# SambaManaged

## Maximize ROI with a fully managed, AI data center

Power your own AI cloud with turnkey inference containers, deployable in as little as 30 days, with proven ROI.

Expand your market and take advantage of the growing demand for high performance AI inference. Rapidly launch a fully managed, high speed AI inference cloud with SambaManaged. SambaManaged delivers extraordinary AI inference performance using existing power and without exotic liquid cooling infrastructure. Easily launch SambaManaged inside your data center or in pre-configured and fully built easily deployable containers.

| 30-Day Deployment | 10kW & Air-Cooled | Optimized Tokenomics |
|---|---|---|
| Deploy production AI in 30 days vs. 18-24 months for traditional builds - containerized solutions eliminate facility constraints entirely. | SambaNova efficiency enables standard air cooling where competitors require exotic liquid cooling infrastructure. | Clear economic models with validated performance data vs. uncertain payback periods and unproven technology risks. |

SambaManaged delivers a fully managed solution that can be quickly and easily deployed in your data center or as a complete container based solution. Based on the innovative Sambanova SN40L RDU, the SambaNova system delivers an unmatched combination of high performance, low latency, extreme efficiency, and flexibility.

Powered by the proven, enterprise grade AI inference platform that delivers industry leading performance on the best and largest open source AI models, SambaManaged delivers an unrivaled combination of high speed inference, efficiency, and flexibility. SambaNova takes care of the deployment, management, and on-going maintenance of the entire solution, enabling your organization to quickly enter the rapidly growing market for AI inference.

**Flexible Deployment**
Immediately scale-up with the solution that delivers the power of an entire data center in a fully configured, plug and play environment or deploy as fully configured rack level systems in your data center.

**Fully Managed**
Rapidly deploy fast inference services to your customers without the need for internal AI expertise. Simply select the models, token pricing, rates, and traffic that matter to you. SambaNova fully manages the systems, so you can focus on your customers

**Fast Inference Performance**
Deliver the fast inference that your customers need. Each SambaNova system provides extraordinary performance for the best and largest open-source models. Racks can be optimized for the inference speeds that customers require in their solution.

**Highly Efficient**
The innovative design of the SambaNova platform means there is no need for costly infrastructure investments. Each system only consumes an average of 10kW and is air cooled.

**Scalable**
Start with a small cluster, scale upto 1MW. Scale also from individual developer (tier) to SLA-driven managed subscriptions tiers to fully dedicated systems

SambaNova is the leading purpose-built AI system for generative and agentic AI implementations, from chips to models that gives enterprises full control over their model and private data. We take the best models, optimize them for fast tokens and higher batch sizes, the largest inputs and enable customizations to deliver value with simplicity.