

GPU コンピューティングカード

目次

はじめに	1
1way サーバ向け製品編	2
1. 機能仕様.....	3
1.1. NVIDIA A シリーズ	3
2. 外観図.....	4
2.1. N8105-64 GPU コンピューティングカード(NVIDIA A2)	4
3. 関連オプション.....	5
3.1. PCI 用 FAN オプション	5
2way サーバ向け製品編	6
1. 機能仕様.....	7
1.1. Tesla シリーズ	7
1.2. NVIDIA A シリーズ	9
2. 外観図.....	11
2.1. N8105-51 GPU コンピューティングカード	11
2.2. N8105-54 GPU コンピューティングカード(Tesla T4)	12
2.3. N8105-55 GPU コンピューティングカード(Tesla V100)/ N8105-56 GPU コンピューティングカード (Tesla V100S).....	13
2.4. N8105-58 GPU コンピューティングカード(Tesla A100)	14
2.5. N8105-62 GPU コンピューティングカード(NVIDIA A2)	15
2.6. N8105-63 GPU コンピューティングカード(NVIDIA A40)	16
2.7. N8105-65 GPU コンピューティングカード(NVIDIA A100 80GB)	17
2.8. N8105-66 GPU コンピューティングカード(NVIDIA A16)	18
3. 関連ケーブル.....	19
3.1. グラフィックスカード電源ケーブル(8Pin.B タイプ)	19
4. 本体装置の冷却設定値について.....	20
4.1. スロットリングについて.....	20
4.2. 本体装置の冷却設定値および設置環境温度について.....	21
4.3. 本体装置における設定変更方法.....	22
5. 注意事項.....	23
5.1. 本体装置のシャットダウンに関する注意事項.....	23
5.2. iLO イベントログに注意ログが記録される	23
6. 補足事項.....	24
6.1. vGPU 利用時の本体装置の設定について	24
商標について.....	25

はじめに

本資料は、Express5800/100 シリーズ用 GPU コンピューティングカード(以下 GPGPU カードと称することがあります)について説明しています。

本資料の対象となる GPGPU カードは次の通りです。(2023 年 08 月末現在)

- ・ Express5800/R120h 以後のシリーズに搭載可能な GPGPU カード
- ・ Express5800/T120h 以後のシリーズに搭載可能な GPGPU カード
- ・ Express5800/T110k 以後のシリーズに搭載可能な GPGPU カード

※グラフィックドライバのバージョンによっては、使用可能な装置と GPGPU カードとの組み合わせが異なる場合があります。グラフィックドライバの Web 公開ページの内容もご確認下さい。

NEC サポートサイト URL <http://jpn.nec.com/express/>

「サポート情報」の「ダウンロード」に入り、キーワード欄に製品コード(例:N8105-58)を入力して下さい。

1way サーバ向け製品編

型番	品名
N8105-64	GPU コンピューティングカード (NVIDIA A2)

1. 機能仕様

1.1. NVIDIA A シリーズ

製品型番		N8105-64	
製品名		GPUコンピューティング カード (NVIDIA A2)	
GPU		NVIDIA GA102	
GPUアーキテクチャ		Ampere	
メモリ		16GB GDDR6	
メモリバス帯域		200GB/s	
CUDAコア数		1280	
Tensorコア数		40	
性能	倍精度演算性能(FP64)	0.00702 TFLOPS	
	単精度演算性能(FP32)	4.5 TFLOPS	
	整数演算性能(INT8)	36 TOPS	
	行列演算性能 (ディープラーニング)	N/A	
PCI-Expressバス		Gen. 4 x8	
占有スロット数		1	
LowProfile対応		○	
寸法*1		167.6mm(L) 68.9mm(W)	
重量		269g*2	
最大消費電力		60W	
補助電源		-	
対応OS	Windows Server 2019		○
	Windows Server 2022		○
	Red Hat Enterprise Linux	6 x86_64	
		7 x86_64	
		8 x86_64	
	VMware	ESXi6	
ESXi7			

※「」（空欄）は未サポート、“-”は未対応であることを示します。

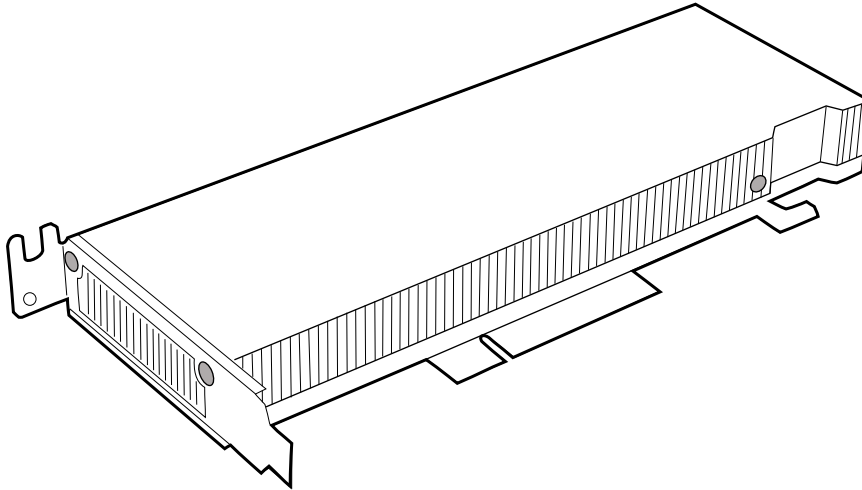
*1. ブラケットは含みません

*2. ロープロファイルブラケット装着時の重量です

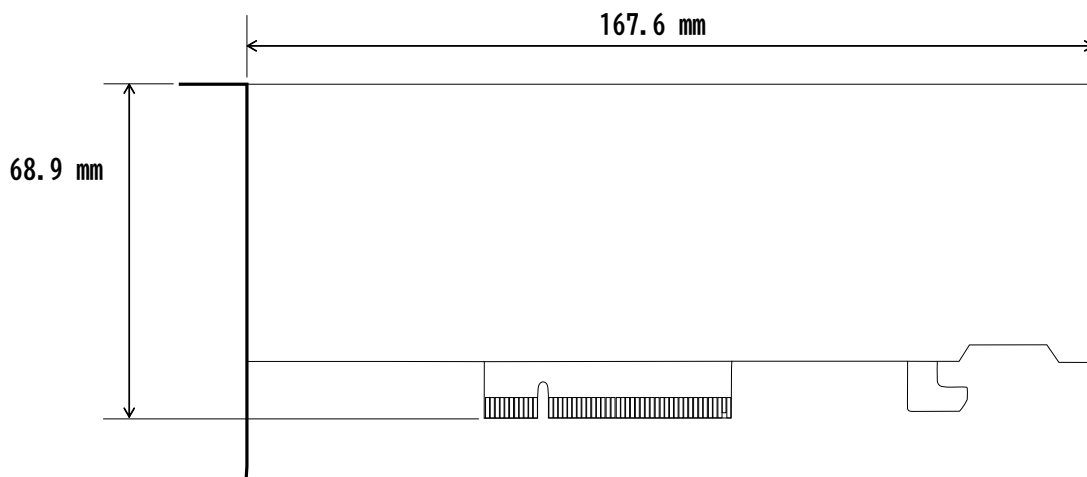
2. 外観図

2.1. N8105-64 GPU コンピューティングカード(NVIDIA A2)

■外観図



■寸法



3. 関連オプション

3.1. PCI 用 FAN オプション

N8181-191 は Express5800/T110k-S(空冷モデル)に対応する PCI 用 FAN オプションです。
 N8105-64 GPU コンピューティングカード (NVIDIA A2)を本体に接続する場合、冷却機能を補うため FAN オプションを取り付ける必要があります。
 取り付け方法は N8181-191 ユーザーズガイドを参照してください。

