

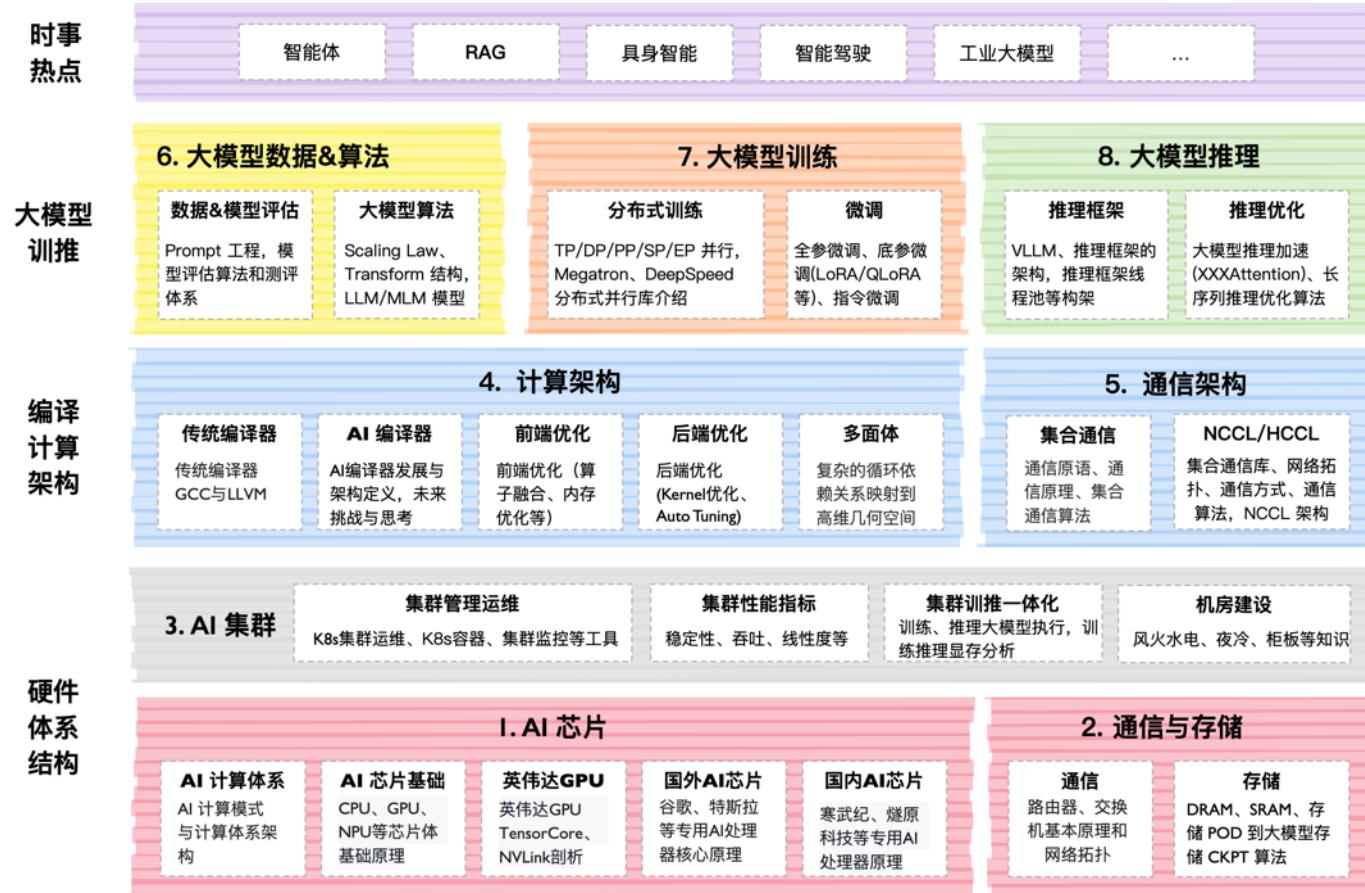
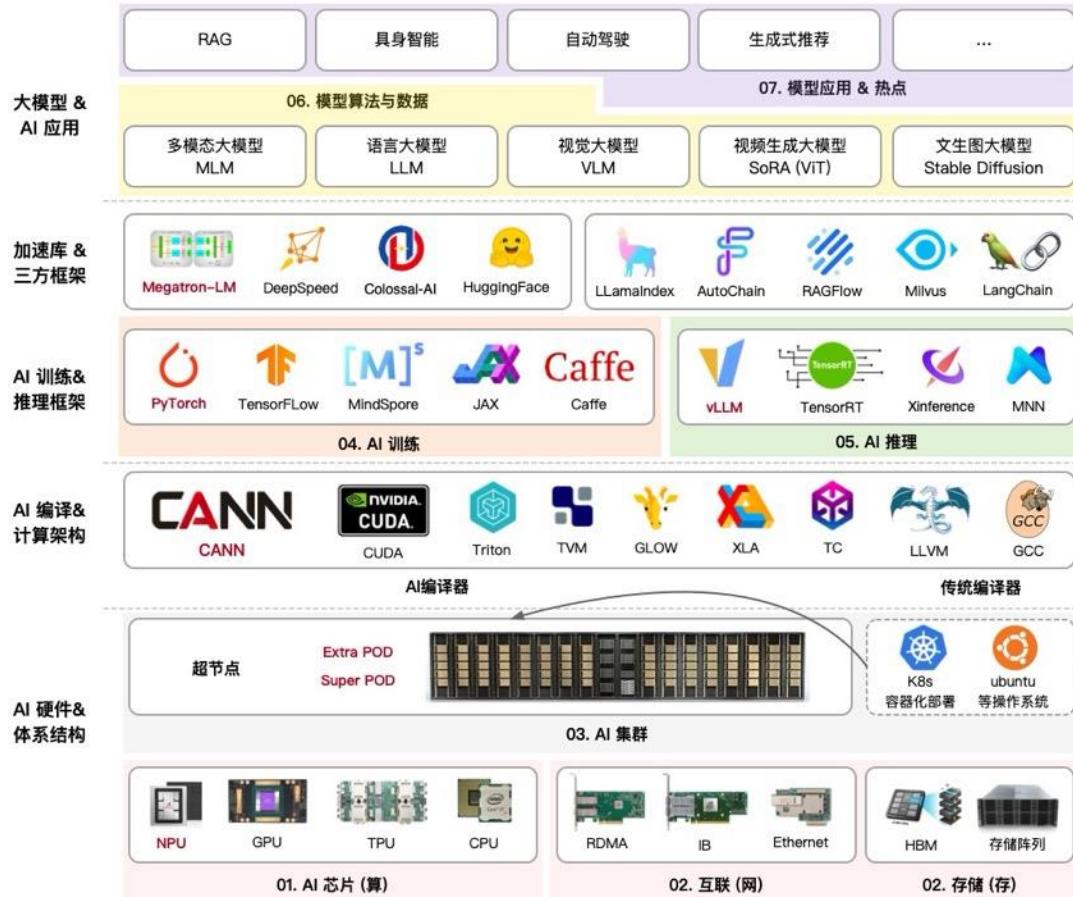
Question?

