

Huawei FusionServer G5500 Server Competitive Positioning

HUAWEI TECHNOLOGIES CO., LTD.



Special Declaration

Usage of This Document	Training materials for Huawei sales personnel, including Huawei sales, channel partners, and pre-sales technical support.
Description	<ul style="list-style-type: none">➤ Introduction to G5500➤ Comparison Between G5500 and Competitor Specifications➤ Advantages and Disadvantages of G5500 and Competitor Specifications (International)➤ Advantages and Disadvantages of G5500 and Competitor Specifications (China)➤ How to Defend
Restriction Scenario	This document is used only for pre-sales learning. The product and vendor specifications provided in this document are compiled according to the information obtained from the network. Actually, the information of each vendor on the network is incomplete or may not be up-to-date. In addition, Huawei may update this course without notifying the customer. Therefore, this course cannot be used as a basis for project bidding or configuration selection.
Revised Version	V1.0 (2017-10)

Contents

1 Introduction to the G5500

2 Comparison Between G5500 and Competitor Specifications

3 Advantages and Disadvantages of G5500 and Competitor Specifications (International)

4 Advantages and Disadvantages of G5500 and Competitor Specifications (China)

5 How to Defend

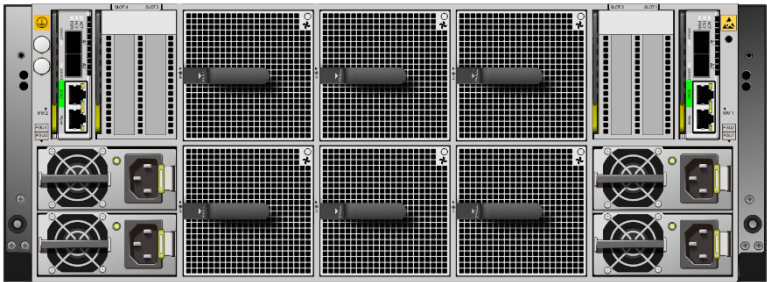
Specifications of the G5500 Full-Width Server

G5500 full-width server

Front view



Rear view



Form factor

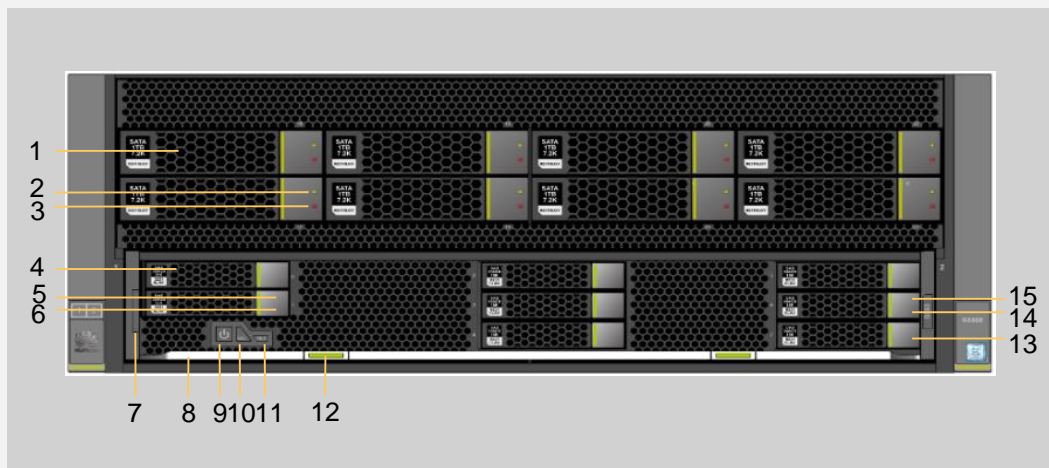
Full-width single-slot 2S blade server

Node	G560, full-width single node
Number of processors	2
Processor model	Intel Xeon E5-2600 v4 series, up to 22 cores
Number of GPUs	8
GPU model	NVIDIA Tesla P40/P100/V100
Memory	24 DDR4 DIMM slots, maximum memory capacity: 3 TB
Number of hard drives	6 x 2.5-inch NVMe SSDs , 2 x 2.5-inch SAS/SATA HDDs, and 8 x 3.5-inch SAS/SATA HDDs
RAID support	Supports RAID 0, 1, 10, 5, 50, 6, and 60 Supports supercapacitors
Power supply units	4 x 2,200 W Platinum PSUs, supporting 220 V AC and 240 HVDC, 2+2 redundancy
Fan modules	6 hot-swappable fan modules, supporting maintenance without opening the chassis; N+1 redundancy
Operating temperature	5°C to 35°C
Dimensions (H x W x D)	175 mm x 447 mm x 790 mm

Highlights

- **Excellent performance:** Supports 8 FHFL dual-slot 350 W GPU accelerator cards, 6 NVMe SSDs, 2 x 2.5-inch and 8 x 3.5-inch SAS/SATA HDDs.
- **Flexible topology:** A single server supports multiple topologies (1:4/1:8). A single node supports one-click switching of single-RC and balanced dual-RC topologies, and supports P2P and GPUDirect RDMA.
- **Easy maintenance:** The CPU and GPU nodes support independent evolution and maintenance. Fan modules and PSUs can be removed and inserted without opening the chassis, and main components can be replaced in one minute.

Front View of the G5500 Full-Width Server



1	3.5-inch hard drive	9	Power button/indicator
2	3.5-inch hard drive fault indicator	10	UID button/indicator
3	3.5-inch hard drive activity indicator	11	Health status indicator
4	2.5-inch hard drive	12	Release button
5	2.5-inch hard drive fault indicator	13	NVMe SSD
6	2.5-inch hard drive activity indicator	14	NVMe SSD activity indicator
7	Slide-out information label	15	NVMe SSD fault indicator
8	Ejector lever		

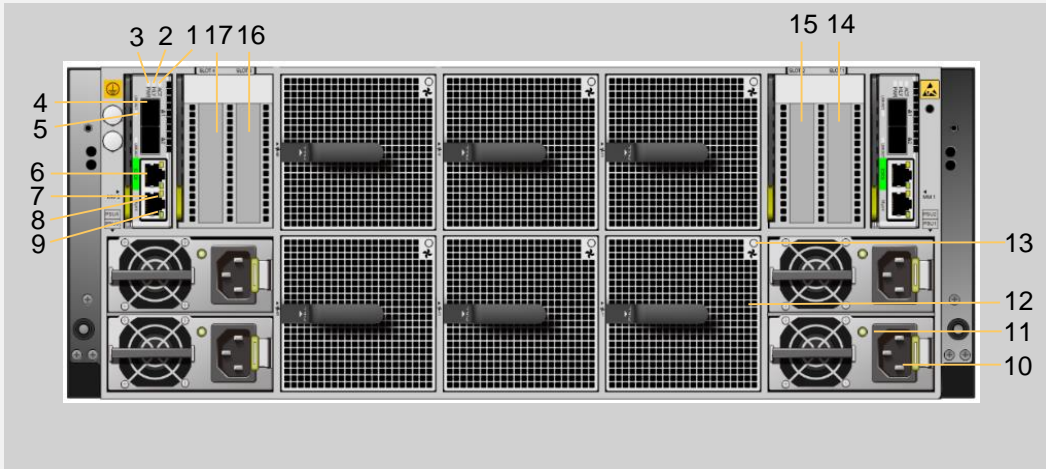
G560

- GPU node, supporting 8 GPU cards
- The GPU supports NVIDIA Tesla P40/P100/V100.
- Compute node, 2S + 24DIMM

Hard Drive

- 6 x 2.5-inch NVMe SSDs
- 2 x 2.5-inch SAS/SATA HDDs
- 8 x 3.5-inch SAS/SATA HDDs
- PCIe x4 NVMe SSD
- 6G/12Gbit/s SAS HDD
- Supports RAID 0, 1, 10, 5, 50, 6, 60, and supercapacitor for cache power-off protection

Rear View of the G5500 Full-Width Server



1	Management module: Active/standby indicator	10	PSU
2	Management module: Health indicator	11	PSU indicator
3	Management module: Power indicator	12	Fan module
4	10GE optical port	13	Fan module indicator
5	10GE optical port LINK/ACT indicator	14	Interface card slot 1
6	Management module: Serial port	15	Interface card slot 2
7	Management module: GE electrical port	16	Interface card slot 3
8	Management module: ACT indicator of the GE electrical port	17	Interface card slot 4
9	Management module: LINK indicator of the GE electrical port	18	

Management/IO Module

- Dual management modules (G560 supports only one module)
- 2 x 10GE optical ports (currently not supported by the G560)
- GE management network port and serial port

Management/IO Module

- 3 or 4 PCIe x16 expansion slots
- Supports HHL PCIe card

Power Supply

- Platinum PSUs, 2+2 redundancy
- Maximum power consumption of a single PSU: 2,200 W
- Supports AC and 240 V HVDC

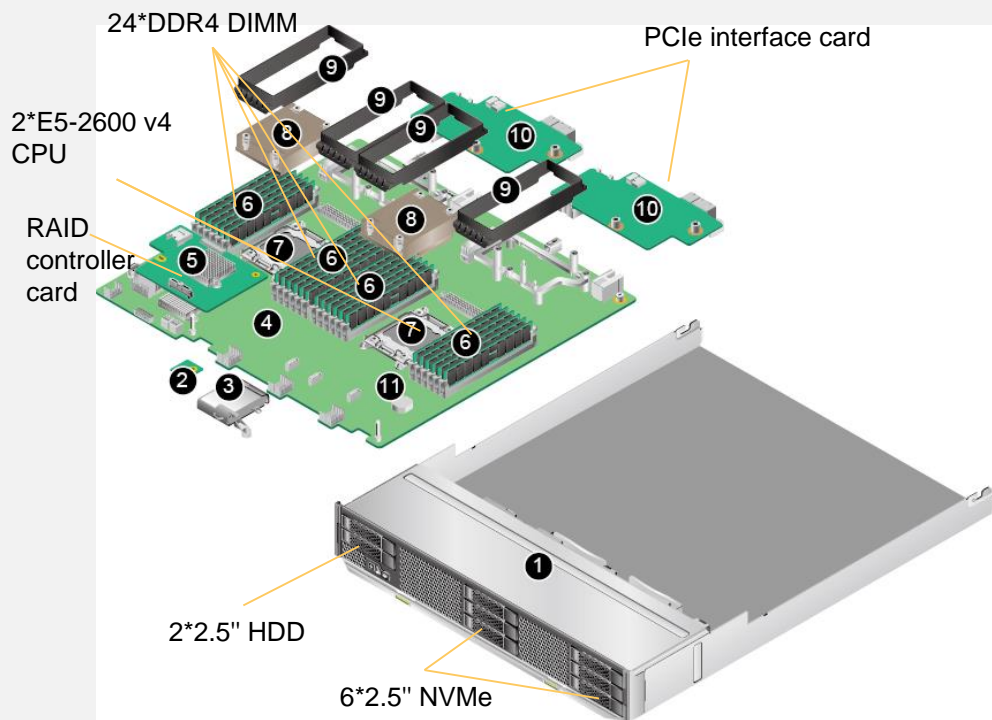
Fan Module

- 6 hot-swappable fan modules, N+1 redundancy
- Supports online maintenance without opening the chassis

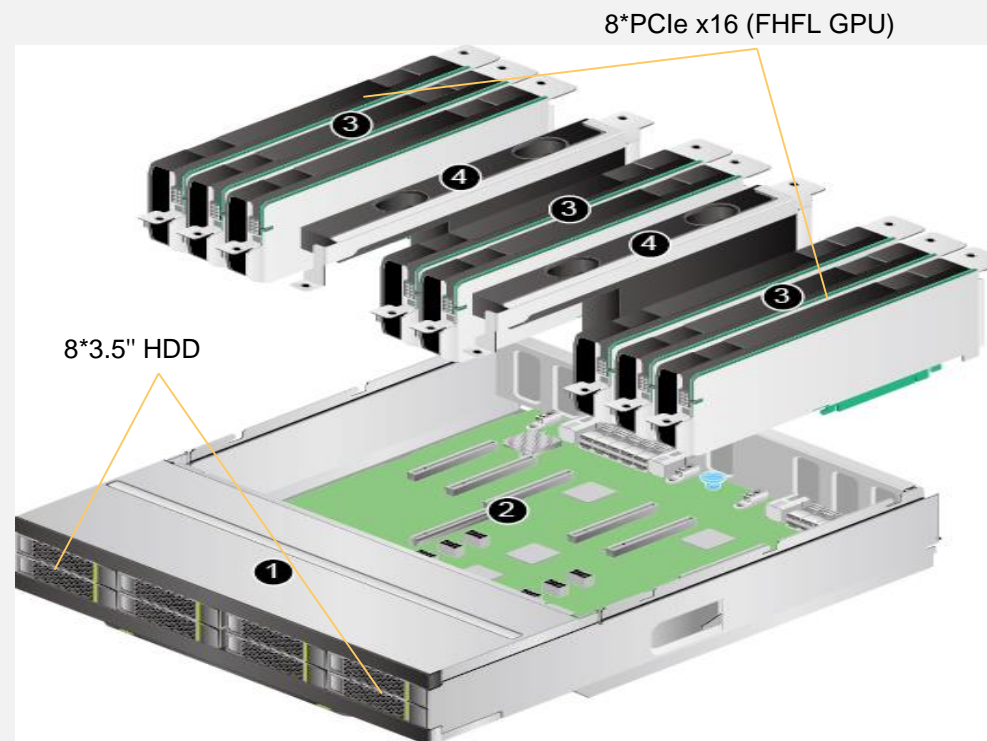
Internal View of the G5500 Full-Width Node

G560 Node

Product diagram



G560 CPU compute module



G560 heterogeneous compute module

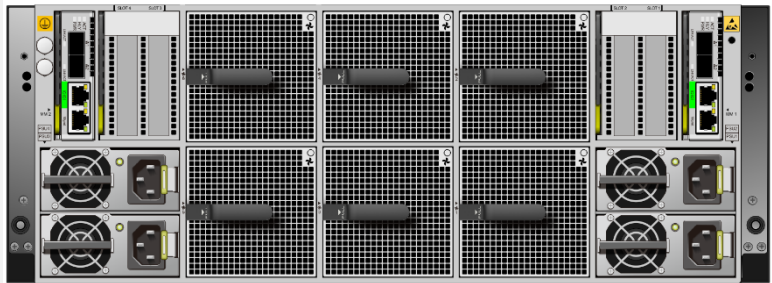
Specifications of the G5500 Half-Width Server

