FUJITSU

FUJITSU AI Zinrai ディープラーニング システム
FUJITSU Server PRIMERGY ご利用ガイド

マルチインスタンス GPU 利用手順

FUJITSU AI Zinrai ディープラーニング システム, FUJITSU Server PRIMERGY をご愛顧いただきありがとうございます。本書では当システム上で NVIDIA® A100 の GPU リソースを分割利用する方法を説明いたします。お客様のご利用環境に合わせて適宜読み替えてご利用ください。

2020 年 12 月

富士通株式会社

030040-20-0059

目次

1. 対象となる製品

FUJITSU AI Zinrai ディープラーニング システム (NVIDIA® A100 搭載対応モデル)

FUJITSU Server PRIMERGY (NVIDIA® A100 搭載対応モデル)
NVIDIA® A100

2. 用語について

本書で使用する略称などについて以下に示します。

略称など	呼称	概要
MIG	マルチインスタンス GPU	NVIDIA 社が提供する NVIDIA Ampere アーキテクチャを搭載した GPU や NVIDIA ドライバの機能
GI	GPU インスタンス	GPU メモリーを共有する分割単位
GIP	GPU インスタンスプロファイル	GPU メモリーと SMs の組合せの識別番号
CI	コンピュートインスタンス	仮想 GPU デバイスの単位
CIP	コンピュートインスタンスプロファイル	GI 内の SMs 割当数に対する識別番号
SM	ストリーミングマルチプロセッサ	GPU の計算処理実行部
SMs	ストリーミングマルチプロセッサ(セット)	14 個の SM をまとめた仮想 GPU に割り当てる単位

3. GPU の分割利用について

AI の事業適用を検討する場合に、社内の複数部門がそれぞれ独自に開発していたりプロトタイピングの延長でワークステーションやクラウド開発環境の利用を行っていて、複数の開発環境が乱立しているような状況にはなっていないでしょうか？このようなAI 開発環境の分散は、各データサイエンティストの知見やプラクティスのサイロ化を招き、業務システムへのプロダクションプロセスやデータパイプライン、セキュリティの観点などからデメリットがあります。

NVIDIA® A100 は高性能で大容量の GPU リソースを持っているため、大規模処理に用いるだけではなく、複数の仮想 GPU に分割して開発環境を統合する用途にも使うことができます。開発環境の統合は、お客様の AI 開発環境に様々なメリットをもたらします。

3.1 主な利点

3.1.1 社内開発基盤の標準化

統合されていない開発環境では各エンジニアが任意のソフトウェア環境を構築できてしまうため、モデル開発～PoC～業務環境への適用といったプロダクションのワークフローを円滑に進めにくい問題があります。開発環境を物理的な統合環境に集約することで、ホスト OS 環境が

一元化され、使用するコンテナイメージの管理も容易になるため、ソフトウェア環境を整備しやすくなりワークフローが円滑になります。

3.1.2 ベストプラクティスの共有

エンジニアの感じる統合環境のメリットとしては、高性能な GPU などの計算資源を柔軟に利用できることがあげられます。これによりワークステーションやクラウド環境では実行しにくかった複雑で大規模な学習を繰り返し実施する事ができ、より自社のビジネスに活用できるモデルの開発に繋がります。こうして開発した優れたネットワークモデルや安定したソフトウェア環境などを統合環境上で速やかに共有できることにより、手戻りの少ない効率的な開発を行うことができます。

3.1.3 コンピューティングリソースの有効活用

統合環境では複数の開発者が資源を分割して利用するだけではなく、コンピューティングリソースを必要に応じて多く割当てて、大規模な処理を行わせるなどの運用が可能です。これにより例えば日中は各エンジニアの開発やデバッグ作業を中心に行い、夜間や休日には大規模な学習処理を繰り返し行うなど、コンピューティングリソースを有効に活用することが可能です。

