

SambaNova DataScale[®] SN30

イノベーションのためのプラットフォーム

特徴

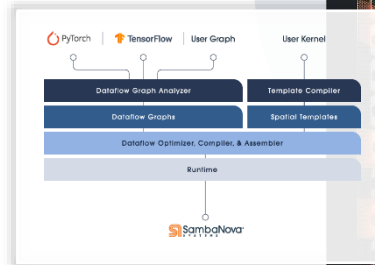
- 再構成可能なデータフローアーキテクチャ (RDA)
- 最新の再構成可能なデータフローユニット (RDU)
- テラバイトのメモリ
- SambaFlow[™]ソフトウェア
- サブスクリプションでの提供

メリット

- 最も要求の厳しいアプリケーションに対応する比類なきパフォーマンス
- 最大規模のモデルを実行し、最大規模のデータを処理する能力
- ベンダー固有のプログラミング言語を使用することなく、一般的なフレームワークを利用可能
- SambaNova再構成可能なデータフローユニット上でモデルの最適なデータフローグラフを自動的に抽出・最適化して実行
- ラック間の帯域とレイテンシーを維持した、シームレスなスケラビリティ
- 学習と推論のためのハードウェアとソフトウェアの統合システム



Cardinal SN30



SambaFlowソフトウェア



DataScale SN30

SambaNova DataScale[®]はハードウェアとソフトウェアが完全に統合されたシステムで、最も要求の厳しいディープラーニング、ファンクションモデル、科学のための AI などのワークロードを学習、展開し、GPT などの最大かつ最も難しいモデルで世界記録のパフォーマンスを達成することが可能です。

DataScaleのデータフローコンピューティングアーキテクチャと大容量のオンチップおよびシステムメモリを組み合わせることで、GPUでは現実的に処理できない以下のような最先端のモデルを管理することができます。



1,000億以上のパラメータを持つ大規模言語モデルのような非常に大規模なモデル



超高解像度の2次元および3次元画像データなどの超大容量データ



RNN、スパースモデル、科学のためのAIワークロードなど、非常に詳細なモデル

最先端の再構成可能なデータフローユニットを搭載したDataScaleは、世界記録の学習時間性能と世界記録のGPT精度を実現し、比類のない使いやすさと相まって、最先端のディープラーニング、ファンクションモデル、科学のためのAIワークロードを強化することができます。

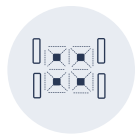
DataScaleは半導体からソフトウェアまで完全に統合されたシステムを提供し、比類のないパフォーマンス、前例のない使いやすさ、シームレスなスケーラビリティを独自の組み合わせで実現することにより、AIプロジェクトのコストと複雑さを軽減します。

SambaNova の再構成可能なデータフローアーキテクチャ

DataScaleはわずか数分でインストールと運用が可能なラックレベルのデータセンターソリューションです。SambaFlowは、並列化のための多大な労力なしに、複雑なAI及びディープラーニングのワークロードの実行を可能とする唯一の実用的なソリューションを提供し、ユーザーはベンダー固有のコーディング言語に依存せずにモデルを実行することが可能です。またPyTorchやTensorFlowなどの一般的なオープンソースの機械学習フレームワークと利用することができます。ローコード/ノーコードAPIは、ベンダー固有の知識を必要とせず、管理を劇的に簡素化します。

世界初のネイティブデータフロープロセッサ

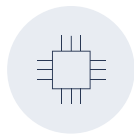
SambaNova DataScaleは、SambaNova Reconfigurable Dataflow Unit™ (RDU) という、ネイティブなデータフロー処理を実現するために作られた次世代プロセッサを搭載しています。特徴は以下の通りです。



SambaNovaの再構成可能なデータフローアーキテクチャは、今日のコアベースのアーキテクチャに固有の頻繁なデータキャッシングと過剰なデータ移動の欠陥をなくし、比類のない効率性を実現することができます。



最適化された計算と通信により、あらゆるサイズと形式のモデル、および学習と推論におけるあらゆるバッチサイズにおいて、すぐに最適な効率で高いパフォーマンスを実現します。



数百MBのオンチップメモリ、数百TFLOPSの処理能力、数TBのオフチップメモリへのダイレクトインターフェースにより、GPUに必要な大規模な並列化作業を行わずに、GPUでは不可能なリッチなコンテキストで、より高度なモデルの開発および展開を行うことができます。



