中科星图技术交流

Security Level: confidential

三大算力需求场景：预训练，微调，日常运营

训练	预训练 Pretrain	GPT-3	GPT-2	GPT-1	···	
	微调 Finetune	SFT监督学习	PPO强化学习	RM奖励模型	迁移学习	···
推理	日常运营	推理计算	参数加载	信息交互	···	

预训练（Pretrain）：预先训练的模型或预先训练模型的过程。训练模型基础语言能力，得到基础大模型。

- 预训练模型就是已经用数据集训练好了的模型。
- 现在我们常用的预训练模型就是他人用常用模型，比如VGG16/19，Resnet等模型，并用大型数据集来做训练集。
- 正常情况下，我们常用的VGG16/19等网络已经是他人调试好的优秀网络，我们无需再修改其网络结构。

微调（Finetune）：将预训练过的模型作用于自己的数据集，并使参数适应自己数据集的过程。进行监督学习、强化学习、迁移学习等二次或多次训练，实现对模型参数量的优化调整。

- 要使用的数据集和预训练模型的数据集相似，如都是自然景物图片。
- 自己搭建或者使用的CNN模型正确率太低。
- 数据集相似，但数据集数量太少。
- 计算资源太少。

日常运营：基于用户输入信息，加载模型参数进行推理计算，并实现最终结果的反馈输出。

预训练显存评估

模型训练的显存需求评估

模型权值显存占用（FP16）	2字节
模型梯度显存占用（FP16）	2字节
优化器参数显存占用（FP32）	3X4=12字节（以Adam为例）

模型训练显存占用=模型参数量*16字节+计算每层中间结果输出*BatchSize

模型	模型参数量（Billion）	显存占用（GB)	最低要求
Bert-Large	0.34	5.44 + 中间结果xBatchSise	A100单卡可完整训练
GPT-2	1.5	24+中间结果xBatchSize	A100单卡可完整训练
Megatron	8.7	139.2+中间结果xBatchSize	A100 8卡可完整训练
Turing-NLG	17	272+中间结果xBatchSize	A100 16卡可完整训练
GPT3	175	2,800 + 中间结果xBatchSize	A100 128（8x16）卡可完整训练

预训练算力评估

每次迭代算力：
$F=96BSlh^2(1+S/6h+V/16lh)$
B: batch size, S: 序列长度, l: transformer layers，h:隐层大小，V: 词表规模

ChatGPT为例：
模型参数175B
Transformer layers:96
序列长度：2048
隐层大小：12288
词表规模：51200
BatchSize:1536
单次迭代算力：$F=96BSlh^2(1+S/6h+V/16lh)=4.5$ EFLOPs
模型收敛所需迭代次数：约95000次
训练ChatGPT模型总算力：单次迭代算力*迭代次数=430 ZFLOPs

一个MFLOPS（megaFLOPS）等于每秒一百万（=10^6）次的浮点运算，

一个GFLOPS（gigaFLOPS）等于每秒十亿（=10^9）次的浮点运算，

一个TFLOPS（teraFLOPS）等于每秒一万亿（=10^12）次的浮点运算，(1太拉)

一个PFLOPS（petaFLOPS）等于每秒一千万亿（=10^15）次的浮点运算，

一个EFLOPS（exaFLOPS）等于每秒一百京（=10^18）次的浮点运算，

一个ZFLOPS（zettaFLOPS）等于每秒十万京（=10^21）次的浮点运算。

训练时间估算：
运行时间=总算力/gpu卡算力/gpu卡数量/计算效率
单张A800算力：312 TFLOPS (FP 16 Tesnor Core)
以1024个gpu卡为例计算（1 ZettaFlops =1024^3 TFlops）
运行时间=430*1024^3/312/1024/50%=2890305秒=2890305/24/3600=33.45天
说明:计算效率=gpu正在做矩阵运算的单元数/gpu所有单元数，AI集群中gpu计算效率大概为50%

微调—需根据具体场景确定

模型迭代带来微调Finetune训练算力需求。

ChatGPT模型并不是静态的，需要不断进行模型调优，以确保模型处于最佳应用状态。

- 需要开发者对模型参数进行调整，确保输出内容不是有害和失真的
- 需要基于用户反馈和PPO策略，对模型进行大规模或小规模的迭代训练
- 微调工作量与微调优化内容、算法优化经验、预期目标强相关
- 算力分配时，建议为微调预留10%~20%的算力

理解微调的意义

Step 1 :假设我们的神经网络符合下面形式：

Y = W * X

Step 2 :现在我们要找到一个W, 使得当 输入X = 2 时，输出Y = 1，也就是希望W = 0.5:

1 = W * 2

Step 3: 首先对W 要进行初始化， 初始化的值服从均值为0，方差为1的分布，假设W初始化为0.1：

Y = 0.1 * X

Step 4 :当 输入X = 2 时, W = 0.1, 输出Y = 0.2, 这个时候实际值和目标值1的误差是0.8 :

1<-----0.2 = 0.1 * 2

Step 5 :0.8 的误差经过反向传播去更新权值W，假如这次更新为W = 0.2，输出为0.4，与目标值的误差为0.6:

1<-----0.4 = 0.2 * 2

Step 6 : 可能经过十次或二十次反向传播，W 终于等于我们想要的0.5：

Y = 0.5 * X

Step 7 : 如果在更新模型最开始有人告诉你，W的值应该在0.47附近:

Y = 0.47 * X

Step 8 : 那么从最开始训练，你与目标值的误差就只有0.06了，那么可能只要一步两步*，就能将w训练到0.5：

1<--------0.94 = 0.47 * 2

总结: Step 7相当于提供一个预训练模型（Pre-trained model），Step 8 就是基于这个模型微调（Fine Tune）。

相对于从头开始训练(Training a model from scatch)，微调省去大量计算资源和计算时间，提高了计算效率,甚至提高准确率。

运营—关注显存满足度

ChatGPT的运营，重点关注GPU显存的匹配度。

需要至少5张 A800（80G显存）GPU卡 。

显存需求估算公式：

model size（占大部分显存） + activation + K/V cache + overhead.

=[model_parameters + b*s*h +b*s*v + 2* n*b*size_per_head*head_num*(s+s_o)]* 2Bytes + overhead

≈357.9 + 0.0053*batch_size*(input_len+output_len) (GB)

Model size：模型参数大小

activation:激活函数

K/V cache:关键字/数值cache

Overhead:基本开销，占比很低，可忽略不计

推理只要能把模型加载进去，就可以进行推理；显存容量主要取决于模型参数量，粗略估算方法：

推理场景显存≈参数量*2Bytes

ChatGPT为例：

显存容量 ≈ 175*10^9*2=350GB　5张A800 80GGPU卡满足推理要求

Batch_size	Input_len	Output_len	Tested Memory
1	128	8	350
16	128	8	371
368	128	8	649
16	128	256	389
150	128	256	644
1	512	128	362
16	512	128	415
80	512	128	643
1	512	512	363
16	512	512	443
48	512	512	613
1	2048	2048	380
12	2048	2048	620

