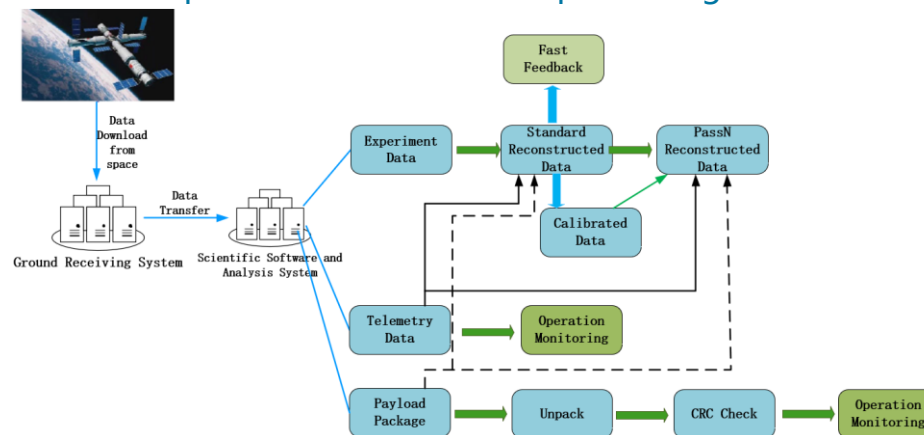
Summary

IHEP local cluster is running smoothly during last 6 months.
- Castor is retired and replaced by EOS-CTA.
- DPM is retiring, and LHC data will migrated to EOS in May 2023.

LHCb IHEP Tier2 is upgrading to Tier1.
- LZU is constructing LHCb Tier2.
- Upgrade will be finished in June 2023.

Some R&D are progressed as plan.
- Quantum computing developing platform.
- AI platform.
- Computational storage server and application.
- HEPS and HERD experiment software and computing system.

Thank you!

Backup

HTCondor Status

Job Statistics (2022.10-2023.03):
◦ **Total Job Number: 78,020,436 Jobs,**
◦ **Total Walltime in hours: 161,129,585 CPU Hours,**

Competed jobs

undefined ucas u07 physics panda offlinerun lhcb lhaaso
junovalprod juno hxmt higgs herdrun herd gecam dyw

Total completed jobs

	Value	Percent
lhaaso	34.63 Mil	45%
physics	20.46 Mil	26%
juno	8.12 Mil	10%
dyw	7.07 Mil	9%
higgs	1.87 Mil	2%
atlas	1.35 Mil	2%
hxmt	1.23 Mil	2%
cms	877.61 K	1%
ucas	773.41 K	1%
offlinerun	410.85 K	1%
lhcb	353.33 K	0%
gecam	155.50 K	0%
alicpt	66.30 K	0%
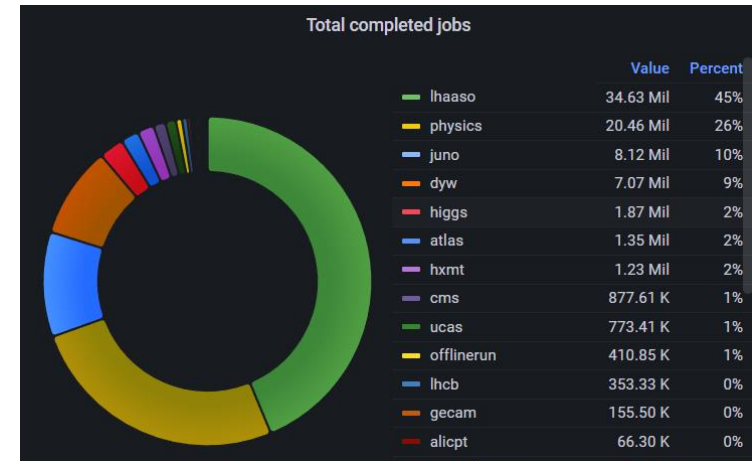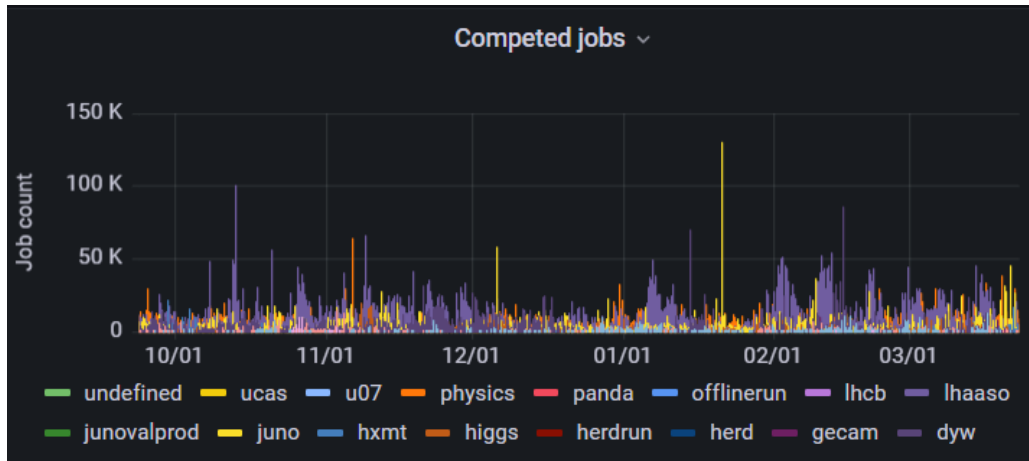
Slurm Cluster Status

