



ビギナーのためのストレージ講座

JDSFエデュケーション部門

開催日時

2026年8月26日(水)13:30－17:15

- 1. ストレージ概論
- 2. ストレージネットワーキングの基礎
- 3. テープストレージのテクノロジー
- 4. バックアップ概論

- 4-1.ストレージネットワークの基礎

ストレージネットワークって何？ SAN？ NAS？

- 4-2.FC-SANとその他のストレージネットワーク

ストレージネットワークの基盤となったFC(**FibreChannel**)のSANと、
その他のストレージネットワークの違い

- 4-3.その他のストレージ技術

オブジェクトストレージ SDS

- 4-4 進化を続けるストレージインターフェース技術

ストレージとの接続技術の進化 NVMe、CXL

- 4-5 ストレージネットワークの進化

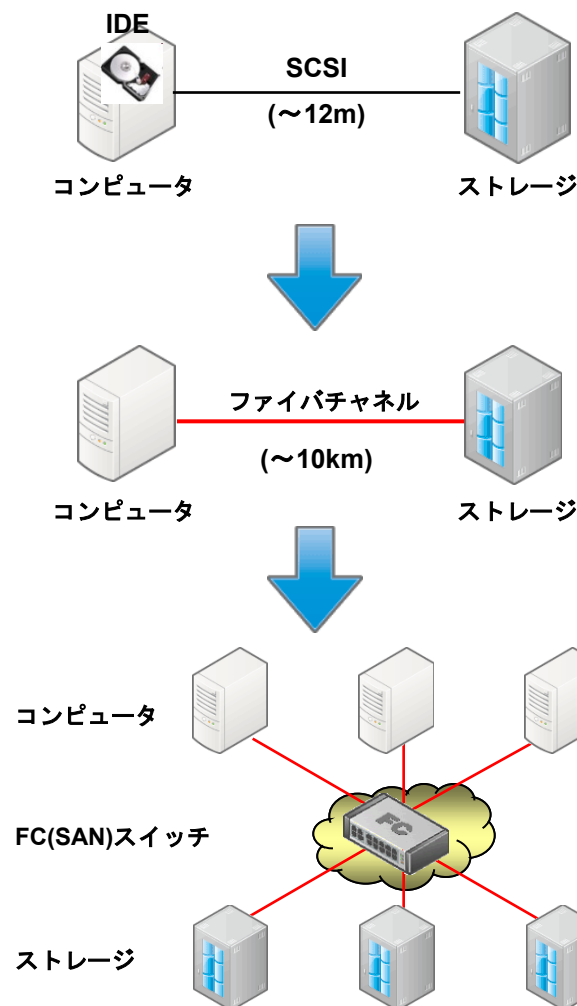
SCSIからNVMe

PCI Expressのファブリック CDI

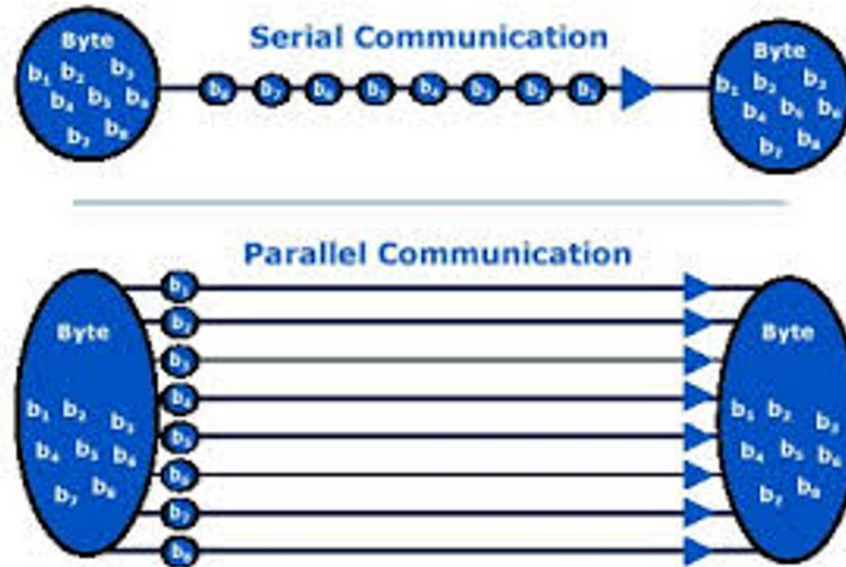
4-1. 進化を続ける接続技術

4-1-1. コンピュータとストレージの関係

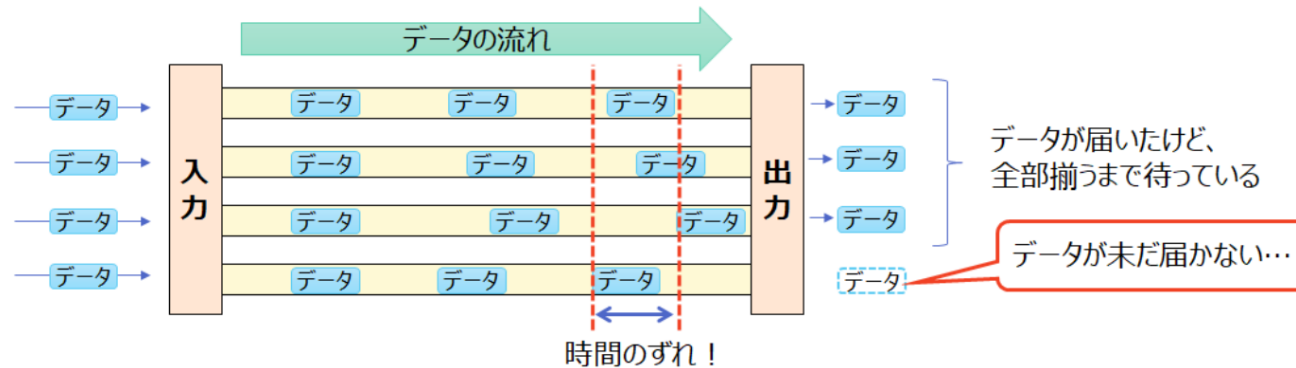
- 1980年代初め頃まで
 - ◆ DAS(Direct Attached Storage):
ホストとストレージが直接接続された状態
 - ▶ IDE (Integrated Device Electronics)/ATAPI
 - ▶ SCSI (Small Computer System interface)
 - ▶ パラレル転送方式
- 1980年代後半～1990年中盤
 - ◆ SCSIの平行アーキテクチャが性能と接続距離の限界に達する
 - ▶ コンピュータシステムの速度が向上
 - ▶ データストレージへの需要増大
- 1990年代後半以降
 - ◆ シリアル転送技術の移行
 - IDE/ATAPI → SATA (SerialATA)
 - SCSI → SAS (SerialSCSI)
 - ◆ Fibre Channelの登場
 - 柔軟性、堅牢性、接続性からSCSIを置き換えはじめた
 - ▶ シリアル転送方式
 - ▶ 光ケーブルにより接続距離を延長
 - ネットワーク上に接続し、デバイス間でストレージを共有
SANの始まり(FC-SAN)
- 2000年代
 - ◆ 高速ネットワークの浸透と進歩 10Gbps 25Gbps 40Gbps
→ **IP-SANへの拡張**
 - ◆ 拡張された技術の発展
 - ▶ iSCSI NICにオフロードエンジン搭載
 - ▶ FCoE (FibreChannel Over Ethernet) ※ **FC-SANの変形**
- 2010年代
 - ◆ SSD(フラッシュ)を使用したストレージが浸透し始める。
このため、処理能力が転送速度を圧倒的に上回る状況となり
高速なネットワークを利用して高速な転送方式が数々生まれてくる
 - ◆ iSER、IBの十八番のRDMAをネットワークで実現(RoCE、iWARP)
 - ◆ SCSIプロトコルからNVMeプロトコルへ
 - ◆ ネットワーク高速化の加速(100、200、400Gbps)
高速ネットワークを活用したストレージネットワークへ変貌



パラレルとシリアル ? ?



- コンピュータ内部(シリコン内部)は64bit並列処理・・・ でもデータを並列で送ると・・・



ややこしい表現が多くてすいませんが



本章では、“**プロトコル**”とか、“**イニシエータ**”“**ターゲット**”というワードを頻繁に使用します。

“**プロトコル**”は、OSIの階層構造に倣ったストレージネットワークでの階層構造のイメージで説明する際に各階層で出てきます

上位層	SCSI、NFS/CIFS、NVMe
通信層	TCP/IP iSCSI
物理層	SCSI FC、Ethernet Infiniband、PCI-Expressなどなど

本章では、装置(デバイス)同士が通信するための手段、方式などの取り決めをプロトコルと表しております。
つまり言語だと思ってください。

装置(デバイス)同士が会話(通信)するためには同じレイア(階層)で同じ言語で会話(通信)する必要があります。

ストレージの世界では装置(デバイス)をイニシエータとターゲットという呼び方をします。

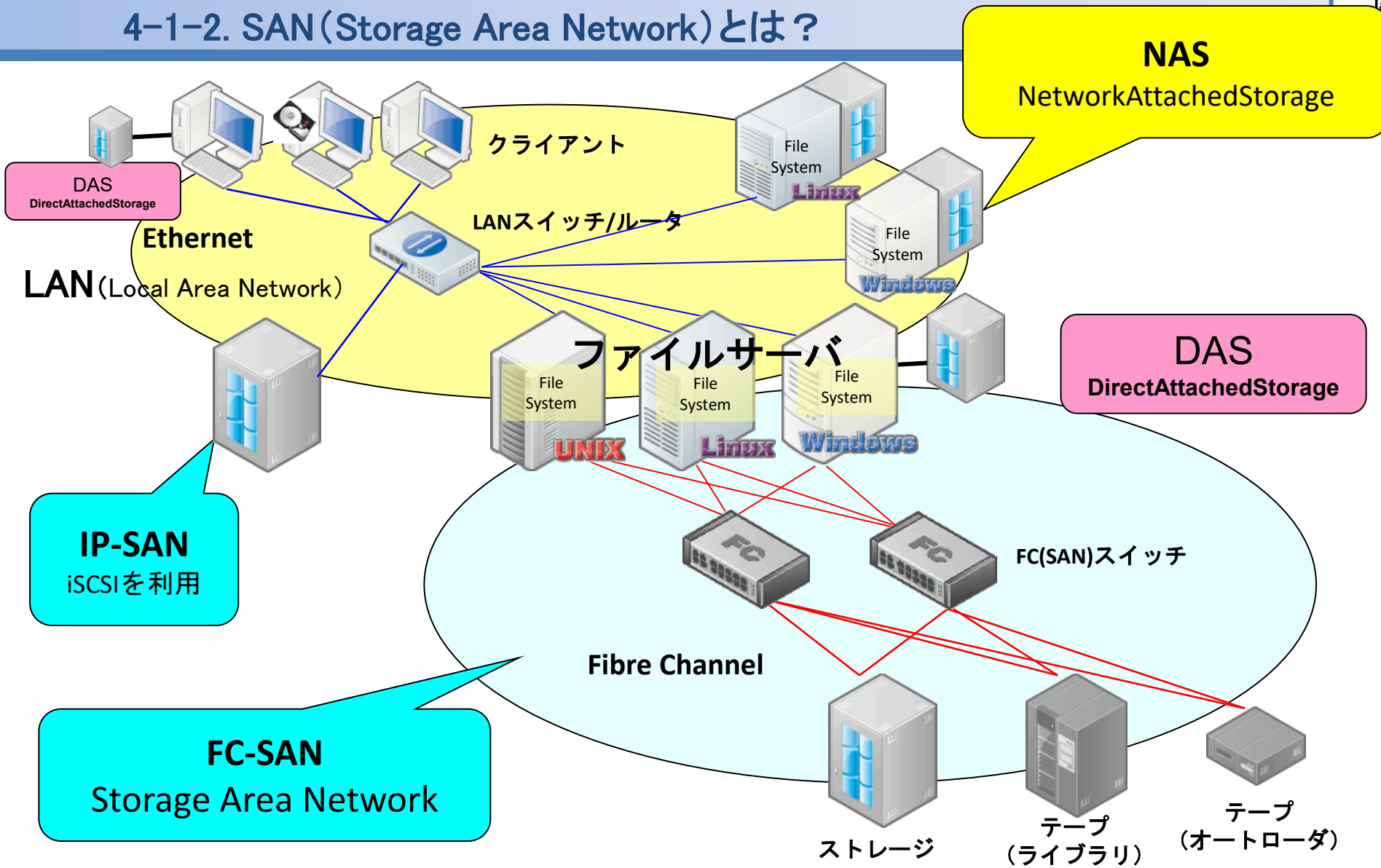
- ・ホスト/サーバ/コンピュータ → イニシエータ : 接続先のデバイスを使う側
- ・ストレージ(HDD/SSD/Tape)装置 → ターゲット : 接続してきたデバイスに使われる側

と呼びます。

従いまして本章では、
“プロトコル”は“物理層(データを運ぶレイア)や、通信、上位層でイニシエータ”と“ターゲット”がデータ通信(会話)を行う手段(言語)として表現し使用します。

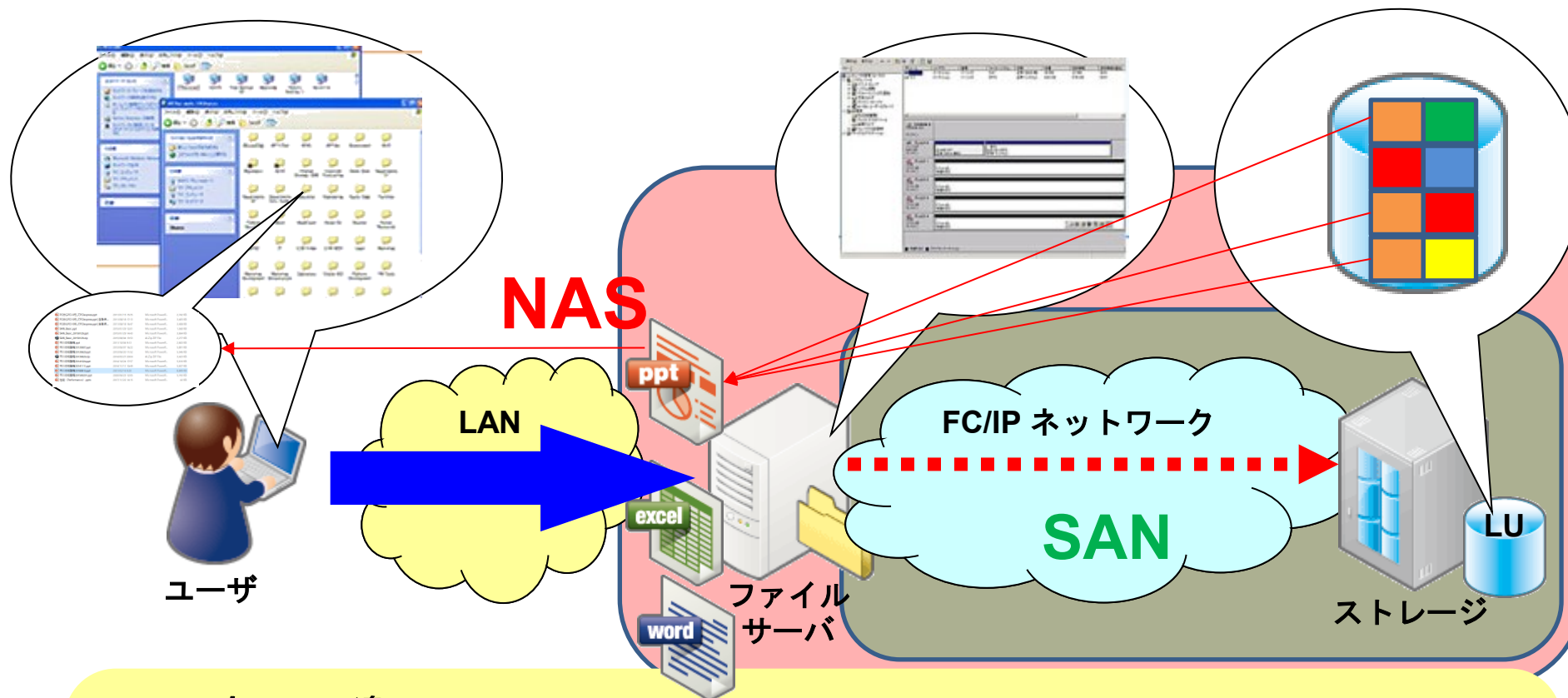
4-1. ストレージネットワークの基礎

4-1-2. SAN(Storage Area Network)とは？



4-1. NASとSANって何が違うの？

ファイルアクセスとブロックアクセスのイメージ



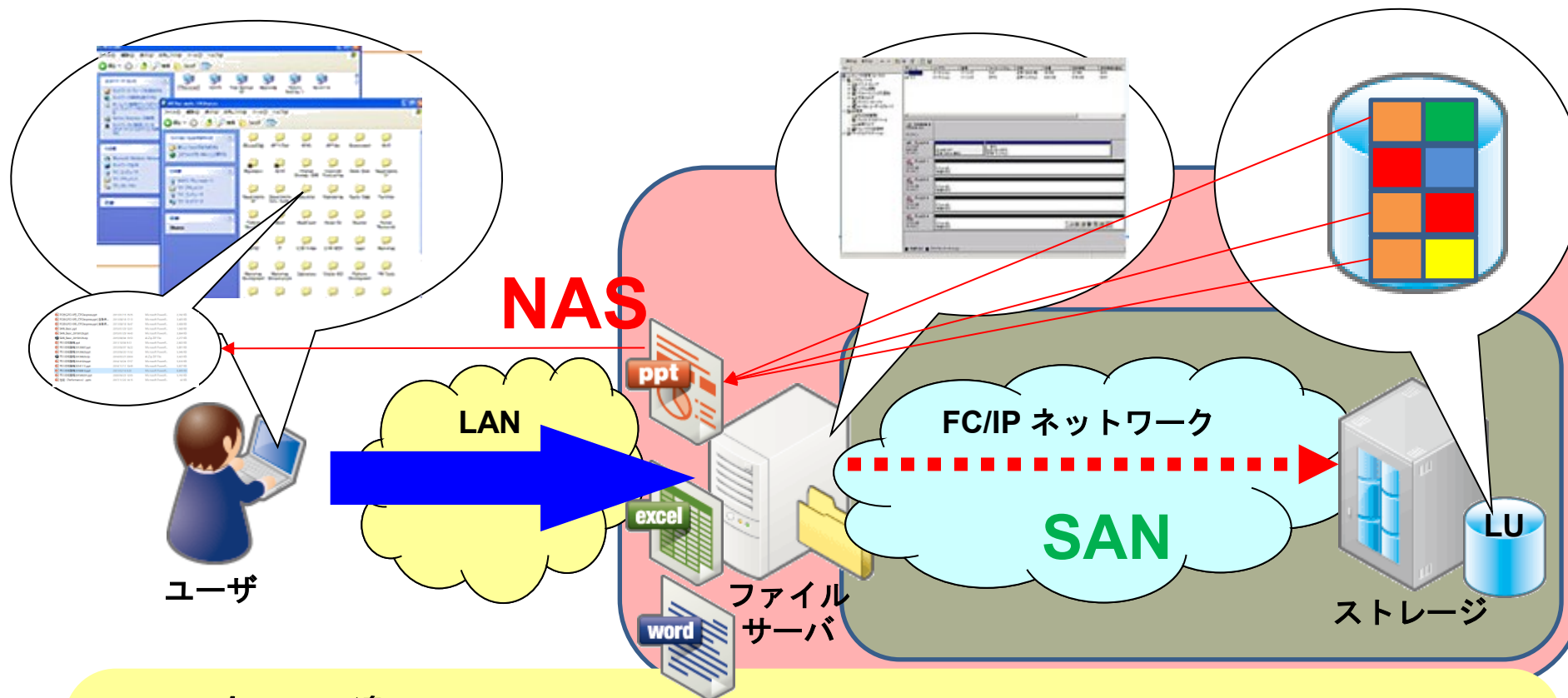
■ NASとSANの違い

- ◆ NAS (Network Attached Storage)
 - IPネットワークに接続するファイルサーバ装置もしくはストレージにファイルサーバを機能を付帯
ファイルアクセス、**ネットワークドライブ**として使う
- ◆ SAN (Storage Area Network)
 - ネットワークのように繋がるストレージのシステム
ブロックアクセス、**ローカルディスク**のように使う

じゃあ全部NASでいいじゃん！

4-1. NASとSANって何が違うの？

ファイルアクセスとブロックアクセスのイメージ



■ NASとSANの違い

◆ NAS (Network Attached Storage)

- IPネットワークに接続するファイルサーバ装置もしくはストレージにファイルサーバ機能を付帯
ファイルアクセス、**ネットワークドライブ**として使う

◆ SAN (Storage Area Network)

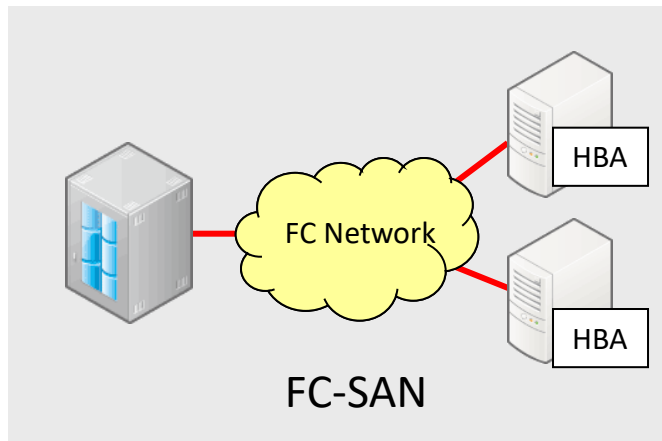
ミッションクリティカルなデータベースサーバなどは、高い処理能力をもつ複数の専用サーバと大容量/高性能のストレージが必要となる

4-1. ストレージネットワークの基礎

4-1-3. SAN(Storage Area Network)とは？

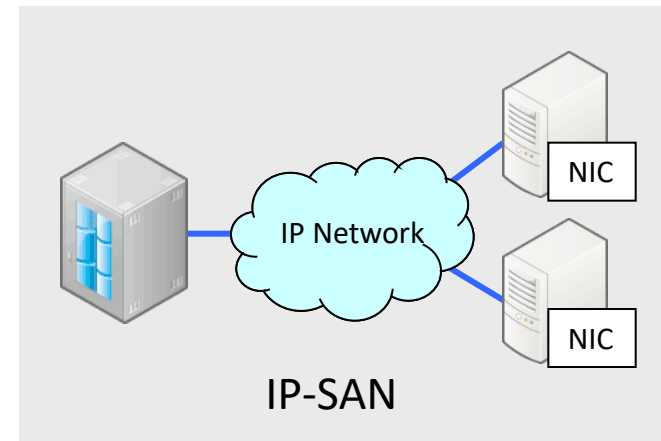
•SANの種類

現在の SAN (Storage Area Network) は、FC-SAN と IP-SAN に大別される



FC-SAN とは

- Fibre Channelプロトコルを使用
- 専用のFCネットワークを構築
 - 主に光ファイバケーブルを使用
- 専用の機器・ケーブルが必要



IP-SANとは

- iSCSIプロトコルを使用
- 既存のIPネットワーク上で動作する
 - ◆ 10Gbps以下は主にUTP (DAC)ケーブルを使用
- 既存のNICやLAN-Switchを流用可能
 - ◆ ソフトウェア iSCSI イニシエータが必要
- 専用NIC (iSCSI-HBA)も使用可能

