

Enhanced Decision Making with SambaNova

World record performance and accuracy for secure generative AI

Challenge

The public sector, including the US government, faces significant challenges when it comes to enhanced decision making. Several key factors contributing to this problem are inefficient decision processes, inadequate data and insufficient analytic capabilities.

Objective

SambaNova allows for the development of custom AI models tailored to specific organizational needs, which can be applied to decision-making processes. SambaNova Suite can process and analyze large datasets quickly which can help analysts make informed decisions while saving time.

Distinguishing Factors

- Model/Data Ownership
- Full Stack Solution
- On-site/Cloud Services
- API based application intelligence
- Web Parsing

Justification

Utilizing SambaNova's platform will allow your organization to expedite decision making processes, saving both time and money. Generative AI can help forecast trends and outcomes based on historical data which can help your organization anticipate future challenges and make decisions accordingly. Our platform can rapidly process and analyze vast amounts of data from multiple sources. This provides decision-makers with deeper, data-driven insights to inform choices.

Solution

SambaNova Suite, combined with the SambaNova 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 CoE 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 CoE 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

