

SambaNova DataScale SN40L

用於 AI 人工智慧運算模型訓練和數據推理的硬體系統

解鎖用於 AI 模型訓練和推理最快的系統，具備運行多個模型的能力，包含最新和最大的開源模型，並提供最高的處理性能。

助力於 SambaNova Cerulean SN40L™ 可重組資料流單元™ (RDU)，SambaNova DataScale® SN40L 在各種運算模型規模下提供前所未有的性能，使政府機構、研究組織和企業能夠訓練和部署最具挑戰性的生成式和代理式 AI 工作負載，並在最大和最具挑戰性的運算模型上實現世界紀錄級的性能。

最快的 AI 推理平台

在最新的大型和小型模型上提供世界紀錄級的性能和準確性。

為擴展而設計

具備數 TB 的可定址記憶體，能夠運行所有運算模型，並在微秒內進行切換。

提供最高的準確性

快速、輕鬆且安全地使用私有數據微調模型，以實現最高的準確性。

DataScale 系統利用獨特的 SambaNova Cerulean SN40L 可重配置數據流單元 (RDU)，在小巧的設備配置中提供卓越的效能。SN40L 能夠實現如此極致的效能，歸功於其革命性的資料流架構和大容量的記憶體配置。

資料流架構

SN40L 專為 AI 而打造，打破了傳統技術的限制。SN40L 採用資料流架構和創新的軟體堆疊，將 AI 演算法映射到處理器，並動態重新配置處理器以達到最佳效能。這種設計消除了 GPU 圖形處理單元架構中固有的冗餘。

三層記憶體架構

SN40L 專為支援最大的 AI 模型而設計，擁有一層記憶體架構，包括超大記憶體、高帶寬記憶體和超高速記憶體。這使得單一系統節點可以支援多達 5 trillion (兆) 參數的運算，並涵蓋多達數百個不同的模型。憑藉數 TB (terabytes) 等級的可尋址記憶體，SN40L 非常適合自訂模型和串聯模型作業，並且可以在微秒內快速切換模型，比傳統 GPU 運算快上數個等級。



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.