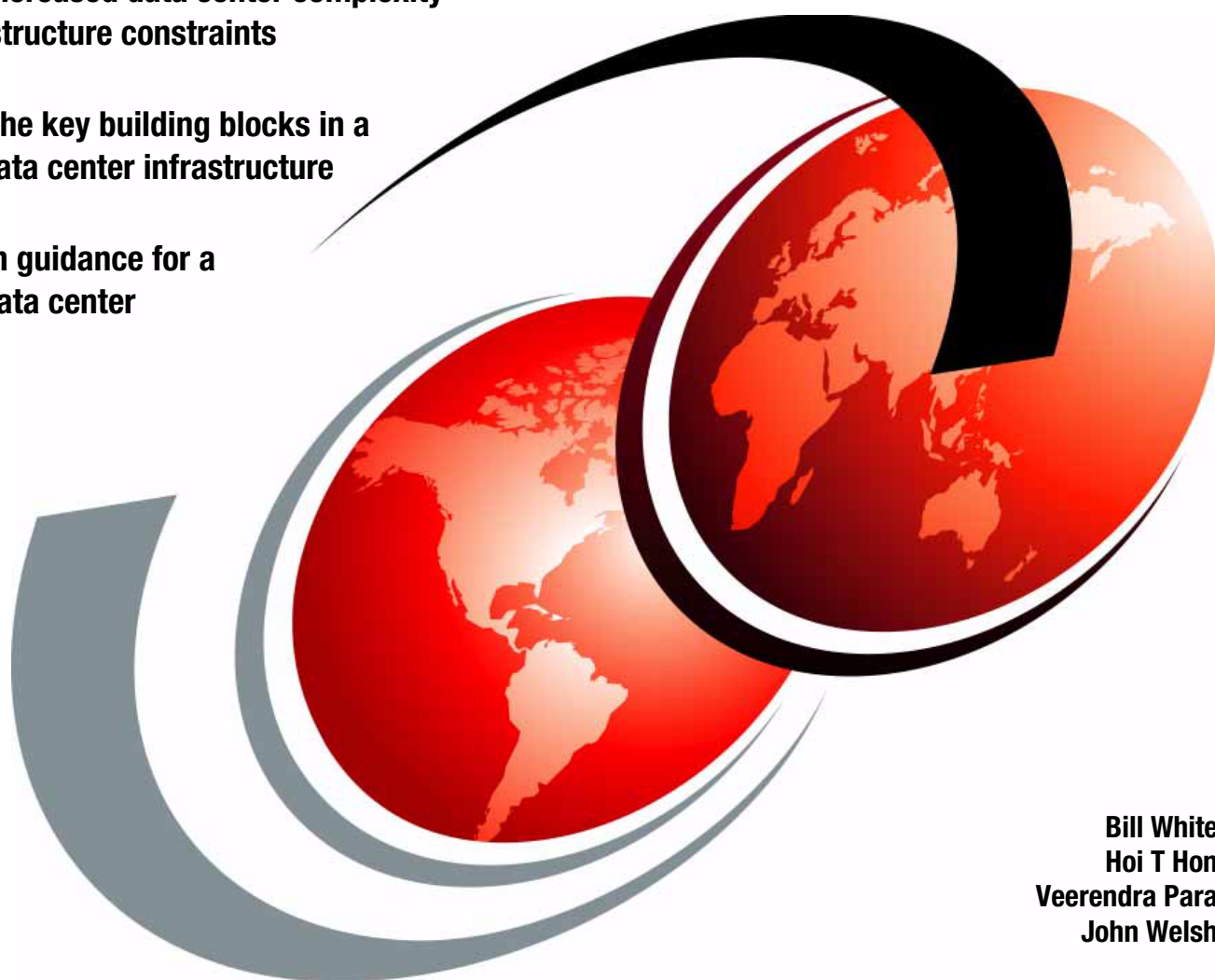


Building a Smarter Data Center with IBM Flex System and Juniper Networks QFabric

Address increased data center complexity
and infrastructure constraints

Discover the key building blocks in a
smarter data center infrastructure

Get design guidance for a
smarter data center



Bill White
Hoi T Hon
Veerendra Para
John Welsh



International Technical Support Organization

**Building a Smarter Data Center with IBM Flex System
and Juniper Networks QFabric**

March 2013

Note: Before using this information and the product it supports, read the information in “Notices” on page v.

First Edition (March 2013)

This edition applies to IBM Flex System and Juniper Networks QFabric.

© Copyright International Business Machines Corporation 2013. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices	v
Trademarks	vi
Preface	vii
The team who wrote this paper	viii
Now you can become a published author, too!	ix
Comments welcome	ix
Stay connected to IBM Redbooks	ix
Chapter 1. The value in creating a smarter data center infrastructure	1
1.1 Challenges for today's data centers	2
1.2 Attributes of a smarter data center infrastructure	3
1.3 Building a smarter data center infrastructure	4
Chapter 2. Understanding the key building blocks for a smarter data center.	7
2.1 Integrated components: The building blocks	8
2.2 Compute and storage domain: IBM Flex System	10
2.2.1 Enterprise chassis	12
2.2.2 Compute nodes	13
2.2.3 I/O modules	15
2.2.4 Expansion nodes	15
2.2.5 Virtual fabric networking	17
2.2.6 FCoE solution capabilities	20
2.2.7 VM-aware networking	20
2.2.8 Network management integration	22
2.2.9 IBM Flex System internal and external storage	24
2.3 High-speed fabric domain: Juniper Networks QFabric	25
2.3.1 QFabric data plane components	27
2.3.2 QFabric management plane component	30
2.3.3 QFabric control plane components	31
2.4 WAN domain: MX Series platform	32
2.5 Security domain: Security services gateways	33
2.5.1 Juniper Networks SRX Series Services Gateway	33
2.5.2 Juniper Networks vGW Series Virtual Gateway	35
2.6 Management domain: Integrated tools	38
2.6.1 IBM Flex System Manager	39
2.6.2 Junos Space	42
2.6.3 IBM Tivoli	45
Chapter 3. Designing a smarter data center infrastructure	47
3.1 Virtualization	48
3.1.1 Hypervisor-based virtualized networking	48
3.1.2 QFabric node groups	53
3.1.3 Providing an integrated solution with IBM Flex System BTO and QFabric	56
3.2 Security	57
3.2.1 Stateful firewall, intrusion detection, and antivirus	58
3.2.2 Integrating SRX Series and vGW to centralize policy management	59
3.2.3 Secure remote access	60

3.3 Provisioning.	61
3.3.1 Physical-level provisioning: Physical QFX Series switches	62
3.3.2 Physical-level provisioning: IBM Flex System BTO and QFX3500.	62
3.3.3 System-level provisioning: IBM Flex System virtual elements	66
3.3.4 System-level provisioning: QFabric logical components.	67
3.4 Management	68
3.4.1 Physical-level planning: Junos Space management network.	68
3.4.2 Physical-level planning: IBM Flex System Manager management network	70
3.5 Automation	71
3.5.1 End-to-end automation tools.	71
3.5.2 Integrated automation: IBM Flex System and QFabric.	72
 Chapter 4. Verifying key client use cases with IBM Flex System BTO and Juniper Networks QFabric	 75
4.1 Business continuity and disaster recovery	76
4.1.1 Description	76
4.1.2 Requirements	77
4.1.3 Proof of technology	78
4.2 Multitenancy	83
4.2.1 Description	83
4.2.2 Requirements	84
4.2.3 Proof of technology.	85
4.3 Virtual machine mobility	89
4.3.1 Description	90
4.3.2 Requirements	91
4.3.3 Proof of technology.	91
 Related publications	 97
IBM Redbooks	97
Online resources	97
Help from IBM	98

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.


COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®	NMotion®	Real-time Compression™
BigInsights™	POWER®	Redbooks®
DB2®	Power Systems™	Redguide™
Easy Tier®	POWER7®	Redpaper™
FlashCopy®	POWER7+™	Redbooks (logo)  ®
Global Technology Services®	PowerVM®	ServerProven®
IBM®	PureApplication™	Storwize®
IBM Flex System™	PureData™	System i®
IBM Flex System Manager™	PureFlex™	System Storage®
IBM PureData™	pureScale®	System x®
InfoSphere®	PureSystems™	Tivoli®
Netcool®	Rational®	VMready®

The following terms are trademarks of other companies:

Netezza, and N logo are trademarks or registered trademarks of IBM International Group B.V., an IBM Company.

QRadar, and the Q1 logo are trademarks or registered trademarks of Q1 Labs, an IBM Company.

Intel, Intel Xeon, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Preface

Data centers must become smarter to meet today's business needs. They need to be more efficient, scalable, and flexible and at the same time keep operational costs in check. A *smarter data center* must seamlessly integrate IT resources, such as servers, storage, and networking, while also responding quickly to change.

Networking plays an essential role in enabling infrastructures for smarter data centers. In dynamic environments with virtualized IT resources, the network must do more than just carry traffic and support the provisioning of new IT services. It must also have the built-in flexibility and capability to adapt quickly while maintaining comprehensive security, visibility, and management. Juniper Networks QFabric offers such an innovative network technology.

In addition, many IT infrastructures cannot handle rapid service delivery and efficient business resilience. Over time, deferred upgrades and extended technology refresh cycles mean that companies are dealing with an aging infrastructure or product obsolescence. Thus, constraints for handling workload mobility within a single data center or across data centers where widespread virtualized environment are complex. IBM® Flex System is up for these challenges with its servers, storage, networking, and virtualization in an integrated stack. The flexible design of the Flex System can meet the needs of varying workloads with independently scalable IT resource pools for higher utilization and lower cost per workload. And the integrated, easy-to-use, management system reduces setup time and complexity, providing a quicker path to return on investment.

IBM Flex System™ build-to-order (BTO) and Juniper Networks QFabric are key building blocks for a smarter data center infrastructure. They are reliable, resilient, and energy efficient resources that seamlessly integrate to provide the capabilities and flexibility needed now and in the future.

This IBM Redpaper™ publication discusses how to build a smarter data center infrastructure with IBM Flex System BTO and Juniper Networks QFabric. It begins with an explanation of how the data center infrastructure has evolved over time and what is needed to remedy the escalating data center challenges in resource utilization, performance, availability, security, provisioning, operational efficiency, and management.

This paper also describes IBM Flex System and Juniper Networks QFabric architectures and associated product portfolio capabilities. And finally, it discusses key client use cases that address today's data center challenges:

- ▶ Business continuity and disaster recovery
- ▶ Multitenancy
- ▶ Virtual machine mobility

This paper is intended for IT management, IT architects, network planners and integrators, and technical specialists.

The team who wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.

Bill White is a Project Leader and Senior Networking and Connectivity Specialist at the International Technical Support Organization, Poughkeepsie Center.

Hoi T Hon is an Infrastructure Architect in the IBM Global Services, US ICS/GTS division, and a member of the Data Center Networking Center of Excellence. He specializes in business consulting, network consulting, optical networking, data center networking, and IBM solutions, using multivendor technologies and products. He has 18 years of experience in the computer networking industry and holds Juniper Networks and many other advanced industry certifications. Hoi has a Bachelor of Engineering degree in Electrical Engineering from CCNY, CUNY. He has led several teams in developing large and complex data center networks, and optical networking and security solutions for fortune 500 companies and large government agencies. Hoi has also written many internal IBM networking reference architecture papers, optical networking technique papers, case studies, and co-authored other IBM Redguide™ publications.

Veerendra Para is a Senior Advisory IT Specialist at the IBM India Software Lab (ISL) in Bangalore, India. He has 12 years of experience in IT Industry and has been directly involved with data center infrastructures for last four years. His current job role is to transform ISL traditional server rooms into energy-efficient smarter data centers, introducing innovative green solutions, strategies, architectures, and timely executing assessment of relevant efficiency elements with reference to industry standard models. Veerendra also shares the success stories with IBM clients through briefings, demonstrations, and education sessions. He has worked in the systems management field, mainly on IBM System i®, POWER® Systems and AIX®. Prior to the IBM Software Group, Veerendra worked for IBM Global Technology Services®. He also has published technical disclosures.

John Welsh is an IBM Certified Architect in IBM Global Technology Services, Strategic Outsourcing Delivery in the Australia/New Zealand (A/NZ) region. He is a member of IBM Technical Experts Council (TEC) A/NZ. John has 16 years of experience in systems, storage, and networking technologies. He started his career at Digital Equipment Corporation in 1997, which eventually became part of Hewlett-Packard, working there for 10 years. John has been with IBM for five years, developing multivendor architectures for Cloud Automation, Systems Management, and large scale virtualized hosting platforms, on both x86 and POWER Systems.

Thanks to the following people for their contributions to this project:

Ella Buslovich, Diane Sherman, Debbie Willmschen
International Technical Support Organization, Poughkeepsie Center

Diep Le, Mark Lewis, Joan Ostojic
IBM

Greg Bassett, Sean Capshaw, Vaishali Ghiya, Lou Owayni, Jeremy Wallace
Juniper Networks

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks® publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

ibm.com/redbooks

- Send your comments in an email to:

redbooks@us.ibm.com

- Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>



The value in creating a smarter data center infrastructure

To respond to the pressures of change in today's business environment, organizations need more flexible and cost-effective IT infrastructures. Many are taking a new approach by consolidating and virtualizing IT resources, such as servers, storage, networks, and even desktops. As the data center evolves, however, data center managers face new challenges. They need to manage and control virtualization sprawl and ensure the security of the increasing number of virtual machines. In addition, they need to address technology constraints while migrating workloads and determine how to manage a network infrastructure that cannot scale with the explosion of network traffic brought about by the consolidation of workloads in virtualized environments.

This chapter presents common data center challenges for IT today and the importance of planning for a smarter data center infrastructure to satisfy future business needs.

1.1 Challenges for today's data centers

Many IT infrastructures were not built to support the explosive growth in computing capacity and network bandwidth use today. Data centers are often highly distributed and somewhat fragmented. As a result, those data centers are limited in their ability to change quickly. They cannot support the integration of new technologies or scale easily to power the business as needed. Data center planning can no longer focus on buying more servers and larger network switches to respond to short-term performance issues. A new data center model that is more efficient, service-oriented, and responsive to business needs is required. The model must also offer improved levels of economy, rapid service delivery, data security, business resilience, and tighter alignment with business goals and technology requirements.

In addition, organizations need to build solutions that are customized to their own workloads. Many workloads require high performance and low latency, which can be difficult to ensure as bandwidth needs change and create shifting I/O bottlenecks. Therefore, scalability is critical, yet the cost of adding capacity can be prohibitive. So, expanding the use of virtualization and accelerating the transition to cloud computing is essential for some businesses.

Virtualization technologies combined with cloud solutions are emerging as a consumption and delivery model for IT solutions. However, the move toward cloud computing, with a service-based acquisition and delivery model, requires data center managers to take a holistic view of the resources that they manage and the actors who access various resources in the data center. Other associated challenges are efficient methods for provisioning and deprovisioning resources and no clear path to optimize or upgrade high-value, non-infrastructure workloads, such as enterprise resource planning (ERP), customer relationship management (CRM), data warehousing, or analytics. Thus, managing cloud resources such as servers, storage, and networking from a single management node can be difficult.

These trends are a driving force in rapid IT consolidation. However, implementing a data center migration can be a complicated resource and time intensive activity that requires an expertise that IT organizations might lack.

Enterprises today face the following common data center challenges:

- System performance, scalability, and flexibility

Many of today's IT infrastructures cannot handle rapid service delivery and efficient business resilience. Over time, deferred upgrades and extended technology refresh cycles mean that companies are dealing with an aging infrastructure or product obsolescence. Thus, constraints for handling workload mobility within a single data center or across data centers in a widespread virtualized environment are complex. In addition, simplified methods for high availability and disaster recovery are lacking.

- Data growth

The explosion in the growth of data means increased costs relative to hardware, software, associated maintenance, administration, and services in order to keep up with storing and managing that data.

- Network congestion and connectivity architecture

To remain competitive, companies require bandwidth-intensive multicast workloads. These types of workloads require more capacity to handle the increased network traffic, which in turn requires resolution of existing network availability, performance, or management issues. Any existing inefficiencies in connectivity make the issue more complex.

All these efforts and challenges define a move from a static design toward a dynamic and scalable design in a *smarter data center*.

1.2 Attributes of a smarter data center infrastructure

Data centers must become smarter, more efficient, scalable, and flexible to meet business needs, while keeping operational costs in check. A smarter data center has the following critical attributes:

- ▶ Seamlessly integrates between different IT resources and provides flexibility and scalability
- ▶ Responds quickly to change by transforming insights into action
- ▶ Extends the life of the existing data center infrastructure and doubles IT capacity
- ▶ Rationalizes the data center infrastructure and improves operational efficiencies while reducing operational costs

Table 1-1 lists the smarter data center attributes that can mitigate the challenges and meet the business needs.

Table 1-1 Smarter data center attributes that address the business needs

Business need	Smarter data center attributes
Deployment	<ul style="list-style-type: none"> ▶ Faster setup time and quicker delivery ▶ Increased workload density and availability in reduced floor space
Scalability	<ul style="list-style-type: none"> ▶ Easily extensible up to superscale size data center without much intervention ▶ Seamless integration with existing data center resources
Simplicity	<ul style="list-style-type: none"> ▶ Integrated management with better control and automated provisioning ▶ Single point for resource pool management
Security	<ul style="list-style-type: none"> ▶ Hardware or software failures do not compromise the entire system solution ▶ Support for regulatory requirements that are associated with security ▶ Information is accessed only by authorized users
Resilience	<ul style="list-style-type: none"> ▶ Built-in redundancy for highly reliable operation ▶ Integrated with risk assessment ▶ Rapidly adaption and response to risks and opportunities to maintain continuous business operations ▶ Support for regulatory requirements that are associated with business resiliency ▶ Response, co-relation, and management of events in near real ime for increased resiliency
Cost	<ul style="list-style-type: none"> ▶ Lower operational costs ▶ Extraction of maximum performance from IT resources with less energy consumption

Many business and IT drivers are forcing data centers to evolve. A smarter data center architecture can be an engine of business growth, yet a simple and scalable method for seamless integration between different IT resources. A smarter data center architecture can provide the flexibility to support the immediate need of business requirements and also future needs for tomorrow's business requirements.

Figure 1-1 depicts a high-level conceptual integration of a smarter data center architecture.

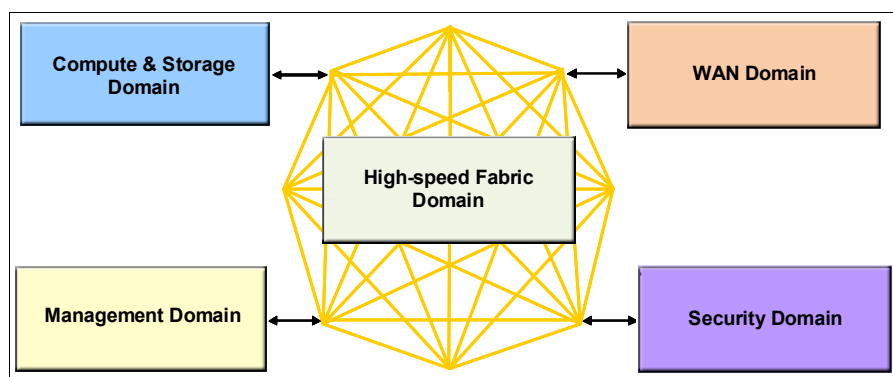


Figure 1-1 Conceptual integration of a smarter data center infrastructure

This architecture uses the following main building blocks:

- ▶ **High-speed fabric domain**
This fundamental domain provides a highly efficient and scalable network with low latency along with cost-efficient resiliency and availability. In addition, this domain supports the rapid service deployment and allows you to use a wide range of devices.
- ▶ **Compute and storage domain**
This domain contains the essential compute and storage infrastructures in an integrated fashion, with built-in expertise to address critical business and operational tasks automatically. Also, it allows you to tune systems in a simplified manner for optimal performance and efficiency.
- ▶ **Management domain**
This domain provides a single user interface to manage entire resource pools. For example, you can reduce the number of steps required to monitor, control, provision, configure, and perform the change management functions for compute, storage, and network.
- ▶ **Wide area network (WAN) domain**
This domain provides external access to data center resources. For example, Internet access, connections to secondary data centers, campus, branch, and cloud environments. These networks enable any-to-any connectivity and interactions that are secure, dynamic, and scalable.
- ▶ **Security domain**
This domain ensures that data is accessed only by authorized users. It protects against evolving threats while enabling agility, and remains available and accessible during disruption, so that it is protected both at rest and in flight.

1.3 Building a smarter data center infrastructure

The smarter data center infrastructure is based on multiple technology domains. However, the IBM Flex System and Juniper Networks QFabric can be pivotal resources when planning a smarter data center, as illustrated in Figure 1-2 on page 5. These resources are reliable, resilient, and energy efficient building blocks for the compute and storage domain and the high-speed fabric domain. Together, they seamlessly integrate within your data center to provide the flexibility, scalability, and capabilities needed now and in the future.

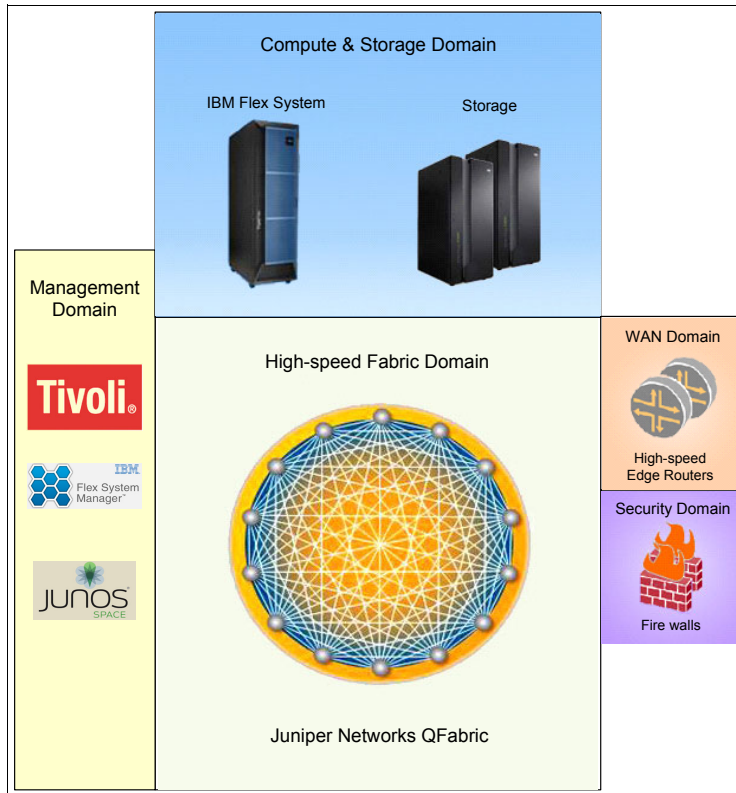


Figure 1-2 Smarter data center building blocks

The design features of IBM Flex System products make it possible for you to configure integrated, customized, and secure solutions that meet the needs of a smarter data center. The scalable hardware features and the energy efficient capabilities of IBM Flex System products also help you optimize hardware utilization, minimize cost, and simplify the overall management of the compute and storage domain.

Juniper Networks QFabric is a packet switched networking technology that can satisfy all high-speed fabric domain requirements. It is specifically designed to create highly efficient, cost-effective, dynamic, and easily managed data centers. QFabric is a scalable product that lets you integrate a wide range of off-the-shelf devices. Those devices connect to QFabric through standard network interfaces, such as Ethernet and Fibre Channel, allowing QFabric to play an integral role in a virtualized, cloud-ready data center network. QFabric is a platform that is immediately ready for deployment in the data center. All QFabric interfaces are equal in terms of latency, bandwidth, and connectivity. The result is a flattened, low-latency, yet highly efficient network, a trending architecture that is replacing the static existing design in many of today's data centers.

In smarter data center environments, automated provisioning of servers, storage, and networks is central to supporting services in a highly-virtualized infrastructure. Integrated management and provisioning is the foundation of service management, such as that supported by the IBM Tivoli® suite, Juniper Networks Junos Space, and IBM Flex System Manager™.

Table 1-2 on page 6 describes the value proposition of combining IBM Flex System and Juniper Networks QFabric to create a smarter data center.

Table 1-2 Value proposition of smarter data center building blocks

Business need	Smarter data center building blocks capabilities
Deployment	IBM Flex System uses optimized configurations that are integrated with advanced IBM hardware and software. Patterns of expertise are included that help you to deploy your solution quickly. Juniper Networks QFabric architecture provides any-to-any connectivity to enhance the connected workload performance.
Scalability	Both IBM Flex System and Juniper Networks QFabric are built by using modular hardware and software components that are highly reliable and scalable. You can add compute and storage capacity to an IBM Flex System without disruption. Juniper Networks QFabric can seamlessly grow to 6144 ports and offers predictable, consistent latency of less than 5 microseconds. Likewise, you can integrate new services, software, and features without disruption to the fabric.
Simplicity	IBM Flex System offers solutions that are pre-integrated, tuned to specific needs, and optimized. These pre-integrated solutions are easier to deploy, streamlining management, thus improving the IT lifecycle with an integrated management of the entire system from a single management view and a broad open ecosystem of optimized solutions. Juniper Networks QFabric provides operational simplicity by behaving as a single switch for the entire network.
Security	IBM Flex System uses Trusted Computing Base (TCB), which secures the entire system's hardware (both systems management and boot firmware). Thus, the foundation of the entire IT infrastructure is secured. Juniper Networks QFabric offers a VM-aware network, which provides visibility and allows network administrators to enforce end-to-end workload-specific security policies when implemented with SRX Series (firewalls) and vGW Series (virtual gateway), thus providing security from the physical layer to the virtual layer.
Resilience	IBM Flex System offers an integrated, easy-to-use management system that provides a predefined resilience policy. This resilience policy is activated when a hardware failure is detected. It also offers you the opportunity to create your own automation plan. You can choose from various monitors and event actions that can help you maintain the health and resilience of your workloads. Juniper Networks QFabric supports VM mobility within the data center. When coupled with MX Series (high-speed edge routers), QFabric offers VM mobility across data centers. Solutions such as IBM Netcool/OMNIBus deliver near real-time, consolidated event management across complex networks and IT domains and provide full management and automation to help deliver continuous uptime of business services and workloads.
Cost	The flexible design of IBM Flex System can meet the needs of varying workloads with independently scalable IT resource pools for higher utilization and lower cost per workload. This design increases security and resiliency to protect vital information and promote maximum uptime. The integrated, easy-to-use management system reduces setup time and complexity and provides a quicker path to return on investment. In addition, the increased workload density in a smaller footprint reduces premium floor space. The Juniper Networks fabric-based design reduces overall network operational cost and complexity that is associated with data center networks. It also protects the existing investment and minimizes total cost of ownership.

To learn more about the value Juniper Networks QFabric can provide, see *Build a Smarter Data Center with Juniper Networks QFabric*, REDP-4830:

<http://www.redbooks.ibm.com/abstracts/redp4830.html?Open>



Understanding the key building blocks for a smarter data center

This chapter expands on the technology domains of the smarter data center, introduced in 1.3, “Building a smarter data center infrastructure” on page 4.

It also provides a high-level overview of how IBM Flex System and Juniper Networks QFabric seamlessly integrate. It explores the options, features, functions, and relationships between the physical and virtual components of IBM Flex System and Juniper Networks QFabric, vGW Series, SRX Series, and MX Series.

2.1 Integrated components: The building blocks

Figure 2-1 depicts the building blocks of a smarter data center infrastructure and the components that align with each technology domain.

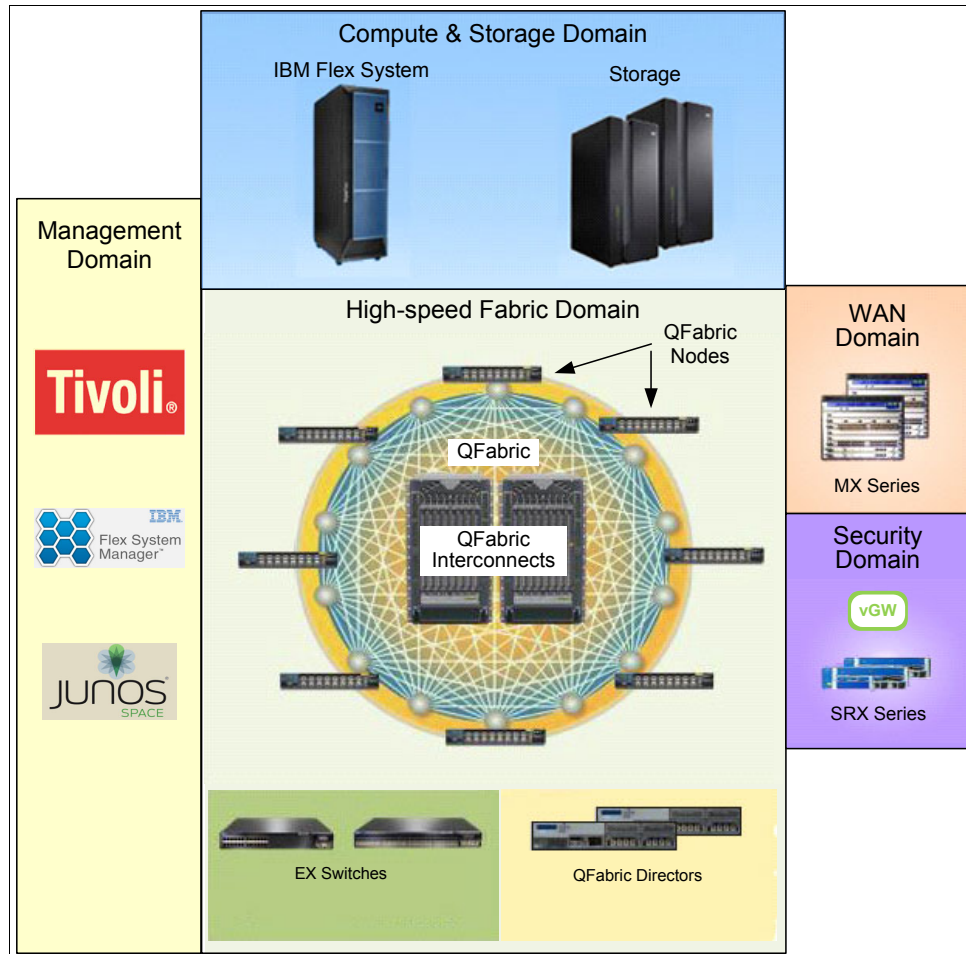


Figure 2-1 Components of a smarter data center

The following components make up each technology domain of the smarter data center infrastructure:

- ▶ Compute and storage domain:
 - IBM Flex System enterprise chassis
 - IBM Flex System compute nodes
 - IBM Flex System I/O modules
 - IBM Flex System storage nodes
 - External storage subsystems
- ▶ High-speed fabric domain:
 - Juniper Networks QFabric Nodes
 - Juniper Networks QFabric Directors
 - Juniper Networks QFabric Interconnects
- ▶ Wide area network (WAN) domain:
 - Juniper Networks MX Series routers

- ▶ Security domain:
 - Juniper Networks SRX Series security gateways
 - Juniper Networks vGW gateway
- ▶ Management domain:
 - IBM Flex System Manager
 - Juniper Networks Junos Space
 - IBM Tivoli

Juniper Networks QFabric creates a high-speed fabric domain and interconnects all the other technology domain components. With QFabric, all servers, storage devices, and network devices are only one hop away from any others, regardless of their location in the fabric, thereby creating an any-to-any high-performance network.

The compute and storage domain contains IBM Flex System, which has many integrated components. IBM Flex System is a flexible, efficient, virtualized, and heterogeneous system platform that easily interfaces with QFabric.

IBM Flex System connects to QFabric seamlessly through interfaces within the IBM Flex System I/O modules and the QFabric Nodes (QFX Series switches). Figure 2-2 shows Juniper Networks QFX3500 top-of-rack (TOR) switches inside an IBM Flex System build-to-order (BTO) configuration.

