



<http://www.deepazurecloud.com>

Densivelo S9867-128DH产品主打胶片

DEEP AZURE CLOUD TECH SOLUTIONS PTE.LTD.

时间：2026.10



目录

1

S9867-128DH产品介绍

2

S9867-128DH亮点技术

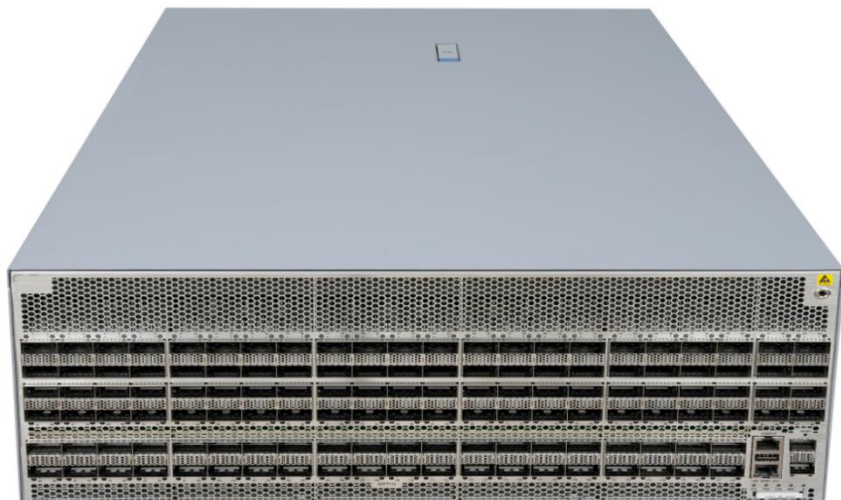
3

S9867-128DH组网方案

4

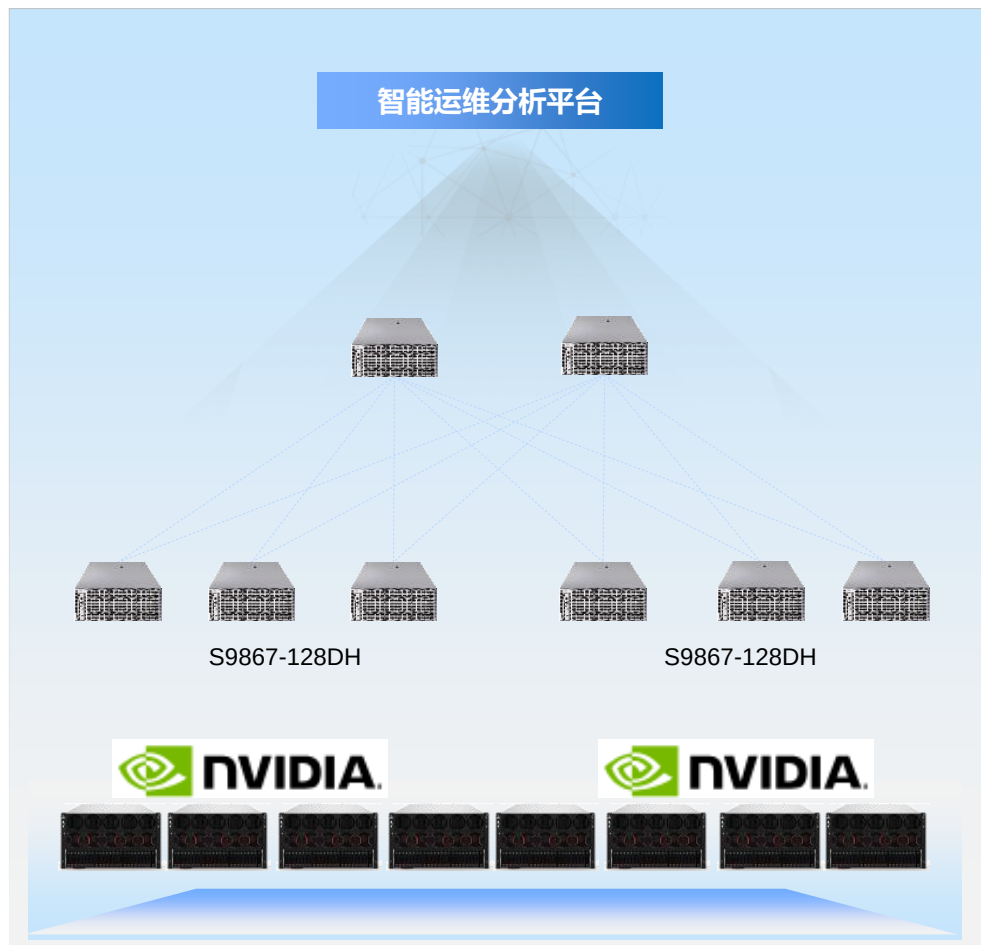
S9867-128DH竞争分析

S9867整机规格



硬件规格	项目	参数
整机参数	外形尺寸mm（高×宽×深）	175×442×760（4U）
	风扇模块	8个，7+1备份
	电源模块	4个，2+2备份
	散热方式	风冷散热，前后风道
	最大功耗	<3500W
整机性能	交换容量	102.4T
	包转发率	33.3Bpps
	转发时延	800ns
技术指标	处理器	8核，2.0GHZ
	DRAM Memory	16GB DDR4
	NOR Flash	64MB
	NAND Flash	240G
接口数量	Console口	1个
	管理网口	1*RJ45
	USB口	1*USB2.0
	400G接口	128*400G QSFP112
	10G接口	2个
环境参数	工作环境温度	0℃～40℃
	工作环境湿度	5%～95%（非凝露）
	存储环境温度	-40℃～70℃
	存储环境湿度	5%～95%（非凝露）

S9867软件规格



软件规格	项目	参数
TABLE	FIB容量	280K
	ECMP Group	4200
	ECMP 成员数	128
	ARP规格	4K
	MAC规格	8K
BFD	BFD session规格	256
	BFD性能	3*100ms
LAGG	LAGG组数	128
	LAGG成员数	128
INTERFACE	VLAN规格	4094
	VLANIF规格	4094
	Loopback口规格	1024
Buffer	缓存规格	160MB
ACL	ACL容量	4K
转发	Jumbo帧长度	9216
	Mirror Group	8

