

VMware および NVIDIA による AI 用 GPU の仮想化

Dell EMC インフラストラクチャ ベース

2021 年 10 月

H18904

設計ガイド

要約

この設計ガイドでは、VMware および NVIDIA と共同で設計された Dell Technologies Validated Design for AI のリファレンス アーキテクチャについて説明します。仮想化 GPU のパフォーマンス特性の導入ガイドラインと導入による結果を示します。

Dell Technologies ソリューション

Dell Technologies

Validated Design



著作権

この資料に記載される情報は、現状有姿の条件で提供されています。Dell Inc.は、この資料に記載される情報に関する、どのような内容についても表明保証条項を設けず、特に、商品性や特定の目的に対する適応性に関する黙示の保証はいたしません。

本書に記載されているすべてのソフトウェアの使用、複写、および配布には、該当するソフトウェア ライセンスが必要です。

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies、Dell、EMC、Dell EMC、ならびにこれらに関連する商標および Dell 又は EMC が提供する製品およびサービスにかかる商標は Dell Inc. Intel、インテル、Intel ロゴ、Intel Inside ロゴ、Xeon は、米国およびその他の国における Intel Corporation またはその子会社の商標または登録商標です。その他の商標は、各社の商標または登録商標です。Published in the USA 10/21 Design Guide H18904.

掲載される情報は、発信現在で正確な情報であり、この情報は予告なく変更されることがあります。

目次

ソリューションの概要.....	4
ソリューションアーキテクチャ.....	5
導入ガイドライン	11
検証と機械学習のパフォーマンス	14
概要	18
参考資料	20

ソリューションの概要

概要

デル・テクノロジーズ、NVIDIA、VMware は、企業全体で AI を民主化して活用するための新しいソリューションを立ち上げることにより、企業に新しい方法を提供しています。この新しい Dell Technologies Validated Design for AI は、組織が AI ワークロードの仮想化のメリットを活用できるように共同で設計および検証されています。Dell Technologies Validated Design for AI には、Dell EMC PowerEdge サーバーおよび VxRail ハイパーコンバージド インフラストラクチャ(HCI)上の NVIDIA AI Enterprise スイートと組み合わせた最新バージョンの VMware vSphere が含まれています。また、この設計には、必要な分析パフォーマンスと大規模な同時処理を提供する Dell EMC PowerScale も含まれており、非常に多くのデータを扱う AI アルゴリズムを一貫してフィードします。

ドキュメントの目的

この設計ガイドでは、Dell Technologies HPC および AI イノベーション ラボで導き出された結果に基づいて、エンタープライズ規模で AI プラットフォームを実装するためのガイドラインを示します。ラボのセットアップには、デル・テクノロジーズのサーバー、ストレージ、ネットワーキング、VMware の仮想化ソフトウェア、NVIDIA の高速化、ネットワーキング、ソフトウェアが含まれており、3 つのパートナーのベスト プラクティスを組み合わせて構成されています。この設計ガイドでは、すべての推奨構成、ネットワーク トポロジー、導入ガイドライン、確認されたパフォーマンスについて説明します。

この設計ガイドの情報を使用すれば、経験豊富な VMware のプロフェッショナルは、既存のスキルを迅速に使用して十分な機能を備えたプラットフォームを導入して運用し、高度な AI ユース ケースをサポートできます。

この設計ガイドは、『VMware および NVIDIA による AI 用 GPU の仮想化』テクニカル ホワイト ペーパーのコンパニオン ガイドです。この Validated Design で使用されるテクノロジー コンポーネントの詳細については、ホワイト ペーパーを参照してください。

対象読者

この設計ガイドは、ソリューション アーキテクト、システム管理者、仮想化プラットフォームでの AI アプリケーションの開発に関心のある方を対象としています。

ソリューションアーキテクチャ

PowerEdge サーバーで NVIDIA GPU 搭載の VMware vSphere を使用して、ニューラル ネットワークトレーニング、推論、モデル開発などの混在ワークロードを実行したいと考えている IT 管理者にとって、このホワイトペーパーの情報は役立ちます。次の図は、2 つの NVIDIA GPU と ConnectX ネットワークアダプターを搭載した PowerEdge サーバーを VMware vSphere クラスターの一部として実装する方法を示しています。

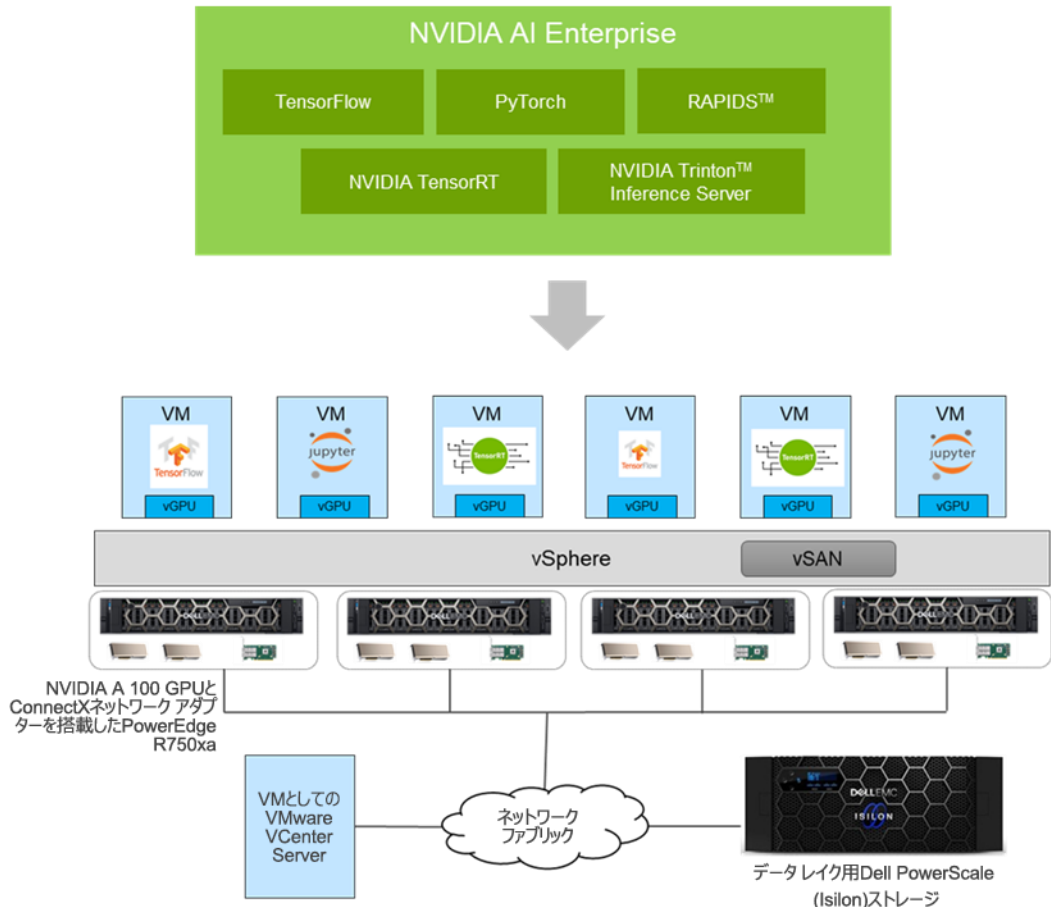


図 1. NVIDIA A100 GPU と VMware vSphere を搭載した PowerEdge サーバー

VMware vSAN を実行しているサーバーは VM とオペレーティング システムの要件に対応するストレージ リポジトリを提供し、PowerScale リソースは分析ワークロード用のデータレイクストレージを提供します。

