

FlashBlade//EXA Site Planning Guide

50-0253-01

Copyright Statement

© 2025 Pure Storage® (“Pure”), Portworx® and its associated trademarks can be found [here](#) and its virtual patent marking program can be found [here](#). Third party names may be trademarks of their respective owners.

The Pure Storage products and programs described in this documentation are distributed under a license agreement restricting the use, copying, distribution, and decompilation/reverse engineering of the products. No part of this documentation may be reproduced in any form by any means without prior written authorization from Pure Storage, Inc. and its licensors, if any. Pure Storage may make improvements and/or changes in the Pure Storage products and/or the programs described in this documentation at any time without notice.

THIS DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. PURE STORAGE SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

Pure Storage, Inc. 2555 Augustine Drive, Santa Clara, CA 95054 USA <http://www.purestorage.com>

Direct comments to DocumentFeedback@purestorage.com.

Conventions and Guidelines (Conventions et lignes directrices)

Conventions

The following symbols appear in Pure Storage documentation.

Les directives suivantes s'appliquent aux produits Pure Storage.

Table 1.1 Documentation Symbols (Symboles de Documentation)

Symbol (Symbole)	Meaning (Signification)
	Additional information Instructions supplémentaires
	Supplementary instruction Meilleure pratique
	Potential for injury, physical damage, data loss, or array configuration loss Risque de blessure, de dommage physique, de perte de données ou de perte de configuration de la matrice
	Important information Renseignements importants
	Indicates a need for login access to a FlashBlade or involvement from Pure Storage Technical Services. Indique un besoin d'accès de connexion à un FlashBlade ou une implication de Pure Storage Technical Services.

-  If this equipment is used in a manner not specified by the manufacturer, the protection provided by the equipment might be impaired.
-  **Mise en GARDE:** Si cet équipement est utilisé d'une manière qui n'est pas précisée par le fabricant, la protection fournie par l'équipement pourrait être compromise.

-  Equipment is intended for installation in a restricted access area.
-  **Mise en GARDE:** Les matériels sont destinés à être installés dans des emplacements à accès restreint.
-  Verify that all electrical connections are grounded before powering on the system.
-  **Mise en GARDE:** Assurez-vous que toutes les connexions électriques sont mises à la terre avant de mettre le système sous tension.
-  Avoid contact with back-plane components and connectors.
-  **Mise en GARDE:** Évitez tout contact avec les composants et les connecteurs du fond de panier.
-  Disconnect power cords before moving chassis and shelves.
-  **Mise en GARDE:** Débranchez tous les cordons d'alimentation avant de déplacer le châssis ou les tablettes ou d'effectuer toute opération de mise hors tension.
-  Install replacement components within two minutes of removing a faulty component to ensure the maximum blade temperatures do not exceed temperature limits.
-  **Mise en GARDE:** Installez les composants de remplacement dans les deux minutes suivant la dépose des composants défectueux. Cela permet de s'assurer que les températures maximales ne sont pas dépassées.
-  Use only CDRH approved Class 1 fiber optic transceiver modules.
-  **Mise en GARDE:** Utilisez uniquement des modules émetteur-récepteur à fibre optique de classe 1 approuvés par le CDRH.
-  Verify all rack electrical distribution systems meet all safety requirements of UL 62368-1, IEC 62368-1, or EN 62368-1. For AC installation provide adequate 50-60Hz AC power, and use over-current protection devices connected to a UL certified (or equivalent 20A) circuit breaker for each

PSU. For DC installation, provide adequate -48-60V DC power, and use over-current protection devices connected to a UL certified 50A (or equivalent 50A) circuit breaker for each PSU.



Vérifiez que tous les systèmes de distribution électrique du rack répondent à toutes les exigences de sécurité des normes UL 62368-1, CEI 62368-1 ou EN 62368-1. Pour l'installation CA, fournissez une alimentation CA adéquate de , 50 à 60 Hz et utilisez des dispositifs de protection contre les surintensités connectés à un disjoncteur certifié UL (ou équivalent de 20 A) pour chaque bloc d'alimentation. Pour une installation CC, fournissez une alimentation CC adéquate de -48 à 60 V et utilisez des dispositifs de protection contre les surintensités connectés à un disjoncteur certifié UL de 50 A (ou l'équivalent de 50 A) pour chaque bloc d'alimentation.



- Do not use bifurcated ("Y") power cords.



Mise en GARDE: N'utilisez pas de cordons d'alimentation bifurqués ("Y").



- Do not open chassis, shelf enclosures, or other components unless you are a qualified service representative performing a procedure approved by Pure Storage.



Mise en GARDE: N'ouvrez pas le châssis, les boîtiers d'étagère ou tout autre composant à moins d'être un représentant qualifié qui effectue une procédure d'entretien conformément aux instructions publiées par Pure Storage.



- Do not use component handles to lift the chassis. Use the chassis lift handles to lift a chassis.



Mise en GARDE: N'utilisez pas de poignées des composants pour soulever un châssis. Utilisez les poignées de levage du châssis pour soulever un châssis.



- When servicing rack-mounted equipment, remove only one chassis at a time.



Mise en GARDE: Lors de l'entretien de l'équipement monté sur bâti, prolonger un seul châssis à la fois.



- Use a mechanical lift or or more people to lift the FlashBlade chassis.



Mise en GARDE: Utilisez un élévateur mécanique ou personnes ou plus pour soulever le châssis FlashBlade.



- Do not replace batteries due to the risk of exploding if mishandled.



Mise en GARDE: Ne remplacez pas les piles en raison d'un risque d'explosion si elles sont manipulées incorrectement.



- Do not insert anything into any FlashBlade or connectors except as specified in installation or repair instructions.



Mise en GARDE: N'insérez rien dans le connecteur FlashBlade or sauf dans les directives d'installation ou d'entretien.

Contacting Pure Storage Technical Services

Support is available 24 hours a day, 7 days a week. For assistance, contact Pure Storage Technical Services.

- Visit <https://www.purestorage.com/support>
- Email support@purestorage.com
- Call (800) 379-7873 (USA). For international support, visit <https://www.purestorage.com/support>.