Resources:
- 233(+5) worker nodes,
- 10K CPU cores, 250(+40) GPU cards.

Applications:
- 9 CPU APPs, 11 GPU APPs.

Job Statistics (2022.10-2023.03):
- 137.9K CPU jobs, 7.2M CPU hours,
- 113.1K GPU jobs, 541.3K GPU hours.

Num. of GPU Jobs of GPU_APP groups
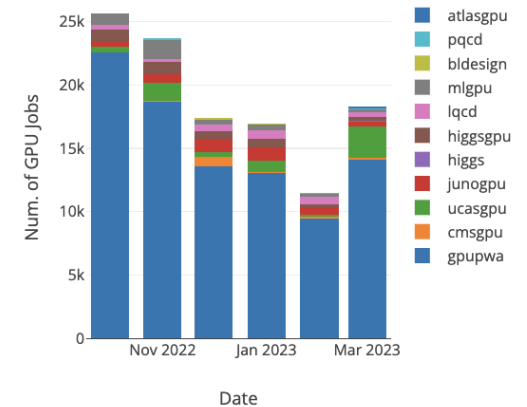
Num. of CPU Jobs of CPU_APP groups
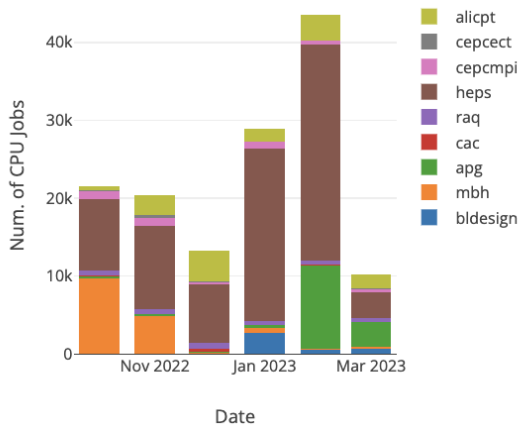
CPU Hours of CPU_APP groups
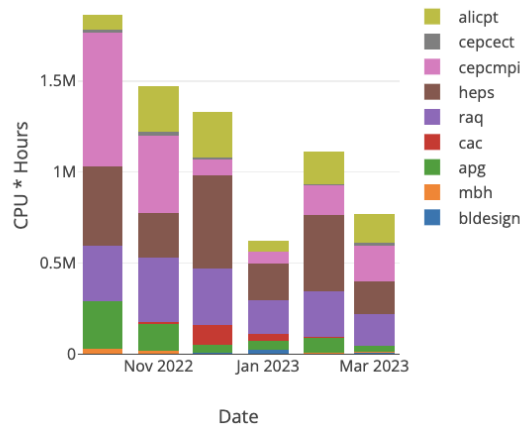
GPU Hours of GPU_APP groups
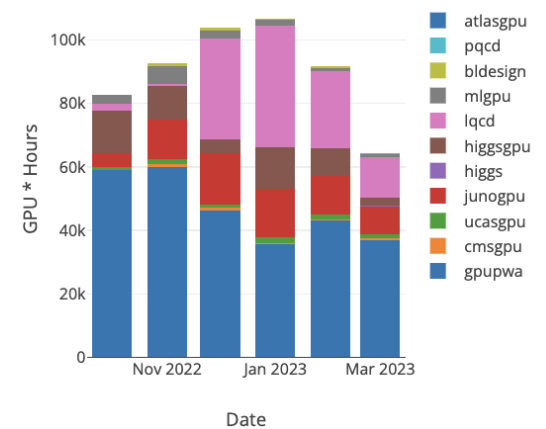
Distributed Computing

