

NVIDIA DGX SuperPOD

Deployment Guide

Featuring NVIDIA DGX A100 and DGX H100 Systems

DG-11251-001 V15 2024-09-26 BCM 10.24.09

Contents

1.	Introduction	2
2.	Initial Point-to-Point Preparations	3
3.	Initial Cluster Setup	4
4.	Head Node Configuration	.25
5.	High Availability	.29
6.	Slurm Setup	.45
7.	Upgrade Infiniband Switches	.49
8.	SuperPOD Validation	.51

1. Introduction

This document details deploying NVIDIA Base Command[™] Manager (BCM) on NVIDIA DGX SuperPOD[™] configurations.

Physical installation and network switch configuration must be completed before deploying BCM. In addition, information about the intended deployment should be recorded in a site survey.

The DGX SuperPOD Deployment Guides contain essential information and should be consulted before deploying the configuration. The DGX H100 version of the deployment guide is available here: <u>https://docs.nvidia.com/dgx-superpod/index.html</u>

Note: NVIDIA DGX SuperPOD only supports Slurm and Kubernetes with RunAI.

2. Initial Point-to-Point Preparations

Follow the instructions provided to automatically assign MAC addresses from the given switch and switch ports. If you prefer to utilize the old method of mapping MAC addresses to IP addresses, you can skip this section.

When using the p2p-superpod-template, navigate to the first tab labeled "Notes," where you'll find four buttons.

- 1. "Create Table of Contents": This button generates a new tab called "Contents," listing all tabs with links (this option is for generating purposes only).
- "Generate P2P": This button creates a new tab called "p2p_ethernet" automatically populating content from existing tabs such as OOB, MGMT-InBand, and DGX-InBand into a single sheet.
- 3. "**Search & Replace**": This function utilizes the tab labeled "Alias" to search for text in the "p2p_ethernet" tab's Column C and replace it with the corresponding text from Column A.
- 4. **"Populate**": Automatically populate the FLOW and ALOCATE_IP columns based on data from the remaining columns.

Below button will generate		Below button will search and replace	Below button will generate
P2P_ETHERNET (new tab)			
t will combine data from columr	1 [F,H,J,K]	data from the P2P_ETHERNET	Additional data fillup
OOB		based on <u>Alias</u> tab	data in the column FLOW and ALLOCATE_IP
DGX-InBand		Seach column [C]	
MGMT-InBand		Replace value column [A]	

Next, navigate to the **p2p_ethernet** tab. From there, access the *File* menu and choose the "*Downloads*" option, selecting the (.csv) format. Save the file and transfer it to a USB stick, along with Cumulus OS and IBSW OS.

3. Initial Cluster Setup

The deployment stage of a DGX SuperPOD consists of using BCM to provision and manage the Slurm cluster.

1. Configure the NFS server.

User home directories (home/) and shared data (cm_shared/) directories must be shared between head nodes (such as the DGX OS image) and must be stored on an NFS filesystem for HA availability. Because DGX SuperPOD does not mandate the nature of the NFS storage, the configuration is outside the scope of this document. This DGX SuperPOD deployment uses the NFS V3 export path provided in the <u>site</u> <u>survey</u>: /var/nfs/general.

The following parameters are recommended for the NFS server export file /etc/exports.

/var/nfs/general *(rw,sync,no_root_squash,no_subtree_check)

- 2. Configure the DGX systems to PXE boot by default.
 - a. Using either KVM or a crash cart, connect to the DGX system, enter the BIOS menu, and configure Boot Option #1 to be [NETWORK].

	– Copyright (C) 2021 Amer: t Security Boot Save & B	
Boot Configuration Setup Prompt Timeout Bootup NumLock State	1 [0n]	Sets the system boot order
FIXED BOOT ORDER Priori Boot Option #1 Boot Option #2 Boot Option #3 Boot Option #4 Boot Option #5 Boot Option #6	ties [Network] [Disabled] [Disabled] [Disabled] [Disabled] [Disabled]	++: Select Screen ↑↓: Select Item

b. Ensure that other Boot Options are [Disabled] and go to the next screen.

c. Set Boot Option #1 and Boot Option #2 to use IPv4 for Storage 4-2 and Storage 5-2.

Boot Option #1	[UEFI: PXE IPv4 Mellanox Network	Sets the system boot order
	Adapter –	
Boot Option #2	0C:42:A1:0A:33:4B] [UEFI: PXE IPv4	
DUUL UPLIUN #2	Mellanox Network	
	Adapter –	
	0C:42:A1:74:F3:1F]	
Boot Option #3	[Disable]	
Boot Option #4	[Disable]	
Boot Option #5	[Disable]	++: Select Screen
Boot Option #6	[Disable]	↑↓: Select Item
Boot Option #7	[Disable]	Enter: Select
Boot Option #8	[Disable]	+/-: Change Opt.
Boot Option #9	[Disable]	F1: General Help
Boot Option #10	[Disable]	F2: Previous Values
Boot Option #11	[Disable]	F3: Optimized Defaults
Boot Option #12	[Disable]	F4: Save & Exit
		ESC: Exit

- d. Ensure that other Boot Options are [Disabled].
- e. Select Save & Exit.
- 3. On the failover head node and the CPU nodes, ensure that Network boot is configured as the primary option. Ensure that the Mellanox ports connected to the network on the head and CPU nodes are also set to Ethernet mode. This is an example of a system that will boot from the network with Slot 1 Port 2

and Slot 2 Port 2.



- 4. Download the BCM installer ISO.
- Burn the ISO to a DVD or to a bootable USB device.
 It can also be mounted as virtual media and installed using the BMC. The specific mechanism for the latter will vary by vendor.
- 6. Ensure that the BIOS of the target head node is configured in UEFI mode and that its boot order is configured to boot the media containing the BCM installer image.
- 7. Boot the installation media.



- 8. At the grub menu, choose Start Base Command Manager Graphical Installer.
- 9. Select Start installation on the splash screen.



10. Accept the terms of the NVIDIA EULA by checking I agree and then select Next.

1	, 5 5	
Ø 📰 O	10.180.217.154	
🥸 NVIDIA.	Base Command Manager installer	v10.23.11 (UBUNTU2204)
IVDIA EULA	NVIDIA	
ernel modules	NVIDIA AI PRODUCT AGREEMENT	*
lardware info	IMPORTANT NOTICE - PLEASE READ AND AGREE BEFORE USING NVIDIA AI PRODUCTS.	
nstallation source		
Cluster settings	This AI Product Agreement is entered into between the entity you represent or you individually if you do not designate an entity ("Customer") and NVIDIA	
Workload manager	Corporation ("NVIDIA"). This AI Product Agreement consists of the terms and conditions below and all documents attached to or referenced in this AI Product	
Network topology	Agreement (together, the "Agreement"). The AI Product catalogs include products that can be used without payment and paid products and services. Key terms are	
lead node	defined in Section 17.	
Compute nodes	By using or registering to use AI Products, Customer is affirming that Customer has read the Agreement and agrees to its terms. If Customer does not have the	
BMC configuration	required authority to enter into the Agreement or if Customer does not accept all the terms and conditions below, do not use (or register to use) AI Products.	
Networks	1. AI PRODUCTS OFFERINGS.	
Head node interfaces	1.1 Grant.	
	Subject to the terms of the Agreement, Customer's Order Agreement and	
Compute nodes interfaces	Subscription or Perpetual license parameters, and payment of applicable fees, NVIDIA grants Customer a non-exclusive, non-transferable, non-sublicensable	
Disk layout	(except as expressly provided in the Agreement) license to do the following for the duration of the license:	
Disk layout settings	1.1.1 install and use copies of AI Products.	
Additional software	1.1.2 create Derivative Samples and Derivative Models to develop and test	
Summary	services and applications,	
Deployment	 1.1.3 configure the AI Product using the configuration files provided (as applicable), 	
	📄 I agree	
	Continue remotely Load config Si	how config Back Next
1024 x 768		rði 🗛 ssl. 🥥 🔿 🍐