Table of Contents

- Chapter 1 Overview..... 9**
- Chapter 2 Site Requirements..... 10**
- Chapter 3 Hardware Specifications..... 12**
 - FlashBlade//EXA and XFM-8400R2 Modules Specifications..... 12
 - Data Node, Switch, and Client Specifications..... 15
 - Supported Cables and Optical Transceivers..... 18
- Chapter 4 Network Architecture..... 19**
 - Layer 2 Network Configuration..... 20
 - Layer 3 Network Configuration..... 22
- Chapter 5 General and Management Network Information Requirements..... 26**
 - Data Network Requirements..... 28
 - Network Settings Worksheet..... 29
- Appendix A FlashBlade//EXA and Data Nodes VIP ECMP Routing..... 31**
- Appendix B RoCE Configuration..... 33**

Chapter 1

Overview

This guide outlines the necessary steps to prepare a customer site for the installation of a FlashBlade//EXA or an upgrade to an existing FlashBlade//EXA system. The site planning tasks are categorized into two distinct areas:

- Physical requirements:
 - Inspect physical location for sufficient clearance and rack space before installing system.
 - Identify any obstructions or layout issues.
 - Determine the required rack space and power requirements.
- General site planning:
 - Review the shipment manifest and report any missing items.
 - Assess the available space for unpacking shipment.
 - Review the customer network.
 - Return the replaced components to the company's headquarters.

Chapter 2

Site Requirements

Before installing a FlashBlade//EXA system, plan, conduct a site survey, and arrange for the safe lifting of equipment. Use a mechanical lift or 4 or more people to lift the chassis. Ensure the site adheres to the following requirements.

Restrictive Access

Ensure the installation and operational status of the FlashBlade//EXA storage system is in a restricted access location to mitigate potential disruptions and security breaches.

Rack Mounting Clearance and Support Requirements

- **Work space and clearance:** Verify the operating location provides sufficient clearance and spacing for installing a FlashBlade//EXA system. Verify that the front and rear mounting posts are between 61.0 cm (24 in.) and 91.5 cm (36 in.). Additionally, there should be 5 RU (22.3 cm/8.75 in.) of contiguous vertical space available for the chassis.
 - The recommended clearance for the front (cold aisle) is 91.5 cm (36 in.). The minimum is 81.5 cm (32 in.).
 - The recommended clearance for the rear (hot aisle) is 91.5 cm (36 in.). The minimum is 81.5 cm (32 in.).
 - No special clearance is required on the top, bottom, or sides of the rack.
 - The recommended front-to-rear clearance for rack posts is 74 cm (29 in. is default on most racks). Minimum is 61 cm (24 in.). Maximum is 81.25 cm (32 in.).
 - Verify the rack can support the weight of all installed equipment, including cables and rails for a fully loaded chassis—approximately 100 kg (220 lb)—and the weight is within the floor loading limit. Depending on the system configuration, the system can weigh more than 80 kg (176 lb).
 - The in-rack weight of a fully loaded FlashBlade//EXA chassis (includes 10 blades and 40 DirectFlash Modules) is approximately 100 kg (220 lb).
 - Each XFM-8400R2 module weighs approximately 14.22 kg (31.34 lb).

- **External Power Distribution Units (PDU):** Some racks have externally located PDUs. Ensure FlashBlade//EXA PSU cables lengths are sufficient and able to connect to the PDUs.

Electrical Power Requirements

Ensure the location provides sufficient electrical power. FlashBlade//EXA require 200-240 V AC, 50/60 Hz. For data node power consumption, refer to vendor technical specifications for details.

- Verify all rack electrical distribution systems meet all safety requirements of UL 62368-1, IEC 62368-1, or EN 62368-1. For AC installation, provide adequate 200-240V 50-60Hz AC power, and use over-current protection devices connected to a UL certified (or equivalent 20A) circuit breaker for each PSU.
- Verify power distribution units (PDU) are equipped with separate power banks and circuit breakers for redundancy when connecting FlashBlade//EXA systems.
- Confirm the electrical outlets in the building provides protective earth connection for connecting the FlashBlade//EXA system and preventing electrical shock. If a protective earth connection is unavailable, arrange for one to be installed.
- Recommend installing the FlashBlade//EXA chassis within 3 m (9.85 ft) of the network switches or patch panels.

Each FlashBlade//EXA chassis is equipped with four 2000W AC PSU with IEC C13/C14 power cords.

 **Note:** Power cords may vary depending on national regulations.

Environmental Requirements

- Ensure the operating temperature of the physical environment is within range of 5° to 35°C when the rack is fully populated.
- Verify airflow ducts are not aiming directly at the FlashBlade//EXA chassis.
- Install FlashBlade//EXA chassis with ports and power supplies facing in the direction of the hot aisle.
- Install XFM modules with ETH ports facing in the direction of the hot aisle.
- For data node operation, refer to vendor technical specifications for details.

Chapter 3

Hardware Specifications

This section describes the hardware specifications for FlashBlade//EXA and associated components.

FlashBlade//EXA and XFM-8400R2 Modules Specifications

FlashBlade//EXA Chassis Physical Dimensions and Weight

Figure 3.1 Chassis (Front View)

