



南京大学开源镜像站——开源基础设施

NJU Mirror - Open-Source Infrastructure



姚 舸

南京大学

人工微结构科学与技术协同创新中心

Collaborative Innovation Center of Advanced Microstructures

yaoge@nju.edu.cn my@yaoge123.com

YAO Ge

Nanjing University





Open Source



- 持续增长近乎无限的需求
- 持续降低永远不够的经费
- 人永远不够
- 一人全栈
- 通用、标准





e-Science Center



- 面向校内外的数字化服务
- ✓ 科研
- ✓ 教学
- ✓ 管理

人工微结构科学与技术协同创新中心

服务 sci.nju.edu.cn

Icon	Service Name	Description
	云盘	为您的数据保驾护航
	协同表格	一起来填表管理信息
	LaTeX	在线编译, 开箱即用
	代码托管	保留代码的点点滴滴
	超级计算	高性能科学与人工智能计算
	开源镜像	极速访问, 又全又大
	数据存储	稳定持久, 海量无忧
	远程控制	安全快捷, 远程直达
	网络测速	网速慢么?
	密码管理	一密掌控, 严防泄露
	网络授时	北斗铸时, 精确同步
	文档中心	帮助都在这里

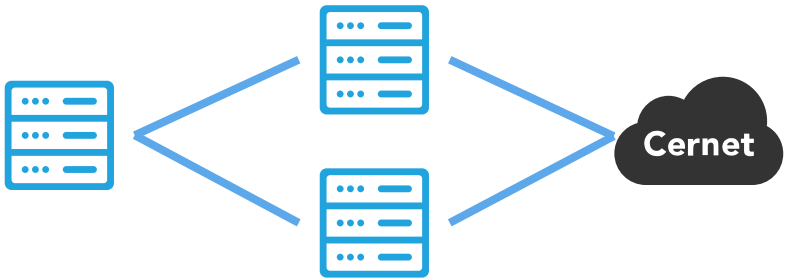


Open-Source Mirror Site



Domain name	IPv4(A)	IPv6 (AAAA)
mirror.nju.edu.cn	Y	Y
mirrors.nju.edu.cn	Y	Y
help.mirror.nju.edu.cn	Y	Y
help.mirrors.nju.edu.cn	Y	Y
*.mirror.nju.edu.cn	Y	Y
*.mirrors.nju.edu.cn	Y	Y
ipv4.mirror.nju.edu.cn	Y	N
ipv4.mirrors.nju.edu.cn	Y	N
v4.mirror.nju.edu.cn	Y	N
v4.mirrors.nju.edu.cn	Y	N
ipv6.mirror.nju.edu.cn	N	Y
ipv6.mirrors.nju.edu.cn	N	Y
v6.mirror.nju.edu.cn	N	Y
v6.mirrors.nju.edu.cn	N	Y

Domain name	IPv4(A)	IPv6 (AAAA)
repo.nju.edu.cn	Y	Y
*.repo.nju.edu.cn	Y	Y
docker.nju.edu.cn	Y	Y
gcr.nju.edu.cn	Y	Y
ghcr.nju.edu.cn	Y	Y
glcr.nju.edu.cn	Y	Y
k8s.nju.edu.cn	Y	Y
ngc.nju.edu.cn	Y	Y
nvcr.nju.edu.cn	Y	Y
quay.nju.edu.cn	Y	Y





Open-Source Mirror Site



HTTP(s) Statistics

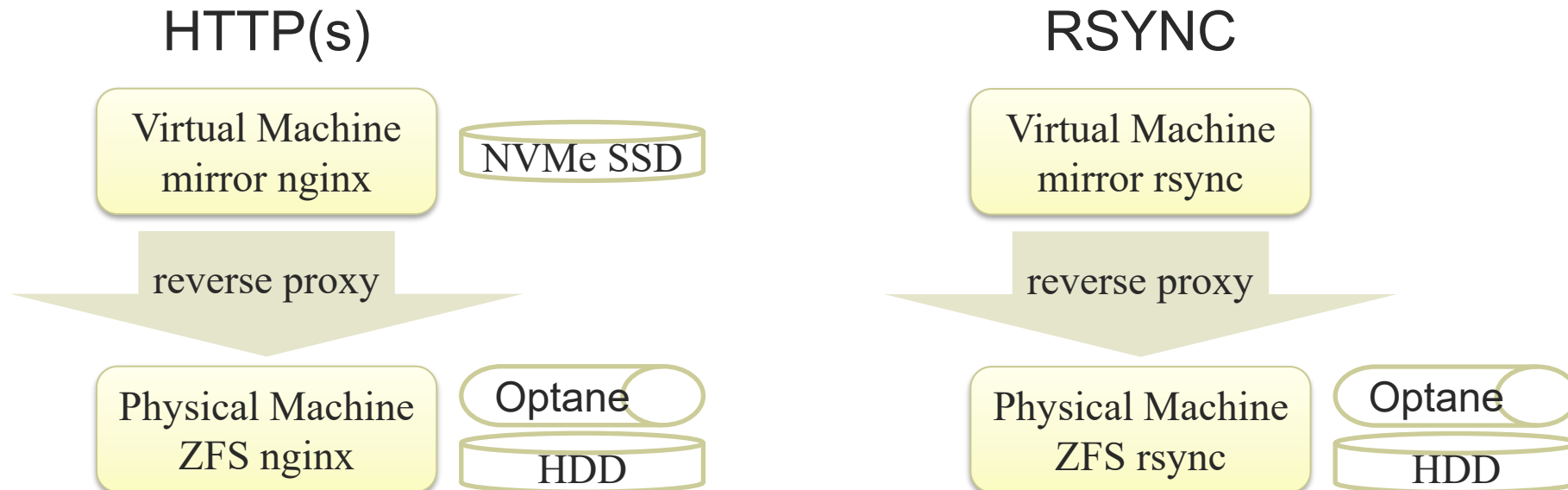
Mirror	Outbound traffic	Unique IP	Unique UA	Unique URL	Count
05/01~05/31	553.0 TB	11,449,026	3,711,323	37,824,966	2,634,019,480
06/01~06/30	427.7 TB	10,138,783	2,541,068	17,947,242	752,361,957
07/01~07/31	1,013.4 TB	5,855,293	2,505,396	22,359,394	402,760,580
08/01~08/31	1.7 PB	6,052,040	1,514,859	35,576,447	546,055,440
09/01~09/30	1.5 PB	4,786,754	3,374,614	34,209,089	740,073,079
10/01~10/27	419.8 TB	3,536,479	5,393,293	15,551,740	256,596,637
Repo	Outbound traffic	Unique IP	Unique UA	Unique URL	Count
05/01~05/31	45.5 TB	99,618	24,806	469,629	135,333,970
06/01~06/30	48.8 TB	110,520	25,507	626,701	128,619,842
07/01~07/31	49.8 TB	103,971	25,513	382,116	137,010,917
08/01~08/31	28.1 TB	109,489	23,618	369,145	155,208,283
09/01~09/30	42.0 TB	102,360	24,014	1,266,812	148,345,314
10/01~10/27	39.2 TB	92,504	20,045	1,205,472	134,427,633



Open-Source Mirror Site



- Frontend: Public IP, reverse proxy to backend, HTTP caching
- Backend: Storage, Synchronization
 - Reduce backend ZFS HDD load and shorten Rebuild and Scrub times