3.2 マルチインスタンス GPU(MIG)による統合環境

NVIDIA® A100 GPU はマルチインスタンス GPU(MIG)機能を搭載しており、1GPU を最大 7 台の仮想 GPU に分割することが可能です。これにより複数の開発環境を 1 台のサーバーに統合するなどお客様環境に合わせた柔軟な運用環境を構築することができます。
テクノロジーの詳細については NVIDIA 社の公式ドキュメントをご参照ください。

3.3 前提条件

NVIDIA® A100 GPU が搭載可能な Zinrai ディープラーニング システム、又は PRIMERGY に NVIDIA® A100 GPU を 1 枚以上搭載している構成を対象としています。また NVIDIA® A100 GPU をサポートするドライバ(450.80.02 以降)が適切にインストールされている必要があります。

3.4 MIG とは

MIG は NVIDIA 社が提供する NVIDIA Ampere アーキテクチャを搭載した GPU や NVIDIA ドライバの機能で、1 台の GPU を仮想的に分割する機能です。分割には GPU 資源(GPU メモリーなど)を共有する単位の GPU インスタンスと、GPU インスタンスを仮想 GPU デバイスの単位に分割して利用者に仮想 GPU デバイスとして見せるコンピュートインスタンスの 2 つの要素があります。用途に合わせてこれらを組み合わせて割当てを行います。下図は 3 人の利用者がそれぞれ別の使い方を行う様に割り当てる例です。左の利用者は 4 台の仮想 GPU を使い、中央の利用者

は大き目の 1 つの GPU を割り当てられています。残ったリソースを最後のひとりがデバッグ用に使っている運用例です。

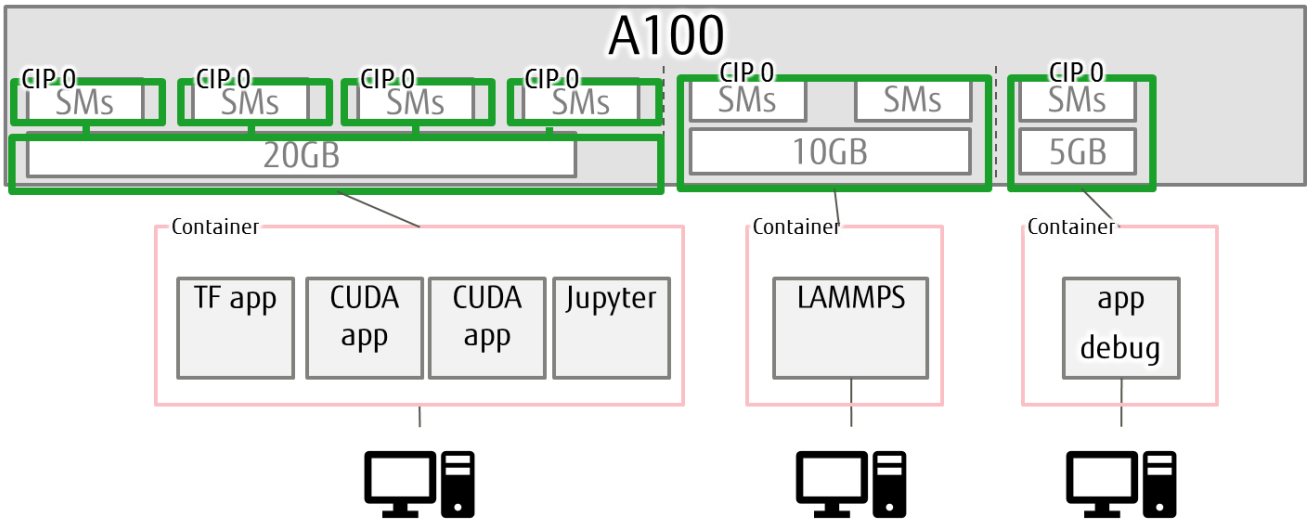

Figure 1 MIG 設定例

MIG の構成は柔軟に変更できますので、日中は多人数での開発作業に使い、夜間は転移学習のバッチ処理に割り当てるなど、利用形態に応じた運用を行えます。

3.4.1 GPU インスタンス (GI)

GPU インスタンスは GPU メモリーを共有する分割単位です。BERT や ResNet50 の転移学習など比較的多くのメモリー容量を必要とする場合は大きく、バッチサイズ 1 の推論や対話式学習によるモデル開発など常時大量のメモリーを必要としない場合は粒度を細かく分けることができます。またストリーミングマルチプロセッサ(SM)という計算処理を実行する資源もあります。GPU インスタンスはこの GPU メモリーと SM の組合せで構成を決定します。この構成は GPU インスタンスプロファイル(GIP)という番号で識別します。

