JWanFS: A WAN Distributed Object Storage for Massive LHAASO Scientific Data

IHEP Computing Center

Siqi Hou

2026/10/24

Outline

1. Background & Challenges
2. Storage Mode Comparison
3. How to use JWanFS
4. Scientific Workflow Example with JWanFS
5. Latest Deployment Update

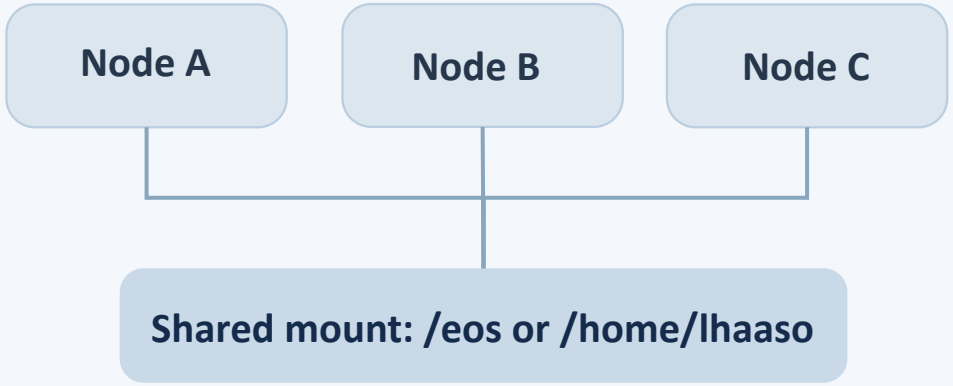
1 Background & Challenges

- LHAASO produces a large amount of scientific data.
- The data is produced, analyzed and reused in multiple places.
  - On a computing cluster
  - On a personal computer
  - With remote collaborators
- Previous solution:
  - Distributed share file system, connected to computing environment.
- User problem: How do I access and share data without being tied to one cluster?



2 Storage mode Comparison – EOS / Luster VS JWanFS

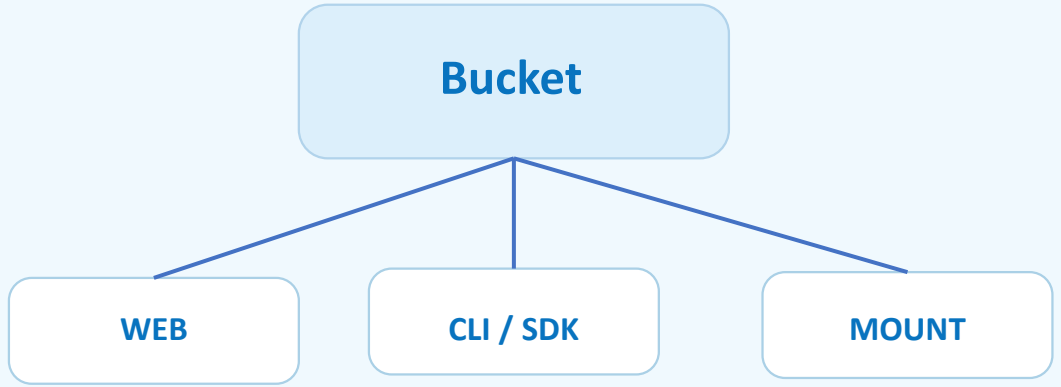
TRADITIONAL SHARED FILESYSTEM



PRE-MOUNTED ON COMPUTING NODE

- access files directly through familiar paths
- standard filesystem operations

JWanFS / OBJECT STORAGE



MOUNT ONLY IF NEEDED

- more freedom to manage their storage space
- access data from different environments

Users can choose different access methods for different workloads.
JWanFS complements Lustre / EOS by providing a more flexible, user-managed storage service.

3 How to use JWanFS

Provide Different Ways to Use It

Website

Browse, manage and preview

Browser · No install

Share URL

Send a result to a colleague

Controlled download link

jcli

Move files in batch jobs

Command line / automation

S3 SDK

Use data inside Python

Programmatic access

FUSE

Work through a directory

Optional local mount

AI conversational assistance is already available when you need help choosing or configuring an access method.