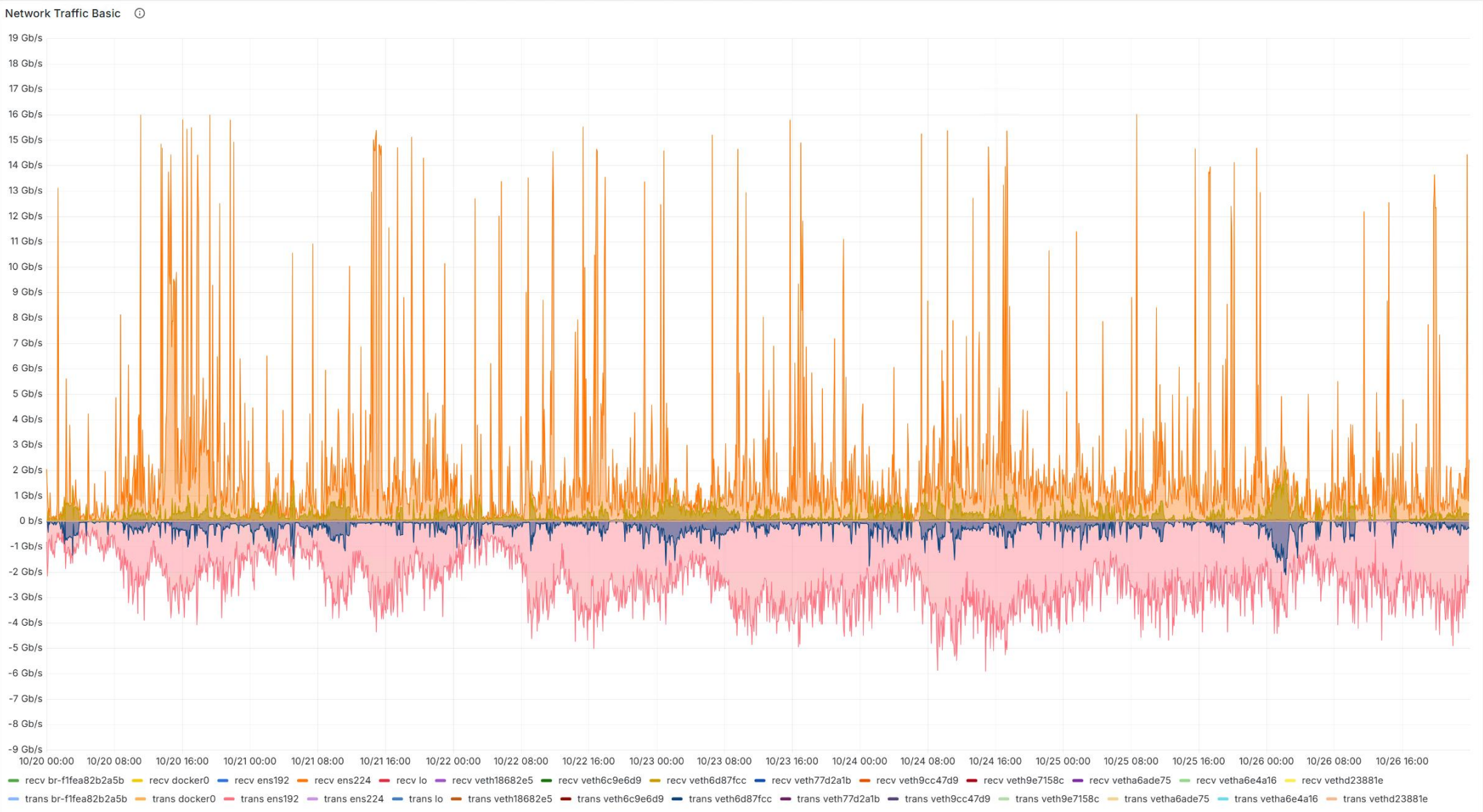


Open-Source Mirror Site



Nginx Proxy Cache

Mirror Network Traffic bits out (-) / in (+)





Open-Source Mirror Site



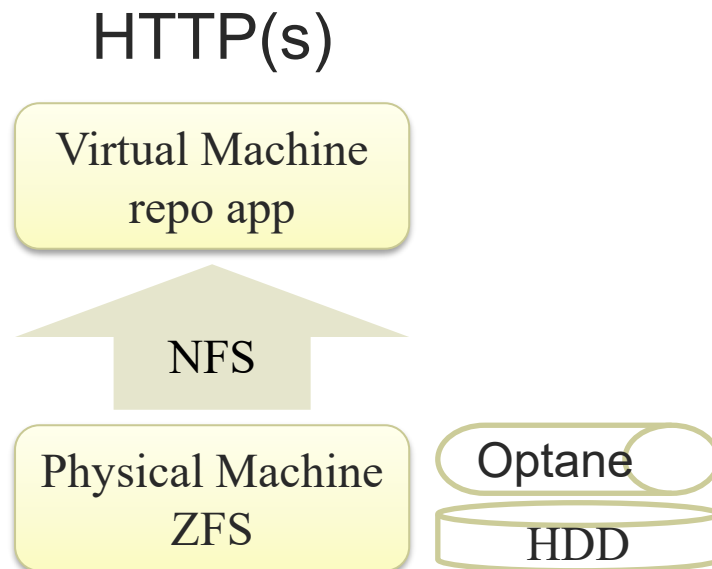
- Synchronization
 - tunasync
 - tunasync-scripts, tsumugu, shadowmire, mirror-clone
- Post-Synchronization Processing
 - Clear frontend node Nginx cache
 - Get size from ZFS
 - Repo synchronize dependencies



Open-Source Mirror Site



- Frontend: Public IP, Application, Database
- Backend: Storage



Repo	App
container	registry
go	goproxy
npm	artipie
maven	reposilite



Open-Source Mirror Site



☆ The largest mirror site in CERNET

- Repos: 457
- Size: ~397 TiB
- Requests: 2035.4 million
- Traffic: 5121.7 TiB
- Cache Hit: 83.7%

Repo	Requests M	Traffic TiB	Cache %
ubuntu-cdimage	247.3	2672.6	98.1%
julia-releases	199.3	317.4	98.8%
rocky	186.8	195.3	97.0%
eclipse	73.5	43.7	89.0%
pypi	57.0	95.1	39.9%

Repo	Requests M	Traffic TiB	Cache %
ubuntu-cdimage	247.3	2672.6	98.1%
julia-releases	199.3	317.4	98.8%
openjdk	3.3	313.6	98.4%
rocky	186.8	195.3	97.0%
deepin-cd	9.1	148.5	95.9%

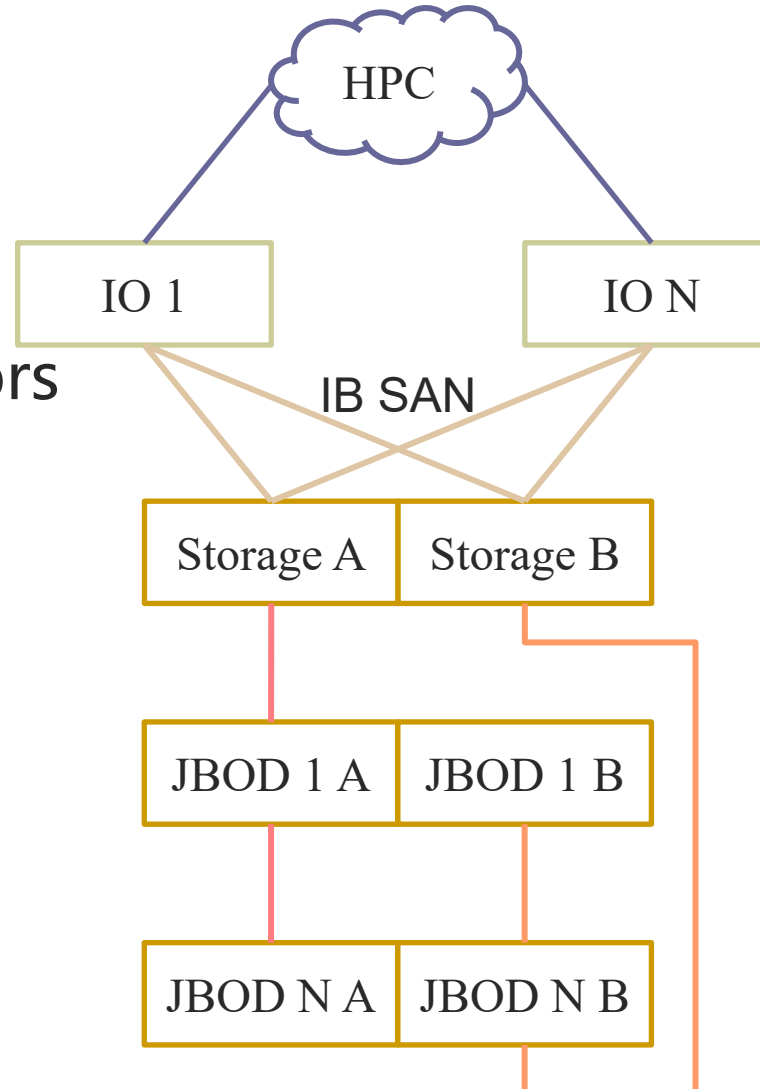
Repo	Requests M	Traffic TiB	Cache %
julia-releases	199.3	317.4	98.8%
kali-images	6.6	78.7	98.4%
openjdk	3.3	313.6	98.4%
armbian-releases	16.9	125.3	98.3%
ubuntu-cdimage	247.3	2672.6	98.1%



