GPU カードの取り付け

この付録では、サポートされている GPU カードの設定ルールと取り付け手順について説明します。

- サーバ ファームウェアの要件 （1 ページ）
- GPU カードの構成規則 （2 ページ）
- すべての GPU の要件：4 GB を超えるメモリマップド I/O （5 ページ）
- シングル幅の GPU カードの取り付け （6 ページ）
- 倍幅 GPU カードの取り付け （8 ページ）
- P シリーズおよび T シリーズ GPU 用 NVIDIA GRID ライセンス サーバーの使用 （14 ページ）
- GPU カードをサポートするドライバのインストール （23 ページ）

サーバ ファームウェアの要件

次の表に、サポートされている GPU カードのサーバ ファームウェアの最小バージョンを示します。

GPU カード	Cisco IMC/BIOS の最低限必要なバージョン
C240 C240 M5 SN、SNEBS、S、SD、および SX モデルの Cisco A100-80 （UCSC-GPU-A100-80）	4.2(2)
NVIDIA Tesla A100	4.1(3)
NVIDIA Tesla M10	3.1(1)
NVIDIA Tesla P40	3.1(1)
NVIDIA Tesla P100 12GB	3.1(1)
NVIDIA Tesla P100 16GB	3.1(1)
AMD FirePro S7150 X2	3.1(1)

GPU カード	Cisco IMC/BIOS の最低限必要なバージョン
NVIDIA Tesla M60	3.1(2c)
NVIDIA Tesla P4	3.1(3)
NVIDIA Tesla V100 16 GB	3.1(3)
NVIDIA Tesla V100 32GB	4.0(1)
AMD Radeon Pro V340	4.0(2) （注） この GPU は、現時点ではスタンドアロンモードでのみサポートされています。
NVIDIA T4	4.0(2e) （注） この GPU をサポートする Cisco UCS Manager の最小バージョンは 4.0(2c) です。

GPU カードの構成規則

GPU カードを使用して、サーバを設定するときは、次の規則に注意してください。

注意 このサーバーで NVIDIA テスラ P100 12GB または P100 16GB GPU カードを使用する場合、特別な温度要件があります。倍幅 GPU カードの取り付け （8 ページ）を参照してください。

注意 このサーバーで NVIDIA テスラ P40 GPU カードを使用する場合、特別な温度要件があります。倍幅 GPU カードの取り付け （8 ページ）を参照してください。

注意 このサーバーで NVIDIA テスラ V100 16GB または V100 32GB GPU カードを使用する場合、特別な温度要件があります。倍幅 GPU カードの取り付け （8 ページ）を参照してください。

- サーバーの設定に基づいて必要な電力を判別するには、リンク http://ucspowercalc.cisco.com で UCS Power Calculator を使用してください。
- 最大 2 個の倍幅 GPU カードが PCIe ライザー 1、スロット 2 と PCIe ライザー 2、スロット 5 でサポートされています。
  - 倍幅の GPU カードはすべての PCIe ライザー オプションでサポートされていません。倍幅の GPU カードをサポートするのは以下のライザー オプションのみです。
    - PCIe ライザー 1（UCSC-R1-A100-M5）

- PCIe ライザー 1B (UCSC-RIS-1B-240M5)
- PCIe ライザー 2A（UCSC-R2A-A100-M5）
- PCIe ライザー 2B（UCSC-R2B-A100-M5）
- PCIe ライザー 2C (UCSC-RIS-2C-240M5)

- NVIDIA Tesla A100 GPU は、次のライザー オプションでサポートされています。
  - PCIe ライザー 1（UCSC-R1-A100-M5）
  - PCIe ライザー 2A（UCSC-PCI-2 a-C240M5）
  - PCIe ライザー 2B（UCSC-PCI-2 a-C240M5）

- スロット 2 に取り付けられたダブル幅 GPU カードは、スロット 4 もカバーします。スロット 5 に取り付けられたダブル幅 GPU カードは、スロット 6 もカバーします。
- NVIDIA テスラ P4：最大 6 個の P4 シングル幅 GPU カードがすべての PCIe スロットでサポートされています。P4 シングル幅の GPU カードは、すべてのライザー オプションでサポートされています。
- NVIDIA T4：
  - 4.0(4) よりも前のバージョンの Cisco IMC では、サーバーは 4 個の NVIDIA T4 GPU のみをサポートできます。この GPU をスロット 2 または 5 に取り付けないでください。サーバに CPU が 1 つしかない場合は、PCIe ライザー バージョン 1B を使用する必要があります。サーバーに 2 つの CPU がある場合は、ライザー バージョン 1 または 1B、および 2A、2B、または 2C を使用できます。
  - Cisco IMC バージョン 4.0(4) 以降では、サーバーは最大 6 個の NVIDIA T4 GPU をサポートできます。ただし、6 個のこれらの GPU をサポートするには、バージョン 2 の PCIe ライザーが必要です。次のバージョン 2 ライザーでは、サーバに最大 6 個の NVIDIA T4 GPU をロードすることをサポートしています。

発注 PID	説明
UCSC-RIS-1-240M5	ライザー 1、バージョン 2
UCSC-RIS-1B-240M5	ライザー 1B、バージョン 2
UCSC-RIS-2A-240M5	ライザー 2A、バージョン 2
UCSC-RIS-2B-240M5	ライザー 2B、バージョン 2
UCSC-RIS-2C-240M5	ライザー 2C、バージョン 2

- NVIDIA Quadro RTX 6000 および RTX 8000：

  次のライザー カードは、UCSC-GPU-RTX6000 および UCSC-GPU-RTX6000 GPU 用です。

発注 PID	説明
UCSC-RIS-1-240M5	ライザー 1、バージョン 2
UCSC-RIS-1B-240M5	ライザー 1B、バージョン 2
UCSC-RIS-2A-240M5	ライザー 2A、バージョン 2
UCSC-RIS-2B-240M5	ライザー 2B、バージョン 2
UCSC-RIS-2C-240M5	ライザー 2C、バージョン 2

- UCSC-C240-M5SN はライザー 1 の GPU のみをサポートします
- サーバで複数のブランドまたはモデルの GPU を混在させないでください。
- 同じライザーに GPU カードと Cisco UCS VIC を取り付けることができます。スロット 2 に GPU カードを取り付けると、ライザー 1 での NCSI のサポートはスロット 1 に自動的に移ります。スロット 5 に GPU カードを取り付けると、ライザー 2 での NCSI のサポートはスロット 4 に自動的に移ります。
- GPU によっては、サーバーで 1 TB 以上のメモリをサポートできるかどうかに制限があります。

表 *1*: サーバー メモリの合計サポート

GPU	サーバーでの 1 TB 以上のメモリのサポート
NVIDIA Tesla A100	はい
NVIDIA M シリーズ	非対応
NVIDIA P シリーズ	対応
NVIDIA T シリーズ	対応
NVIDIA V シリーズ	対応
AMD FirePro S7150 X2	非対応
AMD Radeon Pro V340	対応

