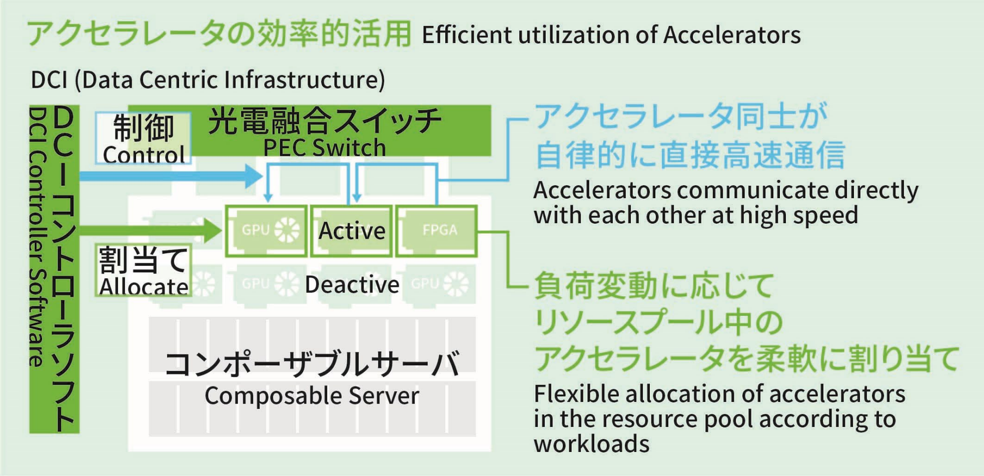
IOWN
INTEGRAL
NTT R&D FORUM 2024

DEVELOPMENT
δ 02-02

Efficient utilization of accelerators and connection technologies with DCI

Power reduction in DCs by efficient utilization of accelerators and low-latency remote connection

#Green Transformation



アクセラレータの効率的活用 Efficient utilization of Accelerators

DCI (Data Centric Infrastructure)

DCIコントローラソフト
DCI Controller Software

制御
Control

光電融合スイッチ
PEC Switch

割当て
Allocate

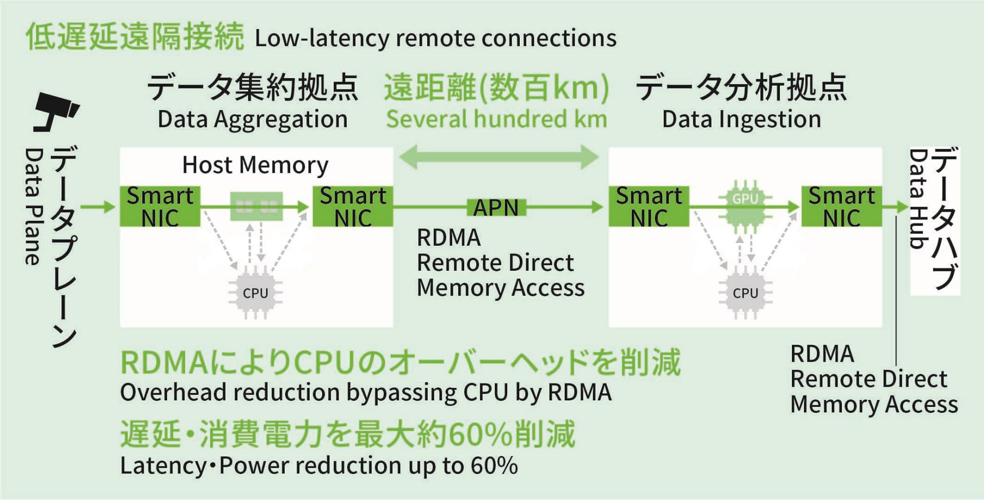
GPU
Active
FPGA

GPU
Deactive
GPU

コンポーザブルサーバ
Composable Server

アクセラレータ同士が
自律的に直接高速通信
Accelerators communicate directly with each other at high speed

負荷変動に応じて
リソースプール中の
アクセラレータを柔軟に割り当て
Flexible allocation of accelerators in the resource pool according to workloads



低遅延遠隔接続 Low-latency remote connections

データプレーン
Data Plane

データ集約拠点
Data Aggregation

遠距離(数百km)
Several hundred km

データ分析拠点
Data Ingestion

データハブ
Data Hub

Host Memory

Smart NIC

Smart NIC

CPU

APN

RDMA
Remote Direct
Memory Access

Smart NIC

GPU

Smart NIC

CPU

RDMA
Remote Direct
Memory Access

RDMAによりCPUのオーバーヘッドを削減
Overhead reduction bypassing CPU by RDMA

遅延・消費電力を最大約60%削減
Latency・Power reduction up to 60%

///Technical Issue

CPU performs GPU control and communication, which consumes a large CPU power and causes communication delays.

///Research Goal

By efficiently using GPUs in remote locations to reduce cost and reduce power consumption by 1/8 compared to 2020.

---Technology

- Low-latency remote communication by RDMA over APN.
- CPU offloading, efficient accelerators connection.
- Controller software which allocates computing resources appropriately according to the situation.

---Novelty

Enables low latency transmission of large size video streaming by RDMA to remote locations. Reduce the number of servers by up to 70% and power consumption by up to 60% in IOWN Global Forum PoC by RDMA NIC and GPU direct transmission.

---Applicable Business

In ICT industry and service industry, use cases for security monitoring and behavior analysis using multiple cameras across various locations, which is required for the realization of smart cities and digital twins. (around 2027) [market size: $18B*]
* 2026 market size forecast for composable infrastructure.

Related Exhibition＝ δ02-04／δ03-02　Exhibitors＝NIPPON TELEGRAPH AND TELEPHONE CORPORATION　Related URL＝NTT Technical Review
Contact URL　Outside Partner＝Fujitsu Limited