ALTIMA
A Macnica Division Company

データセンタインフラの「いま」と「これから」

2019/3/28

取扱い製品(担当)

Network

- ストレージネットワークに最適な、**低遅延**、**高機能バッファ搭載**スイッチ
- ハイパースケールネットワーク(Cumulus)
- Smart NIC, 高機能アダプタカード(GPU-Direct, RDMA, OVS, VXLAN)
- LincX ケーブル、モジュール製品

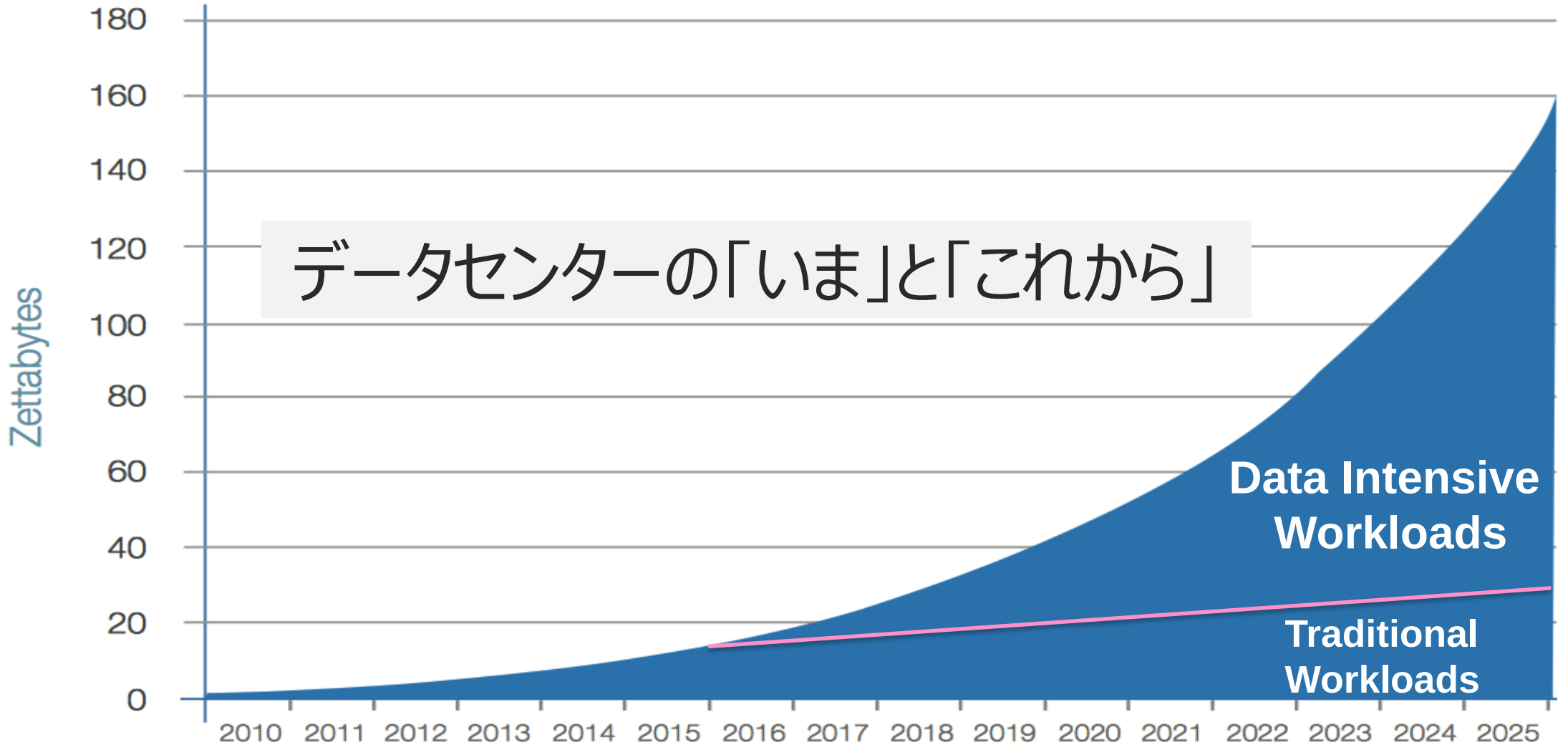
Compute

- FPGAアクセラレータカード
  Deep Learningアクセラレーション
  DataBaseアクセラレーション
- マルチコア、高密度サーバ(Xeon, AMD)
- GPU搭載サーバ
- All NVMe Flashストレージ
- 各種GPU製品
- ML/DLオンプレ統合管理ソフトウェア

Storage

- NVMe SSD
- 大容量オブジェクトストレージシステム
- JBOD/JBOF
- 次世代NVMeスケールアウトストレージ
  ソフトウェアソリューション
- コンポーザブル・インフラストラクチャ
  統合管理ソフトウェア

本日のお話



https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

ALTIMA A Macnica Division Company

Data Intensiveなアプリケーションの加速

AIを制するものが未来を制する　by ソフトバンク孫社長

AIは減衰期に突入　Hype Cycle for Emerging Technologies, 2018 from Gartner

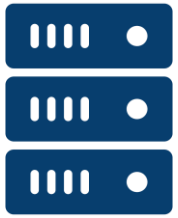
AI
- Machine Learning
- Deel Learning

Big Data
- Real Time Analytic
- Off-line Analytic

IoT
many...

Data Intensiveなアプリケーションはこれまでのインフラで十分か？？

データセンターインフラの構成要素 「いま」と「これから」

	Traditional	Data Intensive
SERVER 	CPU	**メニーコア** **GPU** **FPGA**
STORAGE 	アプライアンスストレージ スケールアップ	**Software Defined Storage** **スケールアウト、PBスケール**
NETWORK 	ベンダー製スイッチ 独自プロトコル	**ホワイトボックススイッチ** **シンプルネットワーク(標準プロトコル)** **Linux OS**

データセンターインフラの構成要素 「いま」と「これから」

- ヘテロジニアスな環境
  - CPU, GPU, FPGA, etc
- スケーラビリティのあるストレージシステム
  - SDS
  - コモディティ―ハード
  - マルチデバイスへの対応(HDD, SSD, NVMe)
- 新しいワークロードに対応できるネットワークファブリック
  - 標準プロトコル
  - ホワイトボックス > ベンダロックイン
  - エコシステム (Linux)