01 AIGC算力需求评估

02 GPU卡技术参数

03 超聚变产品介绍

04 推荐组网方案

05 讨论

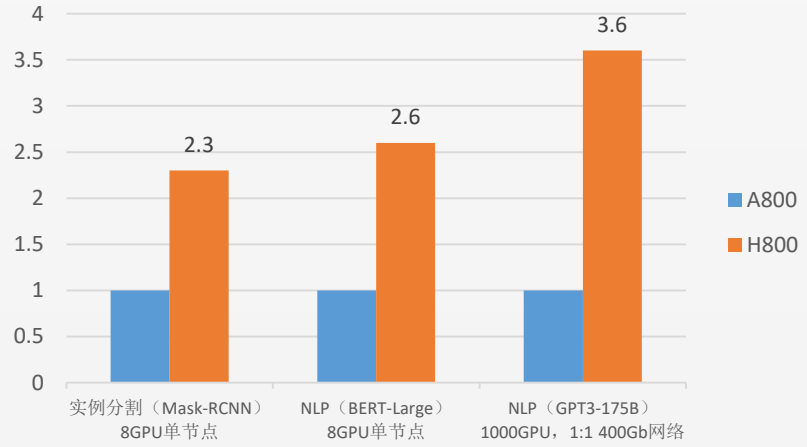
ChatGPT 推荐GPU型号规格表

缺省A800 PCIE参数

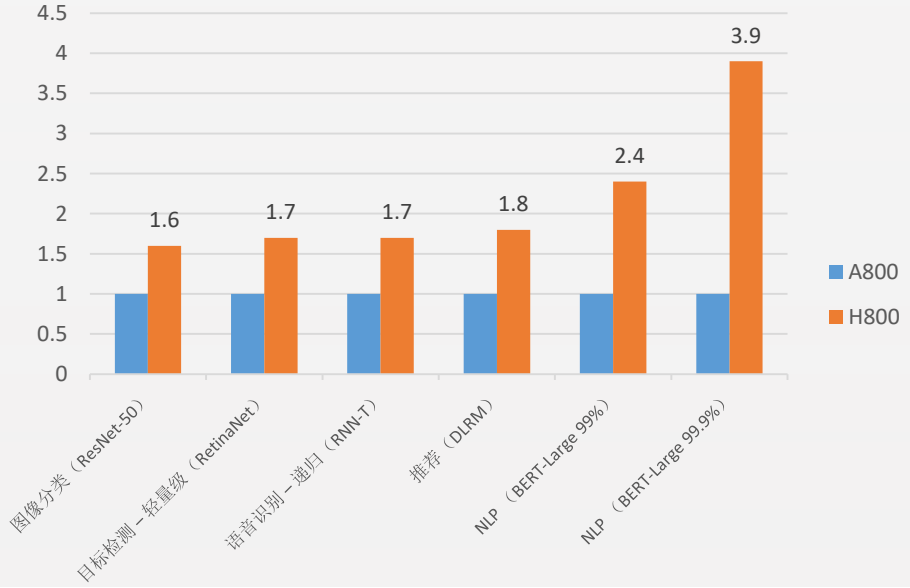
GPU	H100 SXM/H800 SXM	H100 PCIe/H800 PCIe	A100 80G PCIE	A100 80GB SXM	A800 SXM	H800 VS A800
架构	Hopper		Ampere			
FP32	67 TFLOPS	51 TFLOPS	19.5 TFLOPS	19.5 TFLOPS	19.5 TFLOPS	3.44↑
FP16 Tensor Core	1979 TFLOPS	1,513 TFLOPS	312 TFLOPS	624 TFLOPS	312 TFLOPS	6.34↑
INT8 Tensor Core	3958 TOPS	3026 TOPS	624 TFLOPS	1248 TFLOPS	624 TFLOPS	6.34↑
GPU 显存	80GB	80GB	80GB	80GB	80GB	1→
GPU 显存带宽	3.35TB/s	2TB/s	1935 GB/s	2039 GB/s	1935 GB/s	1.73↑
NVLINK速率	900GB/s / 400GB/s	**400GB/s**	600GB/s	600GB/s	**400GB/s**	1→
最大热设计功率	700W	300-350W	300W	400W	300W	

H800相对A800方案性价比更高，AI综合性能2.45倍，成本仅1.74倍

训练性能成倍提升。



推理性能成倍提升。



性价比提升40%

AI综合性能：2.45倍（H800性能/A800性能）。

指标项	H800/A800倍数
推理性能提升倍数	2.18
推理占比	59.50%
训练性能提升倍数	2.83
训练占比	40.50%
综合性能倍数	**2.45**

假定三个主流训练应用等比例分布，六个主流推理应用等比例分布，按2023年推理59.5%，训练40.5%加权，H100性能是A100的2.45倍。

物料成本：仅1.74倍（H800典配成本/A800典配成本）。

机型	G8600 V7	
GPU	H800 8GPU模组	A800 8GPU模组
CPU	2*8462Y	
内存	1TB（32*32GB）	
NVMe盘	8*3.2TB U.2NVMe	
SSD盘	2*480GB SATA SSD	
IB卡	4*HDR200	
机箱&机框	机箱+系统框+复合框+GPU框+Riser框+滑轨	
电源	6*54V PSU+2*12V PSU	

01 AIGC算力需求评估

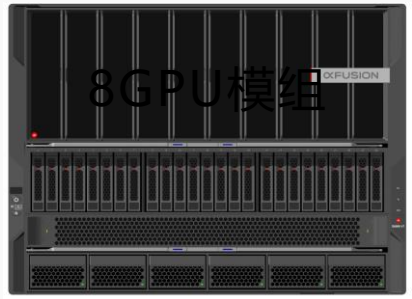
02 GPU卡技术参数

03 超聚变产品介绍

04 推荐组网方案

05 讨论

全系AI—V7 GPU服务器产品总览

基础型	存储型	主流型	旗舰级
AI推理	需大容量存储的 中规模AI训练&AI推理	AI训练&AI推理	大规模AI训练
			
2288H V7	G5200 V7	G5500 V7	G8600/G8600E V7
GPU:支持4个双宽GPU卡 2个Intel SPR, 单处理器最高350W 硬盘:45个2.5寸盘	GPU:支持4个双宽GPU卡或10个单宽GPU卡 CPU:1或2个Intel SPR, 单处理器最高350W 硬盘:32个3.5寸 SAS/SATA; 4个NVMe	GPU:支持10个双宽GPU卡或10个单宽GPU卡 CPU:2个Intel SPR, 单处理器最高350W 硬盘:24个3.5寸/2.5寸 SAS/SATA; 12个NVMe SSD	GPU：8GPU模组 CPU:2个Intel SPR, 单处理器最高350W 硬盘:25个2.5寸 SAS/SATA; 8个NVMe SSD

