

数字经济“算力网络”系列 8

推荐（维持）

AI 拉动多少需求，光模块弹性空间量化测算

不同组网架构中将使用不同的网卡、交换机和单元数量，GPU 与光模块的对应数量关系也将不同。经测算，2023 和 2024 年 AI 预计将为光模块带来 13.8/49.7 亿美金的 AI 增量市场空间。

□ 网卡型号、交换机型号和单元数量是影响光模块实际用量的主要因素。市场中流传着不同版本的光模块与 GPU 数量比例的测算，各种版本的数字对不上主要原因是因为不同组网架构下所需要使用光模块的用量都不尽相同，光模块的实际使用数量主要取决于组网架构中使用的网卡型号、交换机型号和单元数量。网卡主要包含 ConnectX-6 和 ConnectX-7；交换机主要包含 QM9700 系列和 QM8700 系列；单元数量影响交换架构层级，数量少时仅用两层架构，数量多时采用三层架构：H100 SuperPOD 最大支持 4 个单元组成集群，使用两层交换架构，A100 SuperPOD 最大支持 7 个单元组成集群，超过 5 个单元，需要三层交换架构。

□ 不同组网架构所需的光模块用量各不相同。在 A100+ConnectX6+QM8700 三层网络架构中，140 台服务器，共对应 140\*8=1120 片 A100，共需要 56+56+28=140 个 QM8790 交换机，需要 1120+1120+1120=3360 根线缆，需要 3360\*2=6720 个 200G 光模块，A100 与 200G 光模块对应关系为 1120/6720=1:6。在 A100+ConnectX6+QM9700 二层网络架构中，140 台服务器，1120 片 A100，共需要 12+9=21 台交换机，及 560+280=840 个 800G 光模块及 1120 个 200G 光模块。A100 与 800G 光模块对应关系为 1120:840=1:0.75，A100 与 200G 光模块对应关系为 1:1。在 H100+ConnectX7+QM9700 两层网络架构中，共需 512+1024=1536 个 800G 光模块需求和 1024 个 400G 光模块需求，总共 4\*32\*8=1024 片 H100。GPU 与 800G 光模块的对应关系为 1:1.5，与 400G 光模块的对应关系为 1:1。在 H100+ConnectX8+QM9700 三层网络架构中，假设 H100 也升级到 800G 网卡后，对外接口将从 4 个 OSFP 接口提升至 8 个 OSFP 接口，每层之间都用 800G 相连接，整个网络架构与第一种方案类似，仅将 200G 光模块更换为 800G 光模块，GPU 与光模块的需求比例为 1:6。

□ 光模块增量市场空间广阔。假设 2023 年出货 30 万片 H100+90 万片 A100，总共将带来 315 万支 200G+30 万支 400G+78.75 万支 800G 需求。假设 2024 年出货 150 万片 H100+150 万片 A100，总共将带来 75 万支 200G+75 万支 400G+675 万支 800G 需求。以 2023 年 1 美元/GB 均价，2024 年 0.85 美元/GB 均价简单计算，AI 预计为光模块带来 13.8/49.7 亿美金的 AI 增量市场空间。

□ 投资建议：我们重点推荐 800G 技术领先，与英伟达 Infiniband 体系有直接合作关系或有潜在合作可能的光模块公司，重点推荐中际旭创、新易盛、天孚通信，建议关注华工科技（电子组覆盖）、剑桥科技、联特科技、光迅科技、德科立、通宇通讯；同时建议关注产业链中优质公司，光芯片环节建议关注源杰科技、仕佳光子、光库科技，光器件环节建议关注太辰光、中瓷电子、博创科技、腾景科技。

□ 风险提示：AI 应用发展不及预期、参数假设与实际情况偏离、模型假设过于简化。

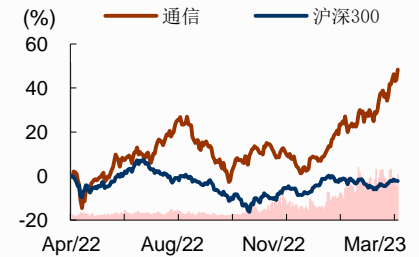
TMT 及中小盘/通信

行业规模

|          |       | 占比% |
|----------|-------|-----|
| 股票家数（只）  | 180   | 3.6 |
| 总市值（亿元）  | 30282 | 3.5 |
| 流通市值（亿元） | 21422 | 2.9 |

行业指数

| %    | 1m   | 6m   | 12m  |
|------|------|------|------|
| 绝对表现 | 15.4 | 51.3 | 46.9 |
| 相对表现 | 12.1 | 43.0 | 48.9 |



资料来源：公司数据、招商证券

相关报告

- 《数字经济“算力网络”系列 7：一温控液冷，AIGC 按下加速键》2023-04-11
- 《运营商系列报告 5：解析运营商产业链，把握数字经济全景图》2023-03-13
- 《数字经济系列报告 3：数字中国规划推出，或驱动光模块技术升级与需求爆发》2023-03-02

梁程加 S1090522060001  
 liangchengjia@cmschina.com.cn  
 刘浩天 研究助理  
 liuhaotian@cmschina.com.cn

**重点公司主要财务指标**

|      | 股价    | 21EPS | 22EPS | 23EPS | 22PE | 23PE | PB  | 评级   |
|------|-------|-------|-------|-------|------|------|-----|------|
| 中际旭创 | 68.57 | 1.09  | 1.54  | 1.84  | 44.5 | 37.3 | 4.6 | 强烈推荐 |
| 天孚通信 | 62.49 | 0.78  | 1.01  | 1.34  | 61.9 | 46.7 | 9.9 | 强烈推荐 |