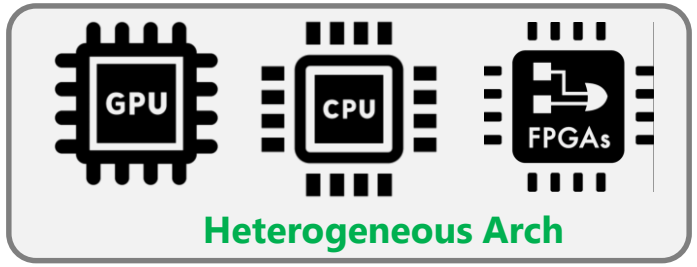
Data Intensive

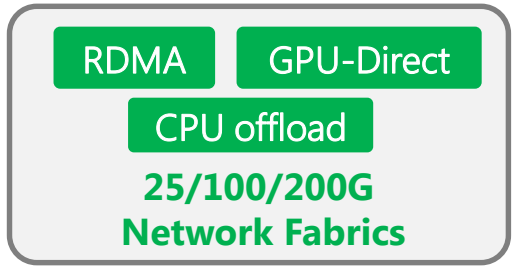
メニーコア
GPU
FPGA

Software Defined Storage
スケールアウト、PBスケール

ホワイトボックススイッチ
シンプルネットワーク(標準プロトコル)
Linux OS

「これから」のテクノロジー要件





ヘテロジニアスな コンピューティングHW

アプリケーション要件を満たすHWの選択
Ex:
マルチコアプロセッサ
マルチノードプロセッシング
各種オフロード技術

大容量、低レイテンシー + ファブリック接続

JBOD/JBOF
NVMeスケールアウトストレージ
インメモリ処理
ストレージクラスメモリ
etc

広帯域(100/200/400G)、低遅延 + Application Aware Network

ストレージトラフィック最適化
→RDMA/RoCE、ロスレス

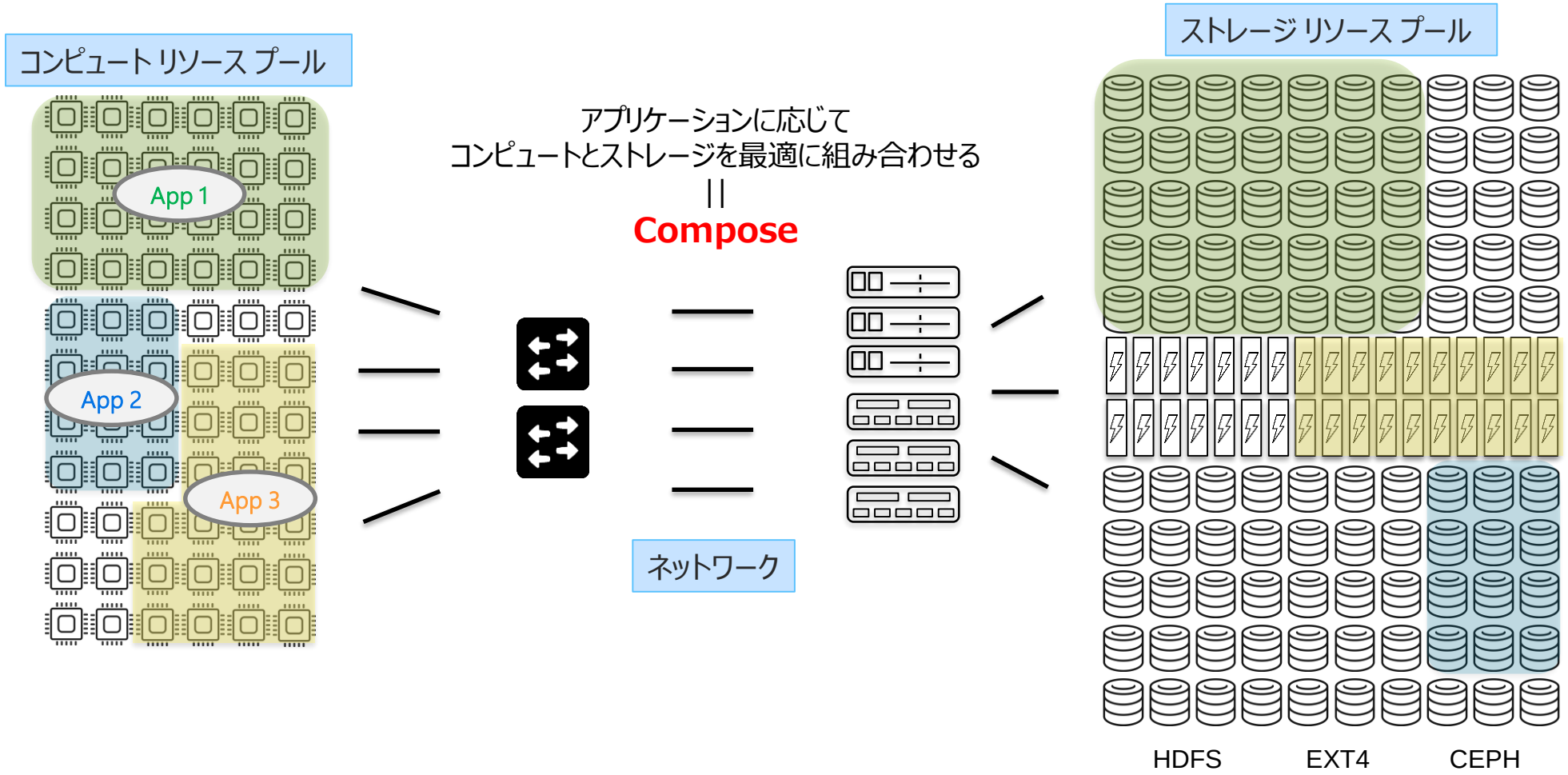
コンピューティングオフロード
→TCP/IP, RDMA, DPDK, OVS offload

スケーラビリティ
→IP CLOS、SDN(VXLAN)、etc

コンポーザブル　インフラストラクチャ

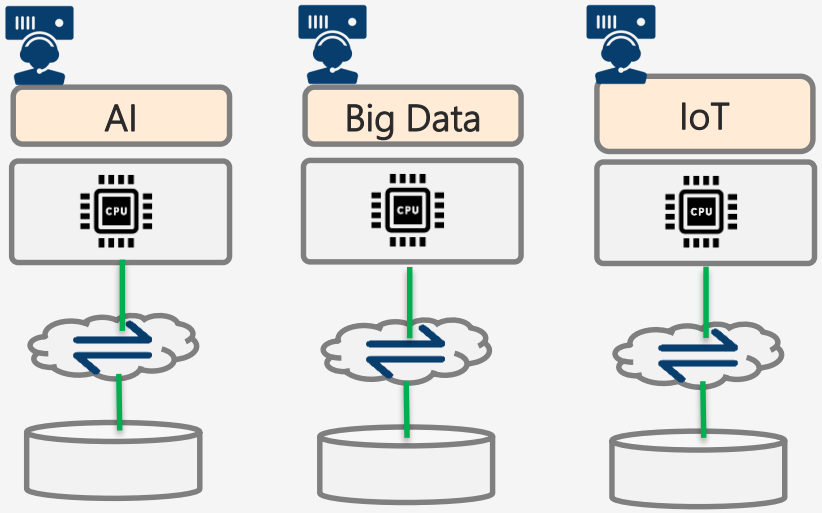
コンポーザブル インフラストラクチャとは？

Composable Infrastructure
分離されたコンピューティング、ストレージ、およびネットワークファブリックのプールを使用して、
サーバを柔軟に構成する次世代サーバ・インフラストラクチャ