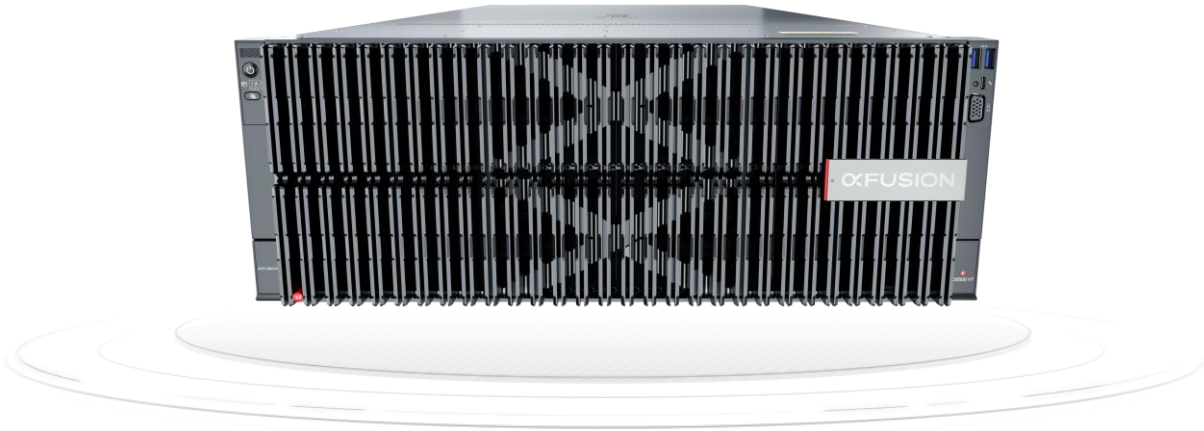
高效算力创新技术
自制电源：功率密度高33%，故障率低50%
自研风扇：散热风量提升50%，45度高温稳定运行
智能节能：同等配置和负载节能8%，SPEC Power第一
运维：远程在线运维，移动运维

智能算力创新技术
AI内存故障自愈：宕机率降低66%
AI硬盘故障预测：提前7-30天预测风险盘

主流GPU服务器产品G5500 V7，均衡灵活，人工智能首选

全新PCIe GPU服务器

图像处理&AI推理



极致性能

- 支持10块双宽GPU卡
- 支持CPU到GPU x32双倍通信带宽，满足CPU与GPU高带宽场景要求
- 支持2个英特尔®第四代至强®可扩展处理器，最高350W，120核通用算力
- 24块3.5寸盘或12个NVMe SSD灵活配置

极致灵活

- 一键切换拓扑：级联型和均衡型两种拓扑，灵活适用多种场景
- 简化链路：免PCIe retimer设计，降低了PCIe通路的延迟和系统功耗
- 缩短链路：支持GPUDirect Storage/RDMA/P2P，适配集群规模部署

极致可靠

- 宽工作温度：精湛散热设计，支持40℃工作温度*
- 电源冗余高效：4个3000W高效钛金级电源，支持N+N/N+M冗余
- 风扇冗余：6个8080+定制风扇，支持N+1冗余

智能管理

- FDM技术：故障诊断，定位准确率达93%
- 智能硬盘故障预测：提前7-30天预测故障
- DEMT 2.0 节能技术：最高节能18%
- 智能内存故障自愈技术：降低宕机率50%
- 五大智能技术，运维效率提升30%

FUSION

G5500 V7，4U空间支持10张双宽GPU卡，业界最高

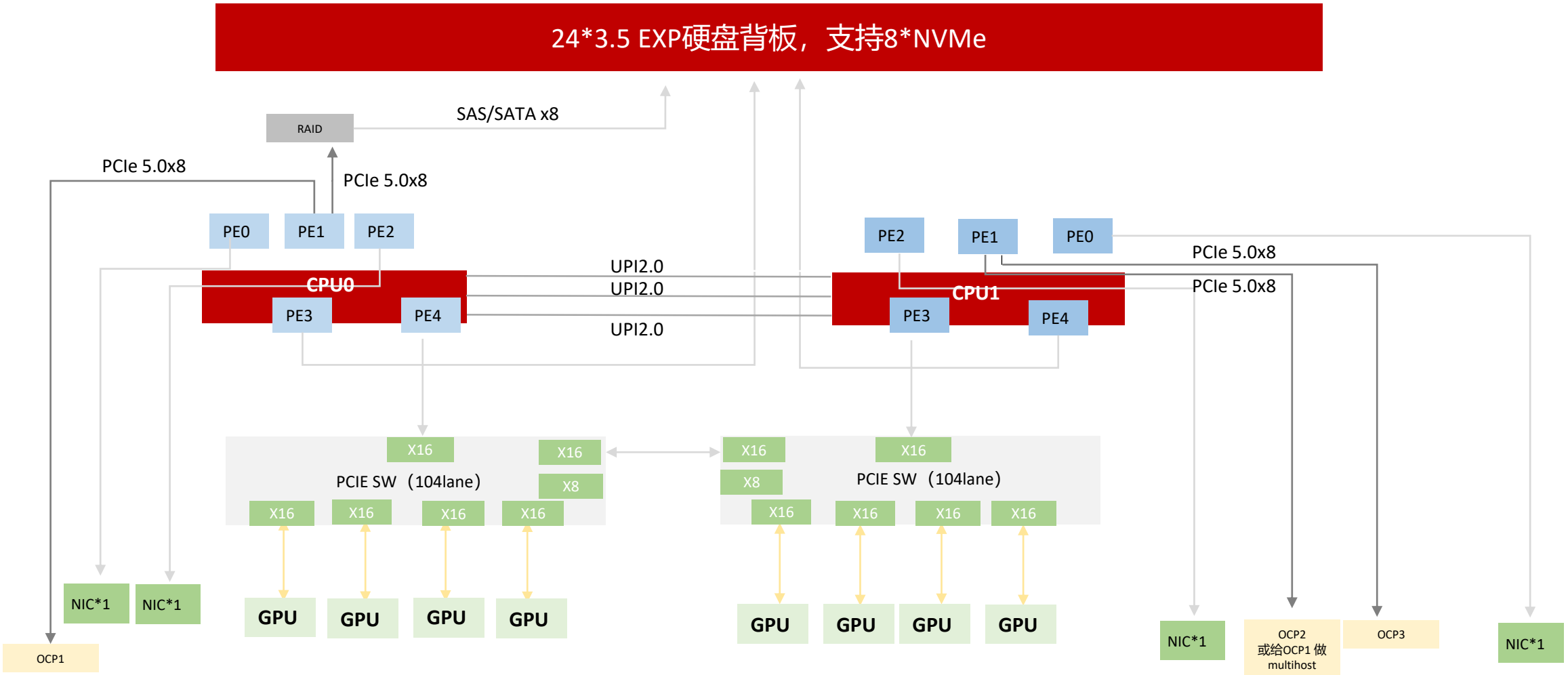
G5500 产品图

G5500 V7 前视图

G5500 V7 后视图

G5500	
应用场景	大规模AI训练/推理，高性能/均衡型两种机型
类别	4U2P GPU服务器
GPU	支持10个双宽GPU卡（推荐4/8个）
CPU	2个第四代英特尔®至强®可扩展处理器（Sapphire Rapids），支持单处理器最大350W功率
硬盘	支持多种不同的硬盘配置，硬盘支持热插拔 • 最多可配置24个3.5/2.5英寸SAS/SATA硬盘 • 最多可配置12个NVMe SSD盘 • 2个M.2 SSDs
RAID	可选配支持RAID0、1、10、1E、5、50、6、60等，支持Cache超级电容保护
IO	最多14个PCIe 5.0 X16标准槽位，最多支持10个GPU卡专用槽位，均支持PCIe5.0 X16 最多支持3*OCP3.0，支持PCIe 5.0 X8，其中一个OCP槽位支持Socket Direct
电源	可配置4个热插拔电源模块，支持N+N/N+M冗余
风扇	6个热插拔对旋风扇，支持N+1冗余
温度	5℃~35℃
尺寸	175mm×447 mm×898 mm（高x宽x深）