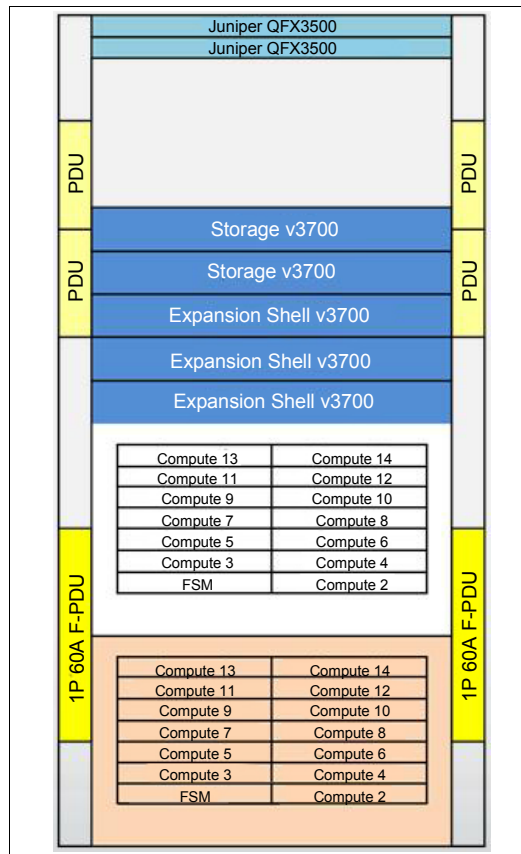


Figure 2-2 Juniper QFX Series TOR switches inside IBM Flex System BTO

Both the IBM Flex System and QFabric interfaces support industry-standard, open protocols for which traditional Ethernet or today’s widely accepted Converged Enhanced Ethernet (CEE) protocols can be configured (see Figure 2-3).

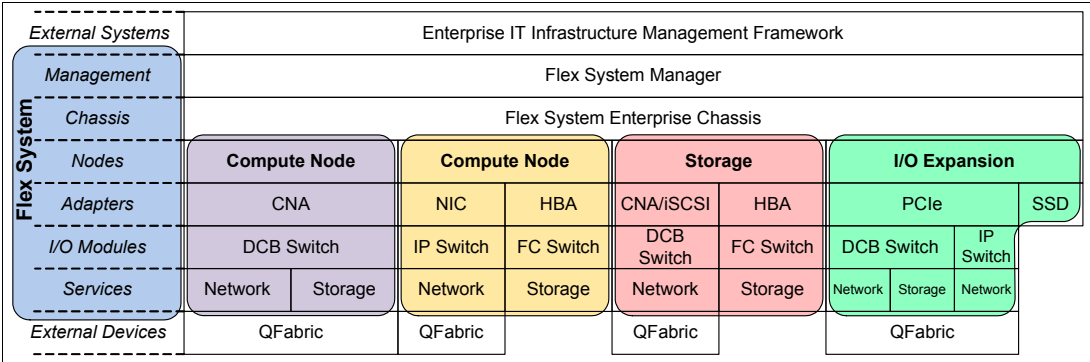


Figure 2-3 IBM Flex System networking and storage interfaces

All the components of the IBM Flex System and Juniper Networks QFabric are discussed further in subsequent sections.

2.2 Compute and storage domain: IBM Flex System

IBM Flex System is built with modular hardware and software components from reliable and scalable IBM technologies that support open protocols. The IBM Flex System family of compute, network, storage (internal and external), and systems management products can be ordered either as fully integrated IBM PureSystems™, or as a customized build-to-order (BTO) IBM Flex System configuration (see Figure 2-4).

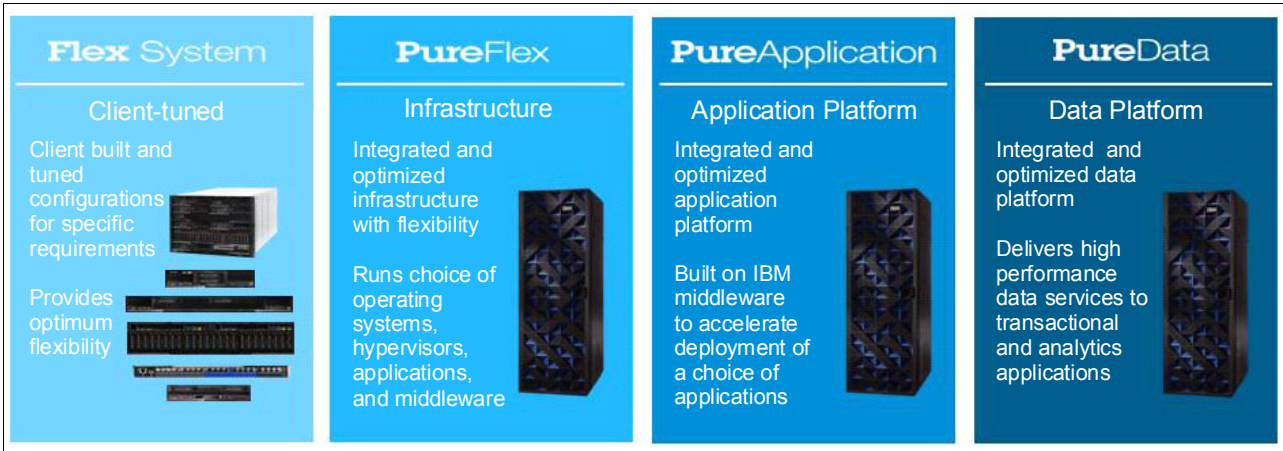


Figure 2-4 IBM Flex System and IBM PureSystems

IBM PureSystems consist of IBM PureFlex™ System, IBM PureApplication™ System, and IBM PureData™ System. The following IBM PureSystems are currently available:

- ▶ IBM PureFlex System Express
Designed for small and medium businesses, this solution is the most affordable entry level solution.
- ▶ IBM PureFlex System Standard
Optimized for application servers with integrated storage and networking, this solution is designed to support your key independent software vendor (ISV) solutions.
- ▶ IBM PureFlex System Enterprise
Optimized for scalable cloud deployments, this solution has built-in redundancy for highly reliable and resilient operation to support your most critical workloads.
- ▶ IBM PureApplication System
Designed and tuned specifically for transactional web and database applications, this workload-aware, flexible platform is designed to be easy to deploy, customize, safeguard, and manage.
- ▶ IBM PureData System
This integrated, optimized, ready-to-run data platform is designed and tuned exclusively for transactional and analytics applications:
 - IBM PureData System for Transactions provides tier-1 database capability by using IBM DB2® pureScale® as the underlying database engine.
 - IBM PureData System for Analytics is powered by Netezza® technology.
 - IBM PureData System for Operational Analytics is powered by IBM Power Systems™ servers and the DB2-based IBM InfoSphere® Warehouse software.
 - IBM PureData System for Hadoop is a standards-based system that integrates IBM InfoSphere BigInsights™ Hadoop-based software, server, and storage into a single, easy-to-manage system.

IBM PureSystems are delivered with integrated virtual fabric switches, storage area network (SAN) switches (configurable in switching or pass-through modes), Flex System Manager, enterprise chassis, and a rack to house all components. If more than one enterprise chassis is ordered in either configuration, a top-of-rack (TOR) Ethernet switch is required. Depending on the integrated solution chosen, a TOR SAN switch might also be required.

IBM Flex System BTO and IBM PureFlex System are included with your choice of compute, expansion, and storage nodes. Many I/O options also exist, allowing for a seamless integration into the smarter data center infrastructure.

The significant difference between IBM PureFlex System and IBM Flex System BTO configurations is the additional flexibility of component level choice. PureFlex System is integrated with IBM virtual fabric technologies and switches; IBM Flex System can be ordered with your choice of switches and pass-through I/O modules, which provides a solution that is easily integrated into Juniper Networks QFabric using QFX3500 or QFX3600 as a TOR switch.

IBM PureFlex System Express and standard configurations with a single enterprise chassis can connect to Juniper Networks QFX3500 or QFX3600 as an access layer switch. Such configurations are also applicable to the use cases in Chapter 4, “Verifying key client use cases with IBM Flex System BTO and Juniper Networks QFabric” on page 75.

Note: IBM PureSystems are listed to show the complete IBM Flex System portfolio. The focus of this publication is only on IBM Flex System BTO and does not provide further details about other solutions.

For more information about the IBM PureFlex System, see *IBM PureFlex System and IBM Flex System Products and Technology*, SG24-7984:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247984.pdf>

2.2.1 Enterprise chassis

The IBM Flex System enterprise chassis, depicted in Figure 2-5, is the backbone of the integrated system. The enterprise chassis provides physical slots for 14 nodes in the front, and four I/O modules in the rear, of various combinations to support your infrastructure requirements.

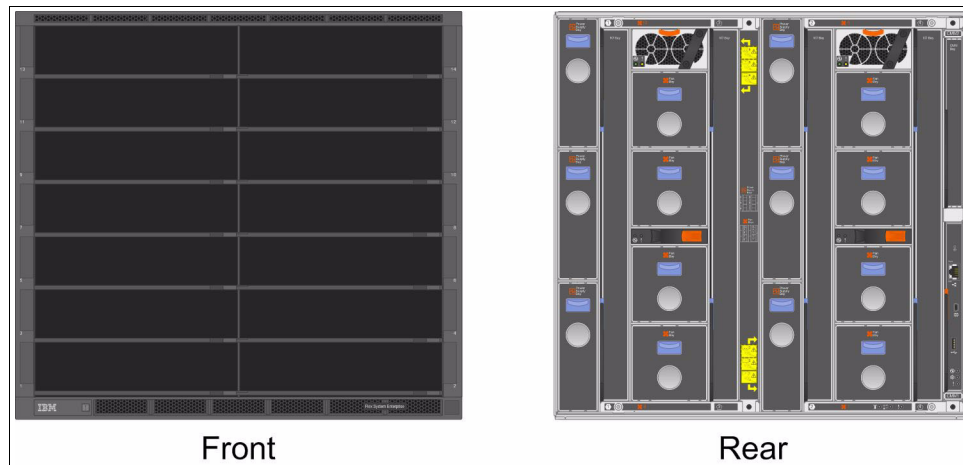


Figure 2-5 IBM Flex System enterprise chassis

The IBM Flex System architecture is agile, allowing for multiple IBM Flex System enterprise chassis to be inter-connected to provide scalable, highly available solutions nondisruptively with minimal effort.

The IBM Flex System includes the following items as part of the enterprise chassis:

- ▶ Nodes
- ▶ Chassis midplane
- ▶ I/O modules
- ▶ Chassis control modules
- ▶ Fan modules
- ▶ PSU modules

At the center of the enterprise chassis is the midplane. The midplane is physically split into two domains, providing redundant, non-blocking passive connections of all components in the enterprise chassis.

2.2.2 Compute nodes

An IBM Flex System compute node is a IBM POWER7® or Intel x86 processor, memory, and I/O controllers integrated onto one hot-swappable module that is installed in the front of the IBM Flex System enterprise chassis. Figure 2-6 shows the currently supported form factors:

- ▶ Half-width
- ▶ Full-width

Apart from form factor, the key difference between the half-width and full-width compute nodes is the quantity of CPU sockets, available slots for installation of memory modules, and quantity of high speed I/O lanes present for the connection of PCIe I/O adapters.

The half-width compute nodes provide up to 16 channels of high speed I/O lanes, and the full-width compute nodes provide up to 32 channels of high speed I/O lanes.

Resiliency is ensured in both half-width and full-width compute nodes. Both form factors contain redundant PCIe slots for connection of redundant I/O adapters, ensuring no single points of failure exist through the redundant planes of the enterprise chassis midplane, to the I/O modules, and onto the TOR switching infrastructure.

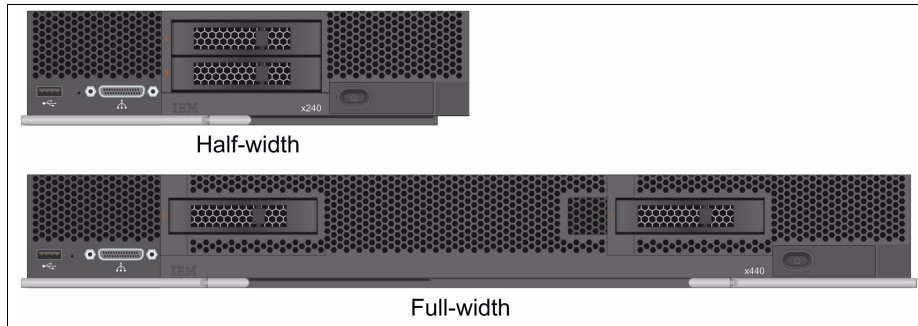


Figure 2-6 Compute node form factor options

The following compute node models are currently supported:

- ▶ Half-width:
 - IBM Flex System x240 Compute Node, a two socket Intel Xeon
 - IBM Flex System x220 Compute Node, a cost-optimized two-socket Intel Xeon
 - IBM Flex System p260 Compute Node, a two socket IBM POWER7 or IBM POWER7+™ processors
 - IBM Flex System p24L Compute Node, a two socket IBM POWER7, optimized for Linux
- ▶ Full-width:
 - IBM Flex System x440 Compute Node, a performance-optimized four-socket Intel Xeon
 - IBM Flex System p460 Compute Node, a four socket IBM POWER7

Each model can be customized to support various memory and I/O adapter options.

Compute node I/O adapters

Several Flex System compute node models include integrated local area network (LAN) on motherboard (LOM) I/O adapters, which can be removed to support the installation of an optional I/O adapter. The LOM simply provides a link between the integrated adapter on the compute node to the Flex System enterprise chassis.

The following I/O adapters are currently supported:

- ▶ IBM Flex System EN2024 4-port 1 Gb Ethernet Adapter
- ▶ IBM Flex System EN4132 2-port 10 Gb Ethernet Adapter
- ▶ IBM Flex System EN4054 4-port 10 Gb Ethernet Adapter
- ▶ IBM Flex System CN4054 4-port 10 Gb Converged Adapter
- ▶ IBM Flex System CN4058 8-port 10 Gb Converged Adapter
- ▶ IBM Flex System EN4132 2-port 10 Gb RoCE Adapter
- ▶ IBM Flex System FC3172 2-port 8 Gb FC Adapter
- ▶ IBM Flex System FC3052 2-port 8 Gb FC Adapter
- ▶ IBM Flex System FC5022 2-port 16 Gb FC Adapter
- ▶ IBM Flex System IB6132 2-port FDR InfiniBand Adapter
- ▶ IBM Flex System IB6132 2-port QDR InfiniBand Adapter

Figure 2-7 depicts the I/O architecture between the LOM, I/O adapter, and the I/O modules through passive connections on the enterprise chassis midplane.

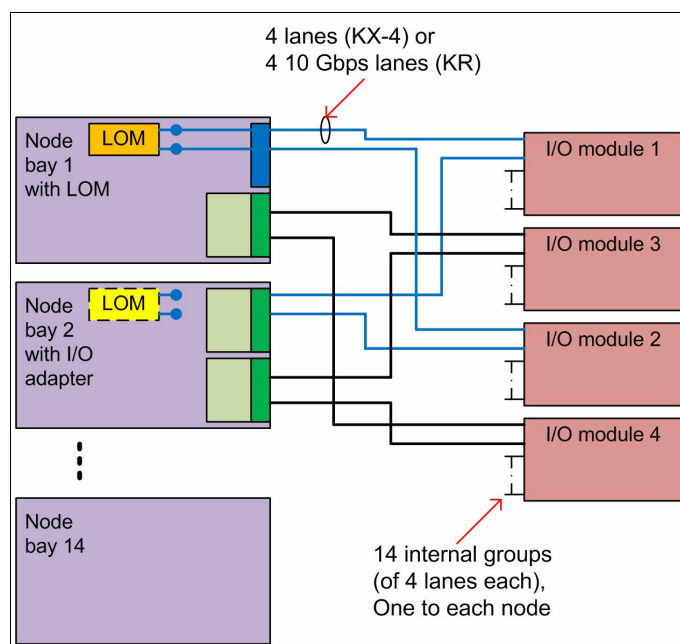


Figure 2-7 LOM, I/O adapter, and I/O modules connectivity

If a node has a two-port integrated LOM as standard, modules 1 and 2 are connected to this LOM. If an I/O adapter is installed in the node's I/O expansion slot 1, modules 1 and 2 are connected to this adapter.

I/O modules 3 and 4 connect to the I/O adapter that is installed within I/O expansion bay 2 on the node. These I/O modules provide external connectivity, and connect internally to each of the nodes within the chassis.

2.2.3 I/O modules

Four slots in the rear of the IBM Flex System enterprise chassis accommodate various combinations of I/O modules as depicted in Figure 2-8.

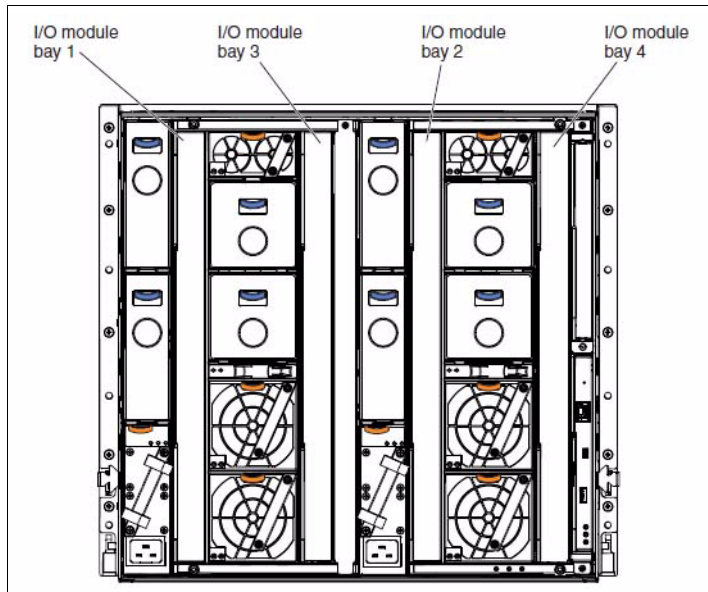


Figure 2-8 Enterprise chassis I/O module bays

IBM Flex System I/O modules are installed in redundant pairs and are available in two topologies as follows:

- ▶ Switched:
 - IBM Flex System Fabric CN4093 10 Gb Converged Scalable Switch
 - IBM Flex System Fabric EN4093 and EN4093R 10 Gb Scalable Switch
 - IBM Flex System EN2092 1 Gb Ethernet Scalable Switch
 - IBM Flex System FC5022 16 Gb SAN Scalable Switch
 - IBM Flex System FC3171 8 Gb SAN Switch
 - IBM Flex System IB6131 InfiniBand Switch
- ▶ Pass-through:
 - IBM Flex System EN4091 10 Gb Ethernet pass-through
 - IBM Flex System FC3171 8 Gb SAN pass-through

All I/O modules are the same form factor, with the difference being their topology and port configurations.

2.2.4 Expansion nodes

IBM Flex System expansion nodes are inserted into the front of the enterprise chassis and provide additional functionality to the compute nodes.

The following IBM Flex System expansion nodes are currently supported:

- ▶ IBM Flex System PCIe expansion node
- ▶ IBM Flex System storage expansion node

The IBM Flex System PCIe and Storage expansion nodes are cabled directly to a half-width compute node through an interposer cable. The interposer cable connects to an expansion

connector that is physically wired to an I/O controller integrated into processor 2 of the compute node. For this reason, the half-width compute node must have both processors installed.

Figure 2-9 depicts an installed I/O expansion node onto an x240 compute node.

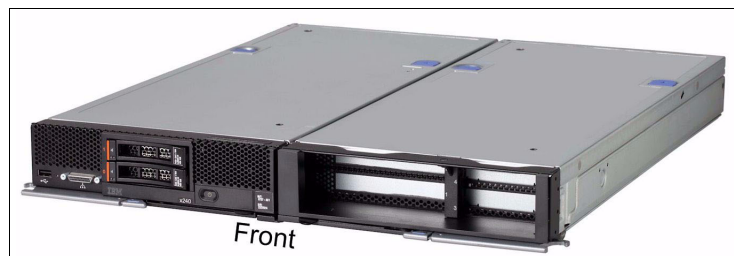


Figure 2-9 PCIe expansion node attached to a compute node

The PCIe expansion node supports both PCIe 2.0 and PCIe 3.0 adapters. PCIe 3.0 adapters are software set to PCIe 2.0 mode. Virtual fabric, Ethernet, Fibre Channel (FC), and InfiniBand adapters can be ordered separately.

The IBM Flex System storage expansion node contains a network-attached storage (NAS) controller and a 12-slot 2.5" disk drive expansion chassis, which when cabled to a half-width compute node, provides dedicated storage to the installed compute node. The NAS controller supports Redundant Array of Independent Disks (RAID) 0, 1, 5, 6, 10, 50, and 60 that are configured using hot swappable SAS/SATA or SSD drives. Cache options include 512 MB or 1 GB modules.

Figure 2-10 depicts an IBM Flex System storage expansion node that is attached to a compute node.

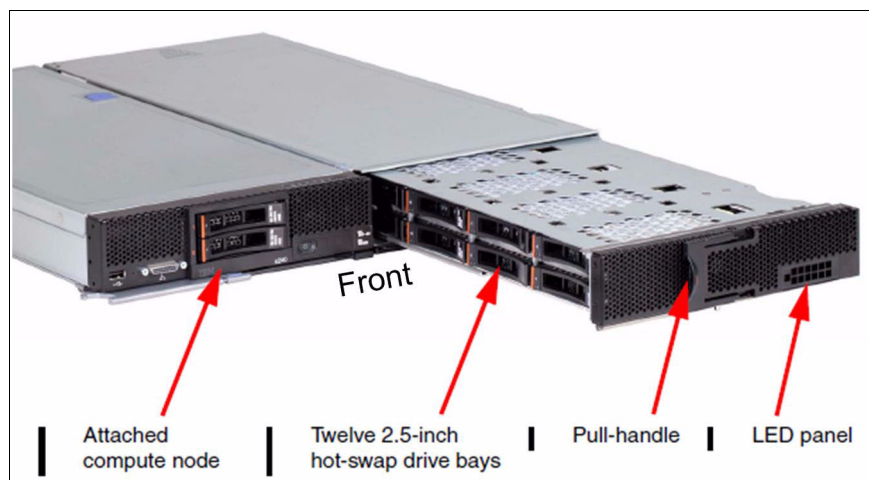


Figure 2-10 IBM Flex System storage expansion node attached to compute node

Consult the IBM ServerProven® website for current lists of components that are supported in the enterprise chassis:

<http://ibm.com/systems/info/x86servers/serverproven/compat/us/flexsystems.html>

For information about any of the IBM Flex System components detailed in this chapter, see the IBM Redbooks Product Guides for IBM Flex System:

<http://www.redbooks.ibm.com/portals/puresystems>

2.2.5 Virtual fabric networking

Currently, deployment of server virtualization technologies in data centers requires significant effort to provide sufficient network I/O bandwidth to satisfy the demands of virtualized workloads and services. For example, every virtualized system can host several dozen workloads and services, and each of these services requires bandwidth to function properly. Furthermore, because of various network traffic patterns relevant to various service types, these traffic flows might interfere with each other. This interference can lead to serious network problems, including the inability of the service to run its functions.

The IBM virtual fabric virtual network interface card (vNIC) solution addresses these issues. The solution is based on 10 Gb Converged Enhanced Ethernet (CEE) infrastructure. It takes a 10 Gb port on a 10 Gb virtual fabric I/O adapter and splits the 10 Gb physical port into four vNICs. This configuration allows each vNIC or virtual channel to be allocated between 100 Mb and 10 Gb in increments of 100 Mb. The total of all four vNICs cannot exceed 10 Gb. For example, when a 10 Gb virtual fabric I/O adapter is split into four vNICs, each vNIC or virtual channel allowed bandwidth is between 100 Mb and 2.5 Gb in increments of 100 Mb.

The vNIC solution is a way to divide a physical NIC into smaller logical NICs (or partition them). This configuration allows the operating system (OS) to have more possible ways to logically connect to the infrastructure. The vNIC feature requires a 10 Gb virtual fabric adapter, or an embedded virtual fabric adapter.

All vNIC modes have the following common elements:

- ▶ vNIC modes are supported only on 10 Gb connections.
- ▶ A NIC can be divided into up to four vNICs per physical NIC (can be less than four, but not more).
- ▶ The I/O adapter must support vNIC modes.
- ▶ When creating vNICs, the default bandwidth is 2.5 Gb for each vNIC. However, you can configure the bandwidth to be anywhere from 100 Mb up to the full bandwidth of the NIC.
- ▶ The bandwidth of all configured vNICs on a physical NIC cannot exceed 10 Gb.

Two primary forms of vNIC modes are available:

- ▶ Switch independent vNIC mode. This mode allows the connection of a TOR for the termination of CEE or Layer 2 networks at the compute node I/O adapter. This mode requires the compute node operating system to manage vNIC fail-over through the installation of a supported I/O adapter driver.
- ▶ Switch dependent vNIC mode. This mode configures redundant vNIC uplinks between the I/O module switch and the compute node I/O adapters. The CEE or Layer 2 networks can be extended from the I/O modules switch to a TOR switch to permit external network communications.

Switch independent vNIC mode

Switch independent vNIC mode is a flexible IBM Flex System option that specifically permits the connection of a TOR switch. Switch independent vNIC mode is enabled on the IBM Flex System compute node I/O adapter. Currently the CN4054 or CN4058 compute node I/O adapters support switch independent vNIC mode. In this mode, the IBM Flex System I/O module is unaware of the networks virtualized within the TOR switch.

This mode is typically used in conjunction with an IBM Flex System EN4091 pass-through I/O module.

Figure 2-11 depicts an IBM Flex System BTO with CN4054 compute node I/O adapters that are configured in switch independent vNIC mode, connecting to TOR switches that use EN4091 pass-through I/O modules.

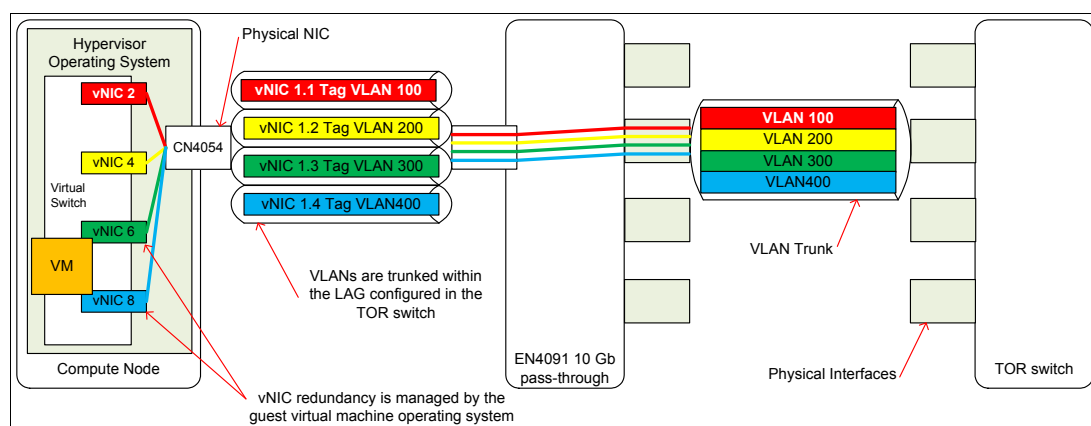


Figure 2-11 Virtual fabric switch independent vNIC mode

See 3.1.3, “Providing an integrated solution with IBM Flex System BTO and QFabric” on page 56 for more details about the integration of IBM Flex System EN4091 and TOR switches.

Switch dependent vNIC mode

Switch dependent vNIC mode (or virtual fabric vNIC mode) depends on the capability of the I/O modules that are installed. Specifically, the I/O module that supports this mode of operation today in the enterprise chassis is the IBM Flex System Fabric EN4093 10 Gb scalable switch.

Switch dependent vNIC mode is configured on the EN4093 switch, which negotiates and sets the interface topology with the I/O adapters on the IBM Flex System compute node. These links between the EN4093 switch and I/O adapters are called uplinks.

Switch dependent vNIC mode supports the following uplink types:

- Dedicated uplink mode
- Shared uplink mode

Uplinks are assigned to groups of vNICs, called vNIC groups. Each vNIC group contains the physical ports of the compute node I/O adapters. The EN4093 switch creates a virtual fabric switch for each vNIC group defined in the IBM Flex System chassis.

How the vNIC groups are used is the primary difference between the dedicated uplink mode and the shared uplink mode.

Dedicated uplink mode

Dedicated uplink mode, as depicted in Figure 2-12, is the default mode when vNIC is enabled on the I/O module. Each vNIC group must have its own dedicated physical or logical (aggregation) uplink. This mode does not allow you to assign more than a single physical or logical uplink to a vNIC group. In addition, it assumes that high availability is achieved by some combination of aggregation on the uplink and NIC teaming on the server.

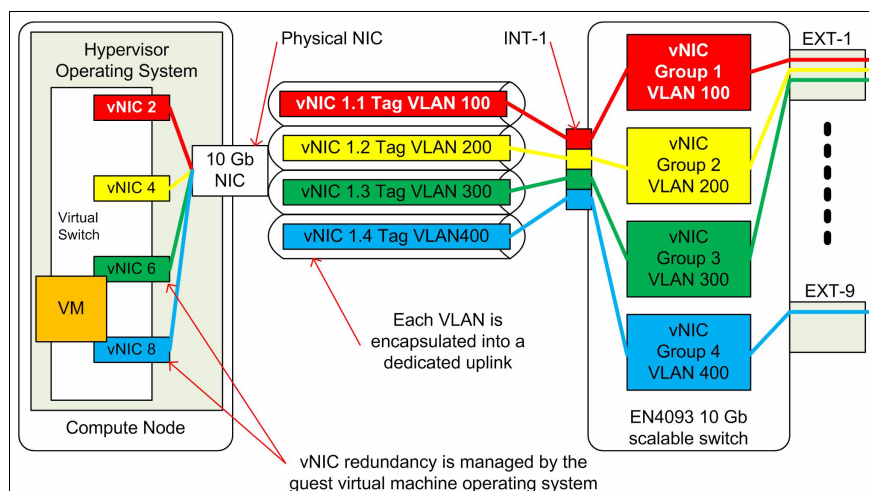


Figure 2-12 Virtual fabric vNIC dedicated uplink mode

Shared uplink mode

Shared uplink mode, as depicted in Figure 2-13, is a global option that can be enabled on the EN4093 switch. Shared uplink mode shares available uplink between all vNIC groups defined in the EN4093 switch.

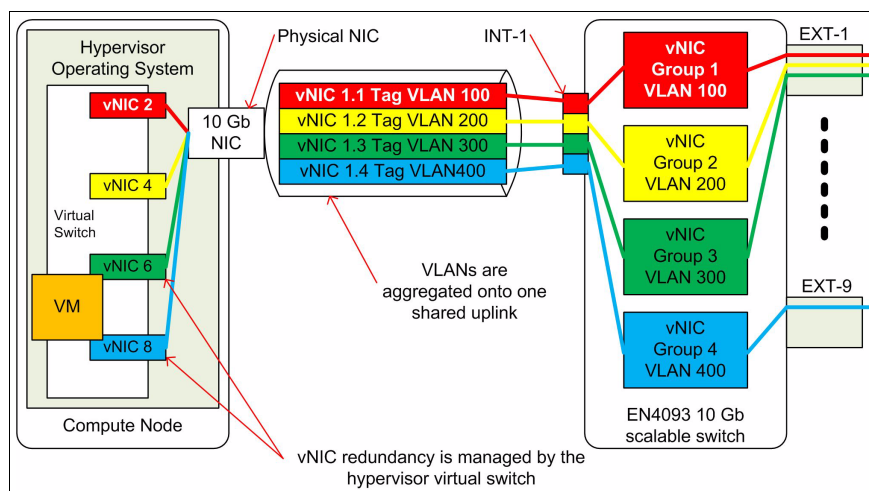


Figure 2-13 Virtual fabric vNIC shared uplink mode

2.2.6 FCoE solution capabilities

One common way to reduce management points in an environment is by converging technologies that were implemented on separate infrastructures. Similar to collapsing office phone systems from a separate cabling plant and components into a common IP infrastructure, Fibre Channel networks also are experiencing this type of convergence. Like phones that move to Ethernet, Fibre Channel also is moving to Ethernet.

Fibre Channel over Ethernet (FCoE) removes the need for separate host bus adapters (HBAs) on the servers and separate Fibre Channel cables flowing out the back of the server or chassis. Instead, a converged network adapter (CNA) is installed in the server. The CNA presents what appears to be both a NIC and an HBA to the operating system, but the output from the server is 10 Gb Ethernet.

2.2.7 VM-aware networking

Virtual machines running on physical servers must be visible to, and recognized by the network to ensure that end-to-end workload-specific (or VM-specific) security and quality of service (QoS) policies are enforced. Hence, the network must be VM-aware.

IBM offers two ways to make the network VM-aware:

- ▶ IBM VMready® functionality embedded into the network switch establishes virtual ports for every VM inside the switch and treats these ports as regular physical ports for end-to-end policy enforcement
- ▶ IBM System Networking Distributed Switch 5000V is an orderable product that replaces the standard vSwitch inside the VMware vSphere 5 hypervisor, thus extending its functionality by supporting access control lists (ACLs), QoS, link aggregation, and switch port analyzer, therefore enforcing required policies end-to-end

VMready capabilities

IBM VMready enables the network to be VM-aware. By using VMready, configuration and management of networks can be achieved at the virtual port layer, rather than just the physical port layer. VMready allows for a “define one, use many” configuration, meaning that network attributes are bundled with a virtual port. The virtual port belongs to a VM and is movable. If the VM migrates, even to a different physical host, the network attributes of the virtual port remain the same.