S9867产品光模块配套

种类	产品代码	产品描述	小类
400G	QSFP112-400G-DR4-SM1310	400G QSFP112单模光模块(1310nm,500m,DR4,MPO12/APC,支持一分四)	DR4
400G	QSFP112-400G-FR4-WDM1300	400G QSFP112单模光模块(1300nm,2km,FR4,LC)	FR4
400G	QSFP112-400G-SR8-MM850	400G QSFP112光模块(850nm,100m OM4,SR8,MPO16/APC)	SR8
400G	QSFP112-400G-VR4-MM850	400G QSFP112光模块(850nm,50m OM4,VR4,MPO12/APC)	VR4
400G	QSFP112-400G-D-AOC-10m-MM	400G QSFP112 to 400G QSFP112 10m多模有源光缆	多模有源光缆
400G	Q112-OSFP-RHS-400G-AOC-10m-MM	400G QSFP112 to 400G OSFP-RHS 10m多模有源光缆	多模有源光缆
400G	Q112-400G-2Q112-200G-AOC-10m-MM	400G QSFP112 to 2x200G QSFP112 10m多模有源光缆	多模有源光缆1分2
200G	QSFP56-200G-SR4-MM850	200G QSFP56 光模块-(850nm,100m OM4,SR4,MPO12/UPC)	SR4
200G	QSFP56-200G-D-AOC-5m	200G QSFP56 to 200G QSFP56 5m有源光缆	有源光缆
100G	QSFP-100G-SR4-MM850	100G QSFP28光模块-(850nm,100m,SR4,MPO)	SR4
100G	QSFP-100G-LR4-WDM1300	100G QSFP28光模块(1310nm,10km,LR4,WDM,LC)	LR4
100G	QSFP-100G-LR4L-WDM1300	100G QSFP28光模块(1310nm,2km,LR4L,CWDM4,LC)	LR4L
10G	SFP-XG-SX-MM850-D	SFP+ 万兆模块(850nm,300m,LC)	

目录

1

S9867-128DH产品介绍

2

S9867-128DH亮点技术

3

S9867-128DH组网方案

4

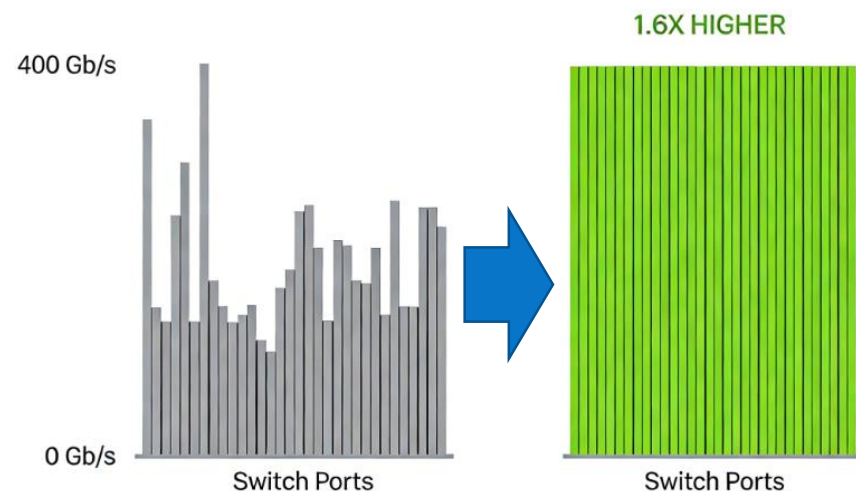
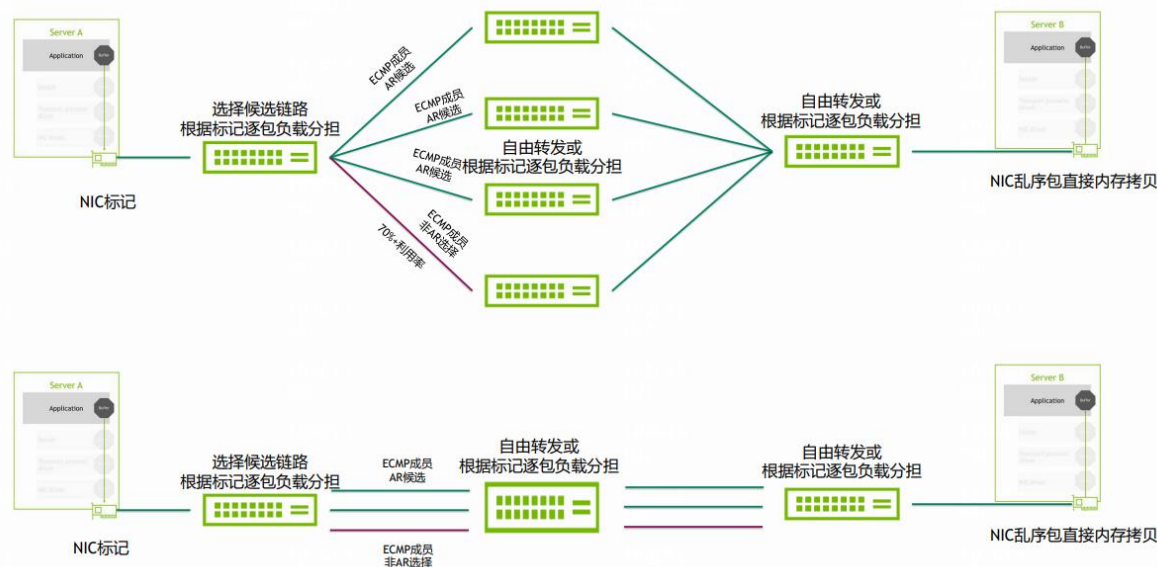
S9867-128DH竞争分析

亮点1：AR自适应路由，实现均匀负载分担

■ **基于端网协同**，AR自适应路由可以实现多平面多路径下的均匀负载均衡，大幅提升网络吞吐量。相对传统负载分担方式，AR可以将吞吐提升到原来的160%。

➢ **均匀负载**：通过评估出口队列负载情况，选择ECMP组中负载最轻的端口转发流量，从而实现均匀负载分担，提升带宽利用率。

➢ **降低时延**：每次转发选择拥塞最小的出口，由于队列长度较小，可以大幅降低转发时延。



相对传统负载分担方式，AR将带宽利用率从60%提升到95%

■ AR自适应路由实现原理

- **在发送端**：网卡动态标记AR流量，并逐包分发到不同平面。
- **在交换机上**：在负载分担时，交换机基于AR标记选择其出口队列负载最小的端口；由于同一数据流的不同数据包通过网络的不同路径传输，它们到达目的地时可能出现乱序。
- **在接收端**：网卡自动处理乱序数据包，并按顺序将数据传输给应用程序。这使得AR自适应路由的魔力对受益应用程序完全透明。