データセンタデザイン 「いま」と「これから」

Now

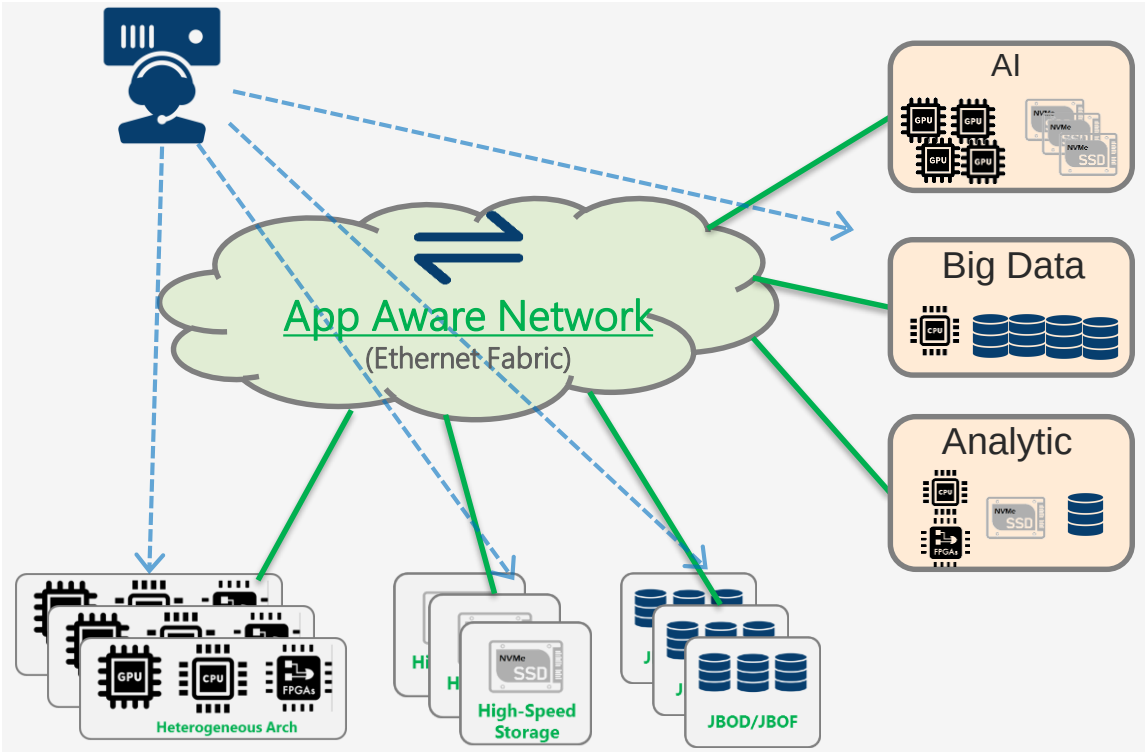


ワークロードに合わせて個々にインフラを構築・管理

× スケーラビリティ
× 管理が複雑
× 汎用性
× 運用コスト

Future

Composable + API



✓ ワークロード要件に応じたなリソース管理
✓ スケーラブルで単一コントロールプレーンによる統合管理
✓ 導入コスト、運用コスト→大幅削減

「これから」を実現するソリューション紹介

ヘテロジニアスな
コンピューティングHW

大容量、低レイテンシー
+ ファブリック接続

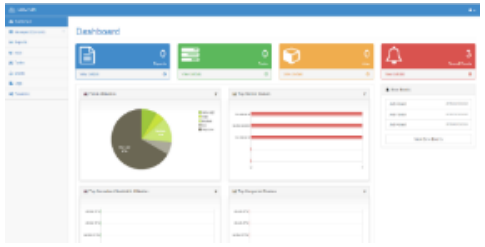
広帯域(100/200/400G)、**低遅延**
+ Application Aware Network

+

コンポーザブル　インフラストラクチャ

すべてを繋ぐ、MellanoxのEnd-to-Endポートフォリオ

Software



Switch

Spectrum™ Spectrum-2™

広帯域、低レイテンシーなデータセンタースイッチ

~400GE見据えたロードマップ

オープンプラットフォーム

Adapter

ConnectX®

10GE～200GE対応

各種アクセラレーション対応

(RoCE, OVS offload, GPU Direct, etc)