※搭載時の注意

BMC ファームウェア*1	リビジョン 01.15 以降
BIOS*2	バージョン F10 以前の場合は BIOS SETUP の AC-LINK の設定を「Stay Off」に設定

*1. 最新の BMC ファームウェアは以下よりダウンロードできます。BMC ファームウェアのリビジョンの確認方法は、BMC ファームウェアアップデートモジュールに添付されている手順書に記載されています。

<https://www.support.nec.co.jp/ListModuleDownload.aspx>
 →モデル名 (T110k-S) で対象の BMC ファームウェアを検索

*2. BIOS バージョンの確認方法や AC-LINK の設定変更方法は、メンテナンスガイドの『便利な機能』『システム BIOS の詳細』をご参照ください。

2way サーバ向け製品編

型番	品名
N8105-51	GPU コンピューティングカード
N8105-54	GPU コンピューティングカード (Tesla T4)
N8105-55/56	GPU コンピューティングカード (Tesla V100(S))
N8105-58	GPU コンピューティングカード (Tesla A100)
N8105-62	GPU コンピューティングカード (NVIDIA A2)
N8105-63	GPU コンピューティングカード (NVIDIA A40)
N8105-65	GPU コンピューティングカード (NVIDIA A100 80GB)
N8105-66	GPU コンピューティングカード (NVIDIA A16)

1. 機能仕様

1.1. Tesla シリーズ

製品型番		N8105-51	N8105-54
製品名		GPUコンピューティングカード	GPUコンピューティングカード (Tesla T4)
GPU		NVIDIA GP104	NVIDIA TU104
GPUアーキテクチャ		Pascal	Turing
メモリ		8GB GDDR5	16GB GDDR6
メモリバス帯域		192GB/s	320GB/s
CUDAコア数		2560	2560
Tensorコア数		—	320
性能	倍精度演算性能(FP64)	N/A	N/A
	単精度演算性能(FP32)	5.5 TFLOPS	8.1 TFLOPS
	整数演算性能(INT8)	22 TOPS	130 TOPS
	行列演算性能 (ディープラーニング)	N/A	N/A
PCI-Expressバス		Gen.3 x16	Gen.3 x16
占有スロット数		1	1
LowProfile対応		○	○
寸法*1		167.0mm(L) 68.9mm(W)	169.53mm(L) 68.9mm(W)
重量		250g*2	318g*2
最大消費電力		75W	70W
補助電源		—	—
対応OS	Windows Server 2016		○
	Windows Server 2019		○
	Red Hat Enterprise Linux	6 x86_64	
		7 x86_64	○7.9
		8 x86_64	○8.2, 8.4, 8.6
	VMware	ESXi6	
		ESXi7	

※「」（空欄）は未サポート、「—」は未対応であることを示します。

*1. ブラケット、リテーナは含みません

*2. フルハイトブラケット装着時の重量です。

製品型番		N8105-55	N8105-56	N8105-58
製品名		GPUコンピューティング カード (Tesla V100)	GPUコンピューティング カード (Tesla V100S)	GPUコンピューティング カード (Tesla A100)
GPU		NVIDIA GV100	NVIDIA GV100	NVIDIA GA100
GPUアーキテクチャ		Volta	Volta	Ampere
メモリ		32GB HBM2	32GB HBM2	40GB HBM2
メモリバス帯域		900GB/s	1134GB/s	1555GB/s
CUDAコア数		5120	5120	6912
Tensorコア数		640	640	432
性能	倍精度演算性能(FP64)	7 TFLOPS	8.2 TFLOPS	9.7 TFLOPS
	単精度演算性能(FP32)	14 TFLOPS	16.4 TFLOPS	19.5 TFLOPS
	整数演算性能(INT8)	N/A	N/A	624 TOPS
	行列演算性能 (ディープラーニング)	112 TFLOPS	130 TFLOPS	N/A
PCI-Expressバス		Gen.3 x16	Gen.3 x16	Gen.4 x16
占有スロット数		2	2	2
LowProfile対応		-	-	-
寸法*1		266.7mm(L) 111.2mm(W)	266.7mm(L) 111.2mm(W)	266.7mm(L) 111.2mm(W)
重量		1260g*2	1260g*2	1296g*2
最大消費電力		250W	250W	250W
補助電源		要*3	要*3	要*3*4
対応OS	Windows Server 2016	○	○	○
	Windows Server 2019	○	○	○
	Red Hat Enterprise Linux	6 x86_64		
		7 x86_64	○7.6, 7.7	○7.7
		8 x86_64		○8.2, 8.3
	VMware	ESXi6		
		ESXi7		

※「」（空欄）は未サポート、“-”は未対応であることを示します。

*1. ブラケット、リテーナは含みません

*2. フルハイトブラケット装着時の重量です。

*3. Express5800/R120h シリーズに接続する場合 K410-387(00) グラフィックスカード電源ケーブル(8Pin. B タイプ)が必要です。

*4. Express5800/R120i シリーズに接続する場合 K410-477(00) グラフィックスカード電源ケーブル(8Pin. B タイプ)が必要です。

1.2. NVIDIA A シリーズ

製品型番		N8105-62	N8105-63	N8105-65
製品名		GPUコンピューティング カード (NVIDIA A2)	GPUコンピューティング カード (NVIDIA A40)*2	GPUコンピューティング カード (NVIDIA A100 80GB)
GPU		NVIDIA GA107	NVIDIA GA102	NVIDIA GA100
GPUアーキテクチャ		Ampere	Ampere	Ampere
メモリ		16GB GDDR6	48GB GDDR6	80GB HBM2e
メモリバス帯域		200 GB/s	696 GB/s	1935 GB/s
CUDAコア数		1280	10752	6912
Tensorコア数		40	336	432
性能	倍精度演算性能(FP64)	0.00702 TFLOPS	0.58464 TFLOPS	9.7 TFLOPS
	単精度演算性能(FP32)	4.5 TFLOPS	37.4 TFLOPS	19.5 TFLOPS
	整数演算性能(INT8)	36TOPS	299.3 TOPS	624 TOPS
	行列演算性能 (ディープラーニング)	N/A	N/A	N/A
PCI-Expressバス		Gen.4 x8	Gen.4 x16	Gen.4 x16
占有スロット数		1	2	2
LowProfile対応		○	—	—
寸法*1		167.6mm(L) 68.9mm(W)	267.7mm(L) 111.2mm(W)	267.7mm(L) 111.2mm(W)
重量		274g*3	1042g*3	1222g*3
最大消費電力		60W	300W	300W
補助電源			要*4	要*4
対応OS	Windows Server 2019		○	○
	Windows Server 2022		○	○
	Red Hat Enterprise Linux	6 x86_64		
		7 x86_64		
		8 x86_64	○8.4, 8.6	○8.4, 8.6
	VMware	ESXi6		
		ESXi7		

