

N-Port Virtualization in the Data Center

What You Will Learn

N-Port virtualization is a feature that has growing importance in the data center. This document addresses the design requirements for designing a virtual environment to support growing storage needs. Three deployment scenarios are discussed—blade server deployments, top-of-rack fabric designs, and VMware—along with the advantages of N-Port virtualization in these situations.

Challenges

An emerging trend in data center design is server virtualization, or the use of virtual machine technology to prevent proliferation of physical servers. Each virtual server, to be managed as a unique entity on the storage area network, requires a separate address on the fabric. The N-Port virtualization feature supports this need for independent management of virtual machines and the increased use of aggregation devices in the data center.

Introduction

With the increased use of blade center deployments and top-of-rack aggregation devices in customer storage area network (SAN) environments, the deployment and use of aggregation switches is becoming more widespread. Because of the nature of Fibre Channel technology, several concerns need to be addressed when deploying large numbers of edge switches. One major concern when designing and building Fibre Channel–based SANs is the total number of switches or domains that can exist in a physical fabric. As the edge switch population grows, the number of domain IDs becomes a concern. The domain is the address of a physical switch or logical virtual fabric; the domain ID is the most significant byte in an endpoint Fibre Channel ID (Figure 1).

Figure 1.

Fibre Channel ID		
8 Bits	8 Bits	8 Bits
Domain	Port	Port

The switch uses this Fibre Channel ID to route frames from a given source (initiator) to any destination (target) in a SAN fabric. This 1 byte allows up to 256 possible addresses. Some domain addresses are used for well-known addresses, and others are reserved for future expansion. The Fibre Channel standard allows for a total of 239 port addresses; however, qualification of such a fabric size is nonexistent.

Another design concern is interoperability with third-party switches. In the past, different SAN fabric vendors interpreted the Fibre Channel addressing standard differently. In addition, some vendor-specific attributes used for switch-to-switch connectivity (or expansion port [E-Port] connectivity) made connection of switches from different vendors challenging, leading customers to implement edge switch technology that matched the core director type in the fabric.

Management of this complex system becomes a concern as smaller form-factor devices are used to aggregate multiple host connections. Typically, this complexity leads to shared responsibility between platform engineering or server operations and fabric operations. The delineation of management responsibilities is blurred.

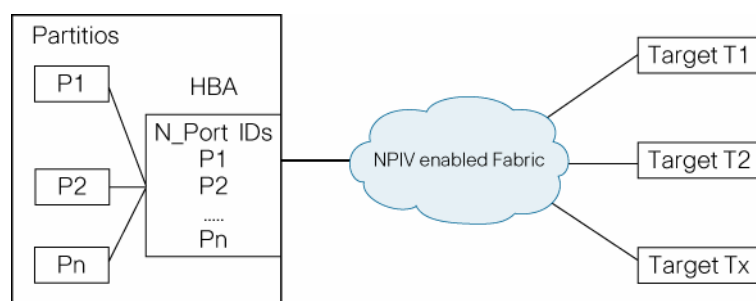
To address these concerns, two features, N-Port ID Virtualization (NPIV) and N-Port Virtualizer, were developed.