11. Accept the terms of the Ubuntu Server UELA by checking I agree and then select

Ø 🔤 O	10.180.217.154
📀 NVIDIA.	Base Command Manager installer v10.23.11 (UBUNTU220
NVDIA EULA	Ubuntu Server 22.04
Kernel modules	
Hardware info	Licensing Ubuntu is a collection of thousands of computer programs and documents created by a range of individuals, teams and companies.
Installation source	Each of these programs may come under a different licence. This licence policy describes the process that
Cluster settings	we follow in determining which software will be included by default in the Ubuntu operating system.
Workload manager	Copyright licensing and trademarks are two different areas of law, and we consider them separately in Ubuntu. The following policy applies only to copyright licences. We evaluate trademarks on a case-by-case
Network topology	basis.
Head node	Categories of software in Ubuntu The thousands of software packages available for Ubuntu are organised into four key groups or components:
Compute nodes	main, restricted, universe and multiverse. Software is published in one of these components based on whether or not it meets our free software philosophy, and the level of support we can provide for it. In addition,
BMC configuration	or not it meets our free sortware philosophy, and the level of support we can provide for it. In addition, software may be published for Ubuntu as a universal Linux snap package, in which case licenses are determined by the snap publisher and documented in the snap store.
Networks	This policy only addresses the software that you will find in main and restricted, which contain software
Head node interfaces	that is fully supported by the Ubuntu team and must comply with this policy.
Compute nodes interfaces	Ubuntu 'main' component licence policy All application software included in the Ubuntu main component:
Disk layout	Nust include source code. The main component has a strict and non-negotiable requirement that application
Disk layout settings	software included in it must come with full source code. Nust allow modification and distribution of modified copies under the same licence. Just having the source
Additional software	code does not convey the same freedom as having the right to change it. Without the ability to modify software, the Ubuntu community cannot support software, fix bugs, translate it, or improve it.
Summary	Ubuntu 'main' and 'restricted' component licence policy
Deployment	All application software in both main and restricted must meet the following requirements: Must allow redistribution. Your right to sell or give away the software alone, or as part of an aggregate software distribution, is important because:
	I agree
	Continue remotely Load config Show config Back Nex
1024 x 768	(6) A SSL 🔗 🔿 🔿

12. Unless instructed otherwise, select Next without modifying the kernel modules to be loaded at boot time.

	10.180.217.154	
	Base Command Manager installer	v10.23.11 (UBUNTU2204
NVDIA EULA	nel modules	
	r to be able to use all the hardware, it is important that the correct se	
	he hardware in this machine has been probed and the kernel module stances it is not necessary to modify the kernel modules selection, bu	
	kernel modules here.	, , , , , , , , , , , , , , , , , , , , , , ,
Cluster settings		€ ⊕
Workload manager		Path
Network topology		· Ø
Head node		• Ø
Compute nodes		· Ø
aesni_ir BMC configuration		· Ø
Networks		· Ø
Head node interfaces	n -	• Ø
cec		· Ø
		· Ø
	dif pclmul -	· Ø
	=1	• Ø
Additional software cryptd		· Ø
	sina -	• Ø
Deployment dca		· Ø
drm	ns helper -	- Ø
drm kn	ns neiper -	- (/)

13. Verify that the Hardware info is correct and then select Next.

VDIA EULA Hardware info ardware info The following hardware has been detected. If not all hardware has been recognized, you may go back to the kernel modules configuration screen to load extra kernel modules. stallation source	🕸 🚃 💿 🕺 NVIDIA.			0.180.217.154 hand Manager installer	v10.23.11 (U	□ 2 ⁴ × BUNTU2204)
uster settings C Cd Rom C Cd Rom Keyboard Mouse /dev/input/event2 HP Virtual Keyboard etwork topology Mouse /dev/input/mice HP Virtual Keyboard etwork topology Mouse /dev/input/mice HP Virtual Keyboard etwork topology Mouse /dev/input/mice HP Virtual Keyboard etwork topology Mouse /dev/input/mice HP Virtual Keyboard etwork topology Mouse /dev/input/mice HP Virtual Keyboard Network interface [bnxt_en] Network interface [mlx5_core] ms2np0 Network interface [mlx5_core] (mk5_core[ib_joib]] Mice /dev/sdb 30GB Bisk layout Storage /dev/sda /dev/nvme2n1 Mice Mice /dev/nvme2n1 Mice Mice /dev/nvme2n1 Mice Mice Mice /dev/nvme2n1		The f	rdware info	n detected. If not all hardware l		ĺ
orkload manager Mouse /dev/input/mice HP Virual Keyboard ead node ens10f0np0 Network interface [bnxt_en] ompute nodes ens10f1np1 Network interface [bnxt_en] ead node ens10f1np1 Network interface [bnxt_en] ead node ens10f2np1 Network interface [bnxt_en] ead node ens1np0 Network interface [mlx5_core] ead node interfaces ibs3 [mlx5_core[ib_ipoib]] ead node interfaces /dev/sdb 30GB DataTraveler 3.0 ofklayout jdev/sda 959GB MR416i-a Gen10+ isk layout settings /dev/nume2n1 MZXLR6T4HALA-000H3 idditional software /dev/nume0n1 MZXLR6T4HALA-000H3 ummary ensemption ensemption	nstallation source Cluster settings	-		Device	Model	
ead node ompute nodes MC configuration etworks ead node interfaces ompute nodes interfaces ompute nodes interfaces etworks ead node interfaces etworks ead node interfaces etworks ead node interfaces etworks ead node interfaces etwork interfaces etwork interfaces etworks etwork interfaces etwork inte	Vorkload manager Ietwork topology				-	
/dev/sdb 30GB DataTraveler 3.0 /dev/sdb 30GB /dev/sdb 30GB /dev/sda 959GB MR416i-a Gen10+ /dev/nvme2n1 MZXLR6T4HALA-000H3 /dev/nvme1n1 MZXLR6T4HALA-000H3 /dev/nvme0n1 MZXLR6T4HALA-000H3 /dev/nvme0n1 MZXLR6T4HALA-000H3 /mmary - eployment -	fead node Compute nodes BMC configuration Hetworks	Ð	Network Interfaces	ens10flnp1 ens1np0 ens2np0	Network interface [bnxt_m_ Network interface [mlx5_co Network interface [mlx5_co Network interface	l re]
eployment .	ompute nodes interfaces isk layout isk layout settings dditional software	Ð	Storage	/dev/sdb /dev/sda /dev/nvme2n1 /dev/nvme1n1	30GB 959GB MR416i-a Gen10+ MZXLR6T4HALA-000H3 MZXLR6T4HALA-000H3	
	,	Ð	Storage Controllers	• • •	• • •	

For example, that the target storage device and the cabled host network interfaces are present (in this case three NVMe drives are the target storage device, and ens1np0 and ens2np01 are the cabled host network interfaces).

14. On the Installation source screen, choose the appropriate source and then select Next. Running a media integrity check is optional.