HPC Storage



- 2017-2020:
 - HPC Home
 - Open Source Mirrors
 - Backup
- ✓ Raid6
- ✓ Metadata: SAS SSD
Data: NL-SAS HDD



ZFS Storage

- 2020-now:
 - Open Source Mirrors
 - Backup
 - Archive
- ✓ Raidz3
- ✓ Compression
- ✓ ARC, ~~L2ARC~~
- ✓ ZIL

ZFS Storage

JBOD 1

JBOD N



Optane Persistent Memory

NVMe SSD

22T SATA

18T SATA

14T SAS

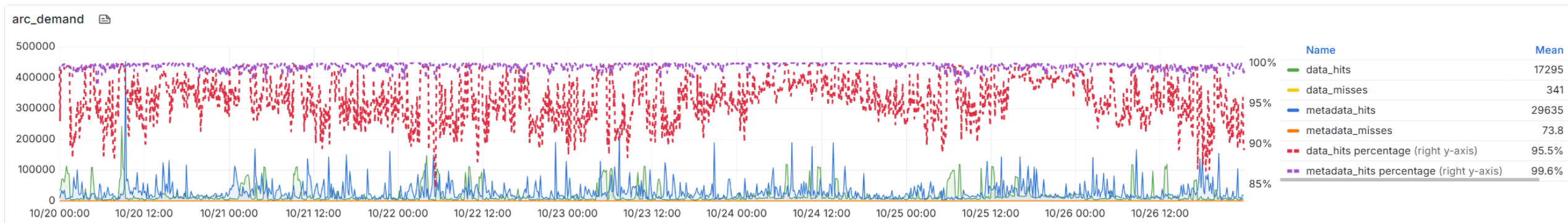


ZFS Storage



■ ZFS ARC (~1.3 TiB)

ARC Demand Metadata & Data



ARC Prefetch Metadata & Data



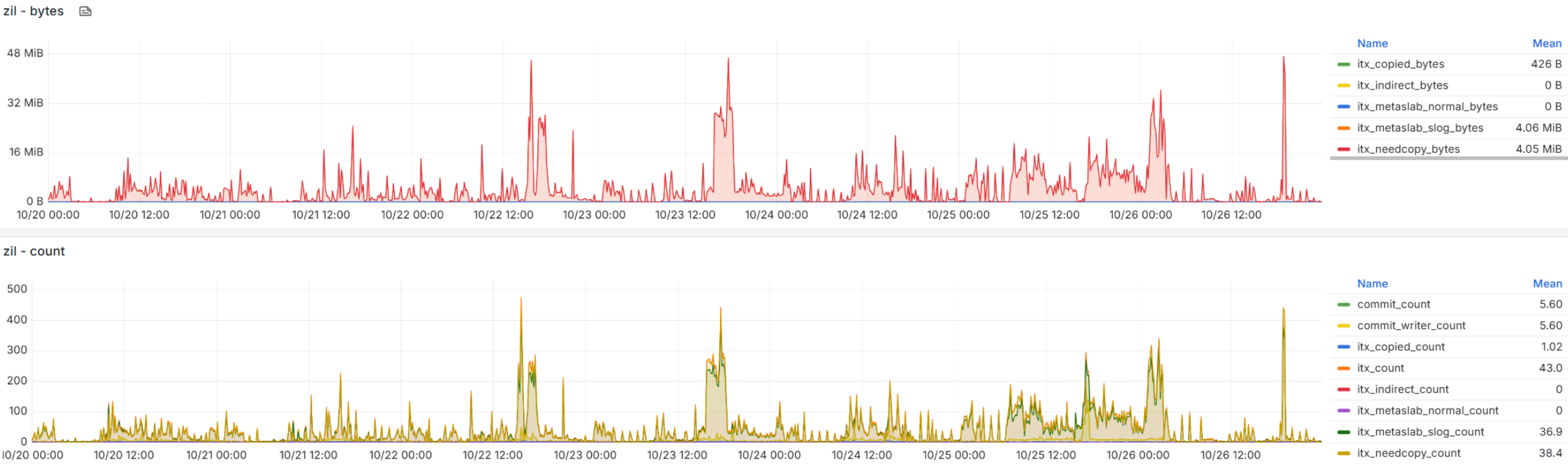


ZFS Storage



■ ZFS ZIL (18 GiB*2)

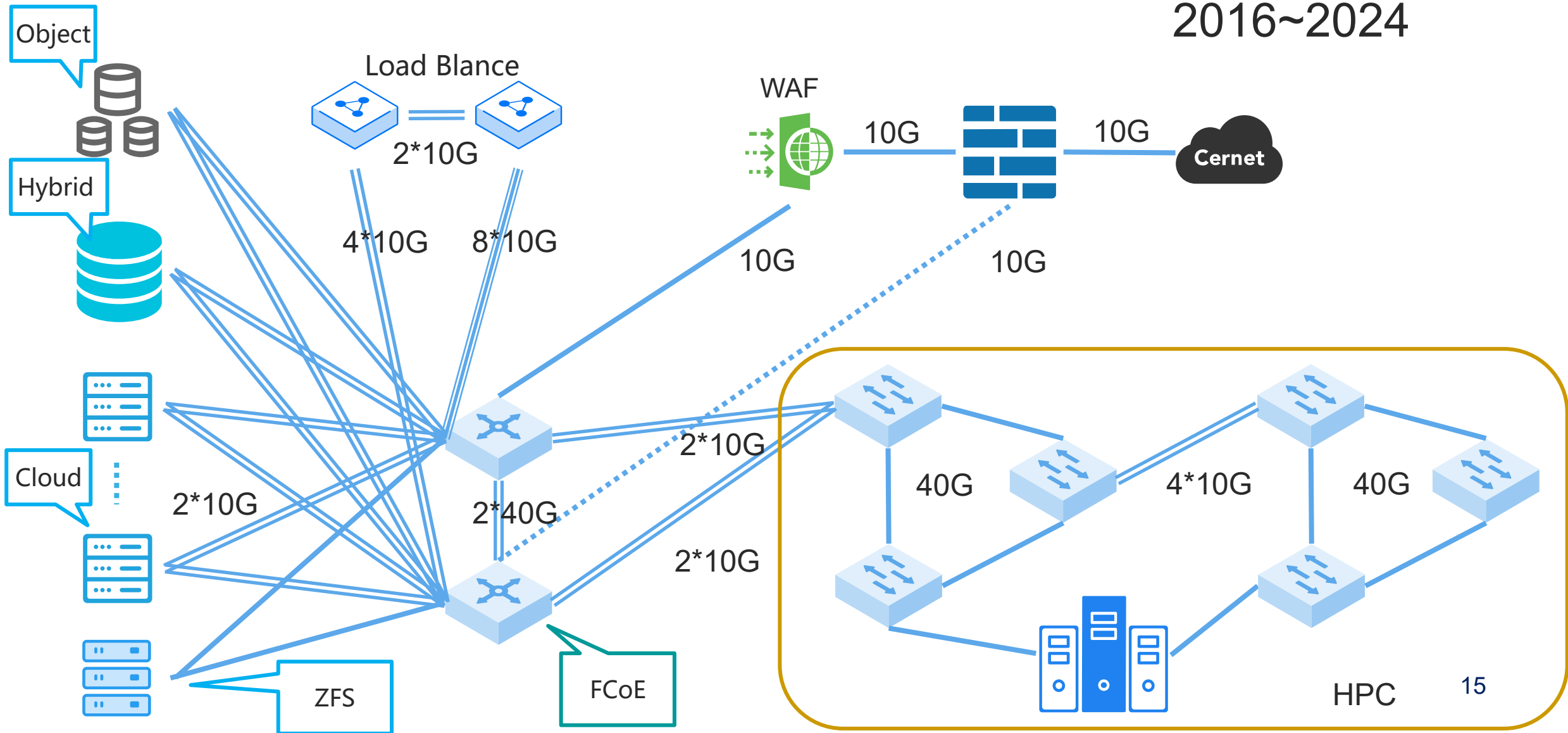
MemoryType	DDR	PMemModule	Total
Volatile	0.000 GiB	1476.000 GiB	1476.000 GiB
AppDirect	–	36.000 GiB	36.000 GiB
Cache	384.000 GiB	–	384.000 GiB