资料来源：公司数据、招商证券

## 正文目录

|  |    |
|--|----|
| 一、光模块实际用量的影响因素 .....                     | 4  |
| 二、不同组网架构所需的光模块用量 .....                   | 6  |
| 1、A100+ConnectX6+QM8700 三层网络 .....       | 6  |
| 2、A100+ConnectX6+QM9700 二层网络 .....       | 7  |
| 3、H100+ConnectX7+QM9700 两层网络 .....       | 8  |
| 4、H100+ConnectX8（尚未发布）+QM9700 三层网络 ..... | 10 |
| 三、光模块与 GPU 数量比例的测算结论 .....               | 10 |

## 图表目录

|  |    |
|--|----|
| 图 1 NVIDIA ConnectX-6 智能网卡 .....       | 4  |
| 图 2 NVIDIA ConnectX-7 智能网卡 .....       | 4  |
| 图 3 NVIDIA Quantum-2 QM9700 系列 .....   | 5  |
| 图 4 NVIDIA Quantum QM8700 系列 .....     | 5  |
| 图 5 交换架构示意图 .....                      | 6  |
| 图 6 NVIDIA A100 共有 8 个计算接口 .....       | 6  |
| 图 7 第一层网络架构示意图 .....                   | 7  |
| 图 8 不同单元数量对应的各层交换机数与线缆连接数 .....        | 7  |
| 图 9 1 拖 4 的 QSFP 转 OSFP 接口 .....       | 8  |
| 图 10 NVIDIA H100 合并成 4 个 800G 接口 ..... | 8  |
| 图 11 第一层网络架构示意图 .....                  | 9  |
| 图 12 第二层网络架构示意图 .....                  | 9  |
| 图 13 不同单元数量对应的各层交换机数与线缆连接数 .....       | 10 |
| 图 12: 通信行业历史 PEBand .....              | 12 |
| 图 13: 通信行业历史 PBBand .....              | 12 |
| 表 1: 不同网络架构对应的光模块需求 .....              | 10 |
| 表 2: 预计 2023 和 2024 年光模块需求量 .....      | 11 |

## 一、光模块实际用量的影响因素

网卡型号、交换机型号和单元（Scalable unit SU）数量是影响光模块实际用量的主要因素。市场中流传着不同版本的光模块与 GPU 数量比例的测算，各种版本的数字对不上主要原因是因为不同组网架构下所需要使用光模块的用量都不尽相同，光模块的实际使用数量主要取决于组网架构中使用的网卡型号、交换机型号和单元数量。

### ■ 网卡型号

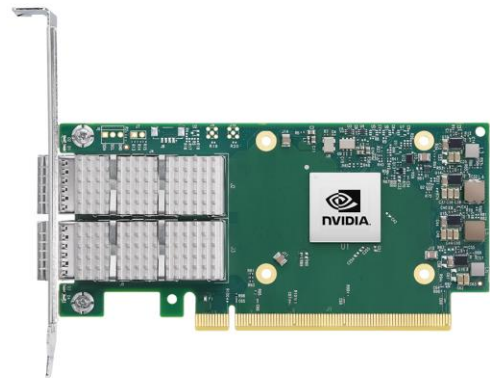
主要包括两种网卡，即 ConnectX-6（200Gb/s，主要配合 A100 使用）和 ConnectX-7（400Gb/s，主要搭配 H100 使用）。ConnectX-6 支持两个 200Gb/s 以太网连接端口，低于 800 纳秒的延迟和每秒 2.15 亿条消息，为要求最苛刻的应用程序和市场提供更高性能和更灵活的解决方案。ConnectX-6 是对 ConnectX 系列行业领先的网卡的开创性补充，提供多项增强功能，可进一步提高性能和可扩展性，并增加了对 200/100/50/40/25/10/1GbE 以太网速度和 PCIe Gen4 的支持。此外，ConnectX-6 以太网卡可以连接多达 32 通道的 PCIe，实现 200Gb/s 的带宽，即使在 3.0 代 PCIe 系统上也是如此。

图 1 NVIDIA ConnectX-6 智能网卡

#### 双端口 100GbE/单端口 200GbE SmartNIC

ConnectX-6 Dx SmartNIC 是业内非常安全、先进的云网络接口卡，用于加速任务关键型数据中心应用程序，如安全、虚拟化、SDN/NFV、大数据、机器学习和存储。SmartNIC 提供最多两个 100Gb/s 端口或一个 200Gb/s 以太网连接的单端口，并提供任何智能网络接口卡中最高的投资回报（ROI）。ConnectX-6 Dx 是 NVIDIA 享誉世界、屡获殊荣的 ConnectX 系列网络网卡的成员，该系列网络网卡由领先的 50Gb/s（PAM4）和 25/10Gb/s（NRZ）SerDes 技术和新功能提供支持，可加速云和数据中心有效负载。

[了解详情](#)



资料来源：NVIDIA 官网、招商证券

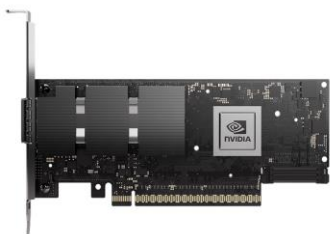
NVIDIA ConnectX-7 智能网卡提供多达 4 个连接端口和 400Gb/s 的吞吐量，可为云、电信、AI 和企业工作负载提供数据中心规模的硬件加速网络、存储、安全和管理服务。ConnectX-7 具有加速交换和数据包处理（ASAP2）、高级 RoCE、GPUDirect Storage，以及用于 TLS、IPsec 和 MACsec 加密与解密的 in-line 硬件加速等功能，能够为敏捷、高性能的网络解决方案提供助力。ConnectX-7 让组织能够在高带宽和高密度环境中满足目前和未来的网络需求。

