# Rethink Possibilities with Dynamic Infrastructure

Software Defined Infrastructure

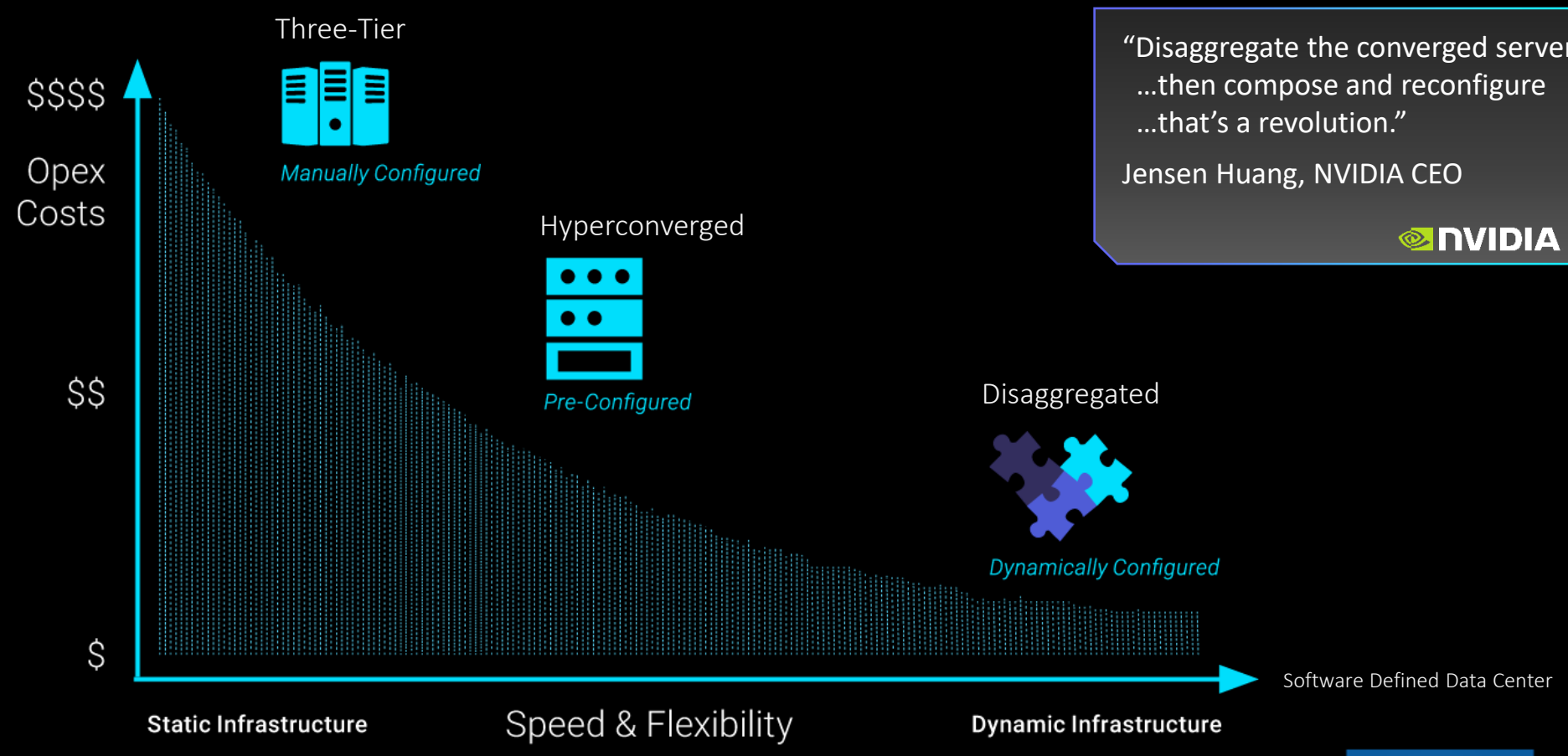
Crafting server resources on demand for the unknown mission workload

Scott Houppermans

Senior Solutions Architect

Federal and Enterprise

# Datacenter Transformation




"Disaggregate the converged server …then compose and reconfigure …that's a revolution."

Jensen Huang, NVIDIA CEO

NVIDIA

Gartner

Source: Gartner




LIQID

# Dynamic Datacenter Infrastructure

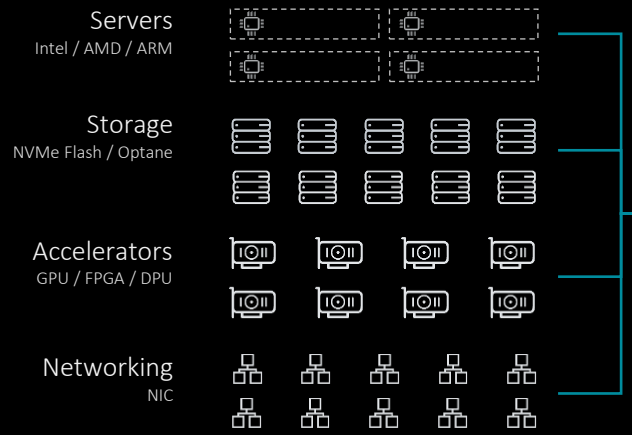
## Legacy Infrastructure

Static Servers

Statically Configured

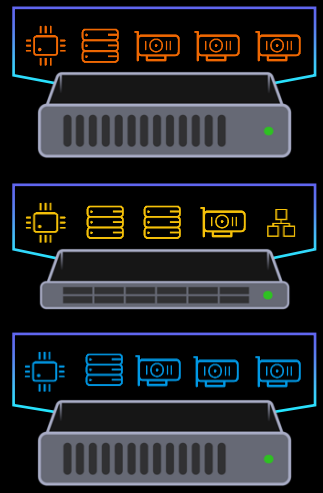
## Modern Infrastructure

Disaggregated Resource Pools




Dynamically Configurable Software Defined Hardware

Dynamically Configured Servers






LIQID

# Liqid Portfolio

## Orchestration Software


1U

**Liqid Matrix Software**
**on Liqid Director**

## Fabric and Expansion Chassis

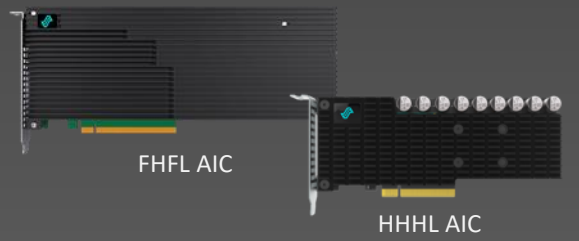
4U

**48/24-Port Switch**

**Liqid EX 4400 Series 10 Slots**

GPU | FPGA | SSD | NIC

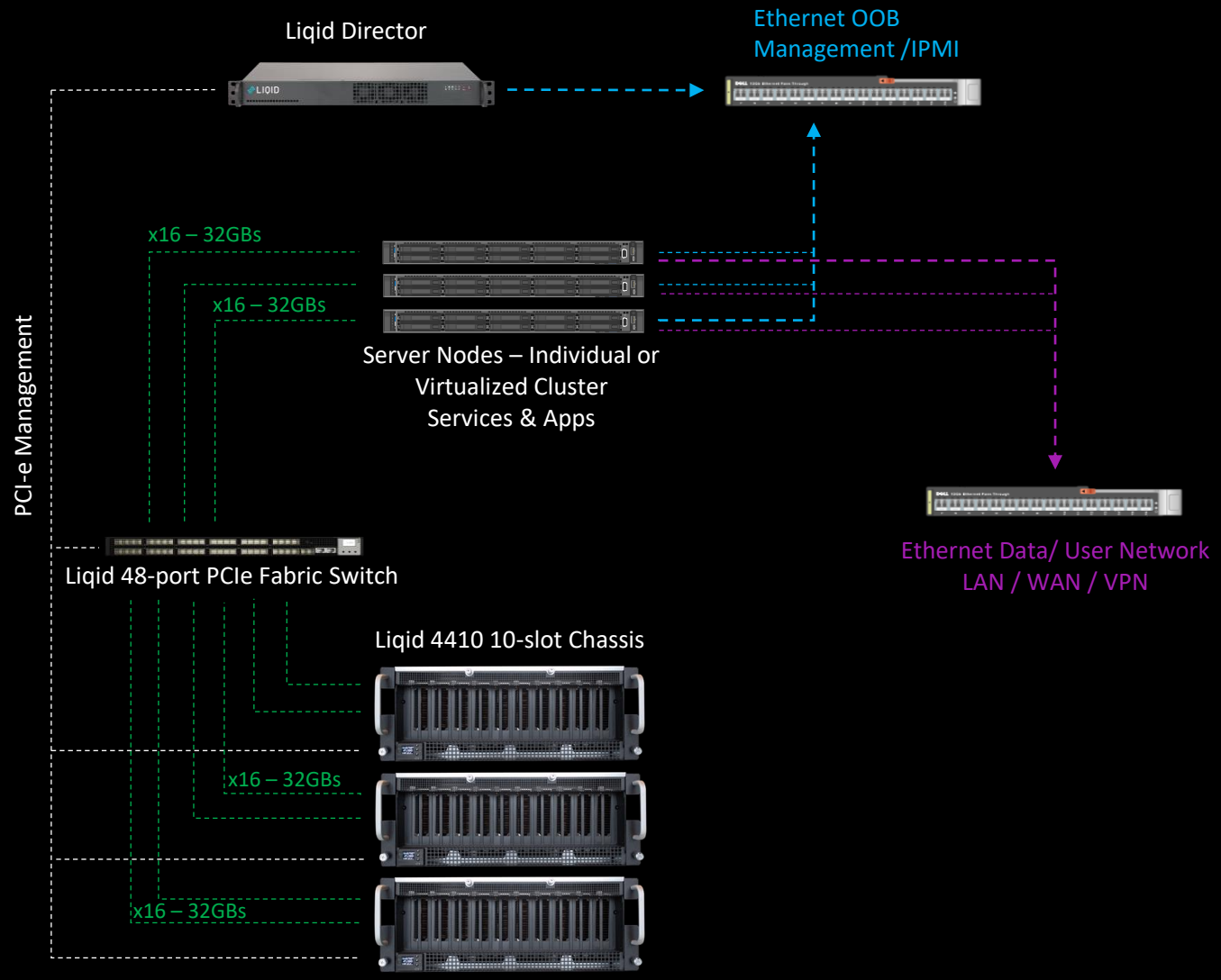
## NVMe Storage




**Liqid NVMe PCIe Storage**

SSD or Storage Class Memory

# SX-3006-4 30-slot SmartStack



# Liqid Portfolio

Liqid SmartStack
Flexibility & Density on Demand

Supports 8 to 30 devices
Supports 1 to 16 server hosts

1U / 2U
Intel / AMD / ARM
+ MX7000

## In Server

NVMe Storage


**Liqid NVMe PCIe Storage**

SSD or Storage Class Memory

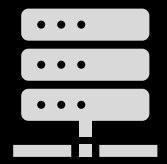
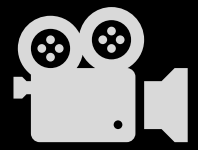
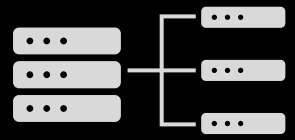
## Or on the Fabric