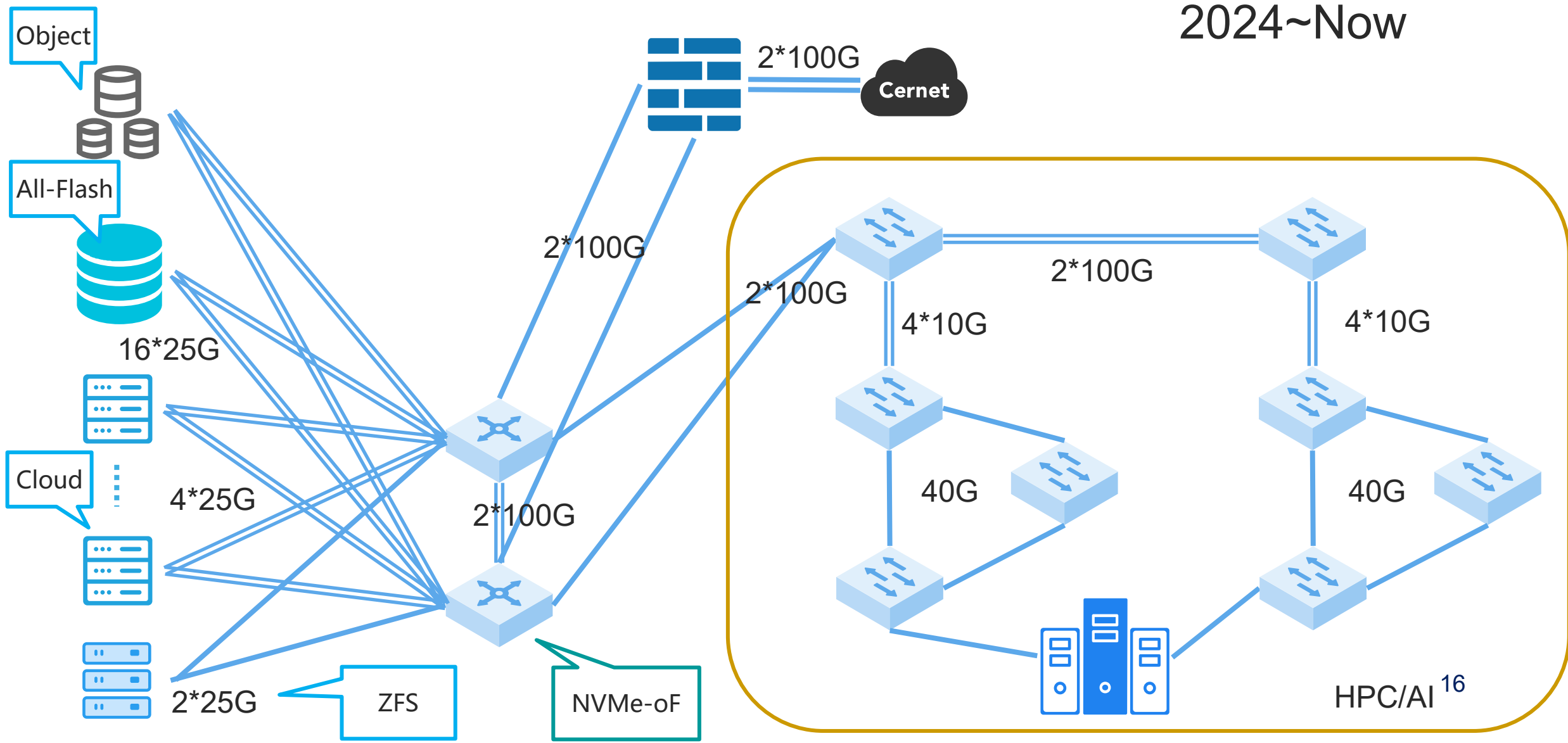
Network

2016~2024



Network

2024~Now





Open-Source Mirrors Site



■ Web Frontend

- <https://github.com/nju-lug/Mira>
- 2021/12: v1
- 2023/01: v2
- 2025/10: v2.3

The screenshot displays the NJU Mirror web interface. The main content area shows a table of mirrors with columns for mirror name, status, size, last sync, and next sync. The status column uses green checkmarks for '同步成功' (sync success) and blue circles for '正在同步' (syncing in progress). The left sidebar contains navigation links for news, mirrors, and various services like cloud storage and document management. The top navigation bar includes links for download, help, news, and about.

