GMO GPUクラウド

のシステム構築事例紹介　&
事例をもとにGPUネットワークのトレンドを
NVIDIAさんに聞いてみよう

GMO INTERNET

NVIDIA

登壇者

GMOインターネット　友 源輝

NVIDIA　木戸 大二郎

アジェンダ

GMO GPUクラウド

1. 登壇者自己紹介
2. GMO GPUクラウドのシステム構築事例紹介
   2-1. GMOインターネットの紹介
   2-2. システム構成
   2-3. ネットワーク構成
3. NVIDIAさんに聞いてみよう
   3-1. NVIDIAの紹介
   3-2. NVIDIAさんに質問をしてみよう
   3-3. 会場を交えての議論

自己紹介

GMO GPUクラウド

友 源輝　Genki Tomo

所属：ネットワーク技術チーム

出身：大分県

趣味：アウトドア、釣り🐟、狩猟採取🦝🍄

2007年 上京してネットワークエンジニアになる

2007年～2017年 NW運用、SIerで中小の構築に携わる

2018年 GMOインターネット入社

ConoHaなどクラウド商材のNWインフラ業務に携わる

NWインフラの設計・構築・運用まで一通り担当

2024年 GPUクラウドプロジェクトにジョイン、GPUクラウド環境のNW設計・構築を担当

自己紹介

GMO GPUクラウド

木戸 大二郎

NVIDIA 合同会社

出身：北海道 千歳市、苫小牧市

趣味：PCゲーム

現職：NVIDIA ネットワーキング担当 ソリューションアーキテクト

前職：Cisco 10年在籍 (WAN Routing/ Datacenter Switching/ Cloud Networking/ AI Networking)

