

データセンター ワークロード向け アクセラレータ カード

AVNET
Reach Further™

概要

Alveo™ データセンター アクセラレータ カードは、常に進化し続けるデータセンターの要件に素早く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。

サーバ・ワークロードの先端アルゴリズム開発は加速度的に複雑化し、ハード機能固定しているGPUやASICのシリコン進化スピードでは追い付けない状況となっています。

ザイリンクスの16nm UltraScale™アーキテクチャ、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要件に柔軟に適応することができ、トータルコストを削減しながら全てのワークロードに対応が可能です。

Alveo アクセラレータカードをお客様が容易に利用できるよう、ザイリンクスはメジャーなデータセンタ・ワークロードを解決するエコパートナーのソリューションを提案いたします。

また、ザイリンクス Alveo 用のアクセラレーション処理開発ツール [SDAccel™](#) と [Machine Learning Suite](#) はお客様アルゴリズムの実装をサポートし早期市場投入のお役に立ちます。

ハイライト

高速・最も早いパフォーマンス

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

適応性・高レベルのデザインを加速

- 同じアクセラレータ カードが、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し、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速く適応するため、リコンフィギュラブル ハードウェアを使用

クラウド ↔ オンプレミス に移行可能

- クラウドで評価したソリューションをアプリケーションに合せオンプレミスに相互移行可能
- エコパートナーのメジャーワークロード向け [アプリケーション](#) とお客さま [アルゴリズム開発環境](#) をご用意



U200 | U250



全ワークロード要件に対応

- 高速データベース検索・分析
- 金融向け低遅延処理(HFT)
- 機械学習
- ストレージ高速化・圧縮/伸長
- 動画像処理
- トランスコーディング
- ゲノム解析、等

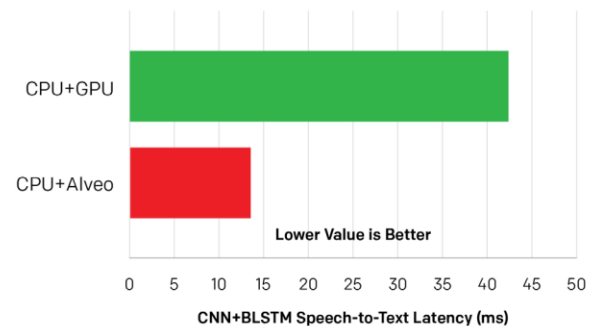
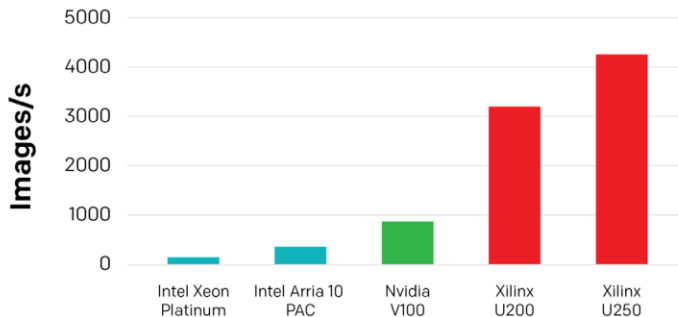
1: BlackLynx Elasticsearch on Alveo versus EC2 c4.8xlarge
2: Based on CapEx & OpEx savings for DNN inference on Alveo vs dual-socket Intel Xeon Platinum servers
3: Accelerating DNNs with Alveo Accelerator Cards White Paper
4: Measured on CNN+BLSTM Speech-to-Text ML inference against Nvidia P4

ベンチマーク ～全ワークロード要件に対応～

Market	Partner	ALVEO Acceleration vs CPU
DBサーチ・分析	BlackLynx	Unstructured Data Elastic Search 90X
金融・コンピューティング	Maxeler	Value-at-Risk (VAR) Calculation 89X
マシンラーニング	XILINX	Real Time Machine Learning 20X
画像処理・トランスコーディング	NGCodec	HEVC Video Encoding 12X
ゲノム	Falcon	Genome Sequencing 10X

CPU Comparisons: 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

Increase Real-Time Machine Learning* Throughput by 20X Reduce ML Inference Latency by 1/3



FEATURES	ALVEO U200 Accelerator Cards	ALVEO U250 Accelerator Cards
INT8 ピーク性能	18.6 TOPS	33.3 TOPS
DDR メモリ帯域幅	77GB/s	77GB/s
内蔵 SRAM 帯域幅	31TB/s	38TB/s
LUT数	892,000	1,341,000
Thermal Operations	強制・受動	強制・受動

更なる情報は www.xilinx.com/alveo サイト若しくは 弊社営業/技術にご連絡頂けますよう宜しくお願い致します。
[Nimbix cloud services](#) で評価が可能、オンプレミス評価用のAlveo [U200](#) or [U250](#) は此方から購入可能です。