

Rucio Status at IHEP

Xuantong Zhang (zhangxuantong@ihep.ac.cn)

Computing Center of Institute of High Energy Physics,
Chinese Academy of Sciences





Outline

Introduction

Rucio deployment at IHEP

Experiences integrations and development at IHEP

Future plan

Summary

IHEP Introduction



The largest fundamental research center in China.

42 K CPU cores, 254 GPU cards to for more than 10 experiments

- HTCondor cluster runs for HTC jobs (34K CPU cores)
- Slurm cluster runs for HPC jobs (8K CPU cores + 250(+40) GPU)
- Distributed computing, WLCG, DIRAC etc., (~100 kHS06, 2K cores at IHEP)

81.6 PB disk storage, 55.1 PB tape storage

- Lustre (31.01 PB) and EOS (60.58 PB),
- EOSCTA for tape storage (55.1 PB).

Network

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

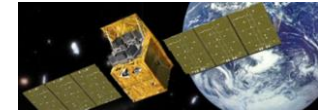
Chinese located or IHEP driven experiments



BESIII (Beijing Spectrometer III at BEPCII)



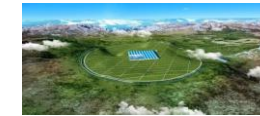
JUNO (Jiangmeng Underground Neutrino Observatory)



HXMT (Hard X-Ray Moderate Telescope)



SNS (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



Rucio History at IHEP



IHEP Computing Center start investigating and researching Rucio since early 2019.

Timeline:

- **2019:**
 - Start investigating and researching.
- **2020:**
 - Invitation talk for JUNO from Rucio Community.
 - Deploy Rucio testbed for JUNO.
- **2022:**
 - Deploy Rucio testbed for HERD.
 - Start IHEP Rucio extensions development.
- **2023:**
 - Move to Rucio production instance for HERD.
 - Deploy Rucio testbed for CEPC.
- **2024**
 - First time join Rucio Workshop.

Rucio Groups at IHEP:

- **Stuff (2):**
 - Xuantong Zhang, since 2020.
 - Xiao Han, since 2024.
- **Experiment coordinator (1):**
 - Xiaomei Zhang, since 2019, for JUNO/CEPC.

Rucio Deployment



3 instances:

- **HERD in production.**
- JUNO, CEPC in testbed.

Rucio system:

- With **v1.29.4.**
- Managed by **docker compose.**
- Daemons list (right figure shows).

Environment:

- For each instance, all daemons lives **in 1 physical machines.**
- Would like to try k8s, but no enough machines and experiences.

```
"rucio-abacus-account"  
"rucio-abacus-collection-replica"  
"rucio-abacus-rse"  
"rucio-activemq"  
"rucio-atropos"  
"rucio-conveyor-finisher"  
"rucio-conveyor-poller"  
"rucio-conveyor-submitter"  
"rucio-conveyor-throttler"  
"rucio-graphite"  
"rucio-hermes"  
"rucio-init"  
"rucio-judge-cleaner"  
"rucio-judge-evaluator"  
"rucio-judge-injector"  
"rucio-judge-repairer"  
"rucio-kronos"  
"rucio-minio"  
"rucio-necromancer"  
"rucio-reaper"  
"rucio-ruciodb"  
"rucio-server"rucioAPI  
"rucio-transmogrifier"  
"rucio-undertaker"  
"rucio-webui"
```

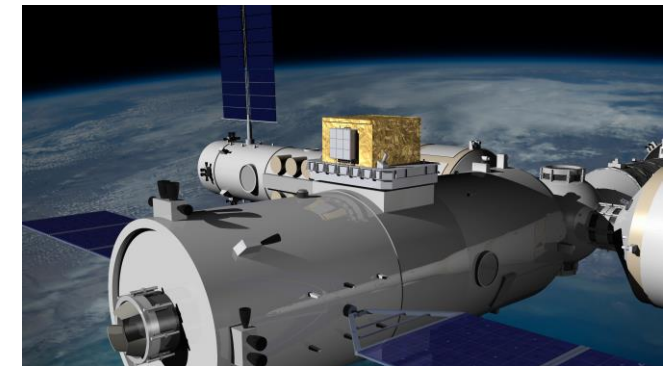
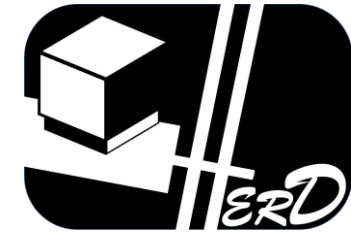
HERD Experiment Introduction