4-1. ストレージネットワークの基礎

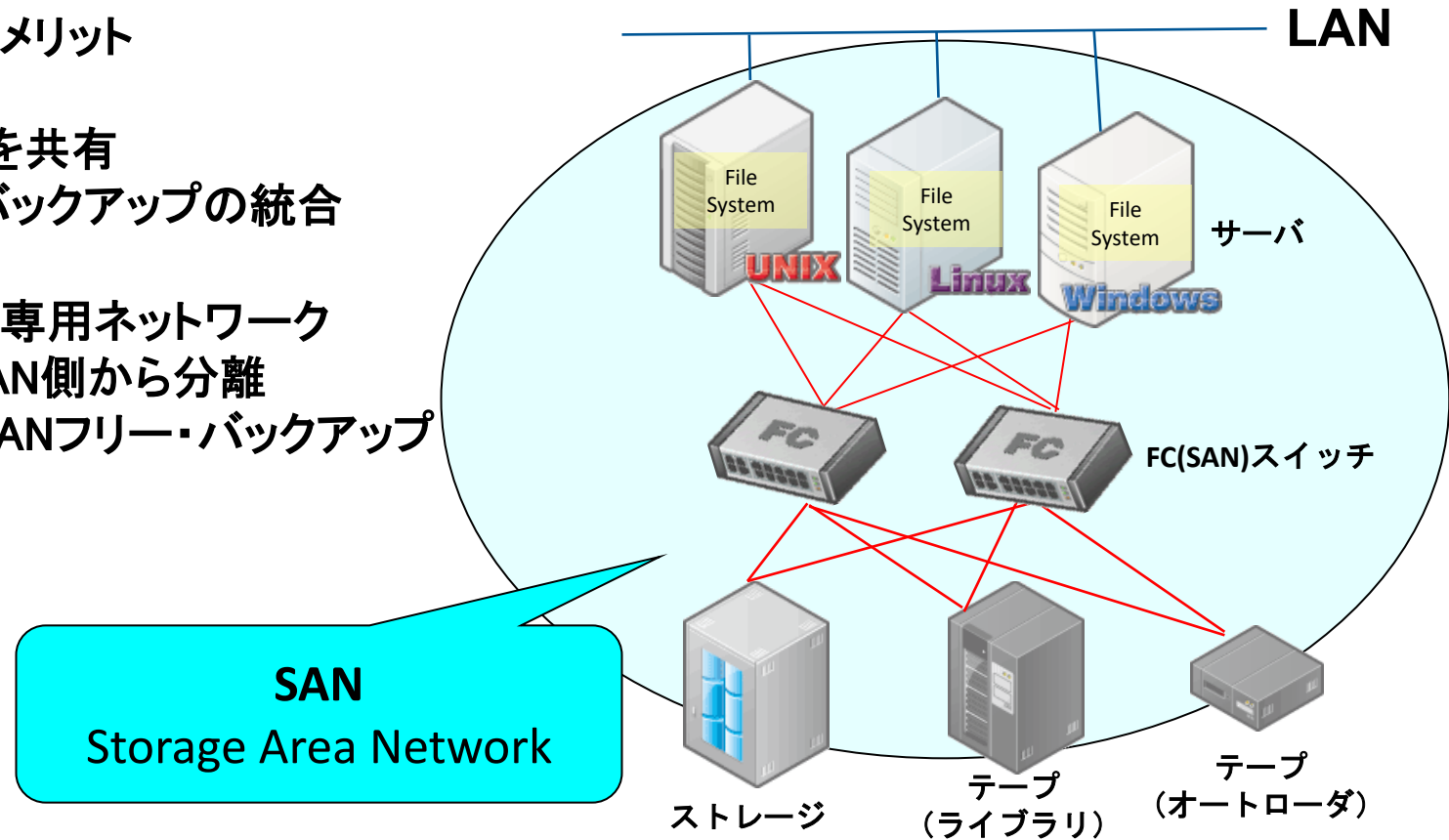
4-1-4. SAN(Storage Area Network)とテープ装置

SAN環境でのバックアップ

現在のテープ装置は、接続性(対応インターフェース)、処理速度、容量とSAN環境に接続する面で懸念することは、ほとんど無くなった。

SAN環境でのメリット

- テープ装置を共有
→バックアップの統合
- バックアップ専用ネットワーク
通信用のLAN側から分離
→LANフリー・バックアップ



SAS/SATA Tape ライブラリ/ドライブの FC-SAN/IP-SANへ接続

ネイティブにFibreChannelやiSCSI(Ethernet)のインターフェースを持たないSAS 12Gのテープドライブも
接続可能！ メーカーサイト <https://www.atto.com/products/xstreamcore-intelligent-bridges/>

■ ATTO XstreamCORE FC7600 32G-FC ↔ 12G SAS



■ ATTO XstreamCORE ET8200T/ET8100T 40G Ethernet(iSCSI) ↔ 12G SAS



国内代理店：<https://www.servants.co.jp/product/sub-gateway/?maker=atto>

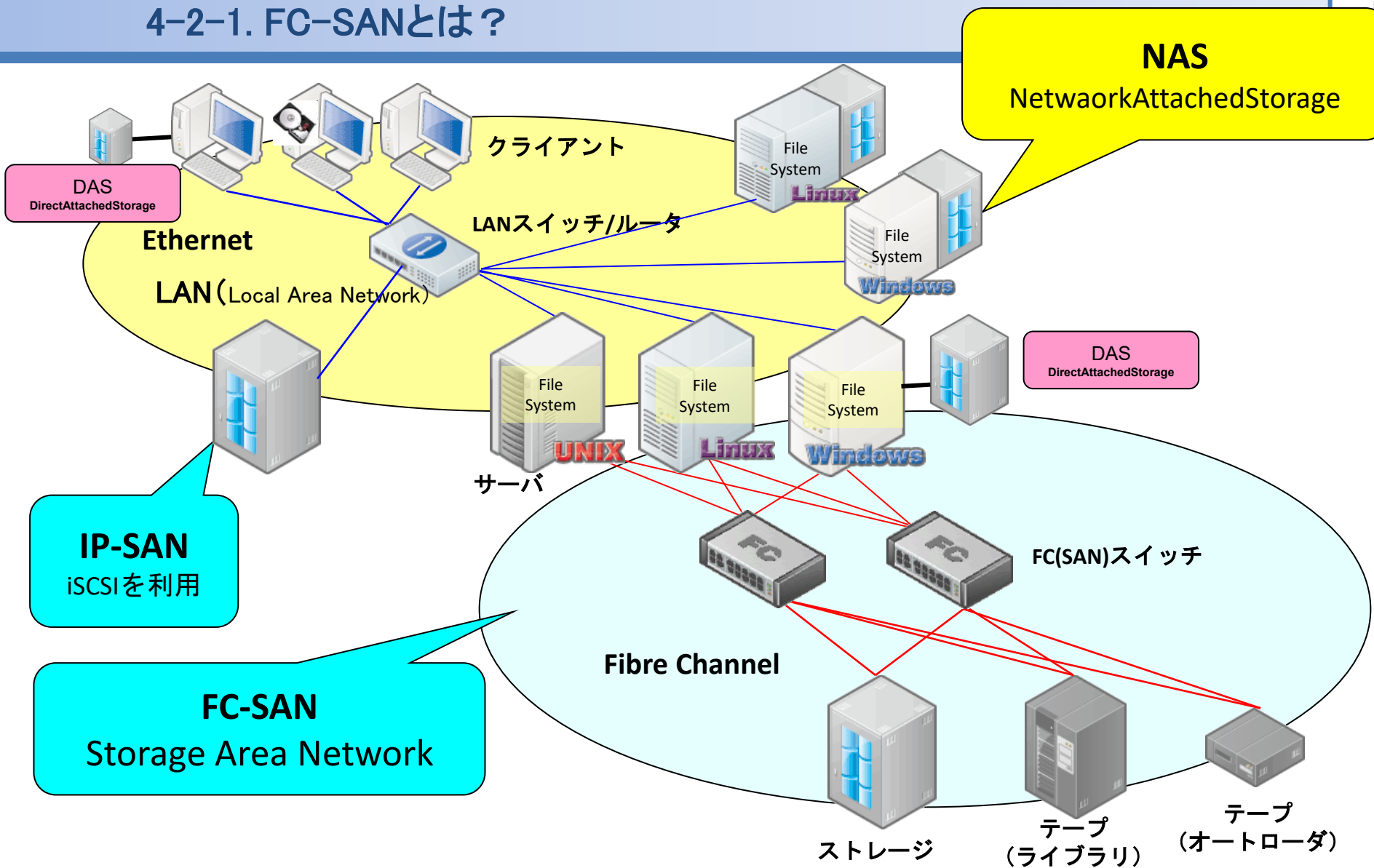
技術コラム：私が検証したインプレッションを掲載してます

8200T：<https://www.servants.co.jp/technology/column/technology/tec-atto/13196/>

8100T：<https://www.servants.co.jp/technology/column/technology/tec-atto/13162/>

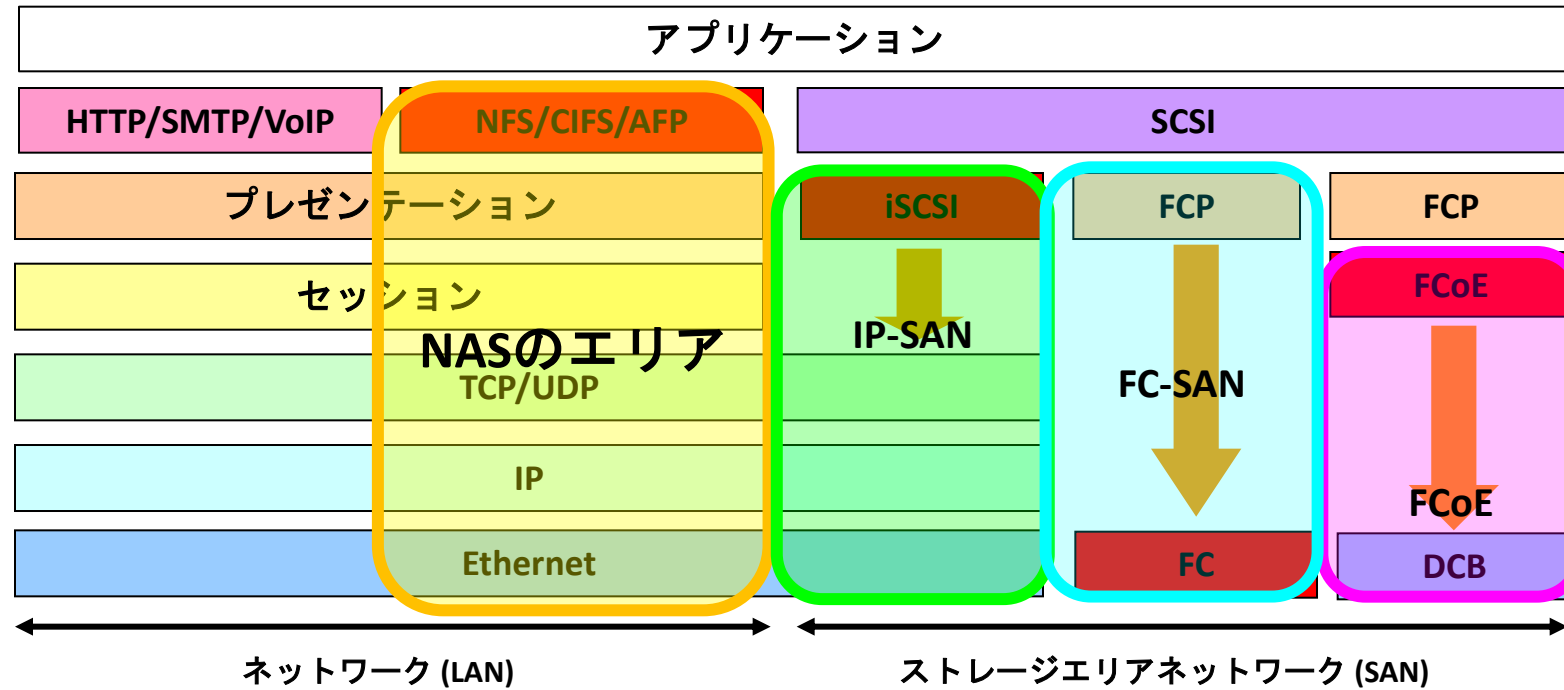
4-2. FC-SANとその他のストレージネットワーク

4-2-1. FC-SANとは？



4-2. FC-SANとその他のストレージネットワーク

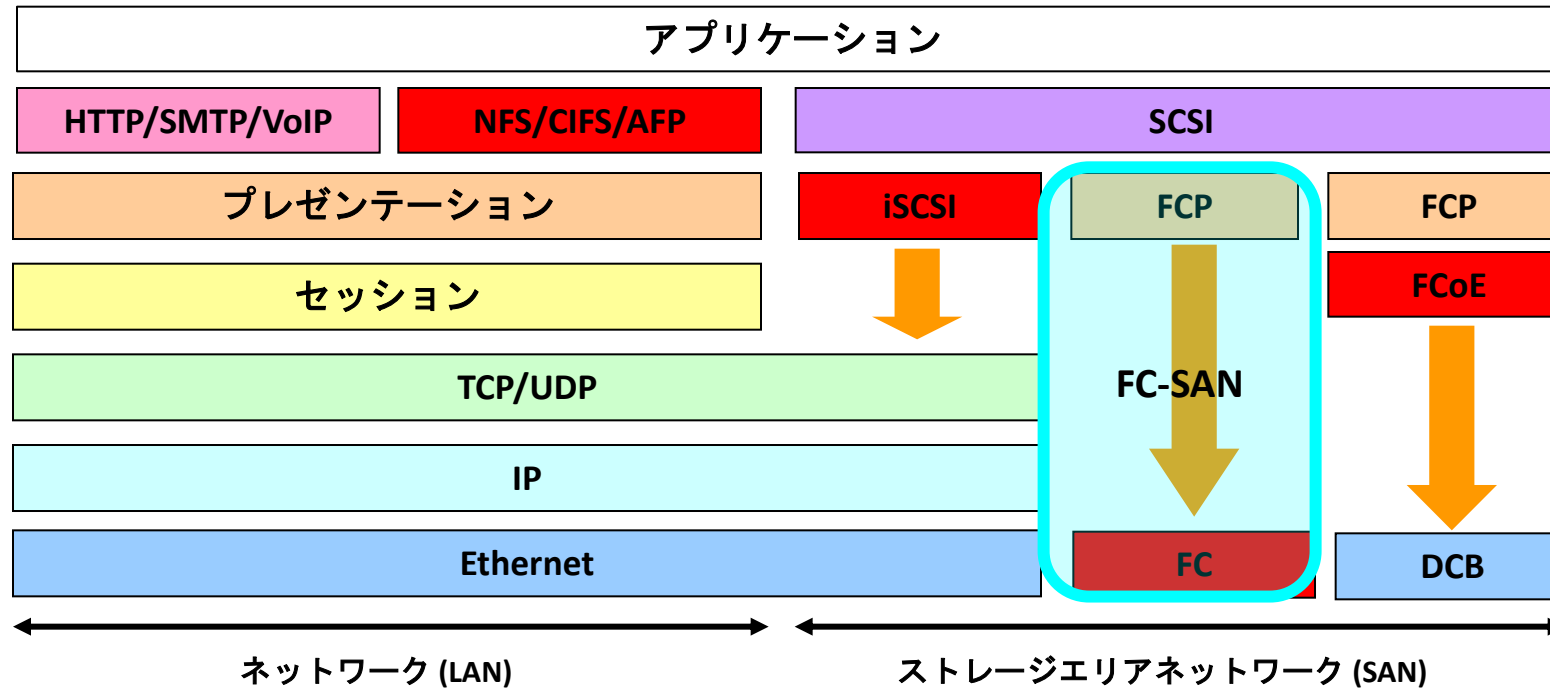
4-2-2. FC-SANを構成するChannelプロトコルとは？



- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。**TCP/IPは利用しない。**

4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？



- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。**TCP/IPは利用しない。**

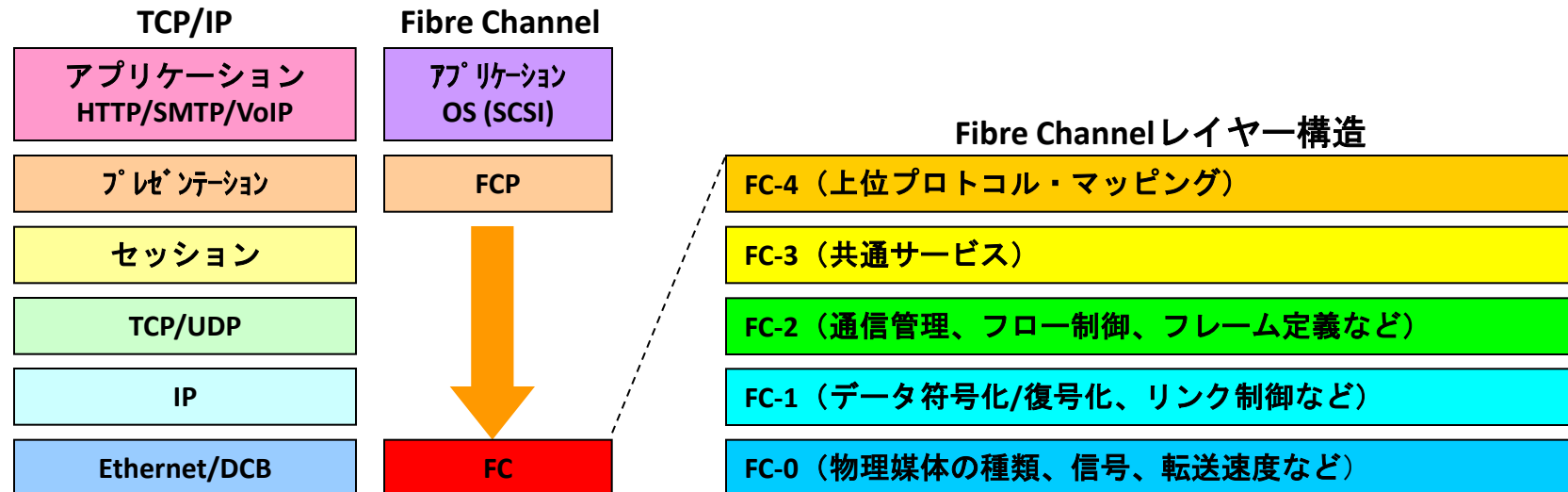
4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？

Fibre Channelは1998年にANSI T11にて標準化がスタート

■ 特徴

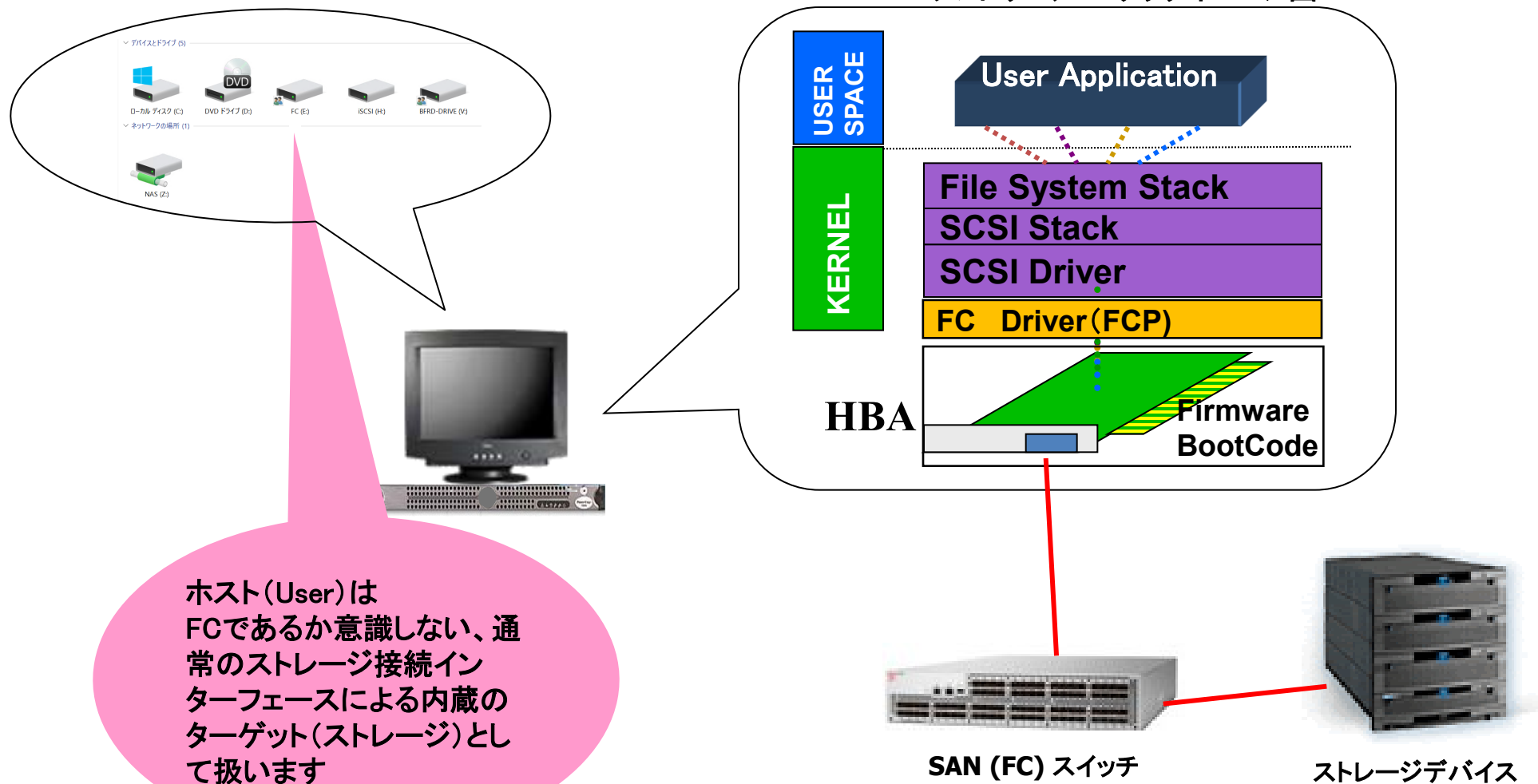
- ◆ Fibre Channelプロトコルを用いたストレージ専用ネットワーク
- ◆ 物理データ伝送速度（Max: Gen7:64Gbpsの製品 Gen8:128Gbpsが策定されている）
- ◆ 全二重シリアル伝送方式
- ◆ プロトコル依存性が無い
- ◆ 最長10kmまで接続可能（長波長、シングルモードケーブル）
- ◆ 基幹業務システムなどでデータベース用に利用されることが多い



4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？

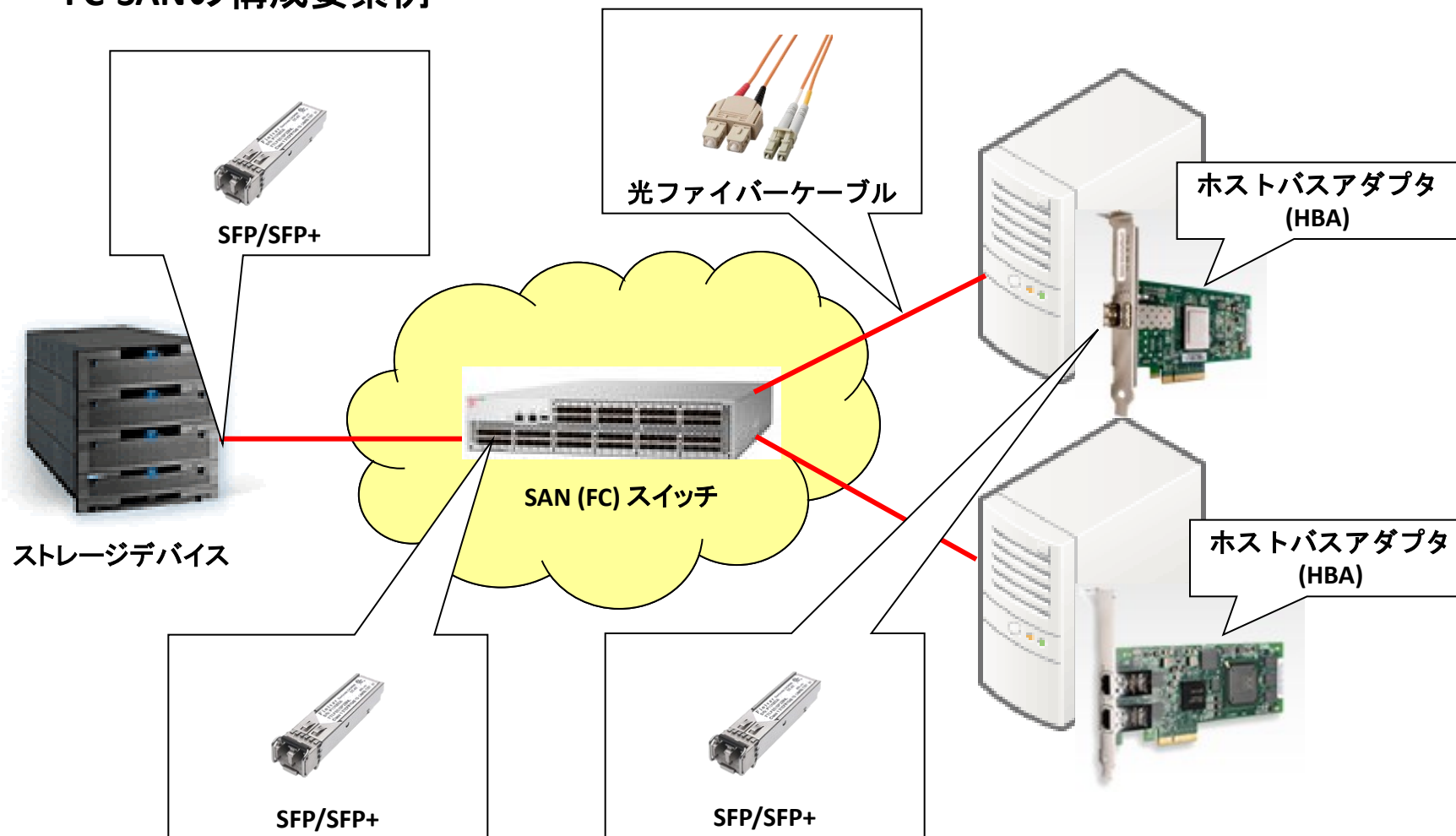
ソフトウェア・スタックイメージ図



4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するFibreChannelプロトコルとは？

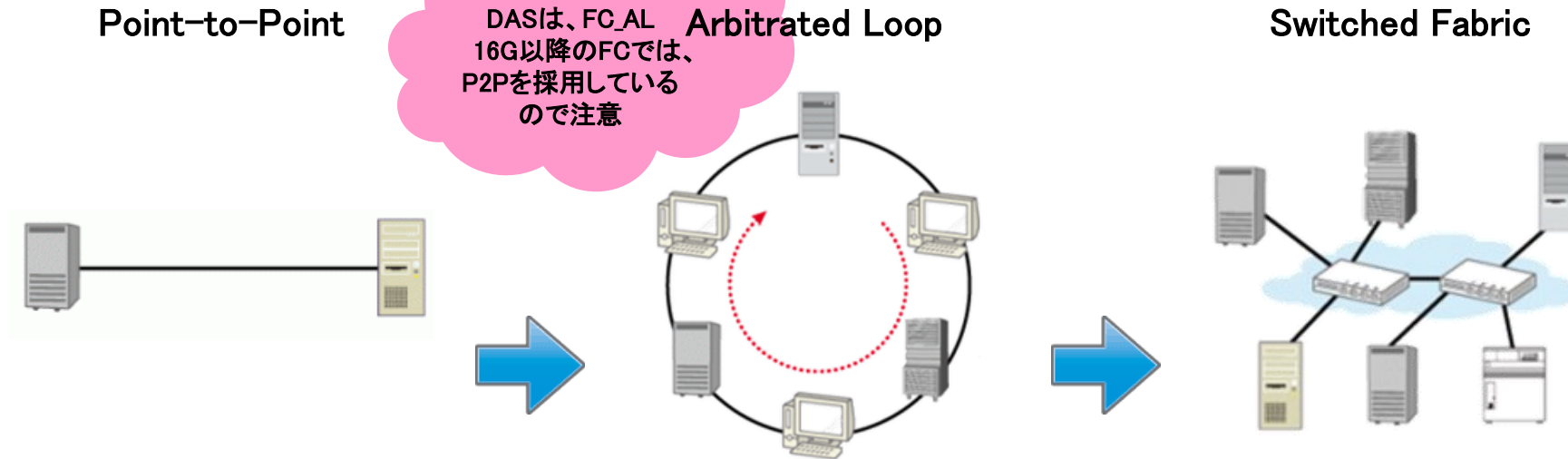
•FC-SANの構成要素例



4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？

FC-SANのトポロジー



SANではない(DAS)