Services:
- VO: BESIII, JUNO, HERD, CEPC.
- Services: DIRAC and Rucio.

Resources at IHEP:
- ~2000 cores CPU, ~500 cores ARM.
- ~4.3 PB disk storage, ~2 PB tape storage.

Statistics (since 2022.10):
- 2.62 million jobs.
- 199 TB data transfers.

Development:
- Rucio starts to support HERD with an API integrated to experiment software.
- IAM with Sci-token and VOMS AA is ready for JUNO and HERD.

Cumulative Jobs by Site
25 Weeks from Week 39 of 2022 to Week 12 of 2023

Max: 2.62, Min: 0.00, Average: 0.55, Current: 2.62

|					
GRID.JINR-CONDOR.ru	1.0	GRID.MANCHESTER.uk	0.0	CLOUD.IHEPCLOUD.cn	0.0
GRID.IHEP.cn	0.9	CLUSTER.CSNS.cn	0.0	GRID.JINR.ru	0.0
GRID.INFN-CNAF.it	0.4	CLOUD.INFN-PADOVANA.it	0.0		
GRID.IN2P3.fr	0.3	CLUSTER.SDU.cn	0.0		

Generated on 2023-03-26 09:55:03 UTC

Transfered data by Channel
25 Weeks from Week 39 of 2022 to Week 12 of 2023

Max: 199, Min: 0.00, Average: 76.0, Current: 199

DIRAC.Client.cn -> IHEP-JUNOEOS	60.7	CNAF-HTTPSTORM -> CNAF-TAPE	4.3
GRID.IHEP.cn -> IHEP-JUNOEOS	36.6	IN2P3-DCACHE -> IN2P3-TAPE	4.3
GRID.JINR-CONDOR.ru -> JINR-EOS	24.2	IN2P3-TAPE -> DIRAC.Client.fr	3.1
GRID.INFN-CNAF.it -> CNAF-STORM	13.6	CLOUD.INFN-PADOVANA.it -> CNAF-STORM	2.9
JINR-EOS -> IHEP-JUNOEOS	11.8	CNAF-TAPE -> DIRAC.Client.it	2.1
GRID.IN2P3.fr -> IN2P3-DCACHE	11.2	IN2P3-DCACHE -> CNAF-HTTPSTORM	2.1
CNAF-STORM -> IHEP-JUNOEOS	6.8	DIRAC.Client.cn -> IHEP-STORM	0.7
CLUSTER.CSNS.cn -> IHEP-JUNOEOS	6.4	DIRAC.Client.it -> IHEP-JUNOEOS	0.5
IN2P3-DCACHE -> IHEP-JUNOEOS	5.2	... plus 105 more	

Generated on 2023-03-26 09:54:18 UTC

Disk Storage: Lustre

Instances:
- BES, JUNO, HXMT, CEPC, HEPS, etc.
- 22 instances (+1 new)

Resources:
- +9PB capacity, +41M files

Statistics:
- Read: 15.6 GB/s peak
- Write: 604 MB/s peak

Read Throughput

		Mean	Last *	Max
■	BES	5.90 GB/s	13.5 GB/s	15.6 GB/s
■	DYB	112 MB/s	28.6 MB/s	1.36 GB/s
■	JUNO	47.3 MB/s	7.68 MB/s	4.29 GB/s
■	Ali	20.9 MB/s	101 kB/s	645 MB/s
■	MBH	318 kB/s	0 kB/s	50.1 MB/s
■	HXMT	1.72 MB/s	5.83 MB/s	71.2 MB/s
■	HEPS	1.49 MB/s	428 kB/s	62.9 MB/s

