Open Source 西门子

Beijing · 2026

Oct. 29
Industrial Open Source

Oct. 30
Open Source Robotics

Agenda

开源基础设施
Open Source Infrastructure

2023



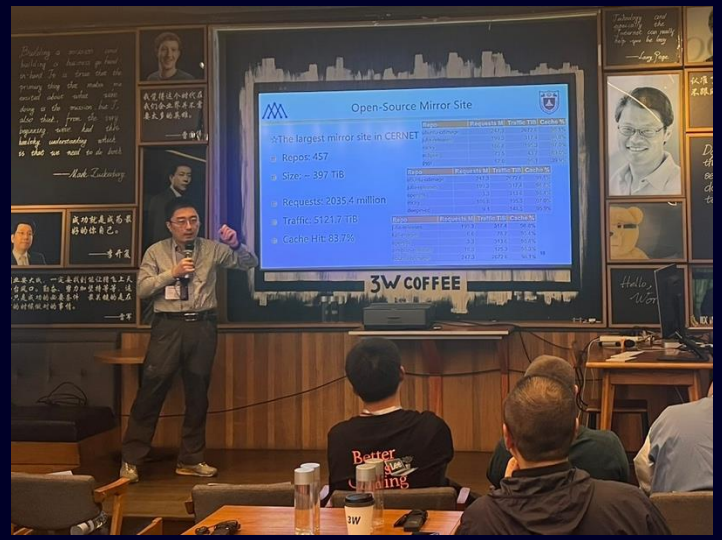
Shengqi Chen from Tsinghua

2024



Zhang Shiming DaoCloud

2025



Yao Ge from Nanjing University

Unrestricted | © Siemens 2026 | Daniel Bovensiepen | FT RPD CPC | 2026-10

SIEMENS

USTC Mirror Site: Past, Present, and Future

Qingyu Wu 吴擎宇
vertsineu@ustclug.org
Open Source @ SIMENS 2026
2026/10/29 @ Beijing

Outline

- What's USTC Mirror?
- History of USTC Mirror
- Current Tech Stack & Scale
  - Network & Hardware
  - Storage & Cache
  - Synchronization
  - Access Limit
- Future of USTC Mirror

About USTC Mirror

- Mirror site of **U**niversity of **S**cience and **T**echnology of **C**hina
- A leading open-source software mirror site in mainland China
- Maintained by USTCLUG (USTC Linux User Group)
- Site: https://mirrors.ustc.edu.cn
- Includes 150+ repositories:
  - Linux distros (Debian is our favorite)
  - linux.git
  - CTAN, VLC, uv...



Start from Debian enthusiasts...

- 2000: Some Debian enthusiasts at USTC started providing a mirror service with machines from their labs...
  - only included Debian
  - the machine moved between labs...
- 2002.5: The Debian mirror got its official domain debian.ustc.edu.cn!



瀚海星云 - 文章阅读　讨论区：**Linux** [Linux系统]　板主: SoRSs bible

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

发信人: Chern_Yf (thinknerd)，原信区: ustcbbs
标　题: New Debian GNU/Linux mirror
发信站: 中国科大BBS站 (Sun, 19 May 2002 16:22:46)，站内信件

由于 Yong_Q 将要毕业离开科大, debian mirror 以后将由我来维护.
现在工作已经完成, 小范围测试工作正常, 欢迎大家测试. sources.list:

ftp://debian.ustc.edu.cn/sources.list

这个 sources.list 针对 sid, 你可以自己修改一适应你的需求. 感谢
James_Zhang 提供域名.

--
※ 来源: 中国科大BBS站 [bbs.ustc.edu.cn]

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

Call for donations...