- ・2デバイスのみ
(直接接続)
- ・1:1の接続
- ・帯域は独占使用
- ・拡張性はない

初期のSANの形

- ・126デバイスまで
(FCハブ)
- ・帯域を共有
- ・アドレス調停が必要

現在のSANの形

- ・1600万デバイスまで
(FCスイッチ)
- ・帯域はポートごと
- ・拡張性がある

4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？

FC-SANのトポロジー

Point-to-Point

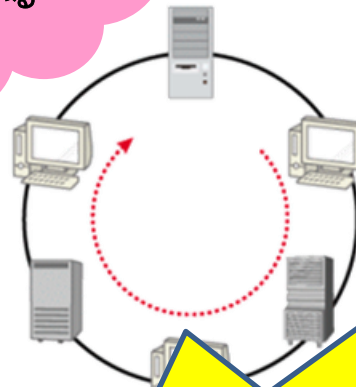


SANではない(DAS)

- ・2デバイスのみ
(直接接続)
- ・1:1の接続
- ・帯域は独占使用
- ・拡張性はない

2-8GのFCでは、
DASは、FC_AL
16G以降のFCでは、
P2Pを採用している
ので注意

Arbitrated Loop



Switched Fabric



**16GのFC製品は
Gen5とGen6対応の
製品があるので注意
が必要！！**

現在のSANの形

- ・1600万デバイスまで
(FCスイッチ)
- ・帯域はポートごと
- ・拡張性がある

4-2 FC-SAN

FC-SANを構成するChannelプロトコルとは？

最大の特徴としてFCには、ポートと言う概念があります
それぞれのポート同士がログインという動作をして情報を共有します。

接続ポート種別

- スイッチのポート
 - F_Port : ノードとのPoint-to-Point接続におけるFCスイッチのポート
 - FL_Port: ノードとのループ接続におけるFCスイッチのポート
 - E_Port: FCスイッチ間接続をするFCスイッチのポート
 - G_Port: E_Port, F_Portになり得るポート
 - GL_Port: E_Port, FL_Port, F_Portになり得るポート
- ノード(接続機器: HBAやストレージ)のポート
 - N_Port: Point-to-Point(Fabric)接続をサポートしたノードのポート
 - NL_Port: ループトポロジをサポートしたノードのポート

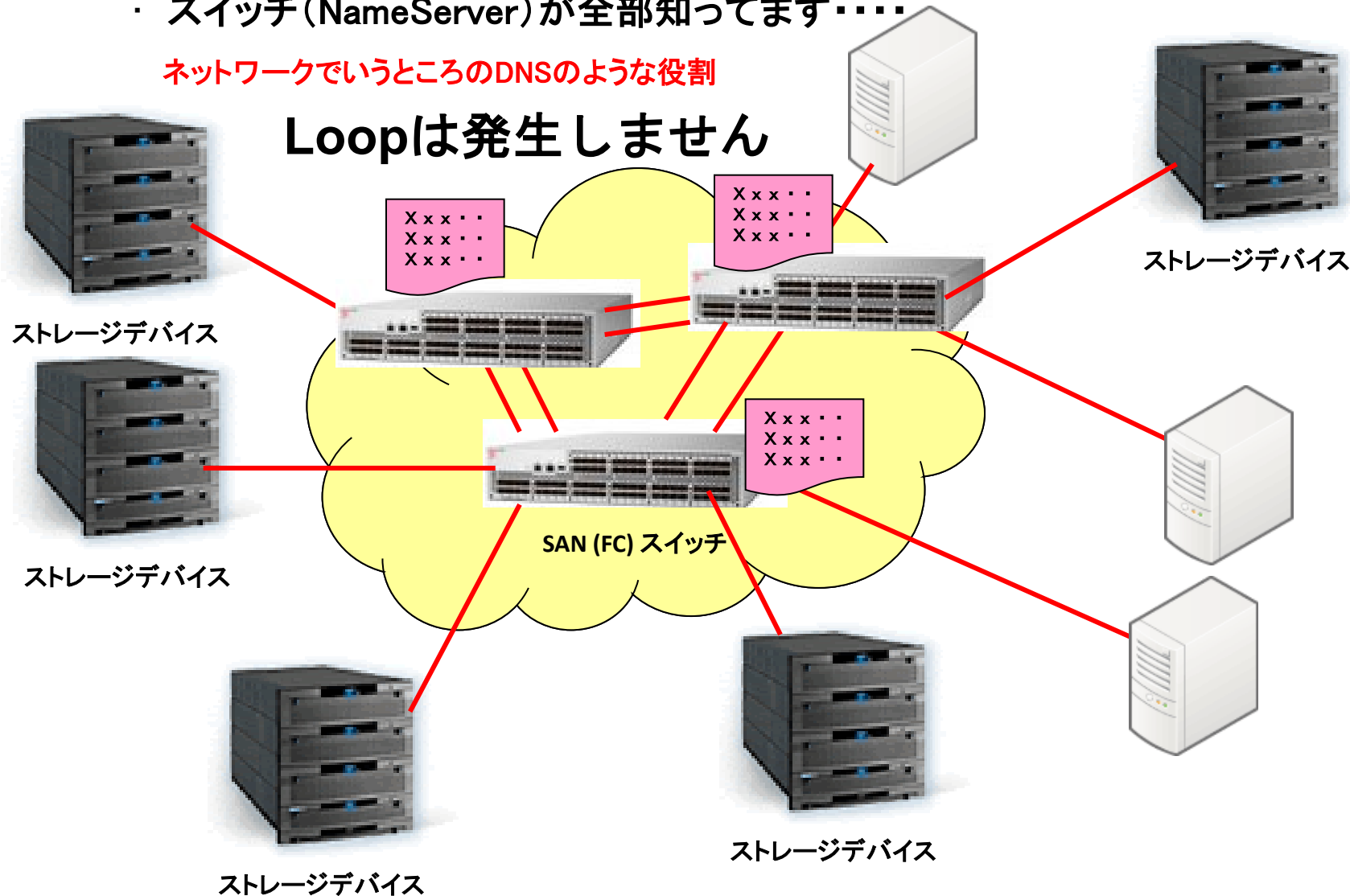
4-2. FC-SANとその他のストレージネットワーク

4-2-2. FC-SANを構成するChannelプロトコルとは？

- ・ スイッチ (NameServer) が全部知ってます……

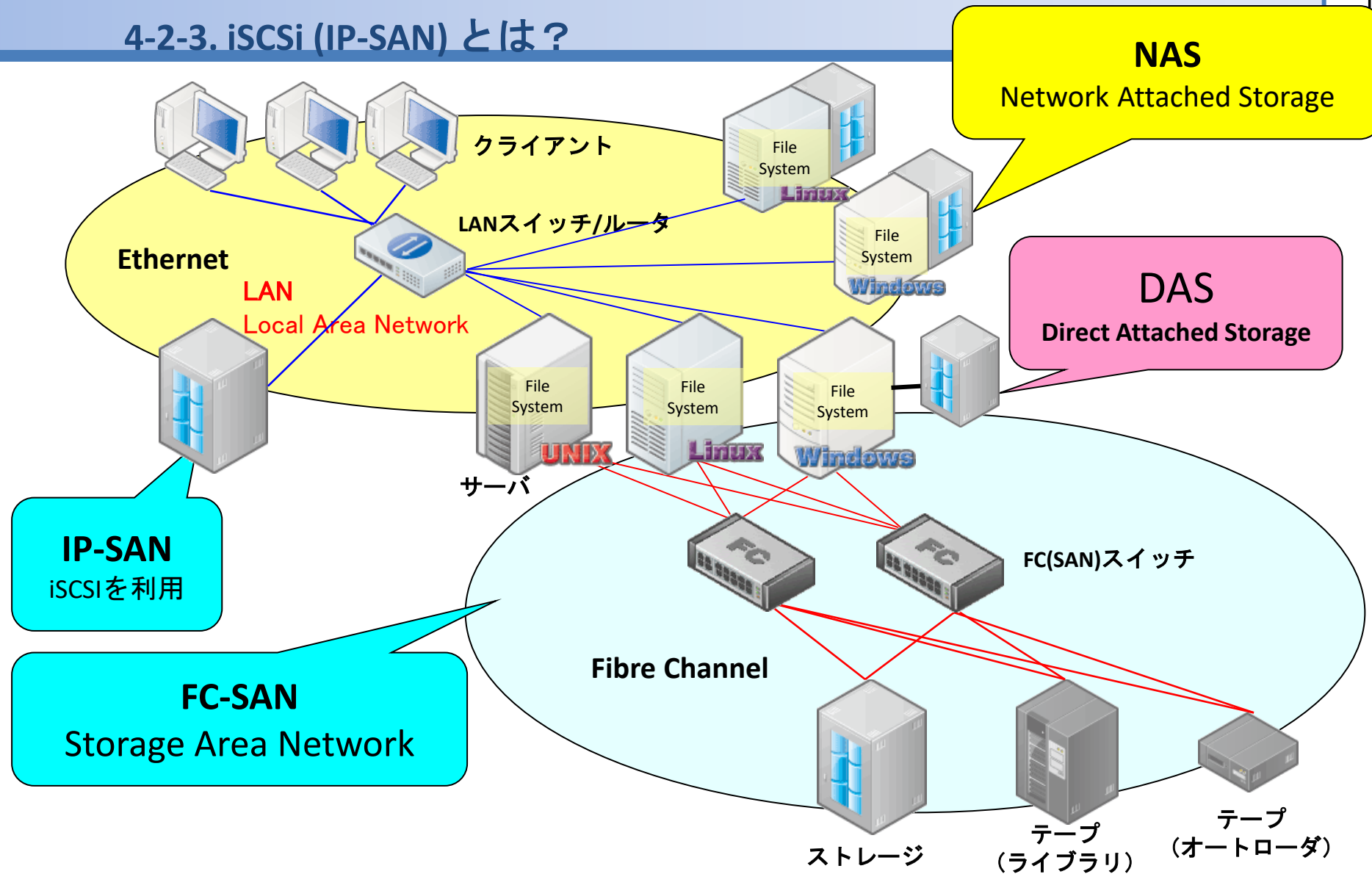
ネットワークでいうところのDNSのような役割

Loopは発生しません



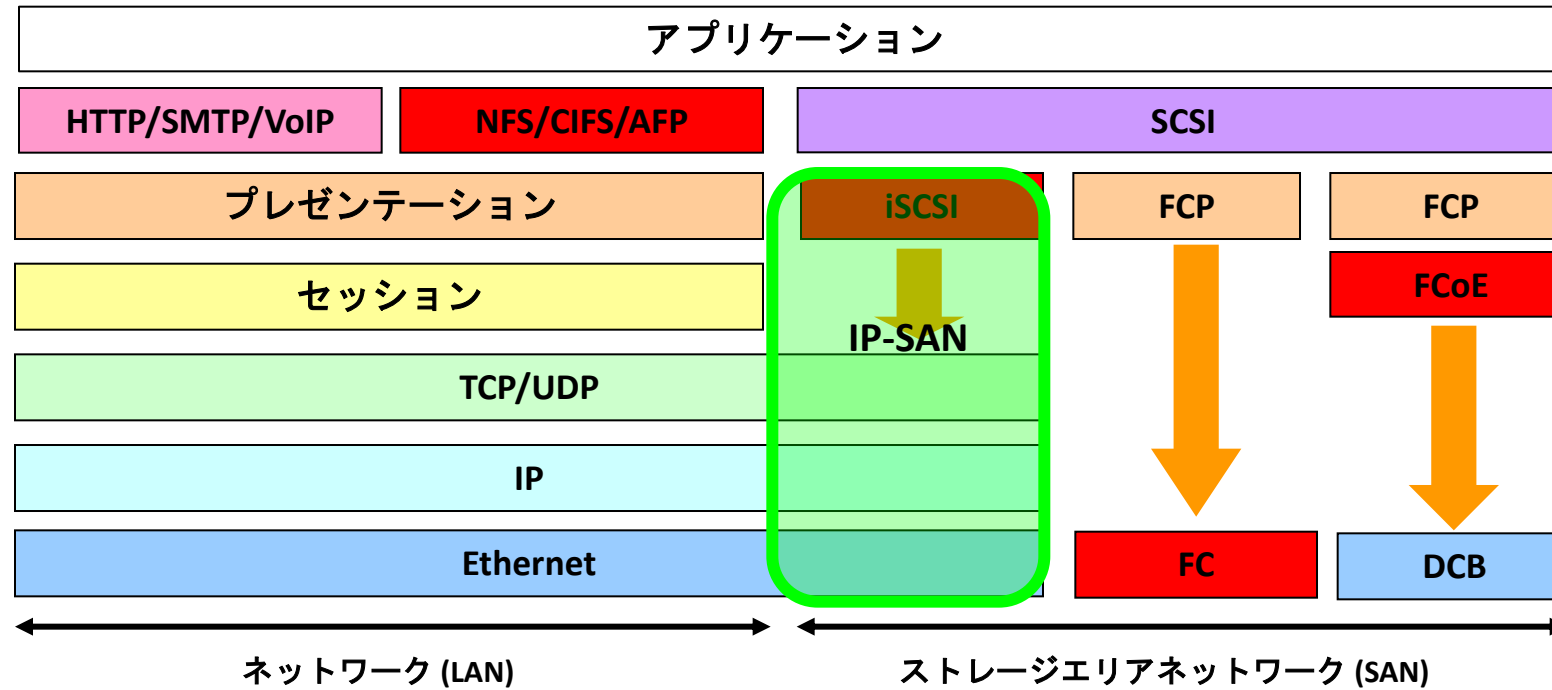
4-2. FC-SANとその他のストレージネットワーク

4-2-3. iSCSI (IP-SAN) とは？



4-2. FC-SANとその他のストレージネットワーク

4-2-3. iSCSI (IP-SAN) とは？



- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。TCP/IPは利用しない。

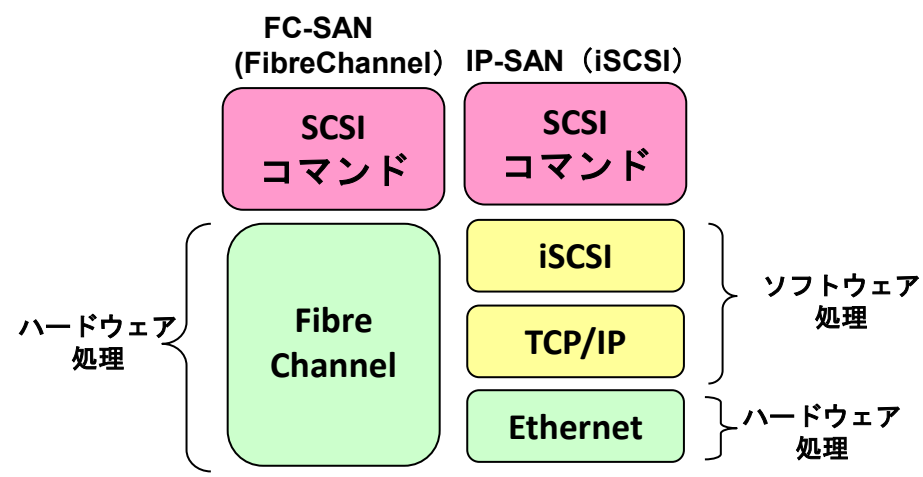
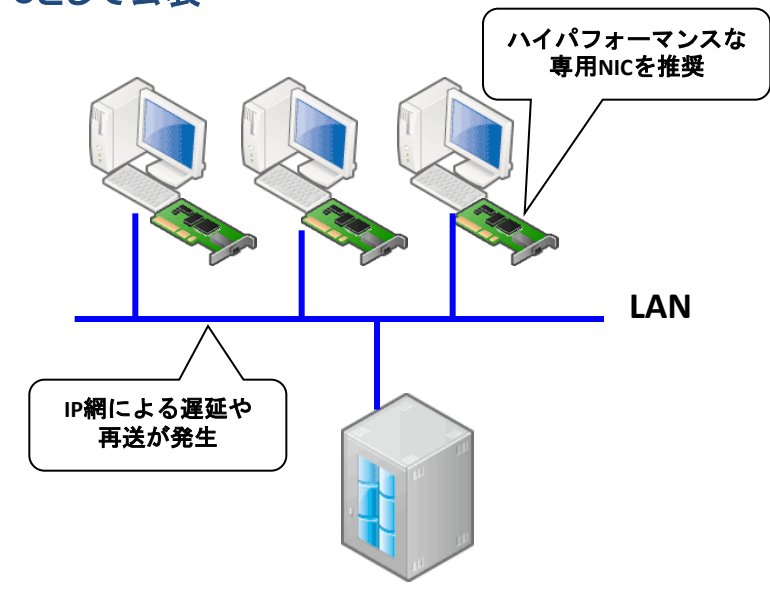
4-2. FC-SANとその他のストレージネットワーク

4-2-3. iSCSI (IP-SAN) とは？

2003年2月にIETFによってRFCとして公表

■ FC-SANとIP-SANの違い

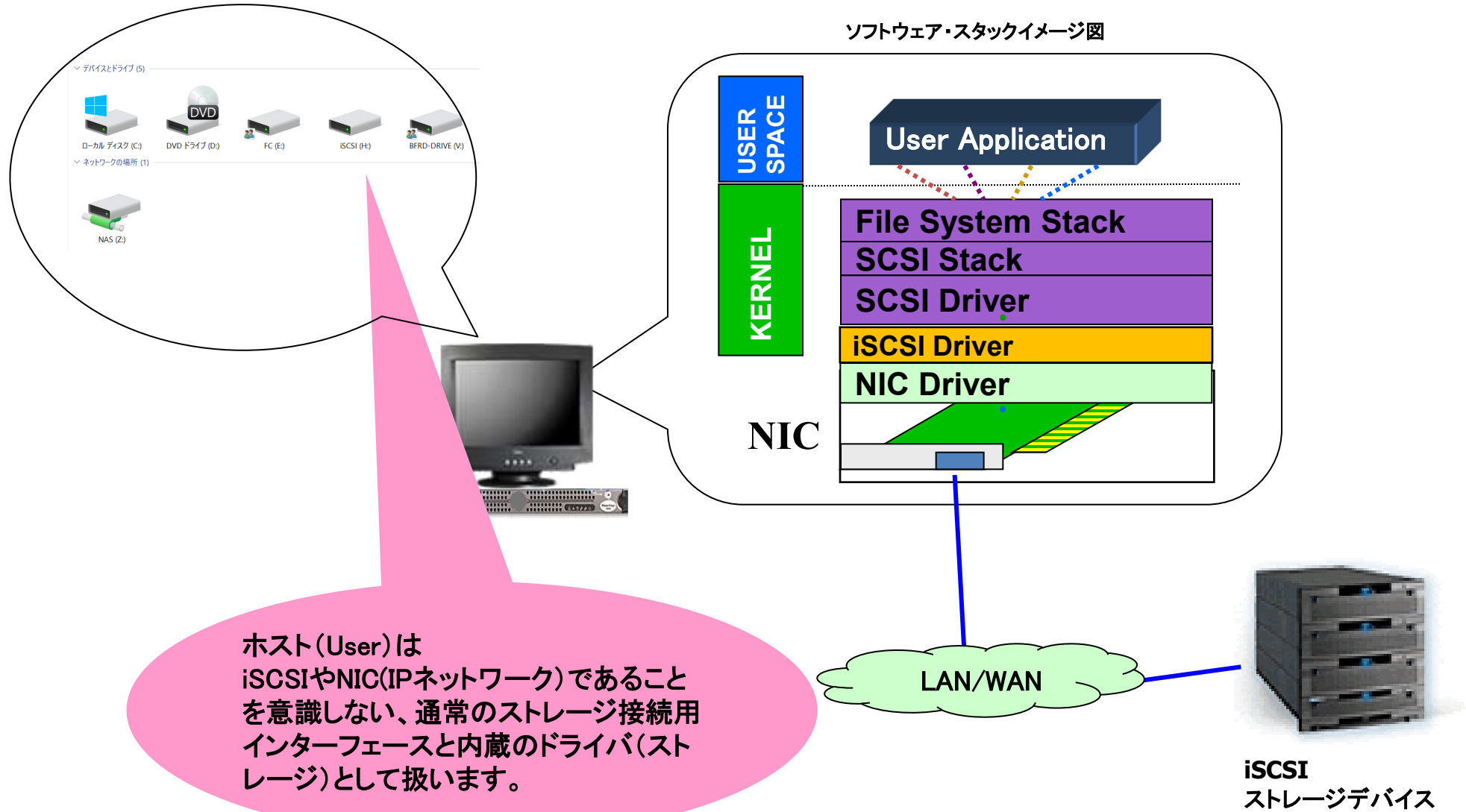
- ◆ IP-SANはFC-SANと比較すると、安価に構築できる。
- ◆ 安定した運用を行うには、下記の点に考慮する必要がある。
 - ▶ 安定したスループットを出せるIP網の整備
 - ▶ Jumbo Frame、全二重通信、Gigabit Ethernet以上のパフォーマンス
 - ▶ TOE (TCP Offload Engine) 搭載のiSCSI HBAを使用してCPUの負荷を軽減して全体のパフォーマンスを上げる



	FC-SAN	IP-SAN
プロトコル	Fibre Channel	iSCSI
ケーブル (物理層)	主に光ファイバケーブル	主にUTPケーブル
安定性	◎	○
価格	△	○

4-2. FC-SANとその他のストレージネットワーク

4-2-3. iSCSI (IP-SAN) とは？



4-2. FC-SANとその他のストレージネットワーク

4-2-4 FC-SANとiSCSI (IP-SAN) の使い分け

■ FC-SANとiSCSI (IP-SAN) の使い分け

- ◆ FC, iSCSIなどのプロトコルは、SCSI(ストレージプロトコル)で通信するための手段である。SCSIはバス通信技術である
- ◆ 適応範囲の選定基準
 - ▶ FC: 信頼性重視
 - ▶ 利点: 高い信頼性とハイパフォーマンス
 - ▶ 欠点: 高コスト
 - ▶ iSCSI: 低コスト
 - ▶ 利点: IPネットワーク網を活用できる
 - ▶ 欠点: パフォーマンス、安定性面では同じ帯域の場合、FCに軍配が上がる

ハイブリッド思考！

どちらかではなく、重要度、運用レベルを考慮し、必要な条件を見極めそれらに合わせたソリューションを実現させる

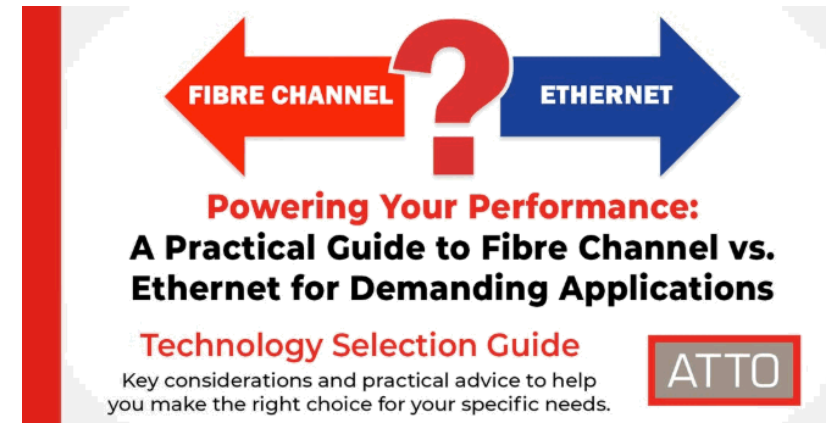
おまけ: FibreChannel VS Ethernet

ATTO Technology社が面白いドキュメントを公開してます。

<https://www.atto.com/FCvEthernetGuide>

ATTO Technollogy社とは？

<https://www.atto.com>



1990年台後半から2000年前半

QLogic / Emulex / McData / Brocade / JNI、そしてCiscoと、各社が製品を続々発表して全盛期を迎えていたFibreChannel業界に**MacOS**に対応した製品を強みとして参入してきたのがATTO Technology社です。

現在、FibreChannel製品(特にHBA)は、OEMが主流となり、チャネル(一般販売)には、ほとんど出回らなくなっている中、いまなおFC製品を活発に啓蒙、販売しています。FCだけでなくEthernet製品も多数取りそろえているメーカーになります。TB-FC, TB-Ether, TB-SASに変換する装置もラインナップしてます。