Liqid UltraStack
High Density Single node
AI Powerhouse

8/16/20 L40S GPU
Ultra Fast Networking

# Use Cases & Customers

## GPU Centric Workloads

| HPC | Enterprise & Cloud | Media & Ent. | Edge | Virtualization |
|---|---|---|---|---|
|  |  |  |  |  |

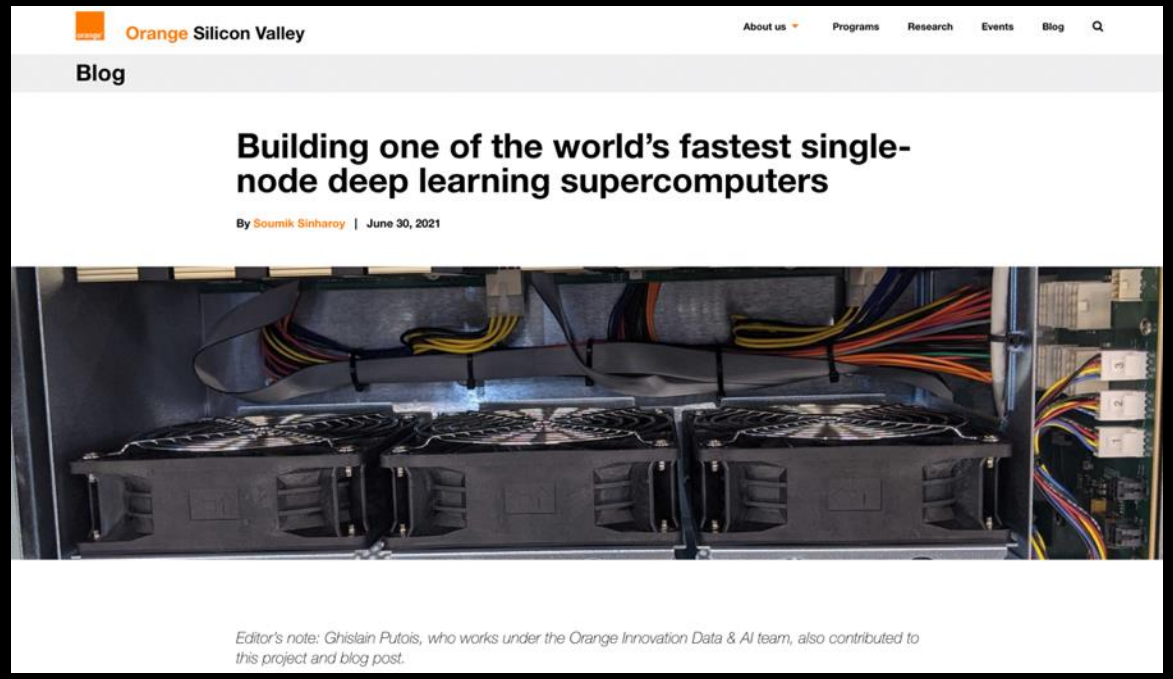
OAK RIDGE National Laboratory · SONY · LUMEN · comcast · servicenow · Raytheon · abbvie

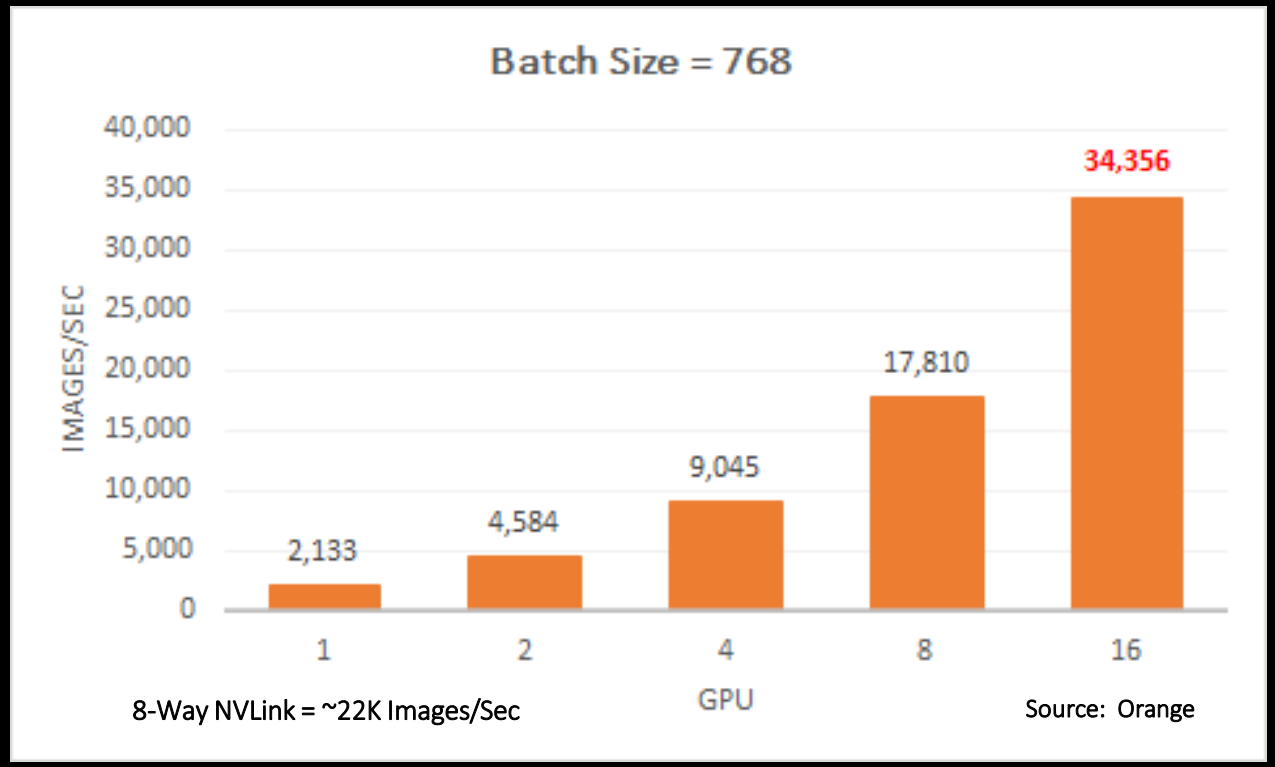
IBM Watson · FBI · HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY · AMOBEE · VANDERBILT · UC San Diego · UIC UNIVERSITY OF ILLINOIS CHICAGO

U.S.ARMY · orange · U.S. NAVAL RESEARCH LABORATORY · VISA · AMERICAN EXPRESS · Durham University · TACC · TEXAS A&M UNIVERSITY

GENERAL DYNAMICS Electric Boat · DOD HPC MODERNIZATION PROGRAM · NAVSEA NAVAL SEA SYSTEMS COMMAND Naval Sea Logistics Center · NORTHROP GRUMMAN · KESSELRUN

# World Record AI Benchmark Performance




8-Way NVLink = ~22K Images/Sec

Source: Orange

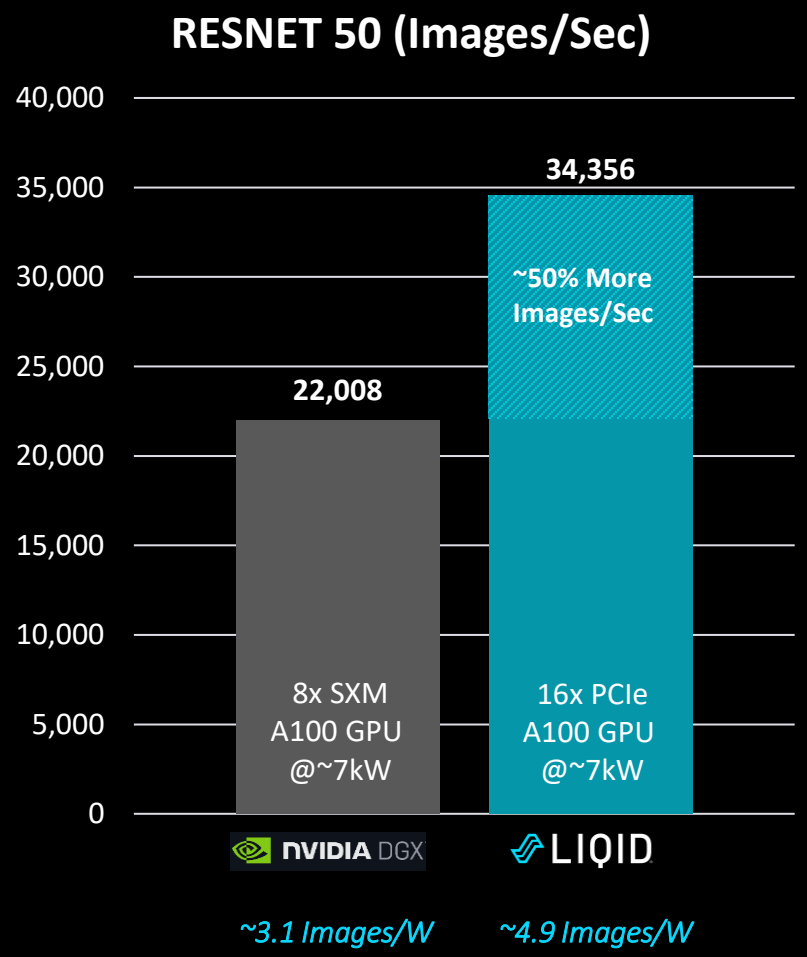
## Highest ResNet Published Performance for Nvidia Based Solution

https://www.liqid.com/blog/orange-silicon-valley-composed-fastest-single-node-gpu-supercomputers-with-liqid-matrix