- AI 集群的组网方式，为什么不直接抄 HPC 的？要不全都是用胖树结构，要不全都在抄英伟达的组网方案？

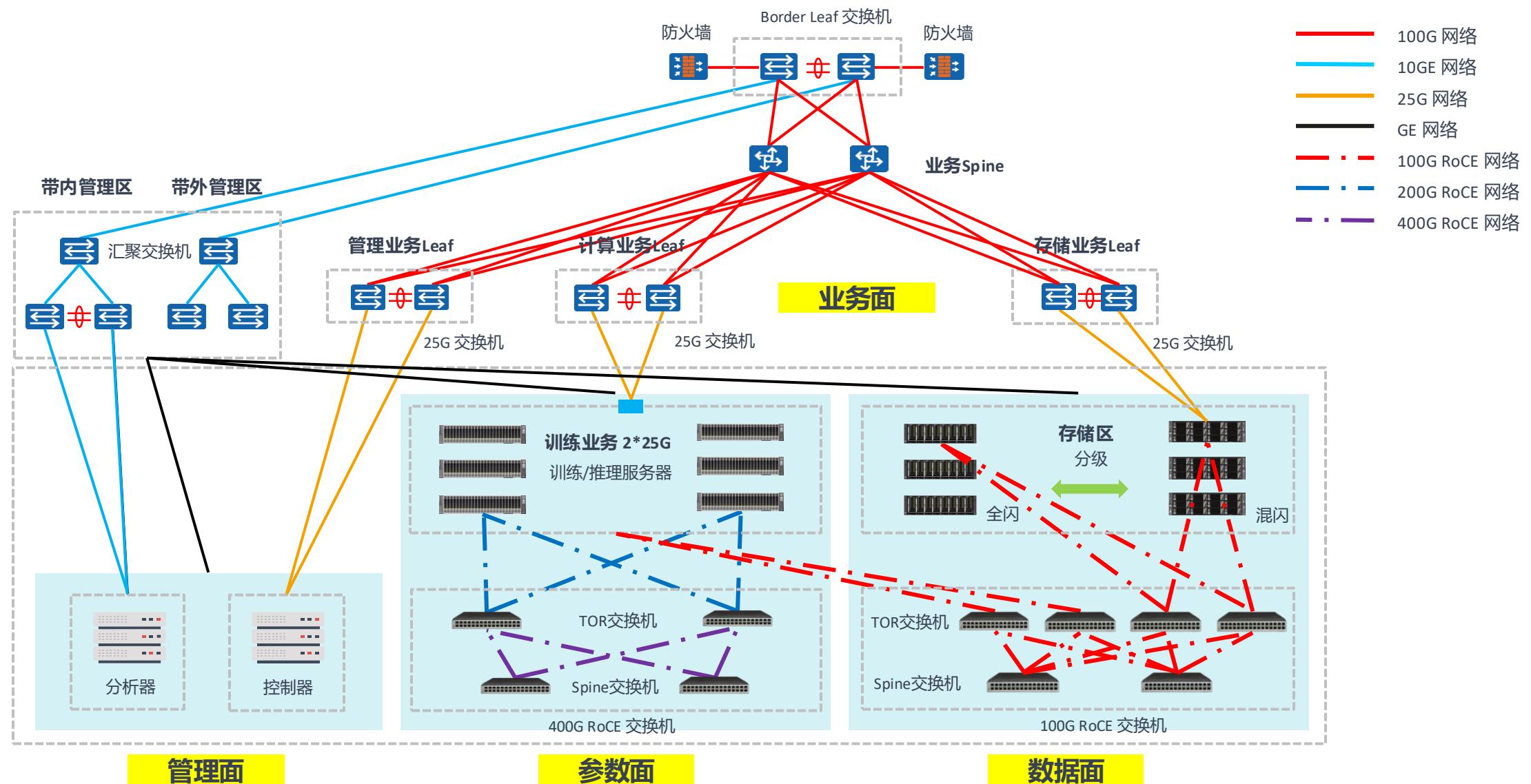


Content github.com/Infrasys-AI/AIInfra

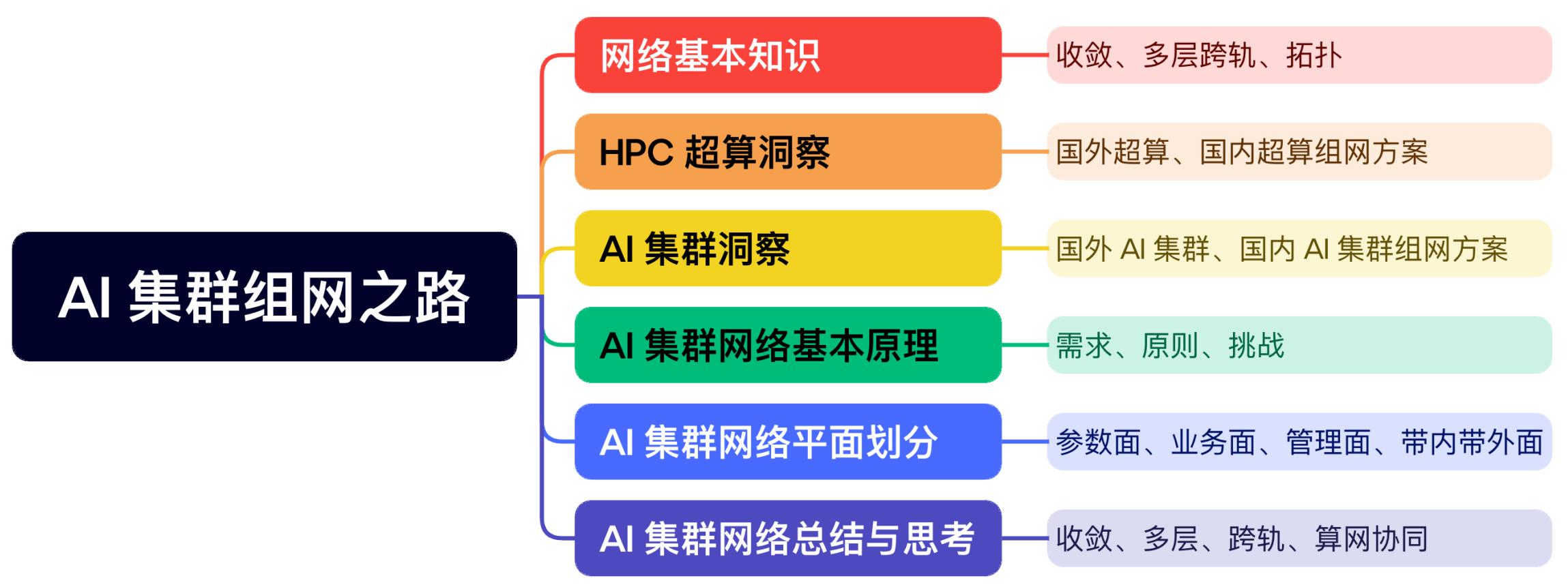
AI 系统 + 大模型全栈架构图



L2 算力底座：网络



Content



目录

1. 美国 AI 集群组网方式
2. 中国 AI 集群组网方式
3. 思考 & 总结

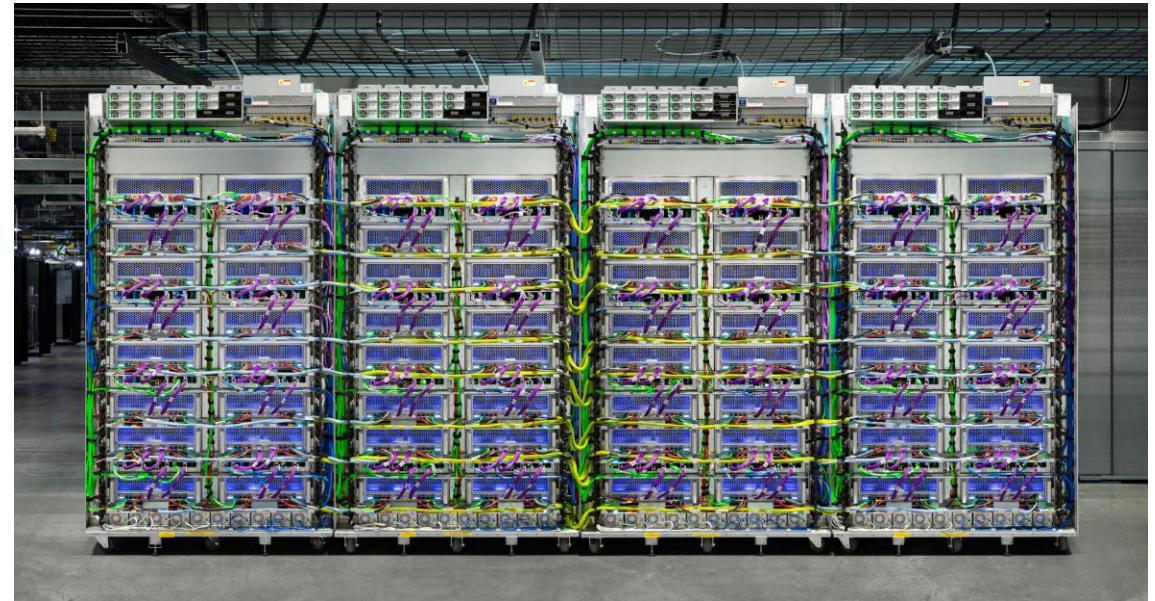
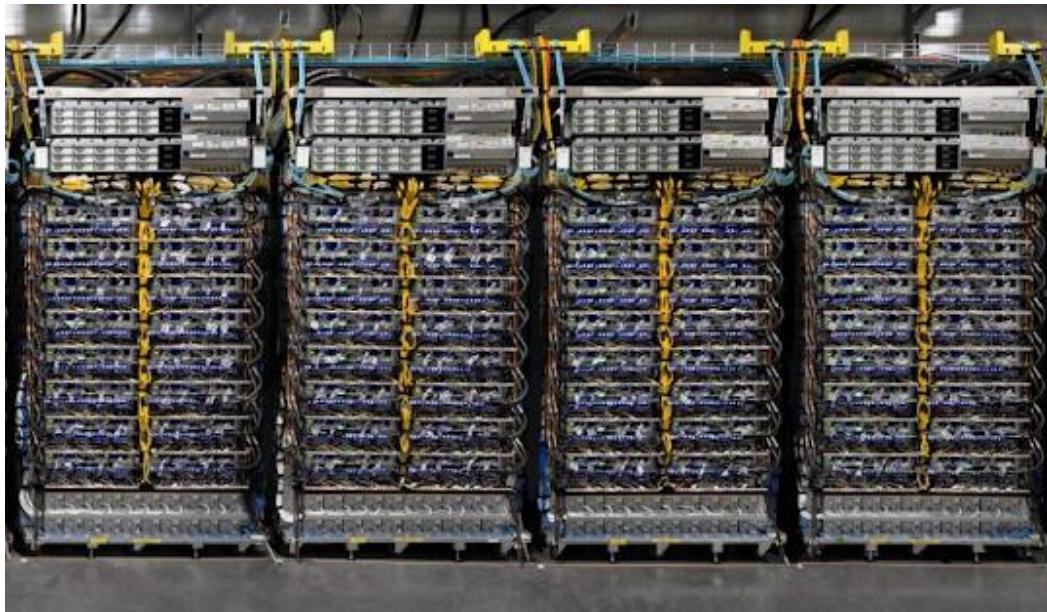


01 国外 AI 集群组网



Google TPU 洞察

- TPU v4 主打高性能训推一体，首次提出采用 OCS 光交换提升可靠性，通过 3D Torus 拓扑降低成本和功耗
- TPU v5e 主要面向推理，采用二维环面拓扑 (2D Torus) + 光路交换机 (OCS) 优化，相比 TPU v4 推理性价比提升 2.5x



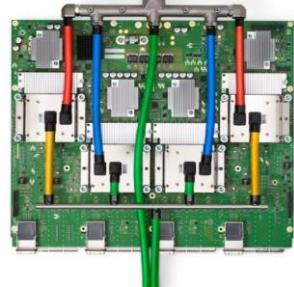
Google TPU 洞察

- TPU v4：主打高性能训推一体，采用 OCS 光交换提升可靠性，降低成本和功耗
- TPU Supercomputer：4096 TPU 通过 3D Torus 相连

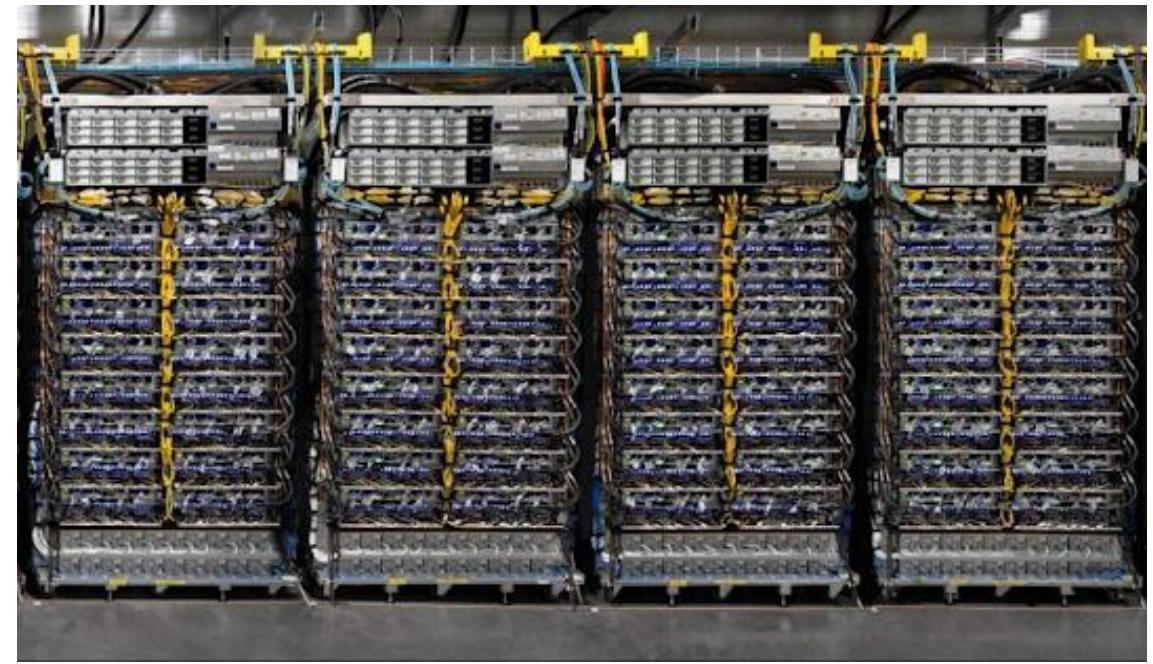
Chip



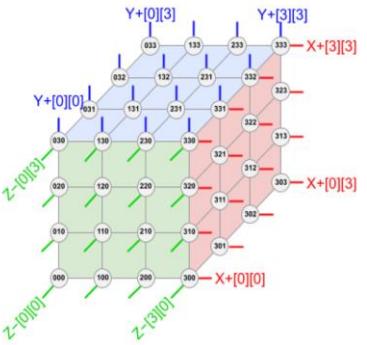
Board



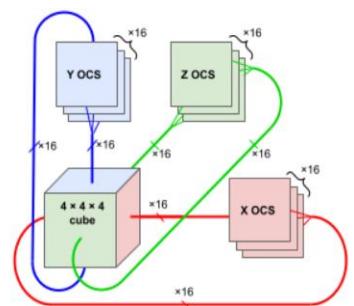
SuperComputer: 4*4*4 cube, 4096 TPU



Cube

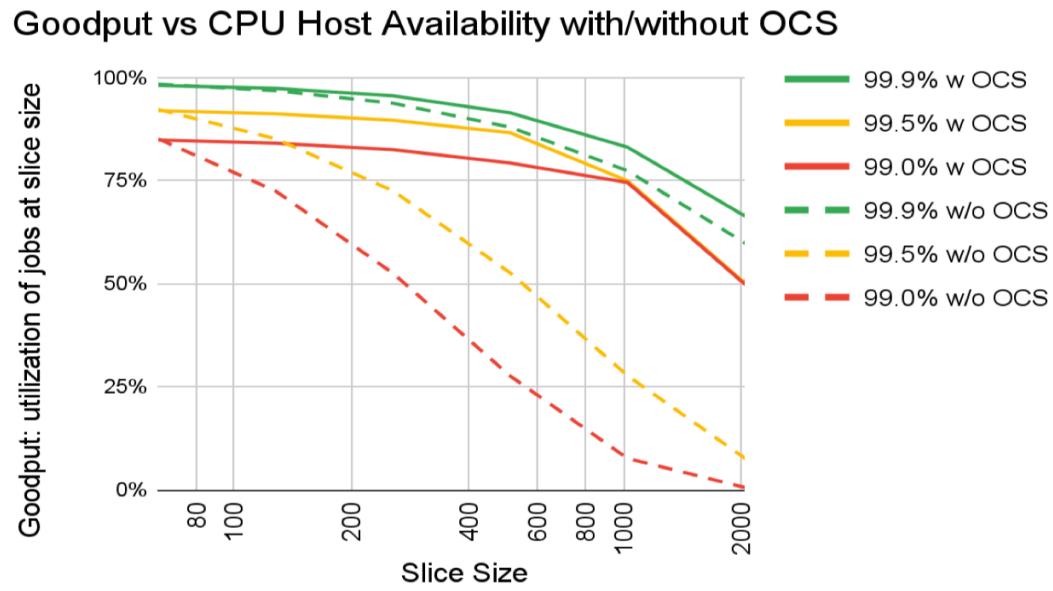


SuperComputer



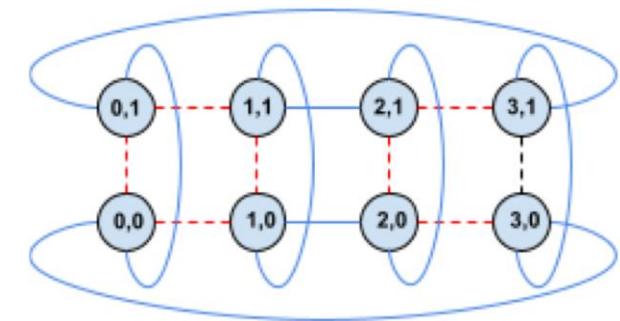
Google TPU 洞察

- 在 TPU 超算系统中部署 OCS，可以动态配置网络，绕过故障芯片，使得集群系统稳定性得以大幅提高；相同切片规模下，带有 OCS 的吞吐量更高

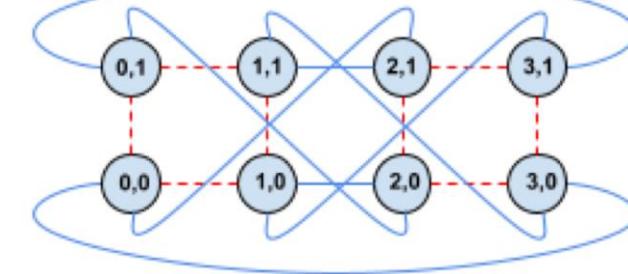


- OCS 光交换可重配光互连，可快速绕过故障节点，提升可靠性；降低成本和功耗。
- 降低系统成本和功耗：OCS 成本占 TPU 超算系统 5%，功耗占整个系统 3%