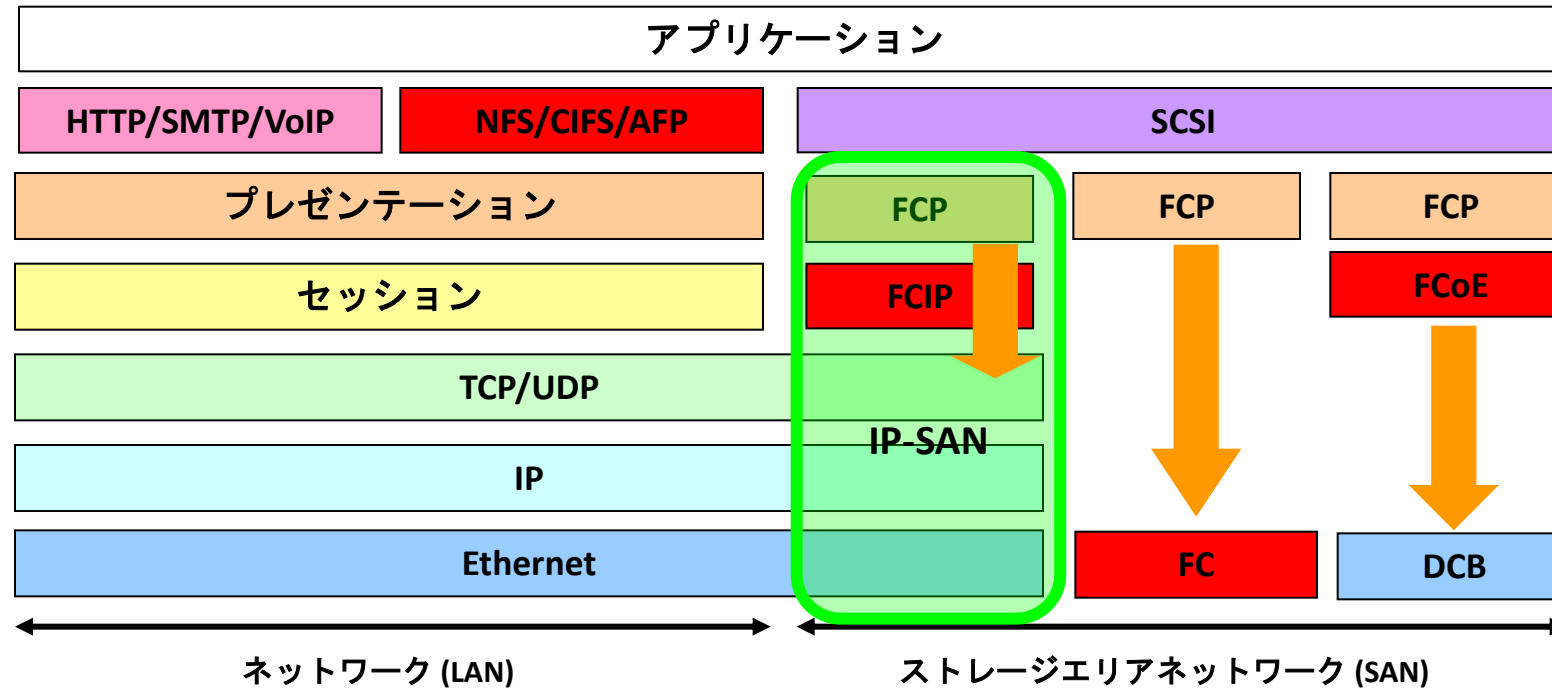


弊社技術コラムでも、FibreChannelの進歩とATTO Technologyというコラムを掲載してます。

<https://www.servants.co.jp/technology/column/technology/tec-atto/14316/>

4-2. FC-SANとその他のストレージネットワーク

4-2-6. FCIP (IP-SAN) とは？



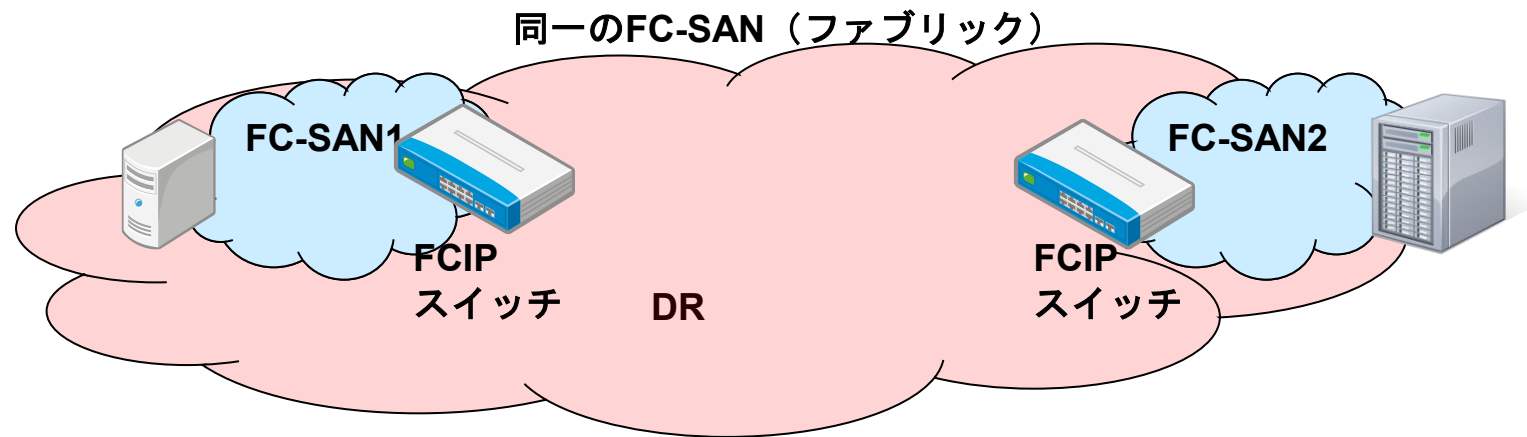
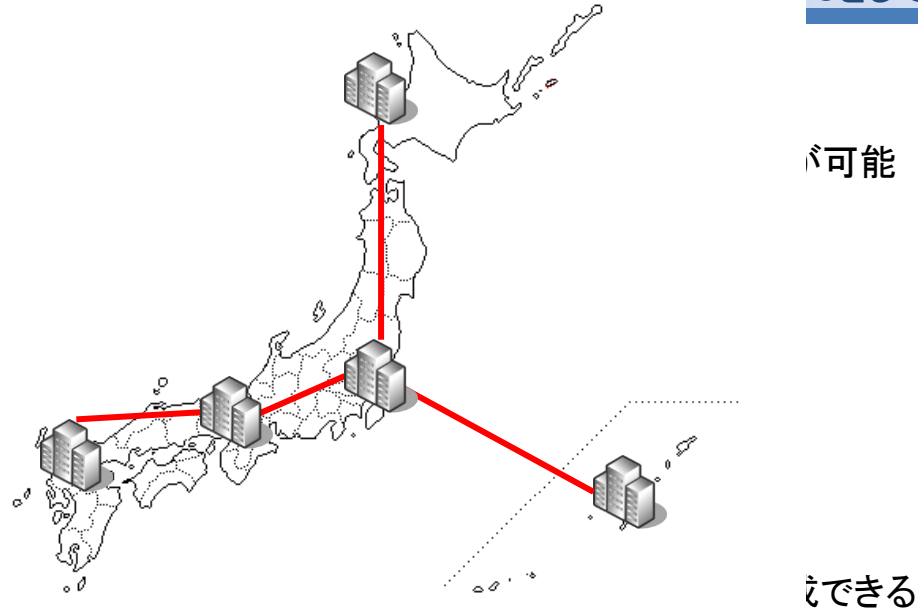
- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。TCP/IPは利用しない。

4-2. FC-SANとその他のストレージネットワーク

4-2-6. FCIP (IP-SAN) とは？ 2003年2月にIETFによってRFCとして公表

■ FCIPのメリット

- ◆ 高いFC-SANと遠隔地同志の
 - アプリケーション
 - FCの機能
- ◆ 高速化、対遅延
 - データの圧縮
 - 帯域のトラフィック制御
- ◆ コストパフォーマンス
 - WAN回線
 - 既存構成

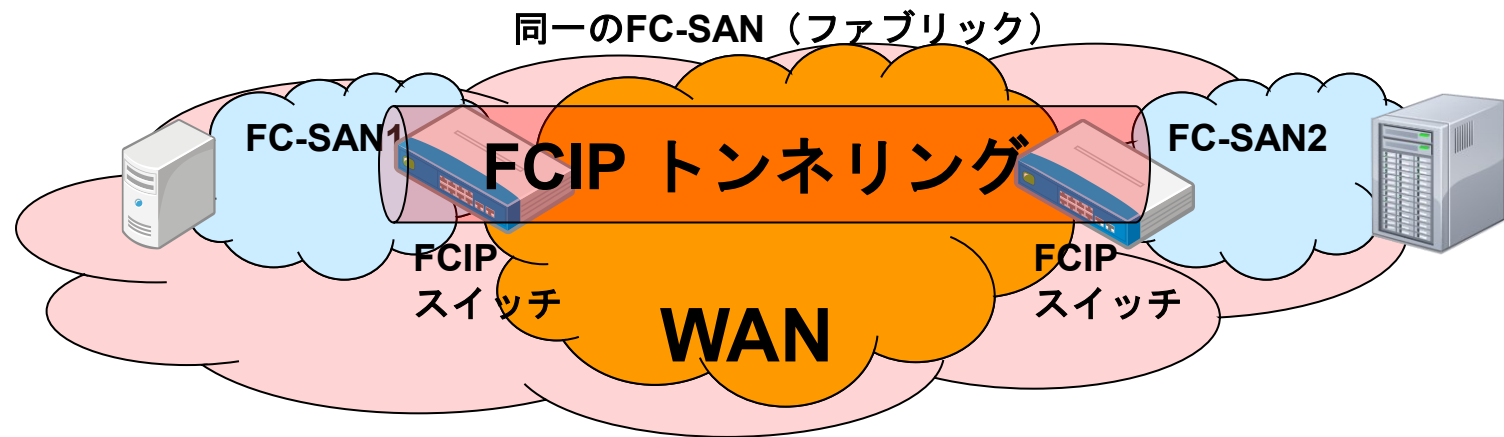
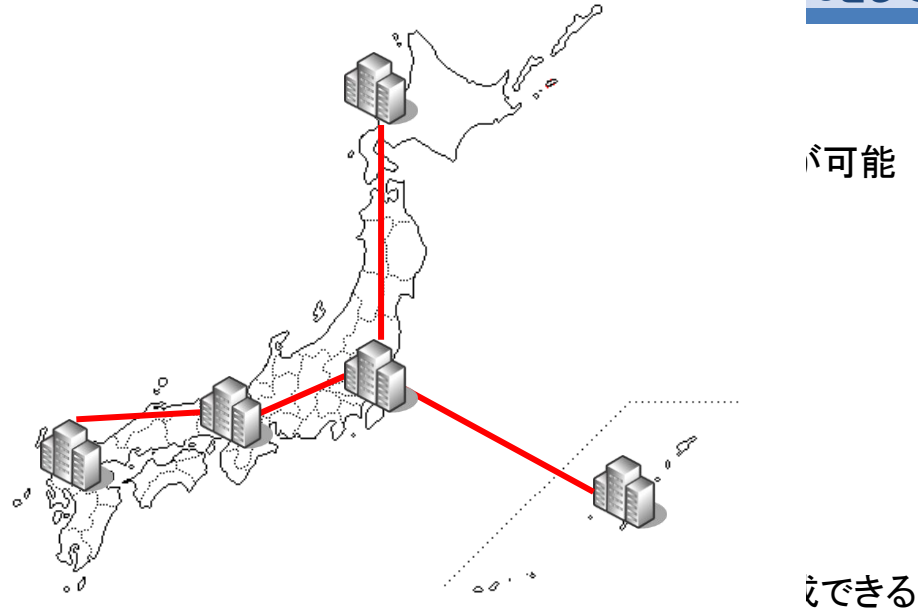


4-2. FC-SANとその他のストレージネットワーク

4-2-6. FCIP (IP-SAN) とは？ 2003年2月にIETFによってRFCとして公表

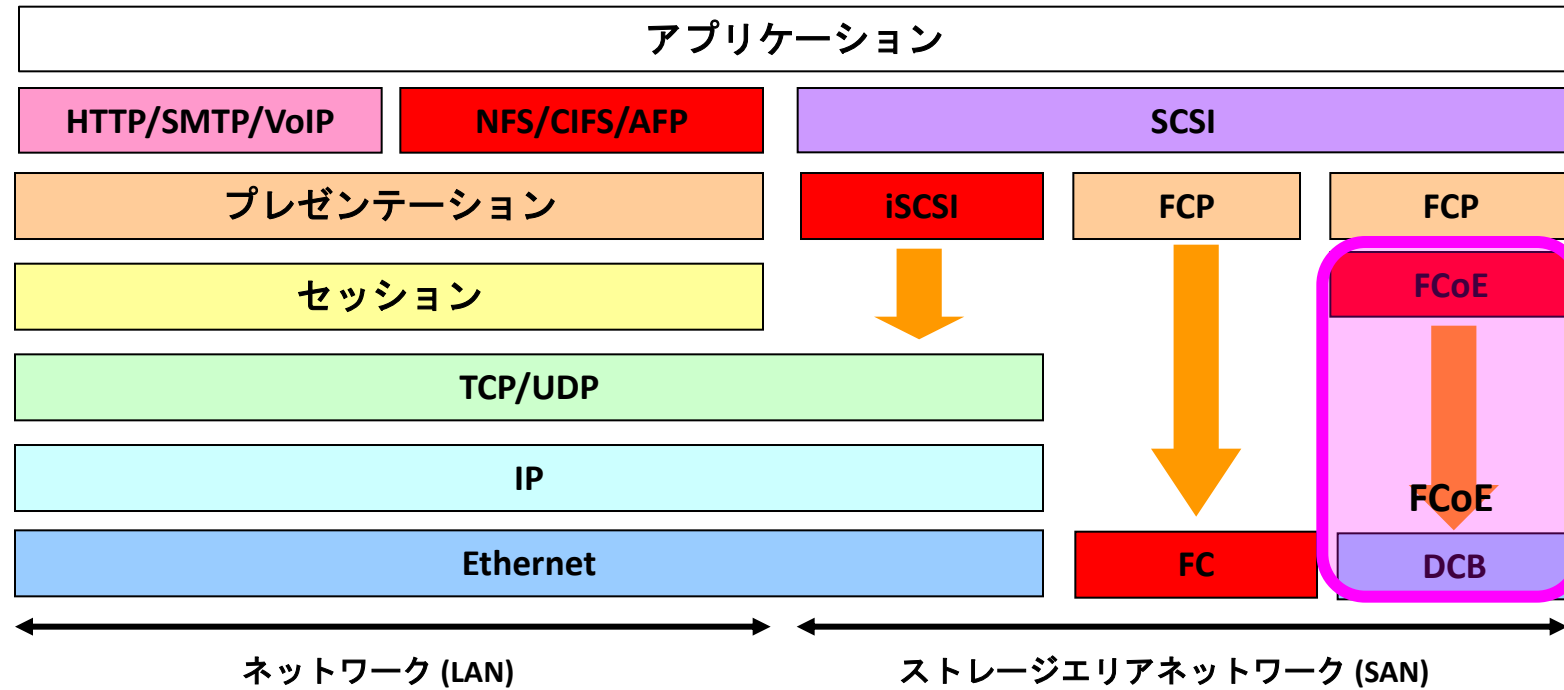
■ FCIPのメリット

- ◆ 高いFC-SANと遠隔地同志の
 - アプリケーション
 - FCの機能
- ◆ 高速化、対遅延
 - データの圧縮
 - 帯域の最適化
- ◆ コストパフォーマンス
 - WAN回線
 - 既存構成



おまけ：その他のストレージネットワーク

もう1つのIP-SAN FCoE/DCB とは？



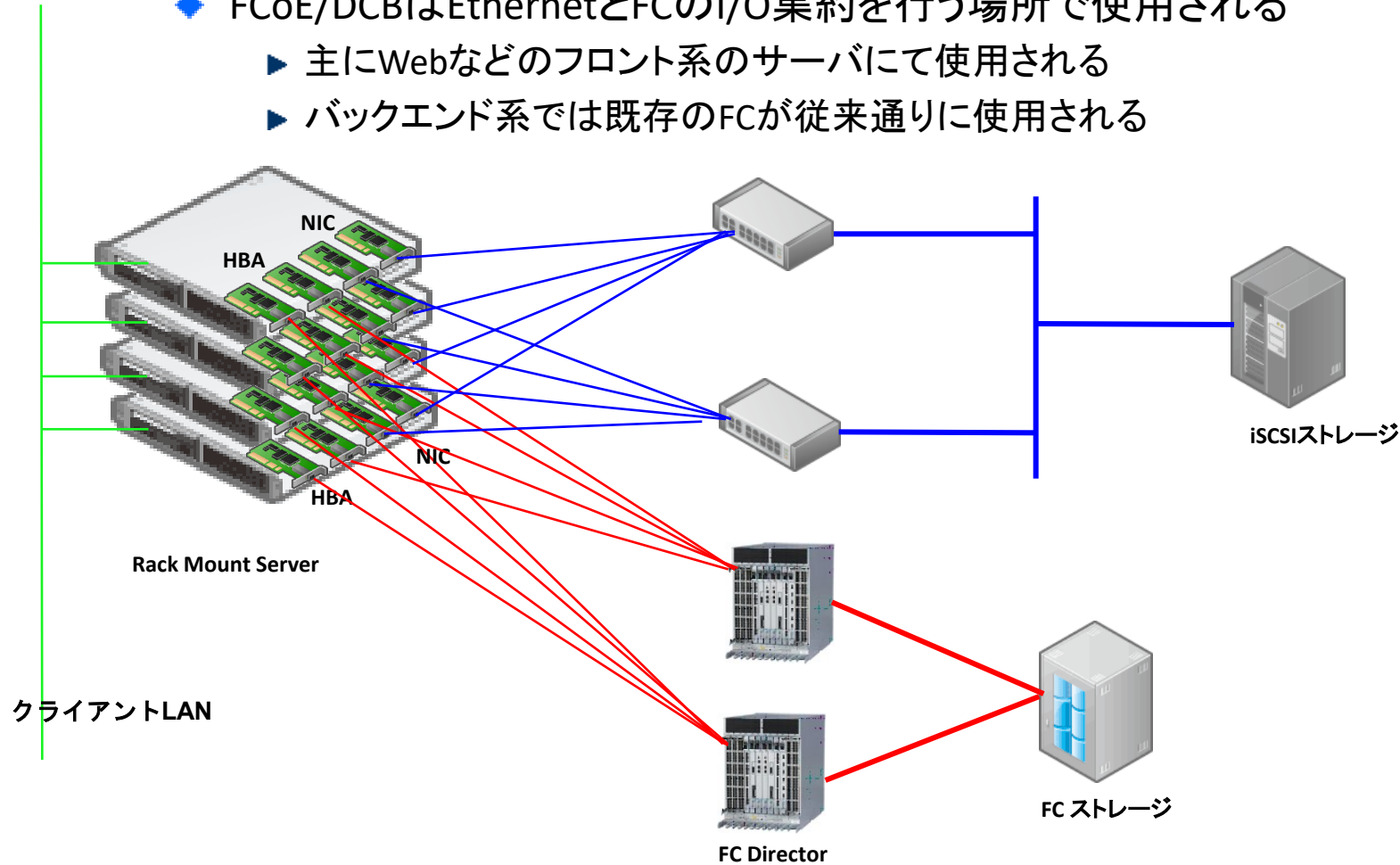
- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。TCP/IPは利用しない。

参考資料 その他のストレージネットワーク

4-2-7. FCoE/DCB とは？

■ FCoE/DCBとFC-SANの関係

- ◆ FCoE/DCBはEthernetとFCのI/O集約を行う場所で使用される
 - ▶ 主にWebなどのフロント系のサーバにて使用される
 - ▶ バックエンド系では既存のFCが従来通りに使用される

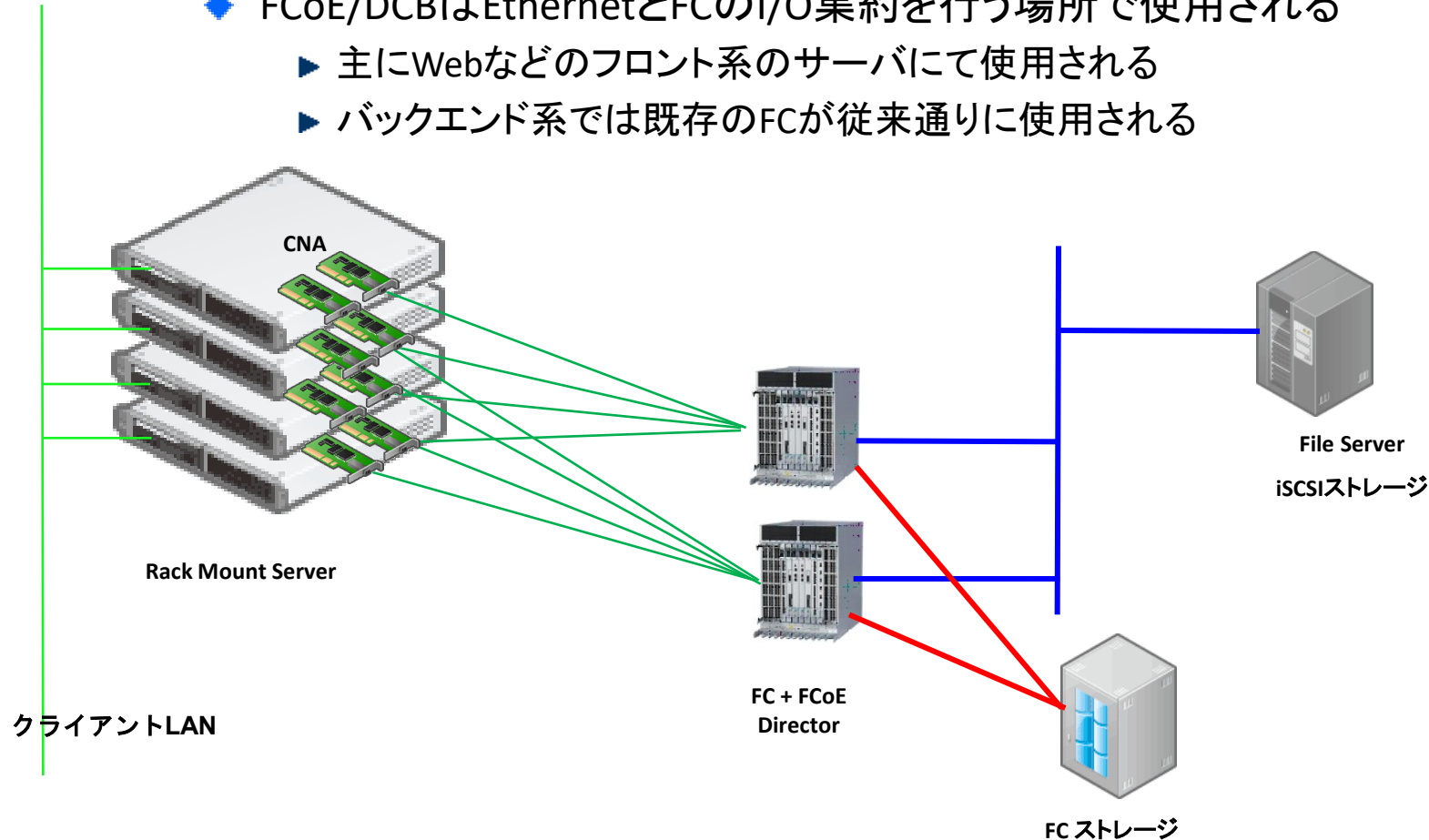


4-2-7. FCoE/DCB とは？

■ FCoE/DCBとFC-SANの関係

◆ FCoE/DCBはEthernetとFCのI/O集約を行う場所で使用される

- ▶ 主にWebなどのフロント系のサーバにて使用される
- ▶ バックエンド系では既存のFCが従来通りに使用される



4-2. FC-SANとその他のストレージネットワーク まとめ

■ SANで利用されるプロトコル

- ◆ FC,iSCSI,FCIP,FCoE 2010年代後半にSSDが台頭してくるまでは、ほぼ「**SCSI**」プロトコルを利用

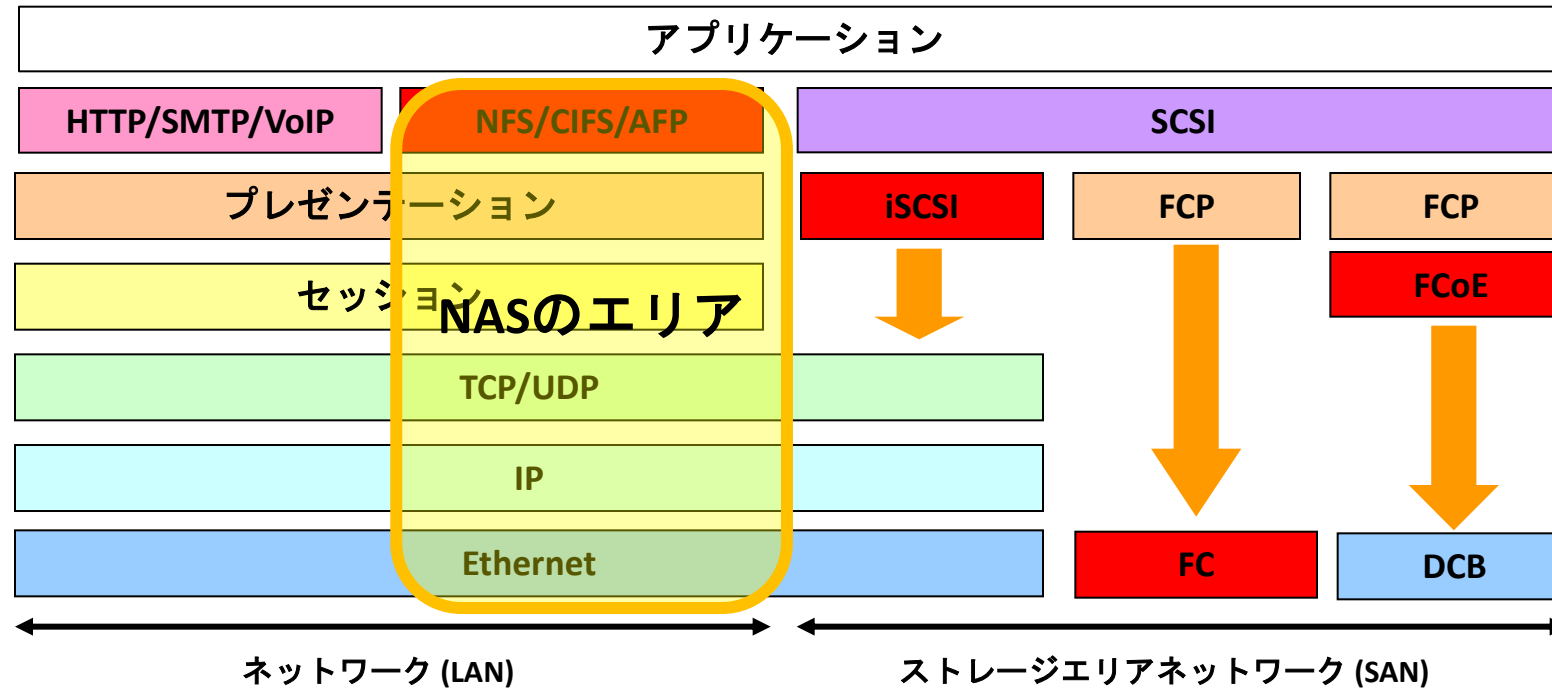
■ FC-SANとIP-SANの違い

- ◆ プロトコル どちらも使うのは“SCSI”
 - FC-SAN: FCP-SCSIにてSCSIに変換
 - IP-SAN: TCP/IPにて、SCSIプロトコルをTCP/IPフレームにカプセル化する
- ◆ 物理層
 - FC-SAN: FibreChannel環境
 - IP-SAN: Ethernet いわゆるLAN
- ◆ 安定性
 - FC-SAN: 専用のFCネットワークで堅牢なシステム向き
 - IP-SAN: 既存のLANも使用を考慮できる、このため拡張は容易であるが安定性はFCに譲る
- ◆ 費用
 - FC-SAN: 割高は否めない
 - IP-SAN 10Gのコストの低下から廉価に構築可能

