

# NVIDIA A40

## NVIDIA vGPU ソフトウェア

### 動作検証レポート



VMware は、米国およびその他の地域における VMware, Inc.の登録商標または商標です。VMware, Inc.の商標に関する詳細は <https://www.vmware.com/jp/help/trademarks.html> を参照ください。

Microsoft、Windows および 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.登録商標または商標です。

NVIDIA 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。

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。

免責条項: 本書または本書に記述されている製品や技術に関して、日本電気株式会社またはその関連会社が行う保証は、製品または技術の提供に適用されるライセンス契約で明示的に規定されている保証に限ります。このような契約で明示的に規定された保証を除き、日本電気株式会社およびその関連会社は、製品、技術、または本書に関して、明示または黙示を問わず、いかなる種類の保証も行いません。

## 目次

<b>NVIDIA A40 動作検証について</b> .....	<b>3</b>
1 ご利用にあたっての注意事項.....	3
2 NVIDIA A40 の概要.....	3
3 検証目的.....	3
4 動作検証の準備.....	4
4.1 動作検証システム構成.....	4
4.2 動作検証済のサーバ構成 (R120i-2M).....	4
4.3 NVIDIA A40 増設手順.....	5
4.4 vGPU ソフトウェア適用手順.....	7
5 検証結果.....	8
6 関連リンク.....	8
7 お問い合わせ先.....	8
8 改版履歴.....	8

# NVIDIA A40 動作検証について

## 1 ご利用にあたっての注意事項

本レポートは動作検証レポートであり、弊社が動作保証するものではありません。  
動作確認情報は、各ページに掲載されている評価環境での検証結果に基づいたものです。  
導入に際しては個々の環境で十分な確認を実施してください。

## 2 NVIDIA A40 の概要

NVIDIA A40 をサーバに実装することで、3D 映像を滑らかに表示するなどの高度なグラフィックス処理に加え、大量の並列処理を実行する GPU コンピューティングの活用により、科学技術計算、ディープラーニングなどで利用することができます。

## 3 検証目的

今回の検証では、Express5800 シリーズに NVIDIA A40 を搭載し、後述の手配物品、搭載手順、搭載条件を満たした VMware vSphere®(VMware ESXi™)での仮想環境下にて、CUDA を使用した GPGPU の基本動作検証を実施した結果を記載します。

## 4 動作検証の準備

### 4.1 動作検証システム構成

弊社において検証済みの構成を掲載いたします。なお、下記は一例ですので、お客様の環境や用途にあわせてシステムを構成してください。

### 4.2 動作検証済のサーバ構成 (R120i-2M)

本章では、動作検証を実施した R120i-2M についてのサーバ手配構成 / 構成に応じた電源ユニットの選択方法 / 動作検証条件について説明します。

#### 4.2.1 R120i-2M サーバ手配構成

製品名	対象型名	補足事項
Express5800/R120i-2M	N8100-2919Y	8x 2.5 型ドライブモデル
CPU ボード / 増設 CPU ボード	---	CPU TDP が 235W 以下の CPU のみ搭載可能。 (GPU x1 枚搭載時の場合)
高性能ファン	N8181-198	必須手配品。
高性能 CPU ヒートシンク	N8101-1821	必須手配品。(CPU の手配数と同数必須)
グラフィックスカード電源ケーブル (8Pin. B タイプ)	K410-477(00)	必須手配品。 NVIDIA A40 に電源供給するために必要。(1 セット で 3 本の補助電源ケーブルが含まれる)
1st ライザカード (2xPCI. B タイプ)	N8116-96	
2nd ライザカード (2xPCI. B タイプ)	N8116-99	2CPU 構成が必須。
3rd ライザカード (1xPCI)	N8116-102	2CPU 構成が必須。
GPU コンピューティングカード搭載 キット(1st/2nd ライザ)	N8116-103	1st/2nd ライザカードの利用時に必要。 <sup>1</sup> (1 枚搭載時: 1 個必要、2 枚搭載時: 2 個必要)
GPU コンピューティングカード搭載 キット(3rd ライザ)	N8116-104	3rd ライザカードの利用時に必要。 <sup>1</sup> (2 枚搭載時: 1 個必要)
電源ユニット(1600W)	N8181-162A	必須手配品。(本製品は AC200V 環境のみ使用可能) 2 台手配必須。(冗長化必須)