The hypervisor manages the various virtual entities on the host server, such as the VMs, virtual switches, and other virtual entities. Currently, VMready functions support up to 2048 virtual entities in a virtualized data center environment. The switch automatically discovers the virtual entities that are attached to switch ports and distinguishes between regular VMs, service console interfaces, and kernel or management interfaces in a VMware environment.

Virtual entities can be placed into VM groups on the switch to define communication boundaries. Virtual entities in the same VM group can communicate with each other, but virtual entities in different groups cannot. VM groups also allow for configuring group-level settings, such as virtualization policies and ACLs.

The administrator can also pre-provision virtual entities by adding their MAC addresses (or their IPv4 addresses or VM names in a VMware environment) to a VM group. When a virtual entity with a pre-provisioned MAC address becomes connected to the switch, the switch automatically applies the appropriate group membership configuration. In addition, VMready, together with IBM NMotion®, allows seamless migration or failover of VMs to different hypervisor hosts, preserving network connectivity configurations.

VMready works with all major virtualization products, including VMware, Hyper-V, Xen, and KVM and Oracle VM, without modification of virtualization hypervisors or guest operating systems. A VMready switch can also connect to a virtualization management server to collect configuration information about associated virtual entities. It can automatically push VM group configuration profiles to the virtualization management server. This process in turn configures the hypervisors and virtual entities, providing enhanced virtual entity mobility.

Distributed Virtual Switch 5000V capabilities

The IBM Distributed Virtual Switch (DVS) 5000V is an advanced, feature-rich distributed virtual switch for VMware environments with policy-based VM connectivity. The DVS 5000V operates as a VM within the vSphere cluster.

The DVS 5000V has two components:

- ▶ IBM DVS 5000V controller

This controller is a VMware compliant open virtual appliance (OVA) file which is installed within a vSphere 5 cluster. The controller contains the core functionality of the DVS 5000V manages virtual switches across multiple VMware ESXi server hypervisors so that the DVS 5000V is visible to the physical network as a distributed virtual switch.

- ▶ IBM DVS 5000V host module

This module is a VMware compliant vSphere installation base (VIB) which is embedded into each ESXi server hypervisor within a vSphere 5 cluster. The host module is a Layer 2 virtual switch, integrated with vCenter to create a distributed virtual switch across all members of the cluster.

The DVS 5000V works with VMware vSphere 5.0 (or later) and interoperates with any 802.1Qbg-compliant physical switch to enable switching of local VM traffic in the hypervisor or in the upstream physical switch. The DVS 5000V enables a large-scale, secure, dynamic integrated virtual and physical environment for efficient VM networking that is aware of server virtualization events, such as VMotion and Distributed Resource Scheduler (DRS).

Configuration is performed through the DVS 5000V Controller and is automatically propagated so that administrators can define configurations for all virtual switches from a single interface to enable simplified management of virtual machine traffic. Private VLANs enable VM traffic separation, ACLs provide VM traffic control, and local port mirroring and remote port mirroring enable advanced VM traffic visibility and troubleshooting.

The DVS 5000V enables network administrators who are familiar with IBM System Networking switches to manage the DVS 5000V just like physical switches by using advanced networking, troubleshooting, and management features so that the virtual switch is no longer hidden and difficult to manage.

Network administrators within a Juniper Networks environment can manage virtual switches from Junos Space Virtual Control using Link Layer Discovery Protocol (LLDP). This approach allows network administrators visibility to virtual switches within an IBM Flex System that hosts a VMware vSphere environment. Junos Space Virtual Control can be used to centralize configuration management, reporting and automating repeatable management tasks. See “Junos Space Virtual Control” on page 44 for more details.

Edge Virtual Bridging (EVB) is an IEEE standard that specifies the interaction between virtual switching environments in a hypervisor and the first layer of the physical switching infrastructure. The EVB 802.1Qbg standard addresses the lack of network management, monitoring, and security with conventional vSwitches. Typically, a conventional vSwitch is invisible to and not configurable by the network administrator. Additionally, any traffic handled internally by the vSwitch cannot be monitored or secured.

Support for the EVB 802.1Qbg standard enables VM traffic management in the physical/virtual network through Virtual Ethernet Port Aggregation (VEPA) and Virtual Station Interface (VSI) Discovery and Configuration Protocol (VDP). This enables live VM mobility across the network through automatic creation, migration, and deletion of VM-specific network port profiles.

The DVS 5000V offers the following capabilities:

- ▶ Supported on VMware vSphere 5
- ▶ Manageability: Telnet, Secure Shell (SSH), Link Layer Discovery Protocol (LLDP), Simple Network Management Protocol (SNMP), TACACS+, RADIUS, and Industry Standard CLI
- ▶ Advanced networking features: L2-L4 ACLs, Static and Dynamic port aggregation, PVLAN, QoS, and EVB (IEEE 802.1Qbg)

See more information about DVS 5000V at the following locations:

- ▶ <http://www.ibm.com/systems/networking/switches/virtual/dvs5000v/index.html>
- ▶ <http://www-304.ibm.com/support/docview.wss?uid=isg3T7000509&aid=1>

To learn about other virtual switch types, see “Virtual switches” on page 48.

2.2.8 Network management integration

The IBM Flex System Manager node relies on internal network connectivity to manage the IBM Flex System. The IBM Flex System Manager has a basic Layer 2 switch that is integrated into the IBM Flex System Manager hardware and that provides the following features:

- ▶ Management network switching within the enterprise chassis
- ▶ Management network connection to external interface on chassis management module (CMM)
- ▶ Management network connection of integrated management module (IMM)
- ▶ Data network connectivity of IBM Flex System Manager for compute node operating system management

The integrated switch is connected to the CMM through the chassis internal management network. The CMM is connected to a top-of-rack switch that effectively extends the chassis management network outside of the chassis and provides a central connection point for all enterprise chassis hardware to be managed by the IBM Flex System Manager management software. Figure 2-14 on page 23 shows the management network.

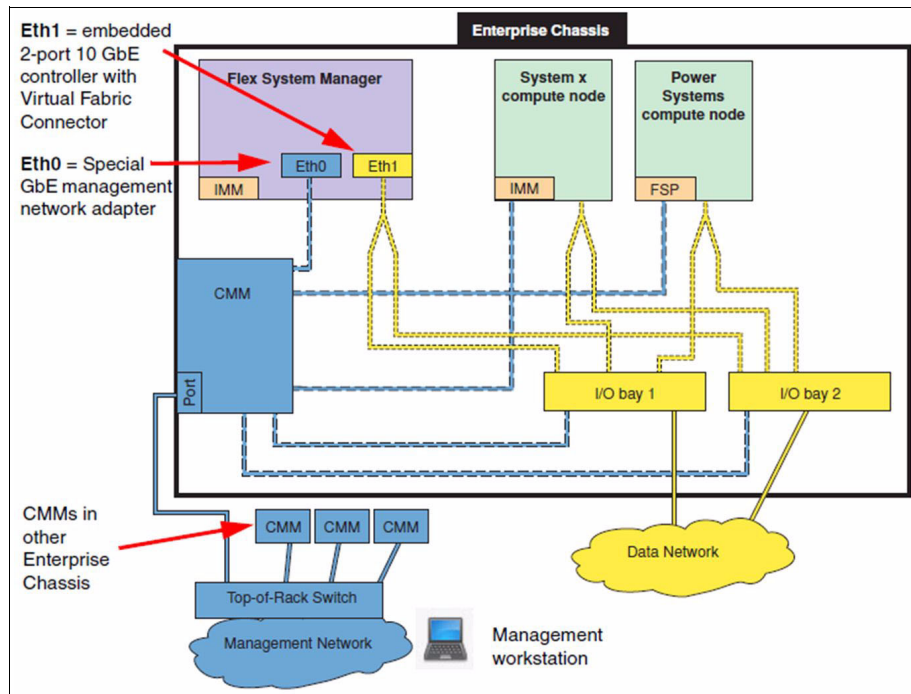


Figure 2-14 IBM Flex System Manager management network

The management network is used to complete management-related functions for the various endpoints that are managed by the IBM Flex System Manager management software, such as other enterprise chassis and compute nodes. During initialization, the management software discovers any enterprise chassis on the management network. The management node console can be connected to the management network or to the data network.

The IBM Flex System Manager node Eth1 interfaces (two 10 Gb Ethernet ports) are connected to the chassis I/O modules that are installed in I/O bay 1 and I/O bay 2.

The key function of the data network connection to the IBM Flex System Manager is automated discovery of IBM Flex System compute node operating system (OS) types and versions. The IBM Flex System Manager discovers a supported OS on a compute node, imports the OS as a managed system, and then allows the system administrator to automate software updates through the data network.

The integrated management module (IMM) is a web-based graphical user interface (GUI) from which the systems management operations are executed.

See 2.6.1, "IBM Flex System Manager" on page 39 for more information about IBM Flex System Manager.

2.2.9 IBM Flex System internal and external storage

The IBM Flex System suite of storage products provide flexible and extensible solutions from the point of deployment to integration with existing data center storage environments. IBM Flex System storage are available in the following forms:

- ▶ Internal storage
 - Comprises two integrated storage subsystems, IBM Storwize® V7000 and IBM Flex System V7000 storage node
- ▶ External storage
 - Supports connection of various Fibre Channel, iSCSI, and network-attached storage storage subsystems

Internal storage

Currently, the following integrated options exist within the IBM Flex System:

- ▶ Storwize V7000
- ▶ IBM Flex System V7000 storage node

Storwize V7000 delivers essential storage efficiency technologies and ease of use and performance in a modular design. Storwize V7000 is considered external storage from an IBM Flex System perspective.

Storwize V7000 architecture is the same as that of the IBM Flex System V7000 storage node and is managed from the IBM Flex System chassis management module (CMM) or IBM Flex System Manager node.

IBM Flex System V7000 storage node is inserted into the front of the IBM Flex System enterprise chassis to enable rapid storage deployment and management simplicity. This class of storage system combines the virtualization, efficiency, and performance capabilities of Storwize V7000. It helps simplify and speed IBM Flex System infrastructure deployment with server and storage management integration to automate provisioning and to help achieve responsiveness to business needs while reducing costs.

IBM Flex System storage provides the following benefits:

- ▶ Supports IBM Flex System compute nodes across multiple chassis.
- ▶ Provides automated deployment and discovery.
- ▶ Supports growing business requirements while controlling costs.
- ▶ Includes FCoE optimized offering (plus FC and iSCSI).
- ▶ Supports thin provisioning, IBM FlashCopy®, IBM Easy Tier®, IBM Real-time Compression™, and nondisruptive migration.
- ▶ Provides performance improvement with automatic migration to high-performing solid-state drives.
- ▶ Enables near-continuous availability of applications through dynamic migration.
- ▶ Supports faster and more efficient data copies for online backup, testing, or data mining.
- ▶ Offers flexible server and storage management with easy to use GUI for block and file storage management.

IBM Flex System V7000 storage node and the IBM Flex System enterprise chassis offer several possibilities for integration into existing data center storage infrastructures, such as Fibre Channel, iSCSI, and CEE.

External storage

You can use one of the following options to attach external storage systems to an enterprise chassis:

- ▶ Storage area networks (SANs) based on Fibre Channel technologies
- ▶ SANs based on iSCSI
- ▶ Converged Networks based on 10 Gb CEE

The connection to the external storage devices is achieved using I/O modules inserted into the rear of the enterprise chassis.

Fibre Channel-based SANs are the most common design of external storage infrastructure. They provide high levels of performance, availability, redundancy, and scalability.

Fibre Channel SANs typically consist of the following components:

- ▶ Server HBAs
- ▶ Fibre Channel switches
- ▶ Fibre Channel storage subsystems
- ▶ Fibre Channel tape subsystems
- ▶ Optical cables for connecting these devices to each other

iSCSI-based SANs provide all the benefits of centralized shared storage in terms of storage consolidation and adequate levels of performance but use traditional IP-based Ethernet networks.

iSCSI SANs typically consist of the following components:

- ▶ Server hardware iSCSI adapters or software iSCSI initiators
- ▶ Traditional IP network components, such as switches and routers
- ▶ Storage subsystems with iSCSI interfaces, such as IBM System Storage DS3500 or IBM N Series

Converged networks can carry both SAN and LAN traffic over the same physical infrastructure, thus decreasing cost and increasing efficiency in implementation and management. Converged networks are the foundation of a smarter data center network.

For the latest support matrix information for storage products, see the storage vendors interoperability guides. You can find IBM storage products in the IBM System Storage® Interoperation Center (SSIC):

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

2.3 High-speed fabric domain: Juniper Networks QFabric

When designing QFabric, Juniper Networks took a physical design approach as opposed to a protocol design approach. The physical design involved re-architecting the physical layers of the data center network and applying the following key principles:

- ▶ Create a lossless, loop-free, and converged Ethernet network with a single switch view
- ▶ Reduce the number of network tiers to optimize data traffic
- ▶ Separate the data and control plane inside the data center network
- ▶ Move toward a flat data plane that provides any-to-any non-blocking connectivity
- ▶ Enable full utilization of the connectivity by eliminating old approaches
- ▶ Move toward a new control plane that enables multi-pathing

- ▶ Simplify management by decreasing the number of discrete devices
- ▶ Centralize management and provisioning to support automation
- ▶ Move toward a single management plane to manage the network as though it were one big switch

QFabric is built using modular hardware and software components that are highly reliable and scalable; together they provide the ability to interconnect and manage all interfaces in the fabric like a single logical switch.

As illustrated in Figure 2-15, the QFabric infrastructure consists of three planes; the data plane, the control plane, and the management plane:

- ▶ A single, flat data plane with the following component models:
 - QFabric Nodes (QFX3500 and QFX3600)
 - Redundant QFabric Interconnects (QFX3008-I or QFX3600-I)
- ▶ A distributed control plane:
 - Control plane elements consist of the QFabric Directors, QFabric Interconnects, and QFabric Nodes
 - The control plane infrastructure uses virtual chassis switches (EX4200-24T, EX4200-24F, or EX4200-48T) to interconnect the control plane elements
- ▶ A single management plane with the following component model:
 - Redundant QFX3100 directors

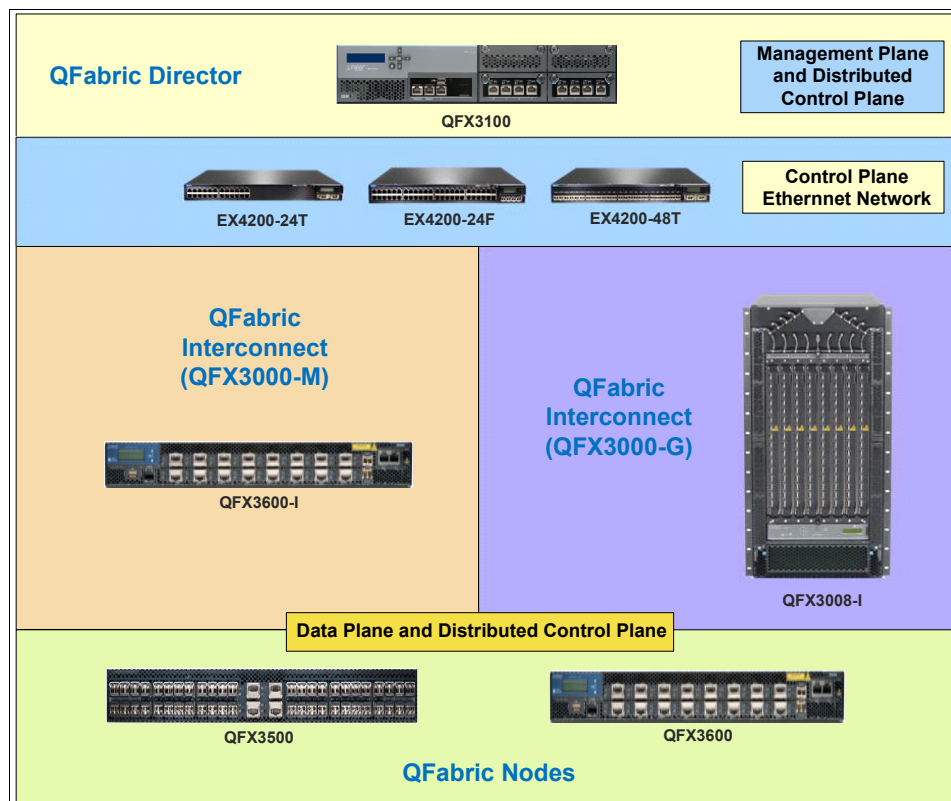


Figure 2-15 QFabric infrastructure

2.3.1 QFabric data plane components

In QFabric, data traffic from servers, storage, and network devices is transported across a redundant, high-performance, and scalable data plane that consists of two component types:

- ▶ QFabric Nodes, which are equivalent to line cards in a switch, offering ports and route engine functions.
- ▶ QFabric Interconnects, which are related to the backplane of a switch, allowing quicker forwarding of packets inside the QFabric.

Each QFabric Node is connected to all QFabric Interconnects in the fabric, establishing redundancy for high availability and multiple paths for low latency.

The QFabric data plane supports transport of data that uses these high-speed connections:

- ▶ QFabric Nodes support 1 Gb or 10 Gb Ethernet or 2 Gbps, 4 Gbps, or 8 Gbps Fibre Channel connections to servers, storage, and network devices
- ▶ Connections between the QFabric Nodes and the QFabric Interconnects are provided through 40 Gbps quad, small form-factor pluggable plus (QSFP+)

QFabric is available in the following configurations:

- ▶ QFX3000-M is designed for small and medium data center environments. It supports up to 768 10 GbE ports with an average latency of 3 microseconds port to port.
- ▶ QFX3000-G is designed for large enterprises, service providers, and cloud environments. It supports up to 6,144 10 GbE ports with an average latency of 5 microseconds from port to port.

QFabric Nodes (QFX3500 and QFX3600) can be deployed interchangeably in QFX3000-M and QFX3000-G environments.

QFabric QFX3000-M System Interconnect

QFabric with the interconnect type of QFX3600-I is referred to as QFX3000-M. The QFX3600-I acts as a backplane that interconnects the QFX3500 and QFX3600 nodes.

The QFabric QFX3600-I has the following key specifications and features:

- ▶ 1U high fixed configuration
- ▶ 16 QSFP+ ports
- ▶ Redundant AC or DC power supplies
- ▶ Front to back (AFI) and back to front (AFO) air flow (separate SKUs)
- ▶ Fiber or copper control plane ports

QFabric QFX3000-G System interconnect

QFabric with the interconnect type of QFX3008-I is referred to as QFX3000-G. The QFX3008-I provides the high-speed transport that serves as the backplane that interconnects the QFX3500 and QFX3600 nodes.

The QFX3008-I includes the following key specifications and features:

- ▶ Size of 21U with an industry-standard 19-inch rack-mount enclosure
- ▶ Up to 16 cards (16-port QSFP+); eight in the front and eight in the rear
- ▶ Control boards (in the rear):
 - Two control boards (master and backup)
 - Four Gigabit Ethernet copper SFPs (two on each board), control plane ports

- Six AC power supplies
- Front to back (AFI) air flow

Table 2-1 lists the key specification and supported feature differences between the QFX3000-M and QFX3000-G systems.

Table 2-1 QFX3000-M and QFX3000-G comparison

Criteria	QFX3000-M	QFX3000-G
QFabric Interconnect model	QFX3600-I (1U)	QFX3008-I (21U)
QFabric Director model	QFX3100	QFX3100
QFabric Nodes models	QFX3500 & QFX3600	QFX3500 & QFX3600
Number of Interconnects^a	2 or 4	2 or 4
Types of QFabric Node client device connections	1/10 GbE or 2/4/8 Gbps Fibre Channel	1/10 GbE or 2/4/8 Gbps Fibre Channel
QFabric Node to QFabric Interconnect connections	Up to four 40 GbE uplink from each Node to Interconnect	Up to four 40 GbE uplink from each Node to Interconnect
Maximum QFabric Nodes supported	16	128
Number of Directors	2	2
Number of links in LAG between Directors	2	2
Maximum 1 GbE ports	16 Nodes x 36 ports = 576	128 Nodes x 36 ports = 4,608
Maximum 10 GbE ports	16 Nodes x 48 ports = 768	128 Nodes x 48 ports = 6,144
SAN connectivity	FCoE transit capable and FCoE/FC Gateway	FCoE transit capable and FCoE/FC Gateway
OS support	JUNOS	JUNOS
MAC table size	Up to 120,000 for each Node	Up to 120,000 for each Node
Number of VLANs supported	Up to 4096	Up to 4096

a. Depends on the number of QFabric Node ports and oversubscribe configuration

Both the QFX3008-I and QFX3600-I run the same Juniper Networks JUNOS operating system as other Juniper switches, routers, and security devices. All provisioning and management is done through the Juniper Networks QFX3100 QFabric Director. The control plane between the QFX3008-I, QFX3600-I, and QFabric Director is established over a redundant out-of-band copper or control plane network using either redundant EX4200 switches with the QFX3000-M or EX4200 in a virtual chassis (VC) configuration with the QFX3000-G. The QFX3600-I Interconnect has two 1000BASE-T ports and two 1 GbE ports, either of which can be used to connect the QFabric Interconnect to the QFabric Director.

QFabric Nodes

The QFX3500 and QFX3600 are the QFabric Nodes that interconnect with servers, storage, and networking devices. Table 2-2 lists the specifications for these QFabric Nodes.

Table 2-2 QFX3500 and QFX3600 specifications

QFX3500	QFX3600
1U high fixed configuration	1U high fixed configuration
Four QSFP+ ports: <ul style="list-style-type: none">▶ Four 40 GbE uplink ports 48 SFP+/SFP ports: <ul style="list-style-type: none">▶ 36 1/10 GbE ports▶ 12 FC capable (2/4/8 Gbps) ports^a	Sixteen QSFP+ ports: <ul style="list-style-type: none">▶ Four 40 GbE uplink ports▶ 12 40 GbE ports or 48 10 GbE ports
Redundant AC or DC power supplies	Redundant AC or DC power supplies
Front to back air flow (AFI)	Front to back (AFI) or back to front (AFO) air flow (separate SKUs)

a. Can be configured as either FC ports or 10 GbE ports

QFX3500 is typically deployed as a QFabric access edge deployment to connect compute and FC/FCoE storage devices in the following data center environment:

- ▶ High-performance Layer 2 and Layer 3 access for enterprise, high performance computing (HPC), and cloud computing
- ▶ High-performance data center bridging (DCB), storage, and I/O convergence environments:
 - FCoE transit switch
 - FCoE-FC gateway
 - iSCSI transit switch
- ▶ Low-latency peering router between co-location exchange
- ▶ Layer 3 aggregation switch

The QFX3600 is an ideal platform for the following items:

- ▶ High-speed, low-latency, storage and I/O-converged networking solution
- ▶ 40 GbE aggregation switch deployments in a two-tier switching infrastructure
- ▶ High-performance Ethernet Layer 2 and Layer 3 access environments
- ▶ Standards-based FCoE transit switch

QFabric Nodes can run in different modes depending on the function needed, for example, routing and spanning tree activated, single node, or dual nodes in a virtual edge. QFabric Nodes can be as one of three node group types: network, server, and redundant server.

Network node group

A network node group is a grouping of QFabric Nodes that allows the connectivity of external network devices. A network node group supports external network devices that require the use of Layer 2 and Layer 3 protocols, for example xSTP, OSPF, BGP, and other such devices.

One network node group is used per QFabric network. Inside the network node group, all the QFabric Nodes, with a maximum of eight nodes, are seen as one logical entity, which simplifies network management.

Server node group

The server node group connects server and storage endpoints to the QFabric system. A server node group does not provide cross-node resiliency. Server node groups have the following characteristics:

- ▶ Support link aggregation group (LAG) between interfaces on the same QFabric Node to provide a redundant connection between the server and the QFabric system.
- ▶ Do not run network protocols, such as STP, PIM, and OSPF.
- ▶ Contain mechanisms, such as BPDU guard and Storm control, to detect and disable across-port loops.
- ▶ The local CPU in the QFabric Node in SNG performs routing engine and Packet Forwarding Engine (PFE) functions. The forwarding functions are local to the server node group, which is the default mode.

Redundant server node group

A redundant server node group provides the same QFabric functionality as a server node group; however, a redundant server node group provides resiliency by spanning redundant connections across two QFabric Nodes.

2.3.2 QFabric management plane component

The management plane function is under the control of the QFabric Director, which provides all management services for the QFabric architecture. QFabric management plane includes the following components:

- ▶ QFX3100 Director
- ▶ QFabric route engine

QFX3100 Director

QFX3100 Director is the management device within the QFabric management plane, for both the QFX3000-M and QFX3000-G systems. QFX3100 QFabric Director is always deployed in pairs for redundancy. It is the central routing engine of the QFabric. The QFabric Director provides control and management services to QFabric, acting as a central point to manage and control all network components as a single logical switch. This approach significantly reduces the operational costs typically associated with managing a data center network.

The QFX3100 is a 2U device featuring GbE ports to connect to the QFabric Interconnects and Nodes. The Director also interfaces with the network management ecosystem using standards-based protocols such as XML/NETCONF, SNMP, or command-line interface (CLI).

QFabric route engine

The route engine is a QFabric-specific processing entity that implements QFabric control plane functions, routing protocols, system management, and user access. Route engines reside in the QFabric Nodes and Directors and work together to provide Layer 2 learning of local hardware, the propagating of information and status, and the routing of data through the optimum path.

Because QFabric runs only Ethernet protocols at the QFabric Nodes, a simple transport protocol is used to move data upstream to the QFabric Interconnects and downstream to the QFabric Nodes. Routing within the fabric is based on a MAC reachability table. The MAC reachability table has an entry for each QFabric Node and enumerates all the possible paths

for reaching every QFabric Node in the fabric through the QFabric Interconnects. The QFabric Interconnect looks into only the fabric header and does not do a MAC address lookup.

QFX3100 Directors are connected directly to each other and also to the QFabric Interconnects and Nodes through redundant EX4200 switches.

2.3.3 QFabric control plane components

The control plane is used for auto-discovery of all devices, provisioning, and image upgrades for various elements in the system. All of these functions are fully automated and do not require user configuration. Control plane services are provided by the QFabric Director.

The QFabric control plane is an out-of-band network, built on EX4200 switches that interconnect QFabric Directors to the QFabric Interconnects and QFabric Nodes for carrying control traffic.

The EX4200-24T, EX4200-24F, and EX4200-48T are the control plane switches for the QFX3000-M and QFX3000-G systems. Table 2-3 lists the details for these switches.

Table 2-3 QFabric Control Plane Switches

EX4200-24T	EX4200-24F	EX4200-48T
<ul style="list-style-type: none"> ▶ 1U high fixed configuration ▶ Supports virtual chassis ▶ 24-port 10/100/1000BASE-T ▶ Redundant AC or DC power supplies 	<ul style="list-style-type: none"> ▶ 1U high fixed configuration ▶ Supports virtual chassis ▶ 24-port 10/100/1000BASE-X (SFP) ▶ Redundant AC or DC power supplies 	<ul style="list-style-type: none"> ▶ 1U high fixed configuration ▶ Supports virtual chassis ▶ 48-port 10/100/1000BASE-T ▶ Redundant AC or DC power supplies
Supports QFX3000-M system	Supports both QFX3000-M and QFX3000-G systems	Supports QFX3000-G system

A pair of control plane switches (EX4200) is deployed in the QFX3000-M system as shown in Figure 2-16. Two management interfaces from each QFabric Director, QFabric Interconnect, and QFabric Node are connected to a separate control plane switch to provide resiliency.

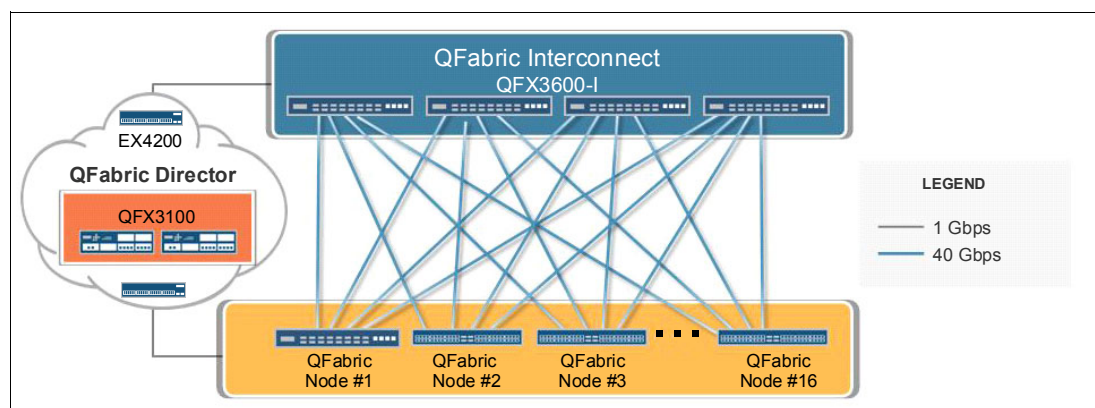


Figure 2-16 QFX3000-M system connectivity

In a maximum QFX3000-G system configuration with 128 QFabric Nodes and four QFabric Interconnects, two virtual chassis groups containing four EX4200 switches each are required (see Figure 2-17 on page 32). Two management interfaces from each QFabric Director, QFabric Interconnect, and QFabric Node are connected to a separate control plane switch in

each virtual chassis group to provide resiliency. The two virtual chassis groups connect to each other across a 10 Gb Ethernet LAG link to provide maximum resiliency for the QFabric control plane.

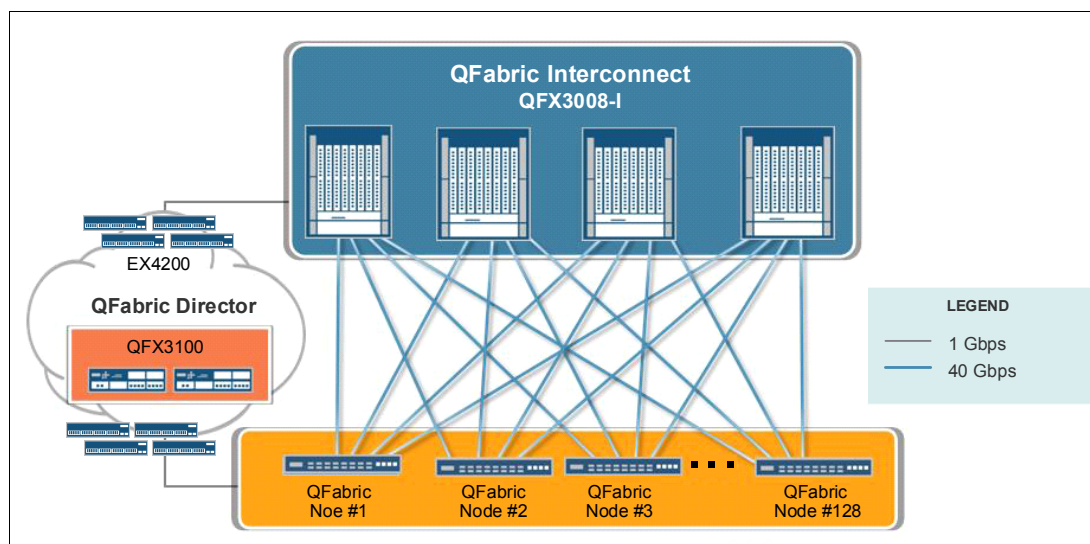


Figure 2-17 QFX3000-G system connectivity

More information about components in the Juniper Networks QFabric is at the following site:

<http://www.juniper.net/us/en/products-services/switching/qfx-series/qfabric-system/>

2.4 WAN domain: MX Series platform

The Juniper Networks MX Series edge routers can do both switching and routing. They run Junos operating system (Junos OS), enabling a wide range of LAN and WAN services. Each router provides full duplex, high-density Ethernet interfaces and high-capacity throughput for Layer 2 and Layer 3 data transport.

The MX Series family of edge routers can suit many applications:

- ▶ **MX240**
The MX240 router enables a wide range of small to medium business applications and services, including high-speed transport and VPN services, next-generation broadband multiplay services, and high-volume Internet data center Internet working.
- ▶ **MX480**
The MX480 is a dense and highly redundant platform to provide edge WAN services in a medium to large data center network. MX480 offers flexibility and reliability to support advanced services and applications. All MX Series Ethernet routers separate control and forwarding functions to provide maximum scale and intelligent service delivery capabilities.
- ▶ **MX960**
The MX960 router gives service providers the ability to offer the widest range of business and residential applications and services, including high-speed transport and VPN services, next-generation broadband multiplay services, and high-volume Internet data center Internet working.