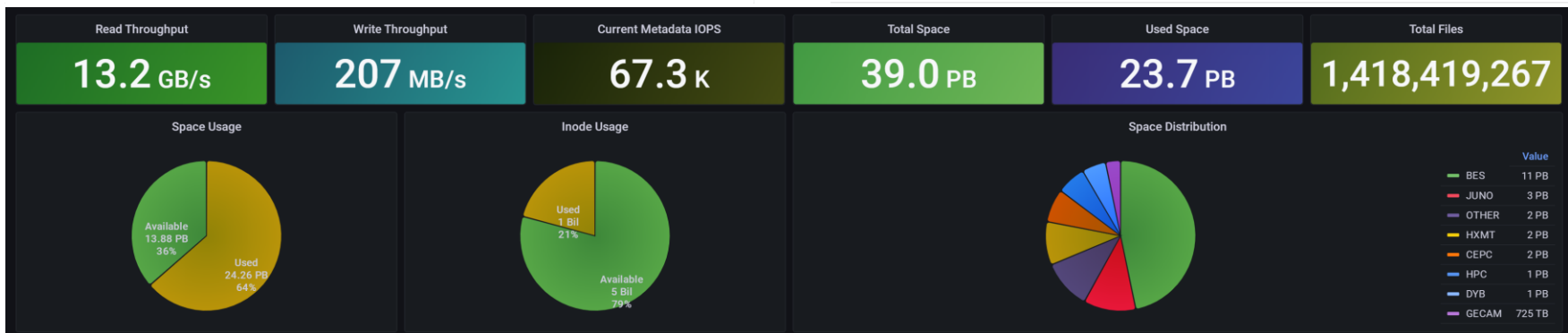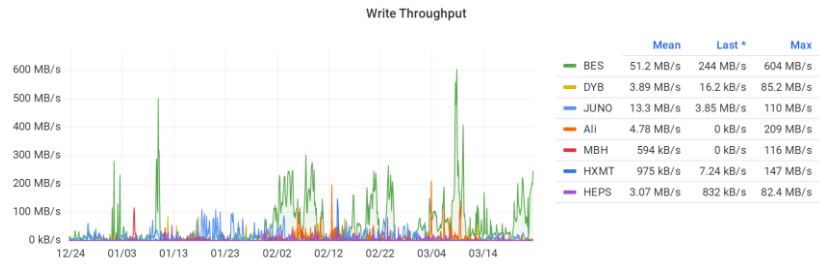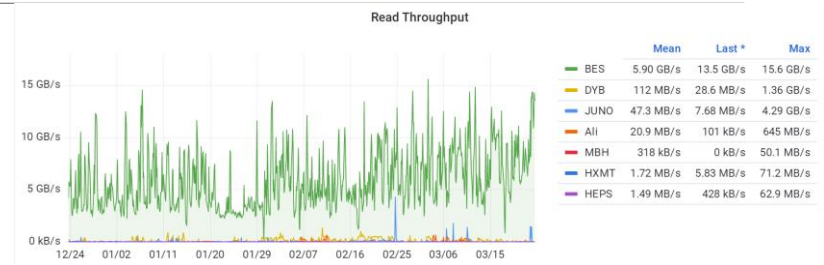Write Throughput

		Mean	Last *	Max
■	BES	51.2 MB/s	244 MB/s	604 MB/s
■	DYB	3.89 MB/s	16.2 kB/s	85.2 MB/s
■	JUNO	13.3 MB/s	3.85 MB/s	110 MB/s
■	Ali	4.78 MB/s	0 kB/s	209 MB/s
■	MBH	594 kB/s	0 kB/s	116 MB/s
■	HXMT	975 kB/s	7.24 kB/s	147 MB/s
■	HEPS	3.07 MB/s	832 kB/s	82.4 MB/s

Read OPS

Write OPS

Read Throughput	Write Throughput	Current Metadata IOPS	Total Space	Used Space	Total Files
13.2 GB/s	207 MB/s	67.3 K	39.0 PB	23.7 PB	1,418,419,267