High Energy cosmic-Radiation Detection facility (HERD),

- A space particle astrophysics experiments, will run in the Chinese Space Station for **>10 years since 2027.**

Science Goal	Type	Contribution to Physics	Methods
Precision measurement of cosmic ray electron flux and dark matter search	Core	Key contribution to solve one of the most important puzzle for astronomy and physics: dark matter	Precision flus measurement of high energy electron and gamma.
Origin, acceleration and propagation of cosmic rays	Core	Key contribution to the origin of cosmic rays	Measurement of cosmic ray nuclei up to $Z=28$ to the highest energy
High energy gamma rays all-sky survey and monitoring		Search and identify gamma ray source , understand the physics of extreme conditions in the universe; search for new physical signals	Wide energy range, High precision measurement of Gamma rays





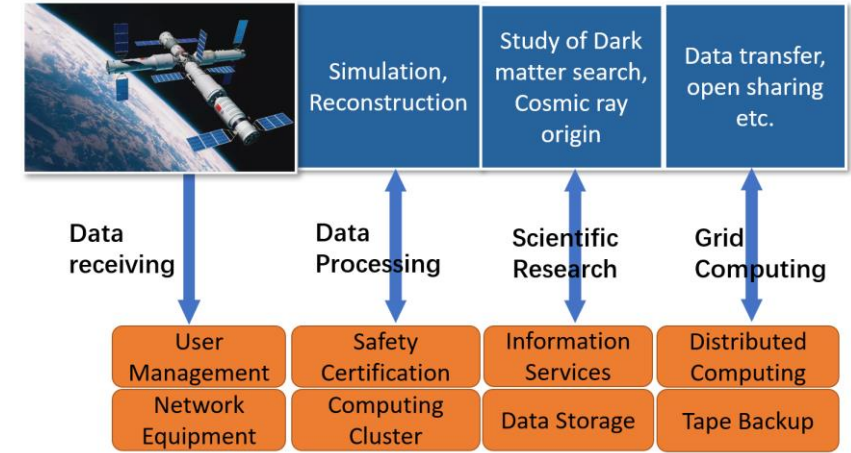
HERD Computing Model

Tier Model for HERD experiment,

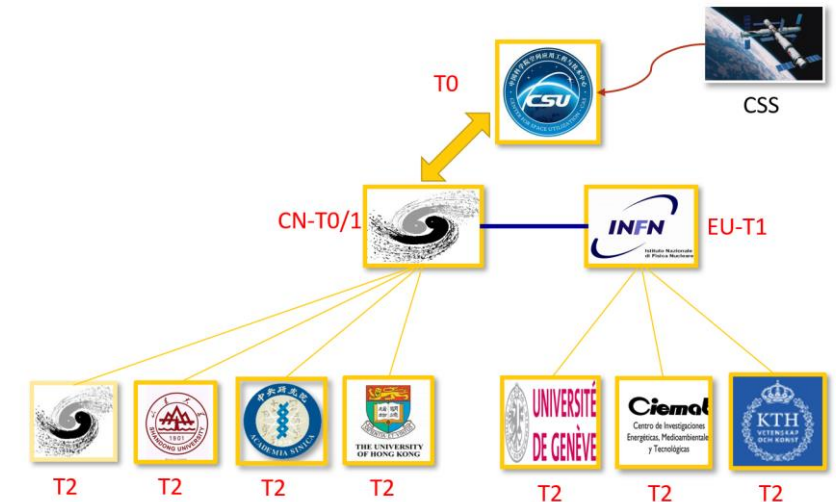
- Tier-0: raw data acquisition and storage (CSU, IHEP),
- Tier-1: regional center sites for SIM and REC data storage (IHEP-CN, INFN-EU),
- Tier-2: data processing sites (CN, EU sites).

Data Volumes,

- **>30PB in 5 years** or **>90PB in 10 years**,
- Rucio work as main data management system involved in data processing and distribution.