ConnectX®-5

ConnectX®-6

Interconnect

LinkX™

各種トランシーバ

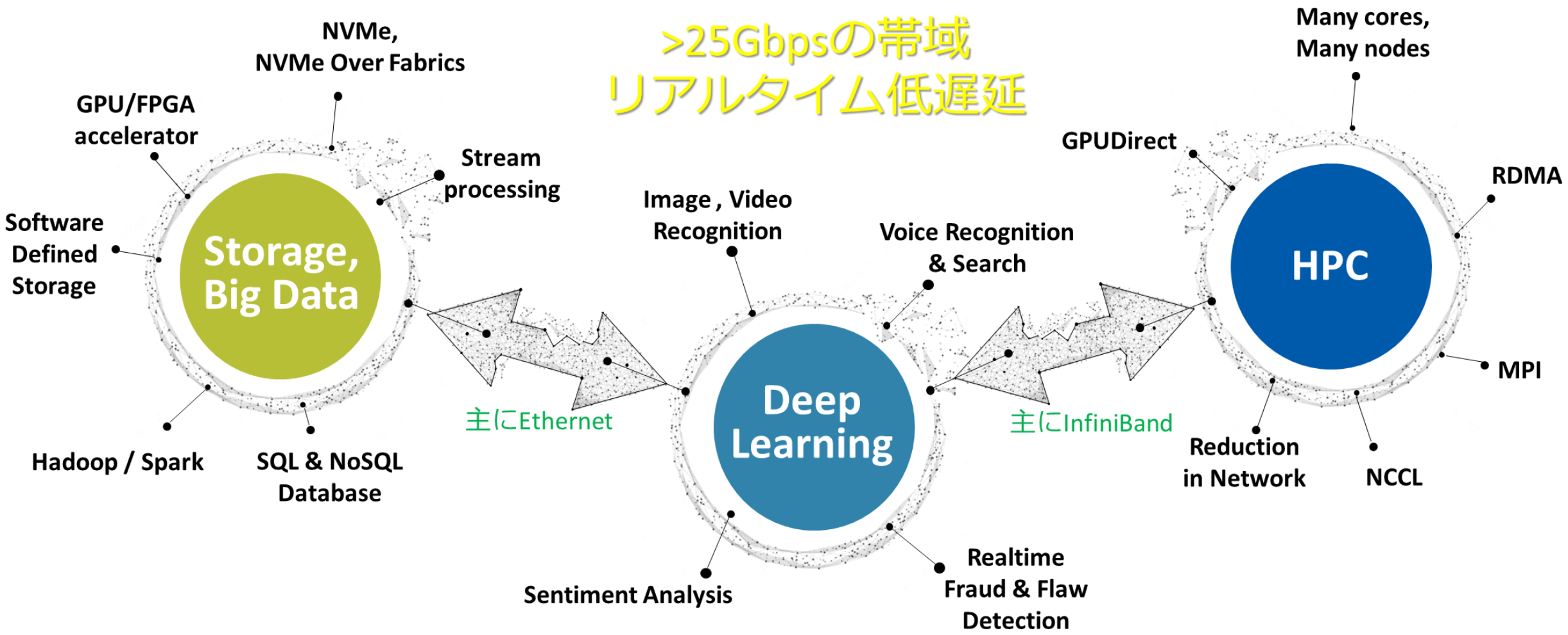
Active Optical and Copper Cables

10 / 25 / 40 / 50 / 56 / 100GbE

VCSELs and Copper

ALTIMA
A Macnica Division Company

Mellanox が成長している分野



多くのアプリケーションに高性能 Interconnect 技術が要求されている

オープンプラットフォーム & 予測可能なイーサネットファブリック

Application

Open Source Routing / puppet labs / CHEF / sFlow / ANSIBLE / Net-SNMP / ZABBIX / InfluxDB / vmware NSX / openstack

