IDC LINK
Real-time IDC Research® opinion on industry news, trends and events

SambaNova、生成 AI モデルの Samba-1 を発表

March 05, 2024

By:Matthew Eastwood

クイックテイク

SambaNova Systems が Samba-1 を発表した。生産性の向上とデータプライバシーおよびデータセキュリティの懸念への対応を目的とした AI 完全対応のインフラストラクチャへの移行を強調するこのモデルは、AI（Artificial Intelligence：人工知能）と ML（Machine Learning：機械学習）市場のトレンドに関する IDC の調査と合致する。1 兆パラメーターモデルを誇る SambaNova の Samba-1 モデルは、パブリッククラウドの拡張性にプライベートインフラストラクチャの制御を融合したソリューションを提供することで、エンタープライズ AI のニーズを体現している。これは、IDC の AI 導入に関するハイブリッドモデルの予測と一致している。このモデルの CoE（Composition of Experts）戦略と SN40L チップのような最新のテクノロジーの統合は、急速に技術が進化する中で、柔軟で適応可能なインフラストラクチャのニーズを反映している。このアプローチは、革新的なアプローチに不可欠な要素としてガバナンス、コラボレーション、戦略計画を重視する IDC の視点と一致しており、エンタープライズが AI 導入の複雑さを適切に管理できるようにする。また、変化し続ける AI 市場環境にうまく適応することを目指しているテクノロジーバイヤーにとって重要なインサイトも提供する。

製品発表のハイライト

SambaNova Systems は、革新的な Generative AI（生成 AI）モデル、Samba-1 を発表したが、その中核に 50 種類以上の高品質のオープンソース生成 AI モデルの基盤を含む 1 兆パラメータは注目に値する。エンタープライズ用途に合わせて開発された Samba-1 は、AI 分野におけるプライバシー、セキュリティ、運用効率の品質を向上させることを目的としている。さまざまなビジネスの異なる要件に応えるように設計されており、オンプレミスとプライベートクラウドの両方への導入オプションを提供して、データプライバシーを最大化し、市場効率の向上を図る。このモデルは、Accenture や NetApp など業界リーダーとの戦略的パートナーシップの恩恵を受けている。また、必要なハードウェアインフラストラクチャを最小限に抑える SambaNova SN40L チップとの統合によってパフォーマンスが大幅に向上している。

Samba-1 は、CoE 戦略を採用することで、効率と精度を重視する幅広いビジネス用途に特化した AI モデルの力を利用する。このモデルによって、企業にはデータプライバシーの向上、精度、コスト削減、管理の合理化、推論費用の削減など、多くのベネフィットが提供される一方、ユーザーにはデータとモデル展開への自主性が与えられる。SN40L チップにより強化された SambaNova Suite は、ハードウェアやソフトウェアの枠を超えて拡大し、大企業と政府の両方の用途に適した汎用性のある生成 AI プラットフォームを提供する。

単一の包括的なモデルを通じて大規模言語モデル（LLM：Large Language Model）を開発する従来のルートとは異なり、SambaNova は、モジュール形式でリソース効率の高い代替手段を導入した。特定の機能に対象を絞った事前トレーニング済みモデルの統合に重点を置いているため、トレーニングと推論における広範なリソースの必要性が軽減され、スケーラブルなエンタ

ープライズソリューションを提供している。LLM へのこの革新的なアプローチは、従来の AI 開発標準に挑戦するだけでなく、安全でプライベートな高水準の AI 導入モデルの代替手段をエンタープライズに提供する。SambaNova が戦略的に重視するのは、フォーチュン 500 とグローバル 2000 企業である。その目標は、従来のモノリシックモデルに代わる費用対効果の高い代替手段を提供することで、エンタープライズ AI 市場における存在感を強固にすることである。これは、現代のデジタルビジネスの複雑な AI 需要に応えながら、インフラストラクチャスタック全体に渡り革新を続けるという SambaNova のコミットメントを示している。