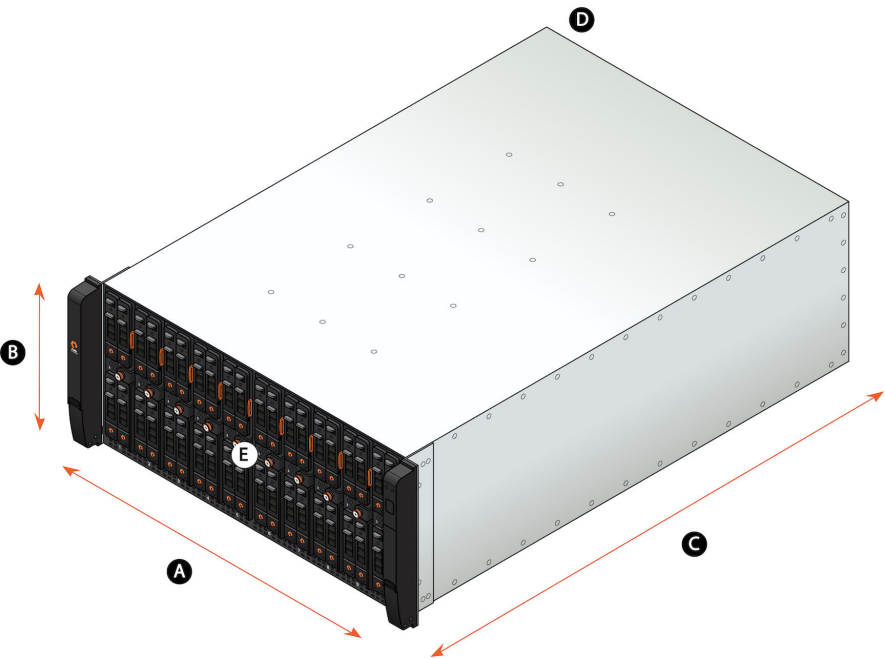


Table 3.1 Chassis Dimensions

Callout	Specification	Dimensions	Notes
A	Chassis width	44.3 cm (17.44 in.)	Not including rack ears.
	Overall width	48.2 cm (19.00 in.)	Including rack ears and bezel.

Table 3.1 Chassis Dimensions (continued)

Callout	Specification	Dimensions	Notes
B	Height	21.9 cm (8.62 in.)	5 RU of contiguous vertical space is required.
C	Chassis length	81.3 cm (32.0 in.)	From rack mount surface.
	Overall length	86.0 cm (34.0 in.)	Includes rack ears and bezel.
D	Rear depth allowance	3.2 cm (1.3 in.)	Clearance for PSU handle not including power cord.
E	Individual blade (without DFMs)	Length: 57.5 cm (22.7 in.) Width: 4.2 cm (1.65 in.) Height: 19.8 cm (7.80 in.)	Width value does not include EMC gaskets. Height value does not include EMC gaskets or ejector.
Not shown	Front bezel allowance	5 cm (2.0 in.)	Rack mount surface to the front bezel.
 Note: The direction of the airflow enters from the front and exits through the rear of the chassis.			

Table 3.2 Component Weight

Callout	Component	Quantity (per system)	Weight (per component)
A	Chassis		28.0 kg (61.8 lb)
B	Fabric I/O Module (FIOM-1000)	2 per chassis	5.8 kg (12.8 lb)
C	AC Power supply unit (PSU)	4 per chassis	1.0 kg (2.3 lb)
	Power supply module (Housing unit for PSU)	4 per chassis	0.6 kg (1.4 lb)
D	Blade (without DFMs)		4.6 kg (10.2 lb)
E	DirectFlash Module		0.2 kg (0.4 lb)

XFM-8400R2 Module Dimensions and Weight

Figure 3.2 XFM-8400R2 module

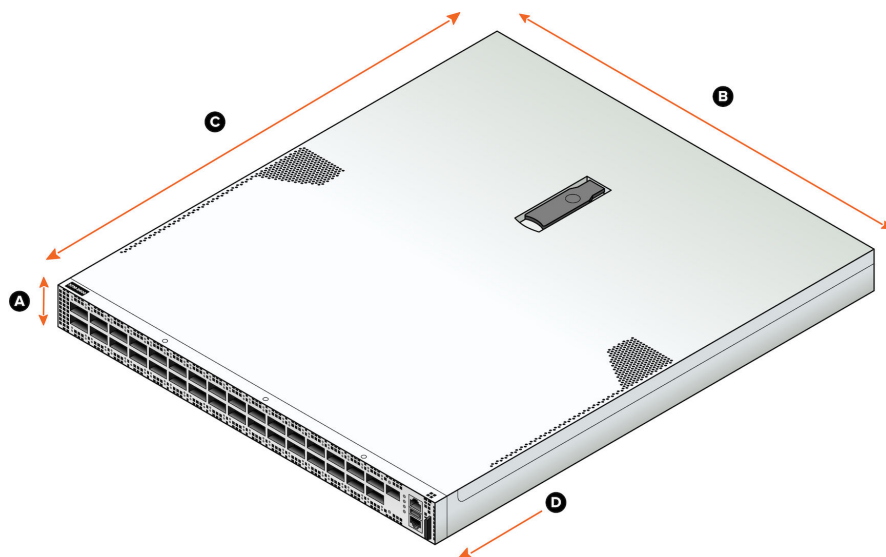


Table 3.3 Physical Dimensions

Callout	Specification	Dimension	Notes
A	Height	4.3 cm (1.70 in.)	
B	Width	43.9 cm (17.26 in.)	
C	Length	65.8 cm (25.89 in.) (Not including fan/PSU latches)	
D			Direction of airflow

Table 3.4 Weight

Component	Quantity	Weight
Chassis (module)	1	14.22 kg (31.34 lb)
Power supply unit (hot-swappable)	2	1.16 kg (2.56 lb)

Data Node, Switch, and Client Specifications

The following table lists the data node and client specifications.

Data Node

⚠ Important! By using your own third-party servers for hosting services, you acknowledge and agree that it is solely your responsibility to ensure the physical and cyber security of your network infrastructure. This includes, but is not limited to, securing access to your physical hardware, implementing appropriate firewalls, encryption, and other security measures to protect against unauthorized access, data breaches, and cyber threats. Pure Storage recommends you to always follow best practices for both physical and cyber security in order to maintain the integrity of your data and systems. Pure Storage is not liable for any damages, losses, or unauthorized access to your servers, data, or networks that may occur due to improper system maintenance or insufficient security measures.