- 次の NVIDIA GPU は、Second Generation Intel Xeon Scalable processor ではサポートされていません。
  - NVIDIA Tesla P4
  - NVIDIA Tesla P100 12G
  - NVIDIA Tesla P100 16G

すべての GPU の要件：4 GB を超えるメモリマップド I/O

サポートされるすべての GPU カードは、4 GB を超えるメモリマップド I/O（MMIO）を許可する BIOS 設定が必要です。

- スタンドアロン サーバ：サーバをスタンドアロン モードで使用する場合、この BIOS 設定はデフォルトで有効です。

  `[Advanced] > [PCI Configuration] > [Memory Mapped I/O Above 4 GB] [Enabled]`

  この設定値を変更するには、ブートアップ時に指示されたら F2 を押して BIOS セットアップ ユーティリティを使用します。

- サーバが Cisco UCS Manager と統合され、サービス プロファイルによって制御される場合に GPU が存在していると、この設定はデフォルトによりサービス プロファイルで有効になります。

  この設定を手動で変更するには、次の手順を使用します。

ステップ 1 サービス プロファイルの設定方法については、使用しているリリースの『Cisco UCS Manager Configuration Guide』（GUI または CLI）をご覧ください。

『Cisco UCS Manager Configuration Guides』

ステップ 2 「Configuring Server-Related Policies」にある「Configuring BIOS Settings」の章を参照してください。

ステップ 3 PCI 構成の BIOS の設定用のプロファイルのセクションで、[Memory Mapped IO Above 4GB Config] を次のいずれかに設定します。

- [Disabled]：64 PCI デバイスを 64 GB 以上のアドレス空間にマッピングしません。
- [Enabled]：64 ビット PCI デバイスの I/O を 64 GB 以上のアドレス空間にマッピングします。
- **Platform Default**：ポリシーはサーバの BIOS のデフォルトに含まれるこの属性の値を使用します。これは、サーバ BIOS がこの項目にデフォルトの有効化設定を使用するように設定されていることがわかっている場合にのみ使用します。

ステップ 4 サーバをリブートします。

（注） Cisco UCS Manager は、BIOS ポリシーまたはデフォルトの BIOS 設定による BIOS 設定の変更を Cisco Integrated Management Controller（CIMC）バッファにプッシュします。これらの変更はバッファ内にとどまり、サーバがリブートされるまでは有効になりません。

シングル幅の GPU カードの取り付け

以下に示すサポート対象のシングル幅の GPU カードの取り付けまたは交換には、次の手順を使用します。

- NVIDIA Tesla P4

  このサーバーは最大 6 つのシングル幅の NVIDIA Tesla P4 GPU カードをサポートできます。これらのハーフハイト、ハーフレングス（HHHL）GPU カードはすべての PCIe スロットでサポートされています。

- NVIDIA T4

  - 4.0(4) よりも前のバージョンの Cisco IMC では、サーバーは 4 個の NVIDIA T4 GPU のみをサポートできます。この GPU をスロット 2 または 5 に取り付けないでください。サーバに CPU が 1 つしかない場合は、PCIe ライザー バージョン 1B を使用する必要があります。サーバーに 2 つの CPU がある場合は、ライザー バージョン 1 または 1B、および 2A、2B、または 2C を使用できます。

  - Cisco IMC バージョン 4.0(4) 以降では、サーバーは最大 6 個の NVIDIA T4 GPU をサポートできます。ただし、6 個のこれらの GPU をサポートするには、バージョン 2 の PCIe ライザーが必要です。次のバージョン 2 ライザーでは、サーバに最大 6 個の NVIDIA T4 GPU をロードすることをサポートしています。

発注 PID	説明
UCSC-RIS-1-240M5	ライザー 1、バージョン 2
UCSC-RIS-1B-240M5	ライザー 1B、バージョン 2
UCSC-RIS-2A-240M5	ライザー 2A、バージョン 2
UCSC-RIS-2B-240M5	ライザー 2B、バージョン 2
UCSC-RIS-2C-240M5	ライザー 2C、バージョン 2

ステップ 1 サーバのシャットダウンと電源切断の説明に従って、サーバをシャットダウンして電源を切ります。

ステップ 2 上部カバーを取り外せるように、サーバをラックの前方に引き出します。場合によっては、リア パネルからケーブルを取り外して隙間を空ける必要があります。

注意 コンポーネントへの接触や確認作業を安全に行えない場合は、ラックからサーバを取り出してください。

ステップ 3 サーバ上部カバーの取り外しの説明に従ってサーバから上部カバーを取り外します。

ステップ 4 交換するシングル幅の GPU カードを取り外します。

a) 青色のライザー ハンドルと、ライザーの前端にあるつまみ部分を両手で上にあげてつかみ、まっすぐに持ち上げます。
b) ライザーの底部にある、固定プレートを固定しているリリース ラッチを押し、ヒンジ付き固定プレートを回して開きます。
c) カードのリア パネル タブを固定しているヒンジ付きカード タブ固定具を開きます。
d) シングル幅の GPU カードの両端を均等に引いて、PCIe ライザーのソケットから取り外します。

   ライザーにカードがない場合は、ライザーの背面の開口部からブランク パネルを取り外します。

ステップ 5 新しいシングル幅の GPU カードの取り付け：

（注） 現時点では、サーバーは 4 つの NVIDIA T4 GPU のみをサポートできます。この GPU をスロット 2 または 5 に取り付けないでください。サーバに CPU が 1 つしかない場合は、PCIe ライザー バージョン 1B を使用する必要があります。サーバーに 2 つの CPU がある場合は、ライザー バージョン 1 または 1B、および 2A、2B、または 2C を使用できます。

a) ヒンジ付きカードタブ固定具を開いた状態で、新しいシングル幅の GPU カードを PCIe ライザーの空ソケットの位置に合わせます。
b) カードの両端を均等に押し下げて、ソケットにしっかりと装着します。
c) カードの背面パネルタブがライザーの背面パネルの開口部に対して水平であることを確認したら、カードの背面パネル タブ上でヒンジ付きカード タブ固定具を閉じます。
d) ライザーの底部でヒンジ付き固定プレートを回して閉じます。プレートのクリップが、カチッと音がしてロック位置に収まったことを確認します。
e) PCIe ライザーを、マザーボード上のソケットとシャーシの位置合わせチャネルの上に配置します。
f) PCIe ライザーの両端を慎重に押し下げて、コネクタをマザーボード上のソケットにしっかりと収納します。

