

IPU-POD₁₆ DA



AIの新たな飛躍的進歩に取り組む革新家のために



AIの新たな領域の探求を開始

IPU-POD₁₆ DA（ダイレクトアタッチ）は、探求、革新、開発に最適なプラットフォームです。これにより、AIチームは機械知能にさらなる飛躍的進歩を遂げることができます。ホストサーバーでサポートされる4台のIPU-M2000は、お手頃価格のコンパクトな5Uシステムで、トレーニングと推論の両方の作業負荷に対応する強力な4ペタフロップスのAI演算性能を提供します。

ダイレクトアタッチのプラグアンドプレイ

IPU-POD₁₆ DAは、すぐに稼働できるように設計されています。データセンターに設置準備済みの承認されたホストサーバーへ直接接続されたIPU-M2000を備えたターンキーシステムです。広範な書類とサポートは、GraphcoreのAI専門家およびエリートパートナーネットワークの両方から提供されます。

小さく始めて大きくスケールアップ

IPU-POD₁₆ DAは、パワフルなスタンドアロンAI演算リソースです。また、条件に合わせて成長の機会も提供します。IPU-POD₁₆ DAシステムへの投資は後で、大きなIPU-PODシステムへと拡大することができます。

スケールアップ可能なAIインフラストラクチャ

大きなAI作業負荷の通信要件に合うように特別に設計されたIPU-Fabricは、業界標準のIT機器を使用したGraphcoreの革新的で低遅延のジッターがない相互接続です。システム全体で効率的かつ決定論的、網羅的なIPU相互接続を、サイズに関係なく、サポートします。

システムの仕様

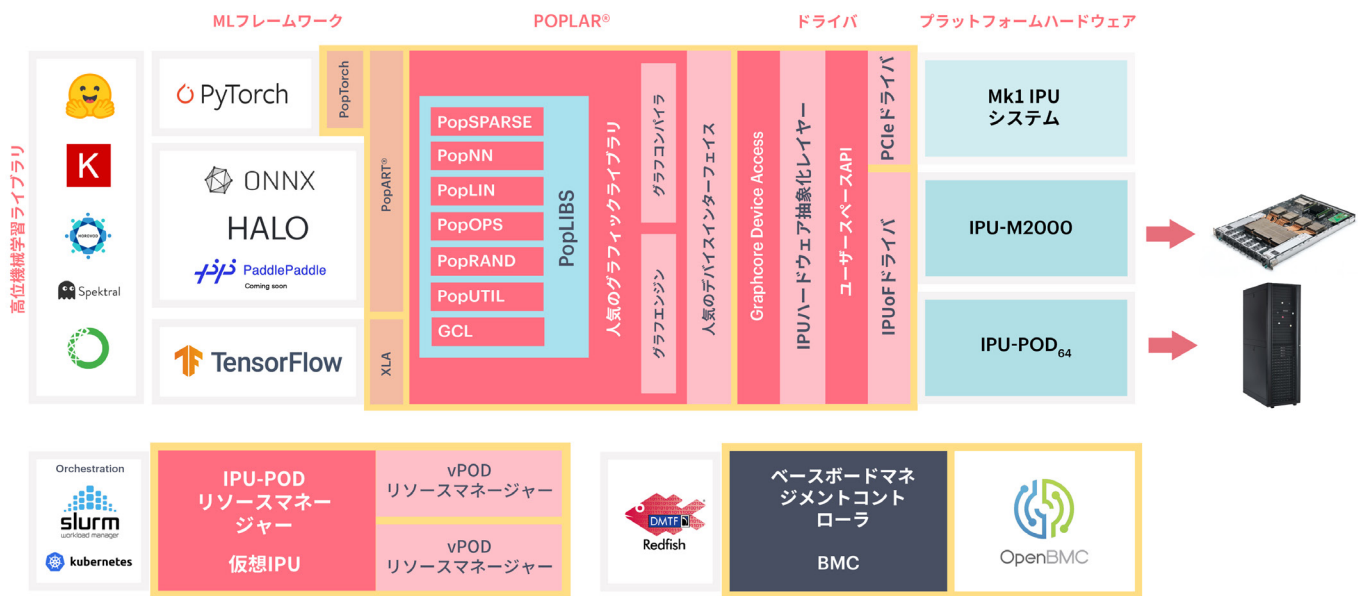
IPU	16 x GC200 Mk2 IPU
IPU-M2000	4 x IPU-M2000
交換メモリ	526.4GB（14.4GBのプロセッサ内メモリと512GBのストリーミングメモリ）
性能	4ペタフロップスFP16.16 1ペタフロップスFP32
IPU Core	23,552
スレッド	141,312
IPU-Fabric	2.8Tbps
ホストリンク	100 GE RoCEv2
ソフトウェア	Poplar
システム重量	66 kg + ホストサーバー
システム寸法	5U
ホストサーバー	Dell、Supermicro、Inspurを含む、Graphcoreパートナーから承認されたホストサーバーの選択肢
熱	空冷式
オプションのスイッチパネーション	Graphcore営業部までご連絡ください

ソフトウェア第一

完全統合で、IPUの最適化された当社のPoplarソフトウェアは、IPUアーキテクチャ固有の特性を活用して、比類なき性能と柔軟性を備えたAIアプリケーションを構築します。Poplarがあれば、開発を複雑にせずモデルのスケールを1から数百へとアップでき、アプリケーションの精度と性能に集中することができます。

AI開発者向けに構築

TensorFlow、PyTorch、その他の人気の高いMLフレームワークは、コミュニティ推進型の連携および革新のために、包括的なPopLibsライブラリとともに、オープンソースとしてサポートされ、利用可能になっています。完全制御して最大性能を引き出したい開発者のために、PoplarはC++での直接IPUプログラミングを可能にします。



展開用に構築

Poplar SDKツールとフレームワークイメージを備えたプリビルドDockerコンテナで、すばやく起動し、実行することができます。IPU-POD₁₆ DAには、IPUリソースの管理を簡素化するための使いやすい直感的なウェブGUIがあります。ここから、ステータスの管理、システムテストの実行、作業負荷に対するIPUのプロビジョニングを行うことができます。

AI専門家へのアクセス

インストール、実稼働展開、アプリケーション開発についての豊富な経験とサポートは、GraphcoreのAI専門家とエリートパートナーネットワークからグローバルに利用できます。

機械知能で次のレベルを体験する用意はできていますか？

AIインフラストラクチャの要件とソリューションの適合性を評価するため、以下の当社パートナーとつながってください。ご不明な点がまだありますか？Graphcoreまで直接ご連絡ください。 info@graphcore.ai