	Base Command Manager installer		v10.23.11 (UBUNTU2204)	
NVDIA EULA	DVD/ISO/USB			-
Kernel modules	Select from the list below:			
Hardware info	 /dev/sdb (BCM Install Media) 			
Installation source	Run media integrity check			
Cluster settings				
Workload manager				
Network topology				
Head node				
Compute nodes				S
BMC configuration	•			
Networks				
Head node interfaces				
Compute nodes interfaces				
Disk layout				
Disk layout settings				
Additional software				
Summary				
Deployment				
		Continue remotely	Show config Back Next	
				-

15. On the Cluster settings screen, enter the required information and then select Next.

🕸 📟 💿	10.180.217.154				
<mark>></mark> NVIDIA.	Base Command Manager installer		v10.23.11	(UBUNT	J22
IVDIA EULA	General cluster settings				
Cernel modules	Cluster name:				
lardware info	BCM 10.0 Cluster				
nstallation source	Organization name:				
luster settings	NVIDIA				
Vorkload manager Ietwork topology	Administrator email:				
lead node					
ompute nodes	Send email to the administrator on first boot				
MC configuration	Time zone:				
letworks	(GMT-08:00) America/Los_Angeles				+
ead node interfaces	Time servers:				
ompute nodes interfaces	× 0.pool.ntp.org × 1.pool.ntp.org × 2.pool.ntp.org			>	< •
isk layout	Nameservers:				
isk layout settings	× 8.8.8.8				< =
dditional software	Leave this field empty if you intend to use DHCP for external network				
ummary	Search domains:				
eployment					
	Leave this field empty if you intend to use DHCP for external network				
	Co	ntinue remotely	Show config	Back	N

16. On the Workload manager screen, choose None and then select Next.



17. On the Network topology screen, choose the network type for the data center environment and then select Next.



18. On the Head node screen, enter the Hostname, Administrator password, choose Other for Hardware manufacturer, and then select Next

Ø 0	10.180.217.154		
	Base Command Manager installer	v10.23.11 (UBUNTU2204)	
NVDIA EULA	Head node settings		
Kernel modules	Hostname:		
Hardware info	head-01		
Installation source	Administrator password:		
Cluster settings	•••••	Ø	
Workload manager	Confirm administrator password:		
Network topology Head node		Ø	
Compute nodes	Hardware manufacturer:	42	
BMC configuration	Other		
Networks			
Head node interfaces			
Compute nodes interfaces			
Disk layout			
Disk layout settings			
Additional software			
Summary			
Deployment		k	
	Continue remotel	ly Show config Back Next	
1024 x 768		🖄 🔒 SSL 🛛 🥪 🔿 🍅	

19. Accept defaults in the Compute nodes and then select Next.

\sim

🚳 NVIDIA.	Base Command Manager installer v10.23.11 (UBU	NTU2204)
NVDIA EULA	Compute nodes settings	
Kernel modules	Number of racks:	
Hardware info	1	
Installation source	Number of nodes:	
Cluster settings		
Workload manager	1	
Network topology	Node start number:	
Head node	1	
Compute nodes	Node base name:	
BMC configuration	node	
Networks	Node digits:	
Head node interfaces	3	
Compute nodes interfaces		
Disk layout	Hardware manufacturer:	
Disk layout settings	Other	-
Additional software	R:	
Summary		
Deployment		
	Continue remotely Show config Bac	k Next
1024 x 768	ල් 🔒 SSL	

20. Ensure that the Node base name is node. Other values will be updated later in the installation.

21. On the BMC Configuration screen, choose No for both Head Node and Compute Nodes, and then select Next.

<mark>⊘</mark> NVIDIA.	Base Command Manager installer		v10.23.11 (UB	UNTU2204)
	MC configuration			
Kernel modules	ine configuration			
Hardware info	ead Node	Compute Node	S	
Installation source Will BM0	l head node have IPMI/iDRAC/iLO/CIMC compatible Cs?	Will compute nodes ha compatible BMCs?	ive IPMI/iDRAC/iLO/CI	MC
	Yes No	○ Yes ● No		
orkload manager				
etwork topology				
ead node				
ompute nodes				
4C configuration				
tworks				
ad node interfaces				
npute nodes interfaces				
k layout				
k layout settings				
ditional software				
ımmary				
eployment				
		Continue remotely	Show config Ba	ack Next
1024 x 768			ି 🔒 SSL	0 🙂

These will be updated later in the post install stages.

22. On the Networks screen, enter the required information for internalnet, and then select Next.

	Base Command Manager installer	v10.23.11 (UBUNTU2204)
NVDIA EULA	Networks	
Kernel modules	The following IP networks have been pre-configured. Using the controls below,	the network settings may be altered.
Hardware info	internalnet	\oplus
Installation source	Name:	U
Cluster settings	internalnet	
Workload manager	Base IP address:	
Network topology	10.180.115.128	
Head node	Netmask:	
Compute n <mark>o</mark> des		× •
BMC configuration	255.255.255.192(/26)	Χ *
Networks	Dynamic range start:	
Head node interfaces	10.180.115.168	
Compute nodes interfaces	Dynamic range end:	
Disk layout	10.180.115.183	
Disk layout settings	Domain name:	
Additional software	eth.cluster	
Summary	Gateway:	
Deployment	10.180.115.129	ĩ
	By default the head node will be used as the default gateway.	1
	MTU: Continue remote	ly Show config Back Next
1024 x 768	Continue remote	
1024 X 708		ici 🔒 ssl. 🥝 🔿 🍎

Since a Type 2 network was specified, there are no other network tabs (for example, externalnet Or ipminet).

23. On the Head node interfaces screen, ensure that one interface is configured with the head node's target internalnet IP, and then select Next.

		Base Command Manag	ger installer	v10.23	.11 (UBUNTU2204)
NVDIA EULA	Head nod	de network inter	faces		
Kernel modules					•
Hardware info	Interface	Network	IP addre	355	
nstallation source	ens1np0	× 👻 internalnet	× - \$0.180	.115.189	× - Ō
luster settings					
/orkload manager					
letwork topology					
lead node					
ompute nodes					
MC configuration					
letworks				- 9	
ead node interfaces					
mpute nodes interfaces					
sk layout					
sk layout settings					
lditional software					
immary					
Deployment					
	_		Continue remo	tely Show confi	ig Back Next
1024 x 768				රෝ	A SSL 🔗 🔿 🍅

Other interfaces will be configured by the post install script.

24. On the Compute node interfaces screen, leave the default entries, and then select Next.

	Ba	se Commar	nd Manager insta	aller	v10.	.23.11 (UBUNTU	J2204)
NVDIA EULA	Compute n	odes ne	twork inter	faces			
Kernel modules							۲
Hardware info	Interface		Network		IP offset		
Installation source	BOOTIF	× =	internalnet	\times $-$	0.0.0.0	× -	Ċ
Cluster settings	IP addresses for eacl	of the interfa	aces on all nodes wil	l be assigned autom	atically from a cor	secutive range of	IP
Workload manager	addresses. The start network. For exampl	ing address in	the range is determ	nined by adding the	specified offset to	the base address o	of the
Network topology	to be assigned 10.14					o war cause ule lit:	at node
Head node							
Compute nodes							
BMC configuration							
Networks							
Head node interfaces							
)isk layout							
isk layout settings							
dditional software							
ummary							
eployment							
				Continue rem	notely Show c	onfig Back	Ne🏝
1024 x 768					<u>6</u>] 🔒 SSL 🥏	0 🈃

These will be updated post install.

25. On the Disk layout screen, select the target install location (in this case nvme@n1) and then select Next.



26. On the Disk layout settings screen, accept defaults and then select Next.



These settings will be updated later in the post installation steps.

