

# Fujitsu Server PRIMERGY CDI GPU性能スケーラビリティについて

富士通は従来のサーバー製品とは一線を画す新シリーズ「PRIMERGY CDI」を発表しました。この製品は、計算サーバー、PCIe fabric switch、PCIe boxから構成されています。GPU、SSD、NICなどのデバイスは計算サーバーの筐体内ではなく外部のPCIe boxに収納します。

PRIMERGY CDIの最大の特徴は、PCIe box内のデバイスを複数の計算サーバーへ自由に割り当てられることです。これにより、ワークロードの負荷に応じて、デバイスをどのサーバーが使用するのかを自由に切り替えることが可能となります。

これまで発表したホワイトペーパーでは、GPU間通信の性能、プロファイラ[1]を用いた分析結果、MLPerf[2]のルールに従って取得したベンチマークの結果について説明しました。

本ホワイトペーパーでは、ResNetベンチマークプログラムにおいて、スループットがGPUの数に応じて向上していることを示します。本書の構成は以下の通りです。

- 測定に使用したPRIMERGY CDIの構成
- それぞれのGPUの数における性能
- まとめ

[1] <https://developer.nvidia.com/nsight-systems>

[2] MLPerf™ name and logo are trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.

---

本資料を第三者に転送したり、本資料記載の内容をWebサイトへアップしたりするなど、情報の再配信はお断りします。

著作権は富士通株式会社、またはその情報提供者に帰属するため、記載内容を許可なく転載することを禁じます。

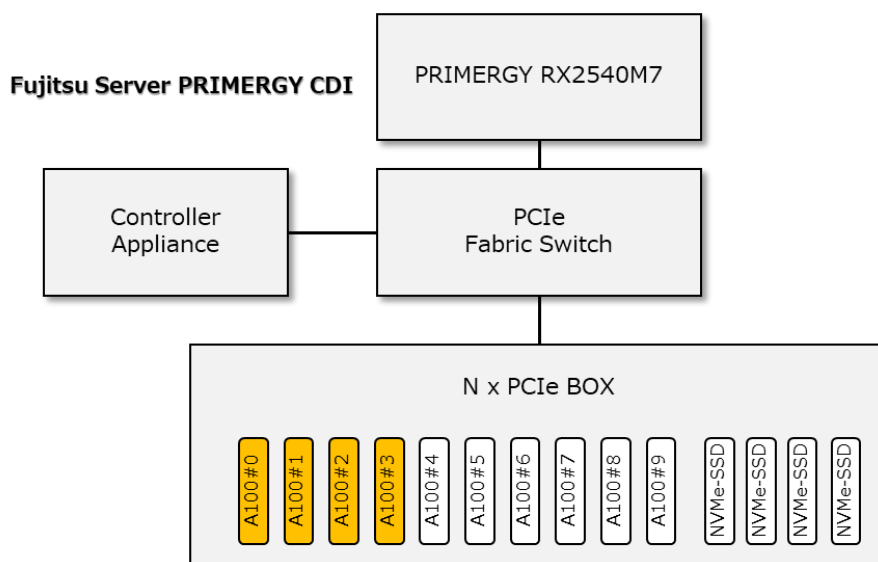
---

# Fujitsu Server PRIMERGY CDI GPU性能スケーラビリティについて

## 1. システム構成

### Block Diagram

測定に使用したPRIMERGY CDIの構成は以下の図のとおりです。計算ノードとしてRX2540M7を使用し、PCIe fabric switch、PCIe box、controller applianceを使用します。PCIe boxには10枚のGPU (NVIDIA® A100 PCIe 80GB) および4枚のNVMe-SSDを搭載します。測定では使用するGPUの数を2, 4, 8, 10にした場合でベンチマークプログラムを動作させてスコア並びにログを取得します。図では4枚のGPUを使用する場合を示しています。



### システム仕様

<b>Server</b>	Fujitsu Server PRIMERGY CDI
<b>CPU</b>	Intel(R) Xeon(R) Gold 6430x2
Frequency	2.1GHz
Core Count	32
<b>Memory</b>	16GBx16
<b>Storage</b>	8x 800GB NVMe SSD
<b>Interconnect</b>	PCIe 4.0
<b>GPU</b>	NVIDIA A100-PCIe-80GBx2~10
<b>OS</b>	Red Hat Enterprise Linux release 8.6 (Ootpa)
<b>Software</b>	CUDA: 12.1.0.023 cuda_driver_version: 530.30.02
<b>HBA</b>	PCIe HBA Card for CDI (Bandwidth 64GB/s (Bidirectional) )
<b>PCIe Fabric Switch</b>	PCIe Fabric Switch (48port) for CDI x1 (Total Bandwidth 768 GB/s Bidirectional 48 port)
<b>PCIe BOX</b>	PCIe Box for CDI xN (Maximum Port Bandwidth 128GB/s (Bidirectional) )
<b>Director</b>	Controller Appliance for CDI