前前職：Cyberagent でデータセンターネットワーク設計、構築、運用

それ以前：キャリア系ネットワーク設計、構築、運用を8年ほど

GMO GPUクラウド

GMO GPUクラウド システム構成

GMOインターネット 会社紹介

GMO GPUクラウド

GMO INTERNET

ConoHa by GMO

主要事業

インターネットインフラ事業

- ドメイン・レンタルサーバー（ホスティング）事業
- インターネット接続（プロバイダー）事業

ドメイン取るなら 公式登録サービス
お名前.com .com .jp .co.jp by GMO
https://www.onamae.com

とくとくBB by GMO

GMO GPUクラウド

- インターネット広告・メディア事業

AI指示文・プロンプトなら
教えて.AI by GMO

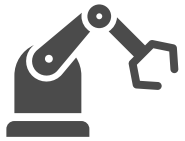
GMO GPUクラウドの紹介

GMO GPUクラウド

サービス名：GMO GPUクラウド
提供開始日：2024年11月22日

生成AIモデル学習や **機械学習**に最適化されたGPUクラウドサービス

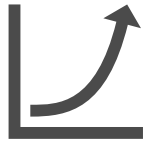
インフラチューニング不要な計算資源を提供



生成AIの急速な進化

さらなる需要拡大

▶



計算リソースの需要の

増加

▶

GMO GPUクラウド

日本のAI研究開発の加速を促し

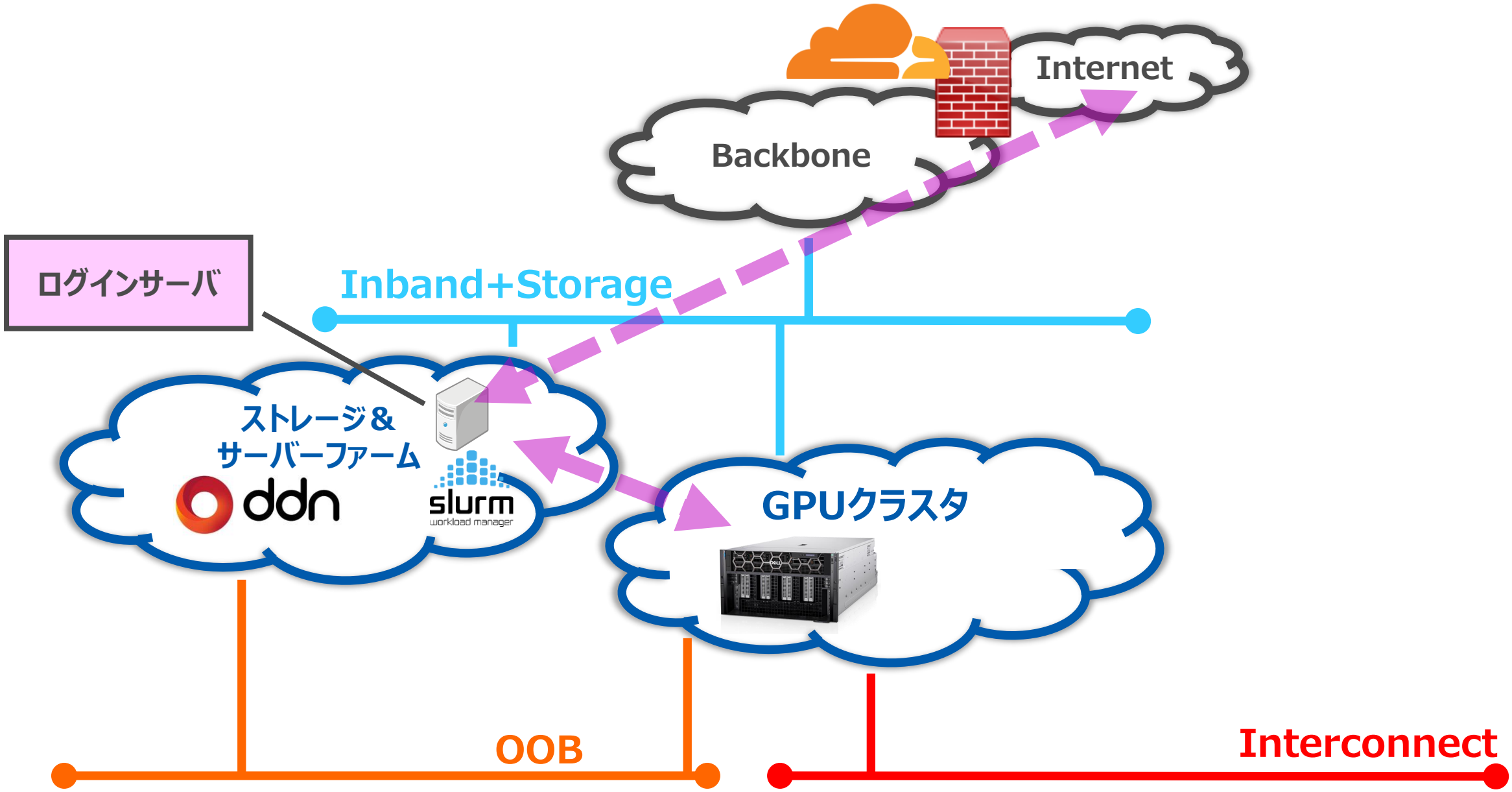
国際競争力を高めるために世界最高水準

の計算リソースの提供を目指す

システム構成

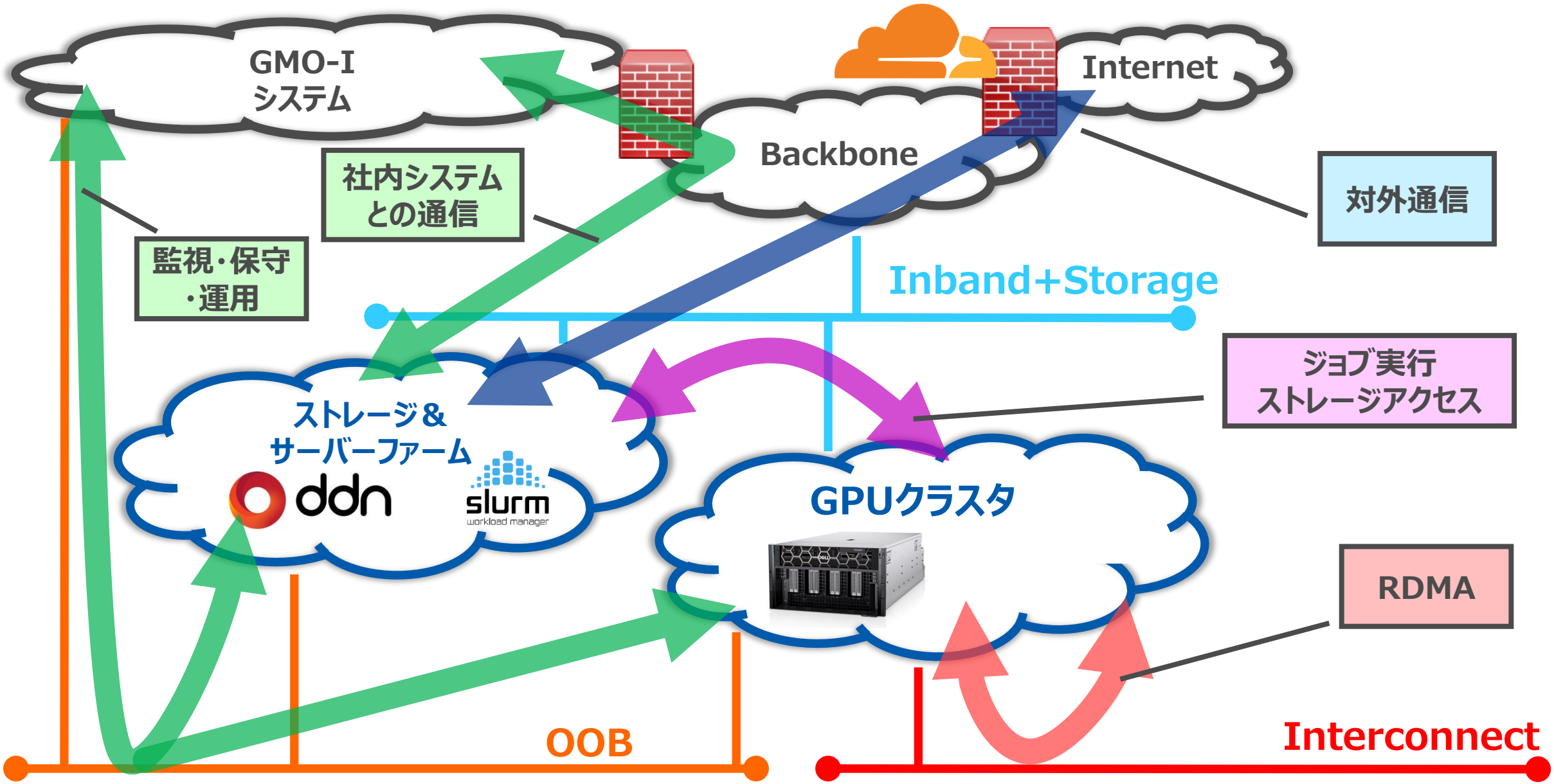
システム構成

GMO GPUクラウド



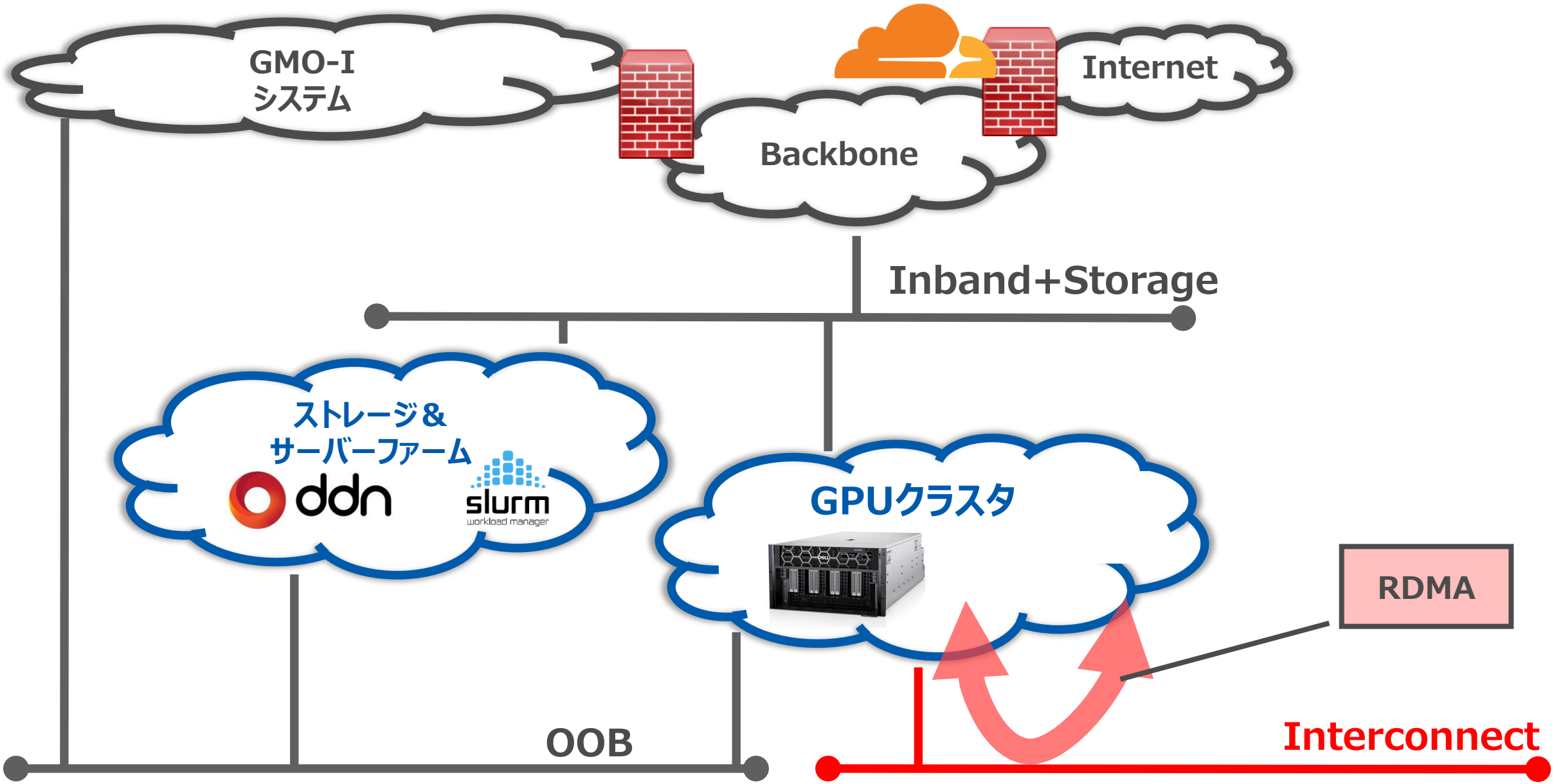
システム構成

GMO GPUクラウド



システム構成

GMO GPUクラウド



システム構成

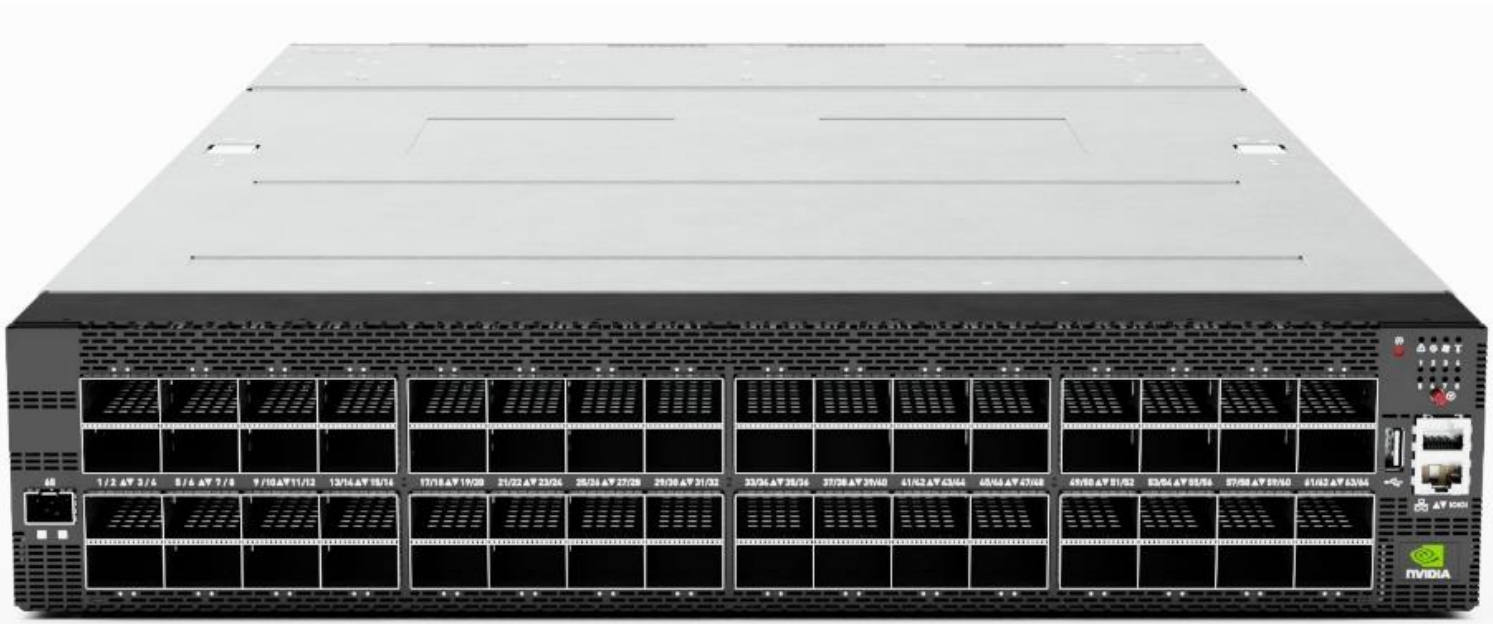
GMO GPUクラウド