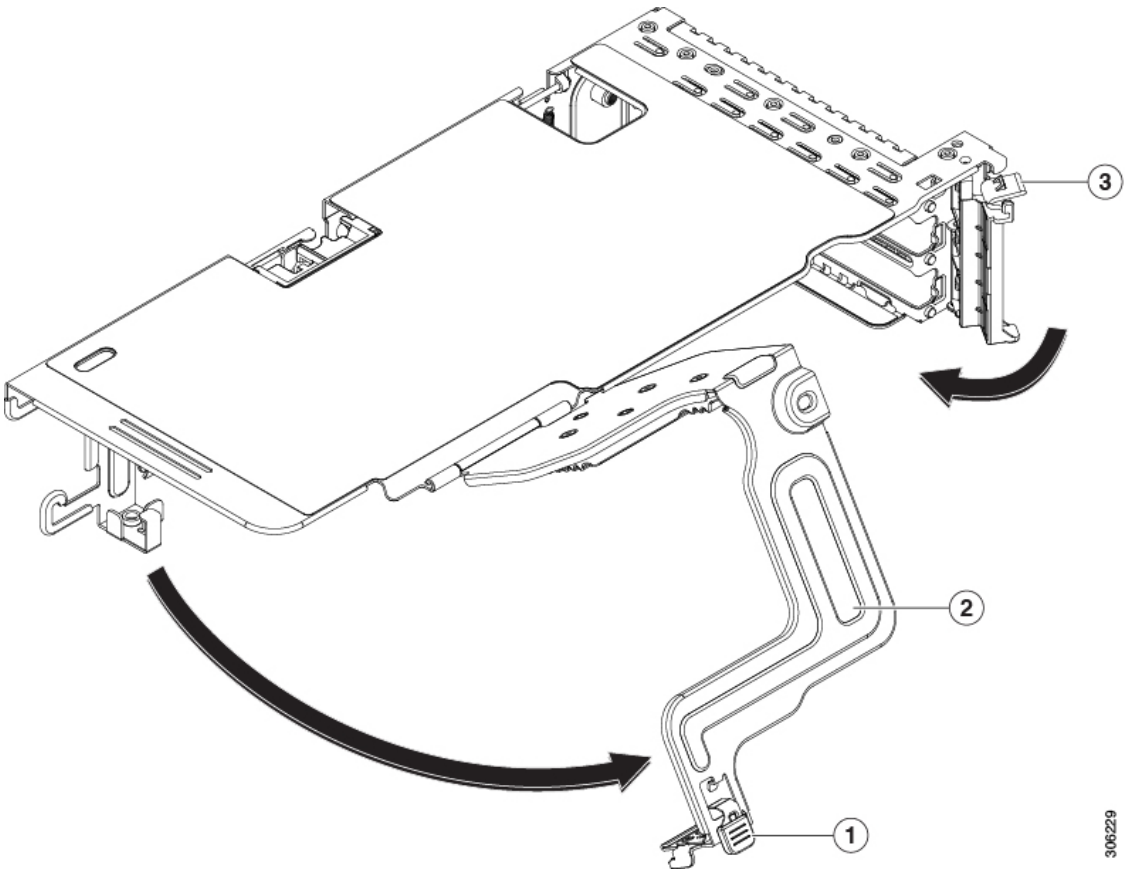
ステップ 6 上部カバーをサーバに再度取り付けます。

ステップ 7 サーバをラックの元の位置に戻し、ケーブルを再度接続したら、電源ボタンを押してサーバの電源を入れます。

ステップ 8 オプション：GPU カードをサポートするドライバのインストール （23 ページ）に進みます。

（注） NVIDIA Tesla M シリーズまたは P シリーズ GPU を取り付けた場合は、GRID 機能を使用するため GRID ライセンスをインストールする必要があります。「P シリーズおよび T シリーズ GPU 用 NVIDIA GRID ライセンス サーバーの使用 （14 ページ）」を参照してください。

図 *1 : PCIe* ライザー カードの固定機構



1 ヒンジ付き固定プレートのリリース ラッチ	3 ヒンジ付きカード タブ固定具
2 ヒンジ付き固定プレート	-

倍幅 GPU カードの取り付け

以下に示すサポート対象の倍幅 GPU カードを装着または交換するには、次の手順を使用します。

- NVIDIA Tesla A100
- NVIDIA Tesla M10
- NVIDIA Tesla M60
- NVIDIA Tesla P40
- NVIDIA Tesla P100 12 GB

- NVIDIA Tesla P100 16 GB
- NVIDIA Tesla V100 16 GB
- NVIDIA Tesla V100 32GB
- AMD FirePro S7150 X2
- AMD Radeon Pro V340

（注） この GPU は、現時点ではスタンドアロン モードでのみサポートされています。

（注） このサーバーで NVIDIA テスラ P100 12GB または P100 16GB GPU カードを使用する場合、次の表で説明されているように特別な温度要件があります。

注意 このサーバーで NVIDIA テスラ P40 GPU カードを使用する場合、次の表で説明されているように特別な温度要件があります。

（注） このサーバーで NVIDIA テスラ V100 16GB または V100 32GB GPU カードを使用する場合、次の表で説明されているように特別な温度要件があります。

表 *2 :* 倍幅 *GPU* カードの *C240 M5* 動作温度要件

GPU カード	最高サーバ動作温度（吸気口温度）
NVIDIA Tesla A100	32°C（89.6°F）
NVIDIA Tesla M10	35 ° C（95.0 ° F）
NVIDIA Tesla M60	35 ° C（95.0 ° F）
NVIDIA Tesla P40	32°C（89.6°F）
NVIDIA Tesla P100 12GB	28°C（82.4°F）
NVIDIA Tesla P100 16GB	28°C（82.4°F）

NVIDIA Tesla V100 16GB	サーバのバージョンによって異なります。 UCSC-C240-M5S：28°C（82.4°F） UCSC-C240-M5SX：28°C（82.4°F） UCSC-C240-M5SN：28°C（82.4°F） UCSC-C240-M5L（最大 8 個の LFF ドライブのみ。8 個以上の LFF ドライブはサポートされていません）：24°C（75.2°F）
NVIDIA Tesla V100 32GB	サーバのバージョンによって異なります。 UCSC-C240-M5S：28°C（82.4°F） UCSC-C240-M5SX：28°C（82.4°F） UCSC-C240-M5SN：28°C（82.4°F） UCSC-C240-M5L（最大 8 個の LFF ドライブのみ。8 個以上の LFF ドライブはサポートされていません）：24°C（75.2°F）
AMD FirePro S7150 X2	35 ° C（95.0 ° F）
AMD Radeon Pro V340	35°C（95.0°F）

（注） **NVIDIA GPU**：NVIDIA GPU カードは、ストレート ケーブルと Y 字型ケーブルの 2 本の電源ケーブルが付属していることがあります。このサーバで GPU カードに電源を接続するにはストレート ケーブルを使用します。Y 字型ケーブルは使用しないでください。Y 字型ケーブルは、外部デバイスの GPU カードを接続する場合にのみ使用します（Magma シャーシなど）。

AMD GPU：このサーバーの正しい電源ケーブルは Y 字型ケーブルです。

次の表では、GPU で使用されるケーブルがリストされています。また、ケーブルが GPU BOM に含まれているかどうか、または別途注文する必要があるかどうかも示されています。