N-Port ID Virtualization

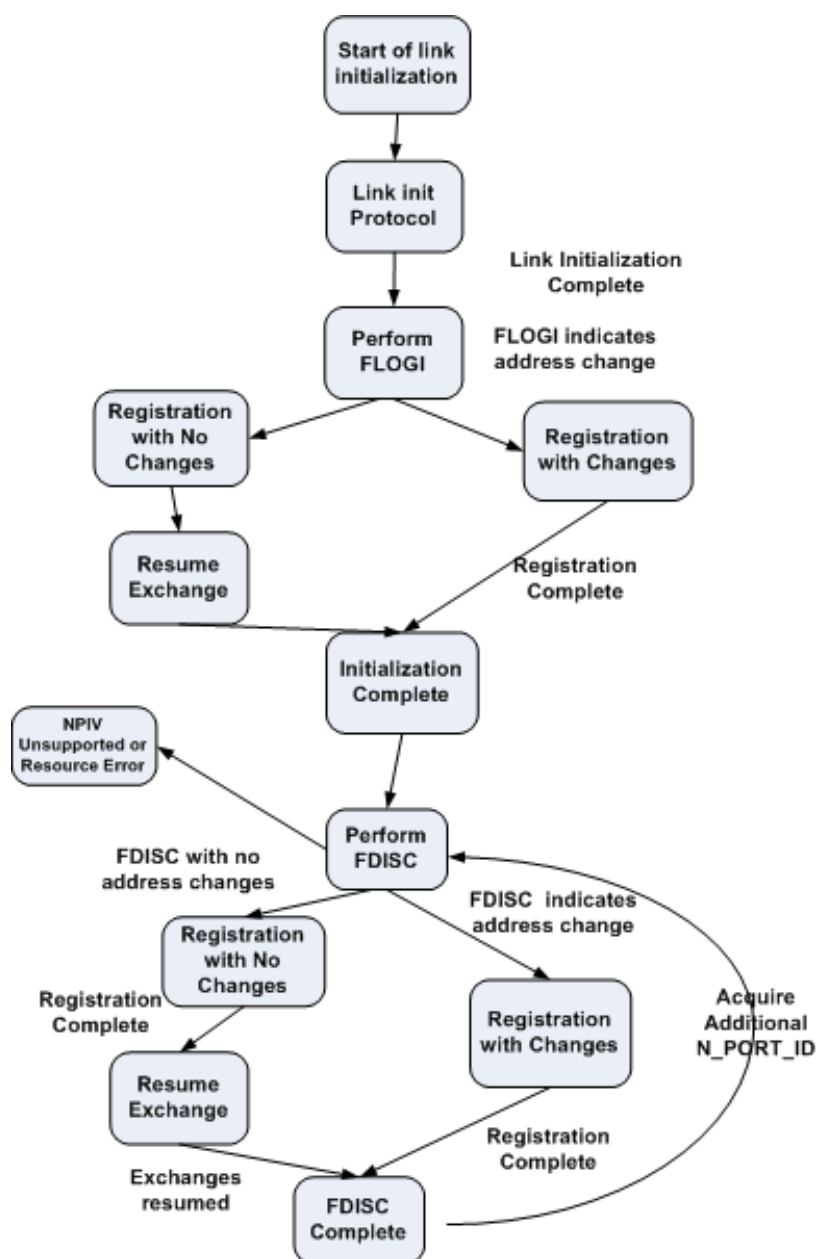
NPIV allows a Fibre Channel host connection or N-Port, to be assigned multiple N-Port IDs or Fibre Channel IDs (FCIDs) over a single link. All FCIDs assigned can now be managed on a Fibre Channel fabric as unique entities on the same physical host. Different applications can be used in conjunction with NPIV. In a virtual machine environment where many host operating systems or applications are running on a physical host, each virtual machine can now be managed independently from zoning, aliasing, and security perspectives.

Figure 2 shows an example of an NPIV-aware host connection.

Figure 2. NPIV-Aware Server Host Connection



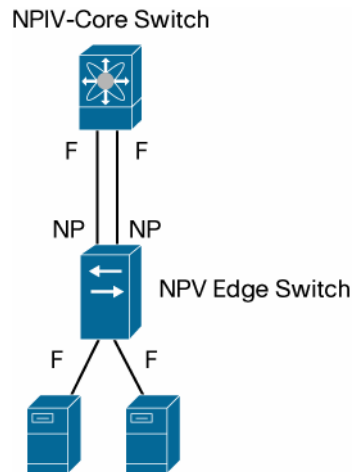
A host bus adapter (HBA) that supports the NPIV feature follows the standard login process. The initial connection and login to the fabric is performed through the standard F-Port login (FLOGI) process. All subsequent logins for either virtual machines or logical partitions on a mainframe are transformed into FDISC login commands. The FDISC logins follow the same standard process. Figure 3 steps through the login process of an NPIV uplink and the local logins to the NPIV-enabled adapter.