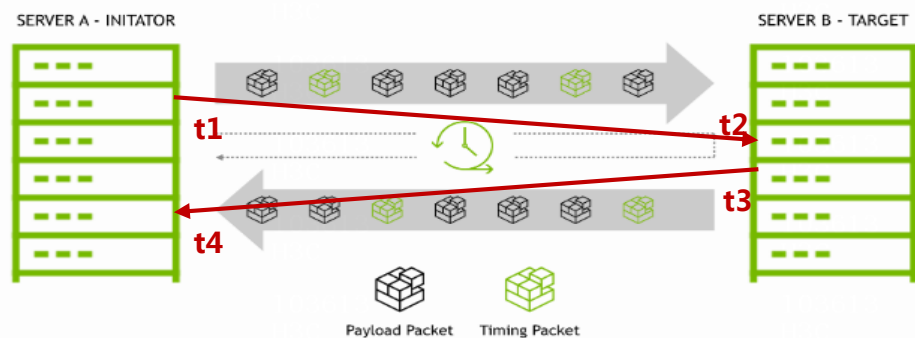
亮点2：端网协同，基于RTT实现实时拥塞控制

■ 基于RTT探测，SPX-CC可以实现**端到端拥塞感知**，并在端侧快速精准降速，降低网络拥塞。

➢ 拥塞感知快：相对DCQCN，RTT不需要发生丢包即可以感知拥塞，感知速度更快。不仅可以感知拥塞发生，也可以感知拥塞解除。若RTT变大，则可以快速启动降速；若RTT变小，则也可以快速恢复业务。

➢ 流量控制准：相对PFC端口级反压降速，SPX-CC只需要降低产生拥塞的流量即可，流量控制更精准。

LATENCY MEASUREMENT FOR NETWORK CONGESTION CONTROL



■ 实现原理：基于RTT进行拥塞控制

- 源端发送一个RTT探测报文，打上时间戳 t_1
- 目的端收到RTT报文时打上时间戳 t_2 ；并立即发出RTT回应报文，打上时间戳 t_3
- 源端收到RTT回应报文后，记录时间戳 t_4 ；计算 $RTT = (t_4 - t_1) - (t_3 - t_2)$
- 远端判断RTT变大，则认为出现拥塞，降低发送速率；RTT降低，认为拥塞解除，提升发送速率。

备注：RTTCC功能不需要1588配合

■ 业务需求

- A、B、C发流量到E；D发流量到F。

■ 预期效果

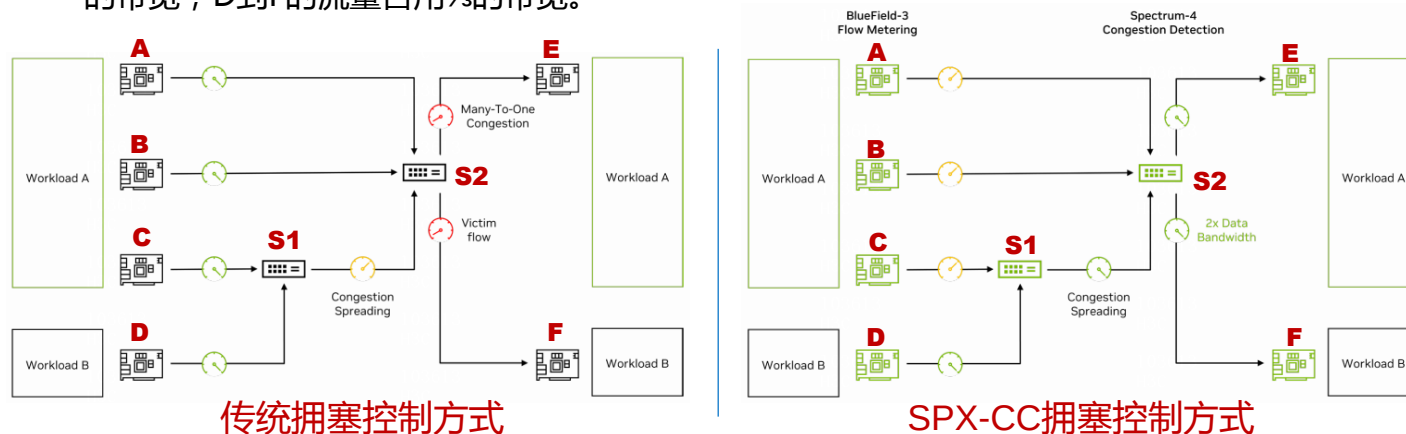
- A、B、C到E的流量各占用 $\frac{1}{3}$ 的带宽，D到F的流量占用 $\frac{1}{3}$ 的带宽。

■ 传统拥塞控制方式存在的问题

- 在S2上存在三打一，导致接口拥塞，从而会反压到S1，从而导致D到F的流量也下降到 $\frac{1}{3}$ 。

■ 部署SPX-CC的效果

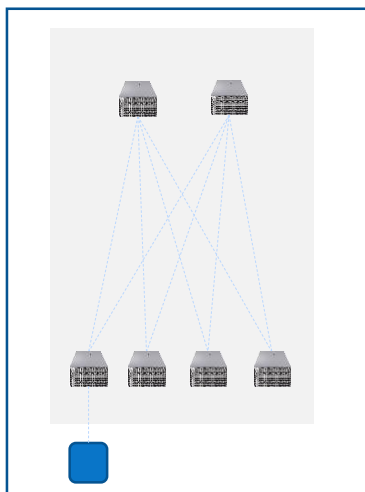
- 基于带内遥测数据，网卡感知网络拥塞状态，自动调整发送速率，A、B、C到E的流量各占用 $\frac{1}{3}$ 的带宽，D到F的流量占用 $\frac{1}{3}$ 的带宽。