Table 3.5 Data Node Specifications

Component	Description
Server model	All server model from variety of vendors must meet hardware requirements listed in this table.
Scalability	Minimum: 7 nodes Maximum: 500 nodes
Minimum CPU, DRAM	32 cores, 192GB DRAM
Capacity per data node	Minimum of 12 NVMe PCIe Gen4+ drives per node ⚠ Important! All drives installed in data nodes must be uniform (same vendor and capacity). Mixing drive vendors is not supported. 📝 Note: PCIe Gen5 drives recommended for best performance. The installed drives must be from same vendor.
Boot drive	M.2 NVMe PCIe 4.0 SSD
Boot mode	UEFI only
BIOS	(Optional) Enable PXE boot in BIOS for all data nodes

Table 3.5 Data Node Specifications (continued)

Component	Description
RAID configuration	RAID10 (performance) or RAID50 (capacity)  Important! Mixed RAID groups are not supported across data nodes within the same data node group.
Connectivity	2 x 400G Ethernet NICs (For best performance) Network Interface Card (NIC) - Minimum: 1 x 100G - Recommendation: 2 x 400G  Note: Network Interface Cards (NICs) must be RoCEv2 compatible. Pure Storage strongly recommends using NICs manufactured by NVIDIA. Cables - MPO-8 or MPO-12  Note: NICs and optics must transmit signals that are compatible (matching transmission signals). Optics 100G - 400G DR4 optics
Physical	- Minimum dimensions: Determined by data node - Drive form factor: Determined by data node - Power: Determined by data node
Software	DNOS (provided by Pure Storage)
Deployment and configuration solutions	One of the container engines is required during deployment: - Docker (Recommended) - Podman

Table 3.5 Data Node Specifications (continued)

Component	Description
Host server	Host running FB//EXA Services Container needs 10GB (for container) + 1GB per data node. For example, a 10-data node deployment will require 20GB [10GB + (1GB x 10 data nodes)] of storage.

Network Switch

Network switches deployed in a FlashBlade//EXA configuration may vary. Ensure the network switch meets the minimum requirements.

Table 3.6 Network Switch Specifications

Component	Description
Power requirements	Refer to vendor technical specifications for details.
Operating requirements	Refer to vendor technical specifications for details.
Connectivity	- Minimum: 100G uplink - Recommended: 400G uplink

Client

Table 3.7 Client Specifications

Component	Description
Software	Ubuntu 24.04
Kernel	6.8+ or newer (Recommended)  Note: If the kernel version is older than 6.14, Pure Storage provides a DKMS package for patching and enabling proper pNFS support.

Table 3.7 Client Specifications (continued)

Component	Description
Installation method	Use Ubuntu mainline kernel tools

Supported Cables and Optical Transceivers

FlashBlade//EXA system supports a variety of network cables and optical transceivers. For more information, visit support.purestorage.com.

Chapter 4

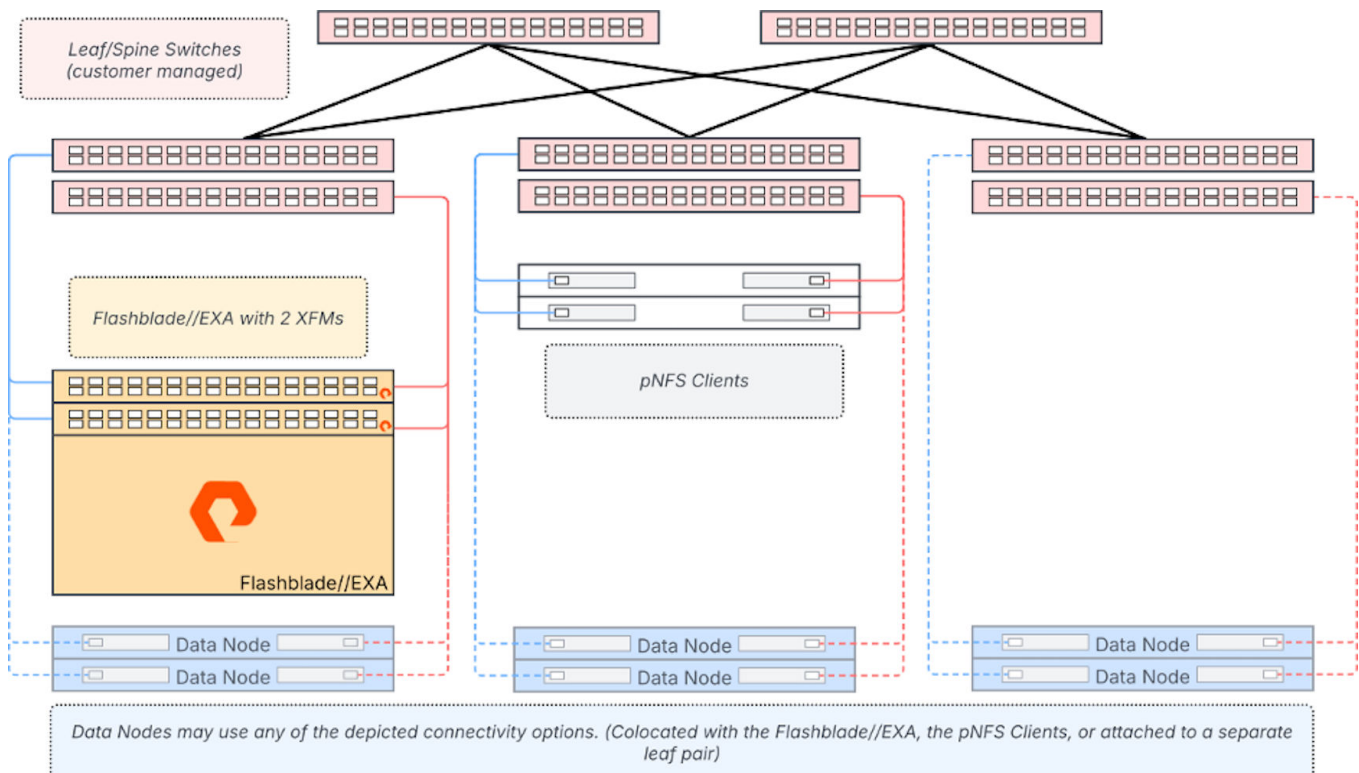
Network Architecture

This section outlines the configuration settings for high-speed configurations on FlashBlade//EXA, data nodes, and pNFS clients. FlashBlade//EXA and data nodes are additionally connected to a 1G management network, which is not depicted in the accompanying diagrams. BMC connectivity is also not explicitly shown.

FlashBlade//EXA Physical Connectivity

Optimum I/O performance may vary based on a customer's network environment. The diagram shows a typical deployment with FlashBlade//EXA system and generic data nodes. The following sections discuss two data node connectivity options and either option can be deployed for a specific network environment.