拓扑灵活，AI标准型拓扑，支持一键拓扑切换



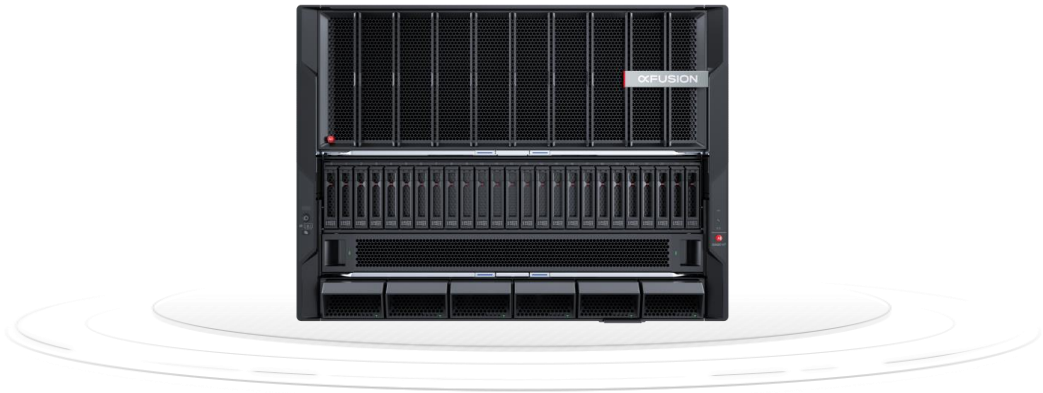
2288H，G5200，G5500支持的GPU卡及应用场景

厂商	形态	场景	型号
NVIDIA	PCIe标卡	计算：训练/HPC/数据分析	H100
NVIDIA	PCIe标卡		H800
NVIDIA	PCIe标卡		A100 80GB
NVIDIA	PCIe标卡		A800 80GB
NVIDIA	PCIe标卡-液冷		A100 80GB液冷
NVIDIA	PCIe标卡	计算：AI推理/中等HPC	A30
NVIDIA	PCIe标卡	计算&图形：边缘AI/小型推理/边缘视频处理/移动云游戏	A2
NVIDIA	PCIe标卡		T4
NVIDIA	PCIe标卡		L4
NVIDIA	PCIe标卡	图形：元宇宙/云渲染 最快光速追踪/最大渲染模型	A40
NVIDIA	PCIe标卡		L40
NVIDIA	PCIe标卡	图形：虚拟工作站/视频会议/4K云游戏	A10
NVIDIA	PCIe标卡	图形：虚拟桌面，虚拟工作站，编解码 最多的编解码视频流	A16
NVIDIA	PCIe标卡	工作站级别显卡，图形处理	RTX A6000
NVIDIA	PCIe标卡		RTX 6000 Ada
NVIDIA	PCIe标卡		RTX A4000
Intel	PCIe标卡	图形：Windows云游戏/虚拟桌面/安卓云游戏	M150
寒武纪	PCIe标卡	推理：高能效比云端AI推理，128T(INT8) 支持INT16/INT8/INT4/FP32/FP16	MLU270-S4-16GB
瀚博	PCIe标卡	推理：推理/200T(INT8)/视频编解码	VA1A PCIe-16GB
瀚博	PCIe标卡	推理/图形：通用AI推理/视频编解码	VA1 PCIe-16GB
瀚博	PCIe标卡	视频/图形：云端智能视频加速/视频编解码/支持图像检测/分类/分割/增强/超分	VA1V PCIe-16GB
海思	PCIe标卡	推理/图形：搜索推荐/内容审核/OCP系统	Atlas 300I Pro
壁仞	PCIe标卡	计算：训练/HPC/数据分析	BR104P

新一代人工智能旗舰G8600 V7

新一代NVLink GPU服务器

用于AI训练、HPC、大数据、关键计算、科学计算等高性能场景



旗舰性能

- 支持业界最高算力的8GPU模组
- 兼容NVIDIA的高性能和均衡两种拓扑
- 支持2个英特尔®第四代至强®可扩展处理器，最高350W，120核通用算力
- 支持32个4800MT/s DDR5内存插槽
- 支持25个2.5英寸硬盘和8个NVMe SSD盘，获得超大容量和超高速存储

极致能效

- 自研钛金电源能效高，单台最高节省160W
- GPU模块的54V与计算模块的12V电源双分区，少一道电源转换，单台最高节省76W
- 业界独创结合GPU的MPC算法，风扇功耗再降1.1%
- 超聚变风扇8080++，散热能力提升20-30%，能效提升16%

极致可靠

- 无需下架可更换模块，维护时间缩短2.5倍（模块如：GPU模块，计算模块，风扇，电源，网卡IO）
- 54V双总线供电，更少电源更高冗余
- GPU模块和计算模块风扇N+1冗余
- GPU模块的54V电源N+N冗余
- 计算模块的12V电源N+1冗余

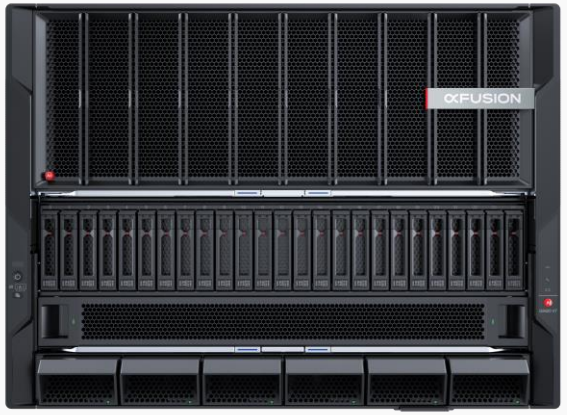
智能管理

- FDM技术：故障诊断，定位准确率达96%
- 智能硬盘故障预测：提前7-30天预测故障
- 智能节能技术：专利动态能效管理技术，整机相比业界最高节能8%
- 智能内存故障自愈技术：降低宕机率66%
- 五大智能技术，运维效率提升30%