- 2005.6.20: USTCLUG launched an initiative to raise donations for the Debian mirror.
  - by 2005.10.6, 2922.05 CNY had been raised



瀚海星云 - 文章阅读　讨论区: **Linux** [Linux系统]　板主: SoRSs bible

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

发信人: Galoisx (在边沿上思索的人), 信区: Linux
标　题: 科大Debian站捐款最新更新[LUG]
发信站: 瀚海星云 (2005年10月01日13:46:03 星期六), 站内信件

根据LUG主页最新消息(10月1日):
http://lug.ustc.edu.cn/

News: 为科大Debian站募捐活动暂时停止 [2005.10.1]

截至2005年10月1号下午，我们总共收到2922.05￥捐款，暂时我们认为已经达到购买科大Debian站服务器的预算费用。所以，从现在起停止接受捐款，若有想要捐款的朋友，请和我们联系，我们对您的好意表示感谢。今后，若还需要捐款，我们将再发布公告。

从2005年6月份我们发布请求社区捐款的公告(倡议书)以来，我们得到科大，全国许多地方和国外朋友的热心关注和大力帮助。在此捐款暂时结束，即将购买新服务器的时候，我们对所有的朋友表示由衷的感谢，感谢大家对自由软件，对Linux，对Debian，对社区的帮助，对科大Debian站和科大Linux Users Group的帮助。谢谢你们，并祝福你们和你们的家人朋友身体健康，工作顺利，心情愉快。

关于最新的捐款信息，捐款使用和其他有关信息，请查看捐款主页http://lug.ustc.edu.cn/donate/lookup.html

PS. 祝大家节日愉快

--
※ 来源:·瀚海星云 bbs.ustc.edu.cn·[FROM: ip已隐藏]

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

Not only Debian...

- 2008.12.25: Another OSS mirror site, oss.ustc.edu.cn, went live with a server donated by Fengguang Wu (吴峰光); Novell also provided a 1.5 TB disk
- 2009.12: Cheng Zhang (张成) donated a 1 TB disk to the OSS mirror
- 2010.6: USTCLUG added a 2 TB disk to the OSS mirror


瀚海星云 - 文章阅读　讨论区：**Linux** [Linux系统]　板主：SoRSs bible

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

发信人：VF (VisualFool.net)，信区：Linux
标　题：OSS.ustc.edu.cn 开张大吉
发信站：瀚海星云 (2008年12月25日10:18:44 星期四)，站内信件

感谢吴峰光老大、Novell 公司和网络中心，一个新的镜像站点今天正式启用了。欢迎
大家通过 HTTP/FTP/Rsync 方式访问：

  IPv4域名：oss.ustc.edu.cn
  IPv6域名：oss6.ustc.edu.cn

目前上面有 OpenSuse 的镜像，以及 Debian 镜像的每日快照。我们准备再提供一些
其他的常用的开源软件或 Linux 发行版的镜像，为校内和教育网内同学提供方便，
促进开源软件的推广工作。目前具体的列表还没有确定，所以大家可以大声说出你希
望我们镜像什么你需要使用的开源软件或 Linux 发行版，我来根据硬盘空间的大小
和需求量的大小做出决定。

圣诞节快乐！

--
※ 来源：·瀚海星云 bbs.ustc.edu.cn·[FROM：ip已隐藏]

回复本文 | 上一篇 | 本讨论区 | 下一篇 | 转寄朋友 | 转至别板 | 同主题阅读 | 删除本文 | 修改本文 | 添加附件

The unified mirror site...

- 2011.4.8: USTCLUG gained the domain mirrors.ustc.edu.cn and a new machine (mirrors1) from USTC NIC (Network Information Center, 网络信息中心)
- 2011.4.15: The mirror site started providing services to the public, and the Debian and OSS mirrors were retired...
- 2013.1.16: New disk array from USTC NIC
- 2016.12.29: New machine (mirrors2) from USTC NIC
- 2019.6: An old machine (mirrors3) for the Library (returned in 2025)
- 2020.3.24: New machine (mirrors4) from USTC NIC
- ...

Machines currently in use

mirrors2 (2016) and mirrors4 (2020) are used for the USTC mirror

- mirrors2: secondary; cold repos, e.g. centos-vault, ubuntu-old-releases
- mirrors4: primary; almost all repos, e.g. Linux distros, open-source software