GPUノード筐体は高密度GPUサーバ
DELL Power Edge XE9680 を採用。

NVIDIA H200 GPUを 8枚搭載
NVIDIA BlueField-3 SuperNICを 8枚 IC接続

GMO GPUクラウド

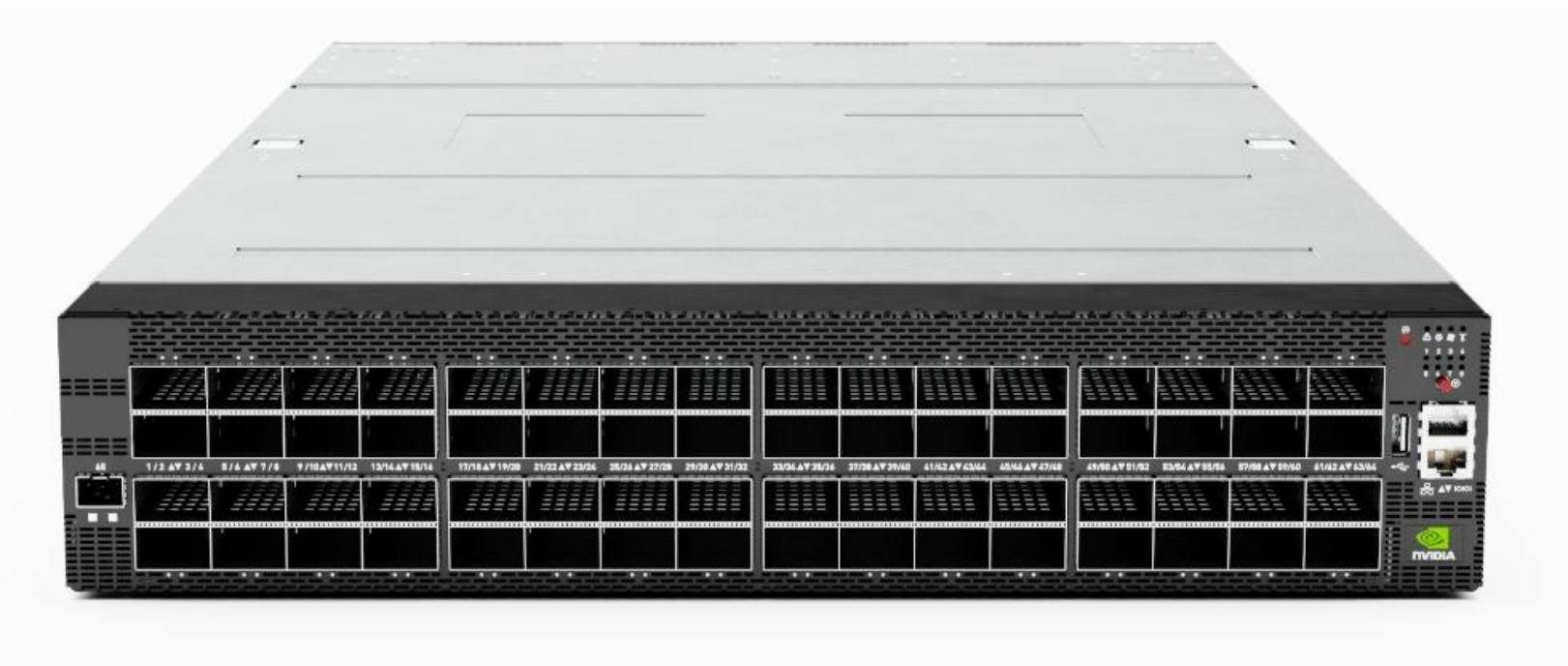
Interconnect-SWに、
NVIDIA Spectrum-X対応 SN5600を採用



Ethernet-SW
800G OSFP　×64

Q. イーサネットとインフィニバンド、どっち選べばいいの？

Interconnect-SWに、
NVIDIA Spectrum-X対応 SN5600を採用

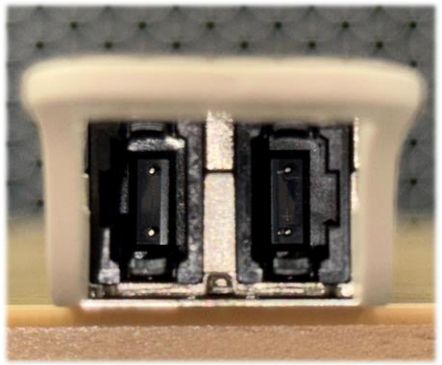


Ethernet-SW
800G OSFP ×64

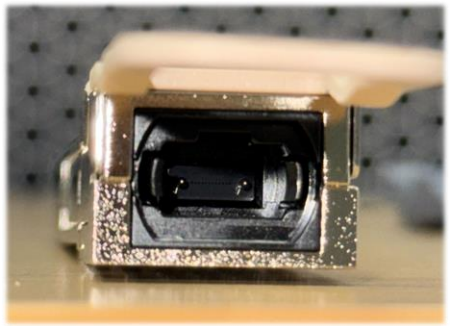
システム構成

GMO GPUクラウド

Interconnect-SW側：
OSFP **400G 2xSR4**



GPUノード側：
QSFP112 **400G-SR4**



ケーブル：
MPO12 マルチモード

システム構成

GMO GPUクラウド

NVIDIA Spectrum-X対応

- **RoCEv2**
  →RDMA通信のIBパケットをカプセル化
  　加えて、ECN・PFC などを利用し、輻輳制御

- **アダプティブルーティング**
  →インターフェイスごとの通信を監視し、動的にバランシング

- **NVIDIA独自の輻輳制御**
  →BlueField-3 SuperNICと連携し、エンドツーエンドの輻輳制御

Q. 輻輳制御って そんなに大事なの？

NVIDIA Spectrum-X対応

- **RoCEv2**
  →RDMA通信のIBパケットをカプセル化
  加えて、ECN・PFC などを利用し、輻輳制御

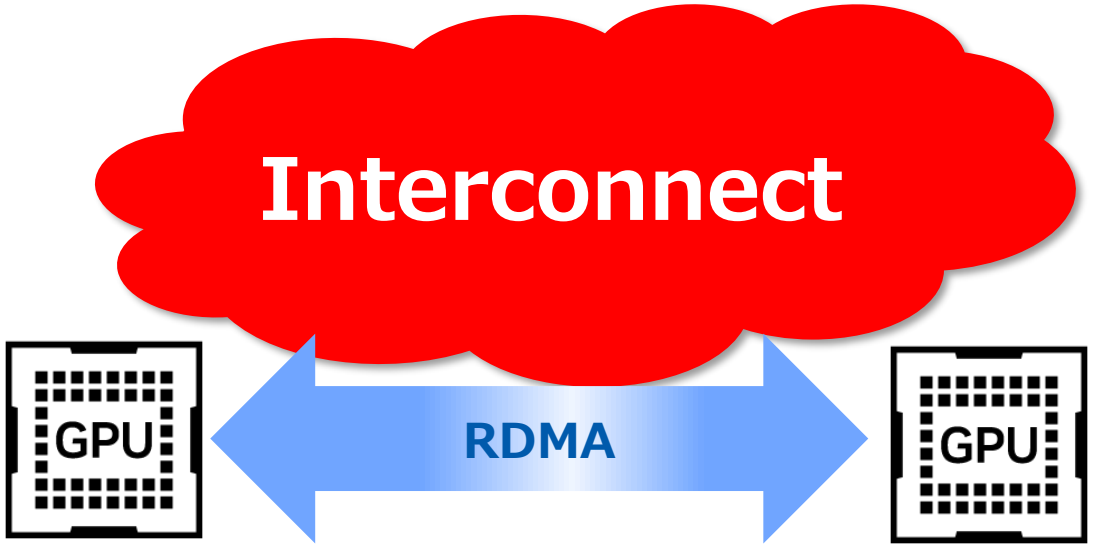
- **アダプティブルーティング**
  →インターフェイスごとの通信を監視し、動的にバランシング