Figure 4.1 FlashBlade//EXA Connectivity



Layer 2 Network Configuration

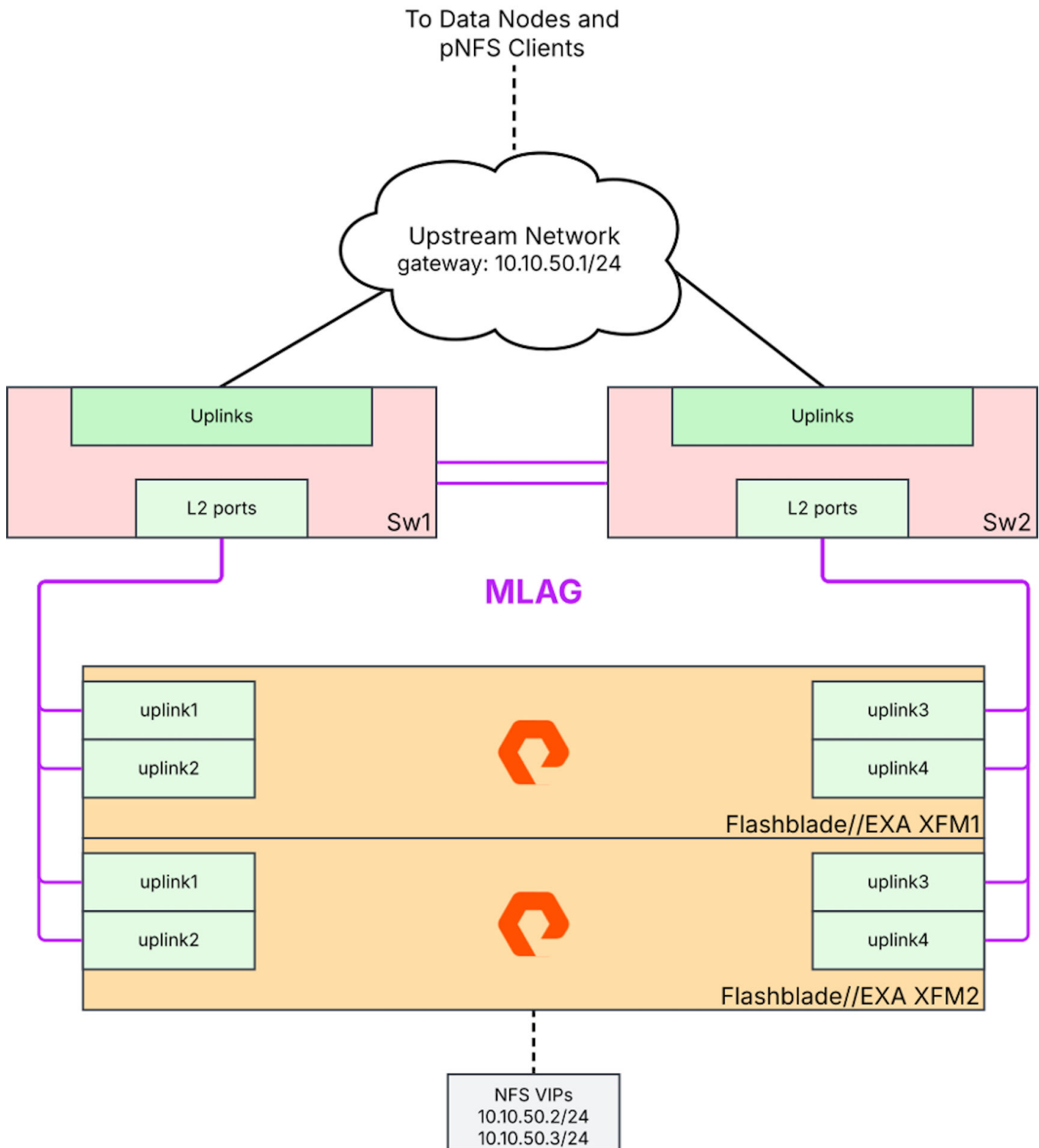
FlashBlade//EXA Connectivity

The Layer 2 network configuration ensures optimal performance and reliable connectivity for data transfer within a network segment. For physical connectivity, FlashBlade//EXA adopts a leaf/spine networking architecture to maximize bandwidth between data nodes and pNFS clients.

A standard L2 configuration for a FlashBlade//EXA storage system requires the following:

- FlashBlade//EXA Metadata Node and data nodes must be assigned to the same VLAN. The subnet assigned to data VIPs on the metadata node must not overlap with the subnet used by data path NICs on the data nodes.
- Data nodes (DN) and pNFS clients should each have two interfaces, or alternatively, clients can have two 2-port network interface cards (NICs).
- The uplink ports from the metadata core are configured as Multi-Chassis Link Aggregation (MLAG), using a redundant pair of switches.
- There must be at least two uplinks per server.
- A gateway for the subnet must be configured. The following illustrates a subnet gateway configured in an upstream network.
 - FlashBlade//EXA storage system subnet should not overlap the subnets used by data nodes.
 - The subnet used by FlashBlade//EXA and data nodes must be mutually reachable.
- Do not install the system away from the majority of clients on the network that are many hops away.
- Do not use the IP ranges 172.16.0.0/28 and 198.18.0.0/15 because they overlap with internal ranges. These ranges are used for internal communication and can cause the fabric modules to experience connectivity issues with other components.