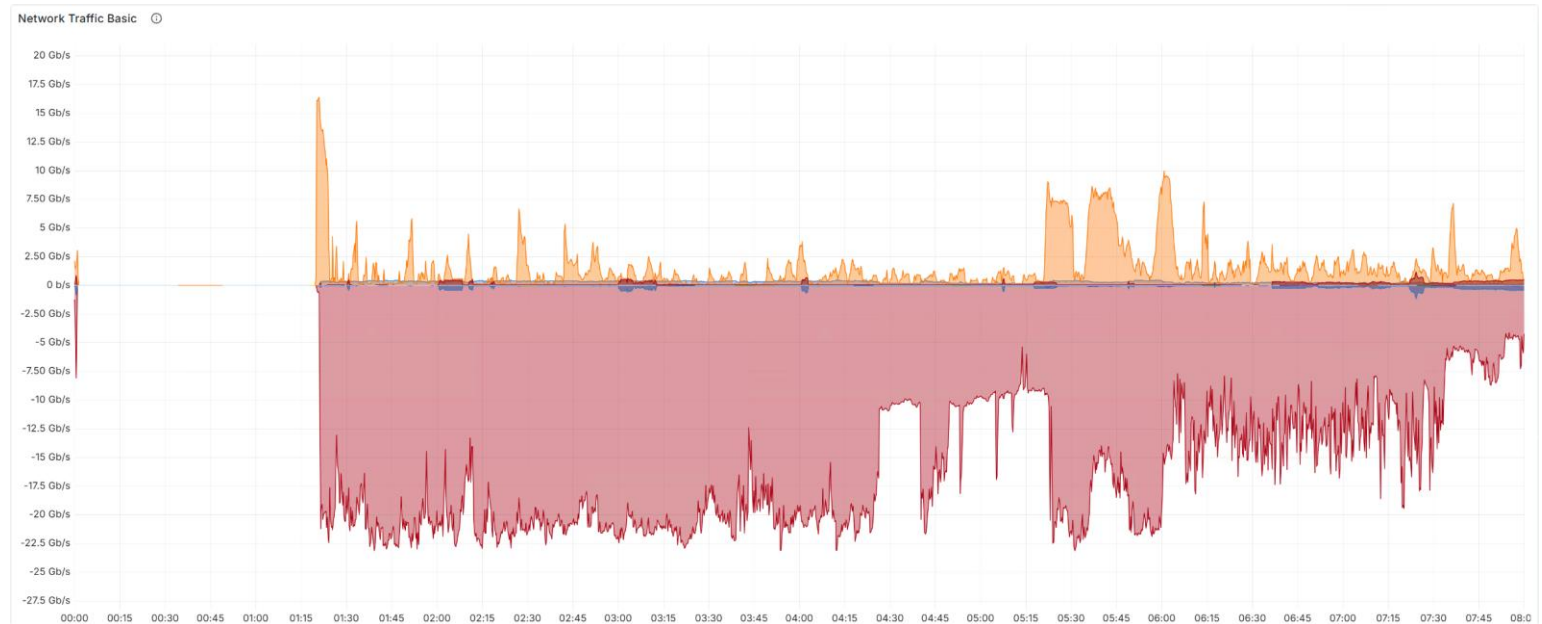
镜像名	状态	大小	上次同步	下次同步
7-zip	同步成功	477M	14 小时前	10 小时后
adobe-fonts	同步成功	63.4G	18 小时前	6 小时后
adoptium	同步成功	202G	2 小时前	1 天后
afrinic	同步成功	14.7G	3 小时前	21 小时后
alhp	同步成功	141.27G	13 小时前	11 小时后
almalinux	同步成功	1.44T	2 小时前	1 小时后
almalinux-elevate	同步成功	733.77M	1 小时前	2 小时后
almalinux-kitten	同步成功	819.37G	3 小时前	26 分钟后
almalinux-vault	同步成功	16.90T	40 分钟前	2 小时后
alphafold	同步成功	838G	-	-
alpine	同步成功	4.73T	26 分钟前	1 天后
anaconda	同步成功	21.3T	1 小时前	11 小时后
analis	同步成功	54.03T	19 小时前	5 小时后
anthon	正在同步	2.18T	33 分钟前	-
aosp	正在同步	1.5T	1 个月前	-
aosp-monthly	同步成功	289.34G	18 小时前	6 小时后
apache	同步成功	318.97G	3 小时前	9 小时后
apnic	同步成功	402G	2 小时前	1 天后
apptainer	同步成功	-	13 小时前	11 小时后
arch4edu	同步成功	36.89G	19 小时前	5 小时后
archlinux	同步成功	120.70G	2 小时前	25 分钟后
archlinuxarm	同步成功	148.22G	19 小时前	5 小时后



Open-Source Mirror Site



- NO Load Balancing, NO High Availability
 - ZFS
 - ZFS ARC
 - ZFS ZIL
 - Nginx proxy cache

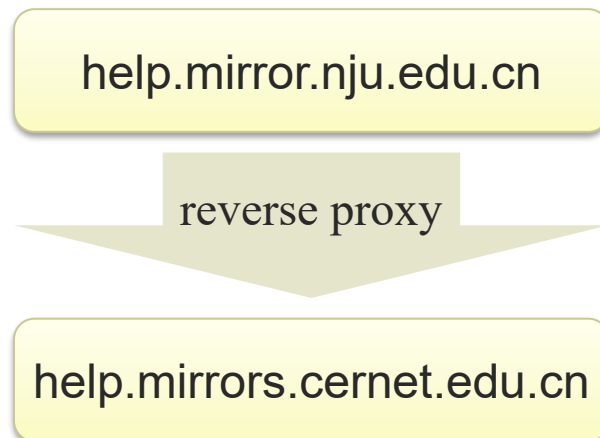




MirrorZ



- <https://mirrors.cernet.edu.cn/>
- <https://mirrorz.org/>



NJU



USTC



ZJU



SJTUG - Zhiyuan



SJTUG - Siyuan



NYIST



BJTU



BFSU



TUNA.NANO



TUNA.NEO



JCUT



SDU



CQUPT



CQU



JLU



LZUOSS



SUSTech CRA



XJTU



DNU



ISRC-ISCAS



NJTech



PKU



QLUT



SCAU



ShanghaiTech GeekPie



TUNA



UESTC LUG



WSYU



2023年：第一届开源软件论坛



- 主要面向学生
- 一天15小时



第零届北京大学高性能计算综合能力竞赛

- 世界首个创新形式的“普惠型”超算比赛
- 参赛人数超 800 人，参赛高校超 60 所
- 享誉全球，收获广泛关注
- 赛中使用超 3,000,000 核时超算资源

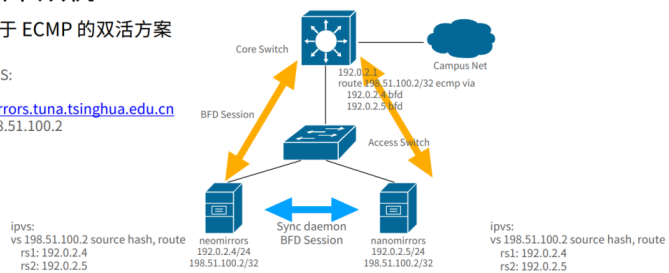


云下双机

基于 ECMP 的双活方案

DNS:

mirrors.tuna.tsinghua.edu.cn
198.51.100.2





2024年：第二届开源软件论坛





2025年：第三届开源软件论坛



■ 直至近24点

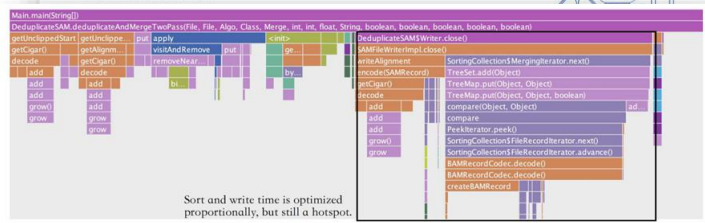
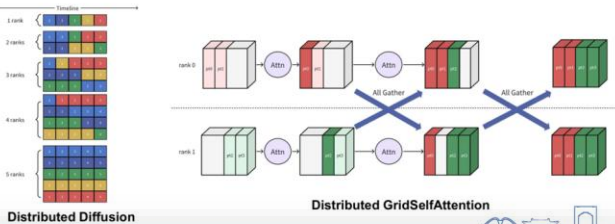
应对与防范策略

- 制定 AI 使用规范
 - 对生成内容进行协议与版权检查
 - 分级管理，明确哪些项目可以使用 AI 生成的代码
- 自动化检测
 - 许可证扫描工具自动化检测
 - 部署“协议防火墙”，拒绝高风险协议代码进入生产库
- 替代方案
 - 使用 AI 生成伪代码 / 架构思路，由人工重写实现
 - 内部部署私有模型，限定数据来源



5. 多机推理优化

- 采用 oneCCL 作为通讯后端
- Diffusion 部分数据并行，设计任务调度系统在保持负载均衡的同时最小化通讯量与通讯次数
- GridSelfAttention 部分切分 seq_len，使计算与通讯可以互相掩盖





开源软件论坛



■ 2026年继续

今年的开源论坛上看到了很多熟悉的面孔，看到了有的学生在一个方向持续精进的过程，也看到了有的学生天马行空又注重落地的项目实践，一年一度的开源软件论坛像极了一帮有共同爱好的人的年度聚会。