G8600 V7，8GPU模组机型，超多IO，灵活存储方案

G8600 产品图

G8600 V7 前视图



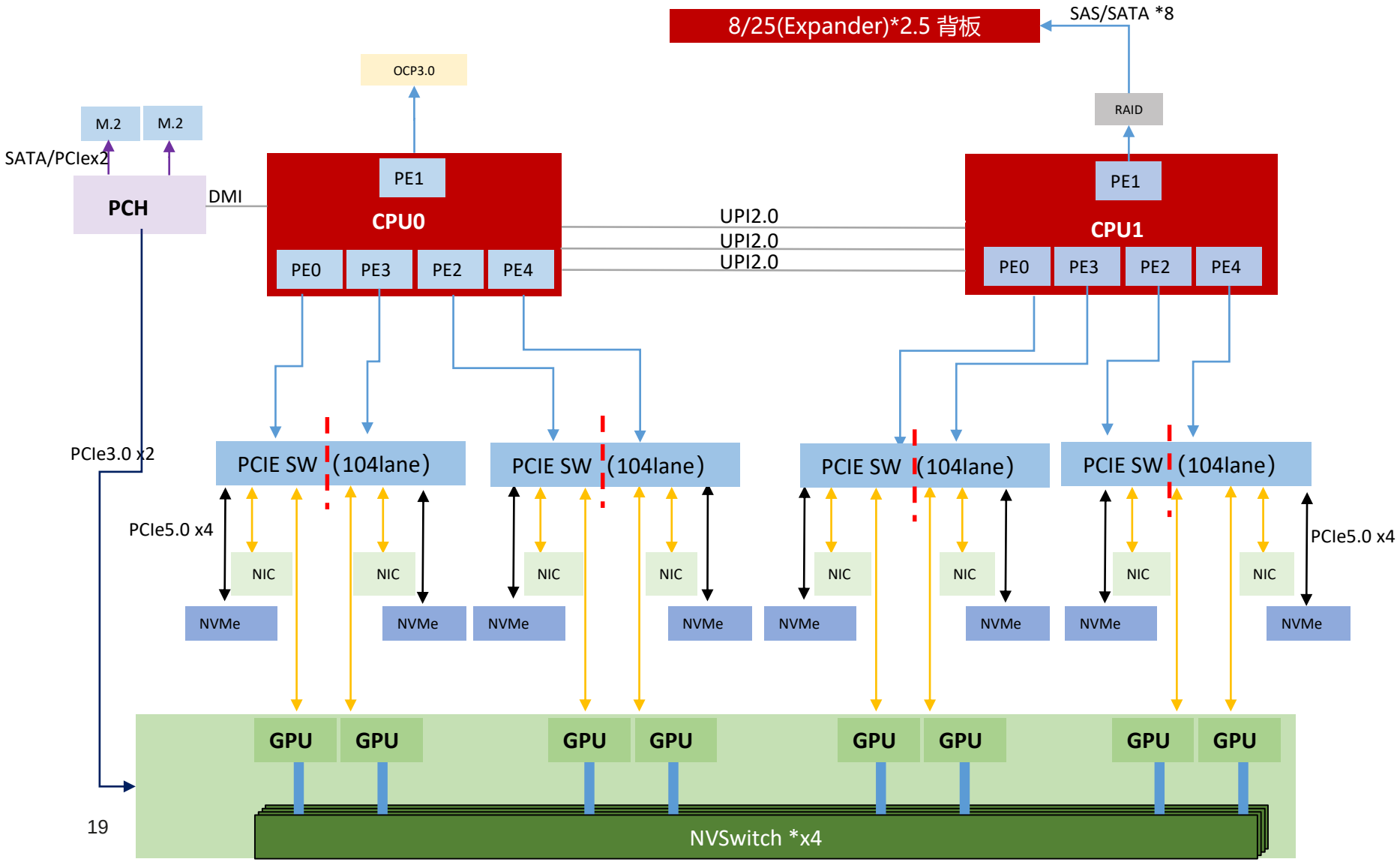
G8600 V7 后视图

G8600	
应用场景	超大规模AI训练
类别	8U2P GPU服务器
GPU	NVLink模组：8GPU模组
CPU	2个第四代英特尔®至强®可扩展处理器（Sapphire Rapids），支持单处理器最大350W功率
硬盘	支持多种不同的硬盘配置，硬盘支持热插拔 •最多可配置25个2.5英寸SAS/SATA硬盘 •最多可配置8个NVMe SSD盘 • 2个M.2 SSD
RAID	可选配支持RAID0、1、10、5、50、6、60等，支持Cache超级电容保护
IO	高性能拓扑：8*PCIe标卡，1/2*OCP3.0卡，8*NVMe 5.0，其中8张NVMe盘可以替换成8张标卡 均衡拓扑：12*PCIe标卡，1/2*OCP3.0卡，8*NVMe 5.0，其中8张NVMe盘可以替换成8张标卡
电源	6个54V双输入热插拔电源模块，2个12V热插拔电源模块，均支持N+N冗余
风扇	GPU区域：10个54V风扇，支持N+1冗余 CPU区域：5个12V风扇，支持N+1冗余
温度	5℃~35℃
尺寸	356mm×447 mm×898 mm（高x宽x深）

*注：支持2*OCP需要更换模块支持

灵活拓扑，高性能拓扑，满足CPU到GPU高带宽通信需求

- **CPU到GPU通信带宽达到512GB/s，相比均衡拓扑带宽高一倍**



- CPU-GPU 通信带宽8*X16 PCIe5.0
- GPU:IB:NVMe=1:1:1
- 支持NVMe SSD x8，每个Switch下挂2个
- 支持8张 x16 PCIe5.0网卡+1张OCP扩展，支持400G网卡

CPU inter-connection

PCIe 5.0 x16

PCIe 5.0 x16

PCIe 5.0 x4

PCIe 3.0 x8

01 AIGC算力需求评估

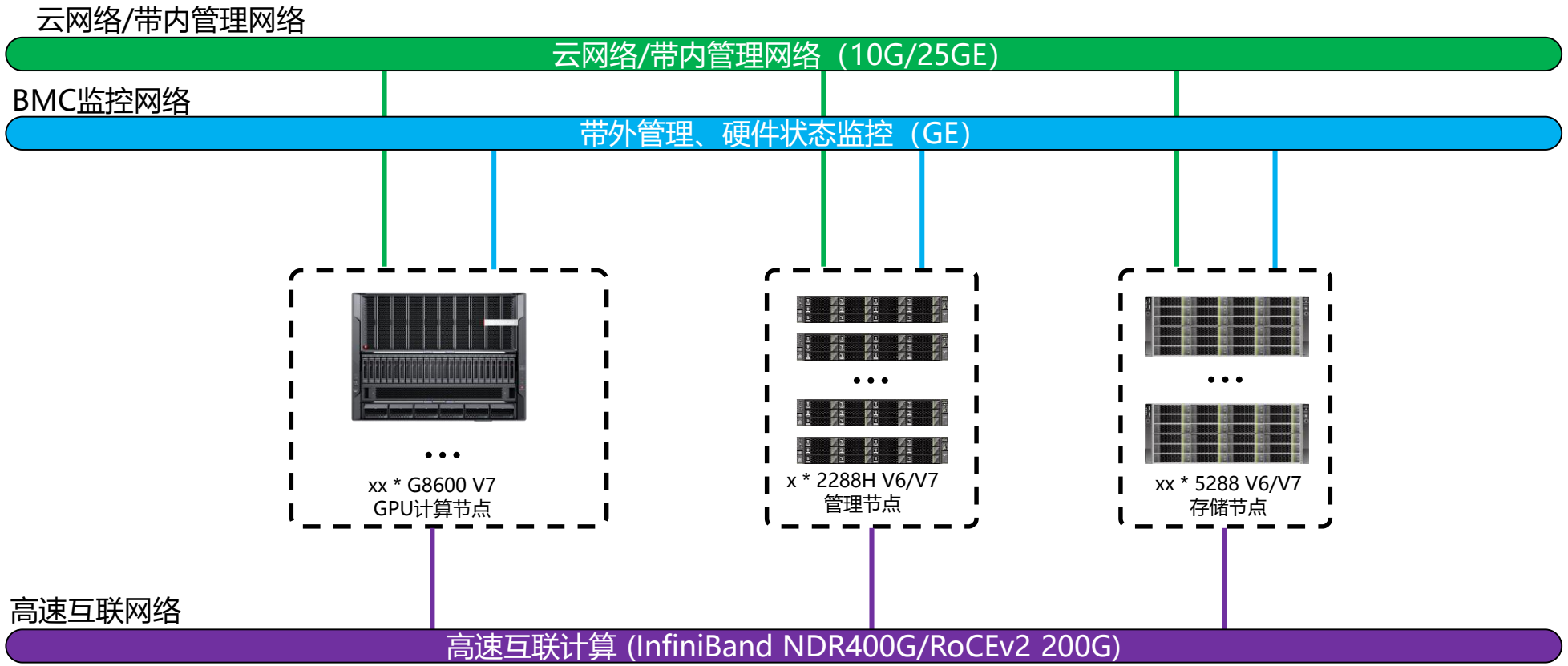
02 GPU卡技术参数

03 超聚变产品介绍

04 推荐组网方案

05 讨论

整体方案概览



GPU训练服务器推荐配置

机型	G8600/G8600E V7	备注
GPU	H800 8GPU模组	
CPU	2*84xx（Intel SPR）	每个GPU卡 8核，推荐32核以上CPU 高性能CPU满足后续扩展诉求
Memory	≥1TB	1.5倍以上显存容量
IB/RoCE	4个NDR 400或8个RoCE 200G	
NVME	4块 3.2TB	如果有大量小文件，需要NVME作为远端存储缓存盘，满足后续多场景需要。
SSD盘	2*960GB SATA SSD	OS盘