Figure 4.2 FlashBlade//EXA Connectivity Diagram - Layer 2

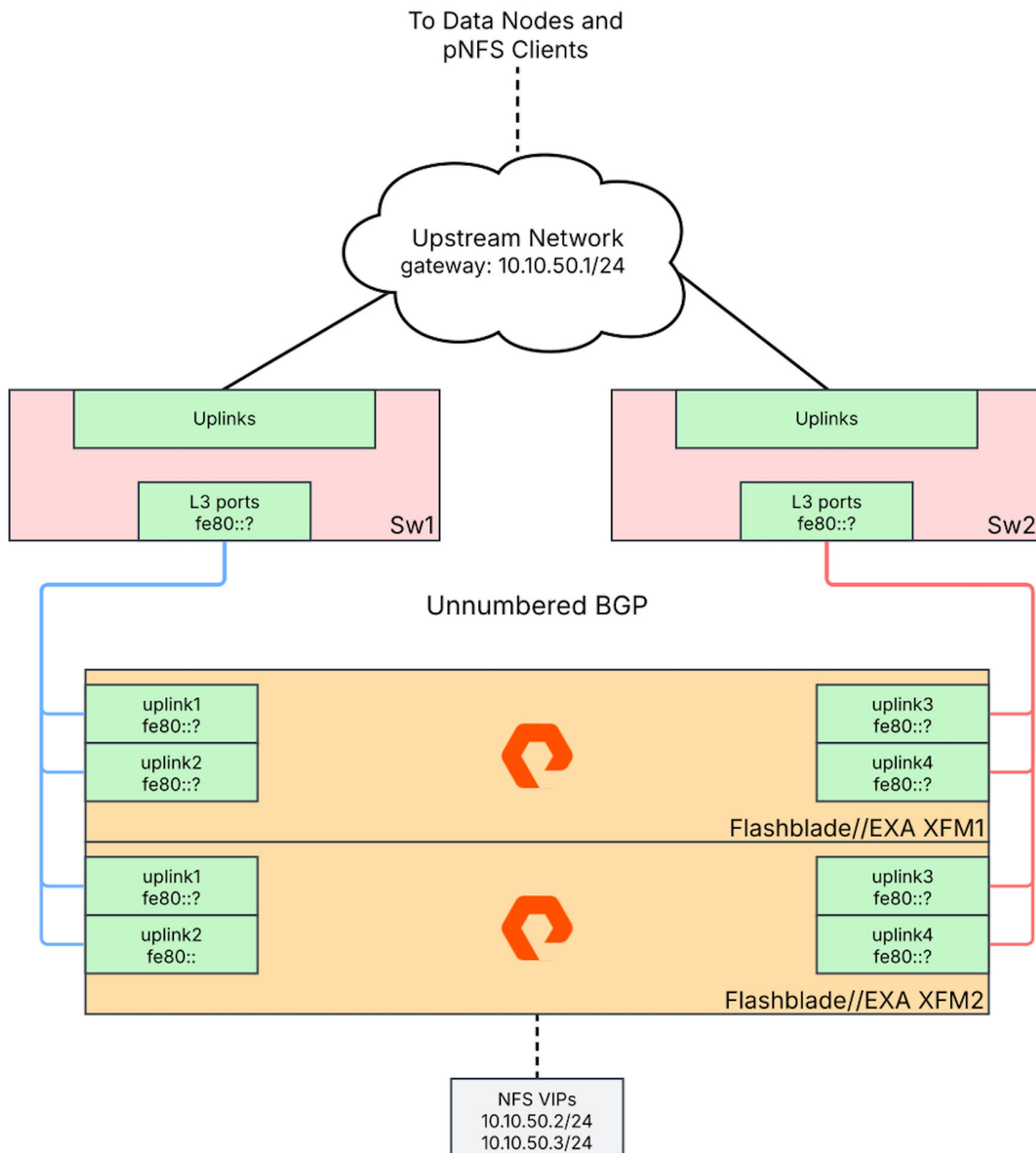


Layer 3 Network Configuration

FlashBlade//EXA Connectivity

FlashBlade//EXA connects to leaf switches with unnumbered BGP. Each XFM module must have at least two uplinks. Upstream switches should advertise the default route to the metadata core. The FlashBlade//EXA advertises its NFS VIPs to the switches.

Figure 4.3 FlashBlade//EXA Connectivity Diagram - Layer 3



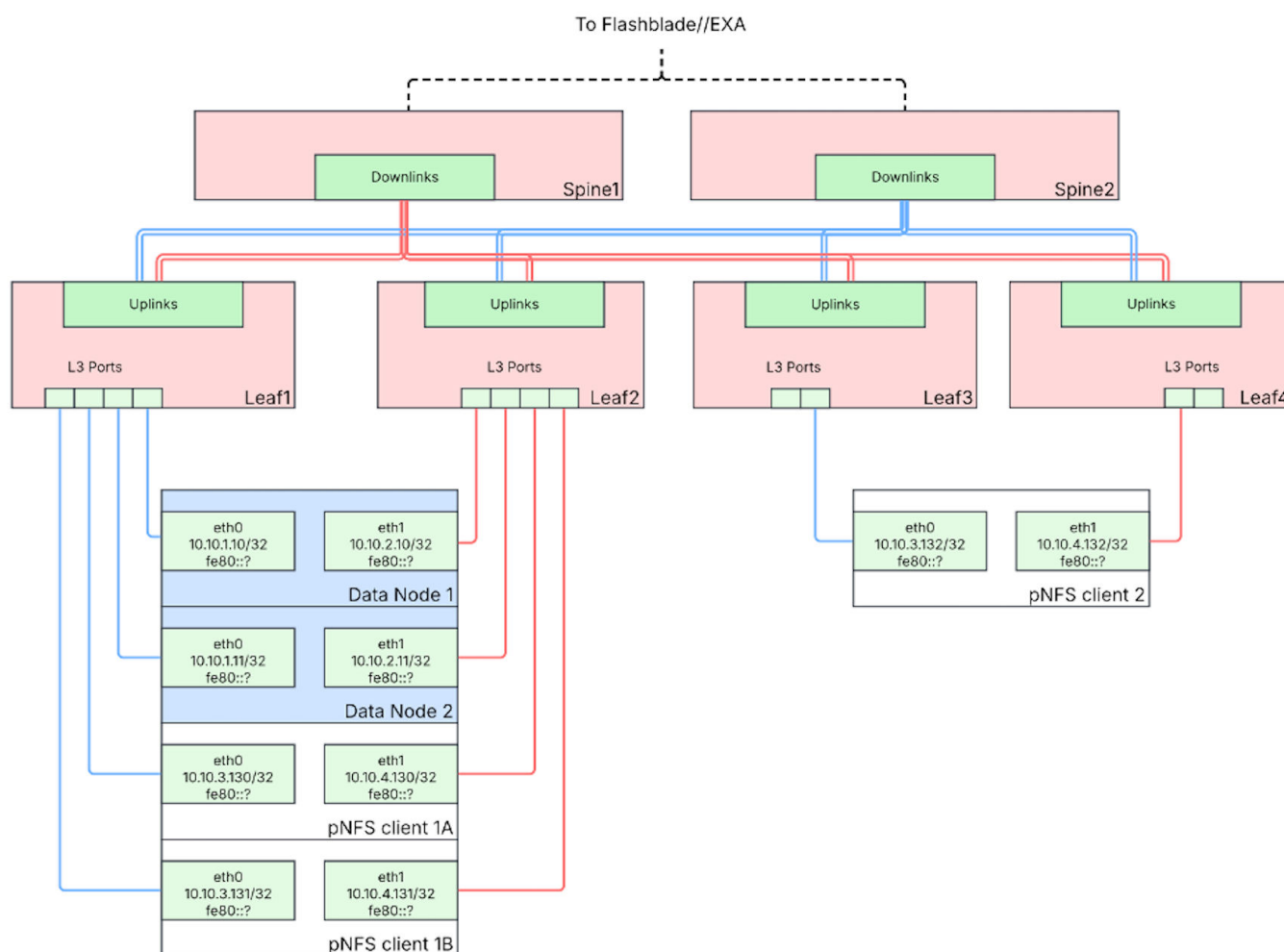
- Layer 3 network for data nodes (unnumbered BGP)