3 How to use JWanFS —— In Your Browser



Quick scientific inspection

- Regard every **bucket** as a disk(cloud drive).
- Explore the file list, check the content.
- Open a ROOT or HDF5 file without downloading.
- Decide what to download or analyze next.

3 How to use JWanFS —— By Sharing a URL

Two separate ways to sharing data

SEND ONE FILE

Share a file link

- Select an object in the Web UI
- Choose Share; set expiry and password
- Send the generated URL to a collaborator

File-sharing links are for files, not folders.

PUBLISH FILES OR FOLDERS

Resource Plaza

- **Publish** a file or directory as a resource
- Add descriptive metadata for discovery
- Useful for reusable group datasets

Review data sensitivity and permissions.

3 How to use JWanFS —— From the Command Line

What can jcli do?

- The official client for common OS
  - Linux, Windows, MacOS
- Authentication (jcli auth)
  - **AK/SK** access flexibly
- file management
  - list, upload, download
- Used in **shell scripts:** before and after analysis program

```
$ jcli help

$ jcli auth

$ jcli ls

$ jcli get ...

$ jcli put ...
```

Configure the selected gateway and bucket during authentication.
Follow “jcli help” for command arguments.

3 How to use JWanFS —— In Your Code

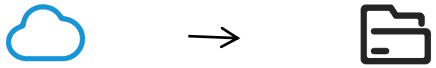
```
import os, boto3
from botocore.config import Config
s3 = boto3.client("s3",
  endpoint_url="https://fgws3-gfs.ihep.ac.cn",
  aws_access_key_id=os.environ["GFS_AK"],
  aws_secret_access_key=os.environ["GFS_SK"],
  region_name="us-east-1",
  config=Config(s3={"addressing_style":"path"}))
s3.download_file("my-bucket", "data/input.root", "input.root")
s3.upload_file("result.root", "my-bucket", "results/result.root")
```

Standard S3 Protocol

- Supported Programing Languages:
  - Python, Go, Java.
- Access storage directly in the code.
- Download dataset and upload checkpoints.
- Ask Your AI Coding assistant for help.
- Useful for scientific analysis and AI applications.
- Note: enable **path-style addressing**

3 How to use JWanFS —— As a Local Filesystem

Mount your storage bucket



Linux example

```
jcli auth
mkdir -p ~/gfs
jcli mount -f ~/gfs --daemon
ls ~/gfs
```

Once mounted, use familiar file operations such as ls and cp. Just like a local disk / directory.

Good for

Lightweight file access, occasional file copies , and manual organization.