27. In the Additional software screen, do not choose anything and then select Next.

NVIDIA.	Base Command I	Manager installer v10.23.11 (UBUNTU2	2204)
DIA EULA	Additional software		
nel modules	Please select from available so	ftware	
dware info	CUDA		
tallation source	OFED stack		
ster settings	OFED Stack		
rkload manager			
twork topology			
ad node			
npute nodes			
C configuration			
tworks			
ad node interfaces	k		
mpute nodes interfaces			
k layout			
k layout settings			
ditional software			
nmary			
ployment			
	formation on the s	© ≙ عد ⊗ ummary screen and then select Nex:	0 0
DVIDIA. DIA EULA rnel modules	Base Command I		
DINFIRM THE IN NUDIA. DIA EULA rnel modules rdware info	Base Command 1 Summary Below is a brief summary of some of t Primary internal interface IP:	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189	
DINFIRM THE IN NUDIA. DIA EULA rnel modules rdware info tallation source	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers:	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8	
DINFIRM THE IN NUIDIA. DIA EULA reel modules rdware info tallation source ster settings	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Time servers	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org	
INFIRM THE IN NUDIA. DIA EULA nel modules dware info callation source ster settings rkload manager	Base Command 1 Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 3.5.8.5 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other	
nfirm the in: NVIDIA. NA EULA deal modules dware info allation source ter settings kload manager work topology	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers: Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other Other	
INFIRM THE IN NUDIA. NA EULA nel modules dware info callation source ster settings rkload manager work topology id node	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other	
INFIRM the in: INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA. INVIDIA.	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
nfirm the in: NVIDIA . NA EULA nel modules dware info allation source ster settings kload manager work topology d node spute nodes c onfiguration	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
Infirm the in: NUDIA. DIA EULA nel modules rdware info tallation source ster settings rkload manager work topology ad node npute nodes c configuration works	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
DINFIRM THE IN DIA EULA rel modules rdware info tallation source ster settings rkload manager twork topology ad node mpute nodes C configuration tworks ad node interfaces	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
Infirm the in: Invibia. DIA EULA nel modules dware info callation source ster settings rkload manager work topology ad node npute nodes c configuration works ad node interfaces npute nodes interfaces	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
Dinfirm the in: Din EULA real modules rdware info tallation source tallation source tallation source twork topology ad node mpute nodes IC configuration tworks ad node interfaces mpute nodes interfaces ik layout	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Nex: Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other Other /dev/nyme0n1 No BMC	
Difirm the in: Difirm the in: Difield EULA rel modules rdware info tallation source ster settings rkload manager twork topology ad node mpute nodes C configuration tworks ad node interfaces mpute nodes interfaces k layout k layout settings	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org Other Other Other No BMC No BMC No BMC	
DINFIRM THE IN: NUIDIA. DIA EULA real modules rdware info tallation source ster settings rkload manager twork topology ad node mpute nodes c configuration tworks ad node interfaces mpute nodes interfaces k layout k layout k layout settings ditional software	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8.8 America/Los Angeles 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org None Other /dev/nyme0n1 No BMC	
Onfirm the in NVIDIA . DIA EULA rmel modules	Base Command I Summary Below is a brief summary of some of the Primary internal interface IP: Primary internal interface Netmask: Nameservers: Timezone: Time servers Workload manager Head node hardware vendor Compute nodes hardware vendor Install drives Head node BMC type	Summary Screen and then select Next Manager installer v10.23.11 (UBUNTU2 the installation settings that were selected: 10.180.115.189 26 8.8.8 America/Los_Angeles 0.pool.ntp.org, 1.pool.ntp.org Other Other Other No BMC No BMC No BMC	

The Summary screen provides an opportunity to confirm the Head node and basic cluster configuration before deployment begins. This configuration will be updated/modified for DGX SuperPOD after deployment is complete. If any values do

not match expectations, use the Back button to navigate to the appropriate screen to correct any mistakes.

29. Once the deployment is complete, select Reboot.

NVDIA EULA Installation Kernel modules Overview of install Hardware info Parsing build Installation source Parsing build Cluster settings Partitioning he Workload manager Partitioning he Network topology Partitioning he Network topology Partitioning he Network topology Configuring kenel and setting up bootloader Head node Installing bear divertion packages to software images(s) BMC configuration Installing BCM packages to software images(s) BMC configuration Installing films election of Python packages to software images(s) Networks Creating node installer NFS image Networks Creating node installer NFS image Networks Prinatzing installation Notores Finalizing installation Sommary Installation is complete Parloyment La/14 Chubmatically reboot after installation is complete		B	ase Command Mana	iger installer				
kernel modules Overview of install Hardwars info Parsing build Reboot now? Cluster settings Partitioning he No N	NVDIA EULA	Installation		×				
installation source Parsing build Reboot now? installing CD Partitioning CD<!--</td--><td>Kernel modules</td><td></td><td>completed!</td><td></td><td></td><td></td><td></td><td></td>	Kernel modules		completed!					
installation source • Mounting CD Cluster settings • Partitioning h workload manager • Installing Ubu Network topology • Installing head node BCM packages Network topology • Installing kernel and setting up bootloader Head node • Installing Date distribution packages to software image(s) Compute nodes • Installing BCM packages to software image(s) BMC configuration • Installing BCM packages to software image(s) Retworks • C creating node installer NF image Head node interfaces • Initializing management daemon Compute nodes interfaces • Initializing management daemon Disk layout • Initializing management daemon Summary • Automatically reboot after installation is complete Paployment • Automatically reboot after installation is complete	Hardware info	✓ Parsing build	Report now?					
Head node Configuring kernel and setting up bootloader Installing Uburtu Server 22.04 base software image(s) Compute nodes Installing DBurtu Server 22.04 base software image(s) BMC configuration Installing BCM packages to software image(s) Installing BCM packages to software image(s) Networks Creating node installer NF5 image Creating management daemon Disk layout Initializing management daemon Additional software Summary Deployment Automatically reboot after installation is complete Show config Nuternatically reboot after installation is complete 	Installation source	Mounting CD/	Reboot now:					
Head node Configuring kernel and setting up bootloader Installing Ubuntu Server 22.04 base software image(s) Compute nodes Installing Dackages to software image(s) Installing BCM packages to software image(s) BMC configuration Installing Offine selection of Python packages Creating node installer NFS image Creating node installer NFS image Initializing management daemon Disk layout Initializing management daemon Objek layout Initializing management daemon Disk layout Initializing management daemon Objek layout Initializing management daemon Disk layout settings Initializing management daemon Summary Initializing management daemon Initializing management daemon Initializing management daemon Initializing management daemon Disk layout settings Initializing management daemon Disk layout settinge	Cluster settings			_				
Head node Configuring kernel and setting up bootloader Installing Uburtu Server 22.04 base software image(s) Compute nodes Installing DBurtu Server 22.04 base software image(s) BMC configuration Installing BCM packages to software image(s) Installing BCM packages to software image(s) Networks Creating node installer NF5 image Creating management daemon Disk layout Initializing management daemon Additional software Summary Deployment Automatically reboot after installation is complete Show config Nuternatically reboot after installation is complete 	Workload manager			No Yes				
Head node Configuring kernel and setting up bootloader Installing Uburtu Server 22.04 base software image(s) Compute nodes Installing DBurtu Server 22.04 base software image(s) BMC configuration Installing BCM packages to software image(s) Installing BCM packages to software image(s) Networks Creating node installer NF5 image Creating management daemon Disk layout Initializing management daemon Additional software Summary Deployment Automatically reboot after installation is complete Show config Nuternatically reboot after installation is complete 	Network topology	🗸 Installing head	d node BCM packages					
Compute nodes Installing base distribution packages to software images(s) Installing BCM packages to software images(s) Installing dfline selection of Python packages Creating node installer NFS image Creating node installer NFS image Finalizing installation Initializing management daemon Initializing management daemon<	Head node							
BMC configuration Installing offline selection of Python packages Creating node installer NFS image Finalizing installation Initializing management daemon Initializing management daemon Initializing management daemon Additional software Summary Deployment Automatically reboot after installation is complete Show config Installing Back Reboot Rebot Rebot Rebot	Compute nodes							
Networks Creating node installer NFS image Finalizing installation Initializing management daemon Disk layout Initializing management daemon Disk layout settings Initializing management daemon Additional software Initializing management daemon Summary Initializing management daemon Deployment Initializing management daemon Summary Initializing management daemon Summary Initializing management daemon Summary Initializing management daemon Beployment Initializing management daemon Summary Initializing management daemon Summary Initializing management daemon Beployment Initializing management daemon Summary Initializing management daemon Summary Initializing management daemon Beployment Initializing management daemon Initial construction Initial construction Initial construction Initial construction Initial construction Initial construction Sum	BMC configuration							
Additional software Initializing management daemon Summary Initializing management daemon Deployment Initializing management daemon 14/14 Initializing management daemon Summary Initializing management daemon Deployment Initializing management daemon 14/14 Initializing management daemon Initializing management daemon Initializing management daemon Summary Initializing management daemon Deployment Initializing management daemon Show config Install log Back Reboot	Networks			ackages				
Compute nodes interfaces Disk layout Disk layout settings Additional software Summary Deployment Id / 14 I Automatically reboot after installation is complete Show config Install log Back Reboot	Head node interfaces	✓ Finalizing inst.	allation					
Disk layout settings Additional software Summary Deployment <u>14/14</u> Automatically reboot after installation is complete Show config Install log Back Reboot	Compute nodes interfaces	🖌 Initializing ma	nagement daemon					
Additional software Summary Deployment Automatically reboot after installation is complete Show config Install log Back Reboot	Disk layout							
Summary Deployment Automatically reboot after installation is complete Show config Install log Back Reboot	Disk layout settings				8			
Deployment Automatically reboot after installation is complete Show config Install log Back Reboot	Additional software							
14/14 Automatically reboot after installation is complete Show config Install log Back Reboot	Summary							
Automatically reboot after installation is complete Show config Install log Back Reboot								
		Automatically re	boot after installation is	The second se				
1024 x 768 (Ö A SSL 🖉 🔿 🖒					Show config	Install log Back	Reboot	
	1024 x 768					(c) A SSL	00	