Regular Torus



Twisted Torus

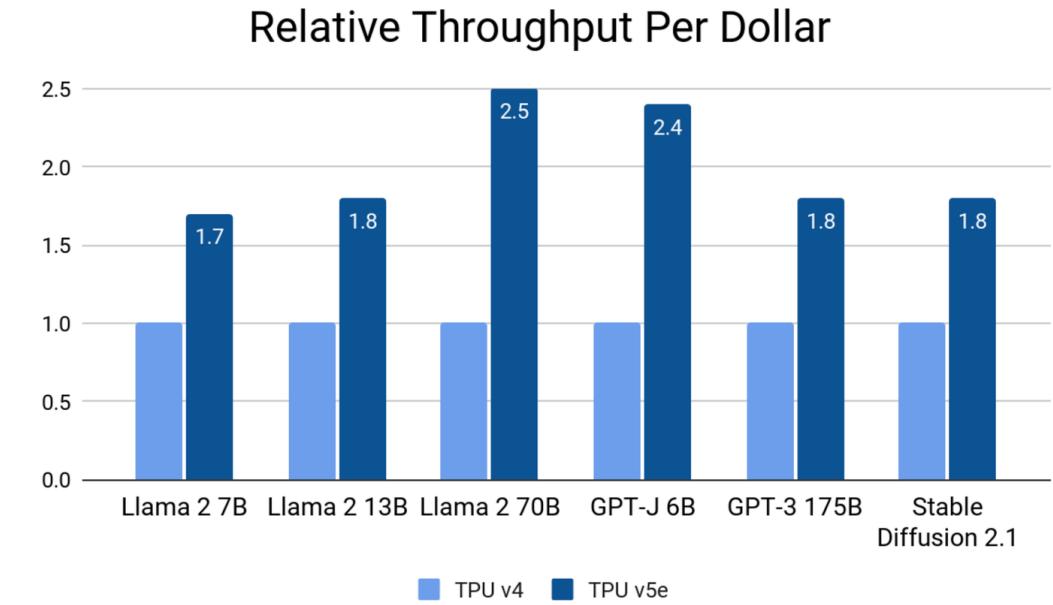


Google TPU 洞察

- TPU v5e：是TPU v5 lite版本，主打云上高性价比AI推理，兼中小模型训练

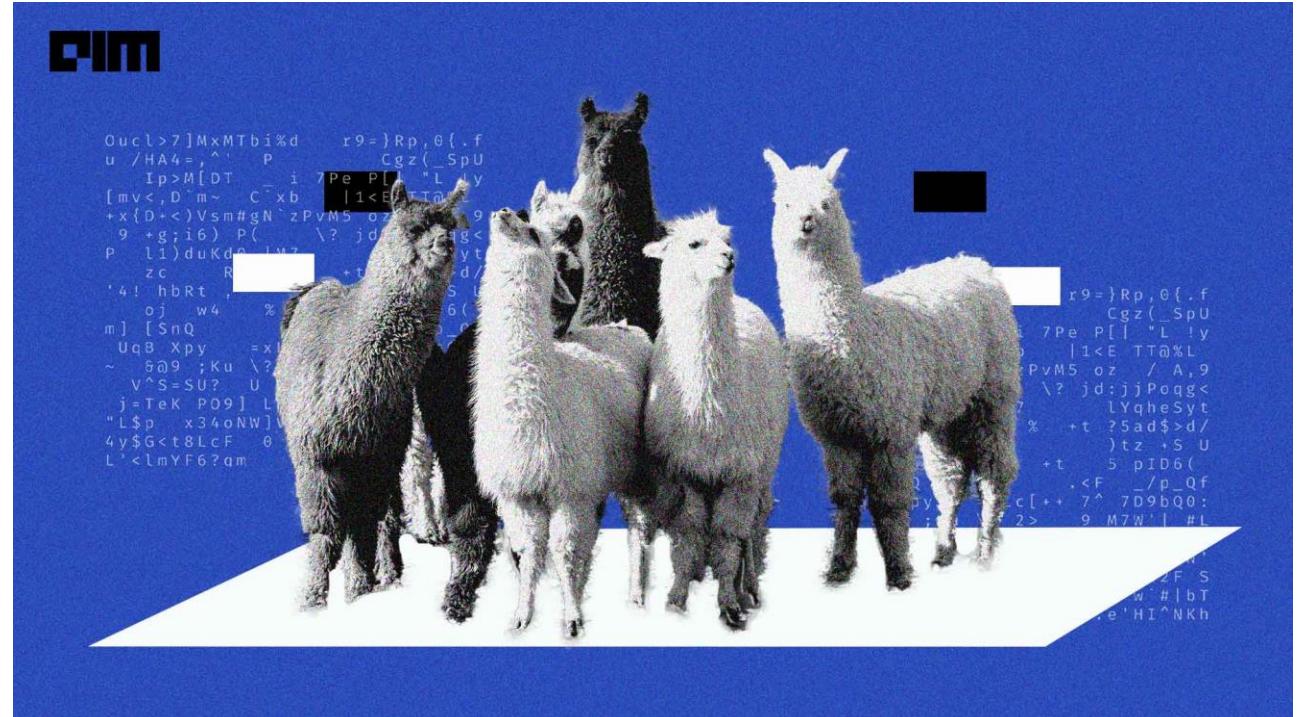
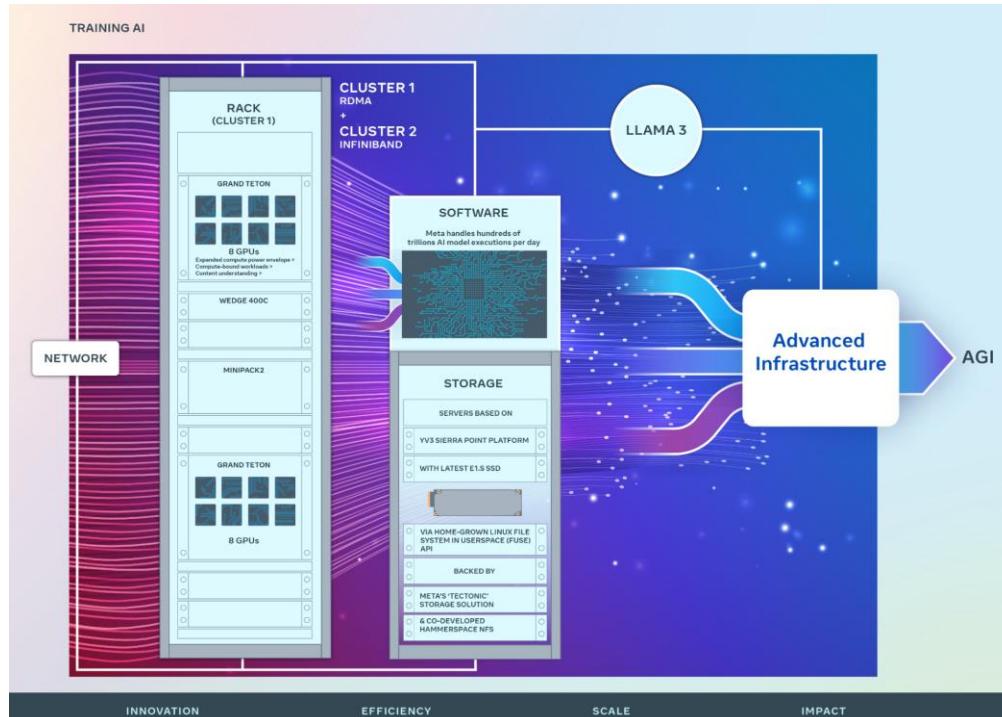
	TPU v5e	TPU v4
主要规格	单Die, TSMC 5nm	双Die, TSMC 7nm
峰值算力 BF16	197 TFLOPs	275 TFLOPs
HBM 容量 & 带宽	16GB, 819GBps	32GB, 1200GBps
Pod 规格	256 chip	4096 Chip
互连拓扑	2D Torus, w/o OCS w/o twisted Torus	3D Torus, w OCS w twisted Torus
片间互连带宽	1600Gbps	?
定位	推理为主，兼中小模型训练	训推一体

- v5e 与前代 V4 相比，每美元训练性能提升高达 2 倍
- 每 \$ 推理性能提升 2.5 倍，且 TPU v5e 成本不到 TPU V4 一半



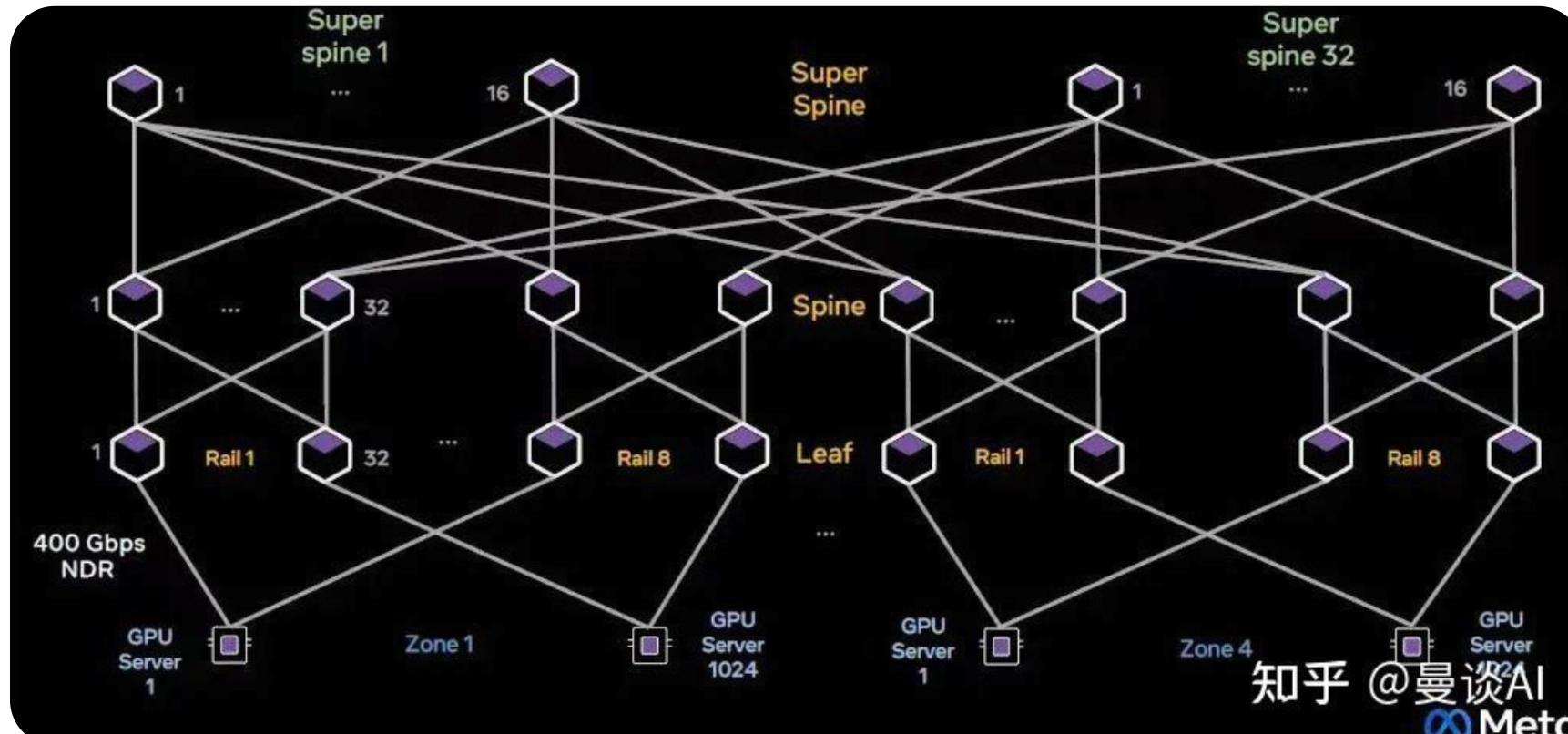
Meta 32K GPU 组网洞察

- 两个 GPU 集群，每个集群 3w+ H100，分别用 RoCE/InfiniBand 网络（2024/07/23）；
- LLaMA3 在以太网的24K 集群上训练出来的；
- 2024 年底，Meta AI 基础设施建设有 3.6w x 2 张 H100 GPU；



Meta 32K GPU 组网洞察：IB

- 通过Nvidia Quantum2 InfiniBand 网络构建 32K 集群，三层胖树组网，多轨架构实现
- L3 实现轨间互访，为规模做了分组分平面；L1/L2 配合接入 8K GPU 形成一个 Zone(x4)

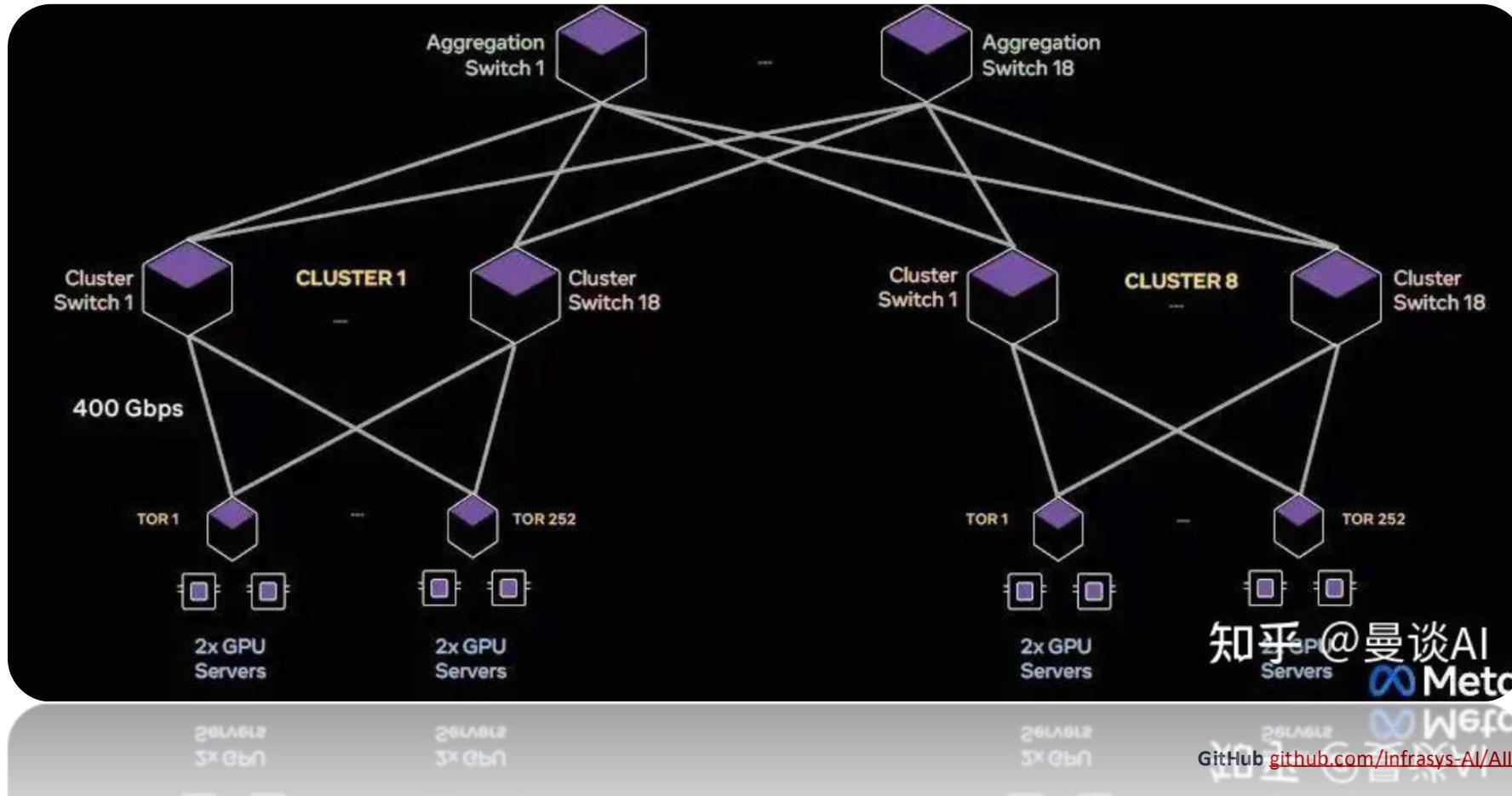


知乎 @曼谈AI
Meta

GitHub github.com/Infrasys-AI/AllInfra book [infrasys-ai.github.io](https://github.com/infrasys-ai/github.io)

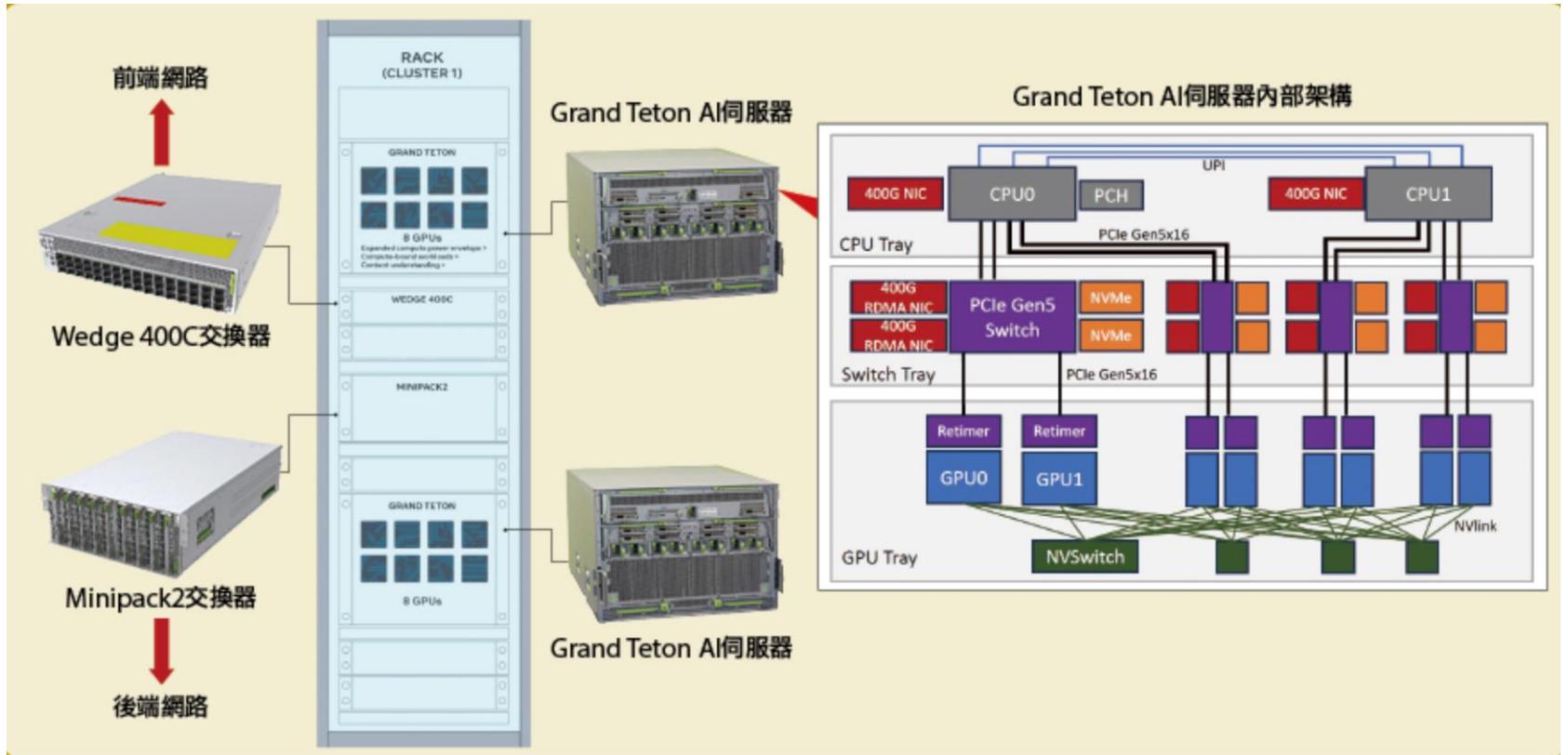
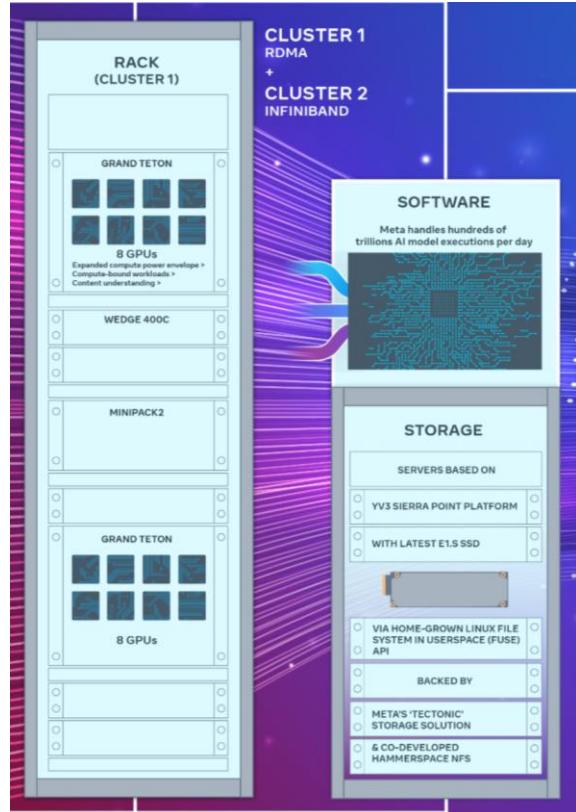
Meta 32K GPU 组网洞察：RoCE