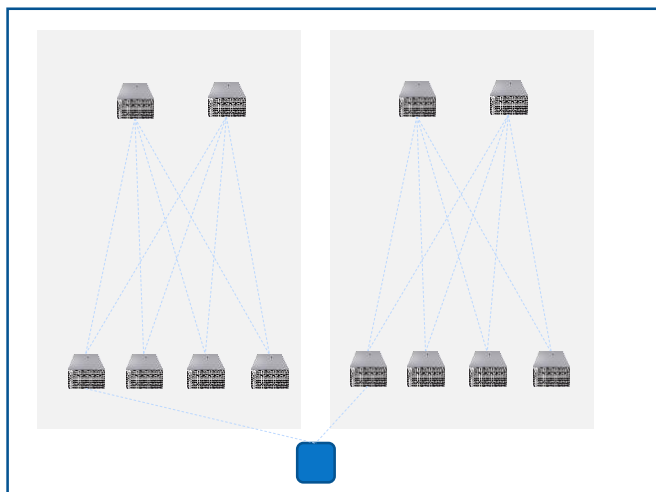
相对传统拥塞控制方式，SPX-CC更灵活更精细，实现端侧快速精准降速。

亮点3：多种组网方式，满足不同规模智算网络建设需求

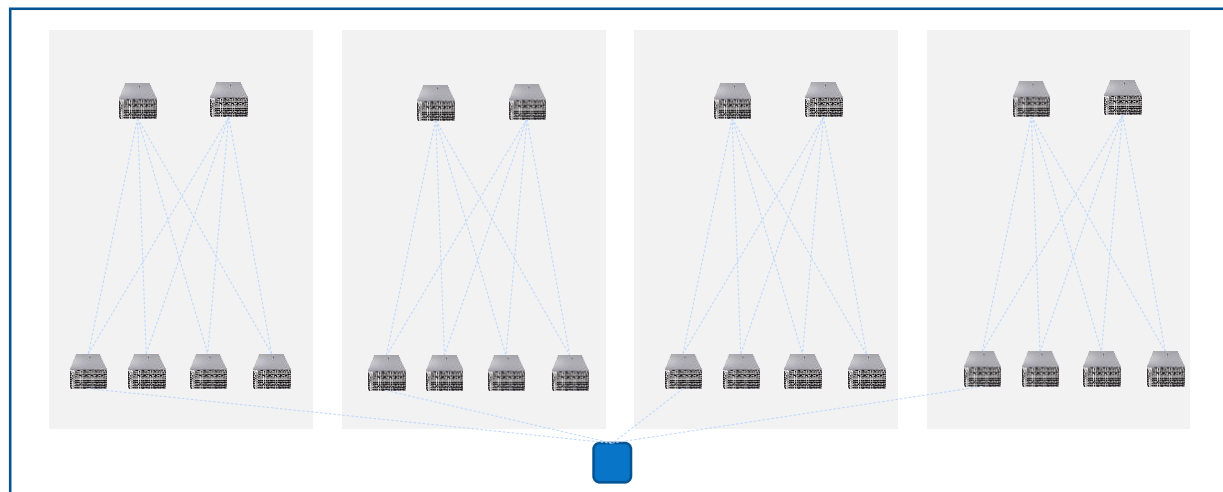
单平面组网



双平面组网



四平面组网



组网方式		适配网卡	集群规模	部署方案
单平面组网		CX7-1*400G	≤8192卡	端侧AR，网侧AR
双平面组网	S-Mlag组网	CX7-2*200G	≤15360卡	端侧开启bonding和AR，网侧逐流负载分担。
	软件双平面	CX8-C8180、CX8-C8240	≤8192卡	1、端侧：安装插件，拆分两个端口，配置两个IP，开启AR。 2、网侧：开启AR，平面间不需要互联。
	硬件双平面	CX8-C8180	≤7680卡	1、端侧：拆分1个VF和2个PF，配置一个IP，开启AR。 2、网侧：开启AR，平面间8*400G互联。
四平面组网		CX8-C8180	≤16000卡	1、端侧：拆分1个VF和4个PF，配置一个IP，开启AR。 2、网侧：开启AR，平面间8*400G互联。

亮点4：智能运维分析平台，全局掌控，智能运维

智能运维分析平台



控制



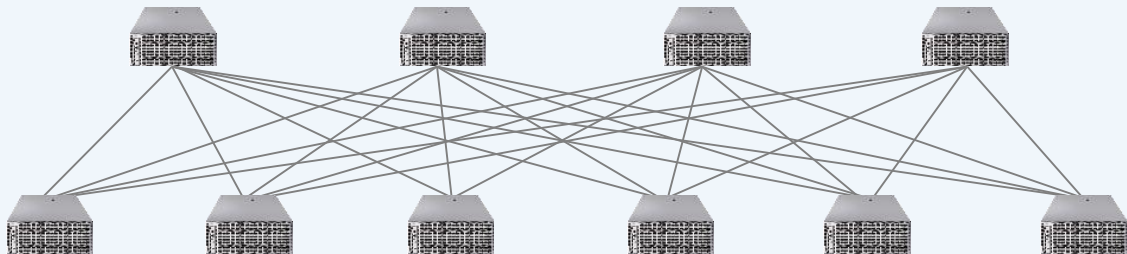
分析



监听



管理



网络域



计算域



存储域

智算中心架构

■ 智能运维分析平台

- **控制组件**：负责智算网络、计算节点，存储设备的自动化配置和流量调优等。
- **分析组件**：聚焦于算网存全域运维分析、流量分析和故障推理等。
- **监控组件**：负责网络、计算、存储的基础设施信息监控。
- **管理组件**：服务器管理系统，负责批量服务器部署、组件升级管理。



快速开局

自动部署，极速开局

- 服务器批量上线；
- 网络自动化部署；
- 端网协同，RoCE策略下发；

70%

节省开局时间



端到端流量
调优

端网协同，端到端流量调优

- AR自适应动态路由
- 逐流HASH
- 端到端拥塞控制

60%

训练效率提升



全流程
运维

智能运维，全流程护航

- 训前健康巡检；
- 训中实时监控；
- 训后跨域问题定位；

80%

排障效率提升

目录

1

S9867-128DH产品介绍

2

S9867-128DH亮点技术

3

S9867-128DH组网方案

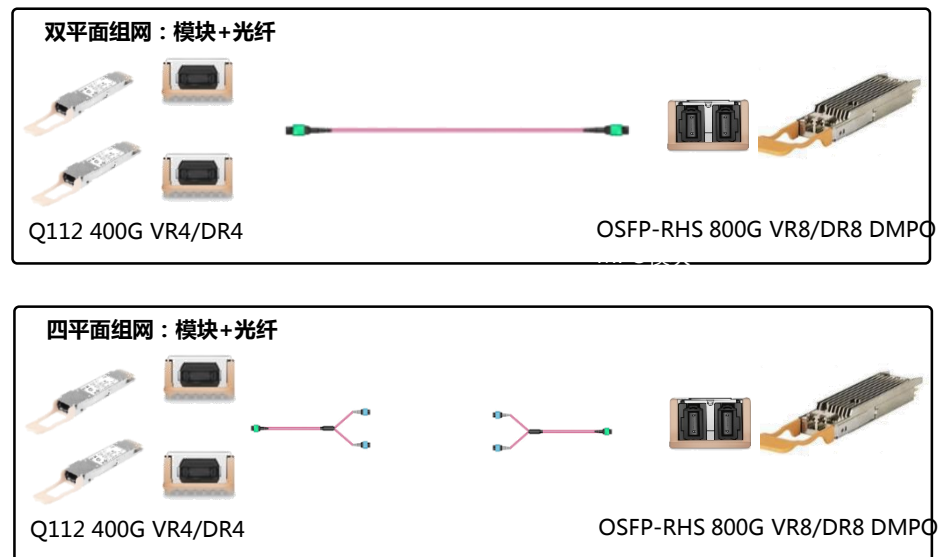
4

S9867-128DH竞争分析

网卡连接1：CX8-C8180连接方式



+



+

CX8 800G OSFP网卡

C8180 单端口 800G OSFP接口形态 (900-9X81E-00EX-ST0/DT0)