30. License the cluster by running the request-license and providing the product key. $_{\mbox{sudo}\mbox{-i}\mbox{ request-license}}$

Product Key (XXXXXX-XXXXXX-XXXXXX-XXXXXXX-XXXXXX):

31. Options:

4

- 1. If using the old method of MAC to IP allocation, skip line 32.
- 2. If employing the new method: Automatically detect MAC addresses based on switch and switchport, proceed to the next step.
- 3. Before advancing with the execution of the network automation application, certain prerequisites are necessary. Do as following:
 - Copy the "p2p_ethernet.csv" file from the USB stick to the following path: /cm/local/apps/bcm-superpod-network/config/p2p_ethernet.csv mv p2p_ethernet.csv /cm/local/apps/bcm-superpod-network/config/
- 32. Load the bcm-superpod-network module. module load bcm-superpod-network
- 33. Run the bcm-netautogen script.
- bcm-netautogen

Noticed New additional information has been provided.



Data was extracted from the p2p_ethernet.csv file to compute the quantities of Network Switches, DGX, IBSW, and PDUs. Accurate values must be provided during menu execution, which will be updated in future releases to utilize the count as physical cable connections.



The following generated files are important and contain data:

- Site network configuration /cm/local/apps/bcm-superpodnetwork/config/network-configuration.yml
- Site network allocations /cm/local/apps/bcm-superpodnetwork/config/network-allocations.yml

- Switch connection /cm/local/apps/bcm-superpod-network/config/switchconnections.yml
- IP Allocation Readme file /cm/local/apps/bcm-superpodnetwork/config/ip_allocations.md
- 34. Download and move cumulus-linux-5.10.0-mlx-amd64.bin and image-X86_64-3.11.2300.img to the following directory on the head node. Contact your TAM for access to the correct file and move the file to the following directory on the head node

```
mv cumulus-linux-5.10.0-mlx-amd64.bin
/cm/local/apps/cmd/etc/htdocs/switch/image/
mv image-X86_64-3.11.2300.img /cm/local/apps/cmd/etc/htdocs/switch/image/
```

- 35. Load the bcm-post-install module. module load bcm-post-install/
- 36. Run the bcm-pod-setup script.

These parameters were used:

- –C sets the base address of the computenet network.
- –s sets the base address of the storagenet network.
- -I sets the installation source.

bcm-pod-setup -C 100.126.0.0/16 -S 100.127.0.0/16 -I /dev/sdb

37. Check the nodes and their categories.

Extra options are used for device list to make the format more readable. %cmsh

```
%[bcm-head-01]%device list -f hostname:20,category:10
hostname(key) category
bcm-cpu-01 default
bcm-dgx-a100-01 dgx-a100
bcm-dgx-h100-01 dgx-h100
```

38. Confirm the config is correct for bcm-dgx-h100-01 / bcm-dgx-a100-01.

```
[bcm-head-01->device[bcm-dgx-h100-01]]% interfaces
```

[bcm-head-	01->device[bcm-dgx-h100-0	01]->interf	aces]% list		
Туре	Network device name	IP		Network	Start if
bmc	ipmi0		10.0.92.50	ipminet	
always	1910		1010192100	1pm1nc c	
bond	bond0 [prov]		10.0.93.12	dgxnet	
always				-	
physical	enp170s0f1np1 (bond0)	0.0.0.0			always
physical	enp41s0f1np1 (bond0)	0.0.0.0			always
physical	ibp154s0		100.126.5.14	ibnetcompute	always
physical	ibp170s0f0		100.127.2.2	ibnetstorage	always
physical	ibp192s0		100.126.6.14	ibnetcompute	always
physical	ibp206s0		100.126.7.14	ibnetcompute	always
physical	ibp220s0		100.126.8.14	ibnetcompute	always
physical	ibp24s0		100.126.1.14	ibnetcompute	always
physical	ibp41s0f0		100.127.1.2	ibnetstorage	always
physical	ibp64s0		100.126.2.14	ibnetcompute	always
physical	ibp79s0		100.126.3.14	ibnetcompute	always



39. Check the Ethernet Switch are in the devices.

[bcm-head-01->device]% Type	list Hostname (key)	MAC	Category	Ip	Network
HeadNode	bcm-head-01	B8:CE:F6:FF:98:A0		10.0.105.189	internalnet
PhysicalNode	slogin-01	00:00:00:00:00:00	default	10.0.105.140	internalnet
PhysicalNode	slogin-02	00:00:00:00:00:00	default	10.0.105.141	internalnet
PhysicalNode	storage-cpu-01	00:00:00:00:00:00	default	10.0.105.139	internalnet
PhysicalNode	storage-dgx-01	00:00:00:00:00:00	dgx-a100	10.0.0.11	dgxnet1
PhysicalNode	storage-dgx-02	00:00:00:00:00:00	dgx-a100	10.0.0.12	dgxnet1
Switch	IPMI-01	00:00:00:00:00:00		10.0.255.194	loopback
Switch	SJC4-C25-T0R-04	B8:59:9F:49:CC:9C		10.0.105.3	ipminet2
Switch	T0R-01	00:00:00:00:00:00		10.0.128.61	ipminet
Switch	T0R-02	00:00:00:00:00:00		10.0.128.60	ipminet
Switch	ib01	E8:EB:D3:31:FE:D2		10.0.105.41	ipminet2

Validate under the "Type" Switches are added after executing *bcm-pod-setup*

40. Add Switch *credential*, under each IPMI, TOR and SPINE switch.

Example:

[bcm-head-01->device]% use IPMI-01	
[bcm-head-01->device[IPMI-01]]% accesssettings	
[bcm-head-01->device*[IPMI-01*]->accesssettings*]% set username cu	
<pre>[bcm-head-01->device*[IPMI-01*]->accesssettings*]% set password <pre></pre></pre>	pass>

commit

- quit
- 41. To allocate IP via switch port:
 - After running bcm-pod-setup, once all the networks, and devices object is added to the Bright
 - Make sure the IPMI switch is UP in the Bright before moving to the next step
 - Based on the switch and switch port configuration for each node, navigate to the device then nodes, and execute the below command:
 - **setmacviaswitchport** Set the MAC of a device via the MAC found on its switch ports

- It will access the switch and pull the MAC address based on the switch port allocation.



42. To gather UFM metrics

- Add UFM to the Bright with mgmt. IP address
- Make sure UFM Promethus exporter is enable in UFM
- ## You can check by curl command from bright:

curl http://<UFM-IP>:9001/metrics

Configure Bright with following: monitoring setup add prometheus UFM set urls https://<UFM-IP>:9001/metrics set -e NoPostAllowed yes nodeexecutionfilters active commit

Wait for (~2mins) for data to be collected
get measurables

To plot
monitoring labeledentity
list

Using the index value: instantquery <index value>

4. Head Node Configuration

This section addresses configuration steps to be performed on BCM head nodes.

Use the root (not cmsh) shell.

- 1. In /cm/local/apps/cmd/etc/cmd.conf, uncomment the AdvancedConfig parameter.
 AdvancedConfig = { "DeviceResolveAnyMAC=1" } # modified value
- Restart the CMDaemon to enable dependable PXE booting from bonded interfaces.
 # systemctl restart cmd

The cmsh session will be disconnected because of restarting the CMDaemon. Type connect to reconnect after the CMDaemon has restarted. Or enter exit and then restart cmsh.