管理者は NVIDIA A100 GPU の MIG 機能を使用して、vGPU プロファイルを作成して VM に割り当てることができます。これらの VM は、トレーニング用の TensorFlow、インタラクティブなモデル開発用の Jupyter ノートパソコン、推論用の TensorRT など、さまざまなワークロードを実行できます。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

推奨構成 次の表に、お客様の用途例に応じた推奨構成を示します。

テーブル 1. 推奨構成

構成	ミッドレンジのパフォーマンス	HCI アプライアンスを使用したメインストリームのパフォーマンス	ハイ パフォーマンス
コンピューティングサーバ	PowerEdge R7525 または R750	VxRail 670F	PowerEdge R750xa
GPU	1 x NVIDIA A30	2 x NVIDIA A30	4 x NVIDIA A100
ネットワーク アダプター	ConnectX-5 DX 25 GbE ¹	ConnectX-5 DX 25 GbE	ConnectX-5 DX 25 GbE マルチノードトレーニングの場合 : ConnectX-6 DX デュアルポート 100 GbE
ネットワーク スイッチ	Dell S5248F-ON x 2	ネットワーク スイッチ	Dell S5248F-ON x 2 マルチノードトレーニングの場合 : Dell S5232F-ON x 2 または Mellanox SN3700 x 2 (データ)
OOB スイッチ	Dell S4148T-ON x 1		
VMware vSphere	vSphere 7.0 U2 ² 以降		
ESXi 用の内蔵ストレージ	Boot Optimized Storage Subsystem コントローラー カード + M.2 スティック 480 GB (RAID 1) x 2		
vSAN 用のストレージコントローラーまたはハードドライブの構成	<ul style="list-style-type: none">ストレージ コントローラー : HBA 330キャッシュ階層 : 800 GB SSD SAS 書き込み集中型 x 2容量階層 : 960 GB SSD SAS 読み取り集中型 x 12		
ストレージ	<ul style="list-style-type: none">VM ストレージ用の vSANデータレイク用の PowerScale オールフラッシュ (F200、F600、F800) またはハイブリッドノード (H400、H500、H600、H5600)		
NVIDIA ソフトウェア	NVIDIA AI Enterprise (vGPU 13)		

メインストリームのパフォーマンスには、PowerEdge サーバーを使用した構成と HCI アプライアンスを使用した構成の 2 つの構成をお勧めします。メインストリームには、モデル開発と推論が含まれます。これには、対話型 AI、推薦システム、言語処理などのユース ケースが含まれます。ハイ パフォーマンス構成には、NVIDIA A100 GPU などのアクセラレーター用に設計された PowerEdge R750xa サーバーが含まれています。ハイ パフォーマンスのユース ケースには、大規模で複雑なモデルのニューラル ネットワークトレーニングが含まれます。お客様が複数のサーバーで GPU リソースを活用することを希望する場合、ハイ パフォーマンス構成には、マルチノードトレーニング用のネットワーク アーキテクチャも含まれます。

¹ このガイドの公開時点では、デル・テクノロジーズは ConnectX-6 25 GbE ネットワーク アダプターをサポートしていません。本書は、サポートが提供された時点で更新されます。

² NVIDIA A30 GPU で MIG をサポートするには、VMware vSphere 7.0 U3 が必要です。

コンピューティング サーバーの設計に 関する考慮事項

機械学習コンピューティング サーバーの設計に関する考慮事項は次のとおりです。

- **PowerEdge サーバーおよび VxRail モデル** : PowerEdge R7525 または R750 サーバーは、メインストリームのパフォーマンスに最適です。PowerEdge R750xa サーバーはアクセラレーター用に設計されており、PowerEdge サーバー間で GPU 密度が最も高くなります。ターンキーでフル スタックのライフサイクル管理を提供する HCI アプライアンスを必要としているお客様は、コンピューティング サーバーとして VxRail を選択するとよいでしょう。これらのモデルはすべて NVIDIA-Certified Systems です。つまり、高速化されたワークロードに最適なパフォーマンス、拡張性、セキュリティを提供することが検証されています。
- **コンピューティング サーバーの数** : VMware vSphere クラスターあたり最大 4 台のサーバーについて検証されています。VMware は、クラスターあたり最大 64 ノードをサポートします。
- **プロセッサとメモリー** : PowerEdge R750、PowerEdge R750xa、VxRail 670F サーバーにはインテル Xeon Platinum または Gold プロセッサ、PowerEdge R7525 サーバーには AMD EPYC プロセッサを推奨します。メモリー負荷の高い AI ワークロードには、少なくとも 512 GB のメモリーを推奨します。
- **GPU** : PowerEdge サーバーは NVIDIA A100 または A30 GPU で構成できます。A100 GPU は、ディープラーニングや複雑なニューラルネットワークのトレーニングに推奨されます。A30 GPU は、AI 推論、言語処理、対話型 AI、推薦システムに推奨されます。
- **VMware ESXi 用のストレージ** : ESXi server は、SD カードまたは Boot Optimized Storage Subsystem コントローラーにインストールできます。
- **vSAN 用のストレージ コントローラーとハード ドライブ** : vSAN 用のストレージ コントローラーとして Dell HBA330 コントローラーが使用されます。お客様のワークロードと VM の要件によって、ハードドライブの要件が決まります。vSAN キャッシュ階層には 800 GB SSD SAS 書き込み集中型、vSAN 容量階層には 960 GB SSD SAS 読み取り集中型が使用されています。
- **ネットワーク アダプター** : [テーブル 2](#) に示すように、25 GbE 設計では ConnectX-5 ネットワーク アダプターを使用し、100 GbE 設計では ConnectX-6 ネットワーク アダプターを使用します。100 GbE 設計では、管理、vSphere vMotion、100 GbE トラフィックを必要としないその他の VM にインテル X710 10 GbE ネットワーク アダプターも使用します。

管理コンポーネント の設計に関する考 慮事項

VMware vCenter Server は、次のいずれかの方法で導入できます。

- vCenter Server は、1 台の PowerEdge または VxRail サーバー（コンピューティング クラスター）にインストールされます。この導入は、小規模な環境に対してのみ推奨されます。メンテナンス、アップグレード、vCenter Server の可用性に影響を与える可能性があるその他のホスト操作時は注意を払ってください。
- vCenter Server は別の管理クラスターにインストールされ、GPU を使用してコンピューティング クラスターにネットワーク接続できます。

ネットワーク設計に 関する考慮事項

Dell Technologies HPC および AI イノベーション ラボのエンジニアリング チームは、この Validated Design のために次のネットワーク構成を導入しました。

