

万卡AI集群 建设方案



ZOMI

Question?

- 万卡集群，感觉很难，但不就是把 GPU/NPU 都放在一起堆料吗？



Content github.com/Infrasys-AI/AIInfra

AI 系统 + 大模型全栈架构图

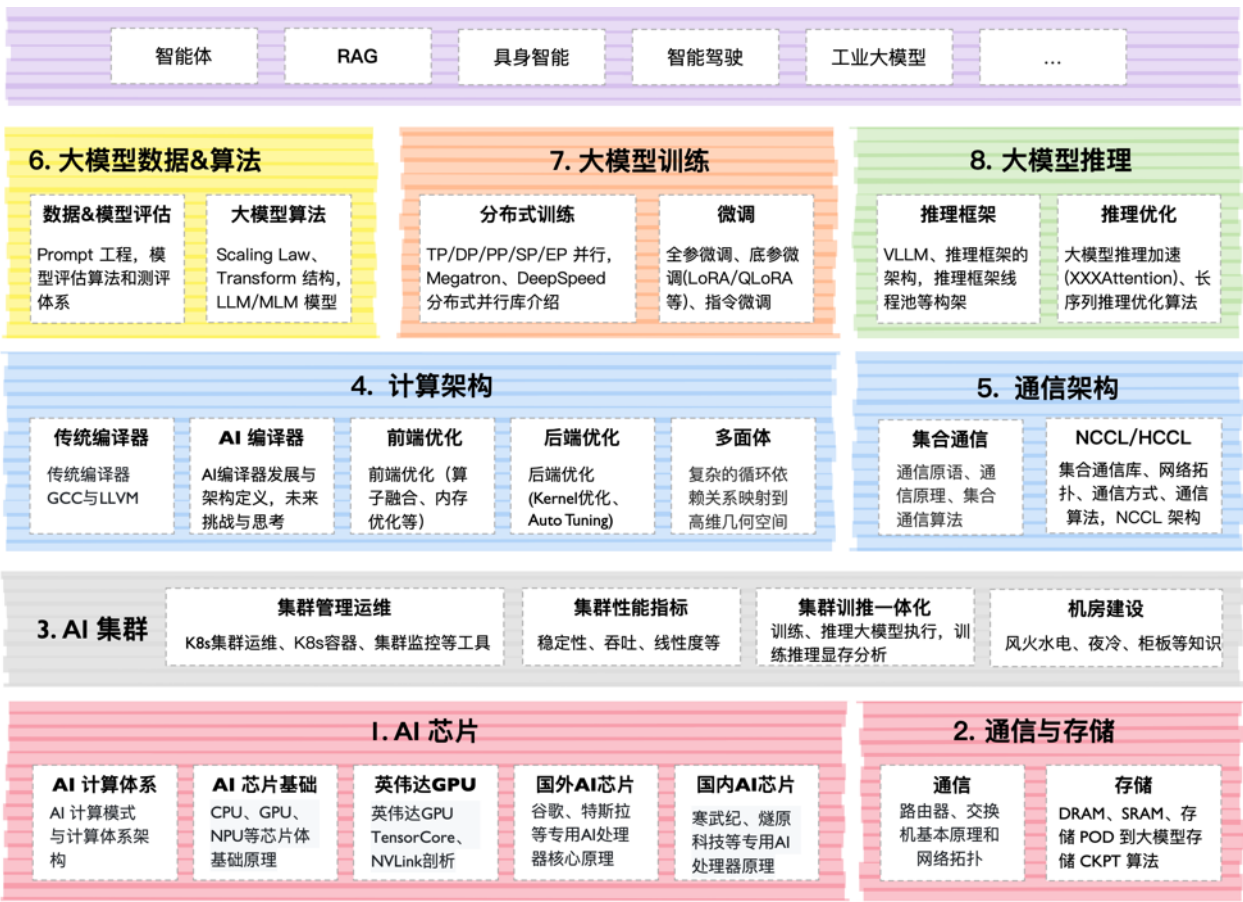


时事
热点

大模型
训推

编译
计算
架构

硬件
体系
结构



Content

集群建设之巅 万卡 AI 集群

万卡 AI 集群建设挑战

存算网、交付周期快、工期紧张等

万卡 AI 集群建设方案

从L0机房布线到L3 上层软件

测试方案与客户万卡场景

万卡性能测试方案？客户真实场景

NVIDIA BlackWell & BG200

NV 计算芯片产品演进与深度分析

XAI 十万卡集群

马斯克 XAI 十万卡集群分析

十万卡集群思考

对构建十万卡集群的思考



Content

1. 整体建设挑战
2. 能耗极限、网络瓶颈、计算效率、系统可靠性



01

整体建设

RoadMap



整体建设 Roadmap

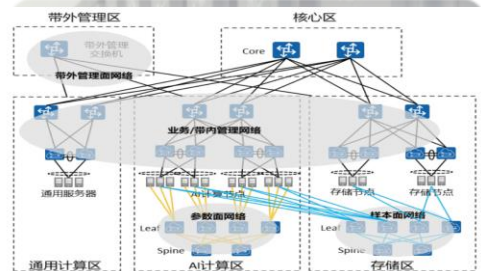
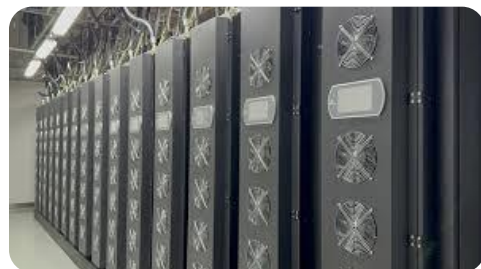
规划设计