Scientific Data Platform



Data type	Data size (PB)			Computing (CPU Core)		
	5 years	10 years	Site	5 year	10 year	Site
Flight Data	2	6	T0, T1	-	-	T0
Standard Reconstruction	2.5	7.5	T0, T1	200	400	T0
Data transmission control system	1	2	T0	300	600	T0
PassN reconstruction	5 (2 version)	15 (2 version)	T0, T1	1000	3000	T0
Simulation data	5	15	T0, T1	4000	8000	T0
Analysis Data	2	4	T1	2000	4000	T1
Summary	15.5+16.5	45.5+47.5		7500	16000	



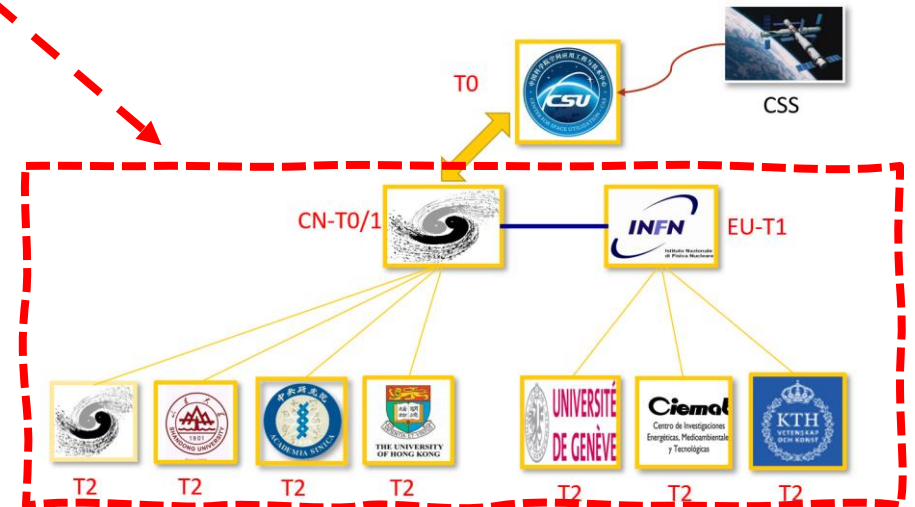
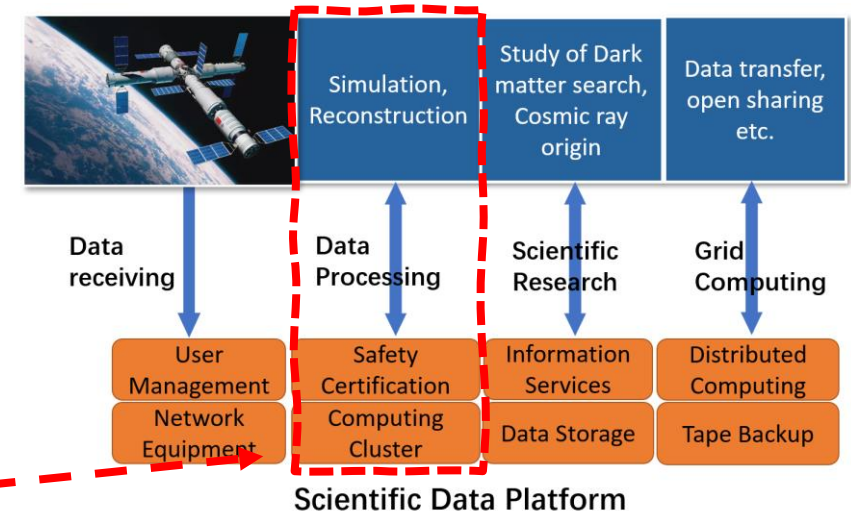
HERD Computing Model

Tier Model for HERD experiment,

- Tier-0: raw data acquisition and storage (CSU, IHEP),
- Tier-1: regional center sites for SIM and REC data storage (IHEP-CN, INFN-EU),
- Tier-2: data processing sites (CN, EU sites).

Data Volumes,

- **>30PB in 5 years** or **>90PB in 10 years**,
- Rucio work as main data management system involved in data processing and distribution.