Space Usage

Available 13.88 PB 36%
Used 24.26 PB 64%

Inode Usage

Used 1 Bil 21%
Available 5 Bil 79%

Space Distribution

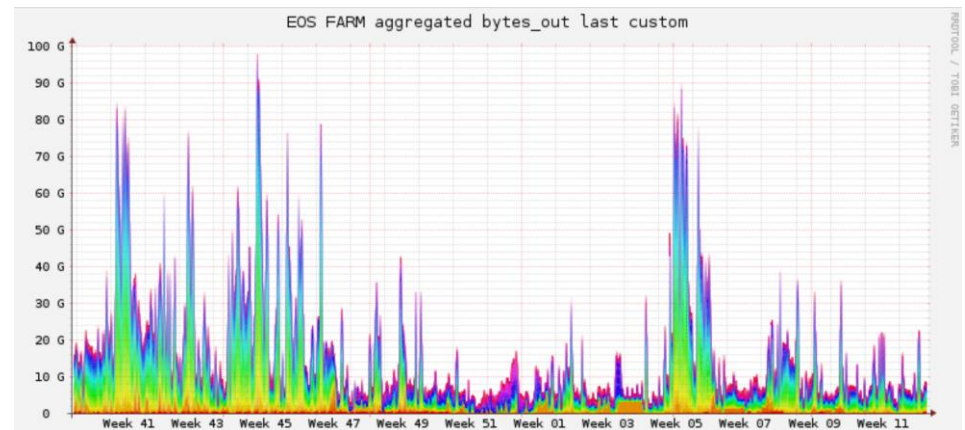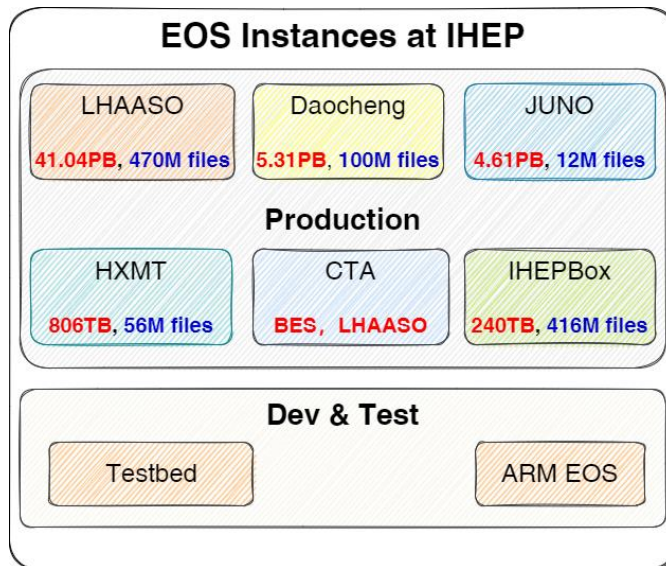|		Value
■	BES	11 PB
■	JUNO	3 PB
■	OTHER	2 PB
■	HXMT	2 PB
■	CEPC	2 PB
■	HPC	1 PB
■	DYB	1 PB
■	GECAM	725 TB

Disk Storage: EOS

Instances:
◦ **6 instances, support 3 experiment, IHEPBox and CTA,**

Resources:
◦ **~50PB Raw capacity,**
◦ **~700M files.**

EOS Instances at IHEP

LHAASO	Daocheng	JUNO
41.04PB, 470M files	5.31PB, 100M files	4.61PB, 12M files

Production

HXMT	CTA	IHEPBox
806TB, 56M files	BES, LHAASO	240TB, 416M files

Dev & Test

Testbed		ARM EOS

EOS FARM aggregated bytes_out last custom

Tape Storage: EOS-CTA

Status:
- All experiments have adopted EOS-CTA for backup,
- LHAASO, HXMT, BESIII , JUNO & HEPS.

Tier1 Construction:
- Instances for LHCb TIER 1 & WLCG under construction,
- Using EOS as the Storage Element.

CTA Transmission System(CTS):
- For backup from EOS/Lustre,
- Integrated with Redis, Kafka and RabbitMQ.

Exp	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