多様なLinux用OSS/商用ツールを自由に活用

NOS

cumulus

LinuxベースNOS

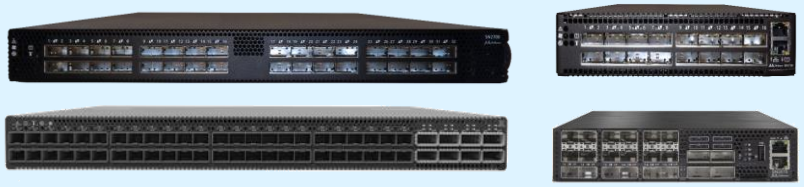
CumulusVX
ネットワークシミュレーション仮想アプライアンス

Cumulus NetQ
ファブリック全体の検証ツール、 障害を防ぎ、対応を高速化

Cumulus RoH
コンテナ・マイクロサービスに最適なホスト用ソフトウェアパッケージ

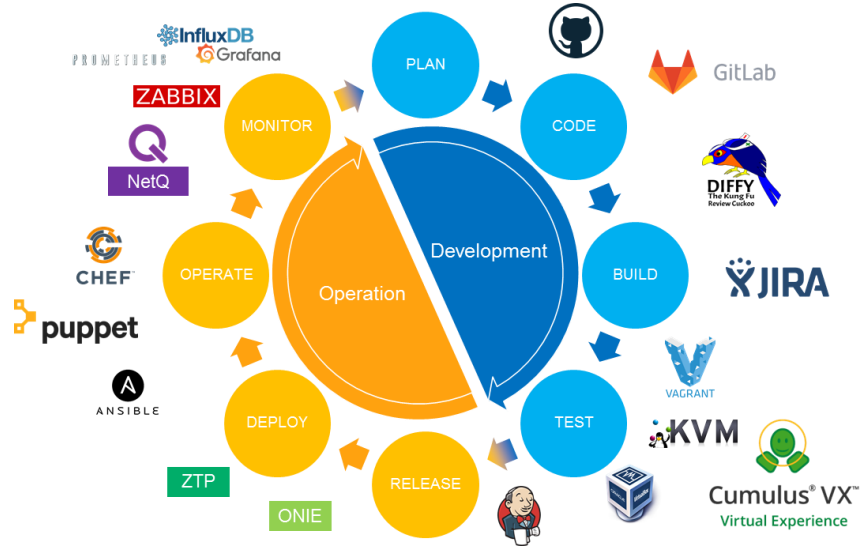
HW

Mellanox TECHNOLOGIES

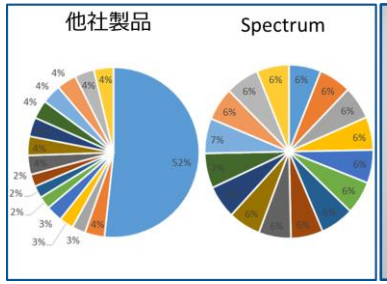


低遅延、高性能バッファリングのスイッチング性能を提供

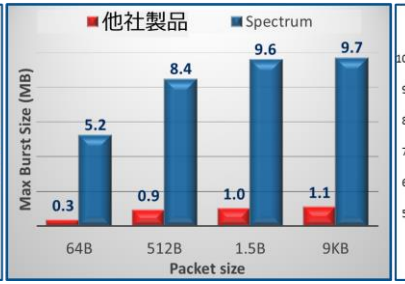
OPEXの最適化、効率化

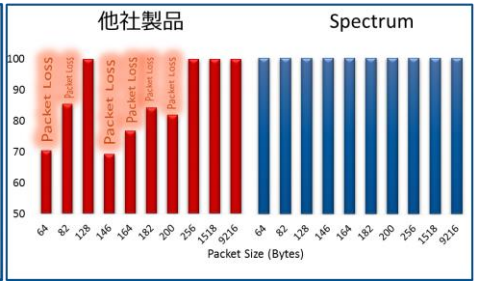


スケーラブルで高性能なアンダーレイ基盤



スマートバッファリング





低レイテンシー

ヘテロジニアスな
コンピューティングHW

大容量、低レイテンシー
+ ファブリック接続

広帯域(100/200/400G)、低遅延
+ Application Aware Network

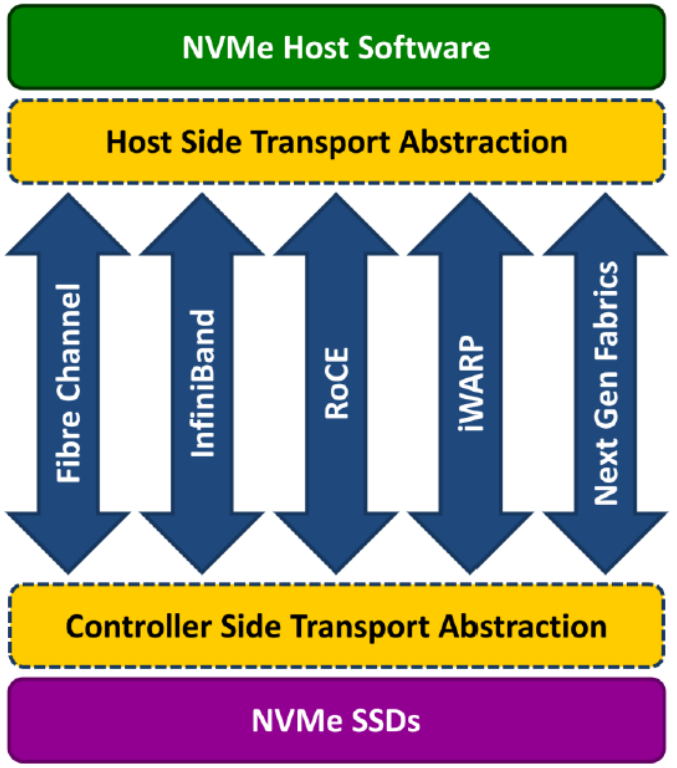
コンポーザブル　インフラストラクチャ

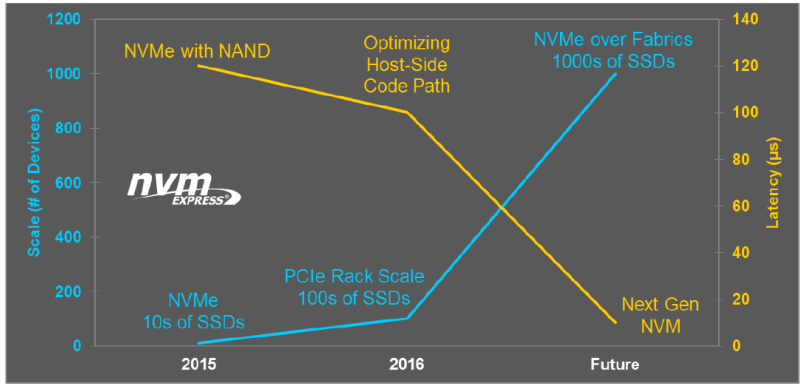
キーワードはNVMe over Fabrics

- NVMeをローカル利用する課題
  - PCIスロット以上には拡張できない
  - ノード単位の容量制限以上をアプリケーション利用できない

- NVMe over Fabricとは
  - NVMeストレージを複数ノードで共有する仕組み
    - マルチプロトコルサポート
    - データプレーン、マネジメントプレーン機能提供
  - NVMe over Fabrics標準規格化
    - NVMe over Fabrics 1st Version 1.0 (2016/June)
    - NVMe over Fabrics Latest Version 1.0a (2018/July)

- NVMe over Fabricが目指すところ
  - 低レイテンシー性能(over Fabric)
  - ファブリックスケール





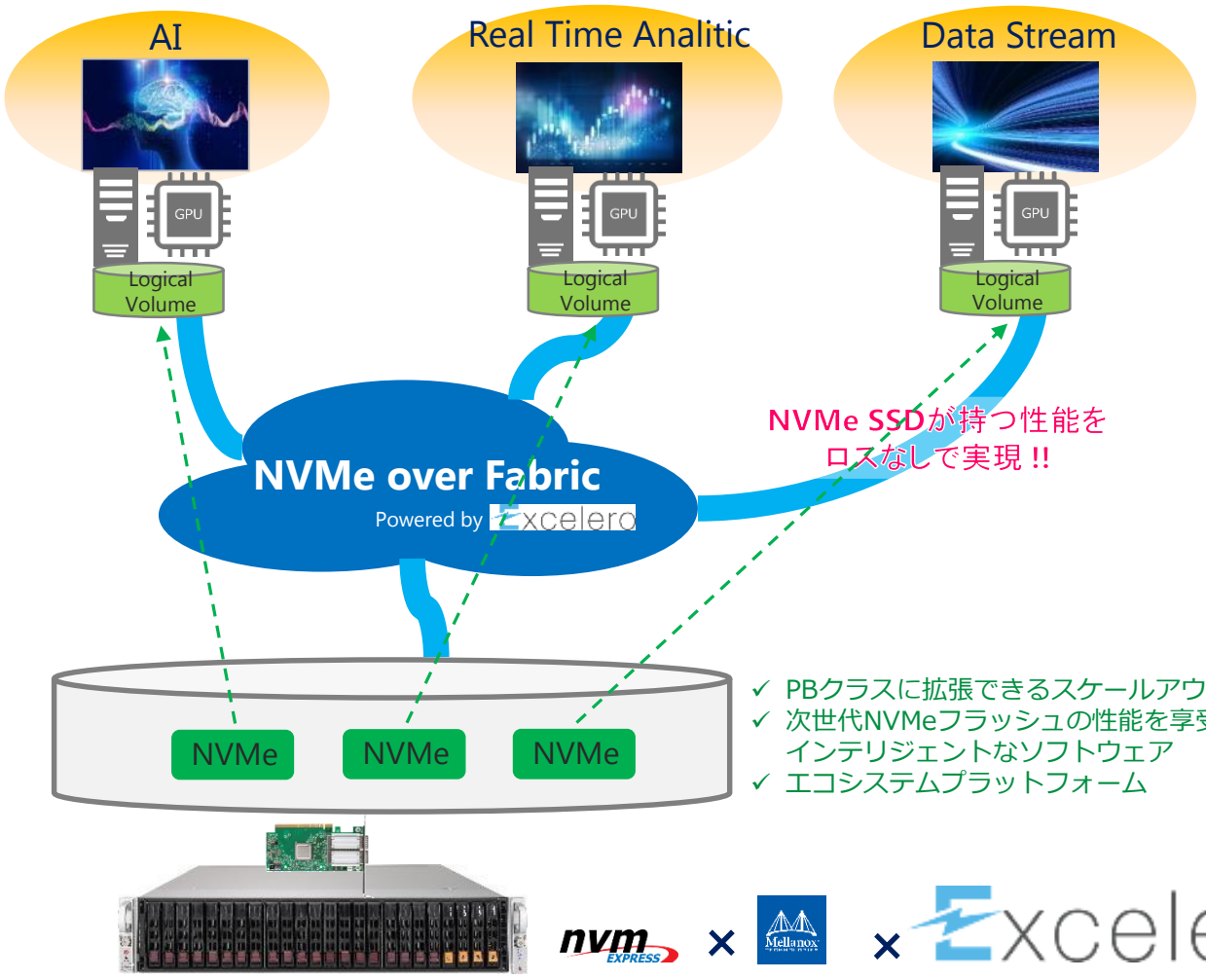
補足) NVMe ドライブの実力

- NVMe ドライブ性能の例
  - HGST社 SN200
    - 7.6TB
    - 3,350MB/s (=26.8Gbps)
    - 835K IOPS