LIQID

# Rack Scale Generative AI Platform




*2x GPU Density With Improved Power Efficiency*

16x GPU = 150%

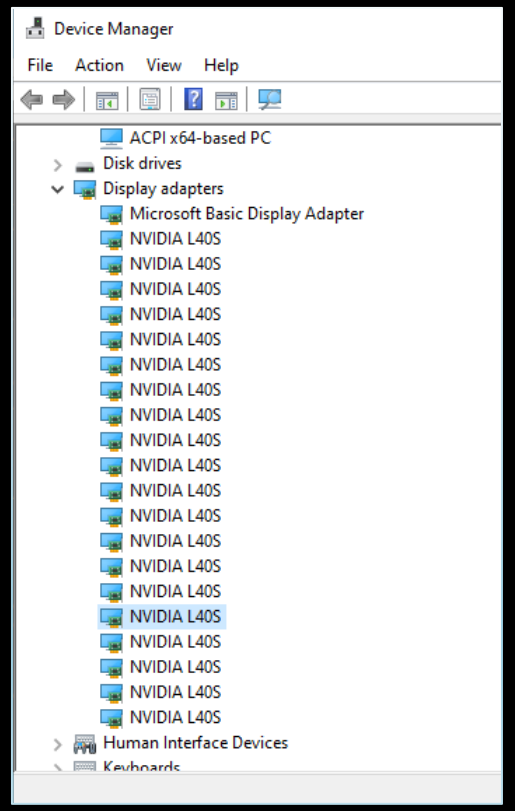
Performance @ similar cost/power

"Thanks for your help in putting together one of the most ridiculous GPU boxes in the world."

VASU AGRAWAL

# Case Study: Meta – 20x L40S UltraStack for VR





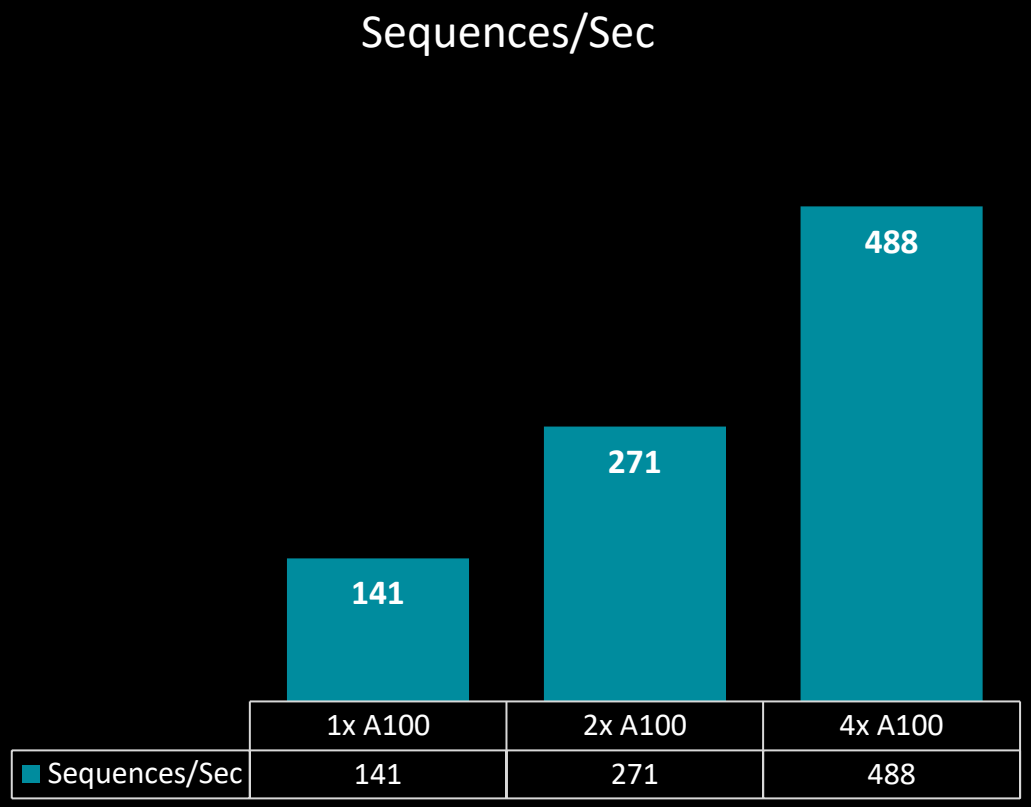

| | |
|---|---|
| **Use Case:** | • Required a high-powered server with 20 GPUs at line rate for VR Development. |
| **Challenge:** | • Needed 20 GPUs to a single socket to avoid any NUMA confusion and increase performance<br>• No vendor could deliver what they needed |
| **Solution:** | • Liqid and Dell partnered to deliver 20x NVIDIA L40S GPUs into a single Dell R7515 server |
| **Results:** | • High-fidelity walkable environment reconstruction and rendering for VR<br>• A new dataset with 11 new room-scale scenes, captured in 8K HDR with up to 22 camera<br>• Presented at SIGGRAPH Asia in December 2023 |

Meta Demo: https://www.youtube.com/watch?v=EohIA7QPmmE&t=187s

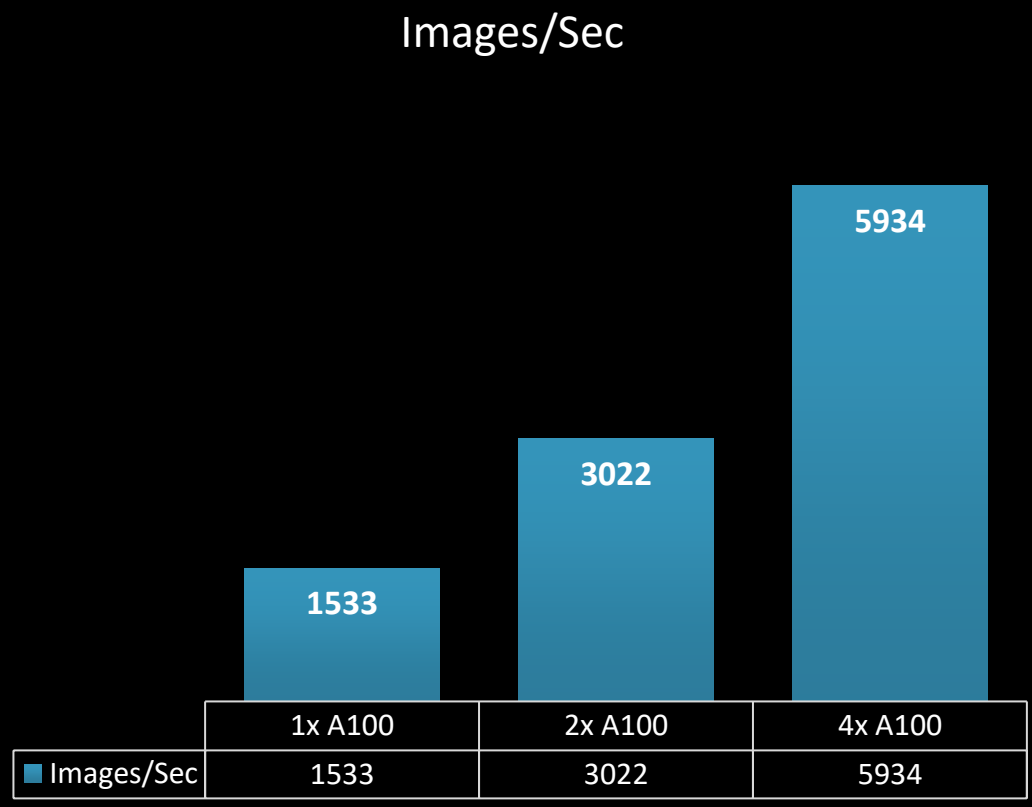
# Performance Scalability: NVIDIA A100 80GB

Composable Configuration: NVIDIA A100 80GB GPUs (PCIe Gen4) +  Dell PowerEdge R750 (PCIe Gen4) + Liqid EX-4410 (PCIe Gen4)