感谢南京大学姚主任提供的机会，明年再会。

收起



精彩回顾 | e-Science中心成功举办第三届开源软件论坛

公众号 · 微结构eScience

一个从四年前高校开源镜像站管理员的群聊中诞生的活动，“暑假能不能找个地方线下小聚”的设想最终演化成了五十余人（第一届）、百余人（第二三届）的论坛。非常非常感谢南大姚老师，是他在繁忙工作之外的义务组织，包括贡献了自己的不少科研经费，才让这件事成为可能。

希望国内受益于开源软件的各类商业组织机构，尤其是高校镜像站的重度使用者们，都能充分认识到这背后到底有多少人、多长时间的无偿付出。请尽可能地提供帮助。捐赠服务器硬件、赞助社区活动（无论是给 TUNA、USTCLUG、LCPU、南大、或者其他大小组织），都是对国内开源生态的重要支持。朋友圈的各位 UR、PR、CTO、负责人、主理人等，无论对哪个组织有何种意向，我都可以代为传达或者协助联系，有一份力出一份力。

长期有效。

收起



精彩回顾 | e-Science中心成功举办第三届开源软件论坛

公众号 · 微结构eScience



服务



网站	服务	推出时间 ↑	OTP	用户量
	高热密度机房	2008		
hpc.nju.edu.cn	超级计算	2011	强制	~ 1 千
box.nju.edu.cn	云盘	2017	可选	~ 5.2 万
mirror.nju.edu.cn	开源镜像	2017		~ 1 千万独立IP / 月
git.nju.edu.cn	代码托管	2018	可选	~ 1.1 万
entry.nju.edu.cn	远程控制	2018	强制	~ 1.1 千
test.nju.edu.cn	网络测速	2019		~ 7 万次 / 月
stor.nju.edu.cn	数据存储	2020/2025		
table.nju.edu.cn	协同表格	2020	可选	~ 8 万
tex.nju.edu.cn	LaTeX	2021	可选	~ 1.8 万
doc.nju.edu.cn	文档中心	2021	编者强制	~ 3 万独立IP / 月
speed.nju.edu.cn	内网测速	2024		~ 7 百次 / 月
pass.nju.edu.cn	密码管理	2025	可选	~ 1 百
time.nju.edu.cn	网络授时	2025	~ 6% 中国DNS应答(NTP pool)	



HPC



CPU	Intel Xeon 5150, E5430, X5550, X7542, E5620, E5645, X5650, E5-2643, E5-2660, E5-2692v2, E5-2680v3, E5-2630v4, E5-2640v4, E5-2650v4, E5-2660v4, E5-2680v4, E5-2682v4, E7-4820v4, E7-4850v4, Gold 6140, Gold 6148, Platinum 8163, Platinum 8168, Gold 5218, Gold 6248, Platinum 9242, Gold 5220R, Gold 6226R, Gold 6230R, Gold 5318H, Gold 6330, Gold 6338, Platinum 8358, Gold 6458Q. AMD EPYC 7302, 7402, 7702, 7B12, 7K62, 7543, 7552, 7K83, 9654, 9754, 9755.
GPU	Nvidia Tesla K40, P100, V100(PCIe), V100(SXM2), A100(SXM4). Nvidia GeForce RTX 2080 Ti, TITAN RTX, GeForce RTX 3080, GeForce RTX 3090, GeForce RTX 4080, GeForce RTX 4090, GeForce RTX 4090 D.
NET	1/10/25/40/100Gb Ethernet, 40Gb iWARP Ethernet. 20/40/56/100/200Gb DDR/QDR/FDR/HDR InfiniBand. 100Gb Omni-Path





Cloud Storage



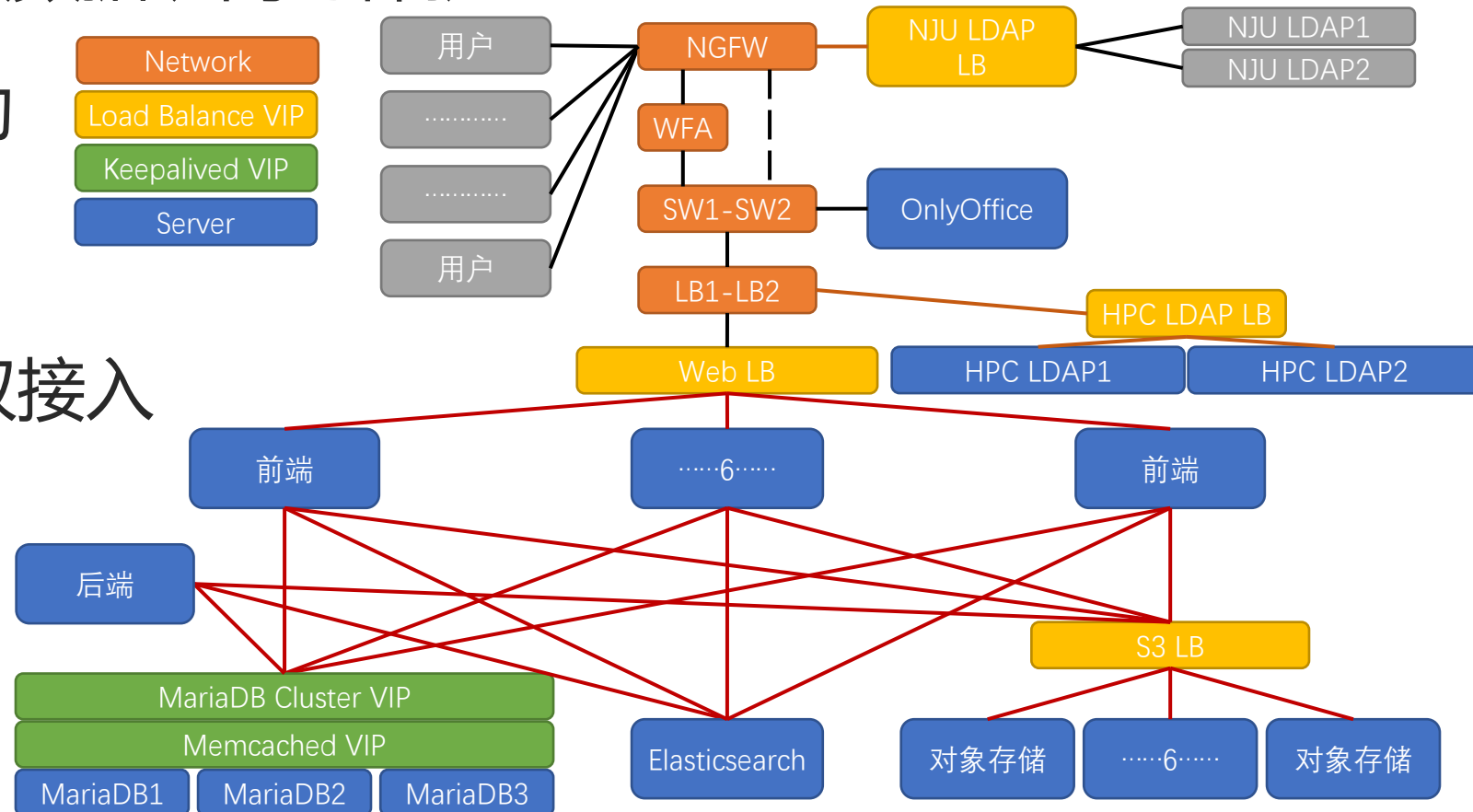
- File Sync&Share and Collaboration Platform
- Seafile
 - Community Edition, **Enterprise Edition**
- Desktop Syncing/Drive Clients, Mobile Clients
 - Windows, Mac, Linux GUI/**CLI**, Android, iOS
- Rclone