- **NVIDIA独自の輻輳制御**
  →BlueField-3 SuperNICと連携し、エンドツーエンドの輻輳制御

インターコネクト ネットワーク 構成

GMO GPUクラウド

GPUネットワークの要、Interconnectネットワーク

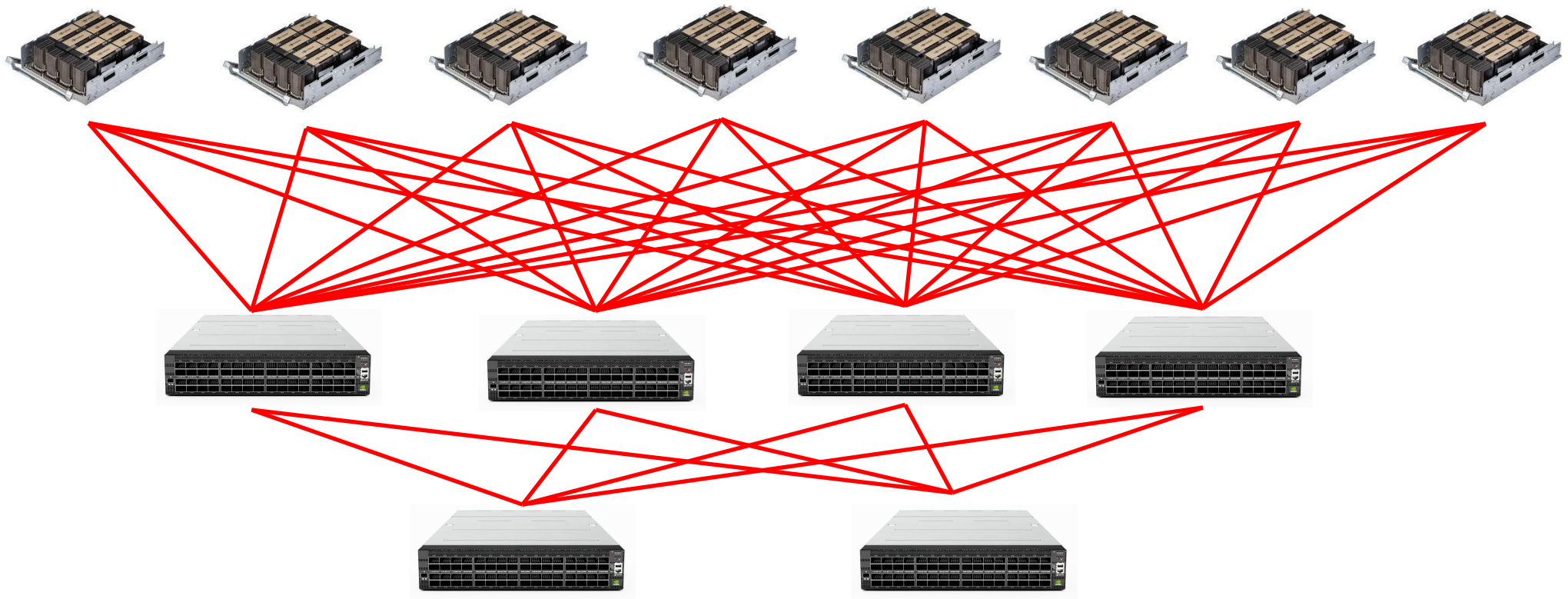


GPUノード間のRDMA
(Remote Direct Memory Access)
を行うネットワーク

パケットロスが計算の失敗に直結、
ロスレスNWを構築する必要がある。

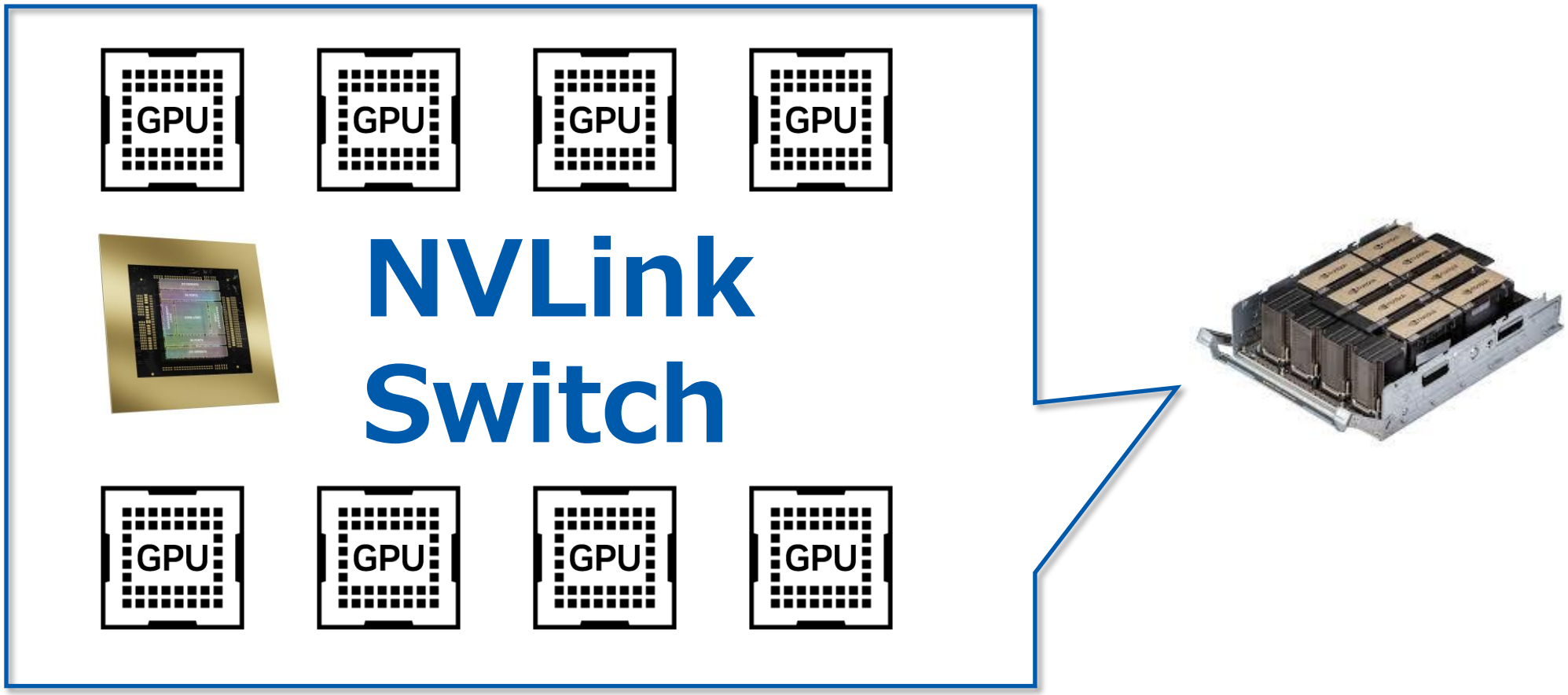
GMO GPUクラウド

ロスレスネットワークの実現のため、
トポロジには Rail optimized topology Full Bisectionを採用



GMO GPUクラウド

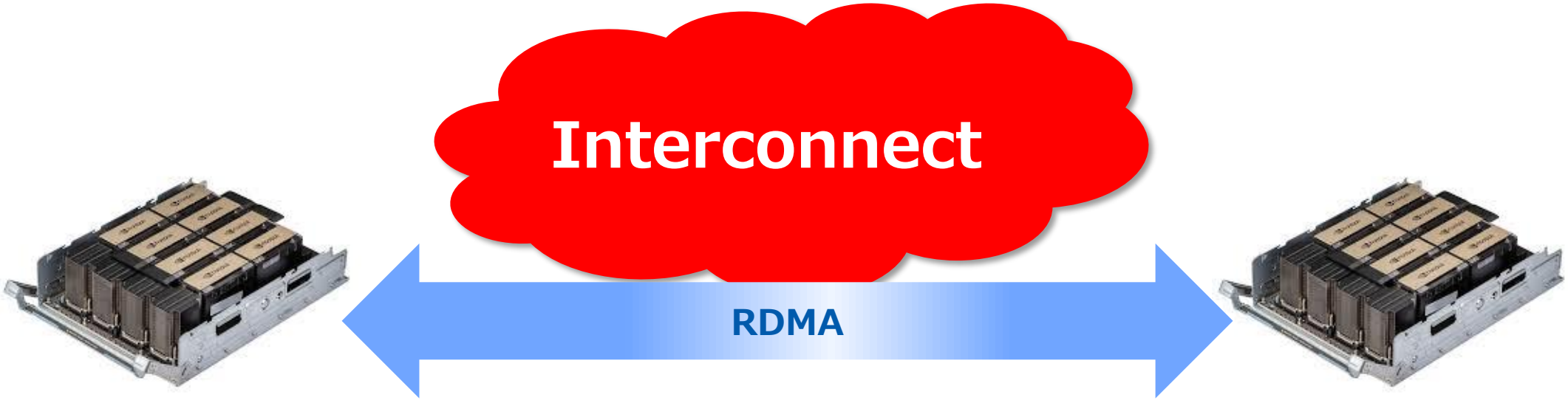
GPUノード内のGPU間通信は、サーバ内蔵のNVLink Switchで処理



ネットワーク構成

GMO GPUクラウド

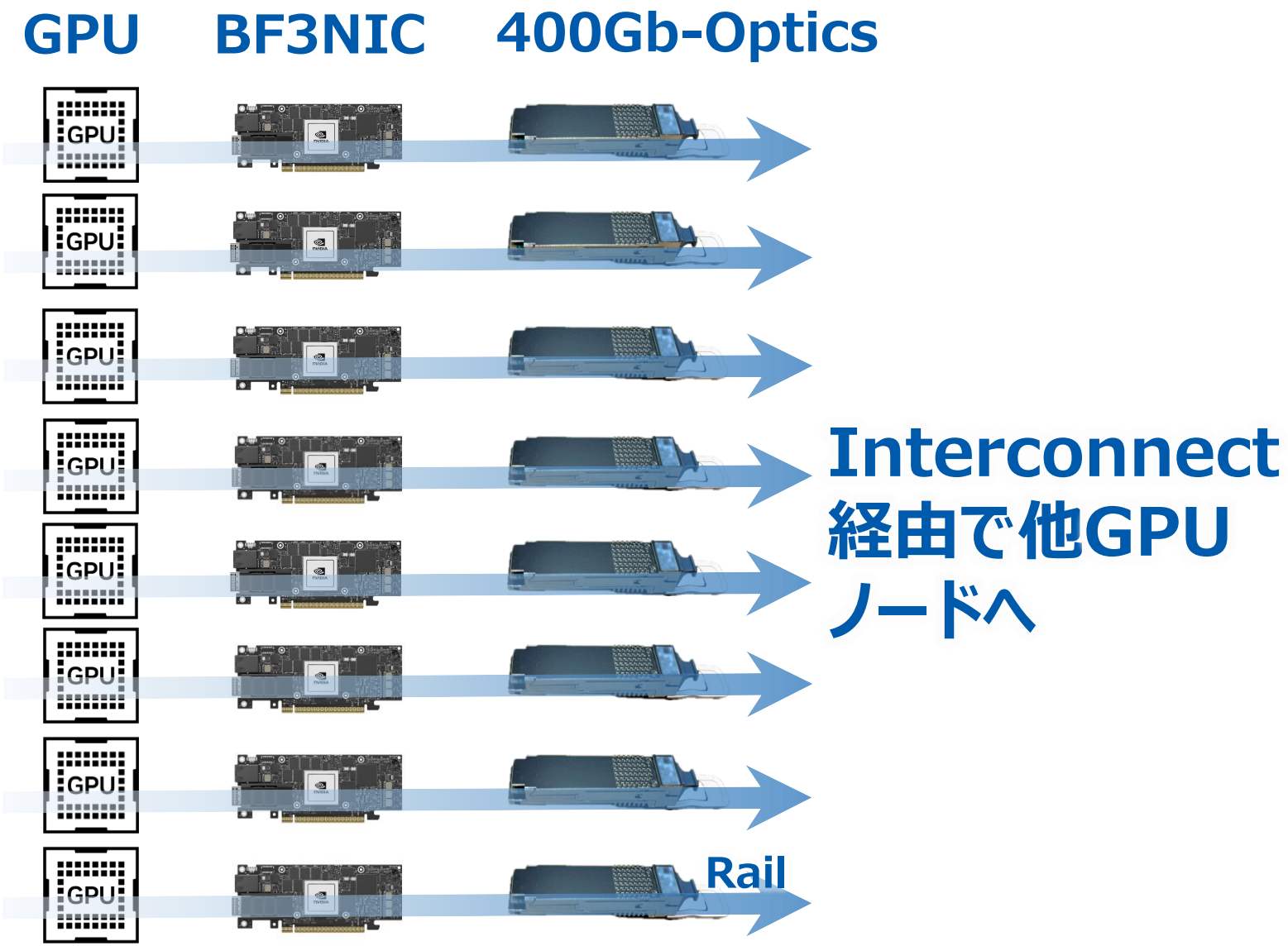
1ノードで処理しきれないほどの大きな計算は、
複数のGPUノードで連携して処理



ネットワーク構成

GMO GPUクラウド

GPUノード間通信には
GPU1枚につき
400Gb/s
の帯域を確保

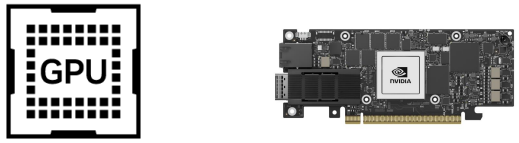


ネットワーク構成

GMO GPUクラウド

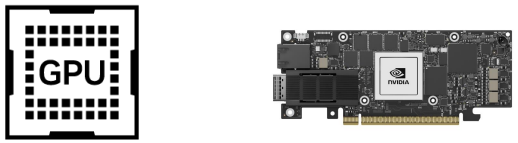
GPUノード#1

GPU#1



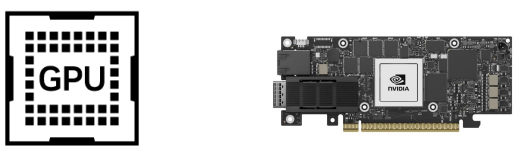
GPUノード#2

GPU#1



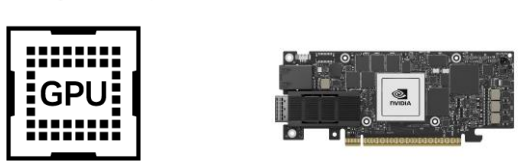
GPUノード#3

GPU#1



GPUノード#4

GPU#1

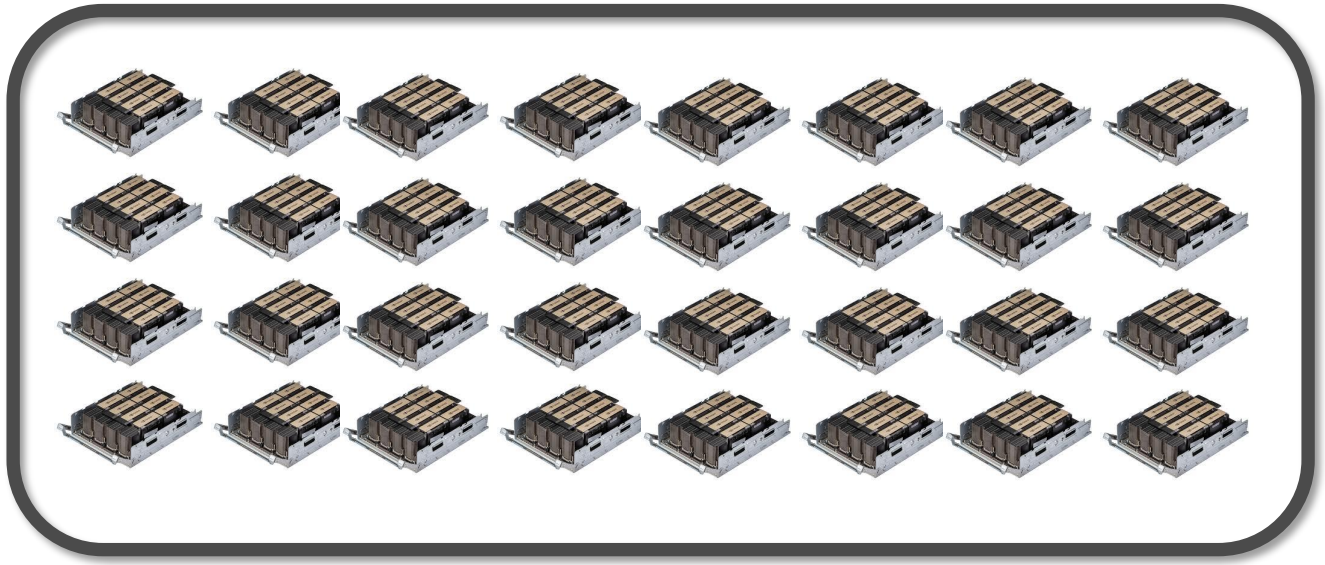


各GPUノードの
同じGPU番号を
同じアップリンクSWへ

Leaf-SW

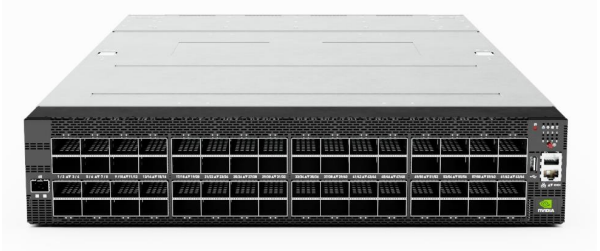
ネットワーク構成

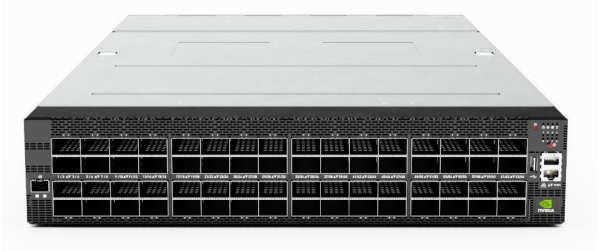
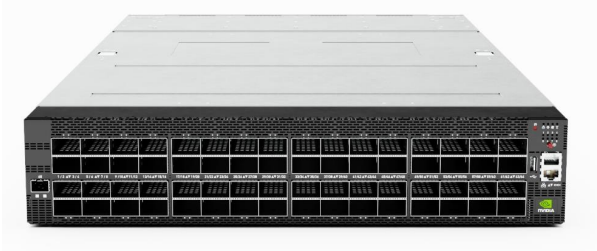
GMO GPUクラウド



GPUノード×32台
＝1SU(スケーラブルユニット)