- 明确业务需求，设计网络、计算、存储架构，并制定L0/L1 基建部署方案



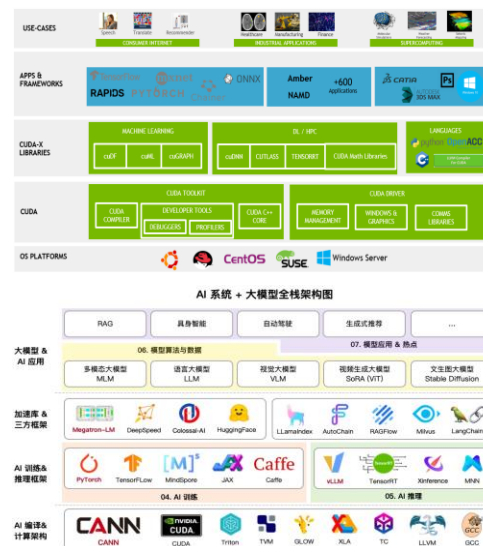
硬件安装

- 按规划安装 IT 服务器和网络设备，进行物理部署与线缆连接联调



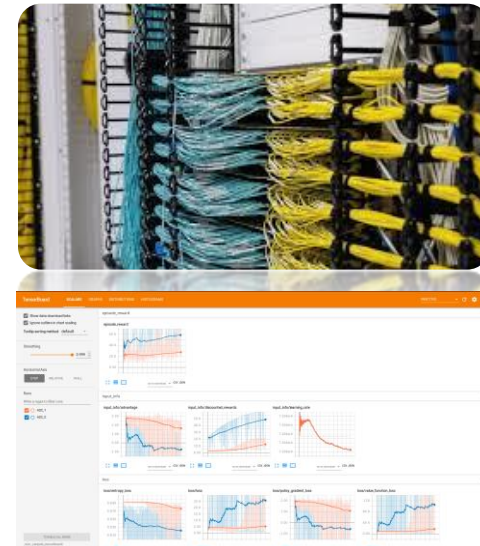
软件安装

- 部署通算操作系统和 AI 基础软件，包括集群管理软件，配置调度系统及作业平台

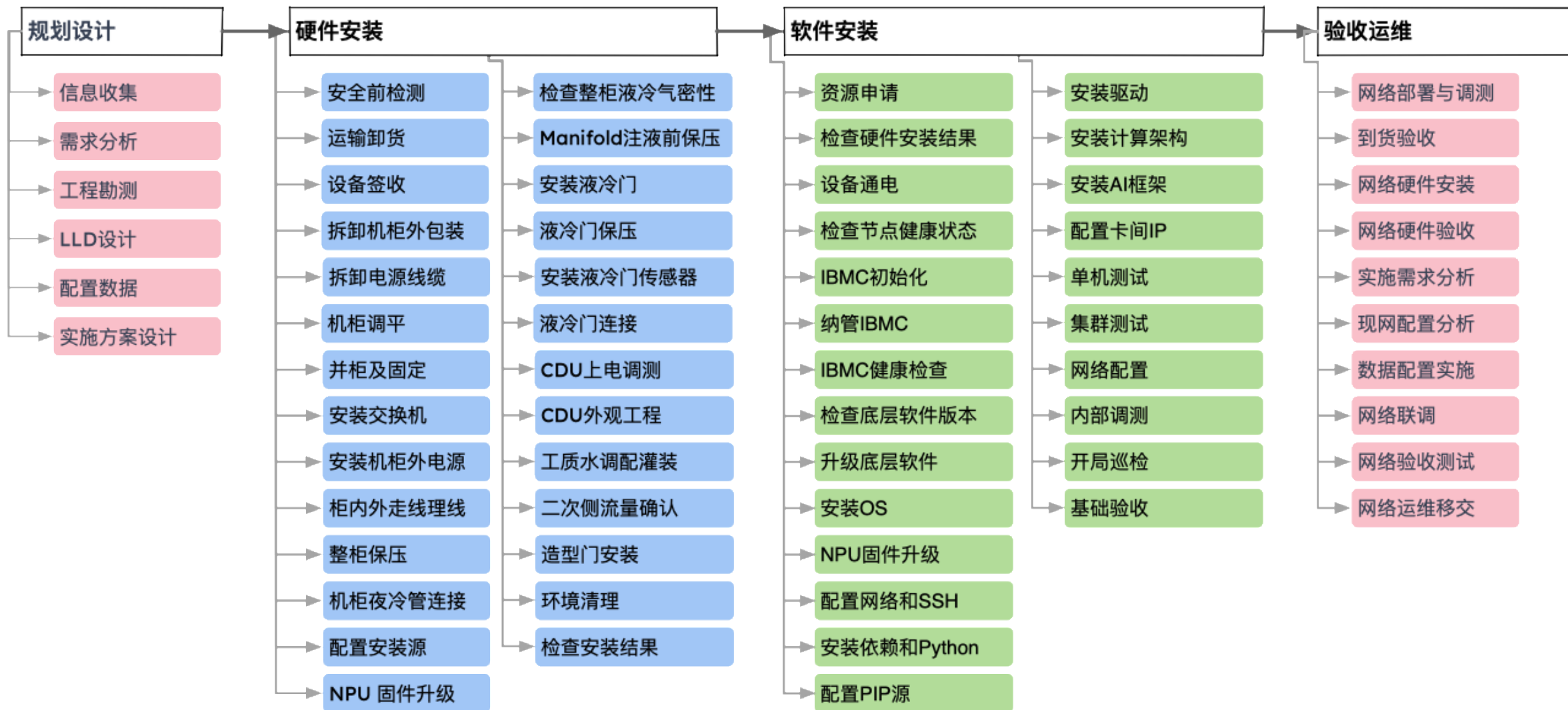