For the purpose of this paper, we focus only on the MX480. The MX480 has the following specifications:

- ▶ 8U high
- ▶ 8-slot chassis
- ▶ Redundant Route Engines
- ▶ Six Dense Port Concentrator (DPCs) or Modular Port Concentrators (MPCs) per chassis
- ▶ System capacity: 1.92 Tbps
- ▶ Redundant AC or DC power supplies

The MX480 also includes the following key features:

- ▶ Link virtualization: VLAN, link aggregation group (LAG), generic routing encapsulation (GRE), and MPLS label-switched path (LSP)
- ▶ MPLS Layer 2 and Layer 3 VPNs
- ▶ Nonstop routing
- ▶ MPLS fast reroute
- ▶ VPLS multi-homing

For a list of Juniper MX Series models and their specifications, see the following site:

<http://www.juniper.net/us/en/products-services/routing/mx-series/>

2.5 Security domain: Security services gateways

The security domain in a data center design primarily consists of firewall, proxy, IDS, antivirus appliances, and security software that protects data center resources based on application and security requirements. Juniper's SRX and vGW provide defense-in-depth protection for application and provide critical security services within the security domain:

- ▶ SRX Series Services Gateway

High-performance security, routing and network solutions for enterprise and service providers data centers. SRX Series gateways offers high port-density, advanced security, and flexible connectivity, into a single, easily managed platform that supports fast, secure, and highly-available, data center and branch operations.

- ▶ vGW Series Virtual Gateway

A comprehensive security solution for virtualized data centers and clouds that can monitor and protect virtualized environments while maintaining the highest levels of VM host capacity and performance.

2.5.1 Juniper Networks SRX Series Services Gateway

Juniper Networks SRX Series Services Gateways offer security, routing, and network solutions for the security domain. The SRX Series Gateways provide a single platform where you can manage advanced security and connectivity for your data center and branch operations.

The SRX Series Service Gateway ranges from the smaller SRX100 model to the large enterprise SRX5800 model. All models are built for small, medium, and large enterprises and service providers. This section discusses two commonly deployed models, the SRX1400 and the SRX5600.

The SRX Series Services Gateway consolidates multiple security services and networking functions in a high availability (HA) appliance. This platform incorporates innovation that improves reliability, enhances network availability, and delivers deterministic performance of concurrent security services at scale. The SRX1400 Services Gateway solution consolidates multiple security services in one chassis under one integrated security policy.

The SRX1400 Services Gateway has the following specifications and key features:

- ▶ 3U high
- ▶ Redundant AC power supply
- ▶ Stateful inspection firewall: 10 Gbps
- ▶ IPS: 2 Gbps
- ▶ IPSec VPN: 2 Gbps
- ▶ Concurrent sessions: 0.5 million
- ▶ Connection establishment rate: 45,000 cps sustained
- ▶ Security policies: 40,000

The SRX 5600 Services Gateway has the following specifications and key features:

- ▶ 8U high
- ▶ Redundant AC or DC power supply
- ▶ Stateful inspection firewall: 100 Gbps
- ▶ IPS: 50 Gbps
- ▶ IPSec VPN: 75 Gbps
- ▶ Concurrent sessions: 60 million
- ▶ Connection establishment rate: 400,000 cps sustained
- ▶ Security policies: 80,000

The SRX Series Services Gateway offers a high performance, scalable, JUNOS-based dynamic services gateway with the following features:

- ▶ Network security
SRX provides VLAN, security zone, and virtual routers that allow enterprises and cloud providers to customize their security policies for different customers and tenants in their cloud infrastructure.
- ▶ Robust routing engine
The routing engine provides logical and physical segregation of data and control planes for flexible routing and security boundary design in data center and cloud environments.
- ▶ Threat protection and VPN services
SRX provides high-end stateful firewall features that can easily integrate into primary and disaster data center design. The high-speed interfaces and high through-put design provide robust packet inspection, network address translation (NAT), intrusion detection and prevention (IDS/IPS), IPSec VPN, GRE tunneling, flow based routing, and quality of service (QoS).

For more information about SRX Series Service Gateway, see the following site:

<http://www.juniper.net/us/en/products-services/security/srx-series/>

2.5.2 Juniper Networks vGW Series Virtual Gateway

Juniper Networks vGW Series Virtual Gateway provides a security solution for virtualized data centers and clouds. It allows you to monitor and protect virtualized environments while maintaining capacity and performance for the virtual machine (VM).

The vGW is a hypervisor-based security solution that includes the following features:

- ▶ Intrusion detection system (IDS)
- ▶ Antivirus
- ▶ Stateful firewall

The vGW engine is integrated into the hypervisor in the VMware kernel. It can allow or deny east-west traffic patterns in the same VMware vSphere ESXi server (such as, VM-to-VM traffic within an IBM Flex System compute node). It can also protect a VM from unauthorized external traffic. The vGW offers management scalability for virtualized data center and cloud security. Traffic patterns (east-west and north-south) are described in “Virtualized network traffic management” on page 52.

Figure 2-18 illustrates the logical relationship of vCenter, vGW Security Design, VM, vGW, hypervisor, and IBM Flex System compute node (System x®).

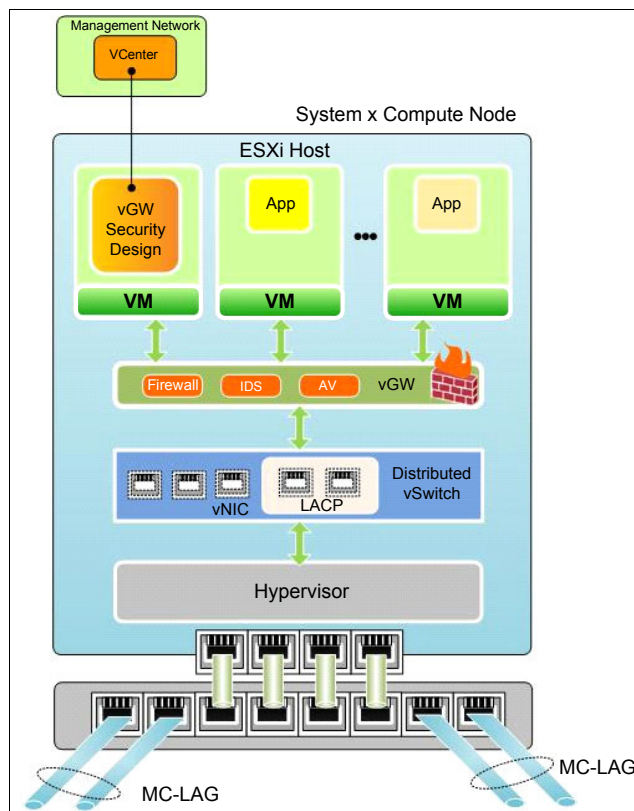


Figure 2-18 vGW installed in an IBM Flex System compute node

Note: In the figure, *vGW Security Design* indicates the management console of vGW.

The vGW can be installed on the compute node in the IBM Flex System. After vGW is installed, Security Design discovers all guest VMs through integration with VMware vCenter. vGW automates enforcement of security rules. vGW can also be integrated with SRX firewalls in zone synchronization policy management and defense-in-depth.

Detailed dynamic security policies, based on the VM guest information, can be created. These dynamic security policy will be assigned to individual VM or VM groups. If new VMs are created, they can automatically be associated with the known VM group when the predefined criteria are matched.

Consider the following information:

- ▶ The vGW stateful firewall controls have the following security policies configuration options:
 - Zones
 - VM groups
 - VMs
 - Applications
 - Ports
 - Protocols
 - Security states
- ▶ The integrated IDS will inspect packets for presence of malware or malicious traffic.
- ▶ vGW provides antivirus protection to deliver highly efficient on-demand and on-access scanning of VM disks and files with the ability to quarantine infected entities.

vGW VM Introspection

Virtualized environments have the potential to increase the risk profile for many types of security breaches, such as unauthorized connection, monitoring, not monitored application login attempts, malware propagation, and various “man in the middle” attacks.

VM Introspection is a hypervisor-based service that examines the internal state of a running VM. VM Introspection provides details about workloads and services that are installed on the VM and also its configuration. VM Introspection does not allow an existing VM to join a different VM group or cluster unless it has a specific OS configuration and hot-fix installed. VM Introspection prevents configuration errors from propagating within a smarter data center.

vGW Security Design keeps in-depth knowledge of the internal security state of each VM within the IBM Flex System. Administrators can use this information to build security policies by using the point-and-click policy editor. This process simplifies security management and ensures efficient enforcement of security policies pertaining to the VM security posture.

As depicted in Figure 2-19, when an existing VM is migrated from the Customer A network to the Customer B network, prior to vMotion executing, vGW validates the setting of the VM to ensure that the security posture is correct before it is allowed to migrate into the Customer B network.

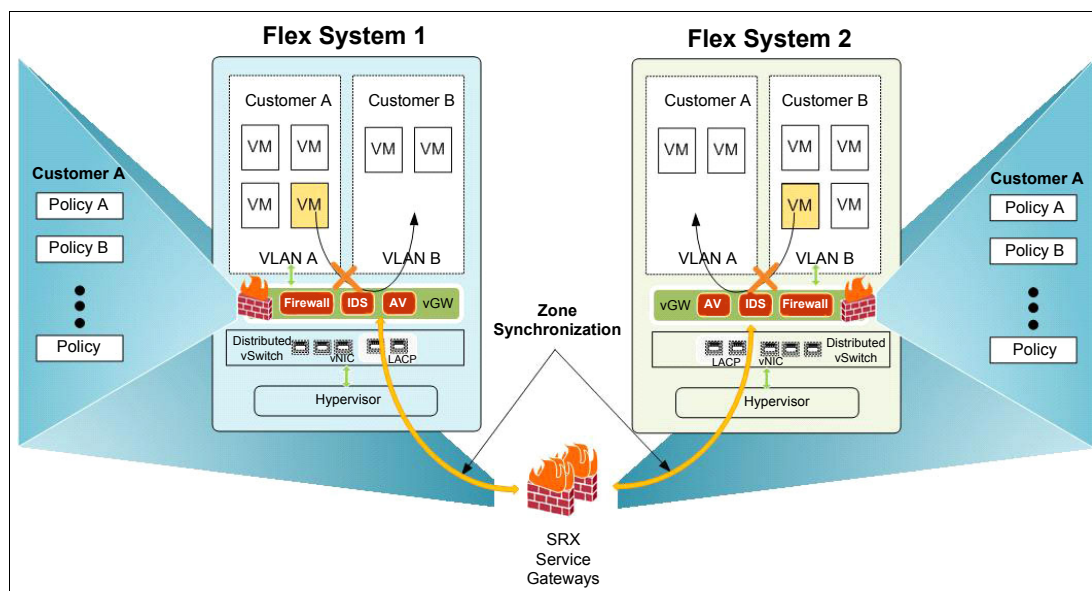


Figure 2-19 vGW IntroSpection prevents VM misconfiguration

vGW Antivirus

The vGW Antivirus adds another layer of defense against malware (such as viruses, worms, and spyware). The vGW Antivirus engine provides optional on-access and on-demand scanning so that administrators can either choose to scan files in real time or use the agentless approach.

The vGW Antivirus features minimize performance impact on the VM and virtual server in both on-access and on-demand modes. These features centralize the scanning on the vGW Security VM that is hosted on each VMware ESX/ESXi server, rather than executing the antivirus functions using thick clients on the VM itself. The vGW Endpoint on the VM passes the file (or in some cases, only the portion of the file that is necessary to determine whether it contains a virus) to the vGW Security VM across the virtualized network for examination whenever the VM accesses or attempts to transmit a file.

On-access antivirus

The vGW on-access scanning protects VMs against malicious content downloading or execution in real time. It detects malware or viruses on VMs, quarantining the infected files or infected guest VMs themselves. In a file-share environment, the vGW automatically scans files when the file is being saved to the VMs local storage. If the file is found to be infected, vGW quarantines and executes the embedded policy-based alerting mechanisms.

On-demand antivirus

This function can conduct full VM disk scans on either a periodic or sequential schedule to significantly diminish antivirus storms. The offline on-demand option scans guest VMs periodically, examining virtual disk files for malicious content. Because the antivirus feature does not need to be deployed on each VM for scanning, it can perform scans on virtual disk files from a centralized location. This method increases the engine's efficiency and allows it to execute scanning "from the outside" relative to the VM, which helps with the detection of attacks such as rootkits.

In this mode (see Figure 2-20), the Security VM mounts a snapshot of the virtual disk of the guest VM and passes it to the scan engine. This approach provides a scalable solution, limited only by the quantity of snapshots within the VMware vSphere cluster or available storage resources.

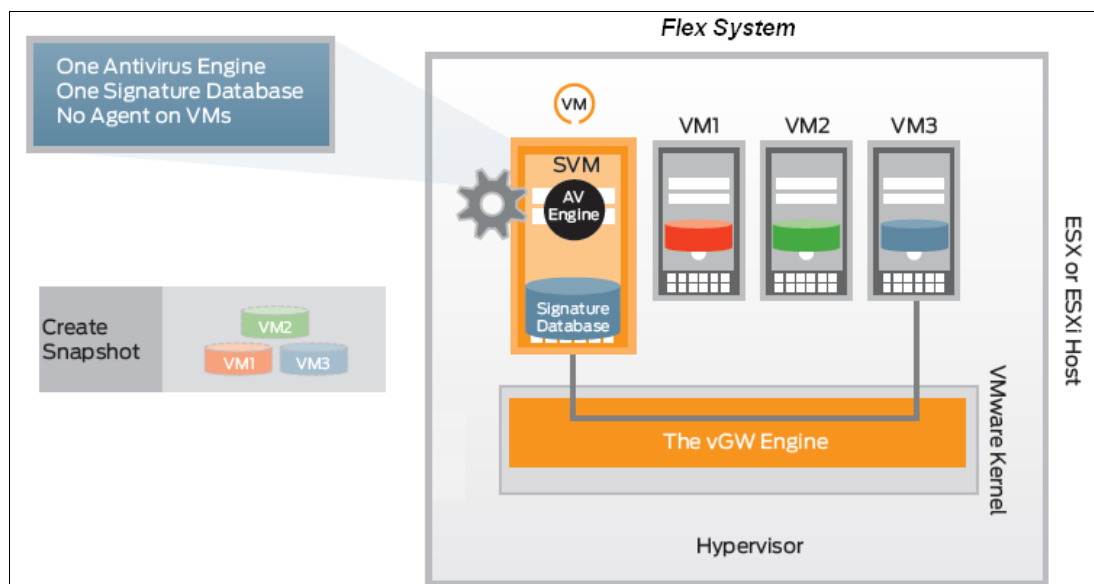


Figure 2-20 vGW antivirus on demand scanning

For more vGW series details, go to the following location:

<http://www.juniper.net/us/en/products-services/software/security/vgw-series/>

2.6 Management domain: Integrated tools

A smarter data center management domain consists of both the IBM Flex System Manager and Juniper Networks Junos Space:

- IBM Flex System Manager

IBM Flex System Manager provides a pre-integrated and virtualized management environment for servers, storage, and networking that is managed from a single interface and drives efficiency and cost savings in the data center.

- Juniper Networks Junos Space

This next-generation, open, secure, and scalable network management software suite includes a set of robust applications that can help manage network infrastructure by simplifying management functions and automating typical management tasks within Juniper Networks' extensive product portfolio.

Both management products can be integrated into existing enterprise IT infrastructure management frameworks, as depicted in Figure 2-21.

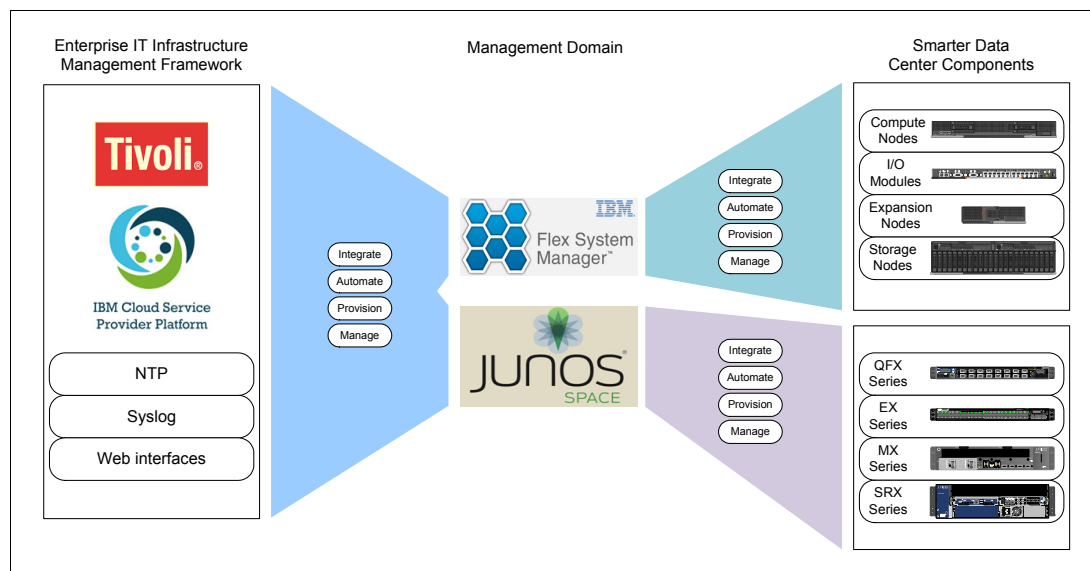


Figure 2-21 Management domain architecture

2.6.1 IBM Flex System Manager

IBM Flex System Manager is designed to help get the most out of IBM Flex System while automating repetitive tasks. IBM Flex System Manager can reduce the number of manual navigational steps for typical management tasks. From simplified system setup procedures with wizards and built-in expertise, to consolidated monitoring for all physical and virtual resources (compute, storage, and networking).

IBM Flex System Manager contains the following components to perform management functions:

- VMControl

Provides facilities for creating server system pools, which enable consolidation of resources and workloads into distinct and manageable groups. Deploy virtual machines and appliances into server system pools and group storage systems together by using storage system pools to increase resource utilization and automation. Manage storage system pools by adding storage, editing the storage system pool policy, and monitoring the health of the storage resources.

- Network Control

Discovers inventory and monitors network devices, launches vendor-specific applications for configuration of network devices, and views groups of network devices. Permits resource virtualization tasks such as assignment of Ethernet MAC and Fibre Channel WWN addresses.

- Storage Control

Creates, deploys, and clones images. Builds relationships between storage and server resources. Executes policy based placement and provisioning. Discovers, collects inventory, and monitors the health of storage devices.

- ▶ **Security Control**
Provisions users within profile groups, assigns roles to users using role-based access control (RBAC), and sets user permissions and access rights. Also provides a real-time view of active users within the IBM Flex System. All communication with the IBM Flex System Manager node require secure communication protocols to ensure secure and trusted connections.
- ▶ **IBM System Director Events**
Forwards events that originate from various physical and logical components such as IBM Flex System enterprise chassis or operating system agents.
- ▶ **Active Energy Manager (AEM)**
Provides simple real-time power and thermal management across compute node, networking and storage. Allows the configuration of thresholds on hardware components to permit proactive shutdown of unused resources through VMControl.

IBM Flex System Manager communicates to the following components:

- ▶ IBM Flex System components to be able to manage the hardware, to enable tasks such as live relocation of workloads
- ▶ Hypervisor element managers to coordinate provisioning tasks such as dynamically provisioning of virtualized server, storage and network resources
- ▶ External V7000 storage devices to more integrated management such as firmware updates, call home support or launching the element manager for provisioning from IBM Flex System Manager.
- ▶ Network services within existing enterprise IT management frameworks such as LDAP to provide centralized user management or to enable secure communications protocols.

Figure 2-22 depicts the IBM Flex System Manager operational model.

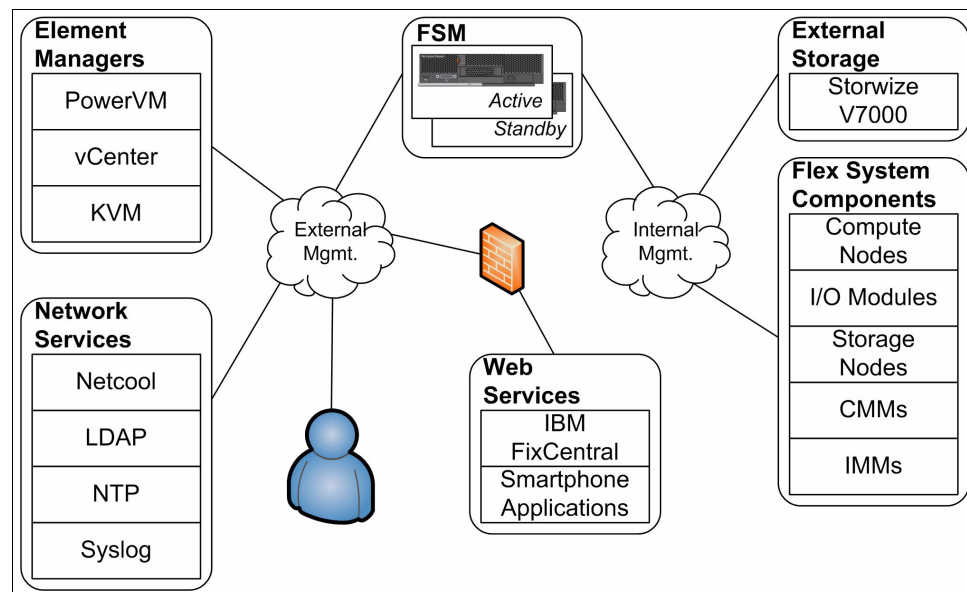


Figure 2-22 IBM Flex System Manager logical operational model

Figure 2-22 shows that the IBM Flex System Manager is accessed by system and network administrators through the external management network. Administrators can also access the IBM Flex System Manager using the IBM published smartphone application, IBM Mobile System Management.

Mobile System Management supports the following functions:

- ▶ Health and status monitoring
- ▶ Eventlog viewing
- ▶ Chassis map views
- ▶ Inventory management views

Download: IBM Mobile System Management application is available to download at no charge for Android, Apple iOS, or BlackBerry operating systems.

IBM Flex System Manager is a physical x240 compute node, which is preloaded with the IBM Flex System management software suite. The IBM Flex System Manager Node has the following fixed hardware specifications:

- ▶ One Intel Xeon Processor E5-2650 8C 2.0 GHz 20 MB Cache 1600 MHz 95 W
- ▶ 32 GB of memory with eight 4 GB PC3L-10600 CL9 ECC DDR3 1333 MHz LP RDIMMs
- ▶ Integrated LSI SAS2004 RAID controller
- ▶ Two IBM 200 GB SATA 1.8" MLC SSD configured in a RAID 1
- ▶ One IBM 1 TB 7.2 K 6 Gbps NL SATA 2.5" SFF HS HDD
- ▶ Dual-port 10 Gb Ethernet Emulex BladeEngine 3 (BE3) controller for data network connections
- ▶ Dual-port Broadcom 5718 controller for internal chassis management network connections
- ▶ Integrated Management Module II (IMM2)

Table 2-4 provides a list of typical IBM Flex System Manager management tasks and their capability within the IBM Flex System Manager of IBM Flex System supported hypervisors

Table 2-4 Supported IBM Flex System Manager management tasks based on hypervisor

Management tasks	IBM PowerVM®	VMware ESXi	Microsoft Hyper-V	Linux KVM
Deploy virtual servers	Yes	Yes	Yes	Yes
Deploy virtual farms	No	Yes	No	Yes
Relocate virtual servers	Yes	Yes	No	Yes
Import virtual appliance packages	Yes	No	No	Yes
Capture virtual servers	Yes	No	No	Yes
Capture workloads	Yes	No	No	Yes
Deploy virtual appliances	Yes	No	No	Yes
Deploy workloads	Yes	No	No	Yes
Deploy server system pools	Yes	No	No	Yes
Deploy storage system pools	Yes	No	No	Yes

IBM Flex System Manager provides the ability to install clients on the compute node operating systems (that is, the managed systems) to perform the central point of control for aggregating and managing discovered systems based on a service-oriented architecture.

IBM Flex System Manager currently includes the following types of client installable agents, each allowing different access and usage patterns:

- ▶ Agentless in-band
Managed systems without any IBM Flex System Manager software installed. IBM Flex System Manager communicates with the managed system through the operating system.
- ▶ Agentless out-of-band
Managed systems without any IBM Flex System Manager software installed. IBM Flex System Manager communicates with the managed system through something other than the operating system, such as a service processor or a hardware management console.
- ▶ Platform Agent
Managed systems with Platform Agent installed. IBM Flex System Manager communicates with the managed system through the Platform Agent.
- ▶ Common Agent
Managed systems with Common Agent installed. IBM Flex System Manager communicates with the managed system through the Common Agent.

For more details about IBM Flex System Manager, see the following site:

<http://www.ibm.com/systems/flex/systems-management/index.html>

2.6.2 Junos Space

Junos Space is an appliance-based network management platform. It provides network operations tools for device discovery and management, topology visualization, device image deployment, job management, audit logging, user administration, and system administration. Junos Space includes the following applications:

- ▶ Ethernet design
Automates the configuration, visualization, monitoring, and administration of large networks by providing a wizard-based interface for realtime device discovery and allows operators to have a broad, topological view of the network including all endpoint devices, link information, bottlenecks and failures, and discovered relationships between network elements.
- ▶ Security design
Permits application identification control with AppSecure, and also firewall, intrusion prevention system (IPS), Network Address Translation (NAT), and VPN security policy management. Also, provides quick and intuitive management of all phases of security policy lifecycle through one centralized web-based interface and eases administration with granular control over global, group, and device level firewall policies.
- ▶ Service Now
Provides automatic detection of problem signatures and collecting relevant diagnostic information at the time the problem occurs. Identifies problem signatures on the devices in real time, and sometimes detects them before the problem actually impacts system operations, creating an early-warning system that keeps a constant vigil for potential and acute problems on enterprise network.
- ▶ Service Insight
Streamlines network operations by automating inventory management and fault management tasks, and collects periodic health and configuration data from network elements. This data is used by Service Insight to deliver targeted bug notifications, identify

which network devices could potentially be impacted, and to perform impact analyses for end of life (EOL) and end of service (EOS) notifications.

- ▶ **Network Activate**

Provides an automated and streamlined MPLS Ethernet provisioning process. Also automates the design, activation and validation of tasks enabling efficient and cost-effective deployment of MPLS Ethernet solutions for L2VPN and L3VPN services.

- Automated network discovery
- MPLS resource management
- Service provisioning, validation, and troubleshooting

- ▶ **Junos Space software development kit (SDK)**

Through the SDK, Junos Space allows for development of open software packages to allow integration with tools that are ready for use. The SDK uses the Junos Space API.

Junos Space is offered as either a hardware appliance (JA1500) or a virtual appliance (VA). Both the JA1500 and VA deliver Junos Space as an appliance, preinstalled to enable first time QFabric configuration by the network administrator. The key differences are as follows:

- ▶ JA1500 is a physical rack-mountable server
- ▶ VA is a VMware template containing a Junos Space image for deployment onto an existing VMware vSphere 4.x or 5.x environment

The JA1500 has the following fixed hardware specifications:

- ▶ One 2U rack-mountable compute chassis
- ▶ Three 1 TB hard disks in hot-swappable RAID 5 array
- ▶ One RAID controller
- ▶ Four RJ-45 10/100/1000 Gigabit Ethernet ports
- ▶ One RJ-45 Serial console port and One USB interface
- ▶ Optional single IOC slot available for I/O card expansion
- ▶ Option for dual-redundant, hot-swappable power supply
- ▶ Redundant cooling fans

Figure 2-23 depicts the front view of the JA1500 appliance.



Figure 2-23 Junos Space JA1500 Appliance

Junos Space communicates to the following components:

- ▶ Network components to manage WAN domain components, such as Edge routers, Service Gateways, and so forth
- ▶ QFabric components to coordinate the tasks for provisioning, monitoring, and diagnosing
- ▶ Network Services within existing enterprise IT management frameworks such as NTP to synchronize time across all fabrics or LDAP to enable authentication

Network and system administrators can communicate to Junos Space by using the following methods:

- External management network to enable proactive maintenance of high-speed data center fabrics
- Development services to develop new application interfaces within Junos Space using the published SDK.

Figure 2-24 depicts a logical operational model of Junos Space.

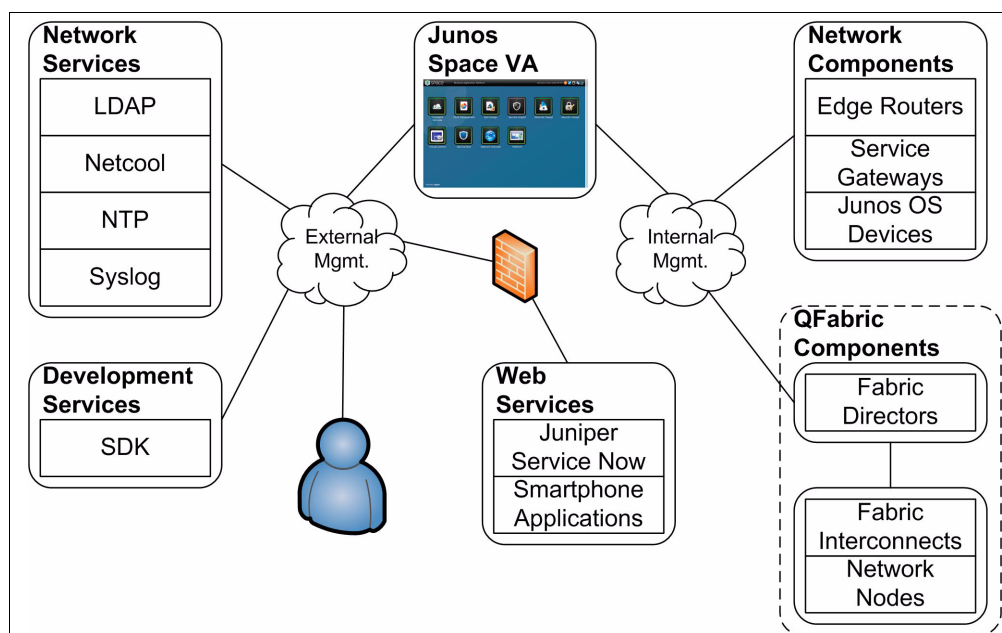


Figure 2-24 Junos Space logical operational model

Junos Space Virtual Control

Junos Space Virtual Control enables administrators to manage, monitor, and control the virtual network environments that run within virtualized servers. Built on the Junos Space Network Management platform, it provides end-to-end visibility and control over both the virtual and physical networks from a single point. Junos Space Virtual Control includes the following features:

- Automated orchestration to minimize errors
The Virtual-to-Physical Mapper and P+V Dynamic Orchestrator features ensure that network administrators can keep their physical and virtual networks in proper alignment, eliminating errors that occur when the network systems are out of sync.
- Simplified data center manageability
Junos Space Virtual Control allows network administrators to discover, manage, and monitor virtual and physical networks. With the VM Locator feature, administrators can locate VMs in the network infrastructure and disable network access when required. Network administrators can also use vSwitch profiles to group common network settings to quickly and easily deploy multiple services in the data center, with the appropriate network policies.

For more information about Junos Space, go to the following location:

<http://www.juniper.net/us/en/products-services/network-management/junos-space-platform/#overview>

2.6.3 IBM Tivoli

IBM Tivoli software provides an intelligent infrastructure management solution that helps network and system administrators understand and proactively manage data center infrastructures.

An overview of these elements and their associated functions is as follows:

► IBM Netcool® Network Manager

Centralizes visibility, control and automation of multivendor network infrastructures. Improves operational efficiency by allowing network administrators to implement change using predefined templates to reduce errors and execute compliance reporting. Key features are as follows:

- Offers policy-based root-cause analysis to reduce the down-time
- Centralizes actionable events for proactive fix before the network goes down
- Provides proactive configuration compliance through command filtering, highlighting human error in deployment of new configurations before they are deployed

► Tivoli Provisioning Manager

Offers provisioning, configuration and maintenance of physical servers and virtual servers, operating systems, middleware, applications, storage and network devices. Key functional areas are as follows:

- Virtualization and image management, including image library and image versioning
- Workload migration using Live mobility of AIX Lpars
- Scalability, including the ability to provision thousands of virtual machines simultaneously
- Provisioning and configuration management of Juniper EX Series, MX Series, SRX Series, and QFX Series platforms
- Regulatory compliance reporting including SOX

► IBM Tivoli OMNIBus

Enables real-time consolidation of enterprise-wide events. Delivers real-time, centralized monitoring of complex networks and IT domains (server, network, security, storage, and other IT management). Bridges the gap across multiple networks and IT silos to help organizations improve the end-to-end availability of their applications and services.