数百TFLOPSの計算能力により、比類のない高速化を実現し、結果を得るまでの時間を大幅に短縮します。

業界最先端のソフトウェア

SambaNova DataScaleは完全なソフトウェアスタックであるSambaFlowを備え、標準的な機械学習フレームワークを利用できます。

- **PyTorchなどの一般的なオープンソースの機械学習フレームワークと完全に統合**され、実行に際してコードの変更は必要ありません。
- **セキュアなマルチテナンシーとマルチグラフの同時実行**により、シームレスなスケールアップとスケールアウトの柔軟性を実現し、無駄なく計算機とメモリのリソースを最大限に活用することができます。
- **簡単な操作でのモデルコンパイル、最適化、実行**により、低レベルのチューニングを必要とせず、すぐに高いパフォーマンスを発揮することができます。
- **データとモデル並列自動マッピング**により、単一デバイスと同じプログラミングモデルを使用してスケーリングを簡素化。特別なプログラミングは不要です。

DataScale SN30 仕様

コンポーネント

ラックシステム	DataScale SN30 <ul style="list-style-type: none">DataScale SN30-8システムを最大3ノードまで搭載
演算器	<ul style="list-style-type: none">DataScale SN30-8システム毎<ul style="list-style-type: none">8基のCardinal SN30™ RDU合計8TBメモリ
ネットワーク	<ul style="list-style-type: none">高性能400/200GbEデータスイッチ48ポートの汎用GbEスイッチシリアルコンソールマネージャーサーバー
追加ハードウェア コンポーネント	<ul style="list-style-type: none">42RUラック冗長電源装置(PDU)×4台ネットワークケーブル
ソフトウェア	<ul style="list-style-type: none">SambaFlowソフトウェアRed Hat Enterprise Linux OSUbuntu Linux OS

環境仕様

構成	1ノード	2ノード	3ノード
システム寸法	高さ 1994 mm 幅 600 mm 奥行 1270 mm		
動作温度	摂氏0度～30度		
動作湿度	20%～80% (結露しないこと)		
動作高度	3000mまで 900m以上で300m毎に動作温度が1度低下		
システム重量 (梱包を除く)	333 kg	477 kg	621 kg

イノベーションのためのプラットフォーム SambaNova Systems DataScale は、アルゴリズムからシリコンまでの Dataflow に最適化された統合システムです。SambaNova DataScale は次世代 AI 技術を迅速に構築し、大規模に展開したい組織のためのコアインフラです。SambaNova Systems Reconfigurable Dataflow アーキテクチャをベースに構築された SambaNova DataScale は、ディープラーニング、ファウンデーションモデル、AI for Science など、幅広いアプリケーションで比類ない効率とパフォーマンスを実現します。SambaNova DataScale はオープンスタンダードとオープンインターフェースで構築されており、既存のインフラストラクチャや環境にシームレスに統合することができます。新たなブレイクスルーの出現により急速に変化する需要にも柔軟かつ効率的に対応することができます。



AIの境界を再定義する SambaNova

SambaNova Systems DataScaleがどのようにAIで組織を加速し、変革する方法について知りたい方はお問い合わせください。

SambaNova についてもっと知るには

 [linkedin.com/company/sambanova](https://www.linkedin.com/company/sambanova)

 @SambaNovaAI_jp

 info@sambanova.ai

AIが実現する世界の需要に対応し、最先端のAIを迅速に展開するためにSambaNovaに注目が集まっています。SambaNovaのエンタープライズAIプラットフォームは、次世代のAIコンピューティングのための技術的支柱です。SambaNovaはデータに閉じ込められた価値あるビジネスインサイトを解き放ちます。主力製品であるDataflow-as-a-Service™は、レガシーテクノロジーの限界を克服し、新しいサービスや収益源を発見し、業務効率を高める大規模で複雑なファンデーションモデルを提供します。カリフォルニア州パロアルトに本社を置くSambaNova Systemsは、業界の著名人や、Sun/Oracleおよびスタンフォード大学のハードウェアおよびソフトウェア設計の専門家により2017年に設立されました。SoftBank Vision Fund 2、BlackRock、Intel Capital、GV、Walden International、Temasek、GIC、Redline Capital、Atlantic Bridge Ventures、Celestaなどから投資を受けています。