



ザイリンクス Alveo ポートフォリオの拡大

Alveo U50 の発表

➤ Alveo プラットフォームの拡大

- ザイリンクス Alveo U50 の発表 - クラウド データセンターおよびエッジ データセンター向けの演算、ネットワーク、ストレージ ワークロードに対応する業界初の適応型アクセラレータ
- 幅広い重要なデータセンター アプリケーションにおいてスループット、レイテンシ、電力効率を劇的に向上
- ソフトウェア パートナーが提供する Alveo エコシステムを拡充し、Alveo ソリューションをスケールアップするための開発者ツールを継続的に強化

ALVEO™



高速

高スループット、超低レイテンシが求められる産業向けに構築

演算、ネットワーク、ストレージのアクセラレーションを実現



適応性

最適化されたドメイン特化アーキテクチャを運用
アルゴリズムの変更に対応



アクセシブル

クラウドまたはオンプレミスで運用可能
高速化用アプリケーションを豊富に提供中



➤ Alveo U50: クラウド データセンターおよび エッジ データセンター向けに構築された 業界初の適応型アクセラレータ

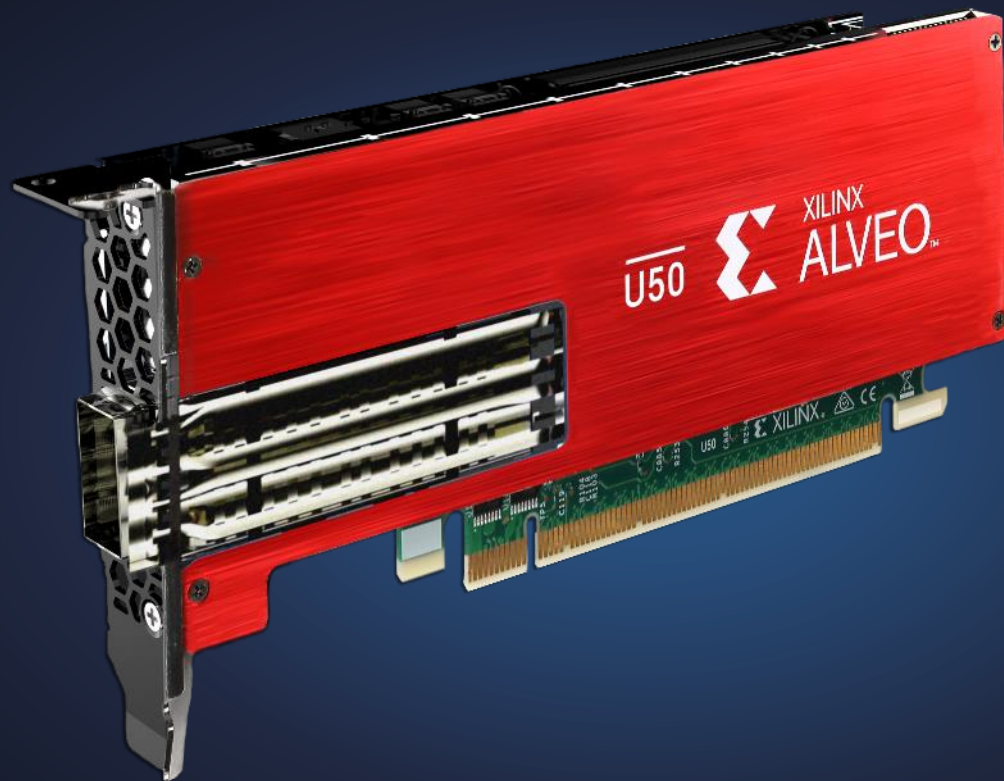
クリティカルなデータセンター アプリケーションの
パフォーマンスと電力効率が 10 倍以上向上