GPU メモリーを共有することから、基本的には GPU インスタンスは利用者ごとに作成する単位と考えるとよいでしょう。

NVIDIA® A100 の場合 GPU メモリーと 1 つの物理 GPU 内に作成できる GPU インスタンスの関係を下図に示します。下図は 1 GPU を一種類の GPU インスタンスプロファイルの GPU インスタンスで最大いくつまで分割できるかを示しています。実際には GIP 0 を除き、SMs が 7 つ及び最大 40GB の GPU メモリーの範囲で複数種類の GPU インスタンスを作成することが可能です。なお、図中の MIG 1g 5GB などの表記は GPU インスタンスプロファイルの説明です。GIP 19 の場合では、1g の部分が SMs の数を 5gb に部分が GPU メモリーの容量を表します。

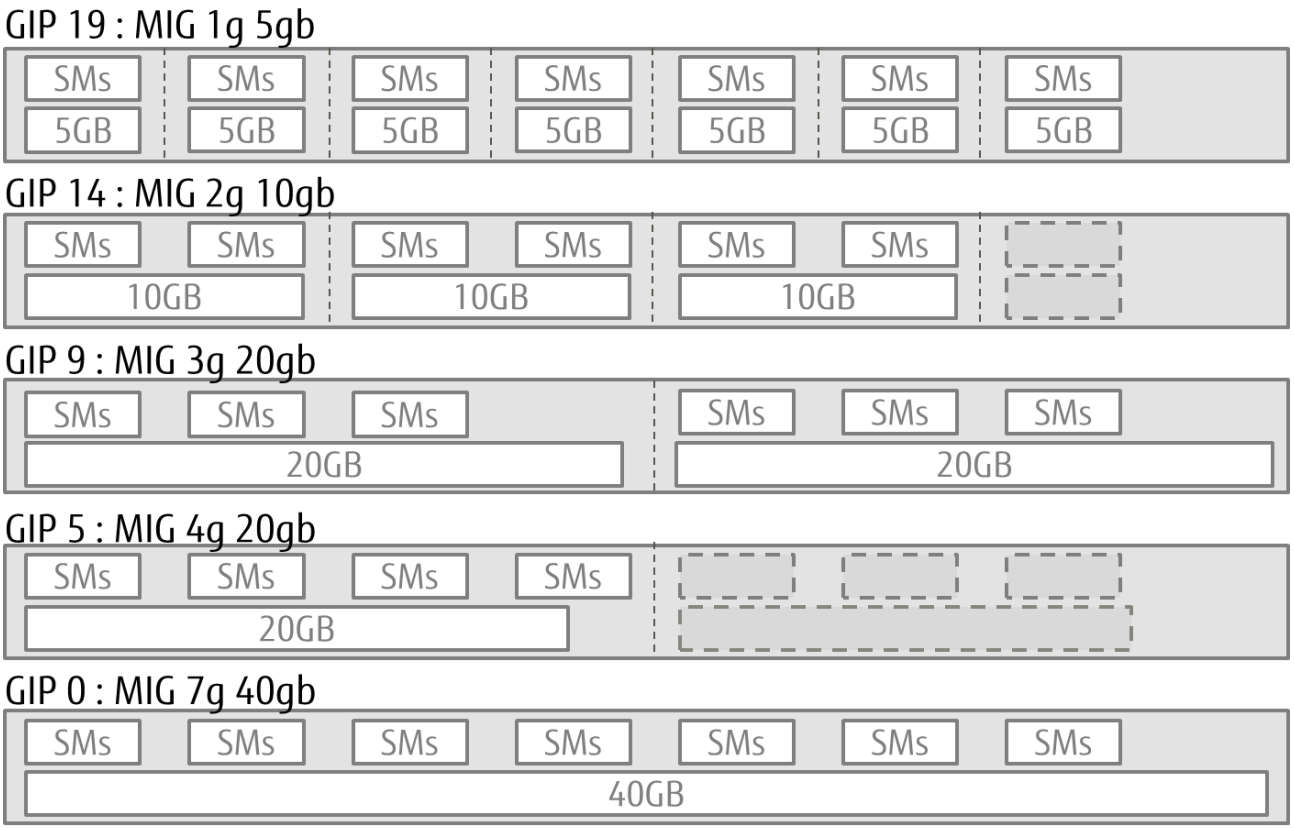

Figure 2 GPU インスタンスプロファイル

例えば Figure 1 *MIG 設定例* の場合、下図のように GIP 5 (4SMs+20GB)、GIP 14 (2SMs+10GB)、GIP 9(1SMs+5GB)の 3 つの GPU インスタンスに分割しています。

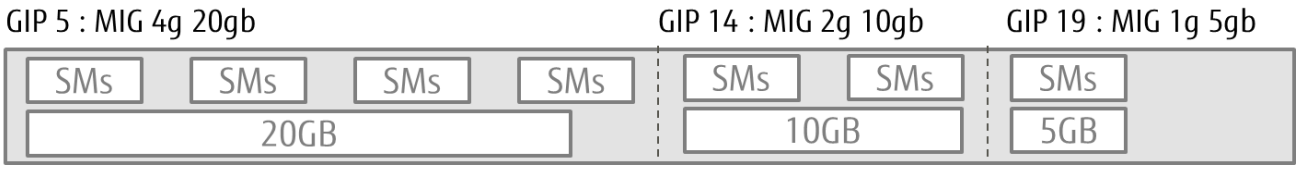

Figure 3 GPU インスタンスの分割例

3.4.2 コンピュートインスタンス (CI)

コンピュートインスタンスは GPU インスタンスを 1 つ以上に分割する仮想 GPU デバイスの単位です。GIP 0 の場合は GPU インスタンスを最大 7 分割、GIP 5 や 9 の場合は最大 3 分割、GIP 14 の場合は最大 2 分割して CI に割り当てることができます。GPU インスタンスを分割した場合は、各コンピュートインスタンスが同じ GPU インスタンス内の GPU メモリーを共有するため、主にひとりの利用者が複数の GPU に異なるタスクを割り当てて使用したい場合などに使うと良いでしょう。同一 GPU インスタンス内の仮想 GPU デバイス間では CUDA IPC(Inter Process Communication)が有効です。推論サーバーのような IPC を使うアプリケーションは同じ GID 配下の仮想 GPU デバイスを構成することで他のワークロードを分離できるので QoS を確保できます。

Figure 1 *MIG 設定例* では下図のように GIP 5 に 4 つの CIP 0 を割当て 20GB の GPU メモリーを共有、GIP 14 には CIP 0 をひとつだけ割当て 2SMs 分の大きな仮想 GPU を持つインスタンスを作成、さらに GIP 19 に残りひとつの SMs を割り当てたインスタンスを作成しています。