- 基于 Arista 公司 Arista 7800 机架交换机以太网 RoCE 构建 32K 集群，采用单轨架构
- 同样三层胖树组网，L1/L2 配合接入 4K GPU 形成一个 PoD，集群共 8 个 PoD



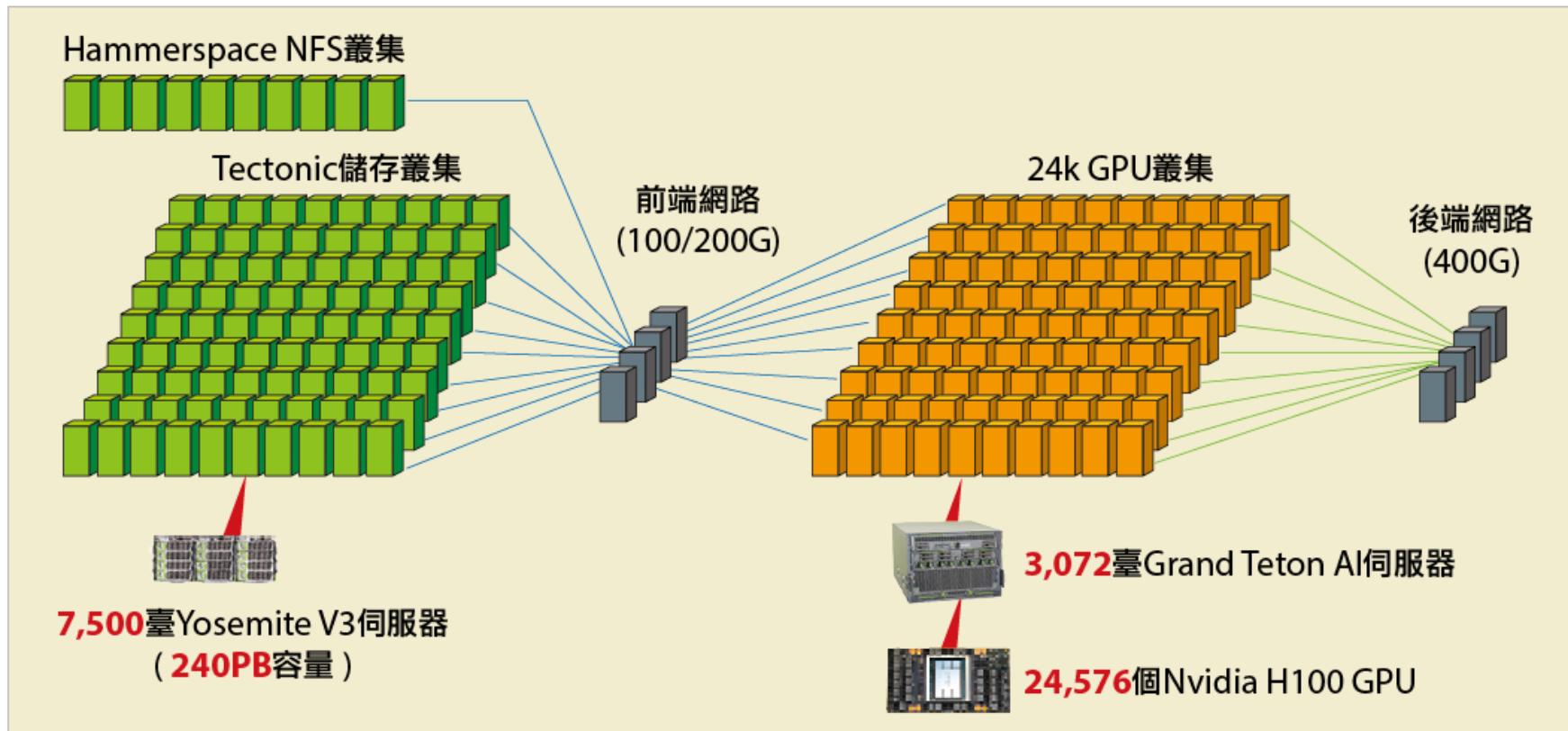
Meta 32K GPU 组网洞察

- 计算柜 Rack: 两个计算节点 (8 张 H100) + wedge 400C 交换机 + Minipack2 交换机
- 节点内 NVLink 互联: 8张 GPU 靠 4 个 Nvsiwtch3 实现 18 个端口的 clos 全互联



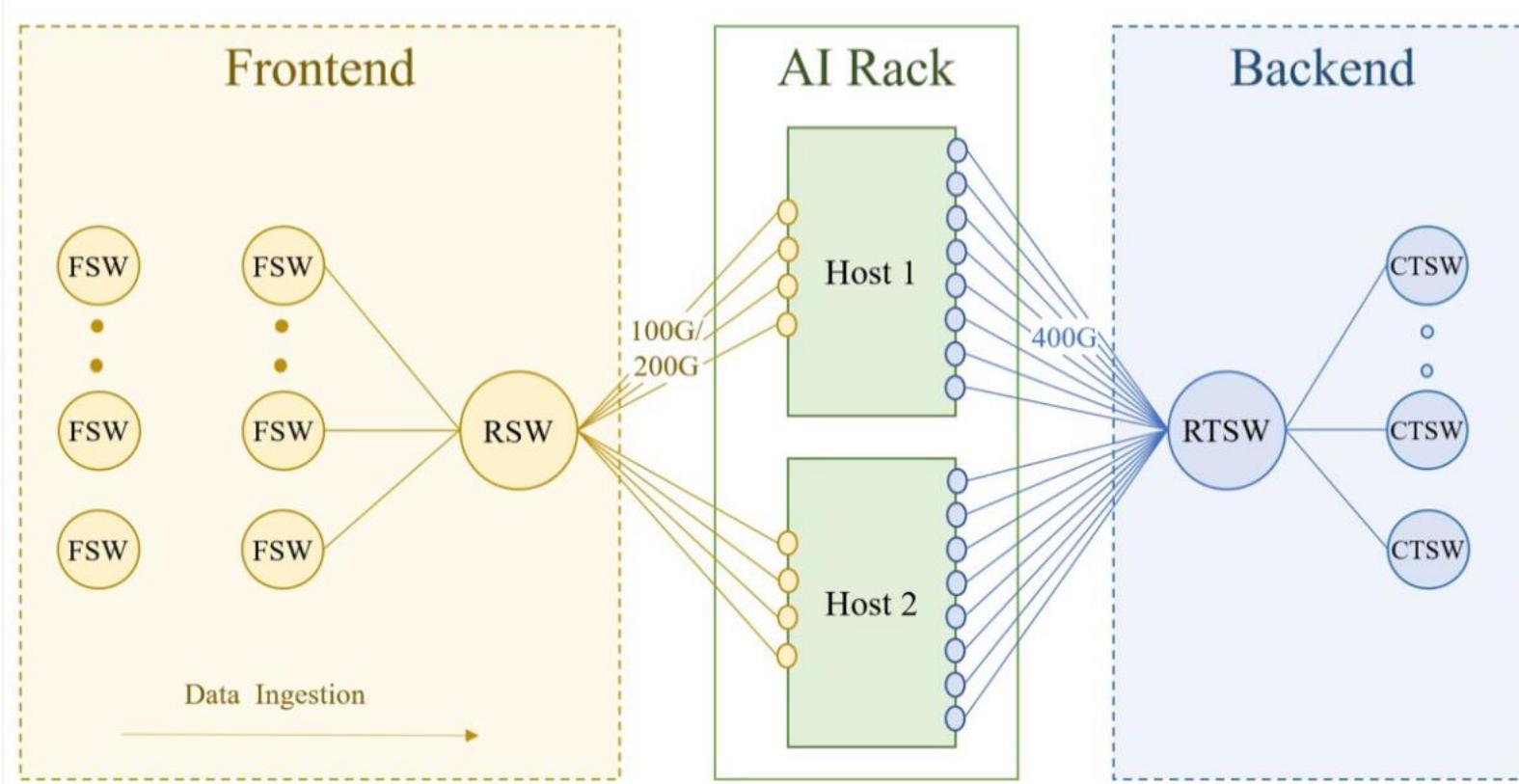
Meta 32K GPU 组网洞察

- 计算柜 Rack: 两个计算节点 (8 张 H100) + wedge 400C 交换机 + Minipack2 交换机
- 节点内 NVLink 互联: 8张 GPU 靠 4 个 Nvsiwtch3 实现 18 个端口的 clos 全互联



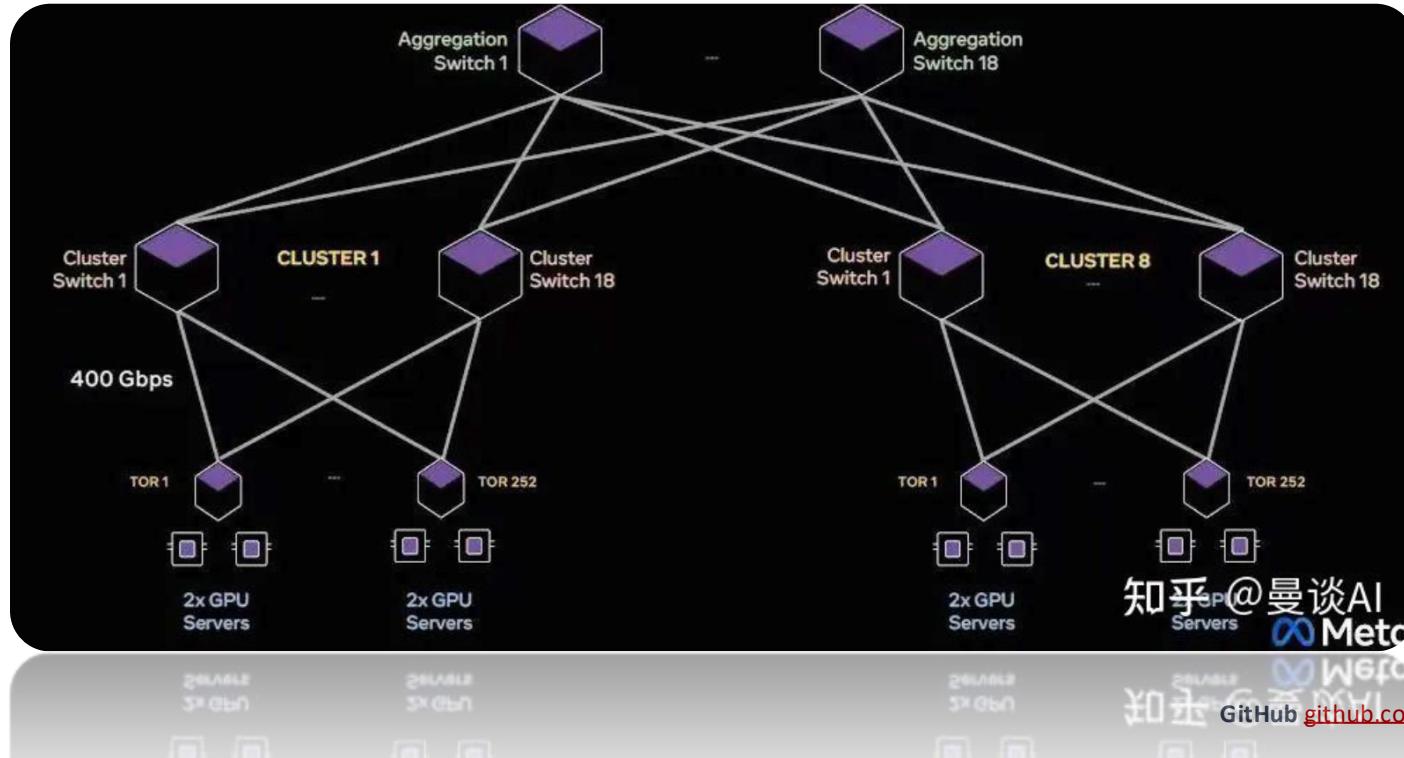
Meta 32K GPU 组网洞察

- 计算柜 Rack: 两个计算节点 (8 张 H100) + wedge 400C 交换机 + Minipack2 交换机
- 节点内 NVLink 互联: 8张 GPU 靠 4 个 Nvsiwtch3 实现 18 个端口的 clos 全互联



Meta 32K GPU 组网洞察

- 节点间单轨，三层组网架构 RoCE 网络互联：
 1. L1 Minipack2：基于TH4构建，单接口板 16*200G/8*400G，最大 8 个接口板两种接入方案，接多框 H100
 2. L2 Arista 7800：单接口板 36 个 400G，选择 8 槽位型号接入 192Rack（3072 张 GPU），构建无收敛 PoD
 3. L3 Arista 7800：选择 8 槽位型号，7:1 收敛接入 8 个 PoD，构建 24K GPU 集群



知乎@曼谈AI



GitHub

github.com/Infrasys-AI/AllInfra

book [infrasys-ai.github.io](https://github.com/infrasys-ai/github.io)



