SambaNova Code Generation

World record performance and accuracy for secure generative AI

Challenge

Defense and intelligence agencies face critical challenges in ensuring secure and efficient generation of code. Current software development practices often lead to vulnerabilities, inefficiencies, and delays. Additionally, maintaining high standards of code quality and security while managing complex, large-scale projects within the constraints of government protocols adds significant pressure. There is a pressing need for a solution that can streamline and automate these processes, enhancing overall code security and development efficiency.

Objective

To implement SambaNova Suite to boost automated secure code generation at the Defense Information Systems Agency (DISA) and the Defense Intelligence Agency (DIA), enhancing code quality, security, and development efficiency through on-site generative AI capabilities.

Distinguishing Factors

- World record performance and accuracy
- Data security & privacy
- Model/data ownership
- Full stack solution
- On-site/cloud services
- IDE copilot plugin

Justification

Implementing SambaNova Suite for automated secure code generation will transform software development practices and improve code security, quality, and efficiency. By utilizing on-premises generative AI, agencies will ensure the creation of secure and compliant code, speed up development timelines, and enhance overall software reliability, leading to more secure and efficient defense IT systems.

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