Functions are as follows:

- Fault detection
- Collect cross-domain events in real-time
- Eliminate duplicate events by policy-based filtering
- Embedded problem escalation engine
- Automated policy-based root cause analysis (RCA)
- Automated event resolution

Figure 2-25 depicts a logical operational model of Tivoli Systems management.

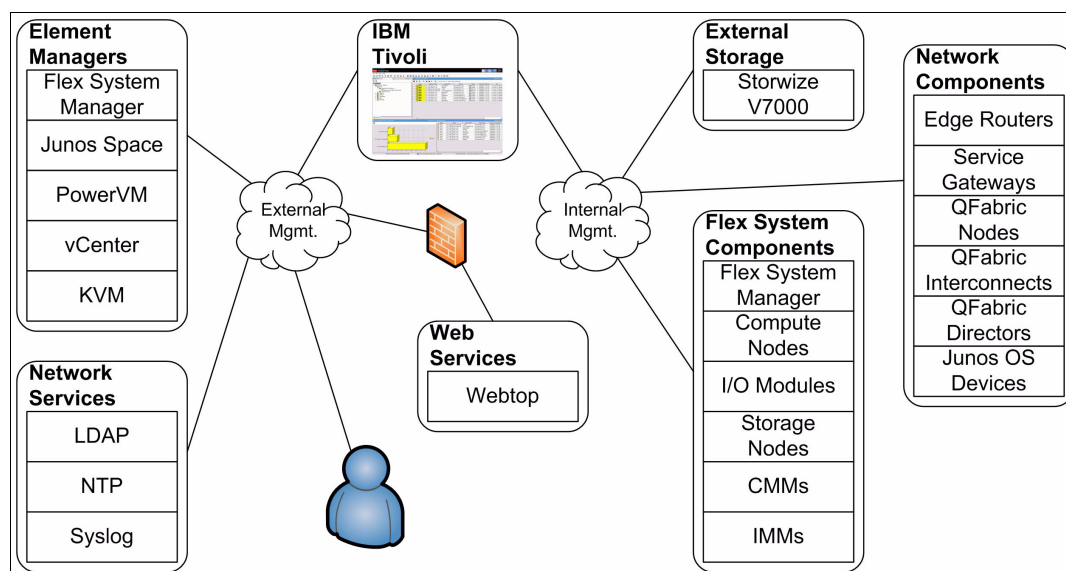


Figure 2-25 Tivoli logical operation model

The various Tivoli management products communicate to the following data center components:

- ▶ Element managers to manage system pool, storage pools or to monitor the resource utilization and to coordinate provisioning tasks such as dynamically provisioning of virtualized server, storage and network resources
- ▶ External V7000 storage devices for more integrated management functions such as firm ware updates or call home support.
- ▶ Network services within existing enterprise IT management frameworks such as syslog to enable centralized management and reporting of devices which support external system logging

Network and system administrators can connect to the various Tivoli components by using the following components:

- ▶ The Tivoli GUI
- ▶ Webtop is a web-based application that processes network events and presents the event data in various graphical formats. It allows administrators remote control of Tivoli products.

For more information about IBM Tivoli suite of products, see the following site:

<http://www.ibm.biz/Bdx2Bg>



Designing a smarter data center infrastructure

When designing a smarter data center infrastructure, standard design methods and common best practices should be applied to all the components that make up the technology domains. Requirements for high availability, scalability, and performance are typically satisfied through a modular design, redundant components (hardware and software), and proven operational procedures. Both IBM Flex System BTO and QFabric allow for data center designs that eliminate single points of failure through all the layers of the OSI model and provide options that incorporate low latency to boost workload performance. In addition, IBM and Juniper Networks management tools can ensure services and workloads are not disrupted in case of planned or unplanned outages.

This chapter outlines design and planning considerations when integrating the building blocks (introduced in Chapter 2, “Understanding the key building blocks for a smarter data center” on page 7). It describes how to use the technologies within IBM Flex System BTO and Juniper Networks QFabric to mitigate the common constraints of the traditional data center infrastructure. The following main areas are covered:

- ▶ Virtualization
- ▶ Security
- ▶ Provisioning
- ▶ Management
- ▶ Automation

3.1 Virtualization

This section provides an overview of the challenges in virtualized server deployments within both virtualized and physical networking environments. It presents key design considerations for hypervisor-based virtual switches and distributed virtual switches. It also describes the QFabric node groups and includes a discussion about integrated solutions to consider for virtual networking in a security zone or multitenant environment when designing a smarter data center network.

3.1.1 Hypervisor-based virtualized networking

In the past, traditional data center networking environments relied on physical and dedicated switching and routing infrastructures to connect compute workloads. In that model, physical server network interface cards (NICs) and Fibre Channel host bus adapters (HBAs) connected directly to physical switches to use data center networking services.

Today's data centers also must consider the following virtualized technologies and solutions in their networking environments:

- ▶ Virtualized switching
- ▶ Network convergence
- ▶ Virtualized routing
- ▶ Traffic shaping and quality of service
- ▶ Intrusion detection and prevention
- ▶ Centralized security policy enforcement

When assessing these technologies for implementation into a smarter data center, it is important to consider the workloads and traffic patterns, location of dependent workload services, and security requirements. A solution that combines these technologies is possible but might not fit all workloads in a smarter data center.

Virtual switches

Hypervisors are designed to optimize the efficiency and experience of the platform that they support. Therefore, many different virtual switches are available depending on the hypervisor.

Two types of virtual switches are currently available:

- ▶ Virtual switch

A virtual switch is limited to the hypervisor of one physical server to allow the aggregation of physical interfaces to virtualize traffic switching amongst collocated virtual machines

- ▶ Distributed virtual switch

A distributed virtual switch spans multiple physical servers to centralize management of virtualized switching. This technology is the foundation of VM mobility functionality in the smarter data center.

The following types of virtual switches and associated hypervisor platforms are available:

- ▶ Microsoft Hyper-V virtual switch
A virtual switch is localized to one Hyper-V x86 server. Consider using for Hyper-V management traffic.
- ▶ Xen server virtual switch
A virtual switch is localized to the Xen x86 server to allow bridging of the physical NICs and paravirtualized network interfaces of the virtual machines (VMs). This bridge connects the VMs to the data center networking services that are external to the Xen server.
- ▶ KVM virtual switch
A virtual switch is localized to the KVM x86 host system to allow bridging of the physical NICs and VM virtual network interfaces (vNICs). This bridge connects the VMs to the data center networking services that are external to the KVM host system.
- ▶ IBM Virtual I/O Server (VIOS) VLAN compatible virtual switch
A virtual switch is localized to one POWER System that is used for all IBM VIOS virtualized network communications.
- ▶ VMware vSphere virtual switch (vSwitch)
A virtual switch is localized to one VMware x86 server. Consider using for VMkernel management traffic.

The following types of distributed virtual switches and associated hypervisor platforms are currently available:

- ▶ Microsoft Hyper-V distributed virtual switch
A distributed virtual switch spans across many Hyper-V x86 servers and is available natively in Hyper-V 3.x. Consider using the distributed virtual switch for all data plane traffic.
- ▶ VMware vSphere distributed virtual switch (dvSwitch)
A dvSwitch spans across many VMware x86 servers and is available natively in VMware vSphere 4.x and 5.x or as a distributed virtual switch appliance. Consider the use of the appliance for all data plane traffic in data center networks that require centralized management capabilities.
- ▶ IBM Distributed Virtual Switch 5000V (DVS 5000V)
The DVS 5000V is an advanced, feature-rich distributed virtual switch for VMware environments with policy-based VM connectivity. The DVS 5000V operates as a VM within the vSphere cluster.

See “Distributed Virtual Switch 5000V capabilities” on page 21 for more information about this switch.

Converged networking

The term *converged networking* refers to the ability to carry both user data (LAN) and storage data (SAN) network traffic over the same unified fabric. Implementing converged networking helps reduce hardware, electrical power, cooling, and management costs by simplifying data center infrastructure and reducing the number of ports (on both server adapters and switches), adapters, switches, and the cabling that is required to build such an infrastructure.

Converged networks require the following key components:

- ▶ Converged network adapters (CNAs)
- ▶ Converged switches

CNAs provide host connections with support for both LAN and SAN data transfer protocols. Converged switches manage flows for these types of traffic by forwarding packets, ensuring reliable in-order delivery, guaranteeing service-level agreements (SLAs) for certain types of traffic with quality of service (QoS), and controlling congestion, among other services.

The 10 Gb Converged Enhanced Ethernet (CEE), Edge Virtual Bridge (EVB) or Data Center Bridging (DCB) technologies are widely adopted and considered an affordable and convenient way to build a converged fabric. CEE supports both Ethernet and Fibre Channel over Ethernet (FCoE) protocols to connect to standard LANs and SANs.

CEE, EVB and DCB are built using several standards defined within the IEEE 802.1Q Virtual LANs standard as follows:

- ▶ Edge Virtual Bridging (802.1Qbg)
- ▶ Congestion Notification (802.1Qau)
- ▶ Enhanced Transmission Selection (802.1Qaz)
- ▶ Priority-based Flow Control (802.1Qbb)
- ▶ MAC Control Frame for Priority-based Flow Control (802.3bd)
- ▶ Bridge Port Extension (802.1BR)

See “Virtualized network traffic management” on page 52 for more information about the functions within each standard.

Figure 3-1 shows a converged network architecture. In this configuration, data center solutions use converged networking designs to achieve a simpler, more efficient infrastructure.

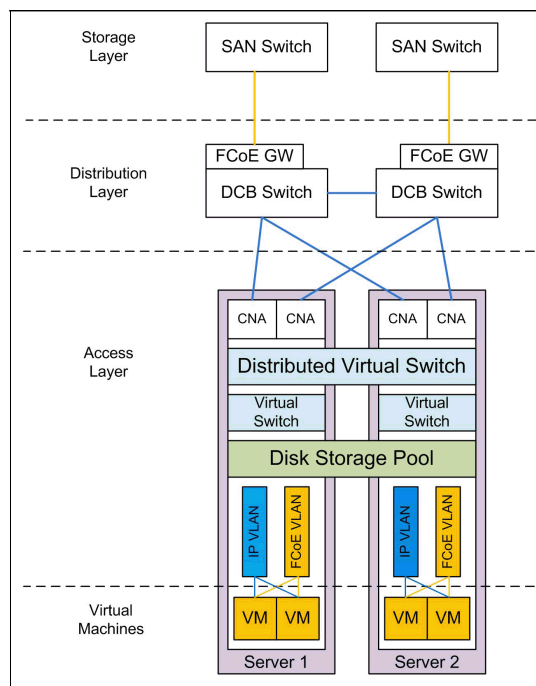


Figure 3-1 Converged networking design

The IBM Flex System EN4091 10 Gb Ethernet pass-through or IBM Flex System Fabric CN4093 10 Gb Converged Scalable Switch connected to an external QFX Series switch

(QFX3500 or QFX3600), along with the following compute node I/O adapters are well suited for the converged networking design:

- ▶ IBM Flex System CN4054 10 Gb Converged Adapter (4-port)
- ▶ IBM Flex System CN4058 10 Gb Converged Adapter (8-port)

For more information about FCoE and FCoCEE, see *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493.

Traditional networking

Figure 3-2 shows the architecture of a traditional virtualized networking design. In this design, the NIC and HBA traffic are physically separated at both the access layer (that is, virtual switches or distributed virtual switches) and the distribution layer (that is, physical network switches). This approach is still valid for solutions that do not require network convergence.

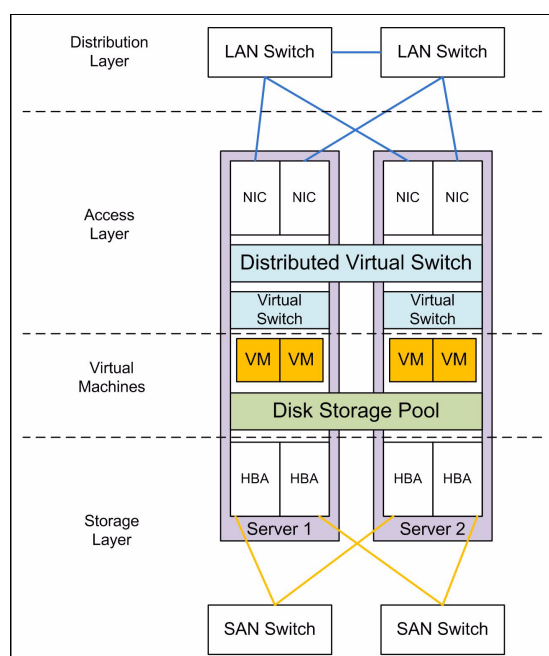


Figure 3-2 Traditional networking design

The following switches and adapters are well suited for the traditional networking design:

- ▶ IBM Flex System EN4091 10 Gb Ethernet pass-through or IBM Flex System Fabric EN4093 10 Gb Ethernet Scalable Switch connected to the external QFabric QFX Series switch (QFX3500 or QFX3600) and the following compute node I/O adapters:
 - IBM Flex System EN4132 10 Gb Ethernet Adapter (2-port)
 - IBM Flex System CN4054 10 Gb Converged Adapter (4-port)
 - IBM Flex System CN4058 10 Gb Converged Adapter (8-port)
- ▶ IBM Flex System FC3171 8 Gb SAN pass-through connected to an external SAN switch or IBM Flex System FC3171 8 Gb SAN Scalable Switch and the following compute node I/O adapters:
 - IBM Flex System EN3172 8 Gb FC Adapter (2-port)
 - IBM Flex System EN3052 8 Gb FC Adapter (2-port)
- ▶ IBM Flex System FC5022 16 Gb SAN Scalable Switch and the following compute node I/O adapter:
 - IBM Flex System EN5022 16 Gb FC Adapter (2-port)

For more information, see the following sections:

- ▶ “Compute node I/O adapters” on page 14
- ▶ “I/O modules” on page 15, “Virtual fabric networking” on page 17
- ▶ “QFabric Nodes” on page 29 for more information.

Virtualized network traffic management

Management of traffic in traditional, non-virtualized data center network topologies was, in the past, limited to north and south bound connections. This traffic flow is because of switching, routing, compute, and storage devices being designated and connected to each-other in measured, carefully planned configurations. Converged networking topologies within the smarter data center changed this model, allowing the collapse of the physical boundaries into logical domains, or building blocks. The collapse of these boundaries, hosting virtualized switches in the compute domain, has created a new form of traffic flow known as *east-west traffic*. East-west traffic can potentially expose VMs to unwanted connections or interference at all levels of the open systems interconnection (OSI) model (that is, from security violations to broadcast traffic).

When designing a smarter data center, consider the following solutions to control east-west traffic:

- ▶ Private virtual local area networks (PVLAN)

Distributed virtual switches provide PVLAN support. This approach permits VM vNICs to be configured in either a community with other VMs to allow east-west traffic or segregated, isolating the VM vNIC and eliminating east-west traffic.

- ▶ Access control lists (ACLs)

ACLs are available in most distributed virtual switches that support Edge Virtual Bridging (EVB). EVB is an IEEE standard (802.1Qbg) that involves the interaction between the virtual switching environments in a hypervisor and the first layer of the physical switching infrastructure. ACLs can be configured for *priority*, *deny*, and *permit* at both OSI Layer 2 (MAC address) and Layer 4 (TCP/UDP).

- ▶ Virtual Ethernet port aggregator (VEPA)

EVB standards include the VEPA option, which forces all to be sent out of the physical NIC and to be handled by the adjacent switch. This approach is called *north-south traffic* and includes local VM to VM traffic on the same physical server. This method is particularly useful for solutions where enforcing data center security policies on the VM is required. When using VEPA, consider the scalability and physical location of the adjacent switch to ensure a sustainable, low-latency data center network design.

- ▶ Virtual security gateway

When considering using virtual security gateway, consider hypervisor-based security, configuration management, and compliance solutions. Juniper Networks vGW Series installs into the hypervisor and has complete visibility of east-west traffic. You can configure the vGW engine with smart groups to isolate traffic to groups of vNICs or to isolate single vNIC traffic, thus securing the VM.

When designing for a converged networking topology, consider the FCoE virtual local area network (VLAN). This design can help to ensure an appropriate QoS based on your specific requirements. You can implement QoS in the following ways:

- ▶ CEE priority-based flow control (PFC)
Eliminate congestion-related frame loss by using an 802.3x PAUSE-like mechanism for individual user priorities.
- ▶ CEE enhanced transmission selection (ETS)
Share bandwidth among different traffic classes more efficiently. When a particular traffic flow does not use all of the bandwidth available to it according to the traffic classification, the unused bandwidth can be allocated to other traffic flows.

3.1.2 QFabric node groups

You can configure Juniper Networks QFabric QFX Series switches with separate operational modes to allow intelligent integration and management of workloads that are virtualized within the compute and storage domain. These operational modes are called *node groups*, which allow grouping multiple QFabric nodes into a single logical entity.

QFabric includes the following types of node groups:

- ▶ Network node groups
One or more QFX Series switches are grouped together to provide redundant, scalable connections to an external QFabric routing engine (running on the QFabric directors) services to support routing between QFabric and external networks.
- ▶ Server node groups
One of more QFX Series switches are grouped together to provide redundant, scalable connections to server and storage devices that are hosting workloads. Key to server node groups is the local virtual routing instances that support routing within the group.
The server node groups also include the following types of node groups:
 - Server node group, which contains one QFX Series node
 - Redundant server node group, which contains two QFX Series nodes in a redundant configuration

Network node groups

A key consideration in the design and implementation of QFabric network node groups is the QFabric topology within the smarter data center. Although support for connectivity to server and storage devices is available, a network node group is typically used to connect edge services, such as wide area network (WAN) routers, security gateways, load balancers (local and global traffic managers), and other networking devices. To allow for a sustainable and resilient smarter data center network, ensure the design of the network node group is redundant in both the node device and the interface connection.

Only one network node group supports up to eight QFX Series node devices. As more nodes are added to the group, scalability and redundancy is enhanced within the network node group because of the non-blocking, high speed, and low latency of the QFabric architecture.

Figure 3-3 depicts the architecture of QFabric network node groups and the location of the QFabric routing engine, which routes to external networks.

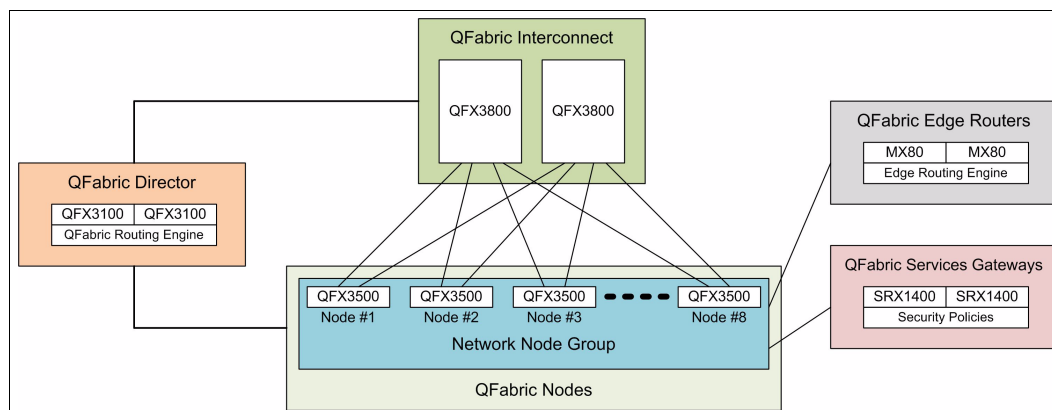


Figure 3-3 The QFabric network node groups

Server node groups

You can configure server node groups in the following redundant and non-redundant configurations:

- Server node group

The configuration is a QFabric node that supports connectivity to server and storage devices, which is contained within one QFX Series switch

- Redundant server node group

The configuration is the same as the server node group but supports two QFX Series nodes in a redundant pair

Consider implementing a redundant server node group in solutions that require QFX node level redundancy. A redundant server node group supports the configuration of two QFX Series switches, which work in tandem to provide redundant connections to workloads that are hosted in the compute and storage domains. You configure these redundant connections as link aggregation groups (LAGs) and span both nodes in the redundant server node group.

See 2.3.1, “QFabric data plane components” on page 27 for more information about using server node groups with QFabric components.

The QFabric node in a redundant server node group must be of the same type, either a QFX3500 or a QFX3600.

Figure 3-4 depicts the architecture of QFabric redundant server node groups and LAG connections to the interfaces of IBM Flex System.

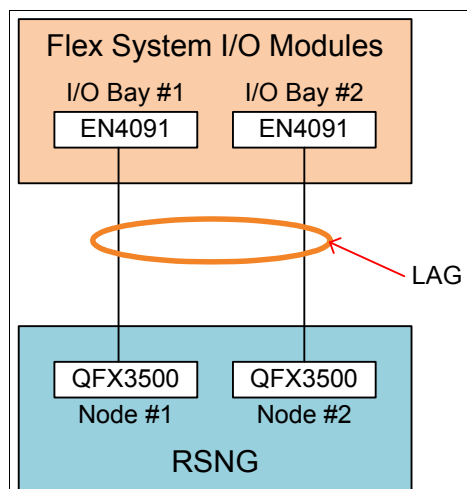


Figure 3-4 QFabric redundant server node group

QFabric uplinks

QFabric server and redundant server node groups support the connection of virtualized server and storage devices using LAG. QFabric node groups support both local node LAG and spanned node LAG.

LAG includes the following options:

- ▶ Server node:
 - 48 LAGs per server node
 - 8 members per LAG (interfaces)
- ▶ Redundant sever node group:
 - 48 LAGs per server node
 - LAGs can be spawned across QF nodes
 - Links are in active/active state
 - 8 members per LAG (interfaces)
- ▶ Network node group:
 - 128 LAGs per network node group
 - LAGs can be spawned across QF nodes
 - 8 members per LAG (interfaces)

When designing uplinks between QFabric nodes and IBM Flex System, consider LAG as a necessary design attribute to support both link redundancy (either on the same QFabric node or between nodes) and bandwidth to support sustained workload traffic.

Connection: Redundancy through cross connectivity is built into the connections between the IBM Flex System compute node I/O adapters and chassis I/O modules (see Figure 2-7 on page 14). Therefore, the uplinks between the IBM Flex System chassis I/O modules and the QFabric nodes (QFX3500 or a QFX3600) should connect only as depicted in Figure 3-4 on page 55.

3.1.3 Providing an integrated solution with IBM Flex System BTO and QFabric

Figure 3-5 depicts an example of an integrated IBM Flex System and QFabric solution for a smarter data center. The solution is a single chassis environment that includes the following compute and QFabric components:

- ▶ One IBM Flex System BTO enterprise chassis
- ▶ Two QFX3500 switches, as top-of-rack (TOR)
- ▶ Four IBM Flex System x440 compute nodes
- ▶ Eight IBM Flex System CN4054 I/O adapters
- ▶ Four IBM Flex System EN4091 pass-through I/O modules
- ▶ Distributed virtual switch

This solution considers the effect of east-west traffic within a single IBM Flex System BTO enterprise chassis and incorporates distributed virtual switch to make the solution VM-aware. The distributed virtual switch allows the solution to satisfy the following requirements:

- ▶ VM traffic separation
- ▶ Centralized management of the following items:
 - Access control lists (ACLs)
 - QoS policies
 - Security policies
 - Network configuration profiles
 - Switch management
 - Network monitoring

See “Virtual switches” on page 48 for more information about the distributed virtual switches that can be used for an integrated environment.

Figure 3-5 shows the traffic flow inside the IBM Flex System BTO through a distributed virtual switch and outside the IBM Flex System through QFX3500 TOR switches.

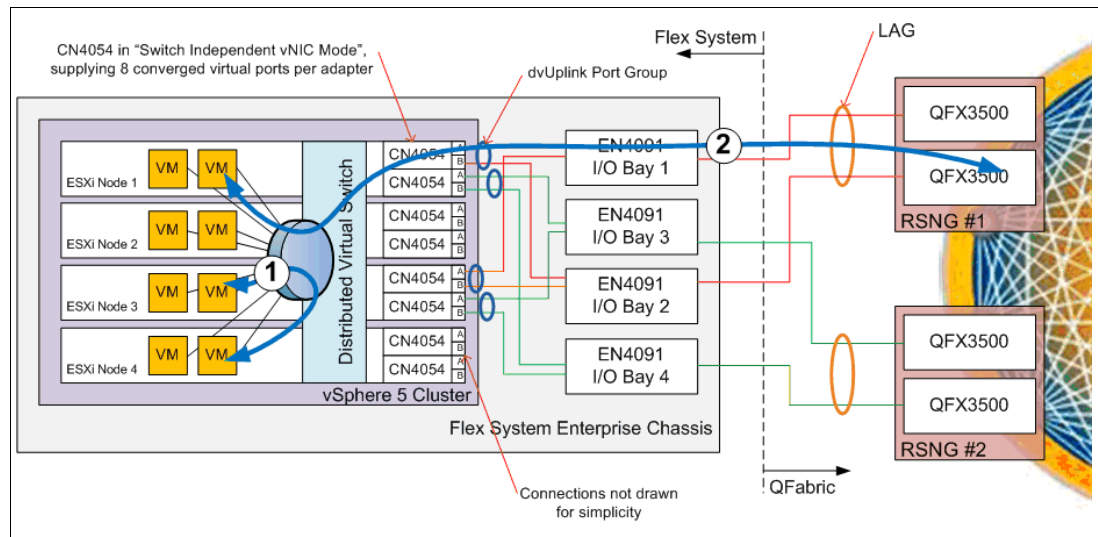


Figure 3-5 Distributed virtual switch logical operational model

Virtualized workloads in today's data center environments typically rely on network services outside of the virtual server cluster they reside. This north-south traffic requires the data center network to provide routing protocols. A design consideration of QFabric virtual router routing instances is that the physical interfaces within the virtual servers can be a member of only one virtual router instance. Currently, the QFX3500 and QFX3600 TOR switches supports as many as 256 virtual router routing instances.

In an IBM Flex System multi-chassis configuration, the enterprise chassis are joined by using redundant connections to the QFX3500 or QFX3600 TOR switches. These physical connections can be of many types of topologies, depending completely on the IBM Flex System I/O modules that are installed in the enterprise chassis.

The solution depicted in Figure 3-5 on page 56 can also be used to represent a multi-chassis configuration, with the difference being additional enterprise chassis. When designing virtual server clusters that consist of eight or more physical servers, consider placing an equal quantity of IBM Flex System compute nodes in each enterprise chassis.

3.2 Security

Stringent security design and considerations are critical to data center planning, especially in today's large and complex virtualized smarter data center environments. Today's data center environments commonly encounter the following internal and external security threats:

- ▶ Improper access to data because of lack of security
- ▶ Undetected and uncontrollable malware outbreaks or insider attacks in the virtual environment
- ▶ Lack of virtual network visibility, including east-west traffic between VMs
- ▶ Inability to enforce policies that isolate VMs, prevent VM sprawl, or secure VM mobility functions
- ▶ Increased network complexity and administrative burden caused by applying established VLAN or firewall technology to the virtual environment
- ▶ Computer virus outbreak
- ▶ Trojan implant
- ▶ Distributed denial of service (DDoS) attacks

Security involves key design considerations of hypervisor-based security policy management, centralized anti-virus infrastructures, VM configuration compliance, enforcement of centralized security policies on virtualized workloads, and intrusion detection services. This section includes design considerations for the following solutions in a smarter data center:

- ▶ Stateful firewall, intrusion detection and antivirus solutions using Juniper Networks vGW Series solutions
- ▶ Centralized security policy enforcement by integrating Juniper Networks SRX and vGW solutions
- ▶ Secure remote access ensuring only authorized administrators have access to components of the smarter data center

3.2.1 Stateful firewall, intrusion detection, and antivirus

To protect VMs in today's data center environments requires software-based firewall, IDS and AV agents that are executing within the operating system (OS) of each server in the data center. Each agent typically requires connection to the Internet or centralized policy orchestrators to synchronize policy and antivirus definition updates. This approach requires compute resources to be used on each server, greatly reducing efficiency.

Consider implementing Juniper Networks vGW Series virtual gateways to provide the following solutions:

- ▶ Stateful virtual firewall

Granular access control and VM isolation through policy enforcement for groups and individual VMs. vGW provides real-time packet inspection, from external to internal and inter-VM communication, to stop improper access.

- ▶ VM introspection

A centralized VM view includes OS, applications, packaged hot fixes, and so forth.

- ▶ Intrusion detection system

Selectable, protocol, and application-specific deep-packet inspection of allowed traffic for malware detection provides IDS protection.

- ▶ VM antivirus

On-demand and on-access scanning of VM disks and files with quarantining of infected entities maintains VM antivirus control.

- ▶ Security compliance, auditing and reporting

Another key design consideration in securing virtualized environments is to incorporate data archiving policies to the data collected by vGW. This can be achieved by leveraging existing data center syslog, security information and event management (SIEM) environments or by integrating Juniper Networks Security Threat Response Manager (STRM). STRM permits capture of logs, events and statistics of all traffic between virtual machines. STRM allows network and security administrators to ensure events are captured, stored and made available for compliance and regulatory requirements.

For more detail about STRM Series devices, see the following site:

<http://www.juniper.net/us/en/products-services/security/strm-series/>

IBM QRadar® Security Intelligence Platform integrates SIEM, Log Management, Network Activity Monitoring and Risk Management technologies. As a result, QRadar enables security professionals to comprehensively prevent, defend, remediate, and analyze exploits and policy violations, all through a unique *one-console* security intelligence approach. For more details, see the following site:

<http://q1labs.com/products/qradar-siem.aspx>

3.2.2 Integrating SRX Series and vGW to centralize policy management

Security design challenges are inside the VMs and virtual servers and also include components such as external physical firewalls, IDS, antivirus appliances, and so forth. Figure 3-6 depicts the architecture of the integrated SRX and vGW Series solution. The SRX zone that holds security policies can be synchronized with the vGW Security Design (SD).

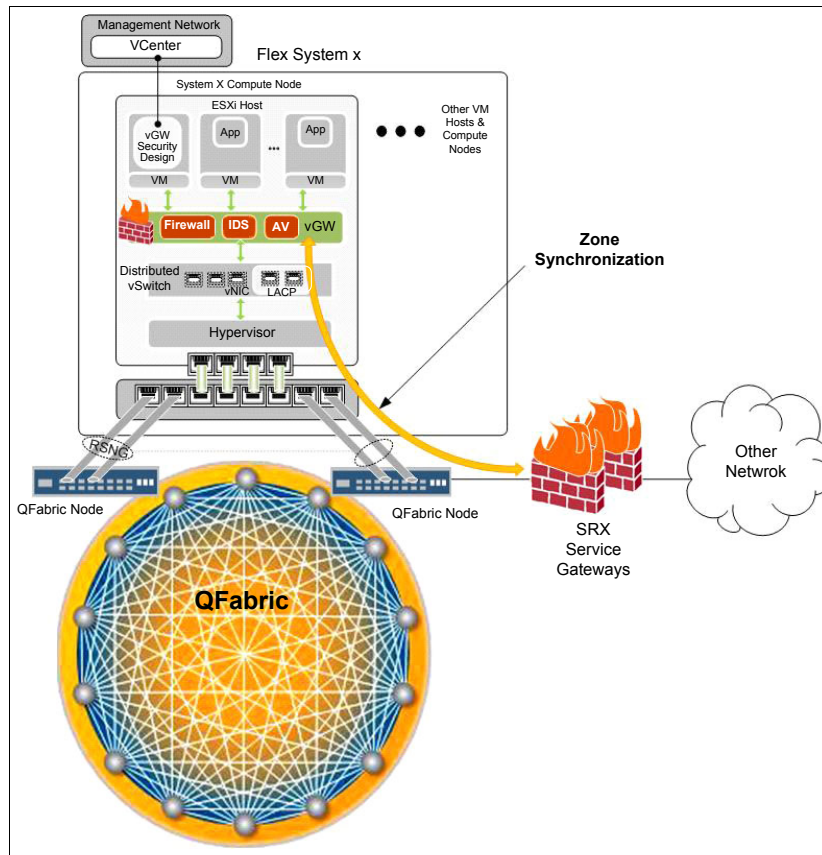


Figure 3-6 vGW and SRX zone information synchronization

A key design consideration is to create security boundaries that permit integration of these key components as follows:

- ▶ Integration of SRX and vGW, by creating smart groups that are aligned to SRX zones to centralize management of security policies
- ▶ Deployment of vGW smart groups and firewall policies to guarantee consistent security enforcement from the boundary perimeter to the VM or interfaces of the VM
- ▶ Implementation of compliance monitoring and enforcement of SRX Series zones within the virtualized environment
- ▶ Automated quarantine of VMs that violate access, regulatory, or zone policies
- ▶ Application of per zone policies at the granularity of individual VMs, VM groups, and vNICs

Security policies are created within SRX zones that are aligned to smart groups. Smart groups can contain either a single virtual machine (VM) or multiple VMs. Each smart group is assigned a firewall policy, which the vGW that is installed on the VMware vSphere ESXi server enforces.

