CDW Documentation

nvidia_bcm-networking

Topic: Networking and Preparing for Installation

What This Unit Covers

* This unit is really about the planning work that has to happen before you install BCM on the head node. NVIDIA's installation and deployment documentation treats networking, licensing, ISO preparation, and site-survey data as foundational prerequisites rather than afterthoughts. In practice, that means you should know your topology, IP plan, DNS/NTP details, BMC design, and head-node licensing path before you ever start the installer. * NVIDIA's current documentation is centered on **Base Command Manager 11**, not 10. BCM 11 is the current major release in NVIDIA's documentation hub, and the BCM 11 release notes state that it includes the same core functionality as BCM 10 unless explicitly noted otherwise. ([NVIDIA Docs][1])

1. BCM Networking Overview

* In BCM, the network design is not just a cabling detail. It directly affects provisioning, PXE boot, BMC access, routing behavior, cluster isolation, and how the head node communicates with regular nodes. NVIDIA describes the internal cluster network as the path used for booting, data storage, and interprocess communication, while the head node is usually the only machine directly connected to the outside world. * At a basic BCM level, the most important logical networks are:

- an **internal network** for node management and provisioning
- an **external network** for outside access to the head node
- optional **BMC / IPMI networks** for out-of-band management
- optional high-speed fabrics such as **InfiniBand** or other dedicated compute/storage networks, depending on the cluster design.

* In current NVIDIA DGX BasePOD deployment docs, these logical networks are often represented as:

- **managementnet (internalnet)** for in-band management and provisioning
- **oobmanagementnet (ipminet)** for out-of-band BMC access
- **computenet (ibnet)** for the high-speed compute fabric
- **externalnet** for upstream connectivity to the customer network. ([NVIDIA Docs][2])

* One correction worth making: the InfiniBand fabric is not really "the network that connects the GPUs together" inside a node. More accurately, in NVIDIA cluster deployment docs it is the **high-speed inter-node compute fabric** used by systems and workloads across the cluster. GPU-to-GPU communication inside a server is a different concept from the external cluster fabric. ([NVIDIA Docs][3])

2. Why Network Planning Matters Before Installation

* NVIDIA explicitly recommends understanding the intended network design before installation. The installer asks you to choose a topology early, and that choice determines which predefined networks BCM creates later in the workflow. If you pick the wrong topology, you are not just labeling interfaces differently, you are shaping the routing and provisioning model of the cluster. * NVIDIA's deployment guides also stress that physical installation and switch configuration should be completed before BCM deployment, and that intended deployment details should be recorded in a site survey first. ([NVIDIA Docs][4])

3. Network Topologies

* BCM installation presents **three network topology choices**. Regular nodes are always located on

an internal network, called **Internalnet** by default. The topology you choose controls how the head node, regular nodes, and outside networks relate to each other.

Type 1 Topology

* Type 1 is the **default** setup and NVIDIA describes it as the **most common and simple** way to run a cluster. The nodes sit on a private internal network, and the head node routes traffic between that private network and the outside network, called **Externalnet** by default. * In this design, the head node provides **DHCP** and **PXE** services to a secondary isolated network for the worker nodes during pre-init boot. This isolates cluster traffic and keeps the external network focused mainly on access to the head node for administration. * The main tradeoff is that broader access to regular nodes from outside the cluster typically requires routing or proxying through the head node. * Because Type 1 defines both an **external network** and an **internal network**, it is the classic “head node as gateway” model.

Type 2 Topology

* In Type 2, regular nodes connect through a **router** to a public network. Traffic from a regular node to outside networks does **not** have to pass through the head node; instead, it goes out through the router. * At the same time, head-node-to-regular-node traffic normally still remains direct because the head node and regular nodes are usually on the same network in a standard Type 2 setup. DHCP and PXE traffic during pre-init boot also normally stays direct in that same-subnet arrangement. * Type 2 has **no Externalnet defined** in BCM’s predefined network list. Instead, BCM defines an internal network only. Routing beyond the router is handled on the router, not by the cluster itself. * NVIDIA also warns that you must avoid DHCP conflicts if the cluster is placed on an existing corporate network that already has a DHCP server. If regular nodes span several subnets, a **DHCP relay agent** may also be needed. * A useful conceptual note is that Type 2 does **not isolate** worker nodes the way Type 1 does. Nodes remain reachable through the main data plane, which NVIDIA notes can be useful for service-hosting use cases such as a web portal.