图 2 NVIDIA ConnectX-7 智能网卡

#### ConnectX-7

来自 NVIDIA Quantum-2 InfiniBand 架构的 ConnectX-7 智能网卡（HCA）可提供超高的网络性能，用于处理极具挑战性的工作负载。ConnectX-7 支持超低时延、400Gb/s 吞吐量和创新的 NVIDIA 网络计算加速引擎，实现额外加速，为超级计算机、人工智能和超大规模云数据中心提供所需的高可扩展性和功能丰富的技术。

[了解详情](#)



资料来源：NVIDIA 官网、招商证券

### ■ 交换机型号

主要包括两种交换机，即 QM9700 系列（32 口 OSFP（2\*400Gb/s），共 64 路 400Gb/s 传输速率，51.2Tb/s 吞吐速率）和 QM8700 系列（40 口 QSFP56，共 40 路 200Gb/s 传输速率，16Tb/s 吞吐速率）。基于 NVIDIA quantum-2 的 QM9700 和 QM9790 交换系统在 1U 标准机箱设计中提供了前所未有的 64 个端口 400Gb/s InfiniBand。单台交换机的总双向吞吐量为 51.2Tb/s，具有超过 665 亿包/秒（BPPS）的容量标志。支持最新的 NVIDIA 高速互连 400Gb/s 技术，NVIDIA Quantum-2 带来了高速、极低延迟和可扩展的解决方案，融合了最先进的技术，如远程直接内存访问（RDMA）、自适

应路由和 NVIDIA 可扩展分层聚合和减少协议(SHARP)™。

图 3 NVIDIA Quantum-2 QM9700 系列



资料来源：NVIDIA 官网、招商证券

NVIDIA Quantum QM8700 和 QM8790 200Gb/s InfiniBand 固定配置交换机提供多达 40 个 200Gb/s 端口和 16 Tb/s 无阻塞带宽，或 80 个 100Gb/s 端口，每个端口具有全双向带宽和极低延迟，是构建任何规模的灵活和可扩展数据中心的理想选择。QM8700 和 QM8790 固定配置交换机旨在通过 NVIDIA 可扩展分级聚合和减少协议(SHARP)™技术使 NVIDIA 网络内计算卸载引擎成为可能。通过使用嵌入式硅交换硬件，该交换体系结构可以使用所有活动数据中心设备来加速通信框架，从而实现数量级的性能改进。

图 4 NVIDIA Quantum QM8700 系列



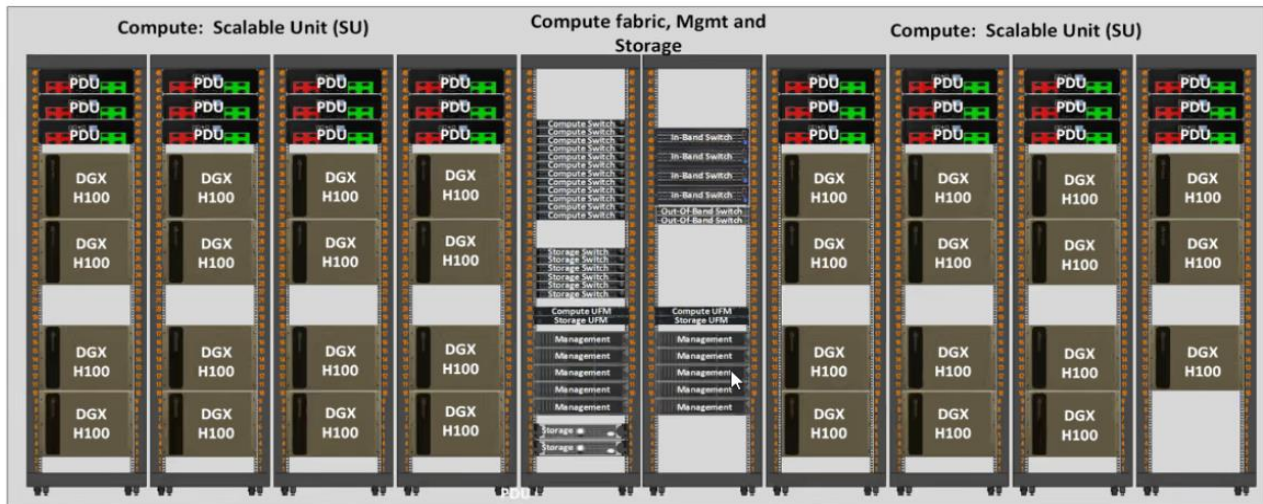
资料来源：NVIDIA 官网、招商证券

## ■ 单元数量

单元数量影响交换架构层级，数量少时仅用两层架构，数量多时采用三层架构。H100 SuperPOD 每个单元包括 32 个节点（DGX H100 服务器），最大支持 4 个单元组成集群，使用两层交换架构。A100 SuperPOD 每个单元包括 20 个节点（DGX A100 服务器），最大支持 7 个单元组成集群，超过 5 个单元，需要三层交换架构。

DGX H100 是 NVIDIA DGX 系统的最新迭代，也是 NVIDIA DGX SuperPOD 的基础，它是由 NVIDIA H100 Tensor Core GPU 突破创新加速的 AI 动力之源。共搭载 8 个 NVIDIA H100 GPU，总 GPU 显存高达 640GB，每个 GPU 配备 18 个 NVIDIA NVLink，GPU 之间的双向带宽高达 900GB/s。NVIDIA DGX A100 是适用于各种 AI 工作负载的通用系统，能够为率先推出的 5 petaFLOPS AI 系统提供之前难以实现的计算密度、性能和灵活性。DGX A100 采用 NVIDIA A100 Tensor Core GPU，使企业能够将训练、推理和分析整合到一个易于部署的统一 AI 基础架构中，该基础架构具备直接联系 NVIDIA AI 专家的功能。