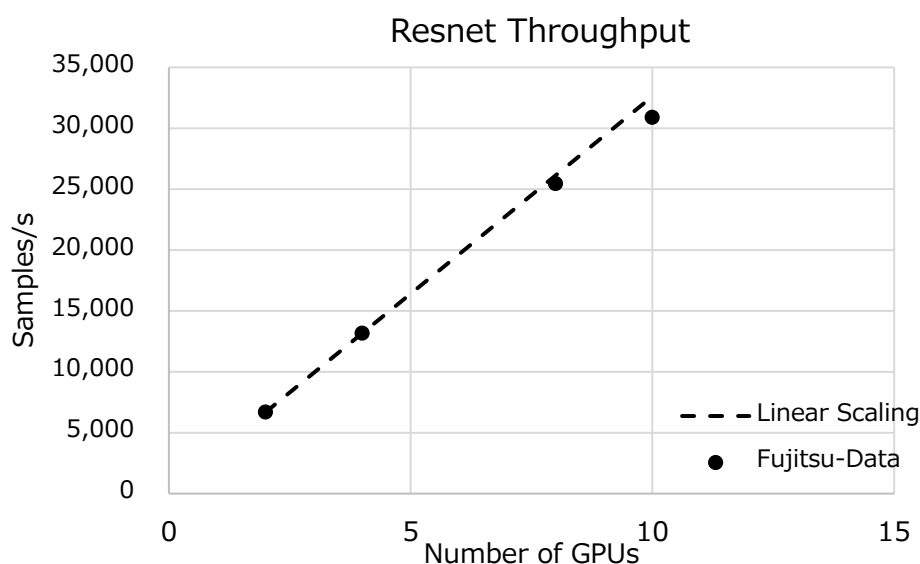
# Fujitsu Server PRIMERGY CDI GPU性能スケーラビリティについて

## 2. 性能スケール結果

使用するGPUの数を変えてResNetを学習させたとき、実際に学習を行った時間 (Training time)、学習が完了するまでに処理した画像の枚数 (Samples to train)、スループット (Throughput) を以下の表に示します。使用するGPUの数が変わると計算内容も変わるため、処理する画像の枚数も変化します。また、スループットは1秒間に処理する画像の枚数であり、この値は処理した画像の枚数を学習に要した時間で割ることで得られます。

GPU	Training time(sec)	Samples to train	Throughput (Samples/s)
2	6890.177	46,122,012	6,694
4	3501.386	46,122,012	13,173
8	1761.707	44,840,845	25,453
10	1451.635	44,840,845	30,890

使用したGPUの枚数に対するスループットの変化を以下のグラフに示します。グラフの黒丸は実測値、破線は2GPUのスループットと4GPUのスループットとを結ぶ直線で、性能がGPUの数に完全に比例した場合の理想的なスループットとなっています。この条件に基づいて10GPUの理想的なスループットと実測値との差を計算すると、実測値は94.73%となります。そのため、ResNetの学習時の性能はGPUの数に対して性能を出せているといえます。



# Fujitsu Server PRIMERGY CDI GPU性能スケーラビリティについて

## 3. まとめ

- 本ホワイトペーパーでは富士通がMLPerf Training™ v3.1に投稿した結果から学習時のスループットを算出し、使用するGPUの数に応じて性能が向上することを確認しました。
- PRIMERGY CDIは最大20枚のGPUを搭載可能です。比較的安価なGPUを多数搭載することで、学習用途にも利用可能です。
- 導入当初は少ないGPUにとどめておき、性能を向上させたい場合にGPUを追加することで対応できます。また、GPUの追加に応じた性能向上の見積もりも容易です。

### 【注意事項】

- 本資料を第三者に転送したり、本資料記載の内容をWebサイトへアップロードしたりするなど、情報の再配信はお断りします。
- 著作権は富士通株式会社、またはその情報提供者に帰属するため、記載内容を許可なく転載することを禁じます。
- 本資料に記載されている性能情報は、お客様システムにおける性能向上を保証するものではありません。

### 【商標について】

- NVIDIA®は米国NVIDIA社の登録商標または商標です。
- Intel、Xeon は米国インテル社の登録商標または商標です。
- その他の記載されている会社名、製品名等は各社の登録商標または商標です。
- その他、本書で記載されている会社名、システム名、製品名等には必ずしも商標表示（®・™）を付記しておりません。