※「」（空欄）は未サポート、“—”は未対応であることを示します。

*1. ブラケット、リテーナは含みません

*2. 本製品はDisplayPortを3基搭載していますが、未サポートです。

*3. フルハイトブラケット装着時の重量です。

*4. Express5800/R120i シリーズに接続する場合 K410-477(00) グラフィックスカード電源ケーブル(8Pin, Bタイプ)が必要です

製品型番			N8105-66
製品名			GPUコンピューティング カード (NVIDIA A16)
GPU			NVIDIA GA107
GPUアーキテクチャ			Ampere
メモリ			4x 16GB GDDR6
メモリバス帯域			4x 200 GB/s
CUDAコア数			4x 1280
Tensorコア数			4x 40
性能	倍精度演算性能(FP64)		4x 0.00702 TFLOPS
	単精度演算性能(FP32)		4x 4.5TFLOPS
	整数演算性能(INT8)		4x 35.9TOPS
	行列演算性能 (ディープラーニング)		N/A
PCI-Expressバス			Gen. 4 x16
占有スロット数			2
LowProfile対応			—
寸法*1			267.7mm(L) 111.2mm(W)
重量			1147g*3
最大消費電力			250W
補助電源			要*4
対応OS	Windows Server 2019		
	Windows Server 2022		
	Red Hat Enterprise Linux	6 x86_64	
		7 x86_64	
		8 x86_64	
	VMware	ESXi6	
ESXi7			