图 5 交换架构示意图



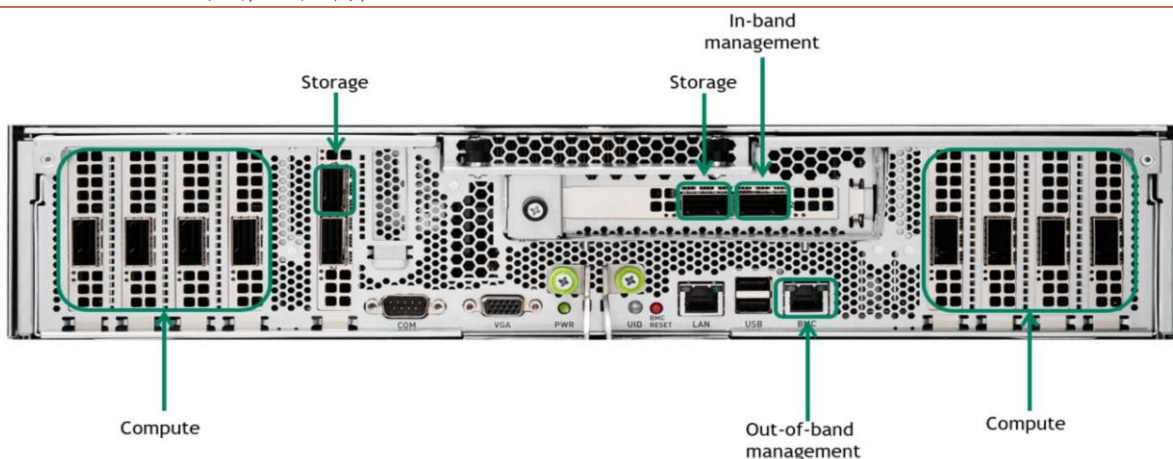
资料来源：NVIDIA 官网、招商证券

## 二、不同组网架构所需的光模块用量

### 1、A100+ConnectX6+QM8700 三层网络

A100 共有 8 个计算接口，左右各四个，目前 A100 出货主要搭配 ConnectX6 对外通讯，接口速率 200Gb/s。NVIDIA A100 Tensor Core GPU 可针对 AI、数据分析和 HPC 应用场景，在不同规模下实现出色的加速，有效助力更高性能的弹性数据中心。A100 采用 NVIDIA Ampere 架构，是 NVIDIA 数据中心平台的引擎。A100 的性能比上一代产品提升高达 20 倍，并可划分为七个 GPU 实例，以根据变化的需求进行动态调整。A100 提供 40GB 和 80GB 显存两种版本，A100 80GB 将 GPU 显存增加了一倍，并提供超快速的显存带宽（每秒超过 2 万亿字节 [TB/s]），可处理超大型模型和数据集。

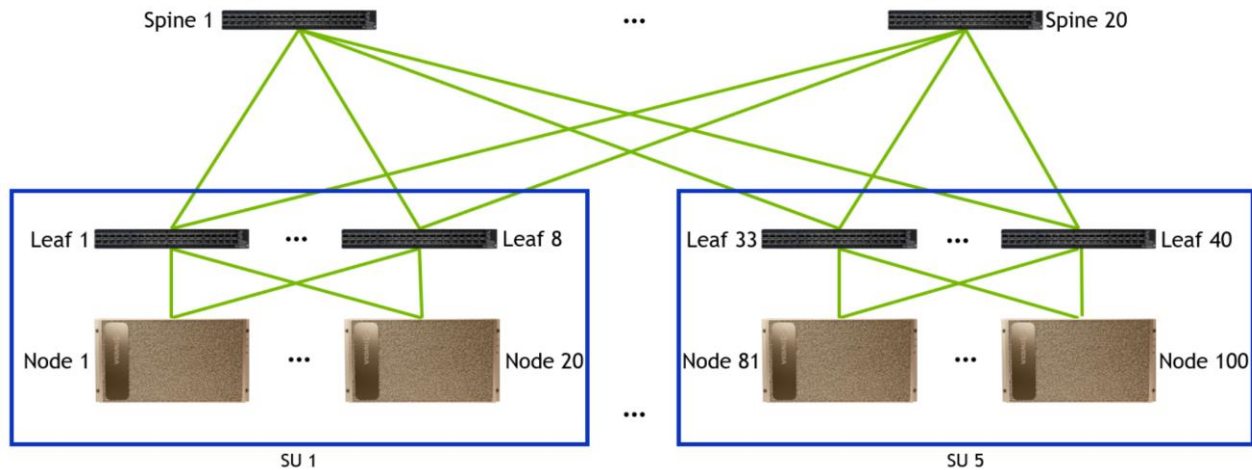
图 6 NVIDIA A100 共有 8 个计算接口



资料来源：NVIDIA 官网、招商证券

在第一层架构中，每个节点（Node）有 8 个接口（Port），每个节点分别连 8 个叶交换机（Leaf），每 20 个节点组成一个单元（SU），因此在第一层总共需要  $8 \times \text{SU}$  个叶交换机，需要  $8 \times \text{SU} \times 20$  条线缆（Cable），需要  $2 \times 8 \times \text{SU} \times 20$  个 200G 光模块。

