How to build a GPU node

Doug Beary
HPC Sales North America, Lenovo
Agenda

• Introduction – Doug Beary
• Lenovo in HPC
• GPU Terms
• GPU variables
• Node Types
• The OEM challenge
• Liquid Cooling
• How to get in touch
• Design Help
Introduction

- Doug Beary
- HPC sales
- Seller / Designer of the #1 system in the Green500
- Top seller at Lenovo for GPU systems
- HPC sales since 2010
# A GLOBAL TECHNOLOGY LEADER

<table>
<thead>
<tr>
<th>57k</th>
<th>#159</th>
<th>7</th>
<th>#1 PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employees serving customers in 160+ countries</td>
<td>Fortune 500</td>
<td>Nationalities in leadership executive council</td>
<td>Provider in the world</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$70B</th>
<th>$1.2B</th>
<th>#150</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortune 500 Company</td>
<td>AI &amp; IOT investment</td>
<td>Forbes 2018 Global 2000 Top Regarded Companies</td>
<td>Markets where we do business worldwide</td>
</tr>
</tbody>
</table>

2018 Lenovo. All rights reserved.
#1 PROVIDER OF SUPERCOMPUTERS IN THE WORLD
1 IN 3 RUN ON LENOVO HPC

168 systems

33% share

19 Markets World-Wide

#1 in 10 Markets

Based on count of listing on the TOP500.org May 2023 list.
Flatiron Institute tops the Green500 List of the world’s most energy-efficient supercomputers*

* Based on attaining the number one position in the latest Green500 List of the most power-efficient supercomputers in the world
Top500 / Green 500 Overview – June 2023

• Retained our #1 Position in System Share (168)
  • Does not include NCI @ #69 (Lenovo/Fujitsu) or LRZ SuperMuc Phase 2 @ #267 (Lenovo/IBM)
• #500 sits at 1.872PF up from 1.729PF
• 44 systems turned over from June
  • HPE: 5 New, 6 Lost (-1)
  • Lenovo: 20 New, 12 Lost (+8)
  • Inspur down to 43 tied with Atos for #3
  • Lenovo has 64 in PRC

• Green500: **Lenovo remains in the #1 position!!**
  • Intel Xeon + nvidia H100 SR670V2

• Lenovo systems in 19 different markets (US, Can, Aus, PRC, Ned, Swe, Ire, UK, JP, BZ, Ger, Swz, Nor, IT, AU, Sing, Spa, SK, France)
GPU types

- Single / Double Wide
- Single precision / Double Precision
- Data Center / Consumer
- Active / Passive
- PCIe Gen 4 / Gen5 – SXM or HGX
- Available / Not available!
- Expensive / More Expensive!
Diverse Data Center GPU offerings from NVIDIA

- **Edge / Inference**
  - T4
  - A10
  - A2
  - 2.7X generative AI performance
  - 4X rendering perf
  - 3.3X ray tracing perf
  - 1.5x larger GPU memory
  - UNDER EMBARGO until GTC23

- **Pro Viz**
  - A16
  - A40

- **Enterprise AI & HPC**
  - A100
  - A30
  - 2x performance for single precision & matrix operations
  - PCIe G4, GDDR6 ECC
  - Quadro Sync II
  - Up to 9x training & 30x inference perf improvement for large language models
  - 5yr subscription for NVAIE included
  - Up to 7x perf for some HPC apps
  - MIG, Confidential computing, DPX instructions
  - 300-350W

- **FSI & Telco**
  - DPU
  - 4X more compute power
  - up to 4X faster crypto
  - 2X faster storage proc.
  - 4X more memory bw
  - 8X faster training (MOE) models.
  - 4th gen Nvlink @ 900 Gbs/sec of GPU-to-GPU interconnect,
  - PCIe Gen6
  - Magnum IO software
  - 3x double precision frops - 236 teraFLOPs FP64 (redstone)
  - Up to 700w

- **HPC & DL Training**
  - HGX 4-GPU

- **HGX 4-GPU**
  - BF3
  - Target SS: May-July
  - Target SS: SR675 V3 (August)
  - Target SS: March (SR670 V2 LeSI)
  - Ship Supported SR670 V2 (LeSI)