Type 3 Topology

* In Type 3, the head node and regular nodes are on **different routed networks**. Regular nodes are on **Internalnet** by default, while the head node is on **Managementnet** by default. * Because communication between the head node and the regular nodes is happening across Layer 3, DHCP’s normal Layer 2 behavior no longer works directly. NVIDIA therefore explains that **DHCP/PXE packets must be relayed**, typically by using a **DHCP relay agent** configured outside BCM by the network administrator or router vendor. * Type 3 is therefore the most networking-dependent of the three common choices. It gives more separation between management and node networks, but it also increases the importance of proper router and relay configuration.

4. Internal Network

* The internal network is the most important network in a BCM cluster because regular nodes are always placed on it. NVIDIA describes the internal cluster network as the one that connects all nodes to the head node and to each other, and compute nodes use it for **booting**, **data storage**, and **interprocess communication**. * In installation terms, the internal network is the default management/provisioning network for regular nodes, and the installer later uses its settings to determine how compute nodes are configured. For Type 1 and Type 2, this is the main cluster-facing network; for Type 3, the regular nodes are still on Internalnet even though the head node is on Managementnet. * NVIDIA’s compute-node interface screen also makes clear that the **BOOTIF** interface is the provisioning path used to pick up the node image. That is one of the clearest reasons

the internal network design matters so much.

5. External Network

* The external network is the connection from the cluster to the outside world and, in the typical BCM cluster model, it is usually the **head node** that is directly connected to it. Regular nodes are not normally directly attached to the external network in the classic cluster model. * On BCM installations using the default firewall model, the head node uses **Shorewall** for firewall and gateway functionality. The internal network is treated as the **nat** zone and the external-facing connection is treated as the **net** zone. * By default, Shorewall denies incoming traffic from the external zone except for explicitly allowed services. NVIDIA states that the cluster responds to **ICMP ping** by default and that these ports are open during installation unless the administrator changes them:

- **SSH**
- **HTTP**
- **HTTPS**
- **port 8081** for access to the cluster management daemon.

* Port **8081** is especially important because NVIDIA documents it as the default HTTPS port used by **CMDaemon** for node management. If needed, the `cm-cmd-ports` utility can move CMDaemon to another HTTPS port.

6. BMC / IPMI / Out-of-Band Management Network

* BCM supports management controllers such as **IPMI**, **iDRAC**, **iLO**, **CIMC**, and **Redfish v1**. These are part of the out-of-band management story and are configured in the installer through the BMC configuration screen. * If BMCs are used, BCM can configure BMC-related networking, and NVIDIA notes that a new Layer 3 subnet can be created for BMC interfaces. In deployment guides this out-of-band network is commonly called **oobmanagementnet (ipminet)**. * NVIDIA explicitly recommends a **dedicated physical BMC interface** where possible. A shared physical interface is supported, but the installation manual warns that it can cause issues during early BIOS checks. * Another useful detail: when BMCs are configured, BCM sets the BMC password to a random value by default for the configured nodes. * In DGX BasePOD network deployment guidance, BCM needs a link to the IPMI network so it can access node BMCs, either directly or indirectly through the customer network. ([NVIDIA Docs][2])