演算、ストレージ、およびネットワークの
アクセラレーションに最適な最も先進的かつ
適応性に優れたプラットフォーム

あらゆるサーバー、あらゆるクラウドに対応



➤ ザイリンクス Alveo U50 の主な仕様



UltraScale+ アーキテクチャ

ロープロファイルフォームファクター

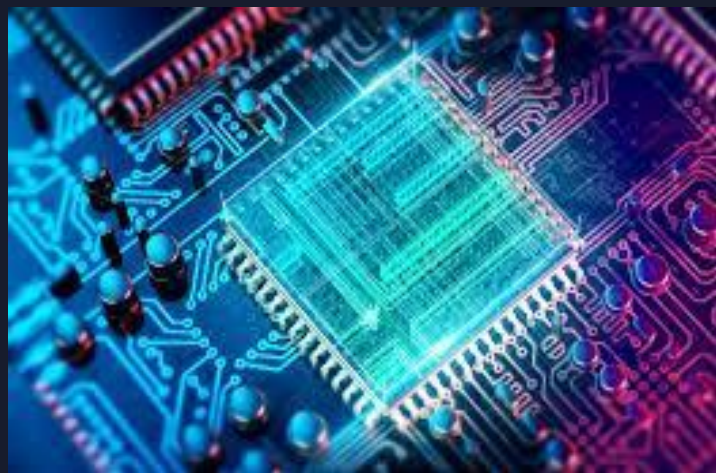
8GB HBM2 メモリ、460GB/秒

PCIe Gen4、CCIX、PCIe Gen3

QSFP 28 (100GbE)