Figure 3. N-Port Virtualization Login Process

In a Cisco® MDS 9000 family environment, each host connection can log in as a single virtual SAN (VSAN). A VSAN is a logical partition of a larger group of physical ports that share a set of fabric services from a management domain. The VSAN architecture is analogous to VLAN deployments in Ethernet networks. When using NPIV in this environment, each subsequent FDISC login will be a part of the same VSAN as the original fabric login.

N-Port Virtualizer

An extension to NPIV is the N-Port Virtualizer feature. The N-Port Virtualizer feature allows the blade switch or top-of-rack fabric device to behave as an NPIV-based HBA to the core Fibre Channel director (Figure 4). The device aggregates the locally connected host ports or N-Ports into one or more uplinks (pseudo-interswitch links) to the core switches.

Figure 4. NPV enables an Edge Switch to Behave as an HBA to the Core Switch

The login process for the N-Port uplink (NP) is the same as for an HBA that is NPIV enabled; the only requirement of the core director is that it supports the NPIV feature. As end devices log into the NPV-enabled edge switches, the FCID addresses that are assigned use the domain of the core director. Because the connection is treated as an N-Port and not an E-Port to the core director, the edge switch shares the domain ID of the core switch as FCIDs are being allocated. The edge NPV-enabled switch no longer requires a separate domain ID to receive connectivity to the fabric. The domain byproduct of designing a SAN using top-of-rack or blade center technology is eliminated using NPV.

When an uplink port is enabled on an N-Port Virtualizer device, it must first perform a fabric login and process login to the core director and register with the Fibre Channel name server (see Figure 3). After this login process is completed for the physical uplink, the end devices can start their login processes. In the Cisco NPV process, an uplink port must first be logged into the core director for the end devices to start their login processes. As an end device initiates its login process into the edge NPV switch, it follows the standard login process with a FLOGI to the NPV edge switch. The edge switch will then convert the FLOGI login to an FDISC login and allow the login to proceed down the NPV uplink. Several NPV uplinks can be connected to either a single or multiple core directors. To date, the Cisco implementation allows connectivity from each edge switch to multiple core directors. The core NPIV directors must be part of the same fabric.

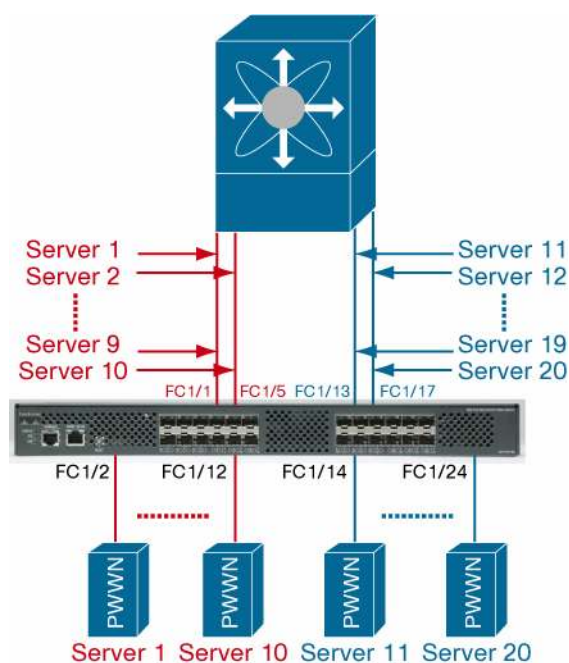
The NPV-enabled device performs proxy logins and also has a mechanism for sending FDISC logins down to the core NPIV director over multiple N-Port uplinks. There are three methods for sending logins down multiple links: a manual setup, a more dynamic load-balancing algorithm, and VSAN-based load balancing. Today, the Cisco implementation uses dynamic and VSAN-based load balancing.