- **PowerSwitch ネットワーク スイッチを使用した 1 つの 25 GbE ベースの設計** : 100 Gb のネットワーク インフラストラクチャに投資することなく、既存のネットワーク インフラストラクチャを使用できるワークロード用にこの設計をお勧めします。この設計は、1 つのノードで実行できるニューラルネットワークトレーニング ジョブ（最大 2 つの GPU を使用）や、GPU の分割を活用するモデル開発や推論ジョブに適しています。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

この設計オプションでは、管理、vSphere vMotion、VM トラフィック用のトップオブラック(ToR)スイッチとして S5248-ON スイッチを使用します。S5248-ON スイッチの代わりに、既存の 10 Gb または 25 Gb Ethernet ネットワーク インフラストラクチャを使用することもできます。

- **GPUDirect RDMA を使用したマルチノード トレーニング用の PowerSwitch および Mellanox Spectrum ネットワーク スイッチを使用した 2 つの 100 GbE ベース設計**：大規模なデータセット（通常は高解像度ビデオまたはイメージ ベースのデータセット）を使用した分散マルチノード モデル トレーニングを必要とするワークロードには、これらの設計を推奨します。

次の図は、PowerSwitch ネットワーキングを使用した 25 GbE Ethernet 設計用のネットワーク トポロジを示しています。

f

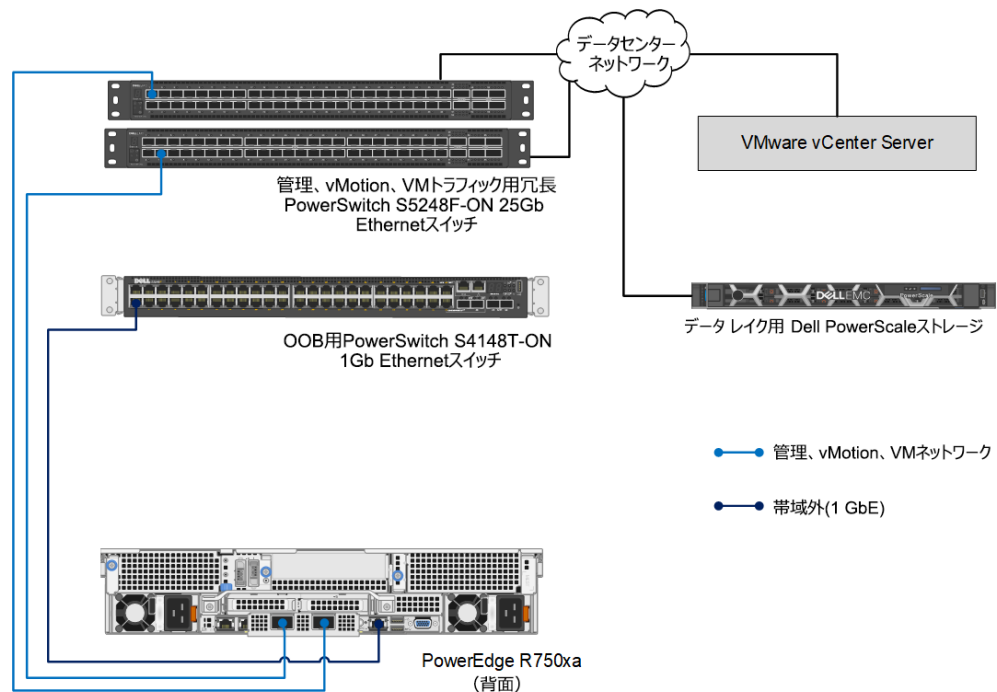


図 2. Dell PowerSwitch ネットワーク スイッチを使用した 25 GbE ベース設計用のネットワーク トポロジ

この例は、1 台のみの PowerEdge サーバーのネットワーク接続を示しています。vSphere クラスタ内の他の PowerEdge サーバーも同様の接続となります。2 つの冗長 S5248-ON スイッチが、25 Gb Ethernet 接続を提供する ToR スイッチとして使用されています。PowerEdge R750xa サーバーの ConnectX-5 DX 25 GbE デュアルポートネットワークアダプターは、ToR スイッチへの接続を提供します。

S3048-ON スイッチは、OOB 接続用に 1 Gb Ethernet を提供します。各 PowerEdge サーバーの iDRAC がこのスイッチに接続されています。

vCenter Server と PowerScale ストレージは、ToR スイッチに接続できます。

次の図は、Dell PowerSwitch S5232F-ON ネットワーキング スイッチを使用した 100 GbE Ethernet 設計用のネットワーク トポロジを示しています。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

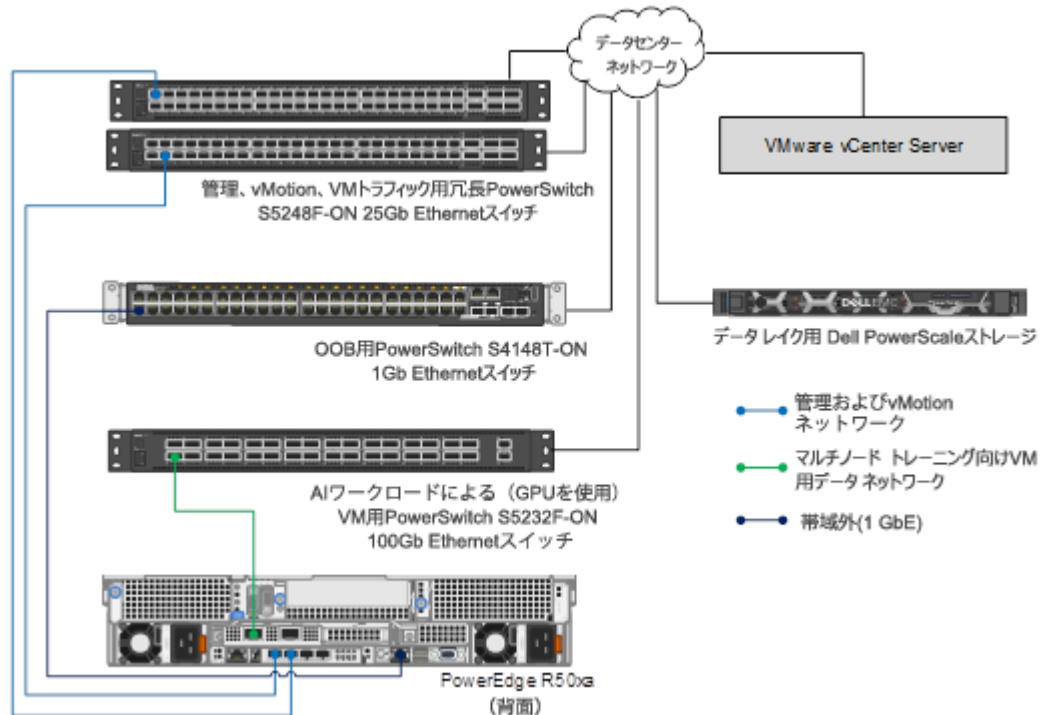


図 3. Dell PowerSwitch ネットワーク スイッチを使用した 100 GbE ベース設計用のネットワーク トポロジー

この例は、1 台のみの PowerEdge サーバーのネットワーク接続を示しています。クラスター内の他の PowerEdge サーバーも同様の接続となります。S5232-ON スイッチは、AI ワークロードを備えた VM に 100 Gb Ethernet 接続を提供する ToR スイッチとして使用されます。PowerEdge R750xa サーバーの ConnectX-6 Dx 100 GbE デュアル ポート ネットワーク アダプターは、100 Gb Ethernet ToR スイッチへの接続を提供します。