验收运维

- 验证性能功能稳定性，正式交付后转移运维主题，持续监控维护更新优化



整体建设 Roadmap

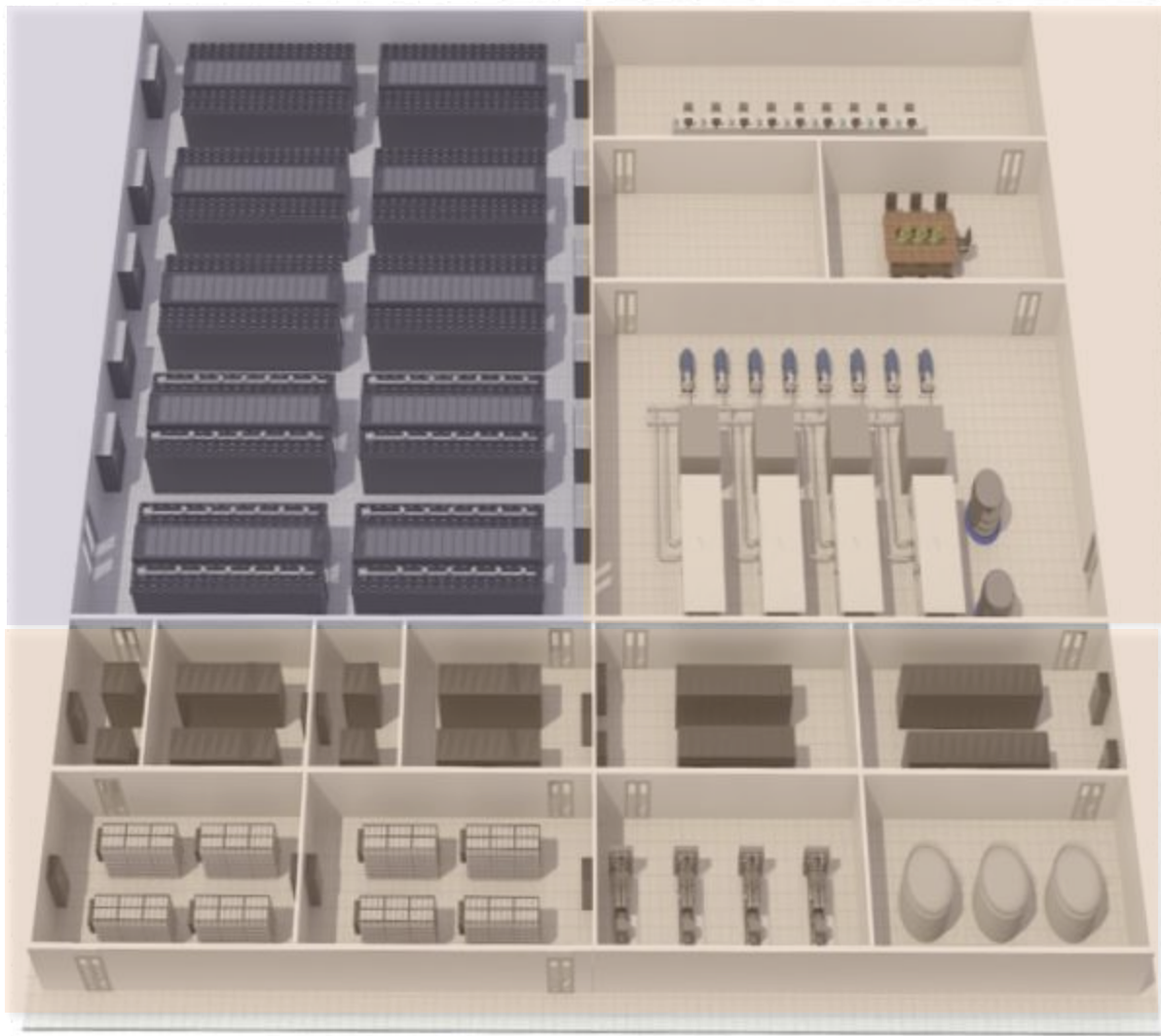


02

基本建设 前沿探索



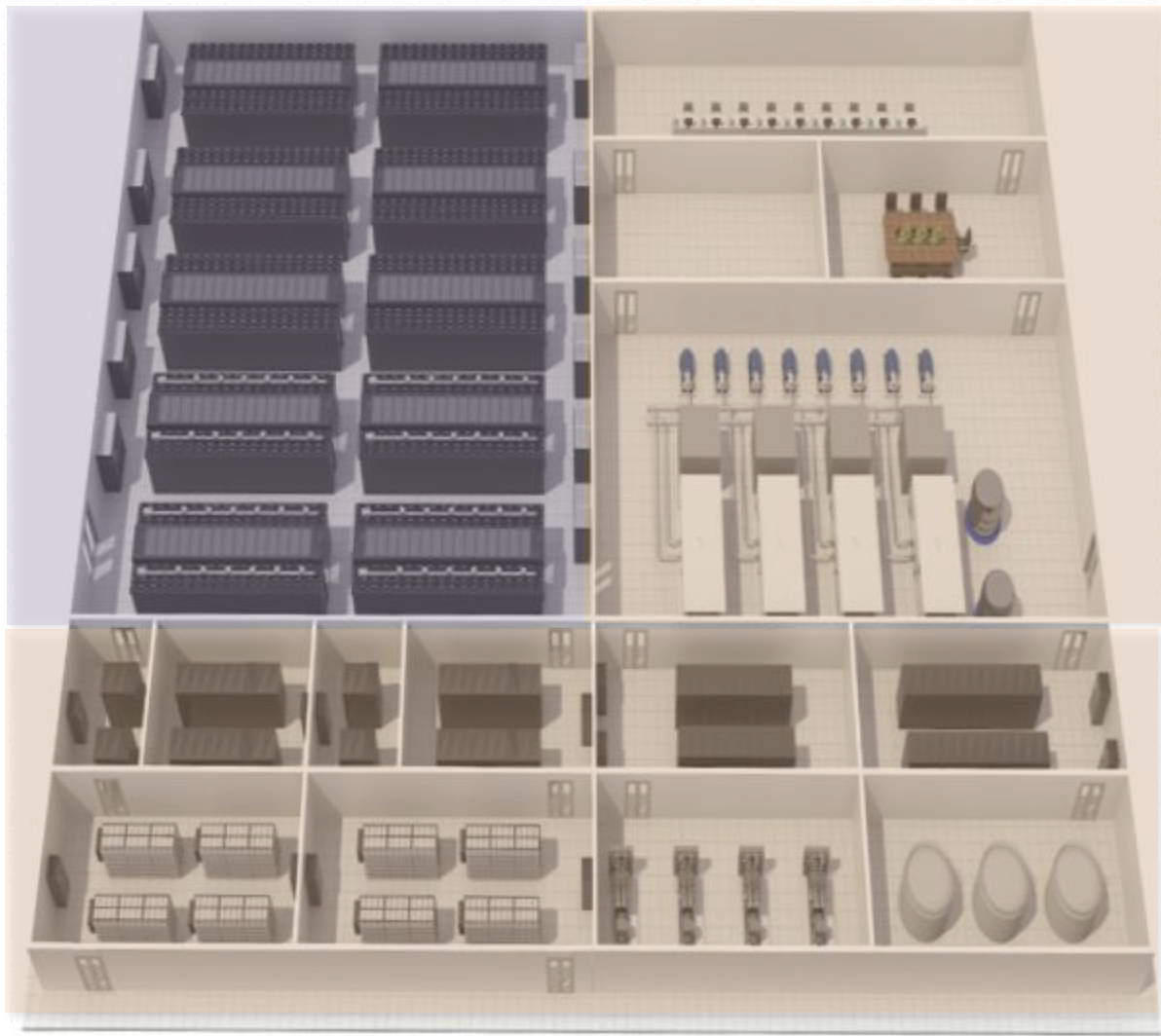
万卡集群机房布局



- 机房主要用于安装主通信设备、通信传输设备、电源等配套设备。
- 为维护和管理上的方便，要求上述设备安排紧凑，并按业务类别不同分别安装在不同的房间。



万卡集群机房布局



- **整体布局**：优先采用单层集中式布局，保障通信效率与运维便捷，其次才是多层和多栋。
- **功能分区**：核心设备居中部署，辅助区域环绕四周，满足高密度配电与制冷需求。
- **机柜扩容**：机房预留扩展空间，支持分阶段部署，提升灵活扩展能力。
- **供电设计**：供电系统分为左右两区，就近供电，避免跨层布线。
- **制冷管理**：采用冷热通道隔离和增强散热措施，确保高密GPU集群稳定运行。