アップリンクケーブルは
8x32=256本



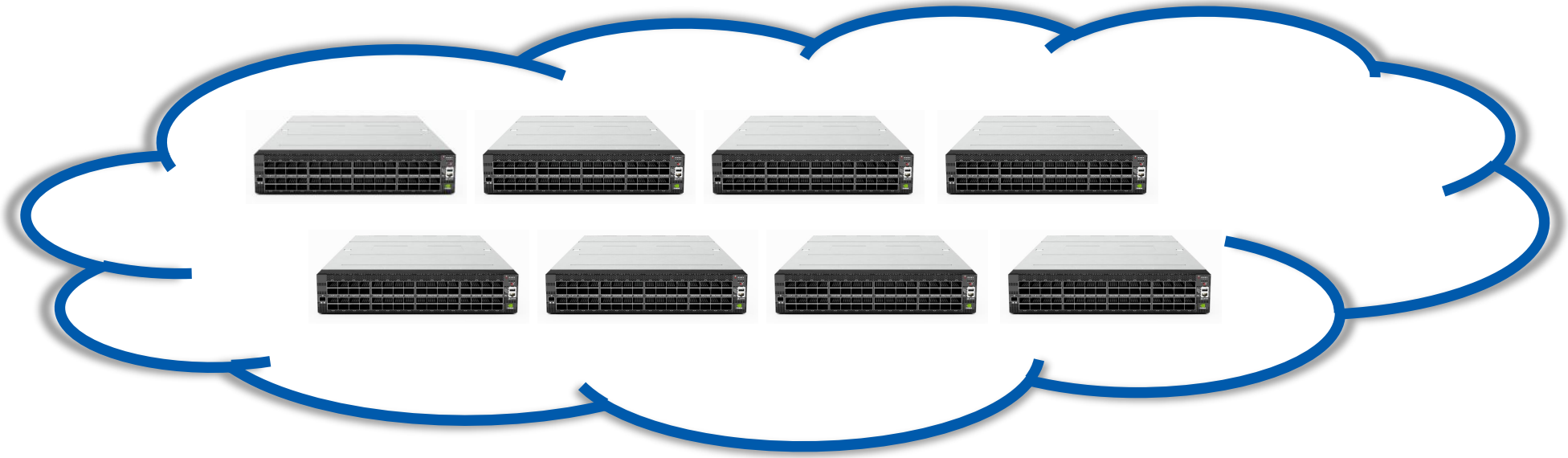



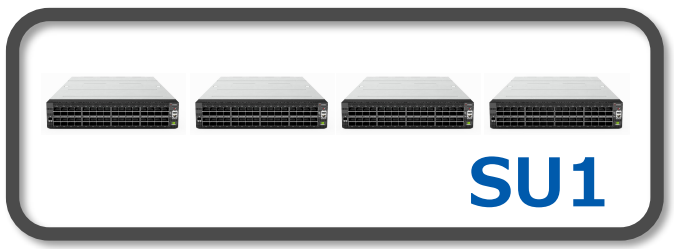
Leaf-SW NVIDIA SN5600 x4 で収容

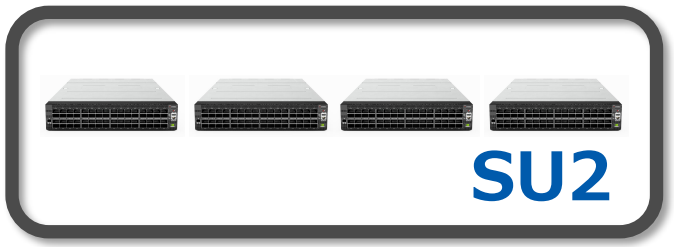
ネットワーク構成

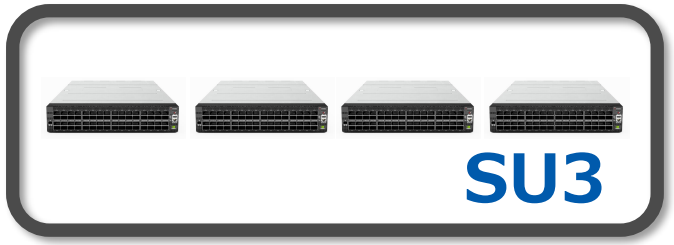
GMO GPUクラウド

SU(スケーラブルユニット)間の通信を処理する
Spine-SWは NVIDIA SN5600 x8 を配置、
4SU(leaf16台)を収容できる構成。





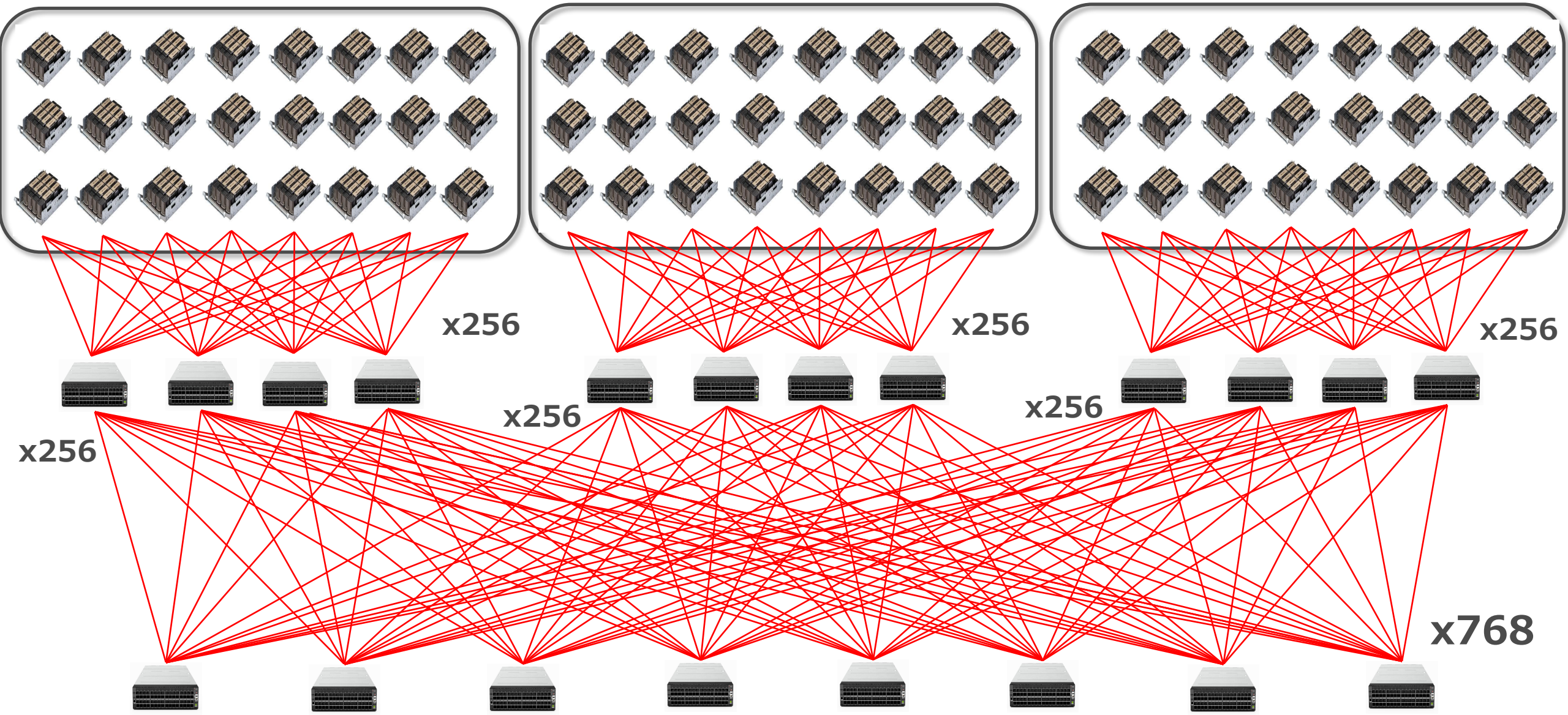




SU4
（予定）

ネットワーク構成

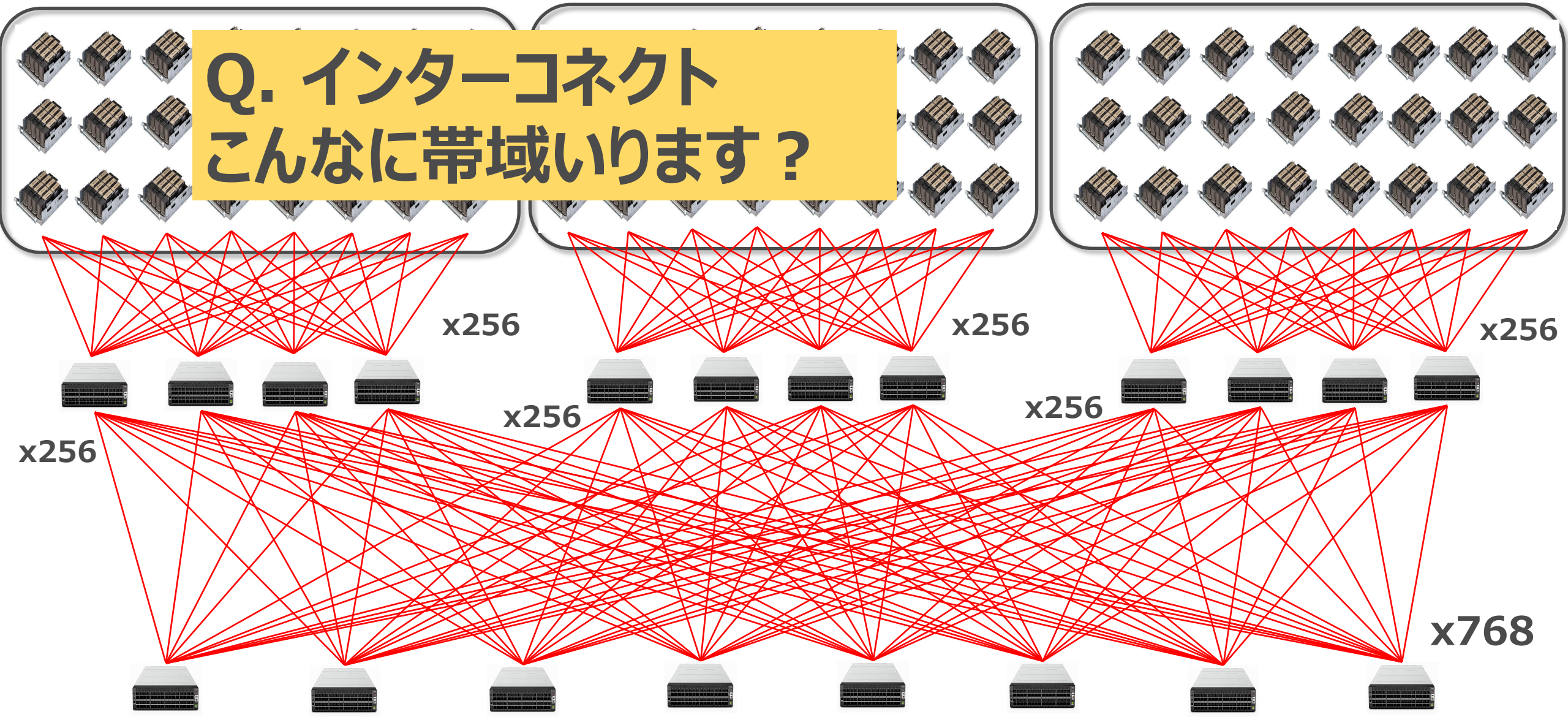
GMO GPUクラウド



ネットワーク構成

GMO GPUクラウド

Q. インターコネクト こんなに帯域いります？



GMO GPUクラウド

サービス提供は、
最短・最速

GMO GPUクラウド

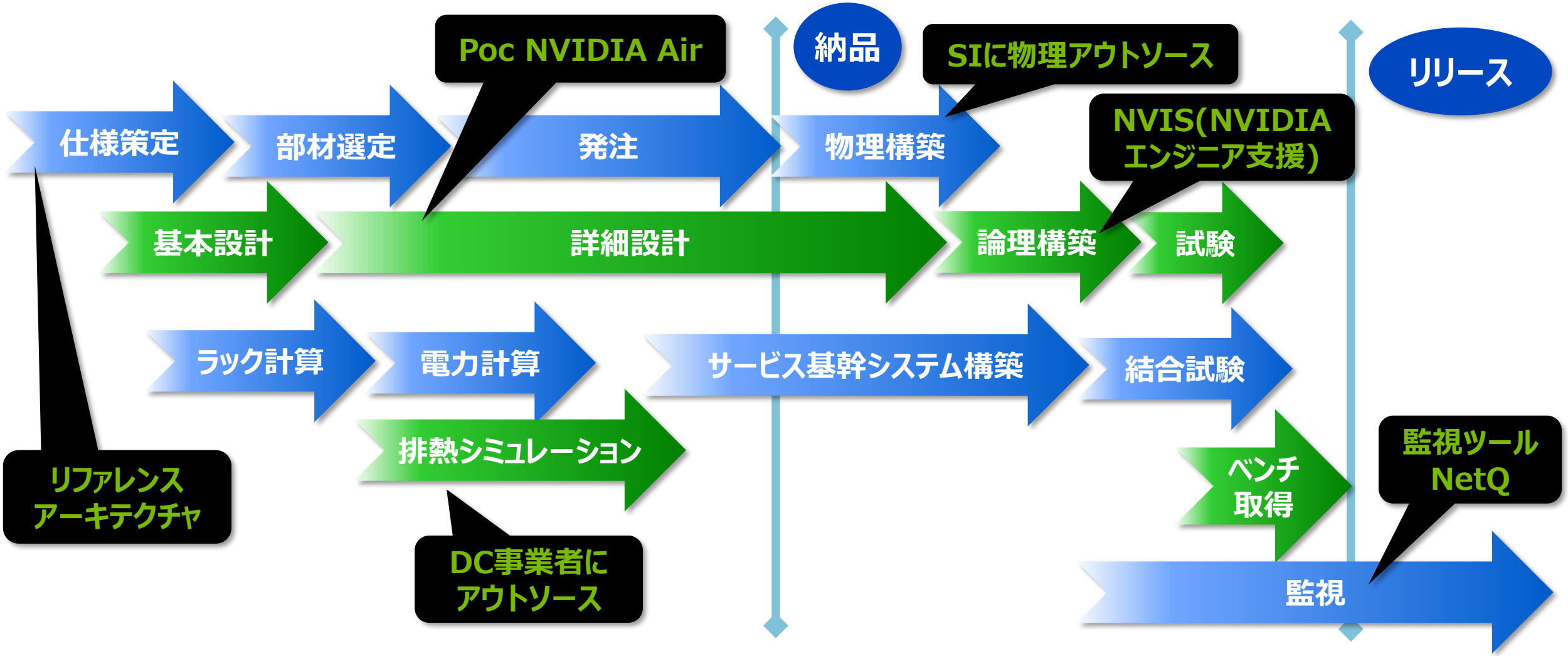
GPUは日に日に性能が上がるため 鮮度が命!!