图 7 第一层网络架构示意图



资料来源：NVIDIA 官网、招商证券

在第二层架构中，由于采用无阻塞架构，上行速率=下行速率，在第一层中总的单向传输速率为 200G\*线缆数量，由于第二层也采用单条线缆 200G 的传输速率，第二层级中线缆数量应与第一层级相同，需要  $8 * \text{SU} * 20$  条线缆 (Cable)，需要  $2 * 8 * \text{SU} * 20$  个 200G 光模块。所需要的主交换机 (Spine) 数量是线缆数量除以叶交换机数量，即需要  $(8 * \text{SU} * 20) / (8 * \text{SU})$  个主交换机。但当 leaf 交换机数量不够多的时，为节约主交换机数量，可在叶和脊之间做两条以上的连接（不能超过 40 个接口的限制）。当单元数量分别为 1/2/4/5 个时所需要的主交换机数量为 4/10/20/20 个，所需要的光模块数量分别为 320/640/1280/1600 个，脊交换机数量不会同比例提升，但光模块数量会同比例提升。

当单元数量达到 7 个时，需要用到第三层架构，由于无阻塞架构，第三层架构所需要的线缆数与第二层数相同。英伟达推荐配置 SuperPOD：英伟达推荐使用 7 个单元进行组网，需要增加第三层架构，并增加核心交换机 (Core)。

图 8 不同单元数量对应的各层交换机数与线缆连接数

| Nodes              | SUs | QM8790 Switches |                    |                   | Cables |       |      |
|--------------------|-----|-----------------|--------------------|-------------------|--------|-------|------|
|                    |     | Leaf            | Spine <sup>1</sup> | Core <sup>1</sup> | Leaf   | Spine | Core |
| 10                 | ½   | 8               | 2                  |                   | 80     | 80    |      |
| 20 (Single SU)     | 1   | 8               | 4                  |                   | 160    | 160   |      |
| 40                 | 2   | 16              | 10                 |                   | 320    | 320   |      |
| 80                 | 4   | 32              | 20                 |                   | 640    | 640   |      |
| 100                | 5   | 40              | 20                 |                   | 800    | 800   |      |
| 120                | 6   | 48              | 80                 | 24                | 960    | 960   | 960  |
| 140 (DGX SuperPOD) | 7   | 56              | 80                 | 28                | 1120   | 1120  | 1120 |

1. To avoid possible performance issues, ports on Spine and Core switches must only be used for inter-switch cabling.

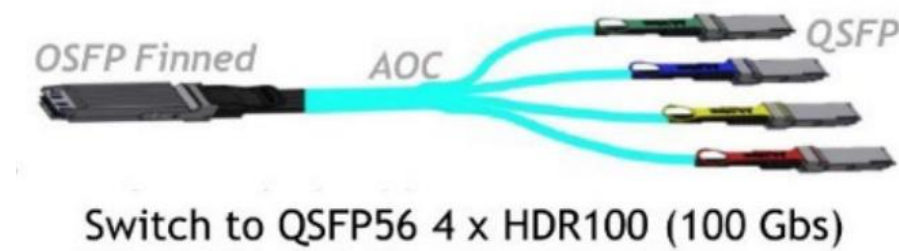
资料来源：NVIDIA 官网、招商证券

140 台服务器，共对应  $140 * 8 = 1120$  片 A100，共需要  $56 + 56 + 28 = 140$  个交换机 (QM8790)，需要  $1120 + 1120 + 1120 = 3360$  根线缆，需要  $3360 * 2 = 6720$  个 200G 光模块，A100 与 200G 光模块对应关系为  $1120/6720 = 1:6$ 。

## 2、A100+ConnectX6+QM9700 二层网络

目前的推荐配置中还没有此方案，但未来可能会有越来越多 A100+QM9700 的组网架构，这将降低光模块使用量，带动 800G 光模块需求。与第一个方案最大的区别，在于第一层连接由对外接 8 根 200G 线缆转换为接两根 1 拖 4 的 QSFP 转 OSFP 接口。

图 9 1 拖 4 的 QSFP 转 OSFP 接口



资料来源：NVIDIA 官网、招商证券

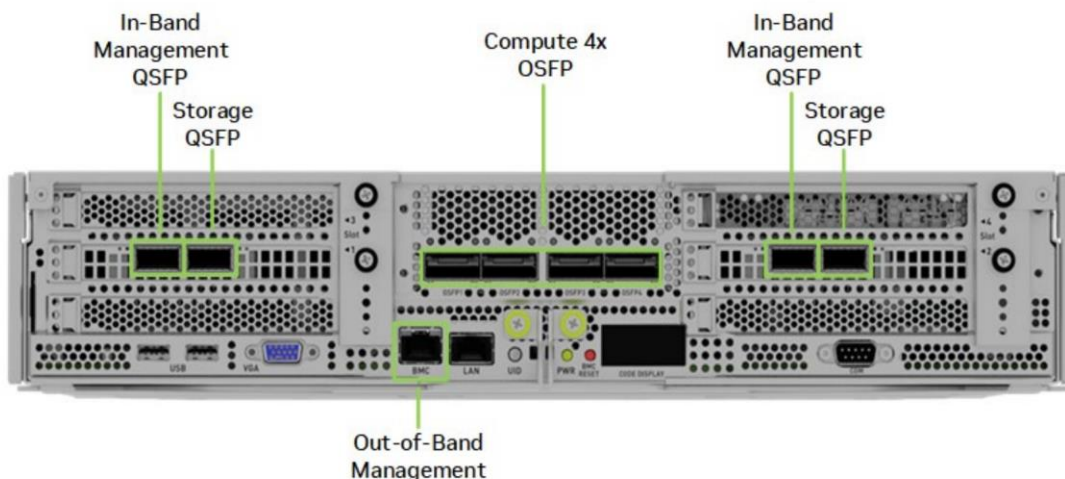
在第一层架构中，对于一个 7 单元的集群，140 台服务器有  $140 \times 8 = 1120$  个接口，共对外接出  $1120 / 4 = 280$  根 1 拖 4 的线，带来 280 个 800G 和 1120 个 200G 光模块需求，共需 12 台 QM9700 交换机。在第二层架构中，仅用 800G 连接，需要  $280 \times 2 = 560$  个 800G 光模块，9 台 QM9700 交换机。

