

## IOWN<sup>®</sup> APN を活用した全国広域分散 GPU 実証環境「GPU over APN Testbed」を提供開始

NTTドコモビジネス株式会社(旧 NTTコミュニケーションズ株式会社、以下 NTTドコモビジネス)は、低遅延・高速大容量通信を実現するネットワーク IOWN<sup>®</sup> APN<sup>※1</sup>を活用し、全国に分散配置された GPU リソースを1つのプラットフォームのように利用できる全国広域分散 GPU 実証環境「GPU over APN<sup>※2</sup> Testbed」(以下 本実証環境)の提供を開始しました。本実証環境は、札幌、金沢、福岡、大阪、首都圏4ヶ所の計8拠点に GPU を分散配置し、分散 AI 学習や分散 AI 推論、RDMA<sup>※3</sup>を活用した拠点間の大容量データ転送などの先進的な AI ワークロード<sup>※4</sup>を、100Gbps 級の低遅延・高速大容量通信と GPU 環境で検証できるものです。これにより、分散された GPU リソースの活用を『体験できる』共同実証の場としてお客さまに提供します。本実証環境の提供を通じて、NTTグループの AI ネイティブインフラである「AIOWN<sup>※5</sup>」の中核の一つである「AI-Centric ICT プラットフォーム<sup>®</sup>」構想<sup>※6</sup>の実現を推進します。



< 「AI-Centric ICT プラットフォーム<sup>®</sup>」構想のイメージ >

### 1. 背景

生成 AI、データ利活用、画像処理などの分野で GPU クラスタの重要性が高まる中、従来の単一データセンターでは、生成 AI のモデルサイズ増大による処理量の変動やリソース確保の制約、データセンターごとのキャパシティや電力供給の制限に応じた運用が求められるなど、さまざまな課題が存在します。こ

これらの課題に対して、データセンターの分散化が注目されており、各地に点在する大量データの高速な同期および転送が必要となっています。NTT ドコモビジネスは、従来の単一データセンター構成の課題に対する新たな取り組みとして、IOWN<sup>®</sup> APN を活用し、地理的に離れた場所に配置された GPU を遅延無く一体的に利用する GPU クラスタの実証実験を進めてきました。

## 2. 全国広域分散 GPU 実証環境「GPU over APN Testbed」について

本実証環境は、広域ネットワーク環境下における分散 GPU 活用の実用性を検証するための実証環境です。IOWN<sup>®</sup> APN による低遅延・大容量通信を活用することで、地理的に離れた拠点間においても GPU が一体的に利用可能となり、分散型の AI 学習や推論、離れた拠点間でのデータ処理に関する検証を実際の環境で行うことができます。本実証環境の提供開始に先行して、複数の企業と連携し、実際のユースケースにもとづく検証も進めています。

本実証環境の主な特長は以下の通りです。

### (1) IOWN<sup>®</sup> APN を活用した全国広域接続

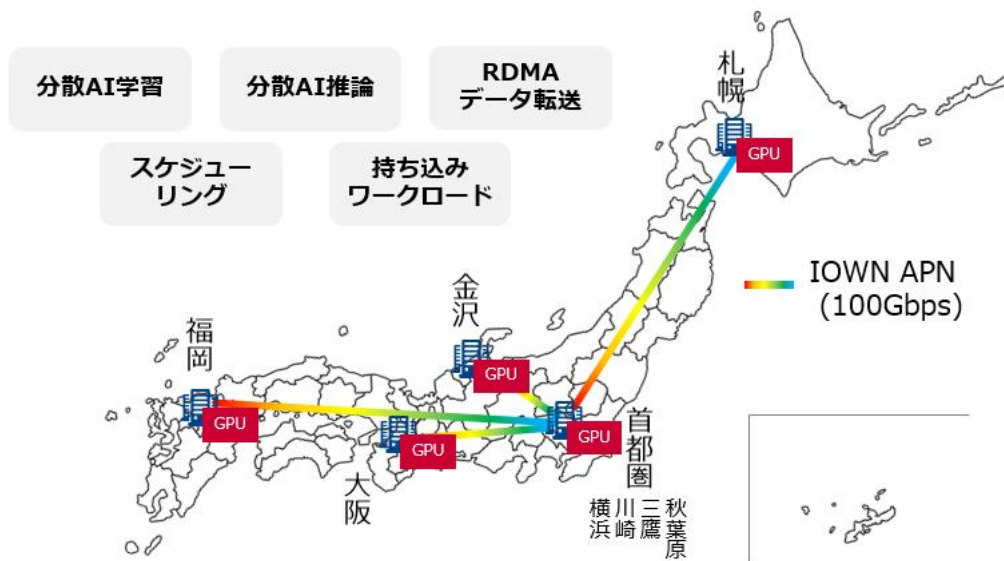
IOWN<sup>®</sup> APN による 100Gbps 級の低遅延・大容量通信を活用し、全国に分散した GPU を 1 つのように連携させることが可能です。従来、距離やネットワークの遅延が制約となっていた、GPU 間の連携において、IOWN<sup>®</sup> APN を用いることで、地理的な距離を意識することなく GPU を活用できる環境を実現しています。