- 別途 = ordering tool プロンプトが表示される場合はケーブルを別途注文する必要があります。
- 付属 = ケーブルは GPU に付属しています。追加のアクションは必要ありません。

表 *3*：倍幅 *GPU* に必要な電源ケーブル

GPU	GPU 電源ケーブル	システムの注文と一緒に GPU カードを注文した場合、ケーブルは含まれていますか。	予備として GPU カードを注文した場合、ケーブルは含まれていますか。
NVIDIA Tesla A100	UCS-P100CBL-240M5	同梱	分離
Nvidia Tesla M10	UCS-M10CBL-C240M5	同梱	分離

GPU	GPU 電源ケーブル	システムの注文と一緒に GPU カードを注文した場合、ケーブルは含まれていますか。	予備として GPU カードを注文した場合、ケーブルは含まれていますか。
Nvidia Tesla P40	UCS-P40CBL-240M5	同梱	分離
Nvidia Tesla P100 12GB	UCS-P100CBL-C240M5	同梱	分離
Nvidia Tesla P100 16GB	UCS-P100CBL-C240M5	同梱	分離
Nvidia Tesla M60	UCS-P40CBL-C240M5	同梱	分離
Nvidia Tesla V100 16GB	UCS-P100CBL-C240M5	同梱	同梱
Nvidia Tesla V100 32GB	UCS-P100CBL-C240M5	同梱	同梱
AMD FirePro S7150 X2	UCS-AMDCBL-C240M5	同梱	分離
AMD Radeon Pro V340	UCS-V340CBL-C240M5	同梱	同梱

ステップ 1 サーバのシャットダウンと電源切断の説明に従って、サーバをシャットダウンして電源を切ります。

ステップ 2 上部カバーを取り外せるように、サーバをラックの前方に引き出します。場合によっては、リア パネルからケーブルを取り外して隙間を空ける必要があります。

注意 コンポーネントへの接触や確認作業を安全に行えない場合は、ラックからサーバを取り出してください。

ステップ 3 サーバ上部カバーの取り外しの説明に従ってサーバから上部カバーを取り外します。

ステップ 4 既存の GPU カードを取り外します。

a) 両手で PCIe ライザーの金属製ブラケットを持ってまっすぐ持ち上げ、マザーボード上のソケットからコネクタを外します。ライザーを静電気防止素材の上に置きます。

b) ライザーの底部で、固定プレートを留めているクリップを押し下げます。

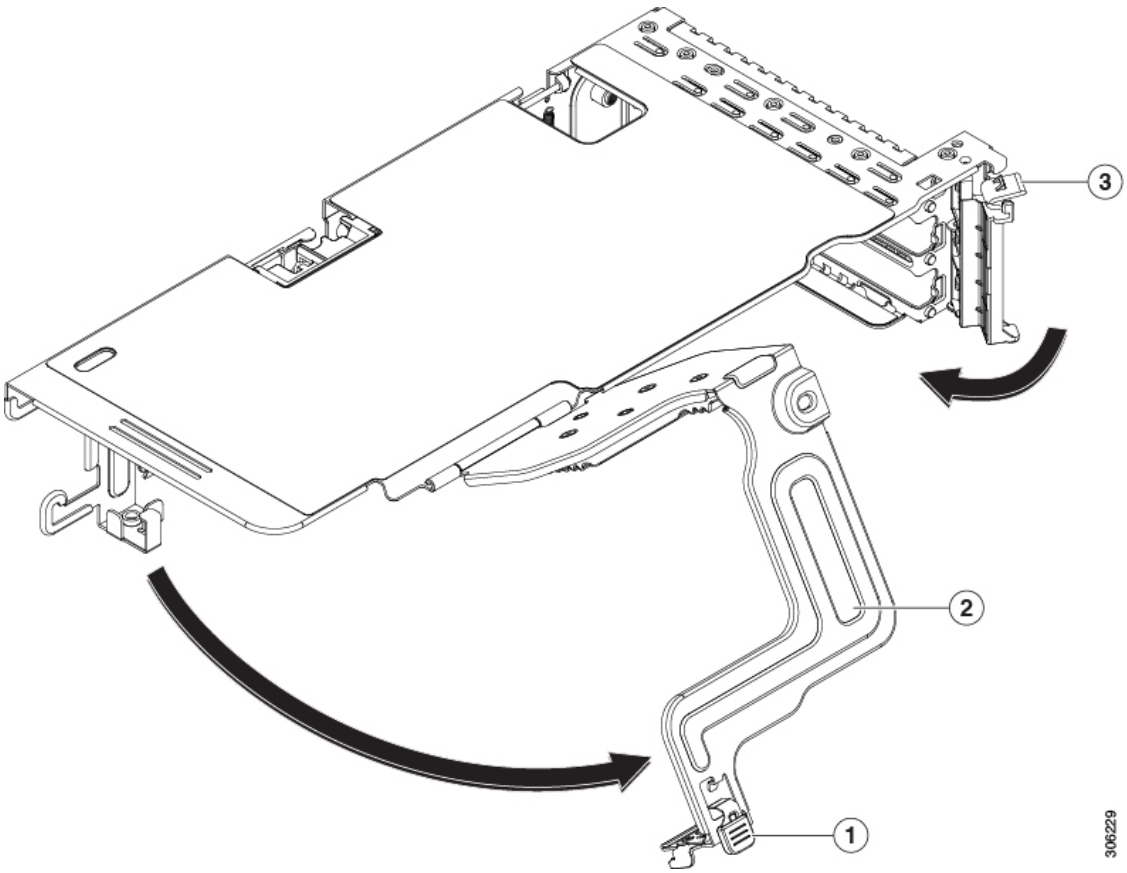
c) ヒンジ付き固定プレートを回して開け、中にアクセスできるようにします。

d) カードの背面パネル タブを固定しているヒンジ付きプラスチック製固定具を開きます。

e) PCIe ライザーの電源コネクタから GPU カードの電源ケーブルを外します。

f) GPU カードの両端を均等に引いて、PCIe ライザーのソケットから取り外します。

図 *2 : PCIe* ライザー カードの固定機構



1 ヒンジ付き固定プレートのリリース ラッチ	3 ヒンジ付きカード タブ固定具
2 ヒンジ付き固定プレート	-

ステップ 5 新しい GPU カードを取り付けます。

（注） GPU カードの構成規則 （2 ページ）で説明されているこのサーバの設定ルールに従います。

a) GPU カードをライザーのソケットの位置に合わせ、カードのエッジ コネクタをソケットにゆっくりと押し込みます。コネクタに支障をきたすことを防ぐためにカードの両隅を均等に押します。

b) GPU の電源コードを接続します。ストレート電源ケーブル コネクタは色分けされています。ケーブルの黒いコネクタを GPU カードの黒いコネクタに、ケーブルの白いコネクタを PCIe ライザーの白い GPU POWER コネクタに接続します。