G5500 half-width server

Front view



Rear view

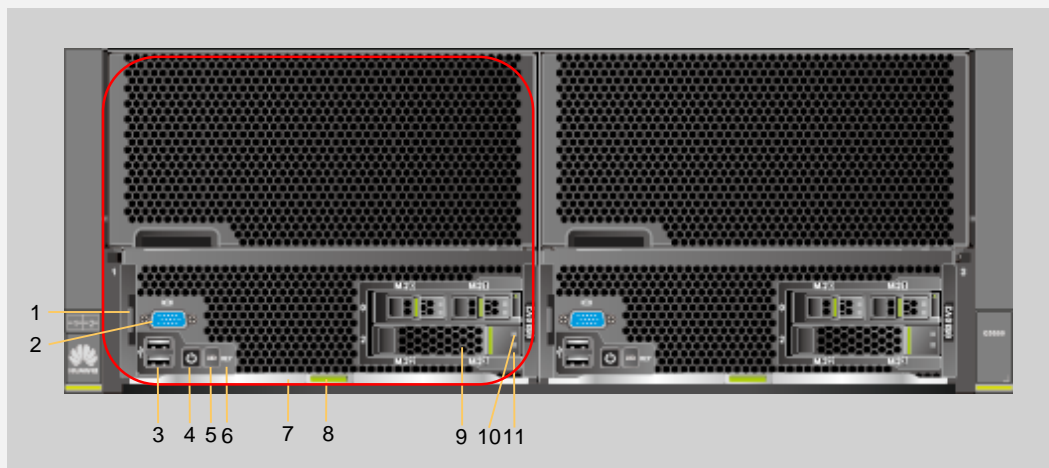


Form factor	Full-width single-slot 2S blade server
Node	G530 V2, half-width dual nodes, 16 GPUs
Number of processors	2
Processor model	Intel Xeon Scalable Processors, up to 28 cores
Number of GPUs	16
GPU model	NVIDIA Tesla P4
Memory	24 DDR4 DIMM slots, maximum memory capacity: 3 TB
Number of hard drives	2 x 2.5-inch SAS/SATA/NVMe or 4 x M.2 SATA SSDs
RAID support	RAID 0, 1, and 10
Power supply units	4 x 2,200 W Platinum PSUs, supporting 220 V AC and 240 V HVDC, 2+2 redundancy
Fan modules	6 hot-swappable fan modules, N+1 redundancy
Operating temperature	5°C to 35°C
Dimensions (H x W x D)	175 mm x 447 mm x 790 mm

Highlights

- **Excellent performance:** A single node supports 16 HHL 75 W GPU accelerator cards and supports 2 NVMe/SAS/SATA or M.2 SATA SSDs.
- **Flexible topology:** A single server supports the 1:8 topology and supports P2P and GPUDirect RDMA.
- **Easy maintenance:** The CPU and GPU nodes support independent evolution and maintenance. Fan modules and PSUs can be removed and inserted without opening the chassis, and main components can be replaced in one minute.

Front View of the G5500 Half-Width Server



1	Slide-out information label	7	Release button
2	VGA port	8	Ejector lever
3	USB3.0 interface	9	2.5-inch hard drive
4	Power button/indicator	10	2.5-inch hard drive fault indicator
5	UID button/indicator	11	2.5-inch hard drive activity indicator
6	Health status indicator		

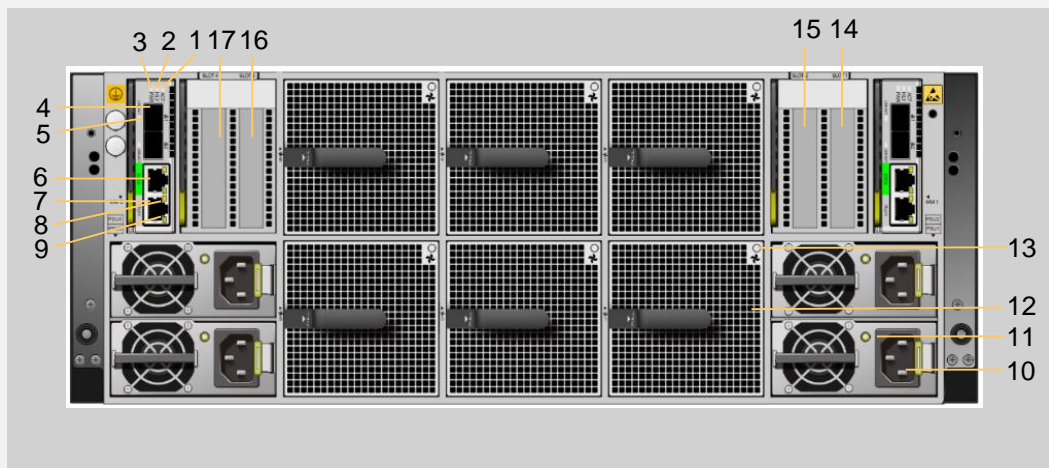
G530 V2

- GPU node, supporting 16 GPU cards
- The GPU supports NVIDIA Tesla P4.
- Compute node, 2S + 24DIMM

Hard Drive

- 2 x 2.5-inch SAS/SATA/NVMe or 4 x M.2 SATA SSDs
- PCIe x4 NVMe
- 6G/12Gbit/s SAS HDD
- RAID 0, 1,10

Rear View of the G5500 Half-Width Server



1	Management module: Active/standby indicator	10	PSU
2	Management module: Health indicator	11	PSU indicator
3	Management module: Power indicator	12	Fan module
4	10GE optical port	13	Fan module indicator
5	10GE optical port LINK/ACT indicator	14	Interface card slot 1
6	Management module: Serial port	15	Interface card slot 2
7	Management module: GE electrical port	16	Interface card slot 3
8	Management module: ACT indicator of the GE electrical port	17	Interface card slot 4
9	Management module: LINK indicator of the GE electrical port	18	

Management/IO Module

- Dual management modules
- 2 x 10GE optical ports
- GE management network port and serial port

Management/IO Module

- 4 PCIe x16 expansion slots
- Supports HHHL PCIe card

Power Supply

- Platinum PSUs, 2+2 redundancy
- Maximum power consumption of a single PSU: 2,200 W
- AC and 240 V HVDC

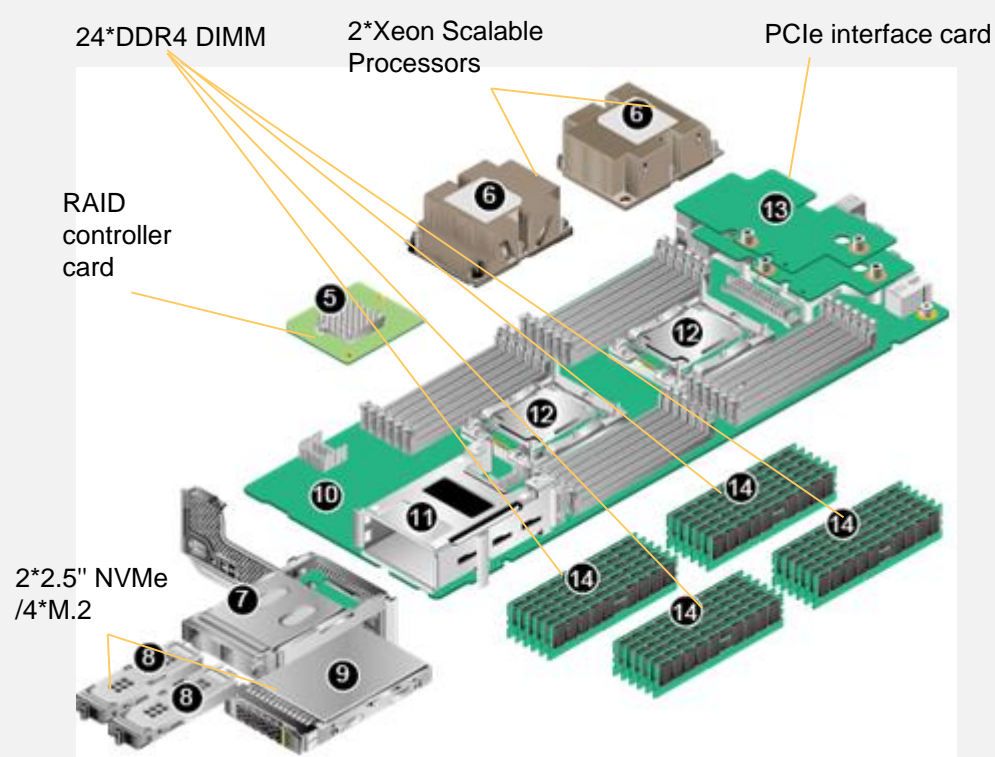
Fan Module

- 6 hot-swappable fan modules, N+1 redundancy
- Supports online maintenance without opening the chassis

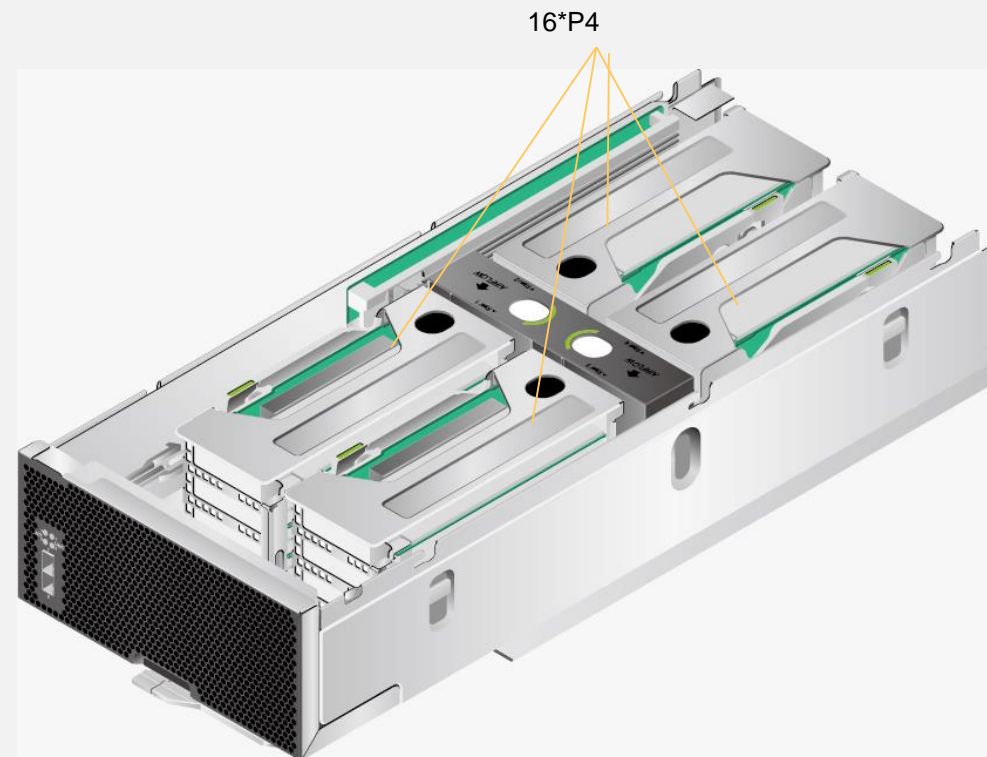
Internal View of the G5500 Half-Width Node

G530 V2 node

Product diagram



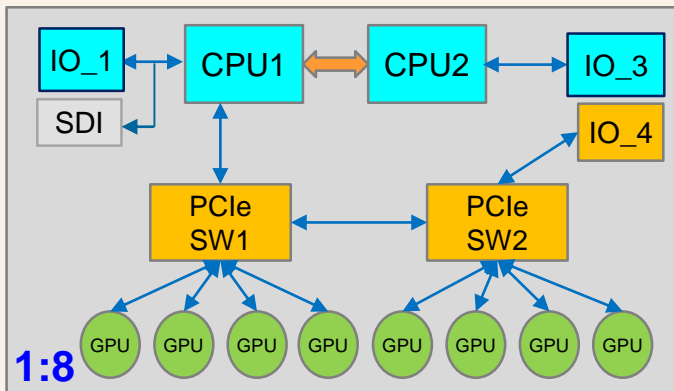
G530 V2 CPU compute module



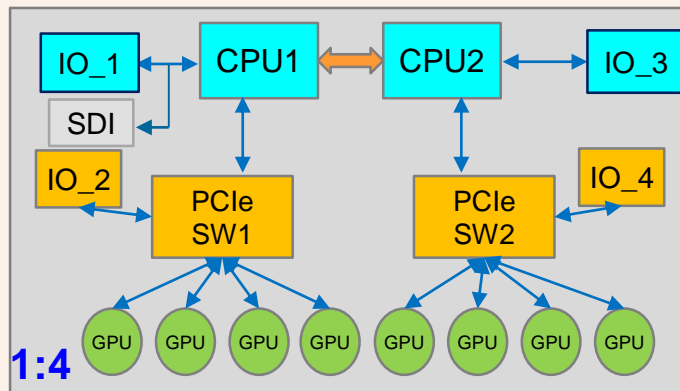
G530 V2 heterogeneous compute module

G5500 Topology Summary

8GPU server (G560 single node)

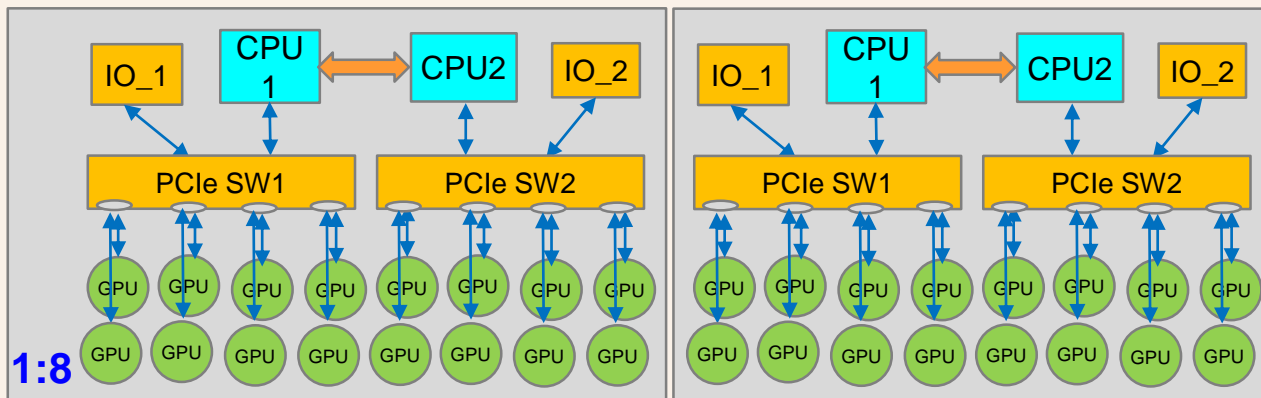


Topology 1: 8-card P2P, single-RC, GPUDirect RDMA; applicable to DL



Topology 2: Balanced, 4-card P2P, dual-RC, GPUDirect RDMA; applicable to public cloud and HPC; 4GPU or 8GPU configurations supported