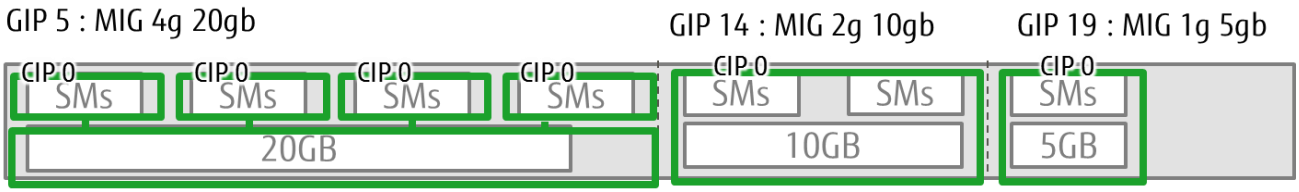

Figure 4 コンピュートインスタンスの割り当て例

コンピュートインスタンスを作成すると、GPU インスタンス内でユニークなコンピュートインスタンス ID(CI ID)と、物理 GPU 内でユニークな MIG デバイスインデックス (MIG Dev)が割り当てられます。コンテナに仮想 GPU デバイスを割り当てる場合は、物理 GPU の GPU デバイスインデックスと、MIG デバイスインデックス(MIG Dev)を指定します。

4. Jupyter Notebook 開発環境を各個人に割り当てる例

Jupyter Notebook はデータ分析やモデル開発の現場で使われる対話型開発環境です。ここでは 7 名のエンジニアに専用の開発環境を割り当てる想定で、下図のような環境を作成する手順を説明します。

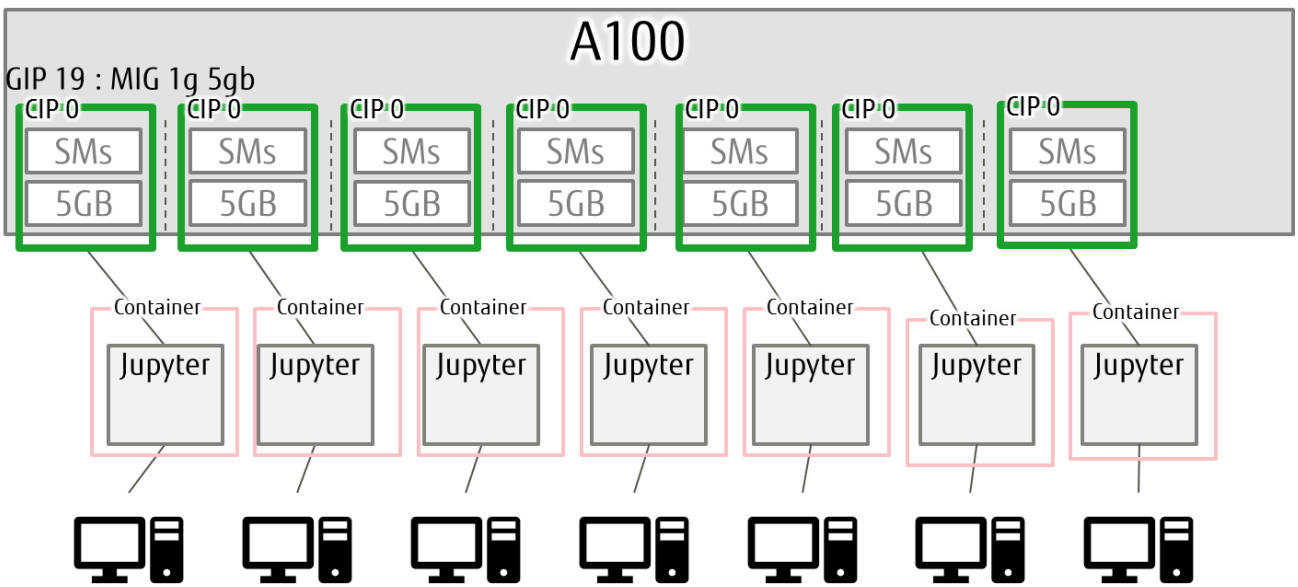

Figure 5 Jupyter 開発環境の例

4.1 MIG の有効化

4.1.1 Persistent Mode 有効化

MIG 利用時は Persistence モードが推奨されていますので、nvidia-smi コマンドで全 GPU の Persistence-M が On になっているか確認してください。

```
# nvidia-smi

+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.80.02    Driver Version: 450.80.02    CUDA Version: 11.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  A100-PCIE-40GB      On   | 00000000:25:00.0 Off |                    On |
| N/A   34C    P0    34W / 250W |                  N/A |     N/A      Default |
|                               |                      |              Enabled |
+-------------------------------+----------------------+----------------------+
#
```