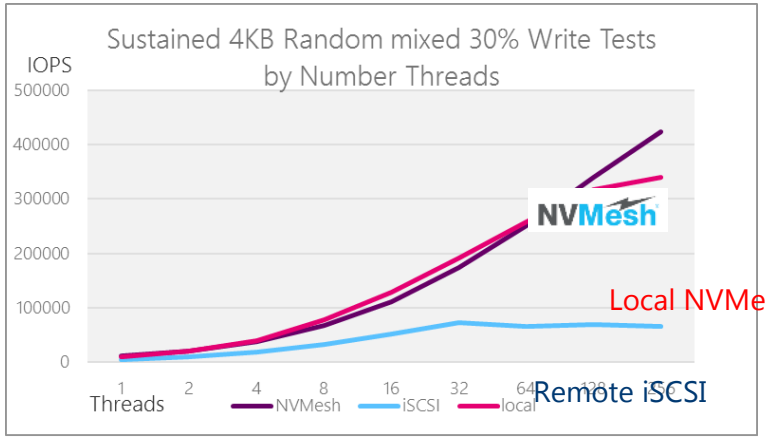
https://www.hgst.com/products/solid-state-solutions/ultrastar-sn200-series

- 4台搭載すれば 100Gbps を超え、軽く 3M IOPS を超える性能を叩き出す

Excelerоで実現するNVMeスケールアウトストレージ



- ✓ **リモート性能 = ローカル性能**
- ✓ **PBスケール**
- ✓ **エコシステムプラットフォーム**



IOPS

NVMeローカル性能に比べ性能低下ゼロ

ALTIMA
A Macnica Division Company

ヘテロジニアスな
コンピューティングHW

大容量、低レイテンシー
+ ファブリック接続

広帯域(100/200/400G)、低遅延
+ Application Aware Network

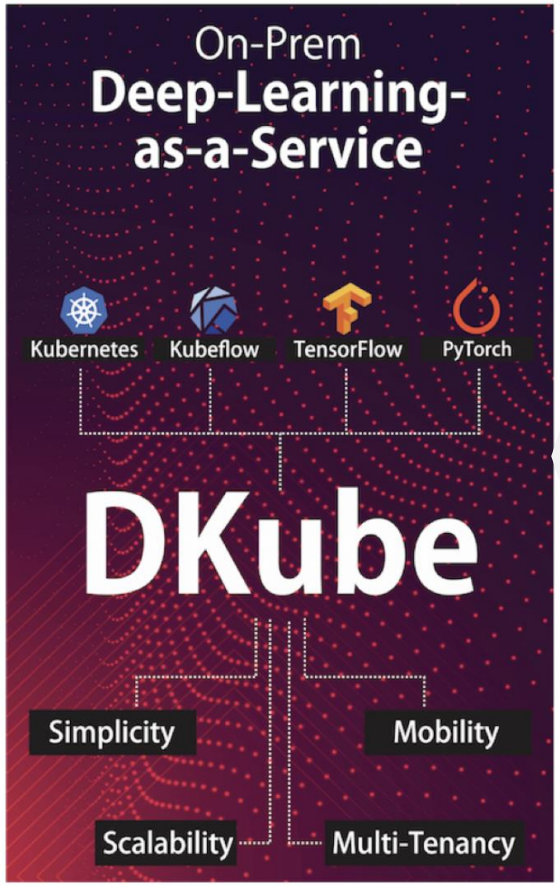
+

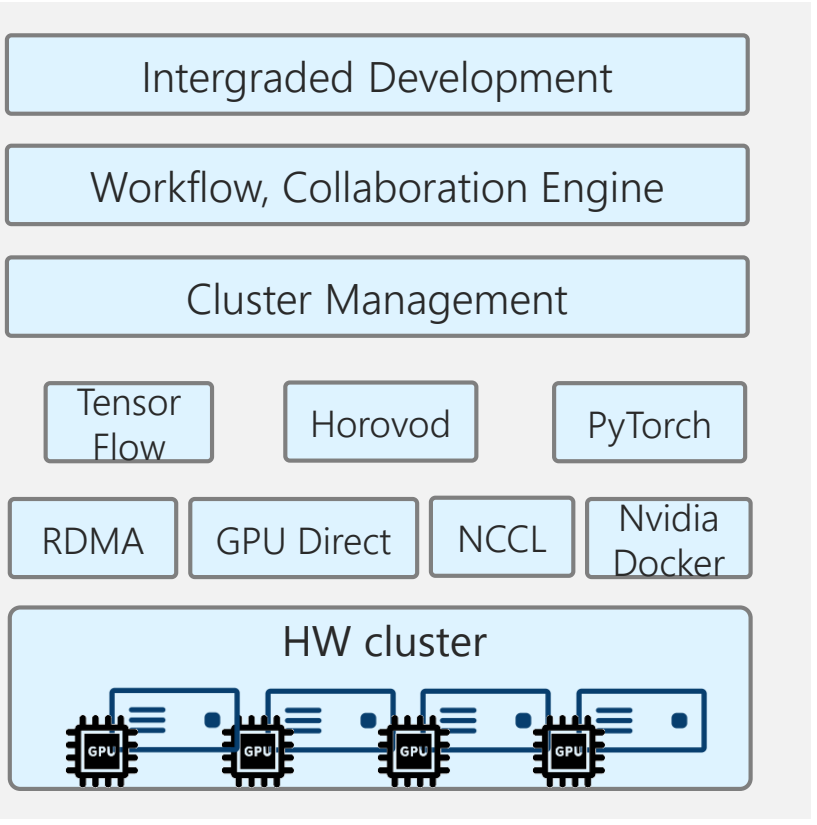
コンポーザブル　インフラストラクチャ

ヘテロジニアスなオンプレ環境の統合管理オーケストレータ

オンプレ基盤でDL-as-a-Serviceを実現するためのオーケストレーションソフトウェア

One Convergence





- プロジェクト管理
- セグメンテ分離、パーティショニング
- リソース、ワークフローのモニタリング

- GPU, マルチノード接続のクラスタ管理

- オープンソースフレームワークのインテグレーション

- ドライバ、ミドルウェアのインストールとメンテナンス

- HWクラスタの抽象化(GPU)