万卡集群机房布线

- **跨层布线分区管理**：按楼层划分光迁区域，避免跨区布线混乱，提升维护效率。
- **就近穿层布线**：光缆穿层路径应最短化，避免绕线，降低延迟和损耗。
- **双层光迁桥架设计**：利用柜间空间设置上下两层桥架，满足高密度布线与冗余部署需求。



英伟达 DGX GB200 超节点



华为 CloudMatrix 超节点



万卡集群机房布线

- **跨层布线分区管理**：按楼层划分光迁区域，避免跨区布线混乱，提升维护效率。
- **就近穿层布线**：光缆穿层路径应最短化，避免绕线，降低延迟和损耗。
- **双层光迁桥架设计**：利用柜间空间设置上下两层桥架，满足高密度布线与冗余部署需求。



万卡机房配电制冷架构

- **挑战：**十万卡AI 集群功耗近 XXXMW。供电投资大（~30%to40% 基建），占地多（~20%to30% 辅助区）
- **选址建议：**电力充足、电价低的地方；气候凉爽，利于自然降温；有扩展空间，方便以后扩容；靠近数据中心聚集区，便于资源共享。
- **单路供电方案：**用一路高质量市电直供，配合UPS或储能系统；省去冗余供电线路，降低初期投资和运维复杂度；适合对可靠性要求不是极端苛刻的AI训练场景。