Data type	Data size (PB)			Computing (CPU Core)		
	5 years	10 years	Site	5 year	10 year	Site
Flight Data	2	6	T0, T1	-	-	T0
Standard Reconstruction	2.5	7.5	T0, T1	200	400	T0
Data transmission control system	1	2	T0	300	600	T0
PassN reconstruction	5 (2 version)	15 (2 version)	T0, T1	1000	3000	T0
Simulation data	5	15	T0, T1	4000	8000	T0
Analysis Data	2	4	T1	2000	4000	T1
Summary	15.5+16.5	45.5+47.5		7500	16000	



HERD Policy (1/2)

Customized grid data file catalog namespace,

- Inspired by BELLE2(DIRAC), to make data logic name closer to local data, follow normal POSIX rules, has 3 types:
 - Data container, to contain other containers and datasets,
 - Dataset, to collect files,
 - Data file, basically ROOT files.
- **Scopes** are working as **data status zones**, so data types could be distinguished by its name,
 - Temp, Valid, Corrupt, etc.
- Don't know if it will be any performance issue.

SCOPE:NAME	[DID TYPE]
temp:/herd/user/z/zhangxt	DIDType.CONTAINER
temp:/herd/user/z/zhangxt/	DIDType.DATASET
temp:/herd/user/z/zhangxt/opt/herd/proton-center-E2.7-1_20TeV-34621161.0.root	DIDType.FILE
temp:/herd/user/z/zhangxt/output1-test.g4mac.root	DIDType.FILE

Namespace Component	HERD Namespace Policy
Name	Linux-like directory and file path
Scope	Defined as data status in data flow(Temp, Valid, Corrupt)
Dataset	Collection of all Files in a directory
Container	Collection of all sub-directories (=datasets) in a directory

HERD Policy (2/2)



User permission is controlled by group identity via Rucio permission plugins,

- HERD only have group user account,
 - As HERD only allow user downloading files from Rucio, all user add their identities to 'normal' user.
 - 'Production' user could add, delete, modify data in HERD.
 - All production users should be well-trained to manage file data.
 - Data processing flow using 'production' user as its identities.

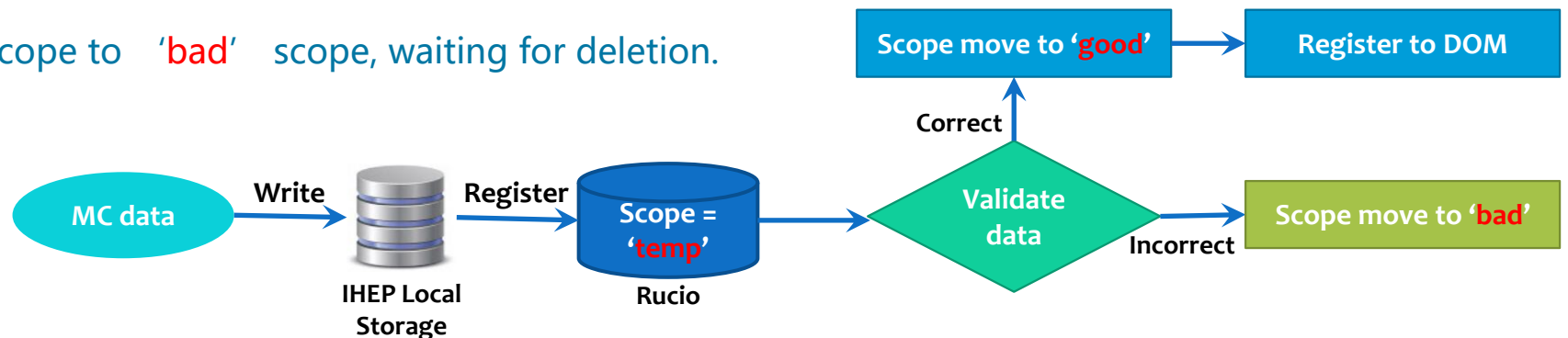
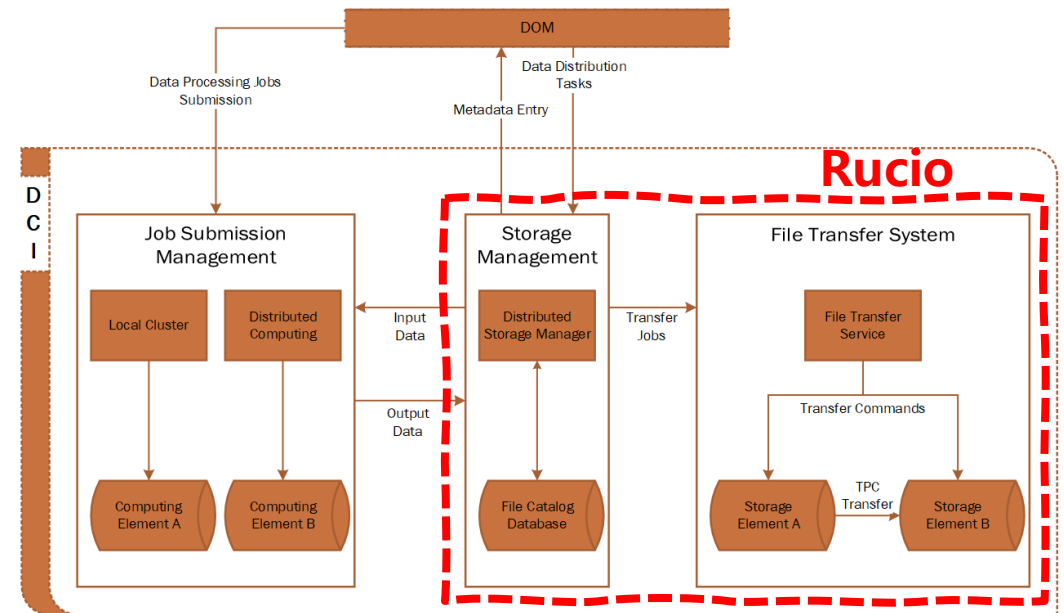
Permissions	Normal user	Production user	VO Administrator
File downloading	Allowed	Allowed	Allowed
File uploading, renaming and deleting	Forbidden	Allowed	Allowed
Metadata creating and modifying	Forbidden	Allowed	Allowed
Rules creating and deleting	Forbidden	Allowed	Allowed
Datasets creating and modifying	Forbidden	Allowed	Allowed
Data status modifying	Forbidden	Allowed	Allowed
Site configuration management	Forbidden	Forbidden	Allowed
Account management	Forbidden	Forbidden	Allowed
System configuration	Forbidden	Forbidden	Allowed



