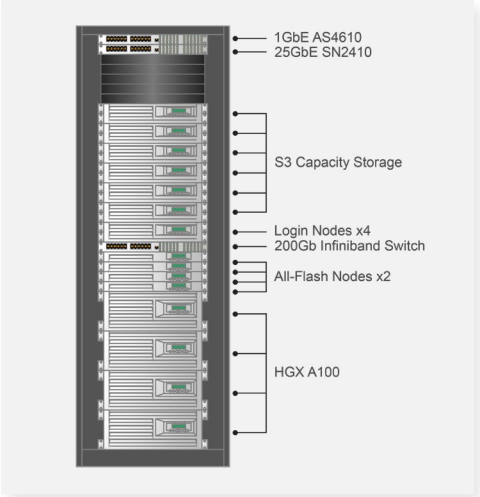
Silicon Mechanics

Silicon Mechanics Atlas AI Cluster™

The next generation in AI supercomputing infrastructure



Everyone's AI workload and organizational needs are unique. But the desire for faster time-to-result and strong ROI isn't.

The Silicon Mechanics Atlas AI Cluster was specifically designed to achieve this combination with powerful building blocks, resulting in a low TCO system that can be tailored to the needs of your organization.

Without any vendor lock-in. Without worries about scaling. And without concerns over how easily you can upgrade.

Welcome to the next generation of AI hardware. Welcome to Silicon Mechanics Atlas.

Benefits

- Pre-configured for AI, reducing time-to-result even on the largest workloads
- Low total cost of ownership compared to traditional supercomputers
- Seamless, linear scaling
- Includes 3rd Gen AMD EPYC™, the world's highest-performing x86 server CPU[i]
- Lower initial cost and no platform lock-in compared to the NVIDIA® DGX A100
- Includes NVIDIA A100 GPUs, providing world's fastest memory bandwidth (over 2 TB/s) to run the largest models and datasets[ii], and GPU partitioning

Ideal Use Cases

- Natural Language Processing
- Machine Learning
- Deep Learning
- Predictive Analytics
- Cybersecurity
- Business Intelligence
- Virtual Assistants
- Robotics

Relevant Industries

- Aerospace/Defense
- Healthcare/Life Sciences
- Design & Manufacturing (inc. Automotive)
- Financial Services
- Retail
- Supply Chain
- Government

GPU-Accelerated Compute Node Components

- 2x AMD EPYC 7742 64-core CPUs (128 cores total), with support for PCIe 4.0
- 9x NVIDIA Mellanox® ConnectX-6 VPI HDR IB 200Gb/s, supports optional GPUDirect RDMA
- NVIDIA Mellanox® Spectrum® SN2000 Gigabit Ethernet HDR switches
- 2x 1.92TB M.2 NVMe Storage (OS) (RAID-1)
- 4x 2200W Power Supplies (3+1)

Option A

- 4U Form Factor 8x NVIDIA HGX™ A100 GPUs with 320GB GPU Memory
- 1TB DDR4 System Memory
- 15TB NVMe Storage (3.84TB U.2 SSD in RAID-0)

Option B

- 4U Form Factor 8x NVIDIA HGX A100 GPUs with 640GB GPU Memory)
- 2TB DDR4 System Memory
- 30TB NVMe Storage (7.68TB U.2 SSD in RAID-0)

- Includes Silicon Mechanics' AI Stack, Silicon Mechanics' Scientific Computing Stack, and support for popular frameworks

Storage Components

- 8 Weka.IO storage nodes + 1 additional Weka.IO node per additional GPU node
- S3 object storage capacity tier. Node count and drive density based on capacity requirements
- Additional optional storage nodes available as needed

Supported Software & AI Frameworks

Includes the Silicon Mechanics AI Stack, Silicon Mechanics' Scientific Computing Stack, and support for popular frameworks.

ubuntu

NVIDIA CUDA

PyTorch

R

Keras

TensorFlow

cuDNN

NVIDIA GPU CLOUD

Quality Assurance for Zero Defects

We build each of our systems to "zero defect" standards in our U.S.-based manufacturing facilities. Then we hand-inspect every order, testing them to ensure they are 100% operational and optimized to support rapid deployment.

Standard 3-Year Warranty

We offer a comprehensive 3-year warranty standard, with every system purchased. But you can add extended or custom warranties if your situation calls for it.

In-House Support

We offer customer support at different levels, each based on a detailed service level agreement (SLA) that fits your needs.

About Silicon Mechanics

Silicon Mechanics is one of the world's largest private providers of high-performance computing (HPC), artificial intelligence (AI), and enterprise storage solutions. Since 2001, Silicon Mechanics' clients have relied on its custom-tailored open-source systems and professional services expertise to overcome the world's most complex computing challenges. With thousands of customers across the aerospace and defense, education/research, financial services, government, life sciences/healthcare, and oil and gas sectors, Silicon Mechanics solutions always come with "Expert Included"℠. Learn more at www.siliconmechanics.com.

[i] MLN-016: Results as of 01/28/2021 using SPECrate®2017_int_base. The AMD EPYC 7763 measured estimated score of 798 is higher than the current highest 2P server with an AMD EPYC 7H12 and a score of 717, https://spec.org/cpu2017/results/res2020q2/cpu2017-20200525-22554.pdf. OEM published score(s) for EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

[ii] https://www.nvidia.com/en-us/data-center/a100/