## BERT_Large Squad FP16



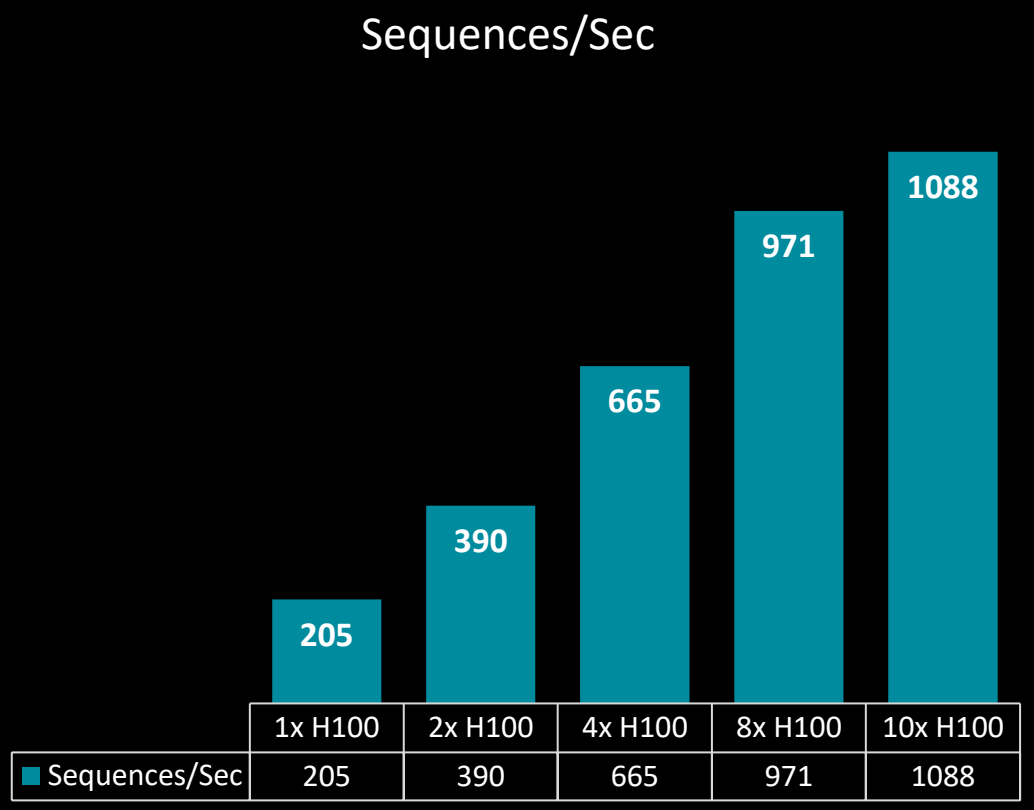

## Resnet50 Training (RESNET50 AMP)



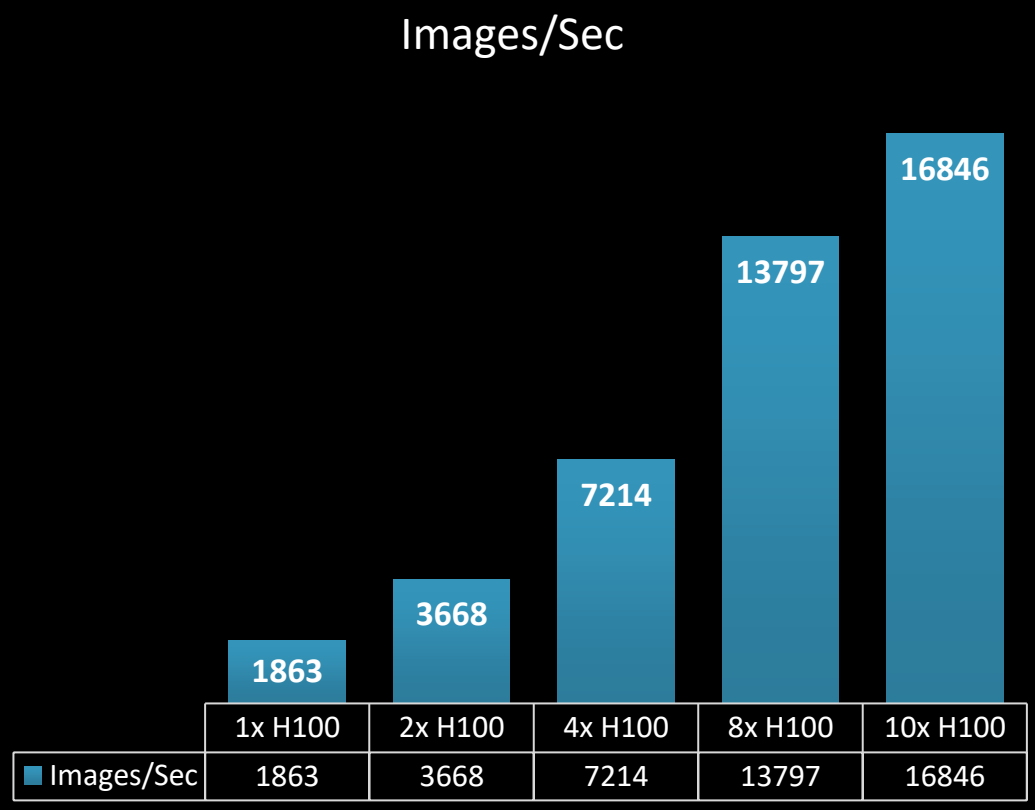
# Performance Scalability: NVIDIA H100

Composable Configuration: NVIDIA H100 GPUs (PCIe Gen5) + Dell PowerEdge R760 (PCIe Gen5) + Liqid EX-4410 (PCIe Gen4)

## BERT_Large Squad FP16




## Resnet50 Training (RESNET50 AMP)



# Composable vs. Non-Composable: NVIDIA H100

Non-Composable Configuration: NVIDIA H100 GPUs (PCIe Gen5) + Dell PowerEdge R760 (PCIe Gen5)
Composable Configuration: NVIDIA H100 GPUs (PCIe Gen5) + Dell PowerEdge R760 (PCIe Gen5) + Liqid EX-4410 (PCIe Gen4)

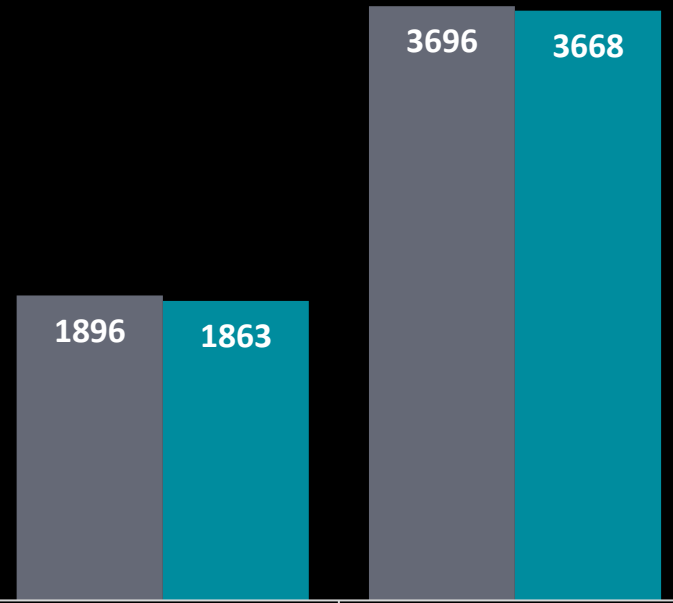
## BERT_Large Squad FP16

Sequences/Sec


| | 1x H100 | 2x H100 |
|---|---|---|
| Non-Composed GPU | 201 | 396 |
| Composed GPU | 205 | 390 |

## Resnet50 Training (RESNET50 AMP)

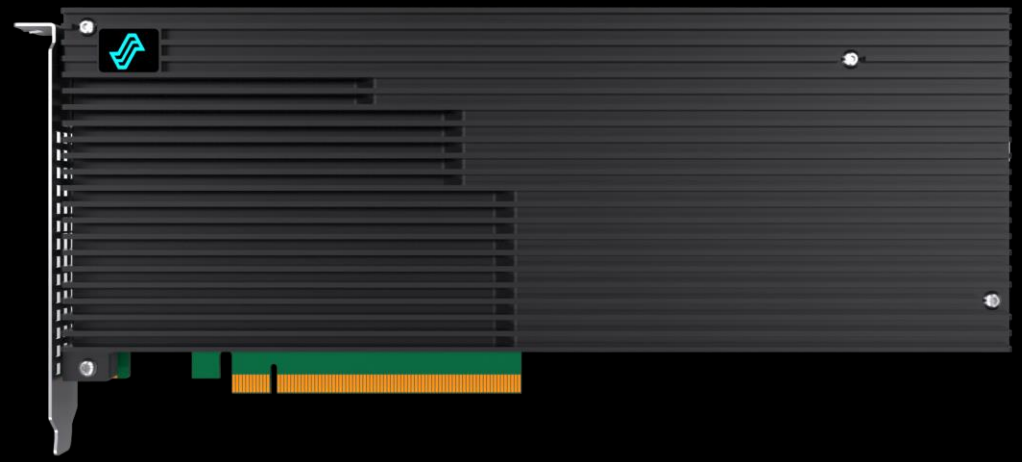
Image/Sec




| | 1x H100 | 2x H100 |
|---|---|---|
| Non-Composed GPU | 1896 | 3696 |
| Composed GPU | 1863 | 3668 |

# LQD4500 Performance Update
# Samsung 9A3 NVMe

| | |
|---|---|
| Random Read 4k | 4.3M IOPS |
| Random Write 4k | 545K IOPS |
| Sequential Read 256k | 26.8 GB/s |
| Sequential Write 256k | 13.0 GB/s |

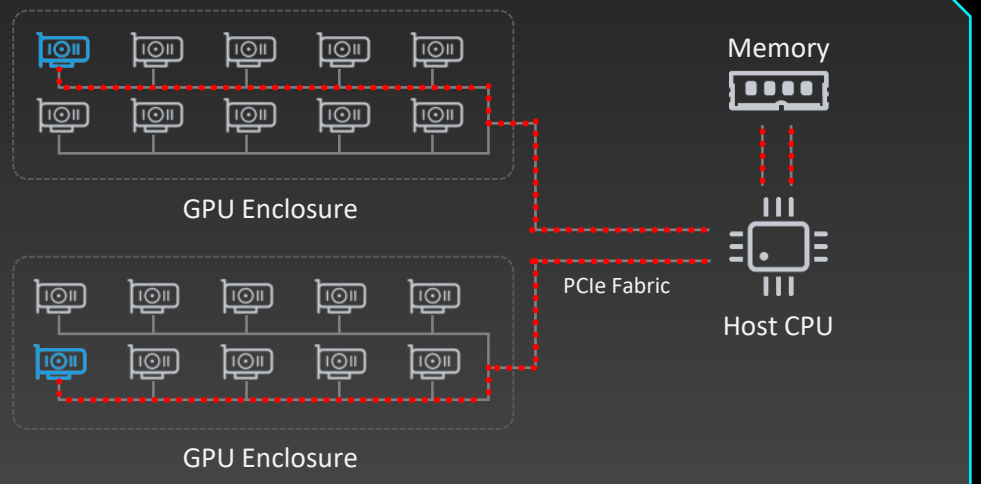


LQD4500 Product Details

| | |
|---|---|
| Product: | Gen4 NVMe SSD AIC |
| Host Interface: | Gen4x16 |
| Form Factor: | FHFL AIC |
| Power Loss Protection: | Yes |
| Capacity: | 8TB to 30TB |