Older method: The steps below are only necessary if you are employing the MAC to IP Allocation method. Newer methods omit assigning MAC addresses based on the ports **** MOVE to STEP 9**

3. The steps that follow are performed on the head node and should be run for all DGX systems.

Note: Double check the MAC address for each interface and the IP number for the bond0 interface. Mistakes here will be difficult to diagnose.

For DGX A100 systems, the commands should be like this code block.

```
# cmsh
% device
% use bcm-dgx-a100-01
% interfaces
% use enp225s0f1np1
% set mac B8:CE:F6:2F:08:69
% use enp97s0f1np1
% set mac B8:CE:F6:2D:0E:A7
% ...
% commit
For DGX H100 systems, the commands should be like this code block.
# cmsh
% device
% use bcm-dgx-h100-01
% interfaces
% use enp170s0f1np1
% set mac B8:CE:F6:2F:08:69
% use enp41s0f1np1
% set mac B8:CE:F6:2D:0E:A7
% ..
% commit
```

4. Verify the configuration.

```
This example is for a DGX A100 system. The output for a DGX H100 system is similar.
   [bcm10-headnode->device[dgx-01]]% get provisioninginterface
   bond0
   % interfaces
   % list
             Network device name IP
                                                             Network Start if
   Туре
                -----
                                                  -----
                                                  10.130.111.68 ipminet
              ipmi0
   bmc
   always
bond bond0 [prov]
physical enp225s0f1np1 (bond0) 0.0.0.0
physical enp97s0f1np1 (bond0) 0.0.0.0
                                             10.130.122.5 internalnet
                                                                              always
                                                                               always
                                                                                 always
5. Identify the nodes by setting the MAC address for the provisioning interface for
   each node to the MAC address listed in the site survey.
   % device
   % use bcm-dax-h100-01
   % set mac b8:ce:f6:2f:08:69
   % use bcm-dgx-h100-02
   % set mac 0c:42:a1:54:32:a7
   % use bcm-dgx-h100-03
   % set mac 0c:42:a1:0a:7a:51
   % use bcm-dgx-h100-04
   % set mac 1c:34:da:29:17:6e
   % foreach -c dgx-h100 (get mac)
   B8:CE:F6:2F:08:69
   0C:42:A1:54:32:A7
   0C:42:A1:0A:7A:51
   1C:34:DA:29:17:6E
6. If all the MAC addresses are set properly, commit the changes.
   % device commit
   % quit
7. Set the MAC addresses for the Ethernet interfaces.
   For control nodes connected to DGX A100 systems, use the following commands.
   % device
   % use bcm-cpu-01
   % interfaces
   % use ens2f0np0
   % set mac 88:e9:a4:92:26:ba
   % use ens2f1np1
   % set mac 88:e9:a4:92:26:bb
   % commit
```

```
For control nodes connected to DGX H100 systems, use the following commands.
```

```
% device
% use bcm-cpu-01
% interfaces
% use enp37s0np0
% set mac 88:e9:a4:92:26:ba
% use enp65s0np0
% set mac 88:e9:a4:92:26:bb
% commit
```

If the head node uses a bonded interface, use the following commands. You may need to reboot the head node and redo request-license steps

% device

```
% use headnode-01
% interfaces
% use ens1np0
% clear ip
% clear network
% add physical ens2np0
% set mac 88:e9:a4:20:18:d8
% add bond bond0
% append interfaces ens1np0 ens2np0
% set mode 1
% set network internalnet
% set ip 10.180.115.189
% ...
% ..
% set provisioninginterface bond0
% interfaces
% use ipmi0
% set ip 10.180.217.154
```

8. Set the IP address for the bond0 interface.

```
% device
% use bcm-cpu-01
% interfaces
% use bond0
% set ip 10.127.3.15
% commit
```

9. Power on and provision the cluster nodes.

For initial provisioning, the cluster nodes must be powered on either directly or by using a KVM. It will take several minutes for the nodes to go through their BIOS. After that, node status progress will be displayed as the nodes are being provisioned. Monitor the /var/log/messages and /var/log/node-installer log files to verify that everything is proceeding smoothly.

Newer method:

Prerequisite

- The switch must be in the **UP** position to retrieve the MAC address from the switch and switchport specified in the CSV file.
 - If any TORs are offline, Bright will be unable to retrieve MAC addresses from the switch.

Verify:

 Ensure that the weather node interfaces have been assigned a switch and switchport. Navigate to cmsh > device > use <node> and execute the "show" command.

 Switch ports
 IPW01091129 (Internation)

 • Interfaces
 IPW01091129 (Internation)

 • Next Step:
 <4 in submode>

- Reboot the Node
 - Detailed MAC addresses per node are not necessary as Bright will automatically detect them based on Switch and Switchport assignments.

5. High Availability

1. Verify that the head node has power control over the cluster nodes.

```
% device
% power -c dgx-h100 status
[-head1->device]% power -c dgx-h100 status
ipmi0 ...... [ ON ] bcm-dgx-h100-01
ipmi0 ...... [ ON ] bcm-dgx-h100-02
ipmi0 ...... [ ON ] bcm-dgx-h100-03
ipmi0 ...... [ ON ] bcm-dgx-h100-04
[bcm-head-01->device]%
```

2. Power off the cluster nodes.

The cluster nodes must be powered off before configuring HA.

% power -c dgx-h100 off				
ipmi0	[OFF] bcm-dgx-h100-01	
ipmi0	[OFF] bcm-dgx-h100-02	
ipmi0	[OFF] bcm-dgx-h100-03	
ipmi0	[OFF] bcm-dgx-h100-04	

- 3. Start the cmha-setup CLI wizard as the root user on the primary head node. # cmha-setup
- 4. Choose Setup and then select SELECT.

Welcome to the Bright Cluster Manager High Availability Setup Utility. Please choose 'Setup' to enter the failover settings menu, 'Shared Storage' to setup shared storage, 'Status' to view the failover status, if high availability has already been setup. Choose 'Help' to see a detailed description of the options available.
SeturConfigure failover seturShared StorageConfigure shared storageStatusView failover statusHelpCmha-setup help
<pre>definition of the second second</pre>

5. Choose Configure and then select NEXT.



6. Verify that the cluster license information found cmha-setup is correct and then select CONTINUE.

The following MAC addresses have been found in the license information:
04:3F:72:E7:67:07 14:02:EC:DA:AF:18
If they are correct, then please press 'Continue'.
If not, one of the following has to be done:
 If you have not activated your Product Key, please run request-license and follow instructions. If you have run out of licenses, please contact your reseller, or contact our support.
Press 'BACK' to go back to the failover setup menu.
SB% -

7. Provide an internal Virtual IP address that is to be used by the active head node in the HA configuration.



8. Provide the name of the secondary head node and then select NEXT.

Name:	basepod-head2	
	< NEXT > < BAG	ск >

9. Because DGX SuperPOD uses the internal network as the failover network, select $_{\mbox{\scriptsize SKIP}}$

This screen takes parameters of a dedicated failover network that will be created. Please keep in mind, to select a network name different from those that have been defined already. Name: Base address: 10.151.0.0 Netmask: 255.255.0.0 Domain name: failover.cluster	S
< NEXT > < SKIP > < BACK >	

10. Configure the IP addresses for the secondary head node that the wizard is about to create and then select NEXT.

ens1f1np1[10.130.122.0/24] ens10f1 [10.130.121.0/24] ipmi0 [10.130.111.64/26]	10.130.121.254	10.130.122.253 10.130.121.253 10.130.111.67
ens10f0 [10.130.111.64/26]		10.130.111.125
< NEXT	> < BACK :	>

11. The wizard shows a summary of the information that it has collected. The VIP that will be assigned to the internal and external interfaces, respectively.