Link to Summary of GPUs available for ThinkSystem Servers
## GPU Variables

<table>
<thead>
<tr>
<th>GPU Name</th>
<th>Precision</th>
<th>Single double wide</th>
<th>Active Passive</th>
<th>Available</th>
<th>Price (EDU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGX H100</td>
<td>DBL</td>
<td>Board</td>
<td>Passive</td>
<td>40Weeks</td>
<td>$90K</td>
</tr>
<tr>
<td>SXM A100</td>
<td>DBL</td>
<td>Board</td>
<td>Passive</td>
<td>52Weeks</td>
<td>$40K</td>
</tr>
<tr>
<td>H100 - 80</td>
<td>DBL</td>
<td>DBL</td>
<td>Passive</td>
<td>40weeks</td>
<td>$25K</td>
</tr>
<tr>
<td>H100 - 94</td>
<td>DBL</td>
<td>DBL</td>
<td>Passive</td>
<td>40Weeks</td>
<td>$24K</td>
</tr>
<tr>
<td>A100 - 80</td>
<td>DBL</td>
<td>DBL</td>
<td>Passive</td>
<td>52weeks</td>
<td>$14K</td>
</tr>
<tr>
<td>L40S</td>
<td>SNGL</td>
<td>DBL</td>
<td>Passive</td>
<td>8weeks</td>
<td>$9K</td>
</tr>
<tr>
<td>L40</td>
<td>SNGL</td>
<td>DBL</td>
<td>Passive</td>
<td>8 weeks</td>
<td>$7K</td>
</tr>
<tr>
<td>A40</td>
<td>SNGL</td>
<td>DBL</td>
<td>Passive</td>
<td>8 weeks</td>
<td>$6K</td>
</tr>
<tr>
<td>A30</td>
<td>SNGL</td>
<td>DBL</td>
<td>Passive</td>
<td>6 weeks</td>
<td>$6K</td>
</tr>
<tr>
<td>A4</td>
<td>SNGL</td>
<td>Single</td>
<td>Passive</td>
<td>8 weeks</td>
<td>$4K</td>
</tr>
<tr>
<td>RTX A6000</td>
<td>SNGL</td>
<td>Single</td>
<td>Active</td>
<td>8 weeks</td>
<td>$4K</td>
</tr>
</tbody>
</table>
Node Types

• Workstation
• Server 1U
• Server 2U
• GPU Node
• Specialty Nodes
The OEM Challenge

- Support wattage
- Picking a winner
- Changing needs / designs

Lenovo ThinkSystem SD665-N V3

PowerEdge XE9680
Experience extreme acceleration for AI/ML/DL training with Dell’s first 8-way GPU server

NVIDIA DGX™ H100
Up to 6x training speed with next-gen NVIDIA H100 Tensor Core GPUs based on the Hopper architecture*
- 8U server with 8 x NVIDIA H100 Tensor Core GPUs
- 1.5x the inter-GPU bandwidth
- 2x the networking bandwidth
- Up to 30x higher inference performance**

*Prior Switch XA/L (2018erman), pending verification
**Performance on 17-billion parameter models training for input sequence length 128, output sequence length 420, 32/4/8/16 DR memory, vs 16 T100/DRS & network.
## HPC&AI 2022/23 Server Portfolio

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Codename(s)</td>
<td>Maui / Kahoolawe</td>
<td>Shanghai / Kauai</td>
<td>New York / Oahu</td>
<td>Portofino</td>
<td>Cobol</td>
<td>Seoul</td>
<td>Florence</td>
<td>Talladega</td>
</tr>
<tr>
<td>Form Factor</td>
<td>1U / 2U</td>
<td>2U</td>
<td>1U</td>
<td>6U12N</td>
<td>6U6N</td>
<td>6U12N</td>
<td>6U6N</td>
<td>3U</td>
</tr>
<tr>
<td>CPU Sockets</td>
<td>1S AMD 320W</td>
<td>2S Intel 350W / 2S AMD 320W</td>
<td>2S Intel 350W / 2S AMD 320W</td>
<td>2S AMD 400W</td>
<td>2S AMD 400W</td>
<td>2S Intel 350W</td>
<td>2S Intel 350W</td>
<td>2S AMD 320W</td>
</tr>
<tr>
<td>DIMM Slots</td>
<td>up to 10 / 32</td>
<td>up to 10</td>
<td>up to 12</td>
<td>up to 4 (NVMe) up to 2 (NVMe)</td>
<td>up to 4 (NVMe) up to 2 (NVMe) up to 8 (NVMe)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drive Slots</td>
<td>3x SW 150W / 3x DW 300W or 6x SW 75W</td>
<td>2x SW 150W or 1x FHHL 150W</td>
<td>n/a</td>
<td>4x 600W SXM (Hopper)</td>
<td>n/a</td>
<td>4x DW 300W or 8x DW 300W / 4x 600W SXM5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooling</td>
<td>Air (A4 ; A2 full perf.)</td>
<td>Air (A4 ; A2 full perf.)</td>
<td>Air (A4 ; A2 full perf.)</td>
<td>Water (A4/W5 full perf.)</td>
<td>Water (A4/W5 full perf.)</td>
<td>Water (A4/W5 full perf.)</td>
<td>Air (A4 ; A2 full perf.)</td>
<td></td>
</tr>
<tr>
<td>Cabling</td>
<td>Rear</td>
<td>Rear</td>
<td>Rear</td>
<td>Front</td>
<td>Front</td>
<td>Front</td>
<td>Front</td>
<td>Front/Rear</td>
</tr>
</tbody>
</table>
Lenovo Neptune Liquid to Air (L2A) module