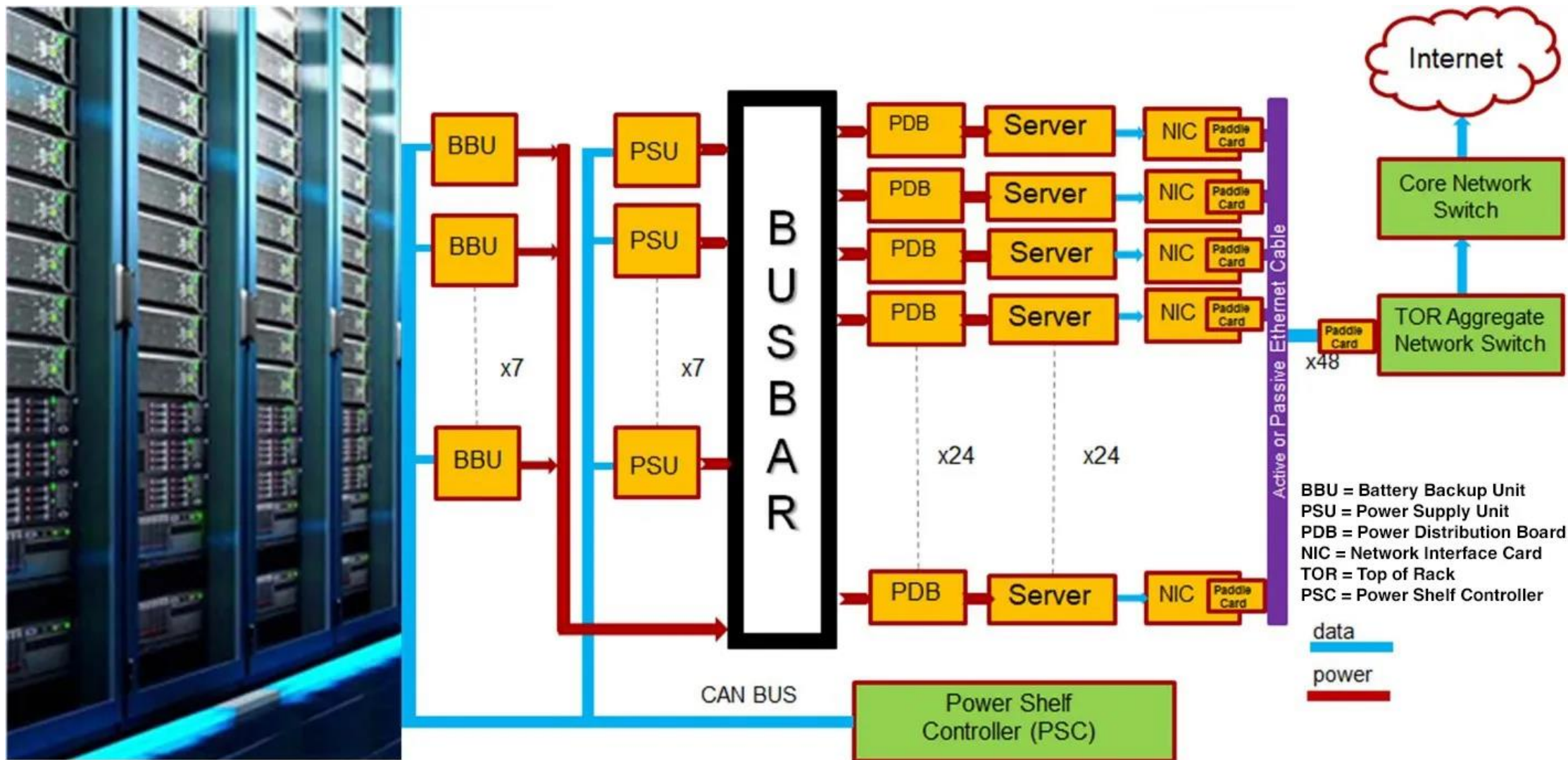
移动哈尔滨计算中心



Google 芬兰数据中心