12. Select Yes to proceed with the failover configuration.

Proceed with failover setup? This will write the new setting to the failover config.
< Y@s > < No >

13. Enter the root password and then select OK.

Please enter the mysql root password: ************************************	
< OK > <cancel></cancel>	

14. The wizard implements the first steps in the HA configuration. If all the steps show OK, press ENTER to continue. The progress is shown here.

or, press enter to continue. The progress is show		icit	
Initializing failover setup on master	[0K]
Updating shared internal interface	[0K]
Updating shared external interface	[0K]
Updating extra shared internal interfaces	[0K]
Cloning head node	[0K]
Updating secondary master interfaces	[0K]
Updating Failover Object	[0K]
Restarting cmdaemon	[0K]
Press any key to continue			

15. When the failover setup installation on the primary master is complete, select ok to exit the wizard

Nov	e failover setup initialization on the primary master is done. w boot the secondary master into the rescue environment and run the llowing command:
/ci	n/cm-clone-installfailover
and	d follow the instructions.
Fa the cor see	ce the installation has begun, select 'Install Progress' from the ilover setup menu, to see e installation progress of the clone machine. When the installation is mplete, and the condary master is up, select 'Finalize' from the Failover setup menu, complete the ilover setup process.
	94% < 0 K >

- 16. PXE boot the secondary head node and then select RESCUE from the grub menu. Because this is the initial boot of this node, it must be done outside of BCM (BMC or physical power button).
- 17. Select RESCUE from the grub menu.



19. After the secondary head node has booted into the rescue environment, run the /cm/cm-clone-install --failover command, then enter yes when prompted. The secondary head node will be cloned from the primary.



20. When cloning is completed, enter y to reboot the secondary head node. The secondary must be set to boot from its hard drive. PXE boot should not be enabled.
21. Wait for the secondary head node to reboot and then continue the HA setup procedure on the primary head node.

I *Welcome to the Cluster Manager rescue environment*	
Creating failover/clone nodes:	
I × Install the secondary head node	
\$ /cm/cm-clone-installfailover	
I * Create a clone of the primary head node	
\$ /cm/cm-clone-installclonehostname=new-hostname	
+ Install the secondary (failover) head node and reboot automatically	
\$ ∕cm/cm-clone-installfailouerreboot	
I ≈ Help	
↓ \$ /cm/cm-clone-installhelp	
<pre>mux ClusterManager 5.13.0-39-generic #44"20.04.1-Ubuntu SMP Thu Mar 24 1 otdClusterManager: The Con/cm-Clone-install</pre>	

22. Select finalize from the cmha-setup menu and then select NEXT. This will clone the MySQL database from the primary to the secondary head node.

		failover setup, 'Finalize' to finalize dary head node has been installed. Choose	
'Clone Ins	stall' to see the insta	allation instructions for the secondary iguration has been completed. Choose	
'Install A	progress', if the secon	ndary head node is being installed. Choose ing failover configuration.	
	Configure Clone Install	Configure failover setup View install instructions	
		View install progress Finalize failover setup	
	Undo Failover	Remove failover setup	
	Main	Main menu	
L			
	< NEXT >	< BACK >	

23. Select CONTINUE on the confirmation screen.

CONFTRM
CONFIRM
firmation is required to proceed.
following tasks will be performed:
under an and an hand and we address in andress der base
Update secondary head node MAC address in cmdaemon database. Clone MySQL database(s) to the secondary head node.
ctone hysic database(s) to the secondary head hode.
ss 'CONTINUE' to proceed with the finalize operation,
press 'BACK' to return to the failover Setup menu.
<pre><continue></continue></pre> < brack >

24. Enter the root password and then select OK.

Please enter the mysql root password: [**********	—
<mark>< OK ></mark> <cancel></cancel>	

25. The cmha-setup wizard continues. Press ENTER to continue when prompted.



Updating secondary master mac address [0K]
Initializing failover setup on bcm-head-02 [0K]
Stopping cmdaemon [0K]
Cloning cmdaemon database [0K]
Checking database consistency [0K]
Starting cmdaemon, chkconfig services[0K]
Cloning workload manager databases [0K]
Cloning additional databases [0K]
Update DB permissions [0K]
Checking for dedicated failover network [0K]
Press any key to continue		

26. The Finalize step is now completed. Select REBOOT and wait for the secondary head node to reboot.



27. The secondary head node is now UP.

,									
% device list -f	hostname:20,cat	egory:12,ip:2	20,sta	tus:1	5				
hostname (key)	category	ip					statı	IS	
bcm-head-01		16	9.130.	122.2	54	[UP]	
bcm-head-02		16	9.130.	122.2	53	[UP]	
bcm-dgx-h100-01	dgx-h100	10.130.122.5	5	[DOWN]			
bcm-dgx-h100-02	dgx-h100	10.130.122.0	5	[DOWN]			
bcm-dgx-h100-03	dgx-h100	10.130.122.7	7	[DOWN]			
bcm-dgx-h100-04	dgx-h100	10.130.122.8	3	[DOWN]			

28. Select Shared Storage from the cmha-setup menu and then select SELECT. In this final HA configuration step, cmha-setup will copy the /cm/shared and /home directories to the shared storage and configure both head nodes and all cluster nodes to mount it.

	Welcome to the Bright Cluster Manager High Availability Setup Utility. Please choose 'Setup' to enter the failover settings menu, 'Shared Storage' to setup shared storage, 'Status' to view the failover status, if high availability has already been setup. Choose 'Help' to see a detailed description of the options available. Setup Status Help Configure failover setup Configure shared storage View failover status Cmha-setup help
_	<pre><guit></guit></pre>

29. Choose NAS and then select SELECT.

The following shared stype of shared storage	torage solutions are supported. Please select from the menu below.	a
DAS	Network Attached Storage Direct Attached Storage Distributed Redundant Block Device	
	< <mark>select></mark> < back >	

- Please select resources that will be shared: [★] /cm/shared ★ /home NEXT > < BACK >
- 30. Choose both /cm/shared and /home and then select NEXT.

31. Provide the IP address of the NAS host, the paths that the /cm/shared and /home directories should be copied to on the shared storage, and then select NEXT. In this case, /var/nfs/general is exported, so the /cm/shared directory will be copied to 10.130.122.252:/var/nfs/general/cmshared, and it will be mounted over /cm/shared on the cluster nodes.

Please fill in NAS parameters	
NAS host: Path to /cm/shared: Path to /home:	10.130.122.252 /var/nfs/general/cmshared /var/nfs/general/home
< NEXT >	< BACK >

32. The wizard shows a summary of the information that it has collected. Select EXIT to continue.

NAS Host: 10,180.116.4 NAS Paramas: Path to /cm/shared: /bcmvikinghome Path to /home: /bcmvikingshared	NAS Params: Path to /cm/shared: /bcmvikinghome Path to /home: /bcmvikingshared		SUMMARY NAS shared storage config
		S	NAS Params:

33. When asked to proceed with the NAS setup, select Yes to continue. This will initiate a copy and update to fsexports.



35. When setup completes, press any key to finish HA setup. The progress is shown here:

1 5				
Copying NAS data	. [OK]	
Mount NAS storage	. [OK]	
Remove old fsmounts	. [OK]	
Add new fsmounts	. [OK]	
Remove old fsexports	. [OK]	
Write NAS mount/unmount scripts	. [OK]	
Copy mount/unmount scripts	. [OK]	
Press any key to continue				

36. cmha-setup is now complete. EXIT the wizard to return to the shell prompt.