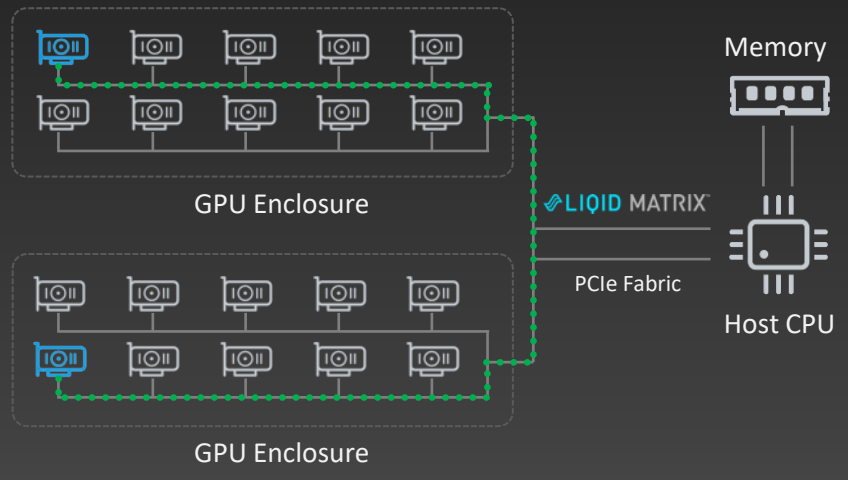
# Peer-to-Peer Technology: GPU-to-GPU

## GPU to GPU w/P2P Disabled




- Bandwidth: 9 GB/s
- Latency: 25 µs

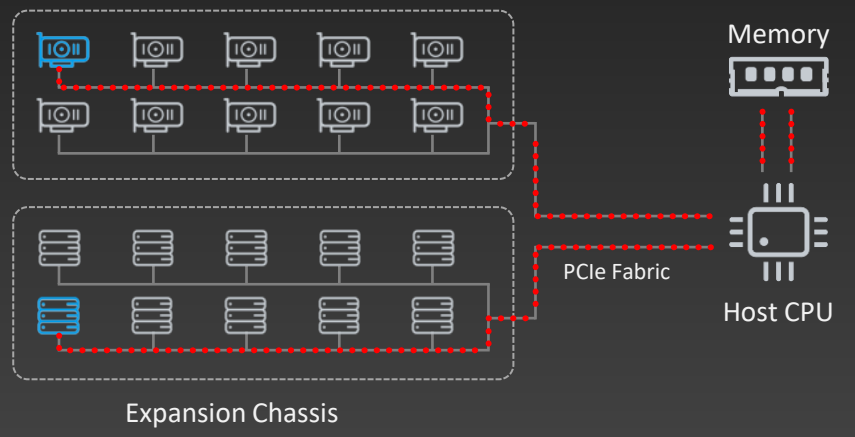
## GPU to GPU w/P2P Enabled




- Bandwidth: 49 GB/s — 474% Higher BW
- Latency: 3 µs — 89% Lower Latency

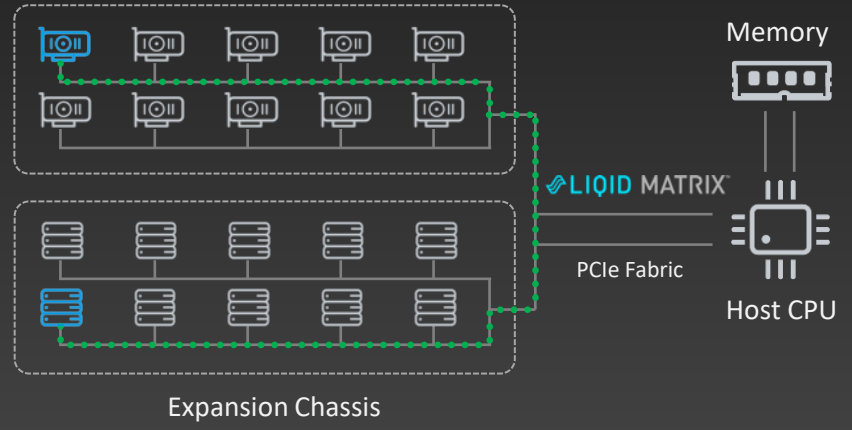
# Peer-to-Peer Technology: GPU-to-SSD (GDS)

## GPU to SSD w/P2P Disabled




- IOPS: 179K
- Latency : 712us

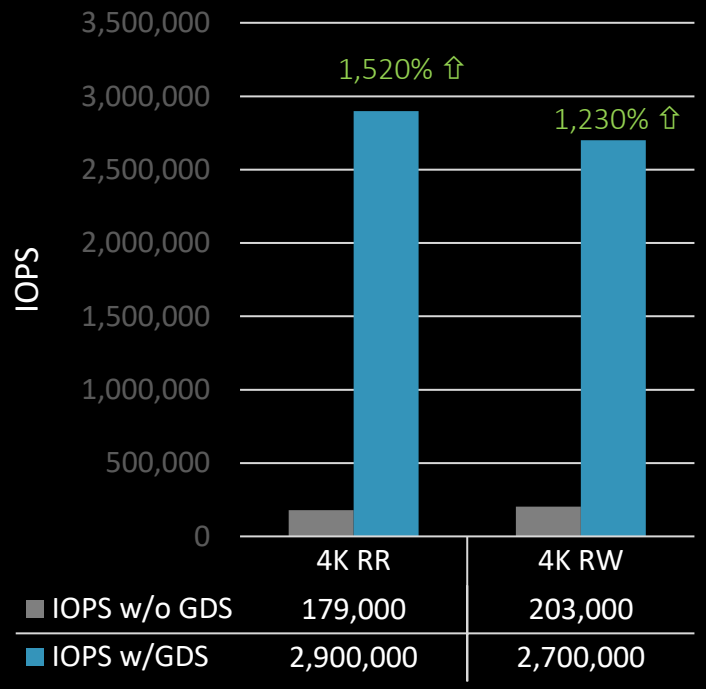
## GPU to SSD w/P2P Enabled



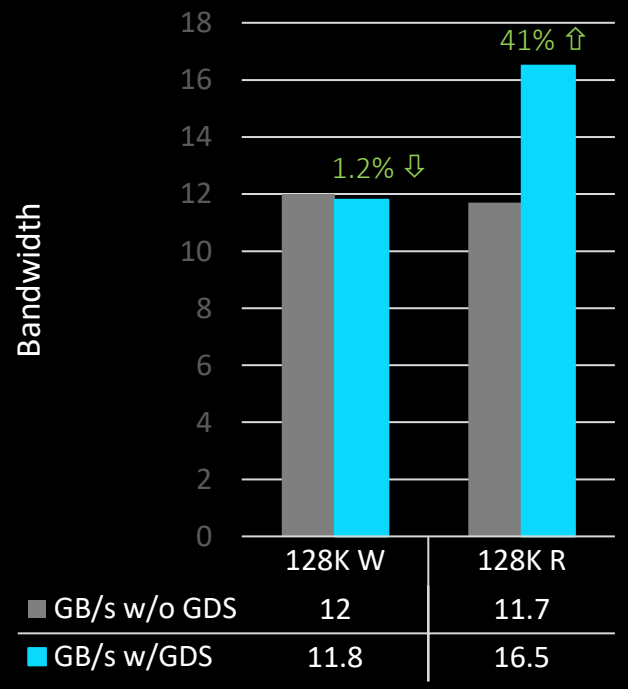

- IOPS: 2,900K 1,520% More IOPS
- Latency: 112us 86% Lower Latency

# Accelerate GPU to Storage Performance

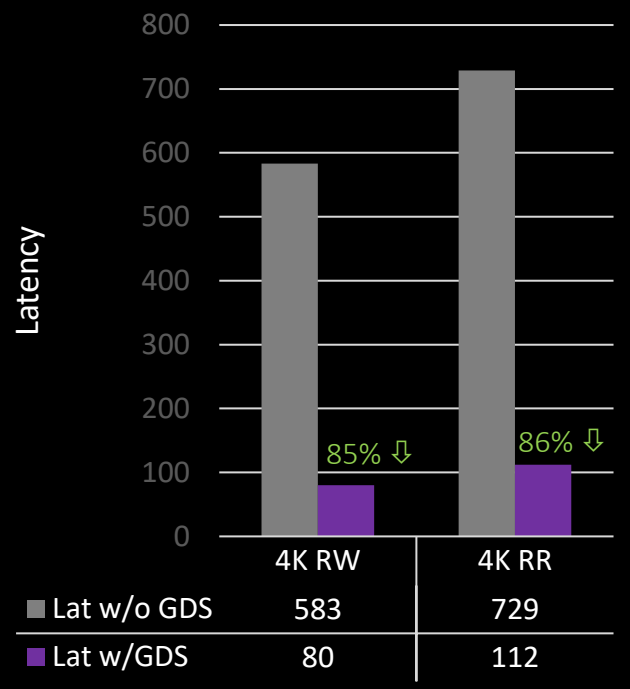
## I/O Per Second (IOPS)