Model	NVIDIA SKU	Port Type	Supported Speed	Form Factor	PCIe Express
C8180	900-9X81E-00EX-ST0	Single-cage OSFP	InfiniBand: 800Gbs XDR IB (Default) Ethernet: 2x400GbE	Half-Height, Half-Length (HHHL)	PCIe 6 x16
	900-9X81E-00EX-DT0	Single-cage OSFP	InfiniBand: 800Gbs XDR IB (Default) Ethernet: 2x400GbE	Half-Height, Half-Length (HHHL)	PCIe 6 x16

注意：C8180网卡端口形态为单端口800G OSFP，以太协议下支持速率为2*400G

连接方式	S9867-128DH	光纤	CX8 1*800G OSFP网卡	传输距离
模块+光纤	QSFP112-400G-VR4-MM850	双平面：多模MPO-12光纤 四平面：多模MPO-12 1分2光纤	800G OSFP-RHS VR8 DMPO	50M
	QSFP112-400G-DR4-SM1310	双平面：单模MPO-12光纤 四平面：单模MPO-12 1分2光纤	800G OSFP-RHS DR8 DMPO	500M

网卡连接2：CX8-C8240连接方式



+



+

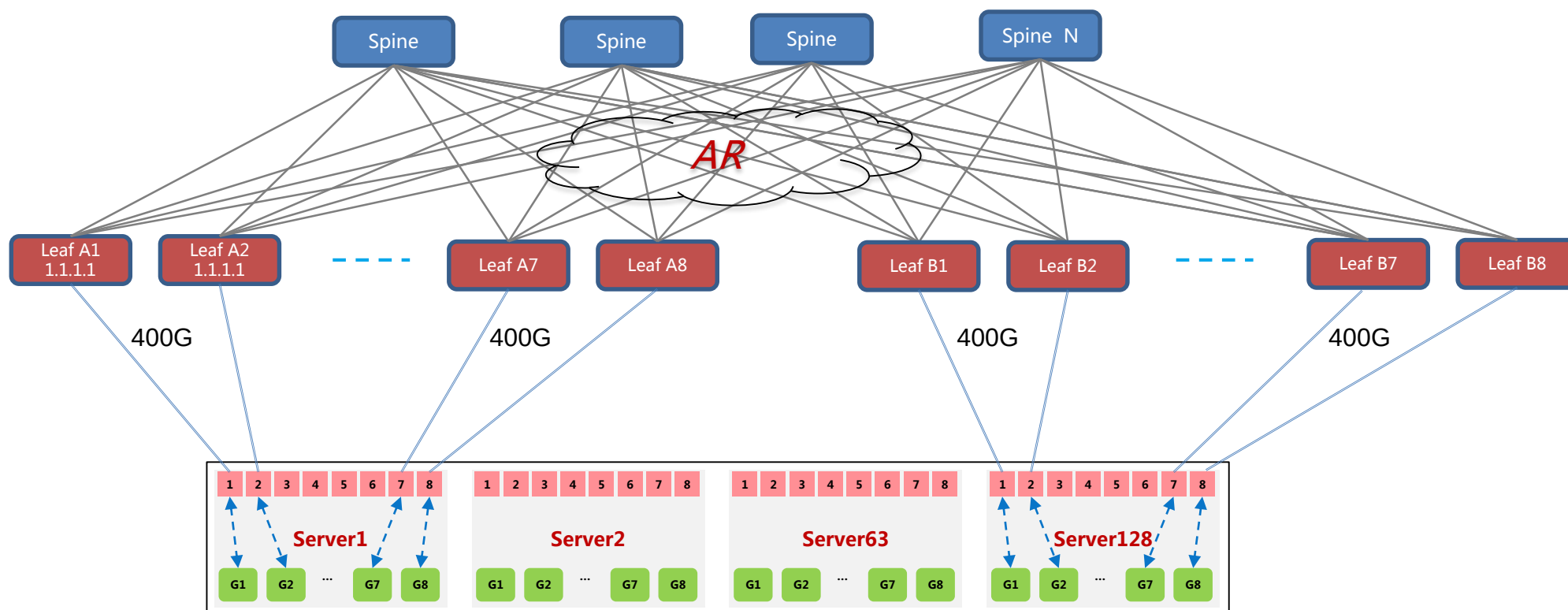
CX8 2*400G QSFP112网卡

C8240 双端口 400G Q112接口形态
(900-9X81Q-00CN-ST0)

Model	NVIDIA SKU	Port Type	Supported Speed
C8240	900-9X81Q-00CN-ST0	Dual-port QSFP112	InfiniBand: NDR IB Ethernet: 400GbE (Default)

连接方式	S9867-128DH	光纤	CX8 2*400G Q112网卡	传输距离
模块+光纤	QSFP112-400G-VR4-MM850	多模MPO-12光纤	UN-QSFP112-400G-VR4-MM850-M	50M
	QSFP112-400G-DR4-SM1310	单播MPO-12光纤	QSFP112-400G-DR4-SM1310	500M