管理、vSphere vMotion、その他の VM トラフィック用に 10 Gb または 25 Gb Ethernet 接続を提供する ToR スイッチとして、2 台の冗長 S5248-ON スイッチが使用されます。PowerEdge R750xa サーバーの Intel または Broadcom 10 Gb Ethernet アダプター（ネットワーク ドーターカード）は、S5248-ON スイッチへの接続を提供します。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

次の図は、NVIDIA Mellanox Spectrum ネットワーク スイッチ SN3700 を使用した 100 GbE Ethernet 設計用のネットワーク トポロジーを示しています。

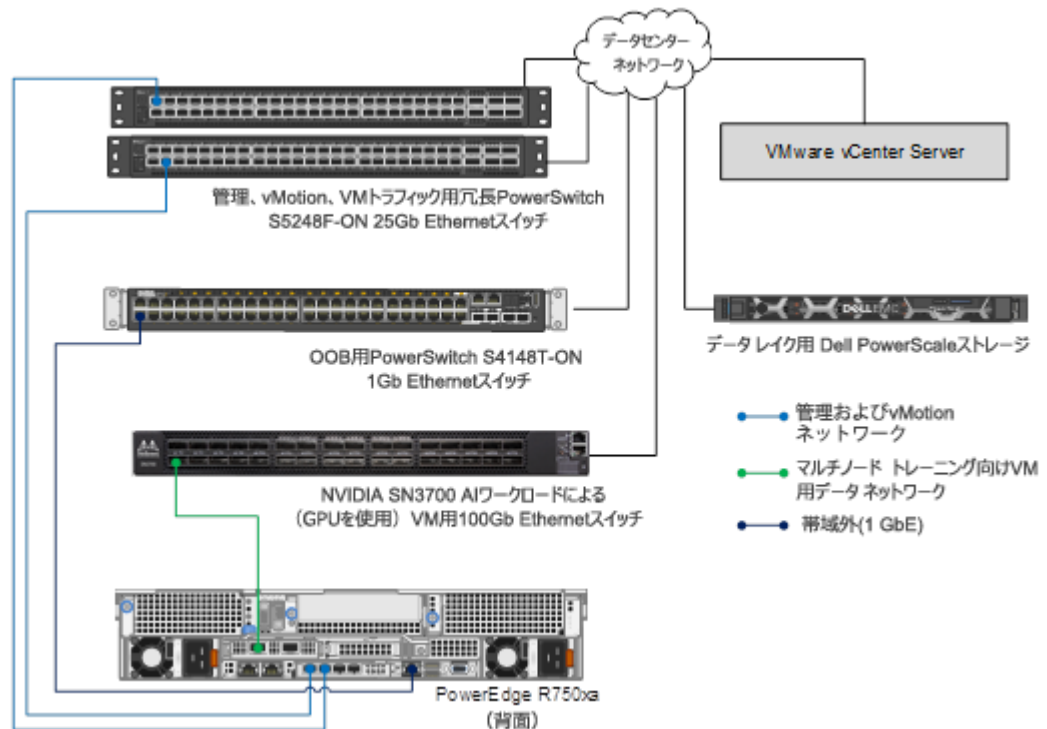


図 4. Mellanox Spectrum ネットワーク スイッチを使用した 100 GbE ベース設計用のネットワーク トポロジー

仮想スイッチの設計に関する考慮事項

25 GbE と 100 GbE の両方の設計では、次の図に示すように、管理用のポート グループ、vSphere vMotion、vSAN、VM を含む 1 つの分散仮想スイッチを作成します。ConnectX-5 DX 25 GbE ネットワーク アダプターは、この分散仮想スイッチへのアップリンク用に割り当てられます。