Off の場合は、以下の方法で有効化してください。

nvidia-persistenced を使う場合 (推奨)

```
# systemctl start nvidia-persistenced.service
```

nvidia-smi コマンドで GPU に設定する場合

nvidia-persistenced が導入されていない環境の場合は下記コマンドで設定してください。

```
# nvidia-smi -pm 1
```

4.1.2 MIG enabling

A100 GPU を MIG 動作させる場合は、MIG mode の有効化が必要です。以下のコマンドで有効化してください。物理 GPU 毎に指定する場合は-i オプションで GPU 番号を指定してください。

```
# nvidia-smi -mig 1
```

nvidia-smi コマンドで指定した GPU の MIG-M が Enabled になっているか確認してください。

```
# nvidia-smi
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.80.02    Driver Version: 450.80.02    CUDA Version: 11.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  A100-PCIE-40GB      On   | 00000000:25:00.0 Off |                    On |
| N/A   34C    P0    34W / 250W |                  N/A |     N/A      Default |
|                               |                      |              Enabled |
+-------------------------------+----------------------+----------------------+
```

4.2 既存の MIG 設定の削除

必要に応じて既存の設定を削除してください。既存の設定が無い場合は No compute instances found: Not Found などエラーメッセージが表示されます。

以降では設定を行う物理 GPU インデックス"0"を明示指定しています。2 台目以降の GPU に設定する場合は、適宜物理 GPU インデックスを読み替えてください。

```
# nvidia-smi mig -i 0 -dci
Successfully destroyed compute instance ID  0 from GPU  0 GPU instance ID  0
# nvidia-smi mig -i 0 -dgi
Successfully destroyed GPU instance ID  0 from GPU  0
```

4.3 GPU インスタンスの作成

GPU メモリ 5GB と 14SM を持つ GI を割り当てます。この仕様の GPU インスタンスプロファイルは Table 1: GPU インスタンスプロファイルを確認すると GIP:19 で 7 つ作成できることがわかりますので、下記のコマンドで 7 つの GI を作成します。

```
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID 13 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID 11 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID 12 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID  7 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID  8 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID  9 on GPU  0 using profile MIG 1g.5gb (ID 19)
# nvidia-smi mig -i 0 -cgi 19
Successfully created GPU instance ID 10 on GPU  0 using profile MIG 1g.5gb (ID 19)
```

または、以下の様に一括して指定することもできます。

```
# nvidia-smi mig -i 0 -cgi 19,19,19,19,19,19,19
Successfully created GPU instance ID 13 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID 11 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID 12 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID  7 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID  8 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID  9 on GPU  0 using profile MIG 1g.5gb (ID 19)
Successfully created GPU instance ID 10 on GPU  0 using profile MIG 1g.5gb (ID 19)
```

なお GPU インスタンスプロファイルは下記コマンドで確認することもできます。すでに上記コマンドで GI を割り当てた後なので Instances Free の値が 0 になっていることがわかります。

```
# nvidia-smi mig -i 0 -lgip
+-----------------------------------------------------------------------------+
| GPU instance profiles:                                                      |
| GPU   Name          ID    Instances   Memory     P2P    SM    DEC   ENC    |
|                           Free/Total   GiB              CE    JPEG  OFA    |
|=============================================================================|
|   0  MIG 1g.5gb     19     0/7        4.75       No     14     0     0    |
|                                                         1     0     0    |
+-----------------------------------------------------------------------------+
|   0  MIG 2g.10gb    14     0/3        9.75       No     28     1     0    |
|                                                         2     0     0    |
+-----------------------------------------------------------------------------+
|   0  MIG 3g.20gb     9     0/2        19.62      No     42     2     0    |
|                                                         3     0     0    |
+-----------------------------------------------------------------------------+
|   0  MIG 4g.20gb     5     0/1        19.62      No     56     2     0    |
|                                                         4     0     0    |
+-----------------------------------------------------------------------------+
|   0  MIG 7g.40gb     0     0/1        39.50      No     98     5     0    |
|                                                         7     1     1    |
+-----------------------------------------------------------------------------+
```