16GPU server (G530 V2 dual nodes)



Topology 5: High-density, 32*P4/NPU; applicable to AI inference and AVI applications

Contents

1 Introduction to the G5500

2 **Comparison Between G5500 and
Competitor Specifications**

3 Advantages and Disadvantages of G5500 and
Competitor Specifications (International)

4 Advantages and Disadvantages of G5500 and
Competitor Specifications (China)

5 How to Defend

Models of International Vendors

8*GPU PCIe



4U-8G Huawei
G5500-G560



4U-8G SuperMicro
SYS-4028GR-TRT2



4U2N-8G HPE
Apollo 6500

8*GPU NVLink



4U-8G SuperMicro
SYS-4028GR-TXRT



3U-8G NVIDIA
DGX-1

4*GPU PCIe



4U2N-16G Huawei
G5500-G530 V2



2U-4G SuperMicro
SYS-2028GR-
TRHT



1U-4G SuperMicro
SYS-1028GQ-TRT



1U-4G Dell
PowerEdge C4130

4*GPU NVLink



1U-4G SuperMicro
SYS-1028GQ-TXRT



1U-4G Dell
PowerEdge C4130

Parameter	G5500-G560	G5500-G530 V2	SuperMicro 4028GR-TRT2	HPE Apollo 6500+	SuperMicro 4028GR-TXRT	NVIDIA DXG-1	Advantage
Dimensions (H x W x D)	4U (175*447*790)	4U (175*447*790)	4U	4U (177*448*960) + 1.5U	4U	3U (131*444*866)	No obvious advantage
Architecture	Modular	Modular, dual-node	Integrated	Modular, dual-node, power shelf required	Modular	Integrated	Structured design, powerful upgrade and evolution capability, and convenient maintenance of components
CPU+DIMM	2S + 24DIMM	Scalable, 4S + 48DIMM	2S + 24DIMM	4S + 32DIMM	2S + 24DIMM	2S + 16DIMM	The G530 V2 dual nodes support Scalable, and a single CPU supports 24 DIMMs.
Maximum CPU power	300 W	205 W	145 W	145 W	145 W	145 W?	The CPU uses independent air channels and has strong evolution capability.
GPU	8*P100/P40	32*P4	10*P100/P40	16*P100/P40	8*P100 NVLink	8*P100 NVLink	The G530 V2 dual nodes support 32*P4.
Maximum GPU power	350 W	350 W	250 W	350 W	300 W	300 W	High-power GPU with strong scalability
CPU:GPU topology	1:8/1:4 orchestratable	1:16	1:8	1:8/1:4	1:8	1:8	G560 full-width servers support flexible orchestration via software.
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	4*2.5" NVMe/SAS /SATA SSD/HDD or 8*M.2	24*2.5" SSD/HDD	Single node: 8*2.5" SAS/SATA SSD/HDD	16*2.5" SSD/HDD (8*NVMe SSD)	4*2.5" SATA SSD	Supports NVMe and 3.5-inch LFF drives.
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10	RAID 0/1/10 Optional: RAID 5/50/6/60	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60	RAID 0	No obvious advantage
PCIe expansion	3 or 4*PCIe x16 LP	4*PCIe x16 LP	1*PCIe x16 + 1*PCIe x8	4*PCIe x16 LP	4*PCIe x16 (2*PCIe x8 and hard drives are optional)	4*PCIe x16 LP	No obvious advantage
On-board network adapter	/	4*10GE	2*GE/10GE electrical port	4*GE	2*GE/10GE electrical port	2*10GE	G530 V2 dual nodes provide 4 x 10GE optical ports.
Power supply	4*2200 W hot swap, 2+2	4*2200 W hot swap, 2+2	4*2000 W hot swap, 2+2	External power shelf, 6*2650W hot swap, 3+3	4*2000 W hot swap, 2+2	4*1600 W hot swap, 2+2, front-access	High power supported and can evolve to 3000 W.
Fan	6 hot swap, 5+1	6 hot swap, 5+1	The built-in fans do not support hot swap. When the fan fails, heat dissipation is affected.	8 hot-swappable	8 hot-swappable, 7+1	4 fans; the built-in fans do not support hot swap.	The fan modules are hot-swappable, which supports easy maintenance and N+1 redundancy.
Temperature	5°C to 35°C	5°C to 35°C	10°C to 35°C	10°C to 35°C, 5*P100 25°C	10°C to 35°C	10°C to 35°C	No restriction for full configuration.

Parameter	G5500-G560	G5500-G530 V2	SuperMicro 2028GR-TR	SuperMicro 1028GQ-TRT	SuperMicro 1028GQ-TXRT	Dell C4130 PCIe	Dell C4130 NVLink	Advantage
Dimensions (H x W x D)	4U (175*447*790)	4U (175*447*790)	2U (89*437*787)	1U (43*437*894)	1U (43*437*894)	1U (43*439*885)	1U (43*439*885)	There is no advantage. The competitor density is high, but the application is limited.
Architecture	Modular	Modular, dual-node	Integrated	Integrated	Integrated	Integrated	Integrated	Structured design, powerful upgrade and evolution capability, and convenient maintenance of components
CPU+DIMM	2S + 24DIMM	Scalable, 4S + 48DIMM	2S + 16DIMM	2S + 16DIMM	2S + 16DIMM	2S + 16DIMM	2S + 16DIMM	The G530 V2 supports Scalable Processors and a single CPU supports 24 DIMMs.
Maximum CPU power	300 W	205 W	145 W	145 W	145 W	145 W	145 W	The CPU uses independent air channels and has strong evolution capability.
GPU	8*P100/P40	32*P4	4*P100/P40	4*P100/P40	4*P100 NVLink	4*P100/P40	4*P100 NVLink	G530 V2 dual-node 4 32*P
Maximum GPU power	350 W	350 W	300 W	250 W	300 W	300 W	300 W	High power consumption GPU with strong scalability
CPU:GPU topology	1:8/1:4 orchestratable	1:16(P4)	1:4	1:4	1:4	1:4/1:2	1:4	G560 full-width servers support flexible orchestration via software.
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	4*2.5" NVMe/SAS /SATA SSD/HDD or 8*M.2	10*2.5" SATA SSD/HDD	4*2.5" SATA SSD/HDD, two of which are built-in	2*2.5" SAS/SATA SSD/HDD, 2*2.5" hard drives are built-in	2*1.8" SATA SSD Configured alternatively with PSU2: 4*2.5" SAS/SATA SSD/HDD	2*1.8" SATA SSD Configured alternatively with PSU2: 4*2.5" SAS/SATA SSD/HDD	Supports NVMe and 3.5-inch LFF drives.
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10	RAID 0/1/10/5 (PCH)	RAID 0/1/10/5 (PCH)	RAID 0/1/10/5/50/6/60	Pluggable RAID controller card (occupying the PCIe expansion slot)	Pluggable RAID controller card (occupying the PCIe expansion slot)	Independent RAID controller card with full functions
PCIe expansion	3 or 4*PCIe x16	4*PCIe x16	1*PCIe x8	2*PCIe x8 LP slots	3*PCIe x16, 1*PCIe x8 LP slot	2*PCIe x16	2*PCIe x16	Supports 4*PCIe x16 cards
On-board network adapter	/	4*10GE	2*GE/10GE	2*10GE	2*10GE	2*GE	2*GE	G530 V2 dual nodes provide 4 x 10GE optical ports.
Power supply	4*2200 W hot swap, 2+2	4*2200 W hot swap, 2+2	2*2000 W hot swap, 2+2	2*2000 W hot swap, 1+1	2*2000 W hot swap, 1+1	2*2000 W hot swap, 1+1	2*2000 W hot swap, 1+1	High power supported and can evolve to 3000 W.
Fan	6 hot swap, 5+1	6 hot swap, 5+1	Non-hot swap (5 fans)	Non-hot swap (11 fans)	Non-hot swap (9 fans)	The built-in fans do not support hot swap. When the fan fails, heat dissipation is affected.	The built-in fans do not support hot swap. When the fan fails, heat dissipation is affected.	The fan modules are hot-swappable, which supports easy maintenance and N+1 redundancy.
Temperature	5°C to 35°C	5°C to 35°C	10°C to 35°C	10°C to 35°C	10°C to 30°C	10°C to 25°C	10°C to 25°C	The temperature range is wide, and no restriction for full configuration

Models of Chinese Vendors

8*GPU PCIe



4U-8G Huawei
G5500-G560



3U-8G Asus
ESC8000 G3



3U-8G ASRock
3U8G C612/V



2U-8G Inspur
NF5288M5(AGX-2)

8*GPU NVLink



2U-8G Inspur
NF5288M5(AGX-2)

4*GPU PCIe

No detailed information is available, which will be supplemented later.

4*GPU NVLink

No detailed information is available, which will be supplemented later.

Parameter	G5500-G560	G5500-G530 V2	Asus ESC8000 G3	ASRock 3U8G-C612/V	Inspur AGX-2	Advantage
Dimensions (H x W x D)	4U (175*447*790)	4U (175*447*790)	3U (130.6*447*759)	3U	2U	There is no advantage, and the 3 U density is high.
Architecture	Modular	Modular, dual-node	Integrated	Integrated	Integrated	Structured design, powerful upgrade and evolution capability, and convenient maintenance of components
CPU+DIMM	2S + 24DIMM	Skylake, 4S + 48DIMM	2S + 24DIMM	2S + 16DIMM	2S + 16DIMM	G530 V2 dual nodes support Skylake and 4S + 48 DIMMs.
Maximum CPU power	300 W	205 W	145 W	145 W	145 W	The CPU uses independent air channels and has strong evolution capability.
GPU	8*P100/P40/P4	8*P100/P40/P4 or 32*P4	8*P100/P40/P4	8*P100/P40/P4	8*P100/V100 NVLink Or 8*P100/P40 PCIe GP U	G530 V2 dual nodes support 32*P4.
Maximum GPU power	350 W	350 W	300 W	300 W	300 W	High power consumption GPU with strong scalability
CPU:GPU topology	1:8/1:4 orchestratable	1:4/1:2/1:16 (P4)	1:8 (1:4)	1:8	1:8	G560 full-width servers support flexible orchestration via software.
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	4*2.5" NVMe/SAS /SATA SSD/HDD or 8*M.2	6*2.5" SATA SSD/HDD. SAS SSD/HDD is supported when the PCIe RAID controller card is configured.	6*2.5" SATA SSD/HDD	8*2.5" NVMe/SAS/SATA SSD/HDD	Supports NVMe and 3.5-inch LFF drives.
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60	PCH: RAID 0/1/10/5 When the PCIe RAID controller card is used: RAID 0/1/10/5/50/6/60	? (No RAID controller card?)	RAID 0/1/10/5/50/6/60	No obvious advantage
PCIe expansion	3 or 4*PCIe x16	4*PCIe x16	1*PCIe x8 LP card; when 8*dual-slot cards are not fully configured, 1*PCIe x8 FHFL card can be configured.	1*PCIe x16 or 2*GE/10GE	NVLink: 4*PCIe x16 LP, 2*PCIe x8 LP (blocking air ventilation) PCIe: 2*PCIe x8 LP (blocking air ventilation)	Supports 4*PCIe x16 cards
On-board network adapter	/	4*10GE	/	/	4*10GE	G530 V2 dual nodes provide 4 x 10GE optical ports.
Power supply	4*2200 W hot swap, 2+2	4*2200 W hot swap, 2+2	3*1600 W hot swap, 2+1	4*1200 W hot swap, 3+1	2*3000 W hot swap, 1+1, front-access power supply	High power supported and can evolve to 3000 W.
Fan	6 hot swap, 5+1	6 hot swap, 5+1	Non-hot swap (14 fans)	Non-hot swap (8 fans)	Non-hot swap (5 fans)	The fan modules are hot-swappable, which supports easy maintenance and N+1 redundancy.
Temperature	5°C to 35°C	5°C to 35°C	10°C to 35°C	10°C to 35°C	5°C to 35°C	No obvious advantage

OEM Relationship Between International/Chinese Vendor Products

Vendor	Model	OEM Vendor	OEM Model	Height		
Microway	?	SuperMicro	4028GR TXRT	4U		
Microway	NumberSmasher	SuperMicro	1028GQ-TXRT	1U		
Microway	Octoputer 4U	SuperMicro	4028GR-TR2	4U		
Sugon	W740-G20	SuperMicro	2028GR-TR	2U		
Sugon	W780-G20	SuperMicro	4028GR-TR2	4U		

Contents

1 Introduction to the G5500

2 Comparison Between G5500 and Competitor Specifications

3 Advantages and Disadvantages of G5500 and Competitor Specifications (International)

4 Advantages and Disadvantages of G5500 and Competitor Specifications (China)

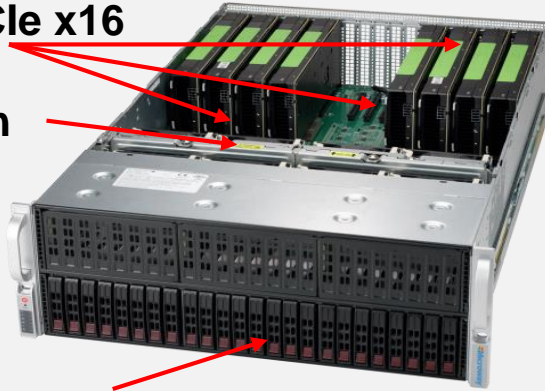
5 How to Defend

G5500 vs SuperMicro SYS-4028GR-TRT2

SuperMicro: SYS-4028GR-TRT2

GPU: PCIe x16

8 built-in fans



24*SAS/SATA



2*10GE/GE
electrical port

4*2000 W
power
supply

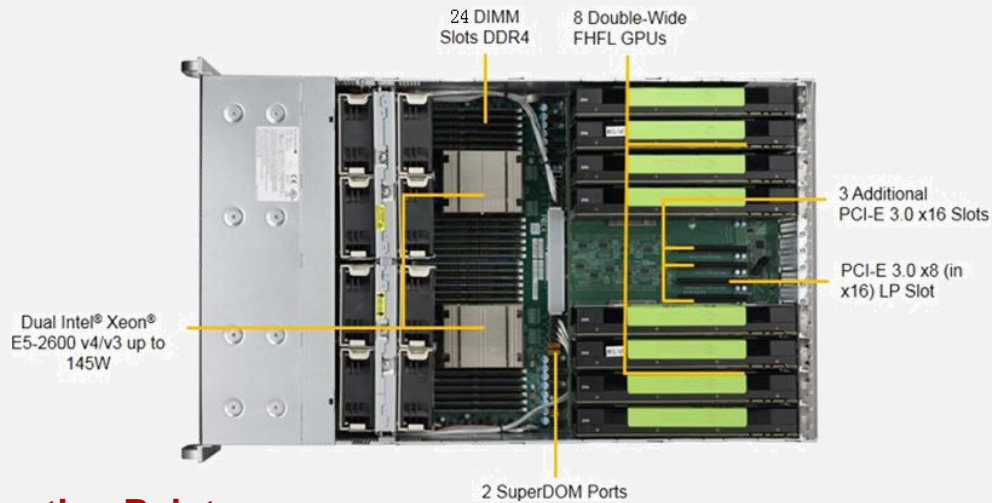
Specifications:

1. 4U/8GPU
2. 2S + 24DIMM, 3*PCIe x16 + 1*PCIe x8, and 24*SAS/SATA 2.5" hard drives
3. 8*92 built-in fans
4. 4*2000 W power supply, 2+2
5. Ambient temperature: 10° C to 35° C

Note:

1. At GTC in September, SuperMicro exhibited the 4029 model that supports the Scalable Processors, but so far the official product specifications are not available.
2. For CPU and GPU topologies, generally the maximum configuration is 8 GPUs.