1. GPU コンピューティングカード(NVIDIA A40)を 2 台手配し、3rd ライザカードを手配する際は、グラフィックスカード搭載キットの手配条件が異なります。下記を手配してください。

- N8116-103 GPU コンピューティングカード搭載キット(1st/2nd ライザ)を 1 個
- N8116-104 GPU コンピューティングカード搭載キット(3rd ライザ)を 1 個

その他増設オプションについては、Express5800/R120i-2M システム構成ガイドを参照の上、手配ください。

<http://jpn.nec.com/express/systemguide/100guide.html>

## 4.2.2 NVIDIA A40 搭載時に搭載可能な電源ユニット

搭載するオプション(メモリ、ディスク等)により、システムに要求される電力量が異なります。導入に際しては個々の環境で十分な確認を実施してください。

### 動作検証条件および搭載制限オプション

分類	GPU 搭載枚数 : 1 枚	GPU 搭載枚数 : 2 枚
CPU	CPU TDP: 235W まで搭載可能	CPU TDP: 185W まで搭載可能
内蔵ドライブ	搭載可能台数 : 16 台以下 NVMe/SAS SSD: 搭載不可	搭載可能台数 : 8 台以下 NVMe/SAS SSD: 搭載不可
メモリ	合計容量が 96GB 以上	合計容量が 192GB 以上
	RDIMM: 制限なし LRDIMM: 搭載不可	RDIMM: 16 枚まで搭載可能 LRDIMM: 搭載不可
増設ドライブケース	N8154-155 8x2.5 型ドライブケース(SAS/SATA)のみ搭載可能	搭載不可
PCI カード	制限なし	
防塵フィルタ	搭載不可	
RAID コントローラ	N8103-238 RAID コントローラ(8GB, RAID 0/1/5/6) 搭載不可	
動作環境温度	N8100-2919Y 8x2.5 型ドライブモデル(標準ドライブケース有) : 25 度以下	

#### 補足事項:

- NVIDIA A40 は PCI ライザあたり最大 1 枚まで搭載可能。
- NVIDIA A40 はサーバ 1 台あたり最大 2 枚まで搭載可能。

## 4.3 NVIDIA A40 増設手順

### 4.3.1 事前準備

Express5800/R120i-2M への各オプションの実装方法については、下記のユーザズガイドの「2 章 準備 1.24 GPU コンピューティングカード」および、搭載する GPU コンピューティングカードの『ユーザズガイド』の「2. 本体装置への取り付けと取り外し」を参照してください。

<https://www.support.nec.co.jp/View.aspx?NoClear=on&id=3170102646>

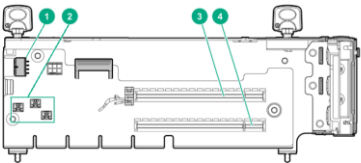
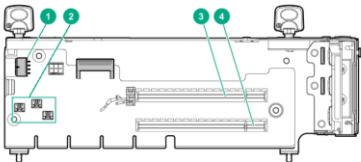
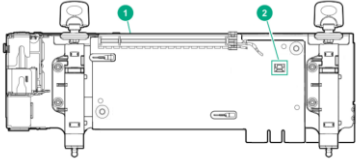
NVIDIA A40 への電源ケーブル接続については、本書 4.3.2 を参照してください。

また、冷却設定について、下記 GPU コンピューティングカードの『テクニカルガイド』を参照の上、Express5800/R120i-2M の設定を実施してください。

<https://www.support.nec.co.jp/View.aspx?id=3140107160>

### 4.3.2 R120i-2M NVIDIA A40 ケーブル接続 (外部電源)

以下の表を参考に、K410-477(00) グラフィックスカード電源ケーブル(8Pin. B タイプ)で NVIDIA A40 を本体装置、およびライザカードの電源コネクタに接続してください。