NAS

4-2. FC-SANとその他のストレージネットワーク

4-2-7. NAS (Network Attached Storage) とは？



- NAS: 各OSに対してファイルアクセスを提供する。
- FCP: FibreChannelへSCSIをマッピングし、ストレージアクセス(SCSI)を提供
- iSCSI: TCP/IP上にSCSIをマッピングする。
- FCoE: 次世代データセンターイーサネット(DCB)上にFCをカプセル化。TCP/IPは利用しない。

4-2. FC-SANとその他のストレージネットワーク

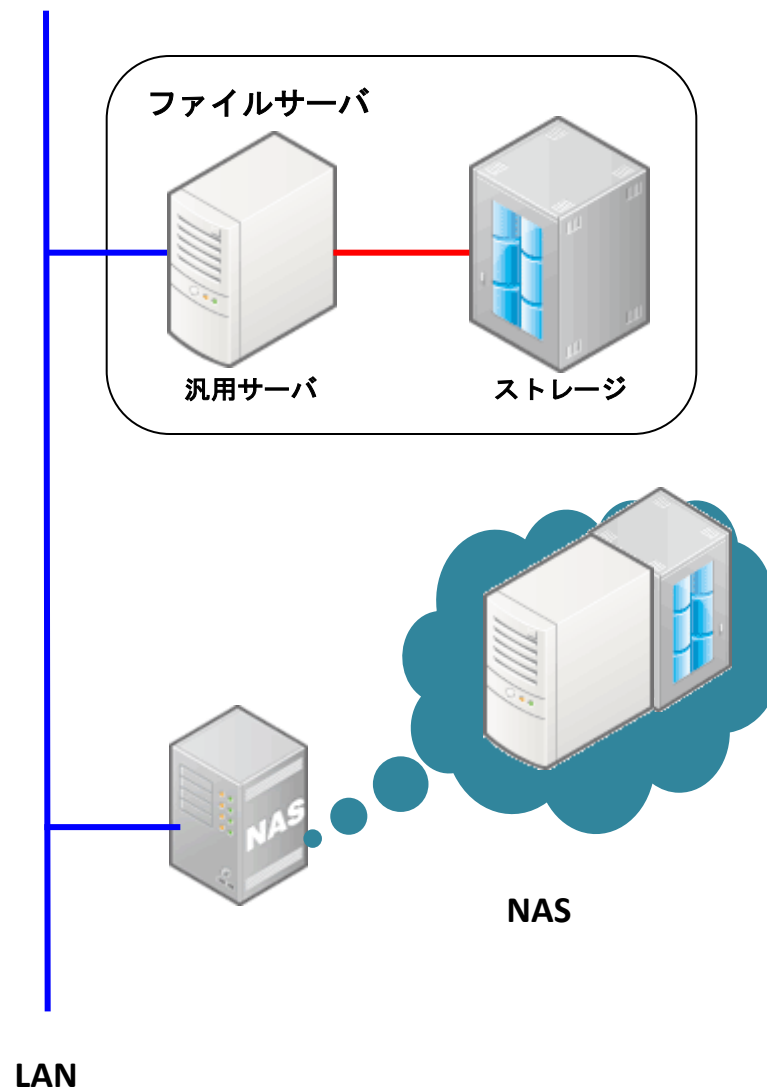
4-2-7. NAS (Network Attached Storage) とは？

■ SANとNASの違いは？

- ◆ SAN (Storage Area Network)
 - ▶ ネイティブなストレージ
 - ▶ ブロックアクセス
- ◆ NAS (Network Attached Storage)
 - ▶ ファイルサーバ機能を持ったストレージ
 - ▶ ファイルアクセス

■ 本来SANとNASは同列で比較する対象ではない。

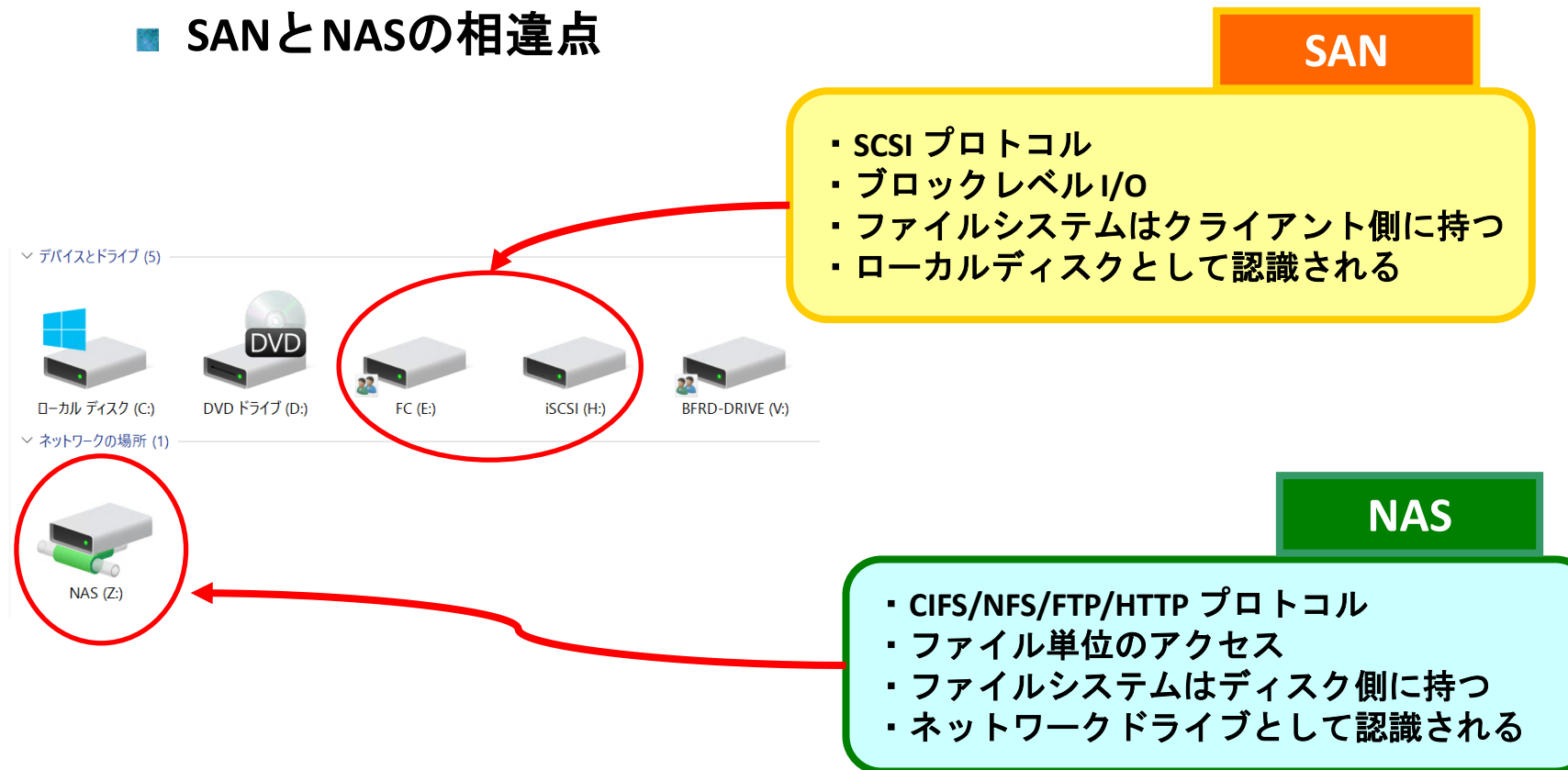
- ◆ NAS: ファイルサーバ装置
- ◆ SAN: ストレージをネットワーク状に配置した構成



4-2. FC-SANとその他のストレージネットワーク

4-2-7. NAS (Network Attached Storage) とは？

■ SANとNASの相違点



テープも、ローカルデバイス (DVDやCDドライブ) のように使えます。

LTFS (Linear Tape File System) 技術

テープにファイルシステムに似たファイル記録形式

ローカル/ネットワークデバイスとしてマウントしてファイルのドラッグ&ドロップが可能！



4-2. FC-SANとその他のストレージネットワーク

4-2-8. InfiniBand

2000年10月にIEEE802.3規格化

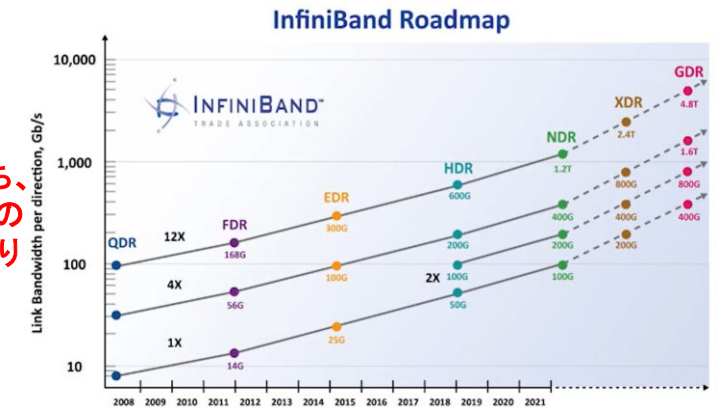
Next Generation I/O (NGIO)とFuture I/Oが統合されて生まれたI/Oアーキテクチャ HCA(ホストチャネルアダプタ)とTCA(ターゲットチャネルアダプタ)により接続する

※高速EthernetNICと兼用もされている(Mellanox VPIカードなど)

特徴

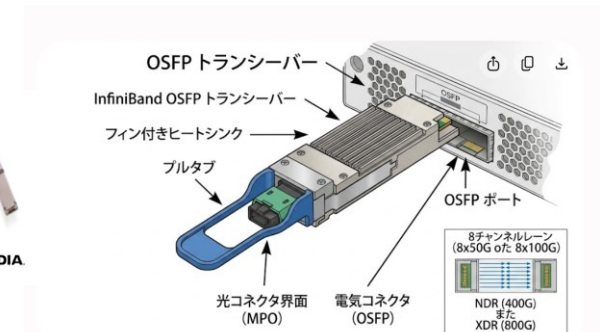
データをプロセッサを介さずに転送する**RDMA**を実現、プロセッサの負荷を劇的に軽減できる。
このため、**低レイテンシー**であり、**通信品質(QoS)**を担保できるので非常に高い**RAS(信頼性、可用性、保守性)**を持ち、
その特徴から**HPC(コンピュータ・クラスターやスーパーコンピュータ)**のノード間接続に使われ、**DeepLearningサーバ**の
インターフェースや、**複数のAI(GPU)サーバ**のノード間通信にも使われており、**昨今ではChatGPT,生成AIの台頭により**
また、注目されている技術になります (対抗馬として**UltraEthernet**が、**コンソーシアム**で策定中)

低レイテンシー 1から1.5 μ 秒程度 ※メモリ(DDR3)が0.1 μ
転送品質(QoS) 物理層(アダプタレベル)で実現、リンクレベルで仮想レーンを提供
符号化 SDR-QDRまでは8/10B FDR-NDRは66/64B
転送レート 2.5Gbps ~400Gbps その先800、1600 も策定されている。
ケーブル 1本(x1)、**4本(x4)**、8本(x8)、12本(x12)とレーンを束ねることができる。
銅線(DAC)で最大10m程度、光ファイバで最大250m程度



InfiniBand 一方向データレート									
	規格発行年	Line code	理論値 (Gbit/s)	スループット (Gbit/s)				レイテンシー (μ s)	
				1x	4x	8x	12x		
SDR	2000 ^[5]	8b/10b	2.5 ^[5]	2	8	16	24	5	
DDR	2004 ^[5]		5 ^[5]	4	16	32	48	2.5	
QDR	2004 ^[5]		10 ^[5]	8	32	64	96	1.3	
FDR	2012 ^[6]	NRZ	14.0625 ^[6]	13.64	54.54	109.08	163.64	0.7	
EDR	2012 ^[6]	PAM4	25.78125 ^[6]	25	100	200	300	0.5	
HDR	2020 ^[7]		53.125 ^[8]	50	200	400	600	<0.6	
NDR	2021 ^[8]		106.25 ^[8]	100	400	800	1200		
XDR	2023 ^[9]		200	200	800	1600	2400		TBD
GDR	TBA	TBD	400	400	1600	3200	4800		
LDR			800	800	3200	6400	9600		

ウィキペディアより抜粋



4-2. FC-SANとその他のストレージネットワーク

4-2-8. InfiniBand

InfiniBand コンポーネント概要

レイヤー2ネットワーク

■ Host Channel Adapter (HCA)

- IBリンク終端に位置しトランスポートレベルの機能を実行し、verbsインターフェイスをサポートするデバイス

■ Switch

- 同じIBサブネットの1つのリンクから別のリンクにパケットをルーティングするデバイス

■ Cable

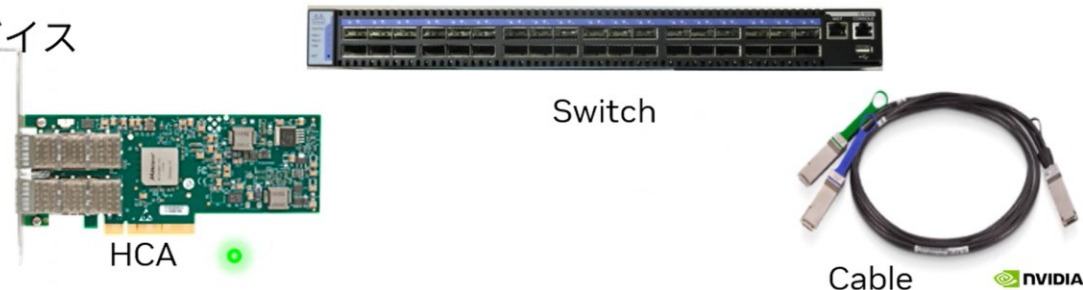
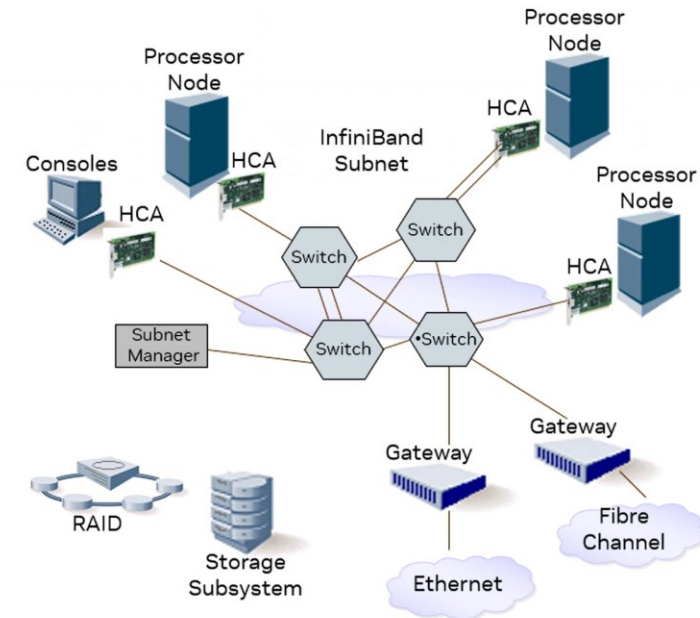
- サーバ間、スイッチ間を接続する導線でOptical and Copperの2 Type ある

■ Router

- 異なるIBサブネット間にパケットを転送するデバイス

■ Bridge

- InfiniBand とEthernetを接続するデバイス



4-2. FC-SANとその他のストレージネットワーク

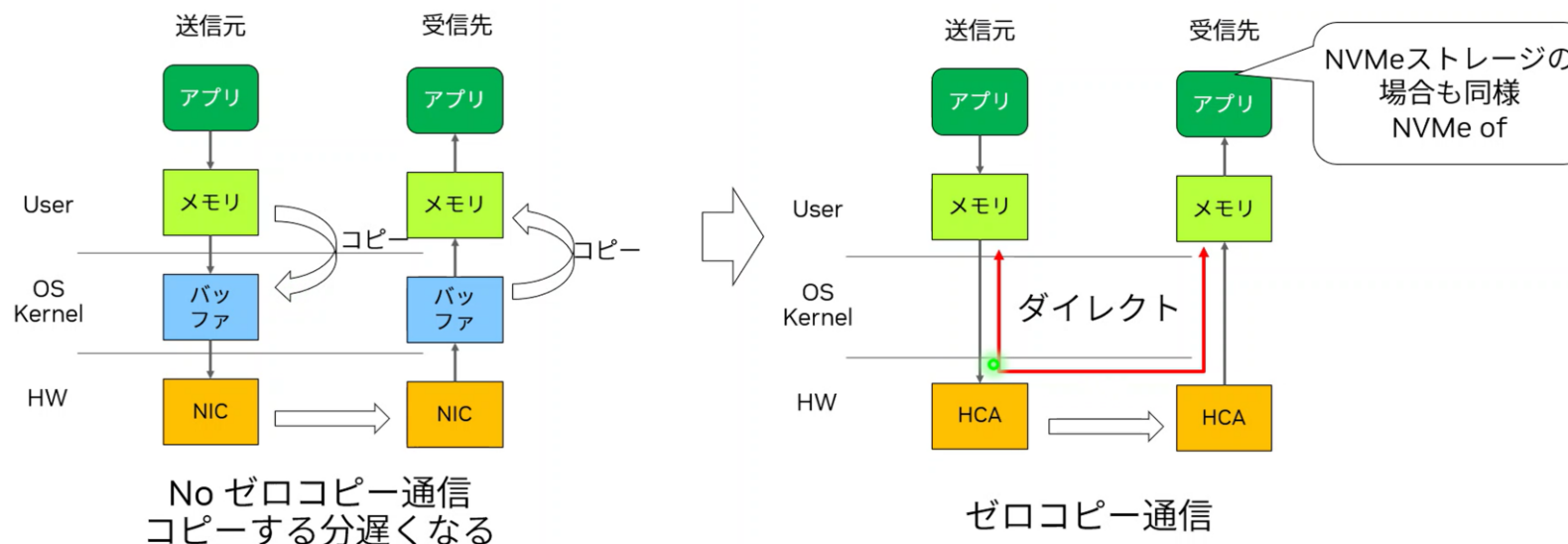
4-2-8. InfiniBand

2000年10月にIEEE802.3規格化

InfiniBandの通信特徴 ゼロコピー(Zero Copy)とRDMA

OS、Kernelが通信に介在しない？

リモート(Remote) ダイレクト(Direct) メモリ(Memory) アクセス(Access) ➡ RDMA通信



OSやKernelのオーバーヘッドによる遅延がない ➡ 速くて遅延が少ない

OSやCPUが処理していること（プロセス）に割り込むことなくリモートマシンのメモリを直接アクセスを可能とする
CPUへの割り込みを無くすことで、I/Oオペレーションのためのコンテキストの切り替えを減らすことにより、従来のネットワーク通信方法に比べ、低遅延かつ高帯域の転送ができる

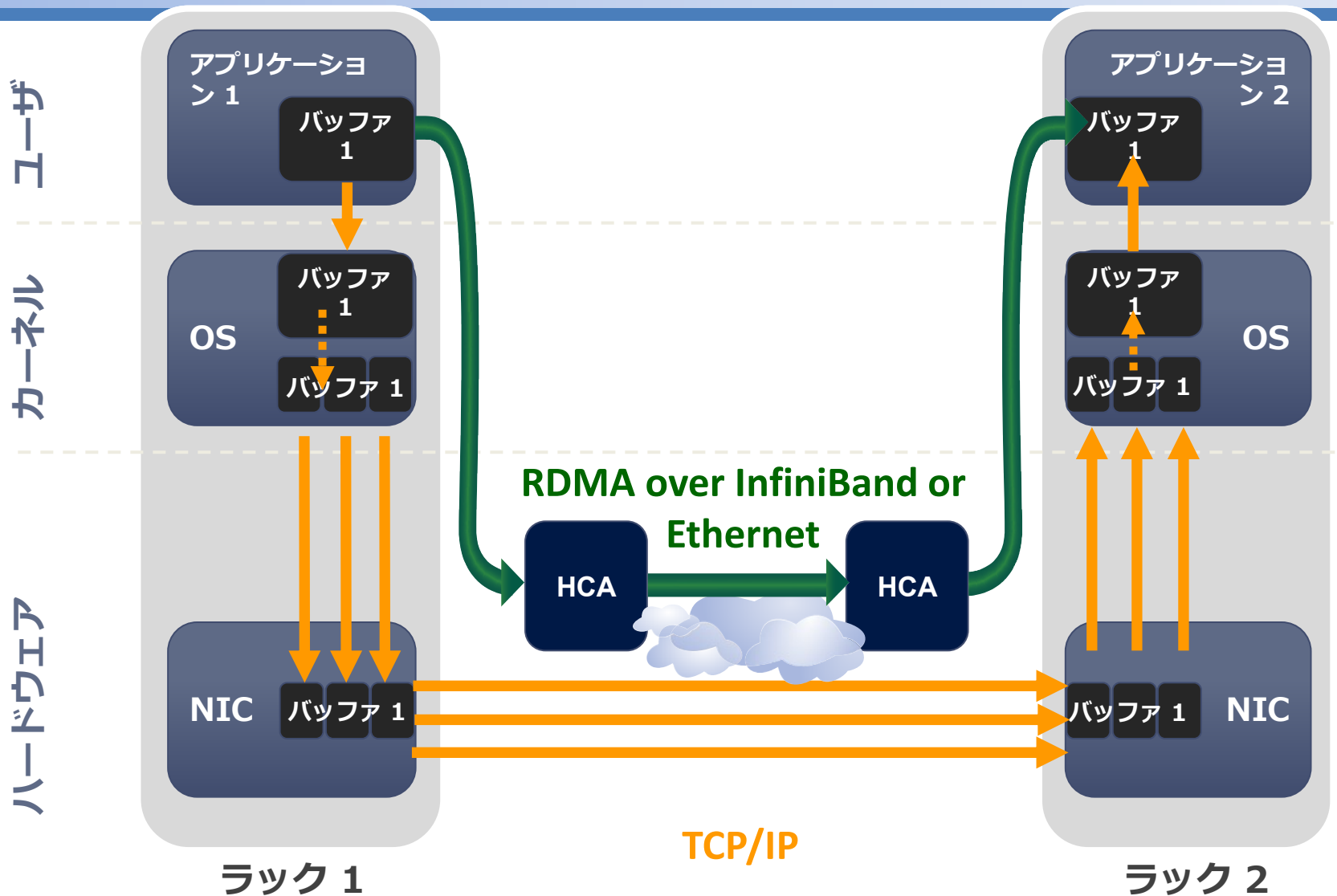
参考ビデオ:

<https://academy.mellanox.com/en/course/rdma-programming-intro/?cm=446>

(英語)

4-2. FC-SANとその他のストレージネットワーク

RDMAの動作



バッファのデータコピー、プロセッサの負担を最小限にする。

ストレージネットワークの変革

現在のストレージネットワークのトレンド

■ SAN

- ◆ サーバ仮想化により接続ポート単位のIOは増大
 - ◆ スイッチ統合の必要性が増加(マルチプロトコル対応)
 - ◆ 仮想化サーバ単位でのI/Oコントロール
 - ◆ ポートの仮想化、QOS
 - ◆ SSDの台頭により転送路の帯域不足が露呈
 - ◆ FCより高速なネットワーク
 - ◆ RDMA技術の浸透
- SR-IOV
 - ネットワークのロスレス化
 - ブロックIOプロトコルの変貌
 - 25~100/200/400/800Gbps Ethernet/IB
 - RoCEv2 iWARP

■ ストレージ環境

- ◆ ハイエンド統合からハイエンド・ミッドレンジの共存環境へ
 - ◆ オールフラッシュストレージ
 - ◆ オブジェクトストレージ
 - ◆ 新しいインターコネクト技術(仕様)の登場
- 処理能力の向上
 - クラウドの浸透
 - NVMe (Non-Volatile Memory Express)
 - SCM(Storage Class Memory)
 - CXL (Compute eXpress Link)

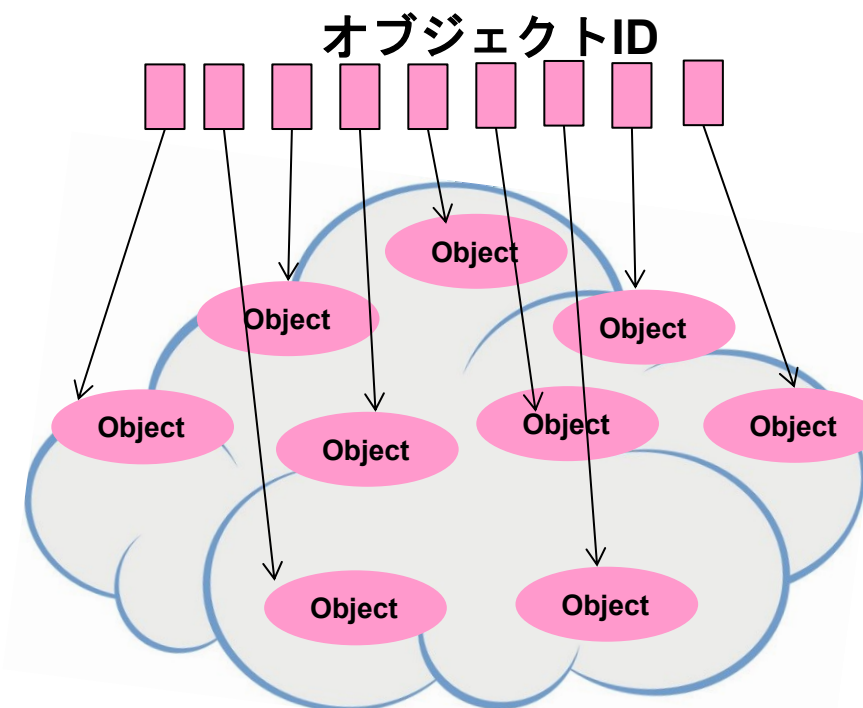
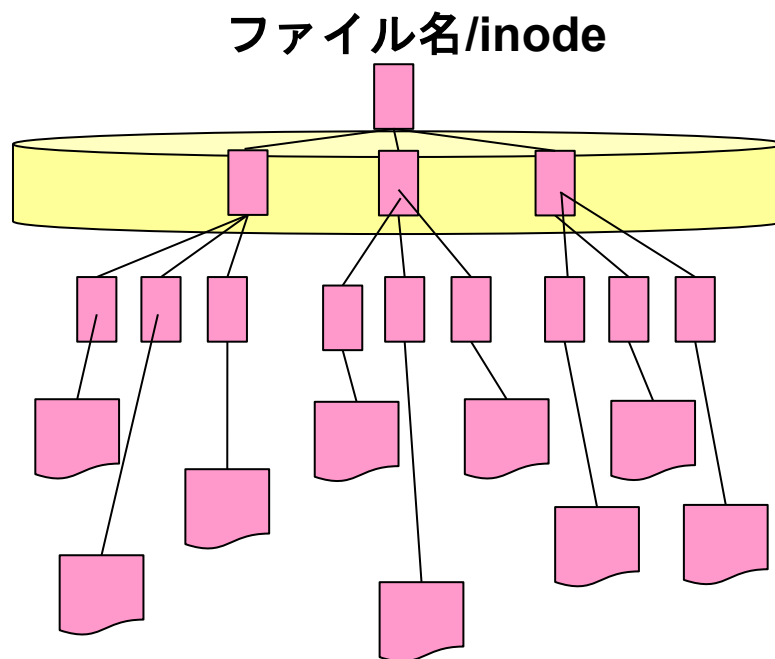
4-3. その他のストレージ技術

4-3-1.オブジェクトストレージ

あらたなるストレージアクセス方法

SCSI(ストレージプロトコル)やNFS/CIFS(ネットワークファイルシステム)使用せずに、REST(Representational State Transfer)やSORP(Simple Object Access Protocol)などのWebプロトコルを使用して、データをオブジェクト形式で保存する。

データを、階層構造で管理せず1つのオブジェクトとして保存



4-3. その他のストレージ技術

4-3-1.オブジェクトストレージ

対照比較

	対照比較： オブジェクト・ストレージ vs. 既存ストレージ		
	オブジェクト ストレージ	ファイルベース ストレージ	ブロックベース ストレージ
トランザクション 単位	オブジェクト、即ち カスタム・メタデータ 付きファイル	ファイル	ブロック
サポート される更新の タイプ	その場で書き換える 更新はサポートせず。 更新は新オブジェクト バージョンとして生成	その場で書き換える 更新をサポート	その場で書き換える 更新をサポート
プロトコル	HTTP上の RESTおよびSOAP	CIFSおよびNFS	SCSI, Fibre Channel, SATA
メタデータ サポート	カスタム・メタデータ をサポート	固定ファイルの 属性	固定ファイルの 属性
最も適合する 対象	比較的静的な ファイル・データ およびクラウド データ	共有ファイル データ	トランザクション データおよび更新が 頻繁なデータ
最大の長所	拡張性および 分散アクセス	簡易なアクセス および共有ファイル の管理	高パフォーマンス
制限事項	頻繁に更新される トランザクション データには不向き 排他制御機能付き 共有プロトコルを 備えていない。	データセンターの 外への拡張が 困難	データセンターの 外への拡張が 困難

JDSFサイトより抜粋

4-3. その他のストレージ技術

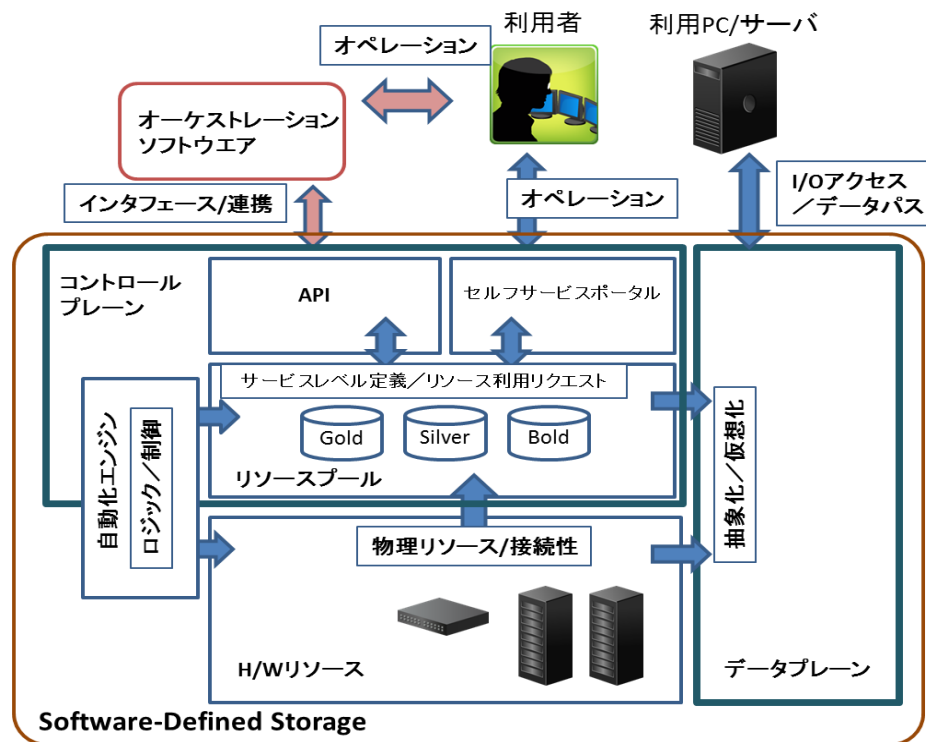
4-3-2 Software-Defined Storage(SDS)

SDSの必須要件

([SNIA White Paper: Software Defined Storage](#)より)

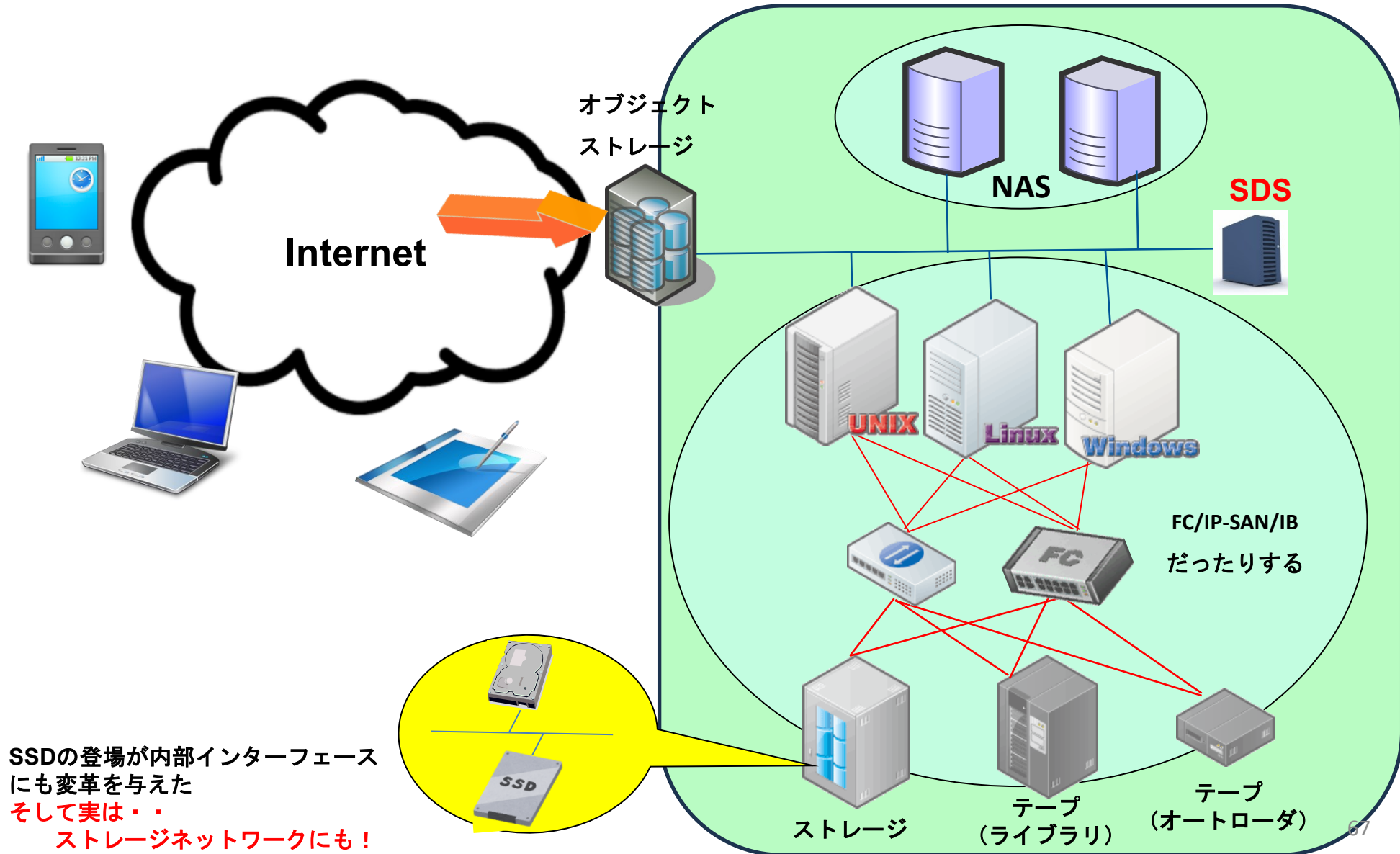
- **自動化**
 - ストレージの運用を自動化し、運用の複雑性やコストを低減できる
- **標準化されたインタフェース**
 - ストレージの管理や活用に関わる操作を標準的なAPIで操作できる
- **仮想化されたデータパス**
 - ブロック、ファイル、オブジェクトといった多様なインタフェースをサポートできる
- **拡張性**
 - 可用性やパフォーマンスに影響無く、柔軟にストレージインフラを拡張できる
- **透過性**
 - ストレージ利用者自身がリソース使用率やコストなどをモニタリングし管理できる

SDSの全体像



つまりは・・・

クラウドストレージサービスもストレージネットワークの使い方(集合体)のひとつ



4-4. 進化を続けるストレージインターフェース技術

4-4-1 SSD用のブロックプロトコルの登場



NVMHCI 1.0 2008/4 → NVMe 1.1 2017/10 現在1.3a <http://www.nvmexpress.org/>

従来のAdvanced Host Controller Interface (AHCI) ではSSDの性能に対して不足
このため NVMHCI (Non-Volatile Memory Host Controller Interface)、NVM Express (NVMe)が登場した。

AHCI と NVMe の仕様の違い^[19]

	AHCI	NVMe
最大キュー深度	1 コマンド キュー 32 コマンド / キュー	65536 キュー 65536 コマンド / キュー ^[21]
Uncacheable register accesses (各々 2000 サイクル)	6 / Non キュー コマンド 9 / キュー コマンド	2 / コマンド
MSI-X and interrupt steering	単一の割り込み ステアリング 無し	2048 MSI-X 割り込み
パラレルizm および、マルチプル・スレッド	コマンドを発行するために、 synchronization lock が必要	No locking
4 KB コマンドの効率	コマンド・パラメータは 2 serialized host DRAM フェッチ必要	コマンド・パラメータは 1 フェッチ (64 Byte) で済む

・NVMeは、SSD接続用として開発されたSCSIやSATA(Serial ATA)と同じストレージを接続するためのインターフェース規格

従来の回転ディスクシステムではなくフラッシュストレージに特化したプロトコル

・4KBの転送に必要なメッセージが2つではなく1つ

・コマンド処理キューが1つではなく6万5536個

※従来のプロトコル：コマンドキューが1つしかなく、最大で64個のコマンドしか保持できない

1台のSSDは5Gbps(800MBs~)以上

多数のディスクI/O要求を同時に処理するサーバーの場合

NVMeにすることにより、大幅なディスクI/Oの高速化が実現できる

※このほかに ATA Express SCSI Expressなどもある

PCIe3.0×4の内部接続で、実行速度は3Gbps以上 SAS-3.0の約3倍程度速い

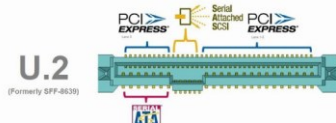
現在の製品仕様

・標準サイズのPCI Express 3.0の拡張カード型



・2.5インチのカード型 U.2コネクタ(SFF-8639)

4レーンのPCI Expressインターフェースを内部に持つ



・M.2コネクタ直接接続型

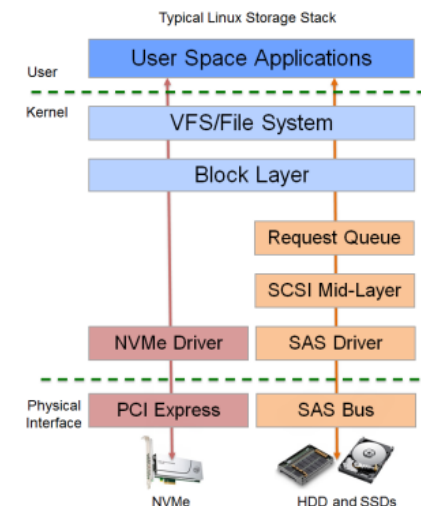
省スペースPCやノートPCのSSDに活用されています



・M.3コネクタ (NGSFF NewGenerationSmallFormfactor)

サーバ(エンタープライズ)用途として、M.2にはないホットスワップ機能やデュアルポート機能を追加したもの

現在はEDSFFに統合された模様・・・



4-5 ストレージネットワークの進化

4-5-1. 高速ネットワークとSSD

■ 2000年代

- ◆ 高速ネットワークの浸透と進歩 10Gbps 25Gbps 40Gbps
→**IP-SANへの拡張**
- ◆ 拡張された技術の発展
 - ▶ iSCSI オフロードエンジン
 - ▶ FCoE (FibreChannel Over Ethernet) ※**FC-SANの変形**

■ 2010年代中盤以降

SSD（フラッシュ）を使用したストレージが浸透し始める。
イニシエータ、ターゲットの両端の処理能力が劇的向上
高速なネットワークを利用した転送方式が数々生まれてくる

- ◆ iSERなどの拡張版iSCSI
- ◆ IBの十八番のRDMAをEthernetにて実現（RoCE、iWARP）

■ 2020年代

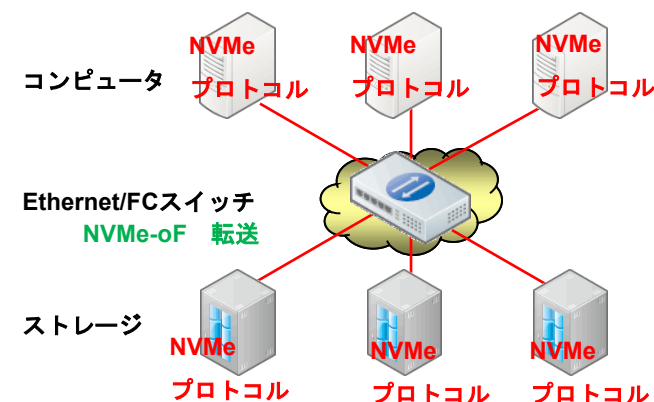
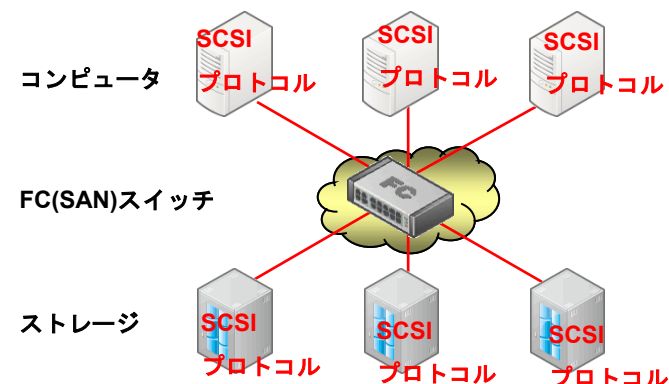
回転ディスク用のSCSIプロトコルに限界……

SCSIプロトコルから NVMeプロトコルへ
NVMeをストレージネットワークへ

NVMe-oF が登場！！

IP-SANを活用（25～200Gbps）

FC-SANを活用（16～128Gbps）



でも・・・やはり
PCI expressバスでは距離の制限が厳しく
ネットワークとしては・・・

高性能なNVMe SSDをストレージネットワーク
として使いたい・・・

4-5. ストレージネットワークの進化

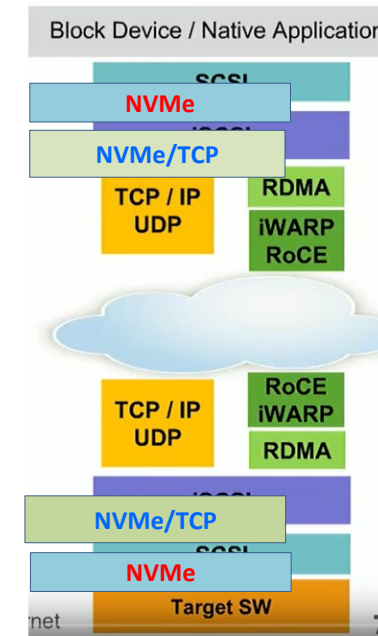
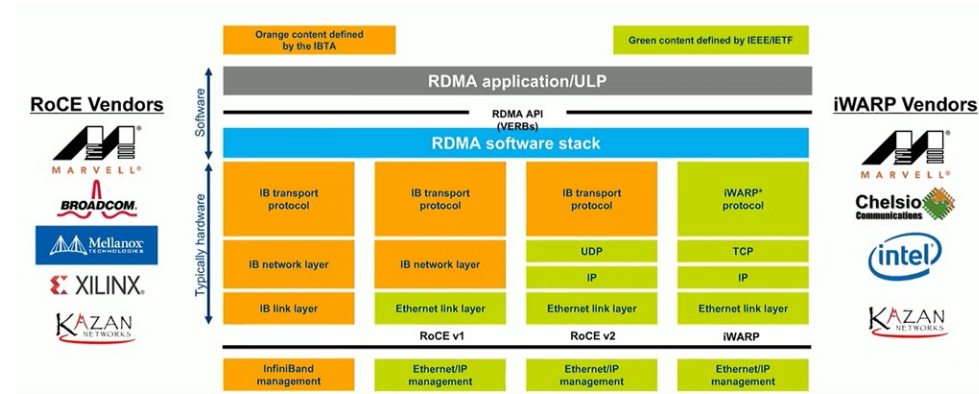
4-5-2 NVMeとRDMA

NVMeはフラッシュストレージに特化したプロトコルでありPCI-Express上で超高速アクセスと多重処理をもたらします。しかしながらPCI-Expressバス上を通るプロトコルであるため、**距離の制限**がありました。これを距離の制限から解放するため・・・NVMe over Fabricが策定されNVMeを距離の制限から解放して幅広い展開を可能します。

特徴

- **NVMe Over RDMA RoCE, Infiniband**

物理コンバージドイーサネット（DCE：データセンターブリッジング・ロスレス・イーサネットネットワーク）でRDMAを実現



- **FC-NVMe**

FCフレーム内にカプセル化されたNVMeコマンドセット
一般的なFCプロセスに準拠、SCSIコマンドと簡単に共存できる。

- **NVMe over TCP (NVMe / TCP)**

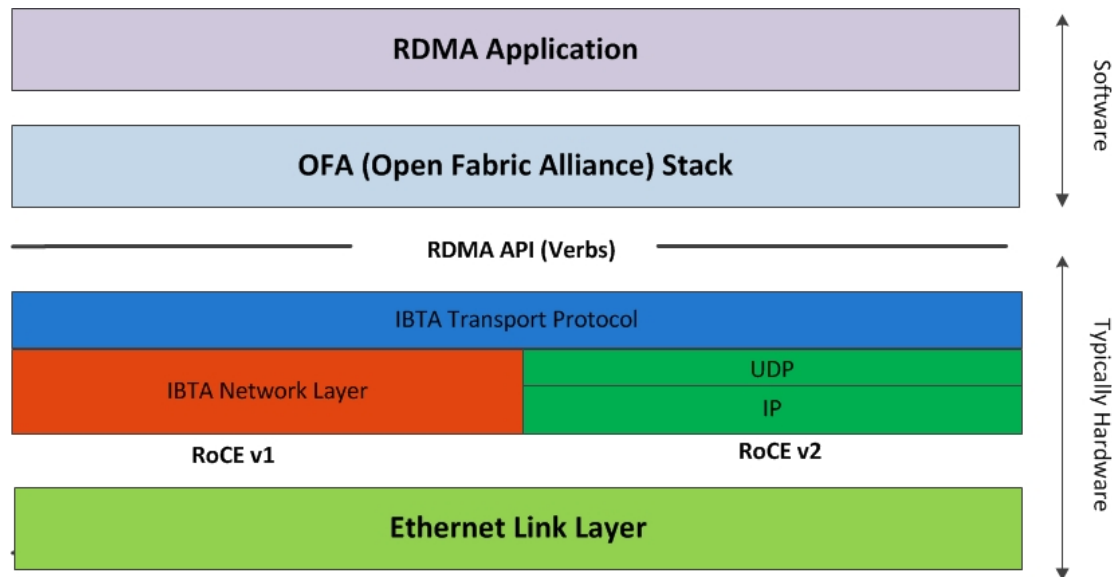
Ethernet上でTCPデータグラム内にNVMeデータ（PDU）をカプセルする。
RoCEとNVMe / TCPは両方ともイーサネットを使用しますが、NVMe / TCPはI/Oにメッセージングセマンティクスを使用するため、NVMe / FCのように動作します。まあ・・・カプセル化なのでiSCSIのような感じです

4-5. ストレージネットワークの進化

4-5-3 RoCE(RoCE、RoCEv2)アーキテクチャ

もともとIBでRDMA転送を行うためのAPI (IB-verb)で
それに、RoCE(RDMA over Converged Ethernet)も加わり、RDMA-verbとなった