## Bandwidth (GB/s)




## Latency (µs)

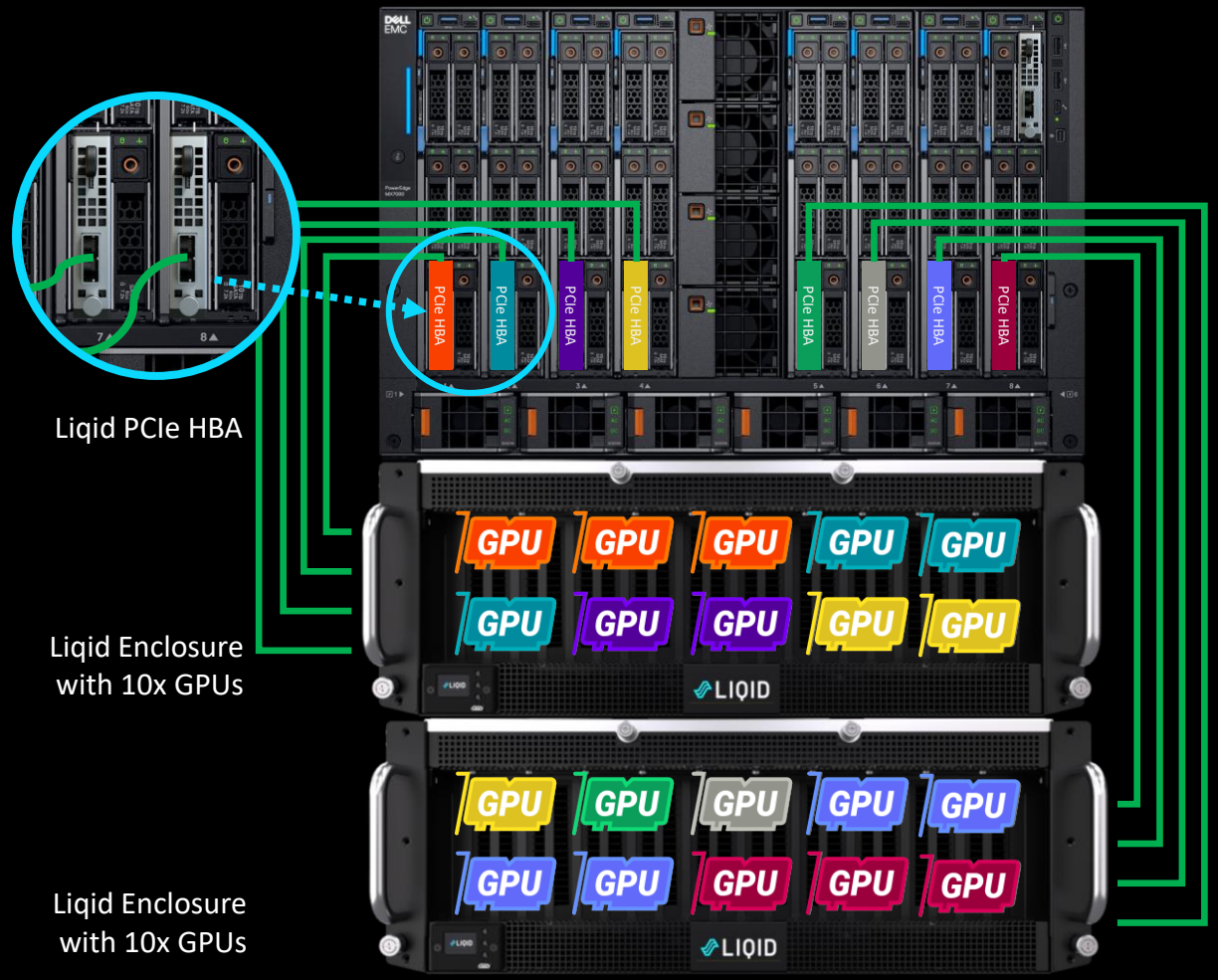



Run on LQD4500 in RAID0 with 128 workers
./gdsio -f /dev/md127 -d 0 -s 10G -i 128k -I 1 -T 60 -x 2 -w 128
./gdsio -f /dev/md127 -d 0 -s 10G -i 4k -I 3 -T 60 -x 0 -w 128




LIQID

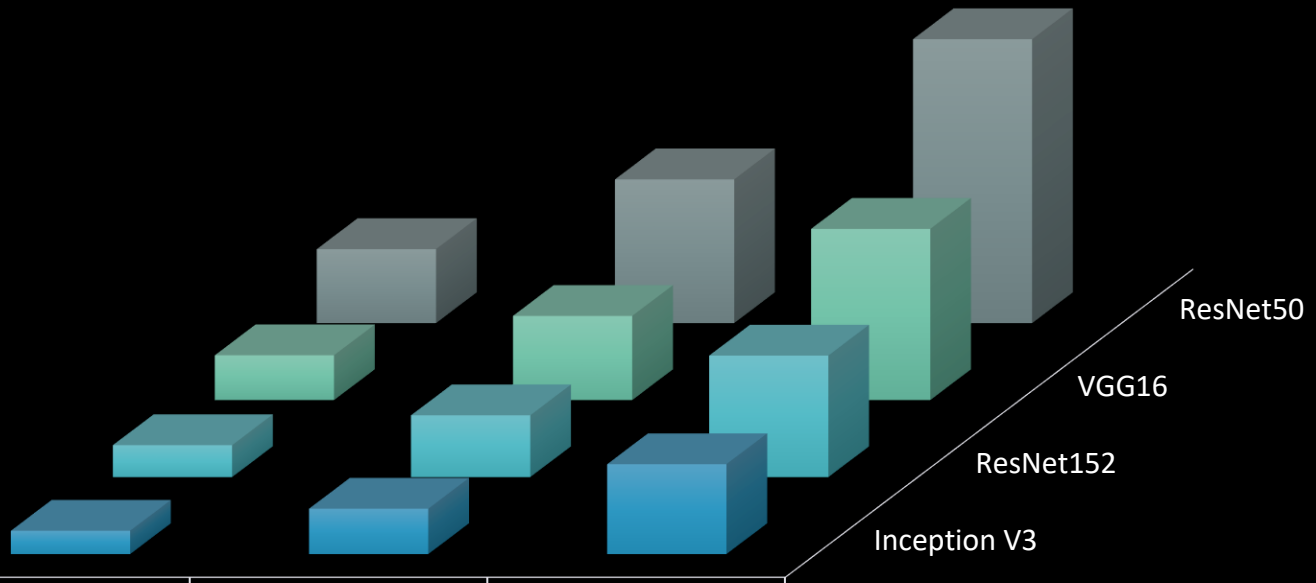
# Transforming MX7000 into an AI Powerhouse




## Deploy GPUs to MX7000 from External Pools via Software

- Scale GPUs up/down in real-time
- Up to 4 GPUs per compute sled
- Up to 30 GPUs per MX chassis
- Supports heterogeneous GPU types
- Compatible with MX740 and MX750

# GPU Performance on MX7000

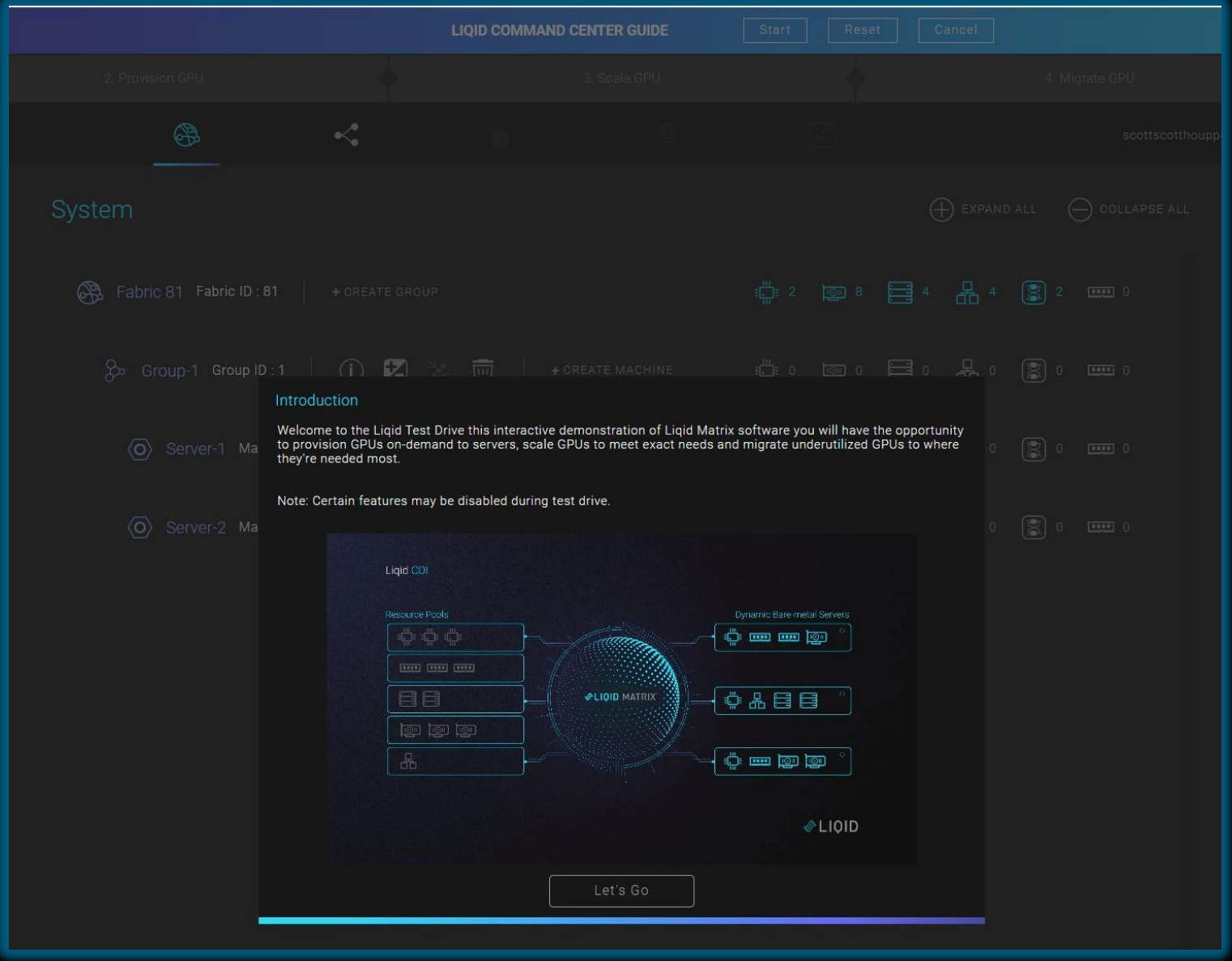



| | 1 x RTX8000 | 2 x RTX8000 | 4 x RTX8000 |
|---|---|---|---|
| Inception V3 | 335 | 651 | 1290 |
| ResNet152 | 458 | 886 | 1743 |
| VGG16 | 643 | 1208 | 2454 |
| ResNet50 | 1060 | 2057 | 4062 |