f

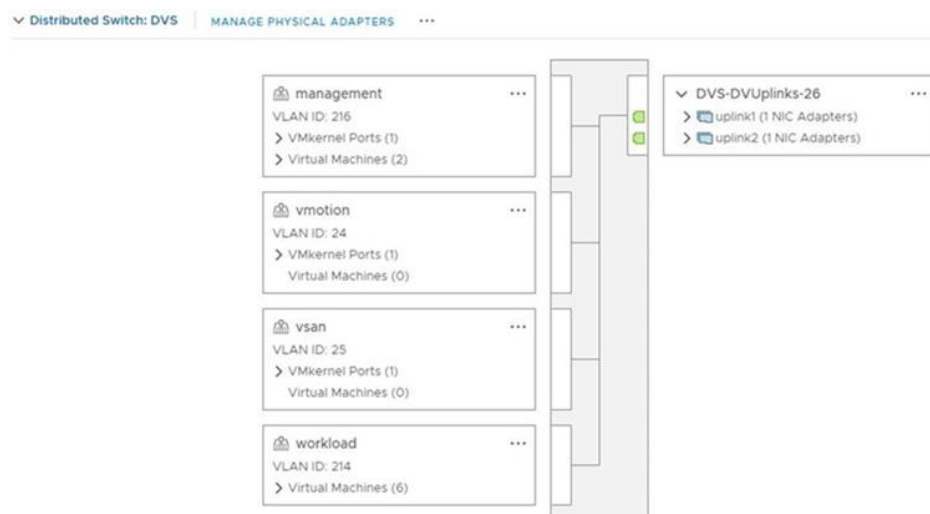


図 5. 25 GbE 設計の分散仮想スイッチ

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

100 GbE 設計では、マルチノードトレーニングを実行する仮想マシンに対してパススルー デバイスとして 100 GbE ConnectX-6 ネットワーク アダプターを割り当てます。

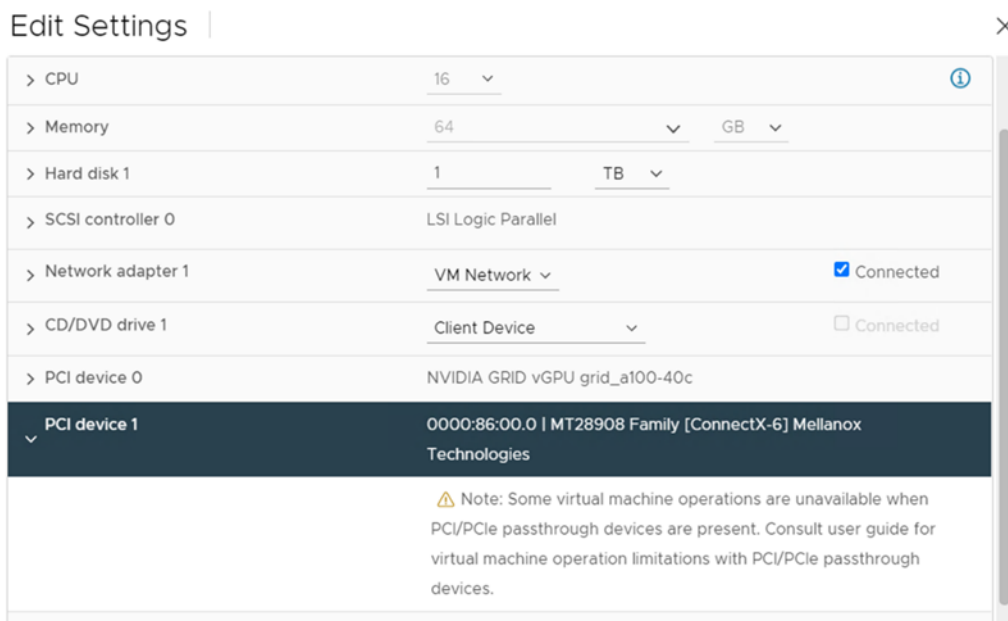


図 6. 100 GbE 設計の仮想マシンへのパススルーNIC の割り当て

ストレージ設計に関する考慮事項

信頼性の向上、パフォーマンスの向上、VM データからの分離性の強化のために、Boot Optimized Storage Subsystem コントローラー カードに VMware ESXi をインストールすることをお勧めします。

vSphere vMotion には共有ストレージが必要です。vSAN は VM に推奨されるストレージです。データレイク ストレージ、つまりニューラル ネットワーク トレーニングに必要なデータの保存には PowerScale ストレージをお勧めします。

導入ガイドライン

このセクションでは、PowerEdge サーバーに A100 または A30 GPU を使用した vSphere 7.0 搭載の GPU クラスターを導入する際のガイドラインについて説明します。MIG を有効にし、vGPU にアクセスして VM を構成することを中心に説明します。ESXi、vCenter、PowerScale ストレージ、vSAN ストレージ、ネットワーキング スイッチを導入する際の標準的なベスト プラクティスに従ってください。このセクションでは説明しません。

NVIDIA Virtual GPU Manager for vSphere のインストールと設定

NVIDIA Virtual GPU Manager for vSphere のインストールと設定の詳細な手順については、『[NVIDIA Virtual GPU Software User Guide](#)』を参照してください。次の大まかな手順について説明しています。

1. vSphere インストール バンドル(VIB)を使用して、NVIDIA Virtual GPU Manager for vSphere をインストールします。
2. 高度な vCenter Server 設定を有効にして、vSphere 用 vGPU を使用して vSphere vMotion を構成します。
3. vSphere でデフォルトのグラフィックス タイプを変更します。

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

4. MIG でサポートされる vGPU の GPU を構成します。デフォルトでは、A100 または A30 GPU では MIG モードは有効になっていません。

MIG を使用するように VM を構成するには、次の手順を実行します。

1. 次の図に示すように、VM を作成し、仮想プロセッサとメモリー構成を指定して、MIG パーティションを割り当てます。

MIG を使用する ための VM の構成

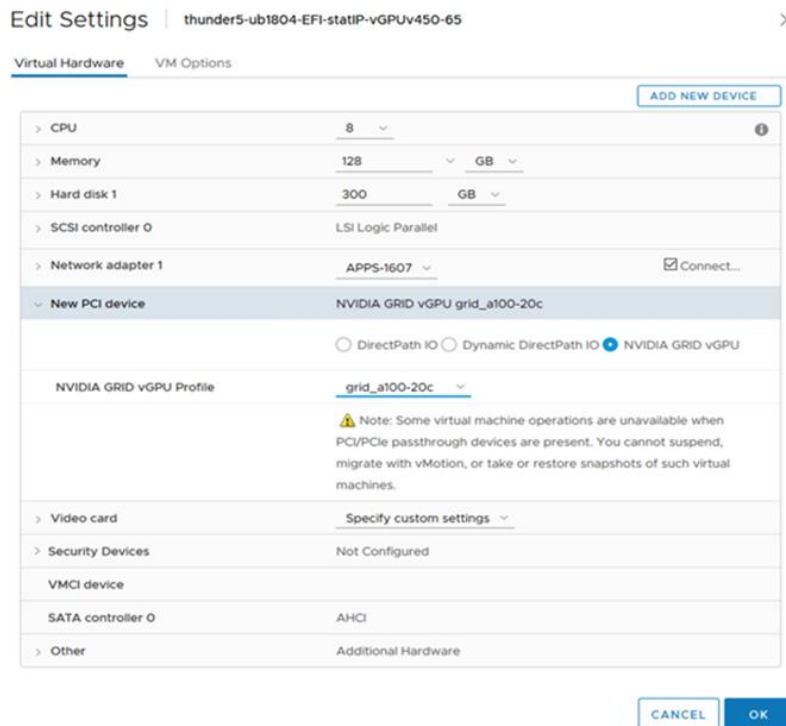


図 7. VM の作成と MIG プロファイルの割り当て

2. 正しい GPU を使用するために、VM の EFI モードまたは UEFI モードでゲスト オペレーティング システムを起動します。
 - a. vSphere Client で VM を選択します。
 - b. **[Edit Settings (設定の編集)]** → **[VM Options (VM オプション)]** → **[Boot Options (起動オプション)]** の順に移動します。[Firmware (ファームウェア)] フィールドで、次の図に示すように、UEFI または EFI が有効になっていることを確認します。

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

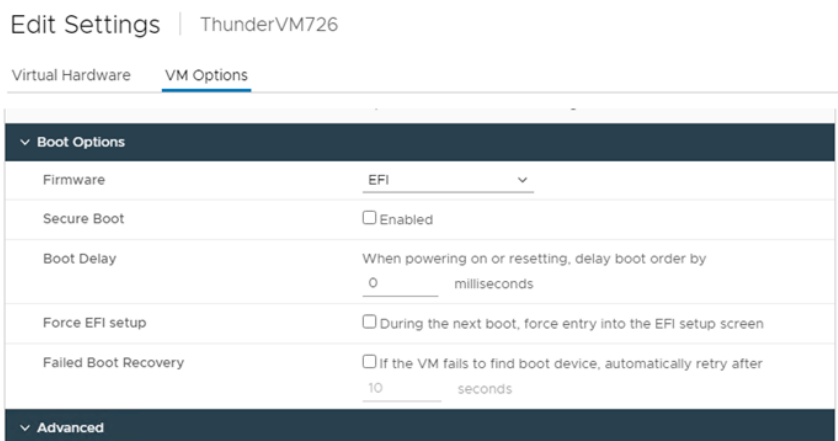


図 8. VM 起動オプション

3. **[Advanced (詳細)]** → **[Configuration Parameters (設定パラメーター)]** → **[Edit Configuration (設定の編集)]** の順にクリックし、次のパラメーターを編集します。

```
pciPassthru.use64bitMMIO: TRUE
pciPassthru.allowP2P: TRUE
pciPassthru.64bitMMIOSizeGB: 64
```