SuperMicro: SYS-4028GR-TRT2



Beating Points

1. The chassis uses an integrated design. The CPUs and GPUs cannot evolve independently.
2. The CPU:GPU topology is fixed and cannot be flexibly configured.
3. NVMe SSD is not supported.
4. The fans are built-in and require opening the chassis for maintenance.
5. The DIMMs locate below the fans; maintaining the DIMMs requires removing the fans first, which is inconvenient.
6. The power supply capability is limited and cannot be evolved.

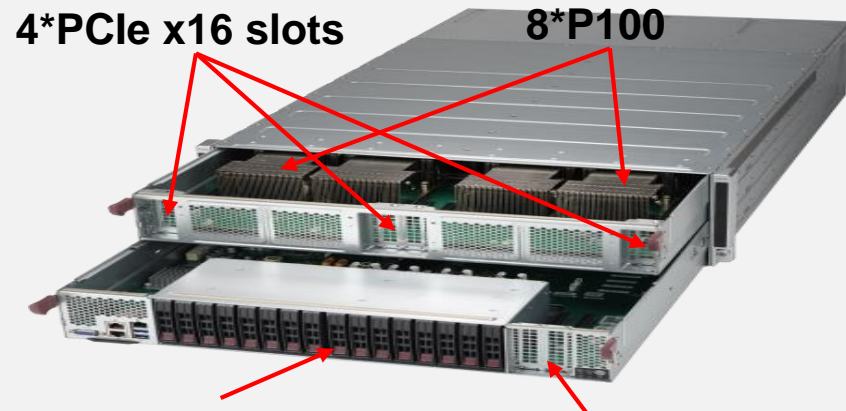
Parameter	G5500-G560	SuperMicro 4028GR-TR2
Dimensions (H x W x D)	4U (175*447*790)	4U (178*437*737)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 24DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	8*P100/P40/P4
Maximum GPU power	350 W	250 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:8
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	24*2.5" SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10 Optional: RAID 5/50/6/60
PCIe expansion	4*PCIe x16	3*PCIe x16 + 1*PCIe x8
On-board network adapter	/	2 x 10GE/GE
Power supply	4400 W hot swap, 2+2	4000 W hot swap, 2+2
Fan	6 hot swap, 5+1	Non-hot swap (hot swap with opening the chassis?)
Temperature	5°C to 35°C	10°C to 35°C

Note:

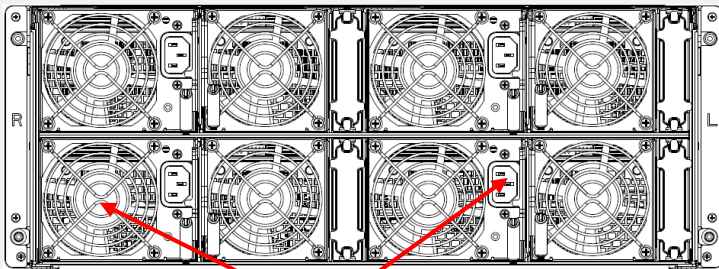
SuperMicro declares that the fans are hot-swappable, but this is generally not possible due to the integrated chassis design. Therefore, SuperMicro's declaration can be regarded as that the fans are hot-swappable without powering off the system, but the chassis still needs to be opened.

G5500 vs SuperMicro SYS-4028GR-TXRT

SuperMicro: SYS-4028GR-TXRT (NVLink)



2*PCle x8 slots



Specifications:

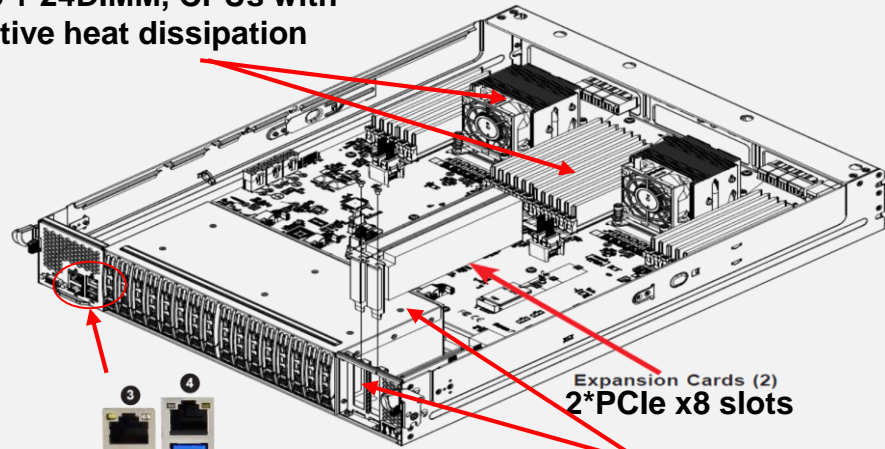
1. 4U/8GPU, supporting NVLink
2. 2S + 24DIMM, 4*PCle x16+2*PCle x8 (alternative support), 16*2.5" SAS/SATA HDDs (eight of them support NVMe) (CPUs use active heat dissipation)
3. 8*80 fans, 4 of them with power supply
4. 4*2000 W power supply, 2+2
5. Ambient temperature: 10° C to 35° C

Note:

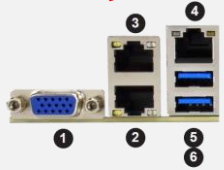
At GTC in September, SuperMicro exhibited the 4029 model that supports the Scalable Processors, but so far the official product specifications are not available.

SuperMicro: SYS-4028GR-TXRT (NVLink)

2S + 24DIMM, CPUs with active heat dissipation

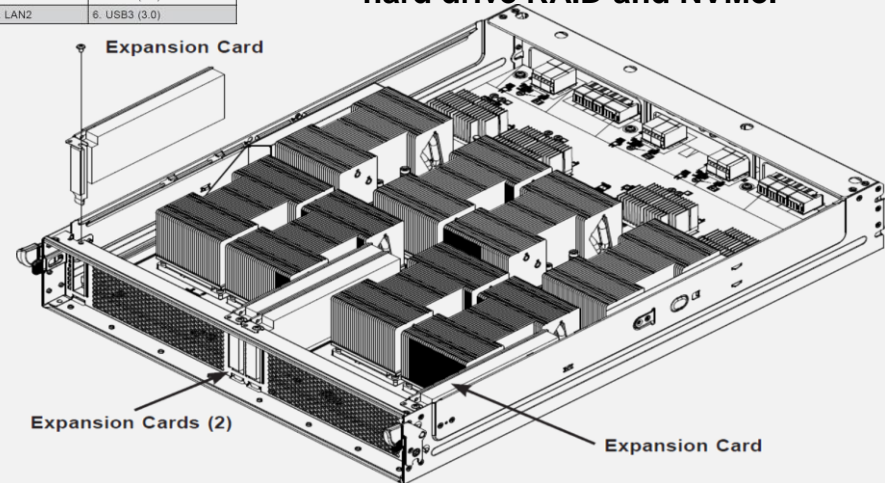


Expansion Cards (2)
2*PCIe x8 slots



1. VGA Connector	4. Dedicated IPMI LAN Port
2. LAN1	5. USB2 (3.0)
3. LAN2	6. USB3 (3.0)

The CPUs provide 2*PCIe x8, which are alternatively supported with the hard drive RAID and NVMe.



Expansion Cards (2)

Expansion Card

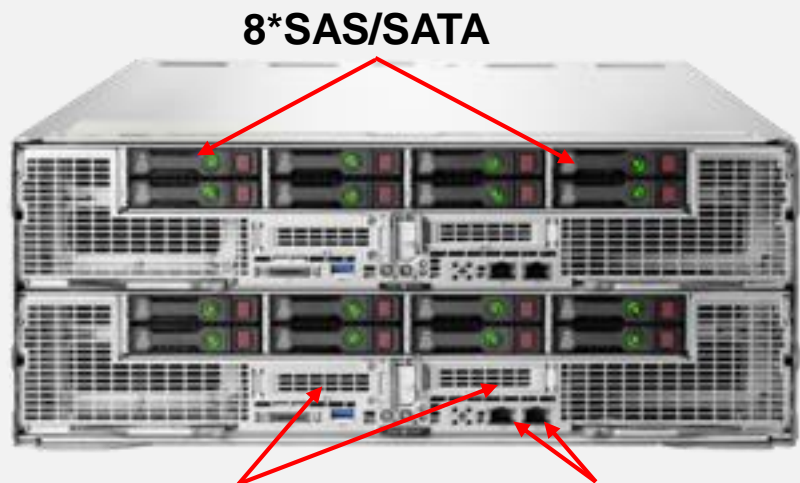
Parameter	G5500-G560	SuperMicro 4028GR-TXRT
Dimensions (H x W x D)	4U (175*447*790)	4U
Architecture	Modular	Modular
CPU	2S + 24DIMM	2S + 24DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	8*P100 NVLink
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:8
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	16*2.5" SSD/HDD (8*NVMe SSD)
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60
PCIe expansion	4*PCIe x16	4*PCIe x16 (2*PCIe x8 and hard drive RAID/NVMe alternatively supported)
On-board network adapter		
Power supply	4400 W hot swap, 2+2	4000 W hot swap, 2+2
Fan	6 hot swap, 5+1	8 hot swap, 7+1
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

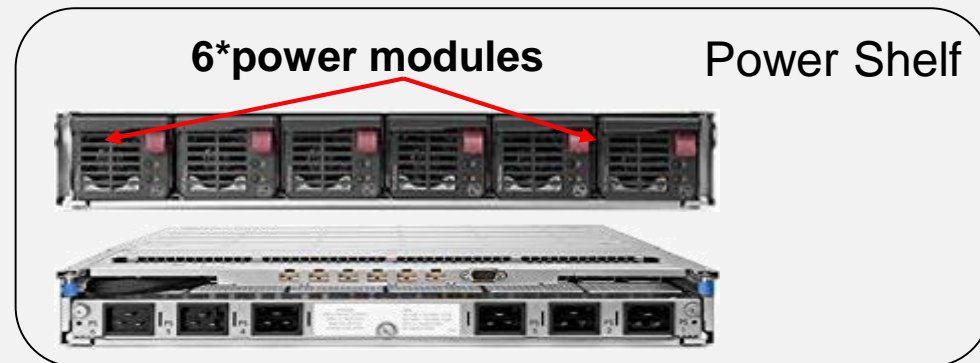
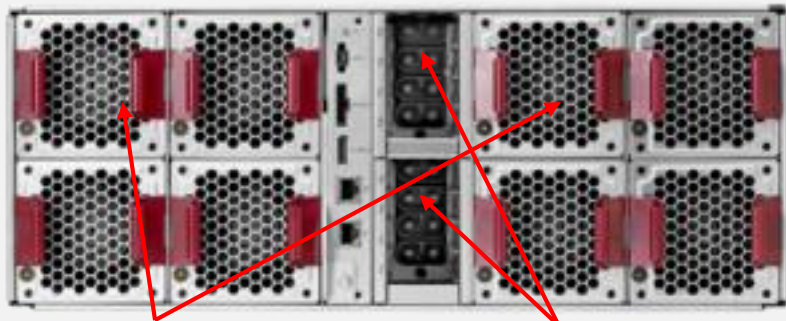
- The 8 NVMe SSDs are restricted by the PCIe x8 links, and only two drives support line rate.
- The power supply capability is limited and cannot be evolved.
- Active heat dissipation for the CPUs (many failure points)
- The power module is integrated into the fan and has to be inserted and removed at the same time during maintenance, which makes maintenance inconvenient.

G5500 vs HPE Apollo 6500+Power Shelf

HPE: Apollo 6500+Power Shelf

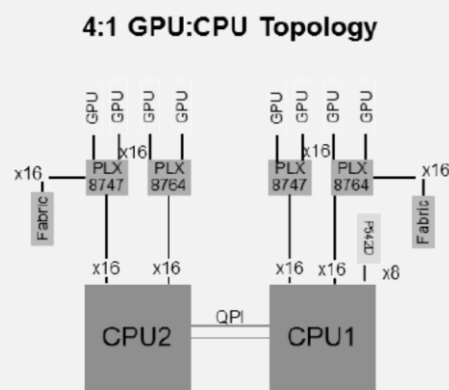
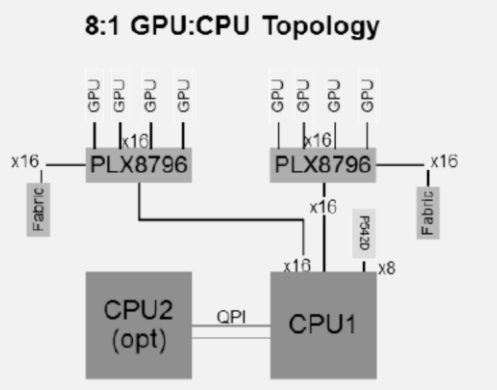
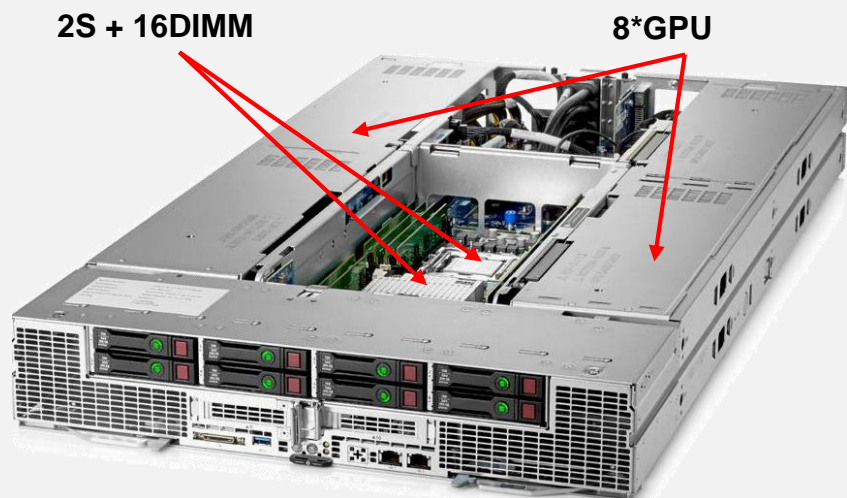


2*PCIe x16 2*10GE/GE electrical port



1. 5.5U/16GPU, including 4U server (two 2U-8GPU nodes) and 1.5U power shelf
2. 2 nodes, each node 2S + 16DIMM, 2*PCIe x16, 8*2.5" SAS/SATA HDDs
3. 8*80 fans
4. The external 1.5U power shelf supports 6*2650W power modules, 3+3 redundancy.
5. Ambient temperature: 10° C to 35° C; **lower specifications with configuration larger than 5 GPUs (P100; supports only 25° C)**

HPE: Apollo 6500+Power Shelf



Parameter	G5500-G560	Apollo 6500
Dimensions (H x W x D)	4U (175*447*790)	4U (177*448*960) + 1.5U (65*448*784)
Architecture	Modular	Modular, dual nodes
CPU	2S + 24DIMM	Single node: 2S + 16DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	Single node: 8*P100/P40
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	Single node: 1:8/1:4 (optional)
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	Single node: 8*2.5" SAS/SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60
PCIe expansion	4*PCIe x16	Single node: 2*PCIe x16 LP + 1*PCIe x8 LP (RAID controller card)
On-board network adapter		Single node: 2 x GE
Power supply	4400 W hot swap, 2+2	External power shelf, 6*2650W hot swap, 3+3
Fan	6 hot swap, 5+1	8 hot swap
Temperature	5°C to 35°C	10°C to 35°C. The specifications lower to 25°C when more than 5 GPUs (P100) are configured.