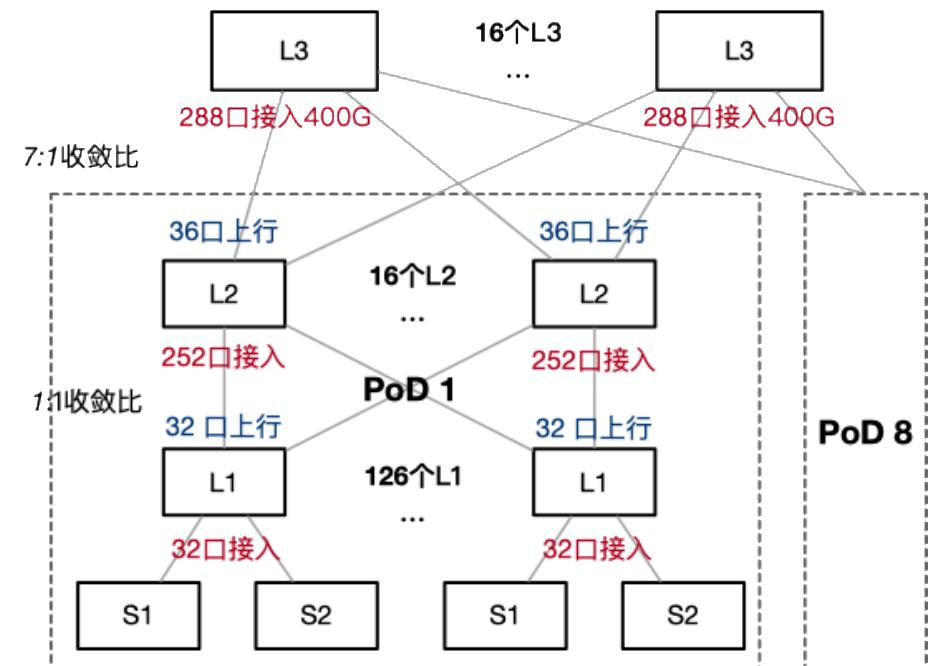
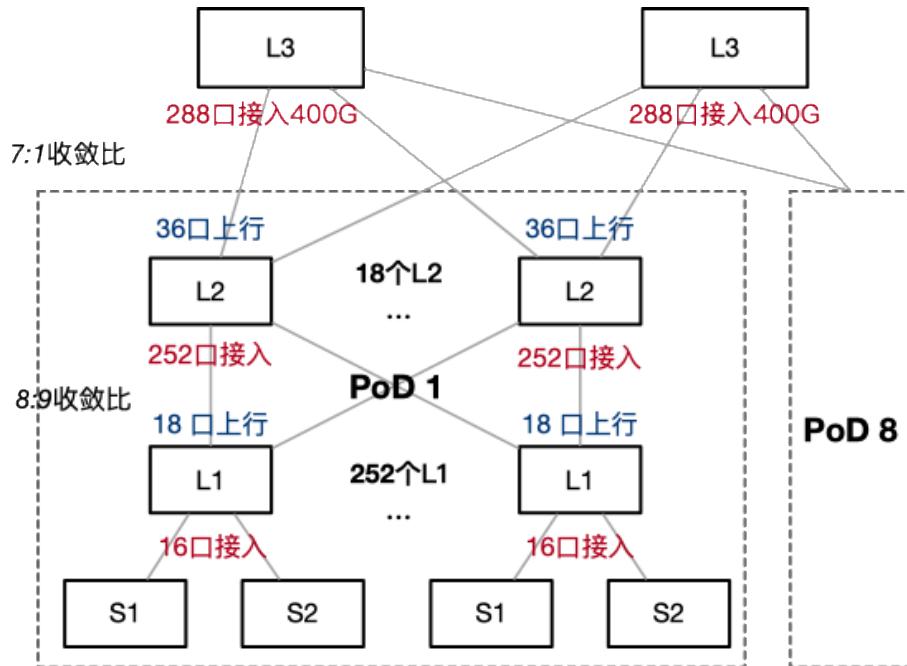
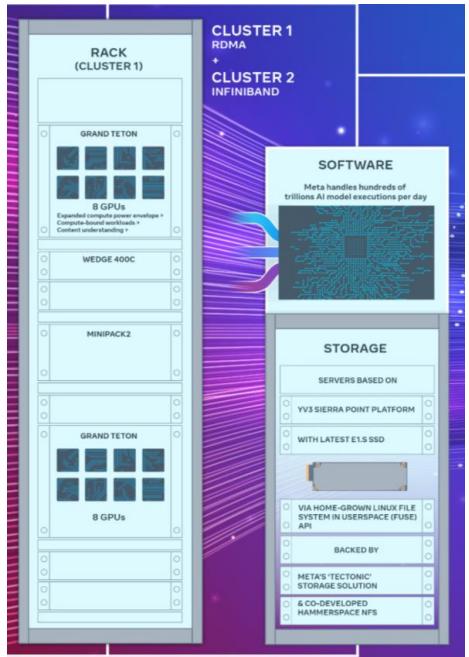
ZOMI

18

Meta 32K GPU 组网洞察

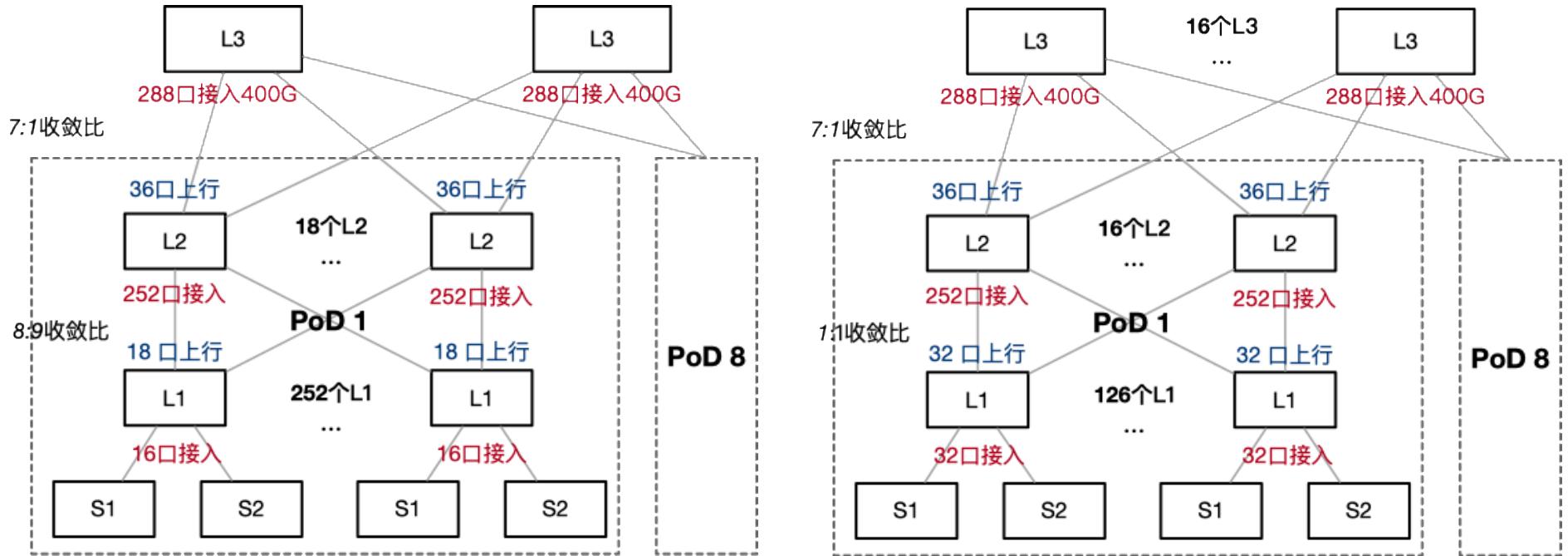
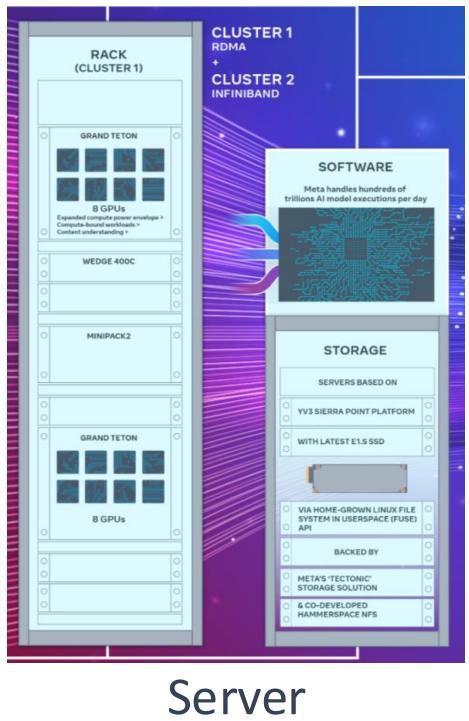
- 节点间单轨，三层组网架构 RoCE 网络互联：

1. L1 Minipack2：基于TH4构建，单接口板 16*200G/8*400G，最大 8 个接口板两种接入方案，接多框 H100
2. L2 Arista 7800：单接口板 36 个 400G，选择 8 槽位型号接入 192Rack (3072张GPU)，构建无收敛 PoD
3. L3 Arista 7800：选择 8 槽位型号，7:1 收敛接入 8 个 PoD，构建 24K GPU 集群



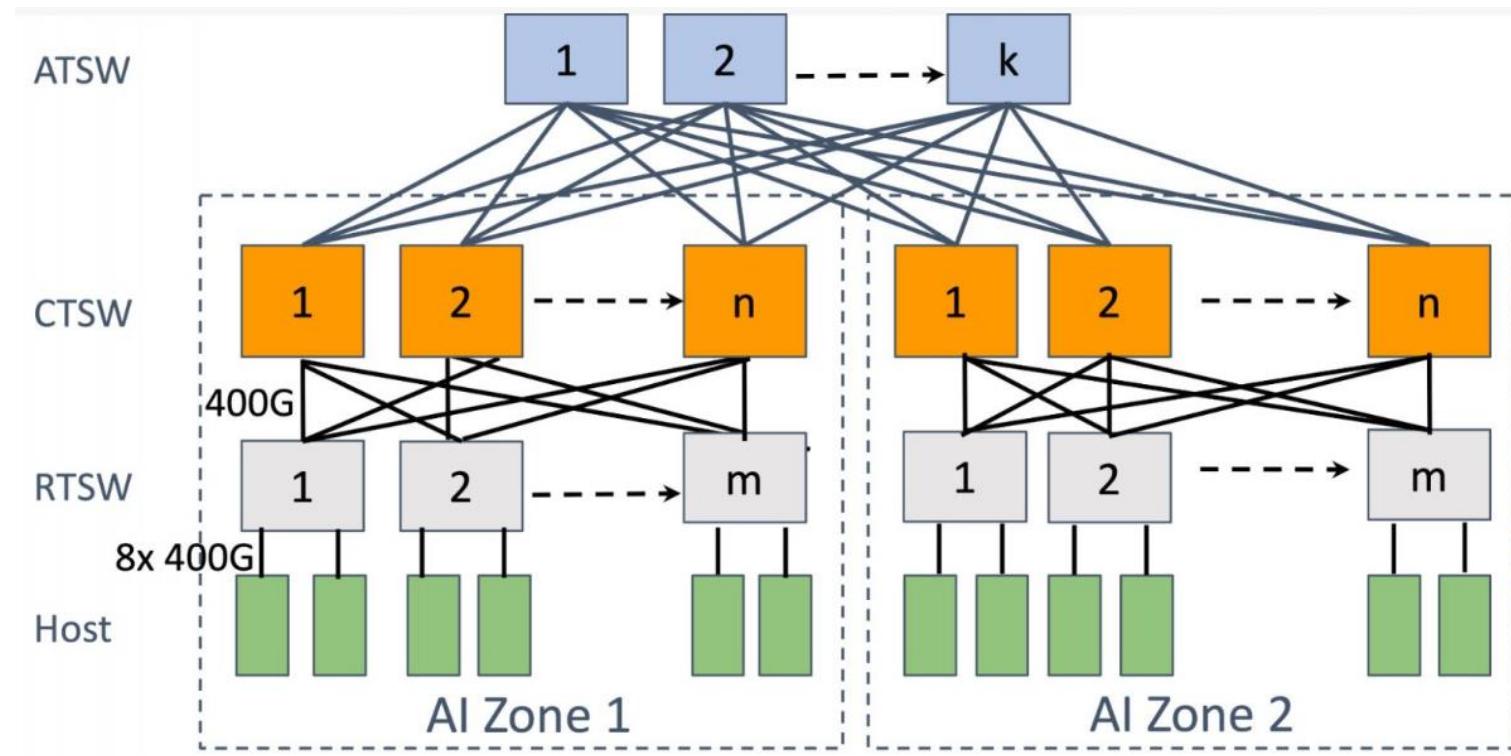
Meta 32K GPU 组网洞察

- 32K 集群，实际训练 Llama3 只用了24K，8 个 PoD，每Pod 252 计算框只用192个
- minipack2 只接入 16 张卡，每个 L1 上行接入 18 个 L2，L3 接入 8 PoD
- 好处：成本降低；坏处：跨框复用 L1，连接复杂度增加



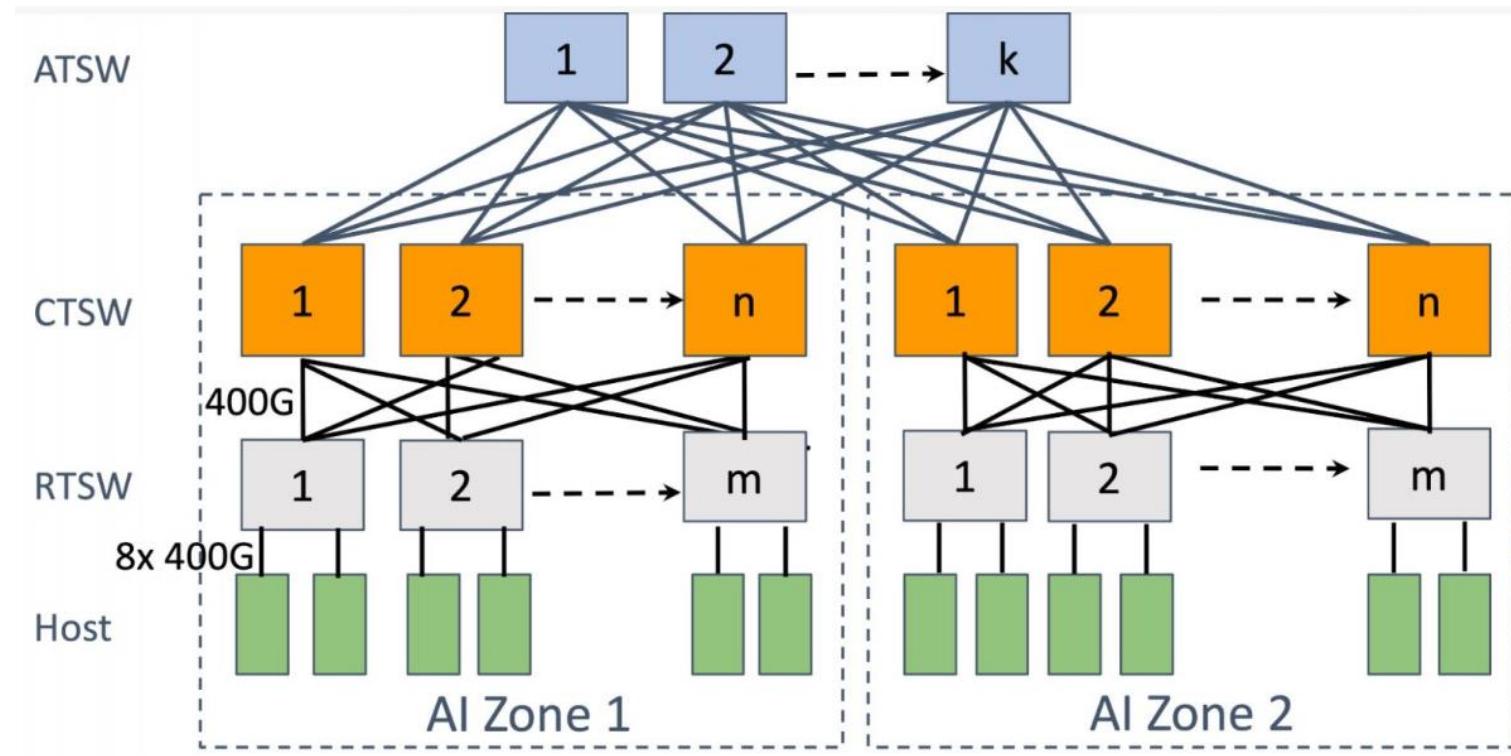
Meta 32K GPU 组网拥塞控制

- 最初采用传统 DCQCN 来进行拥塞控制，发现 400Gbps 网络上效果不佳。
- Spine 交换机使用 Deep-buffer 利用 HBM 来维持多个传输通道作为传输缓冲区，应对集合通信引起瞬时拥塞。