Cloud Storage



- 自主设计建设全分布式微服务架构
- 四校区协同共享，持续数据访问与保护
- 跨学科跨单位数据流动
- 丰富API二次开发
- 学校统一认证与HPC双接入
- HPC同步与备份





Cloud Database



- Like Excel, Like Database
- SeaTable
 - Developer Edition, **Enterprise Edition**

南京大学开源镜像新增镜像申请									
序号	镜像名称	讨论结果	简单介绍	镜像方式	相关链接	申请理由	原因	当前状态	
1	Request-0001	quantum espresso (示例)	完成镜像	一款免费开源的第一性原理	GitHub Release直接镜像	https://github.com/QEF/q-e/releases	免费开源，最广泛用于第一	完成	
2	Request-0002	anaconda/nvidia	不予镜像	深度学习研究基本都用pyto	正常安装方式能成功即可	https://pytorch.org/get-started/locally/	conda install pytorch torch	由于nvidia license限制，我们	已经完成
3	Request-0003	Linux StarCraftII 4.10	不予镜像	用于多智能体强化学习	只需要镜像一个固定的文件	https://github.com/Blizzard/s2client-proto#downloads	该环境在人工智能学院	该页似乎只提供了一个程序接	已经完成
4	Request-0004	南邮课表	完成镜像	这是一款由南京大学校友在	github release会有apk (安	https://github.com/idealclove/NJU-Class-Schedule-Flutter	该开源项目为南京大学校友		已经完成
5	Request-0007	cuda	不予镜像	NVIDIA推出的基于GPU能力			由于在校内有多台服务器存	CUDA并非开源软件，单独镜	已经完成
6	Request-0008	neovim=0.9.0	完成镜像	neovim是一个很好的vim替			neovim是一个很好的vim替	这个版本应该和发行版本有	已经完成



Office on Web



- 2017~2023: Office Online
- 2022~now: OnlyOffice Docs
 - Community Edition, Developer Edition, **Enterprise Edition**
- 为表格SeaTable和云盘Seafile提供在线Office文件查看编辑



Git



■ GitLab

- Community Edition -> Enterprise Edition -> **JiHu Edition**

统计	功能
派生1,108	注册✓ 已启用 ⚙
议题2,618	LDAP未启用 ?
合并请求10,270	Gravatar✓ 已启用 ⚙
备注144,448	OmniAuth✓ 已启用 ⚙
代码片段85	通过电子邮件回复✓ 已启用 ?
SSH Keys5,428	高级搜索✓ 已启用 ⚙
里程碑111	Geo未启用 ⚙
激活用户10,659	容器镜像仓库✓ 已启用 ⚙
	GitLab Pages未启用 ?
	实例 Runners✓ 已启用 ⚙



Speed Test



- <https://github.com/librespeed/speedtest>
- Internet
 - [http\(s\)://test.nju.edu.cn](http(s)://test.nju.edu.cn)
- In-campus
 - <http://speed.nju.edu.cn>



数据存储



■ 用途

- 传递
- 发布
- 归档备份

■ 存储

- HPC
- Cloud
- ZFS

■ ProFTPD



我来助你 | 南京大学 eScience 中心助力天文数据跨国传输

原创 高世杰 微结构eScience 2024年10月29日 13:03 江苏

绿岸射电天文望远镜 ▶ 南京大学
~~OneDrive~~ → ~~NJU Box~~ → FTP ≈ HPC

THE JOURNAL OF
PHYSICAL CHEMISTRY
LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCCL

Letter

Bidirectional Allosteric Mechanism in Catch-Bond Formation of CD44 Mediated Cell Adhesion

Xingyue Guan,[§] Yunqiang Bian,[§] Zilong Guo, Jian Zhang, Yi Cao, Wenfei Li,^{*} and Wei Wang^{*}

■ ASSOCIATED CONTENT

Data Availability Statement

The MD simulation files and data analysis scripts are available at https://mirror.nju.edu.cn/catch-bond_CD44.

Supporting Information
The Supporting Information is available at <https://pubs.acs.org>.

Computation
summarizing

Directory: /catch-bond_CD44/

Type to search...

File Name	File Size	Date
Parent directory/	-	-
All-Atom_Simulation/	-	2024-08-09 14:30:19
Cafemol_Constant_Velocity/	-	2024-08-10 20:50:16
Cafemol_Interaction/	-	2024-08-10 22:34:25
Cafemol_PDBs/	-	2024-08-09 14:43:08
Cafemol_Pulling_Protocols_MAIN_RESULTS/	-	2024-08-10 21:35:23
Cafemol_Pulling_Trajectories_and_Logs_MAIN_RESULTS/	-	2024-08-10 22:36:55
Cafemol_ddV_tuning/	-	2024-08-09 14:43:11
README.md	4189	2024-08-11 15:04:40

Read Me

About This Data Folder

Greetings. This is the raw data folder for our work: *Bidirectional allosteric mechanism of catch-bond effect in cell adhesion*.

Warning: Our complete raw data take up space of several TBs! So here we cannot share all of them. But in Cafemol, once you have our protocol files (given the simulation settings as well as a certain random seed number), you would be able to get EXACTLY THE SAME trajectories compared to our outputs.

So we share the protocol files with settings as well as random seed numbers. But we also show:

- All the trajectories for pulling under 0pN and 12pN (whose comparison maximally shows the catch-bond effects!)
- And also for constant velocity pulling (thus one can see the process for Cterminal to unfold clearly)
- As well as All-atom trajectories (each 5 trajectories for both conditions, not so massive as Cafemol simulations)

Data Folder Structure

The structure of the folder here is listed as follows:

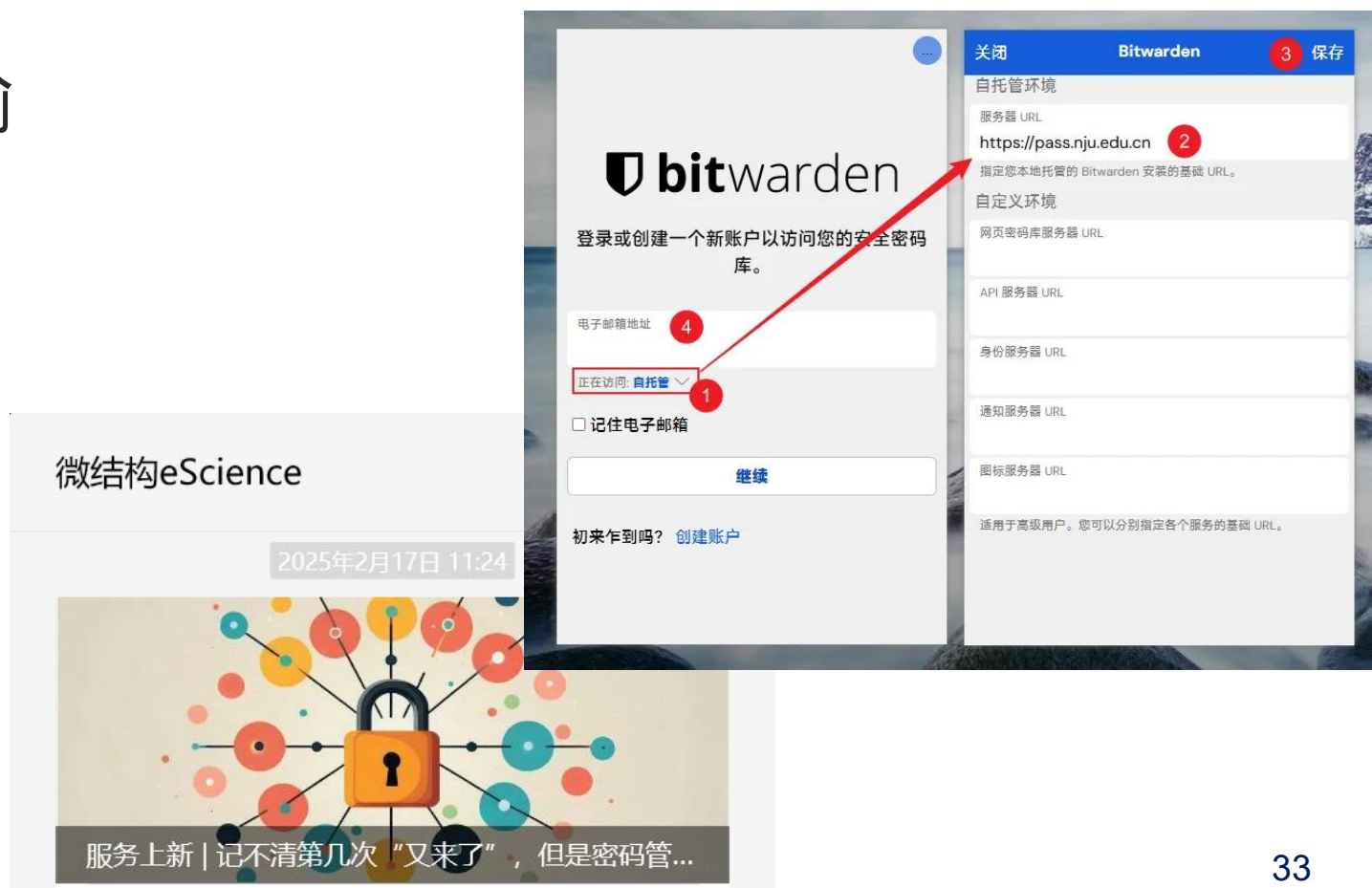
```
catch-bond_CD44/
|
|--- Cafemol_PDBs/          # PDB structures used in coarse-grained simulations.
|   |--- apo/
```