ALTIMA A Macnica Division Company

Dkubeはハード構成からアプリケーション構築までワンストップで提供

One Convergence

On-Prem
Deep-Learning-as-a-Service

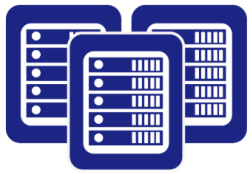
Kubernetes Kubeflow TensorFlow PyTorch

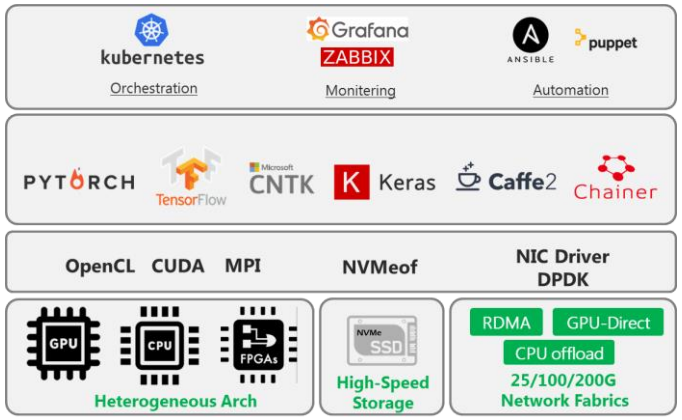
DKube

Simplicity Mobility Scalability Multi-Tenancy

+

オンプレミス





オンプレ + Cloud Agility

- ✓ オンプレの複雑さを解決
- ✓ クラウドのアドバンテージを享受

ALTIMA A Macnica Division Company

ヘテロジニアスな
コンピューティングHW

大容量、低レイテンシー
+ ファブリック接続

広帯域(100/200/400G)、低遅延
+ Application Aware Network

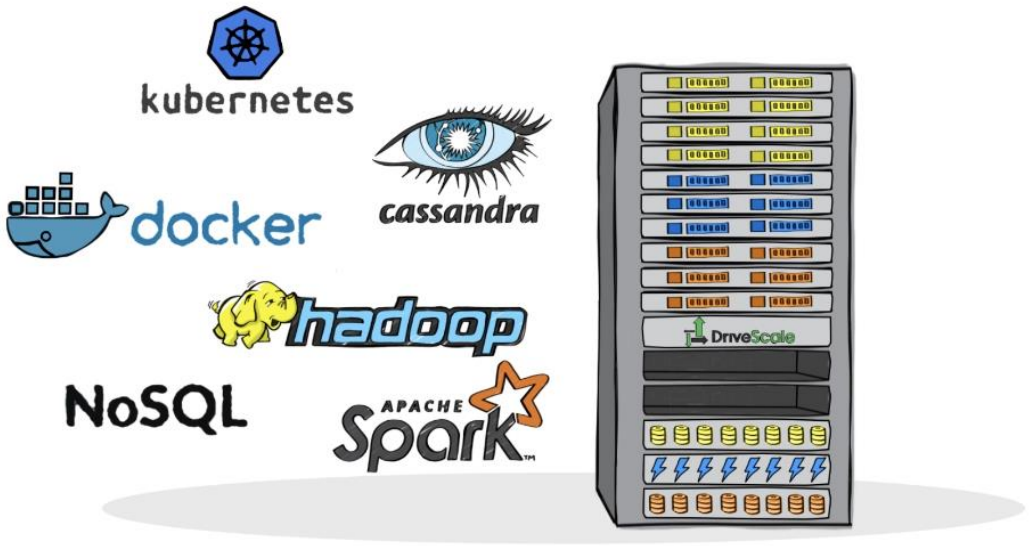
