Video delivery workloads continue to make up a significant portion of current Internet traffic today. As streaming service providers increasingly offer content in 4K and even 8K, or cloud gaming in a higher refresh rate, GPU acceleration with media engines is a must to enable multi-fold throughput performance for streaming pipelines while reducing the amount of data required with better visual fidelity, thanks to the latest technologies such as AV1 encoding and decoding.

Supermicro’s multi-node and multi-GPU systems, such as 2U 4-Node BigTwin system meet the stringent requirements of modern video delivery, each node supporting NVIDIA L4 GPU with the ability to feature plenty of PCIe Gen5 storage and networking speed to drive the demanding data pipeline for content delivery networks.

### Systems

**BigTwin®**
Award Winning Multi-Node System with Resource Saving Architecture

**Large Workload: BigTwin® 2U 4-Node**
- 1 NVIDIA L4 PCIe per node
- 6 2.5" NVMe drives per node
- 16 DIMMs DDR5-4800 per node

SYS-221BT-HNTR / SYS-621BT-HNTR

**CloudDC**
All-in-one Platform for Cloud Data Centers

**Medium Workload: 2U CloudDC**
- 2 NVIDIA L40 PCIe or 4 NVIDIA L4 PCIe
- 12 3.5" SATA drives
- 16 DIMMs DDR5-4800

SYS-521C-NR / AS-2015CS-TNR

**Hyper-E**
High Performance and Flexibility at the Edge

**Medium Workload: 2U Hyper-E**
- 3 NVIDIA L40 PCIe
- 6 NVMe drives
- 32 DIMMs DDR5-4800

SYS-221HE-FTNR / SYS-221HE-FTNRD

### Recommended NVIDIA GPUs

**L40**
- FHFL DW
- PCIe 4.0 x16
- 300W
- 48GB GDDR6

**L4**
- HHHL SW
- PCIe 4.0 x16
- 72W
- 24GB GDDR6
Accelerate **Content Delivery & Virtualization** Workloads
Content Delivery Networks (CDNs), Transcoding, Compression, Cloud Gaming/Streaming

**Opportunities and Challenges:**
- Contents in 4K and 8K, 120Hz+ refresh rate for cloud gaming
- Save data bandwidth and reduce delivery delays
- Faster, more efficient transcoding and compression
- Reduce power consumption and infrastructure cost
- Balancing hot, warm, cold data storage for data throughput and capacity

**Key Technologies:**
- GPU media engines with transcoding acceleration including AV1 encoding and decoding
- NVIDIA RTX GPUs handling both real-time 3D graphic rendering and media streaming for cloud gaming and VDI
- NVIDIA BlueField 2, 3 (DPU) for low latency, secure and fast data management
- Dense, resource-saving multi-node, multi-GPU systems for space and power efficiency
- High-capacity, high-throughput hot-swap storage

**Solution Stack:**
- Red Hat, VMWare
- Container orchestration and management
- SDKs to accelerate and optimize decoding, encoding and transcoding workloads

**Use Cases:**
- Content delivery networks
- 8K, 4K streaming, livebroadcast
- High resolution, high framerate cloud gaming and streaming

---

**GPU Acceleration for Complete Range of Workloads**

- **Large Scale AI Training**
- **HPC/AI**
- **Enterprise AI Inference & Training**
- **Visualization & Design**
- **Content Delivery & Virtualization**
- **AI Edge**

Go to [www.supermicro.com/ai](http://www.supermicro.com/ai) or scan the QR code to download the AI Workload Solution Brochure:

© 2023 Copyright Super Micro Computer, Inc. Specifications subject to change without notice. All other brands and names are the property of their respective owners. All logos, brand names, campaign statements and product images contained herein are copyrighted and may not be reprinted and/or reproduced, in whole or in part, without express written permission by Supermicro Corporate Marketing.