密码管理



- 基于兼容 Bitwarden 客户端API的 Vaultwarden
- 校内部署、存储和传输
- 高级功能
 - 无限组织、用户、集合
 - 集成TOTP
 - 加密附件
 - 紧急访问

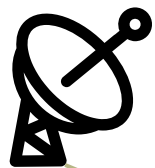




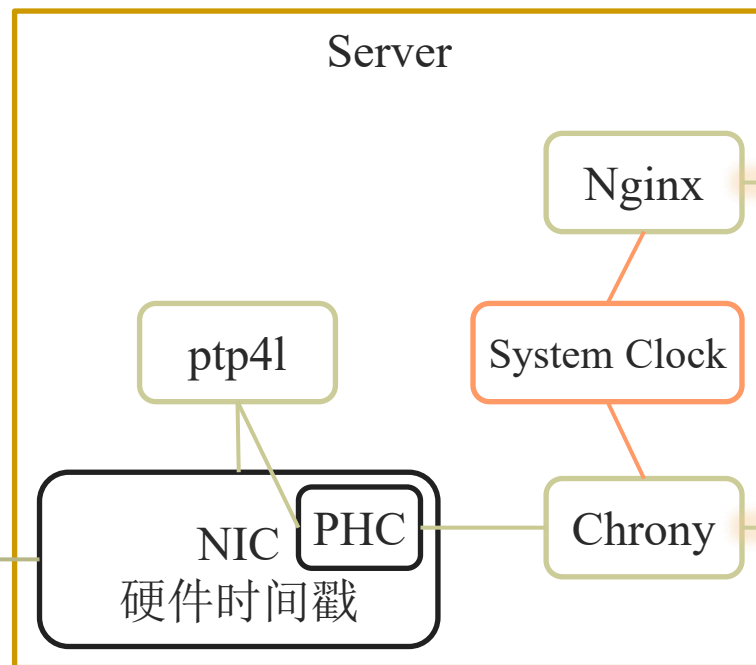
北斗授时



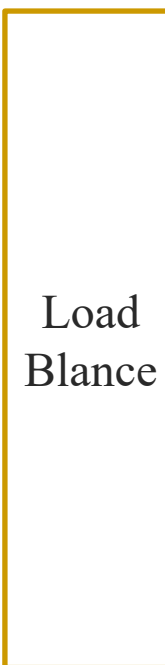
■ 高精度稳定的授时服务



IEEE 1588v2
L2 P2P



x2





Web Site



■ Documentation

- <https://doc.nju.edu.cn>
- <https://github.com/BookStackApp/BookStack>

■ Joomla

- <https://hpc.nju.edu.cn>
-



Logs



- Filebeat -> Elasticsearch -> Kibana
- Log Retention Duration 180 Days

~ 21 billion docs

~ 6.5 TiB

HOST 排名最前值	出站流量 之和	CITY 唯一计数	IP 唯一计数	URL 唯一计数	UA 唯一计数	记录计数
mirror	5.3PB	54,702	29,114,557	109,557,612	13,522,525	5,170,639,031
docker2	230.4TB	10,713	2,341,336	854,899	250,782	3,341,104,469
repo	218.4TB	1,841	457,932	2,453,547	78,201	719,585,987
boxfront5	83.8TB	1,705	132,575	19,731,489	67,830	986,583,823
boxfront1	65.3TB	1,591	116,891	16,059,163	57,946	726,223,735
boxfront6	61.8TB	1,523	109,394	15,217,836	57,314	739,127,552
boxfront3	59.8TB	1,478	105,357	14,034,755	53,905	606,331,027
boxfront4	37.3TB	1,133	62,122	7,570,954	33,857	682,805,772
git	18.0TB	31,705	7,661,742	43,249,652	23,708,652	4,220,838,402
onlyoffice	7.7TB	1,049	108,213	1,829,289	42,274	32,947,568
其他	11.1TB	3,909	1,352,556	6,723,677	312,608	1,221,806,366



Monitoring



- Exporter -> Prometheus -> Grafana
 - ALL: interval 15s, size ~2.3TiB, retention ~868d
 - HPC: interval 30s, size ~2.0TiB, retention ~373d
- Scripts -> InfluxDB -> Grafana
- ~~Cacti~~
- ~~Ganglia~~

Series	Label Pairs
391 961	22 452

Series	Label Pairs
1 279 566	12 434

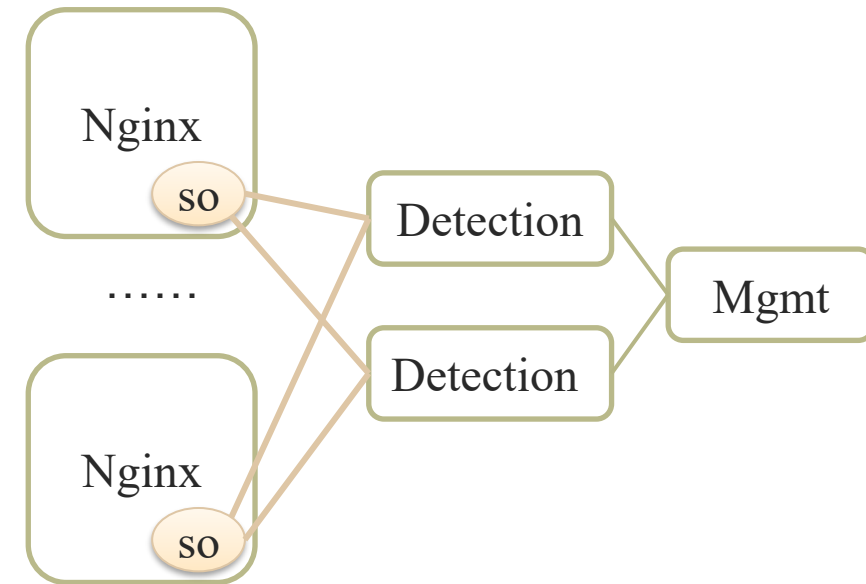


Web Application Firewall



■ SafeLine

- Community Edition, **Enterprise Edition**
- ACME
- TLS





南京大學

谢谢