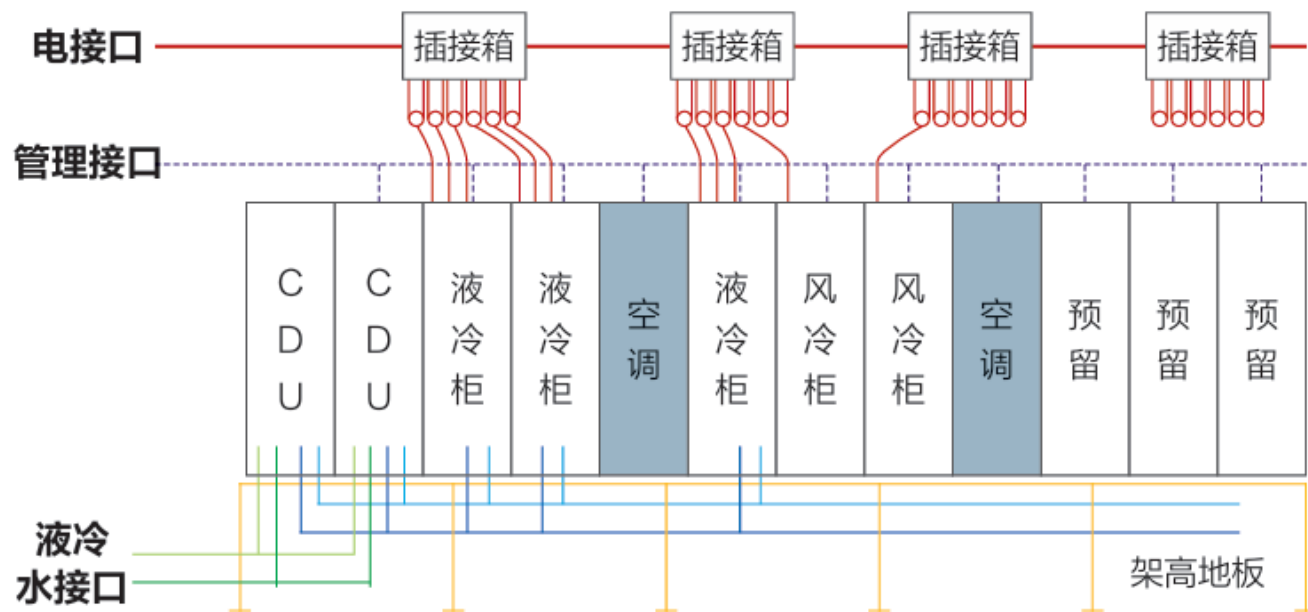


万卡机房配电制冷架构：Open Compute Project (OCP)