140 台服务器，1120 片 A100，共需要  $12 + 9 = 21$  台交换机，及  $560 + 280 = 840$  个 800G 光模块及 1120 个 200G 光模块。  
A100 与 800G 光模块对应关系为 1120:840=1:0.75，A100 与 200G 光模块对应关系为 1:1。

### 3、H100+ConnectX7+QM9700 两层网络

H100 设计比较特殊，虽然 8 个 GPU 配了 8 张 400G 网卡，但接口合并成 4 个 800G 接口，这将带来大量 800G 光模块需求。H100 配备第四代 Tensor Core 和 Transformer 引擎（FP8 精度），与上一代产品相比，可为多专家(MoE)模型提供高 9 倍的训练速度。通过结合可提供 900GB/s GPU 间互连的第四代 NVlink、可跨节点加速每个 GPU 通信的 NVLINK Switch 系统、PCIe5.0 以及 NVIDIA Magnum IO 软件，为小型企业到大规模统一 GPU 集群提供高效的可扩展性。在数据中心级部署 H100GPU 可提供出色的性能，并使所有研究人员均能轻松使用新一代百亿亿次级(Exascale)高性能计算(HPC)和万亿参数的 AI。

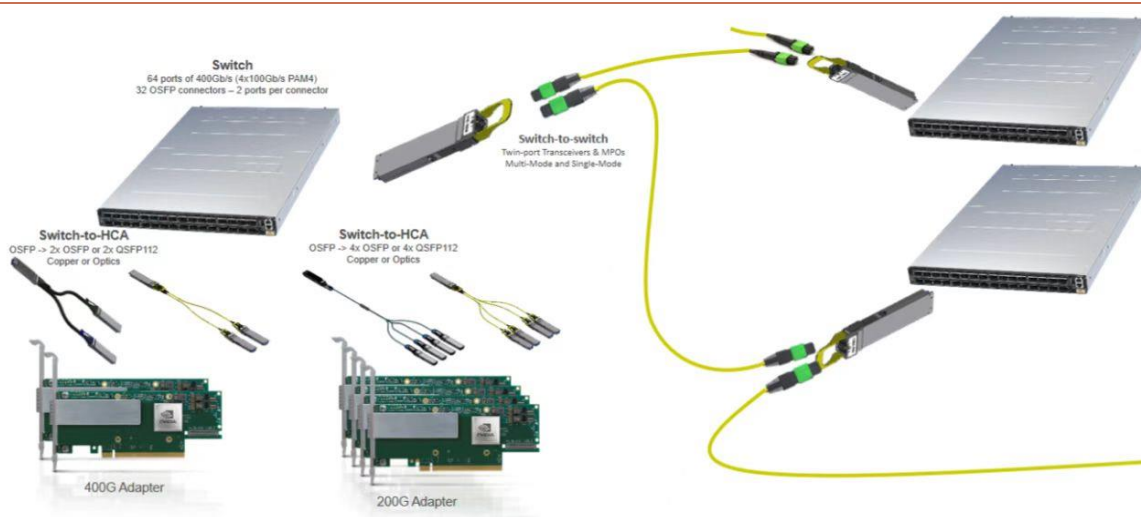
图 10 NVIDIA H100 合并成 4 个 800G 接口



资料来源：NVIDIA 官网、招商证券

在第一层，根据英伟达推荐配置，建议在服务器接口上接 1 个 2\*400G 的 800G 光模块，通过 twin-port 连接两根光缆（MPO），分别插在两个交换机上。