最短最速でサービスインする為の計画を策定

GMOインターネットの事例

GMO GPUクラウド

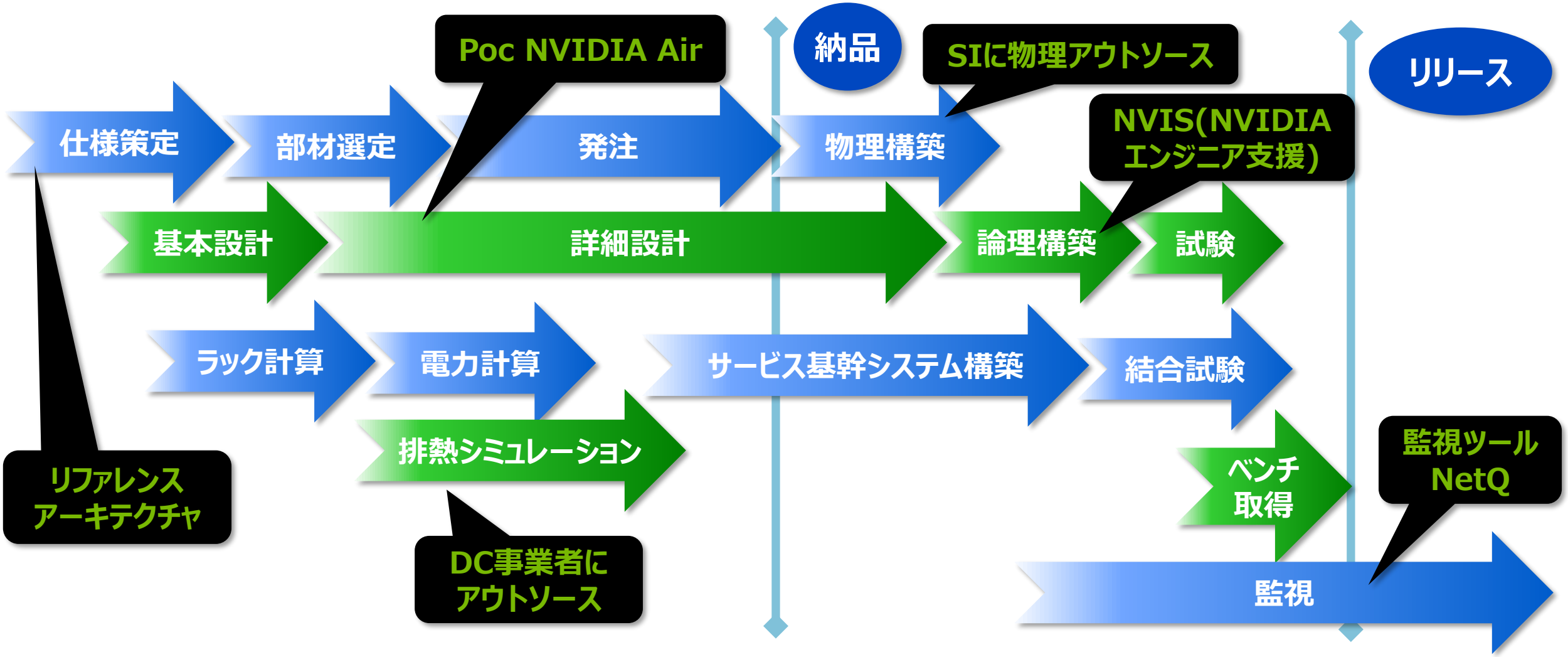
関係各社と連携、必要に応じてアウトソース、NVIDIAフレームワーク活用。納品から3カ月でリリース！



GMOインターネットの事例

関係各社と連携、必要に
NVIDIAフレームワーク活用。納品から3分月でリリース！

Q. 短期リリースが大変！
上手なやり方ないの？



GMO GPUクラウド

NVIDIA さんに聞いてみよう

会社紹介

GMO GPUクラウド

NVIDIA

主な事業分野

- GPU, AI, Network

2019年 Mellanox Technologies 買収

2020年 Cumulus Networks 買収

主なネットワーク製品

- Quantum InfiniBand プラットフォーム
- Spectrum Ethernet プラットフォーム
- ConnectX / BlueField NIC シリーズ
- LinkX トランシーバー&ケーブル ファミリー

InfiniBand を筆頭にスパコン、HPC 分野から続く AI ネットワーキングにも欠かせないインターコネクト技術をリード

Notes

GMO GPUクラウド

GPU ネットワーク = AI ネットワーク = インターコネクト

GPU = アクセラレーター

1Byte = 8bit

GMO GPUクラウド

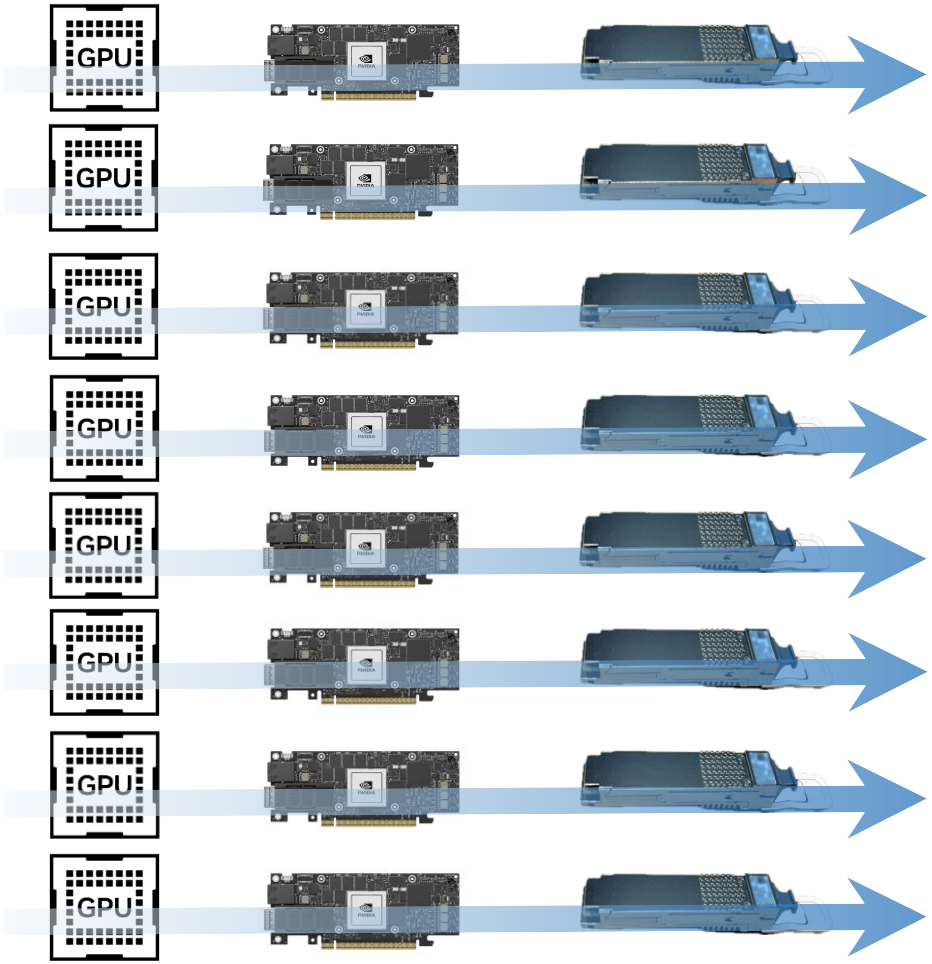
Q.インターコネクトってこんなに帯域いります？

GMOインターネットの事例

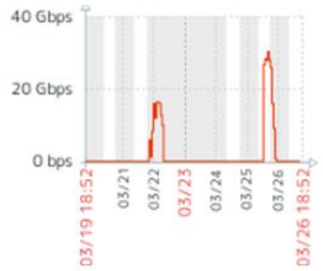
GMO GPUクラウド

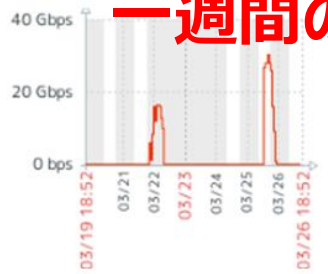
400Gラインで、実際の利帯域は多いときで、55Gほどに見える。

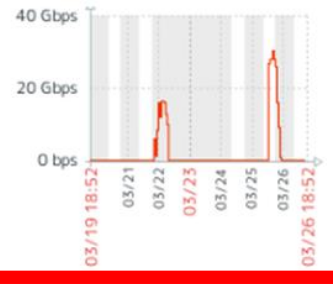
GPU　BF3NIC　400Gb-Optics

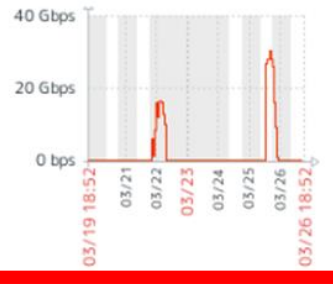


一週間のグラフ

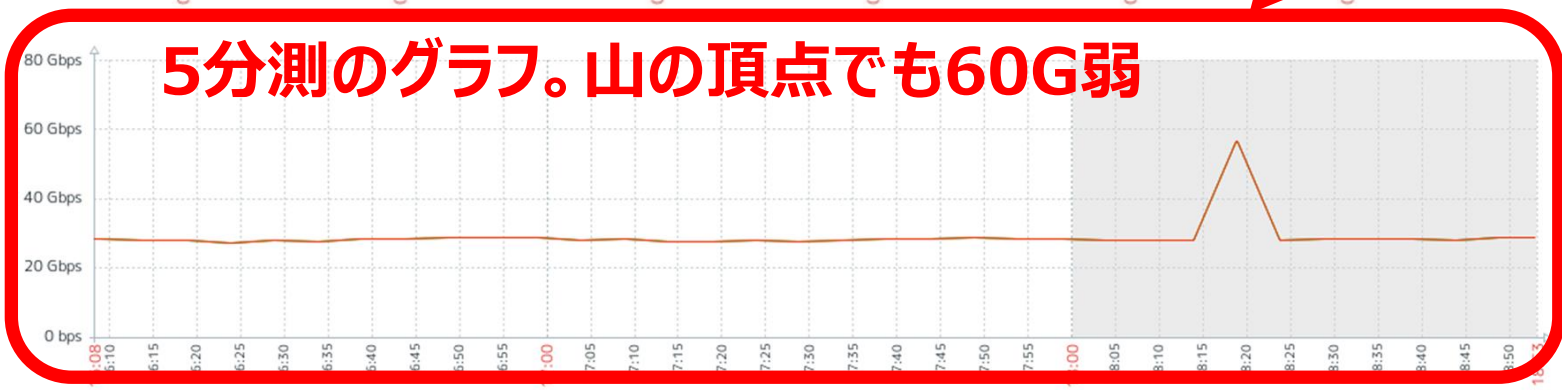








5分測のグラフ。山の頂点でも60G弱



NVIDIAさんに聞いてみよう

GMO GPUクラウド

GMOの顧客だとフルスピード400Gは常に出てない

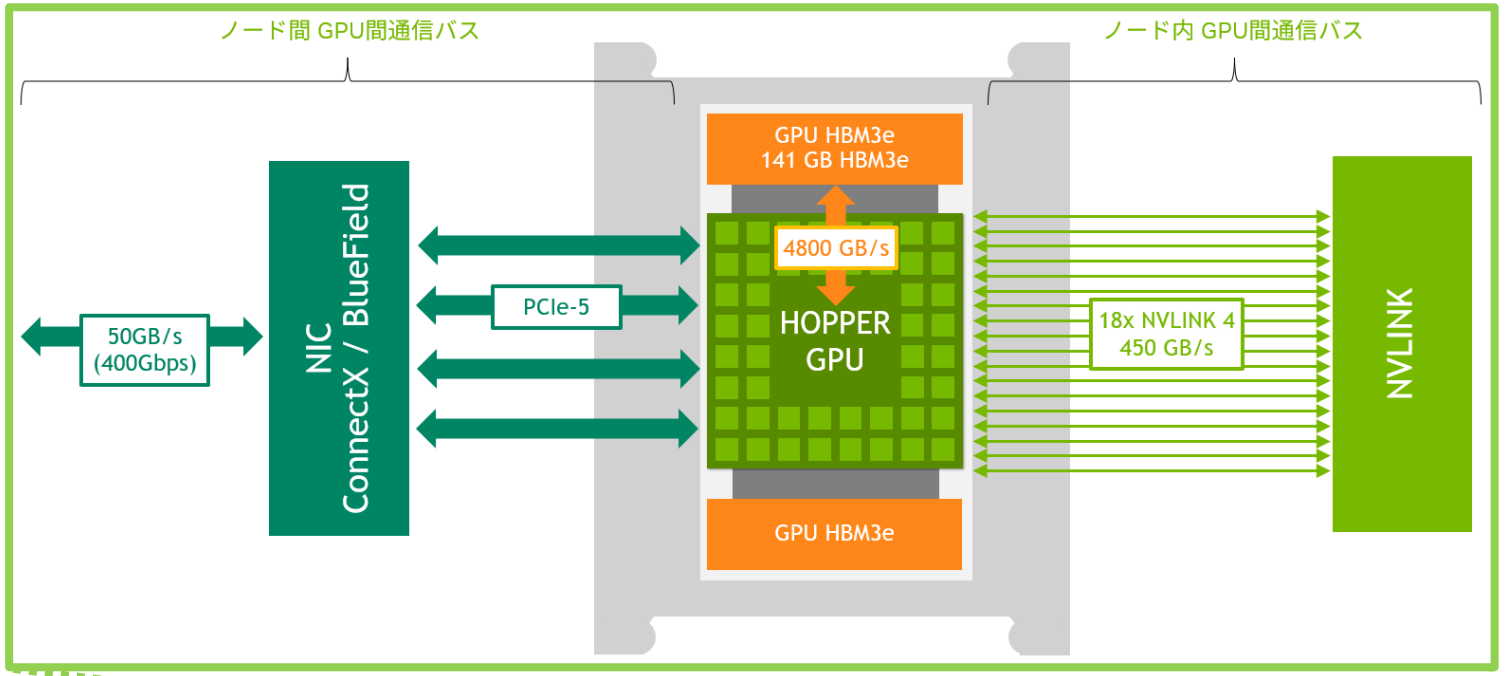
→実際のユーザで400G使うことはあるのでしょうか？

→低速だと何が困るのでしょう？
　例えば100Gとか200Gだと、どうまずいのでしょうか？

NVIDIA 補足資料

GMO GPUクラウド

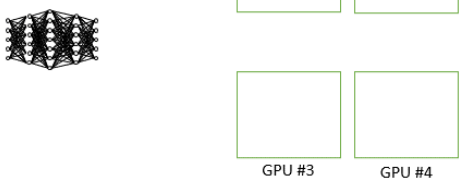
NVIDIA



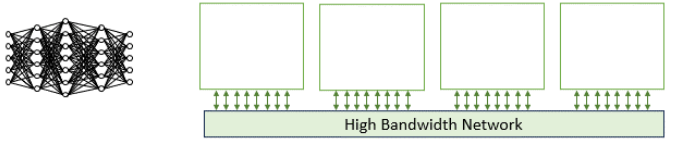
NVIDIA 補足資料

GMO GPUクラウド

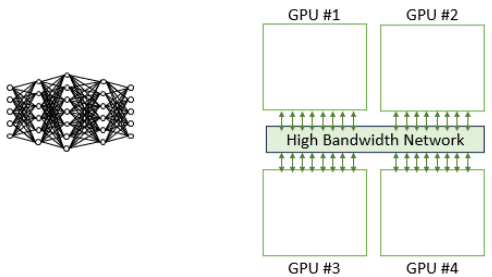
Data Parallel



Pipeline Parallel

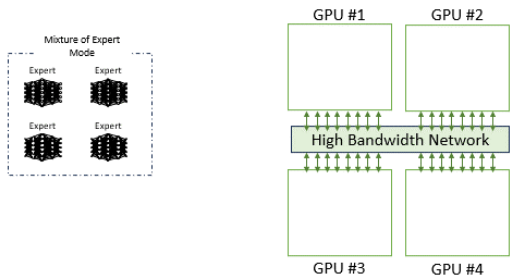


Tensor Parallel

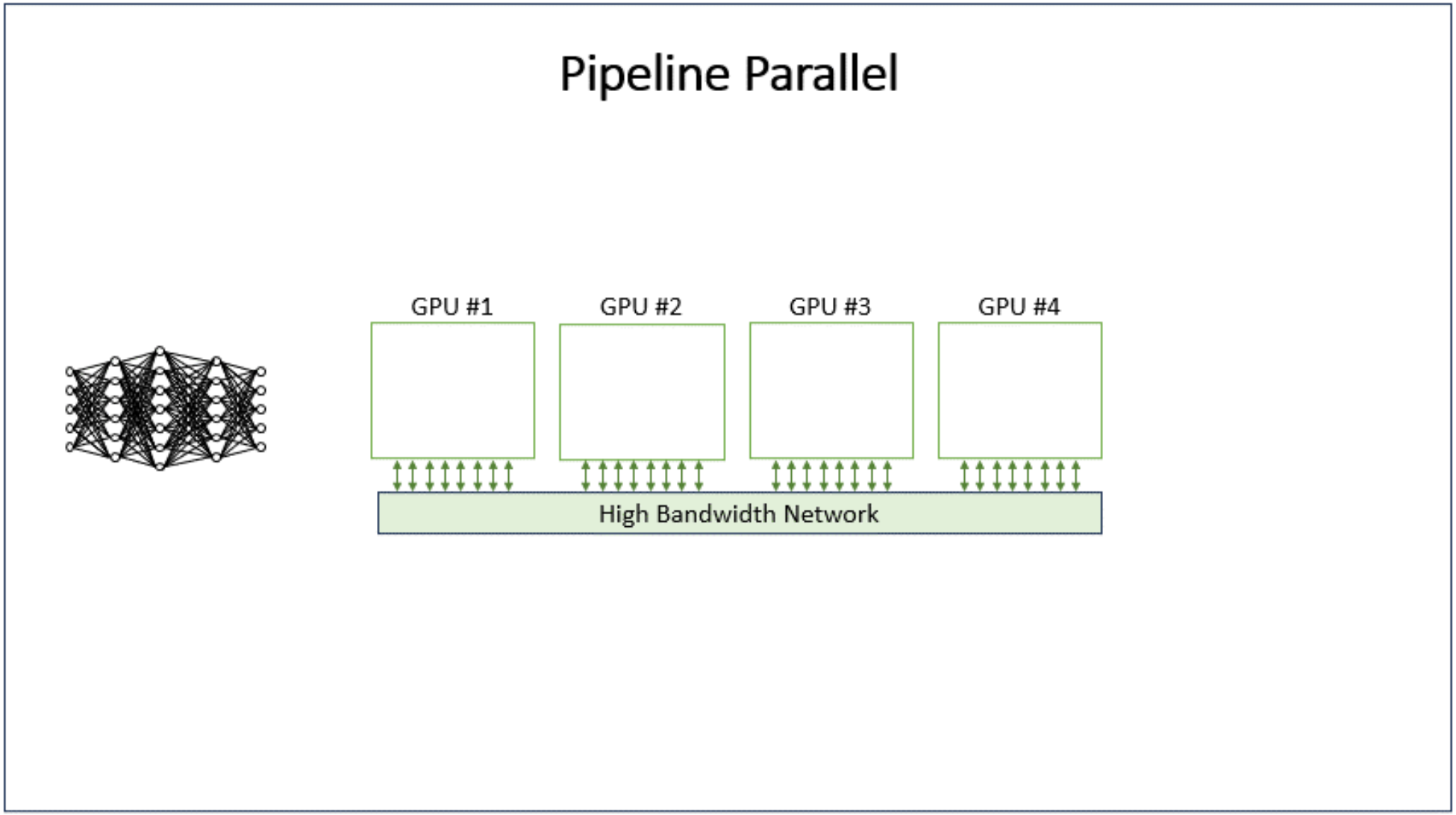


*C2C: Chip-to-Chip

Expert Parallel

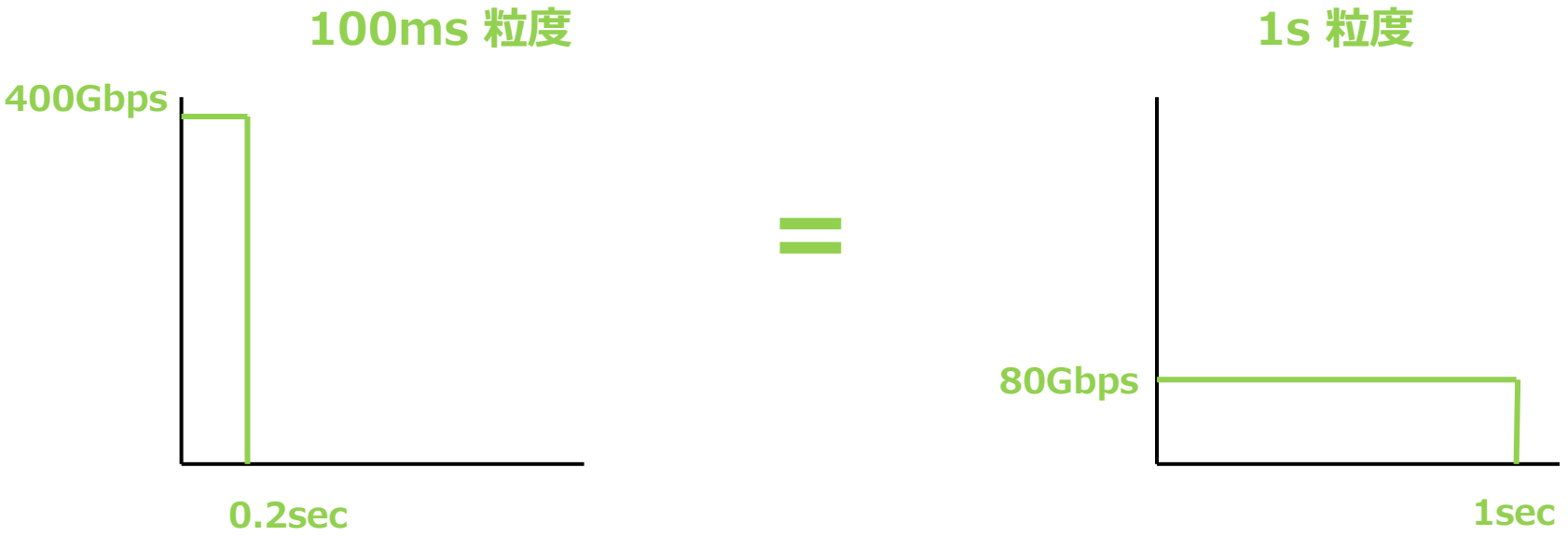


NVIDIA


Pipeline Parallel
GPU #1
GPU #2
GPU #3
GPU #4
High Bandwidth Network

NVIDIA 補足資料

GMO GPUクラウド



NVIDIA

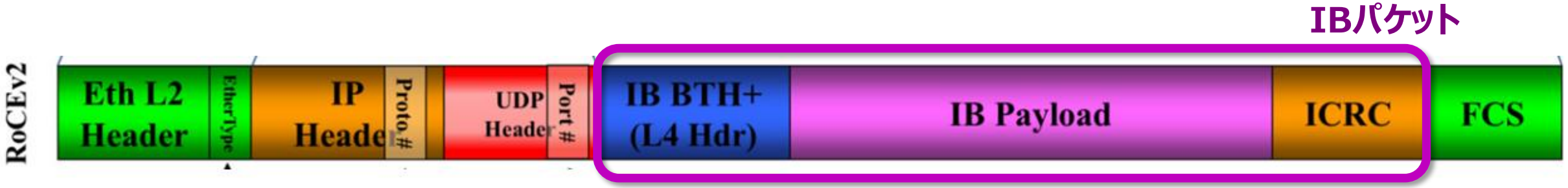
GMO GPUクラウド

Q. 輻輳制御ってそんなに大事なの？

GMO GPUクラウド

RoCEv2 (RDMA over Converged Ethernet ver2) を利用し、RDMA通信のIBパケットをカプセル化