万卡机房配电制冷架构

- **挑战：**传统空调支持功率密度低（<20KW/柜），占用机房面积大（~20%~30%机房）
- **建议方案：主动式液冷**
 - **高散热效率：**液冷直接接触发热部件，适配万卡集群高功耗需求。
 - **节省空间：**取消传统空调风道和冗余空间，释放机房面积用于部署更多算力设备。



03

光链路脏污



光链路脏污问题的核心难点

1. 脏污难以避免

原因	技术挑战
高敏接口要求	200GE/400GE技术对噪声容忍度↓，断面清洁要求达微米级精度
环境限制	机房无法100%无尘，施工中的粉尘、油污无法彻底避免
暴露风险	施工时光模块/光纤必然有暴露时间（难以=0）→ 污染概率↑

2. 脏污难以识别

检测手段缺陷	具体表现
无在线监测	需人工检测，无法实时感知污染状态
指标模糊性	收光功率、信噪比与脏污程度无线性关系
参数失效	发光功率/电压/温度/偏置电流等参数无法反映断面污染物分布
区域差异	断面中心区比边缘对脏污更敏感（需显微镜级仪器定位）



光链路脏污问题的核心难点

3. 清理效果难保证

清理障碍	运维痛点
工具局限性	清洁笔对顽固油污无效 → 必须使用酒精棉签（操作风险↑）
耗时过长	单链路故障需清洁4端面（拉电源/调焦/分工） → 耗时≥15分钟
交叉污染	脏污在光模块与光纤间相互传染（清洁后二次污染率>40%）



光链路脏污问题：解决方案

1. 物理隔离层：施工防护

- **光模块真空封装：**出厂时充氮密封，拆封即插，避免存储/运输污染
- **光纤端面防尘帽：**弹簧自锁式设计（IP6X 防尘），插拔暴露时间 < 0.5秒
- **无尘操作舱：**百级洁净环境（ $\geq 0.3\mu\text{m}$ 颗粒 $\leq 100/\text{m}^3$ ），施工污染风险 $\downarrow 99\%$

2. 主动防御层：环境控制

- **机房正压通风：**维持室内气压 > 外部 5Pa，气流屏障阻隔外部粉尘
- **H14 级空气过滤：**ULPA 超高效过滤（99.999% @0.1 μm ），机房 PM1.0 浓度 < 10 $\mu\text{g}/\text{m}^3$



总结与思考



常见交叉领域问题

交叉领域	交叉点	影响内容
计算 vs 存储	Host 远程挂载存储 PoD	<ul style="list-style-type: none">• 计算节点 IO 问题导致存储性能不达标;• 存储 IO 影响 OS, 包括流控参数;• 存储抢占 OS 系统资源;
计算 vs 网络	网卡、路由、流控参数	<ul style="list-style-type: none">• 网络不通导致集合通信报错;• 计算下发影响网络性能, 出现反压和丢包;• 网络影响计算通信、训练性能不达标;
计算 vs 存储 vs 网络	交换机、流控参数、 模型并行、Host 下发与挂载	<ul style="list-style-type: none">• 存储影响计算、网络拖尾影响计算等互相干扰;• 出现大模型训练性能抖动、性能下降;• Leaf 交换机丢包, 训练精度受影响;





Thank you

把 AllInfra 带入每个开发者、每个家庭、
每个组织，构建万物互联的智能世界

Bring AI Infra to every person, home and
organization for a fully connected,
intelligent world.

Copyright © 2025 [Infrasys-AI](#) org. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. [Infrasys-AI](#) org. may change the information at any time without notice.



ZOMI

GitHub github.com/Infrasys-AI/AllInfra

Book infrasys-ai.github.io



引用与参考

1. <https://support.huawei.com/enterprise/zh/doc/EDOC1100314275/283099c4>
2. <https://www.edn.com/five-major-trends-in-power-supply-design-for-servers/>

PPT 开源在: <https://github.com/Infrasys-AI/AllInfra>

