SambaNova SYSTEMS

Near Real Time Operations Augmentation

SambaNova delivers world record performance and accuracy for secure generative AI

Challenge

Emergency response teams face the challenge of managing and analyzing vast amounts of emergency data in near real-time. Traditional methods often lead to delays in decision-making, reduced operational efficiency, and increased costs. There is a need for a robust solution that can seamlessly integrate into existing systems to provide timely, accurate, and actionable insights, ultimately improving emergency management and public safety.

Objective

To implement SambaNova Suite for State Emergency Response Teams in Florida and Texas. This will enhance the management and analysis of emergency data through on-site generative AI capabilities. By leveraging SambaNova's integrated AI platform, teams can expect improved field operations, smarter decision-making, significant cost savings, and more effective emergency management. The full stack solution ensures data ownership, flexibility in deployment (on-site or cloud), and seamless integration into current workflows.

Distinguishing Factors

- Customizable pipelines
- Composable, domain trained, expert models
- World record performance and accuracy
- Data security & privacy
- Model/dData ownership
- Full stack solution
- On-site/cloud services

Justification

By deploying the SambaNova Suite, the State Emergency Response Teams in Florida and Texas will transform their approach to managing and analyzing emergency data. This will not only enhance field operations and public engagement but also ensure smarter decision-making and significant cost savings, ultimately leading to more effective emergency management and improved public safety.

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