图 11 第一层网络架构示意图

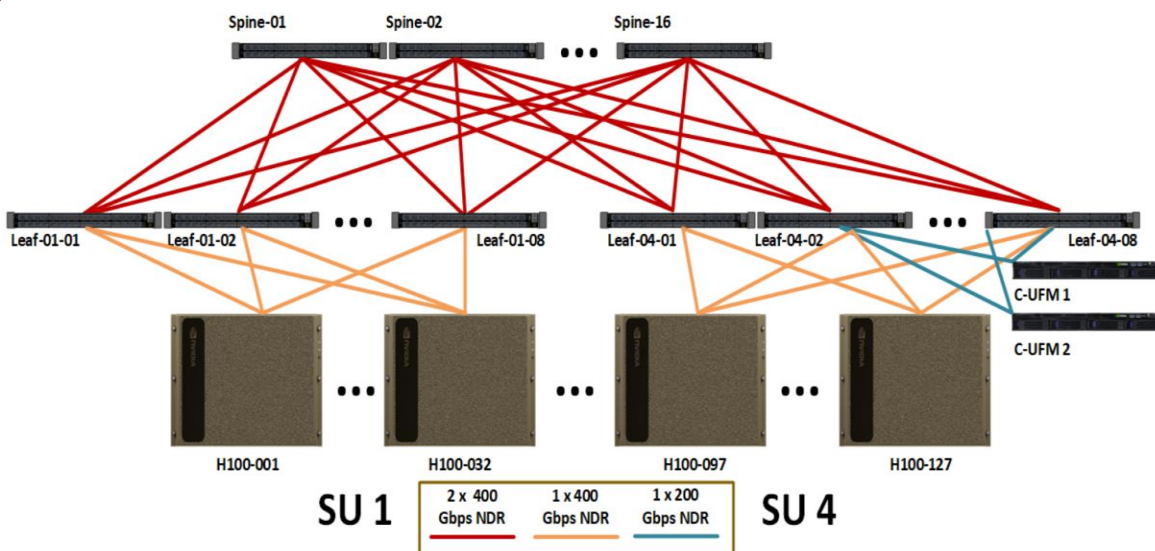


资料来源：NVIDIA 官网、招商证券

第一层一个单元包含 32 台服务器，一台服务器连接  $2 \times 4 = 8$  台交换机，SuperPOD 中包含 4 个单元，共需  $4 \times 8 = 32$  台叶交换机在第一层连接。英伟达建议留出一个节点做管理用途（UFM），由于对光模块的使用量影响有限，我们姑且按照 4 个单元共 128 台服务器简略计算。则在第一层共需要  $4 \times 128 = 512$  个 800G 光模块，和  $2 \times 4 \times 128 = 1024$  个 400G 光模块。

在第二层，交换机之间直接用 800G 光模块连接，一台叶交换机向下连接了  $32 \times 400G$  的单向速率。为保证上行下行速率一致，向上连接需要  $16 \times 800G$  的单向速率，需要 16 台脊交换机，共有  $4 \times 8 \times 16 \times 2 = 1024$  个 800G 光模块需求。

图 12 第二层网络架构示意图



资料来源：NVIDIA 官网、招商证券

在此架构下，两层相加共需  $512 + 1024 = 1536$  个 800G 光模块需求和 1024 个 400G 光模块需求，总共  $4 \times 32 \times 8 = 1024$  片 H100。GPU 与 800G 光模块的对应关系为  $1024/1536 \rightarrow 1:1.5$ ，与 400G 光模块的对应关系为  $1024/1024 \rightarrow 1:1$ 。

图 13 不同单元数量对应的各层交换机数与线缆连接数

| SU Count | Cluster Size # Nodes | Cluster Size # GPUs | Leaf Switch Count | Spine Switch Count | Compute + UFM Node Cable Count | Spine-Leaf Cable Count |
|----------|----------------------|---------------------|-------------------|--------------------|--------------------------------|------------------------|
| 1        | 31 <sup>1</sup>      | 248                 | 8                 | 4                  | 252                            | 256                    |
| 2        | 63                   | 504                 | 16                | 8                  | 508                            | 512                    |
| 3        | 95                   | 760                 | 24                | 16                 | 764                            | 768                    |
| 4        | 127                  | 1016                | 32                | 16                 | 1020                           | 1024                   |

1. This is a 32 node per SU design, however a DGX Node must be removed to accommodate for UFM connectivity.

资料来源：NVIDIA 官网、招商证券

#### 4、H100+ConnectX8（尚未发布）+QM9700 三层网络

此方案尚未发布，但我们假设 H100 也升级到 800G 网卡后，对外接口将从 4 个 OSFP 接口提升至 8 个 OSFP 接口，每层之间都用 800G 相连接，整个网络架构与第一种方案类似，仅将 200G 光模块更换为 800G 光模块。因此在此种架构下 GPU 与光模块的需求比例也为 1:6。

### 三、光模块与 GPU 数量比例的测算结论

对于 A100+ConnectX6+QM8700 三层网络架构，140 台服务器共对应  $140 \times 8 = 1120$  片 A100，共需要  $56 + 56 + 28 = 140$  个 QM8790 交换机，需要  $1120 + 1120 + 1120 = 3360$  根线缆，需要  $3360 \times 2 = 6720$  个 200G 光模块，A100 与 200G 光模块对应关系为  $1120/6720 = 1:6$ 。

对于 A100+ConnectX6+QM9700 二层网络架构，140 台服务器对应 1120 片 A100，共需要  $12 + 9 = 21$  台交换机，及  $560 + 280 = 840$  个 800G 光模块及 1120 个 200G 光模块。A100 与 800G 光模块对应关系为  $1120:840 = 1:0.75$ ，A100 与 200G 光模块对应关系为 1:1。

对于 H100+ConnectX7+QM9700 两层网络架构，两层相加共需  $512 + 1024 = 1536$  个 800G 光模块需求和 1024 个 400G 光模块需求，总共  $4 \times 32 \times 8 = 1024$  片 H100。GPU 与 800G 光模块的对应关系为  $1024/1536 \rightarrow 1:1.5$ ，与 400G 光模块的对应关系为  $1024/1024 \rightarrow 1:1$ 。

对于 H100+ConnectX8+QM9700 三层网络架构，假设 H100 也升级到 800G 网卡后，对外接口将从 4 个 OSFP 接口提升至 8 个 OSFP 接口。每层之间都用 800G 相连接，整个网络架构与第一种方案类似，仅将 200G 光模块更换为 800G 光模块，GPU 与光模块的需求比例为 1:6。

表 1：不同网络架构对应的光模块需求

| GPU  | 网卡速率 | 交换机速率 | 架构层级 | 200G 光模块需求量 | 400G 光模块需求量 | 800G 光模块需求量 |
|------|------|-------|------|-------------|-------------|-------------|
| A100 | 200G | 200G  | 三层   | 1:6         | 0           | 0           |
| A100 | 200G | 400G  | 两层   | 1:1         | 0           | 1:0.75      |
| H100 | 400G | 400G  | 两层   | 0           | 1:1         | 1:1.5       |
| H100 | 800G | 800G  | 三层   | 0           | 0           | 1:6         |