In dynamic load balancing (all uplinks in the same VSAN), as hosts log in to the NPV device, the FDISC logins are sent down the links in a round-robin fashion. The I/O from the end device always follows the NPV path down to the same core switch to which its FDISC login was sent. In a stable fabric, each link should be utilized equally. For example, if there are eight hosts and two uplinks to core switches, then each link will have four initiators using each link. Should an uplink fail, the hosts that were logged in and using this uplink as an uplink to the core switch would be logged out of the fabric and go through the FDISC login procedure down the remaining link. In the initial release of the N-Port Virtualizer feature, there is no preferred uplink configuration that will dynamically send the host back down its original link. The host must be reinitialized for the host to

use the recovered path. This manual assignment of hosts to preferred uplinks will be implemented in a later release of firmware.

The NPV edge devices can be connected to a core Cisco MDS 9000 family switch in multiple VSANs. Each N-Port would follow the login procedure for that specific VSAN's Fibre Channel name server. The host ports on the switch are configured to match the VSAN that contains the uplink. All hosts FDISC logins are sent down the matching VSAN tag of the uplink port. If there are multiple uplinks in the same VSAN, the FDISCs will be round-robin load balanced down those links that match the VSAN ID (Figure 5).

Figure 5. NPV Allows Transparent Connectivity to the Core Switch



The N-Port Virtualizer feature allows transparent connectivity to any core switch that supports the NPIV feature. If the core switch follows the standard NPIV implementation, then the interoperability of different switch vendors is no longer a concern.

Future Enhancements

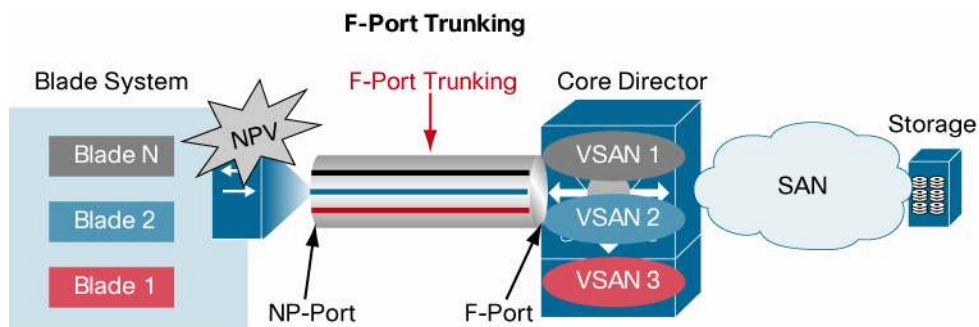
Two new features on the horizon will provide more granular management from an NPV perspective. These are Fabric Port (F-Port) Trunking and F-Port Channeling. A PortChannel is the bundling of multiple physical interfaces into one logical high-bandwidth link. PortChannels provide higher bandwidth, increased link redundancy, and load balancing between two switches. In the Cisco implementation, a port on any module, anywhere on the switch, can be a part of a PortChannel. Route table changes due to a single link failure in the PortChannel interface are unaffected.

The Cisco virtual fabric implementation states that a single F-Port can be a part of one, and only one, VSAN at any given time. An interswitch link (ISL) that is configured in trunk mode, however, can carry multiple VSANs across a single ISL or PortChannel and still keep fabric separation between VSANs. One of the drawbacks of configuring edge devices in NPV mode is that because the link between switches is configured to log in to the fabric as an F-Port, the F-Port uplink can be configured to be in only a single VSAN and carry only a single VSAN across its link. The NPV

device can be configured with multiple uplinks, and each can be placed in a different VSAN. The hosts on the NPV device can then be placed in one of those VSANs, and the hosts will log through the appropriate NP uplink port.

F-Port Trunking will allow a single F-Port to participate in multiple VSANs and follow the trunking protocol that Cisco currently has across an ISL. This feature will allow the consolidation of uplinks ports necessary for extending VSAN connectivity to the NP device. Figure 6 shows the connection and how the network will look with F-Port Trunking enabled.