Beating Points

- The chassis uses a modular design. A power shelf is required.
- In the 1:8 and 1:4 topologies, P2P cannot be implemented between 8 cards and 4 cards.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported.
- When more than 5 P100 GPUs are configured, the maximum temperature supported is only 25°C, which restricts its application.

G5500 vs NVIDIA DGX-1

NVIDIA: DGX-1

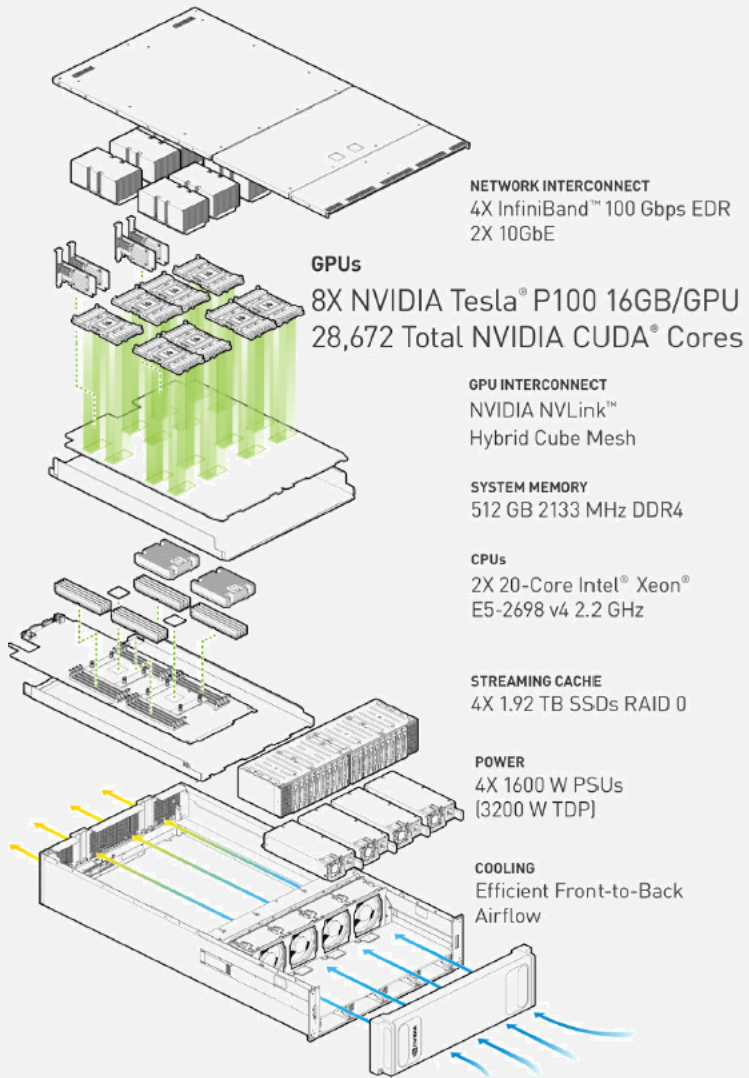


1. 3U/8GPU, supporting NVLink
2. 2S + 16DIMM
3. 4*PCIe x16 LP, 4*2.5" SATA SSD
4. 4*92 fans
5. 4*1600 W power supply, 2+2
6. Ambient temperature: 10° C to 35° C

Note:

1. The power module is removed and inserted from the front of the chassis, and the power cable is connected to the rear of the power module.

NVIDIA: DGX-1



Parameter	G5500-G560	NVIDIA DGX-1
Dimensions (H x W x D)	4U (175*447*790)	3U (131*444*866)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	145 W	145 W?
GPU	8*P100/P40/P4	8*P100/V100 NVLink
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:8
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	4*2.5" SATA SSD
RAID	RAID 0/1/10/5/50/6/60	Raid 0/1/10
PCIe expansion	4*PCIe x16 LP	4*PCIe x16 LP
On-board network adapter	/	2*10GE
Power supply	4400 W hot swap, 2+2	4*1600 W hot swap, 2+2 (power modules locate at the front, cables routed to the rear)
Fan	6 hot swap, 5+1	4 built-in fans; the fans do not support hot
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported.
- The fans are built-in and require opening the chassis for maintenance.
- The power modules locate at the front of the chassis and are routed to the rear for interface output.

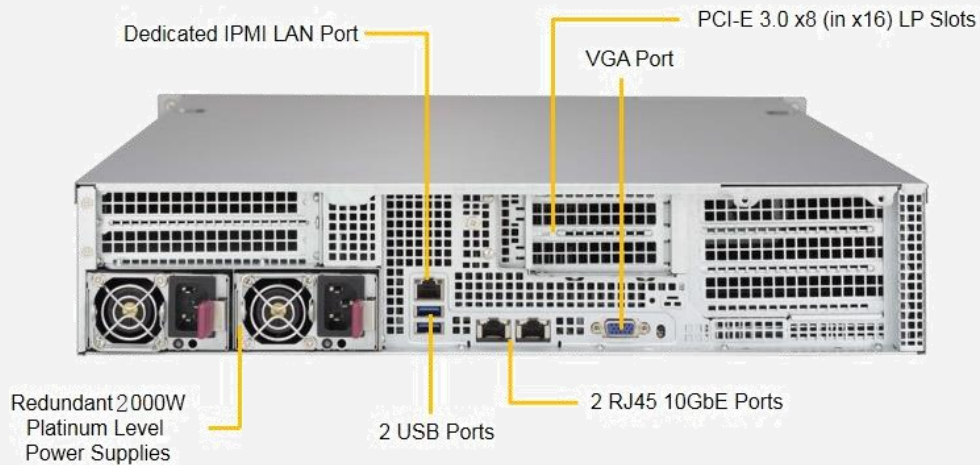
G5500 vs SuperMicro SYS-2028GR-TRHT

SuperMicro: SYS-2028GR-TRHT



10x 2.5" Hot-swap
SATA3 Drive Bays

1. 2U/4GPU (maximum 6GPU)
2. 2S + 16DIMM, 1*PCIe x8, 10*2.5" SATA HDD
3. 5*80 fans, built-in
4. 2*2000 W power supply, 1+1
5. Ambient temperature: 10° C to 35° C



Dedicated IPMI LAN Port

VGA Port

PCI-E 3.0 x8 (in x16) LP Slots

Redundant 2000W
Platinum Level
Power Supplies

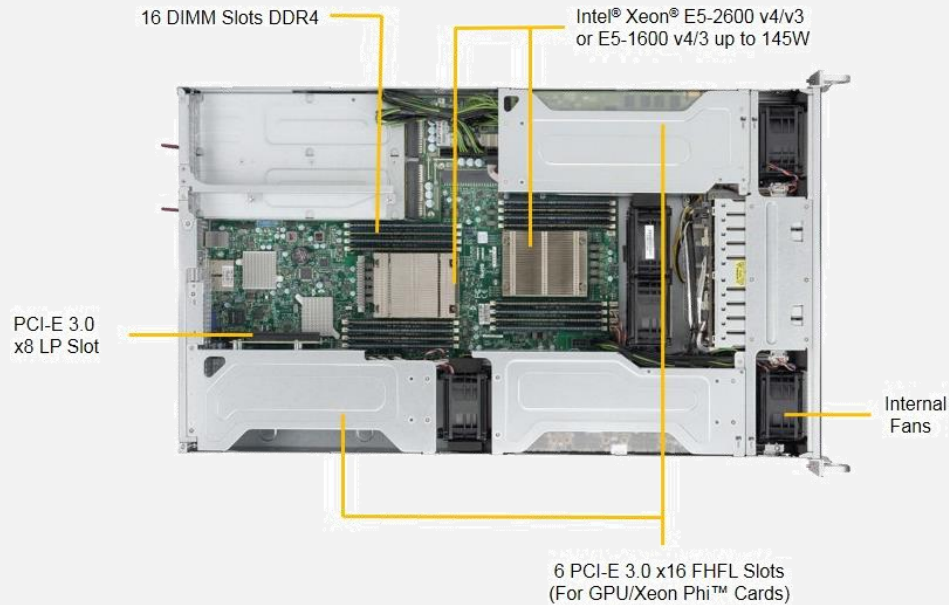
2 USB Ports

2 RJ45 10GbE Ports

Note:

1. The TRHT model supports 6 GPUs. In practice, however, 4 GPUs are used.
2. As shown in the figure, there may be two 40 fans above the power supply, which is not described in other documents; currently, these two fans are not included for analysis.

SuperMicro: SYS-2028GR-TRHT



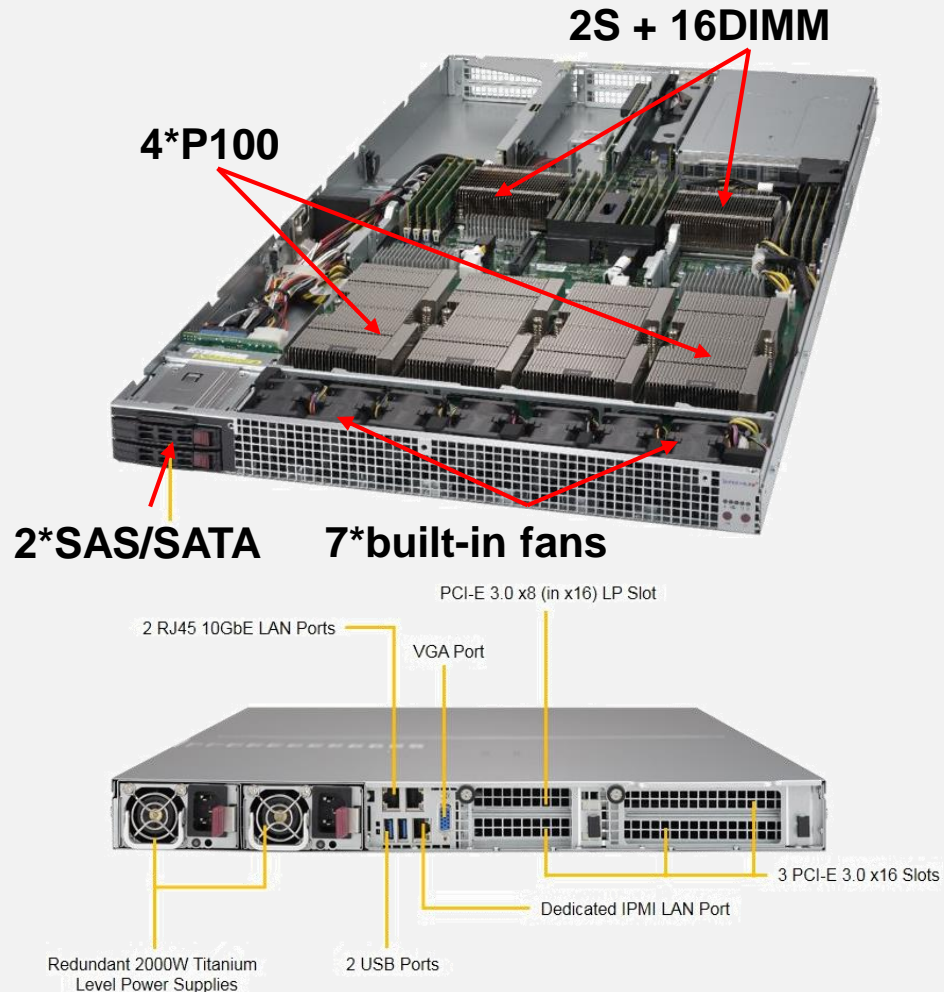
Parameter	G5500-G560	SuperMicro 2028GR-TRHT
Dimensions (H x W x D)	4U (175*447*790)	2U (89*437*787)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	4/6*P100/P40
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:4
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	10*2.5" SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5 (PCH)
PCIe expansion	4*PCIe x16	1*PCIe x8 LP
On-board network adapter		2 x GE/10GE
Power supply	4*2200 W hot swap, 2+2	2000 W hot swap, 1+1
Fan	6 hot swap, 5+1	Non-hot swap (5 fans)
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU supports only up to 145 W, which restricts support for the Scalable Processors in the future.
- The CPU:GPU topology is fixed and cannot be flexibly configured.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported.
- The fans are built-in and require opening the chassis for maintenance.