<https://www.nminoru.jp/~nminoru/network/infiniband/iba-software-stack.html>



- RoCE インターフェースとして Verbsと呼ばれるAPIを提供
- RoCEv1とRoCEv2の違いはネットワークレイヤー構成
- RoCEv1とRoCEv2間の直接通信はできない

RoCE(v1): EthernetのL2フレームにInfiniBandのNetwork Layer を乗せたものであったため、同ドメイン内の転送のみをサポート

RoCEv2 : EthernetのLink Layerの上に一般的なIP/UDP レイヤーを一を乗せる事でネットワーク間を跨ぐことが出来るようになりました。

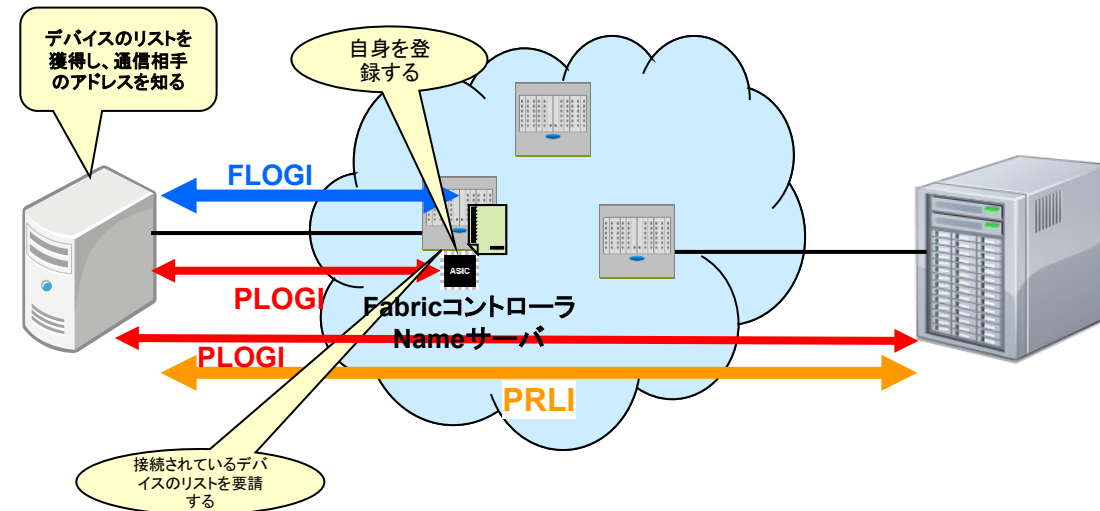
4-5. ストレージネットワークの進化

4-5-4 FC-NVME デバイスディスカバリ、混在可能

FCは、ほとんど変わらずMVMeに対応できる

前述の通り、スイッチがすべて知ってます！

1. ファブリックログインの実行(FLOGI)
2. Name Server にログイン
3. Name Server に自身を登録
4. ファブリックコントローラ へ登録(SCR)



※参考資料

Fabricトポロジでのログインのプロセス
ELS(Extended Link Service) を参照

5. Name Server に GID_FF(Get PortID_FC4_features)を発行
NVMe over FCかつNVMeターゲットポートをサポートするポートIDのリストを取得
6. 獲得したIDリストから accept CT_IU に含まれるそれぞれの Port ID に対して下記の手順を実行する
 - (1) PLOGI の送信(FC2 リンクの確立);
 - (2) PRLI(FC4 接続の確立)
 - (3) NVMe Report Subsystems コマンドを発行
ターゲット NVMe_Port 経由でアクセス可能な NVMe Qualified Names(SUBNQN) のリストを取得

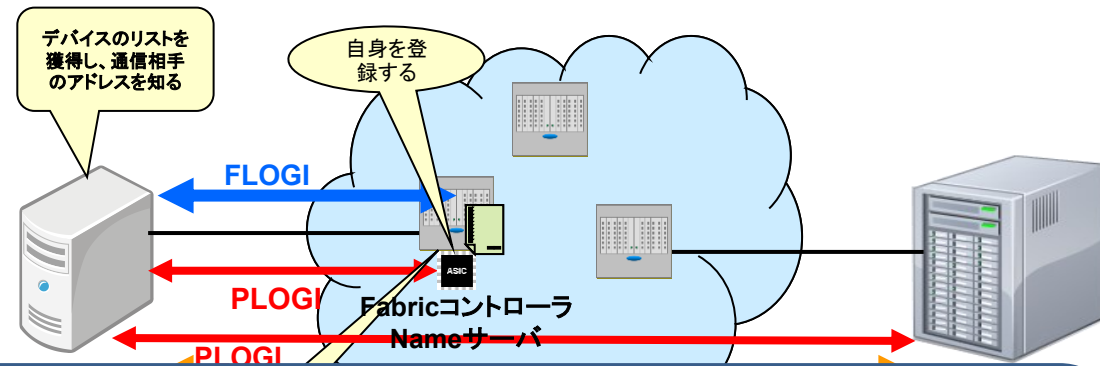
4-5. ストレージネットワークの進化

4-5-4 FC-NVME デバイスディスカバリ、混在可能

FCは、ほとんど変わらずNVMeに対応できる

前述の通り、スイッチがすべて知ってます！

1. ファブリックログインの実行(FLOGI)
2. Name Server にログイン
3. Name Server に自身を登録
- 4.



スイッチがFabric全体を管理しているので、デバイスリカバリ(発見、接続)が容易でFC-SANでもスイッチ内でも複数の上位プロトコル(SCSI・NVMe・TCP/IPなど・・・FC-4)が混在できます。

つまり・・・既存(SCSIプロトコル)と新規(NVMeプロトコル)の共存も可能です。

5.

NVMe over FCからNVMeターゲットポートをリポートするポートIDのリストを取得
6. 獲得したIDリストから accept CT_IU に含まれるそれぞれの Port ID に対して下記の手順を実行する
 - (1) PLOGI の送信(FC2 リンクの確立);
 - (2) PRLI(FC4 接続の確立)
 - (3) NVMe Report Subsystems コマンドを発行
ターゲット NVMe_Port 経由でアクセス可能な NVMe Qualified Names(SUBNQN) のリストを取得

4-5. ストレージネットワークの進化

4-5-5 NVMe-oF 第3の試み

NVMe over Fabricは開発の初期段階では

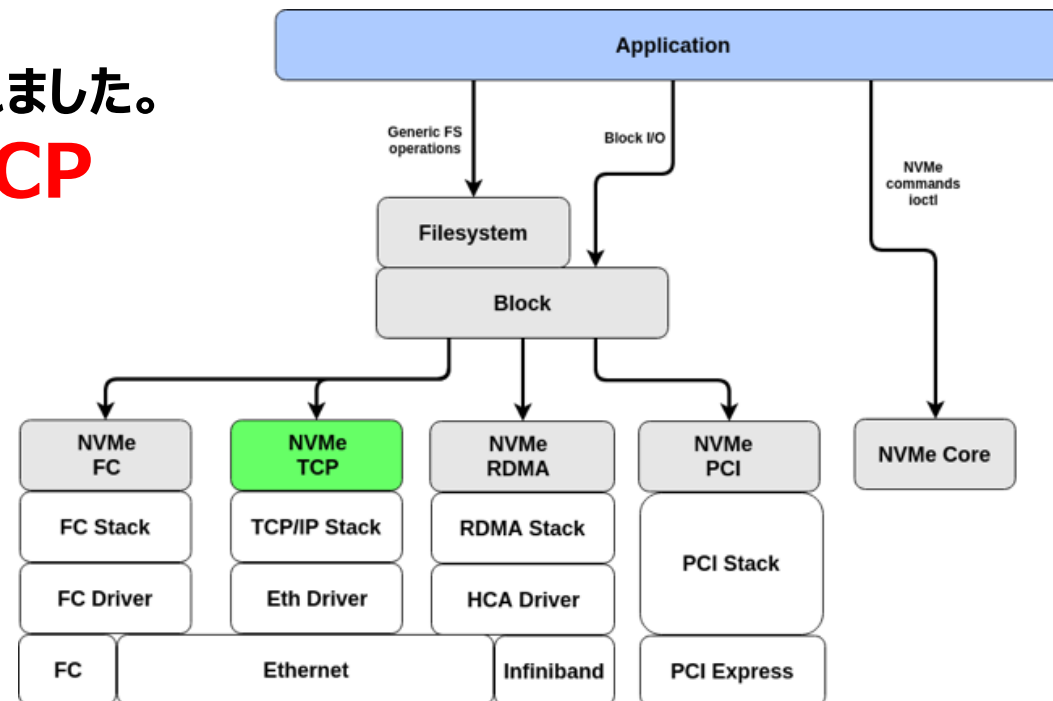
- ・NVMe/RDMA
 - ・NVMe/FC
- RoCE iWARP IB
FibreChannel

を介してNVMeストレージに接続することが考案されてました。

2018年末・・・これにされにもう1つのプロトコルが

NVMe-oF™ 1.1 で追加されました。

NVMe/TCP



4-5 ストレージネットワークの進化

4-5-6 NVMe Over Fabric

長所と短所

NVMe Over RDMA(RoCE)

- 長所 :** NVMe / RoCEは、非常に一般的なネットワーク標準を活用して、転送にイーサネットネットワークを使用します。
RoCE v2製品は、いくつかのエンタープライズストレージベンダーによって開発されています。
- 短所 :** イーサネットは基本的に損失が大きく、信頼性の低いネットワークに対処するように設計されているため、エラー修正と再送信のオプションが多数あり、
この整合性の調整とネットワーク環境はロスレスである必要がある

FC-NVMe

- 長所 :** NVMe-FCは、パフォーマンスの大幅な向上とワークロードレイテンシの削減を実現します。FCプロトコルは安定、成熟しており、効率的です。
一貫して高いパフォーマンスを提供します。FCからNVMe/FCに簡単に移行、FCP/SCSIとの混在も可能。
物理層（FCとPCIe）の親和性が高いのでNVMeとも親和性がある。
- 短所 :** インフラはFC機器が必要で高価

NVMe-TCP

これまで、FCまたはInfiniBandネットワークは高コストであり、NVMe-oFの導入には二の足を踏んでいる状態であることから、市場のこのギャップに対処するために、NVMe.orgの関係者はTCPデータグラムを含むイーサネットLANをトランスポートとして使用する新しいNVMe-oF標準（NVMe / TCP）を開発および公開しました。

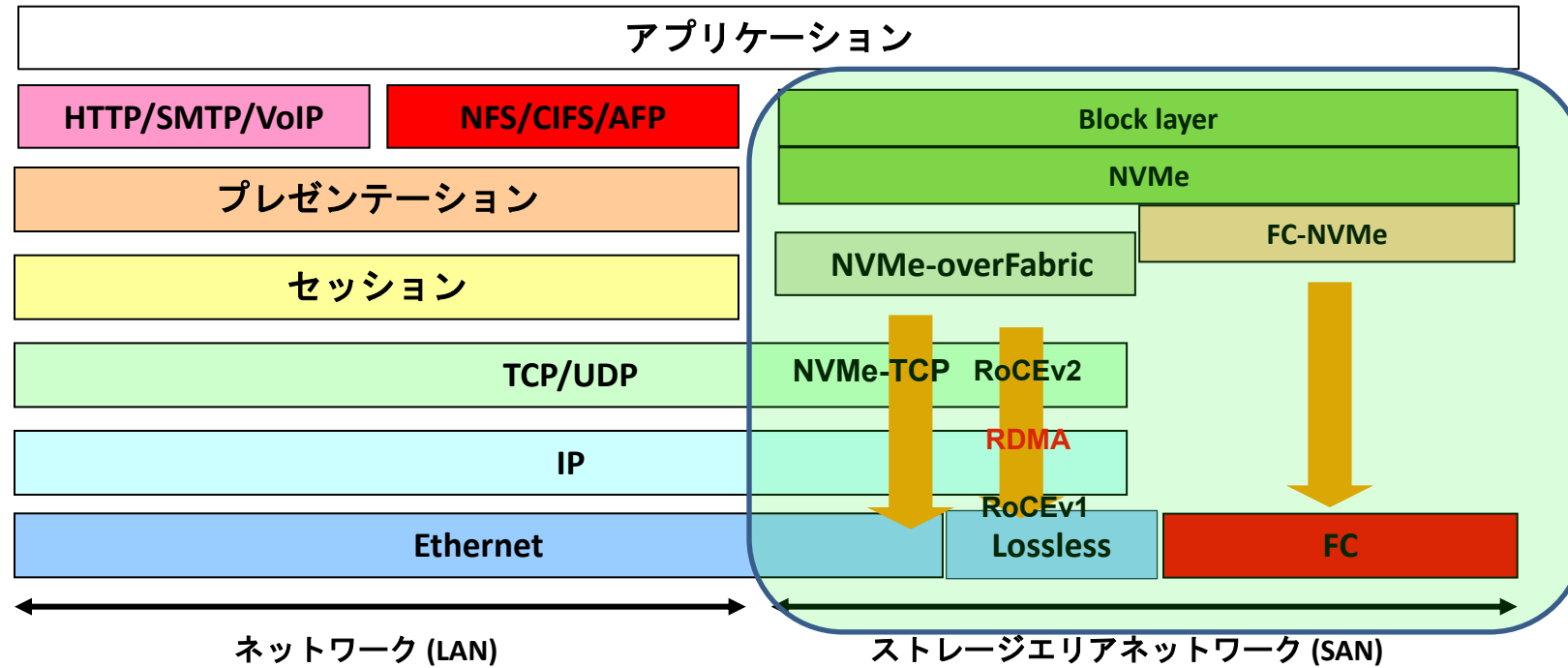
- 長所 :** 既存のネットワーク環境を変更することなく実装できるので、コストパフォーマンスに優れる
- 短所 :** TCP/IPプロトコルに依存すると思われる →これも最終的には、HWオフロードで回避するだろう……

生い立ちはiSCSIと同じようなものですね……1998年末ごろに出てきたiSCSI
最初はTCP/IP上にバスのプロトコルを流すなんて無理なんじゃないか？
という声も……
でも10年もたたず、
iSCSiは汎用性に優れ、あっという間にストレージネットワークの一員になった

	FC-NVMe	NVMe/RoCE	NVMe/TCP
Bandwidth	16-32GFC	10-100GbE	10-100GbE
IOPS	~1M	~1.5M	~1.5M
Network Type	Fibre Channel SAN	Lossless Ethernet	Ethernet Network
Scalability	Yes, dedicated fabric	Limited – 1 to 2 hops	Yes
Adapter Latency	Low	Very Low	High
OS Support	Linux, VMware, Windows	Linux, VMware	Linux, VMware
Security	Yes, dedicated FC network	No	No

4-5. ストレージネットワークの進化

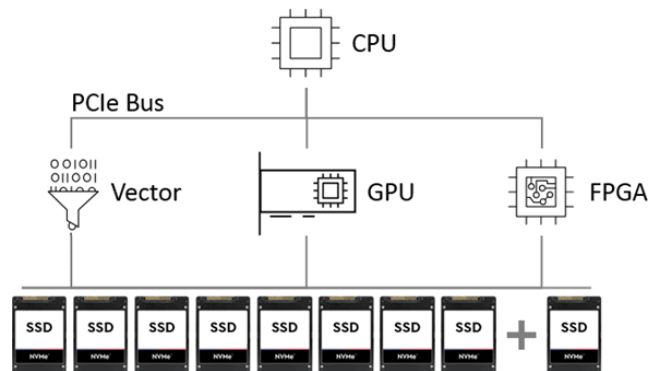
4-5-6 新たなSAN NVMe Over Fabric



NVMeはSCSIに変わってSANでも今後主流になるであろう

NVMeでの装置内（サーバやストレージ内部）でRAIDを構築する場合
RAIDコントローラとは直接PCIeバスで接続します。
従来のSASやSATAのようにカードとSSDをケーブルで接続する必要はありません

※サーバでは全てのストレージベイの-slotでPCIeがサポートされていない場合は、RetimerカードやPCIeライザーカードで、NVMeSSDの接続数を増やす必要がある場合があります



ホストとの接続例



Figure 3-18. Cable Routing Diagram

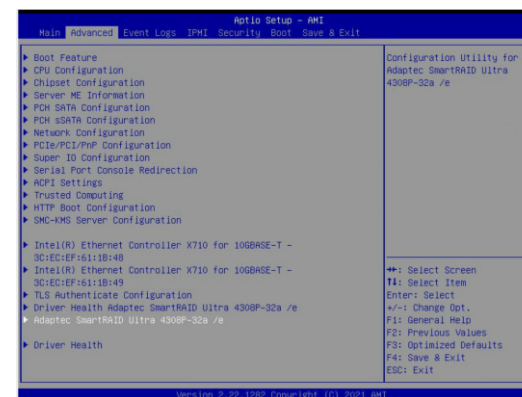


Figure 3-19. Storage Backplane BPN-NVMe4-216N-S24

NVMe RAIDカード

■ Microchip Adaptec SmartRAID4308

UEFIに対応した「HII Configuration Utility」を搭載しているので、NVMe SSDのRAIDからOS起動が可能



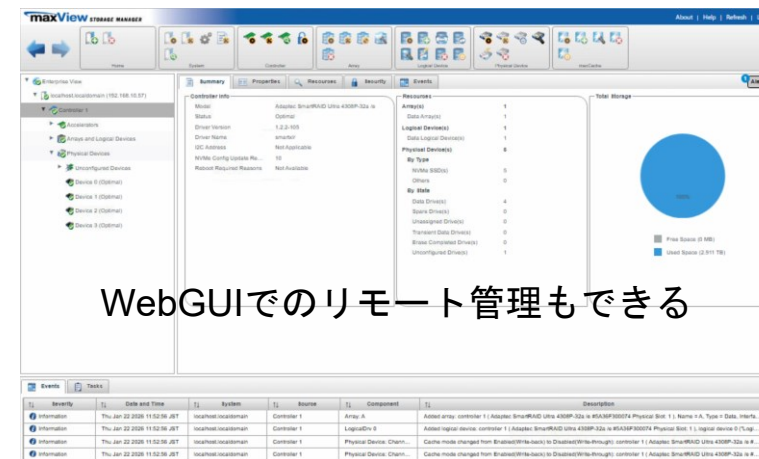
面白い特徴があります！

1. Arrays グループにアサインされているSSDの本数で構成可能なRAIDの種類を選んでLUを作成できる
※LUを作るSSDの本数に変更できない

RAIDグループ



Arraysグループ



WebGUIでのリモート管理もできる

おまけ・・・NVMe-oF デモ

<https://www.servants.co.jp/technology/column/technology/tec-nvidia/7116/>

SuperMicroサーバ PCIe-Gen4x16 (最大256Gbps)

NVIDIA/Mellanox CX-6(200G NIC)
Ubuntu20.04.3 OFEDドライバ

NVMe-oF (RoCEv2)

iSCSI

192.168.200.58

192.168.201.58

200Gbps

200Gbps

NVIDIA/Mellanox SN3700 (200GbpsSwitch)

200Gbps

200Gbps

192.168.200.222

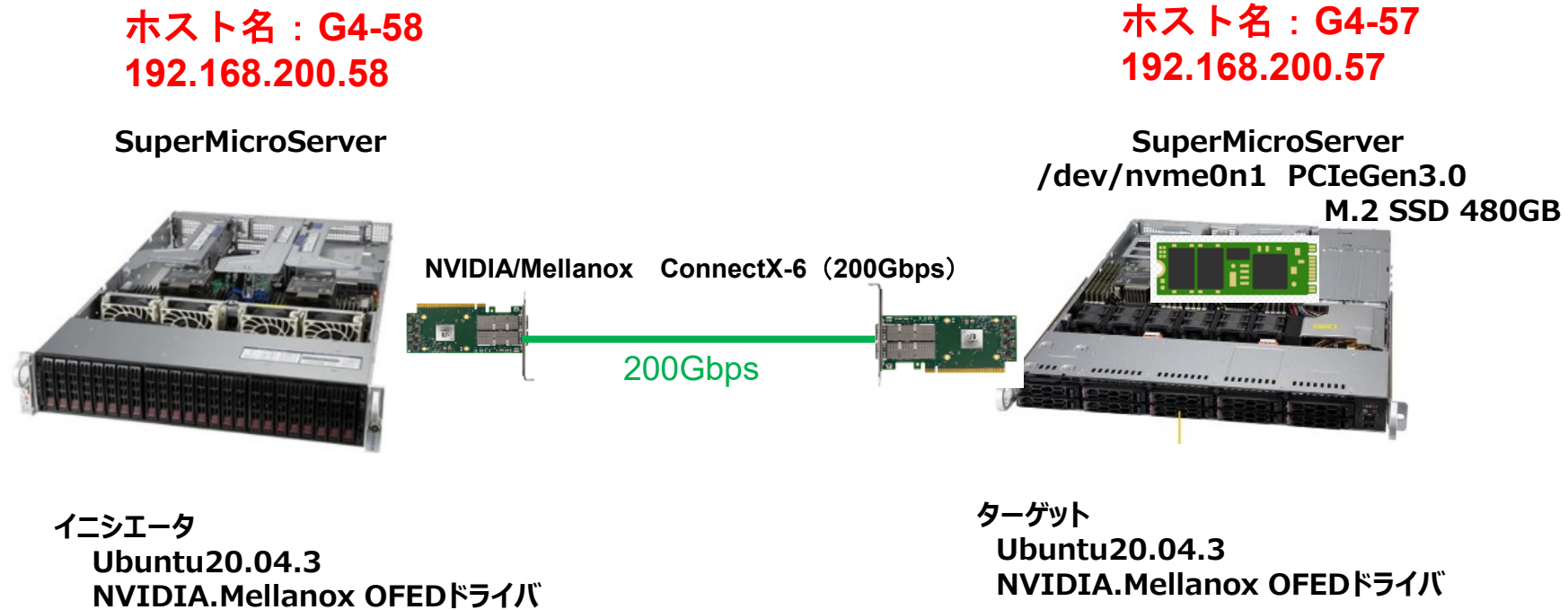
192.168.201.222

SANBlaze ストレージエミュレータ
PCIe-Gen3.0 x16のため
最大128Gbps

NVIDIA/Mellanox CX-6(200G NIC)



NVMe-oF Targetモード デモ



<https://community.mellanox.com/s/article/simple-nvme-of-target-offload-benchmark>

詳細は弊社サイトをご覧ください

<https://www.servants.co.jp/technology/column/technology/tec-nvidia/7116/>



ホーム

メーカーライティング

製品ライティング

新着情報

企業情報

技術情報

お問い合わせ

技術情報

技術コラム

ホーム・技術情報・ダウンロード・技術コラム・弊社社内でのNVMe関連の技術検証

2021.12.21 編集済 NVIDIA

弊社社内でのNVMe関連の技術検証

1. 社内検証 「NVIDIA ConnectX-6 (Ubuntu20.04.3)でのストレージサーバ設定方法」
NVIDIA/Mellanox ConnectX-6のOFEDドライバを使用して、NVMe-oFのストレージサーバの構築方法を検証しました。

カテゴリ

製品技術 (37)

NVIDIA (26)

ATTO (8)

Qlogic (1)

VMware (1)

最新技術動向 (6)

テクニカルラボ・ニュース (4)

最新記事

お手持ちFibre Channel対応ストレージ

NVIDIAスイッチを物理的に接続した
くも糸に、Only OSアップグレード

Benchmark 1 より高速な動作が実現
した Part.2 最終データ動作編

Benchmark 1 より高速な動作が実現
した Part.1 導入編

弊社社内でのNVMe関連の技術検証

2. 社内検証 「NVIDIA/Mellanox ConnectX-6でのNVMe-oFのパフォーマンス検証」
NVIDIA/Mellanox ConnectX-6でNVMe-oFにてSANBLAZEストレージエミュレータへ接続しパフォーマンスを計測してみました。

4-5. ストレージネットワークの進化

4-5-7 Ultra Ethernet

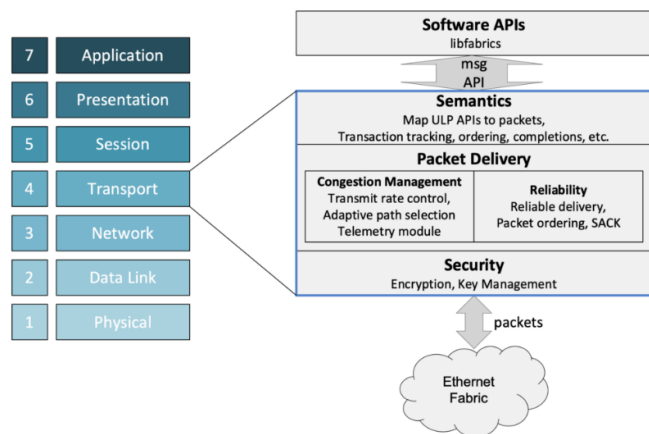


IEEE P802.3dh

所見

まだコミュニティ（策定化前）ではあるが、InfiniBandをお手本としてEthernetの上にトランスポート層までを作りこんで、通信品質（輻輳管理）を担保したり、改良版RDMAを実装したりする。