大模型训练节点数参考

模型参数量(Billion)	模组数(H800)	节点数	运行时间(Days)
6	64	8	9
	128	16	4
50	256	32	16
	512	64	8
175	256	32	65
	512	64	30
	1024	128	14

存储性能参考

场景	好（GBps）	更好（GBps）	最佳（GBps）
单节点读	4	8	40
单节点写	2	4	20
32节点聚合读	15	40	125
32节点聚合写	7	20	62
128节点聚合读	60	160	500
128节点聚合写	30	80	250

好：自然语言处理
更好：压缩图像 ImageNet
最佳：未压缩图片，高清视频

FusionOne DFS：行业领先存储性能，满足高带宽、大IOPS场景应用

极致存储性能

- 300K+ IOPS/PB@8TB SATA HDD
- 25GB/s/PB@8TB SATA HDD
- 600K+ IOPS@48块NVMe SSD
- 44GB/s@48块NVMe SSD

数据来源：实验室数据

兼容性强：兼容多种客户端和多样性数据

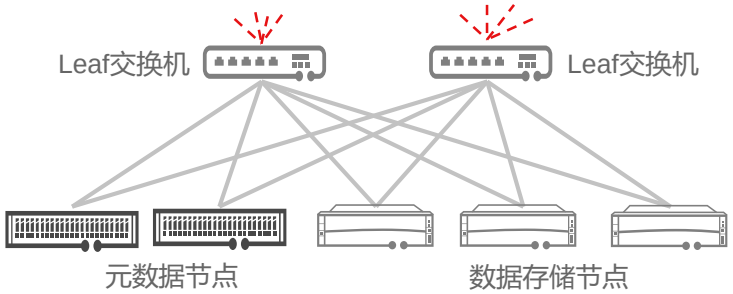
- 兼容多种Linux、Windows客户端，减少手动文件传输
- 文件切片聚合：自研切片、聚合算法，满足多样性数据读写需求