G5500 vs SuperMicro SYS-1028GQ-TXRT

SuperMicro: SYS-1028GQ-TXRT (NVLink)

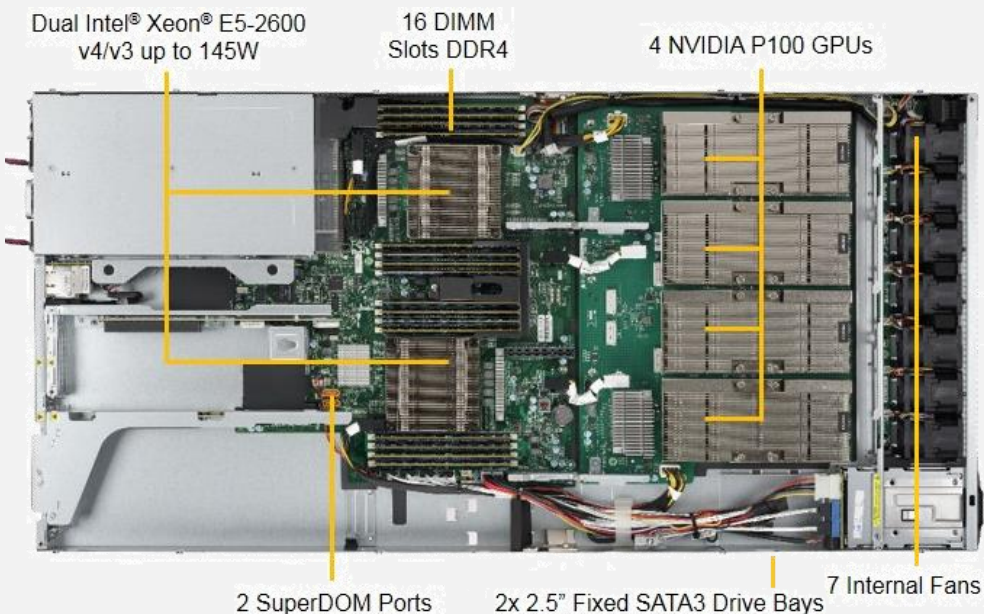


1. 2U/4GPU, supporting NVLink
2. 2S + 16DIMM, 3*PCIe x16, 1*PCIe x8 LP slot
3. 2*2.5" SAS/SATA HDD, 2*2.5" built-in hard drives
4. 7*1U fans
5. 2*2000 W power supply, 1+1
6. Ambient temperature: 10° C to 30° C

Note:

At GTC in September, SuperMicro exhibited the 1029 model that supports the Scalable Processors, but so far the official product specifications are not available.

SuperMicro: SYS-1028GQ-TXRT (NVLink)



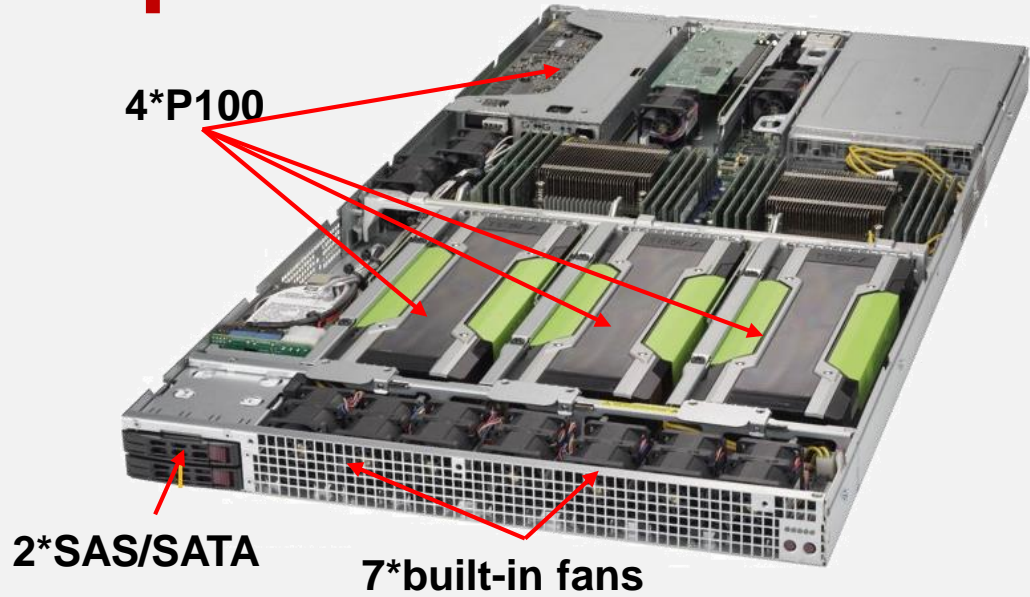
Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU supports only up to 145 W, which restricts support for the Scalable Processors in the future.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported.
- The fans are built-in and require opening the chassis for maintenance.
- **The supported temperature is only 10°C to 30°C, and the application is restricted.**
- The internal GPU cards, PCIe expansion slots, and extended hard drives need to be connected, which makes maintenance complicated.

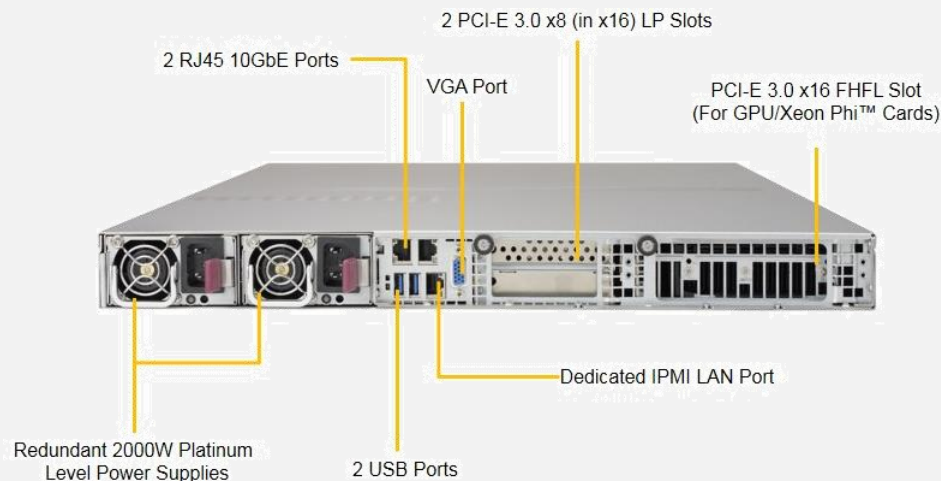
Parameter	G5500-G560	SuperMicro 4028GR-TXRT
Dimensions (H x W x D)	4U (175*447*790)	1U (43*437*894)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	4*P100 NVLink
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:4
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	2*2.5" SAS/SATA SSD/HDD, 2*2.5" built-in hard drives
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60
PCIe expansion	4*PCIe x16	3*PCIe x16, 1*PCIe x8 LP slot
On-board network adapter	/	2*10GE
Power supply	4*2200 W hot swap, 2+2	2000 W hot swap, 1+1
Fan	6 hot swap, 5+1	7 built-in fans
Temperature	5°C to 35°C	10°C to 30°C

G5500 vs SuperMicro SYS-1028GQ-TRT

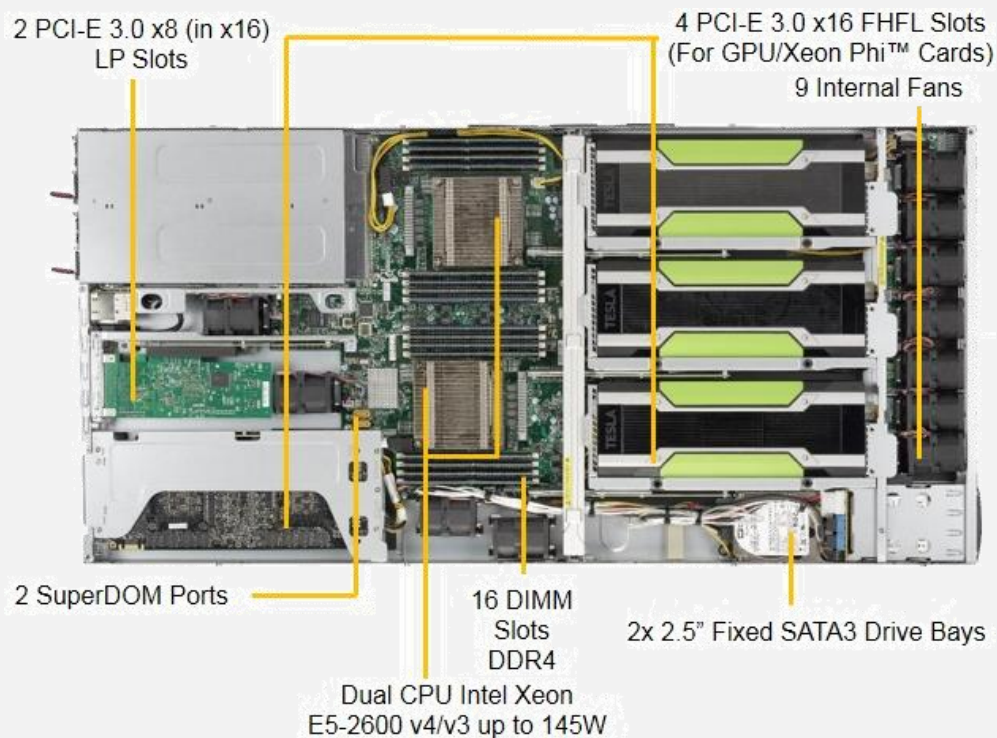
SuperMicro: SYS-1028GQ-TRT



1. 2S + 16DIMM
2. 2*PCIe x8 LP slots
3. 2*2.5" SAS/SATA HDD, 2*2.5" built-in hard drives
4. 7*1U fans
5. 2*2000 W power supply, 1+1
6. Ambient temperature: 10° C to 35° C



SuperMicro: SYS-1028GQ-TRT



Parameter	G5500-G560	SuperMicro 1028GQ-TRT
Dimensions (H x W x D)	4U (175*447*790)	1U (437*894*43)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	145 W	145 W
GPU	8*P100/P40/P4	4*P100/P40/P4
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:4
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	4*2.5" SATA SSD/HDD, two of which are built-in
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5 (PCH)
PCIe expansion	4*PCIe x16	2*PCIe x8 LP slots
On-board network adapter	/	2*10GE
Power supply	4*2200 W hot swap, 2+2	2000 W hot swap, 1+1
Fan	6 hot swap, 5+1	9 built-in fans
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU supports only up to 145 W, which restricts support for the Scalable Processors in the future.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported.
- The fans are built-in and require opening the chassis for maintenance.
- The internal GPU cards, PCIe expansion slots, and extended hard drives need to be connected, which makes maintenance complicated.

G5500 vs Dell PowerEdge C4130

Dell: PowerEdge C4130



2*PCIe x8/x16

2*1.8 SATA
SSD

2*power supply, PSU2/hard
drive tray alternatively
supported



2*GE

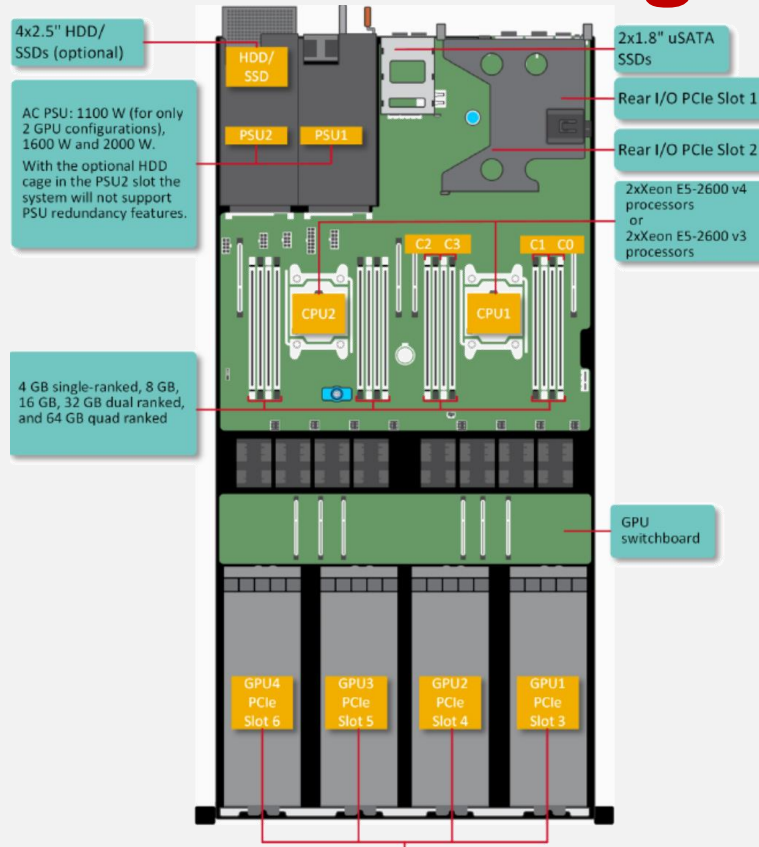


1. 1U/4GPU (PCIe and NVLink, with different extension boards)
2. 2S + 16DIMM
3. 2*1.8" SATA SSD. The 4*2.5" SAS/SATA hard drives can be optionally configured.
4. 2*PCIe x16
5. 8 built-in fans
6. 2*2000 W power supply, 1+1 (without backup when the hard drive tray is configured)
7. Ambient temperature: 10° C to 25° C

Note:

The C4130 supports two modes: PCIe and NVLink. For the NVLink mode, only similar description is found in user manuals and other websites. The description is not found on the Dell official website.