HERD Dataflow Integration

HERD data processing are all based on workflow,

- **DOM, meta database and workflow system.**
 - DOM has workflow synchronizing file catalogs from Rucio.
 - DOM has workflow trigger data processing jobs.
 - In data processing jobs, data are downloaded/uploaded by Rucio.
 - DOM triggers data validation workflow to manage data distribution.
- **MC dataflow as example:**
 - Register all raw MC data to 'temp' scope,
 - Data validation program use APIs to validate whether data are good.
 - If good, move scope to 'good' , then provide it to metadata registering.
 - If not good, move scope to 'bad' scope, waiting for deletion.





HERD RucioAPI

We are developing a HERD workflow oriented API,

- For both experiment software data access in jobs and DOM workflow.
- Merged to HERD software and workflow system.

Provides methods for:

- **Formatted metadata methods for DOM system, keys includes:**
 - Production batch, log file path, job finished time, etc.
 - Which could not get directly from remote jobs.
- **Method not directly provided from Rucio commands:**
 - Scope modification.
 - File removal.
 - Batch files upload with divided backend jobs or submit to local computing cluster.
 - Automatic container creation based on 'HERD' policy.
- **Some daemons:**
 - Automatic account synchronizer from IHEP-SSO and HERD-IAM.
 - Automatic register and rules creation (under development).
- **Other common Rucio methods but packaged in a better model for HERD production.**