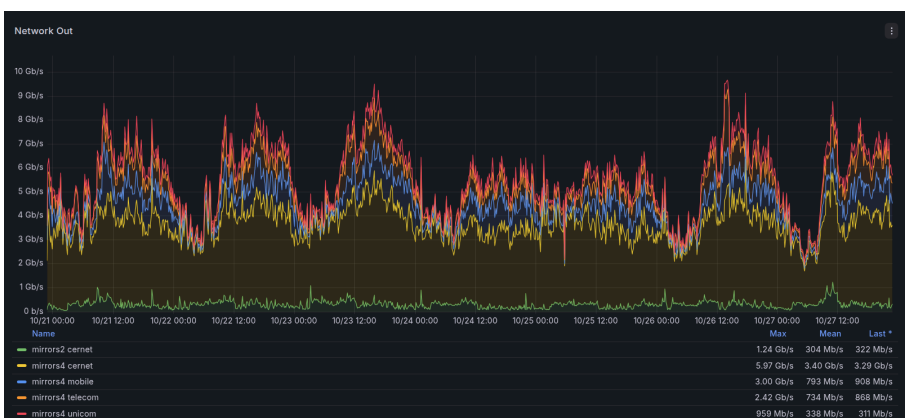
Overview:

- network:
  - 10 Gbps egress bandwidth, CERNET/Telecom/Unicom/Mobile shared
- storage:
  - 1 × 2 TB Samsung PM883 (SSD)
  - 12 × 10 TB HGST HUH721010AL (HDD)
    - 2 × ZFS RAIDZ2 (4 + 2), 80 TB usable

Statistics

All data is from 2026/10/20 to 2026/10/27

- Bandwidth usage:
  - HTTP(S) total: 4.0 Gbps on average, 9.8 Gbps max
    - IPv4: 3.2 Gbps on average, 8.7 Gbps max
    - IPv6: 0.8 Gbps on average, 1.6 Gbps max
  - rsync: 1.0 Gbps on average, 4.4 Gbps max
- Service distribution:
  - 88% HTTPS and 12% HTTP of all HTTP traffic, on average
  - rsync ≈ 16% of total egress (10.5 TiB/day vs 41–48 TiB/day HTTP)



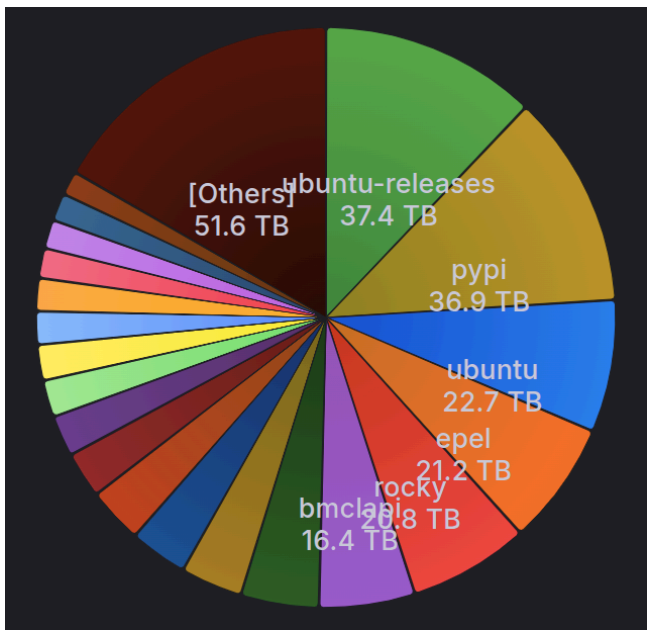
Statistics

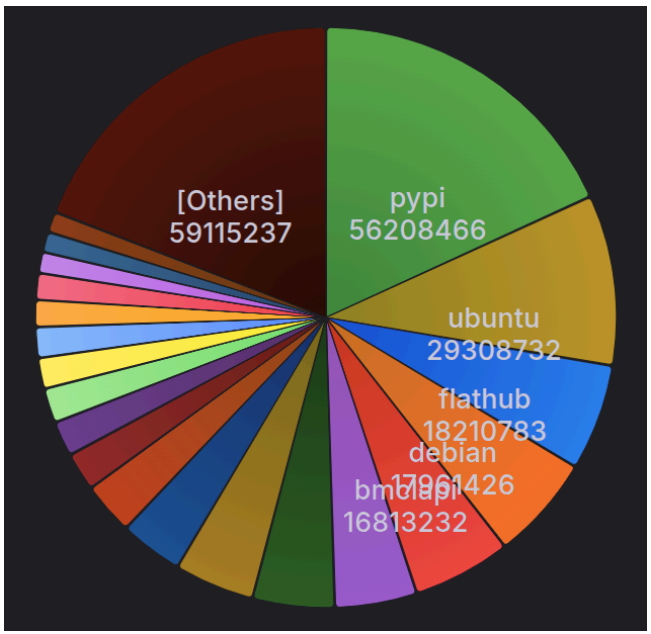
Top 3 repos by bytes:

- Ubuntu releases: 14.0 Mreq, 37.4 TB
- PyPI: 56.2 Mreq, 36.9 TB
- Ubuntu: 29.3 Mreq, 22.7 TB

Top 3 repos by requests:

- PyPI: 56.2 Mreq, 36.9 TB
- Ubuntu: 29.3 Mreq, 22.7 TB
- Flathub: 18.2 Mreq, 4.9 TB



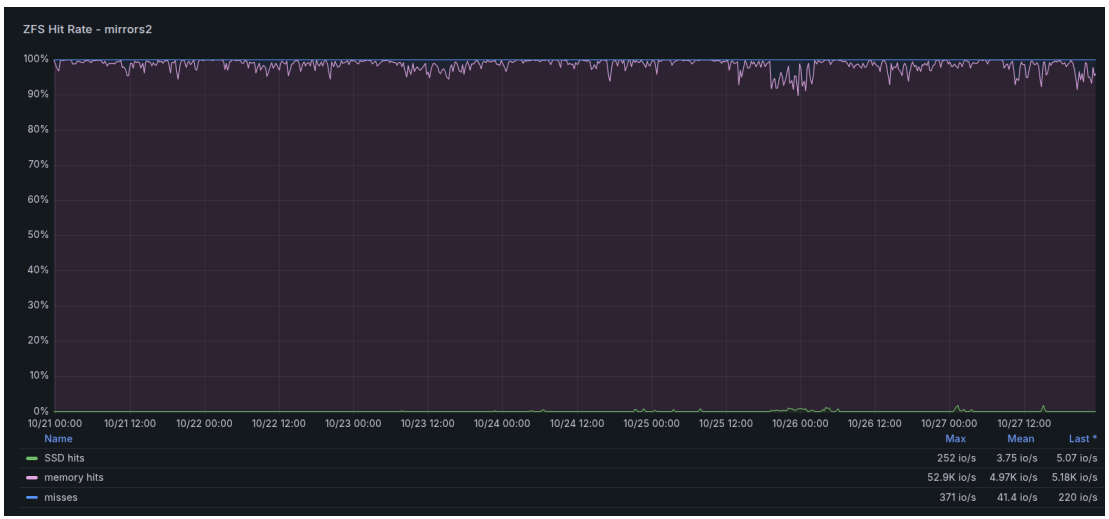


Storage & Cache

We don't have an all-SSD setup, but we can use ZFS for caching!

- ARC: RAM, 160–200 GB
  - 98.56% hit rate (hits 872 G / misses 12.7 G)
- L2ARC: SSD, 1.5 TB
  - read: ~3k IOPS, 14.1 MB/s
  - write: 5 IOPS, 212 KB/s
- ZFS: HDD, 80 TB
  - read: ~1k IOPS, ~450 MB/s
  - write: ~370 IOPS, ~20 MB/s