Dell: PowerEdge C4130



Single processor systems:

- Config. A: 4xGPUs and switchboard, x8 in slot 1, x16 in slot 2
- Config. E: 2xGPUs, no switch board, x8 in slot 1, slot 2 not supported

Dual processor systems:

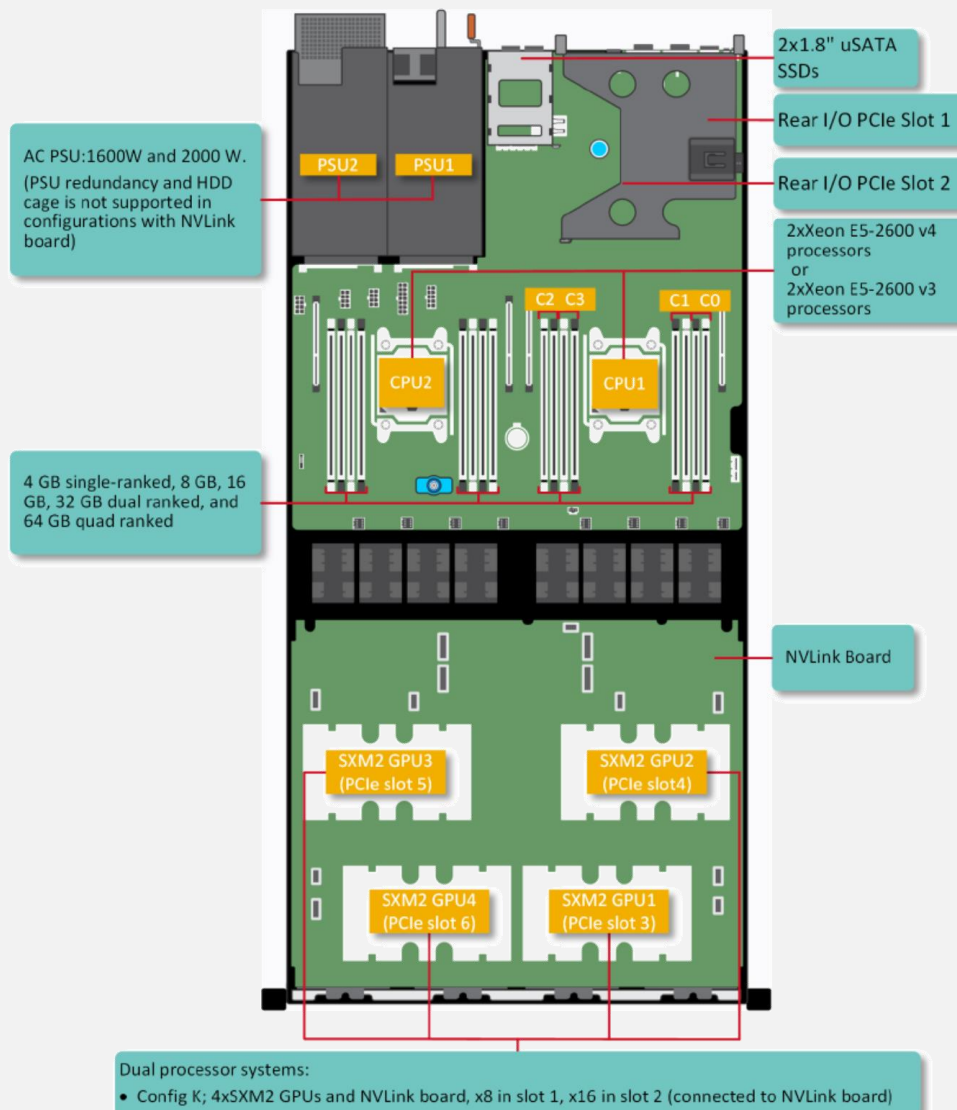
- Config. B: 4xGPUs and switchboard, x8 in slot 1, x16 in slot 2
- Config. C: 4xGPUs, no switchboard, x8 in slot 1, x8 in slot 2
- Config. D: 2xGPUs, no switchboard, x16 in slot 1, x16 in slot 2
- Config. G: 4xGPUs and switchboard with dual GPU virtual mode, x16 in slot 1, x16 in slot 2
- Config. F: 2xGPUs, no switchboard, x8 in slot 1, x8 in slot 2
- Config. H: 3xGPUs no switchboard, x8 in slot 1, x8 in slot 2
- Config. I: 3xGPUs no switchboard, x8 in slot 1, slot 2 not supported

Parameter	G5500-G560	PowerEdge C4130
Dimensions (H x W x D)	4U (175*447*790)	1U (43*439*885)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	300 W	145 W
GPU	8*P100/P40	4*P100/P40
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:4/1:2
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	2*1.8" SATA SSD It can be configured with the PSU2. 4*2.5" SAS/SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	Pluggable Raid card (occupying the extended slot of PCIe)
PCIe expansion	4*PCIe x16	2*PCIe x16
On-board network adapter	/	2*GE
Power supply	4*2200 W hot swap, 2+2	2*2000 W hot swap, 1+1
Fan	6 hot swap, 5+1	The built-in does not support hot swap. When the fan fails, heat dissipation is affected.
Temperature	5°C to 35°C	10°C to 25°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU supports only up to 145 W, which restricts support for the Scalable Processors in the future.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported. (Hard drives and PSU2 alternatively supported)
- The fans are built-in and require opening the chassis for maintenance.
- **The supported temperature is only 10°C to 25°C, and the application is restricted.**
- The internal GPU cards, PCIe expansion slots, and extended hard drives need to be connected, which makes maintenance complicated.

Dell: PowerEdge C4130



Parameter	G5500-G560	PowerEdge C4130
Dimensions (H x W x D)	4U (175*447*790)	1U (43*439*885)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	300 W	145 W
GPU	8*P100/P40	4*P100 NVLink
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:4
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	2*1.8" SATA SSD It can be configured with the PSU2. 4*2.5" SAS/SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	Pluggable Raid card (occupying the extended slot of PCIe)
PCIe expansion	4*PCIe x16	2*PCIe x16
On-board network adapter	/	2*GE
Power supply	4*2200 W hot swap, 2+2	2*2000 W hot swap, 1+1
Fan	6 hot swap, 5+1	The built-in does not support hot swap. When the fan fails, heat dissipation is affected.
Temperature	5°C to 35°C	10°C to 25°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU supports only up to 145 W, which restricts support for the Scalable Processors in the future.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported. (Hard drives and PSU2 alternatively supported)
- The fans are built-in and require opening the chassis for maintenance.
- **The supported temperature is only 10°C to 25°C, and the application is restricted.**
- The internal GPU cards, PCIe expansion slots, and extended hard drives need to be connected, which makes maintenance complicated.

Contents

1 Introduction to the G5500

2 Comparison Between G5500 and Competitor Specifications

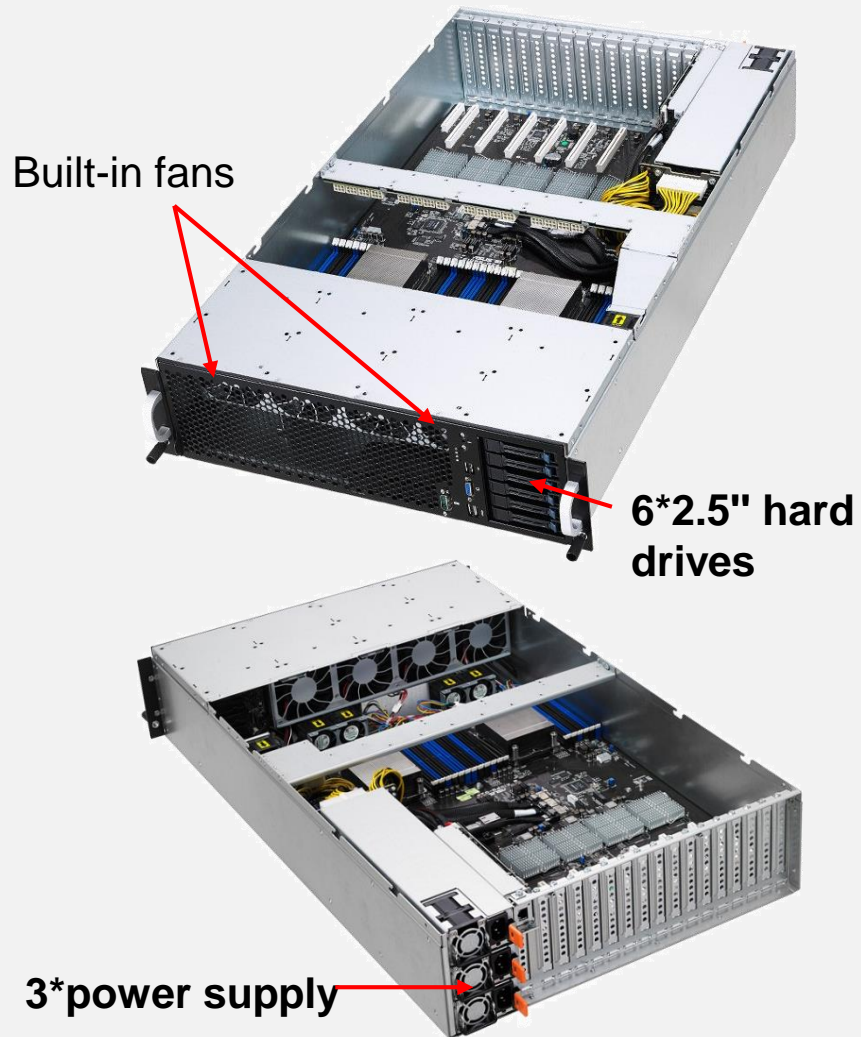
3 Advantages and Disadvantages of G5500 and Competitor Specifications (International)

4 Advantages and Disadvantages of G5500 and Competitor Specifications (China)

5 How to Defend

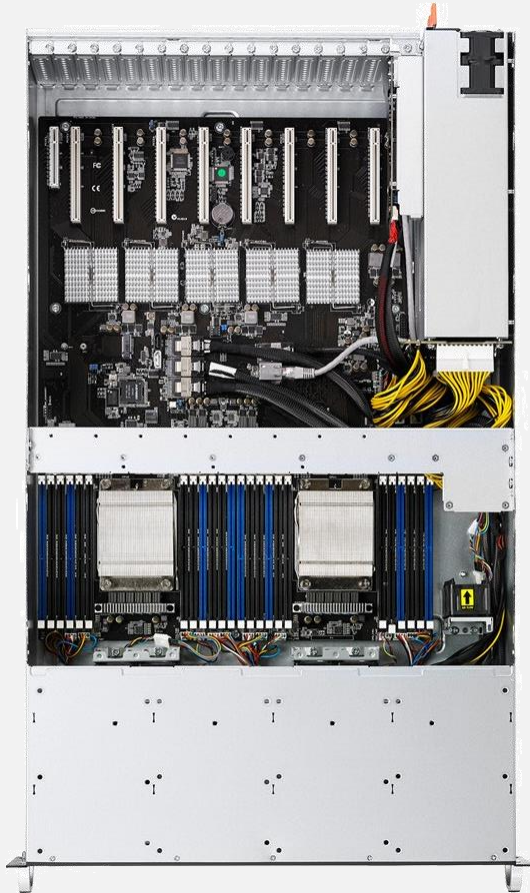
G5500 vs Asus ESC8000 G3

Asus: ESC8000 G3



1. 3U/8GPU
2. 2S + 24DIMM, 6*2.5" SAS/SATA HDDs
3. 1*PCIe x8 LP card. When the 8*dual-slot cards are not fully configured, 1*PCIe x8 FHFL card can be configured.
4. 4*80 built-in fans
5. 3*1600 W power supply, 2+1
6. Ambient temperature: 10° C to 35° C

Asus: ESC8000 G3



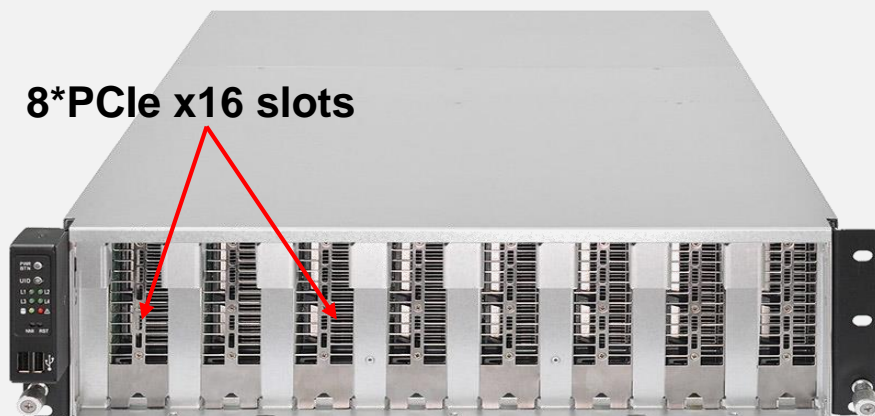
Parameter	G5500-G560	Asus ESC8000 G3
Dimensions (H x W x D)	4U (175*447*790)	3U (130.6*447*759)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 24DIMM
Maximum CPU power	300 W	145 W
GPU	8*P100/P40	8*P100/P40
Maximum GPU power	350 W	300 W
CPU:GPU topology	Flexible 1:8/1:4 Configuration	1:8 (1:4)
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	6*2.5" SATA SSD/HDD, Supports SAS SSD/HH D when the PCIe RAID controller card is configured.
RAID	RAID 0/1/10/5/50/6/60	PCH: RAID 0/1/10/5 When the PCIe RAID card is used: RAID 0/1/10/5/50/6/60
PCIe expansion	4*PCIe x16	1*PCIe x8 LP card. When the 8*dual-slot cards are not fully configured, 1*PCIe x8 FHFL card can be configured.
On-board network adapter	/	2*GE
Power supply	4*2200 W hot swap, 2+2	3*1600 W hot swap, 2+1
Fan	6 hot swap, 5+1	The built-in fans do not support hot swap. When the fan fails, heat dissipation is affected.
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU:GPU topology is fixed and cannot be flexibly configured.
- NVMe SSD is not supported. When the SAS drive is used, it needs to occupy the extended PCIe slot.
- The 14 fans are built-in, and there are two types of fans, which require opening the chassis for maintenance.
- Three 1600 W power modules do not support N+N redundancy, and the capacity expansion is limited and cannot be evolved.
- A large number of cables are connected inside the chassis, which is inconvenient for maintenance.

G5500 vs ASRock 3U8G-C612

ASRock: 3U8G-C612/V



1. 3U/8GPU
2. 2S + 16DIMM, 6*2.5" SATA SSD/HDD
3. 8*80 built-in fans
4. 4*1200 W power supply, 3+1
5. Ambient temperature: 10° C to 35° C

ASRock: 3U8G-C612/V



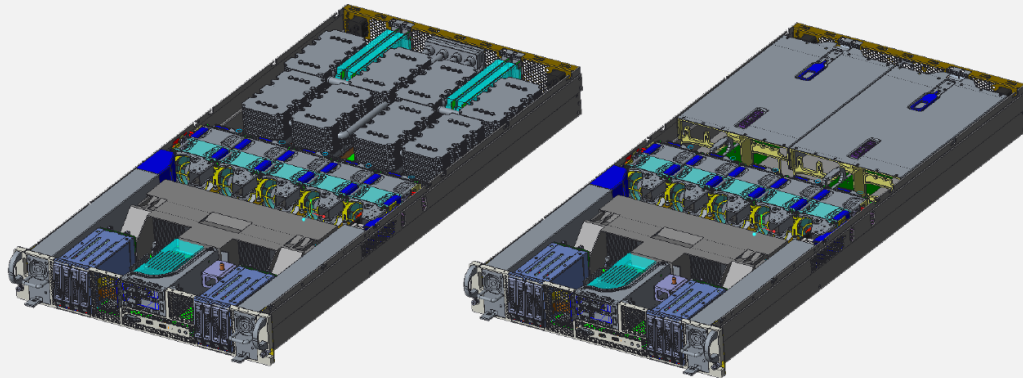
Parameter	G5500-G560	ASRock 3U8G-C612/V
Dimensions (H x W x D)	4U (175*447*790)	3U
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	300 W	145 W
GPU	8*P100/P40/P4	8*P100/P40/P4
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:8
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	6*2.5" SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	? (No RAID controller card?)
PCIe expansion	4*PCIe x16	1*PCIe x16 or 2*GE/10GE
On-board network adapter		/
Power supply	4*2200 W hot swap, 2+2	4*1200 W hot swap, 3+1
Fan	6 hot swap, 5+1	The built-in does not support hot swap. When the fan fails, heat dissipation is affected.
Temperature	5°C to 35°C	10°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- The CPU:GPU topology is fixed and cannot be flexibly configured.
- Only 16 DIMMs are supported, and the memory capacity is limited.
- NVMe SSD is not supported. (Without the RAID function)
- The fans are built-in and require opening the chassis for maintenance.
- The power capability is limited. The N+N redundancy cannot be supported when the power supply is fully loaded.

