

Mayflower®-LLMW33-SP3

19" liquid-cooled GenAI server for AI inference, model fine-tuning & training with NVIDIA® H100 NVL GPU

Features

- **High-Performance with AMD® EPYC™ CPU**
- **Including NVIDIA® H100 NVL powerful GPU (2x on request)**
- **Liquid cooling for CPU and GPU**
- **2x 100 GBit LAN (QSFP28) NVIDIA® ConnectX®-6 Dx**
- **Robust full-industrial components for 24/7 operation**
-  **Discover our free Copilot (open-source ChatBot), with which you can test various LLMs on-premise with your documents**



Configuration example

Further configurations on request!

Mainboard:

Industrial Mainboard (ATX), 24/7 operation, long-term availability

Processor:

AMD® EPYC™ 7003/7002
Liquid cooled

Memory:

8x DDR4 DIMM, max. 3200 MHz
RDIMM: max. 64 GB/DIMM
LRDIMM: max. 128 GB/DIMM

Graphic:

OnBoard AST2500, 512 MB DDR4 KI
NVIDIA® H100 NVL, 2x PCIe x16
188 GB HBM3 RAM, 33792 CUDA Cores
Liquid cooled
On request
2x NVIDIA® H100 NVL, 4x PCIe x16

AI Performance:

FP64: 68 TFLOPs
FP64 Tensor Core: 134 TFLOPs
FP32: 134 TFLOPs
TF32 Tensor Core: 1979 TFLOPs
BFLOAT16 Tensor Core: 3958 TFLOPs
FP16 Tensor Core: 3958 TFLOPs
FP8 Tensor Core: 7916 TFLOPs

I/O:

2x 10 GBit LAN (RJ45)
1x Dedicated IPMI LAN (RJ45)
2x 100 GBit LAN (QSFP28), NVIDIA® ConnectX®-6 Dx, PCIe x16
1x RS-232
2x USB 3.2 (Gen 1)
1x USB-C 3.2 (Gen 2)
1x VGA

Power Supply:

90 ~ 264 VAC, 2000 Watt, 80 Plus
On request: Redundant power supply for AC and DC

Expansions:

7x PCIe x16 (Gen 4), full height (4 slots already used)
One slot of it optional
– PCIe x16 (Gen 4)
– PCIe x8 (Gen 4), second M.2
– no PCIe, second M.2, 2x OCuLink

Mechanical:

Chassis
19 Zoll 4HE Rackmount-Chassis
Dimensions (W x H x D)
430 x 177 x 400 mm
Cooling
Hybrid cooling system with 2x 120 mm fans and liquid cooling for CPU and GPU cards
Remote condenser with 9x 120 mm fans for 19" installation

Drive Bays:

Internal
2x M.2 (one M.2 slot depending on slot configuration)
External
2x 5.25"

Environment:

Operating Temperature
0° ~ 35° C
Storage Temperature
-20° ~ 70° C
Relative Humidity
10 ~ 90 % (non condensing)
IP Protection
IP20

Operating System:

Microsoft® Windows® Server
Linux®

Mayflower®-LLMW33-SP3

Additional illustrations

Remote condenser with
9x 120 mm fans for
liquid cooling



Backside







Mayflower®-LLM18-SP3
installed in a
19" server cabinet

Mayflower®-LLMW33-SP3
with liquid cooling
and remote condenser



PNY Technologies Europe
9 rue Joseph Cugnot, 33708 Mérignac cedex | France
T +33 (0) 5 24 240 240 | pnypro@pny.eu
For more information visit: www.pny.com/en-eu/

InoNet Competences and Services

- 
Consulting
- 
Development
- 
Testing
- 
Customization
- 
Manufacturing
- 
Certifications

InoNet Computer GmbH
Wettersteinstraße 18
82024 Taufkirchen, Germany
www.inonet.com

Published: 02. August 2024, 10:44. Subject to be change without prior notice.



PNY Technologies Europe
9 rue Joseph Cugnot, 33708 Mérignac cedex | France
T +33 (0) 5 24 240 240 | pnypro@pny.eu
For more information visit: www.pny.com/en-eu/