These functions are achieved by sharing of zone information, such as, interface, network, and routing configuration between the vGW SD and SRX zone.

Consider the following options when designing vGW SD to secure the smarter data center:

- ▶ **Split-center**
Allows segmentation of information that is contained in a single VMware vCenter into multiple independently managed vGW Series centers. Allows for improved resource isolation for cloud services and multitenancy. Scales as the VMware vSphere capacity grows, allowing for near limit-less coverage.
- ▶ **Multi-center**
Allows administrators to designate a vGW Security Design VM that is connected to a vCenter at one location as the master. The master vGW Security Design VM ensures that all associated, subordinate vGW Security Design VMs are synchronized to the appropriate security policies. Consider this option when multitenancy or active/active data centers are key functional requirements.

Also consider the following firewall policy modes, which can be applied to either one VM or a group of VMs within a smart group:

- ▶ **No policy**
This option applies to the VM to ensure all associated vNICs are excluded from policy enforcement.
- ▶ **Policy-per-vNIC**
This policy applies at the VM level. It allows administrators to specify whether membership in the group applies to the entire VM or only to the vNICs to which the logic applies, for example whether the interface belongs to a port group or is attached to a VLAN.
- ▶ **Secure-per-vNIC**
An option within policy-per-vNIC allows enforcement of firewall policies on selected vNICs. This option allows specific vNICs to have no firewall policies defined.

For more information about vGW configuration options, see the following location:

<http://www.juniper.net/techpubs/hardware/vgw-series/5.0/vgw-install-admin.pdf>

3.2.3 Secure remote access

Juniper Networks SRX Service Gateway permits the configuration of secure virtual private network (VPN) access as depicted in Figure 3-7 on page 61. With SRX Service Gateway, you can group workload traffic, VLANs, and zones, depending on the requirements of the smarter data center. Based on the specific policy defined, and if integration with vGW is chosen, vGW enforces user separation at the SRX zone layer.

A key design consideration is to ensure that SRX zones aligns to security boundaries that are defined in both the SRX and vGW layers. This alignment might be a zone where a specific customer is hosted or a zone where one organization's business unit workloads are separated from general access (for example, separating an Internet banking administrator from the data warehouse administration access).

Consider integration of the SRX VPN with a role-based access control (RBAC) mechanism, such as remote authentication dial-in user service (RADIUS) or Lightweight Directory Access Protocol (LDAP). These services simplify user authentication, access, and accounting (AAA), ensuring only authorized users have access to critical systems within a smarter data center.

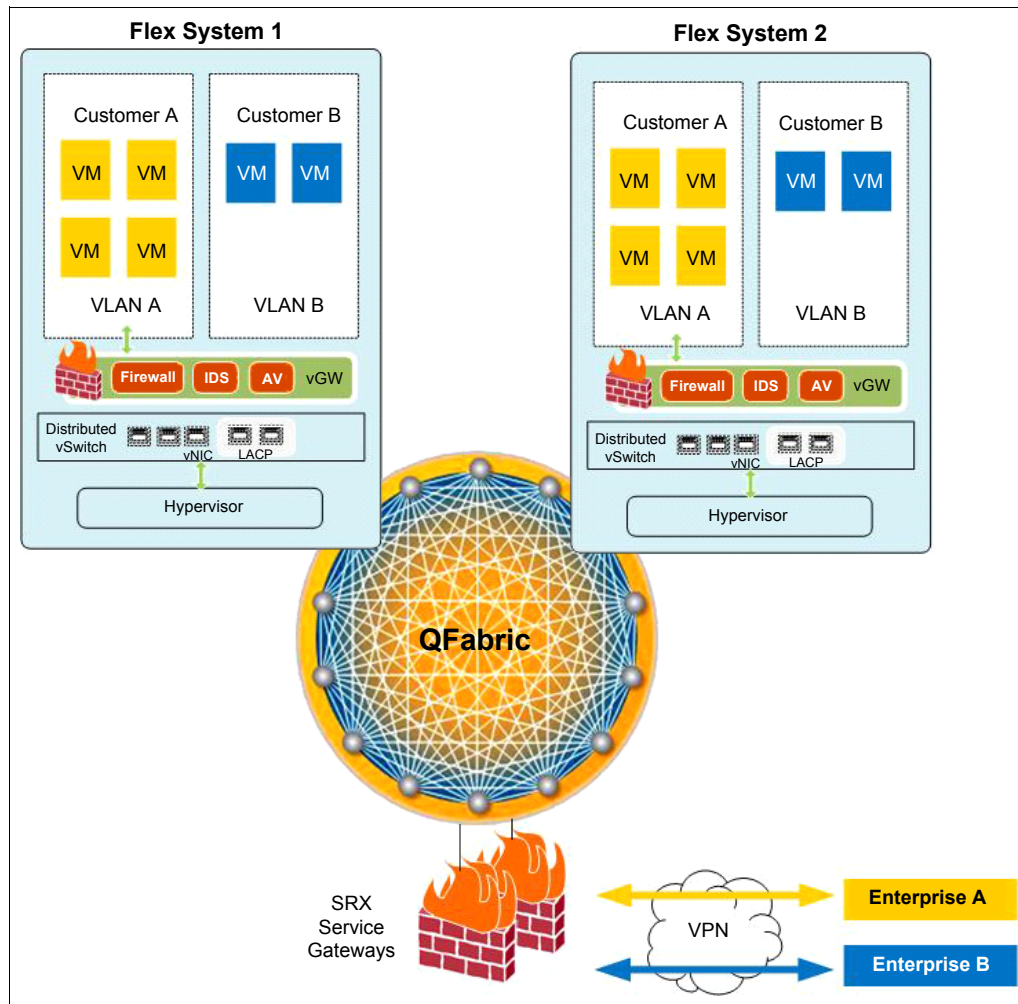


Figure 3-7 SRX and vGW provide a comprehensive solution for remote secure access

3.3 Provisioning

This section provides the following information:

- ▶ An overview of the typical physical level and system level provisioning tasks within today's data center
- ▶ Integrated provisioning tools and considerations when developing the end-to-end provisioning processes
- ▶ Details about the physical connection of an IBM Flex System BTO to QFabric
- ▶ Guidance regarding the most appropriate tool for each element within the building blocks of a smarter data center
- ▶ Several implementation considerations when provisioning QFabric logical elements, such as NNG, SNG, RSNG, VRRP, and LAG

3.3.1 Physical-level provisioning: Physical QFX Series switches

Provisioning of physical devices in data centers requires several decisions to be made based on the following key considerations:

- ▶ WAN services point-of-attachment
- ▶ Physical surface area of the data center
- ▶ Quantity of rows
- ▶ Length of rows
- ▶ Power density
- ▶ Scalability of virtual server platforms

Data center design patterns typically include pre-provisioned backbone cabling for connection of network devices, centralized or federated cable management, and data hall separation based on specific data center services (that is, communications, compute, and so forth). These data center designs require a flexible, high-speed, scalable, and resilient network infrastructure.

Juniper Networks QFabric meets these demands by collapsing core, distribution, and access layers into one high-speed, low-latency flat data plane (that is, distribution and access in one scalable fabric). With QFabric, decisions are simplified to the location and redundancy level of the QFabric nodes:

- ▶ Location:
 - Top-of-rack (TOR) deployment
 - End-of-row (EOR) deployment
 - Center-of-row (COR) deployment
- ▶ Redundancy level:
 - Server node group, which contains only one QFX Series switch
 - Redundant server node group, which contains two QFX Series switches

Because QFabric has a flat data plane, the only difference between these options is cabling cost. QFabric nodes can be placed anywhere within the data center, if maximum distances of interface transceivers are not exceeded, as listed in Table 3-1.

Table 3-1 QFabric interconnect transceiver options

Interface type	Cable type	Maximum distance
QSFP+ 40GBASE-SR4	850nm multi-mode fiber	150 meters
SFP 1000BASE-T	Cat5e (minimum specification)	100 meters
SFP-1G-SR	850nm multi-mode fiber	150 meters

3.3.2 Physical-level provisioning: IBM Flex System BTO and QFX3500

When provisioning an IBM Flex System BTO into an existing QFabric environment, consider the following key attributes:

- ▶ Available interface types
- ▶ Distance between interfaces
- ▶ Diverse path cable routing
- ▶ Single or multi-chassis IBM Flex System configuration
- ▶ Internal or external storage
- ▶ LAN and WAN services required for application workloads

The following example solution uses the IBM Flex System BTO bill of materials (BOM) provided in 3.1.3, “Providing an integrated solution with IBM Flex System BTO and QFabric” on page 56. The key consideration of the physical integration is the I/O module type. In this example, the EN4091 pass-through is used.

The QFabric node that is used depends on the requirements of the workloads and the compute node I/O adapters that are installed. In this case, CN4054 is configured in switch independent uplink mode.

For more information about this configuration, see “Switch independent vNIC mode ” on page 17.

Figure 3-8 depicts the physical connectivity of the IBM Flex System BTO and the QFabric Nodes. The configuration that is chosen is typical of a large environment where the following requirements must be met:

- ▶ No single point of failure (SPOF)
- ▶ Centralize storage

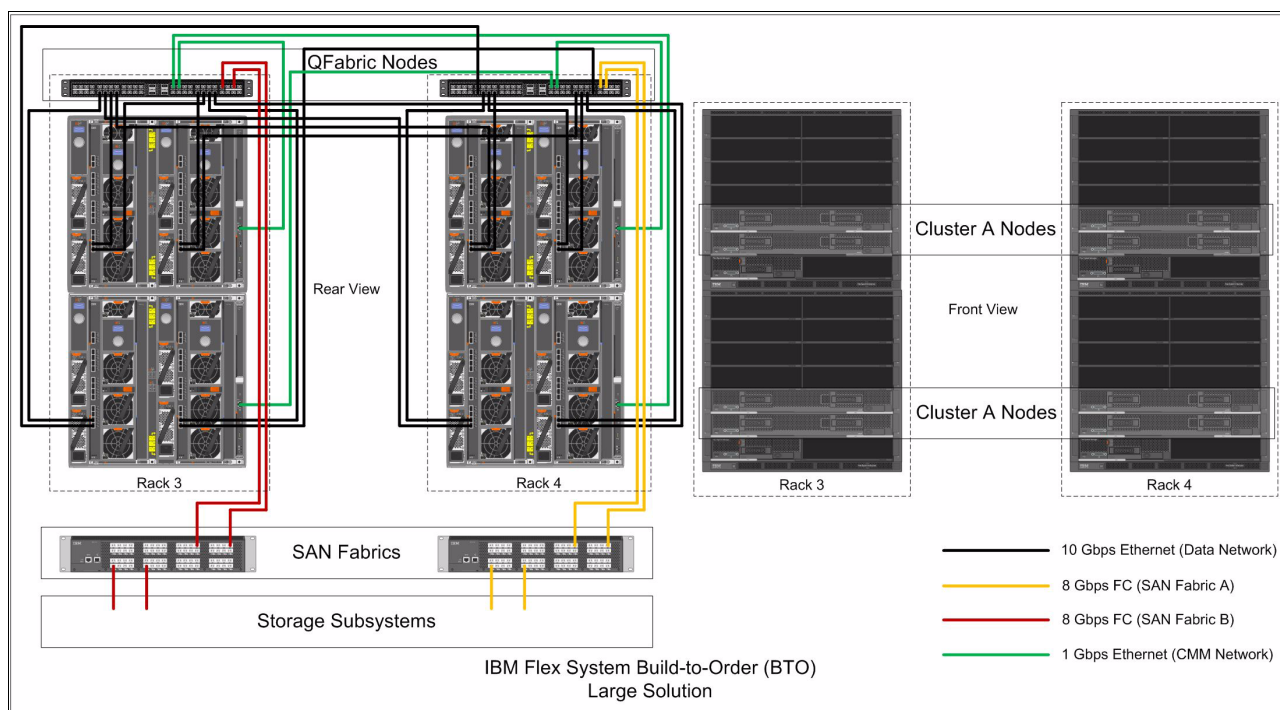


Figure 3-8 Large environment IBM Flex System BTO Layer 1

The EN4091 I/O modules are connected to each QFabric Node (QFX3500), from which a redundant server node group is configured to ensure redundancy. A LAG is also configured between both nodes in the redundant server node group to ensure resiliency and balanced bandwidth for application workload traffic.

The Fibre Channel (FC) connections from the QFX3500 to the SAN switches are not crossed. It is important to maintain complete FC separation at the redundant SAN fabric layer, because the failover algorithms are handled by the multi-pathing driver of the CNA and not the SAN fabrics themselves.

Another important consideration is the connectivity to the SAN fabric via QFabric. The following two FCoE mode types can offer various levels of flexibility when connecting to the SAN environment:

► FCoE-to-FC gateway mode

In this mode, the redundant server node groups (QFX3500s) are performing the FCoE-to-FC gateway function. Because the conversion of FCoE traffic to native FC traffic (and vice versa) is done in the FCoE-to-FC gateway, the servers and the SAN environment must be connected to the same QFX3500.

Figure 3-9 depicts an FCoE-to-FC gateway mode configuration. It has an IBM Flex System BTO multi-pod deployment (with external storage) that spans multiple redundant server node groups (RSNGs), defined as FCoE-to-FC gateways within the same QFabric.

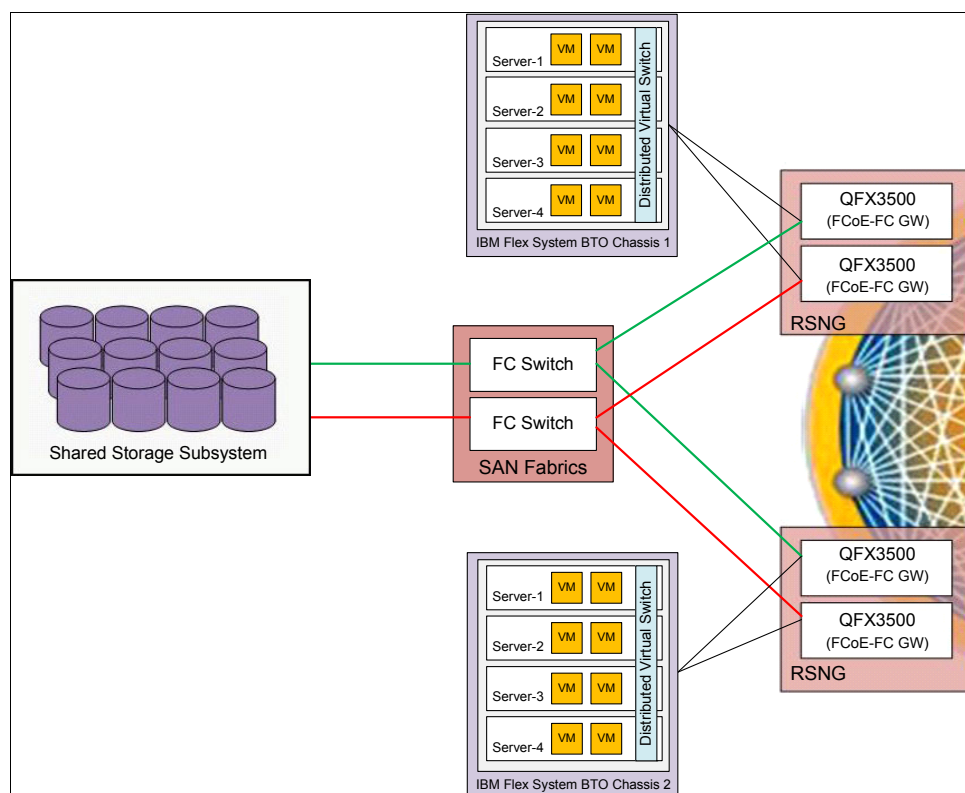


Figure 3-9 Multi-pod deployment using FCoE-to-FC gateway mode

► FCoE transit mode

This is the most common method of deployment. All traffic traverses the QFabric as native IP/Ethernet traffic, and class-of-service options can be configured to ensure the FCoE traffic receives the appropriate priority while traversing and egressing the QFabric.

In this mode, a pair of QFX3500 (in Stand-Alone mode) act as the FCoE-to-FC gateways and connect externally to Network Node Group devices in the QFabric. Therefore, allowing all servers connected to the QFabric access to the SAN environment.

Figure 3-10 depicts an FCoE transit mode configuration. An IBM Flex System BTO multi-pod deployment (with external storage) spans multiple redundant server node groups (RSNGs) that connect to standalone QFX3500s defined as FCoE-to-FC gateways, through network node groups (NNGs) within the same QFabric.

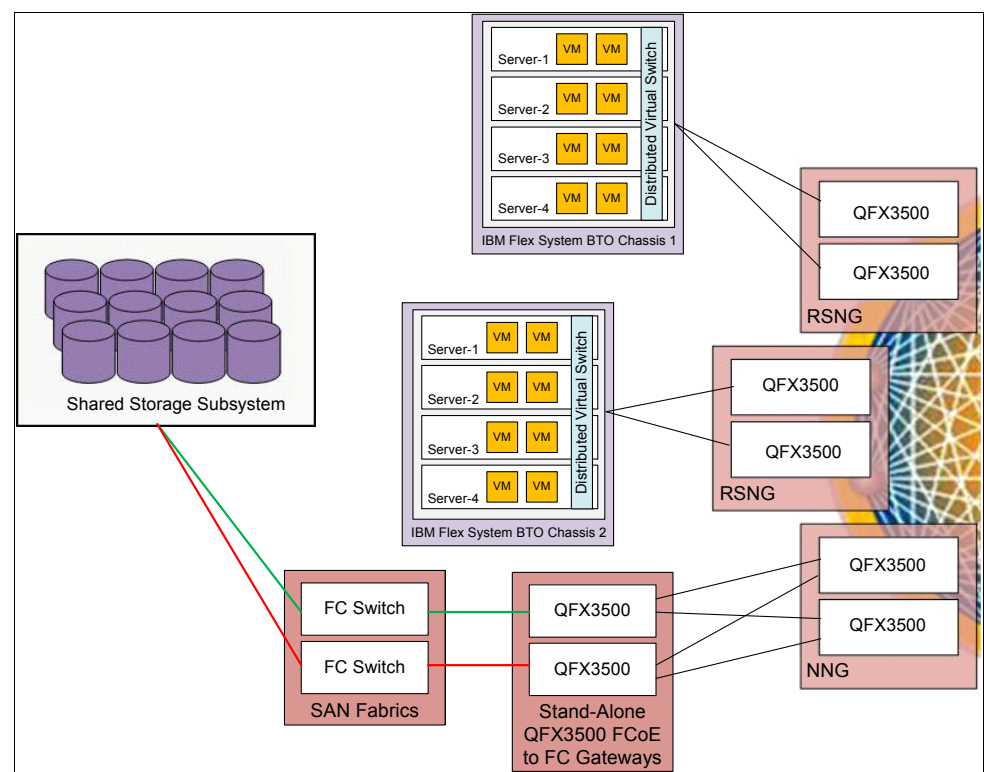


Figure 3-10 Multi-pod deployment using FCoE transit mode

Also see port group descriptions in “FCoE-to-FC gateway” on page 68 and “FCoE transit switch” on page 68 for more details regarding the port group attributes for those FCoE mode types.

Small to medium deployment model

Figure 3-11 depicts a possible variation of the previous solution (with internal storage) for smaller deployments. This variation halves the necessary hardware by removing the redundancy requirements.

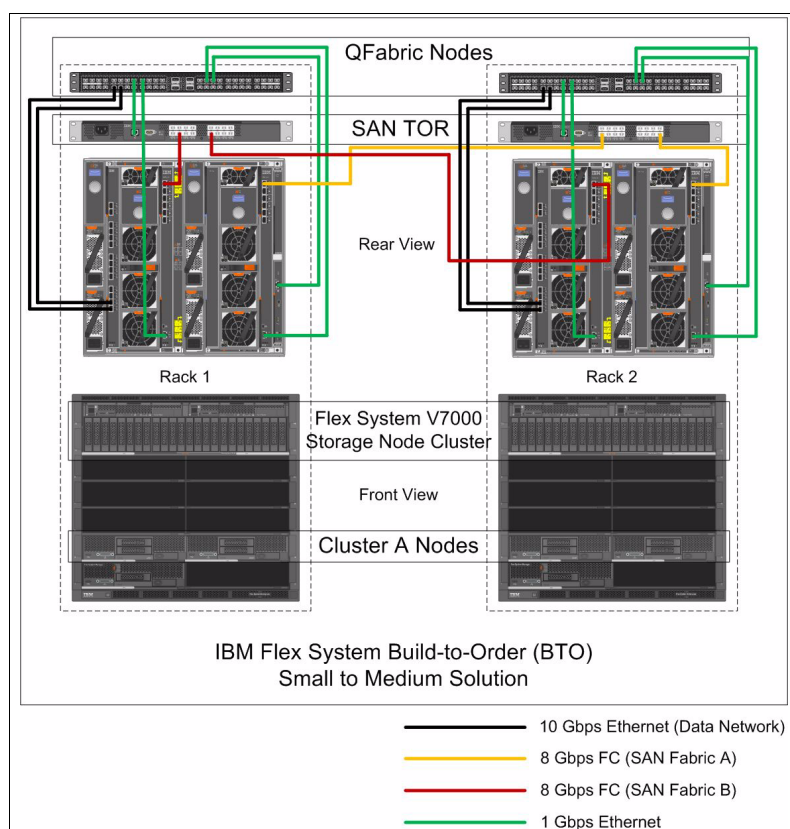


Figure 3-11 Small to medium environment IBM Flex System BTO Layer 1 with no redundancy

One EN4091 I/O module is installed into each IBM Flex System BTO enterprise chassis. Each I/O module is connected to one QFabric Node (QFX3500), from which a QFabric single server node group is configured. A LAG is also configured on the server node group to ensure resiliency and balanced bandwidth for workload traffic. A consideration for scalability when using a single interface/uplink to the QFX3500 is to consider configuring a single interface LAG. This approach allows simple and seamless addition of interfaces as requirements for workload bandwidth change.

3.3.3 System-level provisioning: IBM Flex System virtual elements

The IBM Flex System manager is as an intelligent, optimized resource element manager that allows server and network administrators wizard-based discovery, configuration, and event management functions. The IBM Flex System consists of the following logical elements:

- ▶ Server system pools: Create and modify system pools by using virtual workloads, make dynamic virtual workload adjustments, and move workloads within system pools.
- ▶ Network profiles: Quickly and easily specify network connectivity characteristics of a virtual machine.
- ▶ Storage system pools: Manage images for simple virtual machine creation, deployment, and cloning.

Provisioning of VMs in today's data centers requires careful planning to that ensure like workloads are grouped together to optimize performance and resiliency. VMs are placed in server system pools to permit creation of management policies. These management policies can be as simple as event management attributes or intelligent workload placement profiles, allowing dynamic relocation of workloads based on defined performance characteristics.

3.3.4 System-level provisioning: QFabric logical components

The provisioning of QFabric is made simple with the use of Junos Space Ethernet Design. Ethernet Design provides an efficient way to configure switches based on port profiles that are defined by Juniper Networks. It enables deployment and maintenance of enterprise networks as easily as deploying and maintaining a single switch.

Ethernet Design provides a workspace named EZ Campus Design that permits network administrators to customize and apply port profiles, and a workspace named Campus Analyzer where information about endpoint devices and ports that are available for configuration can be viewed. Deployment of port profiles is executed across multiple node devices. Ethernet Design does this task by allowing the network administrator to group node devices into node bundles and group ports into port groups.

Node bundles are logical groups of QFabric nodes, defined by the network administrator and based on either traffic management or specific workloads. Port groups are logical groups of ports within a node bundle. When a port group configuration change is applied to a node bundle, Ethernet design automatically applies the change to the same ports within all nodes of the bundle. Greatly simplifying provisioning tasks and increasing agility.

The following port groups are available:

- Dual attach

This port group is configured by selecting two ports from the same node device. This port group type uses the server port profile, and can be created on a server node group, redundant server node group, and network node group. A LAG is configured automatically when a dual-attached server port group is created.

- Dual home

This port group is the same in feature and function as dual attach, but cannot be applied to a server node group as the nature of dual home is an active/active resilient connections across redundant node devices.

- Ethernet

This port group can be configured with one of the following port profiles:

- Desktop port profile
- Switched uplink port profile
- Switched downlink port profile
- Server port profile
- Server access port profile
- Wireless access point port profile

A key consideration of this port group is that it supports only 10 GbE and 1 GbE ports.

- ▶ FCoE-to-FC gateway

This port group can be configured with the following port profiles:

- Fibre Channel port profile if all the ports in the port group are Fibre Channel ports
- FCoE gateway port profile if all the ports in the port group are 10 GbE ports

You can add either 10 GbE or FC ports to an FCoE gateway port group but not both simultaneously.

- ▶ FCoE transit switch

This port group can be configured with the following port profiles:

- Network FCoE Transit Port profile
- Server FCoE Transit Port profile
- Converged Server Access profile

This profile acts as a converged server access port and server FCoE transit port. This profile permits customization of the following settings:

- General settings
- CoS settings
- Ethernet switching options
- Maximum packet size

You can add only 10 GbE ports to an FCoE transit switch port group.

These port group types can be easy to apply, and offer a group function that can be customized to the node, thereby greatly simplifying the provisioning of QFabric configurations, specific to the challenges of today's data centers.

3.4 Management

This section provides an overview of typical physical-level and network-level design considerations of IBM Flex System Manager and Junos Space management appliances. It considers management functions of a smarter data center and provides solutions for performance and capacity planning, performance and capacity reporting, event management, and charge back in both small to medium and large scale deployments. It also presents an architecture for the QFabric management network and an architecture for the FSM management network, intended to depict the key design considerations.

3.4.1 Physical-level planning: Junos Space management network

When designing the QFabric management network for the first time, consider the following design attributes:

- ▶ Network topology
- ▶ Physical location within the data center
- ▶ Logical location within the network
- ▶ Available interface types
- ▶ Available compute hosting options
- ▶ Appliance type (physical or virtual)

The chosen network topology (either integrate into the existing management network or use greenfield deployment) will determine the architecture.

Figure 3-12 depicts an example network architecture whereby either a JA1500 or Virtual Appliance is redundantly connected to a Juniper Networks EX Series switches. The virtual switch is the data center's management network. The following management devices can connect to this network:

- ▶ Out-of-band (OOB) interface on the QFabric Directors (QFX3100s)
- ▶ External facing management firewalls
- ▶ Tivoli gateways
- ▶ LDAP services
- ▶ NTP services
- ▶ Syslog services

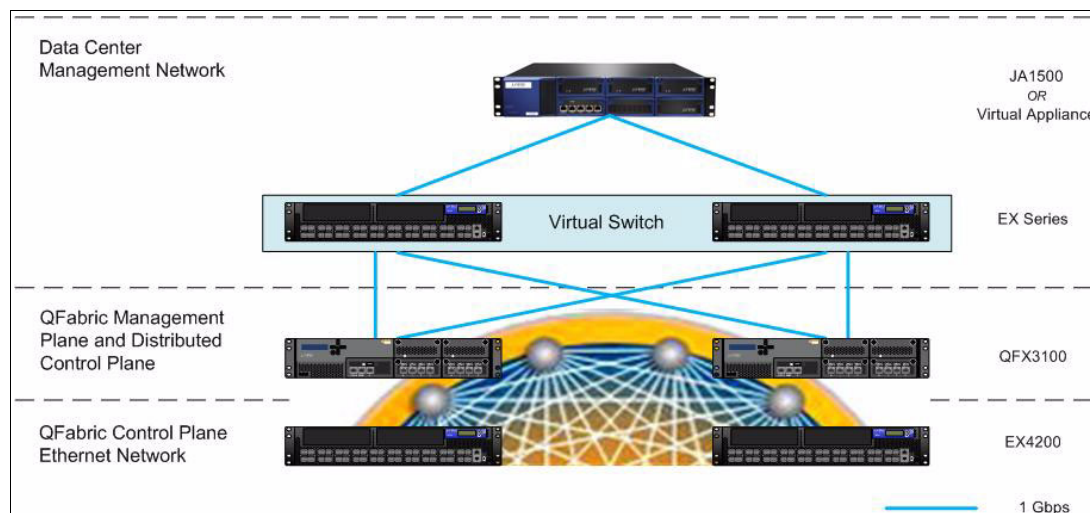


Figure 3-12 Junos Space management network architecture

In this example, the JA1500 directly connects to the management network using redundant 1000BASE-T network interface cards (NICs). The virtual appliance is an option for this solution. However, the virtual appliance requires a VMware vSphere compute environment with available capacity within the management network to host the Junos Space virtual machine.

See 2.6.2, “Junos Space” on page 42 for JA1500 hardware specifications.

Another key consideration is the management network operational model and recovery plan in the case of a significant failure (for example, recovery from a data center wide power failure). Because the data center management network contains the network management devices, it is considered the most critical component of the data center network. Consider the high availability, accessibility, and security requirements when designing the management network.

The following example is of a data center recovery plan in the case of restoring IBM Flex System and QFabric services from a data center wide power outage:

1. Restore power to management network racks.
2. Start Junos Space appliance.
3. Restore power to QFabric Directors.
4. Restore power to QFabric Interconnects.
5. Restore power to QFabric Nodes.
6. Access Junos Space and stabilize QFabric.
7. Restore power to IBM Flex System BTO enterprise chassis.
8. Access IBM Flex System Manager, and stabilize IBM Flex System.

3.4.2 Physical-level planning: IBM Flex System Manager management network

The IBM Flex System BTO enterprise chassis contains a management network to support the following functions:

- ▶ Management network switching within the enterprise chassis
- ▶ Management network connection to external interface on the chassis management module (CMM)
- ▶ Management network connection of an integrated management module (IMM)

These networks provide essential connectivity to IBM Flex System components that are required for management of the IBM Flex System. The external interface on the CMM is key in permitting access to server administrators. Consider installing the second CMM within the enterprise chassis, if eliminating management network SPOFs is a requirement.

The external CMM network connection is a 1000BASE-T Ethernet interface, two interfaces if the second CMM is installed. When using the redundant CMM option consider cabling each interface to a 1000BASE-T interface on a QFX3500 node within an RSNG, as in Figure 3-13.

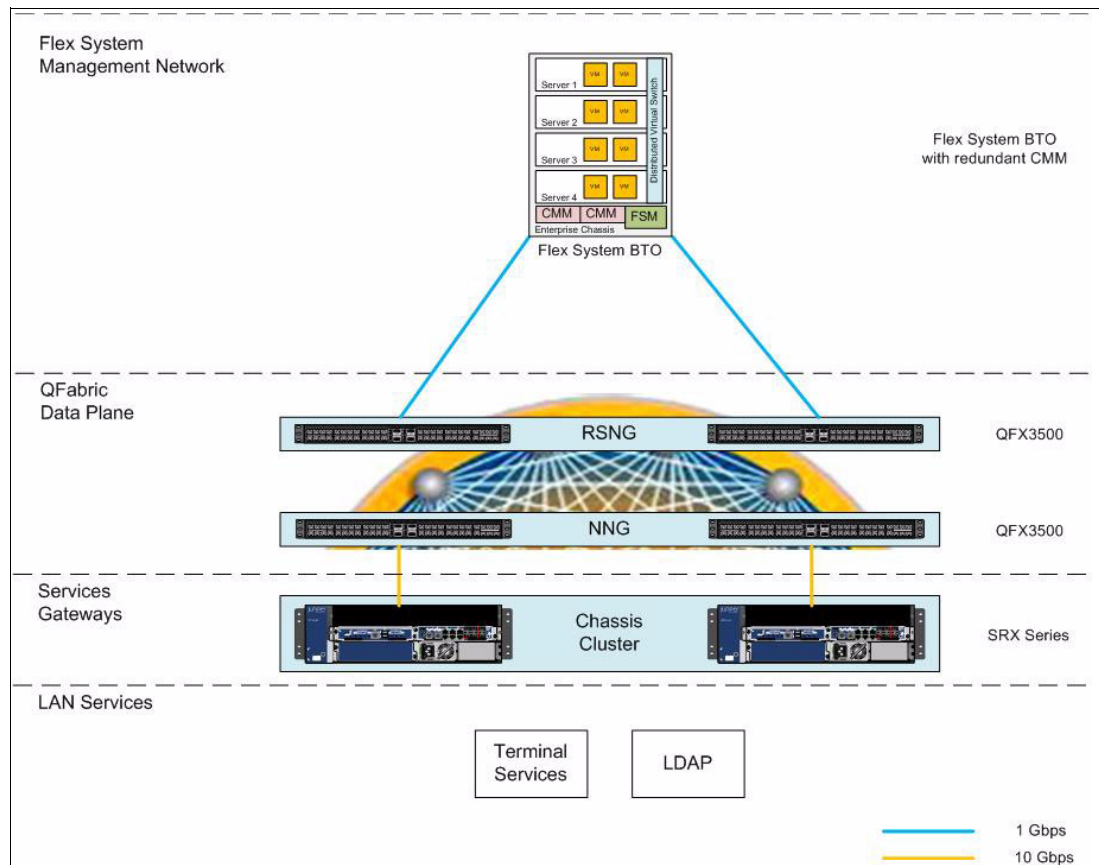


Figure 3-13 IBM Flex System Manager network architecture

This example architecture depicts a network node group that has external network services connected through a redundant SRX Series Service Gateway. The service gateway provides firewall protection from users within the LAN services networks from which services, such as LDAP and terminal services, can be provisioned.