< 75W

データセンターのアクセラレーション



演算



ネットワーク



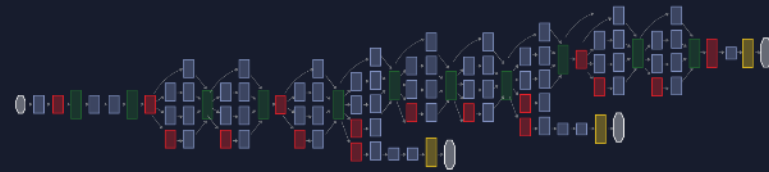
ストレージ

ドメイン特化のアーキテクチャ

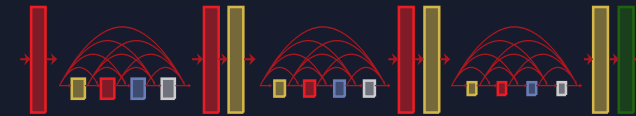
AlexNet



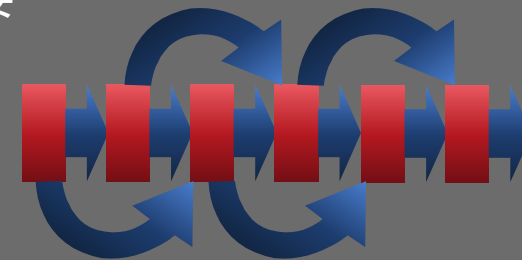
GoogLeNet



DenseNet



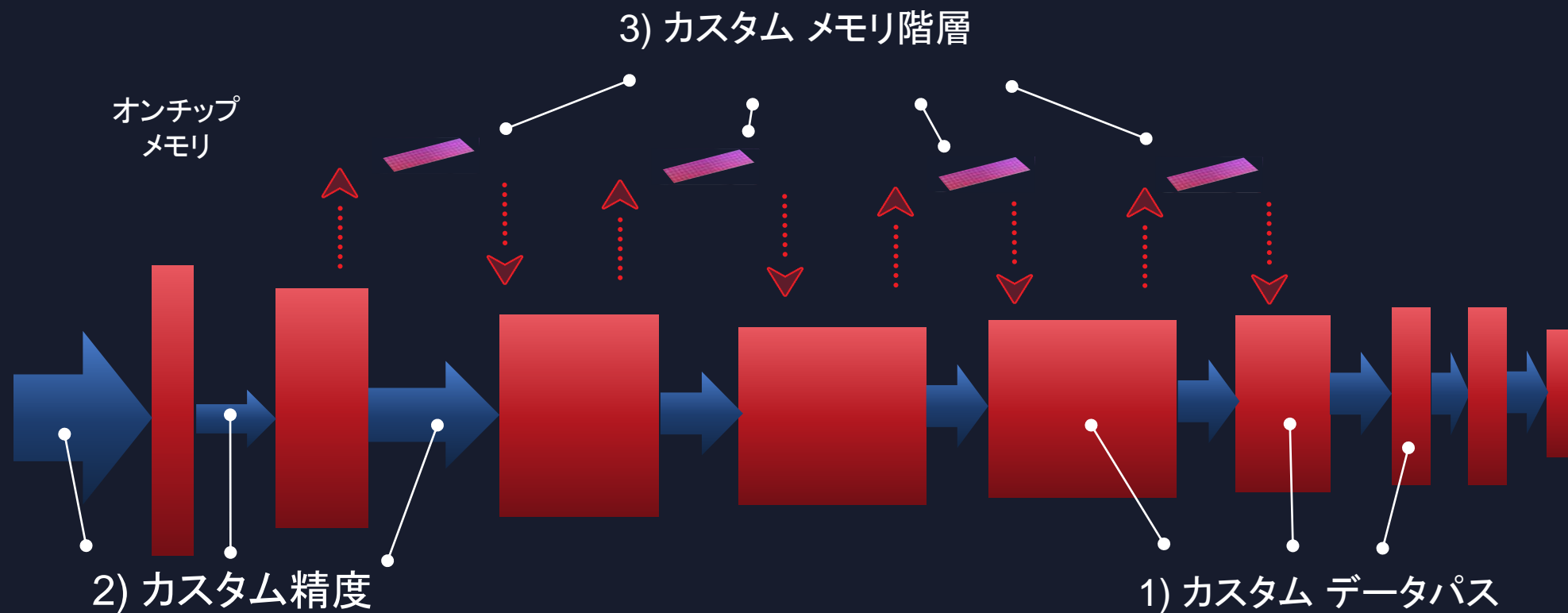
最大のスループット、最小のレイテンシ、最高の電力効率には
さまざまな HW アーキテクチャが必要