- MX7000 Leverages RTX8000 in PCIe Expansion chassis with Gen3x4 - measured with P2P Enabled

# Liqid Test Drive

- Web based test drive of Liqid Matrix OS.
- Actual Liqid Matrix OS software
- Container based environment available anytime
- Upgrade to a structured POC in Liqid Lab.




https://cdi.liqid.com/test-drive


Thank You

# Benefits of Agile Infrastructure

## Accelerate Time-to-Results

*2x* Higher Resource Density

- Scale up to 2x GPUs Per Node/Rack
- Accelerate Workload Performance
- Extend Infrastructure Life Cycle

## Increase Agility

*90%* *Faster Resource Provisioning*

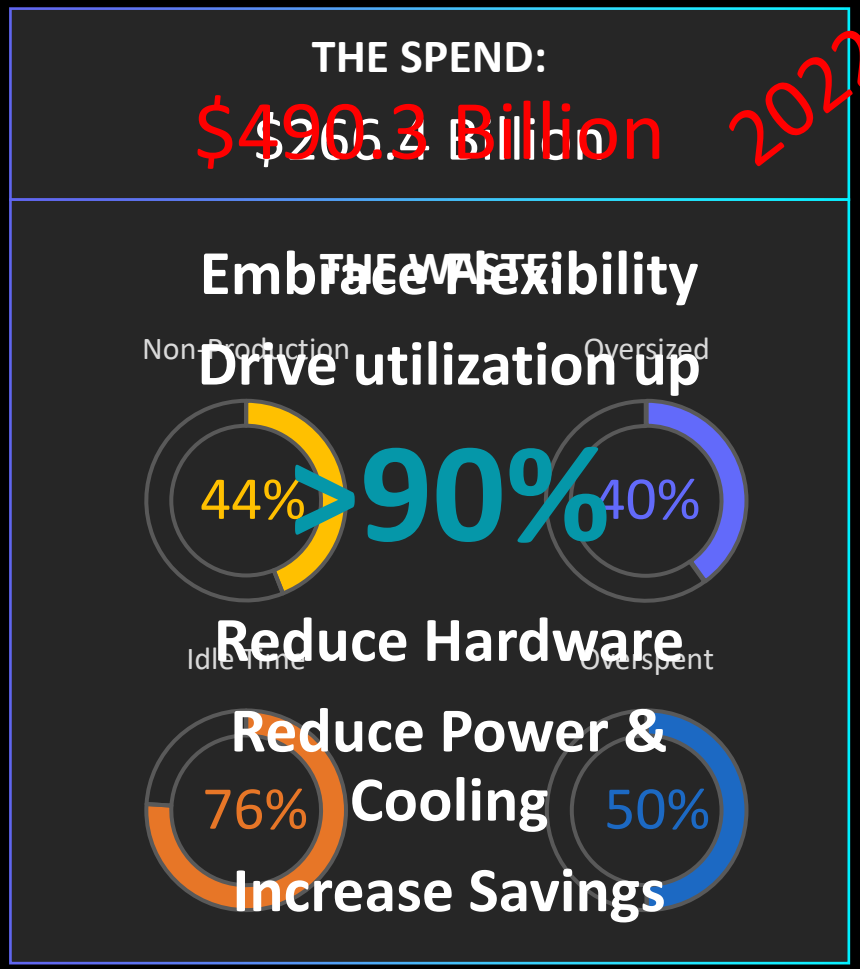
- Instantly Address New Demand
- Scale Resources Independently
- Decouple Purchasing Decisions

## Improve Efficiency

*2x-3x* *Improved Resource Utilization*

- Reduced Infrastructure Costs
- Decrease Power Consumption
- Improve SW License Expense

# Worldwide Cloud Spend & Utilization Rate




Source: https://bit.ly/3uYLoUs
https://tinyurl.com/4fy93zfp

# Performance: Dell MX750c Compute Sled

| FP16 | BERT-Base | BERT-Large | GNMT | NCF | ResNet-50 | Tacotron 2 | Transformer-XL Base | Transformer-XL Large | WaveGlow |
|---|---|---|---|---|---|---|---|---|---|
| 1x A100 | 374 | 119 | 187,689 | 37,422,425 | 1,424 | 37,047 | 37,044 | 16,407 | 198,005 |
| 2x A100 | 638 | 157 | 240,368 | 68,023,242 | 2,627 | 72,631 | 73,661 | 32,694 | 284,709 |
| 3x A100 | 879 | 208 | 313,561 | 85,030,276 | 3,742 | 87,409 | 102,121 | 45,220 | 376,094 |
| 4x A100 | 1,088 | 256 | 379,515 | 98,740,107 | 4,657 | 112,282 | 129,336 | 58,503 | 460,793 |

| FP32 | BERT-Base | BERT-Large | GNMT | NCF | ResNet-50 | Tacotron 2 | Transformer-XL Base | Transformer-XL Large | WaveGlow |
|---|---|---|---|---|---|---|---|---|---|
| 1x A100 | 184 | 55 | 100,612 | 24,117,691 | 891 | 36,953 | 24,394 | 10,520 | 198,237 |
| 2x A100 | 283 | 66 | 115,903 | 38,107,456 | 1,610 | 72,218 | 50,108 | 20,941 | 284,047 |
| 3x A100 | 380 | 88 | 149,359 | 47,133,830 | 2,257 | 84,735 | 66,869 | 28,748 | 370,425 |
| 4x A100 | 464 | 108 | 180,022 | 57,539,993 | 2,840 | 104,398 | 93,394 | 35,927 | 460,492 |

# Performance: Dell MX740c Compute Sled

| FP16 | BERT-Base | BERT-Large | GNMT | NCF | ResNet-50 | Tacotron 2 | Transformer-XL Base | Transformer-XL Large | WaveGlow |
|---|---|---|---|---|---|---|---|---|---|
| 1x A100 | 373 | 120 | 189,738 | 37,371,290 | 1,416 | 37,691 | 37,099 | 16,485 | 195,287 |
| 2x A100 | 737 | 220 | 336,842 | 78,633,579 | 2,775 | 74,613 | 71,591 | 32,574 | 370,720 |
| 3x A100 | 1,082 | 320 | 470,859 | 113,833,170 | 4,184 | 96,308 | 101,962 | 44,914 | 537,716 |
| 4x A100 | 1,421 | 408 | 598,404 | 175,271,653 | 5,553 | 116,697 | 140,794 | 64,826 | 714,114 |

| FP32 | BERT-Base | BERT-Large | GNMT | NCF | ResNet-50 | Tacotron 2 | Transformer-XL Base | Transformer-XL Large | WaveGlow |
|---|---|---|---|---|---|---|---|---|---|
| 1x A100 | 186 | 55 | 101,533 | 24,179,417 | 895 | 37,840 | 24,527 | 10,527 | 202,269 |
| 2x A100 | 350 | 100 | 174,231 | 47,087,138 | 1,755 | 71,998 | 50,188 | 21,085 | 383,209 |
| 3x A100 | 524 | 144 | 241,775 | 67,817,944 | 2,589 | 86,759 | 68,526 | 28,595 | 539,106 |
| 4x A100 | 687 | 188 | 304,852 | 92,692,388 | 3,425 | 114,763 | 91,054 | 41,648 | 710,488 |

# Liqid SmartStack Technical Specifications



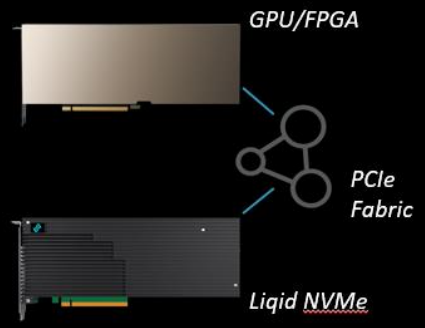

| | SmartStack 10 | SmartStack 20 | SmartStack 30 | SmartStack 30+ |
|---|---|---|---|---|
| **Description** | 10 GPU / 4 Host Capacity | 20 GPU / 8 Host Capacity | 30 GPU / 6 Host Capacity | 30 GPU / 16 Host Capacity |
| **Supported Device Types** | GPU, NVMe, FPGA, DPU | GPU, NVMe, FPGA, DPU | GPU, NVMe, FPGA, DPU | GPU, NVMe, FPGA, DPU |
| **Maximum Devices** | 10x Full-height, full-length (FHFL) 10.5”, dual-slot | 20x Full-height, full-length (FHFL) 10.5”, dual-slot | 30x Full-height, full-length (FHFL) 10.5”, dual-slot | 30x Full-height, full-length (FHFL) 10.5”, dual-slot |
| **Maximum Host** | 4x Host Servers | 8x Host Servers | 6x Host Servers | 16x Host Servers |
| **PCIe Expansion Chassis** | 1x Liqid EX-4410 PCIe Gen4 | 2x Liqid EX-4410 PCIe Gen4 | 3x Liqid EX-4410 PCIe Gen4 | 3x Liqid EX-4410 PCIe Gen4 |
| **PCIe Fabric Switch** | None | 1x 48 Port | 1x 48 Port | 2x 48 Port |
| **PCIe Host Bus Adapter** | PCIe Gen3 x4 Per Compute Sled (1 or more) | PCIe Gen3 x4 Per Compute Sled (1 or more) | PCIe Gen3 x4 Per Compute Sled (1 or more) | PCIe Gen3 x4 Per Compute Sled (1 or more) |
| **Rack Units (without MX)** | 5U | 10U | 14U | 15U |
| **Composable Devices** | Go to liqid.com/resources/library, for a current hardware compatibility list of composable PCIe devices | | | |