4.4 コンピュートインスタンスの作成

7 つの GPU インスタンスには Table 2: コンピュートインスタンスプロファイルにある通り 7～13 の GPU インスタンス ID(GI ID)が割り当てられます。

コマンドでも確認することができます。各 GPU インスタンスごとに割当て可能なコンピュートインスタンス数が 1 つずつある事が確認できます。

```
# nvidia-smi mig -i 0 -lcip
+-----------------------------------------------------------------------------------------+
| Compute instance profiles:                                                              |
| GPU     GPU       Name             Profile  Instances   Exclusive       Shared       |
|       Instance                       ID     Free/Total     SM       DEC   ENC   OFA  |
|         ID                                                          CE    JPEG       |
|=========================================================================================|
|   0      7       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0      8       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0      9       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0     10       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0     11       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0     12       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
|   0     13       MIG 1g.5gb           0*     1/1           14        0     0     0   |
|                                                                      1     0         |
+-----------------------------------------------------------------------------------------+
```

下記コマンドで各 GPU インスタンス ID にそれぞれコンピュートインスタンスプロファイル (CIP)の 0 を割り当ててください。

備考 その GPU インスタンス内で最大の CIP には 0*のようにアスタリスクが付きます。GPU インスタンス内でコンピュートインスタンスを分割する必要が無い場合はアスタリスクが付いた CIP を使用してください。

```
# nvidia-smi mig -i 0 -gi 7 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID  7 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 8 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID  8 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 9 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID  9 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 10 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID 10 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 11 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID 11 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 12 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID 12 using profile
MIG 1g.5gb (ID  0)
# nvidia-smi mig -i 0 -gi 13 -cci 0
Successfully created compute instance ID  0 on GPU  0 GPU instance ID 13 using profile
MIG 1g.5gb (ID  0)
```

4.5 割当て状況の確認

nvidia-smi コマンドで割り当てた結果を確認できます。

```
# nvidia-smi
Thu Oct 29 21:12:56 2020
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.80.02    Driver Version: 450.80.02    CUDA Version: 11.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  A100-PCIE-40GB      On   | 00000000:18:00.0 Off |                   On |
| N/A   35C    P0    33W / 250W |     25MiB / 40537MiB |     N/A      Default |
|                               |                      |              Enabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| MIG devices:                                                                |
+------------------+----------------------+-----------+----------------------+
| GPU  GI  CI  MIG |         Memory-Usage |        Vol|         Shared       |
|      ID  ID  Dev |           BAR1-Usage | SM     Unc| CE  ENC  DEC  OFA  JPG|
|                  |                      |        ECC|                      |
|==================+======================+===========+======================|
|  0    7   0   0  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0    8   0   1  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0    9   0   2  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0   10   0   3  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0   11   0   4  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0   12   0   5  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  0   13   0   6  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+
|  1    0   0   0  |      0MiB / 40537MiB | 98      0 |  7   0    5    1    1 |
|                  |      1MiB / 65536MiB |           |                      |
+------------------+----------------------+-----------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

4.6 コンテナへの割当て

作成した各仮想 GPU デバイスをコンテナに割り当てるには、docker run コマンドの gpus オプションで指定を行います。以下は GPU インデックス"0"の MIG デバイスインデックス"3"を割当てたコンテナで使用できる仮想 GPU デバイスを nvidia-smi コマンドで確認した例です。

```
# docker run -it --rm --network=host --gpus '"device=0:3"' --ipc=host --ulimit memlock=-
1 --ulimit stack=671088
64 nvcr.io/nvidia/mxnet:20.09-py3 bash

===========
== MXNet ==
===========

NVIDIA Release 20.09 (build 16068190)
MXNet Version 1.7.0

Container image Copyright (c) 2020, NVIDIA CORPORATION.  All rights reserved.
Copyright (c) 2015-2020 by MXNet Contributors

Various files include modifications (c) NVIDIA CORPORATION.  All rights reserved.
NVIDIA modifications are covered by the license terms that apply to the underlying
project or file.