※「」（空欄）は未サポート、“-”は未対応であることを示します。

*1. ブラケット、リテーナは含みません

*2. 本製品はDisplayPortを3基搭載していますが、未サポートです。

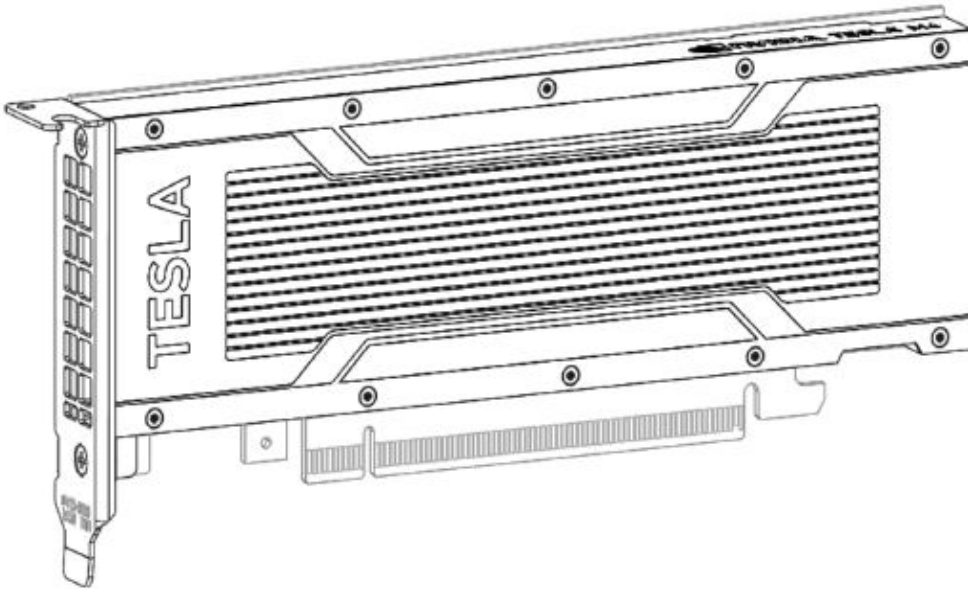
*3. フルハイトブラケット装着時の重量です。

*4. Express5800/R120j シリーズに接続する場合 K410-477(00) グラフィックスカード電源ケーブル(8Pin.Bタイプ)が必要です。

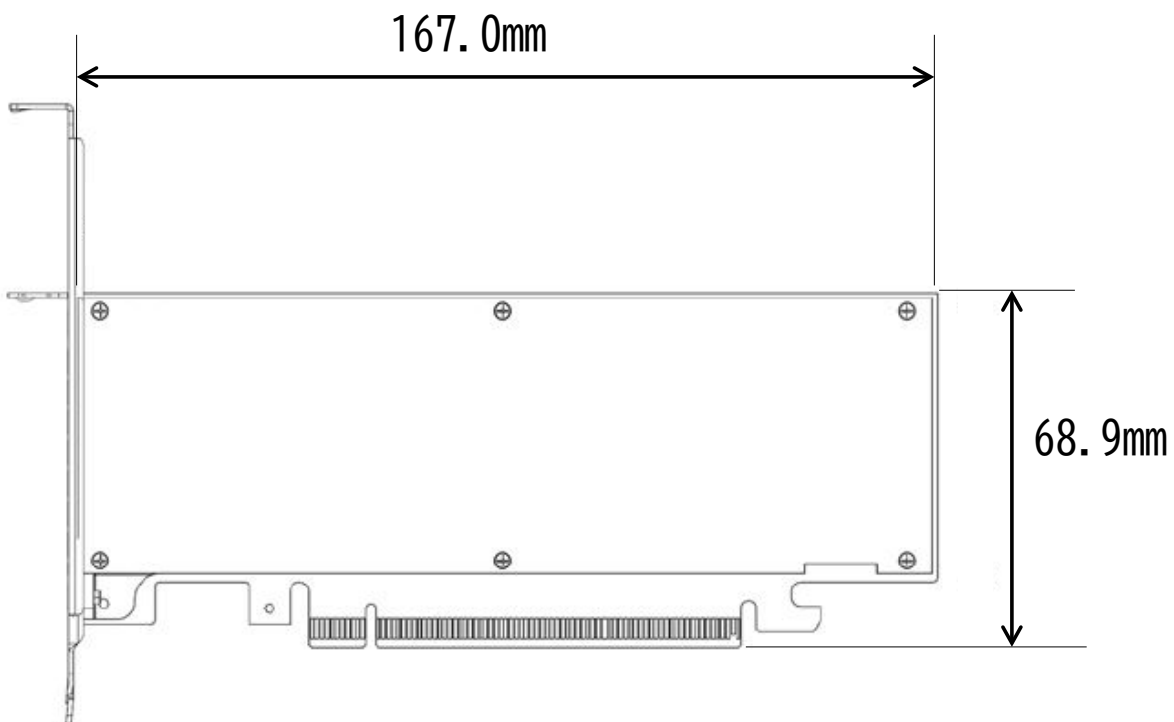
2. 外観図

2.1. N8105-51 GPU コンピューティングカード

■外観図

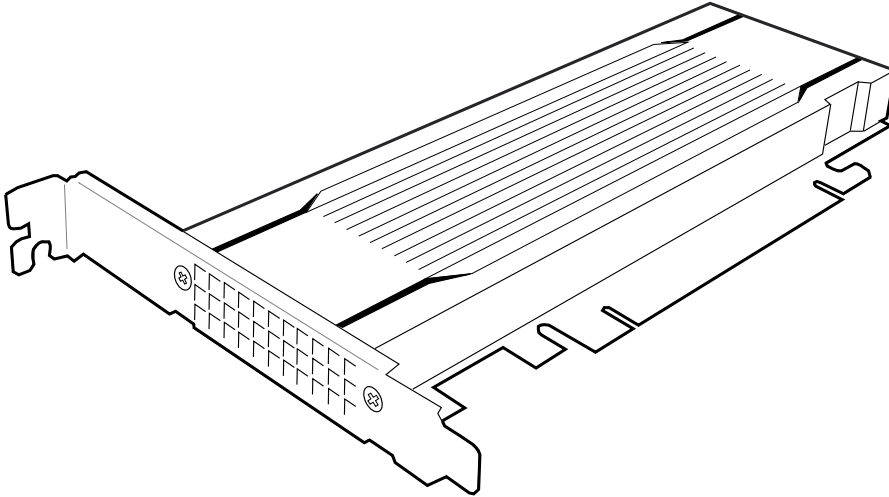


■寸法

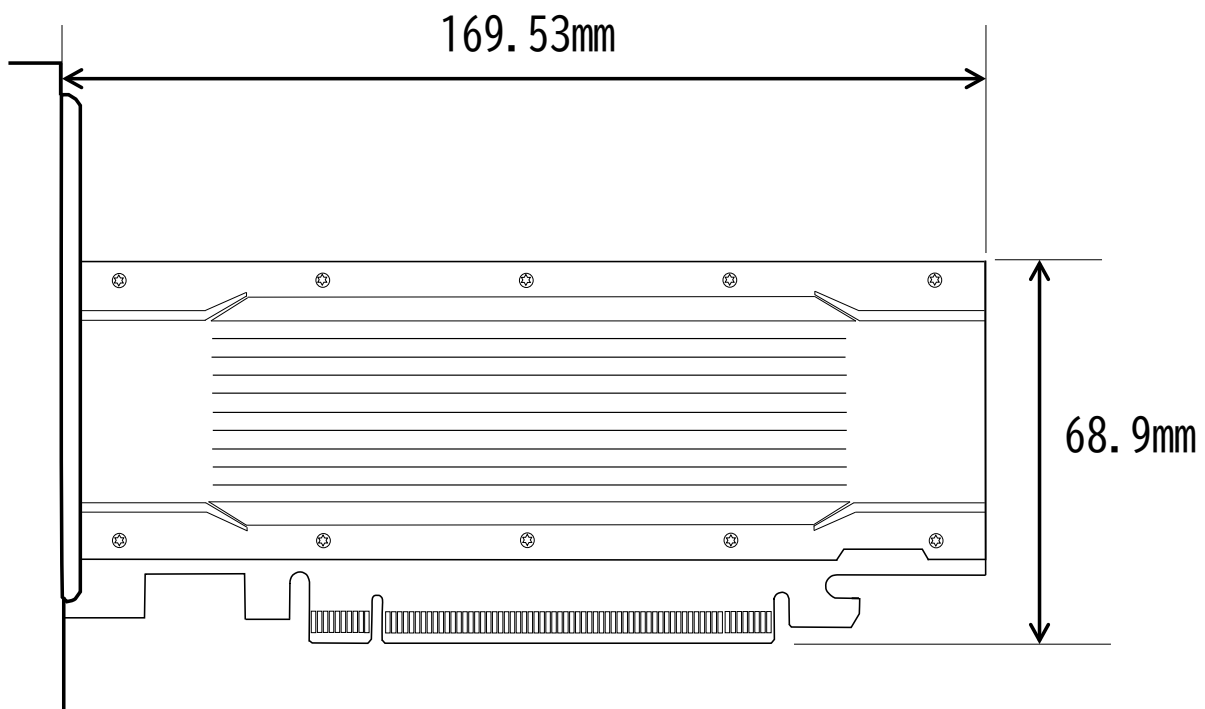


2.2. N8105-54 GPU コンピューティングカード(Tesla T4)

■外観図



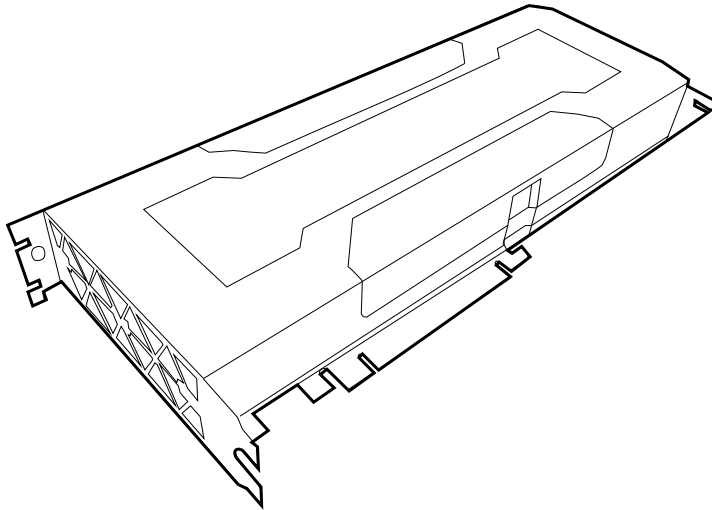
■寸法



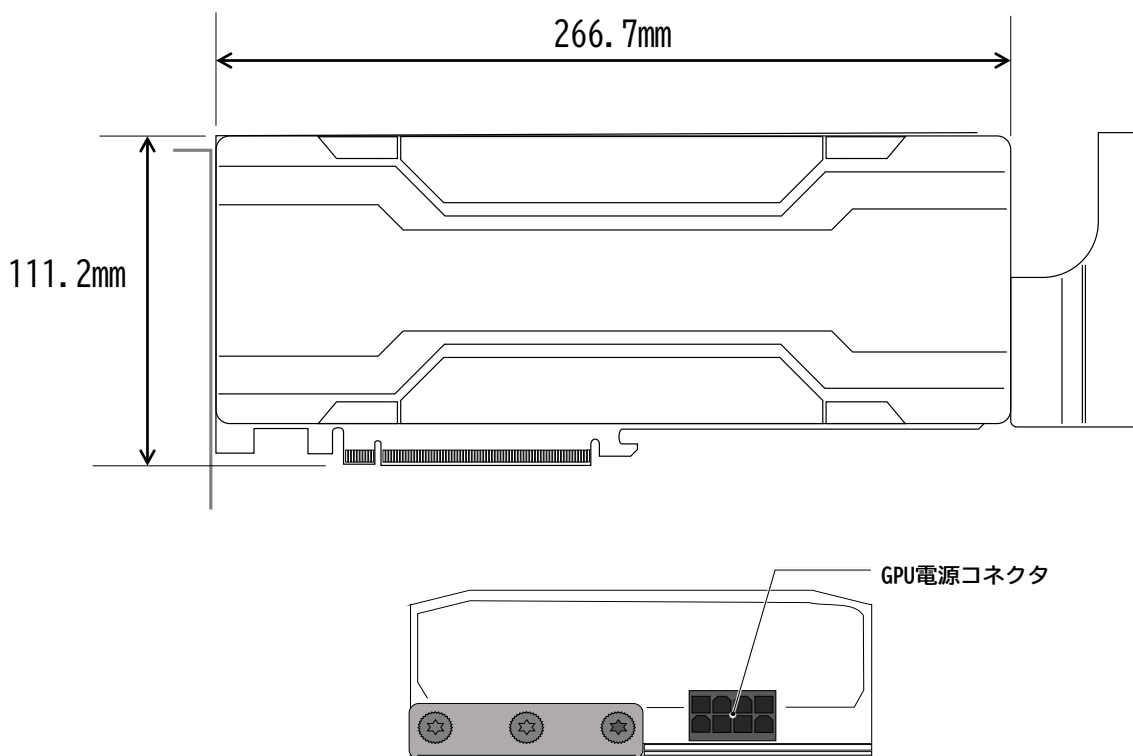
2.3. N8105-55 GPU コンピューティングカード(Tesla V100)/

N8105-56 GPU コンピューティングカード(Tesla V100S)