Meta 32K GPU 组网洞察

- 集合通信基于 NCCL 库的分支 NCCLX，针对大时延网络提高性能；
- 基于拓扑优化并行化配置 TP/CP/PP/DP，PyTorch + NCCLX 强关联，快速异常检测和定位；



Meta洞察

- meta 在 ai-hardware-summit (2024/05) 认为后续方向：
 1. Scale-up 会扩大到节点
 2. Scale-up 会跟 Scale-out 二网合
 3. Scale-up 互访带宽是 scale-out 或 10 倍



02

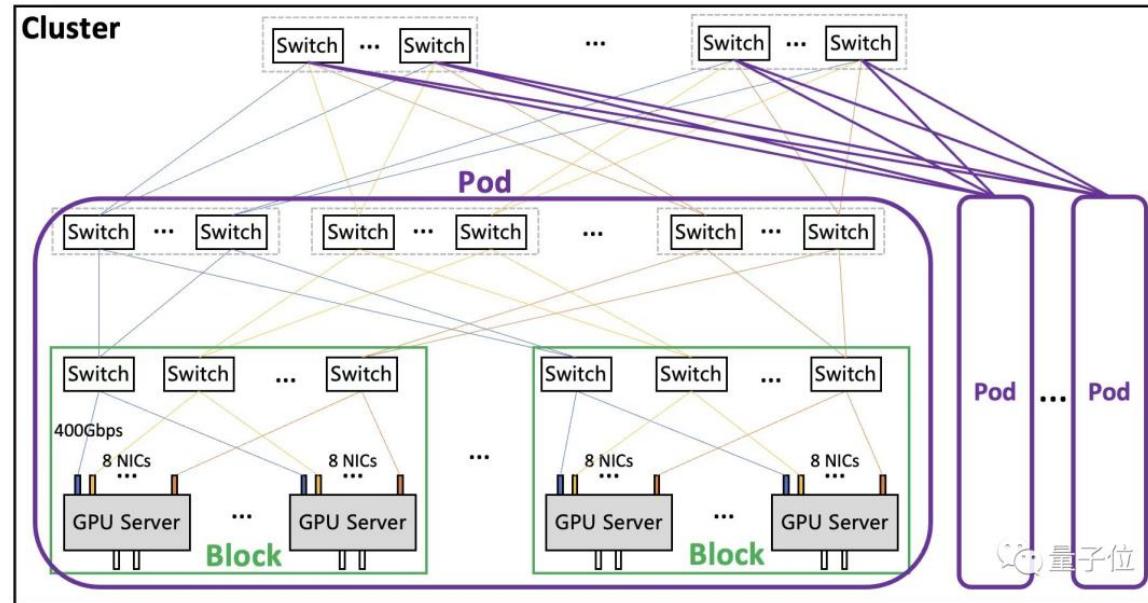
中国

AI 集群组网

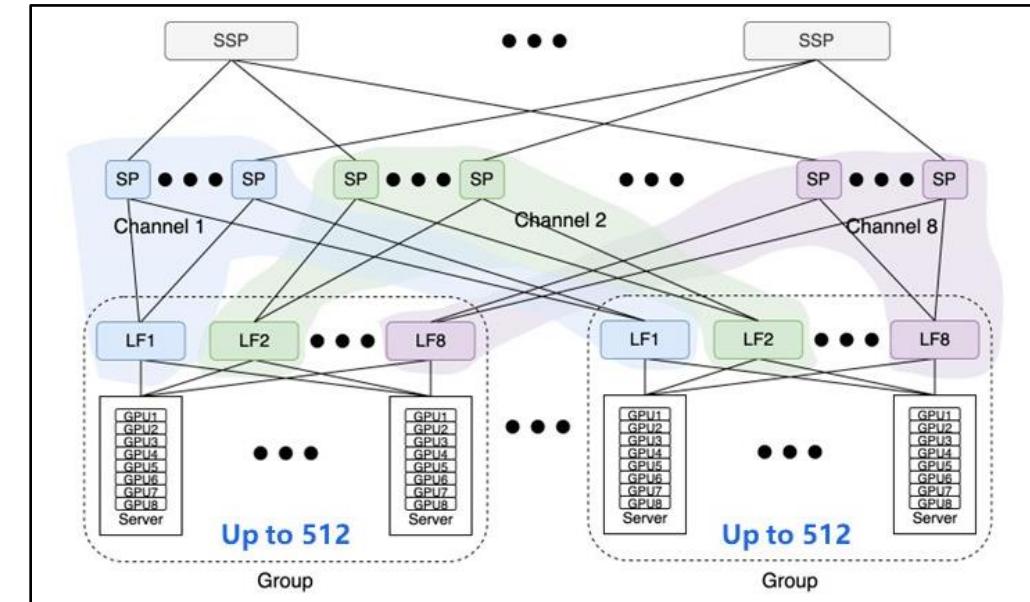


腾讯、百度、字节：三层盒式交换机，无收敛 Clos 组网

- 均采用无收敛 Clos 组网：每 server 8 卡 8 轨道接入
- 差异点：交换机转发容量和每卡接入带宽
- PS：百度认为网络成本在集群整体里相对较小比例，通过无收敛冗余设计带来性能增益优于成本提升



腾讯

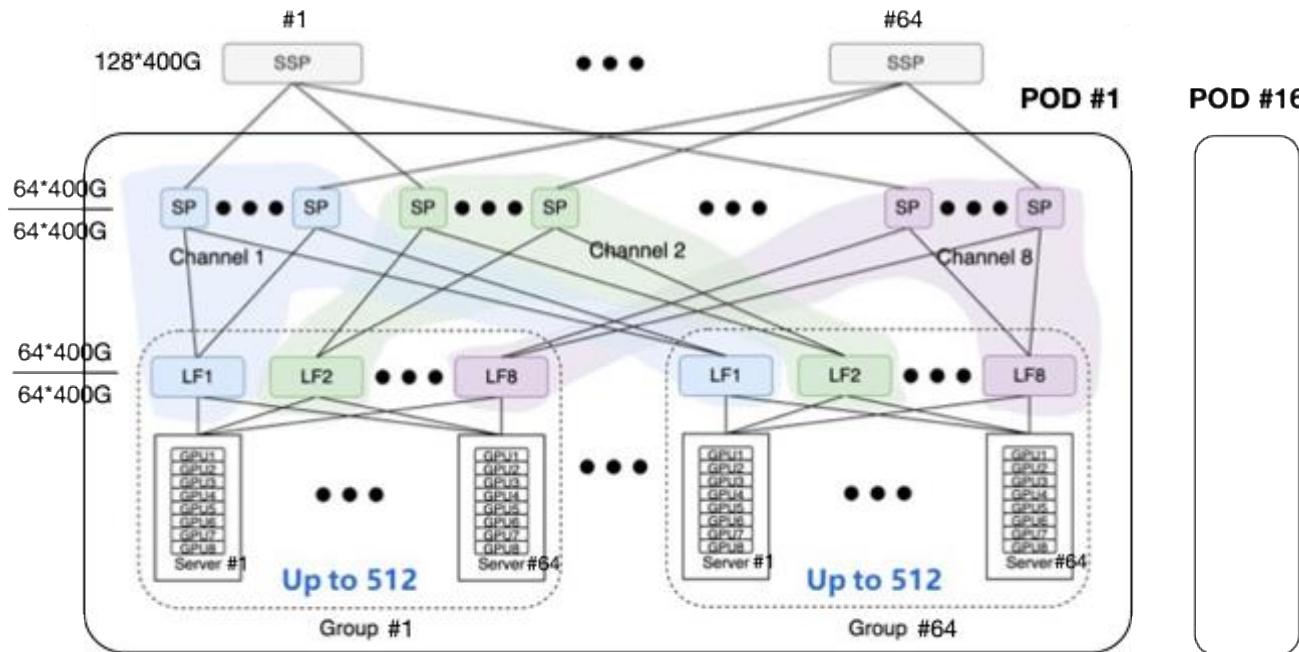


百度



百度：三层盒式交换机，无收敛 Clos 组网

- 采用三层无收敛 CLOS 组网，确保网络高效可靠。
- 设计 8 通道硬件隔离，优化 GPU 通信效率。
- 实现 1:1 无阻塞带宽，避免遇到网络性能瓶颈。
- 选用自研 51.2T 交换芯片，支撑万卡 GPU 超大规模集群。

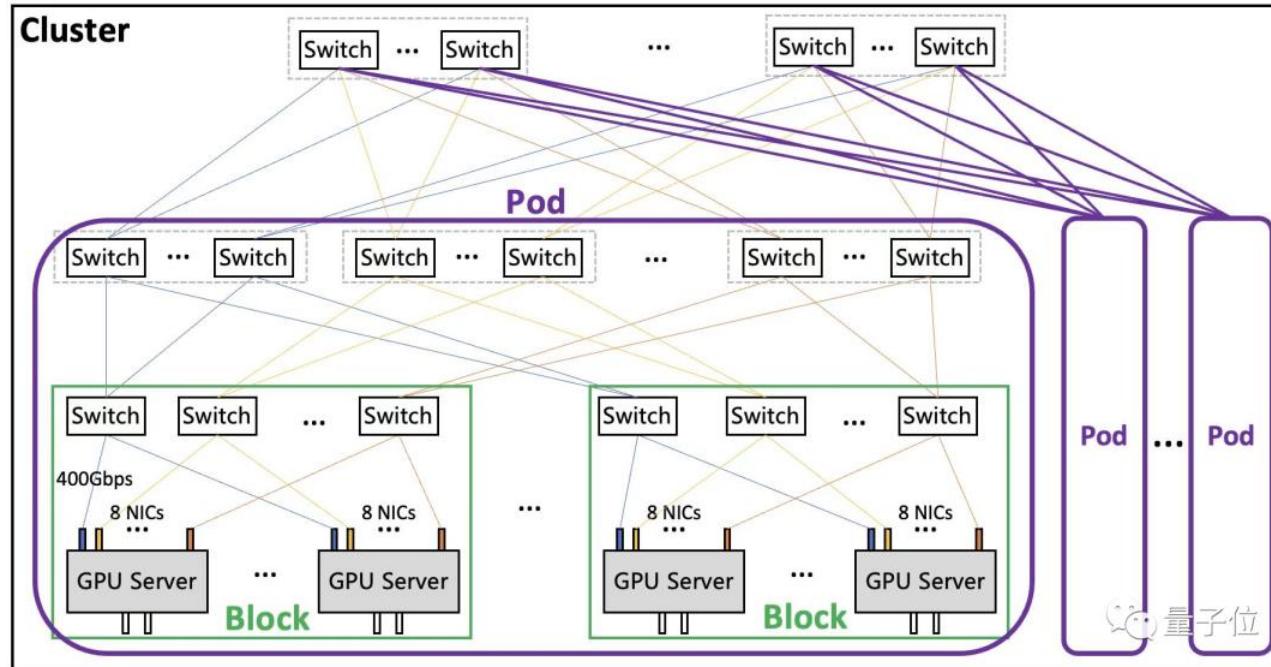


- 每 Server: 8 GPU, 每 GPU 400G 接入
- 每 Group: 64 server (512 GPU)
- 每 Pod: 64 Group (32k)
- Cluster: 16 Pod (512k)
- TOR/Spine/Core: 51.2 Tbps



腾讯：三层盒式交换机，无收敛 Clos 组网

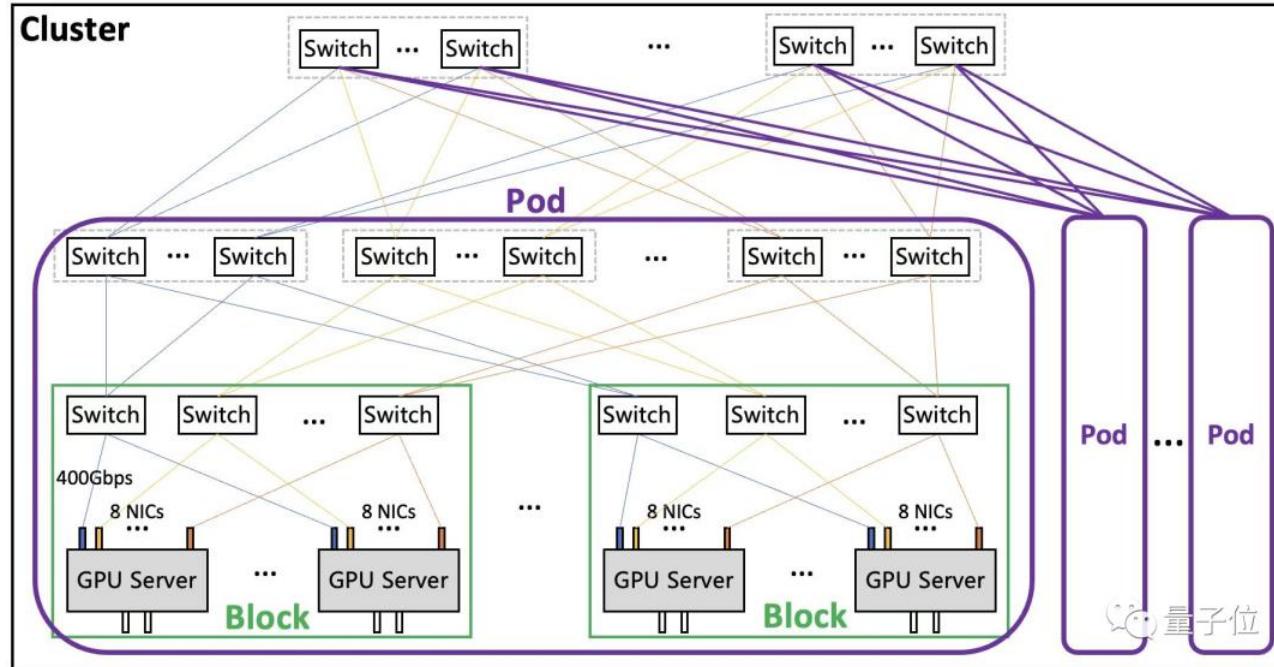
- 基于 NV GPU + Broadcom 芯片构建十万卡集群，三层无收敛胖树，走多轨路线
- 节点内 NVLink 平面互联：节点内 8 张 GPU 4 个 Nvsiwtch3 实现 18 个端口 clos 全互联
- 节点间多轨架构：三层组网架构的无收敛 RoCE 网络互联
- 32 个服务器节点：上行 32 个 400G 光口，分别接入 32 个 L2，Block 接入规模=32*8=256



- 每 Server: 8 卡，每卡 400G 接入
- 每 Block: 32 server ($8 \times 32 = 256$)
- 每 Pod: 64 Block ($256 \times 64 = 16k$)
- Cluster: 16 Pod (256k)
- Spine/Core: 51.2 Tbps
- ToR (推测) : 25.6 Tbps

