2026 年 3 月 30 日

『IOWN APN』を活用した遠隔環境における GPU・ストレージ間 接続性能の実証と結果

GMO インターネット株式会社
NTT 東日本株式会社
NTT 西日本株式会社
株式会社 QTnet

1.実証の概要

1.1 背景と目的

近年の生成 AI や大規模言語モデル（LLM）の普及により、AI 開発基盤への需要が急激に拡大している。従来、AI の演算装置（GPU）と大容量ストレージは物理的な隣接配置が必須とされてきたが、データセンター内の設置スペース制約や、データを自社施設・自組織の管理下に置いたまま国内クラウドの GPU リソースを活用したいという多様なニーズに対応するため、地理的制約を超えた遠隔分散型 AI インフラの実現が求められている。
本実証では、NTT が開発する次世代通信基盤「IOWN（Innovative Optical and Wireless Network）APN（All-Photonics Network）」の高速大容量かつ低遅延性を活用し、GPU とストレージ間の遠隔利用における技術的実現可能性の実証として、東京-福岡で GMO GPU クラウドの性能を評価した。

1.2 各社の役割

GMO インターネット株式会社	GMO GPU クラウドの GPU、およびストレージの提供 アプリケーション実装 データセンター内の実証環境の提供（東京都渋谷区）
NTT 東日本株式会社	IOWN APN 技術提供および実証回線の提供
NTT 西日本株式会社	IOWN APN 技術提供および実証回線の提供
株式会社 QTnet	データセンター内の実証環境の提供（福岡県福岡市）

1.3 検証スケジュール

事前検証：疑似遠隔環境での性能評価（2025 年 7 月実施済み）
本実証：実拠点間での接続検証（2025 年 11 月-2026 年 2 月）

2. 検証環境・構成

2.1 サーバー環境物理構成

- 場所 1 ：QTnet データセンター（福岡県福岡市）
  - GPU：NVIDIA HGX H100
  - ストレージ：DDN AI400X2
  - ネットワークスイッチ：Arista 7050SX3-48YC8
- 場所 2 ：GMO インターネットグループ グループ第 2 本社 渋谷フクラス サーバールーム（東京都渋谷区）

- ストレージ：DDN AI400X2
- ネットワークスイッチ：Arista 7050SX3-48YC8

2.2 ネットワーク環境構成

- 回線：All-Photonics Connect powered by IOWN（100GbE）

2.3 環境の構築手法

福岡県福岡市内のデータセンター内に GPU サーバー(NVIDIA HGX H100)、渋谷フクラスにストレージ（DDN AI400X2）を設置し、All-Photonics Connect powered by IOWN（100GbE）で接続。
また、比較対象として、福岡県福岡市内のデータセンター内に同様のストレージ（DDN AI400X2）を設置し、GPU サーバー(NVIDIA HGX H100)と接続。
これにより、IOWN APN 回線を挟んだストレージの性能測定と福岡データセンター内の性能測定の環境を構築した。

3. 検証シナリオ

3.1 テストワークロード

本実証では、AI 開発における代表的な画像分類と言語学習およびストレージシステムの性能測定を実施

3.2 画像分類タスク：MLPerf® Training Round 4.0 ResNet（※1 以下 ResNet と表記）

- ベンチマーク：ResNet（Residual Neural Network）
- 特徴：ImageNet データセット（約 128 万枚の学習用イメージを内包）の読み込みと処理を実行
- 評価指標：特定ステップ数の実行を完了するまでの学習時間

3.3 大規模言語モデル学習タスク：MLPerf® Training Round 4.1 Llama2 70B（※2 以下 Llama2 と表記）

- ベンチマーク：Llama（Large Language Model Meta AI）2 70B
- 特徴：Llama2 70B モデル本体（約 130GB）に対する学習を実行
- 評価指標：特定ステップ数の実行を完了するまでの学習時間

※本稿で記載している MLPerf 結果は非公式（Unverified）であり、MLCommons Associations に提出し、審査・承認を受けた公式結果ではありません。

4. 実証結果

4.1 ResNet 画像分類タスクの結果

遅延条件	ベンチマークスコア（分）（※1）
ローカル環境	13.72 分
遠隔環境（IOWN 経由）(13.26ms)	14.38 分
(参考)事前実証環境（15ms）	15.55 分

※ Result not verified by MLCommons Association.

4.2 Llama2 大規模言語学習タスクの結果

遅延条件	ベンチマークスコア（分）(※2)
ローカル環境	24.87 分
遠隔環境（IOWN 経由）(13.26ms)	24.99 分
(参考)事前実証環境（15ms）	24.94 分

※ Result not verified by MLCommons Association.