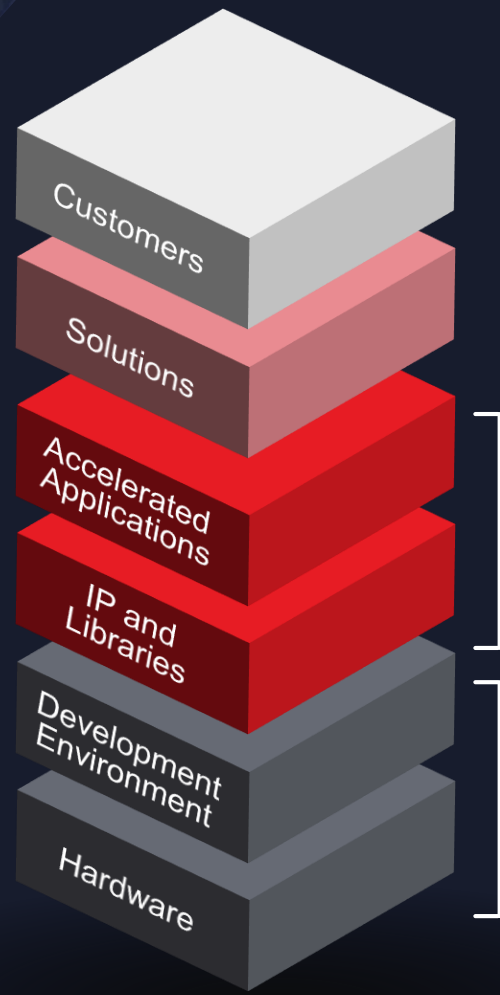
最適化されたパフォーマンス

カスタム メモリおよびカスタム データパスが必要

オフチップ
DDR



Alveo ソリューション スタック



	Data Analytics	Video & Image Processing	Machine Learning	Life Science & HPC	Financial Computing
End Customers					
Solution Providers	Titan IC	NGCODEC	edico genome <small>an Illumina company</small>	MAXELER Technologies <small>HIGH PERFORMANCE COMPUTING</small>	DGEPhi 深盛科技 PLUNIFY
App & IP Developers	XELERA	CTACCEL DEEPOLY	alcom COMPUTING	Mipsology BLACKLYNX	byte LAKE VITESSE DATA
	boon V-NOVA	skreens Nextera Video	swarm64 bigstream	ALGO-LOGIC I-ENIAC inaccel	MEGH COMPUTING BigZetta Systems LegUp mle
Channel Partners	aws		Tencent Cloud	NIMBIX SUPERMICR	inspur DELL
	Alibaba Cloud	Baidu 百度			
	CLOUD				ON-PREMISE

➤ 拡大するクラウドの利用可能地域



ザイリンクス製品ベースのハイパースケール クラウド データセンター

▶ 拡大するエコシステム

Data Analytics

Life Sciences & HPC

Video Processing

Machine Learning

Financial Computing

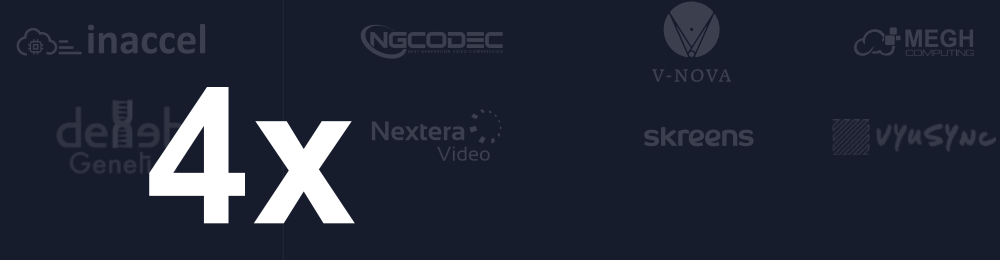
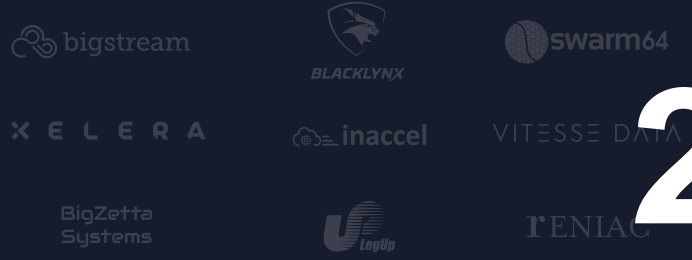
Image Processing