4. VM に Linux オペレーティング システムをインストールします。

5. NVIDIA vGPU Software Graphics Driver を VM にインストールします。

6. VM 上の GRID vGPU のライセンス設定を行います。

手順 4～6 の詳細な手順については、『[Virtual GPU Software User Guide](#)』を参照してください。

NVIDIA AI Enterprise スイートのインストール

AI およびデータサイエンスのアプリケーションとフレームワークは、NVIDIA NGC エンタープライズ カタログを通じて NGC コンテナ イメージとして配布されます。各コンテナ イメージには、アプリケーションまたはフレームワークの実行に必要なユーザースペース ソフトウェア スタック全体が含まれています。つまり、CUDA ライブラリー、cuDNN、必要な Magnum IO コンポーネント、TensorRT、フレームワークが含まれています。NVIDIA AI Enterprise ドキュメントの「[Installing AI and data science applications and frameworks](#)」の章を参照してください。

vSAN のガイドライン

ソリューション検証では、Dell EMC vSAN Ready Nodes を使用して、vSAN クラスター導入の要件をシンプルにしています。VMware vCenter Server Appliance ユーティリティを使用して、vSAN 導入ステップを効率化します。vSAN の導入の詳細については、『[VMware vSAN Design Guide](#)』を参照してください。最初の vCenter インスタンスを導入した後、vCenter QuickStart ウィザードを使用して vSAN クラスターを完了することをお勧めします。vSAN クラスターのプロビジョニングを効率化するには、「[Using Quickstart to Configure and Expand a vSAN Cluster](#)」を参照してください。

PowerScale のガイドライン

ニューラル ネットワーク トレーニングに必要なデータセットを保存するために、Dell EMC PowerScale ストレージを使用します。PowerScale は NFS マウント ポイントとして使用され、ストレージは仮想マシンで実行されているコンテナで使用可能になります。

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

GPUDirect RDMA を使用したマルチノードトレーニング

ニューラル ネットワークの分散トレーニングを実行するように複数の仮想マシンを構成できます。構成の主要ステップには、次のようなものがあります。

- GPU で MIG を無効にし、非 MIG プロファイルを仮想マシンに割り当てる
- ConnectX ネットワーク アダプターを構成し、パススルー モードで VM に割り当てる
- GPU と NIC のペアが同じルート コンプレックスと同じ NUMA ノード上に存在することを確認する
- ネットワーク アダプターとネットワーク スイッチで RDMA over Converged Ethernet (RoCE) を構成する

GPUDirect RDMA を使用してマルチノード トレーニングをセットアップする方法の詳細については、NVIDIA の『[AI Practitioners Deployment Guides](#)』を参照してください。

検証と機械学習のパフォーマンス

HPC および AI イノベーション ラボでは、この設計ガイドに記載されているソリューションの導入および検証を行っています。私たちは、設計ガイドに記載されている PowerEdge サーバーに NVIDIA AI Enterprise という 3 つのネットワーク アーキテクチャを導入しました。ベア メタル GPU と仮想化 GPU を比較してパフォーマンスを把握するための調査も実施しました。また、分割済み GPU のパフォーマンスを明らかにする調査も実施しました。このセクションでは、パフォーマンス調査の結果について説明します。

環境

次の 2 つのワークロードを使用して、パフォーマンス調査を実施します。

- **ResNet 推論** : NVIDIA NGC のコンテナとして使用可能な ResNet50 バージョン 1.5 を使用します。このワークロードは、イメージのデータセットで推論を実行します。
- **ResNet トレーニング** : ここでは使用可能な ResNet50 を使用します。このワークロードは、イメージのデータセットに対して ResNet50 モデル トレーニングを実行します。

パフォーマンスの測定基準として「1 秒あたりのイメージ」を使用します。トレーニングの実行では、VM とベア メタルの両方の構成で CPU スレッドの数が 16 に制限されます。3 回のトレーニングを実施しました。示されている結果は、この 3 回の実行の平均です。

次の表は、ベア メタル サーバーの構成を示しています。

テーブル 2. PowerEdge R740xd サーバーの構成

コンポーネント	構成
CPU	インテル Xeon Gold 6148 CPU 20 コア(2.40 GHz)
メモリー	256 GB
GPU	NVIDIA A100-PCIE-40 GB
NIC	ConnectX-5 DX 25 GbE
ストレージ	PowerStore Isilon H400
ソフトウェア (vSphere サーバーのみ。ベア メタルには適用されない)	VMware vSphere 7.0 U2

次の表は、VM サーバーの構成を示しています。

テーブル 3. VM 構成

コンポーネント	構成
CPU	16
メモリー	64
ストレージ	Isilon ストレージ上の 500 GB シン プロビジョニング 仮想ディスク

すべてのトレーニングの実行に対して、ESXi server で MIG を有効にしました。

結果

結果と所見は次のとおりです。

- **ベア メタル GPU と仮想化 GPU の比較**：この調査では、ResNet トレーニングを使用して、ベア メタル GPU と仮想化 GPU のパフォーマンスを比較しています。次の図に示すように、仮想化 GPU とベア メタル GPU のパフォーマンスの違いは 5%未満です。

