

SambaNova DataScale SN40L

The Hardware System for Model Training and Inference

Unlock the fastest system for AI model training and inference with the capability to run multiple models, including the latest and largest open source models with the highest performance.

Powered by the SambaNova Cerulean SN40L™ reconfigurable dataflow unit™ (RDU), the SambaNova DataScale® SN40L delivers unprecedented performance across all model sizes to enable government agencies, research organizations, and enterprises to train and deploy the most demanding generative and agentic AI workloads and achieve world record performance with the largest and most challenging models.

The fastest platform for inference

Delivering world record performance and accuracy, across the latest large and small models

Designed for scale

With TBs of addressable memory, run all your models and switch between them in microseconds.

Delivering the highest accuracy

Quickly, easily, and securely fine tune models using private data for the highest accuracy

The DataScale system takes advantage of the unique SambaNova Cerulean SN40L Reconfigurable Dataflow Unit (RDU) to deliver exceptional performance in a small footprint. The SN40L is able to deliver this extreme performance thanks to its revolutionary dataflow architecture and large memory footprint.

Dataflow architecture

The SN40L is purpose-built for AI. Breaking free from the limitations of legacy technologies, the SN40L uses a dataflow architecture and revolutionary software stack that maps AI algorithms to the processor and dynamically reconfigures the processor for optimal performance. This eliminates the redundancy inherent to GPU architectures.

Three tiered memory architecture

Purpose-built to power the largest AI models, the SN40L has a three tiered memory architecture that includes very large memory, high bandwidth memory, and very fast memory. The result is that a single system node can support up to 5 trillion parameters consisting of up to hundreds of separate models. With terabytes of addressable memory, the SN40L is ideal for custom and chained models, and can switch between models in microseconds which is orders of magnitude faster than legacy GPUs.



Model Specifications: DataScale SN40L Configurations

Components

Rack system	DataScale SN40L-8 <ul style="list-style-type: none">Up to 2 DataScale SN40L-8 nodes
Compute	DataScale SN40L-8 nodes, each with 8 x Cerulean SN40LTM Reconfigurable Dataflow Unit™ (RDU) chips, 512 GB -1 TB of high-bandwidth total memory, and 6 TB -12 TB DRAM total memory
Networking	<ul style="list-style-type: none">High performance 400/200 GbE data switchGeneral purpose 1 GbE access and management switch, 48-portSerial console manager server
Software	<ul style="list-style-type: none">Red Hat Enterprise Linux OSUbuntu Linux OSSambaFlow™ software

Environmental Specifications	1-node	2-node
System Dimensions	Height: 78.5" (1994 mm) Width: 24.0" (610 mm) Depth: 50" (1270 mm)	
Operating Temperature	59° F to 86° F (15 C to 30° C)	
Operating Humidity	20% to 80% (non-condensing)	
Operating Altitude	Up to 9842ft (3000m); derated by @ 1.8F (1C) per 984ft (300m) above 2952ft (900m)	
System Weight (without packaging)	738 lbs (335 kg)	1056 lbs (480 kg)



SambaNova DataScale®

To learn more about how SambaNova Systems can accelerate and transform your organization with generative AI, [schedule a meeting](#).

Learn more at SambaNova.AI

 [linkedin.com/company/sambanova](https://www.linkedin.com/company/sambanova)

 @SambaNovaAI

 info@sambanova.ai

Customers turn to SambaNova to quickly deploy state-of-the-art AI capabilities to meet the demands of the AI-enabled world. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing. We enable customers to unlock the valuable business insights trapped in their data. Our flagship offering, SambaNova Suite, overcomes the limitations of legacy technology to power the large complex foundation models that enable customers to discover new services and revenue streams, and boost operational efficiency. Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others.