All data is from 2026/10/27

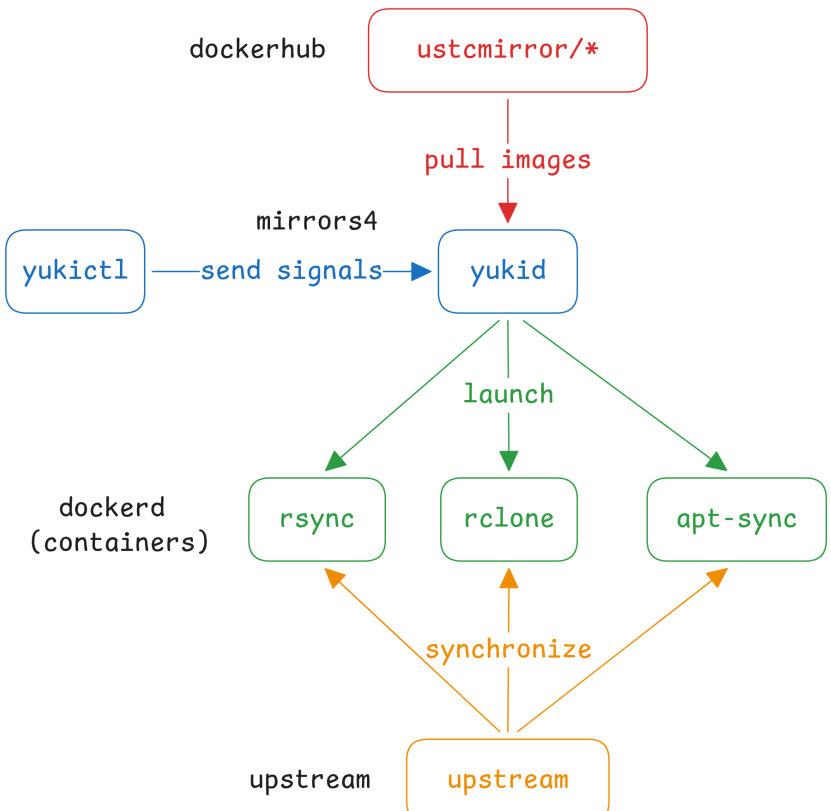


Synchronization

We have open-sourced our standardized synchronization process at github.com/ustclug:

- ustclug/yuki:
  - a scheduler that launches containers to synchronize on a schedule
  - e.g. `cron: 58 3-23/6 * * *`
- ustclug/ustcmirror-images:
  - Docker images that provide an independent, unified environment for synchronization
  - e.g. `docker.io/ustcmirror/rsync`
- rsync, rclone, apt-sync, … (inside the images):
  - the actual synchronization tools and scripts

Feel free to use them to set up your own mirror service!



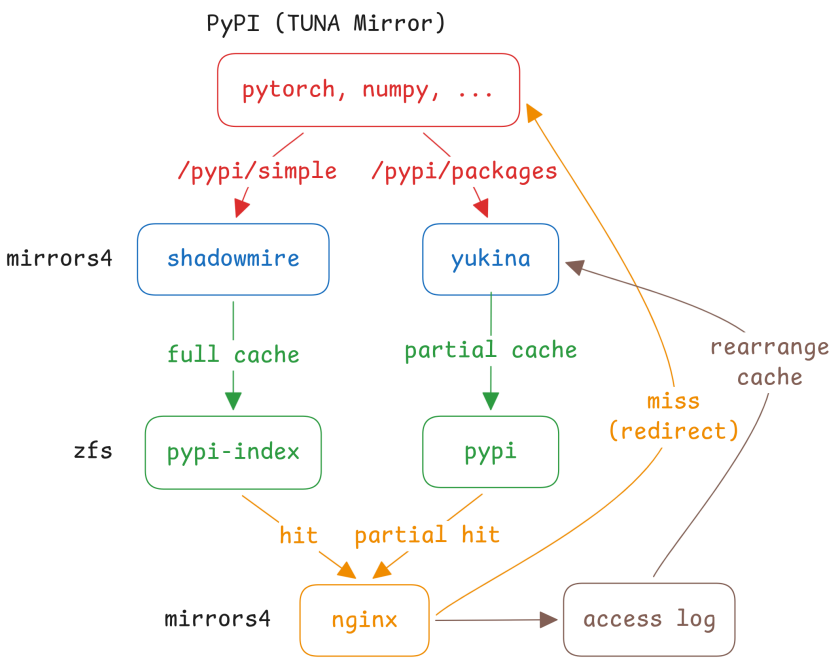
Synchronization

Sometimes, some upstreams are not mirror-friendly:

- ustcmirror/apt-sync: scripts for syncing upstream apt repositories
- ustcmirror/tsumugu: a program that syncs upstreams without HTTP HEAD requests, for lower overhead

Sometimes, some upstreams are too large to mirror:

- ustcmirror/shadowmire + ustcmirror/yukina: programs for syncing the PyPI upstream
  - shadowmire: syncs metadata from TUNA (Tsinghua Mirror)
  - yukina: caches packages from the daily nginx logs



Access Limit

We often encounter unfriendly users who consume too much bandwidth and too many connections, so we have countermeasures:

- iptables (raw): IPv4/IPv6 CIDR prefixes
  - e.g. for ports 80 and 443, at most 12 connections per block of 8 adjacent IPv4 addresses (CIDR /29)
- nginx + lua module (http): global, HTTP HEAD, HTTP Range, per IP…
  - concurrent connection count
  - requests per second
  - requests per day
- ustclug/rsync-proxy: global
  - e.g. at most 60 concurrent and 60 queued connections
- ustclug/git-queue: per IP
  - e.g. at most 4 connections per IP