■外観図

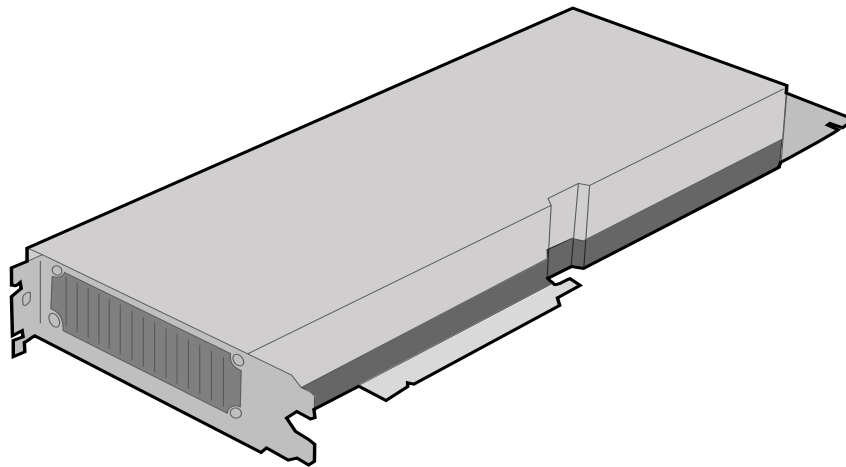


■寸法

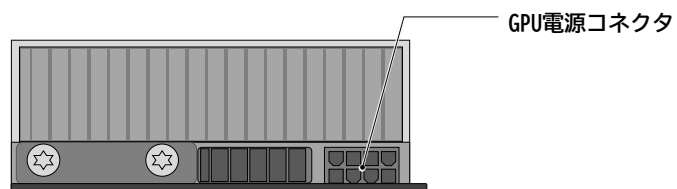
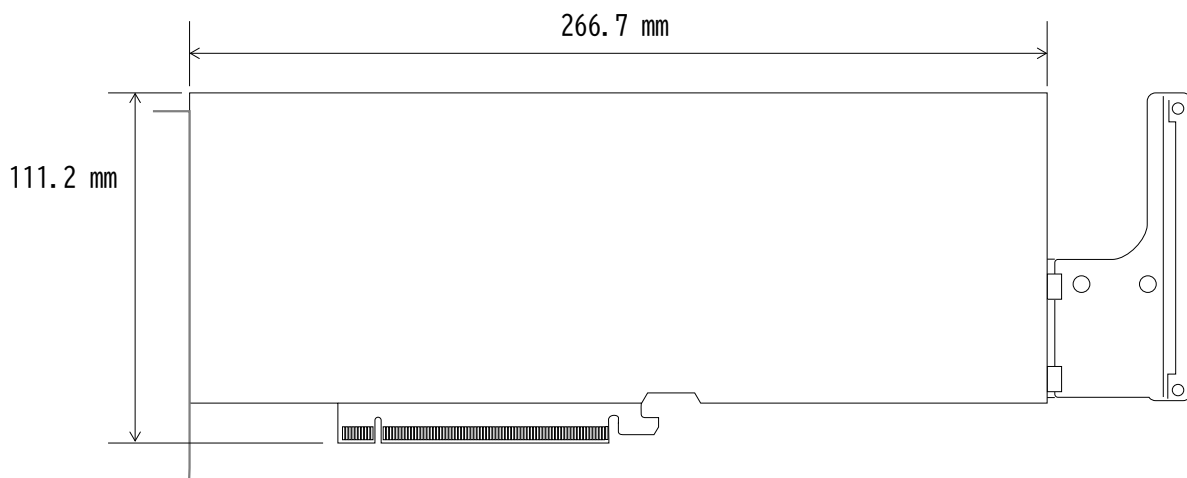