注意 ストレート電源ケーブルを逆向きに接続しないでください。ケーブルの黒いコネクタは GPU カードの黒いコネクタに接続します。コードの白いコネクタを PCIe ライザーの白いコネクタに接続します。

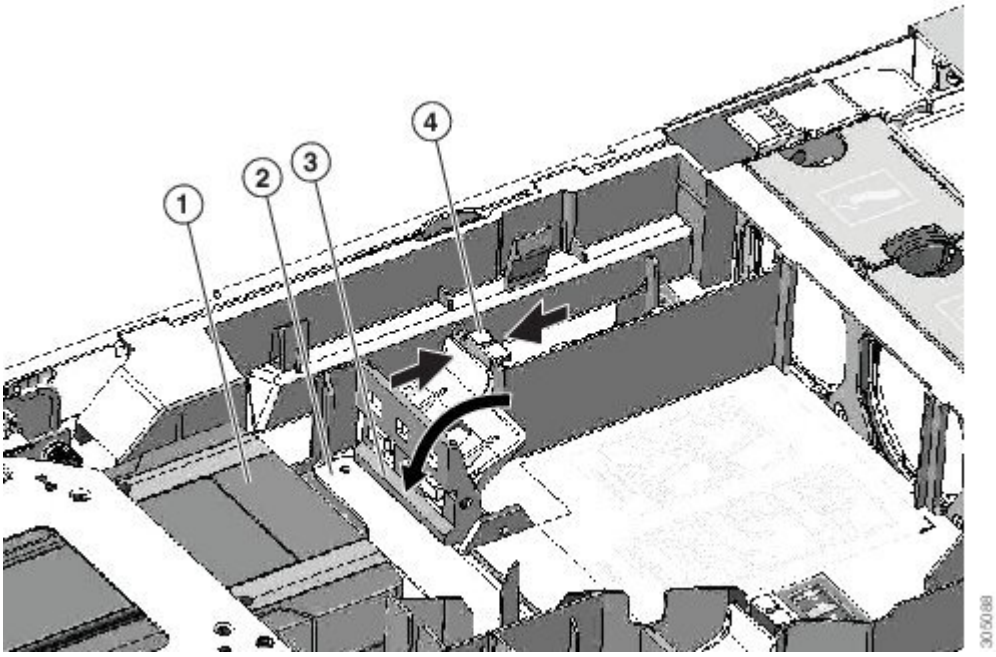
c) カードの端のカード タブ固定具を閉じます。

d) ライザーの底部でヒンジ付き固定プレートを回して閉じます。プレートのクリップが、カチッと音がしてロック位置に収まったことを確認します。
e) PCIe ライザーを、マザーボード上のソケットとシャーシの位置合わせチャネルの上に配置します。
f) PCIe ライザーの両端を慎重に押し下げて、コネクタをマザーボード上のソケットにしっかりと収納します。

同時に、（GPU カードのフロントエンドにある）GPU の前面支持ブラケットを、サーバのエアーバッフルにある固定ラッチの位置に合わせます。

ステップ 6 GPU の前面支持ブラケットをエアー バッフル上のラッチに差し込みます。
a) ラッチ リリース タブをつまみ、ラッチをサーバの前面に向け倒します。
b) ラッチを後方に向けて倒します。これにより、ラッチのへりが、GPU の前面支柱ブラケットの端にかぶるように閉じます。
c) ラッチ リリース タブがカチッと音がしてラッチが所定の位置に固定されたことを確認します。

図 *3*：エアー バッフルの固定ラッチに差し込まれた *GPU* の前面支持ブラケット



1 GPU カードのフロントエンド	3 固定ラッチのへり
2 GPU の前面支持ブラケット	4 固定ラッチ リリース タブ

ステップ 7 上部カバーをサーバに再度取り付けます。

ステップ 8 サーバをラックの元の位置に戻し、ケーブルを再度接続したら、電源ボタンを押してサーバの電源を入れます。

ステップ 9 オプション：GPU カードをサポートするドライバのインストール （23 ページ）に進みます。

（注） NVIDIA Tesla M シリーズまたは P シリーズ GPU を取り付けた場合は、GRID 機能を使用するため GRID ライセンスをインストールする必要があります。「P シリーズおよび T シリーズ GPU 用 NVIDIA GRID ライセンス サーバーの使用 （14 ページ）」を参照してください。