组网方案1：单平面组网



组网：单平面，spine/leaf full-mesh互联

交换机设备数量：192台S9867交换机

服务接入：服务器8个CX7网卡

网卡：CX7，单口400G，开启AR能力，在交换机上开启逐包转发

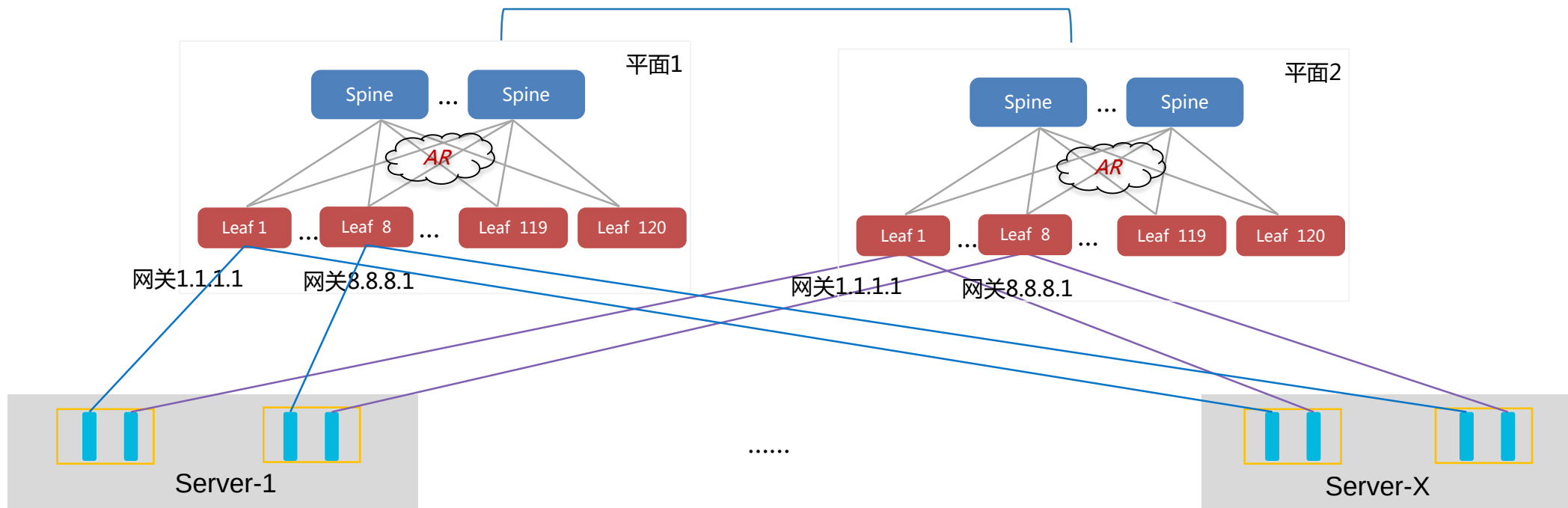
BGP自治系统号：Spine使用同一AS号，每个Leaf的AS号不同。

BGP邻居：Spine与Leaf设备建立EBGP邻居

交换机：全网开启AR

接入交换机：在leaf接入接口配置本地ARP代理，同时发布主机路由。

组网方案2：S-Mlag双平面组网（端侧逐流，网侧逐包）



网卡侧将400G接口拆分2*200G接口，分别连接到两个平面

网卡侧将收到的乱序报文进行重新排序

组网：双平面， spine/leaf full-mesh互联，平面间同号spine互联

交换机设备数量：368台S9867交换机

服务接入：服务器8个CX7网卡

网卡：CX7，单口400G拆分为2*200G，配置为bonding，开启AR

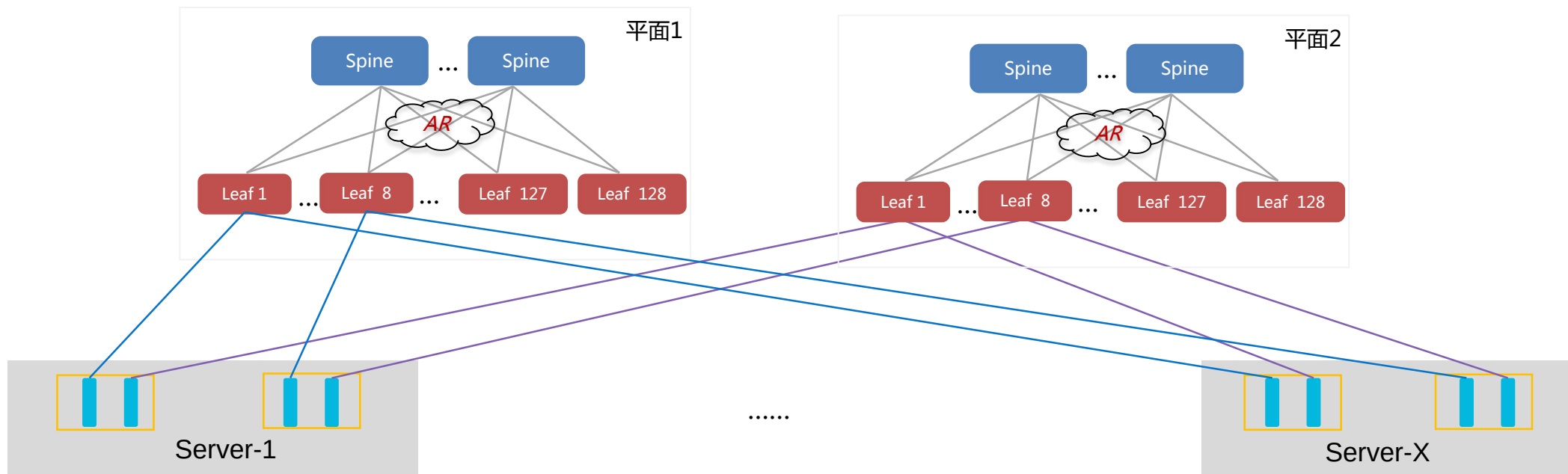
BGP自治系统号：平面内Spine使用相同AS号，每个Leaf的AS号不同；平面间Spine使用不同AS号。

