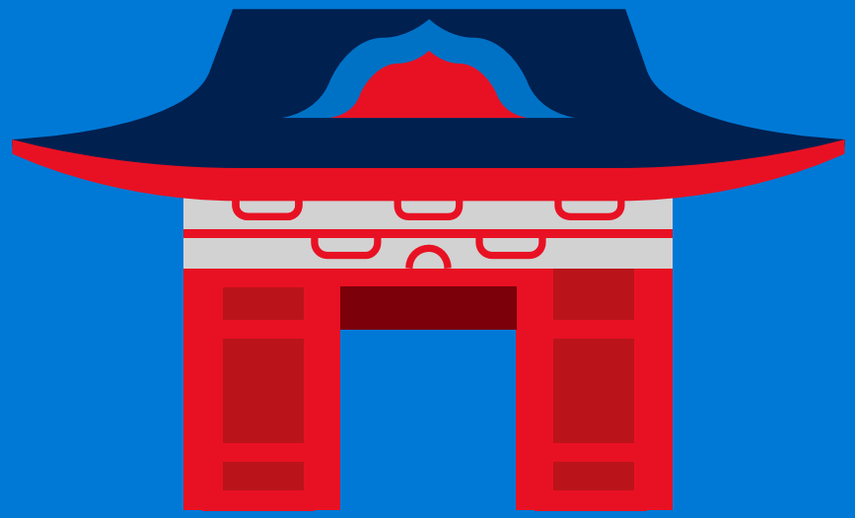
Microsoft

OCPJ Meet-up 2017 夏

Project Olympus Update



日本マイクロソフト株式会社
クラウド&ソリューションビジネス統括本部
クラウドソリューションアーキテクト
真壁 徹

自己紹介

{
  “名前” : “真壁 徹(まかべ とおる)”,
  “所属” : “日本マイクロソフト株式会社”,
  “役割” : “クラウド ソリューションアーキテクト”,
  “経歴” : “大和総研 → HP Enterprise”,
  “特技” : “クラウド & オープンソース”
}

Project Olympus

- OCPでマイクロソフトがリードするプロジェクト
- ハードウェア開発にクラウドのスピード感を
- オープンソース「ソフトウェア」のやり方から学ぶ
- 未完成段階でも検討中の仕様を公開
- フィードバックを得て
  - より良い仕様に
  - よりメーカーが製造しやすく
- On GitHub
  - https://github.com/opencomputeproject/Project_Olympus

Project Olympus Ecosystem Partners

CPU

- intel
- AMD
- QUALCOMM
- CAVIUM

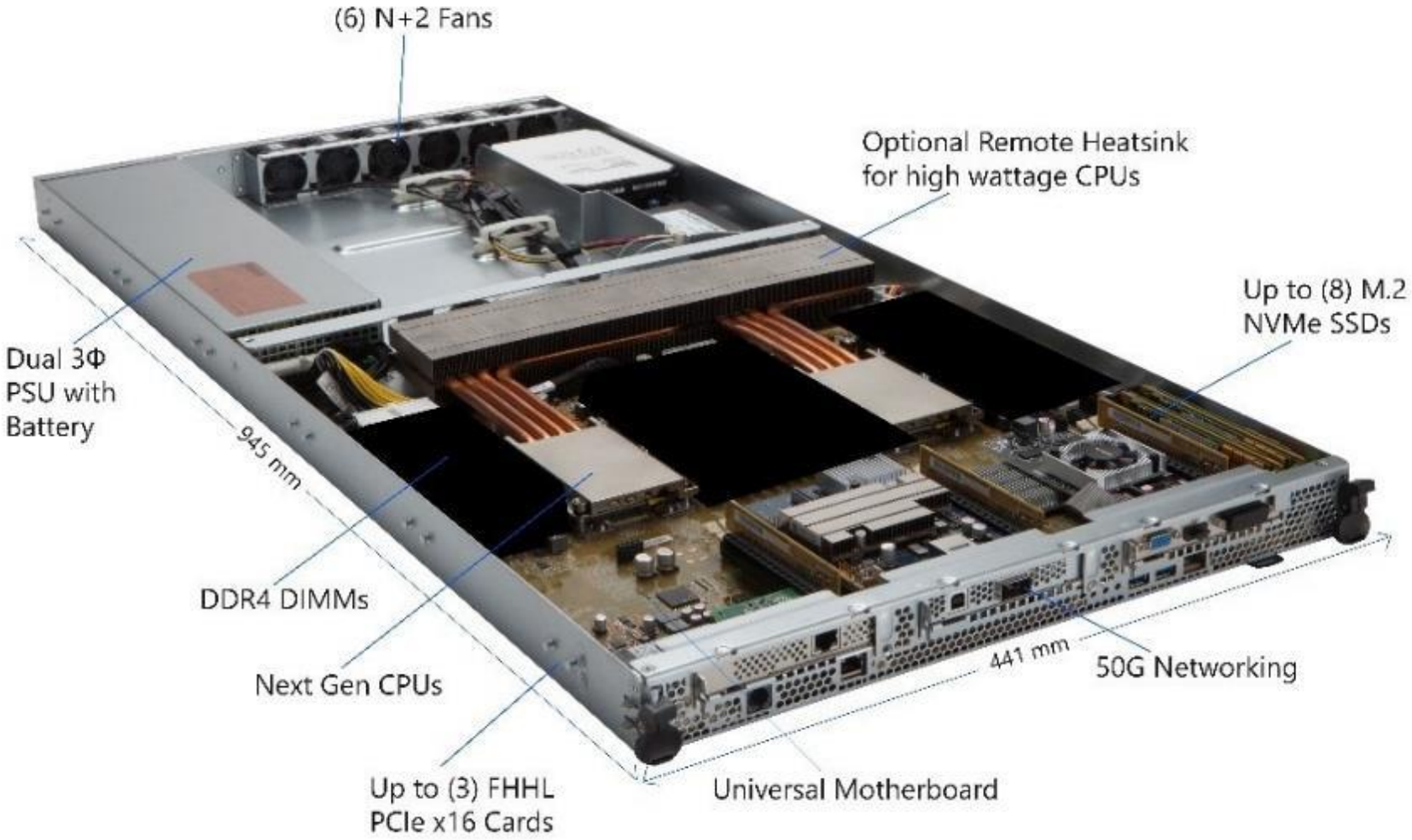
COMPONENTS

- NVIDIA
- BROADCOM
- Mellanox TECHNOLOGIES
- Texas Instruments
- flex
- ARTESYN
- LITEON
- HARTING
- Schneider Electric
- ingrasys
- Inventec
- GEIST
- RITTAL
- hynix
- SAMSUNG
- MARVELL

SYSTEMS

- Hewlett Packard Enterprise
- DELL
- wiwynn
- Quanta
- zt Systems
- inspur

Project Olympus シャーシデザイン



https://schd.ws/hosted_files/ocpussummit2017/f2/OCP17%20Workshop_Microsoft%20Project%20Olympus%20Servers_3_8_2017.pdf

OCP Summit 2017での主な発表

新世代チップへの対応

Intel: Xeon Skylate / Arria 10 FPGA

AMD: Naples

ARM64対応

with Qualcomm, Cavium

GPU拡張シャーシ

HGX-1 with NVIDIA

AMD Naples対応

- Ryzenでにわかに盛り上がるAMD Zenアーキテクチャー
- 伝説の石師 Jim Kellerが開発をリード
- そのサーバー向け実装がNaples

https://schd.ws/hosted_files/ocpussummit2017/f2/OCP17%20Workshop_Microsoft%20Project%20Olympus%20Servers_3_8_2017.pdf

ARM64対応

- Cavium ThunderX2
- Qualcomm Centriq 2400
- モバイルデバイス向けに数が伸び続けるARMへの期待
  - 電力あたり性能
  - エコシステム
- Windows ServerをARM対応中
  - まずはAzure内、ユーザーが意識しないところで採用
  - ソフトウェアとしての販売は未定

https://schd.ws/hosted_files/ocpussummit2017/f2/OCP17%20Workshop_Microsoft%20Project%20Olympus%20Servers_3_8_2017.pdf
http://files.opencompute.org/oc/public.php?service=files&t=5e232f151532e2af3b9754af4ca79f0b

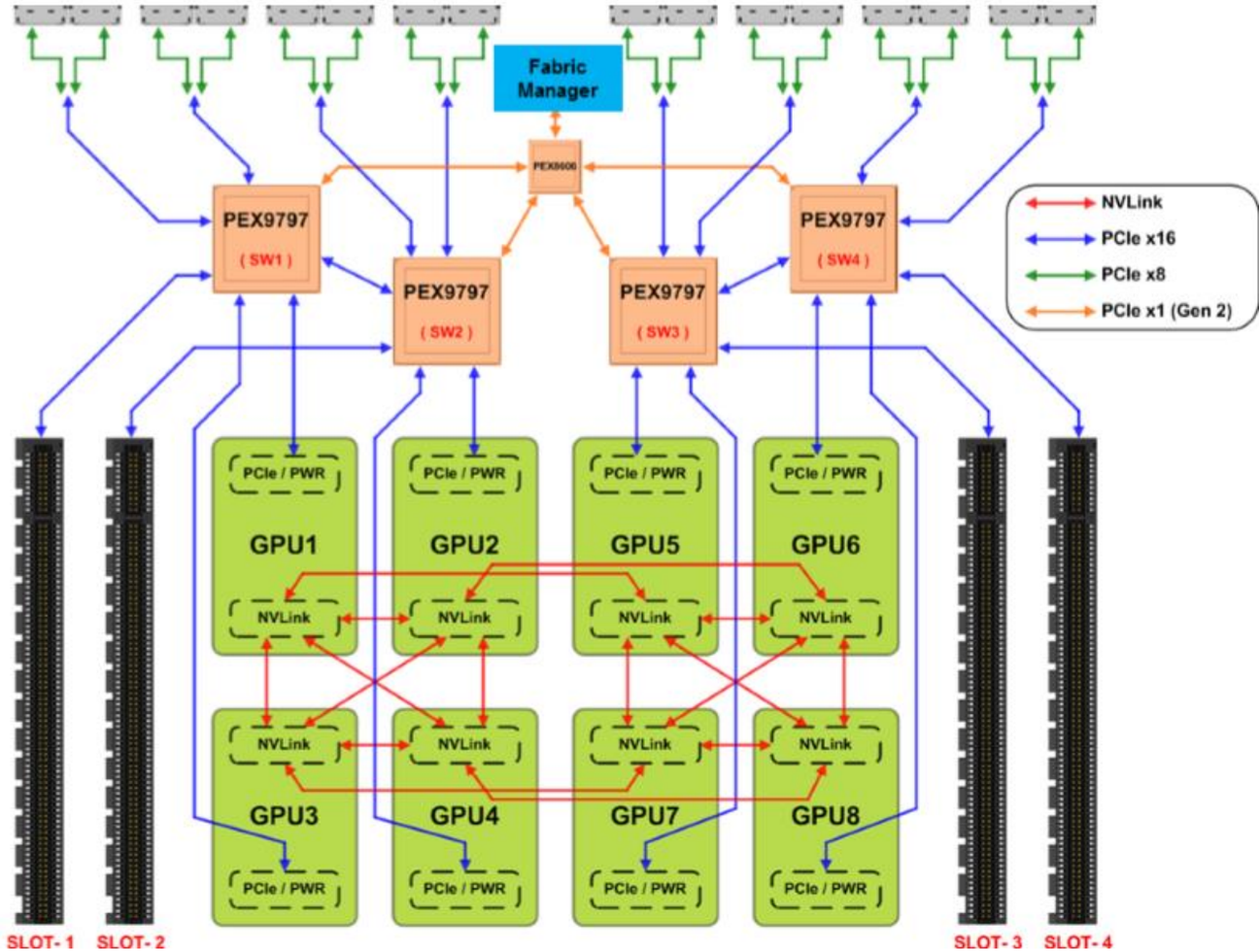
GPU拡張シャーシ HGX-1

- “Hyperscale GPU Accelerator”
- NVIDIAとの共同開発
- Tesla P100 SMX2 x8 & NVLink
- 4シャーシまで拡張可
  - Internal PCIe Fabric Interconnect
- 16サーバーまで共用可

http://files.opencompute.org/oc/public.php?service=files&t=14ab3cf25170b7a0a439e11a3d818c96
https://azure.microsoft.com/en-us/blog/ecosystem-momentum-positions-microsoft-s-project-olympus-as-de-facto-open-compute-standard/

HGX-1 Topology

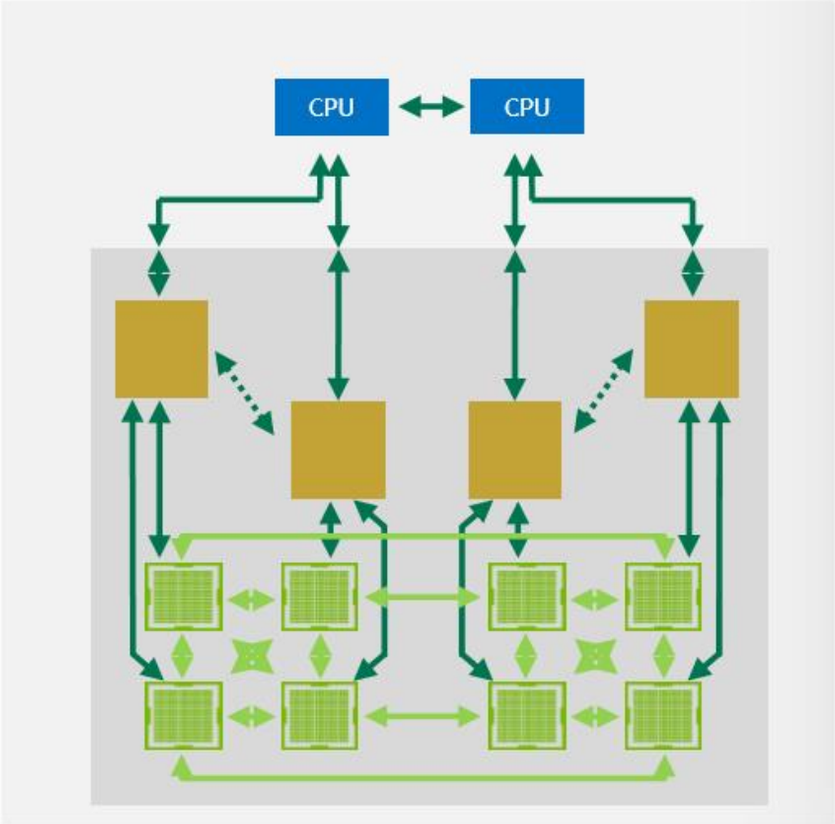
- Flexible PCIe Interconnect Topology
- GPGPU-to-Host via high-BW PCIe Links
- Peer-to-peer without Host interaction
  - GPGPU peer-to-peer via NVLink
  - GPGPU peer-to-peer to IB NICs via x16 PCIe



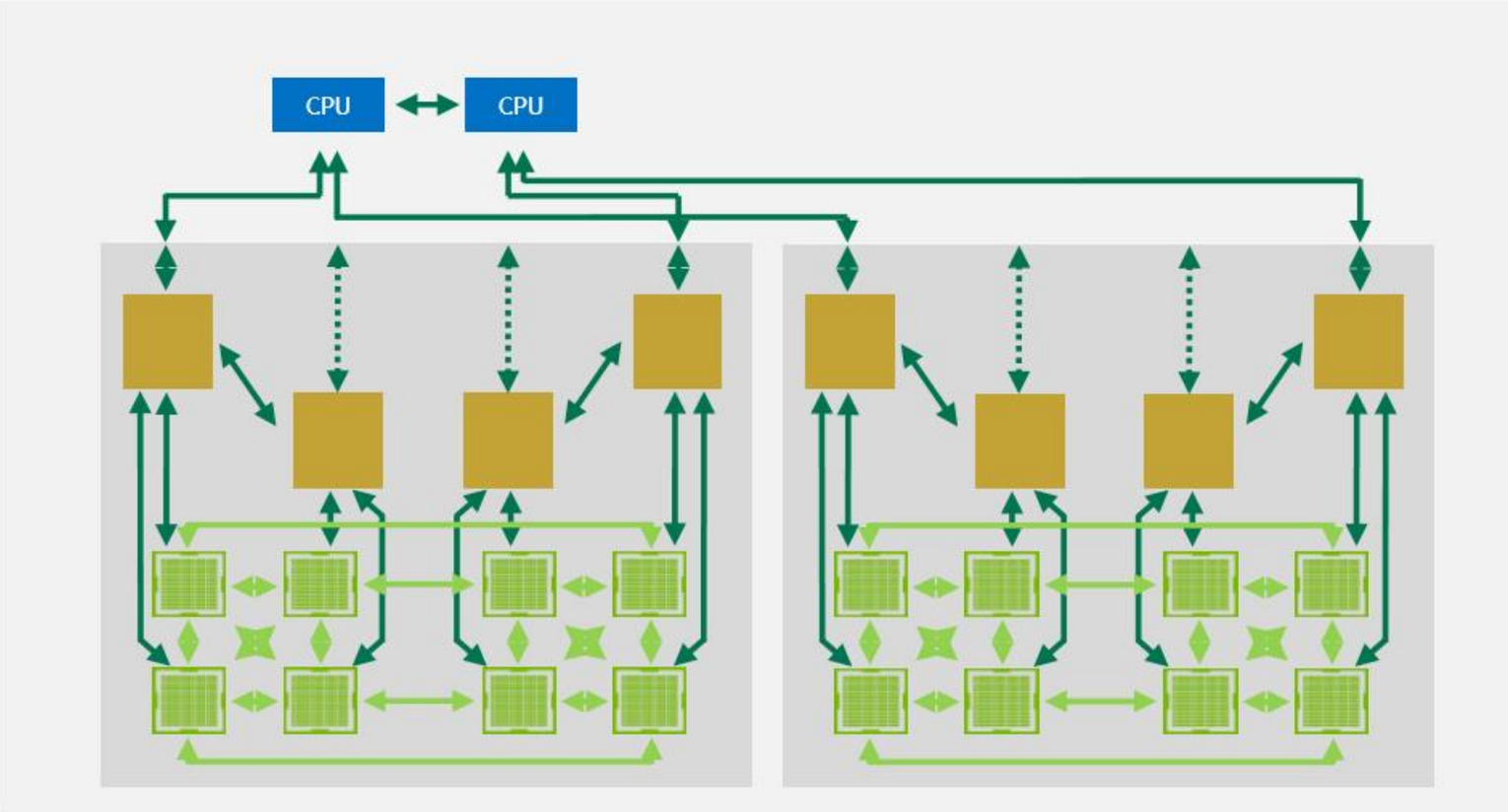
http://files.opencompute.org/oc/public.php?service=files&t=14ab3cf25170b7a0a439e11a3d818c96

DEEP LEARNING

2 CPU : 8 GPU
8x P100 SXM2 | 4x x16 PCIe

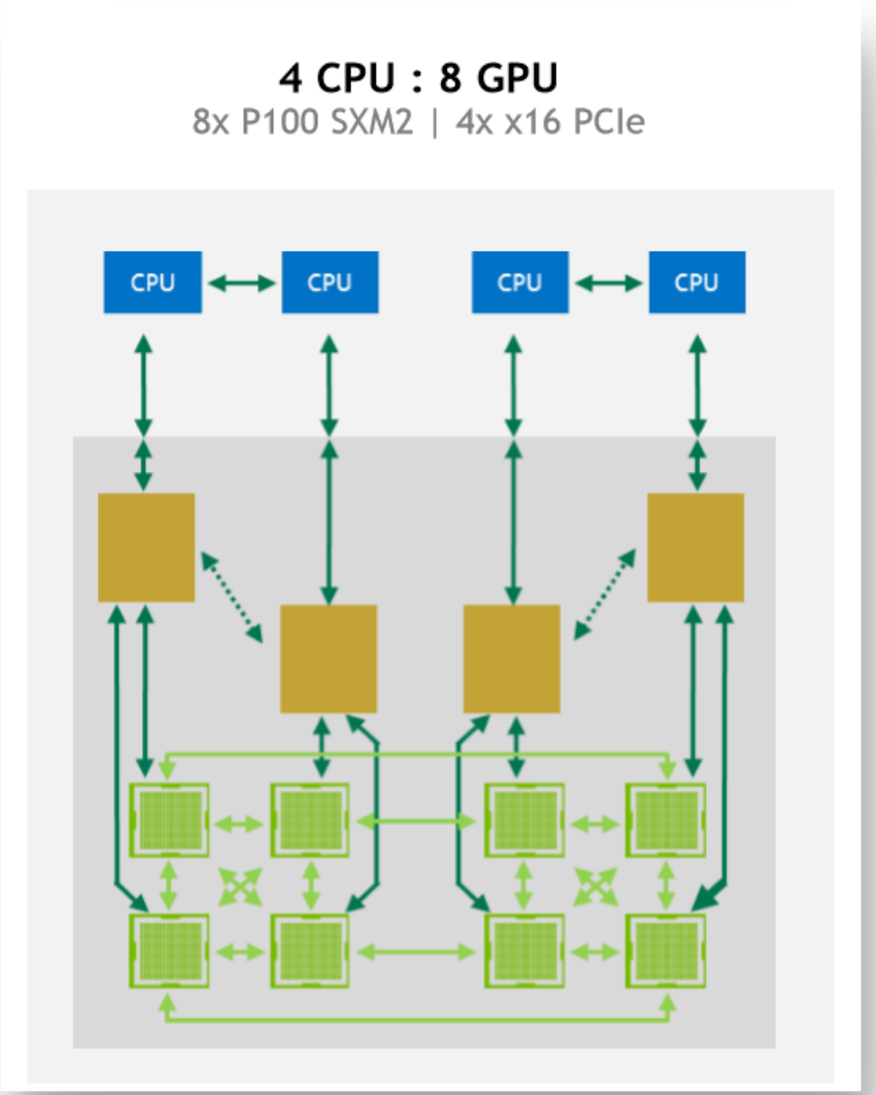


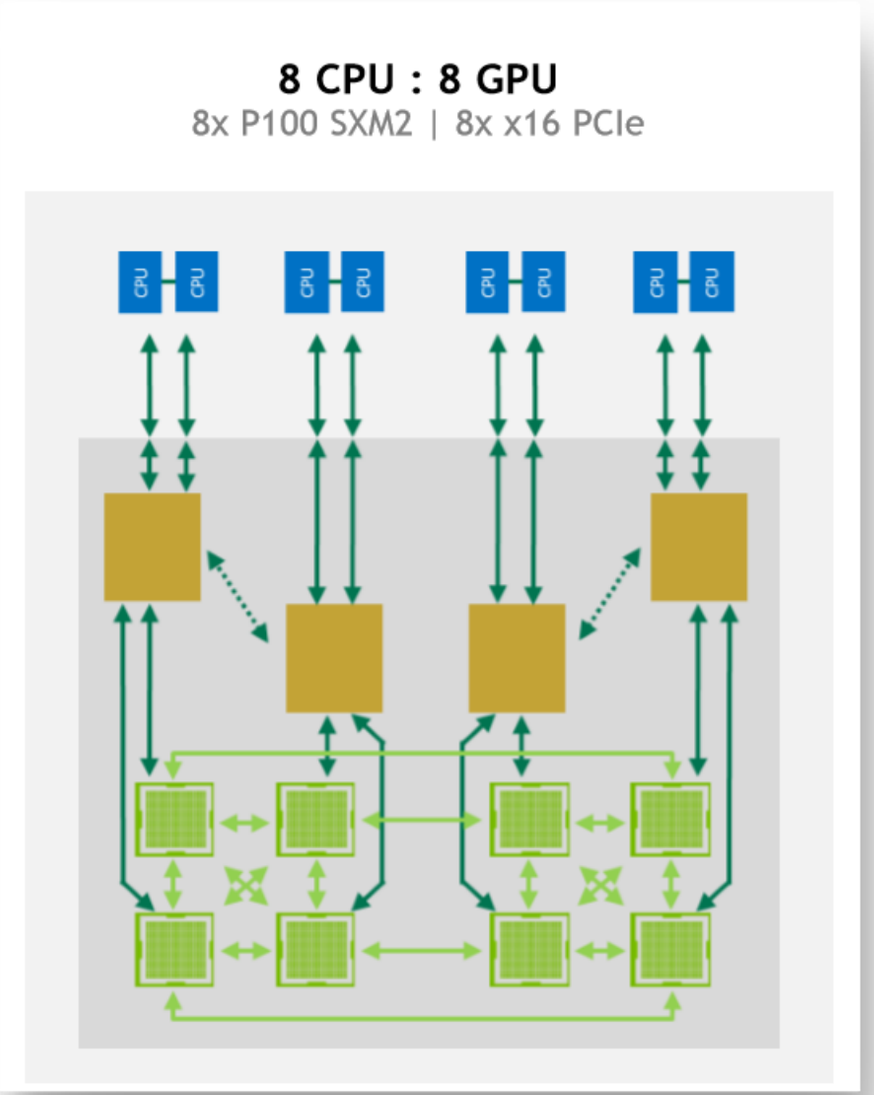
2 CPU : 16 GPU
16x P100 SXM2 | 4x x16 PCIe



http://files.opencompute.org/oc/public.php?service=files&t=14ab3cf25170b7a0a439e11a3d818c96

HPC

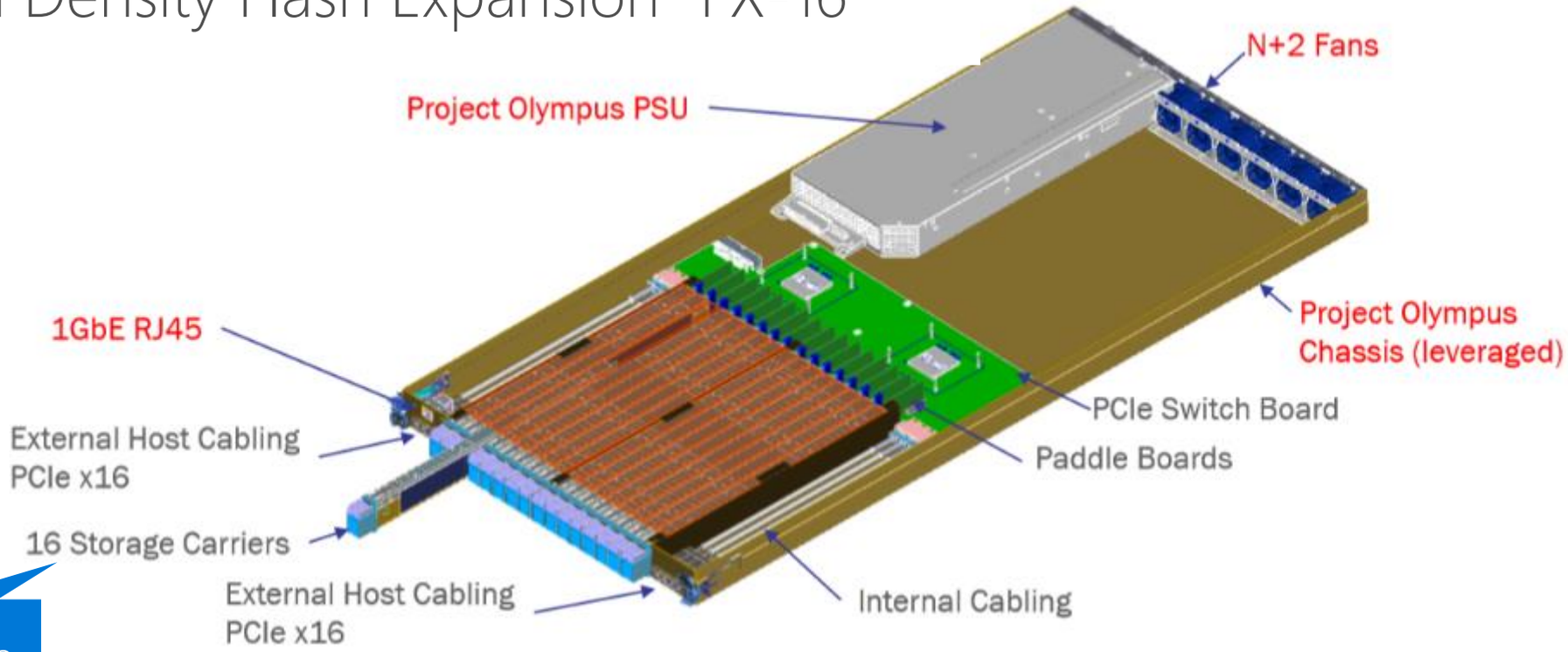




http://files.opencompute.org/oc/public.php?service=files&t=14ab3cf25170b7a0a439e11a3d818c96

Project Olympus Flash Expansion

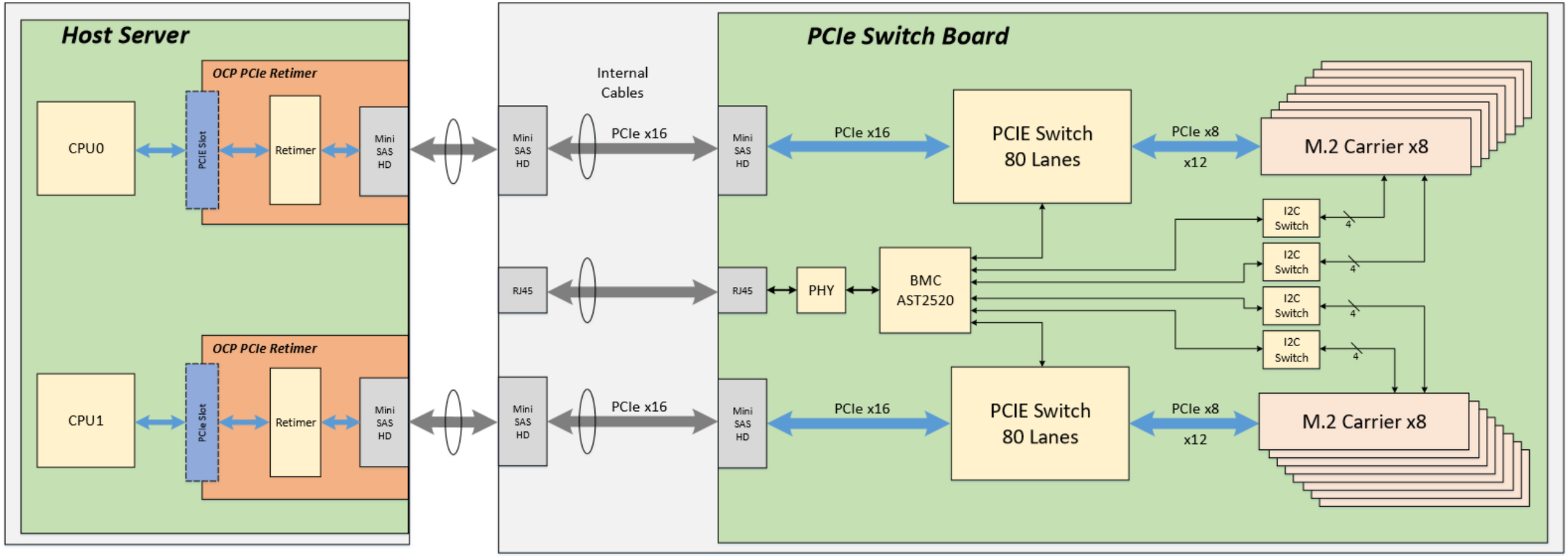
High Density Flash Expansion "FX-16"



M.2 NVMe SSD

http://files.opencompute.org/oc/public.php?service=files&t=309c5b1d58539b56a4cd5d499c5b3a85

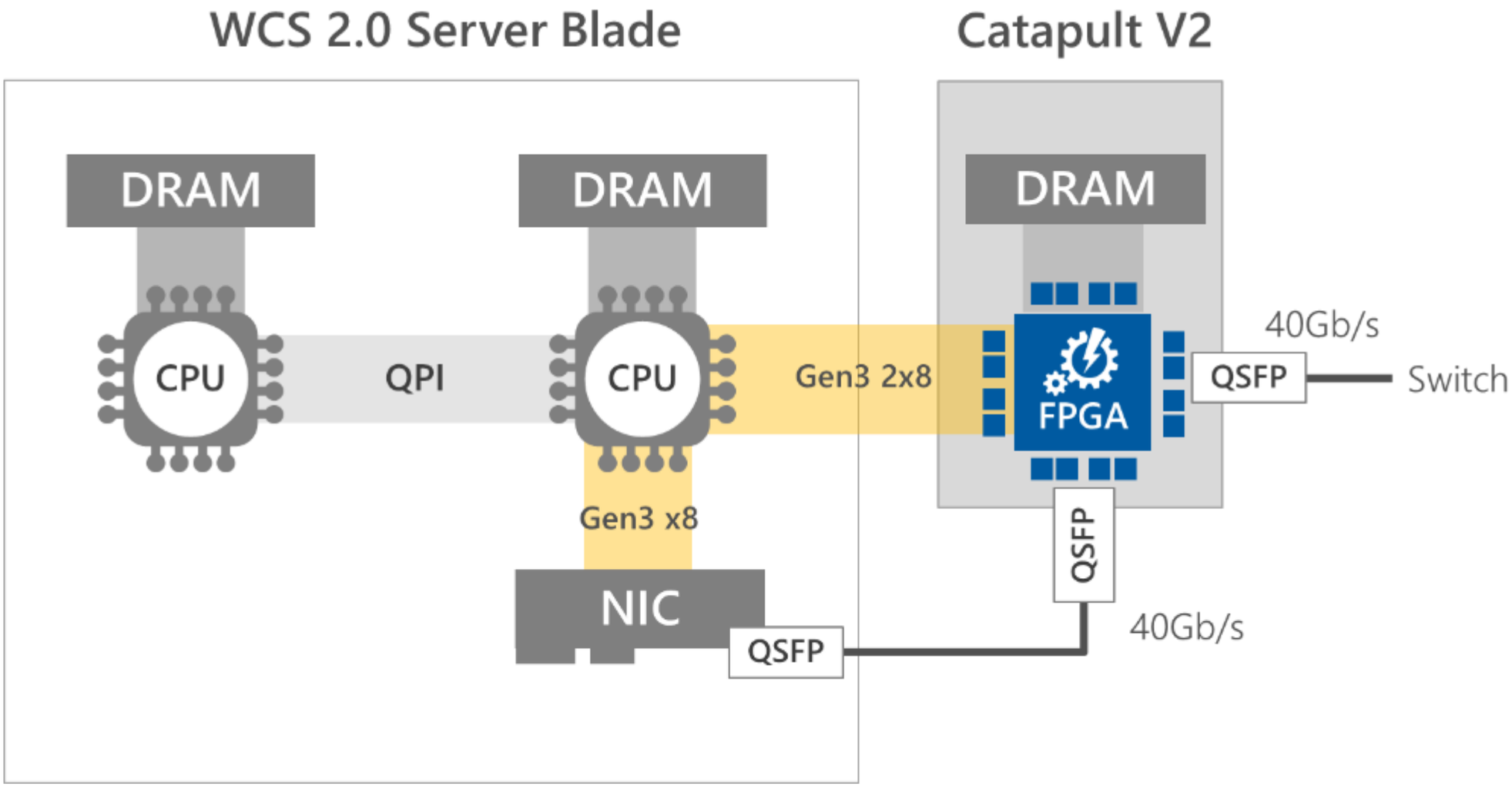
Project Olympus Flash Expansion



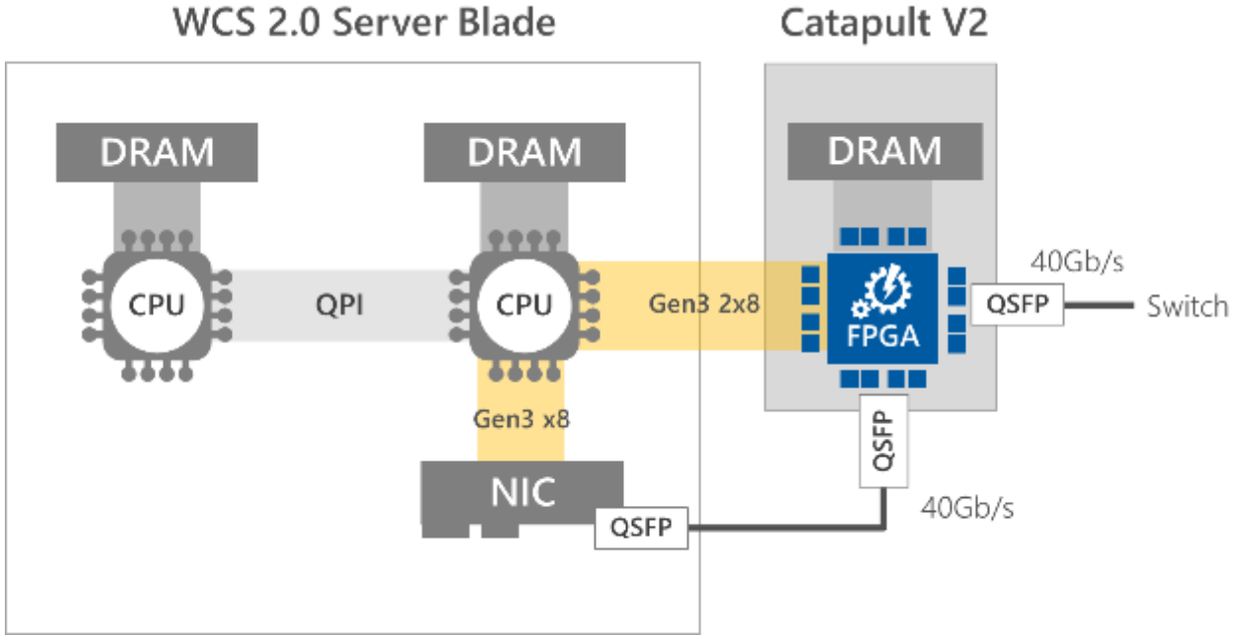
http://files.opencompute.org/oc/public.php?service=files&t=309c5b1d58539b56a4cd5d499c5b3a85

(おまけ)
OCP Server in Azure

Azure “Catapult V2” architecture

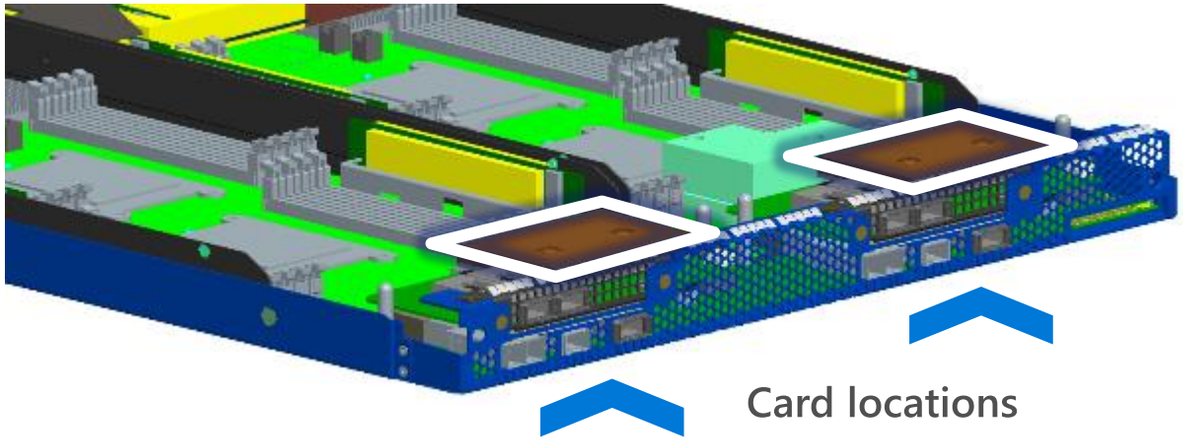


Azure “Catapult V2” architecture



Catapult v2 Mezzanine card

WCS Gen4.1 Blade with NIC and Catapult FPGA



Software Defined Networkにおける挑戦

NICは年々広帯域化:

1G → 10G → 40G → 50G → 100G

VMを集約することでコストを下げられるが、同時に帯域が必要になる

CPUがネットワーク処理で飽和してしまう

新しいネットワーク機能の実装:

BYO IP, BYO Topology, BYO Appliance

今後も多様な処理、役割をSDN、仮想ネットワーク上で実現したい

新しいアイデアを迅速、柔軟に実装したい (ASICでは時間がかかりすぎる、リスクが高すぎる)

"

ハードウェア実装による性能とソフトウェア実装による柔軟性を両立できないか?

Azure SmartNIC

FPGAにネットワーク関連処理を実装

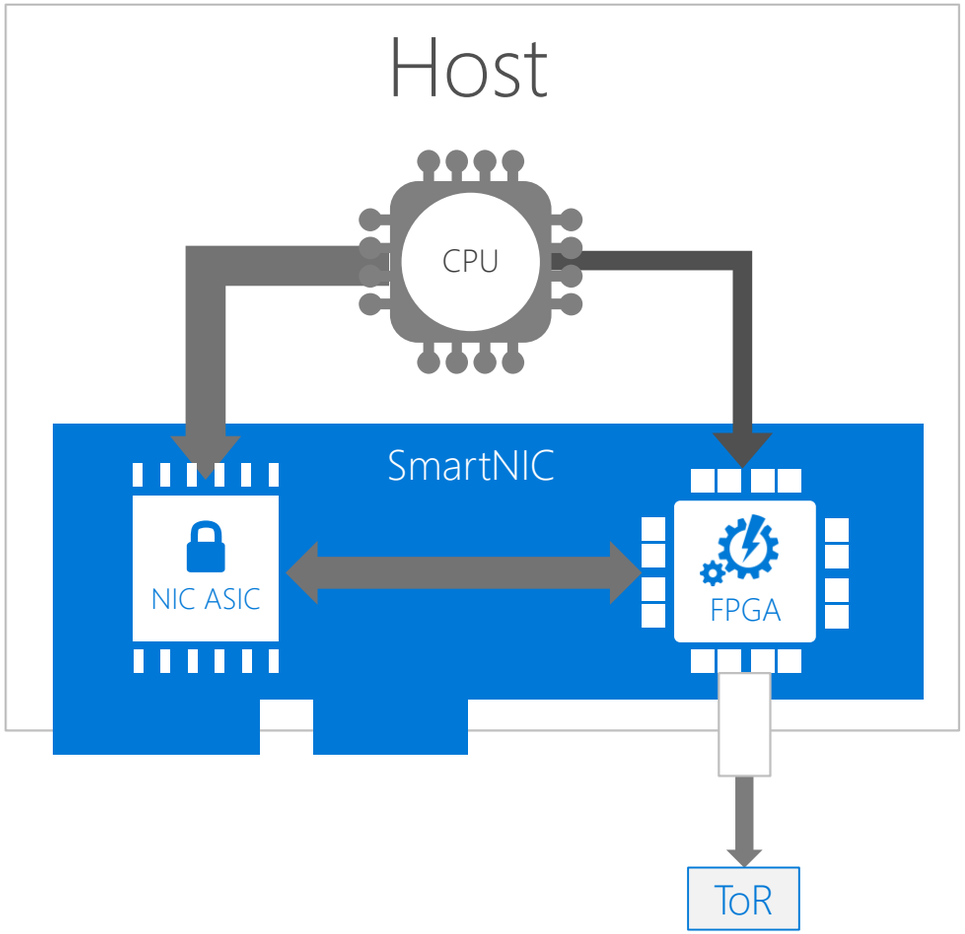
すでにBing (Catapult)でFPGAの知見があった

Generic Flow Tables (GFT)に従いパケット処理

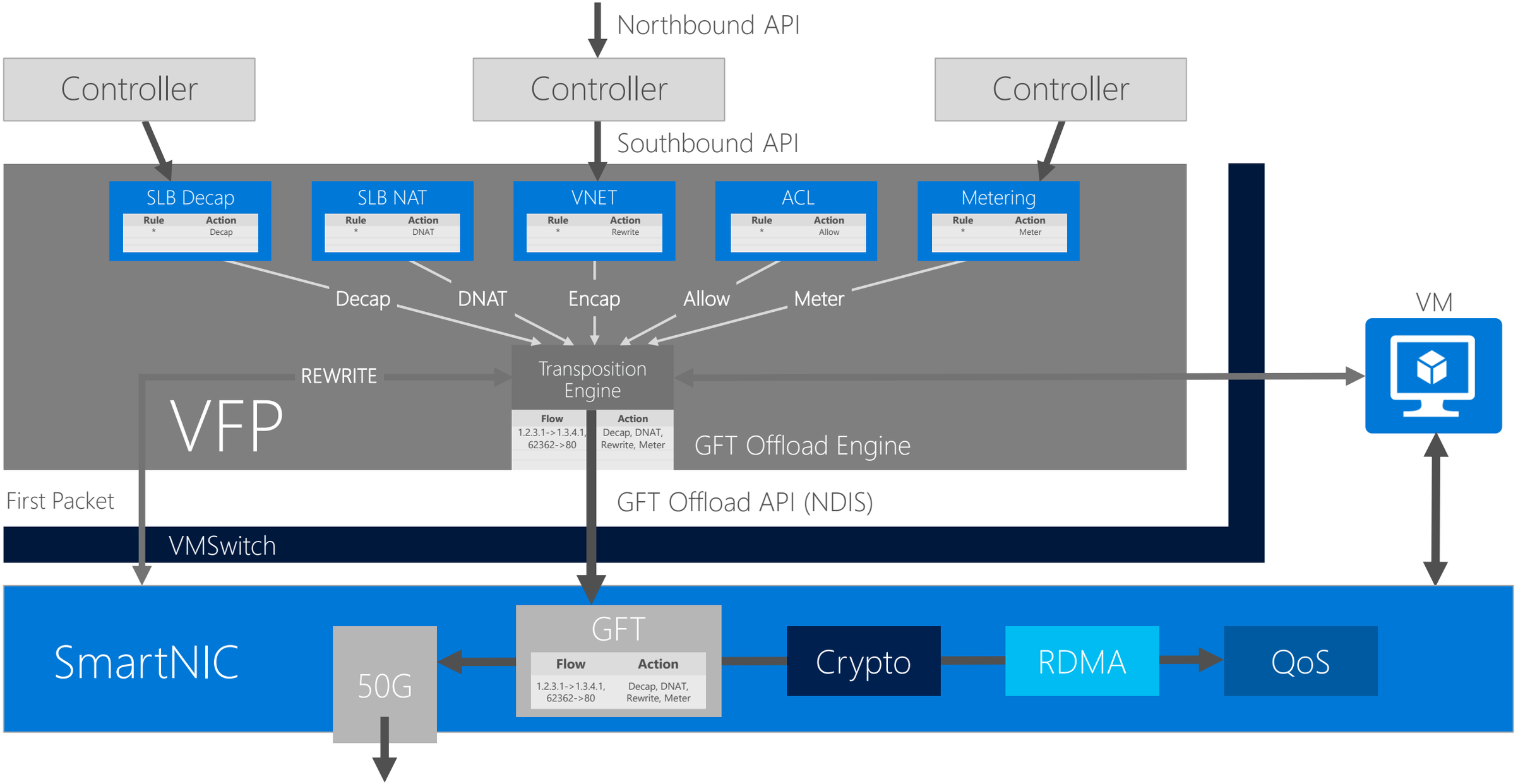
パケットの処理ルールをGFTに書く

2015年後半以降に導入した全サーバーに載っている

暗号化、QoS、圧縮などにも使われている

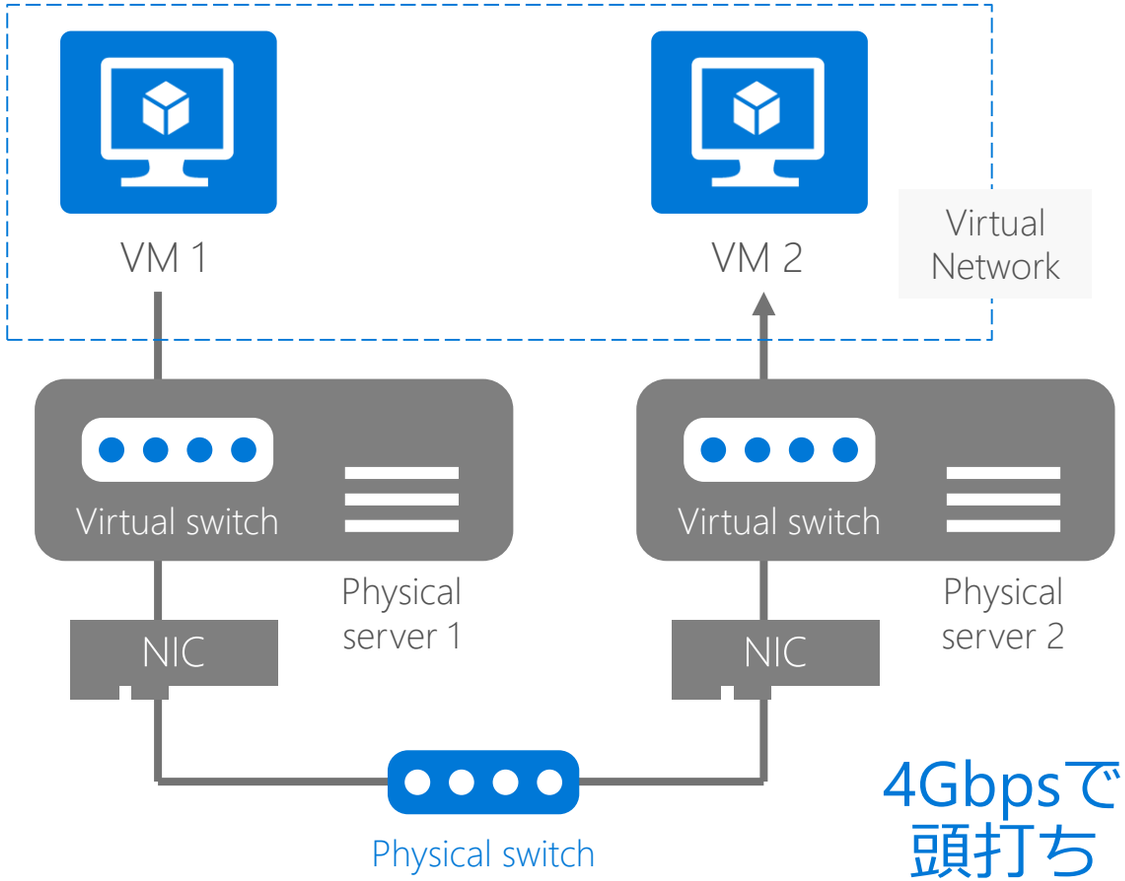


SmartNIC – Accelerating SDN



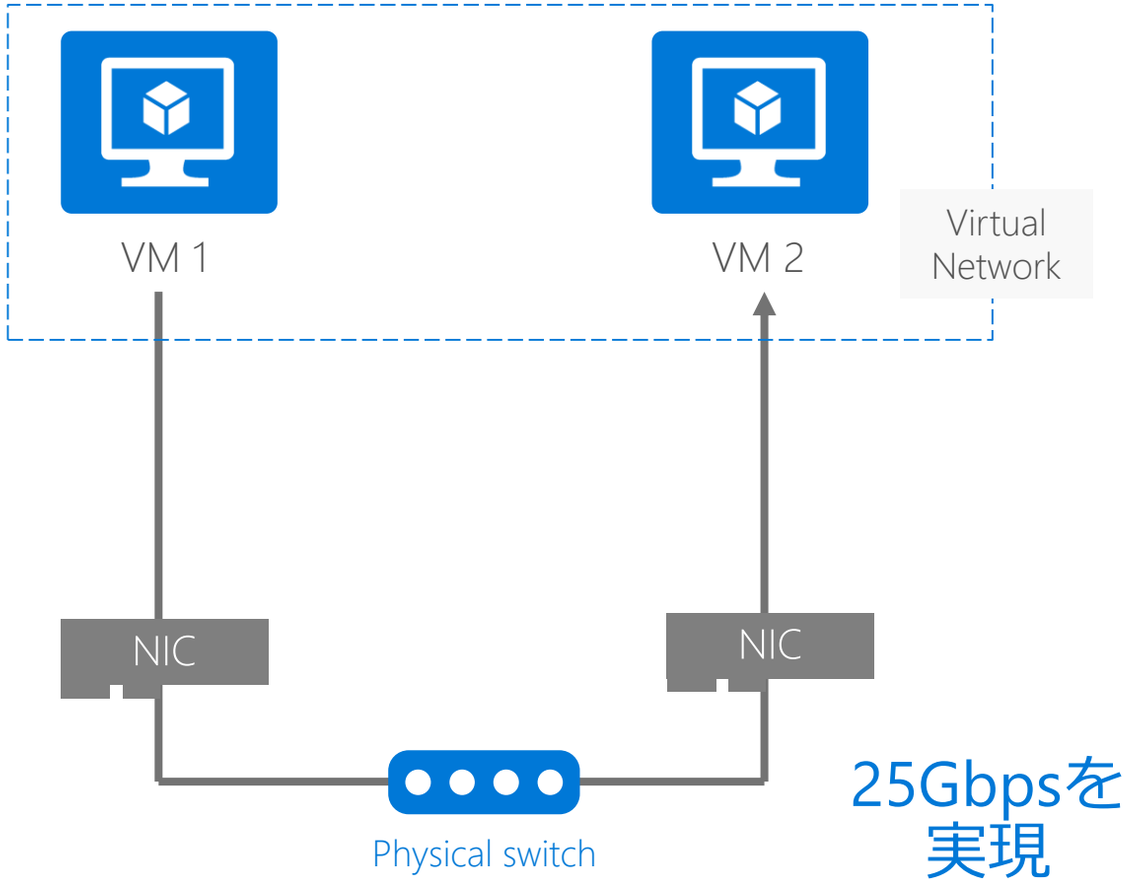
Accelerated networking

従来方式



4Gbpsで頭打ち

Accelerated networking (FPGA & SR-IOV & VFP)

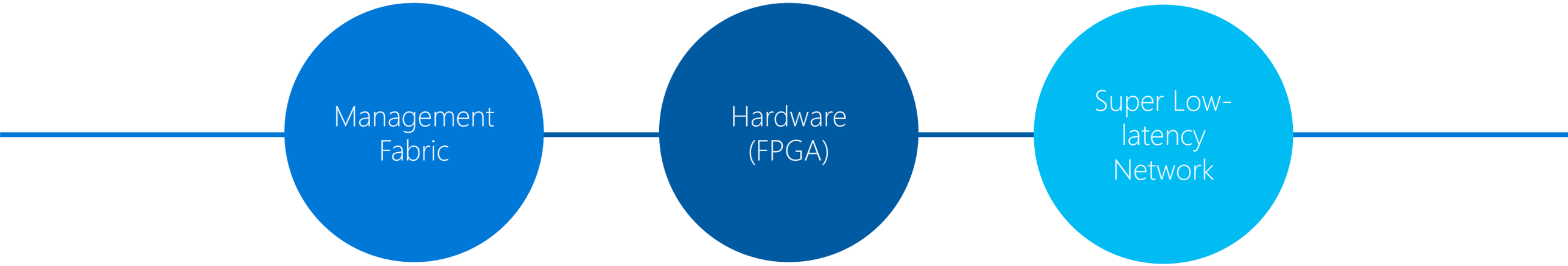


25Gbpsを実現

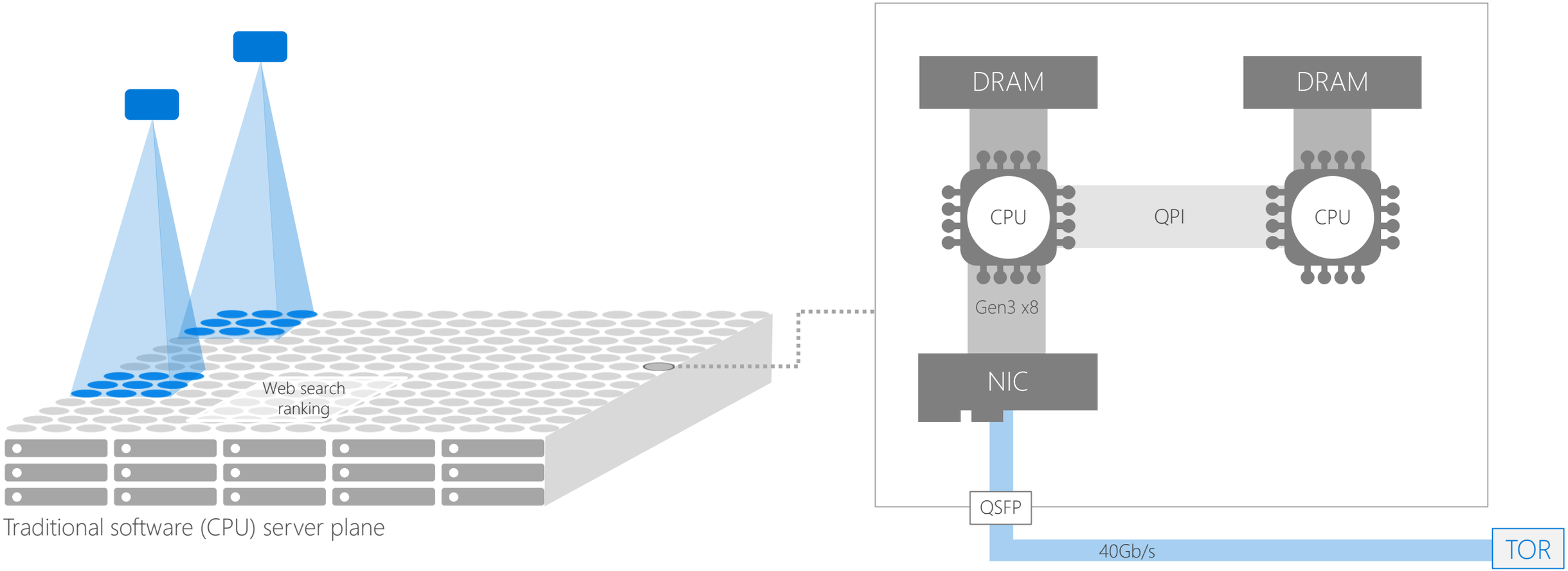
New 50GbE SmartNIC for Project Olympus
(Announced at OCP 2017)

Hardware Microservicesとは?

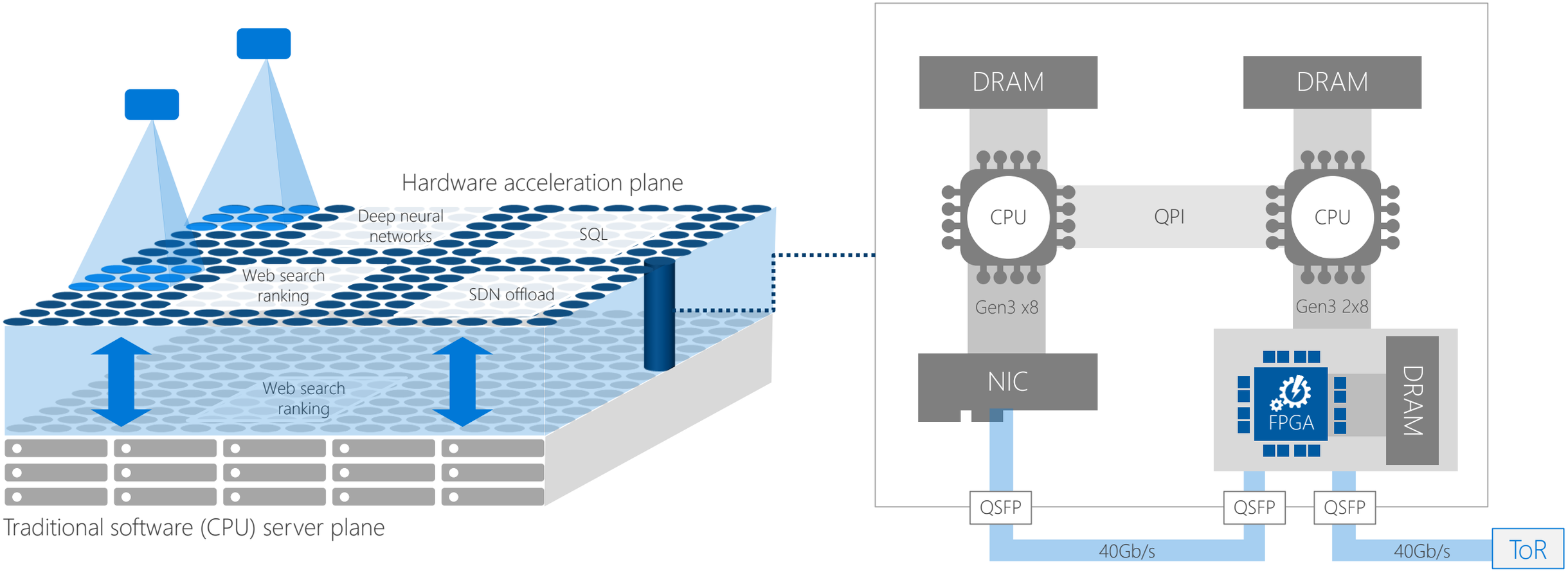
特に性能が求められるサービスをFPGAにのせ、ネットワークでつなぐ。その構成は後から柔軟に変えられる。



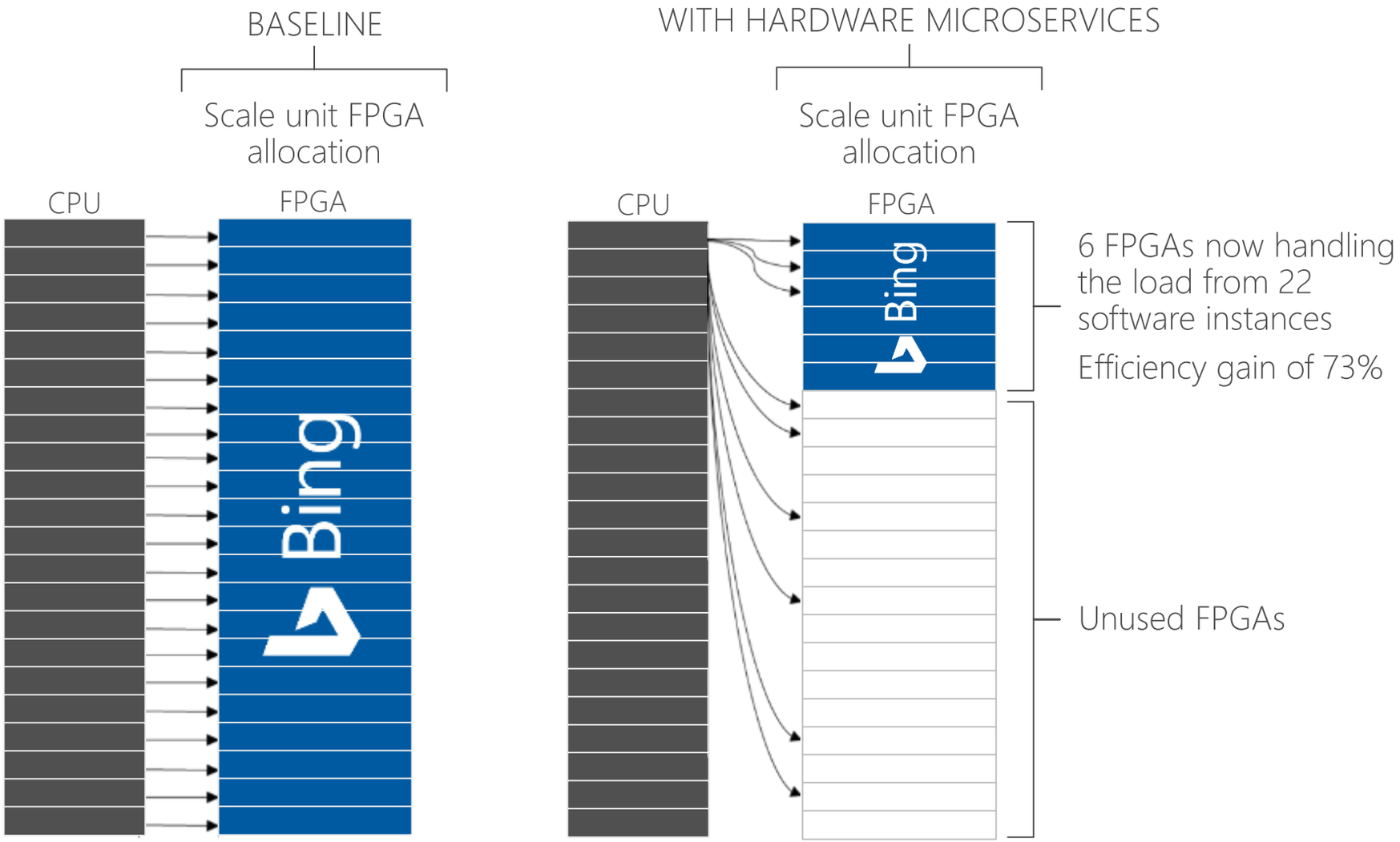
Hardware Microservices適用領域の拡大



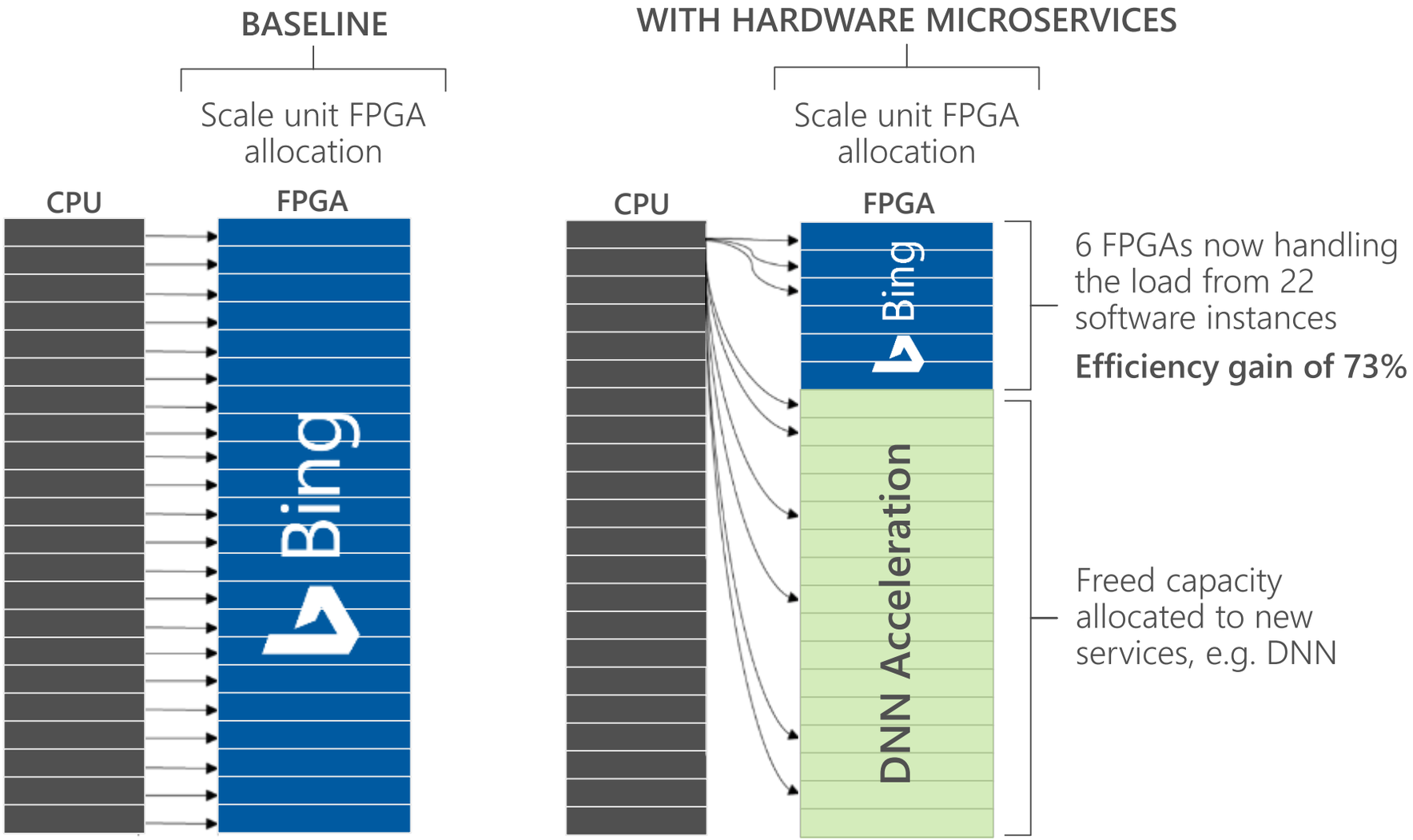
Hardware Microservices適用領域の拡大



再構成を可能にすることで設備効率が向上



例: 空いた領域でDeep Neural Network処理を



© 2017 Microsoft Corporation. All rights reserved.

本情報の内容（添付文書、リンク先などを含む）は、作成日時点でのものであり、予告なく変更される場合があります。