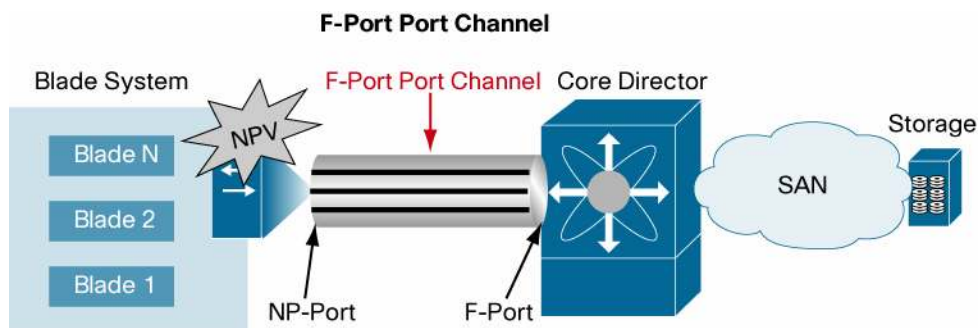
Figure 6. F-Port Trunking



Enabling F-Port Trunking allows any blade server connectivity over any NP uplink across which the VSAN is trunked. This feature can be extended to HBA connections when the HBA vendors enable VSAN trunking within their firmware. F-Port Trunking can enable virtual machines on physical servers to log in and participate in separate VSANs per application or line of business.

The other significant feature on the horizon is F-Port Channeling (Figure 7). As described earlier, when a host logs in to the local switch, its FDISC message is load balanced in a round-robin fashion. Should the link on which the host has its session fail, the host would have to log in again to the fabric, restarting the login process. N-Port uplinks from the NPV device cannot be bundled to take advantage of any type of high availability.

Figure 7. F-Port Channeling



F-Port Channeling will allow the same type of resiliency and availability between switches as an ISL PortChannel has today. In this configuration, multiple NP uplink ports can be bundled or channeled together to form a single logical link. Should a single link in the channel fail, the data frames crossing that single link would be lost, and typical application and network error recovery would take place.

The other concern that F-Port Channeling resolves also relates to link failures. With F-Port Channeling configured, when a link that originally carried the login process from the edge NPV

device to the core fails, the host is no longer required to perform a full login again to the fabric. The login state remains intact even though the link has failed. As long as a single link in the channel is operational, the host does not go through the login process.

In a F-Port Port-Channel logical link, if a single link fails in the channel, the host no longer has to login again to the fabric to get connectivity. The host will remain logged in; the data flow that was on the link that failed will need to go through recovery mode, but the host will stay logged into the network. Data traffic will also be load balanced on a src/dst hash or src/dst/oxid hash to send traffic down the bundled uplinks. This feature removes the need to perform manual load balancing again across NP uplinks during a failure and recovery period.

Conclusion

As customers change their data center architectures to further consolidate data center resources, network-based technology must continue to change to help design and manage the number of devices in the fabric. N-Port virtualization is a feature that has a growing importance in the data center. The three deployment scenarios discussed in this document—blade server deployments, top-of-rack fabric designs, and VMware—all take advantage of NPIV and NPV in different ways. As discussed here, Cisco continues to develop enhancements to NPIV to help customers design, deploy, and manage these fabrics as they grow and evolve.

For More Information

For more information about Cisco MDS 9000 family products, please visit

<http://www.cisco.com/go/storage>.

For a solutions overview, please visit <http://www.cisco.com/go/dc>.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

CCDE, CCVP, Cisco Eee, Cisco StadiumField, the Cisco logo, CDE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn is a service mark; and Access Registrar, Altranet, AnytimeOS, Bringing the Meeting to You, Catalyst, CCDA, CCDE, CCIE, CCIP, CCNA, CCNP, CCSP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IQS, iPhone, IP/TV, IQ Expertise, the IQ logo, IQ Net Readiness Scorecard, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, M3X, Netwerk24, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTlist, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. ©2007