### (2) 広域かつ柔軟な GPU 利用環境

本実証環境では、地理的に離れた札幌、金沢、福岡、大阪、首都圏 4ヶ所の計 8 拠点において、GPU を分散配置し、全国広域にまたがる GPU 利用環境を実現しています。これにより、特定拠点へのリソース集中を避けながら、地域性や設置環境の違いを踏まえた分散 GPU 活用の有効性を検証できます。また、GPU は仮想マシン<sup>※7</sup>および Kubernetes<sup>※8</sup>に対応し、マルチテナント<sup>※9</sup>で提供が可能なため、複数の企業やユーザーが 1 つのハードウェアを効率的に利用できる点も特長です。

### (3) AI ワークロードの技術検証

NTT ドコモビジネスがこれまで取り組んできた実証テーマである、分散 AI 学習、分散 AI 推論、RDMA を活用した拠点間の大容量データ転送などの技術検証についても、実際のネットワークや GPU 環境で検証可能です。また、実際の運用を想定した環境構成を用意することで、分散型の AI 基盤における性能の確認にとどまらず、システム運用面や構成設計上の課題についても検証を行うことができます。これにより、お客さまのシステム環境への将来的な展開を見据え、実践的な分散型の AI ワークロード検証の場を提供します。



<「GPU over APN testbed」のイメージ図>

### 3. お申し込み方法とご利用料金

NTT ドコモビジネス営業担当または本件に関するお問い合わせ先までお問い合わせください。

### 4. 今後の展開

本実証環境の提供を通じて得られた知見や技術的成果を活かし、今後は GPU サービスへの展開を進めていきます。また、お客さまやパートナー企業とも連携しながら、本実証環境を活用したユースケースの創出を継続的に推進します。これにより、分散 AI 基盤の社会実装に向けた課題整理と知見の蓄積を進めるとともに、「AI-Centric ICT プラットフォーム<sup>®</sup>」構想の実現に向けて活用していきます。加えて、NTTグループにおける AI ネイティブインフラ「AIOWN」の展開を推進していきます。

「NTTコミュニケーションズ株式会社」は2025年7月1日に社名を「NTTドコモビジネス株式会社」に変更しました。私たちは、企業と地域が持続的に成長できる自律・分散・協調型社会を支える「産業・地域DXのプラットフォーム<sup>®</sup>」として、新たな価値を生み出し、豊かな社会の実現をめざします。

つながり。驚きを。幸せを。

NTT docomo Business

<https://www.ntt.com/about-us/nttdocomobusiness.html>

※1：IOWN<sup>®</sup> APN とは、IOWN<sup>®</sup>を構成する主要な技術分野の1つとして、端末からネットワークまで、すべてにフォトニクスベース(光)の技術を導入し、エンド・ツー・エンドでの光波長パスを提供する波長ネットワークにより、圧倒的な低消費電力、高速大容量、低遅延伝送の実現をめざすものです。「IOWN<sup>®</sup>」は、NTT株式会社の商標または登録商標です。

※2 : GPU over APN とは、IOWN<sup>®</sup> APN で複数のデータセンターを接続し、1 つの GPU クラスタを稼働させ生成 AI を学習させた実証です。

ニュース 2024 年 10 月 7 日:NTT Com、IOWN APN を活用した分散データセンターでの生成 AI 学習実証実験に世界で初めて成功

※3 : RDMA(Remote Direct Memory Access)とは、サーバー間で CPU 負荷を抑えながらメモリー同士を直接高速にデータ転送する技術であり、低遅延・高スループット通信を実現します。

※4 : AI ワークロードとは、AI を動かすために必要な処理や計算負荷のことです。

※5 : AIOWN とは、お客さまの AI の利用用途に応じて必要な GPU、ネットワーク、電力などのリソースを最適化し、エッジまでを含めたセキュアな利用環境と統合的なオペレーションを実現する NTT グループの AI ネイティブインフラです。

AI 活用の進展に合わせたリソース最適化・オペレーションを実現する AI ネイティブインフラ「AIOWN」の展開 | ニュースリリース | NTT

※6 : AI-Centric ICT プラットフォーム<sup>®</sup>とは、企業が AI を活用して、生産性の抜本的改善、競争力強化やビジネスモデル変革を進める AI 時代に最適な次世代 ICT プラットフォームのことです。

ニュース 2025 年 6 月 10 日:NTT Com、ゲットワークス、NTTPC が戦略的業務提携を締結

※7 : 仮想マシンとは、1 台の物理サーバー上にソフトウェアで複数の独立したコンピューター環境を作り出し、それぞれを別々のサーバーのように利用できる仕組みです。

※8 : Kubernetes とは、アプリケーションを複数のサーバー上に効率的に配置し、自動的に運用・管理するためのオープンソースソフトウェアです。

※9 : マルチテナントとは、1 つのシステムやサービスを複数の利用者(企業・組織)が共用しつつ、データや設定を分離して安全に利用できる仕組みです。