

データセンター ワークロード向けの 適応型アクセラレータ カード

概要

ザイリンクスの Alveo™ データセンター アクセラレータ カードは、進化し続けるデータセンターの要件に迅速に対応するために開発されました。機械学習の推論、ビデオ トランスコーディング、データベース検索/分析などの一般的なワークロードに対して CPU の最大 90 倍の性能を提供します。

シリコンの設計サイクルより早いペースで進化する複雑なアルゴリズムに、固定機能の GPU や ASIC デバイスで対応することは不可能です。ザイリンクスの 16nm UltraScale™ アーキテクチャをベースに構築された Alveo U200/U250 アクセラレータ カードは、継続的なアルゴリズムの最適化に対応する再構成可能なアクセラレーションを提供し、各種ワークロードを加速できるだけでなく、総保有コスト(TCO)も抑えられます。

Alveo アクセラレータ カードの利用は、一般的なデータセンター ワークロードに対応するザイリンクスおよびパートナーが提供するアプリケーションの広範なエコシステムでサポートされます。カスタムソリューションを構築する場合は、ザイリンクスのアプリケーション開発者ツール ([SDAccel™](#)) および [Machine Learning Suite](#) を利用することで差別化したアプリケーションを市場に送り出すことができます。

特長

高速 — 最高性能

- 主なワークロード処理では、CPU*1 と比べて 1/3 のコスト*2で最大 90 倍の性能向上
- GPU ベースのソリューション*3と比べて推論のスループット*4は 4 倍以上、レイテンシは 3 倍向上

適応性 — あらゆるワークロードを高速化

- 同じアクセラレータ カードを使用して、機械学習の推論からビデオ処理まであらゆるワークロードに対応
- ワークロード アルゴリズムの進化に対応するため、固定機能のアクセラレータ カードの製品サイクルよりも迅速に適応可能なリコンフィギャラブル ハードウェアを使用

アクセシブル — クラウドとオンプレミス間のモビリティ

- クラウドまたはオンプレミスでソリューションを相互に運用でき、アプリケーション要件に対応する拡張性にも優れている
- 一般的なワークロード向けの [アプリケーションを利用](#)、または [アプリケーション開発者ツール](#) を使用して独自のアプリケーションを構築

データセンターを刷新



あらゆるワークロードに対応

- データベース検索/分析
- 金融機関のコンピューティング
- 機械学習
- ストレージの圧縮
- ビデオ処理/トランスコーディング
- ゲノム解析

*1: Alveo と EC2 c4.8xlarge での BlackLynx Elasticsearch

*2: DNN 推論における CAPEX と OPEX の削減率に基づく
(Alveo アクセラレータ カード対デュアルソケット Intel Xeon Platinum サーバー)

*3: Nvidia P4 に対して CNN+BLSTM の Speech-to-Text ML 推論で計測

*4: ホワイトペーパー「ザイリンクス Alveo アクセラレータ カードを使用した DNN の高速化」を参照

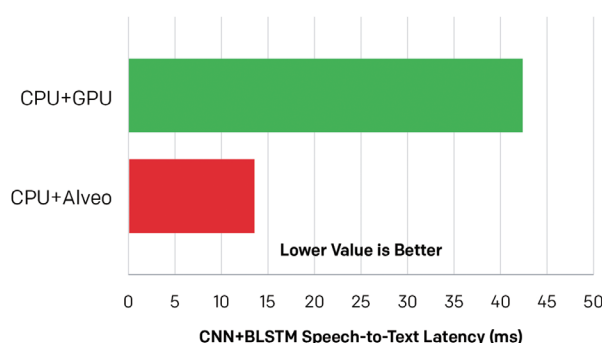
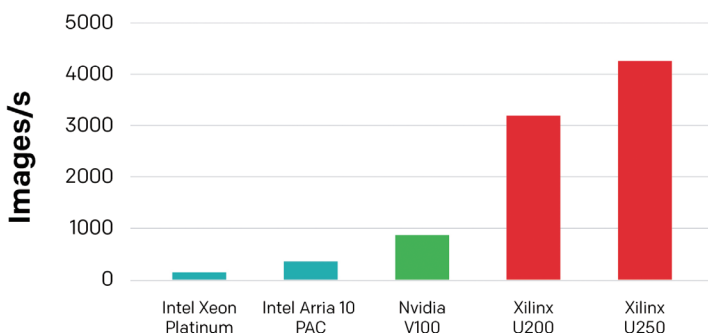
ベンチマーク

あらゆるワークロードを高速化

分野	パートナーのワークロード	ALVEO を使用した場合の加速率(対 CPU)
データベース検索/分析	BlackLynx 社の非構造化データでの Elasticsearch	90 倍
金融機関のコンピューティング	Maxeler 社の VAR(Value-at-Risk) 計算	89 倍
機械学習	ザイリンクスのリアルタイム機械学習推論	20 倍
ビデオ処理/トランスコーディング	NGCodec HEVC ビデオ エンコーディング	12 倍
ゲノム解析	Falcon Computing 社のゲノム解析	10 倍

CPU と比較: Xeon c4.8xlarge AWS | Xeon E5-2643 v4 3.4GHz | Xeon Platinum c5.18xlarge AWS | Dual Socket E5-2680 v3 2.5GHz | Xeon f1.16xlarge

リアルタイム機械学習*の処理能力は 20 倍向上 機械学習推論のレイテンシは 1/3 に低減



*GoogleNet V1: ホワイトペーパー『ザイリンクス Alveo アクセラレータ カードを使用した DNN の高速化』

CPU+GPU: Nvidia P4 + Xeon CPU E5-2690 v4 @2.60GHz(コア数:56)
CPU+Alveo: Alveo U200 または U250 + Xeon CPU E5-2686 v4 @2.3GHz(コア数:8)

機能	ALVEO U200 アクセラレータ カード	ALVEO U250 アクセラレータ カード
INT8 のピーク性能	18.6	33.3
DDR メモリ帯域幅	77GB/s	77GB/s
内部 SRAM 帯域幅	31TB/s	38TB/s
LUT 数	892,000	1,341,000
冷却オプション	パッシブまたはアクティブ	パッシブまたはアクティブ

次のステップ

詳細は、japan.xilinx.com/alveo をご覧ください。データセンター ワークロードのテスト ドライブを開始する場合は [Nimbix 社のクラウド サービス](#) へ、オンプレミスで使用する場合は [Alveo U200](#) または [U250](#) データセンター アクセラレータ カードをご購入ください。

本社
Xilinx, Inc.
2100 Logic Drive
San Jose, CA 95124
USA
Tel: 408-559-7778
www.xilinx.com

ヨーロッパ本社
One Logic Drive
Citywest Business Campus
Saggart, County Dublin
Ireland
Tel: +353-1-464-0311
www.xilinx.com

日本オフィス
ザイリンクス株式会社
〒141-0032
東京都品川区大崎 1-2-2
アートヴィレッジ大崎
セントラルタワー 4 階

アジア パシフィック本社
Xilinx, Asia Pacific
5 Changi Business Park
Singapore 486040
Tel: +65-6407-3000
www.xilinx.com

インド オフィス
Meenakshi Tech Park
Block A, B, C, 8th & 13th floors,
Meenakshi Tech Park, Survey No. 39
Gachibowli(V), Seri Lingampally (M),
Hyderabad -500 084
Tel: +91-40-6721-4747
www.xilinx.com