BGP邻居：Spine与Leaf设备建立EBGP邻居，平面间Spine建立EBGP邻居

交换机：全网开启AR

接入交换机：在leaf部署S-Mlag，配置本地ARP代理，同时发布主机路由；同S-Mlag网关的IP和MAC要配置一致。

组网方案3：软件双平面组网（端侧逐流，网侧逐包）



网卡侧将800G接口拆分为两个逻辑接口，通过逐流分担将流量分担到多个平面

网卡侧将收到的乱序报文进行重新排序

组网：双平面， spine/leaf full-mesh互联，平面间不需要互联

交换机设备数量：384台S9867交换机

服务接入：服务器8个CX8网卡

网卡：CX8，单口800G拆分为2*400G，配置两个IP，开启AR模式

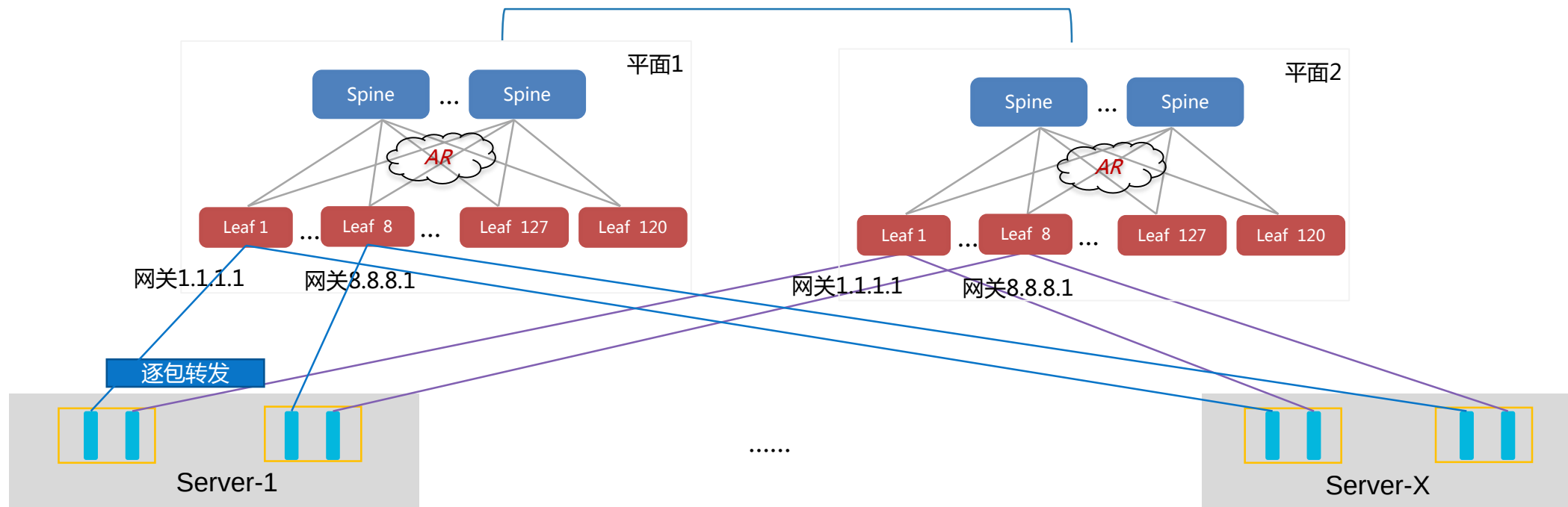
BGP自治系统号：平面内Spine使用同一AS号，每个Leaf的AS号不同；平面间Spine使用不同AS号。