2.4. N8105-58 GPU コンピューティングカード(Tesla A100)

■外観図

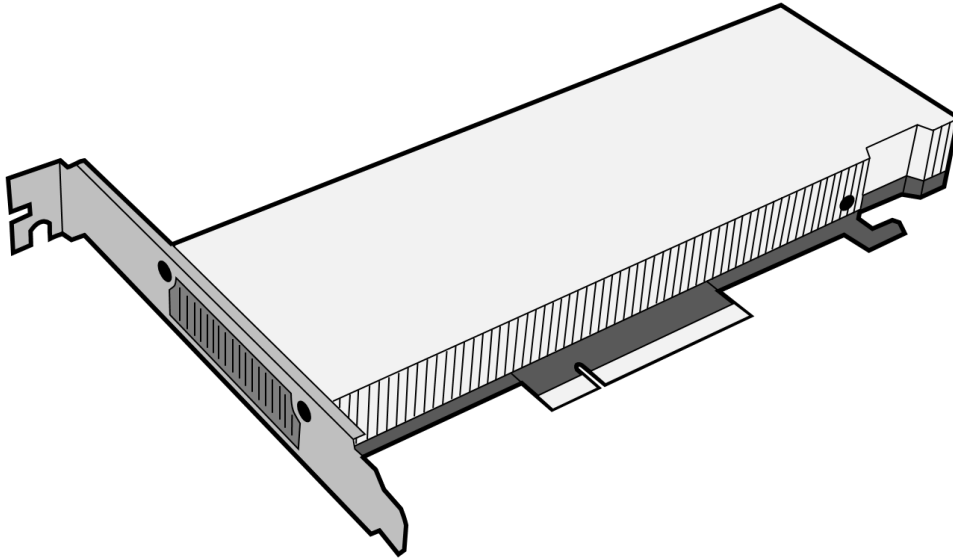


■寸法

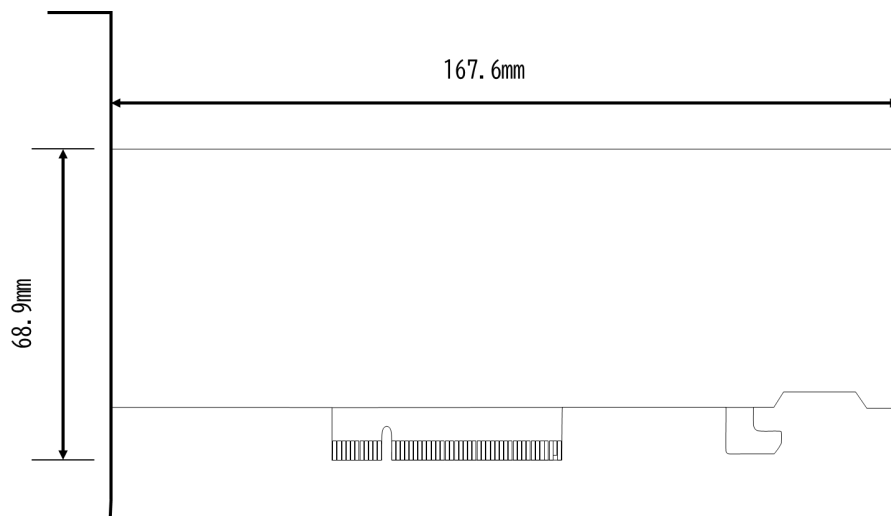


2.5. N8105-62 GPU コンピューティングカード(NVIDIA A2)

■外観図

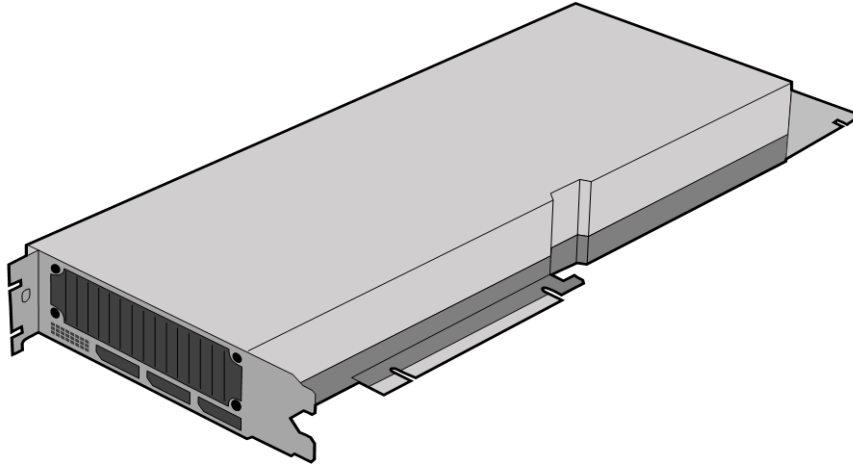


■寸法

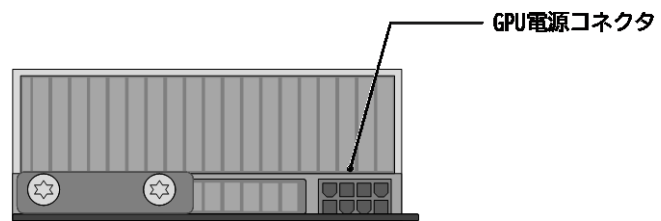
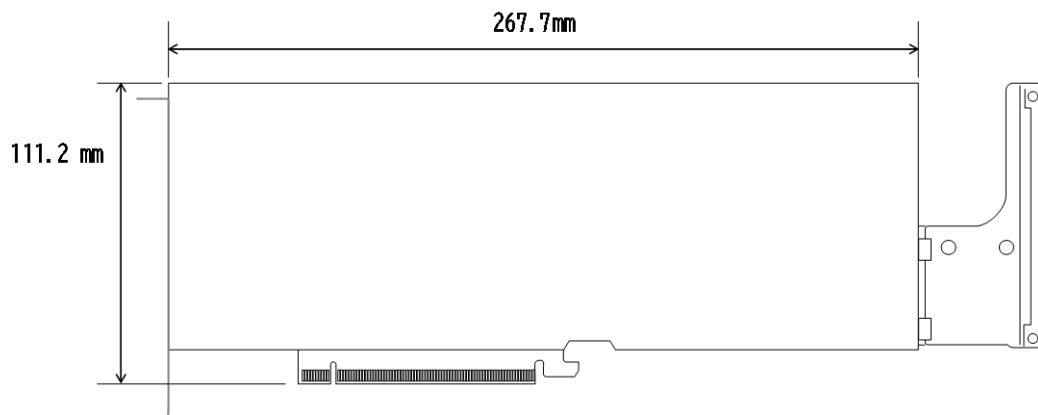


2.6. N8105-63 GPU コンピューティングカード(NVIDIA A40)

■外観図

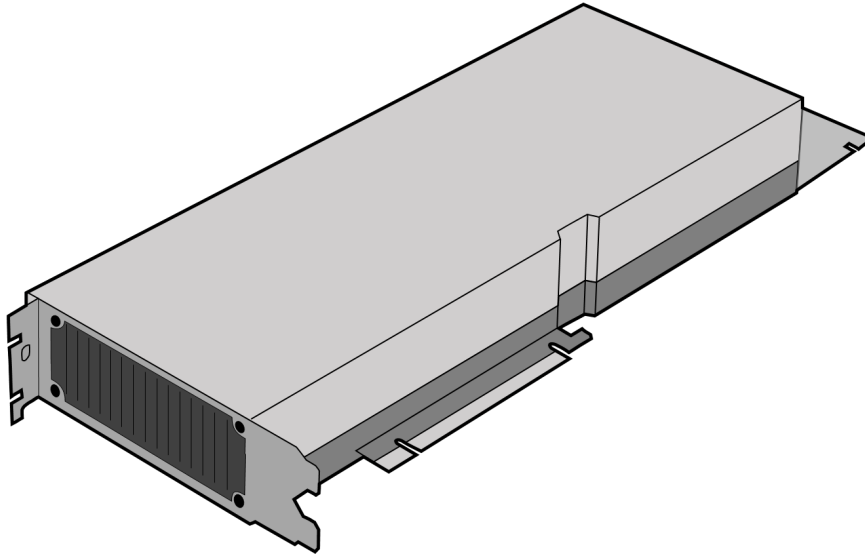


■寸法

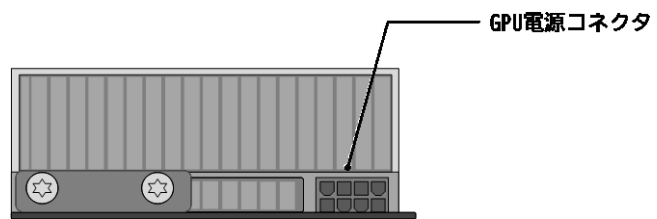
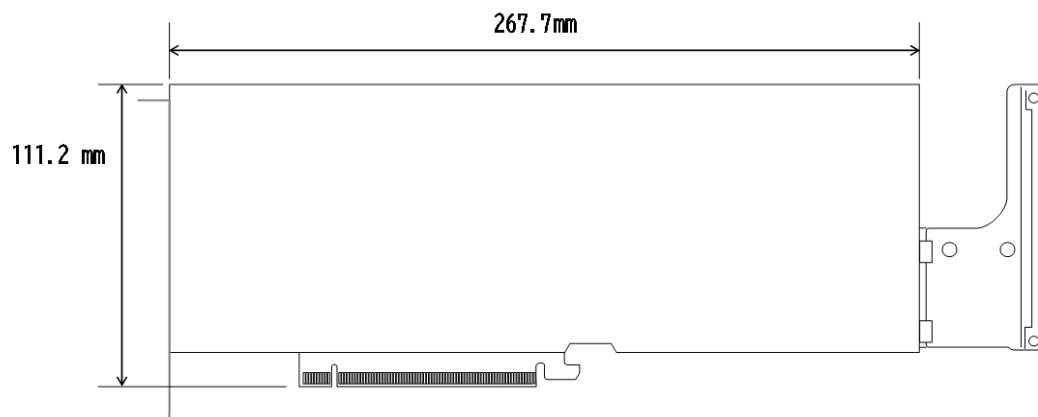