- Interfaces on data nodes use /32 IP addresses and receive routes from leaf switches via unnumbered BGP.
- IP Address ranges and subnets:
 - Leaf switch 1 subnet: 10.10.0.0/24
 - Leaf switch 2 subnet: 10.10.1.0/24
 - Layer 3 subnets for pNFS clients: 10.10.2.0/24 and 10.10.3.0/24

Data Node and pNFS Client Connectivity

A network topology using data nodes (DN) and various L3 pNFS clients is shown in the diagram.

Figure 4.4 Data Node and pNFS Clients Network Connectivity - Layer 3



- **Data nodes:** In the diagram, each data node has 2 NICs. Three /32 IP addresses for each Data Node are needed with two of them are assigned to each NIC port, and the remaining one is a VIP that is assigned to both NIC ports. Each port will have two IP addresses. Each port will also have its own auto-generated IPv6 link-local address. The IPv4 addresses are assigned from three different subnets. In this example 10.10.1.0/24, 10.10.2.0/24 and 10.10.5.0/24. pNFS clients will use the data VIP address to talk to the data nodes and the individual addresses are reserved for debugging and future uses.

The data nodes are BGP speakers and peer with Leaf1 and Leaf2 using unnumbered BGP. They advertise their /32 address and accept incoming advertisements only for the default route. Switches must be configured to advertise the default route to DNs for correct routing.

- **pNFS client 1A/1B:** These two clients are the same TOR switch as the data nodes and learn routes to the DNs from BGP.

In a typical BGP configuration, a client's eth0 will have a path to a data node's eth1 via the spine switch, but it will not prefer to use it because the AS_PATH is longer than using the client's eth1, which traverses only the leaf. Depending on bandwidth availability, users may want to block routes that require traversing the spine with route filtering on the switch or the client. If a client link fails, the pNFS client will use the still reachable addresses and not traverse the spine.

- **pNFS client 2:** This client is similar to 1A/1B but is connected in a way that it always traverse the spine layer.

In a typical BGP configuration, clients learn ECMP routes to all Data Node IP addresses on both eth0 and eth1. Depending on NIC capabilities, the ECMP routes may only compute path hashing using source and destination IP addresses, causing traffic to the DN to be imbalance and favoring one NIC port. To prevent this scenario, clients can filter to accept routes to half the Data Node IP addresses on eth0 and the other half on eth1. For instance, eth0 accepts routes to 10.10.1.0/24, while eth1 accepts routes to 10.10.2.0/24.

Chapter 5

General and Management Network Information Requirements

Opened Network Ports

Network connectivity between FlashBlade//EXA and data nodes require the following ports:

- 1G Management network: TCP 60066
- High-speed network: TCP 111, TCP 2049, TCP 54221
- High-speed network (between pNFS clients and data nodes): TCP 111, UDP 4791, TCP 54221

Installation Server (Jump Host)

An installation server is necessary for hosting Pure Storage installation container and requires the following settings:

- Access to 1G management network
- Access via remote SSH
- Staging capacity (Approximately 50GB)

Management Network Information

See the following table for required general and management network information. For additional details about network management requirements, contact Pure Storage Technical Services.

Table 5.1 General Settings

General Setting	Example
Hostname	test_hostname
Full Domain Name	mydomain.com

Table 5.1 General Settings (continued)

General Setting	Example
FlashBlade//EXA sender domain  Note: The <i>sender domain</i> is the domain name that you are going to use to send mails.	mydomain.com
Time Zone	Americas/Chicago
DNS Nameserver	192.0.2.10, 192.0.2.11
DNS Search Suffix	mydomain.com
NTP Servers	MyNTPServer1.com, MyNTPServer2.com
Watcher Email	aartiealert@example.com
Proxy (If required)	http://proxy.example.com:8080

Table 5.2 Management Network Settings

Management Network Setting	Example
Admin VIP	10.1.1.103
Admin IP 1 (FM1)	10.1.1.104
Admin IP 2 (FM2)	10.1.1.105

Table 5.3 Subnet for Management Network Settings

Subnet for Management Network Setting	Example
Subnet VLAN ID	111
Subnet Netmask	255.255.255.0 (or 24)
Gateway	10.1.1.1

Table 5.3 Subnet for Management Network Settings (continued)

Subnet for Management Network Setting	Example
Subnet Name ¹	Sub01
¹ Used in the UI to identify subnets.	

Data Network Requirements

See the following tables for the example of data network configuration and use the customer worksheet on the following page to record the customer's network information.

You can create data virtual IP addresses (data VIPs) after the initial setup of the FlashBlade. You can add data VIPs to existing (management) or a new subnet. If data VIPs are in a different subnet than the ones already configured, you must add a new subnet first.

Table 5.4 Data VIP Settings

Data VIP Setting	Example
IP	10.2.2.203
Service	NFS
Interface name	DataVip01

Table 5.5 Data Subnet Settings

Data Subnet Setting	Example
Subnet VLAN ID	222
Subnet Netmask	255.255.255.0 (or 24)
Gateway	10.2.2.1
Subnet name	Sub02

Network Settings Worksheet

Use this worksheet to enter your network information and use as reference during the setup process.

