xAI Colossus
世界最大十万卡集群

ZOMI

目录

01 GPU H100 液冷机架

超微液冷 GPU 机架

- **GPU：**
  - Colossus 目前部署 10 万 NVIDIA Hopper GPU，包括 5 万个 H100 和 5 万个 H200；
  - Colossus 基本构建模块是超微液冷机架，每个机架配备八台 4U 服务器；
- **机架配置：**
  - 每台服务器搭载八块 NVIDIA H100 GPU，这样每个机架总共有 64 块 GPU；
  - 每个机架可容纳 64 个 GPU，8个机架组成一个阵列，共计 512 个 GPU；
  - Colossus 共有超过 1500 个机架，接近200个阵列；
  - 每个服务器配备了四个电源，这些电源支持热插拔，并通过三相 PDU 进行供电；

超微液冷 GPU 服务器

- **CPU Tray：**
  - 配备两个 x86 CPU 液冷块 + 一个用于冷却四个 Broadcom PCIe 交换机的定制液冷块；
- **可维护性：**
  - 超微系统采用可维护托盘设计，每个服务器配备了四个热插拔电源；
  - 1U 机架歧管设计有助于引入冷却液和排出热液；
  - 部件被移除后，托盘便可以轻松拉出进行维护；

超微液冷 GPU 服务器

- **网络：**
  - 在机架背面，我们看到用于 GPU & CPU 连接的 400GbE 光纤，以及用于管理网络的铜缆；
  - 每个节点配备 9 个 400GbE（Gigabit Ethernet） 网络连接，总带宽达到 3.6 Tbps；
  - 8 个 NVIDIA BlueField-3 SuperNIC 用于 AI 网络；
  - 1 个 Mellanox ConnectX-7 网卡提供 CPU 端其他网络功能；

02

NVMe 存储系统

存储系统

- **规模：**
  - EB 级存储
- **介质：**
  - NVMe SSD
- **服务器：**
  - 超微 1U 服务器
- **特点：**
  - 满足 AI 大模型训练对存储容量需求；
  - 通过网络交付，供所有 GPU 和 CPU 服务器访问；

03

CPU
计算集群

CPU 计算集群

- **服务器：**
  - 超微 1U 服务器，每机架 42 台
- **CPU：**
  - 采用高速 x86 CPU，具体型号未知；
- **网络：**
  - 每个服务器配备一个 400GbE 网卡；
- **散热：**
  - CPU 服务器采用风冷设计，通过机架后部的热交换器将热量传递到液冷回路中；

04 以太网互联

参数面网络互联

- **技术：**
  - 400GbE 以太网， NVIDIA Spectrum-X 网络解决方案，支持 RDMA
- **交换机：**
  - NVIDIA Spectrum-X SN5600 以太网交换机，每个交换机有 64 个端口；
  - 支持 800Gb/s 速度，并可分割成 128 个 400 GbE 链路；
- **网卡：**
  - NVIDIA BlueField-3 SuperNIC，为每个 GPU 提供专用 RDMA 网络连接

业务面网络互联

- **技术：**
  - 采用 400GbE 以太网， 64 端口 800GbE 以太网交换机；
- **特点：**
  - 以太网而非InfiniBand等技术，以太网具有更好可扩展性，满足万卡规模 scale out 需求；
  - GPU 网络和 CPU 网络分离，以确保高性能计算集群最佳性能；

05

其他基础设施
风火水电

GPU Rack 冷却系统

- **散热方式：**
  - 液冷，
- **CDU 冷却液分配单元：**
  - 机架底部有 CDU（冷却剂分配单元）和冗余泵系统
- **冷却液循环：**
  - 冷却液通过机架分配管道进入每个服务器 CDU 分配器；
  - CDU 类似于大型热交换器，每个机架内都设有一个流体循环系统，为所有计算节点提供冷却服务；
- **其他：**
  - 仍保留风扇系统，用于冷却内存、电源单元、主板管理控制器、网卡等低功耗组件

其他冷却系统

- **CPU 服务器、网络设备和存储系统：**
  - 风冷散热，通过机架后部的热交换器将热量传递到液冷回路中；
  - 热交换器类似于汽车散热器，通过风扇将热空气抽过散热片，并将热量传递给循环水；
- **机房：**
  - 采用冷水循环系统，CDU 将热量传递到循环水中，热水在设施外部冷却后循环利用；
  - 供水管道将冷水引入设施，并循环流经每机架 CDU，吸收热量后，热水被引导至设施外部冷却装置。

电力系统

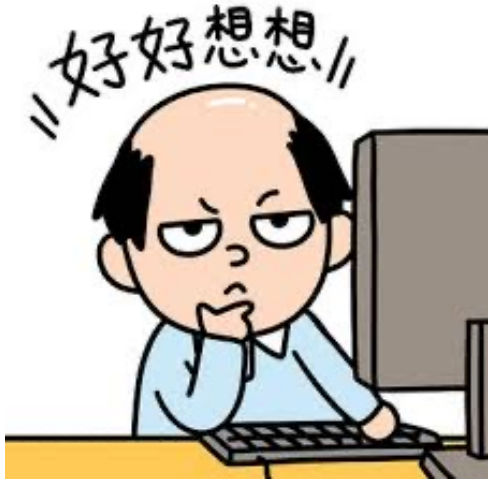
- **供电：**
  - 采用三相电源，每个机架配备多个电源条；
- **储能：**
  - 特斯拉 Megapack 电池组作为超级计算机和电网之间的能量缓冲器；
  - 每个 Megapack 可存储高达 3.9MWh 电能；
  - Megapack 引入解决 GPU 服务器功耗波动对电网造成的压力；