See 2.2.8, “Network management integration” on page 22 for more detail.

3.5 Automation

This section provides guidance about the provisioning tasks that are suited to fulfillment through automation. It describes current challenges in end-to-end service-oriented automation to consider when designing a smarter data center.

Automation in a smarter data center comes in many forms:

- ▶ End-to-end compute, storage and network provision from service catalog
- ▶ Policy-based, performance oriented, dynamic workload relocation
- ▶ Policy-based, business continuity oriented, automatic workload relocation
- ▶ Compute vendor VM provision
- ▶ Network vendor virtual network provision
- ▶ Storage vendor storage pool provision

Automation can streamline and remove human error by provisioning repeatable tasks, such as deploying a VM or applying port configurations to multiple ports within network devices. Automation processes require integration of existing technical-level provisioning and team-oriented workflow to ensure a successful deployment.

3.5.1 End-to-end automation tools

One approach is to use an end-to-end (E2E) automation tool. Typically, this type of tool relies on standards-based, application programming interfaces (APIs) within each hardware vendors resource element managers to coordinate change in each technology domain. IBM Flex System Manager and Juniper Networks Junos Space are resource element managers that have standards-based APIs. These APIs can be exposed to E2E automation tools, such as IBM Cloud Service Provider Platform (CSP²), to execute the following functions:

- ▶ Provide a service portal containing IT services
- ▶ Expose a service request catalog to clients
- ▶ Contain design patterns/blueprints
- ▶ Contain OS images
- ▶ Contain Tier 1, Tier 2, and Tier 3 packages (that is, management tools, middleware, and so forth)
- ▶ Design workflows to allow automation of designated compute, network and storage resources
- ▶ Communicate to resource element managers to execute automation workflows
- ▶ Monitor resources for execution of dynamic workload relocation
- ▶ Report capacity and performance for planning and charge back purposes

Figure 3-14 depicts the component model of IBM CSP².

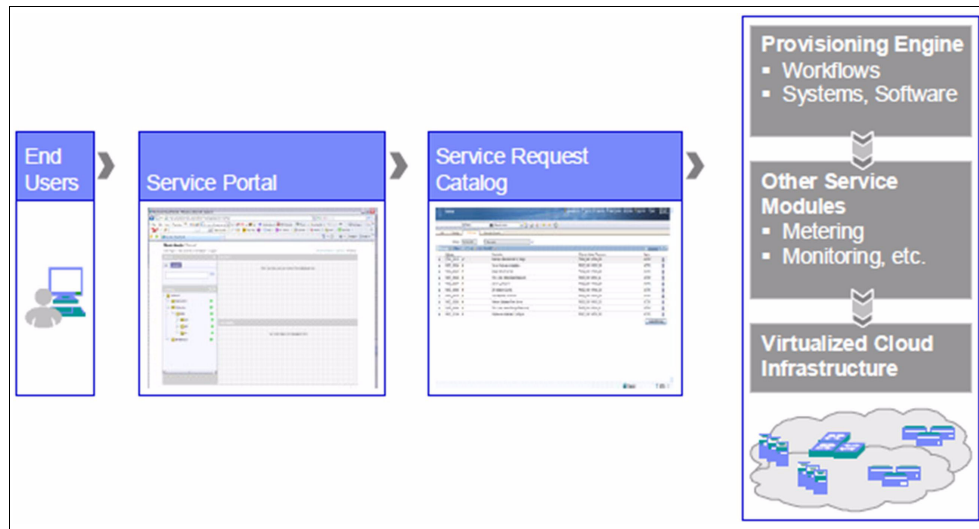


Figure 3-14 IBM CSP² component model

CSP² bundles the following IBM Tivoli products to provide a extensible automation platform that supports many different hardware devices:

- ▶ Tivoli Service Automation Manager

Enables clients to request, deploy, monitor, and manage cloud computing services by using automated workflows. It also provides traceable approvals and processes as part of service request fulfillment.
- ▶ Tivoli Provisioning Manager

Helps organizations optimize efficiency, accuracy, and service delivery by automating data center provisioning tasks.
- ▶ Tivoli Netcool Configuration Manager

Automates routine network configuration management tasks, enhances network security by controlling access by users, devices and commands, and maintains the real-time network state.

Tivoli Provisioning Manager and Tivoli Netcool Configuration Manager are fully compatible with Juniper Networks QFabric. These solutions offer provisioning and configuration management of Juniper EX Series, MX Series, SRX Series, and QFX Series platforms.

3.5.2 Integrated automation: IBM Flex System and QFabric

IBM Flex System and Juniper QFabric provide integrated automation of repeatable tasks for the compute and storage domain and the high-speed fabric technology domain. Consider using the IBM Flex System Manager for IBM Flex System automation and Junos Space for QFabric automation. Figure 3-15 on page 73 illustrates using IBM Flex System Manager and Junos Space workflows when integrating automation workflows.

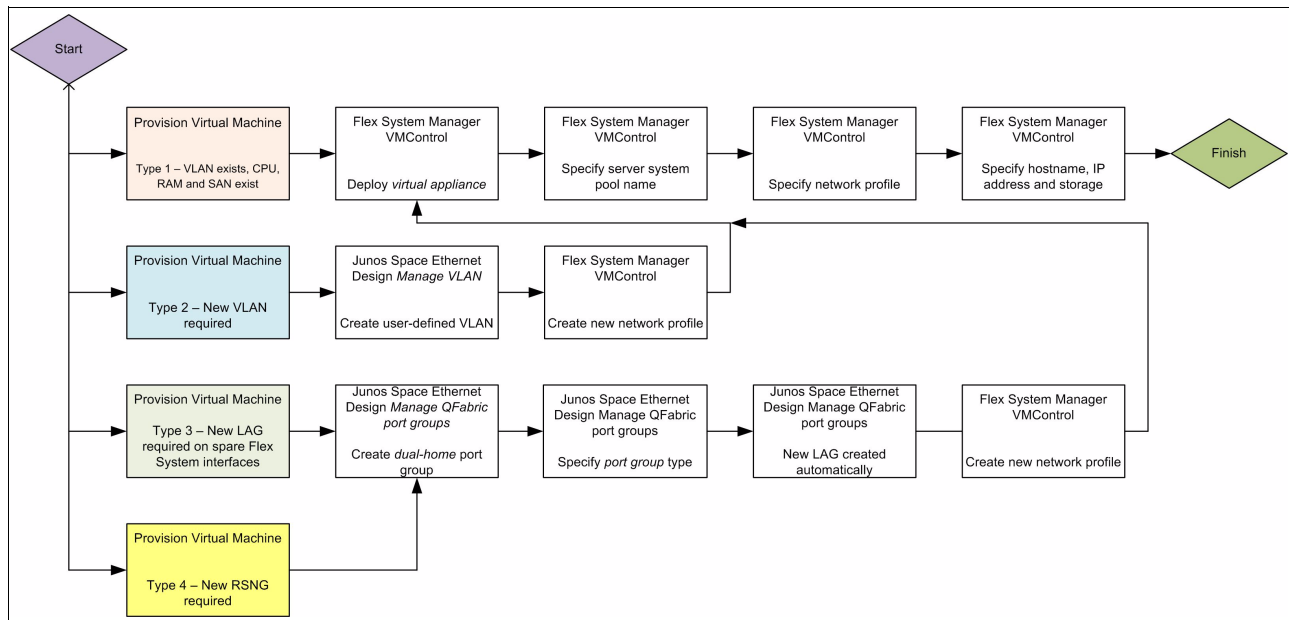


Figure 3-15 E2E IBM Flex System VM provisioning and QFabric automation

The key consideration of this approach to automation is ensuring that a team-oriented workflow is implemented to be able to pass technical-level processes between teams. This type of workflow can be implemented in scheduling type systems or using intelligent process oriented applications like IBM Rational® Method Composer.

The example set of provisioning tasks in Figure 3-15 depicts four typical VM provisioning tasks that are executed within an IBM Flex System that is integrated into QFabric. The tasks themselves are simplified to provide a one page view. The difference between each VM provisioning task, ranging in complexity, is based on the status of the available resources within the data center as follows:

► Type 1

This example provisioning task is the simplest because it requires no additional network resources (that is, VLAN, additional LAG because of spare interfaces on the IBM Flex System and so forth).

► Type 2

This example provisioning task builds on type 1 with the addition of a new VLAN to both the IBM Flex System and QFabric redundant server node group.

► Type 3


This example provisioning task builds on both type 1 and type 2 with the addition of a new LAG. This LAG is created between a new QFabric node group and spare interfaces within the IBM Flex System. This workflow can be used to support different workloads within the IBM Flex System. The Junos Space, Ethernet design, and manage port groups workflow is used here to create a new redundant server node group. Junos Space automatically creates a LAG associated to the new RSNG.

► Type 4

This example provisioning task is superfluous, meaning it is a task that is different from a client point-of-view (that is, within a service catalog). However, it is the same at the technical provisioning level. Consider assessing all provisioning tasks for common technical level workflow to simplify automation.

Provisioning tasks: Some VMware vSphere provisioning tasks in IBM Flex System Manager Network Control and Junos Space Virtual Control are mutually exclusive. When planning for automation, ensure the operational model clearly defines which of the two element managers is used for each provisioning task.

See 2.6, “Management domain: Integrated tools” on page 38 for more information about IBM Flex System Manager Network Control and Junos Space Virtual Control.



Verifying key client use cases with IBM Flex System BTO and Juniper Networks QFabric

As with any novel or innovative design, it is important to validate the design by performing a proof of technology (PoT). A PoT gives an opportunity to demonstrate that the design and the capabilities of its components function as expected.

In this chapter we introduce three use cases that solve some of the key challenges in today's traditional data center environments:

- ▶ Business continuity and disaster recovery

This use case supports the functional and non-functional requirements of business continuity and disaster recovery for vGW high availability deployment and redundant uplinks between IBM Flex System BTO and Juniper Networks QFabric.

- ▶ Multitenancy

This use case supports the typical functional and non-functional requirements of a multitenancy environment. It incorporates the server, network, security, and management architectures that are required to support the requirements.

- ▶ Virtual machine mobility

This use case extends the business continuity and disaster recovery solution by detailing the specific functional and non-functional requirements of virtual machine (VM) mobility.

These use cases further integrate the capabilities of the IBM Flex System BTO and QFabric architectures discussed in previous chapters of this paper.

Each use case completes a PoT with two validation scenarios and a separate interaction map illustrating the integrated components and their capabilities. The validation scenarios verify that the requirements for each use case are satisfied. In addition, the use case PoTs validate the use of QFX3500 and QFX3600 TOR switches with the IBM Flex System BTO configurations.

4.1 Business continuity and disaster recovery

Business continuity and disaster recovery are key attributes for today's computing environments. Organizations require business process resiliency to maintain a competitive advantage and to execute business operations. Business continuity and disaster recovery extend beyond the data center into architecture and design principles, change control, and daily activities, such as help desk operations and procedural maintenance. Business continuity and disaster recovery cannot be implemented at the time of disaster but refer to activities that are performed on a regular basis to maintain stable business operations.

4.1.1 Description

Figure 4-1 depicts an example network and compute architecture to support business continuity and disaster recovery requirements.

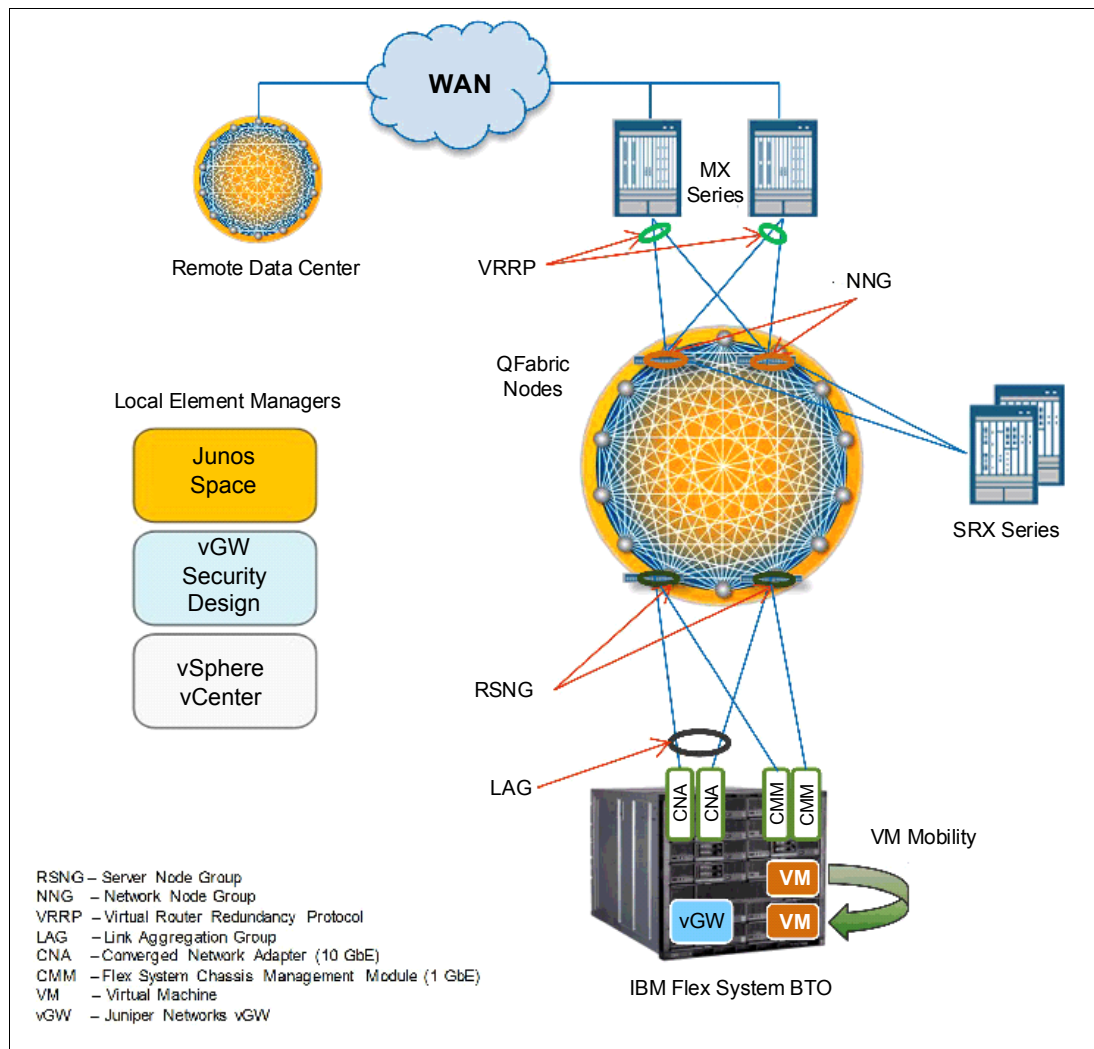


Figure 4-1 Use case 1, architecture overview diagram

This example architecture includes two data centers. Each data center includes a duplicate set of infrastructures in the same architectural configuration. Each data center consists of the following components:

- ▶ Two MX Series routers that support the wide area network (WAN) that spans the data centers
This connection provides redundant MPLS/VPLS networks for communication between infrastructures to be established.
- ▶ Two SRX Series services gateways that enforce centralized security policies
- ▶ One Junos Space virtual appliance that enables network administrators to manage all network devices in the data center
- ▶ One VMware vSphere vCenter that enables server administrators to manage the VMware VMs within IBM Flex System BTO
- ▶ One vGW Security Design VM to communicate with the SRX Series for managing and configuring the vGW policies.
- ▶ Two QFX3500 switches configured in a redundant server node group
The redundant server node group supports redundant uplink connections to IBM Flex System BTO I/O modules
- ▶ Two QFX3500 switches that are configured in a network node group that supports the following connections:
 - Redundant uplink connections to the WAN services through MX Series routers
 - Redundant uplink connections to SRX Series Services Gateways that enforces centralized security policy management
- ▶ One IBM Flex System BTO that contains the following components:
 - Four IBM Flex System compute nodes hosting a VMware vSphere cluster
 - Four vGW virtual machines (one per ESXi server) that enforces SRX zone security policies

4.1.2 Requirements

A typical business continuity and disaster recovery solution within a smarter data center includes the following functional and non-functional requirements:

- ▶ Functional requirements:
 - Minimal disruption
 - Scalable
 - Data center independent
 - Manually initiated disaster recovery
- ▶ Non-functional requirements:
 - No reliance on client software
 - Low latency cross site network
 - Maintain data backup location, optimized for restore
 - No manual configuration changes when relocated to secondary data center (that is, IP address change or other configuration changes)
 - Portable security and routing services to support seamless relocation to secondary data center
 - Load balanced

- Application independent
- Data synchronized with disaster recovery site
- Business continuity, one hour of maximum allowable outage (MAO)
- Disaster recovery, four-hour recovery time objective (RTO)
- Disaster recovery, zero recovery point objective (RPO)

Categorization of critical systems with MAO defined: When planning business continuity and disaster recovery, ensure that critical systems are categorized and have an MAO defined. *MAO* is the total duration of when a critical system can be offline before the business process is impacted. Some critical systems require the MAO to equal the RTO during a disaster recovery event. In this example, disaster recovery is longer than MAO.

4.1.3 Proof of technology

The proof of technology (PoT) is executed for the following scenarios:

- Scenario 1 includes resiliency of security services with vGW Security Design in a disaster recovery data center during a primary data center failure.

This scenario shows technology-level resiliency that is inherent within the architecture of Juniper Networks vGW Security Design.

- Scenario 2 ensures that one Juniper Networks QFabric redundant server node group can be configured with redundant link aggregation groups (LAG) by using interfaces within IBM Flex System build-to-order (BTO) with redundant EN4091 I/O modules.

This scenario details the steps to configure a redundant server node group with both 10 GbE and 1 GbE VLANs to support both Converged Enhanced Ethernet (CEE) and chassis management model (CMM) connections into QFabric.

Validation scenario 1

This scenario validates resiliency of vGW Security Design in the case of a local site failure where disaster recovery failover of vGW Security Design functions is required. It verifies the ability of vGW Security Design to provide key critical security policy functions to VMs during primary data center outage.

Terms of reference

The PoT demonstrates the following criteria:

- vGW Security Design high availability between data centers
- VMware high availability and Distributed Resource Schedule (DRS) settings to restrict vGW Security VMs from being moved through high availability or DRS

Interaction map

Figure 4-2 depicts the interaction map for the technology level steps, validating the terms of reference for this validation scenario.

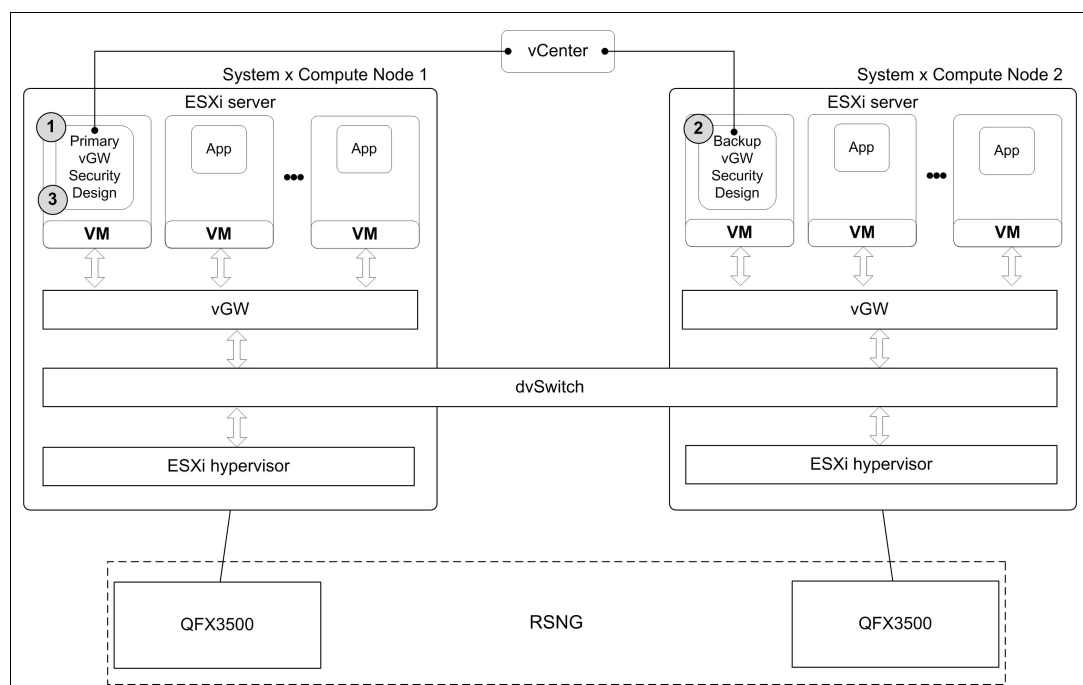


Figure 4-2 Validation of resilient vGW security services during vMotion event

The following steps describe this scenario:

1. Install the primary vGW Security Design in a System x compute node in IBM Flex System 1 in the primary data center.
2. Build another vGW Security Design VM from the vGW Series Open Virtualization Appliance (OVA) file by using the following steps.

Tip: An OVA is a single compressed file in Open Virtual Machine Format (OVF). The secondary vGW Security Design is located in a System x compute node in IBM Flex System 2 in a disaster recovery data center.

- a. Configure the secondary vGW Security Design VM to use the primary system on the Settings module **vGW Application Settings** → **High Availability** page.
- b. Deploy the OVA file for the vGW Security Design VM using the VMware vSphere Client. Use **File** → **Virtual Appliance** → **Import in VMware vCenter**.
- c. First configure the primary vGW Security Design VM for high availability (HA). Then, configure the secondary vGW Security Design VM for HA in the vGW Security Design settings module:
 - i. From the Standby Appliance list, select the vGW Security Design VM to be used as the secondary (standby) vGW Security Design VM.
 - ii. Select the IP address type to assign to the secondary vGW Security Design VM and how it will obtain the address. You can select an IPv4 or IPv6 address.
 - iii. Configure the proxy server and time configuration settings for the secondary vGW Security Design VM.

After you complete this configuration, the secondary vGW Security Design VM is powered on automatically and configured in the System x compute node in IBM Flex System 2 that is located in the disaster recovery data center. This process takes approximately ten minutes. After the operation completes, you can log in to the secondary vGW Security Design VM through the IP address that you specified during the configuration.

The vGW Series monitors connectivity between the two vGW Security Design VM management centers. It initiates promotion of the secondary system if there is no response from the primary one within three minutes.

When the primary vGW Security Design VM is brought back online after it has recovered or the host it was on is repaired, it automatically takes control again. The vGW Series HA is not designed to replace normal backup operations. Rather, it is expected that the primary vGW Security Design VM will be brought back online quickly.

3. The vGW Series automatically sets the VMware high availability and DRS settings to restrict vGW Security VMs from being moved through HA or DRS. It is important that a vGW Security VM not be moved to a new ESX/ESXi servers.

The vGW Series HA for the vGW Security Design VM behaves in the following ways:

- It allows the secondary vGW Security Design VM to continue distributing both antivirus policy and firewall policy until the primary one can be brought back online. When the primary vGW Security Design VM is unavailable, the secondary vGW Security Design VM pushes out the policy database to the vGW Security VMs when they request it. This policy is a copy of what existed in the primary vGW Security Design VM, and it cannot be modified.
- The HA configuration is used in business continuity or disaster recovery design to ensure that new VMs that are powered on ESX/ESXi servers can retrieve policy rather than default to VMsafe failure mode.

It is important to have redundancy at the vGW Security VM level. A vGW Security VM might become inactive, for example, when the vGW Security Design VM is inactive and the secondary will take over immediately. When the VM that hosts the primary vGW Security Design becomes inactive, the secondary Security Design takes over and becomes active in 60 seconds.

Results

The vGW Security Design HA maintains management of the associated vGW gateways on each VMware vSphere ESXi server by ensuring policy synchronization between primary and secondary vGW Security Design servers. This process can be done within a data center or across data center environments.

The activities that are conducted in this scenario verify the ability to protect client data by providing critical security functions to the data center when the primary data center is out of commission. The critical functions include a high-performance hypervisor-based stateful firewall, integrated intrusion detection (IDS), virtualization-specific antivirus protection, and unprecedented management scale for cloud security.

Validation scenario 2

This scenario validates the compatibility and configurations that are required to support connectivity of IBM Flex System BTO to a QFabric redundant server node group with 10 GbE interfaces using a LAG to support CEE protocols. It also validates network node group, redundant server node group, and MX Series routing configurations that are required to support connectivity of 1 GbE interfaces from redundant IBM Flex System CMM.

This scenario is well-suited for the creation of redundant network connections to support high availability solutions within any data center environment.

Terms of reference

The PoT demonstrates the following criteria:

- ▶ LAG compatibility between IBM Flex System and QFabric
- ▶ Configuration of Layer 2 VLANs for CEE connections
- ▶ Configuration of Layer 3 VLANs for connection of the IBM Flex System management network

Interaction map

Figure 4-3 depicts the interaction map for the steps to validate the terms of reference for this scenario.

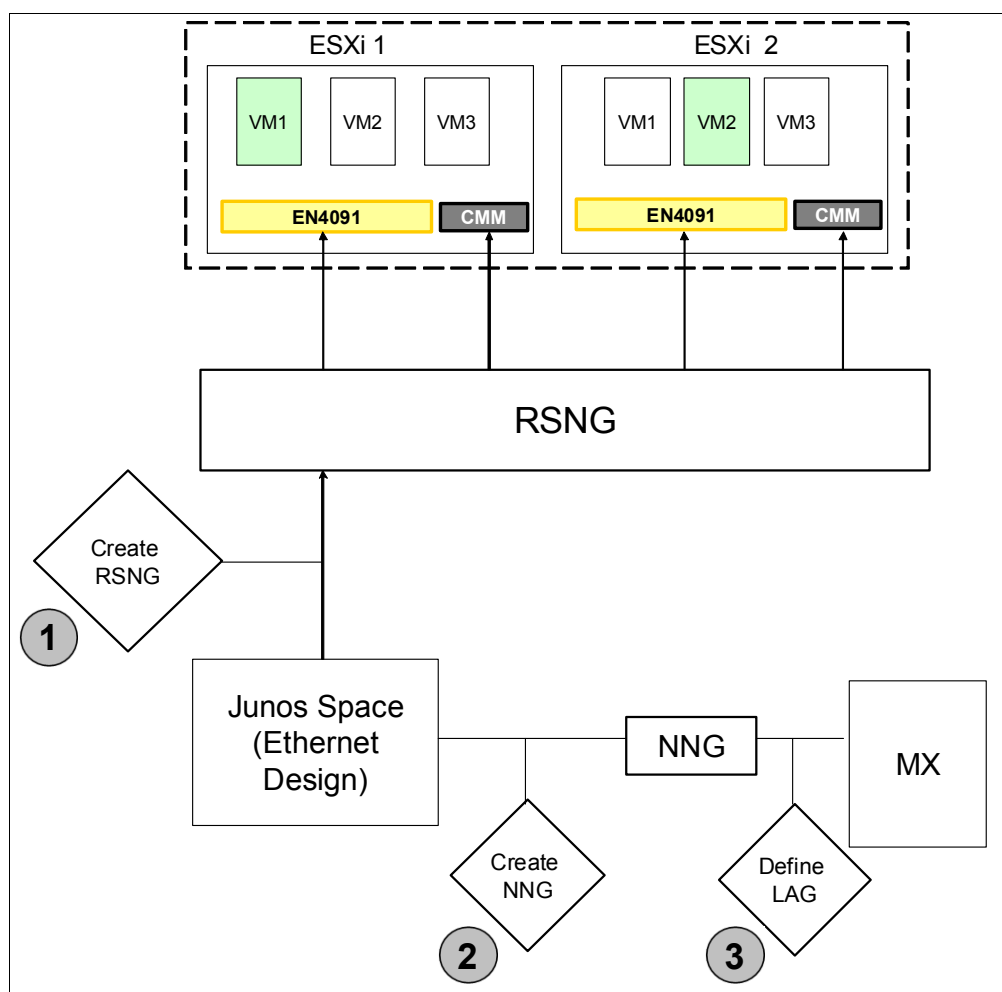


Figure 4-3 Validation of redundant uplink configuration between IBM Flex System BTO and redundant server node group

The following steps detail the configuration of a new redundant server node group and dual-homed LAG (that is, active/active) uplinks to permit the connection of an IBM Flex System BTO Layer 2 10 GbE interfaces, within redundant IBM Flex System BTO EN4091 I/O modules.

Also detailed in the steps is how to configure a new network node group and a LAG to a MX Series router to enable Layer 3 connectivity of the IBM Flex System CMM 1 GbE interfaces to enable IBM Flex System Manager management connectivity to server administrators.

1. Create a redundant server node group (10 GbE Layer 2):
 - a. Define the QFabric node alias and the redundant server node group:
 - i. Set the aliases for the QFX Series switches.
 - ii. Set the QFX Series resources within the redundant server node group.
 - b. Configure the LAG:
 - i. Define the LAG member count.
 - ii. Add the LAG members.
 - iii. Set the Ethernet options.
 - iv. Enable the Link Aggregation Control Protocol (LACP).
 - c. Configure the interface and assign the VLAN membership for a dual-homed server:
 - i. Set the port mode.
 - ii. Add the VLAN range.
2. Create the network node group (1 GbE Layer 3):
 - a. Define the QFabric node alias and the network node group:
 - i. Set the aliases for the QFX Series switches.
 - ii. Set the network domain.
 - b. Define the Layer 2 configuration by setting the VLAN IDs.
3. Define the LAG configuration for the network node group that is connecting to the MX Series device:
 - a. Define the LAG member count.
 - i. Add the LAG members.
 - ii. Set the Ethernet options.
 - iii. Enable the LACP.
 - b. Assign the IP address to the LAG interfaces by setting the IP address on each LAG member.
 - c. Configure the routed VLAN interface (RVI) for two VLANs (one for each CMM):
 - i. Set the IP address on each VLAN.
 - ii. Bind the RVI interface to the VLAN.
 - iii. Set the Layer 3 interface on each VLAN ID.
 - d. Configure the default routes to the MX Series by adding the network-based static route.
 - e. Verify the default route configuration by showing the route terse.

Results

The activities that are conducted in this scenario verify the compatibility of converged Layer 2 and Layer 3 networking standards when integrating QFabric and IBM Flex System BTO. This scenario also validates the ability to manage the IBM Flex System BTO through the IBM Flex System Manager, using the integrated 1 GbE CMM interfaces from within a network outside QFabric.

This scenario proves business continuity (BC) and disaster recovery (DR) through the use of Juniper Networks QFabric redundant server node group with redundant link aggregation groups (LAG) and the interfaces within IBM Flex System build-to-order (BTO) with redundant EN4091 I/O modules.

4.2 Multitenancy

Multitenancy refers to a multi-instance architecture where separate hardware or software instances operate individually within a data center environment. These individual instances can be assigned to separate business units within an organization or separate organizations. Each instance can be manipulated to provide different functions based on individual organizational requirements. Multitenancy is commonly termed *cloud computing*.

4.2.1 Description

Figure 4-4 depicts an example network and compute domain architecture to support a multitenant environment.