➤ 2018年10月リリースからの進展

Data Analytics

Life Sciences & HPC

Video Processing



2x

4x

アプリケーション数

トレーニングを受講した開発者

Machine Learning

Financial Computing

Image Processing



➤ スケールに対応する運用スタック

ALVEO CONTAINERIZED
APPLICATIONS



KUBERNETES
ORCHESTRATION



XILINX ALVEO



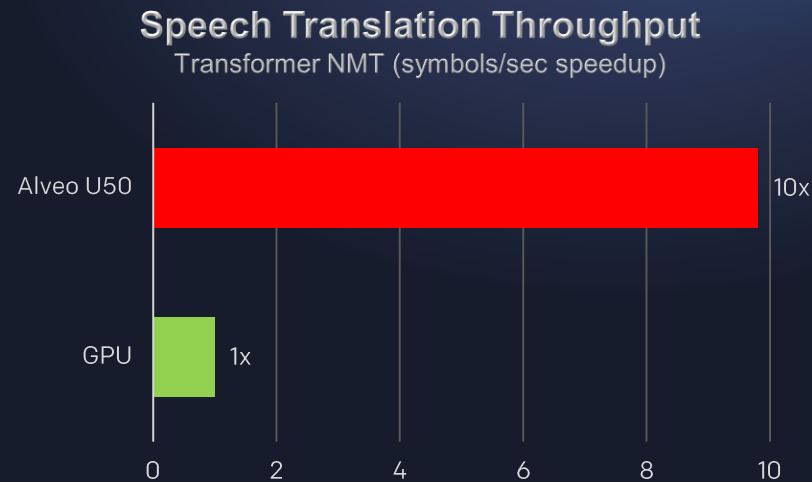
CLOUD ALVEO INSTANCES

ON-PREM ALVEO SERVERS

音声翻訳

高スループットで低レイテンシの推論

CPU や GPU では実現不可能な
高スループットで低レイテンシの推論性能



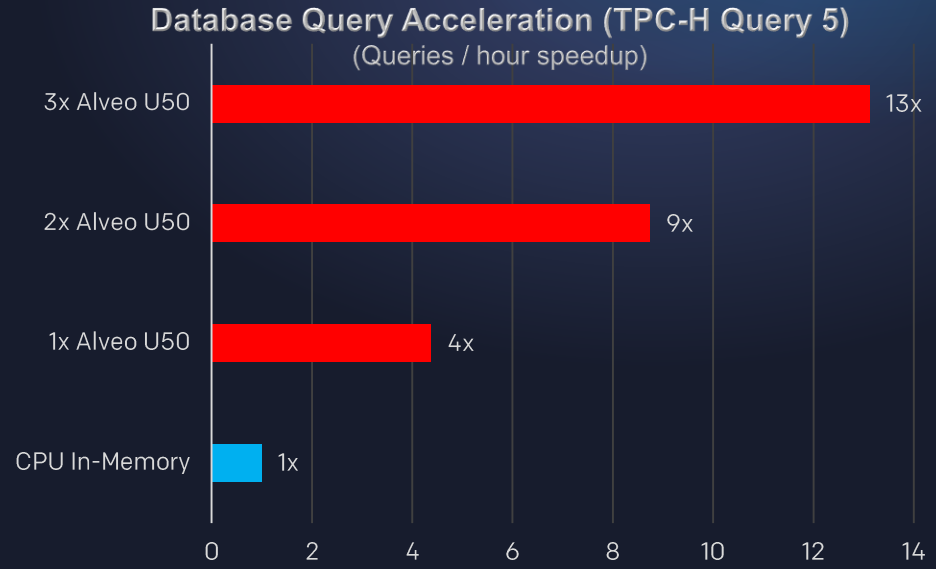
推定データ: Alveo U50 (B=2, L=8)、Tesla T4 (B=8, L=8)

<https://medium.com/syncedreview/tsinghua-university-publishes-comprehensive-machine-translation-reading-list-c3f2df594218>

データベース分析

高スループットのクエリ処理

より高いクエリ スループットと
シンプルになるインフラ構成



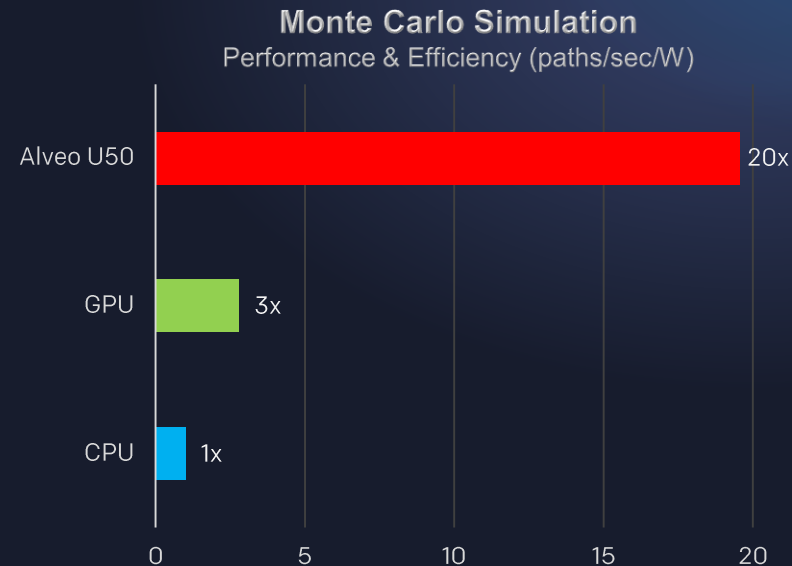
INTEL® XEON® PLATINUM 8260 PROCESSOR (35.75M Cache, 2.40 GHz) 24 core