+

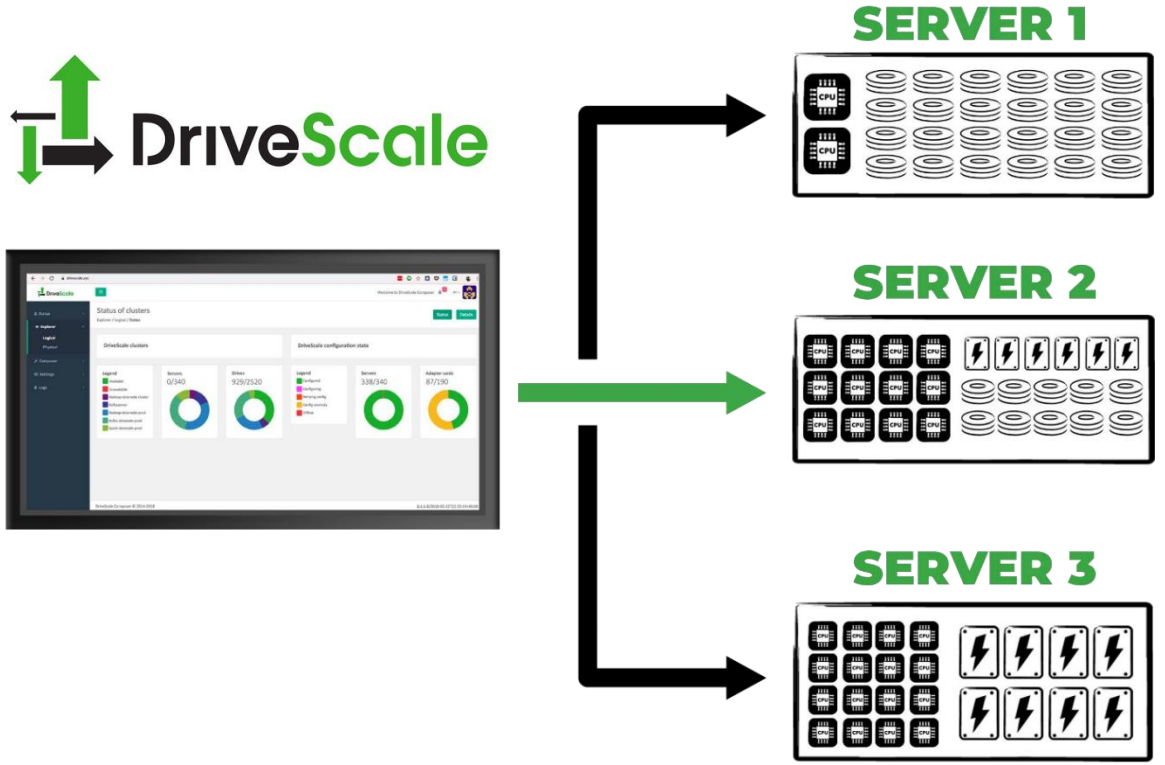
コンポーザブル　インフラストラクチャ

↑

SCIを実現するソフトウェア　DriveScale

Software-Defined Composable Infrastructure

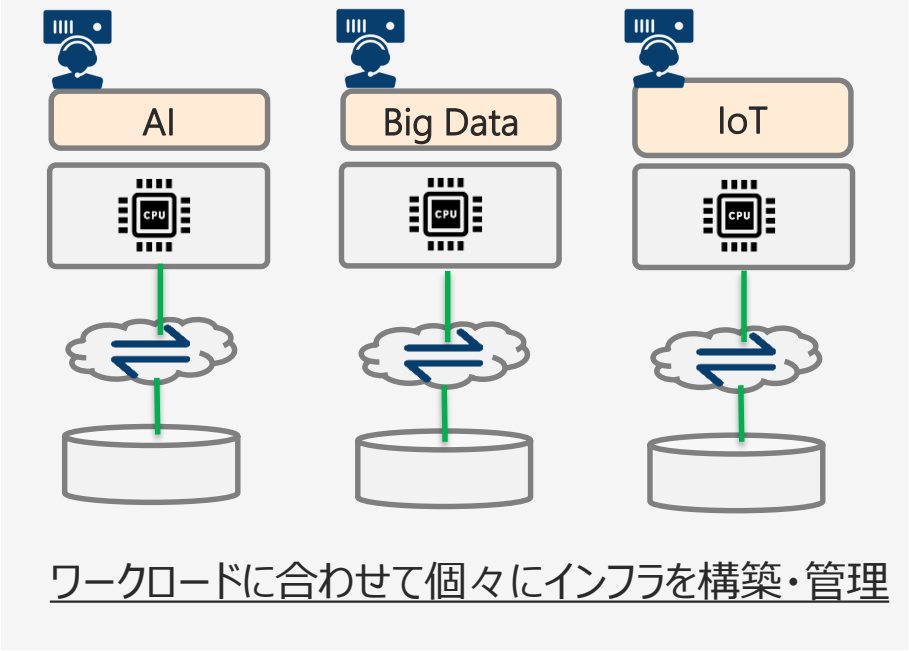




- Software Composable Infrastructure (SCI) を実現するソフトウェア
- ComputeとStorageのリソースを分離して管理可能
  - スケーラブルで柔軟なインフラの実現
- Web GUI、オープンAPIによる統合管理、監視が可能

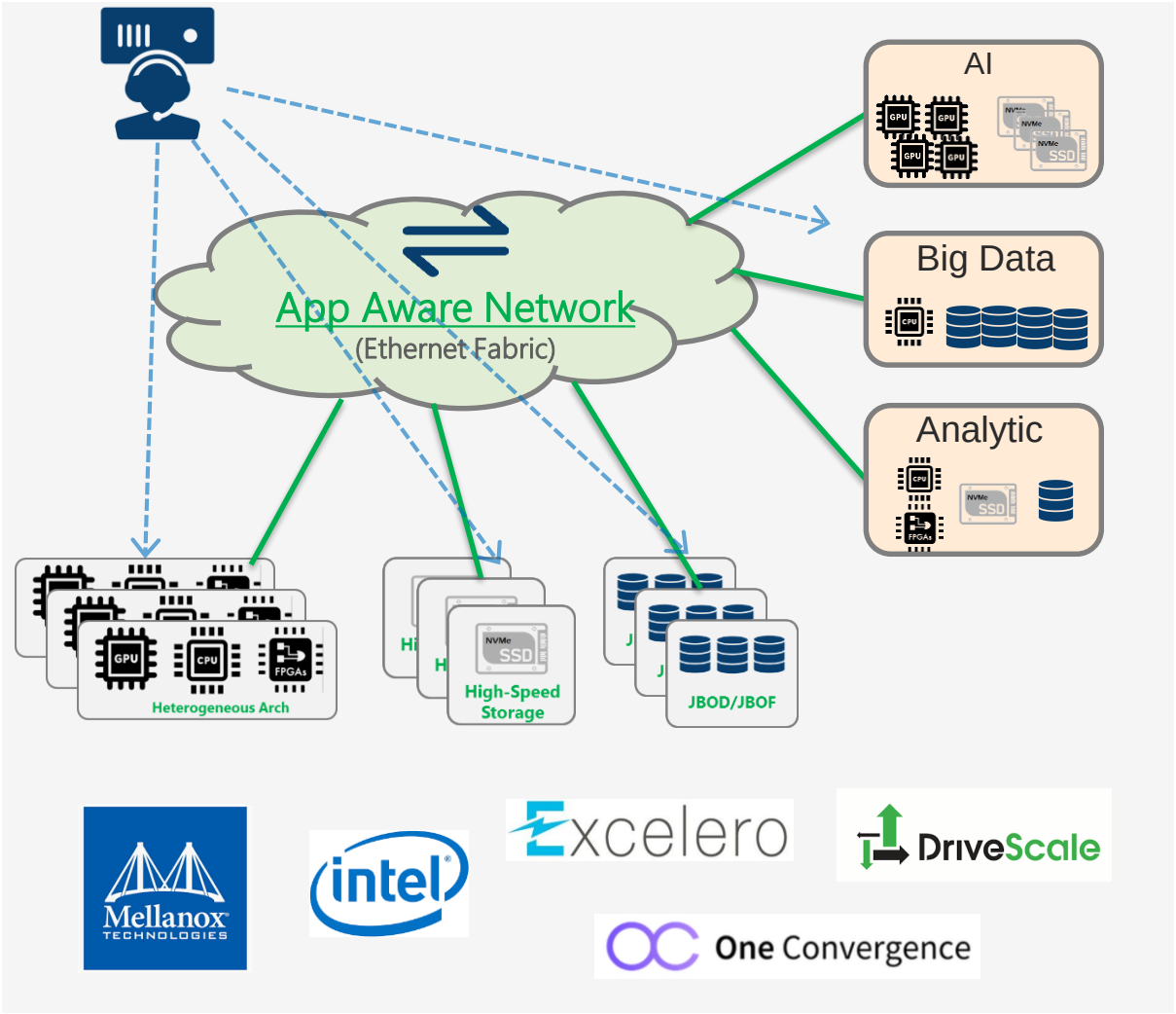
まとめ

Now



Future

Composable + API



オープンで**エコシステム**なインフラのことなら

マクニカがHW/SW含め、ワンストップで

ご提案、技術サポートをご提供

告知

4/24(水) @同セミナー会場

Cumulus x **Mellanox** セミナー開催