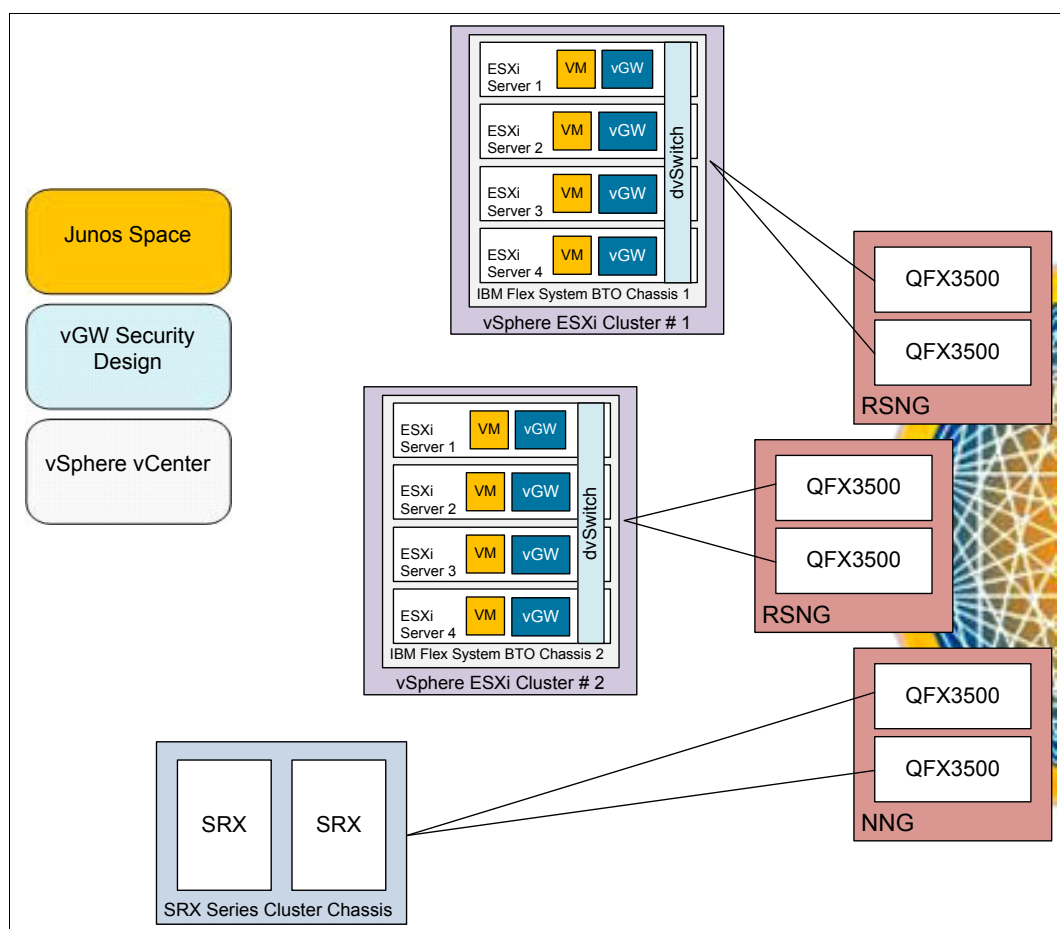


Figure 4-4 Use case 2, architecture overview diagram

The architecture contains the following components:

- ▶ Two SRX Series services gateways that enforce centralized security policies
- ▶ One Junos Space virtual appliance that enables network administrators to manage all network devices in the data center
- ▶ One VMware vSphere vCenter that enables server administrators to manage the VMware VMs within the IBM Flex System BTO
- ▶ One distributed virtual switch built using VMware vSphere distributed virtual switch (dvSwitch) within each ESXi server
- ▶ One vGW Security Design VM that enables configuration of SRX zones for VM-level security policy enforcement
- ▶ Two QFX3500 switches that are configured in a redundant server node group
The redundant server node group supports redundant uplink connections to the IBM Flex System BTO I/O modules.
- ▶ Two QFX3500 switches that are configured in a network node group that supports redundant uplink connections to SRX Series Services Gateways that enforce centralized security policy management
- ▶ Two IBM Flex System BTO, each containing four compute nodes, one VMware vSphere cluster, and four vGW virtual machines (one per ESXi server)

4.2.2 Requirements

Typical of data centers that support either a security boundary-enforced workload separation or multitenant workloads include the following functional and non-functional requirements:

- ▶ Functional requirements:
 - Cost effective
 - Secure
 - Flexible
 - Accessible
 - Simple
 - Resilient
- ▶ Non-functional requirements:
 - Shape traffic, based on application classification
 - Security policies based on application classification
 - Support clustered hosting platforms
 - Isolate and encrypt at rest data
 - Zone or tenant based reporting

4.2.3 Proof of technology

The proof of technology (PoT) is executed for the following scenarios:

- Scenario 1 shows SRX zone synchronization with vGW Security Design for enforcement of similar security policies across security zones.

This scenario enables SRX zone synchronization with vGW Security Design on a secure channel and provides the steps on vGW Smart Groups that enforce zone integrity policies on VMs.

- Scenario 2 provides visibility of dvSwitch VLANs within Junos Space Virtual Control.

This scenario shows the steps to install and configure a dvSwitch within an IBM Flex System BTO. It also includes steps to configure dvSwitch port groups to assign VLAN IDs for networks within the uplink trunk to the QFabric redundant server node group. It ensures that Link Layer Discovery Protocol (LLDP) is enabled on the dvSwitch, to be able to view the virtual network configuration from Junos Space Virtual Control.

Validation scenario 1

This scenario verifies SRX Series zone synchronization with vGW Security Design. It validates the configuration and provides the technology level steps that occur when linking a virtual security layer with a physical security layer. It also validates that vGW enforces the zone integrity on respective VMs using Smart Groups.

This scenario is well-suited for definition of security boundaries within a multitenant environment, or for separation of business units or based on functions within an organization.

Terms of reference

The PoT validates the following criteria:

- Configuration of automatic zone synchronization on vGW
- Policy group configuration on vGW Security Design
- Zone integrity enforcement by vGW gateway

Entry criteria

For this scenario, traffic isolation (security) policies are already defined in SRX Series.

Interaction map

Figure 4-5 depicts the interaction map for the steps to validate the terms of reference for this scenario.

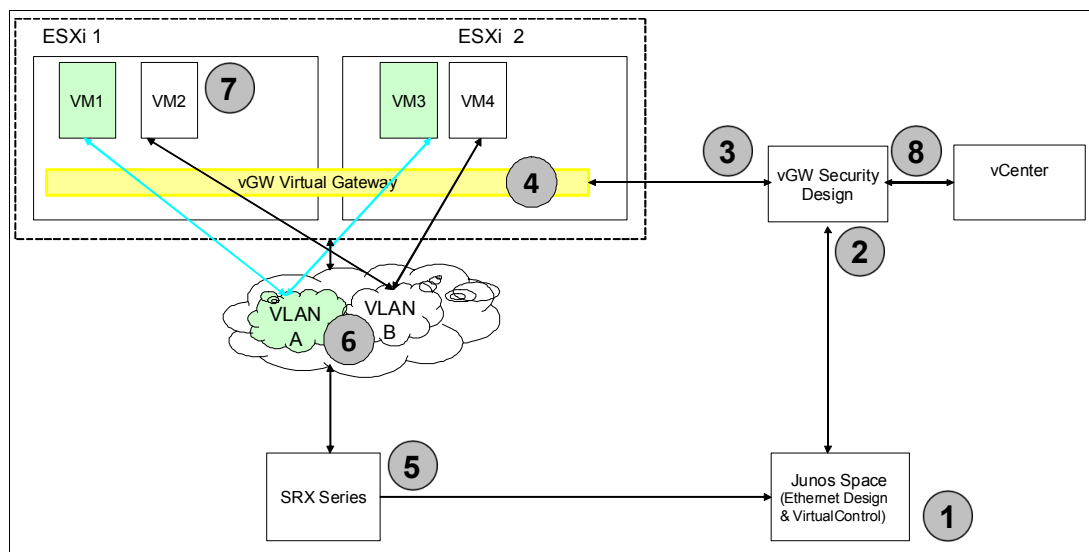


Figure 4-5 SRX and vGW integrated environment

The following steps describe this scenario:

1. Enable secure connection between vGW Security Design and SRX Series interface to read the authorized information from the SRX Series:
 - a. Enable the Junos script XML scripting application interface (API), using the Junos script interface on the SRX Series.
 - b. Configure vGW Virtual Gateway Automatic Zone Synchronization process.
2. The vGW Security Design communicates with the SRX Series to pull the zone definitions and populate them in the vGW firewall for each VMWare hypervisor. Likewise, vGW Security Design communicates with each vGW firewall to get IP Address information from each VM to add to the address book of the SRX Series. Use the following steps:
 - a. Set the frequency for the query to be made to SRX Series for updates.
 - b. If only a subset of the SRX Series interfaces is protecting the virtual network, select only those interfaces, so that only zones that are related to the virtual network are updated.

Note: The vGW Security Design VM is managing and configuring only vGW policies.

3. The vGW Policy Groups dynamically associate each VM to its zone, which can be used for inter-VM policy enforcement and zone compliance validations. The SRX Series zones are created in vGW as VM Policy Groups, using zone information from SRX Series devices, a Policy Group is created based on the following parameters:
 - VLANs are associated with the SRX Series interface.
 - Subnet is defined on the SRX Series interface, and routes are defined within a zone for IP range.

- If the zone sync configuration includes a “VMs associated” selection, the chosen group is included in the Policy Group for VM scope.
- After the zone synchronization is processed, a list of zones is displayed.

In addition particular zones can be selected to import into the vGW as VM zone groupings.

- The vGW uses the information from vGW Security Design to create VM Smart Groups so that users of vGW can see VM-to-zone attachments. Create additional inter-VM zone policies and incorporate zone knowledge into compliance checks:
 - The vGW Security Design VM discovers items, such as installed applications on a VM through VM Introspection, and VMware’s vCenter provides the attributes, such as the port group to which the virtual network interface is connected.
 - Smart Group is created by navigating to the Settings module by selecting **Security Settings** → **Groups** page, and **Add Smart Group**.
 - The vGW Series enforces the zone integrity on respective VMs using Smart Groups.
 - Additional Smart Group for a compliance rule to issue an alert can also be created.
- The SRX Series delivers zone-based segregation at the data center perimeter.
- VLAN A and B connect to their respective VMs in separate ESX servers.
- VMs for separate functionality reside on the same ESX servers, and respective security policies are applied from vGW.
- vGW Security Design server stays in constant communication with the VMware vCenter so that as changes to VMs occur, they are synchronized to the vGW management server.

Results

This scenario proves that the physical security layer can be integrated with a virtual security layer in an automated fashion. Thus, SRX Series traffic isolation policies flow through vGW Series and enforce security policies on respective tenant VMs. This scenario provides an effective and simplified security management process from the data center perimeter to the VM level.

Validation scenario 2

This scenario validates the steps to configure a dvSwitch within IBM Flex System BTO. It also allow visibility of the virtualized networking components to Junos Space Virtual Control.

This scenario is well-suited for multitenant or security zone-separated virtualized-networking environments where centralized network management and reporting functions are required.

Terms of reference

The PoT achieves the following criteria:

- Perform configuration of distributed virtual switch.
- Validate visibility of distributed virtual switch uplinks and VM vNICs using Junos Space.

Interaction map

Figure 4-6 depicts the interaction map for the steps to validate the terms of reference for this scenario.

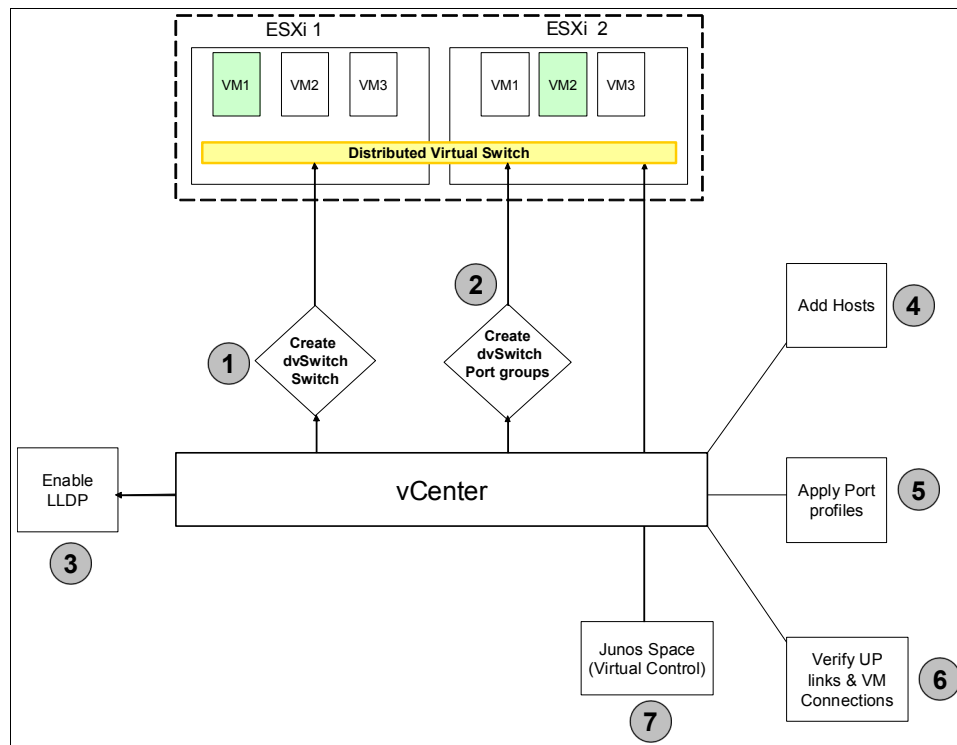


Figure 4-6 Visibility of VMware distributed switch within Junos Space Virtual Control

The following steps detail this scenario:

1. Create a distributed virtual switch:
 - a. Connect to vCenter.
 - b. Select inventory.
 - c. Select networking.
 - d. Create a new dvSwitch.
 - e. Define the dvSwitch name.
 - f. Define a data center name to add the dvSwitch.
2. Create the dvSwitch port groups and assign the VLANs that are configured in the redundant server node group through the vCenter:
 - a. Connect to vCenter.
 - b. Select **Inventory**.
 - c. Select **Networking**.
 - d. Select the new dvSwitch.
 - e. Launch a new port group in the Summary tab.
 - f. Set VLAN type to VLAN trunking.
 - g. Specify the VLAN ID range that was configured on the redundant server node group trunk, connected to the EN4091 I/O modules.
 - h. Finalize the configuration wizard.

3. Enable the LLDP on the new dvSwitch:
 - a. Select **Inventory**.
 - b. Select **Networking**.
 - c. Select **Advanced**.
 - d. Enable LLDP.
4. Add hosts to the new dvSwitch in the vCenter:
 - a. Select **vmnics** in the list to add to the new dvSwitch.
 - b. Apply the configuration.
5. Apply the port profiles to the VMs on the new dvSwitch using the vCenter:
 - a. Edit the VM settings.
 - b. Select the network adapter.
 - c. Change the network connection port profile to the new dvSwitch.
6. Verify the uplinks and connections of the VMs to the new dvSwitch in the vCenter:
 - a. Select the dvSwitch port profile.
 - b. View the configuration.
7. View the dvSwitch in Junos Space Virtual Control:
 - a. Navigate to the Virtual Control from the application chooser page.
 - b. View the dvSwitch and associated VM vNICs in the Virtual Control discovery pane.

Results

The activities that are conducted in this scenario verify the ability to create a distributed virtual switch by using dvSwitch, integrate with IBM Flex System BTO, and ensure network administrator visibility of individual VM vNICs that are nested within port groups on the dvSwitch within Junos Space Virtual Control.

4.3 Virtual machine mobility

VM mobility is a key feature in today's virtualized data center environments. VM mobility offers workload placement flexibility through planned and unplanned events, allowing workloads to migrate around the data center. Migration events can also extend beyond the data center, if the cross-site IP and FC networks are low-latency to support functions such as storage subsystem replication, IBM PowerVM Live Partition Manager (LPM), VMware Site Recovery Manager (SRM), and so on. These technologies permit either live or offline migration between data centers, providing the foundation for the smarter data center.

4.3.1 Description

Figure 4-7 depicts an example network and compute domain architecture to support VM mobility.

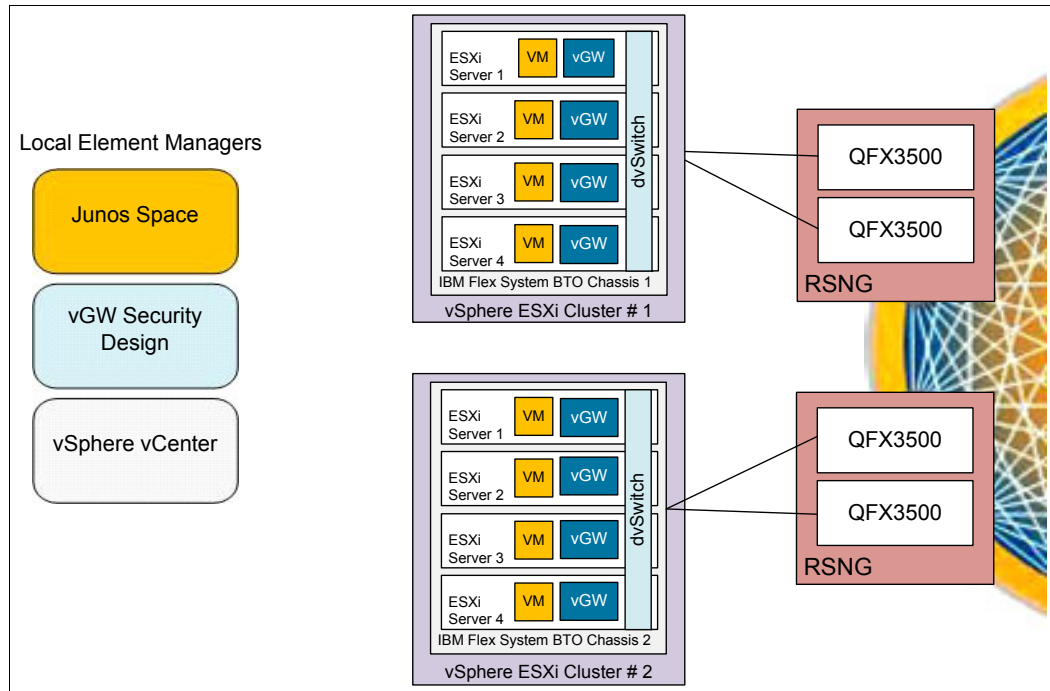


Figure 4-7 Use case 3, architecture overview diagram

The architecture contains the following components:

- ▶ One Junos Space virtual appliance that enables network administrators to manage all network devices in the data center
- ▶ One VMware vSphere vCenter that enables server administrators to manage the VMware VMs within the IBM Flex System BTO
- ▶ One distributed virtual switch built that uses VMware vSphere distributed virtual switch (dvSwitch) within each ESXi server
- ▶ One vGW Security Design VM that enables configuration of SRX zones for VM level security policy enforcement
- ▶ Four QFX3500 switches that is configured into two redundant server node group.
- ▶ Each redundant server node group supports redundant uplink connections to the IBM Flex System BTO I/O modules
- ▶ Two IBM Flex System BTO, each containing four compute nodes, one VMware vSphere cluster and four vGW virtual machines (one per ESXi server)

4.3.2 Requirements

This scenario includes the functional and non-functional requirements pertinent to this specific solution:

- ▶ Functional requirements:
 - Minimal disruption
 - Scalable
- ▶ Non-functional requirements:
 - Optimize routing
 - Automatically load balanced
 - Mitigate outages caused by physical server maintenance

4.3.3 Proof of technology

The PoT executes the following scenarios:

- ▶ Scenario 1 ensures vGW¹ awareness during VM vMotion from one ESXi server to another.

This scenario provides essential information about the vGW advantage feature of assuring security policies are intact during an vMotion event.
- ▶ Scenario 2 validates vMotion of VMs between ESXi servers, within two separate IBM Flex System BTO environments within the same VMware vSphere cluster.

This scenario validates key IBM Flex System BTO and QFabric attributes in support of VM mobility. This scenario also validates an operational model that uses Junos Space Virtual Control.

Validation scenario 1

This scenario verifies vGW tier functions. A post-vMotion event validates that VM Security policies are intact on a migrated node.

Terms of reference

The PoT achieves the following criteria:

- ▶ Perform Security Policy creation at vGW Security Design.
- ▶ Validate vGW security policy enforcement prior to vMotion event.
- ▶ Validate vGW security policy enforcement after to vMotion event.

Entry Criteria

For this scenario, security policies are already defined in vGW Security Design.

¹ vGW is an optional security add-on. It is not required for VM mobility.

Interaction map

Figure 4-8 depicts the interaction map for the technology level steps, validating the terms of reference for this validation scenario.

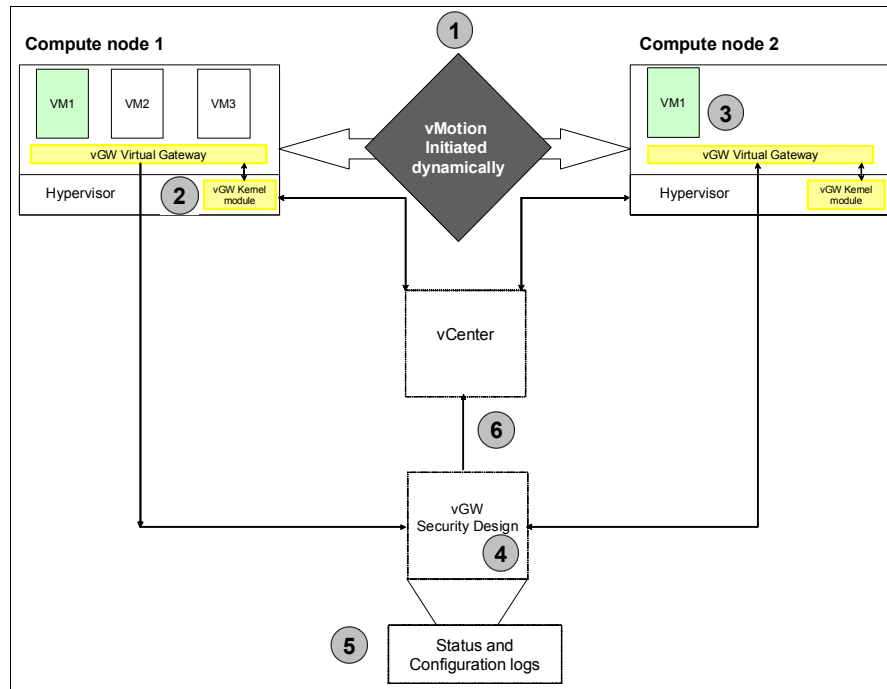


Figure 4-8 vGW ensuring VM security policies during vMotion event

The following steps describe this scenario:

1. vMotion is initiated dynamically from one compute node to compute node two using port TCP 8000.
2. The vGW hypervisor kernel module processes the inspection as follows:
 - a. vGW hypervisor kernel module passes the status to the vGW Virtual Gateway.
 - b. vGW Virtual Gateway records the status in a state table.
 - c. vGW Virtual Gateway uses the state table for the connections and places the appropriate restrictions on the inbound and outbound traffic for the VM.
 - d. vGW Virtual Gateway maintains the open connections and security throughout the event. Also vGW virtual gateway facilitates the communication between the vGW hypervisor kernel module and the vGW Security Design server.
3. vMotion completes at Compute node 2 and change to vGW takes place as follows:
 - a. The vGW hypervisor kernel module detects the newly deployed VM.
 - b. The vGW hypervisor kernel module notifies the vGW Virtual Gateway.
 - c. vGW Virtual Gateway communicates with vGW Security Design server.
4. vGW Security Design server identifies the respective policies for migrated VM.
 - a. vGW Security Design server pushes the respective policies for vMotion VM such as Global, Group, or local VM policies.
 - b. vGW Virtual Gateway receives the security policies for the newly migrated VM.
 - c. vGW Virtual Gateway applies the policies to VM through the vGW hypervisor kernel module.

5. Logs are generated on the vGW Security Design server. You can navigate to the Status and Configuration tab on the vGW Security Design platform to view the logs. Observe the table that lists all of the vGW Virtual Gateway that is associated with migration. Identify the logs for recently moved VM. The log screen is refreshed every 60 seconds.
6. The vGW Security Design server stays in constant communication with the VMware vCenter and synchronizes itself with changes that occur on hosted VMs.

Results

This scenario shows that the vGW Series ensures that appropriate security levels for a VM remain intact throughout its migration. It enables an automated migration of security policies. With this type of scenario, you can avoid compromising or lowering the security levels while migrating VMs.

Validation scenario 2

Some common daily activities in today's data centers can be challenging if the virtualized infrastructures are not flexible to support scalability, either vertical or horizontal, for example, when migrating VMs from physical server to physical server to avoid resource overload. This scenario validates the capability to manually migrate a VM from one ESXi server within one IBM Flex System BTO to one ESXi server within another through IBM Flex System Manager VMcontrol and Junos Space Virtual Control. The Flex Systems reside in a QFabric network.

Manually executed vMotion: IBM Flex System Manager VMcontrol and vSphere vCenter provide manual and policy-based automated vMotion capabilities by using either hardware failure or resource utilization triggers. For simplicity, this procedure is demonstrated by using manually executed vMotion.

Terms of reference

The PoT demonstrates that QFabric and Junos Space Virtual Control support for VM mobility.

Interaction map

Figure 4-9 depicts the interaction map for the steps to validate the terms of reference for this scenario.

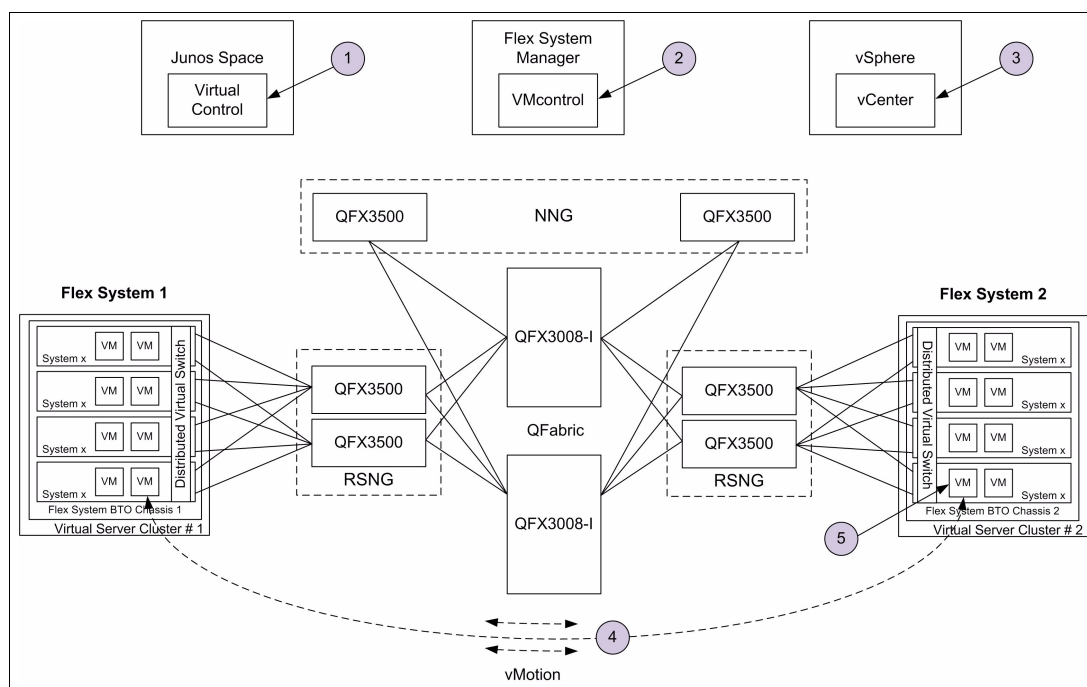


Figure 4-9 VM mobility in a QFabric enabled data center

The following steps describe this scenario:

1. The network administrator ensures that the virtual networks are symmetrical between source and target ESXi servers through Junos Space Virtual Control:
 - a. Click **Virtual Control** on the Junos Space Network Application Platform landing page.
 - b. Select **Virtual Control** from the application switcher.
 - c. Select **vNetworks** → **Manage Hosts**, within virtual control task ribbon, and view the ESXi server inventory.
 - d. Select **Actions** → **View Inventory** to display vSwitch and vNIC configurations.
2. The server administrator initiates a vMotion from one ESXi server to another through IBM Flex System Manager VMcontrol:
 - a. From the Virtual Servers View, right-click the ESXi server that contains the VM to be relocated, and then select **Relocate**.
 - b. In the Relocate menu, select the destination ESXi server and then **Execute**.
3. IBM Flex System Manager VMcontrol communicates with vSphere vCenter to invoke vMotion, which moves the VM from the source to the target ESXi server.
4. vMotion executes relocation from the source to the target ESXi server:
 - a. VM state is encapsulated and stored on shared storage in the form of configuration files.
 - b. The VM memory contents and operating state are transferred at the highest priority through the vMotion network from source to target ESXi servers. Changes in data that is already transferred is tracked in a bit-map for a delta copy to be initiated after the primary copy is complete. The VM is suspended while the delta copy runs.

- c. Junos Space Virtual Control ensures that the proper network profiles are applied and the virtual and physical network infrastructure are in sync.
 - d. The VM network configuration (that is, identity and connections) is preserved.
 - e. The VM is activated on the target ESXi server, which includes its original MAC address.
 - f. VMotion pings the network router to ensure that it is aware of the new MAC address.
5. VM executes on the target ESXi server with no disruption to online operations.

Results

The flattened, fabric-based architecture of Juniper QFabric provides a high-speed, non-blocking and scalable network that supports VM mobility (planned or unplanned). This scenario outlines vMotion events and proves Junos Space Virtual Control, IBM Flex System Manager VMcontrol, and QFabric is up to the task.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- ▶ *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493
- ▶ *Build a Smarter Data Center with Juniper Networks QFabric*, REDP-4830
- ▶ *IBM PureFlex System and IBM Flex System Products and Technology*, SG24-7984

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- ▶ IBM ServerProven website and supported enterprise chassis components:
<http://ibm.com/systems/info/x86servers/serverproven/compat/us/flexsystems.html>
- ▶ IBM Redbooks Product Guides for IBM Flex System:
<http://www.redbooks.ibm.com/portals/puresystems>
- ▶ IBM Distributed Virtual Switch (DVS) 5000V:
 - <http://www.ibm.com/systems/networking/switches/virtual/dvs5000v/index.html>
 - <http://www-304.ibm.com/support/docview.wss?uid=isg3T7000509&aid=1>
- ▶ IBM System Storage Interoperation Center (SSIC):
<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>
- ▶ Juniper Networks QFabric components:
<http://www.juniper.net/us/en/products-services/switching/qfx-series/qfabric-system/>
- ▶ Juniper MX Series models and their specifications:
<http://www.juniper.net/us/en/products-services/routing/mx-series/>
- ▶ SRX Service Gateway Series:
<http://www.juniper.net/us/en/products-services/security/srx-series/>
- ▶ vGW series details:
<http://www.juniper.net/us/en/products-services/software/security/vgw-series/>

- ▶ IBM Flex System Manager:
<http://www.ibm.com/systems/flex/systems-management/index.html>
- ▶ Junos Space:
<http://www.juniper.net/us/en/products-services/network-management/junos-space-platform/#overview>
- ▶ IBM Tivoli suite of products:
<http://www.ibm.biz/Bdx2Bg>
- ▶ STRM Series devices:
<http://www.juniper.net/us/en/products-services/security/strm-series/>
- ▶ IBM QRadar SIEM:
<http://q1labs.com/products/qradar-siem.aspx>
- ▶ vGW configuration options:
<http://www.juniper.net/techpubs/hardware/vgw-series/5.0/vgw-install-admin.pdf>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



Building a Smarter Data Center with IBM Flex System and Juniper Networks QFabric



Address increased data center complexity and infrastructure constraints

Discover the key building blocks in a smarter data center infrastructure

Get design guidance for a smarter data center

Data centers must become smarter to meet today's business needs. They need to be more efficient, scalable, and flexible and at the same time keep operational costs in check. A smarter data center must seamlessly integrate IT resources, such as servers, storage, and networking, while also responding quickly to change.

Networking plays an essential role in enabling infrastructures for smarter data centers. In dynamic environments with virtualized IT resources, the network must do more than just carry traffic and support the provisioning of new IT services. It must also have the built-in flexibility and capability to adapt quickly while maintaining comprehensive security, visibility, and management.

IBM Flex System build-to-order (BTO) and Juniper Networks QFabric are key building blocks for a smarter data center infrastructure. They are reliable, resilient, and energy efficient resources that seamlessly integrate to provide the capabilities and flexibility needed now and in the future.

This IBM Redpaper publication discusses how to build a smarter data center infrastructure with IBM Flex System BTO and Juniper Networks QFabric. It discusses key client use cases that address today's data center challenges:

- ▶ Business continuity and disaster recovery
- ▶ Multitenancy
- ▶ Virtual machine mobility

This paper is intended for IT management, IT architects, network planners and integrators, and technical specialists.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks