A Joint Field Validation of Distributed AI Inference over IOWN-based Distributed Data Centers

A Collaborative Study by Accton × Edgecore × NCHC, NIAR x NTT

Executive Summary

This white paper presents the results of a joint field validation study conducted collaboratively by:

- Accton Technology Corporation ("Accton")
- Edgecore Networks Corporation ("Edgecore")
- National Institutes of Applied Research, Taiwan (hereinafter referred to as "NIAR"), with its executing unit, the National Center for High-Performance Computing (hereinafter referred to as "NCHC"),
- NTT, Inc. ("NTT")

The study focuses on validating distributed AI inference across geographically separated data centers interconnected through IOWN-based All-Photonics Network (APN) infrastructure.

The validation environment spans:

- NTT Musashino R&D Center (Tokyo, Japan)
- NCHC, NIAR Shalun R&D Building (Taiwan)
- NCHC, NIAR Tainan Data Center, STSP (Taiwan)
- Taiwan commercial APN infrastructure environments

This joint effort demonstrates the feasibility of:

- distributed AI inference
- APN-based low-latency interconnection
- distributed GPU resource utilization
- Application and workload deployment
- cross-domain AI infrastructure integration under realistic field-level deployment conditions

Unlike conventional laboratory-only verification, this project validates a practical, multi-party, cross-border architecture involving networking, optical transport, AI infrastructure, orchestration, and operational integration.

The project further demonstrates how IOWN-based infrastructure may enable the evolution, from: “location-bound compute”

into: “location-independent, network-coordinated compute.”

1. Background and Motivation

The rapid growth of AI-driven applications, particularly those involving real-time image analytics, smart city services, industrial automation, and distributed sensor systems, is driving new infrastructure requirements, including:

- ultra-low latency processing
- geographically distributed compute resources
- scalable GPU utilization
- resource orchestration
- high-bandwidth inter-data-center connectivity
- energy-efficient infrastructure architectures

Conventional centralized data center architectures increasingly face limitations in addressing these emerging requirements.

At the same time, the evolution of IOWN (Innovative Optical and Wireless Network), led by NTT, introduces a photonics-centric architecture designed to significantly improve:

- latency performance
- transmission stability
- energy efficiency
- network scalability
- distributed compute coordination

Recognizing the strategic importance of integrating networking, optical transport, distributed AI infrastructure, and orchestration technologies, Accton, Edgecore, CHT, NCHC, NIAR and NTT jointly initiated this collaborative field validation study.

This collaboration also aligns with the ecosystem collaboration framework defined among Accton, Edgecore, CHT, and NCHC, NIAR and NTT regarding IOWN APN computing networks, distributed AI infrastructure, and APN/Data Center Interconnect ecosystem development.

In addition, this study extends the distributed data center validation activities defined in the Joint Experiment Agreement and subsequent amendment among NTT, Accton, and Edgecore, and its subsequent revisions.

2. Collaborative Ecosystem and Roles

This project was executed through a multi-domain collaboration model combining networking, AI infrastructure, optical transport, orchestration, and operational integration expertise.

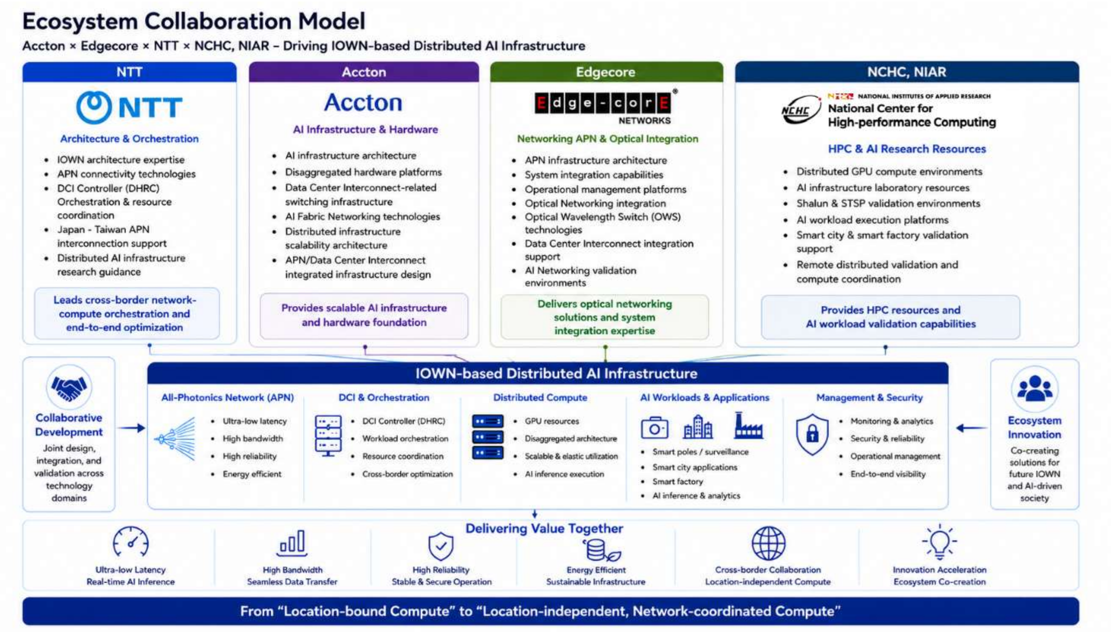

Figure 1. Collaborative Ecosystem and Roles

Accton

Accton contributed:

- AI Infrastructure architecture design
- disaggregated hardware platforms
- DCI-related switching infrastructure
- AI Fabric Networking technologies
- distributed infrastructure scalability architecture
- APN/Data Center Interconnect integrated infrastructure design

Accton additionally contributed architecture validation related to:

- distributed data center feasibility
- GPU resource integration
- AI infrastructure scalability
- future APN/Data Center Interconnect deployment models

Edgecore Networks

Edgecore contributed:

- system integration capabilities
- operational management platforms
- Optical Networking integration
- Optical Wavelength Switch (OWS) technologies
- DCI integration support
- AI Networking validation environments

Edgecore additionally supported:

- interoperability testing
- dynamic optical traffic engineering
- APN/Data Center Interconnect integration validation
- demonstration integration activities

National Center for High-performance Computing, National Institutes of Applied Research (NCHC, NIAR)

NCHC, NIAR contributed:

- distributed GPU compute environments
- AI infrastructure laboratory resources
- Shalun and STSP validation environments
- AI workload execution platforms
- smart city and smart factory validation environments
- distributed inference validation resources

NCHC, NIAR additionally enabled:

- remote distributed validation activities
- geographically separated compute coordination
- practical field-level deployment validation
- AI infrastructure experimentation under operational conditions

The collaboration among infrastructure providers, research organizations, and APN technology contributors demonstrates the importance of ecosystem-level integration across:

- networking
- compute infrastructure
- optical transport
- orchestration

- operational integration
- field deployment coordination

NTT

NTT contributed:

- IOWN architecture expertise
- APN connectivity technologies
- DCI Controller software (“DHRC”)
- orchestration and distributed resource assignment technologies
- Japan-Taiwan APN interconnection support
- distributed AI infrastructure research guidance from past NTT in-house experiment

NTT also led discussions regarding:

- network-compute orchestration
- distributed inference deployment
- cross-border APN operational integration

3. IOWN-based Distributed AI Infrastructure Architecture

IOWN (Innovative Optical and Wireless Network) introduces a photonics-based architecture designed to support next-generation distributed AI infrastructure.

In this project:

- APN was utilized as the interconnection layer between Taiwan and Japan.
- Distributed data centers operated as a logically unified compute environment.
- AI workloads were dynamically distributed across geographically separated compute sites.
- APN/Data Center Interconnect integration enabled low-latency coordination across long-distance connections.

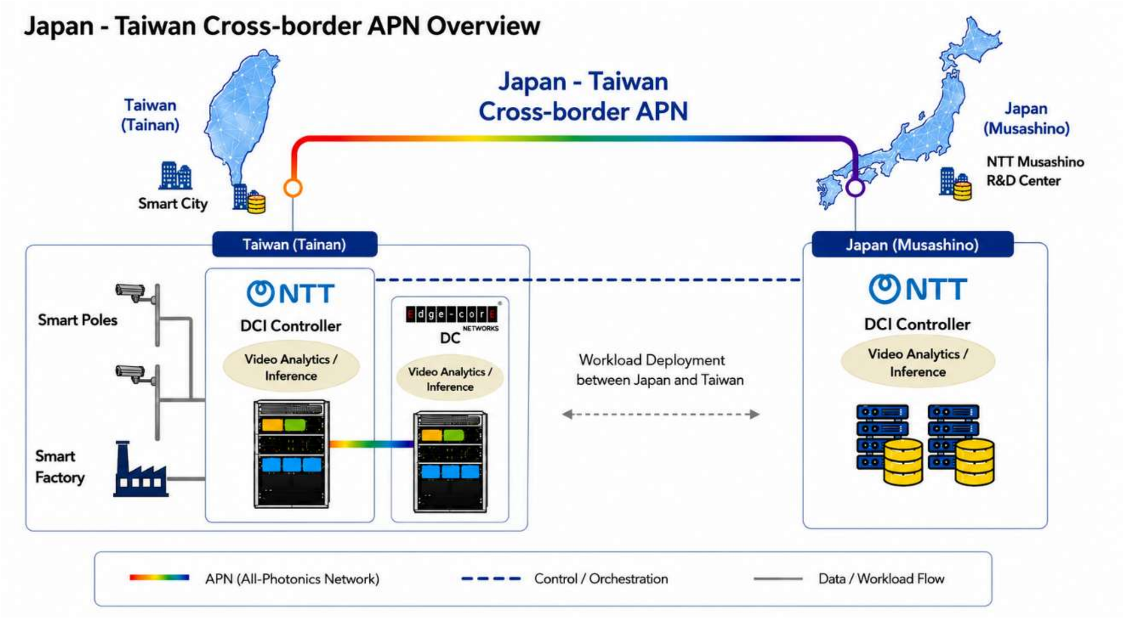

Figure 2. Japan – Taiwan Cross-border APN Overview

The architecture integrated:

- APN connectivity
- DCI infrastructure
- distributed GPU resources
- AI Fabric Networking
- Optical Switching technologies
- orchestration and workload coordination

The field validation environment included:

- Taiwan-Japan APN interconnection
- distributed inference systems
- real-time image analytics workloads
- smart city and smart factory validation scenarios
- GPU workload coordination

The architecture further validated the feasibility of integrating:

- Optical Wavelength Switch (OWS)
- APN/Data Center Interconnect infrastructure
- distributed GPU resource sharing
- network and AI infrastructure coordination under realistic deployment conditions.

4. Demonstration Environment

The project demonstration was conducted as a field-level validation rather than a controlled laboratory-only experiment.

The validation environment included:

Taiwan Sites

- NCHC, NIAR Shalun R&D Building
- NCHC, NIAR Tainan Data Center, Southern Taiwan Science Park (STSP)
- Taiwan APN infrastructure environments
- smart city and smart factory demonstration environments

Japan Site

- NTT Musashino R&D Center in Tokyo

Demonstration Characteristics

The demonstration environment supported:

- distributed image inference
- real-time video streaming workloads
- APN-based low-latency interconnection
- distributed GPU utilization
- DCI orchestration validation
- Optical Networking validation
- cross-border workload coordination

The project also validated:

- interoperability among multi-vendor systems
- AI infrastructure integration across distributed sites
- telecom and compute infrastructure coordination
- network-aware workload distribution

This practical validation approach ensured that the results reflect realistic operational conditions rather than isolated theoretical simulations.

5. System Design and Technical Integration

The distributed image inference system was jointly designed and validated by all participating parties.

Key design principles included:

- separation of inference workloads
- distributed GPU resource utilization
- network-aware workload distribution
- orchestration of compute resources
- APN/Data Center Interconnect coordinated infrastructure integration
- low-latency workload coordination

The validation architecture integrated:

- APN connectivity
- DCI orchestration
- Optical Networking
- distributed compute resources
- AI Fabric Networking
- AI inference systems

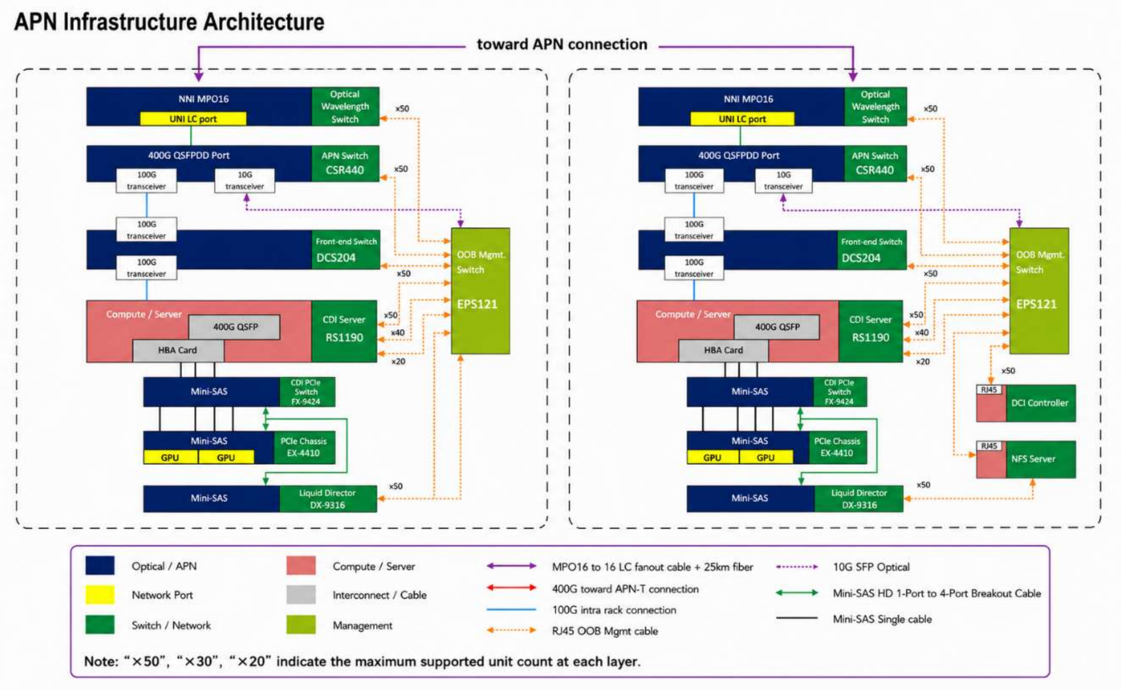

Figure 3. APN Infrastructure Architecture

End-to-End Service Architecture

The APN infrastructure architecture provides the underlying connectivity foundation for distributed AI services. Building upon this infrastructure, an end-to-end architecture was established to support distributed AI inference across geographically separated compute environments.

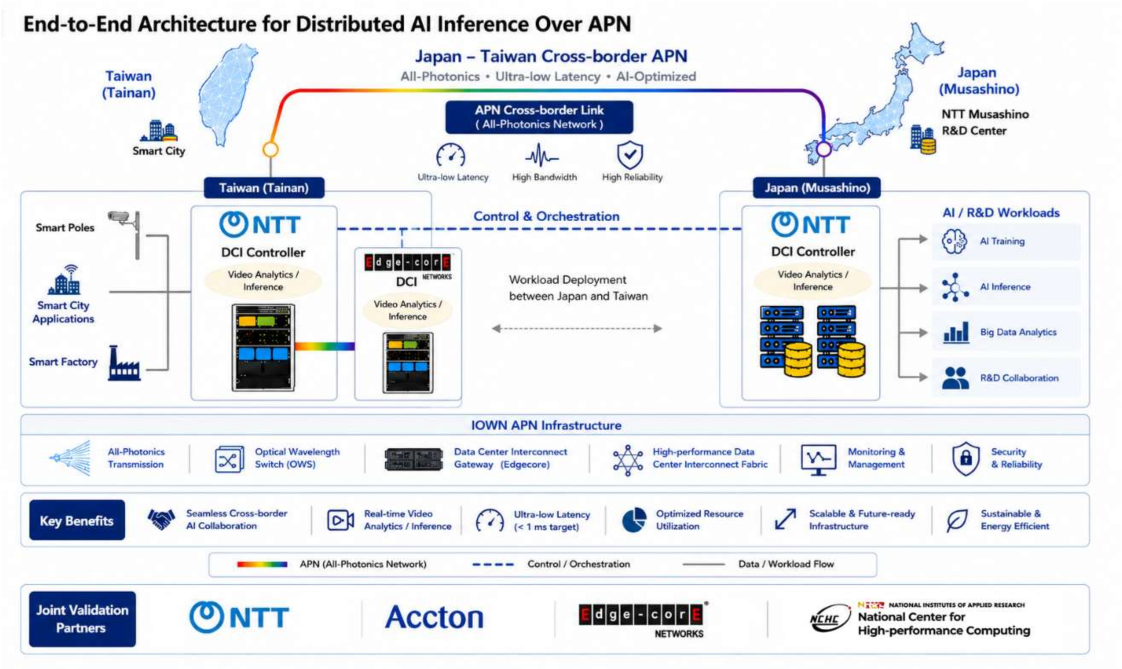

Figure 4. End-to-End Architecture for Distributed AI Inference Over APN

The architecture further evaluated how:

- network latency
- optical transport behavior
- orchestration mechanisms
- Distributed GPU coordination influence distributed inference performance under realistic deployment conditions.

The project additionally validated the feasibility of integrating:

- OWS technologies
- distributed DCI architectures
- telecom infrastructure
- AI orchestration systems within future distributed AI infrastructure environments.

6. Evaluation Methodology

Validation Framework

The Smart Traffic Monitoring demonstration was selected as a practical validation framework rather than a standalone application use case. The objective of the demonstration was not to validate a specific traffic monitoring solution, but to evaluate whether AI services could be dynamically executed across geographically distributed compute resources interconnected through the IOWN All-Photonics Network (APN).

This validation framework enabled the participating organizations to assess cross-region AI inference, distributed GPU utilization, resource disaggregation, low-latency optical networking, and network-centric computing under realistic operational conditions.

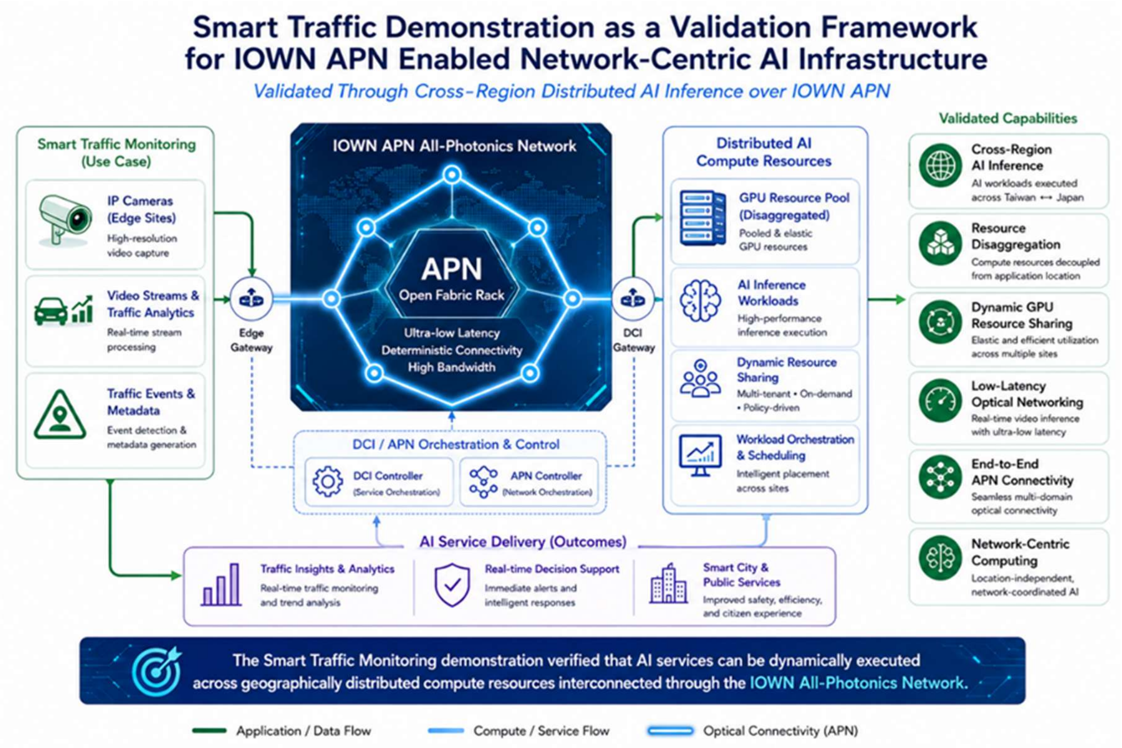

Figure 5. Smart Traffic Demonstration as a Validation Framework for IOWN APN Enabled Network-Centric AI Infrastructure

Validated Capabilities

The Smart Traffic validation framework enabled the participating organizations to verify several key capabilities of IOWN APN-enabled infrastructure, including:

- Cross-Region AI Inference
- Resource Disaggregation
- Dynamic GPU Resource Sharing
- Low-Latency Optical Networking
- End-to-End APN Connectivity
- Network-Centric Computing

The joint validation team defined evaluation metrics aligned with practical deployment scenarios.

Key evaluation items included:

- end-to-end inference latency
- latency stability across long-distance APN connections
- distributed GPU coordination efficiency
- network behavior under continuous workloads
- operational feasibility of distributed GPU utilization
- interoperability between networking and orchestration systems
- Optical Networking effectiveness
- scalability of distributed AI infrastructure

The validation methodology additionally evaluated:

- APN/Data Center Interconnect integration feasibility
- distributed orchestration behavior
- cross-domain operational coordination
- smart city and smart factory application scenarios

The evaluation metrics and methodology were jointly discussed and agreed upon by all participating parties.

7. Validation Results

The field validation demonstrated that:

- distributed AI inference can be executed across geographically separated sites with stable latency performance;
- APN-based connectivity significantly reduces latency variability across long-distance connections;
- distributed GPU resources can be coordinated independently of physical location;
- DCI orchestration can effectively support distributed AI infrastructure environments;
- Optical Networking technologies can improve traffic engineering flexibility and infrastructure scalability.

The validation further demonstrated the feasibility of:

- integrating network infrastructure and AI infrastructure;
- coordinating distributed workloads across Taiwan and Japan;
- enabling network-centric compute architectures;

- supporting scalable distributed AI deployment models.

Importantly, the results reflect the combined effectiveness of:

- networking infrastructure;
- APN connectivity;
- AI infrastructure architecture;
- orchestration technologies;
- Optical Networking integration;
- operational coordination;
- ecosystem collaboration.

No single domain alone would have been sufficient to realize this validation architecture.

8. Strategic Significance of IOWN-based Distributed AI Infrastructure

Open Fabric Deployment Platform

Based on the validation results, the participating organizations further evaluated a deployable infrastructure model capable of supporting future network-centric AI services. The Open Fabric Rack Solution integrates application servers, orchestration systems, APN connectivity, Data Center Interconnect gateways, open networking platforms, and disaggregated compute resources into a modular architecture designed for distributed AI infrastructure deployment.

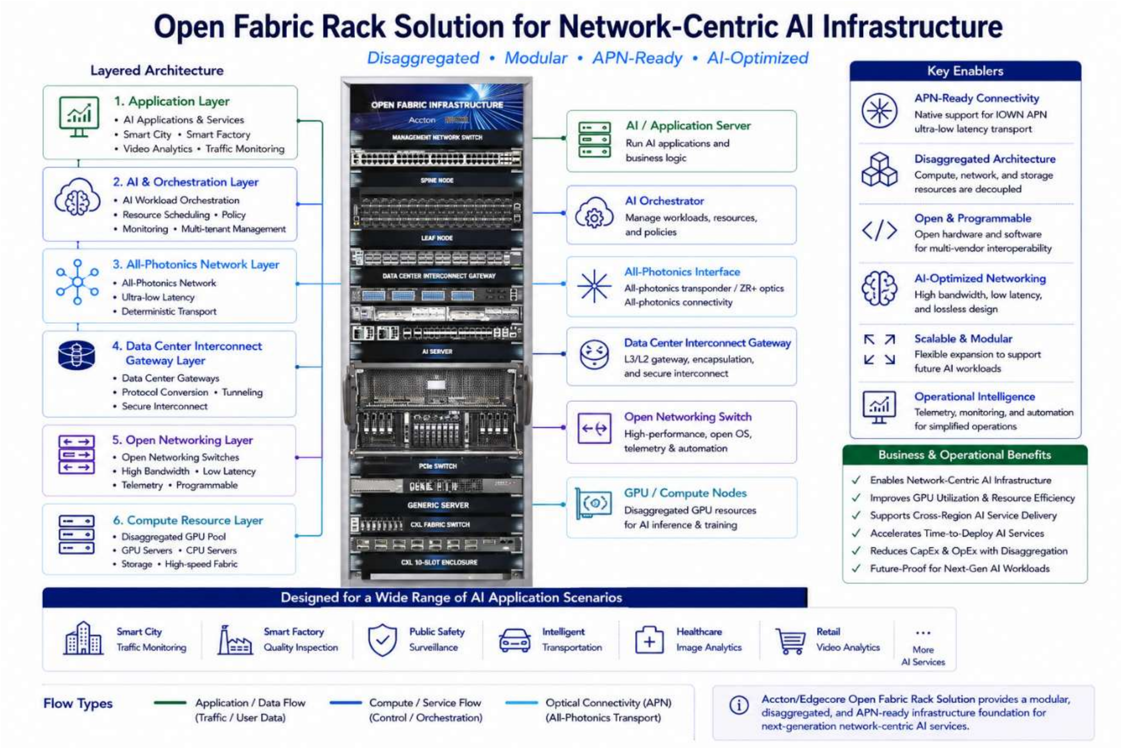

Figure 6. Open Fabric Rack Solution for Network-Centric AI Infrastructure

Infrastructure Transformation

The validation activities and deployment architecture presented in this study demonstrate a transition from traditional device-centric computing toward a network-centric infrastructure model. In this emerging architecture, compute, storage, and networking resources are no longer constrained by physical location and can instead be dynamically orchestrated across geographically distributed environments through the IOWN All-Photonics Network.

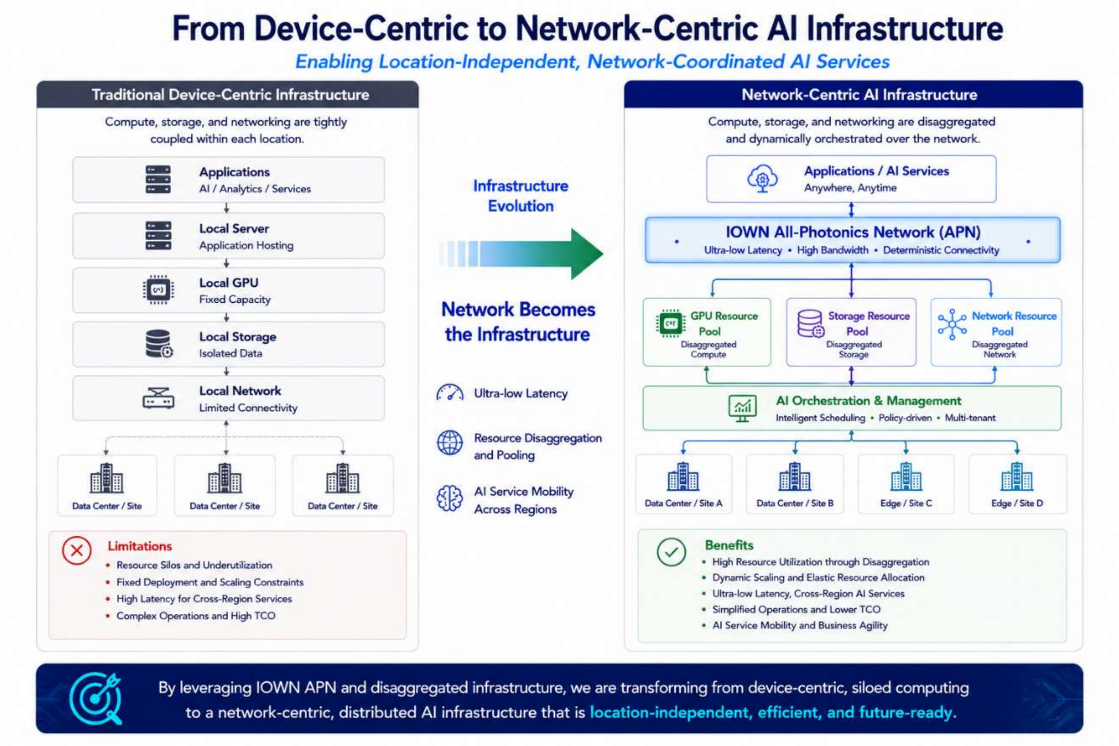

Figure 7. Transformation - From Device-Centric to Network-Centric AI Infrastructure

The transformation illustrated in Figure 7 represents the fundamental architectural shift enabled by IOWN APN, where infrastructure resources become network-coordinated services rather than location-bound assets.

The findings suggest a major architectural shift for future AI infrastructure:

From: Location-bound compute infrastructure
To: Location-independent, network-coordinated AI infrastructure

This project demonstrates how IOWN-based APN infrastructure may enable:

- flexible distributed GPU utilization
- scalable multi-region AI infrastructure
- resource orchestration
- network-cloud-compute integration
- energy-efficient distributed AI systems
- cross-border AI infrastructure coordination

The validation also highlights the importance of:

- multi-party ecosystem collaboration
- network-compute integration

- optical infrastructure innovation
- orchestration interoperability
- distributed infrastructure standardization for future AI-native infrastructure architectures

9. Future Outlook and Ecosystem Expansion

Based on the results of this validation, the participating parties intend to continue discussions regarding:

- expansion of distributed AI use cases;
- smart city and smart factory deployments;
- network-integrated AI infrastructure applications;
- distributed GPU infrastructure deployment;
- APN/Data Center Interconnect ecosystem expansion;
- AI infrastructure orchestration improvements;
- Optical Networking integration;
- future commercialization opportunities.

Potential future directions include:

- AI infrastructure beta-site deployments;
- APN-enabled distributed inference services;
- telecom-integrated AI infrastructure;
- distributed edge AI applications;
- AI infrastructure ecosystem standardization;
- future IOWN Global Forum collaboration activities.

The project also establishes a foundation for future:

- ecosystem collaboration;
- APN/Data Center Interconnect commercialization;
- AI infrastructure deployment;
- international AI infrastructure integration;
- cross-domain operational collaboration.

10. Conclusion

This white paper presented a joint field validation conducted collaboratively by:

- Accton
- Edgecore Networks

- NCHC, NIAR
- NTT

The project successfully demonstrated that:

- IOWN-based APN infrastructure can support distributed AI workloads;
- distributed GPU resources can operate across geographically separated data centers;
- network infrastructure and AI infrastructure can be integrated effectively;
- Optical Networking technologies can enhance distributed AI infrastructure scalability;
- ecosystem-level collaboration is essential for realizing next-generation AI infrastructure.

This validation represents an important step toward:

- network-centric compute architectures;
- distributed AI infrastructure;
- APN/Data Center Interconnect-enabled AI platforms;
- future AI-native infrastructure ecosystems.

The project further demonstrates how collaboration among:

- networking providers;
- telecom operators;
- AI infrastructure vendors;
- orchestration technology providers;
- national HPC organizations;

can accelerate the evolution of globally deployable distributed AI infrastructure.

This white paper is provided for informational purposes only. The content is based on a joint field validation conducted under specific conditions and environments, and may not be representative of all deployment scenarios.

Nothing in this document shall be construed as a commitment, guarantee, or offer to deliver any products, services, or functionalities. All technologies, architectures, and capabilities described are subject to change without notice.

This document contains forward-looking statements regarding future plans, expectations, and potential commercialization opportunities. These statements are based on current assumptions and are subject to risks and uncertainties that could cause actual results to differ materially. The participating organizations undertake no obligation to update or revise any forward-looking statements.

Each participating organization is responsible only for its respective contributions. No single organization assumes responsibility for the entirety of the system or validation results described in this document.

Appendix

Appendix A – System Architecture Diagrams

- APN/Data Center Interconnect topology
- Taiwan-Japan interconnection architecture
- distributed GPU coordination flow
- OWS integration architecture
- VLAN and Optical Networking topology

Appendix B – Hardware and Infrastructure Overview

- Open Fabric Rack Solution overview
- GPU server architecture
- Optical Networking equipment
- AI Fabric Networking configuration
- APN/Data Center Interconnect infrastructure components

Appendix C – Demonstration Scenarios

- Smart Traffic scenario
- Smart Factory scenario
- distributed image inference workflow
- AI workload orchestration flow

Appendix D – Terminology and Abbreviations

• APN (All-Photonics Network)
A next-generation network architecture developed by the NTT Group that introduces photonics-based technologies end-to-end across networks and devices. It enables ultra-low latency, high capacity, high-quality transmission, and significantly reduced power consumption by providing direct optical wavelength paths without repeated electrical conversion.

• IOWN (Innovative Optical and Wireless Network)
A future infrastructure initiative led by the NTT Group that aims to realize a new communication and computing platform based on photonics and advanced

wireless technologies, delivering ultra-low latency, massive capacity, and highly energy-efficient processing.

• DCI (Data-Centric Infrastructure)
An infrastructure paradigm that optimizes the placement, movement, and processing of data across distributed environments, including data centers, edge locations, and networks. It focuses on efficient data utilization, scalability, and integration with high-performance computing and AI workloads.

• Data Center Interconnect

A networking architecture and associated technologies used to interconnect geographically separated data centers, enabling high-capacity, low-latency, and reliable transport of workloads, data, and services across distributed infrastructure environments.

• DHRC (Dynamic Hardware Resource Control)
A software-based control framework that dynamically manages and orchestrates hardware resources—including compute, storage, and network components—across distributed environments. It enables real-time allocation, optimization, and sharing of resources to improve performance, flexibility, and overall system efficiency. (Including future support functions.)

• OWS (Optical Wavelength Switching)
A technology that enables switching and routing of communication signals at the optical wavelength level, allowing high-speed, low-latency data transmission without optical-electrical-optical (OEO) conversion.

• Distributed Inference
A computing approach in which AI inference processing is executed across multiple distributed nodes (e.g., edge, data center, and cloud), enabling improved scalability, reduced latency, and efficient resource utilization for real-time AI applications.

• AI Fabric Networking
A network architecture optimized for AI workloads that provides high-bandwidth, low-latency, and scalable interconnectivity between compute resources such as GPUs, accelerators, storage, and distributed systems.

• GPU Resource Sharing
A mechanism that enables multiple users or workloads to share and utilize GPU resources efficiently across a system or distributed infrastructure, improving utilization rates and reducing overall hardware costs.

NTT

NCHC
NIAR NATIONAL INSTITUTES OF APPLIED RESEARCH
National Center for
High-performance Computing

Accton

Edge-corE®
NETWORKS