ケーブル接続先	PCI ライザ概要	PCI ライザ概略図																				
<b>N8116-96</b> (1st ライザカード)	2xPCI PCI スロット: 2x PCIe 4.0(x16)	 <table border="1"> <thead> <tr> <th>No.</th> <th>Slot</th> <th>PCIe type</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td>-</td> <td>GPU 電源ケーブルコネクタ</td> </tr> <tr> <td>2</td> <td></td> <td>-</td> <td>バックアップ用電源コネクタ</td> </tr> <tr> <td>3</td> <td>2</td> <td>Half-length/Full-height (HL/FH)</td> <td>PCIe4 x16 (16, 8, 4, 2, 1)</td> </tr> <tr> <td>4</td> <td>3</td> <td>Half-length/Full-height (HL/FH)</td> <td>PCIe4 x16 (16, 8, 4, 2, 1)</td> </tr> </tbody> </table>	No.	Slot	PCIe type	Description	1		-	GPU 電源ケーブルコネクタ	2		-	バックアップ用電源コネクタ	3	2	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)	4	3	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)
No.	Slot	PCIe type	Description																			
1		-	GPU 電源ケーブルコネクタ																			
2		-	バックアップ用電源コネクタ																			
3	2	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)																			
4	3	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)																			
<b>N8116-99</b> (2nd ライザカード)	2xPCI PCI スロット: 2x PCIe 4.0(x16)	 <table border="1"> <thead> <tr> <th>No.</th> <th>Slot</th> <th>PCIe type</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td>-</td> <td>GPU 電源ケーブルコネクタ</td> </tr> <tr> <td>2</td> <td></td> <td>-</td> <td>バックアップ用電源コネクタ</td> </tr> <tr> <td>3</td> <td>5</td> <td>Half-length/Full-height (HL/FH)</td> <td>PCIe4 x16 (16, 8, 4, 2, 1)</td> </tr> <tr> <td>4</td> <td>6</td> <td>Half-length/Full-height (HL/FH)</td> <td>PCIe4 x16 (16, 8, 4, 2, 1)</td> </tr> </tbody> </table>	No.	Slot	PCIe type	Description	1		-	GPU 電源ケーブルコネクタ	2		-	バックアップ用電源コネクタ	3	5	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)	4	6	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)
No.	Slot	PCIe type	Description																			
1		-	GPU 電源ケーブルコネクタ																			
2		-	バックアップ用電源コネクタ																			
3	5	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)																			
4	6	Half-length/Full-height (HL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)																			
<b>N8116-102</b> (3rd ライザカード)	1xPCI PCI スロット: 1x PCIe 4.0(x16)	 <table border="1"> <thead> <tr> <th>No.</th> <th>Slot</th> <th>PCIe type</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>7</td> <td>Full-length/Full-height (FL/FH)</td> <td>PCIe4 x16 (16, 8, 4, 2, 1)</td> </tr> <tr> <td>2</td> <td></td> <td>-</td> <td>バックアップ用電源コネクタ</td> </tr> </tbody> </table>	No.	Slot	PCIe type	Description	1	7	Full-length/Full-height (FL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)	2		-	バックアップ用電源コネクタ								
No.	Slot	PCIe type	Description																			
1	7	Full-length/Full-height (FL/FH)	PCIe4 x16 (16, 8, 4, 2, 1)																			
2		-	バックアップ用電源コネクタ																			

## 4.4 vGPU ソフトウェア適用手順

- ソフトウェア(ドライバ/ライセンスファイル) 入手
- ライセンスサーバ設定
- ドライバインストール

適用には上記3つの作業が必要となります。下記 URL の NVIDIA Virtual GPU Software v13.0 through 13.6 QuickStartGuide に記載の手順に従い、適用してください。

<https://docs.nvidia.com/grid/13.0/grid-software-quick-start-guide/index.html>

## 5 検証結果

NVIDIA A40 を搭載した VMware vSphere 上の仮想サーバ(Windows Server 2019、Red Hat Enterprise Linux 8.6)において、CUDA サンプルによるコンピューティング処理の動作検証を行った結果、問題が発生しないことを確認しました。

動作検証時の主要ソフトウェアのバージョン・内容は下記になります。

- NVIDIA vGPU ソフトウェア : 13.6
- NVIDIA vGPU プロファイル : nvidia\_a40-1q/2q/3q/4q/6q/8q/12q/16q/24q/48q
- VMware ESXi : 7.0 Update 3c
- Windows Server : 2019
- Red Hat Enterprise Linux : 8.6

## 6 関連リンク

[PC サーバ Express5800 シリーズ \(NEC\)](#)

[仮想マシンによるサーバ統合 VMware: ソフトウェア \(NEC\)](#)

[Windows Server 2019: ソフトウェア \(NEC\)](#)

[Linux OS のサブスクリプションとサポートサービス Linux サービスセット \(NEC\)](#)

[NVIDIA vGPU \(仮想 GPU\) ソフトウェア \(NVIDIA 社\)](#)

## 7 お問い合わせ先

NEC インフラ・テクノロジーサービス事業部門 コンピュート統括部

[info@gpu.jp.nec.com](mailto:info@gpu.jp.nec.com)

## 8 改版履歴

版数	公開日時	変更内容
第 1 版	2023 年 3 月	第 1 版リリース
第 2 版	2023 年 8 月	「お問い合わせ先情報」変更