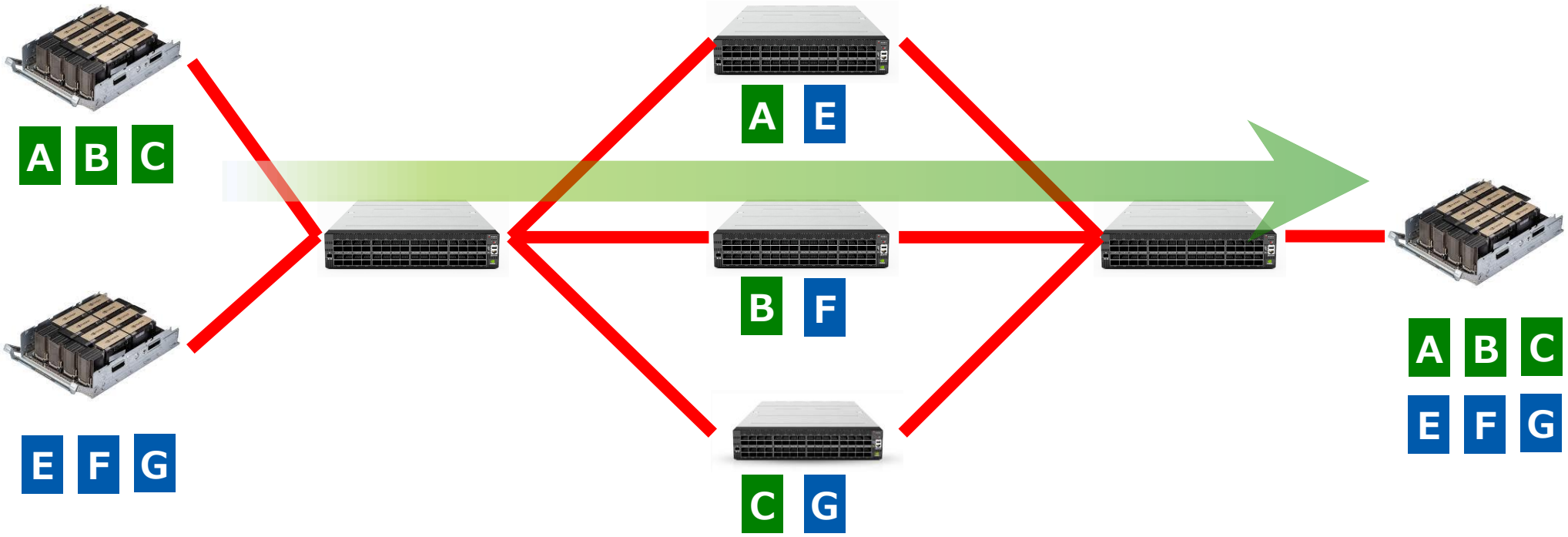
加えて、ECN・PFC などを利用し 輻輳制御、ロスレス通信を目指す。



※NVIDIA DOCS HUB RoCEv2 より引用
https://docs.nvidia.com/networking/display/winofv55053000/rocev2

GMO GPUクラウド

アダプティブルーティングで全帯域を有効活用。
SWでインターフェイスごと通信を監視し、動的にバランシング
トラフィックを均等に転送し、
かつ、順序も維持することで**輻輳回避**を実現。



GMOインターネットの事例

GMO GPUクラウド

Spectrum-Switch & NVIDIA BlueField-3 SuperNIC を搭載した本システムでは、RoCEとアダプティブルーティングに加えてエンドツーエンドの輻輳制御**を実装。GPU間通信の速度を調整し、より低遅延のロスレスNWを実現**

NVIDIAさんに聞いてみよう

GMO GPUクラウド

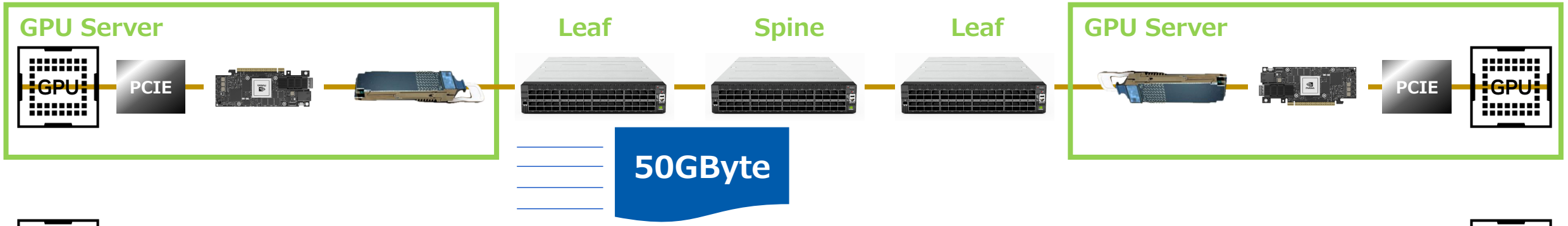
・RoCEとアダプティブルーティングに加えて、
エンドツーエンドの輻輳制御を入れていますが、

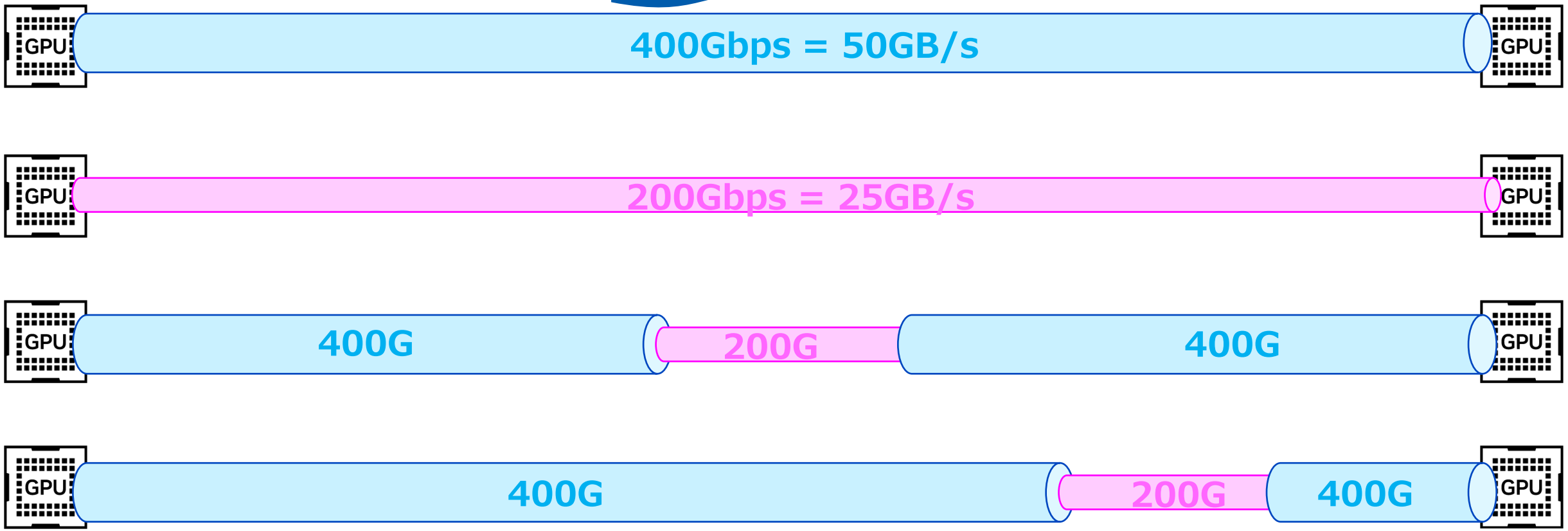
→実装することで、どんな良い事があるのでしょうか？

→実装していないと、どういうシーンで困るのでしょうか？

NVIDIA 補足資料

GMO GPUクラウド



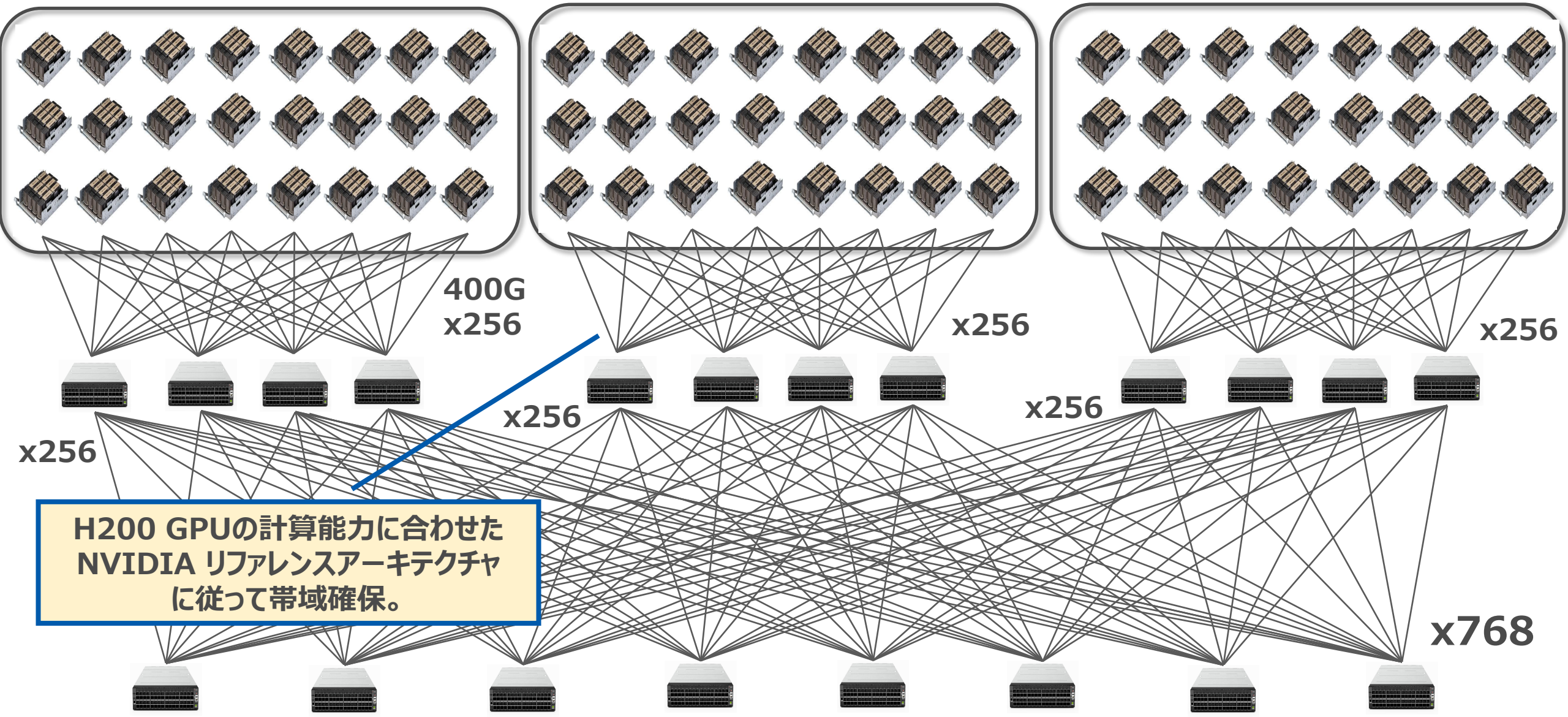


NVIDIA

Q.リファレンスアーキテクチャの次の世代は？

GMOインターネットの事例

GMO GPUクラウド



H200 GPUの計算能力に合わせた
NVIDIA リファレンスアーキテクチャ
に従って帯域確保。

NVIDIAさんに聞いてみよう

GMO GPUクラウド

今後、次世代のGPUを接続するインターコネクト
ネットワークの構築を検討しています

→次世代GPUのリファレンスアーキテクチャは
　どう変わりますか？

→インターコネクトの帯域はどう変わりますか？
　増えます？

GMO GPUクラウド

Q.InfinibandかEthernetか？今後はどっちが主流？

NVIDIAさんに聞いてみよう

GMO GPUクラウド

弊社はマルチテナントなどを理由にEthernetを入れたけど、
Infinibandの場合のメリットも知りたい。

→正直なところ、
　クラウド事業者にはどっちがおすすめですか？

GMO GPUクラウド

■ 議論のポイント

・輻輳制御どうしていますか？
・どのトポロジ採用してる？したい？
・どれくらいの帯域を採用したい？
・構築の時短テクニックはありますか？
・ICケーブルにDACとかAOC採用してる？
・インターコネクト起因のパフォーマンス問題に直面したことはありますか？どのように切り分けましたか？

気になるところあれば、回答できる範囲で答えます！

すべての人にインターネット

GMO

NVIDIA®