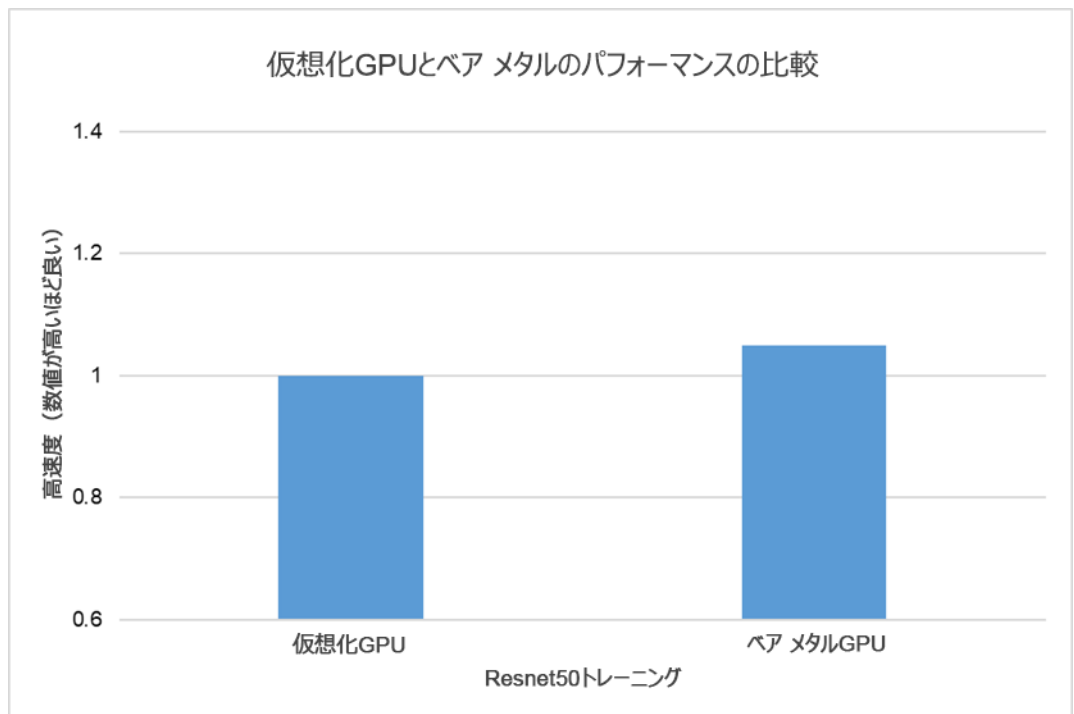


図 9. 仮想化 GPU とベア メタル GPU のパフォーマンスの比較

推奨事項については、パフォーマンスと運用の柔軟性の両方を明示的に評価しました。パフォーマンスの比較は、A100 GPU で有効になっている MIG 機能を使用して完了しました。その結果、ワークロードのパフォーマンスに対して最大で-5%の影響が確認されました。MIG が有効になっている GPU の利点は、ホスト サーバーの運用ダウンタイムが生じることなく、1 つまたは複数の VM で vGPU プロファイルを構成および再構成できる点です。ジョブを実行していない VM に割り当てられたプロファイルは、他の VM が共有 GPU の他のパーティションでワークロードを実行している間に再割り当てまたは変更できます。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

- **ResNet トレーニングを使用した MIG パフォーマンス分析**：ResNet トレーニングを使用して、MIG パーティションのパフォーマンスを調査します。次の図は 3 つのシナリオを示しています。各シナリオでは、x 軸で示すように 1 つの VM に 1 つのプロファイルが割り当てられています。

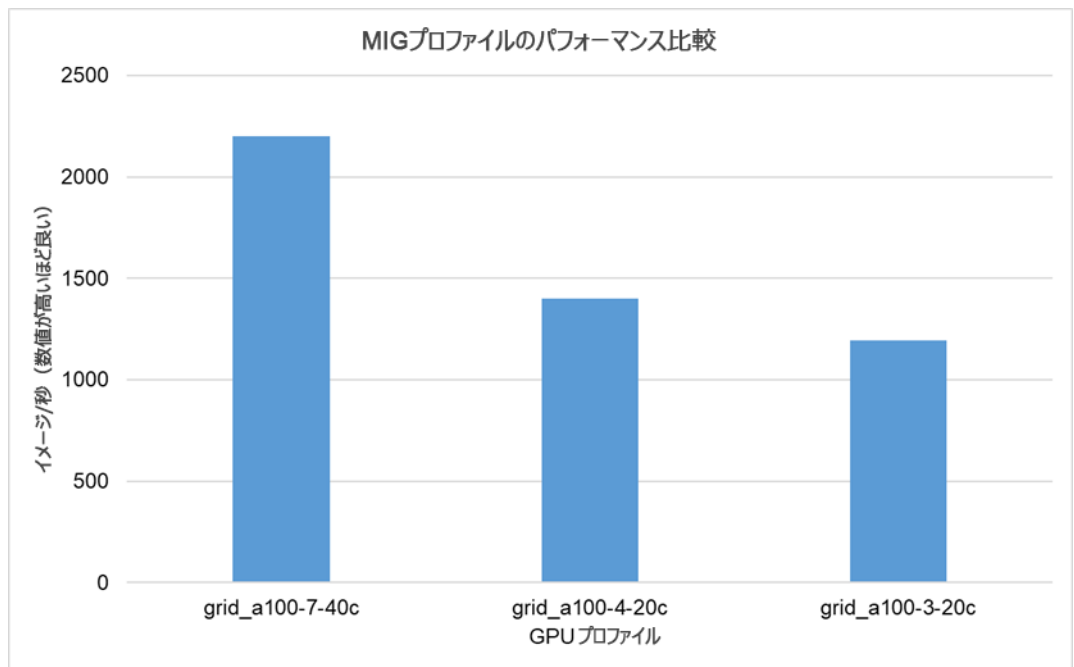


図 10. ResNet トレーニングを使用した MIG プロファイルのパフォーマンスの比較

パーティションは、そのパーティションで使用可能な専用リソースのサイズに対して実行されます。grid_a100-2-10c および grid_a100-1-5c のパーティションでは、ResNet トレーニングを実行できません。これらのパーティションは推論のみに適しており、ニューラル ネットワーク トレーニングには推奨されません。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

- **ResNet 推論を使用した MIG パフォーマンス分析**：ResNet 推論を使用して、MIG パーティションのパフォーマンスを調査します。

次の表に、8 つのシナリオを示します。

テーブル 4. 表 4 テストを実施したシナリオと使用した vGPU プロファイル

シナリオ番号	説明
シナリオ 1	grid_a100-7-40c プロファイルで構成された単一の VM
シナリオ 2	grid_a100-4-20c プロファイルで構成された単一の VM
シナリオ 3	grid_a100-3-20c プロファイルで構成された単一の VM
シナリオ 4	grid_a100-2-10c プロファイルで構成された単一の VM
シナリオ 5	grid_a100-1-5c プロファイルで構成された単一の VM
シナリオ 6	grid_a100-4-20c、grid_a100-2-10c、grid_a100-1-5c プロファイルで構成された 3 つの VM
シナリオ 7	それぞれ grid_a100-2-10c で構成された 3 つの VM と、grid_a100-1-5c プロファイルで構成された 4 つ目の VM
シナリオ 8	grid_a100-1-5c プロファイルで構成された 7 つの VM

次の図は、8 つのシナリオの結果を示しています。

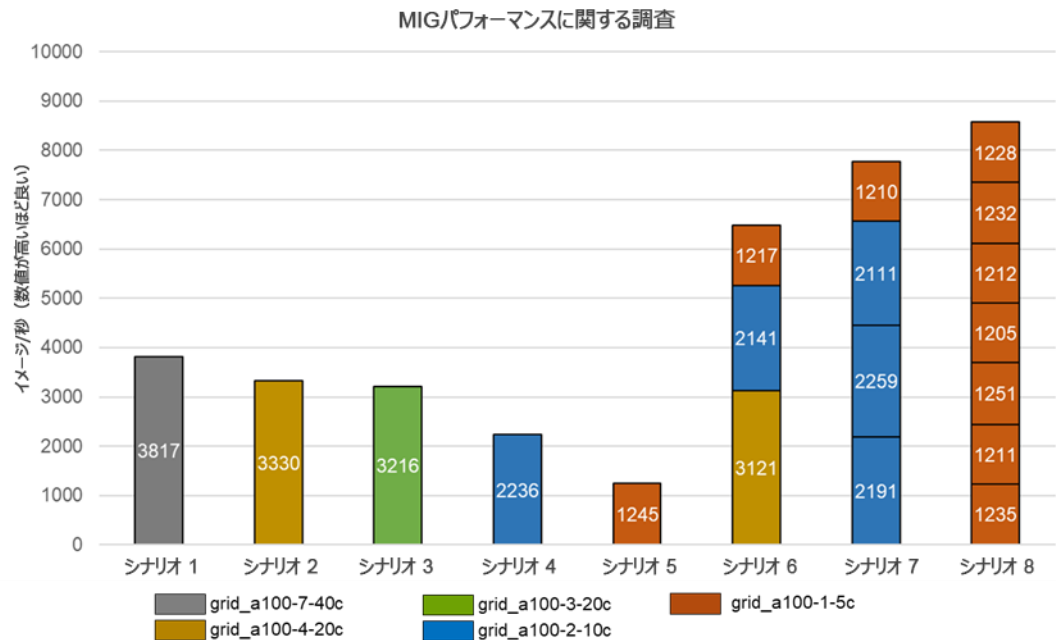


図 11.