CPU Query time = 210ms, 34k query/hr. Alveo U50=24ms, 150k query/hr

金融市場モデリング

超効率的なデリバティブプライシングおよびリスクモデリング

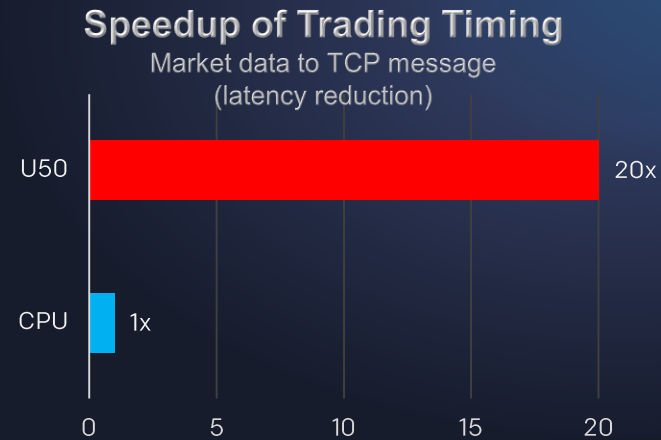
数分の一のコストで、より短時間かつ効率的にデータの抽出が可能



電子取引オペレーション

超低レイテンシのネットワークおよび演算性能

確定的なパフォーマンスで 500ns 未満のレイテンシ



Source: Xilinx Analysis

T2T latency is <0.5usec. Measured from Start of Packet in on Tick (Market Data) to Start of Packet Out on the Order

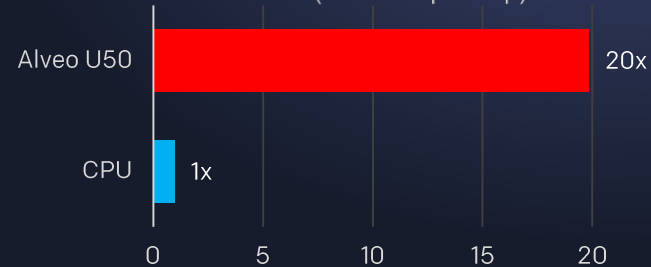
コンピュテーショナル ストレージ

ラインレート データ圧縮の高速化

1つのプラットフォームで圧縮、解凍、消去コード、暗号化すべてに対応



GZIP Compression Throughput
(GB/sec speedup)



Intel Skylake-SP 6152 @2.10GHz 22-core CPU (Ubuntu 16.04), GB/s compression per CPU core = .0229. Alveo U50 = 10GB/s

Hadoop アクセラレーション

ラインレート データ圧縮の高速化

- 20x ノードあたりのスループット
- 2x 少ないノード
- 40% さらなる総コストの削減

Alveo U50
アクセラレーション



2X デュアル CPU サーバー
192Tb、ノードあたり 1GB/秒の
圧縮スループット

2つの Alveo U50 を搭載した Alveo サーバー
96TB (192TB 有効)、
ノードあたり 20GB/秒の圧縮スループット

Intel Skylake-SP 6152 @2.10GHz CPU (Ubuntu 16.04), GB/s compression per CPU core = .0229. Alveo U50 = 10GB/s, Assume 2:1 compression

重要なポイント

市場機会の拡大

- 演算、ネットワーク、およびストレージ向けの初の適応型アクセラレータで、あらゆるサーバー、あらゆるクラウドに対応

性能および TCO における利点

- スループット、レイテンシ、電力効率が 10 ~ 20x 向上
- HBM2 および 100Gbps ネットワークポートを備える、初めての PCIe Gen 4 サポート

