SambaNova SYSTEMS

USE CASE

Vulnerability Analysis with SambaNova

World record performance and accuracy for secure generative AI

Challenge

The US faces challenges with bugs in code and flaws in software design that can be exploited. There are also gaps in security procedures, which can lead to security breaches, especially in software deployed to higher impact levels.

Objective

Implement AI-powered tools on the SambaNova platform to analyze and perform comprehensive vulnerability scans across systems and software. Large language models will be utilized to categorize risks based on criticality and likelihood of exploitation, enabling prioritization of remediation efforts. SambaNova's high-performance AI capabilities will be leveraged to handle large datasets and complex analysis, reducing manual effort and increasing the speed and scale of vulnerability analysis.

Distinguishing Factors

- Domain tuned expert models
- RAG grounding
- World record performance and accuracy
- Data security & privacy
- Model/data ownership
- Full stack solution
- On-site/cloud services

Justification

SambaNova's AI-powered tools can handle the "extraordinary compute needs" required for comprehensive vulnerability scans, as indicated by its "1000 Tokens/sec" performance, which is significantly higher than competitors like Nvidia. This allows for faster and more accurate identification and categorization of risks

Solution

SambaNova Suite, combined with the Samba-1 Composition of Experts (CoE) model, delivers the highest combination of performance and accuracy for generative AI. A complete hardware/software platform, SambaNova Suite can be deployed as a fully configured rack-level solution either on-premises, including air-gapped environments, or as a cloud-based solution. The Samba-1 model provides state-of-the-art accuracy across a wide range of use cases.

Key Features

Performance, Accuracy, and Efficiency
Only SambaNova delivers performance of over 1000 tokens/s at full precision, on as few as 16 sockets. This is an unrivaled combination of performance and accuracy with a small footprint, dramatically reducing power consumption.

Total AI Stack Control
With options for on-premises and air-gapped deployments, SambaNova ensures complete data privacy and security, enabling the use of sensitive data for model training without external exposure.

Model Ownership
Once a model is fine-tuned, it becomes the property of the customer in perpetuity, eliminating model ownership concerns.

Configurable Access and Permissions
The Samba-1 model enables flexible access controls, allowing precise management of model permissions to align with organizational roles, enhancing security and ensuring compliance.

SambaNova Suite delivers the most accurate, integrated full stack generative AI platform, optimized for enterprise and government organizations. Delivered through SambaNova's integrated AI platform, which can be deployed on-premises or through the cloud, SambaNova Suite seamlessly integrates into existing business processes to deliver transformative capabilities. With the capacity to be further refined using customer data, these models can deliver unparalleled accuracy while providing enterprise grade security and data governance.

Learn more