P シリーズおよび T シリーズ GPU 用 NVIDIA GRID ライセンス サーバーの使用

この項の内容は、NVIDIA Tesla P シリーズと T シリーズの GPU に適用されます。

NVIDIA グリッドライセンスを取得し、使用する際は、このセクションのトピックを以下の順序で利用してください。

1. NVIDIA グリッド ライセンス サーバについて理解する。

   NVIDIA グリッド ライセンス サーバの概要 （14 ページ）

2. NVIDIA に製品アクティベーション キーを登録する。

   NVIDIA での製品アクティベーション キーの登録 （16 ページ）

3. グリッド ソフトウェア スイートをダウンロードする。

   GRID ソフトウェア スイートのダウンロード （16 ページ）

4. ホストにグリッド ライセンス サーバ ソフトウェアをインストールする。

   NVIDIA GRID ライセンス サーバ ソフトウェアのインストール （16 ページ）

5. NVIDIA ライセンス ポータルでライセンスを生成し、ダウンロードする。

   NVIDIA ライセンス ポータルからライセンス サーバへのグリッド ライセンスのインストール （19 ページ）

6. グリッド ライセンスを管理する。

   グリッド ライセンスの管理 （20 ページ）

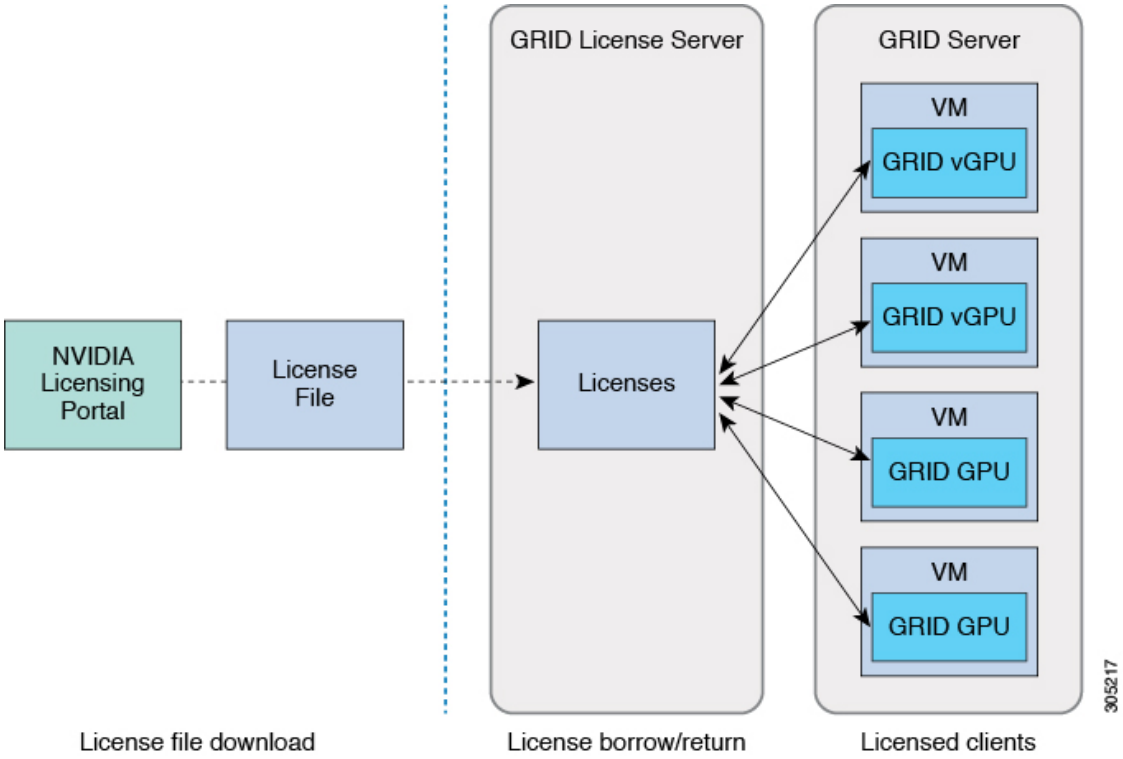
NVIDIA グリッド ライセンス サーバの概要

グリッド vGPU やグリッド仮想ワークステーションなどのライセンス済みグリッド機能が有効になると、NVIDIA M シリーズ GPU は Tesla とグリッド機能を統合します。これらの機能は、OS のブート中に NVIDIA グリッド ライセンス サーバ仮想アプライアンスからネットワーク上で実行するソフトウェア ライセンスを借用して有効になります。ライセンスは、OS がシャットダウンされたときにライセンス サーバに戻されます。

GPU カードの取り付け

グリッド ライセンス サーバで提供されるライセンスを、ダウンロード可能なライセンス ファイルとして NVIDIA のライセンス ポータルから取得します。ユーザはこのライセンス ファイルをグリッド ライセンス サーバに管理インターフェイスを用いてインストールします。

図 *4 : NVIDIA* グリッド ライセンス アーキテクチャ



グリッド ライセンスには、3 つのエディションがあり、それぞれグリッド機能の 3 つの異なるクラスを有効にします。グリッド ソフトウェアは、使用している機能に基づいてライセンス エディションを自動的に選択します。

グリッド ライセンス エディション	グリッド機能
グリッド仮想 GPU（vGPU）	ビジネス デスクトップ コンピューティング向け仮想 GPU
グリッド仮想ワークステーション	ミッドレンジ ワークステーション コンピューティング向け仮想 GPU
グリッド仮想ワークステーション（拡張）	ハイエンド ワークステーション コンピューティング向け仮想 GPU GPU パススルーのワークステーション グラフィック

NVIDIA での製品アクティベーション キーの登録

注文が処理されると、NVIDIA から製品アクティベーション キー（PAK）と、購入したライセンスの種類と数量の一覧が記載されたウェルカム電子メールが送信されます。

ステップ 1 **[ログイン（Log In）]** リンクを選択します。まだアカウントを持っていない場合には **[登録（Register）]** リンクを選択します。

[NVIDIA ソフトウェア ライセンス センター（NVIDIA Software Licensing Center）] > [ライセンス キーの登録（License Key Registration）] ダイアログが開きます。

ステップ 2 [ライセンス キーの登録（License Key Registration）] フォームに入力し、**[ライセンス登録情報を送信（Submit My Registration Information）]** をクリックします。

[NVIDIA Software Licensing Center] > [Product Information Software] ダイアログが開きます。

ステップ 3 追加の PAK があれば、**[Register Additional Keys]** をクリックします。追加の各キーについて、[License Key Registration] ダイアログのフォームに入力し、**[Submit My Registration Information]** をクリックします。

ステップ 4 プロンプトが表示されたら、利用規約に同意し、パスワードを設定します。

GRID ソフトウェア スイートのダウンロード

ステップ 1 [NVIDIA Software Licensing Center] > [Product Information Software] ダイアログボックスに戻ります。

ステップ 2 **[Current Releases]** タブをクリックします。

ステップ 3 **[NVIDIA GRID]** リンクをクリックして、[Product Download] ダイアログにアクセスします。このダイアログには次のダウンロードのリンクが含まれます。

- NVIDIA License Manager ソフトウェア
- gpumodeswitch ユーティリティ
- ホスト ドライバ ソフトウェア

ステップ 4 これらのリンクを使用して、ソフトウェアをダウンロードします。

NVIDIA GRID ライセンス サーバ ソフトウェアのインストール

完全なインストール手順とトラブルシューティングについては、『*NVIDIA GRID LicenseServer User Guide*』を参照してください。また、ご使用のリリースの最新情報については、『*NVIDIA GRID License Server Release Notes*』を参照してください。

http://www.nvidia.com

NVIDIA グリッド ライセンス サーバのプラットフォーム要件

- ホスティング プラットフォームには、物理マシンまたは仮想マシンを使用できます。NVIDIA では、ライセンス サーバの実行専用のホストを使用することをお勧めします。
- ホスティング プラットフォームでは、サポートされている Windows OS を実行する必要があります。
- ホスティング プラットフォームには、固定 IP アドレスが必要です。
- ホスティング プラットフォームには、少なくとも 1 つの固定イーサネット MAC アドレスが必要です。
- ホスティング プラットフォームの日時は、正確に設定する必要があります。

Windows でのグリッド ライセンス サーバのインストール

ライセンス サーバには、Java ランタイム環境と Apache Tomcat のインストールが必要です。Apache Tomcat は、Windows 向け NVIDIA インストール ウィザードを使用するときにインストールされます。

ステップ 1 最新の Java 32 ビットランタイム環境を https://www.oracle.com/downloads/index.html からダウンロードしてインストールします。

（注） プラットフォームが 32 ビットまたは 64 ビット Windows のいずれであっても、32 ビット Java ランタイム環境をインストールします。

ステップ 2 サーバのインターフェイスを作成します。

a) [NVIDIA Software Licensing Center] ダイアログで、**[グリッドのライセンス（Grid Licensing）] > [ライセンス サーバの作成（Create License Server）]** の順にクリックします。

b) [Create Server] ダイアログで、目的のサーバの詳細を入力します。

c) インストール用にライセンス サーバに生成された .bin ファイルを保存します。

ステップ 3 前の手順でダウンロードした NVIDIA ライセンス サーバのインストーラ zip ファイルを解凍して、setup.exe を実行します。

ステップ 4 NVIDIA ライセンス サーバ ソフトウェアと Apache Tomcat ソフトウェアの EULA に同意します。Tomcat は、ライセンス サーバのインストール時に自動的にインストールされます。