Access Deny

Sometimes, access limits are not enough to stop unfriendly access from affecting normal users:

- iptables (raw): IPv4/IPv6 CIDR ranges
  - DROP connections from these ranges
- nginx + lua module (http): User-Agent
  - DROP connections from specific User-Agents, such as crawlers
- iptables (rsync):
  - DROP connections not from CERNET

If you claim to be a browser (UA starts with `Mozilla`), I will challenge you:

- For large files ending in .iso, .exe, etc.
- DROP connections that do not pass the challenge

Problems

We do not — and cannot — mirror everything:

- In ZFS, we have 109 TiB of total space, with 80+ TiB currently used
- We have cleaned up some upstreams that are not frequently used

We cannot afford many connections or much traffic:

- We have reached a load average of 100 (about 100 runnable processes on average)
- We have reached the 10 Gbps physical limit of our uplink
- We have encountered new unfriendly users who bypass our rate limits and blocks

We need more operations and maintenance volunteers to carry it on:

- ~5 members currently participate in maintenance

How to Solve

So, mirrors5 (2026) is coming soon:

- More disks!
- More cores and memory!
- More cables for larger bandwidth!

So, ClickHouse is used to analyze unfriendly traffic:

- From nginx access logs, rsync logs, …

So, we actively seek students who want to participate in maintaining the mirror site.

Linkage

MirrorZ (mirrors.cernet.edu.cn) is a union of mirror sites from many universities, with advantages over a single site:



- load balancing across all member sites
  - bandwidth pressure can be relieved
- automatically chooses the nearest/fastest site for users
- higher availability
  - if one site is occasionally down, just use another one
- ...

Other mirror sites near us:

- mirrors.tuna.tsinghua.edu.cn: Tsinghua University
- mirrors.nju.edu.cn: Nanjing University
- ...

What about AI?

We take a neutral stance on AI. You can:

- use AI to generate and review code
- use AI to start a useful new project

But you should not:

- commit or open a PR without human review
- dump AI slop on maintainers without understanding it
- put AI-generated code directly into production

You should always:

- be responsible for your AI-generated code



commented last month • edited

问题

`sync.sh` Step 4 的 `download_manifest()` 会跳过已存在的 manifest 文件，注释中的假设是 *"Manifest tags are immutable for a given version+revision+rebuild"*。

但这个假设与 Homebrew 的实际发布方式不符：**Homebrew 会向同一版本 tag 的 manifest 追加新的平台条目**。典型场景是新 macOS 发布后的 bottle 引导——例如 2026 年 9 月 Homebrew 为 macOS 27（`arm64_golden_gate`）构建 bottle 时，向 `cmake/manifests/4.4.3`（该版本自 8 月起未变过）追加了 `4.4.3.arm64_golden_gate` 条目。镜像上首次抓取的旧副本从此永远不会被刷新。

影响

自上游追加条目起，镜像上所有"版本号早于追加时间"的包的 manifest 都停留在旧内容。用户配置 `HOMEBREW_BOTTLE_DOMAIN` 指向本镜像、在新系统上安装这些包时，brew 在 manifest 中找不到当前平台的 bottle 条目，安装失败并回退到源码构建（源码路径还需访问 raw.githubusercontent.com，在部分网络环境下会直接失败）。

实测证据（2026-10-03）：

资源	镜像上的状态
`cmake/manifests/4.4.3`	Last-Modified 停留在 2026-08-26，缺 `arm64_golden_gate` 条目（官方 ghcr 同 URL 已含）
`cmake-4.4.3.arm64_golden_gate.bottle.tar.gz`	2026-09-10 已正常同步（bottle 文件本身没有问题）
`node/manifests/26.10.0_1`（09-25 新版本 tag）	条目齐全，安装正常 —— 印证只有"旧 URL"受影响

brew 侧的报错链：

Questions

Feel free to ask anything about USTC mirror!

Thanks for listening!

Open Source 西门子

Beijing · 2026

Oct. 29
Industrial Open Source

Oct. 30
Open Source Robotics

Agenda