シンプルな プログラミング および展開

- ソフトウェアパートナーとアクセラレーションソリューションのエコシステムの拡大
- 開発者ツールとスケールに対応する運用スタックの強化

付録

➤ ザイリンクス Alveo 製品ラインナップ

ALVEO™ U50



ALVEO™ U200



ALVEO™ U250



ALVEO™ U280



UltraScale+ アーキテクチャ

UltraScale+ アーキテクチャ

UltraScale+ アーキテクチャ

UltraScale+ アーキテクチャ

872k LUT

1,182k LUT

1,728k LUT

1,304k LUT

シングル スロット、ハーフ ハイト

デュアル スロット、フル ハイト

デュアル スロット、フル ハイト

デュアル スロット、フル ハイト

8GB HBM2、460GB/秒

64GB DDR、77GB/秒

64GB DDR、77GB/秒

8GB HBM2、460GB/秒

PCIe Gen3、Gen4、CCIX

PCIe Gen3

PCIe Gen3

PCIe Gen3、Gen4、CCIX

1x QSFP 28 (100GbE)

2x QSFP 28 (100GbE)

2x QSFP 28 (100GbE)

2x QSFP 28 (100GbE)

< 75W

< 225W

< 225W

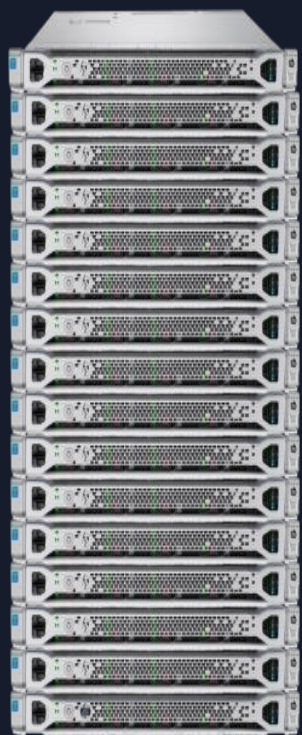
< 225W

Alveo 1 つで幅広い用途に対応

	CPU (逐次処理)	GPU (並列処理)	Alveo (逐次処理 + 並列処理)
サードパーティアプリケーション	●	●	●
高レベルコーディング	●	●	●
複雑なメモリとデータパス			●
適応型ハードウェア			●
AI 推論 + プレ/ポスト プロセス			●
オンボード ネットワーク			●