IHEP Rucio Transfer Plugins

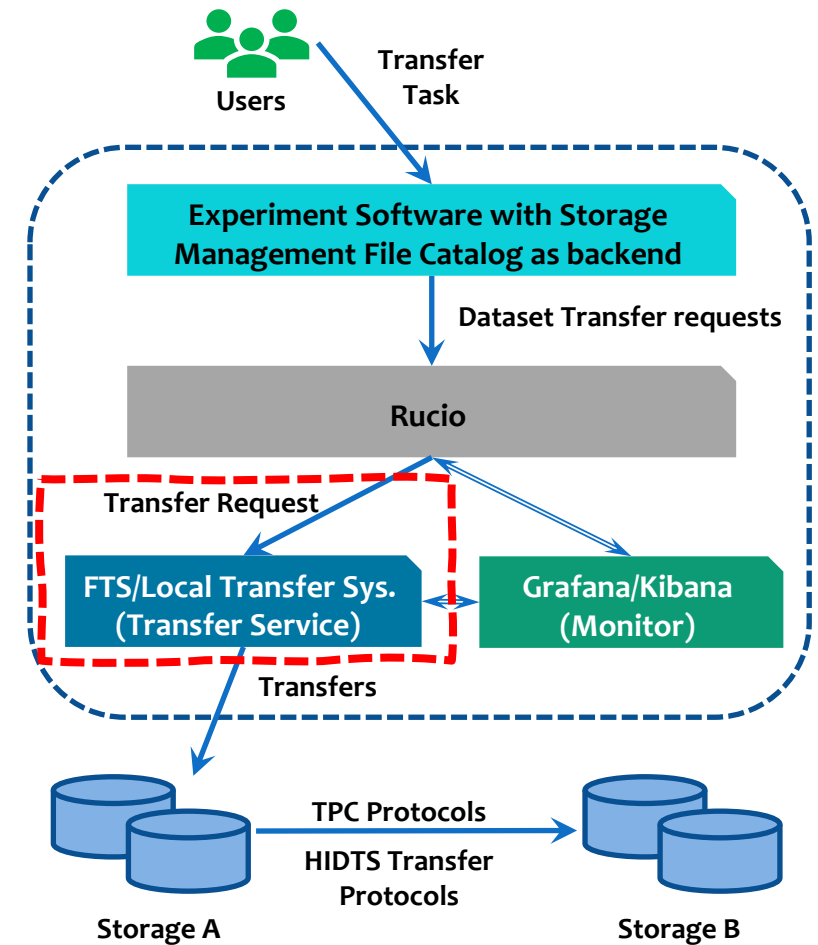
We are also developing some Rucio Transfer Plugins for IHEP local data transfer system (HiDTS).

- Working as another FTS plugins but for HiDTS, an IHEP self-developed data transfer system.
- HiDTS uses a commercial data transfer system as backend,
- But IHEP-CC developed a RESTful API for HiDTS, allowing user submits local storage data transfer.
- We are developing the plugins with allow Rucio use HiDTS as transfer system.

So that we could support more local storages,

- IHEP has lots of storage sites not supporting normal protocol such as Xrootd or WebDAV...
- Serving for future non-WLCG type experiments or big science devices.

The development progress is slow due to lack of man power unfortunately...





HERD Rucio Production Status

Production started since October 2023, running 12 months.

- Working on VOMS-proxy only, no sci-token.
- Only 1 down time for upgrade to Alma Linux 9.

Status so far,

- **RSE: 2 site (IHEP, Shandong),**
- **Accounts:**
 - User accounts: 37 identities,
 - Production accounts: 6 identities,
 - Root accounts: 2 identities.
- **Data volume:**
 - Files in total: 34263
 - Data size in total: 52.8 TB
- **1st user training in August, 2024.**

It is working in a small scale at the beginning of HERD experiment.



JUNO/CEPC Testbed

Rucio in JUNO and CEPC is still running in testbed.

Plan to integration with DIRAC?

- Yes!
- As IHEP is a long-time user in DIRAC community, and DIRAC is chosen as JUNO and CEPC' s distributed computing system.
- We still would like to integrate DIRAC and Rucio following experience of BELLE-2.





Future Plan

Thanks to new man power joining Rucio group at IHEP.

About deployment:

- Upgrade Rucio version.
- Move to k8s.
- Integration with DIRAC.

About development:

- Complete Rucio data management API based on experiment requirement.
- Promote the development of IHEP local transfer system plugins for Rucio.

Promote Rucio wider use:

- As the storage resources are more likely in local storage, data lake needs to based on IHEP local transfer system plugins.
- In future proton sources and AI platform system?



Summary

Rucio application at IHEP start from 5 years ago but in slow progress due to the man power lacking.

We started HERD Rucio in production and JUNO/CEPC Rucio in testbed.

Some experiments oriented Rucio modules and plugins is developed or under developing.

We are pushing Rucio application in more experiments, even more fields at IHEP.

Thanks for your attention!