BGP邻居：Spine与Leaf设备建立EBGP邻居

交换机：全网开启AR

接入交换机：在leaf接入接口配置本地ARP代理，同时发布主机路由。

组网方案4：硬件双平面组网（端侧逐包，网侧逐包）



网卡侧将800G接口拆分为4个逻辑通道，通过逐包分担将流量分担到多个平面

网卡侧将从多个平面收到的乱序报文进行重新排序

组网：双平面， spine/leaf full-mesh互联，平面间需要互联

交换机设备数量：368台S9867交换机

服务接入：服务器8个CX8网卡

网卡：CX8，单口800G拆分为4*200G，配置1个IP，开启AR模式

BGP自治系统号：平面内Spine使用同一AS号，每个Leaf的AS号不同；平面间Spine使用不同AS号。

BGP邻居：Spine与Leaf设备建立EBGP邻居，平面间Spine建立EBGP邻居

交换机：全网开启AR

接入交换机：在leaf接入接口配置本地ARP代理，同时发布主机路由；两个平面同号Leaf的网关的IP和MAC要配置一致。

● ●

网卡：CX8网卡拆分为4*200G，对应800G网卡为一个VF，一个VF对应四个PF，RDMA流量在四个PF之间进行逐包转发。

目录

1

S9867-128DH产品介绍

2

S9867-128DH亮点技术

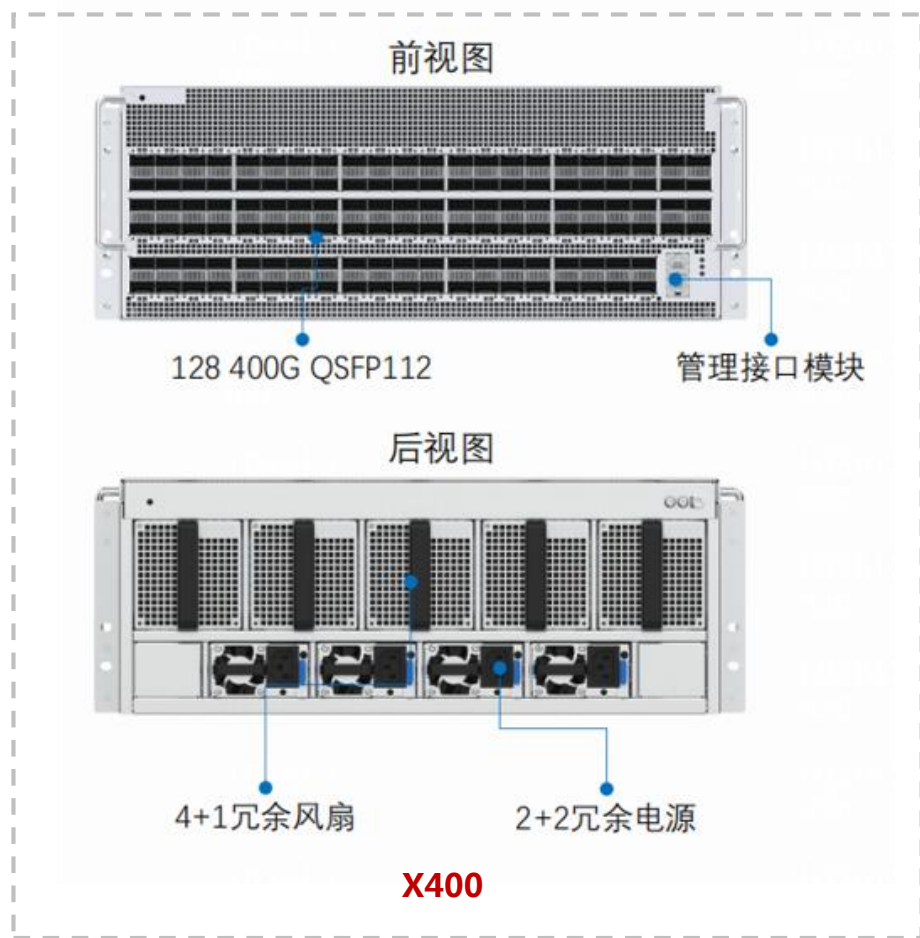
3

S9867-128DH组网方案

4

S9867-128DH竞争分析

元脉X400硬件竞争分析：S9867端口更丰富，可靠性更高

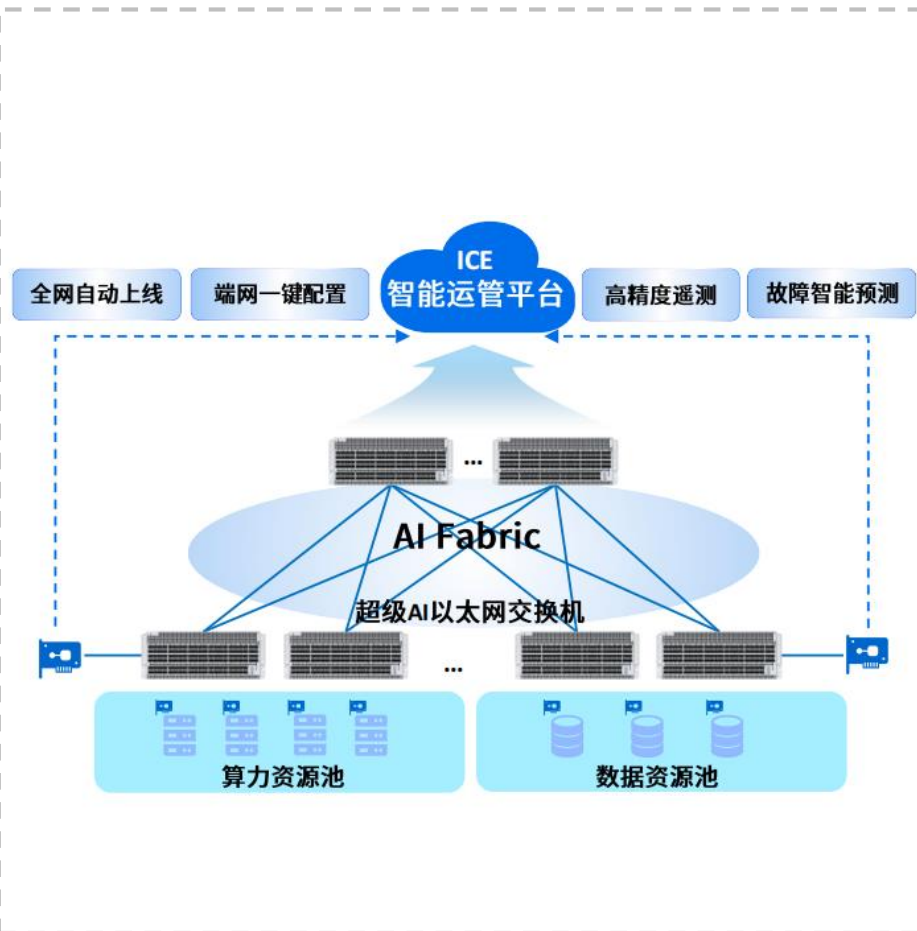


项目	S9867-128DH	X400 (SC8670EL-128QH)
400G端口	128	128
10G端口	2	不支持
交换容量	51.2T	51.2T
包转发率	待刷新	33.3Bpps
转发时延	待刷新	
交换芯片	51.2T ASIC	51.2T ASIC
CPU	2.0GHz@8Cores	暂不了解
内存	缺省16GB	缺省16GB
存储	缺省240GB	缺省240GB
BMC	已预留	支持
电源	2+2冗余	2+2冗余
风扇	7+1冗余	4+1冗余
风道	前后风道	前后风道
最大功耗	< 3500W	< 3500W
尺寸(H×W×D)	175mm×442mm×760mm	173.6 x 446 x 800mm
机箱	支持支持上盖灵活拆装	机箱上盖支持活动小窗，免拆箱更换SSD硬盘和BMC卡
工作温度	0℃ ~ 40℃	5℃~35℃

相对X400，S9867在硬件方面有3大亮点：

- 1、支持2个10GE接口，可以实现端口镜像、Sflow等功能。相对仅支持GE管理网口，可以实现更好的流量监控和设备管理功能。
- 2、风扇冗余度高，支持7+1备份，显著提升了系统在风扇故障时的散热余量和整体可靠性。
- 3、支持更宽的工作温度，设备可靠性更高。

元脉X400软件竞争分析：S9867在软件能力上全面领先



特性	X400支持情况	S9867优势项
操作系统	UNOS基于SONIC开发	支持成熟可靠商用操作系统
Bootware	通过 ONIE 安装 UXOS 镜像时，已经安装的 UXOS 系统会被删除，相应的配置文件也会丢失。	BIOS支持主备分区，并支持主备故障修复，升级版本不影响配置文件
License	当客户需要使用交换机的更多的功能模块或者需要更长时间使用交换机时，需要安装定制 license。	所有功能完全开放支持，不需要License控制
接口	1、不支持基于光模块类型实现速率自适应 2、配置端口拆分后需要重启生效	1、支持基于光模块实现速率自适应 2、端口拆分不需要重启设备。
负载分担	聚合和ECMP不支持逐包HASH	聚合和ECMP均支持逐包HASH
BFD	不支持BFD	支持BFD，512个3*100ms
AR	1、AR模式下，不能再配置和修改 qos 相关的配置。 2、使能AR之后，只有队列 3 使能 pfc。	1、AR模式下，可以配置和修改qos相关的配置。 2、AR模式下，支持4个队列使能pfc。
ACL	不支持用户自定义ACL	支持用户自定义ACL
PFC	在同一个端口最多开启 2 个无损队列。	每端口最多支持4个无损队列
CAR	出方向 ACL counter and meter 暂时不支持。	端口和vlan支持出入两个方向的CAR限速
NQA	不支持NQA	支持NQA，包括芯片硬件发包模式
DHCP	不支持DHCP	支持DHCP relay和DHCP server
Netstream	不支持Netstream	支持Netstream
生成树	不支持STP/RSTP/MSTP	支持STP/RSTP/MSTP
防攻击	不支持防攻击	支持ARP防攻击、攻击防御

相对X400，S9867采用成熟商用操作系统，在软件方面有3大亮点：

- 1、功能支持全：支持BFD、NQA、DHCP、STP等各种功能。
- 2、使用限制少：不需要License、支持灵活修改设备配置，且在调整配置后不需要重启等。
- 3、产品性能强：BFD性能强，支持多个无损队列等。

端网协同，极致性能，打造智算新纪元