G5500 vs Inspur NF5288M5(AGX-2)

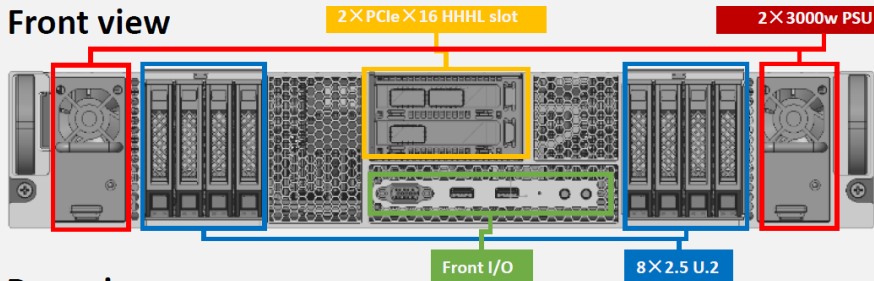
Inspur: NF5288M5(AGX-2)



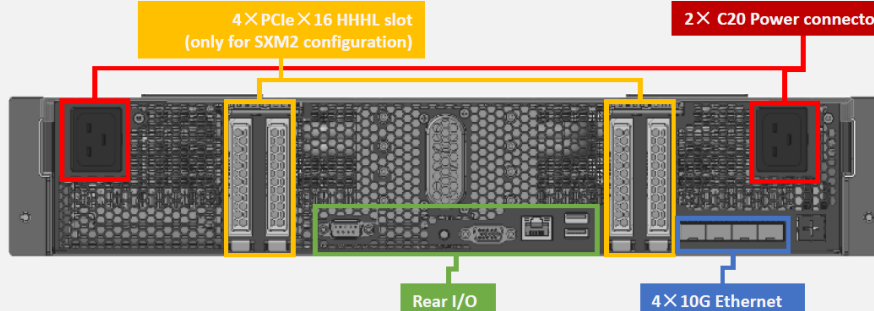
SXM2 Configuration

PCIe AIC configuration

Front view



Rear view

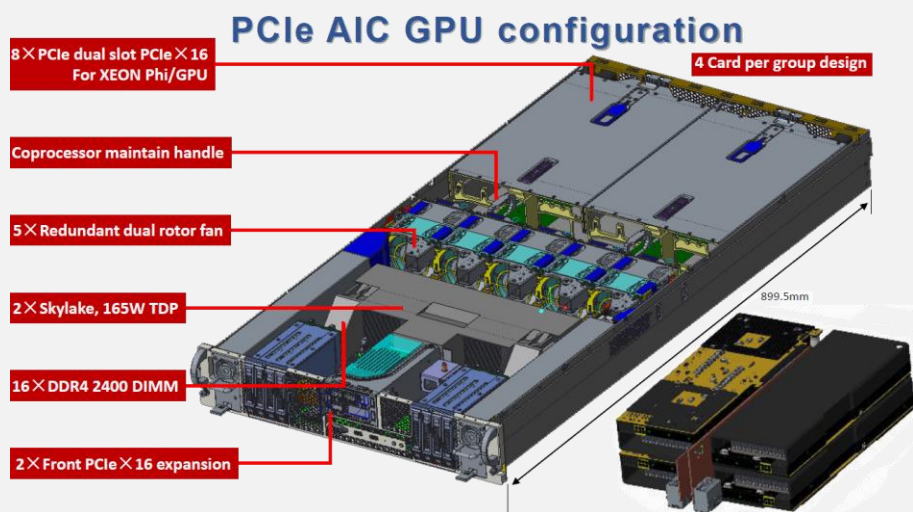
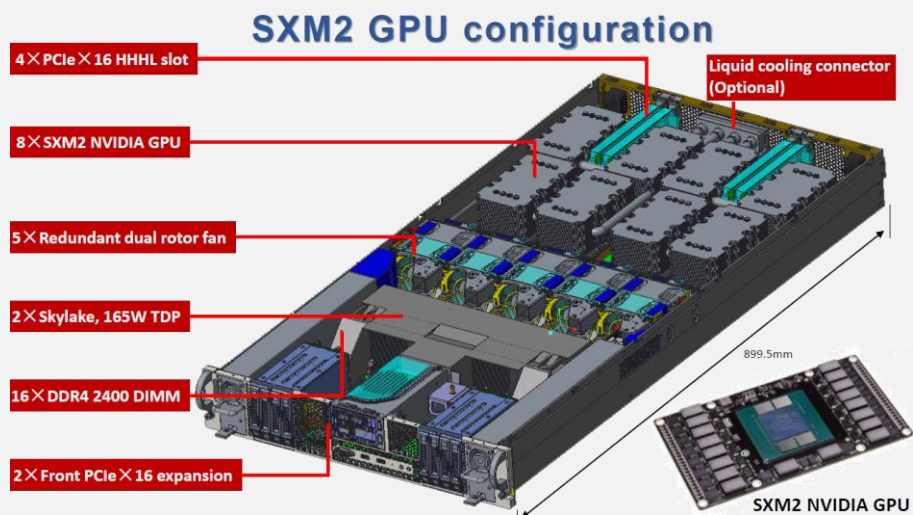


1. 2U/8GPU (PCIe and NVLink, with different extension boards)
2. 2S + 16DIMM; supports Scalable Processors and the maximum power consumption is **165 W**.
3. 6*2.5 U.2, 2*M.2 PCIe & SATA on board
4. 2*PCIe x16 LP (front panel) +4*PCIe x16 LP (provided on the rear panel when NVLink is used; none when PCIe is used)
3. 5*80 built-in fans (or 10 single-layer fans)
4. 2*3000 W power supply, 1+1
5. Ambient temperature: 5° C to 35° C

Note:

The power module is removed and inserted from the front of the chassis, and the power cable is connected to the rear of the power module.

Inspur: NF5288M5(AGX-2)



Parameter	G5500-G560	Inspur NF5288M5(AGX-2)
Dimensions (H x W x D)	4U (175*447*790)	2U (448*899.5*87.5)
Architecture	Modular	Integrated
CPU	2S + 24DIMM	2S + 16DIMM
Maximum CPU power	145 W (the model can be expanded to 205 W)	165 W
GPU	8*P100/P40	8*P100/V100 NVLink or P100/P40 PCIe
Maximum GPU power	350 W	300 W
CPU:GPU topology	1:8/1:4 flexible configuration	1:8
Hard drive	6*2.5" NVMe SSD 2*2.5" SAS/SATA HDD 8*3.5" HDD	8*2.5 NVMe/SAS/SATA SSD/HDD
RAID	RAID 0/1/10/5/50/6/60	RAID 0/1/10/5/50/6/60
PCIe expansion	4*PCIe x16 LP	2*PCIe x16 LP (front panel) + 4*PCIe x16 LP (rear panel in the case of NVLink)
On-board network adapter		4*10GE
Power supply	4*2200 W hot swap, 2+2	2*3000 W hot swap, 1+1 (the power modules locate at the front and are routed to the rear)
Fan	6 hot swap, 5+1	The built-in fans do not support hot swap. When the fan fails, heat dissipation is affected.
Temperature	5°C to 35°C	5°C to 35°C

Beating Points

- The integrated design of the chassis does not support independent evolution of the CPUs and GPUs.
- Scalable Processors are supported but only up to 165 W, which is difficult to expand in the future. (Note that Huawei G5500 supports only v4 processors.)
- PCIe GPU model. Only the front panel supports 2*PCIe x16 LP cards, and the cards block air ventilation.
- The maximum power consumption of the power module is 3000 W. When NVLink GPUs are used, the full-load power consumption exceeds 3000 W, which does not support 1+1 redundancy.
- The power modules locate at the front and are routed to the rear of the chassis.

Contents

1 Introduction to the G5500

2 Comparison Between G5500 and Competitor Specifications

3 Advantages and Disadvantages of G5500 and Competitor Specifications (International)

4 Advantages and Disadvantages of G5500 and Competitor Specifications (China)

5 How to Defend

How to Defend: Advantages of Other Vendors

Server Model	Advantage	How to Defend
SuperMicro 4028GR-TRT2	<ol style="list-style-type: none"> 1. Supports 24*2.5" SSD/HDD. 2. Supports on-board 2*10GE/GE. 	<ol style="list-style-type: none"> 1. The full-width G560 supports 8*3.5" SSD/HDD and the capacity is sufficient. 6*NVMe drives are supported and deliver higher performance. 2. Huawei supports 3 or 4*PCIe x16 expansion cards, and the number of network ports is sufficient.
SuperMicro 4028GR-TXRT	<ol style="list-style-type: none"> 1. 8*NVLlink GPU cards are supported. 	See "Disadvantages" in the next slide.
HPE Apollo 6500	<ol style="list-style-type: none"> 1. High GPU density: 2U-8GPU 2. Bandwidth from the CPU to the GPU: 2*PCIe x16 in the case of 1:8; 4*PCIe x16 in the case of 1:4 	<ol style="list-style-type: none"> 1. High density can be viewed from the cabinet power supply and single power shelf perspective. <ol style="list-style-type: none"> (1) See "Disadvantages" in the next slide. (2) The Apollo 6500 cannot be configured with power modules, and the 1.5U power shelf needs to be configured additionally. The power supply requires separate cabling and the actual density is 3U-8GPU. 2. The GPU server performs high-speed computing on the GPU, with very little interaction with the CPU. For example: In the AI scenario with the 6500 of 1:8 configuration, the forwarding latency between two 4GPU s is relatively high. In the case of 1:4, the forwarding latency between two 2GPU s is also relatively high. (In some HPC scenarios, when a large amount of interaction is required between the CPU and the GPU, the high bandwidth has many advantages.)
Inspur: NF5288M5(AGX-2)	<ol style="list-style-type: none"> 1. High GPU density: 2U-8GPU 	<ol style="list-style-type: none"> 1. High density: Power supply and power modules in the equipment room cannot support full load. <ol style="list-style-type: none"> (1) See "Disadvantages" in the next slide. (2) The NF5288M5 can be configured with up to two power modules of up to 3000 W. The power supply is insufficient in the case of full load.

How to Defend: Disadvantages

No.	Question	Main Considerations	How to Defend
1	The G5500 supports PCIe GPU cards and does not support NVLink GPU cards.	<ol style="list-style-type: none"> NVLink is mainly used in specific scenarios to improve the performance and the price is high. The specifications are being developed. 	<p>Currently, the roadmap and guidance are based on PCIe GPU mainly out of the following considerations:</p> <ol style="list-style-type: none"> NVLink GPU is used to add NVLink high-speed interconnection between GPUs, increase the transmission bandwidth between GPUs, and improve the performance of some specific applications (requiring high-bandwidth communication between GPUs); the performance improvement is not obvious in other applications. The cost of NVLink GPU cards is higher than that of the PCIe GPU cards (estimated to be 30% higher) and the application scope is narrow. The PCIe GPU card specifications are preferentially developed. The 8*NVLink card specifications are expected to be delivered in Q2 2018.
2	The G5500 G560 is 4U-8G, and the density is not high.	<ol style="list-style-type: none"> As cabinet power supply in normal equipment rooms is limited, heat dissipation for some components is difficult if the density is too high. When the density is too high, the chassis power modules will be limited and it will be difficult to support N+N redundancy. 	<ol style="list-style-type: none"> The power consumption of the GPU server is high (up to 250–300 W for a GPU). According to the preliminary tests, after the 8 P100 and 145 W CPUs are pushed to maximum load, the power consumption will exceed 3200 W. In most equipment rooms, the power supply capability of a single cabinet is 6–15 kW. Even if the 15 kW is used, only up to 4 GPU servers can be configured. The high density is not a significant benefit on GPU servers, and the high density also causes difficulties in cabinet heat dissipation and increases difficulty in engineering design. When the 2U height supports 8 GPUs, it is difficult to supply power to the entire chassis. For example, two power modules are generally used for 2U-8GPU. Even if with 16 A power, the maximum power supply is 3000 W with 220 V AC. When the power consumption exceeds 3000 W, power supply cannot work in 1+1 mode and reliability cannot be ensured. An alternative is to separate the power modules from the chassis and use external power modules. However, this will occupy space when the cabinet is deployed and the actual density will decrease. When the height of 8GPU is 2U, the CPUs and GPUs will have to be configured in a front and rear layout due to the limitation of the GPU card size. As a result, the overall heat dissipation channel is cascaded. When the load is full, the CPU and GPU cannot support 35°C. (Some vendors clearly indicate that the temperature specification is reduced to 25°C when the GPU server is fully configured.)
3	The G5500 G560 does not support the Scalable Processors.	<ol style="list-style-type: none"> The GPU server focuses on GPU acceleration. The specifications are being developed. 	<ol style="list-style-type: none"> The GPU server is used for AI and HPC applications, which are mainly used for GPU acceleration and do not require high CPU performance. Therefore, the GPU server preferentially supported the latest GPU card first. Using the modular design, the CPU and GPU can be independently upgraded and replaced. The model supporting Scalable Processors is being developed.
4	When the G5500 G560 is configured with 4 or more SATA SSDs, the performance cannot reach the line rate.	<ol style="list-style-type: none"> For specifications lower than 4 drives, consider using the SATA SSDs to form RAID 5. Consider using NVMe SSD. 	<ol style="list-style-type: none"> When the 4 SATA SSDs are used to form RAID 5, the performance can reach the line rate, which can meet the requirements of most applications. The G560 supports 6 NVMe SSDs, which deliver higher performance.
5	The G5500 G560 does not support the 2G Cache 3108 card.	<ol style="list-style-type: none"> There are only a few customers with such requirements. Guide the customers to use the 1G Cache 3108 card. 	<ol style="list-style-type: none"> According to the 2G and 1G Cache test results, the read performance is not improved when RAID 5 is used. For writes above 16KB, the random write performance is improved by 10–20%. In the case of RAID 10, the read performance is improved obviously for operations above 16KB and the write performance is not improved. It is necessary to confirm the customer's application scenario and guide the customer to use 1G Cache.



Copyright©2017 Huawei Technologies Co., Ltd. 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. Huawei may change the information at any time without notice.