NOTE: MOFED driver for multi-node communication was not detected.
      Multi-node communication performance may be reduced.

root@zdlsserver:/workspace# nvidia-smi
Thu Oct 29 12:25:19 2020
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 450.80.02    Driver Version: 450.80.02    CUDA Version: 11.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  A100-PCIE-40GB      On   | 00000000:18:00.0 Off |                   On |
| N/A   36C    P0    33W / 250W |                  N/A |     N/A      Default |
|                               |                      |              Enabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| MIG devices:                                                                |
+------------------+----------------------+-----------+----------------------+
| GPU  GI  CI  MIG |         Memory-Usage |        Vol|         Shared       |
|      ID  ID  Dev |           BAR1-Usage | SM     Unc| CE  ENC  DEC  OFA  JPG|
|                  |                      |        ECC|                      |
|==================+======================+===========+======================|
|  0   10   0   0  |      3MiB /  4864MiB | 14      0 |  1   0    0    0    0 |
|                  |      0MiB /  8191MiB |           |                      |
+------------------+----------------------+-----------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

5. 補足 : 各インスタンスとインスタンスプロファイル、及び割当て資源の関係

NVIDIA® A100 の場合 GPU メモリーと作成できるインスタンスの関係は下表の通りです。

Table 1: GPU インスタンスプロファイル

GIP	GIP 名	SM(x14)	GPU メモリー	最大作成数	主な用途
19	MIG 1g.5gb	1	5GB	~7	推論/開発
14	MIG 2g.10gb	2	10GB	~3	転移学習
9	MIG 3g.20gb	3	20GB	~2	大規模学習
5	MIG 4g.20gb	4	20GB	1	大規模学習
0	MIG 7g.40gb	7	40GB	1	大規模学習

各 GPU インスタンス ID(GI ID)に対して割当てることができるコンピュートインスタンスプロファイル(CIP)は下表の通りです。

Table 2: コンピュートインスタンスプロファイル

GIP	GI ID	CIP	SM(x14)	最大作成数	GPU メモリー
19	13	0	1	1	5GB
	11	0	1	1	5GB
	12	0	1	1	5GB
	7	0	1	1	5GB
	8	0	1	1	5GB
	9	0	1	1	5GB
	10	0	1	1	5GB
14	5	0	2	2	10GB
		1	1	1	
	3	0	2	2	10GB
		1	1	1	
	4	0	2	2	10GB
		1	1	1	
9	2	0	1	3	20GB
		1	2	1	
		2	3	1	
	1	0	1	3	20GB
		1	2	1	
		2	3	1	
5	1	0	1	4	20GB
		1	2	2	
		3	4	1	
0	0	0	1	7	40GB
		1	2	3	
		2	3	2	
		3	4	1	
		4	7	1	

【留意事項】

- 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。
- 本資料記載の内容について、当社は、その正確性・商品性・ご利用目的への適合性等に関して保証するものではなく、そのご利用により生じた損害について、当社は法律上のいかなる責任も負いかねます。
- 本資料記載の内容は、予告なく変更・廃止されることがあります。
- FUJITSU AI Zinrai ディープラーニング システムについて不明な点は「Zinrai ディープラーニングシステムに関するお問い合わせ」
  https://www.fujitsu.com/jp/solutions/business-technology/ai/ai-zinrai/services/deep-learning/
  よりお尋ねください。
- FUJITSU Server PRIMERGY について不明な点は「PC サーバ FUJITSU Server PRIMERGY お問い合わせ」https://www.fujitsu.com/jp/products/computing/servers/primergy/contact/
  よりお尋ねください。

【商標について】

- NVIDIA®は米国 NVIDIA 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。