7. InfiniBand and High-Speed Fabrics

* BCM supports **NVIDIA InfiniBand HCAs and switches** as part of supported hardware, and NVIDIA cluster deployment documents commonly include a dedicated compute fabric such as **computenet (ibnet)** and, in some designs, a separate storage fabric. * In DGX BasePOD deployment material, the management/provisioning fabric and the compute fabric are treated as separate concerns. The compute fabric carries the high-speed cluster traffic, while managementnet/internalnet handles in-band management and provisioning. ([NVIDIA Docs][2]) * So for study purposes, it is better to think of InfiniBand as a **high-speed cluster interconnect** rather than just “a GPU network.” It supports workload communication across systems, not simply intra-node GPU connectivity. ([NVIDIA Docs][2])

8. Boot and Provisioning Networks

* NVIDIA’s installation manual makes two related points very clearly:

- regular nodes normally **network boot** from the head node
- the **BOOTIF** interface is the interface used to pick up the image for provisioning.

* In Type 1, the head node provides DHCP and PXE services to the isolated worker-node network during pre-init boot. In Type 2, this usually still works directly if the nodes and head node are on the same network. In Type 3, DHCP relay becomes necessary because the traffic is crossing Layer 3 boundaries. * Current NVIDIA Mission Control networking documentation also notes that for `internalnet`, node booting and management are enabled by default, which is consistent with BCM using that network for DHCP and category assignment during provisioning. ([NVIDIA Docs][5])

9. Network Configuration During Installation

* The installer eventually presents a **Networks configuration** screen. Which predefined networks appear depends on the topology and BMC choices made earlier. For Type 1, BCM defines **externalnet** and **internalnet**. For Type 2, BCM defines **internalnet** only. For Type 3, BCM defines **internalnet** and **managementnet**. * NVIDIA notes that network settings are validated when you move forward in the installer, but that validation is only a **sanity check**. Valid values can still be wrong for your environment, so it is wise to confirm them with your network specialist or against the site survey. * The general cluster settings and later network screens ask for details such as:

- cluster name
- administrator email
- time zone
- time servers
- nameservers
- search domains
- base IP addresses
- netmasks
- gateway values
- head-node and compute-node interface assignments.

* BCM also supports **IP offsets** on compute-node interfaces. NVIDIA explains that the offset changes where automatic addressing begins, which is useful when you want to reserve lower addresses in the subnet for gateways, VRRP, or other infrastructure.

10. High Availability Networking

* BCM supports a two-head-node **high availability** model, with an active and passive head node. NVIDIA’s cluster documentation and deployment checklists refer to HA status, manual failover, and propagation of the primary head-node settings to the secondary during setup. * In deployment materials, the site survey includes an HA virtual IP and a failover-network decision field, and HA setup also requires the appropriate head-node MAC information for licensing and failover configuration. ([NVIDIA Docs][6]) * NVIDIA’s BCM status checks for HA include items such as `mysql`, `ping`, and status communication between the head nodes, which reinforces the idea that HA is not only about shared storage or a second node existing, but also about correct inter-head-node communication and monitoring. ([NVIDIA Docs][7]) * One thing I would be careful about is treating the failover link description from older notes as universal. Current NVIDIA docs clearly show HA concepts, shared IP information, and failover checks, but the exact physical heartbeat-cabling recommendation can vary by deployment guide. ([NVIDIA Docs][6])

11. Minimal Hardware Requirements

* NVIDIA’s BCM 11 installation manual lists the **minimal** hardware requirements for a very small cluster of one head node and two regular compute nodes. Those minimums are:

- **Head node**

- x86-64 or ARMv8 CPU
- 4 GB RAM for x86
- 16 GB RAM for ARMv8
- 80 GB disk space
- 2 Gigabit Ethernet NICs for the common Type 1 topology
- DVD drive or USB drive
- **Compute nodes**

- x86-64 or ARMv8 CPU
- 1 GB RAM minimum
- at least 4 GB recommended for diskless nodes
- 1 Gigabit Ethernet NIC.

* NVIDIA also immediately warns that 4 GB on an x86 head node is only a technical minimum and that a standard bare-metal installation runs best with **at least 8 GB RAM**. So for study purposes, memorize the official minimums, but operationally understand that real AI/HPC systems will typically exceed them by a wide margin. * For larger clusters, the same manual points to stronger recommended specs and even suggests significantly higher head-node resources once the cluster grows into the thousands of nodes.