腾讯：三层盒式交换机，无收敛 Clos 组网

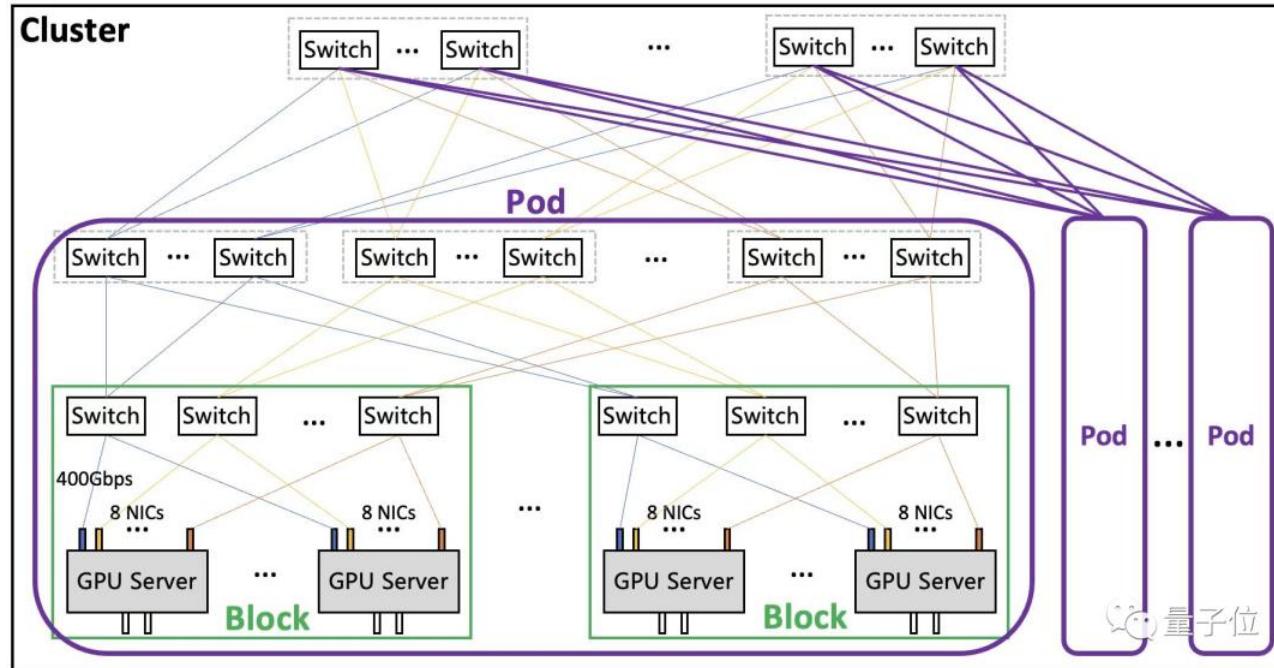
- L1层 25T 芯片设备：64 个 400G，接入 32 个 400G 光口
- L2 50T芯片设备：128 个 400G 光口，接入同轨 9964 个 L1
- 上行 64 个 400G 光口，可接入 64 个 L3 节点，Pod 接入规模=256*64=16K
- 具体使用 TH5 芯片设备、L1/L2 间多链路互联，避免会减少规模



- 每 Server: 8卡，每卡 400G 接入
- 每 Block: 32 server ($8*32=256$)
- 每 Pod: 64 Block ($256*64=16k$)
- Cluster: 16 Pod (256k)
- Spine/Core: 51.2 Tbps
- ToR (推测) : 25.6 Tbps

腾讯：三层盒式交换机，无收敛 Clos 组网

- L3 50T芯片设备：128 个 400G 光口
- 接入 128 个 L2，16 个 PoD*8 个轨道，实现每轨道 8 个 L2，实现轨间互访
- Cluster 接入规模 $16K \times 16 = 256K$



- 每 Server: 8卡，每卡 400G 接入
- 每 Block: 32 server ($8 \times 32 = 256$)
- 每 Pod: 64 Block ($256 \times 64 = 16k$)
- Cluster: 16 Pod (256k)
- Spine/Core: 51.2 Tbps
- ToR (推测) : 25.6 Tbps

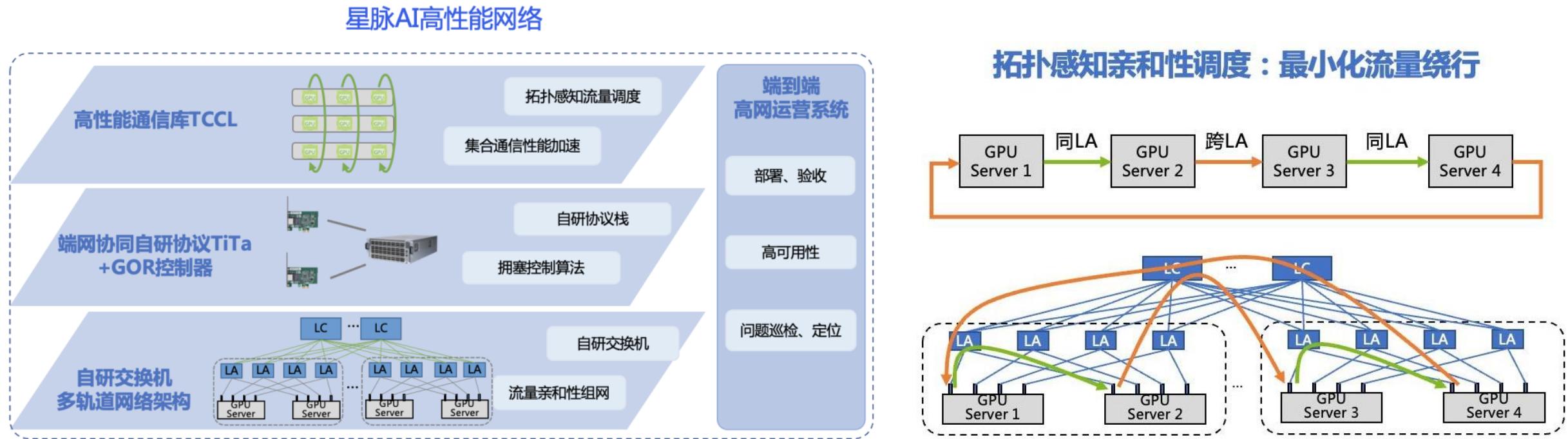
Question

特性	TH4	TH5	意义与提升
互联技术	主要支持 InfiniBand	全面转向以太网技术路线	战略转变：利用更开放、更成熟的以太网生态，降低成本，提高兼容性和可扩展性。
性能	支持 100Gb/s 单端口速率	支持 400Gb/s 单端口速率，并支持下一代 800Gb/s	性能巨大提升：带宽是 TH4 的 4 倍，能应对更庞大的数据交换需求。
关键特性	实现了 RDMA（远程直接内存访问）	在高速以太网上实现了 RDMA，并支持融合以太网上的 RDMA	低延迟：RDMA 允许计算机直接访问另一台计算机的内存，无需 CPU 介入，极大降低延迟和开销。这是高性能互联的核心技术。
应用场景	用于腾讯内部星脉网络第一代，支撑大规模计算	用于新一代星脉网络，支撑万亿参数大模型训练	支撑 AI 未来：TH5 的性能使得在万卡 GPU 集群上进行高效协同训练成为可能，是训练千亿乃至万亿参数大模型的基础设施。



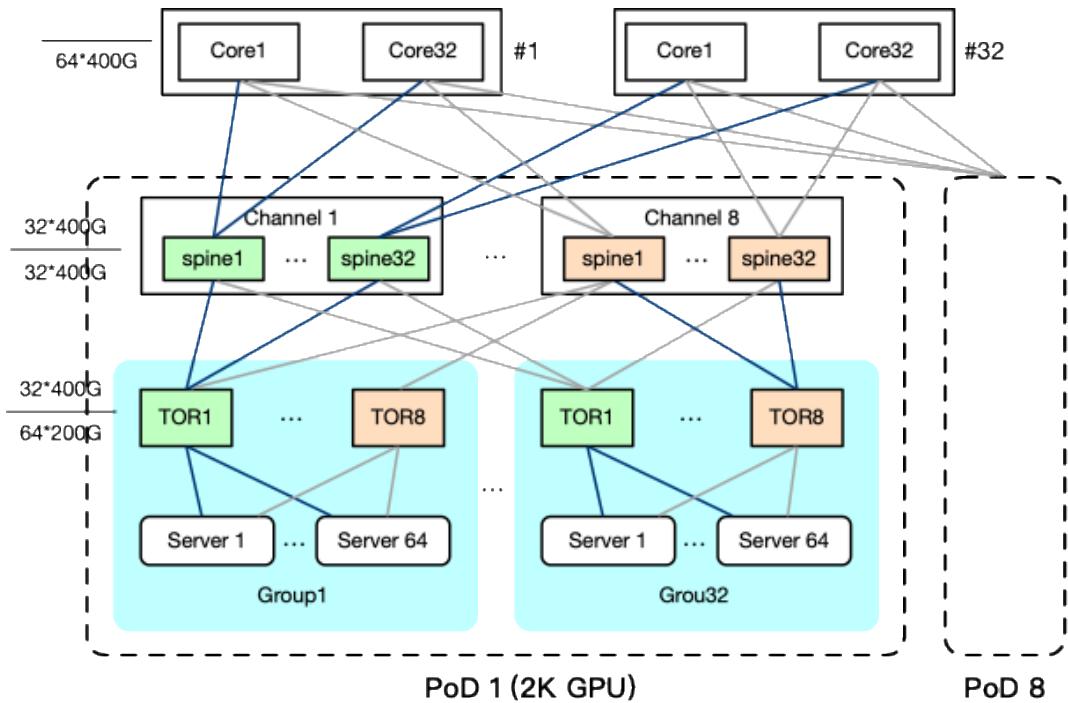
腾讯：TCCL 调度编排、端网协同防拥塞

- Tencent Intelligent Traffic Aware protocol (TiTa)
实现端网协同，基于 telemetry 来调整端侧发包速率和调整交换节点上分流策略
- Tencent Collective Communication Library (TCC L)，基于资源拓扑的资源调度和 Rank 编排



字节：三层盒式交换机，无收敛 Clos 组网

- 基于 NV GPU + Broadcom 芯片构建，多轨接入，三层无收敛胖树架构
- L1：接入 64 个 200G 光口（64 个 A100，使用 NV CX6，1*200G 光口）
- 定制 AOC cable 连接上行 32 个 400G 光口，分别接入 32 个 L2

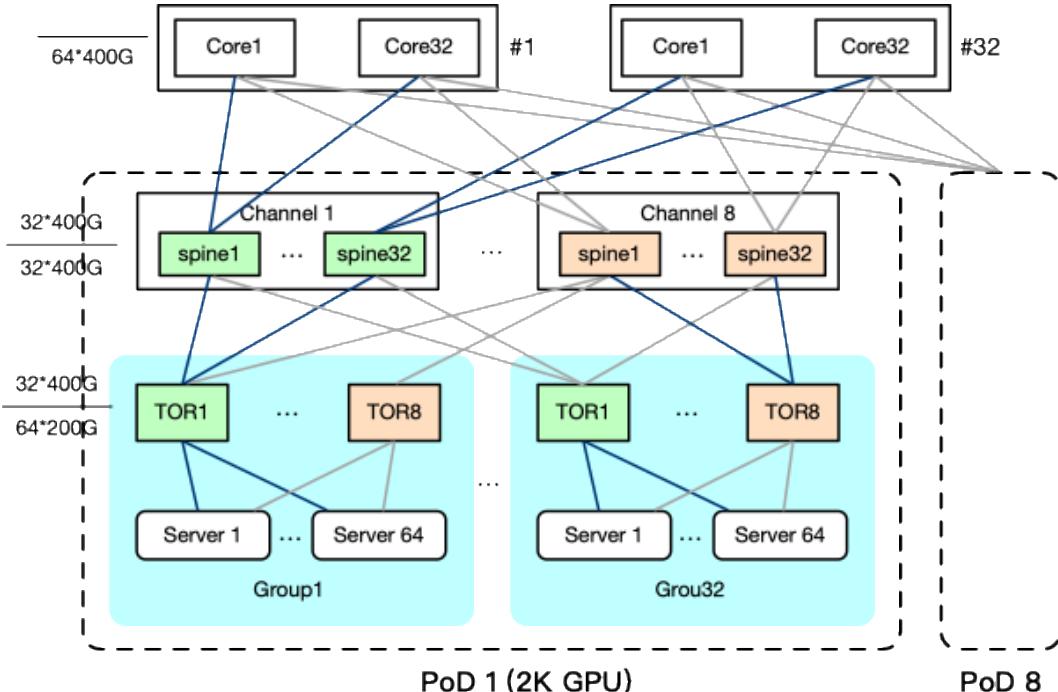


- 每 server: 8 GPU, 每卡 200G 接入
- 每 group: 64 server (512)
- 每 Pod: 32 group (16k)
- Cluster: 8 Pod (128k)
- TOR/Spine/Core: 25.6Tbps



字节：三层盒式交换机，无收敛 Clos 组网

- L2: 32 个 400G 接入 32 个 L1, 上行 32 个 400G 口分别接入 32 个 L3 节点
 - 方案 1: 每 L2 接入同轨 32 个 L1; 方案 2: 每 L2 接入 4 组完全轨道;
- L3: 64 个 400G 口接入 64 个 L2
 - 方案 1: 接入 8 个轨道 (每轨道 8 个 L2); 方案 2: 接入 8 个 PoD, (每 PoD 8 个 L2, 实现 PoD 间互访)

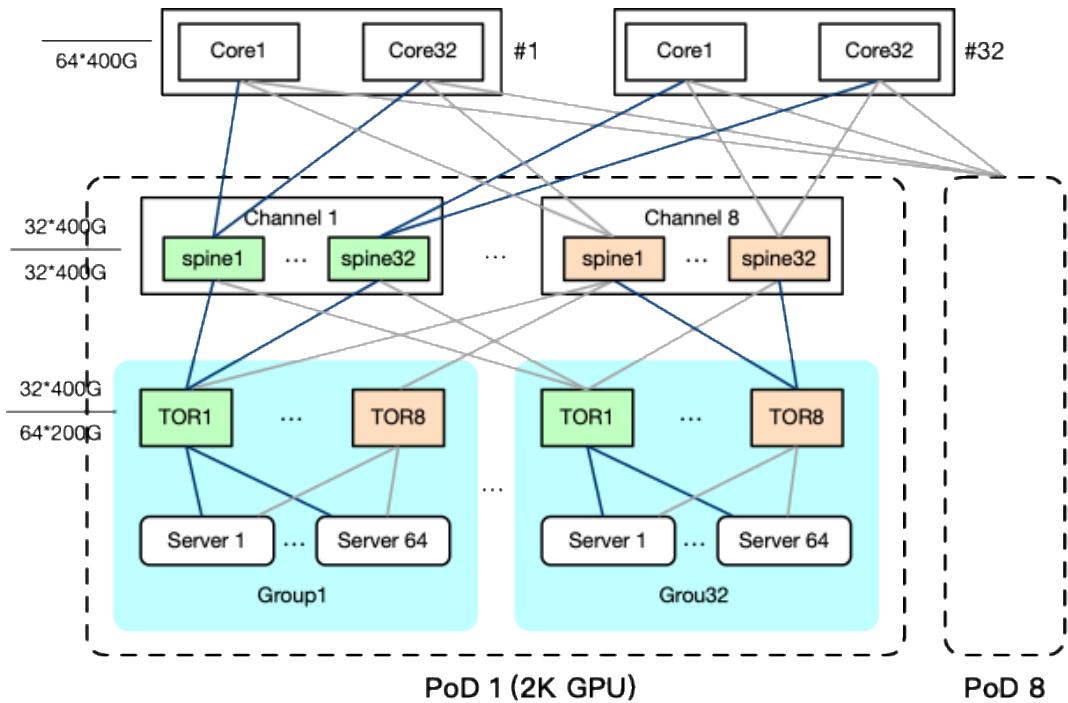


- 每 server: 8 GPU, 每卡 200G 接入
- 每 group: 64 server (512)
- 每 Pod: 32 group (16k)
- Cluster: 8 Pod (128k)
- TOR/Spine/Core: 25.6Tbps