ステップ 5 インストーラ ウィザードを使用して、インストールの手順を実行します。

（注） [ファイアウォール オプションの選択（Choose Firewall options）] ダイアログで、ファイアウォールで開くポートを選択します。NVIDIA では、ポート 7070 を開き、ポート 8080 を閉じたままにしておく、デフォルト設定を使用することを推奨します。

ステップ 6 インストールを確認します。ライセンス サーバ ホストの Web ブラウザを開き、URL http://localhost:8080/licserver に接続します。インストールが正常に完了したら、NVIDIA ライセンス クライアント マネージャ インターフェイスが表示されます。

Linux でのグリッド ライセンス サーバのインストール

ライセンス サーバには、Java ランタイム環境と Apache Tomcat のインストールが必要です。Linux でライセンス サーバをインストールする前に、両方を個別にインストールする必要があります。

ステップ 1 Java が Linux インストール環境にインストールされていることを確認します。次のコマンドを使用します。

```
java -version
```

Java バージョンが表示されない場合は、Linux Package Manager を使用して、次のコマンドでインストールを行います。

```
sudo yum install java
```

ステップ 2 Linux Package Manager を使用して、Tomcat および tomcat-webapps パッケージをインストールします。

a) 次のコマンドを使用して Tomcat をインストールします。

```
sudo yum install tomcat
```

b) 次のコマンドで Tomcat サービスのブート時の自動開始を有効にします。

```
sudo systemctl start tomcat.service
```

c) Tomcat サービスを開始します。

```
sudo systemctl enable tomcat.service
```

d) Tomcat サービスが使用可能であることを確認します。ライセンス サーバ ホストの Web ブラウザを開き、URL http://localhost:8080 に接続します。インストールが正常に完了したら、Tomcat webapp が表示されます。

ステップ 3 ライセンス サーバをインストールします。

a) 次のコマンドを使用して、ライセンス サーバの tar ファイルを展開します。

```
tar xfz NVIDIA linux 2015.09 0001.tgz
```

b) root として、展開したセットアップ バイナリを実行します。

```
sudo ./setup.bin
```

c) EULA に同意し、インストール ウィザードを続行してインストールを終了します。

（注） [ファイアウォール オプションの選択（Choose Firewall options）] ダイアログで、ファイアウォールで開くポートを選択します。NVIDIA では、ポート 7070 を開き、ポート 8080 を閉じたままにしておく、デフォルト設定を使用することを推奨します。

ステップ 4 インストールを確認します。ライセンス サーバ ホストの Web ブラウザを開き、URL http://localhost:8080/licserver に接続します。インストールが正常に完了したら、NVIDIA ライセンス クライアント マネージャ インターフェイスが表示されます。

NVIDIA ライセンス ポータルからライセンス サーバへのグリッド ライセンスのインストール

グリッド ライセンス サーバ管理インターフェイスへのアクセス

ライセンス サーバ ホストの Web ブラウザを開き、URL http://localhost:8080/licserver にアクセスします。

ライセンス サーバへのリモート アクセスを許可するようにライセンス サーバ ホストのファイアウォールを設定した場合は、管理インターフェイスに http://hostname:8080/licserver の URL でリモート マシンからアクセスできます。

ライセンス サーバの MAC アドレスの読み取り

ライセンス サーバのイーサネット MAC アドレスは NVIDIA のライセンス ポータルでライセンス サーバを登録するときに ID として使用されます。

ステップ 1 ブラウザで GRID ライセンス サーバ管理インターフェイスにアクセスします。

ステップ 2 左側の [ライセンス サーバ（License Server）] パネルで **[設定（Configuration）]** を選択します。

[ライセンス サーバの設定（License Server Configuration)] パネルが開きます。**[サーバのホスト ID（Server host ID）]** の横のプルダウン メニューに、選択可能なイーサネット MAC アドレスがリストされます。

ステップ 3 **[サーバのホスト ID（Server host ID）]** プルダウンからライセンス サーバの MAC アドレスを選択します。

（注） NVIDIA のライセンス ポータルでライセンスを生成する場合には、サーバを識別するために一貫して同じイーサネット ID を使用することが重要です。NVIDIA では、プラットフォーム上の、削除できないプライマリ イーサネット インターフェイス用にエントリを 1 つ選択することを推奨しています。

ライセンス ポータルからのライセンスのインストール

ステップ 1 ブラウザで GRID ライセンス サーバ管理インターフェイスにアクセスします。

ステップ 2 左側の [ライセンス サーバ（License Server）] パネルで **[設定（Configuration）]** を選択します。

[ライセンス サーバの設定（License Server Configuration)] パネルが開きます。

ステップ 3 前に生成した .bin ファイルをインストールするには、[License Server Configuration] メニューを使用します。

a) [ファイルを選択（Choose File）] をクリックします。
b) インストールするライセンス .bin ファイルを参照して、**[Open]** をクリックします。
c) [アップロード（Upload）] をクリックします。

ライセンス サーバにライセンス ファイルがインストールされます。インストールが完了すると、「Successfully applied license file to license server」という確認メッセージが表示されます。

使用可能なグリッド ライセンスの表示

インストールされて利用可能なライセンスとそのプロパティを表示するには、次の手順を使用します。

ステップ 1 ブラウザで GRID ライセンス サーバ管理インターフェイスにアクセスします。

ステップ 2 左側の [ライセンス サーバ（License Server）] パネルで **[ライセンス機能の使用（Licensed Feature Usage）]** を選択します。

ステップ 3 **[機能（Feature）]** 列の機能をクリックすると、その機能の現在の使用に関する詳細情報が表示されます。

現在のライセンスの使用状況の表示

現在使用中であり、サーバから交付されているライセンスに関する情報を表示するには、次の手順を実行します。

ステップ 1 ブラウザで GRID ライセンス サーバ管理インターフェイスにアクセスします。

ステップ 2 左側の [ライセンス サーバ（License Server）] パネルで **[ライセンスされたクライアント（Licensed Clients）]** を選択します。

ステップ 3 シングル ライセンス クライアントに関する詳細情報を表示するには、リストの **[クライアント ID（Client ID）]** をクリックします。

グリッド ライセンスの管理

グリッド ライセンスを必要とする機能は、グリッド ライセンスを取得するまで、機能が限定された状態で実行されます。

Windows での GRID ライセンスの取得

ステップ 1 次のいずれかの方法を使用して NVIDIA コントロール パネルを開きます。

- Windows デスクトップを右クリックして、メニューから **[NVIDIA Control Panel]** を選択します。
- Windows コントロール パネルを開き、**[NVIDIA Control Panel]** アイコンをダブルクリックします。

ステップ 2 NVIDIA コントロール パネルの左側のペインで、[Licensing] の下の **[Manage License]** を選択します。

[Manage License] タスク ペインが開き、現在使用されているライセンス エディションが表示されます。グリッド ソフトウェアは、使用している機能に基づいてライセンス エディションを自動的に選択します。デフォルトは、Tesla（ライセンスなし）です。