12. Supported Hardware and BMC-Related Preparation

* BCM 11 supports major Linux platforms such as Rocky Linux 8 and 9, SLES 15, and Ubuntu 22.04 and 24.04, and it runs on both **x86_64** and **arm64 / AArch64** architectures. * Supported management controllers include **IPMI 1.5/2.0**, **iDRAC**, **iLO**, **CIMC**, and **Redfish v1**. Supported InfiniBand hardware includes NVIDIA HCAs and switches. * In DGX BasePOD deployment guidance, NVIDIA also recommends validating that the primary head node sees at least **two Ethernet-mode interfaces** before continuing, and it advises installing the OS on redundant storage such as hardware or software RAID. ([NVIDIA Docs][8])

13. BCM Licensing

* The current NVIDIA documentation is a little more nuanced than the simple statement “BCM licensing is based on GPU count.” What NVIDIA clearly documents today is that BCM uses a **license file** activated by a **product key**, and incorrect license attributes can prevent the cluster from handling the intended number of **GPUs or nodes**. * BCM 11 can be evaluated with a **free license**, and NVIDIA’s free-license FAQ says that there is **no limit on how many nodes or servers** can be in the cluster under that program, but the free license is available for **up to eight accelerators per server/node/system** and does **not** include NVIDIA Enterprise support. ([NVIDIA Docs][1]) * NVIDIA’s installation manual describes product key types including:

- **evaluation product key**
- **subscription product key**
- legacy **hardware lifetime product key**.

* Evaluation licenses are temporary, and the installation manual says evaluation product keys are valid for up to **three months** unless extended. It also notes that evaluation ISO downloads include a temporary built-in license for a very small cluster trial. * The `request-license` workflow prompts for organization and site details such as country, state, locality, organization name, organizational unit, cluster name, and primary head-node MAC address. If HA is being used, the workflow also asks about the second head node. * The activated license is tied to the hardware it was issued for, which is why

MAC addresses matter in the process.

14. BCM Product Key and Download Workflow

* Current NVIDIA docs point to the **NVIDIA Licensing Portal** for generating the BCM product key from your entitlement, and then to the **Base Command Manager Download site** at `customer.brightcomputing.com` for downloading the ISO. The DGX resources page specifically says you will need details such as the product key, Linux version, and optionally a BCM version if not using the most current release. ([NVIDIA Docs][9]) * BCM 11 release notes also state that users can specify their desired Linux distribution on the ISO download page, and that the selected distribution is packaged in the ISO. During installation, BCM installs both that Linux distribution and BCM on the head node, and it also creates a default disk image derived from the same distribution for compute nodes. ([NVIDIA Docs][1]) * So, if your original notes say “latest version is version 10” or imply NGC is the main BCM ISO source, that is outdated relative to the current NVIDIA documentation. The current public NVIDIA docs point to **BCM 11** and the **Base Command Manager Download site** for the ISO workflow. ([NVIDIA Docs][1])

15. Downloading the BCM ISO

* NVIDIA deployment guides describe the practical ISO workflow like this:

- download the BCM ISO from the BCM download site
- verify the checksum
- burn it to DVD or write it to a bootable USB device
- alternatively mount it as virtual media through the appliance BMC virtual console. ([NVIDIA Docs][8])

* The installation manual also says that if using a bootable USB device, you should follow the `README.BRIGHTUSB` instructions inside the ISO and validate the copied image with an MD5 checksum, because corruption can cause subtle problems later. * After booting from the ISO, the correct installer menu entry is **Start Base Command Manager Graphical Installer**.