37. Run the cmha status command to verify that the failover configuration is correct and working as expected.

The command tests the configuration from both directions: from the primary head node to the secondary, and from the secondary to the primary. The active head node is indicated by an asterisk.

```
# cmha status
Node Status: running in active mode
bcm-head-01* -> bcm-head-02
 failoverping [ OK ]
 mysql [ OK ]
            [ OK ]
 ping
 status
            [ OK ]
bcm-head-02 -> bcm-head-01*
 failoverping [ OK ]
             [ OK ]
 mysql
 ping
             [ OK ]
             [ OK ]
 status
```

38. Verify that the /cm/shared and /home directories are mounted from the NAS server.

```
# mount
   . . . some output omitted . . .
   10.130.122.252:/var/nfs/general/cmshared on /cm/shared type nfs4
   (rw, relatime, vers=4.2, rsize=32768, wsize=32768, namlen=255, hard, proto=tcp, timeo=600, retrans=2
   , sec=sys, clientaddr=10.130.122.253, local_lock=none, addr=10.130.122.252)
   10.130.122.252:/var/nfs/general/home on /home type nfs4
   (rw, relatime, vers=4.2, rsize=32768, wsize=32768, namlen=255, hard, proto=tcp, timeo=600, retrans=2
   , sec=sys, clientaddr=10.130.122.253, local_lock=none, addr=10.130.122.252)
39. Login to the head node to be made active and run cmha makeactive.
   # ssh bcm-head-02
   # cmha makeactive
   _____
   This is the passive head node. Please confirm that this node should become
   the active head node. After this operation is complete, the HA status of
   the head nodes will be as follows:
   bcm-head-02 will become active head node (current state: passive)
   bcm-head-01 will become passive head node (current state: active)
   _____
   Continue(c)/Exit(e)? c
   Initiating failover..... [ OK ]
   bcm-head-02 is now active head node. makeactive successful
40. Run the cmsh status command again to verify that the secondary head node has
   become the active head node.
   # cmha status
   Node Status: running in active mode
   bcm-head-02* -> bcm-head-01
     failoverping [ OK ]
    mysql [ OK ]
    ping
                [ OK ]
               [ OK ]
     status
   bcm-head-01 -> bcm-head-02*
     failoverping [ OK ]
     mysql [ OK ]
     ping
                [ OK ]
     status
               [ OK ]
```

41. Manually failover back to the primary head node by running cmha makeactive.

```
# ssh bcm-head-01
   # cmha makeactive
   _____
   This is the passive head node. Please confirm that this node should become
   the active head node. After this operation is complete, the HA status of
   the head nodes will be as follows:
   bcm-head-01 will become active head node (current state: passive)
   bcm-head-02 will become passive head node (current state: active)
   _____
   Continue(c)/Exit(e)? c
   Initiating failover..... [ OK ]
   bcm-head-01 is now active head node, makeactive successful
42. Run the cmsh status command again to verify that the primary head node has
   become the active head node.
                                               # cmha status
   Node Status: running in active mode
   bcm-head-01* -> bcm-head-02
    failoverping [ OK ]
    mysql [ OK ]
              [ OK ]
    ping
    ping [ UK ]
status [ OK ]
   bcm-head-02 -> bcm-head-01*
    failoverping [ OK ]
    mysql [ OK ]
ping [ OK ]
status [ OK ]
43. Power on the cluster nodes.
   # cmsh -c "device ; power -c dgx-h100 on"
   ipmi0 ..... [ ON ] bcm-dgx-h100-01
   ipmi0 ..... [ ON ] bcm-dgx-h100-02
   ipmi0 ..... [ ON ] bcm-dgx-h100-03
   ipmi0 ..... [ ON ] bcm-dgx-h100-04
44. This concludes the setup and verification of HA.
```

6. Slurm Setup

- 1. Update the interface names on the slogin nodes.
 - % device use slogin-01

If the slogin-01 does not have the expected interface names, update the interface names.

% use networkdevicename

% set networkdevicename new-name

- Assign the MAC addresses to the slogin nodes. device use slogin-01 set mac <MAC address>
- 3. Power on and install the slogin nodes.
- 4. Run the bcm-install-slurm script. Use the following parameters:
 - Installation source for the --bcm-media parameter. It can be either a USB or a path to a *.iso file.
 - Use the -A parameter to run the script in air-gapped mode.
 - If CMHA is set up but has failover ping errors, append --ignore-ha-errors.
 - If there is only one slogin node, append --ignore-missing-login-node.

bcm-install-slurm -A --bcm-media <path to installer image or usb device to mount>

5. Confirm that the slurmd file is present in DGX image before provisioning DGX nodes, if not create it.

The same file is needed for both DGX A100 and DGX H100 systems. This example is for DGX H100 systems. It is observed that NCCL tests with PMIX need this file.

```
cat /cm/images/dgx-os-6.3-h100-image/etc/sysconfig/slurmd
PMIX_MCA_ptl=^usock
PMIX_MCA_psec=none
PMIX_SYSTEM_TMPDIR=/var/empty
PMIX_MCA_gds=hash
```

6. Reboot the slogin and compute nodes.

```
cmsh
device
reboot -c slogin
reboot -c dgx-h100
```

7. To simplify the configuration, modify the slurmclient-gpu role to remove the slurm-client role and convert slurm-client-gpu to instead use that name.

```
cmsh
configurationoverlay
remove slurm-client
commit
use slurm-client-gpu
set name slurm-client
commit
```

8. For DGX A100 systems, clear the type value and set the correct core association with each GPU entry for maximum performance.

cmsh configurationoverlay use slurm-client roles use slurmclient genericresources use gpu0 clear type set cores 48-63,176-191 use gpu1 clear type set cores 48-63,176-191 use gpu2 clear type set cores 16-31,144-159 use gpu3 clear type set cores 16-31,144-159 use gpu4 clear type set cores 112-127,240-255 use gpu5 clear type set cores 112-127,240-255 use gpu6 clear type set cores 80-95,210-223 use gpu7 clear type set cores 80-95,210-223 commit

9. For DGX H100 systems, generic resources are set to autodetect.

```
Use this script.
```

```
cmsh
wlm
set gpuautodetect nvml
commit
configurationoverlay
use slurm-client
roles
use slurmclient
set gpuautodetect nvml
commit
genericresources
foreach * (remove)
commit
add autodetected-gpus
set name gpu
set count 8
set addtogresconfig yes
commit
```

Note: addtogresconfig is set by default to "yes" and does not need to be set explicitly.

Which should yield output like this.

The gres.conf file will be updated automatically by BCM—these settings align with the expectations of various scripts and tools in the NVIDIA ecosystem and will then maximize the compatibility of this environment with those scripts and tools.

10. If the /home directory is not mounted on the nodes, increase the number of retries. Due to a race condition between the bond@ interface being up and /home being mounted, sometimes /home will not be mounted. Increasing the number of retries should fix the issue.

```
cmsh
category
use dgx-h100
fsmounts
use /home
set mountoptions "x-systemd.mount-timeout=150,defaults,_netdev,retry=5,vers=3"
```

The pod setup might leave stale repos in an air gapped environment. In which case following files need to be removed manually on the login nodes

```
cd /etc/apt/sources.list.d/
```

Disable following repos:

mv local.list local.list.disabled mv cm.disabled cm.list mv cm-ml.disabled cm-ml.list mv /etc/apt/sources.disabled /etc/apt/sources.list

7. Upgrade Infiniband Switches

Check out the instructional <u>guide</u> on upgrading InfiniBand switches using the UFM appliance.

The ability to update the server software is applicable only for hosts (servers) with the UFM Agent.

To upgrade the software:

- Select a device.
- From the right-click menu, select Software Update.
- Enter the parameters listed in the following table.

Parameter	Description
Protocol	Update is performed via SCP protocol
lb	Enter the BCM head node bonded IP
Path	Enter the parent directory of the SCP directory structure for the Upgrade image. The path should not be an absolute path and should not contain the first slash (/) or trailer slash.
User	Name of the BCM head node root user
Password	Enter the host password

4

- Click Submit to save your changes.
- Right-click InfiniBand to "reboot" the switch.

8. SuperPOD Validation

After the installation is completed, using the bcm-validate-pod command, one can verify whether all the settings are applied correctly according to the specification.

```
root@vikinghead-01:~# module load bcm-post-install/
root@vikinghead-01:~# bcm-validate-pod --verbose
Checking slurm on head node:
Run: /cm/shared/apps/slurm/current/bin/sinfo --version
slurm 23.02.7
Passed
Checking ubuntu on head node:
Run: /usr/bin/lsb_release -a
Distributor ID: Ubuntu
                Ubuntu 22.04.2 LTS
Description:
Release:
                22.04
Codename:
                jammy
Passed
```

This module validates the completed pod setup with the expected configuration.

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Materials (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA'S aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, DGX, DGX NVIDIA SuperPOD, and NVIDIA Base Command Manager are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Authors

AY, DS, JB, HS, LH, MK, ML, RA, RS, and VS.

Copyright

© 2023 NVIDIA Corporation. All rights reserved.