Table 5.6 General Settings

General Setting	Value
Hostname	
Full Domain Name	
Pure1 sender domain	
Time Zone	
DNS Nameserver	
DNS Search Suffix	
NTP Servers	
Watcher email	
Proxy (if required)	

Table 5.7 Management Network Settings

Management Network Setting	Value
Admin VIP	
Admin IP 1 (FM1)	
Admin IP 2 (FM2)	

Table 5.8 Subnet for Management Network Settings

Subnet for Management Network Setting	Value
Subnet Subnet VLAN ID	
Subnet Netmask	

Table 5.8 Subnet for Management Network Settings (continued)

Subnet for Management Network Setting	Value
Gateway	
Subnet Name	

Table 5.9 Data VIP Settings

Data VIP Setting	Value
IP	
Service	
Interface Name	

Table 5.10 Data Subnet Settings

Data Subnet Setting	Value
Subnet VLAN ID	
Subnet Netmask	
Gateway	
Subnet Name	

Appendix A

FlashBlade//EXA and Data Nodes VIP ECMP Routing

From the diagrams in [Network Architecture](#), multi-NIC pNFS clients will have multipath ECMP routes to theFlashBlade//EXA and data nodes.

An L3 client will learn these routes via BGP. For an L2/non-BGP client, these routes must be added manually, for example:

```
root@client $ ip route add 10.10.50.0/24 nexthop via 10.10.3.1 nexthop via 10.10.4.1
```

All clients should set:

```
net.ipv4.fib_multipath_hash_policy=1
net.ipv6.fib_multipath_hash_policy=1
```

This ensures TCP NFS connections to FlashBlade//EXA will balance across available egress ports.

However, in Linux, source IP selection occurs prior to nexthop selection. The above configuration alone will use all paths for egress, but always with the same source IP, and the return traffic will always land on a single interface, which may not be the interface used for egress. To ensure return traffic returns on the correct interface, it is necessary to set the source IP to match the egress interface. If a client has ETH0

with IP 10.10.3.10 and ETH1 with IP 10.10.4.10, this configuration results in correct return paths from both FlashBlade//EXA and data nodes:

```
root@client $ ip route show
...
10.10.50.0/24
    nexthop via 10.10.3.1 dev eth0 weight 1
    nexthop via 10.10.4.1 dev eth1 weight 1
...
10.10.5.0/24
    nexthop via 10.10.3.1 dev eth0 weight 1
    nexthop via 10.10.4.1 dev eth1 weight 1

root@client $ iptables -t nat -A POSTROUTING -o eth0 -j SNAT --to 10.10.3.10
root@client $ iptables -t nat -A POSTROUTING -o eth1 -j SNAT --to 10.10.4.10
root@client $ iptables -t nat -nvL
...
Chain POSTROUTING (policy ACCEPT 3 packets, 822 bytes)
  pkts bytes target     prot opt in     out     source
destination
    0      0 SNAT         0     --  *      eth0    0.0.0.0/0
10.10.50.0/24    to:10.10.3.10
    0      0 SNAT         0     --  *      eth1    0.0.0.0/0
10.10.50.0/24    to:10.10.4.10
```


Appendix B

RoCE Configuration

For network configurations employing RoCE protocols, lossless RoCE connectivity must be configured on pNFS clients and switches utilizing DSCP PFC. Additionally, ECN will require configuration on the switch. The exact configuration varies with switch and NIC vendor. For more information, reference the example [guide](#) from NVIDIA.

Data Nodes enable PFC for priority 3 using DSCP trust mode. DSCP 26 maps to priority 3.

For instance, an initiator with a single RDMA interface called “eth0” can be manually configured with appropriate PFC and DSCP parameters using the following commands:

```
# Enumerate RDMA interfaces
root@host $ rdma link
link mlx5_0/1 state ACTIVE physical_state LINK_UP netdev eth0

# Apply PFC settings
root@host $ modprobe rdma_cm
root@host $ INTF=eth0
root@host $ dcb pfc set dev $INTF prio-pfc 3:on
root@host $ dcb app add dev $INTF dscp-prio 26:3
root@host $ IBNAME=$(ls -l /sys/class/net/$INTF/device/infiniband)
root@host $ mkdir -p /sys/kernel/config/rdma_cm/$IBNAME/ports/1
root@host $ echo 106 > /sys/kernel/config/rdma_cm/$IBNAME/ports/1/
default_roce_tos
```

Ensuring Optimal NIC-to-Switch Connectivity and Compatibility: Optics, Form-Factor and Signaling

When connecting data node NICs to Top-of-Rack (TOR) switches, it is imperative to verify that both the transceivers form-factors (e.g., OSFP, QSFP-DD) and electrical signaling formats (e.g., 4 × 100G PAM4 versus 8 × 50G PAM4) are compatible. Inconsistencies in the signals can cause the links to auto-negotiate to lower speeds, resulting in undesirable bandwidth performance.

⚠ Important! Data nodes must be equipped with RoCE-compatible NICs for high-performance computing.

For instance, NVIDIA ConnectX-7 (CX7) NICs use OSFP ports and anticipate 4 × 100G PAM4 signaling (DR4). Meanwhile, many common TOR switches, such as ones manufactured by Arista or Edgecore, employ QSFP-DD ports that support 8 × 50G PAM4 signaling. If either end lacks DR4-capable, gearbox-enabled optics, the link will establish at 200Gbps instead of 400Gbps.

To ensure full 400Gbps connectivity, follow the recommendations:

- Use matching form-factor optics (e.g., OSFP112 on the NIC side, QSFP-DD on the switch side)
- Ensure both optics support 400G-DR4 signaling
- Employ MPO-8 or MPO-12 fiber compatible with DR4 transceivers



Pure Storage, Inc.

X: @purestorage

2555 Augustine Drive, 1st Floor

Santa Clara, CA 95054

800-379-7873

Sales: sales@purestorage.com

Support: support@purestorage.com

Media: pr@purestorage.com

General: info@purestorage.com