资料来源：招商证券

我们假设 2023 年出货 30 万片 H100+90 万片 A100, 总共将带来 315 万支 200G+30 万支 400G+78.75 万支 800G 需求。假设 2024 年出货 150 万片 H100+150 万片 A100, 总共将带来 75 万支 200G+75 万支 400G+675 万支 800G 需求。以 2023 年 1 美元/GB 均价, 2024 年 0.85 美元/GB 均价简单计算, AI 预计为光模块带来 13.8/49.7 亿美金的 AI 增量市场空间。

表 2: 预计 2023 和 2024 年光模块需求量

|      | A100<br>(万片) | H100<br>(万片) | 200G 交换<br>机比例 | 400G 交换<br>机比例 | 800G 交换<br>机比例 | 200G 光模块需<br>求 (万支) | 400G 光模块需<br>求 (万支) | 800G 光模块需<br>求 (万支) |
|------|--------------|--------------|----------------|----------------|----------------|---------------------|---------------------|---------------------|
| 2023 | 90           | 30           | 37.5%*         | 62.5%          | 0%             | 315                 | 30                  | 78.75               |
| 2024 | 150          | 150          | 0%             | 75%**          | 25%            | 75                  | 75                  | 675                 |

资料来源: 招商证券

\*A100 中一半用 200G 交换机一半用 400G 交换机;

\*\*H100 中一半用 400G 交换机一半用 800G 交换机;

\*\*\*以上对 A100 H100 数量的判断仅是我们的假设, 不代表对未来的预期

## 四、风险提示

### 1) AI 应用发展不及预期

ChatGPT 仍处于推广早期阶段, 处于商业化早期阶段, 尚未形成商业化闭环。由于其所需算力成本高昂, 可能导致其商业化程度不及预期, 推广程度低于我们的预期。若 ChatGPT 等 AI 应用未能如我们所期望的那样普及开来, 那么对算力基础设施的需求可能低于我们的预期。

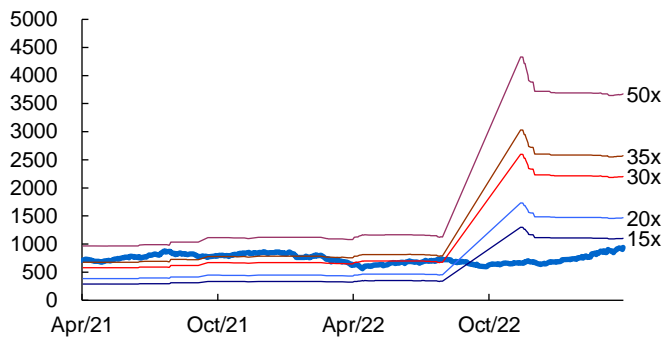
### 2) 参数假设与实际情况偏离

我们对光模块的需求测算基于一系列假设, 用户对网卡速率的选择, 对网卡搭配数量的选择, 交换机型号的选择等等因实际情况不同用户的选择差异较大, 都会对光模块的用量产生影响。若用户实际选择和我们的假设存在较大出入则可能导致光模块实际需求低于我们的预期。

### 3) 模型假设过于简化

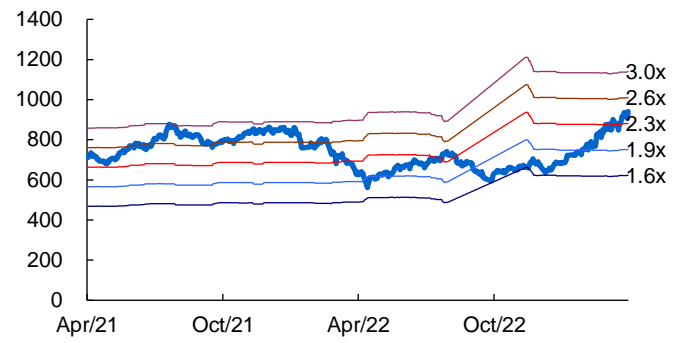
在测算行业市场空间时我们对很多实际情况进行简化处理。我们的测算是基于一般情景下的简要测算, 实际情景远比我们所假设的模型更为复杂, 可能导致我们模型计算的结果与实际情况出现偏差。

图 14: 通信行业历史 PEBand



资料来源：公司数据、招商证券

图 15: 通信行业历史 PBBand



资料来源：公司数据、招商证券

## 分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告中的具体推荐或观点直接或间接相关。

梁程加 招商证券通信行业联席首席分析师 北京邮电大学电路与系统硕士，曾就职于中国移动、长江证券、中信证券、红杉资本，2022 年加入招商证券。

刘浩天 招商证券通信行业研究助理 中央财经大学金融硕士，曾就职于安信证券，2022 年加入招商证券。

## 评级说明

报告中所涉及的投资评级采用相对评级体系，基于报告发布日后 6-12 个月内公司股价（或行业指数）相对同期当地市场基准指数的市场表现预期。其中，A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准；美国市场以标普 500 指数为基准。具体标准如下：

### 股票评级

强烈推荐：预期公司股价涨幅超越基准指数 20%以上

增持：预期公司股价涨幅超越基准指数 5-20%之间

中性：预期公司股价变动幅度相对基准指数介于±5%之间

减持：预期公司股价表现弱于基准指数 5%以上

### 行业评级

推荐：行业基本面向好，预期行业指数超越基准指数

中性：行业基本面稳定，预期行业指数跟随基准指数

回避：行业基本面转弱，预期行业指数弱于基准指数

## 重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对任何人的投资建议。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人均不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。