# LQD4500 “HoneyBadger” IO Accelerator Use-cases and Applications

## Harsh / Challenged

- Performance, Density
- Form-factors, Active Cooling

## “Feed-the-Beast”

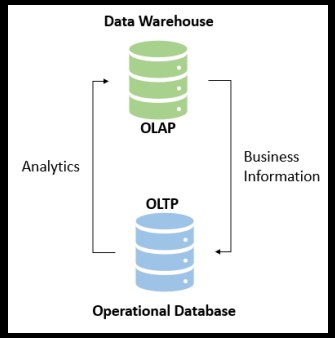



- Storage for GPU/FPGA
- GPU direct storage

## M&E

- Accelerate workflows
- Multiple 8K streams

## Database



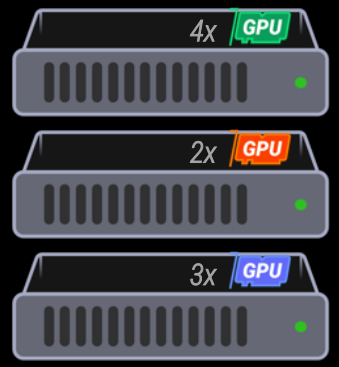

- Rocks/Mongo/MySQL
- Big Data

## Edge

- Fast/Large Data Ingest
- AI/ML at Edge

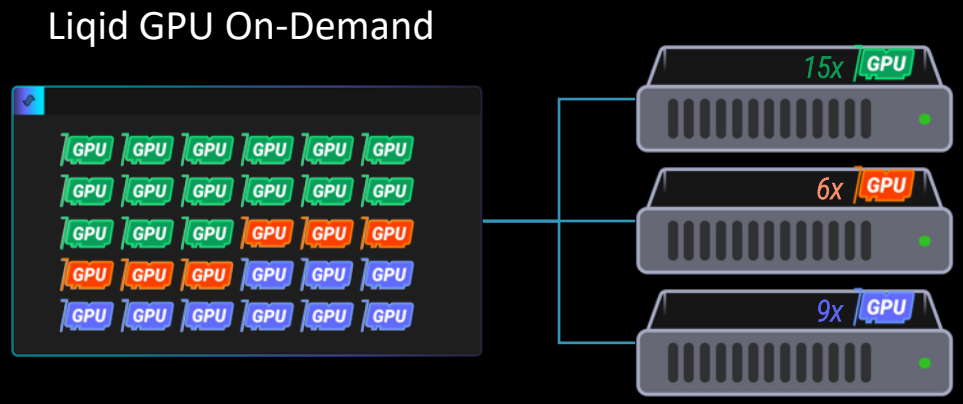
# Use Case: GPU On-Demand Platform

## Statically Configured GPUs



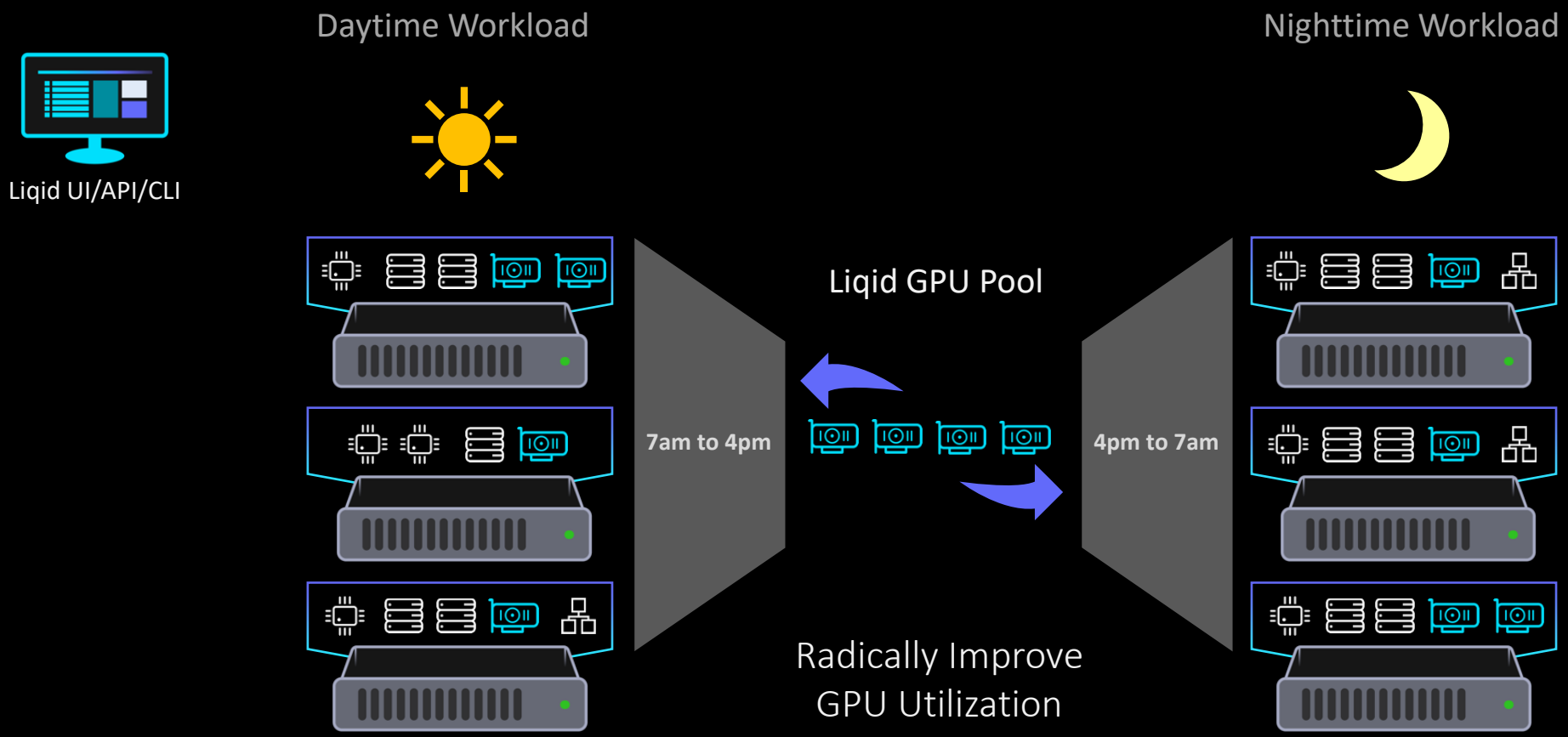

- Limited Scalability
- Inflexible
- Poor Resource Utilization

## Dynamic Configurable GPUs

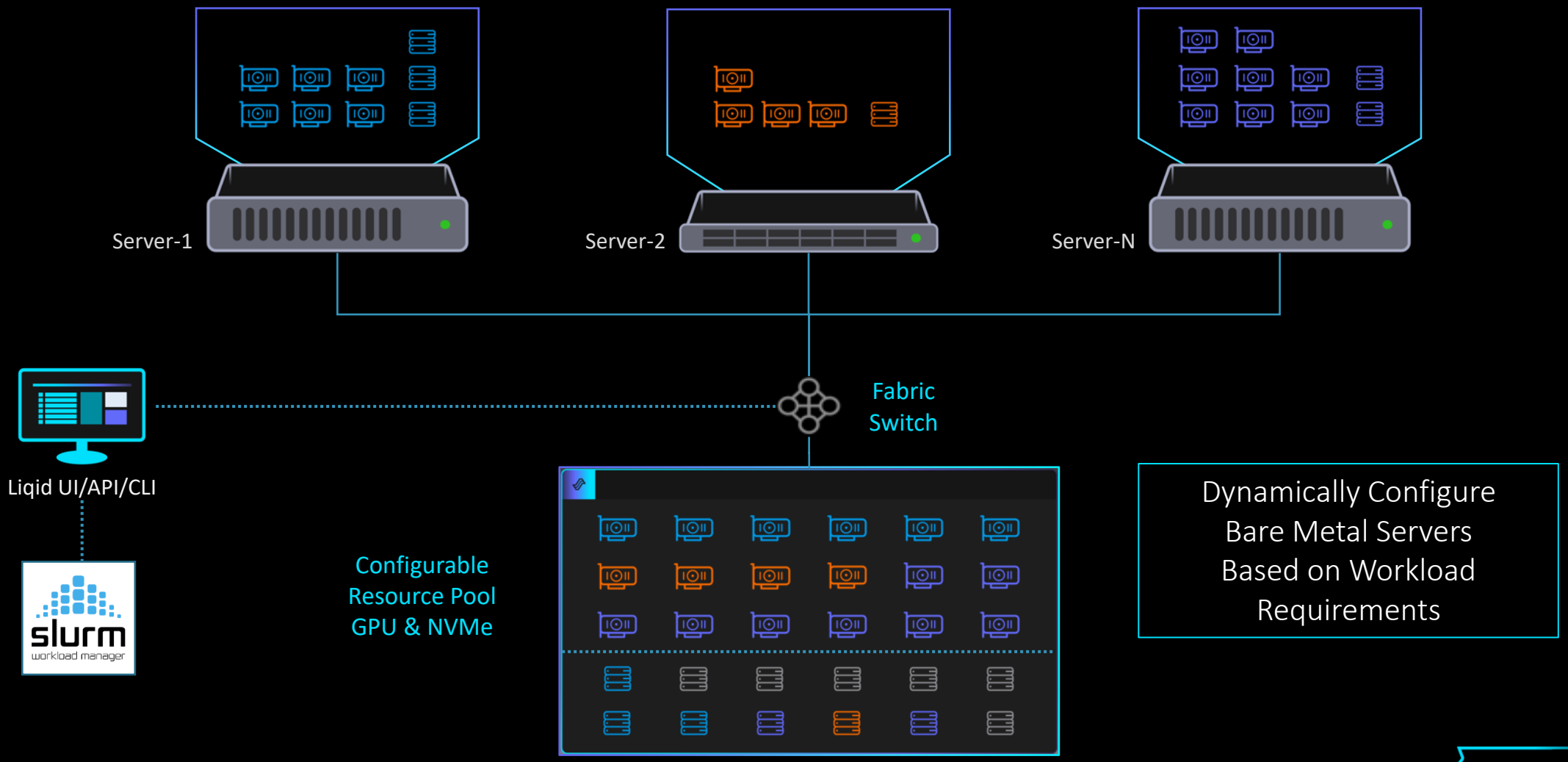



- High GPU Density
- Rapid Deployment and Scaling
- Increased Resource Utilization

# Use Case: GPU Workload Sharing



# Use Case:  HPC Cloud Provisioning




Dynamically Configure Bare Metal Servers Based on Workload Requirements