2.7. N8105-65 GPU コンピューティングカード(NVIDIA A100 80GB)

■外観図

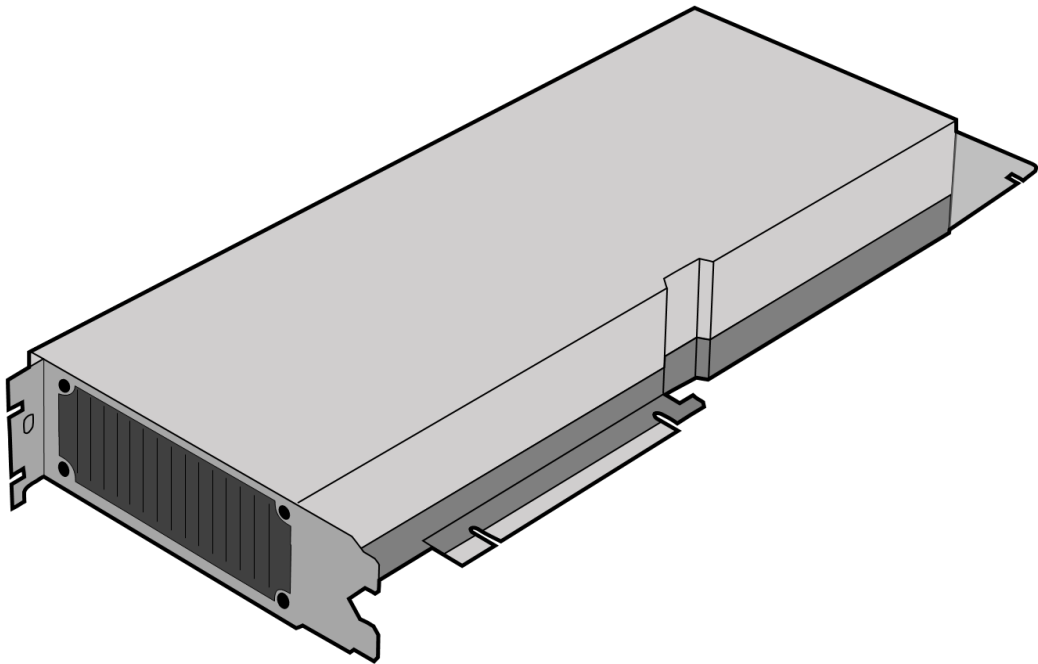


■寸法

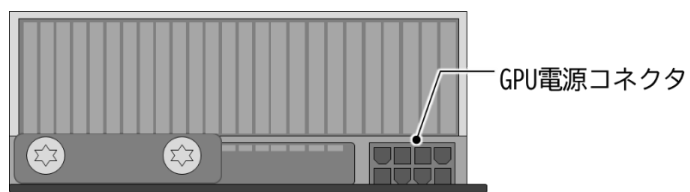
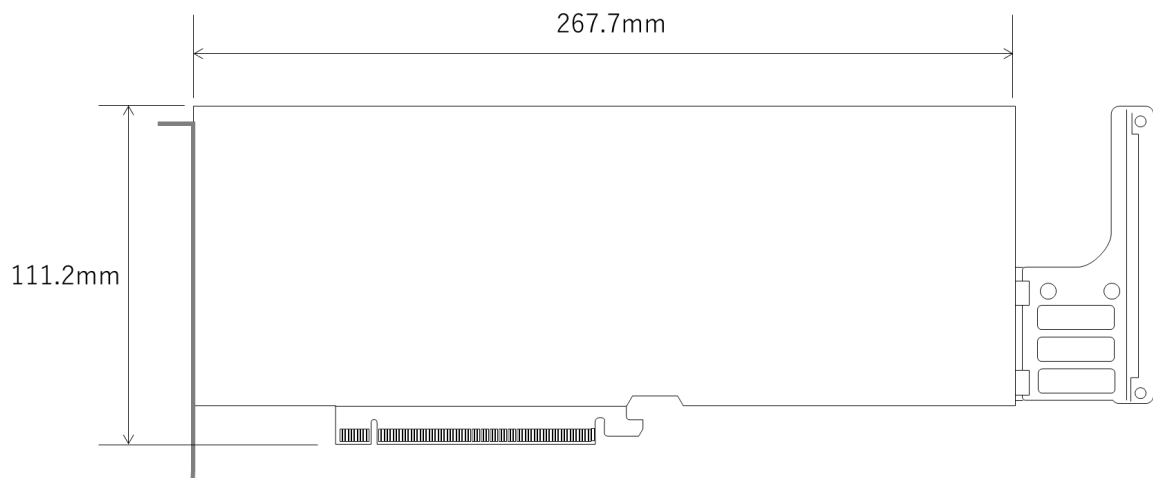


2.8. N8105-66 GPU コンピューティングカード(NVIDIA A16)

■外観図



■寸法



3. 関連ケーブル

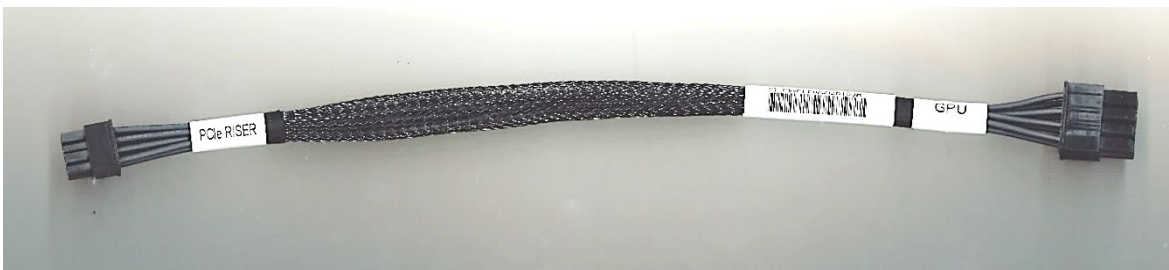
3.1. グラフィックスカード電源ケーブル(8Pin.B タイプ)

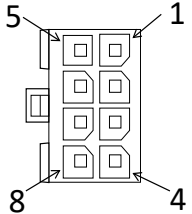
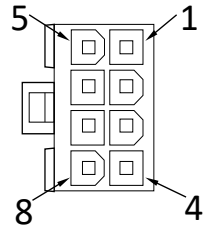
K410-387(00)は Express5800/R120h シリーズに、また K410-477(00)は Express5800/R120i シリーズおよび Express5800/R120j シリーズにそれぞれ対応する電源ケーブルです。

N8105-55/56/58/63/65/66 GPU コンピューティングカード (Tesla V100(S)または Tesla A100、NVIDIA A40 または Tesla A100 80GB、NVIDIA A16)を本体に接続する際、PCI-Express バスからの電力供給では不足するため、補助電源を接続する必要があります。これらのケーブルを用いて N8105-55/56/58/63/65 と本体装置のライザカードに備えられた補助電源コネクタに接続することで、本製品を動作させることができます。

N8105-55/56/58/63/65/66 1 枚につき補助電源ケーブル 1 本の接続が必要です。

K410-387(00)ならびに K410-477(00)には電源ケーブルが 3 本収められていますので、複数枚の N8105-55/56/58/63/65/66 を搭載する場合は搭載枚数に応じて、K410-387(00)、K410-477(00)の手配数量を検討してください。



コネクタ外観	 PCIe RISER 接続側	 GPU 電源コネクタ接続側
	Pin#	signal
	1	12V
	2	12V
	3	12V
	4	SENSE1
	5	GND
	6	SENSE0
	7	GND
	8	GND

4. 本体装置の冷却設定値について

本製品を安全かつ最大性能でご利用いただくには本体装置の冷却設定を変更する必要があります。本体装置の工場出荷設定では GPU コンピューティングカードの性能を十分に発揮することができない場合がありますので、必ず下記の冷却設定値を確認し、本体装置の冷却設定値を変更してから運用してください。

4.1. スロットリングについて

GPU コンピューティングカードには、GPU 高負荷時に発熱量を抑制するために動作クロックを下げる「スロットリング機能」が備わっています。動作環境や GPU 負荷状況により本機能が作動し、一時的に GPU コンピューティングカードの処理性能が低下する場合がありますが、故障などの問題ではありません。なお、本体装置の冷却機能設定を変更(強化)することで、スロットリングの発生を抑えることが可能な場合もあります。変更方法については『4.3 本体装置における設定変更方法』を参照して下さい。