IDC の視点

SambaNova Systems の発表は、IDC が注視してきた業界トレンド、および IDC が IT バイヤー市場に提供してきた推奨事項と合致する。IDC の生成 AI に関する調査は、AI および ML テクノロジーの急速な進歩と、その結果として、エンタープライズニーズがより堅牢な AI 完全対応のインフラストラクチャへ移行することを強調している。この変化する環境の特徴は、ビジネス生産性の向上、顧客のデジタルエクスペリエンスの革新、そしてデータプライバシー、セキュリティ、規制順守の複雑さへの対応がますます重要視されていることである。

SambaNova Systems の新しい Samba-1「1 兆パラメータ」モデルは、今日の AI/LLM 環境で展開している急速な進歩を象徴している。パブリック AI インフラストラクチャとプライベート AI インフラストラクチャが異なる方向に向かっていることに注目する IDC の視点に、このイノベーションはまさに合致しており、パブリッククラウドリソースの拡張性や柔軟性と、プライベートデータセンターインフラストラクチャの制御、セキュリティ、コンプライアンスとの間のギャップを埋めるソリューションの具体例をエンタープライズに提供する。SambaNova のモデルはエンタープライズのニーズを念頭に設計されている。これは、エンタープライズが両方の長所を活用して AI と ML のワークロードを効率的かつ効果的にサポートしようとする中で、ハイブリッド導入モデルを想定する IDC の見解と一致している。

エンタープライズにとって重要なインフラストラクチャの検討事項の微妙なニュアンスをより深く掘り下げることによって、CoE 戦略を通じた AI モデル開発への SambaNova のアプローチは、IDC が推奨するインフラストラクチャの戦略的で長期的な計画策定の実践例を提供する。特化された AI モデルを活用してさまざまなビジネス課題に対応するこの戦略は、急速な技術進歩に直面してますます重要になっている、柔軟で順応性のあるインフラストラクチャの特色を表している。SambaNova が大企業と政府組織の両方に最適化されたフルスタックの生成 AI プラットフォームに重点を置いていることは、AI インフラストラクチャ環境をナビゲートする際のガバナンス、コラボレーション、柔軟性の重要性を強調する IDC の調査と一致している。最後に、SambaNova SN40L チップのような最新のテクノロジーの搭載は、生成 AI ワークロードの展開における革新的なコンピューティング、ストレージ、ネットワーキングソリューションの重要な役割をさらに強調する。

さらに、次世代システムアーキテクチャ向けの UCIe（Universal Chiplet Interconnect Express）や CXL（Compute Express Link）などの新興テクノロジーに関する IDC の調査は、SambaNova のアーキテクチャ上の革新と合致している。よりカスタマイズされた効率的な AI 処理を可能にするこれらのテクノロジーは、次世代 AI ワークロードのインフラストラクチャ需要に最前線で対応する。SambaNova の革新的なモデルと戦術的な導入戦略は、特定のワークロードのニーズと要件に基づいてパブリック AI インフラストラクチャとプライベート AI インフラストラクチャの間

でバランスの取れたアプローチを取るという IDC の推奨と一致する一方、エンタープライズが AI インフラストラクチャ導入の複雑さに対応するための実践的な設計として役立つ。

要約すると、SambaNova のイノベーションは、IDC が注視してきた業界のトレンドと合致しており、AI インフラストラクチャのニーズの大局的な展望を提供している。これは、現代のデジタルエンタープライズの複雑な要求に対応しながら、急速な AI の技術進歩のペースに適応する、インフラストラクチャ計画策定への戦略的でバランスの取れたアプローチの重要性を強調している。この分析は、エンタープライズ AI 市場に対する SambaNova の Samba-1 モデルの当面の影響を強調するだけでなく、こうした展開を進化する AI インフラストラクチャのニーズという、より広範なコンテキストの中に位置づけ、高度にダイナミックな市場における AI 導入の将来をより深く理解し、うまく対応することを目指しているエンタープライズのテクノロジーバイヤーに貴重なインサイトを提供する。

Subscriptions Covered:
AI and Generative AI Infrastructure Stacks and Deployments