ステップ 3 グリッド仮想ワークステーションのライセンスを取得するには、[License Edition] で **[GRID Virtual Workstation]** を選択します。

ステップ 4 **[ライセンス サーバ（License Server）]** フィールドに、ローカルの GRID ライセンス サーバのアドレスを入力します。アドレスには、ドメイン名または IP アドレスを指定できます。

ステップ 5 **[Port Number]** フィールドに、サーバが使用するポート番号を入力するか、またはデフォルト設定（7070）のままにしておきます。

ステップ 6 **[適用（Apply）]** を選択します。

システムは、設定されているライセンス サーバから適切なライセンス エディションを要求します。ライセンスが正常に取得されると、そのライセンス エディションの機能が有効になります。

（注） [NVIDIA Control Panel（NVIDIA コントロール パネル）] でライセンスを設定すると、その設定はリブート後も保持されます。

Linux での GRID ライセンスの取得

ステップ 1 コンフィギュレーション ファイル `/etc/nvidia/gridd.conf` を編集します。

```
sudo vi /etc/nvidia/gridd.conf
```

ステップ 2 ローカル グリッド ライセンス サーバのアドレスを使用して ServerUrl の行を編集します。

アドレスには、ドメイン名または IP アドレスを指定できます。次のファイルの例を参照してください。

ステップ 3 コロンを使用してアドレスの最後にポート番号（デフォルトは 7070）を追加します。次のファイルの例を参照してください。

ステップ 4 ライセンス タイプの整数を使用して FeatureType の行を編集します。次のファイルの例を参照してください。

- グリッド vGPU = 1
- グリッド仮想ワークステーション = 2

ステップ 5 nvidia-gridd サービスを再起動します。

```
sudo service nvidia-gridd restart
```

サービスは自動的に、FeatureType 行に指定したライセンス エディションを取得します。これは、/var/log/messages で確認できます。

（注） NVIDIA コントロール パネルでライセンスを設定すると、その設定はリブート後も保持されます。

サンプル コンフィギュレーション ファイル：

```
# /etc/nvidia/gridd.conf - Configuration file for NVIDIA Grid Daemon
# Description: Set License Server URL
# Data type: string
# Format: "<address>:<port>"
ServerUrl=10.31.20.45:7070

# Description: Set Feature to be enabled
# Data type: integer
# Possible values:
# 1 => for GRID vGPU
# 2 => for GRID Virtual Workstation
FeatureType=2
```

gpumodeswitch の使用

コマンド ライン ユーティリティ gpumodeswitch は、次の環境で実行できます。

- Windows 64 ビットのコマンド プロンプト（管理者権限が必要）
- Linux 32/64 ビット シェル（Citrix XenServer dom0 を含む）（ルート権限が必要）

（注） コンピューティング モードおよびグラフィック モードとの互換性の最新情報については、NVIDIA 製品のリリース ノートを参照してください。

gpumodeswitch ユーティリティでは、次のコマンドがサポートされています。

- -listgpumodes

  このコマンドは、現在の作業ディレクトリにある listgpumodes.txt というログ ファイルに情報を書き込みます。

- --gpumode graphics

  グラフィック モードに切り替えます。プロンプトが表示された際に、特別に指定しない限り、サーバでサポートされているすべての GPU のモードを切り替えます。

- --gpumode compute

  コンピューティング モードに切り替えます。プロンプトが表示された際に、特別に指定しない限り、サーバでサポートされているすべての GPU のモードを切り替えます。

（注） GPU モードを切り替えた後、サーバを再起動して、GPU の修正したリソースがサーバで実行されている OS またはハイパーバイザによって正しく認識されることを確認してください。

GPU カードをサポートするドライバのインストール

ハードウェアの取り付け後、サーバ BIOS を適切なレベルに更新し、ドライバなどのソフトウェアを次の順序でインストールする必要があります。

1. サーバ BIOS を更新します。
2. GPU ドライバを更新します。

1. サーバ BIOS の更新

Host Upgrade Utility を使用して、最新の Cisco UCS C240 M4 サーバ BIOS を Cisco UCS C240 M4 サーバにインストールします。

（注） NVIDIA ドライバを更新する前に、次の手順を実行する必要があります。

ステップ 1 http://www.cisco.com/cisco/software/navigator.html にアクセスします。

ステップ 2 中央の列の [サーバ - ユニファイドコンピューティング（Servers - Unified Computing）] をクリックします。

ステップ 3 右側の列の [Cisco UCS C シリーズラックマウントスタンドアロンサーバソフトウェア（UCS C-Series Rack-Mount Standalone Server Software）] をクリックします。

ステップ 4 右側のカラムでお使いのサーバのモデルの名前をクリックします。

ステップ 5 [Unified Computing System（UCS）サーバソフトウェア（Unified Computing System (UCS) Server Firmware）] をクリックします。

ステップ 6 リリース番号をクリックします。

ステップ 7 [今すぐダウンロード（Download Now）] をクリックして `ucs-server platform-huu-version_number.iso` ファイルをダウンロードします。

ステップ 8 次のページで情報を確認した後、[ダウンロードを続行する（Proceed With Download）] をクリックします。

ステップ 9 次の画面に進んでライセンス契約に同意し、このファイルを保存する場所を参照します。

ステップ 10 サーバ BIOS を更新するには、Host Upgrade Utility を使用します。

Host Upgrade Utility のユーザ ガイドは、『Utility User Guides』を参照してください。

2. GPU カード ドライバの更新

サーバ BIOS を更新したら、ハイパーバイザ仮想マシンに GPU ドライバをインストールできます。

ステップ 1 コンピュータにハイパーバイザ ソフトウェアをインストールします。インストール手順については、ハイパーバイザのマニュアルを参照してください。

ステップ 2 ハイパーバイザ内で仮想マシンを作成します。手順については、ハイパーバイザのマニュアルを参照してください。

ステップ 3 仮想マシンにGPU ドライバをインストールします。ドライバを次のいずれかのサイトからダウンロードします。

- NVIDIA エンタープライズ ポータル、GRID ハイパーバイザ ダウンロード（NVIDIA ログインが必要です）： https://nvidia.flexnetoperations.com/
- NVIDIA パブリック ドライバ エリア：http://www.nvidia.com/Download/index.aspx
- AMD：http://support.amd.com/en-us/download

ステップ 4 サーバを再起動します。

ステップ 5 仮想マシンがGPU カードを認識できることを確認します。Windows では、[デバイス マネージャー（Device Manager）] の [ディスプレイ アダプター（Display Adapters）] から確認します。

翻訳について
このドキュメントは、米国シスコ発行ドキュメントの参考和訳です。リンク情報につきましては、日本語版掲載時点で、英語版にアップデートがあり、リンク先のページが移動/変更されている場合がありますことをご了承ください。あくまでも参考和訳となりますので、正式な内容については米国サイトのドキュメントを参照ください。