- Introduced 2020 with **SR670 V2**
  - Efficiently cooling 4x 500W GPUs in 3U

- Moved forward with **SR675 V3**
  - Efficiently cooling 4x 700W GPUs in 3U

- Applied to **SR630 V3 / SR635 V3 / SR645 V3**
  - Efficiently cooling 350W/400W CPUs in 1U
ThinkSystem SD650 V3
w/ Intel Sapphire Rapids

ThinkSystem SD650-I V3
w/ Intel Ponte Vecchio

ThinkSystem SD665 V3
w/ AMD Genoa

ThinkSystem SD665-N V3
w/ NVIDIA Hopper
Flagship Lenovo Neptune Direct Water Cooling Platform

- Single 6U12N chassis supports multiple technologies from Intel, NVIDIA and AMD
- Provides a low-cost approach to deliver exascale performance in a hybrid architecture
- Leverages industry-leading and proven infrastructure design for liquid cooling

SD650 V2
2 x Intel Xeon "IceLake"

SD650 V3
2 x Intel Xeon "Sapphire Rapids"

SD650-I V3
2 x Intel Xeon "Sapphire Rapids" + Intel Ponte Vecchio

SD665 V3
2 x AMD Epyc "Genoa"

SD665-N V3
2 x AMD Epyc "Genoa" + Nvidia Hopper

SD650-N V2
2 x Intel Xeon "IceLake" + NVIDIA A100

SD665-N V3
2 x AMD Epyc "Genoa" + Nvidia Hopper
Liquid Cooling Has Its Benefits

Efficient cooling
= Highest SKU Frequency
= Shorten runtimes
+
Dense Packaging (DWC)
or
Lower Fan Speeds (with L2A)

Max freq and lowest operating Temp with liquid cooling

Red Line: Liquid cooled GPU freq. during run.
Black Line: Air-cooled GPU freq. during run

Red Line: Liquid cooled temp during run.
Black Line: Air-cooled temp during run

<table>
<thead>
<tr>
<th>Workload</th>
<th>Total Power (W)</th>
<th>SR670 V2 System Fan Power (L2A) (W)</th>
<th>Air Cooled System Fan Power (Est.) (W)</th>
<th>SR670 V2 System Acoustic (L2A) (bels)</th>
<th>Air Cooled System Acoustics (Est.) (bels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVQUAL/HPL</td>
<td>1600</td>
<td>54</td>
<td>134</td>
<td>7.9</td>
<td>8.7</td>
</tr>
<tr>
<td>Typical Applications (0.7*TDP)</td>
<td>1200</td>
<td>32</td>
<td>70</td>
<td>7.4</td>
<td>8.0</td>
</tr>
<tr>
<td>Idle</td>
<td>220</td>
<td>13</td>
<td>20</td>
<td>6.4</td>
<td>7.3</td>
</tr>
</tbody>
</table>

2023 Lenovo Internal. All rights reserved.
How to get in touch

• Doug Beary  
  • dbeary@lenovo.com  
  • 919-961-4777

• Mike Ridout  
  • Systems Design  
  • mridout@lenovo.com

• Mary Langan  
  • Quotes and Order entry  
  • mlangan@lenovo.com

• Need to:  
  • Build a GPU node?  
  • Build an AI cluster?  
  • Accelerate a workload?  
  • Optimize for a workload?  
  • Design a High Speed Network?

• Design Help?  
• Call me
Future technologies and schedules are Lenovo roadmaps subject to change without notice.