▶ ライブ ビデオ トランスコーディング

シンプルで、より低コストのインフラ



Alveo U50 HEVC ビデオ圧縮



20x ノードあたりのスループット

8x HW コストの削減

23x 電カコストの削減



20x デュアル CPU サーバー

40X Xeon Gold

H.265 超高品質

20x 1080p30

1つの Alveo U50 サーバー

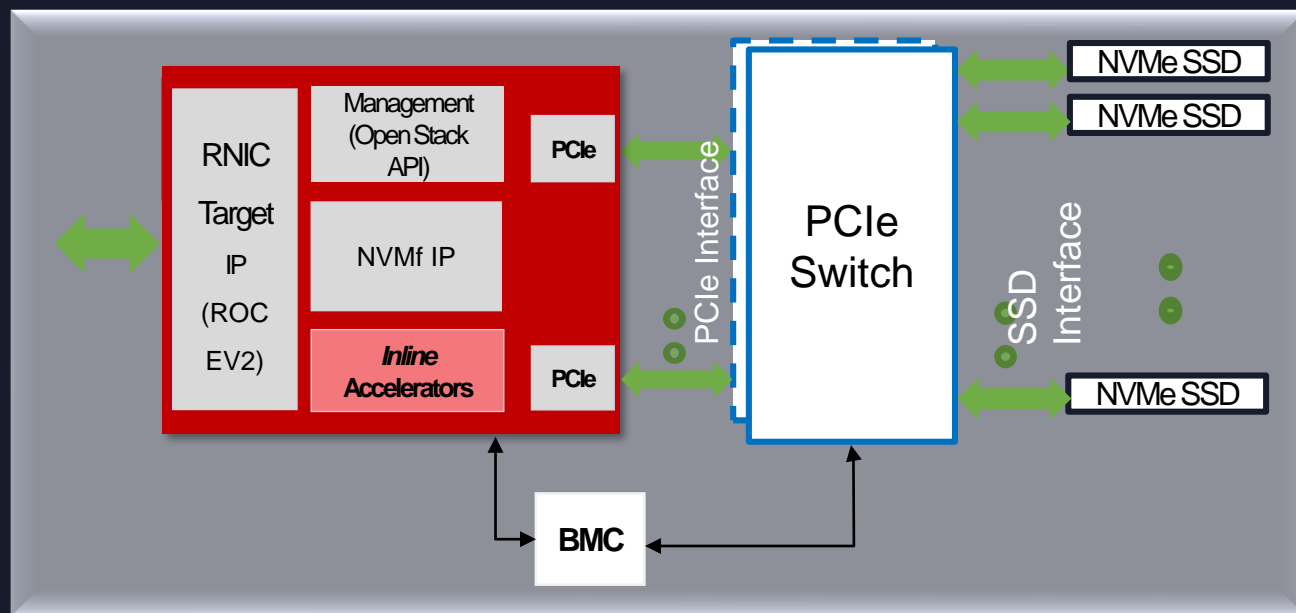
5x Alveo U50

NGCodec HEVC 超高品質

20x 1080p30

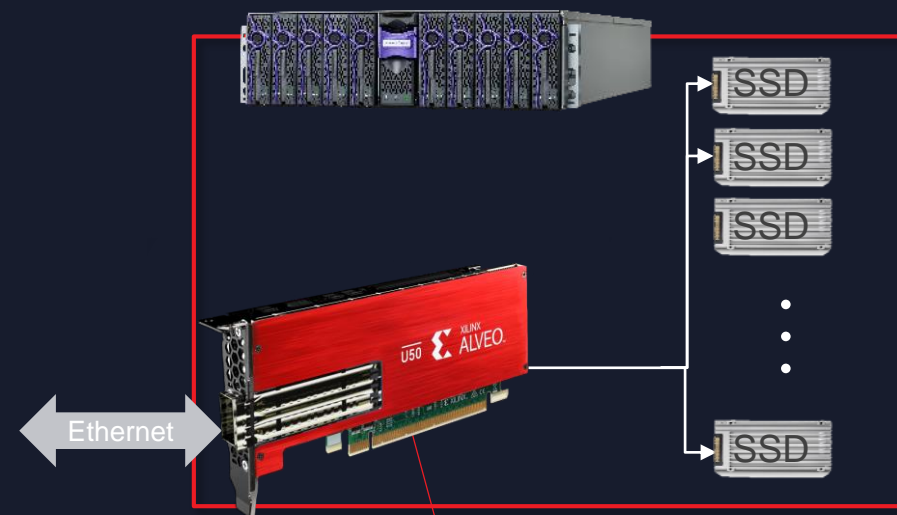
ファブリック上のコンピューテーショナル ストレージ

NVMeoF を利用した高速化を実現



- > SSD にファブリックに接続されたアクセラレータ - データを演算に活用
- > NVMeoF ターゲットは U50 にオフロードし、250 万 IOPS をサポート
- > プログラマブルなインライン ストレージ アクセラレータに対してわずか 1 マイクロ秒の追加レイテンシ

テクノロジープレビュー:
ファブリック上のコンピューテーショナル
ストレージアレイ



インライン アクセラレータの例:
ストレージ サービス:

- ・ 圧縮/解凍
- ・ 暗号化/復号化
- ・ データ保護

データベース アクセラレーション:

- ・ スキャン
- ・ フィルター
- ・ 集計