字节：三层盒式交换机，无收敛 Clos 组网

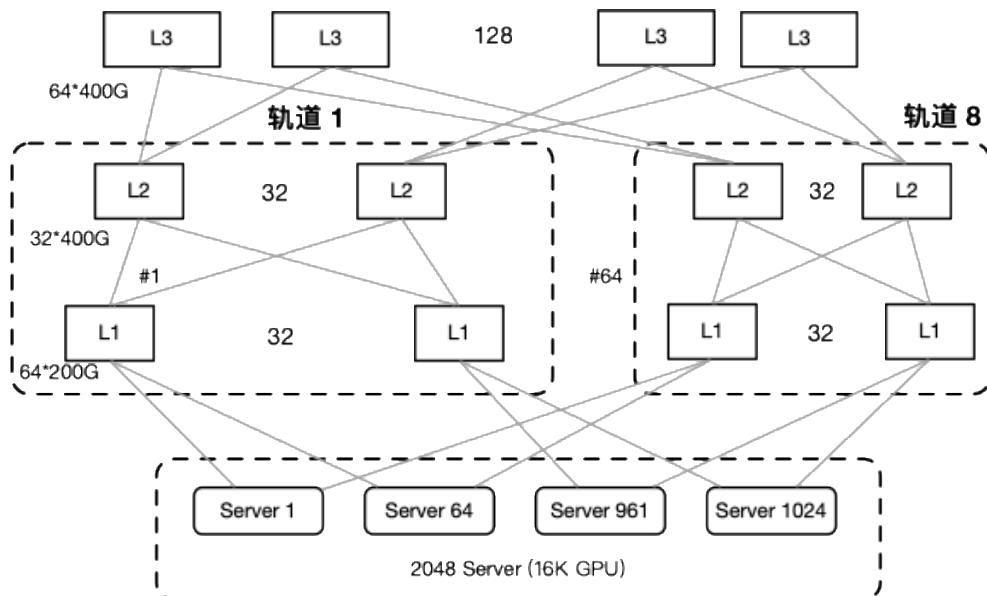
- 方案选择问题：
 - 不同层级、设备节点间，单链路连接可节省设备，双链路连接可增加可靠性，字节如何选择？
 - L2 实现并轨还是 L3 实现并轨，字节如何选择？



字节：三层盒式交换机，无收敛 Clos 组网

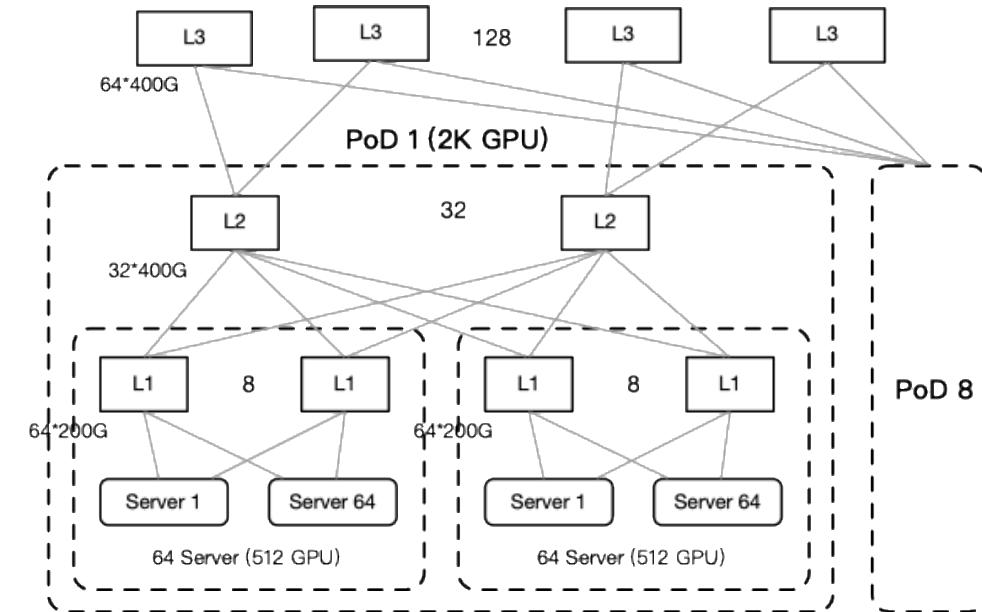
- 方案 1：

- 单轨无收敛，二层胖树接入 $64 \times 32 = 2K$ A100
- 8 轨道在 L3 节点上并轨互访，8 轨道 $2 \times 8 = 16K$



- 方案 2：

- 8 轨道在 L2 上并轨互访
- 单 PoD 完整接入 256 个 server (2K A100)
- 8 个 PoD 通过 L3 互联形成 16K 集群



字节：重并行策略优化、网络可靠性提升

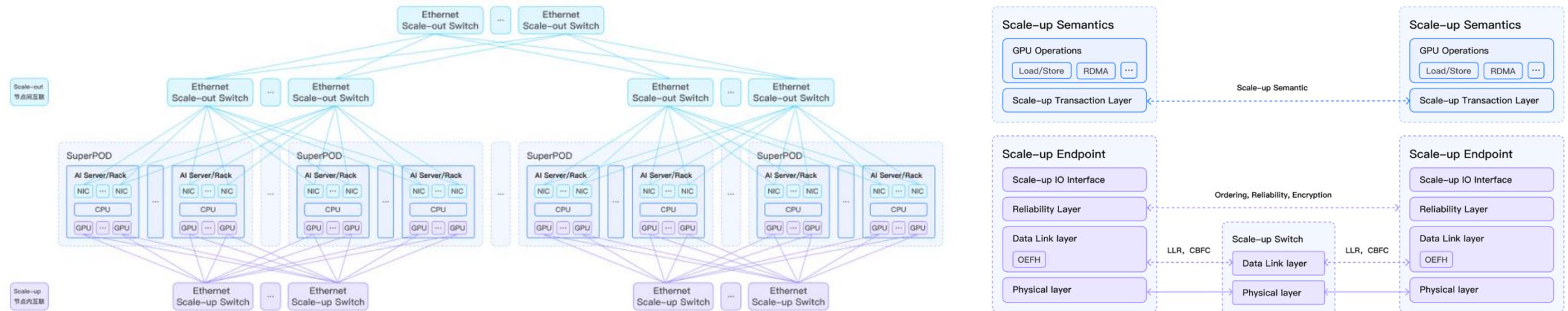
- 优化端到端 CC：结合 Swift 和 DCQCN，精确 RTT 测量和快速响应拥塞 ECN
- 传输层重传优化设计超时时间，对性能差的节点能将任务动态调整
- 基于拓扑优化并行策略 (TP/CP/PP/DP)
- 优化 CKPT IO，减少存放时慢链路等待，减少拉取数据时多访问争抢
- 建立容错机制和故障检测机制，实现实时监测系统状态 (GPU状态、网络性能、训练进度)



字节发布 EthLink

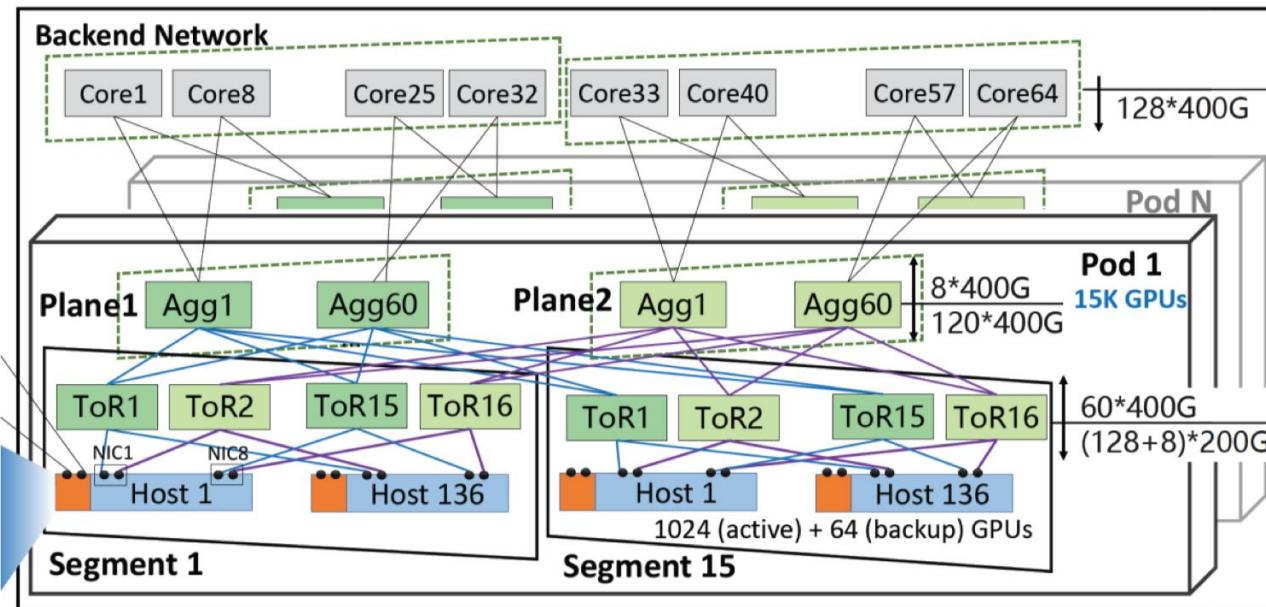
<https://mp.weixin.qq.com/s/aTM5B85wCU6cIbA0lhdpkg>

- 以高带宽、低延迟、强容错为核心，通过自研交换机（B5020）、EthLink 协议及动态优化算法，实现万卡级高效训练。技术亮点：
 - 硬件协同设计：Tomahawk 5芯片 + LPO 光模块降低功耗
 - 协议创新：EthLink 支持双语义，优化 RDMA 效率
 - 全局调度：任务感知的流量均衡与故障自愈



阿里：三层 Clos 组网，Spine 层带宽收敛

- 三层盒式交换机 Clos 组网，每卡 2*200G，16 轨接入，实现双平面扩大 segment
- Spine 层：15:1 收敛比
- Pod 规模：单 PoD 卡数 15k，96.3% 训练任务
- 可在单 Pod 内完成，跨 PoD 任务采用 PP 并行，Spine 层带宽可 15:1 收敛

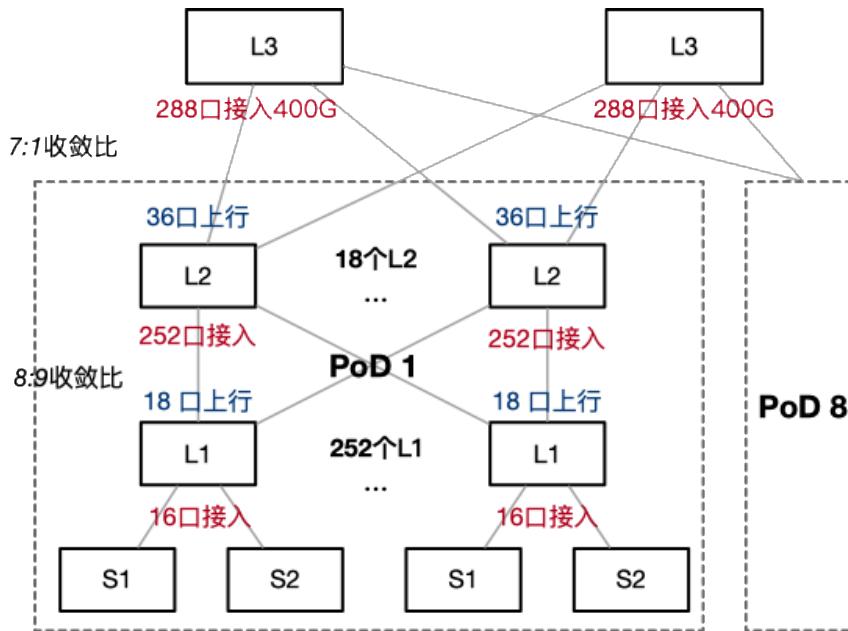
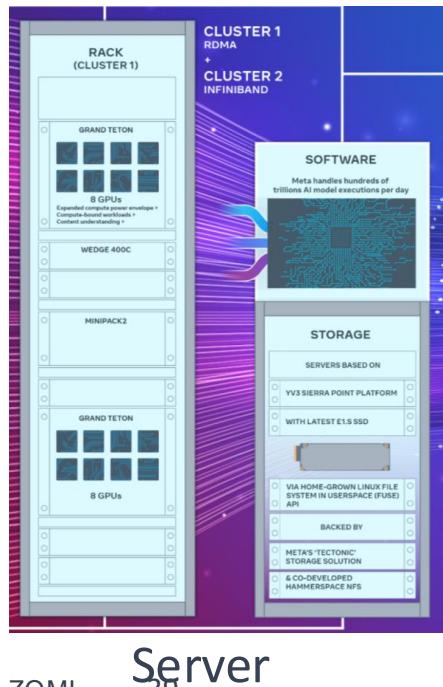


- 每 server: 8 GPU, 每卡 2*200G, 16 轨接入
- 每 segment: 128 server + 8 server (1024 + 64 备)
- 每 Pod: 15 Segment (15k + 960 备用)
- Cluster: 8 Pod (120k + 7.5K)
- 最大组网: 128 Pod (1920k + 120k)
- TOR/Agg/Core 盒式交换机: 51.2Tbps



Meta：三层 Clos 组网，Spine 层带宽收敛

- L1 盒式 L2 框式 Clos 组网 每卡 1*400G，单轨接入，Spine 层 7:1 收敛，跨 PoD 训练任务采用 DP 并行
- TOR 盒式交换机 (Minipack2) : 25.6Tbps, Cluster 框式交换机 (7800R3) : 460Tbps, 最大 576 *400G 光口



- 每 server: 8 GPU, 每卡 1*400G
- 每 Pod: $192 \times 16 = 3k$ 卡
- Cluster: 支持 8 Pod (24k)

总结与思考



AI 集群网络方案

- 国内互联网头部厂商采用 JDM (Joint Design Manufacturer) + CM (Contract Manufacture) 模式自研盒式交换机，并以二层胖树拓扑为主、通过多轨组网减少流量冲突，组网是否收敛按场景划分。

模式	JDM + CM 普及，Mate、字节、百度都自研交换机
拓扑	字节/腾讯：8 轨，阿里：16 轨逻辑双平面
收敛比	计算网络无收敛，存储和带内带外控制按需收敛
交换机	字节 B5020 (800G) 、腾讯 TCS9500 (400G) 、阿里白虎
组网规模	单 Pod \leq 16K GPU，集群物理端口 $<$ 100K



AI 集群网络方案

- 分轨/分平面/层次化收敛来提高 AI 集群规模（Scale Out），集中控制网络/优化拥塞控制/算网协同（通算掩盖）/自适应带宽（软件优化）提升网络能力。

模式	JDM + CM 普及，Mate、字节、百度都自研交换机
拓扑	字节/腾讯：8 轨，阿里：16 轨逻辑双平面
收敛比	计算网络无收敛，存储和带内带外控制按需收敛
交换机	字节 B5020 (800G) 、腾讯 TCS9500 (400G) 、阿里白虎
组网规模	单 Pod \leq 16K GPU，集群物理端口 $<$ 100K





Thank you

把 AIInfra 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/AIInfra

Book infrasys-ai.github.io



ZOMI

43

引用与参考

1. <https://zhuanlan.zhihu.com/p/687147416>
2. <https://www.53ai.com/news/LargeLanguageModel/2024081114950.html>
3. <https://cloud.tencent.com/developer/article/2403455>
4. <https://mp.weixin.qq.com/s/aTM5B85wCU6cIbA0lhdpkg>
5. <https://cloud.baidu.com/article/364290>
6. <https://zhuanlan.zhihu.com/p/640449350>

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