空间利用高：业界最高EC空间利用率

- EC空间利用率94.1%，业界最高
- 可灵活选择EC和多副本技术
- 灵活架构部署：支持对称/非对称式架构灵活部署

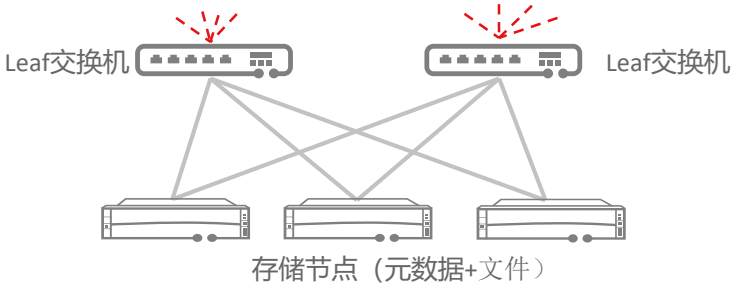
部署方式灵活

① 元数据和数据分离部署，提升元数据访问效率



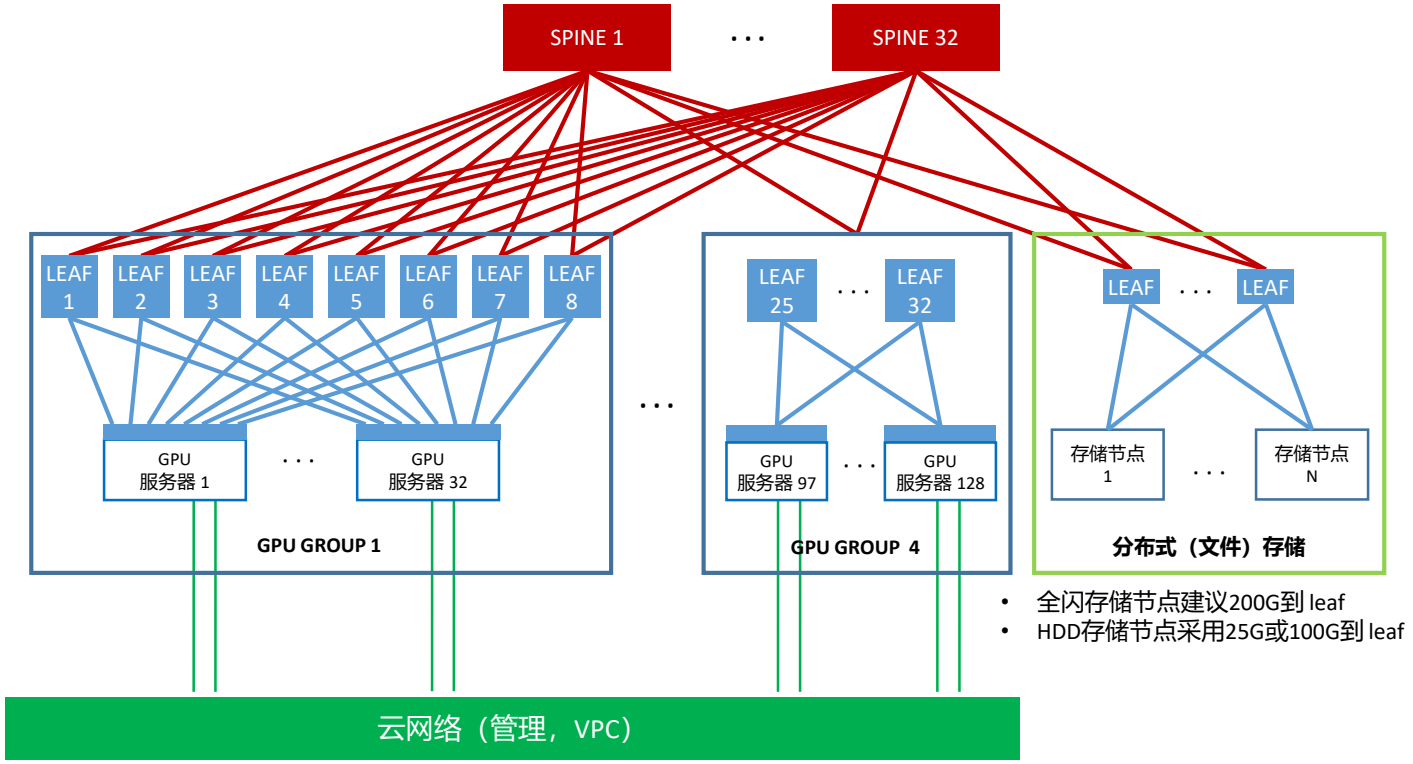
针对海量小文件场景，元数据集中在单独HA节点

② 对称式部署，灵活配置



元数据和数据部署在同一节点，中小规模更灵活

128节点8模组H800 GPU服务器组网方案1--全闪存储



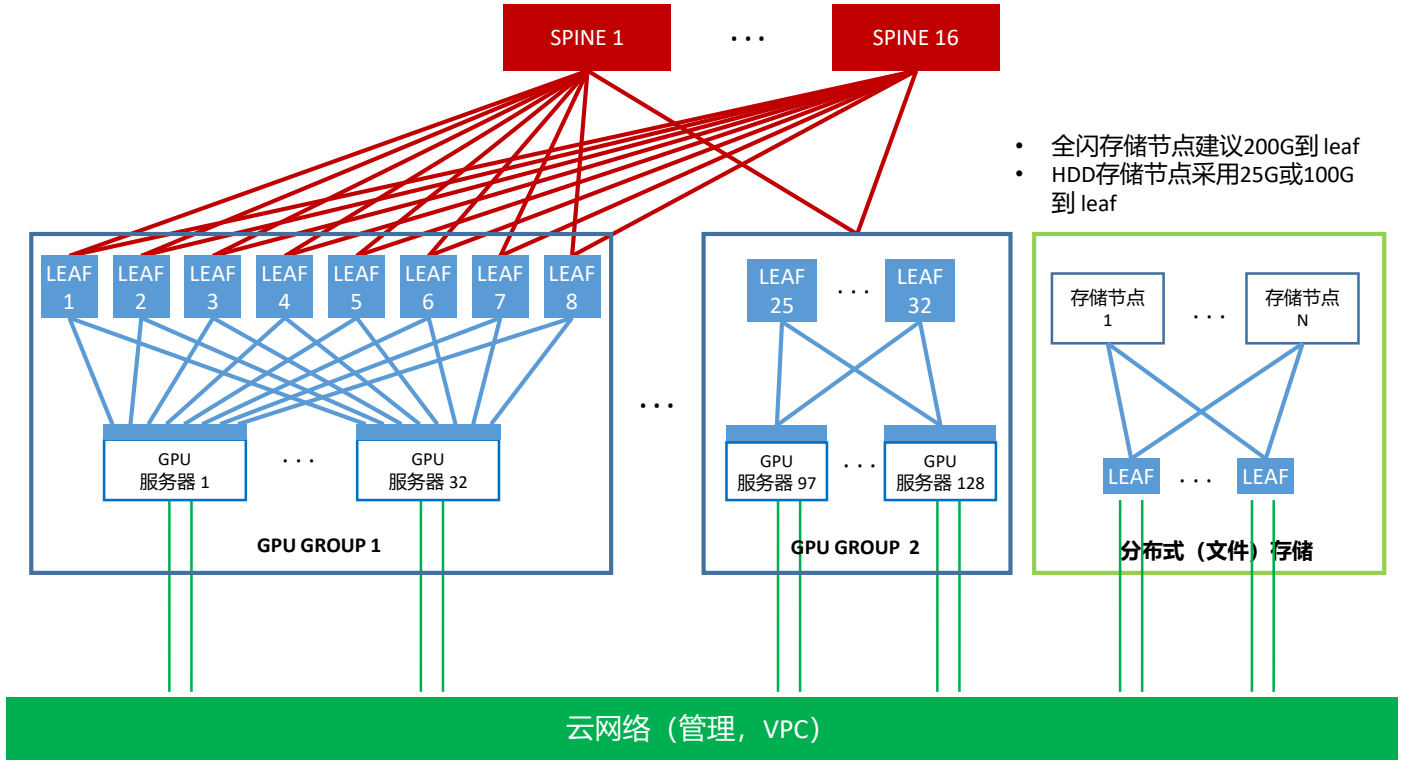
1024个 (Nvidia H800) AI训练集群组网建议

每台AI服务器组网需求：

- 参数平面 8*200GE RoCE
- 存储平面 共用参数平面网络
- 带内管理 + 云网络 2*10GE
- 带外管理平面1*GE（IPMI）

交换平面	建议集采交换机端口数	交换机数量
参数面组网 Spine交换机	32*400GE	32
参数面组网 Leaf交换机	32*200GE（下行） +16*400GE（上行） 收敛比1:1	32
存储平面接入交换机	48*25GE(下行） +6*100GE(上行含M-Lag） 4柜1组（1+1）	视存储规模而定
带内管理 + 云网络接入交换机	48*10GE(下行）+ 2或4 * 100G（上行）	4
带外管理接入交换机	48*GE(下行）+ 2*10GE（上行）	2

128节点8模组H800 GPU服务器组网方案2 –HDD存储



- 全闪存储节点建议200G到 leaf
- HDD存储节点采用25G或100G到 leaf

1024个 (Nvidia H800) AI训练集群组网建议

每台AI服务器组网需求：

- 参数平面 8*200GE RoCE
- 存储平面 共用参数平面网络
- 带内管理 + 云网络 2*10GE
- 带外管理平面1*GE（IPMI）

交换平面	建议集采交换机端口数	交换机数量
参数面组网 Spine交换机	32*400GE	16
参数面组网 Leaf交换机	32*200GE（下行） +16*400GE（上行） 收敛比1:1 16个400G采用1分2线缆成32个200G	32
存储平面接入交换机	48*25GE(下行） +6*100GE(上行含M-Lag） 4柜1组（1+1）	视存储规模而定
带内管理 + 云网络接入交换机	48*10GE(下行）+ 2或4 * 100G（上行）	4
带外管理接入交换机	48*GE(下行）+ 2*10GE（上行）	2

