

开源论坛

热烈欢迎各位嘉宾与同学的到来





人工微结构科学与技术 协同创新中心

中心由南京大学牵头，联合复旦大学、浙江大学、中国科学技术大学、上海交通大学、中国科学院合肥物质科学研究院和华为技术有限公司成立

2014年10月以“科学前沿”类排名第一获得教育部、财政部认定。

e-Science 中心

姚 舸

高级工程师

e-Science 中心 负责人

信息化建设管理服务中心 副主任

南京大学





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



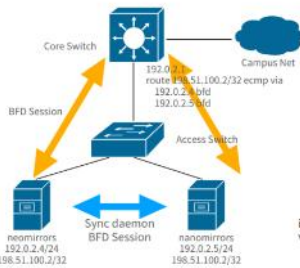
云下双机

基于 ECMP 的双活方案

DNS:

mirrors.tuna.tsinghua.edu.cn
198.51.100.2

ipvs:
vs 198.51.100.2 source hash, route
rs1: 192.0.2.4
rs2: 192.0.2.5



ipvs:
vs 198.51.100.2 source hash, route
rs1: 192.0.2.4
rs2: 192.0.2.5





2024年：第二届开源软件论坛



重建前后两台服务器的磁盘负载比较
左边为重建前，中间为仅 Rsync 服务器重建后，右边为两台服务器均重建后的负载



2025年：第三届开源软件论坛



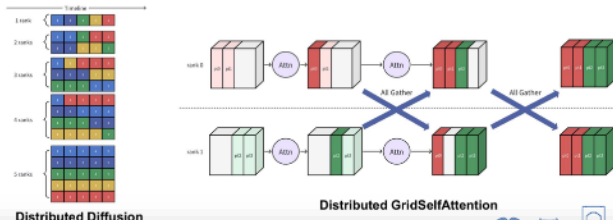
■ 直至近24点

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

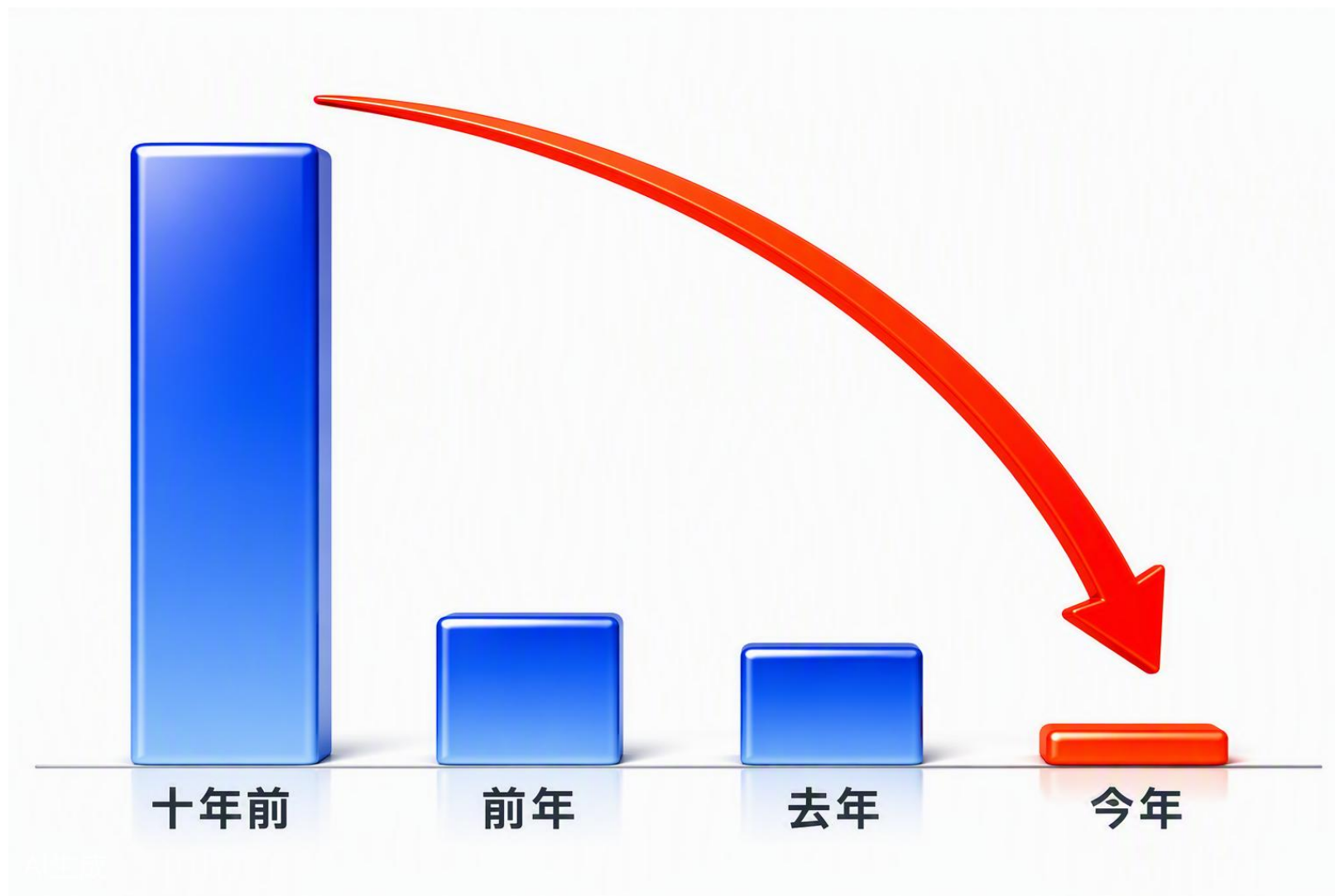
- 采用 oneCCL 作为通讯后端
- Diffusion 部分数据并行，设计任务调度系统在保持负载均衡的同时最小化通讯量与通讯次数
- GridSelfAttention 部分切分 seq_len，使计算与通讯可以互相掩盖



Sort and write time is optimized proportionally, but still a hotspot.









赞助商



Lenovo

联想

TeX *Page*

OG 奥工科技



会务



■ 报告

- 开源相关
- 范围不限制
- 大家兴趣

■ 会务

- 住宿
- 茶歇
- 午饭、晚宴