4.2. 本体装置の冷却設定値および設置環境温度について

性能低下(『4.1 スロットリングについて』参照)を伴わずに運用していただくには、お使いの本体装置に対して適切な冷却設定値を適用する必要があります。

ここでは GPU コンピューティングカードごとに本体装置の冷却設定の推奨値をまとめました。

また GPU コンピューティングカードを搭載する際の本体装置の環境温度についても、本体装置及び構成によって異なる場合があります。詳細はシステム構成ガイドを参照して下さい。

■本体装置冷却機能 推奨設定値 (2023/08 時点)

搭載GPUに適した本体装置の冷却設定値(推奨値) (オプション: Thermal Configuration への設定パラメーター値)				
本体装置 モデル名→ カード型番 ↓	R120j-1M/-2M	R120i-1M/-2M	R120h-2M (2nd-Gen), R120h-2M (3rd-Gen)	T120h (2nd-Gen), T120h (3rd-Gen)
N8105-54 (Tesla T4)	(接続対象外)	(接続対象外)	Increased Cooling *1*2	Increased Cooling *1
N8105-55/-56 (Tesla V100/V100S)	(接続対象外)	(接続対象外)	Maximum Cooling	(接続対象外)
N8105-58 (NVIDIA A100)	(接続対象外)	Maximum Cooling *3	Maximum Cooling	(接続対象外)
N8105-62 (NVIDIA A2)	Increased Cooling *1*4	Increased Cooling *1	(接続対象外)	(接続対象外)
N8105-63 (NVIDIA A40)	(接続対象外)	Increased Cooling *1*3	(接続対象外)	(接続対象外)
N8105-65 (NVIDIA A100 80GB)	(接続対象外)	Increased Cooling *1*3	(接続対象外)	(接続対象外)
N8105-66 (NVIDIA A16)	Increased Cooling *1*5	(接続対象外)	(接続対象外)	(接続対象外)

- *1. 本ボード以外の指示により、冷却設定値の変更の必要がある場合、以下を参照して強度の高い方を優先してください。
Thermal Configuration 強度順: Maximum Cooling > Enhanced CPU Cooling > Increased Cooling > Optimal Cooling
- *2. 複数枚搭載時
- *3. R120i-1M は接続対象外
- *4. R120j-2M は接続対象外
- *5. R120j-1M は接続対象外

4.3. 本体装置における設定変更方法

本体装置における設定変更方法については、本体装置のメンテナンスガイドを参照して下さい。

以下手順により最新版のガイドを参照して適用してください。

- ① 「NEC サポートポータル」の「ハードウェア」ページより、「型番・モデル名から探す」のページにアクセスします。

<http://www.support.nec.co.jp/HWSearchByNumber.aspx>

- ② サーバ本体の型番・またはモデル名を入力して検索を行い、サーバ製品のページを開きます。
- ③ サーバ製品のページで「製品マニュアル」を選択して「ユーザーズガイド」を選びます。
- ④ 「ユーザーズガイド」のページに公開されている「サーバメンテナンスガイド(共通編)」を取得してください。

「1 章 便利な機能」

- 「1. システムユーティリティ」
- 「1.2.2 BIOS/Platform Configuration (RBSU)」
- 「(12) Advanced Options メニュー」
 - 「(a) Fan and Thermal Options メニュー」 → 「Thermal Configuration」

5. 注意事項

5.1. 本体装置のシャットダウンに関する注意事項

オペレーティングシステムのシャットダウンの際、稀に本体装置に PCI Bus に関するエラーが記録されることがありますが、GPU コンピューティングカード及び本体装置の故障を示すものではありません。Windows Server 2016 においては、本事象発生時にはシャットダウンせずに再起動しますので、再起動後に改めてシャットダウンして下さい。

なお、無停電電源装置(UPS)接続時など、確実にオペレーティングシステムをシャットダウンする必要がある場合は、Windows Server 2019 での運用をお奨めします。

5.2. iLO イベントログに注意ログが記録される

下記注意ログが本体装置の iLO 上に記録されることがありますが、本ボード及び本体装置の故障を示すものではありません。

■スロット XX に搭載した場合のエラーメッセージ(iLO の IML(Integrated Management Log))

[Information] - [Integrated Management Log]

The iLO health monitoring status of the device /adapter located in slot XX has OS driver missing or not in persistent mode so read thermal limits is not responsive.

6. 補足事項

6.1. vGPU 利用時の本体装置の設定について

vGPU を利用する場合、本体装置の仮想化支援機能が必要となります。

本体装置マニュアルを参照し、本体装置のシステムパラメーターを確認・設定してください。

■参照ドキュメント

- ① 「NEC サポートポータル」の「ハードウェア」ページより、「型番・モデル名から探す」のページにアクセスします。
<http://www.support.nec.co.jp/HWSearchByNumber.aspx>
- ② サーバ本体の型番・またはモデル名を入力して検索を行い、サーバ製品のページを開きます。
- ③ サーバ製品のページで「製品マニュアル」を選択して「ユーザーズガイド」を選びます。
- ④ 「ユーザーズガイド」のページに公開されている「サーバメンテナンスガイド(共通編)」を取得してください。

■変更パラメーター

システムユーティリティから、「System Configuration > BIOS/Platform Configuration (RBSU) > Virtualization Options」を選択すると、「Virtualization Options」メニューが表示されます。

Virtualization Options メニューで、次のパラメーター設定を確認し、必要に応じて変更してください

オプション	パラメーター(設定値)	出荷時設定
Intel(R) VT-d	Enabled	Enabled
SR-IOV	Enabled	Enabled

■対象製品

vGPU は下記の製品に対応しています。

下記以外の GPGPU 製品で、vGPU の使用はサポートしていません。

型番	品名
N8105-63	GPU コンピューティングカード (NVIDIA A40)
N8105-66	GPU コンピューティングカード (NVIDIA A16)

商標について

<本書内の対象 OS の省略形式>

省略形式	製品名
Windows Server 2016	Microsoft® Windows Server® 2016 Standard Microsoft® Windows Server® 2016 Datacenter
Windows Server 2019	Microsoft® Windows Server® 2019 Standard Microsoft® Windows Server® 2019 Datacenter
Windows Server 2022	Microsoft® Windows Server® 2022 Standard Microsoft® Windows Server® 2022 Datacenter
RHEL6	Red Hat Enterprise Linux 6
RHEL7	Red Hat Enterprise Linux 7
RHEL8	Red Hat Enterprise Linux 8
VMware ESXi 6	VMware ESXi™ 6
VMware ESXi 7	VMware ESXi™ 7

NVIDIA, Tesla, Quadro, PhysX, CUDA, Pascal, Turing, Ampere, vGPU は米国およびその他の国における NVIDIA Corporation の商標または登録商標です。

Microsoft とそのロゴおよび、Windows、Windows Server は米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です。

Linux は Linus Torvalds 氏の日本およびその他の国における商標または登録商標です。

Red Hat、Red Hat Enterprise Linux は、米国およびその他の国における Red Hat, Inc. の登録商標または商標です。

VMware および VMware の製品名は、VMware, Inc. の米国および各国での商標または登録商標です。

Hewlett Packard Enterprise および HPE は米国 Hewlett Packard Enterprise Development LP の米国およびその他の国における登録商標です。

PCI-Express は PCI-SIG の登録商標です。

OpenGL は、米国およびその他の国における Silicon Graphics, Inc. の商標または登録商標です。

OpenCL は、Apple Inc. の商標であり、Khronos Group Inc. の許可を得て使用しています。

X Window System は X Consortium, Inc. の商標です。

その他、記載されている会社名、製品名は、各社の登録商標または商標です。