5. 考察・分析

5.1 性能影響分析

本実証（Phase2）では、事前検証で用いた「疑似遅延付与」に対し、実拠点間での IOWN 回線を用いて、遠隔ストレージ利用時の実運用相当の性能影響を確認した。結果として、ResNet および Llama2 70B の双方でスコアの傾向は事前検証の 15ms 条件の場合と近値であり、IOWN 回線において「意図した遅延条件での性能」を再現できていると判断した。遅延によるスコアの変動の傾向も事前検証と同様であり（後述する事由から）Llama2 70B のほうが、遅延による影響度合が少ない。

5.2 ResNet 画像分類タスク

事前実証でも確認したとおり、ベンチマークが測定開始されてから GPU メモリへ ImageNet データセットの読み込みが行われることから、遅延条件に影響を受けやすく、性能がローカル未満かつ事前実証以上という想定通りの結果となった。ただし、読み込むデータセットは事前に生データ（約 128 万枚の学習用イメージ）を扱いやすい単一のファイル形式へと整形しているため（大ブロックの I/O となり）、性能の落ち込みが軽微なパターンであったと考えられる。

5.3 Llama 大規模言語モデル処理タスク

事前検証でも確認した通り、測定開始される前に GPU メモリへの大規模言語モデルの読み込みが完了しているため、測定開始後は主に GPU 上の演算によって完結する処理が多く、ストレージへの I/O は ResNet 画像分類タスクに比べて少量であることからベンチマークスコアの低下度合いも極めて少なかった。

5.4 まとめ

本実証では画像分類タスクおよび大規模言語モデル処理タスクを対象として事前実証と同様の結果を測定した。事前実証で確認した通り、大きなファイルの読み込みが主となる学習では、その落ち込みは軽微であった。このことから、遠隔ストレージを介した機械学習では遠隔に存在する既存の学習データやあらかじめ整形したデータセットを GPU に読み出す利用形態をとることで、遠隔ストレージを介した機械学習においても、そのメリットを十分に享受することが可能である。本実証にあたり、株式会社データダイレクト・ネットワークス・ジャパン様より、GMO インターネットが所有している DDN AI400X2 と同一モデルを渋谷フクラス側に設置する機器としてご提供いただきました。ここに深く感謝申し上げます。

6. 将来的な社会実装ビジョン

従来、計算資源とデータ（ストレージ）は同一（データセンター）に配置されることが一般的であり、クラウドサービスを利用する等 遠隔地に存在する計算資源を利用する場合はクラウドサービス向けにファイルをコピーして持ち出すことが一般的であった。しかし、IOWN APN 回線を介して計算資源からストレージを遠隔で利用し、データを読

み出すことによってこれらの作業は不要となる。

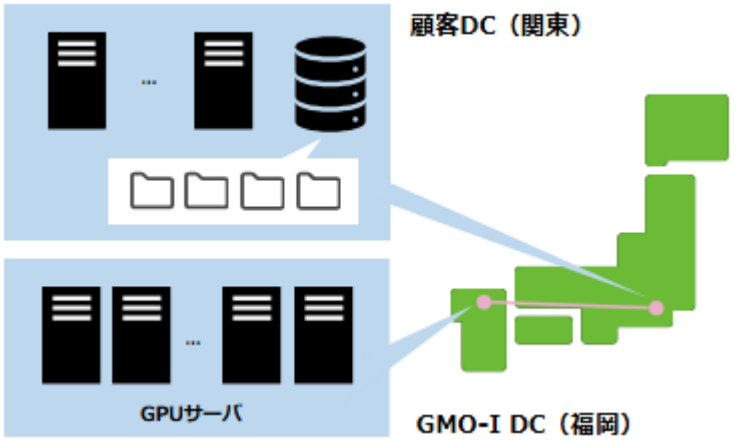
本実証で得られた成果により「計算資源とデータの分離によって生じる課題」を解決できる。

- データ転送時間の削減
- データ重複管理の排除
- 柔軟な計算資源の選定

なお、実際の適用にあたっては、GPU とストレージ間の距離やネットワーク構成等の個別の条件により性能が変動する可能性があるため、ユースケースごとに適用可否を検討していく必要があります。

また、広くあまねく IOWN APN（NTT 東日本・NTT 西日本の「All-Photonics Connect powered by IOWN」）を展開することで、社会的な観点からは、以下のような貢献が期待されます。

1. 分散型 AI 開発基盤の実現：既存のオンプレミス環境とクラウドのハイブリッド活用による全国規模での AI リソース最適配置
2. 災害耐性の向上：分散配置による事業継続性確保



以上

※1 Unverified MLPerf® Training Round 4.0 Closed Resnet offline. Result not verified by MLCommons Association.

※2 Unverified MLPerf® Training Round 4.1 Closed Llama2 70B offline. Result not verified by MLCommons Association.The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See www.mlcommons.org for more information.”

本資料に記載している情報および実証結果は発表日時点のものです。本内容は特定の検証環境下において得られたものであり、いかなる環境においても同等の性能・結果を保証するものではありません。