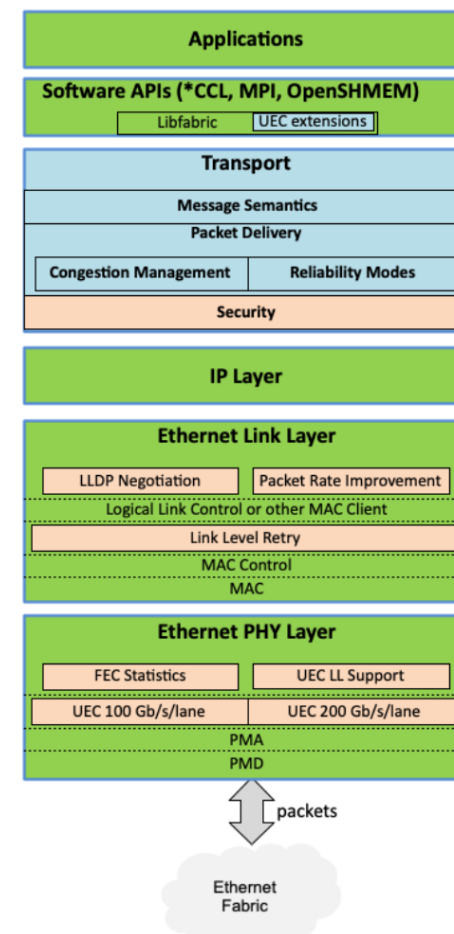
AI、HPCがターゲットとなる モロにIBの対抗馬！！



Ultra Ethernet™
Specification v1.0.2

January 28, 2026

UEC Stack



UECより引用：<https://ultraethernet.org/uec-progresses-towards-v1-0-set-of-specifications/>

4-5 進化を続けるストレージインターフェース技術

4-5-8 PCI Expressバスを活用した高速インターコネクト技術

CXL(Compute eXpress Link)

業界の主要な企業が標準化を進めている、デバイス（CPU,GPU、およびDRAM間）などを接続する新たなインターコネクトの規格 Open industry standard for high bandwidth, low-latency interconnect

- ・現在主流のインターコネクトであるPCI express(以後PCIe) の仕様を流用 (※PCIe Gen.5以降が条件)
- ・CPU,GPU, Smart NIC, FPGA, メモリ(揮発性、不揮発性), ストレージ のようなデバイスの接続でメリット がある
- ・最新仕様はv4.0で、2025/12/4に仕様が公開

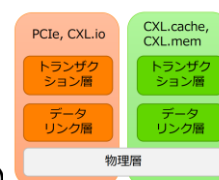


Gen 5.0以降のPCIeを流用

接続(物理層)はPCIeそのままにしつつ、その上位レイヤ をCXL独自のプロトコルを追加することで実現
※ PCIe はGen 5.0以降は、PCIeのバス上で異なるプロトコルの 通信を許容する

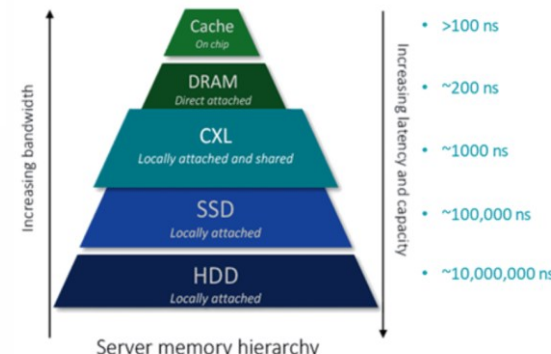
CXLは以下3種類のプロトコルを組み合わせて動作

- CXL.io : CXLデバイスの検知 : PCIeベースのエラー報告 などに使用 (PCIeのプロトコルをそのまま流用)
- CXL.cache : デバイス (アクセラレータ) とCPU間で 主にキャッシュ情報を通信・要求するために利用
- CXL.mem : CPUとデバイス (アクセラレータ) 間で メモリアクセス要求をする際に使用

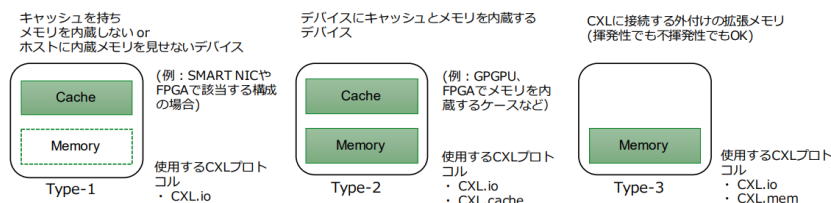


PCIe Gen 3.0	PCIe Gen 4.0	PCIe Gen 5.0	PCIe Gen 6.0
• Maximum Speed: 16 GT/s	• Maximum Speed: 32 GT/s	• Maximum Speed: 64 GT/s	• Maximum Speed: 128 GT/s
• Maximum Bandwidth: 16 GB/s	• Maximum Bandwidth: 32 GB/s	• Maximum Bandwidth: 64 GB/s	• Maximum Bandwidth: 128 GB/s
• Maximum Latency: 100 ns	• Maximum Latency: 100 ns	• Maximum Latency: 100 ns	• Maximum Latency: 100 ns
• Maximum Power: 10 W	• Maximum Power: 10 W	• Maximum Power: 10 W	• Maximum Power: 10 W

(*)アクセラレータ: SmartNICや Computational Storageのように、CPUの処理の一部を代わりに処理できるデバイス



3つのTypeのデバイスを定義

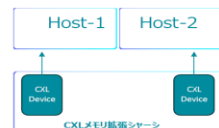


単純メモリー拡張



- ✓ サーバ単体でのメモリー拡張
- ✓ 筐体内蔵
- ✓ 実装容量に上限あり

メモリー・プール



- ✓ 複数ホストでメモリーチャックとして共有
- ✓ CXLメモリー搭載専用筐体を導入
- ✓ サーバ毎で動的プロビジョニング

メモリー・シェア



- ✓ メモリーデバイスの同時アクセス共有
- ✓ アプリケーション毎のインターリーブ可能
- ✓ 秀でた性能、先進機能、TCO削減

CXLの主要な特徴

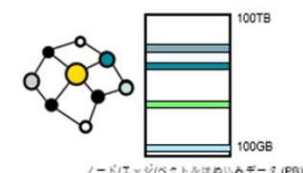
- ・業界での広範な採用
- ・ディ disaggregation
- ・広帯域
- ・低レイテンシー
- ・動的メモリープロビジョニング
- ・メモリープールおよび共有
- ・キャッシュコヒーレンシーサポート
- ・スケーラビリティの改善
- ・低消費電力
- ・TCOの低減
- ・将来的な有効活用を担保

★CXLの活用事例として

RAG (Retrieval-Augmented Generation) にて

グラフ・ベクトル処理の時間がNVMeSSDの8時間からCXLで20分に短縮

グラフ分析とニューラルネットワーク



- ・メモリー拡張: 従来メモリスロットのみだったメモリー実装にPCIe経由での拡張が可能となる
- ・メモリープール: サーバ外部にメモリープールを用意して複数のホストで共有することが可能となる
- ・アプリケーション性能の改善: 搭載可能なメモリー容量の拡張により処理性能の向上が可能となる
- ・メモリー使用率の改善: 個々のサーバにおけるメモリー容量を柔軟かつ動的に最適化できるようになる
- ・ソフトウェアライセンス費用の低減: メモリーのプール化によってサーバ台数の低減が可能となる
- ・VM密度の向上: サーバ当たりのVM (Virtual Machine) 密度の向上が可能となる

4-5 進化を続けるストレージインターフェース技術

4-5-8 PCI Expressバスを活用した高速インターコネクト技術

CXL(Compute eXpress Link)

業界の主要な企業が標準化を進めている、デバイス（CPU,GPU、およびDRAM間）などを接続する新たなインターコネクトの規格 Open industry standard for high bandwidth, low-latency interconnect

- ・現在主流のインターコネクトであるPCI express(以後PCIe) の仕様を流用 (※PCIe Gen.5以降が条件)
- ・CPU,GPU, Smart NIC, FPGA, メモリ(揮発性、不揮発性), ストレージ のようなデバイスの接続でメリット がある
- ・最新仕様はv4.0で、2025/12/4に仕様が公開

Gen 5.0以降のPCIeを流用

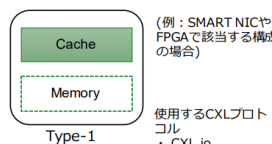
接続(物理層)はPCIeそのままにしつつ、その上位レイヤを独自のプロトコルを追加する
※ PCIe はGen 5.0以降は、PCIeのバス上で異なるプロトコル通信を許容する

CXLは以下3種類のプロトコルを組み合わせて動作

- CXL.io : CXLデバイスの検知 : PCIeベースのエラー
- CXL.cache : デバイス (アクセラレータ) とCPU間で
- CXL.mem : CPUと

3つのTypeのデバイスを定義

キャッシュを持ち
メモリを内蔵しないor
ホストに内蔵メモリを見せないデバイス

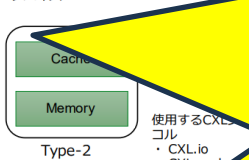


単純メモリー拡張

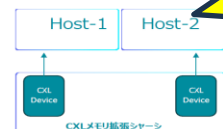


- ✓ サーバ単体でのメモリー拡張
- ✓ 筐体内蔵
- ✓ 実装容量に上限あり

デバイスにキャッシュとメモリを内蔵する
デバイス

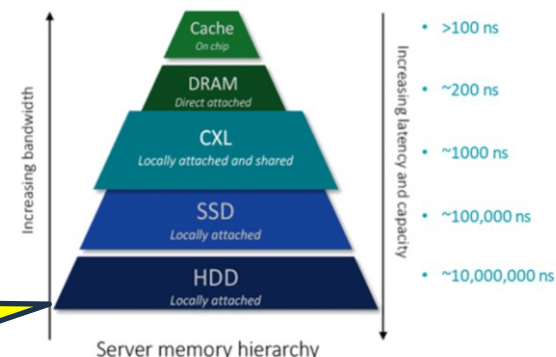


メモリー・プール



- ✓ 複数ホストでメモリーチャックとして共有
- ✓ CXLメモリー搭載専用筐体を導入
- ✓ サーバ毎で動的プロビジョニング

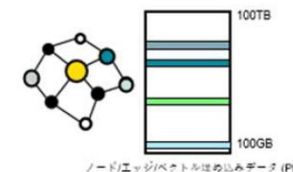
この技術も同様にネットワークへ
CXL over Fabric 策定が始まっています



活用事例として

GenAI-Augmented Generation) にて
AIモデル処理の時間がNVMeSSDの8時間からCXLで20分に短縮

グラフ分析とニューラルネットワーク



- ・従来のメモリー実装に比べて、ホストのみだったメモリー実装にPCIe経由での拡張が可能となる
- ・メモリー拡張: サーバ外部にメモリープールを用意して複数のホストで共有することが可能となる
- ・アプリケーション性能の改善: 搭載可能なメモリー容量の拡張により処理性能の向上が可能となる
- ・メモリー使用率の改善: 個々のサーバにおけるメモリー容量を柔軟かつ動的に最適化できるようになる
- ・ソフトウェアライセンス費用の低減: メモリーのプール化によってサーバ台数の低減が可能となる
- ・VM密度の向上: サーバ当たりのVM (Virtual Machine) 密度の向上が可能となる

4-5. ストレージネットワークの進化

4-5-9 PCI Expressでのファブリック CDI

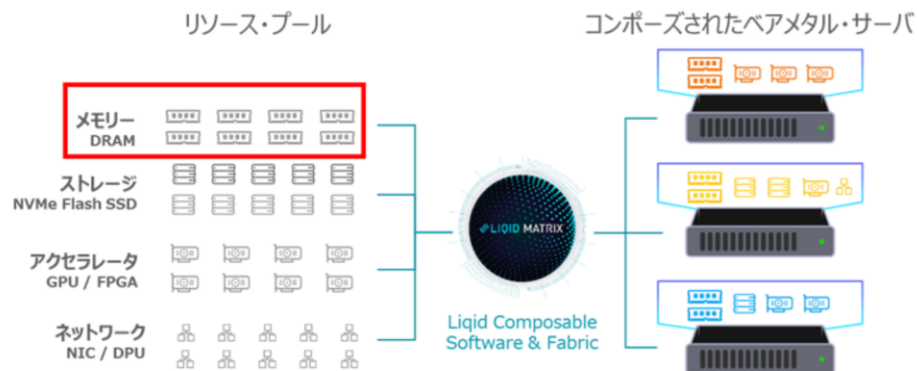
コンポーザブル・ディスアグリ・インフラストラクチャ

従来のサーバ



- ! 低いデバイス密度
- ! 高額なTCO
- ! 柔軟性に制限

動的構成変更が可能なサーバ



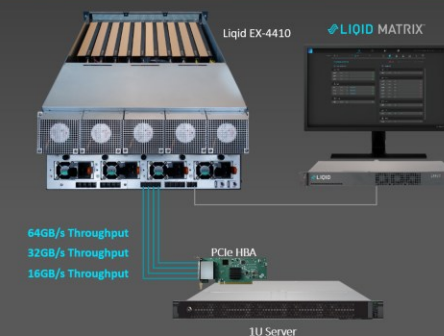
- ✓ 最高のAI性能
- ✓ 効率を最適化
- ✓ 柔軟性を最大化

2025年にCXL2.0もサポートしました！ ※ただしFabricとしては別になります

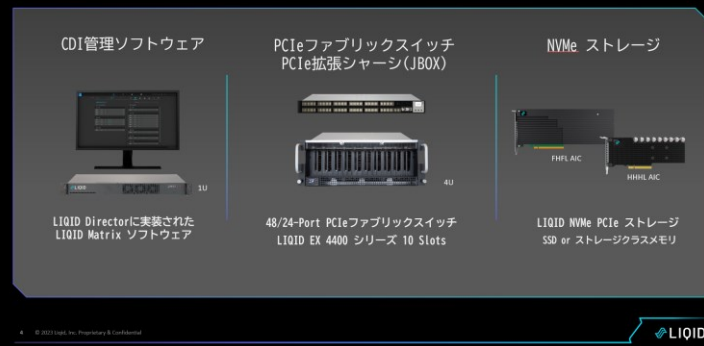
これも**PCI Expressバスを外部へ拡張した**
ストレージネットワーク
とも考えられるかな？

LIQID CDI の PCIe接続&ソフトウェア

- LIQID MATRIXで即座にGPUを動的に割当て/解放が可能
- PCIe拡張シャーシ(JBOX)に実装されたGPU 20基まで各サーバーに接続可
- GPUの使用率90%以上を実現



LIQID ポートフォリオ



詳細は <https://www.liqid.com/> もしくは、

★現在の主な活用方法としてはGPU/NVMeストレージの動的有効活用

サーヴァンツインターナショナル 技術コラムで紹介してます
<https://www.servants.co.jp/technology/column/technology/tec-liqid/9034/>

4-5. ストレージネットワークの進化

4-5-9 PCI Expressでのファブリック CDI

LIQID CDI 概念図

LIQID Matrix (CDI Command Center User Interface)

PCIe接続のネットワークを管理して、
PCIe Expansion Chassisにて登録したリソース (GPUやNVMeストレージなど) を、Fabric内のサーバにオンデマンドに配置して使うことができる

※動的に構成変更が可能である！



LIQID Director

● ディスアグリゲータードリソース ● Liquidファブリックへの接続 ● 即座に導入・拡張

プール化されたリソース

ホスト

Intel / AMD / ARM

ストレージ

NVMe Flash / Optane

アクセラレーター

GPU / FPGA / DPU

ネットワーク

NIC



動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

動的に構成されたサーバー

PCIe

Fabric Switch(PCIe Switch)

PCIe HBA

CXL HBA

PCIe

PCIe

PCIe

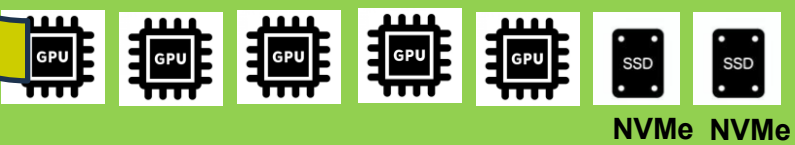
PCIe Fabric
PCI Expressバスのネ

2025年にCXL2.0もサポートしました！
※ただしFabricとしては別になります

CXL Switch
(PCIe Switch)

PCIe

PCIe



NVMe NVMe

メモリー
DRAM



プール PCIe Expansion Chassis

おまけ AI・人工知能Expo2024 LIQID デモ



2024年冬に幕張メッセにて開催されたNexTexh AI・人工知能Expoの弊社ブースにて紹介したLIQIDのCDI（PCIexpressのファブリックでのGPUとSSDの動的アサイン）のデモを以下に掲載しています。

<https://www.servants.co.jp/technology/column/technology/tec-liquid/10850/>

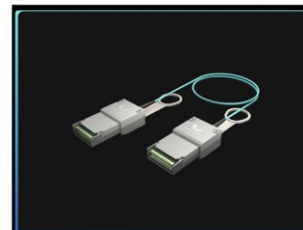


PCI-SIGが「PCIe Optical Workgroup」を結成しPCIeへの光通信技術の採用を検討

従来の銅線によるデータ転送速度が限界を迎えつつある今、そのボトルネックを解消すべく、PCI Express相互規格の策定団体**PCI-SIG**は「**PCIe Optical Workgroup**」の結成を発表し、光通信技術によるPCI-Express相互接続規格の進化の可能性を探ることを決定した。

PCI-SIGとしてはPCIe-Gen7での策定を目指しているが、それに先駆けて、ケーブルベンダーからAOC (Active Optical Cable)による光へ変換するケーブル製品が発表されている

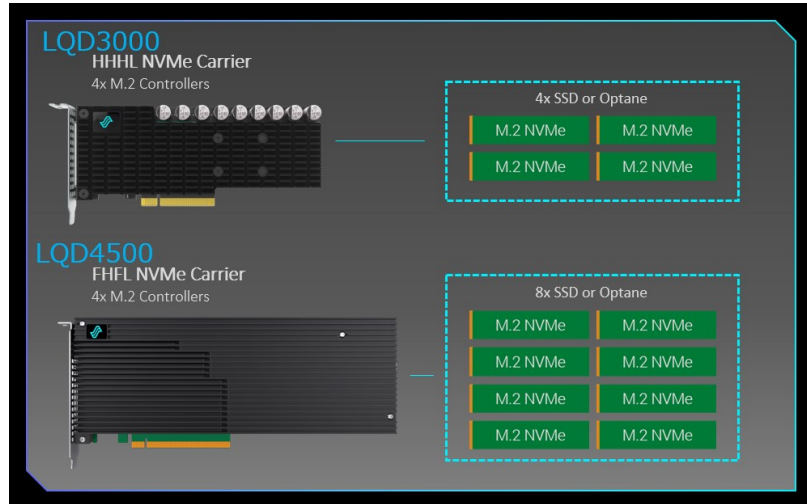
AOC – CDFP x16



Active Optical Cables	
CXL Support	2.0
Latency	<1ns + TDP
Power Consumption	15W
PCIe Generation	5.0
Speed	32Gbps x16
BER	1e-12
Interface	CDFP
Fiber Type	OM3 Multi-mode Fiber
Bifurcation Support	Single x16, Dual x8 and Quad x4

Transfer Rate (*)	PCIe 7.0: 128Gbit/s PCIe 6.0: 64Gbit/s PCIe 5.0: 32Gbit/s + 16Gbit/s (PCIe 4.0) + 8Gbit/s (PCIe 3.0)
Lane (**)	16 (8, 4, 1)
Length (Max)	30m
Modulation	NRZ or PAM-4
Link Scheme [8]	Linear Amplified or Half Retimed
Clock Scheme	Reference Clock or No clock
Diagnostic	PCIe Sideband or No diagnostic

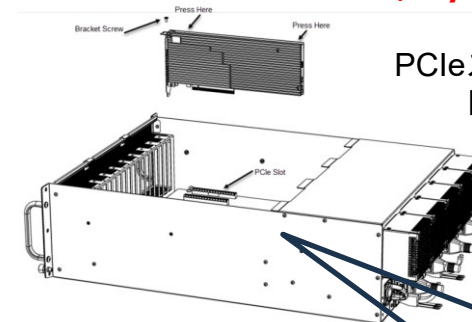
おまけ LIQIDの隠れたヒット商品



- Liquid LQD/AXシリーズは大容量NVMEストレージ製品で、単一デバイスでは世界最高のパフォーマンスを実現！！
SW RAIDで、8つのNVMeSSDを束ねると・・・

爆速！ 48GB/秒、600万IOPS以上

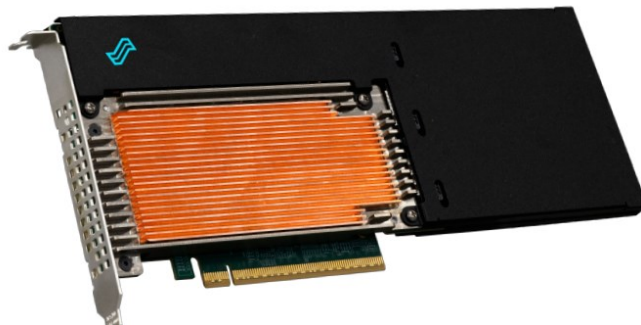
↑ Byteです！



PCIeスロットに差すだけ！
ドライバのインストールや設定は不要です！！

Model	LIQID IO Accelerator AX-5500
Raw Capacity	Up to 64TB
NAND Type	TLC 3D NAND
Read Bandwidth	Up to 48GB/s Throughput
Write Bandwidth	Up to 30GB/s Throughput
Random 4K Read IOPS	Up to 6M IOPS
Random 4K Write IOPS	Up to 1.5M IOPS
Min Read Latency	80µs
Min Write Latency	15µs
Protocol	NVMe 1.4
Bus Interface	PCI Express 5.0 x16
Endurance	Up to 70 PBW / 3 DWPD*
Security	256-bit AES Data Encryption
Weight	20 oz
Warranty	3 years or maximum endurance used
Form Factor	FHFL Single Width
Operating Temperature	Op: 0 to 55 deg C / Non-Op: -40 to 75 deg C
Power	Active: 120W Max
Airflow	Min: 400 LFM
Humidity	5% to 95% (non-condensing)
Altitude	0 ft to 10,000 ft
Operating Systems	Windows, RHEL, SLES, CentOS, Solaris, SUSE, VMware
Agency & Safety	UL, CB, CE, KCC, BSM, VCCI, FCC Part 15, Subpart B and CISPR 32 for a Class A

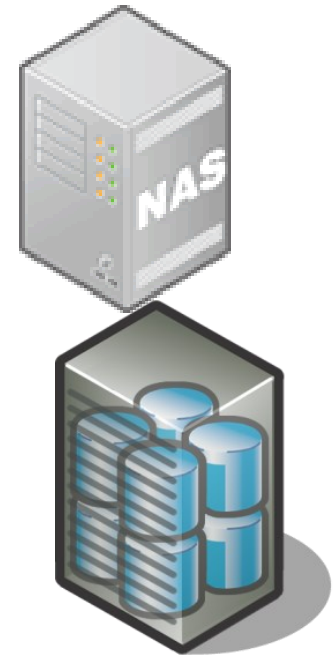
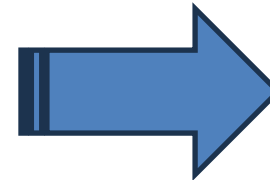
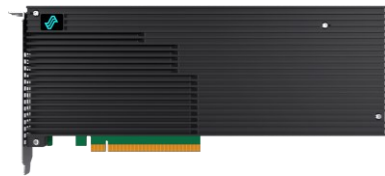
PCIeGen5 AX-5500



- > システム デバイス
- > ディスク ドライブ
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - SAMSUNG MZ1LB960HAJQ-00007
 - WDC WD2003FYYS-02W0B1
- > ディスプレイ アダプター

数分でNASやストレージサーバが構築できますよ！

- LQD4500/AX-5500 + SmartRAID4308/XDP をPCIeスロットに挿してドライバをインストールすれば簡単にNASやストレージサーバの構築が可能！！



 Windows Server 2022



技術的には2Gbpsの時代ですが、基本的な技術は変わってません



個数：

 カートに入れる

 ほしいもののリストに登録する

 店舗の在庫を調べる

ファイバチャネル技術解説書〈2〉

JDSF/ファイバチャネル技術部会【監修・著】

著者名をお気に入りに登録する

価格 **¥9,680** (本体¥8,800)

論創社 (2003/10発売)

ポイント 88pt

提携先に在庫がございます【僅少】

通常、5～7日程度で出荷されます。

※納期遅延や、在庫切れで解約させていただく場合もございます。
(※複数冊ご注文はお取り寄せとなります)

出荷予定日とご注意事項

※上記を必ずご確認ください

●店舗受取サービス(送料無料)もご利用いただけます。

ご注文ステップ「お届け先情報設定」にてお受け取り店をご指定ください。尚、受取店舗限定の特典はお付けできません。[詳細はこちら](#)

サイズ A5判/ページ数 569p/高さ 22cm

商品コード 9784846005429

NDC分類 547.48

Cコード C2055

HBAとFCIPの章は、私が執筆してます・・・・

ご参考 数々の最新情報や検証結果を綴ってます！



その他随時、弊社サイトの**技術コラム**にて検証内容等をアップデートしております。

<https://www.servants.co.jp/technology/>

- LIQID CDI(コンポーザブル・ディスアグリ・インフラストラクチャー)について
- LIQID CXLソリューション
- InfiniBand
- SR-IOV
- NVMe-oF
- Windows10でのRDMA(SMB-Direct)
- こんな時どうする(Nvidia/Mellanox製品のトラブルシューティング) ←ストレージにも役に立つ情報もあります
- Cumulusスイッチの使い方
- LIQID Honey Badger(LQD4500)
- ATTO 製品の紹介 ATTO360Networking、ATTO360Storage、MultiPathDirector(パス管理ソフトウェア)などの無償ツールの検証

Qiita

にも色々な情報を掲載してます！ <https://qiita.com/organizations/servants>