01 AIGC算力需求评估

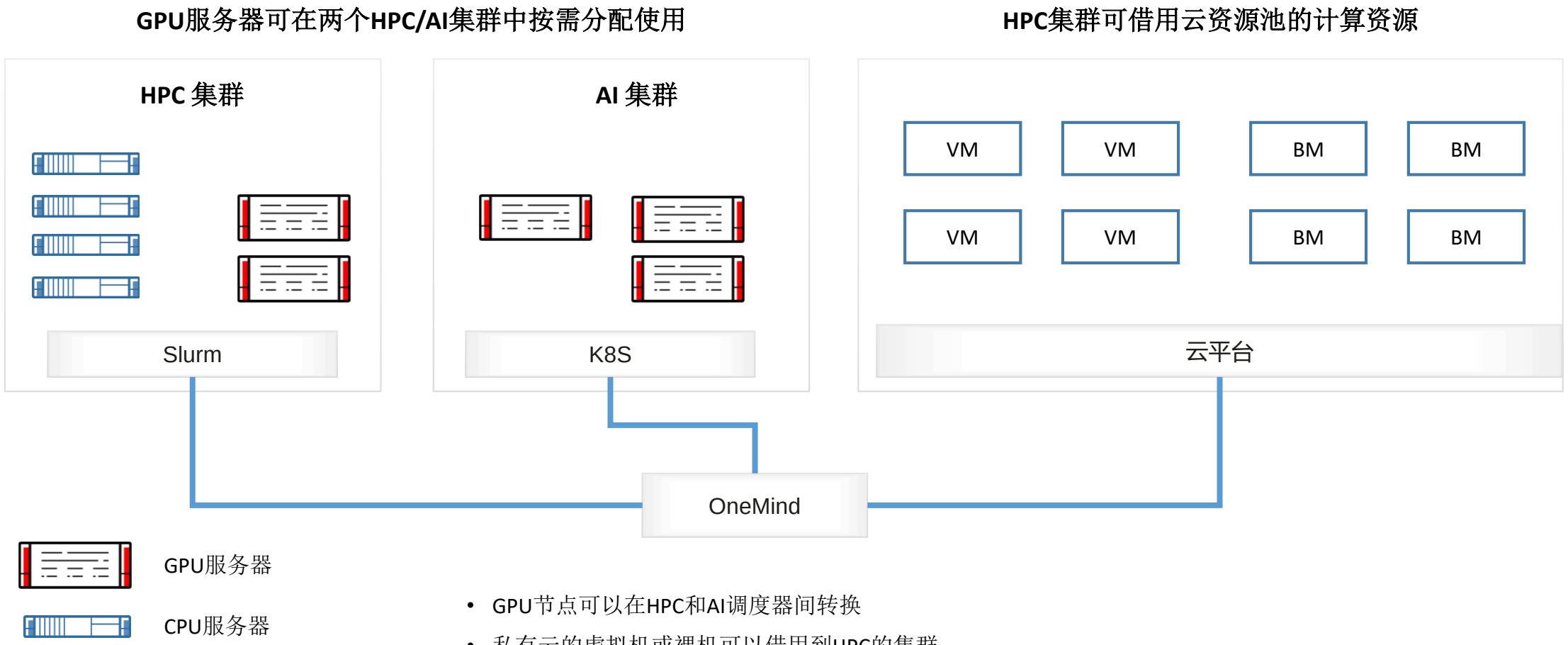
02 GPU卡技术参数

03 超聚变产品介绍

04 推荐组网方案

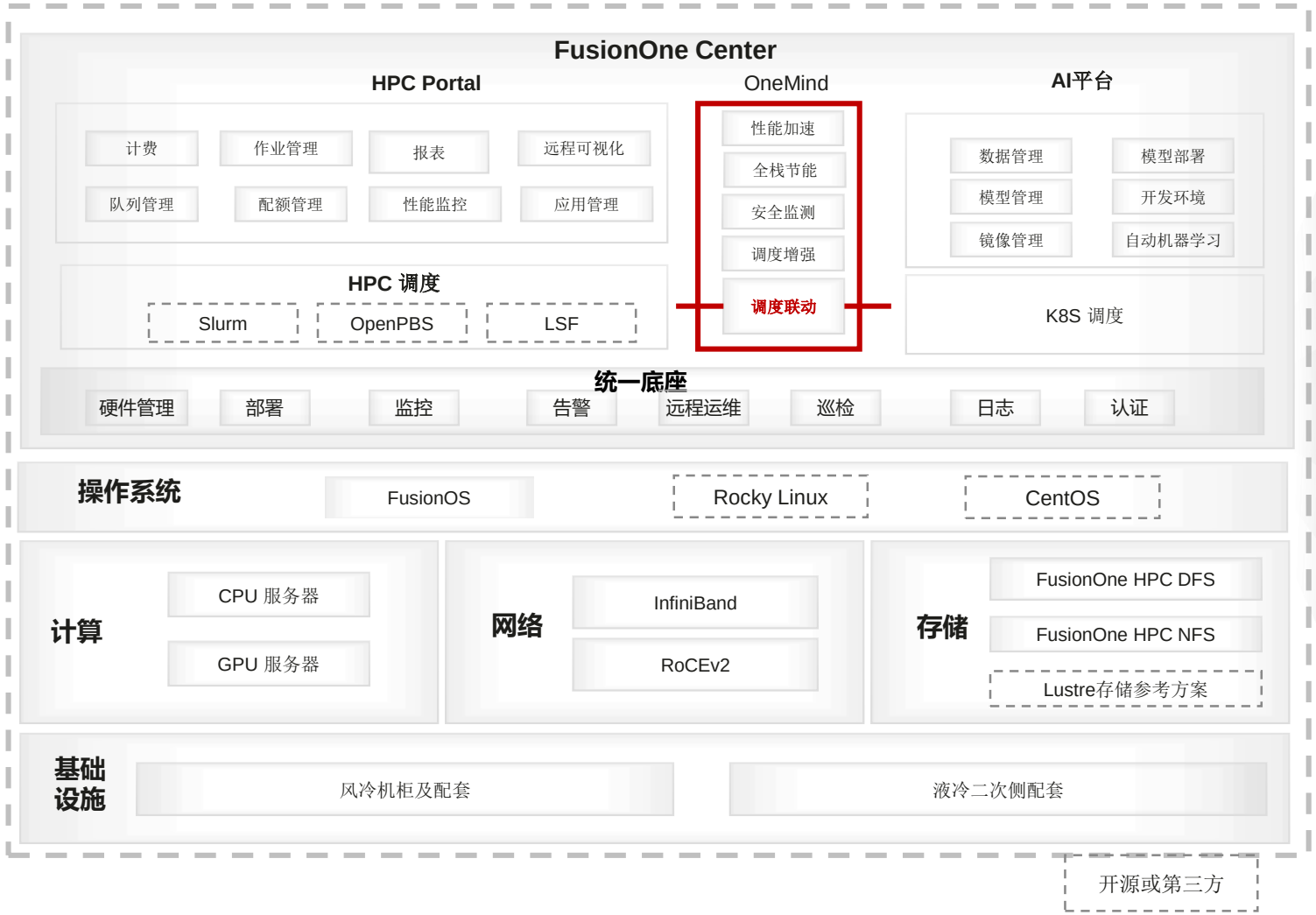
05 讨论

讨论1：HPC、AI和虚拟化三大场景资源打通



- GPU节点可以在HPC和AI调度器间转换
- 私有云的虚拟机或裸机可以借用到HPC的集群

HPC+AI场景



算力互通、业务共存

- 通过OneMind联动 HPC调度、AI调度与云调度，实现三方资源协调调度*
- 统一数据底座，支持HPC/AI/大数据业务共享

高效运行

- OneMind智慧引擎自动调优，应用运行效率**提升10%**
- FusionOne HPC DFS高性能文件存储方案：**25GB/s/PB** @8TB HDD；**30GB/s**@48块NVMe SSD
- CXL， HBM 分级内存*

运维简单

- 异常资源监测，防入侵、防越权、防挖矿
- 20+系统关键数据秒级收集，性能异常实时告警，快速定位系统故障

绿色节能

- OneMind应用感知智能降频&调度技术，整体**节能10%**
- FusionOne HPC全液冷方案，**PUE低至1.1**

讨论2：AI平台

1. AI开发平台支持AI模型全生命周期管理

- 支持AI开发数据管理、算法开发、模型训练、模型管理、推理服务和镜像管理；
- 支持FOC底座统一集成管理
- 支持gpu资源细粒度管理

2. 提供AIGC最佳实践

- GPU服务器：G8600 V7; G5500 V6/V7；FusionPoD 800
- 网络：Infiniband NDR 400G；RoCE v2 200G（配套H800）
- 存储：DFS

Thank you.

让数字世界无限可能

Fusion X, Digital Infinity

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. xFusion may change the information at any time without notice.

xFUSION