16. Site Survey

* The site survey is not just paperwork. NVIDIA deployment guides explicitly say that physical installation and switch configuration should be completed before BCM deployment, and that information about the intended deployment should be recorded in the site survey beforehand. ([NVIDIA Docs][4]) * NVIDIA’s sample site survey includes a lot more than just cluster name and node count. It includes items such as:

- country, state/province, locality
- organization name and organizational unit
- administrator email
- cluster name
- time zone
- HA virtual IP
- NFS server IP and shared paths
- network names
- base IP addresses
- netmasks
- gateways
- name servers

- search domains
- time servers
- hostnames
- BMC IPs
- MAC addresses
- node IP assignments. ([NVIDIA Docs][6])

* NVIDIA’s BasePOD installation guide repeatedly tells the administrator to populate installer values **according to the Site Survey**, including general settings, network definitions, head-node settings, interface selection, and license activation details. That tells you how important the site survey really is: it becomes the authoritative source of truth during installation. ([NVIDIA Docs][8])

17. Practical Installation-Prep Details Worth Remembering

* Before installation, NVIDIA recommends validating:

- hardware info on the head node
- presence of the required Ethernet interfaces
- storage layout
- DNS and NTP details
- node naming and counts
- BMC design
- network addresses and gateways
- whether HA is planned. ([NVIDIA Docs][8])

* The head node is the control point for the cluster and provides critical services such as user management, workload management, DNS, and DHCP. That is why mistakes in head-node planning tend to cascade into everything else. * A bare-metal head-node installation is the recommended path in the BCM installation manual because it avoids inherited issues from an existing OS configuration.

Key Takeaways

* BCM networking is foundational because it determines how nodes boot, how they are managed, how traffic is routed, and how isolated the cluster is from the outside world. * **Type 1** is the default and most common topology, with the head node acting as the gateway between internal and external networks. * **Type 2** exposes worker nodes more directly to the broader network and requires attention to DHCP conflicts. * **Type 3** separates head-node and worker-node networks and usually requires a **DHCP relay agent** because provisioning traffic crosses Layer 3 boundaries. * The **internal network** is the primary provisioning and management path for regular nodes, and the **BOOTIF** interface is used to obtain the node image during provisioning. * Out-of-band management is commonly implemented through **BMC/IPMI networks**, and a dedicated physical BMC interface is preferred where possible. * The current public NVIDIA docs are based on **BCM 11**, and the BCM ISO workflow uses the **NVIDIA Licensing Portal** plus the **Base Command Manager Download site**. ([NVIDIA Docs][1]) * The site survey is critical because it collects the exact data used throughout installation and HA setup. ([NVIDIA Docs][6])

[1]: https://docs.nvidia.com/base-command-manager/bcm-11-release-notes/overview.html “Base Command Manager 11 — Base Command Manager 11 Release Notes 1 documentation” [2]: https://docs.nvidia.com/dgx-basepod/deployment-guide-dgx-basepod/latest/network-deploy.html “Network Deployment — NVIDIA DGX BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems” [3]: https://docs.nvidia.com/dgx-basepod/deployment-guide-dgx-basepod/latest/index.html “NVIDIA DGX BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems — NVIDIA DGX

BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems” [4]: https://docs.nvidia.com/dgx-basepod/deployment-guide-dgx-basepod/latest/network-bcm-intro.html “BCM Introduction — NVIDIA DGX BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems” [5]: https://docs.nvidia.com/mission-control/docs/rack-bring-up-install/2.2.0/bcm-networking.html?utm_source=chatgpt.com “Manual BCM Networking Setup” [6]: https://docs.nvidia.com/dgx-basepod/deployment-guide-dgx-basepod/latest/site-survey.html “Site Survey — NVIDIA DGX BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems” [7]: https://docs.nvidia.com/mission-control/docs/rack-bring-up-install/2.0.0/deployment-summary-validation-checklist.html?utm_source=chatgpt.com “Deployment Summary Validation Checklist” [8]: https://docs.nvidia.com/dgx-basepod/deployment-guide-dgx-basepod/latest/bcm-deploy.html “BCM Headnodes Installation — NVIDIA DGX BasePOD: Deployment Guide Featuring NVIDIA DGX H200/H100 Systems” [9]: https://docs.nvidia.com/dgx-resources/index.html “DGX Resources”