結果から、ResNet 推論ジョブは grid_a100-7-40c および grid_a100-4-20c GPU を十分に使用できないことがわかります。したがって、他のプロファイルと比較して、確認された 1 秒あたりのイメージは多くありません。

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

これらの結果は、GPU でパーティションを使用するメリットを示しています。推論などの機械学習ジョブでは、すべての GPU リソースが消費されるわけではありません。この消費量はシナリオ 1〜3 で顕著に見られます。シナリオ 8 と比較して、1 秒あたりのイメージが少なくなっています。推論ワークロードが一部の GPU リソースを消費しないため、シナリオ 1〜3 は GPU を十分に使用していません。GPU の分割を行って VM にパーティションを割り当てることで、管理者は複数の混在ワークロードを実行できるため、GPU の使用が大幅に増加します。この調査のシナリオ 8 では、推論ワークロードに対して最も高い統合されたスループットが提供されています。

MIG パーティションは、他のパーティションの使用方法に関係なく、予測可能なパフォーマンスを提供します。たとえば、grid_a100-2-10c で構成された VM では、他のプロファイルがアイドル状態でも使用されていても同様の結果が示されます。

概要

データサイエンスのプロフェッショナルは、数十年にわたって IT インフラストラクチャの限界を押し上げてきました。デル・テクノロジーと NVIDIA は、コンピューティング、データストレージ オプション、高速ネットワーキングにおいて数多くの進化を遂げ、その課題を解決してきました。この Validated Design では、これらのハードウェア テクノロジーの進歩により効率的に対応すると同時に、最新バージョンの VMware 仮想化ソフトウェアを使用して仮想化環境において優れたパフォーマンスの向上を維持する方法について説明します。

かつてはグラフィックス処理に主に使用されていたテクノロジーをベースにしたハードウェア アクセラレーターの開発ほど、データサイエンスについてパフォーマンスが急速に向上している分野はありません。NVIDIA の最新世代の Ampere GPU は、旧世代の NVIDIA GPU と比べて最大 20 倍高いパフォーマンスを実現します。NVIDIA A100 GPU は、7 つの GPU インスタンスに分割することができ、需要の変化に合わせて動的に調整できます。GPU パーティションと VMware 仮想化を組み合わせることで、IT プロフェッショナルは、非常に要求の厳しい本番環境の機械学習ワークロードを実行するために必要なツールを利用できます。また、大規模なデータサイエンス ユーザー グループの実験ニーズに単一のプラットフォームで対応します。

「[ソリューションアーキテクチャ](#)」では、エンタープライズ AI ワークロードに推奨される構成について説明しています。25 GbE と 100 GbE の両方の機器の冗長ネットワーキング設計について説明します。また、NVIDIA Virtual GPU Manager for VMware vSphere を導入する際のベスト プラクティスと、A100 および A30 GPU のマルチインスタンス GPU 機能に基づいて GPU パーティションを使用するように VM を構成する方法についても説明しています。

「[検証と機械学習のパフォーマンス](#)」では、仮想化 GPU とベアメタル、複数の GPU 分割アプローチと A100 GPU を比較したパフォーマンス評価をサポートするために実装した実習環境について説明しています。画像分類ツール(ResNet 50 v 1.5)のトレーニングとトレーニング済みの ResNet モデルを使用した推論の両方を実行する A100 GPU に対して、異なるパーティション プロファイルを使用した 8 つのテストの結果を示します。結果から、効率性の向上と比較して、5%未満のモデル トレーニング パフォーマンス低下があることがわかっています。ハードウェアの分割で vGPU を使用すれば、ほとんどのシナリオで、AI 用の機械学習ワークロードを VMware が管理するプラットフォームに導入する IT 組織が増えるでしょう。

エラー! [ホーム] タブを使用して、ここに表示する文字列に Heading 1 を適用してください。

フィードバックを歓迎 いたします

Dell Technologies では、ソリューションやソリューション ドキュメントへのご意見をお待ちしております。Dell Technologies Solutions チームに [E メール](#)を送信するか、[ドキュメント アンケート](#)にコメントを記入してください。

作成者:

Dell Integrated Solutions Engineering : Bala Chandrasekaran、Prem Motgi Pradeep、Phil Hummel、Srinivas Varadharajan、Sarvani Vemulapalli。

貢献者 : Chris Gully (VMware OTCO)、Dharmesh Patel (Dell Integrated Solutions Engineering)、Justin King (Dell Integrated Solutions)。

メモ : このソリューションに関するその他のドキュメントのリンクについては、「[Dell Technologies Solutions Info Hub for AI and Data Analytics Workloads](#)」を参照してください。

エラー! [ホーム] タブを使用して、ここに表示する文字列に **Heading 1** を適用してください。

参考資料

デル・テクノロジーズ ドキュメント

次のデル・テクノロジーズのドキュメントには、追加情報および関連情報が記載されています。これらのドキュメントにアクセスできるかどうかは、お使いのログイン認証情報によって決まります。アクセスできないドキュメントがある場合は、デル・テクノロジーズ担当者までお問い合わせください。

- [HPC および AI イノベーション ラボ](#)
- [Dell EMC VxRail ハイパーコンバージド インフラストラクチャ](#)
- [Dell Technologies PowerSwitch データ センター スイッチ](#)
- [Dell Technologies カスタマー ソリューション センター](#)
- [Dell Technologies Solutions Info Hub for Artificial Intelligence and Data Analytics Workloads](#)

VMware リンク

次の VMware のドキュメントには、追加情報および関連情報が記載されています。

- [vSphere 7 Update 3 - What's New](#)
- ブログ : [Announcing vSphere 7 Update 3](#)
- [vSphere 7 Update 2 vGPU Operations Guide](#)
- [VMware vSAN Design Guide](#)
- ブログ : [vSphere 7 with Multi-Instance GPUs \(MIG\) on the NVIDIA A100 for Machine Learning Applications – Part 1 および Part 2](#)

NVIDIA リンク

次の Microsoft のドキュメントには、追加情報および関連情報が記載されています。

- [NVIDIA AI Enterprise Documentation](#) : このリンクには、NVIDIA AI Enterprise に関する包括的なドキュメントが含まれています。
- [NVIDIA A100 Tensor Core GPU](#)
- [NVIDIA Multi-Instance GPU User Guide](#)
- [NVIDIA-Certified Systems](#)
- [NVIDIA GPU Cloud \(NGC\)](#)
- [NVIDIA Multi-Instance GPU and NVIDIA Virtual Compute Server – Technical Brief](#)
- [NVIDIA AI Enterprise Packaging, Pricing, and Licensing Guide](#)