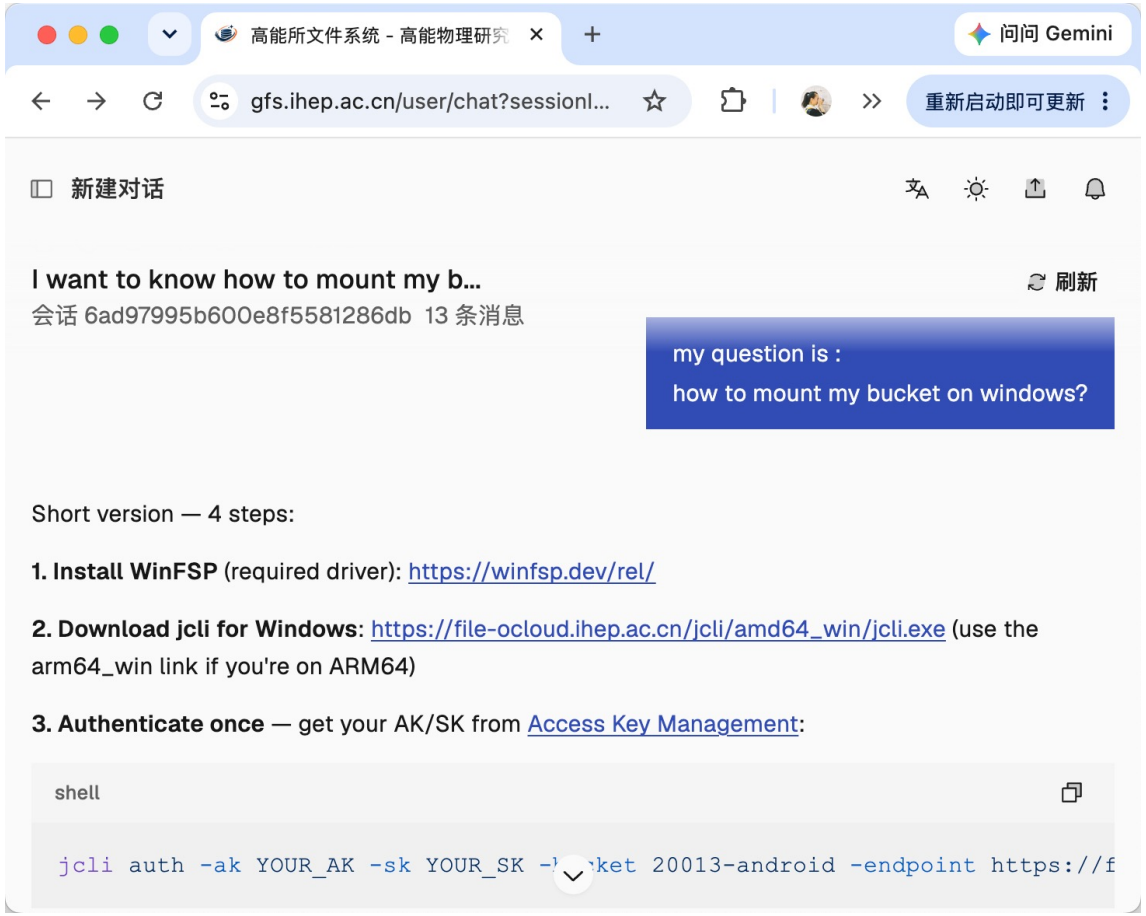
Avoid for

High-concurrency I/O, databases, and frequent small writes.

Mounting is supported on Linux, MacOS and Windows, and it is optimial.

3 How to use JWanFS —— Ask AI Assistant

AI conversational assistance is AVAILABLE NOW.



Why it helps

- **Choose the Right Tool**
  - Find the most suitable access method for your workflow.
- **Get Configuration Guidance**
  - Learn how to set up jcli, mounting, or SDK access.
- **Ask Follow-up Questions**
  - Get explanations tailored to your questions and environment.

3 How to use JWanFS —— Ask AI Assistant

MY TASK	BEST STARTING POINT	WHY
Quick ROOT / H5 inspection	**Website**	No local staging just to preview
Send a figure or file to a colleague	**Sharing link**	Time-limited / password-protected link
Batch data transfers and job scripts	**jcli**	Repeatable shell operations
Read/write inside an AI training script	**S3 SDK**	Programmatic objects and checkpoints
Browse objects as a local directory	**FUSE mount**	Convenient, but not full POSIX semantics
Not sure how to access my data	**AI assistant**	Personalized guidance - available now

CAUTION → High-frequency small writes / databases / tight POSIX requirements: prefer suitable compute storage.

4 Scientific Workflow Example with JWanFS

1. Inspect

ROOT / training datasets in JWanFS

2. Compute

Slurm/ HTCondor
Data Access method
jcli / SDK / FUSE

3. Save

Computing result upload to user‘s bucket

4. Share

Upload results and share with the team

How Does the Computing Job Access JWanFS? (Write it in your job script)

- **jcli — Stage Data to Scratch**

Download input files before analysis and upload outputs afterward. A practical default for traditional ROOT batch jobs.

- **S3 SDK — Direct Access in Code**

Read and write objects programmatically within scientific analysis or AI applications.

- **FUSE — Mount When Needed**

Access files through familiar paths where mounting is permitted; evaluate I/O performance for the workload.

5 Latest Deployment Update

LATEST DEPLOYMENT UPDATE

GFS cluster

8.12 PB

Dedicated **Tianfu cluster**

80TB

+15 storage nodes

added in Daocheng Computing Center

Gateway Update

Daocheng gateway

New gateway intended to support faster on-site data access.
Load Balance supported.

Performance measurements
Single Job writing > 500MB/s .

AVAILABLE NOW

AI conversational support

COMING SOON

New desktop client

Get started: https://gfs.ihep.ac.cn/ | **Guide:** https://cchelp.ihep.ac.cn/storage/gfs/

Thank you for your attention !
Q&A

手机扫码
即可访问