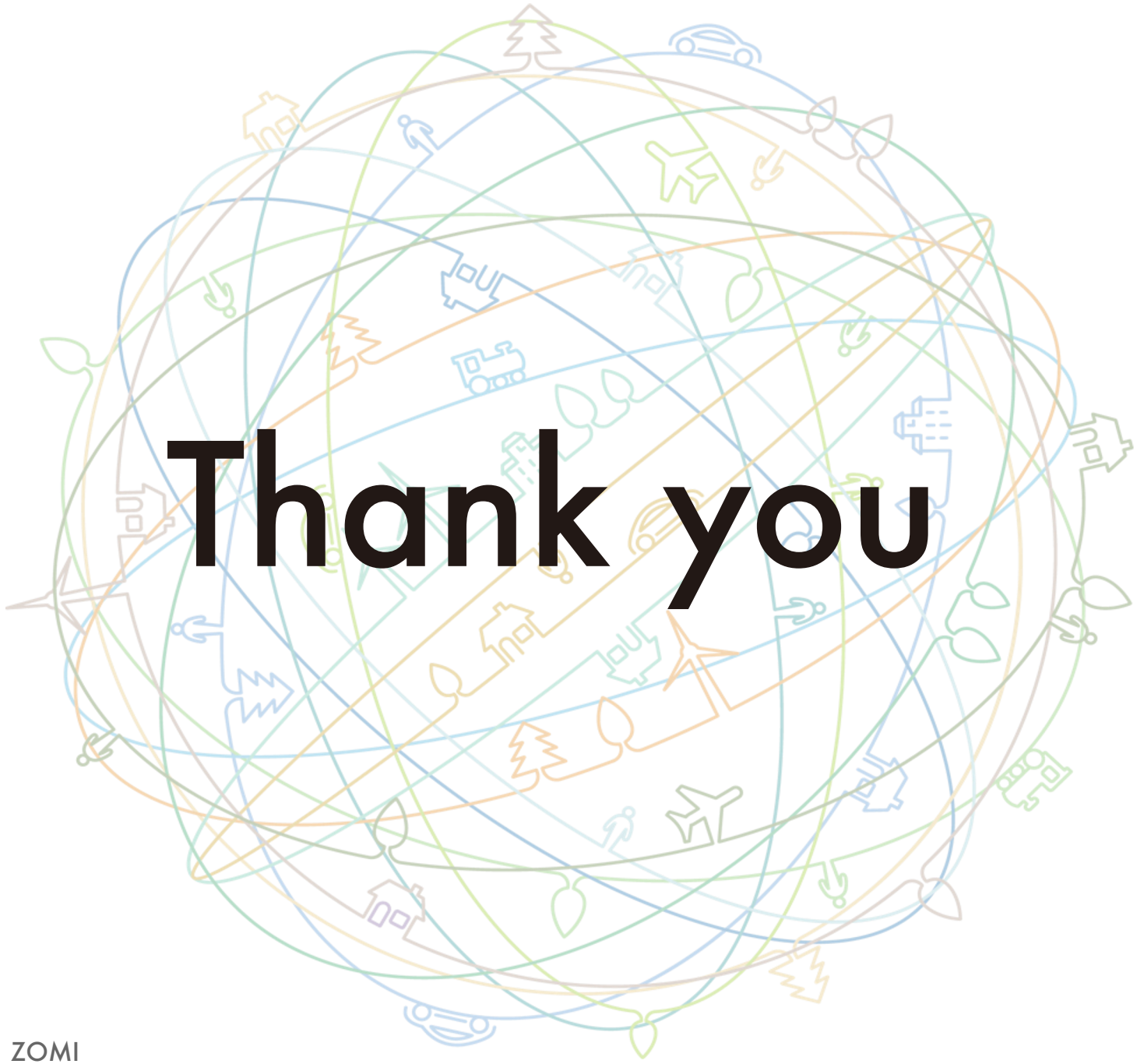
06

十万卡集群思考

思考

- **构建十万卡集群是一项复杂的系统工程：**
  - 不仅意味着算力指数级增长，还涉及复杂的技术和运营挑战
  - 需要解决高效能计算、高能耗管理、高密度机房设计、高稳定性训练等一系列问题
  - 最终能否将算力有效释放，还取决于算法、软件架构的优化与调度能力





Thank you

把AI系统带入每个开发者、每个家庭、每个组织，构建万物互联的智能世界

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

ZOMI

GitHub https://github.com/chenzomi12/AIFoundation

Reference 引用

1. https://www.youtube.com/embed/Jf8EPSBZU7Y
2. https://www.youtube.com/watch?v=Jf8EPSBZU7Y&t=1s